
Contribuições em Multimídia: codificação de vídeo, personalização e adaptação de conteúdo e televisão digital interativa

Rudinei Goularte

Contribuições em Multimídia: codificação de vídeo,
personalização e adaptação de conteúdo e
televisão digital interativa

Rudinei Goularte

Texto sistematizando o trabalho científico do candidato, apresentado ao Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo, como parte dos requisitos para obtenção do título de Livre Docente, junto ao departamento de Ciências de Computação.

USP – São Carlos
Julho de 2011

Este texto foi elaborado para a participação do autor no Concurso Público de Professor Livre Docente junto ao Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (SCC/ICMC/USP). No texto, é apresentada uma sistematização das pesquisas e dos trabalhos realizados pelo autor na área de Multimídia, com foco em vídeo. Primeiramente, são descritas as contribuições na linha de Codificação de Vídeo, cujos fundamentos são aplicáveis em outras linhas. Em seguida são apresentadas as contribuições na linha de Personalização e Adaptação de Conteúdo, a qual concentra os principais esforços do autor e as publicações mais recentes. Por fim, são descritos os trabalhos na linha de Televisão Digital Interativa, na qual o autor possui importantes colaborações em projetos de pesquisa. Neste documento, são descritos os fundamentos e os pressupostos com os quais as linhas de pesquisa têm sido exploradas pelo autor, destacando os trabalhos desenvolvidos e as contribuições nos últimos oito anos de trabalho acadêmico, correspondendo ao período posterior à conclusão do doutorado. Nesse período, o autor publicou sete artigos completos em revistas, incluindo publicações na *ACM Transactions on Multimedia Computing Communications and Applications* e no *Multimedia Systems* (duas vezes), com Qualis A1 e B1, respectivamente. Publicou também 29 artigos em congressos da área, sendo 12 em eventos internacionais e 17 em eventos nacionais.

This document was elaborated to fulfill the requirements of the author's application for a position of Associate Professor in Multimedia, at the Computer Sciences Department of the Institute of Mathematical Sciences and Computing, University of São Paulo (SCC/ICMC/USP). The text is an attempt to systematize the author's research contribution in the field of Multimedia, focusing on video research. First, it presents the contributions on Video Coding, which includes foundations applied in other areas. Then, it presents the contributions on Content Adaptation and Personalization, which concentrates the author's effort in recent years. Finally, it describes the contributions on the Interactive Digital Television area, in which the author has developed important collaborations in the context of some research projects. This text describes the background and the assumptions which are the basis for the author's research, highlighting the work and contributions accomplished in the last eight years, i.e. after the conclusion of the Doctoral Thesis. In this period, the author published seven papers in journals, including one at ACM Transactions on Multimedia Computing Communications and Applications and two at Multimedia Systems. These journals are evaluated as Qualis A1 and B1, respectively. The author also published 29 papers in conferences, where of 12 in international events and 17 in national ones.

Lista de Figuras.....	ix
Lista de Siglas.....	xi
1 Introdução.....	1
2 Pesquisas Desenvolvidas.....	5
2.1 Codificação de Vídeo.....	6
2.1.1 Codificação de Vídeo e MPEG-4.....	6
2.1.2 Vídeo Estéreo.....	8
2.1.3 Resumo.....	14
2.2 Personalização e Adaptação de Conteúdo.....	15
2.2.1 Adaptação de Vídeo.....	15
2.2.2 Personalização de Conteúdo.....	19
2.2.3 Resumo.....	28
2.3 Televisão Digital Interativa.....	28
2.3.1 Contribuições em TV Digital Interativa.....	29
2.3.2 Resumo.....	34
2.4 Considerações Finais.....	34
3 Conclusões.....	37
3.1 Discussão e Reflexões.....	37
3.2 Trabalhos Futuros.....	39
Referências Bibliográficas.....	41
A.....	49
B.....	73
C.....	83
D.....	93
E.....	117
F.....	129
G.....	147
H.....	165
I.....	171
J.....	189

Lista de Figuras

Figura 1 - Processo de compressão espacial de vídeo estéreo proposto.....	10
Figura 2 - Processo de conversão de um par estéreo em vídeo anaglífico.....	11
Figura 3 - Processo de codificação para vídeo estéreo independente do modo de visualização.....	14
Figura 4 - Ferramenta M4Note estendida (retirado de (Manzato, 2011)).....	23
Figura 5 - Distribuição das publicações nas linhas de pesquisa: sinergias.....	34
Figura 6 - Distribuição das publicações nas subáreas.....	35

Lista de Siglas

3D – Terceira Dimensão.

API – *Application Programming Interface*.

CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior.

CBIR - *Content-Based Information Retrieval*.

CENPRA - Centro de Pesquisas Renato Archer.

C.E.S.A.R - Centro de Estudos e Sistemas Avançados do Recife.

CWI - *Centrum Wiskunde & Informatica*.

DCT - *Discrete Cosine Transform*.

FAPESP – Fundação de Apoio a Pesquisa do Estado de São Paulo.

FFCLRP – Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto – Universidade de São Paulo.

FUCAPI - Fundação Centro de Análise, Pesquisa e Inovação Tecnológica.

ICMC-USP – Instituto de Ciências Matemática e de Computação – Universidade de São Paulo.

IMDB - *The Internet Movie Data Base*.

INCT-SEC - Instituto Nacional de Ciência e Tecnologia em Sistemas Embarcados Críticos.

ISDB-TB – *Integrated Services Digital Broadcasting-Terrestrial Brazil*.

ITU-T - *International Telecommunication Union* - Telecommunication standardization sector.

LAN – *Local Area Network*.

LSI - *Latent Semantic Indexing*.

MAN – *Metropolitan Area Network*.

MIPS - *Millions of Instructions Per Second*.

MPEG – *Moving Pictures Experts Group*.

MPEG-J – MPEG Java.

MHP - *Multimedia Home Platform*.

MVC - *Model-View-Controller*.

NCL – *Nested Context Language*.

P&A – Personalização e Adaptação de Conteúdo.

PSNT – *Peak Signal-to-Noise Ratio*.

PUC-RIO – Pontifícia Universidade Católica do Rio de Janeiro.

RGB – *Red, Green, Blue*.

SBC – Sociedade Brasileira de Computação.

TIDIA-Ae - Tecnologia da Informação no Desenvolvimento da Internet Avançada – Aprendizado Eletrônico.

TVDI – Televisão Digital Interativa.

UFBA – Universidade Federal da Bahia.

UFPB – Universidade Federal da Paraíba.

UFSCar – Universidade Federal de São Carlos.

UNICAMP – Universidade Estadual de Campinas.

VIMOS – Vídeo, Mobilidade e Segurança.

XMT – *eXtensible Mpeg-4 Textual format*.

WAN – *Wide Area Network*.

YCbCr – Espaço de cores utilizado em imagens e vídeos digitais. Y representa a luminância enquanto Cb e Cr representam a cromaticidade.

Introdução

Multimídia pode ser definida como o uso simultâneo de diferentes modos digitais de representação de dados, chamados de mídias, durante a comunicação de informações (Halsall, 2001; Vaughan, 2006). Mídias como texto, áudio, imagens, animações e vídeo, são combinadas, por meio de recursos computacionais, produzindo uma apresentação sensorialmente mais rica. Essa é a maior motivação para o uso de multimídia, pois as diferentes representações dos dados exploram de modo simultâneo mais de um canal sensorial humano, possibilitando que o conteúdo seja assimilado mais facilmente (Halsall, 2001). De modo geral, a área de Multimídia pode ser subdividida em quatro subáreas:

Autoria Multimídia: envolve a criação de apresentações multimídia, interativas ou não, estruturadas na forma de um documento que organize (temporal e espacialmente) e sincronize os diferentes elementos de mídia (Hardman et al., 1993).

Codificação Multimídia: evolue técnicas, algoritmos e padrões para representar computacionalmente as informações contidas nas mídias, normalmente utilizando compressão de dados (Gibson, 1998).

Comunicação Multimídia: envolve protocolos, aplicações e infraestrutura para a transmissão e recepção de informação multimídia via redes de comunicação (Halsall, 2001).

Recuperação Multimídia: envolve técnicas, algoritmos, estruturas de dados e aplicações para a descrição e indexação de dados multimídia com o objetivo de permitir que tais dados possam ser eficientemente consultados e recuperados a partir de grandes repositórios (Blanken, 2007).

Durante as últimas décadas a área como um todo vem evoluindo rapidamente, impulsionada por avanços tecnológicos e pela demanda dos usuários por acesso universal a conteúdo multimídia. Tais avanços têm possibilitado o desenvolvimento de aplicações cada vez mais ricas, complexas e interativas, sendo que a lista de possíveis áreas de aplicação para Multimídia é extensa, atingindo vários setores da sociedade: *e-learning*, *e-gov*, *e-science*, telemedicina, comércio eletrônico, bibliotecas digitais, televisão digital interativa, cinema, jogos eletrônicos, entre outros.

Em paralelo ao desenvolvimento de serviços interativos, surgiram novas tecnologias relacionadas a dispositivos e redes que revolucionaram o modo com que os dados são acessados. Primeiramente, o consumo pode ser realizado utilizando diferentes tipos de dispositivos de acesso, incluindo celulares e computadores portáteis, independentemente da localidade. Adicionalmente, tais avanços mudaram o paradigma conhecido em que usuários apenas acessam conteúdo já existente, permitindo a autoria e a disponibilização dos dados na rede por qualquer indivíduo (YouTube¹, por exemplo).

Como resultado desse cenário, o volume de vídeos, áudios, imagens e documentos que são gerados constantemente aumenta muito, dificultando o processamento, armazenamento, recuperação e a disseminação dos mesmos (Blanken, 2007). Os desafios que se apresentam nesse contexto demandam: técnicas para codificação e compressão de dados, para permitir armazenamento e transmissão eficientes; métodos eficazes de filtragem e de seleção de conteúdo, para auxiliar na busca e na recuperação de informação relevante; modelos e ferramentas de autoria que permitam integrar e sincronizar diferentes mídias, auxiliando na produção de documentos multimídia interativos; protocolos e ferramentas de comunicação que garantam que os dados transmitidos cheguem ao receptor dentro de um prazo determinado (Halsall, 2001; Blanken, 2007; Hardman et al., 1993).

Esses problemas estão também no contexto dos Grandes Desafios da Pesquisa em Computação no Brasil: 2006–2016 (SBC, 2006), apontados pela Sociedade Brasileira de Computação (SBC). De acordo com o relatório da SBC, um dos cinco maiores desafios para a área de computação é a gestão da informação em grandes volumes de dados e informação distribuídos, tais como os conteúdos manipulados nas aplicações multimídia citadas.

¹ <http://www.youtube.com>

Esses desafios são alvo de investigações de diversos pesquisadores no Brasil e no mundo, os quais se dedicam às mais variadas linhas de pesquisa ligadas às quatro subáreas mencionadas. Cada linha de pesquisa, evidentemente, possui suas particularidades. Contudo, é comum encontrar intersecções entre diferentes linhas de pesquisa, levando a intersecções entre as subáreas. Isto é um reflexo da característica principal da área de Multimídia de integrar diferentes tipos de dados e informações.

As pesquisas realizadas por este autor, descritas no presente texto, estão divididas nas seguintes linhas: Codificação de Vídeo; Personalização e Adaptação de Conteúdo (P&A) e Televisão Digital Interativa (TVDI). Em todas, vídeo aparece como elemento central. Outra característica das pesquisas realizadas é que o autor busca utilizar as intersecções para, sempre que possível, aplicar os resultados de uma linha em outra, promovendo sinergias.

De modo a apresentar uma descrição das principais contribuições resultantes das atividades de pesquisa realizadas pelo autor, este texto sistematizado foi dividido em mais dois capítulos, além desta introdução. No Capítulo 2 são descritos os trabalhos desenvolvidos nas linhas de pesquisa, dando ênfase aos resultados de orientações e colaborações em projetos de pesquisa. Na Seção 2.1 são apresentados os trabalhos relacionados à Codificação de Vídeo; na Seção 2.2 são apresentados os resultados obtidos na linha de Personalização e Adaptação de Conteúdo e, na Seção 2.3, os resultados obtidos na linha de Televisão Digital Interativa. No Capítulo 3 discute-se as conclusões e são indicados os trabalhos futuros. Nos Apêndices de A a J são incluídas as publicações mais relevantes, resultantes do trabalho aqui reportado.

Pesquisas Desenvolvidas

Neste capítulo são resumidos os trabalhos desenvolvidos pelo autor ou sob sua orientação após a conclusão do doutorado. Os trabalhos são apresentados de acordo com seu enfoque principal, divididos em três linhas de pesquisa correspondentes às subseções deste capítulo. As publicações relacionadas são indicadas ao longo do texto e sumarizadas ao final.

Primeiramente, são apresentados os trabalhos relacionados à linha de Codificação de Vídeo. Em seguida, são apresentados os trabalhos relativos à linha de Personalização e Adaptação de Conteúdo, na qual o autor tem atuado com mais frequência após a conclusão do doutorado. Por fim, são apresentados trabalhos na linha de Televisão Digital Interativa (TVDI), na qual o autor possui participação em vários projetos de pesquisa. Sempre que oportuno, será mencionada a inter-relação de trabalhos desenvolvidos nessas três linhas.

Os trabalhos são apresentados de modo resumido, discutindo-se apenas o contexto no qual se inserem, os conceitos principais e os aspectos que os distinguem e caracterizam sua contribuição principal. Informações adicionais sobre os trabalhos são incluídas nos apêndices de A a J. A referência de cada publicação é incluída em uma nota de rodapé, usando a seguinte convenção: círculos correspondem às publicações em periódicos; os quadrados correspondem às publicações em conferências; publicações nacionais são apresentadas com linha tracejada; as publicações mais relevantes são apresentadas com fundo cinza.

2.1 Codificação de Vídeo

Nesta seção, são apresentados os resultados obtidos na linha de pesquisa Codificação de Vídeo, a qual se insere no contexto da área Codificação Multimídia. Essa área agrega os fundamentos sobre representação computacional dos dados das mídias, necessários para o desenvolvimento de codificadores e para se buscar integração entre aplicações que manipulam diferentes tipos de mídia. Este autor vem se especializando, desde o mestrado, na codificação de vídeo, em geral, e no padrão MPEG-4 (Pereira & Ebrahimi, 2002), em particular. O conhecimento de base tem sido aproveitado para desenvolver trabalhos que contribuam nas outras linhas e, mais recentemente, para desenvolver pesquisa em vídeo estéreo. Sendo assim, esta seção está dividida nessas duas frentes de atuação.

2.1.1 Codificação de Vídeo e MPEG-4

O objetivo da codificação de vídeo é reduzir o volume de dados do mesmo por meio de compressão. De modo geral, a codificação envolve os processos de compressão espacial e de compressão temporal. No primeiro caso, cada quadro do vídeo é comprimido como uma imagem separada. No segundo caso, são codificadas apenas as diferenças entre quadros vizinhos. Ambos processos utilizam técnicas que podem gerar perda de informação e comprometer a qualidade do vídeo (Richardson, 2002).

Ao longo da evolução do vídeo digital surgiram diversos padrões de codificação, desde o H.261, passando pelos padrões do grupo MPEG, chegando ao padrão H.264, considerado o estado da arte para codificação de vídeo (Halsall, 2001; Richardson, 2003). Esses padrões buscam aumentar a taxa de compressão reduzindo a taxa de bits necessária para armazenamento e transmissão. Primordialmente tais padrões codificam apenas vídeo com áudio associado. A exceção importante é o padrão MPEG-4.

O padrão MPEG-4 surgiu para atender às demandas da atual convergência entre aplicações gráficas interativas, Web e Televisão (Pereira & Ebrahimi, 2002; Richardson, 2003). Para tanto, o padrão especifica uma série de codificadores para diferentes tipos de mídia, sendo que cada mídia codificada é representada como um objeto. O padrão estabelece ainda mecanismos de sincronização espacial e temporal entre os objetos, permitindo com isso a composição de apresentações multimídia interativas.

O desenvolvimento de conteúdo interativo em MPEG-4 pode ser realizado de dois modos: por meio da linguagem declarativa XMT (*eXtensible Mpeg-4 Textual format*) ou por meio da API MPEG-J, em linguagem imperativa Java. Em qualquer dos casos, o código para a manipulação interativa dos objetos presentes em uma apresentação está embutido no contêiner da mesma, também como um objeto MPEG-4. Essa característica permite que um vídeo MPEG-4 contenha complexas aplicações multimídia interativas, com potenciais aplicações em Web e Televisão Digital Interativa (Goularte, 2003; Pereira & Ebrahimi, 2002).

O padrão MPEG-4 é um ponto de sinergia entre as três linhas de pesquisa do autor. Resultados de trabalhos com esse padrão, ou o conhecimento técnico sobre o mesmo, contribuíram com orientações, pesquisas e projetos nas outras linhas e serão descritos nas próximas seções. Nesta seção os trabalhos serão analisados apenas sob o ponto de vista da linha Codificação de Vídeo.

ORIENTAÇÕES CONCLUÍDAS

Um dos potenciais benefícios do padrão MPEG-4 para o desenvolvimento de aplicações multimídia interativas é a utilização da API MPEG-J em TV Digital. Têm-se duas situações: a) o Sistema Brasileiro de TV Digital (ISDB-TB) possui um *middleware* Java (Ginga-J (Souza Filho et al., 2007)) para desenvolvimento de aplicações que não manipulam conteúdo, apenas apresentam; b) o padrão MPEG-4 possui uma API em Java, a MPEG-J, para manipulação de conteúdo. Uma vez que tanto o Ginga-J quanto a MPEG-J utilizam a plataforma Java, é possível o desenvolvimento de um codificador (e respectivo *player*) MPEG-J compatível com o ISDB-TB.

Contudo, os codificadores comerciais disponíveis (como Xvid, DIVX e Sorensen) não implementam a parte interativa do padrão MPEG-4, implementam apenas codificadores de áudio e vídeo. Poucas iniciativas de software livre (como o projeto GPAC*) implementam desenvolvimento interativo MPEG-4, mas apenas com linguagem declarativa XMT. Essa situação motivou a orientação do trabalho de Iniciação Científica de Rafael Chies, que desenvolveu um codificador para a API MPEG-J compatível com o *player* do projeto GPAC (Carvalho et al., 2007);^[1]. Apesar do alto teor técnico, o trabalho

* <http://gpac.wp.institut-telecom.fr/>

[1] CARVALHO, D. F.; CHIES, R.; GOULARTE, R. . Esteganografia em vídeos MPEG-4. In: XIII Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia 2007), 2007, Gramado. Anais, Porto Alegre : SBC, 2007. v. 2, p. 5-8. (Qualis B3)

de Chies foi importante para viabilizar o desenvolvimento de um projeto na linha de Televisão Digital Interativa (Seção 2.3.1), assim como contribuiu para o trabalho de mestrado de Diego Fiori de Carvalho.

O mestrado de Carvalho (2008) explorou características do padrão MPEG-4 no desenvolvimento de um método de esteganografia em vídeos digitais. A esteganografia digital tradicionalmente busca esconder informação dentro de mídias como imagens (Singh, 2001) para alcançar sigilo. Uma característica marcante das técnicas aplicadas a imagens é o baixo volume de inserção de dados alcançado. A utilização de vídeo digital tem potencial para aumentar esse volume, uma vez que um vídeo é composto por várias imagens. Contudo, vídeos digitais são normalmente comprimidos utilizando-se técnicas com perdas (Halsall, 2001), afetando a recuperação exata da informação esteganografada. As técnicas desenvolvidas para vídeos em trabalhos relacionados utilizam os padrões MPEG-1 e MPEG-2, conseguindo manter a integridade da informação no processo de recuperação, mas o volume de inserção continua baixo (Carvalho, 2008).

O trabalho de Chies auxiliou Carvalho a desenvolver estruturas de dados em MPEG-J para armazenamento de informação em vídeos MPEG-4 (Carvalho et al., 2007; Carvalho et al. 2008)^[2]. Tais estruturas são codificadas como objetos MPEG-4 não visíveis, não detectáveis (somente pelo destinatário da mensagem) e com alta capacidade de armazenamento (Carvalho & Goularte, 2009)³. O método desenvolvido por Carvalho é flexível, podendo ser aplicado em domínios diversos, como o de controle de versões de documentos multimídia (Carvalho et al., 2008b)^[4].

2.1.2 Vídeo Estéreo

A utilização de vídeo estéreo tem ganhado novo impulso com o retorno do cinema 3D e com o surgimento dos novos televisores 3D (Hutchison, 2008). Esse fato tem incentivado indústria e academia a pesquisar e desenvolver métodos e técnicas que promovam a produção e distribuição desse tipo de vídeo.

^[2] CARVALHO, D. F.; CHIES, R.; GOULARTE, R.. MP4Stego: Esteganografia em Vídeos MPEG-4. In: XIV Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia 2008), 2008, Vila Velha. Anais, Porto Alegre : SBC, 2008. v. 1. p. 154-161. DOI: 10.1145/1666091.1666118. (Qualis B3)

³ CARVALHO, D. F.; GOULARTE, R.. MP4Stego: Steganography in MPEG-4 Videos. International Journal of Advanced Media and Communication, v. 3, n. 4, p. 440-460, 2009. DOI: 10.1504/IJAMC.2009.028712.

^[4] CARVALHO, D. F.; CHIES, R.; FREIRE, A. P.; MARTIMIANO, L. A. F.; GOULARTE, R.. Video Steganography for Confidential Documents: Integrity, Privacy and Version Control. In: ACM International Conference on Design of Communication (SigDoc'08), 2008, Lisboa. Anais, New York : ACM Press, 2008. v. 1, p. 199-206. DOI: 10.1145/1456536.1456578. (Qualis B3)

Em termos técnicos, os vídeos 3D são vídeos estereoscópicos e utilizam métodos que procuram simular o efeito obtido na visão humana pelo fato de nossos olhos estarem distantes horizontalmente um do outro, fazendo com que cada olho tenha um ponto de vista diferente (Azevedo & Conci, 2003). Tais métodos consistem em apresentar um par estéreo de imagens representando, respectivamente, as perspectivas do olho esquerdo e direito. O par estéreo é interpretado pelo cérebro humano criando uma imagem única e tridimensional, provocando a sensação de profundidade e distanciamento. Esse processo de interpretação do cérebro é conhecido como estereopsia (Mendiburu, 2009).

A tecnologia para a produção de vídeos estereoscópicos sofreu avanços tanto na forma de captação quanto na forma de reprodução. Câmeras especiais foram desenvolvidas visando capturar dois pontos de vista diferentes de uma mesma imagem (gerando o par estéreo), ou então gerando um mapa de profundidade das cenas juntamente com o vídeo (Fehn et al, 2002). Também foram desenvolvidas técnicas para conversão e apresentação de vídeo estéreo a partir de vídeos originalmente monoculares (2D) (Tam & Zhang, 2006).

Quanto aos modos de reprodução e visualização, as tecnologias atuais apresentam as duas imagens do par estéreo de duas maneiras distintas. Na primeira delas, as imagens são exibidas ao mesmo tempo, e se utilizam óculos especiais para filtrar e direcionar cada uma ao olho correspondente, causando a estereopsia. É o caso do modo anaglífico, que utiliza lentes com filtros coloridos para separar as imagens (processo detalhado adiante), e do modo polarizador, que utiliza filtros polarizadores de luz para a separação (é o modo dos atuais “cinemas 3D”) (Stereographics, 1997; Mendiburu, 2009).

Na segunda maneira de apresentação as imagens do par estéreo são exibidas alternadamente, a uma taxa suficientemente rápida para que o cérebro as processe de modo a causar estereopsia. É o caso tanto dos óculos obturadores (*shutter glasses*), que abrem e fecham suas lentes em sincronia com as imagens sendo projetadas (como em alguns televisores 3D atuais); quanto dos televisores e monitores autoestereoscópicos, que direcionam a luz da projeção de cada imagem para o olho correto, dispensando o uso de óculos (Fehn et al., 2006; Hutchison, 2008; Sony Corporation, 2010).

Apesar do recente impulso das indústrias do cinema (Mendiburu, 2009; Suppia, 2007) e da televisão (Sony Corporation, 2010), ainda existe necessidade de pesquisa nessa linha. Um exemplo disso é a atual falta de padronização no modo de organizar os dados para fins de armazenamento ou transmissão. Em uma primeira abordagem o par estéreo é simplesmente armazenado em um contêiner (Lipton, 1997), o quê garante utilização dos

dados em qualquer modo de visualização, mas, dobra a necessidade de espaço para armazenamento e requer pós-processamento em alguns casos (Smolic et al, 2009). Em uma outra abordagem, as técnicas tradicionalmente utilizadas para comprimir vídeo 2D são adaptadas para codificação de vídeo estéreo (Smolic et al, 2009), gerando perda de qualidade que pode inviabilizar a correta percepção de profundidade em alguns tipos de visualização (Andrade & Goularte, 2009).

Desse modo, encontra-se uma lacuna na literatura especializada sobre métodos de codificação particularmente projetados para atender vídeo estéreo que forneçam compressão sem perda de qualidade na percepção de profundidade e que, ao mesmo tempo, sejam independentes do modo de visualização. Como resultado dessa lacuna, o conteúdo codificado por uma técnica fica vinculado a um modo específico de visualização/reprodução. Por exemplo, um vídeo codificado pela técnica anaglífica (Siegel et al, 1994) possui o par estéreo fundido em um único fluxo de dados, o que impossibilita a exibição desse conteúdo em sistemas que utilizam óculos obturadores ou polarizadores, pois estes necessitam do par estéreo para realizar a projeção. Ou ainda, vídeos codificados com a técnica de mapa de profundidade (Merkle, 2007; Hutchison, 2008) para exibição autoestereoscópica não possuem o par estéreo, o que impossibilita sua exibição por quaisquer dos outros modos.

ORIENTAÇÕES EM ANDAMENTO

Uma das orientações em andamento nessa linha, o doutorado de Leonardo Antônio de Andrade, tem como objetivo desenvolver uma técnica de compressão espacial especialmente projetada para vídeo estéreo. A estratégia adotada consiste em analisar as etapas principais do processo de compressão espacial de vídeo onde ocorrem perdas: a subamostragem de crominância e a aplicação de transformada matemática com correspondente quantização (Richardson, 2002).

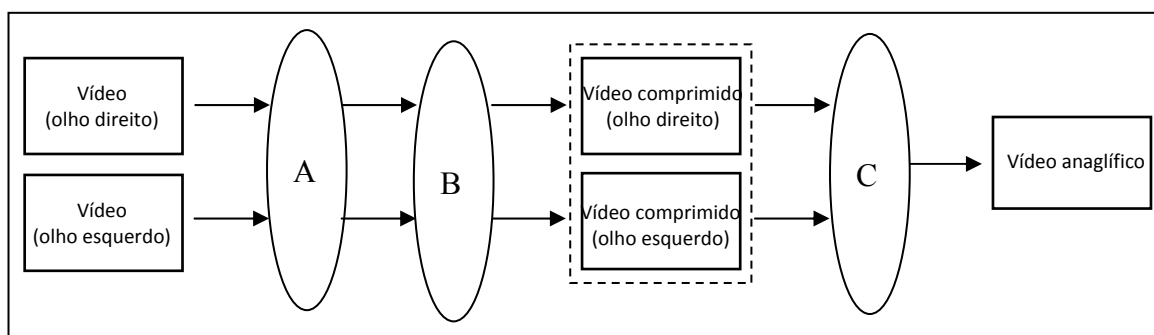


Figura 1 – Processo de compressão espacial de vídeo estéreo proposto.

Tais etapas estão respectivamente ilustradas na Figura 1 por (A) e (B). Nessas etapas devem-se descobrir quais os limites toleráveis para as perdas, permitindo desenvolver técnicas bem ajustadas para o caso estéreo, comprimindo sem perder qualidade na percepção de profundidade. Por fim, para alcançar mais compressão, a proposta é fundir o par estéreo comprimido em um vídeo anaglífico (etapa C na Figura 1).

Um vídeo anaglífico é formado pela fusão do par estéreo em um único fluxo de dados. A Figura 2 ilustra um exemplo dessa conversão. Seja o par estéreo formado por $R_1G_1B_1$ (o vídeo do olho direito) e $R_2G_2B_2$ (o vídeo do olho esquerdo). A conversão se dá, por exemplo, utilizando-se o canal vermelho (R_2) do vídeo do olho esquerdo e os canais verde e azul (G_1 e B_1) do olho direito, formando um terceiro vídeo (o vídeo anaglífico): $R_2G_1B_1$. Como a combinação de cores verde (G_1) e azul (B_1) resulta na cor ciano, o modo anaglífico exemplificado na Figura 2 é conhecido como Vermelho/Ciano. Outros modos anaglíficos possíveis são: Verde/Magenta ($R_1G_2B_1$) e Azul/Amarelo ($R_1G_1B_2$).

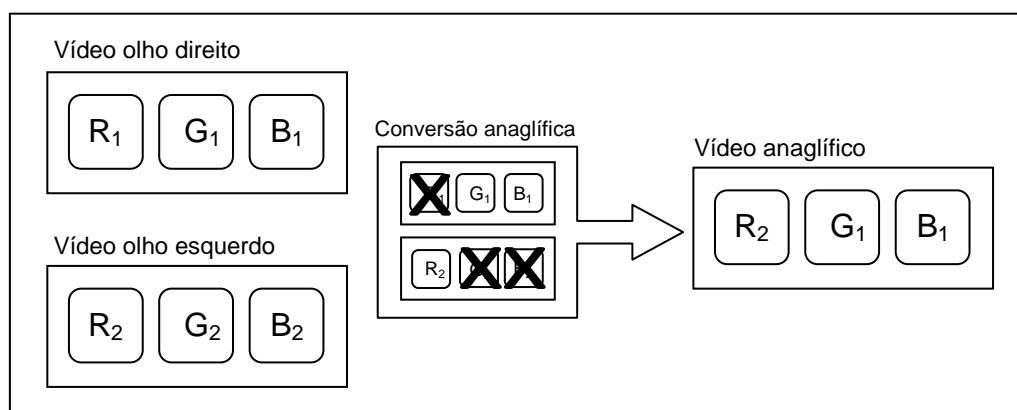


Figura 2 – Processo de conversão de um par estéreo em vídeo anaglífico.

Durante a reprodução do vídeo anaglífico as imagens aparecem sobrepostas e o espectador usa óculos com lentes especiais coloridas atuando como filtros. No caso do exemplo da Figura 2 seriam filtros vermelho e ciano (olhos esquerdo e direito, respectivamente). Com isso consegue-se separar as imagens e direcionar cada uma para o olho correto, criando a estereopsia (Mendiburu, 2009).

A utilização de vídeo anaglífico tem duas motivações: 1) esse tipo de vídeo possui o par estéreo fundido em um único fluxo, reduzindo pela metade o volume de dados; 2) a técnica anaglífica é a mais sensível a perdas dentre todas as disponíveis para visualização estéreo (é dependentes de cores e as outras não) (Mendiburu, 2009; Andrade & Goularte, 2010); logo, se a compressão espacial não causar perda de percepção de profundidade no vídeo anaglífico, também não causará nos outros modos de visualização.

As investigações conduzidas até o momento na etapa B (Figura 1) (Andrade & Goularte, 2009)^[5] demonstraram que a utilização de transformadas *Wavelet* para compressão de vídeo estéreo é preferível à utilização da Transformada Discreta dos Cossenos (*Discrete Cossine Transform – DCT*), tradicionalmente utilizada em compressão de vídeo. Ainda, a análise sugere um fato importante: diferenças de PSNR* maiores que 4dB entre os canais de cor do vídeo causam perda da sensação de profundidade, sendo este um possível limite para as perdas devido a erros de arredondamento impostos pela implementação computacional de transformadas *Wavelets*.

Já os resultados reportados em Andrade & Goularte (2010)^[6], identificaram que a combinação de cores Verde/Magenta apresenta qualidade subjetiva média 10% superior à da combinação Vermelho/Ciano. Ainda, em termos de qualidade objetiva, a combinação Verde/Magenta apresenta em média 3dB de diferença de PSNR entre as componentes de cor, contra 6dB da combinação Vermelho/Ciano, novamente dando indícios de que valores em torno de 4dB podem representar um limite.

Nesse mesmo trabalho (Andrade & Goularte, 2010) analisou-se a etapa de subamostragem de croma, ilustrada pela letra A na Figura 1. A subamostragem é amplamente utilizada em codificação de vídeo e consiste em transformar o espaço de cor dos quadros de vídeo de RGB para YCbCr, reduzindo a resolução espacial das componentes de cor (Cb e Cr) (Richardson, 2002). Tal redução promove compressão, mas, por outro lado, promove perda de qualidade já que as cores são subamostradas. A análise subjetiva demonstrou que a subamostragem 4:2:2 (Y:Cb:Cr) oferece um bom compromisso entre qualidade e compressão. Já a subamostragem 4:2:0, utilizada pelos codificadores de vídeo padronizados e nos trabalhos relacionados, oferece, em média, resultados abaixo do índice crítico para percepção de profundidade. Na análise objetiva, conforme o esperado, a subamostragem 4:2:2 apresentou médias de PSNR maiores que a 4:2:0.

Em outra orientação em andamento, a do mestrado de Matheus Ricardo Uihara Zingarelli, tem-se como objetivo investigar e desenvolver meios de reverter um vídeo anaglífico em seu par estéreo original. Esse processo não é trivial uma vez que a geração

^[5] ANDRADE, L. A.; GOULARTE, R. Percepção Estereoscópica Anaglífica em Vídeos Digitais Comprimidos com Perda. In: XV Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia 2009), 2009, Fortaleza, Brasil. Anais. Porto Alegre : SBC, 2009, v. 1. p. 226-233. DOI: 10.1145/1858477.1858506. (Qualis B3)

* PSNR (*Peak Signal-to-Noise Ratio*) é usada para medir o nível de ruído introduzido na imagem devido à compressão.

^[6] ANDRADE, L. A.; GOULARTE, R. Uma Análise da Influência da Subamostragem de Crominância em Vídeos Estereoscópicos Anaglíficos. In: XVI Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia 2010), 2010, Belo Horizonte, Brasil. Anais. Porto Alegre : SBC, 2010, v. 1, p. 179-186. (Qualis B3)

do vídeo anaglífico implica em perda de informação tanto espacial quanto de cor. A Figura 2 ilustra a perda de informação de cor durante a Conversão anaglífica, onde três dos seis canais originais são descartados. A simples cópia dos canais presentes no vídeo anaglífico para recompor os canais descartados (copiar os dados de G_1 para recompor G_2 , por exemplo) não resolve o problema. Como os pixels de um vídeo do par estéreo estão ligeiramente deslocados em relação ao outro vídeo, devido às perspectivas diferentes de cada olho, a cópia direta eliminaria essa diferença de perspectivas, acarretando a perda da estereopsia.

Em Zingarelli et al. (2011)^[7] foi proposto um método de reversão baseado em imagem anaglífica complementar e em transformação do espaço de cor de RGB para YCbCr. Os resultados obtidos alcançaram uma taxa de compressão de mais de 64% para o vídeo anaglífico gerado. O par estéreo recuperado pôde ser utilizado na geração de novos vídeos sem perdas perceptíveis na sensação de profundidade. Contudo, a qualidade das imagens caiu abaixo do esperado, sendo o PSNR obtido igual a 34,52dB em média. Atualmente Zingarelli está investigando meios de reduzir essa queda de qualidade utilizando modos não tradicionais de realizar a transformação RGB-YCbCr, como a subamostragem 4:4:0.

PROJETOS

O projeto Um Método Genérico para Codificação de Vídeos Estereoscópicos, financiado pela FAPESP (processo número 2011/00422-7) e coordenado pelo autor, tem como meta desenvolver um codificador para vídeo estéreo que seja independente do modo de visualização, preenchendo a lacuna anteriormente discutida.

Com os resultados obtidos até o momento no trabalho de Andrade, tem-se um método de compressão espacial (Figura 1) com taxas de compressão compatíveis com os codificadores atuais para vídeo 2D. O diferencial do trabalho está na qualidade, já que o processo está baseado em particularidades do vídeo anaglífico, conforme discutido ao longo desta seção.

O trabalho de Andrade está restrito ao modo de visualização anaglífico. Contudo, o trabalho de Zingarelli (Zingarelli et al., 2011) provou que é possível realizar a reversão de um vídeo anaglífico em seu par estéreo correspondente. Assim, a combinação dos dois

^[7] ZINGARELLI, M. R. U.; ANDRADE, L. A.; GOULARTE, R. Reversing Anaglyph Videos Into Stereo Pairs. In: XVII Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia 2011), 2011, Florianópolis - SC. (Aceito para publicação como artigo completo). (Qualis B3)

trabalhos (Figura 3) segue em direção à obtenção da referida independência, pois, os outros modos de visualização (por luz polarizada, óculos obturadores e autoestereoscópico) são baseados na projeção do par estéreo.

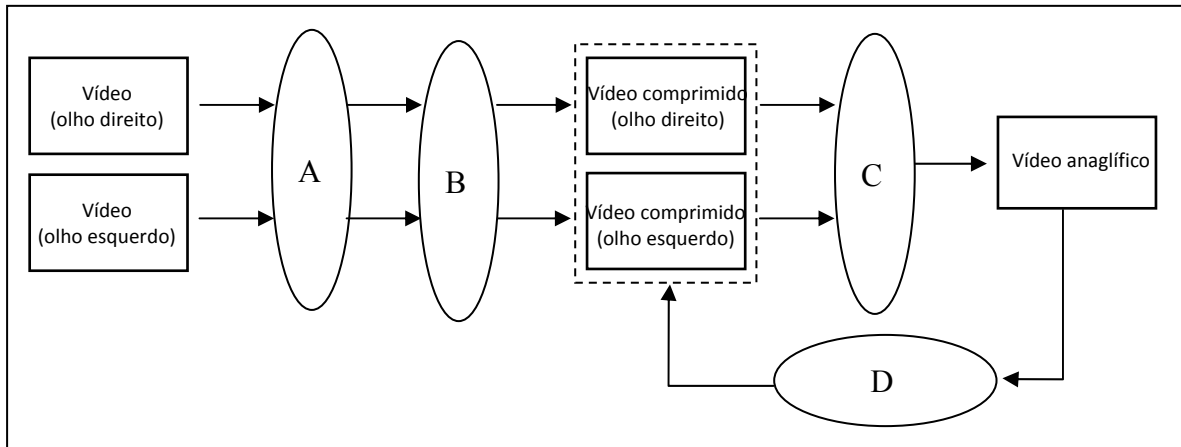


Figura 3 – Processo de codificação para vídeo estéreo independente do modo de visualização. (A) subamostragem de crominância; (B) aplicação de transformada *Wavelet* com correspondente quantização; (C) conversão do par estéreo em vídeo anaglífico; (D) reversão do vídeo anaglífico em par estéreo.

A Figura 3 ilustra o processo em que o projeto está baseado. O par estéreo de vídeos é submetido a compressão espacial, baseada em subamostragem de crominância (etapa A) e aplicação de transformadas *Wavelets* com quantização (etapa B). O par estéreo comprimido é transformado em vídeo anaglífico (etapa C), de modo que se tem apenas um fluxo de dados, facilitando armazenamento e transmissão. O vídeo anaglífico gerado pode ser apresentado utilizando-se o modo anaglífico de visualização. Para utilizar um outro modo de visualização, basta reverter o vídeo anaglífico em par estéreo (etapa D).

Além da formação de recursos humanos qualificados, dois dos resultados esperados deste projeto são: um processo para codificação genérica de vídeo estéreo e a implementação de um codificador correspondente. Como trabalhos futuros o projeto abre espaço para a investigação de diferentes métodos de conversão de vídeo anaglífico em par estéreo.

2.1.3 Resumo

Nessa linha de pesquisa foram orientados um trabalho de mestrado e um trabalho de iniciação científica. Foram publicados: um artigo em periódico internacional, cinco artigos em conferências nacionais e um artigo em conferência internacional. Atualmente a linha conta com duas orientações em andamento, sendo uma de mestrado e a outra de

doutorado, além de um projeto de pesquisa financiado pela FAPESP e coordenado pelo autor.

2.2 Personalização e Adaptação de Conteúdo

Nesta seção, são apresentados os resultados obtidos na linha de pesquisa Personalização e Adaptação de Conteúdo (P&A). Trata-se da linha em que o autor tem atuado com mais regularidade após a conclusão de seu doutorado. As pesquisas nessa linha estão no contexto de Recuperação Multimídia, que busca métodos automáticos e eficientes para recuperar conteúdo. As contribuições nessa área ocorrem, entre outras, na redução do problema recorrente, mas ainda atual, da sobrecarga de informação (Adomavicius & Tuzhilin, 2005).

Como discutido anteriormente, o problema da sobrecarga de informação ocorre devido à rápida evolução dos sistemas multimídia que vem permitindo acesso a conteúdos cada vez mais complexos e com maior volume de dados. Como consequência, o volume de informações disponível para um usuário é maior do que o mesmo conseguiria absorver. Isso, aliado a restrições de infraestrutura de entrega e de recepção (limitação de banda, de memória ou de área de visualização, por exemplo), dificultam a recuperação de itens multimídia específicos por parte do usuário.

Tais problemas vêm sendo estudados na linha de P&A (Magalhães & Pereira, 2004). Um sistema de adaptação procura decidir a versão de conteúdo ideal para apresentação e a melhor estratégia para gerar essa versão (Lum & Lau, 2002). A personalização, por sua vez, conforme argumentam Barrios et al. (2005), é vista como um caso particular da adaptação, quando os dados são adaptados para um único usuário. Na prática, a adaptação procura disponibilizar meios de se acessar conteúdo multimídia a partir de diferentes condições de dispositivos, rede e ambiente computacional. Já a personalização, conforme o próprio nome sugere, estuda meios de se customizar e/ou filtrar os dados segundo as preferências, interesses e necessidades de um usuário específico.

2.2.1 Adaptação de Vídeo

Para a adaptação de vídeo, a literatura reporta três modelos distintos: o modelo de Distribuição por Difusão Simultânea (Conklin et al., 2001); o modelo de Mídia Escalável,

também conhecido como modelo por codificação em camadas (Horn & Girod, 1997; Ohm, 2005) e o modelo de Transcodificação (Vetro et al., 2003; Ahmad et al., 2005). Este último tem potencial para oferecer a melhor relação custo-benefício entre os três modelos, contudo, sua aplicação eficiente depende de métodos para automaticamente obter os parâmetros de recodificação, como taxa de bits e resolução espacial, que irão atender às necessidades do usuário.

ORIENTAÇÕES CONCLUÍDAS

A obtenção automática de parâmetros para realizar transcodificação de vídeo motivou a orientação do trabalho de mestrado de Marcelo Garcia Manzato (Manzato, 2006). Nesse trabalho, propôs-se uma arquitetura para adaptação de vídeo ao vivo no padrão MPEG-4, a qual obtém a largura de banda atual da rede (no momento do início da transmissão) e uma descrição do dispositivo sendo usado para receber o vídeo. Esses parâmetros são utilizados para calcular a taxa de bits possível de ser transmitida, ajustando a resolução espacial dos quadros de vídeo e, assim, adaptando o vídeo segundo as condições atuais do sistema ou necessidades do usuário. Uma das contribuições desse trabalho foi o desenvolvimento de um modelo para a descrição de dispositivos (Manzato & Goularte, 2005)^[8] que garantiu a interoperabilidade dos dados.

No trabalho realizado por Manzato a captura das informações necessárias para a transcodificação era realizada no momento do acesso ao vídeo. Desse modo, se a largura de banda sofresse alterações após a adaptação ter sido iniciada, o recodificador não teria os novos parâmetros para ajustar o fluxo de vídeo dinamicamente, havendo a necessidade de o usuário realizar nova requisição ao servidor. Isso motivou a investigação por um método dinâmico de adaptação, o que levou à orientação do trabalho de mestrado de Robson Eisinger (Eisinger, 2007) e à colaboração com a professora Roseli A. Francelin Romero (ICMC-USP), especialista na área de Aprendizado de Máquina. Eisinger estendeu o modelo de descrição de dispositivos (Eisinger et al, 2005)^[9] com informações sobre características da rede. Tal modelo de representação é alimentado por um método baseado em

^[8] MANZATO, M. G. ; GOULARTE, R. . Live video adaptation: a context-aware approach. In: XI Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia 2005), 2005, Poços de Caldas - MG. Anais, Porto Alegre : SBC, 2005. v. 1 p. 1-8. DOI: 10.1145/1114223.1114230. (Qualis B3)

^[9] EISINGER, R. ; MANZATO, M. G. ; GOULARTE, R. . Devices Descriptions for Context-Based Content Adaptation. In: Third Latin American Web Congress (LA-WEB 2005), 2005, Buenos Aires. Anais, Los Alamitos : IEEE Computer Society Press, 2005, v.1 . p. 121-129. (Qualis B3)

aprendizado de máquina (Eisinger et al, 2008)^[10], desenvolvido para capturar as condições atuais da rede e decidir quais os melhores parâmetros para ajustar dinamicamente a arquitetura de recodificação proposta por Manzato.

PROJETOS

O autor participou do projeto VIMOS[♣], onde uma das tarefas era desenvolver meios de adaptar vídeo durante transmissão e recepção móvel. Nesse contexto, tiveram participação importante os resultados obtidos com os trabalhos de Manzato e de Eisinger no desenvolvimento da arquitetura para adaptação de vídeo MPEG-4.

A utilização de vídeo ao vivo estava prevista em outros dois projetos em que autor participou: projeto Tidia-Ae[♣] (como integrante) e projeto FAPESP 05/58454-0[♣] (como coordenador). Ambos estavam voltados a dar suporte computacional a atividades de ensino a distância, onde os estudantes poderiam fazer uso de diferentes dispositivos (móveis ou não) para acessar o conteúdo das aulas que, eventualmente, deveria ser adaptado. Logo, os resultados dos trabalhos de Manzato e de Eisinger foram, também, aplicados no desenvolvimento desses projetos.

Várias das atividades de suporte previstas nesses projetos envolvem interatividade, que requer um tempo de resposta relativamente curto do sistema – segundo a Recomendação G.114 da ITU-T (*International Telecommunication Union*), um atraso fim-a-fim de 0,150 segundos é aceitável para a maioria das aplicações interativas (ITU-T, 2003). A transcodificação, invariavelmente, impõe atrasos no processo de entrega de vídeo, o que pode tornar inviável o seu uso em sistemas interativos. Resultados obtidos na investigação dessa questão, relatados em Manzato & Goularte (2006)^[11], mostraram que o tempo gasto para recodificar um quadro de vídeo no padrão MPEG-4 é de ordem quatro vezes menor que o tempo limite de 0,150 segundo. Mostraram também que o atraso fim-a-

^[10] EISINGER, R. ; ROMERO, R. A. F. ; GOULARTE, R. . Machine Learning Techniques Applied to Dynamic Video Adapting. In: The Seventh International Conference on Machine Learning and Applications (ICMLA'08), 2008, San Diego. Anais, Piscataway : IEEE Computer Society, 2008. v. 1. p. 819-822. (Qualis B3)

♣ VIMOS - Vídeo, Mobilidade e Segurança. Chamada conjunta SEPIN-CNPq-FINEP-01/2002. (2003-2005).

♣ TIDIA-Ae - Aprendizado Eletrônico sobre Plataforma de Redes de Alta Velocidade. Projeto Fapesp 05/60651-9, Fase 1 (2004-2007) e Fase 2 (2007-2010).

♣ Captura, Transmissão e Adaptação de Vídeo ao Vivo em Sistemas de Apoio ao Ensino a Distância. Projeto Fapesp 05/58454-0 (2006 – 2008).

^[11] MANZATO, M. G. ; GOULARTE, R. . Transcoding Evaluation in Live Video Adaptation. In: IV Latin-American Web Congress (LA-Web 2006), 2006, Cholula, Puebla. Anais, Los Alamitos : IEEE Computer Society Press, 2006. v. 1. p. 203-208. (Qualis B3)

fim superou em até duas vezes esse valor limite, sugerindo que o sistema de transmissão impõe um atraso acumulativo na entrega de vídeo ao vivo.

Investigações mais aprofundadas sobre esse atraso foram conduzidas pelo autor em colaboração com os professores Rodrigo Fernandes de Mello (ICMC-USP) e Laurence T. Yang (St. Francis Xavier University - Canadá), especialistas na área de Sistemas Distribuídos. A meta foi desenvolver meios de identificar o custo de uma tarefa de transcodificação para diferentes padrões de vídeo (MJPEG, MPEG-1, MPEG-2 e MPEG-4) e a capacidade computacional necessária para tanto. Identificando esses pontos é possível prever o comportamento do sistema de acordo com o número de solicitações de clientes ou mudanças em parâmetros de transcodificação de vídeo, permitindo definir um ambiente capaz de corresponder a um amplo leque de aplicações de transcodificação.

As investigações preliminares mostraram que o limite de uma simples máquina pode ser rapidamente alcançado durante transcodificações (Goularte et al., 2007)^[12], e que *clusters* de computadores devem então ser considerados em arquiteturas para transcodificação. No trabalho desenvolvido por Goularte et al. (2007) propôs-se um modelo analítico para quantificar o custo de transcodificação de vídeo. O modelo proposto é baseado em MIPS (*millions of instructions per second*) (Mello & Senger, 2006), que mede o número de instruções necessárias para executar uma operação. Assim, o modelo é independente de hardware uma vez que não utiliza medidas de tempo, as quais podem variar quando se compara uma dada operação em diferentes máquinas. Utilizando o modelo é possível identificar: 1) qual o custo computacional de uma dada tarefa de transcodificação; 2) como definir o limite de processamento do computador para tarefas de transcodificação; 3) quando o uso de computação em *cluster* é preferível ou necessário para lidar com múltiplas tarefas de transcodificação e 4) qual é o impacto da transcodificação na transmissão de vídeo ao vivo fim-a-fim.

Os resultados obtidos mostraram que o modelo proposto consegue dimensionar bem o custo e o limite de processamento das transcodificações e sugeriram que o atraso acumulativo poderia ser tolerado em redes locais, mas talvez não em redes maiores. Novos experimentos, mais abrangentes, foram conduzidos gerando os resultados reportados em

^[12] GOULARTE, R. ; MELLO, R. F. ; DODONOV, E. ; YANG, L. T. . A Model to Estimate Transcoding Costs Applied in Live-Video Transmissions. In: International Conference on Intelligent Pervasive Computing (IPC-07), 2007, Jeju Island. Anais, Washington, DC : IEEE Computer Society, 2007. v. 1. p. 341-348. DOI=10.1109/IPC.2007.13. (Qualis B1)

Mello et al. (2009)⁽¹³⁾, onde se comprovou que o atraso acumulativo se torna proibitivo em 1,8 transcódificações em LANs (redes locais), em quase 1,8 em MANs (redes metropolitanas) e em cerca de 1,2 em WANs (Internet), confirmando que é possível transcodificar e transmitir vídeo ao vivo em ambientes LAN e MAN sem afetar a percepção de interatividade do usuário. Os resultados também confirmaram que o modelo é útil para estudar e definir arquiteturas adequadas para transcodificação de vídeo ao vivo.

2.2.2 Personalização de Conteúdo

Devido à quantidade crescente de conteúdo multimídia disponível atualmente, serviços de personalização se tornam importantes para ajudar usuários a filtrar o que é mais relevante ou de interesse particular de cada indivíduo. Segundo Adomavicius & Tuzhilin (2005), serviços que exploram essas características podem ser categorizados em Seleção de Conteúdo, Sistemas de Recomendação e Sistemas de Sumarização.

Sistemas de personalização, em geral, apresentam uma necessidade em comum: o conhecimento de informações a respeito do conteúdo e do usuário. Esses metadados descrevem: a mídia em si (formato, tipo de compressão e tamanho de arquivo, por exemplo); dados informativos sobre o conteúdo sendo apresentado e dados sobre as preferências e necessidades do usuário. As descrições podem variar em nível de granularidade ou detalhamento, assim como em nível semântico de representação (Snoek & Worring, 2005). Um dado com baixo nível semântico poderia ser o histograma de cores de uma imagem, a média de cores dos blocos que compõem um quadro de um vídeo, a quantidade de movimento existente em uma cena, entre outros. Descrições de dados com alto nível semântico, por outro lado, observam características conceituais do conteúdo e do usuário (objetos reconhecidos, localidade de uma cena, situação em curso na cena, gênero de filme predileto, entre outras).

O processo de obtenção desses metadados a partir de um fluxo multimídia é conhecido por Indexação Multimídia (subárea de Recuperação Multimídia), o qual pode ser automático ou manual (Hampapur et al., 1995; Snoek & Worring, 2005). A indexação automática é aplicada, de modo geral, para extrair dados de baixo nível semântico já que estes podem ser obtidos utilizando-se técnicas bem conhecidas para gerar descrições

⁽¹³⁾ MELLO, R. F. ; GOULARTE, R. ; DODONOV, E. ; YANG, L. T. ; PARK, J. H. ; KIM, T. On modeling and evaluating multicompuser transcoding architectures for live-video streams. *Multimedia Tools and Applications*, v. 43, n. 2, p. 109-129, 2009. DOI: 10.1007/s11042-009-0259-y. (Qualis B2)

automáticas dos mesmos. A limitação está no fato de que tais técnicas tendem a restringir o domínio em que os dados serão analisados (no caso de vídeo, documentário, esportes, etc.). Isso se deve às diferentes propriedades que são inerentes a cada gênero multimídia, fazendo com que a extração de informações semânticas seja difícil de ser realizada sem considerar um modelo visual e auditivo bem definido (Brezeale & Cook, 2007).

No caso da indexação manual, o problema da restrição do domínio de dados inexistente, já que o processo de interpretação do conteúdo é realizado por um especialista. Contudo, além de tal abordagem resultar em um trabalho cansativo, caro e propenso a erros, outro problema existente é que esse tipo de indexação pode explorar diferentes aspectos do conteúdo. Normalmente, profissionais irão criar metadados relacionados a certas características vistas subjetivamente como as mais importantes em uma apresentação. Um produtor, por exemplo, poderá achar que o local onde um carro está estacionado é mais importante que o modelo do carro, que, por sua vez, poderá ser visto como a informação prioritária por outra pessoa (Snoek & Worring, 2005).

Em vídeo digital é comum se utilizar uma etapa anterior à indexação: a Segmentação Multimídia. Essa etapa tem como objetivo dividir o vídeo em porções menores, mais gerenciáveis, facilitando operações como indexação (a automática, em especial), busca, recuperação e personalização (Hanjalic, 2004). As técnicas automáticas para segmentação conseguem bons resultados quando aplicadas para obtenção de estruturas mais simples, como quadros e tomadas. Contudo, para estruturas mais complexas, como cenas, ainda é necessária pesquisa (Hanjalic, 2004; Coimbra, 2011). Cenas são estruturas interessantes, pois, são aderentes ao modelo mental de vídeo utilizado por usuários no mundo real (usuários não fazem buscas por quadros ou tomadas de vídeo, por exemplo) (Coimbra, 2011). As pesquisas em segmentação de vídeo em cenas (como em: Colace et al., 2005; Liu et al., 2007; Liu et al., 2009; Hua-Yong & Tingting, 2009) indicam a necessidade de avanços nos métodos de extração de informação semântica para obter melhores resultados.

Informações semânticas têm potencial para promover uma melhoria nos serviços de personalização, uma vez que podem formar um elo entre a representação computacional de um conteúdo e a interpretação dos dados por um determinado usuário. A falta dessa ligação entre a informação em si e a interpretação do usuário é um problema recorrente em várias áreas, chamado de lacuna semântica (Smeulders et al., 2000). Pesquisadores vêm

centrando esforços no desenvolvimento de técnicas capazes de criar o referido elo entre ambas as partes, entretanto, conforme exposto, a literatura reporta que ainda existem dificuldades a serem superadas no processo de obtenção desses metadados, dificultando o desenvolvimento de aplicações de personalização de conteúdo.

ORIENTAÇÕES CONCLUÍDAS

A investigação sobre extração de informações semânticas para uso em serviços de personalização motivou orientação do trabalho de doutorado de Marcelo Garcia Manzato (Manzato, 2011). Apesar de serem encontradas várias propostas na literatura, as técnicas de indexação automática possuem restrição de domínio e as técnicas de indexação manual são dispendiosas e suscetíveis a erros. Uma possível solução seria explorar anotações colaborativas dos usuários, eliminando a restrição de domínio e reduzindo os esforços da indexação manual. Contudo, a utilização direta dessa abordagem em sistemas de personalização gera um outro problema. Anotações colaborativas são impessoais e, perdendo-se a individualidade das anotações, perde-se informação semântica valiosa para personalização, em especial para sistemas de recomendação de conteúdo (Manzato, 2011).

Assim, o objetivo do trabalho de Manzato foi desenvolver um *framework* para personalização de vídeo que permitisse indexação multimídia sem restrição de domínio, utilizando anotações colaborativas, mas mantendo-se a individualidade dos dados. Para isso, utilizou-se multimodalidade, como um meio de aumentar o volume de informações extraídas e, Anotações por Usuários (*Peer-level Annotations*) (Bulterman, 2004; Goularte et al., 2004; Cesar et al., 2006), para enriquecer semanticamente as descrições provenientes da indexação.

Como primeiro passo para a obtenção de semântica, desenvolveu-se um método para automaticamente segmentar vídeos em tomadas (Manzato & Goularte, 2007)^[14] de modo que as mesmas pudessem ser classificadas com base no *closed caption* geralmente presente nesse tipo de mídia (Manzato & Goularte, 2008)^[15]. A classificação gera informações úteis para personalização, por exemplo: agrupar n tomadas de diferentes vídeos classificadas como “Esporte/Futebol” para compor um vídeo personalizado a ser

^[14] MANZATO, M. G. ; GOULARTE, R. . Shot boundary detection based on intelligent systems. In: XIII Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia 2007), 2007, Gramado. Anais, Porto Alegre : SBC, 2007. v. 1., p. 190-197. (Qualis B3)

^[15] MANZATO, M. G. ; GOULARTE, R. . Video News Classification for Automatic Content Personalization: A Genetic Algorithm Based Approach. In: XIV Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia 2008), 2008, Vila Velha. Anais, Porto Alegre : SBC, 2008. v. 1., p. 36-43. DOI: 10.1145/1666091.1666100. (Qualis B3)

entregue a um usuário específico com predileção por esse tipo de conteúdo. O interesse por *closed caption* se justifica devido a esse tipo de dado poder ser facilmente transformado em texto, que por sua vez possui carga semântica explorável por técnicas bastante difundidas na área Recuperação de Informação (Baeza-Yates & Ribeiro-Neto, 1999), como Indexação Semântica Latente (*Latent Semantic Indexing* – LSI) (Deerwester et al., 1990).

Os resultados indicaram que, embora o suporte de LSI a polissemia* seja visto como uma vantagem em Recuperação de Informação, tal característica tenderia a elevar o número de falsos positivos em tarefas de classificação de tomadas. Assim, estabeleceu-se uma colaboração com a professora Alessandra Alaniz Macedo (FFCLRP - USP), pesquisadora especialista na área de Recuperação de Informação, para realizar uma investigação aprofundada sobre o desempenho do método quando aplicado à classificação de tomadas em diferentes coleções. Os resultados dessa investigação (Manzato et al., 2009)¹⁶ confirmaram que a técnica LSI, ao contrário do que se imaginava inicialmente, não é adequada para classificação. Esses resultados foram importantes para entender os limites da semântica obtida por meio do *closed caption* para atender tarefas complexas de personalização (como recomendação de conteúdo), as quais necessitam de um volume maior e mais rico de informações semânticas (Manzato et al., 2009b; Cesar et al., 2006). Além disso, *closed caption* é um tipo de dado nem sempre presente no vídeo, sendo uma desvantagem limitar sistemas de personalização a extrair informação apenas desse tipo de mídia.

Para tratar a questão do volume dos dados aproveitou-se a experiência do autor em um trabalho anterior na área de Autoria Multimídia, onde um dos resultados foi o desenvolvimento de uma ferramenta multimodal para anotações – a M4Note (Goularte et al., 2004b). A M4Note foi estendida de modo a extrair informações de outras mídias presentes no vídeo (imagens, sons, marcações, etc.), além de *closed caption*. Assim, caso um tipo de mídia não esteja presente (o *closed caption*, por exemplo), ainda existe a possibilidade de obter informações a partir das outras mídias.

Além disso, explorou-se a característica da multimodalidade de oferecer maior liberdade de expressão ao usuário, fazendo-o sentir-se mais à vontade para interagir com a

* Polissemia ocorre quando uma palavra possui múltiplos significados.

¹⁶ MANZATO, M. G. ; MACEDO, A. A. ; GOULARTE, R. . Evaluation of Video News Classification Techniques for Automatic Content Personalization. *International Journal of Advanced Media and Communication*, v. 3, n. 4, p. 383-403, 2009.

aplicação (Shevade & Sundaram, 2003; Goularte et al., 2004b). Esse tipo de interação tem sido pouco investigado no contexto de aplicações de P&A, apesar de poder conter informações semânticas latentes sobre o conteúdo e sobre o interesse do usuário pelo tipo de conteúdo acessado (Manzato et al., 2009b). Então, adotou-se como parte do *framework* uma versão estendida de M4Note capaz de capturar os dados presentes nas interações dos usuários com o conteúdo, estruturando-os em Anotações por Usuários (*Peer-level Annotations*) (Bulterman, 2004). Assim, a semântica latente presente nesse tipo de anotação pôde ser utilizada para enriquecer as descrições de alto e de baixo nível semântico provenientes de Indexação Multimídia, beneficiando aplicações de P&A (Manzato et al., 2009b)¹⁷.

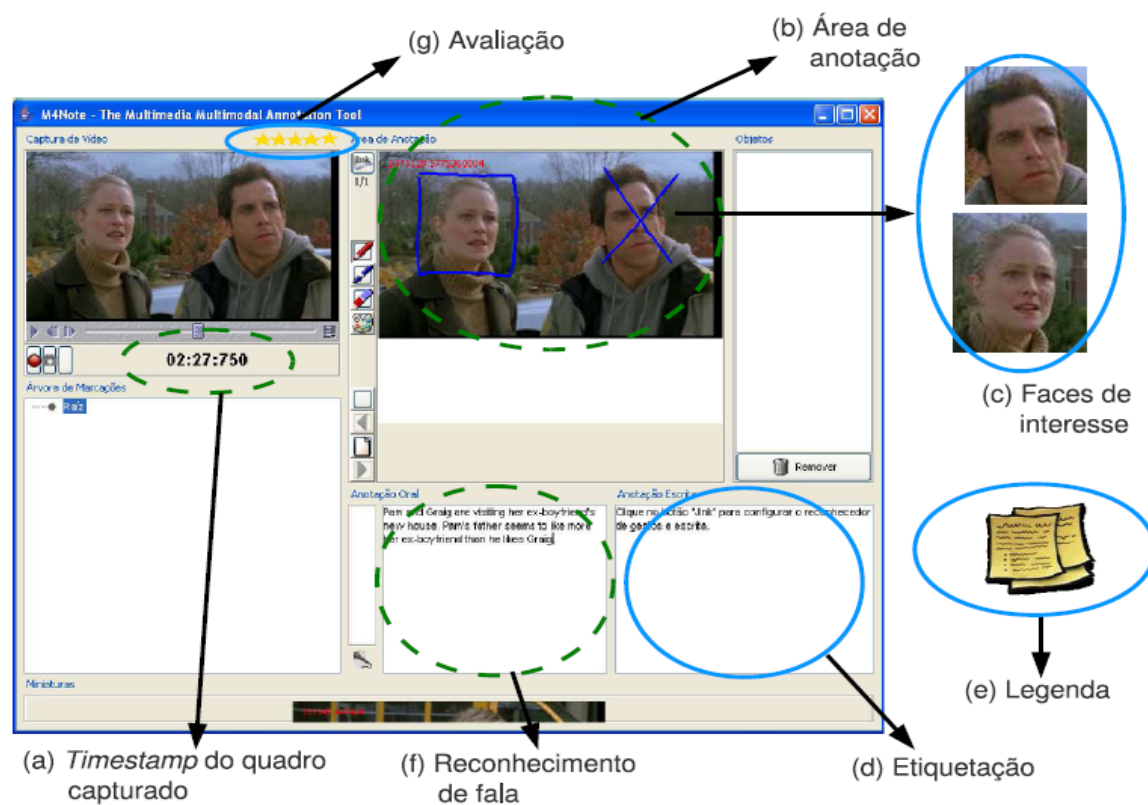


Figura 4 – Ferramenta M4Note estendida (retirado de (Manzato, 2011)).

No contexto do trabalho conduzido por Manzato, as descrições de baixo nível semântico incluem: histogramas de cores; *closed caption*; dados (coordenadas) de anotações com tinta eletrônica. Para a obtenção de descrições de alto nível semântico, Manzato contou com a ajuda do trabalho de iniciação científica de Gilson Araújo do

¹⁷ MANZATO, M. G. ; COIMBRA, D. B. ; GOULARTE, R. . Multimedia content personalization based on peer-level annotation. In: European Interactive TV Conference (EuroITV 2009), 2009, Leuven, Bélgica. Anais, New York : ACM, 2009. v. 1., p. 57-66. DOI: 10.1145/1542084.1542096. (Qualis B3)

Nascimento (Nascimento et al., 2010) na detecção de faces obtidas de quadros-chave. Outras descrições incluem: metadados de filmes (título, gênero, elenco, etc.) e perfil de usuários (identificação, preferências de gênero/assunto, histórico de acesso a itens multimídia, etc.). Ambas podem ser enriquecidas com: coordenadas da área da imagem ao redor de onde o usuário “apontou” (faces); nomes de pessoas cujas faces aparecem em momentos de interação no vídeo; classificação do assunto tratado no vídeo no momento da interação; *time stamps* do vídeo no momento da interação; texto referente a anotações realizadas via voz ou via tinta eletrônica. Essas anotações produzidas pela interação dos usuários com o conteúdo constituem Anotações por Usuários. A Figura 4 ilustra a ferramenta M4Note estendida e as informações extraídas do vídeo que podem ser utilizadas para enriquecer as descrições.

O benefício do enriquecimento das descrições pôde ser comprovado ao utilizar o *framework* no desenvolvimento de um sistema de recomendação de conteúdo baseado em Anotações por Usuários no domínio de filmes (Manzato & Goularte, 2009)^[18]. Esse sistema foi aprimorado para utilizar bases de dados reais (MovieLens Dataset* e IMDB♦) e anotações colaborativas enriquecidas com Anotações por Usuários, comparando os resultados com implementações de abordagens atuais para sistemas de recomendação. As comparações constataram o melhor desempenho da abordagem proposta (Manzato & Goularte, 2010)⁽¹⁹⁾.

Ainda, as abordagens atuais para sistemas de recomendação (filtragem por conteúdo, filtragem colaborativa e filtragem híbrida), segundo Adomavicius & Tuzhilin (2005) e Sen et al. (2009), apresentam problemas recorrentes: a) sobre-especialização (do Inglês, *overspecialization*), que implica em o sistema ser incapaz de recomendar itens diferentes dos presentes no perfil do usuário e b) novo usuário, para o qual o sistema somente irá produzir recomendações eficazes após o usuário fornecer uma quantidade relevante de avaliações. Os resultados obtidos por Manzato & Goularte (2010) demonstraram que o sistema proposto consegue reduzir tais problemas. Isso se deve, em grande parte, ao fato de as Anotações por Usuários eliminarem a impessoalidade das

[18] MANZATO, M. G. ; GOULARTE, R. . Supporting multimedia recommender systems with peer-level annotations. In: XV Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia 2009), 2009, Fortaleza. Anais, Porto Alegre : SBC, 2009. v. 1., p. 202-209. DOI: 10.1145/1858477.1858503. (Qualis B3)

* <http://www.grouplens.org/node/12>

♦ *The Internet Movie Data Base* - <http://www.imdb.com/interfaces>

(19) MANZATO, M. G. ; GOULARTE, R. . Peersommender: A peer-level annotation based approach for multimedia recommendation. *Journal of Information and Data Management*, v. 1, n. 2, p. 277-292, 2010. ISSN 2178-7107.

anotações colaborativas, uma vez que aquelas carregam, mesmo que de modo latente, impressões e características próprias de cada usuário.

Além do serviço de recomendação, utilizou-se também o *framework* no desenvolvimento de um serviço de Seleção de Conteúdo no domínio de telejornais. Os resultados obtidos, reportados em Manzato et al. (2011)²⁰, demonstram que a semântica adicionada pela multimodalidade e pelo enriquecimento de conteúdo melhora a precisão e a revocação das buscas, comprovando que o trabalho pode ser aplicado a diferentes serviços de personalização e domínios.

Apesar dos avanços obtidos, o processo de personalização oferecido pelo *framework* é ainda semi-automático devido à fase de segmentação em cenas ser realizada manualmente. Isso motivou a pesquisa em um novo tópico: segmentação automática em cenas. Uma das principais limitações deste tópico de pesquisa é a interpretação da definição de cena (Coimbra, 2011; Coimbra & Goularte, 2011): cena é definida como uma sequência de tomadas semanticamente relacionadas (Hanjalic, 2004), sendo que o desafio reside exatamente em encontrar essa relação.

Vários trabalhos citados na literatura (Misra et al., 2010; Liu et al., 2009; Wang et al., 2008; Yu et al., 2007; para mencionar alguns) utilizam uma simplificação da definição. No domínio de telejornais, por exemplo, utiliza-se a aparição do âncora (apresentador do telejornal) na imagem para determinar o início das cenas. Contudo, cenas onde mais de um ou nenhum âncora aparece não são detectadas satisfatoriamente. Isso evidencia que as técnicas atuais não conseguem tratar adequadamente segmentos de vídeo com complexidade semântica elevada. Por exemplo, um trecho de telejornal posicionado entre duas aparições do âncora contendo três segmentos, sendo que cada segmento trata de um assunto diferente. Tal trecho seria segmentado como uma única cena, quando o mais apropriado seria segmentar em três cenas diferentes.

Devido à importância do conceito de cenas para os serviços de personalização, o problema da detecção de cenas em segmentos semanticamente complexos se torna relevante. Isso motivou a orientação do trabalho de mestrado de Danilo Barbosa Coimbra

²⁰ MANZATO, M. G. ; COIMBRA, D. B. ; GOULARTE, R. . An enhanced content selection mechanism for personalization of video news programmes. *Multimedia Systems Journal*, v. 17, n. 1, p. 19-34, 2011. DOI: 10.1007/s00530-010-0204-y. (Qualis B1)

(Coimbra, 2011). Coimbra explorou técnicas audiovisuais^[21] para extração de informação semântica do fluxo de vídeo, colaborando com o trabalho de Manzato (Manzato et al, 2009b; Manzato et al., 2011). Baseado nessa experiência, Coimbra desenvolveu uma técnica multimodal para segmentação de vídeo em cenas que combina informações vindas de fontes visuais (histogramas de cor e assinatura *wavelet* da imagem), sonoras (detecção de silêncio) e textuais (*closed caption*). Os resultados relatados em Coimbra & Goularte (2011)^[22] demonstram que tal técnica é capaz de detectar cenas quando o assunto representado no vídeo muda e, também, considerar como uma única cena tomadas adjacentes que possuem imagens muito diferentes cujo assunto, contudo, é o mesmo.

A estratégia desenvolvida para a segmentação em cenas oferece uma solução para o problema dos segmentos semanticamente complexos. Porém, tal estratégia ainda é limitada ao domínio dos telejornais, carecendo de mais investigação para poder ser aplicada em outros domínios. Outro ponto é que, assim como no *framework* como um todo, o método de fusão das modalidades está baseado na utilização de *rankings* para cada modalidade (Manzato et al., 2011). Apesar de ter oferecido bons resultados em termos de precisão e revocação, os custos envolvidos no processamento individual das modalidades para obter os *rankings* ainda necessita ser investigado, assim como outros métodos de fusão. Essas duas questões estão sendo investigadas em dois trabalhos em andamento orientados pelo autor.

ORIENTAÇÕES EM ANDAMENTO

Uma razão pela qual as técnicas atuais não resolvem o problema dos segmentos semanticamente complexos é que as mesmas usam descritores locais (histogramas, assinaturas *wavelet*, coeficientes de Fourier, mapas de gradiente, etc.) para analisar as imagens presentes no vídeo. Descritores locais possuem alto poder discriminativo, sendo adequados para identificar a ocorrência de um objeto (como o âncora) em um segmento, mas não possuem poder de generalização necessário para diferenciar imagens em categorias com diferentes semânticas (Valle & Cord, 2009). Situação semelhante (descritores locais x generalização) vem sendo enfrentada na área de Recuperação de

^[21] COIMBRA, D. B. ; GOULARTE, R. . Identificação de Cenas em Vídeos Digitais Utilizando Características Audiovisuais. In: XV Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia 2009), 2009, Fortaleza - CE. Anais, Porto Alegre : SBC, 2009. v. 1, p. 43-46. DOI: 10.1145/1858477.1858520. (Qualis B3)

^[22] COIMBRA, D. B. ; GOULARTE, R. . Segmentação multimodal de cenas em telejornais. In: XVII Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia 2011), 2011, Florianópolis - SC. (Aceito para publicação como artigo completo). (Qualis B3)

Informação Baseada em Conteúdo (*Content-Based Information Retrieval* – CBIR) com o uso de técnicas *Bag of Features*, conseguindo-se identificar a categoria semântica a que pertence uma imagem independentemente do domínio (Sivic & Zisserman, 2006; Batista et al., 2009; Lazebnik et al., 2006).

A aplicação direta de *Bag of Features* no problema em questão não é suficiente para atingir a independência de domínio. Uma vez que a técnica é orientada ao agrupamento de imagens semelhantes, a mesma não seria capaz de reconhecer segmentos vizinhos visualmente diferentes, mas que tratam do mesmo assunto, como pertencentes à mesma cena. Assim, o trabalho de mestrado sendo conduzido por Bruno Lorenço Lopes tem como objetivo estender a técnica *Bag of Features* para trabalhar com multimodalidade, aproveitando o conhecimento adquirido com o trabalho de Coimbra. Assim, espera-se que os descritores visuais ajudem a aplicar a técnica a outros domínios e, que os outros descritores (sonoros e textuais), ajudem a identificar cenas compostas por segmentos visualmente diferentes.

Outro trabalho em andamento nessa linha de pesquisa é o doutorado (recém iniciado) de Johana Maria Rosas Villena, em co-orientação com a professora Renata Pontin de Mattos Fortes (ICMC-USP). Esse trabalho irá investigar, entre outras coisas, a questão da fusão das modalidades. Em geral, conforme discutido ao longo desta subseção, a utilização de multimodalidade traz dois benefícios para a linha de P&A (Manzato, 2011): 1) possibilita obter um volume maior de dados sobre o conteúdo, gerando a possibilidade de aumento das informações semânticas sobre o mesmo; 2) oferece maior liberdade de expressão ao usuário, fazendo-o sentir-se mais à vontade para interagir com o conteúdo, sendo que as informações sobre essa interação são úteis para serviços de personalização.

Por outro lado, tem-se de processar diferentes mídias, o que aumenta a carga computacional necessária para realizar as tarefas. Ainda, quando mais de uma mídia é utilizada, os problemas presentes no projeto de interfaces de aplicações também aumentam (Talarico Neto, 2011). Estratégias de fusão de modalidades podem reduzir tanto a carga computacional necessária (Snoek & Worring, 2005; Xie et al., 2007) quanto os potenciais problemas de interface (Talarico Neto, 2011), contudo, tendem a diminuir o volume de informações semânticas que se consegue extrair e, conseqüentemente, diminuir a eficiência dos serviços de personalização.

Desse modo, torna-se relevante encontrar um meio de realizar a fusão das modalidades que ofereça equilíbrio entre reduzir a carga computacional necessária e os

potenciais problemas para o projeto de interfaces sem, contudo, diminuir a carga semântica das informações.

2.2.3 Resumo

Na linha de pesquisa Personalização e Adaptação de Conteúdo foram orientados três trabalhos de mestrado, um de doutorado e um de iniciação científica. Foram publicados 15 artigos científicos, sendo quatro em periódicos internacionais, seis em conferências nacionais e cinco em conferências internacionais. Os trabalhos contaram com a participação ativa dos alunos e dos pesquisadores Rodrigo Fernandes de Mello, Laurence T. Yang, Alessandra Alaniz Macedo e Renata Pontin de Mattos Fortes. O autor participou de dois projetos financiados por agências de fomento: o projeto VIMOS (FINEP) e o projeto TIDIA-Ae (FAPESP). Atualmente a linha possui duas orientações em andamento, sendo uma de mestrado e uma de doutorado.

2.3 Televisão Digital Interativa

Nesta seção, são apresentados os resultados obtidos na linha de pesquisa Televisão Digital Interativa (TVDI). A TVDI é parte do processo de convergência digital experimentado nas últimas décadas, que tem se caracterizado pela fusão de tecnologias, estabelecimento de padrões (tais como aqueles da família MPEG) e pelo surgimento de novos produtos e aplicações. Além da melhoria de qualidade do sinal, a transmissão digital permite uma utilização mais eficiente da banda e abre espaço para que uma série de novos programas e novos serviços sejam agregados ao conteúdo televisivo tradicional, dentre eles a interatividade (Srivastava, 2002; Souza Filho et al., 2007). No Sistema Brasileiro de TV Digital (formalmente conhecido como ISDB-Tb)[▲], o *middleware* Ginga[■] é o responsável por prover a interatividade, fornecendo meio de desenvolver e executar aplicações em Java (Ginga-J) e em NCL (Ginga-NCL).

A implantação efetiva de serviços de TVDI implica em alterações de toda a cadeia produtiva de radiodifusão, através da modernização da infra-estrutura da rede de transmissão de sinais, da produção de seus insumos para esta infra-estrutura (software, equipamentos e componentes) e, finalmente, da pesquisa e do desenvolvimento de novos

[▲] Fórum do Sistema Brasileiro de TV Digital Terrestre. <http://www.forumsbtvd.org.br/>

[■] <http://www.ginga.org.br/>

conteúdos digitais (Srivastava, 2002). É nesse último ponto que a pesquisa do autor está inserida, no contexto da área Autoria Multimídia.

Os novos conteúdos digitais correspondem, na prática, a aplicações multimídia convergentes que devem satisfazer a uma série de requisitos clássicos ligados à sincronização de mídias e à adequação ao contexto. Entretanto, existem problemas específicos que devem ser enfrentados para que seja possível oferecer aplicações interativas de modo eficiente (Piccolo & Baranauskas, 2006).

Dentre esse problemas está a limitação imposta pelo uso do controle remoto como mecanismo de interação com as aplicações de TVDI (Piccolo & Baranauskas, 2006). Em paralelo a isso se observa a falta de suporte adequado para atender a um fenômeno recente: usuários de aplicações multimídia interativas deixaram de ser meros consumidores de informação e passaram a atuar também como produtores de conteúdo (Magalhães & Pereira, 2004), criando ou alterando conteúdo e realizando anotações sobre suas impressões pessoais (exemplos de serviços desse tipo são o YouTube* e o Facebook♦).

2.3.1 Contribuições em TV Digital Interativa

As questões levantadas no início da Seção 2.3 têm levado pesquisadores a investigar novos métodos de interação entre o usuário e o conteúdo, explorando interfaces naturais (como gestos, voz e escrita) (Abowd et al., 2002) e multimodalidade (Goularte et al. 2004b). Essas questões levaram também à colaboração do autor com os pesquisadores: Maria da Graça Campos Pimentel (ICMC-USP), especialista em Computação Ubíqua e Autoria Multimídia e César Teixeira (UFSCar), especialista em Multimídia e TVDI. Os resultados dessa colaboração, descritos a seguir, tratam de modo geral dos problemas relativos à captura dos dados vindos da interação, da formalização dos mesmos em estruturas bem definidas para não perder interoperabilidade e da definição de como tais estruturas poderão ser utilizadas na composição de documentos multimídia.

Em Goularte et al. (2004)^[23] discute-se as vantagens em se utilizar multimodalidade para realizar anotações digitais associadas a vídeo. Nesse trabalho descreve-se como anotações por voz e por tinta eletrônica podem complementar a tradicional anotação de

* <http://www.youtube.com>

♦ <http://www.facebook.com>

[23] GOULARTE, R.; CATTELAN, R. G.; CAMACHO-GUERRERO, J. A.; INÁCIO, JR., V. R.; PIMENTEL, M. G. C. Interactive multimedia annotations: enriching and extending content. In: ACM Symposium on Document Engineering (DocEng' 04), 2004, Milwaukee, Wisconsin, USA. Anais, New York : ACM Press, 2004, v. 1, p. 84–86. DOI: 10.1145/1030397.1030414.

vídeo (*video tagging*) baseada em hierarquias pré-definidas de etiquetas. A principal vantagem da multimodalidade, nesse contexto, é tornar a tarefa de anotação menos laboriosa e mais atrativa para o usuário. Em Goularte et al. (2004b)^[24], são apresentados os detalhes da ferramenta desenvolvida para realizar esse tipo de anotação (M4Note). Em particular, demonstra-se como as informações capturadas da interação (como as marcas com tinta eletrônica) podem ser estruturadas em um documento multimídia.

Esses dois trabalhos foram importantes para definir um modelo de estruturação de objetos multimídia que pudesse ser aplicado em TVDI. Tal modelo, o MediaObjectc, é descrito em Goularte et al. (2006)⁽²⁵⁾, tendo como característica ser mais econômico em relações a seus similares na descrição de objetos de mídia sem, no entanto, perder representatividade.

Os conhecimentos adquiridos no desenvolvimento do modelo MediaObject e da ferramenta M4Note motivaram a investigação de novos modos de interação e de autoria em vídeo que pudessem ser aplicados em TVDI, levando a proposição de um projeto de pesquisa – o i2TVD.

PROJETOS

O projeto Serviços e Aplicações Inovadoras para TV Digital Interativa – i2TVD foi aprovado no contexto do edital CAPES RH–TVD 01/2007 (com vigência até 2013), cujo objetivo principal é a formação de recursos humanos qualificados para a produção de conteúdo em TVDI. Por sua vez, o i2TVD tem como meta desenvolver técnicas e aplicações para explorar modos de interação não convencionais (além do controle remoto) entre o usuário e conteúdo em TVDI. Esse projeto, do qual Pimentel é coordenadora e o autor é participante, integra pesquisadores de universidades e institutos de pesquisa brasileiros (USP, UFSCar, FUCAPI) e estrangeiros (CWI – Holanda).

No contexto do i2TVD, a participação do autor se deu colaborando com Pimentel na investigação de uma tarefa comum: assistir vídeos com outras pessoas, realizar pausas e fazer comentários. Esse paradigma de assistir-e-comentar é familiar para os usuários. Por

[24] GOULARTE, R.; CAMACHO-GUERRERO, J. A.; INACIO JR., V. R.; CATTELAN, R. G.; PIMENTEL, M. G. C. M4Note: A Multimodal Tool for Multimedia Annotations. In: 10th Brazilian Symposium on Multimedia and the Web 2nd Latin American Web Congress (WebMedia & LA-Web 2004 Joint Conference), 2004, Ribeirão Preto, SP, Brasil. Anais, Washington, DC : IEEE Computer Society, 2004, p. 142–149. DOI: 10.1109/WEBMED.2004.1348161. (Qualis B3)

(25) GOULARTE, R.; PIMENTEL, M. G. C.; MOREIRA, E. D. S. Context-Aware Support in Structured Documents for Interactive-TV. Multimedia Systems. Springer, v. 11, n. 4, p. 367-382, 2006. DOI: 10.1007/s00530-006-0014-4. (Qualis B1)

outro lado, quando necessitam registrar seus comentários (anotações), compartilhar suas opiniões ou impressões ou então editar o vídeo, os usuários utilizam uma miríade de softwares e ferramentas de edição (Kientz et al., 2007).

Em Pimentel et al. (2007)^[26] discute-se como esse processo laborioso de edição e produção de vídeo pode ser facilitado, em especial para usuários comuns. Nesse trabalho são propostos alguns requisitos para realizar autoria de vídeo de modo transparente, gerando-se vídeo interativo automaticamente com base nas anotações do usuário. Também são propostos métodos para que as anotações, por voz ou por tinta eletrônica, sejam capturadas durante a apresentação de um vídeo e representadas de modo estruturado. A representação inclui uma série de atributos e operações que permitem, entre outras coisas, sincronizar vídeo e anotações, além de permitir a geração de um documento NCL^{*} contendo uma versão interativa do vídeo com as anotações.

A proposta foi implementada (Pimentel et al., 2008)^[27] como uma ferramenta para execução em ambiente TVDI, capturando as anotações realizadas e gerando um vídeo interativo automaticamente. Essa versão da ferramenta foi estendida (Cattelan et al., 2008)⁽²⁸⁾ de duas maneiras: 1) as interações do usuário foram associadas a comandos de edição de vídeo (como *loop*, *seek*, *skip* e *slow motion*), demonstrando que a autoria pode ser realizada de modo transparente; 2) adicionando-se à arquitetura da ferramenta um serviço *Peer-to-Peer* (P2P) integrado ao ambiente de execução NCL, de modo que as anotações e/ou o vídeo dos usuários pudessem ser compartilhados.

Em um outro projeto na linha de TVDI^{*} foi formada uma rede de pesquisa constituída por ICMC-USP-São Carlos, UFSCar, UNICAMP, CENPRA, PUC-RIO, UFPB, C.E.S.A.R., Instituto Genius. Nesse projeto cada instituição da rede esteve responsável por resultados e focos específicos, todos visando a avaliação e a apresentação

[26] PIMENTEL, M. G. C.; GOULARTE, R.; CATTELAN, R. G.; SANTOS, F. S.; TEIXEIRA, C. Enhancing multimodal annotations with pen-based information. In: Workshop on New Techniques for Consuming, Managing, and Manipulating Interactive Digital Media at Home, 2007, Taichung. IEEE International Symposium on Multimedia, 2007. v. 1, p. 207-213. DOI: 10.1109/ISM.Workshops.2007.43.

* NCL – *Nested Context Language*. É a linguagem declarativa do Sistema Brasileiro de TV Digital para a autoria de programas interativos.

[27] PIMENTEL, M. G. C.; GOULARTE, R.; CATTELAN, R. G.; SANTOS, F. S.; TEIXEIRA, C. Ubiquitous Interactive Video Editing via Multimodal Annotations. In: European Interactive TV Conference (EuroITV 2008), 2008, Salzburg. Anais, Heidelberg : Springer-Verlag, (LNCS) v. 5066, p. 72-81, 2008. DOI: 10.1007/978-3-540-69478-6_8. (Qualis B3)

(28) CATTELAN, R. G.; TEIXEIRA, C.; GOULARTE, R.; PIMENTEL, M. G. C. Watch-and-comment as a paradigm toward ubiquitous interactive video editing. ACM Transactions on Multimedia Computing Communications and Applications, v. 4, n. 4, p. 1-24, 2008. DOI: 10.1145/1412196.1412201. (Qualis A1)

* Avaliação do Middleware Ginga. Projeto FINEP Número: 01.07.0110.00. (2007-2008).

de contribuições ao *middleware* Ginga adotado pelo Sistema Brasileiro de TV Digital. A equipe do ICMC-USP-São Carlos ficou responsável pela extensão do *middleware* Ginga para permitir a apresentação de conteúdo MPEG-4 com alta interatividade.

Um dos resultados foi o trabalho descrito em (Manzato et al., 2008)^[29], onde o codificador MPEG-J desenvolvido (Seção 2.1.1) foi utilizado para encapsular programas Java em vídeos MPEG-4 a serem exibidos em ambiente Ginga-J. O protótipo desenvolvido permite que o usuário assista a um vídeo enquanto o programa Java embutido se conecta a um servidor de notícias da Web (RSS*), capturando o fluxo de texto e exibindo a notícia ao usuário. Esse trabalho, em particular, demonstra a sinergia entre as linhas de pesquisa Codificação Multimídia e TVDI.

O projeto xTVD*, no contexto do Programa de Cooperação Acadêmica – Novas Fronteiras da CAPES, é uma cooperação entre a Universidade Salvador (UNIFCAS) em Salvador-BA, as universidades USP e UFSCar em São Carlos - SP e o Instituto CWI em Amsterdã na Holanda. O projeto tem como um de seus objetivos o desenvolvimento de serviços e aplicações para TV Digital Interativa explorando os diferentes recursos dessa nova plataforma multimídia convergente para criar e reproduzir conteúdo interativo estruturado de forma não-linear. Um dos desafios do projeto é adaptar processos de desenvolvimento, métodos, modelos e técnicas utilizados na construção de aplicações computacionais (em especial, aquelas ligadas à multimídia e hipermídia) para a plataforma TVDI.

Um dos resultados desse projeto é a proposta de uma arquitetura para educação a distância utilizando a TVDI (Oliveira et al., 2009)^[30]. Os trabalhos relacionados a esse tema foram projetados em sua maioria para ambiente Web, diferente do ambiente TVDI que possui limitações de canal de retorno e de processamento (Piccolo & Baranauskas, 2006), ou então para o *middleware* europeu MHP (Morris & Smith-Chaigneau, 2005), sendo incompatíveis com o sistema brasileiro. A proposta utiliza conceitos do modelo

^[29] MANZATO, M. G. ; JUNQUEIRA, D. C. ; GOULARTE, R. . Interactive News Documents for Digital Television. In: XIV Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia 2008), 2008, Vila Velha. Anais, Porto Alegre : SBC, 2008. v. 2. p. 1-4. DOI: 10.1145/1809980.1809982. (Qualis B3)

* RSS – *Rich Site Summary*. São notícias encapsuladas em arquivos XML, disponíveis para acesso em portais Web.

♦ xTVD: Explorando Novas Fronteiras na TV Digital Interativa. Edital CAPES Procad-NF 08/2008. (2009-2012).

^[30] OLIVEIRA, L. S.; MARTINS, D. S.; GOULARTE, R.; PIMENTEL, M. G. C. EducaTV: uma arquitetura para acesso a conteúdos educacionais via TVDi. In: XV Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia'09), 2009, Fortaleza. Anais, Porto Alegre : SBC, 2009. v. 2, p. 63-66. DOI: 10.1145/1858477.1858525. (Qualis B3)

MVC (*Model-View-Controller*) (Gamma et al.; 1994) para separar os componentes lógicos e opera sobre a plataforma GINGA-J. Os objetos de aprendizagem utilizados no protótipo são constituídos por apresentações multimídia assistidas pela TV e que podem conter avaliações, glossários, áudio narrativo, vídeo, entre outros. Os objetos de aprendizagem são providos como serviços interativos de valor agregado, associados à programação educativa em vídeo.

Outro resultado no contexto do projeto xTVD foi a proposta de um serviço para auxiliar usuários de TVDI a navegar pelo conteúdo audiovisual recentemente assistido ou gravado. O serviço apresenta ao usuário uma linha de tempo contendo miniaturas de quadros de vídeo associados aos últimos programas assistidos, considerando também trocas de canais. A especificação do serviço, em ambiente GINGA-NCL, pode ser encontrada em Vega-Oliveros et al. (2009)^[31], enquanto a implementação está detalhada em Pedrosa et al. (2009)^[32].

ORIENTAÇÕES EM ANDAMENTO

No contexto do projeto i2TVD o autor está orientando o trabalho de mestrado de Jorge Rodrigues Carvalho. O objetivo é investigar o uso acelerômetros como dispositivo de entrada de dados complementar ao controle remoto em ambiente TVDI. Acelerômetros possuem atualmente custo reduzido, embutidos em diferentes dispositivos (como em consoles para jogos eletrônicos). Esses sensores são capazes de detectar mudança de direção e de aceleração, sendo, portanto, úteis para reconhecimento de gestos e, por consequência, de escrita. Na literatura, até o momento, não foram encontrados modelos de descrição padronizados e interoperáveis para os dados advindos de acelerômetros. Desse modo, Carvalho está investigando meios de tratar esses dados permitindo que sejam manipulados por aplicações diversas. Isto deve ser feito por meio da proposta de uma arquitetura em camadas e de um formato estruturado de dados, de modo que as informações sejam interoperáveis.

^[31] VEGA-OLIVEROS, D. A.; PEDROSA, D. C.; PIMENTEL, M. G. C.; GOULARTE, R. Navegação em Vídeo via Quadros Recentes. In: XV Simpósio Brasileiro de Sistemas Multimídia e Web, 2009, Fortaleza. Anais, Porto Alegre : SBC, 2009. v. 2, p. 15-18. DOI: 10.1145/1858477.1858513. (Qualis B3)

^[32] PEDROSA, D. C.; VEGA-OLIVEROS, D. A.; PIMENTEL, M. G. C.; GOULARTE, R. Uma Aplicação NCL/Lua para Navegação em Vídeo via Quadros Recentes. In: Workshop de Ferramentas e Aplicações - XV Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia'09), 2009, Fortaleza. Anais, Porto Alegre : SBC, 2009. v. 2, p. 136-138.

2.3.2 Resumo

A linha de pesquisa Televisão Digital Interativa é marcada pela forte colaboração do autor com os professores Maria da Graça Campos Pimentel e Edson dos Santos Moreira, ambos do ICMC-USP. Ainda, o autor participa de projetos de colaboração, ambos financiados pela CAPES, que incluem pesquisadores de vários centros: César Teixeira (UFSCar), Celso Saibel Santos (UNIFACS/UFBA), Pablo César e Dick Bulterman (CWI, Holanda). Os trabalhos tiveram participações e influências desses pesquisadores. Foram publicados dois artigos em periódicos internacionais, cinco em conferências nacionais e três em conferências internacionais. Atualmente o autor orienta um aluno de mestrado nessa linha.

2.4 Considerações Finais

A Figura 5 apresenta a distribuição das publicações nas linhas de pesquisa do autor. A numeração corresponde ao número da nota de rodapé utilizada para introduzir a referência no decorrer deste capítulo. Como mencionado anteriormente, os círculos correspondem às publicações em periódicos, enquanto que os quadrados correspondem às publicações em conferências. As publicações nacionais são apresentadas com linha tracejada. As dez publicações mais relevantes são apresentadas com fundo cinza, sendo que cópias dessas publicações podem ser encontradas nos Apêndices de A a J.

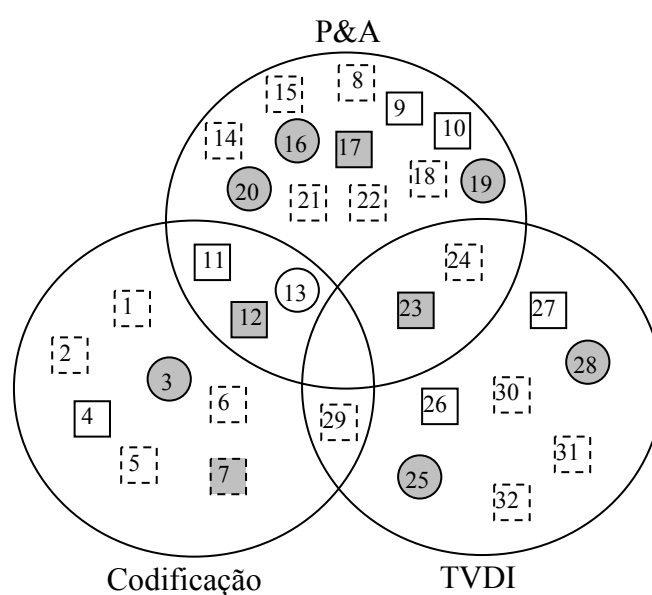


Figura 5 – Distribuição das publicações nas linhas de pesquisa: sinergias.

Os trabalhos estão classificados de acordo com a contribuição principal de cada um. Apesar de vários trabalhos possuírem algum nível de intersecção, conforme é possível verificar ao longo deste capítulo, alguns claramente possuem contribuições em mais de uma linha. Esses estão destacados nas intersecções dos círculos. Isso evidencia a busca do autor em aproveitar características de uma linha em particular que possam beneficiar outras, promovendo sinergias.

Na Figura 6, são apresentadas as publicações obtidas após a conclusão do doutorado, separadas por ano e por subárea dentro da área de Multimídia. Pode-se observar que ao longo de todo o período o autor esteve preocupado em distribuir as publicações entre veículos nacionais e internacionais. Inicialmente, os trabalhos estavam concentrados nas subáreas de Autoria e Codificação Multimídia, coincidindo com os temas de pesquisa dos projetos em que o autor participava. O perfil se altera a partir de 2008, quando aparecem os primeiros resultados na subárea de Recuperação Multimídia, frutos do primeiro doutorado orientado pelo autor investigando temas como Recuperação de Informação e Recuperação Baseada em Conteúdo aplicados à linha de Personalização e Adaptação de Conteúdo.

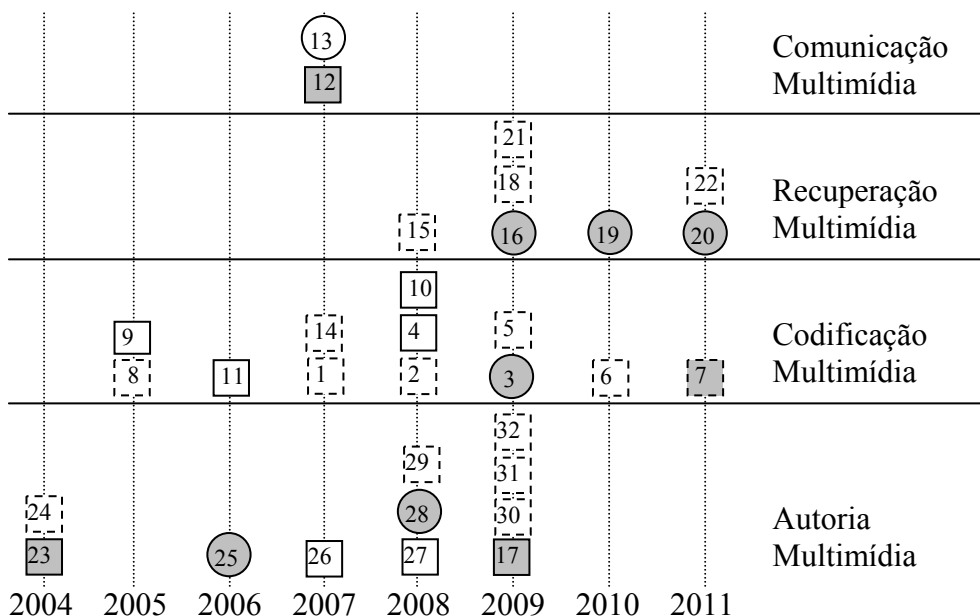


Figura 6 – Distribuição das publicações nas subáreas.

A lacuna observada nos anos de 2010 e 2011 em relação à área de Autoria Multimídia se deve às publicações terem sido classificadas de acordo com a área da contribuição principal. De fato, alguns trabalhos em Recuperação Multimídia (19, 20 e 22,

em particular) possuem intersecções com Autoria uma vez que se baseiam em Anotações por Usuários para produzir novos conteúdos.

Outro ponto a se destacar é que o autor atuou nas quatro subáreas em que Multimídia se divide (conforme ilustra a Figura 6), obtendo-se uma visão ampla da mesma. Vale lembrar que foram investigadas partes dessas subáreas, sempre direcionando os conhecimentos de modo a buscar sinergia entre linhas de pesquisa (Codificação de Vídeo, P&A e TVDI) e sempre centralizando as ações na mídia vídeo.

Conclusões

Este capítulo apresenta as conclusões deste trabalho, discutindo, na Seção 3.1, o engajamento do autor nas linhas de pesquisa aqui reportadas assim como o seu engajamento no cenário acadêmico. A Seção 3.2 aponta direções futuras para dar continuidade às pesquisas do autor.

3.1 Discussão e Reflexões

Após a conclusão do doutorado, o autor obteve resultados importantes na área de Multimídia. Essa importância é demonstrada pela evolução das pesquisas conforme aqui relatadas. Pode-se notar que os trabalhos de orientação e as participações em projetos vêm resultando em publicações cada vez mais significativas em termos de qualificação, segundo a comunidade da área de Computação.

Três linhas de pesquisa em Multimídia foram investigadas, sendo dada maior ênfase à Personalização e Adaptação de Conteúdo (P&A), podendo-se considerar essa como a linha principal do autor. Os resultados relevantes obtidos dizem respeito a como adaptar vídeo, ao vivo em particular, para uso em aplicações interativas. Ainda, dizem respeito a como extrair informações durante a interação do usuário com aplicações que usam vídeo para, entre outras coisas, enriquecer semanticamente o processo de Indexação Multimídia.

Outra linha de investigação em que o autor atuou é a Televisão Digital Interativa (TVDI), explorando multimodalidade e interfaces naturais, como voz e escrita manual com tinta eletrônica, para interagir com vídeo. Os resultados obtidos são métodos e ferramentas

para anotação e edição interativas em vídeo no ambiente TVDI. Os métodos para anotação, em particular, constituem um importante ponto de sinergia com a linha de P&A. Tais métodos são utilizados para enriquecer conteúdo aumentando a carga semântica que se pode extrair do mesmo, ainda que de modo latente, melhorando a resposta em sistemas de Recomendação e Seleção de Conteúdo. Por outro lado, existe demanda por esses tipos de sistemas em ambientes de TVDI.

A última, mas não menos importante, linha de pesquisa do autor é Codificação de Vídeo. Essa linha pode ser considerada base para as outras duas uma vez que o foco de atuação do autor nas três linhas é vídeo. Assim, os conhecimentos sobre padrões de vídeo e sobre técnicas de codificação puderam ser empregados de modo a favorecer o desenvolvimento de métodos de adaptação e de personalização de conteúdo, assim como favorecer o desenvolvimento de métodos de interação com vídeo em ambiente TVDI. Por outro lado, as demandas das outras duas linhas por incluir vídeo em aplicações interativas constituíram campo fértil para aplicação de vídeo codificado, promovendo sinergia.

Um ponto a se destacar é que neste texto foram incluídos apenas os trabalhos desenvolvidos após a conclusão do doutorado e que foram resultados: diretos de orientações, de colaborações sistemáticas do autor com outros pesquisadores ou de colaborações no contexto de projetos de pesquisas. Assim, diversas publicações anteriores à conclusão do doutorado ou resultantes de colaborações pontuais não foram incluídas, as quais totalizam quatro publicações em revistas, dezesseis publicações em conferências e um capítulo de livro.

Com relação às atividades desenvolvidas no ICMC-USP, destaca-se o engajamento do autor no laboratório de pesquisa Intermídia, que agrega pesquisadores importantes das áreas de Hipermídia, Web e Multimídia. Essa participação nas atividades do laboratório tem sido produtiva, refletindo a visão de combinar as experiências vindas dessas áreas (sinergia) e possibilitando uma troca de ideias benéfica. Desse modo, saliento o reconhecimento de que as parcerias com os professores Maria da Graça C. Pimentel, Renata Pontin de Mattos Fortes, Edson dos Santos Moreira e alunos, foram essenciais para os resultados aqui reportados.

Na comunidade científica o autor possui atualmente uma boa inserção, sendo membro do comitê de programa do Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia), o principal evento nacional na sua área de pesquisa. O autor foi *co-chair* do WebMedia na edição de 2009. O autor também é membro do comitê de programa do

European Interactive TV Conference (EuroITV), do *ACM International Conference on Design of Communication (ACM SIGDOC)* e do *ACM Symposium on Applied Computing - Multimedia & Visualization (ACM SAC-MMV)*.

3.2 Trabalhos Futuros

O autor pretende continuar investigado nas três linhas de pesquisa descritas neste texto, com particular interesse em aprofundar as pesquisas em técnicas de extração de características ligadas à Recuperação Baseada em Conteúdo. Espera-se que o uso de multimodalidade possa resultar em novos métodos aplicáveis à linha de P&A, sobretudo para contribuir com as pesquisas já em andamento sobre segmentação de vídeo em cenas e para iniciar a investigação sobre sistemas de Sumarização de Vídeo.

Outro trabalho futuro é explorar vídeo estéreo em ambientes de TVDI. Nota-se atualmente a proliferação dos chamados dispositivos 3D, como televisores, câmeras fotográficas e consoles de jogos eletrônicos. Como as aplicações multimídia vêm experimentando uma tendência de convergência, espera-se que em breve a produção e disseminação de conteúdo televisivo em 3D aumente, criando uma demanda por aplicações interativas 3D em TVDI. Para suprir essa demanda será necessário o desenvolvimento de novas técnicas para manipulação interativa desse tipo de conteúdo, com impactos no *middleware* Ginga, nos processos de autoria de aplicações e nos métodos de codificação de vídeo estéreo. Nesse último ponto, além das pesquisas em codificação espacial (como as que estão em curso no trabalho do autor), serão necessárias também investigações sobre codificação temporal.

Por último, espera-se retomar investigações na subárea Comunicação Multimídia. Recentemente o autor se tornou participante do Instituto Nacional de Ciência e Tecnologia em Sistemas Embarcados Críticos (INCT-SEC), do qual é também integrante o prof. Edson dos Santos Moreira. Moreira foi orientador de mestrado e doutorado do autor e, no contexto do INCT-SEC, uma de suas frentes de atuação é a transmissão de vídeo a partir de veículos aéreos não tripulados. Isso demanda, entre outras coisas, adaptar a taxa de bits do vídeo às condições de transmissão das diferentes redes pelas quais o veículo possa passar durante um sobrevôo.

Agradecimentos

Os pesquisadores e alunos aqui citados tiveram participação fundamental nas atividades de pesquisas desenvolvidas. Em especial agradeço aos professores Maria da Graça C. Pimentel, Renata Pontin de M. Fortes e Edson dos Santos Moreira.

Agradeço à Solange, à Renata e à Graça pelo incentivo.

Agradeço aos funcionários do ICMC-USP pelo suporte.

Agradeço às agências de fomento FAPESP, CAPES, CNPq e FINEP pelo auxílio financeiro aos projetos e bolsas de pesquisa.

Agradeço à Elaine pelo apoio e compreensão.

Referências Bibliográficas

- (Abowd et al., 2002) - ABOWD, G.; MYNATT, E.; RODDEN, T. The human experience. *IEEE Pervasive Computing*, v.1, n. 1, p. 48-57.
- (Adomavicius & Tuzhilin, 2005) - ADOMAVICIUS, G. ; TUZHILIN, A. . Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE Transactions on Knowledge and Data Engineering*, v. 17, n. 6, p. 734–749, 2005. DOI: 10.1109/TKDE.2005.99.
- (Ahmad, 2005) - AHMAD, I.; XIAOHUI WEI; YU SUN; YA-QIN ZHANG. Video Transcoding: An Overview of Various Techniques and Research Issues. *IEEE Transactions on Multimedia*, v. 7, n. 5, p. 793–804, 2005. DOI: 10.1109/TMM.2005.854472.
- (Andrade & Goularte, 2009) - ANDRADE, L. A.; GOULARTE, R. Percepção Estereoscópica Anaglífica em Vídeos Digitais Comprimidos com Perda. In: XV Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia 2009), 2009, Fortaleza, Brasil. Anais. Porto Alegre : SBC, 2009, v. 1, p. 226-233. DOI: 10.1145/1858477.1858506.
- (Andrade & Goularte, 2010) - ANDRADE, L. A.; GOULARTE, R. Uma Análise da Influência da Subamostragem de Crominância em Vídeos Estereoscópicos Anaglíficos. In: XVI Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia 2010), 2010, Belo Horizonte, Brasil. Anais. Porto Alegre : SBC, 2010, v. 1, p. 179-186.
- (Azevedo & Conci, 2003) - AZEVEDO, E.; CONCI, A. *Computação gráfica: teoria e prática*. Editora Campus, Rio de Janeiro, 368 p., 2003. ISBN 8535212523.
- (Baeza-Yates & Ribeiro-Neto, 1999) - BAEZA-YATES, R. A.; RIBEIRO-NETO, B. *Modern Information Retrieval*. Addison-Wesley, New York, USA. 1999. ISBN 0-201-39829-X.
- (Barrios et al., 2005) - BARRIOS, V. M. G.; MÖDRITSCHER, F.; GÜTL, C. . Personalization versus Adaptation? A User-centred Model Approach and its Application. In: 5th International Conference on Knowledge Management (i-Know 2005), 2005, Graz, Austrália. Anais, p. 120–127. Endereço Web: http://i-know.know-center.tugraz.at/previous_i_know/i_know_05/i_know_05_scientific_track (acesso em: 18/07/2011).
- (Batista et al., 2009) - BATISTA, N. C.; LOPES, A. P. B.; ARAUJO, A. A. Detecting buildings in historical photographs using bag-of-keypoints. In: XXII Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI'09), 2009, Rio de Janeiro-RJ, Brasil. Anais, Los Alamitos : IEEE Computer Society Press, 2009. v. 1, p. 276–283. DOI: 10.1109/SIBGRAPI.2009.31.
- (Blanken, 2007) - BLANKEN, H.M.; DE VRIES, A.P.; BLOK, H.E.; FENG, L. (Eds.). *Multimedia Retrieval*. Springer, Berlim, Alemanha, 372 p., 2007. ISBN: 978-3-540-72894-8.
- (Brezeale & Cook, 2007) - Brezeale, D. ; Cook, D. J. Automatic Video Classification: A Survey of the Literature. *IEEE Transactions on Systems, Man, and Cybernetics*, v. 38, n. 3, p. 416–430, 2007. DOI: 10.1109/TSMCC.2008.919173.
- (Bulterman, 2004) – BULTERMAN, D. C. A. Animating Peer-Level Annotations Within Web-Based Multimedia. In: 7th Eurographics Workshop on Multimedia, 2004, Nanjing, China. Anais, p. 49–57, 2004. Endereço Web: <https://diglib.eg.org/EG/DL/WS/EGMM/MM04/049-057.pdf.abstract.pdf;internal&action=paperabstract.action> (acesso em: 19/07/2011).

- (Carvalho et al., 2007) - CARVALHO, D. F.; CHIES, R.; GOULARTE, R. . Esteganografia em vídeos MPEG-4. In: XIII Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia 2007), 2007, Gramado. Anais, Porto Alegre : SBC, 2007. v. 2, p. 5-8.
- (Carvalho, 2008) – CARVALHO, D. F. . Esteganografia em vídeos comprimidos MPEG-4. Dissertação (Mestrado – Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2008. 77 p.
- (Carvalho et al., 2008) - CARVALHO, D. F.; CHIES, R.; GOULARTE, R.. MP4Stego: Esteganografia em Vídeos MPEG-4. In: XIV Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia 2008), 2008, Vila Velha. Anais, Porto Alegre : SBC, 2008. v. 1. p. 154-161. DOI: 10.1145/1666091.1666118.
- (Carvalho et al., 2008b) - CARVALHO, D. F.; CHIES, R.; FREIRE, A. P.; MARTIMIANO, L. A. F.; GOULARTE, R. . Video Steganography for Confidential Documents: Integrity, Privacy and Version Control. In: ACM International Conference on Design of Communication (SigDoc'08), 2008, Lisboa. Anais, New York : ACM Press, 2008. v. 1, p. 199-206. DOI: 10.1145/1456536.1456578.
- (Carvalho & Goularte, 2009) - CARVALHO, D. F.; GOULARTE, R. . MP4Stego: Steganography in MPEG-4 Videos. International Journal of Advanced Media and Communication, v. 3, n. 4, p. 440-460, 2009. DOI: 10.1504/IJAMC.2009.028712.
- (Cattelan et al., 2008) - CATTELAN, R. G.; TEIXEIRA, C.; GOULARTE, R. ; PIMENTEL, M. G. C. Watch-and-comment as a paradigm toward ubiquitous interactive video editing. ACM Transactions on Multimedia Computing Communications and Applications, v. 4, n. 4, p. 1-24, 2008. DOI: 10.1145/1412196.1412201.
- (Cesar et al., 2006) - CESAR, P.; BULTERMAN, D. C. A.; JANSEN, A. J. An Architecture for End-User TV Content Enrichment. Journal of Virtual Reality and Broadcasting, v. 3, n. 9, 2006. DOI: urn:nbn:de:0009-6-7594.
- (Coimbra & Goularte, 2009) - COIMBRA, D. B. ; GOULARTE, R. . Identificação de Cenas em Vídeos Digitais Utilizando Características Audiovisuais. In: XV Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia 2009), 2009, Fortaleza - CE. Anais, Porto Alegre : SBC, 2009. v. 1, p. 43-46. DOI: 10.1145/1858477.1858520.
- (Coimbra & Goularte, 2011) - COIMBRA, D. B. ; GOULARTE, R. . Segmentação multimodal de cenas em telejornais. In: XVII Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia 2011), 2011, Florianópolis - SC. (Aceito para publicação como artigo completo).
- (Coimbra, 2011) – COIMBRA, D. B. . Segmentação de cenas em telejornais: uma abordagem multimodal. Dissertação (Mestrado – Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2011. 126 p.
- (Colace et al., 2005) - COLACE, F.; FOGGIA, P.; PERCANNELLA, G. A probabilistic framework for tv-news stories detection and classification. In: IEEE International Conference on Multimedia and Expo (ICME 2005), 2005, Amsterdam. Anais: Los Alamitos : IEEE Computer Society Press, 2005. v. 1, p. 1350-1353. DOI: 10.1109/ICME.2005.1521680.
- (Conklin et al., 2001) - CONKLIN, G. J.; GREENBAUM, G. S.; LILLEVOLD, K. O.; LIPPMAN, A. F.; REZNIK, Y. A. . Video Coding for Streaming Media Delivery on the Internet. IEEE Transactions on Circuits and Systems for Video Technology, v. 11, n. 3, p. 269–281, 2001. DOI: 10.1109/76.911155.
- (Deerwester et al., 1990) - DEERWESTER, S. ; DUMAIS, S. T. ; LANDAUER, T. ; FURNAS, G. ; HARSHMAN, R.. Indexing by latent semantic analysis. Journal of the Society for Information Science, v. 41, n. 6, p. 391–407, 1990.

- (Eisinger et al., 2005) - EISINGER, R. ; MANZATO, M. G. ; GOULARTE, R. . Devices descriptions for context-based content adaptation. In: Third Latin American Web Congress (LA-WEB 2005), 2005, Buenos Aires. Anais, Los Alamitos : IEEE Computer Society Press, 2005. v.1, p. 121-129. DOI: 10.1109/LAWEB.2005.14.
- (Eisinger & Goularte, 2005) - EISINGER, R. ; GOULARTE, R. . Towards context-based dynamic video adaptation. In: XI Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia 2005), 2005, Poços de Caldas - MG. Anais, Porto Alegre : SBC, 2005. v. 1, p. 1-3. DOI: 10.1145/1114223.1114242
- (Eisinger, 2007) – EISINGER, R. . Adaptação dinâmica de vídeo. Dissertação (Mestrado – Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2007. 115 p.
- (Eisinger et al., 2008) - EISINGER, R. ; ROMERO, R. A. F. ; GOULARTE, R. . Machine learning techniques applied to dynamic video adapting. In: The Seventh International Conference on Machine Learning and Applications (ICMLA'08), 2008, San Diego. Anais, Piscataway : IEEE Computer Society, 2008. v. 1, p. 819-822.
- (Fehn et al, 2006) - FEHN, C.; DE LA BARRÉ, R.; PASTOOR, S. Interactive 3-DTV – concepts and key technologies. Proceedings of the IEEE, v. 94, n. 3, p. 524-538, 2006. DOI: 10.1109/JPROC.2006.870688.
- (Gamma et al.; 1994) – GAMMA, E.; HELM, R.; JOHNSON, R.; VLISSIDES, J. M. Design Patterns: Elements of Reusable Object-Oriented Software. Addison-Wesley, 416 p., 1994. ISBN: 0201633612.
- (Gibson, 1998) - GIBSON, J. D. Digital Compression for Multimedia: Principles and Standards. Morgan Kaufmann, 476 p., 1998. ISBN: 15586036.
- (Goularte, 2003) – GOULARTE, R. Personalização e adaptação de conteúdo baseadas em contexto para TV Interativa. Tese (Doutorado – Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2003. 220 p.
- (Goularte et al., 2004) - GOULARTE, R.; CATTELAN, R. G.; CAMACHO-GUERRERO, J. A.; INÁCIO, JR., V. R.; PIMENTEL, M. G. C. Interactive multimedia annotations: enriching and extending content. In: ACM Symposium on Document Engineering (DocEng' 04), 2004, Milwaukee, Wisconsin, USA. Anais, New York : ACM Press, 2004, v. 1, p. 84–86. DOI: 10.1145/1030397.1030414.
- (Goularte et al., 2004b) - GOULARTE, R.; CAMACHO-GUERRERO, J. A.; INACIO JR., V. R.; CATTELAN, R. G.; PIMENTEL, M. G. C. M4Note: A Multimodal Tool for Multimedia Annotations. In: 10th Brazilian Symposium on Multimedia and the Web 2nd Latin American Web Congress (WebMedia & LA-Web 2004 Joint Conference), 2004, Ribeirão Preto, SP, Brasil. Anais, Washington, DC : IEEE Computer Society, 2004, p. 142–149. DOI: 10.1109/WEBMED.2004.1348161.
- (Goularte et al., 2006) - GOULARTE, R.; PIMENTEL, M. G. C.; MOREIRA, E. D. S. Context-Aware Support in Structured Documents for Interactive-TV. Multimedia Systems. Springer, v. 11, n. 4, p. 367-382, 2006. DOI: 10.1007/s00530-006-0014-4.
- (Goularte et al., 2007) - GOULARTE, R.; MELLO, R. F. ; DODONOV, E. ; YANG, L. T. . A Model to Estimate Transcoding Costs Applied in Live-Video Transmissions. In: International Conference on Intelligent Pervasive Computing (IPC-07), 2007, Jeju Island. Anais, Washington, DC : IEEE Computer Society, 2007. v. 1, p. 341-348. DOI=10.1109/IPC.2007.13.
- (Halsall, 2001) – HALSALL, F. Multimedia Communications; Applications, Networks, Protocols and Standrds. Addison-Wesley, Harlow, Inglaterra, 1033p., 2001. ISBN: 0-201-39818-4.
- (Hampapur et al., 1995) - HAMPAPUR, A. ; WEYMOUTH, T. E. ; JAIN, R. Feature Based Digital Video Indexing. In: Third IFIP WG2.6 Working Conference on Visual Database Systems, 1995, Londres. Anais, London, UK : Chapman & Hall, Ltd, 1995. v. 1, p. 115–141.

- (Hanjalic, 2004) - HANJALIC, A. Content-Based Analysis of Digital Video. Kluwer Academic Publishers, Dordrecht, The Netherlands, 208 p., 2004. ISBN: 1-4020-8114-6.
- (Hardman et al., 1993) – Hardman, L.; van Rossum, G.; Bulterman, D.C.A. Structured Multimedia Authoring. In: ACM International Conference on Multimedia, 1993, New York, USA. Anais, New York : ACM Press, 1993, v.1, p. 291-300. DOI: 10.1145/166266.168406.
- (Horn & Girod, 1997) - HORN, U.; GIROD, B. Scalable Video Transmission for the Internet. Computer Networks and ISDN Systems, v. 29, n. 15, p. 1833–1842, 1997. DOI: 10.1016/S0169-7552(97)00093-7.
- (Hua-Yong & Tingting, 2009) - HUA-YONG, L.; TINGTING, H. Content-based story segmentation of news video by multimodal analysis. In: 6th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD'09), 2009, Tianjin, China. Anais, Washington, DC : IEEE Computer Society, 2009. v. 7, p. 423-426. DOI: 10.1109/FSKD.2009.520.
- (Hutchison, 2008) - HUTCHISON, D. Introducing DLP 3-D TV. Texas Instruments, White Paper, p. 1-5, 2008. Endereço Web: <http://dlp.com/downloads/Introducing%20DLP%203D%20HDTV%20Whitepaper.pdf>. Acesso em: 22/07/2011.
- (ITU-T, 2003) – International Telecommunication Union. Recommendation G.114 – One-way transmission time. Endereço Web: <http://www.cs.columbia.edu/~andrea/new/documents/other/T-REC-G.114-200305.pdf>. Acesso em: 25/07/2011.
- (Kientz et al., 2007) - KIENTZ, J. A.; ARRIAGA, R. I.; CHETTY, M.; HAYES, G. R.; RICHARDSON, J.; PATEL, S. N.; ABOWD, G. D. Grow and know: understanding record-keeping needs for tracking the development of young children. In SIGCHI conference on Human factors in computing systems (CHI '07), 2007. Anais, New York : ACM Press, 2007. v. 1, pages 1351–1360, 2007. DOI: 10.1145/1240624.1240830.
- (Lazebnik et al., 2006) - LAZEBNIK S.; SCHMID, C.; PONCE, J. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR'06), 2006, New York, USA. Anais, Washington, DC : IEEE Computer Society, 2006. v. 1, p. 2169-2178. DOI: 10.1109/CVPR.2006.68.
- (Lipton, 1997) - LIPTON, L. Stereo-Vision Formats for Video and Computer Graphics. In: SPIE, 1997. San Jose, USA. Anais : SPIE, 1997. v. 3021, p. 239-246. DOI: 10.1117/12.274462.
- (Liu et al., 2007) - LIU, H.; HE, T.; ZHANG, H. Nbr: A content-based news video browsing and retrieval system. In: Technologies for E-Learning and Digital Entertainment. Lecture Notes in Computer Science, v. 4469, p. 793-800, 2007. DOI: 10.1007/978-3-540-73011-8_76.
- (Liu et al., 2009) - LIU, W.; YANG, G.; HUANG, X. Semantic features based news stories segmentation for news retrieval. In: International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR, 2009), 2009, Baoding - China. Anais, Washington, DC : IEEE Computer Society, 2009. v. 1, p. 258-265. DOI: 10.1109/ICWAPR.2009.5207491.
- (Lum & Lau, 2002) - LUM, W. Y. ; LAU, F. C. M. . A Context-Aware Decision Engine for Context Adaptation. IEEE Pervasive Computing, v. 1, n. 3, p. 41-49, 2002. DOI: 10.1109/MPRV.2002.1037721.
- (Magalhães & Pereira, 2004) - MAGALHÃES, J. ; PEREIRA, F. Using MPEG Standards for Multimedia Customization. Signal Processing: Image Communication, v. 19, n. 5, p. 437-456, 2004. DOI: <http://dx.doi.org/10.1016/j.image.2004.02.004>.
- (Manzato & Goularte, 2005) - MANZATO, M. G. ; GOULARTE, R. . Live video adaptation: a context-aware approach. In: XI Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia 2005), 2005, Poços de Caldas - MG. Anais, Porto Alegre : SBC, 2005. v. 1, p. 1-8. DOI: 10.1145/1114223.1114230

- (Manzato & Goularte, 2006) - MANZATO, M. G. ; GOULARTE, R. . Transcoding evaluation in live video adaptation. In: IV Latin-American Web Congress (LA-Web 2006), 2006, Cholula, Puebla. Anais, Los Alamitos : IEEE Computer Society Press, 2006. v. 1, p. 203-208.
- (Manzato, 2006) - MANZATO, M. G. . Adaptação de vídeo ao vivo apoiada em informações de contexto. Dissertação (Mestrado – Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2006. 100 p.
- (Manzato & Goularte, 2007) - MANZATO, M. G. ; GOULARTE, R. . Shot boundary detection based on intelligent systems. In: XIII Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia 2007), 2007, Gramado. Anais, Porto Alegre : SBC, 2007. v. 1., p. 190-197.
- (Manzato & Goularte, 2008) - MANZATO, M. G. ; GOULARTE, R. . Video News Classification for Automatic Content Personalization: A Genetic Algorithm Based Approach. In: XIV Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia 2008), 2008, Vila Velha. Anais, Porto Alegre : SBC, 2008. v. 1., p. 36-43. DOI: 10.1145/1666091.1666100.
- (Manzato et al., 2008) - MANZATO, M. G. ; JUNQUEIRA, D. C. ; GOULARTE, R. . Interactive News Documents for Digital Television. In: XIV Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia 2008), 2008, Vila Velha. Anais, Porto Alegre : SBC, 2008. v. 2. p. 1-4. DOI: 10.1145/1809980.1809982.
- (Manzato et al., 2009) - MANZATO, M. G. ; MACEDO, A. A. ; GOULARTE, R. . Evaluation of Video News Classification Techniques for Automatic Content Personalization. International Journal of Advanced Media and Communication, v. 3, n. 4, p. 383-403, 2009.
- (Manzato et al., 2009b) - MANZATO, M. G. ; COIMBRA, D. B. ; GOULARTE, R. . Multimedia content personalization based on peer-level annotation. In: European Interactive TV Conference (EuroITV 2009), 2009, Leuven, Bélgica. Anais, New York : ACM Press, 2009. v. 1., p. 57-66. DOI: 10.1145/1542084.1542096.
- (Manzato & Goularte, 2009) - MANZATO, M. G. ; GOULARTE, R. . Supporting multimedia recommender systems with peer-level annotations. In: XV Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia 2009), 2009, Fortaleza. Anais, Porto Alegre : SBC, 2009. v. 1., p. 202-209. DOI: 10.1145/1858477.1858503.
- (Manzato & Goularte, 2010) - MANZATO, M. G. ; GOULARTE, R. . Peersommender: A peer-level annotation based approach for multimedia recommendation. Journal of Information and Data Management, v. 1, n. 2, p. 277-292, 2010. ISSN 2178-7107.
- (Manzato et al., 2011) - MANZATO, M. G. ; COIMBRA, D. B. ; GOULARTE, R. . An enhanced content selection mechanism for personalization of video news programmes. Multimedia Systems Journal, v. 17, n. 1, p. 19-34, 2011. DOI: 10.1007/s00530-010-0204-y.
- (Manzato, 2011) - MANZATO, M. G. . Adaptação de vídeo ao vivo apoiada em informações de contexto. Tese (Doutorado – Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2011. 170 p.
- (Mello & Senger, 2006) – MELLO, R. F.; SENGER, L. J. Model for simulation of heterogeneous high-performance computing environments. In: 7th International Meeting High Performance Computing for Computational Science (VECPAR 2006), 2006, Rio de Janeiro. Anais, Heidelberg : Springer-Verlag, v. 1, p. 1-13, 2006. ISBN: 978-3-540-71350-0.
- (Mello et al., 2009) - MELLO, R. F. ; GOULARTE, R. ; DODONOV, E. ; YANG, L. T. ; PARK, J. H. ; KIM, T. On modeling and evaluating multicomputer transcoding architectures for live-video streams. Multimedia Tools and Applications, v. 43, n. 2, p. 109-129, 2009. DOI: 10.1007/s11042-009-0259-y.

- (Mendiburu, 2009) - MENDIBURU, B. 3D Movie Making Stereoscopic Digital Cinema from Script to Screen. Elsevier, Burlington, USA, 223 p., 2009. ISBN: 978-0-240-81137-6.
- (Merkle, 2007) - MERKLE, P.; SMOLIC, A.; MUELLER, K.; WIEGAND, T. Efficient Prediction Structures for Multiview Video Coding. *IEEE Transactions on Circuits and Systems for Video Technology*, v. 17, n. 11, p. 1461-1473, 2007. DOI: 10.1109/TCSVT.2007.903665.
- (Misra et al., 2010) - MISRA, H.; HOPFGARTNER, F.; GOYAL, A.; PUNITHA, P.; JOSE, J. Tv news story segmentation based on semantic coherence and content similarity. In BOLL, S.; TIAN, Q.; ZHANG, L.; ZHANG, Z.; CHEN, Y. P. (eds). *Advances in Multimedia Modeling. Lecture Notes in Computer Science*, v. 5916, p. 347-357, 2010. Springer, Berlin / Heidelberg, 806 p. ISBN: 3-642-11300-1.
- (Morris & Smith-Chaigneau, 2005) – MORRIS, S.; SMITH-CHAIGNEAU, A.. *Interactive TV Standards: A Guide to MHP, OCAP, and JavaTV*. Focal Press, 608 p., 2005. ISBN: 978-0240806662.
- (Nascimento et al., 2010) - NASCIMENTO, G. A.; MANZATO, M. G.; GOULARTE, R. Extração de Quadros-Chave como subsídio para personalização em vídeos digitais. In: *Workshop de Iniciação Científica - XVI Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia 2010)*, 2010, Belo Horizonte. Anais, Porto Alegre : SBC, 2010, v. 2. p. 105-107.
- (Ohm, 2005) - OHM, J. R. *Advances in Scalable Video Coding. Proceedings of the IEEE*, v. 93, n. 1, p. 42–56, 2005. DOI: 10.1109/JPROC.2004.839611.
- (Oliveira et al., 2009) - OLIVEIRA, L. S.; MARTINS, D. S.; GOULARTE, R.; PIMENTEL, M. G. C. EducaTV: uma arquitetura para acesso a conteúdos educacionais via TVDi. In: *XV Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia'09)*, 2009, Fortaleza. Anais, Porto Alegre : SBC, 2009. v. 2, p. 63-66. DOI: 10.1145/1858477.1858525.
- (Pedrosa et al., 2009) - PEDROSA, D. C.; VEGA-OLIVEROS, D. A.; PIMENTEL, M. G. C.; GOULARTE, R. Uma Aplicação NCL/Lua para Navegação em Vídeo via Quadros Recentes. In: *Workshop de Ferramentas e Aplicações - XV Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia'09)*, 2009, Fortaleza. Anais, Porto Alegre : SBC, 2009. v. 2, p. 136-138.
- (Pereira & Ebrahimi, 2002) - PEREIRA, F. C. N.; EBRAHIMI, T. *The MPEG-4 Book*. Prentice-Hall, Upper Saddle River, 849 p., 2002. ISBN: 0-13-061621-4.
- (Piccolo & Baranauskas, 2006) - PICCOLO, L. S. and BARANAUSKAS, M. C. 2006. Desafios de design para a TV digital interativa. In *VII Brazilian Symposium on Human Factors in Computing Systems (IHC'06)*, 2006. Anais, New York : ACM Press, 2006. vol. 323, p. 1-10. DOI: 10.1145/1298023.1298025.
- (Pimentel et al., 2007) - PIMENTEL, M. G. C.; GOULARTE, R.; CATTELAN, R. G.; SANTOS, F. S.; TEIXEIRA, C. Enhancing multimodal annotations with pen-based information. In: *Workshop on New Techniques for Consuming, Managing, and Manipulating Interactive Digital Media at Home, 2007*, Taichung. *IEEE International Symposium on Multimedia*, 2007. v. 1, p. 207-213. DOI: 10.1109/ISM.Workshops.2007.43.
- (Pimentel et al., 2008) - PIMENTEL, M. G. C.; GOULARTE, R.; CATTELAN, R. G.; SANTOS, F. S.; TEIXEIRA, C. Ubiquitous Interactive Video Editing via Multimodal Annotations. In: *European Interactive TV Conference (EuroITV 2008)*, 2008, Salzburg. Anais, Hidelberg : Springer-Verlag, (LNCS) v. 5066, p. 72-81, 2008. DOI: 10.1007/978-3-540-69478-6_8.
- (Richardson, 2002) – Richardson, I. E. G. *Video Codec Design*. Wiley, West Sussex, Inglaterra, 303p., 2002. ISBN:0-471-48553-5.
- (Richardson, 2003) – Richardson, I. E. G. *H.264 and MPEG-4 Video Compression*. Wiley, West Sussex, Inglaterra, 281p., 2003. ISBN:0-470-84837-5.

- (SBC, 2006) - Sociedade Brasileira de Computação. Grandes Desafios da Pesquisa em Computação no Brasil 2006-2016. Relatório Técnico, 2006. 26 p. endereço Web: http://www.sbc.org.br/index.php?option=com_jdownloads&Itemid=195&task=finish&cid=11&catid=50. Acesso em: 25/07/2011.
- (Shevade & Sundaram, 2003) – SHEVADE, B.; SUNDARAM, H. Vidya: an experiential annotation system. In: ACM SIGMM Workshop on Experiential Telepresence, 2003, Berkeley – CA, USA. Anais, New York : ACM Press, v. 1, p. 91–98, 2003. DOI: 10.1145/982484.982502.
- (Siegel et al., 1994) - SIEGEL, M. W.; GUNATILAKE, P.; SETHURAMAN, S.; JORDAN, A. G. Compression of stereo image pairs and streams. *Stereoscopic Displays and Virtual Reality Systems*, SPIE, v. 2177, p. 258-268, 1994. DOI: 10.1117/12.173899.
- (Singh, 2001) - SINGH, S. O Livro dos Códigos. Editora Record, Rio de Janeiro, 512 p., 2001. ISBN: 8501055980.
- (Sivic & Zisserman, 2006) - SIVIC, J.; ZISSERMAN, A. Video Google: Efficient Visual Search of Videos. In: *Toward Category-Level Object Recognition. Lecture Notes in Computer Science*, v. 4170, p. 127-144, 2006. Springer, Berlin / Heidelberg, 618 p., 2006. ISBN: 3-540-68794-7. DOI: 10.1007/11957959_7.
- (Smeulders et al., 2000) - Smeulders, A.; Worring, M.; Santini, S.; Gupta, A.; Jain, R.. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 22, n. 12, p. 1349-1380, 2000. DOI: 10.1109/34.895972.
- (Smolic et al, 2009) - SMOLIC, A.; MUELLER, K.; MERKLE, P.; KAUFF, P.; WIEGAND, T. An Overview of Available and Emerging 3D Video Formats and Depth Enhanced Stereo as Efficient Generic Solution. In: *27th conference on Picture Coding Symposium, 2009, Chicago, USA*. Anais, Los Alamitos : IEEE Computer Society Press, v. 1 , p. 1-4, 2009. DOI: 10.1109/PCS.2009.5167358.
- (Snoek & Worring, 2005) - SNOEK, C. G. ; WORRING, M. Multimodal Video Indexing: A Review of the State-of-the-art. *Multimedia Tools and Applications*, v. 25, n. 1, p. 5–35, 2005. DOI: 10.1023/B:MTAP.0000046380.27575.a5.
- (Sony Corporation, 2010) - Sony Corporation. Sony's 3D TV. Endereço Web: <http://www.sony.net/united/3D/#technology/3dtv/>. Acesso em 22/07/2011.
- (Souza Filho et al., 2007) – SOUZA FILHO, G. L.; LEITE, L.E.C.; BATISTA, C. E. C. F. Ginga-J: the procedural middleware for the brazilian digital tv system. *Journal of the Brazilian Computer Society*, v. 13, n. 1, p. 47–56, 2007.
- (Srivastava, 2002) - SRIVASTAVA, H. O. *Interactive TV Technology and Markets*. Artech House, Norwood, 407 p., 2002. ISBN: 1-58053-321-3.
- (Stereographics, 1997) - StereoGraphics Corporation. *Stereographics® Developers' Handbook: Background on Creating Images for CrystalEyes and SimulEyes*, 1997. Endereço Web: <http://www.cs.unc.edu/Research/stc/FAQs/Stereo/stereo-handbook.pdf>. Acesso em: 22/07/2011.
- (Suppia, 2007) - Suppia, A. O Monstro Brasileiro Revive em 3D. *Revista Cultura e Ciência*, v. 59, n. 2, p. 57-59, 2007. Endereço Web: <http://cienciaecultura.bvs.br/pdf/cic/v59n2/a24v59n2.pdf>. Acesso em: 22/07/2011.
- (Talarico Neto, 2011) – Talarico Neto, A. Uma abordagem para projeto de aplicações com interação multimodal da Web. Tese (Doutorado – Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2011. 199 p.

- (Tam & Zhang, 2006) - TAM, W. J.; ZHANG, L. 3D-TV Content Generation: 2D-to-3D Conversion. IEEE International Conference on Multimedia and Expo, 2006, Toronto, Canadá. Anais, Los Alamitos : IEEE Computer Society Press, 2006. v. 1, p. 1869-1872. DOI: 10.1109/ICME.2006.262919.
- (Teevan et al., 2005) – TEEVAN, J. ; DUMAIS, S. T. ; HORVITZ, E. . Personalizing search via automated analysis of interests and activities. In: 28th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '05), 2005, Salvador. Anais, New York : ACM Press, 2006. v.1, p. 449-456. DOI: 10.1145/1076034.1076111.
- (Valle & Cord) - VALLE, E.; CORD, M. Advanced Techniques in CBIR: Local Descriptors, Visual Dictionaries and Bags of Features. In: Tutorials of the XXII Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAP'09 Tutorials), 2009, Rio de Janeiro – RJ, Brasil. Anais, Los Alamitos : IEEE Computer Society Press, 2009. v. 1, p. 72-78. DOI: 10.1109/SIBGRAPITutorials.2009.14.
- (Vaughan, 2006) - VAUGHAN, T. Multimedia: Making It Work. 7a. edição. Osborne/McGraw-Hill, Berkeley, 514 p. ISBN: 978-0072264517.
- (Vega-Oliveros, 2009) - VEGA-OLIVEROS, D. A.; PEDROSA, D. C.; PIMENTEL, M. G. C.; GOULARTE, R. Navegação em Vídeo via Quadros Recentes. In: XV Simpósio Brasileiro de Sistemas Multimídia e Web, 2009, Fortaleza. Anais, Porto Alegre : SBC, 2009. v. 2, p. 15-18. DOI: 10.1145/1858477.1858513.
- (Vetro et al., 2003) - VETRO, A.; CHRISTOPOULOS, C.; SUN, H. Video Transcoding Architectures and Techniques: An Overview. IEEE Signal Processing Magazine, v. 20, n. 2, p. 18–29, 2003. DOI: 10.1109/MSP.2003.1184336.
- (Wang et al., 2008) - WANG, J.; DUAN, L.; LIU, Q.; LU, H.; JIN, J. A multimodal scheme for program segmentation and representation in broadcast video streams. IEEE Transactions on Multimedia, v. 10, n. 3, p. 393-408, 2008. DOI: 10.1109/TMM.2008.917362.
- (Xie et al., 2007) - XIE, L.; NATSEV, A.; TESIC, J. Dynamic multimodal fusion in video search. In: IEEE International Conference on Multimedia and Expo, 2007, Beijing, China. Anais, Los Alamitos : IEEE Computer Society Press, 2007. v. 1, p. 1499 -1502. DOI: 10.1109/ICME.2007.4284946.
- (Yu et al., 2007) - YU, H.; SU, B.; LU, H.; XUE, X. News video retrieval by learning multimodal semantic information. In: 9th International Conference on Advances in Visual Information Systems (VISUAL'07), 2007, Shanghai, China. Anais, Heidelberg : Springer-Verlag, 2007, v. 1, p. 403-414.
- (Zingarelli et al., 2011) - ZINGARELLI, M. R. U.; ANDRADE, L. A.; GOULARTE, R. Reversing Anaglyph Videos Into Stereo Pairs. In: XVII Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia 2011), 2011, Florianópolis - SC. (Aceito para publicação como artigo completo).