RevGlyph – codificação e reversão estereoscópica anaglífica

Matheus Ricardo Uihara Zingarelli

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura:_____

RevGlyph – codificação e reversão estereoscópica anaglífica

Matheus Ricardo Uihara Zingarelli

***Orientador*: Prof. Dr. Rudinei Goularte**

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação - ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Ciências - Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA*.

USP – São Carlos
Novembro de 2013

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados fornecidos pelo(a) autor(a)

Z77r

Zingarelli, Matheus Ricardo Uihara
RevGlyph: codificação e reversão estereoscópica anaglífica / Matheus Ricardo Uihara Zingarelli; orientador Rudinei Goularte. -- São Carlos, 2013.
106 p.

Dissertação (Mestrado - Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional) -- Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2013.

1. CODIFICAÇÃO ANAGLÍFICA. 2. CODIFICAÇÃO DA INFORMAÇÃO. 3. ESTEREOSCOPIA. 4. COMPRESSÃO. I. Goularte, Rudinei, orient. II. Título.

Dedico este trabalho a meus pais, principais coautores e suportes fundamentais para finalizar mais esta etapa de minha vida.

Agradecimentos

Primeiro e acima de tudo agradeço a Deus pelo amparo, guia e proteção. Obrigado por me ajudar a perseverar na caminhada e a me manter sereno mesmo em inúmeros momentos de indecisão e quase desistência.

A meus pais, por mais uma vez suportarem e me apoiarem em mais estes anos de estudo. Obrigado pela confiança e companhia, e me desculpem pelas diversas alterações de humor e ausências que surgiram durante todo o período.

Ao Tobi, pela amizade muda, deitado no tapete ao lado do computador. Sua ausência ainda dói.

A meu orientador, Professor Doutor Rudinei Goularte, por ter me guiado desde meus primeiros passos no universo da pesquisa acadêmica. Obrigado pela paciência, correções e tempo dedicado até a conclusão deste trabalho.

Ao Professor Doutor Leonardo Andrade, cujo trabalho solidificou os rumos de minha pesquisa.

A todo pessoal do Intermídia (e agregados), laboratório no qual desenvolvi o trabalho, pelo companheirismo e momentos de descontração, tornando a jornada mais leve e divertida.

À Universidade de São Paulo, por ter me dado o privilégio de uma formação sólida e oferecido toda estrutura e apoio necessários desde minha graduação. Fico feliz de agora ser parte de sua máquina administrativa e poder colaborar indiretamente na formação de novos alunos.

Ao CNPq, pelo apoio financeiro na maior parte do curso.

Por fim, agradeço a todos os que de alguma forma contribuíram para o andamento e finalização deste Mestrado.

Two roads diverged in a wood, and I —
I took the one less traveled by,
And that has made all the difference.

Robert Frost
The Road Not Taken

Resumo

ZINGARELLI, M. R. U. **RevGlyph – codificação e reversão estereoscópica anaglífica**. 2013. 96f. Monografia de defesa (Mestrado) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2013.

A atenção voltada à produção de conteúdos 3D atualmente tem sido alta, em grande parte devido à aceitação e à manifestação de interesse do público para esta tecnologia. Isso reflete num maior investimento das indústrias cinematográfica, de televisores e de jogos visando trazer o 3D para suas produções e aparelhos, oferecendo modos diferentes de interação ao usuário. Com isso, novas técnicas de captura, codificação e modos de reprodução de vídeos 3D, em especial, os vídeos estereoscópicos, vêm surgindo ou sendo melhorados, visando aperfeiçoar e integrar esta nova tecnologia com a infraestrutura disponível. Entretanto, notam-se divergências nos avanços feitos no campo da codificação, com cada método de visualização estereoscópica utilizando uma técnica de codificação diferente. Isso leva ao problema da incompatibilidade entre métodos de visualização. Uma proposta é criar uma técnica que seja genérica, isto é, independente do método de visualização. Tal técnica, por meio de parâmetros adequados, codifica o vídeo estéreo sem nenhuma perda significativa tanto na qualidade quanto na percepção de profundidade, característica marcante nesse tipo de conteúdo. A técnica proposta, denominada RevGlyph, transforma um par estéreo de vídeos em um único fluxo anaglífico, especialmente codificado. Tal fluxo, além de ser compatível com o método anaglífico de visualização, é também reversível a uma aproximação do par estéreo original, garantindo a independência do método de visualização.

Abstract

ZINGARELLI, M. R. U. **RevGlyph – stereoscopic coding and reversing of anaglyphs**. 2013. 96f. Monografia de defesa (Mestrado) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2013.

Attention towards 3D content production has been currently high, mostly because of public acceptance and interest in this kind of technology. That reflects in more investment from film, television and gaming industries, aiming at bringing 3D to their content and devices, as well as offering different ways of user interaction. Therefore, new capturing techniques, coding and playback modes for 3D video, particularly stereoscopic video, have been emerging or being enhanced, focusing on improving and integrating this new kind of technology with the available infrastructure. However, regarding advances in the coding area, there are conflicts because each stereoscopic visualization method uses a different coding technique. That leads to incompatibility between those methods. One proposal is to develop a generic technique, that is, a technique that is appropriate regardless the visualization method. Such technique, with suitable parameters, outputs a stereoscopic video with no significant loss of quality or depth perception, which is the remarkable feature of this kind of content. The proposed technique, named RevGlyph, transforms a stereo pair of videos into a single anaglyph stream, coded in a special manner. Such stream is not only compliant with the anaglyph visualization method but also reversible to something close to the original stereo pair, allowing visualization independence.

Índice de Figuras

Figura 1 - Exemplo de câmera de gravação estereoscópica 24

Figura 2 - Exemplo de observância da informação de disparidade (STEREOGRAPHICS, 1997). Em (A), quando é focada a visão no dedo polegar, a bandeira aparece duplicada ao fundo. Em (B), quando foca-se na bandeira, o dedo polegar aparece duplicado. Esse efeito de duplicidade é resultado da disparidade binocular 30

Figura 3 - Divisão dos métodos de visualização estereoscópica – adaptado de Urey *et al.*(2011) 32

Figura 4 - Processo de conversão anaglífica verde-magenta. É retirada a cor verde da imagem correspondente à visão do lado direito e as cores vermelha e azul (magenta) da imagem correspondente à visão do lado esquerdo.....33

Figura 5 - Exemplo do efeito de polarização. O filtro situado no meio faz com que a onda de luz siga um padrão linear vertical. Um outro filtro polarizador em outra fonte de projeção pode ser utilizado para fazer com que a onda siga um padrão linear horizontal. Isso possibilita que óculos com lentes polarizadas façam a separação de cada vídeo 35

Figura 6 - Oculus Rift, dispositivo voltado para jogos que promete imersão total em um ambiente de realidade virtual.....37

Figura 7 - Exemplos de sistemas utilizados nos métodos autoestereoscópicos. Em (a), as lentículas são responsáveis por desviar a luz para cada olho. Em (b), pequenas barreiras impedem a passagem da luz em certas direções.....39

Figura 8 - Tipos de subamostragem de crominância 44

Figura 9 - Processo de codificação utilizando vídeo e mapa de profundidades para a formação de um vídeo estereoscópico (SMOLIC et al., 2009).....48

Figura 10 - Predição entrevisões. Uma das visões se torna a base e é utilizada para predizer a outra visão do par estéreo (VETRO; WIEGAND; SULLIVAN, 2011).....51

Figura 11 - Resultado da replicação dos canais de cores de um anáglifo verde-magenta na tentativa de recriar o par estéreo 58

Figura 12 - Exemplo de redundância existente entre as componentes de luminância dos anáglifos formados a partir de um par estereoscópico 61

Figura 13 - Exemplo da aplicação do agrupamento por limiar de aceitação nos valores da Diferença de Luminâncias 62

Figura 14 – RevGlyph: processo de conversão anaglífica utilizando o Cabeçalho de Consulta de Cores e a Diferença de Luminâncias.....63

Figura 15 – RevGlyph: Processo de reversão anaglífica 65

Figura 16 - Comparação qualitativa do par estéreo original (A) e o obtido pelo processo de reversão anaglífica utilizando apenas o Cabeçalho de Consulta de Cores (B). É possível observar o *crosstalk* em (B), representado em tons de verde e magenta.....68

Figura 17 - Comparação entre uma imagem original (A) e sua correspondente processada com subamostragem 4:4:0 (B). Imagem retirada da base Corel1000, idBeach59.jpg 74

Figura 18 - Par estéreo original da base de imagens estereoscópicas. Código de identificação rv03...78

Figura 19 - Par estéreo revertido utilizando a RevGlyph sem limiar. As mudanças visuais são pouco aparentes.....79

Figura 20 - Par estéreo revertido utilizando a RevGlyph - limiar 1.....80

Figura 21 - Par estéreo revertido utilizando a RevGlyph - limiar 2.....81

Figura 22 - Par estéreo revertido utilizando a RevGlyph - limiar 3. Pequenas alterações de cor e textura podem ser notadas nas áreas circuladas. 82

Figura 23 - Par estéreo revertido utilizando a RevGlyph - limiar 4. As alterações começam a ficar mais aparentes 83

Figura 24 - Par estéreo revertido utilizando o RevGlyph - limiar 5. É possível notar com nitidez a formação de texturas com linhas horizontais. 84

Figura 25 - Ampliação de uma região da imagem estereoscópica revertida com diferentes limiares, apresentando a degradação sofrida com o aumento do limiar e a formação de ruídos na forma de linhas horizontais.....85

Figura 26 - Anáglifo verde-magenta criado usando um par estéreo original. Código de identificação rv03 86

Figura 27 - Anáglifo verde-magenta criado usando um par estéreo revertido pela RevGlyph com limiar 2 86

Figura 28 – Uma das visões de cada imagem utilizada para comparação do RevGlyph com o Mapa de Profundidades. Cones (A), Teddy (B), Venus (C) e Sawtooth (D).....88

Figura 29 - Exemplo de par estéreo revertido utilizando Fluxo Óptico 93

Índice de Gráficos

Gráfico 1 - Distribuição do PSNR calculado para imagens da biblioteca Corel1000, processadas apenas pela conversão RGB→YCbCr→RGB.....72

Gráfico 2 - Distribuição do PSNR calculado para imagens da biblioteca Corel1000, processadas pela conversão RGB→YCbCr→RGB e subamostragem de crominância.....73

Gráfico 3 - Gráfico de comparação da taxa de compressão e PSNR obtido em cada limiar aplicado com a técnica RevGlyph 76

Índice de Tabelas

Tabela 1 - Métodos de visualização estereoscópica relacionados a partir de seu custo de aquisição, qualidade visual e compressão.....40

Tabela 2 - PSNR da conversão RGB→YC_bC_r→ RGB 71

Tabela 3 - PSNR da conversão RGB→YC_bC_r→ RGB juntamente com subamostragem de crominância 73

Tabela 4 - Resultados da utilização da técnica RevGlyph com diferentes limiares para a compressão da Diferença de Luminâncias.....76

Tabela 5 - Resultados da aplicação da RevGlyph em amostras do banco de imagens de Middlebury .88

Tabela 6 - Comparação do PSNR entre a RevGlyph e o mapa de profundidades. Valores em dB. 89

Tabela 7 - Comparação da compressão entre a RevGlyph e o mapa de profundidades 90

Sumário

1. Introdução 23
2. Fundamentos da visualização estereoscópica 29
2.1. Informações de profundidade estereoscópicas 29
2.2. Separação do par estéreo 31
2.3. Visualização utilizando métodos por multiplexação 32
2.3.1. Multiplexação de cor – Método Anaglífico 33
2.3.2. Multiplexação por polarização 34
2.3.3. Multiplexação temporal 35
2.4. Visualização utilizando HMD 36
2.5. Visualização utilizando métodos autoestereoscópicos 38
2.6. Considerações finais 40
3. Codificação e compressão estereoscópica 41
3.1. Espaço de cores e subamostragem de crominância 42
3.2. Organização de dados estereoscópicos 45
3.2.1. Método de Lipton 46
3.2.2. Método multivisão 46
3.2.3. Métodos baseados em vídeo e profundidade 47
3.3. Compressão 49
3.3.1. Compressão 2D 49
3.3.2. Compressão Estereoscópica 51
3.4. Trabalhos relacionados 52
3.5. Considerações finais 54
4. RevGlyph 57
4.1. O problema da reversão anaglífica 57
4.2. Processo de codificação anaglífica 58
4.2.1. Cabeçalho de Consulta de Cores 59
4.2.2. Diferença de Luminâncias 60

4.3. Reversão anaglífica 64
4.4. Implementação da técnica 65
4.5. Considerações finais 67
5. Análise da RevGlyph 69
5.1. Análise da conversão de espaço de cores 69
5.2. Análise da subamostragem de crominância 72
5.3. Análise da Diferença de Luminâncias 74
5.4. Comparação entre a RevGlyph e Mapa de Profundidades 87
5.5. Considerações finais 90
6. Conclusões 91
6.1. Oportunidades para novas pesquisas 92
6.2. Resultados 93
6.3. Opinião 94
Referências 97
Glossário 103

1. Introdução

Com a estreia de Avatar[1] em 2009, os chamados "filmes 3D" voltaram ao interesse do público, iniciando uma nova etapa na produção cinematográfica que se estende até hoje. A partir de tal interesse e com o amadurecimento da tecnologia, a indústria do cinema tem aderido ao 3D em suas principais produções como uma das maneiras de reconquistar o público. Seguindo esse interesse, é notável a movimentação de outros ramos do mercado no desenvolvimento para criação de televisores e telas que reproduzam conteúdo 3D com alta qualidade e definição, trazendo essa nova forma de interação para os lares e dispositivos móveis dos consumidores.

A partir daí, as possibilidades são muitas, desde assistir a um simples conteúdo 3D a até mesmo se movimentar pelo ambiente e ver o cenário de diferentes ângulos, inclusive enxergando objetos antes ocultos (CHEN *et al.*, 2012). Num futuro mais distante, pode-se prever total imersão do telespectador no conteúdo reproduzido, aliado a tecnologias como hologramas (SVOBODA *et al.*, 2013) e dispositivos táteis em ambientes de realidade aumentada (STRATER *et al.*, 2012).

O meio acadêmico também demonstra interesse em técnicas estereoscópicas como forma de apoio no desenvolvimento de pesquisas em outras áreas, como, por exemplo, em problemas envolvidos na visão artificial de robôs (KIM *et al.*, 2007) e dispositivos de realidade aumentada (STRATER *et al.*, 2012). Outros ramos da ciência além da computação também utilizam estereoscopia, como a medicina, por meio do desenvolvimento de tecnologias assistivas para deficientes visuais (COSTA *et al.*, 2012) e análise de imagens de raios X (TOMKOWIAK; VAN LYSEL; SPEIDEL, 2013).

Em termos técnicos, os vídeos 3D que possibilitam a percepção de profundidade são definidos como vídeos estereoscópicos (STEREOGRAPHICS CORPORATION, 1997) e são visualizados com a transmissão de dois pontos de vista diferentes para cada olho do telespectador, o que proporciona sensação de níveis de profundidade na cena. Os dois pontos de vista são resultado da captura simultânea de uma mesma cena feita por duas câmeras deslocadas

[1] http://www.imdb.com/title/tt0499549/. Acesso em 19 nov. 2013.

horizontalmente, gerando dois vídeos ligeiramente diferentes que são conhecidos como par estéreo ou par estereoscópico. A **Figura 1**[2] ilustra um exemplo de câmera especial, com duas lentes, utilizada para capturar vídeos estereoscópicos.

Por meio da reprodução e da visualização adequada do par estéreo, cada olho recebe informações visuais ligeiramente deslocadas em relação ao outro. Tais informações são transmitidas para o cérebro, o qual se encarrega de interpretá-las, remover as redundâncias e uni-las em uma única imagem, obtendo também informações de distância e escala que iludem o sentido da visão humano com uma sensação de profundidade do que é visto no plano da tela.

Figura 1 - Exemplo de câmera de gravação estereoscópica

Com o passar dos anos pôde-se observar avanços significativos na captura, reprodução e visualização estereoscópica, visando melhorar a qualidade e a sensação de profundidade, bem como reduzir a fadiga e os problemas na fusão do par estéreo durante a visualização. Dentre as melhorias, pode-se citar:

a) criação de câmeras especialmente desenvolvidas para captura estereoscópica, como exemplificado na **Figura 1**;
b) desenvolvimento dos métodos denominados *Free Viewpoint Video* (FVV) (SMOLIC *et al.*, 2009), permitindo que a visualização não se limite apenas a um ponto de vista, mas sim a vários, e dando liberdade ao espectador de se deslocar pelo ambiente e obter diferentes informações da cena;

[2] http://panasonic.biz/sav/broch_bdf/AG-3DA1_e.pdf. Acesso em 19 nov. 2013.

c) criação e aperfeiçoamento de técnicas de conversão de vídeos 2D[3] para vídeos estereoscópicos foram criadas e aperfeiçoadas (TAM; ZHANG, 2006);
d) surgimento de novos métodos de visualização, com tecnologias que utilizam óculos especiais para separar o par estéreo (STEREOGRAPHICS CORPORATION, 1997), bem como aparelhos denominados autoestereoscópicos, que permitem assistir a conteúdo estereoscópico sem a necessidade de óculos ou qualquer outro dispositivo (UREY *et al.*, 2011).

Apesar dos avanços obtidos nessas áreas, quando comparados com o avanço na codificação de vídeos estereoscópicos é possível notar que este foi inferior àqueles, limitando-se a melhorias e adaptações baseadas na codificação de vídeos 2D. Um reflexo disso é a atual divergência de formatos para organizar dados de vídeos estereoscópicos para fins de armazenamento ou transmissão. As estratégias existentes para tal organização podem ser divididas em três grupos: métodos envolvendo o par estéreo (LIPTON, 1997), que neste trabalho é denominado "método de Lipton", métodos multivisão e métodos envolvendo vídeo e profundidade (SMOLIC *et al.*, 2009).

No método de Lipton o par estéreo é armazenado em um contêiner de vídeo, com ou sem compressão, o que possibilita a reprodução de vídeos estereoscópicos com pouca ou nenhuma modificação dos sistemas de vídeo atuais, mantendo sua interoperabilidade. Os métodos multivisão envolvem mais de um par estéreo, possibilitando diferentes pontos de vista de uma mesma cena. Por fim, métodos envolvendo vídeo e profundidade armazenam apenas um dos vídeos do par estéreo, juntamente com informações auxiliares (de profundidade e/ou redundância entre o par estéreo) que possibilitam a síntese do segundo vídeo; dessa forma, tais métodos ficam limitados para reprodução em sistemas de vídeo e telas específicas, não sendo compatível com sistemas que necessitam do par estéreo (SMOLIC *et al.*, 2009).

Embora simples, o método de Lipton armazena o par estéreo, o que resulta no dobro de informações quando comparado a vídeos 2D. O cenário piora nos métodos multivisão, pois cada nova visão adicionada implica em mais dados a serem capturados e armazenados. Já os métodos baseados em vídeo e profundidade utilizam estratégias para aumento da compressão explorando novas estruturas de dados envolvendo profundidade e relacionamento entre o par estéreo. Todos os métodos podem utilizar de técnicas de compressão de vídeos 2D, porém, por se tratar muitas vezes de compressão com perdas, ocorre a geração de artefatos, os quais

[3] Ao longo deste trabalho, para fins de diferenciação de vídeos estereoscópicos, será denominado vídeo 2D um vídeo capturado por apenas uma câmera.

impactam negativamente na correta percepção de profundidade, notadamente em vídeo anaglíficos (ANDRADE, 2012).

Diante do observado, pode-se notar a falta técnicas projetadas exclusivamente para codificação de vídeos estereoscópicos que forneçam compressão sem diminuir a qualidade na percepção de profundidade, e que, ao mesmo tempo, sejam genéricas, possibilitando seu uso por qualquer método de visualização, tanto os que necessitam de óculos especiais quanto o autoestereoscópico.

Neste trabalho, propõe-se a utilização de codificação anaglífica (**Seção 4.2**) como meio de obter compressão de imagens estereoscópicas e possibilitar a independência de visualização. O método anaglífico transforma um par estereoscópico de imagens em uma única imagem, a qual, desse modo, pode ser utilizada para fins de armazenamento/transmissão, já que o volume de dados é reduzido pela metade. Mais do que isso, é possível obter compressão adicional quando aplicadas técnicas de compressão sem perdas, como métodos estatísticos (Huffman, por exemplo) ou baseados em dicionários (como LZW).

A codificação anaglífica, embora traga compressão, impossibilita a exibição de uma imagem ou vídeo estereoscópico por outros modos de visualização, como, por exemplo, os que utilizam óculos obturadores ou polarizadores, pois estes necessitam do par estéreo para realizar a projeção. Para que a codificação utilizando o método anaglífico seja independente do modo de visualização, é necessário que se possa restaurar o par estéreo a partir de um anáglifo[4].

A restauração do par estéreo por meio de um anáglifo, aqui denominado "Problema da Reversão Anaglífica", não possui solução trivial, pois muita informação é descartada durante a codificação e precisa ser recuperada. Na literatura relacionada, até o momento, não foram encontradas pesquisas associadas a este problema, o que constitui um desafio científico atualmente carente de investigação.

Assim, este trabalho tem como objetivo propor uma solução para o Problema da Reversão Anaglífica por meio do desenvolvimento da técnica RevGlyph. As etapas de desenvolvimento da técnica, os resultados obtidos e a comparação com o estado da arte são descritos nos próximos capítulos.

Os Capítulos 2 e 3 são complementos a esta Introdução, servindo como base para entendimento da área e apresentando uma revisão bibliográfica dos temas centrais. O Capítulo 2 aborda fundamentos da visualização estereoscópica, apresentando os diferentes métodos de

[4] Um anáglifo é um vídeo resultante da aplicação do método anaglífico a um par estéreo.

visualização disponíveis. Já o Capítulo 3 se concentra nas técnicas de codificação e compressão estereoscópica, fornecendo mais detalhes dos problemas de compatibilidade aqui citados e incluindo trabalhos recentes relacionados à área. A técnica RevGlyph é apresentada no Capítulo 4, incluindo as mudanças necessárias no processo de codificação anaglífica, bem como a descrição das duas estruturas desenvolvidas para possibilitar a reversão anaglífica. Por fim, análises e discussão sobre a técnica, além de uma comparação desta com o estado da arte, são feitos no Capítulo 5.

Vale ressaltar que este trabalho contém conceitos e uma revisão do estado da arte associado tanto a imagens quando a vídeos estereoscópicos. Entretanto, considerando o pioneirismo da pesquisa, realizada em uma área até então não explorada, e levando em conta a duração de um programa de Mestrado, a técnica RevGlyph foi implementada e aplicada somente a imagens estereoscópicas.

As análises e exemplos a serem apresentados envolvem observações com relação à qualidade de cores e percepção de profundidade em anáglifos verde-magenta. Dessa forma, é aconselhável que a leitura deste trabalho seja feita por meio de sua versão digital em PDF[5]. Para verificar a percepção de profundidade, é necessária a utilização de óculos anaglíficos com lentes verde (lado esquerdo) e magenta (lado direito).

[5] Uma cópia da versão digital pode ser pedida pelo e-mail: zingarelli.m@gmail.com

2. Fundamentos da visualização estereoscópica

Este Capítulo aborda aspectos fundamentais da área de visualização estereoscópica que foram necessários para o desenvolvimento do trabalho. Serão explicadas características por trás da percepção de profundidade em vídeos estereoscópicos, bem como métodos e dispositivos existentes para realizar a separação e a transmissão correta dos vídeos do par estéreo. Analisando cada método será possível observar que todos têm em comum a necessidade do par estéreo, porém, este pode se apresentar em diferentes formatos, muitas vezes incompatíveis entre métodos.

2.1. Informações de profundidade estereoscópicas

Nossos olhos estão distantes aproximadamente 6,5 cm um do outro, movimentam-se em conjunto para uma mesma direção e cada um possui um ângulo de visão limitado. Por se estarem dispostos na horizontal, os elementos visuais captados por cada olho são semelhantes, porém, chegam a cada retina em posições longitudinais diferentes, ligeiramente deslocados, o que gera dois pontos de vista. A diferença entre as posições dos elementos em cada ponto de vista é denominada disparidade binocular (STEREOGRAPHICS CORPORATION, 1997).

A **Figura 2** ilustra o conceito de disparidade binocular em um exemplo que pode ser facilmente executado. Ao colocar o dedo entre os olhos e um objeto e focar a atenção no dedo, o objeto e quaisquer elementos visuais mais distantes aparecem como que duplicados (**Figura 2 A**). Isso acontece porque o dedo se encontra no ponto de convergência das duas retinas, isto é, ele tem sua imagem projetada no mesmo ponto em cada retina. Já o objeto está fora do ponto de convergência, sendo sua imagem projetada em pontos diferentes em cada retina, dando a impressão de estar duplicado. Esse efeito de duplicidade é resultado da disparidade binocular.

O inverso também ocorre, com um segundo dedo “fantasma” aparecendo se o foco for no objeto (**Figura 2 B**).

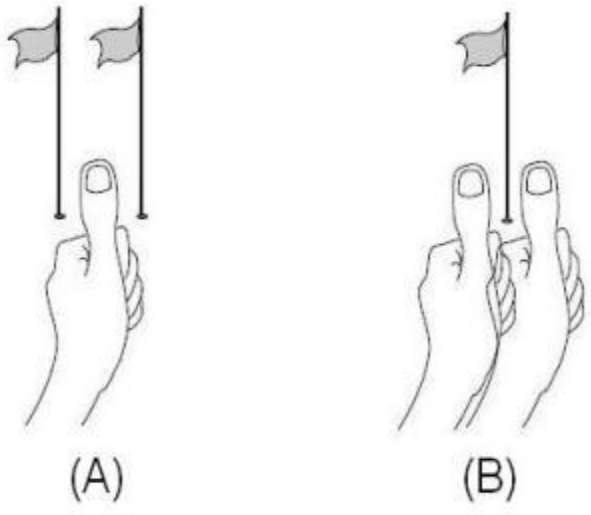


Figura 2 - Exemplo de observância da informação de disparidade (STEREOGRAPHICS, 1997). Em (A), quando é focada a visão no dedo polegar, a bandeira aparece duplicada ao fundo. Em (B), quando foca-se na bandeira, o dedo polegar aparece duplicado. Esse efeito de duplicidade é resultado da disparidade binocular

Pelo fato das informações visuais chegarem a cada olho ligeiramente deslocadas, poderia se chegar à conclusão de que deveriam ser enxergadas duas imagens separadas, e não apenas uma como ocorre na visão humana. Além disso, somos capazes de perceber diferentes níveis de profundidade e distância entre os objetos que vemos. A essa sensação de profundidade, em virtude da disparidade binocular, dá-se o nome de estereopsia, e só é obtida com o uso dos dois olhos (WHEATSTONE, 1838).

Diretamente ligado ao conceito de disparidade existe a paralaxe, que é a distância relativa da imagem de um objeto em um conjunto de figuras (MENDIBURU, 2009). Em outras palavras, dado um par estéreo, a paralaxe seria a diferença entre a posição da imagem de um objeto na figura do lado esquerdo do par em relação ao mesmo objeto na figura do lado direito do par. A diferença entre disparidade e paralaxe é que a primeira é medida na retina e a segunda é medida em um plano de projeção (uma tela ou papel, por exemplo). Ao tentar assistir a um filme 3D no cinema sem o uso dos óculos, é possível perceber a paralaxe na forma de imagens que parecem estar tremidas, com regiões duplicadas e sobrepostas.

A paralaxe é um conceito chave para a produção de vídeos estereoscópicos, pois quando utilizada corretamente é capaz de gerar pontos de vista diferentes de uma mesma imagem para cada olho, tendo como consequência a formação da disparidade, e esta, por conseguinte, produzindo o efeito de estereopsia (STEREOGRAPHICS CORPORATION, 1997).

Quando a paralaxe é utilizada de maneira incorreta ou não calibrada, provoca uma divisão incompleta das informações do par estéreo durante a visualização, o que acarreta na invasão de elementos visuais de uma das figuras do par na outra figura. Isso significa que o olho direito vê também parte dos dados da imagem esquerda e vice-versa. Esse fenômeno é definido como *crosstalk* e causa efeitos indesejáveis e desconforto na visualização estereoscópica (WOODS, 2011). O resultado percebido pelo *crosstalk* é a aparição de um "fantasma" ao redor de alguns objetos de uma imagem, sendo esse fantasma criado pelas informações invasoras advindas da divisão incompleta do par estéreo.

A disparidade binocular, a paralaxe e a estereopsia são classificadas como informações de profundidade estereoscópicas (do inglês *stereoscopic depth cues*), pois com elas é possível obter as informações necessárias para o cérebro interpretar profundidade em pares estereoscópicos. A correta interpretação dessas informações depende da entrega a cada olho da imagem do par estéreo correspondente a ele, e as técnicas disponíveis para isso se encontram descritas na **Seção 2.2**. Existem outras informações visuais que contribuem para a percepção de profundidade, como a informações monoculares (LIPTON, 1982), por exemplo, que estão fora do escopo deste trabalho.

2.2. Separação do par estéreo

A **Seção 2.1** apresentou os três fatores principais envolvidos na visualização estereoscópica. Todos eles apresentam como requisito a utilização dos dois olhos para que a percepção de profundidade aconteça, de modo que cada olho receba um ponto de vista ligeiramente diferente da imagem. A reprodução de um vídeo estereoscópico requer que o par estéreo seja corretamente separado, de modo que seja entregue a informação correta para cada olho, evitando os efeitos do *crosstalk*.

Na literatura é possível encontrar alguns métodos que podem ser utilizados para a separação do par estéreo. Os mais tradicionais envolvem a utilização de óculos especiais ou aparatos semelhantes; já métodos mais recentes, como o autoestereoscópico, realizam a separação sem a necessidade de nenhum instrumento auxiliar. Neste trabalho, utilizaremos a classificação proposta por Urey *et al*. (2011), ilustrada na **Figura 3**.

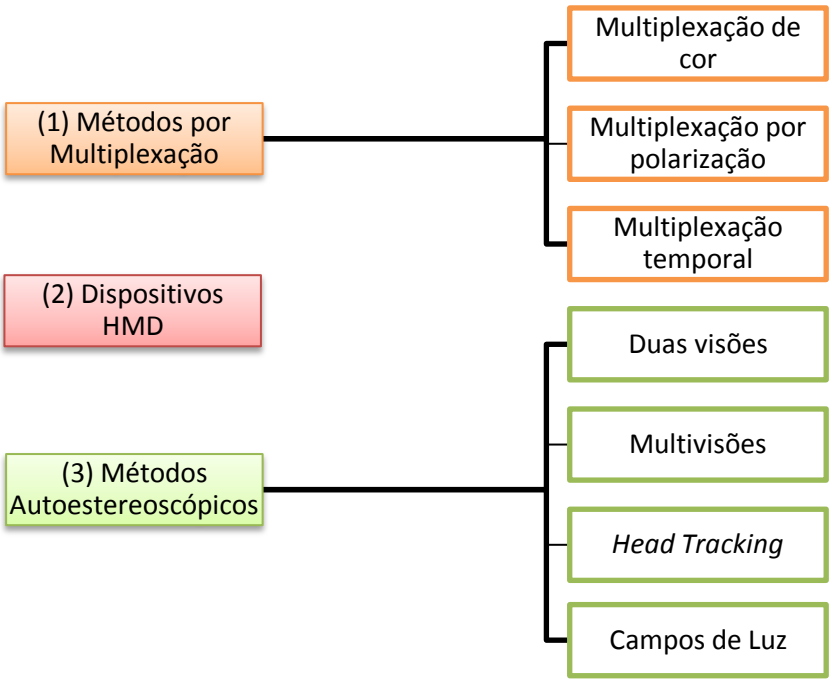


Figura 3 - Divisão dos métodos de visualização estereoscópica – adaptado de Urey *et al.*(2011)

A divisão proposta separa os métodos de visualização estereoscópica em três classes: (1) Métodos por multiplexação, que também denominamos métodos tradicionais, subdivididos em multiplexação de cor, multiplexação por polarização e multiplexação temporal; (2) Dispositivos *Head-Mounted Display* (HMD), um aparato em forma de capacete em que são utilizadas duas telas, uma para cada olho; e (3) métodos autoestereoscópicos, subdivididos pela técnica empregada (duas visões, multivisão, *head tracking*, e campos de luz). Cada classe é discutida nas **Seções de 2.3** a **2.5**, sendo que o método anaglífico, por ser parte fundamental do trabalho, é o exemplo do método de multiplexação de cor descrito.

2.3. Visualização utilizando métodos por multiplexação

Nesta classe de métodos a separação do par estéreo é feita com o auxílio de óculos que possuem lentes especiais, atuando como filtros. São os mais tradicionais e atualmente os mais utilizados no mercado devido à tecnologia conhecida e ao custo baixo envolvido quando comparados às outras duas classes.

2.3.1. Multiplexação de cor – Método Anaglífico

Dentre todos, é o método mais simples e pode ser utilizado em aparelhos de projeção e televisores comuns, não necessitando de hardware adicional. Por se tratar de um método que envolve apenas cores, pode ser feito inclusive em mídia impressa. Foi utilizado na primeira tentativa dos cinemas em reproduzir filmes em 3D durante a década de 1920 (LIPTON, 1982).

O método consiste em fundir as duas imagens de um par estéreo por meio da mistura de canais de cor primária de cada uma, resultando em uma única imagem. A **Figura 4** ilustra esse processo em um par estéreo codificado no espaço de cores RGB. Do par estéreo ilustrado, foi utilizado o canal de cor verde (G_1) da imagem do lado direito e os canais de cor vermelha e azul (R_2 e B_2) da imagem do lado esquerdo para compor uma nova imagem ($R_2G_1B_2$). Pela mistura de canais utilizados, a nova imagem é definida como um anáglifo verde-magenta. Outros tipos de anáglifos possíveis utilizando outras combinações são o vermelho-ciano ($R_1G_2B_2$) e azul-amarelo ($R_2G_2B_1$).

Na visualização, o espectador usa um par de óculos especiais, cujas lentes filtram os canais de cor corretos para cada olho. Seguindo o conjunto de cores escolhidos no anáglifo da **Figura 4**, o par de óculos nesse caso possuiria uma lente verde (para o olho esquerdo) e outra lente magenta, junção da cor vermelha com a cor azul (para o olho direito). Desse modo, o par estéreo é separado e cada olho enxerga as informações espaciais corretas, com as lentes servindo tanto para a separação do par estéreo quanto como complemento às cores faltantes em cada imagem.

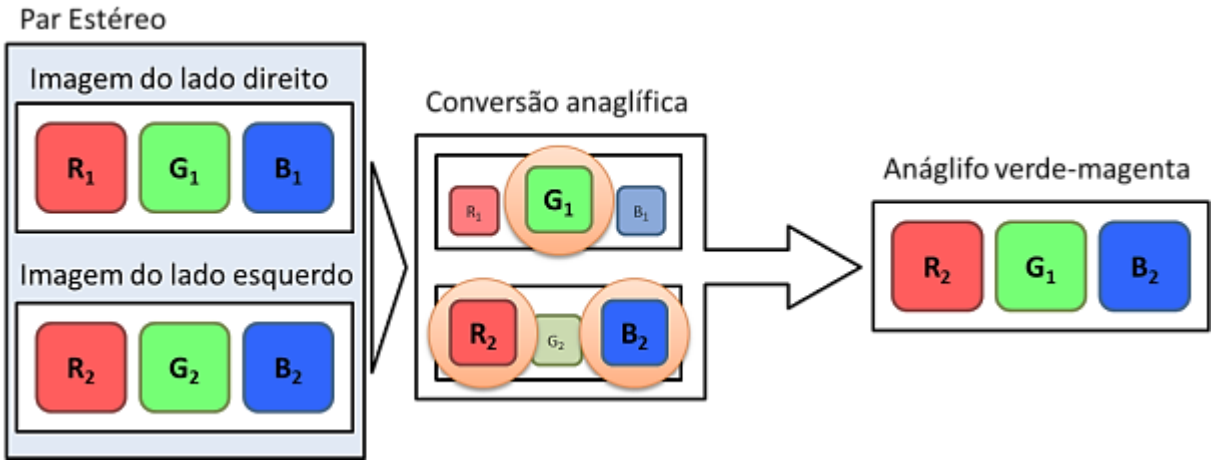


Figura 4 - Processo de conversão anaglífica verde-magenta. É retirada a cor verde da imagem correspondente à visão do lado direito e as cores vermelha e azul (magenta) da imagem correspondente à visão do lado esquerdo

As duas principais vantagens deste método são: (1) o baixo custo para a produção e reprodução desse tipo de vídeo ou imagem, que não requer equipamentos com alta tecnologia e pode ser reproduzido em sistemas de vídeo já existentes; e (2) o volume de dados utilizados para armazenamento, muito menor do que em relação aos outros métodos a serem vistos, já que neste caso há uma redução de metade da quantidade de dados original, o que também dá ao método características de compressão.

Já a principal desvantagem é consequência direta da eliminação de metade das informações do par estéreo original. Pelo fato de se retirar dados de cor do par estéreo e utilizar o filtro das lentes dos óculos como forma de compensação, as cores resultantes da combinação dos dois não são as originais, o que traz diferenças na visualização. Além disso, a remoção de informação impossibilita a reversão do anáglifo para o par estéreo original. A recuperação dessas informações é o objetivo deste trabalho.

Vale ressaltar que, mesmo com a deficiência na reprodução da cor original, o método anaglífico mantém as informações espaciais, isto é, de posição de cada elemento nas duas imagens do par estéreo, conservando os dados de paralaxe e garantindo o funcionamento da disparidade binocular e da estereopsia.

2.3.2. Multiplexação por polarização

Para este método e os métodos seguintes, tem-se como requisito o par estéreo como um todo, e não uma imagem misturada como ocorre no anaglífico. No método de multiplexação por polarização, cada vídeo ou imagem do par é reproduzido por projetores diferentes em uma tela metalizada (UREY *et al.*, 2011). Cada projetor possui um filtro polarizador, que faz com que as ondas de luz sejam transmitidas à tela em padrões diferentes (**Figura 5**[6]). Com o auxílio de óculos possuindo esses mesmos filtros, é possível que cada olho veja apenas a projeção destinada a ele, com a lente bloqueando a outra projeção que segue um padrão de onda de luz diferente.

[6] Figura retirada de http://en.wikipedia.org/wiki/File:Wire-grid-polarizer.svg. Acesso em 19 nov. 2013.

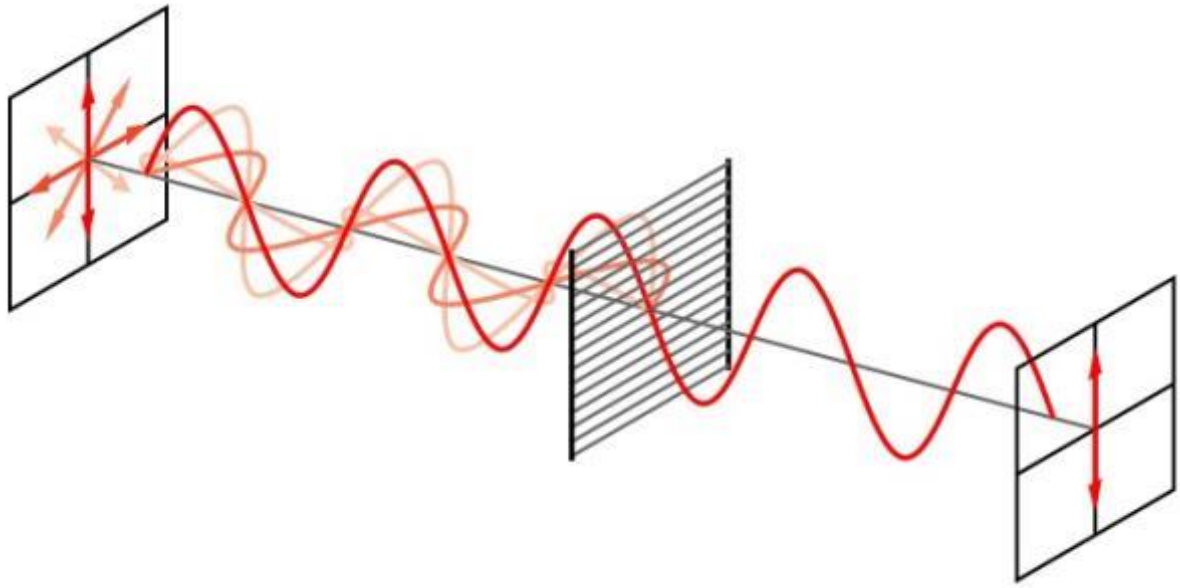


Figura 5 - Exemplo do efeito de polarização. O filtro situado no meio faz com que a onda de luz siga um padrão linear vertical. Um outro filtro polarizador em outra fonte de projeção pode ser utilizado para fazer com que a onda siga um padrão linear horizontal. Isso possibilita que óculos com lentes polarizadas façam a separação de cada vídeo

Como o par estéreo é reproduzido separadamente e de forma íntegra, não há aqui a desvantagem de se perder a cor real da cena. Devido à polarização, as imagens podem ficar mais escuras, porém, isso pode ser corrigido aumentando-se o brilho. Os dispositivos que utilizam a multiplexação por polarização são os que vêm sendo comumente utilizados pela indústria cinematográfica e essa é a tecnologia por trás da maioria dos cinemas 3D atuais. Entretanto, este método introduz complexidades na reprodução: são necessários dois projetores e telas adequadas ao método, sendo que ambos os projetores devem ser calibrados para ficarem em perfeita sincronia, de modo que os vídeos sejam reproduzidos na mesma linha de tempo. Isso tem resultado direto no custo da tecnologia.

2.3.3. Multiplexação temporal

Diferente dos óculos utilizados na multiplexação por cor e polarizada, cujas lentes utilizam filtros para separar as imagens corretas para cada olho, na multiplexação temporal a separação é feita mecanicamente. Por essa razão, a multiplexação temporal é dita utilizar tecnologia ativa, enquanto a polarizada é dita utilizar tecnologia passiva.

Muitos televisores e monitores fazem uso da multiplexação temporal para visualização estereoscópica, que funciona da seguinte forma: o monitor exibe cada lado de imagem do par

estéreo de modo alternado, em uma frequência mais alta do que a normal. O par de óculos ativo, composto por lentes de LCD, também alterna entre si o nível de opacidade de cada lente, na mesma frequência do monitor. Com isso, por uma fração mínima de tempo, uma lente se encontrará opaca e a outra não, e consequentemente, um olho vai enxergar a imagem e o outro não. Como a essa troca ocorre milhares de vezes a cada segundo, nossos olhos não notam a mudança (UREY *et al.*, 2011).

Os principais problemas desta técnica são o alto custo para a produção de cada par de óculos (que possui partes mecânicas e utiliza baterias para funcionamento), dificultando seu uso em cinemas, por exemplo, e a necessidade de maior complexidade envolvendo a sincronia da reprodução com os mecanismos das lentes, para que não ocorra um atraso na visualização que possa causar desconforto e perda da sensação de profundidade.

Vale ressaltar um pequeno embate envolvendo a utilização de tecnologia passiva e tecnologia ativa por alguns fabricantes de televisores[7]. Defensores da tecnologia ativa mostram que a desvantagem da passiva é a redução da resolução do vídeo pela metade. Isso ocorre porque na tecnologia passiva as imagens do par estéreo são reproduzidas simultaneamente em um único quadro (por meio da inserção da informação da imagem direita nas linhas de pixels pares e informação da imagem esquerda nas ímpares, por exemplo), o que reduz a resolução de cada imagem pela metade. Já na tecnologia ativa são transmitidos quadros completos de cada vídeo por meio da alternância entre quadros em uma alta frequência. Entretanto, esse é um ponto de debate, já que na visualização passiva as imagens são interpretadas e combinadas na visão em apenas uma, o que, em teoria, faz com que a resolução vista na tela a resolução total.

2.4. Visualização utilizando HMD

Diferente dos métodos por multiplexação, em que a reprodução é feita em uma tela ou monitor e visualizado com um par de óculos especiais, os dispositivos HMD realizam tanto a reprodução quanto a visualização. Um dispositivo do tipo HMD é um capacete composto por duas telas distintas, cada uma sendo visível a apenas um olho, habilitando a visualização este-

[7] http://planetech.uol.com.br/2011/08/05/debate-3d-acompanhe-o-audio-na-integra/. Acesso em 19 nov. 2013.

reoscópica, com a vantagem de ser portátil e móvel. Além disso, tal dispositivo se caracteriza como sendo um sistema de realidade aumentada, podendo capturar imagens do ambiente e acrescentar informações gráficas adicionais às telas, permitindo maior imersão e disposição de informações.

Os dispositivos HMD têm sido utilizados nas áreas médica, militar e industrial (UREY *et al.*, 2011), possibilitando a criação de um ambiente virtual para treinamentos e simulações, bem como buscando espaço no campo de entretenimento, com o desenvolvimento do Oculus Rift[8] (**Figura 6**)[9], por exemplo.

Figura 6 - Oculus Rift, dispositivo voltado para jogos que promete imersão total em um ambiente de realidade virtual

Os problemas de sincronia e alinhamento entre as telas também continuam nessa classe de visualização. Além disso, ainda existem as dificuldades com o custo elevado e a necessidade de soluções para ampliar o campo de visão e gerar imagens em alta resolução. Vale lembrar que o fato da reprodução ser feita diretamente do dispositivo torna a tecnologia mais individualizada, forçando a necessidade de um equipamento para cada espectador.

[8] http://www.oculusvr.com. Acesso em 19 nov. 2013.
[9] Figura retirada de http://www.theverge.com/2013/3/19/4122388/can-oculus-rift-save-virtual-reality. Acesso em 19 nov. 2013.

2.5. Visualização utilizando métodos autoestereoscópicos

A obrigatoriedade de se utilizar algum aparato especial, como foi visto anteriormente nos métodos por multiplexação e nos dispositivos HMD, se mostra uma abordagem que pode ser incômoda e que quebra o paradigma de como os usuários estão acostumados a assistir vídeos. Visando o descarte desses óculos ou de qualquer outro dispositivo na visualização de vídeos 3D, surgiu a tecnologia envolvida na criação de monitores autoestereoscópicos, os quais são capazes de fornecer diferentes pontos de vista, chamados de visões, para cada olho. Com isso, os métodos autoestereoscópicos permitem a utilização de outro formato de visualização estereoscópica chamado de multivisão (*multiview*).

Para criar uma tela autoestereoscópica é necessário que a própria tela de alguma maneira faça em um só plano a correta separação e envio de visões distintas para cada olho. Isso é feito por meio do uso de componentes ópticos que auxiliam no desvio da luz e criam "caminhos" para cada visão.

Existem diferentes técnicas para visualização autoestereoscópica, cada qual podendo utilizar um sistema diferente de componentes ópticos. A explicação de cada técnica está fora do escopo deste trabalho, e pode ser encontrada em (DODGSON, 2005; HALLE, 1997; UREY *et al.*, 2011). Como exemplo, dois dos sistemas utilizados são:

a) lenticular (**Figura 7 – a**), no qual se coloca na frente da tela um conjunto de lentes cilíndricas, também chamadas de lentículas. Devido ao seu formato, a luz é direcionada em vários ângulos diferentes por refração, o que faz com que cada pixel só possa ser visto de certo ângulo de visão;
b) com barreira de paralaxe (**Figura 7 – b**), em que uma barreira composta de diversas aberturas verticais separadas por uma máscara negra é colocada na frente da tela, o que faz com que a luz passe apenas por essas aberturas. Pedaços de cada imagem do par estéreo são intercaladas, permitindo reproduzir cada uma em direções diferentes.

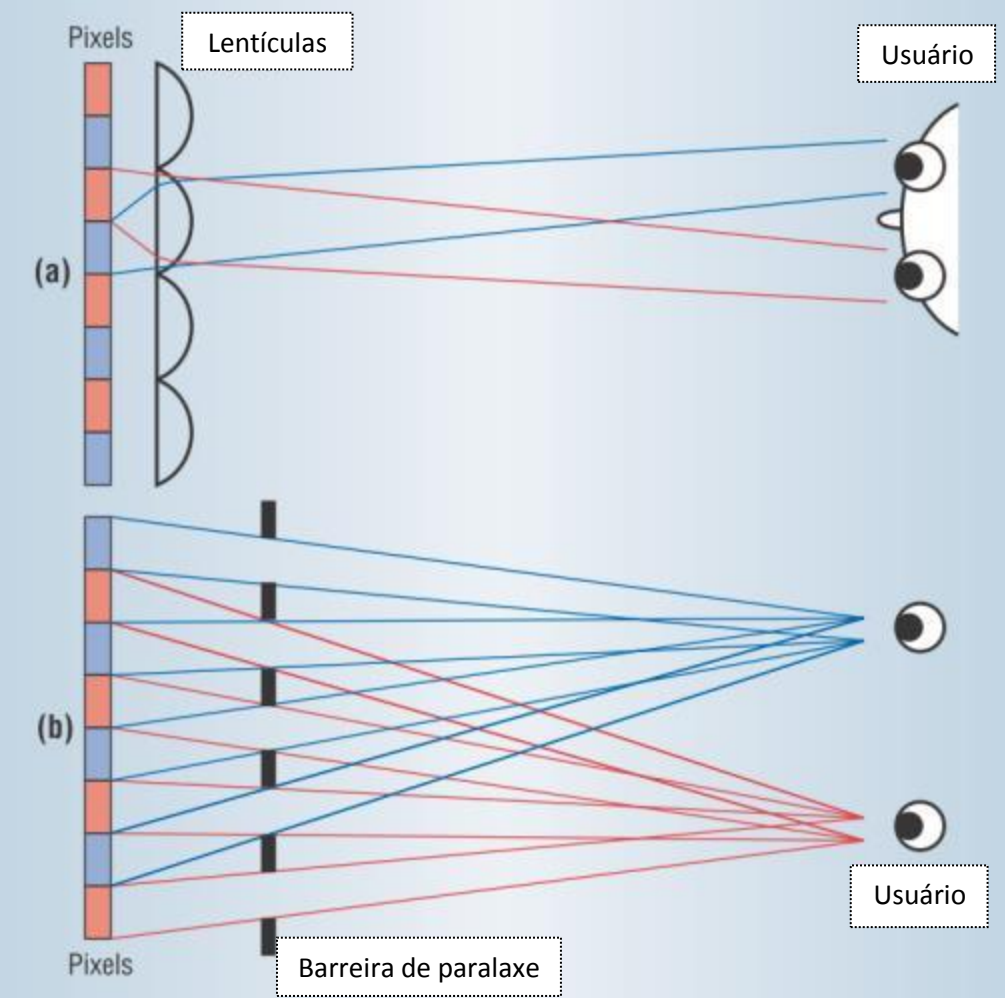


Figura 7 - Exemplos de sistemas utilizados nos métodos autoestereoscópicos. Em (a), as lentículas são responsáveis por desviar a luz para cada olho. Em (b), pequenas barreiras impedem a passagem da luz em certas direções

Os métodos autoestereoscópicos são os mais recentes, sendo a aposta dos fabricantes para o futuro dos televisores 3D. Por esse motivo, ainda há muita pesquisa e desenvolvimento e uma série de deficiências a serem superadas. Uma delas é que o espectador deve se situar em pontos chaves no ambiente, devido ao alcance limitado de cada campo de visão. Fora desses pontos o *crosstalk* se torna muito evidente.

Ainda são necessários mais alguns anos até que televisores autoestereoscópicos sejam produzidos em massa, o que torna o custo de produção elevado. Mesmo assim, algumas soluções para dispositivos móveis, com telas pequenas, já se encontram disponíveis no mercado, como é o caso do smartphone LG Optimus 3D (LG ELECTRONICS, 2011) e do vídeo game Nintendo 3DS (NINTENDO OF AMERICA INC, 2011).

2.6. Considerações finais

Este Capítulo apresentou as principais informações de profundidade envolvidas na visualização estereoscópica. Foram também apresentados diferentes métodos e dispositivos para visualização estereoscópica que, independentes da forma como funcionam, têm em si a mesma premissa de separar o par estéreo e entregar o vídeo correto para cada olho. Um resumo de cada método, relacionando custo de aquisição, qualidade visual e compressão, pode ser visto na **Tabela 1**.

Tabela 1 - Métodos de visualização estereoscópica relacionados a partir de seu custo de aquisição, qualidade visual e compressão

Tipo de Visualização	Custo	Qualidade da visualização	Taxa de compressão
Multiplexação de cor (método anaglífico)	Baixo	Razoável	Alta
Multiplexação por polarização	Alto	Boa	Baixa
Multiplexação temporal	Alto	Excelente	Baixa
Dispositivos HMD	Altíssimo	Excelente	Baixa
Autoestereoscopia	Altíssimo	Boa	Baixa

O método anaglífico é o utilizado pela técnica RevGlyph, a ser descrita neste trabalho, por permitir alta taxa de compressão ao transformar par estéreo em apenas um fluxo de vídeo por meio da combinação dos canais de cor primária de cada vídeo. Dependendo da combinação utilizada, o anáglifo pode ser classificado em verde-magenta, vermelho-ciano e azul-amarelo. Análises de qualidade subjetivas conduzidas por Andrade e Goularte (2010) com grupos de usuários mostram que o anáglifo verde-magenta proporciona maior qualidade em relação à percepção de profundidade, garantindo 10% a mais de qualidade quando comparado ao vermelho-ciano, e foi por isso o escolhido para uso na técnica RevGlyph.

Os diferentes métodos de visualização estereoscópica apresentados podem utilizar o par estéreo em formatos de organização diferentes. Esses formatos podem ser incompatíveis quando utilizados em outro método. Por exemplo, um vídeo codificado pela técnica anaglífica possui o par estéreo fundido em um único fluxo de dados, o que impossibilita a exibição desse conteúdo em sistemas que utilizam óculos obturadores ou polarizadores, que necessitam do par estéreo completo. Essa incompatibilidade gera desafios para se criar uma técnica de codificação independente do método de visualização utilizado.

3. Codificação e compressão estereoscópica

Compressão é uma necessidade constante para o armazenamento e transmissão de dados de forma eficiente. Pelo fato de um par estereoscópico carregar a informação referente a duas imagens ou a dois sinais de vídeo, a necessidade de compressão é duas vezes maior e requer a busca por maneiras mais eficientes de se remover dados sem afetar a qualidade final da percepção de profundidade.

Vídeos possuem redundâncias espaciais e temporais, que são exploradas com a fim de se obter compressão (RICHARDSON, 2003). Os métodos que exploram essas redundâncias (chamados aqui de métodos 2D) vêm sendo aplicados também a vídeos estereoscópicos, o que em alguns casos pode comprometer a percepção de profundidade (ANDRADE; GOULARTE, 2010). Existem outras abordagens que exploram particularidades de um par estéreo, o qual possui redundâncias adicionais a serem exploradas. Por exemplo, existe a redundância entre quadros homólogos de cada vídeo do par estéreo, os quais são guardam muitas informações semelhantes, apenas deslocadas horizontalmente devido à posição da câmera de captura. Tal semelhança pode ser explorada para se obter mais compressão.

Este Capítulo apresenta os conceitos básicos de conversão entre espaço de cores e subamostragem de crominância (**Seção 3.1**), amplamente utilizados em métodos de compressão tanto de imagens quanto de vídeos, e que são explorados de modo particular na técnica RevGlyph. A **Seção 3.2** discute como os dados estereoscópicos são organizados, de acordo com a literatura atual. Tal discussão é importante para que se possa posicionar e comparar a técnica proposta. A **Seção 3.3** apresenta uma discussão sobre os métodos de compressão tradicionais empregados para comprimir vídeo estéreo. Ainda, apresenta um levantamento das principais técnicas atualmente desenvolvidas para codificar vídeo estéreo. Por fim, a **Seção 3.4** faz uma análise crítica dos métodos apresentados, apontando suas limitações e também apresenta trabalhos recentes relacionados a cada um.

3.1. Espaço de cores e subamostragem de crominância

A representação de imagens digitais se dá por meio de tons monocromáticos ou coloridos. Computacionalmente, os tons monocromáticos podem ser representados por um byte, produzindo assim 256 níveis em escala de cinza que representam informações de luminância, ou seja, intensidades da luz em cada ponto da imagem. Quando além de luminância se deseja também incluir informações sobre as cores, é necessário utilizar o chamado espaço de cores, no qual cada cor é representada por uma tripla de valores (x, y, z) (SALOMON, 2008), de acordo com a teoria tricromática (AZEVEDO; CONCI, 2003).

Existem vários modelos de espaço de cores, cada qual apropriado para um tipo de aplicação ou sistema de visualização. Os modelos discutidos nesta Seção são o RGB e o YC_bC_r, por serem amplamente utilizados em codificação de imagens e vídeos, e também utilizados na técnica desenvolvida neste trabalho. Mais detalhes sobre outros espaços de cores podem ser encontrados nos textos de Azevedo e Conci (2003) e Feitosa-Santana *et al.* (2006).

O modelo RGB é baseado na tripla de cores primárias: vermelho, verde e azul. Elas são classificadas como cores aditivas, isto é, por meio da mistura das três são produzidas as outras cores, sendo que o branco é obtido quando as cores primárias são misturadas em sua intensidade máxima. Este modelo é o mais popular e o comumente utilizado por dispositivos de captura, como câmeras fotográficas, e de apresentação, como as telas de LCD (RICHARDSON, 2003).

Já o modelo YC_bC_r traz outra representação de espaço de cores, cujas componentes permitem a separação das informações relativas à luminância (Y) das informações relativas à crominância (C_b e C_r). Os valores de Y, C_b e C_r podem ser extraídos do modelo RGB por um processo de conversão do espaço de cores utilizando a **Equação 1**[10], a qual é uma recomendação da ITU-R (ITU-R, 2011). O inverso é também possível com a **Equação 2**[10].

Independente do espaço de cor utilizado, cada componente é representada na forma de uma matriz cujas dimensões são a mesma da imagem que representam. Cada elemento nas matrizes vai representar um pixel da imagem. No caso do espaço YC_bC_r, o elemento vai possuir informações sobre a luminância e crominância do pixel, e no caso do espaço RGB, informações sobre cada cor primária.

[10] Ambas as fórmulas foram adaptadas de http://compression.ru/video/quality_measure/info_en.html. Acesso em 19 nov. 2013.

$$\begin{aligned} Y &= (0.257 * R) + (0.504 * G) + (0.098 * B) + 16 \\ Cb &= -(0.148 * R) - (0.291 * G) + (0.439 * B) + 128 \\ Cr &= (0.439 * R) - (0.368 * G) - (0.071 * B) + 128 \end{aligned} \quad (1)$$

$$\begin{aligned} R &= 1.164 * (Y - 16) + 1.596 * (Cr - 128) \\ G &= 1.164 * (Y - 16) - 0.391 * (Cb - 128) - 0.813 * (Cr - 128) \\ B &= 1.164 * (Y - 16) + 2.018 * (Cb - 128) \end{aligned} \quad (2)$$

Em que:

Y: matriz resultado com as informações relativas à luminância;

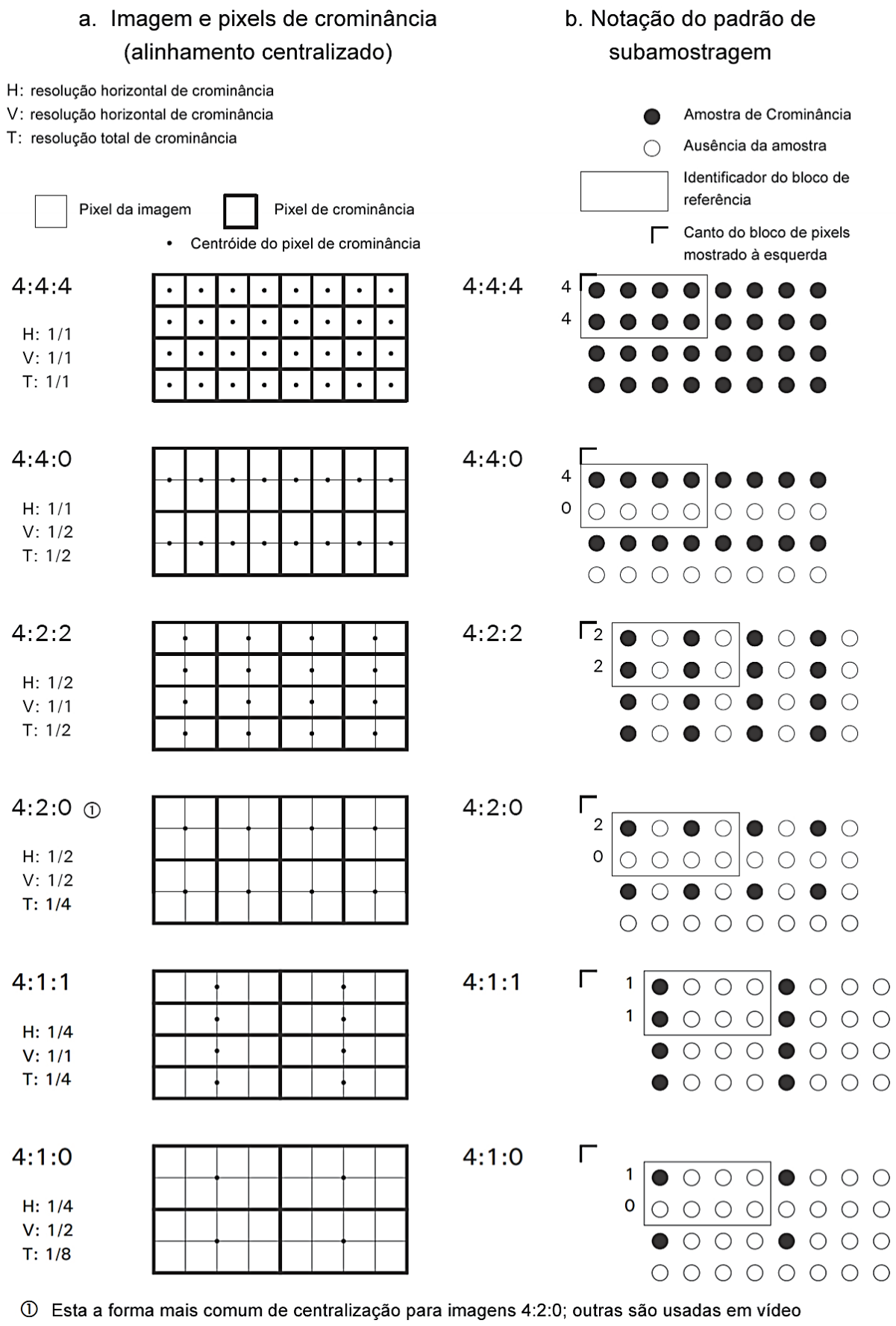
R, G e *B*: matrizes contendo os valores de cor dos pixels de cada componente do espaço de cores RGB;

Cr e *Cb*: matrizes contendo os valores de crominância dos pixels no espaço de cores YC_bC_r.

No modelo RGB a informação de luminância está atrelada ao valor de cada componente de cor do modelo. Isso se torna um inconveniente, pois não torna possível explorar uma propriedade do sistema visual humano: a maior sensibilidade à luminância do que às cores (SALOMON, 2008). Em razão dessa propriedade, a informação relativa às cores (crominância) pode ser representada em uma resolução menor do que a informação relativa à luminância, sem grandes prejuízos de percepção visual na imagem. Essa característica é explorada durante a codificação de imagens e vídeos em uma etapa chamada de subamostragem de crominância (KERR, 2012).

O espaço de cores YC_bC_r separa dados de luminância e cor. Desse modo, pode-se realizar a subamostragem de crominância dentro deste espaço. Como o olho humano é mais sensível às variações de luminância do que de cores, os dados referentes à cor são amostrados a uma taxa menor do que os dados referentes à luminância. Dependendo da subamostragem aplicada à crominância é possível obter grande redução do volume de dados, tendo em contrapartida uma qualidade inferior na imagem subamostrada.

De acordo com a quantidade e a forma em que as componentes de crominância são amostradas, pode-se classificar a subamostragem em diferentes modelos. Alguns modelos são ilustrados na **Figura 8** (KERR, 2012) e descritos a seguir.



① Esta a forma mais comum de centralização para imagens 4:2:0; outras são usadas em vídeo

Figura 8 - Tipos de subamostragem de crominância

No modelo 4:4:4 não há redução da resolução das cores, isto é, para cada amostra de Y, há uma amostra de C_b e uma de C_r. Este modelo mantém a fidelidade das cores da imagem, porém, não contribui na compressão e pode até mesmo não ser classificado como subamostragem. Nos modelos 4:4:0 e 4:2:2, para cada quatro amostras de Y, há duas amostras de C_b e duas de C_r, ou seja, a resolução das cores é reduzida pela metade, sendo a diferença entre esse dois modelos a orientação da subamostragem (vertical no modelo 4:4:0 e horizontal no modelo 4:2:2). O mesmo se aplica para os modelos 4:2:0 e 4:1:1, em que a redução é feita tanto horizontal quanto verticalmente; nestes modelos, a resolução das cores é reduzida a um quarto. Por fim, a subamostragem 4:1:0 realiza a redução mais severa de um oitavo na resolução das cores; isso possibilita maior compressão, porém, acarreta em uma imagem de qualidade muito inferior.

Cabe lembrar que a eliminação de cores feita na subamostragem é irreversível. Na conversão de uma imagem YC_bC_r subamostrada para sua similar em RGB, o retorno aproximado ao modelo 4:4:4 pode ser obtido copiando-se o valor dos pixels vizinhos (ou uma média deles) a cada pixel não amostrado. É necessário retornar ao modelo 4:4:4 para a conversão ao espaço RGB, já que este não possui dados de luminância e crominância separados.

3.2. Organização de dados estereoscópicos

A literatura aponta divergência de formatos para organizar dados de vídeos estereoscópicos (LIU; CHEN, 2010; SMOLIC *et al.*, 2009; VETRO, 2010). Dos formatos existentes, pode-se agrupá-los em três classes: (1) métodos envolvendo o par estéreo, neste trabalho denominado “método de Lipton” por este ser o idealizador dos formatos a serem descritos; (2) métodos multivisão, quando mais de um par estéreo é utilizado; e (3) métodos envolvendo vídeo e profundidade, que utilizam novas estruturas envolvendo dados de profundidade. Cada método é descrito nas Seções a seguir.

3.2.1. Método de Lipton

Em seu trabalho, Lipton (1997) criou dois formatos para vídeos estereoscópicos em uso até hoje, os quais foram desenvolvidos para que pudessem ser utilizados com pouca ou nenhuma alteração na infraestrutura de hardware disponível para visualização. Nestes formatos cada quadro de vídeo contém uma imagem do vídeo esquerdo e a sua correspondente no vídeo direito. Desse modo, as duas imagens que formam um par estéreo são armazenadas como um todo em cada quadro. Esses quadros podem ser então armazenados em um contêiner AVI, por exemplo, bem como serem codificados utilizando técnicas já conhecidas, como o MPEG.

A disposição das imagens do par estéreo em cada quadro gera os formatos acima-abaixo (quando uma imagem é posicionada na parte superior e outra na parte inferior do quadro) e lado-a-lado (quando uma imagem é posicionada à esquerda e outra à direita do quadro), dependendo do sistema em que serão reproduzidos.

O problema desse tipo de organização é o grande volume de dados, já que é necessário armazenar dois sinais de vídeo. Além disso, por colocar o par estéreo em cada quadro, é necessário que as imagens passem por alguma transformação de modo a diminuir sua taxa de aspecto e assim serem acomodadas corretamente no quadro. Já a maior vantagem do método de Lipton é poder ser utilizado independente do tipo de visualização a ser aplicado, pois mantém todas as informações necessárias para a reprodução do vídeo, o que torna essa organização genérica.

3.2.2. Método multivisão

Atualmente, novas tecnologias de telas e monitores são capazes de gerar mais de uma visão ao espectador, dependendo da posição em que ele se encontra em relação à tela. Isso possibilita uma liberdade ao espectador, que pode assistir uma cena com diferentes pontos de vista, o que é denominado FVV. Essa disponibilidade de mais de uma visão da cena é chamada de multivisão e é uma tendência de uso em telas autoestereoscópicas.

Cada visão é resultado de um novo par estéreo, que são capturados em diferentes posições da cena e podem ser armazenados tanto no formato lado-a-lado quanto acima-abaixo. Vídeos organizados por método de multivisão podem ser codificados pela técnica MVC (*Multiview Video Coding*).

A multivisão resulta num aumento crescente no volume de dados, já que cada nova visão implica em agregar um novo par estéreo. Com isso, a compressão eficiente dos dados é essencial. Uma das alternativas envolve explorar redundâncias existentes entre as imagens do par estéreo e é descrita na **Seção 3.3.2**. Alguns padrões disponíveis implementam o MVC, como os do grupo MPEG no MPEG-2 *Multiview Profile* e também o padrão H.264/AVC (SMOLIC *et al.*, 2009; VETRO; WIEGAND; SULLIVAN, 2011).

Mesmo com padrões existentes para codificação multivisão, o processamento do alto volume de dados de vídeo é oneroso e não permite a transmissão eficiente de um grande número de visões utilizando o MVC. Na prática, a quantidade de visões codificadas de maneira eficiente fica limitada a duas ou três visões (MÜLLER; MERKLE; WIEGAND, 2011).

3.2.3. Métodos baseados em vídeo e profundidade

Como visto na Seção anterior, métodos multivisão possuem o problema de se trabalhar com um grande volume de dados. Cada nova visão adicionada resulta em mais dados a serem processados, diminuindo a eficiência da codificação MVC. Métodos baseado em vídeo e profundidade, por sua vez, utilizam somente um dos sinais de vídeo de um par estéreo, juntamente com informações geométricas da cena que possibilitam a síntese de novas visões, chamadas de visões virtuais. Isso propicia uma codificação estereoscópica mais flexível e com um menor volume de dados.

As informações geométricas da cena são inseridas em um sinal de vídeo auxiliar, o qual é formado por tons de cinza e está relacionado ao vídeo capturado. Neste sinal de vídeo auxiliar o valor de cada pixel representa a distância do ponto representado em relação à câmera de captura, agregando dessa forma informações de profundidade da cena. Por esse motivo, o vídeo auxiliar é denominado mapa de profundidades. A **Figura 9** exemplifica essa organização apresentando um quadro de vídeoeo RGB (representando um dos sinais de vídeo de um par estéreo) e seu respectivo mapa de profundidades.

Um dos primeiros trabalhos com esse tipo de organização foi realizado pela Sociedade de Tecnologia da Informação Europeia com o projeto ATTEST – *Advanced Three-Dimensional Television System Technologies* (FEHN *et al.*, 2002). No projeto, dentre diferentes funcionalidades está a utilização de uma câmera com um sistema de infravermelho acoplado. A câmera captura um sinal de vídeo RGB e, ao mesmo tempo, o sistema de infraver-

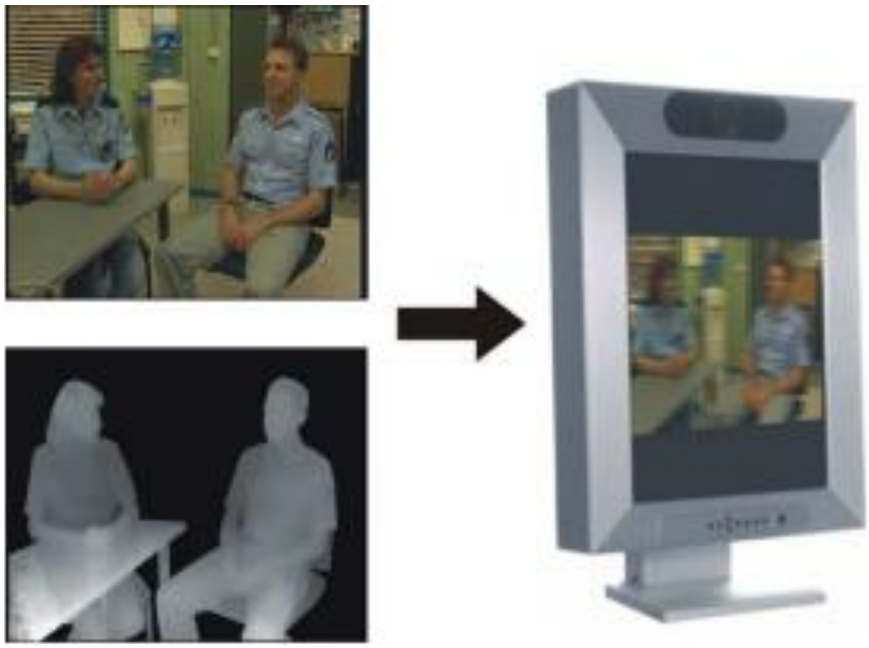


Figura 9 - Processo de codificação utilizando vídeo e mapa de profundidades para a formação de um vídeo estereoscópico (SMOLIC et al., 2009)

melho faz uma varredura do ambiente para gerar outro sinal de vídeo que contém dados de distância dos objetos em relação à câmera.

De posse de um vídeo e seu respectivo mapa de profundidades é possível sintetizar visões virtuais por meio de cálculos baseados nos princípios da geometria projetiva. Esse processo de síntese é feito com uma técnica denominada DIBR – *Depth-Image-Based Rendering* (MÜLLER; MERKLE; WIEGAND, 2011). A DIBR pode ser dividida em duas etapas: na primeira utilizam-se os dados do mapa de profundidades para se projetar cada ponto da imagem em um "mundo 3D". Na segunda etapa, esses pontos são projetados em um plano de uma câmera virtual fictícia, localizada na posição em que a visão virtual supostamente deve estar (FEHN; PASTOOR, 2006).

Nas pesquisas realizadas por SMOLIC *et al.* (2009), métodos baseados em vídeo e profundidade podem ser classificados em três tipos: V+D (*Video plus Depth*), que foi descrito acima; MVD (*MultiView plus Depth*), utilizado quando se necessita da criação e de número maior de visões, similar ao MVC; e LDV (*Layered Depth Video*), uma codificação mais complexa, que adiciona camadas (*layers*) adicionais de vídeo contendo informações auxiliares retiradas de outras visões, utilizadas para a síntese de várias visões com maior qualidade.

SMOLIC *et al.* (2009) também apontam problemas de incompatibilidade entre os diferentes tipos de organização de dados estereoscópicos, tanto os baseados em vídeo e profundidade, quanto os baseados em Lipton e multivisão. Dessa forma, os autores propõem outra

organização chamada de DES (*Depth Enhanced Stereo*), a qual seria um apanhado de todas as outras: o par estéreo completo seria armazenado junto com respectivos mapas de profundidade e camadas auxiliares, de modo a atender a todo tipo de visualização. Entretanto, fica claro um aumento significativo da quantidade de dados a ser armazenada e transmitida.

3.3. Compressão

Nesta Seção serão discutidas as técnicas de compressão espacial e temporal 2D, que podem ser utilizadas tanto para imagens e vídeos 2D quanto para imagens e vídeos estereoscópicos. Além disso, são também apresentadas as técnicas desenvolvidas para explorar redundâncias inerentes aos vídeos estereoscópicos, de acordo com os formatos descritos na **Seção 3.2**.

3.3.1. Compressão 2D

Um fluxo de vídeo é na verdade uma sequência de imagens (chamadas de quadros) que, quando transmitidas a certa frequência, passam uma sensação de movimento. Por se tratar de imagens, o primeiro passo é aplicar em cada quadro métodos de compressão associado a imagens para eliminar as informações de redundância espacial que estas apresentam. Isso pode envolver tanto métodos de compressão sem perdas quanto com perdas, o que influencia na qualidade da imagem resultante (HALSALL, 2000).

O processo de compressão espacial envolve aplicar redução do espaço de cor por meio da subamostragem de crominância, como visto na **Seção 3.1**. Além disso, pode ser aplicada uma transformada, isto é, uma função matemática que vai decompor a informação em componentes de frequência a partir da aplicação de diferentes filtros. Posteriormente, as componentes de frequências mais altas que certo limiar são removidas por meio de um processo denominado quantização. Essa remoção se baseia em características perceptivas da nossa visão: dependendo do limiar estabelecido, o olho humano não percebe diferenças significativas na imagem resultante. Exemplos de transformadas comumente utilizadas são a DCT (*Discrete Cosine Transform*) e DWT (*Discrete Wavelet Transform*) (GONZALEZ; WOODS, 2008).

Com a redução do espaço de cor e uso de transformadas com consequente quantização é possível eliminar redundâncias espaciais e psicovisuais de cada imagem. Além disso, ainda pode ser eliminada a redundância estatística por meio da chamada codificação por entropia, que é uma compressão sem perda, a qual atribui um número de bits para cada valor/padrão conforme a frequência em que aparecem (menos bits para os que aparecem com maior frequência, mais bits para os menos frequentes). As codificações mais conhecidas são Huffman, LZW e por carreira (*run-length*) (HALSALL, 2000).

Os passos descritos acima promovem a compressão de cada quadro de vídeo de maneira isolada, o que não é suficiente em relação ao vídeo como um todo. Como os vídeos são sequências de quadros se alternando durante o tempo, há outro tipo de redundância a ser explorada: a redundância temporal. Esta é caracterizada pela similaridade que há entre quadros vizinhos de uma sequência, resultando em dados que podem ser eliminados. Como os quadros são similares, a compressão temporal se encarrega por codificar apenas alguns e predizer como serão os próximos, armazenando somente as diferenças entre eles.

Descrevemos aqui o processo de compressão temporal baseado no padrão MPEG-1 (CHAPMAN; CHAPMAN, 2004; SAYOOD, 2005). Os quadros são classificados em I, P ou B, de acordo com a maneira em que são codificados. Quadros I (*Intracoded frames*) passam apenas pela compressão espacial utilizando os algoritmos de compressão de imagens. Quadros P (*Predictive frames*) são codificados em relação a um quadro I ou P anterior a ele, obtendo-se uma estimativa do que mudou entre ele e seu antecessor, o que é chamado de estimativa de movimento. Isso significa que um quadro P é excluído e armazena-se apenas com os dados da estimativa de movimento para posterior reconstrução deste durante a decodificação.

O processo de reconstrução de quadros P envolve erros advindos da estimativa de movimento. Por isso, é também codificada uma tabela de compensação de movimento, contendo a diferença entre a posição estimada e a posição real dos objetos. Já que um quadro P pode ser codificado a partir de um quadro P anterior, há propagação de erros. Por essa razão deve-se estabelecer um limite de criação de quadros P consecutivos, chamado de *Prediction Span*.

Por fim, os quadros B (Bidirectional frames) são codificados tanto em relação ao um quadro P ou I anterior a eles quanto em relação a um quadro P ou I posterior, obtendo-se uma taxa maior de compressão, porém impactando o tempo de processamento, já que precisamos esperar os quadros P ou I posteriores serem processados para o cálculo.

Vale ressaltar que tanto a compressão espacial quanto a temporal resultam na perda de dados, sendo que grande parte dessa perda é irreversível e pode afetar a qualidade visual do

vídeo codificado. Um balanço entre qualidade e compressão é sempre necessário ser analisado. O processo de compressão espacial e temporal aqui descrito serve como base para muitos padrões de compressão de vídeos bem conhecidos, como MPEG (1, 2 e 4) e H.264, por exemplo.

3.3.2. Compressão Estereoscópica

Como é possível organizar os dados estereoscópicos de diferentes maneiras (Lipton, multivisão, vídeo e profundidade), a compressão estereoscópica pode seguir diferentes processos. Nesta Seção, são descritas as abordagens existentes para cada tipo de organização.

Vídeos estereoscópicos organizados de modo a armazenar o par estéreo possuem um tipo de redundância adicional em relação a vídeos 2D, chamada de redundância entrevisões (*inter-view*), similar à redundância temporal. As imagens que formam cada quadro de um par estéreo são muito semelhantes, diferenciando-se pelo deslocamento horizontal que ocorre pela distância em que cada uma foi capturada. Dessa forma, não é necessário armazenar as duas imagens: pode-se predizer uma delas baseada na outra.

A **Figura 10** ilustra como se pode explorar a redundância entrevisões, em um processo denominado de predição entrevisões (VETRO; WIEGAND; SULLIVAN, 2011). Um dos vídeos do par estéreo, chamado de visão base, é codificado em quadros I, P e B, conforme descrito na **Seção 3.3.1**. Já o outro vídeo do par, chamado de visão virtual, é formado apenas por quadros P e B, criados a partir de seu quadro correlato na visão base (para quadros P) ou de um quadro correlato na visão base juntamente com um anterior na visão virtual (para quadros B).

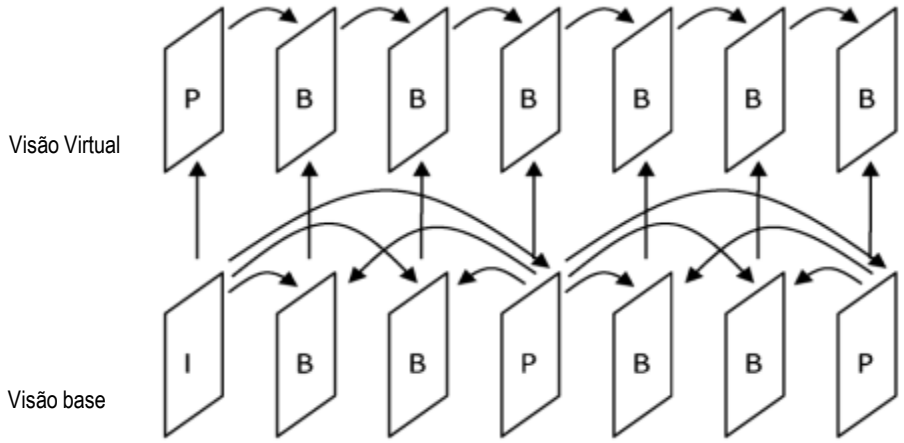


Figura 10 - Predição entrevisões. Uma das visões se torna a base e é utilizada para predizer a outra visão do par estéreo (VETRO; WIEGAND; SULLIVAN, 2011)

A predição entrevisões não se limita a apenas um par estéreo e pode ser escalável para multivisão. Neste caso, uma das visões é a base e o restante é virtual, resultante da mescla entre predição temporal e entrevisões. A visão base, por se utilizar apenas de compressão 2D, possibilita a compatibilidade reversa para dispositivos que não reproduzem vídeos estereoscópicos ou multivisão. O padrão H.264/MPEG-4 implementa a extensão AVC para o MVC, incluindo a predição entrevisões (MÜLLER; MERKLE; WIEGAND, 2011).

Com relação aos vídeos estereoscópicos no formato de vídeo e mapa de profundidades, a própria organização dos dados é uma forma de compressão estereoscópica. O mapa de profundidades é um sinal de vídeo em escala de cinza, ou seja, necessita apenas de uma componente de cor para armazenar os dados, o que causa uma redução de dois terços em comparação a um vídeo RGB.

O padrão MPEG possui uma especificação para codificação utilizando mapa de profundidades com o MPEG-C Parte 3, o qual possui parâmetros adicionais para a interpretação dos valores de profundidade durante a decodificação (ISO/IEC JTC1/SC29/WG11, 2007). Outras abordagens são propostas por alguns pesquisadores, envolvendo codificação utilizando *wavelets* (DARIBO; TILLIER; PESQUET-POPESCU, 2008; MAITRE; DO, 2010) e decomposição de sinal utilizando funções matemáticas denominadas *Wedgelet* e *Platelet* (SMOLIC *et al.*, 2009).

As técnicas e métodos para compressão/codificação aqui apresentados demonstram que cada organização explora aspectos diferentes dos vídeos estereoscópicos para obter compressão, o que torna cada solução específica para certo uso, podendo resultar em incompatibilidades entre formatos dependendo do tipo de visualização aplicado. Mesmo assim, independente da organização dos dados, a mídia a ser processada continua sendo vídeo e por isso pode ser submetida ao processo de compressão 2D descrito na **Seção 3.3.1**.

3.4. Trabalhos relacionados

Como apresentado na **Seção 3.2**, há a falta de um padrão de codificação específico para imagens e vídeos estereoscópicos. Isso traz como consequência pesquisas divergentes, feitas para atender às necessidades de cada formato e método de compressão utilizado. Nesta Seção

citamos trabalhos relacionados na área de codificação e compressão estereoscópica, tanto para multivisão quanto para vídeo e profundidade.

Podem-se observar adaptações na codificação estereoscópica utilizando MVC baseada em melhorias pontuais do padrão H.264/AVC, como é feito no trabalho de LI *et al.* (2009), por exemplo, em que os autores buscam melhorar a eficiência da codificação MVC propondo uma nova estrutura de criação de visões adaptável.

A compressão estereoscópica utilizando predição entrevisões, embora se mostre eficiente e até mesmo escalável para MVC, possui o problema de aumentar o volume de dados a cada adição de uma nova visão. Fora isso, cada visão exige uma nova captura da cena feita por uma câmera em uma posição diferente, o que torna o processo de criação muito custoso se for desejável disponibilizar vários pontos de vista ao espectador, especialmente para telas autoestereoscópicas.

Stefanoski *et al.* (2013) propõe a síntese de novas visões baseada apenas em um par estéreo, com o auxílio de um método chamado *Image-Domain Warping* (IDC). Este método busca explorar a disparidade entre os vídeos do par estéreo, bem como aspectos do sistema visual humano, de forma a sintetizar as visões sem a necessidade de dados complementares, diferente da técnica DIBR descrita na **Seção 3.2.3**. Isso pode trazer a vantagem de produzir conteúdo multivisão sem a sobrecarga de dados adicionais ou a necessidade de captura de vários pontos de vista.

A síntese de visões utilizando IDC se mostra promissora, porém, é feita por meio de cálculos pesados. É necessário mais desenvolvimento e experimentos para que o método ganhe maturidade e robustez no sentido de melhorar o tempo de processamento e diminuir a complexidade computacional envolvida. Isso é necessário para possibilitar o uso da técnica em tempo real, o que ainda não é possível a não ser que se tenha um sistema dedicado.

Na vertente de mapas de profundidade, um dos desafios do método é eliminar a alta interdependência que há entre a criação do mapa, a codificação e a síntese de visões. A qualidade e precisão do mapa, bem como artefatos resultantes da codificação, influenciam na síntese de visões virtuais com relação à qualidade destas. Estes três processos estão muito ligados e sujeitos a propagação de erros, o que prejudica o resultado final (MÜLLER; MERKLE; WIEGAND, 2011).

Atrelado à interdependência dos processos envolvidos nos métodos de vídeo e profundidade, existe também o problema com relação ao conteúdo estereoscópico cuja captura não tenha sido realizada nesse tipo de organização. Mesmo sendo possível computar mapa de pro-

fundidades a partir de um par estéreo, este mapa não tem precisão o suficiente para gerar visões de alta qualidade (STEFANOSKI *et al.*, 2013). Pesquisadores focam tanto na melhoria do processo de aquisição de mapas de profundidade (SCHNYDER *et al.*, 2012), quanto na qualidade das visões sintetizadas (CHEN *et al.*, 2012; NDJIKI-NYA *et al.*, 2010), bem como em tornar o processo como um todo mais eficiente (CHEN *et al.*, 2012).

Outro desafio da área de vídeo e profundidade é com relação a regiões de oclusão na cena. Objetos que se encontram ocultos ou parcialmente ocultos em uma visão irão de alguma forma ficar mais ocultos ou mais descobertos em outras visões. Informação de oclusão não está disponível no vídeo original, o que gera "buracos" nas visões virtuais. Nesses casos, uma solução é usar MVD, LDV ou DES (**Seção 3.2.3**), arcando com o problema de se armazenar mais informação. Outros trabalhos exploram técnicas preditivas caracterizadas como *hole filling*, que realizam cálculos e estimativas para preenchimento desses buracos existentes nas visões virtuais, melhorando sua qualidade. Isto pode ser visto em JIANG *et al.* (2010) e Daribo e Saito (2011).

Independente da codificação utilizada, imagens e vídeos estereoscópicos adicionam uma nova variável a ser considerada quando da avaliação da qualidade: a percepção de profundidade. Estudos realizados por Andrade e Goularte (2009, 2010) demonstraram que os artefatos de compressão resultantes da aplicação de compressão 2D com perdas em vídeos estereoscópicos podem afetar a correta percepção de profundidade destes vídeos. Os autores analisaram as etapas da compressão espacial que afetam a qualidade da estereopsia e também sugeriram uma nova técnica de compressão estereoscópica baseada na codificação anaglífica, projetada com parâmetros que oferecem uma compressão espacial que não afeta percepção de profundidade (ANDRADE, 2012). No entanto, o vídeo codificado é um anáglifo, o que torna a técnica incompatível para ser utilizada com outros tipos de visualização estereoscópica a não ser o método anaglífico.

3.5. Considerações finais

Este Capítulo apresentou um resumo dos aspectos e dos diferentes processos de codificação e compressão que podem ser aplicados a imagens e vídeos estereoscópicos, utilizando-

se tanto dos métodos 2D quanto explorando novas redundâncias encontradas em um par estéreo.

A subamostragem de crominância é um processo de compressão que diminui o volume de dados por meio da redução da crominância, explorando particularidades da visão humana. A redução é variável e pode afetar a qualidade visual do vídeo, dependendo do nível aplicado. Andrade *et al.* (2012), analisaram a influência da subamostragem na percepção de profundidade de imagens estereoscópicas e chegaram à conclusão de que a subamostragem 4:4:0 é a que apresenta a melhor qualidade para a visualização estereoscópica sem afetar a percepção de profundidade. Subamostragens que realizam uma redução mais severa da crominância, como 4:2:0 (utilizada no padrão MPEG-2), por exemplo, se mostram proibitivas para a visualização estereoscópica por prejudicar demais a percepção de profundidade.

Dados de vídeos estereoscópicos podem ser organizados de diferentes formas, utilizando o método de Lipton, multivisão, ou modelos mais complexos envolvendo informação de profundidade. Para compressão, técnicas 2D podem ser utilizadas, porém, não são suficientes, exigindo novas abordagens para compressão mais eficiente, por meio de técnicas como predição entrevisões ou síntese de visões virtuais.

Independente de se utilizar o método de Lipton, multivisão ou vídeo e profundidade, existe o problema de que as soluções apresentadas satisfazem a um formato ou sistema de visualização, não sendo satisfatórias quando aplicadas a outro formato ou sistema. Isso significa que não existe uma técnica de codificação estereoscópica genérica e independente de visualização. Além disso, o uso de compressão 2D gera artefatos que afetam a percepção de profundidade, e é algo que deve ser levado em conta nas técnicas de compressão atuais.

Em suma, temos duas limitações principais na codificação estereoscópica: (1) a falta de um método de codificação que seja genérico e que atenda diferentes formatos e tipos de visualização estereoscópicos; e (2) a aplicação de princípios de qualidade nas técnicas de compressão com relação à percepção de profundidade. Essas duas limitações foram focos de estudo deste trabalho e, em conjunto com alguns dos conceitos abordados neste Capítulo, serviram como base para o desenvolvimento da técnica RevGlyph e são aplicados a ela.

Vale ressaltar novamente que a versão da RevGlyph desenvolvida durante o Mestrado e apresentada no **Capítulo 4** foi aplicada e analisada apenas em imagens estereoscópicas, não sendo explorada as etapas de aplicação de transformadas e quantização, bem como a compressão temporal. Mesmo assim, a RevGlyph pode também ser aplicada a vídeos estereoscópicos por meio da codificação de cada quadro desse vídeo.

4. RevGlyph

Este Capítulo apresenta a técnica RevGlyph, um método reversível de codificação estereoscópica anaglífica. A **Seção 4.1** contextualiza e discute a inserção da RevGlyph na problemática da codificação estereoscópica. A **Seção 4.2** contém o processo de codificação proposto, baseado na transformação do par estéreo em um anáglifo verde-magenta, e também descreve as estruturas criadas para possibilitar a reversão do processo, que é apresentada na **Seção 4.3.** A **Seção 4.4** fornece detalhes de implementação da técnica e as ferramentas de apoio utilizadas. Os testes e discussões sobre a eficiência da técnica são analisados separadamente no **Capítulo 5**.

4.1. O problema da reversão anaglífica

O método anaglífico, descrito na **Seção 2.3.1**, é o método mais simples de visualização estereoscópica: não requer processamento complexo e custoso e o vídeo anaglífico pode ser reproduzido em qualquer tela. Dentre os outros métodos, o anaglífico também leva vantagem por ser capaz de comprimir dados. A redução de metade dos dados na criação de um anáglifo torna o processo de transformação anaglífica relevante para o tópico de codificação estereoscópica.

Entretanto, um anáglifo é uma combinação dos dados das imagens que formam um par estéreo em apenas uma imagem e, por este motivo, não pode ser utilizado por outros métodos de visualização estereoscópica – esses métodos necessitam do par estéreo como um todo. A reversão anaglífica, isto é, a reconstrução do par estéreo a partir de um anáglifo, possibilitaria o uso do par revertido por outros métodos de visualização. Desse modo, com a possibilidade da reversão, o processo de transformação anaglífica poderia ser utilizado como um método de codificação estereoscópica eficiente na compressão e independente de visualização. Essa é a proposta da técnica RevGlyph.

O processo de reversão anaglífica não é trivial, uma vez que na geração de um anáglifo perde-se informação de cor, e essa informação é necessária para a reconstrução do par estéreo. Por exemplo, como ilustrado na **Figura 4** da **Seção 2.3.1**, o anáglifo verde-magenta é resultado da combinação do canal verde (G_1) da imagem do lado direito do par estéreo, com os canais vermelho (R_2) e azul (B_2) da imagem do lado esquerdo, que juntos formam o anáglifo $R_2G_1B_2$. Os dados dos canais de cores restantes, R_1 e B_1 da imagem do lado direito e G_2 da imagem do lado esquerdo, são removidos.

Para que a reversão anaglífica seja possível é necessário reaver as informações contidas em R_1, B_1 e G_2. Uma abordagem errônea seria replicar os dados de R_2, B_2 e G_1 como forma de substituir a informação perdida. Isso não daria certo, já que cada canal traz consigo não somente informação de cor, mas também informação espacial devido ao deslocamento na captura de cada imagem. A simples replicação dos dados de cor faz apenas com que o anáglifo apareça duplicado no par estéreo, como pode ser notado na **Figura 11**.

Figura 11 - Resultado da replicação dos canais de cores de um anáglifo verde-magenta na tentativa de recriar o par estéreo

Dessa forma, o problema se encontra em como recuperar as informações de cor removidas durante o processo de conversão anaglífica, e como distribuí-las de forma correta para a reconstrução do par estéreo. Isso é definido como o "Problema da Reversão Anaglífica" e é solucionado com a técnica RevGlyph através dos processos descritos a seguir.

4.2. Processo de codificação anaglífica

A abordagem utilizada na RevGlyph para solucionar o Problema da Reversão Anaglífica faz com que seja necessário modificar o processo de transformação de um par estéreo em um anáglifo, adicionando duas novas estruturas: o Cabeçalho de Consulta de Cores e a Dife-

rença de Luminâncias. Essas estruturas, descritas a seguir, armazenam dados complementares que são utilizados pela etapa de reversão anaglífica para reconstrução do par estéreo.

4.2.1. Cabeçalho de Consulta de Cores

A abordagem mais simples para recuperar as informações de cores removidas na transformação anaglífica é não removê-las de fato. Elas são armazenadas junto ao anáglifo em uma estrutura especial, para que possam ser consultadas na etapa da reversão anaglífica. Por essa razão, tal estrutura é chamada de Cabeçalho de Consulta de Cores.

Retomando a **Figura 4** da **Seção 2.3.1**, observa-se que a saída do processo de conversão anaglífica é o anáglifo verde-magenta $R_2G_1B_2$. Na abordagem da RevGlyph, os canais de cor não utilizados, R_1, G_2 e B_1, não são removidos e juntos formam um segundo anáglifo. Para efeitos de distinção, o anáglifo $R_1G_2B_1$é chamado de "anáglifo complementar", enquanto o $R_2G_1B_2$ é chamado de "anáglifo principal". É escolhido converter o par estéreo em um anáglifo verde-magenta, já que este foi apontado como o melhor dentre os tipos de anáglifos possíveis (**Seção 2.6**).

O armazenamento do anáglifo complementar no Cabeçalho de Consulta de Cores faz com que o processo de codificação perca sua característica de compressão, já que neste caso o par estéreo como um todo está sendo armazenado com uma distribuição diferente dos canais de cores. Desse modo, é necessário utilizar estratégias que minimizem a quantidade de dados a serem armazenadas no Cabeçalho.

Tendo em vista que o interesse é manter as informações de cor, uma estratégia conveniente é converter o anáglifo complementar para o espaço de cor YC_bC_r. Desse modo, é possível separar a parte interessada, isto é, as componentes de crominância C_b e C_r do anáglifo complementar, e armazenar somente essas componentes no Cabeçalho de Consulta de Cores. Além disso, o espaço YC_bC_r possibilita a realização da subamostragem de crominância, diminuindo a quantidade de dados de C_b e C_r. Como visto na **Seção 3.5**, a melhor opção para vídeos estereoscópicos, sem perdas na percepção de profundidade, é utilizar a subamostragem 4:4:0, que diminui pela metade os dados contidos em cada componente. Utilizando essa estratégia é possível então minimizar a quantidade de dados a serem armazenados no Cabeçalho.

É importante relembrar que a subamostragem 4:2:0, a qual é utilizada em vídeos armazenados em DVD, possibilita maior compressão nos dados de crominância. Apesar disso,

Andrade *et al.* (2012) mostrou que essa subamostragem gera artefatos de compressão que prejudicam a percepção de profundidade. De modo a manter um padrão de qualidade foi decidida a utilização da subamostragem 4:4:0.

Por fim, vale ressaltar que no processo de codificação da RevGlyph, as etapas de conversão para o espaço de cor YC_bC_r e a subamostragem 4:4:0 são também aplicadas ao anáglifo principal, de modo a aumentar ainda mais a compressão. Além disso, ainda é possível utilizar-se de técnicas de compressão sem perdas, tanto no Cabeçalho de Consulta de Cores quanto no anáglifo principal.

4.2.2. Diferença de Luminâncias

Na descrição da estratégia utilizada para reduzir os dados inseridos no Cabeçalho de Consulta de Cores, nada foi dito da componente de luminância Y. Esta é explorada pela estrutura chamada de "Diferença de Luminâncias", trabalhando-se com uma redundância presente nas imagens anaglíficas, explicada a seguir.

As informações de luminância tanto do anáglifo principal, denominada Y_p, quanto do complementar, denominada Y_c, são muito semelhantes entre si. Um exemplo pode ser observado nas duas primeiras imagens da **Figura 12**, que ilustram as componentes de luminância de um anáglifo principal e seu complementar.

De modo a explorar essa redundância existente entre as componentes de luminância, é feita uma operação de subtração entre os pixels homólogos em cada anáglifo, que resulta na terceira imagem da **Figura 12**. O resultado das diferenças entre as luminâncias é denominado Y_d. Pode-se observar que do resultado dessas diferenças vários valores são próximos de zero e também muito próximos entre si. Esses valores são armazenados na Diferença de Luminâncias e a proximidade de valores é explorada para obter compressão.

A abordagem para compressão nos dados armazenados na Diferença de Luminâncias é realizar um agrupamento de intervalos de valores sequenciais em Y_d ao invés de armazenar cada um individualmente. Isso é realizado por meio da criação de blocos de repetição formados por par de coordenadas contendo um valor e a quantidade de vezes que este se repete. Parece ser uma abordagem semelhante ao algoritmo por carreiras (*run-legth*), porém, neste caso, os valores não precisam ser exatamente iguais para serem agrupados, apenas bem próximos. Valores vizinhos cuja diferença fique dentro de determinado limiar são

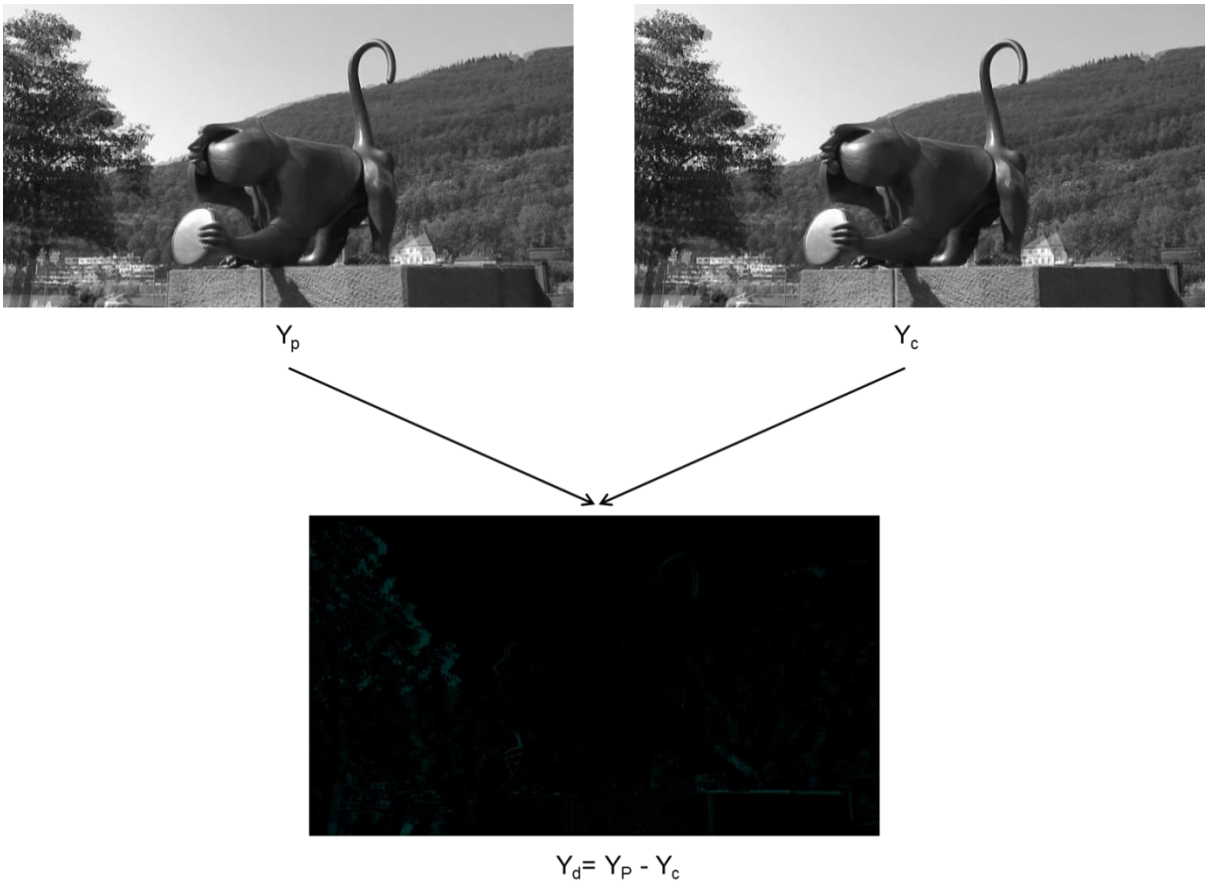


Figura 12 - Exemplo de redundância existente entre as componentes de luminância dos anáglifos formados a partir de um par estereoscópico

considerados iguais e agrupados dentro de um mesmo bloco de repetição, cujo valor final é a média dos valores armazenados. Esse agrupamento por limiar de aceitação fica mais claro com um exemplo.

Suponha que foi realizada a diferença entre as luminâncias de duas imagens anáglifas e um fragmento da cadeia de valores do resultado dessas diferenças se encontra na **Figura 13 (A)**. Na abordagem do agrupamento por limiar de aceitação é fixado o primeiro valor como base de comparação; os valores da sequência são subtraídos dessa base e o resultado absoluto é comparado com um limiar preestabelecido. No caso do exemplo, o limiar é 2. Isso significa que se o resultado absoluto da diferença entre um valor da sequência e a base for menor ou igual a 2, esse valor é considerado como sendo igual à base e pertencente ao mesmo agrupamento. Isso é feito por toda a cadeia até ser encontrado um valor cuja diferença ultrapasse o limiar adotado. No exemplo, isso ocorre quando a base é subtraída do valor 5, cujo resultado absoluto é 4, maior do que o limiar adotado. Nesse caso, encerra-se um bloco de agrupamento e inicia-se outro bloco, cujo valor base agora é 5. O bloco de

agrupamento encerrado vai ser formado por um par de coordenadas contendo o teto da média dos valores existentes no bloco e a quantidade de repetições (**Figura 13 (B)**).

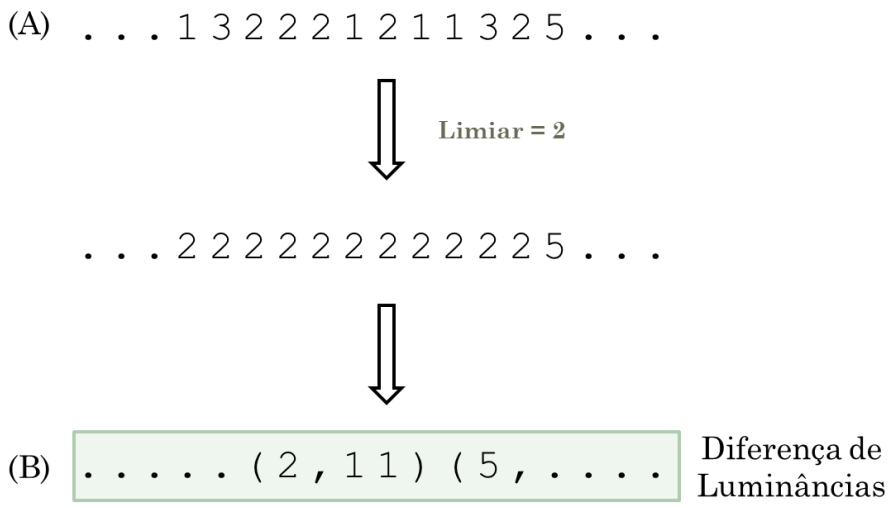


Figura 13 - Exemplo da aplicação do agrupamento por limiar de aceitação nos valores da Diferença de Luminâncias

Vale apontar que a escolha do limiar a ser adotado possibilitará um agrupamento variável, sendo que limiares mais altos poderão acarretar em agrupamentos maiores, já que aceitam uma diferença maior entre os valores vizinhos. Entretanto, isso pode comprometer a qualidade da imagem, visto que um pixel pode se distanciar demais do seu valor original, e deve ser levado em conta na hora da escolha do limiar. No **Capítulo 5** é analisado qual limiar possibilita um melhor balanço entre compressão e qualidade.

O processo completo de codificação estereoscópica, utilizando a transformação anaglífica, o Cabeçalho de Consulta de Cores e a Diferença de Luminâncias, está ilustrado na **Figura 14**. Vale observar que ao final do processo, tanto o anáglifo principal, quanto o Cabeçalho de Consulta de Cores e a Diferença de Luminâncias passam pela etapa de compressão sem perdas e são agrupados em um único arquivo comprimido. Essa última etapa de compressão é feita utilizando-se dos algoritmos LZSS (Lempel–Ziv–Storer–Szymanski) e PPM (*Prediction by Partial Matching*).

Pode ser observado que o Cabeçalho de Consulta de Cores e a Diferença de Luminâncias são na verdade resultado de transformações feitas nas informações descartadas no processo tradicional de conversão anaglífica, de modo que é armazenado apenas o mínimo necessário para a reversão do processo, não comprometendo severamente o nível de compressão adquirido pelo uso da codificação anaglífica. A análise de quanto cada estrutura influencia no tamanho final do arquivo codificado, bem como qual o impacto na qualidade de acordo com o limiar adotado, é apresentada no **Capítulo 5**.

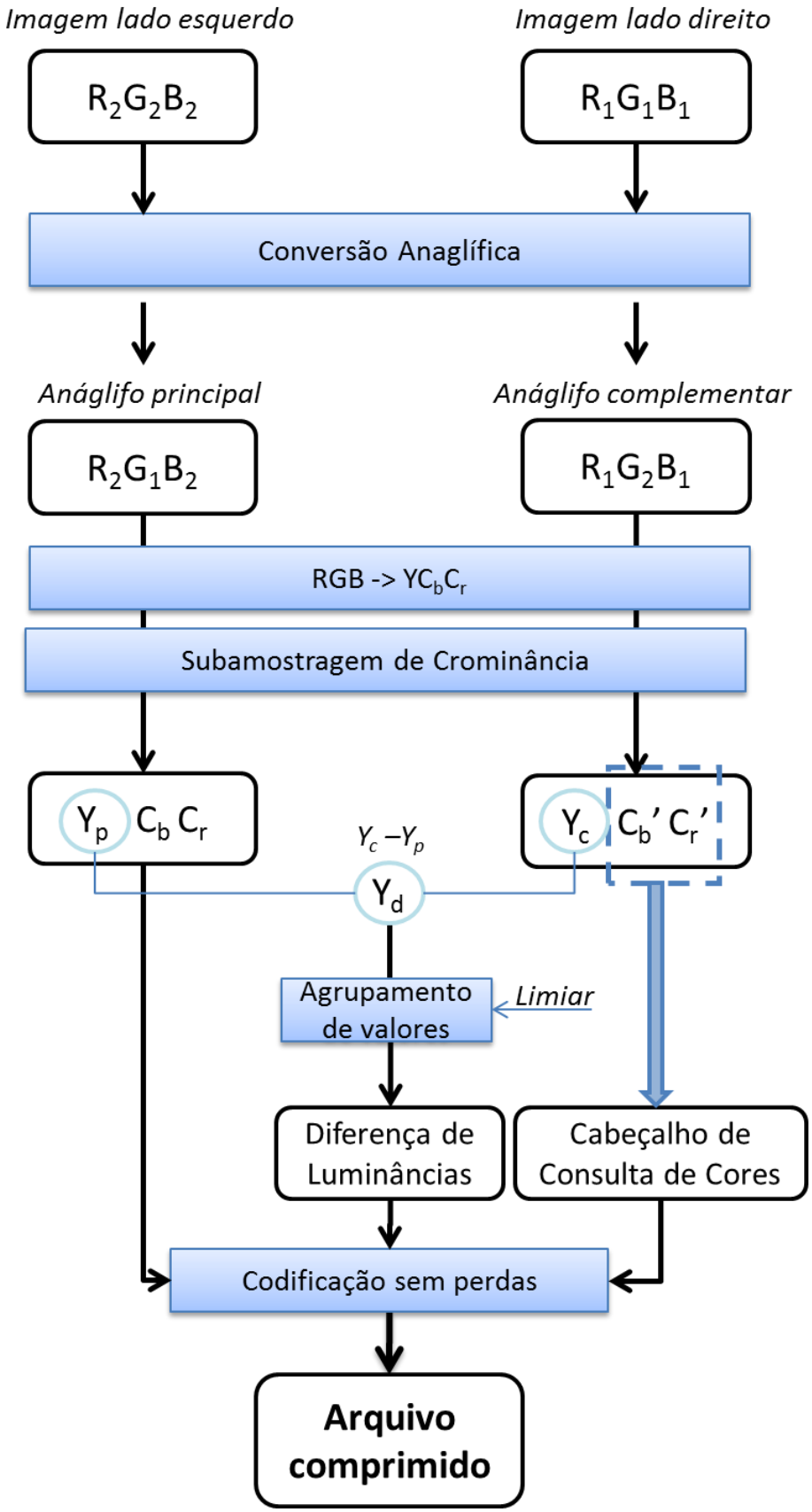


Figura 14 – RevGlyph : processo de conversão anaglífica utilizando o Cabeçalho de Consulta de Cores e a Diferença de Luminâncias

4.3. Reversão anaglífica

A reversão anaglífica é baseada na extração das informações contidas no Cabeçalho de Consulta de Cores e na Diferença de Luminâncias, de modo a recriar o anáglifo complementar e assim obter uma aproximação do par estéreo original correspondente. O processo de reversão é ilustrado na **Figura 15** e descrito a seguir.

Os dados contidos na Diferença de luminância se encontram agrupados pela repetição de valores. A matriz de luminância Y_d é reconstruída repetindo os valores na quantidade de vezes indicadas no seu par de coordenadas. A matriz resultante é somada aos dados de Y_p de modo a reconstruir Y_c. Unindo Y_c com os valores de crominância subamostrados armazenados no Cabeçalho de Consulta de Cores possibilita que o anáglifo complementar seja reconstruído.

Ambos os anáglifos, subamostrados, passam por uma etapa de recomposição de crominância, de modo a voltarem para o modelo 4:4:4. Como a subamostragem 4:4:0 faz uma redução de metade da informação de cada crominância, os valores de pixel existentes são replicados para substituírem os removidos na subamostragem, de modo que as três componentes fiquem com a mesma dimensão e quantidade de pixels, e possam ser revertidas para o espaço de cor RGB. Dessa forma, têm-se seis canais RGB desordenados: $R_2G_1B_2$ (anáglifo principal) e $R_1G_2B_1$ (anáglifo complementar); basta então uma reordenação dos canais para obter o par estéreo.

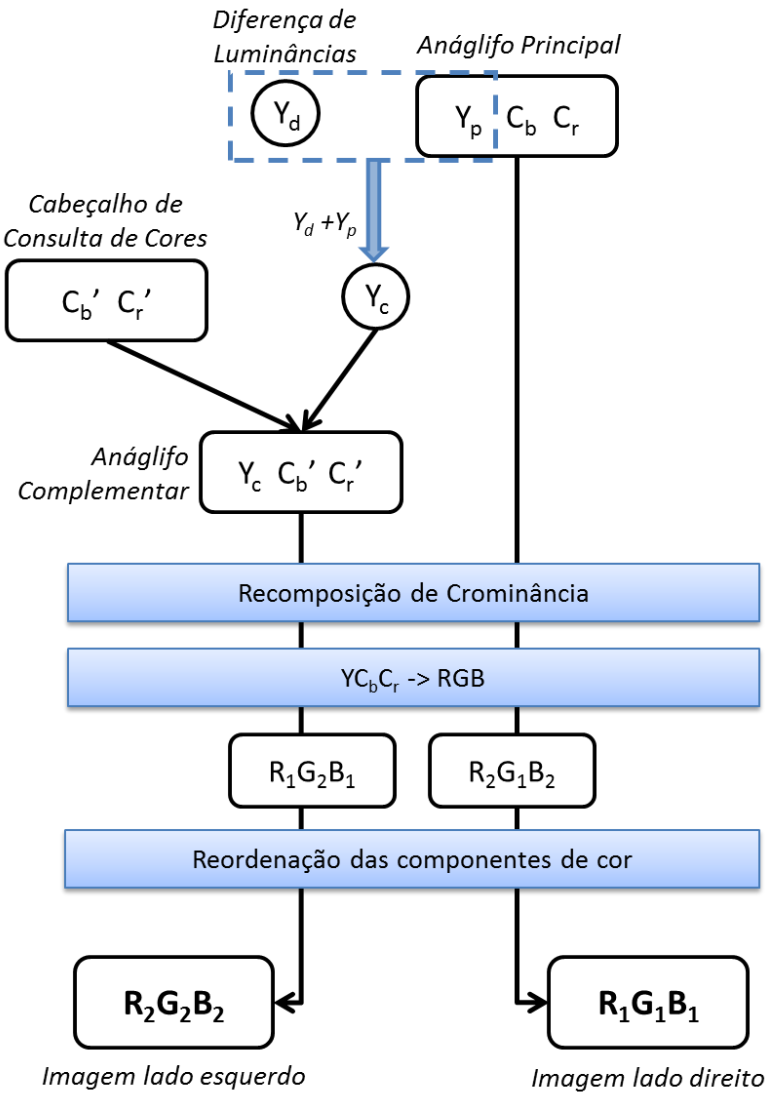


Figura 15 – RevGlyph: Processo de reversão anaglífica

4.4. Implementação da técnica

A RevGlyph foi codificada utilizando a linguagem C, que possui funções e estruturas que dão a liberdade de se acessar a memória do dispositivo em que o programa é executado, alocando e operando com a quantidade de bytes desejada, além de possibilitar a criação de um programa modular e otimizado em termos de processamento.

Como a RevGlyph envolve operações com imagens e armazenamento destas em disco, optou-se por utilizar o suporte da biblioteca OpenCV[11], que é *open source* e contém uma vasta variedade de funções e estruturas úteis, já otimizadas para processamento de imagens e visão computacional, sendo bem difundida e utilizada em áreas correlacionadas. A utilização da biblioteca facilitou o trabalho com estruturas de imagens, permitindo que o foco do desenvolvimento se mantivesse em solucionar o problema. Uma das funções utilizadas foi a conversão de RGB para YC_bC_r e sua função reversa[12], cujo código se apresenta aperfeiçoado para melhorar a qualidade da imagem.

A subamostragem 4:4:0 foi codificada manualmente, com os pixels de crominância amostrados da seguinte forma: os elementos das matrizes de cada componente são lidos da esquerda para a direita, de cima para baixo, a começar do topo superior esquerdo. Os valores de Y são todos lidos e armazenados. Já os valores de C_b e C_r são calculados a partir da média aritmética dos valores de um pixel da linha par e outro pixel logo abaixo dele na linha ímpar da matriz. Dessa forma, a subamostragem é realizada de maneira vertical. O processo para a recomposição da crominância para 4:4:4 também foi codificado manualmente e simplesmente replicava os dados de Cb e Cr a cada linha par e ímpar seguidas.

Todos os processamentos com imagens, desde a abertura e gravação desta, até leitura e operação com valores dos pixels, foi feito utilizando uma estrutura própria fornecida pela biblioteca OpenCV, chamada `IplImage`. Detalhes e exemplos de uso podem ser encontrados em Bradski e Kaehler (2008).

As transformações e operações com os valores dos pixels foram armazenados em caracteres (variáveis do tipo `unsigned char`), com valores arredondados nos casos com resultados envolvendo ponto flutuante; isso de modo a manter um baixo volume de dados e valores num intervalo de 0 a 255 (modelo RGB com 8 bits de profundidade por pixel).

---

[11] A biblioteca OpenCV pode ser obtida em http://sourceforge.net/projects/opencvlibrary/. Acesso em 19 nov. 2013.

[12] Explicações sobre a função de conversão de espaço de cores se encontra em http://opencv.willowgarage.com/documentation/miscellaneous_image_transformations.html?highlight=cvcvtcolor#cvCvtColor. Acesso em 19 nov. 2013.

4.5. Considerações finais

Este Capítulo apresentou a técnica RevGlyph, uma proposta de codificação estereoscópica baseada na conversão e reversão anaglífica que possibilita a codificação de vídeos estereoscópicos com boa compressão, preservando a percepção de profundidade. Devido à possibilidade de reversão anaglífica, a RevGlyph se destaca por ser genérica, podendo ser utilizada por qualquer tipo de visualização estereoscópica.

A reversão anaglífica não é um processo direto, e só foi possível graças à utilização de duas estruturas, o Cabeçalho de Consulta de Cores e a Diferença de Luminâncias, que guardam informações mínimas necessárias para a reconstrução de um par estéreo. A maneira com que as duas estruturas são preenchidas é feita de modo a comprometer minimamente a compressão obtida com a utilização da transformação anaglífica.

Cabe aqui ressaltar a necessidade da existência da estrutura de Diferença de Luminâncias. Se fosse considerada uma abordagem alternativa de codificação anaglífica sem a Diferença de Luminâncias, a reversão ainda seria possível, bastando replicar a informação de luminância do anáglifo principal para a reconstrução do anáglifo complementar. Entretanto, as informações de luminância do anáglifo complementar, embora muito semelhantes com as informações de luminância do anáglifo principal, possuem dados de intensidade deslocados que alteram a crominância da imagem. O resultado final dessa abordagem alternativa resultaria na criação de um par estéreo revertido com ruídos (*crosstalk*), como pode ser visto na **Figura 16**.

Embora esta abordagem alternativa resulte em um par estéreo com *crosstalk*, ela tem o mérito de produzir um arquivo codificado com mais compressão. Mais do que isso, ao se considerar apenas o caso particular da visualização anaglífica verde-magenta, o par estéreo revertido é capaz de criar outro anáglifo verde-magenta de boa qualidade, sem danificar a percepção de profundidade, já que os dados necessários para a produção do anáglifo se mantêm intactos. Por outro lado, para conservar a generalidade da RevGlyph, a Diferença de Luminâncias é necessária.

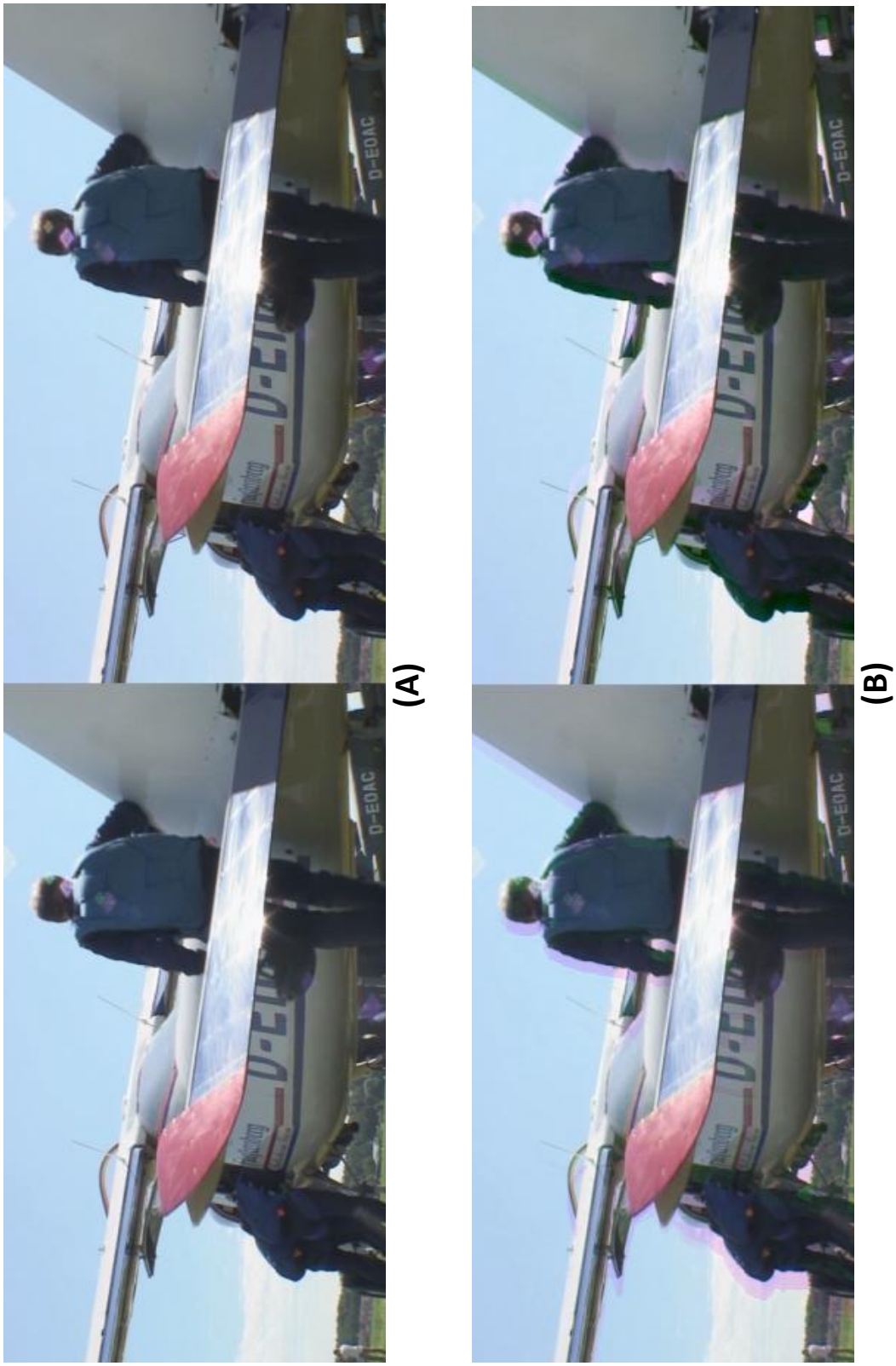


Figura 16 - Comparação qualitativa do par estéreo original (A) e o obtido pelo processo de reversão anaglífica utilizando apenas o Cabeçalho de Consulta de Cores (B). É possível observar o *crosstalk* em (B), representado em tons de verde e magenta

5. Análise da RevGlyph

O **Capítulo 4** fez a apresentação da RevGlyph e das estruturas auxiliares que possibilitam o processo de reversão anaglífica, sendo também justificada a necessidade de cada uma delas. Neste Capítulo analisa-se o impacto dessas estruturas quando aplicadas à técnica, tanto em termos de qualidade de imagem do par estéreo revertido quanto em termos de como elas afetam a taxa de compressão.

Observando novamente a **Figura 14** (**Seção 4.2**), três etapas no processo de conversão anaglífica da RevGlyph são analisadas para verificar seu impacto na qualidade da imagem processada: (1) conversão do espaço de cores; (2) subamostragem de crominância; e (3) diferença de luminâncias (Y_d). Essas etapas são analisadas separadamente nas Seções a seguir.

O Capítulo encerra com uma última análise: a comparação da RevGlyph com outra técnica de codificação estereoscópica estado da arte – o mapa de profundidades –, de modo a observar a competitividade e eficácia da RevGlyph perante outros codificadores.

5.1. Análise da conversão de espaço de cores

O processo de conversão do espaço de cores utilizado pela RevGlyph é a conversão $RGB \rightarrow YC_bC_r$ e sua inversa, $YC_bC_r \rightarrow RGB$. A mudança de um espaço para outro pode ser feita utilizando transformações matemáticas envolvendo as componentes de cor de cada espaço, conforme as **Equações 1 e 2** da **Seção 3.1**.

Por envolver apenas transformação dos dados, é esperado que o resultado final entre conversões não cause perdas, isto é, a aplicação da sequência de conversões $RGB \rightarrow YC_bC_r \rightarrow RGB$ a uma imagem deve ter como resultado final a mesma imagem, com o mesmo valor em cada pixel quando comparada à original. Entretanto isso não se mostra verdadeiro na prática: as equações envolvem cálculos com valores decimais; na implementação da técnica os valores são armazenados em variáveis de caractere, resultando em arredonda-

mento do resultado pelo dispositivo em que o processamento é feito. Esse arredondamento gera ruídos na imagem processada.

Os testes realizados com as conversões RGB→YC_bC_r e YC_bC_r→RGB foram feitos para analisar o quanto esses ruídos influenciam na qualidade de uma imagem. Com isso, é possível isolar a influência da conversão frente às outras componentes da RevGlyph. Os testes foram realizados implementando-se a sequência de conversões RGB→YC_bC_r→RGB e, logo após, calculando-se o valor de *Peak Signal-to-Noise Ratio* (PSNR) resultante da comparação da imagem original com a processada.

O PSNR é uma métrica de qualidade muito utilizada para comparação de imagens processadas (WINKLER, 2005). O cálculo é feito comparando-se pixel a pixel a imagem original com a respectiva processada, sendo o resultado medido em decibéis (dB) numa escala de 0 a 100 dB, de forma que quanto mais próximo de 100 dB, menor o nível de ruído encontrado na imagem processada, ou seja, mais semelhantes elas são (no nível de pixel).

Segundo Ebrahimi, Chamik e Winkler (2004), o PSNR não possui a palavra final em termos de qualidade de imagem, já que não leva em consideração a percepção visual humana. Desse modo, uma imagem com baixo PSNR não tem necessariamente uma baixa qualidade quando visualizada. Por outro lado, a métrica é bem sucedida para analisar introdução de ruído em imagens e, por isso mesmo, é muito utilizada em trabalhos relacionados, servindo como base para comparação entre técnicas.

Dado que a conversão de espaço de cores pode ser aplicada a qualquer imagem, não se restringindo a imagens estereoscópicas, utilizou-se para os testes bibliotecas com grande quantidade de imagens de tipos variados. As bibliotecas utilizadas foram a *Amsterdam Library of Object Images* (ALOI)[13] e a Corel1000[14].

A biblioteca ALOI, criada por Geusebroek, Burghouts e Smeulders (2005), é uma coleção formada por mil objetos capturados em diferentes ângulos e variações de cores em um fundo preto, totalizando 110250 imagens diferentes, salvas em formato PNG, com resolução 384x288 pixels. Para os testes, foi utilizada apenas a coleção "*wide-baseline stereo*", com 2250 itens. A biblioteca Corel1000, utilizada por Wang, Li e Wiederhold (2001), é composta por 1000 imagens variadas, de diferentes resoluções, salvas em formato JPEG. Todas as imagens utilizadas foram convertidas para BMP antes de serem processadas, mantendo a resolução e a profundidade de 8 bits por pixel em cada componente R, G e B.

---

[13] A biblioteca ALOI pode ser obtida em http://staff.science.uva.nl/~aloi/. Acesso em 19 nov. 2013.

[14] A biblioteca Corel1000 pode ser obtida em http://wang.ist.psu.edu/docs/related.shtml. Acesso em 19 nov. 2013.

O resultado dos testes aplicados às duas bibliotecas encontra-se na **Tabela 2**. Os valores de PSNR são apresentados separadamente por componentes, e uma média aritmética das três é apresentada na última coluna. Na última linha da tabela é apresentada outra média, composta pelos valores de PSNR calculados para cada biblioteca.

Tabela 2 - PSNR da conversão RGB→YC_bC_r→ RGB

Base de Imagens	Y(dB)	Cb(dB)	Cr(dB)	Média(dB)
Corel1000	90,67	95,53	94,20	93,47
ALOI *stereo*	99,87	100,00	100,00	99,96
MÉDIAS	95,27	97,76	97,10	96,71

Os dados da **Tabela 2** mostram que os arredondamentos feitos na execução da conversão RGB→YC_bC_r→RGB introduz, ainda que pouco, ruídos na imagem processada em termos de PSNR. A presença de ruídos é maior na biblioteca Corel1000, em cerca de 6,5dB. A biblioteca ALOI apresenta valores de PSNR mais próximos de 100dB, com diferenças significantes apenas na quarta casa decimal. Isso se deve ao fato de que as imagens possuem dimensões menores e também porque cada captura é formada por apenas um objeto em um fundo preto.

O **Gráfico 1** apresenta a distribuição dos resultados de PSNR das imagens da biblioteca Corel1000. Pode-se observar que embora na média o PSNR fique acima de 90dB, ainda há a presença de algumas imagens mais afetadas pela mudança de espaço de cor. Mesmo assim são bem poucas, sendo que apenas 4% delas apresentam resultados inferiores a 80dB. A distribuição de resultados da biblioteca ALOI é omitida, pois a quantidade de resultados inferiores a 99dB (entre 95 e 99dB) foi pouco mais de 1%.

Em termos de programação, a precisão nos cálculos poderia aumentar caso os resultados da conversão fossem armazenados em variáveis de ponto flutuante (`float` ou `double` em C), o que implicaria em menos ruído. Entretanto, isso também implicaria em um aumento no volume de dados a serem armazenados devido à diferença na quantidade de bytes armazenado por valor, o que inviabiliza esse tipo de implementação na RevGlyph. Como na média geral da **Tabela 2** a perda é de pouco mais de 3dB, a aplicação da conversão entre RGB e YC_bC_r armazenando os dados em variáveis de 8 bits (`unsigned char` em C), como é feito na RevGlyph, é aceitável.

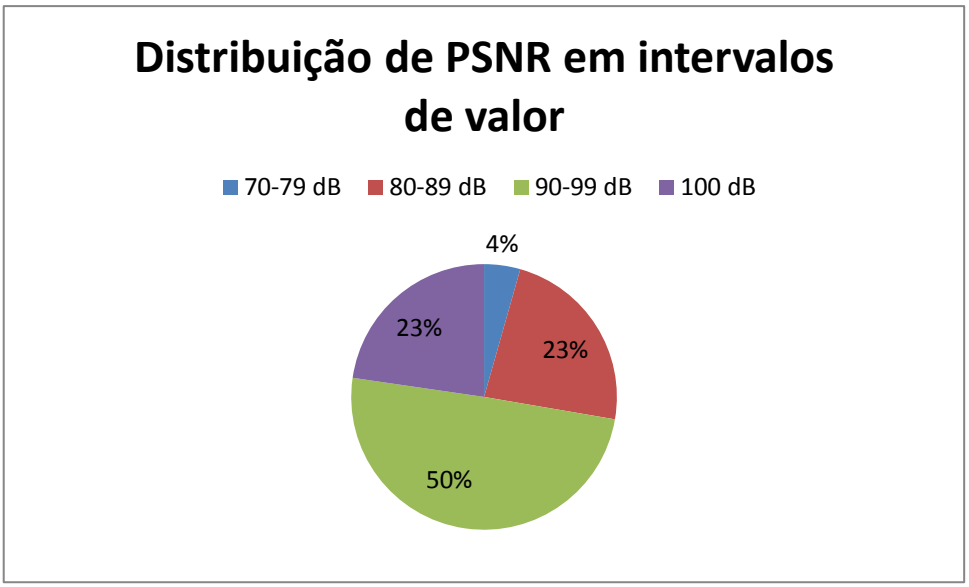


Gráfico 1 - Distribuição do PSNR calculado para imagens da biblioteca Corel1000, processadas apenas pela conversão RGB→YCbCr→RGB

5.2. Análise da subamostragem de crominância

A subamostragem 4:4:0, utilizada na RevGlyph, diminui a resolução das cores pela metade, eliminando 50% dos dados da componente de crominância C_b e 50% dos dados da componente de crominância C_r. O quanto essa redução inclui de ruído à imagem ao refazer a composição para o tipo 4:4:4 é verificada no teste a seguir.

Para o teste, a etapa de subamostragem foi adicionada à implementação descrita na **Seção 5.1**. Utilizaram-se novamente as bibliotecas Corel1000 e ALOI. O procedimento dos testes foi converter cada imagem de RGB para YC_bC_r, aplicar a subamostragem 4:4:0, realizar a recomposição para 4:4:4 e então retornar ao espaço de cores RGB. Por fim, é calculado o PSNR entre a imagem original e a imagem processada.

A **Tabela 3** apresenta os resultados dos testes, com os valores de PSNR divididos por componentes, e uma média aritmética das três na última coluna. A última linha da tabela apresenta uma outra média, feita com os valores de PSNR calculados por componente em cada biblioteca.

Tabela 3 - PSNR da conversão RGB→YC_bC_r→ RGB juntamente com subamostragem de crominância

Base de Imagens	Y(dB)	Cb(dB)	Cr(dB)	Média(dB)
Corel1000	71,99	52,36	52,50	58,95
ALOI *stereo*	76,99	48,06	48,55	57,87
MÉDIAS	74,49	50,21	50,52	58,41

É interessante notar nos dados da tabela que os valores de PSNR em Y também sofreram baixas, reduzindo em média mais de 25dB, mesmo a subamostragem tendo sido aplicada apenas aos valores de crominância. Isso tem sua explicação: o PSNR é calculado no espaço de cores YC_bC_r. Isso significa que a imagem processada (que passou pelo processo de subamostragem e retornou para o espaço RGB) é transformada novamente no formato YC_bC_r de modo a ser calculado o PSNR. O cálculo de Y envolve a utilização dos valores de R, G e B (**Equação 1** da **Seção 3.1**), os quais possuem ruídos já advindos da subamostragem. Com isso, tais ruídos são refletidos nos valores calculados para Y nesta última transformação para YC_bC_r e influenciam no resultado do cálculo do PSNR.

O nível de ruído presente nas imagens ficou em média 58,41dB, com uma queda de quase 50dB em cada crominância, um reflexo da diminuição dos dados de crominância pela metade.

O **Gráfico 2** mostra a distribuição do PSNR das imagens processadas da biblioteca Corel1000, bem diferente da obtida no **Gráfico 1**. Observa-se que o PSNR máximo obtido está na faixa de 70-79dB e corresponde a apenas 1% dos resultados, com a maior parte se concentrando na faixa de 50-59dB.

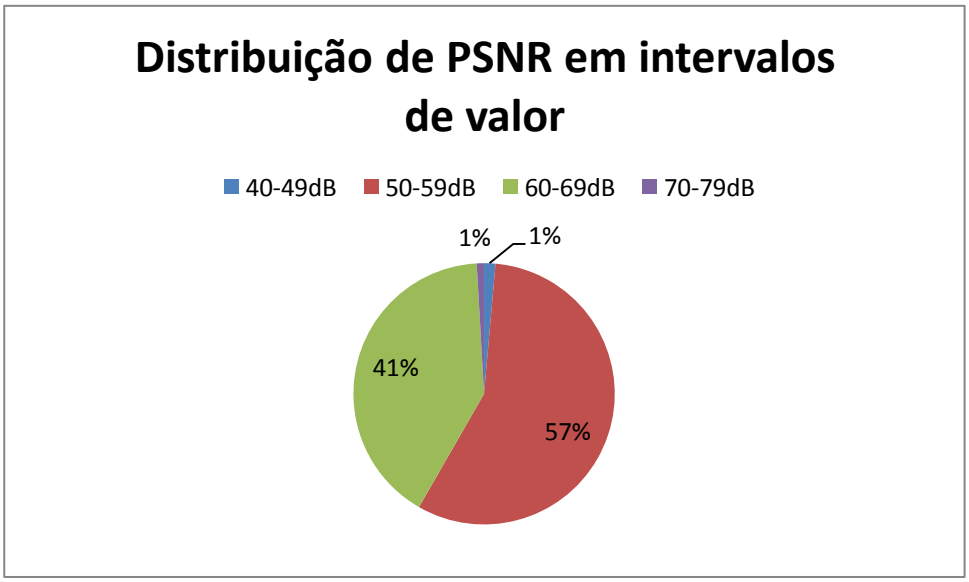


Gráfico 2 - Distribuição do PSNR calculado para imagens da biblioteca Corel1000, processadas pela conversão RGB→YCbCr→RGB e subamostragem de crominância

Os resultados apresentados demonstram como a etapa de subamostragem de crominância afeta severamente a qualidade objetiva de uma imagem, diminuindo o PSNR em torno de 58,41 dB. Entretanto, os efeitos visuais ao se observar uma imagem que passou por subamostragem são mínimos. A **Figura 17** ilustra um exemplo de uma das imagens da biblioteca Corel100 utilizada nos testes. Em (A) se encontra a imagem original e em (B) se encontra a mesma imagem subamostrada. Os valores de PSNR desta imagem foram Y=59,70dB, C_b=53,24dB e C_r=54,45dB. Mesmo com valores baixos, as diferenças entre as duas imagens são praticamente imperceptíveis.

(A) (B)

Figura 17 - Comparação entre uma imagem original (A) e sua correspondente processada com subamostragem 4:4:0 (B). Imagem retirada da base Corel1000, idBeach59.jpg

5.3. Análise da Diferença de Luminâncias

A **Seção 4.2** descreveu como a estratégia da Diferença de Luminâncias foi desenvolvida no contexto da RevGlyph. Essa estratégia, embora traga compressão, necessita ser analisada para verificar o quanto a aplicação de um limiar afeta a qualidade.

Para a análise, foram aplicados limiares da escala de 1 a 5. Isso quer dizer que no limiar 1 é aceitável uma diferença de no máximo 1 entre a base e o valor da sequência. O mesmo raciocínio se estendendo até o limiar 5, em que é aceitável uma diferença de no máximo 5 entre a base de comparação e o valor da sequência. Caso o próximo valor da sequência ultrapasse o limiar, um novo agrupamento é feito pelo algoritmo. Além da escala de 1 a 5, foi também testada a Diferença de Luminâncias sem a aplicação de um limiar, isto é, nenhuma

diferença é aceitável entre números consecutivos (algoritmo *run-length*). Testes preliminares mostraram poucas mudanças na compressão e qualidade objetiva das imagens quando utilizados limiares acima de 5, e por isso decidiu-se por limitar a análise até este limiar.

Utilizou-se para o teste uma base de vídeos construída por Andrade, Cordebello e Goularte (2010)[15] contendo 32 pares de vídeos estereoscópicos no formato lado-a-lado. De cada vídeo da base, foi extraído um quadro, obtendo-se assim uma base de testes formada por 32 imagens estereoscópicas.

Os testes foram feitos em 6 lotes, sendo o primeiro a execução da técnica RevGlyph sem aplicar nenhum limiar na construção da Diferença de Luminâncias. Um limiar incremental foi aplicado aos outros cinco lotes.

Em cada lote de teste o procedimento adotado foi o seguinte: inicialmente é executada a etapa de conversão anaglífica da RevGlyph e extraída a taxa de compressão por meio da divisão do par estéreo processado pelo original; além disso, a quantidade de dados adicionais (*overhead*) utilizados tanto para a construção do Cabeçalho de Consulta de Cores quanto para a Diferença de Luminâncias é previamente calculada antes de sua inserção ao arquivo original, possibilitando saber a porcentagem de cada *overhead* em relação ao tamanho do total do arquivo. Posteriormente é executada a etapa de reversão anaglífica da RevGlyph e calculado o valor de PSNR entre o par estéreo revertido e o original.

A **Tabela 4** mostra os resultados sumarizados de cada lote de teste. A primeira coluna identifica o limiar. A segunda coluna apresenta a taxa média de compressão atingida com a codificação anaglífica aplicada às 32 imagens estereoscópicas. A terceira coluna destaca a quantidade de dados necessária para a criação do Cabeçalho de Consulta de Cores, expresso pela porcentagem que este ocupa no tamanho do arquivo final. A quarta coluna é similar à terceira, apresentando a porcentagem ocupada pelos dados da Diferença de Luminâncias, variáveis de acordo com o limiar adotado. Por fim, a quinta coluna apresenta a média do PSNR calculado após aplicar a reversão anaglífica.

O *overhead* do Cabeçalho de Consulta de Cores se manteve o mesmo em todos os lotes de teste, já que este apenas armazena os dados de crominância subamostrados do anáglifo complementar, os quais não são processados pela abordagem de compressão utilizando os limiares.

[15] A base de vídeos estereoscópicos se encontra disponível em http://200.136.217.194/videoestereo/. Acesso em 19 nov. 2013.

Tabela 4 - Resultados da utilização da técnica RevGlyph com diferentes limiares para a compressão da Diferença de Luminâncias

Limiar	Taxa de Compressão (%)	*Overhead* do Cabeçalho (%)	*Overhead* da Diferença de Luminâncias (%)	PSNR (dB)
0	72,58	6,53	7,06	45,2410
1	74,78	6,53	4,86	40,5443
2	76,02	6,53	3,62	38,9700
3	76,78	6,53	2,87	37,7620
4	77,29	6,53	2,35	36,8650
5	77,67	6,53	1,97	36,2727

Pode-se observar pelos dados da **Tabela 4** que a cada aumento no limiar a taxa de compressão também aumenta; em contrapartida, o valor do PSNR diminui. Isso é esperado, já que cada incremento de limiar remove mais informação da imagem original. Entretanto, do limiar 2 em diante, tanto a taxa de compressão quanto o PSNR parecem manter uma estabilidade, com poucas alterações de valor na transição de limiar. Isso pode ser melhor notado com a ajuda do **Gráfico 3**, em que estão plotados os pontos que relacionam a taxa de compressão (eixo X) com o PSNR (eixo Y) e uma linha de tendência linear para comparar as mudanças.

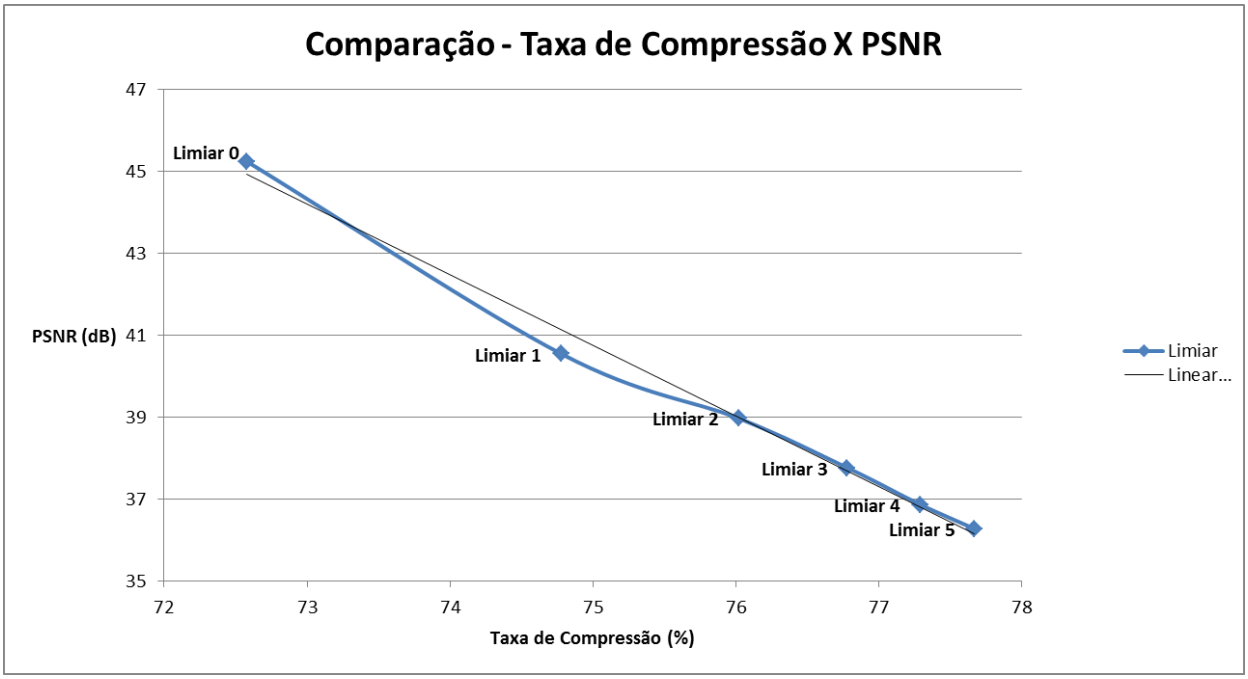


Gráfico 3 - Gráfico de comparação da taxa de compressão e PSNR obtido em cada limiar aplicado com a técnica RevGlyph

É possível verificar no **Gráfico 3** que do limiar 0 para o limiar 1 há uma queda brusca, já entre o limiar 1 e o limiar 2 a queda existente é mais branda e a partir da transição do limiar 2 para o limiar 3 em diante a queda segue uma tendência linear. Os dados no gráfico apontam o limiar 2 como sendo vantajoso em termos de custo-benefício entre qualidade e compressão. Além disso, visualmente é possível perceber um nível de qualidade inferior a partir do limiar 3 em pares estereoscópicos revertidos. Um exemplo disso é ilustrado nas **Figuras de 18 a 25**.

A **Figura 18** apresenta um par estéreo original retirado da base de testes. As **Figuras de 19 a 24** mostram o resultado da reversão utilizando a técnica RevGlyph com a aplicação dos diferentes limiares. A degradação da qualidade visual aumenta a cada transição para um limiar maior, sendo mais perceptível nas regiões em que a mudança de cor ocorre de maneira suave e gradual. Nestas regiões é notada a presença de ruído na forma de linhas horizontais, sendo mais perceptíveis a partir do limiar 3 (**Figura 22**) nas regiões circuladas em vermelho. A **Figura 25** apresenta uma ampliação de uma região da imagem estereoscópica, possibilitando verificar com mais detalhes a degradação sofrida em cada transição de limiar e a formação das linhas horizontais.

O pouco acréscimo na compressão obtido do limiar 3 em diante, tendo em vista as alterações visuais na imagem e em conjunto com o que foi analisado no **Gráfico 3**, leva à escolha da aplicação do limiar 2 à técnica RevGlyph como a melhor em termos de custo-benefício entre qualidade e compressão.

A utilização de limiares para compressão da Diferença de Luminâncias se mostrou efetiva. O *overhead* da Diferença sem utilizar limiar (primeira linha da **Tabela 4**) foi de 7,06%. Já aplicando o limiar 1 houve uma diminuição de 4.86%. No limiar 5 o *overhead* é de apenas 1,97% em relação ao tamanho total do arquivo codificado. Entretanto, a observação de cada imagem da base de testes mostrou degração visual muito aparente para limiares desse grau.

O valor do PSNR médio para o limiar 2 ficou em 38,97 dB. Desses, a mudança de espaço de cores e a subamostragem de crominância já são responsáveis por diminuir o valor do PSNR para 58,41dB. Isso significa que a responsabilidade das estruturas propostas na RevGlyph com relação à diminuição do valor do PSNR é de fato cerca de 20dB.

Figura 18 - Par estéreo original da base de imagens estereoscópicas. Código de identificação rv03.

Figura 19 - Par estéreo revertido utilizando a RevGlyph sem limiar. As mudanças visuais são pouco aparentes.

Figura 20 - Par estéreo revertido utilizando a RevGlyph - limiar 1

Figura 21 - Par estéreo revertido utilizando a RevGlyph - limiar 2

Figura 22 - Par estéreo revertido utilizando a RevGlyph - limiar 3. Pequenas alterações de cor e textura podem ser notadas nas áreas circuladas.

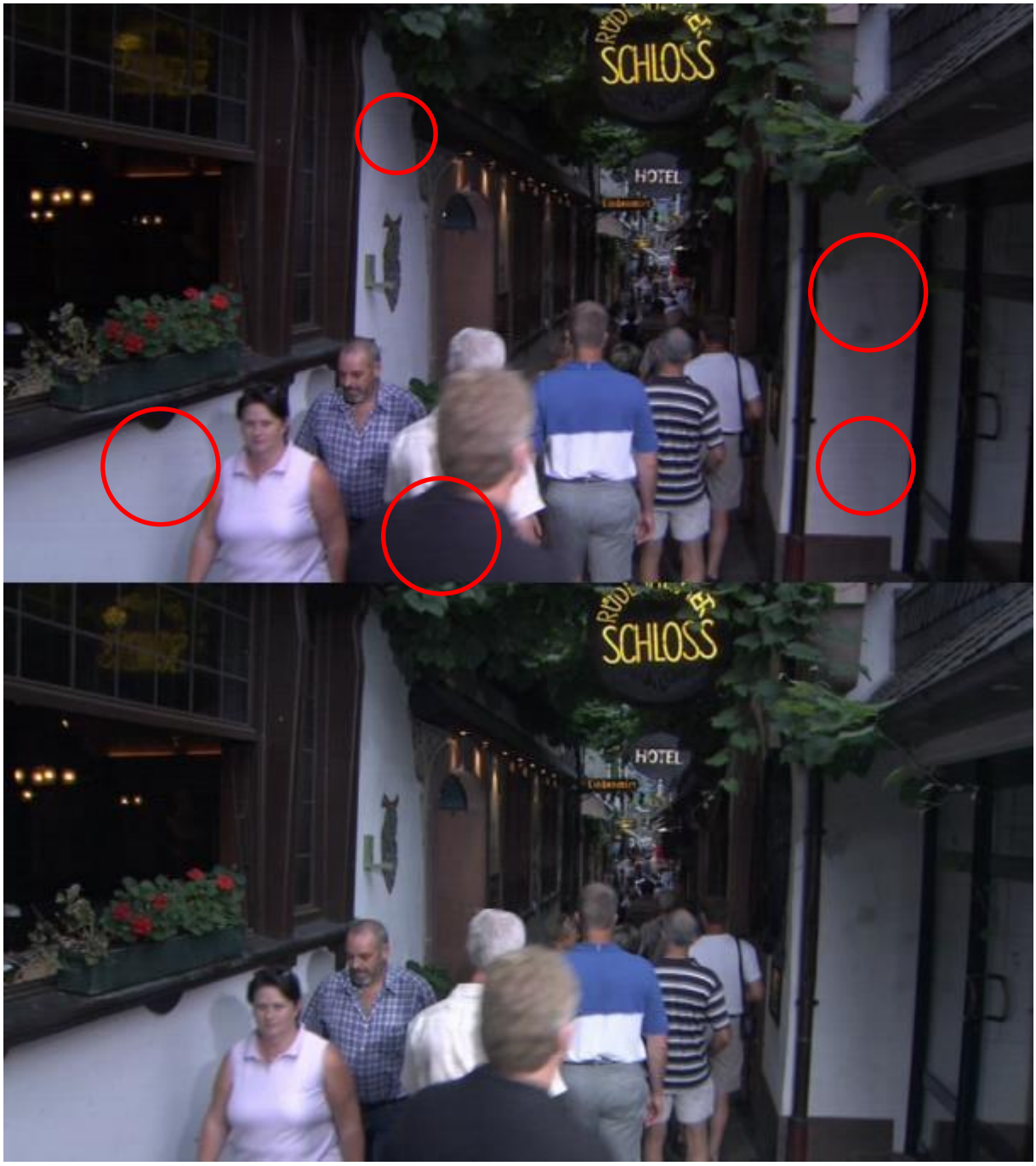


Figura 23 - Par estéreo revertido utilizando a RevGlyph - limiar 4. As alterações começam a ficar mais aparentes

Figura 24 - Par estéreo revertido utilizando o RevGlyph - limiar 5. É possível notar com nitidez a formação de texturas com linhas horizontais.

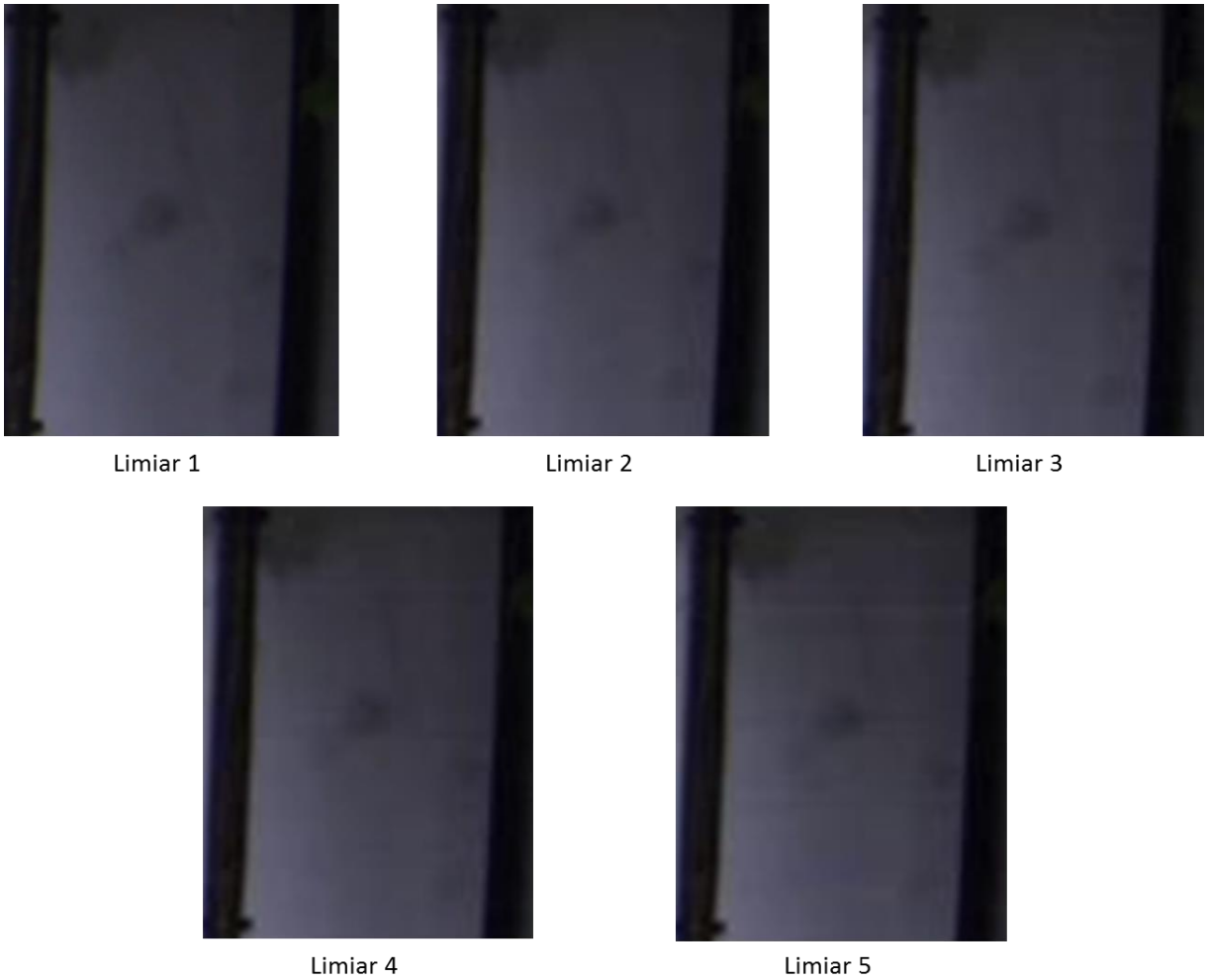


Figura 25 - Ampliação de uma região da imagem estereoscópica revertida com diferentes limiares, apresentando a degradação sofrida com o aumento do limiar e a formação de ruídos na forma de linhas horizontais

Um detalhe importante é ilustrado nas **Figuras 26** e **27**. A **Figura 26** apresenta o anáglifo verde-magenta criado utilizando um par estéreo original e a **Figura 27** apresenta o anáglifo verde-magenta criado utilizando o par estéreo codificado e revertido pela RevGlyph com limiar 2. Utilizando óculos verde-magenta apropriado, é possível observar que a percepção de profundidade se mantém na imagem revertida, o que indica que a aplicação da RevGlyph não tem um impacto severo na percepção de profundidade.

Figura 27 - Anáglifo verde-magenta criado usando um par estéreo original. Código de identificação rv03

Figura 26 - Anáglifo verde-magenta criado usando um par estéreo revertido pela RevGlyph com limiar 2

5.4. Comparação entre a RevGlyph e Mapa de Profundidades

As análises realizadas até agora mostraram a viabilidade da RevGlyph, tanto em termos de compressão quanto em termos de qualidade na imagem revertida. Nesta Seção, a técnica é confrontada com outra técnica de codificação estereoscópica estado da arte. Realiza-se um teste comparativo dos resultados da codificação das duas para verificar a eficiência e utilidade da RevGlyph.

Optou-se por comparar a RevGlyph com a codificação usando mapa de profundidades. Tanto a RevGlyph quanto mapa de profundidades compartilham de características semelhantes: ambas são utilizadas tanto como um mecanismo de compressão quanto de visualização estereoscópica, e armazenam no arquivo codificado apenas uma componente do par estéreo em conjunto com dados complementares.

A criação de mapa de profundidades e a síntese de visões por meio da técnica DIBR ainda são alvos de pesquisa para melhoria dos algoritmos e da qualidade do resultado final. Algumas pesquisas foram citadas nas **Seções 3.2.3** e **3.3.2**. Para este teste comparativo, utilizou-se da pesquisa realizada por Chen *et al.* (2012).

O trabalho dos autores descreve como um mapa de profundidades bem construído influencia na qualidade da síntese de visões, e com isso propõe melhorias que aumentam a precisão no cálculo do mapa e o consequente aumento da qualidade das visões virtuais. O artigo também propõe um algoritmo de menor complexidade computacional em relação aos atuais, diminuindo o tempo de processamento tanto da geração do mapa quanto da síntese.

Os testes realizados pelos pesquisadores foram feitos utilizando amostras de um banco estereoscópico disponibilizado pela Universidade de Middlebury[16] (SCHARSTEIN; SZELISKI, 2002, 2003), com cada amostra formada por 9 visões. Para o teste, os autores utilizaram a 2ª e a 6ª visões para sintetizar a 4ª visão com as melhorias propostas. A visão sintetizada foi então comparada com a original para obter o valor do PSNR.

De modo a estabelecer uma base de comparação entre a RevGlyph e o trabalho de Chen *et al.* (2012), é necessário que a 4ª visão seja processada pela RevGlyph, já que ela foi o alvo de comparação dos pesquisadores. Deste modo, foram criados dois pares estereoscópicos para serem processados pela RevGlyph, um formado pela 3ª e 4ª visões e outro formado pela 4ª e 5ª visões.

[16] http://vision.middlebury.edu/stereo/. Acesso em 19 nov. 2013.

O teste com a RevGlyph fez uso das mesmas imagens e dimensões utilizadas por Chen *et al.* (2012), a saber: Cones, Teddy, Venus e Sawtooth (**Figura 28**). Cada imagem passou pela etapa de conversão anaglífica da RevGlyph, com subamostragem 4:4:0 e limiar 2. Posteriormente, foi executada a etapa de reversão anaglífica e então extraído o valor de PSNR entre o par estéreo revertido e o original.

(A) (B) (C) (D)

Figura 28 – Uma das visões de cada imagem utilizada para comparação do RevGlyph com o Mapa de Profundidades. Cones (A), Teddy (B), Venus (C) e Sawtooth (D)

Os resultados da aplicação da RevGlyph se encontram na **Tabela 5**. Como foram utilizados dois pares estereoscópicos por imagem, adicionou-se ao final do nome de cada amostra o número das visões utilizadas, de forma a diferenciar cada par estéreo. A **Tabela 5** exibe tanto o PSNR obtido em cada componente quanto a média aritmética geral. Pode-se observar que a utilização do par estéreo formado pela 3ª e 4ª visões e o formado pela 4ª e 5ª visões não influenciaram nos resultados de forma significativa, mantendo valores de PSNR muito próximos. Vale a pena também notar que os valores de PSNR obtidos estão próximos do valor obtido na **Tabela 4** da **Seção 5.3** para o limiar 2.

Tabela 5 - Resultados da aplicação da RevGlyph em amostras do banco de imagens de Middlebury

Imagem	PSNR Y (dB)	PSNR C_b (dB)	PSNR C_r (dB)	PSNR Média (dB)
cones34	33,35	36,51	39,85	36,57
cones45	33,26	36,73	39,87	36,62
teddy34	34,36	39,13	39,81	37,77
teddy45	34,22	39,08	39,84	37,71
venus34	33,97	38,50	40,55	37,67
venus45	34,11	38,37	40,57	37,68
sawtooth34	32,85	40,55	43,00	38,80
sawtooth45	32,91	40,45	43,02	38,79

O teste comparativo de qualidade objetiva entre a RevGlyph e os resultados obtidos por Chen *et al.* (2012) pode ser visto na **Tabela 6**. Como para cada imagem processada pela RevGlyph utilizou-se de dois pares estereoscópicos, o PSNR de comparação foi resultado da média aritmética dos valores de PSNR calculados para cada par estéreo. A última linha mostra a diferença absoluta entre os valores de PSNR obtidos em cada técnica.

Tabela 6 - Comparação do PSNR entre a RevGlyph e o mapa de profundidades. Valores em dB.

	PSNR Cones (dB)	PSNR Teddy (dB)	PSNR Venus (dB)	PSNR Sawtooth (dB)
RevGlyph	36,60	37,74	37,68	38,80
Chen *et al.* (2012)	28,73	30,82	32,96	33,05
Diferença	7,87	6,92	4,72	5,75

Como pode ser observado, a RevGlyph superou o mapa de profundidades e síntese de visões proposto por Chen *et al.* (2012) em todas as amostras, atigindo em média cerca de 6,32dB de diferença entre os resultados.

Realizou-se também uma estimativa em termos da comparação na taxa de compressão entre as duas técnicas. Como o trabalho de Chen *et al.* (2012) não menciona nenhuma taxa de compressão, a estimativa foi baseada no conceito de mapa de profundidades: uma imagem auxiliar em escala de cinza, com pixels variando de 0 a 255. Dessa forma, o mapa de profundidades possui um terço do tamanho da imagem RGB original.

Deve-se observar que neste caso não está sendo assumido nenhum pós-processamento envolvendo compressão sem perdas, sendo que por isso, de modo a manter a equivalência, não foi aplicada nenhuma compressão sem perdas ao final do processo de conversão anaglífica da RevGlyph. Os resultados podem ser vistos na **Tabela 7**, com a última coluna apresentando a diferença absoluta entre a taxa de compressão das duas técnicas, e a última linha apresentando a taxa de compressão média obtida em cada técnica. Nesta comparação, pode-se observar que o trabalho de Chen *et al.* (2012) apresentou taxas de compressão maiores, com uma ligeira diferença de em média 3,14% a mais de compressão atingida.

Tabela 7 - Comparação da compressão entre a RevGlyph e o mapa de profundidades

Imagem	Taxa de compressão Revglyph	Taxa de compressão Mapa de Profundidades	Diferença
Cones	27,42%	33,55%	6,13%
Teddy	30,65%	33,55%	2,90%
Venus	32,38%	33,47%	1,09%
Sawtooth	30,99%	33,45%	2,45%
Média	30,36%	33,50%	3,14%

5.5. Considerações finais

Todas as análises realizadas comprovam que a RevGlyph produz bons resultados em termos de qualidade e compressão. Provou-se com a técnica a possibilidade e viabilidade da reversão anaglífica, o que abriu caminho para o desenvolvimento de uma codificação estereoscópica capaz de ser utilizada por qualquer método de visualização.

Um teste comparativo da RevGlyph com a codificação utilizando mapa de profundidades mostrou que a técnica produz resultados equivalentes a um método de codificação estado da arte para vídeos estereoscópicos. Vale lembrar que, diferente do mapa de profundidades, a RevGlyph mantém a simplicidade em todo seu processo, não utilizando nenhum cálculo complexo que onere o processamento dos dados. Além disso, a etapa de reversão utiliza os dados de cor originais para reconstrução do par estéreo, o que possibilita, em tese, melhor qualidade visual. Desse modo, a proposta de codificação e reversão estereoscópica anaglífica apresentado pela técnica RevGlyph mostra ter vantagem competitiva na área de codificação estereoscópica, possibilitando a codificação independente de visualização e com qualidade equivalente a métodos similares.

6. Conclusões

Neste trabalho foi apresentado um panorama da área da estereoscopia em imagens e vídeos digitais, fornecendo detalhes sobre visualização estereoscópica e percepção de profundidade, bem como aspectos técnicos para codificação e compressão, que foi o foco para o desenvolvimento da técnica RevGlyph.

Mostrou-se que, apesar dos avanços obtidos na área, ainda falta um método genérico para armazenamento e transmissão de vídeos estereoscópicos, e que as pesquisas dividem-se em três métodos principais, um envolvendo o par estéreo, um envolvendo multivisão e outro envolvendo vídeo e profundidade. Essa divisão resulta em incompatibilidade entre formatos, dependendo do sistema em que o vídeo é reproduzido. Por esse motivo, surgiu a proposta de criação de um método de codificação genérico, que atendesse a todos os sistemas de reprodução estereoscópica, garantindo compressão e qualidade no resultado final.

Para a compressão dos vídeos utilizou-se da conversão anaglífica, a qual possibilita o armazenamento de apenas metade dos dados do par estéreo. De modo a manter a codificação genérica, compatível com todos os sistemas de visualização, era necessária a reversão do vídeo codificado, recuperando o par estéreo.

A reversão estereoscópica era um tópico ainda não explorado, e ao longo deste trabalho foram apresentadas as estruturas e processos que a possibilitam, o que culminou na criação da técnica denominada RevGlyph. A viabilidade e eficácia da RevGlyph perante outros métodos de codificação foram analisados e mostraram sua utilidade e vantagem competitiva.

Vale lembrar que o que foi apresentado no trabalho foi aplicado somente em imagens estereoscópicas, contendo apenas uma parte do processo completo de codificação e compressão que pode ser realizado em vídeos. Existem outras etapas da codificação que envolvem aspectos de redundância temporal e espacial que ficaram de fora do escopo do trabalho e estão abertas para novas pesquisas e aperfeiçoamentos da técnica.

6.1. Oportunidades para novas pesquisas

A maneira como a técnica RevGlyph é estruturada não vem a ser a única e definitiva para solucionar o problema da reversão anaglífica; foi apenas a primeira abordagem explorada de modo a comprovar a viabilidade da técnica. Um dos principais inconvenientes existentes é a necessidade de se alterar o processo de conversão anaglífica, de modo a disponibilizar um cabeçalho com dados adicionais para que a reversão seja possível. Desta forma, imagens que já se encontram no formato anaglífico não se beneficiam da reversão utilizando a técnica, pois não é possível obter as informações necessárias.

Como explicado na **Seção 4.2**, um anáglifo carrega consigo toda informação espacial do par estéreo, sendo que o problema é a falta de cores removidas durante a conversão. O par estéreo em sua definição é formado por duas imagens semelhantes, que diferem pelo deslocalmento lateral em que foram capturadas. Em tese, as cores dessas duas imagens são iguais, apenas deslocadas. Se de algum modo for possível inferir o valor do deslocamento para cada pixel por meio das informações espaciais contidas em um anáglifo, será possível replicar as cores existentes neste anáglifo paras as componentes faltantes e deslocá-las para sua posição correta. Essa mostra ser a linha de raciocínio mais promissora a ser seguida em estudos futuros.

Uma das apostas para encontrar o deslocamento de cada pixel entre as imagens do par estéreo é a utilização do Fluxo Óptico (D'ANGELO *et al.*, 2011), o qual calcula o movimento aparente dos objetos em uma cena. Dado um anáglifo, é possível separar as componentes de cor pertencentes a cada imagem (G_1 da direita e R_2 e B_2 da esquerda), transformá-las em duas imagens em escala de cinza, e utilizar o fluxo óptico para obter informações de movimentos de regiões de pixel. De posse dessas informações, pode-se replicar os pixels nas componentes de cor faltantes e realocá-los nas posições corretas em cada imagem. A **Figura 29** mostra o resultado de um teste inicial utilizando esta teoria.

Os estudos realizados com o Fluxo Óptico foram bem superficiais, funcionando apenas como prova de conceito. Mesmo assim, trouxeram indícios de possibilidade de reversão utilizando apenas uma imagem anaglífica, sem a necessidade de estruturas auxiliares. É deixado como sugestão o prosseguimento de estudos mais aprofundados nesta direção.

Figura 29 - Exemplo de par estéreo revertido utilizando Fluxo Óptico

A técnica RevGlyph também possui espaço para aprimoramentos de sua estrutura. A abordagem utilizada para a criação da Diferença de Luminâncias faz uso de limiares fixos para o cálculo das diferenças. Uma sugestão é o estudo de limiares dinâmicos, que, baseado no valor dos próximos pixels de uma imagem, ajuste o limiar procurando otimizar a compressão.

Como última sugestão para novas pesquisas, recomenda-se a realização de maiores estudos de análise da qualidade subjetiva de imagens revertidas utilizando-se a RevGlyph, principalmente quando estas são aplicadas a outros métodos de visualização – por polarização, temporal e autoestereoscópico, de modo a averiguar a percepção de profundidade. Para isso, é necessário investimento na aquisição de material e equipamento, bem como em um ambiente apropriado para a realização dos testes e custos envolvendo usuários para o teste.

6.2. Resultados

Toda a informação e conhecimento adquiridos foram essenciais para a atualização do grupo de pesquisa com relação à área de estereoscopia. O desenvolvimento do método para reversão anaglífica traz inovação para o campo de pesquisa e abre oportunidades para novos trabalhos. A implementação da RevGlyph contribui para o estudo e análise do método, e fica disponível para melhorias e otimizações.

Durante o programa de Mestrado, foram publicados dois artigos com resultados da técnica RevGlyph e uma coautoria para um artigo publicado em periódico relacionado a estudos de compressão espacial em vídeos estereoscópicos. Segue a seguir a listagem das publicações:

1. ZINGARELLI, M. R. U.; ANDRADE, L. A.; GOULARTE, R. Reversing Anaglyph Videos Into Stereo Pairs. In: SIMPÓSIO BRASILEIRO EM SISTEMAS MULTIMÍDIA E WEB – WEBMEDIA, 15., 2011, Florianópolis. Anais... Florianópolis, 2011.

2. ZINGARELLI, M. R. U.; ANDRADE, L. A.; GOULARTE, R. RevGlyph: a technique for reverting anaglyph stereoscopic videos. In: PROCEEDINGS OF THE ANNUAL ACM SYMPOSIUM ON APPLIED COMPUTING – SAC, 27., 2012, Riva del Garda. **Anais eletrônicos…** New York: ACM Press, 2012. Disponível em: <http://dl.acm.org/citation.cfm?doid=2245276.2245470>. Acesso em: 19 nov. 2013

3. ANDRADE, L. A. *et al.* A new approach to spatial compression of stereoscopic videos. **Multimedia Tools and Applications**, 8 dez. 2012. DOI: 10.1007/s11042-012-1300-0.

Foi também ministrado um minicurso durante a 14ª Semana da Computação da USP de São Carlos para uma turma de 50 alunos. O minicurso, intitulado "Criação de imagens e vídeos 3D com OpenCV", teve um conteúdo com uma mescla de introdução aos fundamentos da área de estereoscopia e processamento de imagens e vídeos utilizando a biblioteca OpenCV. O material do minicurso se encontra disponível em <https://db.tt/HX6ylgYP> (Acesso em 19 nov. 2013).

6.3. Opinião

Para finalizar este trabalho, registro aqui minha opinião sobre o futuro dos vídeos estereoscópicos. Percebe-se interesse da indústria e também dos consumidores em assistir conteúdo estereoscópico sem a necessidade de qualquer tipo de acessório. Desse modo, ganha destaque o método autoestereoscópico. Mais do que isso, nota-se um direcionamento para o amadurecimento de tecnologias capazes de gerar múltiplas visões de uma cena, atingindo um maior número de usuários e possibilitando observar as cenas de diferentes pontos de vista. Para acompanhar a tecnologia de criação de múltiplas visões, ganha destaque a utilização de mapas de profundidade e da técnica DIBR, modeladas para este objetivo. Porém, ainda é preciso mais estudos de modo a melhorar a qualidade das visões virtuais criadas dessa forma.

Apesar dos avanços e do interesse tanto da indústria quanto do público, o principal ponto preocupante que tem atrasado a adesão dos consumidores aos televisores 3D não é diretamente o preço, e sim (a falta de) conteúdo. Filmes 3D parecem ter se estabelecido bem nos

cinemas, com boa aceitação de público. Entretanto, o conteúdo televisivo produzido em 3D é escasso, o que influencia o consumidor a investir em aparelhos que não possuam esta tecnologia, em geral mais baratos. Enquanto as emissoras de TV não se equiparem para criar conteúdo estereoscópico, considero que a adoção dos televisores 3D não terá apelo significativo.

No caso do Brasil temos um problema adicional. Aqui as emissoras ainda estão em processo de modernização de seus equipamentos para transmissão de sinal digital ou em fase de recuperação do investimento realizado, o que posterga um investimento em equipamentos de captura e transmissão 3D, à exceção de alguns projetos pioneiros, como o da Rede TV, que iniciou a transmissão de alguns programas em 3D a partir de 2010. Além disso, o sinal analógico no Brasil está previsto para ser transmitido até 2018, sendo gradualmente desligado a partir de 2016, o que deixa uma janela de quase 5 anos até o estabelecimento do sinal digital, necessário para a transmissão 3D. Da mesma forma, o número de televisores analógicos presentes nas residências ainda é grande, apesar da tendência no mercado de aquisição de equipamentos digitais. Tudo isso contribui para que a adoção ao conteúdo estereoscópico, no lado do entretenimento, seja estendida por mais alguns anos.

Diante desse cenário, a utilização da técnica RevGlyph se mostra conveniente. O formato anaglífico é simples o suficiente para funcionar em qualquer tipo de tela e não requer equipamento adicional para o envio do sinal de vídeo. Além disso, em um cenário em que a transmissão 3D já estiver consolidada, a técnica ainda se mostra flexível, com a possibilidade da reversão anaglífica disponibilizando o par estéreo para ser codificado para diferentes tipos de visualização disponível em cada aparelho.

Referências

ANDRADE, L. A. **Compressão espacial de vídeos estereoscópicos: uma abordagem baseada em codificação anaglífica**. [s.l.] Universidade de São Paulo - USP, 2012.

ANDRADE, L. A. *et al.* A new approach to spatial compression of stereoscopic videos. **Multimedia Tools and Applications**, 8 dez. 2012.

ANDRADE, L. A.; CORDEBELLO, P. D.; GOULARTE, R. **Construção de uma Base de Vídeos Digitais Estereoscópicos,** São Carlos: 2010

ANDRADE, L. A.; GOULARTE, R. **Anaglyphic stereoscopic perception on lossy compressed digital videos** Proceedings of the XV Brazilian Symposium on Multimedia and the Web - WebMedia '09. **Anais**...New York, New York, USA: ACM Press, 2009 Disponível em: <http://portal.acm.org/citation.cfm?doid=1858477.1858506>. Acesso em: 19 nov. 2013

ANDRADE, L. A.; GOULARTE, R. **Uma Análise da Influência da Subamostragem de Crominância em Vídeos Estereoscópicos Anaglíficos** Proceedings of the XVI Brazilian Symposium on Multimedia and the Web - WebMedia '10. **Anais**...2010

AZEVEDO, E.; CONCI, A. **Computação gráfica: teoria e prática**. [s.l.] Editora Campus, 2003.

BRADSKI, G.; KAEHLER, A. **Learning OpenCV**. [s.l.] O'Reilly Media, 2008. p. 578

CHAPMAN, N.; CHAPMAN, J. **Digital Multimedia**. 3rd. ed. [s.l.] Wiley, 2004.

CHEN, K.-H. *et al.* **Choose your own viewpoint: A high-quality/low-complexity free-viewpoint 3D visual system** 2012 IEEE International Conference on Emerging Signal Processing Applications. **Anais**...IEEE, jan. 2012. Disponível em: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6152432>. Acesso em: 19 nov. 2013

COSTA, P. *et al.* Obstacle Detection using Stereo Imaging to Assist the Navigation of Visually Impaired People. **Procedia Computer Science**, v. 14, p. 83–93, jan. 2012.

D'ANGELO, E. *et al.* **Fast TV-L<inf>1</inf> optical flow for interactivity** 2011 18th IEEE International Conference on Image Processing. **Anais**...IEEE, set. 2011. Disponível em: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6115836>. Acesso em: 19 nov. 2013

DARIBO, I.; SAITO, H. A Novel Inpainting-Based Layered Depth Video for 3DTV. **IEEE Transactions on Broadcasting**, v. 57, n. 2, p. 533–541, jun. 2011.

DARIBO, I.; TILLIER, C.; PESQUET-POPESCU, B. **Adaptive wavelet coding of the depth map for stereoscopic view synthesis** 2008 IEEE 10th Workshop on Multimedia Signal Processing. **Anais**...IEEE, out. 2008. Disponível em: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4665114>. Acesso em: 19 nov. 2013

DODGSON, N. A. Autostereoscopic 3D displays. **Computer**, v. 38, n. 8, p. 31–36, ago. 2005.

EBRAHIMI, F.; CHAMIK, M.; WINKLER, S. **JPEG vs. JPEG 2000: an objective comparison of image encoding quality** Proceedings of SPIE. **Anais**...SPIE, 2004

FEHN, C. *et al.* **An Evolutionary and Optimised Approach on 3D-TV** Proceedings of Inter- national Broadcast Conference. **Anais**...2002

FEHN, C.; PASTOOR, R. S. Interactive 3-DTV-Concepts and Key Technologies. **Proceedings of the IEEE**, v. 94, n. 3, p. 524–538, mar. 2006.

FEITOSA-SANTANA, C. *et al.* Espaço de cores. **Psicologia USP**, v. 17, n. 4, p. 35–62, 2006.

GEUSEBROEK, J.-M.; BURGHOUTS, G. J.; SMEULDERS, A. W. M. The Amsterdam Library of Object Images. **International Journal of Computer Vision**, v. 61, n. 1, p. 103–112, jan. 2005.

GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing**. 3rd. ed. Upper Saddle River: Prentice-Hall, 2008.

HALLE, M. Autostereoscopic displays and computer graphics. **ACM SIGGRAPH Computer Graphics**, v. 31, n. 2, p. 58–62, 1 maio. 1997.

HALSALL, F. **Multimedia Communications: Applications, Networks, Protocols and Standards**. [s.l.] Addison-Wesley, 2000.

ITU-R. **Studio encoding parameters of digital television for standard 4 : 3 and wide-screen 16 : 9 aspect ratios**. Disponível em: <http://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.601-7-201103-I!!PDF-E.pdf>. Acesso em: 19 nov. 2013.

JIANG, L. *et al.* An overview of 3D video representation and coding. **3D Research**, v. 1, n. 1, p. 43–47, 13 ago. 2011.

JTC1/SC29/WG11, I. **ISO/IEC CD 23002-3: Representation of auxiliary video and supplemental information**. [s.l: s.n.].

KERR, D. A. **Chrominance Subsampling in Digital Images**. [s.l: s.n.]. Disponível em: <http://dougkerr.net/pumpkin/articles/Subsampling.pdf>. Acesso em: 19 nov. 2013.

KIM, D. *et al.* **An embodiment of stereo vision system for mobile robot for real-time measuring distance and object tracking** 2007 International Conference on Control, Automation and Systems. **Anais**...IEEE, 2007. Disponível em: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4407049>. Acesso em: 19 nov. 2011

LG ELECTRONICS. **LG Optimus 3D P920 - Apresentação de celular com tela 3D**. Disponível em: <http://www.lg.com/uk/mobile-phones/all-lg-phones/LG-android-mobile-phone-P920.jsp>. Acesso em: 19 nov. 2013.

LI, S. *et al.* **Stereoscopic Video Compression Based on H.264 MVC** 2009 2nd International Congress on Image and Signal Processing. **Anais**...IEEE, out. 2009. Disponível em: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5301218>. Acesso em: 19 nov. 2013.

LIPTON, L. **Foundations of the Stereoscopic Cinema: a study in depth**. New York: Van Nostrand Reinhold Company Inc., 1982.

LIPTON, L. **Stereo-vision formats for video and computer graphics** Proceedings of SPIE. **Anais**...San Jose: SPIE, 1997

LIU, S.; CHEN, C. W. **3D video transcoding for virtual views** Proceedings of the international conference on Multimedia - MM '10. **Anais**...New York, New York, USA: ACM Press, 25 out. 2010. Disponível em: <http://dl.acm.org/citation.cfm?id=1873951.1874080>. Acesso em: 19 nov. 2013.

MAITRE, M.; DO, M. N. Depth and depth–color coding using shape-adaptive wavelets. **Journal of Visual Communication and Image Representation**, v. 21, n. 5-6, p. 513–522, jul. 2010.

MENDIBURU, B. **3D Movie Making: Stereoscopic Digital Cinema from Script to Screen**. Oxford: Elsevier, 2009.

MÜLLER, K.; MERKLE, P.; WIEGAND, T. 3-D Video Representation Using Depth Maps. **Proceedings of the IEEE**, v. 99, n. 4, p. 643–656, abr. 2011.

NDJIKI-NYA, P. *et al.* **Depth image based rendering with advanced texture synthesis** 2010 IEEE International Conference on Multimedia and Expo. **Anais**...IEEE, jul. 2010. Disponível em: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5583559>. Acesso em: 19 nov. 2013.

NINTENDO OF AMERICA INC. **Nintendo 3DS. Apresentação de videogame com tela este- reoscópica**. Disponível em: <http://www.nintendo.com/3ds/hardware>. Acesso em: 19 nov. 2013.

RICHARDSON, I. E. G. **H.264 and MPEG-4 Video Compression – Video Coding for Next-generation Multimedia**. West Sussex: Wiley, 2003.

SALOMON, D. **A Concise Introduction to Data Compression (Undergraduate Topics in Computer Science)**. London: Springer, 2008.

SAYOOD, K. **Introduction to Data Compression**. 3rd. ed. San Francisco: Elsevier, 2005.

SCHARSTEIN, D.; SZELISKI, R. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. **International Journal of Computer Vision**, v. 47, n. 1-3, p. 7–42, 2002.

SCHARSTEIN, D.; SZELISKI, R. **High-accuracy stereo depth maps using structured light** 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings. **Anais**...IEEE Comput. Soc, 2003. Disponível em: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1211354>. Acesso em: 19 nov. 2013.

SCHNYDER, L. *et al.* **Depth image based compositing for stereo 3D** 2012 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON). **Anais**...IEEE, out. 2012. Disponível em: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6365451>. Acesso em: 19 nov. 2013.

SMOLIC, A. *et al.* **An overview of available and emerging 3D video formats and depth enhanced stereo as efficient generic solution** 2009 Picture Coding Symposium. **Anais**...IEEE, maio. 2009. Disponível em: <http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5167358>. Acesso em: 19 nov. 2013.

STEFANOSKI, N. *et al.* Automatic view synthesis by image-domain-warping. **IEEE transactions on image processing : a publication of the IEEE Signal Processing Society**, v. 22, n. 9, p. 3329–41, set. 2013.

STEREOGRAPHICS CORPORATION. **Stereographics® Developers' Handbook: background on creating images for CrystalEyes® and SimulEyes®**. [s.l.] Stereographics Corporation, 1997.

STRATER, L. *et al.* Me and My VE. **Proceedings of the Human Factors and Ergonomics Society Annual Meeting**, v. 56, n. 1, p. 2512–2516, 26 out. 2012.

SVOBODA, J. *et al.* Holographic 3D imaging – methods and applications. **Journal of Physics: Conference Series**, v. 415, p. 012051, 22 fev. 2013.

TAM, W.; ZHANG, L. **3D-TV Content Generation: 2D-to-3D Conversion** 2006 IEEE International Conference on Multimedia and Expo. **Anais**...IEEE, jul. 2006. Disponível em: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4036988>. Acesso em: 19 nov. 2013.

TOMKOWIAK, M. T.; LYSEL, M. S. VAN; SPEIDEL, M. A. **Monoplane stereoscopic imaging method for inverse geometry x-ray fluoroscopy** 13 mar. 2013. Disponível em: <http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2006238>. Acesso em: 19 nov. 2013.

UREY, H. *et al.* State of the Art in Stereoscopic and Autostereoscopic Displays. **Proceedings of the IEEE**, v. 99, n. 4, p. 540–555, abr. 2011.

VETRO, A. Representation and Coding Formats for Stereo and Multiview Video. *In*: [s.l: s.n.]. .

VETRO, A.; WIEGAND, T.; SULLIVAN, G. J. Overview of the Stereo and Multiview Video Coding Extensions of the H.264/MPEG-4 AVC Standard. **Proceedings of the IEEE**, v. 99, n. 4, p. 626–642, abr. 2011.

WANG, J. Z.; LI, J.; WIEDERHOLD, G. SIMPLIcity: semantics-sensitive integrated matching for picture libraries. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 23, n. 9, p. 947–963, 2001.

WHEATSTONE, C. Contributions to the Physiology of Vision.--Part the First. On Some Remarkable, and Hitherto Unobserved, Phenomena of Binocular Vision. **Philosophical Transactions of the Royal Society of London**, v. 128, p. 371–394, 1838.

WINKLER, S. **Digital Video Quality: vision model and metrics**. West Sussex: Wiley, 2005.

WOODS, A. J. **How are crosstalk and ghosting defined in the stereoscopic literature?** Displays. **Anais**...2011.

Glossário

3D

Neste trabalho, toda referência a 3D deve ser entendida com base em vídeo estereoscópico

Anáglifo ou vídeo anaglífico

Vídeo resultante da aplicação do método anaglífico a um par estereoscópico.

Anáglifo verde-magenta

Vídeo resultante da aplicação do método anaglífico, formado pelas componentes de cor vermelha e azul do vídeo do lado esquerdo do par estéreo e pela componente verde do vídeo do lado direito do par estéreo.

Cabeçalho de Consulta de Cores

Estrutura utilizada pela técnica RevGlyph para armazenar informações de crominância das componentes removidas no processo de conversão anaglífica. Tais informações podem ser consultadas para a reconstrução do par estéreo no processo de reversão.

Crosstalk

Fenômeno visual que ocorre quando há invasão de elementos visuais de uma imagem de um lado do par estéreo na imagem do outro lado, causando desconforto na visualização.

***Depth-Image-Based Rendering* (DIBR)**

Técnica de síntese de visões de vídeo, chamada de visões virtuais, a partir de um sinal de vídeo original e respectivo dados de profundidade fornecidos por um mapa de profundidades.

Diferença de Luminâncias

Estrutura utilizada pela técnica RevGlyph; contém o vetor de diferenças entre as luminâncias dos dois anáglifos criados no processo de conversão anaglífica. Auxilia na reconstrução do anáglifo complementar no processo de reversão.

Disparidade binocular

Propriedade devida a cada olho do sistema visual humano estar ligeiramente deslocado um do outro, de modo que o ambiente é observado de dois pontos de vista ligeiramente diferentes, possibilitando que pontos homólogos sejam projetados em posições diferentes em cada retina.

Espaço de cor

Modelo matemático para representação de cores.

Estereopsia

Sensação de profundidade provocada pela disparidade binocular. Embora um ser humano receba duas imagens de um objeto, ele enxerga uma fusão das duas, com sensação espacial e de profundidade.

***Head-Mounted Display* (HMD)**

Método de visualização estereoscópica realizado por meio de um aparato em formato de capacete composto por duas telas distintas, uma para cada olho, de modo que cada vídeo do par estéreo é reproduzido na tela do olho correspondente a ele.

Mapa de profundidades

Sinal de vídeo auxiliar utilizado no método envolvendo vídeo e profundidade. O conteúdo do mapa se encontra em escala de cinza, e o valor de cada pixel representa a distância do ponto representado em relação à câmera de captura.

Método anaglífico

Método de visualização estereoscópica em que o par estéreo é fundido em apenas um sinal de vídeo, por meio da mistura dos canais de cores primárias de cada vídeo. Os canais de cores não utilizados são removidos. A separação do par estéreo é feita com óculos cujas lentes filtram os canais de cor correspondentes a cada olho.

Método autoestereoscópico

Método de visualização estereoscópica em que a própria tela é capaz de separar o par estéreo, direcionando o vídeo correspondente para cada olho, sem a necessidade de óculos especiais ou qualquer outro dispositivo.

Método de Lipton

Método de organização de dados de vídeo estereoscópico em que o par estéreo é armazenado em um container de vídeo, com as imagens do par estéreo dispostas em um único quadro na forma lado-a-lado ou acima-abaixo.

Método envolvendo vídeo e profundidade

Método de organização de dados de vídeo estereoscópico em que apenas um sinal de vídeo do par estéreo é armazenado, juntamente com um mapa de profundidades associado.

Método multivisão

Método de organização de dados de vídeo estereoscópico em que mais de um par estéreo é armazenado, possibilitando ao telespectador diferentes visões de uma mesma cena.

Multiplexação de cor

Ver Método anaglífico

Multiplexação por polarização

Método de visualização estereoscópica em que o par estéreo é separado por meio da polarização das ondas de luz transmitidas por cada vídeo, de modo que cada um é reproduzido em um padrão diferente. De posse de óculos com filtros polarizadores, é possível que cada olho enxergue apenas o padrão de luz do vídeo correspondente a ele.

Multiplexação temporal

Método de visualização estereoscópica em que a imagem de cada lado do par estéreo é transmitida alternadamente em alta frequência. Óculos especiais com uma componente mecânica alternam a transparência e opacidade de cada lente na mesma frequência em que as imagens são trocadas, possibilitando que cada olho enxergue apenas a imagem correspondente a ele.

***Multiview Video Coding* (MVC)**

Técnica de codificação para vídeos estereoscópicos organizados pelo método de multivisão.

OpenCV

Biblioteca de programação *open source* que contém funções e estruturas variadas para processamento de imagens e visão computacional.

Par estéreo ou par estereoscópico

Par de imagens de uma cena, capturadas simultaneamente por duas câmeras ligeiramente deslocadas na horizontal.

Paralaxe

Dado um par estéreo, a paralaxe é a distância de um ponto na imagem direita em relação ao seu homólogo na imagem esquerda do par estéreo.

Predição entrevisões

Técnica de compressão para vídeos estereoscópicos que explora o fato de as duas imagens que formam cada par estéreo possuírem muita informação redundante. Deste modo, apenas um dos vídeos do par é codificado em quadros I, P e B, sendo o outro transformado em quadros P e B com base no primeiro vídeo e em quadros anteriores a ele.

Problema da Reversão Anaglífica

Problema definido em como recuperar as informações de cor removidas durante o processo de conversão anaglífica, e como distribuí-las de forma correta para a reconstrução do par estéreo correspondente.

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

Métrica de qualidade utilizada para análise de qualidade de imagens, envolvendo uma imagem fonte e outra correspondente que passou por algum processo de transformação. Fornece um valor de ruído presente na imagem processada, medido em decibéis (dB). Muito utilizada como base para comparação de técnicas envolvendo processamento de imagens.

RevGlyph

Técnica de codificação de vídeos estereoscópicos baseada na conversão e reversão anaglífica, possibilitando compressão sem perda da percepção de profundidade, e independência da visualização estereoscópica.

RGB

Espaço de cor baseado na tripla de cores primárias vermelho, verde e azul, muito utilizado em dispositivos de captura (câmeras fotográficas, por exemplo) e apresentação (monitores de computador, por exemplo)

Subamostragem de crominância

Técnica de compressão que diminui o volume de dados por meio da redução da crominância, explorando a característica de a visão humana ser mais sensível à informação de luminância do que de cor.

Vídeo 2D

De forma a diferenciar dos vídeos estereoscópicos, vídeo 2D é um vídeo capturado por apenas uma câmera, fornecendo apenas um ponto de vista da cena capturada.

Vídeo estereoscópico

Vídeo que ao ser visualizado proporciona sensação de profundidade na cena. Formado por um par estereoscópico.

YC_bC_r

Espaço de cor que separa informação de luminância (Y) da informação de crominância (C_b e C_r). Possibilita a aplicação da subamostragem de crominância para obter compressão.