

# Percepção Estereoscópica Anaglífica em Vídeos Digitais Comprimidos com Perda

Leonardo Antonio de Andrade  
Universidade Federal de São Carlos,  
Departamento de Artes e Comunicação  
Rod. Washington Luiz, Km 235  
São Carlos, São Paulo, Brasil  
landrade@ufscar.br

Rudinei Goularte  
Universidade de São Paulo,  
Instituto de Ciências Matemáticas e de Computação  
Av. Trabalhador São Carlense, 400  
São Carlos, São Paulo, Brasil  
rudinei@icmc.usp.br

## RESUMO

A estereoscopia anaglífica é um método de baixo custo de reprodução, contrastando com técnicas estereoscópicas baseadas em luz polarizada ou monitores auto-estereoscópicos. A estereoscopia anaglífica necessita apenas um monitor convencional e um par de óculos com filtros de cores para que o usuário perceba a profundidade de um objeto ou cena. Esta técnica que vem sendo explorada desde o final do século XIX, apresenta-se como uma alternativa viável para visualização de vídeos, pois necessita apenas guardar suas informações em um vídeo convencional com três canais de cores. Contudo, a codificação de vídeos estereoscópicos com as técnicas de codificação padrões, faz com ocorram perdas de dados, e estas perdas dificultam a percepção da profundidade nos vídeos codificados. Este trabalho apresenta uma análise de como as transformadas DCT e DWT com seus respectivos processos de quantização afetam a visualização anaglífica. Tal análise utilizou medidas objetivas e subjetivas de qualidade sobre uma base de vídeos estereoscópicos anaglíficos especialmente desenvolvida para esse fim, seguindo critérios padronizados. Os resultados indicaram que existe um limiar para as diferenças dos canais de cor RGB, provenientes das perdas ocorridas na codificação, não interferiram na percepção de vídeos estereoscópicos anaglíficos, e que a DWT é mais indicada para sua codificação desse tipo de vídeo.

## Palavras-Chave

Vídeo Digital, Estereoscopia, Visão Anaglífica Estereoscópica, DWT, DCT.

## 1. INTRODUÇÃO

Os registros e o conteúdo ficcional produzidos desde a invenção da fotografia pelo homem, são majoritariamente bidimensionais, e não possuem uma importante informação de visualização, a profundidade.

A capacidade do homem em interpretar pares de ilustrações ou fotos de uma mesma cena, visualizados por ângulos ligeiramente diferentes, é chamada de estereoscopia, e teve seus fundamentos lançados no século XIX. Desde essa época, passando por livros com fotos [1], pelo cinema [2], as aplicações estereoscópicas têm aumentado sua aplicação em diversas áreas.

A visualização estereoscópica utiliza em alguns métodos o uso de óculos especiais, como no caso da visualização anaglífica (que utiliza óculos com filtros coloridos) e na visualização por luz polarizada (que utiliza óculos com filtros polarizadores de luz).

Em outros, como no caso dos monitores auto-estereoscópicos, a percepção da profundidade pode ser observada sem o uso de óculos.

A visualização anaglífica tem alcançado diversas áreas, tais como a visualização científica [3], a visualização médica [4], ferramenta de ensino [5], e a área de cinema [6], pois seu custo é baixo se comparada à outras técnicas [7]. Ela exige apenas um monitor ou projetor para geração das imagens e um par de óculos com filtros coloridos para a visualização.

As tecnologias de visualização (óculos, monitores, projetores, etc.) vêm se desenvolvendo, evidenciando que a área da estereoscopia está em evolução [8] [9]. Isso contrasta com as técnicas de codificação e armazenamento, que se encontram estagnadas. Desde Lenny Lipton [10] [11], na década de 1990, até os dias de hoje, a gravação de vídeo estereoscópico em um arquivo digital é realizada armazenando o par de vídeos estéreo em um único arquivo de vídeo, colocando-os lado-a-lado ou sobrepostos, sem nenhum tipo de compressão. Nenhuma técnica de codificação para esse tipo de vídeo foi desenvolvida desde então. Isso torna complexos e custosos os processos de produção (captação de imagens), reprodução (exibição de imagens), edição e armazenamento de vídeos estéreo [12].

Os métodos de codificação de vídeo digital empregam, em conjunto, duas técnicas de compressão: com e sem perdas. A utilização da compressão sem perdas apenas não é suficiente para a demanda atual das aplicações multimídia envolvendo vídeos [13]. A compressão com perdas gera diferenças nas cores, geralmente imperceptíveis em um vídeo comum. Porém, tais perdas podem impossibilitar a correta visualização de vídeos estereoscópicos anaglíficos.

Em direção ao estabelecimento de um método para codificação de vídeos estereoscópicos, incluindo os anaglíficos, é necessário analisar todo o processo de codificação de vídeo detectando os pontos onde ocorrem perdas. Uma etapa desse processo que pode afetar diretamente a visualização estereoscópica anaglífica é a quantização, a qual depende da aplicação de transformadas. As transformadas mais comuns são a DCT (*Discrete Cosine Transform*) e a DWT (*Discrete Wavelet Transform*), e seus processos de quantização são distintos.

O objetivo deste trabalho é analisar como as transformadas DCT e DWT com suas quantizações afetam a visualização estereoscópica anaglífica, identificando limiares para a sua utilização. A metodologia empregada visa medir os ruídos gerados em cada canal de cor RGB de forma independente, e seu impacto na percepção de vídeos estereoscópicos anaglíficos.

Este artigo está organizado do seguinte modo: a seção 2 apresenta os conceitos básicos da estereoscopia; a seção 3 mostra a estereoscopia anaglífica aplicada a vídeos digitais; a seção 4 discute os modelos de compressão de vídeo, e as etapas do processo em que podem ocorrer perdas; a seção 5 detalha o desenvolvimento de uma base de testes estereoscópica anaglífica e a escolha de técnicas de codificação utilizadas para testes de codificação; a seção 6 mostra os experimentos realizados utilizando-se medidas objetivas e subjetivas de qualidade aplicada aos vídeos codificados; por último, a seção 7 apresenta as conclusões do trabalho.

## 2. ESTEREOSCOPIA

A visão tridimensional que temos do mundo é resultado da interpretação, pelo cérebro, das duas imagens bidimensionais que cada olho capta a partir de seu ponto de vista e das informações sobre o grau de convergência e divergência. Os olhos humanos estão em média a 65 milímetros um do outro e podem convergir de modo a cruzarem seus eixos em qualquer ponto a poucos centímetros à frente do nariz, ficando estrábicos; podem também divergir ou ficar em paralelo quando se foca algo no infinito [14]. Além de imagens, o cérebro coordena os movimentos dos músculos dos globos oculares e recebe informações sobre o grau de convergência ou divergência dos eixos visuais, o que lhe permite auferir a distância em que os olhos se cruzam em um determinado momento [15].

Podemos dizer que “um visor estereoscópico é um sistema óptico cujo componente final é o cérebro humano” [16]. Isto porque é o cérebro que faz a fusão das duas imagens (uma para cada olho) resultando em noções de profundidade, como é ilustrado na Figura 1. São capturadas duas imagens ligeiramente diferentes relativas à mesma cena, da mesma maneira que vemos o mundo real.

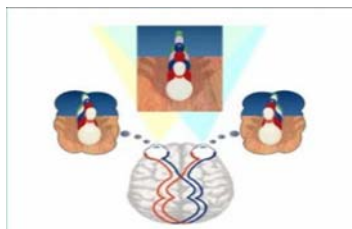


Figura 1. Visões de cada olho da mesma cena.

## 3. ESTEREOSCOPIA ANAGLÍFICA

O método de visualização estereoscópica anaglífica, é o mais simples dos métodos. A primeira experiência cinematográfica tridimensional utilizando esse método data de 1922 com o filme “Power of Love” [2]. Esta técnica caracteriza-se em colorizar com uma cor primária diferente cada uma das imagens referentes a cada olho, de modo que o espectador possa separar cada uma das imagens que se encontram misturadas na tela utilizando óculos com uma lente vermelha e outra ciano [17].

Para a visualização o espectador necessita utilizar óculos com um lado com lente vermelha (esquerda) e o outro com lente ciano (direita), para separar cada uma das imagens que se encontram misturadas no monitor ou projeção.

Para a codificação deste tipo de vídeo estereoscópico é necessário separar os canais RGB dos vídeos do par estereoscópico, sendo

que do vídeo que corresponde à visão do olho esquerdo é extraída a informação do canal vermelho, e do vídeo que corresponde à visão do olho direito extraí-se os canais azul e verde, como mostra a Figura 2. Compõe-se então com a componente vermelha da visão do olho direito e das componentes azul e verde da visão do olho esquerdo um novo vídeo RGB. A imagem anaglífica resultante pode ser observada na Figura 3.

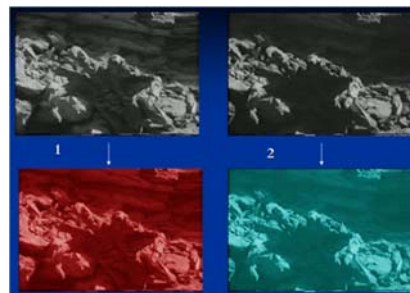


Figura 2. Processo de extração do Canal Vermelho do Vídeo 1 e dos Canais Verde e Azul do Vídeo 2.

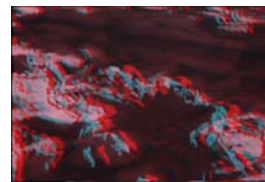


Figura 3. Exemplo de um quadro anaglífico (a percepção de profundidade desta imagem pode ser observada com óculos anaglífico ciano-vermelho).

## 4. MODELOS DE COMPRESSÃO DE VÍDEO

Um arquivo ou fluxo de vídeo digital pode ser comprimido pela remoção da sua redundância. Em um sistema de compressão sem perda, a redundância removida não altera a reconstrução dos dados. Infelizmente a compressão sem perda comprime pouco as imagens e principalmente vídeos digitais, sendo que a grande maioria dos codificadores utiliza os dois métodos em conjunto: compressão com e sem perda [13].

Nos algoritmos de compressão a remoção das redundâncias temporal, espacial e/ou de domínio da frequência é realizada em quatro etapas (modelo utilizado pela maioria dos CODECs – Codificadores/DECodificadores – com alto índice de compressão) [18]:

**1) Redução das informações contidas no espaço de cor:** inicialmente temos um vídeo representado pelo espaço de cor RGB que passa a ser representado pelo espaço de cor YCbCr. A maioria dos codificadores faz essa conversão diminuindo a quantidade de cores (redução não perceptível) e mantendo a informação de luminância, em proporção de 2 para 4 (modelos 4:2:2 e 4:2:0) ou de 1 para 4 (modelo 4:1:1). O modelo 4:4:4 mantém a quantidade de cores.

**2) Aplicação de transformações matemáticas sobre as imagens:** as duas transformadas matemáticas mais utilizadas para comprimir vídeo são a DCT (*Discret Cossine Transform*) e a DWT (*Discrete Wavelet Transform*), que são aplicadas sobre blocos de 8 x 8 ou 16 x 16 pixels das matrizes das imagens. Os

blocos resultantes possuem a informação de redundâncias melhor explicitadas (as matrizes resultantes da transformada possuem mais valores próximos a zero do que a matriz original), o que auxilia processos de compressão posteriores.

**3) Quantização:** neste processo as matrizes resultantes das transformadas passam por um corte de informações, de modo que sejam zerados os valores de cor e luminância para os quais o olho humano é menos sensível (menor percepção).

**4) Remoção da redundância binária:** esta etapa utiliza codificação de informação por entropia, onde não ocorre perda de informação.

Este trabalho tem foco na análise do quanto às transformadas matemáticas e sua posterior quantização influenciam na visualização de vídeos estereoscópicos anaglíficos. Para realizar esta análise escolheu-se CODECs que implementam em software de código aberto técnicas padronizadas de codificação:

- **MJPEG:** utiliza redução do espaço de cor para YCbCr 4:2:2, realiza compressão intra-quadro (com a técnica de compressão JPEG [19]) e utiliza DCT;
- **x264:** utiliza redução do espaço de cor para YCbCr 4:2:0, realiza compressão intra e inter-quadros e utiliza DCT. É a implementação livre do MPEG 4 part10 [20];
- **Rududu:** utiliza redução do espaço de cor para YCbCr 4:2:2, realiza compressão intra-quadro e transformada DWT (baseada na wavelet Antonini 9/7). O projeto Rududu é uma implementação desenvolvida como projeto de código aberto<sup>1</sup>.

## 5. CAPTURA, PROCESSAMENTO E CODIFICAÇÃO DE VÍDEOS ESTÉREO ANAGLÍFICOS

Apesar de a literatura reportar conjuntos de vídeos bidimensionais padronizados [21] [22] para que desenvolvedores possam realizar testes (de conformidade, de qualidade, etc.), não existem tais conjuntos para vídeos estereoscópicos. Assim, neste trabalho, foram produzidas seqüências de vídeos estereoscópicos. Tais seqüências foram produzidas seguindo critérios definidos pelo comitê de TV Digital ITU-T (*International Telecommunication Union*) [23] para a criação de bases de testes, constituindo uma das contribuições deste trabalho. Os critérios seguidos foram:

- Uma seqüência estressando cor.
- Uma seqüência estática (com pouco movimento).
- Uma seqüência estressando luminância;
- Devem existir várias seqüências.
- Devem existir seqüências com detalhes de espaço e movimentação.
- Todas as seqüências originais não podem ter ruído.

### 5.1. Captura

A captura de vídeos estereoscópicos exige câmeras especiais para captura, pois a visão dos dois olhos precisa ser simulada. Para realizar a captura de pares de vídeo estereoscópicos, foi construída uma câmera com dois blocos óticos (distantes 6,5 cm),

sendo que o sinal de vídeo desses blocos foi gravado em dois dispositivos distintos capazes de realizar a gravação de vídeos digitais. Baseados nos critérios do ITU-T, foram produzidas as seqüências de vídeos estereoscópicos descritas na Tabela 1. Todos os vídeos possuem dimensões de 720x480 pixels e 30 quadros por segundo.

**Tabela 1. Vídeos estereoscópicos produzidos com base nos critérios do ITU-T [21]**

No.	Seqüência	Nº de frames	Características
1	Andando pela Rua	350	Movimento em primeiro plano e fundo estático
2	Árvores	600	Movimento e muitas cores
3	Batata	400	Estático
4	Catando Lixo	500	Movimento em primeiro e segundo plano em diferentes direções, poucas cores
5	Gramado	600	Alta luminância, muitas cores
6	Por do Sol	600	Movimento e baixa luminância

### 5.2. Pré-Processamento

Após a captura os vídeos foram sincronizados para serem pré-processados. De posse dos pares estéreo listados na Tabela 1, outros seis pares foram criados, descartando a informação de crominância e mantendo apenas a luminância. Estes novos vídeos em escalas de cinza foram criados para verificar a influência da ausência de cores na percepção da profundidade em vídeos estéreo anaglíficos.

Para criar os vídeos estereoscópicos anaglíficos foi necessário separar os canais de cores dos pares originais e então criou-se um novo vídeo com a metodologia descrita na Seção 3. Foi desenvolvida então uma aplicação com essa finalidade. O código foi desenvolvido em um computador com o sistema operacional Windows e na linguagem Java. Utilizando a API JMF<sup>2</sup>, a partir de dois vídeos que compõe um par estéreo, é produzido um novo vídeo cuja componente vermelha provém do vídeo cuja visão seria do olho esquerdo e as componentes azul e verde provém do vídeo cuja visão seria a do olho direito. Esse processo é ilustrado na Figura 2 e produz a imagem da Figura 3.

### 5.3. Codificação

Depois de realizado o processamento para produzir vídeos estereoscópicos anaglíficos (ilustrados na Figura 4), estes foram codificados com os codificadores escolhidos anteriormente.

O ajuste dos parâmetros dos codificadores pode interferir nos resultados da compressão [24]. A etapa de quantização (que é definida de acordo com a transformada matemática) interfere diretamente nos resultados se for ajustada para que a perda de informação seja grande. Assim, a escolha dos parâmetros foi realizada visando conseguir a menor interferência da etapa de quantização na codificação dos vídeos.

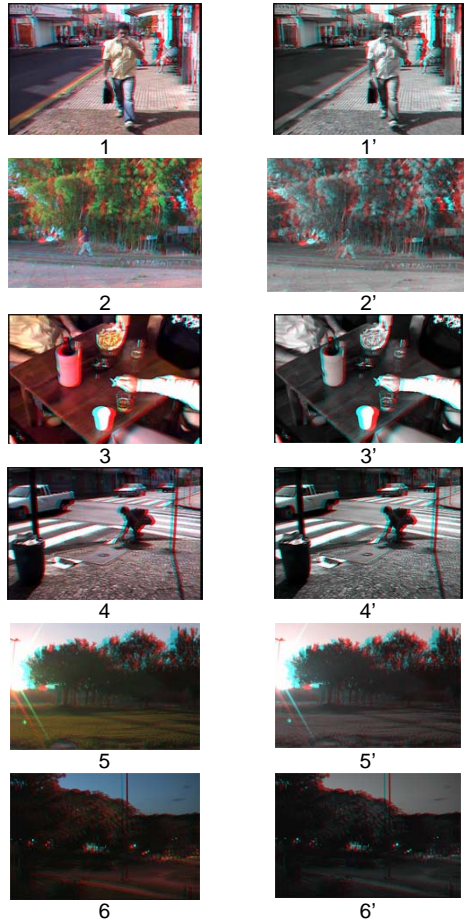
A codificação dos vídeos com os codificadores MJPEG e x264 foi realizada com o software AVIDemux<sup>3</sup> (também funcionado no

<sup>1</sup> <http://rududu.ifrance.com>

<sup>2</sup> <http://java.sun.com/javase/technologies/desktop/media/jmf/>

<sup>3</sup> O software AVI Demux é um software de código aberto que opera em diferentes sistemas operacionais (Windows, Linux e MacOSX) e realiza

sistema operacional Windows), sendo a qualidade do MJPEG ajustada para 100%, enquanto o x264 foi ajustado dois passos de codificação com *bitrate* médio de 8000 Kbps (a máxima qualidade permitida). Para o codificador Rududu, foi utilizado o software Virtual Dub<sup>4</sup> e foi ajustada a máxima qualidade.



**Figura 4. Quadros sem compressão de vídeos estereoscópicos anaglíficos utilizados para testes 5 (a profundidade destes quadros podem ser percebida com óculos anaglíficos, com filtro ciano e vermelho).**

Em uma análise prévia, nos vídeos codificados a partir de pares em escalas de cinza, a percepção estereoscópica pode ser notada sem maiores dificuldades. Contudo nos vídeos originados a partir de pares coloridos, e em especial naqueles que existia maior quantidade de cores, erros de codificação se mostraram mais aparentes.

Alguns desses erros são característicos do processo de quantização aplicado às próprias transformadas: na DCT o processo causa o aparecimento de artefatos (*blockiness*), enquanto

na DWT o processo causa embaçamento (*blur*) [25]. Estes fatores causam interferências na percepção dos vídeos estereoscópicos anaglíficos codificados. A próxima seção (Seção 6) apresenta uma análise realizada para identificar a partir de que momento os ruídos introduzidos pelos processos começam a interferir na visualização. Tal análise constitui um subsídio para saber qual a margem de compressão que se pode aplicar, com segurança, a um vídeo estereoscópico.

## 6. MEDIDAS OBJETIVAS E SUBJETIVAS

Para medir o grau de interferência da quantização das transformadas DCT e DWT na percepção estereoscópica anaglífica foram realizados dois testes: um subjetivo, usando como medida o MOS (*Mean Opinion Score*) [22]; e outro objetivo, usando como medida o PSNR<sup>5</sup> (*Peak to Signal-to-Noise Ratio*) [22].

### 6.1. MOS

A medida subjetiva do MOS foi realizada para averiguar se a dificuldade da percepção da profundidade correspondia à análise prévia realizada. Seu cálculo foi realizado de acordo com as normas do ITU-T [23][26]. Foram montadas estruturas de vídeo ABAB, onde A é o vídeo original e B é vídeo processado. Nos intervalos dos vídeos existe um trecho de tela cinza, como mostrado na Figura 5. O avaliador é instruído a votar durante a visualização do segundo trecho B.



**Figura 5. Estruturas de vídeo utilizadas para medida do MOS.**

Cada avaliação pontua um vídeo de 1 (a pior qualidade) até 5 (a melhor qualidade). Para determinar o MOS é calculada a média dos valores atribuídos aos vídeos. O grupo de avaliadores dos vídeos 1 a 6 e 1' a 6', foi composto de 20 pessoas. Como este experimento tinha relação direta com a percepção de cores todos os indivíduos foram submetidos ao teste de Ishihara, para averiguação de daltonismo<sup>6</sup>. As figuras 6, 7 e 8 mostram o cálculo do MOS para cada uma das seqüências de teste.

Nos gráficos das Figuras 6 à 8 nota-se que o vídeo composto por escalas de cinza tem percepção melhor ou igual ao seu

<sup>5</sup> Esta métrica calcula a relação entre o máximo possível de potência de um sinal, pela potência do ruído, quando comparamos um sinal antes e depois de um processo de degradação, sendo que a unidade utilizada para representá-lo é o dB (decibel). Aplicando este conceito em vídeos e imagens, temos que o PSNR é a relação entre a entrada e a saída de um processo de compressão com perdas, que avalia o quanto a compressão introduziu ruídos na imagem ou quadro original. Matematicamente o PSNR de uma imagem de dimensão  $m$  por  $n$  é representado por:

$$d(X, Y) = 10 \log_{10} \frac{255^2 mn}{\sum_{i,j=1}^m \sum_{k=1}^n (X_{i,j} - Y_{i,j})^2}$$

<sup>6</sup> O teste de Ishihara consiste na exibição de uma série de cartões pontilhados em várias tonalidades diferentes. Esse é o método mais frequentemente utilizado para se diagnosticar a presença do daltonismo, sobretudo nas deficiências envolvendo a percepção das cores vermelho e verde.

[http://genetica.ufcspa.edu.br/seminarios%20monitores/2004/daltonismo\\_texto.pdf](http://genetica.ufcspa.edu.br/seminarios%20monitores/2004/daltonismo_texto.pdf)

conversão de vídeo e áudio com implementações de código aberto. <http://avidemux.org/>

<sup>4</sup> O software VirtualDub é um software de código aberto para o sistema operacional Windows, e realiza codificação de vídeo e de áudio, com os codificadores instalados juntos ao sistema operacional. <http://virtualdub.org>

correspondente colorido. Outro fato ligado a cor, é a percepção do vídeo 2; este vídeo possui uma grande quantidade de cores e teve uma avaliação abaixo dos outros vídeos.

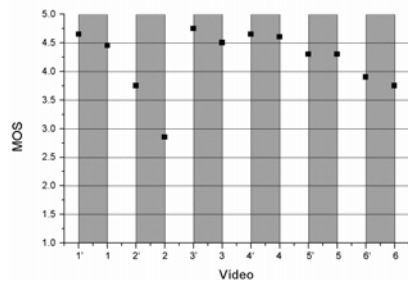


Figura 6. MOS dos vídeos codificados com o codificador MJPEG.

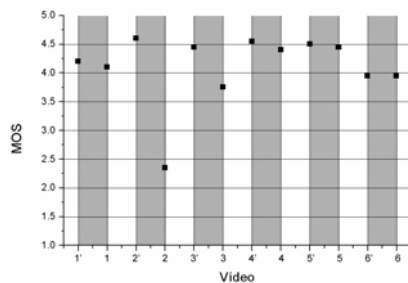


Figura 7. MOS dos vídeos codificados com o codificador x264.

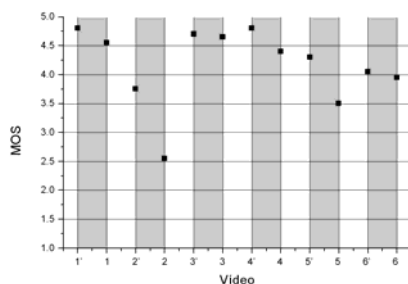


Figura 8. MOS dos vídeos codificados com o codificador Rududu.

## 6.2. PSNR

Após a avaliação subjetiva, a próxima etapa constituiu-se de uma avaliação objetiva de qualidade de vídeo. Como os codificadores trabalham com sistemas diferentes de compressão, no qual variam o espaço de cor YCbCr e o tipo de transformada matemática, a medida objetiva aplicada aos vídeos foi o cálculo de ruído para cada uma das componentes de cor das imagens que compõe o vídeo.

O PSNR é um índice de qualidade objetiva dado pela relação entre a entrada e a saída de um processo de compressão de imagens com perdas. Avalia a quantidade de ruídos a introduzidos na compressão, tendo como base de comparação a imagem original.

O PSNR para cada uma das componentes RGB dos vídeos codificados foi medido com a ferramenta MSU VQMT (*Video Quality Measurement Tool*<sup>7</sup>). Essa ferramenta calcula o valor do PSNR de cada quadro do vídeo codificado (comparando-o com o vídeo original), assim como uma média de PSNR de todos os valores. O valor de máxima qualidade é o valor 100 dB. Os gráficos das Figuras 9, 10 e 11, mostram os valores médios de PSNR para cada componente de cor dos vídeos da sequência de teste.

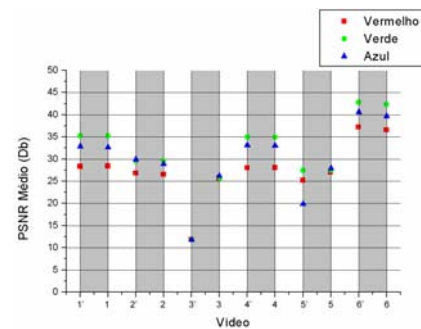


Figura 9. Valores médios de PSNR da codificação dos vídeos estereoscópicos anaglíficos com o codificador MJPEG.

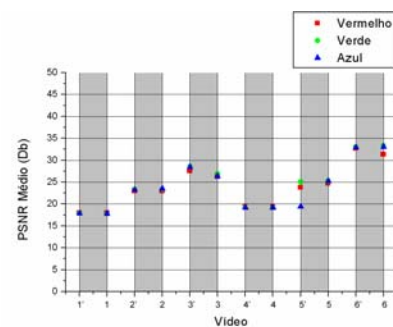


Figura 10: Valores médios de PSNR da codificação dos vídeos estereoscópicos anaglíficos com o codificador x264.

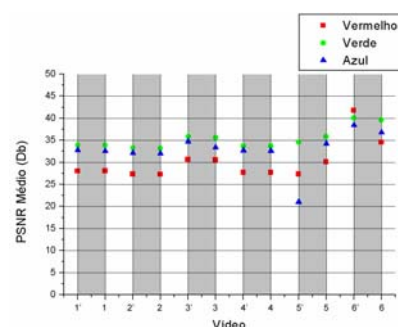


Figura 11. Valores médios de PSNR da codificação dos vídeos estereoscópicos anaglíficos com o codificador Rududu.

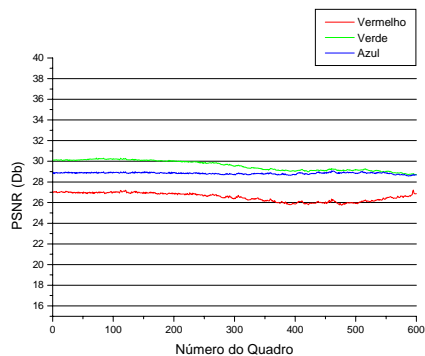
<sup>7</sup> O software MSU VQMT foi desenvolvido pelo grupo de pesquisa MSU Graphics & Media Lab da Universidade de Moscou. [http://compression.ru/video/quality\\_measure/video\\_measurement\\_tool\\_en.html](http://compression.ru/video/quality_measure/video_measurement_tool_en.html)



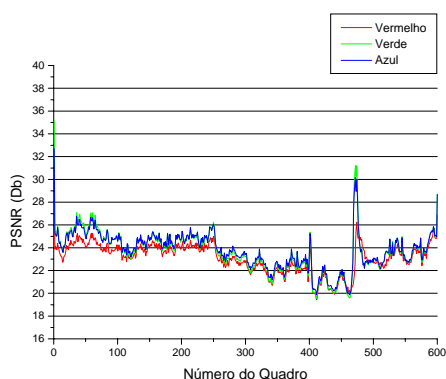
### 6.3. Discussão dos Resultados

Em todos os gráficos de MOS (Figuras de 6 à 8), os vídeos compostos por escalas de cinza tem percepção melhor ou igual ao seus correspondentes coloridos, mesmo quando as diferenças entre os PSNR médios são maiores. Isso mostra que a ausência de cores no par de vídeo estereó original faz com que as diferenças de PSNR não causem alteração na percepção estereoscópica. Isso pode ser verificado no caso dos Vídeos 5 e 5' e nas suas figuras correspondentes: 9, 10 e 11. O PSNR médio se mostrou com maiores diferenças entre as componentes RGB do vídeo composto por escalas de cinza, mas o MOS indicou que a pior percepção desse par continuou sendo o vídeo formado pelo par estéreo com cores.

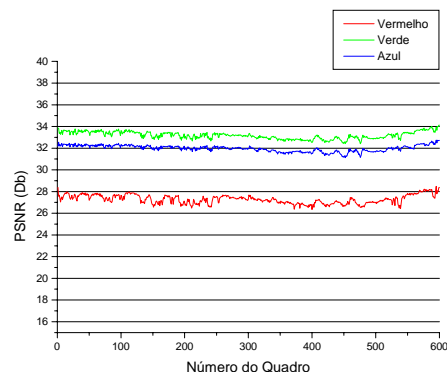
O valor médio de PSNR é indicado para uma avaliação macroscópica de amostras, mas a análise dos valores de PSNR para cada um dos quadros dos vídeos é necessária para uma avaliação pontual. O par de vídeos 2 e 2' apresentou valores de MOS abaixo da média, e por isso as figuras 12, 13 e 14 detalham os valores de PSNR quadro a quadro.



**Figura 12. Valores de PSNR da codificação do vídeo estereoscópico anaglífico 5 com o CODEC MJPEG.**



**Figura 13. Valores de PSNR da codificação do vídeo estereoscópico anaglífico 5 com o CODEC x264.**



**Figura 14. Valores de PSNR da codificação do vídeo estereoscópico anaglífico 5 com o CODEC Rududu.**

Ao analisar os valores de PSNR das três codificações utilizadas, nota-se que nas codificações MJPEG e Rududu (Figuras 12 e 14) as variações são mais uniformes do que na codificação x264 (Figura 13). A codificação x264 utilizada implementa o modelo 4:2:0 para cor, enquanto as outras duas implementam o modelo 4:2:2. Então, a diferença observada comprova que além das transformadas e sua posterior quantização, o modelo de cores utilizado impacta no resultado.

O cálculo de um valor médio de PSNR, considerando os três canais de cor RGB, aponta a codificação Rududu com o melhor resultado, mostrando que a transformada wavelet e seu processo de quantização traz melhores resultados do que as transformadas DCTs e seu respectivo processo de quantização.

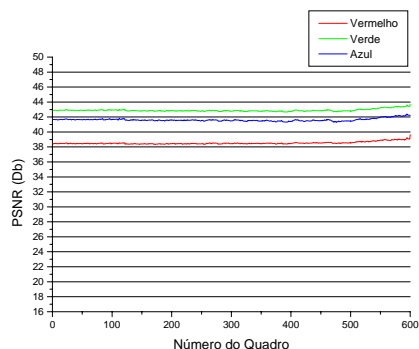
**Tabela 2. Médias Gerais de PSNR da codificação do vídeo estereoscópico anaglífico 5.**

MJPEG	x264	Rududu
28,30 Db	23,24 Db	30,79 Db

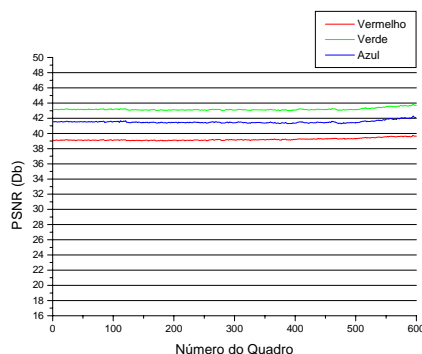
A variação média de PSNR na MJPEG foi de 2 a 3 dB entre as componentes vermelha e azul e vermelha e verde. Essa variação foi menor do que a da Rududu, que variou de 5 dB entre as mesmas componentes. O MOS de ambas as medidas, apesar da variação das medidas MJPEG e Rududu serem diferentes, ficou próxima, indicando que o valor médio de PSNR também influencia a percepção. Seguindo o mesmo raciocínio, ao observar as variações de PSNR na codificação x264, temos picos de 5 dB e pelo MOS nota-se que neste vídeo a percepção foi pior do que a dos vídeos codificados com MJPEG e Rududu.

Como o desempenho da codificação Rududu foi superior, a MJPEG e x264, outros modelos de wavelets foram testados. Foi desenvolvida uma aplicação de codificação que utiliza a biblioteca Rududu que permite codificar imagens com o espaço de cor YCbCr 4:2:2 e pode utilizar as wavelets CDF53, CDF97 e Haar 2. A biblioteca também permite escolher a quantização dos resultados, numa escala de 0 (sem perda) a 16.

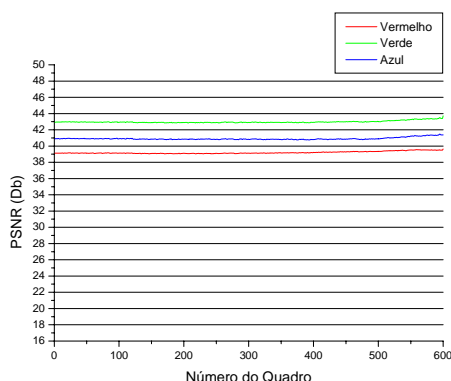
O vídeo 2 foi o que apresentou maior dificuldade na percepção estereoscópica na análise prévia, por isso foi o escolhido para novos testes de PSNR e MOS. As Figuras 15, 16 e 17 mostram o valor do PSNR quadro-a-quadro, enquanto a Tabela 3 mostra o valor de MOS para as amostras.



**Figura 15. Valores de PSNR da codificação do vídeo estereoscópico anaglífico 5 com a transformada wavelet CDF53.**



**Figura 16. Valores de PSNR da codificação do vídeo estereoscópico anaglífico 5 com a transformada wavelet CDF97.**



**Figura 17. Valores de PSNR da codificação do vídeo estereoscópico anaglífico 5 com a transformada wavelet Haar2.**

**Tabela 3. Valores de MOS e de PSNR médio (por canal e geral) da codificação do vídeo estereoscópico anaglífico 5 codificado com aplicação desenvolvida.**

Wavelet	PSNR R médio	PSNR G médio	PSNR B médio	PSNR geral médio	MOS
CDF53	39.21696	43.18164	41.51135	41.30331	4,55
CDF97	38.52734	42.89480	41.61420	41.01211	4,65
Haar2	39.20264	42.97876	40.90741	41.02960	4,55

Os gráficos de PSNR das Figuras 15, 16 e 17 mostram um comportamento linear das transformadas sobre os canais RGB. As diferenças entre o canal vermelho e os canais azul e verde se mostraram inferiores a 4 dB, e isto não resultou em dificuldades em sua percepção, como indicado pelo MOS da Tabela 3.

Os valores médios gerais de PSNR (dados pela Tabela 3), mostram que as codificações realizadas com as transformadas wavelets, se mostraram superiores as médias de codificações realizadas com transformadas DCT (cujos dados estão na Tabela 2).

## 7. CONCLUSÕES

A partir dos resultados apresentados pode-se concluir que a ausência de cores nos pares de vídeo estéreo antes do processamento, facilita a percepção da profundidade em vídeos estéreo anaglíficos. A presença de muitas cores nos pares de vídeo que compõe o vídeo anaglífico pode interferir na percepção da profundidade se este for codificado com perda.

Ao analisarmos os valores de PSNR presentes no gráfico da Figura 11 referentes ao vídeo 5' (composto por vídeos em escalas de cinza) vemos que os valores de PSNR são bem distantes, se tomarmos o canal vermelho como referência. Contudo, o MOS do mesmo vídeo mostrado na Figura 8, mostra que sua percepção foi bem aceitável. Isso revelou que o PSNR a ser avaliado deve ser calculado entre a componente ciano (composta da média das componentes azul e verde) e a componente vermelha.

Realizando a análise das diferenças entre as componentes vermelho e ciano do vídeo 5 nas Figuras 15, 16 e 17, temos alguns momentos em que a diferença se mostra inferior a 4dB, e isso não resultou em dificuldades em sua percepção, como indicado pelo MOS da Tabela 3. Isso mostra que diferenças de 5 dB entre os canais vermelho e o canal ciano ocasiona em dificuldades em sua percepção, o que o torna um valor proibitivo.

Os valores do PSNR médio geral das codificações que utilizam as transformadas wavelets superaram os valores médios das codificações que utilizaram a DCT, indicando que as transformadas wavelets e seu processo de quantização apresentam melhores resultados na codificação de vídeos estereoscópicos anaglíficos.

Ainda como contribuições deste trabalho tem-se uma base de testes para vídeos estereoscópicos e um software para a geração de vídeos estereoscópicos anaglíficos a partir de um par estéreo de vídeos.

Como trabalhos futuros, pretende-se: disponibilizar a base de teste para acesso público; realizar testes e comparações entre os diferentes modelos de cor (4:4:4, 4:2:2, 4:1:1 e 4:2:0) identificando os limiares de perdas toleráveis em cada caso e

analisar como e quanto a etapa de quantização influencia na perda da percepção anaglífica.

## AGRADECIMENTOS

Nossos agradecimentos a Djalma Ribeiro Jr. e Pedro Dolosic, do Departamento de Artes e Comunicação da UFSCar que auxiliaram na montagem da câmera e nas captações estereoscópicas. À FAPESP, pelo apoio financeiro fundamental para o desenvolvimento e divulgação desta pesquisa.

## REFERÊNCIAS

- [1] Parente, J. I. *A Estereoscopia no Brasil 1850-1930*. Ed. Sextante, Rio de Janeiro, 1999.
- [2] Zone, R. *Stereoscopic cinema and the origins of 3-D film, 1838-1952*. The University Press of Kentucky, 2007. ISBN: 0813124611.
- [3] Matsuura, F.; Fujisawa, N. *Anaglyph Stereo Visualization by the Use of a Single Image and Depth Information*. Journal of Visualization archive, v. 11, n. 1, p. 79-86, janeiro de 2008.
- [4] Menezes, M. S. et al. *Estereoscopia Aplicada à Neuroanatomia - Estudo comparativo entre as técnicas de filtro de cores e de polarização*. Arquivo de Neuro-Psiquiatria, v. 60, n. 3B, São Paulo, 2002.
- [5] Stoffer, P. W., Phillips, E., Messina, P. *Anaglyph Image Technology as a Visualization Tool for Teaching Geology of National Parks*. American Geophysical Union, Fall Meeting, 2003.
- [6] Suppia, A. O. *Monstro Brasileiro Revive em 3D*. Revista Cultura e Ciência, ano 59, n.2, p.57-59, 2007. Disponível em: <[http://cienciaecultura.bvs.br/scielo.php?script=sci\\_pdf&pid=S0009-67252007000200024&lng=pt&nrm=iso](http://cienciaecultura.bvs.br/scielo.php?script=sci_pdf&pid=S0009-67252007000200024&lng=pt&nrm=iso)>. Acesso em: 05 mai. 2009. ISBN 85-0009-6725.
- [7] Andrade, L., Goularte, R. *Codificação de Vídeos Estereoscópicos*. Qualificação de Doutorado, ICMC – USP São Carlos, 2009. Disponível em: <[http://www.ies.ufscar.br/leoandrade/uploads/Docs/Doutorado/qualificacao\\_12122008.pdf](http://www.ies.ufscar.br/leoandrade/uploads/Docs/Doutorado/qualificacao_12122008.pdf)>. Acesso em: 05. mai. 2009.
- [8] Dodgson, N. A. *Autostereoscopic Image Compression*. Rainbow Graphics Researchs Group, Computer Laboratory, University of Cambridge, St. Pembroke, 1998. Disponível em: <<http://www.cl.cam.ac.uk/~nad/compr/compr.html>>. Acesso em: 10 mai. 2009.
- [9] Perlin, K. et. al. *Recent Advances in the NYU Autostereoscopic Display*. Proceedings of the SPIE, v. 4297, São José, California, 2001. Disponível em: <<http://www.mrl.nyu.edu/publications/autostereo/spie2001.pdf>>. Acesso em: 10 mai. 2009.
- [10] Lipton, L. *Foundations of the Stereoscopic Cinema, a study in depth*. Nova Iorque, Van Nostrand Reinhold Co., 1982. ISBN: 0442247249.
- [11] Lipton, L. *Stereo-Vision Formats for Video and Computer Graphics*. White Paper, 1997. Disponível em: <[http://www.cours.polymtl.ca/inf6802/stereo/body\\_stereo\\_formats.html](http://www.cours.polymtl.ca/inf6802/stereo/body_stereo_formats.html)>. Acesso em: 10 mai. 2009.
- [12] Andrade, L. A.; *Difusão de Filmes Estereoscópicos*. Revista RUA – Especial 1 – A Era Digital e seus Desdobramentos Estéticos, 2008. ISSN 1983-3725. Disponível em: <<http://www.ufscar.br/rua/site/?p=657>>. Acesso em: 02 mai. 2009.
- [13] Gonzalez, R. C., Woods, R. E. *Digital Image Processing. Third Edition*, Ed. Prentice Hall, 2008. ISBN: 013168728X.
- [14] Okoshi, T. *Three-Dimensional Imaging Techniques*. Academic Press, 1977. ISBN: 0125252501.
- [15] McKay, C. H. *Three-Dimensional Photography - Principles of Stereoscopy*. Revised Edition. American Photography Book Department, Nova Iorque, 1953.
- [16] StereoGraphics Corporation. *Stereographics Developers Handbook: Background on Creating Images for CrystalEyes and SimulEyes*. StereosGraphics Corporation, 1997. Disponível em <[http://www.stereographic.com/support/downloads\\_support/handbook.pdf](http://www.stereographic.com/support/downloads_support/handbook.pdf)>. Acesso em: 25 jun. 2008.
- [17] Mancini, A. *Disparity Estimation and Intermediate View Reconstruction for Novel Applications Stereoscopic Video*. Thesis of Mestre submitted to the Department of Electrical Engineering McGill University, Montreal, Canadá, 1994.
- [18] Richardson, I. E. G. *H.264 and Mpeg-4 video compression*. Editora Wiley. Londres, 2003. ISBN: 0470848375.
- [19] Furht, B. *A Survey of Multimedia Compression Techniques and Standards – Part I : JPEG Standard*. Real-Time Imaging, vol. 1, n. 1, 1995.
- [20] Ebrahimi, F. *The MPEG-4 book*. IMSC Press Multimedia Series, Prentice Hall, 2002. ISBN: 0130616214.
- [21] ITU-T Recommendation J.247 : *Objective perceptual multimedia video quality measurement in the presence of a full reference*. International Telecommunication Union, 2008.
- [22] Winkler, S. *Digital Video Quality*. Ed. Wiley, 2005. ISBN 0-470-02404-6.
- [23] ITU-T Recommendation J.144. *Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference*. International Telecommunication Union, Genebra, Suíça, 2004.
- [24] MSU Graphics & Media Lab (Video Group). *Introduction to Video Codecs Comparison*. Disponível em: <[http://compression.ru/video/codec\\_comparison/introduction\\_en.html](http://compression.ru/video/codec_comparison/introduction_en.html)>. Acesso em: 02 mai. 2009.
- [25] Ebrahimi F., Chamik M., Winkler S. *JPEG vs. JPEG 2000: An Objective Comparison of Image Encoding Quality*. Applications of Digital Image Processing XXVII. Proceedings of the SPIE, v. 5558, p. 300-308, 2004.
- [26] ITU-R Recommendation BT.500-11. *Methodology for the subjective assessment of the quality of television pictures*. International Telecommunication Union, Genebra, Suíça, 2002.