

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Reversão anaglífica baseada em busca local rápida

Juliano Koji Yugoshi

Dissertação de Mestrado do Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional (PPG-CCMC)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Juliano Koji Yugoshi

Reversão anaglífica baseada em busca local rápida

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*

Área de Concentração: Ciências de Computação e
Matemática Computacional

Orientador: Prof. Dr. Rudinei Goularte

USP – São Carlos
Novembro de 2018

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

Y94r Yugoshi, Juliano Koji
 Reversão anaglífica baseada em busca local rápida
 / Juliano Koji Yugoshi; orientador Rudinei
 Goularte. -- São Carlos, 2018.
 77 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, 2018.

1. Vídeo anaglífico. 2. Codificação estereoscópica.
3. Visualização estereoscópica. 4. Vídeo digital. I.
Goularte, Rudinei, orient. II. Título.

Juliano Koji Yugoshi

Anaglyphic reversal based on fast local search

Master dissertation submitted to the Institute of Mathematics and Computer Sciences – ICMC-USP, in partial fulfillment of the requirements for the degree of the Master Program in Computer Science and Computational Mathematics. *FINAL VERSION*

Concentration Area: Computer Science and Computational Mathematics

Advisor: Prof. Dr. Rudinei Goularte

USP – São Carlos
November 2018

À minha família.

AGRADECIMENTOS

A Deus pela minha vida e por sempre colocar pessoas especiais na minha jornada.

Ao meu pai Koiti e minha mãe Tioka pelo amor incondicional.

À minha esposa Ivone pelo amor, cumplicidade e por estar sempre comigo.

À minha filha Natália pelo amor e carinho.

Ao meu orientador Prof. Dr. Rudinei Goularte pela orientação neste mestrado, por tudo que aprendi tecnicamente, pela confiança, pelo exemplo de pessoa e por incentivar a superar diversos desafios.

A minha família, sempre presente em minha vida. Meus irmãos Luciana, Marcelo e Eliane e cunhado Fábio. Meus sogros Yoshio e Ordylette. Minhas cunhadas Elisa e Isabella e os maridos delas Dácio e Renato. Meus sobrinhos Beatriz, Dante, Mateus e Marina que trazem amor, esperança e alegria para nossas vidas.

Aos meus amigos de grupo de pesquisa Felipe Maciel, Johana M. R. Villena, Marcio Funes, Rodrigo Mitsuo Kishi, Tamires T. de S. Barbieri e Tiago Trojahn.

Aos amigos do Laboratório Intermídia e LABIC.

Aos professores de Sistemas de Informação da UFMS.

Aos professores do ICMC-USP.

Aos funcionários do ICMC-USP.

À UFMS pelo afastamento para capacitação concedido.

RESUMO

YUGOSHI, J. K. **Reversão anaglífica baseada em busca local rápida.** 2018. 77 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2018.

A proliferação de conteúdos estereoscópicos atualmente é uma realidade, devido, principalmente, ao interesse e a percepção de valor do público, em geral, como uma tecnologia amigável. Os diversos benefícios trazidos por essa tecnologia, do entretenimento à pesquisa, influenciaram no desenvolvimento de inúmeras técnicas de captação, codificação e reprodução desses vídeos. Tendo em vista a integração com a infraestrutura atual, novas técnicas continuam surgindo e trazendo novas descobertas. No entanto, no campo da codificação, existe um problema que envolve a dificuldade para reproduzir um vídeo sem se conhecer a técnica de codificação que o gerou. Um ponto comum das formas de reprodução é que todas tomam como base um par estéreo, o que, por um lado pode, genericamente, permitir a codificação para operar em modos de reprodução diferentes, mas, por outro lado, traz outro problema, o de duplicar o volume de dados demandados, tornando-o de alto custo para armazenamento e transmissão. Assim, nesta dissertação foi desenvolvida uma nova técnica para reverter um anáglio a uma aproximação do par estéreo original baseada em busca local rápida, utilizando apenas nas informações intracodificadas do vídeo anáglio. A utilização anáglio e da técnica de reversão, reduz o volume dos dados e torna genérico o conteúdo para reprodução. Para mensurar os resultados, foram realizados experimentos submetidos a análise objetiva utilizando o PSNR (*Peak Signal to Noise Ratio*) e a análise subjetiva com o método DSCQS (*Double Stimulus Continuous Quality Scale*). Como resultados foram recuperados aproximações dos pares estéreos originais independentes do modo de visualização com boa qualidade visual e boa percepção de profundidade.

Palavras-chave: Vídeo Anaglífico, Visualização Estereoscópica, Codificação estereoscópica.

ABSTRACT

YUGOSHI, J. K. **Anaglyphic reversal based on fast local search.** 2018. 77 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2018.

The proliferation of stereoscopic content is currently a reality, mainly due to the public interest and perception of value, in general, as a friendly technology. The diverse benefits brought by this technology, from entertainment to research, have influenced the development of numerous techniques for capturing, coding and reproducing these videos. In view of the integration with the current infrastructure, new techniques continue to emerge and bring new discoveries. However, in the field of coding, there is a problem that involves the difficulty of reproducing a video without knowing the coding technique that generated it. A common point of the forms of reproduction is that all are based on a stereo pair, which, on the one hand, can generically allow coding to operate in different reproduction modes, but on the other hand, it brings another problem, that of duplicate the volume of data demanded, making it costly for storage and transmission. Thus, in this dissertation a new technique was developed to revert an anaglyph to an approximation of the original stereo pair based on fast local search, using only the intracoded information of the anaglyph video. The anaglyph use and reversal technique reduces the volume of data and makes the content for reproduction generic. To measure the results, experiments were performed under objective analysis using PSNR (Peak Signal to Noise Ratio) and subjective analysis with the DSCQS (Double Stimulus Continuous Quality Scale) method. As a result, approximations of the original stereo pairs independent of the viewing mode with good visual quality and good depth perception were retrieved.

Keywords: Anaglyphic Video, Stereoscopic View, Stereoscopic Coding.

LISTA DE ILUSTRAÇÕES

Figura 1 – Disparidade binocular	24
Figura 2 – Par estéreo - lado esquerdo e direito	25
Figura 3 – Classificação das dicas de profundidade	25
Figura 4 – Paralaxe de movimento	27
Figura 5 – Dicas de monoculares	28
Figura 6 – Convergência	28
Figura 7 – Primeiro dispositivo estereoscópico. O Estereoscópio de Wheatstone.	29
Figura 8 – Classificação dos métodos de visualização estereoscópicos	29
Figura 9 – Anáglifo	30
Figura 10 – Visualização anaglífica	31
Figura 11 – Codificação anaglífica	31
Figura 12 – Polarização da luz - vertical e horizontal	32
Figura 13 – Visualização por luz polarizada	32
Figura 14 – Multiplexação temporal com óculos obturadores	33
Figura 15 – Visualização utilizando HMD (<i>Head Mounted Display</i>)	34
Figura 16 – Barra de Paralaxe (A) Vs Película lenticular(B)	35
Figura 17 – Princípios de monitores multivisão autostereoscópicos	36
Figura 18 – Esquema de avaliação DSCQS	40
Figura 19 – Escala de avaliação utilizada no DSCQS	40
Figura 20 – MOS	41
Figura 21 – RevGlyph	43
Figura 22 – Visão geral da técnica Recovering Stereo Pairs from Anaglyphs	44
Figura 23 – Técnica proposta por Williem, Raskar e Park (2015)	46
Figura 24 – HaaRGlyph	47
Figura 25 – Codificação anaglífica	49
Figura 26 – Canais de cores preservados	50
Figura 27 – Transferência de cores	50
Figura 28 – Visão geral da abordagem proposta	51
Figura 29 – Separação de canais de cores	52
Figura 30 – Atribuição de coordenadas e cores	53
Figura 31 – Imagens binárias geradas pelo detector de bordas de Canny	54
Figura 32 – Busca de correspondência entre blocos e transferência de cor	55
Figura 33 – Agrupamento dos canais e par estéreo gerado	56

Figura 34 – Busca de correspondência entre macroblocos	56
Figura 35 – Busca de correspondência entre macroblocos sem sobreposição	57
Figura 36 – Busca de correspondência no canal fonte	58
Figura 37 – Resultados de DSCQS e MOS - Fase 1	65
Figura 38 – Média MOS - Fase 1	66
Figura 39 – Resultados de DSCQS e MOS - Fase 2	66

LISTA DE TABELAS

Tabela 1 – Relação dos trabalhos com os <i>datasets</i>	62
Tabela 2 – Descrição de imagens por dataset	62
Tabela 3 – Resultados de PSNR para cada técnica em cada Dataset	63

SUMÁRIO

1	INTRODUÇÃO	19
1.1	Contextualização e Motivação	19
1.2	Organização da dissertação	22
2	FUNDAMENTOS E TRABALHOS RELACIONADOS	23
2.1	Considerações iniciais	23
2.2	Visão estéreo e percepção de profundidade	23
2.2.1	<i>Dicas monoculares</i>	26
2.2.2	<i>Dicas binoculares</i>	28
2.3	Métodos de visualização estereoscópicos	29
2.3.1	<i>Métodos por multiplexação</i>	30
2.3.1.1	<i>Multiplexação de cor - método anaglífico</i>	30
2.3.1.2	<i>Multiplexação por luz polarizada</i>	31
2.3.1.3	<i>Multiplexação temporal</i>	32
2.3.2	<i>Métodos com dispositivos HMD</i>	33
2.3.3	<i>Métodos autoestereoscópicos</i>	34
2.3.3.1	<i>Duas visões</i>	34
2.3.3.2	<i>Multivisões</i>	35
2.4	Codificação estereoscópica	36
2.4.1	<i>Codificação convencional</i>	37
2.4.2	<i>Codificação em múltiplas visões</i>	37
2.4.3	<i>Codificação baseada em vídeo e profundidade</i>	38
2.5	Qualidade de vídeo	38
2.5.1	<i>Medida de qualidade subjetiva</i>	39
2.5.2	<i>Medida de qualidade objetiva</i>	41
2.6	Trabalhos relacionados	42
2.6.1	<i>RevGlyph</i>	42
2.6.2	<i>Recovering Stereo Pairs from Anaglyphs</i>	44
2.6.3	<i>Depth Map Estimation and Colorization of Anaglyph Images Using Local Color Prior and Reverse Intensity Distribution</i>	45
2.6.4	<i>HaaRGlyph: A New Method for Anaglyphic Reversion in Stereoscopic Videos</i>	46
2.7	Considerações finais	48

3	SOLUÇÃO PROPOSTA	49
3.1	Considerações Iniciais	49
3.2	Leitura de imagem de entrada	52
3.3	Processamento de Imagem	52
3.4	Correspondência entre blocos com transferência de cor	54
3.5	Reconstrução do par estéreo	55
3.6	Análise de custo da técnica proposta	56
3.7	Considerações finais	59
4	AVALIAÇÃO EXPERIMENTAL	61
4.1	Considerações Iniciais	61
4.2	<i>Dataset</i> de imagens estéreo	61
4.3	Avaliação objetiva de imagem	62
4.4	Avaliação subjetiva de imagem	64
4.5	Considerações Finais	67
5	CONCLUSÕES	69
5.1	Contribuições Científicas	70
5.2	Limitações e trabalhos futuros	70
	REFERÊNCIAS	73



INTRODUÇÃO

1.1 Contextualização e Motivação

Atualmente os vídeos são uma realidade no cotidiano das pessoas, em parte, pela interatividade social que ele possibilita. Por outro lado, a popularização dos *smartphones*, *tablets* e de outros aparelhos eletrônicos sem fio tornou a edição e o compartilhamento de vídeos uma tarefa rápida e simples.

Estima-se que até 2021 os conteúdos de vídeos representem 81% do total do tráfego de Internet no mundo, acima dos 73% registrados em 2016 ([CISCO, 2017](#)). Além disso, segundo [Harstead e Sharpe \(2015\)](#), desse total de vídeos, de 4% a 6% serão de vídeos 3D. Os conteúdos 3D são utilizados em diversas aplicações que vão do entretenimento à medicina. Isso impulsiona a indústria no desenvolvimento de conteúdos para adequar-se às necessidades das inúmeras aplicações que exploram a característica do 3D, proporcionada pela percepção de profundidade ([KAUFMAN, 1974](#)), que atrai o público consumidor.

O termo 3D, tratando-se de vídeo, representa um conjunto de vídeos que tem por finalidade propiciar a percepção de profundidade para um observador, mimetizando a visão estéreo humana ([AZEVEDO; CONCI, 2003](#)). Um tipo de vídeo desse conjunto é o Estereoscópico, que é composto por um par de vídeos da mesma cena, ligeiramente diferentes, denominado par estéreo ([LIPTON, 1997](#)). Devido ao interesse público, houve um aumento de produção desse conteúdo inicialmente utilizado na indústria cinematográfica e, atualmente, já incorporado em diversos aparelhos eletrônicos, como televisão 3D, *smartphones*, *tablets* e consoles de jogos. Cada um com suporte para diferentes tipos de conteúdos estereoscópicos, consequentemente, novas técnicas de Produção, Codificação e Visualização estão surgindo ou sendo aperfeiçoadas.

No campo da Produção, câmeras com duas lentes foram desenvolvidas para a gravação de duas visões da mesma cena (esquerda e direita) e para a geração do par estéreo de vídeo, com a possibilidade de também gerar um mapa de profundidade para a mesma cena ([FEHN](#)

et al., 2002). Há também técnicas desenvolvidas para a conversão de conteúdos planares em estereoscópicos (TAM; ZHANG, 2006).

No domínio da Visualização, as técnicas que utilizam óculos específicos, funcionando como filtros para separar o par estéreo, com uma imagem para cada olho, são conhecidas como técnicas de multiplexação e podem ser: por cor, por luz polarizada e por tempo (LIPTON, ; LIPTON, 1997). Também está incluso nesse domínio, o dispositivo *Head Mounted Display* (HMD), que é vestível, com uma tela para cada olho, possibilitando a imersão total em ambientes de realidade virtual. Por fim existe a técnica de visualização autoestereoscópica, que permite visualizar conteúdos estereoscópicos sem nenhum tipo de óculos ou dispositivos específicos (SMOLIC *et al.*, 2009).

Os avanços no campo da Produção e Visualização são perceptíveis, porém na Codificação são mais lentos. A Codificação é identificada por dois grupos: (i) o Método de Lipton (LIPTON, 1997) e (ii) os Métodos Vinculados (SMOLIC *et al.*, 2009). No primeiro, o par estéreo é armazenado em um único vídeo recipiente, com ou sem compressão, com o dobro de dados em relação a um vídeo planar (2D). Uma vez que esse método mantém o par estéreo, ele pode ser utilizado por qualquer técnica de visualização. No segundo, tem-se os métodos do vídeo planar aplicados, diretamente, em vídeo estereoscópico, utilizando técnicas clássicas de compressão, como o MPEG-2 e o H.264 (ITUT, 2003), para reduzir a quantidade de dados armazenados e transmitidos. Além disso, cada Método Vinculado é desenvolvido para um sistema de visualização particular, isto é, não são compatíveis com todas as técnicas de visualização (SMOLIC *et al.*, 2009). Ainda, as técnicas de compressão baseadas em vídeos planares alcançam altas taxas de compressão ao serem aplicadas no par estéreo de vídeo, mas ficam com volume de dados expressivo se comparado a um vídeo planar. Um fato relevante é que ao utilizar a compressão com perdas para alcançar altas taxas de compressão, a percepção de profundidade, em alguns casos, pode ser comprometida, especialmente, com vídeos anaglíficos (ANDRADE; GOULARTE, 2009; ANDRADE; GOULARTE, 2010).

Nesse cenário, percebe-se algumas lacunas na literatura relacionadas a (i) codificação do vídeo estereoscópico, de maneira a obter alta taxa de compressão sem a perda significativa da percepção de profundidade, (ii) geração de um conteúdo genérico a qualquer técnica de visualização. Nota-se, também como um desafio, (iii) a grande quantidade de vídeos anáglifos legados na Internet, isto é, que não podem ser visualizados em outro modo que não seja a anaglífica, pois, não possuem mais os respectivos pares estéreos.

Para preencher essas lacunas, nesta dissertação foi utilizado imagens anaglíficas. A codificação anaglífica é um método simples que não requer equipamento caro ou complexo para produção e é computacionalmente barata. As imagens ou vídeos anaglíficos possuem grande parte das informações planares e de cores do par estéreo original com a metade do volume de dados.

No processo codificação anaglífica, apresentado na Seção 2.3.1.1, um par estéreo é

codificado gerando um único vídeo anáglifo, isso, ameniza o problema de volume dos dados para armazenamento e transmissão. No entanto, traz outro problema, o de não poder ser utilizado em outras técnicas de visualização estereoscópica que não seja a anaglífica.

O processo de Reversão Anaglífica ([ZINGARELLI; ANDRADE; GOULARTE, 2011](#)), processo inverso da codificação, não é trivial e os poucos trabalhos relacionados a esse tema, apresentados na Seção [2.6](#), propõem soluções particulares ou computacionalmente caras. A solução particular, nesse caso, é aquela que não pode ser aplicada diretamente em qualquer vídeo anaglífico, pois necessita de informações externas ao vídeo em questão. O custo computacional elevado, em geral, é causado pela utilização de técnicas auxiliares complexas na tentativa encontrar correspondências entre as partes remanescentes do par estéreo presentes no anáglifo. Um problema similar ao problema conhecido como Correspondência Estéreo ([MARR,](#)).

Nesse contexto, surge a questão de pesquisa: é possível recuperar uma aproximação do par estéreo original apenas com as informações intracodificadas no anáglifo e sem perdas significativas na qualidade da percepção de profundidade? O objetivo deste trabalho é responder a essa questão de pesquisa e, por consequência, reduzir as lacunas identificadas.

Assim, nesta dissertação é proposta uma nova técnica de **Reversão Anaglífica Baseada em Busca Local Rápida** (*Anaglyphic Reversal Based on Fast Local Search - ARBFLS*). A ARBFLS é inspirada no algoritmo de correspondência entre blocos (*Block Matching Algorithm - BMA*) ([HANNAH, 1974](#)) para recuperar uma aproximação do par estéreo original. O BMA para estimativa de movimento (*Motion Estimation - ME*) é amplamente adotado pelos padrões de codificação de vídeo (H.261, H.263, MPEG-1, MPEG-2, MPEG-4 e H.264) devido à sua eficácia e simplicidade para implementação.

A motivação para o uso do BMA vem de uma observação interessante. Parte da informação necessária para reconstruir a visão esquerda do par estéreo está presente na visão direita e vice-versa, porém, minimamente deslocadas espacialmente mimetizando movimento. Como um anáglifo possui informações de ambas visões, é provável que se possa calcular as correspondências usando BMA e usá-las para recuperar as informações faltantes de uma visão com informações da outra visão. O BMA mais simples executa a busca global (*Full Search - FS*) ([BHASKARAN; KONSTANTINIDES, 1997](#)) para encontrar correspondências entre blocos e isso é computacionalmente custoso. A ARBFLS para diminuir o custo computacional utiliza uma técnica de busca local baseada na Soma do valor Absoluto das Diferenças de intensidade entre pixels (*Sum of Absolute Differences - SAD*), [Brown, Burschka e Hager \(2003\)](#) afirmam que técnicas de buscas locais são mais eficientes que as técnicas globais. Essa pesquisa explora as vantagens do vídeo anaglífico, do BMA, da busca local e do par estéreo. O vídeo anaglífico possui informações do par estéreo original com a metade do volume de dados, a correspondência entre blocos encontra elementos correspondentes nas informações originais preservadas no anáglifo, a busca local torna a técnica menos custosa computacionalmente e todos os métodos de visualização estereoscópicos tomam por base o par estéreo.

A avaliação da técnica proposta foi realizada utilizando medidas objetivas e subjetivas bem conhecidas da área - *Peak Signal-to-Noise Ratio* (PSNR) e *Double-Stimulus Continuos Quality-Scale* (DSCQS) ([RECOMMENDATION, 1999](#); [RECOMMENDATION, 2002](#)). Os resultados das avaliações experimentais foram comparados com as técnicas relacionadas na literatura e demonstraram que a proposta é promissora, como apresentado no Capítulo 4.

1.2 Organização da dissertação

O restante desta dissertação está organizado do seguinte modo:

Capítulo 2 – Fundamentos e Trabalhos Relacionados. Nesse capítulo são apresentados os conceitos e terminologias utilizados nesta dissertação e trabalhos relacionados necessários para desenvolvimento dessa dissertação.

Capítulo 3 – Solução Proposta. Nesse capítulo é apresentada a técnica denominada ARBFLS proposta nesta dissertação. A técnica proposta foi dividida em 5 etapas: (i) tratamento de imagem de entrada, (ii) processamento de imagem, (iii) correspondência entre blocos, (iv) transferência de cor e (v) aproximação do par estéreo original.

Capítulo 4 – Avaliação Experimental. Nesse capítulo são apresentados os resultados dos experimentos com os datasets utilizados pelos trabalhos relacionados. Na sequência, é apresentada e discutida a comparação dos resultados obtidos com os trabalhos relacionados na literatura.

Capítulo 5 – Conclusões. Nesse capítulo são apresentadas as conclusões desta dissertação. Os objetivos e a questão de pesquisa são retomados e discutidos. Por fim, são descritas as limitações deste trabalho e direções para trabalhos futuros.



FUNDAMENTOS E TRABALHOS RELACIONADOS

2.1 Considerações iniciais

Existe uma variedade de conceitos fundamentais para explicar o fenômeno da percepção de profundidade, sendo que os principais são apresentados na Seção 2.2. Alguns desses conceitos também são explorados para maximizar a percepção humana na visualização de profundidade em monitores ou telas, conforme apresentados na Seção 2.3. Em geral, para que vídeos estereoscópicos possam ser reproduzidos/visualizados digitalmente são utilizados alguns métodos de codificação que são apresentados com mais detalhes na Seção 2.4. Na Seção 2.5 são apresentadas as medidas de qualidade subjetiva e objetiva para avaliar o par estéreo recuperado. Por fim, na seção 2.6 são apresentados os principais trabalhos relacionados ao objetivo dessa dissertação na literatura atual.

2.2 Visão estéreo e percepção de profundidade

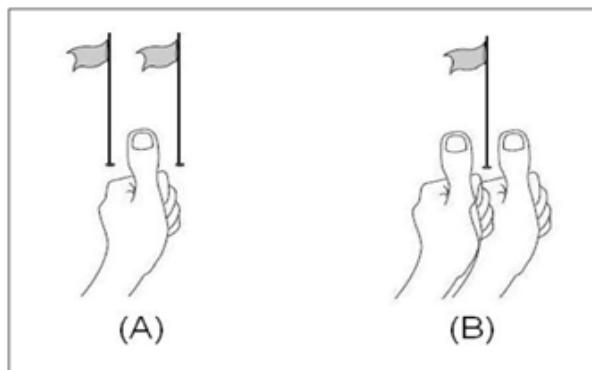
A sensação de profundidade é a capacidade visual para perceber as distâncias entre objetos. Essa habilidade da percepção humana da tridimensionalidade é baseada em imagens planares (2D), projetadas nas retinas (GOLDSTEIN, 2014).

Os olhos humanos estão a aproximadamente 6,5 cm de distância um do outro (humano adulto), movem-se em conjunto na mesma direção e cada olho tem um ângulo de visão limitado. Desde que os nossos olhos estão horizontalmente separados, cada olho tem a sua própria perspectiva do mundo (AZEVEDO; CONCI, 2003). Portanto, o mesmo ponto de uma cena é projetada em posições ligeiramente diferentes na retina de cada olho; e cada olho recebe uma imagem ligeiramente diferente da mesma cena. A distância entre os pontos correspondentes nas imagens projetadas na retina de cada olho é conhecida como a disparidade de retina ou

disparidade binocular.

Com base na disparidade de retina, o cérebro funde as imagens esquerda e direita recebidos por cada um dos olhos, trazendo a sensação de profundidade ao observador. Esse fenômeno é conhecido como estereopsia ([WHEATSTONE, 1838](#)) e é estudado pela área de estereoscopia. A disparidade de retina pode ser melhor compreendida na Figura 1, em que o polegar é colocado entre os olhos e o objeto. Ao concentrar o foco da visão no dedo polegar, ou seja, criando um ponto de convergência das duas retinas no polegar, o objeto que fica atrás desse ponto de convergência, aparece como duplicado (Figura 1 (A)). Isso acontece porque as imagens fora do ponto focal estão sendo formadas em locais diferentes em cada retina. O mesmo acontece ao concentrar o foco no objeto ao fundo, o ponto de convergência fica depois do polegar, duplicando-os (Figura 1 (B)). A distância entre essas imagens duplicadas é a disparidade de retina e o fenômeno ilustrado na Figura 1 (B) é explorado em filmes estereoscópicos para dar a impressão de que os objetos estão “saltando para fora da tela”([LIPTON, 1997](#)).

Figura 1 – Disparidade binocular



Fonte: Adaptada de [Lipton \(1997\)](#).

A paralaxe é o conceito de disparidade de retina aplicado em um monitor ou tela. Com a paralaxe é possível dar um ponto de vista diferente da mesma imagem para cada olho, resultando na formação da disparidade e na visão estéreo. Uma maneira fácil de calcular a paralaxe entre dois pontos é sobrepor uma imagem a outra e medir a distância entre os mesmos pontos nas imagens. É por causa da paralaxe que ao visualizar um vídeo anaglífico sem os óculos adequados, vê-se partes das imagens como duplicadas e sobrepostas. Deste modo, a estereopsia é baseada em métodos que apresentam a um observador um par de imagens planas (2D) do mesmo objeto, denominado par estéreo, conforme a Figura 2, cada uma proporcionando a visão de cada olho com uma perspectiva ligeiramente diferente.

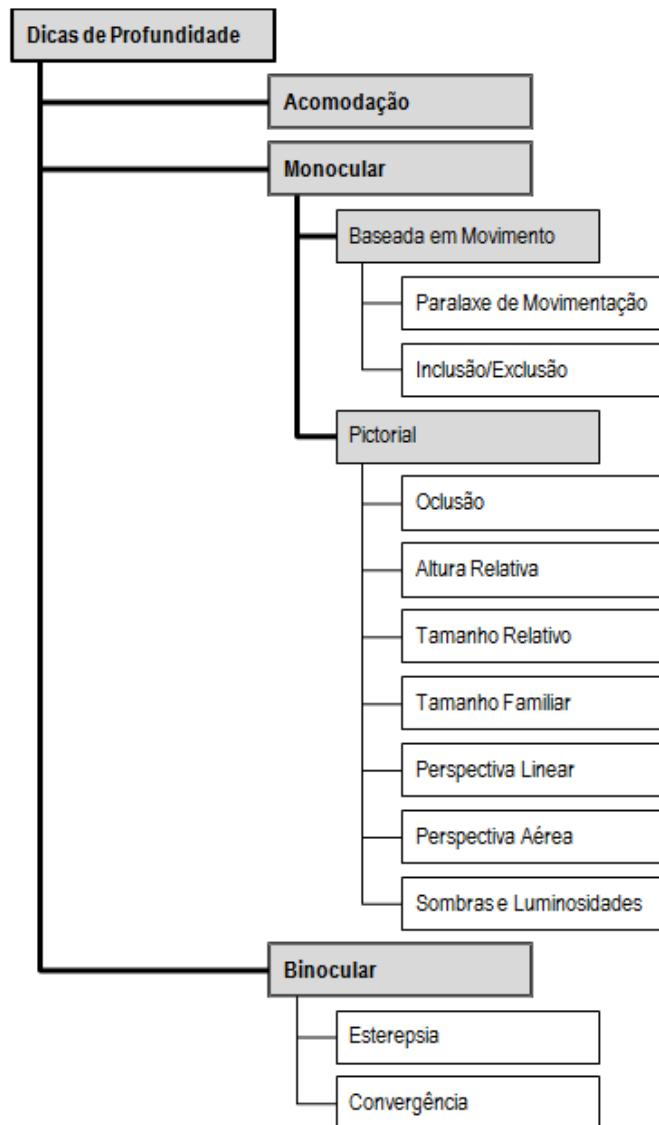
Na Figura 2, é possível observar elementos de imagem (pessoas ou objetos) que foram marcados nas duas imagens. Alguns elementos que estão presentes na imagem do olho esquerdo não aparecem ou estão obstruídas na imagem do olho direito e o contrário também pode ser observado.

Figura 2 – Par estéreo - lado esquerdo e direito



Fonte: Adaptada de Zingarelli, Andrade e Goularte (2012).

Figura 3 – Classificação das dicas de profundidade



Fonte: Adaptada de Goldstein (2014).

Durante anos, pesquisadores identificaram as diferentes dicas de profundidade (*depth cues*) utilizadas para obter informações de profundidade. Uma classificação frequentemente utilizada para dicas de profundidade é dispô-las como dicas monoculares ou binoculares (GOLDSSTEIN, 2014), conforme a Figura 3.

Outra forma comum é agrupá-las como fisiológicas (relativo às funções mecânicas, físicas e bioquímicas do organismo) ou psicológicas (relativo a interpretação mental do que está sendo observado). Essas classificações para as dicas de profundidade não são independentes, por exemplo, a perspectiva linear (apresentada na Subseção 2.2.1) é psicológico e monocular, por outro lado, a disparidade e convergência são fisiológicas e binoculares (apresentadas na Subseção 2.2.2).

2.2.1 ***Dicas monoculares***

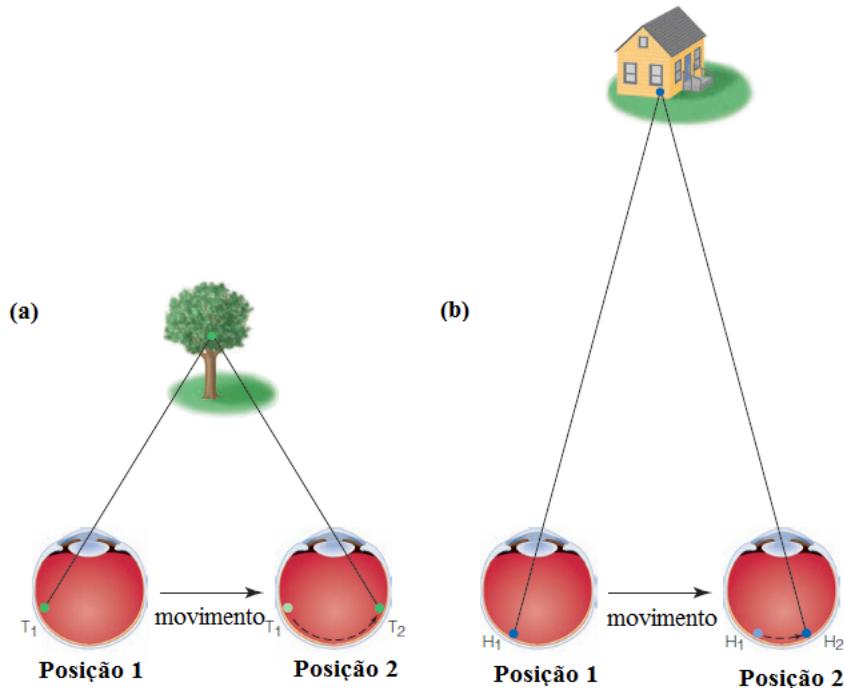
As dicas monoculares são aquelas que podem ser percebidas com apenas um olho. Elas podem ser classificadas como: (i) acomodação; (ii) baseada em movimento; e (iii) pictorial. A acomodação é uma dica de profundidade fisiológica, enquanto que as duas outras são psicológicas.

Acomodação é uma dica que está relacionada com a mudança física (muscular) do foco dos olhos, que ocorre ao focar objetos em diferentes distâncias. Para um objeto próximo, a lente do olho fica mais curva. Os músculos em torno de nossos olhos comprimem-se para alcançar este efeito. O subconsciente capta essa informação e o cérebro interpreta a visão.

As dicas baseadas em movimento (*motion-based*) são aquelas que aferem informação de profundidade criada por um movimento relativo entre o observador e o objeto observado. Dicas baseadas em movimento podem ser: paralaxe de movimento (*motion parallax*), de exclusão (*deletion*) e de inclusão (*accretion*).

A paralaxe de movimento ocorre quando a imagem dentro da retina se movimenta quando o olho muda da posição 1 para a posição 2, conforme Figura 4. A imagem do objeto pode estar próxima ou distante do observador. Para a imagem próxima, a árvore da Figura 4 (a), é possível observar que a imagem movimenta-se por toda a retina de T1 para T2 (seta tracejada) quando o olho muda da posição 1 para a posição 2. Já na imagem distante do observador, representada pela casa na Figura 4 (b), a imagem movimenta-se a uma distância menor, de H1 para H2 (seta tracejada). Assim, como a imagem do objeto próximo percorre uma grande distância na retina à medida que o observador entra em movimento, passa a impressão de que ela movimenta-se rapidamente. A imagem do objeto distante, por percorrer uma distância muito menor na retina, parece mover-se mais lentamente.

Figura 4 – Paralaxe de movimento



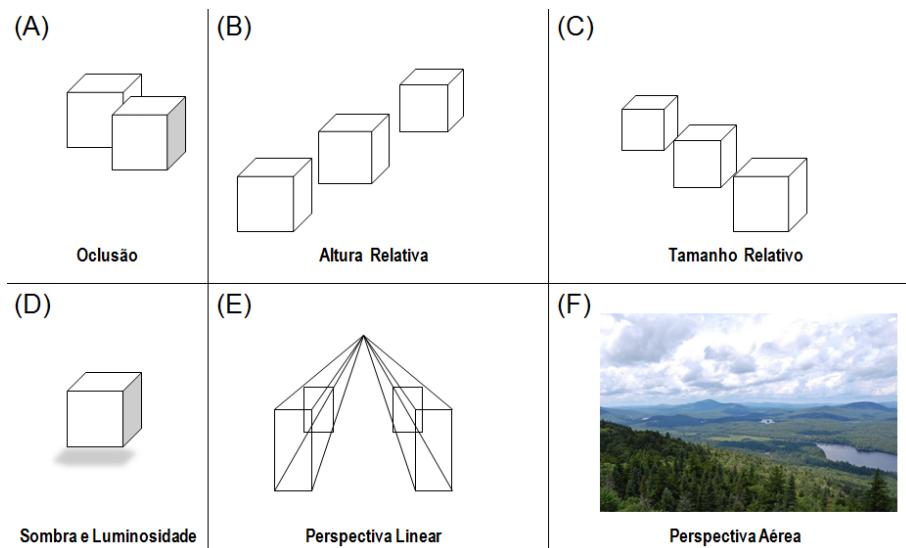
Fonte: Adaptada de [Goldstein \(2014\)](#).

As dicas pictoriais são fontes de informações de profundidade que podem ser representadas em uma figura, como as ilustrações de um livro ou a imagem na retina. Na Figura 5 as dicas pictoriais de profundidade estão representadas por imagens planas 2D. A oclusão também conhecida como interposição, ocorre quando um objeto em primeiro plano bloqueia parcialmente a visualização de objetos que estão ao fundo. Os objetos parcialmente obstruídos são percebidos como mais distantes (Figura 5 (A)). A altura relativa está relacionada com os objetos que estão localizados abaixo e próximos da linha do horizonte. Geralmente, são percebidos como mais distantes (Figura 5 (B)). O tamanho relativo está relacionada ao fato de que objetos próximos cobrem um ângulo visual maior na nossa retina. Esses objetos parecem maiores do que objetos mais distantes. Assim, se dois objetos são de tamanhos iguais, o mais próximo parecerá maior (Figura 5 (C)).

As sombras e a luminosidade estão relacionadas à forma como um objeto lança sombras e como a luz incide sobre a superfície do mesmo. Esta dica também pode ser utilizada para obter informações de profundidade (Figura 5 (D)). A perspectiva linear são linhas paralelas que se fundem em um único ponto de fuga no horizonte. Objetos mais perto do ponto de fuga são percebidos como mais distantes (Figura 5 (E)). Com a dica de perspectiva aérea com esta dica observa-se que os objetos mais distantes possuem menos contraste e saturação. O espectro de cores é também deslocado para a cor azul devido à dispersão da luz na atmosfera (Figura 5 (F)). E por fim a dica do tamanho familiar relacionada ao fato de já ser conhecido alguns tamanhos de objetos, por exemplo, árvores, carros e pessoas. Esses tamanhos relativos dos objetos familiares

são comparados na mesma cena para determinar distância.

Figura 5 – Dicas de monoculares



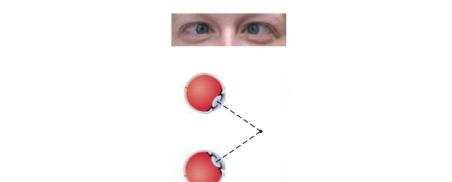
Fonte: Adaptada de [Lebreton et al. \(2014\)](#).

2.2.2 Dicas binoculares

Dicas binoculares dependem de ambos os olhos trabalharem juntos. Duas dicas binoculares importantes são a disparidade binocular e a convergência. A disparidade é uma dica de profundidade psicológica, enquanto que a convergência é fisiológica.

A disparidade binocular, como explicado anteriormente e ilustrado na Figura 1, é a distância entre os pontos correspondentes nas imagens projetadas na retina de cada olho. Já a convergência é uma dica de profundidade que ocorre quando se pretende visualizar objetos próximos. Os olhos movem-se para dentro, convergindo, conforme a Figura 6. O sistema visual humano usa a acomodação (dica de profundidade monocular) combinada com a convergência para corrigir o poder de refração e para assegurar uma imagem clara do objeto visualizado ([GOTCHEV et al., 2011](#)).

Figura 6 – Convergência

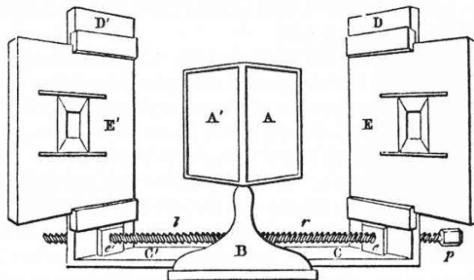


Fonte: Adaptada de [Goldstein \(2014\)](#).

2.3 Métodos de visualização estereoscópicos

Um método de visualização estereoscópica é, talvez, o componente mais importante de um sistema de mídia estéreo. Esses métodos afetam diretamente a qualidade da experiência percebida, fornecendo as dicas de profundidade para o sistema visual humano. Além disso, as tecnologias de reprodução de mídia estéreo, também são afetadas por esses métodos de visualização.

Figura 7 – Primeiro dispositivo estereoscópico. O Estereoscópio de Wheatstone.

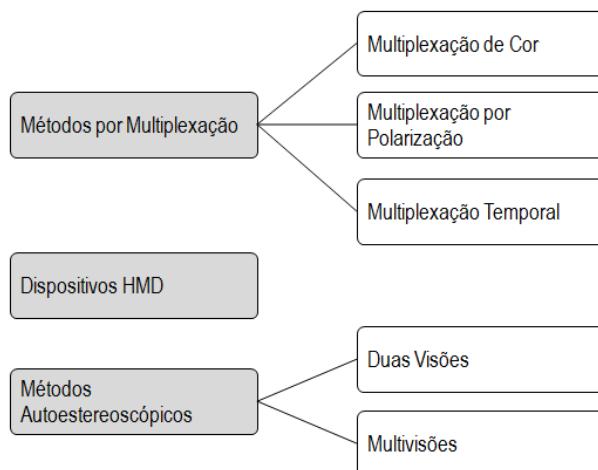


Fonte: [Wheatstone \(1838\)](#).

O primeiro dispositivo estereoscópico foi proposto por C. Wheatstone em 1830 ([WHEATSTONE, 1838](#)), conforme ilustrado na Figura 7. As tecnologias estereoscópicas têm sido usadas em TVs e filmes desde 1940 ([KIM et al., 2014](#)) e ainda são utilizadas no cinema 3D e na TV 3D. Já os métodos mais modernos, os autoestereoscópicos, fornecem imagens diferentes para cada olho sem a utilização de nenhum filtro ou dispositivo auxiliar.

Os métodos de visualização estereoscópicos podem ser divididos em três categorias: (i) métodos por multiplexação (Seção 2.3.1); (ii) métodos com dispositivos HMD (Seção 2.3.2); e (iii) métodos autoestereoscópicos (Seção 2.3.3). Essa classificação é apresentada na Figura 8.

Figura 8 – Classificação dos métodos de visualização estereoscópicos



Fonte: Adaptada de [Urey et al. \(2011\)](#).

2.3.1 Métodos por multiplexação

Os métodos por multiplexação, tradicionalmente usam óculos especiais que funcionam como filtros que induzem a disparidade binocular e de convergência, fornecendo imagens ligeiramente diferentes para cada olho (esquerdo e direito) ([HONG et al., 2011](#)). Esses métodos são os mais utilizados no mercado devido à tecnologia conhecida e ao custo baixo envolvido quando comparado aos métodos HMD e autoestereoscópicos. Nas próximas seções, são apresentadas as multiplexações de cor, por luz polarizada e temporal.

2.3.1.1 Multiplexação de cor - método anaglífico

O método anaglífico, o processo de codificação descarta alguns canais de cor de cada lado do par estéreo de vídeos. Em seguida, os canais de cores restantes do par de vídeos são fundidos gerando um único vídeo anáglifo, conforme Figura 9. Para visualizar esse anáglifo utilizam-se óculos com lentes que possuem os canais de cores preservados, que atuam como filtros e cada cada olho recebe somente uma delas, Figura 10. Assim, como cada olho recebe uma imagem ligeiramente diferente uma da outra, propicia-se a disparidade de retina e a sensação de profundidade.

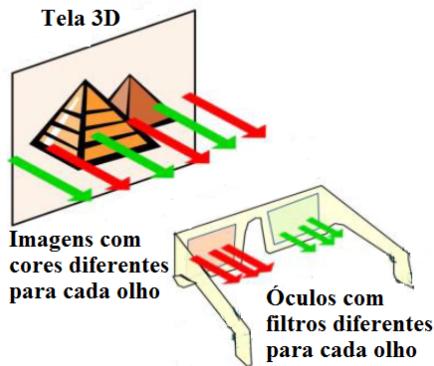
A codificação anaglífica é um método simples e que não requer equipamento caro ou complexo para ser executado ou visualizado. O vídeo anáglifo gerado possui alta taxa de compressão, uma vez que apenas um fluxo de vídeo é gerado, sendo vantajoso para o armazenamento e transmissão.

Figura 9 – Anáglifo



Fonte: [Zingarelli, Andrade e Goularte \(2011\)](#).

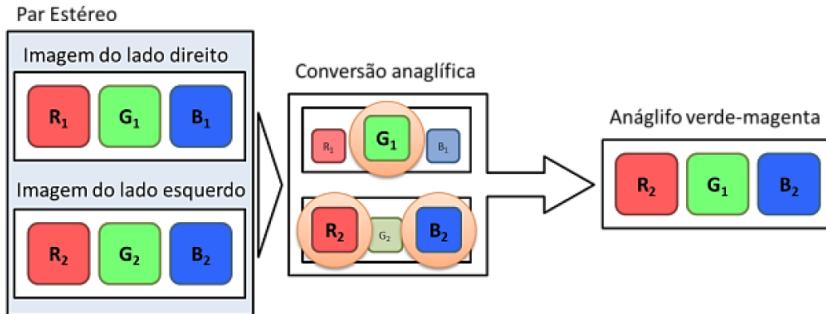
Figura 10 – Visualização anaglífica



Fonte: Adaptada de [Geng \(2013\)](#).

Na Figura 11, o par estéreo de vídeos formado por R1G1B1 e R2G2B2 é submetido ao processo de codificação anaglífica verde-magenta. Ao final do processo de codificação tem-se um anáglifo com o canal verde (G1) do lado direito e os canais R2 e B2 do lado esquerdo.

Figura 11 – Codificação anaglífica



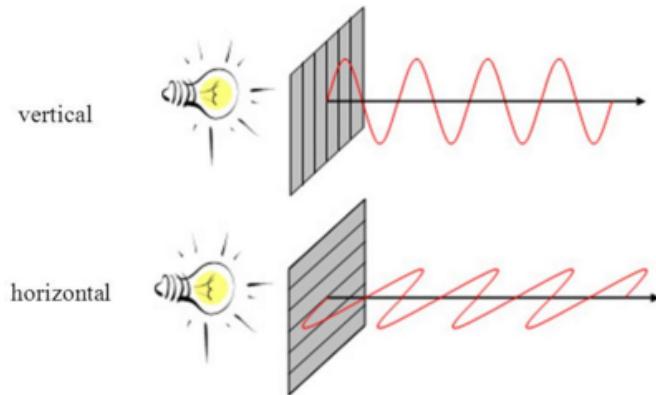
Fonte: [Zingarelli, Andrade e Goularte \(2011\)](#).

O anáglifo gerado, dependendo dos canais que são descartados na codificação, pode ser classificado como verde-magenta, vermelho-ciano ou azul-amarelo. Análises de qualidade subjetiva conduzidas por [Andrade e Goularte \(2010\)](#) com grupos de usuários mostraram 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.

2.3.1.2 Multiplexação por luz polarizada

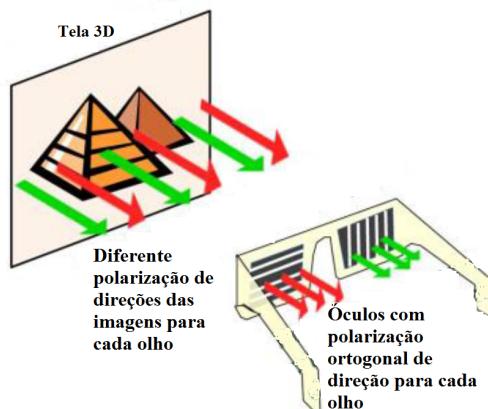
No método de multiplexação por luz polarizada, o par estéreo de vídeos é reproduzido separadamente e de forma integral, o que possibilita manter a qualidade real de cor da cena, porém, duplica o volume de dados armazenado. O processo de polarização, presente nos dois projetores, altera as ondas de luz para vibrarem na vertical e na horizontal (Figura 12) e envia uma imagem para cada olho por ondas de luz que vibram em uma só direção.

Figura 12 – Polarização da luz - vertical e horizontal



Os óculos, utilizados como filtro neste método, estão preparados para que um lado receba as ondas de luz na direção horizontal e do outro na vertical ([MENDIBURU, 2009](#)), como pode ser observado na Figura 13. Essa tecnologia é utilizada na maioria dos cinemas 3D atuais, porém, esse método requer dois projetores perfeitamente alinhados e sincronizados projetando em tela especial metalizada, o que aumenta a complexidade da reprodução dos conteúdos e, consequentemente encarecem a solução.

Figura 13 – Visualização por luz polarizada



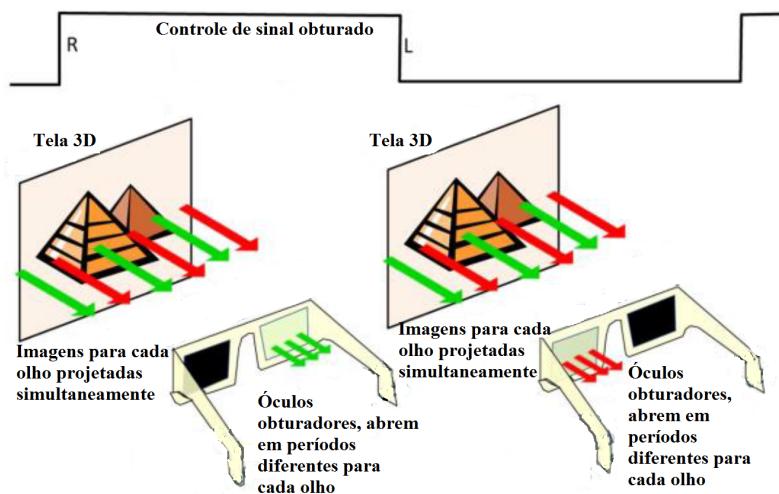
Fonte: Adaptada de [Geng \(2013\)](#).

2.3.1.3 Multiplexação temporal

Diferente dos óculos utilizados em vídeos anaglíficos e por luz polarizada, que filtram as imagens corretas para cada olho, a multiplexação temporal utilizam óculos obturadores (*shutter glasses*) que separam as imagens mecanicamente. Esse método possui um mecanismo de bloqueio que impede o olho direito de visualizar a imagem do olho esquerdo e vice-versa, explorando o efeito memória da imagem, pelo qual o cérebro consegue juntar imagens separadas por um tempo de até 50 milissegundos.

Esta é a tecnologia utilizada por alguns modelos dos atuais televisores 3D e funciona da seguinte maneira: o monitor exibe alternadamente, em alta frequência, as imagens para cada olho e os óculos, compostos por lentes de cristal líquido (LCD - *liquid crystal display*), alternam entre si o nível de opacidade de cada lente na mesma frequência do monitor. Com isso, por uma fração 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 muitas vezes a cada segundo, nossos olhos não notam a opacidade, e o efeito adquirido é a estereopsia (LIPTON, 1997; MENDIBURU, 2009). Este método é ilustrado na Figura 14.

Figura 14 – Multiplexação temporal com óculos obturadores



Fonte: Adaptada de Geng (2013).

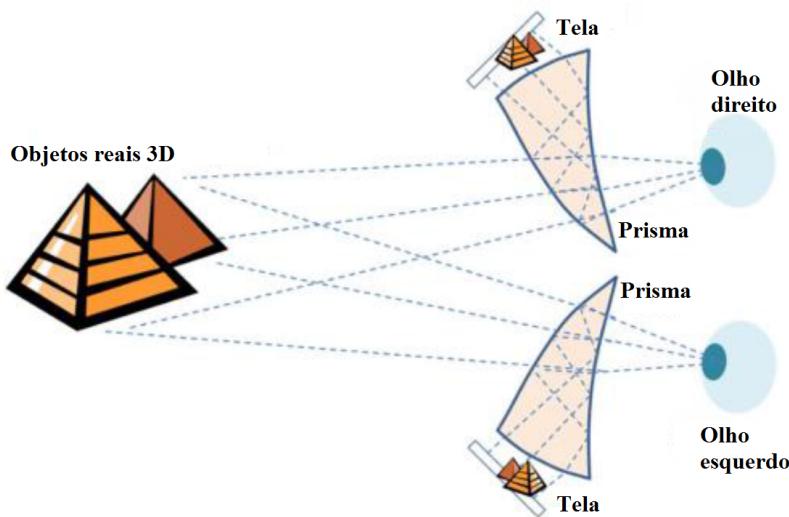
Os principais problemas desta técnica são: (i) alto custo para a produção dos óculos; (ii) necessidade de equipamento para sincronização entre a tela e os óculos; (iii) falta de um padrão para estes equipamentos, dificultando intercâmbio na produção e reprodução de conteúdo entre equipamentos diferentes; e (iv) a perda da resolução ou brilho das imagens, dependendo do padrão de reprodução utilizado para reduzir a impressão de instabilidade da sensação visual (*flicker*) no caso de vídeo entrelaçado (RICHARDSON, 2003).

2.3.2 Métodos com dispositivos HMD

Um dispositivo HMD (*Head Mounted Display*) (SUTHERLAND, 1968) é basicamente um capacete que possui duas telas distintas, uma para cada olho, tendo como objetivo a visualização estereoscópica com as vantagens de ser móvel e compacto (Figura 15). Esses dispositivos são muito utilizados nas áreas médica, militar e industrial (UREY *et al.*, 2011). Nessas áreas as possibilidades de um ambiente de realidade aumentada e de total imersão por parte do usuário são exploradas.

A captura de uma cena ocorre somando imagens do ambiente a outras informações gráficas às telas. Esse método, quando combinado a dispositivos de rastreamento de cabeça fornece aos usuários um efeito de "olhar ao redor", que é muito útil para usuários imersos nesses mundos virtuais. Graças a miniaturização da eletrônica, esses dispositivos estão começando a se tornar disponíveis e em custos mais razoáveis. No entanto, essa tecnologia, apresenta as mesmas limitações de alinhamento e sincronia de outros métodos estereoscópicos de visualização.

Figura 15 – Visualização utilizando HMD (*Head Mounted Display*)



Fonte: Adaptada de [Geng \(2013\)](#).

2.3.3 Métodos autoestereoscópicos

A abordagem autoestereoscópica visa o descarte dos óculos ou qualquer outro equipamento na visualização de vídeo estéreo que não seja o próprio monitor ou tela de projeção. A obrigatoriedade da utilização de óculos especiais, presente nas técnicas apresentadas anteriormente, é vista por esse método como uma abordagem invasiva, dado que o uso de óculos pode gerar certo desconforto ou até mesmo fadiga se utilizados por muito tempo.

2.3.3.1 Duas visões

A tecnologia, como o próprio nome diz, é capaz de gerar sozinha a sensação de profundidade nas imagens reproduzidas. Tal feito é realizado criando-se diferentes visões de uma mesma cena, por ângulos diferentes e limitados a certo segmento do campo de visão do espectador. Na indústria de TV 3D, o termo “monitor autoestereoscópico” é utilizado para referenciar duas tecnologias: a barreira de paralaxe e a película lenticular.

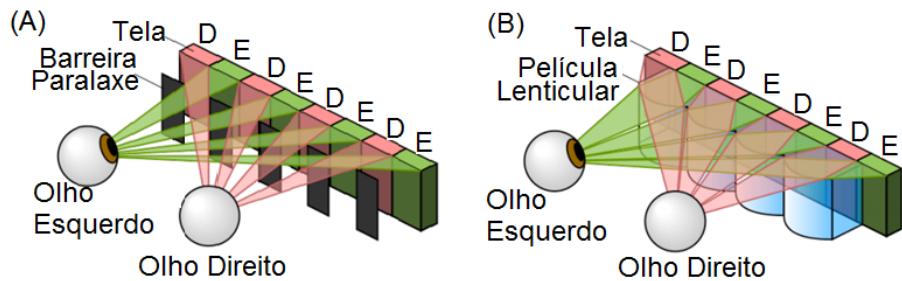
No método que utiliza a barreira de paralaxe (Figura 16 (A)) coloca-se na superfície do monitor uma barreira, composta por fendas estreitas intercaladas com placas opacas que

bloqueiam a luz em certas direções, para que cada olho tenha a visão de diferentes conjuntos de pixels (GOTCHEV *et al.*, 2011).

No outro método (Figura 16 (B)), coloca-se no monitor uma película especial chamada película lenticular, que é formada por pequenas lentes, as lentículas, capazes de direcionar a luz de cada imagem para um ângulo diferente. Além disso, o par estéreo de imagens é submetido a uma técnica de entrelaçamento (*interlacing*), na qual as imagens são cortadas em pequenas partes do tamanho das lentículas e são intercaladas. Com isso, cada pedaço da imagem é direcionado pelas lentículas para o respectivo olho (HUTCHISON, 2008).

Um problema ainda em estudo para monitores autoestereoscópicos é que o espectador deve se situar em pontos chaves para ter a visão estéreo, devido ao alcance limitado do campo de visão fornecido. Fora desses pontos a imagem aparece “borrada” ou não se percebe a profundidade. Além disso, ainda é uma tecnologia em aprimoramento e de alto custo de produção.

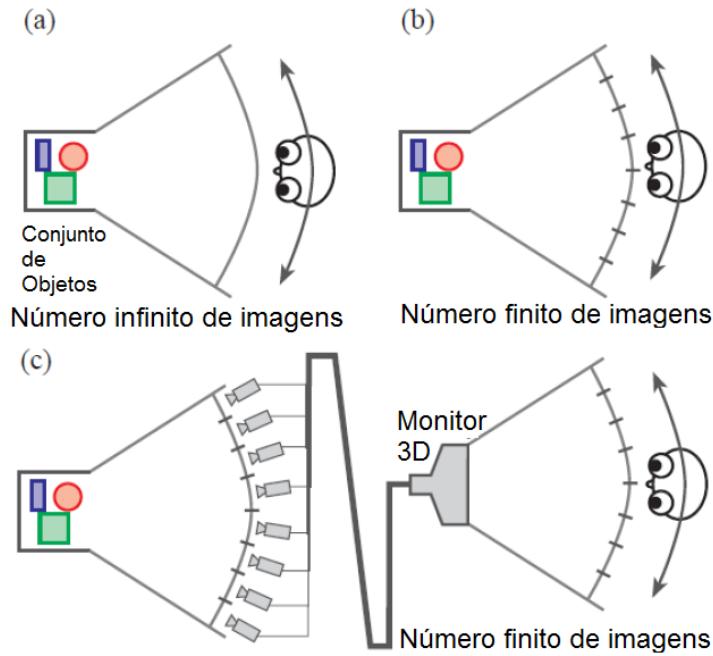
Figura 16 – Barreira de Paralaxe (A) Vs Película lenticular(B)



2.3.3.2 Multivisões

Os métodos multivisão autoestereoscópicos (*multi-view autostereoscopic*) são uma extensão dos métodos utilizados nos monitores autoestereoscópicos. São capazes de mostrar mais de um ponto de vista, conforme a Figura 17, para uma mesma cena e fornecer dicas de profundidade baseadas em movimento. As tecnologias discutidas anteriormente (de barreira paralaxe e película lenticular) podem ser utilizadas em conjunto com este método para fornecer pontos de vista adicionais. Geralmente com os métodos multivisão são providos de 5 até 28 diferentes pontos de visão. Na Figura 17 à medida que o usuário move a cabeça, diferentes feixes de luz, ou seja, diferentes pontos de vista, são percebidos. No mundo real, o número de pontos de vista é infinito (Figura 17 (a)); monitores multivisão autoestereoscópicos fornecem um número finito de pontos de vista do mundo (Figura 17 (b)), que podem ser captadas a partir de múltiplas câmeras (Figura 17 (c)).

Figura 17 – Princípios de monitores multivisão autostereoscópicos



Fonte: Adaptada de [Dodgson \(2002\)](#).

Estudos recentes, propõem métodos supermultivisão ([TAKAKI, 2014](#)) para monitores (SMV - *super-multiview displays*). Nesses estudos o intervalo entre pontos de vista é reduzido para que fique menor que o diâmetro da pupila do olho, resolvendo o conflito de convergência / acomodação na geração de paralaxe horizontal contínua.

2.4 Codificação estereoscópica

Para suprir as necessidades geradas pelas novas tecnologias de produção e reprodução de vídeos 3D, foi necessário o desenvolvimento de novos métodos de codificação de vídeos. Segundo [Smolic et al. \(2009\)](#), existem diversos formatos de vídeo estereoscópico, cada qual para um sistema específico, que exigem estruturas e implementações particulares.

Nesta seção são apresentadas algumas dessas codificações divididas com base na quantidade de pares estéreos que codificam. Assim, a Codificação Convencional (*Conventional Stereo Video - CSV*) codifica apenas um par estéreo, a Codificação em Múltiplas Visões (*Multiview Video Coding - MVC*) codifica mais que um par estéreo e a Codificação baseada em Vídeo e Profundidade (*Video Plus Depth V+D*) codifica um mapa de profundidade com um lado do par estéreo.

2.4.1 Codificação convencional

Este método desenvolvido por [Lipton \(1997\)](#) é o mais conhecido e simples modo de codificação de vídeo estéreo. Os métodos de codificação convencional são em geral adaptações ou extensões de métodos desenvolvidos para codificar vídeos 2D. Utiliza um par de vídeos de uma mesma cena, as imagens desses vídeos apresentam diferentes pontos de vista e têm o objetivo apresentar uma imagem distinta para cada olho. Nesses métodos somente dados sobre cor de pixel são utilizados. Após a captura por duas câmeras, dois sinais de vídeo podem receber algum processamento, como por exemplo, a correção de cor e a normalização. No entanto, nenhuma informação de geometria de cena é incluída. Os sinais de vídeo são destinados, em princípio, a serem exibidos diretamente usando um sistema de visualização, apesar de algum processamento de vídeo também poder estar envolvido antes da exibição ([SMOLIC et al., 2009](#)).

Uma maneira comum para codificar um vídeo CSV é utilizar a abordagem conhecida como quadro compatível (*frame-compatible*). Nessa abordagem, vídeos CSV são codificados como vídeos regulares 2D que multiplexam as visões esquerda e direita em um único quadro ou a sequência de quadros ([VETRO, 2010](#)).

Comparado a outros métodos de codificação de vídeo estéreo, os algoritmos associados ao método convencional são os menos complexos. Podem ser tão simples como apenas aplicar separadamente codificação e decodificação (MPEG-2, por exemplo) a múltiplos sinais de vídeo (dois sinais para vídeo estéreo).

A principal desvantagem do formato CSV é que a quantidade de dados duplica em relação ao vídeo monocular. Além disso, a falta de informação sobre a geometria explícita da cena faz com que seja difícil de implementar recursos de reescala de profundidade, que é uma característica necessária para fornecer uma boa qualidade de experiência em diferentes condições de visualização estereoscópica.

2.4.2 Codificação em múltiplas visões

A codificação em múltiplas visões é uma extensão da CSV e utiliza mais que dois vídeos que capturam uma mesma cena para fornecer diferentes pontos de vista a um observador localizado em frente da tela. Para isso, é necessária a captação sincronizada de diversas câmeras sequencialmente posicionadas a uma determinada distância.

Nesse método há um aumento significativo no volume de dados, pois cada visão possui um par estéreo de vídeo. Essa característica de fornecer diferentes pontos de vista (múltiplas visões) é utilizado em dispositivos autoestereoscópicos. A maioria das técnicas que utilizam MVC, exploram as redundâncias temporais existentes entre os quadros de uma determinada visão e as semelhanças entre os quadros sucessivos, utilizando as abordagens de compressão espacial, temporal e de disparidade.

O problema da MVC está no número de pontos de vista que dependendo do número de

vídeos utilizados influencia no aumento de volume do arquivo final. Geralmente, a quantidade de visões codificadas de maneira eficaz são limitadas em duas ou três visões (MULLER; MERKLE; WIEGAND, 2011).

2.4.3 Codificação baseada em vídeo e profundidade

Na codificação baseada em vídeo e profundidade, somado ao vídeo envia-se um mapa de profundidade, obtido através de cálculos complexos que mapeiam a cena fazendo a estimativa de disparidade e profundidade dos objetos nela contidos. Esses cálculos oneram o dispositivo por adicionarem processos de síntese e *rendering* tanto na codificação quanto na decodificação. Além disso, os algoritmos são complexos e ainda propensos a erros. Esses algoritmos podem ser divididos em três: V+D (*Video plus Depth*), MVD (*MultiView plus Depth*) e LDV(*Layered Depth Video*) (SMOLIC *et al.*, 2009).

O primeiro algoritmo, o V+D que somado ao sinal do vídeo envia um mapa de profundidade que habilita o dispositivo à criação do segundo vídeo tendo em vista a geração de estereopsia (ISO, 2007; ISO, 2008). O MVD é uma extensão do V+D, que envia no sinal de vídeo múltiplas visões de uma mesma cena, cada qual com seu próprio mapa de profundidade. Novas visões podem ser criadas combinando-se duas outras existentes. Com isso, é possível disponibilizar várias visões ao usuário, sendo um bom candidato a ser utilizado por monitores autoestereoscópicos (MERKLE *et al.*, 2007; ISO, 2008). O LDV inclui no sinal, além da camada do vídeo e seu mapa de profundidade, nomeada como visão principal, novas camadas responsáveis por outras visões, como dados contendo informações referentes à cena vista de outras direções. Tudo é processado para a criação de diferentes visões. A complexidade dos algoritmos aumenta, porém, o arquivo final é menor do que o do MVD. As camadas eliminam visões que não conseguem processar.

Como se pode observar, as desvantagens gerais dessas codificações são os algoritmos complexos e ainda propensos a erros, passíveis de um melhor estudo. Além disso, tem-se um processamento intenso tanto no lado transmissor quanto no receptor, exigindo equipamentos mais robustos e caros.

2.5 Qualidade de vídeo

A forma como o ser humano percebe as cores e o movimento é um fator importante para medir a qualidade de vídeo ou imagem. Segundo Winkler (2005), grande parte dos neurônios do cérebro estão envolvidos na percepção visual no complexo Sistema Visual Humano (HVS). Assim, tomar por base apenas a distorção calculada por uma métrica objetiva de qualidade, por exemplo o RMSE (*Root Mean Squared Error*), pode não representar a real qualidade visual de uma imagem. O RMSE pode calcular a quantidade de ruído causado por um artefato presente em uma imagem gerada em relação a imagem original. Esse valor pode ser expressivo, no

entanto, pode não representar problema para o espectador caso o artefato detectado esteja em uma localização que o HVS não perceba.

A avaliação da qualidade de vídeo ou imagem pode ser realizada utilizando medidas subjetivas e objetivas. As medidas subjetivas, geralmente, avaliam uma sequência de vídeos ou imagens. Essas medidas levam em conta a percepção de qualidade feita pelo observador como uma forma confiável para avaliar a qualidade de imagens. Também são eficientes para testar a performance e o desempenho de modelos que tentam simular o sistema visual humano. As medidas de avaliação objetivas são baseadas em modelos matemáticos que preveem, de forma automática, rápida e escalável, a qualidade das imagens. No entanto, a predição da qualidade de vídeo é uma tarefa difícil, devido à complexidade do sistema visual humano.

2.5.1 Medida de qualidade subjetiva

Segundo Winkler (2005) as expectativas do observador, o tipo do *display*, as condições de visualização e a presença de áudio podem influenciar nas notas dos observadores na avaliação de qualidade subjetiva de vídeo ou imagem. Segundo Pinson, Wolf e Gallagher (2004) o impacto introduzido pela tecnologia do *display* em uma avaliação pode ser reduzido utilizando testes subjetivos com dois estímulos (um original e outro processado).

Testes subjetivos para avaliação de qualidade visual foram formalizados nas recomendações ITU-R BT.500-11 (BT, 2002) e ITU-T P.910 (RECOMMENDATION, 1999). Essas recomendações sugeriram condições gerais para a avaliação subjetiva de vídeo (condições de observação, critérios para seleção de observadores e materiais utilizados nos testes, procedimentos de avaliação e métodos para avaliação dos dados obtidos nos testes, por exemplo). Portanto, as recomendações têm o objetivo principal de proporcionar uma forma de obter resultados úteis para futuras comparações.

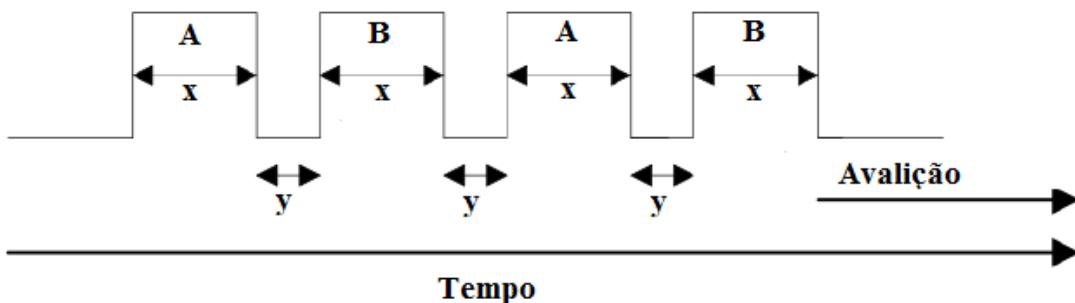
Diversos métodos de testes para avaliar a qualidade subjetiva foram definidos na recomendação ITU-R BT.500-11 como: DSIS, DSCQS, SSCQE, SDSCE e SAMVIQ. O mais comumente utilizado é o método *Double Stimulus Continuous Quality Scale* (DSCQS).

As imagens submetidas ao método DSCQS podem ter a qualidade mensurada subjetivamente utilizando o *Mean Opinion Score* (MOS) (RECOMMENDATION, 1999). Essa medida fornece uma indicação numérica da qualidade percebida pelo espectador. Nesta dissertação foram utilizados o método DSCQS juntamente com o MOS para avaliar a qualidade subjetiva do par estéreo recuperado.

No método DSCQS são apresentados ao observador múltiplas sequências de pares de vídeo ou imagem. Esses pares são compostos pelas imagens teste (imagem gerada) e referência (imagem original). Nesse método, a ordem de exibição das imagens dos tipos referência e teste pode ser apresentada de duas maneiras. Na primeira o observador não é informado sobre os tipos de imagens e a ordem de apresentação é aleatória. Na segunda o observador é informado dos

tipos de imagens e sobre a ordem de apresentação. A Figura 18, ilustra o esquema de avaliação DSCQS, em que imagens (A e B) são exibidas duas vezes, com tempos iguais para exibição (x), tempos iguais de transição (y) e tempo para avaliação no final de cada sequência de imagens.

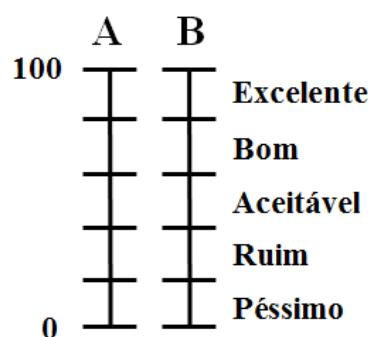
Figura 18 – Esquema de avaliação DSCQS



Fonte: Adaptada de BT (2002).

Após a sequência de exibição, cada imagem é avaliada separadamente com uma escala de qualidade que é de "Péssimo" a "Excelente", conforme a Figura 19. O resultado é obtido pela diferença entre a pontuação das imagens avaliadas, calculado em equivalente escala numérica de zero (Péssimo) a cem (Excelente). Esse cálculo é tipicamente utilizado para avaliações onde existe pouca diferença entre as sequências de teste e referência (ALPERT *et al.*, 1997; WINKLER, 2005).

Figura 19 – Escala de avaliação utilizada no DSCQS



Fonte: Adaptada de BT (2002).

O MOS representa a média das pontuações dadas pelos observadores na avaliação de qualidade subjetiva de vídeo ou imagem. Essas notas são representadas por um único número que varia de 1 a 5, conforme a Figura 20. Assim, a pontuação 5 é equivalente ao conceito de "Excelente", isto é, nenhum defeito foi percebido pelo observador. A pontuação 4, "Bom", o defeito foi percebido e não causou desconforto. A pontuação 3, "Aceitável", o defeito foi percebido e causou desconforto. A pontuação 2, "Ruim", apesar da grande degradação

na imagem o observador conseguiu visualizar alguma informação. E por fim, a pontuação 1, "Péssimo", a imagem tornou-se ininteligível e o observador ficou impossibilitado de extrair alguma informação.

Figura 20 – MOS

Qualidade	MOS
Excelente	5
Bom	4
Aceitável	3
Ruim	2
Péssimo	1

Fonte: Elaborada pelo autor.

2.5.2 Medida de qualidade objetiva

Os métodos para medir a qualidade objetiva de vídeo ou imagem são executados sem a interação humana, ou seja, a imagem original e a imagem a ser avaliada passam por um algoritmo que calcula a diferença entre elas automaticamente. A medida de qualidade objetiva possibilita resultados precisos relativos a quantidade de ruído presente na imagem gerada. Em [Wang et al. \(2003\)](#) são apresentados diversos algoritmos que avaliam a qualidade objetiva de vídeo.

As medidas de qualidade objetiva de vídeo ou imagem podem ser classificadas quanto à disponibilidade do vídeo original. Se o vídeo original estiver disponível, o sistema de avaliação de qualidade objetiva é chamado de Referência Total (FR). Quando não há disponibilidade, o sistema é chamado de Sem Referência (NR). Existe também a avaliação objetiva chamada de Referência Reduzida (RR), que utiliza algumas características do vídeo original auxiliando o sistema na detecção de falhas.

As medidas FR são as mais desenvolvidas e estudadas. Baseiam-se na comparação quadro a quadro entre o vídeo original (referência) e o vídeo processado (teste). Essas medidas necessitam de um alinhamento espacial e temporal preciso entre os vídeos para uma comparação quadro a quadro de boa qualidade. Para as medidas NR, a avaliação objetiva de qualidade é realizada com informações disponíveis na ponta do receptor (decodificador), isto é, não necessitam dos alinhamentos temporais e espaciais, pois, nenhuma comparação quadro a quadro é realizada.

A medida FR mais utilizada na literatura é o Pico Sinal Ruído (PSNR) ([WINKLER, 2005](#)), que baseia-se na diferença, pixel a pixel, entre duas imagens (original e processada), resultando em um valor medido em decibéis (dB), conforme a Equação 2.1. A escala utilizada por essa métrica varia de 0 à 100 dB, em que, quanto maior o valor, menor o nível de ruído encontrado na imagem processada, isto é, a imagem processada está com a qualidade próxima

da imagem original em termos de pixel. Apesar de ser uma métrica muito utilizada em trabalhos relacionados e bem sucedida para analisar a adição de ruídos em imagens, o PSNR não leva em consideração a percepção visual humana. Assim, não é possível afirmar que quanto maior o PSNR melhor a qualidade da percepção visual da imagem.

$$PSNR = 10 \log_{10} \frac{(2^d - 1)^2 WH}{\sum_{i=1}^W \sum_{j=1}^H (p[i, j] - p'[i, j])^2} \quad (2.1)$$

em que:

- d : profundidade de bits do pixel;
- W : largura da imagem;
- H : altura da imagem;
- $p[i, j]$, $p'[i, j]$: pixel na linha i e coluna j da imagem original e da imagem processada, respectivamente.

2.6 Trabalhos relacionados

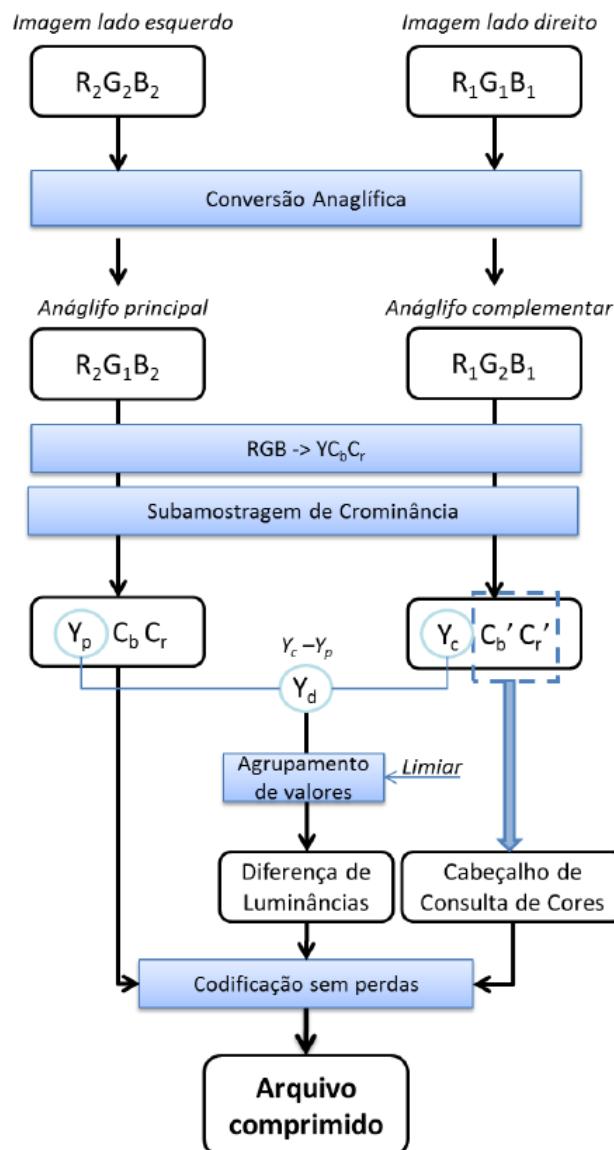
Nesta seção são apresentados os trabalhos que contemplam a utilização do vídeo anaglífico e técnica para recuperar o par estéreo. Atualmente três problemas podem ser observados na codificação de vídeos estereoscópicos. O primeiro é a quantidade de dados a ser armazenada, que dependendo da tecnologia de visualização a ser utilizada, emprega-se o uso de dois ou mais fluxos de vídeo. O segundo é dado pelos métodos tradicionais de compressão de vídeo monocular com perdas, que produzem artefatos prejudicando a percepção de profundidade quando utilizados em vídeos estereoscópicos; do mesmo modo, novas técnicas criadas especificamente para codificação estereoscópica produzem boa taxa de compressão, entretanto, são exclusivas para um método particular de visualização. O terceiro problema, em especial, está relacionado a como recuperar ou gerar as informações das cores dos canais descartados na codificação anaglífica. Os trabalhos descritos nesta seção visam solucionar parte desses problemas e estão relacionados a esta dissertação.

2.6.1 RevGlyph

A abordagem proposta por [Zingarelli, Andrade e Goularte \(2012\)](#), a técnica Revglyph, modificou o processo de codificação anaglífica adicionando duas novas estruturas: o cabeçalho de consulta de cores e a diferença de luminâncias. Essas estruturas, armazenam dados complementares que são utilizados no processo de reversão anaglífica para reconstrução do par estéreo.

No processo de codificação anaglífica, conforme a Figura 21, para separar apenas os componentes de crominância, o anáglifo complementar é convertido para o espaço de cores YC_bC_r e os componentes de crominância C_b e C_r são armazenados no cabeçalho de consulta de cores. Além disso, nesse espaço de cores é possível realizar a subamostragem de crominância, para diminuir a quantidade de dados de C_b e C_r . As informações de luminância do anáglifo principal (Y_p) e do anáglifo complementar (Y_c) são muito semelhantes entre si. Para obter compressão nessas componentes de luminância é feita uma operação de subtração entre os pixels homólogos em cada anáglifo que resulta na estrutura de diferença de luminâncias (Y_d). No 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.

Figura 21 – RevGlyph



Fonte: Zingarelli, Andrade e Goularte (2012).

No processo de reversão anaglífica, os dados contidos na matriz de luminância Y_d são reconstruídos 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 armazenados no cabeçalho de consulta de cores, possibilita que o anáglifo complementar seja reconstruído. Tendo o anáglifo principal ($R_2G_1B_2$) e o complementar ($R_1G_2B_1$), basta reordenar os canais para obter o par estéreo.

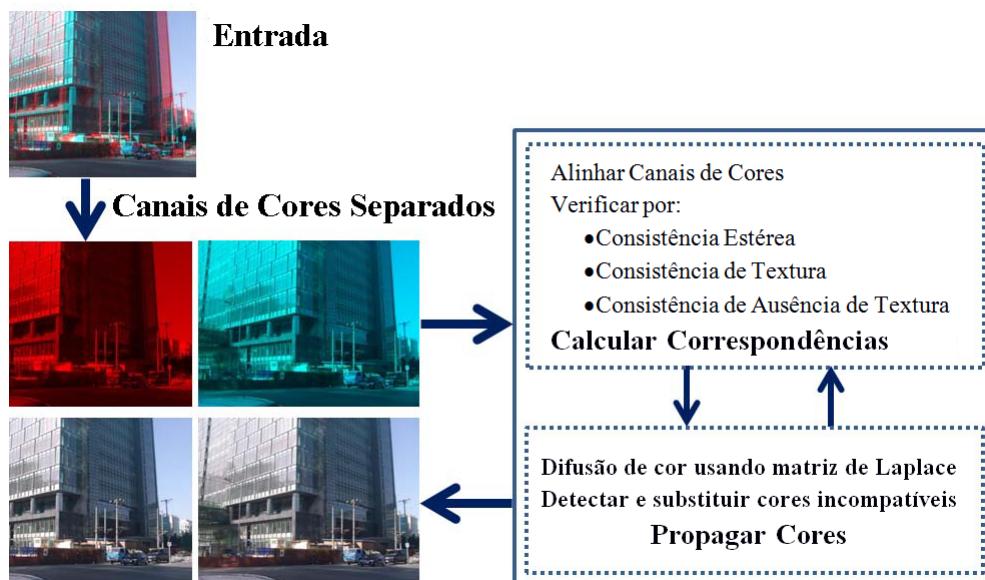
A técnica RevGlyph alcança altas taxas de compressão sem comprometer a qualidade da percepção de profundidade. No entanto, referente a reversão anaglífica, a desvantagem está na necessidade de armazenar informações do par estéreo original.

2.6.2 Recovering Stereo Pairs from Anaglyphs

A abordagem proposta por [Joulin e Kang \(2013\)](#) utilizada para recuperar o par estéreo a partir de um anáglifo consiste em encontrar iterativamente boas correspondências entre duas imagens e recolori-las com base nestas correspondências. Com uma etapa de pós-processamento, são detectadas cores “incompatíveis” em regiões sem correspondência (definidas como as cores que existem em uma imagem, mas não na outra) e para recolori-las foram utilizadas as cores conhecidas mais próximas. Este processo evita que novas cores apareçam nas regiões obstruídas.

O processo, conforme a Figura 22, é dividido em quatro etapas. Na primeira etapa são lidos os dados do vídeo anáglifo vermelho-ciano. Na segunda etapa, são separados os canais de cores do anáglifo. Na terceira etapa, é utilizada uma versão modificada do *SIFT-Flow* ([LIU et al., 2008](#)), chamada posteriormente de A-SIFT (Anaglyph-SIFT), para obter um conjunto inicial de correspondências entre as duas imagens.

Figura 22 – Visão geral da técnica Recovering Stereo Pairs from Anaglyphs



Fonte: Adaptada de [Joulin e Kang \(2013\)](#).

Essas correspondências têm que conter consistências estereoscópicas, de textura e de ausência de textura. Cada canal é encontrado de forma iterativa e independente usando um método de fluxo óptico para produzir novas correspondências. Na última etapa as cores são atualizadas por difusão de cor e as cores incompatíveis são detectadas e substituídas pelas cores conhecidas mais próximas.

A desvantagem dessa técnica é não funcionar bem com camadas sobrepostas, grandes desocluções e estruturas finas. Essas limitações acabam gerando artefatos nas imagens recuperadas. Outra limitação é utilizar um descritor *SIFT-Flow* próprio adaptado o *Anaglyph-SIFT* (A-SIFT). O descritor realiza três etapas: (i) extraír características, (ii) calcular o fluxo óptico e (iii) calcular correspondências. Embora esse descritor seja bastante eficiente, a quantidade de características detectadas e o tamanho dos vetores de características podem tornar a técnica computacionalmente cara. Um exemplo disso é que nesse método o algoritmo de fluxo óptico possui complexidade $O(n^3)$ (LIU *et al.*, 2008).

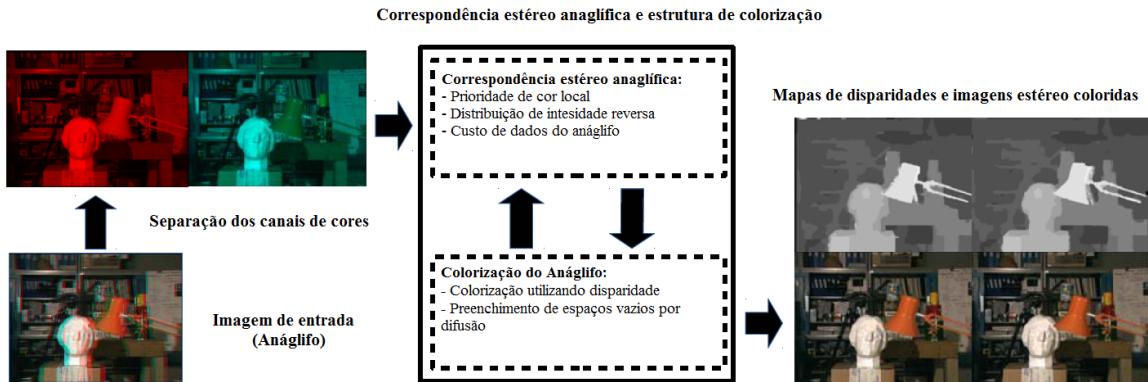
2.6.3 Depth Map Estimation and Colorization of Anaglyph Images Using Local Color Prior and Reverse Intensity Distribution

Na proposta de Williem, Raskar e Park (2015) é desenvolvida uma técnica iterativa para obter um mapa de disparidade preciso e, simultaneamente, colorir as informações de cores ausentes do anáglio para isso utilizou uma função de minimização de energia (BOYKOV; VEKSLER; ZABIH, 2001). Na visão geral da estrutura proposta, conforme a Figura 23, o mapa de disparidade de ancoragem é estimado empregando dois novos custos de dados do anáglio. Esses custos são baseados nas cores locais anteriores e na distribuição reversa de intensidade. Para restaurar a cor ausente, primeiro são transferidas as informações de cores conhecidas, de uma imagem para outra, utilizando a disparidade. Em seguida, são coloridos os pixels restantes em regiões oclusas, utilizando a colorização baseada em difusão. Uma nova função de kernel de peso baseada na similaridade de cores é introduzida para obter uma coloração precisa.

Embora a profundidade exata seja obtida usando os custos de dados propostos, a técnica ainda não pode alcançar a precisão do subpixel e nem recuperar as bordas dos objetos com precisão. Nas imagens geradas foram encontrados artefatos em torno das bordas do objeto com descontinuidade de profundidade.

Os autores assumem que o método é computacionalmente caro para reconstruir as cores ausentes de um anáglio e isto pode ser observado analisando a Figura 23. O método iterativo proposto para localizar as correspondências utiliza a função de minimização de energia de Boykov, Veksler e Zabih (2001) que possui complexidade $O(n^2)$ e enquanto estão processando as correspondências é executada a colorização de Levin, Lischinski e Weiss (2004) de complexidade $O(n^2)$. Então a complexidade desse método é no mínimo $O(n^4)$.

Figura 23 – Técnica proposta por Williem, Raskar e Park (2015)



Fonte: Adaptada de Williem, Raskar e Park (2015).

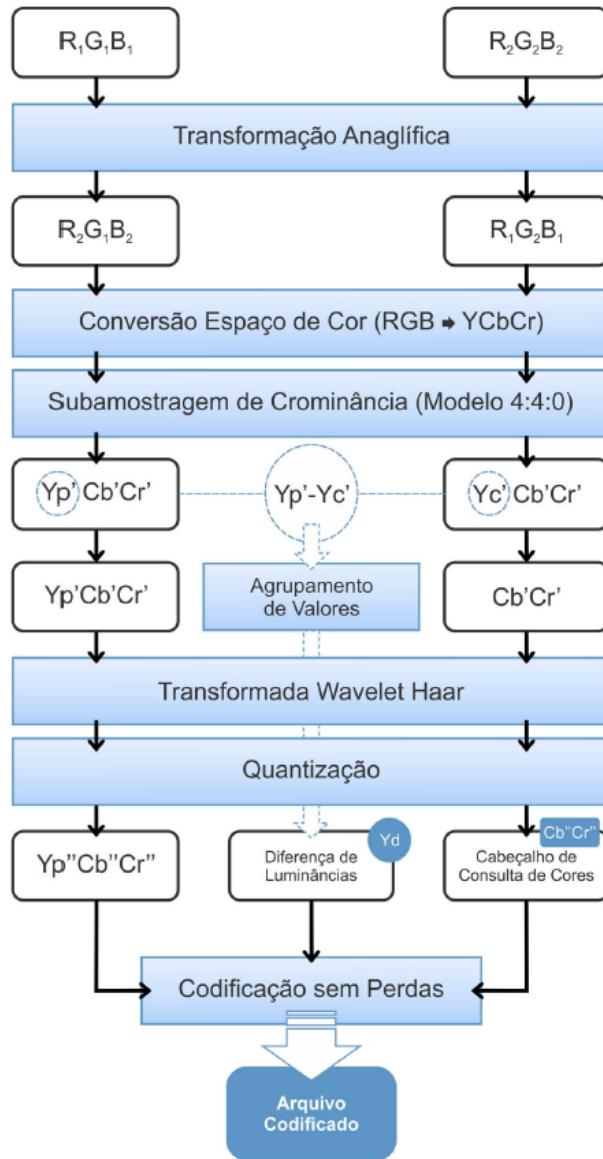
2.6.4 HaaRGlyph: A New Method for Anaglyphic Reversion in Stereoscopic Videos

A técnica HaaRGlyph de Rodrigues, Yugoshi e Goularte (2016) para a codificação de vídeo estereoscópico baseia-se na transformação anaglífica. A técnica desenvolvida utilizou a codificação espacial, com e sem perdas, de modo a atingir melhores taxas de compressão interferindo minimamente na qualidade da percepção de profundidade. Ao mesmo tempo, a técnica preserva informação suficiente e gera uma aproximação de qualidade do par estereópico original.

A Figura 24 ilustra o processo para o método de codificação proposto. A HaaRGlyph recebe como entrada um par estereópico, de imagens ou quadros de vídeo, no espaço de cores RGB, representando as visões do olho esquerdo ($R_1G_1B_1$) e direito ($R_2G_2B_2$). A primeira etapa do processo envolve a transformação do par estereópico em uma imagem anaglífica intitulada imagem anaglífica complementar. Na segunda etapa é realizada a conversão do espaço de cor RGB para YC_bC_r . Na terceira etapa, as duas imagens anaglíficas, principal e complementar, são então subamostradas segundo o modelo 4:4:0 (na Figura 24, respectivamente, $Y_pC_bC_r$ e $Y_c'C_b'C_r'$). Na quarta etapa, é realizada uma codificação por diferença (*Differential Coding*) entre os pixels homólogos de cada imagem $Y'_p - Y'_c$, na Figura 24 e, em vez de se armazenar diretamente os valores das diferenças, um agrupamento de intervalos de valores é realizado. O valor final a ser codificado é representado pela média aritmética dos valores agrupados na sequência. Essa etapa do processo é representada na Figura 24 como “Agrupamento de Valores”. Na quinta etapa, aplica-se uma Transformada Discreta Wavelet (*Discrete Wavelet Transform - DWT*) aos dados codificados anteriormente da imagem anaglífica principal e das componentes de crominância da imagem anaglífica complementar. Na sexta etapa é aplicada a quantização, na qual são quantizados os coeficientes DWT provenientes das imagens anaglíficas, principal ($Y'_pC'_bC'_r$) e complementar ($C'_bC'_r$). Como resultado da quantização obtém-se matrizes esparsas, contendo uma

quantidade grande de valores nulos. A última etapa do processo proposto é aplicar codificação sem perdas aos dados já quantizados na etapa anterior (na Figura 24, $Y_p''C_b''C_r''$ e $C_b''C_r''$) e a Y_d , proveniente da codificação diferencial das componentes de luminância das imagens anaglíficas principal e complementar.

Figura 24 – HaaRGlyph



Fonte: Adaptada de [Rodrigues, Yugoshi e Goularte \(2016\)](#).

A técnica HaaRGlyph consegue alcançar altas taxas de compressão sem comprometer a qualidade da percepção de profundidade. As limitações dessa técnica em relação a reversão anaglífica são o uso de um decodificador próprio e a dependência de informações do par estéreo original.

2.7 Considerações finais

Neste capítulo foram apresentados os principais fundamentos que envolvem a visualização estereoscópica. Foram apresentados alguns métodos de codificação e detalhados alguns conceitos básicos de técnicas que foram utilizadas para avaliar a qualidade do par estéreo recuperado. Também foram apresentados, para o melhor do nosso conhecimento, os trabalhos relacionados a reversão anaglífica presentes na literatura. Tais trabalhos apresentam lacunas como: a dependência de informações externas às imagens anaglíficas para recuperar aproximações do par estéreo; alto custo computacional. A proposta apresentada neste trabalho (Capítulo 3) difere das propostas de ([JOULIN; KANG, 2013](#)) e ([WILLIEM; RASKAR; PARK, 2015](#)) por possuir menor custo computacional e das propostas de ([ZINGARELLI, 2013](#)) e ([RODRIGUES, 2016](#)) por utilizar apenas informações internas às imagens anaglíficas.

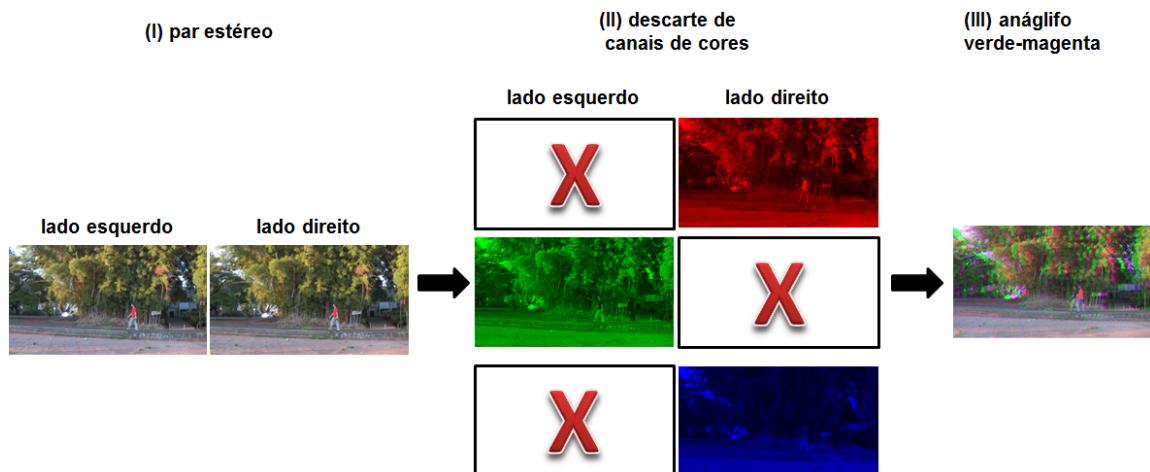
CAPÍTULO
3

SOLUÇÃO PROPOSTA

3.1 Considerações Iniciais

Na codificação anaglífica, Figura 25, um par estéreo de vídeos são codificados em apenas um vídeo anaglífico que ameniza o problema de volume dos dados para armazenamento e transmissão. No entanto, traz outro problema, o de não poder ser utilizado em outras técnicas de visualização estereoscópica que não seja a anaglífica.

Figura 25 – Codificação anaglífica

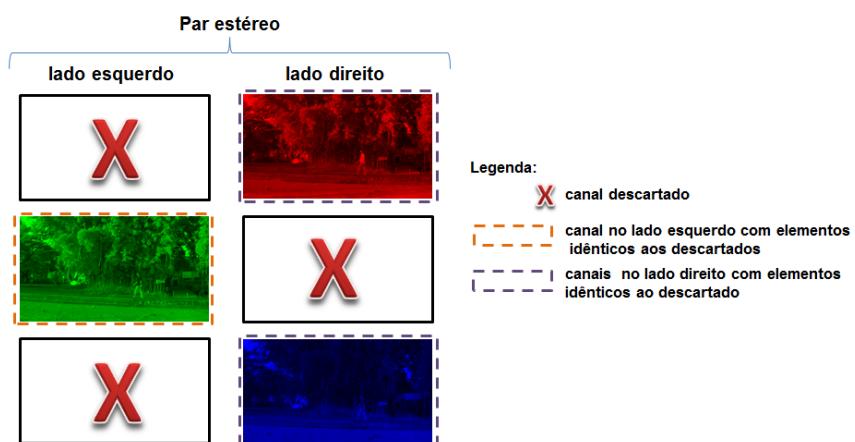


Fonte: Elaborada pelo autor.

Nesse caso, a Reversão Anaglífica torna-se necessária para recuperar uma aproximação do par estéreo original. Como todas as técnicas de visualização estereoscópica são compatíveis com o par estéreo, a reversão auxilia a tornar o conteúdo estereoscópico independente de método de visualização.

O maior obstáculo para realizar a Reversão Anaglífica está no descarte de informação que ocorre na codificação. Como exemplo, a codificação verde-magenta, no processo para gerar o vídeo anaglífico são descartados dois canais de cores do lado esquerdo do par estéreo e um canal do lado direito, conforme a Figura 26. São preservados o canal verde do lado esquerdo e os canais vermelho e azul do lado direito. Os canais descartados tinham os mesmos elementos de imagem dos canais preservados. Isso possibilita a atribuição dos elementos de imagem e de cor dos canais preservados para os canais que foram descartados. Por exemplo, atribuição do canal vermelho do lado direito ao canal verde do lado direito. Essa atribuição resolve o problema de saber qual imagem estava presente nesses canais antes do descarte, mas não o da cor.

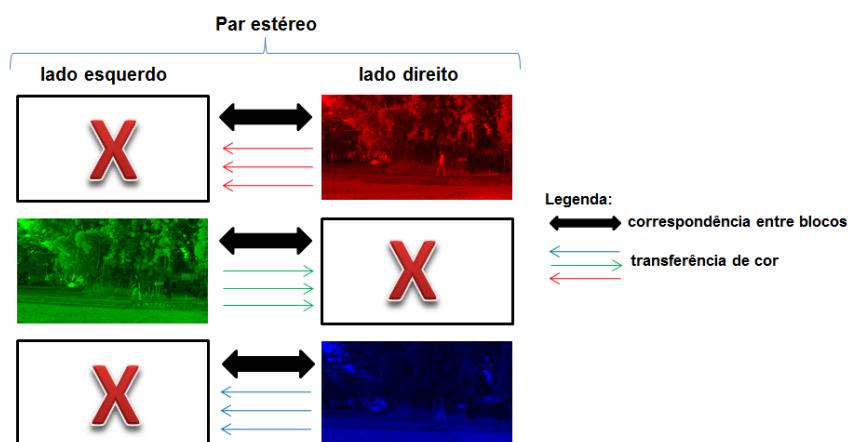
Figura 26 – Canais de cores preservados



Fonte: Elaborada pelo autor.

A maior parte das cores necessárias para complementar cada lado do par estéreo estão presentes nos canais preservados, porém em lados opostos, conforme a Figura 27. O maior desafio é justamente localizar e transferir a cor correta que está de um lado para o lado oposto na posição correta.

Figura 27 – Transferência de cores



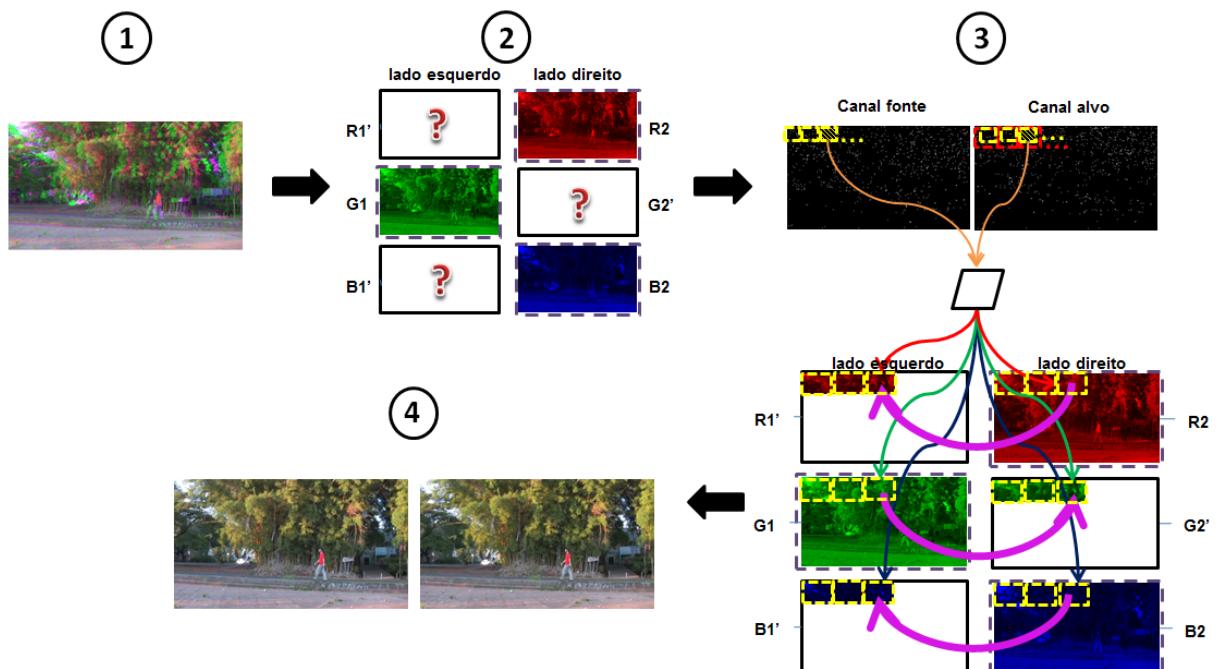
Fonte: Elaborada pelo autor.

Existe uma ligeira diferença de angulação e posicionamento entre as imagens do par estéreo. Essa diferença dificulta a localização e a transferência da cor correta entre canais correspondentes do par estéreo. A abordagem utilizada neste trabalho para atacar este problema é inspirada no algoritmo de correspondência entre blocos (*Block Matching Algorithm* - BMA) para recuperar uma aproximação do par estéreo original, comumente aplicada em compressão de vídeo.

Nesse aspecto, de localizar e transferir a cor, diferencia-se das técnicas propostas do estado da arte. A proposta de [Joulin e Kang \(2013\)](#) utiliza uma adaptação do descritor SIFT Flow adaptado para anáglifos (A-SIFT), a técnica de [Williem, Raskar e Park \(2015\)](#) utiliza a função de minimização de energia ([BOYKOV; VEKSLER; ZABIH, 2001](#)) para gerar um mapa de disparidade e colorir o anáglifo e as propostas de [Rodrigues, Yugoshi e Goularte \(2016\)](#) e [Zingarelli, Andrade e Goularte \(2012\)](#) geraram estruturas com informações de cor complementares ao vídeo anáglifo.

A técnica proposta nessa dissertação para recuperar uma aproximação do par estéreo original, a ARBFLS, apresentada na Figura 28 é dividida em 4 etapas: (i) leitura de imagem de entrada, (ii) processamento de imagem, (iii) correspondência entre blocos com transferência de cor e (iv) aproximação do par estéreo original. Nas seções seguintes são detalhadas cada uma delas.

Figura 28 – Visão geral da abordagem proposta



Fonte: Elaborada pelo autor.

3.2 Leitura de imagem de entrada

Nesta etapa inicial são executados os processos de leitura e ajuste dos parâmetros para a codificação da imagem anaglífica de entrada. São lidas as informações de arquivo e as informações de imagem coletadas diretamente dos respectivos cabeçalhos. Do cabeçalho de arquivo são utilizadas as informações de assinatura de arquivo e a do início da área de dados.

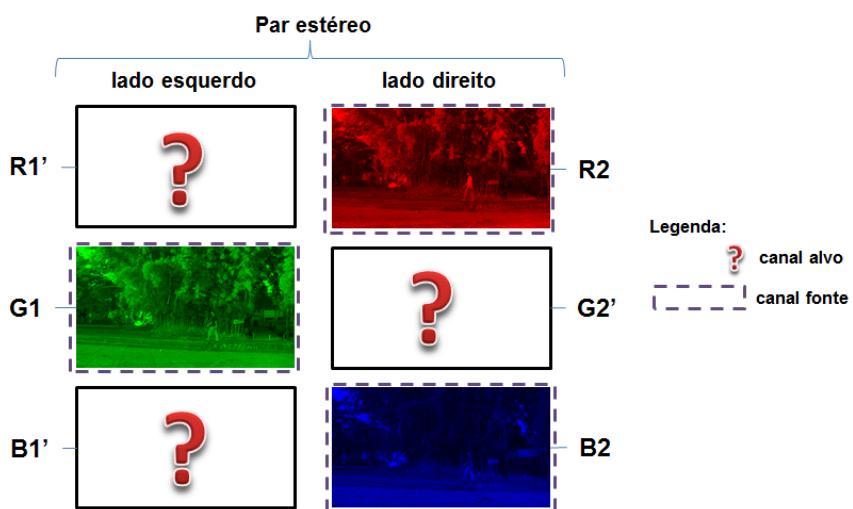
Neste trabalho são utilizadas como entrada imagens no formato de mapa de bits (BMP) e, de seus cabeçalhos de arquivo são extraídas informações de altura e largura da imagem e o número de bits por pixel. Com essas informações pode-se iniciar a técnica de reversão anaglífica proposta nesta dissertação.

3.3 Processamento de Imagem

Esta etapa é dividida em 2 processos: (i) separação dos canais de cores e (ii) aplicação do detector de bordas de Canny. O primeiro visa organizar os canais de cores em alvo ou fonte e o outro torna possível a comparação entre eles.

Na Figura 29 são apresentados os canais separados em cada lado do par estéreo, para o espaço de cor RGB, do anáglifo verde-magenta. Nessa separação é possível identificar o canal verde (G1) do lado esquerdo e os canais vermelho (R2) e azul (B2) do lado direito, esses canais são marcados como fonte. Também são identificados os canais de cores descartados para gerar o anáglifo, do lado esquerdo (R1' e B1') e do direito (G2'), que são marcados como alvo. Também, é possível notar que os canais marcados como alvo complementam os canais fonte para cada lado do par estéreo.

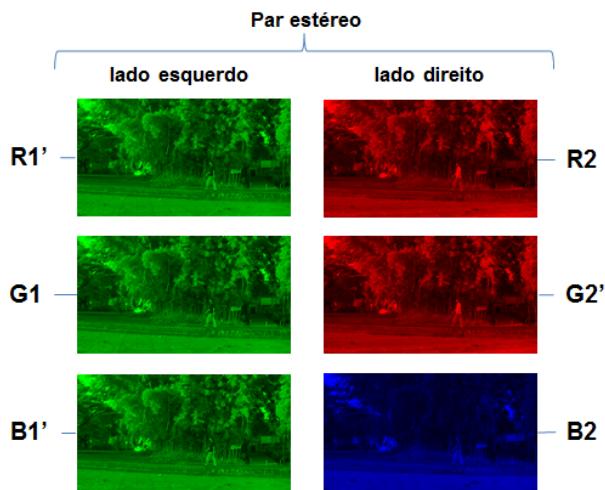
Figura 29 – Separação de canais de cores



Fonte: Elaborada pelo autor.

Após o processo de separação, os canais alvo estão vazios e para completá-los basta preenchê-los com os elementos de imagem contidos nos canais fonte dos respectivos lados. Assim, para os canais alvo do lado esquerdo, Vermelho ($R1'$) e Azul ($B1'$), são atribuídas as coordenadas planas dos elementos da imagem e a cor do canal Verde ($G1$). No lado direito, para o canal alvo Verde ($G2'$) são atribuídas as coordenadas e a cor do canal Vermelho ($R2$). Dessa forma são recuperadas as informações das coordenadas e os elementos de imagem para todos os canais alvo que estavam vazios inicialmente, conforme a Figura 30.

Figura 30 – Atribuição de coordenadas e cores



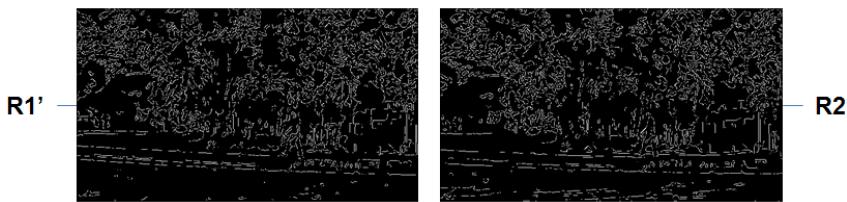
Fonte: Elaborada pelo autor.

A comparação entre os canais alvo ($R1'$, $G2'$ e $B1'$) e os canais fonte ($R2$, $G1$ e $B2$) ainda não pode ser realizada, pois as imagens estão em cores diferentes. O canal $R1'$ está com a cor verde e $R2$ está com a cor vermelha e essa diferença também ocorre com os outros canais alvo e fonte. Para resolver esse problema é realizada a aplicação do detector de bordas de Canny nesses canais. Esse detector de bordas gera como saída imagens binárias, contendo informação de contorno dos objetos presentes nas imagens que possibilita a comparação para fins localização de objetos (partes da imagem correspondentes entre duas imagens). A Figura 31 ilustra os canais alvo e fonte após a aplicação do detector de bordas de Canny.

O detector de bordas de Canny possui três parâmetros. O primeiro parâmetro é o tamanho do filtro Gaussiano, valores menores para esse parâmetro causam menos *blur* (borrão) e permitem a detecção de linhas pequenas e bem nítidas. Um filtro maior causa mais *blur*, o que é mais útil para detectar bordas maiores, mais suaves. O segundo e o terceiro parâmetros são dois limiares iniciais um superior outro inferior. Supondo que as bordas devem ser linhas contínuas, linhas mesmo com pouca intensidade são investigadas, mas evita-se identificar pixels que não constituem uma linha. Assim aplica-se primeiro o limiar superior (*high threshold*). Isto marca as bordas que possivelmente podem ser genuínas. Partindo destas, usa-se a informação direcional para identificar as bordas da imagem. Ao seguir uma linha, usa-se o limiar inferior (*lower*

threshold), permitindo seguir mesmo fracas possibilidades de bordas a partir de um ponto inicial. Nesta dissertação foram testados empiricamente diversos valores de filtro, limiar superior e inferior. Os valores 0.5, 1, 2 para os respectivos parâmetros de filtro, *high threshold*, *lower threshold* obtiveram o melhor resultado. Toda a fundamentação matemática e maiores detalhes sobre o descriptor de bordas de Canny podem ser encontrados no trabalho original de Canny ([CANNY, 1986](#)).

Figura 31 – Imagens binárias geradas pelo detector de bordas de Canny



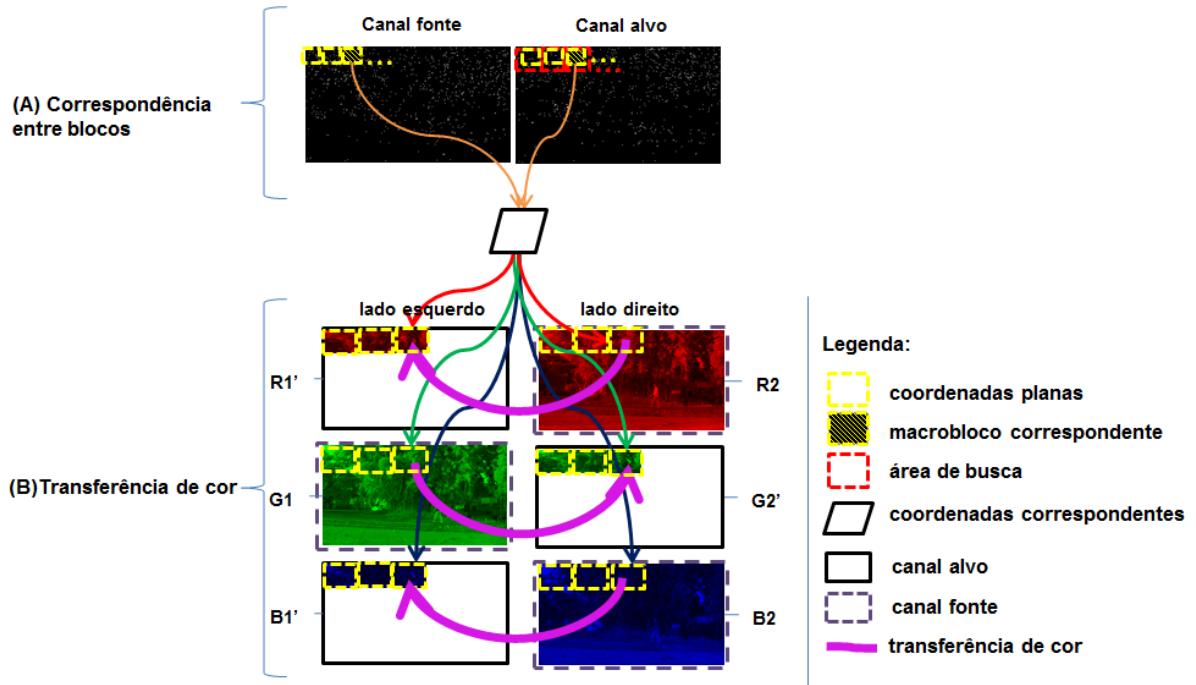
Fonte: Elaborada pelo autor.

3.4 Correspondência entre blocos com transferência de cor

O par estéreo possui duas imagens ligeiramente diferentes uma da outra, sendo que os elementos das imagens diferem em posicionamento e angulação. Ao comparar esses elementos percebe-se que eles estão minimamente deslocados horizontalmente, mimetizando um movimento. Por essa semelhança é possível utilizar o algoritmo de correspondência entre blocos (*Block Matching Algorithm - BMA*) para relacionar os canais alvo e fonte. Dessa maneira pode-se identificar as coordenadas planas da cor no canal fonte e para quais coordenadas essa cor deve ser transferida no canal alvo. O objetivo dessa etapa é gerar a correspondência de coordenadas e transferir a cor entre os canais alvo e fonte simultaneamente.

Para realizar a busca dos blocos correspondentes entre os canais alvo e fonte são definidos macroblocos com dimensões iguais (em pixels) que vão percorrer os canais alvo e fonte. Esses macroblocos podem ter as dimensões selecionadas entre 8x8, 16x16, 32x32 e 64x64. No processo de busca do macrobloco no canal fonte, é definida uma janela de busca com dimensões de 4 pixels a mais na altura e na largura sobre as dimensões do macrobloco no canal fonte. Por exemplo, se o macrobloco for definido com dimensões de 8x8 a janela de busca terá as dimensões de 12x12. Os macroblocos alvo e fonte são comparados até cobrir toda a janela de busca no canal fonte. O menor valor para a medida de correspondência entre os macroblocos SAD ([RICHARDSON, 2004](#)), indica que os macroblocos são correspondentes, isto é, que os elementos de imagem do macrobloco do canal alvo foram encontrados na área de busca do canal fonte, como apresentado na Figura 32 (A).

Figura 32 – Busca de correspondência entre blocos e transferência de cor



Fonte: Elaborada pelo autor.

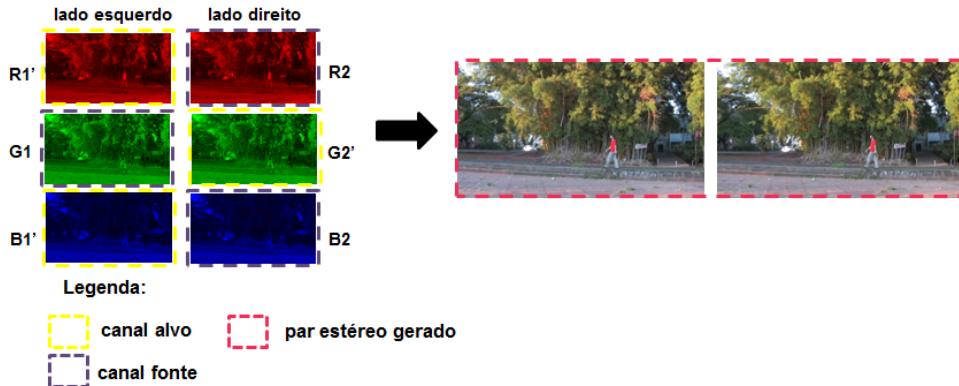
Em seguida é realizada a busca de um novo macrobloco do canal alvo no canal fonte em uma nova área de busca delimitada. Esse processo de selecionar um novo macrobloco e buscá-lo na nova área de busca delimitada é executado até cobrir toda a imagem do canal alvo. Enquanto o processo de busca é realizado o processo de transferência de cor entre os canais correspondentes também é executado, em que são realizadas as transferências de informações, bloco a bloco, entre os canais alvo ($R1'$, $G2'$ e $B1'$) e fonte ($R2$, $G1$ e $B2$), conforme a Figura 32 (B). No final dessa etapa são recuperados uma aproximação dos canais alvo ($R1'$, $G2'$ e $B1'$).

3.5 Reconstrução do par estéreo

O par estéreo é comum a todas as técnicas de visualização estereoscópica. Com a codificação anaglífica são descartados canais que não podiam mais ser recuperados. Assim, o resultado da técnica proposta, ARBFLS, é uma aproximação desses canais descartados. A junção dos canais alvo com os canais fonte resulta em uma aproximação do par estéreo original.

Nessa etapa final para recuperar uma aproximação do par estéreo original é necessário agrupar os canais alvo ($R1'$, $G2'$ e $B1'$) com os canais fonte ($R2$, $G1$ e $B2$), para cada lado do par estéreo. Dessa maneira, para o lado esquerdo são agrupados os canais $R1'$, $G1$ e $B1'$ e para o lado direito os canais $R2$, $G2'$ e $B2$. Cada lado é organizado em um arquivo BMP e são armazenados separadamente gerando uma aproximação do par estéreo original. Conforme apresentado na Figura 33.

Figura 33 – Agrupamento dos canais e par estéreo gerado

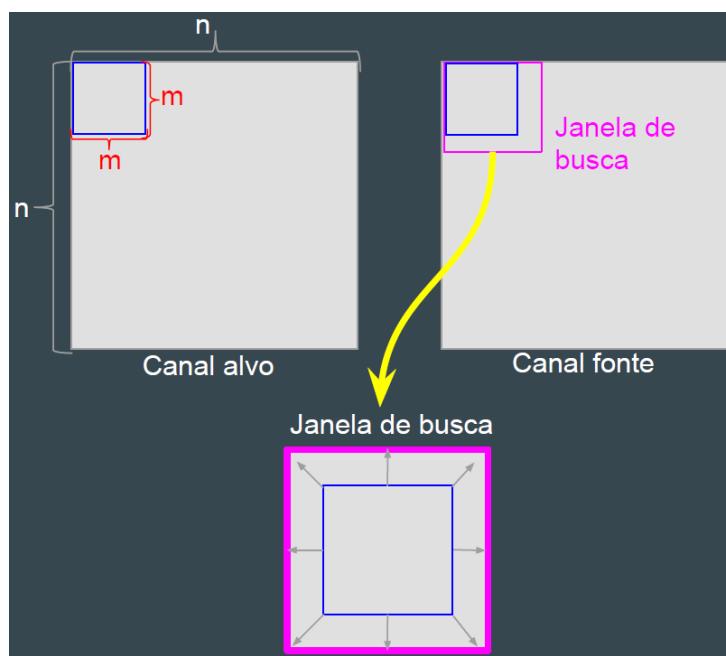


Fonte: Elaborada pelo autor.

3.6 Análise de custo da técnica proposta

O que diferencia esta técnica das demais é a operação de encontrar macroblocos correspondentes nos canais alvo e fonte. O anáglifo possui os canais alvo e fonte minimamente deslocados. Essa característica do anáglifo possibilita reduzir consideravelmente as dimensões da janela de busca. A busca de macroblocos correspondentes utiliza uma técnica local e como a operação de busca do macrobloco não é realizada em janelas de busca anteriores no canal fonte a quantidade de buscas nesse canal é reduzida. O processo de busca de correspondência entre macroblocos é ilustrado na Figura 34.

Figura 34 – Busca de correspondência entre macroblocos

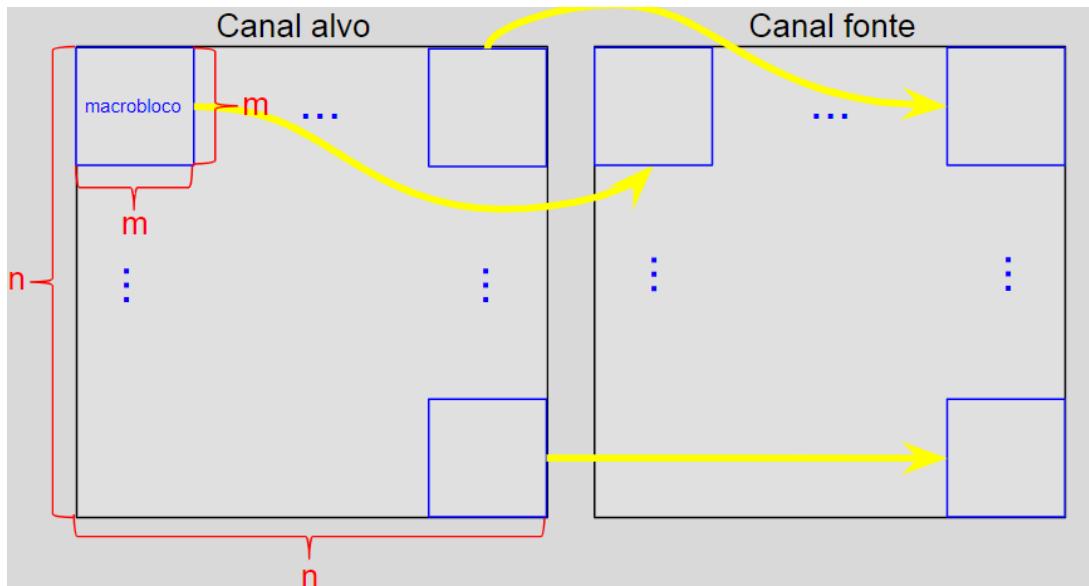


Fonte: Elaborada pelo autor.

Na Figura 34, os macroblocos estão representados na cor azul nos canais fonte e alvo. A janela de busca está representada na cor rosa e possui 4 pixels a mais que as dimensões do macrobloco no canal fonte. Uma visão detalhada da janela de busca e do macrobloco no canal fonte é indicada pela seta amarela.

A busca por correspondência de um macrobloco do canal alvo no canal fonte é realizada sequencialmente e sem sobreposição neste canal. Considerando que o canal alvo tem dimensões $n \times n$ e considerando um macrobloco de dimensões $m \times m$ são realizadas $(n/m)^2$ buscas de um macrobloco do canal alvo no canal fonte, como ilustrado na Figura 35.

Figura 35 – Busca de correspondência entre macroblocos sem sobreposição



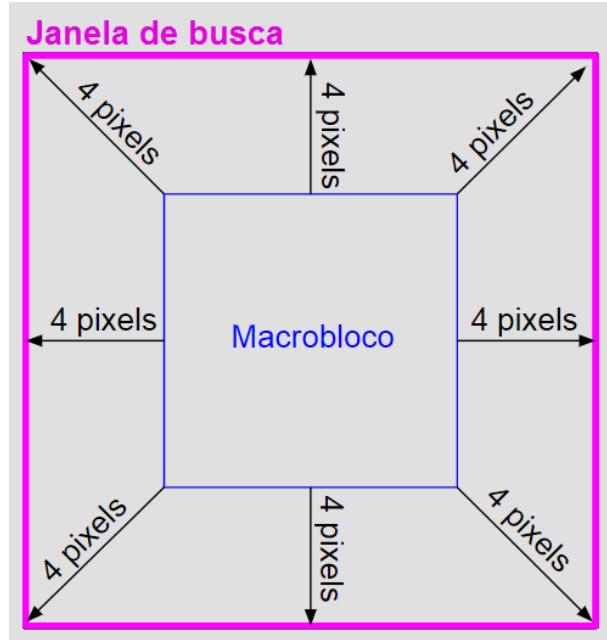
Dessa forma, o custo da técnica ARBFLS, em operações, pode ser definido como:

$$C(\text{ARBFLS}) = (n/m)^2 \times C(\text{busca_fonte}), \quad (3.1)$$

onde $C(\text{busca_fonte})$ é o custo, em operações, da busca de um macrobloco do canal alvo no canal fonte.

Ao procurar um macrobloco do canal alvo no canal fonte, são efetuadas comparações daquele macrobloco com o macrobloco correspondente no canal fonte e também com os macroblocos vizinhos, nas oito possíveis direções. Todos os macroblocos vizinhos, distantes até quatro pixels do macrobloco central correspondente no canal fonte são verificados, conforme ilustrado na Figura 36. Observe que são realizadas $8 \times 4 + 1$ operações.

Figura 36 – Busca de correspondência no canal fonte



O custo $C(\text{busca_fonte})$ é dado então por

$$C(\text{busca_fonte}) = (8 \times 4 + 1) \times C(\text{comp}), \quad (3.2)$$

onde $C(\text{comp})$ é o custo, em operações, da operação de comparação entre dois macroblocos.

As comparações entre dois macroblocos são efetuadas utilizando a medida de correspondência SAD (soma das diferenças absolutas). O cálculo do SAD entre dois macroblocos B_1 e B_2 quaisquer, de mesmas dimensões $m \times m$, é definido na Equação 3.3:

$$\text{SAD } (A, B) = \sum_{i=0}^m \sum_{j=0}^m |A(i, j) - B(i, j)| \quad (3.3)$$

em que $A(i, j)$ é o valor do pixel na posição (i, j) da matriz de pixels do macrobloco A e $B(i, j)$ é o valor do pixel na posição (i, j) da matriz de pixels do macrobloco B . Pode-se observar na Equação 3.3 que o cálculo do SAD, por realizar dois somatórios aninhados de m valores, realiza m^2 operações. Com isso tem-se que:

$$C(\text{comp}) = m^2 \quad (3.4)$$

Substituindo as equações 3.4 em 3.2 e o resultado disso em 3.1, obtém-se:

$$C(\text{ARBFLS}) = (n/m)^2 \times (8 \times 4 + 1) \times m^2 \quad (3.5)$$

$$C(\text{ARBFLS}) = n^2/m^2 \times m^2 \times (33) \quad (3.6)$$

$$C(\text{ARBFLS}) = n^2 \times 33 \quad (3.7)$$

Como $C(\text{ARBFLS}) = n^2 \times 33 \leq O(n^2)$, conclui-se que o método ARBFLS tem complexidade $O(n^2)$.

3.7 Considerações finais

Neste capítulo foi apresentada a técnica ARBFLS de reversão anaglífica em vídeos estereoscópicos que gera como resultado uma aproximação do par estéreo. Para isso, é utilizada a correspondência entre blocos, para localizar e transferir a cor entre canais alvo e fonte.

Os processos de leitura de imagem, separação de canais, busca de correspondências entre blocos, operações com valores de pixels no cálculo do SAD, transferência de cores e geração da aproximação do par estéreo foram desenvolvidos sem o auxílio de bibliotecas particulares voltadas ao processamento digital de imagens e visão computacional. A única exceção é a aplicação do detector de bordas de Canny nos canais alvo e fonte, na etapa de processamento de imagem. Para isso é utilizada uma função nativa da biblioteca *OpenCV* e os seus respectivos detalhes bem como o *download* da biblioteca podem ser encontrados no endereço <https://docs.opencv.org/2.4/modules/imgproc/doc/feature_detection.html?highlight=canny#canny>.

A reversão anaglífica é proposta como um processo direto, isto é, sem precisar de informações complementares ao anáglifo. Isso por um lado é positivo, pois possibilita aplicar essa técnica em qualquer anáglifo, tornando-a genérica. Por outro lado, o par estéreo recuperado é uma aproximação do par estéreo original, em que as cores geradas são transferidas entre canais de imagens ligeiramente diferentes uma da outra, o que pode gerar ruídos. Uma avaliação experimental acerca da qualidade das imagens geradas é apresentada no Capítulo 4, a seguir.



AVALIAÇÃO EXPERIMENTAL

4.1 Considerações Iniciais

Neste capítulo, são apresentados os experimentos realizados com dois *datasets* de imagens anaglíficas submetidos à técnica ARBFLS, proposta no Capítulo 3, para gerar uma aproximação do par estéreo original. Esses experimentos visam avaliar a qualidade objetiva e subjetiva de imagem dos pares estéreos gerados, definidas na Seção 2.5. Os resultados obtidos são comparados com as técnicas do estado da arte apresentadas na Seção 2.6.

Este capítulo está organizado em mais quatro subseções. Na Seção 4.2 são apresentados os *datasets* utilizados nos experimentos. Nas Seções 4.3 e 4.4, são apresentados os resultados da avaliação objetiva e subjetiva de imagens, respectivamente. E por fim, na Seção 4.5 são apresentadas as considerações finais.

4.2 Dataset de imagens estéreo

Nas avaliações experimentais são utilizados dois *datasets* de imagens estéreo no formato BMP (Bitmap). O primeiro, construído por Andrade (2012), composto por 32 pares estéreos foi utilizado por Zingarelli (2013) e depois por Rodrigues (2016). O segundo, o *Middlebury dataset* (SCHARSTEIN; SZELISKI, 2002) com 04 pares estéreos que foram utilizados por Joulin e Kang (2013) e Williem, Raskar e Park (2015). Na Tabela 1 são relacionados os trabalhos do estado da arte com os respectivos *datasets*.

Dos *datasets* de pares estéreos são gerados os anáglifos, pré-ajustados nas dimensões planares em aproximadamente 30 pixels a menos nas bordas, utilizados como entrada na técnica proposta. Do primeiro, dataset foram gerados 32 anáglifos do tipo verde-magenta e o do segundo foram gerados 4 anáglifos do tipo vermelho-ciano. Na Tabela 2 são apresentados para cada dataset os totais de imagens originais, anaglíficas recuperadas que são utilizadas nos experimentos.

Tabela 1 – Relação dos trabalhos com os *datasets*

Trabalhos Relacionados	Datasets	
	Andrade (2012)	Scharstein e Szeliski (2002)
Zingarelli (2013)	X	-
Joulin e Kang (2013)	-	X
Williem, Raskar e Park (2015)	-	X
Rodrigues (2016)	X	-
Técnica proposta	X	X

Fonte: Elaborada pelo autor.

Tabela 2 – Descrição de imagens por dataset

Tipo de Imagem	Dataset		Total
	Andrade (2012)	Scharstein e Szeliski (2002)	
Original	64	8	72
Anáglifo gerado	32	4	36
Total de imagens	96	12	

Fonte: Elaborada pelo autor.

4.3 Avaliação objetiva de imagem

A técnica proposta recupera as aproximações dos pares estéreos originais para cada dataset de anáglifos. Em seguida, cada par estéreo gerado é submetido à avaliação objetiva de imagem. Essa avaliação compara os pares estéreos gerados com os respectivos pares originais, para isso utiliza a métrica PSNR. Essa métrica, apresentada na Seção 2.5.2, calcula a diferença, pixel a pixel, entre a imagem original e a imagem processada, resultando em um valor medido em decibéis (*dB*). A escala utilizada por essa métrica varia de 0 à 100 *dB*, em que, quanto maior o valor, menor o nível de ruído encontrado na imagem processada, isto é, a imagem processada está com a qualidade próxima a imagem original em termos de pixel.

Com o propósito de realizar uma análise comparativa da técnica proposta com os trabalhos anteriores que utilizam anáglifos, são selecionados os seguintes trabalhos e respectivas técnicas propostas:

- Zingarelli, Andrade e Goularte (2011), Zingarelli (2013), a técnica RevGlyph descrita na Seção 2.6.1;
- Joulin e Kang (2013), a técnica de recuperação de par estéreo descrita na Seção 2.6.2;
- Williem, Raskar e Park (2015), o algoritmo DME descrito na Seção 2.6.3;
- Rodrigues, Yugoshi e Goularte (2016), Rodrigues (2016), a técnica HaaRGlyph descrita na Seção 2.6.4.

Os resultados obtidos de PSNR desses trabalhos e da técnica proposta nesta dissertação são apresentados na Tabela 3. Na primeira coluna são apresentados os *datasets* de imagens, na segunda coluna, os trabalhos do estado da arte e na última coluna, os valores médios de PSNR de cada dataset de imagens por trabalho relacionado.

Tabela 3 – Resultados de PSNR para cada técnica em cada Dataset

Dataset	Trabalhos	PSNR
Andrade (2012)	Zingarelli (2013)	38,97
	Rodrigues (2016)	39,26
	Técnica proposta	31,18
Scharstein e Szeliski (2002)	Joulin e Kang (2013)	27,14
	Williem, Raskar e Park (2015)	31,46
	Técnica proposta	31,28

Fonte: Elaborada pelo autor.

Para o dataset de [Andrade \(2012\)](#), conforme apresentado Tabela 3, é possível observar que a técnica de [Rodrigues \(2016\)](#) obteve o maior valor de PSNR ($39,26\text{ dB}$) com uma diferença de $0,29\text{ dB}$ a mais que a técnica de [Zingarelli \(2013\)](#) ($38,97\text{ dB}$) e com $8,08\text{ dB}$ a mais que técnica proposta ($31,18\text{ dB}$). Essa diferença de valores de PSNR em relação a técnica proposta já era esperada, pois conforme apresentado na Seção 2.6, as técnicas de [Zingarelli \(2013\)](#) e [Rodrigues \(2016\)](#), na etapa de conversão anaglífica, armazenam e utilizam informações adicionais do par estéreo original para recuperar o par estéreo no final do processo.

A técnica proposta tem a característica de gerar o par estéreo apenas com as informações do anáglifo de entrada, isto é, sem informações adicionais. Por essa característica ser a mesma nos trabalhos relacionados que utilizaram o dataset de [Scharstein e Szeliski \(2002\)](#) (Tabela 3), elas podem ser consideradas mais próximas a técnica proposta.

Conforme apresentado na Tabela 3 o algoritmo proposto por [Williem, Raskar e Park \(2015\)](#) obteve o maior valor de PSNR ($31,46\text{ db}$) com uma diferença de $0,18\text{ dB}$ em relação a técnica proposta ($31,28\text{ db}$) e com $4,32\text{ dB}$ a mais que a técnica de [Joulin e Kang \(2013\)](#) ($27,14\text{ db}$). A técnica proposta obteve um PSNR $4,14\text{ dB}$ maior que a técnica de [Joulin e Kang \(2013\)](#). Apesar de não ter obtido um PSNR superior a técnica de [Williem, Raskar e Park \(2015\)](#), a técnica proposta tem o diferencial de utilizar recursos mais simples e baratos computacionalmente (apresentados no Capítulo 3) que os trabalhos relacionados a esse dataset, conforme as Seções 2.6.2 e 2.6.3.

No entanto, como já apresentado na Seção 2.5, o PSNR não leva em consideração a percepção visual humana e portanto o valor obtido não está relacionado com a percepção visual da imagem ([EBRAHIMI; CHAMIK; WINKLER, 2004](#)). Assim, uma imagem com valor de PSNR baixo não implica necessariamente a uma baixa qualidade quando visualizada. Porém, essa medida é bem sucedida para avaliar a quantidade de ruído inserido em imagens processadas.

Assim, pelo fato de ser uma métrica simples e escalável, o PSNR foi muito utilizado em trabalhos relacionados, servindo como base de comparação entre as técnicas.

4.4 Avaliação subjetiva de imagem

Nesta seção é apresentada a avaliação da qualidade subjetiva de imagem dos pares estéreos gerados pela técnica proposta. Para isso, os pares estéreos gerados foram avaliados utilizando o método de teste *Double Stimulus Continuous Quality Scale* (DSCQS) juntamente com o *Mean Opinion Score* (MOS) ([RECOMMENDATION, 1999](#); [RECOMMENDATION, 2002](#)). Esses métodos foram apresentados com detalhes na Seção [2.5.1](#).

No método DSCQS são apresentados ao observador múltiplas sequências de pares de vídeos ou imagens. Esses pares são compostos pelas imagens teste (par estéreo gerado) e referência (par estéreo original). Nesse método, a ordem de exibição das imagens, referência e teste, pode ser apresentada de duas maneiras. Na primeira o observador não é informado sobre os tipos de imagens e a ordem de apresentação é aleatória. Na segunda o observador é informado dos tipos de imagens e da ordem de apresentação. O MOS representa a média das pontuações dadas pelos observadores na avaliação de qualidade subjetiva de vídeo ou imagem. Estas notas são representadas por um único número que varia de 1 a 5. Assim, a pontuação 5 é equivalente ao conceito de "Excelente", isto é, nenhum defeito foi percebido pelo observador. A pontuação 4, "Bom", o defeito foi percebido e não causou desconforto. A pontuação 3, "Aceitável", o defeito foi percebido e causou desconforto. A pontuação 2, "Ruim", apesar da grande degradação na imagem o observador conseguiu visualizar alguma informação. E por fim, a pontuação 1, , a imagem tornou-se ininteligível e o observador ficou impossibilitado de extrair alguma informação.

A avaliação experimental foi realizada em 2 fases e contou com a participação de 83 pessoas de ambos os sexos, com idades de 18 a 50 anos e sem experiência prévia com avaliação de imagens. Na primeira fase, as imagens foram avaliadas por 48 pessoas, divididas em cinco grupos. E na segunda fase, por 35 pessoas divididas em quatro grupos.

Os *datasets* de testes utilizados no DSCQS continham um total 128 imagens, sendo 64 originais do dataset de [Andrade \(2012\)](#) e 64 recuperadas pela técnica proposta. Em cada avaliação experimental foi utilizado um dataset de imagens com 64 imagens, sendo 32 originais e 32 recuperadas. Foram também utilizadas 8 imagens para o treinamento dos participantes, essas imagens não faziam parte dos *datasets* de avaliação.

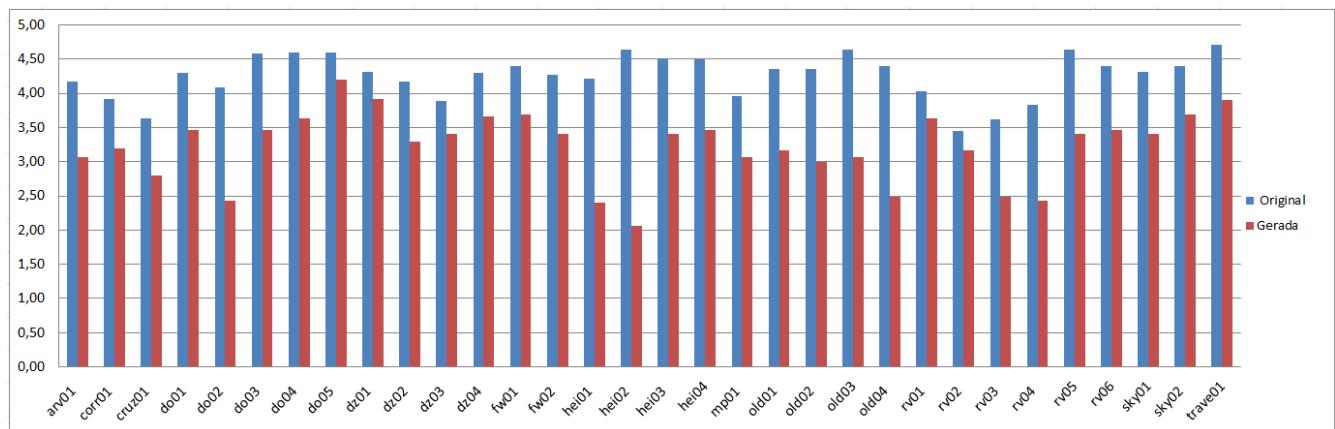
Os experimentos levaram em média 20 minutos para serem realizados com cada grupo de avaliadores, do início ao fim, incluindo o tempo de treinamento. O ambiente dos experimentos era composto por uma sala de 5 x 5 m, dez carteiras escolares dispostas em linha paralela em relação a parede a uma distância de 4m da tela branca, um projetor multimídia com 3.000 lumens de potência com resolução *Full HD* (1024 x 768), as imagens foram projetadas em tela branca

medindo 406 x 304 cm com 200 in de diagonal.

[Andrade \(2012\)](#), após análises subjetivas com vídeos estereoscópicos, indicou o valor 3,5 de MOS como sendo crítico, isto é, os vídeos analisados com valores MOS inferiores a esse interferiram na percepção de profundidade. Nos gráficos das Figuras 37 e 39 são apresentados os resultados de MOS obtidos nas duas fases de avaliação experimental realizadas com os pares estéreos originais e gerados pela técnica proposta.

No gráfico da Figura 37, são apresentados os valores médios de MOS, obtidos na primeira fase da avaliação experimental subjetiva para cada um dos 32 pares de imagens, originais e recuperadas, que compõem o dataset de teste. No eixo horizontal são apresentados os nomes das imagens e no vertical os valores da pontuação MOS. Nessa primeira fase, as imagens originais e recuperadas foram exibidas em ordem aleatória e os avaliadores não foram informados da existência desses dois tipos de imagens e nem da ordem da apresentação, conforme recomendado em ITU-R BT.500-11 ([BT, 2002](#)) e ITU-T P.910 ([RECOMMENDATION, 1999](#)). Nesse contexto, foram pontuadas tanto as imagens originais quanto as recuperadas.

Figura 37 – Resultados de DSCQS e MOS - Fase 1

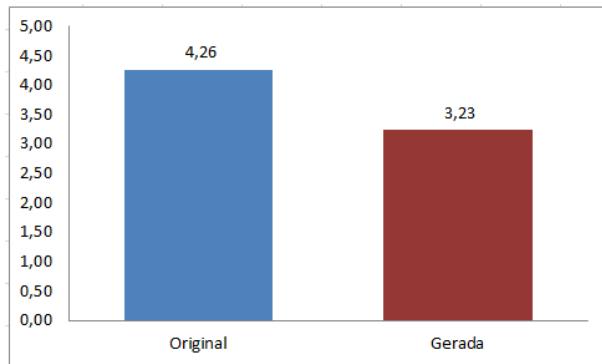


Fonte: Elaborada pelo autor.

Como pode ser observado no gráfico da Figura 37, para o par estéreo original, em azul, existem 31 imagens com as pontuações acima de 3,5, uma imagem com pontuação na faixa de 3,0 a 3,5 e nenhuma imagem com pontuação abaixo de 3,0. Com base nesses valores foi possível avaliar, também, a qualidade subjetiva do dataset de imagens originais de [Andrade \(2012\)](#). Esses valores juntamente com a média MOS de 4,26 (Figura 38), indicam que a base original possui imagens com "Excelente" qualidade visual.

Na análise qualitativa do par estéreo recuperado, em vermelho no gráfico da Figura 37, foi observado que existem 8 imagens com pontuação acima de 3,5, 17 imagens com pontuação na faixa de 3,0 a 3,5 e 07 imagens com pontuação abaixo de 3,0. Esses valores, juntamente com a média MOS de 3,23 (Figura 38), indicam que a base recuperada possui imagens com "Boa" qualidade visual.

Figura 38 – Média MOS - Fase 1



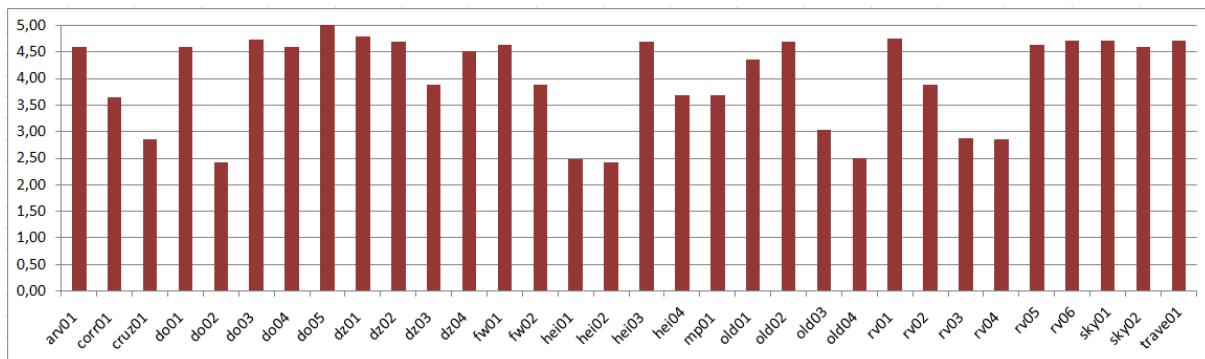
Fonte: Elaborada pelo autor.

Também foi observada uma diferença entre os pares originais e recuperados nas pontuações tanto par a par (Figura 37) quanto na média (Figura 38). Essa diferença, que em 26 pares comparados foi menor que 1,5, pode ter sido causada pelo fato da base de imagens originais também ter sido avaliada.

Para avaliar a qualidade subjetiva de imagem do par estéreo recuperado tendo o par estéreo original como referência foi realizada a segunda fase de avaliação experimental. Nessa segunda fase de experimentos, também seguindo as recomendações de ITU-R BT.500-11 (BT, 2002) e ITU-T P.910 (RECOMMENDATION, 1999), as imagens originais foram exibidas antes das recuperadas e os avaliadores foram instruídos da ordem de apresentação e dos tipos de imagens. Nesse contexto, somente os pares recuperados foram pontuados.

No gráfico da Figura 39, são apresentados os valores médios de MOS, obtidos na segunda fase da avaliação experimental subjetiva para cada um dos 32 pares de imagens, originais e recuperadas, que compõem o dataset de teste. No eixo horizontal são apresentados os nomes das imagens e no vertical os valores da pontuação MOS.

Figura 39 – Resultados de DSCQS e MOS - Fase 2



Fonte: Elaborada pelo autor.

Na análise qualitativa do par estéreo recuperado, Figura 39, foi observado que existem 24 imagens com pontuação acima de 3,5, uma imagem com pontuação na faixa de 3,0 a 3,5 e 07 imagens com pontuação abaixo de 3,0. Esses valores juntamente com a média MOS dessa segunda fase igual a 4,00, indicam que a base recuperada possui imagens com "Boa" qualidade visual.

Ao comparar a pontuação dos pares estéreos recuperados, entre as duas fases de avaliação experimental, na segunda fase houve um aumento de 0,77 na média MOS. Foi observado que para a pontuação acima de 3,5, houve um acréscimo de 16 imagens totalizando 24 imagens. Para o intervalo de 3,0 a 3,5 houve uma redução de 16 imagens ficando com apenas uma e para a pontuação abaixo de 3,0 a quantidade foi de 07 imagens. Embora a média MOS seja um pouco maior na segunda fase a qualidade visual dos pares estéreos recuperados continua sendo a mesma da primeira, isto é de "Boa" qualidade. Ainda comparando as duas fases, é importante observar que a diferença na pontuação MOS entre os pares originais e recuperados foi mantida. Na primeira fase a diferença foi de 1,03 e na segunda foi de 1,00. Assim, com base nos resultados de média MOS e diferença de pontuação MOS é possível concluir que os resultados da segunda fase reforçam os resultados da primeira fase.

Rodrigues, Yugoshi e Goularte (2016) com a técnica HaaRGlyph, também realizaram testes subjetivos com o *dataset* de Andrade (2012). A HaaRGlyph alcançou a pontuação média MOS de 3,30 e a técnica ARBFLS proposta, na primeira fase de avaliação experimental, obteve a pontuação média MOS de 3,23. Ambas são consideradas de "Boa" qualidade visual, porém a HaaRGlyph utiliza informações extras para recuperar uma aproximação do par estéreo original.

4.5 Considerações Finais

Neste capítulo foram apresentados os experimentos para avaliar a solução proposta para reversão anaglífica. Foram realizadas dois tipos avaliações: objetiva e subjetiva. Na avaliação objetiva constatou-se que a solução proposta obteve resultados competitivos com as técnicas do estado da arte, técnica proposta tem o diferencial de utilizar recursos mais simples e baratos computacionalmente (apresentados no Capítulo 3) que os trabalhos relacionados ao *Middlebury dataset* (SCHARSTEIN; SZELISKI, 2002), apresentados as Seções 2.6.2 e 2.6.3. Na avaliação subjetiva foi constatado que a técnica proposta gerou pares estéreos com boa qualidade visual, comparados aos pares estéreos originais.



CONCLUSÕES

Nesta dissertação de mestrado, foi proposta e desenvolvida uma nova técnica de reversão anaglífica para recuperar uma aproximação do par estéreo original. O objetivo da técnica foi investigar a possibilidade de recuperar o par estéreo original apenas com as informações intracodificadas no anáglifo sem perdas significativas na qualidade de imagem. Para isso, a técnica proposta utiliza a correspondência entre blocos (*block matching*) em busca local rápida para recuperar as informações que ainda estão presentes no anáglifo e, em seguida, transferir essas informações para recuperar os canais de cores faltantes a fim de uma aproximação do par estéreo original.

A técnica proposta, denominada ARBFLS, foi testada experimentalmente em dois *datasets* de imagens estéreo utilizados por trabalhos relacionados na literatura. Os resultados obtidos pela técnica proposta para o primeiro dataset indicaram uma vantagem na qualidade objetiva para as duas técnicas anteriores ([ZINGARELLI, 2013](#); [RODRIGUES, 2016](#)) de aproximadamente 8,00 dB. Essa vantagem já era esperada, pois essas técnicas armazenam informações do par estéreo original e as utilizam no processo de reversão juntamente com o anáglifo. No entanto, ainda para esse dataset, para avaliar a qualidade visual de imagem, os pares estéreos recuperados foram submetidos à avaliação de grupos de usuários. Os resultados dessa avaliação subjetiva demonstraram que as imagens recuperadas têm “Boa” qualidade visual em relação ao par estéreo original (média MOS variando entre 3 e 4). Já para o segundo dataset a técnica proposta obteve vantagem de aproximadamente 4,14 dB em relação a técnica de [Joulin e Kang \(2013\)](#) e empate técnico com a outra técnica proposta por [Williem, Raskar e Park \(2015\)](#), aproximadamente 0,19 dB a menos. Assim, constatou-se que a solução proposta obteve resultados competitivos comparados às técnicas dos trabalhos relacionados. Sendo importante destacar que a técnica proposta é mais simples e mais barata computacionalmente que as demais, conforme apresentado na Seção 4.3.

5.1 Contribuições Científicas

A principal contribuição deste trabalho é a nova técnica ARBFLS para reconstruir uma aproximação do par estéreo original à partir de um anáglico baseada em busca local rápida. Também podem ser consideradas contribuições a ferramenta desenvolvida como implementação da técnica ARBFLS, essa ferramenta possibilita que novas funcionalidades sejam facilmente acopladas e a implementação da busca local rápida para imagens anaglíficas. Outras contribuições são dois artigos científicos:

RODRIGUES, F. M.; YUGOSHI, J. K.; GOULARTE, R. Haarglyph: A new method for anaglyphic reversion in stereoscopic videos. In: ACM Proceedings of the 22nd Brazilian Symposium on Multimedia and the Web. [S.I.], 2016. p. 151–158.

YUGOSHI, J. K.; RODRIGUES, F. M.; GOULARTE, R. Anaglyphic Reversal Based on Fast Local Search. [a ser submetido]

O primeiro artigo, desenvolvido em colaboração para a técnica HaaRGlyph já foi publicado em conferência nacional relevante da área. O segundo artigo a ser submetido a um periódico internacional relevante para essa área é referente a técnica ARBFLS com as respectivas avaliações experimentais. Por fim, a formação de recursos humanos qualificados, em nível de mestrado e de iniciação científica, também pode ser citada como uma das contribuições desta dissertação.

5.2 Limitações e trabalhos futuros

A técnica ARBFLS atingiu bons resultados em comparação aos trabalhos do estado da arte, principalmente pelo fato de ser simples e barata computacionalmente. No entanto, não soluciona definitivamente o problema de recuperar uma aproximação do par estéreo e pontos de melhoria podem ser trabalhados. A técnica proposta não funciona bem com camadas sobrepostas, grandes desoclusões, regiões oclusas e com informações de cores inexistentes.

Uma limitação está em utilizar o detector de bordas Canny, aplicado nos canais fonte e alvo do anáglico antes da etapa de correspondência entre blocos que converte as imagens em binárias. Isto é, para as bordas dos elementos de imagem são atribuídos o valor “1” (“255” em 8 bits) e “0” para o restante da imagem. Um trabalho futuro é realizar a conversão dos canais fonte e alvo para tons de cinza para que seja possível melhorar a precisão da correspondência entre blocos.

Uma outra limitação é que as imagens foram pré-ajustadas nas dimensões planares em aproximadamente 30 pixels a menos nas bordas. Um trabalho futuro é aplicar técnicas de colorização que utilize as cores vizinhas para preenchimento de bordas.

Um outro direcionamento para trabalhos futuros é a exploração de outras técnicas para encontrar a correspondência entre os canais preservados do anáglifo, como, por exemplo, algoritmos de correspondência estéreo baseados em otimização, *deep learning* e correlação de cores. Também é um trabalho futuro a investigação dessas técnicas para identificar as mais precisas em detectar regiões e elementos de imagens oclusos.

Por fim, é importante ressaltar que a reversão anaglífica para recuperar uma aproximação do par estéreo original é um tema relevante que ainda apresenta questões em aberto, possui poucas ferramentas implementadas e poucos *datasets* de imagens de *baseline* disponíveis. Nesta dissertação, foram realizadas investigações, desenvolvimento de uma nova técnica e avaliações dos resultados.

REFERÊNCIAS

- ALPERT, T.; BARONCINI, V.; CHOI, D.; CONTIN, L.; KOENEN, R.; PEREIRA, F.; PETERSON, H. Subjective evaluation of mpeg-4 video codec proposals: Methodological approach and test procedures. **Signal Processing: Image Communication**, v. 9, n. 4, p. 305 – 325, 1997. ISSN 0923-5965. MPEG-4, Part 1: Invited Papers. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0923596597000040>>. Citado na página 40.
- ANDRADE, L. **Compressão espacial de vídeos estereoscópicos: Uma abordagem baseada em codificação anaglífica.** 2012. Tese (Doutorado) — Tese de doutorado em ciências da Computação e Matemática Computacional). Universidade de São Paulo, USP, São Carlos, 2012. Citado nas páginas 61, 62, 63, 64, 65 e 67.
- ANDRADE, L.; GOULARTE, R. Uma análise da influência da subamostragem de crominância em vídeos estereoscópicos anaglíficos. In: **Proceedings of the Webmedia–Brazilian Symposium on Multimedia and the Web**. Belo Horizonte, MG, Brasil: ACM, 2010. Citado nas páginas 20 e 31.
- ANDRADE, L. A.; GOULARTE, R. Anaglyphic stereoscopic perception on lossy compressed digital videos. In: **Proceedings of the XV Brazilian Symposium on Multimedia and the Web**. New York, NY, USA: ACM, 2009. (WebMedia '09), p. 29:1–29:8. ISBN 978-1-60558-880-3. Disponível em: <<http://doi.acm.org/10.1145/1858477.1858506>>. Citado na página 20.
- AZEVEDO, E.; CONCI, A. **Computação gráfica: teoria e prática**. São Paulo, SP, Brasil: Elsevier, 2003. Citado nas páginas 19 e 23.
- BHASKARAN, V.; KONSTANTINIDES, K. **Image and Video Compression Standards: Algorithms and Architectures**. 2nd. ed. Norwell, MA, USA: Kluwer Academic Publishers, 1997. ISBN 0792399528. Citado na página 21.
- BOYKOV, Y.; VEKSLER, O.; ZABIH, R. Fast approximate energy minimization via graph cuts. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 23, n. 11, p. 1222–1239, Nov 2001. ISSN 0162-8828. Citado nas páginas 45 e 51.
- BROWN, M. Z.; BURSCHKA, D.; HAGER, G. D. Advances in computational stereo. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 25, n. 8, p. 993–1008, Aug 2003. ISSN 0162-8828. Citado na página 21.
- BT, R. I.-R. Methodology for the subjective assessment of the quality of television pictures. 2002. Citado nas páginas 39, 40, 65 e 66.
- CANNY, J. A computational approach to edge detection. **IEEE Transactions on pattern analysis and machine intelligence**, Ieee, n. 6, p. 679–698, 1986. Citado na página 54.
- CISCO, V. **Cisco Visual Networking Index: Forecast and Methodology 2016–2021.(2017)**. 2017. Citado na página 19.

DODGSON, N. A. **Analysis of the viewing zone of multiview autostereoscopic displays.** 2002. 4660 - 4660 - 12 p. Disponível em: <<https://doi.org/10.1117/12.468040>>. Citado na página 36.

EBRAHIMI, F.; CHAMIK, M.; WINKLER, S. Jpeg vs. jpeg 2000: an objective comparison of image encoding quality. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. **Applications of Digital Image Processing XXVII.** Denver, Colorado, USA, 2004. v. 5558, p. 300–309. Citado na página 63.

FEHN, C.; KAUFF, P.; BEECK, M. O. D.; ERNST, F.; IJSSELSTEIJN, W.; POLLEFEYS, M.; GOOL, L. V.; OFEK, E.; SEXTON, I. An evolutionary and optimised approach on 3d-tv. v. 2, p. 357–365, 2002. Citado na página 20.

GENG, J. Three-dimensional display technologies. **Advances in optics and photonics**, Optical Society of America, v. 5, n. 4, p. 456–535, 2013. Citado nas páginas 31, 32, 33 e 34.

GOLDSTEIN, E. B. **Sensation and perception.** Belmont, CA, USA: Wadsworth Publishing Co, 2014. Citado nas páginas 23, 25, 26, 27 e 28.

GOTCHEV, A.; AKAR, G. B.; CAPIN, T.; STROHMEIER, D.; BOEV, A. Three-dimensional media for mobile devices. **Proceedings of the IEEE**, IEEE, v. 99, n. 4, p. 708–741, 2011. Citado nas páginas 28 e 35.

HANNAH, M. J. **Computer Matching of Areas in Stereo Images.** Tese (Doutorado), Stanford, CA, USA, 1974. AAI7427032. Citado na página 21.

HARSTEAD, E.; SHARPE, R. Forecasting of access network bandwidth demands for aggregated subscribers using monte carlo methods. **IEEE Communications Magazine**, IEEE, v. 53, n. 3, p. 199–207, 2015. Citado na página 19.

HONG, J.; KIM, Y.; CHOI, H.-J.; HAHN, J.; PARK, J.-H.; KIM, H.; MIN, S.-W.; CHEN, N.; LEE, B. Three-dimensional display technologies of recent interest: principles, status, and issues. **Applied optics**, Optical Society of America, v. 50, n. 34, p. H87–H115, 2011. Citado na página 30.

HUTCHISON, D. 3-d tv. **White Paper**, 2008. Citado na página 35.

ISO. ISO, **Representation of Auxiliary Vide and Supplemental Information - Doc N8768 - akech - ocos.** 2007. Citado na página 38.

_____. ISO, **Overview of 3D Video Coding - Doc N9784 - Archamp - França.** 2008. Citado na página 38.

ITUT. recommendation and final draft international standard of joint video specification (itu-t rec. h. 264| iso/iec 14496-10 avc). **Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, JVTG050**, v. 33, 2003. Citado na página 20.

JOULIN, A.; KANG, S. B. Recovering Stereo Pairs from Anaglyphs. In: **CVPR 2013 - IEEE Conference on Computer Vision and Pattern Recognition.** Portland, Oregon, United States: IEEE, 2013. p. 289–296. Disponível em: <<https://hal.inria.fr/hal-01064225>>. Citado nas páginas 44, 48, 51, 61, 62, 63 e 69.

KAUFMAN, L. **Sight and mind: An introduction to visual perception.** Oxford, England: Oxford U. Press, 1974. Citado na página 19.

- KIM, N.; PHAN, A.-H.; ERDENEBAT, M.-U.; ALAM, M.; KWON, K.-C.; PIAO, M.-L.; LEE, J.-H. 3d display technology. **Disp. Imag.**, v. 1, p. 73–95, 2014. Citado na página 29.
- LEBRETON, P.; BARKOWSKY, M.; RAAKE, A.; CALLET, P. L. **3D Video**. Cham: Springer International Publishing, 2014. 299–313 p. Disponível em: <https://doi.org/10.1007/978-3-319-02681-7_20>. Citado na página 28.
- LEVIN, A.; LISCHINSKI, D.; WEISS, Y. Colorization using optimization. In: **ACM SIGGRAPH 2004 Papers**. New York, NY, USA: ACM, 2004. (SIGGRAPH '04), p. 689–694. Disponível em: <<http://doi.acm.org/10.1145/1186562.1015780>>. Citado na página 45.
- LIPTON, L. Foundations of the stereo-scopic cinema a study in depth, 1982. **Von Nostrand Reinhold Company, ed, ISBN: 0-442-24724-9**. Citado na página 20.
- _____. Stereo-vision formats for video and computer graphics. In: **Stereoscopic Displays and Virtual Reality Systems IV**. San Jose, CA, USA: Fisher, S. S. and Merritt, J. O. and Bolas, M. T., 1997. (, v. 3012), p. 239–244. Citado nas páginas 19, 20, 24, 33 e 37.
- LIU, C.; YUEN, J.; TORRALBA, A.; SIVIC, J.; FREEMAN, W. T. Sift flow: Dense correspondence across different scenes. In: FORSYTH, D.; TORR, P.; ZISSEMAN, A. (Ed.). **Computer Vision – ECCV 2008**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. p. 28–42. ISBN 978-3-540-88690-7. Citado nas páginas 44 e 45.
- MARR, D. **Vision: A computational investigation into the human representation and processing of visual information**. MIT Press. Cambridge, Massachusetts, USA: MIT press. Citado na página 21.
- MENDIBURU, B. 3d movie making stereoscopic digital cinema from script to screen. Elsevier, Inc, 2009. Citado nas páginas 32 e 33.
- MERKLE, P.; SMOLIĆ, A.; MÜLLER, K.; WIEGAND, T. Efficient prediction structures for multiview video coding. **IEEE Transactions on Circuits and Systems for Video Technology**, IEEE, v. 17, n. 11, p. 1461–1473, 2007. Citado na página 38.
- MULLER, K.; MERKLE, P.; WIEGAND, T. 3-d video representation using depth maps. **Proceedings of the IEEE**, IEEE, v. 99, n. 4, p. 643–656, 2011. Citado na página 38.
- PINSON, M. H.; WOLF, S.; GALLAGHER, M. D. The impact of monitor resolution and type on subjective video quality testing. Citeseer, 2004. Citado na página 39.
- RECOMMENDATION, I. 709-5, parameter values for the hdtv standards for production and international programme exchange. **ITU Radiocommunication**, 2002. Citado nas páginas 22 e 64.
- RECOMMENDATION, P. I.-T. Subjective video quality assessment methods for multimedia applications. 1999, 1999. Citado nas páginas 22, 39, 64, 65 e 66.
- RICHARDSON, H. **IEG 264 and MPEG-4 video compression: video coding for next-generation multimedia**. Chichester, West Sussex, England: John Wiley & Sons, 2003. Citado na página 33.
- RICHARDSON, I. E. **H. 264 and MPEG-4 video compression: video coding for next-generation multimedia**. Chichester, West Sussex, England: John Wiley & Sons, 2004. Citado na página 54.

RODRIGUES, F. M. **Reversão anaglífica em vídeos estereoscópicos**. Dissertação (Mestrado) — Universidade de São Paulo, 2016. Citado nas páginas 48, 61, 62, 63 e 69.

RODRIGUES, F. M.; YUGOSHI, J. K.; GOULARTE, R. Haarglyph: A new method for anaglyphic reversion in stereoscopic videos. In: **Proceedings of the 22Nd Brazilian Symposium on Multimedia and the Web**. New York, NY, USA: ACM, 2016. (Webmedia '16), p. 151–158. ISBN 978-1-4503-4512-5. Disponível em: <<http://doi.acm.org/10.1145/2976796.2976864>>. Citado nas páginas 46, 47, 51, 62 e 67.

SCHARSTEIN, D.; SZELISKI, R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. **International journal of computer vision**, Springer, v. 47, n. 1-3, p. 7–42, 2002. Citado nas páginas 61, 62, 63 e 67.

SMOLIC, A.; MUELLER, K.; MERKLE, P.; KAUFF, P.; WIEGAND, T. An overview of available and emerging 3d video formats and depth enhanced stereo as efficient generic solution. In: IEEE. **Picture Coding Symposium 2009**. Chicago, IL, USA, 2009. p. 1–4. ISBN 978-1-4244-4593-6, 978-1-4244-4594-3. Disponível em: <<https://doi.org/10.1109/PCS.2009.5167358>>. Citado nas páginas 20, 36, 37 e 38.

SUTHERLAND, I. E. A head-mounted three dimensional display. In: **Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I**. New York, NY, USA: ACM, 1968. (AFIPS '68 (Fall, part I)), p. 757–764. Disponível em: <<http://doi.acm.org/10.1145/1476589.1476686>>. Citado na página 33.

TAKAKI, Y. Development of super multi-view displays. **ITE Transactions on Media Technology and Applications**, The Institute of Image Information and Television Engineers, v. 2, n. 1, p. 8–14, 2014. Citado na página 36.

TAM, W. J.; ZHANG, L. 3d-tv content generation: 2d-to-3d conversion. In: IEEE. **International Conference on Multimedia and Expo**. Toronto, Ontario, Canada, 2006. p. 1869–1872. Citado na página 20.

UREY, H.; CHELLAPPAN, K. V.; ERDEN, E.; SURMAN, P. State of the art in stereoscopic and autostereoscopic displays. **Proceedings of the IEEE**, IEEE, v. 99, n. 4, p. 540–555, 2011. ISSN 0018-9219. Citado nas páginas 29 e 33.

VETRO, A. Representation and coding formats for stereo and multiview video. In: **Intelligent Multimedia Communication: Techniques and Applications**. Springer, Berlin, Heidelberg: Springer, 2010. p. 51–73. ISBN 978-3-642-11686-5, 978-3-642-11685-8. Disponível em: <https://doi.org/10.1007/978-3-642-11686-5_2>. Citado na página 37.

WANG, Z.; SHEIKH, H. R.; BOVIK, A. C. *et al.* Objective video quality assessment. **The handbook of video databases: design and applications**, CRC Press, v. 41, p. 1041–1078, 2003. Citado na página 41.

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**, JSTOR, v. 128, p. 371–394, 1838. Citado nas páginas 24 e 29.

WILLIEM; RASKAR, R.; PARK, I. K. Depth map estimation and colorization of anaglyph images using local color prior and reverse intensity distribution. In: **Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)**. Washington, DC, USA: IEEE Computer Society, 2015. (ICCV '15), p. 3460–3468. ISBN 978-1-4673-8391-2. Disponível em:

<<http://dx.doi.org/10.1109/ICCV.2015.395>>. Citado nas páginas 13, 45, 46, 48, 51, 61, 62, 63 e 69.

WINKLER, S. **Digital video quality: vision models and metrics**. Chichester, West Sussex, England: John Wiley & Sons, 2005. Citado nas páginas 38, 39, 40 e 41.

ZINGARELLI, M. R. U. **RevGlyph-codificação e reversão esteroscópica anaglífica**. Dissertação (Mestrado) — Universidade de São Paulo, 2013. Citado nas páginas 48, 61, 62, 63 e 69.

ZINGARELLI, M. R. U.; ANDRADE, L. A. de; GOULARTE, R. Reversing anaglyph videos into stereo pairs. In: **Proceedings of the 17th Brazilian Symposium on Multimedia and the Web on Brazilian Symposium on Multimedia and the Web - Volume 1**. Porto Alegre, Brazil, Brazil: Brazilian Computer Society, 2011. (WebMedia 2011), p. 27:205–27:212. Disponível em: <<http://dl.acm.org/citation.cfm?id=3021508.3021541>>. Citado nas páginas 21, 30, 31 e 62.

_____. Revglyph: A technique for reverting anaglyph stereoscopic videos. In: **Proceedings of the 27th Annual ACM Symposium on Applied Computing**. New York, NY, USA: ACM, 2012. (SAC '12), p. 1005–1011. ISBN 978-1-4503-0857-1. Disponível em: <<http://doi.acm.org/10.1145/2245276.2245470>>. Citado nas páginas 25, 42, 43 e 51.

