

Métodos de Codificação de Vídeos Estereoscópicos

Leonardo Antônio de Andrade

Orientador: *Prof. Dr. Rudinei Goularte*

Monografia apresentada ao Instituto de Ciências Matemáticas e de Computação - ICMC/USP, para o Exame de Qualificação, como parte dos requisitos para obtenção do título de Doutor em Ciências de Computação e Matemática Computacional.

USP – São Carlos
Dezembro/2008

1. Introdução	1
2. Estereoscopia	6
2.1. Considerações Iniciais.....	6
2.2. Fundamentos	8
2.2.1. Visão Binocular Humana	8
2.2.2. Fundamentos Matemáticos	9
2.2.3. Profundidade	10
2.3. Captação Estereoscópica de Imagens em Movimento	12
2.4. Cinematografia Estereoscópica.....	12
2.5. Exibição de Imagens Estereoscópicas.....	13
2.5.1. Monitores e Projetores de Imagens Bidimensionais	13
2.5.2. Óculos para Visualização Estereoscópica.....	14
2.5.3. Monitores Autoestereoscópicos	15
2.6. Considerações Finais	17
3. Codificação de Vídeo Digital.....	18
3.1. Considerações Iniciais.....	18
3.2. Espaços de Cor.....	18
3.2. Amostragem de Cor	19
3.2.1. Modelo 4:4:4	20
3.2.2. Modelo 4:2:2	20
3.2.3. Modelo 4:1:1	21
3.2.4. Modelo 4:2:0	21
3.3. Formatos de Imagem e Vídeo, Contêineres e CODECs	22
3.3.1. CODECs.....	24
3.3.2. CODECS Espaciais.....	24
3.3.3. CODECS Espaço-Temporais.....	27
3.4. Vídeos Estereoscópicos	30
3.4.1. Vídeos Estereoscópicos Visualizados com Óculos Polarizadores.....	30
3.4.2. Vídeos Estereoscópicos Visualizados com Óculos Obturadores.....	30
3.4.3. Vídeos Estereoscópicos Visualizados com Óculos de Filtros Coloridos.....	33
3.5. Armazenamento de Vídeos Estereoscópicos	34
3.6. Considerações Finais	36
4. Trabalhos Relacionados	38
4.1. Considerações Iniciais.....	38
4.2. Armazenamento de Vídeos Estereoscópicos	38
4.3. Comparação entre Técnicas de Codificação	39
4.4. Problemas na Codificação de Vídeos Estereoscópicos Anaglíficos	40
4.5. Qualidade de Vídeo.....	42
4.6. Considerações Finais	42
5. Testes de Codificação	44
5.1. Considerações Iniciais.....	44
5.2. Medidas de Qualidade e de Compressão em CODECs	44
5.3. Considerações Finais	57
6. Plano de Pesquisa.....	58
6.1. Objetivo.....	58
6.2. Metodologia e Atividades	58
6.3. Cronograma.....	59

6.4. Considerações Finais e Resultados Esperados.....	60
Referências.....	61
Glossário	67

Índice de Figuras

Figura 1.1 – Câmera estereoscópica digital em operação. A câmera possui dois blocos óticos de captação de imagem na câmera, afastados 6,5 cm (simulando a visão humana)	1
Figura 1.2 – Par de imagens a serem exibidas nos monitores da linha WOWvx da <i>Philips</i> , lançada em 2008: a) Imagem Bidimensional e b) Mapa de Profundidade da imagem. O hardware do monitor realiza os cálculos a partir do mapa de profundidade e exibe a imagem tridimensionalmente	2
Figura 1.3 – Projeção de Luz Polarizada. Dois projetores com filtros polarizadores projetam o par estereoscópico de imagens sobre uma tela metalizada. Através do uso de um óculos com filtros polarizadores, o expectador visualiza a imagem tridimensional.....	2
Figura 1.4 – Figura estereoscópica formada pelo processo de cores complementares (uma lente vermelha e outra ciano): a) Anaglifo e b) Óculos utilizados para visão estéereoscópica anáglifica (Wood, 2000)	4
Figura 1.5 – Etapas da compressão de vídeo espaço-temporal	4
Figura 2.1 - Estereoscópio: a) Exemplo de aparelho estereoscópio e b) Esquema básico de um estereoscópio (Kirner & Tori, 2004)	6
Figura 2.2 – Anaglifo de um quadro do filme “Creature from the Black Lagoon”, de 1956	7
Figura 2.3 – Visões de cada olho da mesma cena (Watson, 1998)	8
Figura 2.4 – a) Visões da mesma cena pelos dois olhos e b) Superposição das imagens e a disparidade na retina	9
Figura 2.5 – Tipos de paralaxe: a) Paralaxe zero (ZPS), b) Paralaxe negativa e c) Paralaxe positiva	10
Figura 2.6 – Geometria básica da visão estereoscópica.....	11
Figura 2.7 – Óculos obturadores sincronizados por infra-vermelho e por cabo (Stereographics, 1997)	15
Figura 2.8 – Esquema de funcionamento de um monitor autoestereoscópico	16
Figura 2.9 – Exemplo de visualização em monitor autoestereoscópico, onde nove posições são possíveis para o espectador que observa a mesma cena por ângulos diferentes.	16
Figura 3.1: Modelo de amostragem de cor 4:4:4	20
Figura 3.2: Modelo de amostragem de cor 4:2:2	21
Figura 3.3: Modelo de amostragem de cor 4:1:1	21
Figura 3.4: Modelo de amostragem de cor 4:2:0	22
Figura 3.5: Algoritmo de codificação LZW	25
Figura 3.6: Tipos de quadros dos padrões MPEG. Os quadros I são comprimidos internamente, enquanto os quadros P e B são calculados a partir do quadro I.....	28
Figura 3.7: Compensação de movimento da codificação MPEG.....	29
Figura 3.8: Quadro estereoscópico campo-sequencial.....	32
Figura 3.9: Processo de extração do canal vermelho do vídeo 1	33
e dos canais verde e azul do vídeo 2	33
Figura 3.10: Processo de composição do vídeo anaglífico.....	34
Figura 3.11: Quadro de um vídeo lado a lado (<i>side by side</i>) quadro original: 720 x 480 – quadro lado a lado: 1440 x 480	35
Figura 3.12: Quadro de um vídeo sobreposto (<i>over/under</i>) quadro original: 720 x 480 – quadro sobreposto: 720 x 960.....	35
Figura 4.1: Comparações entre JPEG e JPEG20000. A imagem (a) mostra o gráfico da relação da compressão pela porcentagem de <i>blockiness</i> , enquanto a imagem (b) mostra um gráfico da relação da compressão pela porcentagem de <i>blur</i>	40
Figura 4.2: A imagem A mostra um recorte do quadro de um filme estereoscópico sem compressão. A imagem B o mesmo recorte com compressão MPEG4, com a qualidade máxima permitida pelo método de codificação	41
Figura 4.3: A imagem A mostra um recorte em um quadro de um filme estereoscópico, de sua componente vermelha. A imagem B mostra um recorte em um quadro com compressão MJPEG, com a qualidade máxima permitida pelo método de codificação	42
Figura 5.1: Quadros da animação <i>Elephant’s Dream</i>	45
Figura 5.2: PSNR dos canais Vermelho, Verde e Azul com o CODEC HD Cineform	46
Figura 5.3: PSNR dos canais R, G e B com o CODEC HD MJPEG2000 Morgan	47
Figura 5.4: PSNR dos canais R, G e B com o CODEC MJPEG2000 Quicktime.....	48
Figura 5.5: PSNR dos canais R, G e B com o CODEC Apple Intermediate.....	49
Figura 5.6: PSNR dos canais R, G e B com o CODEC MJPEG (Linux).....	50
Figura 5.7: PSNR dos canais R, G e B com o CODEC x264 (Linux).....	51

Figura 5.8: PSNR dos canais R, G e B com o CODEC MPEG4 (Linux).....	52
Figura 5.9: PSNR dos canais R, G e B com o CODEC Rududu e a transformada wavelet cdf97.....	54
Figura 5.10: PSNR dos canais R, G e B com o CODEC Rududu e a transformada wavelet cdf97	55
Figura 5.11: PSNR dos canais R, G e B com o CODEC Rududu e a transformada wavelet Haar4	56
Figura 5.12: PSNR dos canais R, G e B com o CODEC desenvolvido utilizando a transformada wavelet Daub4	57

Índice de Tabelas

Tabela 3.1: Padrões de Compressão de Imagens e de Vídeo.....	23
Tabela 3.2: Comparações de Custo, Qualidade, Perda de Cor e Compressão de Dados de vídeos estereoscópicos armazenados com o modelo proposto por Lipton, 1997.....	37

Os registros e o conteúdo ficcional produzido 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, pelo cinema e atualmente com monitores que permitem a visualização estereoscópica sem óculos especiais, as tecnologias de visualização vêm se desenvolvendo, contrastando com as técnicas de codificação que estão estagnadas. Os arquivos de vídeo digital que guardam as informações estereoscópicas, resumem-se a guardar o par de vídeos em um único arquivo, o que acarreta em limitações para sua exibição. Este trabalho tem como objetivo propor novas metodologias de armazenamento e de reprodução de vídeos estereoscópicos, realizando uma análise das codificações de vídeo bidimensionais e tridimensionais atuais e seu impacto na reprodução estereoscópica. Levando em conta as diversas técnicas de visualização estereoscópica, pretende-se elaborar protótipos de codificadores e reprodutores de vídeo com as metodologias propostas.

Introdução

Desde o final do século XIX, quando câmeras de película captavam os primeiros filmes, até as atuais câmeras digitais que captam vídeo em alta resolução, as imagens bidimensionais têm sido o suporte para o homem registrar a história e produzir conteúdo ficcional. Com registros bidimensionais uma importante informação para visualização tem sido descartada, a profundidade, presente apenas em poucas obras audiovisuais nos últimos dois séculos (Supia, 2007). Para que a profundidade possa estar presente em um registro visual, é necessário gravar um par de imagens bidimensionais que formam um par estéreo. Um par de imagens estéreo é composto de imagens captadas por olhos (ou câmeras), ligeiramente afastadas, como ilustrado na Figura 1.1.



Figura 1.1 – Câmera estereoscópica digital em operação. A câmera possui dois blocos óticos de captação de imagem na câmera, afastados 6,5 cm (simulando a visão humana)

Os seres humanos conseguem visualizar a profundidade devido à estereoscopia, capacidade de interpretação de um par de imagens estéreo para visualização de uma imagem tridimensional (Machado, 1997). A estereoscopia é processada no cérebro, onde a fusão das duas imagens resulta em informações quanto à profundidade, distância, posição e tamanho dos objetos, gerando uma sensação de visão de 3D (Stereographics, 1997).

Atualmente monitores especiais (Figura 1.2) e salas de cinema (Figura 1.3), trabalham com técnicas que permitem a visualização de imagens em movimento com profundidade. A essas técnicas se dá o nome de visualização estereoscópica, e seu armazenamento e difusão digital são uma realidade.

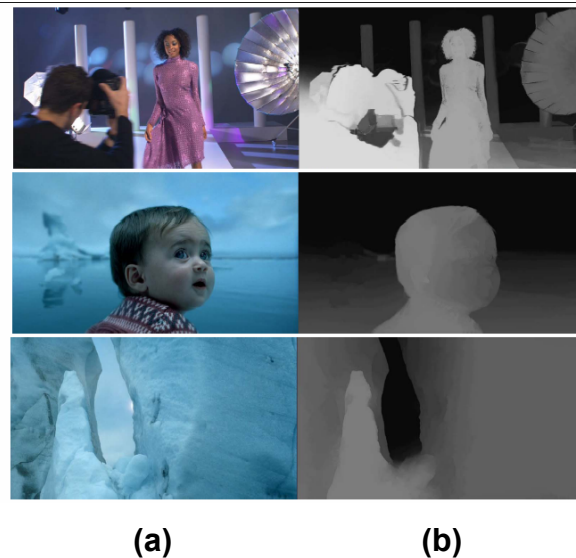


Figura 1.2 – Par de imagens a serem exibidas nos monitores da linha WOWvx da *Philips*, lançada em 2008: a) Imagem Bidimensional e b) Mapa de Profundidade da imagem. O hardware do monitor realiza os cálculos a partir do mapa de profundidade e exibe a imagem tridimensionalmente

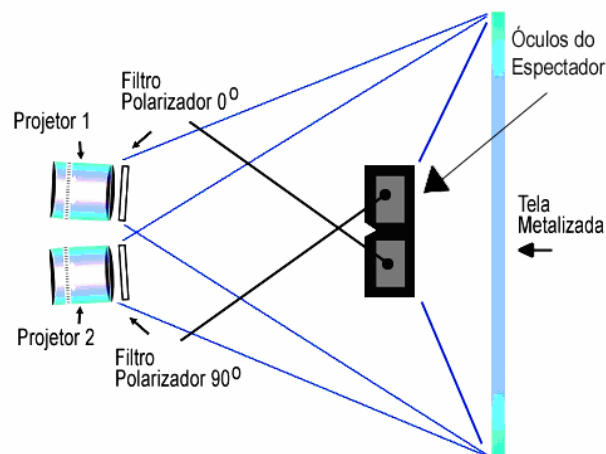


Figura 1.3 – Projeção de Luz Polarizada. Dois projetores com filtros polarizadores projetam o par estereoscópico de imagens sobre uma tela metalizada. Através do uso de um óculos com filtros polarizadores, o expectador visualiza a imagem tridimensional

A representação do espaço tridimensional através de um par de imagens é uma técnica que foi registrada em 1838, com a invenção de um aparelho denominado estereoscópio¹. Na mesma década são datados os primeiros registros fotográficos da história, conforme Patterson (1989). Apesar de serem contemporâneos, o uso do estereoscópio com pares de fotografias data de meados de 1880. Nessa época houve a popularização das máquinas fotográficas que tiravam pares de fotos para livros que exploravam o recurso da estereoscopia para entretenimento (Adams, 2001).

Nas décadas de 1920 e 1950, a estereoscopia cinematográfica foi amplamente explorada. Diversas produções foram captadas e exibidas estereoscopicamente. Na década de 1920 com produções experimentais de grande repercussão e na década de 1950 o cinema investiu em produções estereoscópicas para recuperar o público que experimentava a nova mídia da TV (Supia, 2007).

Apesar de hoje existirem diversas tecnologias de exibição estereoscópica, os métodos de codificação desse tipo de informação não evoluíram do mesmo modo. Os arquivos que guardam as informações estereoscópicas resumem-se a gravar informações dos dois vídeos em um único arquivo de tamanho maior, contendo as informações lado-a-lado (*side by side*), ou um sobre o outro (*under/over*) (Lipton, 1997). O método proposto é efetivo para a visualização de projeção de pares de imagens com luz polarizada, conforme ilustrado na Figura 1.3, e para óculos obturadores², tecnologias de alto custo.

A visualização anaglífica se apresenta como um método de visualização de baixo custo, pois com apenas um monitor ou projetor e óculos de filtros coloridos, é possível visualizar um par de imagens estereoscópicas (Figura 1.4) (Wood, 2000). A desvantagem desse método em relação aos anteriormente citados, é que as cores da visualização, não são as cores reais da captação. Além disso a metodologia proposta por Lipton não especifica um CODEC e a grande maioria dos algoritmos de codificação realiza seus cálculos a partir de uma redução do espaço de cores³ durante a compressão de dados. Isso causa distorções nas cores toleráveis em alguns métodos de visualização, mas proibitivas no método anaglífico.

A compressão de dados para a manipulação de vídeo digital tem dois importantes benefícios. O primeiro deles é tornar possível a utilização de vídeos digitais em transmissões e armazená-los em ambientes que não suportam vídeos descomprimidos. A transmissão de vídeos em tempo real pela Internet, por exemplo, só é possível com altas compressões de dados. O segundo benefício tem relação

¹ Aparelho que permite fazer uso da estereoscopia, a capacidade de reprodução da visão binocular humana a partir de duas imagens bidimensionais (Godoy, 2005). No início essas imagens eram ilustrações. Quatro décadas depois elas foram substituídas por fotografias.

² Segundo essa técnica, o observador, ao visualizar a tela do computador ou televisor, deve utilizar óculos especiais, cujas lentes são feitas de cristal líquido. As lentes podem ficar instantaneamente transparentes ou opacas de acordo com um controle eletrônico. Os óculos obturadores fecham ora a imagem do olho direito, ora a imagem do olho esquerdo, de maneira síncrona com a exibição.

³ Espaço de cor é um modelo abstrato matemático para formalizar a descrição de cores através de tuplas de números, tipicamente formadas por três ou quatro elementos. RGB, YCbCr e CMYK são exemplos de espaços de cor.

com a otimização de recursos de transmissão e de armazenamento. Como exemplo, uma mídia de DVD pode armazenar poucos segundos de vídeo de alta definição sem compressão (Richardson, 2003).

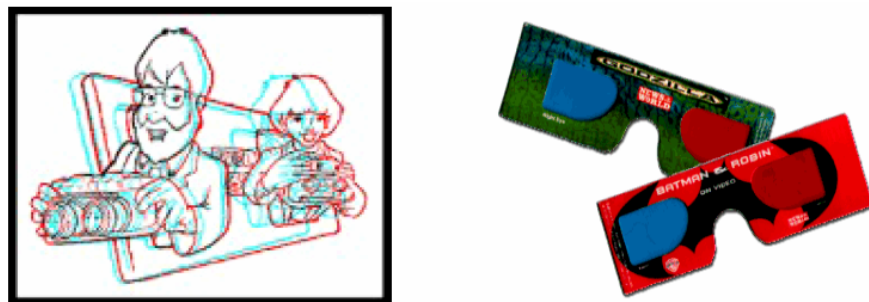


Figura 1.4 – Figura estereoscópica formada pelo processo de cores complementares (uma lente vermelha e outra ciano): a) Anaglifo e b) Óculos utilizados para visão estéereoscópica anáglifica (Wood, 2000)

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 a imagem e principalmente o vídeo digital, e a grande maioria das técnicas de compressão perde parte da informação original visando tamanhos significativamente menores dos arquivos gerados (Richardson, 2003).

Os algoritmos de compressão de vídeo realizam a remoção da redundância temporal, espacial e/ou de domínio da frequência. Na compressão de vídeo espaço-temporal (modelo utilizado em CODECs com grande índice de compressão como MPEG4 e H264), a remoção da redundância é feita em quatro etapas, como mostrado na Figura 1.5. A primeira etapa é a redução das informações contidas no espaço de cor (e como dito anteriormente, uma etapa que compromete a exibição de vídeo estereoscópicos anaglíficos). A segunda está relacionada à aplicação de transformações matemáticas sobre as imagens, o que explicita as redundâncias de informação. O terceiro passo é a quantização das informações resultantes da transformada. A última etapa é a remoção da redundância binária, como o método de compressão LZW, onde não ocorre perda de informação.



Figura 1.5 – Etapas da compressão de vídeo espaço-temporal

Este projeto tem como objetivo propor um novo método de codificação de vídeos estereoscópicos, no qual o armazenamento deverá ser efetuado com compressão significativa de dados

(equivalente às compressões espaço-temporais citadas anteriormente). Além disso, tal método deverá permitir que a reprodução do vídeo possa ser efetuada utilizando diversos tipos de tecnologia, incluindo a visualização anaglífica que é sensível à perda de cores devido aos métodos de compressão atuais.

As próximas seções estão divididas do seguinte modo: o **Capítulo 2** apresenta um breve histórico sobre a estereoscopia, seus fundamentos, a captação de imagens estereoscópicas em movimento além das tecnologias disponíveis atualmente para visualização desses dados; o **Capítulo 3** apresenta conceitos básicos sobre vídeos digitais e seus espaços de representação de cor, sobre codificação de vídeo e armazenamento de vídeos estereoscópicos; o **Capítulo 4** apresenta e discute alguns trabalhos relevantes relacionados a esta pesquisa; o **Capítulo 5** discute os modelos de codificação para vídeos e suas implicações para a visualização de vídeos estereoscópicos, além de apresentar o ambiente de testes de qualidade de CODECs; o **Capítulo 6** apresenta o detalhamento da proposta de projeto, a metodologia a ser seguida para se alcançar os objetivos e o cronograma de atividades proposto.

Estereoscopia

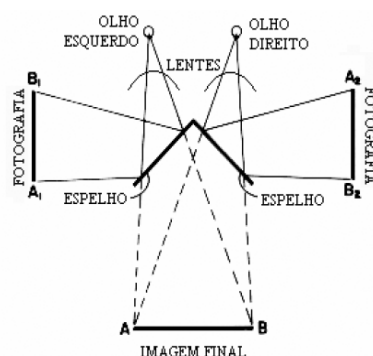
2.1. Considerações Iniciais

A questão da representação do espaço tridimensional foi talvez a mais importante contribuição do período renascentista às artes plásticas. A descoberta do método de representação do espaço tridimensional no espaço bidimensional vem de Filippo Brunelleschi (1387-1446) e Leon Batista Alberti (1404 – 1472), denominada de *perspectiva artificialis*. A técnica de representação teve sua origem em projetos de arquitetura desenvolvidos pelos dois autores. Está ligado à origem da perspectiva um pequeno aparato denominado *tavolleta*, que servia ao desenvolvimento daqueles projetos (Godoy, 2001).

Data de 1838 a patente de um engenho denominado estereoscópio, feita por Sir Charles Wheatstone. O aparato era constituído de um jogo central de espelhos que refletiam dois desenhos de um objeto observado por ângulos ligeiramente diferentes um do outro, de modo que sua observação através dos espelhos permitia a sensação de visão binocular humana. Ou seja, descobre-se nesta época uma outra técnica representacional da tridimensionalidade espacial, utilizando-se de maneira mais avançada as características da visão binocular humana, pois era o cérebro que se encarregava de reconstituir a sensação estereoscópica. A **Figura 2.1** mostra o Estereoscópio e seu funcionamento.



(a)



(b)

Figura 2.1 - Estereoscópio: a) Exemplo de aparelho estereoscópio e b) Esquema básico de um estereoscópio (Kirner & Tori, 2004)

Não tardou para que a fotografia, cuja descoberta data de meados do século XIX, portanto posterior à descoberta de Wheatstone, passasse a utilizar essa técnica para produção de imagens fotográficas estereoscópicas. Tal invenção tornou-se mesmo uma espécie de padrão popular para a fotografia entre os anos 1850 e 1880; e muitos dos ícones fotográficos do século XIX, eram na verdade estereogramas. Cartões fotográficos estereoscópicos, com seus respectivos visores (estereoscópio) eram comercializados e tomaram conta do mercado nas classes média e alta (Adams, 2001). Os estereogramas entraram em declínio comercial a partir do final do século XIX, mas encontraram aplicações científicas na fotogrametria e fotointerpretação (Godoy, 2005).

Na década de 50, com o surgimento da televisão, desencadeou-se o processo histórico do declínio econômico do cinema. As produtoras cinematográficas norte-americanas procuraram formas de atrair a atenção do público, e trazê-lo de volta às salas de exibição. Telas de grande formato (cinemascope), filmes coloridos, e também o cinema em terceira dimensão foram usados nessa reconquista do espaço perdido para a TV.

O surgimento da técnica polaróide 3D, que utilizava filtros polarizadores para a separação das imagens de cada olho, permitiu a visualização de imagens coloridas. Por este processo foram produzidos vários filmes tais como “House of Wax” dirigido por Andre de Toth em 1953, “Creature from the Black Lagoon” dirigido por Jack Arnold em 1954 (a **Figura 2.2** ilustra um quadro deste filme com a tecnologia de exibição anaglífica vermelho-ciano) e “Disque M para Matar” dirigido por Alfred Hitchcock em 1954.



Figura 2.2 – Anaglifo de um quadro do filme “Creature from the Black Lagoon”, de 1956

Outras tentativas surgiram posteriormente, já na década de 70, tais como o “Flesh for Frankenstein” (1973) dirigido por Paul Morrissey apelando para a sexualidade e violência para conquistar algum público.

2.2. Fundamentos

No curso da evolução, alguns animais (inclusive o ser humano) passaram a apresentar olhos posicionados na frente da cabeça. Esses animais perderam o incrível campo visual de, praticamente, 360 graus, proporcionado por olhos laterais e opostos. Por outro lado, eles adquiriram uma nova função: a visão binocular ou estereoscopia, “visão sólida” em grego (Fontoura, 2001).

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. Os eixos visuais dos animais que têm olhos laterais e opostos, obviamente, nunca se cruzam. 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 (Fontoura, 2001).

2.2.1. Visão Binocular Humana

Os animais, em específico os seres humanos, são dotados de dois olhos que enxergam o mundo de forma diferente, pois estão separados por uma distância. As diferenças entre imagens geradas pelo olho direito e pelo olho esquerdo são processadas pelo cérebro nos dando uma noção de profundidade e, com isto, tem-se a idéia de imersão em um ambiente com objetos posicionados a distâncias diferentes.

Podemos dizer que “um visor estereoscópico é um sistema óptico cujo componente final é o cérebro humano” (StereoGraphics 1997). 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 2.3**. São capturadas duas imagens ligeiramente diferentes relativas à mesma cena, da mesma maneira que vemos o mundo real.

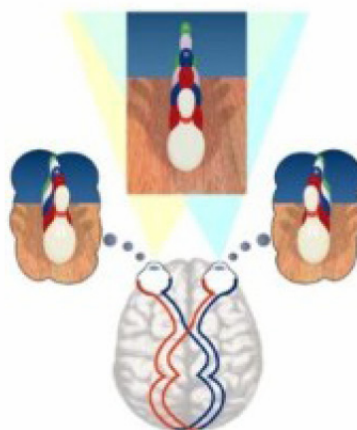


Figura 2.3 – Visões de cada olho da mesma cena (Watson, 1998)

Uma consequência imediata da diferença das imagens capturadas pelos dois olhos é o espaçamento entre o mesmo ponto projetado nas duas retinas, chamado de disparidade na retina. Por exemplo, na **Figura 2.4 (a)**, o olho da esquerda vê a árvore à esquerda do pinheiro, enquanto o olho direito a vê à direita. Estas duas imagens chegam ao cérebro onde é feita uma superposição, resultando na **Figura 2.4 (b)**.

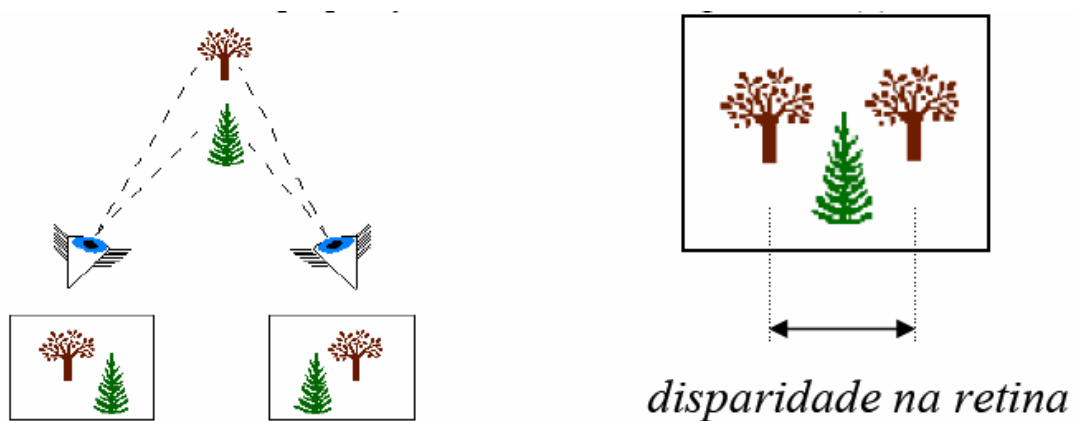


Figura 2.4 – a) Visões da mesma cena pelos dois olhos e b) Superposição das imagens e a disparidade na retina

2.2.2. Fundamentos Matemáticos

Conforme citado anteriormente, existem diferenças entre imagens formadas nas retinas de cada olho quando sobrepostas. Estas diferenças são na **direção horizontal**. A disparidade é zero para objetos onde os olhos convergem. Já a paralaxe é a distância entre os pontos correspondentes das imagens do olho direito e do esquerdo na imagem projetada na tela. Em outras palavras, disparidade e paralaxe são duas entidades similares, com a diferença que **paralaxe é medida na tela do computador** e **disparidade, na retina**. É a paralaxe que produz a disparidade, que por sua vez, produz o estéreo. Os três tipos básicos de paralaxe são:

- **Paralaxe zero:** conhecida como ZPS (do inglês *Zero Parallax Setting*). Um ponto com paralaxe zero se encontra no plano de projeção, tendo a mesma projeção para os dois olhos (**Figura 2.5 (a)**).
- **Paralaxe negativa:** significa que o cruzamento dos raios de projeção para cada olho encontra-se entre os olhos e a tela de projeção, dando a sensação de o **objeto estar saindo da tela** (**Figura 2.5 (b)**).

- **Paralaxe positiva:** o cruzamento dos raios é atrás do plano de projeção, dando a sensação de que o objeto está atrás da tela de projeção (**Figura 2.5 (c)**).

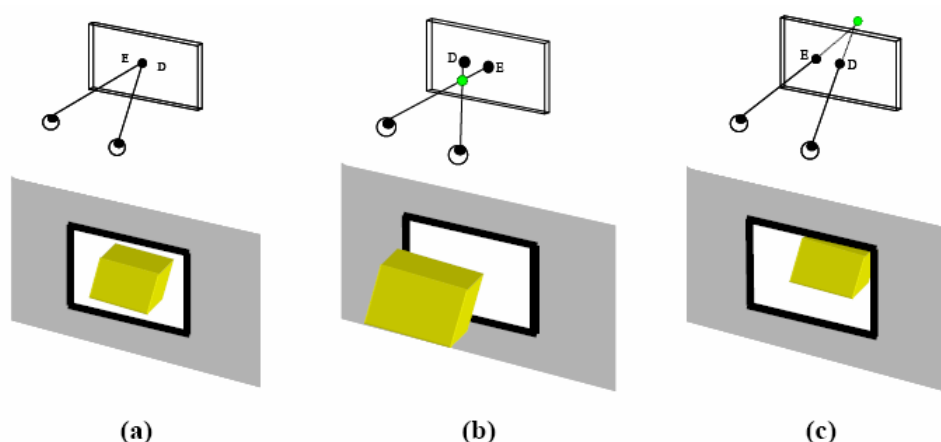



Figura 2.5 – Tipos de paralaxe: a) Paralaxe zero (ZPS), b) Paralaxe negativa e c) Paralaxe positiva

2.2.3. Profundidade

O ponto conjugado de uma imagem é o ponto que corresponde a ele na outra imagem. De posse do par estéreo é possível calcular o ponto conjugado da seguinte forma: Seja $E(i, j)$ um ponto na imagem esquerda e $D(i', j')$ o ponto correspondente a ele na imagem direita, então temos :

$$D = \sum_{i=1}^m \sum_{j=1}^n (I_E(i, j) - I_D(i, j))^2$$

O ponto conjugado é o ponto (i, j) onde D é mínimo. 

Uma vez determinados os parâmetros das câmeras, a posição tridimensional de um ponto na cena, identificado e casado nas imagens estereoscópica, pode ser determinada através da simples triangulação.

Para descrever este cálculo, suporemos que as câmeras estão montadas de modo que seus eixos estejam paralelos e separados por uma distancia b , e que a linha focal é paralela aos planos das imagens, e especificamente às linhas de varredura. Suporemos também que as projeções ortogonais dos focos nos planos das imagens têm coordenadas (x_0, y_0) e que f (a distancia focal) é a distancia do centro da lente ao plano da imagem das duas câmeras. Veja **Figura 2.6** (Macedo, 1995).

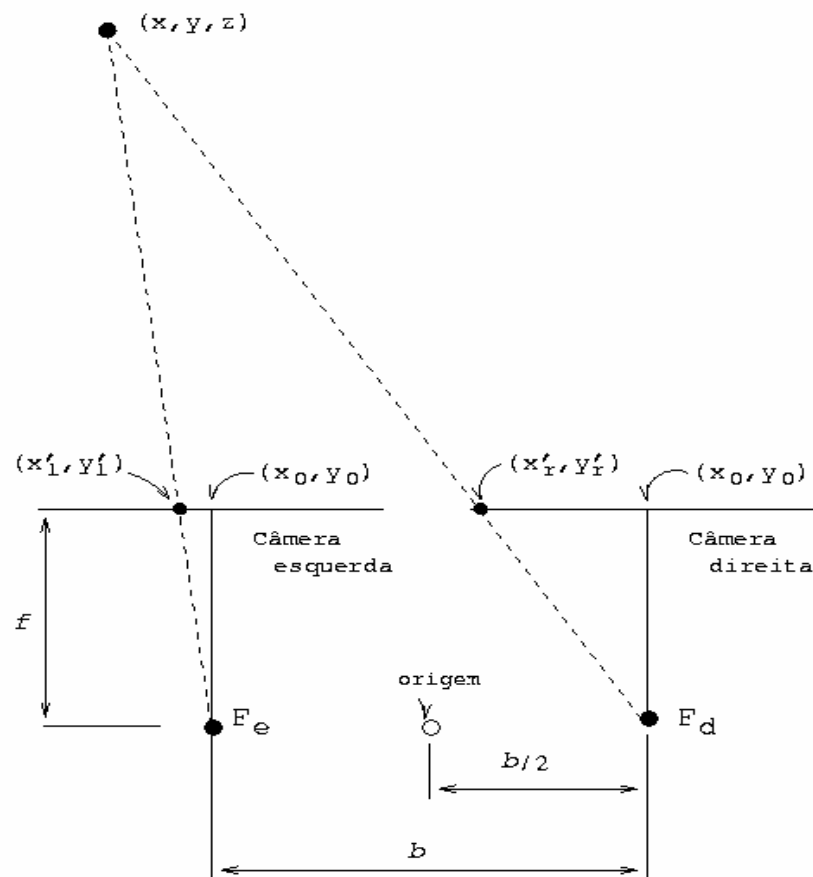


Figura 2.6 – Geometria básica da visão estereoscópica

Sejam X'_L e X'_R as posições horizontais de dois pontos, um em cada imagem, que foram casados na fase de determinação de pontos conjugados. Seja y' a coordenada comum aos dois. Supondo-se que os dois pontos correspondem a um mesmo ponto do objeto, as coordenadas espaciais do mesmo podem ser determinadas pelas fórmulas de estereoscopia:

$$\frac{x'_l - x_0}{f} = \frac{x + b/2}{z} \quad \frac{x'_r - x_0}{f} = \frac{x - b/2}{z} \quad \frac{y'_l - y_0}{f} = \frac{y'_r - y_0}{f} = \frac{y}{z}$$

Rearranjando obtemos as seguintes fórmulas:

$$x = \frac{b}{2} \frac{x'_l + x'_r - 2x_0}{x'_l - x'_r} \quad y = b \frac{y' - y_0}{x'_l - x'_r} \quad z = b \frac{f}{x'_l - x'_r}$$

As coordenadas (x, y, z) calculadas são medidas relativamente a um sistema de coordenadas com eixos paralelos aos eixos das imagens, e cuja origem é situada no ponto médio da linha focal (Macedo,1995).

2.3. Captação Estereoscópica de Imagens em Movimento

Dentre os fatores que implicam na formação da imagem binocular humana é necessário destacar os aspectos referentes à distância inter-ocular e à convergência. A distância média entre olhos é de 65mm. Assim essa deve ser a distância entre as duas objetivas das câmeras (Lipton, 1982).

Considera-se que objetos situados a grandes distâncias perdem sua aparência tridimensional e, portanto, é necessário afastar-se as objetivas para que a aparência do objeto volte a ser tridimensional, apesar de sua grande distância. Essa técnica é utilizada nas fotografias estereoscópicas da superfície terrestre feitas de aviões. Para esses casos, considera-se que a distância inter-ocular deve ter $1/50$ do valor da distância do objeto mais próximo da objetiva. Normalmente com uma distância inter-ocular de 65mm a aparência tridimensional melhor se apresentará em objetos situados a aproximadamente 3 metros de distância das objetivas (Langford, 1990).

Quanto à alteração da convergência das objetivas, esta diz respeito ao enquadramento de objetos situados a curtas distâncias. Neste caso, observa-se a formação de imagens tridimensionais mais adequadas para situações de estúdio onde o fundo neutro não prejudica a perfeita visualização do objeto.

Considerando-se esses aspectos ressalta-se o fato de que as duas câmeras devem ser alinhadas paralelamente na maioria dos casos, mas com a possibilidade de haver modificações na distância interocular. Portanto uma estrutura que dê suporte físico para essas duas câmeras deve permitir o ajuste do posicionamento das câmeras nos três eixos do espaço além de permitir a escolha entre diferentes distâncias inter-oculares. Este aparato permite a utilização de diferentes tipos de câmeras ao invés de se utilizarem apenas câmeras especialmente desenvolvidas para a produção estereoscópica. O método de alinhamento poderá estar utilizando feixes de luz laser que permitirão a precisão das medidas de distância entre objetivas. Esse aparato, além de necessitar precisão necessita também de materiais resistentes que não sofram deformações por ação do peso das câmeras.

2.4. Cinematografia Estereoscópica

Como condição de aprofundamento da fundamentação técnica para a produção das estereoscopias videográficas, é necessário o conhecimento prévio das tecnologias disponíveis para a produção de estereoscopia cinematográfica.

Um dos métodos, o mais antigo deles, é o método anaglífico. A primeira experiência cinematográfica tridimensional utilizando esse método data de 1922 com o filme “Power of Love” (Godoy, 2005). Esta técnica caracteriza-se em colorizar com uma cor primária diferente cada uma das imagens referentes a cada olho (azul e vermelho, ou verde e vermelho), 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, verde ou azul (Mancini, 1998).

O método mais moderno de estereoscopia para projeção é chamado de método de filtragem por luz polarizada, o qual é descrito a seguir: duas câmeras cinematográficas sincronizadas montadas à distância interocular produzem dois filmes referentes respectivamente à visão do olho esquerdo e do olho direito. Em uma tela metalizada é feita projeção sincronizada das duas películas, usando-se filtros polarizadores. O espectador assiste ao filme de óculos, com filtros polarizadores iguais aos da projeção (vide Figura 1.3). Dessa forma as imagens referentes a cada um dos olhos são filtradas pelos filtros polarizadores de modo que cada olho perceba somente a imagem referente à sua lateralidade. Em sistemas mais aprimorados, uma única câmera, através de uma objetiva especial anamórfica, forma duas imagens sobre uma única película; depois, um projetor também equipado com uma objetiva semelhante, reproduz aquela película, gerando duas imagens separadas e polarizadas sobre tela metalizada. Da mesma forma óculos polarizadores são necessários para a separação de cada imagem (Inoue & Ohzu, 1990).

2.5. Exibição de Imagens Estereoscópicas

2.5.1. Monitores e Projetores de Imagens Bidimensionais

A exibição de imagens estereoscópicas exige imagens bem definidas, com um mínimo de ruído (Godoy, 2005). A tecnologia atual à disposição do mercado distingue dois principais tipos de monitoramento de imagens em movimento: a tecnologia CRT (*Catodic Ray Tube*) e a tecnologia LCD (*Liquid Cristal Display*).

A primeira caracteriza-se pelo bombardeio interno de uma tela de vidro por um raio de elétrons, produzido por um canhão eletrônico, que varre a tela produzindo luz com o choque das partículas; é a tecnologia utilizada na maioria dos aparelhos de TV e monitores de vídeo. O tamanho dos pixels é determinado pelo tamanho dos orifícios de uma placa de metal localizada interiormente antes do vidro do monitor. O feixe de elétrons atinge a placa de metal e só o atravessa através desses orifícios. A placa determina o tamanho dos pontos (*dot pitch*) no monitor que estão situados em torno do valor de 0,7mm. Esse tipo de tela se visualizada muito de perto pode parecer muito reticulada e, portanto deve ser observada a certa distância (Mathias & Patterson, 1985).

A outra tecnologia caracteriza-se pela utilização de substâncias classificadas como cristais líquidos dentro de um painel transparente, contidas em células matriciais, com resposta do tipo liga-desliga, que refletem a luz conforme estejam eletrizadas de acordo com as informações dos pixels analógicos ou digitais. Basicamente dois tipos de telas são observadas: as de matriz passiva e as de matriz ativa. Estas últimas têm sido mais utilizadas, devido à velocidade de resposta permitida pelos TFT (*Thin Film Transistor*) localizados em cada cruzamento de malha da matriz e que são capazes de

controlar cada célula com cristal líquido, ou seja, o controle liga-desliga é muito mais preciso feito pixel a pixel. Possuem tela plana, o que evita a distorção da imagem nas bordas. Além disso, esses monitores são muito mais compactos, pois não se utilizam dos tubos de raios catódicos responsáveis pelo aumento de dimensão e peso nos monitores do tipo CRT (Winston, 1996).

No caso dos monitores, a vantagem da tecnologia CRT em relação à de LCD é a luminosidade. A imagem é mais clara e independe em grande medida do ângulo através do qual se visualiza a tela. No caso do LCD os ângulos de visualização são mais estreitos o que pode prejudicar sua utilização em ambientes de maior largura que comprimento. Além disso, telas de maior dimensão são possíveis com o CRT; os sistemas LCD com telas grandes não passam de sistemas de retro-projeção (Hearn & Baker, 2004).

Quanto aos projetores eletrônicos, dois sistemas podem ser encontrados, novamente a tecnologia LCD e a tecnologia DLP (*Digital Light Processing*). No primeiro caso três painéis LCD, são submetidos a três feixes de luz (Vermelha, Verde e Azul) de modo a compor na saída do sistema uma imagem colorida. Neste caso a luminosidade do sistema depende em grande medida da lâmpada utilizada como fonte de luz. Deve ser levado em conta que neste caso grande quantidade de luz é absorvida em função do processo de transparências e de reflexões internas ao sistema. No caso desta tecnologia, a imagem projetada é mais clara, possui maior precisão e saturação de cores, oferecendo uma imagem sem distorção, permitindo projeção em espaços físicos iluminados, sob ângulos de visão mais abertos que a outra tecnologia. Este tipo de tecnologia de projeção vem sendo utilizado em salas de cinema digital de alta definição (Godoy, 2005).

2.5.2. Óculos para Visualização Estereoscópica

Nos métodos de produção de imagens estereoscópicas a partir de projeções ou exibição em monitores de imagens bidimensionais, os óculos representam o dispositivo final para a percepção da tridimensionalidade (Stereographics, 1997). No método estereoscópico anaglífico, são os óculos com lente vermelha/ciano. No método estereoscópico por polarização da luz, são os óculos polarizadores. No método estereoscópico por multiplexação temporal, são os óculos obturadores de cristal líquido (Andrade, 2008).

Do ponto de vista tecnológico, este último aparato é bem mais complexo que os outros dois. Todavia, a complexidade do aparato de visualização vem desonerar a complexidade do sistema de projeção das imagens. Sabe-se que no caso do método da luz polarizada são necessários dois projetores equipados com objetivas polarizadoras; os sistemas de reprodução videográfica têm que estar absolutamente sincronizados para que o efeito tridimensional seja percebido, além de ajustes ópticos de convergência dos dois projetores. Estes fatores são elementos que complicam não somente do processo de exibição, mas também da edição das imagens estereoscópicas (Machado, 1997).

No método que utiliza multiplexação temporal, o observador, ao visualizar a tela do computador ou televisor, deve utilizar óculos especiais, cujas lentes são feitas de cristal líquido. As lentes podem ficar instantaneamente transparentes ou opacas de acordo com um controle eletrônico. Esse controle é sincronizado com sinal de vídeo, de forma a deixar, por exemplo, opaca a lente da esquerda e transparente a da direita quando estiver sendo exibido, na tela, o quadro referente ao olho direito e vice-versa.

O sinal de vídeo deve apresentar, na sequência, as imagens esquerda e direita em sincronismo. Como as taxas de atualização das imagens no vídeo são suficientemente rápidas (60Hz para cada olho), o resultado é que cada olho enxerga uma imagem diferente, resultando no efeito estereoscópico. De forma alternativa, o sinal de vídeo pode ser entrelaçado (linhas pares mostram a imagem esquerda e linhas ímpares mostram a imagem direita), reduzindo o cintilamento da imagem. Existem hoje no mercado diversos óculos obturadores. Os de custo mais elevado são controlados por um emissor de infravermelho que realiza a obturação das lentes dos óculos, enquanto os de baixo custo têm seu controlador conectado ao microcomputador ou à fonte de vídeo e permitem a visualização em estéreo no monitor de vídeo ou na televisão (Santos, 2000). A **Figura 2.7** mostra óculos obturadores sincronizados com infravermelho e por cabo.



**Figura 2.7 – Óculos obturadores sincronizados por infra-vermelho e por cabo
(Stereographics, 1997)**

Esse sistema é capaz de respostas rápidas. Em milésimos de segundos a estrutura do cristal líquido pode transformar-se de modo que para nas durações de 1/60 segundo seja possível alternância entre as duas telas referentes respectivamente ao lado esquerdo e ao lado direito.

2.5.3. Monitores Autoestereoscópicos

Os monitores autoestereoscópicos permitem que um espectador assista a vídeos digitais com profundidade, dispensando o uso dos óculos especiais. Nestes monitores, as visões esquerda e direita são multiplexadas espacialmente, permitindo ao observador visualizar uma imagem tridimensional (Kirner & Tori, 2004). Cada imagem do par estéreo é “fatiada” e reside sobre as colunas pares e

ímpares do monitor. As fatias são direcionadas para o olho do observador por meio de uma película lenticular colocada na superfície do monitor (**Figura 2.8**) ou pelo cálculo de distância e posicionamento dos olhos do observador⁴.

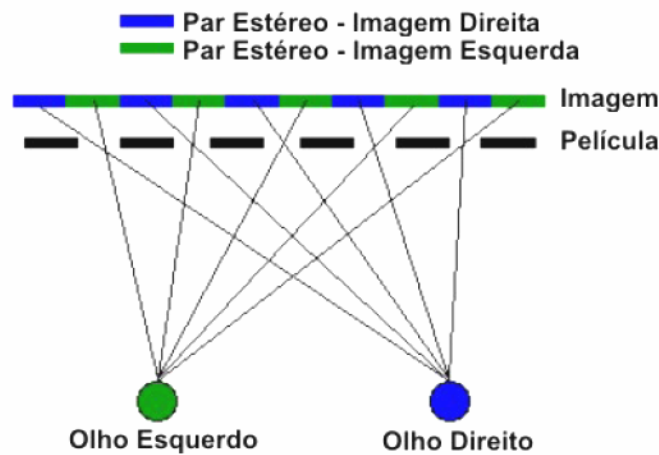


Figura 2.8 – Esquema de funcionamento de um monitor autoestereoscópico

O monitor autoestereoscópico possui uma área de visão delimitada ao espectador. De acordo com o ângulo de visualização (dentro da área permitida), é possível a visualização de uma cena de diferentes posições, como mostrado na **Figura 2.9**.

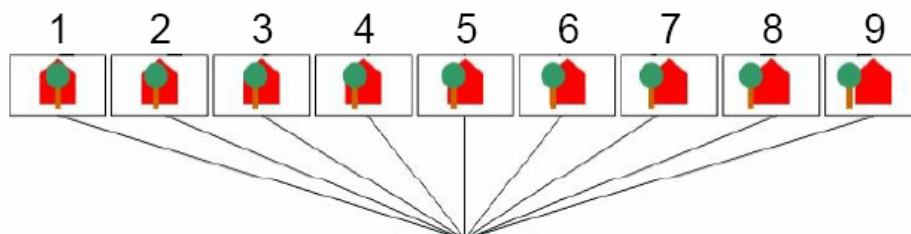


Figura 2.9 – Exemplo de visualização em monitor autoestereoscópico, onde nove posições são possíveis para o espectador que observa a mesma cena por ângulos diferentes.

⁴ Uma explicação detalhada dos componentes pode ser encontrada em:

1. Dodgson, N. A. (1998) Autostereoscopic Image Compression. Rainbow Graphics Researchs Group, Computer Laboratory, University of Cambridge, Pembroke St. Disponível em: <<http://www.cl.cam.ac.uk/~nad/compr/compr.html>>.
2. Perlin, K., Paxia, S. and Kollin, J. (2000) An Autostereoscopic Display; SIGGRAPH 2000 Conference Proceedings. New Orleans, Louisiana. July 23-28. Disponível em: <<http://www.mrl.nyu.edu/publications/autostereo/autostereo.pdf>>.
3. Perlin, K. et. al. (2001) Recent Advances in the NYU Autostereoscopic Display; Proceedings of the SPIE, Vol. 4297. San Jose, California. January 22-24. Disponível em: <<http://www.mrl.nyu.edu/publications/autostereo/spie2001.pdf>>.

O custo dos monitores autoestereoscópicos tem reduzido muito nesta década, permitindo sua fabricação em linha de produções. A empresa Phillips no ano de 2008, lançou uma série de monitores autoestereoscópicos, variando de 22 a 52 polegadas⁵. Com estes monitores jogos 3D são passíveis de utilizar visualização estereoscópica⁶. Para a produção de animações 3D é possível utilizar softwares de modelagem e animação 3D com *plugins* (Maya e 3D Studio). Para a produção de vídeos, é necessária a utilização de um sistema de hardware e software proprietário da Phillips para codificação de um vídeo lado-a-lado ou sobreposto gerando um par de novas imagens, como mostrado na Figura 1.2.

2.6. Considerações Finais

Apesar de diversos aparatos terem sido desenvolvidos nos últimos dois séculos, a visualização de imagens em movimento estereoscópicas ainda não possui uma tecnologia padronizada, e o que dificulta sua disseminação. É importante salientar que quando tecnologias de baixo custo de exibição de fotos estereoscópicas eram utilizadas (em meados de 1880), a disseminação da tecnologia era alta, contrastando com o momento atual, mais de um século depois.

O número de salas de cinema estereoscópicas vêm aumentando no mundo todo (no ano de 2007, uma dessas salas foi instalada na cidade de São Paulo, com a tecnologia de luz polarizada). O mercado cinematográfico tem interesse em que as produções estereoscópicas possam se tornar cada vez mais populares, seguindo o mesmo caminho que o som (que passou de mono a estéreo e depois a *surround*), percorreu nas últimas décadas.

Monitores que permitem a visualização estereoscópica sem o uso de óculos, têm começado a surgir comercialmente, contudo diversas barreiras para a produção de material estereoscópico precisam ser vencidas, como o uso de soluções proprietárias.

Uma solução de baixo custo que poderia dissminar a estereoscopia seria a visualização anaglífica, com os óculos com filtros de cor (como o ciano e vermelho, por exemplo), visualizando vídeos em computadores, TVs de alta definição e projetores digitais. Contudo, os atuais métodos de codificação de vídeo, acabam por diminuir de maneira drástica a qualidade dos vídeos a serem exibidos nesse tipo de dispositivo com os filtros coloridos.

No próximo capítulo, serão apresentados alguns conceitos sobre vídeos digitais que mostram o porquê da perda de qualidade citada.

⁵ Série de monitores Philips WOWvx

⁶ Na década de 90 a tecnologia de visualização por multiplexação temporal (que utiliza de óculos obturadores) chegou ao mercado, nos videogames e nas placas de vídeo de computadores pessoais. A partir dessa década, o número de jogos de computador 3D que permitem visualização estereoscópica, aumentou consideravelmente. Como exemplo, segue uma série de jogos que podem ser visualizados estereoscópicamente: Battlefield 2142, Command & Conquer 3 Tiberium Wars, DiRT, DTM racing 3, FarCry, Need for Speed Carbon, Need for Speed Most Wanted, Rainbow Six Vegas, Tiger Woods PGA TOUR 2007, Tomb Raider Anniversary, Tomb Raider Legend, World of Warcraft, rFactor e Crysis.

Codificação de Vídeo Digital

3.1. Considerações Iniciais

O vídeo digital, e em especial o vídeo digital de alta definição, exige grandes quantidades de bytes de espaço para armazenamento, e uma alta largura de banda para transmissão. Neste sentido, o uso de algoritmos e técnicas de compressão e descompressão torna-se imprescindível para a viabilização do transporte dos mesmos, tanto no que diz respeito à comunicação quanto ao armazenamento dos dados. A sub-amostragem de cor das imagens e o modo como os dados são codificados definem como um CODEC, um codificador e decodificador de vídeo digital, processa seus dados.

Neste capítulo são apresentados os principais espaços de representação de cor para vídeos digitais (RGB e YCrCb) e os modelos utilizados para redução de informação, além de alguns métodos de codificação, que levam em consideração as informações redundantes em imagens separadas e em um determinado conjunto de imagem. Por fim, são apresentados os métodos de armazenamento de dados estereoscópicos nos padrões atuais.

3.2. Espaços de Cor

O elemento básico para a construção de imagens digitais é o pixel (*picture element* – elemento da figura). Uma imagem é uma grade de pixels, normalmente descrita por sua quantidade de elementos verticais e horizontais (como exemplo temos a imagem de uma câmera digital padrão DV, que possui resolução de 720 x 480 pixels). Cada pixel tem um estado, que indica brilho e cor, por exemplo. A quantidade de informações contidas e codificadas em um pixel, tem relação direta com a quantidade de bits que representa o estado de um pixel.

Em imagens branco e preto puras, apenas 1 bit por pixel é suficiente para representar os pixels dessa imagem (atribuindo o valor zero para preto e o valor um para branco, como exemplo). Em imagens com 16 milhões de cores, são necessários 24 bits para cada pixel, sendo que neste tipo de



imagem cada uma das cores básicas vermelho, verde e azul (padrão RGB), possui 8 bits para ser representada. Imagens em escalas de cinza, assim como as componentes de cor do padrão RGB, também possuem 8 bits (1 byte) por *pixel*, que permite uma representação satisfatória à percepção humana de luminosidade (Murray & Willian, 1996). Das três componentes de cor citadas, a cor que o olho humano consegue diferenciar com maior facilidade é o verde (cerca de 250 variações). Com 8 bits por *pixel*, temos valores de 0 a 255 para cada componente, que são suficientes para representar uma quantidade de cores (cerca de 16 milhões) até superior ao que nossos olhos podem diferenciar (Costa, 2007).

O espaço de cor HSB (*Hue, Saturation, Bright*) consiste em combinar Matiz, Saturação e Luminância (sistema HSB – *Hue, Saturation, Bright*). A luminância, também utilizada como sinônimo de brilho é uma unidade de medida da intensidade de uma fonte de luz. Outro padrão utilizado para representar as informações de um pixel, é intitulado de sistema natural de cor ou YCbCr, que separa as informações de brilho da informação de cor. Para a componente Y (brilho ou luminância), são atribuídos valores de 16 a 240. As componentes de cor azul e vermelha em relação a componente verde, Cb e Cr, representam as cores com valores de 16 a 240. Assim como no padrão RGB existem variações, também existem variações do padrão YCbCr. Uma equivalência matemática entre os padrões YCbCr e RGB é mostrada a seguir (Natress & Jordam, 2005):

Do padrão RGB para o YCbCr (aproximadamente):

- $Y = 0.299 R + 0.587 G + 0.114 B$
- $Cb = 0.493 (B - Y)$
- $Cr = 0.877 (R - Y)$

Do padrão YCrCb para o RGB (aproximadamente):

- $Y' = (Y - 16) * 255 / 219$
- $Cb' = (Cb - 128) * 127 / 112$
- $Cr' = (Cr - 128) * 127 / 112$
- $R = Y' + 1.402 * Cr'$
- $G = Y' - 0.344 * Cb' - 0.714 * Cr'$
- $B = Y' + 1.772 * Cb'$

3.2. Amostragem de Cor

Os dispositivos que capturam a imagem de uma câmera (como os CCDs e os CMOS), realizam a aquisição de dados por filtros vermelhos, verdes e azuis. O fósforo em um monitor CRT e os filtros em um monitor LCD ou um projetor de imagens DLP também utiliza filtros com as mesmas cores.

No entanto, o vídeo digital armazenado e manipulado é separado em componentes de luminância e croma, no padrão YCbCr, porque muitas vezes é realizada uma redução de dados nas amostras das componentes de cor Cb e Cr, diminuindo assim o volume de informações a serem processadas.

A seguir são mostrados alguns modelos de amostragem de cor, sendo que o primeiro número representará a luminância e os outros dois a croma (Natress & Jordam, 2005).

3.2.1. Modelo 4:4:4

Neste modelo temos resolução completa, pois temos para cada pixel de luminância dois valores correspondentes de croma (um de Cb e outro de Cr). Este modelo tem correspondência direta com o modelo de representação de imagens RGB, sendo necessária uma unidade de armazenamento com grande velocidade de gravação para conseguir captar vídeos com esse padrão (Supia, 2007). Câmeras de alta definição que utilizam esse formato gravam diretamente em unidades de disco de alta desempenho. A **Figura 3.1** mostra o modelo de amostragem 4:4:4 com a representação YCbCr.

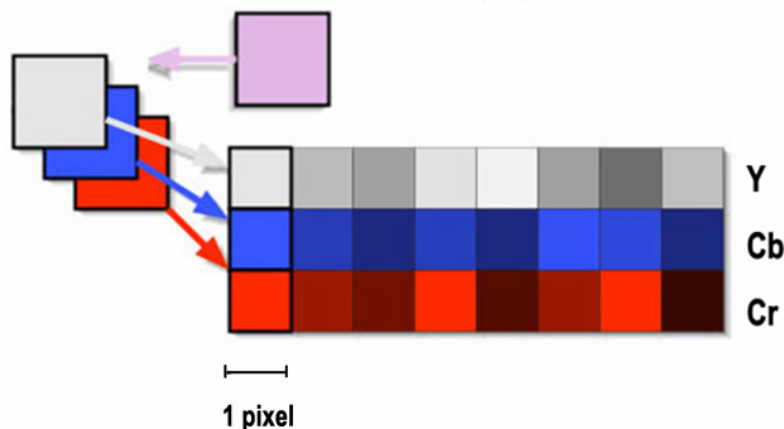


Figura 3.1: Modelo de amostragem de cor 4:4:4

3.2.2. Modelo 4:2:2

Neste modelo as componentes de cor são reduzidas, fazendo com que para cada dois pixels de luminância, sejam associados apenas dois valores de croma (um de Cb e outro de Cr). Isto faz com que este modelo ocupe menos espaço em disco que o 4:4:4, contudo metade da resolução de cores é perdida. Este é o padrão convencional para difusão, utilizado pelos padrões de vídeo DigiBeta e DVCPPro50 (Wilt, 2006). A **Figura 3.2** mostra o modelo de amostragem 4:2:2 com a representação YCbCr.

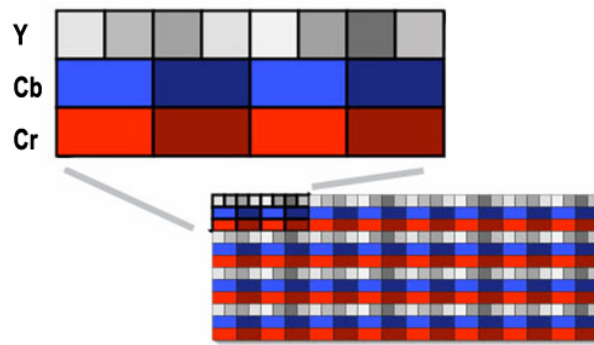


Figura 3.2: Modelo de amostragem de cor 4:2:2

3.2.3. Modelo 4:1:1

Neste modelo as componentes de cor são bastante reduzidas, fazendo com que para cada quatro pixels de luminância, sejam associados apenas dois valores de crominância (um de Cb e outro de Cr). Isto faz com que este modelo ocupe menos espaço em disco do que os dois anteriores, contudo 3/4 da resolução de cores é perdida. Este é o padrão da maioria das câmeras digitais, utilizado pelos padrões de vídeo NTSC DV e PAL DVCPro (geralmente o armazenamento é feito em fita magnética) (Wilt, 2006). A Figura 3.3 mostra o modelo de amostragem 4:1:1 com a representação YCbCr.

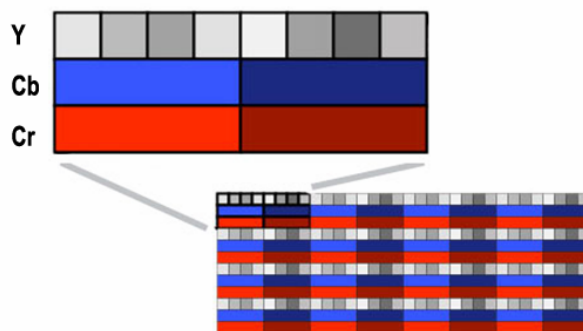


Figura 3.3: Modelo de amostragem de cor 4:1:1

3.2.4. Modelo 4:2:0

O modelo 4:2:0 é o modelo de amostragem mais complexo, com muitas variações (como o vídeo ser entrelaçado ou não). Neste modelo novamente as componentes de cor são reduzidas, fazendo com que para cada dois Y, sejam associados, alternadamente, um de Cb e outro de Cr. Isto faz com que este modelo ocupe espaço em disco equivalente ao modelo 4:2:2, contudo metade da resolução de cores é perdida. Este é o padrão utilizado na autoria de DVDs de vídeo, pelos padrões PAL DV (Wilt, 2006) e MPEG2. A **Figura 3.4** mostra o modelo de amostragem 4:2:2 com a representação YCbCr.

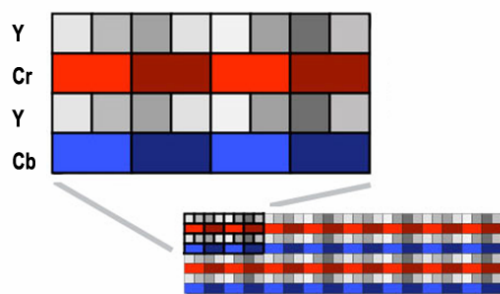


Figura 3.4: Modelo de amostragem de cor 4:2:0

3.3. Formatos de Imagem e Vídeo, Contêineres e CODECs

Um formato de imagem ou de vídeo é um padrão para organizar e armazenar dados de imagens digitais (no caso do vídeo uma sequência de imagens) e define como os dados estão organizados e o tipo de compressão – se existir – é utilizado. Um contêiner de imagens ou de vídeo é similar a um formato de arquivo, mas pode armazenar tipos de imagens ou de vídeo diferentes. Os padrões de compressão definem processos de compressão e descompressão de imagens e de vídeo – com o objetivo de reduzir a quantidade de dados necessária para representar uma imagem ou uma sequência de imagens.

A tabela 3.1 a seguir mostra importantes padrões de imagens e de vídeo em uso, agrupados pelo tipo de informação que contém (Gonzalez & Woods, 2008).

Tabela 3.1: Padrões de Compressão de Imagens e de Vídeo.		
Nome	Organização	Descrição
<i>Imagens Still Binárias</i>		
CCITT Grupo 3	ITU-A	Destinado como um método de facsimile (FAX) para transmissão binária de documentos sobre linhas telefônicas. Suporta codificação <i>run-length</i> 1-D e 2-D e codificação por Huffman.
CCITT Grupo 4	ITU-T	Uma versão simplificada e otimizada para <i>streams</i> do padrão CCITT Grupo 3 e suporta apenas codificação <i>run-length</i> 2-D
JBIG ou JBIG1	ISO/IEC/ITU-T	A sigla JBIG (<i>Joint Bi-level Image Experts Group</i>) define um padrão progressive, sem perda e de dois níveis. Tons contínuos de imagem de até 6 bits/pixel podem ser codificados em base de bits plana. Codificação aritmética é usada em uma imagem inicial de baixa resolução, que pode ser gradualmente melhorada com adição de dados comprimidos.
JBIG2	ISO/IEC/ITU-T	Uma evolução do JBIG1 para imagens de dois níveis para aplicações desktop, Internet e de FAX. O método de compressão usado é baseado em contexto, com métodos baseado em dicionário para texto e Huffman e codificação aritmética para o contexto da imagem. Pode ser configurado com ou sem perda.
<i>Imagens de Tons Contínuos</i>		
GIF	Compuserve	A sigla GIF (<i>Graphics Interchange Format</i>) define um formato sem perda que utilize compressão LZW para imagens de 8 bits por pixel.
JPEG	ISO/IEC/ITU-T	A sigla JPEG (<i>Joint Photographic Experts Group</i>) define um padrão para imagens fotográficas. É uma compressão com perda, e utiliza uma transformada discreta de cosseno (DCT – <i>Discrete Cosine Transform</i>) em blocos de imagem 8x8, Huffman e codificação <i>run-length</i> . É um dos mais

JPEG-LS	ISO/IEC/ITU-T	populares métodos de compressão de imagem para a Internet. Uma codificação que pode ser ajustada para não ter perda ou uma pequena perda para imagens de tom contínuo com predição adaptativa, modelagem de contexto e codificação Golomb.
JPEG-2000	ISO/IEC/ITU-T	Uma evolução do padrão JPEG, desenvolvido para aumentar a compressão de imagens de qualidade fotográfica. Codificação aritmética e uma quantização com transformada wavelet discreta (DWT – <i>Discret Wavelet Transform</i>) são utilizadas. A compressão pode ser com ou sem perda.
PNG	Consórcio World Wide Web (W3C)	A sigla (<i>Portable Network Graphics</i> – Gráficos portáteis para Redes) define um padrão sem perda para imagens com todas as cores e transparência (até 48 bits/pixel) pela codificação das diferenças entre cada valor do pixel e um valor calculado em pixels anteriores.
TIFF	Aldus	Um contêiner que pode suportar compressão sem perda, JBIG2, JPEG, JPEG2000, JPEG-LS, e outros padrões de compressão.
<i>Vídeo</i>		
DV	IEC	<i>Digital Video</i> ou Vídeo Digital. Um padrão desenvolvido para aplicações de vídeo caseiro e semi-profissional. Os quadros são comprimidos independentemente, utilizando uma abordagem baseada em DCTs, similar ao JPEG.
H.261	ITU-T	Um padrão de videoconferência para linhas ISDN. Ele suporta quadros não-intrelaçados de 352x288 e 176x144 de resolução, chamados CIF (<i>Common Intermediate Format</i> – Formato Comum Intermediário) e QCIF (<i>Quarter CIF</i> – Um quarto de CIF), respectivamente. Uma transformada DCT similar ao JPEG é utilizada como redução espacial de redundância. Uma técnica baseada em blocos é utilizada para compensar o movimento entre os frames.
H.262	ITU-T	Veja MPEG-2 a seguir.
H.263	ITU-T	Uma versão melhorada do H.261, desenvolvida para uso em modems telefônicos (isto é, linha de 28.8 Kb/s) com resoluções adicionais: SQCIF (128 x 96), 4CIF (704x576) e 16CIF (1408x512).
H.264	ITU-T	Uma extensão do H.263 para videoconferência, <i>streaming</i> via Internet e para <i>broadcasting</i> de televisão. Ele suporta predição das diferenças entre os quadros, DCT e contexto adaptativo por codificação aritmética.
HDV	Consórcio de companhias	A sigla HDV (<i>High Definition Video</i> – Vídeo de Alta Definição) define um padrão que é extensão do DV para televisões de alta definição, e utiliza uma compressão próxima ao MPEG-2, incluindo redundância temporal.
M-JPEG	Várias companhias	<i>Motion JPEG</i> ou JPEG em movimento é um formato em que cada frame é comprimido de forma independente, utilizando JPEG.
M-JPEG2000	Várias companhias	Motion JPEG2000 ou JPEG2000 em movimento é um formato em que cada frame é comprimido de forma independente, utilizando JPEG2000.
MPEG-1	ISO/IEC	A sigla MPEG-1 (<i>Motion Pictures Expert Group</i>) define o padrão para aplicações em CD-ROM com vídeo não intrelaçado até 1.5 Mb/s. É similar ao H.261 mas as predição dos quadros pode ser baseada no quadro anterior, posterior ou em uma interpolação de ambos. É suportada pela maioria dos computadores atuais e tocadores de DVD.
MPEG-2	ISO/IEC	Uma extensão do MPEG-1, desenvolvido para DVDs com taxas de transferência superiores a 1.5 Mb/s. Suporta vídeo intrelaçado e HDTV.
MPEG-4	ISO/IEC	Uma extensão do MPEG-2 que suporta tamanhos de bloco variáveis e compensação de movimento.

Um contêiner de vídeo possui um arquivo de vídeo com compressão ou sem compressão. Se ele possuir compressão, o contêiner irá guardar os dados de acordo com um CODEC. O mais popular

contêiner de vídeo é o AVI (*Áudio Video Interleaved*) (OpenDML, 1997). Este contêiner foi desenvolvido inicialmente para sistema operacional *Windows* e suas variações, e hoje é um dos mais simples e versáteis padrões para armazenar vídeo, permitindo que um CODEC instalado no sistema operacional possa manipular arquivos de vídeo nesse formato.

Outro contêiner que é muito difundido é o MOV (Apple, 2007), padrão da *Apple* para armazenar vídeos digitais, como DV, H.264, MPEG-1, MPEG-2 entre outros padrões de codificação.

Quando um arquivo AVI ou MOV não possui um CODEC associado, ele disponibiliza seus dados sem qualquer compressão.

3.3.1 CODECs

Vídeos digitais quando não possuem compressão, exigem grandes quantidades de *bytes* de espaço para armazenamento, e uma alta largura de banda para transmissão. Nesse sentido, o uso de algoritmos e técnicas de compressão e descompressão torna-se imprescindível para a viabilização do transporte de vídeo digital, tanto no que diz respeito à comunicação quanto ao armazenamento dos dados. Se considerarmos um vídeo padrão com resolução de 720x480, com 24 bits por *pixel* e aproximadamente 30 quadros por segundo sem nenhuma compressão necessitaremos de uma largura de banda de 248 Mbits/s, e com cerca de 20 minutos teremos mais de 1 GB de dados para serem armazenados.

De fato não é necessário armazenar cada um dos *pixels* de uma imagem. Existem alguns processos que reduzem o número de bits necessários para representação de uma imagem, e depois recuperam essa informação para exibição, sendo intitulados CODECs (CODificadores-DECodificadores). Os CODECs podem alterar as informações na codificação e neste caso, o sistema é dito com perda (*lossy*). Se a entrada e a saída são idênticas, o sistema é dito ser sem perda (*lossless*) (Furht, 1995). Os sistemas *lossy* são importantes, uma vez que aceitar a perda de um pequeno volume de informações pode oferecer uma grande vantagem em termos da taxa de compactação possível.

3.3.2. CODECS Espaciais

Um CODEC espacial é um algoritmo computacional que busca informações redundantes dentro de um quadro do vídeo, comprime a informação computacional do quadro (com perda ou sem perda de qualidade) e posteriormente decodifica essa informação para poder exibir o quadro.

Dois formatos conhecidos de armazenamento de imagens digitais são o GIF (*Graphics Interchange Format* – desenvolvido em 1987 pela empresa CompuServe⁷) e o JPEG (proposto pelo comitê NA –

⁷ <http://www.compuserver.com>

Joint Photographers Expert Group). Os arquivos codificados com o formato GIF utilizam um algoritmo que permite recuperar a imagem com a mesma qualidade da imagem original (sem perda), enquanto nos arquivos de formato JPEG é possível selecionar a qualidade da imagem resultante do processo de codificação (com ou sem perda).

O algoritmo utilizado para codificar com o formato GIF é baseado no algoritmo LZW, que utiliza códigos de comprimento variável para definir trechos redundantes.

A compressão LZW tem suas raízes no trabalho de Jacob Ziv e Abraham Lempel. Em 1977, eles publicaram um artigo sobre compressão de “janelas-deslizantes”, e seguiram com outro artigo em 1978 sobre compressão baseada em “dicionários” (Ziv et al, 1978). Esses algoritmos foram chamados de LZ77 e LZ78, respectivamente. Em 1984, Terry Welch fez uma modificação no LZ78, acrescentando o “W” ao nome desse popular algoritmo de compressão (Welch, 1984). O algoritmo LZW é de fácil compreensão e apresenta excelentes resultados se comparado a outros algoritmos, como seus antecessores (LZ77 e LZ78) ou Huffman (Nelson, 1988). A compressão LZW não apresenta perda de dados (*lossles*).

Os primeiros 256 códigos do dicionário são padronizados, formando um conjunto padrão de caracteres (0 até 255), enquanto os códigos de 256 até 4095 se referem a sub-cadeias existentes no *stream* que estiver sendo processado. Assim, o código gerado pelo algoritmo LZW pode possuir tamanho variado, mas sempre será maior do que o tamanho de um caractere (1 *byte*), do código padrão. O algoritmo de compressão LZW é relativamente simples como mostra a **Figura 3.5**.

```
string = caractere de entrada do stream
ENQUANTO existe string FAÇA
    caractere = próximo caractere de entrada do stream
    SE string + caractere está no dicionário FAÇA
        string = string + caractere
    SENÃO
        gere o código para string
        adicione string + caractere para o dicionário
        string = caractere
    FIM SE
FIM ENQUANTO
gere o código para string
```

Figura 3.5: Algoritmo de codificação LZW

O algoritmo utilizado pelo formato JPEG, aplica uma transformada discreta de cossenos (DCT) em blocos de imagens, e posteriormente um esquema de quantização que elimina altas frequências (Thakrar, 1996). O formato JPEG utiliza 24 bits por *pixel*, sendo um dos melhores formatos para imagens com essa representação.

Há muitos parâmetros para o processo de compressão JPEG. Pelo ajuste desses parâmetros, pode-se obter imagens com grande compressão em detrimento de uma boa qualidade, e vice-versa. Por

exemplo, pode-se obter uma compressão 100:1 de uma imagem original de 24 bits, perdendo muita qualidade, ou uma compressão 3:1, com perda indistinguível.

Geralmente, utiliza-se uma compressão de 10:1 a 20:1 (Wallace, 1991). O padrão JPEG é baseado em um algoritmo com perda (*lossy*), além de extensões opcionais para codificação progressiva e hierárquica. Existe também um modo de compressão sem perda, que gera, tipicamente, uma compressão 2:1. Um efeito notado em arquivos submetidos à alta compressão é o “borramento” das bordas entre elementos distintos da imagem (Steinmetz, 1995).

O padrão JPEG foi dividido em duas partes. A parte 1 especifica os modos de operação, os formatos e o codificador especificado para esses modos. A parte 2 do padrão descreve os testes que determinam se a implementação do codificador ou decodificador está em conformidade com o padrão, para garantir interoperabilidade de sistemas. O esboço do algoritmo de compressão JPEG é descrito a seguir (Wallace, 1991), sendo que o algoritmo de descompressão realiza o processo inverso:

1. Transformar a imagem dentro de uma faixa de cor adequada (caso seja colorida), geralmente da codificação RGB para a codificação YCbCr, o que permite uma perda de informação nos componentes de crominância e luminância a que o olho humano não é sensível;
2. (Opcional) Transformar cada componente pela mediana de grupos de *pixels*, sendo o componente de luminância deixado em resolução completa, enquanto componentes de crominância reduzidos em uma razão de 2:1 horizontalmente e 2:1 ou 1:1 verticalmente, reduzindo o volume de dados pela metade ou em um terço, o que, em termos numéricos, tem grande perda (*loosely*), mas para a maioria das imagens isto não gera grande impacto na qualidade;
3. Agrupar valores de *pixels* para cada componente dentro de blocos 8x8, sendo cada bloco submetido a uma transformada discreta de cossenos (DCT), que é um tipo de transformada de *Fourier*, gerando um mapa de frequências com componentes 8x8, podendo-se descartar informações de alta-frequência sem afetar informações de baixa frequência;
4. Em cada bloco, dividir os 64 componentes de frequência por um “coeficiente de quantização” e arredondar os resultados em função de números inteiros, fundamental para descarte de dados;
5. Codificar os coeficientes reduzidos usando codificação de Huffman ou aritmética, não afetando a qualidade da imagem (*lossless*);
6. Adicionar *headers* com dados da compactação e gerar o arquivo final.

Além de imagens estáticas, esse padrão pode ser aplicado para codificações de vídeo. O MJPEG (Motion JPEG) é um CODEC Espacial que utiliza o algoritmo de compressão JPEG codificando vídeos como seqüências de imagens (Gonzalez & Woods, 2008).

3.3.3. CODECS Espaço-Temporais

Um CODEC espaço-temporal é um algoritmo computacional que busca informações redundantes dentro de um quadro do vídeo e entre quadros consecutivos, comprime a informação redundante desses quadros (com perda ou sem perda de qualidade) e posteriormente decodifica essa informação para poder exibir os quadros. Assim como nos CODECs espaciais, redundâncias presentes no mesmo quadro podem ser eliminados por um método de redundância binária ou por transformação matemática (com posterior eliminação de informações redundantes binárias). Além dessa compressão, uma segunda é realizada, baseada nas semelhanças de informações de quadros posteriores. Dessa forma é possível obter compressão espacial (do mesmo quadro) e através do tempo de sucessão dos quadros (entre quadros sucessivos).

Os modelos que realizam a compressão espacial exploram assimilaridades entre os quadros vizinhos, sejam eles anteriores ou posteriores, visando à eliminação dessas informações. Basicamente procuram-se as informações repetidas presentes em diversos quadros para referenciar apenas um desses quadros e eliminar a codificação da informação nos demais diminuindo a quantidade de informação a ser codificada no *bitstream* final. Dessa forma na codificação espaço-temporal utilizam-se 3 formas de compressão (Richardson, 2003):

- Codificação de menor resolução para os canais de cor;
- Remoção da redundância temporal pela compensação de movimento;
- Remoção da redundância espacial pelas técnicas *Lossy* e *Lossless*.

Como exemplos, a seguir serão analisados alguns padrões de compressão espaço-temporais, criados pelo grupo MPEG (*Motion Picture Expert Group*). Para efeito de redução de informação temporal os padrões MPEG utilizam um mecanismo de diferenciação na codificação de seus quadros visando a maior compressão. Os quadros de um formato típico MPEG são (NA, 1997):

- **Quadro I** (*Intra coded picture*): É um quadro onde apenas a compressão espacial é realizada. Esses quadros aparecem no *bitstream* final em intervalos regulares.
- **Quadro P** (*Predictive coded picture*): codificação preditiva de compensação de movimento entre um quadro I ou P.
- **Quadro B** (*Bidirectionally predictive coded picture*): codificado a partir de I e P, maior compressão de todos os quadros.

O quadro I é codificado independentemente dos demais apresentando apenas a compressão espacial (como no caso de imagens JPEG), alcançando taxas de compressão na ordem de 10:1 e 20:1. Já os quadros P têm a função preditiva sempre aparecendo entre um quadro I ou P, alcançando taxas de compressão na ordem de 20:1 a 30:1. Finalmente o quadro B é codificado a partir da interpolação de um quadro I ou P anterior e um quadro I ou P, alcançando taxas de compressão da ordem de 30:1 a 50:1. Na **Figura 3.6** podemos visualizar a disposição dos tipos de quadros em um vídeo típico MPEG.

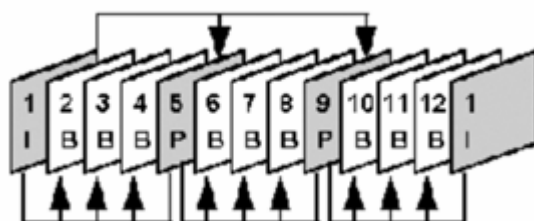


Figura 3.6: Tipos de quadros dos padrões MPEG. Os quadros I são comprimidos internamente, enquanto os quadros P e B são calculados a partir do quadro I.

Ao considerarmos um vídeo, temos em um mesmo plano (seqüência de imagens onde apenas um movimento de câmera é realizado) muitas imagens similares. Em um plano é permitido codificar um quadro através do cálculo do vetor de deslocamento em relação à imagem anterior. No caso do MPEG as imagens são separadas em macroblocos de 16 x 16 pixels e apenas um vetor de movimento é estimado, codificado e transmitido para cada um desses blocos. O erro de estimativa, ou seja, a diferença entre um quadro e sua estimativa também é transmitido. A codificação de quadros P de cada macrobloco alvo é comparado pixel a pixel com o macrobloco correspondente no quadro de referência. Quando a busca encontra macroblocos iguais, dois parâmetros são codificados (Pereira, 1999):

- Vetor de movimento: indica o *offset* (deslocamento) do macrobloco.
- Erro de predição: Três matrizes, uma para cada componente YCrCb contendo diferenças de valores dos pixels do macrobloco e retornando após a busca.

No caso do quadro P elimina-se a redundância espacial e temporal a partir de quadros passados, conforme visto na **Figura 3.7**. Já nos quadros B, é estimado primeiro o vetor de movimento e as matrizes de diferenças usando-se o quadro P ou I anterior. Depois são estimados os mesmos parâmetros usando-se o macrobloco alvo e a média dos valores previstos nos dois passos anteriores. O conjunto com os menores valores é escolhido para ser codificado (Richardson, 2003). Neste tipo de quadro elimina-se a redundância espacial e temporal a partir de quadros passados e futuros. Na codificação dos quadros P e B são gerados vetores de movimento e aplicadas técnicas de codificação

por diferenças e Huffman. Também são geradas matrizes de diferenças que sofrerão compressão DCT, quantização e codificação por entropia.

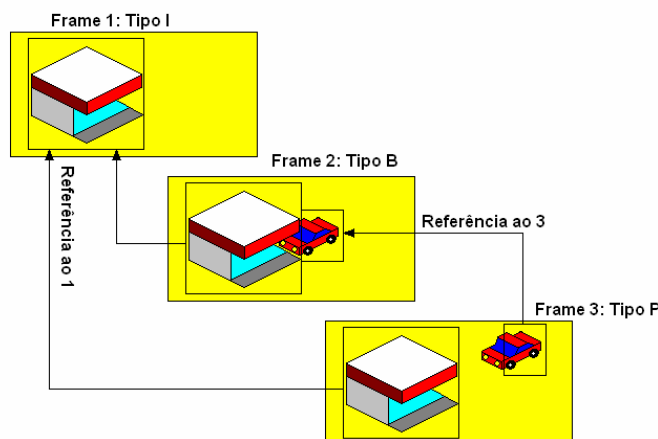


Figura 3.7: Compensação de movimento da codificação MPEG

O primeiro padrão desenvolvido pelo grupo MPEG, intitulado MPEG-1 (NA, 1996), combinou sinais audiovisuais a uma taxa de 1,5Mb/s. A motivação pelo seu desenvolvimento estava vinculado ao fato dos computadores da época (da década de 1990), poder armazenar e reproduzir arquivos de áudio e vídeo em CDs com qualidade equivalente ao armazenado em fitas cassete VHS. Foi finalizado em 1991, possuindo espaço de cores YCbCr (4:2:0) com as seguintes resoluções: 352 x 240 pixels a 30 fps (padrão NTSC) e 352 x 288 pixels a 25fps (padrão PAL). O MPEG-1 foi um padrão inovador, e sua repercussão fez com que várias empresas que até então possuíam departamentos de áudio e vídeo independentes se reorganizassem.

O padrão MPEG-2 manipula a codificação genérica de TV digital e sinais de HDTV (*High Definition TV*) (NA, 1997). Esse padrão foi finalizado em 1994, tendo como objetivo a transmissão em broadcast para TV. Foi o primeiro padrão MPEG a ser dividido em perfis (*profiles*) e níveis (*levels*) para o vídeo. O perfil define a escalabilidade da palavra de bits e resolução do espaço de cores (4:2:2 e 4:2:0). O nível define a resolução da imagem e a máxima taxa de bits por perfil. A taxa de transmissão do padrão MPEG-2 está entre 4 e 9 Mb/s.

O MPEG-2 oferece o padrão de compressão apropriado para redes de televisão, ou seja, desde a captação até a transmissão, passando obviamente pela pós-produção e pela distribuição. Os diferentes perfis e níveis variam consoante a resolução que se pretende obter. Isto é, no número de linhas, de pixels, da taxa de amostragem, e do número de quadros por segundo.

O padrão MPEG-4 foi desenvolvido pela ISO/IEC em 1998, e possui uma enorme lista de funcionalidades que foram criadas após o MPEG-2, visando beneficiar tanto usuários finais quanto criadores e autoradores de vídeo. O MPEG-4 apresenta uma forma diferente de codificação comparada aos padrões anteriores, que utilizam codificação linear de áudio e de vídeo. A codificação MPEG-4 é

baseada em objetos, isto é, as cenas audiovisuais são codificadas em termos de objetos. Um objeto pode ser uma imagem, um vídeo ou um som. A associação de um áudio e um vídeo é chamado de objeto audiovisual.

Basicamente as ferramentas da parte visual do MPEG-4 utilizam compressão de macroblocos baseada em DCT (MPEG-1/2) e inova na compressão temporal diferenciando a compensação de movimento baseada nos objetos da cena (Ebrahimi, 2002).

3.4. Vídeos Estereoscópicos

3.4.1. Vídeos Estereoscópicos Visualizados com Óculos Polarizadores

No que se refere à imagem eletrônica (Evans et al, 1995), um dos métodos para produção do vídeo estereoscópico é constituído pelo que se segue: as imagens são produzidas por duas câmeras de vídeo, os dois sinais de vídeo são gravados em dois dispositivos distintos (fitas magnéticas, discos, entre outros).

Os sinais de vídeo são reproduzidos em dispositivos independentes e sincronizados, cujas imagens são exibidas em dois monitores posicionados segundo ângulo de 90°, equipados com filtros polarizadores. Um espelho especial, localizado no caminho da luz emitida pelos monitores funde as duas imagens e o espectador assiste ao vídeo com óculos que possuem lentes que polarizam a luz de cada um dos projetores. Os monitores podem ser substituídos por dois projetores de vídeo também equipados com filtros polarizadores, e exibidos em uma tela metalizada.

3.4.2. Vídeos Estereoscópicos Visualizados com Óculos Obturadores

No método denominado estereoscopia por multiplexação temporal, um único projetor ou monitor de vídeo mostra uma imagem que apresenta a cada 1/60 seg, referente à visão de cada olho (Tommaselli, 2007). De acordo com a taxa de atualização da imagem no monitor, a fração de segundos pode variar, como no caso do sistema europeu PAL, onde a fração seria 1/50 seg.

O sistema de imagem eletrônica NTSC (*National Television Standard Committee*) produz imagens coloridas na frequência de aproximadamente 30 quadros (*frames*) por segundo (fps), com uma resolução aproximada de 480 linhas horizontais. Os quadros são divididos em dois campos (*fields*) de 240 linhas com uma duração de metade da duração do quadro, aproximadamente 1/60s (StereoGraphics, 1997).

No processo de exibição da imagem, uma TV ou monitor entrelaçado, apresenta inicialmente as 240 linhas ímpares (campo 1) e posteriormente as 240 linhas pares (campo 2). Este processo de formação de imagem videográfica é conhecido como vídeo entrelaçado (*interlaced video*). É interessante salientar que as novas tecnologias digitais utilizam outros métodos de produção de imagens, baseadas

no que se convencionou chamar de vídeo de varredura progressiva (*progressive scan*). Neste caso os quadros são apresentados inteiros sem a formação de artefatos de imagem resultantes do entrelaçamento. Esta é a forma como podem ser exibidos filmes na tela do computador, projetores de imagens computacionais, ou ainda, monitores de televisão digital (Mancini, 1998).

No método denominado estereoscopia ativa, a imagem é obtida por sequencialização dos campos do sinal de vídeo gerado por duas câmeras, assim, o projetor ou monitor de vídeo mostra um único sinal de vídeo que apresenta, a cada 1/60s (no sistema NTSC), um campo com a imagem referente à visão de cada olho. Como se sabe, o olho humano não tem capacidade de discernir dois eventos luminosos consecutivos, ocorridos a intervalos menores que 1/5 de segundo⁸. É por este motivo, que no cinema, imagens projetadas com duração de 1/24s (menores que 1/5s), são entendidas como contínuas; ou ainda na televisão, quadros com a duração de aproximadamente 1/30s são interpretadas como contínuas. Todavia cada um desses quadros é composto por dois subquadros denominados campos, com a duração de 1/60s. Desta forma, é possível simular-se a visão estereoscópica através da exibição para cada olho, de subquadros/campos com a duração de 1/60s. Ou seja, a cada 1/30s projetam-se duas imagens com a duração de 1/60s, uma para cada olho, referentes à sua lateralidade específica. Este método de sequencialização produz um tipo de vídeo estereoscópico que é conhecido por Vídeo Estereoscópico do tipo campo-sequencial (StereoGraphics, 1997), que pode ser observada na **Figura 3.8**.

⁸ O conceito de persistência retiniana é conhecido desde o Antigo Egito e apesar dos trabalhos desenvolvidos por Isaac Newton e o Cavaleiro d'Arcy, só em 1824 é que Peter Mark Roget definiu-o satisfatoriamente como a capacidade que a retina possui para reter a imagem de um objeto por cerca de 1/20 à 1/5 segundos após o seu desaparecimento do campo de visão, ou seja, é a fração de segundo em que a imagem permanece na retina (Belas Artes, 2006).

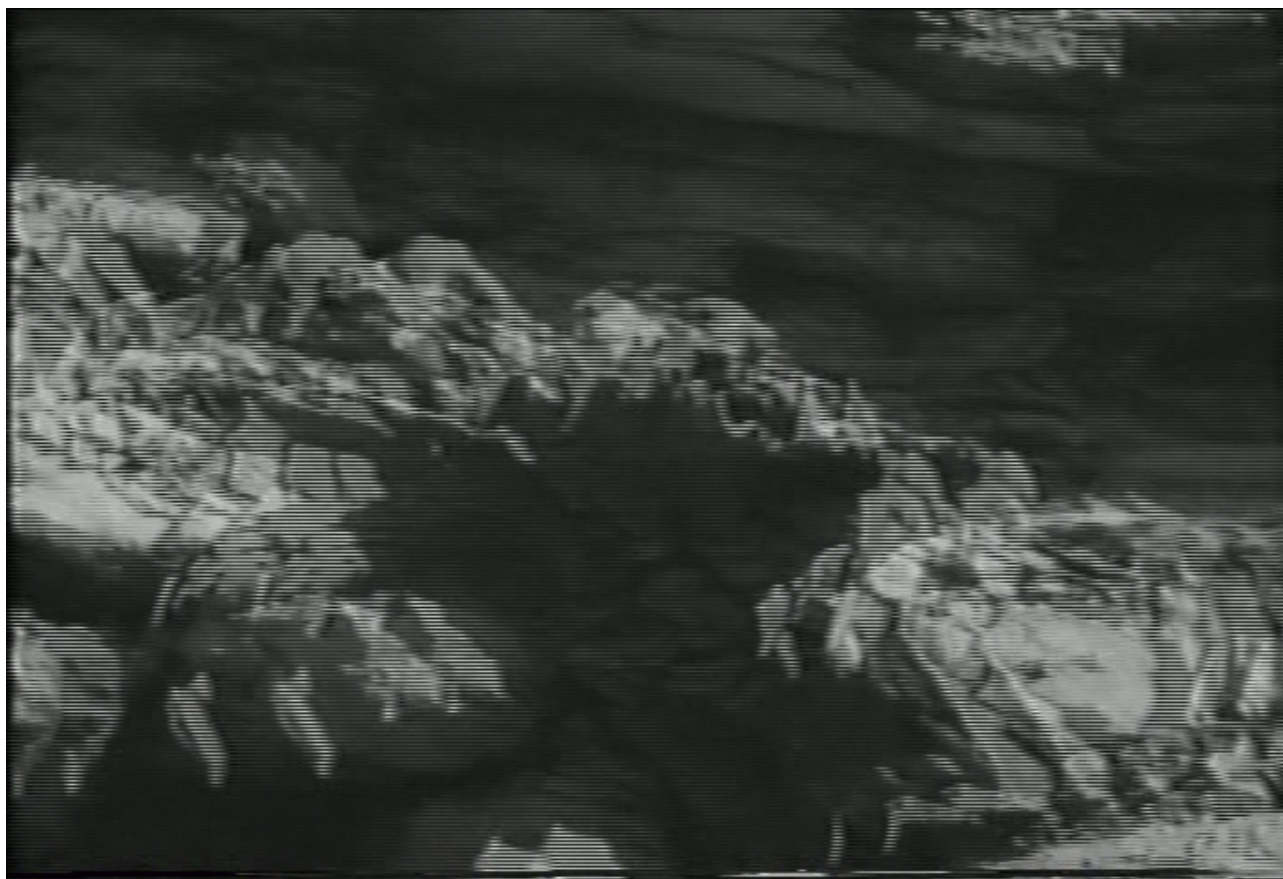


Figura 3.8: Quadro estereoscópico campo-sequencial

Para sua visualização é necessária a utilização pelo espectador de óculos obturadores de cristal líquido (*LCD shutter glasses*) que são capazes de permitir a passagem da luz somente das imagens referentes a cada olho, numa frequência de 60Hz. Desse modo, o sistema eletrônico que comanda a obturação da passagem de luz de cada um dos lados dos óculos abre-se e fecha-se a cada $1/60$ s, sincronizando-se com o vídeo que está sendo exibido, de tal forma que, a abertura do lado esquerdo dos óculos corresponda à projeção do campo referente ao olho esquerdo, e vice-versa. Este sistema só funciona em monitores e projetores do tipo CRT.

A tecnologia dos óculos obturadores já é utilizada largamente, existindo no mercado audiovisual/computacional inúmeras empresas que os comercializam, sejam conectados através de cabos às placas gráficas de vídeo dos computadores, sejam com óculos avulsos, sincronizados apenas através de pulsos de luz infravermelha, produzida por emissores conectados entre a placa e o monitor de computadores. Também é possível encontrar-se sistemas semelhantes que são conectados a aparelhos de DVD.

3.4.3 Vídeos Estereoscópicos Visualizados com Óculos de Filtros Coloridos

Da mesma forma que os dois métodos anteriores, o processo de captação deste tipo de vídeo estereoscópico necessita do armazenamento de dois sinais de vídeo em dois dispositivos distintos, para um posterior processamento antes da exibição.

Mais simples do que os outros métodos, para a codificação deste tipo de vídeo estereoscópico é necessário separar os canais RGB dos dois vídeos, sendo que do primeiro vídeo é extraído a informação do canal vermelho, e do segundo vídeo extraí-se os canais azul e verde (Wood, 2000), como mostra a **Figura 3.9**. Compõe-se então um novo vídeo com os canais extraídos, como é mostrado na **Figura 3.10**.

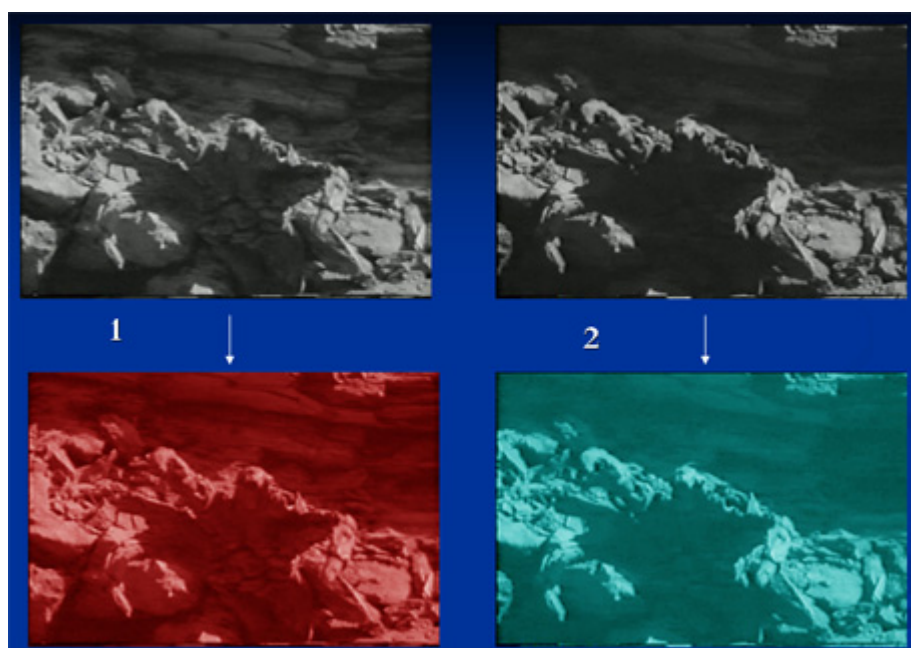


Figura 3.9: Processo de extração do canal vermelho do vídeo 1 e dos canais verde e azul do vídeo 2

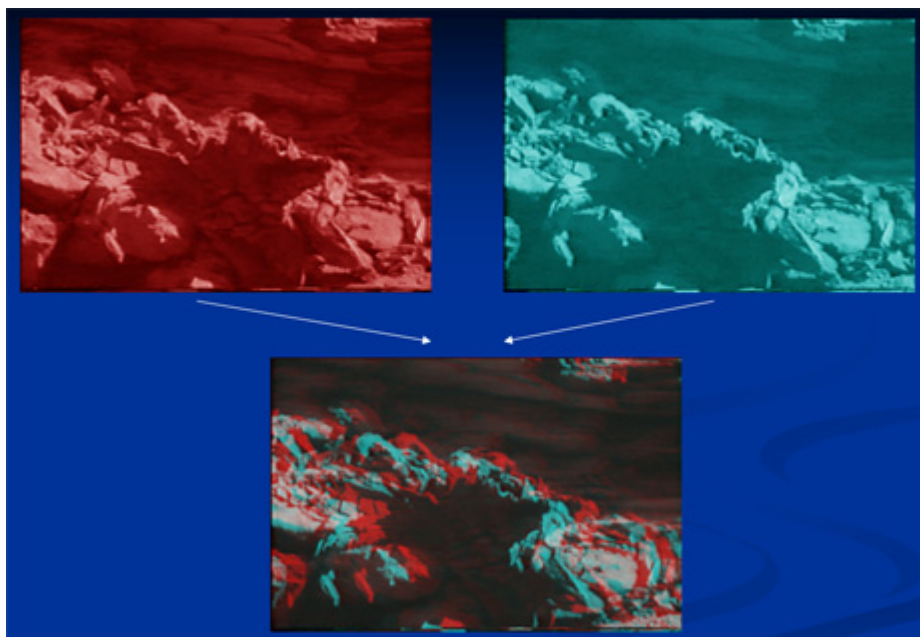


Figura 3.10: Processo de composição do vídeo anaglífico

Para a visualização assim como citado para a exibição cinematográfica de filmes ciano-vermelho, o espectador necessita utilizar óculos com um lado com lente vermelha e o outro com lente ciano, para separar cada uma das imagens que se encontram misturadas no monitor ou projeção.

3.5. Armazenamento de Vídeos Estereoscópicos

Os vídeos estereoscópicos visualizados com óculos polarizadores, são armazenados em dois dispositivos distintos com uma determinada codificação. Neste tipo de armazenamento, cada um dos vídeos pode ter compactação espacial ou temporal, mas nenhuma otimização pode ser realizada entre os vídeos, pois eles devem ser reproduzidos por dispositivos independentes.

Os vídeos estereoscópicos visualizados com óculos obturadores e com filtros polarizadores podem ser armazenados em único arquivo, sendo que o *player* pode se encarregar de reproduzi-lo para determinada visualização. Os dois formatos que não utilizam entrelaçamento de vídeo são (Lipton,1997):

- **Lado a Lado (*Side by Side*):** nesse formato os dois vídeos são gravados lado a lado em um único arquivo que possui no formato de sua janela o dobro da largura de seu tamanho original, como pode ser observado na **Figura 2.14**;
- **Sobrepostos (*Over/Under*):** nesse formato os vídeos são gravados sobrepostos em um único arquivo, que possui no formato de sua janela o dobro da altura de seu tamanho original, como pode ser observado na **Figura 2.15**.



Figura 3.11: Quadro de um vídeo lado a lado (*side by side*) quadro original: 720 x 480 – quadro lado a lado: 1440 x 480



Figura 3.12: Quadro de um vídeo sobreposto (*over/under*) quadro original: 720 x 480 – quadro sobreposto: 720 x 960

Para cada um desses formatos, ainda pode existir uma faixa preta separando as imagens. Uma justificativa para a existência dessa faixa, vertical no formato lado a lado e horizontal no formato sobrepostos, é a imposição de alguns compressores de vídeo em relação à razão de aspecto das imagens. Por isto, não existe um padrão de altura (no caso vertical) ou largura (no caso horizontal) dessa faixa preta. Convencionalmente, busca-se adotar 5% da altura total para a faixa horizontal e 5% da largura total para a faixa vertical (Kirner & Tori, 2004).

O software *TecStereoPlayer*, desenvolvido na PUC-Rio (Carvalho et al, 2004) reproduz arquivos lado a lado e sobrepostos. É importante salientar que a velocidade do disco onde o vídeo está armazenado e a capacidade de processamento da máquina influenciam diretamente na reprodução do vídeo, e mostram-se como um fator limitante para visualização de vídeos estereoscópicos.

No caso dos vídeos estereoscópicos anaglíficos, a codificação com perda (*lossy*), reduz o espaço de cor e isso acarreta em erros de separação de áreas de cor internas a imagem (erro conhecido como *blockiness*) (Ebrahimi et al, 2004) que dificultam a visualização estéreo.

3.6. Considerações Finais

O propósito deste capítulo, no contexto do projeto a ser desenvolvido, foi aprofundar os conhecimentos em técnicas de compressão de imagens e vídeos para facilitar o entendimento do funcionamento de uma codificação padrão de vídeo com as técnicas atuais.

Do ponto de vista da reprodução do vídeo estereoscópico, as melhores experiências com o efeito estéreo são as que utilizam luz polarizada e de monitores autoestereoscópicos, seguidas pela visualização de multiplexação temporal e por fim as que utilizam filtros coloridos. Levando em consideração o custo, as primeiras técnicas necessitam de um alto custo se comparada com a última, que pode ser reproduzida por qualquer monitor de computador acima do padrão VGA. Um quadro comparativo entre custo, qualidade, perda de cores e compressão de dados é ilustrado na Tabela 3.2.

Ao codificar um vídeo estereoscópico anaglífico em um único arquivo, é necessário utilizar uma codificação sem perda para um melhor resultado de visualização dos dados. No caso de optar por uma qualificação com perda, cada uma das etapas de codificação deve ser acompanhada para que não sejam gerados erros diferentes em cada uma das componentes de cor, que prejudicam a visualização anaglífica.

Tabela 3.2: *Comparações de Custo, Qualidade, Perda de Cor e Compressão de Dados de vídeos estereoscópicos armazenados com o modelo proposto por Lipton, 1997*

Visualização Estereoscópica	Custo ¹	Qualidade da Visualização ²	Perda de Cor ³	Compressão de Dados ⁴
Luz Polarizada	Altíssimo	Excelente	Nenhuma	Baixa
Monitor Autoestereoscópico	Alto	Boa	Nenhuma	Baixa
Multiplexação	Regular	Razoável	Nenhuma	Baixa
Anaglífica	Baixo	Razoável	Alta	Alta

1 Quando ao critério custo, partimos do pressuposto que o usuário já possui um computador em casa, e os seguintes parâmetros foram seguidos: baixo – U\$5,00; regular –U\$ 100,00; Alto – U\$ 5.000,00; Altíssimo: U\$ 30.000,00.

2 No critério qualidade da visualização, os seguintes parâmetros foram seguidos: excelente – apesar do óculos, a projeção com filtros de polarização não alteram as imagens; boa - a visualização para monitores estereoscópicos não necessita dos óculos, mas a qualidade está limitada a quantidade de camadas de LCD; razoável – a multiplexação mantém as cores, mas o chaveamento de imagens causa desconforto; baixa – a visualização anaglífica quando exibida a partir de um vídeo sem compressão de cores, possibilita uma visualização estereoscópica com boa noção de profundidade.

3 No parâmetro perda de cor a visualização anaglífica aparece como alta, pois as cores da visualização não são as mesmas da captação.

4 Na escolha de escalas do parâmetro taxa de compressão, com exceção da visualização anaglífica que pode ser armazenada em 3 canais (RGB), os outros métodos precisam armazenar 6 canais (um RGB para cada imagem do par estereoscópico).

Consultando a tabela 3.1 (Gonzalez & Woods, 2008), observamos que na maioria das codificações atuais que apresentam perda, é feita a escolha pelo espaço de cor YCbCr. Em muitos deles temos o uso da DCT enquanto em poucos temos a DWT. Estas duas transformadas matemáticas serão comparadas, a fim de determinar quais as implicações de seu uso na qualidade final do processo de compressão de vídeo.

O Capítulo 4 apresenta alguns trabalhos relacionados, importantes para conduzirem testes com as limitações apresentadas, que serão mostradas no Capítulo 5.

Trabalhos Relacionados

4.1. Considerações Iniciais

Desde a formalização do formato de vídeo estereoscópico no final da década de 90 (Lipton, 1997), os trabalhos de estereoscopia de vídeo digital se concentraram em sua exibição, contrastando com as pesquisas em codificação de vídeo desse tipo que ficaram estagnadas. Por esse motivo, os trabalhos aqui relacionados tratam dos métodos de codificação de vídeos bidimensionais, ressaltando àqueles que avaliam qualidade de vídeo.

No processo de codificação de vídeo com perda, a qualidade final mostra-se variável, devido ao grande número de fatores internos aos métodos. A qualidade visual de um vídeo é primordialmente subjetiva, e é influenciada por muitos fatores que dificultam a obtenção de uma métrica de qualidade precisa. Métodos ditos objetivos realizam um cálculo de erro matemático preciso, mas estes métodos não reproduzem a experiência subjetiva da visão humana ao visualizar um vídeo digital.

Este capítulo mostra comparações entre algumas técnicas de codificação, métricas para mensurar qualidade de vídeo e uma discussão sobre os problemas da codificação de vídeos estereoscópicos anaglíficos.

4.2. Armazenamento de Vídeos Estereoscópicos

O artigo que define os formatos lado-a-lado e sobrepostos para armazenamento de vídeos estereoscópicos, foi proposto por Lenny Lipton em 1997, e é intitulado *Stereo-Vision Formats for Video and Computer Graphics*. Lipton definiu ambos os formatos tendo em vista que a visualização estereoscópica seria por projeção de luz polarizada ou por multiplexação temporal, tecnologias presentes na década de 1990. O trabalho de ajustar a imagem para dois projetores ou realizar a multiplexação temporal sobre os formatos de arquivo propostos, fica a cargo do *player* que trabalha sobre o vídeo.

Na metodologia proposta, o tipo de contêiner de vídeo (AVI, MOV, MPEG) não é definido, nem tampouco o tipo de CODEC ficando a decisão de escolha por conta do usuário. Apesar de ser perfeitamente possível um *player* filtrar os canais de cor para a reprodução anaglífica estereoscópica, os resultados obtidos com a grande maioria de CODECs não produzem bom resultados para este tipo de visualização.

4.3. Comparação entre Técnicas de Codificação

Tendo como objetivo comparar algumas técnicas de compressão de vídeo, em especial as transformadas DCT e DWT, o artigo *JPEG vs. JPEG2000: Na Objective Comparison of Image Encoding Quality* (Ebrahimi et al, 2004), descreve os erros encontrados na codificação JPEG (que utiliza a transformada DCT) e na codificação JPEG2000 (que utiliza DWT).

Utilizando-se de uma estimativa visual da compressão para geração de artefatos, perceptíveis qualitativamente, é mensurado o MOS (*Mean Opinion Score*), um índice de percepção visual subjetiva. Os métodos de codificação utilizados foram baseados em codificações proprietárias (JPEG – Adobe Photoshop e IrfanView; JPEG2000 – JasPer, Kakadu, and IrfanView) com vários níveis de compressão. Com base nos resultados, verificou-se que a qualidade da codificação JPEG2000 mostra-se muito superior à codificação JPEG em relação à quantidade de *blur* (embaçamento) e *blockiness* (quantidade de artefatos de compressão) como ilustrado na Figura 4.1.

Seguindo nessa linha, outro trabalho que deixa claro a superioridade do JPEG2000 sobre o JPEG, é de um grupo russo de pesquisadores intitulado MSU Graphics&Media Lab Video Group⁹. No artigo *JPEG 2000 Image CODECs Comparison* (MSU, 2005), uma comparação é feita entre o JPEG e JPEG2000 mostrando através de medidas de PSNR¹⁰ a superioridade do JPEG2000. Além disso mostra que em relação à qualidade, o JPEG2000 também ganha na compressão de dados, fato também atribuído ao uso da DWT ao invés da DCT.

No artigo *Subjective Comparison of Modern Video CODECs* (Vantolin, 2006), temos uma comparação entre os CODECs DivX 6.0, XviD 1.1.0, x264 svn-352M e WMV 9.0 utilizando duas taxa de *bitrate* (690 e 1024 kbps). Foram analisadas medidas objetivas (PSNR, VQM e SSIM) e subjetivas

⁹ <http://www.compression.ru/video>

¹⁰ Esta métrica é chamada de pico da relação sinal ruído (*peak-to-peak signal-to-noise ratio*), e é uma 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 representa-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 frame original. Matematicamente o PSNR de uma imagem de dimensão m por n é representado por:

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

(MOS), e o resultado apontou para o padrão x264 como superior aos outros dois. É importante salientar que todos os métodos analisados neste artigo utilizam transformada DCT.

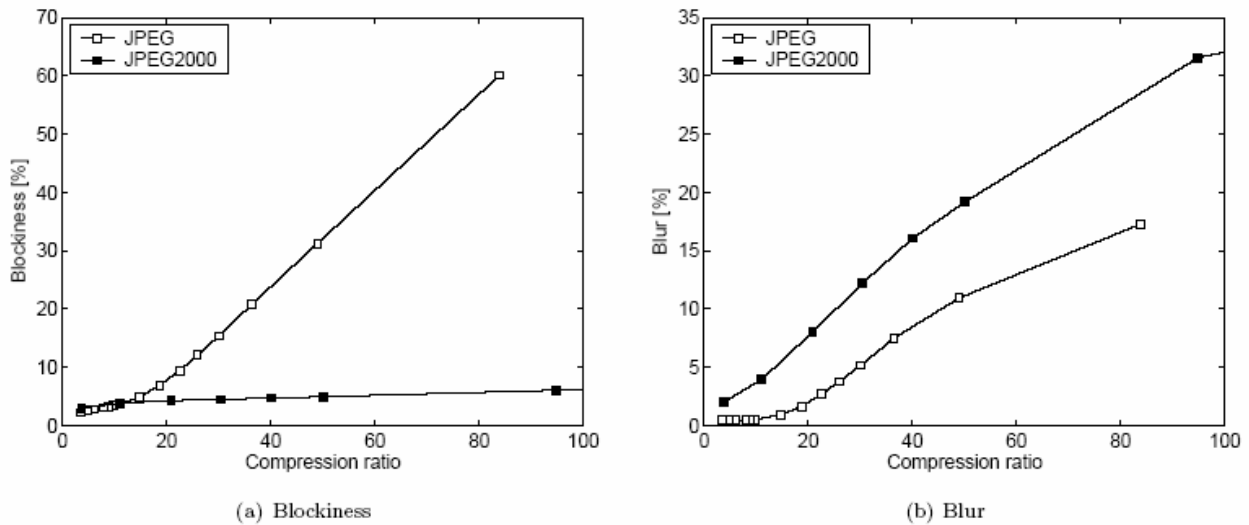


Figura 4.1: Comparações entre JPEG e JPEG2000. A imagem (a) mostra o gráfico da relação da compressão pela porcentagem de *blockiness*, enquanto a imagem (b) mostra um gráfico da relação da compressão pela porcentagem de *blur* (Ebrahimi et al, 2004).

4.4. Problemas na Codificação de Vídeos Estereoscópicos Anaglíficos

A metodologia de codificação proposta por Lipton, não é adequada para trabalhar com vídeos estereoscópicos anaglíficos, pois a maioria dos CODECs que são padrão de uso tem perda de cores e erros gerados pela DCT.

A visualização anaglífica permite erros nos canais de cor RGB, contudo esse erro deve ser igual ou próximo em todos os canais. Diferenças entre o canal vermelho e os canais azul e verde geram desconforto ao espectador, pois fazem com que os canais de cores com mais erros tenham as bordas invadidas pelas outras cores. Isto pode ser observado pelo anaglifo da Figura 4.2, onde a imagem foi codificada com o CODEC MPEG4.

Ao utilizarmos a transformada DCT no processo de codificação, são gerados erros de *blockiness* (Ebrahimi et al, 2004), como podemos visualizar na Figura 4.3. Nesta figura é mostrada a componente vermelha sem compressão e com compressão MJPEG. Esses erros dificultam a visualização estereoscópica anaglífica, causando desconforto ao espectador.

Para mensurar se algum CODEC com perda era adequado para este tipo de vídeo, foi realizado um estudo de cálculo de erro entre o vídeo original e codificado de diversos CODECs com perda, para avaliação da quantidade de erros gerados em cada um dos canais RGB, a ser visto no Capítulo 5.

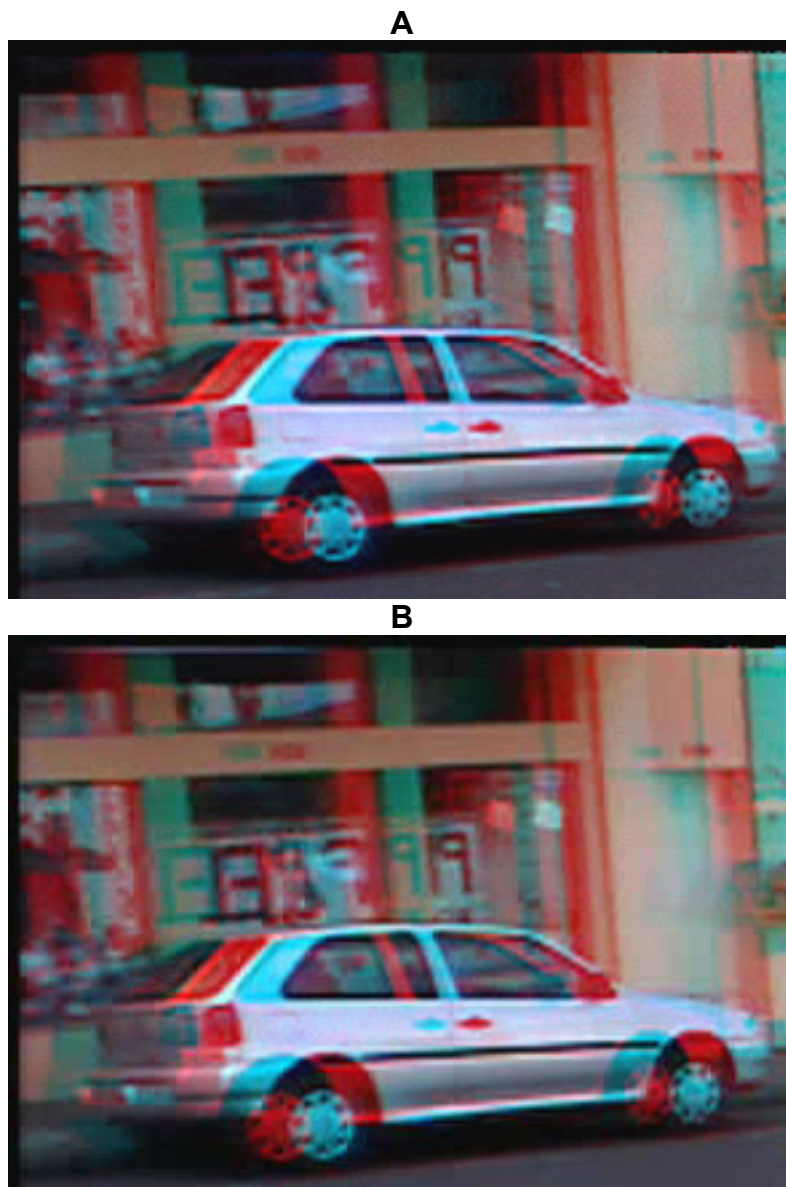


Figura 4.2: A imagem A mostra um recorte do quadro de um filme estereoscópico sem compressão. A imagem B o mesmo recorte com compressão MPEG4, com a qualidade máxima permitida pelo método de codificação¹¹. O PSNR para cada uma das componentes RGB foi de R: 15,96, G: 18,98 e B: 14,88.

¹¹ Essa imagem faz parte de um vídeo gerado com o software AVIDemux, com ajuste para o *profile Ultra High Definition*.



Figura 4.3: A imagem A mostra um recorte em um quadro de um filme estereoscópico, de sua componente vermelha. A imagem B mostra um recorte em um quadro com compressão MJPEG, com a qualidade máxima permitida pelo método de codificação¹² e com o aparecimento de embaçamento e de artefatos de compressão

4.5. Qualidade de Vídeo

Um *overview* sobre as métricas de qualidade de imagem e de vídeo, pode ser encontrado em um artigo feito pelo MSU (MSU, 2008)¹³. Neste artigo temos a descrição e formulação de índices objetivos relevantes, tais como *PSNR*, *Blurring Measure*, *Blocking Measure*, *SSIM Index*, *VQM*. Alguns deles medem um determinado efeito perceptível como às medidas de *blurring* e *blocking*, enquanto outras analisam luminância, contraste e estrutura da imagem (*SSIM Index*) ou a perda de informação de uma ou mais componentes de cor (*PSNR*). O índice *VQM* mede em particular a perda de informação com a transformada DCT.

A medida de qualquer um desses índices pode ser realizada com uma ferramenta desenvolvida pelo MSU. Num computador com o sistema operacional Windows, é possível realizar a instalação do software livre e fazer as comparações entre arquivos no contêiner AVI.

4.6. Considerações Finais

Entre os diversos processos de codificação de vídeo, os que acarretam em alguma perda são aqueles que conseguem maior compressão de dados. Como podemos observar pela qualidade objetiva e subjetiva quanto a DCT e a DWT, as transformadas wavelets ganham em qualidade e em tamanho dos

¹² Essa imagem faz parte de um vídeo gerado com o software AVIDemux, com ajuste para 100% de qualidade.

¹³ http://compression.ru/video/quality_measure/info_en.html

arquivos codificados. Considerando que um dos objetivos desse trabalho é o propor um novo método de codificação para vídeos estereoscópicos, a escolha de transformadas wavelets para essa nova codificação trará melhores resultados de qualidade e de compressão.

No próximo capítulo alguns testes de codificação são mostrados, onde será utilizada uma métrica objetiva (PSNR) para a determinação da perda de cor em cada uma das componentes RGB, em vídeos codificados com DCT e DWT e com os modelos de espaços de cor YCbCr e RGB.

Testes de Codificação

5.1. Considerações Iniciais

No contexto de visualização de vídeos estereoscópicos, a visualização anaglífica se mostra como uma solução de baixo custo, pois permite que informações de um vídeo estéreo sejam armazenadas em um vídeo digital bidimensional. Dessa forma, como apenas um monitor convencional e um óculos de lentes coloridas é possível realizar essa visualização.

Como mostrado no Capítulo 3 (nas Figuras 3.9 e 3.10) a composição de um vídeo anaglífico se dá pela combinação de duas imagens, captadas por blocos óticos ligeiramente afastados e gravadas em disco, fita magnética ou dispositivo de memória *flash*. Poucas são as câmeras digitais que captam vídeos sem compressão, pois isso exige grande espaço de armazenamento e *hardware* especial. A grande maioria das câmeras digitais codifica as imagens captadas por um CCD ou CMOS com um CODEC que reduz as informações do espaço de cor e que utiliza transformada DCT. Assim, ao utilizarmos a gravação de imagens de acordo com uma dessas codificações, o sinal armazenado já contém erros de codificação antes dos processos de composição e edição.

Ao compormos um vídeo anaglífico gravado com câmeras que gravam com CODECs como DV, MPEG2, MPEG4, por exemplo, são encontradas diferenças entre as componentes de cor RGB. Assim como mostrado na Figura 4.2, as diferenças entre essas componentes prejudicam a visualização de vídeos anaglíficos.

Com bases nessas limitações, foi realizado um estudo de cálculo de erro entre o vídeo original e codificado de diversos CODECs com perda, para avaliação da quantidade de erros gerada em cada um dos canais RGB.

5.2. Medidas de Qualidade e de Compressão em CODECs

Para realizar medidas de qualidade de vídeo, a amostra a ser analisada é muito importante, pois os CODECs podem ser ajustados para diferentes tipos de filme, e a qualidade depende dos ajustes

internos do CODEC (MSU, 2008b). Além disso se uma amostra do vídeo possuir algum tipo de compressão prévia, como os vídeos produzidos por quase todos os modelos de câmeras digitais (exceção feita às filmadoras que geram arquivos em discos rígidos sem compressão), artefatos como o *blockiness* podem interferir nas medidas realizadas.

É importante ressaltar que as medidas de compressão de dados realizadas neste trabalho foram realizadas com base em um arquivo codificado sem perdas de dados. O vídeo escolhido para realizar as medidas de qualidade foi uma animação intitulada *Elephants Dream*¹⁴, gerada com o software *Blender*¹⁵. (conforme ilustrado na Figura 5.1.) Nessa animação os personagens têm seu design projetado para serem enquadrados entre o realista e o caricato, enquanto o cenário ficcional enquadra-se como realista. Foi montado um vídeo com os quadros da animação sem qualquer compactação, o que impossibilita o aparecimento que quaisquer artefatos ou erros de codificação.

O erro foi calculado através da medida de PSNR em cada um dos canais de cor RGB da imagem, sendo a ferramenta utilizada *MSU Perceptual Video Quality*. Essa ferramenta funciona no sistema operacional *Windows*, e lê arquivos de vídeo encapsulados em um contêiner AVI. Como teste, foram feitas medidas do arquivo original em relação a ele mesmo, o que gerou um gráfico linear de valor 100 de PSNR para todos os quadros deste vídeo. Além das medidas nas componentes RGB, foi calculada uma média dos valores de PSNR (média aritmética) e também foi avaliada a relação de compressão dos CODECs em relação ao arquivo sem nenhuma compressão e em relação ao CODEC Huffyuv.

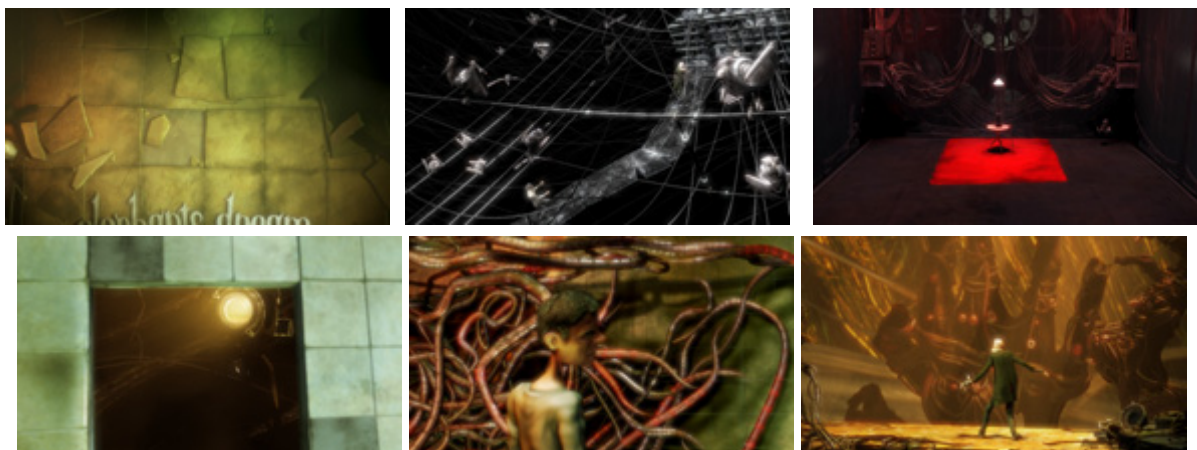


Figura 5.1: Quadros da animação *Elephant's Dream*

¹⁴ *Elephants Dream* é o primeiro filme de animação "aberto" (*open movie*), feito inteiramente com softwares gráficos de código aberto, e com todos os arquivos livremente abertos com a licença *Creative Commons*. Maiores informações em <http://orange.blender.org/>

¹⁵ *Blender* é um software de código aberto, desenvolvido pela Blender Foundation, para modelagem, animação, texturização, composição, renderização, edição de vídeo e criação de aplicações interativas em 3D. Maiores informações em: <http://www.blender.org/>

Seguem os dados do arquivo original:

- Tamanho do Arquivo sem compressão: 95.323.648 KB
- Tamanho do Arquivo codificado com Huffvuv: 32.276.989 KB
- Tamanho do quadro: 1920 x 1080
- *Framerate*: 24 fps
- Número total de frames: 15692
- Duração: 10 min 8 seg

A seguir temos os CODEC analisados.

CODEC: HD Cineform

Contêiner: AVI

Parâmetros de Codificação: 4:4:4, profile: High Definition

Tamanho do Arquivo Codificado: 15.174.282 KB

Relação de Compressão: 6,28 (descomprimido); 2,12 (Huffvuv)

Média do PSNR em cada canal:

- **R:** 9,97763
- **G:** 7,66462
- **B:** 11,52641

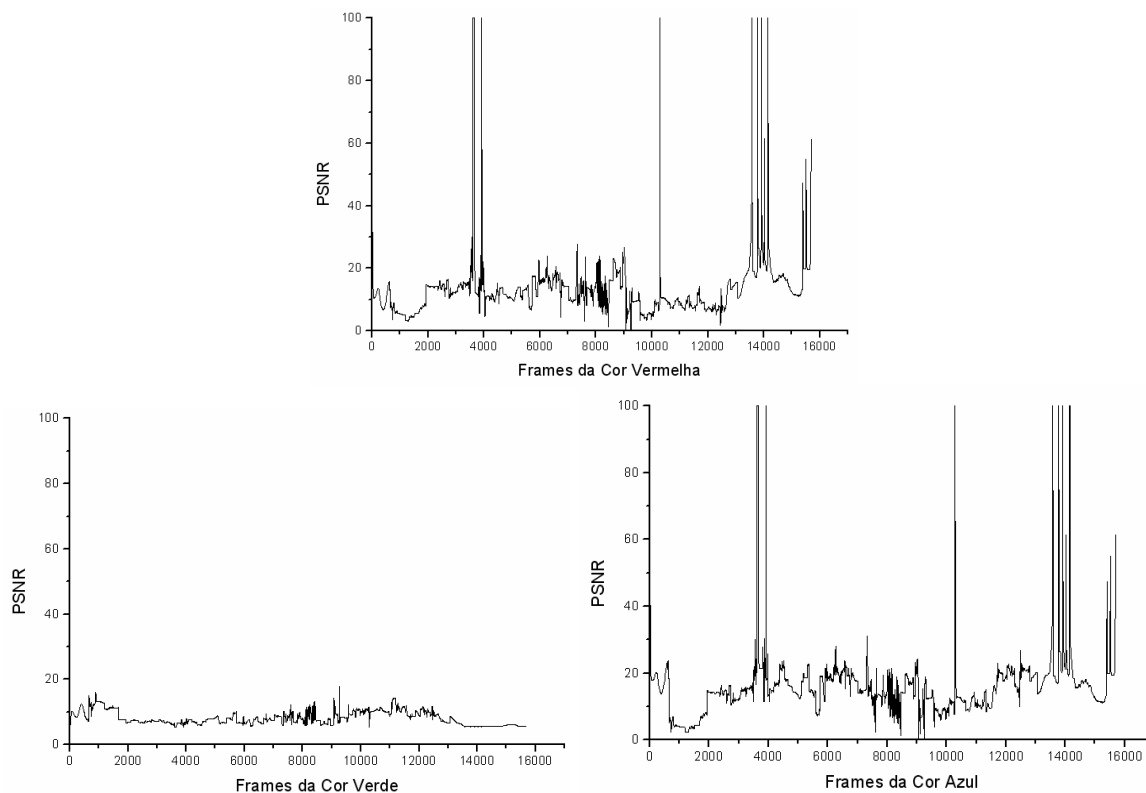


Figura 5.2: PSNR dos canais Vermelho, Verde e Azul com o CODEC HD Cineform

CODEC: MJPEG2000 Morgan

Contêiner: AVI

Parâmetros de Codificação: 4:4:4, constant bitrate

Tamanho do Arquivo Codificado: 8.061.209 KB

Relação de Compressão: 11,8 (descomprimido), 4 (Huffyuv)

Média do PSNR em cada canal:

- R: 19,82675
- G: 19,94227
- B: 19,90121

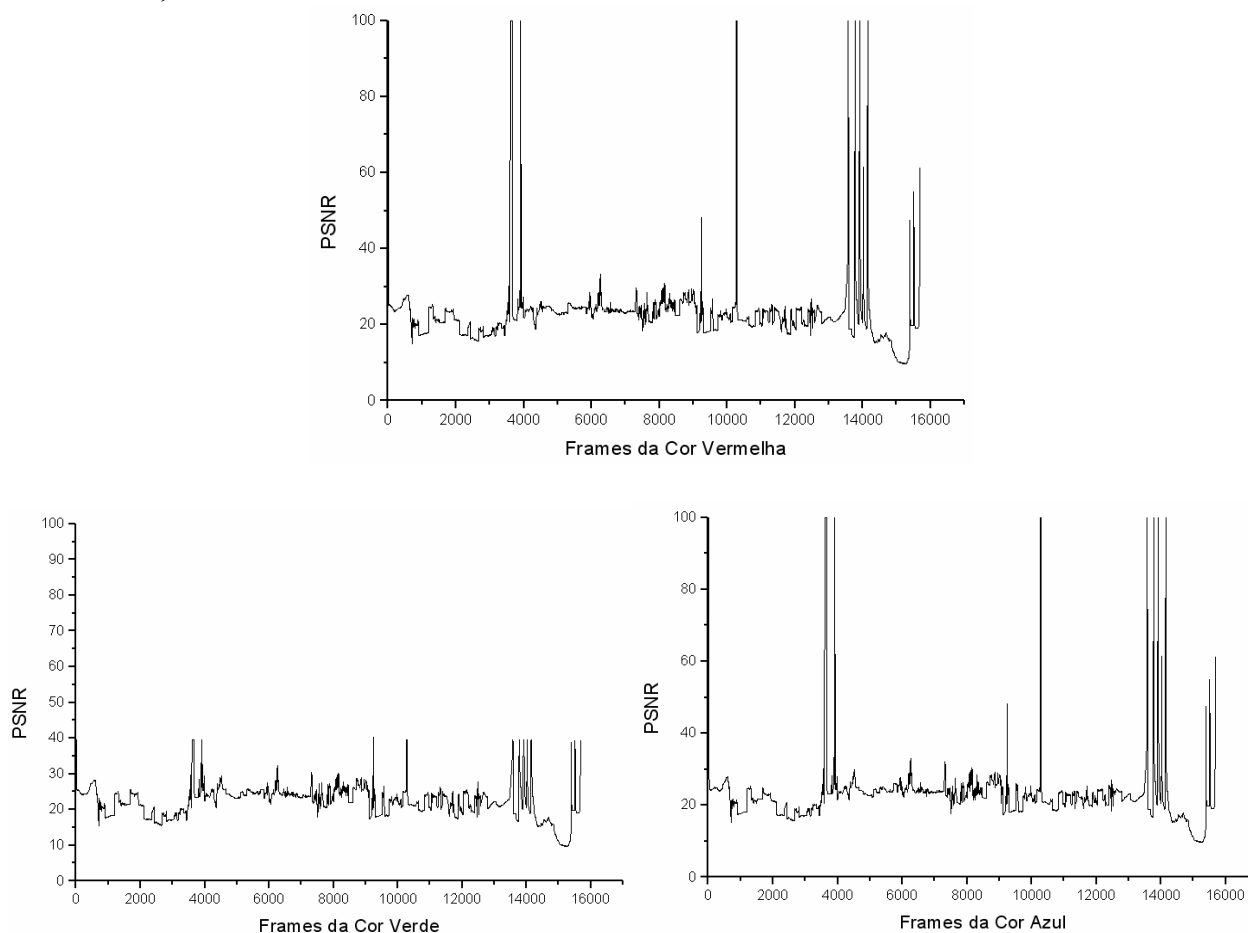


Figura 5.3: PSNR dos canais R, G e B com o CODEC HD MJPEG2000 Morgan

CODEC: MJPEG2000 Quicktime

Contêiner: MOV

Parâmetros de Codificação: 4:4:4, constant bitrate

Tamanho do Arquivo Codificado: 16.675.128 KB

Relação de Compressão: 5,71 (descomprimido), 1,93 (Huffyuv)

Média do PSNR em cada canal:

- R: 16,28834
- G: 16,61420

- B: 17,11141

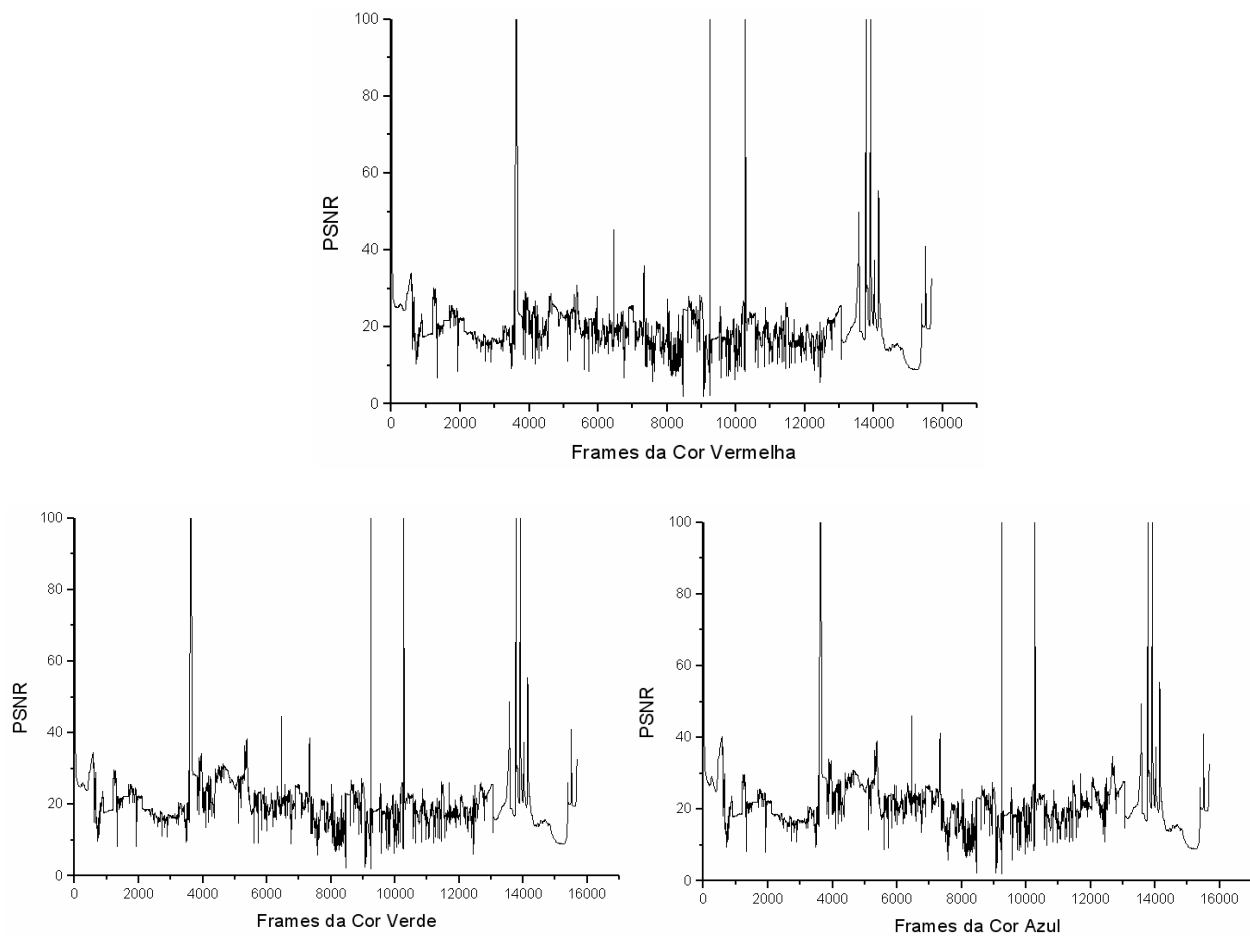


Figura 5.4: PSNR dos canais R, G e B com o CODEC MJPEG2000 Quicktime

CODEC: Apple Intermediate

Contêiner: MOV

Parâmetros de Codificação: 1080p 24fps

Tamanho do Arquivo Codificado: 3.634.339 KB

Relação de Compressão: 26,23 (descomprimido), 8,88 (Huffyuv)

Média do PSNR em cada canal:

- R: 16,03451
- G: 16,38116
- B: 16,94288

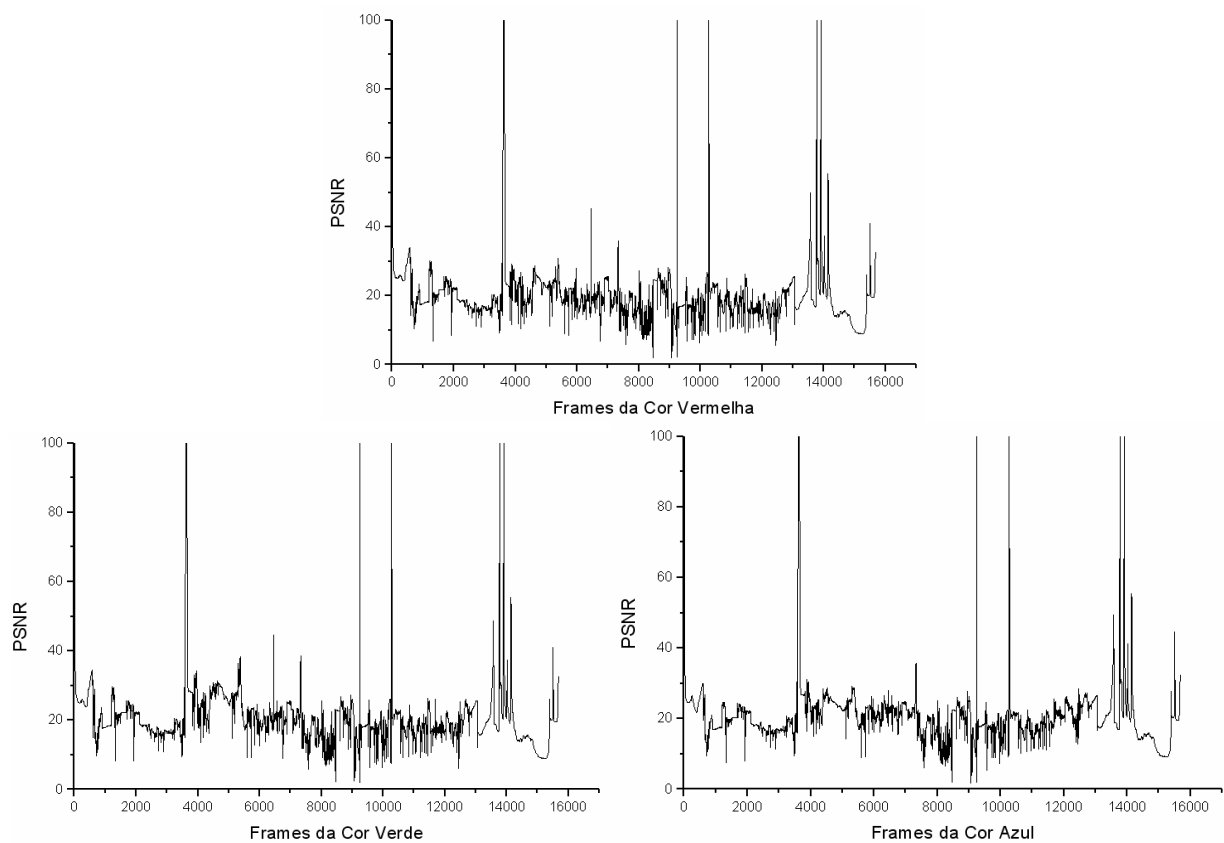


Figura 5.5: PSNR dos canais R, G e B com o CODEC Apple Intermediate

CODEC: MJPEG (Linux)

Contêiner: AVI

Parâmetros de Codificação: *100 quality*

Tamanho do Arquivo Codificado: 3.048.068 KB

Relação de Compressão: 31,27 (descomprimido), 10,59 (Huffyuv)

Média do PSNR em cada canal:

- R: 17,07351
- G: 19,69993
- B: 15,22869

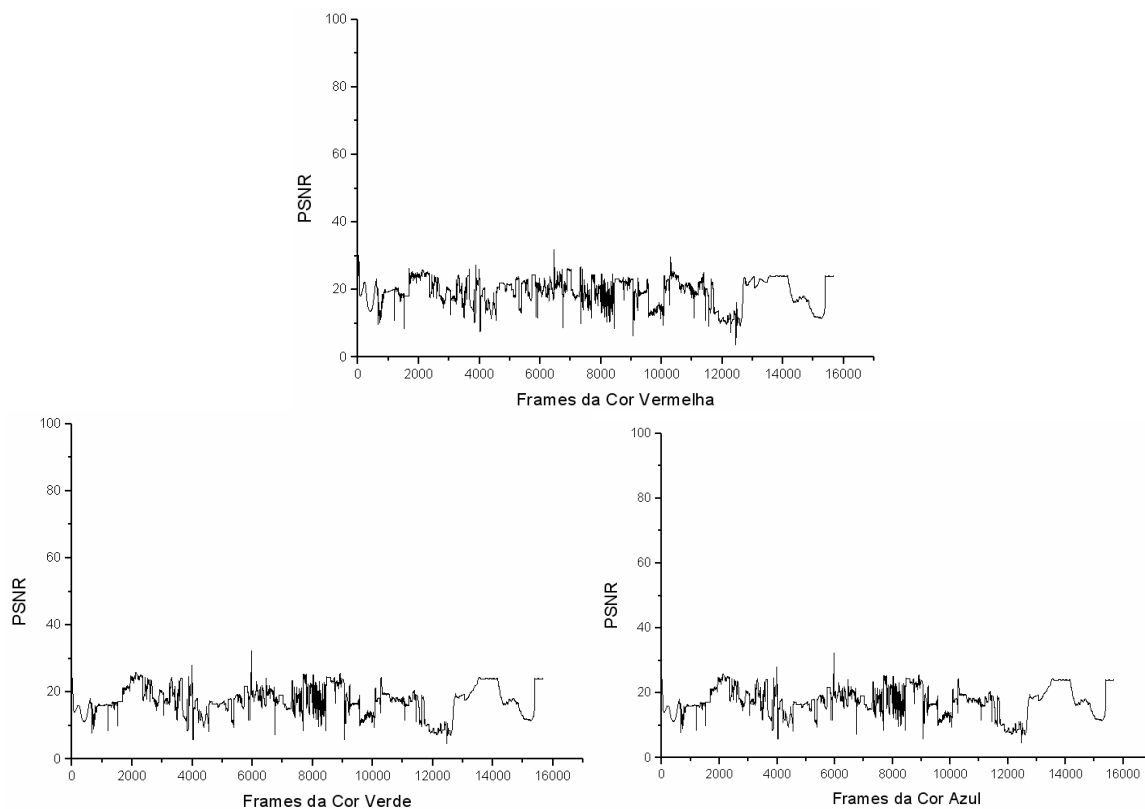


Figura 5.6: PSNR dos canais R, G e B com o CODEC MJPEG (Linux)

CODEC: x264 (Linux)

Contêiner: AVI

Parâmetros de Codificação: *Ultra-high resolution*

Tamanho do Arquivo Codificado: 234.243 KB

Relação de Compressão: 406,94 (descomprimido), 64,78 (Huffyuv)

Média do PSNR em cada canal:

- R: 15,73778
- G: 18,60417
- B: 14,79797

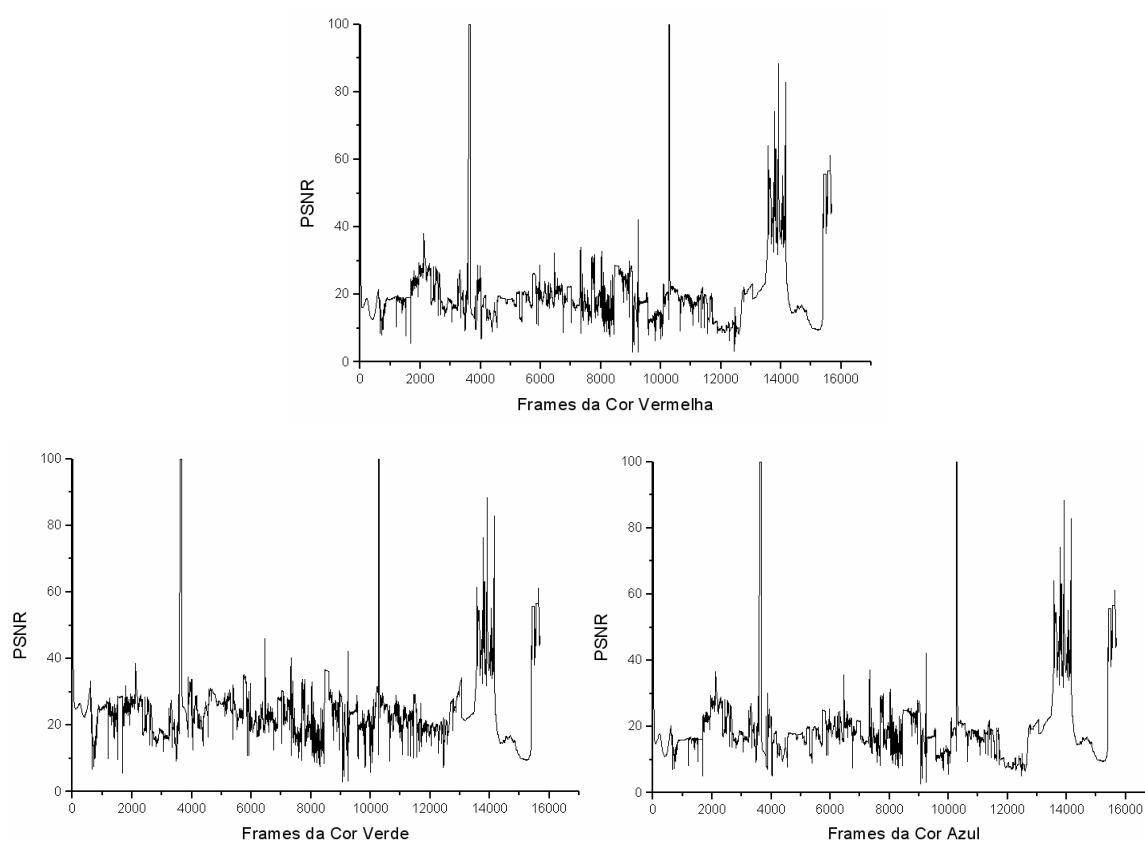


Figura 5.7: PSNR dos canais R, G e B com o CODEC x264 (Linux)

CODEC: MPEG4 (Linux)

Contêiner: AVI

Parâmetros de Codificação: 8000 Kbps

Tamanho do Arquivo Codificado: 717.287 KB

Relação de Compressão: 132,89 (descomprimido), 21,15 (Huffyuv)

Média do PSNR em cada canal:

- R: 15,96220
- G: 18,98243
- B: 14,88759

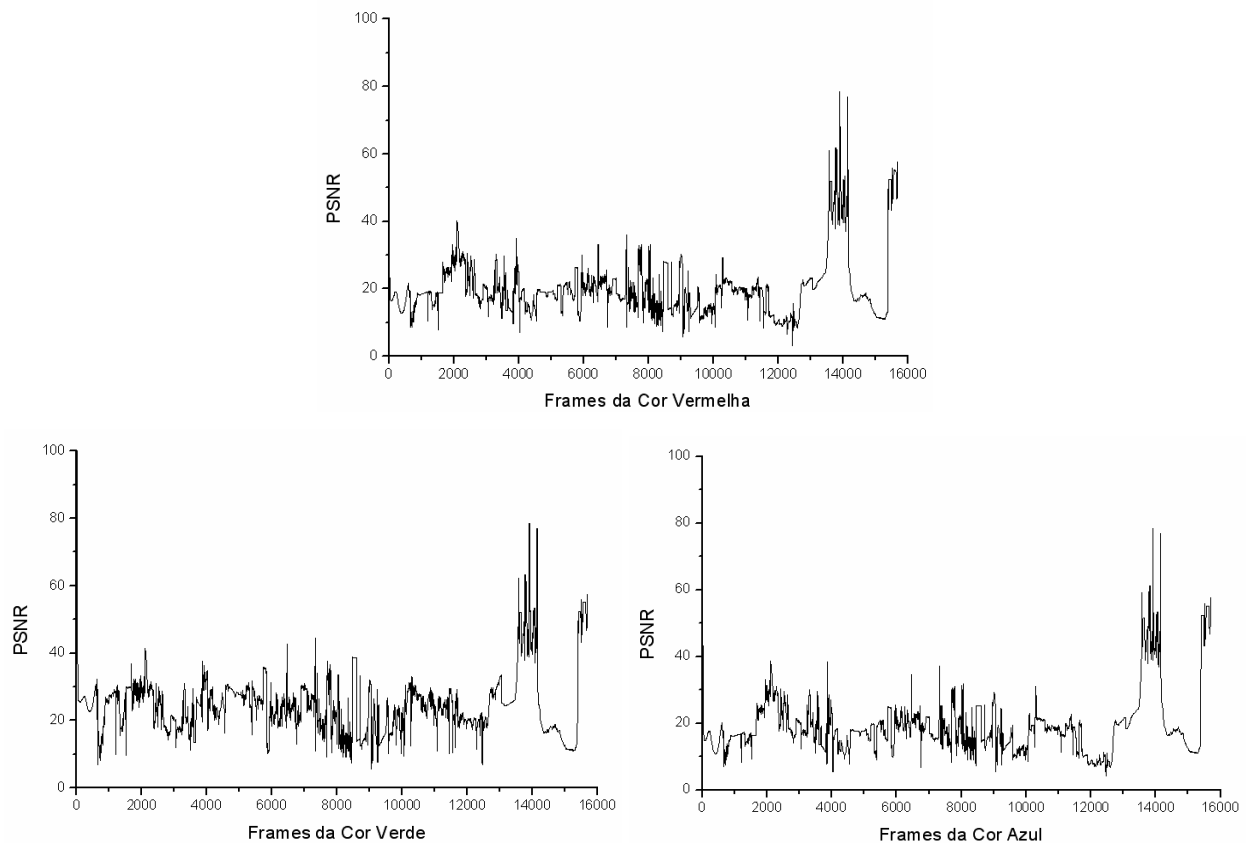


Figura 5.8: PSNR dos canais R, G e B com o CODEC MPEG4 (Linux)

Os CODECs testados foram configurados para que fossem obtidos seus melhores resultados de qualidade. Apenas a codificação MJPEG2000 Morgan se mostrou eficiente para armazenamento de vídeos estereoscópicos anaglíficos, pois o PSNR calculado para as componentes RGB, apresenta valores próximos. Apesar de outros métodos mostrarem uma compactação de dados maior, nenhum CODEC se mostrou com os valores de PSNR tão próximos quanto o MJPEG2000 Morgan, sendo que em segundo lugar quem mais se aproximou foi outro CODEC baseado em transformada wavelet, o CODEC MJPEG2000 Quicktime.

Como os CODECs baseados em transformadas wavelets se mostraram melhores que os baseados em DCTs, os próximos testes mediram a qualidade das componentes de cor com CODECs baseados em DWTs.

Para realizar medidas de PSNR em vídeos baseados em transformadas wavelets inicialmente foi escolhido um trecho do filme *Creature From Black Lagoon* codificado anaglífico em um arquivo descomprimido. O trecho em questão é a abertura do filme, onde o efeito estereoscópico pode ser visualizado com clareza¹⁶. Seguem os dados do arquivo:

¹⁶ Este trecho pode ser acessado via Web (<http://www.ies.ufscar.br/leoandrade/uploads/Docs/Videos/creature.html>). O arquivo está no contêiner FLV e com o CODEC On2VP6 (configurado para o máximo de qualidade, mas ainda assim apresentando alguns erros para a visualização estereoscópica).

- **Arquivo:** Creature_Abertura.avi
- **Número de Frames:** 5293
- **Tamanho sem compressão:** 5263 MB

Para realizar a codificação do arquivo, foi desenvolvido um software que codifica espacialmente as imagens utilizando três tipos de transformadas wavelets. Com base no código de uma implementação francesa DWT, intitulada Rududu¹⁷, foi criada uma aplicação que lê as imagens do vídeo original e as grava em um arquivo de extensão “.ric”.

No código Rududu é possível selecionar três tipos de transformadas wavelets: a transformada Haar (com quatro coeficientes), a transformada cdf97 (padrão do CODEC MJPEG2000), e a transformada a cdf53 (sendo que as duas últimas utilizam índices binários) (Tian et al., 2007). Neste código também é possível selecionar a quantização de dados, que varia de zero (sem quantização) à 16 (quantização máxima). Os resultados apresentados pertinentes à qualidade de vídeo são precisos, contudo os tamanhos dos arquivos comprimidos são aproximações, podendo ser melhorados com um processo de quantização e de redundância de dados adequado.

Seguem os dados obtidos:

Transformada 0 (cdf97)

Quantização: 1

Tamanho dos arquivos (.ric): 348 MB

PSNR:

- R: 43.43644
- G: 47.77644
- B: 46.23037

¹⁷ <http://rududu.ifrance.com/>

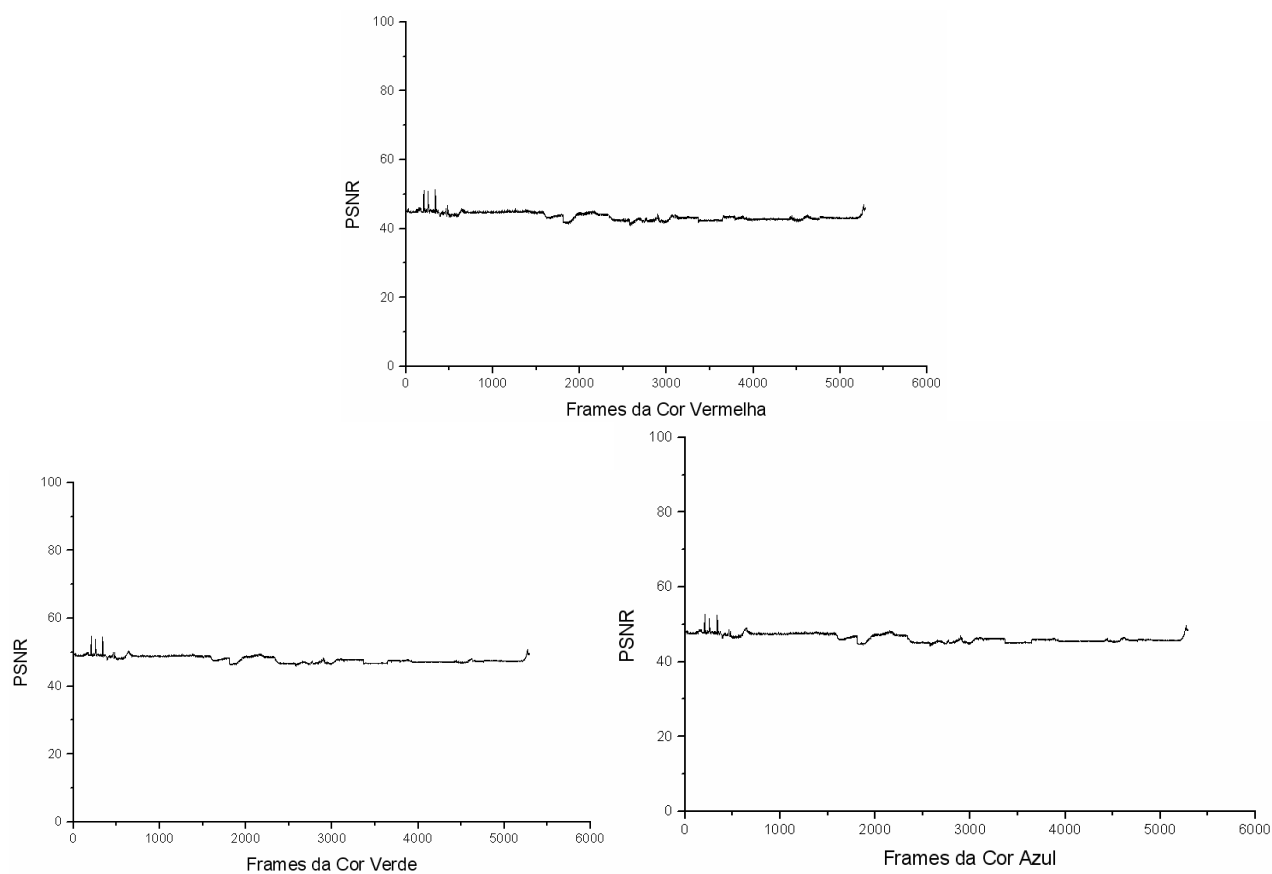


Figura 5.9: PSNR dos canais R, G e B com o CODEC Rududu e a transformada wavelet cdf97

Transformada 1 (cdf53)

Quantização: 1

Tamanho dos arquivos (.ric): 401 MB

PSNR:

- R: 43.23415
- G: 47.72237
- B: 46.54450

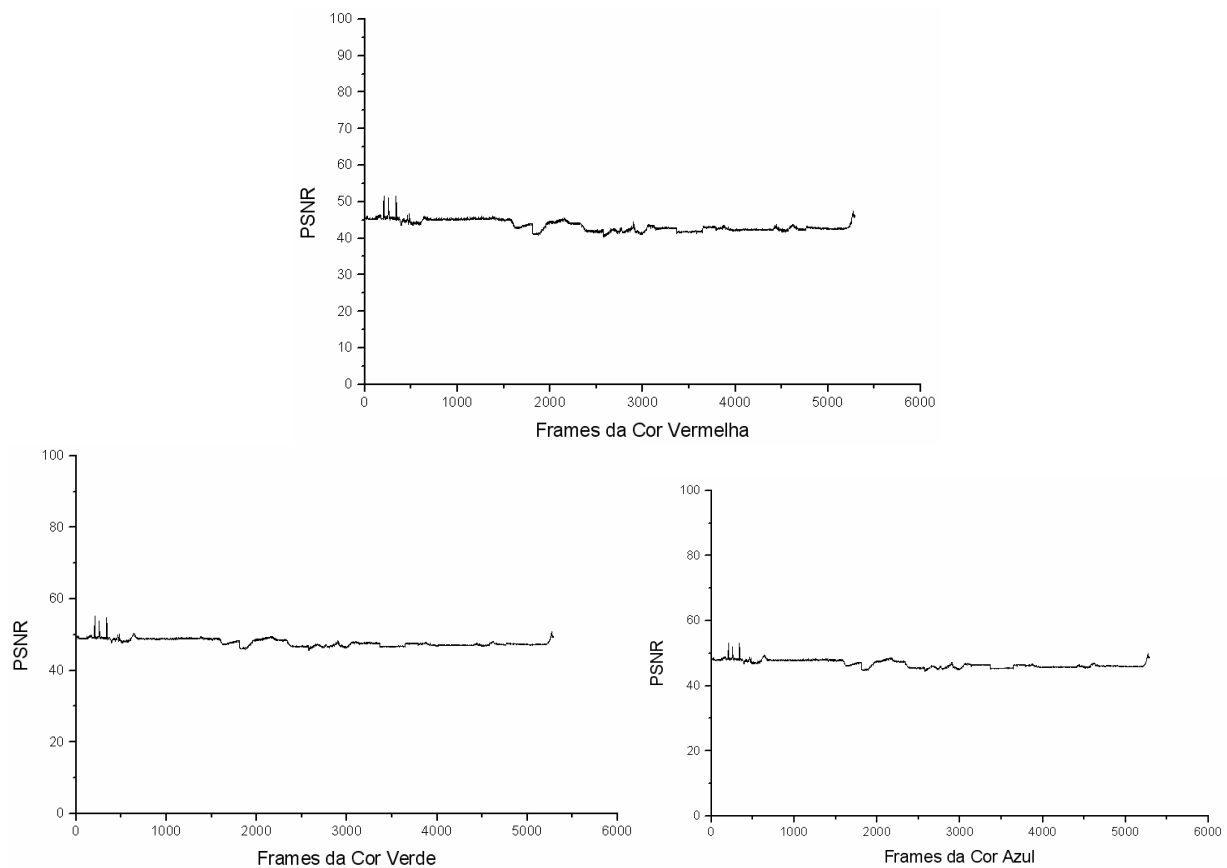


Figura 5.10: PSNR dos canais R, G e B com o CODEC Rududu e a transformada wavelet cdf97

Transformada 2 (Haar4)

Quantização: 1

Tamanho dos arquivos (.ric): 619 MB

PSNR:

- R: 41.65541
- G: 46.07053
- B: 44.29805

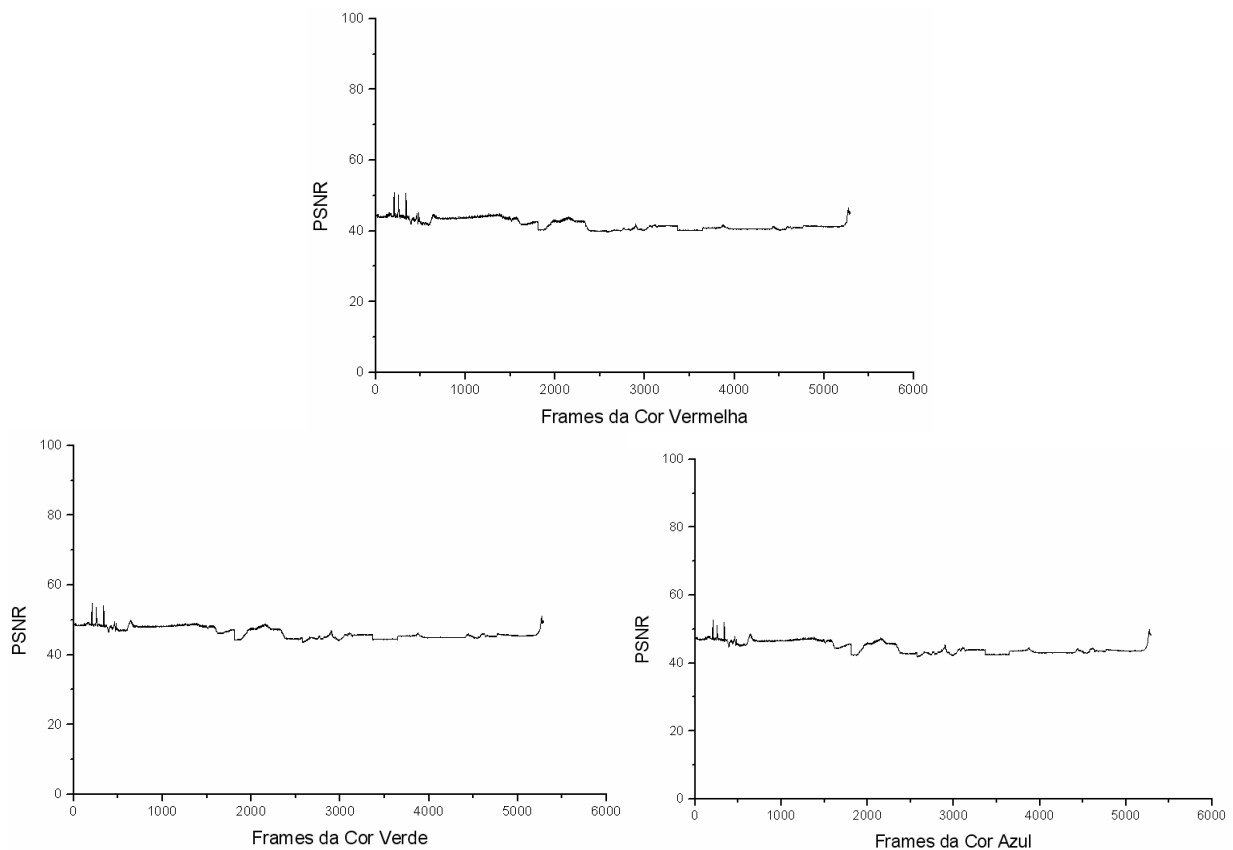


Figura 5.11: PSNR dos canais R, G e B com o CODEC Rududu e a transformada wavelet Haar4

Todos os testes realizados mostraram desigualdades em relação as componentes de cor RGB. Concluiu-se que tal fato se deve ao modelo de cor utilizado ser o YCbCr, que faz com que a quantização acarrete em diferenças de PSNR.

Visando eliminar o problema do modelo de cor, foi feita uma implementação de um métodos de codificação utilizando a transformada Daubechies 4. Neste novo método a codificação ainda foi implementada espacialmente, só que utilizando o modelo de cor RGB. O gráfico de medida de PSNR é mostrado a seguir:

Transformada: Daub4

Quantização: 1

Tamanho do arquivo: não avaliado

PSNR:

- R: 39,23413
- G: 39,93843
- B: 39,67592

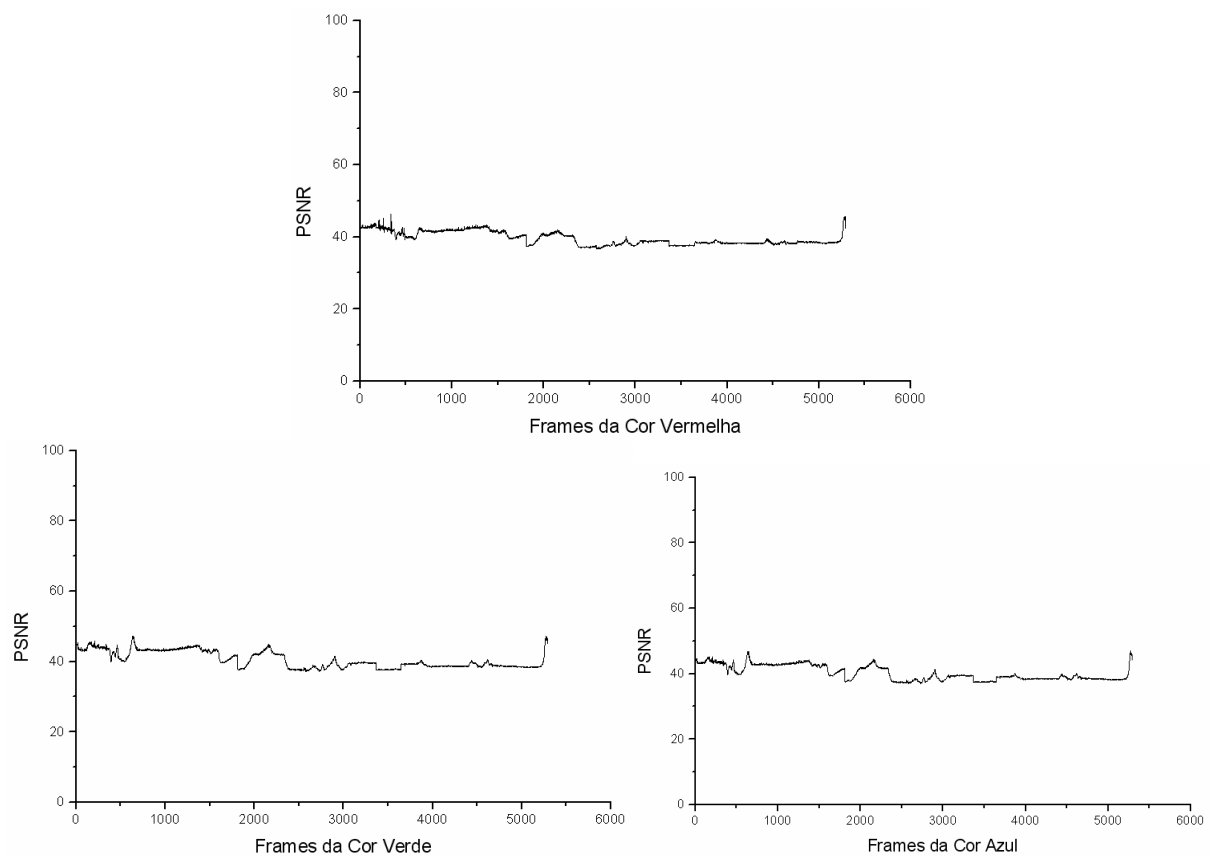


Figura 5.12: PSNR dos canais R, G e B com o CODEC desenvolvido utilizando a transformada wavelet Daub4

5.3. Considerações Finais

Ao compararmos as transformadas wavelets com as baseadas em transformadas de cossenos, temos que as primeiras conseguem atingir níveis de compressão de dados e de qualidade objetiva de imagem muito melhores. Dessa forma este trabalho aponta para a implementação de métodos de codificação de vídeos estereoscópicos com transformadas wavelets.

A escolha do espaço de cor RGB também se mostra necessária para que os erros gerados no processo de compressão se mostrem nivelados, o que não acontece os vídeos com perda no modelo YCbCr.

A implementação realizada com o espaço de cor RGB e transformada wavelet Daub4, mostrou-se promissora nos estudos de codificação de vídeos estereoscópicos.

Plano de Pesquisa

6.1. Objetivo

O objetivo deste projeto é desenvolver um novo método de codificação de vídeos estereoscópicos. Neste método o armazenamento deverá ser efetuado com compressão significativa de dados, utilizando espaço de vídeo RGB e a transformada DWT. A escolha desse conjunto se deve ao fato da codificação resultante aplicada a um vídeo acarreta em qualidade igualitária para todos os canais de cor. Isso fará com que esse método permita que a reprodução do vídeo possa ser efetuada utilizando diversos tipos de tecnologia, incluindo a visualização anaglífica que é sensível à perda de cores devido aos métodos de compressão atuais.

6.2. Metodologia e Atividades

Para este projeto será adotada a pesquisa aplicada, exploratória e experimental. Áreas como visualização estereoscópica, processamento de imagens, e testes de qualidade de vídeos estão entre as principais candidatas a se aplicar a pesquisa para obtenção de resultados que possam ser medidos e quantificados. Por ser exploratória, a pesquisa realizar-se-á por meio do estudo e revisão dos principais trabalhos reportados na literatura relacionados com o tema do projeto, a fim de identificar o estado da arte em codificação de vídeo digital aplicada à estereoscopia. Além disso, serão utilizados livros e artigos de periódicos especializados disponíveis em bibliotecas, bibliotecas digitais da ACM, IEEE, Springer-Verlag, Elsevier e outras fontes de artigos científicos da área. A parte de experimentação resultará em levantar as variáveis que influenciam os resultados obtidos por intermédio das abordagens estudadas.

Para que os objetivos propostos sejam alcançados, a metodologia empregada no estudo da literatura tem como foco pesquisar: visualização estereoscópica; etapas e técnicas de compressão de vídeo; métricas de qualidade de imagens e vídeos. Os estudos orientarão na escolha de passos que possam produzir um método de codificação que possa ser reproduzido com diversas técnicas de visualização.

Este trabalho abordou a sub-amostragem de dados em espaços de cor e suas implicações, assim como comparações entre a DCT e a DWT. Os resultados apontaram para o uso do espaço de cor RGB e da transformada DWT, pois os resultados obtidos mostraram qualidade igualitária para os canais RGB. A próxima etapa do processo de codificação é a etapa de quantização, onde os dados serão transformados em números binários, e posteriormente terão sua redundância eliminada. Além disso serão efetuados testes de medidas de qualidade sobre os métodos de compressão efetuados.

Com o intuito de validar os estudos realizados, pretende-se desenvolver uma aplicação que faça a codificação de vídeos estereoscópicos e sua posterior reprodução com o método proposto, utilizando para tal, técnicas de codificação que possuam boa compressão de dados e fidelidade de cores.

6.3. Cronograma

Para atingir os objetivos e resultados esperados é prevista a seguinte lista de atividades e o correspondente cronograma:

1. Estudos sobre estereoscopia e levantamento de padrões de codificação de vídeo.
2. Testes de qualidade de imagens e de CODECs.
3. Testes de compressão de imagens e de vídeo, com transformadas DCT e DWT.
4. Exame de qualificação.
5. Estudos sobre quantização de dados em codificação de vídeo.
6. Escolha de plataformas para implementação de software que manipule vídeos digitais.
7. Implementação da codificação de vídeos estereoscópicos com o método proposto. Medidas de qualidade, tamanho e tempo de execução.
8. Implementação da aplicação para visualização de vídeos estereoscópicos.
9. Submissão de resultados parciais e completos a simpósios e periódicos da área.
10. Defesa da dissertação.

Atividade	1º semestre 2008		2º semestre 2008		1º semestre 2009		2º semestre 2009		1º. Semestre 2010	
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										

6.4. Considerações Finais e Resultados Esperados

Dentre os resultados esperados, pretende-se obter um referencial sobre a qualidade da imagem nos mais eficientes CODECs atuais, além de poder apontar quais os paradigmas de codificação são importantes para a codificação de vídeos estereoscópicos.

Também é esperada a publicação de resultados parciais e completos em eventos e periódicos da área, além da ampliação dos conhecimentos do grupo sobre técnicas de codificação de vídeo digital.

Por fim depois da definição de um método para a codificação de vídeos estereoscópicos, pretende-se realizar a implementação de protótipos de codificadores para vídeo esteoscópicos.

- (Adams, 2001) Adams, G. O passe de mágica do turismo fantástico: o sistema de viagem estereoscópica de Underwood & Underwood. Anais do XXIV Congresso Brasileiro da Comunicação. Campo Grande, MS. Setembro de 2001. Disponível em: <http://reposcom.portcom.intercom.org.br/bitstream/1904/4646/1/NP7ADAMS.pdf> em 10 de junho de 2007.
- (Andrade, 2008) Andrade, L. A.; Difusão de Filmes Estereoscópicos. Revista RUA – Especial 1 – A Era Digital e seus Desdobramentos Estéticos. ISSN 1983-3725. Disponível em: <http://www.ufscar.br/rua/site/?p=657> em 20 de setembro de 2008.
- (Andrade & Moreira, 2001) Andrade, L. A.; Moreira E. S. Estudos em Transmissões Multicasting de Vídeo Comprimido. Dissertação de Mestrado, ICMC – USP São Carlos, 2001.
- (Apple, 2007) Apple. Developer Guide – Quicktime 7. Disponível em: <http://developer.apple.com/documentation/QuickTime/Conceptual/QT7UpdateGuide/index.html> em 9 de janeiro de 2007.
- (Belas Artes, 2006) Belas Artes, UFMG. Quadro a Quadro – Gênese da Animação. Belo Horizonte, 2006. Disponível em: <http://eba.ufmg.br/panorama/genesis/genese01.html> em 20 de setembro de 2006.
- (Carvalho et al, 2004) (Carvalho et al, 2004) Carvalho, F. G.; Bastos, T. A.; Raposo A.; Corseiul, E. T. L.; Szenberg, F.; Azevedo, L. Uma Ferramenta para a Composição de “Vídeos Interativos”. Disponível em: <http://www.sbc.org.br/bibliotecadigital/download.php?paper=434> em 16 de junho de 2007.
- (Costa, 2007) Costa, A.. C. O Olho Humano. Documento Técnico do Laboratório de Ensino de Ótica. Unicamp, 2007. Disponível em: <http://www.ifi.unicamp.br/~accosta/olhohumano.html> em 22 de março de 2008.
- (Dodgson, 1998) Dodgson, N. A. (1998) Autostereoscopic Image Compression.

- Rainbow Graphics Researchs Group, Computer Laboratory, University of Cambridge, Pembroke St. Disponível em: <http://www.cl.cam.ac.uk/~nad/compr/compr.html> em 30 de agosto de 2008.
- (Ebrahimi, 2002) Ebrahimi, F. Pereira. The MPEG-4 book. IMSC Press Multimedia Series, 2002.
- (Ebrahimi et al, 2004) Ebrahimi F.; Chamik M.; Winkler S. JPEG vs. JPEG 2000: An Objective Comparison of Image Encoding Quality. Proceedings of SPIE, 2004.
- (Evans et al, 1995) Evans, J.P.O.; Robinson, M.; Godber, S.X.; Petty, R.S. The Development of 3-D (Stereoscopic) Imaging Systems for Security Applications, 29th IEEE Intl. Carnahan Conference, pp. 505-511, Sanderstead, Surrey, UK, October, 1995, ISBN 0-7803-2627-X.
- (Furht, 1995) Furht, B. A Survey of Multimedia Compression Techniques and Standards – Part I: JPEG Standard. Real-Time Imaging, vol. 1, n. 1, pp. 49-67, 1995.
- (Fontoura, 2001) Fontoura, F. N. F. Estereoscopia – Curso de Especialização em Informática com ênfase em internet e aplicações de ensino, da Universidade Católica de Pelotas em 2001. <http://atlas.ucpel.tche.br/~magic/compgraf/estereoscopia.html> em 18 de abril de 2003.
- (Godoy, 2001) Godoy, H. Documentário, Realidade e Semiose, os sistemas audiovisuais como fontes de conhecimento. São Paulo, AnnaBlume/FAPESP, 2001.
- (Godoy, 2005) Godoy, H. A. S. A Imagem Tridimensional e o Audiovisual. Publicado no site da Associação Brasileira de Cinema em outubro de 2005. Disponível no site <http://www.abcine.org.br> em 10 de junho de 2007.
- (Gonzalez & Woods, 2008) Gonzalez, R. C.; Woods, R. E. Digital Image Processing. Third Edition. Prentice Hall, 2008.
- (Hearn & Baker, 2004) Hearn, D. Baker, M. P. Computer Graphics with OpenGL, Editora Prentice Hall, 2004.
- (Inoue & Ohzu, 1990) Inoue, T., and Ohzu, H. Measurement of the human factors of 3-D images on a large screen. Large-Screen Projection Displays II, SPIE Vol.1255, 1990.

- (ISO, 1996) ISO/IEC JTC1/SC29/WG11, MPEG-1: Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s, 1996.
- (ISO, 1997) ISO/IEC IS 13818, MPEG-2: Generic coding of moving pictures and audio information, 1997.
- (Kirner & Tori, 2004) Kirner, C.; Tori, R. Realidade Virtual: Conceitos e Tendências. Cap. 11, p.179-201. Editora Mania de Livro, São Paulo, 2004. ISBN 85-904873-1-8.
- (Langford, 1990) Langford, Michael J. Tratado de Fotografia, um texto avanzado para Profesionales. 5ª ed. Barcelona, Omega, 1990.
- (Lipton, 1982) Lipton, L. Foundations of the Stereoscopic Cinema, a study in depth. New York, Van Nostrand Reinhold Co., 1982.
- (Lipton, 1997) Lipton, L. Stereo-Vision Formats for Video and Computer Graphics White Paper. Disponível em http://www.stereographics.com/support/body_stereo_formats.html em 10 de junho de 2007.
- (Macedo, 1995) Macedo, M. R. Visão Computacional a partir de Imagens Estereoscópicas. UNICAMP, 1995.
- (Machado, 1997) Machado, L. S. A Realidade Virtual em Aplicações Científicas. Dissertação de Mestrado em Computação Aplicada, INPE [INPE-6389-TDI/605], São José dos Campos, Abril.
- (Mancini, 1998) Mancini, A. Disparity Estimation and Intermediate View Reconstruction for Novel Applications Stereoscopic Video. Thesis of Mestre submitted to the Departament of Electrical Engineering McHill University, Montreal, Canada.
- (Mathias & Patterson, 1985) Mathias, H.; Patterson, R. Electronic Cinematography, achieving photographic control over the video image. Belmont, Wadsworth, 1985.
- (Murray & Willian, 1996) Murray, James D.; William van Ryper. Encyclopedia of Graphics File Formats, Second Edition. Sebastopol, Calif.: O'Reilly, 1996. ISBN 1-56592-161-5.
- (MSU, 2005) MSU Graphics & Media Lab (Video Group). JPEG 2000 Image CODECs Comparison. CODECs Comparison, Moscou, Setembro 2005.
- (MSU, 2008a) MSU Graphics & Media Lab (Video Group). Quality Measurement

- Tool: Metrics information. Disponível em: http://compression.ru/video/quality_measure/info_en.html, em 16/08/2008.
- (MSU, 2008b) MSU Graphics & Media Lab (Video Group). Introduction to Video CODECs Comparison. Disponível em: http://compression.ru/video/CODEC_comparison/introduction_en.html, em 16 de agosto de 2008.
- (Natress & Jordan, 2005) Natress, Graeme; Jordan, Larry. What is Color Sampling? Larry's Final Cut Pro, issue March 2005. <http://www.larryjordan.biz/> em 5 de novembro de 2005.
- (Nelson, 1991) Nelson, M. The Data Compression Book (Redwood City, CA: M&T Books), 1991.
- (OpenDML, 1997) OpenDML AVI File Format Subcommittee. OpenDML AVI File Format Extensions. Disponível em: <http://www.the-labs.com/Video/odmlff2-avidef.pdf> em 10 de setembro de 2008.
- (Patterson, 1989) Patterson, F. Photography and The Art of Seeing, 1989, Key Porter Books, ISBN 1-55013-099-4.
- (Pereira, 1999) Pereira, G. H. C. Y. Mpeg-2 um estudo do padrão de vídeo. <http://homepages.dcc.ufmg.br/~gpereira/mpeg/mpeg.html>, 1999.
- (Perlin, 2000) Perlin, K., Paxia, S. and Kollin, J. (2000) An Autostereoscopic Display; SIGGRAPH 2000 Conference Proceedings. New Orleans, Louisiana. July 23-28. Disponível em: <<http://www.mrl.nyu.edu/publications/autostereo/autostereo.pdf>>.
- (Perlin, 2001) Perlin, K. et. al. (2001) Recent Advances in the NYU Autostereoscopic Display; Proceedings of the SPIE, Vol. 4297. San Jose, California. January 22-24. Disponível em: <http://www.mrl.nyu.edu/publications/autostereo/spie2001.pdf> em 30 de agosto de 2008.
- (Richardson, 2003) Richardson, I. E. G. H.264 and Mpeg-4 video compression. Editora Wiley. Londres, 2003. ISBN 0-470-84837-5.
- (Santos, 2000) Santos, E. T. Uma Proposta para Uso de Sistemas Estereoscópicos Modernos no Ensino de Geometria Descritiva e Desenho Técnico. Anais do III Congresso Internacional de Engenharia Gráfica nas

- Artes e no Desenho (GRAPHICA 2000), Junho, Ouro Preto, MG. (CD-ROM).
- (Siscoutto & Tori, 2003) Siscoutto, R. A. and Tori, R. Solução de Baixo Custo para Inserção de Vídeo Avatar Estereoscópico em Ambientes Virtuais Tridimensionais. Proceedings of SVR2003 – VI Symposium on Virtual Reality, Ribeirão Preto, v. 1, p. 264-272.
- (Steinmetz, 1995) Steinmetz, R.; et al. Multimedia: Computing, Communications and Applications. Prentice-Hall, Inc., 1995.
- (StereoGraphics, 1997) StereoGraphics Corporation. Stereographics Developers Handbook: Background on Creating Images for CrystalEyes and SimulEyes, 1997. Disponível em: http://www.stereographic.com/support/downloads_support/handbook.pdf em 16 de junho 2007.
- (Supia, 2007) Suppia, A. O. Monstro Brasileiro Revive em 3D. Revista Cultura e Ciência, Ano 59. no. 2. p.57-59. Disponível em: http://cienciaecultura.bvs.br/scielo.php?script=sci_pdf&pid=S0009-67252007000200024&lng=pt&nrm=iso em 10 de junho de 2007. ISBN 85-0009-6725.
- (Thakrar, 1996) Thakrar, K. & Andleigh, P.K. Multimedia Systems Design, Prentice-Hall, New Jersey, 1996.
- (Tian et al, 2008) Tian, Xin; Tan, Yi-Hua; Tian Jin-Wen. A New Design for 9/7 –Tap Wavelet Filter and Its VLSI Implementation. Chinese Journal of Computers, Vol. 31, No. 3, Março de 2008.
- (Tian et al, 2007) Tian, X.; Zhou, Z.; Tan, Y.-H.; Tian, J.-W. Parallel 9/7-tap wavelet based on lifting structure. Electronics Letters, vol 43, issue 11, pages 617-618. ISSN: 0013-5194.
- (Tommaselli, 2007) Tommaselli, A. M. G. Fotogrametria Básica - Estereoscopia e Paralaxe. Apostila. do curso Fotogrametria I da Unesp – Presidente Prudente, 2007. Disponível em: http://www2.prudente.unesp.br/area_doc/foto1/fotogrametria_I/FOTO5_6_2006.pdf em 20 de setembro de 2008.
- (Vantolin, 2006) Vantolin, Dmitriy. Subjective Comparison of Modern Video CODECs. MOS CODECs Comparison, Moscou, Janeiro 2006.

- (Wallace, 1991) Wallace, G.K. The JPEG Still Picture Compression Standard, Communications of the ACM, v. 34, n. 4, p. 30-44, Abril de 1991, disponível em <ftp://ftp.uu.net/graphics/jpeg/wallace.ps.gz>, 20 de novembro de 2000.
- (Welch, 1984) Welch, T. A Technique for High-Performance Data Compression, Computer, June 1984.
- (Wilt, 2006) Wilt, A. J. The DV, DVCAM & DVCPRO Formats. Disponível em: <http://www.adamwilt.com/DV.html> em 10 de setembro de 2008.
- (Winston, 1996) Winston, B. Technologies of Seeing, photography, cinematography and television. Londres, BFI Publishing, 1996.
- (Wood, 2000) Wood, D. 3D Imagery. Página Html - Nv News – Vidia, 2000. Disponível em: <http://www.nvnews.net/articles/3dimagery/introduction.shtml> em 16 de junho de 2007.
- (Ziv et al, 1978) Ziv, J., and Lempel, A. IEEE Transactions on Information Theory (A Universal Algorithm for Sequential Data Compression), vol. IT-24, pp. 530–536, 1978.

Anaglifo

Par de imagens estéreo formatado de maneira especial para fornecer um efeito tridimensional. Para a visualização utiliza-se de óculos de duas cores (cada lente com uma cor diferente – azul e ciano). A imagem é formada por duas camadas de cor sobrepostas, mas com uma pequena distância entre as duas para produzir um efeito de profundidade.

AVI (*Audio Video Interleave*)

Formato ou contêiner de vídeo utilizado para encapsular áudio e vídeo, criado pela Microsoft Corporation.

CCD (*Charge-Coupled Device*)

O dispositivo de carga acoplado é um sensor para captação de imagens formado por um circuito integrado contendo uma matriz de capacitores acoplados.

CMOS (*Complementary Metal-Oxide-Semiconductor*)

É um tipo de sensor de imagem usado comumente em câmeras digitais produzidos com a tecnologia de fabricação de transistores MOSFET, e tem sido utilizado como uma alternativa mais barata aos sensores do tipo CCD.

CODEC (*Codificador/Decodificador*)

Dispositivo de hardware ou software que codifica/decodifica áudio e ou vídeo.

Compressão de Vídeo Espacial

É um algoritmo computacional que busca informações redundantes dentro de um quadro do vídeo, comprime a informação computacional do quadro (com perda ou sem perda de qualidade).

Compressão de Vídeo Espaço-Temporal

É um algoritmo computacional que busca informações redundantes dentro de um quadro do vídeo e entre quadros consecutivos, comprime a informação redundante desses quadros (com perda ou sem perda de qualidade).

CRT (*Cathode Ray Tube*)

O tubo de raios catódicos, também conhecido como cinescópio, é a tela usada em muitos monitores, televisores, e osciloscópios.

DCT (*Discret Cossine Transform*)

A transformada discreta de cossenos é uma simplificação da transformada rápida de Fourier (FFT), sendo aplicada como uma etapa da codificação do vídeo digital.

DivX

É um CODEC de vídeo criado pela empresa DivX Inc que utiliza compressão de vídeo espaço-temporal com a transformada DCT. Utiliza espaço de cor YCbCr com amostragem 4:1:1 e 4:2:2.

DV (*Digital Video*)

É um CODEC de vídeo que utiliza espaço de cor YCbCr e amostragem 4:1:1, com codificação de vídeo espaço temporal e transformada DCT. É a codificação que diversas câmeras digitais com resolução de vídeo 720x480 pixels gravam suas informações em fitas magnéticas.

DWT (*Discret Wavelet Transform*)

A transformada discreta de wavelets é o nome utilizado para qualquer transformada wavelet com amostragem discreta.

Espaço de Cor

É um modelo abstrato matemático para formalizar a descrição de cores através de tuplas de números, tipicamente formadas por três ou quatro elementos.

Estereoscopia

Capacidade de interpretação de um par de imagens estéreo para visualização de uma imagem tridimensional.

Estereoscópio

Aparelho que através do uso de ilustrações ou fotos, permite fazer uso da estereoscopia, a capacidade de reprodução da visão binocular humana a partir de duas imagens que compõe um par estéreo. Foi inventado em 1838.

FFT (*Fast Fourier Transform*)

A transformada rápida de Fourier é um algoritmo matemático que transforma uma onda ou sinal no domínio do tempo em uma seqüência de coeficientes, baseados em uma base ortogonal de pequenas e finitas ondas, em uma somatória de senos e cossenos. A transformada pode facilmente ser estendida a processamentos multidimensionais, como imagens, onde o domínio do tempo passa a ser o domínio do espaço.

GIF (*Graphics Interchange Format*)

É um formato de arquivos de imagem digital, criado em 1987 pela CompuServe. Este formato suporta 8 bits por pixel, permitindo uma palheta de 256 cores distintas. Sua compressão de dados não traz perda de qualidade à imagem (LZW).

H264

H.264 é um CODEC de vídeo baseado no MPEG-4 *Part 10* ou AVC (*Advanced Video Coding*). O padrão utiliza espaço de cor YCbCr com amostragem variável e transformada DCT. A versão final, formalmente chamada por ISO/IEC 14496-10, foi lançada em Maio de 2003.

Huffman

É um método de compressão que usa as probabilidades de ocorrência dos símbolos no conjunto de dados a ser comprimido para determinar códigos de tamanho variável para cada símbolo. Ele foi desenvolvido em 1952 por David A. Huffman.

HUFFYUV

É um CODEC de vídeo que trabalha com espaço de cor YCbCr e amostragem 4:4:4, sobre o qual é aplicada a codificação de huffman, sendo sua codificação sem perda de dados.

JPEG (*Joint Picture Experts Group*)

É um formato de arquivos de imagem digital, que utiliza DCT e quantização variável para compressão de dados. O usuário determina se sua compressão de dados pode ou não ter perda.

JPEG2000 (*Joint Picture Experts Group, v. 2000*)

É um formato de arquivos de imagem digital, que utiliza DWT e quantização variável para compressão de dados. O usuário determina se sua compressão de dados pode ou não ter perda.

LCD (*Liquid Crystal Display*)

Um monitor de cristal líquido é um monitor muito leve e fino, sem partes móveis. Consiste de um líquido polarizador da luz, eletricamente controlado, que se encontra comprimido dentro de celas entre duas lâminas transparentes polarizadoras.

LZW (*Lempel-Ziv-Welch*)

É um algoritmo de compressão de dados, derivado do algoritmo LZ78, baseado na localização e no registro dos padrões de uma estrutura, desenvolvido e patenteado em 1984. Nas imagens, o algoritmo lê os valores de pixels de uma imagem digital e elabora uma tabela de códigos onde se representam os padrões repetidos dos pixels encontrados.

MJPEG (*Motion JPEG*)

É um CODEC de vídeo que utiliza o algoritmo de compressão JPEG codificando vídeos como seqüências de imagens.

MJPEG2000 (*Motion JPEG2000*)

É um CODEC de vídeo que utiliza o algoritmo de compressão JPEG2000 codificando vídeos como seqüências de imagens.

MPEG-1 (*Motion Picture Experts Group, version 1*)

Conjunto de padrões de compressão com perda de áudio e de vídeo, criado em 1988. Sua compressão de dados define sua qualidade como similar a VHS. Utiliza um espaçamento de cor inferior ao YCbCr 4:1:1 e DCT.

MPEG-2 (*Motion Picture Experts Group, version 2*)

Conjunto de padrões de compressão com perda de áudio e vídeo, mas com resultados melhores que seu antecessor (MPEG-1). É utilizado em DVDs e transmissões de TV a cabo e satélite. Utiliza espaço de cor 4:2:2 ou 4:2:0 e DCT, num algoritmo de compressão espaço-temporal de vídeo.

MPEG-4 (*Motion Picture Experts Group, version 4*)

É um conjunto de padrões para arquivos multimídia. Contém diversas camadas, nas quais podemos encontrar dois algoritmos de compressão de vídeo (MPEG4 *Part 2* e MPEG4 *Part 10*). Nessas camadas utiliza-se espaço de cor YCbCr com amostragem variável e DCT, num algoritmo de compressão espaço-temporal de vídeo.

MOS (*Mean Opinion Score*)

É um índice de medida de qualidade subjetiva, calculado pela média da percepção visual de um conjunto de indivíduos. O MOS é expresso num número de 1 a 5, onde 1 seria uma qualidade ruim e 5 a melhor.

MOV

Formato ou contêiner de vídeo utilizado para encapsular áudio e vídeo, criado pela Apple Inc.

MSU

Grupo de computação matemática e cibernética, do departamento da Universidade do Estado de Moscou, responsáveis pelo desenvolvimento de diversas ferramentas para manipulação e testes de vídeo digital.

NTSC (*National Television Systems Committee*)

NTSC é o sistema de televisão analógico atualmente em uso nos Estados Unidos e na maioria dos países da América (o Brasil é uma das exceções que utilizam esse sistema apenas em DVDs) e em alguns países do leste asiático.

PAL (*Phase Alternating Line*)

A linha de fase alternante é uma forma de codificação da cor usada nos sistemas de transmissão televisiva analógica, usado em muitos países da Europa e da Ásia.

Par Estéreo ou Par Estereoscópico

Par de imagens que simulam a captação de imagens dos dois olhos humanos. Esse efeito pode ser simulado com ilustrações ou com imagens captadas por câmeras ligeiramente afastadas (que simulam a visão binocular humana).

Paralaxe

É a distância entre os pontos correspondentes das imagens do olho direito e do esquerdo em uma imagem projetada.

Pixel (*Picture Element*)

É o menor elemento num dispositivo de exibição, ao qual é possível atribuir-se uma cor.

Player (reprodutor de mídia)

Software que reproduz um determinado de mídia armazenada digitalmente.

PSNR (*Peak-to-Peak Noise Ratio*)

É um índice de qualidade objetiva (medido matematicamente), dado pela relação entre a entrada e a saída de um processo de compressão de imagens com perdas. Avalia a quantidade de ruídos introduzidos na compressão, tendo como base de comparação a imagem original.

RGB (*Red, Blue, Green*)

É um espaço de cor aditivo, formado pela combinação das cores Vermelho, Verde e Azul.

SSIM Index

É um índice de qualidade objetiva (medido matematicamente). Seu valor é baseado na medida de três componentes (similaridade de luminância, similaridade de contraste e de similaridade estrutural) combinadas em um valor final.

Visualização Anaglífica ou Visualização por Anaglífos

Técnica de visualização estereoscópica na qual um espectador utilizando de óculos com lentes coloridas observa imagens que compõe pares estereoscópicos sem fidelidade de cores (anaglífos).

Visualização Estereoscópica

Técnica que permite a visualização de imagens em movimento com profundidade, e varia conforme a tecnologia de reprodução. Pode variar de óculos com lentes coloridas em um monitor bidimensional até monitores autoestereoscópicos, onde se pode observar o efeito estereoscópico sem o uso de óculos.

Visualização por Luz Polarizada

Técnica de visualização estereoscópica na qual dois projetores com filtros polarizadores de luz projetam suas imagens sobre uma tela metalizada. O espectador utilizando de óculos com lentes com filtros polarizadores de luz observa imagens que compõe pares estereoscópicos com um alto grau de qualidade.

Visualização por Multiplexação

Técnica de visualização estereoscópica na qual um par de vídeo estereoscópico é codificado para ser exibido em um monitor alternando seus quadros, sendo observado com óculos com lentes de cristal líquido que alterna a visão do olho direito e esquerdo em sincronia com o quadro exibido.

VQM

É um índice de qualidade objetiva (medido matematicamente), que mostra os erros gerados com a DCT no processo de codificação.

X264

É um CODEC de vídeo livre baseado no MPEG-4 Part 10 ou AVC (*Advanced Video Coding*). O padrão utiliza espaço de cor YCbCr com amostragem variável e transformada DCT.

Xvid

É um CODEC de vídeo livre baseado no MPEG-4 Part 2. O padrão utiliza espaço de cor YCbCr com amostragem 4:1:1 ou 4:2:2 e transformada DCT.

Wavelet

Também conhecida por ondaleta, é uma função capaz de decompor e descrever outras funções no domínio da frequência, de forma a podermos analisar estas funções em diferentes escalas de frequência e de tempo. A decomposição de uma função com o uso de wavelets é conhecida como transformada de wavelet e tem suas variantes contínua e discreta. Graças à capacidade de decompor as funções tanto no domínio da frequência quanto no domínio do tempo, as funções wavelet são ferramentas poderosas para a análise de sinais e compressão de dados.

YCbCr

É um espaço de cor composto por luminância (Y) e cromaticidade do verde em relação ao azul (Cb) e ao vermelho (Cr). É comumente utilizado com redução das componentes de cromaticidade, para diminuição do volume de dados processados.