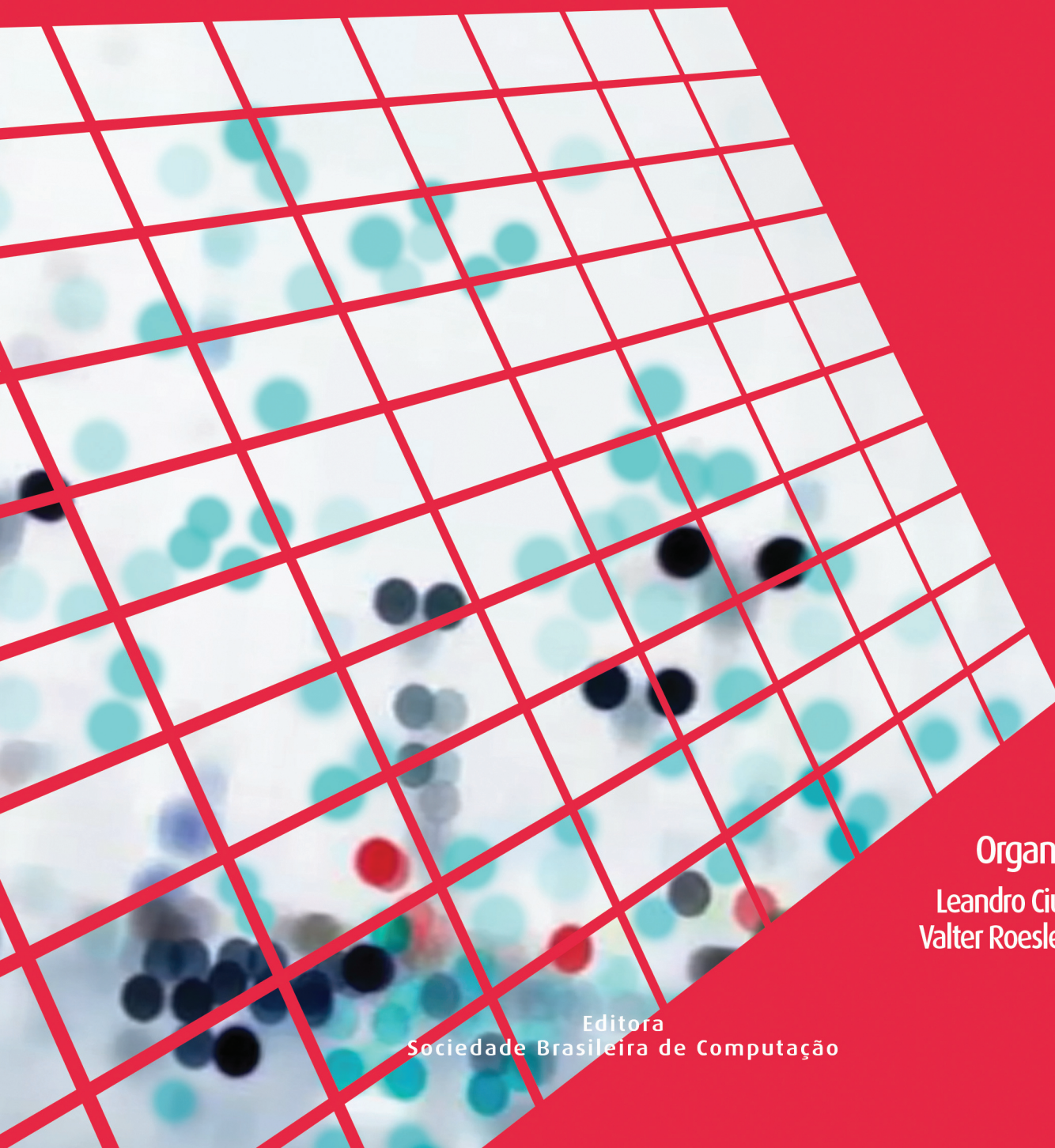


WebMedia2017
XXIII Simpósio Brasileiro de Sistemas
Multimídia e Web

IV Workshop do CT-Vídeo
Comitê Técnico de Prospecção Tecnológica
em Videocolaboração

O futuro da videocolaboração: perspectivas



Organizadores
Leandro Ciuffo (RNP)
Valter Roesler (UFRGS)

Editora
Sociedade Brasileira de Computação



IV Workshop do CT-Vídeo
Comitê Técnico de Prospecção Tecnológica em Videocolaboração



Organização



MINISTÉRIO DA
DEFESA

MINISTÉRIO DA
CULTURA



MINISTÉRIO DA
SAÚDE



MINISTÉRIO DA
EDUCAÇÃO

MINISTÉRIO DA
**CIÊNCIA, TECNOLOGIA,
INOVAÇÕES E COMUNICAÇÕES**



Patrocinadores do WebMedia



FICHA CATALOGRÁFICA

Simpósio Brasileiro de Sistemas Multimídia e Web: Workshop do CT-Vídeo – Comitê Técnico de Prospecção Tecnológica em Videocolaboração (23º: 2017: Gramado, RS).

Anais do XXIII Simpósio Brasileiro de Sistemas Multimídia e Web: Workshop do CT-Vídeo (Comitê Técnico de Prospecção Tecnológica em Videocolaboração), 19 de outubro, 2017, Gramado, Rio Grande do Sul.

368p.

E-book e Impresso.

ISBN: 978-85-7669-381-9.

Evento promovido pela Sociedade Brasileira de Computação (SBC), Porto Alegre, RS. Organização: UFRGS.

1. Multimídia. 2. Telemedicina. 3. Videocolaboração. 4. Visões de futuro.
5. Cinema. 6. Dança. 7. Rede de computadores. I. Título

IV Workshop do CT-Vídeo (Comitê Técnico de Prospecção Tecnológica em Videocolaboração)

19 de outubro de 2017. Gramado, Rio Grande do Sul, Brasil.

Evento paralelo ao 23º Simpósio Brasileiro de Sistemas Multimídia e Web (Webmedia 2017).

Organização

RNP – Rede Nacional de Ensino e Pesquisa

UFRGS – Universidade Federal do Rio Grande do Sul

Coordenação Geral

Leandro Neumann Ciuffo (RNP)

Comitê Organizador

Leandro Neumann Ciuffo (RNP)

Valter Roesler (UFRGS)

Montagem

Rafael Thomé

Arte da Capa

Amanda Areias

Coordenação do CT-Vídeo

Valter Roesler (UFRGS)

Promoção

SBC – Sociedade Brasileira de Computação

Patrocinadores do Webmedia

NIC.br – Núcleo de Informação e Coordenação do Ponto BR

CGI.br – Comitê Gestor da Internet no Brasil

CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico

Prefácio

Estabelecido em 2014 com o apoio a Rede Nacional de Ensino e Pesquisa (RNP), o Comitê Técnico de Videocolaboração (CT-Video) visa ser um fórum aberto para estudar o futuro das aplicações de vídeo e colaboração remota, nas mais variadas áreas, como por exemplo:

- Sistemas de videoconferência e telepresença.
- Sistemas de colaboração remota.
- Cinema digital.
- Aplicações multimídia na saúde.
- Produção, edição e transmissão multimídia.
- Áudio 3D.
- Educação online.
- *Machine Learning* em aplicações de vídeo.
- *Software Defined Networking* em aplicações de vídeo.
- Sistemas de vigilância e segurança.
- Geração de vídeos em ambientes não controlados.

Qualquer pessoa pode participar do CT-Video. Seus membros são formados em sua maioria por: pesquisadores acadêmicos das áreas de multimídia, vídeo digital e redes; profissionais de empresas desenvolvedoras; usuários de serviços de videocolaboração. Para participar, basta se inscrever na lista de discussão¹ do CT-Video e participar das reuniões.

Um dos benefícios do Comitê é proporcionar a interação entre pesquisadores de diferentes instituições para facilitar a troca de conhecimento e eventual integração das aplicações desenvolvidas por cada grupo de pesquisa. Entretanto é importante destacar que o CT-Vídeo está inserido em um projeto amplo de criação de comitês técnicos como forma de ampliar o processo de prospecção tecnológica da RNP. Além do CT-Video, outros dois comitês apoiam a RNP na realização da prospecção tecnológica: o comitê técnico de Monitoramento de Redes (CT-Mon)² e o comitê técnico de Gestão de Identidade (CT-GId), sendo este último o mais antigo de todos, estabelecido em 2010.

O CT-Video realiza, em ciclos anuais, prospecção tecnológica com o auxílio de seus participantes, contribuindo para a construção de visões técnicas que orientem novos investimentos em Pesquisa e Desenvolvimento (P&D). Em geral, isso é materializado na forma de um mapeamento sistemático do estado da arte das pesquisas em andamento nos diversos laboratórios participantes do CT-Video e na elaboração de um relatório anual de Visão de Futuro.

O plano de trabalho anual do Comitê é definido autonomamente por seus membros, em reunião realizada no primeiro trimestre de cada ano. Este livro é resultado do planejamento das atividades de 2017, em que foi decidido a organização de uma chamada de trabalhos para apresentação no IV Workshop do CT-Video: O Futuro da Videocolaboração³, realizado no WebMedia 2017 em Gramado (RS). Os melhores

¹ <http://listas.rnp.br/mailman/listinfo/ct-video>

² <https://wiki.rnp.br/display/ctmon/Home>

³ http://www.inf.ufrgs.br/webmedia2017/?page_id=830&lang=pt

trabalhos aceitos para apresentação foram convertidos em capítulos de livro e compilados nesta publicação.

Laboratórios e Grupos de Pesquisa que contribuíram para este livro

Apresentados na mesma ordem de autoria dos capítulos:

O **Grupo de Pesquisa Design Audiovisual** da UFPB foi criado em 2016 e é composto por professores, doutorandos, mestrandos, graduandos e alunos de iniciação científica de diferentes cursos e universidades, Seu foco está na integração metodológica e teórica dos processos de criação, produção, desenvolvimento e fruição das áreas da Interação Humano Computador com os estudos de audiência, que tem metodologias e propostas similares, apesar de estarem em áreas do conhecimento distintas. Desenvolve pesquisas e projetos com temas variados, tais como: narrativas generativas, conteúdos pedagógicos, jornalismo 360o, visualização de dados, conteúdos para TVs públicas, rádio e interatividade.

O Grupo é autor do conceito de *Design Audiovisual*, um modelo comunicacional metodológico que reúne métodos e conceitos pertinentes a Interação Humano-Computador (IHC), das Ciências da Computação, e os Estudos de Audiência, das Ciências da Comunicação. Em resumo, o Design Audiovisual tem dois alcances práticos: 1) analisar produções audiovisuais feitas com a utilização das ferramentas disponibilizadas em conjunto pela IHC e pelos estudos de audiência; 2) prever interações e propor aplicações inovadoras dessas ferramentas no desenvolvimento de produtos audiovisuais. A principal característica deste modelo é o planejamento da interação durante o processo de produção em áudio e vídeo, com base em quatro linhas que configuram e modelam o processo criativo: o indivíduo, as motivações, a experiência e o conteúdo.

O **Laboratório de Artes Cinemáticas e Visualização (LabCine)** da Universidade Presbiteriana Mackenzie produz pesquisas históricas, teóricas, críticas e tecnológicas no campo das artes audiovisuais, a partir de colaboração com pesquisadores de tecnologias computacionais e de redes. O LabCine desenvolve e pesquisa cinemas expandidos, modos de exibição em espaços contemporâneos, arquivamento e seleção de imagens, narrativas científicas e educacionais, visualizações avançadas de bancos de dados e visualizações avançadas remotas. O LabCine foi criado para dar continuidade a pesquisas em um campo que é mais amplo que o cinema tradicional e por isso a denominação “cinemáticas”. O grupo, registrado no CNPq, tem mais de 20 participantes entre doutores, graduandos, mestrandos, doutorandos, pós-doutorando e é composto por pesquisadores de diferentes universidades como UNICAMP, UFG, UNIFESP e PUC/SP e parceiros de pesquisa da UCSD (Universidade da Califórnia de San Diego), UFPB e USP. O LabCine é um dos laboratórios conveniados com o SAGE2™ (Scalable Amplified Group Environment) da Universidade de Illinois em Chicago.

O LabCine é o resultado de pesquisas, demonstrações e produções de cinemas 4K, 8K e estereoscopia para transmissão em redes de alta velocidade que iniciaram em 2008 em

parceria com o CALIT2 (UCSD), o KMD de Keio, Japão e o FILE (Festival Internacional de Linguagem Eletrônica). A equipe do LabCine participou da coordenação dos eventos CineGrid Brasil, realizados em 2011, no Rio de Janeiro e em 2014, em São Paulo. Desde 2013, o LabCine produziu nove filmes digitais, alguns de super-alta definição, alguns em estereocópia, incluindo filmes para Moocs, filmes de registro de procedimentos médicos, filmes sobre ciência e cultura astronômica. Desenvolve no momento o software SenseMaking, o filme EstereoEnsaio SP (4K3D) e o filme-ensaio, Projeto Marte (ou Loga). Tem obtido apoio de agências de pesquisa como FINEP, FAPESP, RNP, MINC -PROAC, MackPesquisa, Capes e Fiocruz. O Laboratório de Artes Cinemáticas e Visualização é um projeto associado à pós-graduação *Stricto Sensu* (mestrado e doutorado) em Educação, Arte e História da Cultura e à Faculdade de Computação e Informática (FCI) da Universidade Presbiteriana Mackenzie. O Laboratório possui parcerias com diversas instituições de ensino e pesquisa no Brasil e no exterior.

O Grupo de Pesquisa Poéticas Tecnológicas: Corpoaudiovisual (GPPoética)⁴, fundado em 2004, tem como objetivo a investigação teórico-prática no campo da Arte Tecnologia, principalmente ações relacionadas com o “corpo contemporâneo”, seja em dança, *performance* ou outras linguagens artísticas. Desde 2005, desenvolve pesquisa no campo da telemática contanto com vários artigos, projetos artísticos e dispositivos computacionais. Em 2015, publicou na Revista eletrônica MAPAD2 - Mapa e Programa de Artes em Dança Digital⁵, diversos artigos de pesquisadores que colaboraram ao longo dessa trajetória em cada uma das etapas de uma década de investigação. O ponto de vista de músicos, coreógrafos, engenheiros da computação, mídia-artistas e estudiosos do campo está disponibilizado nessa edição de 2015, número 2, promovendo um panorama significativo sobre o uso da Rede como local de troca, de transferência de conhecimento e de processos criativos artísticos e tecnológicos. Uma nova linha de pesquisa do GP Poéticas, denominada Arte-Tecnologia-Cognição, tem sido responsável por investigar o sistema perceptivo em contextos nos quais o corpo está mediado pelas tecnologias digitais. Para avançar com os estudos e aplicações da relação homem-máquina (termo tradicional da área), é preciso explorar tanto o desenvolvimento de novas tecnologias e ferramentas, como também processos de percepção em ambientes remotos. Percebe-se que uma nova compreensão sobre “presença” deve ser investigada de acordo com as compreensões das teorias contemporâneas de Cultura Digital e das Ciências Cognitivas. Com esses estudos, pretendemos analisar os contextos e configurações já testados, bem como prospectar quais aspectos seriam precisos para o avanço de sistemas de Vídeo_Colaboração no campo das Artes.

O Laboratório do PRAV (Projetos em Áudio e Vídeo)⁶ iniciou suas atividades em 1998 e, desde 2006, está situado junto ao Instituto de Informática da Universidade Federal do Rio Grande do Sul, sendo composto atualmente por quatro áreas de especialidade:

⁴ <http://www.poeticastecnologicas.com.br>

⁵ <https://portalseer.ufba.br/index.php/mapad2/issue/view/1184>

⁶ <http://www.inf.ufrgs.br/prav>

- Desenvolvimento de sistemas para videoconferência;
- Telemedicina / Telessaúde;
- Ensino a Distância.

Com uma equipe de mais de 30 pesquisadores e colaboradores, o PRAV conquistou uma posição de referência em tecnologia no Brasil, tendo diversas parcerias com outras Universidades, grupos de pesquisa e empresas.

O laboratório do PRAV foi berço de duas *spin-offs*. A empresa I9Access Tecnologia Ltda trabalha na área de Sistemas de Telessaúde (Sala Cirúrgica Integrada, Sistemas de simulação realística assistida por vídeo, Sistemas de monitoramento de doenças crônicas, tele-ECG, entre outros). A empresa Mconf Tecnologia Ltda tem seu foco no sistema de webconferência Mconf, voltado para reuniões remotas e ensino a distância.

Nosso foco é na melhoria da qualidade de vida do cidadão através do desenvolvimento tecnológico na área de áudio e vídeo. Diversos produtos úteis para a sociedade surgiram dentro do laboratório. Exemplos são o sistema de webconferência Mconf, a Sala Cirúrgica Integrada, e o sistema Interativo de Áudio e Vídeo, entre outros. Atualmente o grupo está desenvolvendo um MCU em software e um sistema de teleoftalmologia, entre outros projetos.

O Núcleo Técnico-Científico (NTC) Telessaúde Brasil Redes Unifesp é uma parceria entre a Universidade Federal de São Paulo (Unifesp), por meio de sua Secretaria de Educação a Distância (SEAD) e o Ministério da Saúde. A parceria é firmada por meio de convênio, em regime de cooperação técnica, administrativa e científica, com o objetivo de manter o Núcleo na Unifesp atuante e com cobertura no Estado de São Paulo, prioritariamente nos municípios de difícil acesso e áreas remotas de maior vulnerabilidade, para melhoria da qualidade da assistência na Atenção Básica no Sistema Único de Saúde, utilizando diferentes estratégias educacionais, formativas e de suporte. O NTC provê acesso a mais de 1000 equipes de Saúde da Família, com atividades que vão desde teleassistências e telediagnósticos por meio de recursos digitais avançados (vídeo ou webconferências síncronas e assíncronas), telediagnósticos com o uso de aplicações avançadas de visualização remota, ações de formação educacional em APS (formação permanente dos profissionais de saúde) e o desenvolvimento de Segundas Opiniões Formativas (SOF's) para apoio à pesquisa e educação permanente a partir da publicação nas bibliotecas virtuais de saúde (BVS's) e bases indexadas nacionais e internacionais.

Sob a coordenação do **Núcleo de Estudos de Tecnologias Avançadas (NETAv)** da Escola de Engenharia da UFF, a iniciativa conta com o apoio do Hospital Universitário Antonio Pedro, da Faculdade de Medicina e da Pro-reitoria de Extensão, que cedeu o espaço onde foram instalados o Centro de Saúde Holográfico e o Consultório de Saúde Virtual. Assim, além de professores, técnicos administrativos, alunos de graduação e de pós-graduação, médicos experientes, internos e residentes de diversas especialidades e outros profissionais de saúde, a equipe conta com de profissionais do mercado, em áreas

como o design de interiores e sistemas multimídia. Parceiros da iniciativa privada também tiveram um papel fundamental na montagem das instalações, com destaques para a Embratel/Star One e a EyeMotion, empresa líder no setor de eventos, detentora da tecnologia holográfica utilizada. O telessaúde é, portanto, um exemplo bem-sucedido de aplicação da “tríplice hélice”, onde governo, academia e o setor privado somam esforços para alcançar o desenvolvimento tecnológico sustentável, promotor de bem-estar social.

O **TeleMídia** é um laboratório de pesquisa afiliado ao Departamento de Informática da PUC-Rio, dedicado à pesquisa e inovação em sistemas hipermídia e multimídia distribuídos. O TeleMídia foi fundado em 1991 pelo Prof. Luiz Fernando Gomes Soares, e é atualmente liderado pelo Prof. Sérgio Colcher. Além da sua vasta contribuição em pesquisas de base e em formação de recursos humanos, nos últimos anos, o laboratório também se destacou com importantes contribuições relacionadas à padronização em sistema de TV digital terrestre e IPTV. Entre as mais conhecidas dessas contribuições estão a linguagem NCL (*Nested Context Language*) e o middleware Ginga, ambos padrões para serviços interativos no Sistema Brasileiro de TV Digital Terrestre e recomendações da União Internacional das Telecomunicações para IPTV e sistemas híbridos *broadcast/broadband*.

O **Laboratório de Aplicações e Inovação em Computação (LApIC)**⁷ do Departamento de Ciência da Computação da Universidade Federal de Juiz de Fora (UFJF) foi criado em 2009 para concentrar ações que visam o desenvolvimento de aplicações avançadas, mais especificamente nas áreas de Web e multimídia, incluindo as várias modalidades de vídeo interativo. O grupo vem experimentando uma expansão continuada e sustentável, apoiada por diversos projetos de pesquisa e sua inserção no Programa de Pós-Graduação em Ciência da Computação da UFJF. Atualmente, suas pesquisas abrangem diversos tópicos na área de Sistemas Multimídia/Hipermídia, tais como:

- Infraestrutura de telecomunicações multimídia, incluindo protocolos, Qualidade de Serviço e nuvem multimídia
- Análise de conteúdo, incluindo indexação e recuperação multimídia
- Multimídia aplicada à educação, incluindo modalidades presencial e à distância
- Arquiteturas e serviços de vídeo interativo, incluindo Internet TV, IPTV, TVD terrestre, sistemas integrados broadcast-broadband e vídeo sob demanda
- Arquitetura de sistemas, modelagem e autoria do conteúdo hipermídia

O LApIC conta com uma equipe de pesquisadores de perfil diversificado e que se guia pela constante colaboração, promovendo alto grau de sinergia entre os projetos e entre seus resultados. Desde sua criação, o grupo vem participando de iniciativas conjuntas de pesquisa, desenvolvimento e transferência de tecnologia junto a órgãos nacionais e

⁷ <http://www.ufjf.br/lapic>

internacionais, que incluem o Fórum do Sistema Brasileiro de TV Digital e a União Internacional de Telecomunicações.

O Video Technology Research Group (ViTech) é um grupo de pesquisa criado em 2017 como uma spin-off do Grupo de Arquiteturas e Circuitos Integrados (GACI) da Universidade Federal de Pelotas (UFPel). O ViTech é formado por cinco professores/pesquisadores com ampla experiência e interesse em tecnologias de vídeo, incluindo técnicas e algoritmos de processamento de sinal, redução e controle de complexidade em codificação de vídeo, entrega e transporte de multimídia em redes, sistemas embarcados e projeto de hardware. Atualmente, o grupo conta com 9 alunos de doutorado, 7 alunos de mestrado e 12 bolsistas de iniciação científica. Os principais interesses de pesquisa do grupo se estendem desde o nível algorítmico até a implementação de sistemas de tempo real, com foco em processamento de vídeo 2D, multivista e 3D, padrões de codificação de vídeo, HDR, nuvens de pontos e light fields. Apesar de ser um grupo recentemente criado, os pesquisadores que lideram o ViTech atuam na comunidade científica há mais de uma década e são responsáveis pela publicação de mais de 300 artigos em eventos internacionais, dezenas de periódicos, livros, patentes e softwares registrados.

O Grupo de Multimedia e Processamento de Sinal do Pólo de Coimbra do Instituto de Telecomunicações (Portugal) desenvolve actividades de pesquisa, formação e consultoria especializada nas áreas relacionadas com o processamento digital de sinais. Os seus pesquisadores estão envolvidos de forma ativa na resolução de problemas científicos focados no processamento de vídeo, processamento de imagem, processamento da fala, processamento de sinais de telecomunicações e implementação de sistemas de processamento usando hardware programável, processadores multicore e processadores gráficos. Atualmente, conta com seis investigadores permanentes e sete investigadores auxiliares, contando ainda com mais de trinta alunos de pós-graduação e pesquisadores contratados.

O grupo de pesquisa em gerenciamento e transmissão de vídeos surgiu em 2005 como um subgrupo de Laboratório de Arquitetura e Redes de Computadores (LARC) do Departamento de Engenharia de Computação e Sistemas Digitais da Escola Politécnica da Universidade de São Paulo. Sob a coordenação da Profa. Regina Melo Silveira, o grupo desenvolveu projetos que trouxeram contribuições relevantes para a área de rede, e que permitiram a implementação de serviços que têm beneficiado a comunidade acadêmica, como os projetos financiados pela RNP: GT-GV (Gerência de Vídeo) (2006-2007), GT-Overlay (2008-2009), GT-IMAV (Instrumentação e monitoração para aplicações de vídeo) (2011-2012). Estes projetos resultaram na implantação dos serviços video@rnp e IPTV USP, que estão em operação desde 2007. Como resultado da necessidade contínua de evolução destes serviços, foi criada a spin-off TVoD (TV on

Demand Consultoria e Soluções) fundada por alguns integrantes deste grupo. Esta empresa teve grande importância na evolução destes serviços ao longo dos últimos anos, imprimindo inovação e integrando tecnologias que propiciam o melhor desempenho da transmissão do acervo de vídeo de instituições de ensino e pesquisa. Importantes serviços de distribuição de vídeo com escopo nacional, e suas evoluções, como o vídeo@rnp, o IPTV USP, o e-Aulas USP e o sistema da TV Escola são de responsabilidade da TVoD. Atualmente a empresa mantém forte parceria com o LARC para o desenvolvimento de projetos de pesquisa, como o projeto “Sistema para Rede de Distribuição de Conteúdos considerando tecnologia SDN”, projeto PIPE financiado pela FAPESP, com desenvolvimento tecnológico de ponta alinhado a pesquisas que se encontram no estado da arte da área de redes de computadores.

Sumário

Prefácio	05
-----------------------	----

Videocolaboração no Cinema e nas Artes

Capítulo 1: Arquitetura de design colaborativo para a imersão temporal e espacial em vídeos de altíssimas resoluções e HFR.....	13
Capítulo 2: Cinema em Definição.....	54
Capítulo 3: Presenças do nosso futuro. O avanço dos sistemas de videocolaboração para uma Dança Expandida.....	95

Videocolaboração na Medicina

Capítulo 4: Tecnologias emergentes em telemedicina. Ênfase em videocolaboração.....	118
Capítulo 5: Análise dos processos videocolaborativos do Núcleo Telessaúde Brasil Redes Unifesp na Atenção Básica.....	149
Capítulo 6: Interiorização da medicina utilizando um sistema de telepresença holográfico.....	171

Sistemas Multimídia e TV Digital

Capítulo 7: Projetos atuais e visão de futuro do laboratório TeleMídia/PUC-Rio em videocolaboração.....	206
Capítulo 8: Uma abordagem abrangente para a colaboração multimídia.....	231

Codificação de vídeo e rede de computadores

Capítulo 9: Aplicações de Aprendizado de Máquina na Codificação de Vídeo HEVC.....	267
Capítulo 10: Distribuição de Vídeo na era da Infraestrutura Definida por Software....	295

Padrões de videocolaboração

Capítulo 11: Mapeamento de soluções de videocolaboração da RNP.....	337
---	-----

Capítulo

1

Arquitetura de design colaborativo para imersão temporal e espacial em vídeos de altíssimas resoluções e HFR

Rafael Toscano^{1,2}, Valdecir Becker^{1,2}, Laíz Ferreira²,
Samara Coutinho², Leonardo Burgos^{2,3}

¹ Universidade Federal da Paraíba, Programa de Pós-Graduação com Computação,
Comunicação e Artes (PPGCA)

² Grupo de Pesquisa Design Audiovisual (GPDA)

³ Universidade Federal da Paraíba, Programa de Pós-Graduação em Jornalismo (PPJ)

rafaeltoscano@lavid.ufpb.br, valdecir@ci.ufpb.br,
laizfederal@hotmail.com, samara.coutinho@lavid.ufpb.br,
leonardomattosburgos@gmail.com

Abstract

The narrative of the moving image from its earliest manifestations recognizes time and space as essential elements in the understanding of the films. Over the years, aesthetic innovations have been added to audiovisual products, however, the handling of time and space remain predominantly determined by the directors. This is due to the choice of narratives and technological constraints created in analog environment and pass through products and markets that, instead of incorporating elements of digital media into the creation process, only updated their devices. In order to understand the impacts of this aspects and the collaboration between users and systems, this paper studies strategies that assign new meanings to time and space in the digital universe: temporal opportunities, exploratory field, historicity and generability. In terms of video collaboration, the interactions can be parallel or sequential, decentralized and realtime (from two to even thousands of users). Linked to the survey of strategies was developed an experimental case study with eight technical productions, which allow to identify, in addition to the narrative aspects, technical demands of image processing and data traffic in video sharing networks, data mining systems, recommendation and content generation. Finally, it is perceived that time and space act as interface of the process of

fruition in the interactive media, that is, they have the function of mediating the narrative understanding through the affordances of the system with the purpose of attending the most varied roles of interaction, be it audience, synthesizer, modifier or producer. In this way, the interaction and the immersion can present different characteristics, with varied structural demands of video processing and network traffic, since the process of narrative enjoyment is increasingly interconnected to the use of complex computational systems.

Resumo

A narrativa da imagem em movimento desde suas primeiras manifestações reconhece o tempo e espaço como elementos essenciais na compreensão dos filmes. Ao decorrer dos anos, inovações estéticas foram acrescentadas ao audiovisual, contudo, o manuseio do tempo e espaço permanecem predominantemente determinados pelos diretores. Isso se deve à escolhas das narrativas e à restrições tecnológicas que vem do ambiente analógico e perpassam produtos e mercados que ao invés de incorporar elementos da mídia digital ao processo de criação, apenas atualizaram seus dispositivos. Com o objetivo de compreender quais os impactos do compartilhamento dessa escolha e da colaboração entre usuários e sistemas, este trabalho faz um estudo das estratégias que atribuem novos significados ao tempo-espaço no universo digital: oportunidades temporais, campo exploratório, historicidade e geratividade. Em termos de videocolaboração, as interações podem ser paralelas ou sequenciais, descentralizadas e em conjunto (a partir de dois a até milhares de usuários). Arelado ao levantamento das estratégias foi desenvolvido um estudo de caso experimental com oito produções técnicas, que permitem identificar, além dos aspectos narrativos, demandas técnicas de processamento de imagens e tráfego de dados nas redes de compartilhamento dos vídeos, sistemas de mineração de dados, recomendação e geração de conteúdo. Por fim, percebe-se que o tempo e espaço atuam como interface do processo de fruição nas mídias interativas, isto é, ocupam a função de mediar a compreensão narrativa por meio de affordances do sistema com o objetivo de atender os mais variados papéis de interação, seja audiência, sintetizador, modificador e produtor. Dessa forma, a interação e a imersão podem apresentar diferentes características, com variadas demandas estruturais de processamento do vídeo e tráfego da rede, já que o processo de fruição narrativa se apresenta cada vez mais interligado à utilização de sistemas computacionais complexos.

1.1 Introdução

Desde a concepção inicial de cinema, que surge como apropriação de um dispositivo e depois engloba o elemento narrativo, diversos recursos tecnológicos e comunicacionais passaram a ocupar espaços de transmissão de informação e contação de histórias. A evolução técnica e estética empregada posteriormente à narrativa cinematográfica se estruturou ao longo dos anos como um acervo de códigos comuns, de modo a conceber a noção de linguagem (BLOCK, 2010). Dentre os itens que compõem essa matriz linguística, dois elementos podem ser destacados: o tempo e o espaço.

Tradicionalmente as histórias cinematográficas são contadas por diretores e produtores usando recursos técnicos e estéticos baseados na sala escura, isto é, a construção de uma supervalorização da imagem que privilegia a ação passiva do público. Costa (2003) acredita que esse tipo de cenário sofre constantes alterações porque quem produz o conteúdo é invariavelmente influenciado pelas possibilidades que a tecnologia oferece. Ou seja, as modificações técnicas nos dispositivos refletem na forma como os produtores irão encarar os conteúdos, realizando novas experiências estéticas.

O estado da arte da narrativa audiovisual contemporânea, majoritariamente digital, mostra uma condição de estreitamento dos formatos cinema, TV, vídeo, internet e infográfico. De modo geral, é possível identificar processos de migração dos dispositivos e/ou das interfaces que buscam entender o cenário de exposição do conteúdo; das imagens que englobam cada vez mais dados e, principalmente; dos indivíduos que assumem novos papéis de interação diante da narrativa. Podemos citar a inclusão de novos recursos, como: sistemas computacionais que moldam os produtos à experiência prévia do espectador, mídias pervasivas e ubíquas, altíssimas resoluções (4K, 8k), *High Frame Rates* (HFR, taxas maiores que 48 quadros por segundo), ambientes imersivos em 360 graus, são meios que constroem um cenário em que as barreiras entre as mídias perdem seus contornos e tornaram-se permeáveis. Dessa forma, torna-se pertinente analisar como os novos recursos de reprodução e interação, baseados em uma postura ativa das pessoas, afetam a extensão do universo narrativo quando justapostos, e conseqüentemente, quais são os impactos nas infraestruturas tecnológicas que suportam a troca de informações entre sistemas e usuários.

Analisando a vasta literatura sobre a interação entre indivíduo e mídia audiovisual (CANNITO, 2010; GOSCIOLA, 2008; MACIEL, 2009; MANOVICH, 2001; MONTEZ; BECKER, 2005; MURRAY, 2003; NOGUEIRA, 2010; PERCY, 2011; PRIMO, 2000; URICCHIO, 2012) e dos estudos sobre tempo e espaço fílmico (ALVARENGA, 2011; AUMONT, 1993; COSTA, 2003; DELEUZE, 2005; LÉVY, 1996; TARKOVSKI, 1998), percebe-se que a utilização do tempo e espaço na narrativa é compreendida como possibilidade de criação estética do autor para um consumo passivo, projetivo e até mesmo reativo. Fazendo um comparativo entre o estado da arte das narrativas digitais e a literatura, percebeu-se a importância de compreender quais os impactos da interação temporal e espacial pelo usuário nas narrativas audiovisuais digitais, para então projetar impactos nas infraestruturas de transmissão e troca de informações.

Com base nesse cenário de inquietações, o presente trabalho parte da hipótese de que o tempo e o espaço atuam como interfaces de fruição em sistemas audiovisuais digitais. Ou seja, o manuseio temporal e espacial ofertado em sistemas computacionais desencadeia processos que potencializam o ato da compreensão narrativa. Dessa forma, novos requisitos tecnológicos tornam-se representativos, impactando a relação entre produtor e consumidor de conteúdo. Do ponto de vista da fruição, novos papéis surgem, gerando novos modelos de interação e demandas de infraestrutura.

Dessa forma, este capítulo apresenta um estudo sobre as estratégias conceituais e técnicas de atribuição do papel de manipulação temporal e espacial para as pessoas, sua relação com as mídias digitais e os processos de convergência de formatos estéticos e modos de fruição. Além disso, este estudo baliza as análises de evolução das narrativas

audiovisuais que, por sua vez, fundamentam a análise da visão de futuro por meio de dois cenários.

1.2. A imagem movimento

Apesar de relatos de experimentos mais antigos, a narrativa de imagens em movimento se consolidada no cinema, com a premissa do recorte visual da câmera e anos depois com o sequenciamento de instantes por meio do processo de montagem. O cineasta russo Dziga Vertov introduziu questões como fusões (*dissolve*), fracionamento de imagens (*split screen*), imagens congeladas (*freeze frames*), câmera lenta (*slow motion*) e acelerações (*fast motion*). Já os representantes da escola soviética dos anos 1920, principalmente Pudovkin e Kuleshov, puseram-se a experimentar e teorizar sobre uma montagem construtiva, uma vez que essa junção de imagens produz significados que não estão em cada plano, mas nas relações entre aqueles que foram estabelecidos pelo diretor-montador (COSTA, 2003).

O poder construtivo da sequência para Kuleshov é tão expressivo que o cineasta atribui ao processo de montagem a capacidade de construir uma geografia criativa, ou seja, com a justaposição de cenas mesmo que de fragmentos de lugares diferentes. Em detrimento da contribuição dos cineastas apresentados, este processo passa a ocupar um lugar de destaque no fazer cinematográfico.

Esse lugar estabelecido pela montagem é marcado pela dualidade de escolas estéticas. A escola americana defende o ocultamento (fantasmagoria) dos dispositivos técnicos e um discurso linear que promove o encantamento visual, numa espécie de culto à imagem. Enquanto isso, a escola soviética defendia a coalização como proposta visual, isto é, ideias, choques e rupturas com o imaginário que permitem lembrar ao público que ele está assistindo um filme e não contemplando a vida real. Essa discussão perpassa os anos e constrói um debate sobre a relação entre transparência do fazer cinematográfico e uma hiperconsciência das funções e métodos do cinema (GOSCIOLA, 2008).

Assim como os dispositivos técnicos passaram a delimitar novos espaços e sentidos à imagem em movimento, o amadurecimento sobre o papel da audiência, mesmo que limitado, também alcança o fazer cinematográfico. Para Aumont (1993) o público pode ser considerado um parceiro da obra já que o projeto desenvolvido pelo diretor constrói seu sentido ao ser projetado e percebido pelos aspectos emocionais e cognitivos das pessoas. Sendo assim, a imagem constrói o espectador e o espectador constrói a imagem.

Se por um lado cabe ao cineasta o papel de criar a obra, do outro lado cabe às pessoas a ação interpretativa e projetiva. Essa segunda ação ocorre em dois modos: 1) no sentido de acrescentar elementos ou ideias; 2) restituir as partes omissas, ocultas de objetos e situações representadas. Com o intuito de garantir uma coerência da recepção da mensagem, o diretor, ou autor da obra, assume o papel de estabelecer os valores principais do universo narrativo de modo a evitar projeções e interpretações incongruentes por parte do público. Como estratégia de construção dessa forma de percepção da obra cinematográfica, dois elementos foram bastante utilizados: tempo e espaço.

1.2.1. Compreender o tempo

É possível perceber o tempo como a condição intrínseca do cinema que reflete a experiência discursiva e afetiva dos filmes (NOGUEIRA, 2010). Aumont (1993) compreende essa relação entre tempo e imagem por meio de duas categorias: 1) Não temporalizadas ou estáticas; 2) Temporalizadas. A imagem temporalizada altera os dados de representação do instante em pelo menos dois modos: primeiro, ao permitir a representação de um acontecimento sem ter que recorrer à síntese arbitrária de apenas um de seus instantes, como a fotografia; segundo, ao multiplicar os instantes, formando uma sequência de blocos temporais.

Com base no estudo de Aumont (1993), é possível identificar que existe um tempo que pertence à imagem e aquele que é do espectador. Outros autores atribuem a essa relação os termos diegético/extradiegético, explícito/implícito, fílmico, narrativo e físico. Porém, essas classificações têm em comum a definição da multiplicidade de tempo que envolve a fruição de uma obra.

O conceito de tempo também pode ser utilizado tecnicamente para atingir funções estritamente cognitivas e revelar aspectos da ação que são tênues à percepção do olho como, por exemplo, pequenas mudanças que apenas o *slow motion* permite notar ou grandes mudanças que só o *fast motion* permite evidenciar, bem como saltos de tempo por elipses (NOGUEIRA, 2010). Dessa forma, as variações de sentido da percepção do tempo, sejam em aspectos técnicos como *frame rate*, captação e reprodução, ou em discussões estéticas, como montagem e ritmo, constroem a dimensão temporal da obra audiovisual.

O cineasta tradicionalmente atribui a responsabilidade do fluxo de tempo para a densidade e ritmo de sequenciamento dos fotogramas, como um artifício de conduzir o olhar do público e construir a instância narrativa do autor. Dessa forma, o sentido de fruição do filme é guiado através da quantidade de imagem em um enquadramento, bem como a amarração entre esses quadros ao longo de uma sequência.

1.2.2. Compreender o espaço

Desde os primeiros experimentos da cinematografia, a composição do espaço é destacada por produtores como recurso narrativo. Inicialmente, o cinema ofertava uma percepção semelhante ao ponto de vista do espectador de teatro. Com a introdução da montagem, a câmera ganhou mobilidade em cena e assumiu novos papéis narrativos, como pontuar e conduzir a fruição do espectador.

A construção do quadro (recorte visual) parte do objetivo da cena, que pode existir por diversos parâmetros: movimentação dos agentes, significação da luz, superfície e profundidade. A respeito da construção do espaço como elemento da fruição, Aumont (1993) retoma o princípio introduzido pelas imagens não temporalizadas e fixas, como a pintura e a fotografia. Para o autor, essas obras ocasionam uma ação de exploração ocular (*scanning*), uma vez que a compreensão do conteúdo se dá pelo processo de percorrer a profundidade da imagem e identificar os elementos que a compõem. Tal prática pode ser percebida em obras como “Retábulo de Ghent” de Eyck (1432) que, por meio de uma justaposição de elementos, condensam narrativas numa mesma relação espacial. Para exemplificar o conceito de Aumont (1993), foram adicionados, sobre a imagem da pintura de Eyck, indicadores gráficos

que revelam alguns dos diversos pontos focais presentes no super enquadramento¹ da pintura.



Figura 1.1. Composição visual obra de Jan van Eyck (EYCK, 1432) - montagem e inserção gráfica nossa.

Como é possível ver na figura 1, item 1, a construção da ambiência passa pelo agrupamento de diversos pontos focais. Por questões técnicas como resolução, tempo de compreensão e uma linguagem que se propunha a delimitar um único caminho de percepção, essa exposição de múltiplos eventos espaciais foi majoritariamente transposta da fotografia ao cinema pela variação de enquadramento e movimento de câmera. Uma vez que a fruição fílmica é conduzida pelo encadeamento temporal de imagens, o cinema transforma o simultâneo em sequencial.

Apesar da construção segmentada e temporal dominar as produções cinematográficas, podem ser encontrados experimentos estéticos dessa espacialização aplicada ao cinema. Por exemplo, o filme *Cidadão Kane*, de Orson Welles (1941), constrói uma profundidade visual por meio do espaço profundo, escolha técnica em que todos os elementos do primeiro, segundo plano e *background* estão em foco e perspectiva (BLOCK, 2010). Aumont (1993) destaca que a profundidade do quadro traz uma temporalidade própria e uma infinidade de enquadramentos que são traçados pelo espectador ao percorrer a obra.

1.3. Da sala escura às novas mídias

A incorporação de dispositivos tecnológicos impacta a estética cinematográfica desde sua origem. A montagem, o som e a cor foram os primeiros agentes que estimularam um amadurecimento no acervo de técnicas e recursos e, desse modo, constituíram o formato ou a forma, cinema. Tal modelo pode ser definido como integração de filme (narrativa visual), espectador, instituição (dinâmica cultural e econômica) e dispositivo (DUBOIS, 2013). O entendimento acerca da estrutura do cinema tradicional também pode ser obtido por outra classificação: arquitetônica (condições de projeção de imagens); tecnológica (produção, distribuição das imagens); e discursiva (MACIEL, 2009). O autor atribui o termo pós-cinema aos processos de migração de imagem,

¹ Termo adotado por Aumont (1993) para designar a diversidade de enquadramentos que existem dentro do quadro geral que é a pintura.

dispositivo e público que rompem com modelo tradicional. O pós-cinema engloba a saída da sala escura, passando por galerias de arte, televisores domésticos, computadores pessoais e dispositivos conectados e em trânsito como as novas mídias.

Essa evolução compreende o conceito de multimídia como sendo um formato capaz de lidar com mais de um tipo de meio, na relação entre pelo menos um tipo de mídia digital contínua e outras mídias estáticas (MONTEZ; BECKER, 2005). Esse conceito de múltiplas instâncias midiáticas se torna abrangente e acaba ganhando outras conotações, por exemplo: a) os meios, no sentido de aparelhos; b) os modos de apresentação; c) os sentidos implicados na recepção da mensagem.

Dessa forma, a multiplicidade de mídias deriva também no conceito de linguagem hipermidiática, uma construção baseada nos hyperlinks textuais que têm a capacidade de estabelecer relações multimídia. O mercado de obras hipermídia tem seus primeiros registros em meados da década de 1980, mas foi intensificado a partir de 1991 com o lançamento da World Wide Web. De acordo com Gosciola (2008), as obras audiovisuais de caráter hipermidiático exploram, e até expandem, elementos da linguagem cinematográfica clássica, introduzidos por Edwin Potter e Griffith.

O espaço cinematográfico se torna dinâmico em dimensões variadas, da tela do computador aos domos interativos. Para Gosciola (2008), essas novas mídias são um campo delimitado pela natureza digital, e a hipermídia é o meio e a linguagem em que ele se encontra. Sendo assim, para o autor, a hipermídia é uma composição de meios cujo encontro potencializa as funções comunicacionais daqueles que são audiovisuais. Essa expansão das habilidades do filme pode ser identificada pela interação entre obra e público.

O vídeo digital, por sua vez, instaura novas formas e suportes de exibição, bem como as mídias locativas e móveis. Essa discussão se baseia em três variáveis: tela, tempo e trânsito, (MÉDOLA; OLIVEIRA, 2016). Isto é, tamanho da tela em que o conteúdo é exibido; tempo de que o usuário dispõe para consumir o conteúdo oferecido; e trânsito, uma vez que o dispositivo é carregado para aonde se vai. Por fim, esse entrelaçamento de estéticas pode ser considerados como produções que não utilizam o meio apenas como método de distribuição, mas incorporam sua flexibilidade.

Apesar do surgimento e consolidação das produções videográficas, a experimentação continuou ativa no campo cinematográfico. Aly (2012) desenvolveu um estudo que identifica sete desdobramentos do cinema contemporâneo: Cinema Expandido (Sinestésico); Cinema Expandido Digitalmente; Cinema Interativo; Cinema Quântico e Neurocinema; Transcineamas; Cinema Digital; Live Cinema. Essas manifestações têm em comum dois fatores já destacados neste trabalho: 1) a busca por uma aproximação dos dispositivos com a audiência; 2) a inclinação por uma ocupação de novos espaços e interações.

É possível reconhecer uma tendência de transformação das produções por meio de uma relação de escalabilidade entre os desdobramentos da cinematografia e pós-vídeo. De modo geral, identifica-se uma tendência nas experimentações do cinema, por meio do levantamento de Aly (2012), que identifica uma inclinação por expandir os limites da tela e, desse modo, construir experiências pervasivas numa escala macro, ocupando novos espaços e mantendo o conceito de imersão natural. Enquanto isso, as artes do vídeo têm caminhado para uma direção de exploração de mundos existentes por

meio da tela (micro). Gosciola (2008) conceitua essa jornada de navegação em sistemas não lineares como um modo de imersão. Essa, por sua vez, pode ocorrer por meio de uma ambientação gráfica bidimensional ou tridimensional. Todavia, é preciso ressaltar que essa tendência, apontada pelos autores, não implica que uma manifestação não perpassa o “espaço” da outra.

A construção dessa malha de relação entre os formatos e suportes de mídia instauram questionamentos como de Lopes, Montañó e Klipp (2014), “Qual é hoje a tela do audiovisual? A das salas de cinema, a da TV, a dos mobiles, a parede dos edifícios, a camiseta, o corpo desnudo?” (ibidem, 2014, p. 2). Esse diálogo constante entre os meios que têm como essência o audiovisual faz com que nenhum deles se apresente de forma pura. Pelo contrário, esses formatos mantêm uma relação de extremo intercâmbio e convergência.

A partir desse ponto cabe a discussão de Santaella (2005) que aponta o experimentalismo como a nova ordem de linguagem. Além disso, as barreiras entre as artes e as mídias perderam seus contornos, tornaram-se permeáveis. O papel do produtor do conteúdo entra em processo de mudança. Maciel (2004) interpreta que a ideia da tela como o infinito cinematográfico foi redefinida pelo surgimento de novas camadas tecnológicas. Em adição, dialogam entre si na concepção de novos moldes e molduras da narrativa. Para Figueiredo e Mendes (2015), neste cenário, o roteirista (ou diretor) já não responde mais pelo processo, mas sim pelo sistema, ou seja, já que o usuário é cada vez mais responsável pelo desenvolvimento de uma obra audiovisual nas mídias digitais, o papel do roteirista e diretor recai sobre o estabelecimento das regras e possibilidades que servem de guia para os encaminhamentos do processo como um todo.

1.3.1. Interfaces temporais de fruição e interação

Sobre esta dimensão de contato com os diversos atributos de representação do tempo é preciso lembrar que: 1) a maioria dos estímulos visuais varia com a duração ou se produz sucessivamente; 2) nossos olhos estão em constante movimento, o que faz variar a informação percebida pelo cérebro; 3) a própria percepção não é um processo instantâneo, pois tem certos estágios rápidos, outros lentos, mas o processamento da informação se faz sempre no tempo. (AUMONT, 1993).

Somado a isto, a disposição técnica dos sistemas computacionais de armazenamento e manipulação de dados permite que dentro de um mesmo produto audiovisual possam ser incorporados conteúdos relativos a diversas formas de representação da história. Essa relação entre múltiplas versões de representação pode ser vista pela dimensão estética no estudo de Delleuze (2005) como imagem cristal² ou numa visão de IHC como a integração dos filmes à sistemas de interação e banco de dados proposto por Manovich (2001).

A diversidade de dispositivos de interação com produtos de mídia audiovisual dispõem de atributos de representação e manipulação do tempo de modos distintos. Por exemplo, nas salas de cinema a audiência está familiarizada com a recepção da obra por uma baixa taxa de quadros (24fps), cenas com maior duração para a contemplação das

² Para o autor, as produções contemporâneas marcam um paradigma de estreitamento das relações de passado, presente e futuro, ou seja, uma realidade multiplicada, acessível e mutável.

imagens e o uso recorrente do *slow motion* como recurso estético. Já nos dispositivos de vídeo doméstico, o consumo de conteúdo com menor duração e maiores taxas de reprodução como 30fps, 60fps ou 120fps, no caso de jogos eletrônicos, também já compõem uma espécie de convecção de consumo.

Todavia, apesar de uma certa “relação” com múltiplos modos de representação temporal em mídias audiovisuais, experimentações e inovações que buscam romper com essas “convenções”, tem recebido certa resistência³. Nesse sentido, pesquisas tem se preocupado, por exemplo, em avaliar os impactos de altíssimas taxas de reprodução de imagens às capacidades da percepção visual humana.

A recorrente utilização em estudos sobre audiovisual, dos conceitos de persistência de visão e movimento aparente, tem se demonstrado insuficientes para descrever este cenário de impacto das altíssimas taxas de reprodução. Em contrapartida, estudos como o de Watson (2013) tem sugerido um conjunto de medições denominado janela de visibilidade como uma abordagem de relacionar as potencialidades, limitações e efeitos da exposição visual do homem às novas possibilidades de exibição de imagens.

Tag et al. (2016) propõem uma avaliação de stress das pessoas a uma variação de *frame rate* e propõem que o entendimento dos efeitos desses padrões permitam a resposta emocional intensificada ao conteúdo de vídeo, maior envolvimento e, por consequência, histórias mais imersivas. Apesar de estudos identificarem o convívio das pessoas com diferentes dispositivos e contextos de atribuição do tempo, ainda é necessário aprofundar discussões sobre os impactos do manuseio destes atributos controlados pelos indivíduos na medida que utilizam os sistemas audiovisuais.

Dado a evolução técnica e oferta de maiores taxas de amostragem, percebe-se que altas taxas de *frame rate* tendem a melhorar a experiência de percepção de artefatos em movimento. Segundo Mackin, Noland e Bull (2017) a necessidade de aumentar a taxa de amostragem e reprodução das imagens mantém uma correlação com o aumento da resolução espacial das imagens. Os autores destacam que formatos de vídeos com resoluções iguais ou superiores a 8k dispõem de uma melhor representação de detalhes e movimentação de objetos com taxas acima de 120fps.

1.3.2. Interfaces espaciais de fruição e interação

Em relação aos aspectos espaciais da construção da imagem, os contextos e requisitos técnicos constroem uma espécie de moldura de modo que são criadas convenções para a fruição. Como já discutido anteriormente, o desenvolvimento da linguagem cinematográfica privilegiou a montagem temporal, ou seja, a construção dos eventos pelo processo de segmentação e deformação em sequências. Anos depois, com o surgimento da computação e popularização das interfaces web, a montagem espacial volta a produção audiovisual.

De modo semelhante, a diversidade de amostragens de representação temporal, as resoluções espaciais distintas estão presentes no cotidiano das pessoas. O cinema em linhas gerais ocupa um lugar de destaque na qualidade de som e imagem, e possibilita à audiência a compreensão do espaço em atributos como cor, resolução e profundidade.

³ Uma dos exemplos dessa reação foi o lançamento do filme o Hobbit no ano de 2012 que utilizou uma taxa de 48fps e desencadeou diversas críticas e grupos de apoio à experimentação. Outro contexto que têm disponibilizado essa variação de taxas de reprodução também é serviços de vídeo sob demanda.

Já os dispositivos domésticos de vídeo como TV apresentam cenários que variam desde a transmissão analógica (Standard Definition – SD) a sinais de imagem HD e Full HD. Outro contexto são as centrais multimídia e serviços de vídeo sob demanda que permitem às pessoas com infraestrutura específica o consumo de conteúdos com maiores resoluções e amostragens de cor High Dynamic Range (HDR). Por fim, a crescente popularização e evolução dos dispositivos móveis permitem que as pessoas disfrutem de dispositivos com alto poder de processamento e reprodução de vídeo.

Todo esse cenário de diversos dispositivos de interação, somado à evolução técnica das imagens possibilita a retomada da montagem espacial e exploração da imagem por meio de uma postura mais ativa da audiência, que cada vez mais relaciona as interfaces gráficas com as telas de exibição e interação de produtos audiovisuais. Aly (2012) destaca que “o que a câmera e o processo de edição fazem com o tempo (subverter, comprimir, alterar, acelerar e/ou retardar), acontece agora com o espaço.” (Ibidem, 2012, p. 61).

Esse reordenamento espacial da narrativa vem contrapor a lacuna criada pelos elementos que estavam sendo desconsiderados como objetos de interesse.

O campo é o espaço contido no enquadramento fotográfico. Reconhecendo a imagem fotográfica como produto análogo ao obtido pela prática do olhar, o espectador adiciona ao campo um *fora de campo*, ou seja, uma continuidade espacial que extravasa os limites da tela. Em se tratando de imagens em movimento, podemos falar de dois tipos de fora de campo: o fora de campo *já visto*, ou seja, aquelas áreas do espaço que o plano contemplou em um instante do seu movimento, mas que em outro momento deixa de sê-lo, e o fora de campo, *não visto*, que em momento nenhum foi apresentado pelo plano e permanece desconhecido. (ALVARENGA, 2011, p. 42, grifos do autor).

Na figura 1, por exemplo, em que há destaque para a composição visual complexa da obra de Jan van Eyck, existem relações de macro e micro narrativas expostas simultaneamente. Entretanto, ao analisar a forma tradicional de exposição dos vídeos, ocorre uma situação pertinente à figura 2, item 1. Portanto, um fragmento justaposto sequencialmente a outros constrói a impressão da narrativa.

Já o item 2 da Figura 2 é um fotograma da instalação de Jeff Desom, que reconstruiu o cenário da vila e inseriu todos os fragmentos isolados do filme *Janela Indiscreta* num espaço. A experiência continua linear do ponto de vista temporal, contudo, Desom conseguiu trazer a exploração ocular à montagem espacial dos fatos. Tal item tem fortes semelhanças com o quadro destacado na Figura 1, uma vez que ambos englobam diversos pontos focais e constroem uma espécie de super enquadramento. Esses novos modos de exposição, evidenciados pelo trabalho de Desom, refletem a utilização do percurso e manuseio do espaço como uma instância narradora.

“A narratividade espacial implica pensar a ação física do espectador (seu percurso) como performance. Uma performance produtora de significado narrativo. É o percurso do visitante que vai ‘contar a história’” (DUBOIS, 2014, p. 147). Assim, surgem as instalações-exposições, onde um espectador/montador é transformado em um caminhante-narrador (DUBOIS, 2014).



Figura 1.2. Janela indiscreta: (1) Fotograma do filme; (2) Fotograma do experiment. (DESOM, 2012)

Esse audiovisual, que constrói um espaço em que o usuário “entra” na narrativa e tem suas habilidades físicas e cognitivas potencializadas, traz de volta o conceito de imersão “natural” que estava presente na sala escura. A discussão da utilização do espaço como interface de fruição é tão expressiva que pode ser considerada como uma nova forma do audiovisual.

1.3.3. Affordances de interação no espaço-tempo

Por *affordances* podemos entender, por alto, um conjunto de características de um ambiente ou sistema que possibilita a capacidade de identificação e manipulação de parâmetros possíveis de serem executadas dentro deste ambiente. Neste sentido, foram identificadas estratégias conceituais que possam ser utilizadas como *affordances* em sistemas audiovisuais e que viabilizem a transferência de controle de aspectos do tempo e espaço para o indivíduo à medida que ele desenvolve sua relação com a mídia.

Neste levantamento foram identificadas duas estratégias diretas que atribuem processos temporais e espaciais à fruição do usuário, são elas: oportunidades temporais e campo exploratório. Em virtude da forte relação entre sistema e usuário, as abordagens conceituais de historicidade e geratividade também foram identificadas como atributos ofertados por sistemas que dialogam com essa capacidade de manuseio do espaço-tempo. A seguir, detalhe-se todos esses aspectos.

1.3.3.1. Oportunidades temporais

Sawhney; Balcom; Smith (1997) compreendem a obra hipermidiática como um conjunto de oportunidades de ações que apontam para *links* relacionais e complementares à narrativa. Os autores destacam o manejo do tempo atual, a visão histórica da interação e a inserção de conteúdos extra diegéticos como os atributos do *hypervideo*. “Ele oferece aos seus utilizadores um caminho a seguir, fornece momentos narrativos que determinam o que vem pela frente e explica o que veio anteriormente” (SAWHNEY; BALCOM; SMITH, 1997, p. 30.).

A sequência narrativa para Sawhney; Balcom; Smith (1997) serve como um conceito unificador para a criação de tópicos de história e de uma colagem de elementos. Desse modo, o corpo de sequências narrativas pode ser considerado um "mapa" por meio do qual o usuário navega pelas oportunidades que surgem dinamicamente com base no contexto da trama e do percurso tomado.

Há dois grandes grupos de oportunidades temporais: 1) *Intra-frames*, ou seja, oportunidades que surgem com base em elementos implícitos no interior do quadro narrativo; 2) *Extra-frames*, oportunidades com base em elementos exteriores ao enquadramento. Estes por sua vez, podem ser indicados por mudanças no cursor, instruções textuais, índices sonoros ou critérios de interação estabelecidos pela narrativa.

Sobre as oportunidades *Intra-frames*, os pesquisadores evidenciam a complexidade e o potencial deste elemento, uma vez que em obras audiovisuais, os elementos internos do quadro podem se deslocar ao longo do tempo. Esta condição é reconhecida como uma semântica espacial implícita da narrativa

1.3.3.2. Campo exploratório

A relação entre mundo real e mundo da tela se dá pelo recorte visual escolhido pelo diretor. Logo, pontos de interesse do usuário podem ficar de fora (BLOCK, 2010). “Desse modo, elementos do extraquadro são desconsiderados como parte consumível. [...]. No entanto, há cenários em que os elementos tanto do extraquadro quanto do intraquadro possuem relevância” (TOSCANO; BECKER, 2015, p.1).

A utilização de vídeos com o campo exploratório pode ser compreendida como uma nova forma de mídia (Carlier et al. 2011). Esse modo de interação permite que o usuário delimite uma seleção de interesse e o destaque numa perspectiva de escalabilidade, ampliação ou redução de certos elementos em cena. Essa relação é funcional porque a diferença entre a resolução do vídeo e a das telas dos dispositivos móveis permite um estado de proporcionalidade técnica e estética.

Com os recentes sistemas de captura de vídeo UHD, novos tipos de experiências de mídia são possíveis, em que os usuários finais têm a possibilidade de escolher o seu nível de direção, de visualização e *zoom*. (REDI; D'ACUNTO; NIAMUT, 2015, p.2)

A relação entre a imagem captada e o recorte visual estabelecido pelo usuário pode ser identificada pela Figura 3.

O processo de exploração do quadro ocorre por meio de malhas. Essa relação permite agrupar diferentes resoluções da imagem e quadrantes não utilizados não precisam ser processados pelo sistema. O campo exploratório pode ser aplicado em produções bidimensionais, contudo, essa manipulação espacial se torna expressiva com os formatos audiovisuais em 360°, também nomeados por esféricos ou panorâmicos.

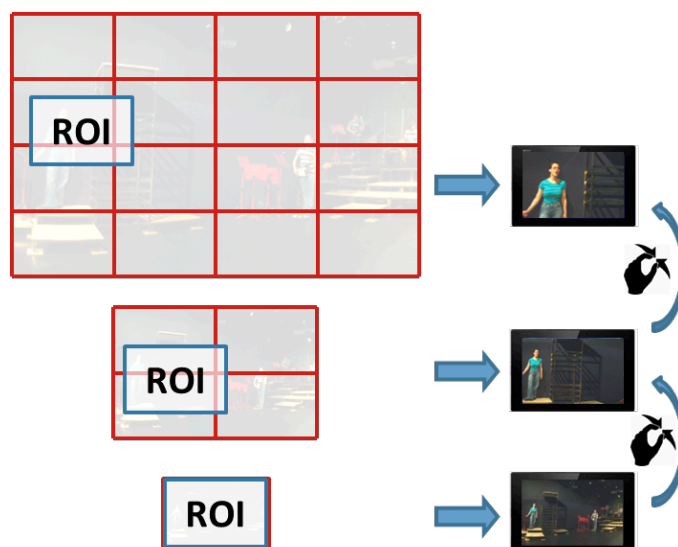


Figura 1.3. Segmentação do vídeo em quadrantes para escalonamento. (REDI; D'ACUNTO; NIAMUT, 2015)

1.3.3.3. Historicidade

A interação nos filmes possibilita o surgimento de novos modos de fruição. Rafaeli (1988) reconhece a interação entre sistemas computacionais e usuário como uma relação previamente programada de eventos e possibilidades. Sendo assim, sugere que o processo interacional entre homem e máquina deve ser percebido pela responsividade ao invés da conversacionalidade. Ou seja, o autor compreende a interação como uma variável de mutação contínua, ao invés de uma condição plena.

O processo interacional é entendido por Rafaeli (1988) mediante três categorias: 1) Comunicação bidirecional; 2) Comunicação reativa; 3) Comunicação interativa. No primeiro modelo de comunicação estabelece-se uma relação entre dois agentes que trocam mensagens entre si, porém não são impactados pelo processo. Recebem e enviam mensagens sem processar sua natureza semântica. Essa relação de dupla via é reconhecida pelo autor como não interativa. Em seguida, sugere-se o modelo de comunicação reativa. Nessa relação, os agentes introduzem o sistema de resposta direta à mensagem recebida, contudo, o processo é limitado a responder com base apenas no último estímulo. Por fim, Rafaeli (1988) introduz um modelo interacional que incorpora o histórico de mensagens e desenvolve uma resposta cumulativa e adaptada ao processo percorrido de cada agente.

A obra audiovisual pode ser percebida então como um agrupamento de módulos que são interligados e modificados em detrimento do histórico de intervenção dos indivíduos. Para Rafaeli (1988), essa capacidade do sistema de interpretar os eventos e

responder de acordo com seu repositório certifica o *status* responsivo da trama. Essa especificidade da interação é complexa tendo em vista a necessidade do autor projetar previamente todo o conteúdo para o processo de interação.

1.3.3.4. Geratividade

A arte generativa compreende o desenvolvimento de projetos artísticos por meio da programação computacional através da implementação de código, da mineração de dados de serviços, processos, dispositivos e deslocamentos, bem como também da hibridização de ecossistemas computacionais e artefatos físicos e analógicos. A geratividade ou generatividade de conteúdo surge inicialmente a partir de um conjunto finito de instruções porém podem atingir uma grande possibilidade de resultados (GALANTER, 2004).

A arte generativa como um campo de pesquisa vem se desenvolvendo desde a década de 1990 e tem sido incorporado em áreas como arquitetura e design. Em relação aos experimentos audiovisuais, Berga et al. (2016) e Carreras e Minguillon (2015) reconhecem o estudo de Manovich (2001) como a base de construção dos sistemas audiovisuais generativos.

Manovich (2001) compreende que a produção do filme é, antes de tudo, a relação do banco de dados gerado pela equipe de produção e sistematizado pelo editor. Desse modo, o resultado é uma possibilidade de ligação entre as diversas opções possíveis. Entende ainda que a inserção do computador no processo de produção dos filmes deve impactar na construção da obra. Desse modo, Manovich (2001) propõe a discussão sobre relação produtiva ente narrativa e banco de dados.

Essas estruturas audiovisuais com base nos dados do usuário são práticas potencias para o *design* de conteúdo dinâmico e personalizado (TOSCANO; BECKER, 2016). A utilização de sistemas, ao invés da intuição artística pelo diretor da obra, é reconhecida por Eigenfeldt et al. (2015) como um estado de meta criação. Uma vez que, ao utilizar ferramentas de inteligência artificial (aprendizado de máquina) o software desenvolve processos de relações criativas, comumente realizados por seres humanos. No entanto, mesmo com a geratividade de múltiplos resultados, o sistema precisa de regras e comandos pré-estabelecidos pelo olhar artístico do autor e dos inputs de interação dos indivíduos.

1.4. Análise comparativa de obras

Com base no levantamento das quatro estratégias: oportunidades temporais, campo exploratório, historicidade e geratividade, descritas na seção anterior, foi realizada uma análise comparativa e conceitual com o intuito de verificar se a relação entre os três itens (referencial, estratégias e produções) convergem, divergem ou emergem significados entre si. Outro objetivo desta análise, é verificar como os autores tem disponibilizado estas abordagens na perspectiva de *affordances* dos sistemas afim de expandir os processos de fruição.

Ao longo deste trabalho, diversos atributos da representação temporal e espacial tem sido utilizados como objeto de discussão. Dentre os principais estão: resolução geométrica espacial, enquadramento ou campo, profundidade de imagem, *frame rate*, duração, fluxo e ritmo. No decorrer das análises, serão apresentadas oito produções que dispõem de alguma contribuição nos atributos elencados. As produções foram

identificadas pelo levantamento de artigos científicos, reviews técnicos e repositórios de conteúdo especializados como: DocuBase, IFA Doc ou comerciais como o Youtube, Facebook e Eko.

1.4.1. Oportunidades temporais

Sawhney; Balcom; Smith (1997) propõem que o projeto de interação de uma obra audiovisual seja desenvolvido com base nas oportunidades temporais que a narrativa e o suporte tecnológico podem oferecer. As oportunidades *intra-frame* podem ser relacionadas à discussão de Tarkovski (1998), que compreende a narrativa não apenas como uma colagem de instantes realizada pelo editor, mas como um resultado de variações que a própria imagem (história) promove. Se para Sawhney; Balcom; Smith (1997) um discurso oferece a oportunidade de expansão narrativa por meio de um link a outros conteúdos, para Tarkovski (1998) a cena a seguir de um filme vem porque um esgotamento da atual se instaurou, logo um novo instante (enquadramento, diálogo, ponto de vista) é necessário para avançar o enredo.

Com relação às oportunidades *extra-frame*, a compreensão de Sawhney; Balcom; Smith (1997) se aproxima de Eisenstein (2002), que defende a autoria do diretor e montador como agentes que inserem sobreposições simbólicas, normalmente atreladas a metáforas, que expandem a diegese fílmica. No ambiente das mídias digitais essa inserção de conteúdo externo que complementa a narrativa principal pode ser manifestada por ambiências gráficas, bidimensionais e tridimensionais, que estimulam os processos de imersão na tela. Afim de compreender possíveis aplicações deste conceito, foram elencadas as mídias interativas “Exhausting a Crowd” e “Vivid 360° Party”.

1.4.1.1. Exhausting a Crowd

A produção audiovisual “Exhausting a Crowd” do diretor Kyle McDonald (2015) expõe um acervo de registros em locais movimentados das cidades de Londres e Amsterdã. Todo o material é reproduzido ininterruptamente de modo que o acesso à obra tem uma correspondência ao horário registrado pelo sistema, ou seja, o horário do acesso tem influência sobre o trecho da gravação que será exibido. A primeira impressão do usuário ao se deparar com o filme é de que o tempo pertencente à imagem e o tempo do espectador se assemelham. Porém, ao interagir com o conteúdo, essa relação de correspondência temporal perde a “sincronia”, uma vez que quanto maior o número colaborações, maior é tempo necessário para compreender as narrativas implementadas.

Exhausting a Crowd oferta a possibilidade de interação por meio de duas ações: (1) retroceder o tempo em ciclos de dez segundos ou (2) inserir anotações na imagem. A oferta de atribuições destas formas de manipulação pode ser comparada aos conceitos de *intra* e *extra-frame*. Ao longo da duração dos eventos desdobramentos narrativos são implementados ou projetados pelo usuário tanto pelos fatos registrados quanto pelas colaborações de outras pessoas que já acessaram a aplicação.

Quanto ao papel do autor, percebe-se que este cria a estrutura de alimentação das interações e contribuições diante das imagens, de modo que os indivíduos que utilizem a plataforma agreguem novos sentidos aquele conjunto de imagens ofertadas. Ainda sobre o produtor do conteúdo, percebe-se que ele estabelece *affordances* gráficas e de

colaboração para que durante toda a experiência de interação, o indivíduo possa sair da perspectiva contemplativa e assuma um papel mais ativo em relação a obra.

1.4.1.2. Vivid 360° Party

O vídeo “Vivid 360° Party”⁴ é uma propaganda de um serviço de internet publicada na plataforma web Youtube. Analisando a obra é possível identificar um percurso de leitura do vídeo “360° que demonstra a utilização da manipulação do tempo como recurso complementar a experiência.

Vamos considerar um primeiro caso: o indivíduo está assistindo ao filme, observando as pessoas da festa dançando, percebe que a música para e que, logo em seguida, os convidados olham para uma mesma direção. Ao se deslocar espacialmente para a direção em que todos olham percebe que a protagonista conversa com o DJ, mas não consegue assimilar o que motivou a parada da música. Nos próximos segundos, o vídeo termina e mesmo que compreenda a chamada publicitária do vídeo, carregará consigo lacunas sobre os eventos que compõem a história.

Já num segundo caso: vamos considerar um indivíduo que durante sua exploração percebeu algo (a música parou) e tomou a seguinte decisão: reposicionar o ponto de vista, voltar no tempo e ajustar a sequência para a reprodução a velocidade de 50% e com isso observar os detalhes do fato. Ao término do vídeo, este indivíduo que desempenhou uma ação entre as ferramentas disponíveis obteve uma maior compreensão dos eventos da narrativa.

Vivid 360° Party utiliza o princípio de oportunidade temporal de modo distinto ao proposto por Sawhney; Balcom; Smith (1997). Ao invés de propor um indicador gráfico, dispõe de uma experiência à deriva em que a manipulação da imagem espacial é a interface que media os instantes pregnantes da narrativa.

Esse exemplo é uma aplicação do conceito de como a manipulação do tempo pode ser útil para expandir a fruição. O campo de possibilidades da instância narradora dos filmes panorâmicos é diverso, inclusive o mesmo filme pode ser visto sem manipulação do tempo num papel “menos explorador”. Nesse aspecto, o tempo real do filme coincide com o tempo de compreender os eventos-chave da história. Já a utilização de elementos escondidos em oportunidades temporais *intra-frame*, incita subunidades temporais, que atreladas à manipulação de velocidade de reprodução, podem expandir a experiência de uso e compreensão narrativa.

1.4.2. Campo exploratório

A construção de espaços complexos para uma fruição espacial, destacada como campo exploratório por Carlier et al. (2011), retoma a discussão proposta por Aumont (1993) sobre narrativas de exploração ocular (*scanning*). Nessa modalidade a compreensão da história se dá pelo processo de percorrer a profundidade da imagem e identificar os elementos de cena. Nas mídias digitais, esse movimento ocular é transposto pela navegação em ambiências gráficas por meio de *zoom*, *panorâmica* e movimentação tridimensional.

⁴ Disponível no endereço: <https://www.youtube.com/watch?v=4HaWlfHpaBo>

O papel do diretor é projetar o universo de variações e segmentações possíveis que tenham sentido como um todo, mas que permitam uma expansão da trama (LININGTON, 2013). Sobre a relação ponto focal e campo geral, Manovich (2014) apresenta uma resposta que relaciona a divisão de superfície como estratégia de manter a dimensão macro e micro ao mesmo tempo. “Em outras palavras, como você continua a ver a imagem inteira enquanto examina os detalhes, o sentido de contexto em que cada detalhe se encaixa permanece” (MANOVICH, 2014, p.112). A compreensão da narrativa pela capacidade de percorrer do intra ao extra quadro proposta por Toscano; Becker, (2015) dialoga com o pensamento de compreensão pela escalabilidade de Manovich (2014) e do redirecionamento de pontos de interesse proposto por Carlier et al. (2011). Para Newton; Soukup (2016) é preciso projetar significados em todos os espaços e conduzir relações de detalhe e contexto para a trama.

Essa multiplicidade espacial tradicional da pintura e fotografia para Lopes; Montaña; Klipp (2014) perde sua força no cinema tradicional, mas é retomada pelas novas mídias, uma vez que os usuários estão acostumados a realizar tarefas com inúmeras janelas e atividades simultâneas, isto é, a narrativa retoma um modelo cognitivo que dá conta de unir diferentes estados de atenção.

1.4.2.1. Video Collections in Panoramic Contexts

O experimento desenvolvido por Tompkin et al. (2013) propõe um sistema audiovisual que utiliza imagens panorâmicas (semelhantes ao Google Street View) para reunir vídeos e processá-los computacionalmente numa visualização conjunta em um ambiente panorâmico e interativo. A aplicação integra registros de relações temporais e espaciais de produções em vídeo que ocorrem no mesmo local da imagem numa interface unificadora em que é possível explorar diversos eventos com relações de tempo e espaço dinâmicos. Dessa forma, a obra desconstrói a lógica de representação do cinema linear ao reproduzir de modo simultâneo o que habitualmente é reproduzido sequencialmente.

Na primeira etapa, o usuário tem a disponibilidade de escolher um panorama para navegação e logo em seguida navegar por um ambiente 360°. Nesse instante é possível estabelecer regiões de interesse e iniciar uma jornada de exploração. Pela interface gráfica é possível identificar os seguintes elementos: 1) Faixa dinâmica dos vídeos disponíveis e habilitados; 2) Foco ou enquadramento dos vídeos selecionados; 3) Timeline do vídeo selecionado; 4) Play/pause; 5) Timeline global, que oferta o controle de dias, meses e anos com conteúdo disponível; 6) Controle de sobreposição de identificação dos vídeos; 7) Ajuste de perspectiva de visualização; 8) Bloquear a reprodução do vídeo.

Esses elementos ofertados pela interface constroem um ambiente de exploração que estabelece uma relação narrativa pelo deslocamento de conteúdo dos vídeos e interação. Este experimento emerge significados às oportunidades temporais à medida que integra num mesmo instante a possibilidade de variação temporal e espacial dos eventos durante a fruição do conteúdo.

1.4.2.2. Michelle Obama 360° interview

O vídeo Michelle Obama 360° interview, é uma entrevista⁵ realizada em janeiro de 2016, com a então primeira dama dos Estados Unidos. A construção narrativa empregada ao vídeo associada às funcionalidades de interação espacial pelo *player* de vídeo mobile do Facebook expandem a experiência de assistir uma entrevista.

O primeiro ponto a ser levantado, é o controle por parte do usuário do ângulo de visão da cena. Conforme a figura 4, é possível perceber que da exibição normal (frame 2) ao *zoom in* (frame 1) e *zoom out* (frame 3), visualizações diferentes podem ser ofertadas, seja o *intra* quadro, do *frame 1*, ou o contexto de sequenciamento espacial do infográfico no *frame 3*. Esse recurso exploratório proporciona ao usuário uma expansão da compressão narrativa por meio de uma instância dinâmica e manuseável.

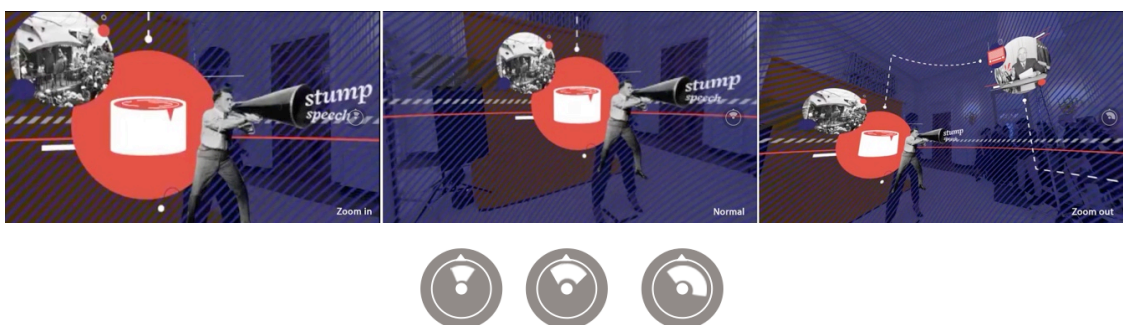


Figura 1.4. Variação perspectiva de visão e bússola de referência. Fonte: Captura de tela e montagem realizada pelos autores

A bússola destacada na Figura 4 norteia o usuário em seu posicionamento no espaço e sua relação com o ponto central estabelecido pelo autor da obra. Além disso, a indicação do ângulo de visão ou *zoom* também é ofertada pela bússola, como podemos ver na Figura 4. Por fim, ao clicar no ícone referente à bússola, o *player* do vídeo responde a ação conduzindo o usuário ao posicionamento e ângulo *default* definidos pelo diretor.

Apesar da manipulação espacial *default* dos vídeos panorâmicos, a solução disponibilizada nesta plataforma possibilita atribuições estéticas que evidenciam que, mesmo dentro de um vídeo 360°, é possível oferecer à audiência o acesso a recortes e direcionamentos de interesse pessoal. O campo no 360° é amplo, porém ao adicionar o ângulo de visão dos objetos ou *zoom* cria-se mais uma camada de aprofundamento de exploração na imagem.

1.4.3. Historicidade

O ato de incorporar o histórico das interações e adequar as respostas a esse percurso é compreendido por Rafaeli (1988) como historicidade. Cirino (2012) propõe que a lógica da historicidade de uma obra deve começar no processo de roteiro, de modo que o compartilhamento de cenas, prática comum em filmes reativos, seja substituída pela incorporação de significados atribuídos ao longo do percurso narrativo. Além disso, o

⁵ Disponível em: <<https://www.facebook.com/verge/videos/1055823707787305/>>

autor destaca que, tanto em narrativas com tramas complexas (hiper-histórias), quanto em narrativas espaciais, a responsividade em detrimento da historicidade pode ser aplicada.

1.4.3.1 The sound of energy

O vídeo interativo “The sound of energy”⁶ é uma ação publicitária da empresa Shell. A proposta da campanha perpassa outras instâncias midiáticas, contudo, nossa análise se delimita à experiência narrativa proposta ao vídeo interativo. A experiência do vídeo consiste na disponibilização de conteúdo audiovisual que vai sendo justaposto a elementos que o usuário insere. Essa junção de conteúdo ao longo do vídeo promove um mix de imagens e sons sobre formas de energia. O sistema desenvolvido tem a capacidade de processar a historicidade das escolhas do usuário e estabelecer a responsividade dos blocos de modo a criar um clipe personalizado para cada execução.

As interfaces do sistema representam o campo de possibilidades que promove a variabilidade do produto, uma vez que as escolhas do usuário resultam em novas composições de conteúdo (tipos de energia) e estética por meio de enquadramento, edição de sons e imagem.

Cirino (2012a) compreende que a interface media a relação entre usuário e estrutura narrativa fílmica com relações cada vez mais significativas de fruição. No vídeo “The sound of energy”, a interface ocupa o papel de norteador do caminho percorrido e das possibilidades temporárias de expansão. Além disso, o gráfico indicador das escolhas feitas durante a exibição do filme é reformulado ao término da narrativa para expor ao usuário seu percurso.

Em “The sound of energy” a construção da narrativa é oriunda do processo cumulativo das escolhas, ou seja, de modo semelhante à premissa de Rafaeli (1988), a obra desencadeia o conteúdo com base no percurso adotado pelo usuário. A obra se alterna entre estágios de: 1) narrativa base, primeiras cenas do vídeo; 2) expansão, oferta de conteúdo; 3) Switch, alternância e pré-visualização de pontos de vista; 4) Compilação final, construção de clipe personalizado com opções compartilhamento e exibição do resultado;

1.4.3.2. Crowdsourcing Video

Carlier et al. (2011) propõe um experimento em vídeo que minera os dados da manipulação espacial e estabelece um processo de recomendação de pontos de vista ao usuário. Esse experimento compreende a historicidade como uma estratégia de expandir o vocabulário do sistema que acopla uma narrativa para a geração de respostas.

A mineração dos dados de navegação é constituída de sete elementos: (1) Vídeo original; (2) Mapa de importância; (3) Vídeo para análise com grupo; (4) Mapa de interesses; (5) Junção dos mapas e; (6) Aplicação dos indicadores gráficos relacionados a enquadramentos. Cada um desses itens é processado por um sistema computacional até que pontos de vista possam ser recomendados ao usuário no (7) Vídeo final.

O mapa de importâncias predefinido pela equipe de pesquisa considera as pretensões do autor sobre os elementos em cena que merecem ser destacados. Ao expor

⁶ Disponível no site: <http://interlude.fm/home/videos>. Acesso em 10 de novembro de 2016.

a narrativa a um grupo preliminar de usuários o autor obtém um mapa de navegação de interesses. A indicação de cada etapa da coleta de informações de feedback dos usuários é denominada *Crowdsourcing*. Assim, o sujeito enunciador (autor) fornece elementos e o sujeito atualizado (usuário) realiza parte das possibilidades e introduz novos sentidos.

As recomendações de ponto de vista são promovidas pela interface gráfica, que além de indicar pontos de interesse permite ao usuário explorar a imagem e demarcar um novo ponto focal. O campo exploratório é um gerador de dados, uma vez que cada usuário ao demarcar novos posicionamentos (escalonamentos), também computa um recorte e sugere um novo ponto de vista.

Embora exista a opção de automatizar o processo de acompanhamento dos objetos de interesse em movimento, o sistema utiliza o histórico de interação dos usuários no campo exploratório para computar uma biblioteca de interesses. Os dados da biblioteca expandem os atributos do sistema à medida que os levantamentos dos dados dos reais interesses da audiência informam ao sistema como ele deve se portar nas respostas e automações. Ao término do processo de mineração e junção dos mapas de importância (autor) e interesse (usuário), Carlier et al. (2011) atribui ao sistema modelos de interpretar os dados e repondê-los com base em práticas de montagem.

1.4.4. Geratividade

A geratividade, segundo Eigenfeldt et al. (2015), pode ser considerada um estágio de meta criação, uma vez que o sistema passa a desenvolver relações combinativas (criativas) para fins estéticos. Se por um lado o autor da obra partilha sua visão com o usuário, nas ditas mídias digitais interativas o sistema também participa desse processo, seja nas relações de reprodução a construções autônomas de sentidos à narrativa.

O entendimento de que a narrativa pode gerar signos e significados que estão acoplados em sua essência mas que ainda não foram expostos, se assemelha à percepção de Pimenta (2001), que considera a narrativa hipermidiática como uma matriz criativa e geradora oriunda das relações de tempo e espaço. A relação de construção de sentido pela relação de imagens autônomas a contextos preestabelecidos passa pelo experimentalismo proposto por Gaudenzi (2013) e relaciona o processo de montagem construtiva e geografia criativa, uma vez que, ao justapor imagens, o sistema oferta uma lógica de relação entre as imagens.

1.4.4.1. The wilderness downtown

“The wilderness downtown”⁷ é um projeto do diretor Chris Milk, em parceria com a Google Data Arts, para produzir o videoclipe da música “We used to wait” da banda canadense Arcade Fire. A narrativa de “The wilderness downtown” utiliza dois modos de interação: a geratividade de conteúdo baseada na localização definida pelo usuário e a participação na produção textual e gráfica de um cartão postal durante a exibição do clipe.

A interface do clipe é composta por pontos focais dinâmicos, que por meio de uma montagem espacial de *pop-ups* integra o conjunto de cenas programadas pelo diretor a dados de localização retirados do Google maps (Ver figura 6). Em “The wilderness downtown”, a narrativa principal predefinida pelo diretor é ambientada e

⁷ Disponível em: <<http://www.thewildernessdowntown.com/>>.

complementada pela interação. Desse modo, a trama não tem sentido sem a participação do usuário de modo que a narrativa estaria sem módulos de ambientação.

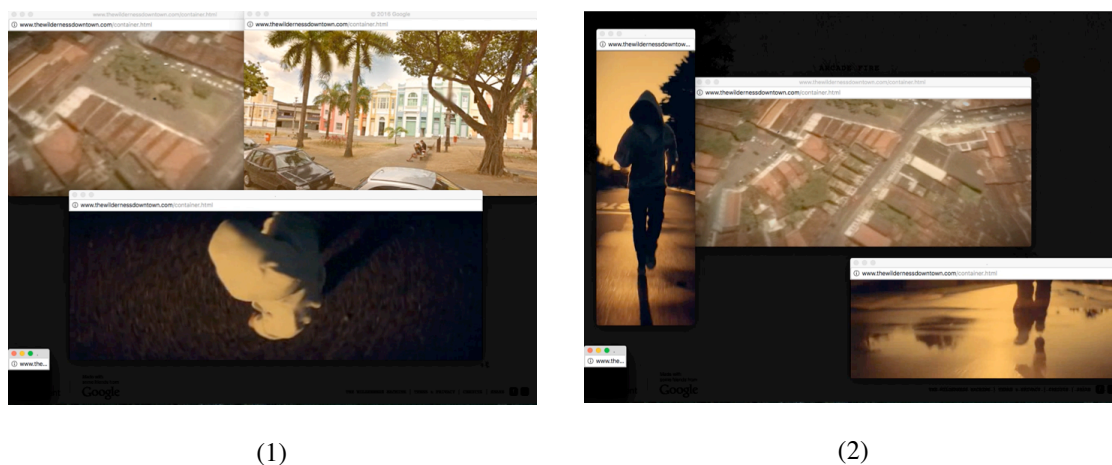


Figura 1.5. Interação entre narrativa e dados do usuário. (MILK, 2011)

Na Figura 6, é possível identificar dois instantes com 3 *pop-ups* gerados pelo sistema. Para ilustrar a experiência narrativa foi inserido no início da aplicação o endereço de um ponto turístico da cidade, a praça Antenor Navarro⁸, no centro histórico de João Pessoa – PB. No instante destacado pela Figura 6, item 1, o personagem está perdido e por isso olha em sua volta. Enquanto isso, o sistema insere com base na localização informada dois modos de visualização do Google Maps e adiciona um movimento de rotação para a construção de uma associação do movimento. Essa solução utilizada pelo diretor pode ser identificada pelo princípio da montagem construtiva e geografia criativa proposta por Kuleshov (NOGUEIRA, 2010).

“The wilderness downtown” além de ofertar a geratividade de módulos de conteúdo à trama por meio da montagem espacial dos elementos também aplica a geratividade estética sobre as informações do usuário. A manipulação estética dos dados inseridos pelo usuário pode ser percebida como a meta-criação do sistema para a narrativa, uma vez que, baseado em requisitos estabelecidos pelo diretor e programador, o sistema interpreta os dados coletados e adiciona uma camada estética de significado aos dados.

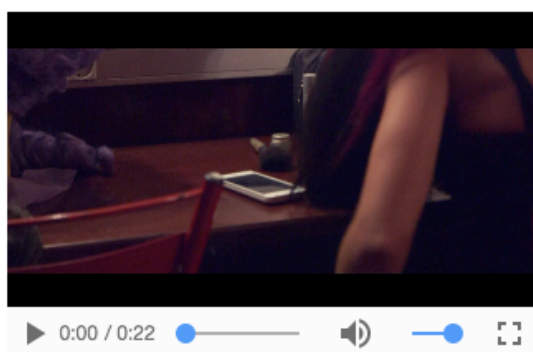
O número de janelas de exibições é variável, bem como seu arranjo e conteúdo. A narrativa de base projetada pelo diretor é expandida por oportunidades *extra-frames* que surgem no ambiente gráfico para complementar o sentido do conteúdo principal. Quanto à lógica de interação, “The wilderness downtown” apresenta no momento inicial da narrativa, uma ação configurativa, tendo em vista que é por meio da contribuição da audiência que as especificidades da narrativa serão expostas. Do ponto de vista da fruição, o usuário está diante de uma meta-realidade que envolve seu contexto e o mundo da obra.

⁸ Endereço inserido na plataforma: Maciel Pinheiro, 16 – Varadouro, João Pessoa – PB.

1.4.4.2. I' am

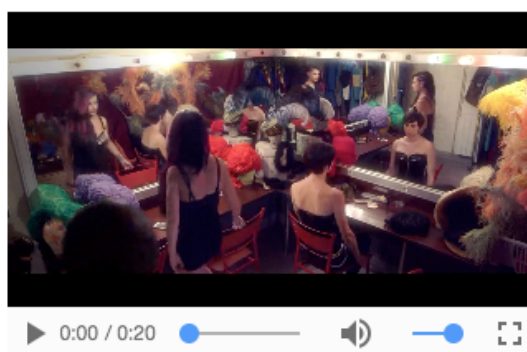
“I' AM” é uma plataforma online, desenvolvida por Berga et al. (2016), capaz de gerar em tempo real clipes audiovisuais de 24 segundos, sempre originais e inéditos. Cada clipe se refere a situações cotidianas que demonstraram como as novas tecnologias estão mudando o nosso cotidiano pela vasta quantidade de informação disponível. “I' am” é um produto que muda o paradigma da construção de narrativas uma vez que os papéis de autor, sistema e usuário são impactados.

Por meio de 223 *takes* sobre seis situações, somados a 9 trilhas sonoras e 32 frases diferentes, a plataforma desenvolvida combina os arquivos aleatoriamente de forma estruturada e consistente, gerando uma cena inédita a cada uso. Como interface desse processo, Berga et al. (2016) desenvolveu um infográfico animado que revela o processo de seleção randômica do conteúdo. Ao passar o cursor do mouse é possível identificar o conteúdo respectivo.



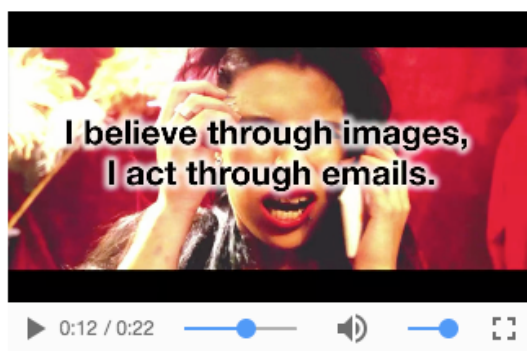
Generat a 2016-11-07 15:33:48

Clipe 765 – Frame 1



Generat a 2016-10-30 16:37:10

Clipe 764 – Frame 1



Generat a 2016-11-07 15:33:48

Clipe 765 – Frame 2



Generat a 2016-10-30 16:37:10

Vídeo 764– Frame 2

Figura 1.6. Comparativo de cenas entre vídeo gerados. (BERGA, 2013)

Apesar da plataforma interativa “I' am” dispor de 6 situações, nos dois clipes gerados (764-765) foi ofertada a mesma situação. Apesar da temática em comum, o modo de representação escolhido para cada clipe modifica a compreensão dos eventos. No clipe 765 a narrativa inicia por um plano fechado, enquanto no clipe 764 por um plano mais aberto, onde podemos perceber o contexto da cena.

Já aos 12 segundos da cena, os cliques mantêm uma similitude na temática, porém as relações estéticas (cor e enquadramento) e discursivas (textos) diferenciam a narrativa. Essa é a relação de variabilidade e automação do sistema.

De modo semelhante à perspectiva de Galanter (2004) sobre a necessidade de estruturar requisitos para produção de arte generativa, foi identificado na narrativa de “I am” uma estrutura de eventos: 1) Entre 0 e 5 segundos a temática é apresentada; 2) de 5 à 12 segundos um evento rompe com a normalidade e desencadeia uma reação; 3) Entre 12 à 20 segundos surge o ponto alto da narrativa; 4) Entre 20 à 25 segundos tem o desfecho e o retorno ao estado inicial. A estrutura identificada contém, dentro de um intervalo de 25 segundos, todo o arco narrativo tradicionalmente utilizado no audiovisual. Esta pesquisa compreende que a divisão elencada anteriormente pode divergir do conjunto de regras atribuídos ao sistema pelo diretor e programador, porém ao comparar o produto gerado pelo sistema, percebe-se a existência de uma estrutura modular que induz no sistema a criação da narrativa. As regras que regem a narrativa são implementadas pelo autor, mas a montagem da trama acontece pelo sistema como meta-criador do conteúdo.

1.5. Arquitetura de design colaborativo

O desenvolvimento de conteúdos audiovisuais se tornou mais complexo à medida que novos recursos de interação, de compartilhamento, diferentes telas e formas de visualização e novas formas de transmissão foram se popularizando. A experiência de ver televisão ou assistir a um filme se tornou mais complexa, podendo envolver desde pesquisas na internet a complexos algoritmos de recomendação e adaptação de conteúdos. O próprio conceito de obra audiovisual atualmente é fluido, com exemplos que se adaptam ao momento da fruição. Da mesma forma, discutir e trocar informações sobre determinado programa, antes feito pessoalmente com vizinhos e colegas, atualmente pode envolver compartilhamentos de opiniões e críticas nas redes sociais até o download da obra e edição de trechos para gerar novas versões. Dessa forma, consideramos que o termo adequado que melhor explica a relação dos indivíduos com sistemas audiovisuais é fruição, que inclui o ato de ver ou assistir, o processo psicológico de apreensão do conteúdo e a incorporação de interação e compartilhamento.

Uma característica deste cenário é a presença e o uso de softwares para compor a experiência de fruição das obras audiovisuais. As interfaces de interação e as interfaces gráficas ganharam papel central no audiovisual, podendo ser consideradas, elas mesmas, como mídias. Em determinadas obras, a experiência de compartilhamento se torna mais importante do que o próprio momento de audiência, havendo uma troca de informações constante entre diferentes papéis dos indivíduos. De certa forma, todo processo de fruição audiovisual passa a ser colaborativo, com uma relação próxima - e de troca, entre indivíduo e sistema e indivíduo e produtores.

Para compreender o desenvolvimento e a fruição de sistemas audiovisuais é preciso organizar os papéis que um indivíduo pode desempenhar em relação a estas obras. Uma forma de analisar este processo é integrar duas áreas distintas, que se complementam neste novo cenário: a Interação Humano Computador (subárea da Ciência da Computação) e os Estudos de Audiência (também chamada de estudos de

recepção por algumas linhas de pensamento, é uma subárea das Ciências da Comunicação).

Em publicações anteriores foi desenvolvido o conceito de Design Audiovisual (DA), um modelo teórico e metodológico que integra teorias de IHC com os estudos de audiência, gerando teorias, métodos e ferramentas de criação e disponibilização de conteúdos em redes digitais. O modelo é representado por um fluxo de trabalho gráfico que permite o reconhecimento do fluxo dinâmico da produção audiovisual considerando uma variedade de cenários e papéis desempenhados por indivíduos (BECKER; GAMBARO, RAMOS, 2017).

Nos processos de desenvolvimento de software, o usuário usualmente é um indivíduo abstrato, comumente identificado por arquétipos. Por outro lado, nos estudos de audiência, a pessoa diante dos meios de comunicação não é tratada individualmente, mas na forma de grupos ou percentuais de um todo. Para integrar essas visões e atender a demandas novas de comportamento no mundo digital, no modelo do DA uma pessoa pode desempenhar diferentes papéis em diferentes momentos: Audiência, Sintetizador, Modificador, Jogador e Produtor. O modelo é descrito em forma de conjuntos (Figura 7), para facilitar a compreensão de grandes grupos de indivíduos que contêm outros como subconjuntos.

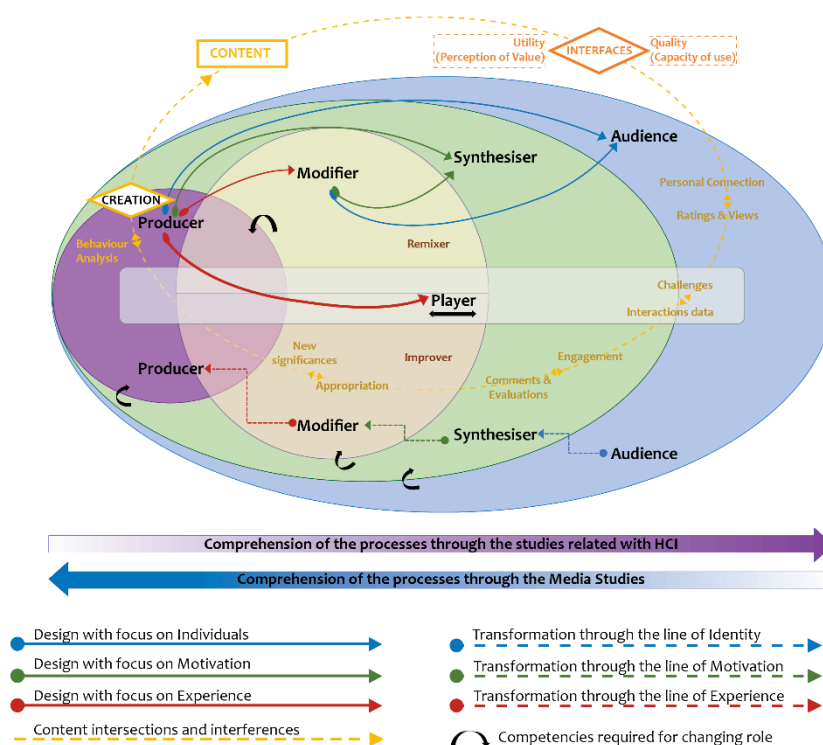


Figura 1.7. Fluxo processual para o design audiovisual (BECKER; GAMBARO; RAMOS, 2017)

Audiência: esta é a base para todos os papéis que um indivíduo pode assumir. Indica baixo nível ou ausência de interação durante o consumo de mídia. É o

comportamento passivo usualmente associado ao modelo Broadcast, abordando as interfaces digitais através da seleção de canais, busca e reprodução de conteúdo, assinatura de um *feed* ou canal, etc. A relação dos indivíduos com o conteúdo ocorre no nível do gosto pessoal e permanece relativamente particular.

Sintetizadores: os indivíduos que compõem esse papel apresentam competências para compilar, classificar, comentar, recomendar e compartilhar o conteúdo que eles gostam, geralmente para criar e "identidade", ou seja, um perfil encenado em uma rede social. O papel do sintetizador considera a noção de engajamento, o vínculo emocional que permite às pessoas expressarem algo sobre si mesmas usando o conteúdo a que se relacionam.

Modificador: este é um subconjunto dos sintetizadores com mais competências e habilidades. Dominam softwares para manipular e recriar conteúdos aos quais eles se identificam, ampliando assim a noção de engajamento para a apropriação. Eles podem ser subdivididos em dois grupos principais - embora possa ser possível identificar outros: os Melhoradores, ou seja, aqueles que se apropriam do conteúdo para mudar algo que não lhe agrada; e Remixadores, que criam novos conteúdos a partir de conteúdos originais.

Produtor: uma pessoa ou grupo de pessoas que cria conteúdo original (inspirado ou não por outras mídias), autônomo e independente ou como colaborador em grandes corporações de mídia. Embora cada produtor seja um sintetizador (eles mantêm entre suas competências a distribuição de conteúdo), as competências do Modificador só se aplicam quando o conteúdo desenvolvido industrialmente é uma adaptação a outra. Assim, a superposição de ambos os papéis é apenas parcial, conforme representado na Fig. 10.

Jogador: este papel se sobrepõe a todos os outros, refere-se aos indivíduos que utilizam completamente as ferramentas disponíveis em cada nível. Suas ações, especialmente aquelas não previstas no projeto do produto audiovisual, alimentam o Produtor em desenvolvimentos futuros. Em outras palavras, eles podem perceber e desenvolver recursos não previstos no seu nível. Os jogadores perseguem conteúdo desafiadores que os faz - mesmo que individualmente, pensar e executar uma ação. Pode-se identificá-los (não restritivamente) com os *early adopters* ou *early users*, ou seja, aqueles que assumirão o risco de usar uma nova tecnologia e, assim, contribuir para o seu desenvolvimento. O modelo DA tenta prever o comportamento de todos os usuários, mas no caso do jogador, abre a possibilidade de usos imprevistos e inovadores.

Um importante recurso do DA é o planejamento da produção e análise da fruição audiovisual através de quatro linhas conceituais de design que moldam o processo e a experiência diante de sistema midiáticos: Identidade, Motivação, Experiência e Conteúdo. Através destas quatro linhas é possível identificar e analisar as mudanças de papel do indivíduo, as habilidades e competências necessárias em cada papel, assim como os impactos na infraestrutura gerados pela troca de informações e inclusão de novas tecnologias.

Linha de Identidade: a relação do indivíduo com o produto audiovisual é basicamente uma identificação pessoal. Os indivíduos escolhem o que querem assistir ou ler com base nas preferências pessoais, na disponibilidade do conteúdo e no tempo vago para a fruição. Os produtores precisam mapear as preferências individuais para

criar conteúdo adequado para o público-alvo. O processo inclui um conjunto analítico de dados sobre os indivíduos que devem marcar o desenvolvimento da história ou das interfaces de interação de forma adequada a cada contexto de uso.

Linha de Motivação: o planejamento de um produto focado no sintetizador deve considerar a intencionalidade do que um indivíduo pode querer comunicar sobre si mesmo. O ato de compartilhar conteúdos em redes sociais está atrelado à criação de uma identidade ou um perfil que conscientemente ou não, disponibiliza uma série de informações pessoais. Consequentemente, tanto a forma quanto o conteúdo devem permitir o desenvolvimento de competências que despertem o engajamento dos indivíduos.

Linha de Experiência: o envolvimento com o conteúdo é conduzido pelo uso avançado de recursos interativos que devem fornecer experiências práticas com conteúdo e interface. Dessa linha, deriva a necessidade de propor desafios complexos aos jogadores e modificadores visando o engajamento.

Linha de Conteúdo: a metodologia DA considera conteúdo como produção audiovisual junto com as interfaces através das quais o indivíduo tem acesso a este conteúdo. As relações variam dependendo do papel desempenhado pelo usuário: relacionamentos pessoais no nível da Audiência; engajamento no Sintetizador; apropriação no Modificador; desafios no jogador; Criação no produtor.

Finalmente, é importante destacar que os indivíduos podem se relacionar de diferentes formas em cada linha de design, bem como acumular papéis ou alterná-los durante o processo de fruição. A natureza complexa da relação entre os indivíduos e sistema audiovisual, passa pelas linhas de design, *affordances* e papéis apresentados no DA de modo que a fruição como experiência vai sendo construída.

Do ponto de vista da arquitetura de produção, de um lado o Produtor constrói um artefato (sistema audiovisual) com um conjunto de *affordances*, enquanto que o indivíduo constrói sua relação com a mídia mediante seus fatores humanos (motivações, experiências) e conduz sua ação para atingir objetivos e atividades. Sendo assim, o Design Audiovisual como arquitetura oferta uma base teórica que pode ser aplicada em práticas de produção de conteúdo.

Com relação aos aspectos da dimensão de tempo e espaço, antes controlados pelos produtores de conteúdo, percebe-se o potencial de utilização dessas estratégias como *affordances* para as pessoas explorarem (oportunidades temporais e campo exploratório) o universo narrativo, gerando dados ao longo do processo (historicidade) que podem retornar ao produtor como necessidades de produção imediatamente ou em produções futuras (geratividade).

1.6. Visão de futuro

Tempo e espaço atuam como interface do processo de fruição nas mídias interativas, isto é, ocupam a função de mediar a compreensão narrativa por meio de *affordances* do sistema com o objetivo de atender os mais variados papéis de interação, seja audiência, sintetizador, modificador ou produtor. Dessa forma, a interação e a imersão podem apresentar diferentes características, com variadas demandas estruturais de processamento do vídeo e tráfego da rede, já que o processo de fruição narrativa se apresenta cada vez mais interligado a utilização de sistemas computacionais complexos.

A partir da fundamentação discutida neste capítulo, da análise das obras e da arquitetura de design colaborativo apresentada na seção anterior, discutiremos agora possíveis e potenciais evoluções técnicas, tecnológicas e narrativas do ecossistema audiovisual. O foco da discussão será centrado na fruição audiovisual, o que engloba a exibição e transmissão das obras. Trata-se de uma visão de futuro do Grupo de Pesquisa Design Audiovisual, baseada na experiência de seus pesquisadores, no domínio da literatura pertinente, nas pesquisas em andamento e nos projetos em prospecção.

A visão de futuro sobre a evolução das narrativas audiovisuais e seu impacto nas infraestruturas de exibição e transmissão será apresentada em dois cenários, visando minimizar a margem de erro da análise. A análise por cenários é interessante porque permite segmentar os temas e discuti-los independentemente. O uso de cenários é útil para antever o que pode acontecer se um conjunto diferente de previsões se confirmar. Autores sugerem que sejam feitos cenários a partir da previsão mais provável, com pelo menos uma variação (VAN DER HEIJDEN, 1996; BUARQUE, 2003). No presente caso, vamos trazer um cenário com impactos limitados em infraestrutura e outro com impactos elevados, mudando a forma como indivíduos se relacionam com mídias audiovisuais.

O foco dos cenários será no entretenimento doméstico, que guia a maior parte da indústria do entretenimento, sistemas educativos pedagógicos e sistemas de telessaúde, que têm demandas científicas e sociais altas. O primeiro cenário foca na evolução da qualidade da imagem, e não depende, necessariamente, de telas maiores, ou do aumento dos recursos de interação e interatividade. Já o segundo cenário analisa a evolução dos sistemas audiovisuais a partir do aumento da resolução dos sistemas de exibição, com incremento substancial de recursos para geração e fruição das obras. Ambas as visões complementam a visão de futuro deste trabalho. Os dois cenários não são excludentes, podendo se materializar parcial ou integralmente.

1.6.1. Cenário 1: desenvolvimento da profundidade da imagem

Os recursos audiovisuais tradicionalmente são considerados mediadores entre o objeto, ou universo, representado e os indivíduos que assistem as obras. As pessoas entram em contato com universos narrativos, paisagens, eventos esportivos, países, fatos ou sonhos através da televisão, do cinema, da internet. Personagens se tornam tão familiares como se participassem do dia a dia dos indivíduos; lugares se tornam conhecidos ou desejados a partir de filmes ou novelas.

Dessa forma, o audiovisual é a representação de um mundo que pode ser conhecido através de uma tela. Esses mundos, narrativamente, fazem sentido a partir do momento em que os indivíduos se identificam e se motivam a pensar, refletir ou sonhar, tornando-se uma janela para experiências novas e diferentes. Muitas dessas experiências são impossíveis de serem vivenciadas no mundo real, seja pela impossibilidade física (viajar ou estar presente no local da notícia), seja graças a impossibilidades lógicas (o universo não existe no mundo real).

Se do ponto de vista narrativo e imagético, a televisão pode ser considerada como uma janela para o mundo, fisicamente ela está longe de atingir tal objetivo. O espectro de luz possível captado pelas câmeras de TV ou de cinema, e exibido pela televisão, está longe do espectro visível pelo olho humano. Em outras palavras, a televisão aquém de apresentar imagens próximas às captadas pelo olhar dos indivíduos.

A CIE (*International Commission on Illumination*) define quantitativamente espaços de cor entre cores físicas puras no espectro eletromagnético visível e cores percebidas fisiologicamente pela visão humana. O espaço de cor definido pelo Rec.2020 pode reproduzir cores que não são presentes na recomendação anterior (Rec.709 – HDTV). As cores primárias RGB (vermelho, verde e azul) usadas pela Rec.2020 são equivalentes às fontes de luz monocromáticas do lócus espectral do CIE 1931. O comprimento de onda das cores primárias do Rec.2020 é de 630nm para a cor primária vermelha, 532nm para a cor primária verde e 467nm para a cor primária azul. Para efeitos de comparação, o espaço de cor do Rec.2020 cobre 75% do espectro visível, o de um projetor de referência de cinema digital cobre 53%, o do Adobe RGB cobre 52,1% e o do REC. 709 cobre 35% por cento de seu espaço de cor. Ou seja, na televisão de alta definição atual, é possível perceber apenas pouco mais de 1/3 das cores.

Mesmo considerando as resoluções 4 e 8K, a imagem ainda está longe de compreender todas as cores visíveis. A olho nu os indivíduos enxergam pelo menos 25% a mais de detalhes de cor. Por isso, é impossível uma pessoa confundir, por exemplo, uma televisão ou uma tela de cinema, com uma janela.

A Figura 8 reforça este comparativo. Uma das melhores câmeras no mercado atualmente é a Sony F65, usada largamente na indústria de cinema e com sensor preparado para captar imagens na resolução 8K. É a câmera hoje com maior capacidade de gravação de variação de comprimentos de onda do espectro visível. É a primeira câmera que consegue um desempenho superior à película de cinema, utilizada na indústria cinematográfica durante mais de um século.

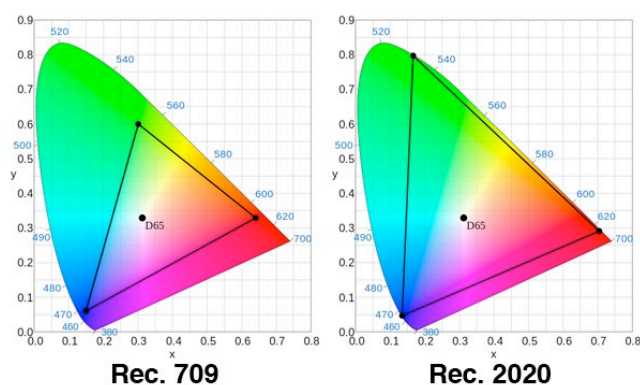


Figura 1.8. Comparação da diferença de abrangência de cores entre o espaço de cor do Rec.709 e do Rec. 2020

Também podemos ver a diferença em relação ao Digital Cinema Initiatives (DCI), que hoje é padrão mundial de projeção nas salas de cinema comercial. A quantidade de luz do espectro visível que os projetores atualmente são capazes de projetar na tela é um pouco mais da metade do que o olho humano seria capaz de enxergar se a pessoa estivesse presente na gravação dos filmes.

Comparativamente, a diferença entre o cinema e o HD é muito menor do que entre o cinema e a película ou entre o cinema e o 4K. Dessa forma, podemos afirmar que, além da limitação das câmeras na hora da gravação das imagens, ainda há uma redução da quantidade de cor que compõe a imagem na hora da projeção.

O padrão Rec. 2100, de 2016, introduziu o High Dynamic Range (HDR), o que melhora a latitude das imagens, especialmente a relação de contraste entre elementos muito claros e outros, muito escuros, na mesma imagem. Trata-se de outro recurso importante para a melhora estética das imagens.

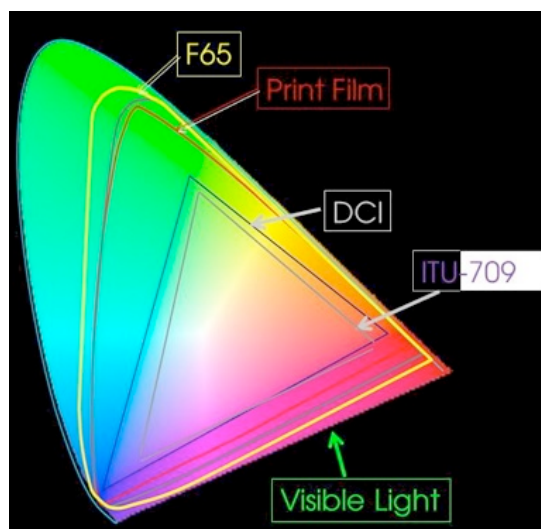


Figura 1.9: Comparativo da quantidade de luz captada pela câmera F65 com a capacidade de gravação em negativo, padrão DCI e o Rec.709

A evolução da representação e percepção de cores nas imagens depende de três elementos: a evolução do padrão Rec. 2020, com a padronização de novos comprimentos de onda incluídos; a oferta de câmeras aderentes a estes padrões; e a oferta de receptores capazes de exibir o novo padrão. Dessa forma, é possível projetar pouca evolução neste aspecto no prazo de dois ou cinco anos. Se considerarmos a velocidade da expansão da oferta e diminuição dos custos de câmeras cinematográficas profissionais, como as da linha RED, é possível projetar um aumento considerável na captação do espectro visível, mesmo que a padronização não acompanhe o processo. Tanto as câmeras da linha RED, quanto a F65 da Figura 9, captam luzes além do padrão Rec. 2020, especialmente nas regiões mais quentes, próximas ao magenta. Dessa forma, é possível vislumbrar uma evolução concreta da captação da imagem em um prazo de 10 anos, com uma aproximação da luz captada com a visível pelo olho humano.

Além disso, o aumento da taxa de quadros, já comum em televisores comerciais, tende a se acelerar nos próximos anos. A percepção de que imagens com taxas de frames mais altas são mais confortáveis é um forte indicativo de que este recurso deveria ser amplamente desenvolvido tanto na produção, quanto na recepção. Dessa forma, mais um recurso de melhora da qualidade da experiência do usuário a partir de imagens técnica e esteticamente melhores tende a ser incorporado nos próximos 10 anos.

Essa evolução é mais avançada e rápida nos dispositivos móveis, que não dependem de padronizações relacionadas às telas de exibição. Dessa forma, *smartphones* e *tablets* naturalmente são ferramentas de teste e indução de inovações tecnológicas. Como o tamanho das telas é limitado pela portabilidade do dispositivo, a melhora da qualidade da imagem se torna ainda mais relevante do que na televisão ou outras telas maiores. Além da qualidade das imagens, os dispositivos móveis estão no centro da fruição pessoal no ecossistema audiovisual que será discutido no cenário dois.

1.6.1.1. Impacto narrativo

Como consequência deste cenário e considerando especificamente a evolução narrativa no entretenimento doméstico, que guia a maior parte da indústria do entretenimento, o Grupo de Estudo Design Audiovisual acredita que as próximas mudanças em 2, 5 e 10 anos na imersão de narrativa audiovisual não necessariamente seja conduzida por instalação de grandes telas em casas e apartamentos (cada vez menores e mais compactos) mas a condução de trilhas narrativas que possibilitem a imersão do usuário em narrativas microcômicas escolhidas para serem consumidas desde telas a partir de 30 polegadas à telões e estendendo a imersão da trama à equipamentos mobiles como *tablets* e *smartphones*.

Atualmente o cenário de narrativas de grandes produtoras de conteúdo, como a exemplo da Netflix, vê-se o surgimento de obras audiovisuais a partir de tramas centrais que ganham ramificações em séries subsequentes possibilitando uma extensão de historicidade e uma infinita possibilidades de enredos entrelaçados. A exemplo disso, temos a série *Breaking Bad* (2008-2013) considerada um filme de 62 horas, eleito o programa com mais reviews positivos da história da empresa, no qual não finaliza com o último episódio, muito pelo contrário, tem o enredo estendido a partir da criação da série do personagem secundário, Saul Goodman (Bob Odenkirk) destacando vários outros personagens presentes na trama primária, possibilitando assim o surgimento de um microcosmo da dramaticidade. Assim, é fato que hoje há a aceitabilidade do público na extensão de uma narrativa a partir de sua subdivisão formando trilhas em torno de uma história central, aprofundando seus detalhes de construção de personagens e aumentando a imersão e envolvimento de narrativa de uma determinada trama.

Mediante essas situações o Grupo acredita que, numa perspectiva de dois anos, esse tipo de estratégia de narrativa estendida com criação de microcosmos da dramaticidade ganhe mais volume em grandes empresas de produção de conteúdo voltado ao entretenimento do segmento do audiovisual, sendo consumidas em TVs de *led* de tamanhos variados conforme a condição do consumidor.

Em 5 anos, acreditamos numa ampliação de possibilidades na construção dessas narrativas. A indústria do entretenimento poderá passar a construir esses microcosmos da dramaticidade como narrativas que suportem a condição técnica de receptividade da historicidade de cada público. Por exemplo: o personagem Saul Goodman poderá ser sequestrado por um de seus clientes depois que perdeu uma causa e ter sido preso em uma cela pequena onde teve que criar técnicas para sair do isolamento e conseguir sobreviver. O enredo gera todo no interior da cela com imagens em detalhes e que podem ser assistidas em pequenos dispositivos. Essa narrativa está à parte da trama, foi um dos episódios de vida do personagem que não interfere na trama geral e que pode ser assistida de um equipamento mobile não exigindo qualidade superior. Numa tela maior, o episódio pode ser modificado para que essa mesma experiência do personagem possa acontecer numa selva com várias cenas de sobrevivência numa mata e seus vários desafios e belos cenários naturais e qualidade maior de imagem.

Em dez anos, com o avanço da tecnologia, essa criação de tramas gerais com imersão em microcosmos da dramaticidade poderá ser ainda mais real com a possibilidade de impressão de objetos que compõem o cenário da trama. Por exemplo: atualmente as telenovelas, filmes e produções de massa impactam costumes. Um exemplo disso é a moda. Pessoas observam modelos de roupas ou bijuterias usados

pelos personagens nas tramas e vão em busca de adquirir esses objetos nas lojas. Em uma década, acreditamos que a inteligência artificial será capaz de efetuar a leitura de qual enredo e por qual dispositivo o usuário fará a trilha por uma determinada historicidade, com base em seus gostos e disposição temporal para isso e mais: será possível imprimir produtos desejáveis que compõem o cenário daquela trama.

Além disso, acreditamos que a experiência do usuário pela imersão nesses microcosmos da dramaticidade desperte nele o interesse na participação da construção do cenário e o mesmo atue como co-criador dessa imersão alternativa. Por exemplo: se as gravações de um determinado episódio acontecem na Basílica de Nossa Senhora das Neves, em João Pessoa, um fã da série, pode ir até o local, registrar imagens do local que não foram mostradas durante a gravação do filme e disponibilizar para os outros usuários que não conhecem o lugar como ele. Essas imagens são enviadas para a plataforma e disponibilizadas como um novo microcosmo da dramaticidade criado a partir da colaboração do público. Essa disponibilidade materializa a participação do usuário na trama abrindo espaço para a interação dentro das condições de quem oferta o conteúdo e de quem o consome, não impactando a indústria de broadcast e ampliando seu consumo para todos os públicos.

1.6.2. Cenário 2: desenvolvimento da fruição

Avaliando os sistemas online, tecnologias e dispositivos de produção, transmissão e recepção de mídias audiovisuais, percebe-se uma inserção tecnológica diferente de cada meio. A difusão da tecnologia de realidade virtual, por exemplo, apesar de antiga, ganhou evidência fora do contexto acadêmico e experimental apenas nos últimos dois anos por meio dos novos dispositivos de interação, como óculos, em papéis de alto engajamento, como Jogador e Modificador.

Todavia essa experiência de uso, percepção de valor e utilidade ainda é rudimentar na Audiência. Nesse aspecto, percebe-se a movimentação de empresas e centros de pesquisa, como o Google e Facebook, em fomentar um percurso técnico e estético (na perspectiva de consumo de mídia) que parte do vídeo tradicional 2D, ou *flat* vídeo, para a experiência imersiva da realidade interativa e virtual. Essas empresas têm utilizado uma estratégia similar à introdução da GUI na informática, já que ao invés de apresentar um cenário completamente novo, têm construído um percurso técnico e de consumo progressivo de mídias espaciais, seja pelas fotos panorâmicas, vídeos em 360° e manipulações da mídia. Ou seja, existe uma construção de *affordances* midiáticas, que constroem percepções de valor e demanda de novos produtos e serviços, desenvolvendo, assim, a utilidade e qualidade do sistema. Apesar de citarmos um exemplo da transição entre o *flat* vídeo e a realidade virtual, é preciso destacar que esta evolução não significa o fim de qualquer mídia, mas a convivência de mais um recurso audiovisual dentro do ecossistema como um todo.

A partir das pesquisas realizadas, acredita-se que os sistemas audiovisuais futuros terão experiências estéticas cada vez mais entrelaçadas, aumentando a importância da manipulação de software e hardware via rede. Neste aspecto, percebe-se que a constante busca técnica por resolução de imagem e *frame rate* caminha para o estabelecimento de novos sentidos de mercado, alcançando não apenas como atributo técnico de reprodução ou transmissão de dados, mas sim como interface de interação

entre indivíduo nos diferentes papéis e contextos narrativos, seja para fins de entretenimento, técnico ou científico.

A capacidade de o sistema audiovisual dispor de *affordances* de interação em rede, permitirá desenvolverá novas formas de interação com a mídia, tanto no âmbito individual (um para um), quanto broadcast (um para muitos) e coletivo (muitos para muitos). Dados sobre preferências em comum podem gerar grupos de consumo e compartilhamento por meio de associações semânticas e significativas oriundas da constante mineração e geração de significados inovadores.

Dessa forma, este cenário não diferencia tipos de vídeo, resoluções, infraestruturas de transmissão ou exibição, e nem formas de interação. Consideramos que todas as tecnologias e meios de fruição poderão conviver, dependendo apenas das opções do indivíduo em cada papel. Para isso, descrevemos o cenário como uma forma de interação entre o indivíduo e mundo real, mediada pelas tecnologias de informação e de comunicação, focando nas oportunidades temporais e campo exploratório discutidas anteriormente. A análise das estratégias de historicidade e geratividade ficam como sugestão e trabalhos futuros desde já.

A Figura 10 mostra este cenário. Partindo do princípio de que o audiovisual é um mediador, consideramos as telas como interfaces dessa mediação. O mundo real, físico, pode ser percebido pelos indivíduos em 360 graus tridimensionalmente. Dessa forma, a interface mais simples de mediação é um sistema de projeção também de 360 graus, tridimensional. Pode ser uma caverna digital, totalmente imersiva, ou um *display* montado na cabeça do indivíduo (HMD, do inglês Head Mounted Display) e que simula esse ambiente. HMDs em forma de óculos de realidade virtual já são comuns no mercado, e tendem a ser foco da indústria de consumo nos próximos dois anos.

Seguindo a representação virtual do mundo real, onde é possível se afastar de um objeto para visualizá-lo melhor, o zoom pode ser utilizado neste ambiente de fruição. O afastamento da tela 360 aumenta a quantidade de objetos que podem ser exibidos. Dessa forma, neste cenário podemos ter inúmeras camadas de visualização, onde as mais afastadas, ou com menor zoom, se aproximam de uma experiência do mundo real. Por outro lado, as camadas mais próximas ou com maior zoom, aproximam o foco dos objetos com recortes mais próximos em exibições menores.

Neste contexto, a resolução ganha importância, apesar de relativa. Em telas pequenas as altíssimas resoluções não acrescentam muito valor para papéis menos engajados, como Audiência e Sintetizadores. 4K, por exemplo, precisa de telas acima de 120 polegadas para uma experiência plenamente perceptível nestes papéis. Por outro lado, considerando projetos pedagógicos e de arte, a visualização pode ser criada a partir de entradas de fluxos de vídeos de diferentes origens e resoluções, compondo uma experiência nova.

Dessa forma, é possível desenvolver usos que partem de ambientes complexos, mostrados integralmente nas camadas mais afastadas, cujos objetos podem ser aproximados através do zoom para as camadas mais próximas. Neste caso, pode inclusive ocorrer uma interação com dispositivos pessoais, como um smartphone. Parte do universo 360 pode ser representado por um vídeo *flat* nesta tela. Trata-se de uma interface pessoal de fruição.

Essa fruição pessoal pode, em algum momento, se tornar coletiva em uma tela dentro de casa, por exemplo. Neste caso, a camada de fruição passa a ser doméstica, com uma tela maior (uma TV na sala ou uma projeção na parede da sala). Entre essas duas camadas, uma intermediária representa uma tela de uso pessoal e compartilhado, alternadamente (uma TV no quarto ou na cozinha).

Estas duas camadas de fruição deverão ser foco do desenvolvimento tecnológico dentro do prazo de cinco anos. Recursos e ferramentas de interação entre o ambiente pessoal e doméstico devem ser introduzidos, especialmente relacionados a acesso e gestão dos conteúdos. Além disso, telas com resoluções superiores a 4K deverão ser lançadas, seguindo a linha de oferta de conteúdos captados por smartphones, cujas câmeras estão em processo de evolução rápida. Projetores e ambiente *home cinema* devem acompanhar este processo.

Em um primeiro momento, o papel do Jogador deverá ter destaque na disseminação deste ambiente, devido ao perfil inovador e curioso. Uma vez dominada a tecnologia por parte do Jogador, Modificadores terão a oportunidade de adequar a experiência para usos mais simples, despertando a participação dos sintetizadores.

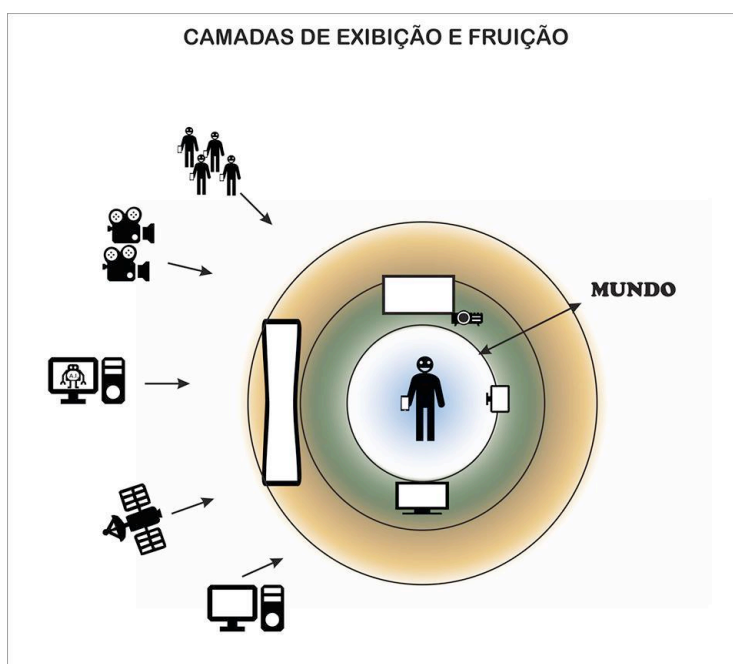


Figura 1.10. Camadas exibição e fruição.

Na terceira camada, a fruição é social. As telas são maiores, como numa sala de cinema ou projeções ao ar livre. Neste caso, a linha do conteúdo precisa considerar a fruição compartilhada com produtos adequados à telas maiores e altas quantidades de indivíduos compondo o papel da audiência simultaneamente em um mesmo lugar. A experiência da produção cinematográfica é central aqui, agregada de novos usos em ambientes externos. Neste caso, resoluções 8K ou maiores, têm demanda e acrescentam valor à experiência audiovisual.

Dentro deste contexto é importante considerar o uso pedagógico e científico. Neste ponto, este cenário converge com o anterior, uma vez que a qualidade da imagem é essencial para a correta visualização das informações, assim como são pertinentes a resolução e o tamanho da tela. Em serviços de telemedicina e telessaúde, por exemplo, há conteúdos que dependem de resolução e de profundidade da imagem, como segunda opinião e cirurgias realizadas remotamente. Dessa forma, altíssimas resoluções, com controles de zoom e *frame rate* no momento da exibição são recursos pertinentes.

Devido ao alto impacto científico e social, acreditamos que este tema se desenvolverá consideravelmente nos próximos anos. O barateamento dos equipamentos de captação de imagem em 4K, com lançamento de webcams e softwares de transmissão simples e robustos, tendem a disseminar essa tecnologia em um curto espaço de tempo, como dois anos. Em cinco anos, esses sistemas de captação deverão estar integrados aos focos cirúrgicos das salas de cirurgia, com recursos de operação simplificada a tal ponto de ser incorporada na atividade dos médicos e enfermeiros.

Em cinco anos também deverão começar os testes com 8K, integrados com novos sistemas de fruição, que não dependem de telas gigantes para visualização. Considerando as camadas de fruição, o vídeo 8K deverá chegar no dispositivo pessoal “fatiado”, de acordo com a demanda do indivíduo. A imagem poderá ter focos de interesse, que se adaptam em função do uso semântico da imagem.

Este cenário, considerado viável para um período de 10 anos, tem impactos incomensuráveis atualmente na infraestrutura. Narrativa e pedagogicamente, as potencialidades são diversas, dependendo da criatividade do Produtor. Já em termos de tráfego de dados, processamento de imagens, adaptação e recomendação de conteúdos ou de narrativas generativas, as consequências podem comprometer o desenvolvimento neste período de tempo. Apesar disso, trata-se de um uso viável neste período.

Sob o ponto de vista da produção e da narrativa, a Linha do Conteúdo pode ser desenvolvida de diferentes formas. No modo mais simples, o papel é exercido por uma entidade apenas, que pode ser uma pessoa, uma empresa ou um conjunto de empresas gerando um conteúdo oferecido de forma integrada à Audiência. É como acontecem as produções televisivas e cinematográficas atualmente.

Por outro lado, a produção pode ser compartilhada ou oferecida simultaneamente para a audiência, que faz o filtro do que quer visualizar. Em um cenário extremo, milhares de vídeos de diferentes origens e resoluções podem estar disponíveis na camada de visualização e fruição. Ao invés da troca de canais ou do acesso a diferentes canais do Youtube sequencialmente, todos os fluxos podem estar disponíveis simultaneamente neste ambiente e o indivíduo escolhe qual visualizar através do zoom.

Dentro deste contexto, pode ocorrer uma troca entre os papéis do Produtor e da Audiência, que se mesclam. Se considerarmos que todos os indivíduos com smartphone são potenciais produtores e que esta produção pode ter relevância em determinados contextos, o fluxo de dados e de processamento tende ao infinito. Como exemplo, podemos considerar uma partida de futebol, onde milhares de pessoas captam e transmitem a partida com seus smartphones. Todos os fluxos de vídeo gerados podem chegar na primeira camada de interação, sendo filtrados e apenas os mais relevantes chegam na camada doméstica ou pessoal. O mesmo exemplo pode ser expandido para

videoconferências com milhares de pessoas ou sistemas de segurança onde todas as câmeras disponíveis representam fontes de informação. No entanto, este cenário tende a ser restrito a pequenas interações dentro do prazo de 5 anos, devido ao enorme poder de processamento computacional e tráfego de dados demandados. Em 10 anos, com o desenvolvimento de políticas de gestão para tratar este tipo de tráfego em rede, com o aumento da velocidade e do poder de processamento computacional disponível, é possível a materialização deste tipo de sistema.

Finalmente, é preciso ressaltar que boa parte da confirmação de todo este cenário depende da interação. Ou seja, da capacidade de fruição que indivíduo terá diante do ecossistema audiovisual. Essa relação pode ser representada em termos de qualidade, onde a experiência do indivíduo é tão, ou mais, relevante quanto a valor simbólico dos conteúdos. Como exemplo, podemos citar os HMDs necessários à experiência imersiva em ambientes tridimensionais, que comprometem a boa experiência do indivíduo ao limitarem a fruição. O uso de óculos ou outros sistemas de imersão tira parte da liberdade de movimentos. Dessa forma, a imersão em ambientes virtuais domésticos depende da evolução das ferramentas de interação, especialmente da computação vestível.

1.7. Conclusão

O presente estudo sistematizou práticas que indicam potencialidades de uso do tempo e espaço para a narrativa audiovisual contemporânea. A junção de diferentes recursos, configurações e agrupamento de formatos de interface e estética podem gerar sistemas audiovisuais, viabilizando uma infinidade de recursos narrativos. Toda essa atribuição de significados na relação entre a interação entre sistemas computacionais, narrativa e usuário atuam como elemento integrador e catalizador da fruição do conteúdo.

As estratégias oportunidades temporais; campo exploratório, historicidade; geratividade, evidenciam que a manipulação do tempo e espaço instauram novas instâncias de narração e fruição. Essa “nova” fruição não se limita a um conjunto de ações para interpretação de outrem, mas sim, como produto da interação entre narrativa, usuário e sistema.

Os experimentos audiovisuais discutidos tornam-se paulatinamente sistemas audiovisuais complexos, a medida que incorporam as estratégias e experiências de interação identificadas. Em comum, os objetos analisados possuem três características: 1) O posicionamento ativo do usuário; 2) A capacidade de manipular, adaptar e reordenar a estrutura de acordo com ações de seu contexto de consumo; 3) A convergência de novos elementos e apropriações estéticas.

Analisando possíveis evoluções da narrativa audiovisual, algumas características podem ser identificadas com base no levantamento realizado entre referencial, estratégias e produções técnicas. A exploração, tanto de oportunidades temporais quanto de campos já vistos e não vistos, é um elemento central no desenvolvimento de novas práticas estéticas e narrativas. Além disso, a capacidade dos sistemas em dialogar com a historicidade do usuário, e deste modo, ofertar um conteúdo personalizado através da relação construída na plataforma, incita a construção de hiper-histórias. Do mesmo modo, a geratividade de conteúdo individualiza a experiência fílmica, criando obras esteticamente diferentes a cada visualização. Neste ponto, pode-se inclusive questionar conceitos consolidados na teoria cinematográfica. Temas centrais, como filme e cinema,

carecem de releituras analíticas no âmbito das novas tecnologias e do pós-cinema. Com base nas análises percebe-se que o tempo e espaço atuam como interface do processo de fruição nas mídias interativas, isto é, são mediadores do processo interacional percorrido pelo usuário para compreender a narrativa.

Em síntese das análises desenvolvidas neste trabalho, compreende-se que uma nova estrutura surge oriunda dos processos de manipulação temporal e espacial. Essa nova estrutura percebe a narrativa como um campo de possibilidades que são potencializadas por um sistema computacional, a medida que oferta experiências de interação de imergir e emergir da tela. Dessa forma, as potencialidades identificadas no presente trabalho expandem a capacidade de contar histórias (e compreendê-las) quando ofertadas em paralelo, podendo ainda agregar valor se somadas a campos de atuação como realidade virtual, neurocinema ou cinema quântico e internet das coisas.

Observamos, porém, que os usos de estratégias de manipulação do tempo e espaço não concluem o estudo sobre interação em mídias digitais audiovisuais. É natural que o surgimento de novas obras amplie as fronteiras do gênero, e funções diferentes poderão ser encontradas de acordo com esse avanço. Além disso, as funções levantadas foram baseadas nos exemplos analisados neste trabalho, por isso, não acreditamos que todas as possibilidades de desenvolvimento tenham sido contempladas. Estudos futuros podem revelar outras funções além das apresentadas nesta pesquisa. Entretanto este trabalho compreende que a manipulação temporal e espacial pelo usuário, instaura novos modos de produção e compreensão da narrativa audiovisual digital interativa.

Referências

- Alvarenga, A. L. De (2011). A Geografia Criativa do Cinema: O Papel da Montagem na Construção dos Espaços Fílmicos. Espaço Aberto, v. 1, n. 2, p. 39–54.
- Aly, N. (2012). Desdobramentos contemporâneos do cinema experimental. Teccogs, v. 6, p. 61–92.
- Aumont, J. and Marie, M. (1993). Dicionário teórico e crítico de cinema. Tradução Eloisa Araújo Ribeiro. Papyrus ed. Campinas, SP.
- Becker, V.; Gambaro, D. and Ramos, T.S. (2017). Audiovisual Design and the Convergence Between HCI and Audience Studies. In: International Conference on Human-Computer Interaction. Springer, Cham, p. 3-22.
- Berga, Q. et al (2016). Case study of a generative editing audiovisual project. X Computation Communication Aesthetics & x. Anais... Bergamo, Itália.
- Block, B. (2010). A narrativa visual: criando a estrutura visual para cinema, TV e mídias digitais. Tradução Cláudia Mello Belhassof. São Paulo: Elsevier.
- Buarque, S. C. (2003). Metodologia e técnicas de construção de cenários globais e regionais. Brasília: IPEA.
- Cannito, N. G. (2010). A televisão na era digital: Interatividade, convergência e novos modelos de negócio. Elsevier ed. São Paulo.

- Carlier, A. et al. (2011). Combining content-based analysis and crowdsourcing to improve user interaction with zoomable video. Proceedings of the 19th ACM international conference on Multimedia - MM '11, p. 43.
- Carreras, Q. B. and Minguillon, J. (2015) An online short-film editing machine with a fixed structure and pseudo-infinite combinations. GA2015 – XVIII Generative Art Conference.
- Cirino, N. N. (2012). Cinema interativo: Problematizações de linguagem e roteirização. Recife: Universidade Federal de Pernambuco.
- Costa, A. (2013). Compreender o cinema. Tradução Nilson Moulin Louzada. São Carlos, SP: v. 3.
- Deleuze, G. (2005). A imagem tempo. Tradução Eloisa de Araujo Ribeiro. Brasiliens ed. São Paulo: v. Cinema II.
- Desom, J. (2017). Rear Window Timelapse. Disponível em: <<http://jeffdesom.com/hitch/>>. Acesso em: 28 ago. 2017.
- Dubois, P. (2017). Aula Magna com Philippe Dubois. Disponível em: <<https://www.youtube.com/watch?v=bq8nYh6DWhA>>. Acesso em: 28 ago. 2017.
- Dubois, P. (2014). A questão da “forma-tela”: espaço, luz, narração, espectador. In: Gonçalves, O. *et al.* (Ed.). . Narrativas Sensoriais: Ensaio sobre cinema e arte contemporânea. 1. ed. Rio de Janeiro: Editora: Circuito. p. 123–157.
- Eigenfeldt, A. et al. (2015). Applying Valence and Arousal Values to a Unified Video, Music, and Sound Generative Multimedia Work. XVIII Generative Art Conference.
- Eisenstein, S. (2002). A forma do filme. Tradução Teresa Ottoni. Rio de Janeiro: Jorge Zahar Editor.
- Figueiredo, C. D. de and Mendes, A. R. da S. (2015). Roteiros para dispositivos de mídias móveis: Tela, tempo e trânsito como elementos contingentes. *Geminis, Abordagens multiplataformas*, v. 6, n. 2, p. 165–182.
- Galanter, P. (2004). Generative art is as old as art. An interview with Philip Galanter. *artificial.dk*.
- GAUDENZI, S. *The Living Documentary: from representing reality to co-creating reality in digital interactive documentary*. London,: University of London, 2013.
- Gosciola, V. (2008). Roteiro para as novas mídias - Do cinema às mídias interativas. 2. ed. São Paulo: Senac - SP.
- Lévy, P. (1996). *O que é o virtual?* Tradução Paulo Neves. São Paulo: Ed. 34.
- Lopes, T. R. C.; Montañó, S. and Klipp, S. (2014). Montagem espacial e potencialidades do audiovisual locativo no cenário urbano *Spatial*. *Ecopós*, v. 17, n. 2, p. 1–11.
- Maciel, K. (2004). Por um cinema sensorial: o cinema e o fim da “moldura”. *Conexão – Comunicação e Cultura UCS*, v. 3, n. 6, p. 61–71.
- Maciel, K. (2009). A Forma Cinema: Variações e Rupturas. In: *Transcineamas. Contra cap ed. Rio de Janeiro: [s.n.]. p. 431.*

- Mackin, A.; Noland, K. C. and Bull, D. R. (2017) High Frame Rates and the Visibility of Motion Artifacts. *SMPTE Motion Imaging Journal*, v. 126, n. 5, p. 41–51.
- Manovich, L. (2001). *The Language of New Media*. MIT press.
- Manovich, L. (2014). Efeitos de escala. *Revista Trama interdisciplinar*, v. 5, n. 1, p. 107–114.
- Médola, A. S. L. D. and Oliveira, B. J. de. (2016). Audiovisual panorâmico para dispositivos móveis: reconfigurações em tela , imersão e instância instance. Associação Nacional dos Programas de Pós-Graduação em Comunicação XXV Encontro Anual da Compós. Anais...Goiás, Goiânia: XXV Encontro Anual da Compós.
- Montez, C. and Becker, V. (2005). *TV Digital Interativa: conceitos, desafios e perspectivas para o Brasil*. Ed. da UFS ed. Florianópolis: Editora da UFSC Florianópolis. v. 2
- Murray, J. H. (2003). *Hamlet no Holodeck - o futuro da narrativa no ciberespaço*. Tradução Elisa Khoury Daher; Marcelo Fernandez Cuzziol. São Paulo: Itáu Cultural: UNESP.
- Newton, K. and Soukup, K. (2017). *The Storyteller's Guide to the Virtual Reality Audience*. Disponível em: <<https://medium.com/stanford-d-school/the-storyteller-s-guide-to-the-virtual-reality-audience-19e92da57497#.ikwe1f78x>>. Acesso em: 28 ago. 2017.
- Nogueira, L. (2010). *Planificação e montagem*. 1. ed. Covilhã: Série: Estudos em comunicação. v. 1
- Percy, M. (2017). *The Internet Native Filmmaker's Manifesto*. Disponível em: <<https://web.archive.org/web/20130517090205/http://internetnativemanifesto.posterous.com/>>. Acesso em: 28 ago. 2017.
- pimenta, F. (2001). O conceito de Virtualização de Pierre Lévy e sua aplicação em hipermídia. *Lumina*, v. 4, n. 1, p. 85–96.
- Primo, A. (2000). Interação mútua e reativa: uma proposta de estudo. *Famecos*, v. 12, p. 81–92.
- Rafaeli, S. and Sudweeks, F. (1998). Interactivity on the net. *Network and Netplay: Virtual Groups on the Internet*, p. 173–190.
- Redi, J.; D'acunto, L. and Niamut. (2015). O. Interactive UHDTV at the Commonwealth Games. *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video - TVX '15*. Anais...New York, New York, USA: ACM Press.
- Santaella, L. (2005). *Por que as comunicações e as artes estão convergindo?* Coleção Qu ed. São Paulo.
- Sawhney, N.; Balcom, D. and Smith, I. (1997). Authoring and navigating video in space and time. *IEEE Multimedia*, v. 4, n. 4, p. 30–39.

- Tag, B. et al. (2016). In the Eye of the Beholder. Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '16. Anais...New York, New York, USA: ACM Press.
- Tarkovski, A. (1998). Tarkovski - Esculpir o tempo. Tradução Jefferson Luiz Camargo. 2. ed. São Paulo: Livraria Martins Fontes. v. 1
- Tompkin, J. et al. (2013). Video collections in panoramic contexts. Proceedings of the 26th annual ACM symposium on User interface software and technology - UIST '13. Anais...New York, New York, USA: ACM Press.
- Toscano, R. and Becker, V. (2015). Potenciais narrativos de interatividade para vídeos de altíssimas resoluções no campo da videocolaboração. Rio de Janeiro: II Workshop: O Futuro da Videocolaboração, RNP.
- Toscano, R. and Becker, V. (2016). Sistemas audiovisuais inteligentes: um levantamento das práticas que apontam a dinâmica da narrativa responsiva e generativa. XXXIX Congresso Brasileiro de Ciências da Comunicação. Anais...Caruaru, PE: Intercom.
- Van Der Heijden, K. (1996). Scenarios: the art of strategic conversation. Nova York: John Wiley & Sons.
- Watson, A. B. (2013). High Frame Rates and Human Vision: A View through the Window of Visibility. SMPTE Motion Imaging Journal, v. 122, n. 2, p. 18–32.

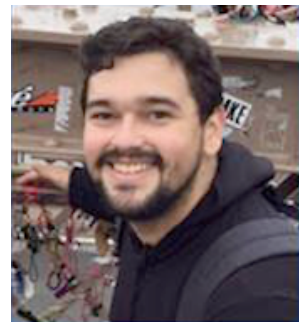
ANEXO 1: biografia dos autores

<p>Rafael Toscano da Nóbrega - Designer de comunicação, graduado em Comunicação em Mídias Digitais pela Universidade Federal da Paraíba (UFPB). Mestrando no programa de Pós-Graduação em Computação, Comunicação e Artes (PPGCCA) da UFPB. Bolsista no laboratório de aplicações em vídeo digital (Lavid) e colaborador do Projeto de pesquisa e extensão DIAS. Desenvolve estudos por meio de dois grupos de pesquisa, Design Audiovisual do DSC/UFPB e Comunicação, Tecnologias e Multimídias para a Educação do DEMID/UFPB. É autor artigos científicos em áreas relacionadas a sistemas audiovisuais, interação e novas mídias.</p>	
<p>Valdecir Becker - É jornalista, mestre em Engenharia e Gestão do Conhecimento (2006,UFSC) e doutor em Ciências (Engenharia Elétrica, 2011, USP). Professor no Centro de Informática e nos Programas de Pós-graduação em Jornalismo (PPJ) e Pós-Graduação em Computação, Comunicação e Artes (PPGCCA), da Universidade Federal da Paraíba (UFPB). Também é coordenador do PPGCCA, líder do Grupo de Pesquisa Design Audiovisual e pesquisador do Lavid - Núcleo de Aplicações de Vídeo Digital. Ao longo dos anos, participou de vários projetos acadêmicos relacionados à digitalização da TV, estudando novos formatos de conteúdos convergentes e multiplataformas e os impactos das tecnologias digitais nos modelos de conteúdo e de negócio. É autor de livros e artigos científicos sobre TV digital, interatividade, audiência, mercado audiovisual e IHC.</p>	
<p>Laíz Silveira Ferreira - Graduação em Comunicação Social - com habilitação em Relações Públicas pela Universidade Federal da Paraíba, Especialização em Telejornalismo pela Fesp Faculdades e Mestrado em Jornalismo pela UFPB. Atualmente cursa especialização em Tecnologias e Inovações para a Web no Centro Universitário Senac Santo Amaro -SP, é Bolsista do CNPQ como membro do grupo de pesquisa Design Audiovisual do DSC/UFPB e Comunicação, é proprietária e diretora de conteúdo da empresa especializada em produção de conteúdo jornalístico digital de Turismo: Jornalismo de Viagem e corporativo B2B conteúdos. Professora substituta do curso de Relações Públicas da UFPB.</p>	

Samara Coutinho - Estudante de Relações Públicas, pesquisadora do Grupo de Pesquisa Design Audiovisual, focando em temas como interação, interatividade e modelo de negócios.



Leonardo Burgos - Concluiu em 2011 graduação em Comunicação Social com Habilitação em Jornalismo na Universidade Católica de Pernambuco (Unicap). É especialista em Direitos Humanos pela mesma universidade desde 2014. Atualmente, é aluno do Mestrado Profissional em Jornalismo pela Universidade Federal da Paraíba (UFPB), onde começou a estudar em 2016. Atua como Editor de Texto da Rede Globo Nordeste desde 2012.



Capítulo

2

Cinema em Definição

Jane de Almeida¹, Maria Amelia Eliseo¹, Helena Prates¹, Vic Von Poser¹, Patricia Gimenez¹, Amanda Areias¹, Leandro Oliveira¹, Mateus Teixeira¹, Cicero I. da Silva²

¹ Universidade Mackenzie: Laboratório de Artes Cinemáticas e Visualização (LabCine)

² Universidade Federal de São Paulo. NTC Telessaúde Redes (Telessaúde/Unifesp)

janeidealmeida@post.harvard.edu, mamelia@gmail.com, helena.prates@gmail.com, vicvonposer@gmail.com, patigimenez@gmail.com, amandaareias@givemeflour.com, leaoliveira@me.com, mateushrteixeira@gmail.com, cicero.silva@unifesp.br

Resumo

Este artigo apresenta a produção de quatro anos do Laboratório de Artes cinemáticas e Visualização, LabCine, da Universidade Presbiteriana Mackenzie e os projetos em elaboração pensados como projetos futuros com potencial de videocolaboração. Com o propósito de refletir sobre a imagem em movimento no contexto digital, o LabCine incentiva a produção de filmes e múltiplas derivações cinemáticas para propor soluções a partir de um contexto histórico e cultural das tecnologias atuais. O artigo, portanto, apresenta a pesquisa histórica de tecnologias exploradas pelos projetos para descrever o percurso de cada proposta. A intenção do Labcine é propor a aprendizagem pela prática, mas em consonância com a pesquisa internacional pensando em tecnologias futuras.

Abstract

This paper presents the four-year production of the Laboratory of Cinematic Arts and Visualization, LabCine, from Mackenzie Presbyterian University and the projects under development intended as future projects with the potential of video collaboration. In order to reflect on the moving image in the digital context, LabCine encourages the production of films and multiple kinematic derivations to propose solutions from a historical and cultural context of the current technologies. The article, therefore, presents the historical research of technologies explored by the projects to describe the course of each proposal. The intention of LabCine is to propose learning by practice, but in line with international research thinking on future technologies.

2.1. Introdução

O cinema digital de altíssima definição ainda é um desafio para ser produzido, não apenas por causa da quantidade de dados envolvidos em todo o seu processo – da captação à exibição, mas também por causa do esforço narrativo e imagético, considerando que cada materialidade tecnológica requer a construção de uma linguagem própria. Essa é a premissa da qual parte o LabCine, Laboratório de Artes Cinemáticas e Visualização da Universidade Presbiteriana Mackenzie e sua principal linha de investigação, Linguagens e Tecnologias, que oferece pontes entre o universo da imagem em movimento e o da computação¹. Considerando a complexidade contemporânea do processo de produção do cinema digital de super-alta definição, nossas pesquisas pretendem também enfrentar questões a respeito das padronizações, das dificuldades de arquivamento e exibição e, conseqüentemente, das políticas públicas a serem adotadas a respeito destes critérios. Dentro dessa perspectiva, uma questão crucial para as pesquisas do Labcine é que não se conhecem ainda as limitações do “digital”, especulando assim se este limite partirá da estabilidade dos sistemas computacionais ou da capacidade da visão humana.

O capítulo proposto pelo grupo de pesquisa do LabCine para esta coletânea apresenta portanto questões a respeito do embate entre o cinema, a imagem estática e em movimento e suas relações com a computação. Mais especificamente, descreve como o LabCine tem lidado com tais questões, às vezes propondo projetos que tentam avançar a discussão de forma teórica e outras vezes por meio de projetos que apresentam propostas práticas de soluções tecnológicas para determinados impasses apresentados pelo digital.

Uma das metodologias do LabCine é conhecer a história das tecnologias envolvidas durante o processo de pesquisa e então contextualizá-las a respeito do estado da arte internacional, para daí propor soluções. No entanto, tais soluções devem ser adequadas à realidade brasileira e aos limites das tecnologias e do conhecimento dos envolvidos. Algumas pesquisas do LabCine, ao apresentarem a arqueologia das tecnologias envolvidas, apenas sugerem caminhos, como exercício criativo para os impasses apresentados.

O artigo está dividido em duas partes: uma sobre projetos realizados e em desenvolvimento, e outra sobre projetos futuros. Na primeira parte, são apresentados os filmes estereoscópicos, o projeto Marte, a pesquisa e filme Pixel Race, os fast-frame films e o SenseMaking. Na segunda parte chamada Cinemas Futuros (ou Visão do Futuro) são apresentados os projetos para Filmes de Bancos de dados, Fotogrametria e o Streaming de orquestra. Estes projetos refletem o que o Labcine pensa sobre um certo futuro das tecnologias de imagem em movimento no contexto do digital e também da videocolaboração.

2.2. Parte Um: filmes de projeto

Esta parte do artigo é composta pela descrição da produção de filmes e os desafios enfrentados durante esses processos. Os filmes produzidos pelo Labcine apresentam a característica pouco comum na tradição brasileira, pois são “filmes de projeto”, feitos em condições de laboratório de pesquisa, envolvendo pesquisadores-professores, além de

¹ Ver website do LabCine: www.labcine.com

mestrandos e doutorandos. No entanto, parte da equipe é composta por profissionais de mercado, gerando um grupo peculiar de profissionais, pesquisadores e estudantes envolvidos em uma troca de conhecimento fértil, dinâmica e interdisciplinar. Como a maior parte dos projetos tem sido realizada com parceiros de universidades ou profissionais do exterior, tratam-se de pesquisas e trabalhos com cunho internacional, com capacidade de competição e projeção em grupos de pesquisa bastante avançados no exterior.

Nesta parte, serão apresentados os projetos: “EstereoEnsaio Rio de Janeiro” (2011), “EstereoEnsaio São Paulo” (2017), “Pixel Race” (2014), “Openmouthed” (2015), “O Mergulho” (2015), “Cirurgia Oftálmica” em 4k a 1000 fps (2015), “Projeto Marte” e “Projeto Sensemaking”.

2.3. Estereocopia de super-alta definição

Para refletir sobre os limites do sistema digital de super-alta definição, o retorno da estereoscopia, agora como efeito do cinema digital, apresenta-se como uma forma singular de oportunidade, pois é a antiga técnica com as possibilidades infinitas da imagem digital e de alta definição. O LabCine produziu dois filmes estereoscópicos (em 3D) em super-alta definição desde a captação até a projeção e transmissão, passando por todas as etapas do processo digital de execução filmica.

O primeiro filme do LabCine, realizado em 2011, junto com o Grupo de Trabalho “Aplicações Avançadas de Visualização Remota” da RNP (Rede Nacional de Ensino e Pesquisa), chama-se *EstereoEnsaio : cinco ou seis ensaios estereoscópicos à procura de uma narrativa* e captou cenas clichês do Rio de Janeiro. Como desafio tecnológico, o filme partiu da proposta de realizar, produzir, sistematizar e catalogar o processo de criação de conteúdo audiovisual em formato 4K 3D (imagem com definição mínima de 4096×2160 pixels). Os resultados dessa pesquisa encontram-se no endereço www.estereoensaios.com.br.

O projeto deveria produzir um filme com a duração de 8 a 10 minutos no padrão 4096×2160 pixels por frame e disponibilizar um relatório circunstanciado relativo a todas as fases de produção do conteúdo, tais como sistema de câmeras em 4K, estereoscopia, produção, captação, finalização e formatos de output de vídeos (JPEG 2000, H.264, MPEG-4), além de testes e escolhas de formatos mais adequados à transmissão de conteúdos que poderiam ser transmitidos por redes fotônicas de alta performance, propiciando testbed para as redes. A partir deste conteúdo e relatório seria possível criar massa crítica, conhecimento do processo e *know how* científico e tecnológico no campo do cinema e das redes fotônicas.

“EstereoEnsaio” foi pensado como um ensaio audiovisual de tecnologia que explora imagens em ultra-definição. Assim como no lendário “Vistas da Bahia da Guanabara” (1898), dos irmãos Segreto, o Rio de Janeiro foi novamente a cidade para onde câmeras pioneiras se dirigiram. Com a diferença de que, agora, as “vistas” podem ser observadas em três dimensões e altíssima definição. O filme foi dividido em 5 atos, além de um prólogo e um epílogo. O prólogo e o Ato I resgatam mais explicitamente a ideia de “vistas” do Rio de Janeiro, planos do Pão-de-Açúcar, da ponte Rio-Niterói e da Baía de Guanabara, em tomadas aéreas e marítimas. Caminhões, guindastes e navios – todos os elementos em dimensões gigantescas e plenos de movimento. Em 3D, o movimento contínuo das máquinas molda os planos que configuram o espaço

estereoscópico. Como disse o estereografista americano Keith Collea sobre as máquinas: “vejam, elas trabalham para nós!”. Em oposição, os contêineres coloridos e gigantes formam uma cadeia de fundo, fechados em seus segredos sobre os objetos que trazem de regiões distantes. O ato III explora as ruas labirínticas da comunidade Tavares Bastos e celebra os jogadores locais de futebol. Os pés, a bola, os movimentos fortuitos dos garotos são elementos de composição dos planos que se misturam aos movimentos da dança do Ato V. Os pés, as mãos e os movimentos dos bailarinos da Mangueira se integram aos movimentos da bola, em um jogo desprezioso e espontâneo, cujo principal objetivo é promover uma nova experiência visual. Todo o filme é envolto por uma trilha sonora que coloca o espectador de fora do visor estereoscópico, observando as cenas belas e marcantes da memória do Rio de Janeiro e do Brasil. A trilha foi composta por Livio Tragtenberg e o *design* de som por Beto Ferraz, dois profissionais reconhecidos da indústria sonora do cinema.

O projeto, desde o seu início, enfrentou os desafios tecnológicos dos grandes laboratórios de pesquisa. Com um par único de câmeras, equipamento ainda desconhecido dos especialistas, passando pelo processamento de imagens, com cerca de 10 milhões de pixels de definição, até a exibição do filme em projetores especiais, ainda em fase de estabilização. Na projeção estereoscópica do filme (3D), a resolução chega a 20 milhões de pixels por frame na tela, somando as imagens correspondentes aos olhos esquerdo e direito. Este experimento remonta aos desafios enfrentados pelos pioneiros do cinema e recupera a fantasia dos aparatos estereoscópicos do século XIX. Em 1891, Edson já tinha a intenção de dar profundidade estereoscópica às imagens do Kinetógrafo que seriam vistas pelo Kinetoscópio. Em 1856, com apenas dois anos de existência, a London Stereoscopic Company chegou a vender um milhão de cartões fotográficos estereoscópicos (Adams, 2003). Desde meados do século XIX, a fotografia estereoscópica trazia uma experiência nova de percepção da realidade, oferecendo aos espectadores um novo nível de imersão nas imagens, como se estivessem viajando a lugares remotos e participando pessoalmente de eventos distantes².

2.3.1. Arqueologia da estereocopia cinematográfica

Vistas estereoscópicas da cidade e outras localidades já eram exibidas ao público em 1905, e o panorama fotográfico Fotoplastikon Warszawski na Al. Jerozolimskie 51, em Varsóvia é a única atração do gênero que ainda funciona em seu endereço original. Em 1935, Louis Lumière refez a *Chegada do trem à estação* (“L’arrivée d’un train en gare de La Ciotat”, 1895) em versão estereoscópica, *remake* que talvez tenha sido o verdadeiro responsável pelo famoso espanto com o realismo das imagens que geralmente se atribui à projeção pública de *L’arrivée d’un train en gare de La Ciotat*, em 1896.

O historiador e filósofo alemão, Walter Benjamin relata mais de uma vez a sua experiência estereoscópica com o Kaiserpanorama em Berlim em 1900, principalmente em seu livro “A infância em Berlim por volta de 1900” (Benjamin, 2012). O pensador alemão comenta o fascínio do público com as fotografias de viagens e usa o estereoscópio como metáfora da nova configuração da modernidade, com imagens “dispostas como se saíssem de gavetas”, cenas e objetos de colorido intenso em oposição a um fundo com

² O pesquisador Gavin Adams organiza um compilado referencial sobre o tema da estereoscopia: <http://www.scielo.br/pdf/anaismp/v6-7n1/10.pdf>

fortificações e edifícios opressores, herança czarista na cidade de Riga. Mais tarde, Benjamin chama a atenção para o tempo dos detalhes que as novas cenas tridimensionais exigem, e sugere que o “olhar estereoscópico” deva ser cultivado para o enfrentamento em profundidade das “sombras históricas”. Toda essa “herança estereoscópica” das imagens fixas e em movimento inspira o projeto de “EstereoEnsaaios”: cenas que nos remetem à historicidade do aparato com “vistas”, o movimento fortuito da natureza, das máquinas e do corpo humano, plenos de sensualidade frente ao dispositivo estereoscópico.

Se antes havia o “cinema-olho” (no singular), lembrando do termo cunhado pelo cineasta soviético Dziga Vertov, e a centralidade da perspectiva monocular que acompanha a sua história, a imagem estereoscópica permite que se perceba o mundo visual ao nosso redor através de uma “stereopsis”. Em outras palavras, trata-se de um deslocamento da câmera de cinema, cíclope (um só olho), para um olhar de câmeras duplas, *estereóptico*. Nesse sentido, como pensar em todo um novo mundo das imagens em movimento mecânica e eletronicamente captadas a partir de uma perspectiva binocular? Em termos mais simples e diretos, como filmar e montar um filme tridimensional, narrativamente falando? O que nossos olhos, saturados por mais de 110 anos de imagem em movimento, suportam – e esperam – das imagens revisitadas pela estereoscopia em dimensões colossais?

Antes de tudo, a questão-tema de “EstereoEnsaaios” é a seguinte: o 3D é uma linguagem ou um efeito? Sabe-se que a estereoscopia, ou técnica popularmente chamada de 3-D é a produção da ilusão da profundidade, utilizando 2 imagens do mesmo tema, filmadas e exibidas simultaneamente e processadas pelo cérebro para a construção da profundidade. O historiador americano Ray Zone (1947-2012), conhecido como o pioneiro nos métodos de convergência de imagens, em seu livro *3-D Revolution*, define em 4 separadas eras a evolução da estereoscopia cinematográfica, a primeira a *Novidade*, quando surge a tecnologia, seguida pela Era da Convergência, marcada pela década de 1950 com o surgimento de novas tecnologias e o grande sucesso de público; a Era da Imersão, marcada pelo surgimento dos grandes formatos e a quarta o Cinema 3-d Digital, com a mudança de película para digital, possibilitou melhor qualidade técnica nos filmes e reduziu os custos. Existem ainda autores que acreditam que a única maneira de usar o 3-D na indústria cinematográfica é pela via econômica e outros ainda profetizam que a estereoscopia está condenada a uma vida cíclica em vir e ir a cada “nova geração de espectadores” (Doherty, 2011).

2.3.2. Produção estereoscópica atual

As mesmas questões apresentadas pela estereoscopia de alta definição do filme de 2011 apresentam-se agora no filme “EstereoEnsaaios São Paulo”, sendo ainda finalizado. Trata-se de um projeto realizado com apoio do PROAC – Governo do Estado de São Paulo e apoiado pela EletroPaulo, pela lei de isenção fiscal. Desde o início, o filme foi concebido com pesquisadores e profissionais de cinema.

Diferentemente do filme anterior, não são mais os clichês da cidade do Rio de Janeiro que estão sendo explorados, mas o aspecto de gigantismo da cidade de São Paulo, cuja inspiração é um filme realizado em 1929, chamado “São Paulo, Sinfonia da Metrôpole”. No final da década de 1920, dois diretores, Rudolf Rex Lustig e Adalberto Kemeny, inspirados no gênero “filme-sinfonia” eternizaram uma São Paulo que vivia ainda a expansão econômica do ciclo do café e a ascensão do Modernismo. “São Paulo

Sinfonia da Metrópole” é um longa-metragem que mostra até hoje como era a cara da maior metrópole brasileira no ímpeto de sua industrialização. No início do século XX, diversos marcos cinematográficos tiveram a cidade grande como objeto de sua arte. Um gênero que fez arte das cenas das metrópoles foi o documental, com as conhecidas “sinfonias da cidade” que apresentavam suas imagens em movimento com som ao vivo de orquestras. Manhattan, Paris, Moscou e Berlim tiveram seus filmes-sinfonia³.

Em 2016, quase um século após a estreia de “São Paulo Sinfonia da Metrópole” (1929)⁴, o curta-metragem “EstereoEnsaio São Paulo” foi realizado também com as mesmas condições tecnológicas do filme de 2011. O filme revisita a obra de 1929 fazendo da estereoscopia uma proposta de linguagem. O curta-metragem se utilizou da tecnologia de câmeras 5K 3D (super-alta-definição e estereoscópica) para explorar o gigantismo da capital paulista. O filme tem a duração de 15 minutos e tem a intenção de ser exibido com orquestra ao vivo, composta por músicos imigrantes da cidade de São Paulo.

Em sua composição técnica, utilizou-se um par de câmeras Red Epic suportadas e posicionadas em um *rig*, ferramenta que segura e estabiliza as duas câmeras, para produzir a estereoscopia. Desta forma, o trabalho foi realizado no limiar da tecnologia fílmica e, ao mesmo tempo, mais uma vez, em condições muito parecidas com a dos pré-cinemas e/ou do cinema silencioso. Algumas cenas de “São Paulo Sinfonia da Metrópole” estão presentes na montagem de “EstereoEnsaio SP”, proporcionando ao espectador apreciar, analisar e comparar os comportamentos de duas épocas distintas da mesma metrópole. No início do século XXI, assim como há quase cem anos, cenas da cidade de São Paulo foram registradas, com o objetivo de se resgatar a memória paulistana e deixar um legado para futuras gerações, explorando a mais alta tecnologia cinematográfica atualmente existente.

A cidade de São Paulo propicia imagens com incontáveis “camadas” (níveis ou planos) com seus inúmeros edifícios, pessoas e carros. Estas camadas são essenciais para se explorar a linguagem 3D. Um dos procedimentos estéticos do documentário é a montagem poética que privilegia os ritmos mais do que os sentidos das imagens. Não se pretende divulgar ideologicamente a cidade, mas trazê-la em relevo estereoscópico como o documento de uma época⁵.

³ Os dois curtas EstereoEnsaio foram inspirados nos filmes: “Berlim Sinfonia da Metrópole” (1927) de Walter Ruttmann, “São Paulo Sinfonia da Metrópole” (1929) de Adalberto Kemeny e Rudolf Lustig, “Visões Paulistas” (1972) de Elie Politi e “O homem com uma câmera” (1929) de Dziga Vertov, que são filmes-ensaios: objetos não identificados em termos de gênero cinematográfico. Com imagens documentais, mas sem compromisso documental, com imagens ficcionais, mas sem compromisso narrativo ficcional. Filme-ensaio: não é um documentário, nem uma ficção, nem um docudrama. Os ensaios foram os seguintes: (1] Fluxos – dedica-se ao estudo do movimento na grande metrópole, (2] Arquitetura – concentra-se no “design” da maior metrópole brasileira, (3] Pessoas – centra foco sobre o “elemento humano” na cidade e (4] Imagens – os aspectos imagéticos da grande metrópole, pontes, vistas aéreas dos prédios. O principal objetivo de EstereoEnsaio é visitar formas, gêneros e tradições do Cinema Silencioso com “olhar” contemporâneo.

⁴ Baseado no clássico “Berlim - sinfonia da cidade”, os húngaros Adalberto Kemeny e Rodolfo Lustig, donos do laboratório de cinema Rex Filmes, produziram este documentário sobre a cidade de São Paulo, em 1929. Link para o filme: <https://www.youtube.com/watch?v=JZUPyq10q9I>.

⁵ A estreia do filme está prevista para o segundo semestre de 2017, ao som de uma orquestra ao vivo, sob a regência do maestro Livio Tragtenberg, que também fez a composição sonora da obra, mantendo a referência de formas, gêneros e tradições do Cinema Silencioso com um olhar contemporâneo.

Na captação de “EstereoEnsaio São Paulo” foram utilizadas lentes grande-angulares para mergulhar o espectador dentro da metrópole, assim como propor uma imersão nas diversas camadas da imagem em 3-D procurando não estender o efeito a ponto de incomodar ou causar estranheza, mas utilizando-o de forma branda e muitas vezes sutil, como é o caso nas cenas de trânsito vistas de cima. Este tipo de utilização tem a proposta de produzir um efeito se dá muito mais dentro do que fora da tela, de forma a valorizar a imersão e propor sensações ao espectador que vive cotidianamente em vizinhança com os 12 milhões de habitantes da cidade.



Figura 2.1 - Cena do filme “EstereoEnsaio São Paulo”



Figura 2.2 - Cena do filme “EstereoEnsaio São Paulo”

Para a captação foram utilizadas duas câmeras, disparadas ao mesmo tempo e sincronizadas pelo dia e hora, capturando a mesma imagem. Nessa captação com o *rig* chamado Stereotec *light weight*, as câmeras são separadas por um espelho que reflete a mesma imagem de forma que na finalização se dá a estereoscopia, como acontece com nossos olhos pois cada câmera representa tecnologicamente um olho humano. Posteriormente, durante a filmagem, toda a estrutura com câmeras de cinema e *rig* era ainda muito pesada, mesmo optando pelos modelos mais leves possíveis. Este aspecto era evidente quando o operador de câmera utilizando um *steadycam* se deslocava em um *segway*, como é o caso das cenas na Bienal de São Paulo, no Ibirapuera, no trem, na Ciclofaixa e na rua 25 de Março. Estes planos tiveram uma duração de tempo limitada pelo peso do equipamento.



Figura 2.3 - Imagem de um dos equipamentos utilizado nas filmagens de “EstereoEnsaio São Paulo”

A partir do teste aéreo verificou-se que seria preciso uma estrutura de câmeras muito robusta com uma “cabeça estabilizada” por causa da grande trepidação que as câmeras iriam sofrer tanto do helicóptero como do vento. Novamente seria preciso pensar no fator peso de toda a estrutura e foi então utilizado um *rig* desenvolvido pela empresa especializada em filmagens aéreas Aircam para filmagens em estéreo.



Figura 2.4 - Imagem do helicóptero com rig para captação das imagens aéreas de “EstereoEnsaio São Paulo”

A velocidade de gravação da câmera foi 48fps, ou seja duas vezes mais lenta do que a velocidade normal. Trata-se de um tipo de recurso para auxiliar a observação e mergulho do espectador nas cenas apresentadas, a fim de que ele possa apreciar com tempo a imagem estereoscópica que exige mais atenção que a imagem bidimensional. Este também é um recurso comumente utilizado para se dar um ritmo poético às imagens, mesmo para imagens bidimensionais. As câmeras utilizadas foram duas Red Epic Dragon, no formato 5k (5120x2700 pixels) no sentido de obter uma margem para re-enquadramento e ajuste do efeito para compor a estereoscopia. Mas, o filme foi finalizado em 4k (4096x2160 pixels) por causa da resolução dos projetores de cinemas e festivais.

Estes dois filmes estereoscópicos haviam sido precedidos pelo projeto de filme 2014k, apresentado em Joanesburgo, durante a Copa do Mundo de 2010. O projeto 2014k foi uma iniciativa dos pesquisadores Jane de Almeida, Eunezio de Souza (Thoroh), Cicero Silva, entre outros que, em parceria com o CPqD (Centro de Pesquisa e Desenvolvimento em Telecomunicações) e com apoio financeiro da FINEP (Financiadora de Estudos e Projetos), propuseram a captação do primeiro jogo de futebol em 4K 3D para posterior exibição da captação durante a realização da Copa do Mundo de Futebol da FIFA de 2010, na Casa Brasil. O projeto também tinha como objetivo criar uma infraestrutura de redes avançadas de 10 Gbps para permitir que salas de cinema e redes de TV pudessem ter acesso e capacidade de transmissão dos jogos na Copa do Mundo do Brasil de 2014 em 4K 3D. O objetivo era deixar um legado de redes e sistemas cabeados para distribuição

de dados, audiovisual, ciência e, principalmente, um parque tecnológico para a pesquisa e desenvolvimento⁶.

2.4. Projeto Marte: imagens e o imaginário com material de arquivo

O “Projeto Marte” foi iniciado com um apoio do fundo MackPesquisa em 2013, com a proposta de produzir um filme em 4k para ser transmitido por redes fotônicas. Após a realização do filme em 2016, não foi realizada a fase de transmissão na Universidade Mackenzie, por dificuldades com a rede de 1GB/s dedicada ao laboratório, já que ela fica bloqueada por firewall da universidade. No entanto, o filme gerou uma série de curiosidades de linguagem sobre o seu tema, além do desafio de trabalhar imagens de arquivo como fotos da internet e filmes do youtube de qualidade baixa em imagens de alta definição, esteticamente aceitáveis ao espectador de filmes 4k – um espectador digital.

Trata-se de um filme de tese e científico de super-alta definição que aborda as crenças norte-americanas de poder, as utopias soviéticas e a influência do planeta Marte nas expressões culturais brasileiras. As cenas são compostas com material de arquivo, histórias curiosas sobre o gigante vermelho e com entrevistas realizadas *in loco* com engenheiros e pesquisadores do Mars Institute, da NASA e da Moon Express. A tese do filme é que “imagem” tem relações muito mais próximas com o “imaginário” do que com a realidade. Pergunta-se se as imagens de Marte, hoje produzidas pela Nasa, são provenientes de um novo imaginário – o da pós-produção – e menos da realidade, como geralmente se supõe. Ou seja, o imaginário do fim do século XIX, de habitantes belicosos foi deslocada por imagens desérticas do atual temo do desastre ecológico do nosso planeta. Dessa forma, Marte continua sendo espelho dos nossos desejos e medos.



Figura 2.5 Cena do filme “Loga”

⁶ Sobre o projeto 2014k ver: <<http://www.witz.com.br/2014k/media.html>> . Acesso em: 14 jul. 2017.



Figura 2.6 Cena do filme “Loga”

“Projeto Marte”, ou “Loga” (como o filme agora se chama) tem sido visualmente composto por imagens de arquivo, estáticas (fotografias) e em movimento (vídeos) provenientes de diversas fontes, resoluções e formatos. Algumas imagens inéditas foram produzidas em 4K, especialmente para o filme, além de *Motion Graphics* (letreiros) que reafirmam ou complementam informações relevantes. O áudio é composto por uma narração em voz *off* que guia o vídeo, além de uma trilha sonora original e músicas sobre Marte que compõem o imaginário brasileiro.

A partir de pesquisa inicial e da estruturação do roteiro seguiu-se para um pesquisa online de imagens estáticas e vídeos que contemplassem os temas, que abordam desde as primeiras observações telescópica, até a exploração americana. Muitos estudos e experiências ainda estão sendo realizados para a construção do filme na ilha de edição com base em imagens estáticas e vídeos resgatados da internet. Os maiores desafios deste projeto estão relacionados aos diferentes formatos e qualidade das imagens usadas e as formas de construir um produto coeso com materiais de arquivo digital.

Com a inspiração nos fragmentos de informação sobre Marte recebidos pela NASA optou-se por não trabalhar as imagens estáticas com movimentos geralmente usados em produtos cinematográficos como: zoom in-out, movimentos verticais ou horizontais. No processo de edição, foi usado o padrão de cortes secos e muitas vezes são apresentados recortes, pedaços quase abstratos, antes de revelar a integralidade da imagem. Alguns artifícios estão sendo estudados para melhorar a qualidade das imagens em baixa resolução, são eles: a interpolação e o *upscaling*, a utilização de filtros e ajuste de cor e até a tentativa de vetorização das imagens para se obter um material homogêneo.

No caso das imagens em movimento, vídeos, estão sendo usados dois tipos de material. O material capturado em 4K com a JVC 4K pela equipe de pesquisadores, especialmente para o filme, apresentando entrevistas com Michael Sims (diretor do Instituto Marte na NASA Ames Research Center) e John Welch (diretor do Connect Lab na NASA AMES Research Center). Foram também capturadas imagens de Throna Pinnacle, na Califórnia – uma locação muito utilizada por Hollywood para representar

Marte e espaços extraterrestres. Foi também filmada uma aula fictícia de Marciano. Este material está sendo editado em full HD, mas será substituído pela sua resolução original (4K), na finalização do filme. Dentro das imagens em movimento também o filme é composto de vídeos de arquivo disponíveis online, dependendo da resolução original do material, os trechos selecionados são trabalhados de duas formas, ou na tela inteira como parte do filme ou repetido várias vezes formando um mosaico em movimento. Estas possibilidades ainda estão sendo estudadas para melhor trabalhar os vídeos de uma forma que ele não perca sua textura e temporalidade. Após o primeiro corte, que foi feito para o relatório final do projeto, o filme continua no processo de pós-produção independente.

2.5. Filmes Frames

Filmes Frames são pequenos filmes em alta resolução que têm a característica de serem filmados em *slow motion*, para verificar a capacidade de visualização de detalhes de cada *frame*. Eles foram realizados em 2015 e 2016 com diferentes câmeras. O filme “O Mergulho” com um minuto foi produzido em diversos formatos com o objetivo de se perceber a relação entre resolução e tempo dos frames. O mergulhador⁷ foi filmado quatro vezes em resoluções diferentes, com a preocupação de que os enquadramentos fossem muito semelhantes. Dessa forma pode-se perceber o movimento “devagar” do *slow*, em máxima velocidade que a câmera permite, em resolução menos alterada e em resolução baixa, além do número de pixels por linhas x colunas. Foram capturados 4 mergulhos, em full Hd com 241f/s, 2K com 160 f/s, 4K com 121f/s e 6K com 83f/s. A câmera ao mudar de resolução se utiliza de uma parte do sensor correspondente ao tamanho da imagem e normalmente gera um *crop/up scaling* desta imagem. O filme “Open Mouthed” (Boquiabertos) foi feito no formato 6k (6144x3160) a 83 frames por segundo (fps)⁸. Esta é a velocidade máxima que a câmera suporta para um formato 6k. Uma das propostas do estudo também era trabalhar com uma imagem em alta definição a muitos quadros por segundo para entender quais seriam as dificuldades de seu processamento durante a finalização, questão não apenas do ambiente acadêmico, mas também internacional, do mercado sobre o processamento desta imagem.

O projeto também deveria servir de conteúdo para *streaming* ao vivo para testes da rede acadêmica, uma vez que a resolução da imagem estava alterada pela sua velocidade. A câmera usada em ambos os filmes foi a Red Epic Dragon, com iso 400 para melhor proteger as altas luzes e a curva de cor REDGamma4. Também foi utilizado shutter a 90° e 45°, para permitir maior definição a cada quadro, já que a cena possuía rápidos movimentos, tanto de cartas quanto o splash da água.

A respeito da montagem, o filme “Open Mouthed” ou *Boquiabertos*, seguiu uma narrativa tradicional construída pela relação de plano-contraplano das crianças com o mágico, personagens do filme que homenageavam o cineasta soviético Dziga Vertov. Os planos de diferentes truques de mágica foram intercalados com as reações das crianças que se deslocam de uma "desconfiança" até um "deslumbramento". Esta montagem tem o intuito de causar a empatia no espectador para que ele também se encante com as imagens como a criança. Já a montagem do filme “O Mergulho” foi feita para melhor

⁷ Performance de John Flavin. Trilha sonora de André Olzon. Ne endereço: https://www.youtube.com/watch?v=Z_SFP5PK76c.

⁸ Performances dos mágicos Henry e Melinski e das crianças Joaquim e Cecilia Silva. Trilha sonora de André Olzon. No endereço: <https://www.youtube.com/watch?v=I65E5cKvhc0>.

visualizar as diferentes resoluções e os efeitos de *Slow Motion* gerados pela quantidade de frames capturados na filmagem. Mesmo que a versão final do filme esteja em Full Hd 23,97 f/s para melhor se adequar aos players padrões, pode-se perceber e comparar a qualidade das imagens capturadas nos diferentes formatos e a velocidade que cada um atinge com diferentes quantidades de frames.

Depois de editado o material foi substituído pelo original e exportado em 'super-alta' resolução em uma sequência de imagens estáticas como é característico do formato TIFF (a 23,97 fps), cada imagem correspondeu a um frame do vídeo final. A escolha em exportar os filmes em uma sequência de imagens em alta resolução e não em um formato natura de vídeo (.mov .mp4 etc) ocorreu como uma solução compatível com o computador de edição usado na época, no caso utilizamos um servidor da empresa de players de vídeos Zaxel Systems⁹. Atualmente com o grande avanço da tecnologia, principalmente referente a manipulação de imagens em 4K esse procedimento já pode ser substituído pela edição, finalização e transmissão do material em seu formato original. Esses filmes serviram de conteúdo de streaming em diversos festivais, tais como o TIFF (Tokyo International Film Festival) e o CineGrid Amsterdam e foram exibidos em playback no Cinegrid em San Diego, em 2015.

2.6. Cirurgia Refrativa em 1000 fps

Em 2015, a equipe do LabCine, em parceria com o Núcleo Técnico-científico Telessaúde Brasil Redes e com o Departamento de Oftalmologia da Escola Paulista de Medicina da Unifesp realizou a captação de uma cirurgia laser com uma câmera capaz de captar 1000 frames por segundo em 4K, o que permite que se visualizem detalhes imperceptíveis a olho nu. A equipe de captação realizou a filmagem durante um dia de trabalho no Departamento de Oftalmologia da Unifesp. A captação da cirurgia com o tipo de câmera utilizado é bastante delicada, pois, além de toda a preparação do ambiente, em relação a segurança do paciente, entre outros, o próprio equipamento demanda uma capacidade computacional associada bastante significativa.



Figura 2.7 Câmera com recurso de captação de 1000 fps em 4K posicionada no centro cirúrgico do Depto. de Oftalmologia da Unifesp. (Foto: Cicero da Silva)

⁹ Ver em: <<http://www.zaxel.com/>>.

Para se ter uma ideia do volume de dados, a câmera, quando captando a 1000 fps em 4K, gera 3.5 Terabytes de dados a cada 3 minutos. Isso significa que a cada 4 minutos é necessário descarregar os dados dos cartões, o que, com o uso de computadores sofisticados, com taxas de transferências altas e capacidade de memória bastante avançada, a transferência durou cerca de 8 minutos. Além disso, a quantidade de *hard drives* externos para armazenamento dos dados também foi significativa. A captação foi bem-sucedida e foi possível visualizar a ação do laser de maneira precisa.

A equipe que realizou a captação da cirurgia com a câmera pretende realizar a transmissão de uma cirurgia com fluxo de vídeo de 1000 frames por segundo em 4K, porém, estima-se que no momento atual seria necessária uma banda de 106 Gbps de conectividade diretamente conectada à câmera. Os resultados dessa captação foram apresentados no CineGrid Workshop 2015, realizado anualmente no CALIT2 da Universidade da Califórnia, San Diego (UCSD). A proposta deste trabalho parte do desafio de captar imagens que o olho humano não vê, explorando a área da medicina oftalmológica, especificamente com a utilização da máquina a laser de correção oftalmológica chamada Lasic. Para se obter tal efeito, seria necessário filmar em velocidade com muitos quadros por segundo. As imagens quando reproduzidas mostram de forma lenta aquilo que nosso olho não capta (o que chamamos de *slow*). Neste experimento foi utilizada a câmera Phantom Flex 4K, uma câmera ultra-rápida capaz de filmar em alta definição. Ela tem sido usada com frequência em comerciais, assim como em séries televisiva (como a série “Timewarp”) e até em longas-metragens.

Foram realizados diversos testes anteriores à filmagem com câmeras mais simples, para estudar enquadramento, luz, exposição e velocidade. Mesmo que essas câmeras não superem a velocidade de 60 fps, os testes eram válidos pois não haveria facilmente uma segunda chance, caso alguma coisa desse errado.

Os testes cirúrgicos mostraram que a velocidade precisava ser muito maior do que 60 fps para avaliar o processo do laser e chegou-se à conclusão de que seria preciso 1000 fps em alta-definição, no formato 4K. Além disso, como todo o frame deveria ser preenchido com o objeto de interesse (no caso o olho), optou-se por utilizar a lente Arri Macro 100mm. A iluminação da sala era uniforme, e ao iniciar o procedimento cirúrgico as luzes eram apagadas restando apenas um feixe de luz que incidia no rosto do paciente. Esse era um problema, pois se fosse muito forte, não seria possível registrar a incidência do laser nem mesmo a 1000 fps. Assim, a lente deveria ser mais luminosa possível. No caso as lentes Arri macro que partem do diafragma T=3, diafragma pouco luminoso e sensível para este tipo de trabalho, sendo que as lentes mais usuais são as Arriflex Super Speed que partem de T=1.3. Foi usado um obturador menor que o padrão (180°) e decidiu-se pelo de 90°, para que o desenho do laser fosse mais definido possível. A sensibilidade padrão da câmera é ISO 800, mas por causa da lente e do obturador, foi necessário utilizar ISO 1250, mais sensível que a câmera permite, sem perder qualidade das altas e baixas luzes.

O resultado das imagens da cirurgia a laser foi muito satisfatório, pois possibilitou visualizar o laser, o caminho que ele percorre no olho do paciente e quão veloz é esta aplicação. Considera-se que este material seja útil para os estudantes de oftalmologia que poderão utilizar estas imagens para consulta ou mesmo futuros estudos.

2.8.1. Arqueologia do procedimento super slow

A cirurgia refrativa e os filmes frames fazem parte da mesma lógica. Os usos dessa câmera e da câmera anterior possibilitaram uma série de estudos filmicos de movimento e imagens não visíveis a olho nu com balões, fogo e água nos moldes dos filmes de pesquisa dos primeiros cinemas. Observar um mergulho em câmera lenta ou visualizar detalhes minuciosos do rompimento de um balão de água são investigações iniciadas ainda no século XIX e que tinham como base a conquista da nitidez e precisão do instantâneo (Rosenblum 2007 p.15). Acredita-se que a considerada primeira fotografia fixada com sucesso em placa de metal (Rosenblum 2007 p. 19), *View from the Window at Le Gras* realizada por Joseph Nicéphore Niépce no ano de 1827, exigiu mais de 8 horas de exposição para ser concluída. O daguerreótipo, processo fotográfico anunciado publicamente em 1839 por Louis Jacques Mandé Daguerre, discípulo de Niépce, já apresentava uma redução do tempo de exposição que podia variar de “cinco a sessenta minutos, dependendo da cor do referente e da potência da luz” (Rosenblum 2007, p.17). Tanto Rosenblum (2007, p. 248) como Roiullé (2009, p. 39) apontam que o interesse pela exploração da visibilidade do mundo para além da visão humana impulsionou, nos anos subsequentes, o desenvolvimento de novos aparatos, o aperfeiçoamento da técnica e a conquista de cortes temporais cada vez mais rápidos.

A ideia de ver mais e além do que o olho vê é a premissa por trás dos estudos iniciais sobre o movimento. Endelman (1988, p. 02) considera o experimento de William Henry Fox Talbot, em 1851, como sendo a primeira tentativa bem-sucedida de se congelar uma ação em uma imagem fotográfica. Numa sala escura e com a ajuda da iluminação de faíscas de garrafas de Leiden, Talbot expôs um pequeno pedaço de jornal em movimento usando uma velocidade de obturador de 1/2000 de segundos, produzindo uma imagem legível do referente em questão (p. 02). Segundo Rosenblum (2007 p. 249), apesar do sucesso de Talbot, experimentos com luz artificial e estouros de *flashes* voltaram a ser observados com maior atenção apenas em 1880, pelo cientista austríaco Ernst March.

De maior importância para o desenvolvimento da cinematografia de alta velocidade foi o estudo do movimento pela produção de sequências fotográficas. A conquista do instantâneo – ou a quase anulação do tempo na imagem – despertaria, paradoxalmente, um interesse entre cientistas dos mais diversos campos pela produção das imagens em série e levaria, conseqüentemente, a um resgate do tempo na imagem. Para além da curiosidade, os cientistas e médicos imediatamente perceberam a importância dessas aplicações. No campo da astronomia, destaca-se o francês Pierre Jules Cesar Janssen e seu “revólver fotográfico” (Almeida et. al 2016 p. 3), instrumento “operado por um dos integrantes de sua equipe, o astrônomo brasileiro Francisco Antonio de Almeida” (McKernan apud Almeida et. al 2016 p. 04) durante uma expedição para o Japão em dezembro de 1874. Endelman (1988 p. 04) acrescenta que o aparato foi capaz de registrar 47 daguerreótipos da passagem do planeta Vênus sobre o sol em intervalos de 70 segundos.

Segundo a pesquisa histórica de Rosenblum (1997, p. 249), a primeira grande produção de fotografias sequenciais foi realizada por Eadweard J. Muybridge, fotógrafo inglês radicado na Califórnia. Entre os anos de 1872 e 1885, Muybridge já havia produzido mais de 100 mil imagens de ações como “caminhar, correr, jogar bola, dar piruetas, movimentos de reverência, empilhar tijolos, entre outras atividades” (p. 250). Endelman (1988, p. 03) explica que a técnica do fotógrafo consistia na utilização de 24

câmeras equipadas com filmes de placas de vidro expostas a 1/1000 de segundos. Posicionadas paralelamente à linha do movimento, as câmeras eram disparadas por um circuito programado para “liberar o obturador a tempos predeterminados de 1/100 a 1/6000 de segundo” (Endelman, 1988, p. 03).

Apesar de Muybridge já utilizar um sistema de medição de tempo na captação de fotos foi o médico fisiologista francês Etienne-Jules Marey quem cunhou o termo cronofotografia. Diferente de Muybridge, Marey trabalhava com uma única câmera (Figura 1), tendo transformado o revólver fotográfico de Janssen em um aparato muito mais rápido e mais compacto, capaz de fixar 12 fotos em tempo de exposição de 1/720 de segundo, “velocidade permitida pelo uso de gelatinas secas, significativamente mais sensíveis do que as antigas placas fotográficas” (Sanz 2014 p. 449). A escolha do fisiologista por realizar múltiplas exposições em uma única placa de filme tinha como objetivo diminuir os intervalos espaciais entre um movimento e outro e, como aponta Sanz (2014 p. 457), revelar com maior eficácia a temporalidade fisiológica da locomoção de um corpo, conquista que para Marey não era possível pelo sistema de câmeras separadas usado por Muybridge.

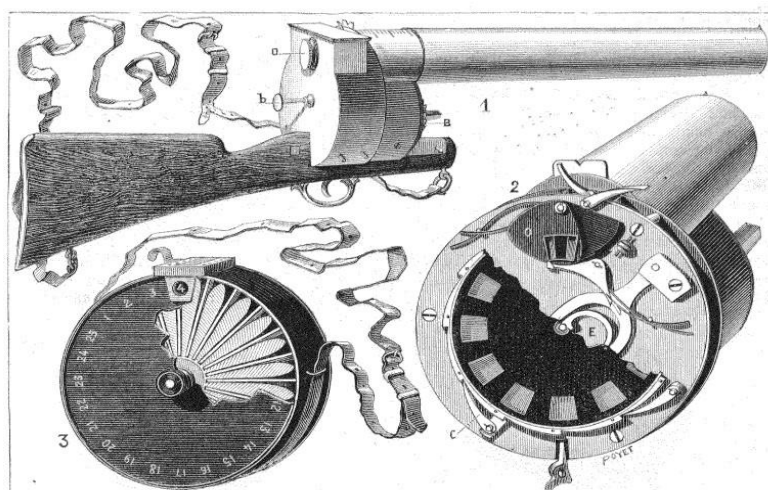


Fig. 2. Mécanisme du fusil photographique.
1 Vue d'ensemble de l'appareil. — 2. Vue de l'obturateur et du disque à fenêtre. — 3. Boîte contenant vingt-cinq plaques sensibles.

Figura 2.8: Detalhes do fuzil fotográfico de Marey, gravura publicada em um artigo do fisiologista para a revista La Nature n°464, em 22 abril de 1882. Disponível em <https://commons.wikimedia.org/wiki/File:Fusil_photographique_Marey2.png> Acesso em: 27 jul. 2017.

As imagens de Marey mostram um corpo em deslocamento (Figura 2), ocupando diferentes posições em um único espaço¹⁰. Segundo Sanz (2014 p. 458), essa busca do fisiologista por compreender a relação existente em cada momento entre a distância e o tempo transcorridos, o levou a diminuir “progressivamente os intervalos entre um disparo e outro, entre um e outro instante fotográfico” (Sanz, 2014 p. 458). Michaelis (1955 p. 123) afirma que foi Marey, no ano de 1890, o primeiro a apresentar a cinematografia de alta velocidade ao capturar o voo de uma pomba a 60 f.p.s. Endelman (1988 p. 04) acrescenta que seu fuzil fotográfico foi mais tarde aperfeiçoado, chegando a capturar até 100 f.p.s.

¹⁰ Disponível em <https://www.youtube.com/watch?v=11KItGNuiY>.

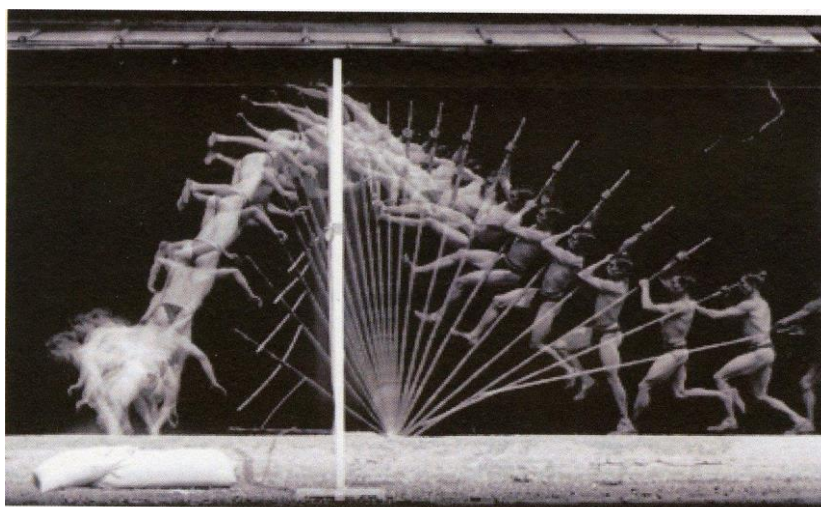


Figura 2.9: Exemplo de um dos muitos experimentos realizados por Marey durante seus estudos sobre a fisiologia humana. Disponível em <<https://theredlist.com/wiki-2-16-601-798-view-pioneers-profile-marey-etienne-jules.html>> Acesso em: 27 ago. 2017.

As conquistas de Marey no campo das imagens em alta velocidade foram expandidas por seu assistente Lucien Bull. No filme *Incunables du cinéma scientifique* (1984), apresentado na videoteca virtual do *Le Centre national de la recherche scientifique*, é possível observar estudos feitos por Bull entre os anos de 1904 e 1909, dentre eles a observação do movimento de enguias, o voo e vibração da asa de insetos e o estudo de balística. Um de seus trabalhos de 1904¹¹ apresenta uma bala perfurando uma bolha de sabão. A animação capturada a 1500 f.p.s. demonstra que a bolha reconstitui-se quando a bala entra e rompe-se apenas com a saída do projétil (Figura 3). Michaelis (1955 p. 125) discorre ainda sobre o sucesso de Bull em desenvolver uma câmera estereoscópica ativada por descargas elétricas. O aparato composto por “dois tambores giratórios, cada um munido de 54 negativos de 35mm” (p. 125) era capaz de produzir imagens a uma frequência de 1/2000 f.p.s. e foi usado com sucesso para analisar os movimentos das asas de libélulas e consequentemente desvendar seu mecanismo de voo¹².

As tecnologias digitais têm possibilitado a criação de visualizações cada vez mais complexas e avançadas. O exemplo disso são as imagens de ultra definição criadas por laboratórios de supercomputação em centros de pesquisa de física, tais como o CERN na Suíça e o EVL (Electronic Visualization Laboratory) em Chicago. Esta capacidade de geração de imagens permite visualizar fenômenos tanto em macro quanto micro escalas, como as imagens geradas em nanoescala de proteínas ou bactérias. Daí a associação de imagens científicas e as super resoluções da imagem em movimento. Tais imagens são campo fértil para discussões sobre a forma como fenômenos siderais podem ser compreendidos pelo sujeito contemporâneo, já que, em sua maioria, são imagens geradas em laboratórios de pós-produção com ferramentas digitais a partir de interpretações acerca dos fenômenos. Os cientistas, por sua vez, têm buscado o apoio de artistas do campo das artes cinemáticas para auxiliá-los na produção de imagens, criando narrativas criativas para as imagens geradas nos laboratórios. Acredita-se que este seja um excelente

¹¹ Disponível em <https://www.youtube.com/watch?v=j0do7IQH3ms>

¹² Disponível em <https://www.youtube.com/watch?v=rfmGzjIpUa4>.

campo de produção a ser pesquisado, que envolvem as tecnologias 4k e acima, temáticas científicas e a transmissão por redes fotônicas.

2.7. Pixel Race¹³

A equipe do LabCine iniciou, no ano de 2014, o projeto *Pixel Race*, que teve como resultado a realização de um documentário em 4K que documentou os trabalhos de gravação, transmissão e exibição de jogos da Copa do Mundo no Brasil com resolução 8K. O projeto foi organizado pela rede de televisão japonesa NHK. A resolução de 8K, ou Ultra Full HD (FUHD), é uma resolução de ultra-alta definição (UHD) da ordem de 7680 x 4320 pixels, ou 33.177.600 pixels (33 mega pixels), com 22.2 canais de áudio. A emissora pública japonesa NHK liderou a pesquisa e investimento em tecnologia de resolução 8K a partir de 1995. Desenvolvida pela Sony, a tecnologia 8K utilizada pela NHK foi batizada de Super Hi-Vision®. Paralelamente, a NHK também desenvolveu tecnologia de som surround 22.2, ou seja, com 22 canais de áudio, estimando que esse seja o padrão comercial adotado até 2032. Em maio de 2009 um projetor com resolução full 8K é apresentado, durante o *open house* do NHK Science & Technology Research Laboratories, em paralelo a um experimento de transmissão ao vivo de vídeos no padrão Super Hi-Vision® em múltiplos canais via satélite. Em maio de 2010, uma câmera de resolução full 8K é apresentada no *open house* do NHK Science & Technology Research Laboratories. Um projetor 8K compacto é apresentado em janeiro de 2011.

Em fevereiro de 2012, um sensor de imagem com *frame rate* de 120Hz é desenvolvido, e em maio do mesmo ano uma câmera 8K compacta, do mesmo tamanho de uma câmera HDTV, é finalmente desenvolvida. Em julho de 2012, começa a aplicação de maior fôlego do Super Hi-Vision® em coberturas esportivas, com o sistema sendo empregado em visualizações públicas de eventos nos Jogos Olímpicos de Londres. Um mês depois, o sistema é recomendado como padrão internacional para a televisão pelo ITU-R. Em fevereiro de 2014, dando sequência à sua estratégia de capitalizar sobre a visibilidade de grandes eventos esportivos, o Super Hi-Vision® é aplicado à cobertura de eventos nos Jogos Olímpicos de Sochi (Olimpíadas de Inverno). Em julho de 2014, o Super Hi-Vision® é utilizado na cobertura de alguns jogos da Copa do Mundo no Brasil. Em setembro do mesmo ano, um plano produzido com uma câmera 8K Super Hi-Vision®, com frequência de *frames* de 120Hz, é exibido pela primeira vez ao mundo na IBC2014 em Amsterdã, na Holanda.

O acompanhamento e documentação do trabalho dos técnicos da NHK, bem como a cobertura dos processos de captação, distribuição e recepção pública das imagens 8K dos jogos acentuou e ampliou um campo de indagações que tem norteado as atividades de nossa equipe de pesquisa. Este registro foi realizado com ajuda de custo de hospedagem e câmera 4K providenciados pela RNP. Ou seja, apenas por interesse da própria equipe em registrar o evento, entendendo sua importância nos campos da tecnologia e da linguagem do cinema.

8k é uma configuração que parece estar sendo gradativamente redefinida pelo mercado, com a chegada de monitores de TV domésticos que prometem resolução 4K

¹³ Pixel Race foi parcialmente publicado no livro CineGrid: futuros cinemáticos (CINUSP). Filme disponível em <https://goo.gl/62f5td>.

(nativa ou aproximada, os atuais televisores UHD), e até mesmo superior (8K), enquanto se anunciam projeções experimentais, públicas ou coletivas de imagens em movimento com resolução 8K ou superior (16K), mais dirigidas ao parque exibidor cinematográfico. Os pesquisadores da NHK afirmam que no ano de 2020, durante as Olimpíadas de Tóquio, os jogos serão transmitidos pela TV aberta com qualidade 8K.

No documentário *Pixel Race*, o especialista em processamento e compressão de imagens em vídeo, Kazuhisa Iguchi, foi entrevistado como o pesquisador responsável pela transmissão das imagens 8K gravadas no Rio de Janeiro para o Japão. O sinal de transmissão comprimido partia do IBC (Instituto Brasileiro de Ciências), sendo encaminhado para o Japão através da rede acadêmica da RNP em tempo real e resolução 8K, no padrão Super Hi-Vision®. Pode-se concluir, portanto, que a resolução 8K representa o limite para telas caseiras, mas não para a situação cinema. No âmbito da cobertura de eventos esportivos, especificamente de jogos de futebol, a tecnologia 8K acena com padrões diferenciados de decupagem e edição de conteúdos. Dada a altíssima resolução, as possibilidades de aproximação e distanciamento (*zoom in* e *zoom out*) podem vir a ganhar em amplitude e dinamismo. A alta profundidade de campo oferece também uma relação visualmente singular entre objetos no primeiro e no segundo planos (forma/fundo).

2.8. Sensemaking, um editor de vídeos *On the Fly*

A transmissão em tempo real de vídeos tem permitido que pesquisas e ações docentes possam ser acessadas por diversas pessoas em localidades geográficas distintas. Contudo, as transmissões criam uma quantidade significativa de conteúdos para serem armazenados e um fluxo de imagens em tempo real, praticamente não “assistíveis” pelo espectador, cujo tempo e interesse exige uma edição do material. Além do mais, a transmissão em tempo real de imagem em movimento com alta resolução resulta no excesso de imagens e de dados que tornam a armazenagem e a recuperação dos dados (conteúdos) armazenados complexa e cara. A transmissão de procedimentos cirúrgicos em tempo real, por exemplo, gera quantidades de dados muito superiores à capacidade de armazenamento padrão de conteúdos extensos sendo transmitidos, o que torna difícil a busca por determinados conteúdos, a seleção de algum trecho, podendo comprometer a visualização de maneira dinâmica pelos interessados. Estes procedimentos cirúrgicos, tais como cirurgias cardíacas de campo aberto, podem durar em torno de 6 horas, sendo que, para atividades de ensino e pesquisa, o procedimento que de fato interessa pode ter uma duração bem menor, gerando vídeos editados de uma hora ou mesmo de 10 minutos, para divulgação e registro da pesquisa. No campo do ensino, sabe-se que várias atividades gravadas no *streaming*, que foram importantes na pesquisa compartilhada, não serão posteriormente necessárias à utilização dos docentes.

Levando em consideração estes aspectos, a proposta deste projeto é a criação de uma plataforma de edição em tempo real (*on the fly*) de *streaming* de vídeos, proporcionando a possibilidade de edição digital em tempo hábil para ser disponibilizada e não esquecida, como acontece com o material transmitido geralmente. A ideia é desenvolver um editor de vídeo amigável e manejável que possa ser utilizado durante o andamento do streaming em repositórios com o objetivo de oferecer o material para acesso público a ser usado no ensino e pesquisa de forma geral.

Acredita-se que este dispositivo seja um recurso prático para diminuir o excesso de imagens de uma transmissão *online* de aulas, palestras, cirurgias, procedimentos científicos e observações com longa duração. Pretende-se, portanto, produzir uma ferramenta com a qual a equipe responsável por uma atividade de ensino ou de pesquisa, que utiliza e necessita de ferramentas de streaming de vídeo em tempo real possa, ao mesmo tempo em que transmite o conteúdo, editar e selecionar as partes do conteúdo. A ferramenta de edição de vídeos *on the fly* permitirá que o usuário crie uma narrativa que faça sentido ao conteúdo (por isso, SenseMaking) que está sendo captado e visualizado em tempo real, por meio da possibilidade de se inserir metadados, imagens de abertura, créditos, imagens estáticas, trilha sonora, computação gráfica e legendas nos vídeos.

A princípio, este projeto teve como perspectiva gerar um editor que pudesse ser disponibilizado no repositório `video@RNP`, pois a plataforma de Vídeos da Rede Nacional de Ensino e Pesquisa – RNP tem sido utilizada por instituições acadêmicas no país como base de armazenamento de dados, como ferramenta de recurso de incorporação de vídeos em bases dedicadas ao ensino e pesquisa, como nos casos dos repositórios de vídeos utilizados nas salas de aula virtuais pelas universidades, entre outras aplicações.

2.8.1. Histórico e Estado da Arte dos editores de vídeo

O processo de montagem ou edição de um produto audiovisual consiste em selecionar, ordenar, e ajustar os planos ou momentos a fim de alcançar o resultado esperado. Antes da era digital, pode-se comparar o processo de edição com a escrita de um texto em uma máquina de escrever. Antes de datilografado, o texto deveria ser bem esquematizado, rascunhado e revisado; qualquer alteração posterior ao texto no papel era complicada, demorada e com altos custos. Assim eram também feitas as edições de produtos audiovisuais, em que se pensava o produto final como um todo antes de ir para o processo de pós-produção. No início do cinema, esses eram processos físicos como corte e colagem das películas. Com a evolução da tecnologia, as gravações passaram a utilizar fitas magnéticas (BETAMAX, VHS), para selecionar e regravar, o que gerava perda da qualidade do material. Com esses processos não havia nenhuma abertura para o erro ou para o acaso, tudo tinha que ser pensado e planejado com antecedência. A esse tipo de edição chamamos “edição linear”.

Com a chegada da era computacional, surgem os “softwares de edição” funcionando com os hardwares (Placas de Vídeos, HDS), com versões inclusive para computadores pessoais. O primeiro sistema adotado pelo mercado foi a mescla de hardware e software criado pela empresa NewTek, chamado de Video Toaster para o computador Amiga 2000. O mecanismo apresentava uma interface inspirada nos processos analógicos, como o ícone de tesoura representando a ferramenta de corte das antigas moviolas do cinema de película. Na época, isso representou um avanço em questão de mobilidade das imagens dentro do vídeo, como maior facilidade em apagar ou substituir imagens. Consiste ainda de uma série de efeitos e ferramentas, transições e aprimoramento de imagens de baixo custo, que antes eram complicados de realizar. Grandes conceitos apresentados por este mecanismo ainda resistem nos softwares de edição modernos, como ferramentas de corte digital e o conceito de Timeline. Porém, não era mais necessário de pensar em um produto final integral e muitos experimentos podem ser realizados, inserindo-se e retirando-se imagens dentro da linha de edição. Esse seria o início do conceito de “edição não-linear”.



Figura 2.10 Video Toaster rodando no Computador Amiga 2000. Fonte: <<https://goo.gl/iSB6YS>>

A edição de vídeo não-linear é, portanto, comparável ao texto produzido com um editor de texto digital, podendo ser escrito e reescrito, podendo ser testado em várias ordens diferentes e aleatórias, experimentando novas possibilidades e, conseqüentemente, dando a abertura para o acaso e o erro. Em um único computador, podem-se editar diferentes formatos, mudar e editar o áudio, além de acrescentar “artes” sem alterar o bruto original.

Com o passar dos anos, houve um aprimoramento destas tecnologias e dos softwares. Atualmente, há um grande número de softwares de edição digitais, linear e não-linear que oferece ferramentas capazes de se adaptarem às necessidades do usuário. Foram analisados os seguintes softwares de edição: Windows Movie Maker (Microsoft), iMovie (Apple), Final Cut (Apple), Premiere Pro (Adobe), Avid Media Composer (Avid) e Smoke (Autodesk) e conclui-se que, mesmo apresentando características específicas próprias partem de um princípio básico: um espaço para a visualização do material anterior à montagem, uma Timeline com espaço para áudio e vídeo, um espaço para artes e efeitos e uma tela para visualizar do produto final que será exportado em forma de arquivo ou enviado para algum site de armazenamento de vídeo.

Tradicionalmente, a realização de um produto audiovisual passa por três etapas. A primeira é destinada a concepção e planejamento do filme, chamada de pré-produção, que se resume a criação e organização do roteiro e dos diversos elementos que envolvem a sua produção eficaz. Muitas vezes em uma ficção ela já é entendida como um produto final, as cenas do roteiro já devem ser estruturadas na sequência do filme editado, e muitas vezes são usados *storyboards* (desenhos ou animações) para pré-visualização das sequências em conjunto. Já no universo do documentário esse processo pode ser mais flexível, pois é a etapa de estruturação das ideias, pesquisa de personagens, pré-entrevistas e organização das filmagens que interessa ao gênero. A segunda etapa do processo, tanto de uma obra de ficção quanto a de documentário está destinada à produção. O momento de captura das imagens para o produto final, geralmente, quando bem pré-produzida é a etapa mais rápida de todo o processo.

O processo de edição do filme/vídeo é considerado parte da pós-produção. Com as imagens já produzidas, é o momento em que o diretor e o montador trabalham em

conjunto para construir o filme como um produto coeso. Em uma obra de ficção eles se baseiam no roteiro para cortar, colar e experimentar as linguagens e efeitos, criando ritmo e sentido para os trechos de imagens desconexas. Nos documentários, o processo de montagem está junto com o roteiro, pois só depois do material capturado o diretor saberá como quer trabalhar com as imagens para contar a história. Após a montagem/edição o filme ainda poderá passar pela finalização, que se constitui nos processos de composição (interferências gráficas ou efeito especiais digitais); Mixagem de som (equalização de volume e efeitos sonoros); Coloração que deixará o filme com a cor e textura desejada pelo diretor.

No caso do Sensemaking, fica previsto que a etapa de pré-produção se mantém igual, mas a etapa de produção (captura do material) em paralelo com a pós-produção (organização, montagem e finalização). Enquanto o evento está sendo captado e transmitido ele poderá ser editado e finalizado, assim, otimizando o tempo e a produtividade do produtor audiovisual.

O Sensemaking possui três componentes principais a serem disponibilizados: *streaming*, interface do usuário e o banco de dados. O componente streaming e player de vídeo recebe o *streaming* vindo da câmera de vídeo e o transfere para a interface. O banco de dados armazena os dados referentes ao vídeo já editado (url do vídeo, autor, palestrante, tempos de duração, início e término, título, categoria, palavras-chave, data, local). Para o desenvolvimento da interface do Sensemaking, foi feita uma análise das interfaces de editores de vídeo presentes no mercado, como Windows Movie Maker, iMovie, Apple Final Cut, Adobe Premiere Pro, Avid Media Composer, Autodesk Smoke e Youtube Video Editor. O objetivo foi buscar suas principais funcionalidades e identificar a forma como os diferentes perfis de usuário interage com cada uma delas.

A partir destas observações foi desenvolvida a primeira versão da interface do Sensemaking. Com este primeiro protótipo da interface, foram realizados testes de usabilidade com potenciais usuários do Sensemaking para nos certificarmos da clareza e eficiência da interface. A tarefa que os usuários realizaram foi criar, em tempo real, fragmentos do vídeo em *streaming* e transferí-los para a linha do tempo, onde seriam editados. A edição não foi testada, pois esta funcionalidade não foi integrada à interface. A partir de um questionário para verificar a eficiência e o entendimento da interface, investigou-se a opinião dos potenciais usuários. Participaram do teste inicial seis pessoas que na sua maioria tinham pouca ou nenhuma ou experiência com edição de vídeos. Essa seleção dos participantes levou em consideração que o objetivo do Sensemaking, que é ser uma plataforma de edição voltada para os usuários que criam conteúdo de vídeo casuais ou voltados para pesquisa, portanto, não são profissionais em edição de vídeo. A partir do *feedback* recebido percebeu-se os diversos problemas que a interface apresentava: dificuldade em entender para que a interface seria utilizada, dificuldade na compreensão dos ícones que informam as funcionalidades da interface. Foi então construída uma nova versão de interface para os usuários dividida em quatro áreas: a área dos vídeos, a timeline, as abas de edição e a área de informação sobre os vídeos. Na área dos vídeos, a janela maior, à esquerda, recebe o fluxo de vídeo do webcam em tempo real. A janela menor mostra o vídeo que está sendo editado. Os focos destas janelas podem ser alterados conforme o desenvolvimento da edição.

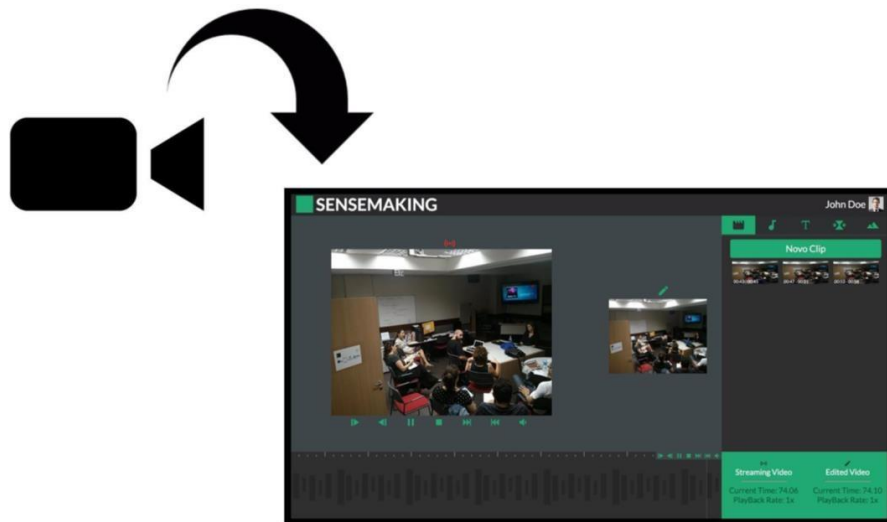


Figura 2.11 - Interface dos usuários, que recebe o fluxo de vídeo a partir da webcam.

Enquanto o sinal de vídeo original, proveniente da câmera de vídeo está sendo recebido, o usuário poderá separar fragmentos deste vídeo para realizar a edição desejada. Podem ser selecionados quantos trechos de vídeo forem necessários, ou seja, novos clips de vídeo a partir do fluxo que está sendo recebido.

A barra lateral esquerda está reservada para recursos que o Sensemaking disponibilizará no futuro, como a opção de incluir trilha sonora, cartelas e titles, efeitos de transição e filtros para aprimorar as imagens captadas. Na *Timeline*, cada clip selecionado poderá ser editado, conforme mostra a Figura 2. A edição irá contemplar cortes do clip, inserção de textos, legendas, cartelas, áudio, efeitos especiais, imagens. No canto inferior direito, há uma área com informações a respeito dos vídeos, tanto aquele que está sendo recebido (*Streaming Video*), quanto o editado (*Edited Video*).

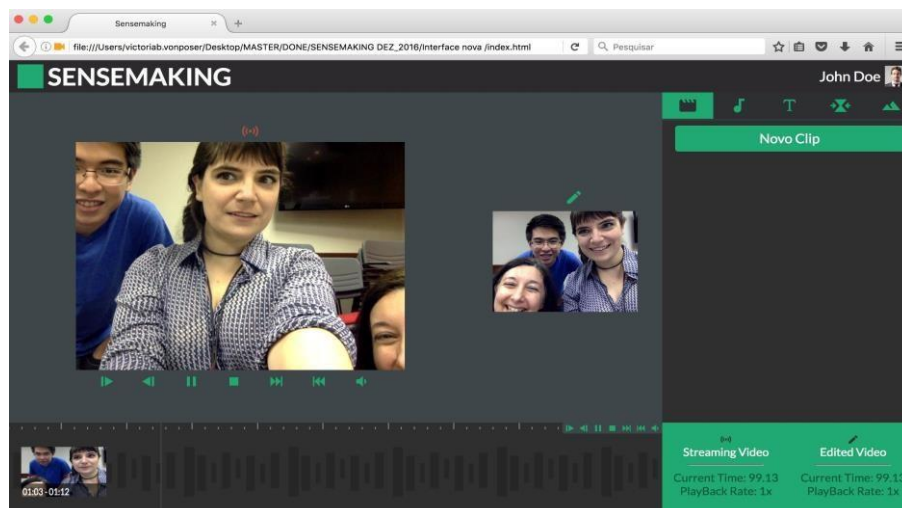


Figura 12 - Interface do usuário, com o clip capturado na Timeline para a realização da edição.

2.9. Parte Dois: Cinemas Futuros ou Visão do Futuro

Como “visão do futuro”, imagina-se os “possíveis futuros” a partir do que se desenvolveu até o momento. Destacam-se os projetos de bancos de dados de filmes, a fotogrametria como proposta de arquivamento digital e o streaming de orquestras. Esta parte do artigo contempla os projetos que ainda estão sendo desenvolvidos para serem submetidos a agências de fomento. Aposta-se em um “futuro” com o desenvolvimento de bancos de dados filmicos, bancos de dados de arquivos para memória, registro e ensino. Acredita-se também que a estereoscopia e a fotogrametria – imagens “tresderizadas”, sejam também campos de pesquisa a serem desenvolvidos para colaboração no contexto digital.

2.9.1. O Cinema de Bancos de Dados

O sistema digital tem proporcionado a criação de formas de visualização que se tornam a cada instante mais desenvolvidas e complexas. Um dos resultados mais significativos dos últimos anos é um Cinema do Banco de Dados que surge da lógica e da estética computacional. Bancos de dados são formados por um conjunto de arquivos que estão relacionados entre si, como coleções de dados organizados e que tenham alguma relação entre elas com o objetivo de trazer sentido a um montante de informações. Um Cinema de Banco de Dados é aquele distribuído via tecnologia *streaming* (Netflix e Youtube, por exemplo) e que também é produzido com uma estética de Banco de Dados (uso de imagens de arquivos, mistura de gêneros cinematográficos, e ritmos narrativos e de montagem cada vez mais acelerados e interconectados – influenciados pelas novas narrativas advindas das redes sociais – são alguns dos exemplos deste padrão estético).

Para Lev Manovich, a sociedade computadorizada atual, se assemelha a “uma coleção interminável e desestruturada de imagens, textos e outros registros de dados”(Manovich in Vesna 2007 p. 40. *Ttadução nossa*). Se levarmos em conta a lógica da qual Manovich fala, pode-se considerar a Netflix como um grande banco de dados que, assim como os demais bancos de dados de *streaming*, inova a forma de distribuição do conteúdo audiovisual. Assim, “O banco de dados se torna o centro do processo criativo na era da computação”(Manovich in Vesna 2007 p. 45).

O termo “estética de banco de dados” tornou-se uma palavra-chave do reino digital e coloca questões semânticas interessantes que parecem delinear o campo da pesquisa e da arte em relação às próprias bases de dados. “O que exatamente queremos dizer com ‘estética de banco de dados’? No discurso sobre a arte digital, o termo é frequentemente usado para descrever os princípios estéticos aplicados na imposição da lógica do banco de dados a qualquer tipo de informação, filtragem de coleta de dados e visualização de dados”(Paul in Vesna 2007 p. 95). Nesse sentido, a estética de banco de dados muitas vezes se torna um potencial conceitual e forma cultural – um modo de revelar padrões (visuais) de conhecimento, crenças e comportamento social. Porém, apesar de aparentemente bastante contemporânea, de acordo com Manovich (2000), a *estética do banco de dados* no cinema foi inaugurada em 1929, pelo filme “O Homem com uma câmera” de Dziga Vertov.

Uma das grandes inovações do cinema do Banco de Dados de uma plataforma como a Netflix é o fim de uma relação programática entre o espectador e o conteúdo, que não precisa esperar o horário programado de conteúdos para usufruir de filmes e vídeos que o interessam. Mérito não da plataforma em si, mas da Internet, um meio de telecomunicação, de acordo com Manovich (2000), que revolucionou as formas de

distribuição de conteúdo, e não somente isso, mudou também a lógica da sociedade contemporânea. O fim da relação programática entre o espectador e o conteúdo também se deu devido à tecnologia *streaming*, desenvolvida há mais de vinte anos, como afirmam Baochun Li, Zhi Wang, Jiangchuan Liu, Wenwu Zhu, em seu artigo “Two Decades of Internet Video Streaming: A Retrospective View” (2013).

Para concluir, de acordo com Manovich, “(...) A transformação rápida da cultura nos anos 1990 em e-cultura, dos computadores em portadores de cultura universal, dos meios de comunicação em novas mídias, exige que repensemos nossas categorias e modelos” (Manovich 2011 p. 06, tradução nossa). Assim, o LabCine pretende contribuir com a reflexão sobre o Cinema de Banco de Dados, suas novas formas de distribuição e produção, através da criação da plataforma de serviço *streaming* do Laboratório de Artes Cinemáticas e Visualização.

Dois projetos estão sendo desenvolvidos no momento: o primeiro para facilitar o acesso aos filmes 4k realizados pelo LabCine, e o segundo para organizar e promover o acesso a filmes científicos do Brasil.

2.9.2. Repositórios de vídeo online

Os vídeos são um meio de comunicação que envolvem o ser humano ao integrar o ver, o visualizar, o ouvir, colocando diante do espectador cenários, situações, cores, relações espaciais, diálogos, efeitos sonoros. Estas mensagens múltiplas, proporcionadas pelos meios audiovisuais tendem a ser um processo de comunicação viável na medida que facilitam a compreensão de determinado assunto. Diante destas facilidades, repositórios de vídeo online têm se intensificado na web. Reúnem dados de maneira organizada relacionados a produções científicas ou áreas temáticas e trazem uma série de benefícios tanto para os pesquisadores quanto às instituições ou sociedades científicas, proporcionando maior visibilidade aos resultados de pesquisas e possibilitando a preservação da memória científica. Pode-se citar como exemplos o VideoLectures.NET e o Video@RNP.

VideoLectures.NET reúne um acervo de vídeos educacionais a partir de palestras ministradas por estudiosos ilustres e pesquisadores em eventos relevantes no campo da Ciência. O portal visa promover a ciência, a troca de ideias e a partilha de conhecimento, fornecendo conteúdos didáticos de alta qualidade, não só para a comunidade científica, mas também para o público em geral. Todas as palestras, documentos anexos, informações e links são sistematicamente selecionados e classificados por meio do processo editorial levando em conta também as observações dos usuários. O Video@RNP disponibiliza um conteúdo relacionado às atividades-fim das organizações usuárias da RNP (Rede Nacional de Ensino e Pesquisa): ensino, pesquisa, saúde e cultura. O portal permite ao usuário visualizar, buscar ou publicar vídeos; agendar transmissões de streaming de eventos: buscar e recuperar as vídeo-aulas publicadas no serviço Videoaula@RNP. Também permite às emissoras de TV transmitirem seus canais pela rede.

Seguindo nesta linha, este projeto pretende estudar, implementar e validar um repositório online e colaborativo para armazenamento, transmissão e compartilhamento de Vídeos de Ultra Alta Definição (UHDV). Os UHDVs priorizam a informação visual devido à alta qualidade das imagens geradas, sendo propício para manter a fidelidade da representação. Em algumas áreas de ensino e pesquisa, como medicina, astronomia,

geografia, entre outras, preservar estas características é relevante e prioritário. O repositório se constituirá em um portal online, aberto à comunidade científica e público geral, promovendo trocas de conhecimento científico em várias áreas do conhecimento. Pretende-se que o repositório atenda às instituições acadêmicas no país para consulta e incorporação de vídeos. Os vídeos serão sistematicamente selecionados e categorizados de forma a facilitar a busca e a visualização. A implementação do repositório de UHDVs oferece oportunidade de pesquisa em três frentes:

- explorar as limitações da largura de banda e da infraestrutura de rede na transmissão de UHDVs e garantir sua qualidade da entrega na arquitetura Cliente- Servidor;
- estudar a elaboração de uma base de dados dinâmica e colaborativa específica para UHDVs que envolve a categorização;
- investigar formas de interação do usuário em portais de vídeos para a elaboração de uma interface adequada ao repositório de UHDVs, considerando as heurísticas de usabilidade.

Como os arquivos de vídeos UHD geram grande quantidade de dados e demandam uma enorme capacidade de armazenamento, cabe ressaltar a importância de um estudo dos algoritmos de compressão, como o H.264/AVC (Advanced Video Coding) e o padrão High Efficiency Video Coding (HEVC) para reduzir essa quantidade sem comprometer a qualidade da imagem. Como o repositório será colaborativo, ou seja, aceitará o envio de novos vídeos por usuários específicos, a criação de uma base de dados multimídia, dinâmica e específica para UHDV será necessária. Cada vídeo estará relacionado a uma categoria autoria, assunto, título, procedência e outras informações para facilitar a indexação e a busca do vídeo ou trecho de vídeo desejado pelo usuário. A base de dados deverá considerar aspectos relevantes para garantir a eficiência, eficácia e a consistência dos dados. Pretende-se utilizar uma arquitetura Cliente-Servidor híbrida, onde o cliente (front-end) continua tendo a característica de executar as tarefas do aplicativo através da interface do usuário. Já a camada servidor será uma solução híbrida, tendo uma camada inicial responsável pela resposta ao usuário, seguida de uma camada de back-end, onde serão integradas diversas soluções consolidadas para a garantir a responsividade necessária para este tipo de aplicação (streaming de vídeos de UHDV).

Um dos desafios do projeto é construir uma interface para o usuário, a partir de uma análise de usabilidade de interfaces de repositórios de vídeo existentes, como VideoLectures.NET, Video@RNP, Youtube, Vimeo, Netflix, Hulu Plus, Plataforma Vídeo Brasil, Cinegrid. Além da busca e seleção de vídeos, a interface deverá permitir a incorporação de novos vídeos para usuários que tenham esta permissão, exibir a categorização de assuntos disponíveis nos vídeos e facilitar o compartilhamento das informações.

Como contribuições para a sociedade, pretende-se disponibilizar um portal composto por um repositório *online* e colaborativo de vídeos de ultra alta definição para consulta e compartilhamento com conteúdos relacionados ao ensino, pesquisa, saúde e cultura. O acesso será público, mas a postagem de vídeos será limitada a instituições de ensino e pesquisa. O repositório será utilizado como base de armazenamento para incorporação de vídeos a ser utilizado em salas de aula presenciais ou virtuais pelas universidades, entre outras aplicações, não se limitando a este uso. O banco de dados

produzido durante esta pesquisa contribuirá para a organização de uma produção de filmes UHD, categorização dos assuntos e criação de metadados multimídia para facilitar a busca não só pelo filme em si, mas por trechos de vídeo.

2.9.3. Labcine Interviews com SenseMaking

A interface do LabCine tem como uma de suas subdivisões o LabCine Interviews, um banco de dados de vídeo-entrevistas focadas no tema a que o grupo de pesquisa se propõe, o embate entre o cinema, a imagem estática e em movimento e suas relações com a computação. Resultado do trabalho colaborativo de pesquisadores associados ao Labcine, o arquivo, em agosto de 2017, já conta com três entrevistas em fase de edição. Com o objetivo de padronizar e criar uma coerência estética para o produto final, foi escolhida como locação para as entrevistas a própria sede do LabCine. Na captação das imagens digitais em full HD (1920x1080) foram utilizadas duas câmeras, uma JVC Cinema 4K e uma Canon 5D Mark III. A primeira câmera foi posicionada diretamente na frente do entrevistado e permaneceu fixa durante todo o processo de captação da entrevista. A segunda câmera, responsável por enquadramentos mais fechados, foi posicionada à esquerda do entrevistado, a um ângulo de 45°, e manipulada por uma cinegrafista. A entrevistadora foi posicionada entre as duas câmeras para dirigir o diálogo e guiar o olhar dos entrevistados, mas permaneceu fora do enquadramento do vídeo.

A escolha dos entrevistados e, conseqüentemente, o conteúdo das entrevistas, reflete o cenário híbrido e interdisciplinar que permeia o estudo das imagens cinematográficas. Através do cinema estudam-se tanto as práticas de visualização como novos modos de exibição, arquivamento e seleção de imagens, narrativas científicas e educacionais. Soma-se a tais discussões a relação da imagem na construção de identidades e representações, seu papel na história da cultura e da cultura visual. A primeira entrevista foi concedida por Maxine Brown, diretora do Laboratório de Visualização Eletrônica (EVL, na sigla em inglês) da Universidade de Illinois em Chicago (UIC). Brown é a idealizadora e desenvolvedora do sistema de colaboração remota SAGE (Scalable Amplified Group Environment) e uma das mais importantes pesquisadoras nas áreas de visualização avançada e *big data*. Em sua entrevista a pesquisadora discorreu sobre realidade virtual (VR) e cinema, as diferenças e trocas entre ciência e ficção científica e como sistemas de visualização avançada podem moldar as salas de aula e as pesquisas colaborativas no futuro. Brett Stalbaum, o segundo entrevistado, é artista e coordenador do curso de Design Especulativo e do Curso Interdisciplinar em Computação e Artes da UCSD. Stalbaum tem desenvolvido pesquisas há mais de 10 anos com o LabCine. O principal tópico abordado pelo artista-pesquisador diz respeito ao “Earth computing” e inclui as diferenças do conceito em relação ao “Cloud computing” bem como a sua relação com as artes visuais.

A terceira entrevista foi concedida por Stephanie Dennison, professora de Estudos Brasileiros na Universidade de Leeds, Inglaterra. Dennison atualmente coordena a pesquisa “Soft Power, cinema and the BRICS”, financiado pela AHRC - Arts, Humanities and Research Council. A pesquisadora comentou sobre o tema principal de sua pesquisa, a cultura do cinema brasileiro e seu lugar num contexto mais amplo do cinema mundial.

Para pensar as dinâmicas que cercam a relação entre cinema e o mundo contemporâneo no campo da visualização, todos os entrevistados, após responderem questões sobre suas áreas específicas de atuação, foram convidados a refletir e opinar sobre a “transparência” do computador, como máquina de ação do cinema digital. O

diálogo entre esses campos integrados busca assim alimentar e instigar o desenvolvimento de pesquisas sobre transmissão de dados e aplicação de novas tecnologias além de ajudar a traçar um eixo que determine o que é o cinema, as transformações pelas quais vem passando com o advento do digital e novas configurações passíveis de serem adotadas pelas artes cinematográficas num futuro próximo.

Após serem editados, os vídeos das entrevistas serão armazenados no canal do Labcine no Youtube, com a resolução máxima de 1080p. A fim de uma melhor apresentação estética e operacional deste conteúdo, os vídeos do canal do labcine serão centralizados no site readymag.com/labcine/interviews. Esta plataforma foi escolhida, pois propicia o design e a estruturação de informações visuais, sonoras e textuais de forma rápida e intuitiva, bem como, a integração de plugins externos, como o Youtube e o Vimeo para vídeos, Soundcloud para áudios, e Flickr para imagens. Nesse sentido, a plataforma funcionará como uma *gallery wall* virtual que, além das entrevistas em vídeo, agrupará imagens no formato gif, citações, links e outras informações relevantes com o propósito de ilustrar e contextualizar os conteúdos discutidos nas conversas com o entrevistado. Os bancos de dados, além de armazenarem uma história, propiciam a disponibilização de conteúdos para ensino de forma não programática – não é preciso esperar o horário do broadcasting. Para efetivar a colaboração a ferramenta SenseMaking deverá ficar disponível na plataforma de entrevistas, propiciando ao receptor o corte de cada entrevista para apresentar para seu público. Assim, o fluxo de cada câmera poderá ficar disponível para ser cortado pelo espectador.

2.9.4. A fotogrametria no cinema, nos games e nos museus

Dentre as diversas formas contemporâneas de captação de imagem que podem ser representadas tanto pelo cinema de alta definição quanto pelas imagens cinematográficas de alta velocidade, a fotogrametria é uma ferramenta que se destaca, pois possibilita através do uso de no mínimo duas fotografias de determinado espaço ou objeto – tiradas de diferentes ângulos e com uma margem de sobreposição entre elas – a representação tridimensional deste mesmo espaço-objeto.

Embora, no passado, os processos fotogramétricos dependessem, segundo Brito e Coelho (2007), de enormes aparelhagens mecânicas, câmeras métricas e demandassem, além disso, um corpo de trabalho específico e treinado, na contemporaneidade, este cenário tem mudado. Como afirmam os autores, isso ocorre devido a automatização digital da fotogrametria, que é capaz, através do uso de computadores, de processar o enorme montante de dados resultantes dos procedimentos fotogramétricos, exigindo apenas o controle e, ocasionalmente, a intervenção humana.

Softwares fotogramétricos como Agisoft Photoscan, Reality Capture e Autodesk Remake se baseiam, geralmente, em um mesmo fluxo de trabalho semiautomático. Remondino (2011) classifica essa sequência de tarefas em: aquisição das imagens, calibração/orientação, geração da nuvem de pontos 3D, estruturação/modelagem e, por fim, a obtenção do modelo tridimensional. Portanto, primeiramente, os softwares precisam ser abastecidos com uma série de fotografias de determinado objeto, as quais necessitam cobrir o máximo de lados possíveis deste item; caso o objeto de interesse esteja sob um fundo neutro, o software conseguirá distinguir mais facilmente o que é a figura principal e o que não é; em uma próxima etapa, o programa alinha as imagens tentando reconstituir virtualmente o ângulo, no qual cada fotografia foi tomada em relação ao objeto fotografado; em seguida, é criada uma nuvem de pontos, representando a

superfície externa do objeto; logo, essa nuvem é transformada em uma malha tridimensional rígida; e, por fim, gera-se uma textura fotorrealística que recobre essa malha 3D. O objeto finalizado pode ser exportado para softwares de renderização e animação – como Blender, Cinema4d ou Maya –, para programas de modelagem – como 3D Coat, Mudbox, Zbrush – e para plataformas de arquivamento, visualização e compartilhamento de modelos 3D – como p3d.in ou Sketchfab.

Willis (2016) relata que, tendo sido usada, no século passado, principalmente para fins de mapeamento geográfico, a fotogrametria, hoje, tornou-se um processo indispensável na criação de gráficos tridimensionais, na composição de sequências de efeitos visuais em filme, além de, também, ser útil no campo da cinematografia virtual e em outros processos de modelagem 3D.

Um dos usos mais marcantes desta técnica, no campo do cinema, foi a sua aplicação na criação dos cenários de fundo mostrados durante o efeito cinemático *bullet-time* de Matrix (1999). Como visto, em um vídeo¹⁴ que revela o processo de produção deste longa-metragem, o *bullet-time* é criado em um ambiente totalmente revestido por um *chroma key* verde, no qual um ator é fotografado, ao mesmo tempo, em um movimento de ação, por diversas câmeras posicionadas a sua volta. Na pós-produção, essas imagens fotográficas são sequenciadas e mescladas a fundos criados por fotogrametria, ou seja, por fotos provenientes do mundo real processadas no computador e transformadas em ambientes tridimensionais fotorrealísticos.

Porém, esse truque que utiliza a técnica fotogramétrica não foi descoberto ao acaso, na verdade, como Willis (2016) esclarece, o efeito usado por Matrix (1999) foi influenciado pela pesquisa de doutorado de Paul Debevec sobre fotogrametria, realizada na Universidade da Califórnia, em Berkeley, nos Estados Unidos, anos antes. O curta-metragem *The Campanile Movie* (1997)¹⁵ – apontado por Willis como sendo parte das pesquisas de Debevec – mostra o pesquisador andando pelo campus universitário com uma torre em miniatura nas mãos. Em determinado ponto do filme, o pequeno modelo se mescla a um modelo tridimensional virtual da torre real do campus que é, de forma vertiginosa, sobrevoada por uma câmera que explora suas várias facetas. Willis revela que para a realização do filme, o pesquisador e seus parceiros fotografaram a torre do campus da universidade de diferentes ângulos, inclusive por meio de tomadas aéreas realizadas com a ajuda de pipas. Em seguida, dados geométricos foram retirados destas imagens e elas foram processadas por um software chamado *Facade*, a fim de se obter um modelo tridimensional completo da área ao redor da torre da universidade. Debevec também criou um percurso de uma câmera virtual que, no curta, se mistura, em determinados cortes, com o movimento da câmera real.

Foster e Halbstein (2014) chamam atenção para o fato de que, no passado, os artistas se aproximaram da matemática e da perspectiva linear para dar as suas obras bidimensionais uma aparência tridimensional; hoje, nas áreas da computação gráfica e no design digital, uma aproximação entre a arte e a ciência ocorre mais uma vez, quando os artistas se apropriam da tecnologia fotogramétrica para transformar objetos tridimensionais em *artworks*. Para os autores, a absorção do uso da fotogrametria, no

¹⁴ The Making of "The Matrix" - What Is Bullet Time? - Disponível em: <<https://www.youtube.com/watch?v=uPNBdDNZbYk>>. Acesso em: 15 ago. 2017.

¹⁵ The Campanile Movie. Disponível em: <https://www.youtube.com/watch?v=RPhGEiM_6IM>. Acesso em: 15 ago. 2017.

campo da computação gráfica 3D, objetiva a criação rápida de determinados modelos tridimensionais, com o intuito de se obter imagens baseadas no mundo real. Nesse sentido, pontua-se que, no efervescente campo dos jogos eletrônicos, a ambição dos estúdios de videogames por replicar o mundo real no virtual recorre, cada vez mais, aos procedimentos fotogramétricos digitais.

Na Game Developers Conference de 2016¹⁶, ocorrida na Califórnia, nos Estados Unidos, o diretor técnico de arte Kenneth Brown e o *environment artist* Andrew Hamilton demonstraram o uso da fotogrametria digital na produção do jogo Star Wars: Battlefront (2015). A tecnologia foi usada em etapas significativas de sua produção a fim de se poupar tempo na criação de modelos 3D, além do mais, a técnica também foi escolhida por proporcionar a captura complexa e realista de elementos naturais – como árvores, rochas e penhascos – e itens artificiais – como roupas, acessórios e armamentos – necessários à composição das ambientações do jogo de Star Wars. Também foi noticiado que a equipe teve acesso às peças do Lucasfilm Cultural Arts Museum¹⁷ – local que abriga vestimentas, armaduras e objetos provenientes das trilogias fílmicas de Star Wars; nesse espaço, eles puderam fotogrametrar parte desses itens, exceto pela icônica armadura do vilão Darth Vader, que não conseguiu ser bem capturada pelo software fotogramétrico, devido ao seu brilho e a sua cor escura – fatores que, comumente, confundem o funcionamento desse tipo de programa.

Além do campo do entretenimento, a fotogrametria também começa a adentrar ao campo da museologia virtual, sendo representada, principalmente, pelos esforços na digitalização e no compartilhamento online de importantes acervos museológicos, como os reunidos na plataforma “Sketchfab for Cultural Heritage”¹⁸. Instituições como o The British Museum, com 220 obras digitalizadas em 3D, e o Museu d'Arqueologia de Catalunya, com 96, são exemplos disso. No Brasil, apenas o Museu de Arqueologia e Etnologia da USP é listado nesta plataforma, com 7 obras disponíveis em 3D.

Nesse aspecto, Manovich (2010) comenta sobre a resistência demonstrada por determinadas instituições culturais, como os museus, em relação ao modo de apresentação e a representação de seus dados culturais, que ainda seguem uma mentalidade unidimensional e desconsidera fatores como fluidez, multiplicidade, hibridismo e mutação no trato de suas informações. O pesquisador aponta que, diferentemente de áreas como o design e a arquitetura, que sofreram grande influência da mentalidade fluída da computação gráfica, os campos culturais desperdiçam a potencialidade desse tipo de pensamento.

Portanto, são inúmeras as possibilidades ofertadas pelos objetos gerados pela fotogrametria: ao contrário do que é, comumente, proibido em museus, esses objetos podem ser tocados e manipulados virtualmente; podem ser ampliados para que pequenos detalhes sejam vistos; podem ter seus volumes, cores, texturas alterados; podem ganhar vida, ao ser animados em outros softwares; podem guardar informações textuais, tags e

¹⁶ Star Wars: Battlefront and the Art of Photogrammetry. Disponível em: <https://www.youtube.com/watch?v=U_WaqCBp9zo>. Acesso em: 15 ago. 2017.

¹⁷ How We Used Photogrammetry to Capture Every Last Detail for Star Wars Battlefront. Disponível em: <<http://starwars.ea.com/starwars/battlefront/news/how-we-used-photogrammetry>>. Acesso em: 15 ago. 2017.

¹⁸ Sketchfab for Cultural Heritage. Disponível em: <<https://sketchfab.com/museums>>. Acesso em: 15 ago. 2017.

áudio; podem ser compartilhados em redes sociais; podem ser replicados em impressoras 3D etc. É válido lembrar que na indústria cinematográfica, no campo dos games e na museologia, a fotogrametria digital ainda é um processo em contínua expansão, que não pretende anular outras formas de captação do real, mas, ao contrário, junto a elas, objetiva a ampliação, cada vez maior, dos alcances das visualidades contemporâneas.

Um experimento recente que reuniu fotogrametria e museu foi a projeção de quatro objetos do Museu Memória do Bixiga – MUMBI – no evento cultural 13 na 13, ocorrido no bairro da Bela Vista em São Paulo, em maio de 2017, com o apoio do grupo de pesquisa Arquivo, Memória e Cidade da Universidade Presbiteriana Mackenzie. Para isso, foram fotogrametrados quatro objetos pertencentes ao espaço museológico: uma caneca comemorativa da escola de samba Vai-Vai, um rádio *vintage* da Emerson Radio and Television, uma antiga mala de viagem e um gravador *vintage* Mitsubishi. Os objetos foram captados com uma câmera fotográfica FujiFilm XT10 com uma lente de 35mm, sendo que, foram tiradas, ao redor de cada objeto, cerca de 100 fotos. Nesse processo, o Labcine foi o espaço escolhido para se realizar a digitalização dos objetos, pois possui equipamentos informáticos robustos para lidar com processamento gráfico de alta complexidade, portanto, foram utilizados tanto o *software* Agisoft Photoscan na digitalização dos objetos quanto o *software* Adobe Photoshop na animação 3D destes itens. Em um futuro breve, este projeto fotogramétrico pretende se desdobrar para novas captações fotogramétricas, com o intuito de criar, abastecer e compartilhar um acervo de objetos tridimensionais provenientes de espaços museais, como o MUMBI.

2.9.5. Streaming de Orquestras



Fig. 13 - Streaming de vídeo captado no Roy Thomson Hall, em Toronto, Canadá. Foto: Cicero da Silva.

A transmissão ao vivo em audiovisual de concertos sinfônicos é um empreendimento de história notável, e remonta aos primeiros *broadcastings* para televisão. Um dos primeiros relatos a respeito é escrito pelo diretor britânico Arnold Roger Manvell (1909 - 1987),

primeiro diretor da British Film Academy. Antes de completar um ano de “experimentos”, Manvell viu a necessidade de registrar os primeiros resultados e algumas perspectivas do que poderia vir a ser o “Terceiro Programa” da grade diária produzido pela “British Broadcasting Corporation” (BBC) para a televisão pública na Grã-Bretanha. Neste pequeno artigo, de julho de 1947, acaba por documentar que ao lado de “conferências, longa-metragens, variedades de primeira linha, peças de teatro”, eram os “concertos de alta qualidade (realizados frequentemente pela própria orquestra sinfônica da BBC)” que formavam a programação diária disponibilizada ao público.

Ali, entre questões prementes de acesso e difusão no que argumentam ser “uma oportunidade de (...) quebrar as barreiras do preconceito, e destruir a pernicioso distinção que cresceu entre o *highbrow* e o *lowbrow*”, imagina-se que o objetivo da programação é permitir que “o nível do gosto (médio) gradualmente aumente”. E Manvell chama atenção a questão da qualidade da imagem, que se coloca como uma das necessidades técnicas primordiais.

“*Experiments in broadcasting and television*” pode ser entendido como um estudo seminal para a área. Embora não objetive termos específicos do “problema sinfônico” em *broadcasting*, o fato é que já aponta ali diversos entre os elementos que seguem objetivamente em desenvolvimento da performance ao vivo para as atuais transmissões para internet de concertos de música clássica: primeiro por chamar a atenção para os aspectos não necessariamente sistematizados do projeto, com todo um ambiente criativo que faz dele uma (nova?) modalidade artística a ser eventualmente esmiuçada em sua gramática e técnica; a oportunidade, que segue potencialmente aberta, e virtualmente desejável, para uma comunicação de massa com alto poder de participação, prazer e aprendizado; a legitimação por elementos típicos da performance presencial, como a intimidade - uma questão de apresentação e temporalidade vivas.

Manvell faz questão de citar longamente Dennis Johnston (1901 - 1984), autor e diretor de teatro envolvido desde os primeiros momentos na produção de programas para televisão. A citação é oportuna, pois Johnston demonstra uma sensibilidade não pequena na descrição desta adaptação necessária entre mídias - e as diferenças entre televisão e cinema, televisão e teatro. O centro da diferença, segundo ele, é o *live broadcasting* e toda a gramática cujo viés é dado indiscutivelmente pelo aparato tecnológico. À diferença das tantas discussões teóricas a respeito, Johnston restringe-se – a nosso ver de modo inequivocamente estimulante para o desenvolvimento objetivo das discussões sobre o assunto – a questões pragmáticas. O comando do espetáculo deixa de ser realizado pelo diretor de cena ou música, e passa para o que chama de “Producer” – aquele responsável pela palavra final do que vai a tela do telespectador. Tudo que deve ter a atenção da audiência é organizado pela direção de câmera, e o “Producer” tem, em resumo, todos os recursos do diretor de cinema a sua disposição, inclusive a possibilidade, à diferença do teatro, de “aproximar” a visão dando maior ou menor ênfase a objetos e personagens. Por outro lado, não há qualquer modo de trabalhar sob a estimulante e laboriosa flexibilidade do filme: o *live broadcasting* é um empreendimento onde a possibilidade de novas tomadas ou testes com material para opções de diferentes efeitos não é possível. “Como no teatro, estamos frente a prospectiva de uma tomada irremediável, a partir da qual precisamos seguir ou cair” (Manvell 1947 p. 391).

Não curiosamente é acompanhando as publicações sobre os primeiros *live broadcasting* de orquestras para televisão que se encontram de forma bem organizada e original os termos que fundam as reflexões dos eventos contemporâneos de *real time streaming* para internet. Em 1953, Burton Paulu, diretor desde 1938 do KUOM¹⁹ vice-presidente do "National Association of Educational Broadcasters" e conselheiro da presidência do "Educational Television and Radio Center", publica o artigo "Televising the Minneapolis Symphony Orchestra". Tratando longamente da defesa do financiamento do projeto – argumentando entre outros elementos do potencial educativo e de difusão da mídia – Paulu (1953) trata de forma direta das questões técnicas e artísticas originais que acabam por estarem envolvidas no empreendimento.

As perguntas que originam o artigo ("devem orquestras sinfônicas serem televisionadas? Ou o *broadcast* de música instrumental é tarefa para a rádio?") não são respondidas diretamente.

Paulu (1953) descreve assim os procedimentos da apresentação de nove concertos televisionados entre janeiro e abril de 1953. Intitulado "A grande Orquestra Sinfônica e a região a qual ela serve", o projeto previa não apenas experimentar o *broadcasting* de música clássica tais como poderiam acontecer em grandes cidades, mas organizá-las sob condições financeiras reduzidas. Seu objetivo é portanto mais diretamente relacionado à realidade do *webcasting* no mercado internacional atual que, produzido ou financiado pelas próprias instituições sinfônicas, acaba por ver-se constrito a soluções econômicas de baixo custo o que pode acabar por afetar objetivamente a qualidade artística de todo o projeto²⁰. E é exatamente sobre metodologias para a melhoria das soluções artísticas nesta realidade que Paulu desdobra suas reflexões. Talvez a questão mais estimulante do trabalho esteja em sua teorização sobre os problemas da transmissão por tecnologia audiovisual de música sinfônica – a primeira referência documental da questão encontrada por esta pesquisa.

Algumas questões se apresentam para nós e, estando de algum modo latentes mas não desenvolvidos nos trabalhos supracitados, serão parte da tarefa da presente pesquisa. A primeira, é lógico, sobre a dita pertinência visual da música sinfônica, que a nosso ver pode ser melhor argumentada e atualizada a partir da mais recente discussão não apenas sobre os elementos objetivos da performance presencial como o da chamada "digital performance"²¹; acreditamos que esta discussão se dê a partir de dois eixos principais: os elementos de compreensão das propriedades desta nova performance audiovisual em seu *technological bias* (assumidos nos termos de Innis e atualizados por Manovich), quanto à codificação da performance para câmera e sua decodificação por parte da audiência através da tela; a conceituação de *live* e *real time* a partir das categorias de interatividade, sugeridas por Dixon (2007, pp. 563-598). O resultado deve reforçar a prerrogativa da diferença entre a experiência de um concerto presencial e aquele difundido por meio de tecnologias audiovisuais, não só pela ruptura dos paradigmas de visualização dada pela

¹⁹ A estação educativa licenciada para a University of Minnesota Twin Cities, que organiza-se como uma plataforma estudantil não comercial.

²⁰ O valor total para o projeto foi \$18,630.00 dólares americanos, que rendem atualizados \$168,459.86 para os nove programas, ou cerca de \$18.000 dólares por concerto. O valor não inclui gastos de pessoal do canal de televisão, assim como de transporte.

²¹ Este questionamento passa por toda uma literatura de *medium theory* que, como veremos, tem desdobramentos atuais nas obras de Lev Manovich ("The language of new media", 2001) e Steve Dixon (Digital Performance, 2007].

perspectiva do palco italiano, e sua formalização na sala de concertos (ou “caixa preta”), mas pela própria reorganização da experiência de presença e interatividade.

A segunda questão que a presente pesquisa deverá ater-se diz da não irrelevante problemática da situação do gênero ao qual devemos situar tais transmissões, o que para nós está no arcabouço de todo o problema artístico deste empreendimento. Seja entendendo-o como uma nova modalidade, o "Digital Broadcast Cinema" como considera por bem convencionar Paul Heyer(2007) seja assumindo-o como um gênero híbrido entre live theater e televisão, como sugere James Steichen (2009) a atualização da questão deve partir da discussão proposta por Paulu, ou seja: para efeito da gramática e retórica da transmissão audiovisual destas performances, seus elementos lógicos de construção e estilo de expressão, há de se preservar a distinção entre concertos sinfônicos de música instrumental e os gêneros dramáticos-musicais narrativos como a ópera e o balé.

Paulu justifica os músicos na tela por seis motivos: 1) o interesse de uma audiência em ver e entender o trabalho dos músicos, o que, ademais, fortalece a sensação de participar, “estar dentro” da orquestra; 2) dependendo da habilidade do roteiro, pode tornar o reconhecimento dos temas musicais mais fácil quando temos os instrumentos executantes desses temas na tela; 3) reforça, a partir do reconhecimento dos movimentos de corpo do músico, aspectos morfológicos ou emocionais da música; 4) a oportunidade de permitir a melhor inteligibilidade de relações formais - como a concatenação de temas musicais - na medida em que a montagem pode passear entre instrumentos distantes no espaço e descrever visualmente os timbres executantes; 5) por mimese, as reações emocionais da audiência acabam por conectar-se às reações emocionais dos instrumentistas; 6) finalmente, todo o repertório fotográfico do cinema (close-ups, superposições e movimentos de câmera) reforça efeitos de clímax e outras nuances dinâmicas (Paulu 1953 pp. 159-160).

Reivindicando elementos de pré-produção e planejamento, com mapa da orquestra, o posicionamento das câmeras coadunado à arquitetura da sala de concertos, e roteiro de edição que toma por base a análise da partitura musical (inclusive com exemplos graficamente pertinentes), o trabalho de Paulu segue como um rico e interessantíssimo detalhamento sobre os procedimentos de filmagem, assim com a solução de cada parte do processo de filmagem para alguns repertórios específicos trabalhados no projeto. Trata-se de uma referência inequivocamente importante para a história da transmissão audiovisual de orquestras sinfônicas e a discussão sobre o processo e relevância da filmagem daqueles concertos, em Minneapolis na década de cinquenta, tange aspectos os mais ricos e ainda pertinentes quanto à relevância e impacto, o potencial de novos serviços institucionais, visão artística e crítica, assim como a formação de plateia e novos públicos nos dias de hoje. De algum modo, ao pensarmos a atual transmissão ao vivo de orquestras sinfônicas, e decidimos os termos artísticos de sua exibição, estamos por fim a atualizar os termos tal como postos por Manvell e Paulu em seus estudos seminiais.

Com o avanço da tecnologia, diversos modelos de gravação dos concertos acabaram por ser disponibilizados – alguns deles já em fase de teste e execução em grandes orquestras do mundo. As dificuldades tecnológicas, no entanto, seguem prementes. A necessidade de silêncio por parte de maquinário e operadores fez com que câmeras controladas remotamente passassem a ser uma necessidade recorrente. Suas limitações artísticas, no entanto, seguem evidentes tanto pelo plano fixo, quanto pela

dificuldade de movimentação coordenada necessária para um *real time streaming*. A necessidade de "proximidade" com os músicos e as limitações de posicionamento no palco fizeram com que o investimento em lentes e câmeras de alta definição criassem um novo padrão, mas as restrições orçamentárias óbvias do universo da Cultura criam desafios para a aquisição de tais equipamentos como parte rotineira das instituições sinfônicas. Mas a capacitação artística dos produtores deste novo gênero segue como um desafio maior para difusão do espetáculo mantendo a sua qualidade, sobretudo entre orquestras de médio e pequeno porte como aquelas da América Latina, ou de cidades médias norte-americanas.

A pesquisa aqui proposta se vale de experiências com diversos modelos de gravação, em que as câmeras contam com o uso de operadores, como no caso das experiências junto à Orquestra Sinfônica do Estado de São Paulo (cinquenta concertos com a produção da TV Cultura e o uso de sete câmeras) e à Orquestra Sinfônica de Toronto (sete concertos com a produção da Symmetrica Inc., e o uso de doze a quinze câmeras, sem operadores, pré-fixadas). A estas devemos somar alguns estudos que permitam mapear novos modelos, todos em o fim de produzir uma metodologia própria que dê subsídios à capacitação e treinamento de equipes artisticamente qualificadas para uma transmissão sinfônica de alto nível. Ainda em fase de pré-projeto, ao se produzir tal metodologia de filmagem durante o *streaming* de alta resolução de imagem e som, prevendo os problemas apresentados, pretende-se produzir soluções para a melhoria de um espetáculo conhecido por "telespectadores", mas em construção ainda no ambiente digital. Em uma instância mais complexa, pretende-se ao final sugerir uma linguagem inovadora para este tipo de espetáculo partindo do princípio que há um público à espera de um espetáculo contemporâneo.

2.10. Considerações finais

A pesquisa sobre estereoscópica da imagem em movimento tem se mostrado como a mais estável do grupo de pesquisa do LabCine. A estereoscopia teve mais uma onda de sucesso no cinema de entretenimento de extração hollywoodiana, e especula-se ter agora sido esgotada. Acredita-se que só a América Latina ainda vende filmes 3D para o público de cinema tradicional. Mas a estereoscopia não deixa de ser um instrumento importante na ciência. Assim como foi usada para visualizar as armas inimigas na segunda guerra mundial pela *Photographic Reconnaissance Unit* na Inglaterra ela foi essencial para combater os Mísseis V da base de Peenemunde na Alemanha, invisíveis pela fotografia aérea tradicional. Nos dias de hoje, a rover Curiosity fotografa em 3D para produzir profundidade de campo estereoscópico do planeta Marte. Ou seja, a estereoscopia tem ainda muito campo a ser explorado como visão remota. Assim como a fotogrametria que pode ser usada para reconstruir objetos inexistentes (de museus, por exemplo) ou objetos imaginários, supostos por teorias científicas como vírus, bactérias, proteínas etc. Mesmo que nossa pesquisa tenha se aprofundado no campo da cultura, a estereoscopia reorganiza o olhar tradicional fotográfico e proporciona outro espaço psíquico da realidade. Os sistemas colaborativos de visualização terão muitos desafios a partir dos conteúdos estereoscópicos. Imagina-se que elaborações científicas teóricas, arquitetura, museologia, exploração espacial, reconstrução corporal sejam campos prolíficos para o uso colaborativo da estereoscopia e fotogrametria.

A agenda da colaboração no campo sonoro sempre fica em segundo plano e com o projeto de *streaming* de orquestras, mesmo que visando uma metodologia de filmagem adequada ao sistema sinfônico, o LabCine pretende avançar a respeito de qualidade de som sendo transmitido via internet. Além, claro, da perspectiva de um roteiro imagético que se organiza a partir do som e não o contrário - que é o tradicional. Finalmente, o editor de vídeo SenseMaking pode representar uma plataforma econômica e dinâmica de produção de vídeos didáticos. Mas, mais do que isso, o SenseMaking poderá ser um sistema de colaboração de pesquisa e ensino, se chegar a se tornar uma plataforma de visualização de artigos científicos acessados durante as palestras e demonstrações.

As pesquisas do LabCine (Laboratório de Artes Cinemáticas e Visualização) de Universidade Presbiteriana Mackenzie produziu reflexões, pesquisas e soluções que envolvem tecnologias de imagem (e de som) de forma contínua e prolífica. Com poucas bolsas de pesquisa, o LabCine tem conseguido aprovação de projetos em diferentes agências de apoio: Finep, RNP, Capes, Secretaria de Cultura do Estado de São Paulo. Durante seus 4 anos de existência, foram publicados 02 livros, 17 artigos em periódicos internacionais e 10 artigos em anais de eventos. Foram realizados 5 filmes em resolução 4K, sendo dois estereoscópicos. Em andamento estão sendo realizados dois filmes e duas plataformas de bancos de imagem em movimento. O editor SenseMaking ainda procura um parceiro financiador, mas percebe-se grande potencial de aplicabilidade para ensino e pesquisa e também como plataforma de visualização de dados acadêmicos.

2.11. Referências

- Adams, Gavin. “Um Balanço Bibliográfico e de Fontes da Estereoscopia”. São Paulo: Anais do Museu Paulista (2003).
- Adeyemi-Ejeye, A., Alreshoodi, M., Walker, S.D. “Implementation of 4kUHD HEVC-content transmission”. *Multimedia Tools and Applications*, p.1-20 (2016).
- Almeida, Jane de et. al. “Passages on Brazilian scientific cinema”, In: *Public Understanding of Science*. <https://doi.org/10.1177/0963662516683638>, (2017).
- Eliseo, M., da Silva, C., Prates, H., Poser, V., Stalbaum, B. and Furtado, N. “Sensemaking: A Proposal for a Real-Time on the Fly Video Streaming Platform”. In: *Creative Education*, 7, p. 2515-2523. doi: 10.4236/ce.2016.716238 (2016)
- “4K: imagens de quatro quilates”, In: Perissinotto, P. & Barreto, R. *FILE 2008 000 000*, São Paulo, IMESP (2008).
- André, T.; Almeida, J.; Silva, C. (orgs.). *CineGrid: Futuros Cinemáticos*. São Paulo, PRCEU-USP (2016).
- Boeke, K. *Cosmic view: the Universe in forty jump*. New York: The John Day Company, 1957.
- Brito, Jorge Nunes; Coelho Luiz. *Fotogrametria Digital*. 1ª ed. Rio de Janeiro: EdUERJ (2007).
- California Institute of Technology. “PIA16051: first high resolution color mosaic of curiosity’s mastcam images” (2012).
- Costa, Antonio. *Compreender o cinema*. 2. ed. São Paulo: Globo (1989).

- Dixon, Steve. "Digital Performance". The MIT Press, Cambridge Massachusetts (2007).
- Doherty, Thomas. "3-D Is Comin' at Ya!". In: The Chronicle of Higher Education (2011). Academic OneFile. Web. 23 Nov. (2011).
- Eliseo, M. A.; Casac, B. S.; Gentil, G. R., "A comparative study of video content user interfaces based on heuristic evaluation," 2017 12th Iberian Conference on Information Systems and Technologies (CISTI), Lisbon (2017) pp. 1-6. doi: 10.23919/CISTI.2017.7975820.
- Endelman, Lincoln. "A Brief history of high speed photography 1851-1930". In: Rochester Institute of Technology RIT Scholar Works (1988). <http://scholarworks.rit.edu/article/369/>. Acesso em: 25 jul. 2017.
- Evans, A. "Science fiction in France: a brief history". Fiction Science Studies, v. 16, n.3, p. 254-276, (Nov. 1989).
- Foster, Shaun; Halbstein, David. Integrating 3D modeling, photogrammetry and design. London: Springer (2014).
- Freeman, D., Santosa, S., Chavalier, F., Balakrishnan, R. and Singh, K. LACES: "Live Authoring through Compositing and Editing of Streaming Video", In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI'14, Toronto, (2014), pp. 1207-1216.
- Gomes, R. D. et al. "A solution for transmitting and displaying UHD 3D raw videos using lossless compression". In: Proceedings of the 19th Brazilian symposium on Multimedia and the web, WebMedia 2013, Salvador, p. 173-176 (2013).
- Graf, Herbert. "Opera in Television in NBC Columbia University Broadcasting Series" (1946). Music in Radio Broadcasting Literary Licencing, LLC (2013).
- Hamidouche, W.; Cocherel, G.; Feuvre, J. Le; Raulet, M.; Déforges, O. "4K real time video streaming with SHVC decoder and GPAC player", In: Multimedia and Expo Workshops (ICMEW), 2014 IEEE International Conference on (2014).
- Herschdorfer, Nathalie (Edited by). The Thames & Hudson Dictionary of Photography. Thames & Hudson, London (2015).
- Heyer, Paul. "Live from the met: digital Broadcast Cinema, Medium Theory, and Opera for the Masses". In: Canadian Journal of Communication, vol. 03 (2008).
- Hunt, Neil; Uribe, Carlos A. Gomez-. "The Netflix Recommender System: Algorithms, Business Value, and Innovation". In: Association for Computing Machinery (ACM). DOI: <http://dx.doi.org/10.1145/2843948>. (2015).
- Incunables du Cinéma Scientifique. Produção: ICS, CNRS AV. França (1984). Online. http://videotheque.cnrs.fr/index.php?urlaction=doc&id_doc=22&rang=3.
- ISO, 9241-210:2010. "Ergonomics of human-system interaction -- Part 210: Human-centred design for interactive systems", In: The International Organization for Standardization, 2010. http://www.iso.org/iso/catalogue_detail.htm?csnumber=52075. Kitamura, M.; Shirai, D.; Kaneko, K. et al. "Beyond 4k: 8k 60p live video streaming to multiple sites., Future Generation Computer". Future Generation Computer Systems, Vol 07 (2011) pp. 952-959. <https://keio.pure.elsevier.com/en/publications/beyond-4k-8k-60p-live-video-streaming-to-multiple-sites>

- Krippendorff, K., Butter, R. "Semantics: Meanings and Contexts of Artifacts", In Schifferstein, H.N.J.; Hekkert, P. (Eds.). *Product experience*. New York: Elsevier (2007). p.353-376.
- Li, Baochun; Wang, Zhi; Liu, Jiangchuan; Zhu, Wenwu. "Two Decades of Internet Video Streaming: A Retrospective View" (2013), In: Association for Computing Machinery (ACM). Doi: 10.1145/2505805.
- Lin, Y.; Wang, H., Lu, H. "A New Online Video Editing Tool", In: Proceedings of the 10th WSEAS international conference on applied informatics and communications, and 3rd WSEAS international conference on Biomedical electronics and biomedical informatics. Taipei, (2010), p. 465-470.
- Manovich, Lev. *The Language of new Media*. The MIT Press, Cambridge Massachusetts, (2001).
- _. "Teoria dos Nurbs". PERISSINOTTO, Paula; BARRETO, Ricardo.(orgs.), In: *Teoria digital*. Trad. Cicero Inacio da Silva & Jane de Almeida. São Paulo: Imprensa Oficial do Estado de São Paulo, 2010.
- _. *Data Draft. Archiving Media and Data Art in the 21st Century*. Latvia: Riga (2015).
- Manvell, Roger. "Experiments in Broadcasting and Television", In: *Hollywood Quarterly*", Vol. 2, No.4. University of California Press (1947). pp: 388-392. Meng, Horace J.; Zhong, Di; Chang, Shih-Fu. "Searching and editing MPEG-compressed video in a distributed online environment". *Journal Multimedia Systems*, Volume 7, Issue 4, pp 282-29 (1999).
- Michaelis, Anthony R.. *Research Films in Biology, Anthropology, Psychology, and Medicine*. Academic Press Inc, New York, (1955).
- Nielsen, J., Loranger, H. *Prioritizing Web Usability*. Pearson Education (2006). Paulu, Burton. "Televising the Minneapolis Symphony Orchestra", In *The Quarterly of film Radio and Television*, Vol 8 N. 2 (1953).
- Petrangeli, Stefano; Hooft, Jeroen van der; Wauters, Tim; Huyssegems, Rafael; Alface, Patrice Rondao; Bostoën, Tom; Turck, Filip De. *Live Streaming of 4K Ultra-High Definition Video over the Internet*. In *Proceedings of the 7th International Conference on Multimedia Systems (MMSys 2016)*, Klagenfurt, Austria, Article No. 27 (2016).
- Remondino, Fabio. *Accurate and Detailed Image-Based 3D Documentation of Large Sites and Complex Objects*. In: *Digital Imaging for Cultural Heritage Preservation: Analysis, Restoration, and Reconstruction of Ancient Artworks*. STANCO, Filippo; Battiato, Sebastiano; Gallo, Giovanni (orgs.) Boca Raton: CRC Press (2011).
- Rosenblum, Naomi. *A world history of photography*. New York: Abbeville Press, (2007).
- Rouille, A. *A fotografia: entre o documento e a arte contemporânea*. São Paulo: Editora SENAC (2009).
- Sanz, Cláudia Linhares. "Entre o tempo perdido e o instante: cronofotografia, ciência e temporalidade moderna". In: *Bol. Mus. Para. Emílio Goeldi. Cienc. Hum.*, Belém, v. 9, n. 2, p. 443-462, maio-ago. (2014).
<http://www.scielo.br/pdf/bgoeldi/v9n2/a11v9n2.pdf>

- Smite, Rasa; Smits, Raitis; Steichen, James. "The Metropolitan Opera Goes Public: Peter Gelb and the Institutional Dramaturgy of the Met: Live in HD". In: *Music and the Moving Image*, Vol. 2 Summer (2009).
- Stalbaum, Brett; da Silva, Cicero I. The sensor network is the computer. Internet Freedom Festival. Valencia, 2016. DOI: <http://dx.doi.org/10.13140/RG.2.1.4137.6402/1>.
- Vesna, Victoria. "Database Aesthetics". Minnesota: University of Minnesota Press (2007).
- Wankel, Charles. *Streaming Media Delivery in Higher Education: Methods and Outcomes*. New York: IGI Global (2011).
- Whittaker, R. Video Editing, Part VII (2017). <http://cybercollege.com/tvp056.htm>.
- Wien, Mathias. *High Efficiency Video Coding*. Berlin: Springer-Verlag Berlin Heidelberg (2015).
- Willis, Holly. *Fast Forward: The Future(s) of the Cinematic Arts*. New York: Columbia University Press (2016).
- Ye, Yan; He, Yuwen; Xiu, Xiaoyu. Manipulating Ultra-High Definition Video Traffic. In *IEEE MultiMedia*, vol. 22, no. 3, pp. 73-81 (2015).
- Zone, Ray. *Stereoscopic Cinema and the Origins of 3-D Film 1838-1952*. Kentucky: The University Press of Kentucky (2007).

ANEXO 1: biografia dos autores



Jane de Almeida: Professora orientadora do Mestrado e Doutorado em Educação, Arte e História da Cultura da Universidade Mackenzie e coordenadora do Laboratório de Artes Cinemáticas (LabCine).



Maria Amelia Eliseo: Doutora pelo Instituto Tecnológico de Aeronáutica (ITA). Professora da Faculdade de Computação e Informática (FCI) da UPM. Pesquisa: Web Semântica, Mundos Virtuais Semânticos, Realidade Virtual e Aumentada.



Helena Prates: Jornalista, especialista em Jornalismo Cultural pela PUC/SP, mestranda em Educação, Arte e História da Cultura pela Mackenzie SP. Pesquisa: o audiovisual na internet, tendo a plataforma Netflix como estudo de caso. Foi bolsista CAPES/PROSUP.



Vic Von Poser: Mestre do Programa de Educação, Artes e História da Cultura na Universidade Presbiteriana Mackenzie- SP. Graduada em Comunicação e Mídias pela PUC-SP (2011). Atualmente, é mestranda em Fine Arts Digital da UAL Camberwell College.



Patrícia Gimenez: É formada em Cinema pela FAAP. Atua no mercado cinematográfico brasileiro há mais de 18 anos. Mestre em Educação, Arte e História da Cultura pela Universidade Mackenzie.



Amanda Areias: Fotógrafa, mestranda no programa de EAHC na Universidade Mackenzie. Bolsista CAPES/PROSUP.



Leandro Oliveira: Produtor de conteúdo para eventos e publicações para empresas e instituições culturais como OSESP, entre outros. Colunista de cultura do jornal “O Estado de São Paulo”. Realiza doutorado em Educação, Arte e História da Cultura pela Universidade Mackenzie. Bolsista CAPES/PROSUP.



Mateus Teixeira: Mestrando do programa de Pós-Graduação em Educação, Arte e História da Cultura da Universidade Mackenzie (UPM). Graduado em Artes Visuais pela Faculdade de Educação e Cultura Montessori. Bolsista CAPES/PROSUP.



Cicero I. da Silva: Professor e pesquisador da Universidade Federal de São Paulo (Unifesp). Coordena o Núcleo Telessaúde Brasil Redes Unifesp. Pesquisa mídias digitais aplicadas à educação, saúde, cultura e sociedade.

Capítulo

3

Presenças do nosso futuro. O avanço dos sistemas de videocolaboração para uma Dança Expandida

Ivani Santana¹

¹ Universidade Federal da Bahia (UFBA). Instituto de Humanidades, Artes e Ciências Prof. Milton Santos, Programa de Pós-graduação em Artes Cênicas (vice-coordenadora), Grupo de pesquisa Poéticas Tecnológicas: Corpoaudiovisual (líder)

ivani@ufba.br

Abstract

The purpose of this article is to demonstrate that videocollaboration systems can be used in the performing arts to contribute to the development of telematic dance. In the light of situated cognition we assume that the mind is embodied, and that experience and the sensory-motor process are important in the construction of the conceptual system of the individual. With this, this language is understood as "expanded dance" providing new notions of presence. The research is based on my experience of more than a decade in the creation, direction and performance in telematic dance between several Brazilian cities and abroad, most of them built with the Poetic Technological Research Group: Corpoaudiovisual.

Resumo

O objetivo deste artigo é demonstrar que sistemas de videocolaboração podem ser utilizados nas artes cênicas para contribuir com o desenvolvimento da dança telemática. À luz da cognição situada assumimos que a mente é corporificada e que a experiência e o processo sensório-motor são importantes na construção do sistema conceitual do indivíduo. Com isso, compreende-se essa linguagem como "dança expandida", irrompendo novas noções de presença. A pesquisa está fundamentada na minha experiência de mais de uma década na criação, direção e atuação em dança telemática, em várias cidades brasileiras e no exterior, a maioria construída com o Grupo de Pesquisa Poéticas Tecnológicas: Corpoaudiovisual.

3.1. Introdução

Este texto aborda o campo da dança realizada com a utilização de tecnologias digitais, ao qual designamos “dança com mediação tecnológica”, uma vez que o foco não está no uso da tecnologia na cenografia, na iluminação, na música ou em qualquer outro campo correlato à dança. O foco está nos processos artísticos cuja relação com as tecnologias digitais é determinante para a própria ação do bailarino e para a criação da obra coreográfica ou de improvisação, contextos em que a dança apenas ocorre nessa articulação. Desta forma, delimitamos nossa pesquisa e reflexão nos contextos em que o bailarino e a tecnologia estão implicados através de alguma interface, seja por dispositivos acoplados ao corpo do sujeito ou responsáveis por tornar o ambiente sensível, por sistemas de captação, gravação, transmissão, dentre outros. O importante é que a dança realizada nesse contexto apenas pode ocorrer pelo acionamento interdependente desses dois sistemas: o biológico e o não biológico.

Para isso, é preciso compreender, primeiramente, o que estamos denominando aqui como “dança com mediação tecnológica”. Vejamos assim sua origem histórica. No final do século XIX, com a chegada da iluminação elétrica, os palcos do teatro ganharam novas formas de visualização e, conseqüentemente, novos conceitos e estéticas. A luz não estava mais ali apenas para iluminar o ambiente: ela permitiu que o palco se transformasse em um mundo a parte, delimitado até a quarta parede, cuja existência se separava daquela realidade da platéia com seres do cotidiano. Com a tecnologia da iluminação cênica, a separação desses dois universos – do imaginário e da realidade, palco e platéia respectivamente –, o olhar do público pôde ser direcionado, uma vez que poderia perceber apenas aquilo que o iluminador apresentava para ser visto, tornando a iluminação um elemento importante na criação da própria narrativa da obra. Conforme consta no blog¹ do *lighting designer* Valmir Perez, a iluminação cênica proporcionou novas descobertas visuais, tais como a invenção do ciclorama, originalmente denominado *Kuppelhorizont* pelo seu criador, o cenógrafo Mariano Fortuny (1871-1949), o qual propiciou uma nova sensação da cena quanto à sua altura, à arquitetura do cenário, à sensação de infinito, entre outros aspectos. Dessa forma, a luz elétrica foi uma das grandes responsáveis pela mudança radical da estrutura teatral.

As inovações quanto à iluminação elétrica reverberaram na criação da bailarina e criadora americana Loïe Fuller (1862/1928), uma das pioneiras da dança moderna, considerada a primeira artista a perceber que a tecnologia poderia estar além do uso apenas funcional, podendo servir como um elemento da própria produção do corpo e da estética da obra. Fuller criou as *Serpentines dances*, cuja existência dependia da articulação entre seus movimentos do corpo, vestido em grandes túnicas esvoaçantes, com a refração da iluminação. Portanto, a energia elétrica não estava apenas iluminando a cena: para além disso, ela servia de mediação tecnológica para que uma poética emergisse daquele sistema. Assim como em Fuller, os espetáculos, instalações e outros formatos de dança com mediação tecnológica existem apenas quando essa relação se estabelece.

No caso deste artigo, o campo de interesse é a dança com mediação tecnológica e, mais especificamente, a dança distribuída, dança telemática ou dança em rede,

¹ Laboratório de Iluminação da UNICAMP, <http://www.iar.unicamp.br/lab/luz/dicasemail/dica26.htm>

nomenclaturas utilizadas para o contexto no qual bailarinos são postos em locais distintos, conforme será apresentado na próxima sessão. Essa configuração artística apresenta uma série de aspectos e condições que propiciam novas formas de criação, execução e fruição da dança; para tanto, sistemas computacionais devem ser considerados, assim como nas *Serpentine dances* de Fuller, para além da ideia de mera ferramenta, pois uma mediação tecnológica implica todos os elementos desse contexto, o qual ganham novas concepções, novos conceitos artísticos e estéticos.

Para avançar com os estudos e aplicações sobre a relação da tecnologia de rede com a dança é preciso explorar tanto os dispositivos tecnológicos, seus sistemas e ferramentas, como os processos de percepção em ambientes remotos. O interesse desse artigo é refletir sobre uma nova compreensão de “presença”, que será investigada de acordo com teorias das ciências cognitivas, pela vertente da cognição situada. A partir dos conceitos *embodied*, *embeddedness* e *extended cognition*, que integram o sistema sensorio-motor como parte importante do processo de conhecimento e, portanto, da nossa cognição, poderemos afirmar que as condições telemáticas estão implicadas na maneira como o criador pode perceber e agir nesse contexto. Como já mencionado, para um artista da dança (criador, coreógrafo, bailarino, diretor), a interação entre sujeitos remotos provoca uma série de alterações na forma como ele cria, pois deve levar em conta não apenas o corpo e o movimento, mas a forma como o dispositivo manipula, controla, dispara, e como será visto, escutado e sentido pelo outro e pelo público. Tais aspectos modificam o modo como o movimento é realizado uma vez que cada tecnologia implicará certas possibilidades, assim como trará também algumas restrições, e todas essas condições determinarão a fruição do público. Com esses estudos, pretendemos analisar os contextos e configurações já testados, bem como prospectar quais aspectos seriam preciosos e oportunos para o avanço de sistemas, como a videocolaboração no campo das Artes Cênicas, principalmente da Dança.

A seção 3.2 deste artigo faz uma breve contextualização do campo da dança telemática; a seção 3.3 apresentará os pontos principais dos espetáculos, produzidos pelo Grupo de Pesquisa Poéticas Tecnológicas: Corpoaudiovisual e que são relevantes para a discussão; a seção 3.4 pretende fundamentar a reflexão por meio dos conceitos das ciências cognitivas provenientes da perspectiva corporificada da mente; a seção 3.5 pretende contribuir com algumas visões de futuro; e, por fim, na seção 3.6, faremos uma breve conclusão.

3.2. Contextualização

A dança começou sua articulação entre corpos distantes geograficamente com o projeto de Kit Galloway e Sherry Rabinowitz que utilizaram satélite da NASA para tal feito. Apesar de o primeiro ter sido com dança, o *Satellite arts project* (1975), a obra mais reconhecida desses artistas, nesse sentido, foi *Hole in spaces* (1980), que interligou duas cidades dos Estados Unidos, a saber, Nova York (*Lincoln Center for Performing Arts*) e Los Angeles (na frente da loja *The Broadway*). Sem nenhum aviso prévio, os pedestres que passavam por esses lugares começaram a perceber que havia uma reação da imagem projetada na parede e, de forma complementemente espontânea, descobriram que aquele era um local que poderia “furar” o espaço, a distância e inserir pessoas remotas em um mesmo ambiente virtual, um ciberespaço. Conforme afirmou Maria Chatzichristodoulou (2012), essa era a primeira vez que o limite da presença física era desafiado, colocando em tempo real a imagem de pessoas separadas pela distância e que, a partir de então, poderiam

estabelecer projetos colaborativos. Justamente por essa razão, esses pioneiros definiam esse contexto como espaço sem fronteiras.

Telematic dream (1992)², de Paul Serman, é outro projeto de destaque nessa área. O artista sempre teve interesse na pesquisa artística com os sistemas de telecomunicação e essa obra foi um dos seus grandes marcos. O projeto tinha como proposta estabelecer uma relação entre o público e a *performance*, por meio da transmissão de imagem em tempo real. A bailarina ficava em uma cama e sua imagem era projetada em cima de outra cama acessível ao público, o qual tanto mantinha uma relação de *voyeur*, como de coadjuvante interagindo com aquele corpo virtual.

Nessa mesma época, Lisa Naugle e John Crawford iniciaram vários projetos que denominaram dança distribuída, dança telemática ou *performance* em rede. Sobre a obra *Janus/ghost stories* (1999),³ realizada entre universidades americanas, Naugle afirma:

Networked performance is a synchronous approach to communication; that is, a shared activity between two or more people who are collaborating at the same time. Collaborators may be located at the same place or in different places. Using video-conferencing systems, people at different locations can see and hear each other simultaneously. This can be a two-way or multipoint method of communication. The basic system consists of computer, monitor, videocamera, microphone, and speakers at each site. [Naugle 2002, 56]

No meu texto *Novas configurações da dança em processos distribuídos das Redes* [Santana 2013], faço uma contextualização sobre o campo, não apenas indicando alguns dos primeiros pioneiros, como Galloway/Rabinovitz, Serman, Naugle/Crawford, como também cito a respeito da *Association of Dance and Performance Telematic, body>data>space*⁴ (ADAPT), dentre outros, e ainda procuro explicar o interesse sempre existente na humanidade pela comunicação à distância. Abordo também os vários títulos que a arte em rede recebe, sem encontrar uma concordância entre os pesquisadores e artistas da área, até mesmo porque não há necessidade de compreender da mesma forma todas as organizações provenientes das diversas linguagens artísticas; quer dizer, a configuração, os objetivos, os conceitos desejados na Música em Rede podem não ser os mesmos que os da Dança em Rede, ou dos trabalhos criados pelas Artes Visuais, ou ainda, da Arte Tecnologia. Para exemplificar, cito aqui algumas nomeações, muitas utilizadas apenas em inglês e outras apenas utilizadas nas artes cênicas: *ciberperformance*, *networked art*, *networked performance*, arte digital, ciberteatro, arte em rede, dança telemática, dança distribuída, dentre outras.

Conforme veremos na próxima sessão, venho desenvolvendo a dança telemática no Brasil desde 2005, buscando encontrar metodologias, conceitos, estéticas, configurações, preparação corporal e processos colaborativos que auxiliem no desenvolvimento de projetos dessa natureza, de acordo com os objetivos que norteiam minha pesquisa geral no campo, bem como as metas específicas de cada investigação que venho realizando.

² <http://www.paulsermon.org/dream/>

³ <http://dance.arts.uci.edu/linaugle/files/janus/index.html>

⁴ Disponível em: <http://www.bodydataspace.net/>

Para finalizar essa breve contextualização, apresento aqui algumas das ferramentas mais utilizadas nos projetos de arte em rede, a saber: *LoLa*⁵ [Drioli 2013], *Open Broadcaster Software*⁶, *Scenic*⁷, *Snowmix*⁸, *UltraGrid*⁹ [Gharai et al. 2006]; [Vieira et al. 2012], dentre outras internacionais. No Brasil, participamos do Grupo de Trabalho em Mídias Digitais e Artes (GTMDA), coordenado pela Dra. Tatiana Aires Tavares e com suporte da Rede Nacional de Ensino e Pesquisa (RNP), em 2009 e 2010, para o desenvolvimento da *Arthron*¹⁰ [Vieira et al. 2012], a qual utilizamos até 2013. Em 2014, utilizamos a ferramenta computacional *Telecorpo*, criada por Pedro Lacerda, bolsista do Grupo de Pesquisa Poéticas Tecnológicas: corpoaudiovisual, para a transmissão das imagens do Chile, Portugal e Brasil na obra *Personare*.

3.3. Experiências telemáticas do Grupo de Pesquisa Poéticas Tecnológicas: corpoaudiovisual

Nesta sessão, abordo alguns dos principais projetos realizados pelo Grupo de Pesquisa Poéticas Tecnológicas: Corpoaudiovisual: a obra *VERSUS*, realizada especialmente para o lançamento da Rede Ipê em 2005; apresento alguns experimentos realizados ao longo de uma década; e os últimos trabalhos com parceira internacional. Minha experiência com a telemática começou nos Estados Unidos, durante meu doutoramento no campo da Dança com mediação tecnológica, em 2001. Minha estadia no *Environments Lab* para uma residência artística, a convite do então coordenador, o Dr. Johannes Birringer¹¹, tinha como primeiro objetivo o estudo em captura de movimento (*motion capture*) que pude realizar no *Advanced Computing Center for the Arts and Design (ACCAD)*¹². Entretanto, como o *Environments Lab* estava sediado na *Ohio State University*, uma das universidades participantes da ADAPT, as sessões de dança telemática tornaram-se também parte das minhas atividades durante a residência artística. Contudo, a forma como realizavam os experimentos não me parecia coerente com o entendimento de Internet e de rede que eu havia estudado, uma vez que era feito como forma de transmissão de um para muitos, ou seja, uma compreensão ainda da tecnologia de massa como a TV. As sessões, àquela época, consistiam de apresentações individuais de cada instituição, por 10 minutos, e, ao final, todos realizavam suas performances ao mesmo tempo. Tal configuração não me parecia estar preocupada, ou mesmo ter interesse, que os bailarinos encontrassem movimentos e ações para interagir com seus parceiros remotos explorando as possibilidades da própria rede.

Ao final da minha temporada no *Environment Lab*, concebi a instalação performativa *DRYWET* (2001), realizada entre dois espaços: Studio V e Sullivant Theater. Essa obra foi importante para a realização do meu espetáculo *Pele* (2002)¹³, realizado durante o Ateliê de Coreógrafos Brasileiros - Ano 1, o qual fui contemplada como uma das cinco coreógrafas escolhidas em todo Brasil. Ressalto ainda que, ao

⁵ www.conservatorio.trieste.it/artistica/lola-project/lola-low-latency-audio-visual-streaming-system

⁶ <https://obsproject.com/>

⁷ <https://code.sat.qc.ca/redmine/projects/scenic/wiki>

⁸ <http://snowmix.sourceforge.net/>

⁹ www.ultragrid.cz/

¹⁰ <http://gtavcs.lavid.ufpb.br/downloads/>

¹¹ <http://www.brunel.ac.uk/people/johannes-birringer>

¹² <https://accad.osu.edu/>

¹³ <http://ivanisantana.net/videos/>

retornar dos Estados Unidos, tentei participar das sessões da ADAPT, mas a rede acadêmica naquela época não suportava projetos dessa natureza.

Apenas em 2005, a convite da RNP para o lançamento da Rede Ipê, que ocorreu junto com o evento em homenagem ao aniversário do Ministério de Ciência e Tecnologia, foi possível realizar a primeira obra da dança telemática através das redes acadêmicas de telecomunicação. O espetáculo *VERSUS* foi realizado com a parceria do Laboratório de Vídeo Digital (LAVID), coordenado pelo Dr. Guido Lemos. Iniciava então uma longa, duradoura e eficiente parceria entre o GPPoética e o LAVID. A obra foi realizada entre três cidades brasileiras: Salvador e Brasília com bailarinos e João Pessoa, onde estavam os músicos que fazia a trilha sonora em tempo real. Conforme informado anteriormente, meu interesse na telemática era estabelecer uma conexão multipontos, na qual os bailarinos realmente buscassem uma interação de acordo com as possibilidades do contexto remoto (Figura 3.1). Minha concepção, preparação com os bailarinos e execução da obra teve que ser pensada por outros caminhos e referências distintos do que tive como experiência nos Estados Unidos.



Figura 3.1 - VERSUS (2005), entre Brasília, Salvador e João Pessoa

Como trabalhava com mediação tecnológica desde a década de 1990, contando com várias experiências quanto ao uso de câmera e imagem em tempo real, utilização de sensores, dentre outros, criei a obra procurando uma verdadeira articulação entre os bailarinos, a partir dos conhecimentos com os quais já contava. Todavia, era preciso descobrir ainda como contar com a distância, com o atraso da imagem por conta da latência da transmissão (*delay*), como explicar de forma clara para os engenheiros de rede e de computação as metas do projeto para que eles tivessem informações suficientes para avançar com a tecnologia, de que forma fazer a mudança das imagens de uma tela para outra de projeção e assim por diante.

Desde o início do processo criativo de *VERSUS*, meu grupo foi dividido em dois e procuramos ensaiar sempre de forma distribuída, mesmo que, num primeiro momento, as imagens fossem enviadas diretamente da câmera para o projetor do outro ponto. O problema era que o envio da imagem em direto (via cabo RCA, à época) não tinha como indicar o *delay* que teríamos no espetáculo. Uma solução parcial para esse problema foi encontrada no *software* Isadora, com o qual poderíamos simular a latência da transmissão da imagem. De qualquer forma, trabalhar constantemente com as imagens, sejam elas em

direto ou com a simulação do *delay*, foi uma boa estratégia para preparar os bailarinos que tiveram a condição de compreender os tipos de relação que podiam ser estabelecidos com um corpo remoto. Essa noção de presença e de corporificação do contexto da telemática será discutida na próxima sessão. Por agora, o importante é mencionar a importância da vivência do corpo com a tecnologia, pois é dessa experiência que uma efetiva articulação ocorre. Não podemos pensar que basta colocar o bailarino num sistema sensível e seu corpo estará apto a utilizá-lo de forma criativa, pois é necessária a compreensão sensório-motora, como discutiremos adiante.

Minha experiência com o uso de câmera de vídeo no ambiente da dança, seja para transmissão de imagem ao vivo, ou mesmo criação de videodança, foi essencial para que os bailarinos tivessem condições de criar imagens interessantes para a articulação com o parceiro remoto. Não apenas os bailarinos, mas também aqueles que estavam responsáveis pelo uso da câmera precisavam compreender como utilizar o dispositivo naquele contexto específico da telemática. A construção da imagem e a relação corpo-câmera foram exploradas também no que tange à discussão sobre o corpo virtual, ou seja, interessava-nos mostrar as especificidades do corpo nesse contexto, como, por exemplo, a inexistência da gravidade nesse contexto (Figura 3.2).



Figura 3.2. VERSUS (2005). Uso da latência (delay) como elemento estético.

Em *VERSUS*, procuramos explorar a questão da imagem do corpo. Sendo assim, em Salvador instalamos dois espaços cênicos, um em que o corpo era captado e transmitido diretamente e outro pelo qual era enviada, através do *software* Isadora, a imagem processada dos bailarinos que lá estavam (Figuras 3.3 e 3.4). Por essa razão, contando com o espaço lotado de equipamentos, cabos e dispositivos que aquela época eram enormes, preferimos não ter público em Salvador, mas havia uma projeção do espetáculo em uma outra sala para aqueles que assistiam nessa cidade. Apenas em Brasília contamos com um público presente fisicamente. O espetáculo foi transmitido em tempo real e qualquer pessoa que acessasse o link poderia assistir pela internet.



Figura 3.3. VERSUS (2005). Uso do software Isadora.

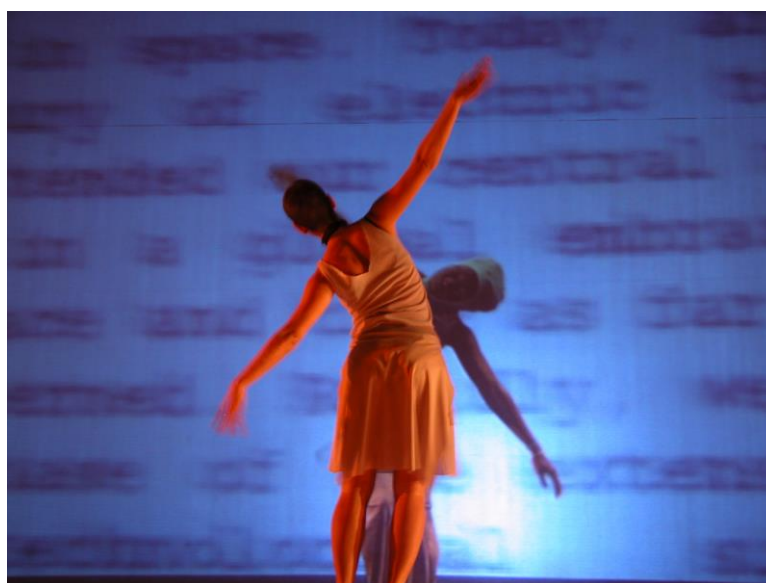


Figura 3.4. VERSUS (2005). Uso do software Isadora.

O espetáculo *VERSUS* serviu como um excelente início de investigação no campo da telemática pelo grupo de pesquisa, à época designado Poéticas Tecnológicas na Dança. Fundado em 2004 e contando com uma produção contínua e dinâmica, cada vez envolvendo profissionais, pesquisadores e estudantes de várias áreas, resolvemos alterar o nome do grupo para que ficasse mais coerente com nossa equipe, com os projetos e produções, passando, assim, a designar-se *Grupo de Pesquisa Poéticas Tecnológicas: Corpoaudiovisual*, sendo este neologismo acrescido para designar o corpo como elemento principal das investigações, cuja meta sempre esteve voltada para a mediação tecnológica, a qual, ao fim e ao cabo, refere-se à produções imagéticas e/ou sonoras, podendo contar ou não com telemática, interatividade, imersão, simulação ou outras propostas da Arte Tecnologia.

Após *VERSUS*, a relação corpo-câmera continuou sendo explorada no estudo *Por onde Cruzam Alamedas* (2006), criado entre dois espaços da Universidade Federal da Bahia, a saber, a Escola de Dança e a Reitoria. Tanto a construção da imagem, como a utilização do corpo desenvolveram os potenciais da relação corpo-câmera já explorados em *VERSUS*. Por exemplo, os espaços vazios eram investigados por esses dois termos, quer dizer, o bailarino dificilmente era centralizado ou se postava no centro da imagem, uma vez que o espaço devia ser preenchido com os dois parceiros remotos, sendo assim, o outro estava sempre como uma nova camada (Figura 3.5). O resultado dessa investigação demonstrou que, por meio desse processo, poderíamos explorar de forma mais efetiva a profundidade da imagem, bem como a complexidade da edição em tempo real, uma vez que nos aproximávamos da ideia de edição pelo jogo de sobreposição criado com as imagens transmitidas em tempo real e as videocenografias que poderiam ou não utilizar os bailarinos na gravação.



Figura 3.5. Por onde cruzam alamedas (2006).

O processo de camadas contava também com a relação de planos de câmera, pois eram criadas as “janelas” (Figura 3.6), conforme denominamos os espaços construídos com o corpo em primeiro plano. Demandas como essa, que valorizavam a articulação entre primeiro e segundo plano, ou ainda, o uso da diagonal para explorar a profundidade, propiciavam novos acionamentos sensório-motores. Por essa razão, afirmamos que não se trata apenas de efeitos especiais, mas de condições inéditas com as quais os bailarinos deviam desenvolver suas *performances*.



Figura 3.6. Por onde cruzam alamedas (2006). Enquadramento de câmera com primeiro e segundo plano para composição da telemática por camadas.

Em 2006, fui convidada pela coreógrafa e diretora espanhola Salud Lopes, do grupo *En lugar de creación* para participar do *Proyecto Paso* que abordava Declaração Universal dos Direitos Humanos e que contou também com a participação dos Estados Unidos. O projeto foi apresentado na Bienal de Artes de Sevilha e, em 2007, reapresentado como *Nukonén, paso ao Chile*, justamente por contar também com a participação dos nossos parceiros desse país. Com uma abordagem completamente diferente da minha, o objetivo não estava na articulação entre os bailarinos remotos, mas apenas na justaposição das imagens pela composição gráfica concebida e executada por Laura Hernández e Sergio Moreno (Figura 3.7).

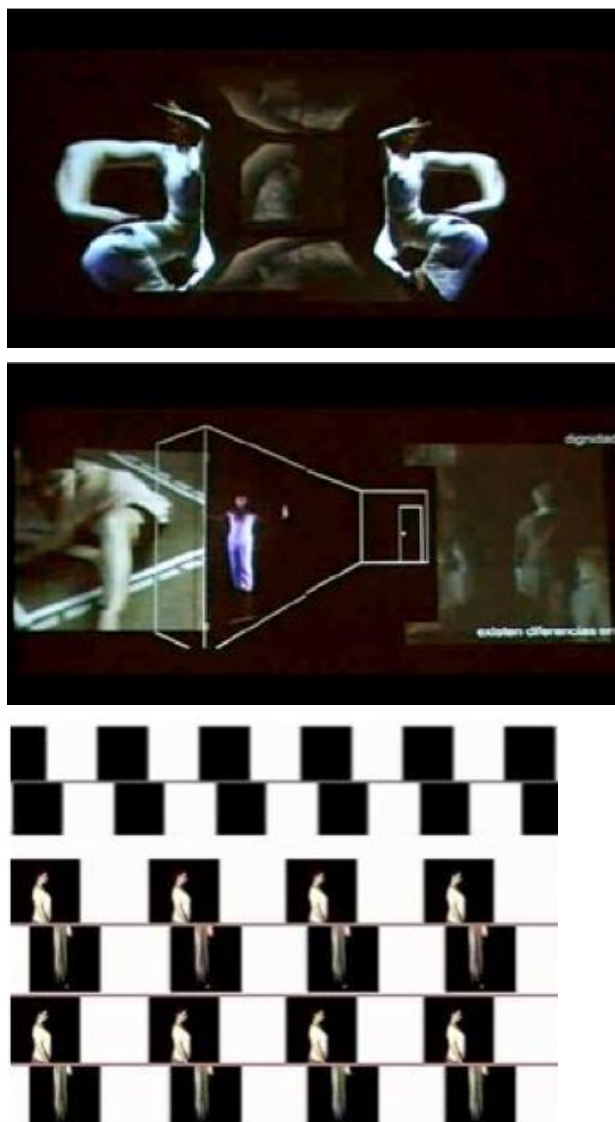


Figura 3.7. *Proyecto Paso* (2006), composição gráfica concebida e executada por Laura Hernández e Sergio Moreno.

Em 2009, participamos do Grupo de Trabalho em Mídias Digitais e Artes (GTMDA), coordenado pela Dra. Tatiana Aires Tavares. O GTMDA contava com a participação do LAVID, do GPPoética, dentre outras instituições, parceiros e redes de telecomunicação nacionais e internacionais. Como já tínhamos experiência no envio de dados, imagem e áudio em alta resolução, utilizando a rede nacional, minha sugestão ao grupo foi de começar uma parceria internacional, posto que precisávamos avançar na investigação sobre conectividade e projetos colaborativos em rede. Portanto, meu objetivo era explorar um processo criativo construído inteiramente através da Internet. Para tanto, foi necessário o desenvolvimento de uma metodologia através da qual eu tivesse condições de transferir conhecimento para poder cooperar com os grupos parceiros.

Na parte tecnológica, o GTMDA desenvolveu a ferramenta Arthron, baseada no sistema de gerenciamento de vídeo de alta resolução já concebido pelo LAVID, bem como foi estruturado a partir dos meus conceitos sobre espetáculos distribuídos. Talvez se o Arthron tivesse sido criado a partir do trabalho de Salud Lopes, a parte de composição gráfica do fluxo de informação seria mais importante a ser desenvolvida. Contudo, sendo a parceira estabelecida desde *VERSUS*, a ferramenta Arthron foi estruturada justamente para a determinação e gerenciamento de entradas (codificadores/*encoders*) e saídas (decodificadores/*decoders*) de imagem que poderiam ser combinadas e recombinadas sem limite, podendo ainda deixar programada cada organização temporal de *encoders-decoders* como uma cena (Figura 3.8).

Do ponto de vista tecnológico, esse gerenciamento, o qual poderia ser realizado também remotamente, facilitava a operação e diminuía a demanda de um engenheiro ou técnico por computador, trabalhando assim de forma mais coerente com a telemática. A possibilidade de combinação entre codificadores e decodificadores permitia uma valorização maior e um processo mais dinâmico da relação a) corpo1/câmera1, b) projeção1, c) corpo2/câmera2, d) imagem final. Para ficar mais claro e relacionando com os desenvolvimentos e resultados já apresentados acima, podemos explicar que: a) codificadores são as câmeras que gravam o bailarino, articulação essa que, conforme já mencionado, permite minha aplicação do processo de camadas e janelas; b) imagens que podem ser transmitidas em qualquer uma das telas distribuídas no espaço, contribuindo assim com c) a relação espacial do bailarino em d) diálogo remoto. Desta forma, não se trata apenas de ter um sistema com *input* e *output*, mas uma ferramenta computacional concebida de forma interdisciplinar que comprova a importância e necessidade de articulação entre as ciências exatas, humanas e as artes.

Do ponto de vista das artes e da dança mais especificamente, todo o processo do GTMDA e do desenvolvimento do Arthron foram determinantes para a construção de estratégias metodológicas para a concepção, preparação e execução de dança telemática que temos desenvolvido ao longo desses mais de dez anos. Cada ensaio, processo, apresentação e projeto criavam novas demandas para o corpo e para a tecnologia, num fluxo mútuo de aprendizado e construção.

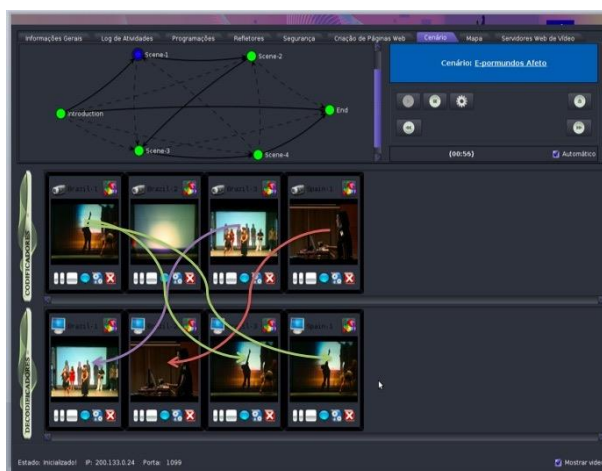


Figura 3.8. Arthron (GTMDA, 2009, 2010) e Pormundos Afeto (2009), entre Brasil e Espanha.

Todos os grupos e artistas que convidamos para os nossos projetos já trabalhavam com Arte Tecnologia, mas não tinham ainda experiência de processos artísticos em telemática. Para o GTMDA, convidei o grupo catalão Konic Thtr, com o qual foi possível uma pesquisa continuada durante três anos. Sem dúvida alguma, para o crescimento do uso da tecnologia de rede em processos artísticos, é importante o incentivo e fomento de pesquisas continuadas e não apenas de eventos esporádicos.

Outros processos colaborativos foram estabelecidos com o intuito de transferir conhecimento, pois acreditamos que um campo apenas pode crescer e se desenvolver se houver um número significativo de profissionais empenhados no tema. Além disso, cada projeto tinha seus próprios objetivos específicos que necessitava de parceiros apropriados, como no caso do *Laboratorium MAPAD2 de Arte Telemática*, contemplado com o edital da VIVO Lab e realizado durante dez meses no ano de 2011, finalizando com a apresentação da obra *Frágil*, entre o Museu de Arte Moderna do Rio de Janeiro e o Teatro da Universidade Federal do Ceará, como parte do evento Desafio de Arte em Rede, promovida pela RNP e pelo Ministério da Cultura, atividade que abriu o Fórum de Cultura Digital daquele ano. O projeto selecionou um grupo artístico em parceria com outro tecnológico, sendo essas duplas sediadas em Fortaleza e no Rio de Janeiro.

O *Laboratorium MAPAD2 de Arte Telemática* ficou estabelecido da seguinte forma: Grupo de Pesquisa Poéticas Tecnológicas (GPPoética), em parceria com a equipe do professor Celso Saibel, à época vinculado ao Departamento de Computação da Universidade Federal da Bahia, os quais coordenavam o projeto na parte artística e tecnológica; Laboratório de Poéticas Cênicas e Audiovisuais (LPCA), da Universidade Federal do Ceará, com profissionais do Departamento de Computação da mesma instituição, os quais estariam colaborando com as áreas do cinema e do teatro (mais especificamente voz); e do Núcleo de Arte e Novos Organismos (NANO), da Universidade Federal do Rio de Janeiro, com a colaboração do Grupo Telemídia da PUC Rio, com o qual foi possível uma primeira articulação com as artes plásticas. Sendo assim, esse projeto ampliava a delimitação de telemática utilizada até então, pois reunia em rede as várias linguagens artísticas, as quais, por sua vez, criavam demandas específicas para o desenvolvimento da arte telemática.

A estrutura metodológica do *Laboratorium MAPAD2* foi significativa e coerente, pois contou com um processo contínuo, com encontros semanais e, talvez o aspecto mais importante, realizou reuniões presenciais com todos os grupos, em quatro etapas distintas, cada uma ocorrendo em uma cidade. Esses encontros permitiram maior imersão e mais articulação das ideias entre as distintas linguagens, conseguindo avançar de forma eficiente nas resoluções de cada proposição que surgia. Talvez a presença física ainda seja significativa para avançarmos juntos no desenvolvimento de algum tema ou, talvez, o problema é que não estamos habituados a permanecer tanto tempo conectados, como era a proposta desses encontros que duravam praticamente um dia inteiro. Ademais, o ambiente tecnológico em rede, contendo as diversas possibilidades perceptivas que um encontro dessa natureza exige, não é acessível nas universidades e muito menos nos centros artísticos. Acredito que, tecnologicamente, já temos como ampliar as possibilidades perceptivas dos sistemas que atualmente estão centradas quase que exclusivamente na visualização e na escuta. Já é possível desenvolver um ambiente em que se possa sentir o outro, indo para além de apenas ver e escutar. Entretanto, um espaço completamente conectado como esse ainda é desejo de ficção científica, ao menos para as universidades brasileiras e, arrisco em dizer, para grande parte das instituições no

mundo. Desde o livro de Mark Dery, de 1996, *Escape velocity: Cyberculture at the end of the century*, as formas de ciberconexão já estavam apontadas. E agora, no século XXI, que estamos para além de conexões via sensores acoplados no corpo que permitem a interação remota, as tecnologias estão cada vez menores e mais ambientadas, ou seja, antes portávamos um *wii-remote* para interagir com a máquina, hoje um *xbox kinect* captura nosso movimento e faz a simulação e interação.

Como veremos na próxima sessão, o sentido de presença depende da relação que estabelecemos com o ambiente e com o outro e da forma como podemos agir em relação ao outro para percebê-lo, portanto, não se trata apenas do estado de estar aqui-e-agora e em carne-e-osso que faz sentido para a noção de presença no século XXI. Esses ambientes sensíveis transportam os sujeitos, possibilitando outras vias de articulação que não mais o contato físico. É essa forma de perceber o mundo e agir que está implicada no processo de aquisição de conhecimento.

Outro projeto colaborativo importante nessa trajetória foi o *Embodied en varios Darmstadt 58* (EVD58), realizado em 2013 e 2014. Esse projeto tinha como proposta o desenvolvimento do conceito “corpo (tele)sonoro” ao qual me dedicava desde 2010. Em 2013, além de convidar novamente o grupo espanhol para manter a continuidade na parceria, foi chamada a equipe da coreógrafa mexicana Rebeca Sánchez, que contou com a colaboração do Translab, coordenado por Minerva Rodriguez. Em 2014, tendo como meta avançar na investigação sonora, convidamos o músico e professor da Universidade de Santiago, Rolando Cori, e a equipe do professor Daniel Tércio, da Faculdade de Motricidade Humana, de Portugal. Esse desdobramento do EVD58 culminou com a performance *Personare* e foi realizada com outra ferramenta tecnológica chamada *Telecorpo*, desenvolvida por Pedro Lacerda, membro do GP Poética. Na parte de áudio, continuamos com a utilização do *Jacktrip*.

A descrição e reflexão sobre esses e outros projetos estão na edição número 2 de 2015 da *Revista Eletrônica MAPA D2*¹⁴, que homenageia os 10 anos de produção em dança telemática pelo Grupo de Pesquisa Poéticas Tecnológicas: corpoaudiovisual. Como citado anteriormente, o enorme avanço obtido quanto ao entendimento dos conceitos e delimitação do campo e quanto ao modo de conceber, atuar e executar a dança telemática ocorreu graças à longevidade dessa pesquisa que contou com contribuições de parceiros distintos e com aspectos importantes para investigação.

Vale a pena indicar aqui os passos metodológicos de processos colaborativos de arte em rede. Como esse não é o foco principal deste artigo, apresentamos a seguir alguns itens principais, fornecendo uma visão resumida das estratégias metodológicas utilizadas:

- Concepção do projeto indicando qual a proposta da arte em rede ou dança telemática, lembrando que são indissociáveis corpo e tecnologia numa pesquisa artística que pretende a mediação tecnológica.
- Definição dos parceiros, considerando a necessidade da proposta do projeto.
- Definição clara de um cronograma de trabalho para a equipe considerando:
 - pesquisa local, uma vez que a proposta deve ser discutida e testada localmente pelo grupo interdisciplinar;

¹⁴ <https://portalseer.ufba.br/index.php/mapad2/issue/view/1184/showToc>

- periodicidade nos encontros em rede para verificação da sua estabilidade, dos dispositivos e da integração entre os bailarinos remotos;
- Ensaios com sistema e conexão específicos que serão realmente utilizados na obra;
 - ➔ cronograma próprio de montagem, ensaio geral, apresentação e desmontagem, considerando o tempo de configuração de cada parte.
- Discussão sobre o tema, a configuração e estética da obra com os parceiros artísticos e tecnológicos de todas as equipes.
- Apresentação da versão inicial de um *storyboard* ampliado que inclui indicação para bailarinos, músicos e responsáveis pela execução dos *encoders* e *decoders*. No meu caso, a criação artística tem apenas pontos-chave, determinados *a priori*, para que cada grupo tenha autonomia estética no seu ponto de presença. Vale acrescentar que, ainda assim, considero que todos os pontos fazem parte de uma única obra, mesmo que haja distinções entre as partes.

O *storyboard* é desenvolvido ao longo do processo, de acordo com os resultados de cada ensaio. Quando a relação é criada com mais de dois pontos, o ideal é fazer encontros primeiramente entre duplas, verificando todos os problemas e demandas, para depois colocar todos em rede. Essa condição pode ser seguida a cada etapa de estruturação da obra.

- Executar os ensaios finais necessariamente com todos os dispositivos e rede que serão utilizados na apresentação para testar a estabilidade daquele determinado contexto.
- Montagem: acredito que uma melhor organização para montagem deve obedecer a seguinte ordem: cenário, caso exista; posicionamento dos projetores e telas; colocação dos equipamentos de áudio, vídeo e rede, principalmente quando utilizados no ambiente cênico; montagem e afinação da iluminação, considerando o recebimento e transmissão de imagem; teste de rede; reconhecimento do palco e trajetória dos movimentos em cada cena, com principal atenção nas transições; verificação do sistema de som (o ambiente deve estar completamente silencioso); teste de rede com os outros pontos de presença, caso seja mais de um: neste caso, melhor é fazer o teste com um ponto de cada vez, sendo sempre o ponto-gerente com um outro e, ao final, com todos os problemas e demandas resolvidos, a rede completa deve ser testada; ensaios geral considerando primeiro as passagens das cenas, com marcações de trajetória em relação à câmera e à projeção, para depois fazer um teste geral com uma apresentação completa; apresentação propriamente dita; desmontagem.

3.4. Corporificado, enraizado e expandido, conceitos das Ciências Cognitivas aplicados à Arte em Rede.

Meus estudos sobre a dança com mediação tecnológica encontrou, nos conceitos das Ciências Cognitivas, uma coerência com as reflexões obtidas empiricamente durante toda minha investigação. A vertente que me interessa desse campo interdisciplinar é o da cognição situada (*situated cognition*), podendo ser mencionada também como a perspectiva da mente expandida ou *embodied perspective*. Essas teorias estão contribuindo muito nos dois campos de interesse deste artigo: da dança e da tecnologia. Isso porque os estudos da percepção e cognição humana podem indicar caminhos para o desenvolvimento de dispositivos, sistemas e processos que auxiliam na relação do sujeito

com seu mundo contemporâneo e, nesse caso, do homem com as máquinas digitais. Lembrando que a Inteligência Artificial foi de grande importância nas primeiras abordagens das Ciências Cognitivas – e ainda é – a visão computacional da mente, entretanto, encontrou discordância de vários teóricos que não acreditavam na metáfora da mente como um *software* e o corpo como um *hardware*. Para esses críticos, essa visão demonstra um neodualismo, uma visão atualizada do cartesianismo que separa o corpo da mente, atribuindo a importância apenas a esta última no processo de aquisição de conhecimento. Pela visão da cognição situada, mente e corpo são integrados e indissociáveis e o sistema sensorio-motor ganha um papel fundamental, demonstrando que o desempenho cerebral é apenas uma parte do processo e não mais "a" parte exclusiva e fundamental da cognição humana. “Percepção e ação, sensorium e motorium, estão ligados entre si como sucessivamente emergentes e mutuamente selecionando padrões” [Varela et al. 1991, 163] Por essa perspectiva, o sujeito e sua cognição estão implicados de forma mútua. O sujeito está engajado no seu mundo cultural, social e físico de forma dinâmica e em constante exploração, pois não há nada dado, pronto ou fixo. Este é um processo ininterrupto em que o fluxo de informação afeta o corpo (no sentido corporeamente), o mundo e o próprio processo de engajamento. É inerente à cognição situada envolver percepção e ação e essa é a condição para o humano conhecer e apreender o meio ao qual pertence, seus elementos e a si próprio.

Our experience of the world is not separate from our conceptualization of the world. Indeed, in many cases (by no means all!), the same hidden mechanisms that characterize our unconscious system of concepts also play a central role in creating our experience. This does not mean that all experience is conceptual (far from it!); nor does it mean that all concepts are created by hidden mechanisms that shape experience. However, there is an extensive and important overlap between those mechanisms that shape our concepts and those that shape our experience. [Lakoff and Johnson 1999, 509]

Minha argumentação principal é de que a tecnologia não está como adorno ou elemento indiferente ao corpo e à estética da obra; ao contrário, acredito que o corpo, a dança e seu contexto digital são completamente interdependentes, uma vez que consideramos os pressupostos da cognição situada. Do ponto de vista da dança, para uma construção e desenvolvimento efetivos desse processo, é preciso que o bailarino esteja implicado com o sistema; isto significa que os dispositivos e o ambiente estejam "corporificados" no sujeito, que este sujeito esteja “enraizado”, engajado, nesse contexto e, conseqüentemente, que esses elementos estejam como "artefatos cognitivos" que contribuam na formação do bailarino, "expandindo" o processo corpo-mente. Na língua inglesa dos textos fundamentais desta pesquisa, de autores como Lakoff e Johnson (1999), Varela, Thompson e Rosch (1991), Andy Clark (1997, 2003), Alva Noë (2004, 2012), apenas para citar alguns, esses conceitos são respectivamente: *embodied*, *embeddedness* e *extended cognition*¹⁵.

Portanto, fica claro aqui que o sistema sensorio-motor de um bailarino imerso em um sistema sensível, construído pelas novas tecnologias, trará implicações na forma como esse sujeito adquire o conhecimento proposto nesse contexto. Ambientes imersivos,

¹⁵ Respectivamente com o entendimento *daquilo que é da incorporado, *enraizado no e com o ambiente, *termo cunhado por Varela do espanhol “en acción” que coloca a ação do indivíduo como protagonista na percepção e cognição, e a *cognição estendida através dos dispositivos e do meio ambiente.

interativos e telemáticos não devem ser considerados apenas espaços tecnológicos onde os sujeitos podem entrar e sair sem nenhuma implicação (considerando aqui processos que atuam pela mediação tecnológica, como já explicado neste artigo). Se o interesse é que o bailarino vivencie o máximo possível o sistema sensível, descobrindo como perceber o meio e o outro, encontrando formas de como se relacionar com o contexto, então, sua percepção e ação estarão implicadas na aquisição de conhecimento daquele ambiente.

Conforme já mencionado, a presença é um dos conhecimentos postos em questão em sistemas de telemática, uma vez que, em sua essência, carrega a proposição de corpos distantes que entram em sinergia; sendo assim, é preciso compreender tais condições de contato e conectividade, já não mais balizados pelo estado físico de estar aqui-e-agora em carne-e-osso. Na telemática, o bailarino vivencia outras experiências perceptivas que refletem na formação dos seus sistemas conceituais. Nesses espetáculos, o bailarino tem consciência de que a câmera é um dos pontos de interligação entre ele e seu parceiro remoto e que esse fluxo de informação é transferido via rede, cujo *delay* estará sempre presente. Essas informações não são apenas dados teóricos, mas são sensações e vivências conhecidas no corpo. Segundo o filósofo Alva Noë, trata-se de um “[...] conhecimento conceitual, mas também uma forma mais prática de conhecimento que inclui o que chama de conhecimento sensório-motor” [Noë 2012, 15].

Portanto, o processo mediado pela tecnologia digital propicia outros conhecimentos sensório-motores distintos de um espetáculo cênico não telemático. Minha afirmação sobre a necessidade de um processo continuado para o aprendizado e criação da dança telemática está relacionada com as proposições desses autores, com a compreensão que existe um conhecimento sensório-motor que precisa ser estabelecido para a efetiva realização desse campo. Estar continuamente em condição telemática faz com que o corpo exercite sua forma de perceber e agir naquele contexto e em relação ao outro; ignições motoras inéditas podem surgir para dar conta das demandas encontradas; compreensões distintas ocorrem sobre a noção de espaço-tempo, de distância e proximidade e da própria ideia de contato. Sendo assim, é preciso que o processo seja corporificado, isto é, que aqueles dispositivos estejam conhecidos pelo corpo de forma sensório-motora; que o sistema corpo-tecnologia esteja enraizado, ou seja, que a troca mútua de informação entre ambos altere e fortaleça os vínculos entre os dois e que o estar remotamente seja uma condição compreendida, formando então seu sistema conceitual. “Um conceito corporificado é uma estrutura neutra que realmente faz parte, ou faz uso do sistema sensório-motor do nosso cérebro. Muito da inferência conceitual é, portanto, uma inferência sensório-motora” [Lakoff and Johnson 1999, 20].

Interessa saber como o próprio bailarino lida com uma dança em que o seu parceiro é percebido em uma dessas condições (de presença), com as quais ele estará em relacionamento contínuo durante uma improvisação telemática. À luz das reflexões de Alva Noë, essas condições são consideradas variedades de presença, afirmando que nossa experiência perceptiva assume verdadeiramente distintas formas de estar presente, mesmo que frágeis, uma vez que existem várias gradações como qualidades e modalidades de presença. Pensar e perceber essas variedades de presença é uma questão que ele nomeia de estilo e o contato com esse outro é uma questão de disponibilidade de ambos. Nas palavras do filósofo:

The proposal, then, is this: perceptual consciousness is a special style of access to the world. But access is not something bare, brute or found. The

ground of access is our possession of knowledge, understanding, and skills. Without understanding, there is no access and so no perception. My emphasis here is on a special kind of understanding that distinctively underwrites our perceptual access to objects and properties, namely, sensorimotor understanding. We can see what there is when it is there, and what makes it the case that it is there is the fact that we comprehend its sensorimotor significance. Sensorimotor understanding brings the world into focus for perceptual consciousness. [Noë 2012, 20]

A percepção da presença não é uma questão de estar próximo, em carne-e-osso, mas uma questão de *disponibilidade* que se instaura por haver um *entendimento sensório-motor* [Noë 2012] por parte de todos os envolvidos. Segundo Noë, a experiência perceptiva da presença oscilará em sua apreensão, de acordo com duas condições: a) dependente do movimento, ou seja, “[...] a relação com o objeto é claramente controlada pelo movimento do corpo”; b) dependente do objeto, ou seja, “[...] o movimento do objeto claramente controla as características da relação objeto/percebedor” [Noë 2012, 22]. Isso quer dizer que quanto maior for o grau de existência dessas duas condições, maior será a força dessa presença. A presença perceptiva será mais efetiva quanto mais concreta forem as duas condições: como o sujeito alcança o objeto e como esse se manifesta, ou seja, um acordo entre o potencial do objeto em ser percebido e da ação do percebedor.

A relação entre duas pessoas fisicamente próximas pode até ser mais intensa e ter as condições “a” e “b” de forma mais concreta. Contudo, sujeitos corporificados e engajados em processos de telemática encontram novas ignições para a condição “a”, assim como o desenvolvimento de processos artísticos e tecnologias específicas para a arte em rede permitem uma maior possibilidade de ocorrência da condição “b”. O desenvolvimento da dança telemática tem contribuído justamente para o fortalecimento dessas duas condições.

3.5. Visão de Futuro

Após uma década ou mais de pesquisa e produção no campo da telemática, testando configurações sempre distintas e inéditas, buscando compreender de forma aprofundada cada tecnologia em curso, podemos apontar algumas dificuldades e indicar alguns caminhos:

- A grande maioria dos trabalhos realizados contou com a parceria das redes acadêmicas nacionais e internacionais, entretanto, muitas vezes algumas instituições apresentaram dificuldades para a compreensão de projetos artísticos, dificultando o uso da rede e, conseqüentemente, o processo criativo. Conseguir redes abertas e específicas para um fluxo continuado no campo das artes seria uma condição ideal para o desenvolvimento da telemática. Se as argumentações deste artigo forem procedentes, apenas quando tivermos ciberespaços específicos para utilizar (assim como quem utiliza uma sala de ensaio, um estúdio de gravação ou um ateliê) é que poderemos ter um crescimento exponencial nesse campo. Tenho realizado diversos cursos e orientado vários alunos interessados na arte em rede, mas eles ainda ficam na dependência de um projeto institucional para poder realizar seus experimentos.
- As ferramentas para transmissão de audiovisual em alta resolução ainda dependem de um profissional especializado, ao menos as indicadas nesse contexto. Muitas possuem uma operação simplificada, o que facilita a utilização de um aluno ou de um alguém sem conhecimento em rede ou computação, mas qualquer problema específico com a

configuração, interfaces e conectividade, dependerá da interferência de um profissional. Assim como temos aplicativos populares para transmissão de vídeo, da mesma forma deveria ser com as ferramentas digitais que utilizamos em projetos dessa natureza.

- A transmissão a partir de dispositivos cabeados ainda é uma premissa para projetos em telemática, condicionando assim os locais de sua realização. Muitos equipamentos culturais e até mesmo teatros, estúdios e auditórios acadêmicos, não contam com ponto de rede ou estão em um local de conexão reduzida. O ideal seria conseguir transmitir por sistemas não cabeados, através de hot-spots, que poderiam ser instalados com um simples modem, sem com isso, é claro, perder a qualidade e a velocidade de transmissão. Tal proposta daria condição para os vários projetos de intervenção ou ocupação urbana que a cada dia ganham mais força.

A visão de futuro para a utilização de sistemas de videoconferência para a produção de processos artísticos, sejam produções de dança telemática, como um produto, seja como uma forma de cooperação para aprendizado e troca de informação no campo das Artes Cênicas, dependeria da resolução dos três pontos indicados acima: rede específica para o livre trânsito de projetos artísticos, ferramentas de fácil uso para não especialistas e conexões que não dependam de cabeamentos. Talvez não sejam um problema exclusivamente tecnológico, mas também operacional e financeiro. Entretanto, não temos como pensar nos dispositivos sem pensar no sistema como um todo.

A videocolaboração tem um grande potencial nas artes em geral. No campo da dança não é diferente. Talvez essa linguagem artística traga apenas maior complexidade, uma vez que depende do próprio corpo e não de algo que o corpo produz, como as demais linguagens convencionalmente podem fazer. Mesmo o ator pode interpretar uma rádio-novela e sua voz terá maior relevância do que o corpo como um todo.

Tive a oportunidade de assistir algumas aulas normalmente realizadas entre a New York University, nos Estados Unidos, e sua filial em Abu Dhabi. Nessas aulas de música realizadas por sistemas telemáticos, as salas eram bem equipadas na parte acústica e contavam com um atraso de transmissão quase inexistente. Para acompanhamento, há uma câmera e uma tela com a projeção do parceiro. No contexto da música, a ferramenta LoLa tem sido muito utilizada por conseguir uma maior velocidade de transmissão, com grande qualidade e baixíssima taxa de latência. Com isso, conseguem uma boa condição para enviar e receber a sonoridade entre os pontos de presença.

No caso da dança, teríamos que definir qual configuração de sistema atenderia as demandas do corpo em sua complexidade para conseguir uma condição eficiente e satisfatória de envio e recepção. Para essa linguagem artística, interessa uma visão tridimensional do corpo, seu peso e sua energia, apenas para citar alguns aspectos. Seria, então, possível um sistema de videocolaboração no futuro que pudesse enviar não uma imagem videográfica apenas transmitida em uma tela plana, mas que tivesse condições de enviar também seu avatar, com informação tridimensional desse corpo e sem latência? O peso e a energia de cada parte do corpo poderia ser sentida pelo parceiro remoto ou demonstrado de alguma forma para todo o sistema, em seus vários ambientes remotos? Essas configurações aumentariam a disponibilidade, no sentido atribuído por Alva Noë (2012), para que as condições “a” (dependente do movimento) e “b” (dependente do objeto) alcançassem um grau maior de efetivação.

A ideia de videocolaboração como desenvolvimento e avanço conceitual e tecnológico dos antigos sistemas de videoconferência deve levar em conta “um ambiente” e não uma tela, assim como buscar uma maior articulação entre a realidade física com a virtual. A dança telemática, por sua complexidade, deve ser considerada não apenas pelo seu viés artístico, mas também por sua própria natureza, pelo seu foco no corpo em si mesmo. Resultados nesse sentido poderão auxiliar as diversas áreas que desejam colocar pessoas em interação via rede, de forma mais humana, com um maior grau da variedade de presença, nesse contexto.

3.6. Conclusão

A videocolaboração pode auxiliar o campo da dança por diversos fatores. De forma mais óbvia, podemos citar a possibilidade de incremento no campo de ensino e de aprendizagem entre grupos, aspecto bem-vindo em um país de dimensão continental e que não conta com cursos superiores (por exemplo), em todos os locais. Esse contexto tem sido contemplado por programas de ensino à distância (EAD), entretanto, pelo que tenho de conhecimento, o mesmo tipo de ferramenta é utilizado no EAD para toda e qualquer área, independente das especificidades em jogo. Além disso, por diminuir custos, a possibilidade de troca de informação e transferência de informação via rede pode ajudar no intercâmbio entre outros países, entre centros culturais, principalmente por considerarmos que as artes e a cultura, em geral, quase sempre carecem de recursos financeiros para seu desenvolvimento. Existe atualmente um grande interesse em fortalecer os vínculos entre países iberoamericanos ou, apenas, sulamericanos. Seja qual for a delimitação, o sistema de videoconferência poderá auxiliar em muitas propostas como esta. Como exemplo de projeto de integração latinoamericano, mas com coletivo interdisciplinar, podemos citar o grupo colaborativo formado durante o último Encontro Latinoamericano de eCiência, realizado durante a Sétima Conferência TICAL2017, na Costa Rica, de 3 à 5 de julho de 2017. Essa experiência como integrante desse coletivo tem demonstrado também, como já mencionado, que não adianta apenas a existência de um sistema tecnológico (neste caso, a plataforma *Colaboratorio*¹⁶), pois é preciso saber qual o objetivo e a proposta de cooperação. Nesse sentido, uma das primeiras metas do grupo foi a organização de um colóquio para abordar esse tema e, assim, poder evoluir com outras atividades. Paralelamente, e por um subgrupo de pesquisadores das artes, estamos propondo uma disciplina no curso de Pós-Graduação em que lecionamos, estabelecendo um vínculo interinstitucional com participação do Brasil, Chile e Colômbia, com o objetivo de desenvolver o estudo sobre arte em rede, que contará com pesquisadores e alunos das áreas de teatro, dança e música. Essa proposta vai de encontro a outra possibilidade de utilização da videocolaboração: a pesquisa e o desenvolvimento de projetos artísticos.

As duas propostas acima – para o ensino-aprendizagem e para processos artísticos – poderão ser muito beneficiadas pelo sistema de videocolaboração, assim como este ganhará novas ideias para seu desenvolvimento, desde que esteja atento a demandas específicas dos campos em questão. Se há uma estreita relação entre arte/tecnologia, sujeito/dispositivos, é preciso pensar em sistemas mais flexíveis nos aspectos gerais e mais engajados nas particularidades de cada contexto. O processo de enraizamento

¹⁶ <http://www.aireuropa-ontheair.com/>

(*embeddedness*), conforme explicado neste artigo, depende de projetos continuados para atingir uma efetivação satisfatória.

Como discutido e argumentado, a videocolaboração deve ser compreendida como outra forma de acesso ao outro, através de outra noção de presença que ocorre, de acordo com as condições “a”, dependentes do movimento, ou “b”, do objeto. Portanto, o desenvolvimento da videocolaboração deve escapar da cópia fiel de um mundo físico e encontrar no próprio contexto telemático a forma de funcionamento perceptivo do indivíduo, ou seja, de como ele atua nesse ambiente. É preciso compreender que ele é um corpo expandido por artefatos tecnológicos que alteram seu sistema conceitual. Justamente por essa razão, argumentada ao longo deste artigo, é que não se pode mais pensar nos dispositivos e nos sistemas tecnológicos construídos, desenvolvidos, programados “para” o sujeito e, sim, deve ser pensado “com” o sujeito.

Compreender a relação corporificada, engajada e expandida da dança com mediação tecnológica, de acordo com os conceitos da cognição situada discutidos nesse artigo, contribui para o fortalecimento de projetos interdisciplinares que efetivamente trabalhem em conjunto, sabendo dessa construção mútua de conhecimento. Perceber a importância dos estudos sobre a cognição e percepção humana para o desenvolvimento da dança telemática promove outro caminho para o processo colaborativo entre os campos artístico e tecnológico que não é mais de usuário e desenvolvedor, pois esses estão em um fluxo contínuo e mútuo de transformação.

Referências

- CHATZICHRISTODOULOU, M. (2012) “Cyberformance? Digital or Networked Performance? Cybertheaters? Or Virtual Theatres? ...or all of the above?” In: CYPOSIUM - CYBERPERFORMANCE SYMPOSIUM. <www.cyposium.net/wpcontent/uploads/2012/09/maria_text.pdf>, Agosto.
- CLARK, A., Being there: putting brain, body, and world together again. Cambridge, London: Bradford Book, MIT Press, 1997.
- _____. Natural born-cyborg. Oxford: Oxford University Press, 2003.
- Dery, M., Escape velocity: Cyberculture at the end of the century. New York: Grove Press, 1996.
- DRIOLI, C. and BUSO, C. A. (2013) “Networked performances and natural interaction via LOLA: Low latency high quality A/V streaming system.” In: Information Technologies for Performing Arts, Media Access, and Entertainment. Springer Berlin Heidelberg, p. 240-250. <https://www.researchgate.net/publication/259221698_Networked_Performances_and_Natural_Interaction_via_LOLA_Low_Latency_High_Quality_AV_Streaming_System>. Agosto.
- GALLOWAY, K. and RABINOWITZ, S. (1994) “Ecafe Manifesto. The challenge: we must create at the same scale as we can destroy.” In: Telematic Connections. You are in space and time. <http://telematic.walkerart.org/timeline/timeline_ecafe.html>, Setembro.

- Hansen, M., *New philosophy for new media*. Cambridge: The MIT Press, 2004.
- Gharai, L. et al. (2006) “Experiences with high definition interactive video conferencing.” In: *Multimedia and Expo, 2006 IEEE International Conference on*. IEEE.
<http://eprints.gla.ac.uk/3585/1/High_Definition_Interactive_Video_Conferencing.pdf>, Setembro.
- Lacerda, P. (2015) “TeleCorpo: mesa de corte de vídeo para redes de computadores”. In: *Revista Eletrônica MAPA D2 - Mapa e Programa de Artes em Dança (e Performance) Digital*, 2(2): 272–84. <<https://portalseer.ufba.br/index.php/mapad2/article/view/14996/10260>>, Agosto.
- Lakoff, G. and M. Johnson, *Philosophy in the flesh: the embodied mind and its challenge to western thought*. New York: Basic Books, 1999.
- Naugle, L. "A Study of Collaborative Choreography Using LifeForms and Internet Communication." Ph.D. Dissertation. New York University, Department of Music and Performing Arts Professions, The Steinhardt School of Education, 2002.
- Noë, A., *Varieties of Presence*, Cambridge, Massachusetts/ London, England: Harvard University Press, 2012.
- Noë, A., *Action in Perception*, Cambridge, MA, and London: MIT Press, 2004.
- Santana, I. (2013) “Novas configurações da dança em processos distribuídos das redes. 2013”. In: *Plataforma Eletrônica Internacional Xanela Comunidad Tecno Escenica*. <<http://www.xanela-rede.net>>, Setembro.
- Varela, F.J., Thompson, E. and Rosch, E., *A mente incorporada: ciências cognitivas e experiência humana*. Porto Alegre: Editora Artmed, 1991.
- Vieira, E. et al. (2012) “Uma Ferramenta para gerenciamento e transmissão de fluxos de vídeo em alta definição para telemedicina”. In: *Salão de Ferramentas do Simpósio Brasileiro de Redes de Computadores 2012, Ouro Preto. Anais 2012, v. 1*. <http://gtavcs.lavid.ufpb.br/wp-content/uploads/2011/11/ARTHRON_2-0_salao-ferramentas-SBRC-2012_final.pdf>, Setembro.

ANEXO - Biografia da autora

Ivani Santana: pesquisa dança mediada pelas novas tecnologias desde a década de 1990. A partir de 2001, nos Estados Unidos, e de 2005 no Brasil, desenvolve projetos interdisciplinares para criação de ferramentas, metodologias e produtos artísticos no campo da telemática. Coordenadora do Grupo de Pesquisa Poéticas Tecnológicas: Corpoaudiovisual, através do qual tem difundido estudos e aplicações no campo da arte em rede, propiciando uma multiplicação do saber e, com isso, impulsionando o desenvolvimento de dissertações sobre essa temática, a capacitação de novos profissionais para a área, dentre outros. Conta com mais de 10 projetos de Arte em Rede e com diversas parcerias de âmbito mundial. É Vice-coordenadora do Programa de Pós-graduação em Artes Cênicas, Professora do Instituto de Humanidades, Artes e Ciências Prof. Milton Santos, Bolsista Produtividade em Pesquisa PQ Nível 1 D. Realizou mestrado e doutorado no Programa de Pós-graduação em Comunicação e Semiótica (PUC-SP) e pós-doutorado no Sonic Arts Research Center, Queen's University Belfast, Irlanda do Norte, Reino Unido, com a pesquisa *Dramaturgias do corpo (tele)sonoro* (2012/2013).

Capítulo

4

Tecnologias emergentes em telemedicina - ênfase em videocolaboração

Valter Roesler¹, Erno Harzheim²

¹ Universidade Federal do Rio Grande do Sul (UFRGS). Instituto de Informática

² Universidade Federal do Rio Grande do Sul (UFRGS). Medicina Social

roesler@inf.ufrgs.br, erno.harz@terra.com.br

Resumo

Este capítulo apresenta uma visão sobre tecnologias emergentes em telemedicina no País com ênfase em videocolaboração. Inicialmente, será efetuado uma introdução ao tema e, em seguida, será apresentado o panorama tecnológico da telemedicina no País. Depois será apresentado um estudo coordenado pela ABDI (Associação Brasileira de Desenvolvimento Industrial) que, através de vários encontros com especialistas na área, elegeu 34 tecnologias emergentes para telemedicina, dividindo-as em “Viável (ou não) comercialmente”, “Atrativa (ou não) em 5 ou 15 anos”, “Relevante (ou não) crítica ou prioritária”. Haverá um detalhamento maior para as tecnologias relacionadas à videocolaboração, que é o foco deste texto. Em seguida, serão apresentados projetos atuais que os autores estão envolvidos detalhando o desenvolvimento e uso de tecnologias que buscam atender as demandas relevantes, prioritárias e atrativas eleitas no estudo da ABDI. Finalmente, será apresentada uma visão de futuro dos autores em relação à telemedicina e videocolaboração.

Abstract

This chapter presents an overview of emerging technologies in telemedicine with an emphasis on videoconferencing. Initially, an introduction will be made to the topic and then the technological panorama of telemedicine in Brazil will be presented. Then, it will be presented a study coordinated by ABDI (Brazilian Association of Industrial Development). This study, through several meetings with specialists in the area, Chose 34 emerging technologies for telemedicine, dividing them into "Viable (or not) commercially", "Attractive (or not) in 5 or 15 years," "Relevant (or not) critical or priority" There will be an emphasis for technologies related to videocolaboration, which is the focus of this text. Next, we will present current projects that the authors are involved, with more details in the development and use of technologies that seek to meet

the relevant, priority and attractive demands chosen in the ABDI study. Finally, we will present a future vision of the authors regarding telemedicine and videocollaboration.

4.1 Introdução

O Brasil, como muitos outros países do mundo, apresenta um contexto epidemiológico complexo, caracterizado pela manutenção de agravos ligados ao seu histórico e lento desenvolvimento socioeconômico – agravos materno-infantis e problemas de saúde relacionados à desnutrição e a doenças infectoparasitárias – somados à epidemia de doenças crônicas e ao envelhecimento da população, além dos altos índices de agravos relacionados a acidentes e à violência. A expectativa de vida hoje está em torno dos 73 anos, mas será de 82 anos em 2030, quando o número de idosos terá ultrapassado o número de jovens (BRASIL, 2011b).

Associadas ao envelhecimento populacional, as doenças crônicas não transmissíveis correspondem a 72% das causas de morte no país, com maior impacto sobre a população mais vulnerável socioeconomicamente (BRASIL, 2011a). Além do impacto pessoal e populacional, trazem consequências negativas sobre o orçamento da saúde, pois muitas vezes o tratamento é bastante especializado, necessitando deslocamento para grandes centros urbanos onde normalmente estão os maiores especialistas, ampliando muito os gastos em saúde (BODENHEIMER, 2008; DAVIES et al., 2006; HOFMARCHER, 2007; SCHOEN et al., 2011). A concentração do conhecimento, muitas vezes centralizado em grandes centros urbanos, acarreta também impacto no orçamento individual/familiar, pois além do afastamento das atividades produtivas, implica em longos deslocamentos das pessoas, tanto no sistema público, como no sistema privado e na saúde suplementar.

Uma forma de otimizar a relação custo-efetividade dos tratamentos em saúde é através da disseminação do uso de e-Saúde, através das Tecnologias de Informação e Comunicação (TICs), que permitem a interação à distância, sendo para efetuar diagnóstico remoto, segunda opinião médica, entre outros. Essa disseminação poderia inclusive evitar a necessidade de realização de um percentual de procedimentos médicos de alta complexidade, por meio da resolução de problemas de saúde antes de sua evolução para estágios de maior gravidade. Além disso, doenças pouco frequentes, incluindo as condições consideradas como raras, podem se beneficiar de estratégias de e-Saúde, dada a distribuição ainda mais concentrada de profissionais especializados no diagnóstico e manejo dessas condições.

No contexto da organização de sistemas de saúde mantendo sustentabilidade econômico-financeira dos mesmos, três aspectos são centrais nessa discussão: **acesso (escala), qualidade e custo** das ações em saúde. Se quiser acesso e qualidade, aumenta o custo. Se quiser qualidade e baixo custo, limita o acesso. Se quiser acesso e baixo custo, diminui a qualidade.

É justamente nesta interação entre acesso, qualidade e custo que tecnologias de Telemedicina podem representar um grande avanço na prestação de cuidados em saúde. Conforme o *Institute of Medicine*, Telemedicina é “o uso de tecnologias eletrônicas de informação e comunicação para prover e/ou apoiar cuidado em saúde quando a distância separa os participantes” (SCHWAMM, 2014). Ao ultrapassar barreiras de

acesso físico, ofertando intervenções efetivas e reguladas por mecanismos promotores de equidade e que previnam uso indevido de intervenções médicas (prevenção quaternária¹) aliadas a um custo adequado, tecnologias de Telemedicina têm papel estratégico na consolidação de Redes de Atenção à Saúde e melhoria da saúde da população (BASHSHUR et al., 2014; NORMAN, 2009; GERVÁS, 2006).

Um segmento fundamental das tecnologias para telemedicina é de **videocolaboração**, permitindo a interação através de vídeo. Essa classe de tecnologias traz uma especialização à simples troca de texto ou imagens, abrindo um conjunto de aplicações utilizadas em diversos contextos, como a educação a distância, visualização remota, consultas remotas, diagnósticos remotos, entre outras.

A palavra-chave em Telemedicina é interação. Interação entre profissionais de saúde, entre profissionais de saúde e pacientes, entre gestores e profissionais de saúde, entre gestores e pacientes, ou entre diferentes gestores.

Entre 2014 e 2016 foi executada a Agenda Técnica Setorial (ATS) no âmbito do Plano Brasil Maior. No setor da saúde, um dos tópicos avaliados foi a Telemedicina. O objetivo foi identificar tecnologias relevantes para a competitividade setorial no horizonte de 5 e 15 anos. Os resultados com foco em videocolaboração serão apresentados neste capítulo de livro.

Este capítulo é dividido da seguinte forma: a seção 2 apresenta algumas definições básicas do setor de telemedicina. A seção 3 apresenta um resumo do panorama tecnológico brasileiro em telemedicina. A seção 4 detalha algumas das tecnologias emergentes com foco em videocolaboração. A seção 5 apresenta alguns projetos em videocolaboração sendo desenvolvidos pelos autores deste texto. A seção 6 apresenta uma visão de futuro na área, e a seção 7 as considerações finais do trabalho.

4.2 Definições básicas de telemedicina

Telemedicina pode ser definida como a prestação de serviços de saúde por profissionais de saúde, onde a distância é um fator crítico, usando Tecnologias de Informação e Comunicação (TICs) para o intercâmbio de informações válidas para o diagnóstico, tratamento e prevenção das doenças e lesões, pesquisa e avaliação, e para a educação continuada dos profissionais de saúde; tudo no interesse de promover a saúde dos indivíduos e suas comunidades (WHO, 1998). Ao construir essa definição, o grupo de trabalho da Organização Mundial da Saúde (OMS) vislumbrava que as TICs proporcionariam uma oportunidade sem precedentes para melhorar a qualidade e equidade do acesso aos cuidados de saúde (WHO, 1998).

A palavra-chave em Telemedicina é *interação*, com foco no cuidado em saúde, mediada por TICs. A amplitude da Telemedicina nos remete a diferentes contextos de interação humana em saúde mediada por TICs. Atividades de teleeducação em saúde prescindem, muitas vezes, de atividades sincrônicas, o que gera menor necessidade de estruturas robustas de conectividade entre os participantes. Por outro lado, ações de

¹ Prevenção quaternária refere-se ao conjunto de ações médicas ou de profissionais de saúde que atenuam ou evitam as consequências do intervencionismo médico excessivo, evitando a realização de ações médicas desnecessárias à manutenção da saúde e/ou do bem-estar do paciente.

telemonitoramento em um ambiente crítico, como um Centro de Terapia Intensiva (CTI), exigem uma segurança e uma estabilidade da estrutura de conectividade muito diferentes do exemplo anterior. Dessa forma, as diversas ações possíveis de serem realizadas em Telemedicina exigem diferentes requisitos tecnológicos, com diferentes cadeias produtivas.

Telessaúde é o uso das modernas tecnologias da informação e telecomunicações para atividades a distância relacionadas à saúde em seus diversos níveis (primário, secundário e terciário). Possibilita a interação entre profissionais de saúde ou entre estes e seus pacientes, ou entre gestores e qualquer dos dois grupos antes citados, bem como o acesso remoto a recursos de apoio diagnósticos ou até mesmo terapêuticos (por exemplo, através da robótica).

A fim de evitar confusões terminológicas, neste documento, utilizaremos o termo Telemedicina como um termo amplo, que engloba a prestação de serviços de saúde ou atividades educacionais relacionadas a tais serviços, por meio das tecnologias de informação e comunicação, incluindo sob seu guarda-chuva referências a termos correlatos como Telessaúde, e-Saúde (e-Health), m-Health (*mobile health*) ou Saúde Conectada (*connected health*).

As principais ações de Telemedicina podem ser classificadas como teleassistência ou teleeducação e subdivididas em:

- **Teleconsulta:** é a realização de consulta médica (ou de outro profissional de saúde) a distância por meio de tecnologia de informação e comunicação, isto é, interação a distância entre profissional de saúde e paciente¹.
- **Teleconsultoria:** consulta registrada e realizada entre profissionais e gestores da área de saúde, por meio de instrumentos de telecomunicação bidirecional, com o fim de esclarecer dúvidas sobre procedimentos clínicos, ações de saúde e questões relativas ao processo de trabalho, podendo ser de dois tipos: síncrona (realizada em tempo real, geralmente por web, videoconferência ou telefone) ou assíncrona (realizada por meio de mensagens off-line) (BRASIL, 2011c).
- **Telediagnóstico:** serviço que utiliza as Tecnologias de Informação e Comunicação para realizar Apoio ao Diagnóstico através de distâncias geográficas e/ou temporais, incluindo diversos subtipos como Telerradiologia (raio-X, ultrassonografia, ressonância magnética), Telecardiologia (ou TeleECG), Tele-espirometria, Telepatologia, Teledermatologia (normalmente fotos digitalizadas), Teleoftalmologia, entre outras.
- **Telecirurgia:** realização, a distância, de cirurgia (ou procedimento) por um cirurgião, por meio de um sistema robótico; conhecida também como cirurgia remota (*remote surgery*).

¹ A teleconsulta ainda não é permitida no Brasil pelo Art. 37 do Código de Ética Médica, salvo em situações de emergência, o que é regulado pela Resolução 1.643/02 do Conselho Federal de Medicina. Essa questão vem trazendo bastante polêmica no País, visto que o avanço das tecnologias permite claramente uma efetiva interação médico-paciente, como já está regulado em outros países.

- **Telemonitoramento:** monitoramento a distância de parâmetros de saúde e/ou doença de pacientes por meio de TICs, incluindo a coleta de dados clínicos do paciente, sua transmissão, processamento e manejo por um profissional de saúde por meio de um sistema eletrônico (BASHSHUR et al., 2014). Pode incluir dispositivos e sistemas de monitoramento relacionados a cuidado domiciliar (homecare), cuidado intra-hospitalar ou até mesmo vigilância epidemiológica. Sua utilização tem sido associada à redução de consultas em emergências, redução de hospitalizações e redução da duração da internação, sendo uma alternativa para suporte ao cuidado de pacientes portadores de doenças crônicas (PARÉ, 2007).
- **Teleducação:** conferências, aulas, cursos, ou disponibilização de objetos de aprendizagem interativos sobre temas relacionados à saúde ministrados a distância por meio de TICs (BRASIL, 2011c). Este segmento inclui tecnologias de comunicação síncronas, como videoconferências, bem como tecnologias específicas para fins educacionais, como ambientes virtuais de aprendizagem e objetos de aprendizagem virtuais (atlas virtuais 3D por exemplo). Além disso, o uso de manequins robóticos pode aprimorar muito o aprendizado.
- **Segunda Opinião Formativa (SOF):** resposta sistematizada, construída com base em revisão bibliográfica das melhores evidências científicas e clínicas e no papel ordenador da atenção básica à saúde, a perguntas originadas das teleconsultorias, e selecionadas a partir de critérios de relevância e pertinência em relação às diretrizes dos serviços de saúde (BRASIL, 2011c).

4.3 Panorama resumido de telemedicina no Brasil

O Brasil conta com um sistema público de saúde de base universal, isto é, um conjunto de serviços de saúde que deveriam ofertar ações em saúde condizentes com as necessidades epidemiológicas da sua população, sem restrição de acesso aos cidadãos, custeado por meio de impostos. Este sistema, o Sistema Único de Saúde (SUS) é organizado como uma rede de serviços de saúde, contando com serviços de Atenção Primária à Saúde (APS) e serviços de urgência como suas portas de entrada.

A porta de entrada refere-se ao atendimento inicial. Em relação à atenção básica ou atenção primária, o objetivo é orientar sobre a prevenção de doenças, solucionar os possíveis casos de agravos e direcionar os mais graves para níveis de atendimento superiores em complexidade. A atenção básica funciona, portanto, como um filtro capaz de organizar o fluxo dos serviços nas redes de saúde, dos mais simples aos mais complexos. No Brasil, há diversos programas governamentais relacionados à atenção básica, sendo um deles a Estratégia de Saúde da Família (ESF), que leva serviços multidisciplinares às comunidades por meio das Unidades Básicas de Saúde (UBSs), por exemplo. Consultas, exames, vacinas, radiografias e outros procedimentos são disponibilizados aos usuários nas UBSs. A atenção básica também envolve outras iniciativas, como: as Equipes de Consultórios de Rua, que atendem pessoas em situação de rua; o Programa Melhor em Casa, de atendimento domiciliar; o Programa Brasil Sorridente, de saúde bucal; o Programa de Agentes Comunitários de Saúde (PACS), que busca alternativas para melhorar as condições de saúde de suas comunidades etc. (FIOCRUZ PENSESUS, 2017).

As ações de assistência podem ocorrer nos próprios serviços de porta de entrada ou em outros pontos da rede assistencial: hospitais de diferente incorporação tecnológica, serviços de apoio diagnóstico, serviços de cuidado especializado, etc.

Para essa rede funcionar com efetividade e eficiência é imprescindível que haja um bom sistema de comunicação entre os pontos. Uma das diretrizes do SUS é a integralidade, isto é, apresentar um grande leque de serviços capaz de atender a todas as necessidades em saúde da população. Num país de dimensão continental, fica inviável, do ponto de vista do custo e da qualidade (economia de escala), ter toda a oferta de serviços de saúde em todos os municípios do país. Neste contexto, a distância geográfica e a necessidade de comunicação entre os diversos pontos (serviços) assistenciais são grandes desafios a serem superados nos quais tecnologias em Telemedicina podem representar soluções de qualidade e custo-efetivas (SCHWAMM, 2014).

Em nível nacional, o quadro atual é de uma APS e uma rede de serviços ambulatoriais especializadas com resolutividade limitada, baixa incorporação tecnológica e praticamente ausência de mecanismos de coordenação assistencial. Isto gera uma demanda insustentável por atenção nos hospitais de grande porte. Fatos que comprovam esta análise são: a superlotação dos serviços de urgência e emergência, as listas de espera para consultas especializadas e procedimentos diagnósticos com volume crescente e inferior à oferta, grande proporção de usuários aguardando por anos para ter necessidades atendidas e, por outro lado, tanto consultas ambulatoriais, como internações em hospitais de grande porte destinadas a resolver problemas de saúde que poderiam ser resolvidos na APS ou em serviços ambulatoriais especializados.

Na área da saúde suplementar (atendimento privado em saúde, realizado ou não através de um convênio), a progressiva e positiva regulamentação exercida pela Agência Nacional de Saúde Suplementar (ANS) em defesa dos direitos dos beneficiários dos planos de saúde tem levado o setor a uma diminuição no número de planos registrados, mas com aumento da qualidade dos planos sobreviventes, a um custo elevado. Esse custo em elevação se deve não só ao aumento da regulamentação, mas ao contexto epidemiológico já citado – a epidemia de doenças crônicas, à cultura de *shopping around* dos pacientes (consumir diversas ações em saúde, mesmo quando não necessárias ou repetidas vezes além do razoável), ao exagerado intervencionismo médico e à prática de uma medicina defensiva, entre outras condutas médicas que transcendem a legalidade.

Entretanto, o maior fator da escalada de custos, na área da saúde suplementar, é a fragmentação de suas redes de serviços e saúde, semelhante ao que ocorre no SUS, e a fragilidade do sistema de comunicação desta “quase-rede” de serviços de saúde. O dilema que surge do aumento da regulamentação, em que as necessidades dos pacientes por consultas e procedimentos médicos especializados têm prazo curto para serem realizados, e a fragmentação e centralização em poucos centros urbanos de grande parte de médicos especialistas, hospitais e centros de diagnóstico e terapia, abre uma grande oportunidade para a Telemedicina. Porém, a oportunidade tem sua porta fechada pelo fato de não haver regulamentação acerca da forma de pagamento das ações em Telemedicina no escopo legal da ANS.

Neste contexto existem grandes oportunidades e iniciativas para a Telemedicina, Muitas não se consolidam, por entraves financeiros, falta de condições estruturais ou dificuldades de gestão dos próprios projetos, outras têm se consolidado e aumentado sua

capacidade instalada e servido de exemplo de inovação em Telemedicina, inclusive apresentando economia de escala.

Em 2011, o MS ampliou a base de financiamento para iniciativas de Telessaúde, incluindo a Secretaria de Atenção à Saúde como financiadora, além da Secretaria de Gestão do Trabalho e da Educação em Saúde (SGTES), e transformou o Projeto Piloto de Telessaúde no Programa Nacional Telessaúde Brasil Redes. Este está presente atualmente em todas as regiões do Brasil, com um ou mais núcleos de Telessaúde financiados pelo SUS e com foco na APS nos estados do Amazonas, Tocantins, Maranhão, Ceará, Paraíba, Pernambuco, Rio Grande do Norte, Goiás, Mato Grosso do Sul, Mato Grosso, Espírito Santo, Minas Gerais, Rio de Janeiro, São Paulo, Paraná, Santa Catarina e Rio Grande do Sul.

O Programa Telessaúde Brasil Redes objetiva não apenas fomentar as atividades de educação continuada, aproximando-as das equipes de APS localizadas em qualquer ponto do país, mas ofertar estratégias de apoio assistencial, teleconsultorias e telediagnóstico, que fortaleçam a integração entre os serviços de saúde, ampliando a resolutividade dos mesmos.

Os Núcleos de Telessaúde utilizam as mais diversas tecnologias para realizar suas ações, como Plataforma de Fluxo e Monitoramento de Teleconsultorias, Sistemas de Fluxo de Imagens com base DICOM, Plataforma de Fluxo e Monitoramento de Exames Diagnósticos, Imagens Virtuais Anatômicas Realístico Morfofuncional, entre outras. A maioria dessas tecnologias é de produção nacional, centrada nesses Núcleos Universitários de Telemedicina.

Outra iniciativa importante do governo brasileiro, capitaneada pela Rede Nacional de Ensino e Pesquisa (RNP), do Ministério de Ciência e Tecnologia, é a Rede Universitária de Telemedicina (RUTE), uma iniciativa que objetiva apoiar o aprimoramento de projetos em telemedicina já existentes e incentivar o surgimento de futuros trabalhos interinstitucionais. Dispõe de infraestrutura de alta capacidade do backbone nacional da RNP, a rede Ipê, e das Redes Comunitárias Metropolitanas de Educação e Pesquisa (Redecomep). Essa iniciativa complementa o esforço coordenado para prover uma infraestrutura fim-a-fim (nacional, metropolitana e institucional), adequada ao uso de aplicações avançadas de rede. Através do link da RNP com a Rede Clara (Cooperação Latino-Americana de Redes Avançadas), as instituições participantes contam com a colaboração de redes-parceiras na América Latina, Europa, Japão, Austrália e nos Estados Unidos (REDE UNIVERSITÁRIA DE TELEMEDICINA, 2017). A RNP financia, e a rede RUTE utiliza, o sistema MConf, um sistema de webconferência de código aberto cuja arquitetura permite escalabilidade para milhares de participantes simultaneamente, desenvolvido pelo Instituto de Informática da UFRGS (ROESLER, 2013).

No âmbito desses projetos em rede, pode-se destacar o projeto Minas Telecardio, que implantou serviço de ECG a distância em municípios de Minas Gerais. Médicos de hospitais universitários, incluindo docentes e especialistas em cardiologia, analisam os eletrocardiogramas enviados e realizam discussões online de casos clínicos. Da mesma forma, o programa Tele-ECG digital do Centro de Telessaúde do Instituto de Cardiologia do Rio Grande do Sul disponibiliza laudos de eletrocardiograma online, através de videoconsultorias realizadas por uma equipe de cardiologistas. Conta com 90

municípios em acompanhamento cardiológico virtual 24 horas, além de um programa de capacitação e educação continuada desenvolvida para aperfeiçoar a assistência em cardiologia (FUNDAÇÃO UNIVERSITÁRIA DE CARDIOLOGIA, 2012).

Na área da telerradiologia, o TeleRX, desenvolvido em parceria do Núcleo do Programa Telessaúde Brasil Redes no Rio de Janeiro com a Universidade Estadual do Rio de Janeiro (UERJ), possibilita que exames radiológicos de tórax realizados em serviços de APS sejam digitalizados e enviados para avaliação de radiologistas teleconsultores em Hospitais Universitários como os da UERJ, da UFF e da UFRJ (FERREIRA et al., 2009). Em Porto Alegre, uma iniciativa da Secretaria Municipal de Saúde ampliou o acesso a ultrassonografias obstétricas com a implantação de um serviço de telemedicina que conecta postos de saúde ao Hospital Materno-Infantil Presidente Vargas (HMIPV). Através de uma unidade móvel, as gestantes da região podem fazer, na unidade básica local, ultrassonografia obstétrica monitorada em tempo real por médicos especialistas do Hospital Materno-Infantil Presidente Vargas (ROESLER, 2012), como mostrado na Figura 4.1.

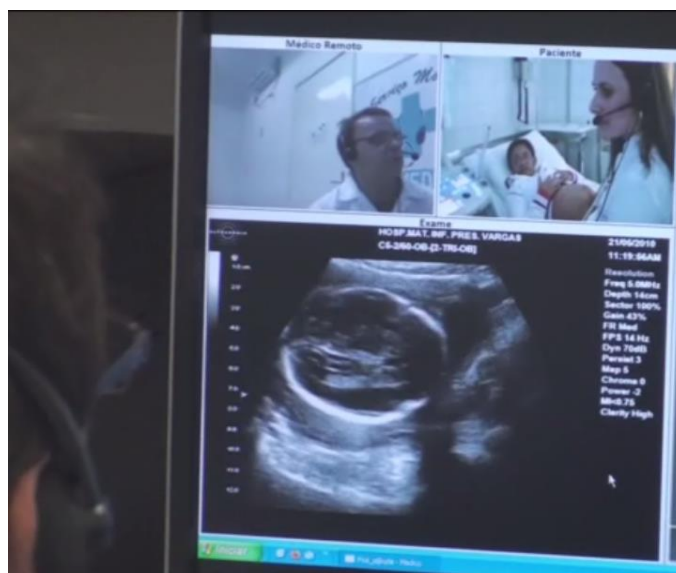


Figura 4.1. Interação médico-médico-paciente no sistema de Teleultrassonografia Obstétrica.

Santa Catarina desenvolve uma rede de transmissão de exames, entre eles, eletrocardiograma digital, tomografia computadorizada, ultrassom e ressonância magnética. Por meio deste projeto, exames obtidos em hospitais do interior do estado são laudados por médicos especialistas em centros colaboradores. Esta tecnologia foi desenvolvida pelo Grupo Cyclops, do Departamento de Informática e Estatística, do Centro Tecnológico da Universidade Federal de Santa Catarina, com apoio do governo do Estado, e atualmente está disponível em quase 90% dos municípios catarinenses (SAVARIS et al., 2011). A ênfase de desenvolvimento desse grupo de pesquisa é direcionada para tecnologias emergentes no campo dos sistemas de comunicação e arquivamento de imagens médicas – *Picture Archiving and Communication System* (PACS), em consonância com padrões internacionais, especialmente o padrão *Digital Imaging and Communications in Medicine* (DICOM) e suas extensões.

Outros projetos na área da tele dermatologia e telerradiologia foram realizados em áreas urbanas, indígenas e rurais de forma permanente ou temporária (CHAO et al., 2003; MESSINA et al., 2010). O Núcleo Goiás do Programa Telessaúde Brasil Redes vem desenvolvendo um projeto na área da tele oftalmologia com objetivo de rastreamento de retinopatia diabética. Avaliações iniciais demonstraram bons resultados (MAGRIÇO et al., 2011).

No Brasil, até o momento, o Conselho Federal de Medicina apresenta restrições ao uso da Telemedicina para a realização de consultas ou procedimentos cirúrgicos por meio de TICs fora do âmbito da pesquisa (CONSELHO FEDERAL DE MEDICINA, 2002). Entretanto, a realização de provas diagnósticas e monitoramento clínico de sinais vitais de pacientes a distância já é uma realidade nacional, com alguns centros muito desenvolvidos na área de telediagnóstico.

Os desafios no campo do telediagnóstico abrangem a garantia do envio dos dados, a interpretação do mesmo, bem como da confidencialidade que exige qualquer procedimento médico. Além disso, a realização indiscriminada de telediagnóstico em larga escala, sem indicação clínica baseada em evidências científicas sólidas, é um risco para a saúde dos pacientes por propiciar uma cascata de exames adicionais desnecessários, incluindo o potencial de iatrogenia, isto é, o dano à saúde causado por procedimento médico. Obviamente, os dois últimos aspectos ainda trazem um gasto financeiro adicional ao sistema de saúde, o que pode comprometer ainda mais sua sustentabilidade.

4.4 Tecnologias emergentes em telemedicina com ênfase em videocolaboração

Entre 2014 e 2016 foi executada a Agenda Técnica Setorial (ATS) dentro do Plano Brasil Maior. No setor de saúde, um dos tópicos avaliados foi *Telemedicina*. Essa ATS foi proposta pelas entidades ABDI (Associação Brasileira de Desenvolvimento Industrial)¹, CGEE (Centro de Gestão e Estudos Estratégicos)² e MS (Ministério da Saúde)³. A diretriz foi a mudança no padrão de concorrência da indústria nacional através do conhecimento e fortalecimento corporativo, com maior inserção nas áreas de tecnologia avançada, com fortalecimentos de políticas de atenção à saúde. As principais orientações foram:

- Promover a inovação e o desenvolvimento tecnológico
- Criar e fortalecer competências críticas da economia nacional
- Aumentar o adensamento produtivo e tecnológico das cadeias de valor
- Ampliar os mercados interno e externo das empresas brasileiras
- Ampliar os níveis de produtividade e competitividade da indústria brasileira

¹ www.abdi.com.br

² www.cgge.org.br

³ www.saude.gob.br

- Garantir crescimento socialmente inclusivo e ambientalmente sustentável.

O objetivo foi identificar as tecnologias relevantes para a competitividade setorial no horizonte de 5 e 15 anos.

A metodologia organizou o setor de telemedicina em equipes, sendo que cada uma delas tinha um especialista econômico, um especialista tecnológico e um comitê técnico de até 8 pessoas. Os dois autores deste capítulo participaram ativamente do estudo, sendo que o Erno Harzheim foi o especialista tecnológico, e Valter Roesler compôs o comitê técnico.

O Comitê Técnico foi composto por especialistas em Telemedicina e áreas afins, como informática em saúde, bioengenharia, redes computacionais, etc. Foram realizados quatro encontros presenciais, intercalados por trabalho individual ou em pequenos grupos, em que cada um dos especialistas sugeria novos itens e revisava a lista de tecnologias emergentes com objetivo de identificar e definir cada uma das tecnologias emergentes, enfatizando sua aplicação e objetivo, assim como o desafio tecnológico representado por ela.

O resultado do estudo foi uma lista de tecnologias (ou grupos de tecnologias), de produtos e/ou processos, novos produtos, novos usos de produtos existentes, novos processos produtivos, novos materiais e componentes em fase pré-comercial, de desenvolvimento ou pesquisa exploratória, num horizonte de até 15 anos.

A Figura 4.2 mostra a ideia geral do estudo (ABDI/CGEE/MS, 2014), que tem o formato de um funil que vai eliminando as de menor interesse. Inicialmente elegem-se as *Tecnologias Emergentes*. Destas, eliminam-se as consideradas não factíveis, restando as *Tecnologias Factíveis* (podem ser feitas no mundo em até 15 anos). Destas, eliminam-se as não viáveis economicamente, restando as *Tecnologias Viáveis* (possibilidade comercial em até 15 anos). Destas, elegem-se as Tecnologias de *Elevado Potencial de Difusão* no Brasil. Destas, elegem-se as *Tecnologias Relevantes Comercialmente* (com grande potencial de sustentabilidade). Nesse momento, separam-se ainda as *Tecnologias Prioritárias*, ou seja, que o Brasil possui competência para produção.

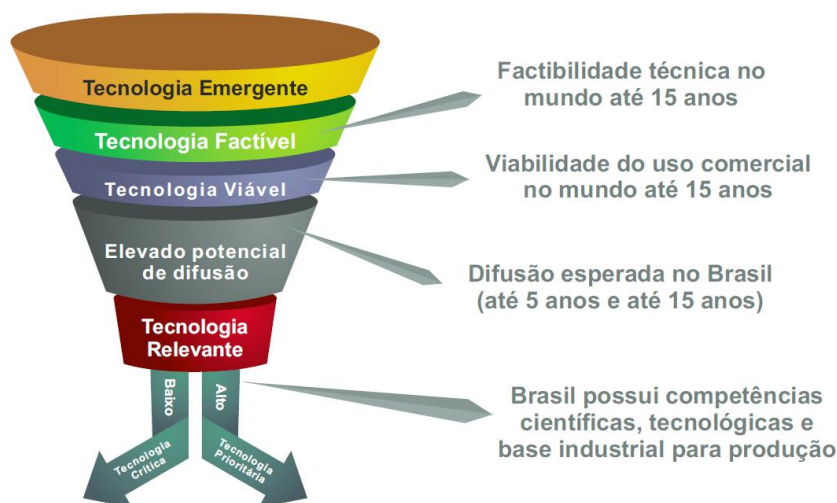


Figura 4.2. Metodologia da Agenda Tecnológica Setorial – Plano Brasil Maior – consulta estruturada

A Figura 4.3 apresenta a mesma ideia através de uma metodologia de *consulta estruturada* (ABDI/CGEE/MS, 2014). As tecnologias emergentes vão sendo eliminadas de acordo com os critérios, buscando escolher as mais relevantes economicamente e socialmente, dentro das possibilidades de produção do Brasil.

Consulta Estruturada

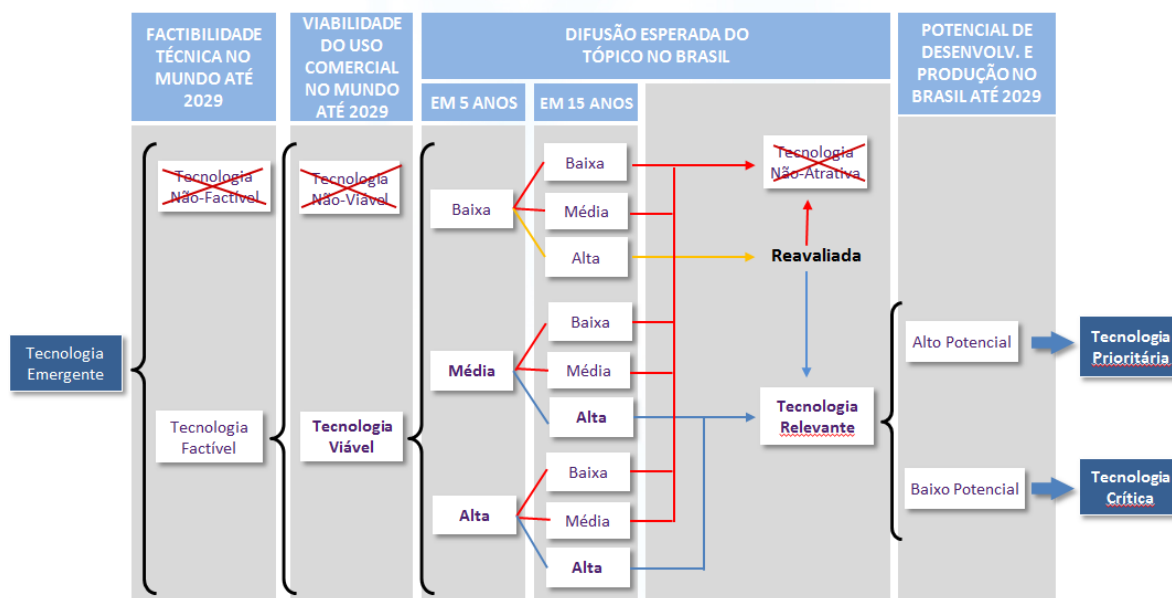


Figura 4.3. Metodologia da Agenda Tecnológica Setorial – Consulta estruturada

O texto a seguir representa um resumo explicativo feito pelos autores em relação às tecnologias escolhidas, não sendo idêntico ao texto final do estudo. As tecnologias relacionadas à videocolaboração relevantes, prioritárias e com alto potencial de difusão eleitas pelo estudo estão descritas nos próximos parágrafos.

As primeiras três tecnologias eleitas estão relacionadas a sistemas de videoconferência. São três pois representam detalhamentos dos sistemas, como a questão de que alguns necessitam conexão segura, outros necessitam uma conexão estável, outros necessitam maior banda, outras necessitam integração com sistemas de regulação, algumas com obrigatoriedade de imagem. Nesse caso entram os sistemas de cirurgia robótica controlados (pelo cirurgião) à distância, centros de simulação realística para aprimorar o ensino, videoconferências com qualidade adaptável para debates de técnicas cirúrgicas, salas cirúrgicas integradas, teleconsultoria, teleconsulta, debates em grupos de interesse (SIGs – *Special Interest Groups*), telediagnóstico, entre outros. O texto resumido dessas três tecnologias é o seguinte:

Uso de sistemas síncronos (podendo ter ou não vídeo), multiplataformas e interoperáveis (ambientes heterogêneos), integrados a registros eletrônicos de saúde (com vídeo, áudio e dados integrados), aplicados em ambientes críticos (urgência e emergência, como tele acompanhamento em tempo real e tele UTI, incluindo ambiente cirúrgico) e também aplicados em ambientes não críticos (sem situações de urgência e emergência), visando à colaboração, ensino e melhoria da qualidade da

interação para troca de informações de saúde através de múltiplas plataformas com conectividade dependente da aplicação.

Detalhando o texto acima, pode-se ressaltar os seguintes pontos:

- **Uso de sistemas síncronos:** reflete a necessidade da existência de tecnologias de colaboração em tempo real. Basicamente trata de sistemas de conferência, que pode ser através de chat (somente texto), áudio (telefonia via redes de computadores) ou vídeo (videoconferência).
- **Multiplataformas e interoperáveis:** reflete a necessidade da tecnologia permitir, dependendo da aplicação, a comunicação em ambientes heterogêneos, como dispositivos móveis, notebooks, plataformas dedicadas, e assim por diante. Uma área consolidada atualmente para isso é o uso do navegador web, através do WebRTC (*Web Real Time Communication*), no HTML5. Utilizando o navegador web para comunicação, se consegue facilmente o uso de múltiplas plataformas e sem necessidade de instalação de programas na máquina. Atualmente algumas necessidades de vídeo em ultra-alta definição ainda não são possíveis, mas os autores acreditam que em alguns anos isso será viável tecnologicamente.
- **Integrados a Registros Eletrônicos de Saúde:** a tecnologia deve permitir a integração com padrões de registro dos procedimentos, com o objetivo de compatibilizar o compartilhamento de dados digitais entre diferentes centros e serviços relacionados à assistência em saúde. Essa questão é de tal importância que recebeu um item dedicado numa tecnologia emergente definida pelo grupo.
- **Aplicado em ambientes críticos:** existem aplicações críticas que exigem a colaboração com diversos requisitos mínimos de qualidade. Por exemplo, uma cirurgia robótica à distância requer alta disponibilidade do ambiente (tolerância a falhas), alta banda (para permitir a transmissão de vídeo em alta definição ou ultra-alta definição) e baixo atraso (para que os movimentos do cirurgião reflitam rapidamente no equipamento remoto e sejam visualizados também rapidamente). Outro exemplo é acompanhamento de UTI, que possui uma exigência de alta disponibilidade, porém os requerimentos de banda e atraso são bem menos importantes que no caso da cirurgia.
- **Aplicado em ambientes não críticos:** existem aplicações não críticas que também possuem requisitos de qualidade. Videoconferências para a troca de experiências, por exemplo, possuem um foco maior no compartilhamento de tela, e não tanto na imagem das pessoas. A tolerância a falhas é desejável, porém não é um fator crítico. O áudio é bastante importante para que a comunicação seja bem entendida por todos. Nesse caso o uso de várias plataformas é bastante desejável, pois algumas pessoas acessam a conferência por dispositivos móveis, outros pelos notebooks, e assim por diante.

Outro grupo de tecnologias emergente eleitas pelo grupo pode ser resumido na seguinte frase, que ressalta os sistemas de telediagnóstico em dois aspectos: a) regulação descentralizada, buscando padronizar os mecanismos de acesso a exames diagnósticos; b) uso de sistemas de pré-laudo para apoiar o médico na confecção do laudo.

Uso de software multiplataforma de telediagnóstico envolvendo diferentes tipos de sinais (dados, áudio, vídeo e imagens), com mecanismos de segurança e confidencialidade, aplicados a redes de serviços de saúde (unidades básicas, hospitais, homecare) e diretamente a pessoas visando:

- *descentralização de métodos diagnósticos regulados por protocolos clínico-assistenciais.*
- *melhoria (acurácia, no tempo) na interpretação diagnóstica por meio do reconhecimento de padrões (pré-laudo).*

Nessa linha também foram ressaltadas tecnologias de telemonitoramento, seja para doenças crônicas (com comunicação através de conferência) ou para sistemas de bem-estar (wellness), que frequentemente utilizam vídeos para ilustrar a boa alimentação, o exercício ou similares. A seguinte frase resume a tecnologia.

Uso de software multiplataforma de telemonitoramento envolvendo diferentes tipos de sinais (dados, áudio, vídeo e imagens) e dispositivos, com mecanismos de segurança e confidencialidade, aplicado a pacientes e redes de serviços de saúde (unidades básicas, hospitais, homecare), visando ao telemonitoramento integrado de dados biométricos, relacionados a hábitos (alimentação) ou outros.

Além dos softwares de telemonitoramento, também se considerou como tecnologia emergente os sensores que alimentam o software com informações, como câmeras de vídeo em ambientes caseiros, sensores de presença, entre outros. O uso dessas tecnologias é muito variada, indo desde o monitoramento remoto como apoio a pessoas que requerem cuidados especiais (ex: dificuldade de movimentação / apoio para reabilitação). Essas tecnologias podem ser resumidas como:

Uso de dispositivos e sensores para captação de sinais vitais, medidas bioquímicas ou detecção de movimentos. Um dos tipos de sensores são os vestíveis não invasivos (relógios, óculos e pulseiras) interconectáveis com capacidade de telecomunicação sem fio. Tais sensores seriam aplicados a pessoas em seu domicílio, unidades básicas de saúde ou grandes hospitais, visando o acompanhamento e/ou monitoramento remoto de pessoas/pacientes.

Ainda na linha do biofeedback e telemonitoramento, o grupo sugeriu o seguinte conjunto de tecnologias emergentes:

Uso de dispositivos interconectáveis para captação de imagem em uso domiciliar, integrados com sistemas de telemedicina, aplicados ao monitoramento e interação com pessoas que requerem cuidados especiais, visando à reabilitação física (coordenação motora, atividades físicas ou movimentação).

Na área da educação, o grupo sugeriu também algumas tecnologias emergentes visando a potencialização do processo de ensino-aprendizagem que incluem educação à distância e repositório de conhecimentos (laboratório de habilidades), aplicados ao ensino em saúde. Essas tecnologias podem ser resumidas como:

Uso de objetos digitais de aprendizagem, incluindo simuladores, e-books interativos, realidade virtual, realidade aumentada, jogos etc..

Uso de Atlas Anatômico 3D Realístico Morfofuncional multiplataforma (tablets, smartphones, impressora 3D, esculpadora com braço robótico etc.), baseado em computação gráfica 3D, integrado ou não à realidade aumentada e a manequins de habilidades.

Uso de vídeos interativos de computação gráfica, de curta duração, para orientação específica de procedimentos.

4.5 Pesquisas em videocolaboração sendo efetuadas pelos autores

Esta seção detalha alguns projetos sendo efetuados pelo grupo e relacionados às tecnologias emergentes com foco em videocolaboração vistas no item anterior.

4.5.1 TelessaúdeRS-UFRGS

O TelessaúdeRS-ufrgs (www.ufrgs.br/telessauders) é um projeto de pesquisa do Programa de Pós-Graduação em Epidemiologia da Faculdade de Medicina da Universidade Federal do Rio Grande do Sul.

As ações do projeto são dirigidas à população, por meio de apoio aos profissionais da atenção primária à saúde e dos demais níveis assistenciais do Sistema Único de Saúde.

O TelessaúdeRS-UFRGS é uma estratégia para qualificação da Atenção Primária à Saúde (APS). As ações de teleconsultoria, telediagnóstico e teleducação, assim como de desenvolvimento de sistemas, são voltadas a todos os profissionais que trabalham na APS (médicos, enfermeiros, odontólogos, técnicos de enfermagem, técnicos e auxiliares em saúde bucal, agentes comunitários de saúde) e aos demais profissionais de outros níveis assistenciais.

Seu objetivo principal é qualificar o trabalho das equipes de atenção primária à saúde, fortalecendo os atributos da APS: acesso de primeiro contato, longitudinalidade, coordenação, integralidade, orientação familiar, orientação comunitária e competência cultural.

Um vídeo do TelessaúdeRS-UFRGS está disponível em <https://youtu.be/t11ZtIPN8Fk>.

Os principais serviços são de Teleconsultoria, Desenvolvimento (principalmente da Plataforma de Telessaúde e vários aplicativos móveis), Telediagnóstico, Teleducação e Apoio a núcleos e à informatização.

4.5.2 Sala MIR (Multimedia Integrated Room)

A sala cirúrgica integrada MIR (*Multimedia Integrated operating Room*) consiste de um sistema de software que se integra a salas cirúrgicas, tornando-as multimídia, e permite transmissão interativa em tempo real entre a sala e pontos remotos com as seguintes características:

- Videoconferência interna de alta qualidade via IP (Internet Protocol) entre salas de cirurgia e um ou mais pontos remotos (auditórios ou salas multi-uso). Essa transmissão pode ser integrada com a webconferência e ganhar um caráter

mundial, permitindo acesso em qualquer local do mundo via Internet, através do navegador web, seja em desktops, notebooks, tablets e dispositivos móveis (via app). Tanto canais de vídeo do ambiente quanto vídeos gerados por equipamentos médicos (por exemplo, videolaparoscópio, foco cirúrgico, robô Da Vinci, outros geradores de imagens analógicas ou digitais).

- Possibilidade de gravação das transmissões em alta qualidade para edição (anotações no vídeo) e uso em congressos, catalogação, conservação e apoio no ensino e pesquisa.
- Possibilidade de gravação e disponibilização em site web das webconferências.
- Possibilidade de transmissão e visualização de imagens médicas progressivas acessadas on-line através de um visualizador DICOM.

O sistema foi desenvolvido no âmbito de um edital FINEP (Financiadora de Estudos e Projetos), que financiou a maior parte do projeto.

A Sala Cirúrgica Multimídia Avançada possui uma ampla gama de utilizações, desde cirurgias videolaparoscópicas locais em alta definição (como utilizado nas maiorias das salas cirúrgicas integradas) até usos avançados, como orientação remota, cirurgias assistidas a distância, transmissão de uma cirurgia para diversos locais, entre outros.

Em termos de melhorias proporcionadas pela sala, pode-se citar:

- Otimização do tempo dos médicos especialistas: médicos especialistas podem otimizar seu poder de ensino, pois estarão acompanhando remotamente uma ou mais cirurgias efetuadas por médicos residentes, chegando às localidades mais remotas que tenham infra-estrutura de redes.
- Melhoria no treinamento dos novos cirurgiões: o treinamento de cirurgiões não precisa ficar mais limitado a grandes centros, podendo ser efetuado em localidades menores. O cirurgião pode, também, efetuar a cirurgia e transmitir a mesma em tempo real, explicando os procedimentos e respondendo perguntas.
- Utilização de técnicas minimamente invasivas: com o sistema de videolaparoscopia, o tempo de recuperação do paciente é muito mais rápido, ocupando menos tempo no leito, tendo menos risco de infecção, melhorando sua qualidade de vida.
- Maior durabilidade dos equipamentos e maior facilidade de limpeza: como a sala é fixada no teto, todo o cabeamento vem por cima, tornando o ambiente muito mais limpo, fácil de higienizar e eliminando riscos de quebra de cabeamento pela passagem de carrinhos pela sala.
- Facilidade no posicionamento dos equipamentos e conforto do cirurgião: com as estativas e o número de monitores e TVs, o cirurgião fica muito mais confortável no ambiente, tendo uma ampla gama de possibilidades de posicionamento do equipamento de acordo com sua necessidade.
- Fomento da indústria nacional: sendo um produto brasileiro, haverá redução da dependência estrangeira em produtos similares, que custam muito mais, gerando empregos em áreas de altíssimo nível.

A Figura 4.4 apresenta a sala cirúrgica em pleno funcionamento, e pode-se ver com mais detalhes o seu uso. Na figura da esquerda, pode-se observar um dos monitores pendurados na estativa apresentando a imagem do videolaparoscópio em tela cheia. A TV (detalhe na figura da direita) está apresentando a imagem do videolaparoscópio (superior à esquerda), ambiente da sala via câmera 1 (inferior à esquerda), do corpo do paciente via câmera 2 (superior à direita) e do médico preceptor remoto (inferior à direita).



Figura 4.4. Sala desenvolvida pelo grupo no âmbito de um edital FINEP, com parceria da Santa Casa de Misericórdia de Porto Alegre (ISCMPA). O grupo desenvolveu a sala cirúrgica completa, desde apoio ao desenvolvimento das estativas (pendentes do teto), até a integração com videolaparoscópios. O foco do projeto, entretanto, foi no software de videoconferência.

A Figura 4.5 mostra o ambiente da Santa Casa com duas TVs lado a lado, uma com a imagem do videolaparoscópio em tela cheia, e outra com as imagens radiológicas do paciente.



Figura 4.5. Uma das paredes da sala cirúrgica implantada na Santa Casa. A primeira TV está exibindo a imagem do videolaparoscópio em Full-HD, e a segunda imagens radiológicas do paciente buscadas no PACS da Santa Casa.

A Figura 4.6 mostra um exemplo do sistema sendo utilizado para transmissão de eventos. No caso, o VI SIVA (Simpósio Internacional de Videocirurgia Avançada), que é efetuado na Santa Casa de Porto Alegre de dois em dois anos.



Figura 4.6. Foto do auditório Hugo Gerdaud na Santa Casa mostrando a transmissão no VI SIVA (Simpósio Internacional de Videocirurgia Avançada), em março de 2014, utilizando o software desenvolvido. Observe que cada imagem possui aproximadamente 5m de largura e é Full HD. A projeção da esquerda mostra a imagem interna do paciente, e a da direita mostra o cirurgião debatendo o procedimento com o público. Eram 3 salas sendo transmitidas simultaneamente, cada uma com a transmissão interna e externa.

Em 2016, o sistema foi utilizado para efetuar a transmissão do VII SIVA (Simpósio Internacional de Videocirurgia Avançada), um evento organizado a cada dois anos na Irmandade da Santa Casa de Misericórdia de Porto Alegre (ISCMPA – <http://www.santacasa.tche.br>), um grande complexo de 7 hospitais, localizado em Porto Alegre, RS.

O objetivo do simpósio, que tem a duração de 3 dias, é a discussão de técnicas cirúrgicas, com a transmissão simultânea de 3 salas cirúrgicas ao vivo para um teatro remoto, e debates sobre os procedimentos. Cirurgiões de várias partes do Brasil e do mundo participam do evento. Duas cirurgias são exibidas no teatro através de dois projetores Full-HD. Normalmente a terceira sala está em preparação.

A Figura 4.7 apresenta uma visão do teatro do CHC (Centro Histórico Cultural) durante o evento. Pode-se perceber as duas transmissões acontecendo simultaneamente, sendo que o operador pode escolher, via tablet de controle, quais transmissões devem chegar ao palco. As possibilidades são várias, como a interna e externa de qualquer sala, a interna de uma sala e a externa de outra, e assim por diante. É possível também apresentar a imagem interna e externa no mesmo projetor, dividindo as telas em até 8 imagens simultâneas. Esse recurso foi utilizado em alguns momentos para mostrar várias imagens de duas salas, mas quando possível, passava-se para tela cheia a fim de ter um maior detalhamento da imagem.

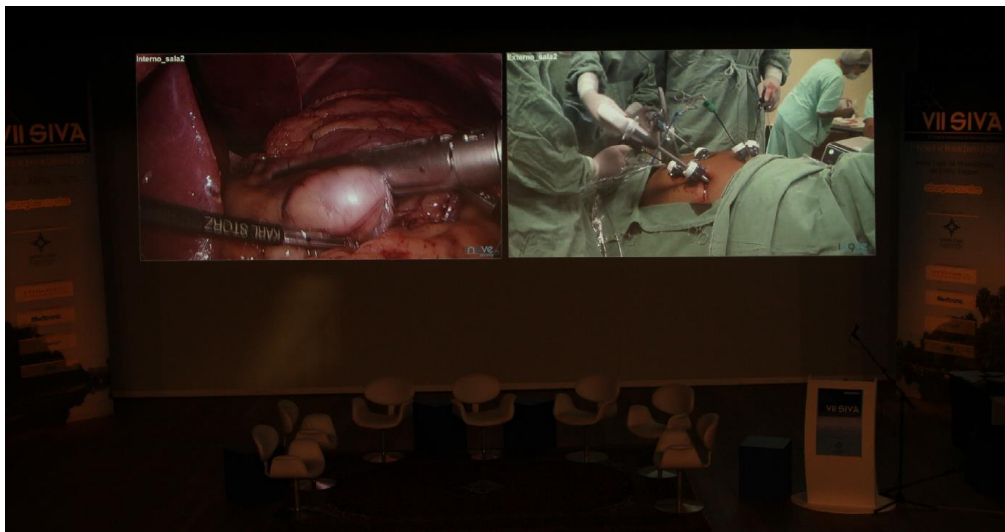


Figura 4.7. Visão do teatro do CHC com a transmissão acontecendo em tempo real

A interação entre as dezenas de cirurgiões participantes no teatro e o cirurgião na sala cirúrgica, executando e debatendo sobre sua técnica, foi bastante proveitosa. Em vários momentos houve debates calorosos sobre as técnicas utilizadas, mostrando que a tecnologia estava inserida de forma transparente e apoiando a comunicação dessa aplicação de e-Saúde. Isso só foi possível porque o atraso do sistema é bastante baixo, na ordem de 125ms, similar à sensação do ser humano numa conversa presencial. Além disso, a qualidade da transmissão permitia a visualização do procedimento de forma muito melhor do que estando na própria sala.

4.5.3 Projeto de Multipresença

O sistema de videoconferência denominado “Multipresença” está sendo executado com o patrocínio da RNP (Rede Nacional de Ensino e Pesquisa) no âmbito dos programas de Grupos de Trabalho.

O objetivo do GT-Multipresença é a interoperação transparente de pelo menos o seguinte:

- Sala de telepresença em alta definição (Full HD – 1920x1080p).
- Sala de ultra-telepresença em ultra alta definição (UHD 4K – 3840x2160).
- Acesso através de sistemas de videoconferência de sala (padrões SIP e H.323, integrando com Polycom, Cisco, e outros).
- Acesso em alta definição através de programa aplicativo no computador pessoal.
- Acesso através de webconferência.
- Acesso através de dispositivos móveis.
- Compartilhamento de conteúdo.

Um diferencial do sistema de Multipresença em relação aos outros sistemas, além do modelo de integração explicado acima e do baixo custo, é a possibilidade de adaptar o espaço físico de acordo com as necessidades do usuário.

A interoperação das múltiplas tecnologias propostas pelo sistema de Multipresença permite diferentes aplicações compartilharem a mesma sala, como por exemplo: a) aulas presenciais e a distância; b) dinâmica de grupo local e remota; c) reuniões em U; d) reuniões de diretoria; entre outras. Cada aplicação tem suas próprias demandas de comunicação em termos de qualidade, interoperação e colaboração.

Em agosto de 2015 um protótipo da solução foi apresentado durante o evento “Forum RNP”, em Brasília, Brasil. A Figura 4.8 mostra uma imagem do painel. As quatro TVs da esquerda mostram o compartilhamento de conteúdo através do SAGE2 (Scalable Amplified Group Environment, explicado a seguir), onde havia compartilhamento de logos, imagens e aplicações entre Brasília e Porto Alegre (2.000km de distância). As duas TVs do canto inferior direito mostram a videoconferência em alta definição ponto a ponto entre Brasília e Porto Alegre. As duas TVs do canto superior direito mostram a integração com o sistema de webconferência Mconf, com a TV da esquerda mostrando dois vídeos (um usuário conectado via tablet e outro via laptop) e a TV da direita mostrando slides de uma apresentação. A TV maior no canto direito da imagem mostra uma transmissão em tempo real em ultra-alta definição (4K) de Porto Alegre, onde um arquivo YUV de 3Gbit/s (simulando uma câmera ao vivo) era comprimido e enviado para Brasília em tempo real a aproximadamente 20 Mbit/s.



Figura 4.8. Foto do painel durante o Fórum RNP

É importante lembrar que ambos locais (Brasília e Porto Alegre) estavam conectados através do backbone da RNP, suportando toda a banda requerida para todas essas comunicações em tempo real.

Há um vídeo de demonstração do sistema Multipresença no endereço <https://www.youtube.com/watch?v=lud2KwzstCM> (com legendas em inglês). Este vídeo foi gravado durante o evento “Fórum RNP”, em agosto de 2015.

A comunicação com alta definição, incluindo a integração com sistemas legados de videoconferência, é efetuada através de um módulo chamado “PRAV Player”.

O Mconf é o módulo de webconferência da solução do Multipresença (ROESLER 2016). Ele é responsável pela integração com dispositivos móveis (Android e iOS), bem como pela integração com telefones SIP.

O módulo de compartilhamento de conteúdo foi criado através de uma integração do Multipresença com o sistema de código aberto de colaboração chamado SAGE2 (SAGE, 2015), desenvolvido pelas universidades de Illinois e do Havaí.

O SAGE2 é uma ferramenta de software que permite aos usuários compartilhar suas telas, arquivos e aplicativos através de redes IP, criando um campo virtual onde participantes remotos e locais podem compartilhar e visualizar os conteúdos.

A Figura 4.9 mostra o painel instalado na UFRGS durante uma reunião. O painel possui 6 TVs. As 3 TVs de baixo estão configuradas para alta definição, e estão se comunicando respectivamente com o INCA (Instituto Nacional do Câncer) no Rio de Janeiro, com a Escola Bahiana de Medicina e Saúde na Bahia, e com a Escola Superior de Redes em Brasília. As duas TVs do canto superior esquerdo mostram os participantes conectados através de webconferência (no caso são 7 participantes mais o vídeo da sala aonde a foto foi tirada, portanto 8 vídeos), e a TV do canto superior direito mostra a apresentação de slides sendo efetuada.



Figura 4.9. Painel do Multipresença instalado na UFRGS

Um caso de uso do sistema em saúde foi feito com apoio do Instituto Nacional do Câncer (<http://www2.inca.gov.br>) está atualmente utilizando o sistema Multipresença para transmissões de cirurgias robóticas em alta-definição e via web. Apesar de ainda ser um projeto piloto, a ideia do instituto é criar uma rotina de transmissões, com objetivo de disseminar conhecimento em técnicas cirúrgicas.

A Figura 4.10 mostra uma transmissão de cirurgia em tempo real realizada em 2017. As duas telas de baixo mostram a cirurgia em tempo real (visão interna e externa do paciente), enquanto a TV de cima mostra algumas pessoas conectadas através de webconferência, inclusive um tablet dentro da sala de cirurgia.



Figura 4.10. Transmissão de cirurgia integrando alta definição e webconferência

A Figura 4.11 mostra a mesma transmissão sendo visualizada através do sistema de webconferência. As pessoas que assistem via navegador web possuem as mesmas imagens de quem está com o painel, porém a qualidade da transmissão é inferior, devido ao fato do acesso ser adaptada para baixa banda.

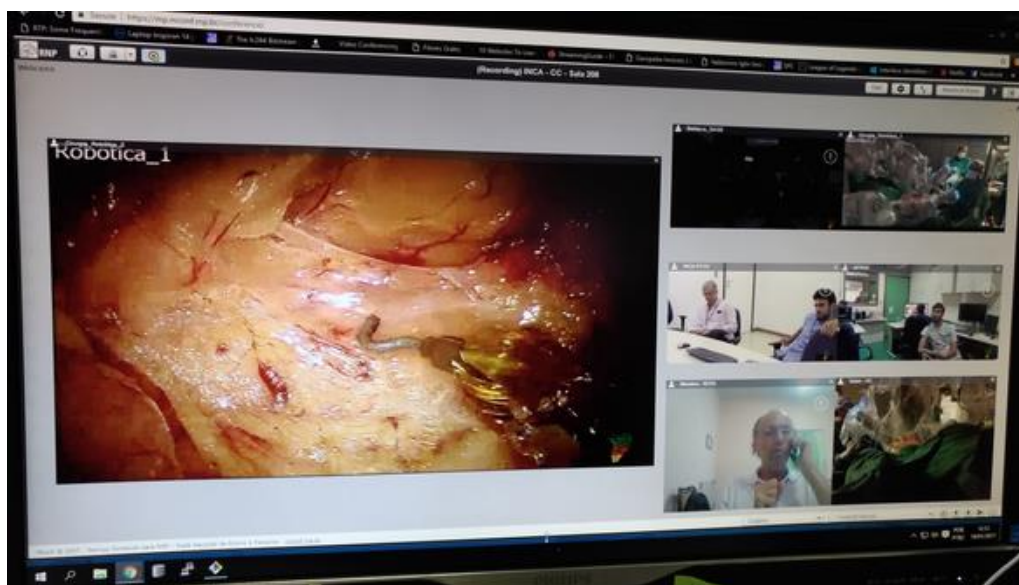


Figura 4.11. Visão da cirurgia através de webconferência (navegador web)

4.5.4 Projeto de Teleoftalmologia

A Figura 4.12 mostra o projeto Teleoftalmo – Olhar Gaúcho, inaugurado dia 10 de julho de 2017. O projeto permite telediagnósticos em oftalmologia, e prevê oito consultórios remotos no Estado do Rio Grande do Sul, com um centro de controle em Porto Alegre onde os médicos analisam diversos aspectos em relação à visão do paciente.

O projeto foi liderado pelo TelessaúdeRS-UFRGS com o patrocínio do Hospital Moinhos de Vento no âmbito do programa PROADI-SUS do Ministério da Saúde e da Secretaria estadual de Saúde do Rio Grande do Sul. Também estão envolvidos o Ministério da Saúde e prefeituras dos municípios envolvidos. A parte de comunicação por videoconferência, vista na figura, foi desenvolvida pelo laboratório do PRAV (Projetos em Áudio e Vídeo)¹ do Instituto de Informática da UFRGS².

O oftalmologista necessita da imagem em alta definição para verificar diversas questões relativas ao olho, como foco, acuidade visual, entre outras. O paciente sai do pólo remoto já com diagnóstico completo e até receita de óculos estará disponível com o médico solicitante, caso paciente necessite.



Figura 4.12. Consultório remoto da Restinga. Foto: Guilherme Longoni/PRAV

Essa inovação visa reduzir a espera por atendimento na especialidade de oftalmologia, uma das mais procuradas pela comunidade. A fila de pacientes aguardando uma consulta é enorme: são 9 mil pessoas que levam mais de um ano para serem atendidas e diagnosticadas. O projeto tem como objetivo diminuir drasticamente esse problema.

O PRAV, representando o INF, está envolvido na parte de transmissão de imagens para a videoconferência entre a médica e o consultório remoto. A tecnologia desenvolvida é totalmente nacional, e permite a transmissão de imagens em alta definição e praticamente sem atraso (a interação é tão natural como se o médico estivesse localmente no consultório remoto)

O funcionamento do teleoftalmo é simples, podendo ser resumido da seguinte forma:

- a) O médico da atenção primária envia solicitação de atendimento especializado via Plataforma de Telessaúde (plataformatelessaude.ufrgs.br)

¹ www.inf.ufrgs.br/prav

² www.inf.ufrgs.br

- b) A equipe do TelessaúdeRS-UFRGS faz o agendamento com o paciente.
- c) O exame é realizado remotamente pelos oftalmologistas do projeto, com apoio presencial da equipe de enfermagem. Os seguintes exames são realizados: Teste de acuidade visual; Refração; Medida da pressão intraocular; Avaliação das pálpebras; Motilidade ocular; Reflexos pupilares; Documentação do segmento anterior e do fundo do olho; Exame síncrono, em que o oftalmologista comanda o equipamento médico a distância, ao mesmo tempo em que acompanha, por câmeras, o que ocorre nas salas.
- d) O paciente já sai com o resultado da avaliação impresso. O laudo também é enviado pela Plataforma de Telessaúde/MS para o médico solicitante com recomendações de conduta.
- e) Se houver necessidade, os óculos serão fornecidos sem custo ao paciente, confeccionados por ótica contratada pelo projeto.

Cada consultório remoto terá capacidade de realizar aproximadamente 500 atendimentos por mês. Dessa forma, o sistema com 8 consultórios possui capacidade para 4.000 pacientes por mês. O objetivo é eliminar as filas do SUS na especialidade. Segundo a Secretaria Estadual da Saúde, pelo menos 9 mil pessoas esperam atendimento na área, não incluindo nesse número a demanda de cada município, o que pode ampliar esse dado para 15 mil pacientes, em um tempo de espera para consulta que ultrapassa um ano.

A oftalmologia adulta está entre as três especialidades com maior demanda reprimida na cidade, e o projeto deve facilitar o acesso ao atendimento, que também inclui a confecção de óculos gratuitamente para os pacientes atendidos pelo sistema. Cada consultório é equipado com três câmeras, computadores e equipamentos de diagnóstico oftalmológico. Serão atendidas crianças a partir dos oito anos e adultos preferencialmente moradores da macrorregião de saúde em que está instalado. Dependendo do diagnóstico, o paciente poderá ter seu problema resolvido na própria unidade básica de saúde e poderá fazer os óculos após a avaliação com o médico no posto de saúde. Pacientes com doenças como glaucoma e retinopatia diabética, muitas vezes ainda não diagnosticadas, terão prioridade no encaminhamento ao oftalmologista.

A Figura 4.13 mostra uma imagem no Centro de Controle localizado no Telessaúde-RS. Cada médico oftalmo deverá controlar intercaladamente duas salas, otimizando seu tempo e agilizando o atendimento. A figura mostra parte do ambiente de controle, onde a médica verifica os dados dos instrumentos remotos e também interage através da videoconferência.

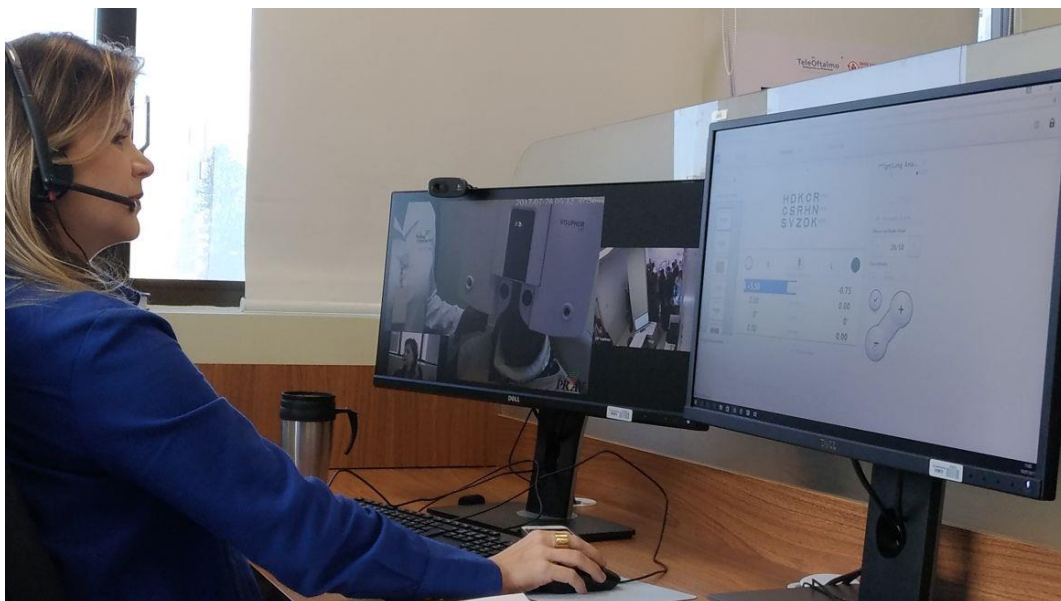


Figura 4.13. Centro de controle no TelessaúdeRS-UFRGS. Foto: Guilherme Longoni/PRAV

4.5.5 Ambiente de Simulação Realística

Apesar de não estar sendo desenvolvido diretamente pela UFRGS, é importante mencionar o sistema de Simulação Realística denominado Vtraining, derivado dos projetos do laboratório do PRAV.

O sistema Vtraining permite a criação de um ambiente para treinamento médico e simulação realística assistido por vídeo, e está sendo desenvolvido e comercializado pela empresa i9access Tecnologia Ltda (que é uma spin off do Laboratório do PRAV).

O sistema Vtraining tem os seguintes principais modos de operação:

- Mosaico: Acompanhamento simultâneo de todas as salas de simulação; Envio de áudios ao vivo e gerenciamento de gravações para futuro debriefing
- Sala de Aula: Visualização da sala de simulação para acompanhar treinamentos. Anotações a serem compartilhadas com a turma (atuantes na simulação, participantes presenciais e também remotos via internet).
- Controle: Sincronização das salas e treinamentos

Outras funcionalidades do sistema incluem Gravação, Modo Histórico, Modo de Cadastros, entre outros.

A Figura 4.14 mostra uma imagem da tela do modo Mosaico, que possui os seguintes quadrantes:

- Área principal: área central vista na figura. Permite visualização dos vídeos, escuta de áudio, tela cheia e swap.
- Área superior à direita: permite escolher quais vídeos serão vistos na área principal. Permite também escolher quais salas podem receber áudio. Permite

também verificar o status das gravações nas salas cadastradas e iniciar ou parar uma gravação manualmente.

- Área inferior à direita: permite a gravação e o envio de mensagens de áudio pré-gravadas padrão, pré-gravadas customizadas ou gravadas na hora.

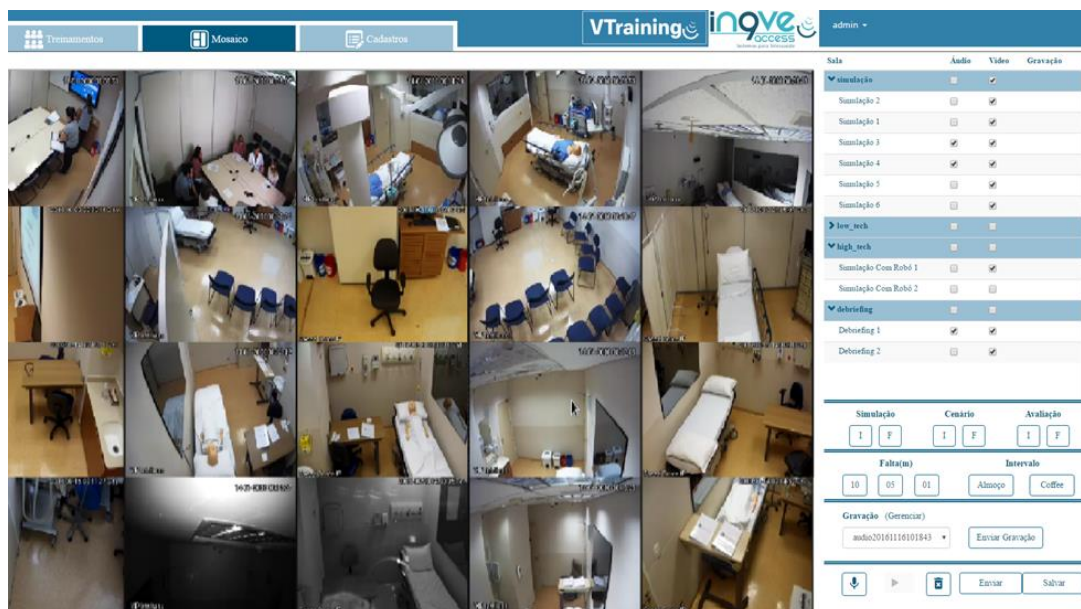


Figura 4.14. Visão no modo Mosaico

A Figura 4.15 mostra a aba de Cadastros, que é composta por 6 ícones clicáveis, que são: Salas, Computadores, Câmeras e Som, Usuário, Perfil, Treinamentos. O sistema é dividido nos seguintes grupos de atividades:

- Cadastro de equipamentos e infraestrutura: permite a criação, visualização, atualização e remoção de salas, computadores, câmeras de vídeo, microfones, entre outros.
- Cadastro de contas: permite a criação, visualização, atualização e remoção de usuários e perfis no sistema. Cada usuário possui os direitos definidos no seu perfil. São permitidos 3 perfis de usuários: Administrador, Operador e Instrutor.
- Treinamentos: permite cadastrar treinamentos para que os mesmos sejam gravados e armazenados no modo histórico.



Figura 4.15. Modo de cadastro do sistema Vtraining

A Figura 4.16 mostra o modo de sala de aula, onde o professor e parte do grupo de alunos visualiza e escuta o que está acontecendo no ambiente de simulação, onde estão presentes um ou mais alunos. O professor pode fazer anotações de pontos relevantes para apresentar o vídeo aos alunos quando eles retornarem à sala de visualização, no momento de debriefing, ou debates sobre o procedimento efetuado.



Figura 4.16. Visualização remota de um procedimento com manequim

4.6 Visão de futuro

A própria lista de tecnologias com foco em videocolaboração apresentada neste documento é um sumário da tendência de evolução tecnológica da Telemedicina em nível nacional e internacional. Ao revisar literatura internacional sobre o tema verificou-se a grande similaridade entre as tecnologias citadas aqui e as tendências citadas por especialistas internacionais. Muitas dessas tecnologias realmente serão incorporadas aos

serviços de saúde caso haja um progresso na qualidade, tamanho, estabilidade e segurança nas redes de transmissão de dados wireless, requisito fundamental para a evolução da Telemedicina (INNOVATION WORKING GROUP, 2013; FONG, 2010; SCHWAMM, 2014; BASHSHUR, 2014).

Entre as principais tendências, estão áreas consideradas prioritárias por vários especialistas e representadas por tecnologias emergentes citadas neste documento:

- Tecnologias para cuidado de idosos: sensores de presença ligados a centrais de monitoramento a distância.
- Tecnologias que permitam teleconsulta: robôs que integrem videoconferência e múltiplos dispositivos biomédicos (esfigmomanômetro digital, oxímetros, etc), além de sensores somatosensoriais que permitam a percepção dos sentidos a distância, simulando perfeitamente uma consulta presencial.
- Tecnologias baseadas na “Internet das Coisas”: como dispensadores eletrônicos de medicamentos ligados a dispositivo de aviso aos pacientes e demais tecnologias que evoluam para *smart homes* e *smart hospitals*, em que medidas de proteção e segurança, além de monitoramento, são tomadas automaticamente pela conexão de dispositivos e sensores aos próprios objetos da casa ou de um leito hospitalar (controle automatizado de gotejamento de infusão endovenosa frente a parâmetros clínicos, por exemplo). Pode-se imaginar algum momento futuro em que um sensor de sinal vital, por exemplo temperatura, poderá acionar automaticamente a administração endovenosa de um antitérmico a um paciente monitorado em uma Unidade de Cuidados Intensivos sem a intervenção humana, com exceção da programação desta ação.
- Tecnologia para simulação de interatividade humana: robôs interativos para pessoas idosas ou pessoas portadoras de transtorno mental a fim de estimular comunicação e cognição.
- Tecnologias “vestíveis”: dispositivos *wearable* de monitoramento de sinais vitais e de medidas bioquímicas ou de parâmetros clínicos.

4.7 Agradecimentos

Gostaríamos de agradecer o trabalho executado pela ABDI (Associação Brasileira de Desenvolvimento Industrial) na organização do grupo de Telemedicina no contexto da Agenda Setorial do Plano Brasil Maior. Também gostaríamos de agradecer ao financiamento de alguns projetos pela RNP (Rede Nacional de Ensino e Pesquisa) no âmbito dos seus grupos de trabalho. Também à FINEP (Financiadora de Estudos e Projetos) e CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) pelo financiamento da Sala Cirúrgica Integrada. Também ao Ministério da Saúde e Secretaria do Estado do Rio Grande do Sul, pelo apoio e financiamento do Telessaúde-RS. Sem esse apoio o País estaria numa situação bem pior do que está hoje.

4.8 Referências

- ABDI/CGEE/MS. Plano Brasil Maior 2011/2014. Agenda Técnica Setorial. Apresentação técnica. 2014.
- BASHSHUR, R. L. et al. The empirical foundations of telemedicine interventions for chronic disease management. *Telemedicine Journal and e-Health*, Larchmont, v. 20, p. 769-800, 2014.
- BODENHEIMER, T. Coordinating Care: a perilous journey through the health care system. *New England Journal of Medicine*, Boston, v. 358, n. 10, p. 1064-1071, 2008.
- BRASIL. Ministério da Saúde. Plano de ações estratégicas para o enfrentamento das doenças crônicas não transmissíveis (DCNT) no Brasil, 2011-2022. Brasília: Ministério da Saúde, 2011a. (Série B. Textos Básicos de Saúde).
- BRASIL. Ministério da Saúde. Plano Nacional de Saúde – PNS: 2012-2015. Brasília: Ministério da Saúde; 2011b. (Série B. Textos Básicos de Saúde). BRASIL. Ministério da Saúde. Portaria nº 2.546, de 27 de outubro de 2011. Redefine e amplia o Programa Telessaúde Brasil, que passa a ser denominado Programa Nacional Telessaúde Brasil Redes (Telessaúde Brasil Redes). *Diário Oficial da União*, Brasília, DF, 28 out. 2011, Seção 1, p. 50. 2011c.
- CHAO L. W. et al. Evaluation of an Internet-based teledermatology system. *Journal of Telemedicine and Telecare*, London, v. 9, suppl. 1, p.9-12, 2003.
- CONSELHO FEDERAL DE MEDICINA. Resolução nº 1.643, de 7 de agosto de 2002. Define e disciplina a prestação de serviços através da telemedicina. *Diário Oficial da União*, Brasília, DF, 26 ago. 2002. Seção 1, p. 205.
- DAVIES, G. et al. Coordination of care within primary health care and with other sectors: a systematic review. Sydney: University of New South Wales, Australian Primary Care Research Institute, 2006.
- FERREIRA, L. V. et al. TIPIRX: tele-integração por imagens de raio-x. In: CONGRESSO BRASILEIRO DE TELEMEDICINA E TELESSAÚDE, 4., 2009, Belo Horizonte. Anais... Belo Horizonte, 2009.
- FIOCRUZ PENSESUS: <https://pensesus.fiocruz.br/atencao-basica>.
- FONG, B.; FONG, A. C. M.; LI, C. K. Telemedicine technologies: information technologies in medicine and telehealth. Chichester: John Wiley & Sons, 2010.
- FUNDAÇÃO UNIVERSITÁRIA DE CARDIOLOGIA. Instituto de Cardiologia. Na vanguarda de ações inovadoras em cardiologia: como uma das áreas mais desenvolvidas da telemedicina, a telecardiologia encontra destaque no cenário internacional devido ao grande impacto causado na saúde das populações, 2012. Disponível em: <<http://www.cardiologia.org.br/siteConteudo.aspx?id=8>>. Acesso em: 19 jan. 2012.
- GERVÁS, J.; PÉREZ-FERNANDEZ, M. El fundamento científico de la función de filtro del médico general. *Revista Brasileira de Epidemiologia*, São Paulo, v. 9, n. 1, 2006.

- HOFMARCHER, M. M.; OXLEY, H.; RUSTICELLI, E. Improved health system performance through better care coordination. OECD Health Working Papers, Paris, n. 30, Dec. 2007.
- INNOVATION WORKING GROUP. Roadmap for telemedicine: key considerations and recommendations. [Documento eletrônico]. [s.l.]: IWG, 2013. Disponível em: <<http://catai.net/blog/wp-content/uploads/2014/09/roadmapfortelemedicineiwgasia2014-140919210153-phpapp02.pdf>>. Acesso em: 01 setembro 2017.
- MAGRIÇO, A. et al. Consulta de diabetes ocular: primeiros resultados do rastreio da retinopatia diabética por câmara não midriática. *Oftalmologia*, Lisboa, v. 35, n.1, p. 67-73, jan/mar. 2011
- MESSINA, L. et al. Telediagnóstico e a Formação de Redes de Assistência. São Paulo: Sociedade Brasileira de Informática em Saúde, 2010.
- NORMAN, A. H.; TESSER, C. Prevenção quaternária na atenção primária à saúde: uma necessidade do sistema Único de Saúde. *Cadernos de Saúde Pública*, Rio de Janeiro, v. 25, n. 9, 2009.
- PARÉ, G.; JAANA, M.; SICOTTE, C. Systematic review of home telemonitoring for chronic diseases: the evidence base. *Journal of the American Medical Informatics Association: JAMIA*, [S.l.], v. 14, n. 3, p. 269-277, 2007.
- REDE UNIVERSITÁRIA DE TELEMEDICINA [Homepage]. Disponível em: <<http://rute.rnp.br/>>.
- Roesler, Valter; COELHO, L. ; LONGONI, G. ; MARINS, A. ; DARONCO, Leonardo Crauss ; CIUFFO, L. ; DUARTE, R. . Multipresence: towards videoconference and collaboration in multi-use environments. In: TNC16 Networking Conference, 2016, Praga. 32th TNC Networking Conference, 2016.
- ROESLER, VALTER; Daronco, Leonardo Crauss; CECAGNO, F.; MARINS, A. Mconf: collaboration proposal to form a global infrastructure for web conferencing based on open source. *Proceedings of the Asia-Pacific Advanced Network*, v. 35, p. 28, 2013.
- ROESLER, V.; Cirano Iochpe; CUNHA, A.; BINOTTO, A. A Real-Time Collaborative Tele-Ultrasonography System Applied to Underserved Communities. *IEEE Potentials*, v. 31, p. 22-27, 2012.
- SAGE (2015). Electronic Visualization Laboratory, University of Illinois at Chicago; Laboratory for Advanced Visualization & Applications, University of Hawai'i at Manoa. SAGEnext: Next Generation Integrated Persistent Visualization and Collaboration Services for Global Cyberinfrastructure. Si2 PI Workshop Forum, National Science Foundation. February 17, 2015.
- SAVARIS, A. et al. O uso da telemedicina assíncrona em larga escala no setor público de saúde. São Paulo: Sociedade Brasileira de Informática em Saúde, 2011.
- SCHOEN et al. New 2011 survey of patients with complex care needs in eleven countries finds that care is often poorly coordinated. *Health Affairs*, 2011;30(12):2437-2448.

SCHWAMM, L. H. Telehealth: seven strategies to successfully implement disruptive technology and transform health care. *Health affairs*, Bethesda (MD), v. 33, n. 2, p. 200-206, 2014.

TELESSAÚDERS/UFRGS. [Homepage]. Porto Alegre: TelessaúdeRS/UFRGS, 2017. Disponível em: <<http://www.ufrgs.br/telessauders>>.

WHO. A health telematics policy in support of WHO's Health-for-all strategy for global health development : report of the WHO Group Consultation on Health Telematics. Geneva: World Health Organization, 1998. Disponível em: <<http://apps.who.int/iris/handle/10665/63857?mode=full>>.

ANEXO I – biografia dos autores



Valter Roesler: Possui graduação em Engenharia Elétrica pela Universidade Federal do Rio Grande do Sul (1988), mestrado (1993) e doutorado (2003) em Ciências da Computação pela Universidade Federal do Rio Grande do Sul. Atualmente é Professor Associado na Universidade Federal do Rio Grande do Sul. Tem experiência na área de Ciência da Computação, com ênfase em Redes de Computadores, atuando principalmente nos seguintes temas: Telemedicina, Teleducação, Multimídia e Redes de Computadores. Coordenador do laboratório do PRAV (Projetos em Áudio e Vídeo) - www.inf.ufrgs.br/prav. Líder do grupo de pesquisa de Telessaúde, no diretório dos grupos de pesquisa CNPq. Coordenou diversos projetos relacionados à telemedicina e teleducação, como o Multipresença, o Ambiente de Simulação Realística e o Sala Cirúrgica Inteligente. Coordena a parte de videocolaboração do projeto de teleoftalmologia.



Erno Harzheim: Possui graduação em Medicina pela Universidade Federal do Rio Grande do Sul (1998), residência médica em Medicina de Família e Comunidade pelo Grupo Hospitalar Conceição (2001), doutorado em Medicina Preventiva e Saúde Pública pela Universidade de Alicante, Espanha (2004), e pós-doutorado em Epidemiologia pela Universidade Federal do Rio Grande do Sul (2006). É professor associado do Departamento de Medicina Social da Faculdade de Medicina da UFRGS e professor permanente do Programa de Pós-Graduação em Epidemiologia da mesma instituição. Tem experiência na área de Medicina com ênfase em Medicina de Família e Comunidade. Atua principalmente nos seguintes temas: Medicina de Família e Comunidade, Epidemiologia, Atenção Primária, Avaliação de Serviços de Saúde, Telemedicina e Telessaúde, Doenças Cardiovasculares e HIV/AIDS. Coordenou, por cerca de dez anos, o TelessaúdeRS-UFRGS, núcleo de pesquisa vinculado ao Programa de Pós-Graduação em Epidemiologia da UFRGS que tem como objetivo melhorar a qualidade do atendimento da Atenção Primária à Saúde do Sistema Único de Saúde (SUS). Atualmente é Secretário Municipal de Saúde de Porto Alegre.

Capítulo

5

Análise dos processos videocolaborativos do Núcleo Telessaúde Brasil Redes Unifesp na Atenção Básica

Cicero I. da Silva¹, Claudia Galindo Novoa¹, Alberto Cebukin e Jane de Almeida²

¹ Universidade Federal de São Paulo: NTC Telessaúde Redes (Telessaúde/Unifesp)

² Universidade Mackenzie: Laboratório de Artes Cinemáticas e Visualização (LabCine)

cicero.silva@unifesp.br, claudia.novoa@unifesp.edu.br, cebukin@unifesp.br,
janedalmeida@post.harvard.edu

Abstract

This article analyzes and presents the projects carried out by the research groups linked to the Telehealth Brazil Networks Program, from the Federal University of São Paulo (Unifesp) and the Laboratory of Cinematic Arts and Visualization (LabCine) from Mackenzie University. Secondly, the article also proposes to evaluate current and future applications in the field of video collaboration related to the use of video-based streaming environments for the so-called "synchronous teleconsulting", which are decision support processes in the health area that use video chat resources between two health professionals applied to Primary Health Care.

Resumo

O presente artigo analisa e apresenta os projetos realizados pelos grupos de pesquisa ligados ao Núcleo Telessaúde Brasil Redes, da Universidade Federal de São Paulo (Unifesp) e ao grupo de pesquisa do Laboratório de Artes Cinemáticas e Visualização (LabCine) da Universidade Mackenzie. Em um segundo momento, o artigo também propõe avaliar aplicações atuais e futuras no campo da videocolaboração relacionadas à utilização de ambientes de streaming baseados em vídeo para as denominadas "teleconsultorias síncronas", que são processos de apoio à decisão na área da saúde que utilizam recursos de chat de vídeo entre dois profissionais de saúde aplicadas à Atenção Primária em Saúde.

5.1. Introdução

A utilização de recursos de vídeo nas ações ligadas ao cuidado em saúde têm cada vez mais ocupado um espaço na agenda de gestores e provedores de saúde. A potencialidade

de comunicação entre dois profissionais de saúde, ou mesmo entre gestores, por meio do uso de recursos de vídeo, tem possibilitado inúmeras trocas e avanços no diagnóstico, tratamento e acesso às redes referenciadas em saúde. A troca de informações por meio do vídeo também tem ganhado espaço nas ações de teleeducação em saúde, tendo em vista a possibilidade de interação em tempo real, por meio dos recursos audiovisuais, entre o professor e o aluno, o que torna a ação didática mais próxima de uma dinâmica da sala de aula. Para que esse contato por meio do uso de recursos audiovisuais aconteça, existe uma necessidade de infraestrutura específica, que permite que os usuários possam se comunicar sem muitos percalços. Nesse sentido, o uso de protocolos de *streaming* de vídeo tem se tornado um importante elemento para a pesquisa contemporânea, pois permite que atividades científicas, sociais, culturais, pedagógicas e didáticas possam ser compartilhadas por diversos pesquisadores ao mesmo tempo, debatidas e avaliadas em tempo real. Contudo, o *streaming* gera um volume de dados intenso criando a necessidade de aumento de recursos de hardware para armazenagem dessas gravações, além de um extenso material “bruto”, muitas vezes sem narrativa que o torne acessível. Como exemplo pode-se pensar na transmissão de procedimentos cirúrgicos em tempo real, utilizando recursos de Ultra Alta Definição (UHD), 4K ou acima, que geram quantidades de dados muito superiores à capacidade de armazenamento padrão. Os conteúdos transmitidos são extensos, tornando difícil a busca, a seleção e, por fim, a visualização de maneira dinâmica pelos interessados. Atualmente diversos serviços de exibição de vídeos têm disponibilizado ferramentas de transmissão de vídeos em tempo real, tais como o Facebook Live e, mais recentemente, o YouTube, que disponibilizou um serviço experimental de *streaming* para 10 mil usuários nos Estados Unidos¹. Para o teórico da visualização Lev Manovich, dissertando sobre as novas formas narrativas a partir da visualização massiva de dados, afirma que:

Assim, por exemplo, se queremos criar uma amostra representativa de filmes do século XX, podemos usar o IMDb, que contém informações sobre 3,4 milhões de filmes, programas de TV (incluindo episódios separados). Da mesma forma, podemos criar uma boa amostra de páginas de jornais norte-americanos históricos usando o acervo do *Historical American Newspaper* com milhões de páginas digitalizadas da Biblioteca do Congresso Americano. (...) No caso de "grandes dados culturais", o cultural e o social se sobrepõem. (...) Grandes grupos de pessoas de diferentes países e contextos socioeconômicos (perspectiva sociológica) compartilham imagens, vídeos, textos e, ao fazerem isso, estão determinando escolhas estéticas. (Manovich, 2017, pgs. 28 e 29).

Como exemplo de aplicações futuras dos resultados que serão desenvolvidos por este artigo, podemos pensar nas ações de educação digital existentes no Brasil, tais como a Universidade Aberta do Brasil, a Universidade Aberta do SUS (Unasus), a Rede Universitária de Telemedicina (RUTE), o Programa Telessaúde e as próprias universidades, que necessitam cada vez mais de sistemas que possibilitem o envio e a recepção de atividades em tempo real por meio da videodifusão on-line. Um exemplo prático do uso da transmissão educacional de vídeos são os procedimentos cirúrgicos

¹ Ver em <https://arstechnica.com/business/2017/02/youtube-now-lets-creators-with-10000-subscribers-live-stream-video-on-mobile/>. Acessado em 03/04/2017.

realizados nas faculdades de medicina, tais como cirurgias neurológicas de campo aberto que duram em torno de 8 horas, sendo que, para atividades de ensino e pesquisa, o procedimento que de fato interessa pode ter uma duração bem menor, gerando vídeos de uma hora ou mesmo de 10 minutos para divulgação e registro da pesquisa. Outro exemplo é a preparação do paciente em procedimentos cirúrgicos, que pode levar algumas horas e que não é essencial na área do ensino de um procedimento cirúrgico, assim como a finalização deste procedimento. No campo do ensino, sabe-se que várias atividades gravadas no *streaming*, que foram importantes na pesquisa compartilhada, não serão posteriormente necessárias à utilização dos docentes.

Nesse sentido, o uso cada vez mais intenso de captação de vídeo por meio das tecnologias digitais aponta para uma nova baliza temporal em termos de arquivamento, como afirma Sobrinho:

...consideramos as mídias digitais, ou os ecossistemas digitais, como parâmetros de transformações e, ao mesmo tempo, sinalizadoras de uma fronteira histórica que coloca uma baliza no tempo: o passado analógico, o presente e o futuro digitais. (Sobrinho, 2016, pg. 11).

Dessa maneira, o presente artigo tem como perspectiva analisar as plataformas de transmissão de vídeos em tempo real (*streaming*) que têm sido utilizadas por inúmeras instituições de saúde no Brasil, com foco principalmente no Estado de São Paulo, mais precisamente no território no qual o Núcleo Telessaúde Brasil Redes da Unifesp atua, e onde hoje encontra-se um ambiente acadêmico propício à avaliação das pesquisas e inovações na área de videodifusão on-line em tempo real. O presente artigo leva em consideração aspectos como a base de armazenamento dos dados, as ferramentas de recurso de incorporação de vídeos em bases dedicadas ao ensino e pesquisa, como nos casos dos repositórios de vídeos utilizados nas salas de aula virtuais pelas universidades, entre outras aplicações. O artigo também analisa os recursos de transmissão de vídeos que vêm sendo desenvolvidos por publicações internacionais, como o grupo SAGE (Sage Video), Elsevier e JoVE (*peer-reviewed scientific video protocols to accelerate biological, medical, chemical and physical research*).

A Universidade Federal de São Paulo (Unifesp), por sua vez, possui um número significativo de pesquisas e experimentos na área de transmissão, edição, codificação e visualização de dados na área de vídeo. A Unifesp já desenvolveu, em parceria com o LabCine da Universidade Mackenzie, alguns protocolos de transmissão em tempo real e sistemas de visualização. No LabCine também acontecem eventos internacionais na área de videodifusão, como os Seminários sobre Videocolaboração do próprio CT-Vídeo, entre outros. O LabCine há alguns anos promove o desenvolvimento analítico-crítico da transmissão de mídias digitais (vídeos, visualizações científicas avançadas) por meio de redes fotônicas avançadas em ultra alta velocidade. O NTC Telessaúde Brasil redes, por sua vez, possui diversos experimentos na área que estudam o uso intensivo das tecnologias da informação e da comunicação no campo do vídeo.

5.2. Videocolaboração e transmissão (*streaming*) de vídeos em UHD



Figura 5.1: exemplo de visualização de conteúdos de vídeos em uma HiperWall pertencente ao grupo de estudos do software e *cultural analytics*. O sistema está localizado na Universidade da Califórnia, San Diego (UCSD), no CALIT2. Foto: Cicero I. da Silva.

Com o evento da transmissão de vídeos por meio de plataformas digitais, os procedimentos inventivos estão agora relacionados ao universo computacional, em uma cadeia de produção bastante extensa e complexa que inclui a visualização científica avançada: o que pode também ser pensado como uma espécie de retorno ao fotográfico, devido a sua interface estática, porém com capacidade de interações de outras mídias. No campo das narrativas, as tecnologias digitais de alta definição têm possibilitado a criação de visualizações cada vez mais sofisticadas. O exemplo disso são as imagens de ultra alta definição (*Ultra high definition*, ou UHD) criadas em centros de supercomputação para laboratórios de pesquisa, tais como o as produzidas pelo EVL (*Electronic Visualization Laboratory*), em Chicago. A capacidade de geração de imagens permite visualizar

fenômenos tanto em macro quanto em micro escalas, como as imagens geradas de proteínas ou bactérias em dimensões muito pequenas. Os cientistas têm buscado o apoio de artistas do campo das artes cinemáticas para auxiliá-los na produção de imagens, criando narrativas criativas para as imagens geradas nos laboratórios. Este é um exemplo que envolve as tecnologias 4k, 8K e acima, e também as temáticas científicas, além de sua transmissão por redes acadêmicas de fibras ópticas, também conhecidas como redes fotônicas.

Nesse sentido, a visualização de transmissões de filmes digitais de altíssima definição ainda é um desafio para ser produzido, não apenas por causa da quantidade de dados envolvidos em todo o seu processo, da captação à exibição, mas também por causa da materialidade tecnológica que requer a construção de uma linguagem própria. O cientista Richard Weinberg, professor do curso de Artes Cinemáticas da *University of Southern California* (USC), repetindo seus antecessores, produziu um filme científico chamado *MicrOrganisms* em 2010, usando uma câmera 4k RedOne acoplada a um microscópio para visualizar micro-seres (diatomáceas, hydras e amebas) de gotas de água de um lago em Los Angeles (Figuras 02 e 03).

Além do aumento significativo de resolução, esta imagem, com qualidade cinematográfica pode ser transmitida on-line em redes fotônicas de super alta velocidade, levando adiante o desenvolvimento científico que agora pode ser compartilhado com outros pesquisadores.



Figura 5.2 – Câmera Red One acoplada ao microscópio.



Figura 5.3: Filme MicroOrganims de R. Weinberg (2010). Repositório do Cinegrid Exchange. Link: <https://www.youtube.com/watch?v=1nFk-TPFI3U>

Acrescenta-se aí a possibilidade de gravar o evento e editá-lo posteriormente. De acordo com de Almeida:

A imagem de filmes 4k possui mais de oito milhões de pixels por quadro e tem sido largamente difundida. Em 2008, a resolução 4k foi estabelecida como a imagem “padrão” do cinema digital recomendada pela DCI (Digital Cinema Initiatives), uma associação dos sete estúdios principais de Hollywood. 4K refere-se ao número de pixels horizontais, 4.096 multiplicado por 2.160 pixels verticais, gerando 8.847.360 pixels. Trata-se de uma imagem quatro vezes mais precisa do que a HD, 24 vezes mais precisa que a televisão tradicional e tem sido considerada suficientemente definida para substituir o cinema 35 mm, de película. (de Almeida, 2010, p. 124)

5.2.1 As definições da imagem UHD

Algumas questões também surgiram quanto ao processo de finalização de filmes em 4K e dizem respeito ao suporte tecnológico. Por exemplo, se um usuário finaliza um filme em 4K (utilizando o padrão digital de captura da imagem) e o exporta para uma película 35mm para projeção, pelos padrões adotados pela DCI, esse usuário supostamente não estará assistindo a um filme em 4K, isso porque 4K é a “identidade” do arquivo digitalizado, projetado diretamente de um “player” que rode 4K, tanto faz se no padrão JPEG 2000, TIFF, RAW ou H.264. A questão comercial também vem sendo um dos tópicos centrais em relação ao 4K 3D (estereoscópico). Como sabemos, a indústria do audiovisual está testando várias tecnologias e empresas como Sony e Dolby buscam achar a melhor forma de negociar com as tecnologias de projeção. Por isso vemos diversos padrões de imagens (arquivos JPEG, padrão H.264, H.265, entre outros) sendo utilizados

por diversos players, tais como Zaxel, Intopix, entre outros. Contudo, é importante apontar que a estereoscopia, ao menos em 2017, ano em que este artigo foi escrito, não está recebendo muita adesão por parte dos fabricantes de televisores. É importante lembrar que todos os fabricantes comerciais, ao menos no Brasil, pararam de produzir televisores com recursos em 3D. Contudo, os cinemas continuam recebendo aportes significativos de recursos para produção e projeção de filmes em 3D, e que o retorno e a margem de lucro do investimento ainda tem sido superior ao custo de produção. Também é importante lembrar que na área da saúde, principalmente no campo da telemedicina, o uso de recursos estereoscópicos têm uma peculiar importância. Atualmente já existem no mercado, voltados à área de telemedicina e telessaúde, monitores 3D sem óculos, para centros e salas cirúrgicas. A empresa Dimenco, produtora de monitores especializados, foi capaz de criar uma tela que produz o efeito 3D (estereoscópico) sem a utilização dos inconvenientes óculos.

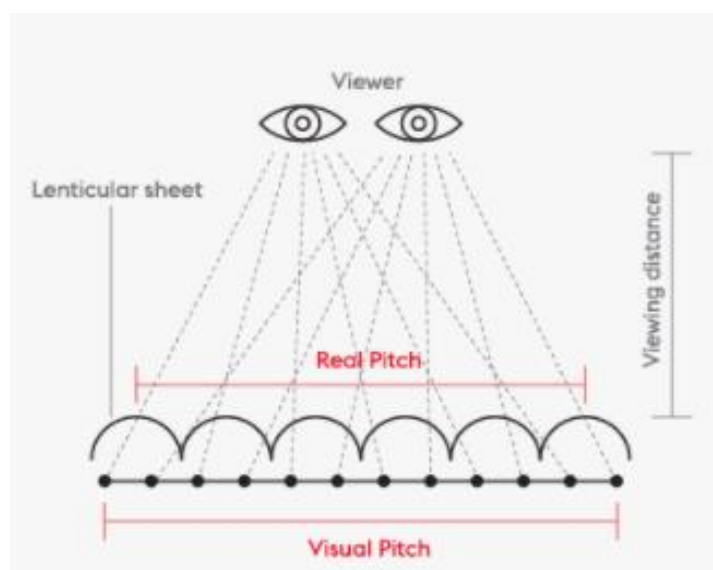


Figura 5.4: diagrama que explica a "mágica" obtida pela Dimenco para criação do efeito 3D sem a necessidade de óculos.

O efeito 3D sem óculos é descrito no endereço da empresa como tendo sido produzido a partir da

"...percepção de profundidade que é criada ligando uma sobreposição lenticular especialmente desenvolvida a uma tela LCD, de tal forma que a luz projetada é transmitida em diferentes direções. Com a introdução de interferências no que cada olho percebe como resultados da sensação de profundidade em uma superfície plana. O olho esquerdo e direito do observador vêem imagens diferentes na tela, que o cérebro funde em uma única imagem sem a necessidade de óculos 3D. A tecnologia 3D da Dimenco permite aos espectadores sentar-se em qualquer lugar e ver uma imagem 3D personalizável...e que pode ser observada de diferentes ângulos."²

² Ver em <https://www.dimenco.eu/technology.html> . (Acessado em 21 de julho de 2017).

5.2.2. Histórico de pesquisa sobre a imagem UHD

Em 2008, durante o Festival Internacional de Linguagem Eletrônica (FILE) que acontece na FIESP, o grupo de pesquisa ligado ao LabCine realizou a projeção de 12 filmes experimentais (a maior parte animações computacionais) com um servidor local, no caso um Zaxel. No ano de 2009, foi realizada a primeira transmissão de um filme em 4K para Japão e Estados Unidos utilizando servidores Zaxel e diversos routers fornecidos pela NTT. O experimento foi o primeiro que realizou a *première* de um filme em três países ao mesmo tempo utilizando redes de alta performance (1 Gbps para cima)³.

Em 2010 surgiu um outro fato importante relacionado à definição de imagem em 4K: o Youtube anunciou que a partir daquele ano proveriam suporte para filmes 4K. Esse fato foi, à época, bastante comentado e comemorado, pois poderia popularizar o termo e fazer com que vídeos de ultradefinição pudessem ser veiculados pelas redes comuns, o que já é esperado. Além disso, foi um fato importante para a indústria do audiovisual, principalmente no Brasil, que em muitos circuitos considerava o fato de você apertar um botão no Brasil e rodar um filme no Japão algo do campo da ficção científica, ou seja, impossível e não interessante do ponto de vista comercial. Parece bastante óbvio que essas crenças tiveram de ser revistas diante do avanço das redes.

Contudo existem alguns poréns que não foram ser levados em consideração. Se você observasse no site do Youtube, eles ofereciam primeiramente um upload de 2Gb com 10 minutos, e se você quisesse fazer um upload de algo maior, eles pediam para você enviar via um app java e não especificavam o tamanho, mas podemos supor que não deviam ser muitos Gigas e que, se você quisesse subir um volume um pouco maior do que dez minutos em 4K você precisaria, com uma conexão de 30 Mb, no mínimo de uma semana. Ou seja, um filme de dez minutos em 4K (com 8 milhões de pixels/frame) a 120 fps sem compressão (com a quase totalidade de qualidade, que é o que diferencia mesmo o 4K de um 2K ou Full HD) gera em média um arquivo em torno de 2Tb. Se quisermos ir um pouco além, vamos pegar como exemplo um filme de 60 minutos, como o filme projetado e enviado em 2009 durante o FILE, o Enquanto a Noite não Chega de Beto Souza e Renato Falcão, teremos algo em torno de 6Tb a 8Tb. Agora o que importa é: o Youtube faz 4Kp (progressivo) ou 4ki (entrelaçado)? Na época tivemos acesso à entrevista do engenheiro do Youtube, Ramesh Sarukkai, na qual essas perguntas não foram respondidas e não nos informaram o quanto eles estavam entregando como “4K” real, pois sabemos que para fazer um streaming de 8Tb em 60 minutos ainda levaremos um bom tempo. Ainda ficaram muitas outras questões em aberto em relação ao 4K no Youtube que talvez eles esclareçam no futuro, pois até os vídeos de demonstração que eles subiram são muito comprimidos e a definição em um projetor 4K da Sony (o SXR, por exemplo) não segurava, em termos de pixelização, todos os detalhes.

O que o Youtube provocou à época com essa “novidade”, ao dar “suporte” para filmes em 4K no meio que trata de vídeos de ultradefinição, foi a necessidade de se explicar uma coisa importante e que muda um pouco o que se falava no senso comum: um vídeo 4K não é somente um vídeo no formato 4096x2304, mas sim um vídeo não comprimido nesse formato, com um número de frames por segundo específico, em um padrão de cores específico e de preferência não entrelaçado. É possível fazer um vídeo de

³ Ver em <http://ucsdnews.ucsd.edu/archive/newsrel/science/08-09Film4K.asp> . (Acessado em 09/07/2017).

10 minutos em “4K” com 100Mb, já que a denominação 4K refere-se apenas a quantidade de linhas horizontais.

Agora, só isso é 4K? Nesse sentido, entendemos que o Youtube percebeu que era hora e que o 4K realmente seria algo importante, portanto quis sair na frente e jogar algo para cinco, seis anos, no mínimo.

As discussões sobre padrões de 4k também foram motivo de muitos debates, por exemplo no evento Cinegrid de 2007 e 2008 foram quase que tópicos centrais. O que era 4k e o que não era? As câmeras Reds, por meio de seus representantes, eram atacadas ferozmente pelo pessoal da Sony. Na ocasião Nori Suzuki, da Zaxel Systems, demonstrou por meio de fórmulas matemáticas de composição de cor que “provou” que a Red não filmava 4:4:4 (quatro quadrantes) mas sim fazia uma “aplicação” matemática que não capturava a totalidade.⁴ Isso pode ser visto nos vídeos do FILE de 2008, na FIESP, no CineGrid 2008, na mesa coordenada por Naohisa Ohta e nos artigos da Future Generation Computer Systems publicados posteriormente a esses Cinegrid's. Nesse sentido, diversos artigos explicam e se retrucam, mas valem como discussão sobre a distribuição do 4k e o futuro da imagem UHD nas suas mais variadas potencialidades e aplicações.

5.3. Metadados e metodologias para recursos audiovisuais

Os sistemas de distribuição, exibição, compartilhamento, armazenamento e produção de imagens em movimento no campo do audiovisual tem passado por uma transformação frente aos desafios impostos pelos processos de digitalização. Um conteúdo audiovisual, para ser validado no campo da pesquisa experimental e poder ser utilizado como referência científica na área da saúde, por exemplo, necessita passar por um processo de avaliação, catalogação e indexação referenciada para ser localizável em uma vasta base de dados de conteúdos de vídeos. Nesse sentido, existem inúmeros modelos e padrões para se realizar essa catalogação, armazenamento e recuperação desses dados. Parte desse conjunto de dados referencial é denominado de "metadado". Bases de dados correlacionais associadas aos objetos "vídeos" para fins de criação de catalogações vastas (*big data*) vêm sendo desenvolvidos no Brasil, criando um paralelo crítico analítico com os modelos já existentes na União Européia, Japão e Estados Unidos, propondo um paradigma aberto e universal para a leitura, recuperação e arquivamento dos registros audiovisuais, o que torna os objetos audiovisuais importantes recursos didáticos, de pesquisa e desenvolvimento, como é o caso de recursos produzidos na área da saúde.

A padronização dos conteúdos audiovisuais de super alta resolução em formato digital tem demandado uma série de estudos e análises por parte dos setores responsáveis pela criação, distribuição e principalmente fomento dessas obras. Diversos setores do campo do audiovisual têm se deparado nos últimos meses com uma acelerada necessidade de se pensar em alguns padrões e sistemas de leitura, arquivamento e distribuição dos conteúdos armazenados em sistemas digitalizados, sem ainda obter um padrão estável para os conteúdos que vêm sendo produzidos nesses formatos. Além disso, com a possibilidade de integração dos conteúdos armazenados de forma digital em universidades e centros de pesquisa em todo mundo, por meio da conexão desses espaços às redes de comunicação fotônicas de alta velocidade, existe pela primeira vez a possibilidade real de distribuição dos conteúdos para outras localidades que estão com

⁴ Ver vídeo de Nori Suzuki sobre as câmeras Red: <https://www.youtube.com/watch?v=EIJmGEUrzxc&feature=youtu.be>

essas redes de 1 a 180 Gbps já ativas em suas salas e centros cirúrgicos em diversos hospitais universitários ou mesmo unidades de saúde. Esse cenário é ainda recente e talvez fique estável em 20 ou 30 anos.

A partir dessas possibilidades iniciadas pela digitalização, já estão em formação, em vários países, comitês de especialistas no campo do audiovisual (cinema, vídeo e áudio) que vêm discutindo possíveis cenários para esse momento de migração dos recursos audiovisuais no campo da pesquisa e desenvolvimento em saúde. Na União Européia foi nomeada uma comissão especial, que se reúne desde 2008, com o objetivo de pensar a migração dos conteúdos audiovisuais para o formato digital. A Comissão Européia para Políticas de mídia e audiovisual (Comissão de especialistas para cinema) foi formada pelo parlamento europeu tendo em mente toda a diversidade da produção audiovisual trazida pela democratização que as redes e os sistemas digitais têm operado no campo da imagem em movimento.

Outros grupos também têm se dedicado a essas discussões, tais como a associação CineGrid, com base na Califórnia (EUA), que desde 2006 tem se dedicado a analisar modelos de produção, distribuição e acesso a conteúdos de super alta definição (4K) ligados em redes de computação de alta velocidade (também conhecidas como redes fotônicas baseadas em fibra óptica). Outros países que estudam a padronização de seu patrimônio audiovisual são Argentina, Chile, Japão e Coréia, que já buscam um diálogo com os sistemas para promover uma possível interoperabilidade entre os países e suas bases de dados audiovisuais, criando pela primeira vez um modelo de visualização de dados audiovisuais globais em rede. Contudo, ainda é preciso quebrar diversas barreiras técnicas e conceituais para se integrar às bases de dados digitalizadas e poder ter, de fato, um modelo de acesso mais aberto e democrático, que não seja meramente comercial e que não seja pautado somente nos interesses de grandes distribuidores de conteúdos audiovisuais, como o modelo hoje vigente e que favorece somente poucos países na que tange a distribuição e acesso. Esses modelos podem servir também para se pensar nos processos de arquivamento dos conteúdos (filmes e vídeos) gerados por meio de incentivos públicos ou mesmo para promover o acesso às bases de dados audiovisuais digitalizadas que estão em domínio público em diversas universidades em vários países do mundo. Os metadados levam em consideração, por exemplo, os formatos de arquivos digitais, os seus recursos de leitura e gravação, os padrões de acesso, envio, registro, direitos autorais e segurança. Esses itens ainda não são listados por nenhum órgão de padronização e ainda não existem normas elaboradas para esses processos.

A padronização aqui sugerida leva em consideração os modelos teóricos de código aberto e livre, portanto, pode ser lida como um “repositório de mídias digitais 4K universal”, no qual poderão ser depositados “todos e quaisquer” padrões de mídia UHD (4k, 8k ou acima). Isso quer dizer que a sugestão que estamos elaborando neste artigo parte do princípio de que é possível criar um “repositório universal de mídias audiovisuais em UHD”, sem prejuízo de qualidade e nem restrições quanto aos direitos e formatos. É óbvio que já existem interesses por parte dos grandes distribuidores científicos (Elsevier, SAGE, entre outros) de somente arquivar em seus repositórios determinados padrões de softwares e formatos de finalização. Isso é até esperado por parte das empresas que lidam com a produção científica e que querem manter as restrições que antes eram impostas ao formato de distribuição física (CDs, DVDs, entre outros). Contudo, essas restrições perdem completamente o sentido quando os formatos de armazenamento dos conteúdos passa a ser digital, portanto, acessível e de rápida distribuição e circulação. Não estamos dizendo que os repositórios devam desconsiderar os direitos de exploração e circulação

dos conteúdos audiovisuais captados em procedimentos de telemedicina, o que propomos é, por outro lado, a possibilidade de se armazenar qualquer formato (JPEG 2000, TIFF, RAW, MPEG4, entre outros), e que esse sistema consiga catalogar e sistematizar os padrões de acesso e leitura e permitir que pesquisadores de todo o mundo possam ter acesso à esses registros audiovisuais de maneira aberta, rápida e fácil. Portanto, a criação de uma metodologia de "metadados" para conteúdos audiovisuais produzidos a partir dos centros de pesquisa em telessaúde e telemedicina no país será também um importante passo na educação permanente de profissionais da saúde nas unidades básicas de atenção primária à saúde, pois permitirá que esses profissionais possam vir a ter acesso a esses recursos, em qualquer lugar, a qualquer hora e por meio de qualquer equipamento conectado a uma rede de dados.

5.3.1. O uso de recursos colaborativos audiovisuais em saúde

Diversas aplicações têm sido cada vez mais utilizadas no campo da saúde no que tange o uso de recursos audiovisuais para atividades fins, tais como telemedicina e telessaúde. Uma das ações que tem adotado esse procedimento, de maneira muito bem sucedida no Brasil é a Rede Universitária de Telemedicina (RUTE), fomentada pela Rede Nacional de Ensino e Pesquisa (RNP). A RUTE funciona por meio das SIG (Special Interest Groups), que se reúnem, por meio de plataformas de colaboração audiovisual on-line (normalmente a RNP provê acesso por meio da ferramenta Adobe Connect e, mais recentemente, por meio da ferramenta MConf).

A colaboração na área da saúde têm possibilitado inúmeros avanços nas pesquisas que necessitam intensa troca de dados. A RUTE tem demonstrado que a videocolaboração é uma ferramenta importante e fundamental para o avanço da ciência, e que a colaboração por meio do uso de recursos de vídeo permite uma troca mais intensa de informações e protocolos, que de certa forma talvez levassem mais tempo até serem formalizados por meio de publicações solipsistas que talvez redundassem nas mesmas conclusões. Nesse sentido, a videocolaboração permite um avanço na formação dos cientistas e dos pesquisadores na área da saúde.⁵

5.4. Videocolaboração na Atenção Básica

5.4.1. Primórdios

O apoio às ações em saúde por meio da utilização de recursos tecnológicos audiovisuais encontra local na história desde os anos 1950. No Brasil, a Unifesp é uma das pioneiras no uso de tecnologias comunicacionais da informação na área da saúde, com foco

⁵ Para um maior detalhamento das ações de videocolaboração por meio da RUTE, ver o trabalho de pesquisa de Thiago Delevidove de Lima Verde Brito intitulado Análise da colaboração nos Grupos de Interesse Especial (SIG) da Rede Universitária de Telemedicina (RUTE). Link: <http://repositorio.unifesp.br/handle/11600/41275> (Acessado em 22 de julho de 2017).

intenso na telemedicina. A Escola Paulista de Medicina abrigou um dos primeiros departamento de Informática em Saúde do país, cuja fundação data de meados de 1985.



Figura 5.5: imagem retratando o uso de uma "teleconsultoria" por meio de recursos audiovisuais

5.4.1.2. Estudo de caso: o projeto Favela (DIS/Unifesp)

Entre 1998 e 1999 o Departamento de Informática em Saúde da Escola Paulista de Medicina/Unifesp desenvolveu uma ação com foco na atenção primária em saúde para comunidades carentes e de alta vulnerabilidade na região da Vila Mariana, no município de São Paulo. A ação foi denominada de "Projeto Favela" (BARSOTTINI et al., 2000) e tinha como objetivo: a) atender uma comunidade de alto risco; b) ampliar as ações e competências dos profissionais de saúde que atuavam na prevenção, solução e gestão dos problemas; c) investigar a utilização de sistemas de comunicação simples e aplicações de baixo custo; d) adequar a tecnologia ao tipo de serviço de saúde oferecido; e) avaliar seu impacto na assistência primária à saúde; f) o cadastramento e Prontuário Eletrônico na WEB; g) criar uma rotina de agendamento e encaminhamento; h) prover a consulta aos resultados de exames por meio do uso de recursos digitais de comunicação; i) a supervisão

de especialistas e segunda opinião; j) atendimento primário à comunidade carente; k) educação dos profissionais envolvidos; e l) a análise e acompanhamento de informações.

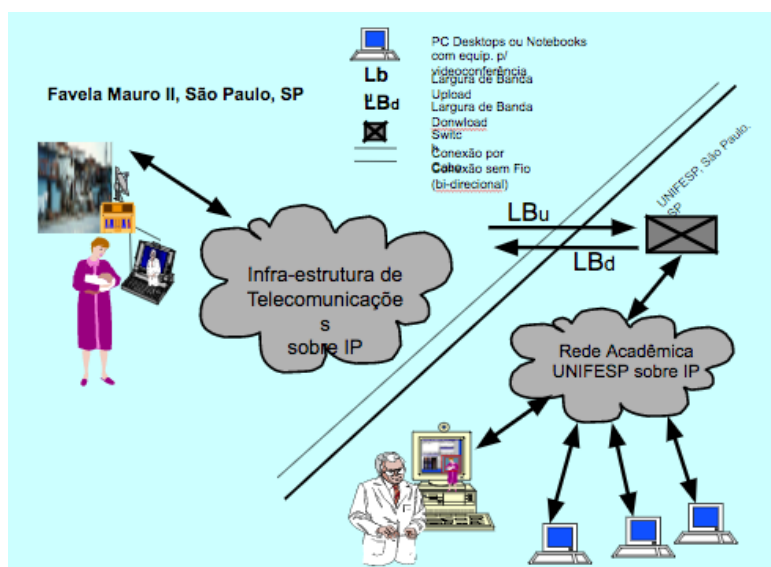


Figura 5.6: diagrama inicial das ações do "projeto Favela"

Um dos recursos tecnológicos utilizados à época foi o software de comunicação NetMeeting, que permitia a transmissão de conteúdo audiovisual, mesmo que em baixa definição e qualidade, mas que já demonstrava a potencialidade do uso de ferramentas de vídeo nos primórdios da atenção primária em saúde mediada por recursos comunicacionais da informação.



Figuras 5.7 e 5.8: detalhe do software NetMeeting com imagem sendo transmitida da comunidade para as dependências da Unifesp. No detalhe, a Profa. Dra. Claudia Nova atendendo uma família do projeto.

5.4.2 O UHD (ultra alta definição) na Educação Permanente em Saúde

Recentemente o uso de recursos de ultra alta definição tem ocupado parte das pesquisas no campo da telemedicina e telessaúde. Entende-se que o aumento significativo da qualidade e definição da imagem tem potencializado tomadas de decisão mais seguras por meio dos profissionais de saúde, que podem optar por visualizar, de forma mais ágil e com uma qualidade muito superior, detalhes de ferimentos, ou mesmo exames de

imagem que permitem que o profissional consiga observar detalhes que, com equipamentos mais simples de captação de imagens, antes não seriam observáveis por meio da imagem captada. O uso de câmeras de vídeo com definição 4K, por exemplo, em clínicas, salas de cirurgia, entre outros espaços de saúde, têm permitido que médicos, enfermeiros e técnicos de enfermagem, possam observar, captar e até mesmo transmitir imagens para outras localidades, com o intuito de trocar informações profissionais acerca de casos específicos. A Educação Permanente em saúde, por exemplo, tem sido uma das áreas que mais utilizam esses recursos de videocolaboração, pois cada vez mais os centros de formação têm tido a oportunidade de realizar reuniões de formação, capacitação, de trabalho, de gestão e de trocas de informações, por meio do uso de recursos tecnológicos baseados em videocolaboração, como com as ferramentas Adobe Connect e, mais recentemente, com o uso do sistema MConf. Algumas unidades que possuem melhores condições de conectividade, já estão fazendo uso de ferramentas colaborativas de compartilhamento de imagens, como é o caso do sistema SAGE2 (Scalable Amplified Group Environment)⁶, que permite o compartilhamento de telas em tempo real em definições como 4k e 8k. A tecnologia SAGE2 já está disponibilizada em alguns centros no Brasil, por meio do apoio da Rede Nacional de Ensino e Pesquisa (RNP), que disponibiliza tanto o acesso ao recurso tecnológico do software (instalação, *setup* etc.) quanto, em alguns casos, acesso também aos recursos de hardware, no caso aos Optiportais (conjunto de 4, 8 ou 12 telas) que permitem a visualização de dados de alta escala e definição.

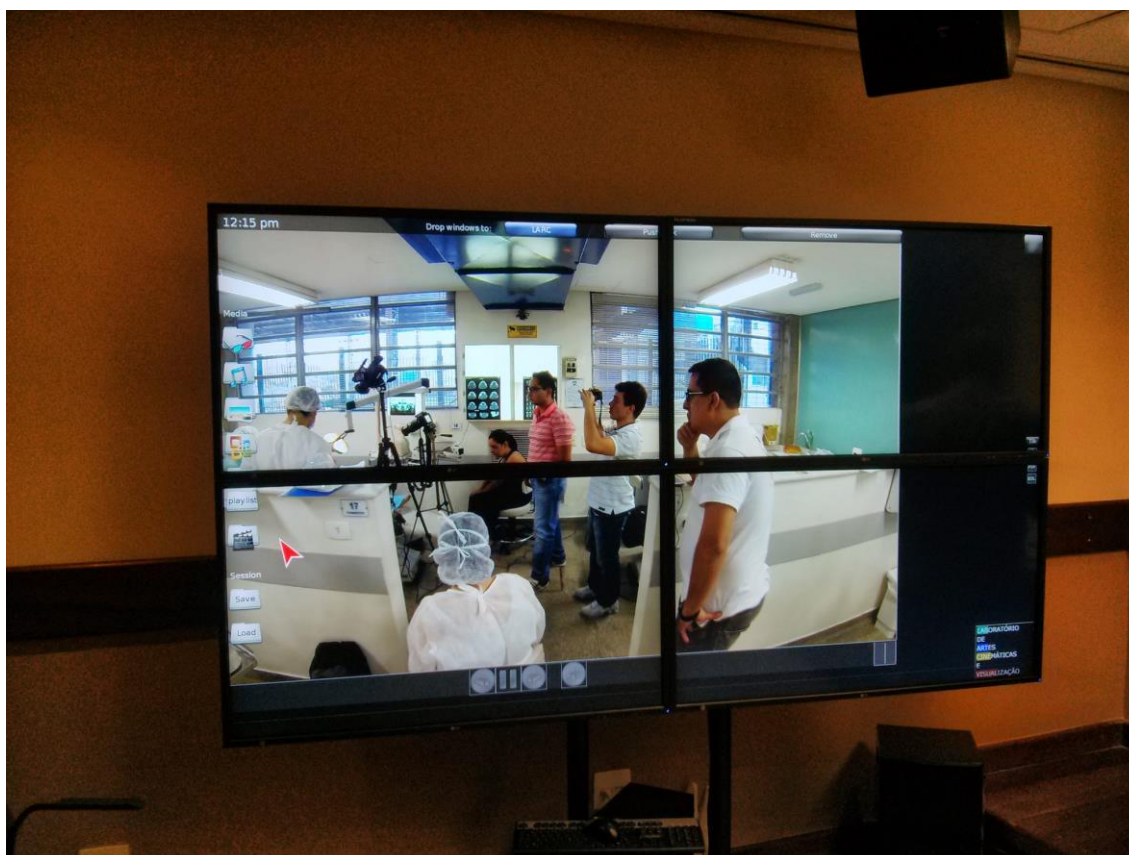


Figura 5.9: imagem de um Optiportal transmitindo uma cirurgia bucal em tempo real em 4k.

⁶ Ver mais em <http://sage2.sagecommons.org/>. (Acesso em 22 de julho de 2017).

5.4.3 Recursos tecnológicos audiovisuais como Sistema de Apoio à Decisão

Os sistemas de apoio à decisão permitem que profissionais de saúde possam ter mais dados à sua disposição para a formação de uma hipótese diagnóstica. Nesse sentido, o uso de recursos tecnológicos têm possibilitado aos profissionais acesso à base de dados científicas, por exemplo, ou mesmo consultas simples a sistemas de dados sobre protocolos da Atenção Básica, que fornecem informações específicas sobre casos recorrentes, ou que permitem que o profissional possa ter acesso a outros colegas e troque informações clínicas preciosas para sua tomada de decisão. Contudo, até muito recentemente, a maior parte dos sistemas de apoio à decisão, por falta de recursos de infraestrutura tecnológica de redes (internet cabeada), entre outros, não suportava o tráfego de dados de vídeos ou imagens em ultra alta definição. As redes não trafegavam nem mesmo conteúdos de vídeo por meio dos sistemas comerciais de comunicação, tais como Skype ou Hangout, o que limitava os sistemas de apoio somente ao texto escrito e, em alguns casos, ao envio e recebimento de imagens de baixa ou média qualidade, impossibilitando, muitas vezes, uma análise mais detalhada de uma foto ou mesmo vídeo.

Desde a implantação de recursos de transmissão de dados com maior qualidade, tem sido possível a utilização de recursos tecnológicos de videocolaboração em unidades de saúde até então desprovidas de acesso à internet. Essa conectividade, embora ainda recente e não muito robusta, já permite que algumas Unidades Básicas de Saúde (UBS's), por exemplo, tenham acesso a ambientes de videocolaboração para realização de Teleassistências (teleconsultorias síncronas). Uma teleconsultoria síncrona é a utilização de um sistema de videocolaboração (Adobe Connect ou Mconf) para a troca, por meio de uma conversa em vídeo, de informações acerca de casos clínicos, processos de trabalho ou questões relacionadas à regulação, além de protocolos da Atenção Básica, entre outras tantas questões que podem ser trocadas por meio de uma teleconsultoria síncrona por meio do recurso audiovisual. No caso do Núcleo Técnico-científico Telessaúde Brasil Redes da Unifesp, existe à disposição das equipes de teleconsultores em saúde, salas de videocolaboração providas pela Diretoria de Tecnologia da Informação (DTI) para uso específico de webpalestras e capacitações dos profissionais em saúde. Essas salas de videocolaboração hoje utilizam o sistema Adobe Connect. Existem também salas de videoconferência, hoje equipadas com sistemas da Cisco.



Figura 5.10: equipes dos Núcleos de Telessaúde da Prefeitura de São Paulo e da Unifesp durante capacitação de equipe de profissionais de saúde em uma UBS para o Programa Telessaúde Brasil Redes.

É importante observar que o uso de recursos audiovisuais poderá auxiliar no apoio às ações em saúde em áreas de difícil acesso e de alta vulnerabilidade, isso porque profissionais de saúde das mais variadas áreas poderão ter acesso a recursos de apoio à decisão, como teleconsultas com profissionais especializados em Atenção Básica, como médicos de família, entre outros, que poderão sanar dúvidas e prover segundas opiniões, por meio do uso de recursos audiovisuais, de forma mais precisa e segura.

5.5. Visão de futuro: aplicações do audiovisual na Atenção Básica

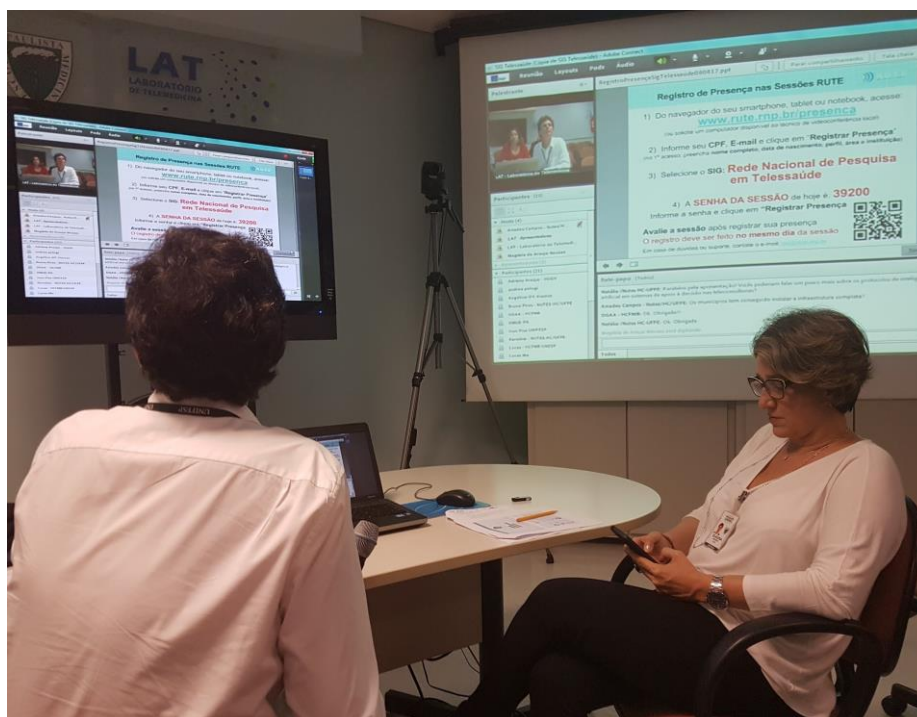


Figura 5.11: uso de ferramentas de video colaboração durante uma reunião do SIG Telessaúde da RUTE.

A partir da experiência adquirida em 2015 quando a equipe do NTC Telessaúde Brasil Redes Unifesp, em parceria com a equipe do Laboratório de Artes Cinemáticas da Universidade Mackenzie realizou a captação de uma cirurgia laser com uma câmera capaz de captar 1000 frames por segundo em 4K, o que permite que se visualizem detalhes imperceptíveis a olho nu, o grupo de pesquisa pretende ampliar ainda mais o uso de recursos audiovisuais na Atenção Básica. Nesta experiência, a equipe de captação realizou a filmagem durante um dia de trabalho no Departamento de Oftalmologia da Unifesp. A captação da cirurgia com essa câmera é bastante delicada, pois, além de toda a preparação do ambiente, em relação a segurança do paciente, entre outros, o próprio equipamento demanda uma capacidade computacional associada bastante significativa.

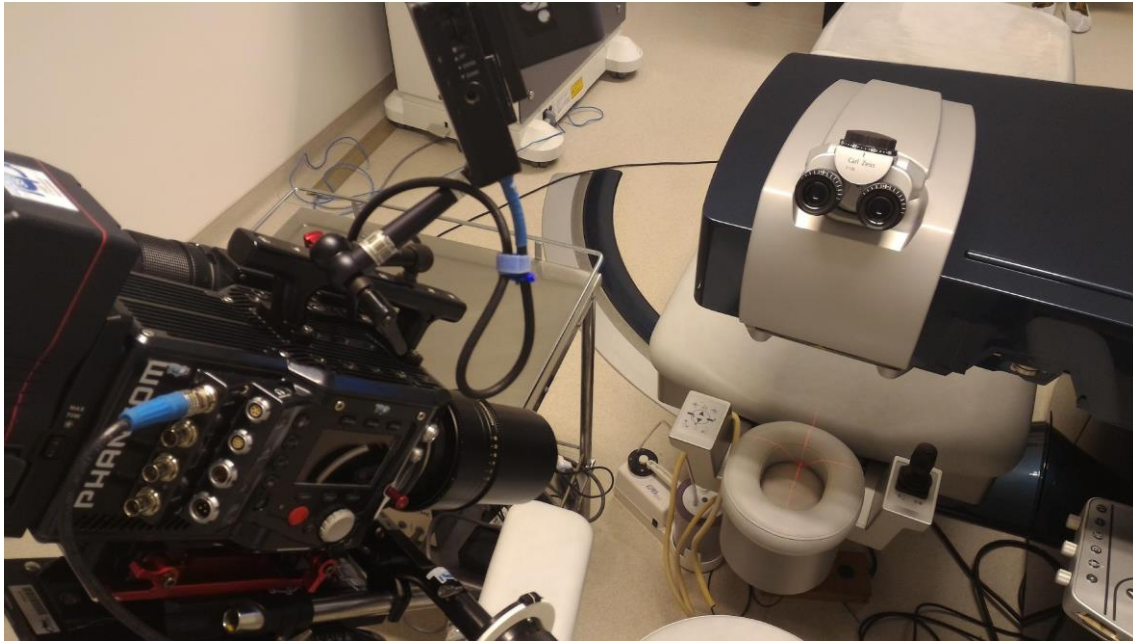


Figura 5.12:: câmera com recurso de captação de 1000 fps em 4K posicionada no centro cirúrgico do Depto. de Oftalmologia da Unifesp. (Foto: Cicero da Silva)

Para se ter uma ideia do volume de dados, a câmera, quando captando a 1000 fps em 4K, gera 3.5 Terabytes de dados a cada 3 minutos. Isso significa que a cada 4 minutos é necessário descarregar os dados dos cartões, o que, com o uso de computadores sofisticados, com taxas de transferências altas e capacidade de memória bastante avançada, a transferência durou cerca de 8 minutos. Além disso, a quantidade de *hard drives* externos para armazenamento dos dados também foi significativa. A captação foi bem-sucedida e foi possível visualizar a ação do laser de maneira precisa. A equipe que realizou a captação da cirurgia com a câmera pretende realizar a transmissão de uma cirurgia com fluxo de vídeo de 1000 frames por segundo em 4K, porém estima-se que no momento atual seria necessária uma banda de 106 Gbps de conectividade diretamente conectada à câmera. Os resultados dessa captação foram apresentados no CineGrid Workshop 2015, realizado anualmente no CALIT2 da Universidade da Califórnia, San Diego (UCSD).

5.5.1. Projetos audiovisuais futuros

As equipes que trabalham na redação deste artigo têm trabalhado no limite da geração de imagens em ultra alta definição. Acredita-se que o próximo desafio seja a captação e transmissão de cirurgias em 8K, lembrando que uma imagem com essa resolução possui mais de 33 milhões de pixels por frame e será necessário processar uma quantidade de dados bastante alta. A equipe da Unifesp já trabalha na elaboração das adaptações necessárias à captação das imagens, preparando tanto os equipamentos quanto a rede acadêmica, para suportar tal taxa de dados necessária ao fluxo de vídeo em 8K. O outro desafio, que ainda levará algo em torno de 5 anos para ser realizado, será a captação e transmissão de uma cirurgia em 8K 3D (estereoscópica), duplicando a quantidade de dados por frame (mais de 66 milhões de pixels por frame) e, por consequência, a ampliando para mais de 500 Gbps a necessidade de banda dedicada para a transmissão da imagem. Além disso, é intenção do grupo trabalhar também na integração mais intensa

dos recursos do sistema de colaboração SAGE2 com as ferramentas de videocolaboração hoje existentes, tais como Mconf, entre outras, e também ampliar a resolução das imagens hoje transmitidas por meio desses sistemas, que em poucos casos, suportam resoluções próximas ao 4k. A intenção do grupo é desenvolver cada vez mais análises sistemáticas e padronizações no campo da Atenção Primária em Saúde do uso de imagens em ultra alta definição, pois entendemos que a videocolaboração em alta definição nas Unidades Básicas de Saúde (UBS's) poderá trazer grandes benefícios para os profissionais de saúde, não só em termos de apoio à tomada de decisões, mas também na forma de ampliação da educação permanente em saúde por meio de videoaulas e processos videocolaborativos em saúde.

O grupo de pesquisa pretende, com essa investigação, desenvolver tecnologias de imagem que possam ampliar a capacidade de ensino e aprendizagem remota dos alunos que, de alguma forma, podem ter acesso às aulas, cirurgias, procedimentos médicos e sistemas de saúde, buscando a melhoria da qualidade das ações desenvolvidas pela Unifesp no campo da videocolaboração em saúde.

Referências

Almeida, J. de et. al. Passages on Brazilian scientific cinema. *Public Understanding of Science*. 2016. <https://doi.org/10.1177/0963662516683638>.

Almeida, J. de. 4K: imagens de quatro quilates, in Perissinotto, P. & Barreto, R., FILE 2008 000 000, São Paulo, IMESP, 2008.

André, T.; Almeida, J.; Silva, C. (orgs.). *CineGrid: Futuros Cinemáticos*. São Paulo, PRCEU-USP, 2016

Barsottini, C. G. N., Lopes, P. R. L., Ramos, M. P., Anção, M. S., Sigulem, D. Telemedicina: uma nova forma de levar a saúde às favelas brasileiras In: VII Congresso Brasileiro de Informática em Saúde, 2000, São Paulo. *Anais do VII Congresso Brasileiro de Informática em Saúde*, 2000.

Abreu B.C.F. Apresentação. *Cadernos do CHDD [Presentation. CHDD Journal]*. Centro Histórico de Documentação Diplomática [Diplomatic Documentation Historical Center] 11(20): 3–5, 2012.

Apud Canales J. Photogenic venus: The ‘cinematographic turn’ and its alternatives in nineteenth-century France. *Isis* 93(4): 585–613, 2002.

Baptista, T. Il faut voir le maître: A recent restoration of surgical films by E.L. Doyen (1859-1916), *Journal of Film Preservation* 70: 42–50, 2005.

Bazin, A. Science film: Accidental beauty. In: Bellows, A.M. and McDougall, M. (eds) *Science is Fiction: The Films of Jean Painlevé*. San Francisco, CA: Brico Press (Translated from *Le Film scientifique: Beauté du hazard* (21 October), vol. 121. Paris: L’Ecran français, pp. 144–147), 1947.

Berteme, C. and Dahlen, N. Back to the future: Early cinema and late economy of attention. An interim report about the crazy cinematographe. In: Loiperdinger M (ed.)

Early Cinema Today: The Art of Programming and Live Performance. New Barnet: John Libbey Publishing Ltd., 2012.

Braile, D.M. & Godoy, M.F. *Revista Brasileira de Cirurgia Vascular*. Brazilian Journal of Vascular Surgery 27(1–3), 2012.

Canales, J. *A Tenth of a Second: A History*. Chicago, IL: The University of Chicago Press, 2009.

Catani, A.M. *Anhemi e a crítica de cinema—1950-1962: Estudos de cinema da Socine [Anhemi and cinema critique—1950–1962. Socine cinema studies]. Socine cinema studies.*]. São Paulo: Annablumme, 2000.

Christensen, T.C., Mestdagh, B., Schleich, F. and Welter, J. D2.1: Specification of Processes and Digital Formats. Frankfurt: European Film Gateway, 2012.

de Almeida, J. Cinema over photonic networks. In: Proceedings of the ISEA 2011 Istanbul Conference: Leonardo Electronic Almanac, Istanbul, 14–21 September, vol. 18, issue 4, pp. 379–391. Cambridge, MA: The MIT Press, 2011.

Didi-Huberman, G. *Invention of Hysteria. Charcot and the Photographic Iconography of the Salpêtrière*. Cambridge, MA: MIT Press, 2003.

Doyen, E-L. *Le Cinématographe et l'enseignement de la chirurgie [The cinematographer and the surgery teaching]*. *La Science française* 248(27): 145–148, 1899.

Duarte, B.J. *O Filme científico*. *Revista Filme Cultura [The scientific film. Revista Filme Cultura]*, vol. 14, issue no. 1. Rio de Janeiro: Instituto Nacional do Cinema, pp. 34–39, 1970.

Finkielman, J. *The Film Industry in Argentina: An Illustrated Cultural History*. Jefferson, NC: McFarland, 2004.

Gouyon J-B. Science and film-making. *Public Understanding of Science* 25: 17–30, 2015.

Gunning, T. The cinema of attractions: Early films, its spectator and the Avant-garde. In: Stam R and Miller T (eds) *Film and Theory: An Anthology*. Oxford: Blackwell, pp. 229–235, 2000.

Janssen, J. *Ouvres Complètes I. Oeuvres scientifiques [Complete Works I, Scientific Works]*. Paris: Société d'éditions géographiques, maritimes et coloniales, pp. 312–313, 1929-1930.

Launay, F. *The Astronomer Jules Janssen: A Globetrotter of Celestial Physics*. Paris: Observatoire de Paris, 2012.

Lefebvre, T. *La Chair et le celluloid: Le cinéma chirurgical du docteur Doyen [Flesh and celluloid: The surgical cinema of Doctor Doyen]*. Paris: Jean Doyen éditeur, pp. 140–166, 2004.

- Lefebvre, T. Les 'films-leçons' du docteur Comandon. *Cinéma et science. Cinéma scientifique et médical* [The 'film-lessons' of doctor Comandon. *Cinema and science. Scientific and Medical Cinema*]. *Alliage* 71(1): 26–34, 2012.
- Machado, A. O Cinema Científico [The Scientific Cinema]. *Significação: Revista de Cultura Audiovisual* 41(42): 15–29, 2014.
- Mackeith, R.C. ; Engel C.E. The film pamphlet. *Journal of Medical Education* 30(2): 300–302, 1955.
- Mannoni, L. *The Great Art of Light and Shadow: Archeology of the Cinema*. Exeter: University of Exeter Press, 2000.
- Mannoni, L. Archeology of cinema/pre-cinema. In: Abel R (ed.) *Encyclopedia of Early Cinema*. New York, NY: Routledge, p. 33, 2005.
- Mourão, R. O Brasil e a passagem de Vênus entre o Sol e a Terra [Brazil and the Venus Transit between the Sun and the Earth]. *Revista ECO* 21(91). Available at: <http://www.eco21.com.br/textos/textos.asp?ID=808>, 2004a (acessado em Março de 2016).
- Mourão, R. The Brazilian contribution to the observation of the transit of Venus. In: Kurtz W (ed.) *Transits of Venus (IAU C196): New Views of the Solar*. Cambridge: Cambridge University Press, pp. 154–160, 2004b.
- Okamoto, M. The Brazilian discourse about Japan via France. Immigration, identity and racial prejudice (1860-1945). Ph.D. dissertation. São Paulo: University of São Paulo, 2010.
- Oliveira, B.J. Cinema e imaginário científico: História, Ciências [Cinema and scientific imaginary. *História, Ciências*]. *Saúde* 13(2): 133–150, 2006.
- Silva, M.R.B. O filme de temática científica: Possibilidades de uma documentação histórica. *Cadernos de História da Ciência* [The scientific thematic film: Possibilities of a historical documentation. *Cadernos de História da Ciência*]. *Instituto Butantan* 3(2): 13–36, 2007.
- Smarr, L. *The Grid: Blueprint for a New Computing Infrastructure*. San Francisco, CA: Morgan Kaufmann Publishers, 1999.
- Tosi, V. Les origines du cinéma scientifique 1 [The origins of scientific cinema 1]. Paris: CNRS Broadcast, 1997.
- Weinberg, R. Producing and streaming high resolution digital movies of microscopic subjects. *The International Journal of Future Generation Computer Systems* 27(1): 906–913, 2011.
- Chen, H., Chen, Y, Finkelstein, A., Funkhouser, T., Li, K., Liu, Z., Samanta, R. Wallace, G. Data distribution strategies for high-resolution displays, *Computers & Graphics* 25 (5) (2001) 811-818.

Freeman, D., Santosa, S., Chavalier, F., Balakrishnan, R. and Singh, K. LACES: “Live Authoring through Compositing and Editing of Streaming Video”. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI’14, Toronto, (2014), pp. 1207-1216.

Herschdorfer, Nathalie (Edited by). The Thames & Hudson Dictionary of Photography. Thames & Hudson, London (2015).

Incunables du Cinéma Scientifique. Produção: ICS, CNRS AV. França (1984). Online. http://videotheque.cnrs.fr/index.php?urlaction=doc&id_doc=22&rang=3.

ISO, 9241-210:2010. “Ergonomics of human-system interaction -- Part 210: Human-centred design for interactive systems”, In: The International Organization for Standardization, 2010. http://www.iso.org/iso/catalogue_detail.htm?csnumber=52075.

Sanz, C. L. “Entre o tempo perdido e o instante: cronofotografia, ciência e temporalidade moderna”. In: Bol. Mus. Para. Emílio Goeldi. Cienc. Hum., Belém, v. 9, n. 2, p. 443-462, maio-ago. (2014). <http://www.scielo.br/pdf/bgoeldi/v9n2/a11v9n2.pdf>.

Lin, Y.; Wang, H., Lu, H. “A New Online Video Editing Tool”. In: Proceedings of the 10th WSEAS international conference on applied informatics and communications, and 3rd WSEAS international conference on Biomedical electronics and biomedical informatics. Taipei, (2010), p. 465-470.

Michaelis, A. R.. Research Films in Biology, Anthropology, Psychology, and Medicine. Academic Press Inc, New York, (1955).

Leigh, J. et al., The global lambda visualization facility: an international ultra-high definition wide-area visualization collaboratory, Future Generation Computer Systems 22 (2006).

B. Jeong, L. Renambot, R. Jagodic, R. Singh, J. Aguilera, A. Johnson, J. Leigh, High-performance dynamic graphics streaming for scalable adaptive graphics environment, in: Proceedings of SC06, Novembro de 2006.

Sigulem, D., M. Ramos, and C. Barsottini. "Sistemas de apoio à decisão em medicina. Atualização terapêutica-Manual prático de diagnóstico e tratamento." (1998).

T.A. DeFanti, J. Leigh, L. Renambot, B. Jeong, A. Verlo, L. Long, M. Brown, D.J. Sandin, V. Vishwanath, Q. Liu, M.J. Katz, P. Papadopoulos, J.P. Keefe, G.R. Hidley, G.L. Dawe, I. Kaufman, B. Glogowski, K.-U. Doerr, R. Singh, J. Girado, J.P. Schulze, F. Kuester e L. Smarr, The OptiPortal, a scalable visualization, storage, and computing interface device for the OptiPuter, in The International Journal of Future Generation Computer Systems Grid Computing: theory, methods & applications, vol. 25, issue 2, Feb. 2009, pp. 114.

ANEXO 1: biografia dos autores



Cicero I. da Silva

Professor e pesquisador da Universidade Federal de São Paulo (Unifesp). Coordena o Núcleo Telessaúde Brasil Redes Unifesp. Pesquisa mídias digitais aplicadas à educação, saúde, cultura e sociedade.



Claudia G. Novoa

Professora da Universidade Federal de São Paulo no Departamento de Informática em Saúde. Coordenadora Adjunta do Núcleo Técnico Científico Telessaúde Redes da UNIFESP.



Alberto Cebukin

Secretário de Educação a Distância da Universidade Federal de São Paulo (Unifesp). É médico e doutorando em Informática em Saúde pela Unifesp.



Jane de Almeida

Professora orientadora do Mestrado e Doutorado em Educação, Arte e História da Cultura da Universidade Mackenzie e coordenadora do Laboratório de Artes Cinemáticas (LabCine).

Capítulo

6

Interiorização da medicina utilizando um sistema de telepresença holográfico

Ana Carolina Beaklini¹, André Luis de Oliveira Fonseca¹, Eduardo Rodrigues Vale¹, Yolanda E.M. Boechat², Julio Cesar Rodrigues Dal Bello¹, Rene Pestre Filho¹, Ricardo Campanha Carrano¹, Natalia Castro Fernandes¹

¹ NETAv - Escola de Engenharia - Universidade Federal Fluminense (UFF)

² HUAP - Faculdade de Medicina - Universidade Federal Fluminense (UFF)

netav@vm.uff.br

Abstract

This chapter presents the telehealth project developed by Universidade Federal Fluminense for remote medical care using a holographic telepresence system. We developed a system that allows a board of medical experts to assist patients in remote areas of the country, where the medical service is usually precarious and deficient in specialties. The system is based on the premise that the doctor needs a comprehensive and detailed view of the patient in order to assist professionals who are in remote locations. The doctor also needs good quality communication, but without imposing discomfort on the patient due to the interaction with the image and sound equipment. The proposal is already implemented and in use. We are now expanding the project scope, counting with the support of the Brazilian Armed Forces.

Resumo

Esse capítulo apresenta o projeto de telessaúde desenvolvido pela Universidade Federal Fluminense para atendimento médico remoto utilizando um sistema de telepresença holográfico. A ideia consiste em um sistema que permita a uma junta de médicos com mais experiência auxiliar o atendimento de pacientes que estejam em áreas remotas do país, nas quais o serviço médico costuma ser precário e deficiente em diversas especialidades. O sistema se baseia na premissa de que, para auxiliar os profissionais que se encontram nos locais remotos, o médico precisa de uma visão global e detalhada do paciente, tendo

facilidade de comunicação, mas sem impor ao mesmo um desconforto na interação com os equipamentos de imagem e som. A proposta já está implementada e em uso, situando-se em uma fase de ampliação da sua abrangência, contando para isso com o apoio das Forças Armadas brasileiras.

6.1. Introdução

Os problemas na prestação de serviços médicos, em especial nos locais mais remotos do Brasil, são bem conhecidos. Faltam profissionais de um modo geral, e especialistas em particular, fazendo com que boa parte da população sofra com a ausência de um atendimento médico adequado. Uma das principais causas é a dificuldade de fixar médicos experientes em pequenas cidades com pouca infraestrutura e localizadas em locais de difícil acesso. Mesmo nas iniciativas das Forças Armadas para atender as populações carentes em locais remotos [Marinha do Brasil 2016], em geral, é clara a ausência de médicos experientes nos atendimentos.

Dado esse cenário, surgiu a ideia de suplantar essas necessidades de médicos especialistas através de um suporte remoto ao atendimento. Tal ideia foi convertida em projeto pela Universidade Federal Fluminense, através do Núcleo de Estudos de Tecnologias Avançadas (NETAv), em parceria com a Marinha do Brasil, financiado originalmente pela FAPERJ [Faperj 2017]. A ideia do projeto consistia em criar um Consultório de Saúde Virtual (CSV) por meio de telepresença e holografia [Dal Bello et al. 2012]. A proposta é que o médico, para fazer uma boa assistência remota, precisa ter a sensação de estar no mesmo local que o paciente, podendo ver integralmente desde sua face até seu aspecto geral, percebendo movimentos, reações, entre outros fatores sensíveis para a realização de um bom atendimento. Assim, a transmissão apenas da voz ou só da face do paciente, como ocorre nos sistemas de teleconferência tradicionais, não seria admissível para esse tipo de aplicação [Buvik et al. 2016, Theurer et al. 2017].

A primeira fase do projeto consistiu de uma validação da ideia, pela implantação de um Centro de Saúde Holográfico (CSH), local aonde o CSV remoto é projetado por meio de holografia. A ideia é que a imagem captada no CSV pudesse ser transmitida por qualquer tipo de meio físico, seja rede terrestre ou satélite, dependendo da infraestrutura disponível no local remoto. Além disso, os médicos especialistas atendendo remotamente deveriam se sentir confortáveis durante um dia inteiro de atendimentos, não podendo sentir efeitos como o excesso de luminosidade que seria gerado por telas usualmente utilizadas para a telepresença.

Após o sucesso do protótipo, iniciou-se a segunda fase do projeto, atualmente em andamento. Nessa fase, está sendo feita a instalação de pilotos dos CSVs em diversos locais, permitindo o atendimento assistido com profissionais especializados de diferentes áreas do Hospital Universitário da UFF. Além disso, nessa fase, busca-se o aprimoramento tecnológico pelo atendimento de outros requisitos:

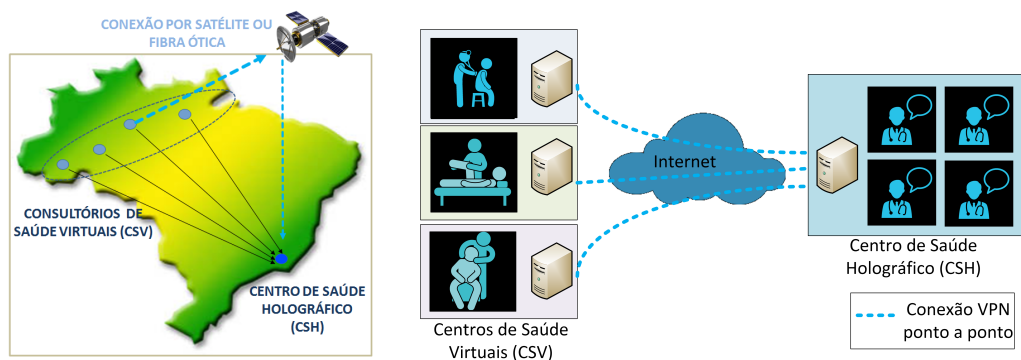
- Baixo custo de instalação e operação do CSV, pois é uma estrutura a ser instalada em diversos locais remotos, precisando de baixos custos em escala, manutenção e operação, devido às dificuldades de acesso e à ausência de técnicos no local;
- Humanização do ambiente do CSV, pois o paciente não pode se sentir oprimido

pelas técnicas de captura de som e imagem;

- A automatização do controle de som, áudio e iluminação nos centros de saúde virtual e holográfico, já que a infraestrutura deve ser utilizada exclusivamente por médicos e pacientes, sem auxílio de técnicos;
- Redução da quantidade de equipamentos do CSV, visando o aumento da disponibilidade, por meio do uso intensivo de soluções por software.
- A utilização apenas de software livre, para redução de custos viabilizando a implantação do sistema por todo o país;
- Disponibilidade de alta qualidade de imagem com baixo uso de banda passante, já que não se pode contar com infraestrutura de rede de qualidade nos locais mais remotos do Brasil;
- Criação de um ambiente de transmissão seguro e integrado com sistemas hospitalares, para que a troca de prontuários e exames seja natural durante a consulta;
- Inserção de novos tipos de câmeras médicas na captura e transmissão das imagens, para aumentar a gama de especialidades médicas que podem utilizar o sistema;
- Busca por parcerias para instalação e uso médico dos CSVs nos locais que tem necessidade no Brasil, sendo os principais parceiros as Forças Armadas.

É possível verificar que o projeto que está sendo desenvolvido e colocado em operação é bastante relevante principalmente em um país de magnitude continental como o Brasil, que tem problemas graves em saúde pública e distribuição de renda. Esse tipo de implementação possibilita que, com uma infraestrutura relativamente simples do lado do paciente, populações remotas possam passar a ter um bom atendimento médico, aumentando a qualidade de vida.

Dessa forma, esse capítulo visa à apresentação do projeto, das tecnologias desenvolvidas e dos padrões em uso, dos principais resultados obtidos até o momento e uma visão dos próximos passos, tanto em termos de tecnologia quanto de atendimento médico. Assim, esse capítulo está organizado como descrito a seguir. Na Seção 6.2, é descrita a arquitetura da solução projetada. Na Seção 6.3, são apresentados os cenários vislumbrados para uso da tecnologia desenvolvida, explicando também como ela vem sendo utilizada para consultas reais. Na sequência, na Seção 6.4, são apresentadas brevemente as tecnologias utilizadas para desenvolver o projeto e, na Seção 6.5, são apresentados os detalhes do piloto desenvolvido. Na Seção 6.6, é apresentada uma visão sobre como estão sendo as consultas médicas realizadas com o sistema, apresentando uma visão dos efeitos para o médico e para o paciente. Para finalizar, a Seção 6.7 apresenta os trabalhos em andamento para melhoria do sistema, a Seção 6.8 apresenta as conclusões desse trabalho e a Seção 6.9 apresenta a visão para o uso do sistema nos próximos anos.



(a) Exemplo de distribuição geográfica entre os CSVs e o CSH

(b) Arquitetura de conexão entre os CSVs e o CSH.

Figura 6.1: Visão geral da solução, com os centros de saúde virtuais (CSV), que são consultórios remotos, e o centro de saúde holográfico (CSH).

6.2. Arquitetura

A arquitetura do sistema é mostrada na Figura 6.1. Uma vez que o sistema visa fazer o atendimento remoto de pacientes, ele é composto de dois elementos principais, sendo eles o Centro de Saúde Virtual (CSV) e o Centro de Saúde Holográfico (CSH). O CSH é o local onde a junta médica se reúne para prestar atendimento ao profissional de saúde e paciente que estão na localidade remota. É no CSH que é feita a projeção da imagem holográfica, para que a junta médica tenha a sensação de estar no mesmo local que os pacientes. Em geral, é esperado que o CSH esteja em um grande centro urbano com bom suporte às diversas especialidades. Já o CSV é o consultório remoto, no qual diferentes especialidades podem estar sendo atendidas. A imagem e o som gerados no CSV e no CSH são transmitidos entre os dois locais por meio de uma *Virtual Private Network* (VPN), de forma a garantir que apenas quem esteja no CSV ou no CSH terá acesso ao som ou à imagem da consulta. Essa conexão pode ser feita por meio de *backbones* de fibra ótica ou por meio de conexões por satélite, no caso de o CSV estar instalado em uma localidade sem acesso de banda larga à Internet. Com isso, é possível estabelecer uma conversação com imagem entre o paciente, o profissional de saúde na localidade remota e a junta médica no centro urbano.

6.2.1. Centro de Saúde Virtual - CSV

O centro de saúde virtual (CSV) é um consultório médico tradicional com algumas modificações que permitem a geração e a captura da imagem holográfica e do som. Além disso, ele também deve ser capaz de apresentar a imagem e o som gerados pela junta médica. A Figura 6.2 mostra o CSV montado no CRASI/Hospital Universitário Antônio Pedro.

Em termos de equipamentos, o consultório remoto tem a mesa do médico e uma maca, além um computador, uma webcam HD, um microfone de ambiente, caixas de som e uma televisão para mostrar a imagem do CSH para o paciente. Com relação ao ambiente, a sala deve ter a área que aparece na imagem com paredes e piso pretos. A parede pode ser pintada ou apenas coberta com um painel preto. Isso é necessário para a produção do efeito 3D na holografia.

Além disso, existe a necessidade de uma iluminação especial, para evitar sombras no rosto e nas pernas do paciente. Essa iluminação, contudo, não pode ser incômoda para o médico ou para o paciente.

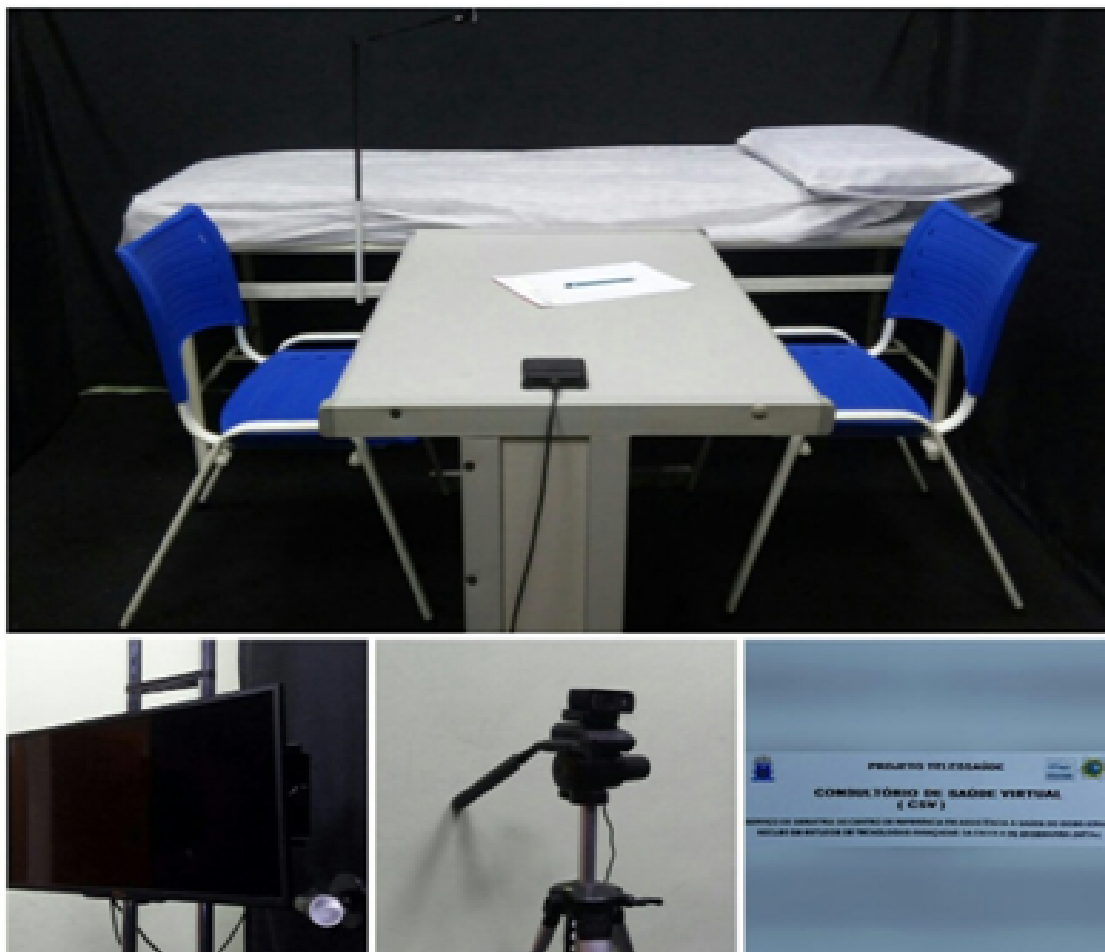


Figura 6.2: Exemplo de centro de saúde virtual, com seus equipamentos, no CRASI/Hospital Universitário Antônio Pedro (HUAP), na UFF.

Atualmente, o projeto já tem dois CSVs instalados e funcionais, um no HUAP e outro no Hospital Central do Exército (HCE), que é parceiro da UFF na fase atual do projeto. Uma imagem projetada no CSH do consultório do HCE pode ser vista na Figura 6.3.

Um dos conceitos na elaboração do CSV é que ele precisa ser uma unidade de baixo custo e de fácil manutenção e operação, pois será instalado e operado em locais remotos, onde usualmente não existem técnicos em informática disponíveis. A primeira versão do CSV contava com câmera profissional, mesa de som e caixa de som profissional, além dos demais equipamentos já citados. Além disso, ela requiritava o uso de CODECs em hardware, os quais são equipamentos caros, quando comparados a todos os demais equipamentos utilizados no ambiente. Assim, além de ser um ambiente caro, era um ambiente frágil e de difícil manutenção. Por essa razão, existiu todo um trabalho para modificar a tecnologia utilizada para obter a mesma qualidade de imagem e som, mas



Figura 6.3: Imagem projetada do centro de saúde virtual instalado no Hospital Central do Exército (HCE).

com equipamentos mais simples de usar e mais baratos. A descrição da tecnologia usada nesse novo ambiente será feita na Seção 6.4.

Os equipamentos citados fazem parte do conjunto básico para compor o CSV. Contudo, determinadas especialidades podem demandar mais alguns equipamentos específicos. Por exemplo, na dermatologia, se faz necessário o uso de imagens com alta resolução das lesões na pele. Para tanto, optou-se pelo uso de softwares implementados em celulares para obtenção dessas imagens, com controle de zoom e foco feito pelo médico à distância, como mostrado no exemplo da Figura 6.4. Com isso, se tornou possível realizar os exames utilizando um equipamento que todo médico possui, bastando apenas a instalação do software. Cabe adicionar que outros tipos de câmeras/sensores podem ser necessários, dependendo da especialidade que for ser atendida.

6.2.2. Centro de Saúde Holográfico - CSH

O CSH é uma sala de projeção holográfica, na qual uma junta médica irá interagir em tempo real com o paciente no CSV. A Figura 6.5 mostra o CSH instalado dentro das dependências da UFF. Em suma, trata-se de uma sala com paredes escuras e um palco, onde é projetada a imagem holográfica. A tecnologia usada para a produção dessa imagem é descrita na Seção 6.4.2.

O CSH, devido à produção da imagem holográfica, tem um custo mais elevado. Isso, contudo, não representa um problema, já que a previsão é a existência de múltiplos CSVs sendo atendidos em um único CSH. Nessa sala, existe um conjunto de equipamentos incluindo microfones direcionais, mesa de som, mesa de iluminação, roteador, computador, data show, além de uma tela em material transparente especial para produção da holografia. Assim, espera-se que os CSHs tenham um técnico responsável pela operação



Figura 6.4: Imagem de lesões de pele para análise pelo médico remoto. No sistema, tais imagens são projetadas no tamanho aproximado de uma pessoa, dando boa precisão no exame.

dessa sala, assim como pela conexão com os CSVs, operados remotamente. Contudo, uma vez preparada a sala, não há mais a necessidade da presença do técnico, o que é essencial para a preservação da privacidade durante a consulta.

6.3. Cenários de uso

O projeto telessaúde foi idealizado para promover a interiorização da medicina dentro do país. Dentro desse cenário, existem diversas formas de aplicação da ideia, algumas já em andamento.

6.3.1. Navios da Marinha do Brasil

Inicialmente, o projeto foi desenvolvido em parceria com a Marinha do Brasil, objetivando uso com o projeto de atendimento médico das populações ribeirinhas da Amazônia e também dos pesquisadores da Estação Antártica Comandante Ferraz, os quais só têm acesso ao serviço médico provido pela Marinha [Marinha do Brasil 2017].

A Marinha do Brasil disponibiliza três navios de assistência hospitalar na Amazônia, todos muito bem equipados com dispositivos para realizações de diversos tipos de exame. Tais navios contam com profissionais com formação em poucas especialidades,



Figura 6.5: Centro de saúde holográfico instalado no CRASI/Hospital Universitário Antônio Pedro, na UFF.

o que se mostra eficaz na maioria dos casos atendidos, porém insuficiente para atendimento a casos mais complexos. Com isso, a aplicação do sistema proposto junto com esses navios é de grande interesse, pois diferentes tipos de casos médicos são atendidos sem a presença de especialistas. O sistema permitiria uma assistência de maior qualidade à população, simplificando e dando maior segurança para o serviço dos médicos no navio.

Esses navios são operados pela Marinha com apoio do Ministério da Saúde, juntamente com o apoio de alguns governos da Região Norte. O uso do sistema holográfico dependeria apenas da disponibilidade de banda no Satélite Geoestacionário de Defesa e Comunicação Estratégica (SGDC).

6.3.2. Cidades do interior e da fronteira do Brasil

A segunda fase do projeto foi realizada em parceria com o Exército do Brasil, visando o atendimento das bases militares e populações nas fronteiras do Brasil. Nesse caso, a ideia é prover auxílio no atendimento pela mesma razão observada nos navios: em geral, são médicos militares recém-formados que servem nas fronteiras, sendo que, nessas localidades, muitos casos com grande nível de dificuldade precisam ser atendidos. Assim, o apoio de médicos especialistas dos hospitais militares de grande porte seria muito interessante para melhorar a qualidade do atendimento e para evitar remoções desnecessárias.

A conexão, nesses cenários, poderia ser feita por meio da rede militar cabeada, nas

cidades onde essa estiver disponível, por algum provedor privado de Internet ou, ainda, por satélite, caso as duas opções anteriores não sejam viáveis.

Na fase atual do projeto, já foi instalado um CSV dentro do Hospital Central do Exército - HCE, para que os médicos militares possam avaliar e aprender a usar o sistema. Consultas dermatológicas estão sendo realizadas utilizando o CSH da UFF.

6.3.3. Uso em grandes centros

Está em desenvolvimento uma iniciativa para dar assistências aos estudantes de medicina da UFF no atendimento dos postos de saúde de Niterói, no Rio de Janeiro. Embora sejam locais próximos, é de grande importância a assistência dos preceptores, que são os professores responsáveis e que tem vasta experiência nas modalidades dos atendimentos. Com isso, se garante um atendimento de maior qualidade, uma assistência mais dinâmica a uma quantidade maior de pacientes e um melhor aprendizado por parte dos alunos.

Essa iniciativa passará pela instalação de novos CSVs em alguns postos de saúde de Niterói e está planejada para ser executada em 2018.

6.4. Tecnologias utilizadas

As tecnologias utilizadas no projeto compõem um papel chave para que o atendimento ao paciente seja o mais natural possível, onde algumas características são necessárias:

- **Custo Baixo:** O primeiro desafio foi construir um sistema capaz de transmitir fluxos de áudio e vídeo com um custo baixo, se comparado aos equipamentos que estão disponíveis no mercado para realizar tal tarefa. O objetivo é realizar a transmissão de áudio e vídeo em tempo real com qualidade e segurança para que a videoconferência seja simples e natural como uma conversa do dia a dia.
- **Possibilidade de expansão:** Outro desafio foi criar um sistema capaz de interconectar diversos dispositivos para que os médicos tenham mais possibilidades de diagnosticar e interagir com o paciente para trocar informações de maneira eficaz e eficiente.
- **Usabilidade:** O uso da tecnologia deve ser o mais natural possível, sem distrair a atenção dos profissionais de saúde.
- **Segurança:** Um ponto importante é que toda informação gerada pelo sistema no momento da videoconferência seja trocada entre os médicos e o paciente com mínimo de segurança. Dados do paciente estarão trafegando em rede pública, portanto é imprescindível que o sistema esteja seguro de ameaças externas.

A maior parte dos softwares usados/desenvolvidos no projeto é livre, ou seja, além de possuir o código fonte aberto, os usuários têm a permissão de modificar e distribuir livremente. Isso reduz o custo e flexibiliza o desenvolvimento do projeto ao longo do tempo. Para compreender melhor o sistema, essa seção é organizada da seguinte maneira:

- **Áudio e Vídeo:** descreve a tecnologia usada na transmissão do áudio e vídeo dentro do sistema;

- **Holografia:** descreve a tecnologia usada para projetar a imagem de forma holográfica no sistema.
- **Segurança:** descreve a tecnologia usada para garantir a privacidade e segurança ao sistema;
- **Gerência e operação:** descreve as ferramentas usadas para gerenciar e operar o sistema;

6.4.1. Áudio e Vídeo

O melhor aproveitamento do setup desenvolvido exige um cuidadoso estudo da teoria de vídeo, garantindo encontrar a configuração ótima que permita uma transmissão robusta e confiável mesmo com um canal com banda, perda e latência variáveis. Cada etapa do processo deve ser cuidadosamente estudada: captura, codificação/decodificação e compressão/decodificação do mesmo. A escolha cuidadosa de cada parâmetro pode melhorar a qualidade do conteúdo transmitido, como qual codec de áudio e vídeo utilizar, taxa de bit, tamanho do *Group of Picture* (GOP), taxa de quadros, resolução e protocolo de transporte. Todos esses fatores impactarão na qualidade da experiência do usuário e a escolha deve ser feita sempre buscando minimizar perda de pacote e atraso na transmissão. Mais do que isso, assume-se uma restrição na banda passante disponível, dado que, em geral, os CSVs ficarão em locais remotos.

Na etapa da captura, a escolha da câmera é importante. A câmera deve atender no mínimo as configurações que se pretende ter com o codificador: quanto maior a qualidade do material capturado, melhor será a qualidade resultante na etapa de codificação.

A captura é a parte do processo responsável pela digitalização do material e duas etapas determinarão a qualidade do mesmo: amostragem e quantização [Tektronix 2009b]. A amostragem é a etapa que transformará o sinal capturado contínuo em uma sequência de amostras, tensões em diferentes intervalos de tempo, de modo que permita que o sinal analógico seja totalmente recuperado. A quantização é a etapa que define a quantidade de níveis de tensão que será utilizado para reconstruir o sinal. Para o vídeo digital em alta definição, é recomendado o uso de 10 bits (1024 níveis) [Tektronix 2009b]. O processamento de uma quantização inferior a 10 bits pode gerar artefatos e perda de resolução nos contornos da imagem.

Após a fase de captura, o conteúdo de vídeo precisará então ser preparado para o canal de transmissão pretendido. Para isso, o vídeo será codificado e comprimido, de modo que possa ser expresso em menor quantidade de dados, atendendo a uma largura de banda inferior.

Existem diferentes técnicas para codificar um vídeo e a escolha do codec deve ser feita cuidadosamente garantindo que se atinja o objetivo. A codificação pode ser feita reduzindo a quantidade de informação a ser enviada, eliminando a redundância através de mecanismos que codifiquem nos modos com ou sem perda a informação [Tektronix 2009b]. A compressão pode ser feita buscando e eliminando as repetições, comparando apenas cada quadro de vídeo (compressão intraframe ou espacial) ou buscando redundância entre quadros de vídeo (compressão interframes ou temporal) [Tektronix 2009a].

A informação considerada única é chamada de entropia do sinal e o papel do codificador é buscar separar a entropia e eliminar a redundância do mesmo [Tektronix 2009a]. Se a eliminação de redundância não reduz os dados tanto quanto necessário, alguma informação terá que ser descartada. Sistemas de compressão com perdas alcançam a redução de dados pela remoção de informação irrelevante ou de relevância menor. Para decidir o que pode ser considerado informação menos relevante, é importante que a análise seja feita no contexto da aplicação. No caso de vídeo e áudio, deve ser considerado o que os sistemas visual e auditivo dos seres humanos não são capazes de perceber [Tektronix 2009b]. Um sistema ideal conseguiria buscar toda a redundância espacial e temporal e eliminá-la, transmitindo apenas a entropia. Porém, na prática, isso causaria um longo atraso para processar essa busca em toda a redundância temporal, tornando o sistema complexo, caro e lento [Tektronix 2009a], o que não é aceitável na aplicação desse projeto, que exige um baixo atraso para garantir a interação em tempo real.

6.4.1.1. Padrões de compressão de vídeo

Durante anos, dois grupos foram os pioneiros e os principais responsáveis pelo desenvolvimento de padrões de compressão de vídeo: o MPEG (*Moving Picture Experts Group*) e a ITU (*International Technological University*) [Tektronix 2009a]. A Figura 6.6 ilustra a linha do tempo dos padrões de vídeo onde podemos verificar as contribuições dos grupos ITU-T VCEG e MPEG WG 11 [Tektronix 2009a]. Os CODECs resultantes são publicados por ambos os grupos, no grupo MPEG com o nome de MPEG-X e os publicados pelo grupo da ITU com o nome de H26X.

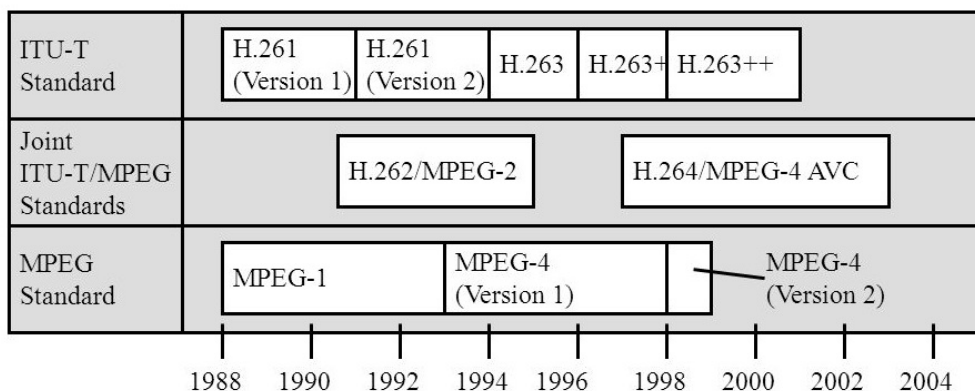


Figura 6.6: Padrões de vídeo desenvolvidos nas últimas décadas.

O padrão MPEG-4 é um algoritmo de referência, mas, por ser excessivamente complexo, foi posteriormente remodelado pela ITU e chamado de H.264/AVC. O H.264/AVC possui uma compressão considerada de excelente qualidade, razão pela qual esse padrão foi adotado nas implementações do projeto. É comum chama-lo de H.264/AVC ou H.264/MPEG-4 AVC ou ainda MPEG-4 part 10. Esse padrão trata exclusivamente do vídeo [Tektronix 2009a].

A escolha do CODEC é de grande importância, pois interfere diretamente na qualidade do vídeo e na taxa de transmissão, os quais são dois fatores sensíveis na aplicação

proposta. Seu desempenho é medido baseado em quantos bits são necessários para alcançar a qualidade visual pretendida para um dado vídeo. Essa qualidade, embora um tanto quanto subjetiva, depende da nitidez, sincronização entre áudio e vídeo, suavidade de contornos e movimentos e fidelidade das cores [Tektronix 2009a]. Na Figura 6.7, pode-se observar a comparação entre os diferentes padrões desenvolvidos com foco em vídeo SD/HD, o que torna claro a nossa escolha pelo H.264. Esse padrão é capaz de comprimir em menor taxa de bits por segundo uma maior quantidade de informação.

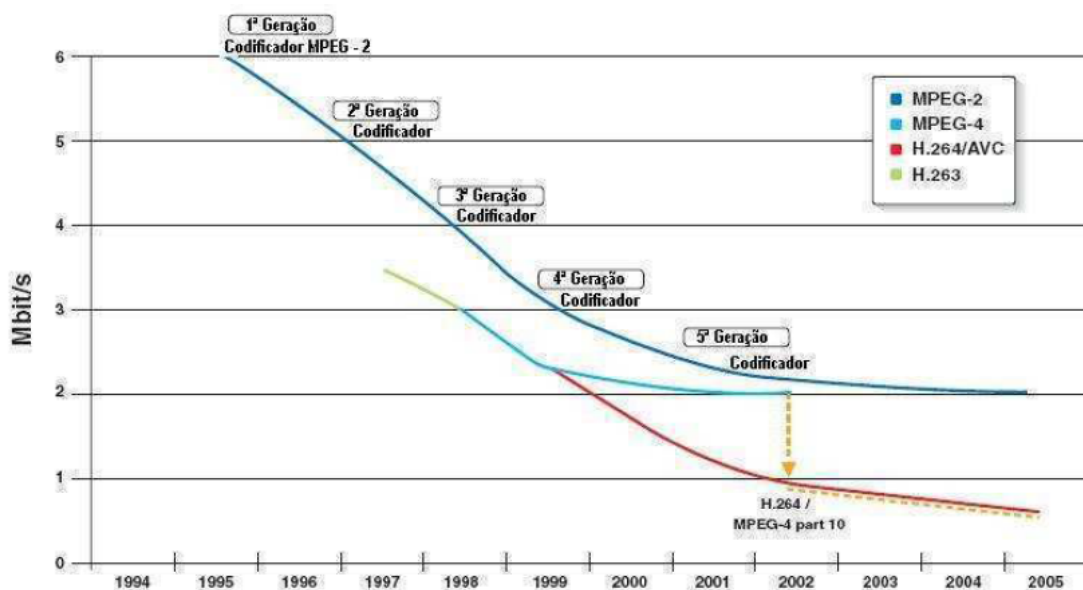


Figura 6.7: Comparativo dos padrões de vídeo ao longo dos anos.

A codificação H.264 é um algoritmo de compressão com perdas, que busca eliminar informações do vídeo que não possam ser percebidas ao olho humano. O H.264 possibilita transmitir o dobro de informação na mesma taxa quando comparado ao seu antecessor MPEG-2, por isso é o mais utilizado para transmissões em alta resolução no mundo. Dentre os mecanismos utilizados para eliminação de informação de menor relevância, está a sub amostragem de croma, devido a uma característica do olho humano que é mais sensível a informações de luminância que de crominância. É possível reduzir a amostragem de sinais de croma pela metade na horizontal quase sem prejuízo à imagem, gerando uma relevante economia na taxa de informação. O particionamento do quadro em macro blocos de informação de tamanhos variáveis é outro método de compressão muito relevante do H.264. Este método permite particionar a imagem de modo a transmitir apenas uma vez a informação relevante e, através desses macros blocos, recriá-las utilizando compressão intraframes. Na compressão interframe, utiliza-se os macro blocos como referência e adicionam-se informações de vetores de movimento para reposicionar a imagem nos *frames* subsequentes. É possível subdividir um vídeo em Grupos de Imagens (*GOP – Group of Pictures*) e, através de um *frame* de referência, utilizar mecanismos de predição e vetores bidirecionais para reproduzir as imagens daquele grupo. Nesse caso, o *frame* I (Intra) contém a maior parte da informação e do peso e os demais B (bilateral) e P (preditivo) carregam pouquíssima informação.

6.4.1.2. Processo de codificação e decodificação do sistema desenvolvido

Todo o processo de codificação e decodificação de áudio e vídeo é realizado através de uma aplicação multimídia de *streaming*, chamada GStreamer. O GStreamer é um arcabouço em software livre, licenciado sob GNU GPL, com código fonte aberto. Esse *framework* possibilita o desenvolvimento de qualquer tipo de aplicação multimídia de *streaming*, mesmo aquelas que dependem de edição de áudio e vídeo [Taymans et al. 2017]. O GStreamer é uma estrutura extremamente poderosa, versátil e modular, pois pode incorporar diversos *plugins* de forma transparente [Boulton et al. 2017]. Através de uso de *pipelines*, o GStreamer executa as aplicações de *streaming* com alta velocidade e baixa latência. O GStreamer é multiplataforma, trabalhando com a maior parte dos sistemas operacionais [GStreamer 2017]. Dadas as suas especificações e o bom desempenho nos testes realizados, o GStreamer foi adotado como ferramenta definitiva na solução desenvolvida. Outras opções como o VLC [VideoLAN 2017] e o ffmpeg [FFmpeg.org 2017] também foram testadas, mas o GStreamer se mostrou mais eficiente e estável nos testes.

Uma parte importante para este sistema é uma câmera compatível com este software. No caso, foram selecionadas *webcams*, usualmente utilizadas para videoconferências via Internet. Entretanto, essas *webcams* precisam ter algumas peculiaridades:

- capacidade de captura de vídeo em qualidade *High Definition* (HD);
- capacidade de compressão de vídeo no padrão H.264;
- ser compatível com sistema operacional Linux; e
- possuir conectividade pela porta USB.

No sistema projetado, através de um computador, se executa o software de *streaming* no Linux. O esquema é representado pela Figura 6.8, seguindo o fluxo abaixo:

- **Passo 1:** O processo se inicia com a captura do vídeo pela câmera em H.264 e do áudio pelo microfone de ambiente, ambos conectados ao computador;
- **Passo 2:** O áudio é codificado pelo software usando o codec Speex [Xiph.org 2017] e o vídeo é transcodificado para uma taxa mais baixa em H.264;
- **Passo 3:** Áudio e vídeo são injetados na Internet;
- **Passo 4:** O processo no computador remoto decodifica os dados recebidos da Internet;
- **Passo 5:** O áudio e vídeo são projetados.

Esse processo é bidirecional entre CSV e CSH, pois há a necessidade de interação direta entre as pessoas nas duas localidades. Inicialmente, optou-se pela passagem apenas do áudio na direção do CSH para o CSV, pois se acreditava que era suficiente o áudio dos médicos especialistas. Contudo, houve uma demanda por parte dos médicos para que a imagem da junta de médicos remota também fosse transmitida, para dar maior segurança ao paciente sobre com quem ele está falando.



Figura 6.8: Processo de codificação e decodificação do sistema.

6.4.2. Holografia

Este trabalho descreve como, através do uso do conceito de holografia, é possível capturar, codificar e transmitir vídeo e áudio entre um paciente em localização remota e um especialista em um hospital que possua a tecnologia necessária para reprodução de uma imagem holográfica com qualidade suficiente para uma consulta médica [Poon, T. C. 2007, Poon, T. C.; Liu, J. P. 2014]. Com o objetivo de conseguir uma imagem do paciente com a maior fidelidade possível, porém sem necessidade de uma banda passante muito elevada, a tecnologia escolhida foi aquela conhecida como *Pepper's Ghost*, que embora não gere efetivamente hologramas, produz uma imagem tridimensional do paciente. Tal tecnologia foi desenvolvida em 1962 por John Henry Pepper [Steinmeyer 2013]. Na Figura 6.9, encontra-se o modelo de reconstituição da imagem do paciente no sistema holográfico.

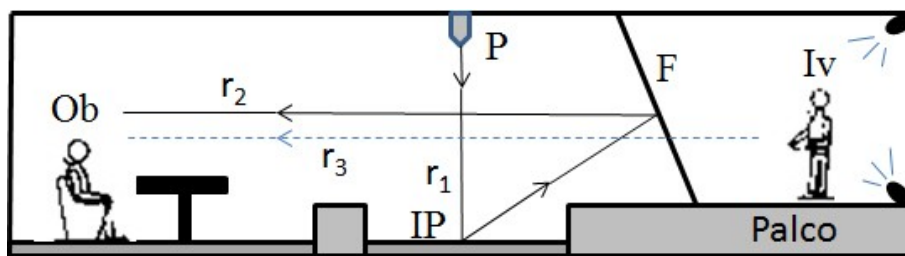


Figura 6.9: Modelo de reconstituição da imagem no Sistema holográfico.

Os componentes presentes na Figura 6.9 são:

- **Ob (Observador)** - Caracteriza-se pelo especialista que observa a imagem virtual do paciente;
- **Iv (Imagem Virtual)** - É a imagem do paciente distante vista pelo Observador;
- **P (Projetaor)** - Projeta a imagem do paciente;
- **IP (Imagem do Paciente)** - Local, denominado tecnicamente por "piscina", onde é projetada a imagem do paciente oriunda do projetor;
- **F (Foil)** - Película que reflete os raios provenientes da imagem do paciente e refrata os raios provenientes do Palco, dirigindo-os para o observador.

A ideia é simples: o projetor projeta a imagem sobre a piscina, que reflete os raios, os quais atingem o *foil*. Por ser um material transparente, os raios refletidos do *foil*

formam uma imagem virtual. Assim, através da reflexão e difração dos feixes R2 e R3, o observador verá uma imagem aparentemente tridimensional do paciente e do ambiente em que ele está inserido. Contudo, essa sensação de 3D depende que o ambiente de projeção seja escuro. Outro fator que ajuda na sensação de 3D é a aplicação de determinados feixes de luz que promovem a sensação de profundidade no palco, apesar do ambiente escuro.

No ambiente do consultório (CSV), algumas particularidades deverão ser atendidas para garantir uma boa imagem holográfica no destino. A principal delas, como já comentado, é a cor escura das paredes. Além disso, o CSV precisa ser muito iluminado, para diminuir o impacto das sombras na projeção em um local escuro como o CSH. Assim, iluminação é uma preocupação especial nesse tipo de ambiente.

Outro ponto relevante é que a sensação de 3D está vinculada ao posicionamento da imagem projetada sobre o piso do palco. Uma imagem projetada que aproveita a largura do ambiente, mas não a profundidade, tende a ser mais facilmente projetada no local correto. Por exemplo, se se tem duas pessoas em pé, uma ao lado da outra, é simples ajustar a imagem para que ambos fiquem com os pés no chão do palco. Contudo, se existem duas pessoas, uma atrás da outra, pode acontecer de os pés da pessoa de trás não tocarem o chão do palco. No caso do CSV, existem objetos que são projetados e que estão em profundidades diferentes, como mostrado na Figura 6.10: a maca, que fica no fundo da sala, e a mesa, que fica de frente para a câmera. Dependendo do tamanho do consultório e da distância entre a câmera e a mesa, pode não ser possível acertar a projeção da maca sobre o palco, criando a sensação de um objeto voando.

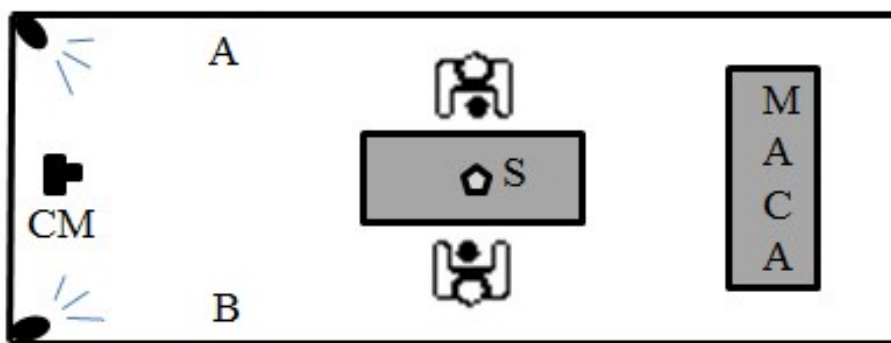


Figura 6.10: Modelo do consultório remoto.

Na Figura 6.10, observam-se todos os elementos do ambiente do consultório (CSV):

- **CM (Câmera)** - Responsável pela captura das imagens do profissional de saúde e do paciente;
- **Caixas de som (A e B)**: emissor de áudio (*full duplex*) recebido pela sala de projeção (Figura 6.9).
- **Microfone (S)** - Equipamento responsável por captar o som ambiente;
- **Maca** - A ser utilizada quando não se justifica o atendimento do paciente na mesa de consulta.

Além desses elementos, existe ainda o plano de iluminação, que conta com focos de luz a meia altura em A e B, sobre a maca, embaixo da mesa e no teto. O ambiente deve possuir paredes escuras para intensificar o contraste das imagens transmitidas.

6.4.3. Segurança da informação

A segurança da informação é muito importante nas aplicações de telessaúde, pois a transmissão da consulta poderia expor informações dos pacientes que estão sendo consultados pela junta médica. É crucial para o sucesso da aplicação garantir a privacidade do paciente. Assim, o sistema precisa estar o mais protegido possível contra ameaças externas [Silva 2003]. Este sessão aborda os aspectos de segurança e as tecnologias usadas para esse fim.

6.4.3.1. Conceitos Tradicionais de Segurança

De acordo com [ABNT NBR ISO/IEC 27001 2006] e [Stallings 2015], a segurança da informação passa pela garantia de autenticidade, do não repúdio, da confidencialidade, da integridade e da disponibilidade de um sistema de informação. Dentro do ambiente do sistema proposto, algumas considerações podem ser feitas:

- **Autenticidade:** esse conceito diz respeito à confirmação de uma identidade dentro da rede. Isso significa que a fonte é autêntica em cada informação prestada, sendo possível verificá-la. Dentro do telessaúde, temos duas fontes de informação: o CSV e o CSH. É preciso garantir que as informações venham sempre dessas fontes, o que torna necessário introduzir mecanismos que permitam verificar a origem dos dados.
- **Confidencialidade:** é a garantia que as informações sejam privadas, ou seja, não acessíveis por usuários não autorizados. No caso, todos os dados trafegados relativos à consulta precisam ser confidenciais, exigindo o uso de técnicas criptográficas e de autenticidade adequadas.
- **Integridade:** assegura que informação seja íntegra, onde somente o proprietário é autorizado modificá-la ou manipulá-la. No caso da aplicação proposta, não apenas imagem e som trafegam, mas também prontuários e exames. A responsabilidade pela edição do prontuário precisa ser respeitada, garantindo que as informações serão armazenadas e transmitidas sem modificações por terceiros.
- **Disponibilidade:** trata-se do conceito relativo ao fato de que o sistema tem que estar sempre disponível para o usuário autorizado. Ser houver perda de disponibilidade, o usuário fica sem o acesso ao sistema ou às informações. Por se tratar de uma aplicação sensível, pois o médico no CSV pode estar em meio a uma emergência, é importante garantir uma alta disponibilidade no sistema. Contudo, há sempre que se considerar que o sistema depende da internet, o que já estabelece um teto para a disponibilidade do sistema.

6.4.3.2. *Virtual Private Network - VPN*

Para garantir boa parte das características de segurança citadas, utilizam-se redes virtuais privadas (VPN) associadas à certificação digital.

Segundo [Silva 2003], uma VPN é a implementação de uma rede de comunicação privada utilizando uma rede pública, que geralmente é a Internet. A principal ideia é utilizar os protocolos padrões da Internet para promover a comunicação entre dois pontos da rede por meio de um artifício chamado de tunelamento.

Como a VPN faz uma transmissão via rede pública, em geral, é requerido o uso de criptografia para garantir a confidencialidade, a integridade e a autenticidade das informações. Além disso, o uso de uma VPN vem usualmente acompanhado de um *firewall* para proteger as redes internas.

Entre as vantagens do uso da VPN, tem-se a redução de custo, pois se elimina a necessidade de enlaces dedicados de longa distância [Tanenbaum 2011, Silva 2003]. Por outro lado, a VPN produz um maior fluxo de informação, se comparada com uma rede real, necessitando de uma conexão com boa qualidade para garantir a interoperabilidade entres os computadores e dispositivos conectados a ela [Hosner 2004].

A VPN é aplicada na solução desenvolvida para realizar a conexão entre redes locais. Portanto os computadores da rede de um hospital ou consultório em um local remoto passam a se comunicar, como se estivessem na mesma rede local, com os computadores de um hospital de referência em algum grande centro urbano. Conforme [Nakamura and de Geus 2007], essa configuração é também chamada de “*gateway-to-gateway VPN*”.

Uma VPN pode ser classificada de acordo com os protocolos que utiliza (Figura 6.11):

- **PPTP:** chamado de *Point-to-Point Tunneling Protocol*, foi desenvolvido pela Microsoft em conjunto com outras empresas [Silva 2003], para criação da VPN em camada 2. Os dados são encapsulados e podem ser criptografados para serem enviados através de redes públicas como a Internet.
- **L2F:** chamado de *Layer Two Forwarding Protocol*, foi desenvolvido pela Cisco e outra empresas. A proposta do protocolo é oferecer acesso remoto discado para redes privadas. Tem o mesmo propósito que L2TP e, segundo [Silva 2003], o mesmo se tornou obsoleto e foi descontinuado.
- **L2TP:** chamado de *Layer Two Tunneling Protocol*, foi desenvolvido pela *Internet Engineering Task Force* (IETF) [Silva 2003]. Os dados podem ser criptografados e enviados através de uma conexão ponto a ponto.
- **IPSec:** esse conjunto de protocolos e padrões de segurança pode ser usado para construção de uma VPN sobre uma rede IP. O IPSec é denso e contém muitas opções a serem configuradas e administradas de forma segura por um especialista [Hosner 2004].

- **SSL/TLS:** chamado de *Secure Socket Layer/Transport Layer Security*. Quando o IETF assumiu o desenvolvimento e o gerenciamento do SSL, ele foi renomeado para *Transport Layer Security* [Hosner 2004]. O protocolo cria um canal criptografado entre um servidor web e um navegador (browser) para garantir que os dados transmitidos sejam seguros e privados.

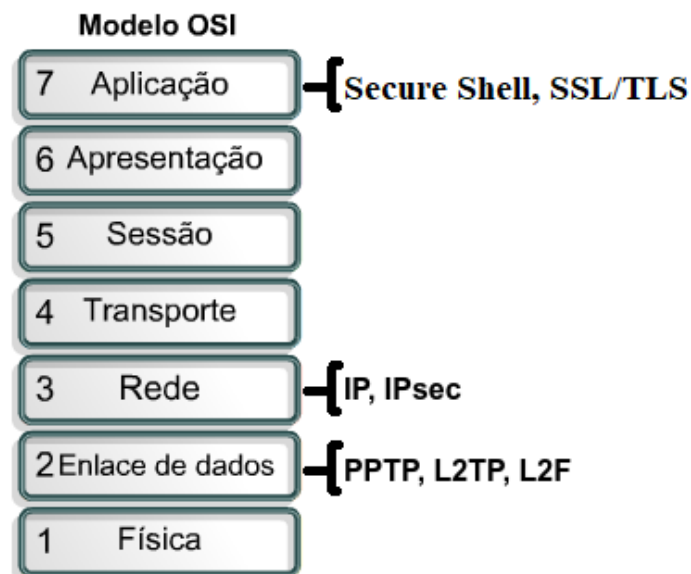


Figura 6.11: Distribuição dos protocolos no modelo OSI [Silva 2003].

Na implementação do projeto, optou-se pelo uso de VPNs baseadas em SSL/TLS, por serem mais flexíveis na configuração e administração, além de garantirem um alto nível de segurança [Hosner 2004]. Para tanto, utilizou-se a ferramenta OpenVPN [Hosner 2004, Matos et al. 2015]. O OpenVPN foi construído com um design fortemente modular. Toda a criptografia é tratada pela biblioteca SSL e toda a funcionalidade de tunelamento IP é fornecida através do drive virtual de rede TUN / TAP. Os benefícios desta modularidade podem ser vistos, por exemplo, na forma como o OpenVPN pode ser vinculado dinamicamente com uma nova versão da biblioteca SSL e imediatamente tem acesso a qualquer nova funcionalidade fornecida na nova versão. Da mesma forma, a interface do OpenVPN permite ser portada para qualquer sistema operacional que inclua um *drive* de rede virtual TUN/TAP [Hosner 2004]. O OpenVPN foi rigorosamente concebido e testado para funcionar de forma robusta em redes não confiáveis.

6.4.4. Gerência e operação do sistema

Esta seção discute as tecnologias usadas para a gerência e operação do sistema, o que inclui o consultório remoto (CSV) e consultório projeção holográfica (CSH).

Cabe observar que se colocou como objetivo do projeto que todo o sistema de software dos CSVs pudesse ser controlado a partir do CSH. Assim, o profissional de saúde no consultório remoto não precisa saber operar o sistema, mas apenas executar a sua consulta.

Para acessar os CSVs remotamente, optou-se pelo uso das ferramentas SSH e VNC. Além disso, as máquinas estão configuradas para *Wake-on-LAN*, para que possam ser ligadas/reiniciadas à distância.

O Secure Shell (SSH) foi criado 1995 [Silva 2003] e possui as mesmas funcionalidades do Telnet para acesso remoto, só que de forma mais segura. A porta 22 é padrão para acesso ao servidor auxiliado pelo TCP/IP para transporte da informação. Assim, a inicialização da captura, codificação e transmissão foi automatizada por meio de scripts usando SSH.

Para ajuste de áudio e realização de debugs, utiliza-se o *Virtual Network Computing* (VNC) [Richardson et al. 1998], que disponibiliza acesso remoto ao desktop. Assim, ele permite que o usuário se conecte a outro computador e exiba sua área de trabalho remota ao vivo ou controle o computador remotamente com o mouse e o teclado, como se você estivesse sentado em frente ao computador [TightVNC 2012]. O protocolo principal usado no VNC é *Remote Frame Buffer* (RFB) [Richardson and Levine 2011] e as portas padrão são 5900 e 5800 [TightVNC 2012]. Este protocolo permite somente a transmissão de imagens e não permite a transmissão de som ou arquivos. O VNC pode ser usado com criptografia para garantir maior segurança.

Portanto, o passo a passo para a operação do sistema consiste em:

1. Ligar computador, som, imagem e iluminação no CSH;
2. Estabelecer a VPN com o CSV escolhido;
3. Iniciar os scripts para captura, transmissão e reprodução de mídia no CSV;
4. Iniciar os scripts para captura, transmissão e reprodução de mídia no CSH;
5. Realizar ajustes de áudio, já que existem variações entre os CSVs.
6. Caso exista uso de câmeras ou sensores auxiliares, inicializar/configurar os softwares apropriados;
7. Liberar a sala para os médicos, os quais são os únicos autorizados a participar das consultas.

6.5. Pilotos Implantados

Essa seção descreve um pouco mais em detalhes as escolhas feitas na implementação dos pilotos, de forma a demonstrar as opções que foram tomadas.

6.5.1. Hardware

- **Computadores utilizados:** Todos os computadores utilizados executam sistema operacional Linux, na distribuição Ubuntu, com o fim de reduzir custos. Nenhuma outra restrição é necessária na configuração dos computadores;
- **Ponto de acesso:** Optou-se pelo uso de pontos de acesso tanto no CSV quanto no CSH. Esses pontos disponibilizam o acesso sem fio, caso seja necessário, NAT

para a VPN, a própria conexão da VPN, além de permitirem o gerenciamento de ligar/desligar os computadores. Nos pilotos atuais, faz-se uso do modelo TP-Link TL-WR842ND v1, roteador capaz de instalação do OpenWRT e configuração das VPNs ponto a ponto;

- **Câmera :** Optou-se pelo uso da Webcam Logitech HD Pro C920, que captura o vídeo direto em H.264.
- **Microfone e caixa de som:** Para o microfone do CSV, foram testados alguns modelos e todos aqueles sensíveis o suficiente para capturar sons em qualquer parte do consultório remoto são apropriados. A caixa de som também não possui fortes requisitos. A única restrição com relação a esses equipamentos é o seu posicionamento dentro do CSV para evitar a microfonia. Já no CSH, optou-se pelo uso de microfones com curto alcance ligados a uma mesa de som, já que não tem pacientes envolvidos que possam se sentir intimidados por falar diretamente no microfone;
- **Projeção holográfica:** as instalações para projeção holográfica foram feitas pela empresa Eyemotion, única distribuidora dessa tecnologia no Brasil, a qual também é parceira no desenvolvimento desse projeto.

6.5.2. Software

Dos softwares utilizados, apenas o GStreamer demanda a criação de scripts para realização das atividades. Muitas configurações são possíveis, dependendo da qualidade e tipo da rede, entre outros. Apenas para exemplificar, são mostradas algumas das opções de configuração utilizadas:

1. **Codificação de vídeo:** *“gst-launch-1.0 v4l2src device=/dev/video1 ! video/x-h264, width=1920, height=1080, framerate=30/1 ! h264parse ! rtph264pay ! udpsink host=localhost port=9000”*.

Nesse exemplo, o processo “gst-launch-1.0” captura o vídeo do dispositivo “/dev/video1” pelo drive “v4l2src” no formato h264 e passa para o codec “h264parse”, que codifica os dados recebidos para serem injetados na rede pelo “rtph264pay” (Protocolo RDP). Na sequência, são transferidos para o “localhost” (IP local) na porta 9000 pelo “udpsink” (Protocolo UDP).

2. **Codificação de áudio:** *“gst-launch-1.0 pulserc ! audioconvert ! 'audio/x-raw, rate=16000, width=16, channels=1' ! speexenc ! rtpspeexpay ! udpsink host=localhost port=9001”*.

Nesse exemplo, o processo “gst-launch-1.0” captura o áudio da fonte “pulserc” (Servidor áudio do Linux), onde “audioconvert” converte no formato Raw (Sem Padrão) e passa para o codec “speexenc” (Padrão de áudio comprimido), que codifica os dados para serem injetados na rede pelo “rtpspeexpay” (Protocolo RDP). Em seguida, os dados são transferidos para o “localhost” (IP local) na porta 9000 pelo “udpsink” (Protocolo UDP).

3. **Decodificação de vídeo:** `gst-launch-1.0 udpsrc port=9000 ! application/x-rtp, payload= 96 ! rtpjitterbuffer ! rtpH264depay ! avdec_h264 ! xvimagesink sync=false`”.

Nesse exemplo, o processo “gst-launch-1.0” recebe o vídeo pela fonte “udpsrc” na porta 9000 (Protocolo UDP), onde o “rtpjitterbuffer” (Fila de pacotes RDP) ordena os pacotes e entrega a mensagem em ordem para “rtpH264depay”, o qual traduz a informação para o codec “avdec_h264” decodificar e reproduzir o vídeo no display pelo “xvimagesink”.

4. **Decodificação de áudio:** `gst-launch-1.0 udpsrc port=9001 caps='application/x-rtp, media=(string)audio, clock-rate=(int)16000, encoding-name=(string)SPEEX, encoding-params=(string)1, payload=(int)110' ! rtpjitterbuffer ! rtpSpeexdepay ! speexdec ! audioconvert ! pulsesink sync=false`”.

Nesse exemplo o processo “gst-launch-1.0” recebe o áudio pela fonte “udpsrc” na porta 9000 (Protocolo UDP), onde o “rtpjitterbuffer” (Fila de pacotes RDP) ordena os pacotes e entrega a mensagem ordenada para “rtpSpeexdepay”. Em seguida, a informação é enviada para o codec “speexdec” decodificar e passar para o “audioconvert” converter e reproduzir o áudio pela fonte “pulsesink”.

Os exemplos dados acima são para um ambiente Linux. Embora o GStreamer funcione em diversos sistemas operacionais, os scripts não são interoperáveis entre sistemas. Portanto, tais comandos podem variar em diferentes sistemas operacionais. No Linux, o *drive* “V4l2” cuida da lista de câmeras compatível com sistema e “Pulse” é serviço responsável pela captura e reprodução do áudio no sistema, sendo assim os mesmos não estão presente em outros sistemas operacionais.

6.5.3. Experimentos

Para mensurar o desempenho do sistema de captura, exclusivamente o vídeo, foi realizado um experimento para determinar a latência de propagação da imagem via rede. É muito difícil medir latência exata, pois isso exigiria um dispositivo de sincronização GPS com precisão de relógio atômico. Assim, para uma estimativa inicial, utilizou-se uma metodologia mais simples de medir a latência: filmar com a câmera um display exibindo um cronômetro e exibir imagem capturada com tempo em outro display. A partir disto, se pode tirar a diferença entre o tempo real do cronômetro e o tempo que é exibido na imagem capturada, como mostra na Figura 6.12.

Ambiente dos experimentos: Foram usados dois computadores, cujo hardware possui um processador de 3 GHz com 4 núcleos e memória de 4 GB DDR3. Em ambos, foi instalado o sistema operacional Linux Ubuntu 16.04.02 LTS e configurado o GStreamer versão 1.0 e o VLC versão 2.2.2.

6.5.3.1. Experimento de captura

Esse teste consiste em determinar a latência de captura do GStreamer e do VLC localmente, onde a imagem capturada do cronômetro é exibida no mesmo display. Para

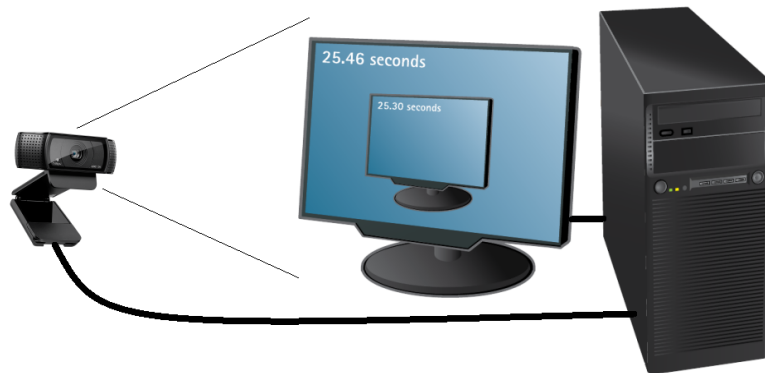
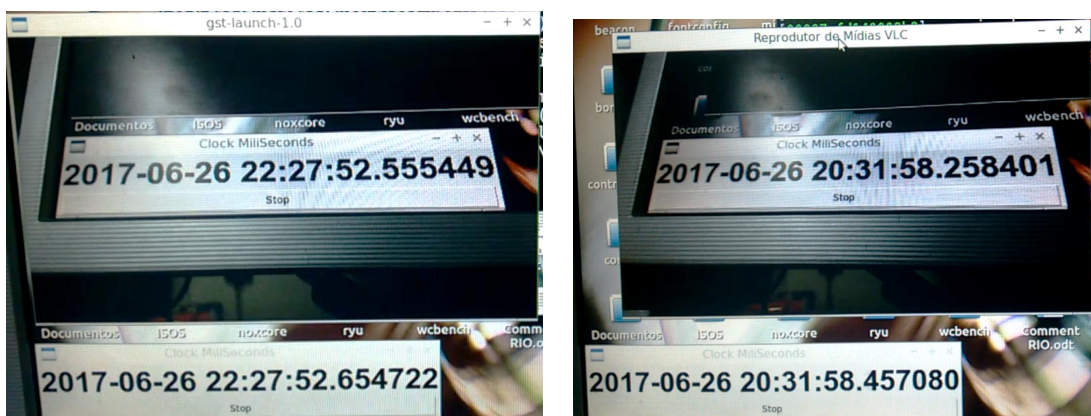


Figura 6.12: Metodologia simplificada para estimar a latência na transmissão do vídeo.

esse experimento, foi necessário somente um computador conectado a câmera. A captura foi realizada codificando o vídeo em H.264 com uma resolução de 848x480 (480p), 1280x720 (720p) e 1920x1080 (1080p) com *frame rate* igual a 30 FPS (*frames* por segundo). Algumas amostras em 480p são mostradas na Figura 6.13.



(a) GStreamer - atraso aproximado de 100 ms.

(b) VLC - atraso aproximado de 200 ms.

Figura 6.13: Amostras da medida de latência em 480p.

6.5.3.2. Experimento de propagação via rede

Esse teste consiste em determinar o impacto da latência da rede sobre o vídeo. Para esse experimento foram necessários dois computadores conectados em rede a um roteador TP-Link TL-WR842ND, um computador com a câmera para captura e transmissão do vídeo e outro computador para receber e exibir o vídeo. Para esse teste, observou-se a latência de captura do GStreamer via rede, onde a imagem capturada do cronômetro (Tempo real), com uma pré-visualização para determinar o tempo de captura, é transmitida e exibida em outro computador na rede para determinar o tempo de propagação via rede.

A transmissão foi realizada codificando o vídeo em H.264 com uma resolução de 848x480 (480p), 1280x720 (720p) e 1920x1080 (1080p), com *frame rate* igual a 30 FPS (*frames* por segundo), em 3 Mbps de taxa útil através do protocolo RTP (*Real Time*

Protocol) em unicast para o receptor. A Figura 6.14 apresenta uma amostra em 480p para captura do GStreamer via rede: no display direito, está o cronômetro e a latência de captura em pré-visualização; e no display esquerdo está a latência de propagação do vídeo via rede.

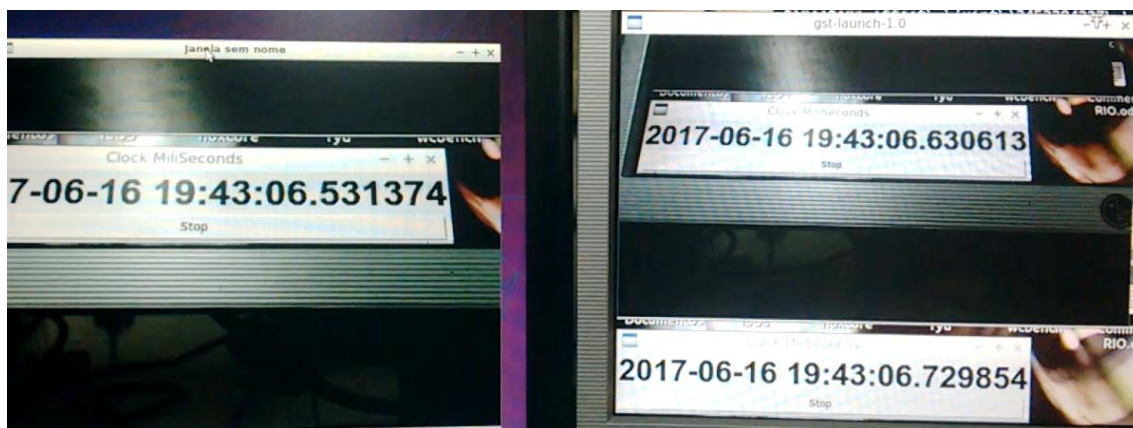


Figura 6.14: Amostra de Latência do GStreamer via rede em 480p.

6.5.3.3. Resultados dos testes de vídeo

Os resultados dos experimentos são mostrados nas Figuras 6.15 e 6.16. Pelos resultados obtidos (Figura 6.15), a captura do GStreamer apresenta um atraso próximo de 100 ms para todas as resoluções citadas, o que é um resultado adequado para videoconferências em tempo real. Já o VLC não obteve um resultado satisfatório, pois a menor latência é de aproximadamente 200 ms para resolução 480p e aumenta de acordo com o tamanho da resolução configurada. Os resultados de latência para propagação via rede do GStreamer (Figura 6.16) foram de aproximadamente 200 ms para todas as resoluções citadas, o que ainda é considerado aceitável para *streaming* de vídeo em tempo real.

Esses resultados preliminares serviram para auxiliar na escolha da ferramenta. A validação foi posteriormente finalizada pelo uso da ferramenta dentro dos dois pilotos desenvolvidos para o projeto.

6.6. Resultados preliminares com os pacientes

Num país de dimensões continentais como o nosso, levar assistência em saúde a todos os seus recantos é um desafio. Na realidade, muitos pacientes não terão acesso a serviços de saúde a não ser que o profissional vá até às áreas remotas cujos recantos somente o Exército e a Marinha tem acesso. Nestes locais, em geral, é o profissional inexperiente e recém-formado que chega. Dar a ele suporte supervisonal poderá salvar muitas vidas. Deste modo, a busca de um meio de levar supervisão a este profissional que trabalha em locais distantes feita através da parceria entre a Medicina e a Engenharia de Telecomunicação fez com que o projeto Telemedicina surgisse como área de inovação científica.

Nos últimos cinco meses, foram programadas 18 consultas holográficas, sendo realizadas 16 avaliações de pacientes através do sistema holográfico. Algumas imagens das

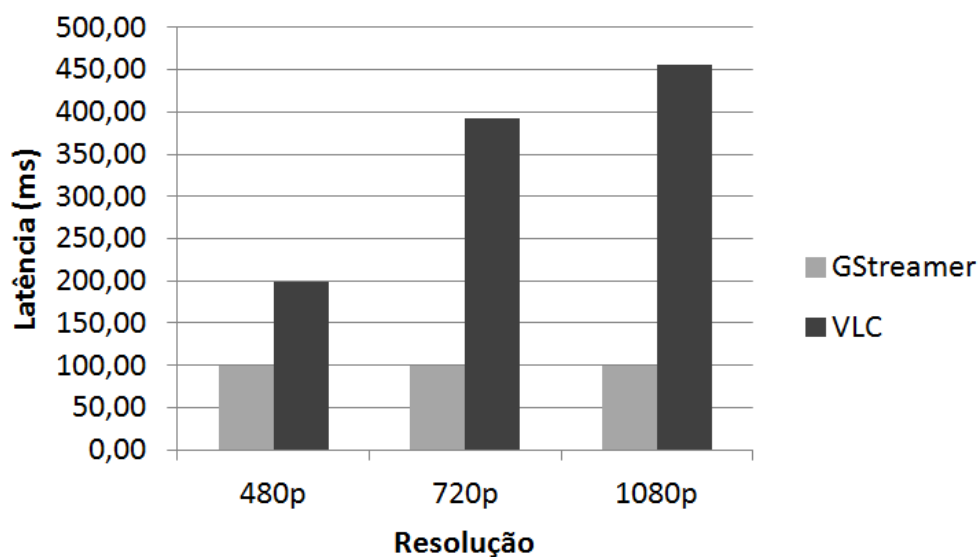


Figura 6.15: Latência de captura de imagem para VLC e GStreamer.

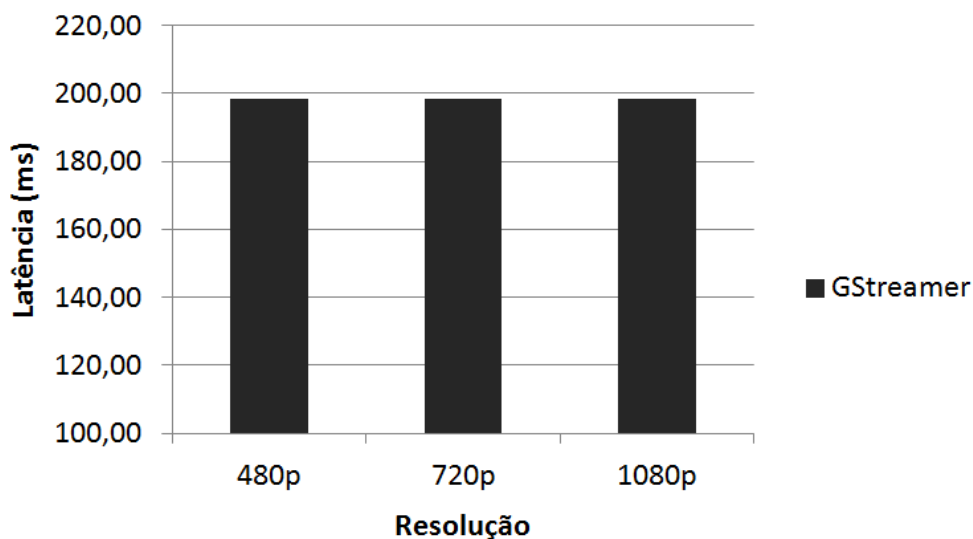


Figura 6.16: Latência de codificação e decodificação do GStreamer via rede.

consultas, divulgadas com autorização dos pacientes, estão na Figura 6.17. As consultas não realizadas foram suspensas apenas por dificuldades de rede e não por restrição do paciente quanto ao ambiente holográfico. Estas consultas são marcadas mensalmente e, antes de sua realização, o paciente é informado sobre sua estrutura, para que o mesmo opte por ser avaliado ou não através da holografia. Desta forma, é explicado ao idoso que, na consulta, o médico residente o avaliará, sendo supervisionado no centro de saúde holográfico por seu preceptor. É explicado ao paciente que a consulta holográfica é realizada como uma consulta médica normal, diferindo apenas por ter em sua sala o equipamento de captura da imagem e por ter as paredes do consultório de cor escura para que imagem holográfica seja formada com nitidez. Desta forma, todos os pacientes que foram

até o momento avaliados não referiram qualquer constrangimento quanto ao ambiente do consultório holográfico, uma vez que o mesmo possui o mobiliário regular de qualquer consultório médico clínico.

Como este é um projeto de pesquisa, este ambiente holográfico está em constante adaptação para as demandas médicas, pois, após cada consulta, discutem-se as facilidades e dificuldades do sistema, no sentido de qualificá-lo. Este é um interessante processo, uma vez que pacientes e alunos dão suas opiniões, o que permite o aprimoramento cada vez maior do sistema. Inicialmente, as dificuldades apontadas pelos pacientes foram o impacto da cor da sala, o que no desenrolar da consulta era esquecido, desaparecendo o impacto inicial. Segundo, a troca de informações sem a presença física do instrutor, o que se tornava natural à medida que se iniciava a interação do paciente com o binômio aluno-professor. Este impacto foi encerrado a partir do momento em que se introduziu na sala a presença da imagem de retorno do preceptor. Por fim, o uso do microfone, que em primeiro momento era do modelo *Gooseneck*, exigindo do paciente atenção ao responder aos questionamentos do médico, foi apontado como um problema. Essa questão, posteriormente, foi eliminada pela troca do aparelho por um microfone para captura de som ambiente, não exigindo mais do paciente especificamente falar ao microfone. Dessa forma, o paciente até se esquecia de que havia um microfone na mesa, passando a falar naturalmente na consulta.

Estas adaptações só foram possíveis em virtude da participação ativa do paciente no processo de construção do consultório virtual, dando opiniões e exigindo coisas que lhe dariam mais conforto em sua avaliação médica. Estes pontos também foram apontados pelos profissionais que trabalharam na construção desse ambiente holográfico. De modo geral, os pacientes se sentiram tranquilos e bastante à vontade nas consultas. De fato, alguns até relataram que após a explicação da equipe médica prévia à consulta, acharam que entrariam numa espécie de cápsula espacial, se surpreendendo positivamente ao ver o consultório virtual.

Os pós-graduandos do Serviço de Geriatria e alunos de graduação da Faculdade de Medicina da UFF que participam desta experiência fizeram diferentes apontamentos. Uma das dificuldades iniciais foi aprender a lidar com o atraso do sistema, pois este traz um pequeno eco da voz. Essa questão hoje já foi superada pela equipe, já quase não interferindo com a qualidade da consulta. Outros apontamentos que surgiram foram: falta de visualização dos pés do paciente, dificuldade de ver a outra face do paciente, dificuldade de ver a lesão devido à coloração amarela do foco de luz inicialmente usado e a necessidade de melhorar a intensidade da voz no retorno do som. Essas observações permitiram a equipe de telecomunicação realizar ajustes de luz, som e imagem, facilitando progressivamente a consulta holográfica e chegando a uma imagem do paciente que hoje nos autoriza a fazer consultas até dermatológicas através do sistema.

Alguns pontos negativos ainda tem que ser superados e vêm sendo trabalhados pela equipe. De fato, todo o desenvolvimento do sistema vem sendo feito em um processo espiral, onde melhorias são introduzidas conforme os *feedbacks* de médicos e pacientes surgem. Entre as dificuldades ainda não superadas, está a dificuldade para os médicos lidarem com a troca de fluxos de vídeo da câmera principal para câmeras auxiliares, como a câmera do celular, usada para avaliar lesões dermatológicas. Esse e outros apontamentos



(a) Exemplo de atendimento com idosa e residentes.



(b) Exemplo de atendimento com apenas uma residente.

Figura 6.17: Fotos de imagens holográficas de consultas reais. Para fim de preservação da privacidade dos pacientes, seus rostos foram distorcidos.

estão sendo usados para o desenvolvimento de um sistema de automação com interface amigável para os médicos interagirem com os equipamentos.

6.7. Trabalhos em andamento

A lista de desafios técnicos do Projeto Telessaúde é potencialmente gigantesca. Na medida em que avançam as funcionalidades do sistema e que se vislumbram novas possibilidades, surgem acréscimos e ocorrem as fusões e reordenamentos dos tópicos de investigação, como é natural em qualquer projeto de pesquisa. Um dos diferenciais do projeto é justamente o fato destes diversos aspectos estarem em fases diferentes na cadeia de produção científica e tecnológica. Há assuntos com graus de amadurecimento distintos – o Telessaúde é, atualmente, um *testbed* de experimentação e desenvolvimento de novas ideias, mas é também um projeto piloto, em processo de implementação.

Dentre as áreas de investigação contempladas pelo projeto, destacam-se:

- Sistemas Multimídia – Equipamentos, software e métodos para a captação, processamento e transmissão do áudio e do vídeo, incluindo os domínios analógico e digital e a conversão, codificação e compressão dos fluxos multimídia;
- Desenvolvimento de um Painel de Automação - É o sistema de interface amigável para operação, configuração e controle do ambiente e dos demais subsistemas do CSH e do CSV;
- Questões Ambientais - Dizem respeito ao ambiente e inclui as questões de ergonomia, conforto, adequação aos métodos de captação de imagem e som, cores, materiais e design de interiores;
- Monitoramento do Paciente - Além das informações transmitidas sob a forma de imagem, é possível capturar e transmitir medidas biométricas (pressão arterial, pulso, oximetria, dentre outras) que podem prover informações úteis e complementares, auxiliando no diagnóstico;
- Monitoramento do Ambiente – Dados do meio ambiente (temperatura, umidade e outras) podem auxiliar na manutenção do sistema;
- Comunicação – Diz respeito às componentes ativas de comunicação instaladas no CSV, como switches, roteadores, roteadores sem fio, e cabeamento de rede. Trata da especificação destes elementos, de seu monitoramento e gerência. Além disso, há as questões de interligação entre CSV e CSH, que pode se dar por sistemas de comunicação satelitais ou terrestres.
- Segurança da informação – o conjunto de técnicas e dispositivos usados para garantir a privacidade, a integridade e a autenticidade das informações trafegadas, em se considerando não apenas os requisitos de segurança preconizados nas melhores práticas de TIC, mas também as questões da ética médica.
- Disponibilidade do sistema – olhar holístico sob o sistema, para identificação de pontos de falha críticos e projetos de contingência.

- Gestão do conhecimento – Inclui as técnicas de gestão da informação e a documentação, com os objetivos de permitir a manutenção do sistema, facilitar o treinamento da equipe, sobretudo de membros ingressantes, oferecer base de dados de apoio aos atendimentos, com o objetivo de auxiliar a equipe de saúde, manter histórico dos atendimentos, com o objetivo de gerar métricas e embasar relatórios e publicações científicas.

Todos esses temas já estão em desenvolvimento e devem estar disponíveis nas próximas versões do sistema, por meio de trabalhos de iniciação científica, trabalhos de conclusão de curso, dissertações de mestrado e teses de doutorado.

6.8. Conclusão

O Projeto Telessaúde tem despertado o interesse de autoridades de diversos setores. Dentre as diversas possibilidades de desdobramentos, destacam-se duas tratativas em estágio avançado. Tem sido profícua a parceria com Exército Brasileiro, mais especificamente com o Hospital Central do Exército, onde já foi instalado um Consultório Virtual (CSV-HCE). Nesta fase inicial, objetiva-se colocar em operação o Sistema de Saúde Holográfico, em tempo real, para diagnóstico, acompanhamento e supervisão de ações de saúde, pelos médicos do HCE, a partir do Centro de Saúde Holográfico da UFF (CSH-UFF), com os pacientes do Exército sendo atendidos no Consultório de Saúde Virtual do HCE (CSV-HCE). Uma vez que seja realizado um grande número de atendimentos, será possível ajustar e refinar os parâmetros técnicos do sistema de telecomunicações terrestre do Sistema de Saúde Holográfico (rede ótica) para atender à qualidade necessária aos atendimentos médicos.

Futuramente, este projeto poderá ser expandido para outras instalações do Exército, como, por exemplo, nos Pelotões de Fronteira, onde as dificuldades de atendimento médico são extremas.

Está também em tratativas a criação do “Projeto de Telemedicina para Integração do Hospital Universitário Antônio Pedro (HUAP) com a Rede Básica de Saúde da Região Metropolitana 2”, com financiamento de emenda parlamentar. A premissa é que a rede básica de saúde é dedicada ao atendimento primário da população na sua região residencial, sendo os casos de maior complexidade encaminhados ao Hospital Universitário Antônio Pedro (HUAP). Atualmente, este fluxo, até o atendimento efetivo no HUAP, é complexo e prolongado, não se identificando nem priorizando os casos de maior gravidade e emergenciais. Através da instalação de CSVs nos Postos de Saúde da Região, esta realidade pode mudar.

Entre ações preconizadas para o sistema tem-se:

- Consultoria em medicina, neuropsicologia e telecomunicações;
- Diagnóstico *in loco* de diferentes morbidades;
- Diagnóstico neuropsicológico;
- Diagnósticos em telecomunicações;

- Diagnósticos em diferentes áreas, como farmácia, odontologia, nutrição, fisioterapia, terapia ocupacional, enfermagem e outras;
- Implantação de cursos online para educação continuada;
- Implantação de testes online em diversas áreas do saber;
- Ativação de comunicação em rede agilizando processos e suporte de equipamentos;
- Reuniões de serviço distantes;
- Parcerias em pesquisa clínica e laboratorial;

Este projeto trará integração entre distintos setores de saúde e de outros campos do saber, aproximando universidades e pesquisadores, permitindo trabalhos multicêntricos de interesse nacional. Essas experiências compartilhadas com a unificação do conhecimento dos *experts* podem solucionar situações dando maior resolutividade e maior velocidade às ações dos projetos de pesquisa que auxiliarão atitudes governamentais no apoio às comunidades atendidas. Além disto, ele poderá ser um instrumento de educação permanente e treinamento profissional. Ele permitirá, ainda, redução das despesas com pesquisa, evitando deslocamento de pesquisadores, o que por vezes se torna dispendioso, além de evitar deslocamento desnecessário de pacientes de seus locais de residência. Através do projeto, pode-se construir uma rede de serviços de prevenção em saúde, qualificando o planejamento de ações institucionais, o que trará como resultado final uma qualificação para as atividades da atenção básica em prefeituras, empresas públicas e outros setores da sociedade civil. Através da holografia, será possível compartilhar dados de diferentes locais, construir uma rede de diagnóstico em tempo real e levar serviços especializados a profissionais que se encontram distantes dos grandes centros, compartilhando em tempo real arquivos, exames e diagnósticos especializados para populações distante das universidades. Desta forma, este trabalho terá impacto científico, tecnológico, social e até mesmo econômico.

Em síntese, o Projeto Telessaúde tem grande potencial como plataforma para interiorização da medicina em geral e, mesmo nos centros urbanos, seu potencial tem sido reconhecido pela comunidade médica e por autoridades civis e militares.

6.9. Visão do Futuro

Existem diversos tipos de abordagens para se procurar obter uma visão de futuro de uma determinada situação no presente. Aqui se optou, devido à complexidade do assunto, analisar eventos históricos, ocorridos em outros países, procurando adequá-los ao nosso caso presente e, então, a partir daí, propor algumas projeções futuras.

Inicialmente, será considerado o desenvolvimento da Telemedicina, uma vez que o Telessaúde, como descrito neste trabalho, se apresenta como uma evolução tecnológica, a partir da inclusão de recursos de Holografia na Telemedicina tradicional. De fato, a Telemedicina iniciou-se há mais de 60 anos, durante o programa espacial americano, quando a NASA implementou meios, com recursos de vídeo, para possibilitar o diagnóstico e o tratamento de astronautas em missões espaciais.

De acordo com a Organização Mundial de Saúde (OMS), “a Telemedicina compreende a oferta de serviços ligados aos cuidados com a saúde, nos casos em que a distância é um fator crítico. Tais serviços são prestados usando tecnologias de informação e de comunicação para o intercâmbio de informações válidas para diagnósticos, prevenção e tratamento de doenças e a contínua educação de prestadores de serviços em saúde, assim como para fins de pesquisas e avaliações”. Ressalte-se que, no caso da NASA, citado anteriormente, a distância era um fator crítico. Com o avanço da tecnologia, observa-se um incremento das tecnologias de informática e telecomunicações no ambiente da saúde. Surge, em decorrência, a figura do técnico em Telemedicina, responsável por dar suporte, local ou remoto, aos profissionais de saúde que fazem uso de videoconferências. Ele é o responsável por montar os equipamentos de vídeo, áudio, internet, assim como dar um treinamento básico para os profissionais de saúde que ainda não estão familiarizados com essas novas ferramentas. No futuro próximo, então, pode-se supor que irá crescer a demanda por este tipo de técnico, que deverá incorporar conhecimentos em áreas tecnológicas mais recentes aplicadas à Medicina, como a Holografia.

No exterior, observa-se um interesse crescente na Telemedicina, principalmente naqueles países esparsamente povoados, aonde a distância é um fator crítico. Um bom exemplo é a Groenlândia, que vem aplicando esta tecnologia para pacientes remotos desde 2010. De fato, o seu sistema foi copiado do seu congênere no Alaska. Na Groenlândia, em média, 22% da população reside em vilarejos localizados a centenas de quilômetros de distância do centro de saúde mais próximo. Durante a maior parte do tempo, a realização de viagens em seu território é muito difícil, senão impossível. Atualmente, a Telemedicina na Groenlândia é um programa ativo de governo, considerado essencial e prioritário.

Em 2011 a população mundial atingiu sete bilhões de pessoas. Dados de 2012 do Instituto Brasileiro de Geografia e Estatística (IBGE) revelam que o Brasil ultrapassou os 193 milhões de habitantes. À medida que a população cresce e envelhece, a assistência médica torna-se cada vez mais necessária. A qualidade no atendimento continuará a se deteriorar se a medicina continuar a ser praticada da forma tradicional em consultórios e hospitais. Considera-se que somente o emprego da Telemedicina poderá vir a alterar este quadro. Da tribuna do Senado, a senadora Vanessa Grazziotin fez as seguintes afirmações:

1. Existem municípios da Região Amazônica que contam com a presença de um médico somente a cada quinquena;
2. A média de médicos no Brasil é de 1,95 para cada grupo de mil habitantes, quase o dobro da recomendada pela Organização Mundial de Saúde (OMS), de um médico para cada mil habitantes. Esses profissionais, no entanto, estão concentrados, em sua maioria, nos grandes centros urbanos.

Em entrevista para a Rede Globo, o senador Eduardo Braga afirmou que "nós na Amazônia temos grandes e graves carências de médicos especialistas no interior da nossa região. Não temos ginecologistas, ortopedistas, pediatras, obstetras, neurocirurgiões, anestesistas, enfim, faltam médicos especialistas no interior. Não adianta colocarmos salários de até R\$ 25 mil, porque esses médicos ganham isso ou quase isso em outras cidades e preferem morar em Ipanema do que em Tabatinga, por exemplo".

Tendo em vista o quadro anteriormente delineado, pode-se afirmar que a Telemedicina apresenta-se como uma solução para os problemas de saúde decorrentes do aumento e envelhecimento da população, concomitantemente ao atendimento de populações de difícil acesso, por exemplo aquelas situadas na Região Amazônica.

Adicionalmente, pode-se prever o uso da Telemedicina também para o apoio às tripulações de navios das Marinhas de Guerra e Mercante. Em resumo, pode-se prever o seguinte quadro futuro para a Telemedicina:

1. Incorporação de técnicas avançadas, baseadas em Holografia, na concepção atual da Telemedicina, levando aos sistemas denominados por Telessaúde. Com este procedimento, permite-se, ao especialista, uma visão em 3D do paciente, com as vantagens decorrentes;
2. Formação de técnicos especializados, o que levará à definição de uma nova carreira técnica, com o objetivo de atender os profissionais de saúde no manejo dos equipamentos eletrônicos concernentes ao Telessaúde;
3. Utilização cada vez mais intensiva dos sistemas de Telessaúde, para atendimento de populações esparsas e apoio aos sistemas médicos para atendimento às tripulações das Marinhas de Guerra e Mercante;
4. Utilização dos sistemas de Telessaúde também para a atividade de treinamento médico.

Devido às razões apresentadas, em um futuro entre 2 a 5 anos, espera-se a proposta e o desenvolvimento de novos sistemas de telessaúde, com a implantação dos primeiros serviços pelo país. Há também expectativa para a evolução da legislação para apoio dos médicos que prestam ou recebem esse tipo de assistência remota. Algumas ações já foram tomadas pelos conselhos regionais e federais, mas ainda há muito que se discutir sobre o tema. Contudo, em um período relativamente curto, entre cinco e dez anos, é esperado que muitas iniciativas de telessaúde já sejam realidade no Brasil, de forma ampla, assim como já são em outras partes do mundo.

Bibliografia

- [ABNT NBR ISO/IEC 27001 2006] ABNT NBR ISO/IEC 27001 (2006). Tecnologia da informação – técnicas de segurança – sistemas de gestão de segurança da informação – requisitos.
- [Boulton et al. 2017] Boulton, R. J., Walthinsen, E., Baker, S., Johnson, L., Bultje, R. S., Kost, S., Müller, T.-P., and Taymans, W. (2017). Gstreamer plugin writer's guide.
- [Buvik et al. 2016] Buvik, A., Bugge, E., Knutsen, G., Småbrekke, A., and Wilsgaard, T. (2016). Quality of care for remote orthopaedic consultations using telemedicine: a randomised controlled trial. *BMC health services research*, 16(1):483.
- [Dal Bello et al. 2012] Dal Bello, J., Vale, E. R., Rebello, C. A. C., Silva, L. E., Lima, J. A. M. C., Boechat, Y. E. M., Dal Bello, M. A. O., Carvalho, M. L. O., Mendes,

- N. R. S., Baltar da Silva, E., and Benttenmüller Pereira, G. (2012). Pedido de patente inpi para o processo de telessaúde holográfico.
- [Faperj 2017] Faperj (2017). Saúde ao vivo, a cores e em 3D. *Rio Pesquisa*, 9(38):26–28.
- [FFmpeg.org 2017] FFMpeg.org (2017). FFMpeg - A complete, cross-platform solution to record, convert and stream audio and video. Disponível online em <https://www.ffmpeg.org/>.
- [GStreamer 2017] GStreamer (2017). Gstreamer. Acessado em: jul.2017.
- [Hosner 2004] Hosner, C. (2004). Openvpn and the ssl vpn revolution. Acessado em: ago.2017.
- [Marinha do Brasil 2016] Marinha do Brasil (2016). Navio da Marinha leva atendimento de saúde a comunidades ribeirinhas do Acre. Disponível online em <http://www.defesa.gov.br/noticias/17938>.
- [Marinha do Brasil 2017] Marinha do Brasil (2017). Assitência hospitalar às comunidades ribeirinhas na Amazônia. Disponível online em <https://www.mar.mil.br/asshop/navios.htm>.
- [Matos et al. 2015] Matos, F. H. S., Ragnini, D. F., de Oliveira, P. d. T. C., and de Almeida, F. M. (2015). Implementação de segurança em redes wi-fi com a utilização de VPN. *LINKSCIENCEPLACE-Interdisciplinary Scientific Journal*, 2(1).
- [Nakamura and de Geus 2007] Nakamura, E. and de Geus, P. (2007). *Segurança de Redes em Ambientes Cooperativos*. Novatec.
- [Poon, T. C. 2007] Poon, T. C. (2007). *Optical Scanning Holography with Matlab*. Springer.
- [Poon, T. C.; Liu, J. P. 2014] Poon, T. C.; Liu, J. P. (2014). *Introduction to Modern Digital Holography*. Cambridge University Press.
- [Richardson and Levine 2011] Richardson, T. and Levine, J. (2011). The remote framebuffer protocol. Internet Requests for Comments.
- [Richardson et al. 1998] Richardson, T., Stafford-Fraser, Q., Wood, K. R., and Hopper, A. (1998). Virtual network computing. *IEEE Internet Computing*, 2(1):33–38.
- [Silva 2003] Silva, L. S. d. (2003). *Virtual Private Network: VPN*. Novatec, 1 edition.
- [Stallings 2015] Stallings, W. (2015). *Criptografia e segurança de redes: princípios e práticas*. Pearson Brasil, 6 edition.
- [Steinmeyer 2013] Steinmeyer, J. (2013). *The Science Behind the Ghost*. Hahne.
- [Tanenbaum 2011] Tanenbaum, A. (2011). *Redes de computadores*. Pearson Prentice Hall, 5 edition.

- [Taymans et al. 2017] Taymans, W., Baker, S., Wingo, A., Bultje, R. S., and Kost, S. (2017). Gstreamer application development manual.
- [Tektronix 2009a] Tektronix (2009a). *Um Guia para Fundamentos de MPEG e Análise de Protocolo*. Tektronix.
- [Tektronix 2009b] Tektronix (2009b). *Um Guia para Sistemas e Medidas de Televisão Digital*. Tektronix.
- [Theurer et al. 2017] Theurer, L., Bashshur, R., Bernard, J., Brewer, T., Busch, J., Caruso, D., Coccaro-Word, B., Kemalyan, N., Leenknecht, C., McMillan, L. R., et al. (2017). American telemedicine association guidelines for teleburn. *Telemedicine and e-Health*, 23(5):365–375.
- [TightVNC 2012] TightVNC (2012). Tightvnc for windows: Installation and getting started. Acessado em: ago.2017.
- [VideoLAN 2017] VideoLAN (2017). Vlc. Acessado em: ago.2017.
- [Xiph.org 2017] Xiph.org (2017). Speex: A free Codec for Free Speech. Disponível online em <https://www.speex.org/>.

Anexo 1: biografia dos autores

Ana Carolina Ferreira Beaklini



É formada pela UFF em Engenharia de Telecomunicações. Está cursando o Mestrado em Engenharia Elétrica e de Telecomunicações na UFF. Atualmente, atua como engenheira de sistemas de televisão da Rede Globo de Televisão. Possui interesse nas áreas de redes de computadores e avaliação de equipamentos de transmissão de televisão.

André Luis de Oliveira Fonseca



Possui graduação em Ciência da Computação pelo Centro Universitário de Barra Mansa (2008). Está cursando o Mestrado em Engenharia Elétrica e de Telecomunicações na UFF. Atualmente, atua como Técnico de Tecnologia da Informação da UFF. Possui interesse nas áreas de redes de computadores e multimídia.

Eduardo Rodrigues Vale



Possui Mestrado em Engenharia Elétrica pela PUC-RJ (1975) e Doutorado, também pela PUC-RJ, em 2006. É professor da UFF desde 1983, lecionando no Departamento de Telecomunicações. Atualmente é professor Titular da UFF. Exerce o cargo de Coordenador Científico do Núcleo de Estudos Avançados da UFF. Tem experiência na área de Engenharia Elétrica, com ênfase em Sistemas de Telecomunicações, atuando principalmente em redes de dados, redes de nova geração, IoT e sistemas móveis de terceira e quarta gerações.

Julio Cesar Rodrigues Dal Bello



Realizou curso de graduação em Engenharia de Telecomunicações no IME (1979), curso de mestrado em Engenharia Elétrica no IME (1984) e curso de doutorado em Engenharia Elétrica na PUC-RJ (1998). É autor de centenas de estudos, perícias e projetos-técnicos de engenharia de telecomunicações, que foram realizados para governos dos três níveis e dezenas de empresas, nacionais e transnacionais. Desde 1982 é professor do Departamento de Engenharia de Telecomunicações da UFF, se dedicando a pesquisas nas áreas de Sistemas Modernos de Telecomunicações, Tele-educação e Telemedicina.

René Pestre Filho



Possui graduação em Engenharia de Telecomunicações pela UFF (1976). Atualmente é professor da UFF. Tem experiência na área de Engenharia Elétrica, com ênfase em Telecomunicações. Atuou em empresas por 40 anos, com destaque para grande prestadora de serviços de telecomunicações. Possui MBA Executivo pela COPPEAD-UFRJ. Mais de 30 anos de experiência como professor no curso de Engenharia de Telecomunicações da UFF.

Ricardo Campanha Carrano



É Coordenador Geral do NETAv da Escola de Engenharia da UFF e professor adjunto do Departamento de Engenharia de Telecomunicações, onde é chefe do Setor de Comunicação de Dados e Vice-Coordenador da Graduação. Possui os títulos de Engenheiro de Telecomunicações (1995), Mestre em Engenharia de Telecomunicações (2009) e Doutor em Computação (2013), todos pela UFF. Seus interesses de pesquisa incluem Internet das Coisas, Redes de Computadores, Redes sem Fio, Redes de Sensores sem Fio, e Telemedicina, já tendo atuado em diversos projetos de pesquisa relacionados a esses temas, financiados por órgãos de fomento como Faperj, RNP e Aneel.

Natalia Castro Fernandes



Possui graduação em Engenharia Eletrônica e de Computação (2006), mestrado em Engenharia Elétrica (2008) e doutorado em Engenharia Elétrica (2011) pela UFRJ. Atualmente, é professora adjunta da Escola de Engenharia da UFF e subchefe do Departamento de Engenharia de Telecomunicações. Já participou de diversos projetos de pesquisa financiados pelo CNPq, FAPERJ e RNP. Tem experiência na área de redes de comunicação, atuando principalmente nos seguintes temas: telessaúde, segurança, redes sem-fio e Internet do Futuro.

Yolanda Eliza Moreira Boechat

Possui Mestrado e Doutorado em Medicina pela UFF. Atualmente, é Professora Adjunta de Geriatria do Departamento de Medicina Clínica, Geriatria, Supervisora de Residência Médica em Geriatria e Residência Multiprofissional - Área de Concentração - Atenção à Saúde do Idoso do Hospital Universitário Antonio Pedro e Coordenadora do Serviço de Geriatria/Centro de Referência em Assistência a Saúde do Idoso (CRASI/PIGG) da UFF. Tem experiência principalmente nos seguintes temas: idoso, atendimento multiprofissional, riscos da hospitalização, cognição.

Capítulo

7

Projetos Atuais e Visão de Futuro do Laboratório TeleMídia/PUC-Rio em Videocolaboração

Sérgio Colcher¹, Alan L. V. Guedes¹, Roberto Gerson de A. Azevedo¹,
Guilherme F. Lima¹, Rodrigo C. M. Santos¹, Antonio J. G. Busson¹

¹Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio)

{colcher, aguedes, razevedo, glima, rsantos, abusson}@inf.puc-rio.br

Abstract

Since the early 1990's, the TeleMídia laboratory of the Department of Informatics of PUC-Rio researches and trains people in the areas of multimedia and hypermedia, which include interactive videos and video collaboration as important subareas. This chapter presents the main research lines and projects of the laboratory and discusses the future vision of the authors about interactive videos and video collaboration.

Resumo

Desde a década de 1990 o laboratório TeleMídia do Departamento de Informática da PUC-Rio realiza pesquisas e forma recursos humanos em sistemas multimídia e hipermídia, dos quais vídeos interativos e videocolaboração são importantes subáreas. Este capítulo apresenta as principais linhas e projetos de pesquisa do laboratório e a visão de futuro dos autores sobre vídeos interativos e videocolaboração.

7.1. Introdução

Desde 1991 o laboratório TeleMídia, um dos laboratórios temáticos do Departamento de Informática da PUC-Rio, desenvolve pesquisas e forma recursos humanos em sistemas multimídia e hipermídia. Durante esses anos, vários têm sido os temas de pesquisa abordados pelo laboratório, incluindo diversas aplicações de sistemas multimídia, tais como sistemas de teleconferência, e-mail multimídia, TV e rádio digitais, IPTV, streaming de mídia sob demanda, e, mais recentemente, experiências imersivas e internet das coisas. Vários desses trabalhos estão relacionados aos temas de vídeos interativos e videocolaboração, sendo que muitas das tecnologias e dos padrões desenvolvidos no laboratório podem servir de base para a implementação de diferentes cenários de aplicações avançadas nestas áreas.

Além do seu papel no desenvolvimento científico e na formação de pessoas, o laboratório TeleMídia também tem participado ativamente na transferência de tecnologia para a indústria e contribuído com órgãos de padronização nacionais e internacionais, tais como a Associação Brasileira de Normas Técnicas (ABNT) e a União Internacional das Telecomunicações (UIT). No que se refere à padronização, o laboratório tem como destaques a criação do *middleware* Ginga e da linguagem NCL que suportam o desenvolvimento de aplicações multimídia interativas para ambientes de TV digital, IPTV e sistemas integrados *broadcast/broadband* (IBB). Tanto o *middleware* Ginga como a linguagem NCL são hoje padrões ABNT para sistemas de TV digital terrestre (ABNT 2016) e recomendações UIT-T para sistemas de IPTV e sistemas IBB (ITU-T 2014).

Este capítulo apresenta as principais linhas de pesquisa do TeleMídia e detalha os projetos em andamento relacionados ao tema de videocolaboração (Seção 1.2). Além disso, o capítulo apresenta a visão de futuro dos membros do laboratório (Seção 1.3) sobre como avanços tecnológicos recentes podem viabilizar novos cenários de videocolaboração, e sobre como cenários hoje inviáveis podem motivar o desenvolvimento de tecnologias ainda inexistentes. Ao final do capítulo, são apresentadas as conclusões (Seção 1.4).

7.2. Linhas de pesquisa e projetos atuais

O laboratório TeleMídia possui três linhas de pesquisa relacionadas à videocolaboração: (1) *Representação e autoria de cenas multimídia interativas*; (2) *Arquitetura e implementação de middlewares para sistemas multimídia*; e (3) *Experiências imersivas*. Esta seção apresenta os objetivos, resultados e possíveis aplicações dos trabalhos desenvolvidos em cada uma dessas linhas no contexto de vídeos interativos e de videocolaboração.

7.2.1. Representação e autoria de cenas multimídia

O objetivo desta linha de pesquisa é desenvolver modelos, linguagens e ferramentas para a autoria de aplicações multimídia interativas. Uma aplicação multimídia interativa é composta pela apresentação sincronizada de múltiplos objetos de mídia (textos, áudios, vídeos, animações e, mais recentemente, mídias não-tradicionais, como mídias olfativas, táctis, etc.) e responde à interação do usuário (ou usuários, no caso de aplicações multiusuário).

Modelos e linguagens projetados especificamente para a construção de aplicações multimídia interativas (ou *modelos e linguagens multimídia*) são uma base adequada para a especificação de diversos tipos de aplicações de vídeo interativo e videocolaboração. A principal vantagem de se usar esses modelos e linguagens em vez de linguagens de programação de propósito geral, é que, em geral, eles oferecem abstrações de alto nível mais próximas do domínio de aplicações multimídia interativas.

Por exemplo, em uma linguagem multimídia, especificar o sincronismo entre o vídeo principal da aplicação e objetos de mídia secundários é normalmente uma tarefa simples, já que a linguagem oferece construções deliberadamente projetadas para esse fim. (Em um contexto de videocolaboração, podemos imaginar que o vídeo principal é aquele compartilhado entre os usuários colaboradores, e os objetos de mídia secundários representam os controles necessários para realizar uma determinada tarefa.)

Por outro lado, implementar cenários similares em uma linguagem de programação de propósito geral exige que o programador escreva essa mesma especificação em um nível de abstração mais baixo, utilizando primitivas mais próximas do modelo de execução da máquina, e distantes do domínio da aplicação. Assim, a tendência é que o programador acabe “reinventando a roda”. Isto é, em vez de trabalhar diretamente na especificação da solução do problema (no caso, na especificação do sincronismo entre o vídeo e os objetos secundários) usando um vocabulário no nível de abstração adequado, ele acaba implementando esse vocabulário.

Há na literatura inúmeros exemplos de aplicações baseadas em vídeo que tiram proveito de linguagens e modelos multimídia. Algumas são aplicações locais para construção e reprodução de artefatos interativos em um mesmo dispositivo (Meixner 2017), enquanto outras são aplicações distribuídas onde diversos usuários colaboram simultaneamente para a realização de uma determinada tarefa (Jansen et al. 2015).

No caso de aplicações de vídeo interativo e de videocolaboração, os requisitos de distribuição de conteúdo e de controle tornam particularmente atrativos o uso de modelos e linguagens com suporte a múltiplos dispositivos (Van Deventer et al. 2016) e múltiplos usuários (Guedes, Á. L. et al. 2016). Um dos modelos que possui tais características é o NCM, base conceitual da linguagem NCL. Ambos, NCM e NCL, foram desenvolvidos no laboratório TeleMídia.

O restante desta seção apresenta o NCM e a NCL e discute a aplicabilidade dessas tecnologias no contexto de videocolaboração. Em seguida, são apresentadas as principais ferramentas de apoio à criação de aplicações NCL desenvolvidas no laboratório: NCL Eclipse, ISB Designer e NCL Composer.

NCM e NCL

O modelo NCM (*Nested Context Model* ou Modelo de Contextos Aninhados) (Casanova et al. 1991), cuja primeira versão data de 1991, é um modelo conceitual para a especificação e apresentação de documentos hipermídia. O NCM combina as ideias usuais de nós e elos hipermídia com o conceito de contextos aninhados, que permite a definição de escopo e uma melhor organização das informações representadas. A Figura 7.1 apresenta as principais entidades do NCM 3.0 (versão atual do modelo).

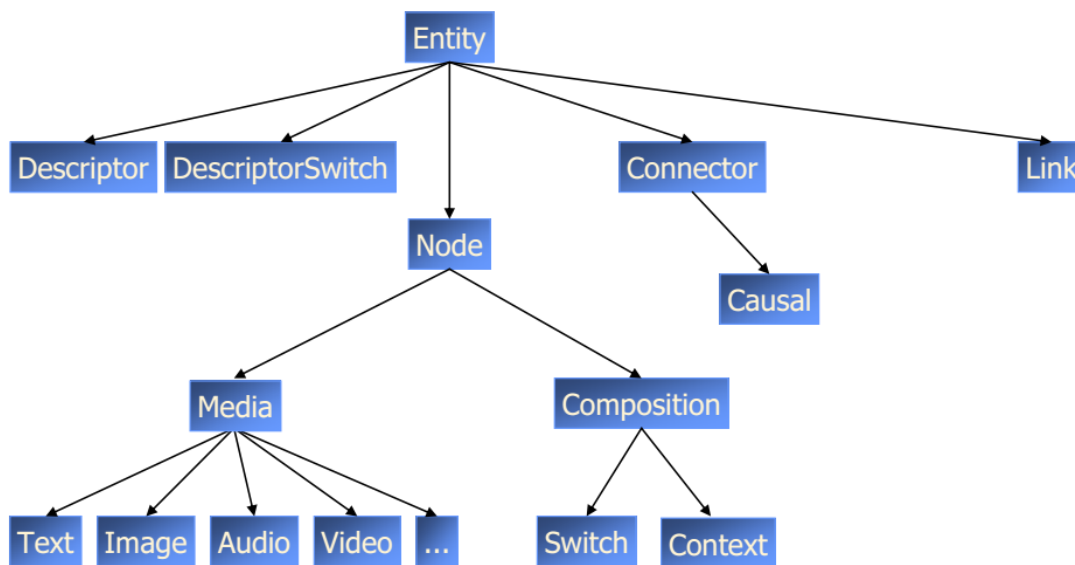


Figura 7.1. Entidades básicas do NCM 3.0 (Luiz Fernando Gomes and Barbosa 2009).

NCL (*Nested Context Language* ou Linguagem de Contextos Aninhados) (Soares et al. 2006) é uma linguagem declarativa para a especificação de aplicações multimídia interativas baseada no NCM. A primeira versão da linguagem, NCL 1.0, foi desenvolvida em 2000 (Antonacci et al. 2000), e a versão mais recente, NCL 3.0, foi lançada em 2006 (ABNT 2016). NCL 3.0 é hoje a linguagem declarativa padrão para a especificação de aplicações interativas no sistema brasileiro de TV digital (ABNT 2016) e recomendação UIT-T para sistemas IPTV (ITU-T 2014).

Um programa NCL 3.0 é um código XML (*eXtensible Markup Language*) que descreve uma apresentação multimídia interativa. Essa descrição consiste, basicamente, da especificação de um conjunto de objetos mídia (textos, imagens, áudios, vídeos, etc.) e de um conjunto de elos de sincronismo, que determinam o comportamento dos objetos de mídia durante a apresentação. Objetos de mídia e elos agrupados em contextos, os quais podem ser aninhados, possibilitando uma melhor organização do código. Em NCL 3.0, os elos são relacionamentos causais, isto é, pares da forma condição-ação que relacionam eventos.

Além de primitivas para descrição do conteúdo (mídias) e do comportamento de uma aplicação multimídia, NCL 3.0 possui funcionalidades adicionais particularmente úteis para o desenvolvimento de aplicações de videocolaboração. A primeira delas é o suporte à aplicações cuja apresentação pode ser distribuída em múltiplos dispositivos em uma rede local. Em NCL 3.0, esse tipo de aplicação é especificado de forma declarativa—o autor da aplicação precisa apenas especificar as classes de dispositivos suportadas e o em qual classe de dispositivo cada objeto será apresentado (Batista 2013; Soares et al. 2009). O suporte da linguagem à múltiplos dispositivos é discutido em detalhe na Seção 1.2.3.

A segunda funcionalidade de NCL 3.0 interessante para o desenvolvimento de aplicações de videocolaboração é o suporte da linguagem à comandos de edição.

Comandos de edição são instruções que modificam dinamicamente o código ou o estado de uma aplicação NCL. Há comandos para inserir ou remover objetos de mídia, inserir ou remover elos, iniciar ou parar partes específicas da aplicação, etc. Tais comandos podem ser gerados tanto pela emissora (no caso de aplicações de TV digital terrestre ou IPTV) quanto programaticamente usando *scripts* Lua (Sant'Anna et al. 2008), a linguagem de *script* de NCL. Adicionalmente, também é possível usar Lua para enviar e receber dados diretamente via TCP/IP e HTTP (Sant'Anna et al. 2008). Tais funcionalidades, possibilitam, por exemplo, o desenvolvimento de aplicações de TV Social (Cesar and Geerts 2011; Fachine 2017), onde usuários podem colaborar enquanto assistem um determinado programa de TV, e.g. comentando

Também podem ser úteis para o desenvolvimento de aplicações de videocolaboração extensões recentes (ainda não padronizadas) que adicionam à NCL 3.0 suporte à especificação declarativa de mídias estereoscópicas (Azevedo 2015) e interação multimodal (Guedes, A. L. et al. 2016; Guedes, Á. L. et al. 2016). Ambas essas extensões são resultado de teses desenvolvidas recentemente no laboratório TeleMídia e são detalhadas na Seção 1.2.3.

Além de desenvolver pesquisas relacionadas à extensão da NCL para novos cenários de aplicação, o laboratório TeleMídia mantém um projeto dedicado à evolução do padrão ABNT e da recomendação UIT-T da NCL. O objetivo principal desse projeto é desenvolver o perfil completo e o perfil restrito da versão 4.0 de NCL.

O *perfil completo* da NCL 4.0 deverá incluir:

1. um melhor suporte à aplicações cientes de contexto;
2. um melhor suporte à aplicações distribuídas em múltiplos dispositivos, incluindo um melhor suporte à serviços de TV social;
3. suporte declarativo ao desenvolvimento de aplicações com conteúdo dinâmico;
4. suporte à inclusão de objetos e eventos 3D;
5. suporte à relacionamentos de restrição, em adição aos de causalidade;
6. suporte a templates (integrando as funcionalidades de TAL (Soares Neto et al. 2012)).

Já o *perfil restrito* da NCL 4.0, chamado de Raw Profile, deverá incluir:

1. a especificação do perfil de NCL 4.0 excluindo todos os elementos redundantes do perfil completo da NCL 4.0 que possam afetar o desempenho e robustez do *player* NCL;
2. o desenvolvimento de um conversor do perfil completo para o perfil Raw;
3. o desenvolvimento de conversores de outras linguagens declarativas para a criação de aplicações multimídia (e.g. SMIL (Bulterman and Rutledge 2016) e) para o perfil NCL 4.0 Raw;
4. a implementação de referência do Ginga-NCL seguindo o perfil Raw;
5. o desenvolvimento de conversores do perfil NCL 4.0 Raw para exibição em *players* de outras linguagens declarativas.

A maioria dos itens acima, em especial os do perfil completo, estão relacionados ao contexto de videocolaboração. Por exemplo, melhor suporte à aplicações cientes de contexto implica em aplicações que se adaptam melhor às peculiaridades de cada usuário, o que é um requisito importante para aplicações multiusuário. Um argumento similar se aplica ao melhor suporte à aplicações distribuídas—um sistema de videocolaboração remota é, por definição, um sistema distribuído.

NCL Eclipse, ISB Designer e NCL Composer

As principais ferramentas de autoria multimídia desenvolvidas pelo laboratório TeleMídia são o NCL Eclipse, o ISB Designer e o NCL Composer (vide Figura 7.2).

O *NCL Eclipse* (Azevedo et al. 2011) é um *plugin* para a IDE Eclipse¹ que provê facilidades para ajudar programadores a desenvolver aplicações NCL. Essas facilidades incluem coloração de código, sugestão de código (*autocomplete*), navegação no código por meio de hiper-elos, visualização programática, verificação sintática e integração com diferentes *players* NCL. Atualmente, o NCL Eclipse é a ferramenta mais utilizada na América Latina para o desenvolvimento de aplicações interativas para o *middleware* Ginga.

O *ISB Designer* (Araújo and Soares 2014) é uma ferramenta de apoio ao projeto de narrativas interativas baseada no conceito de *storyboards* interativos—uma extensão dos *storyboards* tradicionais utilizados por *designers* e profissionais de TV para a descrição de narrativas. A vantagem dos *storyboards* interativos é que eles são fáceis de editar, e permitem que o autor construa narrativas não-lineares de forma gradual (contendo artefatos ainda incompletos) e independente de uma linguagem-alvo específica. O ISB Designer também permite a edição e visualização de cadeias temporais interligadas por eventos não-determinísticos, cada uma delas compondo um *storyboard* com pontos de eventos não-previsíveis (por exemplo, a interação do usuário). A ferramenta permite exportar o projeto para NCL e HTML, e pode ser utilizada de forma *standalone* ou como um *plugin* do NCL Composer (Araújo and Soares 2014).

O *NCL Composer* (Azevedo et al. 2012) é um ambiente de autoria para aplicações NCL que facilita o uso dos conceitos definidos pelo NCM e pela NCL por não-programadores. Além da visão textual, que traz funcionalidades similares às do NCL Eclipse, o NCL Composer também traz diferentes visões gráficas, tais como a visão estrutural e de leiaute, que permitem que usuários não-programadores utilizem abstrações visuais para o desenvolvimento de aplicações multimídia. Cada uma dessas visões é implementada como um *plugin* da ferramenta.

Os seguintes *plugins* compõem a distribuição oficial do NCL Composer:

- *Visão textual*: provê as mesmas funcionalidades hoje oferecidas pela ferramenta NCL Eclipse para a edição do código do documento.
- *Visão estrutural*: oferece uma representação gráfica da estrutura do documento.
- *Visão de leiaute*: permite a edição gráfica das propriedades de posicionamento e dimensionamento de objetos de mídia NCL.

¹ Disponível em <http://eclipse.org>.

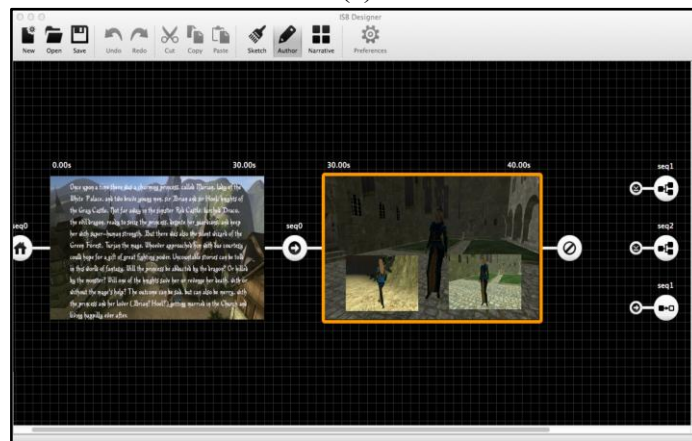
- *Visão de propriedades:* permite a edição das propriedades de objetos de mídia.
- *Plugin de validação:* verifica a sintaxe do documento.
- *Visão de execução:* permite apresentar um documento NCL dentro da própria ferramenta.

```

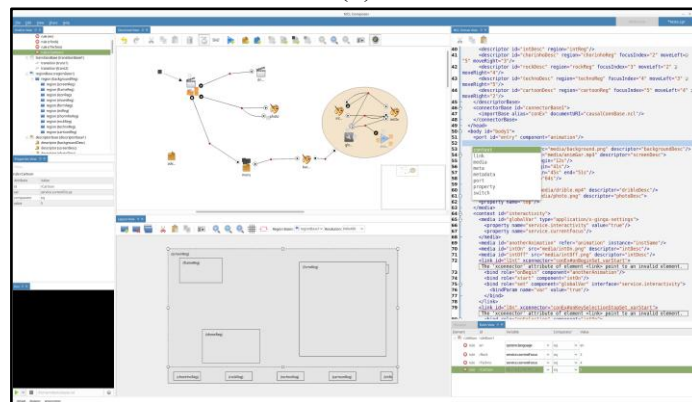
1 <?xml version="1.0" encoding="ISO-8859-1"?>
2 <!DOCTYPE ncl SYSTEM "http://www.ncl.org.br/NCL3.0/EDTVProfile">
3 <!--=====
4 ! CABEÇALHO DO PROGRAMA
5 =====>
6 <thead>
7 <!--=====
8 ! BASE DE REGRAS:
9 ! definem regras utilizadas em switches para a seleção de rbs
10 =====>
11 <ruleBase>
12 <rule comparator="eq" id="r1" value="1" var="opcao"/>
13 <rule comparator="eq" id="r2" value="2" var="opcao"/>
14 <rule comparator="eq" id="r3" value="3" var="opcao"/>
15 <rule comparator="eq" id="r32" value="32" var="opcao"/>
16 <rule comparator="eq" id="r4" value="4" var="opcao"/>
17 </ruleBase>
18 <!--=====
19 ! BASE DE REGIÕES:
20 ! define as regiões na tela onde as mídias são apresentadas
21 =====>
22 <regionBase>
23 <region height="100%" id="rgTV" width="100%" />
24 <region id="rgAgt" width="100%" height="15.5%" top="20%" zIndex="2"/>
25 <region id="rgBola" width="100%" height="40%" top="22%" zIndex="2"/>
26 <region id="rgAresta" width="51%" height="17%" top="13%" left="40%" zIndex="
27 <region id="rgLogo" width="22.5%" height="19%" top="19%" zIndex="
28 <region id="rgResultado" width="50%" height="50%" left="40%" top="30%" zInd
29 <region id="rgFundo" width="22.5%" height="19%" top="20%" zIndex="2"/>
30 <region id="rgOpcao1" width="90%" height="98%" left="5%" top="11%" zIndex=
31 <region id="rgOpcao2" width="90%" height="98%" left="5%" top="11%" zIndex=
32 <region id="rgOpcao3" width="90%" height="98%" left="5%" top="11%" zIndex=
33 <region id="rgOpcao3" width="90%" height="98%" left="5%" top="11%" zIndex=
34

```

(a)



(b)



(c)

Figura 7.2. Ferramentas de autoria multimídia desenvolvidas no laboratório TeleMídia: (a) NCL Eclipse, (b) ISB Designer, e (c) NCL Composer.

A arquitetura do NCL Composer é baseada em um núcleo mínimo, que gerencia as mensagens trocadas pelos diferentes *plugins*, e que pode ser estendido por *plugins* de terceiros. Os *plugins* de terceiros podem ser novas visões (gráficas ou textuais) para auxílio à autoria, ou novos geradores de dados em uma dada sintaxe de transferência para os vários sistemas de TV digital, ou em um dado formato de armazenamento. Atualmente, existem diversos *plugins* desenvolvidos por terceiros para facilitar a criação de aplicações interativas de domínio específico (Bernal and Mejía 2015; Ferreira and Souza Filho 2014; Mejía et al. 2014; Santos et al. 2013).

O núcleo do NCL Composer tem se mostrado estável, escalável e flexível. Além da manutenção do núcleo e dos *plugins* oficiais mencionados acima, o laboratório TeleMídia mantém as seguintes frentes de evolução do NCL Composer:

1. Implementação de um *plugin* de visão temporal, isto é, uma visão que represente no tempo os eventos associados à apresentação da aplicação.
2. Implementação de *plugins* para a codificação e transmissão da aplicação em sistemas de TV digital, rádio digital e IPTV.
3. Suporte dinâmico à diferentes perfis de usuários (de especialistas a não-programadores) e ambientes (de *high-end* a *low-end*).
4. Suporte à edição interativa da aplicação, isto é, permitir que o autor modifique-a em tempo de execução, enquanto testa o que já foi implementado.
5. Implementação de um *plugin* para autoria WYSIWYG (*What You See Is What You Get*) de aplicações para múltiplos dispositivos.
6. Suporte à distribuição de *plugins* em múltiplos processos e múltiplas máquinas, com foco em suportar a autoria colaborativa de documentos multimídia.

7.2.2. Arquitetura e implementação de sistemas multimídia

O objetivo desta linha de pesquisa é desenvolver exibidores (*players*) e arquiteturas de referência que permitam executar os modelos e linguagens multimídia apresentados na Seção 1.2.1 de forma correta e eficiente. Também está relacionado a esta linha de pesquisa o desenvolvimento de algoritmos e mecanismos de sinalização e de transporte que permitam obter maior qualidade de experiência (QoE) tanto em aplicações locais quanto distribuídas.

Middleware Ginga

As abstrações oferecidas por modelos e linguagens multimídia são úteis para a modelagem teórica de vários tipos de aplicações em videocolaboração. Entretanto, é o desenvolvimento de *players* (ou exibidores) que torna esses modelos e linguagens úteis na prática. Um exibidor de um modelo ou linguagem multimídia é um programa que interpreta as instruções de alto nível do documento de entrada e as transforma em uma apresentação multimídia equivalente.

Apesar de essencial, o exibidor é apenas um dos componentes de uma solução multimídia integrada que garanta uma boa QoE para o usuário. Tal solução integrada também deve incluir módulos adicionais que, por exemplo, dêem suporte para: (1) a recepção e transmissão de fluxos de dados (conteúdo e controle) via redes diversas; (2) a sincronização dos fluxos recebidos e enviados; (3) a descoberta e o pareamento de

dispositivos em redes locais; e (4) a distribuição da aplicação entre esses diversos dispositivos.

Nesse sentido de solução multimídia integrada, a principal contribuição do laboratório TeleMídia é o *middleware* Ginga (Gomes Soares et al. 2010), que hoje é o *middleware* padrão do sistema brasileiro de TV digital terrestre (ABNT 2016) e recomendação UIT-T para sistemas IPTV (ITU-T 2014). O *middleware* Ginga é o resultado de anos de pesquisa em sistemas multimídia e envolveu o esforço de vários alunos de graduação, mestrado e doutorado que passaram pelo laboratório TeleMídia, além de diversas contribuições de outros grupos de pesquisa.

A Figura 7.3 apresenta a arquitetura do *middleware*. O Ginga é dividido em dois módulos principais: Common-Core e Ginga-NCL. O Common-Core é o módulo responsável por instanciar os *players* dos tipos de mídia suportados em uma determinada plataforma (TV digital terrestre, rádio digital ou IPTV), intermediar a comunicação com as diversas redes e dispositivos remotos, e gerenciar o ciclo de vida das aplicações recebidas. Já o Ginga-NCL é o exibidor NCL propriamente dito, isto é, o módulo responsável por apresentar as aplicações NCL.

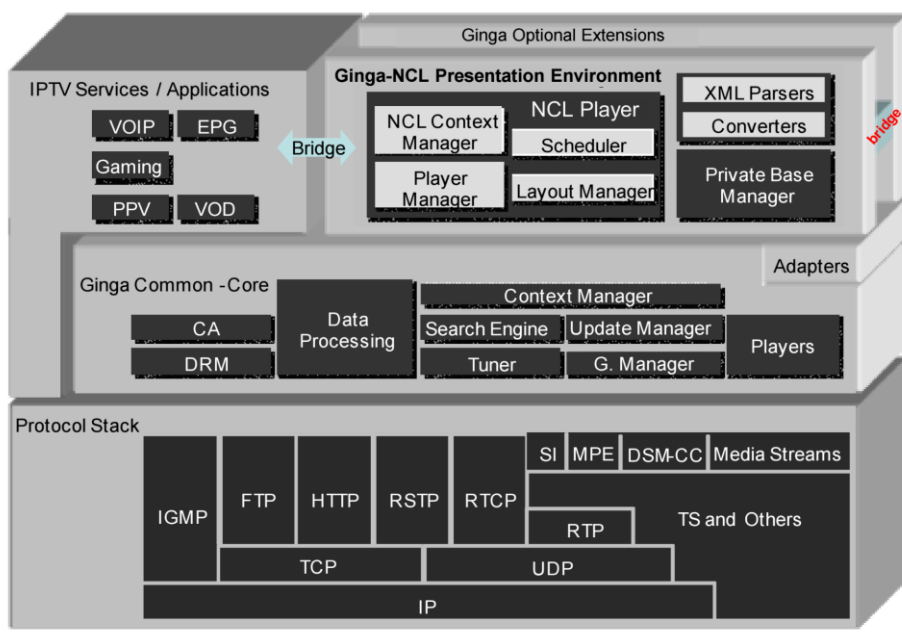


Figura 7.3. Arquitetura do *middleware* Ginga.

Dando continuidade ao desenvolvimento das próximas gerações do *middleware* Ginga, atualmente o laboratório TeleMídia mantém o projeto Ginga 2.0. Esse projeto tem como objetivo principal aperfeiçoar a implementação de referência do Ginga e do exibidor Ginga-NCL, incorporando as novas funcionalidades previstas pela NCL 4.0, além de promover a evolução do *middleware* em diferentes plataformas de TV digital terrestre, rádio digital e IPTV. São objetivos específicos do projeto Ginga 2.0 integrar ao *middleware* suporte às seguintes funcionalidades:

1. Pré-busca inteligente de dados sob demanda (*pulled data*).
2. Pré-busca inteligente de dados sem solicitação (*pushed data*) em carrossel de objetos.

3. Negociação antecipada de QoS (*Quality of Service*) para dados sob demanda (*pulled data*).
4. Extensão do modelo de apresentação do formatador para escalonamento de aplicações iniciadas em qualquer ponto do tempo de sua duração.
5. Atualização dinâmica dos componentes do *middleware*.

Ginga IBB, BemTV e audio watermarking

Outro tema recorrente nas pesquisas do laboratório TeleMídia, relacionado ao tópico de arquitetura e implementação de sistemas multimídia, é o desenvolvimento de *players* de objetos de mídia e de protocolos de transmissão ou sincronização que resolvam problemas específicos de um determinado domínio multimídia. Os exemplos mais recentes de projetos desse tipo desenvolvidos pelo laboratório são: (1) a integração de suporte à IBB (*Integrated Broadcast/Broadband*) ao padrão ABNT e à recomendação UIT-T do Ginga; (2) a biblioteca BemTV para balanceamento de carga via comunicação P2P (*peer-to-peer*) em fluxos de vídeo; e (3) um sistema para sincronismo de mídia em larga escala via *audio watermarking*.

Atualmente, serviços de TV Conectada (ou Smart TV) provêem a distribuição de conteúdo por redes de transmissão de TV (satélite, terrestre ou cabo) e internet (por exemplo, via Ethernet DSL, Ethernet no CATV ou LTE). Entretanto, eles não necessariamente oferecem serviços convergentes que fazem o uso conjunto de conteúdos de ambas as redes de distribuição. Os serviços IBB (ITU 2016) têm sido propostos com o objetivo de suportar uma convergência real entre as redes de transmissão de TV e a internet, e permitir, por exemplo, a vinculação do conteúdo oferecido por redes via satélite, terrestre ou cabo ao conteúdo da internet. Nesse sentido, o *middleware* Ginga também suporta serviços IBB, por meio de sinalizações no fluxo de transporte de TV que permitem a sincronização dos dados no fluxo de transmissão aos serviços de conteúdo da internet. Através desta sinalização é possível, por exemplo, iniciar o portal de conteúdo do radiodifusor do canal correspondente, o que permite que o espectador do canal acesse serviços de vídeo sob demanda relacionados ao conteúdo que está sendo transmitido.

O *BemTV* (Barbosa and Soares 2014) é uma biblioteca desenvolvida no laboratório TeleMídia para o balanceamento de carga na distribuição de fluxos HLS (*HTTP Live Streaming*). A biblioteca utiliza comunicação P2P para compartilhar *chunks* comuns entre os clientes de um mesmo vídeo e, dessa forma, diminuir o número de requisições à CDN (*Content Delivery Network*). Essa técnica é particularmente atrativa nos casos em que o vídeo é captado ao vivo, em alta qualidade, e precisa ser distribuído em tempo real a um grande número de clientes parcialmente sincronizados (por exemplo, partidas de futebol, concertos, etc.).

A Figura 7.4 a seguir apresenta o resultado de um experimento em que dois *players* requisitam *chunks* à CDN. O primeiro *player* utiliza apenas a CDN; o segundo utiliza o modelo híbrido CDN-P2P do BemTV. Enquanto o primeiro *player* (linha contínua) realizou 7457 requisições à CDN, o segundo (linha tracejada) realizou apenas 4482, ou seja, uma redução de 39,89% no total de requisições.

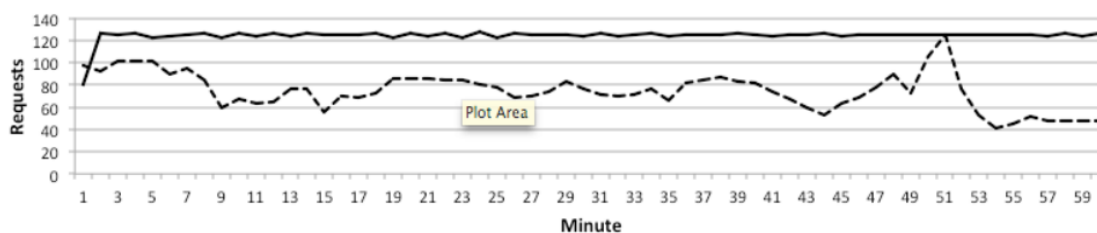


Figura 7.4. Requisições à CDN e requisições aos *peers* (Barbosa and Soares 2014).

Em algumas aplicações a sincronização—e não a distribuição—do conteúdo é que precisa ser feita em larga escala. Esse é o caso, por exemplo, de aplicações em que um relógio global precisa ser distribuído a um grupo de dispositivos confinados (próximos uns dos outros) mesmo na ausência de redes tradicionais. Uma forma de atender a esse requisito (assumindo que os dispositivos possuem microfone) é através de ondas sonoras via *audio watermarking*.

Audio watermarking é uma técnica utilizada para a adição de informações adicionais em fluxos de áudio. É comumente utilizada na indústria para incorporação de metadados em faixas musicais, por exemplo, título da música, autor, copyright, etc. Alguns trabalhos recentes propõem o uso dessa técnica para sincronizar conteúdo multimídia em ambientes controlados. Em (Arnold et al. 2014), por exemplo, o *audio watermarking* é aplicado ao som transmitido pela TV e é utilizado para sincronizar dispositivos móveis. Já em (Lemos et al. 2016) a técnica é utilizada para sincronização local de aplicações Web.

Diferente desses trabalhos, o laboratório TeleMídia tem pesquisado a aplicação de *audio watermarking* para sincronização de dispositivos em ambientes abertos, onde há uma grande aglomeração de pessoas e não há rede, ou a rede não é confiável. Cenários comuns desses ambientes são shows e eventos para multidões em ambientes abertos. São alvos dessa pesquisa a investigação da usabilidade da técnica (incômodo gerado e efeitos da exposição prolongada) e da sua viabilidade tecnológica quando considerada a transmissão de dados em altas frequências, isto é, frequências entre 18KHz e 20KHz.

7.2.3. Experiências imersivas

O objetivo desta linha de pesquisa é estender as representações de cenas (Seção 1.2.1) e exibidores multimídia (Seção 1.2.2) para possibilitarem uma maior imersão dos usuários durante a apresentação. Com esse objetivo, o laboratório TeleMídia vem realizando pesquisas sobre como aumentar a QoE do usuário através de novas formas de visualização e interação. Na parte de visualização, essas pesquisas consideram o uso de múltiplas telas de exibição e de displays 3D. Já na parte de interação, os estudos desenvolvidos consideram o uso de múltiplas modalidades de interação (por exemplo, reconhecimento de gestos e voz, e outras mídias sensoriais em geral.)

Múltiplos dispositivos e telas secundárias

O ambiente de TV é um ambiente de assistência coletiva, isto é, é comum que várias pessoas assistam o mesmo programa ao mesmo tempo. Nesse cenário, quando uma dos telespectadores interage com a TV, por exemplo, para responder um questionário interativo, a qualidade de experiência dos demais telespectadores (aqueles que querem

apenas continuar assistindo o programa principal) pode ser prejudicada. Uma solução para esse problema é delegar a interação à dispositivos secundários (ou de segunda tela) permitindo assim que cada telespectador tenha uma interação personalizada. Exemplos de dispositivos secundários são *smartphones*, *tablets* e *smartwatches*.

O *middleware* Ginga suporta aplicações com dispositivos secundários desde a sua primeira versão (Soares et al. 2009). Para prover esse suporte, o Ginga utiliza um modelo hierárquico de distribuição de conteúdo e de controle entre os múltiplos dispositivos. Os dispositivos secundários registram-se como filhos do dispositivos base (em geral, a TV) e passam a fazer parte de uma classe de dispositivos, conforme ilustrado na Figura 7.5. Dessa forma, o dispositivo pai pode controlar o conteúdo que é exibido nos dispositivos filhos. NCL 3.0 define duas classes de dispositivos: *classe passiva*, em que todos os dispositivos da classe apresentam simultaneamente o mesmo conteúdo; e *classe ativa*, em que a navegação e conteúdo exibido é individualizada para cada dispositivo da classe.

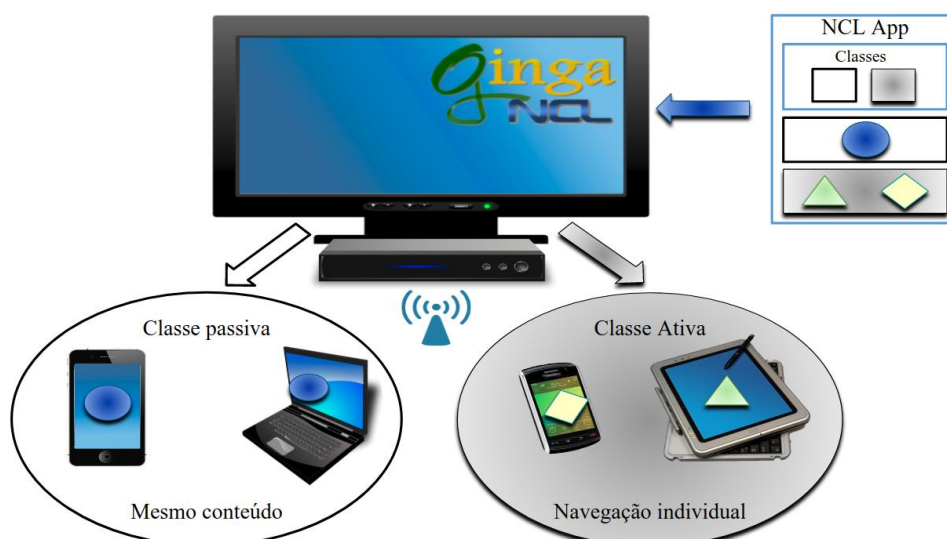


Figura 7.5. Modelo hierárquico de controle de dispositivos do Ginga.

A Figura 7.6 abaixo apresenta exemplos de aplicações que utilizam múltiplos dispositivos secundários desenvolvidas em NCL 3.0 para o *middleware* Ginga.



Figura 7.6. Exemplos de aplicações de segunda tela usando o Ginga.

Diversos tipos de aplicações de video colaboração podem ser implementadas em NCL 3.0 utilizando o suporte da linguagem à múltiplos dispositivos. Isso é especialmente interessante no caso de dispositivos da classe ativa. Por exemplo, durante

a apresentação de um vídeo na tela principal, os usuários da aplicação podem colaborar utilizando seus smartphones para realizar uma determinada tarefa, ou podem entreter-se por meio de jogos colaborativos relacionados com o vídeo principal (como os exemplos da Figura 7.6).

Como discutido na Seção 1.2.1, a próxima versão da linguagem NCL permitirá, além das classes passivas e ativas, que novas classes de dispositivos customizadas sejam criadas pelos autores. Uma proposta inicial de como é possível a definição de novas classes de dispositivos (ainda não padronizada) é discutida em (Batista 2013).

Displays 3D e vídeos 3D

Outras tecnologias com grande potencial para aumentar a imersão e melhorar a colaboração entre usuários em sistemas de videocolaboração são aquelas relacionadas à *displays* e vídeos 3D. Um exemplo de sistemas de videocolaboração com esse tipo de suporte são os sistemas de videoconferência 3D que tiram proveito de tecnologias de *displays* e vídeos 3D estereoscópicos, e que vem sendo pesquisados há bastante tempo.

Displays 3D estereoscópicos são *displays* que fornecem ao usuário dicas binoculares de profundidade através de óculos especiais. Apesar dessa tecnologia existir há bastante tempo, nos últimos anos a produção desse tipo de *display* cresceu consideravelmente impulsionada pela produção massiva de filmes 3D. *Displays* 3D estereoscópicos, entretanto, possuem algumas limitações importantes, por exemplo, a necessidade do uso de óculos especiais e a possibilidade de provocar náuseas em algumas pessoas.

Com o objetivo de solucionar algumas dessas limitações, novas tecnologias de *displays* 3D vêm sendo desenvolvidas. Exemplos dessas tecnologias incluem *head-mounted displays* (HMD), *displays* autoestereoscópicos e *displays* baseados em *light field*. Os dois últimos fornecem dicas binoculares aos usuários sem a necessidade de óculos especiais.

São exemplos de *head-mounted displays* o Google Cardboard (Figura 7.7.a) e o Microsoft HoloLens (Figura 7.7.b). Já o HoloVizio (Figura 7.7.c) é um exemplo de *display light field*. A Figura 7.7.d apresenta, esquematicamente, o funcionamento de um *display light field*: múltiplos projetores são usados para recriar um padrão de raios de luz de tal forma que o espectador possa ver uma imagem 3D sem a necessidade de óculos especiais e com paralaxe de movimento, isto é, o espectador percebe uma imagem diferente na medida em que movimenta a sua cabeça (Lawton 2011).

Simultaneamente, para dar suporte à essas novas tecnologias de *displays* 3D e possibilitar o uso eficiente das redes de distribuição, novas representações de vídeo 3D também têm sido propostas. São exemplos dessas novas representações os formatos de vídeo estereoscópico convencional (Figura 7.8.a), vídeo com profundidade (Figura 7.8.b), vídeo multivista (Figura 7.8.c), vídeo multivista com profundidade (Figura 7.8.d), e vídeo com camadas e profundidade (Figura 7.8.e).

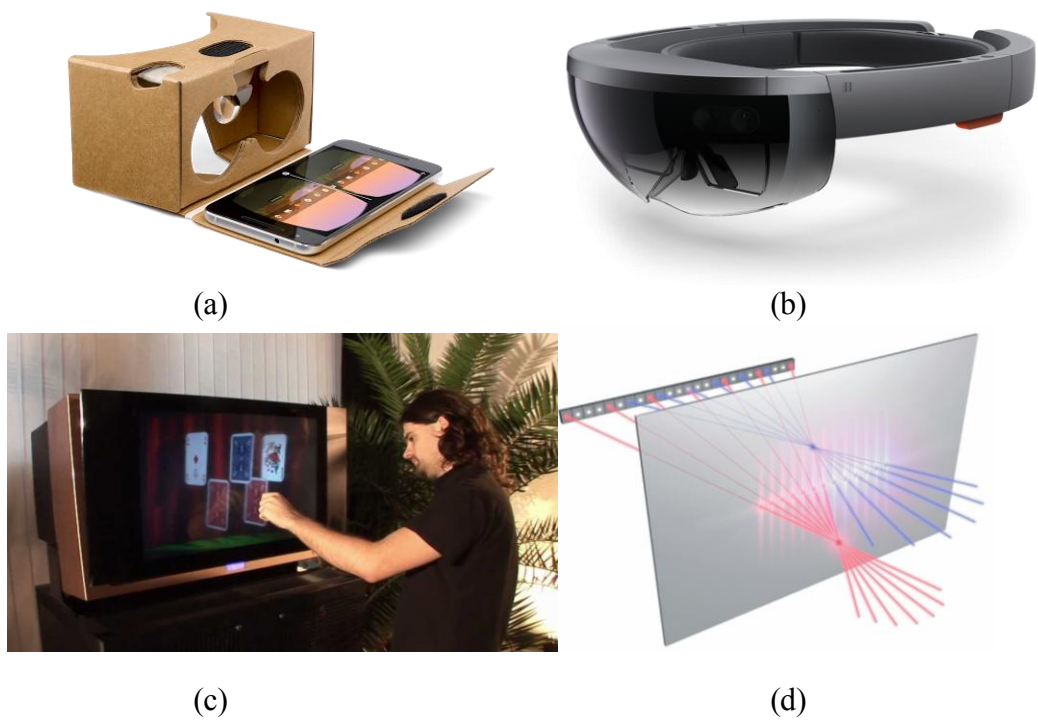


Figura 7.7. Exemplos de displays 3D: (a) Google Cardboard, (b) Microsoft HoloLens, (c) HoloVizio; e (d) esquema de funcionamento um *display light field*.

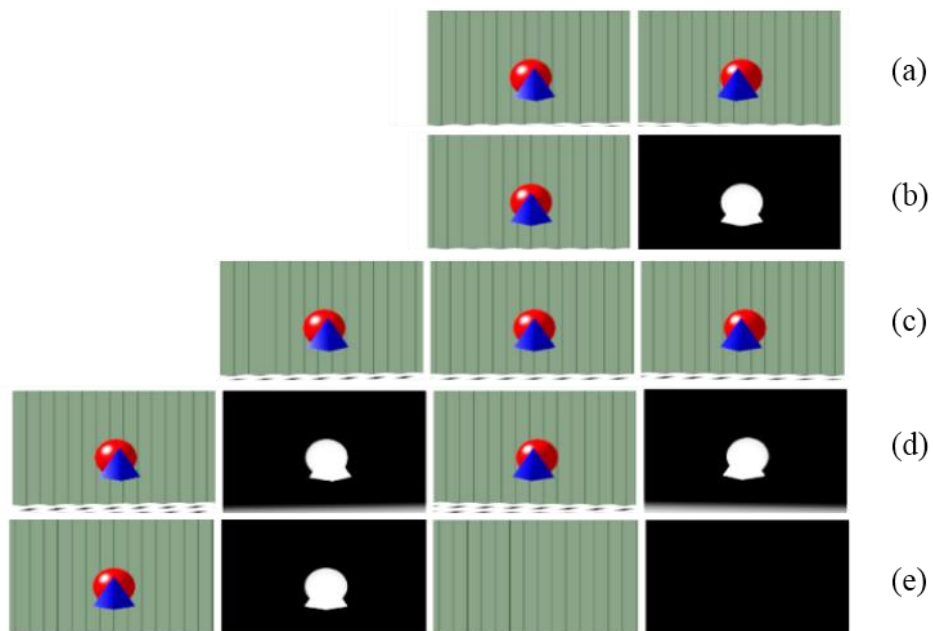


Figura 7.8. Diferentes representações de vídeo 3D.

Desde o início da década de 2010, o laboratório TeleMídia tem aplicado algumas dessas tecnologias de *displays* 3D e vídeos 3D ao contexto de aplicações multimídia interativas. Entre os trabalhos desenvolvidos estão a renderização em tempo

real de vídeos com profundidade (Azevedo et al. 2014) e o suporte à autoria e execução de aplicações multimídia baseadas em profundidade (Azevedo 2015; Azevedo and Lima 2016; Azevedo and Soares 2013).

No que se refere a videocolaboração, quando comparadas aos sistemas baseados em vídeo estereoscópicos convencionais, as tecnologias de vídeo 3D baseadas em profundidade (como as discutidas em (Azevedo et al. 2014)) tem as seguintes vantagens: (1) utilizam menos banda, uma vez que um quadro de profundidade armazena menos informação do que um quadro de cor; (2) permitem ao cliente adaptar a profundidade às características de diferentes *displays* estereoscópicos e preferências do usuário; e (3) permitem ao cliente gerar vistas adicionais da cena, suportando assim *displays* auto estereoscópicos.

A inclusão de cenas interativas em ambientes de videoconferência 3D também é um tema que vem sendo pesquisado no laboratório. Além de possibilitar a reprodução de cenas de forma mais fiel ao mundo real e permitir cenários de realidade aumentada, onde objetos virtuais podem ser inseridos em vídeos 3D capturados, outra oportunidade interessante ao integrar-se cenas interativas em ambientes de videoconferência 3D é a possibilidade do uso da profundidade real para ressaltar características da interface com o usuário (Azevedo 2015).

Interações multimodais e mídias sensoriais

Recentes avanços em tecnologias de reconhecimento de sinais e o uso de novos dispositivos como sensores e atuadores vêm impulsionando o desenvolvimento de interfaces de usuário não-tradicionais, como, por exemplo, as interfaces multimodais. Interfaces multimodais são caracterizadas pelo uso (possivelmente simultâneo) de múltiplas modalidades dos sentidos humanos e por poderem combinar múltiplos dados de entrada (e.g., fala, caneta, toque, gesto, olhar e movimentos de cabeça e corpo) e saída (e.g., sintetização de voz e efeitos sensoriais) (Turk 2014). Comparada às tecnologias atuais de mouse, teclado e *displays*, as tecnologias de interface multimodal permitem o desenvolvimento de interfaces com usuários mais naturais para a linguagem e comportamento humano.

Até recentemente, entretanto, linguagens multimídia como NCL e HTML, não suportavam interações multimodais. Tradicionalmente, essas linguagens privilegiam a criação de aplicações que usam elementos audiovisuais e que adotam um paradigma de interação WIMP (*Windows, Icons, Menus, Pointers*), pressupondo o uso de dispositivos de interações tradicionais como teclado, mouse e *touchscreens*.

Relacionado ao tema de interações multimodais, o laboratório TeleMídia mantém uma frente de pesquisa que visa incorporar extensões a linguagens multimídia que permitam: (1) o uso de novas modalidades de entrada, como reconhecimento de voz e de gestos através de descrições de reconhecimento como SRGS e GML; (2) o uso de novas modalidades de saída, como fala sintetizada e atuadores através de descrições como o SSML (*Speech Synthesis Markup Language*) (W3C 2004) —linguagem para especificação de aplicações de sintetização de voz— e SEDL (*Sensory Effect Description Language*) (ISO 2016) —linguagem parte da suíte de padrões MPEG-V para a especificação de aplicações que descrevem efeitos sensoriais como iluminação, ventilação, vibração, etc.; e (3) a especificação da sincronização entre essas diferentes modalidades de entrada e saída.

Como resultado preliminar dessas pesquisas, propostas de extensões multimodais para NCL 3.0 são detalhadas em (Guedes, A. L. et al. 2016). A Figura 7.9 apresenta uma versão alternativa da aplicação com múltiplos dispositivos da Seção 1.2.2 (Figura 7.5), agora com suporte à interação por gestos.



Figura 7.9. Exemplo de aplicação multimodal com reconhecimento de gestos.

Relacionado especificamente a novos tipos de atuadores, em (Guedes, Á. et al. 2016), é discutido um exemplo de aplicação que apresenta um vídeo de forma sincronizada a efeitos de iluminação, cheiro e vento (Figura 7.10). Enriquecer aplicações multimídia com o uso de objetos que estimulam outros sentidos de forma sincronizada pode contribuir para aumentar a sensação de imersão dos usuários dessas aplicações. Tais aplicações são algumas vezes chamadas de aplicações multi-sensoriais ou *MulseMedia* (Sulema 2016),

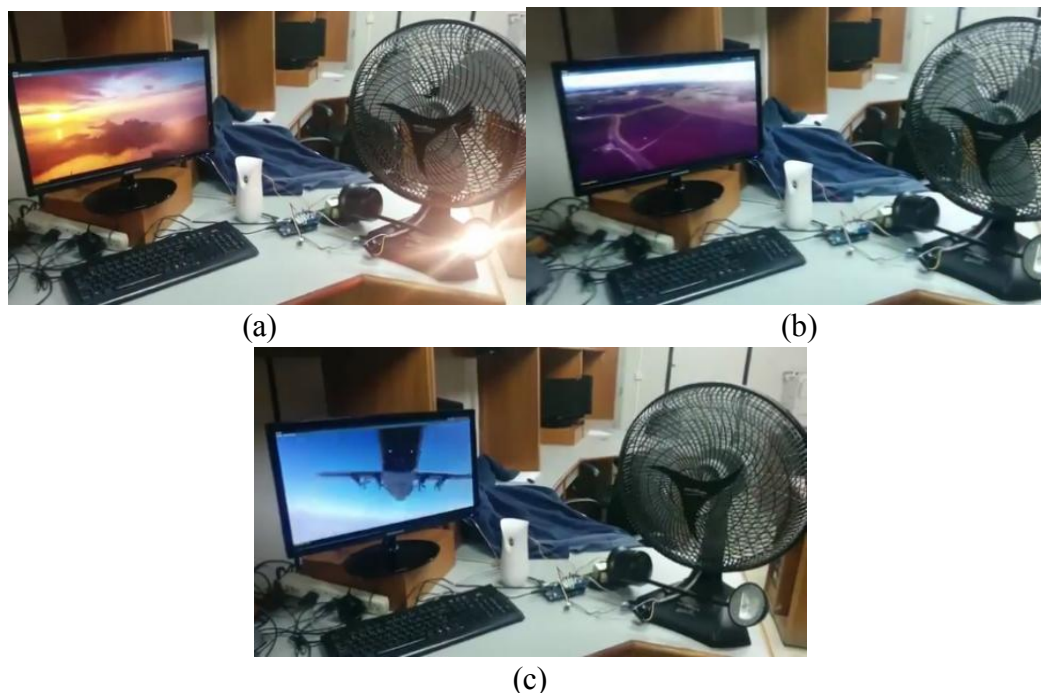


Figura 7.10. Aplicação com efeitos sensoriais de luz (a), cheiro (b) e vento (c).

7.3. Visão de futuro em videocolaboração

Considerando as tecnologias e pesquisas discutidas nas seções anteriores, esta seção apresenta quatro possíveis cenários de evolução para aplicações de vídeo interativo e de videocolaboração.

7.3.1. Videocolaboração em ambientes domésticos

A comunicação intermediada por computador tem se tornado uma parte indispensável das interações diárias da maioria das pessoas. Em ambientes empresariais e de telemedicina, sistemas de videoconferência de alta qualidade e telepresença, por exemplo, já são uma realidade. Tais sistemas, em geral, são implantados utilizando-se redes privadas gerenciadas e hardware dedicado. No entanto, avanços atuais em infraestrutura de rede e poder computacional vêm possibilitando o desenvolvimento de novas aplicações de videocolaboração também em ambientes domésticos.

Cenários de videocolaboração usando redes locais—e.g. aplicações de segunda tela em que os usuários estão em uma mesma rede local (Buchner et al. 2014)—já são possíveis de serem implantados com boa qualidade de serviço, incluindo aqueles que demandam sincronismo fino entre os objetos a serem apresentados nos diversos dispositivos. Entretanto, a manutenção desse sincronismo em aplicações em que usuários estão geograficamente distribuídos e usando redes não-gerenciadas ainda é um problema para a tecnologia atual.

Aplicações de *Social TV* são exemplos de aplicações de videocolaboração em que o suporte ao sincronismo fino mesmo entre usuários geograficamente distribuídos é indispensável. Em aplicações de Social TV, usuários geograficamente distantes assistem ao mesmo conteúdo ao vivo enquanto interagem uns com os outros em tempo real. O YouTube, por exemplo, já suporta transmissões de eventos ao vivo e permite que os espectadores dessas transmissões troquem mensagens enquanto assistem a um vídeo. Porém, as mensagens e o vídeo do YouTube não são sincronizados, e não há qualquer garantia de que todos os clientes estão sincronizados (assistindo exatamente o mesmo quadro do vídeo). Possibilidades de que os usuários possam controlar a transmissão junto com seus pares, por exemplo, avançando e retrocedendo o vídeo para que possam comentar sobre um determinado evento da transmissão, também são exemplos de desenvolvimentos futuros em videocolaboração e *Social TV*.

Outro exemplo de aplicação de videocolaboração que pode se beneficiar do suporte ao sincronismo fino entre usuários geograficamente distribuídos são os ambientes de aprendizagem online (*e-learning*). Atualmente plataformas como Coursera², MIT OpenCourseWare³ e Udacity⁴ oferecem serviços de vídeo interativo em salas de aula virtuais em que professores e alunos interagem por meio de fóruns (interação tipicamente assíncrona). Esses serviços poderiam ser aperfeiçoados através de técnicas que garantem que professores e alunos compartilhem a mesma experiência remotamente, de forma que as mensagens trocadas e o conteúdo exibido estejam sempre sincronizados em todos os clientes, o que não ocorre hoje.

² <https://www.coursera.org/>

³ <https://ocw.mit.edu/>

⁴ <https://www.udacity.com/>

7.3.2. Videoconferência 3D e telepresença

Há pouca dúvida de que um dos caminhos na evolução dos sistemas de videoconferência é a utilização cada vez maior de tecnologias de *displays* e vídeo 3D, as quais estão constantemente evoluindo. Esses avanços têm como objetivo promover uma experiência cada vez mais próxima da comunicação face-a-face que temos diariamente.

Os vários avanços dos últimos anos em tecnologias de captura, codificação, transmissão, renderização e *displays* 3D de fato já tornou viável o uso de tecnologias estereoscópicas tradicionais em sistemas de videoconferência de baixo custo. Mesmo assim, tecnologias estereoscópicas tradicionais ainda não são utilizadas em larga escala em videoconferência, principalmente, quando consideramos cenários domésticos. De fato, o principal motivo de ainda não termos o uso rotineiro dessas tecnologias parece estar relacionado à baixa QoE dos *displays* estereoscópicos tradicionais (vide Seção 1.2.3.). Nesse sentido, em um curto e médio prazo possivelmente veremos aplicações de videoconferência baseadas principalmente em *head-mounted displays* e *displays* baseados em *light field*, os quais resolvem alguns dos problemas dos *displays* estereoscópicos tradicionais.

Head-mounted displays possibilitam uma sensação de imersão bem maior do que os *displays* estereoscópicos tradicionais, o que é bastante interessante para aplicações de videoconferência 3D e telepresença. De fato, atualmente, existe bastante interesse no uso de *head-mounted displays* e vídeos 360/VR em videoconferência. Muito desse interesse está relacionado à disponibilidade de vários novos tipos de *head-mounted displays* de baixo custo e com boa qualidade. Além disso, a crescente disponibilização de *smartphones* também tem permitido o reuso de tais dispositivos como telas em dispositivos VR de baixo custo, e.g., Google Cardboard.

Atualmente, pesquisadores e grandes *players* da indústria (Google e Facebook) já estudam e desenvolvem tecnologias relacionadas à captura, transmissão e renderização de vídeos 360/VR. Por exemplo, várias tecnologias para codificação e transmissão de vídeos 360 (e.g., por meio de técnicas adaptativas (Hosseini and Swaminathan 2016)) têm sido desenvolvidas e podem ser facilmente adaptadas para sistemas de videoconferência. Essas tecnologias, principalmente aquelas baseadas em Web, têm potencial para se tornarem produtos de videoconferência e irão poderão, em breve, estar disponíveis ao usuário final.

Com relação aos *displays* baseados em *light field*, a principal vantagem desses *displays* é que é possível ter uma visão estereoscópica e com paralaxe de movimento sem a necessidade de vestir nenhum dispositivo (e.g. óculos). Embora tais *displays* já estejam disponível em pequena escala, e.g. HoloVizio (Balogh et al. 2007), eles ainda são caros e requerem muito mais dados do que os *head mounted displays*. De qualquer forma, há um crescente interesse de pesquisa em codificação e compressão tanto de representações para *light fields* (Perra 2016; Zare et al. 2017) quando de representações para *displays* múltiplas vistas com profundidade (Purica et al. 2016; Zhang et al. 2017) que, em um médio prazo, podem viabilizar o uso dessa tecnologia também em aplicações de videoconferência.

7.3.2. Videocolaboração com experiências multi-sensoriais

Atualmente, alguns cinemas, também chamados de cinemas 4D, buscam oferecer uma sensação realista para usuários utilizando, além de vídeos e *displays* 3D, também efeitos sensoriais adicionais, tais como vibração de assentos, iluminação e cheiro. Tais experiências, entretanto, ainda estão restritas a teatros especializados, com dispositivos específicos e que não possuem a capacidade de transmissão de conteúdo ao vivo. Baseado nesses exemplos hoje já existentes, é possível vislumbrar que essas novas experiências sensoriais também possuem grande potencial de aumentar a imersão em sistemas de videocolaboração ao permitir ao público remoto uma experiência mais próxima à de estar no local do evento. Exemplos de aplicações que podem se beneficiar dessa maior imersão são as transmissões ao vivo de eventos esportivos e artísticos.

A reprodução realista de um local em outro pode ser alcançada com a apresentação coordenada do conteúdo audiovisual e de objetos reais, que podem ser controlados por meio de sensores e atuadores. Enquanto os sensores capturam características do ambiente, os atuadores reproduzem essas características, tentando assim reproduzir a experiência sensorial do local de origem no local de destino. Exemplos de atuadores que podem ser integrados em uma aplicação de videoconferência são: áudio direcional, vibração, efeitos luminosos e vento.

Entretanto, o uso de forma efetiva desses efeitos em ambientes de videocolaboração, ainda depende de estudos mais detalhados de como os usuários percebem e em quais circunstâncias tais efeitos sensoriais adicionais podem de fato contribuir para aumentar a qualidade de experiência dos usuários. Adicionalmente, efeitos sensoriais como aqueles baseados em cheiro e paladar possuem requisitos de sincronismo que ainda estamos longe de entender completamente e é difícil acreditar que poderão ser utilizados em ambientes reais em um curto período de tempo.

7.3.4. Aplicações multimídia sensíveis à semântica do conteúdo

Recentes avanços nas áreas de inteligência artificial e visão computacional buscam reconhecer automaticamente o conteúdo semântico de vídeos. Algumas abordagens utilizam aprendizagem de máquina para detectar características visuais de baixo nível no vídeo (Laptev et al. 2008; Ye et al. 2015) enquanto outras utilizam modelos semânticos (ontologias) para descrever conceitos de alto nível (Liu et al. 2013; Ye et al. 2015). Apesar desses avanços, o desenvolvimento de aplicações multimídia ainda foca na captura e apresentação de conteúdo audiovisual e ignora a semântica desse conteúdo.

Aplicações multimídia sensíveis à semântica do conteúdo são aplicações que conseguem reconhecer os conceitos que estão sendo apresentados e disparar eventos em decorrência disso. Por exemplo, ao reconhecer um monumento histórico no vídeo que está sendo apresentado, o *player* de um portal de vídeos poderia apresentar textos, links ou vídeos relacionados a esse monumento.

Os principais desafios relacionados ao desenvolvimento desse tipo de aplicação são, primeiro, como fazer a aplicação entender o que é o conceito “monumento histórico”? E como fazê-la reconhecer o objeto “monumento histórico” no conteúdo no vídeo? A primeira questão é explorada pela comunidade de Web semântica, que busca definir vocabulários e ontologias que representam conceitos e relações semânticas entre

eles. Já a segunda questão, é estudada pela área de visão computacional que pesquisa técnicas de reconhecimento de padrões para classificação automática de conteúdo.

Outra possibilidade interessante é desenvolvimento de aplicações que, além da estrutura e do conteúdo audiovisual, considerem a representação e a extração em tempo real da semântica do conteúdo apresentado ou do conteúdo gerado pelos usuários. Por exemplo, em sessões de videoconferência de grupos de trabalho ou de ensino a distância, o sistema poderia reconhecer os conceitos que estão sendo discutidos e exibir automaticamente informações adicionais sobre esses conceitos ou conceitos relacionados.

7.4. Conclusão

Desde a sua fundação o laboratório TeleMídia vem atuando em pesquisas em sistemas multimídia e hipermídia, focando-se especialmente em problemas de representação de cenas multimídia interativas e no projeto e implementação de arquiteturas para sistemas multimídia.

Relacionado à representação de cenas multimídia interativas, as pesquisas do laboratório resultaram no modelo conceitual NCM, na sua linguagem NCL e em várias ferramentas de autoria. Atualmente, diversos trabalhos do laboratório estão focados na evolução tanto do modelo NCM quanto da linguagem NCL, e no desenvolvimento de novas ferramentas e aprimoramento das já existentes.

Relacionado às arquiteturas de sistemas multimídia, as pesquisas do laboratório estão voltadas principalmente para o acesso convergente e sincronizado de diferentes canais de distribuição (*Integrated Broadband Broadcast*), e para cenários que envolvam experiências imersivas. Esses cenários consideram o uso de experiências envolvendo múltiplos dispositivos e o uso de novas modalidades de mídia e interação, e.g. via vídeos 3D, efeitos sensoriais e interação multimodal.

No futuro, o laboratório pretende pesquisar cenários de sistemas de multimídia e videocolaboração que ofereçam uma experiência de comunicação cada vez mais próxima àquela da experiência face-a-face, ou mesmo cenários que superem a experiência real. Esses novos cenários se apoiam na visão de *Sincronismo das Coisas* (Soares 2014a, 2014b), que considera que o requisito de sincronismo pode ocorrer em diferentes níveis em um sistema multimídia, por exemplo, entre fluxos mídia oriundos de diferentes canais distribuição (IBB), entre dispositivos geograficamente distribuídos, entre interações de usuário via novas modalidades entrada (voz e gestos), etc. Além disso, os novos cenários consideram sistemas que se adaptam dinamicamente dependendo da semântica do conteúdo que está sendo apresentado e da semântica da entrada fornecida pelos usuários do sistema.

Referências

- ABNT (2016). ABNT NBR 15606-2:2016 EN Digital terrestrial television - Data coding and transmission specification for digital broadcasting Part 2: Ginga-NCL for fixed and mobile receivers - XML application language for application coding. <http://www.abntcatalogo.com.br/norma.aspx?ID=361857>.
- Antonacci, M. J., Saade, D. C. M., Rodrigues, R. F. and Soares, L. F. G. (jun 2000). NCL: Uma Linguagem Declarativa para Especificação de Documentos Hipermídia na Web. In *VI Simpósio Brasileiro de Sistemas Multimídia e Hipermídia - SBMidia2000*.
- Araújo, E. C. and Soares, L. F. G. (2014). Designing iDTV Applications through Interactive Storyboards. In *Proceedings of the 20th Brazilian Symposium on Multimedia and the Web - WebMedia '14*. <http://dx.doi.org/10.1145/2664551.2664572>.
- Arnold, M., Chen, X.-M., Baum, P., Gries, U. and Doerr, G. (2014). A Phase-Based Audio Watermarking System Robust to Acoustic Path Propagation. *IEEE Transactions on Information Forensics and Security*, v. 9, n. 3, p. 411–425.
- Azevedo, R. G. A., Araújo, E. C., Lima, B., Soares, L. F. G. and Moreno, M. F. (2012). Composer: meeting non-functional aspects of hypermedia authoring environment. *Multimedia tools and applications*, v. 70, n. 2, p. 1199–1228.
- Azevedo, R. G. A., De Salles Soares Neto, C., Teixeira, M. M., Santos, R. C. M. and Gomes, T. A. (2011). Textual authoring of interactive digital TV applications. In *Proceedings of the 9th international interactive conference on Interactive television - EuroITV '11*. <http://dx.doi.org/10.1145/2000119.2000169>.
- Azevedo, R. G. de A. (2015). Supporting Multimedia Applications in Stereoscopic and Depth-based 3D Video Systems. PUC-Rio.
- Azevedo, R. G. de A., Ismério, F., Raposo, A. B. and Soares, L. F. G. (2014). Real-Time Depth-Image-Based Rendering for 3DTV Using OpenCL. *Lecture Notes in Computer Science*. p. 97–106.
- Azevedo, R. G. de A. and Lima, G. F. (2016). A graphics composition architecture for multimedia applications based on layered-depth-image. In *2016 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*. <http://dx.doi.org/10.1109/3dtv.2016.7548882>.
- Azevedo, R. G. de A. and Soares, L. F. G. (2013). NCL+depth: Estendendo NCL para Displays 3D Estereo/Autoestereoscópicos. In *Proceedings of the 19th Brazilian symposium on Multimedia and the web - WebMedia '13*. <http://dx.doi.org/10.1145/2526188.2526203>.
- Balogh, T., Kovács, P. T. and Megyesi, Z. (2007). HoloVizio 3D Display System. In *Proceedings of the ImmersCom*. <http://dx.doi.org/10.4108/icst.immerscom2007.2145>.
- Barbosa, F. R. N. and Soares, L. F. G. (2014). Towards the Application of WebRTC Peer-to-Peer to Scale Live Video Streaming over the Internet. In *Anais do IX Workshop de Redes P2P, Dinâmicas, Sociais e Orientadas a Conteúdo - Wp2p+*.

- Batista, C. E. C. F. (2013). *Ginga-MD: Uma Plataforma Para Suporte à Execução de Aplicações Hipermidia Multi-Dispositivo Baseada em NCL*. Pontifícia Universidade Católica do Rio de Janeiro.
- Bernal, I. and Mejía, D. (2015). Building a Basic Hardware and Software Infrastructure for Developing Ginga-NCL Interactive Applications. *Applications and Usability of Interactive TV*. Springer, Cham. p. 74–89.
- Buchner, K., Lissermann, R. and Holmquist, L. E. (2014). Interaction techniques for co-located collaborative TV. In *Proceedings of the extended abstracts of the 32nd annual ACM conference on Human factors in computing systems - CHI EA '14*. <http://dx.doi.org/10.1145/2559206.2581257>.
- Bulterman, D. C. and Rutledge, L. (2016). *SMIL 3.0: Flexible Multimedia for Web, Mobile Devices and Daisy Talking Books*. Springer.
- Casanova, M. A., Tucherman, L., Lima, M. J. D., et al. (1991). The nested context model for hyperdocuments. In *Proceedings of the third annual ACM conference on Hypertext - HYPERTEXT '91*. <http://dx.doi.org/10.1145/122974.122993>.
- Cesar, P. and Geerts, D. (2011). Past, present, and future of social TV: A categorization. In *2011 IEEE Consumer Communications and Networking Conference (CCNC)*. <http://dx.doi.org/10.1109/ccnc.2011.5766487>.
- Fechine, Y. (2017). Social TV: A contribution towards defining the concept. *Revista Contracampo*, v. 36, n. 1.
- Ferreira, T. P. and Souza Filho, G. L. (2014). Uma extensão da visão estrutural do NCL Composer para integração de código imperativo.
- Gomes Soares, L., Soares, L. G., Moreno, M., Neto, C. S. S. and Moreno, M. (2010). Ginga-NCL: Declarative middleware for multimedia IPTV services. *IEEE Communications Magazine*, v. 48, n. 6, p. 74–81.
- Guedes, Á., Cunha, M., Fuks, H. and Colcher, S. (2016). Using NCL to Synchronize Media Objects, Sensors and Actuators. In *In: Workshop Internacional de Sincronismo das Coisas (WSoT), 1, 2016, Teresina. Anais do XXII Simpósio Brasileiro de Sistemas Multimídia e Web. Porto Alegre: Sociedade Brasileira de Computação, 2016. v. 2. WebMedia '16*.
- Guedes, A. L., Azevedo, R. G. A. and Barbosa, S. D. J. (2016). Extending multimedia languages to support multimodal user interactions. *Multimedia tools and applications*, v. 76, n. 4, p. 5691–5720.
- Guedes, Á. L., Azevedo, R. G. de A., Colcher, S. and Barbosa, S. D. J. (2016). Extending NCL to Support Multiuser and Multimodal Interactions. *Proceedings of the 22nd Brazilian Symposium on Multimedia and the Web - Webmedia '16*.
- Hosseini, M. and Swaminathan, V. (dec 2016). Adaptive 360 VR Video Streaming: Divide and Conquer. In *2016 IEEE International Symposium on Multimedia (ISM)*. IEEE.
- ISO (2016). ISO/IEC 23005-3:2016 - Information technology -- Media context and control -- Part 3: Sensory information. http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=65396, (accessed on Sep 5).

- ITU (2016). Report BT.2267-6. <http://www.itu.int:80/en/publications/ITU-R/Pages/publications.aspx>.
- ITU-T (2014). ITU Recommendation H.761: Nested context language (NCL) and Ginga-NCL for IPTV services. <http://handle.itu.int/11.1002/1000/12237>.
- Jansen, J., Frantzis, M. and Cesar, P. (2015). Multimedia Document Structure for Distributed Theatre. In *Proceedings of the 2015 ACM Symposium on Document Engineering - DocEng '15*. <http://dx.doi.org/10.1145/2682571.2797087>.
- Laptev, I., Marszalek, M., Schmid, C. and Rozenfeld, B. (2008). Learning realistic human actions from movies. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*. <http://dx.doi.org/10.1109/cvpr.2008.4587756>.
- Lemos, V. S., Ferreira, R. F., Costa Segundo, R. M., Costalonga, L. L. and Santos, C. A. S. (2016). Local Synchronization of Web Applications with Audio Markings. In *Proceedings of the 22nd Brazilian Symposium on Multimedia and the Web - Webmedia '16*. <http://dx.doi.org/10.1145/2976796.2976853>.
- Liu, J., Cheng, H. and Javed, O. Et al (2013). SRI-Sarnoff AURORA System. In *TRECVID 2013 Multimedia Event Detection and Recounting*.
- Luiz Fernando Gomes and Barbosa, S. D. J. (2009). *Programando em NCL 3.0: desenvolvimento de aplicações para middleware Ginga: TV digital e Web*.
- Meixner, B. (6 mar 2017). Hypervideos and Interactive Multimedia Presentations. *ACM Computing Surveys*, v. 50, n. 1, p. 1–34.
- Mejía, D., Bernal, I. and Becerra, F. (1 feb 2014). Plugin para Composer NCL y Aplicación Interactiva para TV Digital Orientada a Educación Superior. *Revista*, v. 33, n. 1.
- Perra, C. (2016). Light field image compression based on preprocessing and high efficiency coding. In *2016 24th Telecommunications Forum (TELFOR)*. <http://dx.doi.org/10.1109/telfor.2016.7818930>.
- Purica, A. I., Mora, E. G., Pesquet-Popescu, B., Cagnazzo, M. and Ionescu, B. (2016). Multiview Plus Depth Video Coding With Temporal Prediction View Synthesis. *IEEE Transactions on Circuits and Systems for Video Technology*, v. 26, n. 2, p. 360–374.
- Sant'Anna, F., Cerqueira, R. and Soares, L. F. G. (2008). NCLua. In *Proceedings of the 14th Brazilian Symposium on Multimedia and the Web - WebMedia '08*. <http://dx.doi.org/10.1145/1666091.1666107>.
- Santos, R. C. M., Neto, J. R. C., De Salles Soares Neto, C. and Teixeira, M. M. (2013). Incremental validation of NCL hypermedia documents. *Journal of the Brazilian Computer Society*, v. 19, n. 3, p. 235–256.
- Soares, L. F. G. (2014a). Sincronismo das Coisas não é apenas expressão da moda. <http://video.rnp.br/porta/video.action?idItem=23850>, (accessed on Sep 6).
- Soares, L. F. G. (2014b). Sincronismo das Coisas Não É Apenas Expressão da Moda. In *Proceedings of the 20th Brazilian Symposium on Multimedia and the Web*. <http://www.lbd.dcc.ufmg.br/colecoes/wvcwebmedia/2014/009.pdf>, (accessed on Sep 6).

- Soares, L. F. G., Costa, R. M. R., Moreno, M. F. and Moreno, M. F. (2009). Multiple exhibition devices in DTV systems. In *Proceedings of the seventeen ACM international conference on Multimedia - MM '09*.
<http://dx.doi.org/10.1145/1631272.1631312>.
- Soares, L. F. G., Rodrigues, R. F., Costa, R. R. and Moreno, M. F. (2006). Nested context language 3.0: Part 9--ncl live editing commands. *Monografias em Ciência da Computação do Departamento de Informática, PUC-Rio*, v. 6, p. 36.
- Soares Neto, C. de S., De Salles Soares Neto, C., Soares, L. F. G. and De Souza, C. S. (2012). TAL—Template Authoring Language. *Journal of the Brazilian Computer Society*, v. 18, n. 3, p. 185–199.
- Sulema, Y. (2016). Mulsemmedia vs. Multimedia: State of the art and future trends. In *2016 International Conference on Systems, Signals and Image Processing (IWSSIP)*.
<http://dx.doi.org/10.1109/iwssip.2016.7502696>.
- Turk, M. (jan 2014). Multimodal interaction: A review. *Pattern recognition letters*, v. 36, p. 189–195.
- Van Deventer, M. O., Stokking, H., Hammond, M., Le Feuvre, J. and Cesar, P. (mar 2016). Standards for multi-stream and multi-device media synchronization. *IEEE Communications Magazine*, v. 54, n. 3, p. 16–21.
- W3C (2004). Speech Synthesis Markup Language (SSML) Version 1.0.
<https://www.w3.org/TR/speech-synthesis/>, (accessed on Sep 5).
- Ye, G., Li, Y., Xu, H., Liu, D. and Chang, S.-F. (2015). EventNet: A Large Scale Structured Concept Library for Complex Event Detection in Video. In *Proceedings of the 23rd ACM international conference on Multimedia - MM '15*. ACM Press.
- Zare, A., Balogh, T., Bregovic, R. and Gotchev, A. (2017). Architectures and Codecs for Real-Time Light Field Streaming. *Electronic Imaging*, v. 2017, n. 5, p. 54–66.
- Zhang, Q., Huang, K., Wang, X., Jiang, B. and Gan, Y. (2017). Efficient multiview video plus depth coding for 3D-HEVC based on complexity classification of the treeblock. *Journal of Real-Time Image Processing*,

ANEXO 1: Biografia dos autores



Sérgio Colcher é professor do quadro principal do Departamento de Informática (DI) da PUC-Rio desde 2001 e coordenador do laboratório TeleMídia. Obteve os títulos de Engenheiro de Computação (1991), Mestre em Ciências em Informática (1993) e Doutor em Ciências em Informática (1999), todos pela PUC-Rio, além do Pós-Doutorado (2003) no ISIMA (Institute Supérieur D'Informatique et de Modelisation des Applications — Université Blaise Pascal, Clermont Ferrand, França). Trabalhou no Centro Científico da IBM-Rio e na divisão de desenvolvimento de hardware da COBRA (Computadores Brasileiros S/A). Foi também professor dos cursos de MBA em Gerência de Telecomunicações e MBA em e-Business da Fundação Getúlio Vargas. Suas áreas de interesse incluem redes de computadores, análise de desempenho de sistemas computacionais, sistemas multimídia/hipermídia e sistemas de TV digital.



Alan L. V. Guedes é pesquisador associado ao laboratório TeleMídia. Possui doutorado (2017) em Informática pela PUC-Rio, bacharel em Ciência da Computação (2009) e de mestre em Informática (2012) pela UFPB, onde também atuou como pesquisador associado no laboratório Lavid. Seus interesses de pesquisas incluem as áreas de sistemas multimídia e multimodal, e a autoria de aplicações para estes sistemas.



Roberto Gerson de A. Azevedo é pesquisador associado ao laboratório TeleMídia. Possui doutorado (2015) e mestrado (2010) em Informática pela PUC-Rio e é bacharel em Ciência da Computação pela UFMA (2008). Seus interesses de pesquisa incluem: representação e autoria de cenas multimídia interativas; e representação, codificação, transmissão e renderização de vídeos 3D.

Guilherme F. Lima é pesquisador associado ao laboratório TeleMídia. Ele possui doutorado (2015) e mestrado (2011) em informática, e o título de bacharel (2009) em Sistemas de Informação, todos pela PUC-Rio. Seus principais interesses de pesquisa são linguagens de programação e modelos para sincronização multimídia, em particular, na interseção entre linguagens síncrona e multimídia.



Rodrigo C. M. Santos é aluno de doutorado na PUC-Rio e pesquisador associado ao laboratório TeleMídia. Ele recebeu o título de bacharel em Ciência da Computação (2010) e mestre em Ciência da Computação (2012) pela UFMA, onde trabalhou como pesquisador associado ao laboratório LAWS. Seus interesses de pesquisas incluem as áreas de linguagens de programação para multimídia, modelos de sincronismo distribuído e sistemas multimídia distribuídos.



Antonio José G. Busson é aluno de doutorado na PUC-Rio e pesquisador associado ao laboratório TeleMídia. Ele recebeu o título de bacharel em Ciência da Computação (2013) e de mestre em Ciência da Computação (2015) pela UFMA, onde atuou como pesquisador associado ao laboratório LAWS. Seus interesses de pesquisa incluem: sistemas multimídia/hipermídia, modelos de hiperdocumentos, reconhecimento de padrões e sistemas de TV Digital.

Capítulo

8

Uma abordagem abrangente para a colaboração multimídia

Marcelo F. Moreno¹, Eduardo Barrére¹ and Jairo F. de Souza¹

¹ Universidade Federal de Juiz de Fora. Dep. Ciência da Computação (DCC/UFJF)

{moreno, eduardo.barrere, jairo.souza}@ice.ufjf.br

Abstract

LapIC is a research group in the Multimedia/Hypermedia Systems area, at the Federal University of Juiz de Fora (UFJF). The group gathers researchers with diversified background and profiles, but with a high degree of synergy in their works, leading to a comprehensive approach for the resolution of problems related to video collaboration. This is being concretized through their recent results, like the conception of a convergent, collaborative, multimedia IPTV platform, whereon other group's works are being integrated, applied and validated. Such works include the modelling and assistance for hypermedia content authoring; multimedia information indexing and retrieval; content delivery and audience measurement; and the innovative use of multimedia in scenarios like education.

Resumo

LapIC é um grupo de pesquisa na área de Sistemas Multimídia/Hipermídia, da Universidade Federal de Juiz de Fora (UFJF). O grupo reúne pesquisadores de perfis diversificados, mas com alto grau de sinergia em seus trabalhos, levando a uma abordagem abrangente para a resolução de problemas relacionados à videocolaboração. Isso vem se concretizando com seus recentes resultados, como a concepção de uma plataforma IPTV, multimídia, colaborativa e convergente, sobre a qual outros trabalhos do grupo vêm sendo integrados, aplicados e validados. Tais trabalhos incluem a modelagem e auxílio à autoria de conteúdo multimídia/hipermídia; a indexação e recuperação da informação multimídia; a entrega do conteúdo e medição de sua audiência; e o uso inovador de multimídia em cenários como a educação.

8.1. Introdução

O Laboratório de Aplicações e Inovação em Computação (LApIC)¹ do Departamento de Ciência da Computação (DCC) da Universidade Federal de Juiz de Fora (UFJF) foi criado em 2009 para concentrar ações que visam o desenvolvimento de aplicações multimídia avançadas, mais especificamente nas áreas de Web e também nas várias modalidades de TV Digital Interativa. Desde então, o grupo vem experimentando uma expansão continuada e sustentável, apoiada por sua inserção no Programa de Pós-Graduação em Ciência da Computação (PPGCC) da UFJF. Atualmente, suas pesquisas abrangem diversos tópicos na área de Sistemas Multimídia/Hipermídia, dentre estes, principalmente:



- Infraestrutura de telecomunicações multimídia, incluindo protocolos, Qualidade de Serviço (QoS) e nuvem multimídia
- Análise de conteúdo, incluindo indexação e recuperação multimídia
- Multimídia aplicada à educação, incluindo modalidades presencial e à distância
- Arquiteturas e serviços de vídeo interativo, incluindo Internet TV, IPTV, TVD terrestre, vídeo sob demanda
- Arquitetura de sistemas, modelagem e autoria do conteúdo hipermídia

O LApIC conta com uma equipe de pesquisadores de perfil diversificado e que se guia pela constante colaboração, promovendo alto grau de sinergia entre os projetos e entre seus resultados. Tal sinergia conduziu a uma natural integração dos trabalhos, de forma que, com uma visão sistêmica e estratégica, o grupo se vê concretizando uma abordagem abrangente para a resolução de problemas relacionados à colaboração multimídia.

Seus resultados recentes em projetos de pesquisa e desenvolvimento incluem uma plataforma multimídia colaborativa e convergente, sobre a qual outros trabalhos do grupo vêm sendo aplicados, integrados ou validados. Tais trabalhos incluem desde a modelagem e autoria de conteúdo hipermídia até questões sobre a entrega desse conteúdo; desde a indexação e recuperação da informação multimídia até a medição de sua audiência; desde a infraestrutura de comunicação multimídia até a sua adequação para uso inovador em cenários como a educação apoiada por multimídia.

Neste contexto, o presente capítulo apresenta um *roadmap* dos trabalhos de pesquisa do LApIC, abordando temas relevantes para a composição de serviços multimídia, tendo a vídeo colaboração como caso especial.

8.2. Sinergia do LApIC

O LApIC vem experimentando uma expansão continuada e sustentável de seu quadro de pesquisadores, formado por professores doutores, alunos de mestrado e de graduação. O Laboratório conta atualmente com três professores permanentes do Programa de Pós-graduação em Ciência da Computação (PPGCC) da UFJF (Mestrado), seis mestrandos e seis alunos de graduação, em iniciação científica. Já passaram pelas bancadas do LApIC cerca de 25 alunos de mestrado e graduação, participantes em projetos de pesquisa e

¹ <http://www.ufjf.br/lapic>

desenvolvimento em áreas diversificadas e, ao mesmo tempo complementares, em que seus membros exercem relevante papel no cenário nacional e internacional.

Tendo como Sistemas Multimídia/Hipermídia e a Informática na Educação como áreas guarda-chuva em seus trabalhos, o LApIC busca, constantemente, agregar multidisciplinaridade em seu quadro de pesquisadores e em suas parcerias. Ao reunir professores especialistas em busca e recuperação da informação, bancos de dados, informática na educação, comunicação social, redes, middleware, plataformas e autoria multimídia, o LApIC experimenta plenamente uma sinergia que leva a trabalhos com alto nível de integração, reuso, evolução, validação e/ou testes. Este Capítulo é pautado nessa diversidade de pesquisas sinérgicas conduzidas no Laboratório, que de alguma forma lidam com a videocolaboração em sua infraestrutura, plataformas, protocolos ou aplicações.

O Laboratório mantém parcerias com diversas instituições de pesquisa e desenvolvimento, com as Universidades Federais do Maranhão (UFMA), Fluminense (UFF) e da Paraíba (UFPB), Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), além de órgãos assessores ou regulatórios como a União Internacional de Telecomunicações e o Fórum do Sistema Brasileiro de TV Digital. Além disso, o laboratório promove, constante colaboração com o Laboratório de Mídias Digitais (LMD), da Faculdade de Comunicação da UFJF, ponto de debate interdisciplinar envolvendo a Comunicação Social, Computação, Artes/Design e Educação.

Dentre os vários projetos coordenados ou que tiveram participação do LApIC, pode-se citar como os principais: Grupo de Trabalho da RNP - Busca Avançada por Vídeo, CAPES-UAB - Recomendação de mídias baseada em transcrição de áudio e anotação semântica; Grupo de Trabalho RNP – IpêTeVê: Serviço de Televisão IP de alcance global; Ginga Internacional: Transferência de tecnologia do middleware Ginga em nível internacional; Iniciativa Global de Padrões em IPTV (IPTV-GSI); Iniciativa IPTV.br; GingaFrEvo & GingaRAP: Evolução do Middleware Ginga para Múltiplas Plataformas (Componentização) & Ferramentas para Desenvolvimento e Distribuição de Aplicações Declarativas, entre outros.

Dentre os principais veículos de publicações estão o Simpósio Brasileiro de Informática na Educação, Simpósio Brasileiro de Sistemas Multimídia e Web, a Revista Brasileira de Informática na Educação, o *ACM Symposium on Document Engineering* o periódico *Multimedia Tools and Applications* da *Springer*, entre outros.

Atualmente, o LApIC encontra-se instalando nas dependências do PPGCC da UFJF, como uma área de cerca de 30m², 12 estações de trabalho, e um parque de PCs, notebooks, switches e servidores que atendem as demandas de pesquisa e desenvolvimento, sempre atualizado com as concessões dos novos projetos.

8.3. Uma Abordagem Abrangente sobre colaboração multimídia

Com uma visão estratégica de trabalho conjunto e em sintonia com as tendências mais inovadoras tanto da comunidade acadêmica, quanto do mercado, o grupo se vê concretizando uma abordagem abrangente para a resolução de problemas relacionados à colaboração multimídia.

Seus resultados recentes em projetos de pesquisa e desenvolvimento incluem uma plataforma multimídia colaborativa e convergente, seguindo padrões abertos internacionais, baseada em *Internet Protocol Television* (IPTV). Apesar do nome, a plataforma IPTV não é restrita a serviços televisivos, pois atualmente há técnicas bem estabelecidas para a extensão da plataforma para qualquer tipo de serviço multimídia, com facilidades para a descoberta automática de serviços e conteúdos, Qualidade de Serviço (QoS), e uma grande variedade de novas funcionalidades que podem ser agregadas por se tratar de plataforma aberta. Especificamente, os resultados atingidos permitem a colaboração na criação e disponibilização de conteúdos e via aplicações multimídia interativas ricas em recursos aos usuários. A Seção 8.4 descreve mais detalhes dessa iniciativa.

Essa flexibilidade da plataforma tem também permitido seu uso não somente nos diversos pilotos já conduzidos no próprio projeto, como também sob a forma de *testbed* para outros trabalhos do grupo. Validações, integrações, aplicações e avaliações de QoE de novas proposições do laboratório têm sido realizadas tomando IPTV como base.

Ao mesmo tempo, a plataforma dá ampla liberdade aos pesquisadores para investigar problemas inerentes às suas áreas de maior afinidade, conferindo assim uma abrangência de temas de pesquisas que contribuem, cada um, com resultados graduais para evolução consistente dos serviços multimídia, sendo a videocolaboração uma das beneficiadas. Sobre cada ponto dessa abordagem abrangente para a colaboração multimídia encontram-se as discussões deste capítulo.

A Seção 8.5 descreve os trabalhos de modelagem de documentos hipermídia, que incluem a incorporação de recursos avançados como adaptação de conteúdo, sincronismo, interação multimodal e apresentações distribuídas.

A Seção 8.6 descreve a solução Bavi, que lida com a extração da informação de vídeo e áudio, para o enriquecimento semântico e indexação de conteúdo multimídia para facilitar sua posterior recuperação;

A Seção 8.7 apresenta trabalhos aplicados à entrega e consumo de conteúdo hipermídia, incluindo estratégias da transmissão em modo *push* de acordo com as características da rede de comunicação utilizada, a medição de audiência em sistemas de TV Digital e IPTV e a concepção de máquinas de apresentação de conteúdo hipermídia;

Finalmente, a Seção 8.8 Descreve como o LApIC vislumbra a evolução dos trabalhos apresentados e da abordagem abrangente de ataque aos problemas inerentes à colaboração multimídia. Serão discutidos os novos rumos de pesquisa, os pontos em aberto, as novas áreas de aplicação e novos casos de uso.

8.4. Uma plataforma multimídia convergente e colaborativa

A possibilidade de oferecimento de serviços multimídia distribuídos e de alta qualidade esbarra no fato de que o mercado se encontra carente de soluções fim-a-fim interoperáveis e abertas. No entanto, esforços recentes de padronização vêm tornando realidade a provisão de uma ampla gama de serviços por meio da plataforma IPTV (*Internet Protocol Television*), convergente e, dependendo do ferramental disponível, de forma colaborativa.

Segundo a definição da União Internacional de Telecomunicações (UIT), IPTV é um serviço multimídia abrangendo televisão, áudio, vídeo, dados entregues sobre redes baseadas em Protocolo Internet, gerenciadas para prover os níveis requeridos de Qualidade de Serviço (QoS), Qualidade de Experiência (QoE), segurança, interatividade e confiabilidade [UIT-T Y.1910].

O esforço de padronização da UIT na área de IPTV visa especificar cada subsistema que compõe um serviço realmente aberto, tal que qualquer ator da cadeia de valor de IPTV possa manter foco e independência em suas ações, além de poder contar com equipamentos e serviços padronizados, capazes de serem encontrados em estabelecimentos comerciais em geral.

Incrivelmente, serviços IPTV possuem pouca penetração no Brasil. Os atuais Provedores de Rede do país preferem investir em infraestrutura paralela e terceirizada (redes de satélites) para prover serviços televisivos, em detrimento de ampliar a capacidade de suas redes de acesso em banda larga.

No entanto, no contexto acadêmico, o país possui uma infraestrutura em constante evolução, provida pela Rede Nacional de Ensino e Pesquisa (RNP). A RNP oferece a conectividade e diversos serviços às instituições usuárias, serviços estes que vêm evoluindo e demandando um uso cada vez mais avançado da Rede.

Neste contexto, IpêTeVê é uma iniciativa da UFJF, PUC-Rio e RNP para o desenvolvimento de uma solução fim-a-fim para o “Suporte a Provedores de Serviços IPTV voltados a Educação e Pesquisa”. Assim, instituições usuárias podem se tornar provedores de serviços IPTV, oferecendo grades de canais e conteúdos VoD interativos segmentados em áreas do conhecimento.

Uma vez capazes de criar seus próprios provedores de serviços IPTV, as instituições e comunidades de ensino e pesquisa ligadas à RNP poderão buscar nessa plataforma novas dinâmicas na colaboração e no compartilhamento do conhecimento. Os serviços podem ser explorados para o oferecimento de videoaulas interativas, eventos ao vivo, salas de aula inteligentes, e agregação de conteúdos educativos advindos de diversas fontes, etc.

A solução desenvolvida pelo GT-*IpêTeVê* abrange desde o dispositivo terminal IPTV até as ferramentas de administração de serviços. Servidores de prateleira vem sendo utilizados para a entrega de dados e metadados aos dispositivos terminais. O protocolo IPv6 é adotado por concepção, buscando facilidades para comunicação em grupo (*multicast*), configurações automáticas de interfaces de rede e suporte a QoS.

Diversos requisitos nortearam as decisões de projeto no desenvolvimento da solução. Notadamente, os diferentes perfis de usuários de cada componente da solução foram levados em conta. O sucesso de um futuro serviço baseado nesta solução depende das facilidades, conveniência e nível de colaboração oferecida aos administradores (para a criação de serviços e publicação de conteúdo) e aos usuários finais (para a descoberta dos serviços, seleção e consumo dos conteúdos).

A Figura 8.1 ilustra a Arquitetura Geral *IpêTeVê*, cujos componentes possuem funções relacionadas a três camadas de provisão do serviço (áreas dispostas horizontalmente na figura): Entrega de Conteúdo (laranja); Entrega de Metadados sobre o Conteúdo (azul) e Gerência do Serviço (rosa).

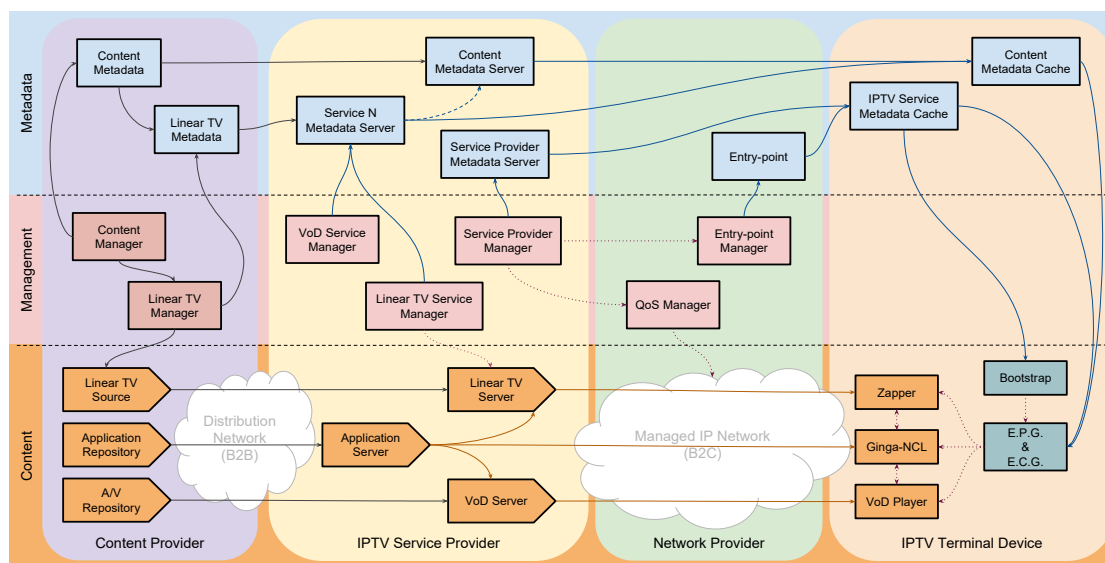


Figura 8.1 - Arquitetura Geral IpêTeVê

A figura também identifica como os componentes são distribuídos entre os diversos atores da cadeia de valor do IPTV (áreas dispostas verticalmente na figura): Provedor de Conteúdo (roxo claro); Provedor de Serviço IPTV (amarelo claro); Provedor de Rede (verde claro) e Usuário do Serviço (laranja claro). Dentre os componentes ilustrados, Gerente de QoS e os componentes localizados no Provedor de Conteúdo são alvos de desenvolvimento atualmente em andamento.

O primeiro protótipo desenvolvido pelo grupo implementa as funções de Dispositivo Terminal IPTV, conforme descritas na Recomendação UIT-T Y.1910, em especial um conjunto mínimo de funcionalidades que permitam o consumo dos serviços IPTV definidos para o modelo básico de dispositivo terminal IPTV [UIT-T H.721].

A Figura 8.2 ilustra a arquitetura do dispositivo terminal. Os componentes da arquitetura estão implementados em linguagem C e de forma isolada, como processos independentes que se comunicam por meio de um middleware, com baixo nível de acoplamento.

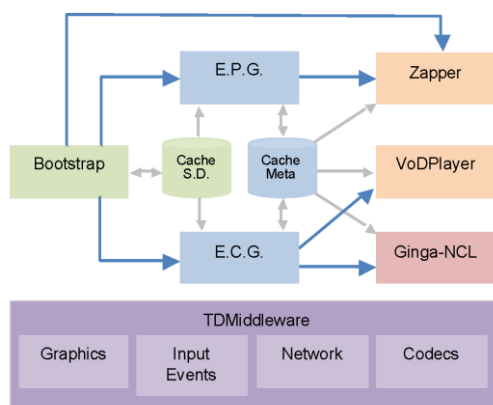
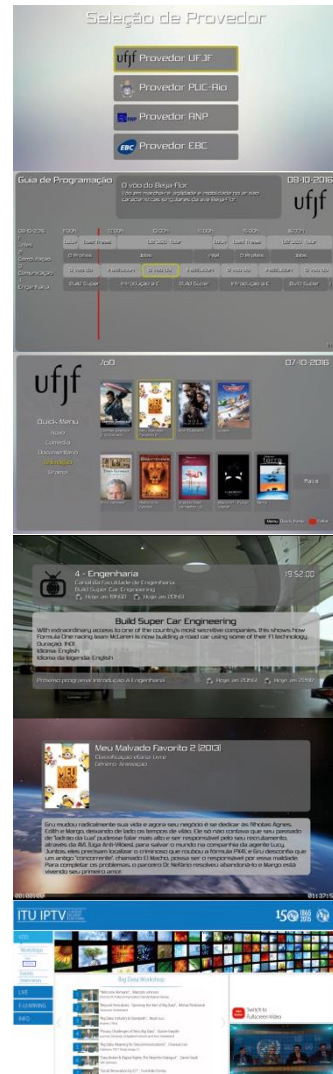


Figura 8.2 - Arquitetura do Dispositivo Terminal

A seguir, encontra-se uma breve descrição dos componentes do dispositivo terminal.

- *TDMiddleware*: Camada intermediária entre os componentes e o sistema operacional do dispositivo terminal. Inclui as funcionalidades de controle do terminal, assim como oferece uma API de suporte gráfico, codecs, protocolos e eventos de entrada.
- *Bootstrap*: Inicializa o dispositivo terminal na rede IPTV, oferecendo ao usuário uma interface para seleção dos provedores descobertos e serviços IPTV de interesse.
- *EPG* (Guia Eletrôn. de Programação): Disponibiliza ao usuário interface gráfica para a seleção de conteúdo de TV Linear (canais TV) dentre os descobertos no provedor de serviços IPTV em uso.
- *ECG* (Guia Eletrônico de Conteúdo): Disponibiliza ao usuário interface gráfica para a seleção de conteúdo sob demanda, dentre os descobertos para o provedor de serviços IPTV em uso.
- *Zapper*: Exibidor do serviço de TV Linear (canais IPTV com programação pré-definida). É capaz de exibir áudio e vídeo nos formatos recomendados e permite troca de canais numérica e incremental.
- *VoDPlayer*: Exibidor do serviço de Vídeo Sob Demanda (VoD). É capaz de exibir áudio e vídeo nos formatos recomendados e permite controle de exibição de mídia, tal como pausa e retrocesso.
- *Ginga-NCL Player*: Aplicação para a apresentação de documentos hipermídia escritos em NCL (*Nested Context Language*) e Lua [UIT-T H.761], voltados para o oferecimento de serviços interativos IPTV.



Já os servidores para a entrega de metadados e de conteúdo são baseados em soluções de prateleira, uma vez que há no mercado opções, inclusive em software livre. Conforme ilustrado na arquitetura geral (Figura 8.1), os servidores são os seguintes:

- Servidor de Descoberta de Entry-point: Anuncia o endereço do Ponto de Entrada do Provedor de Rede. Protocolo: DNS. Software: DNS Bind.
- Servidor Entrypoint: Anuncia os Provedores de Serviços IPTV presentes no Provedor de Rede, bem como os seus serviços. Protocolo: HTTP/HTTPS. Software: Apache Web Server.
- Servidor de Metadados: Anuncia Metadados sobre conteúdos de um serviço. Protocolo: HTTP/HTTPS. Software: Apache Web Server.
- Servidor de Conteúdo TV Linear: Entrega streaming de áudio/vídeo por meio de comunicação em grupo (*multicast*). Protocolo: RTP Multicast. Software: VLC Server.

- Servidor de Conteúdo Vídeo sob Demanda: Entrega streaming de áudio/vídeo por meio de comunicação individualizada (*unicast*). Protocolo: RTSP/RTP Unicast. Software: Live555.

Os servidores de prateleira para entrega de conteúdo (tanto TV Linear quanto VoD) têm se mostrado ineficientes para a provisão do serviço em larga escala. O grupo planeja o desenvolvimento ou adaptação de solução existente para melhor suprir suas necessidades.

Por fim, a Ferramenta de Gerência é essencial para a viabilidade da solução como um potencial serviço da RNP, uma vez que ela oferece abstrações gráficas e facilidades de automação que facilitam a criação de Provedores de Rede e Provedores de Serviços IPTV. A Ferramenta é implementada como serviço Web e foi concebida sobre PHP e o framework AngularJS. A Ferramenta de Gerência é composta dos seguintes módulos, acessíveis por meio de uma interface única e adaptável a cada perfil de usuário administrador:

- Local Admin Manager, que permite a criação/edição de contas de usuários.
- Network Provider Manager, que permite: descoberta automática de entry-points; criação/edição de provedor de rede; aprovação de Provedores de Serviços IPTV; edição de Entry-Points.
- IPTV Service Provider Manager, que permite: descoberta automática de Provedores de Rede; descoberta automática de servidores de metadados; descoberta automática de servidores de streaming (TV linear e VoD); criação/edição de Provedor de Serviços IPTV; solicitação de aprovação de Provedor de Serviços IPTV.
- Linear TV Service Manager, que permite: descoberta automática de servidores de streaming (TV linear); criação/edição de grades de programação; publicação de grades de programação; suspender/retomar entrega do serviço
- On-Demand Service Manager, que permite: descoberta automática de servidores de streaming (VoD); criação/edição de catálogos de VoD; publicação de catálogos de VoD; suspender/retomar entrega do serviço;

Portanto, a solução IpêTeVê pode ser vista como uma plataforma multimídia flexível, baseada no protocolo convergente IPv6 e outros padrões abertos, e, por isso, pode ser explorada de várias formas. O serviço proposto de “Suporte a Provedores de Serviços IPTV voltados a Educação e Pesquisa”, no âmbito da RNP, pode dar às instituições interessadas a capacidade de se tornarem provedores de serviços IPTV, oferecendo grades de canais e conteúdo VoD interativos segmentados em áreas do conhecimento. O alcance geográfico desse conteúdo vai depender da viabilidade técnica e do acordo firmado entre a RNP e a instituição usuária.

Uma vez capazes de criar seus próprios provedores de serviços IPTV, as instituições de ensino e pesquisa ligadas à RNP poderão buscar novas dinâmicas na colaboração e no compartilhamento do conhecimento, explorando o uso de videoaulas interativas, eventos ao vivo, salas de aula inteligentes, agregação de conteúdos educativos advindos de diversas fontes, entre muitos outros.

Notadamente, as possibilidades de se construir programações de canais, ou de disponibilizar conteúdo sob demanda, com a qualidade de serviço inerente à definição

de IPTV e a partir de fontes de conteúdo diversas, trarão novos usos aos conteúdos disponíveis em serviços como Vídeo@RNP, Videoaula@RNP e Portal de Intercâmbio de Conteúdos Digitais (ICD).

IpêTeVê tem como uma de suas premissas o caráter agregador de conteúdo, ao estabelecer, de fato, uma nova plataforma de disseminação e consumo de vídeo, desta vez atendendo a requisitos como QoS e usabilidade de TV. Além disso, há todo o suporte necessário para a exploração de recursos avançados de interatividade e liberdade de formatos de conteúdo (e.g. 3D, UHDTV, Segunda Tela...).

O serviço proposto deve promover a colaboração entre diferentes instituições e estabelecer regras claras de governança do serviço. A colaboração é possível, uma vez que a Ferramenta de Gerência por si é um sistema distribuído e permite que, por exemplo, um administrador de TV Linear delegue a responsabilidade de preenchimento de slots da programação a administradores autorizados de outras instituições. Regras de negócio podem ser incluídas por cada instituição e, principalmente, pela RNP, para que as grades de programação já venham com reservas de horários, seja para conteúdo institucional da RNP, seja para a implantação de políticas públicas, ou qualquer propósito de interesse da RNP. O mesmo pode ser feito nos catálogos de conteúdo sob demanda, que incluem não somente vídeo, mas também aplicações.

8.5. Contribuições em modelagem da autoria hipermídia

Recentemente, o LApIC vem também trabalhando na concepção de um modelo hipermídia denominado STorM e uma Linguagem de Domínio Específico (DSL) denominada STorML [Freesz Jr., 2017]. A proposta reúne as melhores características de modelos conhecidos como o NCM [Soares, 2005] e o AHM [Hardman, 1994] para facilitar a autoria de conteúdo hipermídia voltado especialmente para mídias digitais fora de casa (*Digital out-of-home* - DOOH), como a sinalização digital e a TV corporativa. Tais veículos possuem requisitos diferenciados para a autoria hipermídia, uma vez que os terminais de exibição do conteúdo encontram-se normalmente em locais públicos ou privados com maior concentração de pessoas. A plataforma DOOH tem grande potencial para a colaboração e treinamento em ambiente corporativo, assim como pode levar experiências colaborativas entre usuários ou destes com serviços em locais públicos.

Avanços nos métodos de produção, distribuição e consumo de conteúdo DOOH se tornam cada vez mais importantes no cenário global. No entanto, várias soluções atuais para DOOH mostram-se excessivamente restritivas, ao tratar a autoria de conteúdo como uma simples criação de *playlists* de objetos de mídia. Nesses casos, os conteúdos usualmente veiculados acabam se limitando a uma apresentação de objetos de mídia predominantemente discreta, ainda que com suporte a vídeo/animações, mas como uma espécie de *slideshow*.

E quando há recursos mais elaborados, a representação da informação acaba por se tornar complexa, fazendo uso de abstrações não relacionadas ao domínio DOOH e até mesmo de linguagens baseadas em *script* de propósito mais geral. Dessa forma, a representação dos documentos vem a ser ilegível por seus criadores e até mesmo de difícil extração de informação e semântica, o que seria fundamental para a definição de *workflows* abrangentes que tratassem de maneira fim-a-fim o problema da produção e distribuição de mídias de DOOH.

Além disso, futuras modalidades de serviços DOOH devem não somente considerar os aspectos básicos da exibição de um conteúdo como uma sequência de objetos de mídia sincronizados, mas também os aspectos avançados que permitam o oferecimento de múltiplas modalidades de interação, leiautes cientes de arranjos de telas em *videowalls*, integração com dispositivos móveis dos usuários e adaptabilidade do conteúdo ao contexto da audiência. Porém, tais aspectos avançados encontram-se ainda incipientes nas ferramentas e sistemas DOOH. Nesse sentido, STorM se propõe a não somente auxiliar a autoria hipermídia, mas também suportar funcionalidades essenciais para experiências engajadoras junto aos terminais de DOOH.

STorM (*Scene-and-Track-oriented hypermedia Model*) busca reduzir a distância entre os conceitos utilizados nas abstrações gráficas de autoria e os elementos utilizados na representação de documentos hipermídia. Para isso, STorM e sua respectiva linguagem de representação, denominada STorML (*STorM Language*), apresentam entidades de mais alto nível, com conceitos próximos aos existentes no contexto da indústria do audiovisual, como cenas e trilhas, de modo a respeitar o perfil dos criadores de conteúdo. Além disso, STorM inclui suporte a formas avançadas de interação e de personalização do conteúdo, possibilitando seu uso na autoria textual de apresentações para DOOH.

Para a concepção de STorM, buscou-se encontrar conceitos e entidades que se aproximassem do vocabulário utilizado pelos profissionais da área de comunicação social que, em grande parte das corporações/instituições, são os responsáveis pela criação de conteúdos DOOH. Portanto, conceitos similares aos utilizados nos softwares de edição audiovisual foram abordados. Não somente levou-se em conta esses softwares, como também as diversas ferramentas gráficas usadas na criação de conteúdos DOOH estudadas. Dessa forma, buscou-se facilitar a assimilação dos conceitos com o intuito de favorecer a autoria de instâncias STorM.

STorM utiliza a abstração de cenas e trilhas como pilares para organizar o documento hipermídia. As entidades de STorM encontram-se ilustradas na Figura 8.3. As cenas são comparadas aos contextos de NCL, como composições com escopo restrito, com o diferencial de apresentarem semântica de paralelismo incorporada (como composições *<par>* de AHM) e um nível de abstração maior. Uma cena representa a apresentação em paralelo de uma coleção de trilhas, que por sua vez correspondem à apresentação em sequência de uma coleção de objetos de mídia. O nó de cena é representado pela entidade *Scene*, que não permite aninhamento de outras cenas. Nota-se, assim, que o comportamento de uma apresentação baseada em STorM pode ser comparada ao resultado da edição de um filme, produzido através de uma sequência de cenas, em que apenas uma cena é exibida por vez.

Cada trilha (*Track*) que compõe uma cena representa uma sequenciação de objetos de mídia (Content Nodes). As trilhas são exibidas em uma ou, de forma replicada, mais regiões associadas. Conceitualmente, as trilhas têm um comportamento de *timeline*, inspirado nas trilhas de editores de audiovisual. Além disso, sua semântica se assemelha à composição *<seq>* utilizada em AHM. O início das trilhas está relacionado ao início de sua respectiva cena e os objetos de mídia internos às trilhas apresentam um comportamento sequencial, na ordem em que são especificados. Cada trilha possui duração implícita determinada pela agregação das durações dos objetos de

mídia internos. No entanto, quando a duração explícita é definida, sua duração implícita é ignorada.

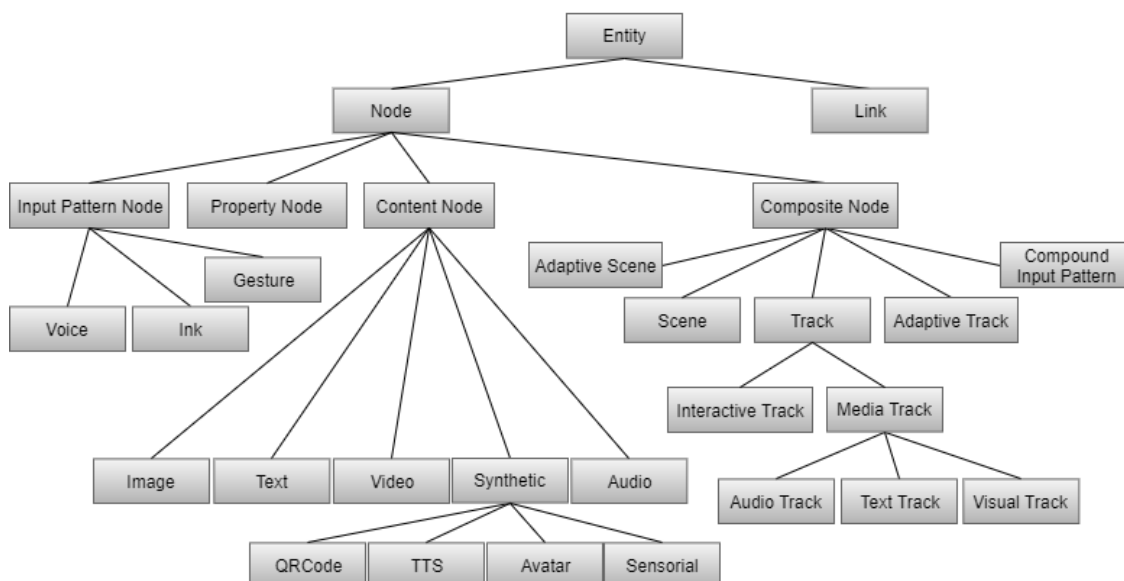


Figura 8.3 – Entidades STorM

Algumas exceções a essas definições se aplicam a um tipo especial de trilha: as trilhas interativas (*Interactive Track*). Trilhas interativas apresentam um comportamento diferente das trilhas de mídias contínuas (*Audio Track*, *Visual Track* e *Text Track*), as quais contêm objetos de mídia relacionados sequencialmente, de modo predeterminado. No caso das trilhas interativas, os relacionamentos ocorrem por meio do paradigma de eventos de causalidade, igual ao utilizado em NCL. Dessa forma, trilhas interativas podem ser comparadas a contextos NCL e qualquer relacionamento explícito em uma trilha interativa deve ser especificado por meio de entidades elo (*Link*). No entanto, além da utilização de relacionamentos síncronos e assíncronos, a trilha interativa poderá ser utilizada para especificar aplicações autocontidas, denominadas *widgets*. Esses arquivos de *widgets* devem estar em conformidade com a especificação de aplicações empacotadas para o serviço de IPTV definida em [UIT-T H.765]. De forma similar às trilhas interativas, as trilhas de texto também poderão ser utilizadas para referenciar arquivos compostos por fragmentos de textos, cada qual com marcas temporais (*timestamp*) definidas para exibição. Esses arquivos são comumente utilizados na especificação de legendas (e.g. SRT, WebVTT, TTML, etc.).

Destaca-se, ainda, que as trilhas e seus objetos de mídia internos possuem propriedades implícitas, que podem ser alteradas ou utilizadas em relacionamentos durante a exibição da cena, de modo a trazer maior dinamicidade e controle do conteúdo exibido. Dentre essas propriedades, destacam-se: A região (*region*) em que determinada trilha será exibida; o atraso para que uma mídia seja iniciada (*delay*); o volume de uma trilha ou mídia (*sound level*), dentre outras.

Com relação à adaptabilidade de conteúdo, STorM propõe os conceitos de cenas adaptativas (*Adaptive Scene*) e trilhas adaptativas (*Adaptive Track*). Tais entidades são tratadas como uma cena ou trilha que irá se adaptar ao contexto do terminal. Portanto, cena adaptativa é uma composição de cenas alternativas, das quais apenas uma será

exibida, havendo, nesse caso, uma adaptação do todo de uma cena. Quando se pretende adaptar apenas parte do conteúdo de uma cena e não sua totalidade, utiliza-se a trilha adaptativa. Para auxiliar essas adaptações, os nós de propriedade (*Property*) são especificados no modelo para permitir um maior controle dos conteúdos que devem ser exibidos a partir da avaliação dessas propriedades explícitas, que representam, normalmente, variáveis de ambiente.

Além das mídias comumente utilizadas em outros modelos hiperídia, como áudio, vídeo, texto e imagem, STorM propõe a utilização de mídias sintéticas (*Synthetic*), como especializações dos nós de conteúdo. Nós sintéticos são mídias geradas em tempo de execução com base em linguagens de marcação específicas, as quais podem representar vozes, avatares 3D, efeitos sensoriais e *QR Codes*. Os sintetizadores são os componentes de uma máquina de apresentação STorM responsáveis por interpretar essas especificações.

Máquinas de apresentação STorM também incluem componentes reconhecedores, responsáveis pela geração de eventos baseados em especificações de padrões de entrada (*Input Pattern*). Padrões de entrada são fragmentos de informações (*node*) que representam um gesto, um comando de voz ou um movimento por meio de caneta. Além de sua utilização com base em uma entrada simples do usuário, o modelo permite a composição desses padrões de entradas (*Compound Input Pattern*), de modo a possibilitar diversas combinações (interações multimodais), como, por exemplo, um gesto em paralelo a uma fala.

Em STorM os padrões de entrada são associados às trilhas ou às mídias nos relacionamentos (*Link*), sendo, portanto, entidades distintas que não necessitam ser iniciadas. Ou seja, essas entidades sempre estarão associadas a um objeto de primeira classe do modelo e, portanto, a identificação do padrão de entrada ocorrerá quando o objeto associado estiver ocorrendo.

STorML é uma linguagem baseada em XML especificada para a representação de instâncias de STorM. Para ilustrar rapidamente sua sintaxe e algumas de suas funcionalidades, toma-se como base um caso de uso específico, no qual um terminal DOOH móvel, comumente encontrado em transportes coletivos, circula no âmbito da UFJF. A proposta é fornecer ao público da instituição um conteúdo dependente da localização do terminal. Assim, um conteúdo com informações sobre o ICE (Instituto de Ciências Exatas) será mostrado caso o veículo esteja próximo a esse Instituto. Da mesma forma, um conteúdo próprio será exibido para o ICH (Instituto de Ciências Humanas). Caso o veículo esteja em qualquer outro ponto da instituição, um conteúdo padrão com informações gerais sobre a Universidade será exibido. Além do conteúdo diferenciado para cada localização, o caso de uso aborda a disponibilização do cardápio semanal do Restaurante Universitário em horário comumente dedicado ao almoço dos estudantes.

Para representar esse caso de uso em STorML, duas cenas principais foram definidas: uma responsável por exibir informações gerais sobre a Universidade e outra responsável pela exibição adaptada, de acordo com a localização do terminal móvel. Essa última trata-se de uma cena adaptativa, a qual irá se especializar em uma das duas cenas internas, uma responsável por exibir o conteúdo referente ao ICE e outra referente ao ICH. A primeira cena, ilustrada pela Figura 8.4, contém um vídeo institucional com

legenda sobre a inauguração do Centro de Ciências da UFJF, notícias gerais e informações básicas de temperatura e horário.



Figura 8.4 – Cena sobre a UFJF (Centro de Ciências)

A segunda cena deve ser uma cena adaptativa, que irá se especializar em duas cenas alternativas. Essas cenas estão ilustradas na Figura 8.5. Cada uma delas contém um vídeo específico com legenda, notícias próprias de cada Instituto e também informações básicas de temperatura e horário.



Figura 8.5 – Cenas alternativas ICE e ICH/ UFJF, dependentes da localização

Além destas exibições padrões de cada cena, em determinado horário cada cena terá seu conteúdo alterado na região onde normalmente são exibidas as notícias. Nessa região, no horário de 11:00 às 14:00, informações do cardápio semanal serão mostradas, além disso, essa mesma informação será exibida nos dispositivos secundários, exemplificados por *smartphones* na Figura 8.6, para a localização do ICH. Da mesma forma acontecerá com as demais cenas. Em STorML, essas adaptações internas a uma cena são possíveis por meio do conceito de trilhas adaptativas.



Figura 8.6 – Cena com a trilha de cardápio no horário de almoço

Para tratar esse caso de uso, dois *layouts* foram definidos. O primeiro especifica as regiões do terminal principal. Esse layout será reutilizado por todas as cenas por apresentarem a mesma composição espacial. O outro *layout* se refere à exibição das informações de cardápio no dispositivo secundário. Esse *layout* é composto por duas regiões. A primeira região delimita toda a tela do dispositivo secundário e será usada para exibir o cardápio diário. A outra região especifica os locais onde as imagens responsáveis pela interação estarão presentes. A listagem a seguir ilustra esses *layouts*.

```

<storml>
  <head>
    <layoutBase>
      <layout id="lyVideoWall" dev="main">
        <region id="rgFull" top="0%" left="0%" width="100%" height="100%" z-index="1"/>
        <region id="rgVideo" top="0%" left="0%" width="60%" height="80%" z-index="2"/>
        <region id="rgLeg" top="70%" left="10%" width="40%" height="10%" z-index="3"/>
        <region id="rgMaNews" top="0%" left="60%" width="40%" height="80%" z-index="2"/>
        <region id="rgWid" top="80%" left="0%" width="100%" height="20%" z-index="2"/>
      </layout>
      <layout id="lyMobRU" dev="secondary(1)">
        <region id="rgMobRU" top="0%" left="0%" width="100%" height="100%" z-index="1"/>
        <region id="rgNavMobRU" top="20%" left="0%" width="100%" height="15%" z-index="2"/>
      </layout>
    </layoutBase>
  </head>

```

O restante do documento STorML necessário para representar este caso de uso pode ser encontrado na próxima listagem. Existem três cenas que serão exibidas sequencialmente (*presentationMode*), cada qual com duração máxima de 5 minutos e composta por sete trilhas. A primeira trilha é uma trilha visual responsável por exibir o vídeo institucional de cada cena. A segunda trata-se de uma trilha de texto responsável por representar a legenda referente ao vídeo institucional. Nessa trilha, um objeto de mídia texto é referenciado e esse objeto é composto por diversos trechos de texto, cada qual com seu momento de exibição autocontido. A terceira trilha trata-se de uma trilha adaptativa que irá determinar qual o conteúdo para a região identificada por (*rgMaNews*) a partir da análise de expressão (*expr*) presente nas duas trilhas visuais internas. Com base no horário, apenas uma trilha será exibida: uma é responsável pelas notícias institucionais (*default*) ou outra responsável pelas informações de cardápio. Observa-se que, pelo tempo de duração agregado das mídias internas, a reavaliação das expressões irá ocorrer a cada 60 segundos (*repeatCount="indefinite"*) até que a cena atinja sua duração máxima. Essa avaliação ocorre da mesma maneira em todas as outras cenas desse caso de uso.

A partir do momento em que a trilha visual interna, responsável pela exibição do cardápio no dispositivo principal, iniciar, ou seja, quando o horário determinado acontecer, duas trilhas responsáveis por prover o mesmo conteúdo nos dispositivos secundários irão iniciar por meio de um elo causal definido pela trilha interativa, identificada por *itStartMobTracks*. Uma destas trilhas é visual e contém as mesmas imagens da trilha identificada por *vtCardRU*. Por conter os mesmos objetos de mídia, uma cópia foi criada por meio do atributo *refer*. No entanto, a propriedade de região foi sobrescrita por um novo valor (*rgMobRU*). A outra trilha é uma trilha interativa que define por meio de elos o recurso de navegação entre os cardápios diários. Portanto, o

recurso de interatividade estará disponível apenas nos dispositivos secundários, de modo a garantir a liberdade dos usuários na navegação entre as mídias. Observa-se que, dentre todas as trilhas presentes na cena, apenas essas duas não iniciarão no mesmo instante ao da cena, pois o atributo *startMode* foi definido como *deferred*, sendo exibidas apenas por meio do elo causal citado, o qual necessita de uma condição (*onBegin*) para executar duas ações em paralelo (||). A última trilha, identificada por *wgtTempHour*, é utilizada para exibir as informações de temperatura e horário e são representadas como *widgets*, sendo todo seu conteúdo autocontido. Todas as outras cenas fazem uso dos mesmos layouts e também de trilhas parecidas. Portanto, outras cópias das trilhas similares foram definidas por meio do atributo *refer*.

```

<body presentationMode="sequential">
  <!-- Cena sobre UFJF -->
  <scene maxDur="300s" id="scMainUFJF">
    <visualTrack id="vtUFJF" region="rgVideo" repeatCount="indefinite">
      <media id="vidUFJF" type="video" src="..." dur="60s"/>
    </visualTrack>
    <textTrack id="ttLegendaVidUFJF" src="..." type="text/str" dur="60s"
      repeatCount="indefinite" region="rgLeg" />
    <adaptiveTrack id="adtInfoICE" repeatCount="indefinite">
      <visualTrack id="vtCardRU" expr="system.time>='11:00'&&system.time<='14:00"
        region="rgMaNews">
        <media id="imgCardapioDia1" type="image" src="..." dur="12s"/>
        <media id="imgCardapioDia2" type="image" src="..." dur="12s"/>
        <media id="imgCardapioDia3" type="image" src="..." dur="12s"/>
        <media id="imgCardapioDia4" type="image" src="..." dur="12s"/>
        <media id="imgCardapioDia5" type="image" src="..." dur="12s"/>
      </visualTrack>
      <visualTrack id="vtNewsUFJF" region="rgMaNews">
        <media id="imgNewsUFJF1" type="image" src="..." dur="20s"/>
        <media id="imgNewsUFJF2" type="image" src="..." dur="20s"/>
        <media id="imgNewsUFJF3" type="image" src="..." dur="20s"/>
      </visualTrack>
    </adaptiveTrack>
    <visualTrack id="vtCardRUMob" startMode="deferred" refer="vtCardRU"
      repeatCount="indefinite" region="rgMobRU"/>
    <interactiveTrack id="itCardRU" startMode="deferred" region="rgNavMobRU">
      <autoStart media="btnNext"/> <autoStart media="btnPrev"/>
      <media id="btnNext" type="image" top="0%" left="0%" width="10%" height="100%"
        dur="indefinite"/>
      <media id="btnPrev" type="image" top="0%" left="90%" width="10%" height="100%"
        dur="indefinite"/>
      <link>onSelection btnNext then seek next vtCardRUMob</link>
      <link>onSelection btnPrev then seek prev vtCardRUMob</link>
    </interactiveTrack>
    <interactiveTrack id="itStartMobTracks" region="rgFull" >
      <link>onBegin vtCardRU then start vtCardRUMob || start itCardRU end</link>
    </interactiveTrack>
    <interactiveTrack id="wgtTempHour" type="application/x-itu-iptv-widget"
      src="..." region="rgWid" repeatCount="indefinite" dur="60s"/>
  </scene>
  <adaptiveScene id="adsICE or ICH">
    <!-- Cena sobre ICE -->
    <scene maxDur="300s" id="scICE" expr="system.location=='-21.775304, -43.371353,10'">
      <visualTrack id="vtICE" region="rgVideo Images">
        <media id="vidICE" type="video" src="..." dur="60s"/>
      </visualTrack>
      <textTrack id="ttLegVidICE" src="...str" type="text/str" dur="60s" region="rgLeg"/>
      <adaptiveTrack id="adtNewsICEorCardapio" repeatCount="indefinite">
        <visualTrack id="vtCardRUICE" refer="vtCardRU"/>
        <visualTrack id="vtNewsICE" region="rgMainNews">
          <media id="imgNewsICE1" type="image" src="..." dur="20s"/>
          <media id="imgNewsICE2" type="image" src="..." dur="20s"/>
          <media id="imgNewsICE3" type="image" src="..." dur="20s"/>
        </visualTrack>
      </adaptiveTrack>
      <visualTrack id="vtCardRUMob2" startMode="deferred" refer="vtCardRU"

```

```

        repeatCount="indefinite" region="rgMobRU"/>
<interactiveTrack id="itCardRU2" startMode="deferred" region="rgNavMobRU">
  <autoStart media="btnNext2"/>
  <autoStart media="btnPrev2"/>
  <media id="btnNext2" type="image" top="0%" left="0%" width="10%" height="100%"
    dur="indefinite"/>
  <media id="btnPrev2" type="image" top="0%" left="90%" width="10%" height="100%"
    dur="indefinite"/>
  <link>onSelection btnNext2 then seek next vtCardRUMob2</link>
  <link>onSelection btnPrev2 then seek prev vtCardRUMob2</link>
</interactiveTrack>
<interactiveTrack id="itStartMobTracks2" region="rgFull" >
  <link>onBegin vtCardRUICE then start vtCardRUMob2 || start itCardRU2 end</link>
</interactiveTrack>
<interactiveTrack id="wgtTempHour2" refer="wgtTempHour"/>
</scene>
<!-- Cena sobre ICH -->
<scene maxDur="300s" id="sclCH" expr="system.location=='-21.773223,-43.366100,10'">
  <visualTrack id="vtICH" region="rgVideo">
    <media id="vidICH" type="video" src="..." dur="60s"/>
  </visualTrack>
  <textTrack id="ttLegendaVidICH" src="..." type="text/str" dur="60s" region="rgLeg"/>
  <adaptiveTrack id="adtNewsICHorCardapio" repeatCount="indefinite">
    <visualTrack id="vtCardRUICH" refer="vtCardRU"/>
    <visualTrack id="vtNewsICH" region="rgMaNews">
      <media id="imgNewsICH1" type="image" src="..." dur="20s"/>
      <media id="imgNewsICH2" type="image" src="..." dur="20s"/>
      <media id="imgNewsICH3" type="image" src="..." dur="20s"/>
    </visualTrack>
  </adaptiveTrack>
  <visualTrack id="vtCardRUMob3" startMode="deferred" refer="vtCardRU"
    repeatCount="indefinite" region="rgMobRU"/>
  <interactiveTrack id="itCardRU3" startMode="deferred" region="rgNavMobRU">
    <autoStart media="btnNext3"/>
    <autoStart media="btnPrev3"/>
    <media id="btnNext3" type="image" top="0%" left="0%" width="10%" height="100%"
      dur="indefinite"/>
    <media id="btnPrev3" type="image" top="0%" left="90%" width="10%" height="100%"
      dur="indefinite"/>
    <link>onSelection btnNext3 then seek next vtCardRUMob3</link>
    <link>onSelection btnPrev3 then seek prev vtCardRUMob3</link>
  </interactiveTrack>
  <interactiveTrack id="itStartMobTracks3" region="rgFull">
    <link>onBegin vtCardRUICH then start vtCardRUMob3 || start itCardRU3 end</link>
  </interactiveTrack>
  <interactiveTrack id="wgtTempHour3" refer="wgtTempHour"/>
</scene>
</adaptiveScene>
</body>
</storuml>

```

A proposta de STorM/STorML é buscar unir os benefícios das composições com semânticas incorporadas, como as utilizadas em AHM/SMIL, com a expressividade do paradigma de causalidade utilizado em NCM/NCL, propondo entidades com um maior nível de abstração. Esse alto nível de abstração se refere aos conceitos utilizados nas entidades, comumente vistos nas ferramentas audiovisuais usadas na área da comunicação social, a qual detém grande parte da mão de obra na autoria de conteúdos DOOH. Além disso, apesar de ser um modelo voltado inicialmente para o domínio de DOOH, entende-se que sua utilização é possível em outras áreas da comunicação social, devido a familiaridade com os conceitos usados em outros domínios pertencentes a mesma área.

8.6. Contribuições em indexação e recuperação de multimídia

Em pesquisa realizada no ano de 2013², o YouTube conta com mais de 40 milhões de espectadores no país, sendo que mais de 40% deles se interessam por vídeos que contenham entrevistas, documentários ou tutoriais (Barrère, 2014). Outros repositórios de vídeos também mostram um grande crescimento, mas todos eles apresentam um desafio em comum: como facilitar a busca por vídeos? Cada um adota estratégias diferentes, mas que basicamente envolve o cadastro de informações básicas (título, palavras-chave etc), metadados (principalmente no caso de Objetos de Aprendizagem), legendas, entre outras soluções.

Através do Grupo de Trabalho em Busca Avançada de Vídeos (GT-BAVi), da RNP, o LApIC para indexação não só nas informações básicas de um vídeo (título etc.) e metadados, mas também na transcrição do áudio e posterior análise semântica do mesmo. Isto é feito para ampliar as possibilidades do usuário encontrar um vídeo que contemple sua busca ou para que mecanismos de recomendação possam gerar uma indicação mais adequada para aquele usuário. A abordagem proposta pode ser resumida no fluxo de trabalho abaixo.



Figura 8.7 - Fluxo de trabalho para indexação e busca de vídeos

Na solução proposta, informações do vídeo são extraídas para indexação do conteúdo. Dentre as informações, os metadados dos vídeos e, principalmente, o áudio são utilizados para informação sobre os principais tópicos abordados no vídeo. O áudio é processo para gerar transcrições automáticas. Para este fim, o laboratório tem trabalhado com ferramentas para reconhecimento automático de fala e gerado modelos de transcrição especializados nos cenários de aplicação do GT-BAVi. Atualmente, os modelos para reconhecimento automático de fala gerados no LApIC possuem taxa de erro de palavras (WER) de aproximadamente 25% para a base de vídeos utilizados para teste, o que é próximo da taxa de erro dos principais serviços de transcrição automática disponíveis.

Através do conjunto de informação textual gerado, são aplicadas técnicas para anotação automática do texto. Neste caso, a técnica de anotação automática precisa ser definida de forma a lidar com textos ruidosos, visto que a transcrição automática é um processo que pode gerar frases semanticamente distintas do que foi falado no áudio. Assim, técnicas de recuperação de informação e de processamento de linguagem natural podem gerar baixa acurácia. Neste contexto, o laboratório tem explorado técnicas para anotação automática de vídeos sensíveis a ruídos na transcrição. A abordagem disponibilizada pelo laboratório permite que os tópicos principais abordados nos falantes do vídeo sejam anotados automaticamente. De forma a manter uma solução

² <http://www.proxima.com.br/home/negocios/2013/07/25/Brasil-e-um-dos-paises-que-mais-cresce-em-consumo-de-videos-online.html>

adaptável a um novo conhecimento de mundo, esta abordagem utiliza a base de conhecimentos da DBpedia. Para tal, os recursos (conceitos) da DBpedia são processados, utilizando seus rótulos, ligações com outros recursos, propriedades e a hierarquia de categorias. Com essas informações, é gerado um espaço vetorial de assuntos presentes na base de conhecimento, o qual é comparado com o espaço vetorial de possíveis assuntos gerado através dos tokens presentes no texto extraído do áudio do vídeo, o que permite determinar qual o tópico principal que o vídeo está tratando e quais os prováveis recursos (conceitos) abordados pelo vídeo. Como este processo é dependente da informação contida na base de conhecimento, novas entidades ou relações que venham a surgir nesta base podem contribuir, de forma transparente, para novas associações de entidades aos vídeos processados pela solução. Vale ressaltar, também, que as técnicas utilizadas para anotação automática do transcrito podem ser adaptadas para outras bases de conhecimento abertas (como o Freebase ou Schema.org) ou proprietárias (como ontologias ou *thesaurus* institucionais).

Uma vez que os vídeos são anotados automaticamente, é possível utilizar essa informação para encontrar vídeos de conteúdos similares, de forma a recomendar vídeos a um usuário. Porém, vídeos podem ser considerados similares por serem anotados com as mesmas entidades ou por compartilharem de um mesmo assunto, embora tratando de tópicos distintos deste assunto. Para retornar vídeos similares ao usuário, é utilizado o grafo de categorias da DBpedia, de forma a encontrar nós em comuns que ligam entidades presentes em cada vídeo. Dessa forma, é possível reconhecer como semelhantes vídeos com entidades distintas, mas que possuem alguma interseção nos tópicos abordados. O algoritmo utilizado para determinar a similaridade entre vídeos com base nas suas anotações, cria um conjunto de categorias associadas ao vídeo da seguinte forma: para cada anotação presente no vídeo, são recuperadas todas as categorias associadas, na DBpedia, a esse recurso. Para cada categoria c , são recuperadas todas as categorias mais abrangentes e mais específicas de c . O mesmo é feito para todos os outros vídeos do repositório e os conjuntos são comparados utilizando o coeficiente de Dice. Como um recurso da DBpedia pode estar ligado a diversas categorias e como as ligações entre categorias da DBpedia forma um gráfico que pode conter diversos ciclos, o conjunto gerado para cada vídeo representa, de forma mais adequada, uma gama de tópicos pertinentes ao assunto principal do vídeo. Nossos experimentos mostram que a abordagem alcança uma acurácia de até 93% de cobertura e 67% de TopN em vídeos da RNP (Dias et al, 2017).

A Figura 8.8 mostra, de forma resumida, os resultados alcançados pelo GT-BAVi para cada vídeo da base de avaliação utilizada (41 vídeos no total), todos vídeos do portal VideoAula@RNP. Os resultados estão ordenados de forma decrescente por precisão ou cobertura, de forma a mostrar a curva de decaimento dos resultados. Assim, quanto mais constante (valor 1) for a reta, melhor. As linhas de precisão e de cobertura são independentes neste gráfico. Nota-se, então, que mais da metade dos vídeos testados alcançaram uma cobertura máxima (valor 1), enquanto a menor cobertura alcançada foi de 40%. Em relação à precisão, cerca de 25% da base alcançou precisão máxima (ou seja, foram relacionados apenas os vídeos esperados, sem gerar falsos positivos), enquanto houve vídeos em que a precisão ficou próximo de zero. Tal resultado mostra que em todas os testes foi possível relacionar os vídeos corretos, embora, em alguns casos, foram relacionados mais vídeos do que o esperado (falsos positivos).

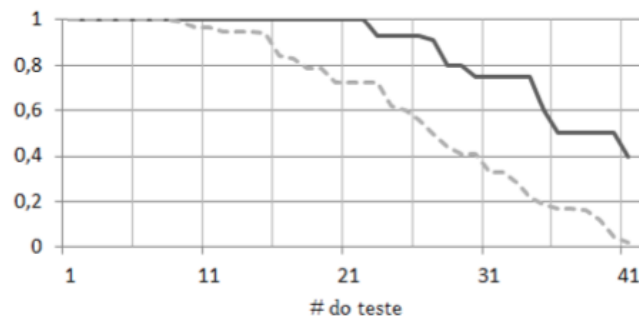


Figura 8.8 - Cobertura (linha contínua) e Precisão (linha tracejada) alcançada para cada vídeo da base de avaliação (eixo X) (Dias et al, 2017).

Outra contribuição nessa área está relacionada à segmentação de cenas em vídeos educacionais, videoaulas. Tal processo faz uso das etapas de transcrição e anotação semântica citadas acima e é composto por uma arquitetura, vide Figura 8.9, formada por três módulos principais (Soares e Barrére, 2017): ASR, Anotação Semântica e *Framework* de geração de cenas.

O ASR (Reconhecimento Automático de Fala - *Automatic Speech Recognition*) é responsável pela transcrição do áudio da videoaula. Esse áudio é segmentado em tomadas (*shots* - trecho de áudio com fala ininterrupta, geralmente representa frases ou parágrafos).

O módulo de Anotação Semântica associa conceitos existentes em uma base de conhecimento aos textos transcritos pelo módulo ASR. Para cada texto são associados termos existentes em uma base de dados ligados e feita uma expansão da anotação semântica com a finalidade de determinar as categorias e subcategorias de cada termo encontrado.

Por fim, o *framework* de geração de cenas utiliza as informações resultantes da anotação semântica para gerar as cenas da videoaula. Todos termos, categorias e subcategorias obtidos para todos os textos de transcrição formam um dicionário de palavras. A partir desse dicionário e sabendo a ocorrência de cada palavra em cada tomada, obtém-se uma matriz, na qual as linhas representam as palavras do dicionário e as colunas as tomadas de áudio. Nessa matriz são computados os pesos de cada palavra do dicionário em cada tomada de áudio. Cada coluna da matriz representa um vetor de características. Medidas de distância ou similaridade entre os vetores de características podem ser calculadas para guiar o algoritmo para agrupar tomadas adjacentes e similares, gerando assim as cenas.

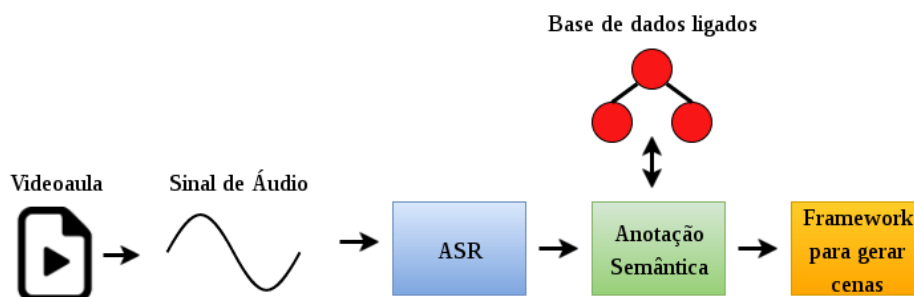


Figura 8.9 - Arquitetura para o processo de segmentação automática de videoaulas em cenas (Soares e Barrére, 2017).

Como produto final, obtêm-se um conjunto de cenas (divisão temporal do vídeo) com seus respectivos termos de busca, viabilizando que ações como permitir ao usuário acessar a cena na qual o termo buscado está contextualização sejam possíveis. Assim é possível: navegação do vídeo por cenas, divisão do vídeo em pequenos vídeos contextualizados, entre outras aplicações.

8.7. Contribuições sobre a entrega e consumo de multimídia

8.7.1. Estratégias de transmissão multimídia em modo push

Alguns tipos de aplicações multimídia, notadamente aplicações hipermídia, dão suporte à interação do usuário, permitindo que este decida por um dentre diferentes caminhos de reprodução do conteúdo. Cada caminho de reprodução pode ser composto por um conjunto distinto de objetos de mídia, definido pelo autor da aplicação. Não somente nos pontos de interatividade, mas também nas decisões de adaptação ao contexto do usuário, o autor acaba por especificar diferentes caminhos de reprodução.

Quando aplicações hipermídia são transmitidas em modo *push* através de um único canal de comunicação em grupo (*broadcast* ou *multicast*), o servidor não tem conhecimento sobre qual caminho de interação seus receptores irão seguir, e, por isso, deve enviar todos os objetos de mídia da aplicação. Nesse caso, cada receptor do grupo irá receber todos os objetos de mídia, independente do caminho de reprodução seguido, levando a desperdícios de memória no dispositivo receptor e de vazão do canal de transmissão. Esse cenário é típico de plataformas de colaboração baseadas em IPTV e TV Digital Terrestre.

Com o objetivo de evitar tal desperdício de recursos, Josué (2017) propõe extensões e refinamentos a um framework introduzido em [Josué, 2016], que, até então, suportava apenas a análise do comportamento determinístico da aplicação, sem considerar os diferentes caminhos de reprodução que ela pode oferecer. Josué (2017) propõe outras duas estratégias de transmissão de aplicações hipermídia através do envio do conteúdo em múltiplos grupos *multicast*. As estratégias propostas se baseiam na análise da especificação da aplicação, a fim de se obter os possíveis caminhos de reprodução do conteúdo, para que então os instantes de transmissão de cada objeto de mídia sejam estabelecidos.

Aplicações hipermídia que contêm eventos não-determinísticos podem ser apresentadas de diferentes maneiras para diferentes usuários, ou para um mesmo usuário

em reproduções distintas. Nos cenários onde o conteúdo é transmitido utilizando um único canal, o emissor deve enviar todos os objetos de mídia que compõem a aplicação, independente do caminho de reprodução que está sendo escolhido pelo usuário. Assim, o receptor potencialmente armazenará objetos de mídia que não serão utilizados pelo usuário, levando a desperdício de memória.

Em cenários que dão suporte a transmissão de conteúdo em múltiplos canais, a análise de eventos não-determinísticos pode ser empregada para evitar esse desperdício no lado receptor. Neste contexto, Josué (2017) explora a instanciação de cenários nos quais o plano de transmissão é construído com base nos eventos determinísticos e não-determinísticos definidos na aplicação.

O framework proposto especifica componentes que atuam no gerenciamento da transmissão (Figura 8.10), além da entrega e recepção do conteúdo hiperídia. A análise da aplicação é definida na classe `{ApplicationAnalyzer}` onde a especificação da aplicação (método `LoadApplication`) é carregada para extrair suas características, como o tempo de início e fim de cada objeto de mídia, e as relações temporais entre eles.

Essas informações possibilitam escalonar a transmissão dos objetos de mídia e construir um Grafo Hiperídia Temporal (HTG) [Costa, 2008], capaz de modelar o comportamento de aplicações hiperídia. Em seguida, inicia-se o preenchimento do plano de transmissão (`TransmissionPlan`), que terá seus dados complementados pelo método `updateTransmissionTime()`, de acordo com uma estratégia definida em `TransmissionStrategy`, gerando o Plano de Transmissão final. A partir dessa estrutura é possível visualizar os objetos de mídia necessários em um intervalo de tempo, para que o conteúdo seja enviado no tempo correto de sua exibição nos clientes, evitando que haja perda de sincronização.

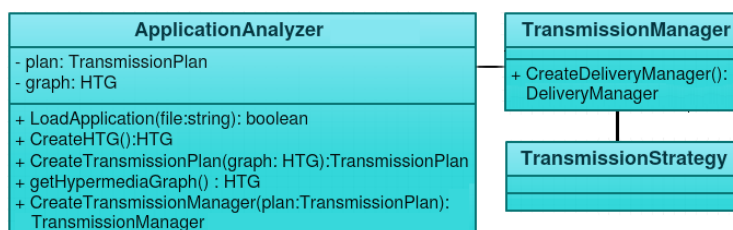


Figura 8.10. Componentes de análise e gerenciamento da transmissão (servidor)

Como o framework dá suporte a diferentes cenários de provisão de serviços hiperídia, diferentes estratégias de transmissão podem ser empregadas, através da especialização do componente `TransmissionStrategy`.

Já as tarefas relacionadas à entrega do conteúdo são implementadas pela classe `DataSender`, que se comunica com o gerenciador de entrega `DeliveryManager` e implementa o protocolo de entrega do conteúdo (Figura 8.11). A definição e manipulação da estrutura utilizada para envio dos dados entre o servidor e o receptor é especificada pelo gerenciador de entrega, que pode ser especializado de acordo com o mecanismo de entrega escolhido.

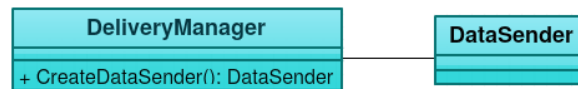


Figura 8.11. Componentes de gerenciamento da entrega de dados (servidor)

No lado cliente, o componente *DataReceiver* estabelece a conexão com o emissor para dar início ao recebimento de dados, e implementa o protocolo de entrega de conteúdo. *ApplicationMiddleware* representa o componente que controla a apresentação da aplicação hipermídia, coordenando o carregamento de players de mídia, as interações do usuário com o conteúdo, entre outras atividades (Figura 8.12).

Algumas estratégias de transmissão podem alterar o documento hipermídia como parte da sua técnica de envio dos objetos de mídia. Essas alterações podem fazer com que a especificação não seja entendida diretamente pelo middleware, necessitando ser pré-processada por um componente intermediário. Então, este trabalho propõe a adição de um novo componente *ReceptionManager*, que é acoplado ao middleware para possibilitar sua integração aos diferentes sistemas hipermídia, com características de entrega de conteúdo distintas.

Quando o receptor obtém o documento hipermídia, o gerenciador de recepção realiza um pré-processamento da aplicação, para obter informações como o local de onde buscar os objetos de mídia, podendo, inclusive, alterar o documento recebido para maior adequação ao cenário de provisão do serviço.

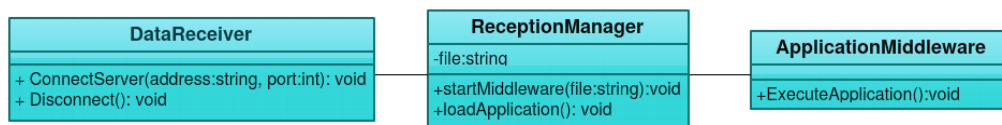


Figura 8.12. Componentes de gerenciamento da recepção de dados (cliente)

A transmissão de conteúdo hipermídia pode se dar em diferentes meios, seja por radiodifusão, redes gerenciadas, ou até mesmo através da Internet. Josué (2017) analisou quatro cenários de transmissão em modo *push*, listados a seguir.

- Cenário 1: Os dados que compõem o conteúdo hipermídia são enviados por *broadcast*, sem empregar nenhuma análise dos eventos definidos na aplicação para gerenciamento da transmissão.
- Cenário 2: Os dados são enviados por *broadcast* ou *multicast*, e considera apenas os eventos determinísticos definidos pelo autor da aplicação para gerenciar a entrega conteúdo.
- Cenário 3: Os dados são enviados por *multicast*, utilizando grupos de comunicação múltiplos e independentes.
- Cenário 4: Os dados são enviados por *multicast*, utilizando grupos de comunicação múltiplos e complementares entre si.

Esses cenários são obtidos através da combinação dos métodos de transmissão de conteúdo hipermídia (*broadcast* e *multicast*) com estratégias de transmissão, que são componentes especializados de *TransmissionStrategy* definido pelo framework adaptável.

Nos Cenários 1 e 2, a transmissão do conteúdo hiperímia se dá através de um único canal, por onde todos os objetos de míia que compõem a aplicação devem ser enviados. Por isso, nesses casos nenhuma melhoria relacionada a eventos não-determinísticos pode ser empregada para a entrega de conteúdo hiperímia. Já nos novos cenários 3 e 4, como a transmissão pode ocorrer através de múltiplos canais de comunicação, são considerados eventos determinísticos e não-determinísticos definidos na aplicação, para gerenciar a entrega do conteúdo.

No envio por *multicast* com grupos independentes (Cenário 3) cada grupo *multicast* é composto por objetos de míia relacionados a uma opção de caminho a ser seguida ao longo da apresentação. Esses caminhos são identificados no HTG pela cadeia principal e pelas cadeias secundárias. Para cada cadeia secundária, deve-se criar um novo grupo de comunicação *multicast*, contendo os novos objetos de míia que compõem aquele caminho, além dos objetos de míia da cadeia principal.

Ao enviar conteúdo hiperímia em um cenário de *multicast* com grupos complementares (Cenário 4), um caminho de reprodução do conteúdo é dado pela junção de um ou mais grupos *multicast*. Nesse caso, não há repetição de objetos de míias entre os grupos. Para tal cenário, os objetos de míia definidos na cadeia principal do HTG são colocados em um grupo base, e o conjunto de novos objetos de míia relacionado à cada cadeia secundária, é disposto em novos grupos de comunicação.

A especialização do framework para abrigar as diferentes estratégias de gerenciamento da transmissão é ilustrada pela Figura 8.13. Na primeira estratégia (*TS_Basic*), todos os objetos de míia que compõem a aplicação são enviados durante seu intervalo de transmissão. Então, o instante de inserção de todos os objetos de míia é configurado no Plano de Transmissão com o valor "0", e o instante de exclusão com valor igual ao instante de fim da transmissão da aplicação.

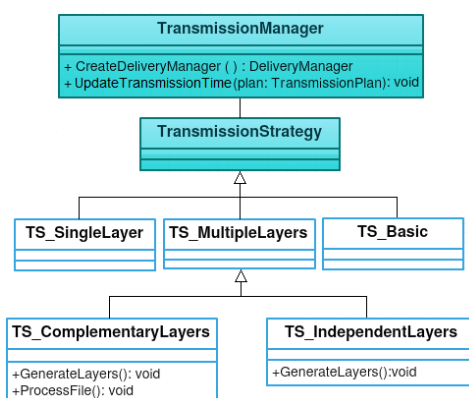


Figura 8.13. Especialização do framework para os 4 cenários

Já a estratégia *TS_SingleLayer* considera a ocorrência de eventos determinísticos para construir o Plano de Transmissão, com base no cálculo do momento em que cada objeto de míia é necessário na apresentação do conteúdo hiperímia, e o tempo necessário para transferi-la.

Na estratégia *TS_IndependentLayers*, os objetos de míia que compõem a aplicação são enviados em grupos *multicast* independentes. E na estratégia *TS_ComplementaryLayers*, os objetos são enviados em grupos *multicast* que se

complementam. A distribuição dos objetos de mídia entre as camadas de conteúdo é dada pelo caminhamento no HTG, considerando as cadeias secundárias do grafo. No caso de envio em camadas independentes (*TS_IndependentLayers*}), cada camada é composta por objetos de mídia relacionados a uma opção de caminho a ser seguida ao longo da apresentação. Desse modo, cada caminho é associado a uma camada que é enviada em um grupo *multicast*. A associação entre o grupo *multicast* e a camada de envio é feita pelo gerenciador da transmissão, e enviada ao receptor, para que este possa definir a qual grupo se conectar, conforme o comportamento de execução da aplicação.

No envio por camadas complementares (*TS_Complementary-Layers*}), um caminho de reprodução da aplicação é dado pela junção de uma ou mais camadas com objetos de mídia. Nesse caso, é possível estabelecer uma relação entre mídia e grupo de comunicação *multicast*, de modo que a localização de cada objeto de mídia na especificação da aplicação deve ser alterado para apontar para um endereço específico (*processFile()*). Para esta estratégia, os objetos de mídia definidos na cadeia principal do HTG são colocados em uma camada comum, e o conjunto de novos objetos de mídia relacionado a cada cadeia secundária é disposto em novas camadas de conteúdo.

As quatro estratégias foram aplicadas aos seus respectivos cenários e o consumo de banda e utilização da rede foram avaliados a partir da métrica ocupação média do carrossel, ou seja, a quantidade de dados transportada ao longo da transmissão da aplicação. Nos cenários de transmissão que empregam o envio em multicamadas, foi observada também a quantidade de dados recebida pelo receptor.

Os resultados apresentados em (Josué, 2017) mostram que, no caso de aplicações sem eventos não-determinísticos, o desempenho dos Cenários 2, 3 e 4, tratando-se da ocupação do carrossel, é o mesmo. Isso porque nesses três cenários a construção do Plano de Transmissão se baseia na ocorrência de eventos determinísticos definidos na aplicação. Já no Cenário 1, onde a aplicação é enviada sem qualquer gerenciamento, o carrossel é estático, ou seja, seu tamanho é total e não varia ao longo da aplicação, independente de suas características de sincronismo temporal.

Com aplicações que possuem eventos não-determinísticos, no *multicast* com camadas independentes (Cenário 3), pode ocorrer repetição de objetos de mídia entre elas. Isso ocorre porque todas as camadas transmitem os objetos de mídia que compõem a cadeia principal do HTG. Essa característica foi observada pela diferença a maior na quantidade média de dados enviada nos carrosséis, entre os Cenários 3 e 4.

Por fim, além da economia na ocupação do carrossel, devido à análise de eventos determinísticos, os cenários que utilizam o envio por *multicast* com múltiplas camadas reduzem a quantidade de dados recebida pelos usuários, pois torna-se possível direcionar os dados em resposta à ocorrência de eventos não-determinísticos, conforme o comportamento dos usuários.

8.7.2. Medição de audiência em sistemas de TV Digital e IPTV

A maioria das emissoras ao redor do mundo são de capital privado e por isso veem na venda de seu espaço publicitário a principal forma de obter lucro. Neste cenário, a medição e análise de audiência se torna fundamental para as emissoras, e mais ainda para os anunciantes que desejam que sua propaganda atinja o maior número de pessoas

possíveis. Além disso, um ponto importante é a preocupação que a propaganda possa atingir o público-alvo ao qual a se destina.

O IBOPE é o principal instituto de pesquisa em audiência na TV aberta no Brasil e utiliza basicamente duas metodologias de trabalho. A mais obsoleta é chamada caderno, na qual o telespectador deve preencher um formulário informando a programação assistida em intervalos de 15 minutos num período de duas semanas. A segunda forma, mais sofisticada, utiliza um dispositivo chamado *peoplemeter*, que é instalado na casa do telespectador. Devido ao seu elevado custo, o IBOPE limita o uso deste aparelho a poucos domicílios. Estas duas formas de medir audiência no Brasil não são muito eficientes. No caso do caderno, a medição depende totalmente do usuário para informar à programação assistida, o que pode gerar dados que não correspondem à realidade. Em relação ao *peoplemeter*, o seu custo elevado faz com que apenas algumas poucas residências tenham seus dados mensurados. Isso acarreta em uma medição de audiência que não representa bem o cenário em todo país.

Com o advento da Televisão Digital, Televisão Digital Interativa e IPTV, novas possibilidades na medição de audiência surgem. Esta seção está dedicada a discutir as possibilidades de medição de audiência nesses novos cenários.

8.7.2.1 Medição de audiência em sistemas de TV Digital

O LApIC criou soluções para a implementação de provedores de serviço de análise de interação e audiência (Basilio, 2013a), como uma evolução tecnológica que pode ser adotada por institutos de pesquisa desta área. Foram propostas e comparadas duas abordagens para a captura de dados de audiência e de comportamento dos usuários de sistemas de TV Digital Interativa (TVDI). Além disso, foi proposto o IAASP (*Interaction and Audience Analysis Service Provider*) como um conjunto de ferramentas para análise de interação e audiência, que permitem a manipulação dos dados originados das interações dos telespectadores com seus receptores digitais, independentemente do modelo de captura escolhido. Tais propostas estão alinhadas às recomendações da União Internacional de Telecomunicações (UIT) para TV terrestre, a cabo e IPTV.

A primeira abordagem para captura dos dados requer o desenvolvimento de extensões aos padrões de *middleware* já existentes, enquanto a segunda faz uso apenas dos recursos comumente suportados por estes padrões. Um benefício imediato do uso destas soluções para captura de dados é o fato de que qualquer telespectador que possuir um receptor digital nos referidos padrões e com um canal de retorno habilitado, é um participante em potencial para a análise. Podendo contar com amostras maiores, a análise pode se tornar mais abrangente, precisa e dinamicamente estruturada. Além disso, custos de *hardware* e logística podem ser substituídos pelo custo incorrido no desenvolvimento de *software* e transmissão dos dados, o que de fato permite uma variedade de modelos de negócio. Finalmente, o uso de recursos já presentes nos receptores digitais dos telespectadores torna latente aos usuários o fato de que sua audiência está sendo mensurada. Obviamente, por motivos de privacidade, os usuários devem estar cientes e permitir a captura antes que ela se inicie.

A principal contribuição da proposta do LApIC é que qualquer receptor digital, especialmente aqueles em conformidade com as recomendações da UIT, se torna capaz de realizar a captura das interações dos usuários. Ainda que vários artigos abordem a

questão de medição de audiência e proponham novos modelos de medição, nenhum considerava por completo as interações dos usuários em sistemas de TV Interativa.

Além disso, a arquitetura do IAASP proposta e suas estruturas de dados permitem à próxima geração de análise de mídias uma maior abrangência e uma habilidade única de analisar múltiplas plataformas através de uma perfeita harmonização dos dados provenientes de diferentes fontes de conteúdo como TV terrestre, a cabo, satélite, IPTV e serviços de TV via Internet.

A Figura 8.14 apresenta uma visão geral da arquitetura para análise de interação e audiência, conforme proposto em (Basilio, 2013a), identificando os principais atores, como o provedor de serviços de televisão, os receptores digitais e o IAASP. O provedor de serviços de televisão é o componente responsável por transmitir o conteúdo a ser consumido, incluindo áudio, vídeo e dados (aplicações). Exemplos de provedores de serviços de televisão incluem emissoras de TV aberta, TV por assinatura e por Internet.

Na figura, a seta entre o provedor de serviços de televisão e o dispositivo terminal representa a rede de distribuição. O dispositivo terminal representa o receptor de conteúdo do usuário (por exemplo o receptor de TV digital, o set-top box de IPTV, o computador pessoal, etc.).

A arquitetura apresentada difere nas duas propostas no que se refere às formas de captura dos dados, as quais foram nomeadas: Extensão do Middleware (Figura 8.14 (a)) e Aplicações Interativas (Figura 8.14 (b)). Essas abordagens podem ser escolhidas como métodos alternativos ou complementares já que as estruturas para armazenamento dos dados capturados são iguais para ambas abordagens. A escolha entre uma abordagem ou outra, ou mesmo ambas, é uma negociação entre o IAASP e os provedores de serviços de televisão, considerando seus respectivos modelos de negócio e o impacto sobre os telespectadores participantes.

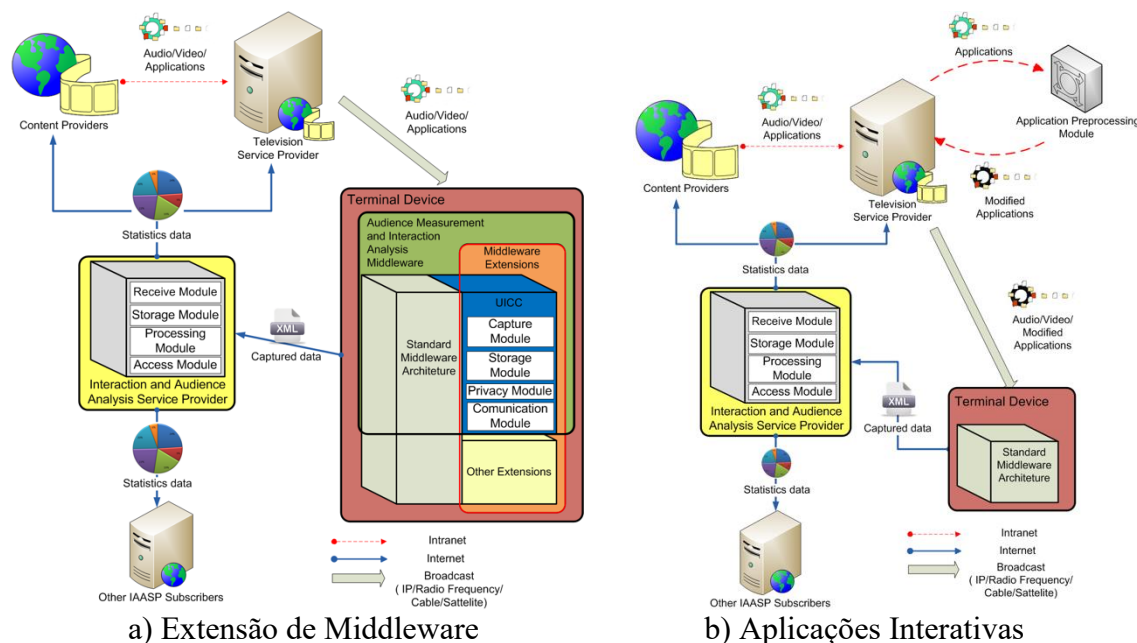


Figura 8.14 – Visão Geral da Arquitetura para Análise de Interação e Audiência

Dependendo do modelo de negócio adotado, o provedor de serviços pode ou não requerer um módulo para tratar as aplicações interativas. Este módulo é responsável por manipular as aplicações interativas antes do seu envio, para que assim sejam capazes de realizar a captura dos dados para a análise de interação e audiência.

Da mesma forma que o provedor de serviços deve se adaptar ao modelo de negócio escolhido, o receptor digital do usuário também deve estar preparado de acordo com o modelo de negócio escolhido. Para o usuário, esse preparo pode ser transparente ou não. Independentemente do modelo de negócio escolhido, dentre os propostos, fica a cargo do receptor digital capturar os dados para a análise de interação e audiência.

Se o modelo adotado utilizar extensões do *middleware*, o receptor digital terá um componente anexado (usualmente instalado através de atualização do *firmware*) que será responsável pela captura, armazenamento e transmissão dos dados das interações e da audiência. Se o modelo adotado não utilizar extensões instaladas no receptor digital, as próprias aplicações interativas informarão ao IAASP as interações dos usuários. Neste caso não há necessidade de novos módulos serem instalados no receptor digital do usuário.

O Provedor de Serviços de Análise de Interação e Audiência é responsável por receber, armazenar, processar e disponibilizar os dados das interações dos usuários e sua audiência. Para cada uma destas funções, o IAASP tem um módulo específico.

O Módulo de Recebimento provê uma interface onde qualquer tipo de receptor digital pode enviar os dados capturados independentemente do modelo de negócio escolhido. Logo após o recebimento, o IAASP, através do Módulo de Armazenamento, armazena os dados recebidos de forma a preservar a estrutura original dos dados, sem perder ou adicionar qualquer informação.

O Módulo Processador é responsável por adicionar dados externos na base de dados. Alguns dados são úteis para alguns tipos de consulta, mas não estão disponíveis no receptor digital do usuário. Esse tipo de dado deve ser inserido diretamente na base de dados. É também função do Módulo Processador transformar os dados da base original se necessário, para que assim novas e otimizadas consultas possam ser realizadas. Esse processo é conhecido como ETL (*Extract, Transform and Load* - Extração, Transformação e Carga). O Módulo de Acesso disponibiliza os dados das bases originais e modificadas via interfaces web.

A tabela abaixo compara os modelos de medição propostos. Maiores detalhes podem ser encontrados em (Basilio, 2013b)

	Extensão do Middleware	Aplicações Interativas	Peoplemeter
Custo HW/SW	Médio	Baixo	Alto
Dependência do provedor de serviço de TV	Nenhuma	Total	Nenhuma
Tamanho do espaço amostral	Pequeno	Grande	Muito pequeno
Portabilidade	Não	Sim	Não
Tipo de canal de retorno	Internet	Internet	Internet ou canal dedicado
Característica do	menos transferências	mais transferências	mais transferências

tráfego gerado	maior tamanho	menor tamanho	menor tamanho
Evento capturáveis	Zapping, interações, volume, EPG...	Zapping, interações	Zapping

8.7.2.2 Medição de audiência em IPTV

A própria característica do IPTV de funcionar sob uma rede IP, fornece um canal de retorno entre o dispositivo terminal e servidores de medição de audiência, o que torna a troca de dados de audiência mais natural. Porém, apesar dessa facilidade, existem complicações neste cenário. Nos serviços IPTV várias tecnologias proprietárias são utilizadas, como codificações de áudio, vídeo e dados. Além disso, cada provedor IPTV pode fazer a medição de audiência de uma forma diferente, o que traz muitas dificuldades na hora de agregar todos esses dados vindos de diferentes provedores de medição de audiência utilizando tecnologias distintas.

Para amenizar as dificuldades no cenário de medição de audiência em IPTV, este trabalho propõe a implementação de uma arquitetura para o serviço de medição de audiência, conforme padronização definida pela UIT (União Internacional de Telecomunicações). Nesta arquitetura, a captura dos dados se baseia em eventos, possibilitando assim, além da medição de audiência em TV linear e vídeo sob demanda, também capturar as interações do usuário com aplicações interativas.

Outra vantagem da medição aqui proposta, é que ela é feita toda por *software*, não necessitando assim, de nenhum equipamento adicional para medir a audiência e interatividade dos usuários. O canal de comunicação entre o terminal IPTV, onde as capturas são feitas, e o servidor de medição de audiência, onde os dados são processados, é a internet. Visto que para o usuário acessar um serviço IPTV é necessária uma conexão banda larga, sem nenhuma infraestrutura adicional.

A arquitetura proposta por Soares e Barrére (2016) é composta por três módulos principais: TDMiddleware, *Terminal Device Audience Measurement Functions* (TD-AMFs) e *Aggregation Functions* (AFs). Estes módulos se comunicam através de redes IP, sendo suportados tanto IPv4 quanto IPv6. No caso da comunicação entre TDMiddleware e TD-AMFs é feita uma comunicação local no dispositivo terminal, já na comunicação entre TD-AMFs e AFs pode ser utilizada a internet. A Figura 8.15 apresenta um esquema da arquitetura e suas comunicações. A captura dos dados de audiência e interatividade do usuário ocorre no terminal IPTV e os mesmos são enviados através da internet para o servidor de medição de audiência. O servidor, por sua vez, funciona como um intermediário entre os dados capturados no terminal e os interessados nos dados de audiência e interatividade.

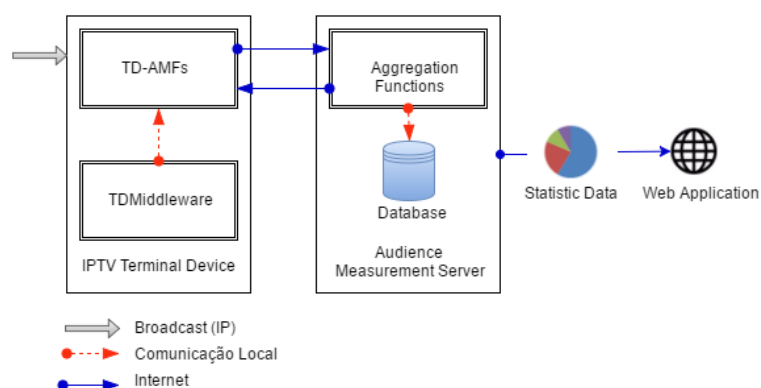


Figura 8.15 - Arquitetura do serviço de medição de audiência e interatividade em IPTV.

O TDMiddleware é o middleware para o terminal IPTV desenvolvido pelo LApIC, conforme recomendação ITU-T H.721, sendo responsável pelo controle do ciclo de vida do dispositivo terminal, gerenciando todos seus módulos. Desta forma, para realizar a captura baseada em eventos, basta que o módulo onde ocorreu o evento notifique o TD-Middleware e este, por sua vez, notifique o módulo TD-AMFs.

O Terminal Device Audience Measurement Functions - TD-AMFs é responsável por gerar os relatórios de medição de audiência e interatividade com base nas notificações de evento recebidas do TDMiddleware.

Já o Aggregation Functions - AFs é o módulo executado no servidor de medição de audiência e interatividade. Responsável por enviar as configurações de medição de audiência citadas na subseção anterior para o módulo TD-AMFs dos dispositivos terminais dos clientes que aceitaram ter sua audiência mensurada. Também tem a função de receber os reports desses terminais, processá-los e agregar informações adicionais que podem ser relevantes, como feriados, grandes eventos que estão acontecendo no momento da medição de audiência, localidades, etc.

A captura de dados no terminal IPTV deve ter a permissão do usuário. No momento em que o terminal IPTV descobre o serviço de medição de audiência na rede, é perguntado ao usuário se ele deseja ter sua audiência mensurada, em seguida é perguntado qual nível de permissão o usuário deseja atribuir à sua permissão. São definidos 4 níveis pela UIT em ITU-T H.741.1:

- Nível 0: Apenas dados de medição serão incluídos nos *reports* de medição.
- Nível 1: Identificadores anônimos para distinguir um usuário do outro são adicionados, além dos dados de medição de audiência.
- Nível 2: Informações genéricas do usuário são adicionadas aos *reports*, como sexo, idade, além das informações dos níveis anteriores.
- Nível 3: Além de toda informação dos outros níveis, são incluídas informações identificadoras do usuário.

A abordagem proposta utiliza apenas recursos do dispositivo terminal IPTV para realizar a captura dos dados de audiência. Além disso, ela segue as recomendações da UIT (ITU-H.741.0 a ITU-H.741.4).

8.7.3. Máquinas de apresentação de conteúdo hipermídia.

Apesar das facilidades que ferramentas de autoria e linguagens declarativas podem oferecer no processo de codificação de documentos hipermídia, existe a necessidade de avaliações repetitivas do comportamento observável da apresentação do documento. Essa etapa não consiste apenas em reproduzir o documento hipermídia por meio de um exibidor (ou máquina de apresentação), mas também conceder ao desenvolvedor informações e controle de todos os elementos que compõem a apresentação.

Diferentemente do desenvolvimento com linguagens imperativas, ferramentas de depuração são pouco exploradas no âmbito de linguagens declarativas. Isso se deve ao fato de que este tipo de linguagem possui normalmente um alto nível de abstração, apresentando desafios conhecidos, como a suspensão de uma aplicação realçando o trecho do código relacionado a tal instante ou até mesmo a verificação de correção da aplicação depender de eventos imprevisíveis. Sendo assim, mais do que uma arquitetura para a simples reprodução do conteúdo, a máquina de apresentação voltada à depuração precisa estar apoiada em componentes que permitam a captura de informações relativas aos elementos da apresentação em cada instante, além de permitir o controle destes elementos.

Dessa forma, o LAPIC vem definindo uma arquitetura para máquinas de apresentação hipermídia que possibilitem a depuração avançada, incluindo tanto o reconhecimento quanto a alteração dos estados das instâncias de entidades que formam o modelo hipermídia adotado, com baixo acoplamento entre os módulos componentes de tal arquitetura. Recentemente, a etapa de especificação da arquitetura foi finalizada, dando início ao desenvolvimento de uma prova de conceito, que toma o modelo NCM como caso de uso.

A arquitetura está dividida em 3 seções: *Presentation Engine*, *Resources* e *Monitoring & Controlling* (Figura 8.16). *Presentation Engine* é composta pelos módulos responsáveis pela tradução, representação e reprodução do documento hipermídia. *Resources* é a seção que contém os possíveis recursos de entrada e saída, como os players de mídia, teclado alfanumérico e controle remoto. *Monitoring & Controlling* é a seção responsável por escutar os meios de comunicação entre os módulos ou até mesmo gerar mensagens nestes meios para modificar o comportamento de uma apresentação. Além disso, existe um meio de comunicação que é responsável por interligar as seções *Presentation Engine* e *Resources*, que é chamada de *Resource Broker*.

Como o propósito da arquitetura é auxiliar o desenvolvimento, os módulos e as seções da arquitetura se comunicam através de meios compartilhados. Assim, o monitoramento das mensagens que são trocadas se torna possível, bem como a injeção de mensagens para os módulos, o que permite se ter o processo de depuração.

O fato da arquitetura possuir um meio de comunicação compartilhado e um baixo acoplamento entre os módulos faz com que possua uma elevada escalabilidade.

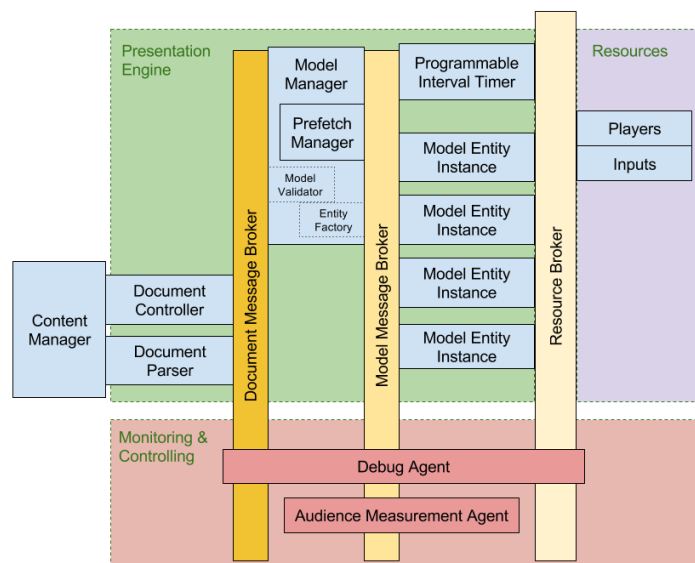


Figura 8.16. Arquitetura de Máquina de Apresentação para Depuração

Na arquitetura estão definidos três meios de comunicação, que são: *Document Message Broker (DMB)*, *Model Message Broker (MMB)* e *Resource Broker (RB)*. Com o intuito de aumentar o desempenho dos canais de comunicação, todos eles são baseados no *pattern Publish-subscribe*. Dessa forma, os módulos podem dizer ao *Broker* quais tipos de eventos são de seu interesse, fazendo com que estes eventos sejam repassados apenas aos interessados. Nesse paradigma é possível se ter uma comunicação em grupos por interesse, onde um conjunto de *subscribers* pode se inscrever com intuito de receber a notificação de um evento específico.

O *Document Parser* é responsável por carregar uma aplicação hipermídia, seja ela escrita em NCL, SMIL ou outra linguagem suportada e agir para a criação do modelo que representa esta aplicação. Cada elemento da linguagem poderá ser responsável pela geração de instâncias de entidades do modelo que a representa. Essas instâncias são geradas pelo *Model Manager*, que tem o papel de não apenas gerá-las, mas também validar se as regras do modelo são respeitadas, como *Namespaces*, identificadores, papéis em relacionamentos e afins. Como o *Model Manager* é responsável por criar estas instâncias, ele tem as informações necessárias para a criação do Plano de pré-busca. Neste aspecto, o sub-módulo *Prefetch Manager* fica responsável por notificar às instâncias para iniciarem o processo de pré-busca no momento correto, baseando-se em um plano que ele mesmo mantém mediante as ações do *Model Manager*.

Além do modelo representado pelo documento hipermídia, comandos de edição podem ser gerados com o propósito de manipulações mais sofisticadas, como a modificação em tempo real, colaboração ou até mesmo a geração da aplicação. Estes comandos são recebidos pelo módulo *Document Controller* e repassados ao *Model Manager*, podendo alterar o comportamento da apresentação.

Os componentes *Model Entity Instance* representam diferentes instâncias de entidades do modelo. Cada instância terá seu papel para gerenciar o comportamento da apresentação mediante ações devido a eventos espaciais, temporais ou mesmo não-determinísticos, como interações do usuário. Essa instância pode representar uma mídia,

o conteúdo em si, ou até mesmo um elo que interliga as mídias no documento hipermídia.

O módulo *Programmable Interval Timer* é responsável por sincronizar os relógios dos recursos com os de *Model Entity Instances*. Além disso, ele é responsável por notificar à uma instância a ocorrência de um evento temporal solicitado, como por exemplo notificar quando um determinado intervalo de tempo tenha se passado a partir do instante atual.

Nota-se que a *Presentation Engine* possui dois meios compartilhados de comunicação: o DMB e MMB. O DMB tem a finalidade de promover a comunicação entre os módulos que fazem modificações no modelo através do *Model Manager*. Outro meio de comunicação é o MMB, que tem a finalidade de estabelecer a comunicação entre o *Model Manager* e as *Model Entity Instances* geradas. Assim como as *Model Entity Instances* comunicam entre si, elas podem ter interesse em controlar recursos, como *players*. Sendo assim, existe um meio de comunicação com a seção *Resources*, que é a *Resource Broker* (RB).

O *Content Manager* é responsável por gerenciar os documentos e os conteúdos hipermídia, como as aplicações e suas respectivas mídias de vídeo e áudio. Ele é responsável por notificar o *Document Parser* para realizar a análise de um documento ou mesmo notificar o *Document Controller* sobre edições a serem feitas na apresentação. Apesar do *Content Manager* se comunicar com o *Document Controller* e o *Document Parser*, ele é um módulo que está fora da seção *Presentation Engine*, por não estar no escopo da máquina de apresentação em si. Caso algum *player* gere comandos ao *Document Controller*, este *player* irá se comunicar com o *Document Controller* através do *Content Manager*.

A seção *Resources* contém todas as representações de recursos de entrada e saída. Nela está contida *Players* de conteúdos como áudio, vídeo ou mesmo de aplicação hipermídia. Nesta abordagem, um *player* de aplicação hipermídia é uma nova instanciação da seção *Presentation Engine* que será responsável por interpretar e apresentar tal parte da aplicação. Além dos *players*, *Resources* contém as entradas de dados, como mouse, toque, teclado e controle remoto. Esses recursos são representados como componentes *Input*. Tanto *Players* como *Inputs* podem ser utilizados pela *Presentation Engine*, bastando fazer a solicitação de uso.

Todos os eventos gerados pelos recursos, sejam de entrada ou saída, como o pressionamento de uma tecla ou o fim natural de um conteúdo, são notificados à instância contida na respectiva *Presentation Engine*, que fez a solicitação de uso de tal recurso. Tanto a solicitação de recursos quanto a notificação de eventos são feitos através da RB.

Pelo fato dos meios de comunicação entre os módulos serem compartilhados, todas as mensagens geradas podem ser ouvidas. Além disso, mensagens também podem ser geradas de forma sintética nesse meio. Sendo assim, a seção *Monitoring & Controlling* inclui os módulos que desejam ouvir estes meios e interpretar as mensagens de acordo com a sua função. Um desses módulos é o *Debug Agent*, que fica responsável por notificar a usuários testadores/desenvolvedores de toda e qualquer informação relevante obtida a partir das mensagens trafegadas durante a apresentação. Como este módulo deseja receber todas as notificações decorrentes da geração e da apresentação hipermídia, ele se inscreve nos *brokers* DMB, MMB e RB, recebendo todas os eventos

gerados, seja de criação e edição do modelo, da apresentação ou até mesmo da utilização dos recursos. Através do *Debug Agent* também pode-se fazer alterações comportamentais da apresentação, como alteração de propriedades de mídia, como volume, visibilidade ou área que ocupam. Como mensagens podem ser geradas, o *Debug Agent* também cria instâncias da entidade do modelo, bastando notificar o *Model Manager* o interesse em fazê-lo. Como existem diferentes perfis de desenvolvedores, desde os que criam o código ou até mesmo os que utilizam geradores de códigos, o *Debug Agent* precisa ter a capacidade de filtrar e traduzir as mensagens de tal forma que seja legível ao perfil adequado ao usuário.

Resource Broker (RB) é um meio de comunicação que tem a função de estabelecer a comunicação dos recursos com os módulos que os utilizarão. Esse *broker* tem um comportamento similar ao *pattern Publish–subscribe* e possui um controlador interno que permite repassar as mensagens aos destinatários corretos. Quando um evento de entrada é gerado por um recurso do tipo Input, uma mensagem é gerada neste broker e é repassada ao Player que está ativo naquele momento. Caso o player ativo não tenha interesse na mensagem, ela é então repassada ao nível acima de uma possível hierarquia de players, como ocorre em documentos aninhados.

8.8. Uma visão do futuro para a colaboração multimídia

Nota-se que boa parte das iniciativas do Grupo tem em comum proposições que levam ao aumento das facilidades ou da qualidade de experiência do usuário de aplicações multimídia, entre as quais inclui-se qualquer modalidade de serviços de streaming de dados audiovisuais ligados, interativos e colaborativos (e.g. jogos online, TV interativa, ensino à distância e em salas de aula inteligentes, visualização remota...). Há também especial preocupação com a viabilidade de modelos de negócio para aplicações multimídia, sempre visando proposições técnicas que levam em conta, por exemplo, aspectos regulatórios, aspectos de mercado e análise de cadeia de valor de um serviço. A experiência do grupo em esforços de extensão inovadora e de transferência de tecnologias o habilita para um caminho consistente de pesquisas aplicadas.

Além disso, o Grupo vê com grande importância, para viabilizar suas propostas, o oferecimento de ferramentas que automatizam ou ao menos facilitem qualquer aspecto da provisão de um serviço multimídia. Acredita-se que todo metadado relevante ao conteúdo deve ser inserido na cena tanto de forma automática, como as transcrições propostas pelo GT-Bavi, quanto para automatização de qualquer passo de um workflow de videocolaboração. Enfim, metadados normalmente inseridos manualmente e repetidas vezes desde a produção até a publicação do conteúdo constituem grande esforço de trabalho e deverão em futuro próximo ser associados automaticamente e de forma compatível com edições colaborativas no conteúdo. Além disso, ao oferecer conteúdo para publicação, ferramentas colaborativas e com alto grau de automação para a construção de serviços são necessárias, principalmente no âmbito de plataformas convergentes como a IpêTeVê.

Para trilhar um caminho rumo à provisão de serviços multimídia de alta qualidade, em altíssimas resoluções de áudio e vídeo, o LApIC vislumbra a necessidade de investir em pesquisa na infraestrutura de comunicação e processamento. Nos próximos 10 anos, deverá haver uma mudança no paradigma da Internet, na qual

recursos de hardware e de software poderão ser descobertos e negociados sob a demanda do usuário, no momento em que inicia sua interação com um serviço multimídia. Tal funcionalidade se tornará mais evidente com a implantação da real Internet das Coisas (IoT), agregada às facilidades de computação em nuvem. Somente dessa forma poderão ser oferecidos os níveis necessários de QoS e QoE que aplicações avançadas vão demandar.

Trata-se de um grande desafio, dada a complexidade dos mecanismos, heterogeneidade das plataformas de hardware e de software, granularidade de gerenciamento e escalabilidade. A solução deve passar por um gerenciamento avançado de recursos em redes de comunicação e sistemas finais e possivelmente explorando os benefícios de virtualização e Redes Definidas por Software (SDN).

Além disso, parte-se do pressuposto nas pesquisas do grupo que toda comunicação deve ser feita por meio de IPv6, de forma a habilitar recursos avançados para comunicação em grupo (*multicast*), descoberta de vizinhança, QoS, e, evidentemente, possibilitar um número elevado de dispositivos conectados e globalmente alcançáveis. E, para isso, vale a pena também investigar modelos de negócios que permitiriam o uso massivo de *multicast* e QoS em alcance global.

Enfim, o LAPIC acredita que a partir dessas iniciativas está colaborando para a obtenção de serviços multimídia de alto desempenho e, ainda, ubíquos. A cada novo dispositivo ou novo recurso disponível ao alcance do usuário, o serviço pode ter mudado seu comportamento, de modo a entregar a melhor forma possível de consumo e interação, com qualidade. Todo esse cenário habilitará o uso inovador de recursos multimídia colaborativos no entretenimento, em espaços públicos e privados e, principalmente, na educação.

Referências

- Barrére, E. Videoaulas: aspectos técnicos, pedagógicos, aplicações e bricolagem. In: Maria Augusta Silveira Netto Nunes, Elizabeth Matos Rocha. (Org.). Anais da Jornada de Atualização em Informática na Educação. 1ed. Dourados: EaD-UFGD, 2014, v. 1, p. 70-105.
- Basilio, S. C. A.; Moreno, M. F.; Barrére, E. Supporting interaction and audience analysis in interactive TV systems. In: the 11th european conference, 2013, Como. Proceedings of the 11th european conference on Interactive TV and video - EuroITV '13. New York: ACM Press. p. 23.
- Basilio, S. C. A.; Moreno, M. F. ; Barrére, E. Análise de Audiência e Interação de Usuários de Sistemas de TV Interativa. Revista de RadioDifusão, v. 7, p. 60-70, 2013.
- Costa, R. M. R.; Moreno, M. F.; Soares, L. F. G. Intermedia synchronization management in DTV systems. In: Proceeding of the eighth ACM symposium, 2008, Sao Paulo. Proceeding of the eighth ACM symposium on Document engineering - DocEng '08. New York: ACM Press. p. 289.

- Dias, L., Barbosa, J., Barrère, B., Souza, J.F. An Approach to Identify Similarity Among Educational Resources Using External Knowledge Bases. *Brazilian Journal of Computers in Education*. 2017, Vol. 25, No 2.
- Freesz Jr., M. A.; Yung, L.; Moreno, M. F. STorM: A Hypermedia Authoring Model for Interactive Digital Out-of-Home Media. In: *Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia)*, 2017, Gramado. *WebMedia '17 23rd Brazilian Symposium on Multimedia and the Web*. New York, NY, USA: ACM, 2017.
- Hardman, L., Bulterman, D., Van Rossum, G. "The Amsterdam Hypermedia Model: adding time and context to the Dexter model". *Communications of the ACM*, Vol. 37, N° 2, February 1994.
- Josué, M. I. P.; Moreno, M. F.; Costa, R. M. R. Hypermedia Content Transmission Plan. In: *the 22nd Brazilian Symposium, 2016, Teresina. Proceedings of the 22nd Brazilian Symposium on Multimedia and the Web - Webmedia '16, 2016*. p. 87.
- Josué, M. I. P.; Moreno, M. F. Costa, R. M. R. An Adaptable Transmission Management Framework for Push-mode Hypermedia Content. In: *Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia)*, 2017, Gramado. *WebMedia '17 23rd Brazilian Symposium on Multimedia and the Web*. New York, NY, USA: ACM, 2017.
- Soares, E. R; Barrère, E. Medição de Audiência e Interatividade em IPTV. In: *Workshop de Iniciação Científica do Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia)*, 2016.
- Soares, E. R; Barrère, E. An Approach for Automatic Segmentation of Scenes in Educational Videos through the use of Audio Transcription and Semantic Annotation. In: *Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia)*, 2017, Gramado. *WebMedia '17 23rd Brazilian Symposium on Multimedia and the Web*. New York, NY, USA: ACM, 2017.
- Soares, L. F. G.; Rodrigues, R. F. Nested context model 3.0: Part 1 - NCM core. *Monografias em Ciência da Computação do Departamento de Informática, PUC-Rio*, n. 18/05. Rio de Janeiro, 2005.
- União Internacional de Telecomunicações. Recomendação UIT-T Y.1910. IPTV Functional Architecture. Genebra. Setembro, 2008.
- União Internacional de Telecomunicações. Recomendação UIT-T H.721. IPTV Terminal Devices: Basic Model. Genebra. Abril, 2015.
- União Internacional de Telecomunicações. Recomendação UIT-T H.761. Nested Context Language (NCL) and Ginga-NCL. Genebra. Novembro, 2014.
- União Internacional de Telecomunicações. Recomendação UIT-T H.765. Packaged IPTV application (widget) service. Genebra. Abril, 2015.

ANEXO 1: Biografia dos autores



Marcelo F. Moreno é Professor Adjunto do Departamento de Ciência da Computação da UFJF desde 2011, onde integra o quadro permanente do Programa de Pós-Graduação em Ciência da Computação. Possui Mestrado (2002) e Doutorado (2008) em Informática pela PUC-Rio. É Relator (líder) da Questão 13/16 (Multimedia application platforms and end systems for IPTV) da União Internacional de Telecomunicações (UIT-T) e Co-chair do Grupo Intersetorial em Integração Broadcast-Broadband (IRG-IBB). É também coordenador do GT de Middleware e membro suplente do Conselho Deliberativo do Fórum do Sistema Brasileiro de TV Digital (SBTVD). Sua principal área de interesse é Sistemas Multimídia, com foco em infraestrutura para comunicação multimídia, qualidade de serviço, máquinas de apresentação e modelagem/autoria hipermídia.



Eduardo Barrére é Professor Associado do Departamento de Ciência da Computação da Universidade Federal de Juiz de Fora (UFJF), desde 2009, onde integra o quadro permanente dos Programas de Pós-Graduação em Ciência da Computação e Educação Matemática. Possui graduação em Bacharelado em Ciência da Computação - UFSCar (1996), Mestrado em Ciência da Computação - UFSCar (1998) e Doutorado em Engenharia de Sistemas e Computação - COPPE/UFRJ (2007). Desenvolve pesquisas nas áreas de tecnologias para EaD, IPTV e Multimídia, como foco principal na mídia vídeo.



Jairo F. de Souza é Professor Adjunto do Departamento de Ciência da Computação da UFJF desde 2009, onde integra o quadro permanente do Programa de Pós-Graduação em Ciência da Computação. Possui Mestrado em Engenharia de Sistemas e Computação pela COPPE/RJ e Doutorado em Informática pela PUC-Rio. Desenvolve pesquisas em recuperação de informação, processamento de linguagem natural e tecnologias semânticas para web.

Capítulo

9

Aplicações de Aprendizado de Máquina na Codificação de Vídeo HEVC

Guilherme Corrêa¹, Mateus Grellert², Sergio Bampi², Luis A. da Silva Cruz³

¹ Universidade Federal de Pelotas – Programa de Pós-Graduação em Computação – Video Technology Research Group (ViTech)

² Universidade Federal do Rio Grande do Sul – Programa de Pós-Graduação em Computação – Grupo de Microeletrônica (GME)

³ Universidade de Coimbra – Departamento de Engenharia Electrotécnica e de Computadores – Instituto de Telecomunicações (IT)

gcorrea@inf.ufpel.edu.br, mgsilva@inf.ufrgs.br, bampi@inf.ufrgs.br, lcruz@deec.uc.pt

Abstract

This chapter presents an overview of the main works that employ Machine Learning techniques to solve problems related to video coding, especially focusing on the complexity reduction of the High Efficiency Video Coding (HEVC) standard. Initially, the HEVC encoding tools and techniques are reviewed and the main sources of its complexity are exposed. Then, a survey on references that solve different video coding problems using learning-based techniques is provided, followed by a more detailed review of solutions that employ Machine Learning specifically for HEVC complexity reduction. Finally, a discussion on the future of learning-based video coding closes this chapter.

Resumo

Este capítulo apresenta uma visão geral dos principais trabalhos que utilizam técnicas de Aprendizado de Máquina para resolver problemas relacionados à codificação de vídeo, com foco especial na redução de complexidade do padrão HEVC. Inicialmente, as principais ferramentas do HEVC são apresentadas e as razões para a sua alta complexidade são discutidas. Em seguida, uma revisão dos trabalhos que solucionam diferentes problemas da codificação de vídeo através de técnicas de aprendizado é apresentada, seguida por uma descrição das soluções focadas na redução de complexidade do HEVC. Finalmente, uma discussão sobre o futuro da codificação de vídeo com base em Aprendizado de Máquina encerra este capítulo.

9.1. Introdução

Graças ao grande avanço tecnológico pelo qual passa a indústria de semicondutores nas últimas décadas, dispositivos eletrônicos vêm sendo cada vez mais utilizados em grande escala para os mais diversos fins. Mais recentemente, os dispositivos eletrônicos móveis, como computadores portáteis, smartphones e tablets, passaram a incorporar funções antes desempenhadas apenas por dispositivos fixos com maior poder computacional, substituindo-os em determinadas funções, especialmente aquelas que envolvem telecomunicações. Foi a partir da expansão das redes de alta velocidade, entretanto, que a popularização de vídeos digitais tomou força. Atualmente, os vídeos digitais podem ser reproduzidos em diversos tipos de dispositivos eletrônicos para fins de comunicação, entretenimento, educação, entre outros. Uma estimativa da CISCO indica que até 2019 cerca de 82% do tráfego de dados na internet serão relacionados a conteúdo de vídeo digital [CISCO 2017].

Entretanto, devido ao enorme volume de dados presentes em um vídeo digital, a sua compressão é mandatória para que tais dispositivos possam manipulá-los, transmiti-los em tempo real e armazená-los, quando necessário. Esta necessidade por altas taxas de compressão torna-se ainda mais latente nos últimos anos, com o aumento da resolução espacial e temporal dos vídeos em busca de uma melhor experiência de visualização do usuário. Com foco neste problema, tanto a indústria quanto a academia vêm realizando amplas investigações nas últimas décadas em busca de técnicas mais avançadas e eficientes de compressão de vídeo. O *High Efficiency Video Coding* (H.265/HEVC) [Kim 2013], definido pelo *Joint Collaborative Team on Video Technology* (JCT-VC) e lançado em 2013, tornou-se recentemente o estado-da-arte em compressão de vídeo e vem gradualmente substituindo o seu predecessor, o padrão H.264/AVC [Wiegand 2003], em aplicações comerciais e em dispositivos multimídia. Em comparação com o H.264/AVC, o HEVC provê taxas de compressão entre 40% e 50% mais elevadas, mas apresenta complexidade até 500% mais alta [Correa 2012]. Torna-se, portanto, cada vez mais necessário o estudo e o desenvolvimento de estratégias que permitam reduzir a complexidade da codificação de vídeo, especialmente para utilização em dispositivos móveis com recursos energéticos limitados e em aplicações de tempo real que necessitem de codificação e transcodificação rápida compatível com as taxas de captura e de transmissão de dados.

Diversos trabalhos que focam na redução de complexidade do codificador HEVC e seus predecessores mais recentes podem ser encontrados na literatura. A maior parte baseia-se em heurísticas resultantes de análises estatísticas sobre características dos vídeos e variáveis intermediárias da codificação, eliminando etapas custosas que não necessariamente resultam em ganhos expressivos em eficiência de codificação. Entretanto, muitas dessas técnicas levam a perdas significativas, de forma que abordagens mais inteligentes baseadas em análises mais profundas e metódicas sobre os dados são necessárias. Os recentes avanços na área de Aprendizado de Máquina permitiram uma vasta aplicação de modelos inteligentes para solucionar problemas difíceis de resolver com estruturas de dados e algoritmos tradicionais. Especificamente na pesquisa de codificação de vídeo, trabalhos recentes mostram que modelos baseados em Aprendizado de Máquina podem auxiliar na redução do espaço de otimização relativo à decisão do modo dos codificadores, diminuindo significativamente o esforço computacional desse processo e, ao mesmo tempo, mantendo a eficiência de codificação elevada. Além disso, o uso de Aprendizado de Máquina abre espaço para uma grande

variedade de técnicas e algoritmos que ainda não foram aplicados no contexto de codificação de vídeo, ampliando o espectro da pesquisa significativamente.

Este capítulo apresenta uma visão geral sobre a utilização de técnicas de Mineração de Dados e Aprendizado de Máquina na redução de complexidade em codificadores de vídeo, mais especificamente com foco no padrão HEVC. A seção 9.2 apresenta um estudo sobre as principais fontes de complexidade nos codificadores HEVC. A seção 9.3 apresenta e discute brevemente os principais trabalhos publicados na literatura com foco na aplicação de técnicas de Aprendizado de Máquina no processo de codificação de vídeo. Os trabalhos mais recentes na área publicados pelos autores e seus grupos de pesquisa são revisitados na seção 9.4. Por fim, a seção 9.5 apresentará tendências e prospecções de trabalhos futuros, os quais envolvem a utilização de outros modelos, algoritmos e metodologias de implementação e aplicação de Aprendizado de Máquina para solucionar o problema da complexidade de codificação de vídeo com maior eficiência e a utilização de técnicas similares nas gerações futuras de codificadores de vídeo.

9.2. Codificação de Vídeo HEVC

Assim como seus antecessores, o codificador HEVC é baseado em um esquema híbrido composto principalmente pelas etapas de estimação/compensação de movimento e de transformadas. Entretanto, o HEVC introduziu uma série de novas ferramentas mais avançadas de processamento de sinais para atingir taxas de compressão elevadas sem causar efeitos indesejáveis na qualidade da imagem. Como esperado, o aumento na eficiência de codificação foi obtido às custas de um aumento significativo na complexidade do processo de codificação, principalmente resultante de estruturas de particionamento muito mais complexas e algoritmos de otimização exaustivos [Correa 2012].

9.2.1. Estruturas de Particionamento

Uma das funcionalidades mais importantes e mais complexas introduzidas no padrão HEVC consiste nas estruturas de particionamento de quadros baseadas em árvores quadráticas [Sullivan 2012]. Inicialmente, cada quadro de um vídeo é dividido em uma série de blocos quadrados de tamanhos iguais chamados *Coding Tree Units* (CTUs), os quais são utilizados como raízes de árvores quadráticas de particionamento chamadas *Coding Trees*. As *Coding Trees* obedecem uma estrutura de particionamento recursiva, na qual cada nó, chamado de *Coding Unit* (CU), pode ser subparticionado recursivamente em quatro novos nós até que a menor dimensão possível de CU seja atingida. Normalmente, codificadores HEVC permitem CUs de tamanho 8×8 , 16×16 , 32×32 e 64×64 . A parte inferior da Figura 9.1 apresenta uma CTU dividida recursivamente em uma série de CUs. As folhas da *Coding Tree* (nós em azul) representam a decisão final de particionamento daquela CTU em CUs.

Cada CU pode ser dividida em duas ou quatro *Prediction Units* (PUs), que são os blocos básicos utilizados nas etapas de predição inter-quadros e intra-quadro. A parte superior da Figura 9.1 apresenta todos os particionamentos possíveis de uma CU em PUs. As possibilidades são chamadas de modos de particionamento de PU e a quantidade de modos disponíveis para utilização varia de acordo com o tamanho da CU em questão. As CUs no nível mais profundo das *Coding Trees* admitem apenas PUs quadradas e retangulares simétricas (chamados de modos $2N \times 2N$, $2N \times N$, $N \times 2N$ e $N \times N$),

enquanto que as CUs nos níveis superiores admitem todos os modos de particionamento de PU. Além disso, a predição intra-quadro só pode ser realizada sobre PUs quadradas, enquanto que a predição inter-quadros é realizada sobre PUs quadradas e retangulares.

Finalmente, as últimas estruturas de particionamento são as *Transform Units* (TUs), que também são organizadas em árvores quadráticas (assim como as CUs) chamadas *Residual Quadrees* (RQTs). As TUs são as estruturas de particionamento básicas de processamento nas etapas de transformada e quantização do codificador HEVC. Assim como as CUs, as TUs são particionadas recursivamente em TUs quadradas menores, até que o tamanho mínimo de TU seja atingido. O tamanho máximo e mínimo de TU são definidos na configuração do codificador, mas o maior e menor tamanhos permitidos no padrão HEVC são 32×32 e 4×4 , respectivamente.

Além das estruturas de particionamento, diversas outras ferramentas de alta complexidade foram adicionadas ao padrão HEVC ou são modificações em relação às ferramentas existentes no padrão H.264/AVC. A predição intra-quadro, por exemplo, que no padrão H.264/AVC era composta por 9 modos de predição, passou a ser composta no padrão HEVC por 35 modos [Lainema 2012]. O filtro redutor de efeito de bloco no padrão H.264/AVC foi complementado por um novo filtro, o *Sample Adaptive Offset* (SAO) [Fu 2012], no padrão HEVC. A predição inter-quadros foi complementada pelo modo *Merge* e pelo *Motion Vector Prediction* [Helle 2012]. Novos tamanhos de transformadas foram incorporados ao padrão para suportar as TUs de tamanhos variados. Entretanto, apesar de todas as ferramentas adicionadas ao padrão HEVC, nenhuma delas é tão impactante em termos de eficiência de codificação e de complexidade quanto as estruturas de particionamento, que poderiam ser classificadas um nível acima das ferramentas na hierarquia de processamento. Note que todas as novas ferramentas são executadas repetidamente ao testar cada uma das possibilidades de divisão das CUs, de forma que cada novo teste de particionamento incorre em uma grande complexidade associada. Por esta razão, a maioria das estratégias que visam reduzir a complexidade dos codificadores HEVC focam, de alguma forma, na simplificação do processo de decisão das estruturas de particionamento [Correa 2011].

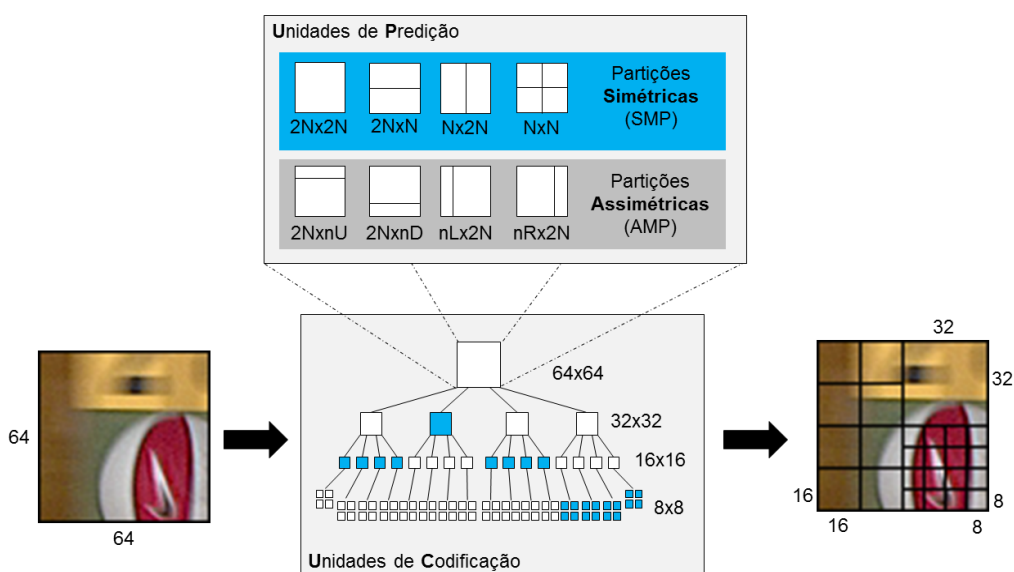


Figura 9.1. Exemplo de particionamento de CTU em CUs e de CUs em PUs simétricas e assimétricas.

9.2.2. Decisão por Otimização Taxa-Distorção

A escolha do melhor particionamento de CTU em CUs, do melhor modo de particionamento de cada CU em PUs e do melhor particionamento de cada CU em TUs é uma tarefa complexa e essencial para o bom desempenho do codificador HEVC. No HEVC e nos codificadores que o precedem, a escolha ótima dos modos é realizada através do cálculo do custo taxa-distorção, que consiste em uma função que especifica um custo em termos da taxa de bits e perda de qualidade resultante da codificação de um bloco da imagem de acordo com um modo [Sullivan 1998]. Após serem calculados os custos taxa-distorção para todos os modos possíveis, o codificador escolhe aquele com menor custo. Note que a complexidade desta decisão é multiplicativa, já que todas as possibilidades de particionamento de PUs (e de TUs) são testadas dentro de cada possibilidade de particionamento de CUs.

A tarefa de encontrar o melhor modo, conhecido como Otimização Taxa-Distorção (*Rate-Distortion Optimization* – RDO) [Sullivan 1998], é geralmente vista como um problema de otimização que pode ser resolvido através do método de Lagrange, no qual os termos de distorção e de taxa de bits são ponderados com um termo λ . A minimização lagrangiana é representada na Equação 1, onde $J(i)$ é o custo taxa-distorção, $D(i)$ é a distorção e $R(i)$ é a taxa de bits resultante da codificação de um bloco com um determinado modo i . O termo λ é o multiplicador de Lagrange, que permite aumentar ou reduzir a influência da taxa de bits no custo taxa-distorção. O melhor modo i^* é, portanto, definido como menor custo taxa-distorção J encontrado entre todas as S possibilidades testadas.

$$i^* = \underset{i \in S}{\operatorname{arg\,min}} J(i) , \quad (\text{Eq. 1})$$

onde $J(i) = D(i) + \lambda \cdot R(i)$

Além das decisões de particionamento em CUs, PUs e TUs, o processo de codificação de vídeo HEVC envolve diversas outras decisões, como a decisão do melhor modo de predição intra-quadro, a decisão entre modo intra-quadro e modo inter-quadro, e a decisão do melhor vetor de movimento durante a predição inter-quadros, conforme mencionado na seção anterior. A técnica RDO permite que as melhores taxas de compressão sejam atingidas às custas das menores perdas de qualidade de imagem, dado um conjunto específico de possibilidades testadas. Entretanto, se cada combinação de modos de predição e particionamento forem avaliados por busca exaustiva, a complexidade do processo de codificação se torna proibitiva, especialmente se a codificação em tempo real for o objetivo final. Na prática, os codificadores implementam ou permitem a utilização de apenas um subconjunto de modos, de forma a reduzir a complexidade do processo de codificação para níveis aceitáveis. Entretanto, a seleção de tal subconjunto não é uma tarefa trivial e é essencial para o bom desempenho dos codificadores de vídeo.

9.2.3. Complexidade da Codificação e da Transcodificação HEVC

Em [Correa 2012], os autores apresentam uma análise do padrão HEVC que tem o objetivo de identificar as ferramentas que mais afetam a eficiência de compressão e a complexidade do codificador *HEVC Model* (HM) como um todo. Um conjunto de 16 configurações diferentes foram criadas para testar o impacto de cada ferramenta em comparação com o mesmo codificador na sua configuração mais básica. Os resultados em [Correa 2012] revelam que o HEVC apresenta uma complexidade entre 10% e

500% maior que a do seu antecessor, o codificador H.264/AVC, dependendo da configuração utilizada. O trabalho [Correa 2015] apresenta uma análise similar, entretanto coloca um foco maior na complexidade introduzida pelas novas estruturas de particionamento do HEVC. A análise conclui que a habilitação de todas as possibilidades de CU aumenta a complexidade de codificação em 74,1%, enquanto que a habilitação de todas as possibilidades de TU aumenta a complexidade em 15,3%, em comparação com uma configuração básica. A habilitação dos modos de particionamento assimétricos de PU representa um aumento de 8% na complexidade do codificador.

O problema da alta complexidade também é crítico em transcodificadores de vídeo. Os transcodificadores de vídeo são, na sua forma mais trivial, compostos por um decodificador conectado em sequência a um codificador de vídeo, com o intuito de realizar alguma modificação no formato do bitstream original. A transcodificação heterogênea H.264/AVC-para-HEVC, por exemplo, consiste em um decodificador H.264/AVC conectado a um codificador HEVC. Este tipo de transcodificação é extremamente importante em períodos de transição entre padrões, já que a adoção completa de um novo padrão pode levar vários anos devido à enorme quantidade de conteúdo legado do seu antecessor. Por esta razão, há uma demanda real para conversão de bitstreams H.264/AVC em conteúdo compatível com reprodutores HEVC, permitindo interoperabilidade entre codificadores e decodificadores que seguem os dois padrões. A transcodificação homogênea para adaptação de taxa de bits ou de resolução, composta por um decodificador e um codificador que seguem o mesmo padrão conectados em sequência, também é extremamente importante do ponto de vista de adaptabilidade do conteúdo para o receptor. Em sistemas de transmissão de vídeo em fluxo contínuo adaptativo, a transcodificação homogênea para adaptação de taxa de bits é essencial, já que possibilita ao servidor de conteúdo realizar a adaptação do vídeo às características do reprodutor no cliente ou às características da rede na qual o cliente está conectado. Obviamente, por ser composto por um decodificador e um codificador de vídeo em sequência, um transcodificador de vídeo na sua implementação mais trivial, seja ele homogêneo ou heterogêneo, apresenta uma complexidade proibitiva em sistemas que requerem processamento em tempo de reprodução do vídeo.

9.3. Aprendizado de Máquina Aplicado à Codificação de Vídeo

As seções anteriores mostraram que a codificação de vídeo é um processo de alta complexidade que envolve uma série de métodos e técnicas para ser realizado de forma eficiente. Recentemente, com a popularização da área de Inteligência Artificial e o aumento do poder computacional para processamento de grandes quantidades de dados, diversas abordagens surgiram com o objetivo de resolver problemas derivados da codificação de vídeo com o auxílio de técnicas de Aprendizado de Máquina. As seções a seguir apresentam as principais contribuições encontradas na literatura, separadas por tópico.

9.3.1. Codificação de Alta Qualidade

Alguns trabalhos desenvolvidos têm por objetivo propor modelos capazes de codificar vídeos com melhor qualidade de imagem, ou ainda com menor taxa de bits para a mesma qualidade. As soluções desenvolvidas baseiam-se na substituição de ferramentas, como filtros de pós-processamento e interpoladores sub-pixel, por modelos preditivos capazes de gerar amostras de melhor qualidade. Os pesos dos filtros de

interpolação e de pós-processamento utilizados no codificador HEVC são desenvolvidos com base em análise estatística, portanto modelos preditivos podem ser mais eficazes. Para medir os ganhos em qualidade, os autores normalmente utilizam métricas objetivas como *Peak Signal to Noise Ratio* (PSNR), *Structural Similarity* (SSIM), ou ainda uma métrica amplamente utilizada na comunidade de codificação de vídeo chamada de *Bjontegaard Delta*, a qual pode ser medida com base na taxa de bits (BD-rate) ou no PSNR (BD-PSNR) [Bjontegaard 2003].

O trabalho apresentado em [Chauvet 2016] descreve um filtro de pós-processamento adaptativo baseado em Redes Neurais para o HEVC. A técnica proposta utiliza dados da codificação (como modo de predição, presença de borda e presença de resíduos significativos) para ajustar os parâmetros do filtro, o qual é implementado com uma Rede Neural Convolutiva (*Convolutional Neural Network* - CNN) de duas camadas treinada offline. Os resultados apresentados mostram que o filtro proposto traz um ganho médio de 0,1 dB quando aplicado em conjunto com os filtros do HEVC.

Uma estratégia similar é encontrada em [Park 2016], em que uma CNN é treinada para servir como filtro de pós-processamento, chamada *In-Loop Filtering CNN* (IFCNN). Este trabalho baseia-se na metodologia de [Dong 2014] para gerar um filtro capaz de reduzir os artefatos produzidos pela quantização, substituindo o filtro SAO. Os autores observaram que o parâmetro de quantização (QP) utilizado nas amostras de treino influi diretamente na performance dos modelos, portanto uma versão foi gerada para QPs entre 20 e 29, e outra para valores entre 30 e 39. Os resultados experimentais apontam que a IFCNN desenvolvida supera a implementação padrão do codificador HEVC com reduções entre 1,6% e 2,8% na taxa de bits.

Outra solução de filtro de pós-processamento para substituir o SAO com base em CNN é apresentada em [Dai 2017], chamada *Variable-filter-size Residue-learning CNN* (VRCNN). Essa solução se baseia no trabalho de Dong (2015), o qual propõe uma CNN utilizada para redução de artefatos após a compressão JPEG, estendendo as técnicas para suportar as transformadas de tamanho variável do HEVC. Além disso, os autores aplicam um treinamento baseado em otimização do resíduo entre a entrada e a saída da rede, acelerando a convergência durante o treino. Comparada com a implementação HEVC de referência, a solução proposta apresenta uma redução de 4,6% na taxa de bits.

O trabalho de Wang (2017) também apresenta um filtro baseado em CNN para o decodificador HEVC, porém aplicado juntamente com o SAO ao invés de servir como alternativa. Os autores utilizam a noção de *compressive sensing* para fundamentar sua proposta, segundo a qual é possível aumentar a qualidade de um vídeo decodificado através de alguns algoritmos de restauração. Segundo os autores, a parte mais interessante desta abordagem é que esta não exige modificações no codificador. Para implementar o filtro, os autores utilizam uma rede chamada *Deep CNN-based Auto Decoder* (DCAD) com 10 camadas de profundidade e *Rectifier Linear Units* (ReLU) em cada nodo. Os resultados obtidos apontam reduções médias entre 5% e 6,4% na taxa de bits, dependendo da configuração do codificador.

Em [Ma 2016], um método para adaptação dinâmica do multiplicador de Lagrange no modo de decisão do HEVC é proposto, com o intuito de gerar vídeos comprimidos com maior qualidade subjetiva. Os autores treinam modelos baseados em *Support Vector Machines* (SVM), utilizando descritores como energia temporal e

espacial para estimar o fator de ajuste mais apropriado para o multiplicador de Lagrange. Os resultados apresentados mostram um ganho médio de 0,5 em SSIM em 2 dB em PSNR. Todavia, os resultados de bitrate não foram apresentados, portanto não é possível fazer uma análise mais profunda do impacto no tradeoff compressão-qualidade.

Um modelo adaptativo para as transformadas é proposto em [Puri 2016]. Inicialmente, os blocos residuais são classificados de acordo com seu custo de codificação, e as transformadas mais adequadas para cada grupo são geradas usando um processo iterativo baseado em minimização da norma l_0 . O critério de seleção se baseia na estimativa do custo necessário para transmitir os novos parâmetros de cada transformada somado ao custo dos coeficientes quantizados. Duas abordagens foram desenvolvidas: a primeira é chamada *Adaptive Directional Transforms* (ADT) e a segunda é uma derivação que acrescenta um parâmetro de tolerância para permitir uma exploração maior do espaço de soluções, chamada *Annealing-Based ADT* (AADT). Os resultados apresentados mostram uma redução na taxa de bits de 1,4% para a configuração *Random Access* e 2,1% para a configuração *All Intra* do codificador HEVC.

A solução proposta por Pan (2017) dedica-se a manter um nível de qualidade estável em aplicações de controle de taxa de bits. Os autores defendem que a estrutura de codificação baseada em QPs diferentes entre quadros (utilizada na codificação HEVC) leva a flutuações na qualidade visual, especialmente quando vídeos de conteúdos diferentes são combinados. Para contornar esse problema, os autores desenvolveram um sistema de controle de taxa que estima o valor de QP necessário para cada quadro através de um modelo de regressão baseado em SVM. Além disso, os autores aplicam um filtro de Kalman para ajustar a taxa alvo dinamicamente. Dessa forma, é possível manter a qualidade estável mesmo quando a diferença entre a taxa de bits alvo e a taxa de bits real é alta. Os resultados de qualidade apontam ganho médio de 0,04 dB com acréscimo de 31,8% no tempo de codificação para a configuração *Low Delay B* e perda de 0,28 dB com acréscimo de 39,85% no tempo para a configuração *Random Access*. No entanto, os autores defendem que a solução proposta é a que possui maior estabilidade entre as demais.

Em [Yan 2017], um filtro de interpolação para pixels fracionários com base em uma CNN é proposto. A primeira dificuldade encontrada pelos autores foi definir quais seriam os valores de referência (*ground truth*) utilizados para treinar os modelos, visto que as amostras fracionárias são geradas sinteticamente. A solução encontrada foi aplicar um filtro de suavização (*blurring*), gerando amostras também sintéticas, mas de alta correlação com as amostras reais. Após a geração dos dados de treino, três CNNs foram geradas, uma para cada tipo de amostra fracionária (horizontal, vertical e diagonal). Com essa abordagem, uma redução de 0,9% na taxa de bits foi obtida. Neste trabalho somente os pixels fracionários de precisão $\frac{1}{2}$ (*half pixels*) são gerados, mas os autores têm intenção de estender a metodologia para precisão $\frac{1}{4}$ em trabalhos futuros.

9.3.2. Detecção de Regiões de Interesse

Outro problema comum em sistemas de codificação consiste em detectar quais regiões do quadro devem ser codificadas com maior esforço, chamadas regiões de interesse (*Region of Interest* - ROI). Regiões de movimento intenso, por exemplo, normalmente requerem um particionamento mais detalhado em comparação a regiões estáticas. Sistemas de controle de taxa de bits e de complexidade, por exemplo, geralmente

utilizam essa informação para alocar recursos de acordo com a importância de cada bloco. No entanto, realizar essa classificação em tempo de codificação não é uma tarefa trivial, de forma que alguns pesquisadores dedicam-se a desenvolver algoritmos eficientes para essa tarefa. As soluções que utilizam Aprendizado de Máquina para esse fim são baseadas no conceito de saliência visual, ou seja, os modelos desenvolvidos são capazes de detectar as regiões em que o usuário (visualizador) vai focar sua atenção.

O trabalho de Judd (2009) apresenta um extenso estudo das regiões de interesse em imagens e serviu de base para outros trabalhos voltados para aplicações de vídeo. Os autores trazem duas contribuições importantes: a primeira consiste em uma base de dados contendo mais de 1000 imagens, juntamente com os dados de saliência visual (obtidos através de um dispositivo de rastreamento visual, ou *eye tracking*) e a segunda é um modelo SVM para computar a saliência visual de imagens com base em um extenso conjunto de atributos. Os atributos utilizados incluem tanto descritores de alto nível, como detecção de faces e pessoas, quanto de baixo nível, especificamente medidas de contraste, orientação, intensidade luminosa, entre outras. Os resultados apresentados pelos autores mostram que utilizar descritores de alto, médio e baixo nível permite gerar modelos mais eficientes, comparados a outros trabalhos que focam em um único nível, atingindo um índice de acertos acima de 85%.

Souly (2016) apresenta um método não supervisionado para geração de modelos de detecção de saliência visual em vídeos. A técnica dos autores consiste em dividir os vídeos em cubóides (a profundidade corresponde a quadros sucessivos), os quais são vetorizados em uma matriz cujas colunas correspondem à intensidade dos pixels. Além disso, o vídeo é segmentado utilizando um algoritmo hierárquico baseado em grafos proposto por Grundmann (2010). Essa segmentação serve para agrupar a informação entre regiões similares no espaço e no tempo (chamadas super-voxels), de forma a diminuir representações redundantes. Finalmente, os valores resultantes do dicionário (simplificados pela segmentação) são decompostos em uma matriz de saliência (S) e uma matriz de baixa saliência (L) utilizando a técnica *Robust Principal Component Analysis* (RPCA). O mapa de saliência resulta da soma entre os valores normalizados da matriz S com os valores de saliência dos super-voxels obtidos na segmentação. Os resultados apresentados mostram algum ganho com relação a outros trabalhos similares, com um índice AUC (*Area Under ROC Curve*) de 0,71.

O trabalho de [Xu 2017] propõe um modelo para detecção de saliência através de atributos extraídos da codificação HEVC. Os autores criaram sua própria base de dados de *eye tracking*. Durante a análise estatística dos dados gerados, os autores perceberam que os visualizadores prestam mais atenção em regiões que são mais particionadas durante a codificação, assim como aquelas que possuem maior alocação de bits e maiores vetores de movimento. Além disso, três outras descobertas são descritas no artigo: (1) a fixação da atenção em um objeto novo na cena ou que começa a se mover demora alguns microssegundos; (2) a atenção tende a se concentrar em novos objetos que aparecem na cena; (3) objetos que se movem na direção contrária dos demais à sua volta tendem a atrair mais atenção. Os autores criaram um conjunto de descritores derivados dessas observações e treinaram um modelo de classificação com base em SVMs. Os resultados apontam um índice AUC de 0,823, superando outros trabalhos existentes na literatura.

Em [Zhang 2017], um esquema de controle de taxa de bits para codificação HEVC que prioriza as ROI é proposto. Inicialmente, os autores utilizam regressão linear para gerar um modelo de classificação de CUs em ROI ou não-ROI, utilizando como base a informação dos coeficientes da DCT. Em seguida, os autores treinam uma Rede Neural para estimar a distorção de uma CU de acordo com o valor do parâmetro de quantização. Para facilitar a modelagem das redes, os modelos foram separados entre CUs dentro e fora da ROI. As rede neurais utilizadas contêm uma função de ativação de base radial (*Radial Basis Function* - RBF). Os autores justificam o uso dessa função demonstrando que a relação entre distorção e a quantização não é linear e que a função RBF é mais indicada para essa aplicação. Em seguida, os autores desenvolvem um novo modelo para computar o custo taxa-distorção, que é aplicado de acordo com a profundidade da CU e o grupo de ROI. Esse modelo é então utilizado em um sistema de controle de taxa em vários níveis (GOP, quadro, CU). Os resultados demonstram que a solução proposta supera o algoritmo de controle de taxa de bits utilizado no software de referência do codificador HEVC, o *HEVC Test Model* (HM) [Kim 2013], em termos de qualidade de imagem e precisão de controle.

9.3.3. Estimativa de Métricas de Qualidade

A qualidade do vídeo codificado é normalmente mensurada durante a codificação através de métricas objetivas (como PSNR e SSIM), porém a correlação dessas medidas com a qualidade subjetiva observada pelos usuários não é sempre direta [Liu 2013]. A qualidade subjetiva é comumente aferida através do *Mean Opinion Score* (MOS), mas a metodologia aplicada para obter o valor de MOS envolve uma base de usuários aptos a visualizar as diferentes versões do vídeo para posterior avaliação. Isso se torna inviável para a maioria dos casos, nos quais é necessário estimar a qualidade durante a codificação ou a transmissão do vídeo. Com base nisso, algumas métricas altamente correlacionadas com o índice MOS foram propostas na literatura, dentre as quais podemos destacar a *Video Quality Model with Variable Frame Delay* (VQM-VFD) [Pinson 2014]. Os trabalhos apresentados nesta seção apresentam modelos de Aprendizado de Máquina para cálculo de qualidade.

Os pesquisadores da Netflix desenvolveram recentemente um modelo SVM para estimativa da qualidade subjetiva de vídeos [Li 2016]. O modelo é treinado com base em métricas recentes de trabalhos relacionados da literatura: a *Visual Information Fidelity* (VIF) de [Sheikh 2006], baseada na perda de fidelidade entre o sinal original e o sinal distorcido (por exemplo, o vídeo reconstruído após uma codificação com perdas), e a *Detail Loss Metric* (DLM) de [Li 2011], uma métrica de qualidade de imagem baseada em medir separadamente a perda de detalhes, que afeta a visibilidade do conteúdo, e a perda obtida através de conteúdo redundante na imagem, o qual distrai a atenção do usuário para regiões menos importantes da cena. Além disso, a diferença média de pixels co-localizados em quadros adjacentes também é computada para representar as características de movimento. Os resultados apresentados no relatório mostram que a técnica proposta supera o VQM-VFD, modelo considerado por muitos como referência atualmente, atingindo um coeficiente de correlação de 0,963 com métricas subjetivas. O resultado é superior ao atingido pelo VQM-VFD (0,934).

Vega (2017) apresenta uma abordagem para avaliação de qualidade em ambientes de streaming de vídeo com base em Redes Neurais Profundas (*Deep Neural Networks* - DNN). Processos de aprendizagem profunda utilizando *Restricted*

Boltzmann Machines, por serem mais custosos, são aplicados offline pelo servidor de conteúdo. Os modelos são gerados por métodos de aprendizado não supervisionado, o que torna o modelo escalável e facilmente adaptável a novos conteúdos (visto que não é preciso rotular novos dados). O cliente é responsável pela extração dos valores utilizados como atributos para treino. Esses valores consistem em métricas do tipo *No-Reference*, que não necessitam do vídeo original para serem computadas. Dentre os atributos utilizados estão informação de movimento, taxa de ruído e nível de efeito de bloco. Os resultados apontam uma correlação entre 78% e 91% com a métrica VQM.

9.4. Redução de Complexidade HEVC Baseada em Aprendizado de Máquina

Esta seção apresenta os principais trabalhos desenvolvidos pelos autores e seus grupos de pesquisa nos últimos anos com foco na redução de complexidade dos codificadores e transcódificadores de vídeo HEVC utilizando técnicas de mineração de dados e aprendizado de máquina.

Técnicas de mineração de dados preditiva são utilizadas para determinar o valor de variáveis chamadas dependentes através da observação do valor de alguns atributos no conjunto de dados, identificando regularidades e construindo regras de generalização que são expressas como modelos. Existem diversos métodos de mineração de dados preditiva, as quais variam entre si em termos de eficiência, complexidade e aplicabilidade. As árvores de decisão [Quinlan 1993] são modelos frequentemente utilizados devido às seguintes características: (i) geralmente atingem altos níveis de precisão depois de treinadas, (ii) são de fácil compreensão para humanos e, portanto, facilmente implementadas, (iii) existem muitos algoritmos eficientes para construí-las, (iv) podem lidar tanto com valores categóricos quanto numéricos e (v) uma vez implementadas, executam muito rapidamente.

A primeira e a quinta características listadas acima são essenciais em soluções que buscam resolver ou reduzir o problema da alta complexidade da codificação de vídeo, já que tanto a precisão das decisões quanto a rapidez com que estas são tomadas são extremamente importantes para o funcionamento dos codificadores. Por esta razão, a maioria dos trabalhos apresentados nesta seção tem, como estratégia principal, a utilização de algoritmos para construção de árvores de decisão que são capazes de substituir o processo exaustivo de decisão baseado em RDO ou pelo menos simplificar este processo, adicionando interrupções prematuras nos testes realizados.

9.4.1. Decisões Rápidas de Particionamento Utilizando Árvores de Decisão

Os autores em [Correa 2015] propõem a utilização de um esquema composto por três conjuntos de árvores de decisão que permitem decidir se o algoritmo RDO deve ser terminado prematuramente ou executado até o final para a escolha dos melhores particionamentos de CU, PU e TU. Assim, através da utilização das árvores, a decisão de particionamento da estrutura em questão (seja ela CU, PU ou TU) passa a ser vista como um problema de classificação binária, onde as duas classes possíveis correspondem a *Continuar* e *Terminar* o algoritmo RDO.

A partir de dados intermediários extraídos do processo de codificação e de atributos dos vídeos, três conjuntos de árvores de decisão foram construídos e implementados no codificador de referência do padrão HEVC, o *HEVC Model* (HM)

[Kim 2013]. Todos os dados para a construção das árvores de decisão foram coletados *offline*, ou seja, em uma etapa de codificações prévias à mineração de dados e à construção dos modelos. Mais de 40 atributos para cada CU, PU e TU foram coletados neste processo, incluindo a classificação final para aquela estrutura de particionamento (isto é, *Continuar* ou *Terminar* o processo de decisão). São exemplos de atributos os custos taxa-distorção de cada modo de particionamento de PU, a profundidade máxima e mínima das CUs vizinhas à CU atual, o melhor modo de particionamento das PUs vizinhas à PU atual, a variância e o gradiente das amostras que compõem um bloco, entre outros [Correa 2015].

A ferramenta *Waikato Environment for Knowledge Analysis* (WEKA) [Hall 2009] foi utilizada para auxiliar no processo de mineração dos dados coletados e o codificador HM original (isto é, sem qualquer modificação) foi utilizado para codificar um conjunto de 10 sequências de vídeo definidas nas Condições Comuns de Teste (CCT) [Bossen 2011] do JCT-VC. A entrada para a ferramenta WEKA consiste em arquivos *Attribute-Relation File Format* (ARFF), exemplificados na Figura 9.2(a) [Correa 2015]. Os dados coletados durante esta fase foram organizados em arquivos ARFF e utilizados tanto na etapa de mineração de dados quanto na etapa de treinamento das árvores de decisão. No processo de mineração de dados, a utilidade de cada atributo coletado foi avaliado através do método de *Information Gain Attribute Evaluation* (IGAE) [Cover 1991], que mede o ganho de informação de cada atributo em relação à classificação final (*Continuar/Terminar*). Com base no método IGAE, os atributos mais relevantes observados na mineração foram selecionados para a fase de treinamento das árvores de decisão.

O algoritmo C4.5 [Quinlan 1993], que também se baseia no método IGAE para determinar os testes realizados pelas árvores de decisão, foi utilizado para o treinamento de dez árvores de decisão, conforme descrito a seguir. Um exemplo de árvore de decisão obtida é apresentado na Figura 9.2(b) [Correa 2015], onde ‘0’ corresponde a uma decisão de interrupção prematura (classe *Terminar*) e ‘1’ corresponde a uma decisão de continuação do processo de decisão baseado em RDO (classe *Continuar*). Após treinadas as árvores, as suas precisões foram observadas através da aplicação de um processo de validação cruzada com 10 *folds* com a ferramenta WEKA. Em média, as árvores atingiram uma precisão de classificação de 87%.

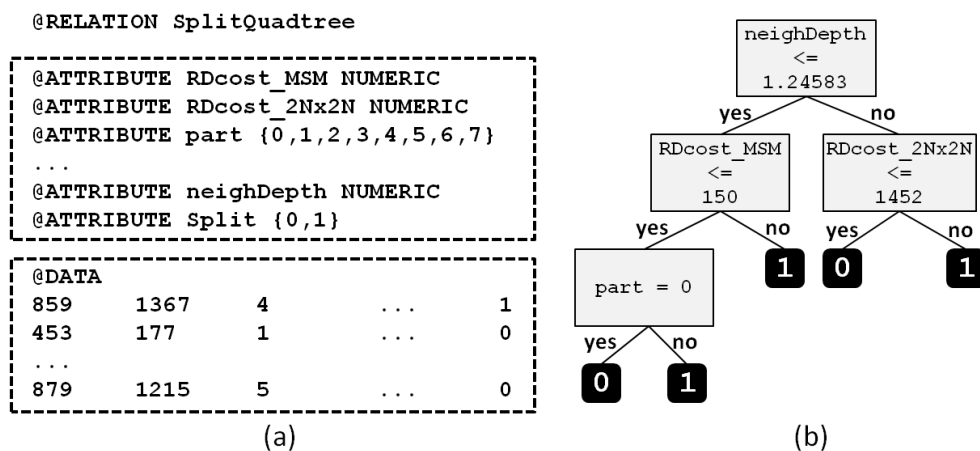


Figura 9.2. Exemplo de um (a) arquivo ARFF e (b) árvore de decisão resultante [Correa 2015].

Dentre as três estruturas de particionamento básicas do HEVC, as CUs têm um papel central devido à sua relação com as demais estruturas. Como discutido na seção 9.2 deste capítulo, todos os modos de particionamento de PUs são testados para cada CU testada na definição da *Coding Tree*. Da mesma forma, todas as possibilidades de TUs são testadas para cada CU considerada. Assim, o primeiro esquema de terminação prematura proposto em [Correa 2015], o *Coding Tree Early Termination* (CTET), consiste em decidir se o particionamento de uma CU em quatro CUs menores deve ser testado. Se o resultado da árvore de decisão for ‘1’, o próximo nível da *Coding Tree* é testado. Caso contrário, a CU atual não é dividida e as suas quatro sub-CUs não são consideradas no processo de decisão baseado em RDO. Três árvores de decisão compõem este primeiro esquema. A primeira árvore de decisão é aplicada à CU 64×64 e, caso a terminação prematura não aconteça, a segunda árvore de decisão é aplicada às quatro CUs 32×32 que a compõem. Por fim, a terceira árvore de decisão é aplicada às quatro CUs 16×16 que compõem todas as CUs 32×32 que não sofreram terminação prematura.

No segundo esquema de terminação prematura proposto em [Correa 2015], o *Prediction Unit Early Termination* (PUET), as árvores de decisão são projetadas para decidir se a busca pelo melhor modo de particionamento em PUs inter-quadros deve continuar após alguns modos serem testados. Uma análise prévia revelou que a maior parte das CUs inter-quadros são previstas como uma única PU, portanto as árvores de decisão do esquema PUET são aplicadas após as previsões com o modo de particionamento $2N \times 2N$ serem testadas. Caso a decisão seja pela terminação prematura, a escolha do melhor modo é realizada conforme o processo RDO entre o modo *Merge-Skip* e o melhor vetor de movimento para a PU $2N \times 2N$. Caso contrário, os modos restantes são todos testados. Quatro árvores de decisão compõem este esquema, correspondentes às CUs 64×64, 32×32, 16×16 e 8×8.

Finalmente, o esquema *RQT Early Termination* (RQTET) consiste em decidir se o particionamento de uma TU em quatro TUs menores deve ser considerado. Se as árvores de decisão resultarem em ‘1’, o próximo nível da RQT deve ser testado. Caso contrário, o processo de particionamento é interrompido para a TU atual. Três árvores de decisão compõem este esquema. A primeira árvore de decisão é aplicada à TU 32×32 e, caso a terminação prematura não aconteça, a segunda árvore de decisão é aplicada às quatro TUs 16×16 que a compõem. Por fim, a terceira árvore de decisão é aplicada às quatro TUs 8×8 que compõem todas as TUs 16×16 que não sofreram terminação prematura.

A avaliação da eficiência dos esquemas foi realizada separadamente, de forma que os conjuntos de árvores foram primeiro implementados separadamente no codificador HM e, depois, implementados em conjunto. Foram utilizadas dez sequências de vídeo, também especificadas nas CTC do JCT-VC [Bossen 2011], mas diferentes daquelas utilizadas na fase de treinamento. Os resultados experimentais apresentados na Tabela 9.1 [Correa 2015] indicam a perda na eficiência de codificação e a redução de complexidade atingida por cada implementação. Os resultados em termos de eficiência de codificação são apresentados com a métrica Bjontegaard-Delta (BD)-rate [Bjontegaard 2001], que indica o aumento da taxa de bits para codificar uma sequência de vídeo com a mesma qualidade de imagem que o codificador HM original. Os resultados em termos de redução de complexidade são apresentados como a diminuição de tempo de execução em relação ao codificador HM original. Quando

implementados separadamente, os esquemas atingiram reduções de complexidade de 7,2% até 50%, com uma perda marginal entre 0,05% e 0,56% na eficiência de codificação. Quando combinados os três esquemas, uma redução média de 65% na complexidade foi atingida, com uma perda de eficiência de codificação de apenas 1,36%.

Os esquemas propostos em [Correa 2015] podem ser ainda ativados dinamicamente para possibilitar níveis distintos de redução de complexidade no codificador. Além disso, podem ser combinados com outras técnicas complementares de redução de complexidade que atuam sobre outras funcionalidades do codificador HEVC, a fim de permitir um maior espectro de complexidade ajustável. Em [Correa 2016], aos autores propõem a utilização dos esquemas em um sistema de controle dinâmico de complexidade do codificador HEVC com 15 pontos de operação, os quais possibilitaram um espectro dinâmico de redução de complexidade entre 10% e 90%.

Tabela 9.1. Resultados experimentais dos métodos de terminação prematura propostos em [Correa 2015].

Esquema Proposto	BD-rate (%)	Redução de Complexidade (%)
CTET	+0.28	37
PUET	+0.56	50
RQTET	+0.05	7.2
CTET + PUET	+1.33	63
PUET + RQTET	+0.7	54
CTET + RQTET	+0.35	43
CTET + PUET + RQTET	+1.36	65

9.4.2. Decisão Rápida de Particionamento Através de Técnicas de Agrupamento

O trabalho apresentado nesta seção faz parte de uma pesquisa em andamento com foco na redução de complexidade do processo de particionamento em codificadores HEVC (semelhante ao discutido na seção anterior). A diferença dessa solução, entretanto, está no fato de empregar uma técnica de Aprendizado por Agrupamento (*Ensemble Learning*) [Dietrich 2000] para aumentar a precisão dos modelos gerados. Nesta técnica, diversas árvores de decisão são treinadas e agrupadas através de um algoritmo conhecido como *boosting*. Vários valores relativos à codificação foram extraídos durante a compressão de vídeos com diferentes características e resoluções. Dentre os valores coletados, destacam-se o número de bits necessários para codificar uma CU, o modo de predição escolhido, o modo de particionamento de PU utilizado, além de valores relativos ao contexto espacial e temporal, como média das profundidades da vizinhança. Além disso, os valores de particionamento foram também coletados para servirem de classes durante o treino. Neste trabalho, as classes correspondem a dois valores: *SKIP*, quando a CU correspondente aos valores extraídos foi particionada durante a decisão de modo, e *SPLIT*, quando a CU foi particionada em blocos menores.

A primeira análise, apresentada na Figura 9.3, mede a relação entre os doze principais atributos coletados e o particionamento da CU. A métrica escolhida para essa análise foi o Ganho de Informação (GI), pois é geralmente utilizada nos algoritmos de treinamento de árvores de decisão. O valor médio de GI para cada atributo é apresentado no gráfico nas caixas preenchidas em cinza.

A partir do gráfico da Figura 9.3, percebe-se que os valores relativos aos custos de taxa e distorção são relevantes para decisão de particionamento (*SKIP* ou *SPLIT*). Além disso, valores relativos ao particionamento das CUs vizinhas no espaço e no tempo (*SplitFlag_{Ctx}*, *SplitFlag_{Coloc}* e *AvgDepth_{Ctx}*) também obtiveram altos índices de GI. Isso ocorre porque muitas vezes existe uma alta semelhança da informação de textura e movimento entre blocos vizinhos. Uma última observação sobre a Figura 9.3 é que a relevância de alguns atributos é diferente em cada tamanho de bloco. Para CUs 64×64, por exemplo, o atributo *AvgDepth_{Ctx}* é o terceiro mais importante, enquanto que, para blocos menores, o atributo ocupa a sétimo (32×32) e a oitava (16×16) posição no ranking.

A mineração de dados apontou que a importância de alguns atributos está ligada ao tamanho do bloco em questão. Além disso, valores como número de bits e distorção também são sensíveis ao tamanho do bloco do codificado. Esses fatos indicam a necessidade de se treinar modelos diferentes para a decisão de particionamento de cada tamanho de CU: 64×64, 32×32 e 16×16. De fato, essa abordagem está alinhada com outros trabalhos da literatura, como [Correa 2015]. Note que CUs 8×8 estão fora do escopo deste trabalho, pois elas correspondem ao menor bloco de codificação, portanto nunca são particionadas.

Para avaliar a eficiência dos atributos extraídos na geração de modelos preditivos, bem como a eficiência das árvores de decisão, uma análise inicial foi realizada com um conjunto reduzido de amostras para treino (mil amostras). A Figura 9.4 mostra a margem de decisão gerada por uma árvore de decisão treinada com dois atributos de alto GI. Considere que *SKIP* corresponde a CUs que não foram particionadas, e *SPLIT* representa aquelas que foram particionadas.

É possível observar que a margem de decisão das árvores de decisão se adapta bem aos dados, mesmo quando eles estão misturados no espaço de amostragem. Isso acontece porque o modelo gerado é baseado em uma série de condicionais que testam o valor dos atributos e os categorizam de acordo com o resultado. A precisão atingida pelo modelo (85,1%) é alta, especialmente quando considerado que apenas dois atributos foram utilizados.

Para treinar as árvores, duas estratégias foram abordadas: a primeira baseou-se no algoritmo padrão de treino disponível no software *C5.0* [Kuhn 2014], enquanto que a segunda utilizou a técnica de *Boosting*, gerando modelos compostos por diversas árvores. Na técnica *Boosting*, cada árvore toma sua decisão e a decisão final é computada pela votação da maioria. A lógica por trás dessa técnica é que a chance de várias árvores tomarem uma decisão errada é menor do que a chance de uma única árvore cometer o erro. A Tabela 9.2 mostra a precisão dos modelos gerados para cada tamanho de CU. O conjunto de treino foi formado com 10 mil amostras extraídas das mesmas sequências de treino utilizadas em [Correa 2015]. Em seguida, a Tabela 9.3 lista as características e detalhes dos modelos com e sem *Boosting*.

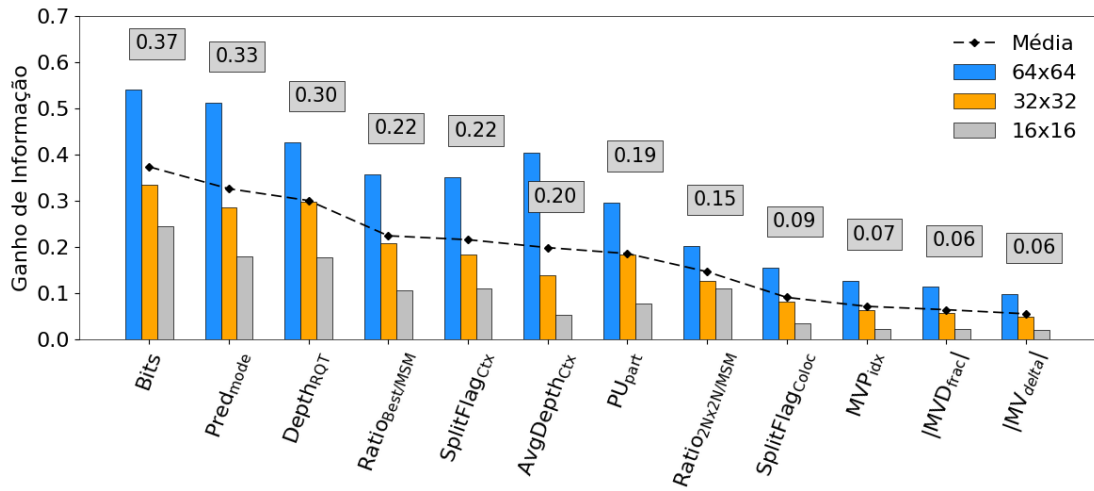


Figura 9.3. Ganhos de Informação de diferentes atributos para os três particionamento de CU.

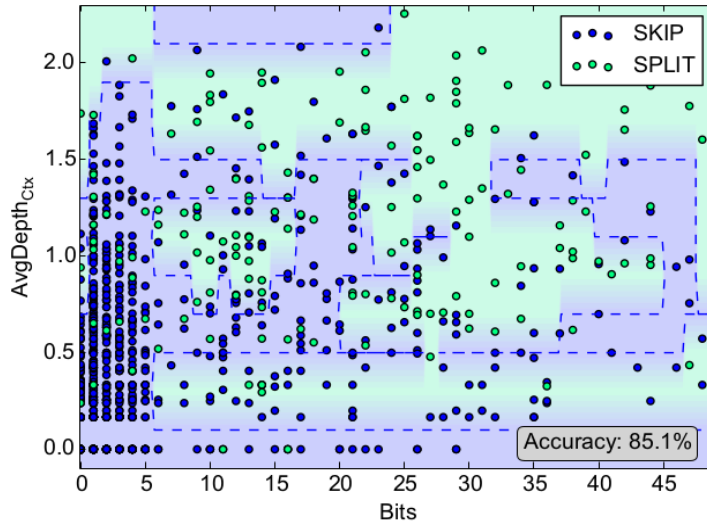


Figura 9.4. Margem de decisão para uma árvore de decisão treinada a partir de dois atributos (CU 64×64).

É possível notar na Tabela 9.2 que a precisão aumenta com a técnica de *Boosting* no treinamento, mas permanece praticamente a mesma quando consideramos o conjunto de teste. Em muitas aplicações, isso seria motivo suficiente para considerar o modelo com *Boosting* ineficiente, mas, de qualquer forma, decidiu-se por prosseguir à análise seguinte, na qual a eficiência das árvores de decisão na codificação é avaliada. Na Tabela 9.3, é possível perceber que o número de nós dos modelos com *Boosting* é diretamente proporcional ao número de árvores utilizadas. Isso seria um problema se o cálculo dos atributos demorasse um tempo relativamente alto do tempo total da aplicação. No entanto, os valores utilizados nos testes de cada nó já são pré-computados, portanto o *overhead* computacional ocorre somente nas operações lógicas utilizadas para percorrer cada árvore.

Tabela 9.2: Resultados de precisão para o conjunto de treino e de teste para cada tamanho de CU utilizando árvores de decisão com e sem *Boosting*.

<i>Sem Boosting</i>			
Tamanho da CU	64×64	32×32	16×16
Precisão Treino	91,20%	89,40%	84,90%
Precisão Teste	86,50%	81,00%	75,30%
<i>Com Boosting</i> (Número de Árvores = 10)			
Tamanho da CU	64×64	32×32	16×16
Precisão Treino	97,40%	98,10%	94,60%
Precisão Teste	87,30%	80,80%	76,10%

Tabela 9.3: Características dos modelos gerados com e sem *Boosting*.

<i>Sem Boosting</i>			
Tamanho da CU	64×64	32×32	16×16
Número de Nós	141	305	304
Profundidade	13	17	11
<i>Com Boosting</i> (Número de Árvores = 10)			
Tamanho da CU	64×64	32×32	16×16
Número Máximo de Nós	277	461	478
Número Médio de Nós	214	395,5	408,2
Profundidade Máxima	26	24	21
Profundidade Média	20,3	18,3	16,2

Tabela 9.4: Resultados de eficiência de compressão (BD-rate), redução de complexidade (TS) e a razão entre os dois valores.

<i>Sem Boosting</i>			<i>Com Boosting</i>		
BD-rate	TS	BDR/TS	BD-rate	TS	BDR/TS
0,73%	43%	1,96	0,22%	39%	0,64

A Tabela 9.4 mostra os resultados da codificação HEVC utilizando os modelos gerados para decisão do particionamento das CUs. Os vídeos foram codificados com QPs 22, 27, 32 e 37, e os valores de BD-rate e redução média de complexidade (*Time*

Savings - TS) foram computados. Os resultados atingidos pelo modelo com uma única árvore já são bastante interessantes, atingindo um incremento de BD-rate inferior a 1% para uma redução de complexidade de 43%. Quando consideramos a versão com *Boosting*, uma melhora significativa na eficiência de compressão pode ser observada. Neste caso, a redução de complexidade foi levemente reduzida, enquanto que a perda de BD-rate diminuiu por um fator igual a 3,3. Esses resultados mostram que técnicas de aprendizado por agrupamento podem gerar modelos muito mais eficientes para a codificação, mesmo quando eles não trazem ganhos significativos de precisão.

9.4.3. Estimação de Movimento Rápida Utilizando Árvores de Decisão

A estimação de movimento é uma das etapas mais complexas da codificação de vídeo, independente do padrão em consideração e das estruturas de particionamento utilizadas, pois envolve a busca, em quadros previamente codificados, do bloco que melhor se assemelha ao bloco atualmente em codificação. Assim como no caso da decisão das melhores estruturas de particionamento, o melhor bloco de referência do ponto de vista de eficiência de codificação só pode ser encontrado através do exaustivo processo RDO sobre todas as possibilidades de localizações candidatas. Entretanto, para tornar a estimação de movimento exequível, as possibilidades de teste devem ser limitadas.

O algoritmo *Test Zone Search* (TZSearch ou TZS) foi empregado no codificador de referência do HEVC, o HM [Kim 2013], desde o início do processo de padronização. O TZS emprega um conjunto de heurísticas para buscar o melhor casamento de bloco em quadros de referência, aproveitando-se da correlação espacial e temporal em áreas vizinhas de um bloco. Quatro estágios compõem o algoritmo TZS: *Motion Vector Prediction*, *First Search*, *Raster Search* e *Refinement*. Em sua execução normal, todas as etapas do TZS são executadas para cada PU. O único estágio que pode ser eventualmente evitado, segundo as regras do TZS, é o estágio de *Raster Search*. Mesmo com todas as heurísticas do TZS, a estimação de movimento é ainda muito complexa, já que todas as etapas do algoritmo precisam ser executadas para todas as possibilidades de particionamento de PU durante o processo de codificação.

Em [Gonçalves 2017], os autores analisam a distribuição dos melhores vetores de movimento (*Motion Vector* - MV) encontrados pelo algoritmo TZS de acordo com a etapa na qual foram encontrados. A Figura 9.5 [Gonçalves 2017] mostra esses resultados para vários tamanhos de PU e permite concluir que, em média, 87% dos melhores MVs determinados pelo TZS são encontrados logo após a primeira etapa do algoritmo (*Motion Vector Prediction*), enquanto que 11% são encontrados após a segunda etapa (*First Search*). Apenas os restantes 2% dos melhores MVs são encontrados nas etapas *Raster Search* e *Refinement*. Por outro lado, Gonçalves (2017) conclui em uma análise de complexidade que estes 2% são responsáveis pela maior parte da complexidade da estimação de movimento com o algoritmo TZS. A Figura 9.6 mostra o tempo de execução de cada uma das etapas. É possível perceber que a etapa *Raster Search* é, de longe, a mais complexa de todas, sendo responsável por 81% do tempo total de execução do algoritmo TZS, em média, apesar de ser responsável por encontrar apenas 0,4% dos melhores MVs. As etapas *Motion Vector Prediction* e *First Search*, apesar de serem responsáveis por encontrar 98% dos melhores MVs, representam uma complexidade de apenas 12% do algoritmo TZS.

A análise permite concluir que, em muitos casos, o custo de processamento de etapas complexas do TZS é desperdiçada, já que a maior parte dos MVs são

encontrados logo nas duas primeiras etapas. Assim, Gonçalves (2017) propõe uma estratégia que permite identificar esses casos e pular as etapas finais, reduzindo significativamente a complexidade do algoritmo sem perdas na eficiência da codificação. Um novo esquema de terminação prematura múltipla para o TZS, chamado de e-TZS, foi proposto em [Gonçalves 2017], construído com base em técnicas de mineração de dados e aprendizado de máquina. O esquema e-TZS busca identificar os estágios do algoritmo TZS nos quais os melhores casamentos de bloco têm maior chance de serem encontrados e, então, pular a execução dos demais estágios dinamicamente. Assim como nos trabalhos previamente apresentados neste capítulo, um processo exaustivo de mineração de dados sobre mais de 40 parâmetros antecedeu a fase de treino de um conjunto de árvores de decisão, cada uma delas dedicada especificamente a um estágio do TZS. Além dos mais de 40 parâmetros, as decisões de *Continuar* ou *Terminar* o algoritmo TZS foram salvas ao fim de cada uma das etapas. Dentre os parâmetros calculados, citam-se os custos taxa-distorção obtidos com o melhor MV encontrado após cada uma das etapas do TZS, o valor de *Sum of Absolute Differences* (SAD) entre a PU atual e o melhor bloco de referência após cada etapa do TZS e as distâncias euclidiana, de *city-block* e *chessboard* [González 2002] entre a PU atual e o melhor bloco de referência após cada etapa do TZS.

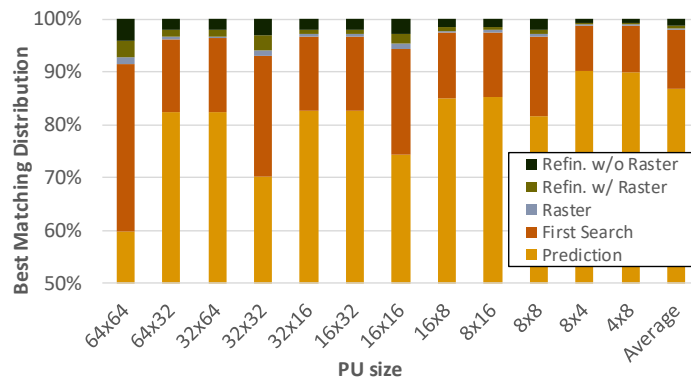


Figura 9.5. Etapa na qual o melhor MV foi encontrado no algoritmo TZS [Gonçalves 2017].

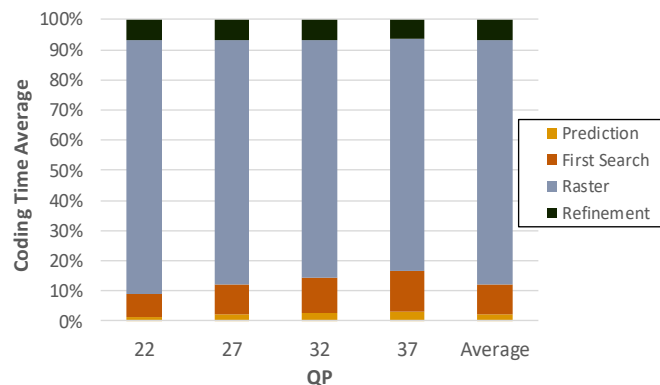


Figura 9.6. Análise de complexidade das etapas do algoritmo TZS [Gonçalves 2017].

Três árvores de decisão foram treinadas com o algoritmo C4.5 [Quinlan 1993] e incorporadas ao algoritmo TZS, implementado no codificador HM [Kim 2013], conforme ilustra o esquema da Figura 9.7 [Gonçalves 2017]. A primeira árvore de decisão do esquema, nomeada *Prediction Early Termination* (Pr-ET), realiza uma série

de testes sobre os atributos coletados em tempo de codificação e decide se o algoritmo TZS deve ser interrompido logo após a etapa de *Motion Vector Prediction*. Caso a árvore decida por *Terminar*, o algoritmo é interrompido. Caso decida por *Continuar*, a etapa de *First Search* é executada. Similarmente, a segunda árvore de decisão, nomeada de *First Search Early Termination* (FS-ET), permite interromper o processo antes da execução da etapa de *Raster Search* ou de *Refinement*, dependendo da decisão interna do algoritmo TZS de executar ou não o *Raster Search*. Por fim, a árvore *Raster Search Early Termination* (RS-ET) permite interromper o algoritmo TZS após o *Raster Search*, evitando a etapa de *Refinement*.

A taxa média de precisão das árvores atingiu 94,2%. Gonçalves (2017) relata que, como esperado, a árvore Pr-ET apresenta menor precisão, pois está localizada numa etapa inicial do algoritmo TZS, na qual poucos atributos da estimação de movimento estão disponíveis para testes. Quando implementado no HM, o esquema e-TZS possibilitou uma redução de 62,53% na complexidade do algoritmo TZS, causando uma pequena perda na eficiência de codificação de apenas 0,49%.

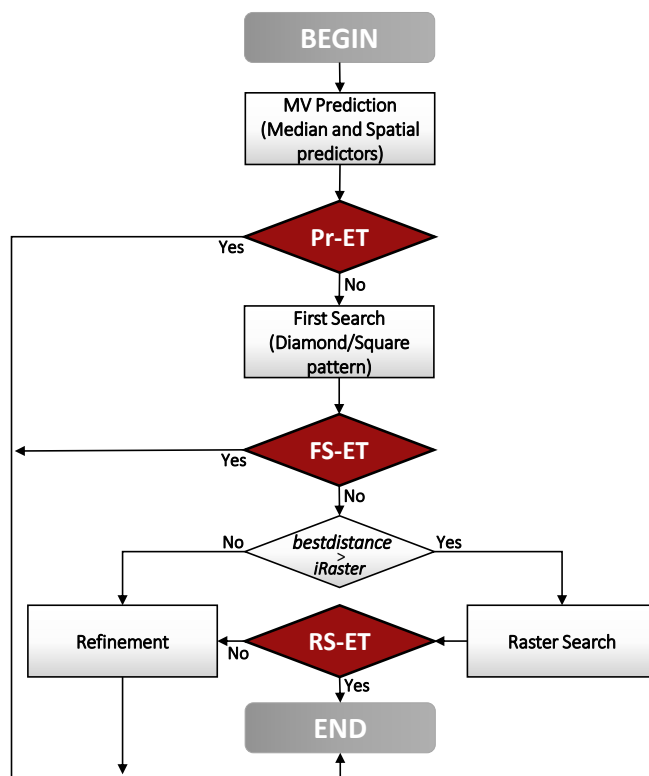


Figura 9.7. Fluxograma do esquema e-TZS proposto em [Gonçalves 2017].

9.4.4. Transcodificação H.264-para-HEVC Rápida Utilizando Árvores de Decisão

Apesar de o padrão HEVC estar gradualmente substituindo o seu antecessor, há uma enorme quantidade de conteúdo legado codificado com o padrão H.264/AVC que deve ser convertido ao HEVC para que seja compatível com novos dispositivos e reprodutores, de forma que a transcodificação H.264-para-HEVC é essencial nesta fase de transição. Conforme apresentado na seção 9.2.3, a complexidade do processo de transcodificação de vídeo é também um problema importante a ser investigado, já que esta tende a superar a da codificação.

O trabalho em [Correa 2016a] apresenta uma abordagem, também baseada em aprendizado de máquina, que visa reduzir a complexidade da transcodificação H.264-para-HEVC utilizando informações coletadas do decodificador H.264/AVC que integra o transcodificador para auxiliar nos processos de decisão do codificador HEVC que também integra o transcodificador, conforme ilustra o processo da Figura 9.8 [Correa 2016a]. A partir de uma estratégia similar às apresentadas nas seções anteriores, um conjunto de atributos foi coletado durante o processo de decodificação H.264/AVC e utilizado pelo algoritmo C4.5 [Quinlan 1993] para construir um conjunto de árvores de decisão que simplificam o particionamento de CUs na codificação HEVC.

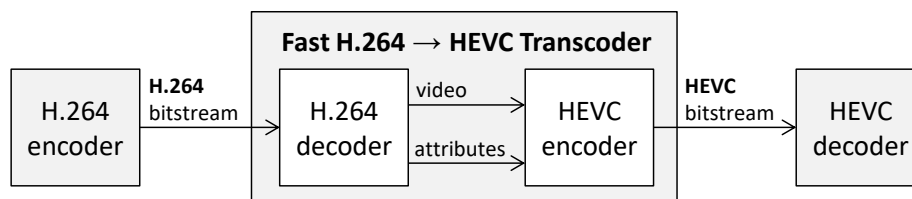


Figura 9.8. Transcodificação rápida H.264-para-HEVC proposta em [Correa 2016a].

A Figura 9.9 [Correa 2016a] mostra os quatro atributos, dentre todos aqueles coletados na decodificação H.264/AVC, que apresentam a melhor correlação com as decisões de particionamento de CU realizadas na codificação HEVC. Note na Figura 9.9(a) que as regiões codificadas com as maiores partições no padrão H.264/AVC (*SKIP*, P16×16 e I16MB) são aquelas raramente particionadas em CUs menores na codificação HEVC, enquanto que as regiões codificadas com as menores partições no H.264/AVC (P8×8, I4MB) são mais frequentemente particionadas em CUs menores no HEVC. A média dos valores de todos os coeficientes diferentes de zero pertencentes a um macrobloco (MB) (área de 16×16 pixels no padrão H.264/AVC), obtidos logo após a etapa de quantização inversa no processo de decodificação, é o parâmetro apresentado na Figura 9.9(b) e também apresenta comportamento similar à Figura 9.9(a). Fica claro que um aumento na ocorrência de particionamento de CUs na codificação HEVC é coincidente com o aumento da média de coeficientes diferentes de zero na área correspondente do vídeo H.264/AVC.

Os dados na Figura 9.9(c) apresentam o número de MBs codificados no modo *SKIP* em cada região 32×32 do vídeo H.264/AVC decodificado. Note-se que o número de MBs *SKIP* é inversamente proporcional à probabilidade de particionar a CU 32×32 localizada na mesma região durante a codificação HEVC. Quando todos os quatro MBs dentro da região 32×32 são codificados como *SKIP*, as chances de particionar a CU 32×32 durante a codificação HEVC reduzem para menos de 9%. Por outro lado, quando nenhum MB é codificado como *SKIP*, as chances de particionar a CU aumentam para 85%. Por fim, a Figura 9.9(d) apresenta dados similares àqueles da Figura 9.9(b), mas agora considerando regiões 64×64 no vídeo codificado com o padrão H.264/AVC. É possível notar que quando a soma dos coeficientes dentro da região é superior a 2.000 na decodificação H.264/AVC, a grande maioria das CUs 64×64 é particionada em CUs menores. Quando este valor fica abaixo de 2.000, a maioria das CUs não é particionada.

Assim como nos trabalhos referenciados anteriormente, a utilidade de cada parâmetro foi analisada através do método IGAE na ferramenta WEKA. Após a etapa de mineração de dados, o algoritmo C4.5 [Quinlan 1993] foi utilizado para gerar três árvores de decisão que geram, como retorno, as decisões *Split* e *Not Split*, que

respectivamente indicam se a CU atual deve ser particionada em quatro CUs menores ou não. Assim como o trabalho revisto na seção 9.4.1, as três árvores de decisão são separadamente construídas e aplicadas às CUs 64×64 , 32×32 e 16×16 . As três árvores de decisão atingiram precisão de 76%, 81% e 83,4%, respectivamente.

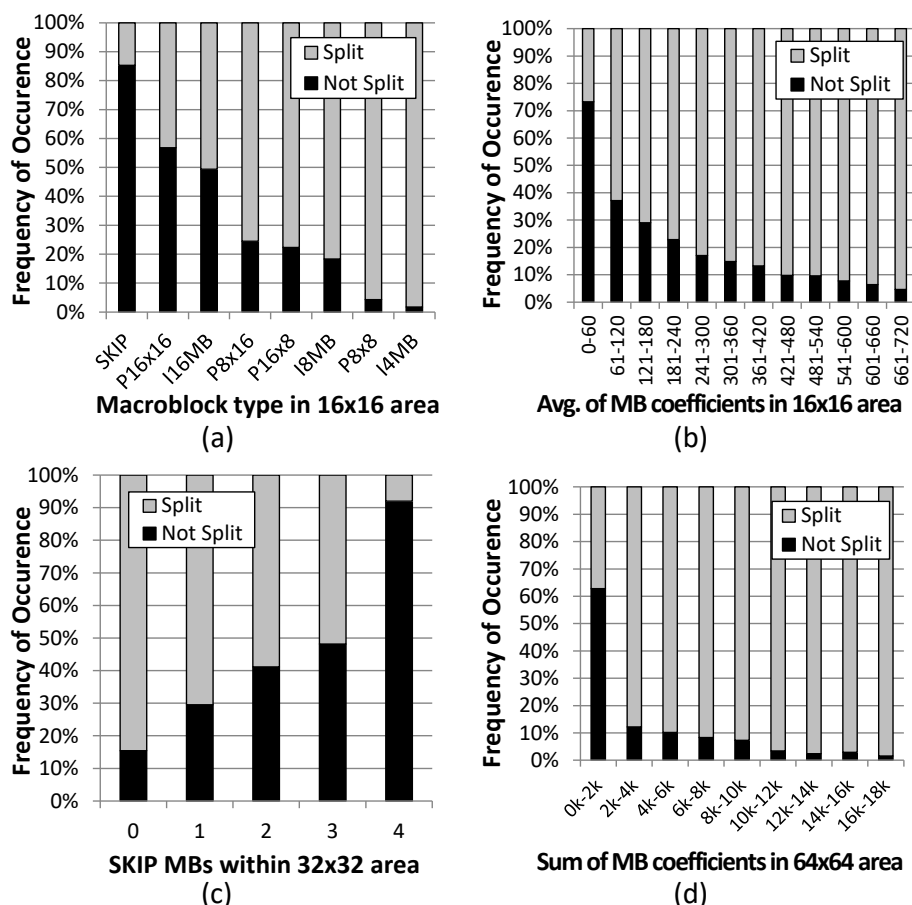


Figura 9.9. Correlação entre ocorrência de casos *Split* e *Não-Split* de CUs na codificação HEVC e (a) tipo de MB, (b) média de coeficientes não-zero no MB, (c) número de MBs *SKIP* em uma área 32×32 , (d) soma dos coeficientes que compõem uma área 64×64 no vídeo H.264 decodificado [Correa 2016a].

Conforme ilustra a Figura 9.8, durante a transcodificação rápida proposta pelos autores em [Correa 2016a], o processo de decodificação opera normalmente, coletando os atributos enquanto decodifica cada MB do vídeo. O codificador então recebe os atributos e a sequência de vídeo, calcula os resultados das árvores de decisão e aplica tais decisões no fluxo de codificação HEVC. Os resultados experimentais apresentados em [Correa 2016a] mostraram que a utilização do esquema leva a uma redução de 44% no tempo de transcodificação H.264-para-HEVC, com uma perda de apenas 1,67% na eficiência da codificação.

9.5. Visão de Futuro

As seções anteriores descrevem aplicações de Aprendizado de Máquina em várias operações de processamento de vídeo, com ênfase na codificação e transcodificação de vídeo para representação no padrão HEVC. Sendo este padrão o

estado da arte em termos de desempenho de codificação, estão já em curso atividades exploratórias de melhoramentos parcelares e globais das várias ferramentas de processamento de sinal incorporadas no HEVC, com vista à proposta de futuras gerações de codificadores de vídeo. Dentre as atividades em curso, são de destacar as que estão sendo iniciadas pelo *Joint Video Exploration Team* (JVET) fundado em outubro de 2015 e constituído por membros dos organismos/grupos de padronização ITU-T VCEG e ISO/IEC MPEG. O objectivo principal destas atividades é obviamente incrementar o fator de compressão atingido pelo padrão atual (HEVC) e apesar de estarem ainda por definir quais as tecnologias a utilizar no sucessor do HEVC, todas as tecnologias e melhoramentos já apresentados nos estudos preliminares têm como denominador comum o aumento da complexidade computacional, comparativamente à do HEVC. A título de exemplo, o método proposto para a predição intra, além de outras modificações, aumenta o número de modos direcionais que passam de 35 para cerca de 250. Esta e outras propostas fazem antever que o próximo padrão de codificação de vídeo seja ainda mais exigente do ponto de vista computacional, o que, conjugado com a adoção de sinais de vídeo com resoluções espaciais, temporais e cromáticas cada vez maiores, consolidará a necessidade de encontrar heurísticas e modificações algorítmicas capazes de reduzir as exigências computacionais até níveis comportáveis pelos dispositivos a usar na codificação e transcodificação de vídeo. Neste contexto de aumento substancial da complexidade computacional dos métodos de compressão de vídeo, que é parte resultante de algoritmos mais complexos caracterizados por um espaço paramétrico mais vasto, a pesquisa manual com forte base intuitiva de heurísticas de aceleração dos processos de compressão não se afigura viável, sendo por isso inevitável o recurso a técnicas automáticas de análise de dados (e.g. mineração de dados) e métodos de aprendizagem automática com ou sem supervisão para projeto de algoritmos para aceleração da compressão de vídeo. É de prever que técnicas de aprendizagem automática recentes, como por exemplo classificadores multivista [Sun 2013], venham a ser úteis no desenvolvimento destas soluções.

Todas as aplicações descritas nas seções anteriores pressupõem que as operações de processamento de vídeo associadas à compressão de vídeo, sejam elas pré ou pós-processamento, codificação ou transcodificação, são executadas localmente sem qualquer distribuição das tarefas de codificação entre vários agentes. A título de exemplo, neste paradigma as operações de codificação de vídeo a decorrer num *smartphone* são executadas integralmente nesse dispositivo, que não é coadjuvado por qualquer outro processador de informação. Com a disponibilização de tecnologias mais eficientes de comunicação e computação distribuída é de esperar que haja uma evolução (ou revolução) também na forma de codificar vídeo, desenvolvendo e implementando algoritmos que facilitam a distribuição das várias operações de processamento de sinal em que se baseia a codificação por sistemas computacionais não-locais. Com efeito, as tecnologias de comunicação com e sem fio atuais exibem já latências suficientemente baixas para permitir que certas operações como a estimação de movimento sejam parcialmente deslocalizadas do codificador para co-processadores remotos com maior poder computacional e sem restrições energéticas sérias. Neste contexto, são particularmente interessantes os sistemas de comunicação LTE com latências extremamente baixas (alguns milissegundos) e as tecnologias de computação distribuída dista “Edge Computing”, que em conjunto possibilitam operações complexas a efectuar durante a codificação sejam realizadas em servidores de computação especializados (MEC - mobile edge computers) localizados nas estações base LTE (e-nodes). Alguns

trabalhos como [Beck 2015] e [Baktir 2017] exploram já este tipo de solução, demonstrando a viabilidade dos conceitos. Estes desenvolvimentos são muito interessantes também do ponto de vista das aplicações de aprendizagem automática à codificação e transcodificação de vídeo, pois permitem que todo o processo de aprendizagem seja efetuado de forma contínua (isto é online) e que modelos mais complexos (por exemplo redes neurais convolucionais profundas) sejam utilizados sem que as limitações computacionais de dispositivos de baixo custo e baixo desempenho (por exemplo webcams sem fio operadas a bateria) se façam sentir. A adoção de métodos de aprendizagem contínua é muito importante, pois um dos principais problemas dos sistemas pré-treinados é o seu fraco desempenho quando aplicados a dados com características muito diferentes dos usados na fase de treino. Curiosamente, a transferência do processo de aprendizagem para servidores tipo *edge* teria como benefício adicional o facto de poder observar dados de *streams* de vídeo variados provenientes de vários utilizadores, com características diferentes, o que provavelmente permitiria treinar continuamente algoritmos de aceleração muito mais flexíveis e performantes.

9.6. Agradecimentos

O levantamento bibliográfico e os trabalhos apresentados ao longo deste capítulo são resultados de uma pesquisa financiada pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS) e pela Fundação para a Ciência e a Tecnologia (FCT-Portugal).

Referências

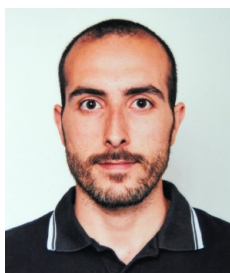
- Baktir, A. C., Ozgovde, A., e Ersoy, C. (2017). How Can Edge Computing Benefit from Software-Defined Networking: A Survey, Use Cases & Future Directions. *IEEE Communications Surveys & Tutorials*.
- Beck, M. T., Feld, S., Fichtner, A., Linnhoff-Popien, C., e Schimper, T. (2015). ME-VoLTE: Network functions for energy-efficient video transcoding at the mobile edge. In *Intelligence in Next Generation Networks (ICIN), 2015 18th International Conference on* (pp. 38-44). IEEE.
- Bjontegaard, G. (2001). Calculation of average PSNR differences between RD-Curves. Proceedings of the ITU-T Video Coding Experts Group (VCEG) Thirteenth Meeting.
- Bossen, F. (2011). Common test conditions and software reference configurations. In Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, 5th meeting, Jan. 2011.
- Chauvet, Antoine, et al. (2016). "Adaptive Post Filter for Reducing Block Artifacts in High Efficiency Video Coding", International Conference on Multimedia Systems and Signal Processing (ICMSSP).
- CISCO (2017). "White paper: Cisco VNI Forecast and Methodology", 2015-2020, 2016. Disponível em: <www.cisco.com/>. Acesso em: 19 de abril de 2017.

- Correa, G., Assuncao, P., Agostini, L., & da Silva Cruz, L. A. (2011). Complexity control of high efficiency video encoders for power-constrained devices. *IEEE Transactions on Consumer Electronics*, 57(4), 1866-1874.
- Correa, G., Assuncao, P., Agostini, L., e da Silva Cruz, L. A. (2012). Performance and computational complexity assessment of high-efficiency video encoders. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(12), 1899-1909.
- Correa, G., Assuncao, P. A., Agostini, L. V., & da Silva Cruz, L. A. (2015). Fast HEVC encoding decisions using data mining. *IEEE transactions on circuits and systems for video technology*, 25(4), 660-673.
- Correa, G., Assunção, P. A., Agostini, L. V., & da Silva Cruz, L. A. (2016). Pareto-based method for high efficiency video coding with limited encoding time. *IEEE Transactions on Circuits and Systems for Video Technology*, 26(9), 1734-1745.
- Correa, G., Agostini, L., & da Silva Cruz, L. A. (2016a). Fast H. 264/AVC to HEVC transcoder based on data mining and decision trees. In *Circuits and Systems (ISCAS), 2016 IEEE International Symposium on* (pp. 2539-2542). IEEE.
- Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. John Wiley & Sons.
- Dai, Y., Liu, D., e Wu, F. (2017). A convolutional neural network approach for post-processing in hevc intra coding. In *International Conference on Multimedia Modeling* (pp. 28-39). Springer, Cham.
- Dietterich, T. G. (2000). Ensemble methods in machine learning. *Multiple classifier systems, 1857*, 1-15.
- Dong, C., Loy, C. C., He, K., e Tang, X. (2014, September). Learning a deep convolutional network for image super-resolution. In *European Conference on Computer Vision* (pp. 184-199). Springer, Cham.
- Dong, C., Deng, Y., Change Loy, C., e Tang, X. (2015). Compression artifacts reduction by a deep convolutional network. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 576-584).
- Fu, C. M., Alshina, E., Alshin, A., Huang, Y. W., Chen, C. Y., Tsai, C. Y., ... & Han, W. J. (2012). Sample adaptive offset in the HEVC standard. *IEEE Transactions on Circuits and Systems for Video technology*, 22(12), 1755-1764.
- Gonçalves, P., Correa, G., Porto, M., Zatt, B., Agostini, L. (2017). Multiple Early-Termination Scheme for TZSearch Algorithm based on Data Mining and Decision Trees. In *Proceedings of the IEEE 19th International Workshop on Multimedia Signal Processing* (pp. 1-6). IEEE.
- Gonzalez, R. C., & Woods, R. E. (2002). *Digital image processing*. Prentice Hall. Upper Saddle River, NJ.
- Grundmann, M., Kwatra, V., Han, M., e Essa, I. (2010). Efficient hierarchical graph-based video segmentation. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on* (pp. 2141-2148). IEEE.

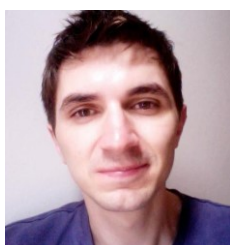
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1), 10-18.
- Helle, P., Oudin, S., Bross, B., Marpe, D., Bici, M. O., Ugur, K., ... & Wiegand, T. (2012). Block merging for quadtree-based partitioning in HEVC. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(12), 1720-1731.
- Judd, T., Ehinger, K., Durand, F., e Torralba, A. (2009). Learning to predict where humans look. In *Computer Vision, 2009 IEEE 12th international conference on* (pp. 2106-2113). IEEE.
- Katsenou, A. V., Afonso, M., Agrafiotis, D., e Bull, D. R. (2016). Predicting video rate-distortion curves using textural features. In *Picture Coding Symposium (PCS), 2016* (pp. 1-5). IEEE.
- Kim, I. K. (2013). "High efficiency video coding (HEVC) test model 10 (HM10) encoder description. Doc. JCTVC-L1002." 12th Meeting: Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG. Vol. 16.
- Kuhn, M., Weston, S., Coulter, N., & Quinlan, R. (2014). C50: C5. 0 decision trees and rule-based models. R package version 0.1. 0-21, URL <http://CRAN.R-project.org/package=C50>.
- Lainema, J., Bossen, F., Han, W. J., Min, J., & Ugur, K. (2012). Intra coding of the HEVC standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(12), 1792-1801.
- Li, Z., Aaron, A., Katsavounidis, I., Moorthy, A. e Manohara, M. (2016). Toward A Practical Perceptual Video Quality Metric. *Netflix Technology Blog*, Junho.
- Li, S., Zhang, F., Ma, L., e Ngan, K. N. (2011). Image quality assessment by separately evaluating detail losses and additive impairments. *IEEE Transactions on Multimedia*, 13(5), 935-949.
- Liu, T., Lin, Y., Lin, W., e Kuo, C. (2013) "Visual quality assessment: Recent developments, coding applications and future trends", *APSIPA Transactions on Signal and Information Processing*, 2.
- Ma, C., Naser, K., Ricordel, V., Le Callet, P., e Qing, C. (2016). An adaptive Lagrange multiplier determination method for dynamic texture in HEVC. In *Consumer Electronics-China (ICCE-China), 2016 IEEE International Conference on* (pp. 1-4). IEEE.
- Pan, X., e Chen, Z. (2017). Multi-Layer Quantization Control for Quality-Constrained H. 265/HEVC. *IEEE Transactions on Image Processing*, 26(7), 3437-3448.
- Park, W. S., e Kim, M. (2016). CNN-based in-loop filtering for coding efficiency improvement. In *Image, Video, and Multidimensional Signal Processing Workshop (IVMSP), 2016 IEEE 12th* (pp. 1-5). IEEE.
- Pinson, M. H., Choi, L. K., e Bovik, A. C. (2014). Temporal video quality model accounting for variable frame delay distortions. *IEEE Transactions on Broadcasting*, 60(4), 637-649.

- Puri, S., Lasserre, S., e Le Callet, P. (2016). Annealed learning based block transforms for HEVC video coding. In *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on* (pp. 1135-1139). IEEE.
- Quinlan, J. R. (1993). *C4. 5: programs for machine learning*. San Francisco: Morgan Kaufmann Publishers.
- Sheikh, H. R., e Bovik, A. C. (2006). Image information and visual quality. *IEEE Transactions on image processing*, 15(2), 430-444.
- Souly, N., e Shah, M. (2016). Visual saliency detection using group lasso regularization in videos of natural scenes. *International Journal of Computer Vision*, 117(1), 93-110.
- Sullivan, G. J., & Wiegand, T. (1998). Rate-distortion optimization for video compression. *IEEE signal processing magazine*, 15(6), 74-90.
- Sullivan, G. J., Ohm, J., Han, W. J., & Wiegand, T. (2012). Overview of the high efficiency video coding (HEVC) standard. *IEEE Transactions on circuits and systems for video technology*, 22(12), 1649-1668.
- Sun, S. (2013). A survey of multi-view machine learning. *Neural Computing and Applications*, 23(7-8), 2031-2038.
- Vega, M. T., Mocanu, D. C., Famaey, J., Stavrou, S., e Liotta, A. (2017). Deep Learning for Quality Assessment in Live Video Streaming. *IEEE Signal Processing Letters*, 24(6), 736-740.
- Wang, Z., Bovik, A.C., Sheikh, H.R. e Simoncelli, E.. (2014). Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* 13. 600-612.
- Wang, T., Chen, M., e Chao, H. (2017). A Novel Deep Learning-Based Method of Improving Coding Efficiency from the Decoder-End for HEVC. In *Data Compression Conference (DCC), 2017* (pp. 410-419). IEEE
- Wiegand, T., Sullivan, G. J., Bjontegaard, G., & Luthra, A. (2003). Overview of the H. 264/AVC video coding standard. *IEEE Transactions on circuits and systems for video technology*, 13(7), 560-576.
- Xu, M., Jiang, L., Sun, X., Ye, Z., e Wang, Z. (2017). Learning to detect video saliency with HEVC features. *IEEE Transactions on Image Processing*, 26(1), 369-385.
- Yan, N., Liu, D., Li, H., e Wu, F. (2017). A Convolutional Neural Network Approach for Half-Pel Interpolation in Video Coding. *arXiv preprint arXiv:1703.03502*.
- Zhang, Z., Jing, T., Han, J., Xu, Y., e Zhang, F. (2017). A New Rate Control Scheme for Video Coding Based on Region of Interest. *IEEE Access*.

ANEXO 1: Biografia dos Autores



Guilherme Corrêa é Doutor em Engenharia Electrotécnica e de Computadores (2015) pela Universidade de Coimbra – Portugal, Mestre em Computação (2010) pela Universidade Federal do Rio Grande do Sul (UFRGS) e Bacharel em Ciência da Computação (2009) pela Universidade Federal de Pelotas (UFPel). É Professor Adjunto da UFPel desde 2016, onde atua nos cursos de graduação na área de Computação e no Programa de Pós-Graduação em Computação (PPGC). Suas atividades de pesquisa estão vinculadas ao *Video Technology Research Group* (ViTech) da UFPel, onde atua nos temas de codificação de imagens e vídeos digitais, entrega e transporte de multimídia em redes e projeto de sistemas digitais multimídia.



Mateus Grellert recebeu o título de Mestre em Ciência da Computação da Universidade Federal do Rio Grande do Sul (UFRGS) em 2014, e é atualmente aluno de Doutorado na mesma instituição. Ele atua na pesquisa em sistemas multimídia desde 2007. Seus trabalhos publicados incluem tópicos como arquiteturas reconfiguráveis, desenvolvimento de sistemas eficientes em energia, arquiteturas de memória, e sistemas de codificação de vídeo. Seus interesses englobam algoritmos eficientes de codificação e transcodificação de vídeo com base em técnicas de Machine Learning. Atualmente, atua como pesquisador no Grupo de Microeletrônica (GME) da UFRGS e no Instituto de Telecomunicações (IT) da Universidade de Coimbra (Portugal).



Sergio Bampi recebeu os títulos de Engenheiro Elétrico e Bacharel em Física pela Universidade Federal do Rio Grande do Sul em 1979. Recebeu também os títulos de Mestre e Ph.D. em Engenharia Elétrica pela Universidade de Stanford em 1982 e 1986, respectivamente. Ele é professor titular do Instituto de Informática da Universidade Federal do Rio Grande do Sul desde 1981. Já foi presidente da Sociedade Brasileira de Microeletrônica, da agência de financiamento FAPERGS e também diretor técnico do CEITEC. Em 2009 e 2010, foi um *Distinguished Lecturer* da IEEE CAS. Sua produção inclui mais de 270 artigos em áreas como design de circuitos CMOS analógicos, digitais e de RF, e arquiteturas e algoritmos para codificação de vídeo. Atua como pesquisador no Grupo de Microeletrônica (GME) da UFRGS.



Luis Alberto da Silva Cruz recebeu os títulos de Licenciado e Mestre em Engenharia Electrotécnica da Universidade de Coimbra – Portugal em 1989 e 1993, respectivamente. Também recebeu os títulos de Mestre em Matemática e Ph.D. em *Electrical Computer and Systems Engineering* do Rensselaer Polytechnic Institute (RPI), Troy, NY, US em 1997 e 2000, respectivamente. Desde 1990, atua como professor no Departamento de Engenharia Electrotécnica e de Computadores da Universidade de Coimbra. Como pesquisador, atua no pólo de Coimbra do Instituto de Telecomunicações nos temas de processamento e codificação de vídeo, processamento de imagens médicas e comunicações sem fio.

Capítulo

10

Distribuição de Vídeo na era da Infraestrutura Definida por Software

Samuel Kopp¹, Michael Prieto Hernandez¹, Lucas Chigami¹,
Reinaldo Matushima², Bruno Klava² Regina Melo Silveira¹

¹Laboratório de Arquitetura e Redes de Comutadores da
Escola Politécnica da Universidade de São Paulo (USP) – São Paulo, SP – Brazil

²TVoD - TV on Demand Consultoria e Soluções – São Paulo, SP – Brazil

samuel@larc.usp.br, mhernandez@larc.usp.br,
lchigami@larc.usp.br, regina@larc.usp.br,
bruno@tvod.com.br, reinaldo@tvod.com.br

Abstract

The evolution of video consumption in recent years has generated a great demand for network infrastructures, so that the video distribution services have played a key role in enabling services such as Netflix, YouTube and Video@RNP. This context has stimulated the research around the CDNs and SDN to optimize the use of the resources and of greater dynamicity in the operation of the networks, so that the Infrastructure Defined by Software has been highlighted. This chapter addresses this issue by bringing the state of the art to the use of network infrastructures to support video applications and prospecting how they will be used and improved in the coming years.

Resumo

A evolução do consumo de vídeo nos últimos anos tem gerado uma grande demanda para as infraestruturas de rede, de modo que os serviços de distribuição de vídeo têm exercido um papel fundamental para viabilizar serviços, como Netflix, YouTube e Video@RNP. Este contexto tem estimulado as pesquisas em torno das CDNs e SDN para otimizar do uso dos recursos e de maior dinamicidade na operação das redes, de modo que a Infraestrutura Definida por Software tem recebido destaque. Este capítulo trata desta questão, trazendo a tona o estado da arte no uso das infraestruturas de rede para o dar suporte às aplicações de vídeo e fazendo uma prospecção de como estas serão usadas e aperfeiçoadas nos próximos anos.

10.1. Introdução

A utilização de vídeo na Internet tem se tornado cada vez mais popular. Atualmente, os conteúdos em vídeo não são gerados apenas por produtores profissionais - seja no âmbito educacional, de entretenimento ou jornalístico. A disponibilização de ferramentas amigáveis para produção de conteúdo, inclusive através de *smartphones*, tem proporcionado a “postagem” de vídeos por qualquer pessoa, a qualquer momento e de qualquer lugar.

Por outro lado, os avanços das tecnologias para produção de conteúdos de alta qualidade proporcionam o cenário adequado para que sejam exploradas formas mais intensas para o consumo de vídeos de entretenimento de alta resolução e com uso de óculos de realidade virtual, como também de aplicativos de vídeo colaboração. Este cenário tem demandado, tanto o aumento da capacidade dos serviços de distribuição, como também o desenvolvimento de novos serviços de mídias que possam dar suporte à estas evoluções.

O reflexo disto é que, segundo análise prospectiva da Cisco[Cisco 2017], até o final de 2021 82% do tráfego da Internet será constituído por fluxos de vídeo. Este estudo também aponta que a entrega destes fluxos só tem alcançado eficiência devido às CDNs (Content Delivery Networks – Redes de Distribuição de Conteúdo), sendo que deverão ser responsáveis por mais de 71% do tráfego total da Internet até 2021.

Por outro lado, as pesquisas para a evolução da arquitetura da Internet tem proporcionado novas oportunidades através do surgimento da SDN (*Software-Defined Networking* - Rede Definida por Software)[McKeown 2009]. Através desta, é possível implementar mecanismos para otimizar o uso da rede utilizando técnicas de engenharia de tráfego com uma visão holística. Adicionalmente, o uso da Computação em Nuvem (*Cloud Computing*) permite implementar sistemas com provisionamento dinâmico dos recursos computacionais de acordo com a demanda solicitada ao serviço. A associação destas - as Redes Definidas por Software e a Computação em Nuvem - resultou na SDI (*Software-Defined Infrastructure* - Infraestrutura Definida por Software) [Kang et al. 2013]. Estes fatores impulsionam um novo ciclo de evolução na distribuição de conteúdo na Internet, em especial para fluxos de vídeo [Fan et al. 2017].

Assim este capítulo tem o objetivo prover uma visão geral sobre o futuro da distribuição de vídeo. Para isso, primeiramente são apresentados os projetos voltados à distribuição de vídeo desenvolvidos pela parceria LARC/TVoD subsidiados pela RNP e FAPESP no últimos anos, os quais foram alinhados com estas perspectivas de evolução e demandas da distribuição de conteúdo. Na seção 10.3 é feita a conceituação da infraestrutura utilizada atualmente para a distribuição de conteúdos, considerando evolução, variações de arquiteturas e integrações. Na seção 10.4 é apresentado um levantamento do estado da arte com relação a interface com o usuário, serviços de distribuição de conteúdo e infraestruturas definidas por software. Posteriormente, são descritas na seção 10.5 algumas das infraestruturas em operação que fazem uso destas tecnologias, as quais são utilizadas como casos de referências para implementação dos serviços propostos. Finalmente, nas seções 10.6 e 10.7 são apresentadas respectivamente uma visão de futuro da distribuição de vídeo tendo em vista a era da infraestrutura definida por software e as considerações finais.

10.2. Projetos

Em 2005, o LARC (Laboratório de Arquitetura e Redes de Computadores) iniciou o *Grupo de Trabalho de Gerência de Vídeo*, fomentado pela RNP, na qual propôs uma *Plataforma de Gerência de Vídeo* [Uchôa e Silveira 2007].

Este projeto se baseou na percepção de que, apesar de já haverem pesquisas no Brasil na área de vídeo, como o *Sistema para Distribuição de Conteúdo Multimídia* [Silveira, Fonseca et al. 2003] e o *Grupo de Trabalho de Vídeo Digital* [Lemos 2002], havia a necessidade de ferramentas adequadas para a publicação e consumo de vídeos. Isto é, que permitissem de forma amigável ao usuário a publicação, busca e visualização desses conteúdos através de uma interface para navegadores web. O tratamento destes aspectos foram grandes desafios na época, e atualmente são de grande relevância em qualquer plataforma de vídeo.

Na mesma época era lançada uma das principais plataformas de distribuição de vídeo em uso atualmente, o YouTube, que revolucionou o consumo de vídeos na Internet. Um dos motivos deste sucesso foi tratar a questão da usabilidade de acesso. Segundo Roelof Botha, “*grande parte do sucesso do YouTube se deu devido a facilidade de uso do site*”. Entre estas facilidades, destaca-se o suporte a upload de qualquer formato de vídeo com posterior adaptação (automática) para o formato “padrão de visualização” - o *Flash* [Wasserman 2015].

A evolução desta plataforma desenvolvida pelo LARC/RNP deu-se mediante o *GT Overlay - Grupo de Trabalho em Redes de Serviços Sobrepostos* [Uchôa e Silveira 2007] durante os anos 2007-2009. O principal objetivo desta etapa foi permitir a criação de uma Federação de Rede de Distribuição de Vídeo como forma de tratar um dos grandes gargalos, a construção de Redes de Conteúdos escaláveis [Lao et al. 2007]. Além disso, o amadurecimento da plataforma deu origem ao serviço de distribuição de vídeo - *Video@RNP* - provido pela RNP às instituições associadas.

Em 2009, o grupo de pesquisa que desenvolveu estes projetos formou uma *spin-off*, a TVoD, com o objetivo de prover a manutenção e evolução destes projetos. Foram lançados então diversos serviços que nasceram do amadurecimento destes projetos, dentre eles o IPTV-USP, eAulas-USP, Zappiens.br e o TVEscola. Cada um destes serviços estão inseridos em contextos distintos, com evoluções tecnológicas independentes.

A continuidade destes trabalhos deu origem ao projeto “FAPESP/PIPE - Rede de Distribuição de Conteúdo integrada a Rede Definida por Software” em 2014. Este projeto teve como principal objetivo a melhoria da qualidade de serviço na transmissão de vídeo e otimização do uso da rede, através do novo paradigma de construção de redes - o SDN [McKeown 2009].

A seguir são apresentadas algumas questões relevantes destes trabalhos que auxiliam no entendimento da evolução do uso de vídeo.

10.2.1. Plataforma de Gerência de Vídeo

O desenvolvimento da Plataforma de Gerência de Vídeo teve como principal objetivo a elaboração de uma ferramenta baseada na web que permitisse fazer o gerenciamento do consumo de vídeo através de três enfoques distintos (i) o publicador de conteúdo; (ii)

o administrador do sistema e (iii) e o gerente da infraestrutura. Desta forma, diferentes perfis de usuários tinham acesso a diferentes funcionalidades. Para prover publicação, catalogação, indexação, busca e consumo de vídeos de forma eficiente foram utilizados metadados padronizados utilizando os padrões MPEG-7 e MPEG-21.

Tal plataforma foi integrada ao sistema de distribuição de conteúdo utilizado até então pela RNP, a RVD (Rede de Vídeo Digital), desenvolvida pelo *GT-Video Digital* [Lemos 2002]. Os elementos que compunham este sistema eram: (i) Portal (interface web de acesso), (ii) o Gerente, que fazia o papel de encaminhador de requisições (*request-router*) e (iii) os refletores (JD-VoD e JD-Live), que executavam a função de servidores de borda (*surrogate/edge servers*) os quais davam suporte, respectivamente, à transmissão de vídeo sob demanda e vídeo ao vivo.

Assim, a primeira versão da plataforma, lançada em 2006, foi integrada à rede já existente na RNP e é utilizada até hoje como o serviço denominado Video@RNP. Uma segunda versão da mesma plataforma foi adaptada para as necessidades de distribuição e consumo da Universidade de São Paulo, tendo o serviço sido batizado de IPTV USP, e lançado em agosto de 2007. Esta plataforma conta hoje com um acervo de mais de 20.000 títulos produzidos na USP, sendo o ponto de acesso único de toda a produção relevante de vídeo da instituição.

Nas primeiras versões da plataforma, a visualização dos vídeos era feita através de *plugins* para os navegadores (ex: Windows Media Player, VLC), mas ao decorrer da evolução das tecnologias de visualização, foi sendo adotado o “*Flash Player*”, posteriormente o *player* HTML 5 e recentemente o HLS e DASH.

10.2.2. Federação de Rede de Distribuição de Vídeo

Os trabalhos voltados diretamente para a questão de distribuição de vídeo foram evoluindo no sentido de permitir a expansão escalável da abrangência e capacidade de distribuição. Para isso, o projeto *GT-Overlay* estruturou a RVD de forma a construir uma federação de redes sobrepostas de distribuição de vídeo.

O primeiro grande desafio deste projeto foi a elaboração de mecanismos para estabelecer a interconexão das redes sobrepostas. Para isso, foi desenvolvida uma arquitetura que permitisse a interconexão entre múltiplas redes de distribuição de vídeo para realizar a transmissão através de uma federação. Em [Uchôa, Kopp et al. 2009] foram apresentados os resultados do experimento realizado em uma federação composta por nós em 3 redes (USP, RNP e RedClara).

O segundo foi o desenvolvimento de um protocolo de segurança para redes sobrepostas [Pimentel et al. 2015]. Uma das principais preocupações foi garantir a autonomia e sigilo na operação de cada rede, ou seja, apesar da construção da rota utilizada para a transmissão do vídeo ser realizada através de um processo cooperativo entre os *request-routers* das redes envolvidas na transmissão, os elementos da rota internos de cada rede são “visíveis” apenas pelos demais elementos da mesma rede.

10.2.3. Rede de Distribuição de Conteúdo integrada com Rede Definida por Software

O projeto “FAPESP/PIPE - Rede de Distribuição de Conteúdo integrada a Rede Definida por Software” tem como objetivo a elaboração, especificação e desenvolvimento de uma solução para CDN que atue de forma integrada com as SDNs, possibilitando usufruir das funcionalidades providas por estas para otimizar a distribuição de conteúdos na Internet.

Este projeto está fundamentado no princípio da *Segmentação da arquitetura em camadas administrativas coesas com o modelo operacional da Internet*. A ideia principal é permitir que entidades administrativamente distintas possam gerenciar de forma autônoma os elementos da arquitetura pertinentes a sua atuação no contexto de negócio na Internet. Assim, o gerenciamento da distribuição do conteúdo é realizado pelo operador da CDN, e as operações de engenharia de tráfego pelo operador da rede SDN. Deste modo, apesar de existirem mecanismos para a realização de operações dinâmicas entre as camadas, é necessário que haja a autonomia entre elas.

Para isso foi definido um elemento de intermediação entre os módulos da CDN e o controlador da SDN - o *Módulo de Abstração de Rede*. Assim é possível criar uma camada de abstração acima do serviço de topologia dos controladores SDN, baseado nas políticas de negócio dos ISPs. Adicionalmente, a interação deste com um *Gerenciador de Topologia* da CDN permite ao serviço de distribuição de conteúdo solicitar ao *Módulo de Abstração de Rede* a criação dos circuitos virtuais que interligam os servidores de réplica. Isto é, definir proativamente um conjunto de regras nas tabelas de fluxo do *OpenFlow* que conectam os servidores que efetuam a distribuição do conteúdo. A ideia fundamental é associar o conceito de caminho/rota sobreposta (*overlay path/route*) da CDN com o de circuito virtual em SDN. A consequência desta associação é a capacidade de aumentar a escalabilidade do conjunto de regras nas tabelas de fluxos. Deste modo, cada regra na tabela de fluxos representa a conexão entre dois servidores de réplica e não apenas uma única transmissão, atendendo ao envio de múltiplos conteúdos.

Como resultado, será possível ao serviço de distribuição de conteúdo garantir maior desempenho e eficiência ao definir a melhor rota a ser utilizada para o encaminhamento do conteúdo ao usuário solicitante.

10.3. Distribuição de Conteúdo na Internet

A evolução das tecnologias de rede e as aplicações oferecidas aos usuários da Internet estão diretamente relacionadas [Fluckiger 1995]. Desde o desenvolvimento da tecnologia digital começaram a surgir os serviços de transmissão de vídeo. A ideia de um serviço de transmissão com garantia de entrega surgiu na década de 90 com a tecnologia ATM (Asynchronous Transfer Mode) [Orzessek e Sommer 1997] padronizada pelo ITU-T. Graças a sua arquitetura, a utilização de circuitos virtuais e a comutação de pequenos pacotes de dados (denominados células), a rede ATM é capaz de transmitir simultaneamente diferentes tipos de tráfego, como: voz, vídeo e dados. No entanto, a maneira como a rede lida com cada um desses tráfegos depende das características do fluxo e dos requisitos da aplicação. Desta forma, é definido um conjunto de parâmetros de qualidade de serviço (QoS), assim como um conjunto de parâmetros de perfil de tráfego para cada aplicação, que são usados para a definição dos recursos alocados no circuito virtual.

A proposta do serviço de vídeo sob demanda [Committee 1997], estimulou pesquisas sobre estratégias de recodificação do fluxo e vídeo em caso de restrição de recursos [Silveira e Ruggiero 2001] e otimização do processo de controle de admissão [Cheng e Chang 1997]. No entanto, as tecnologias como o Frame Relay e o ATM não se integravam adequadamente às aplicações IP, exigindo grande adaptação, e conseqüentemente, perda de desempenho. Além disso, o alto custo dos equipamentos ATM e o advento da tecnologia Gigabit Ethernet fez com que esta tecnologia tivesse um espaço bastante limitado no mercado. Apesar do insucesso comercial da tecnologia ATM na rede pública de dados, alguns dos conceitos propostos em sua especificação foram inspiradores para o surgimento de outras tecnologias de rede. Por exemplo, o uso de mecanismos para controle de qualidade de serviço (QoS) para o atendimento à entrega de serviços específicos, como o caso dos Serviços Integrados (IntServ) e Serviços Diferenciados (DiffServ). Outros conceitos correlatos à QoS surgiram e trabalhos relacionados à qualidade de atendimento da rede passaram a considerar mecanismos de tratamento do tráfego como os parâmetros de QoE (Quality of Experience) [Nam et al. 2014; Goran, Hadzialic e Begovic 2014] e QoT (Quality of Transmission) [Cugini et al. 2008].

Estas tecnologias influenciam fortemente o mercado de telecomunicação para o oferecimento de serviços em geral, incluindo o de entrega de fluxos de vídeo. No entanto, o grande desafio dos provedores de serviço de rede (ISPs) é a otimização da entrega de conteúdo aos seus usuários através de soluções que apresentem alto retorno de investimento, flexibilidade e facilidade nos processos de evolução, manutenção e operação dos sistemas [Yue 2013]. Assim, apesar de haverem diversos métodos para distribuição de conteúdo, as CDNs têm sido amplamente utilizadas. Um dos principais motivos é a relativa facilidade de implantação, visto que não depende de uma tecnologia de rede específica para sua operação devido à sua atuação basicamente na camada de aplicação. Além disso, caso haja disponibilidade de utilizar mecanismos de otimização de tráfego, pode ser utilizado de forma independente em cada segmento de rede, sem que isso afete sua arquitetura.

10.3.1. Evolução da Distribuição de Conteúdo na Internet

O termo CDN surgiu nos anos 90 com a necessidade de criar meios de prover desempenho e escalabilidade aos serviços Web, a fim de evitar que *sites* com grande volume de acesso apresentassem problemas de congestionamento e conseqüente indisponibilidade. Este tipo de problema ficou conhecido como *Flash Crowd* [Arlitt e Jin 2000] ou *SlashDot Effect* [ADLER 1999], que consiste em um aumento repentino e significativo da demanda a um determinado conteúdo em um sistema, ocasionando a indisponibilidade do mesmo.

No início da transmissão de vídeo através da Internet, era necessária a utilização de plugins de players de vídeos (ex: Apple QuickTime, Microsoft Media Player e RealPlayer da RealNetworks) para a visualização destes conteúdos. Para o controle de transmissão dos conteúdos, os protocolos mais utilizados eram: o Microsoft Media Server (MMS) e o Real Time Streaming Protocol (RTSP) - este último desenvolvido pela RealNetworks, Netscape e Columbia University e posteriormente padronizado pelo IETF [Schulzrinne, Rao e Lanphier 1998]. A transmissão dos fluxos de vídeo por sua vez era realizado através de UDP (User Datagram Protocol) ou via TCP (Transmission Control Protocol) com “*http chunked*”.

Posteriormente, emergiu uma nova geração de serviços de vídeo, no qual o YouTube ganhou grande notoriedade. Esta onda foi impulsionada por algoritmos de compressão mais eficientes (como H.263 e o H.264), o protocolo de streaming RTMP e o plugin Flash, que tornou-se o “padrão” para visualização de vídeos na Web. Além disso, o download progressivo e o recurso de busca (*seek*) no vídeo foram mecanismos que também trouxeram benefícios para o consumo de vídeo.

Com o surgimento do HTML5 (com a *tag video*) foi possível visualizar os vídeos diretamente pelo navegador, isto é, sem a utilização de plugins. Assim, começou um novo ciclo de evolução na distribuição de vídeo. Este movimento foi impulsionado com o desenvolvimento do HLS (HTTP Live Streaming) pela Apple. Com a remoção do suporte ao Flash nos *smartphones* da Apple, deu-se a mudança do padrão para HTML5. Com o suporte nativo no navegador à visualização do vídeo e a requisição dos conteúdos usando métodos padrões para a solicitação aos vídeos (*http byte ranges*), foi possível utilizar servidores HTTP padronizados para implementar os serviços de distribuição das CDNs. Ademais, com o HTTP Adaptativo (HLS/DASH), foi possível oferecer o suporte à múltiplas resoluções de um mesmo vídeo de forma dinâmica, o que permitiu uma melhoria significativa na QoE de consumo dos vídeos. Por fim, recentemente está emergindo uma nova geração de Redes de Conteúdos, impulsionada pelas tecnologias de infraestrutura, como a Computação em Nuvem e as Redes Definidas por Software. Estas evoluções são o alvo das seções 10.4 e 10.6.

10.3.2. Arquitetura das CDNs

O princípio básico do funcionamento das CDNs é fazer um balanceamento de carga entre diversos servidores com réplicas do conteúdo original e direcionar o usuário para um servidor mais próximo deste, de forma a minimizar a latência no acesso ao conteúdo, maximizando assim a taxa de transmissão devido aos efeitos do da janela deslizante do TCP. Em suma, as CDNs são constituídas por um conjunto de elementos de rede que podem ser organizados de forma centralizada, hierarquizada ou descentralizada. A arquitetura básica de uma CDN, segundo [Buyya, Pathan e Vakali 2008], é composta por diversos componentes ilustrados na figura 10.1 e descritos a seguir.

- *Entrega de conteúdos* - constituído pelo servidor de origem, que é a fonte dos conteúdos, e um conjunto de servidores de réplica (também conhecidos como servidores de borda) que são os servidores que efetivamente entregam os conteúdos aos clientes;
- *Encaminhamento de requisição* - faz o redirecionamento das requisições dos clientes aos servidores de réplica apropriados;
- *Distribuição* - responsável por replicar o conteúdo do servidor de origem, disponibilizado pelos clientes da CDN, para os servidores de réplica e garantir a consistência dos conteúdos em todo o sistema;
- *Contabilização* - mantém o registro das requisições dos clientes e estatísticas de uso, que podem ser utilizados para realizar a monetização dos serviços prestados e controlar o consumo dos clientes da CDN.

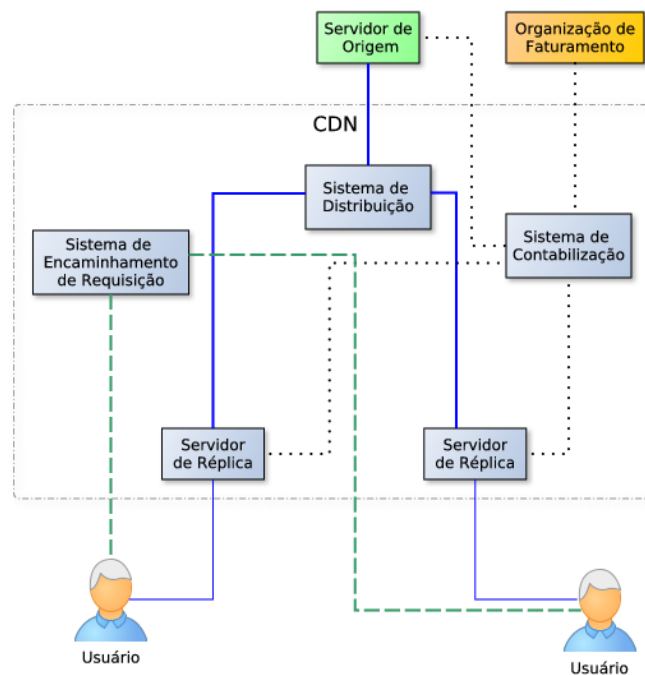


Figura 10.1. Visão geral da arquitetura de uma CDN (adaptado de [Buyya, Pathan e Vakali 2008]).

Além disso, ao analisar como estes componentes são distribuídos na infraestrutura, podemos diferenciar dois tipos de estratégias básicas de implantação [Kurose e Ross 2012]:

- *Enter deep* - alocar um conjunto de servidores em diversos provedores de Internet - *Internet Service Providers (ISPs)*, a fim de se posicionar o mais perto possível do usuário final. A situação ideal seria ter servidores alocados nos ISPs de cada usuário final, minimizando assim a necessidade de desempenho nas comunicações de longas distâncias.
- *Bring home* - estabelecer grandes infraestruturas de armazenamento e distribuição, interligados por um *backbone*, conectando os ISPs até esta infraestrutura. Normalmente estas instalações ficam próximas de PTTs (Pontos de Troca de Tráfego) ou PoP (Pontos de Presença), facilitando assim a conexão com os ISPs e empresas de comunicação.

Contudo, independentemente da estratégia de implantação, pode-se dizer de forma objetiva que a principal funcionalidade de uma CDN é a capacidade de atender as requisições de acesso aos conteúdos de forma escalável, isto é, prover um serviço que através de mecanismos de distribuição de carga, possa atender a uma grande demanda de requisições. Para isso, é necessário primeiramente uma forma de validar a requisição com o objetivo de verificar: (i) se o conteúdo solicitado é válido e (ii) se o usuário está autorizado a acessá-lo. Em seguida é preciso que o encaminhador de requisições escolha o(s) servidor(es) de réplica que irá(ão) atender a requisição. Esta seleção pode considerar

diversos fatores como: (i) a localização topológica do cliente, dos servidores de réplica e do servidor de origem, (ii) níveis de utilização da infraestrutura, tanto da rede como dos servidores e (iii) custo operacional. Finalmente, é necessário um mecanismo para encaminhar o cliente ao servidor de réplica escolhido. Adicionalmente, pode ser agregado mecanismos de controle de QoS da rede.

Gerenciamento da topologia

O gerenciamento da topologia da CDN é uma etapa fundamental para a otimização da distribuição de conteúdo. Esta atividade é realizada de forma assíncrona com o consumo do conteúdo. No entanto, a consistência e acuracidade deste processo implica diretamente na robustez da CDN, pois baseado nesta, é que o encaminhador de requisições irá realizar as decisões para o atendimento das requisições de consumo do conteúdo.

A atividade mais básica no gerenciamento da topologia é mapear os servidores de réplica, verificando se estão disponíveis para efetuar a distribuição do conteúdo. Estas informações são cruciais para que o encaminhador de requisições possa escolher um dos servidores de réplica mapeados para atender a requisição. Adicionalmente é essencial haver um modo de consolidar o estado de utilização dos servidores de réplica através de serviços de monitoração dos seus recursos computacionais. Estas informações representam uma parcela muito importante na tomada de decisão, pois permitem realizar um balanceamento de carga mais preciso.

Além disso, o mapeamento da infraestrutura de rede é essencial. As informações mais importantes nesta atividade são a identificação das regiões de redes e os endereços de rede que então alocados em cada região. Um exemplo de região pode ser a rede de um Sistema Autônomo ou uma subárea deste. Consequentemente, os endereços de rede alocados para esta região podem ser identificados pelos CIDRs - *Classless Inter-Domain Routing* definidos para este Sistema Autônomo ou subárea. Finalmente, o monitoramento das informações do estado da rede representa um importante fator adicional de otimização na escolha dos servidores de réplica. Estas informações devem ser coletadas e consolidadas para identificar quais enlaces estão mais sobrecarregados, ou se algum destes apresenta alguma falha operacional. Assim, de acordo com essas informações, será possível designar melhores rotas para determinados tipos de fluxos a serem transmitidos pela CDN.

Mecanismo de encaminhamento de requisição

Um aspecto muito importante na arquitetura de uma CDN é o mecanismo de encaminhamento de requisição, pois afeta diretamente diversas questões da CDN como eficiência e qualidade de experiência. Os métodos básicos encontrados na literatura para implementar estes mecanismos são: *DNS*, *HTTP/RTSP Redirect*, *URL Rewrite*, *IP Anycast* e *VIP(A) - Virtual IP (Address)*. Outras técnicas como: *Global Server Load Balancing (GSLB)* e *CDN peering* também são utilizadas, mas internamente utilizam alguns dos métodos supracitados [Buyya, Pathan e Vakali 2008]. Adicionalmente, novos métodos estão sendo pesquisados voltados essencialmente para as novas arquiteturas da Internet (ex: SDN) .

Cada um destes mecanismos possuem vantagens e desvantagens, por exemplo, o mecanismo de *HTTP Redirect* apresenta uma latência maior que o DNS, isto é, o início da transmissão efetiva dos dados demora mais devido ao processo de estabelecimento da conexão HTTP/TCP com mais de um elemento. Por outro lado, o *HTTP Redirect* permite uma escolha mais precisa e personalizada do que o método baseado em DNS, conseqüentemente, possibilita definir um servidor que apresente um menor RTT (Round Trip Time) até o cliente.

Estes fatores têm um efeito muito grande no QoE do usuário ao acessar diferentes tipos de conteúdos, principalmente devido ao efeito do *sliding window* do TCP (que é usado pelo HTTP). Além disso, conforme o mecanismo utilizado, é possível coordenar o acesso aos serviços da CDN de forma mais precisa, refletindo diretamente na eficiência do uso da infraestrutura. Exemplo, através do *HTTP redirect*, é possível integrar a lógica de escolha do servidor de réplica com a rota sobreposta do *multicast overlay* de forma a minimizar as interações necessárias de acesso à fonte do conteúdo [Pimentel et al. 2015].

Os efeitos associados ao QoE do usuário no acesso a diferentes tipos de conteúdo ficam evidentes quando são comparadas a distribuição de um arquivo HTML e a de um vídeo. No caso de uso do mecanismo de DNS, proporciona ao usuário uma melhor percepção do acesso a arquivos pequenos (páginas web), pois o DNS apresenta uma latência menor em relação ao *HTTP redirect*. Por outro lado, no caso do acesso a um vídeo, o seu volume de dados é relativamente maior quando comparado a um arquivo HTML. Assim, um RTT (Round Trip Time) menor permite alcançar taxas de transmissão maiores e mais estáveis em conexões que se mantenham ativas por mais tempo, devido ao controle de fluxo do TCP (janela deslizante). Portanto, o mecanismo de *HTTP Redirect* proporciona melhores resultados de QoE ao usuário nestes cenários.

Concluindo, à vista destes aspectos, é comum que uma CDN utilize uma abordagem que integre mais de um mecanismo a fim de extrair as vantagens de cada um destes.

Controle de QoS

Os mecanismos para controle de QoS tem como objetivo garantir a entrega de pacotes, como fluxos de vídeo e voz, atendendo aos requisitos de cada tipo de tráfego. Estes mecanismos baseiam-se na classificação, priorização de tráfego e/ou reserva de recursos na infraestrutura de rede. Adicionalmente, podem utilizar diferentes protocolos de transmissão baseado nas estratégias de otimização.

Os primeiros serviços de distribuição de vídeo optavam pelo uso do UDP (User Datagram Protocol), visto que o desempenho do protocolo TCP para a transmissão de fluxo contínuo não era satisfatório, principalmente devido ao seu longo “*handshake*” e seu controle de fluxo (*sliding window*). Assim, o uso de um protocolo mais simples, como o UDP era a opção mais adequada para a transmissão de áudio e vídeo. Neste contexto, os protocolos RTP (Real Time Protocol) e RTCP (Real Time Control Protocol) foram especificados para dar suporte a garantia de sincronismo e melhor controle do fluxo na camada de aplicação e do serviço oferecido. Além disso, o uso do IP Multicast permitia otimizar o uso da infraestrutura na transmissão para múltiplos usuários. Adicionalmente, foi definido o RSVP (Resource Reservation Protocol) para garantir que os recursos necessários

para a transmissão do conteúdo estariam disponíveis.

Todavia, o uso destas tecnologias, apesar de adequados para contextos mais restritos, como o serviço de IPTV de um ISP, não foram bem aceitos para o provimento de serviços na Internet de forma ampla. Entre os principais motivos, destacam-se as questões de: segurança, interoperabilidade no IP Multicast, NAT (Network Address Translation) e fatores mercadológicos[Lao et al. 2007].

Desta forma, a solução atual é oferecer fluxos de áudio e vídeo utilizando o protocolo HTTP na camada de aplicação, isto é, retornando ao uso do protocolo TCP na camada de transporte. Como forma de tratar os aspectos de QoS, foram propostas técnicas de *multicast overlay* [El-Sayed 2004] e a aceleração transparente do TCP (Transparent TCP acceleration)[Ladiwala, Ramaswamy e Wolf 2009]. Finalmente, mecanismos de rede para priorização de tráfego também podem ser aplicados como técnicas complementares a fim de prover um serviço com melhor QoE.

10.3.3. Interconexão de CDNs

Inicialmente as CDNs tradicionais montavam uma infraestrutura altamente distribuída, e potencialmente dentro das redes de terceiros, capazes de entregar o conteúdo de seus clientes aos usuários finais em todo o mundo. No entanto, o surgimento das tecnologias de computação em nuvem facilitou a implantação de CDNs construídas em ambientes virtuais. Desta forma, começaram a surgir CDNs instaladas em servidores virtualizados em vários *data centers* de grande porte, nos quais a alocação dos recursos é feita de forma dinâmica, baseada na demanda de acesso. Além disso, emergiram pequenas CDNs para atender cenários específicos, geralmente com um escopo de atuação regional.

As CDN construídas sobre serviços de computação em nuvem têm sido nomeadas de Cloud-CDNs. Esta arquitetura que tem sido recentemente alvo de estudos, procurando fazer uma integração mais transparente e otimizada entre a infraestrutura de borda da CDN e os servidores instanciados “na nuvem”. O principal objetivo desta proposta é diminuir os custos, aumentar a escalabilidade, melhorar a proximidade do conteúdo com o usuário final e melhorar o gerenciamento do consumo do conteúdo para o cliente das CDNs [Lin et al. 2011]. Um exemplo de Cloud-CDN é o serviço de distribuição de conteúdo da Amazon - o CloudFront.

Em vista da grande oportunidade de explorar este tipo de serviço, os ISPs notaram que poderiam aproveitar a sua própria infraestrutura de rede para construir e operar os seus próprios sistemas de distribuição de conteúdo. Este tipo de abordagem tem sido chamado de CDNs operadas por ISPs. Contudo, devido às limitações decorrentes da abrangência da sua própria rede, essas CDNs não são construídas de forma distribuída através de múltiplas redes, e portanto, não operam globalmente. Para superar este problema, diversas iniciativas têm surgido para trabalhar na integração de múltiplas CDNs a fim de formar Federações de CDN. O grande benefício desta estratégia é a possibilidade da redução de custos de implantação, operação e gerenciamento de estruturas adicionais pelas CDNs para atender de forma satisfatória os usuários que não estão conectados diretamente em suas próprias infraestruturas [Frank et al. 2013]. Além disso, este tipo de associação permite expandir a atuação da operadora de rede, permitindo que esta possa agregar maior valor à sua infraestrutura.

Como consequência do surgimento de diversos serviços de CDNs, em que cada serviço pode oferecer benefícios diferentes ou capacidades de atendimento diferenciados, figurou-se um novo modelo, a Meta-CDN [Broberg, Buyya e Tari 2009]. Basicamente, esta atua como um *broker* na seleção de outras CDNs, usando diversos critérios para estimar qual é a melhor CDN para cada caso. Para isso, uma das estratégias é a distribuição de pequenos arquivos com códigos que incorporam o código-fonte de diversos sites populares dentro da rede da CDN, assim, quando os clientes se conectarem a ela, serão geradas estatísticas de acesso, que alimentarão as métricas de desempenho. Apesar da semelhança entre as CDNs Federadas e a Meta-CDN, o grande diferencial é o modo como estas se relacionam. Na estratégia de federação, nota-se um modelo de cooperação planejado na distribuição do conteúdo, por outro lado, com a Meta-CDN a estrutura é hierarquizada, em que o *broker* não participa na distribuição do conteúdo, apenas atua no processo de delegação entre as diversas CDNs, que em princípio, podem nem atuar de forma conjunta na transmissão do conteúdo.

De forma complementar a estes modelos, as CDNs Híbridas com Redes P2P são constituídas de forma em que tanto os computadores dos usuários como os servidores da CDN atuam como *peers* de uma rede P2P. Um dos benefícios desta abordagem é que, de acordo com [Ross et al. 2008], até 80% do tráfego de distribuição de conteúdo pode ser realizado através dos usuários finais, sem degradação significativa no tempo total de download. Além disso, quando comparado com redes P2P puras, a disponibilidade de conteúdo pode ser gerenciada pela CDN, devido à existência de servidores na CDN dedicados para atuarem como *peers*. Alguns exemplos deste tipo de abordagem podem ser vistos em [Aditya et al. 2012], em que a Akamai oferece uma solução de CDN híbrida aos seus clientes. Esta solução tem como objetivo a distribuição de arquivos de grande porte, tais como atualizações de software.

Apesar de existirem vários modelos de funcionamento e implantação de CDNs, estes carecem de padronização na intercomunicação. Com foco neste problema, o IETF propôs o *Framework for Content Distribution Network Interconnection* (CDNI) [Peterson, Davie e R. van Brandenburg 2014] que especifica as interfaces requeridas para a intercomunicação entre duas ou mais CDNs. Através destas interface é possível realizar: (i) inicialização da interconexão; (ii) encaminhamento de requisições, (iii) distribuição de metadados (ex: blocos de endereços atendidos e políticas de controle de acesso) e (iv) informações das requisições atendidas - *logging*. A seção 10.4.2 descreve de forma mais detalhada este protocolo.

10.4. Estado da Arte

A percepção de que o “Conteúdo é o Rei” antevista por Bill Gates [Gates 1996] fica cada vez mais evidente, visto que existem iniciativas que tornam “o Conteúdo” o ponto central da rede [Wang, McGee e Huang 2009] - caso das Redes Centradas na Informação (Information Centric Network - ICN) [Ghodsí et al. 2011]. No entanto, como apontado por Odlyzko em “Content is not King” [Odlyzko 2000], apesar da inegável importância, existem também outros problemas nas redes de computadores que têm motivado diversas pesquisas no sentido de remodelar a Internet [Conti et al. 2011].

Frente a este contexto, apesar do conteúdo não ser “o Soberano”, nota-se a grande

relevância da pesquisa na área das Redes de Conteúdo, pois como demonstrado em [Cisco 2017] este representa uma parcela muito significativa do tráfego total da Internet. A consequência disto é que as pesquisas nessa área terão cada vez mais sinergia com a infraestrutura de rede e computação em nuvem, principalmente na integração com o novo paradigma das redes - a SDN [Passarella 2012].

Como resultado, tem emergido a percepção que o atendimento às demandas de distribuição de conteúdo devem ser tratadas de forma conciliada com as demandas de outros serviços (ex VoIP, email, etc). Além disso, é importante considerar a autonomia do gerenciamento da infraestrutura da rede. Desta forma, tem surgido o entendimento de estruturar a rede em uma abordagem de Redes Centradas em Serviços - SCN (Service-Centric Networking) [Simoens et al. 2017].

Paralelamente, a evolução das tecnologias de virtualização proporcionou um grande ganho no aspecto de otimização de recursos. Isto porque, em vez de adquirir um equipamento que atendesse a um conjunto específico de necessidades, as quais não eram compartilhadas, a virtualização proporcionou a maximização do uso destes recursos. Todavia, esta estratégia não é recente. Ao analisar a evolução das redes de circuitos, passando pela era das redes convergentes, até o momento atual das redes definidas por software, podemos ver que a otimização do uso da infraestrutura converge com o compartilhamento dos recursos entre as diversas demandas.

Atualmente existem diversas tecnologias capazes de prover mecanismos para alocar e configurar dinamicamente o uso dos recursos da infraestrutura, tanto de rede, como de computação. Assim, o principal desafio é o desenvolvimento de sistemas e algoritmos capazes de gerenciar todos estes recursos baseado em uma visão holística [Silveira 2016]. Isto é, mediante a demanda aos serviços disponibilizados e baseado nos estados de uso de cada um dos recursos da infraestrutura disponível, orquestrar todo o sistema visando prover o melhor QoE ao usuário, além de minimizar os custos de operação. Enfim, prover uma Infraestrutura Definida por Software.

Tendo em vista este cenário, diversas pesquisas têm sido desenvolvidas para tratar os desafios existentes, sendo descritos algumas destas a seguir, classificadas de acordo com o foco de atuação.

10.4.1. Interface com o Usuário

O primeiro aspecto a ser abordado é como o usuário interage com o conteúdo, isto é, publica e/ou consome. Neste é considerado tanto os dispositivos pelo qual ele visualiza o conteúdo, como os *codecs* de vídeo, ferramentas de interação e protocolos associados.

Para obter ganhos na qualidade da imagem, com o objetivo de aumentar a sensação de realidade, tem havido grande investimento no aumento da resolução das imagens. O UHD TV (Ultra High Definition TV) é o formato de imagem com no mínimo, 3840 x 2160 pixels, quatro vezes maior que o Full HD e conhecido como 4K, mas também abrange resoluções maiores, sendo no momento o 8K o representante de mais alta resolução da tecnologia com 7680 x 4320 pixels disponível no mercado. Dentro desta perspectiva, verifica-se que atualmente os *smartphones* já não apresentam mais limitações quanto a capacidade de visualização de mídias, seja pela resolução da tela ou pela capacidade

de processamento. Ademais, associados com óculos de realidade virtual, permitem experiências de consumo de vídeos superiores aos oferecidos por outros dispositivos (computadores, televisores, etc...). Diversos exemplos deste tipo de experiência são descritos em [Laver et al. 2012], dentre estes se destacam as inovações nos contextos de educação, desde crianças com necessidades especiais à treinamento de astronautas pela NASA.

Adicionalmente, estes dispositivos também são muito úteis para a geração dos conteúdos. Atualmente, já existem diversos *smartphones* capazes de gerar conteúdos em 4K, de modo que muitas produções de baixo custo podem ser realizadas com este tipo de equipamento. Por outro lado, para a produção de conteúdo *Hi-End*, já há disponibilidade no mercado de câmeras com características como 3D e 360° graus, permitindo gerar conteúdos imersivos de alta qualidade.

Outro aspecto na questão da interface com o usuário são as ferramentas que permitem a interatividade com o vídeo. Ou seja, que possibilite ao usuário interagir com a mídia, provendo diferentes desfechos de um enredo, acessando conteúdos auxiliares ou comunicando com outros usuários, através de aplicativos de vídeocolaboração. Entre os benefícios destas técnicas destaca-se o maior engajamento dos usuário com o conteúdo. Em [Schoeffmann, Hudelist e Huber 2015] é apresentado um levantamento dos trabalhos recentes nesta área e classificadas as diversas pesquisas baseadas em fatores como: o objetivo da interatividade e os meios pela qual esta é feita.

Para dar suporte a estas evoluções, estão emergindo novos *codecs* de vídeo: o H.265/MPEG-HEVC proposto pela ITU-T/ISO e o AV1 desenvolvido pela AOM (*Alliance for Open Media*). Existem diversos fatores que podem ser considerados para comparar estas propostas. Em [Grois, Nguyen e Marpe 2017; Barman e Martini 2017] são analisados aspectos de eficiência e desempenho, que indicam que o H.265 apresenta melhores resultados em relação ao AV1. Contudo, conforme apresentado em [Pozdnyakov 2017], as versões mais recentes do AV1 apresentam resultados equivalentes ao H.265. No entanto, pela perspectiva de adoção da tecnologia pelo mercado, existe uma incógnita, pois o H.265 é a alternativa que representa a evolução da tecnologia padrão de mercado, o H.264, sendo a opção “denominador comum” dos *codecs* suportados pelos navegadores com o uso do HTML5. Em contrapartida, a *Alliance for Open Media* tem entre seus diversos participantes, grandes empresas como: Adobe, Amazon, AMD, ARM, Cisco, Google, Intel, Microsoft, Mozilla, Netflix, NVidia. Além disso, diferentemente do H.265, é *royalty-free*, diferencial bastante atrativo para desenvolvedores de aplicações e dispositivos.

Finalmente, de modo complementar, são necessários protocolos para tratar a transmissão dos conteúdos de forma adequada a fim de permitir tanto o envio com o recebimentos dos vídeos.

Entre os protocolos mais utilizados atualmente, o RTMP (Real Time Messaging Protocol) foi o primeiro a ser desenvolvido. Em 2002, a Macromedia (posteriormente adquirida pela Adobe), lançou o RTMP para atuar juntamente com o Flash. Este era usado tanto para a transmissão propriamente dos fluxos de vídeo, como também de outros objetos associados. Após alguns embates jurídicos de quebra de licença com a Wowza, a Adobe publicou em 2012 a especificação do protocolo de forma livre de *royalties*. Durante alguns anos o RTMP foi utilizado amplamente para a distribuição do conteúdo até o

usuário final. No entanto, atualmente a principal utilização deste protocolo é para a transmissão do fluxo de vídeo do *encoder* para o serviço de distribuição de conteúdo, o qual adapta o conteúdo para o formato a ser entregue aos usuários finais, utilizando protocolos como o HLS (HTTP Live Streaming) e o DASH (Dynamic Adaptive Streaming over HTTP). A importância deste protocolo atualmente é que, apesar de existirem protocolos para função semelhante, como o RTSP (Real Time Streaming Protocol) [Schulzrinne, Rao e Lanphier 1998], o RTMP atende de forma mais “descomplicada” o cenário de envio de fluxos ao vivo. Isto é, devido ao modo *push* de envio do conteúdo (do *encoder* para o servidor), usando TCP, não sofre com problemas de NAT (Network Address Translation). Como resultado, atualmente o RTMP é o principal protocolo para a transmissão de fluxos de vídeo ao vivo para os serviços de distribuição de conteúdo. Todavia, no que se refere ao envio de conteúdos “VoD”, o *upload* via HTTP POST é o protocolo padrão.

Por outro lado, a visualização de vídeos pelo usuário tem sido atualmente predominada pelo uso do HLS[Pantos e May 2012] e do MPEG-DASH[Sodagar 2011].

O HLS, desenvolvido inicialmente pela Apple e posteriormente padronizado pelo IETF, revolucionou o consumo de vídeo na Web, assim como o Adobe Flash há anos atrás. Pois com mecanismo simples e flexível para dar suporte a taxa de bits adaptativa do vídeo, proporcionou uma sensível melhora no QoE do usuário. Além disso, a descontinuidade do suporte ao Flash nos iPhones acelerou o processo de adoção deste padrão. Paralelamente, o grupo MPEG desenvolveu o Dynamic Adaptive Streaming over HTTP, conhecido como MPEG-DASH. Este também é um protocolo para transmissão de mídia na Internet utilizando como base o protocolo HTTP, semelhante ao HLS. Inicialmente sofreu com a falta de suporte nos dispositivos dos usuários, mas este cenário tem mudado, sendo que atualmente o MPEG-DASH tem sido usado tanto pelo YouTube como pelo Netflix. Além disso, a Adobe e a Microsoft também desenvolveram tecnologias concorrentes, respectivamente, o HDS (HTTP Dynamic Stream) e o HSS (HTTP Smooth Streaming).

Em suma, estas tecnologias compartilham o mesmo paradigma: segmentar o fluxo do vídeo em diversos arquivos de diferentes qualidades (MPEG-TS ou MP4), indexados em arquivos de metadados, possibilitando a comutação entre os fluxos. Em outras palavras, fornece informações para o cliente alternar dinamicamente entre *bitrates* diferentes do mesmo vídeo, além de possibilitar alternar os fluxos de áudio (ex: de línguas diferentes) ou a seleção de vídeo entre diferentes ângulos de câmera. Apesar de utilizar mecanismos bem consolidados de transmissão - o HTTP - estas tecnologias têm aberto novas oportunidades de pesquisa na distribuição de conteúdo. Um exemplo interessante é o apresentado em [Mao, Netravali e Alizadeh 2017] no qual aplica um algoritmo de inteligência artificial baseado em aprendizagem por reforço para melhorar o QoE do usuário, mitigando problemas de *rebuffering*.

Finalmente, outro protocolo que tem ganhado muito destaque recentemente é o Web Real-Time Communication (WebRTC), definido pelo W3C (World Wide Web Consortium) e o IETF (Internet Engineering Task Force) [Bergkvist et al. 2012]. Este permite que aplicações Web estabeleçam comunicação em tempo real, inclusive de forma direta entre dois navegadores. Esta possibilidade abre novas oportunidades de uso, flexibilizando e simplificando a troca de informação através de conexões peer-to-peer (P2P).

Outro grande benefício é que permite o acesso à fluxos de áudio e vídeo nativamente, através de uma API (Application Programming Interface) JavaScript, isto é, sem a necessidade de *plugins*. Em setembro de 2016 o WebRTC já era suportado por vários navegadores e ambientes como: Google Chrome, Mozilla Firefox, Opera, Android, iOS. Portanto, cerca de 67% de usuários de navegadores *desktop* podem usar WebRTC. Esta oportunidade permitiu que várias aplicações que estavam baseadas em uma arquitetura cliente-servidor, passassem a usar o WebRTC para que, através de conexões peer-to-peer, ganhassem escalabilidade e eficiência. Um exemplo é o Google Hangouts (aplicação de videoconferência), que passou a usar WebRTC no Chrome desde 2014. Logo depois, a Mozilla anunciou o uso de WebRTC no seu novo programa de videoconferência chamado Hello, nativamente integrado no Firefox. Confirmando esta tendência a favor do WebRTC, o Skype anunciou em julho de 2016 a versão Alpha do seu cliente para Linux, usando WebRTC.

10.4.2. Serviços de Distribuição de Conteúdo

Outro aspecto importante são os serviços de distribuição de conteúdo. Neste sentido, muitos trabalhos tem ressaltado a importância da integração com a infraestrutura da rede. Em [Aggarwal, Feldmann e Scheideler 2007] é proposto que os ISPs ofereçam um “Oráculo” para os clientes das redes P2P, de maneira que a seleção dos *peers* possa ser realizada segundo critérios relacionados diretamente com a topologia física da rede. Assim como este, diversos outros trabalhos têm pesquisado a questão do compartilhamento de informações da topologia de rede pelo ISP, como [Choffnes e Bustamante 2008] e [Poese et al. 2012], no qual foi desenvolvido o PaDIS, um sistema para otimizar os mecanismos de redirecionamento de requisição das CDNs através da monitoração do estado dos enlaces da rede.

Com o objetivo de padronizar estas iniciativas, o IETF especificou o “Application Layer Traffic Optimization” - ALTO [IETF 2008]. Através deste, os provedores de rede (ex: ISP) podem prover informações referentes à topologia de rede para as aplicações, dentre elas, redes P2P e CDNs. Assim, visa fornecer uma visão da infraestrutura de rede conforme a conveniência dos provedores de serviços (“*My Internet View*”). Deste modo, contribui com a tomada de decisões em serviços que possuam mais de uma fonte de conteúdo e operem utilizando o protocolo HTTP disponibilizando recursos *RESTfuls*. Para isso, as informações sobre a infraestrutura de rede podem ser observadas como um conjunto de agregações (*hosts*, sub-rede, conjunto de subredes, área metropolitana) e os custos associados entre estas agregações. O custo definido pelo protocolo é genérico, deixando ao fornecedor de serviço a implementação do mesmo através de políticas internas (banda disponível, atraso, distância, etc). Estas informações compõem o *Mapa da Rede* e o *Mapa de Custos*. Baseado nestes mapas, o protocolo fornece formas de filtragem visando a entrega de informações relevantes para as aplicações. Um exemplo seria a classificação de servidores de réplica de uma CDN baseada no custo de transmissão até o cliente final. A utilização deste protocolo de abstração tem sido amplamente indicado em diversas pesquisas, como apresentadas em [Faigl, Szabo e Schulcz 2014; Gurbani et al. 2012; Ahmad et al. 2016; Xie et al. 2012; Arumathurai et al. 2013]. Um dos benefícios deste é que o modelo de dados definido é bem flexível, atendendo às diversas necessidades, inclusive facilitando a definição de extensões para atender mais cenários, como a

proposta apresentada em [Kopp et al. 2017]. Adicionalmente, estão sendo realizadas pesquisas para a interação entre as CDNs, pois apesar de existirem vários modelos de funcionamento e implantação de CDNs, estas carecem de padronização na intercomunicação. Com foco neste problema, o IETF propôs um framework [Peterson, Davie e R. van Brandenburg 2014] que especifica as 4 interfaces requeridas para a intercomunicação entre duas ou mais CDNs.

- A interface de controle (*Control interface - CI*) deve fornecer métodos que viabilizem o gerenciamento da intercomunicação entre as CDNs, através do qual é feita a inicialização da interconexão e a parametrização das demais interfaces. Adicionalmente, também contempla mecanismos para a publicação/invalidação de conteúdos e seus metadados.
- A interface de encaminhamento de requisições (*Request Routing interface*) é dividida em duas subinterfaces, que apesar de oferecer operações distintas, são correlacionadas. A primeira (*Footprint and Capabilities Advertisement interface - FCI*) possibilita que a CDN *downstream* (dCDN) exponha a sua abrangência de atuação e capacidades de atendimento em um processo assíncrono à requisição do conteúdo. Esta característica contribui na escolha de uma dCDN por parte da CDN *upstream* (uCDN). A segunda (*Request Routing Redirection - RI*) viabiliza o redirecionamento da requisição realizado pela uCDN para a dCDN, isto é, de forma síncrona com a requisição do usuário. Os dois principais mecanismos para a realização deste direcionamento são: via DNS ou redirecionamento na camada de aplicação (ex: *HTTP Redirect*). Um aspecto importante em relação a este redirecionamento é que pode ser realizado de forma interativa ou recursiva. Na interativa, a uCDN direciona o usuário a uma dCDN, que por sua vez irá tratar a requisição do usuário direcionando para um servidor apropriado ou reencaminhando para outra dCDN. Por outro lado, no modo recursivo, a uCDN irá consultar o sistema de encaminhamento de requisição da dCDN para eleger o servidor que irá atender ao usuário (ou requisitar novamente a um outro dCDN recursivamente), e ao receber a resposta, a uCDN direciona o usuário ao servidor definido.
- A interface de metadados (*Metadata interface - MI*) é responsável por viabilizar a distribuição de metadados feita pela uCDN até a dCDN. Os metadados transportados podem ser: restrições sobre blocos de endereços atendidos, políticas de controle de acesso, dentre outras. Esta interface pode ser utilizada para facilitar o acesso ao conteúdo da dCDN através das informações de autorização requeridas, assim como o fornecimento de várias fontes de conteúdo.
- A interface de logging (*Logging interface - LI*) é necessária para que a uCDN tenha acesso às informações dos conteúdos atendidos que foram redirecionados para uma dCDN. Com estas informações a uCDN pode contabilizar tanto o consumo de conteúdo servido pela dCDN, assim como elaborar relatórios sobre as estatísticas de volume do tráfego gerado pelos fornecedores de conteúdo.

Finalmente, outra linha de pesquisa explora a capacidade de realizar de forma nativa (com o uso do OpenFlow) operações na rede para otimizar a distribuição do

conteúdo. A iniciativa descrita em [Chanda, Westphal e Raychaudhuri 2013] investiga a implementação de mecanismos para a engenharia de tráfego na distribuição de conteúdo (como vídeo sob demanda), baseada em Redes Centradas na Informação (*Information Centric Network* - ICN) [Ghodsí et al. 2011]. Em [Woo et al. 2014] por sua vez, também explora a otimização da distribuição de conteúdo associada a SDN através das Redes de Dados Nomeada (*Named Data Networking* - NDN) [Afanasyev et al. 2014].

Adicionalmente, nota-se a crescente busca de estratégias para implementar mecanismos de controle de QoS na rede, de modo que as aplicações informem os requisitos necessários para a reserva de recursos para tratar tais requisições. Nesta linha, podemos destacar o trabalho apresentado em [Egilmez et al. 2012], que propõe um arcabouço em que um serviço de distribuição de vídeo pode requisitar o atendimento à determinados parâmetros de QoS, que irão resultar na definição de regras *OpenFlow* para condicionar o tráfego dos pacotes mediante o SLA estabelecido. Egilmez [Egilmez et al. 2012] apresenta o OpenQoS, que integrado ao controlador *OpenFlow*, utiliza um mecanismo de diferenciação de tráfego que permite tratar dinamicamente, de forma diferenciada, os fluxos de vídeo escolhendo a rota que mais se adequa aos parâmetros de QoS requisitados. Ademais, em Diorio [Diorio e Timóteo 2016] é apresentado uma abordagem semelhante, mas com o benefício de que não é necessário realizar modificações no controlador SDN, pois estas operações são realizadas através de um *gateway multimídia*.

10.4.3. Infraestrutura Definida por Software

Por fim, o último aspecto deste panorama é a infraestrutura para dar suporte à estes serviços. Para alcançar este objetivo é importante ter em vista quais são os principais elementos que compõem este ecossistema complexo: recursos de computação, funções de rede e conectividade. Contudo, é fundamental considerar como estes elementos são cada vez mais interdependentes, por exemplo na SDN. Sem os recursos de computação não é possível prover conectividade, nem implementar as funções de rede.

Recursos de Computação

O termo *computação em nuvem* pode ser entendido como um sistema de alocação de recursos de computação (processamento, memória e armazenamento) para a implantação de serviços que atendem as demandas dos usuários - sejam estes indivíduos ou outros sistemas. No entanto, para atender adequadamente esta demanda, é necessária uma arquitetura escalável associada a uma infraestrutura robusta. Além disso, é notável a necessidade de alocação de recursos adicionais para atender adequadamente quando ocorrem picos de demanda. Para isso, a equipe de operações tem sob sua responsabilidade um número considerável de ações que devem ser realizadas para que este dimensionamento seja realizado de forma estável, como: monitoração, alocação e reconfiguração de novos recursos [Jain e Paul 2013].

Para tratar esta questão, surgiu recentemente o conceito de serviços elásticos [Szabo et al. 2015], que propõe um modelo de gestão de recursos mais eficiente, flexibilizando o uso da infraestrutura computacional e automatizando os processos de configuração, diminuindo custos de operação. Segundo [Dustdar et al. 2011] o conceito

de elasticidade pode ser aplicado não só para provisionamento de recursos computacionais, mas também para qualidade de serviço (QoS) e custo da oferta de um serviço. A ideia central é dinamizar e automatizar estas operações de forma que a continuidade do serviço não dependa tanto de ações de profissionais e ainda traga diminuição de custos com infraestrutura.

Neste sentido, é apresentado em [Um et al. 2014] uma proposta de arcabouço para implementar uma CDN Virtual baseada em computação em nuvem. Entre os principais aspectos, é a definição do *CloudBroker* que permite a alocação de recursos computacionais em diversos provedores de computação em nuvem. Desta forma, de acordo com a demanda de serviços, pode ser alocado mais recursos nos *datacenters* que permitem atender, de forma mais adequada, as novas demandas. Outra função deste é coletar o estado de monitoração dos servidores em execução, e aplicando um algoritmo de alocação de recursos, identificar qual localidade precisa instanciar novas máquinas virtuais.

Tendo em vista as tecnologias envolvidas neste contexto, destaca-se o papel dos orquestradores de nuvem. A orquestração permite um melhor gerenciamento tanto de recursos físicos quanto de serviços, permitindo além da otimização dos recursos oferecidos, a adequação aos requisitos cada vez mais dinâmicos aos quais os serviços estão sendo submetidos atualmente. Como mostrado em [Bousselmi, Brahmí e Gammoudi 2014] e [Tosatto, Ruiu e Attanasio 2015] podemos classificar os sistemas de orquestração em dois tipos, de *software (SaaS)* e de *hardware (IaaS)*.

- A *orquestração de serviços de software (SaaS)* realiza o gerenciamento de processos de negócio que interagem com outros serviços de *SaaS*, além de estarem associados diretamente com os recursos físicos onde serão executados dentro do provedor de nuvem.
- A *orquestração de recursos de hardware (IaaS)* realiza o gerenciamento de recursos computacionais (i.e., memória, CPU, armazenamento, rede) com o objetivo de atender aos requisitos de usuários ou cumprir as necessidades computacionais do provedor de nuvem.

Estes sistemas, em sua maioria, realizam o gerenciamento e o provisionamento de recursos em conjunto com o sistema do provedor da nuvem. Assim, permite que sejam alocadas máquinas virtuais em provedores de recursos computacionais, com configurações pré-definidas, de acordo com a necessidade de um determinado serviço. Esta integração permite a otimização no uso dos recursos providos pela nuvem, além de permitir uma elasticidade ao serviço, através da definição de custos relacionados ao uso dos recursos da nuvem.

Funções de Rede

O modelo de operação das redes é hoje em dia muito diferente dos modelos teóricos em que há relação direta entre topologia física e lógica. Isto acontece porque a organização lógica das redes refletem os objetivos de comunicação, enquanto a topologia física é construída de forma a otimizar os custos, a partir do compartilhamento de recursos.

Nesta nova organização, em que existe um número muito grande de funções de rede, os equipamentos se tornam muito complexos e de alto custo. Tendo em vista este cenário e a dinamicidade da rede, em conjunto com a facilidade de utilização de recursos da computação em nuvem, foi proposto a virtualização das funções da rede (NFV – *Network Function Virtualization*). Através desta é possível implantar as funções de rede por meio de software que são executados sobre uma infraestrutura denominada NFVI [ETSI 2013]. Assim, estas funções podem ser instanciadas em vários locais da rede dependendo da conveniência, sem a necessidade da instalação de novo equipamento.

O NFV é bastante complementar ao SDN, mas não são necessariamente dependentes entre si [Bhaumik et al. 2014]. Apesar de parecerem conceitos semelhantes, NFV é a virtualização das funções de rede, enquanto SDN é a separação das funções de controle e encaminhamento. Portanto, o SDN pode facilitar a implementação de NFV, provendo o fatiamento da camada física e a virtualização da rede, o que facilita a operação e manutenção dos procedimentos da NFV. Desta forma, é fácil verificar que a combinação das duas soluções pode potencializar os benefícios. Considera-se que uma implementação NFV para alcançar o sucesso deve suportar as seguintes características: virtualização, abstração, programabilidade e orquestração. A seguir é detalhada a arquitetura de uma infraestrutura para NFV, ilustrada na figura 10.2.

- As funções de rede virtualizadas (VNFs – Virtual Network Functions): são as várias implementações de software das funções de rede para diferentes propósitos que são executados sobre a NFVI;
- A Infraestrutura de Virtualização de Funções de Rede (NFVI – Network Functions Virtualization Infrastructure): que inclui diversos recursos de hardware e a virtualização de cada um deles;
- Gerenciamento e orquestração NFV (MANO – Management and Orchestration): que inclui o gerenciamento e orquestração de todos os módulos, tanto físicos como de software que dão suporte à virtualização.

O desacoplamento entre hardware e software na virtualização das funções de rede é realizado pela camada de virtualização, que abstrai os recursos de hardware da NFVI. As funções virtualizadas rodam sobre a camada de virtualização, que faz parte da NFVI, e pode envolver recursos de computação, armazenamento e rede. Além disso, as VNFs podem ser implementadas de forma independente ou como cadeia de VNFs, mantendo conectividade entre si para executar funções correlacionadas. Esse é o caso de funções associadas ao serviço Web, como: firewall, NAT e balanceamento de carga. Estas cadeias de funções são representadas como um grafo de encaminhamento de funções de rede virtualizadas, ou uma VNF-FG (Virtualized Network Function - Forwarding Graph). Um exemplo interessante de uso desta estratégia é o trabalho apresentado em [Herbaut et al. 2017]. Neste, é proposto a vCDN, uma implementação de CDNs como uma Função de Rede Virtual (*Virtual Network Function* - VNF). Além disso, propõe um conjunto de passos para implementar e operar este serviço de forma a ajustar dinamicamente os diversos elementos desta infraestrutura.

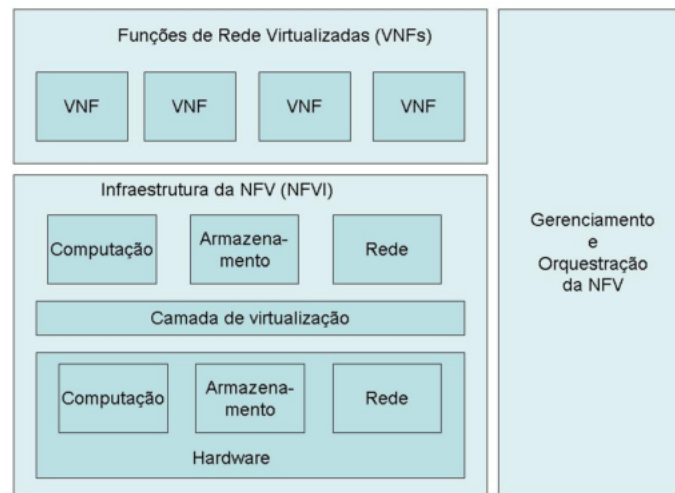


Figura 10.2. Visão de alto nível da arquitetura para NFV (adaptada de [Quittek et al. 2014])

Conectividade

Finalmente, a conectividade desta infraestrutura pode ser entendida como a parte deste ecossistema que permite o tráfego de dados entre os demais elementos. De forma simplista poderia ser considerado apenas como “a rede”, todavia, conforme apresentado, os elementos necessários para realizar esta função não são restritos apenas ao roteadores e *switches*, pois é composta também pelos controladores SDN, as quais são aplicações executadas nos recursos computacionais da infraestrutura.

A ideia fundamental do conceito das SDNs é a separação entre o plano de controle e a infraestrutura de rede, ou plano de dados. Esta separação significa que as funções dos dispositivos de rede estão simplificadas ao encaminhamento de pacotes e disponibilização de estatísticas. A inteligência da rede é logicamente centralizada nos controladores SDN. Como resultado, cada um dos comutadores da rede, que podem ser de diferentes fabricantes, possuem um único ponto lógico de gerenciamento.

O paradigma SDN simplifica as operações destes comutadores, pois não é necessário compreender e processar diversos protocolos, senão apenas receber instruções dos controladores SDN [ONF 2012]. Assim, centralizando o estado da rede no plano de controle, a SDN fornece flexibilidade para configurar, gerenciar, criar mecanismos de segurança e otimizar os recursos da rede através de aplicações de *software*. Consequentemente, possibilita que a infraestrutura de rede seja usada de forma mais inteligente, pois permite que o funcionamento da rede seja programado através de uma visão holística. Além disso, permite que as aplicações possam ter acesso à informações relevantes a fim de aperfeiçoar funcionalidades de Qualidade de Serviço (QoS) e Engenharia de Tráfego (TE). No entanto, atualmente o grande desafio está no desenvolvimento de abstrações adequadas para que estas possam atender a um grande número de aplicações de classes distintas.

O Intents [Han et al. 2016] surge neste contexto como um modelo arquitetural para SDN que define uma forma de relacionamento entre as aplicações e a infraestrutura

da rede, porém, mantendo uma abstração dos mecanismos de rede para as aplicações. Diversos projetos têm trabalhado no sentido de definir uma camada de abstração a fim de permitir que as aplicações possam informar as suas necessidades de uso da rede, sem que esta exponha os detalhes de sua infraestrutura.

Dentre estes projetos podemos destacar o trabalho realizado por Pham [Pham e Hoang 2016] que define interfaces no controlador SDN com foco na semântica e em modelos de dados. Este provê às aplicações uma interface simples de acesso à rede para configurar a mesma de acordo com suas necessidades. Outro projeto importante que também explora esta abordagem é o OpenStack, que permite a definição de um grupo de políticas, denominado Group Based Policy (GBP), como uma coleção de *endpoints* que possuem políticas semelhantes e que podem compartilhar e reutilizar as mesmas regras de controle. O grupo OpenDayLight possui dois projetos neste sentido: o Intent-based Network Modelling (IBNEMO) [Zhang et al. 2016], que define uma linguagem estruturada para descrição de cenários de rede mais simples e voltada ao usuário; e o Network Intent Composition (NIC) [OpenDayLight 2015], que tem a visão do Intent como um serviço de protocolo de roteamento. Ademais, o projeto do ONOS também especificou um *framework*, o ONOS Intents [ON.LAB 2016], para expressar o “desejo” sobre o comportamento da rede na forma de políticas ao invés de descrever explicitamente como os mecanismos devem implementá-la.

Outra linha de pesquisa são os trabalhos voltados à integração dos serviços oferecidos aos usuários e a infraestrutura de rede. Dentre estes, destaca-se o trabalho apresentado em [Simoens et al. 2017], que introduz o conceito de Redes Centradas em Serviços (Service-Centric Networking) como forma de prover uma perspectiva de integração entre os serviços oferecidos, como CDNs e a infraestrutura de rede a fim de usufruir dos benefícios das SDNs.

Para dar suporte a estas operações, o protocolo mais utilizado atualmente para que o controlador da SDN orquestre os equipamentos de comutação de pacotes é o OpenFlow. Em sua arquitetura, o controlador da SDN é responsável por: (i) reconhecer os fluxos de pacotes, (ii) determinar uma ou várias ações sobre eles e (iii) atualizar as tabelas de fluxos nos comutadores. Além disso, fornece um conjunto de estatísticas que permitem que o controlador conheça a topologia e o estado dos nós da rede (*switches*) em tempo real, possibilitando o tratamento dinâmico de fluxos. Contudo, a evolução deste protocolo tem tornado-o complexo. Desde a criação do protocolo OpenFlow, a lista de cabeçalhos suportados aumentou de 12 para 41 (em outubro de 2013) cobrindo mais de 40 protocolos. Assim, vários trabalhos foram propostos visando contribuir na especificação do que seria OpenFlow 2.0. O P4 (*Programming Protocol-Independent Packet Processors*) é uma proposta alternativa, na qual é especificada uma linguagem de programação que fornece uma interface aberta e flexível para o reconhecimento de pacotes e processamento dos cabeçalhos. Através do P4, é possível modificar a forma em que os comutadores processam os pacotes sem considerar o *hardware* utilizado [Bosshart et al. 2014].

Dentre os controladores, os mais difundidos são o *OpenDaylight* (ODL) [OpenDaylight 2013] e o *Open Network Operating System* (ONOS) [Berde et al. 2014]. O ODL é um projeto *open source* criado pela Linux Foundation que fornece um *framework* para implementar serviços de controle e gerenciamento de forma centralizada. Este é subsidi-

ado por grandes empresas como: CISCO, Brocade, Intel, Ericsson, Microsoft e IBM. Por sua vez, ONOS é o primeiro controlador SDN de *open source* direcionado especialmente para os fornecedores de serviços. O desenvolvimento deste controlador visa garantir altos níveis de disponibilidade, desempenho e escalabilidade. Além disso, outras plataformas tem sido desenvolvidas para atender a necessidade por controladores capazes de implementar funcionalidades específicas de forma ágil. O projeto brasileiro Kytos, patrocinado pela Huawei, é um exemplo. Este desenvolveu uma plataforma *open source* para SDN com o propósito de facilitar a sua instalação e utilização, bem como a criação e desenvolvimento de aplicações de rede utilizando o conceito de NApps (Network Applications). Uma característica importante é de não ser vinculada a um protocolo SDN específico, havendo a possibilidade de estender o suporte a outros protocolos, como o P4.

10.5. Infraestruturas em Operação

Esta seção tem como objetivo apresentar uma visão geral das infraestruturas atualmente em operação vinculadas a este tema. Assim, inicialmente são descritos algumas das principais CDNs as quais oferecem o serviço de distribuição de conteúdo. Em seguida, são retratadas algumas infraestruturas de rede em operação que estão evoluindo no sentido de criar Infraestruturas Definidas por Software.

10.5.1. CDNs

A maioria das CDNs em operação atualmente são desenvolvidas por empresas com fins comerciais e foram se consolidando ao longo do tempo com aquisições e expansões. A seguir, são descritas algumas das principais CDNs em operação.

CDNs Comerciais

A Akamai é considerada a líder de mercado no serviço de distribuição através de CDN de uso geral, ou seja, para distribuição de conteúdos estáticos (como textos, documentos, imagens e páginas web) e dinâmicos (como animações, scripts e DHTML), como também mídias contínuas como áudio e vídeo. A primeira versão da CDN da Akamai foi criada em 1998 e atualmente cerca de 30% do tráfego da Internet utiliza o serviço da Akamai.

A Akamai adota a estratégia “*enter deep*”, colocando cluster de servidores de acesso dentro das redes de acesso dos ISPs por todo o mundo, com o objetivo de conseguir proximidade com o usuário final, e consequentemente acesso mais eficiente, com melhoria no atraso e na taxa de transferência.

Segundo Nygren [Nygren et al. 2010] a rede da Akamai pode ser vista como um conjunto de redes, cada uma delas otimizada para a entrega de um tipo de conteúdo. Estas redes, do ponto de vista macro, compartilham uma arquitetura similar, mas um olhar mais detalhado mostra diferenças de implementações e tecnologias que permitem adequar a necessidade de entrega, seja de páginas estáticas, fluxos de mídias ou aplicações dinâmicas. A elaboração do sistema considerou questões relevantes como confiabilidade, garantindo alto grau de redundância para recuperação rápida de falhas, escalabilidade, distribuindo um grande número de servidores geograficamente o mais perto possível dos usuários finais.

Os servidores Akamai se comunicam através de uma rede *overlay* estabelecida entre eles de forma a acelerar a recuperação de conteúdos e minimizar os efeitos de perda de pacotes [Su et al. 2009]. Para fazer o balanceamento de carga, a Akamai monitora continuamente o estado do serviço e seus servidores e rede. Para tanto, são usados agentes que simulam continuamente o comportamento dos usuários distribuídos geograficamente, fazendo downloading de conteúdos Web e medindo as taxas de falha e os tempos da transmissão. Estas informações são utilizadas para monitorar o desempenho de todo o sistema e automaticamente detectar problemas em centro de dados e servidores [Buyya, Pathan e Vakali 2008]. Cada servidor de conteúdo reporta sua carga para a aplicação de monitoração que agrega todas as informações e a repassa para o servidor DNS local, e a partir desta informação o DNS pode definir para qual servidor deve ser direcionada a solicitação. Em caso de um servidor muito carregado, outro servidor é alocado para responder as solicitações, diluindo problemas de sobrecarga.

A LimeLight, fundada em 2001, atualmente faz a distribuição de serviços Web e conteúdos digitais como fluxos de vídeo e música, arquivos de jogos, *software* e interação com redes sociais. Nesta, os provedores de conteúdo podem depositar seus conteúdos diretamente em um servidor da LimeLight ou utilizar seu servidor próprio, conectado diretamente a rede de borda da LimeLight. Quando o conteúdo é solicitado por um usuário final, o sistema da LimeLight localiza o conteúdo e o distribui para um ou mais clusters de servidores, alimentando os servidores espalhados geograficamente. O conteúdo é entregue ao usuário através de um ISP.

A estratégia da LimeLight, conhecida como “*bring home*”, é construir cluster enormes em um menor número de lugares-chave e conectando-os a uma rede privada de alta velocidade.

Estes clusters normalmente são estabelecidos próximos a PoPs de ISPs importantes, o que permite fazer uma conexão eficiente com estes. Além disso, utiliza o redirecionamento através de DNS para direcionar as solicitações dos clientes para o cluster mais próximo. Para isso, usa um mecanismo de mapeamento utilizando BGP associado às informações de sua monitoração através de *traceroutes* de vários pontos da rede.

CDNs privadas

Recentemente, grandes produtores de aplicações e conteúdo têm desenvolvido suas próprias CDNs, pois algumas aplicações podem gerar tanto tráfego, que torna-se mais adequado criar uma CDN específica para atender estas demandas.

Um exemplo deste caso é o Google, que implantou um conjunto de Datacenters interligados com redes de backbone de alta velocidade, além de conecta-los com um grande número de ISPs via IXPs ou peering privado. Adicionalmente, também lançou o Google Global Cache (GGC), que pode ser instalado dentro das redes dos ISP, reduzindo o custo do trânsito de pequenos ISPs e daqueles que estão localizados em áreas com conectividade limitada, por exemplo, na África. Os servidores GGC são fornecidos sem custo para os ISPs que desejarem instalá-los e mantê-los. GGC também permite que um ISP anuncie através de BGP os prefixos de usuários que cada servidor GGC deve atender.

Para suportar seus diversos serviços a Google implementou uma CDN privada em uma arquitetura de três camadas de agrupamento de servidores [Kurose e Ross 2012]: (i) Datacenters primários, responsáveis por servir conteúdos dinâmicos incluindo resultado de busca e mensagens de email; (ii) Dezenas de clusters do tipo “*bring home*”, em localidades distribuídas ao redor do mundo, normalmente próximo a múltiplos PPTs de ISP camada 1. Estes agrupamentos são responsáveis por servir conteúdo estáticos, incluindo os vídeos do YouTube; (iii) Centenas de cluster do tipo “*enter deep*”, localizados em ISPs de acesso, os quais implementam mecanismos de *aceleração transparente do TCP* e servem conteúdos estáticos, incluindo partes estáticas das páginas da Internet dos resultados de busca.

De forma semelhante, o Netflix que inicialmente contratava serviços CDNs para realizar a distribuição de vídeo aos seus clientes, desde junho de 2012 utiliza a sua própria CDN, denominada Open Connect Network. Esta oferece uma interface em que os provedores de internet podem anunciar, via BGP, suas preferências a respeito de que subredes devem participar da distribuição do conteúdo do próprio Netflix.

Através do Open Connect, os ISPs podem escolher se conectar diretamente com a Netflix, ou podem instalar um dos servidores Open Connect da Netflix em sua própria rede. Isso lhes permite armazenar em cache o conteúdo localmente para que ele não tenha que ser transferidos através da rede sempre que for solicitado, trazendo economia em recursos de rede, diminuindo o tráfego gerado pelos clientes Netflix. Para o Netflix isso significa um menor custo de entrega, e uma melhor experiência geral do usuário.

10.5.2. SDNs

Para dar suporte às novas demandas de infraestrutura, inclusive para os serviços de distribuição de conteúdo, muitas instituições (comerciais e de pesquisa) têm iniciado a implementação das redes definidas por software em suas infraestruturas.

Comerciais

Um das infraestruturas de maior destaque é a infraestrutura de rede da Google. Esta é responsável por aproximadamente 25% do tráfego da Internet [Vahdat e Koley 2017]. Motivados pela necessidade de possuir maior controle sobre os diversos serviços que a empresa provê, assim como da infraestrutura que os suporta, foi projetada uma estratégia de implantação do paradigma SDN que contempla quatro pilares observados na Figura 10.3. Cada um deles são responsáveis por atender demandas específicas de QoS assim como de escalabilidade, resiliência e disponibilidade.

O gerenciamento eficiente do grande volume de dados trafegados nos *datacenters* da Google em termos de configuração, distribuição balanceada de tráfego e custos financeiros, demandou a transformação gradual dos equipamentos usados nestes datacenters, assim como o conjunto de protocolos utilizados na camada de controle, que não possuíam um suporte adequado para encaminhamento considerando múltiplos caminhos de igual custo [Singh et al. 2016]. Como consequência destas transformações, foram desenvolvidas diversas versões de especificações (i.e., Four-Post CRs, Firehose 1.0, Firehose 1.1, Watchtower, Saturn), até que a versão final foi o projeto *Jupiter*. Com esta inovadora

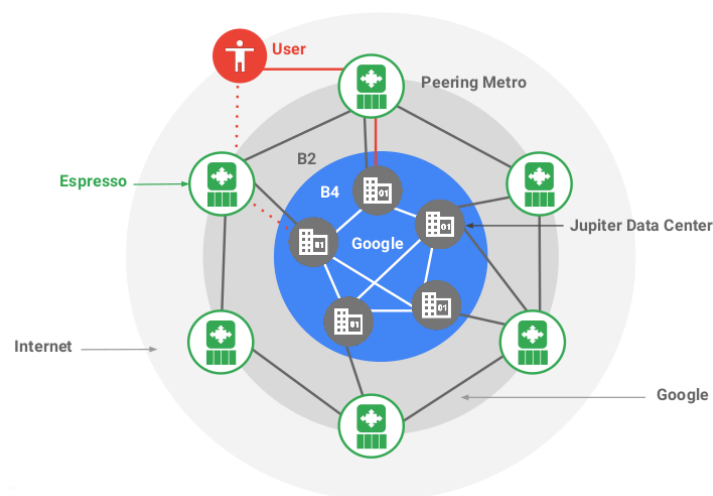


Figura 10.3. Ecossistema SDN da Google [Vahdat e Koley 2017].

forma de conceitualizar um *datacenter*, Jupiter é capaz de suportar mais de 100000 servidores e 1.3 Pb/s de largura de banda total no fornecimento dos serviços da empresa, sendo este o primeiro pilar.

O *B4* [Jain, Kumar et al. 2013], segundo pilar na pilha SDN projetado pela Google, visa prover níveis de escalabilidade, tolerância à falhas, eficiência e controle necessários que não podiam ser atendidos através da arquitetura tradicional de uma rede de longa distância. Desenvolvido para fornecer engenharia de tráfego de forma centralizada, *B4* aumentou os níveis de utilização dos enlaces que interconectam os *datacenters* para aproximadamente 100% do tempo, sendo capaz de manter estes níveis por períodos prolongados. Na Open Network Summit [Google 2012], foram apresentadas as experiências adquiridas na implantação de *B4* e na aplicação do conceito e arquitetura de SDN.

O terceiro pilar é o *Andromeda*, responsável por prover uma maior escalabilidade dos serviços através de NfV. Este visa fornecer as mesmas funcionalidades das aplicações nativas, mas em forma de *containers* e máquinas virtuais executados na nuvem. Além disso permite o isolamento das operações e recursos alocados pelos diversos usuários que utilizam a nuvem da empresa.

Para aumentar a velocidade de atendimento dos usuários finais, foi definido o *Espresso*, como o quarto pilar da estratégia SDN da empresa Google. Enquanto *B4* desempenha a função de interconectar os *datacenters* garantindo a alocação de recursos de acordo com as demandas de diversas aplicações, *Espresso* estende o paradigma SDN até a borda da rede, na qual é delegado o atendimento dos usuários aos provedores de internet externos, como pode ser observado na Figura 10.3. Os mecanismos utilizados no *Espresso* permitem a Google determinar dinamicamente desde qual ponto do planeta é mais apropriado para servir um usuário baseado em medições realizadas em tempo real.

A Microsoft por sua vez, para implantar o paradigma SDN, projeta, desenvolve e opera *datacenters* de escala mundial para o oferecimento de serviços como *Microsoft Azure*. Para isso, desenvolveu duas tecnologias: SmartNICs e Virtual Filtering Platform

(VFP). Com SmartNICs, obtêm os benefícios de programabilidade da SDN e o performance de *hardwares* dedicados usando Field Programmable Gate Arrays (FPGAs) integrados aos servidores da nuvem [Firestone 2017a]. VFP por sua vez é um comutador de pacotes programável que expõe interfaces de fácil programação para os agentes da rede que executam as tarefas delegadas pelos controladores da rede. Microsoft utiliza VFP para abstrair os *hosts* Hyper-V visando prover funcionalidades SDN no núcleo dos serviços de rede do Azure [Firestone 2017b].

De forma semelhante, Facebook busca, com a adoção do paradigma SDN, aumentar a escalabilidade e programabilidade dos *datacenters* que ela gerencia. No ano 2013 Facebook anunciou o *Open Compute Project* (OCP) focado na especificação de um comutador que pode ser utilizado em diversos sistemas operacionais, diminuindo dessa forma a dependência que os operadores dos *datacenters* tem atualmente com os diversos *softwares* distribuídos nos equipamentos de diferentes fabricantes.

O OCP visa a separação entre o *software* e *hardware* da rede, e neste sentido foi desenvolvido FBOSS, um conjunto de aplicações que são executadas dentro de um ambiente Linux mas não atrelada a uma distribuição específica. FBOSS possui as aplicações necessárias para tratar diretamente com o *application-specific integrated circuit* (ASIC) de encaminhamento assim como para automatizar e configurar os comutadores de forma remota através de APIs. Outras aplicações são as responsáveis pela configuração da tabela de encaminhamento e a aplicação de políticas de rede.

Pesquisa

As redes orientadas a pesquisa e educação (RENs) historicamente têm fornecido infraestrutura e serviços de produção com grande qualidade para as instituições de ensino e pesquisa, enquanto mantém a liderança em termos de inovação nas redes. Nesse contexto as RENs são um ambiente ideal para a implantação de SDN, possibilitando a inovação nas redes enquanto são mantidas as redes de produção tradicionais.

Nos últimos três anos uma implantação global de SDN está sendo realizada por diversas RENs, como pode ser observado na Figura 10.4. Utilizando o controlador ONOS e algumas de suas aplicações, foi possível criar uma rede com conectividade L2, L3, na qual cada REN atua como um sistema autônomo administrativamente isolado, que troca rotas BGP's com os outros RENs participantes.

Neste experimento, o SDN-IP é a aplicação mais amplamente utilizada na implantação global de SDN. Esta aplicação foi criada com o intuito de facilitar a migração gradual para SDN, provendo conectividade L2, L3 entre sistemas autônomos sem a necessidade de utilizar equipamentos legados no núcleo da rede. A tabela 10.1 detalha as redes participantes deste experimento.

Americas Lightpaths (AmLight) é um projeto do programa IRNC (International Research Network Connections) que, desde a sua concepção no ano 2012, visa apoiar a pesquisa científica e a educação entre os Estados Unidos e América Latina. Com a colaboração de instituições como a Universidade Internacional da Florida (FIU), a Rede Acadêmica de São Paulo (ANSP) e a Rede Nacional de Ensino e Pesquisa (RNP), Am-

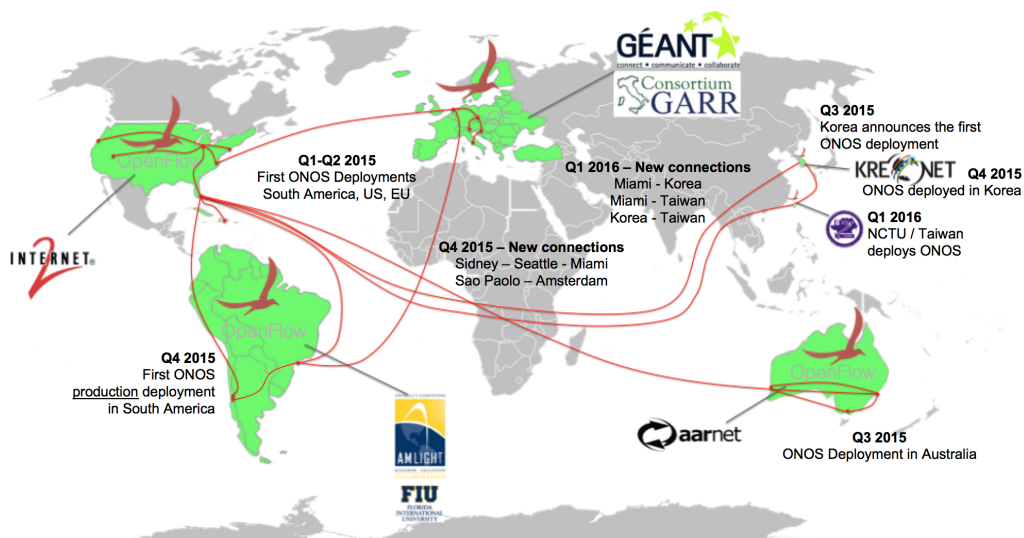


Figura 10.4. ONOS/SDN-IP [ON.Lab 2017].

Tabela 10.1. Participantes na implantação global de SDN através de ONOS-SDN-IP

Pais/Região	Rede
Austrália	AARNet
Brazil	RNP e ANSP através de AmLight
Caribe	CKLN através de AmLight
Chile	Reuna através de AmLight
Estados Unidos	Clemson University, Duke University, Florida International University, Georgia Tech, Indiana GigaPoP, University of Huston, University of Maryland, University of Oklahoma at Norman, University of Utah - através de Internet2
Europa	GEANT
Itália	CREATE-NET Università Roma Tor Vergata (GARR e GEANT)
Korea	KREONET
Taiwan	National Chiao Tung University - NCTU

Light opera múltiplos enlaces internacionais que conectam redes de pesquisa e educação dos Estados Unidos com outras redes similares em Latino América como pode ser visto na Figura 10.5.

Com o intuito de simplificar os processos de provisionamento e gerenciamento da rede assim como de reduzir os esforços de coordenação entre diferentes operadores desde o ano 2014 foi adotado o paradigma SDN/Openflow [Ibarra et al. 2015]. A utilização do orquestrador de rede Exchange Software Suite (OESS) com visão global da infraestrutura possibilitou o provisionamento de circuitos locais e inter-domínios através de On-demand Secure Circuits and Advance Reservation System (OSCARS) sem precisar de prévias operações de coordenação entre as instituições envolvidas na alocação dos recursos.

Com relação a aspectos de programabilidade, a rede AmLight, através de SDN, provê aos pesquisadores um ambiente de experimentação no qual podem ser implantadas



Figura 10.5. Topologia de rede da AmLight [Ibarra et al. 2015].

aplicações de rede com as mais diversas particularidades e requisitos de QoS sem isto afetar outros experimentos ou a rede de produção. Com o uso de Internet2's FlowSpace Firewall (FSF) a AmLight possibilitou o particionamento da rede garantindo uma rede virtual dedicada (*slice*), na qual um usuário pode desenvolver experimentos usando portas e faixas VLAN específicas.

No contexto nacional, destaca-se o FIBRE (Future Internet Brazilian environment for Experimentation) [Sallent et al. 2012]. Este é uma infraestrutura de pesquisa focada em experimentação disponível para o uso de estudantes e pesquisadores no Brasil. Um de seus maiores desafios é oferecer uma visão unificada de recursos geridos por diferentes organizações, por meio de uma federação de *testbeds* independentes.

Essa federação é gerenciada através do Centro de Operações do FIBRE (NOC), operado pela Rede Nacional de Ensino e Pesquisa (RNP), que atua como um ponto de interconexão entre as instituições provedoras de recursos para o *testbed*, denominadas *ilhas de experimentação*, localizadas em diferentes organizações de ensino e pesquisa no Brasil e em redes parceiras internacionais. Uma ilha típica do FIBRE possui nós wireless, NetFPGA's, comutadores (tradicionais e *Openflow*) e sistemas de monitoramento. Cada ilha é conectada ao POP mais próximo da RNP e virtualmente conectada à rede do FIBRE, denominada de FIBREnet. A infraestrutura atual é composta pelas ilhas: RNP (Brasília), USP (São Paulo), UFPE (Recife), UFRJ (Rio de Janeiro), UFSCar (São Carlos), UFF (Niterói), UNIFACS (Salvador), UFG (Goiânia), UFPA (Belém), CPqD (Campinas) e Ampath (Miami). Ademais, está sendo realizada a inclusão de mais quatro instituições de pesquisa nacionais (UFES, UFRGS, UFU e UFMG) [Ciuffo et al. 2016]. A topologia atual é ilustrada na figura 10.6.

10.6. Visão de Futuro

Para projetarmos uma perspectiva de como será o contexto da distribuição do conteúdo nos próximos anos é importante identificar quais foram os movimentos que direcionaram a evolução deste cenário nos últimos anos.

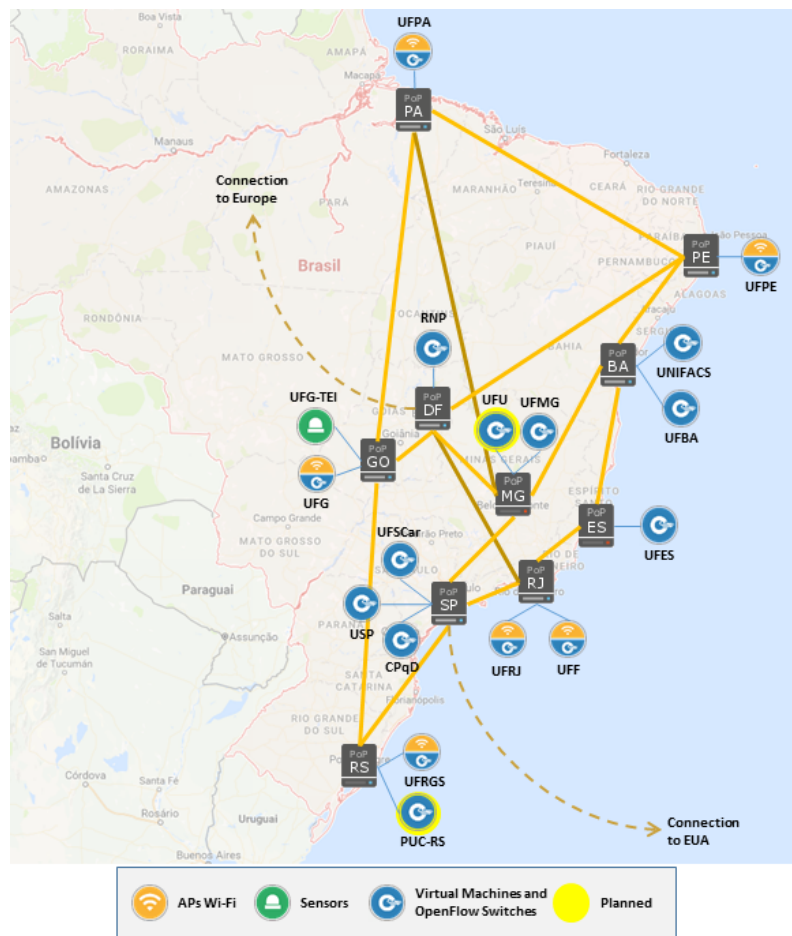


Figura 10.6. Topologia FIBRE [FIBRE 2017].

Apesar de produções fonográficas fazerem parte de um segmento pouco utilizado como caso de estudo nas pesquisas, observar a trajetória deste mercado e as tecnologias envolvidas podem nos dar diversos indícios de como o “mundo” do vídeo pode evoluir nos próximos anos. A evolução do mercado fonográfico partiu de um estado em que as grandes produtoras dominavam os negócios. Mas com o surgimento do MP3, Napster, iPod, iTunes, YouTube e recentemente do Spotify, este setor mudou completamente. Hoje em dia, vemos frequentemente, artistas “independentes” lançando *singles* em plataformas online e estabelecendo contratos diretamente com serviços de distribuição voltado ao usuário final, como o Spotify. Da mesma forma, podemos ver o setor de vídeo neste movimento. Hoje em dia vemos a revolução do Netflix, que promove diretamente a geração de “grandes produções” para disponibilizar em sua plataforma multi-dispositivos. Além disso, possibilita acompanhar “séries de TV” através de *smartphones* no transporte público.

Concluindo, o que se poder vislumbrar para os próximos anos é um acesso cada vez mais ubíquo aos conteúdos, os quais serão disponibilizados diretamente pelos produtores. Assim, a seguir, são apresentadas as tendências de como a evolução da distribuição dos conteúdos se dará nos próximos anos mediante este cenário de desenvolvimento tecnológico em que a Infraestrutura Definida por Software será a base para a implementação

das soluções.

10.6.1. Interface com o Usuário

Nas questões que tangem a interação do usuário, podemos distinguir alguns aspectos, os quais influenciam diretamente as soluções a serem adotadas. O primeiro é em relação ao tipo do conteúdo - educacional, de entretenimento, jornalístico, esportivo, pessoal.

Dado este aspecto, o formato, o meio de acesso ao conteúdo e as ferramentas auxiliares diferem muito. Neste sentido, o que se espera é que os conteúdos de cunho educacional façam uso de ferramentas interativas de forma cada vez mais frequente afim de potencializar o aprendizado dos alunos. Este tipo de conteúdo tem grande potencial para ser utilizado com ferramentas de interações mais elaboradas, em que se espera um maior comprometimento do usuário. Assim, existe um grande potencial para ferramentas que permitam verificar de forma eficaz a compreensão do conteúdo exposto, visto que através desta, poderia ser aferido legalmente o cumprimento de metas estudantis. Por outro lado, na cadeia de produção de conteúdos educacionais, existe um grande potencial para a geração de vídeos com a utilização dos *smartphones*. Por exemplo, um professor poderia fazer a gravação do conteúdo de forma independente, isto é, sem o auxílio de uma grande equipe de apoio técnico.

Os conteúdos jornalísticos e esportivos por sua vez, poderão usufruir da interatividade para criar uma relação mais forte do espectador com o emissor do conteúdo, assegurando a sua fidelidade e potencializar as vendas de produtos correlatos (ex: camisas de times). Todavia, como o propósito do consumo deste tipo de conteúdo tem motivações bem diferentes das educacionais, principalmente na questão de comprometimento, o tipo de ferramenta de interação deve ser adequado para este tipo de cenário, considerando inclusive o meio de acesso e o envolvimento social decorrente do conteúdo. Por exemplo, uma transmissão de um jogo de futebol geralmente é vista de forma coletiva através de uma projeção em uma TV. Contudo, o meio de interação poderia ser realizado através de *smartphones*. As possibilidades para este cenário são muitas. Se o consumo deste conteúdo é realizado em um estabelecimento comercial (como um restaurante), poderiam ser incluídas facilidades para compra de itens disponibilizados neste, inclusive com a associação do produtor do item, construindo uma estratégia de marketing integrada. Por outro lado, a combinação deste tipo de conteúdo com óculos de realidade virtual pode proporcionar uma experiência com grande potencial de interesse, pois apesar de fisicamente estar isolado, poderia vivenciar emoções semelhantes de estar dentro do estádio, mas sem os incômodos de deslocamentos e filas para estar presente.

Todavia, este cenário apresenta diversos desafios a serem superados. O primeiro é referente a latência entre o evento (um gol) até a reprodução do vídeo na tela. Este é um fato que decorre da latência inerente a todo processo desde a digitalização, passando pela transmissão até a reprodução no dispositivo do usuário. O problema é que esta latência aumenta significativamente com a introdução de maiores resoluções de vídeo e de tecnologias como o HLS/DASH, associado ao fato de que para este tipo de conteúdo, a latência representa um fator muito importante ao QoE.

Assim, um dos grandes desafios para atender satisfatoriamente este cenário é a capacidade de garantir uma baixa latência por meio das tecnologias mais proeminentes.

Para isso é importante considerar as questões dos protocolos de transmissão do conteúdo (RTMP) e de consumo (HLS/DASH), assim como fatores ligados diretamente com a codificação do vídeo. Ao analisar estes protocolos que dão suporte à distribuição de vídeo, percebemos um ciclo de evolução para atender da melhor forma o aspecto do consumo do conteúdo, através do HLS e DASH. Entre estes, apesar do HLS ser o pioneiro em trazer estes avanços aos usuários, o amadurecimento do DASH tem sido baseado em fatores importantes, como ser independente ao formato do vídeo e o suporte a DRM de forma mais aberta. Sendo assim, o movimento que tem sido visto é que gradualmente o DASH venha a ser o protocolo mais utilizado para o consumo de vídeos.

Por outro lado, atualmente o protocolo para a transmissão de vídeo ao vivo usado de forma ampla é o RTMP. Apesar de atender satisfatoriamente as demandas atuais, este foi projetado para um cenário com requisitos de uma solução proprietária, tornando assim uma solução amarrada às tecnologias (ex: *codecs* de vídeo) consideradas neste contexto. Sendo assim, é desejável que seja desenvolvido um novo protocolo que atenda as demandas desta nova geração de serviços de vídeo, e que, preferencialmente, esteja baseada em tecnologias abertas. Uma característica desejável seria que este protocolo suportasse o transporte de fluxos de vídeo de forma independente ao codec utilizado.

Uma outra tendência que deve ser observada é o aumento do uso do WebRTC para o desenvolvimento de novas aplicações devido à sua facilidade de utilização e de forma sinérgica com o movimento de padronização guiado pelo W3C e IETF.

10.6.2. Serviço de Distribuição de Conteúdo

Em relação aos serviços de distribuição de conteúdo, estamos vendo nos últimos anos diversas revoluções, em que o Netflix é um caso de uso muito interessante a ser estudado. Entre os aspectos relevantes para o contexto desta seção, destaca-se o processo evolutivo em que partiu de um serviço que delegava a distribuição do conteúdo a outros serviços, como Akamai, e começou a construir sua própria infraestrutura.

Este caso demonstra como a evolução dos serviços de distribuição de conteúdo tende a progredir. A perspectiva é de que cada vez mais os grandes provedores de conteúdo busquem não apenas serviços de distribuição de conteúdo que ofereçam funcionalidades inovadoras, mas também que possam ajustar de forma otimizada a alocação dos recursos necessários para atender as demandas de seus clientes.

Neste sentido, a construção de CDNs Virtuais, como a proposta apresenta em [Um et al. 2014], permitiria minimizar os custos de operação do provedor de conteúdo. Além disso, ao associar as tecnologias de orquestração de nuvens e as virtualizações de funções de redes (NFV), é possível construir serviços de CDNs elásticas ajustadas para atender as demandas de cada provedor de conteúdo.

No entanto, para dar suporte a esta estratégia, é indispensável sistemas de monitoração adequados para estas novas infraestruturas. Apesar deste tópico estar associado fortemente com a infraestrutura, a monitoração das aplicações que provem estes serviços podem contribuir na estimativa de demanda a ser considerada no provisionamento dos recursos, tanto computacionais como de conectividade.

Neste sentido, soluções como o Prometheus [Prometheus 2016] provem funcio-

nalidades muito interessantes para este cenário. Dentre as facilidades está a possibilidade de agregar informações monitoradas através de uma linguagem funcional, que associado ao sistema de alerta que permite disparar eventos para tratar casos que podem afetar o funcionamento ótimo do sistema.

De forma complementar, são necessários algoritmos que permitam especificar de forma programática como interpretar cada um dos parâmetros monitorado e, através de uma composição, quais recursos devem ser reservados a fim de tangenciar o ponto ótimo de alocação. Um exemplo de pesquisa neste sentido é o apresentado em [Silveira 2016] em que é proposto um algoritmo de alocação de recursos baseado em um limiar deslizante.

Para garantir redes virtualizadas e serviços de redes virtualizados com funcionamento dinâmico e eficiente há a necessidade do uso cada vez mais intenso de mecanismos de processamento baseados em BigData para extrair informações relevantes da monitoração da infraestrutura, da mesma forma que deve-se fazer o uso de processos relacionados à inteligência artificial para a automatização e elasticidade desta infraestrutura.

10.6.3. Infraestrutura Definida por Software

Conforme apresentado na seção 10.5, existe um grande empenho tanto de grandes empresas como de centros de pesquisa para implantação das SDNs. Entre os fatores que motivam este empenho, destaca-se a otimização do uso da infraestrutura de rede, pois permite uma engenharia de tráfego mais dinâmica e a redução dos tempos de configuração dos equipamentos no provimento de serviços.

Todavia, um dos grandes desafios para esta implantação é como realizar a transição destas infraestruturas de forma gradual, permitindo uma evolução progressiva dos serviços implementados, sendo que a troca dos equipamentos deste porte implica em um grande investimento. Ademais, é necessário garantir que não haja interrupção dos serviços que atendem os clientes.

Assim, muitas estratégias tem sido propostas para a realização deste processo [Das et al. 2015; Basta et al. 2015; Nelson et al. 2015; Ibarra et al. 2015]. Esta diversidade provém dos diferentes objetivos a serem alcançados em cada situação. Por exemplo, em casos que existe uma eminência de troca de equipamentos para suprir a crescente demanda da rede, podem ser adotadas estratégias de substituir os equipamentos tradicionais, por equipamentos de maior capacidade, mas que atue de forma transparente à rede. Por outro lado, se o objetivo é o provisionamento de novos serviços, ou melhoria dos existentes, utilizando as funcionalidades provenientes da SDN, podem ser criadas redes sobrepostas à rede em operação, pela qual o tráfego dos novos serviços são encaminhados, operada usando o paradigma SDN.

Finalmente, com a integração das SDNs, com os serviços de orquestração (*SaaS* e *IaaS*) e a virtualização das funções de rede, é possível a implementação de redes otimizadas para cada serviço. Isto é, provisionar de forma elástica os recursos computacionais e configurando de forma automática toda infraestrutura, atendendo as demandas de cada serviço de forma a prover QoE esperada e minimizando o custo operacional.

10.7. Considerações finais

As análises de prospecção do uso futuro da Internet apontam para um crescente consumo de fluxos de vídeo, o que vai ao encontro com o comportamento observado dos usuários das novas gerações. O surgimento do UHD TV, com a possibilidade de produção de vídeos com altas resoluções, associado às aplicações imersivas e de vídeocolaboração, propiciam o aumento cada vez maior da sensação de presença e de realidade ao utilizar aplicações de vídeo. Este cenário indica uma conseqüente tendência ao maior consumo de recursos da rede, assim como uma grande expectativa de bom atendimento dos serviços de vídeo.

Este capítulo procurou fazer um levantamento do cenário atual do consumo de vídeo, considerando os novos serviços, novos dispositivos e o potencial de consumo esperado. Para tanto, levou-se em consideração o estado da arte das tecnologias que surgiram recentemente que dão suporte aos mecanismos de distribuição de vídeo e as CDNs.

Procurou-se também contextualizar os serviços de distribuição de vídeo dentro do panorama de redes e serviços virtualizados, considerando as tecnologias SDN e NFV. A partir deste olhar foi possível traçar um horizonte a respeito do futuro do uso da rede de distribuição de vídeo na era da infraestrutura definida por software do ponto de vista da interface com o usuário, do serviço e da infraestrutura de rede.

10.8. Agradecimentos

Esse trabalho teve apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo processo nº 2015/50479-6.

Referências

- Aditya, Paarijaat et al. (2012). “Reliable client accounting for P2P-infrastructure hybrids”. Em: *NSDI’12: Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation*.
- ADLER, Stephen (1999). “The Slashdot effect: an analysis of three Internet publications”. Em: *Linux Gazette Issue*.
- Afanasyev, Alexander et al. (2014). “Named Data Networking”. Em: *ACM SIGCOMM Computer Communication Review* 44.3, pp. 66–73. ISSN: 19435819.
- Aggarwal, Vinay, Anja Feldmann e Christian Scheideler (2007). “Can ISPS and P2P users cooperate for improved performance?” Em: *ACM SIGCOMM Computer Communication Review* 37.3, p. 29. ISSN: 01464833.
- Ahmad, Ijaz et al. (2016). “New Concepts for Traffic , Resource and Mobility Management in Software-Defined Mobile Networks New Concepts for Traffic , Resource and Mobility Management in Software-Defined Mobile Networks”. Em: *JANUARY*, pp. 41–48.
- Arlitt, M. e T. Jin (2000). “A workload characterization study of the 1998 World Cup Web site”. Em: *IEEE Network* 14.3, pp. 30–37. ISSN: 08908044.
- Arumathurai, Mayutan et al. (2013). “Evaluation of ALTO-enhanced request routing for CDN interconnection”. Em: *IEEE International Conference on Communications*, pp. 3519–3524. ISSN: 15503607.

- Barman, Nabajeet e Maria G. Martini (2017). “H.264/MPEG-AVC, H.265/MPEG-HEVC and VP9 codec comparison for live gaming video streaming”. Em: *2017 9th International Conference on Quality of Multimedia Experience, QoMEX 2017*.
- Basta, Arsany et al. (2015). “Towards a dynamic SDN virtualization layer: Control path migration protocol”. Em: *2015 11th International Conference on Network and Service Management (CNSM)*. IEEE, pp. 354–359. ISBN: 978-3-9018-8277-7.
- Berde, Pankaj et al. (2014). “ONOS”. Em: *Proceedings of the third workshop on Hot topics in software defined networking - HotSDN '14*. New York, New York, USA: ACM Press, pp. 1–6.
- Bergkvist, Adam et al. (2012). “WebRTC 1.0: Real-time communication between browsers”. Em: *Working draft, W3C* 91.
- Bhaumik, Partha et al. (2014). “Software-defined optical networks (SDONs): A survey”. Em: *Photonic Network Communications* 28.1, pp. 4–18. ISSN: 1387974X.
- Bosshart, Pat et al. (2014). “P4”. Em: *ACM SIGCOMM Computer Communication Review* 44.3, pp. 87–95. ISSN: 01464833.
- Bousselmi, Khadija, Zaki Brahmi e Mohamed Mohsen Gammoudi (2014). “Cloud services orchestration: A comparative study of existing approaches”. Em: *Proceedings - 2014 IEEE 28th International Conference on Advanced Information Networking and Applications Workshops, IEEE WAINA 2014*. IEEE, pp. 410–416. ISBN: 9781479926527.
- Broberg, James, Rajkumar Buyya e Zahir Tari (2009). “MetaCDN: Harnessing 'Storage Clouds' for high performance content delivery”. Em: *Journal of Network and Computer Applications* 32.5, pp. 1012–1022. ISSN: 10848045.
- Buyya, Rajkumar, Mukaddim Pathan e Athena Vakali (2008). *Content Delivery Networks*. Vol. 49. 1, p. 429. ISBN: 9783540778868.
- Chanda, Abhishek, Cedric Westphal e Dipankar Raychaudhuri (2013). “Content based traffic engineering in software defined information centric networks”. Em: *Computer Communications Workshops (INFOCOM WKSHPS), 2013 IEEE Conference on*. IEEE, pp. 357–362.
- Cheng, R.-G. R-G e C.-J. C-J Chang (1997). “Neural-network connection-admission control for ATM networks”. Em: *IEE Proceedings-Communications* 144.2, pp. 93–98. ISSN: 13502425.
- Choffnes, David R e Fabián E Bustamante (2008). “Taming the torrent: a practical approach to reducing cross-isp traffic in peer-to-peer systems”. Em: *ACM SIGCOMM Computer Communication Review* 38, pp. 363–374. ISSN: 0146-4833.
- Cisco (2017). “The Zettabyte Era - Trends and Analysis.” Em: *Cisco* June, pp. 2016–2021.
- Ciuffo, Leandro et al. (2016). “Testbed fibre: Passado, presente e perspectivas”. Em: *Anais do WPEIF 2016 Workshop de Pesquisa Experimental da Internet do Futuro*, pp. 3–6.
- Committee, A T M Forum Technical et al. (1997). *Audiovisual multimedia services: Video on demand specification 1.1*.
- Conti, Marco et al. (2011). “Research challenges towards the Future Internet”. Em: *Computer Communications* 34.18, pp. 2115–2134. ISSN: 01403664.

- Cugini, Filippo et al. (2008). “Enhancing GMPLS signaling protocol for encompassing quality of transmission (QoT) in all-optical networks”. Em: *Journal of Lightwave Technology* 26.19, pp. 3318–3328. ISSN: 0733-8724.
- Das, Tamal et al. (2015). “Insights on SDN migration trajectory”. Em: *IEEE International Conference on Communications*. Vol. 2015-Sept. IEEE, pp. 5348–5353. ISBN: 9781467364324.
- Diorio, Rafael Fernando e Varese Salvador Timóteo (2016). “Per-Flow Routing with QoS Support to Enhance Multimedia Delivery in OpenFlow SDN”. Em: *Proceedings of the 22nd Brazilian Symposium on Multimedia and the Web - Webmedia '16*, pp. 167–174.
- Dustdar, Schahram et al. (2011). “Principles of elastic processes”. Em: *IEEE Internet Computing* 15.5, pp. 66–71. ISSN: 10897801.
- Egilmez, H. E. et al. (2012). “OpenQoS: An OpenFlow controller design for multimedia delivery with end-to-end Quality of Service over Software-Defined Networks”. Em: *Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference*, pp. 1–8.
- ETSI (2013). “Network Functions Virtualisation (NFV)”. Em: *Architectural Framework*.
- Faigl, Zoltan, Zsolt Szabo e Robert Schulcz (2014). “Application-layer traffic optimization in software-defined mobile networks: A proof-of-concept implementation”. Em: *2014 16th International Telecommunications Network Strategy and Planning Symposium (Networks)*. IEEE, pp. 1–6. ISBN: 978-1-4799-6515-1.
- Fan, Qilin et al. (2017). “Video delivery networks: Challenges, solutions and future directions”. Em: *Computers & Electrical Engineering*, pp. 1–10. ISSN: 00457906.
- FIBRE (2017). *FIBRE Overall Map Architecture*. URL: <http://fibre.org.br/infrastructure/resources/>.
- Firestone, Daniel (2017a). “Accelerated SDN in Azure”. Em: *Open Networking Summit*.
- (2017b). “{VFP}: A Virtual Switch Platform for Host {SDN} in the Public Cloud”. Em: *14th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 17)*. Boston, MA: {USENIX} Association, pp. 315–328. ISBN: 978-1-931971-37-9.
- Fluckiger, François (1995). *Understanding Networked Multimedia: Applications and Technology*. Hertfordshire, UK, UK: Prentice Hall International (UK) Ltd. ISBN: 0-13-190992-4.
- Frank, Benjamin et al. (2013). “Collaboration Opportunities for Content Delivery and Network Infrastructures”. Em: *Recent Advances in Networking* 1, pp. 305–377.
- Gates, Bill (1996). *Content is king*. URL: <http://www.sick-site-syndrome.com/articles/content-is-king.shtml>.
- Ghodsi, Ali et al. (2011). “Information-centric networking”. Em: *Proceedings of the 10th ACM Workshop on Hot Topics in Networks - HotNets '11*, pp. 1–6.
- Google (2012). “Inter-Datacenter WAN with centralized TE using SDN and OpenFlow”. Em: *Open Networking Summit*, pp. 1–2.
- Goran, N, M Hadzialic e A Begovic (2014). “Real time assuring QoE in the lowest OSI/ISO layers during delivering of IPTV services”. Em: *2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. IEEE. IEEE, pp. 532–535. ISBN: 978-953-233-077-9.

- Grois, Dan, Tung Nguyen e Detlev Marpe (2017). “Coding efficiency comparison of AV1/VP9, H.265/MPEG-HEVC, and H.264/MPEG-AVC encoders”. Em: *2016 Picture Coding Symposium, PCS 2016*.
- Gurbani, Vijay K. et al. (2012). “Abstracting network state in Software Defined Networks (SDN) for rendezvous services”. Em: *2012 IEEE International Conference on Communications (ICC)*. IEEE, pp. 6627–6632. ISBN: 978-1-4577-2053-6.
- Han, Yoonseon et al. (2016). “An intent-based network virtualization platform for SDN”. Em: *2016 12th International Conference on Network and Service Management (CNSM)*. IEEE, pp. 353–358. ISBN: 9783901882852.
- Herbaut, Nicolas et al. (2017). “Dynamic Deployment and Optimization of Virtual Content Delivery Networks”. Em: *IEEE MultiMedia*, pp. 1–1. ISSN: 1070-986X.
- Ibarra, Julio et al. (2015). “Benefits brought by the use of OpenFlow/SDN on the AmLight intercontinental research and education network”. Em: *Proceedings of the 2015 IFIP/IEEE International Symposium on Integrated Network Management, IM 2015*. IEEE, pp. 942–947. ISBN: 9783901882760.
- IETF (2008). *IETF ALTO WG Charter*. URL: <http://datatracker.ietf.org/wg/alto/charter/>.
- Jain, Raj e Subharthi Paul (2013). *Network virtualization and software defined networking for cloud computing: A survey*.
- Jain, Sushant, Alok Kumar et al. (2013). “B4: experience with a globally-deployed software defined wan”. Em: *ACM SIGCOMM Computer Communication Review* 43.4, pp. 3–3–14–14. ISSN: 0146-4833.
- Kang, Joon Myung et al. (2013). “Software-defined infrastructure and the Future Central Office”. Em: *2013 IEEE International Conference on Communications Workshops, ICC 2013*, pp. 225–229.
- Kopp, Samuel et al. (2017). “Content Delivery Networks integrated with Software Defined Networks”. Em: p. 4.
- Kurose, J e K Ross (2012). *Computer Networking: A Top Down Approach, 4e*.
- Ladiwala, Sameer, Ramaswamy Ramaswamy e Tilman Wolf (2009). “Transparent TCP acceleration”. Em: *Computer Communications* 32.4, pp. 691–702. ISSN: 01403664.
- Lao, Li et al. (2007). “A scalable overlay multicast architecture for large-scale applications”. Em: *IEEE Transactions on Parallel and Distributed Systems* 18.4, pp. 449–459. ISSN: 10459219.
- Laver, K. et al. (2012). “Virtual Reality for Stroke Rehabilitation”. Em: *Stroke* 43.2, e20–e21. ISSN: 0039-2499.
- Lemos, Guido (2002). *Grupo de Trabalho de Video Digital*. URL: <https://memoria.rnp.br/pd/gts2002-2003/gt-vd.html>.
- Lin, Chia-Feng Feng et al. (2011). “The study and methods for cloud based CDN”. Em: *2011 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*. IEEE, pp. 469–475. ISBN: 9780769545578.
- Mao, Hongzi, Ravi Netravali e Mohammad Alizadeh (2017). “Neural Adaptive Video Streaming with Pensieve”. Em: *Mohammad Alizadeh MIT Computer Science and Artificial Intelligence Laboratory*.
- McKeown, Nick (2009). “Software-defined networking”. Em: *INFOCOM keynote talk* 17.2, pp. 30–32.

- Nam, Hyunwoo et al. (2014). “Towards QoE-aware video streaming using SDN”. Em: *2014 IEEE Global Communications Conference, GLOBECOM 2014*. IEEE, pp. 1317–1322. ISBN: 9781479935116.
- Nelson, Tim et al. (2015). “Exodus: Toward Automatic Migration of Enterprise Network Configurations to SDNs”. Em: *Proceedings of the 1st ACM SIGCOMM Symposium on Software Defined Networking Research - SOSR '15*, pp. 1–7.
- Nygren, Erik et al. (2010). “The Akamai Network: A platform for high-performance Internet applications”. Em: *ACM SIGOPS Operating Systems Review* 44.3, pp. 2–19. ISSN: 0163-5980.
- Odlyzko, Andrew (2000). “Content is not King”. Em: *SSRN Electronic Journal*, pp. 1–21. ISSN: 1556-5068.
- ONF (2012). *Software-Defined Networking : The New Norm for Networks*. URL: <https://www.opennetworking.org/images/stories/downloads/sdn-resources/white-papers/wp-sdn-newnorm.pdf>.
- ON.LAB (2016). *Intent Framework - ONOS - Wiki*. URL: <https://wiki.onosproject.org/display/ONOS/Intent+Framework> (acesso em 17/08/2015).
- ON.Lab (2017). *Global SDN Deployment Powered by ONOS*. URL: <https://wiki.onosproject.org/display/ONOS/Global+SDN+Deployment+Powered+by+ONOS>.
- OpenDayLight (2015). *Network Intent Composition*. URL: https://wiki.opendaylight.org/view/Network%7B%5C_%7DIntent%7B%5C_%7DComposition:Main.
- OpenDaylight (2013). *OpenDaylight: A Linux Foundation Collaborative Project*. URL: <http://www.opendaylight.org> (acesso em 29/04/2015).
- Orzessek, Michael e Peter Sommer (1997). *ATM & MPEG-2: integrating digital video into broadband networks*. Prentice-Hall, Inc.
- Pantos, R e W May (2012). *HTTP Live Streaming*. Rel. téc., pp. 1–33. URL: <http://tools.ietf.org/html/draft-pantos-http-live-streaming-07>.
- Passarella, Andrea (2012). “A survey on content-centric technologies for the current Internet: CDN and P2P solutions”. Em: *Computer Communications* 35.1, pp. 1–32. ISSN: 01403664.
- Peterson, L., B. Davie e Ed. R. van Brandenburg (2014). *Framework for Content Distribution Network Interconnection (CDNI)*. Rel. téc. RFC Editor, pp. 1–58. URL: <http://www.rfc-editor.org/rfc/pdf/rfc7336.txt.pdf>.
- Pham, Minh e Doan B. Hoang (2016). “SDN applications - The intent-based Northbound Interface realisation for extended applications”. Em: *2016 IEEE NetSoft Conference and Workshops (NetSoft)*. IEEE, pp. 372–377. ISBN: 978-1-4673-9486-4.
- Pimentel, Hécio M. et al. (2015). “OCP: A protocol for secure communication in federated content networks”. Em: *Computer Communications* 68, pp. 47–60. ISSN: 01403664.
- Poese, Ingmar et al. (2012). “Improving content delivery with PaDIS”. Em: *IEEE Internet Computing* 16, pp. 46–52. ISSN: 10897801.
- Pozdnyakov, Andrey (2017). *AOM AV1 vs. HEVC*. URL: https://www.elecard.com/page/aom_av1_vs_hevc.
- Prometheus (2016). *The Prometheus monitoring system and time series database*. URL: <https://github.com/prometheus>.
- Quittek, Jürgen et al. (2014). “Network Functions Virtualisation (NFV); Management and Orchestration”. Em: *Gs Nfv-Man 001 V1.1.1* 1, pp. 1–184.

- Ross, Keith W. et al. (2008). “Understanding hybrid CDN-P2P”. Em: *Proceedings of the 18th International Workshop on Network and Operating Systems Support for Digital Audio and Video - NOSSDAV '08*, p. 75.
- Sallent, Sebastia et al. (2012). “FIBRE Project: Brazil and Europe Unite Forces and Testbeds for the Internet of the Future”. Em: *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering*. Vol. 44 LNICST. Springer, Berlin, Heidelberg, pp. 372–372. ISBN: 9783642355752.
- El-Sayed, Ayman (2004). “Application-level multicast transmission techniques over the internet”. Tese de doutorado. Institut National Polytechnique de Grenoble.
- Schoeffmann, Klaus, Marco A. Hudelist e Jochen Huber (2015). “Video Interaction Tools”. Em: *ACM Computing Surveys* 48.1, pp. 1–34. ISSN: 03600300.
- Schulzrinne, Henning, Anup Rao e Robert Lanphier (1998). *Real Time Streaming Protocol (RTSP)*. RFC 2326. RFC Editor. URL: <http://www.rfc-editor.org/rfc/rfc2326.txt>.
- Silveira, Regina Melo (2016). “Rede Elástica de Distribuição de Vídeo Baseada em Virtualização de Serviços”. Livre Docência. USP, p. 96.
- Silveira, Regina Melo, Rubens Ramires Fonseca et al. (2003). “TVoD: Sistema Multimídia sob Demanda para Distribuição de Material Digital das TVs Educativas”. Em: *4o. Workshop da RNP2*, pp. 1–4.
- Silveira, R.M. e W.V. Ruggiero (2001). “Server of quality system for video on demand over ATM”. Em: *Joint 4th IEEE International Conference on ATM(ICATM'01) and High Speed Intelligent Internet Symposium. ICATM 2001 (Cat. No.00EX486)*. IEEE, pp. 118–122. ISBN: 0-7803-7093-7.
- Simoens, Pieter et al. (2017). “Service-Centric Networking for Distributed Heterogeneous Clouds”. Em: *IEEE Communications Magazine* 55.7, pp. 208–215. ISSN: 0163-6804.
- Singh, Arjun et al. (2016). “Jupiter rising”. Em: *Communications of the ACM*. Vol. 59. 9. ACM, pp. 88–97. ISBN: 9781450335423.
- Sodagar, Iraj (2011). “The MPEG-DASH Standard for Multimedia Streaming Over the Internet”. Em: *IEEE Multimedia* 18.4, pp. 62–67. ISSN: 1070-986X.
- Su, Ao Jan et al. (2009). “Drafting behind akamai: Inferring network conditions based on CDN redirections”. Em: *IEEE/ACM Transactions on Networking* 17.6, pp. 1752–1765. ISSN: 1063-6692.
- Szabo, Robert et al. (2015). “Elastic network functions: opportunities and challenges”. Em: *IEEE Network* 29.3, pp. 15–21. ISSN: 0890-8044.
- Tosatto, Andrea, Pietro Ruiu e Antonio Attanasio (2015). “Container-Based Orchestration in Cloud: State of the Art and Challenges”. Em: *Proceedings - 2015 9th International Conference on Complex, Intelligent, and Software Intensive Systems, CISIS 2015*. IEEE, pp. 70–75. ISBN: 9781479988709.
- Uchôa, D e R M Silveira (2007). “Oversea: Towards a scalable and effective architecture for overlay networks”. Em: *XIII Brazilian Symposium on Multimedia and the Web (WebMedia)*.
- Uchôa, Daniel C., Samuel Kopp et al. (2009). “An overlay application-layer multicast infrastructure”. Em: *2009 International Conference on Advanced Information Networking and Applications*, pp. 233–240. ISSN: 1550445X.

- Um, Tai-Won et al. (2014). “Dynamic Resource Allocation and Scheduling for Cloud-Based Virtual Content Delivery Networks”. Em: *ETRI Journal* 36.2, pp. 197–205. ISSN: 1225-6463.
- Vahdat, Amin e Bikash Koley (2017). *Espresso makes Google cloud faster, more available and cost effective by extending SDN to the public internet*. URL: <https://www.blog.google/topics/google-cloud/making-google-cloud-faster-more-available-and-cost-effective-extending-sdn-public-internet-espresso/>.
- Wang, L, J McGee e S Huang (2009). “Networking Named Content”. Em: *Sumtweets Googlecode Com*, pp. 1–12. ISSN: 00010782.
- Wasserman, Todd (2015). *The revolution wasn't televised: The early days of YouTube*. URL: mashable.com/2015/02/14/youtube-history (acesso em 18/08/2017).
- Woo, Honguk et al. (2014). “A Virtualized, Programmable Content Delivery Network”. Em: *2nd IEEE International Conference on Mobile Cloud Computing, Services, and Engineering*. ISBN: 9781479944255.
- Xie, H et al. (2012). *Use Cases for ALTO with Software Defined Networks*. Rel. téc. draft-xie-alto-sdn-use-cases-01.txt, pp. 1–29.
- Yue, Frank (2013). *Network functions virtualization-everything old is new again*.
- Zhang, Yali et al. (2016). *NEMO NETwork MOdeling Language*. Internet-Draft. Internet Engineering Task Force. URL: <https://tools.ietf.org/html/draft-xia-sdnrg-nemo-language-04>.

ANEXO 1: biografia dos autores



Samuel Kopp Pesquisador colaborador do LARC (Laboratório de Arquitetura e Redes de Computadores) da Escola Politécnica da Universidade de São Paulo e Sócio Diretor da TVoD. Formado em Ciências da Computação pela Universidade de Santo Amaro, com MBA e Mestrado pela Escola Politécnica da Universidade de São Paulo. Nos últimos anos tem trabalho em projetos nas áreas de tecnologia de redes com foco no desenvolvimento de soluções para distribuição de vídeo digital com qualidade de serviço (QoS) e Redes Definidas por Software. Atualmente é coordenador de um projeto FAPESP pelo programa PIPE - Pesquisa Inovativa em Pequenas Empresas, que tem como objetivo o desenvolvimento de uma CDN integrada à SDN.



Michael Prieto Hernández Possui graduação em Engenharia em Ciências Informáticas pela Universidade das Ciências Informáticas, Cuba 2009 e mestre em Ciências pela Escola Politécnica da Universidade de São Paulo no ano 2017. Tem experiência na área de Redes Definidas por Software com foco na qualidade de serviço e engenharia de tráfego. Atualmente é consultor especialista no contexto do projeto PIPE "Sistema para Rede de Distribuição de Conteúdos integrado à tecnologia SDN".



Lucas Takeshi Chigami Lucas Chigami, formado pela Universidade de São Paulo no curso de Engenharia Elétrica, trabalha como consultor e coordenador de projeto desde 2007 no Laboratório de Arquitetura e Redes de Computadores. Possui experiência em diversas tecnologias de infraestrutura e aplicações para dispositivos móveis. Possui grande interesse em DevOps, gerenciamento de infraestruturas e técnicas de virtualização.



Reinaldo Matushima Sócio Diretor da TV on Demand Consultoria e Soluções S/S Ltda. Formado em Engenharia de Computação pela Escola Politécnica da Universidade de São Paulo, com mestrado pela mesma entidade. Possui MBA em Gestão da Tecnologia da Informação pela FIA (2012) e Especialização em Gestão de Inovação pela USP (2014). Nos últimos anos tem trabalhado em projetos nas áreas de segurança, e principalmente no desenvolvimento de soluções para armazenamento, gerenciamento e distribuição de vídeo digital. Na TVoD atua como Gerente de Projetos, Coordenador Técnico, Arquiteto de Software e Consultor especializado.



Bruno Klava Sócio Diretor da TV on Demand Consultoria e Soluções S/S Ltda. Formado em Ciências da Computação pelo Instituto de Matemática e Estatística da Universidade de São Paulo (IME-USP), com Mestrado e Doutorado pela mesma entidade. Pesquisador consultor com ampla experiência em metodologia e desenvolvimento de software, com grande atuação nesta área nos últimos oito anos.



Regina Melo Silveira Professora Associada e pesquisadora da Escola Politécnica da Universidade de São Paulo, onde está vinculada ao LARC (Laboratório de Arquitetura e Redes de Computadores). Atua na área desde 1995, desenvolvendo projetos na área de aplicações multimídia para redes de alta velocidade, compressão e transmissão de vídeo digital, sistema de Vídeo sob Demanda, TV Digital, indexação e gerenciamento de mídias digitais. Participou dos Projetos Poli-Virtual, Sistema Multimídia sob Demanda, RMAV-SP (Internet 2 de São Paulo), Tidia-Ae, KyaTera, TV Interativa em parceria com a TV Cultura e de Projetos para o Sistema Brasileiro de TV Digital (SBTVD) em parceria com outras instituições de ensino e pesquisa. Atuou como Coordenadora do Grupo de Trabalho de Gerência de Vídeo, Grupo de Trabalho em Redes de Serviços Sobrepostos, e Grupo de Trabalho em Instrumentação e Monitoração para Aplicações de Vídeo, todos financiados pela RNP (Rede Nacional de Pesquisa e Ensino). Co-fundadora da empresa TV on Demand Consultoria e Soluções criada em 2009. Livre-docente pela Escola Politécnica da USP (2016). Pós-dourada pela University of California – Santa Cruz como bolsista da Fapesp (2007-2008). Doutora pela Escola Politécnica da USP (2000) e Mestre em Física Experimental pelo Instituto de Física da Universidade de São Paulo (1994). Bacharel em Física pela Pontifícia Universidade Católica de São Paulo (1988). Tem atuado principalmente nos seguintes temas: rede de distribuição de vídeo, aplicação multimídia, redes ópticas e qualidade de serviço e de experiência para web.

Capítulo

11

Mapeamento de soluções de videocolaboração da RNP

Leandro Ciuffo, Marcelino Cunha, Christian Miziara de Andrade,
Graciela Leopoldo Martins, Jean Carlo Faustino, Rafael Valle,
Fausto Vetter, Clayton Reis da Silva, Iara Machado, Luiz Coelho,
Antônio Carlos Nunes, Helder Vitorino, Paulo Roberto de Lima Lopes

Rede Nacional de Ensino e Pesquisa (RNP)

```
{leandro.ciuffo,marcelino.cunha,christian.miziara,graciela.martins,  
  jean.carlo,rafael.valle,fausto.vetter,clayton.reis,  
  iara,luiz.coelho,antonio,helder}@rnp.br, paulo.lopes@rute.rnp.br
```

Abstract

In 2017, the RNP Working Groups programme was reformulated, starting to call for projects to either evolve or expand some of RNP's services, including the video-collaboration ones. This made the need to respond to a previous demand from the CT-Video members, which is to make public technical information or guidelines on existing video-collaboration services, aiming at facilitating the development of new features and identifying opportunities for interoperability between services. As a first and simplified approach to meet this demand, this chapter presents a mapping of technological components of some video-collaboration solutions currently supported by RNP, which resulted in a graphic map available in a poster format.

Resumo

Em 2017 o Programa de Grupos de Trabalho da RNP foi reformulado, passando a também solicitar propostas de projeto para evoluir ou ampliar alguns dos serviços da RNP, incluindo os serviços de videocolaboração. Isso tornou mais latente a necessidade de atender a uma demanda já evidenciada pelos membros CT-Video, que é a publicação de informações sobre os serviços de videocolaboração existentes com o objetivo de facilitar a elaboração de propostas de desenvolvimento de novas funcionalidades, além da identificação de sinergias e possíveis interoperabilidades entre serviços. Como uma primeira abordagem simplificada para atender a essa demanda, este capítulo apresenta um mapeamento de componentes tecnológicos das soluções de videocolaboração existentes na RNP, que resultou na elaboração de um mapa gráfico disponibilizado no formato de pôster.

1. Introdução

A Rede Nacional de Ensino e Pesquisa (RNP) tem como missão promover o uso inovador de redes avançadas. Uma das formas de materialização dessa missão é através do desenvolvimento de aplicações e serviços inovadores, com foco no atendimento de demandas específicas de sua comunidade usuária, composta majoritariamente por universidades, institutos de pesquisa, museus e hospitais universitários brasileiros. Nos departamentos de computação e tecnologia da informação (TI) dessas instituições existe diversos grupos de pesquisa em áreas relacionadas a redes de computadores, engenharia de redes e sistemas distribuídos.

Como uma forma de promover maior interação entre as atividades executadas na RNP e esses grupos de pesquisa e, ainda, a inovação contínua dos produtos e serviços da RNP, o Programa de Grupos de Trabalho (GT-RNP) foi criado em 2002. O Programa GT-RNP visa desenvolver projetos colaborativos que possam demonstrar a viabilidade da utilização de novos protocolos, serviços e aplicações na rede que atendam às necessidades das instituições educacionais e de pesquisa conectadas à rede da RNP. Desde então, essa iniciativa tem atraído, através de editais lançados anualmente, o interesse de grupos de pesquisa de todo o Brasil, desafiados a desenvolver aplicações e serviços inovadores. As propostas são avaliadas por um comitê com representantes internos e externos à RNP e as melhoras propostas são selecionadas para o programa. Em geral, o ciclo de desenvolvimento dos projetos aprovados é organizado em até três fases de 12 meses de duração cada. Desde 2002, vários temas foram abordados pelo programa, como: telefonia através da Internet, redes sem fio, desempenho de redes, gestão de identidade, sistemas colaborativos, educação à distância e outros. Ao longo dos anos, muitos projetos resultantes do Programa de GTs transformaram-se em serviços em produção da RNP.

Em 2017 o Programa de Grupos de Trabalho foi reformulado, passando a sugerir temas de interesse da RNP e sendo dividido em dois programas: Internet Avançada e Serviços Avançados¹. O objetivo do primeiro foi prospectar propostas de produtos ou serviços que ajudem a otimizar ou automatizar a infraestrutura de redes, assim como aplicações de análise de monitoramento e novas arquiteturas que permitam a evolução da rede acadêmica. O objetivo do programa Serviços Avançados foi o de buscar propostas que pudessem contribuir para a evolução ou ampliação do portfólio de serviços avançados da RNP, incluindo os serviços de videocolaboração.

Possivelmente esses programas passarão por novas reformulações e aperfeiçoamentos nos editais dos próximos anos, sendo um deles a divulgação de informações técnicas sobre os serviços ofertados pela RNP, a fim de facilitar a proposição de projetos com a finalidade de evoluí-los. Um primeiro passo nessa direção é o mapeamento de componentes tecnológicos de cada serviço, apresentado neste trabalho.

O restante deste capítulo está estruturado da seguinte maneira: a seção 2 apresenta e categoriza os serviços de videocolaboração da RNP que foram selecionados para fazerem parte do mapeamento inicial. A seção 3 descreve a metodologia utilizada e apresenta a primeira versão do mapa. A seção 4 encerra o capítulo apresentando novas funcionalidades previstas em alguns dos serviços.

¹ <https://www.rnp.br/noticias/rnp-lanca-edital-programas-pd-2017>

2. Catálogo de serviços da RNP e as soluções de videocolaboração

O Programa de Grupos de Trabalho constitui uma das formas de prospecção e oferta de novos serviços pela RNP, mas não é a única. Existem outros programas de pesquisa e desenvolvimento (P&D), e diversos projetos de tecnologia de comunicação e colaboração (TIC) que visam implantar soluções específicas para determinadas comunidades de usuários, que eventualmente podem gerar novos serviços.

De modo geral, pode-se dizer que os serviços oferecidos pela RNP são resultado de processos de inovação e prospecção em atividades de análise de cenários e tendências, com parceiros como a academia, o setor empresarial e as principais redes acadêmicas mundiais, que facilitam e promovem a comunicação, a colaboração a distância e a disseminação de conhecimento, englobando desde a estratégia até a operação.

Atualmente os serviços da RNP podem ser agrupados em oito categorias: (1) serviços de Comunicação e colaboração, (2) Hospedagem e armazenamento, (3) Gestão de identidade, (4) Ensino, (5) Conectividade, (6) Segurança, (7) e-Ciência e (8) serviços para Experimentação. As soluções de videocolaboração enunciadas neste capítulo referem-se a aplicações interativas que trafegam dados de áudio, imagens e/ou vídeo via Internet (rede IP). Desta forma as soluções apresentadas neste trabalho representam um subconjunto dos serviços de colaboração (Comunicação, Cooperação e Coordenação) em Ensino e Pesquisa. Essas soluções são descritas a seguir seguindo um agrupamento de acordo com o nível de maturidade.

Os “Serviços em produção” são as soluções disponibilizadas para todos os clientes da RNP e com suporte técnico profissional 24x7 através do Service Desk da RNP². São denominados “Serviços Experimentais” os serviços oriundos de programas de P&D que ainda encontram-se em fase de modelagem e pré-produção em parceria com um conjunto limitado de instituições cliente. Por fim, os “Serviços sob demanda” são aqueles que já encerraram seu ciclo de desenvolvimento com a RNP e encontram-se disponíveis para serem empregados para atender demandas específicas de determinados clientes ou eventos.

2.1. Serviços em produção

2.1.1. Vídeo@RNP

O Vídeo@RNP é o serviço para disponibilização e armazenamento de conteúdo audiovisual a partir de um portal web Vídeo@RNP³, que também integra os serviços de Transmissão de Vídeo ao Vivo e Transmissão de Sinal de TV. Mais do que um ambiente para disponibilização e visualização de vídeos, com interface amigável e infraestrutura de servidores distribuídos que otimiza o acesso pelo território nacional, ele oferece conteúdo audiovisual de alta qualidade, produzido pela comunidade acadêmica e instituições parceiras da RNP.

Além disso, o Vídeo@RNP conta, também, com uma versão com interface desenvolvida especificamente para *tablets* e smartphones que é visualizada quando o usuário faz o acesso a partir desse tipo específico de dispositivo.

² <https://www.rnp.br/servicos/servicos-avancados/service-desk>

³ <http://video.rnp.br/>

Mas o principal diferencial desse serviço em relação a outras soluções semelhantes disponíveis no mercado encontra-se na sua infraestrutura de rede, que implementa a Rede de Vídeo Digital (RVD) - um tipo de CDN (*Content Delivery Network*) restrita ao tráfego audiovisual, que faz com que as instituições clientes do serviço possam disponibilizar seu conteúdo, seja ele sob demanda ou por streaming (transmissão contínua), ao mesmo tempo que economizam o processamento de seus computadores e o uso de banda de acesso à internet.

2.1.2. Rede de Vídeo Digital (RVD)

Prover uma infraestrutura de servidores distribuídos ao longo de todo o território nacional com o objetivo de propiciar uma distribuição online e otimizada de vídeos para as instituições clientes. Esse é o objetivo do serviço de distribuição de vídeo sob demanda e ao vivo (para transmissões de eventos e canais de TV e distribuição de rádio), que conta com suporte a *multicast overlay*, “cacheamento” e funcionalidades associadas para monitoração e gerenciamento.

Com esse serviço, os clientes que desejam transmitir ao vivo um evento realizado em sua organização não precisam ter um servidor robusto ou uma grande capacidade de banda para dar conta dos inúmeros acessos simultâneos de usuários e espectadores.

2.1.3. Conferência Web

Realizar reuniões virtuais entre participantes remotos com a utilização de recursos de áudio, vídeo, texto, quadro de notas, chat e imagens, além do compartilhamento da tela do computador ou de aplicativos específicos. Este é o objetivo principal do serviço, que se destaca pela facilidade de uso e mobilidade. Voltado para clientes que demandam uma solução de menor custo para reuniões virtuais, treinamentos e palestras a distância, o serviço também se destaca pela facilidade de uso.

Para utilizar toda a integração e a interatividade disponibilizadas pelo Conferência Web⁴, basta um computador, celular ou tablete conectado à internet, um navegador (browser) e um conjunto de microfone e fone de ouvido, sem necessidade adicional de hardware ou software. Além desses recursos, o serviço possui a funcionalidade de gravar reuniões, que podem ser disponibilizadas para visualização ou baixadas para armazenamento.

A solução é fruto de um desenvolvimento iniciado em 2010 no âmbito do Programa de Grupos de Trabalho, que teve como premissa a criação de um sistema de webconferência de código aberto, baseado no arcabouço BigBlueButton⁵, distribuído, escalável e federado.

2.1.4. Fone@RNP

Serviço que permite a interconexão VoIP (voz sobre IP) entre diferentes organizações. Com isso, os clientes deste serviço conseguem oferecer a seus usuários a possibilidade de realizar chamadas telefônicas gratuitas para outras instituições distantes geograficamente

⁴ <https://conferenciaweb.rnp.br/> - também referenciado como “Webconferência”

⁵ <https://bigbluebutton.org/>

no Brasil ou no exterior, dentro ou fora da rede VoIP da RNP. O benefício das ligações gratuitas para o usuário final se estende através da capilaridade do fone@RNP⁶ - com presença em todos os estados brasileiros - e das conexões VoIP com a rede pública de telefonia de algumas instituições clientes, o que propicia ao usuário final ligar para um número telefônico comum a partir de um ramal VoIP. Destaque também para a interconexão da rede VoIP da RNP com outras redes VoIP dentro e fora do país. Além das ligações telefônicas gratuitas, as instituições clientes podem acessar um sistema centralizado de estatísticas (estatisticafone.rnp.br), que propicia ao gestor da infraestrutura local dispor de dados sobre como o serviço tem sido utilizado na sua instituição ou em outras que integram o fone@RNP.

2.1.5. Videoaula@RNP

Serviço integrado para elaboração, armazenamento e disponibilização pela web de videoaulas produzidas pelas instituições clientes⁷. Tais organizações passam a ter, por meio desse serviço, um meio para acesso e armazenamento de um amplo e interessante material didático formado por múltiplas mídias (vídeo, áudio, animações, roteiro e arquivos de apoio), que pode ser utilizado como apoio ao ensino a distância ou presencial. O upload das videoaulas digitais é feito em um sistema com redundância e alta disponibilidade, resultando em um acesso rápido e fácil a partir de um simples navegador web.

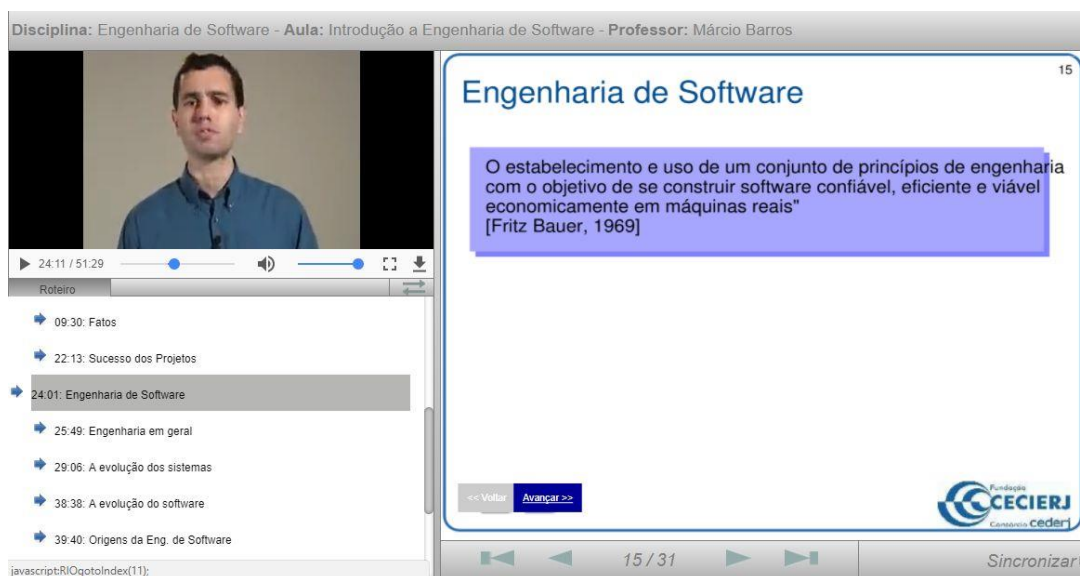


Figura 11.1. Videoaula de Engenharia de Software produzida pelo Consórcio CEDERJ

2.1.6. Intercâmbio de Conteúdos Digitais (ICD)

O principal objetivo é facilitar o processo de organização e disponibilização do acervo produzido por TVs universitárias, além de colaborar para a estruturação de grades de programação a partir dos conteúdos disponibilizados.

⁶ <https://www.rnp.br/servicos/servicos-avancados/fonernp>

⁷ <http://videoaula.rnp.br/>

A solução é implementada na forma de uma plataforma distribuída que tem como foco o intercâmbio de conteúdos digitais através da Internet (rede IP). São oferecidos vários recursos, como catalogação, armazenamento, classificação, compartilhamento, busca e transmissão de ativos digitais.

2.1.7. Exibidor de Conteúdo Digital e Controlador de Sessões

Parte do Programa Soluções Digitais para Cultura (SDC), o projeto “Rede de Cinemas Digitais ” tem como principal objetivo o estabelecimento de uma rede de distribuição e compartilhamento de conteúdos audiovisuais entre seus participantes, mais especificamente entre os cinemas e salas de exibição das instituições conectadas à rede operada pela RNP.

A execução do projeto viabilizou o estabelecimento de um circuito de salas compartilhadas para exibição de conteúdos de caráter cultural/alternativo e, conseqüentemente, um ambiente para a experimentação e realização de eventos e festivais de cinema distribuídos. Para tanto, a Rede constituída, referenciada como Rede de Cinemas é suportada por um conjunto bem definido de ferramentas, além de treinamento, documentação e suporte às salas de exibição associadas. Tais ferramentas, destacadas neste trabalho de mapeamento de soluções, são:

- Controlador: Software responsável por selecionar a playlist (criada no ICD) e executá-la para a realização da sessão de cinema. Permite visualizar suas informações, o status da disponibilidade dos conteúdos da playlist no ICD e também controlar a exibição dos conteúdos no projetor do cinema por comandos (play, pause, stop, etc.).Exibidor: Artefato de hardware/software responsável pela exibição de conteúdos digitais (de resolução HD a 4k) em salas de cinemas. Esse artefato se comunica com o ICD para obter os conteúdos, decodifica-os e os entrega para o projetor e sistema de som da sala de cinema.

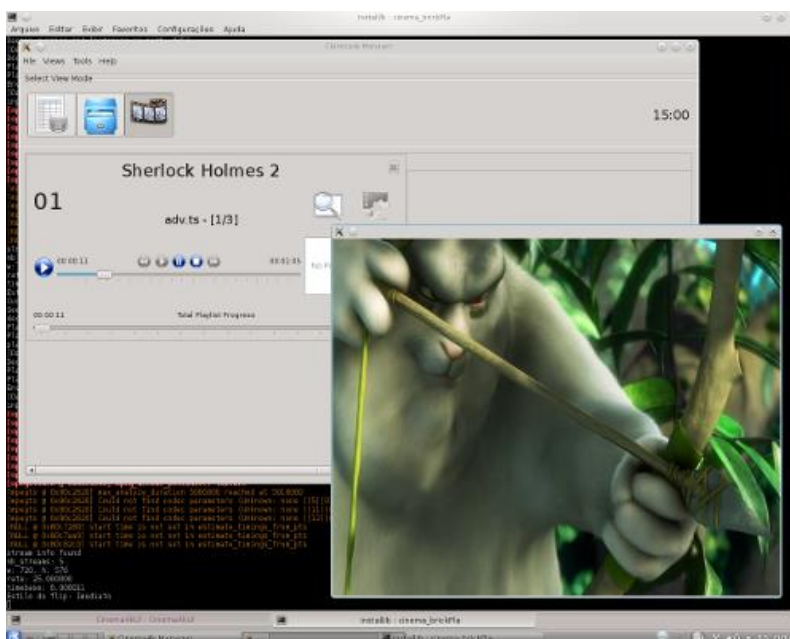


Figura 11.2. Reprodução de um vídeo pelo exibidor de cinema

2.1.8. Videoconferência

O principal objetivo desse serviço é fornecer salas virtuais para viabilizar a realização de videoconferências multiponto para as instituições clientes sem que estas precisem arcar com o ônus de adquirir uma solução local para isso. O serviço também oferece recursos de gravação, streaming e integração com sistemas de videoconferência que utilizem conexão por Integrated Service Digital Network -ISDN, além da inclusão de participantes com apenas áudio a partir de uma chamada por telefonia convencional.

Atualmente esse serviço, no núcleo (MCU, Gravador, Streaming) faz uso de uma solução comercial fechada, fornecida pela empresa Polycom. Em função dessa característica não foi possível obter todas as informações previstas no mapeamento. Apesar disso, é importante relacioná-la neste capítulo por ser um dos principais serviços de videocolaboração ofertados pela RNP.

2.2. Serviços experimentais ou em fase de pré-produção

2.2.1. AAAS 2.0 (Acessibilidade como um serviço versão 2)

A audiodescrição (AD) é um importante recurso de acessibilidade para pessoas cegas. Ela consiste na narração de ações, imagens e detalhes visuais e é essencial para inclusão das pessoas cegas em experiências visuais como vídeos, aulas, televisão, cinema, etc.

Essa solução permite a criação automática e semi-automática de audiodescrição de vídeos. O objetivo é expandir o conceito de Acessibilidade como um Serviço, explorado no projeto “GT-AaaS”, criando um serviço capaz de gerar trilhas de audiodescrição em vídeos, de forma automática ou semi-automática, tomando como base a análise do vídeo ou do seu roteiro.

A atual fase de desenvolvimento prevê a incorporação de estratégias de reconhecimento de objetos em imagens utilizando técnicas de aprendizado profundo para diminuir a dependência do roteiro, legenda e dos metadados relacionados ao vídeo.

2.2.2. Multipresença

O sistema videocolaboração que tem por objetivo permitir a interoperação de forma transparente entre diversas tecnologias e padrões de comunicação, tais como:

- Sala de telepresença em ultra alta definição (UHD);
- Troca de conteúdo entre os participantes remotos;
- Acesso pelos sistemas de videoconferência legados (Polycom, Cisco e outros);
- Acesso em alta definição por programa aplicativo no computador pessoal;
- Acesso por webconferência;
- Acesso por dispositivos móveis e fones SIP.

Os principais sistemas integrados são: i) o PRAV Player, originário de um sistema de sala cirúrgica do laboratório do PRAV (Projetos em Áudio e Vídeo) da UFRGS; ii) o Mconf, software de webconferência utilizado no serviço Conferência Web; e iii) o middleware SAGE2 (*Scalable Amplified Group Environment*), apresentado na seção 2.3.1.

A interoperação entre diferentes tecnologias junto ao sistema de controle pelo *tablet* facilita a criação de salas multiuso, onde é possível configurar o leiaute do ambiente rapidamente. Isso permite, por exemplo, os seguintes modelos: a) telepresença; b) dinâmica de grupo; c) reunião presencial ou remota; d) evento sendo transmitido; e) aula remota; f) outros. Essa possibilidade otimiza os custos, visto que a sala será multifuncional e poderá ser utilizada com mais frequência em relação a salas dedicadas com leiaute fixo.

O sistema Multipresença adapta-se automaticamente ao número de monitores ou TVs conectadas às máquinas que executam o sistema. Dessa forma, é possível ter um painel com 4 TVs e outro com 8 TVs, mantendo a interação entre ambos. Além disso, é possível a integração também com pessoas que possuem um simples notebook, ou mesmo um dispositivo celular, onde é feita a escolha do sinal que se quer visualizar naquele instante. Em 2017, o projeto iniciou a fase experimental, em que instituições selecionadas estão validando o sistema.

2.2.3. BAVi: Busca avançada por vídeos baseada em transcrição de áudio, metadados e anotação semântica

Ainda em desenvolvimento, a solução tem como objetivo implantar uma arquitetura distribuída e escalável que permite a transcrição de áudio (oriundos de vídeos ou não), a anotação semântica de textos (transcritos ou naturais) e a recomendação dos textos anotados semanticamente, visando à categorização desse conteúdo (conforme a DBpedia) e relacionamento com outros conteúdos digitais do mesmo serviço.

Todas essas etapas podem estar encadeadas, ou serem chamadas de forma isolada, via API. Neste cenário, será possível agregar valor a conteúdos digitais disponibilizados por serviços da RNP, como o vídeo@RNP, através da ampliação dos termos de busca (quantidade e relevância) e relacionamento entre conteúdos do mesmo serviço, permitindo uma maior visibilidade a esses conteúdos.

2.2.4. Busca automática de vídeos

O software para busca automática de vídeos (*Video Search*), permite a localização automática de vídeos dentro de um repositório a partir de uma amostra desses vídeos.

Essa ferramenta foi desenvolvida em parceria com a empresa Dynavideo, no âmbito da cooperação com a Agência Nacional do Cinema (Ancine), com a finalidade de se integrar à plataforma MP-SeAC para localizar obras audiovisuais em um acervo gravado, verificar se uma obra foi veiculada em um determinado canal, comprovar a veiculação de uma obra declarada pelo canal, dentre outras possibilidades.

2.2.5. Sistema de captura do sinal da TV Aberta

A parceria entre a RNP e a Agência Nacional do Cinema (Ancine) teve como um de seus desdobramentos o desenvolvimento inovador de uma placa PCI-Ex que permite a captura e decodificação de até 8 canais simultaneamente de TV Digital Aberta. O equipamento compõe uma solução inovadora para captura simultânea de múltiplos canais de TV Digital aberta, associada a uma arquitetura para processamento, indexação, armazenamento local e transmissão para armazenamento remoto desse conteúdo, o que

foi amplamente testado em laboratório para auxiliar a Ancine na formação de um acervo de obras audiovisuais publicitárias veiculadas nos canais de TV Digital, e a RNP, na disponibilização desse conteúdo para fins de ensino e pesquisa.

A placa foi desenvolvida entre março e dezembro de 2016, por uma equipe formada por pesquisadores da Universidade Estadual de Campinas (Unicamp), da Universidade de São Paulo (USP) e do Instituto Presbiteriano Mackenzie.



Figura 11.3. Placa OctoTuner - permite a captação e decodificação de até oito canais de TV simultaneamente

2.2.6. STB Scan

O STB-Scan é uma ferramenta de coleta de sinais, estatística de acesso, auxílio à predição de cobertura do sinal digital terrestre e diagnóstico para instalações de TV Digital, como API para Middleware Ginga. Essa ferramenta é capaz de coletar amostras de sinais moduladas em OFDM (*Orthogonal Frequency Division Multiplexing*), além de verificar o número de pacotes não corrigidos pelos decodificadores corretores de erro, detectar nível de sinal, relação sinal ruído, taxa de transmissão, tipo de modulação, entre outros parâmetros que irão auxiliar em um diagnóstico da qualidade do sinal, da instalação e do canal de transmissão. A coleta de dados será feita por amostragem durante o período em que o telespectador estiver sintonizando determinada frequência UHF (*Ultra High Frequency*), o tempo de sintonia também será utilizado para a estatística de audiência em determinados canais.

Os dados coletados poderão ser enviados para um servidor de dados via internet utilizando diferentes tecnologia de acesso. O pós-processamento dos dados enviados de diferentes regiões irá auxiliar como calibrador de ferramentas de predição, fornecendo informações precisas do sinal recebido em diferentes ambientes e localizações. Os dados coletados poderão ser utilizados como resposta em tempo real para qualquer ajuste na transmissão, analisando o desempenho da rede e indicando os melhores parâmetros de modulação para a melhor cobertura de sinal, auxiliando desta forma na implantação do

SBTVD Sistema (Brasileiro de TV Digital) e desenvolvendo o middleware Ginga a partir da criação e desenvolvimento de API's (Interface de Programação de Aplicativos).

2.2.7. CDN (*Content Distribution Network*)

Também é importante mencionar neste mapeamento a existência de uma prova de conceito de CDN, desenvolvida entre 2012 e 2014 a partir de uma cooperação com o antigo Ministério das Comunicações (MC), atual MCTIC.

Instalada em 10 PoPs (pontos de presença) e com um servidor central para monitoramento e controle, a CDN é composta por um conjunto de servidores distribuídos que disponibilizam vídeos e páginas da web aos usuários com base em sua localização geográfica, reduzindo a latência e aliviando a carga de tráfego na rede.

2.3. Produtos alocados sob demanda

2.3.1. Painel de Colaboração e Visualização SAGE2

O middleware SAGE2⁸ (*Scalable Amplified Group Environment*) é uma solução desenvolvida pelas universidades de Illinois em Chicago e do Haváí (EUA), com financiamento da agência federal norte-americana de fomento à pesquisa (NSF) e disponibilizado gratuitamente para usos não comerciais. A solução provê um ambiente colaborativo que permite a usuários locais e remotos acessar, visualizar e compartilhar múltiplos conteúdos via rede IP. A interação é controlada por meio de um navegador (Chrome ou Firefox) executado em notebooks, *smartphones* ou *tablets*.

O nome “Painel de Colaboração e Visualização” foi cunhado pela comunidade brasileira de usuários do SAGE2, com a finalidade de deixar mais explícito o propósito da ferramenta e aproveitando do fato da maioria das instalações existente no Brasil até então serem compostas de um painel (ou vídeo-wall) com 4 monitores full-HD.

Essa solução objetiva atender à necessidade crescente de novas ferramentas que permitam o trabalho colaborativo em rede, tanto para conteúdo de alta definição como para exibição simultânea de múltiplos conteúdos, cobrindo dois cenários comuns que não são atendidos pelas ferramentas tradicionais de colaboração:

- Colaboração em reuniões locais, com visualização e análise de múltiplos conteúdos simultâneos, provenientes de um ou mais usuários (por exemplo, análise em paralelo de conteúdos de diferentes participantes de uma reunião presencial ou de múltiplos alunos em sala de aula compartilhando a tela de seus notebooks);
- Colaboração remota, que envolve grupos de usuários distribuídos geograficamente, com fontes de dados locais e remotas (por exemplo, dois ou mais grupos que analisam dados provenientes de sensores, simulações e experimentos). Ou, por exemplo, um professor ou palestrante remoto que compartilha, visualiza e manipula arquivos em um Painel SAGE2 via rede IP.

⁸ <http://sage2.sagecommons.org>

Painéis de duas instituições parceiras também podem ser interligados, para permitir o espelhamento de conteúdo.

Atualmente, a RNP auxiliou na instalação do SAGE2 em 11 instituições brasileiras⁹, sendo a segunda maior comunidade mundial de usuários.

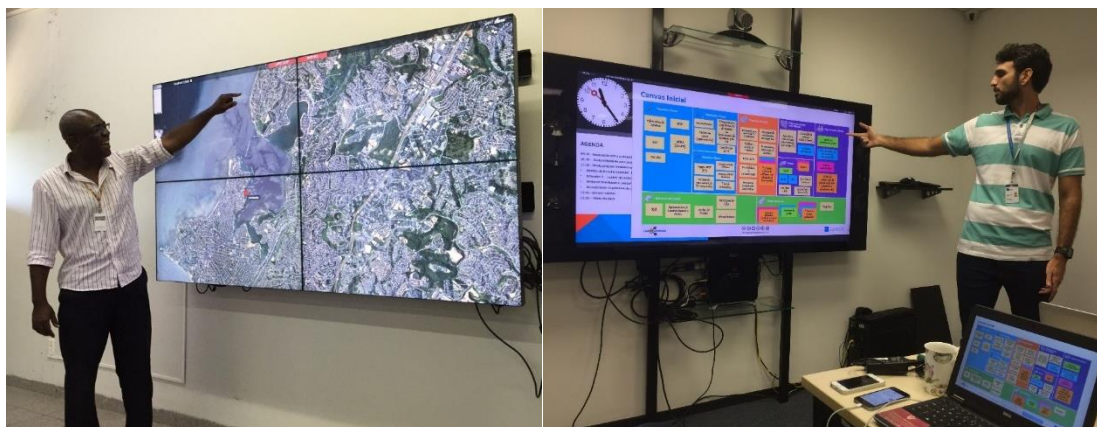


Figura 11.4. Instalações do Painel na Universidade Católica de Salvador e RNP-Rio.

2.3.2. Arthron

O Arthron é um sistema que facilita a execução de performances artísticas que utilizam representações midiáticas e o compartilhamento de espaços reais e virtuais, promovendo a convergência entre arte e tecnologia.

A solução é composta por uma infraestrutura de hardware e software com gerência remota para a captura e distribuição segura de múltiplos fluxos simultâneos de áudio e vídeo, a fim de prover suporte a diversos cenários de videocolaboração, em especial para o suporte a manifestações artísticas e culturais distribuídas.

2.3.3. Fogo Player 4K

O Fogo é uma suíte de software responsável pela manipulação de vídeos UHD. O objetivo da plataforma é reduzir o custo computacional associado à manipulação deste tipo de mídia. Para tal o Fogo implementa uma arquitetura distribuída e escalável que mantém a plataforma no estado da arte das altas definições. Em 2016 a solução foi empregada para reproduzir vídeos na resolução 8K (7680 × 4320 pixels) com áudio em 5.1 canais.

Adicionalmente, a suíte de software conta com uma variedade de ferramentas implementadas para facilitar a experiência do usuário que opera a solução, tais como diversos codificadores e decodificadores de vídeo, ferramentas para análise de sincronia de quadrantes e testes da qualidade da rede.

A empregabilidade dessa solução tem se destacada na área de telessaúde, em especial na formação de novos profissionais da saúde, pois propicia que seja visualizado

⁹ <https://wiki.rnp.br/pages/viewpage.action?pageId=89582755>

em alta definição e em grandes dimensões detalhes de órgãos do corpo humano durante um procedimento cirúrgico, por exemplo. A solução também propicia a redução do risco de infecção hospitalar, com a diminuição de pessoas no bloco cirúrgico¹⁰.

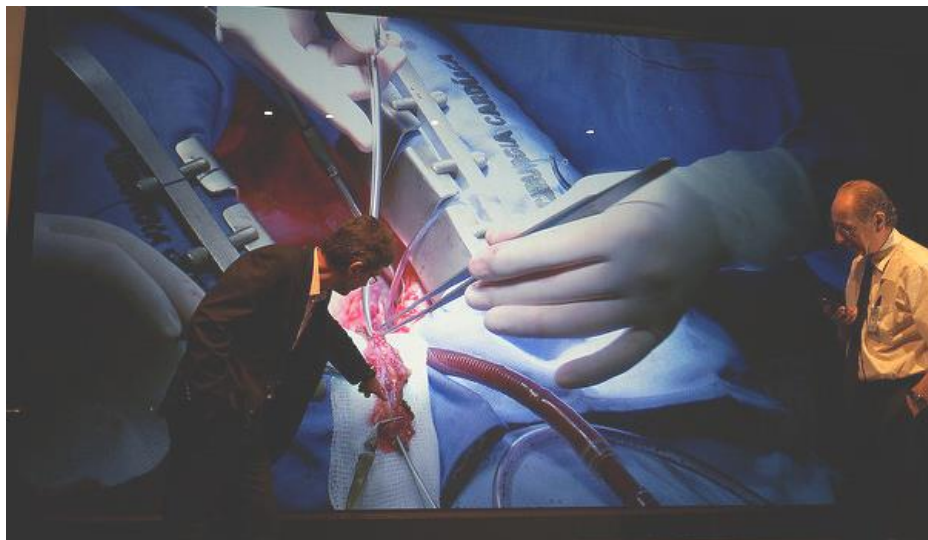


Figura 11.5. Uso do Fogo Player para transmissão 4K de cirurgia realizada no Hospital Universitário da UFRN (HUOL) em 2013.

2.3.4. Ambiente Colaborativo em UHD para Procedimentos de Alta Complexidade

É uma solução integrada que combina elementos do Fogo Player, a ferramenta Arthron e o ICD, motivada por uma demanda de transmissão de procedimentos complexos na área da saúde. Em particular neste momento o Hospital Universitário Lauro Wanderley (HULW) da UFPB, busca uma nova plataforma de videocolaboração em Ultra Alta Definição (UHD) aplicado ao Ensino de Procedimentos Complexos.

A Gerência de Ensino de Pesquisa do HULW (GEP), hospital da rede EBSE¹¹, vem investigando com sua comunidade acadêmica da área de saúde a validade e impacto da transmissão em 4K, em procedimentos cirúrgicos e procedimentos de curativos de feridas complexas, como estudos de caso. Foram instaladas duas câmeras na sala do bloco cirúrgico, proporcionando maior interatividade com discentes, docentes e técnicos da área da saúde que se encontram no auditório. Trata-se do processo de consolidação da área de telemedicina no âmbito do hospital, onde o uso de tecnologia é uma ferramenta pedagógica que pode contribuir com a formação em saúde.

Imagens em alta definição e em tempo real têm potencial para possibilitar que os discentes vejam o procedimento com detalhes com maior segurança ao paciente, e com isto, melhorem a sua compreensão do processo. Enquanto hospital universitário, a GEP quer fazer parte de um cenário onde as tecnologias digitais possam a ser pensadas como um recurso que possa potencializar em nossos alunos a capacidade de compreensão.

¹⁰ <https://www.rnp.br/noticias/cirurgia-4k-abre-1o-mutirao-nacional-rede-ebserh>

¹¹ Empresa Brasileira de Serviços Hospitalares (www.ebserh.gov.br) é uma empresa pública vinculada ao Ministério da Educação.

2.3.5. Cacuriá

Cacuriá é uma ferramenta de autoria, projetada para permitir que o próprio professor possa criar objetos de aprendizagem multimídia sem a necessidade de conhecimentos prévios de programação.

Seu desenvolvimento foi originado em um projeto de P&D denominado VoA (Vídeos sob Demanda como Objetos de Aprendizagem), com o propósito de permitir a criação de objetos de aprendizagem com conteúdo multimídia interativo para o uso no ensino a distância (EaD).

2.3.6. AAAS (Acessibilidade como um Serviço)

Solução para tradução automática e semi-automática de conteúdos digitais em Português para Libras. Nesse projeto, é explorado o conceito de Acessibilidade como um Serviço, através da modelagem e desenvolvimento de um serviço para viabilizar o acesso de usuários surdos a conteúdos digitais multimídia. Através do serviço é possível traduzir vídeos digitais, sob demanda, para Libras (língua brasileira de sinais) usando a legenda ou um áudio em Português como entrada. Um *plugin* para navegadores (Google Chrome, Firefox e Safari) faz tradução automática dos conteúdos solicitados pelos usuários em páginas da web. Como forma de atender às preferências de cada usuário, eles podem ainda configurar parâmetros como o tamanho da janela de Libras e a velocidade de sinalização do avatar que apresenta a tradução para Libras.

O projeto teve como desdobramento a criação da Suíte VLibras¹² que hoje é a solução oficial do Governo Federal utilizada para tradução automática do conteúdo dos seus sites. Atualmente o VLibras está sendo utilizado em mais de 1500 sites, incluindo os sites do Governo Federal, Câmara Federal e Senado Federal, acessíveis para pessoas surdas, e possui mais de 100 mil downloads.



Figura 11.6. Avatar apresentando a tradução para Libras.

¹² <http://vlibras.gov.br/>

3. Mapeamento dos componentes tecnológicos

Um dos principais motivadores deste trabalho foi a ausência de diretrizes (ou “*guidelines*”) para orientar o desenvolvimento de novas funcionalidades em serviços/soluções existentes, em especial quando pretende-se fazer isso por meio de chamadas abertas de programas de P&D.

Adicionalmente, o Comitê de prospecção tecnológica em Videocolaboração (CT-Video)¹³, em anos anteriores, já havia elencado a necessidade de realizar um trabalho de mapeamento dos padrões e componentes tecnológicos adotados pelas diversas soluções de videocolaboração fomentada pela RNP, com o intuito de identificar lacunas de desenvolvimento e facilitar a interoperabilidade entre as soluções. Uma chamada aberta foi lançada em 2015¹⁴, mas na ocasião não houve alunos interessados em realizar o mapeamento.

A reformulação do Programa de Grupos de Trabalho em 2017, apresentada no capítulo 1, trouxe novo senso de urgência para essa necessidade. A seguir é descrito o trabalho realizado.

3.1. Metodologia

Apesar da RNP ofertar diversos tipos de soluções, como segurança, conectividade, gestão de identidade e outros, para o trabalho aqui apresentando foi escolhido apenas o subconjunto de soluções relacionadas à videocolaboração.

Definido este recorte, foram estipuladas as seguintes premissas:

1. Coloque-se no lugar do desenvolvedor: O mapeamento deveria buscar responder as dúvidas técnicas primordiais que um desenvolvedor precisa saber para propor um projeto de evolução de uma solução existente.
2. Simplicidade e informação visual: o resultado do mapeamento deve ser comunicado na forma de um mapa visual, a exemplo do pôster “Network Communication Protocols Map” (Javvin 2007).

Cada serviço em operação, solução tecnológica ou produto em desenvolvimento possui um gerente ou coordenador de P&D responsável na RNP. Este, por sua vez, coordena as atividades da equipe externa que desenvolve e/ou faz manutenção na respectiva solução. Essas equipes são formadas por parceiros da academia ou por empresas fornecedoras, em sua maioria pequenas empresas ou *startups*. O primeiro na construção do mapeamento foi realizar um levantamento interno na RNP de quais soluções seriam incluídas no mapa, independentemente do nível de maturidade, e consequentemente identificar os gerentes e desenvolvedores de cada solução.

Para atender à primeira premissa, foi elaborado um questionário para coleta de informações (vide Anexo I) contendo um conjunto de perguntas técnicas a serem respondidas pelo desenvolvedor de cada solução. Foi tomado o cuidado de deixar o formulário de fácil preenchimento, de forma a não onerar demasiadamente o tempo do respondente. Uma primeira versão do formulário foi enviada para a avaliação dos membros do CT-Video. Um conjunto de sugestões de melhorias foi incorporada numa segunda versão do formulário, que por sua vez foi submetido para revisão de um conjunto

¹³ <http://ctvideo.rnp.br/>

¹⁴ <https://wiki.rnp.br/display/ctvideo/Chamada+de+projetos+-+2015>

de gerentes da RNP. Nessa segunda iteração foi sugerida a inserção de algumas questões não-técnicas, como informações sobre registro de software.

Uma vez finalizado o formulário, o mesmo foi encaminhado para que cada gerente e coordenador de P&D ficasse responsável pelo preenchimento junto ao seu respectivo parceiro desenvolvedor.

O mapa resultante foi diagramado na forma de uma tabela, com cada solução apresentada em uma coluna e cada questão do formulário representada a em uma linha. As Tabelas 1-a, 1-b, 2 e 3 apresentam parte da informação disponível no mapa, de maneira textual para facilitar a leitura neste livro.

Tabela 1-a: Parte das informações disponíveis no mapa sobre os serviços em produção

Solução	Video@RNP	Videoaula@RNP (sistema RIO)	Conferência Web (Mconf)	Fone@RNP
Versão	E.201708.1	2.21.21	Mconf-Live: 1.0 Mconf-Web: 2.4.0-5	Gateway: 1.0.115 PBX: 2.0.178 SRL: 1.0.101 SRC: 1.0.62
Licença	Principais componentes: Proprietária Demais módulos: GPL v2, LGPL v2.1, Apache v2.0 e BSD	QPL	Mconf-LB: Proprietária Demais módulos: LGPL, AGPL	Ainda não definida
Resolução máxima	Full HD (1920 x 1080)	Full HD (1920 x 1080)	HD (1280 x 720)	-
Interface para o usuário	Navegador web	Navegador web	Navegador web	Navegador web
APIs implementadas	API REST para requisições de aplicações externas	SAML para comunicação com a CAFe	1) API de gerenciamento de salas. 2) API para envio de comandos e recebimento de callbacks do cliente Flash 3) API para registrar no Mconf-Live e receber callbacks	API para criação e manipulação de ramais e usuários no PBX-IP
Formatos de codificação de vídeo	H.264, VP6, VP8, WMV3, HLS	H.264, VP8	H.264, VP8	-
Protocolos de transmissão	HTTP(s)	HTTP(s), RIO	TCP, UDP, RTMP, RTP	RTP
Bibliotecas ou ferramentas utilizadas	Ffmpeg, Varnish, Grafana, Collect, OpenTSDB, Pentaho BI, Mondrian, Pentaho Data Integration	Ffmpeg, Pentaho	Ffmpeg, Nginx, Tomcat, Grails, Red5, Akka, Monit, FreeSWITCH, Redis, LibreOffice, SWFTools, ImageMagick, GhostScript, Node.js, D3.js, GeolP Lite, Resque	Asterisk, OpensipS, bind
Linguagens de programação	Java	C, C++, Java, Javascript. HTML5	Java, Scala, Ruby on Rails, Javascript, ActionSript	C, Perl, PHP
Banco de dados	PostgreSQL, HBASE	-	MySQL	MySQL, OpenLDAP
Sistema Operacional	Ubuntu (14.04)	Ubuntu (14.04)	Ubuntu (14.04)	Ubuntu (14.04)
Fornecedor/Desenvolvedor	TVoD	B&M	Mconf Tecnologia	PoP-SC

Tabela 1-b: Parte das informações disponíveis no mapa sobre os serviços em produção (continuação)

Solução	Rede de Vídeo Digital (RVD)	Intercâmbio de Conteúdos Digitais (ICD)	Rede de Cinemas: Exibidor de Conteúdo Digital	Rede de Cinemas: Controlador de Sessões
Versão	E.201708.1	4.3.13	1.0	1.0
Licença	Principais componentes: Proprietária Demais módulos: GPL v2, LGPLv2.1 e Apache v2.0	GPL v3	GPL v3	GPL v3
Resolução máxima	Definida pela aplicação que usa a RVD	4K (3840 x 2160)	4K (3840 x 2160)	Limitada pelo Exibidor de Conteúdo
Interface para o usuário	Definida pela aplicação que usa a RVD	Navegador web	Software instalado	Software instalado
APIs implementadas	OCP (comunicação segura entre RVD e aplicação cliente)	APIs para o gerenciamento de conteúdos e autenticação e autorização de usuários.	-	-
Formatos de codificação de vídeo	MPEG, WMV	MPEG-1, MPEG-2, MPEG-4, H.264, H.265/HEVC, WMV, VP8, VP9 Theora Real Video	H.264, MPEG-4	-
Protocolos de transmissão	TCP, UDP, HTTP(s), RTMP, HLS, MPEG-DASH	HTTP(s), SCP	TCP	TCP, UDP
Bibliotecas ou ferramentas utilizadas	FFmpeg, Grafana, Collect, OpenTSDB, VLC	FFmpeg, PrimeFaces, GlassFish 4.1	FFmpeg, JlibCPP, libvlc, libpthread	Libssh, Libcurl, jlibcpp, libvlc, QT, qcreator
Linguagens de programação	Java, Python	Java, C++	C++	C++
Banco de dados	PostgreSQL, HBASE	PostgreSQL, H2	-	-
Sistema Operacional	Ubuntu (14.04) Dynaencoder: Windows 10	Ubuntu (14.04) Red Hat (6.x)	Ubuntu (12.04)	Ubuntu (12.04)
Fornecedor/Desenvolvedor	TVoD	Dynavideo	Dynavideo	Dynavideo

Tabela 2: Parte das informações disponibilizadas no mapa sobre os serviços experimentais

Serviço experimental	Multipresença	AAAS 2.0	BAVi	Busca Automática de Vídeos	Sistema de captura do sinal da TV Aberta	STBScan
Versão	1.2.1.13	1.1	1.0	vhc-gen: 2.0.0 vhc-searcher: 1.0.5	2.0	1.0
Licença	LGPL	Ainda não definida	Ainda não definida	GPL v3	Proprietária	Proprietária
Resolução máxima	4K (3840 x 2160)	Full HD (1920 x 1080)	N/A	Full HD (1920 x 1080)	Full HD (1920 x 1080)	Full HD (1920 x 1080)
Interface para o usuário	Navegador web	Navegador web	WebService REST	WebService REST	Navegador web	Navegador web
APIs implementadas	-	API para solicitação de geração de audiodescrição	API para envio e consulta de requisições	-	API RESTful Service	API RESTful Service
Formatos de codificação de vídeo	H.263, H.264	H.264, MPEG-2, WMV	-	H.264, MPEG-TS	H.264, H.265	H.264
Protocolos de transmissão	RTP	HTTP	HTTP	-	HTTP, UDP, SFTP	HTTP, SOAP, CoAP
Bibliotecas ou ferramentas utilizadas	FFmpeg, QT, OpenGL, NodeJS, RubyonRails, PJSIP, Directshow	FFmpeg, NodeJS, Apache Lucene	FFmpeg, SOX, Plotly, Apache MQ	libavcodec, libavformat, libavutil, libswscale, libswresample, libpthread, libboost_regex, libfftw3, libpq	FFmpeg, VLC, GENTLELLA, JAX-B, JERSEY, MediaInfo	Google Maps
Linguagens de programação	C, C++, Ruby, Javascript	C++, Python	Java, PHP, Scala	C++	Java, C, VHDL, PHP, Javascript	Java, C, PHP, Javascript, Ginga
Banco de dados	PostgreSQL, SQL Lite	MongoDB	Blazegraph, MySQL	PostgreSQL	PostgreSQL	PostgreSQL, MySQL, Oracle
Sistema Operacional	Ubuntu (14.04)	Ubuntu (14.04)	Ubuntu (14.04)	Ubuntu (12.04)	Ubuntu (16.04)	Ubuntu (16.04)
Fornecedor/Desenvolvedor	UFRGS	UFPB	UFJF	Dynavideo	Kasko P&D Tecnologia	Kasko P&D Tecnologia

Tabela 3: Parte das informações disponibilizada no mapa sobre as soluções que são alocadas sob demanda.

Solução	Painel de Colaboração e Visualização SAGE-2	AAAS	Arthron	Fogo Player	Ambiente colaborativo em UHD para procedimentos de Alta Complexidade	Cacuriá (VoA)
Versão	2.0.107	4.0	1.0.5	1.0	1.0	1.6
Licença	Baseada em BSD	GPL v3	GPL v3	GPL v3	GPL v3	Ainda não definida
Resolução máxima	Escalável nK	Full HD (1920 x 1080)	Full HD (1920 x 1080)	8K (7680 x 4320)	8K (7680 x 4320)	Full HD (1920 x 1080)
Interface para o usuário	Navegador web	Navegador web	Software instalado	Software instalado	Software instalado	Software instalado
APIs implementadas	-	API RESTful Service	-	-	-	-
Formatos de codificação de vídeo	-	H.264, MPEG-2, WMV	MPEG-1, MPEG-2, MPEG-4, H.264	MPEG-2 TS	MPEG-2 TS	-
Protocolos de transmissão	HTTP, WebSocket	HTTP	UDP, TCP	UDP, TCP	UDP, TCP	-
Bibliotecas ou ferramentas utilizadas	FFmpeg, Node.js, OpenGL, ImageMagick, Google Chrome	FFmpeg, Unity, Node.js, Aelius POS-Tagger	VLCJ, libVLC, Jung, JFreeChart, NimROD, Look & Feel	FFmpeg, OpenGL, SDL	FFmpeg, OpenGL, SDL, Libssh, libcurl, jlibcpp, libvlc, QT, qtcreator	QT
Linguagens de programação	Electron, Javascript	C+, Python, C#, Javascript	Java, C++	C++	C++	C++
Banco de dados	-	PostgreSQL, MongoDB	-	-	-	-
Sistema Operacional	Ubuntu (14.04) OpenSuSE (13.x) Windows 10	Ubuntu (14.04) Windows 7	Ubuntu (12.04)	Windows 7	Ubuntu (12.04) Windows 7	Ubuntu (16.04) OpenSuSE (42.1) Red Hat (7.2) Windows 7
Fornecedor/Desenvolvedor	UIC - VL	UFPB	Dynavideo	Dynavideo	Dynavideo	UFMA

3.2. O mapa

Além do conteúdo apresentado nas Tabelas 1-a, 1-b, 2 e 3, a versão final do mapa também inclui informações sobre a instituição desenvolvedora da solução e sobre a “genealogia da solução”, isto é, o histórico de projetos de pesquisa que deram origem à solução de video colaboração. A Figura 11.7 mostra a versão final do mapa após diagramação. O tamanho original para impressão é 170x84cm. Uma versão em alta resolução do mapa pode ser descarregada através do site do CT-Video (<http://ctvideo.rnp.br>) ou solicitada através do endereço de e-mail contato@ct-video.rnp.br

Mapeamento das soluções de video colaboração da RNP (2017)																	
	Em produção						Experimental ou pré-produção						Sob demanda ou "na prateleira"				
	Projeto RNP	Projeto RNP (Solução RNP)	Projeto RNP (Solução RNP)	Projeto RNP (Solução RNP)	Projeto RNP (Solução RNP)	Projeto RNP (Solução RNP)	Projeto RNP (Solução RNP)	Projeto RNP (Solução RNP)	Projeto RNP (Solução RNP)	Projeto RNP (Solução RNP)	Projeto RNP (Solução RNP)	Projeto RNP (Solução RNP)	Projeto RNP (Solução RNP)	Projeto RNP (Solução RNP)	Projeto RNP (Solução RNP)	Projeto RNP (Solução RNP)	
Descrição	
Objetivo	
Impacto	
Tipologia de Solução	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Identificação	
Identificação para usuários	
URL	
Identificação de parceiros	
Identificação de patrocinadores	
Identificação de colaboradores	
Identificação de parceiros	
Identificação de patrocinadores	
Identificação de colaboradores	

Figura 11.7. Visão geral do mapeamento, diagramado na forma de um pôster de 170 X 84cm

4. Visão de futuro

O resultado do trabalho aqui apresentado deve encorajar novos mapeamentos de outros tipos de soluções da RNP. É desejado também que o mapa de soluções de videocolaboração seja constantemente atualizado, em periodicidade a ser definida, e que cumpra seu papel de fornecer informações para facilitar o desenvolvimento de novas funcionalidades e a interoperabilidade entre as soluções. Indo mais além, ao identificar lacunas de desenvolvimento, o mapa poderá auxiliar na elaboração de *roadmaps* para a evolução dos serviços de videocolaboração.

Enquanto não existe um *roadmap* integrado, as equipes de cada serviço planejam individualmente suas respectivas listas de novas funcionalidades a serem implantadas. A seguir é apresentada algumas das funcionalidades previstas para serem implementadas em 2018.

4.1. Video@RNP

- Robô YouTube (disponibilizada máquina para aplicação na Cloud e finalizada homologação. Para apontar para produção, verificada necessidade de atualização do portal de produção com correção necessária para pegar adequadamente bit rate de vídeos MKV).
- Gateway RTMP para suporte a transmissões ao vivo via Push.
- Ajuste player canais de rádio.
- Grade de Programação nos Canais.
- MPEG DASH - Troca de Player.
- Melhorias Canais de TV e Rádio.
- Melhorias relatórios - Suporte a medição de audiência simultânea.
- Melhorias no chat.
- Integração com solução BAVi.

4.2. Conferência Web

- Controle de permissão para visualização de uma gravação (atualmente qualquer um pode assistir uma gravação, se tiver acesso ao link desta. É um risco se o conteúdo for classificado.)
- Tornar o teste de uso de uma sala de webconferência mais simples e acessível (a ferramenta de teste para avaliar o ambiente do cliente é muito técnica e pouco amigável. Não ajuda os clientes mais leigos.)
- Exibir quota de gravação da instituição por hora no portal (a quota atual é exibida em GBytes, o que não dá uma ideia exata de tempo de gravação.)
- Integração com o serviço Filesender@RNP (permitir que as gravações das salas de reunião sejam enviadas através do Filesender de forma integrada sem a

necessidade de qualquer interação manual. O usuário deverá enviar essa gravação através de um único clique em um botão.)

- Integração com o serviço `edudrive@RNP` (permitir que os clientes façam backup das gravações no serviço de armazenamento em nuvem da RNP. Pode funcionar como uma extensão da cota de gravação (cota pessoal de gravação))
- Integração com o serviço `Videoaula@RNP` (Muitos clientes usam o serviço para entregar Ensino a Distância e gravam as aulas para que os alunos consumam o conteúdo. Exportar uma gravação do Mconf no formato reconhecido do portal do serviço `Videoaula@RNP` vai reforçar a utilidade de ambos os serviços e facilitar a localização das aulas gravadas.)
- Administrador institucional remover usuário "não aprovado" (o administrador institucional não tem a opção de remover um usuário que se autenticou pela CAFe, mas que não terá o acesso liberado. Isto dificulta a gestão do serviço pelo lado do cliente.)
- Exibir informações de uso do serviço para o cliente (Atualmente os administradores institucionais não tem uma visão do uso do serviço dentro da instituição. É necessário entrar em contato com o Service Desk da RNP e aguardar o envio das estatísticas.)
- Alertar quando a cota de gravações estiver terminando (facilitar o gerenciamento de cota de gravação)
- Proibir usuário ou comunidade sem instituição associada (atualmente é possível desassociar uma instituição de uma comunidade ou usuário. O resultado prático é a criação de objetos órfãos no portal do serviço. Precisamos manter a integridade relacional das entidades existentes no sistema.)
- Permitir *login* federado nas salas virtuais (um convidado pode informar qualquer nome ao acessar uma sala de webconferência. Não é possível garantir a identidade dele.
É uma forma de garantir a identidade dos participantes de uma reunião que possuem credencial de acesso federado em uma instituição que já aderiu a CAFe.)
- Notificar cliente quando instituição não está cadastrada no portal (melhorar a mensagem, com objetivo de tornar mais claro, o motivo de uma autenticação federada bem-sucedida não ter sucesso de acessar o portal).
- Lista única de todas as reuniões com acesso as gravações (melhorar a experiência de uso do usuário, facilitar visualização das informações e simplificar o uso do serviço).
- Integração com SIP (convergir os serviços de Comunicação e Colaboração da RNP, permitindo acessar videoconferências através do Serviço de Conferência Web).

4.3. Fone@RNP

4.3.1. Para o SRC - SIP Router Central

- Gerar estatísticas de confiabilidade de um “*peering*” que poderá ser construída com (Disponibilidade e % de completagem de uma rota) - Base de reputação de rotas destino.
- Suporte a IPv6 (administração do ambiente e encaminhamento de chamadas).
- Estudo de viabilidade para o suporte a SIPS e autenticação mais robusta dos “*peerings*” (IP+User+Senha+TLS(se for o caso)).
- Melhorias na segurança (fail2ban, por ex.) sem integração ao ambiente. Será utilizado para segurança IP e não SIP.
- Monitoramento SNMP através de MIB privada estendida.
- Reorganização da interface gráfica e ajustes na nomenclatura utilizada.

4.3.2. Para o SRL - SIP Router Local

- Todas as melhorias do SRC
- Melhorias do media Proxy (Sempre com SRC (SIM ou Não), Comunicação interna Por Peer/Matriz). Incluir também a possibilidade de indicar um endereço IP diferente do SRL, mas que tenha preenchimento automático com o IP do SRL.
- Testes com media-proxy para integrar IP + IPv6 no SDP.
- Gerar estatísticas de confiabilidade de um “*peering*” que poderá ser construída com (Disponibilidade e % de completagem de uma rota).
- Suporte a atualização via interface gráfica.
- Monitoramento SNMP através de MIB privada estendida.
- Possibilidade de habilitar gravação de ligação (para auditoria).
- Monitoração dos PEERS quanto a disponibilidade e qualidade para ajustes na tabela de rotas.
- Implementar mecanismo de atualização em data e horário predefinido para realizar as atualizações críticas.
- Reorganização, ajustes na nomenclatura utilizada na interface e configuração.
- Importar rotas do SRC associar com o PEER local e aplicar configurações.
- Documentação de instalação operação, para administração e gerência do ambiente.
- Ajuste no filtro da exportação de rotas via ENUM para permitir a coexistência de mais de um GWT/SRL por par (*peer*).

4.3.3. Para o GWT - Gateway Transparente

- Melhoria no algoritmo de reputação.
- Suporte a IPv6.
- Acesso a CLI Asterisk via HTTP.
- Reorganização, ajustes na nomenclatura utilizada na interface e configuração.
- Homologar GWT com maior densidade de portas GWT180, GTW240.
- Implementar e Homologar GWTV (Gateway virtual) com EBS da Khomp até GWT120.
- Homologar GWTV (Gateway virtual) com EBS da Khomp até GWT240.
- Implementar a definição de perfis de configuração da interface E1 considerando, (PBX | operadora), sinalização (ISDN | R2), e outros para facilitar a configuração das interfaces E1 - Wizard - RNP irá consolidar a documentação das instalações feitas no fone@2012.

4.3.4. Para a PBX - IP Educacional

- Suporte a IPv6.
- Possibilidade de habilitar gravação de ligação de ramal (para auditoria).
- Manter log da origem dos registros dos telefones IP. Localização geográficas de onde estão registrados.
- Gerar mapa a partir do log de registro de localização.
- Implementação do Softfone RNP com possibilita de instrumenta-lo para medição da qualidade.

4.3.5. Para a PBX - IP Corporativa

- Suporte a IPv6.
- Homologação de telefones IP de outros fabricantes Khomp, Intelbras e Digtro.
- Homologação de ATA - Nacional - 2 linhas.
- Homologação de ATA (SIP) de grande porte - entre 20 e 40 linhas.
- Homologar função de aprovisionamento de telefones IP de outros fabricantes Khomp, Intelbras e Digtro.
- Implementar na PBX-IP Core a função de aprovisionamento de telefones IP de outros fabricantes Khomp, Intelbras e Digtro.
- Implementar na PBX-IP WEB a função de aprovisionamento de telefones IP de outros fabricantes Khomp, Intelbras e Digtro.
- Implementar a função de aprovisionamento para telefones Polycom para vídeo-chamadas.

- Homologação de telefones Polycom para vídeo chamadas.
- Monitoramento SNMP através de MIB privada estendida.
- Possibilidade de habilitar gravação de chamadas com diversas filtros (para auditoria).
- Implementação da facilidade de chefe-secretária.
- Implementar Portal da telefonista / secretaria.

4.4. RVD

- Implantação do MPEG-DASH

Agradecimentos

A elaboração deste trabalho seria impossível sem a contribuição de todos os parceiros, fornecedores e colaboradores que dedicaram tempo respondendo o questionário para coleta de informações (Anexo I). A seguir o nome de todos os respondentes em ordem alfabética.

- Alex Galhano Robertson (RNP)
- Bernardo Netto (B&M)
- Carlos de Salles Soares Neto (UFMA)
- Celso Capovilla (RNP)
- Denio Mariz (Dynavideo)
- Diogo Gará Caetano (KASCO P&D Tecnologia)
- Diego Pessoa (Dynavideo)
- Eduardo Barrére (UFJF)
- Felipe Cecagno (Mconf Tecnologia)
- Fernando Frota Redigolo (USP)
- Giuliano Maia (Dynavideo)
- Henrique Daniel Ferraz (RNP)
- Jairo F. de Souza (UFJF)
- Maxine Brown (UIC)
- Reinaldo Matushima (TvOD)
- Sindolfo Miranda (Dynavideo)
- Tiago Maritan (UFPB)
- Valter Roesler (UFRGS)

Referências

- Messina, L.A.; Ribeiro Filho, J.L. e Lopes, P. R. Lima. (2014) “RUTE 100: as 100 primeiras unidades de telemedicina no Brasil e o impacto da Rede Universitária de Telemedicina (RUTE)”. 1 ed. – Rio de Janeiro: E-papers, 2014. 506 p. ISBN 978-85-7650-439-9.
- Valle,R., Marins, A, Brauner,D., Stanton, M. e Machado, I. (2014) “Criando Serviços Avançados na América Latina em Colaboração com a Comunidade Acadêmica e as Redes”, In: TICAL 2014
- Ciuffo, L.N., Souza Filho, G.L. et al. (2013) “A Software-Based Solution for Distributing and Displaying UHD Video Content over Packet Networks with Applications to Telemedicine and Culture”, In. UbuntuNet-Connect 2013.

- RNP. (2016) “Relatório de Gestão MCTIC – RNP – edição anual”.
<https://www.rnp.br/aceso-informacao>
- RNP. (2012) “Catálogo de Serviços da RNP”,
http://memoria.rnp.br/_arquivo/servicos/catalogo_servicos_novo.pdf
- Javvin Company (2007). “Network Communications Protocols Map” (URL indisponível)
- Fuks H, Raposo AB, Gerosa MA. (2003) “Do Modelo de Colaboração 3C à Engenharia de Groupware”. Simpósio Bras Sist Multimídia e Web [Internet]. 2003;1:8.
Available at: <http://groupware.les.inf.puc-rio.br/public/papers/Webmedia2003.pdf>

Anexo I: Questionário para coleta de informações

<p>Mapeamento tecnológico do produto/solução:</p> <p>Nome do serviço</p>	<p>Responsável pela solução na RNP: Fulano de Tal</p> <p>Responsável pelo preenchimento: Nononono</p> <p>Data do preenchimento: 15/08/2017</p>
<p>Resumo/pequeno parágrafo explicativo (preenchido pelo gerente da RNP)</p>	
<p>Exemplo: “Middleware que cria um ambiente colaborativo para visualização de múltiplos conteúdos e compartilhamento de telas de usuários locais e remotos via rede IP.”</p>	
<p>Versão atual do código / release</p>	
<p>Identifique a versão do solução na qual as respostas deste questionário serão baseadas. Exemplo: “Versão 1.2.13 “</p>	
<p>Licenças de software</p>	
<p>A versão atual da solução é distribuída sob qual licença? No caso de haver mais de uma licença, explicitar. Exemplo: “Módulo cliente = LGPL. Módulo servidor = Apache 2.0”</p>	
<p>Resolução máxima suportada (se aplicável)</p>	
<p>Caso a solução manipule e exiba imagens/vídeos, qual a resolução máxima suportada? Exemplos: 1280 x 720 (HD) 1920 x 1080 (Full HD) 2048 x 1080 (2K) 3840 x 2160 (4K)</p>	
<p>Interface para o usuário</p>	
<p>Como o usuário acessa a solução? Exemplo: “Software instalado na máquina cliente ou Navegador Web”</p>	
<p>Linguagens de programação</p>	
<p>Informe a linguagem de programação utilizada no desenvolvimento da solução. Caso mais de uma linguagem tenha sido usada no desenvolvimento de diferentes módulos, identifique a linguagem usada em cada módulo/componente. Exemplo: “Interface do cliente = C++”.</p>	

Bibliotecas
<p>Relacione as ferramentas ou bibliotecas de propósito geral, desenvolvidas por terceiros, que a solução faz uso.</p> <p>Exemplos: FFMPEG, OpenGL, NodeJS</p>
APIs
<p>Relacione as APIs implementadas pela solução que são usadas para comunicação com outros componentes / serviços / aplicações (se houver).</p>
Banco de dados
<p>Identifique o software de gerenciamento de banco de dados utilizado pela solução (se houver). No caso de mais de um, informe resumidamente que tipo de informação cada banco de dados armazena.</p> <p>Exemplos: PostgreSQL, SQL Lite</p>
Protocolos de Transmissão
<p>Relacione os protocolos de transmissão empregados.</p> <p>Exemplos: UDP, RTP</p>
Formatos de codificação
<p>Informe os formatos de áudio e vídeo suportados pela solução.</p> <p>Exemplo: "Video = H.264, MPEG-2, WMV. Audio = G.722, AAC, Opus"</p> <p>Se aplicável, faça distinção dos codecs e formatos que a solução recebe como entrada e gera como saída.</p>
Sistema operacional
<p>A solução requer qual sistema operacional (e versão mínima suportada) para ser executada?</p> <p>No caso de uma solução que é executada na Web ou "na nuvem", especificar o sistema operacional que deve ser instalado no servidor/VM da solução.</p> <p>Exemplo: Ubuntu 14.04 , openSUSE 13.x, Windows (8.x/10) 64bit</p>
Registro de software
<p>A solução possui registro de software? Em caso afirmativo, informar o número do processo e a data da criação. Informe também caso o processo de registro tenha sido aberto mas ainda não esteja concluído.</p> <p>Exemplo: BR 91 2016 001572-1 Criação em 1/11/2016</p>

Anexo II: Biografia dos autores



Leandro Neumann Ciuffo

Trabalha na Rede Nacional de Ensino e Pesquisa (RNP) como gerente de P&D para comunidades e aplicações avançadas. De 2011 a 2015 coordenou o Programa de P&D em Aplicações Avançadas de Visualização Remota, que inclui transmissões de conteúdos em ultra-alta definição pela Internet. Atualmente também atua na coordenação de projetos relacionados a ambientes para experimentação, sendo responsável pela gerência da operação do testbed FIBRE (www.fibre.org.br). Entre 2006 e 2009 trabalhou no *Istituto Nazionale di Fisica Nucleare* (INFN, Itália), atuando na coordenação de projetos relacionados à e-Ciência e Grid Computing. Leandro possui mestrado em Computação pela Universidade Federal Fluminense (2005).



Marcelino Nascentes da Cunha

Bacharel em Administração de Sistemas de informação, pós-graduado em Gestão de projetos, com 22 anos de experiência na área de TI, sendo 17 anos implantando e coordenando projetos de vídeo na RNP. Atualmente ocupa a função de gerente de serviços.



Christian Miziara de Andrade

Gerente de Soluções da RNP, é responsável pela gestão de projetos para desenvolvimento e implementação de serviços e soluções para clientes estratégicos da Organização. Possui graduação em Engenharia Elétrica pela Universidade Federal de Goiás (UFG), com especialização em sistemas de comunicação digital pela Universidade de Brasília (UnB), e MBA em Administração Estratégica de Sistemas de Informação pela Fundação Getúlio Vargas (FGV).



Graciela Leopoldo Martins

Mestre em Ciência da Computação pela Universidade de São Paulo (USP) e graduada na mesma área pela Universidade Estadual Paulista (UNESP). Ela tem especialização em Gestão Estratégica de Inovação Tecnológica pela Unicamp desde 2008 e é certificada pelo Project Management Institute (PMI) em gerenciamento de projetos desde 2013. Ela trabalha na RNP desde julho de 2000, onde já atuou como administradora de redes e especialista em aplicações de comunicação e colaboração na Internet. Atualmente, é Gerente de Projetos na RNP e atua na modelagem e implementação de soluções de tecnologia de informação e comunicação (TIC)



Rafael Valle

Coordenador de Pesquisa e Desenvolvimento da Rede Nacional de Ensino e Pesquisa - RNP. Possui graduação em Engenharia de Telecomunicações pela Universidade Federal Fluminense (UFF) e mestrado em Telecomunicações com ênfase em Redes de Computadores pela mesma instituição. Está na RNP desde 2011 e, atualmente, tem coordenado projetos de P&D nas áreas de Internet Avançada, Redes Sem Fio, Redes Definidas por Software, IPTV, Aplicações de Vídeo e Acessibilidade.



Fausto Vetter

Coordenador de Pesquisa e Desenvolvimento (P&D) no Departamento de P&D (DPD) da RNP desde 2011. Ele tem experiência profissional em gerenciamento e operação de redes. Ele tem coordenado grupos de pesquisa e desenvolvimento para entregar serviços inovadores. Antes de entrar na RNP, ele trabalhou na DANTE envolvido principalmente no perfSONAR e na infraestrutura de monitoramento do backbone da GÉANT. Ele é Bacharel em Sistemas de Informação pela Universidade Federal de Santa Catarina (UFSC) e possui um Master of Business Administration (MBA) em Gestão de Projetos pela Fundação Getúlio Vargas (FGV).



Clayton Reis da Silva

Trabalha na Rede Nacional de Ensino e Pesquisa (RNP) como coordenador de P&D para comunidades e aplicações avançadas, desde 2012, coordenando projetos de P&D envolvendo internet avançada, aplicações de visualização avançada e grandes transferências de dados. Clayton possui Bacharel e Mestrado em Computação pela Universidade Federal Fluminense (UFF), respectivamente em 2009 e 2012.



Iara Machado

Diretora Adjunta de P&D na RNP. Suas atividades envolvem colaboração com a comunidade de pesquisa em Redes de Computadores e Sistemas Distribuídos, no desenvolvimento de Redes e Aplicações avançadas. Graduada em Física pela UFRJ e Mestre em Redes de Computadores e Sistemas Distribuídos pela UFF.



Luiz Eduardo de Souza Coelho

Diretor-adjunto de serviços da RNP. Possui MBA em Gerenciamento de Projetos pelo IAG Escola de Negócios da PUC-Rio, cursou engenharia e processamento de dados. Trabalhou com projetos de redes corporativas e segurança em empresas como Alcatel/Xylan, IBM e, por último, Sonicwall.



Antônio Carlos Fernandes Nunes

Diretor Adjunto de Soluções da Rede Nacional de Ensino e Pesquisa (RNP). Mestre em Engenharia Elétrica pelo Grupo de Teleinformática e Automação (GTA) do Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia (COPPE) da Universidade Federal do Rio de Janeiro (UFRJ), e graduado em Engenharia Elétrica com ênfase em Sistemas e Computação pela Universidade do Estado do Rio de Janeiro (UERJ). Como Gerente de Projetos Especiais da Diretoria Geral da RNP conduziu vários projetos interministeriais, sendo o responsável pela estruturação, crescimento e coordenação do Ponto Federal de Interconexão Redes (FIX/PTTMetro de Brasília), e responsável pela criação do Internet Data Center (IDC) da RNP. Como Diretor Adjunto de Gestão de Serviços

da RNP, foi responsável pela estruturação e ampliação das ofertas dos serviços de comunicação e colaboração, disponibilização de conteúdos digitais, gestão de identidade, hospedagem estratégica, e suporte à rede acadêmica. Em 2016 assumiu a Diretoria-Adjunta de Gestão de Soluções onde é responsável pela gestão dos diversos programas e projetos da RNP com parceiros, desenvolvidos a partir da aplicação de conhecimentos técnicos e metodológicos na concepção, desenvolvimento, integração e implantação de soluções inovadoras com o uso intensivo de TIC.



Paulo Roberto de Lima Lopes

Doutor em Ciências, pelo Programa de Pós-graduação em Gestão e Informática em Saúde da Universidade Federal de São Paulo (2013); Mestre em Ciências em Reatores Nucleares de Potência e Tecnologia do Combustível Nuclear (1995). Engenheiro Eletrônico formado pela Faculdade de Engenharia Industrial - FEI (1988). Atualmente, está Gerente de Inovação da Comunidade de Saúde e as Redes colaborativas de Ensino e Pesquisa:

Rede Universitária de Telemedicina - RUTE e Rede Global de Ensino, Pesquisa e Extensão em Nutrição, Soberania e Segurança Alimentar e Nutricional - NUTRISSAN, da Rede Nacional de Ensino e Pesquisa - RNP. Participou em diversos comitês e grupos de trabalho em TIC em Saúde e é especialista do GT2 - Interoperabilidade de Sistemas e Dispositivos da Comissão Especial de Estudos em Informática em Saúde da Associação Brasileira de Normas Técnica - CEE-78/ABNT e da Pesquisa TIC Saúde no CETIC.br/CGI.br.