

Estudo Comparativo de Codificação Tridimensional para o SBTVD

Adenilson J. A. Tomé*
 Universidade Federal do ABC
 Rua Santa Adélia, 166
 Santo Andre, SP, 09210-580,
 Brasil
 adenilson.tome@ufabc-
 .edu.br

Celso S. Kurashima†
 Universidade Federal do ABC
 Rua Santa Adélia, 166
 Santo Andre, SP, 09210-580,
 Brasil
 celso.kurashima@ufabc-
 .edu.br

Mario Minami‡
 Universidade Federal do ABC
 Rua Santa Adélia, 166
 Santo Andre, SP, 09210-580,
 Brasil
 mario.minami@ufabc-
 .edu.br

ABSTRACT

The Brazilian Digital TV System (SBTVD) started signal transmission in São Paulo on 2 December 2007. Currently, all the Brazilian main cities already have DTV channels available. The adoption of the H.264/MPEG-4 AVC standard provides advanced technologies for encoding digital video signals with high definition video. This standard, also allows transmission of three-dimensional video, but this is not considered by SBTVD yet. This research aims to study the encoding of three-dimensional video content with H.264/MPEG-4 AVC standard, and the insertion of these encoded signals in the SBTVD broadcasting stream. Was investigated techniques for capturing video from two and more cameras, encoding three-dimensional video streams, and transmission of 3D video in the Brazilian Digital TV system. The multiple views were coded with the side-by-side technique as well as with multiple-view coding. The videos were transmitted, decoded, and analyzed in a laboratory simulation environment. The objective quality of each video was measured within a wide range of transmission bit rate, and we concluded that good quality 3D video is possible to be allocated in the transmission stream.

Categories and Subject Descriptors

I.3.7 [Computing Methodologies]: Three-Dimensional Graphics and Realism; I.4.1 [Computing Methodologies]: Digitization and Image Capture—*Camera calibration*

General Terms

Experimentation, Measurement, Theory and Verification

*Programa de Pós Graduação em Engenharia da Informação (UFABC). Dissertação de Mestrado. Data de início: FEV.2010. Data da defesa: JUL.2012.

†Orientador.

‡Coorientador.

Keywords

Video encoding, 3DTV, SBTVD, Digital TV

1. INTRODUÇÃO

A implantação da TV Digital no Brasil iniciou oficialmente em 2007 e mudanças significativas na televisão brasileira surgem com a tecnologia digital. A transmissão de imagens em alta definição, áudio com qualidade superior ao de CD e a possibilidade de interatividade com os usuários são uma pequena parte das vantagens que o sistema pode propiciar.

A *International Telecommunication Union* (ITU) definiu um roteiro para a implantação da TV 3D. Esse roteiro é composto de três gerações, sendo que a primeira geração já está sendo utilizada por algumas emissoras, a segunda geração são estudos para um futuro próximo e a terceira prevê um tempo de desenvolvimento de 15 a 20 anos [7].

Foi realizado um estudo sobre captura de fluxos de vídeo a partir de N câmeras *firewire*, de modo sincronizado, visando permitir a codificação de vídeos tridimensionais no formato *side by side* (lado a lado) e *Multiview Video Coding* (MVC), Codificação de Vídeo Multi-vistas, apoiados pela norma H.264/MPEG-4 AVC. A camada de transporte foi considerada para a inserção dos sinais de vídeo tridimensional dentro do feixe de transmissão digital do SBTVD, o conteúdo 3D produzido foi encapsulado em fluxos de transportes e transmitido no SBTVD. Os conteúdos transmitidos tiveram sua qualidade medida através da métrica PSNR e analisada posteriormente.

1.1 Objetivo

O objetivo geral da dissertação é investigar e produzir contribuições sobre a viabilidade da utilização de conteúdo tridimensional no SBTVD. Atualmente as normas do SBTVD não preveem a transmissão de conteúdo tridimensional, devido a isso foram realizados estudos sobre a codificação de vídeos tridimensionais e inserção desses sinais de vídeo dentro do fluxo de transporte do SBTVD, efetuando a multiplexação de fluxos de bits válidos.

1.2 Contribuições Esperadas

A implantação do SBTVD no Brasil que deve abranger todo território nacional até 2016 extinguindo a transmissão do sistema analógico de TV país. A produção cada vez maior de aparelhos de TV com suporte à tecnologia 3D assim como

de filmadoras e câmeras digitais é motivo de reflexão quanto à utilização de técnicas eficientes para a disponibilização de conteúdo 3D de qualidade de modo gratuito ao usuário final, possibilitando abranger todas as classes sociais. As tecnologias a serem adotadas devem garantir além de robustez e qualidade que, sejam o estado na arte em 3D e que permitam possíveis expansões futuras. O estudo de técnicas de codificação de vídeo 3D, como a *side by side* e a MVC, voltadas para a transmissão no sistema aberto de TV digital brasileiro é estudada a fim de gerar resultados que visam propiciar fundamentos e conclusões sobre quais técnicas de codificação 3D que podem ser utilizadas para transmissão no SBTVD. As transmissões consideraram a camada de transporte e a largura de banda disponível por canal, podendo assim ter uma visão sobre as transmissões para dispositivos móveis e, TVs e receptores com resolução HD e Full HD.

1.3 Metodologia Adotada

Para dar suporte à esta dissertação e atingir os objetivos propostos a seguinte metodologia foi seguida: (i) Revisão bibliográfica abordando os assuntos relevantes para a elaboração dessa dissertação como técnicas de captura de imagens com múltiplas câmeras sincronizadas, composição de conteúdo 3D estereoscópico, sistemas de TV Digital, Multiplexação de dados e Serviços de Informação aplicados à TV Digital; (ii) Desenvolvimento de um aplicativo de captura de vídeo sincronizado a partir de 2, 4 e 8 câmeras *firewire*; (iii) Composição e codificação de conteúdo 3D estereoscópico e com múltiplas vistas com diversas taxas de bits; (iv) Encapsulamento dos fluxos elementares de vídeo 3D dentro de um fluxo de transporte válido para o SBTVD; (v) Transmissão dos fluxos de transporte no SBTVD; (vi) Análise de qualidade dos fluxos elementares de vídeo após a transmissão com base na resolução e taxa de bits.

As etapas descritas acima compõem um ambiente *offline*, necessário para análises individuais de cada etapa de um ambiente de TV 3D. Cada etapa será melhor detalhada a seguir. Todas as codificações foram realizadas com CPU Intel® Core™ i7-960 3.2Ghz, 6GB de memória RAM e GPU Nvidia GTS 250 de 1GB.

1.4 Estágio Atual do Trabalho

A dissertação encontra-se em fase final de análise final dos resultados e composição dos gráficos e tabelas.

2. CONTEXTO TEÓRICO

2.1 Captura de vídeo

Todo sistema multimídia envolve diversos fatores que devem ser considerados para o tratamento e manipulação dos dados e um sistema de captura de vídeos não é diferente. No processo de aquisição dos fluxos de vídeo os dados devem ser tratados de forma que possibilitem sua manipulação para construção de conteúdos tridimensionais [10]. Uma das principais distorções é a radial, que devido ao seu formato, provoca uma distorção da localização dos *pixels* próximos das bordas das imagens [3]. O processo de calibração de câmeras fornece um modelo da geometria da câmera e um modelo de distorção da lente. Estes dois modelos informativos definem os parâmetros intrínsecos da câmera. Através desses parâmetros é possível efetuar a correção da distorção causada pelas lentes.

A composição de conteúdo tridimensional requer que a captura dos fluxos de imagens e vídeos seja efetuada de modo sincronizado, para que as diversas imagens das câmeras capturem o mesmo cenário ao mesmo tempo. Para a aquisição sincronizada de conteúdo faz-se necessário uma arquitetura de hardware robusta combinada com alguns recursos de softwares que possibilitem esse funcionamento.

2.2 Embasamento Teórico Tridimensional

Ao olhar ao nosso redor os dois olhos humanos nos dão dois pontos de vista simultâneos dos objetos em uma cena. Por um processo chamado de estereopsia, o cérebro funde essas e gera automaticamente informações de profundidade, e é isso que nos permite executar operações do nosso dia a dia com segurança, por exemplo, dirigir ou praticar esportes [4]. A visão 3D se destina a diversos métodos de visualização, onde os *displays* 3D se dividem em duas categorias: (i) estereoscópicos e (ii) autoestereoscópicos. A visualização estereoscópica requer dois pontos de vista enquanto os autoestereoscópicos necessitam de oito ou mais pontos de vista para fornecer a funcionalidade de ponto de vista livre, também conhecido como *Free-View Point* [11].

A estereoscopia é a técnica utilizada para a criação de um par de imagens estereó planares. Plano-estereoscópico é o termo exato para descrever *displays* 3D que conseguem reproduzir um efeito de profundidade binocular, proporcionando ao espectador uma perspectiva com imagens ligeiramente diferentes em uma tela planar comum [8]. Na técnica conhecida como *side by side* cada quadro é composto por duas imagens, uma para cada olho. Cada quadro é reduzido horizontalmente, assim, cada imagem terá metade da resolução na horizontal, sem alterar a resolução vertical. Ao receber a imagem nesse formato, o *display* divide o quadro e redimensiona cada metade para o formato original e os exibe de modo sequencial conforme ilustrado na Figura 1.

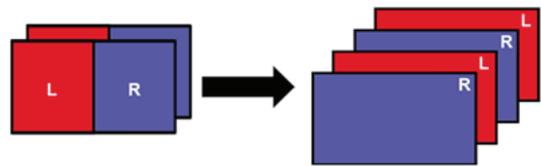


Figure 1: Sequência de quadros *side by side* transformados em quadros sequenciais.

O H.264/ AVC é um padrão de codificação de vídeo (considerado atualmente o estado da arte), desenvolvido em conjunto pela ITU-T *Video Coding Experts Group* (VCEG) e a ISO/IEC *Moving Pictures Experts Group* (MPEG). O H.264/ AVC foi desenvolvido para fornecer compressão eficiente e alta confiabilidade na transmissão de vídeo [9]. O desenvolvimento mais recente dos grupos VCEG e MPEG é o projeto *Joint Video Team* (JVT), as atividades desse projeto concentram-se na codificação com múltiplas vistas, também conhecido como *Joint Multiview Video Coding* (JMVC). O MVC é um codificador que possibilita representar uma cena tridimensional a partir da captura e codificação de duas ou mais câmeras utilizando ângulos ligeiramente diferentes da mesma cena [8].

2.3 SBTVD

O padrão base utilizado no SBTVD é o *Integrated Services Digital Broadcasting Terrestrial* (ISDB-T), desenvolvido pelo Japão. Assim, o sistema adotado no Brasil é o ISDB-TB, a sigla "B" foi adicionada para identificar que o sistema é o do Brasil [5]. Diversas atualizações foram acrescentadas no SBTVD, para codificação de vídeo foi adotado o padrão MPEG-4 parte 10, também conhecido como H.264, e para codificação de áudio o HE-AAC v2. O padrão MPEG-4 foi desenvolvido pela ITU e pelo *Video Coding Experts Group* (VCEG) em conjunto com a ISO/IEC MPEG. A ferramenta de compressão de vídeo do sistema de televisão digital terrestre brasileiro deve obrigatoriamente estar de acordo com a ITU-T *Recommendation H.264* [2].

No sistema de transmissão foi utilizada a mesma técnica de modulação do ISDB-T, o *Orthogonal Frequency Division Multiplexing* (OFDM). O canal de 6 MHz é dividido em 14 segmentos de 428,57 kHz, dos quais 13 segmentos podem ser utilizados na transmissão de até três serviços distintos e simultâneos. A largura de banda de cada segmento do ISDB-TB varia de acordo com o tipo de modulação e das configurações dos parâmetros utilizados. Cada segmento pode ter taxa máxima entre 280 kbps e 1787 kbps. Na soma dos 13 segmentos a largura de banda máxima fica entre 3,6 Mbps e 23,2 Mbps [1]. O serviço de multiplexação do SBTVD é padronizado pelas normas ABNT NBR: 15602-1, 15602-2 e 15602-3. A principal referência para estas normas é a norma de camada de transporte, criada em conjunto pela ITU e pela ISO/IEC, respectivamente com os nomes de ITU-T H.222 e ISO/IEC 13818-1 [6].

3. RESULTADOS

Os resultados obtidos foram divididos em várias partes que juntos compõem o cenário desenvolvido para atingir os objetivos da dissertação em questão. A etapa de captura visou a captura dos fluxos de vídeos utilizando N câmeras *firewire*. A aplicação desenvolvida, designada *MultipleSync-Capture* é capaz de capturar fluxos de vídeos a partir de 2, 4 e 8 câmeras. O aplicativo é capaz de efetuar a sincronização das câmeras pelo barramento IEEE 1394a, efetuar a calibração das câmeras extraíndo os parâmetros intrínsecos e de distorção e assim efetuar a correção de distorção das lentes. Por fim o aplicativo realiza a gravação de um vídeo no formato AVI sem compressão.

3.1 Codificação de vídeo 3D

As codificações de todos os fluxos de vídeo seguiram a norma MPEG-4 Parte 10, também conhecida como H.264 AVC. Para realizar as medições os fluxos foram codificados utilizando a mesma resolução, tendo como entrada duas vistas de 160x240 *pixels*.

Os fluxos de vídeo foram codificados com várias taxas de bits para o MVC e o *side by side*. As codificações dos fluxos elementares MVC foram efetuadas utilizando o software JMVC, software de referência para codificação MVC desenvolvido pelo JVT. As codificações MVC com duas vistas foram realizadas utilizando os parâmetros de entrada sugeridos no manual do JMVC, com exceção do comprimento do GOP, resolução e taxa de quadros dos vídeos, que devem ser exatamente informados de acordo com a codificação desejada. Para possibilitar fluxos com varias taxas de bits para o mesmo fluxo de vídeo o parâmetro de quantiza-

ção do codificador JMVC foi alterado, assim cada cenário foi codificado variando o QP. Os valores de QP utilizados para os cenários foram: 2, 6, 9, 12, 15, 18, 21, 24, 30, 36 e 42. Cada valor gera uma taxa de bits e uma qualidade diferente no vídeo codificado. O processo de decodificação é mais simples que o processo de codificação bastando apenas informar ao decodificador o *bitstream* a ser decodificado e o número de vistas. As codificações MVC demandam elevado tempo, o que justifica a utilização de baixas resoluções nas codificações.

A composição dos fluxos elementares *side by side* foi realizada através dos mesmos fluxos elementares de vídeo utilizados na codificação MVC, ou seja, o codificador recebeu como entrada dois fluxos elementares de vídeo com resolução de 160x240 *pixels*, como no *side by side* os quadros são colocados lado a lado e somente são rescalados na decodificação a resolução final de saída do codificador é o um fluxo elementar com resolução de 320x240 *pixels*. Um fator importante a ser considerado para a realização da codificação *side by side* é que o codificador JMVC não gera fluxos elementares de vídeo com taxa de bits constantes, ou seja, não é possível determinar os valores exatos. Os valores das taxas de bits são gerados pela variação do parâmetro de quantização do codificador JMVC. Devido a isso, as taxas de bits utilizadas na codificação *side by side* tiveram como ponto de partida as taxas geradas pelo codificador JMVC para cada fluxo elementar codificado. Foram gerados fluxos de vídeo elementares utilizando as mesmas taxas de bits geradas pelo codificador JMVC, sendo assim, para cada fluxo elementar de vídeo MVC foi codificado um fluxo elementar de vídeo *side by side* correspondente, com os mesmos parâmetros utilizados na codificação MVC.

A transmissão de um conteúdo de vídeo 3D, ou seja, um fluxo elementar de vídeo 3D dentro do SBTVD deve ser feito seguindo todas as normas técnicas de tal forma que o ambiente de testes se aproxime o máximo possível de uma transmissão realizada por uma emissora de televisão. Os fluxos elementares devem ser encapsulados em fluxos de transporte e multiplexados seguindo a norma internacional ISO/IEC 13818-1 (ITU-T *Recommendation H.222*) e as normas ABNT NBR 15603-1, 15603-2 e 15603-3. Visando simular um ambiente real utilizado pelas emissoras de TV digital, todas as normas foram seguidas e as tabelas obrigatórias de *Serviço de Informação* (SI) e *Programação de Informação e Sistema* (PSI) foram inseridas em todos os fluxos de transporte gerados.

3.2 Medidas PSNR

Para os cálculos de qualidade foi utilizada a métrica *Peak Signal-to-Noise Ratio* (PSNR), efetuada com o apoio do software *MSU Video Quality Measurement Tool*¹, sempre analisando o componente de luminância. Para realizar os cálculos PSNR foram selecionados 8 seqüências de vídeo com conteúdo capturados a partir de duas câmeras (estéreo). Cada seqüência possui características diferentes que visam representar diversos tipos de cenários possíveis. Para cada seqüência de vídeo foram realizadas 11 codificações MVC variando o parâmetro de quantização. Cada uma dessas 11 codificações geraram uma taxa de bits diferente. O valor medidas

¹http://www.compression.ru/index_en.htm

PSNR foram efetuadas comparando o fluxo elementar codificado com o seu respectivo fluxo elementar original. O valor das medidas de qualidade foram calculadas de maneira independente para cada vista para os fluxos MVC e depois calculada a média aritmética. O valor PSNR do fluxo elementar de vídeo *side by side* foi calculado sobre os quadros contendo as duas vistas. As análises visam comparar somente os métodos MVC e *side by side*, o que pode ser avaliado através de testes objetivos. Os gráficos das seqüências com as medidas PSNR das seqüências *side by side* são demonstradas na Figura 2 e a MVC na Figura 3.

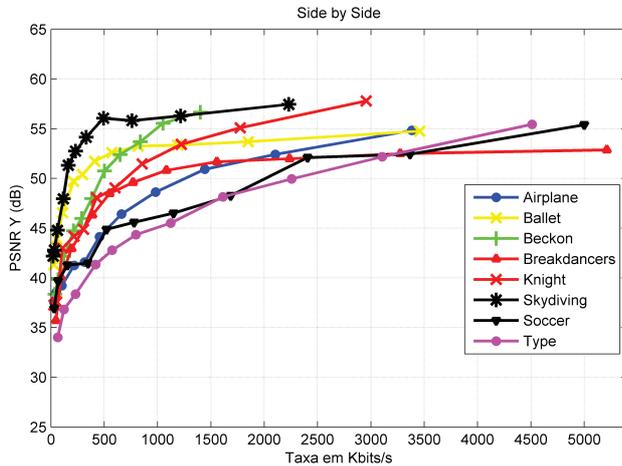


Figure 2: PSNR x taxa de bits das seqüências *Side by Side*.

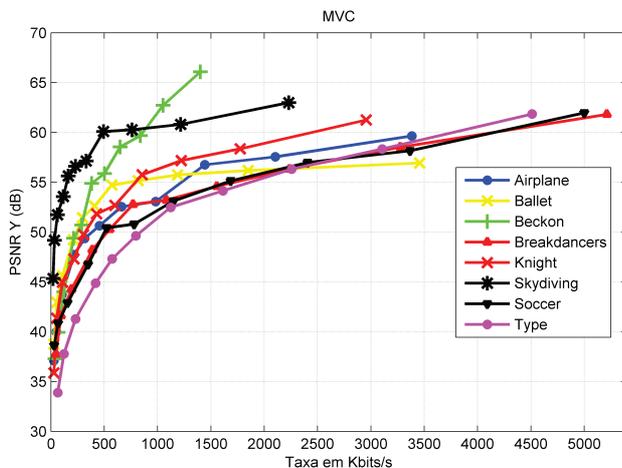


Figure 3: PSNR x taxa de bits das seqüências MVC.

4. CONCLUSÕES

A codificação e análise das medidas PSNR dos fluxos de vídeo estereoscópicos *side by side* e MVC permitiram realizar uma comparação entre os dois codificadores. Essa comparação é importante uma vez que o conteúdo *side by side* é utilizado por algumas emissoras na Europa, nos Estados Unidos e por uma emissora aberta no estado de São Paulo/BR. Os resultados das medidas PSNR dos fluxos de vídeo mostram que a codificação *side by side* conseguiu ser melhor que a MVC somente em taxas de bits baixas. Em

cinco dos oito fluxos 3D foi possível visualizar essa diferença nas taxas baixas, porém, cada fluxo foi codificado com onze taxas de bits diferentes e em quatro seqüências a codificação *side by side* superou a MVC na taxa mais baixa. Dessas quatro seqüências o melhor desempenho da codificação *side by side* foi apenas 3,77%, quando a taxa de bits foi de 28 Kbps.

Em 80 das 88 comparações efetuadas a codificação *side by side* obteve resultados piores que a MVC. O ganho de qualidade das seqüências nem sempre segue um padrão linear a medida que a taxa de bits é acrescida. É possível ver que em determinadas seqüências, como a *Airplane*, a *Knight* e a *Skydiving* a maior diferença de qualidade entre as codificações se situa em taxas de bits intermediárias, variando de acordo com a seqüência: *Airplane* (316 Kbps), *Knight* (305 Kbps), *Skydiving* (61 Kbps).

5. AGRADECIMENTOS

À CAPES, pelos recursos e bolsa cedidos no projeto RHTVD de No. 231/2008, através do programa de formação de recursos humanos em televisão digital.

6. REFERENCIAS

- [1] Associação Brasileira de Normas Técnicas, Rio de Janeiro. *NBR 15601: Televisão digital terrestre - Sistema de transmissão*, nov. 2007.
- [2] Associação Brasileira de Normas Técnicas, Rio de Janeiro. *NBR 15602-1: Televisão digital terrestre - Codificação de vídeo, áudio e multiplexação*, nov. 2007.
- [3] A. Bradski, G.; Kaehler. *Learning OpenCV*. O'Reilly Media, Inc, first edition edition, 2008.
- [4] C. Connolly. Stereoscopic imaging. In *Sensor Review*, volume Vol. 26 Iss: 4, pages 266–271. Emerald Group Publishing Limited, 2006.
- [5] DTV. História da tv digital no brasil. <http://www.dtv.org.br/index.php/informacoes-tecnicas/historia-da-tv-digital-no-brasil>, Acesso em jun. 2012.
- [6] Institute of Electrical and Electronics Engineers. *ITU-T Recommendation H.222*, Sep. 1995.
- [7] ITU. 3d tv moves into focus. http://www.itu.int/newsroom/press_releases/2010/-01.html, Acesso em jun. 2012.
- [8] D. Minoli. *3DTV content capture, encoding and transmission: building the transport infrastructure for commercial services*. John Wiley Sons, Inc., September 2010.
- [9] O. Nemeié, S. Rimac-Drlje, and M. Vranjes. Multiview video coding extension of the h. 264/avc standard. *52nd International Symposium ELMAR*, Sep. 2010.
- [10] O. B. Neto. Captura e modelagem de mãos com rastreamento de movimentos. Master's thesis, Universidade Federal do ABC, 2010.
- [11] C. Peng, V. Ollikainen, and S. Tomperi. Stereoscopic 3-d tv webcast. In *Proceedings of the 9th international interactive conference on Interactive television*, EuroITV '11, pages 177–180, New York, NY, USA, 2011. ACM.