UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Reversão anaglífica - um novo método baseado em cálculo de correspondências robusto a diferenças radiométricas

André Luiz Dutra Costa

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: _____

André Luiz Dutra Costa

Reversão anaglífica - um novo método baseado em cálculo de correspondências robusto a diferenças radiométricas

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
Outubro de 2021

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

C837r Costa, André Luiz Dutra
Reversão anaglífica - um novo método baseado em
cálculo de correspondências robusto a diferenças
radiométricas / André Luiz Dutra Costa; orientador
Rudinei Goularte. -- São Carlos, 2021.
110 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, 2021.

1. Reversão Anaglífica. 2. Estereoscopia. 3.
Anáglifo. I. Goularte, Rudinei, orient. II. Título.

André Luiz Dutra Costa

**Anaglyph reversion - a new method based on stereo
matching computation robust to radiometric differences**

Dissertation submitted to the Instituto de Ciências
Matemáticas e de Computação – ICMC-USP – in
accordance with the requirements of the Computer
and Mathematical Sciences Graduate Program, for
the degree of Master in Science. *FINAL VERSION*

Concentration Area: Computer Science and
Computational Mathematics

Advisor: Prof. Dr. Rudinei Goularte

USP – São Carlos
October 2021

"Eu não quero acreditar, eu quero saber."

Carl Sagan

AGRADECIMENTOS

Agradeço, primeiramente, à minha esposa e amiga, Bárbara, por ter me incentivado a ingressar neste desafio e por não me deixar desistir. Também por seu amor, sua companhia e por me fazer querer ser cada vez melhor.

Aos meus pais, Joselis e Antônio, pelo amor, pelos valores, pela liberdade, pelo apoio e por todos os esforços que fizeram pela nossa família.

Aos meus irmãos, Gustavo, Laís, Fábio, e à minha sobrinha, Alice, que são fonte constante de inspiração, força, alegria, cumplicidade e parceria.

À minha segunda família, Bruno, Camila, Emne, Flávia, Fred e Letícia, pela amizade, pelo companheirismo e por dividirem tantos momentos felizes comigo.

À minha cunhada, Fernanda Monedeiro, pela inspiração e pela contribuição com a análise estatística apresentada no Capítulo 5.

Ao meu orientador, Rudinei Goularte, pela oportunidade, pela confiança, por me guiar na execução deste mestrado e pelo tempo e esforço investidos.

Aos docentes e funcionários técnicos e administrativos tanto do ICMC quanto da USP como um todo, que tornam esta Universidade tão grandiosa e motivo de orgulho.

RESUMO

COSTA, A. L. D. **Reversão anaglífica - um novo método baseado em cálculo de correspondências robusto a diferenças radiométricas**. 2021. 106 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, 2021.

O processo de recuperação de um par estéreo a partir de um anáglifo, chamado de *reversão anaglífica*, promove alguns benefícios como: a visualização de imagens e vídeos anaglíficos legados, a independência de dispositivos de visualização, e a economia de volume de dados para fins de armazenamento e transmissão de conteúdo estéreo. Contudo, as técnicas de reversão anaglífica necessitam comparar os canais de cor do anáglifo buscando por correspondências a fim restaurar os canais eliminados durante a codificação anaglífica, o quê é uma tarefa difícil devido às diferenças radiométricas presentes nesses canais. Os trabalhos relacionados, até onde se tem conhecimento, não oferecem meios especificamente projetados para lidar com essas diferenças, limitando a qualidade final das imagens recuperadas. Assim, este trabalho propõe uma nova técnica de reversão anaglífica que realiza os cálculos de correspondência estéreo por meio de um descritor de pixel robusto às variações radiométricas existentes entre os canais do anáglifo. Testes objetivos e subjetivos realizados indicam que a técnica proposta supera em qualidade os trabalhos considerados estado-da-arte com custo computacional compatível.

Palavras-chave: Reversão Anaglífica, Estereoscopia, Anáglifo, Correspondência Estéreo.

ABSTRACT

COSTA, A. L. D. **Anaglyph reversion - a new method based on stereo matching computation robust to radiometric differences**. 2021. 106 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, 2021.

The recovering a stereo pair from an anaglyph, called *anaglyph reversion*, promotes some benefits such as viewing legacy anaglyphic images and videos, independence of viewing devices, and save data volume for stereo content storage and transmission purposes. However, anaglyphic reversion techniques need to compare the anaglyph color channels searching for matches in order to restore the channels eliminated during anaglyphic encoding, which is a difficult task due to the radiometric differences in these channels. Related works do not offer means specifically designed to deal with these differences, limiting the final quality of the recovered images. Thus, this work proposes a new anaglyphic reversion technique that performs stereo correspondence calculations using a pixel descriptor robust to the radiometric variations existing between the anaglyph channels. Objective and subjective tests performed indicate that the proposed technique exceeds in quality state-of-the-art works with similar computational cost.

Keywords: Anaglyph Reversion, Stereoscopy, Anaglyph, Stereo Matching.

LISTA DE ILUSTRAÇÕES

Figura 1 – Dica binocular e oculomotora <i>convergência</i>	24
Figura 2 – Dicas monoculares <i>oclusão, altura relativa, tamanho relativo e convergência perspectiva</i>	25
Figura 3 – Dica monocular <i>tamanho familiar</i>	26
Figura 4 – Dicas monoculares <i>perspectiva atmosférica e gradiente de textura</i>	26
Figura 5 – Dica monocular <i>sombras</i>	27
Figura 6 – Exemplo de par estéreo: imagens vistas por cada um dos olhos	28
Figura 7 – Estereoscópio de Wheatstone	29
Figura 8 – Métodos de visualização por multiplexação	30
Figura 9 – Dispositivos HMD	31
Figura 10 – Métodos de visualização auto estereoscópicos	32
Figura 11 – Problema de visualização <i>interferência</i>	32
Figura 12 – Problema de visualização <i>ghosting</i>	33
Figura 13 – Problema de visualização <i>rivalidade de retina</i>	33
Figura 14 – Processo de codificação anaglífica vermelho-ciano	34
Figura 15 – Esquemas de cores X Tipos de óculos	34
Figura 16 – Anáglifo vermelho-ciano codificado no modo <i>colorido</i>	35
Figura 17 – Anáglifo vermelho-ciano codificado no modo ' <i>verdadeiro</i> '	36
Figura 18 – Anáglifo vermelho-ciano codificado no modo <i>cinza</i>	37
Figura 19 – Anáglifo vermelho-ciano codificado no modo <i>semicolorido</i>	38
Figura 20 – Mapa de disparidades	39
Figura 21 – Par estéreo com diferentes intensidades	40
Figura 22 – Colorização de imagens a partir da transferência de cor	43
Figura 23 – Colorização a partir de anotações	43
Figura 24 – Esquema de avaliação do DSCQ	45
Figura 25 – Escala de avaliação do DSCQS	45
Figura 26 – Técnica proposta por Joulin e Kang (2013)	50
Figura 27 – Técnica proposta por Williem, Raskar e Park (2015)	51
Figura 28 – Técnica proposta por Yugoshi (2018)	52
Figura 29 – Técnica proposta por Kunze (2019)	53
Figura 30 – Separação e mistura dos canais do anáglifo	57
Figura 31 – Bloco de tamanho 3x3	59
Figura 32 – Blocos com diferenças radiométricas	60

Figura 33 – Secionamento da matriz de custo de correspondência	63
Figura 34 – Seção da matriz de correspondência com $d=25$	64
Figura 35 – Vetor de custos de correspondência agregados para o pixel $x=231, y=120$. .	65
Figura 36 – Mapas de disparidades	65
Figura 37 – Vetor de custos de correspondência para o pixel $x=84, y=243$	66
Figura 38 – Máscaras de reciprocidade	68
Figura 39 – Refinamento do mapa de disparidade da V_E	69
Figura 40 – Recuperação por correspondência direta	70
Figura 41 – Processo de colorização de um pixel	72
Figura 42 – Bordas das regiões válidas	74
Figura 43 – Par estéreo recuperado	74
Figura 44 – Mapas de disparidades e máscaras de oclusão	79
Figura 45 – Exemplo da identificação de regiões oclusas	80
Figura 46 – Mapas de texturas e máscaras de baixas texturas	81
Figura 47 – Exemplo da identificação de regiões de baixa textura ($\alpha = 3, b = 13$)	83
Figura 48 – Exemplo de diferenças radiométricas	84
Figura 49 – Mapas e máscaras de diferença radiométrica	85
Figura 50 – Exemplo da identificação de regiões de grandes diferenças radiométricas . .	86
Figura 51 – Análise subjetiva detalhada da visão direita da imagem <i>Cones</i>	97
Figura 52 – Análise subjetiva detalhada da visão direita da imagem <i>Teddy</i>	97

LISTA DE TABELAS

Tabela 1 – Métricas para correspondência	42
Tabela 2 – Comparação entre métodos de reversão anaglífica	53
Tabela 3 – Colorização incremental	75
Tabela 4 – Imagens selecionadas do <i>dataset Middlebury</i>	78
Tabela 5 – Identificação de oclusão	80
Tabela 6 – Identificação de textura ($\alpha = 1$, $b=13$)	83
Tabela 7 – Identificação de diferenças radiométricas ($\beta = 64$)	86
Tabela 8 – PSNR dos canais esquerdo e direito recuperados	87
Tabela 9 – PSNR médio entre as visões esquerda e direita	89
Tabela 10 – Teste de Shapiro-Wilk	89
Tabela 11 – Teste de Levene	90
Tabela 12 – Teste de t de Student	90
Tabela 13 – PSNR das visões esquerda e direita recuperadas por recuperação direta	91
Tabela 14 – PSNR médio entre as visões esquerda e direita recuperadas por recuperação direta	92
Tabela 15 – p-value para os testes de relevância	92
Tabela 16 – PSNR médio entre as visões esquerda e direita recuperadas por colorização	93
Tabela 17 – p-value para os testes de relevância	93
Tabela 18 – MOS das imagens originais e recuperadas	95
Tabela 19 – p-value para os testes de relevância - DSCQS/MOS	96

SUMÁRIO

1	INTRODUÇÃO	19
1.1	Contextualização	19
1.2	Organização do texto	22
2	FUNDAMENTOS TEÓRICOS	23
2.1	Visão humana	23
2.1.1	<i>Dicas oculomotoras</i>	24
2.1.2	<i>Dicas monoculares</i>	24
2.1.2.1	<i>Dicas monoculares pictoriais</i>	24
2.1.2.2	<i>Dicas monoculares de movimento</i>	27
2.1.3	<i>Dicas binoculares</i>	28
2.2	Estereoscopia	28
2.2.1	<i>Métodos de visualização estereoscópicos</i>	29
2.2.2	<i>Problemas de visualização</i>	32
2.2.3	<i>Método anaglífico</i>	33
2.2.3.1	<i>Anáglifos coloridos (Color anaglyphs)</i>	35
2.2.3.2	<i>Anáglifos 'verdadeiros' (True anaglyphs)</i>	36
2.2.3.3	<i>Anáglifos cinzas (Gray anaglyphs)</i>	36
2.2.3.4	<i>Anáglifos semicoloridos (Half color anaglyphs)</i>	37
2.3	Correspondência estéreo x Correspondência anaglífica	38
2.4	Colorização de imagens	42
2.5	Qualidade em imagens e vídeos digitais	44
2.5.1	<i>Avaliação subjetiva de qualidade</i>	44
2.5.2	<i>Avaliação objetiva de qualidade</i>	46
2.6	Considerações finais	47
3	REVISÃO DE LITERATURA	49
3.1	Técnicas baseadas em informações internas ao anáglifo	49
3.2	Considerações finais	54
4	TÉCNICA PIXRADIO	55
4.1	Recuperação por correspondência direta	55
4.1.1	<i>Carregamento, separação e mistura de canais</i>	56
4.1.2	<i>Obtenção de representações</i>	58

4.1.3	<i>Cálculo do custo de correspondência</i>	60
4.1.4	<i>Agregação dos custos de correspondência</i>	62
4.1.5	<i>Obtenção do mapa de disparidades</i>	64
4.1.6	<i>Verificação de reciprocidade</i>	66
4.1.7	<i>Refinamento do mapa de disparidades</i>	67
4.1.8	<i>Recuperação do par estéreo</i>	69
4.2	<i>Colorização</i>	71
4.2.1	<i>Processo de colorização de um pixel</i>	71
4.2.2	<i>Colorização incremental</i>	73
4.3	<i>Considerações finais</i>	76
5	AVALIAÇÃO EXPERIMENTAL	77
5.1	<i>Base de Imagens</i>	77
5.1.1	<i>Identificação de oclusão</i>	78
5.1.2	<i>Identificação de textura</i>	81
5.1.3	<i>Identificação de diferenças radiométricas</i>	83
5.2	<i>Avaliação Objetiva</i>	87
5.2.1	<i>Par estéreo recuperado</i>	87
5.2.2	<i>Recuperação por correspondência direta</i>	90
5.2.3	<i>Recuperação por colorização</i>	92
5.3	<i>Avaliação Subjetiva</i>	93
5.4	<i>Considerações finais</i>	97
6	CONCLUSÕES	99
6.1	<i>Contribuições</i>	99
6.2	<i>Limitações e Trabalhos Futuros</i>	100
	REFERÊNCIAS	101
	GLOSSÁRIO	105

INTRODUÇÃO

1.1 Contextualização

A facilidade de produção de conteúdo *multimídia* através de *smartphones*, a popularização de serviços de *streaming* e a alta capacidade de difusão desse tipo de conteúdo proporcionada pelas redes sociais causaram um aumento de tráfego multimídia na internet. Isto pode ser comprovado com os dados publicados pela [Cisco \(2018\)](#), que estima que até 2022 cerca de 82% do tráfego de internet no mundo será desse tipo de conteúdo, sendo que em 2016 essa taxa era de 73%. Desse total, [Harstead e Sharpe \(2015\)](#) estimam que de 4% a 6% serão de conteúdo em 3D. Além de vídeos e imagens estáticas, as aplicações 3D também podem ser interativas como no caso de jogos imersivos de realidade virtual e aumentada, podem ter fins didáticos, como em aplicações em medicina que utilizam imagens estereoscópicas para a simulação de cirurgias¹, ou ainda serem usadas em visão computacional e robótica².

O termo 3D - *3 dimensões* - se refere à capacidade de reprodução da sensação de profundidade de uma cena utilizando uma técnica chamada *estereoscopia*. Nessa técnica um par de imagens ligeiramente distintas da mesma cena são exibidas ao observador, sendo cada uma delas dedicada a um dos seus olhos. A esse par de imagens dá-se o nome de *par estéreo* ([LIPTON, 1997](#)). A visualização individual de cada imagem do par estéreo por seu respectivo olho permite a sensação de profundidade por meio da *estereopsia* ([GOLDSTEIN, 2013](#)), um efeito proporcionado pelo cérebro ao interpretar as duas imagens observadas simultaneamente. A produção de conteúdo estereoscópico utiliza câmeras especiais com duas lentes ou um par de câmeras comuns dispostas de modo a permitir a captura simultânea de duas visões da mesma cena para a geração do par estéreo ([LIPTON, 1982](#)). Existem também técnicas para a conversão

¹ Exemplos do uso de imagens estéreo em medicina podem ser encontrados nos trabalhos de [Beurden, IJsselsteijn e Juola \(2012\)](#) e [Nam et al. \(2012\)](#)

² Exemplos do uso de imagens estéreo em robótica podem ser encontrados nos trabalhos de [Murray e Little \(2000\)](#) e [Goldberg, Maimone e Matthies \(2002\)](#)

de conteúdos planares (imagem única 2D) em estereoscópicos (ZHANG; VÁZQUEZ; KNORR, 2011).

Para a visualização de conteúdos estereoscópicos é necessário que cada um dos olhos do espectador veja sua respectiva imagem do par estéreo. Diversos métodos foram desenvolvidos para permitir essa visualização e dentre esses métodos, que serão apresentadas no Capítulo 2.2.1, destaca-se o método *anaglífico*, que consiste na criação de uma imagem única pela sobreposição das imagens do par estéreo, descartando-se canais de cor complementares das visões direita e esquerda e sobrepondo-se os canais restantes. Esse processo descarta 50% das informações de cor do par estéreo e resulta em uma imagem composta pelas duas visões chamada de *anáglifo*. Para a visualização da imagem estereoscópica a partir do anáglifo é necessário o uso de óculos especiais que o filtrem, permitindo que cada um dos olhos veja uma sua respectiva imagem.

A codificação anaglífica tem como vantagem o baixo custo para ser utilizada, já que pode ser aplicada a projetores e monitores comuns, necessitando apenas o uso dos óculos com filtros de cor que normalmente são de baixo custo. Outra vantagem decorre do descarte de metade da informação de cor presente no par estéreo, que reduz o volume de dados, facilitando o armazenamento e a transmissão desse tipo de conteúdo.

Um problema causado pelo descarte de canais do par estéreo é que as imagens codificadas pelo método anaglífico somente podem ser visualizadas por meio desse método. Assim, diversas imagens e vídeos anaglíficos legados disponíveis na Web não podem ser visualizados por meio de outros métodos, como o de luz polarizada - que hoje em dia é a método mais utilizado em salas de cinema - nem em televisores que normalmente utilizam multiplexação temporal. É, então, desejável que haja meios de se recuperar o par estéreo original a partir do anáglifo, para possibilitar a visualização via outros métodos, atingindo independência de dispositivos de visualização, ou até mesmo para a visualização do conteúdo planar (2D) de cada uma das visões por usuários que não possuam dispositivos de visualização estereoscópica.

A esse processo de recuperação dá-se o nome de *reversão anaglífica*. A reversão anaglífica é possível pelo fato de um par estéreo possuir informações das duas visões, esquerda e direita, muito parecidas. Isso permite a recuperação de uma aproximação do par estéreo original a partir da comparação entre essas visões, fazendo a recuperação dos canais que foram descartados no processo de codificação. Contudo, esse não é um processo trivial. Em geral, algoritmos de reversão anaglífica no estado da arte utilizam descritores de pixels e adaptações de técnicas de correspondência estéreo para encontrar correspondências entre pontos semelhantes nos canais do anáglifo e realizar a cópia de informações para a recuperação dos canais faltantes no par estéreo.

A área de *visão computacional* estuda técnicas de *correspondência estéreo*, sendo uma área de pesquisa ativa com um grande número de trabalhos sendo realizados (JACQUESSON *et al.*, 2020; ULHAQ *et al.*, 2020; FENG *et al.*, 2019; SCHARSTEIN; SZELISKI; ZABIH, 2001). A correspondência estéreo visa encontrar pixels correspondentes em imagens estéreo para o cálculo da disparidade, que posteriormente pode ser utilizada para o cálculo da distância entre a câmera

e objetos retratados nestas imagens. Técnicas de correspondência estéreo tradicionais buscam a correspondência entre cada pixel do par estéreo por meio da comparação entre canais de mesma cor ou entre a versão em tons de cinza das imagens do par estéreo, sendo que essas técnicas assumem, ainda, que as condições de iluminação dos pixels correspondentes são as mesmas nas duas imagens comparadas, ou seja, que as imagens são *radiometricamente semelhantes* (KUMARI; KAUR, 2016; SCHARSTEIN; SZELISKI; ZABIH, 2001).

Por outro lado, a *correspondência estéreo anaglífica* (ou correspondência anaglífica³), precisa fazer comparações entre canais de cores diferentes, pois os canais da mesma cor nas visões opostas foram eliminados durante a codificação anaglífica. Canais de cor diferentes possuem grandes diferenças radiométricas e isto impede o uso de técnicas tradicionais de correspondência estéreo para a correspondência anaglífica. Mesmo existindo técnicas para a correspondência estéreo em imagens com diferenças radiométricas (KUMARI; KAUR, 2016), esses métodos não podem ser diretamente aplicados a reversão anaglífica, devido às diferentes naturezas das imagens estéreo e anaglífica (Williem; RASKAR; PARK, 2015). Assim, percebe-se que a correspondência anaglífica é um problema mais difícil que a correspondência estéreo.

Em regiões de imagens que possuem baixa textura a correspondência anaglífica também é uma tarefa difícil, pois nessas regiões há pouca variação na intensidade dos pixels com relação aos seus vizinhos. Por essa falta de variação, essas regiões não possuem informações suficientes que permitam a comparação de seus pixels com a visão oposta de maneira confiável, o que pode causar correspondências incorretas.

Essas dificuldades, juntamente com a impossibilidade de se encontrar correspondência em regiões oclusas⁴, fazem com que restem regiões contendo pixels sem correspondência definida. Nessas regiões uma abordagem comumente utilizada é a colorização, que estima qual cor deve ser usada para um pixel que não pode ser recuperado via correspondência com base nas cores encontradas nos seus vizinhos que foram recuperados via correspondência. Entretanto, os métodos de colorização utilizados pelas técnicas em estado da arte normalmente são computacionalmente caros, além de a colorização apresentar resultados inferiores à recuperação por correspondência anaglífica. Isso aumenta a necessidade de que a técnica de correspondência maximize a quantidade de pixels recuperados via correspondência anaglífica, pois isso tanto melhora a qualidade das imagens obtidas quanto torna a técnica de reversão mais eficiente.

Todas as técnicas de reversão anaglífica encontradas na literatura lidam com o problema das diferenças radiométricas e das regiões de baixa textura. Entretanto, as soluções propostas utilizam descritores para os pixels que não foram projetados especificamente para tratar esses problemas. Além disso, como será discutido no Capítulo 2, essas soluções muitas vezes apresentam alto custo computacional ou baixa qualidade da imagem recuperada. Assim este trabalho tem

³ Alguns trabalhos também utilizam o termo "alinhamento de pixels" para a correspondência estéreo anaglífica

⁴ regiões visíveis em apenas uma das imagens do par estéreo devido à diferença de ângulo entre as duas visões

o objetivo de propor uma nova técnica de reversão anaglífica baseada em um descritor de pixels projetado para tratar as diferenças radiométricas e as regiões com baixa textura, contribuindo para melhorar a qualidade da imagem recuperada e reduzir a necessidade de colorização.

Testes objetivos e subjetivos demonstraram a capacidade da técnica de gerar imagens com qualidade superior àquelas geradas por técnicas no estado da arte.

1.2 Organização do texto

Este texto está organizado em seis Capítulos. No Capítulo 1 foi apresentada uma introdução ao tema da pesquisa, contextualizando o problema a ser investigado e apresentando os objetivos. No Capítulo 2 é apresentada uma revisão dos principais fundamentos sobre visão estéreo e percepção de profundidade, uma introdução à estereoscopia, aos métodos de codificação e visualização estereoscópicos, às técnicas de correspondência estéreo e anaglíficas, às técnicas de colorização e, por fim, às métricas de qualidade que serão aplicadas para avaliar os resultados obtidos por este trabalho. Em seguida, o Capítulo 3 detalha os trabalhos em estado-da-arte relacionados. O Capítulo 4 descreve a técnica proposta, denominada PixRadio, e o Capítulo 5 apresenta testes objetivos e subjetivos com a finalidade de avaliar o método proposto. Por fim, no Capítulo 6 são apresentadas as conclusões obtidas por este trabalho de mestrado.

FUNDAMENTOS TEÓRICOS

Neste Capítulo serão apresentados os conceitos necessários para o entendimento tanto do problema da reversão anaglífica quanto da proposta deste trabalho. A Seção 2.1 trata de aspectos da visão humana, dando ênfase à sua capacidade de perceber profundidade. Em seguida, a Seção 2.2 explica a estereoscopia e os métodos estereoscópicos. Já a Seção 2.3 introduz conceitos de correspondência estéreo que ajudam a entender a abordagem proposta. A Seção 2.4 trata sobre técnicas de colorização e, por fim, a Seção 2.5 sobre os métodos e métricas que serão utilizadas para a avaliação da técnica proposta.

2.1 Visão humana

O sistema visual humano capta a luz emitida ou refletida por objetos em um ambiente por meio dos olhos - a luz atravessa a pupila, passa pelo cristalino e é focada no plano da retina, que contém receptores sensíveis à luz, os cones e bastonetes. Esses receptores contêm compostos químicos que geram sinais elétricos ao serem estimulados. Tais sinais são capturados por neurônios na retina e são enviados pelo do nervo ótico até o cérebro, onde são interpretados.

Para prover a percepção da profundidade de uma cena o cérebro interpreta um conjunto de diversas *dicas de profundidade* (*Depth Cues*) que são obtidas do sistema visual. Essa interpretação pode ser feita usando informações de apenas um ou dos olhos ou de ambos. As dicas de profundidade possuem diversas naturezas, podendo ser decorrentes das características das imagens projetadas nas retinas, das alterações oculares necessárias para a visão, da diferença entre as projeções em cada uma das retinas e até da diferença entre as imagens projetadas na mesma retina em momentos distintos, quando há movimento. Goldstein (2013) divide as dicas de profundidade em três grupos: dicas oculomotoras, dicas monoculares e dicas binoculares, que serão discutidas, respectivamente, nas Seções 2.1.1, 2.1.2 e 2.1.3.

2.1.1 Dicas oculomotoras

As dicas oculomotoras se relacionam às alterações fisiológicas que ocorrem nos olhos ao observarem objetos em diferentes distâncias. Esse tipo de dica é mais perceptível quando a distância entre o observador e o objeto é pequena.

- Acomodação

A acomodação é uma dica que se relaciona com a ação muscular necessária para alterar o formato do cristalino do olho e ajustar o foco da visão a um objeto. O cérebro interpreta essa informação e a utiliza para ajudar na compreensão da profundidade.

- Convergência

Para observar objetos próximos é necessário que os olhos se movam para dentro, convergindo, conforme pode ser visto na Figura 1a. Objetos distantes necessitam de menos convergência do que objetos mais próximos, conforme a Figura 1b. Essa convergência também é utilizada pelo cérebro para percepção de profundidade.

Figura 1 – Dica binocular e oculomotora *convergência*

Fonte – Adaptado de Goldstein (2013)

2.1.2 Dicas monoculares

As dicas monoculares podem ser percebidas utilizando apenas um dos olhos. A acomodação citada na Seção 2.1.1 é, além de uma dica oculomotora, uma dica monocular. As dicas monoculares também podem ser *pictoriais*, que são baseadas em imagens estáticas, ou *baseadas em movimento*.

2.1.2.1 Dicas monoculares pictoriais

Essas dicas são obtidas pelo cérebro pela interpretação da imagem projetada em uma das retinas, possuindo natureza psicológica. Esse tipo de dica é explorado por pintores e fotógrafos

na tentativa de reproduzir a noção de profundidade em suas obras.

Figura 2 – Dicas monoculares *oclusão*, *altura relativa*, *tamanho relativo* e *convergência perspectiva*

Fonte – [Goldstein \(2013\)](#)

- Oclusão

Ocorre quando um objeto bloqueia parcialmente a visão de outros objetos. Os objetos obstruídos são percebidos como mais distantes do observador. É possível notar que na Figura 2 o cacto ao lado direito da imagem bloqueia a visão dos arbustos e que os arbustos obstruem a visão da montanha. A oclusão não dá nenhuma informação sobre a distância entre os objetos, ou seja, é possível perceber que um objeto está atrás do outro, mas não é possível indicar o quão distantes estão.

- Altura relativa

Objetos que estão abaixo da linha do horizonte e possuem a sua base em um ponto mais alto do campo de visão normalmente são percebidos como estando mais distantes. Na Figura 2, a base da moto mais distante está mais alta na imagem que a base da moto mais próxima. Quando os objetos estão acima da linha do horizonte o efeito é o oposto: os objetos que possuem sua base em um ponto mais alto são percebidos como mais próximos.

- Tamanho relativo

Quando dois objetos que possuem o mesmo tamanho e estão a diferentes distâncias do observador, o objeto mais distante aparenta ser menor. Essa dica depende em parte do conhecimento prévio do observador quanto ao tamanho familiar dos objetos observados. Na Figura 2 as motos, que possuem tamanhos aproximadamente iguais, aparentam ser de tamanhos diferentes. A moto que se encontra mais distante aparenta ser menor.

- Convergência perspectiva

Linhas paralelas aparentam convergir no horizonte ao se distanciarem do observador. Ainda na Figura 2 é possível perceber que as linhas da estrada parecem se aproximar conforme a distância aumenta.

- Tamanho familiar

O fato de sabermos as dimensões de objetos permite ao cérebro estimar a distância com base no tamanho do objeto visto na imagem, haja visto que objetos distantes tender a parecer menores. Essa ideia está retratada na Figura 3. Sabemos que a moeda de 1 centavo é menor que a moeda de 1 real. Dessa forma, a moeda de 1 centavo, por estar maior que a moeda de 1 real na imagem, aparenta estar mais próxima.

Figura 3 – Dica monocular *tamanho familiar*

Fonte: Elaborada pelo autor.

- Perspectiva atmosférica

A atmosfera dispersa a luz provinda de objetos distantes, fazendo que esses objetos apresentem um tom azulado e sejam menos nítidos quando comparados a objetos próximos. A Figura 4a demonstra essa dica. Pode-se notar que as pedras mais distantes apresentam desfoque e uma tonalidade azulada quando comparadas com as pedras no primeiro plano.

Figura 4 – Dicas monoculares *perspectiva atmosférica* e *gradiente de textura*

(a) Perspectiva Atmosférica

(b) Gradiente de Textura

Fonte – Adaptado de Goldstein (2013)

- Gradiente de textura

O aumento na distância entre o observador e um objeto altera a percepção de textura desse objeto. Conforme a distância aumenta, a densidade da textura parecer estar aumentando. Na Figura 4b é possível perceber que a textura do chão parece variar conforme a aumenta a distância.

- Sombras

As sombras associadas a objetos em uma cena podem fornecer informações sobre a localização desses objetos. A Figura 5a mostra esferas em um tabuleiro. Nessa figura a localização das esferas é incerta, sendo que elas podem estar na superfície do tabuleiro ou flutuando sobre ele. Já na Figura 5b, as sombras ajudam a perceber que as esferas da esquerda estão apoiadas no tabuleiro e as da direita estão flutuando. Sombras aumentam a tridimensionalidade de objetos, como as sombras que fazem os círculos da Figura 5 parecerem ser esferas.

Figura 5 – Dica monocular *sombras*

Fonte – Goldstein (2013)

2.1.2.2 Dicas monoculares de movimento

As dicas pictoriais vistas na seção anterior podem ser percebidas mesmo quando não há movimento relativo entre o observador e o objeto. Quando existe esse movimento relativo, surgem novas dicas de profundidade.

- Paralaxe de movimentação

Quando nos movemos, objetos próximos parecem se deslocar rapidamente na nossa frente, enquanto objetos distantes parecem mover mais lentamente. Assim conseguimos estimar distância de objetos com base na velocidade em que eles se deslocam.

- Inclusão e exclusão

Durante o movimento relativo entre observador e objeto, alguns objetos podem ser encobertos e outros descobertos. Assim, tem-se a noção de profundidade por inclusão/exclusão de objetos em movimento no campo de visão. Essa dica está relacionada tanto à dica pictorial *oclusão* quanto à *paralaxe de movimento*.

2.1.3 Dicas binoculares

As dicas binoculares são aquelas que são percebidas utilizando-se simultaneamente ambos os olhos. A dica oculomotora de convergência é também é uma dica binocular. Os olhos de um ser humano adulto estão localizados a aproximadamente 6,5cm de distância um do outro, isso faz com que um determinado ponto de uma imagem seja projetado nas retinas de cada olho em pontos ligeiramente diferentes. A Figura 6 mostra as imagens vistas pelos olhos esquerdo e direito. Pode-se notar que as posições dos elementos variam entre as Figuras 6a e 6b. A distância entre os pontos correspondentes nas imagens projetadas na retina de cada olho é conhecida como *disparidade de retina*. O cérebro é capaz de fundir as imagens individuais de cada um dos olhos e, a partir da disparidade de retina, sintetizar a sensação de profundidade por meio de um efeito chamado de *estereopsia* (WHEATSTONE; WHEATSTONE, 1838). A estereopsia é uma das dicas de profundidade mais importantes.

Figura 6 – Exemplo de par estéreo: imagens vistas por cada um dos olhos

(a) Imagem vista pelo olho esquerdo

(b) Imagem vista pelo olho direito

Fonte – Goldstein (2013)

2.2 Estereoscopia

Como demonstrado na Seção 2.1, a estereopsia, juntamente com as dicas de profundidade pictoriais e de movimento, são os fatores que permitem a percepção da profundidade de uma cena. Imagens e vídeos tradicionais/planares conseguem reproduzir com fidelidade as dicas monoculares pictoriais e por movimento (no caso de vídeos). Entretanto, esse tipo de conteúdo não é capaz de reproduzir as dicas binoculares nem a estereopsia, que depende da disparidade de retina e, conseqüentemente, de uma imagem para cada olho. Assim, a reprodução da estereopsia necessita de uma técnica diferente da tradicional para a captura de imagens e vídeos. Essa técnica é a *estereoscopia*.

A estereoscopia estuda métodos para a criação e visualização de imagens que reproduzem a sensação de profundidade permitindo a estereopsia. O processo de captura de imagens e vídeos estereoscópicos é feito com o uso de duas câmeras montadas com os centros óticos de suas lentes alinhados horizontalmente e 6,5cm distantes entre si, a fim de reproduzir a distância entre os olhos de um ser humano. A captura de imagens estereoscópicas exige técnicas de fotografia específicas para que as imagens capturadas não causem desconforto ao espectador. Existem trabalhos que descrevem os detalhes técnicos sobre como capturar imagens estereoscópicas otimizadas, como os de [Lipton \(1982\)](#), [Lipton \(1997\)](#). O par de imagens capturado é chamado de *par estéreo*. Um exemplo de par estéreo pode ser visto na Figura 6.

Para a visualização de imagens estereoscópicas é necessário utilizar um método ou dispositivo que permita que cada um dos olhos veja a imagem do par estéreo que é destinada a si. Esses métodos serão apresentados na Seção 2.2.1.

2.2.1 Métodos de visualização estereoscópicos

O primeiro dispositivo para visualização estereoscópica foi proposto por C. Wheatstone ([WHEATSTONE; WHEATSTONE, 1838](#)) e consistia em um arranjo de espelhos que permitia a projeção de duas imagens de forma independente para cada um dos olhos do observador. A Figura 7 ilustra um estereoscópio de Wheatstone. As imagens do par estéreo são fixadas nos pontos E e E' e são refletidas pelos espelhos A e A' . O observador se posiciona frente aos espelhos e cada um de seus olhos vê, individualmente, pelo do reflexo dos espelhos, a imagem destinada a si.

Figura 7 – Estereoscópio de Wheatstone

Fonte – [Lipton \(1982\)](#)

A partir da invenção do estereoscópio, outras tecnologias foram desenvolvidas e hoje existem diversos métodos de visualização de imagens estereoscópicas. Os principais métodos de

visualização são:

- Multiplexação de cor – método anaglífico

Nesse método alguns componentes de cor de cada imagem do par estéreo são descartados e os componentes restantes são unidos formando uma única imagem: o anáglifo. Para separar os canais da imagem anaglífica e visualizar as imagens distintas para cada olho é necessário utilizar óculos com lentes que filtram a imagem formada separando as duas visões. Por ser fundamental a este trabalho, esse método será detalhado na Seção 2.2.3.

- Multiplexação por luz polarizada

Esse método possibilita a exibição integral de ambas as imagens, sem perdas na qualidade de cor da cena. Nessa modalidade, as imagens do par estéreo são projetadas simultaneamente em uma tela, sendo que cada uma das imagens passa por um filtro polarizador em uma orientação (vertical ou horizontal) diferente. Para a visualização utiliza-se óculos especiais polarizados de modo que o olho esquerdo veja uma direção da polarização e o olho direito veja a outra. Essa é a tecnologia utilizada atualmente na maioria das salas de cinemas 3D e requer dois projetores perfeitamente alinhados e sincronizados, além de uma tela especial. Uma desvantagem dessa técnica é a perda em luminosidade.

A Figura 8a mostra a técnica de multiplexação por luz polarizada. A tela reproduz uma imagem composta pela sobreposição do par estéreo com polarizações vertical e horizontal. Os óculos especiais permitem a filtragem das imagens e a visualização correta de cada componente dedicado ao seu respectivo olho.

Figura 8 – Métodos de visualização por multiplexação

- Multiplexação temporal

Nesse método as imagens do olho esquerdo e do olho direito são exibidas sequencialmente e alternadamente. Para a visualização, óculos especiais sincronizados com o dispositivo projetor alternam a opacidade entre a lente direita e a lente esquerda na mesma frequência em que a TV ou dispositivo projetor exibe as imagens. Dessa forma, cada olho vê apenas

os *frames* destinados a si. Devido ao efeito de memória da imagem, onde o cérebro consegue unir imagens exibidas em um intervalo de até 50 milissegundos, se vê uma imagem contínua e não se percebe que por alguns instantes a visão de cada olho está sendo bloqueada. Essa técnica é utilizada por alguns modelos de televisores 3D e é relativamente cara por necessitar de óculos ativos.

A Figura 8b exemplifica esse método. Nela pode-se notar que enquanto a tela exibe a imagem esquerda do par estéreo a lente esquerda fica transparente e a lente direita fica opaca. Dessa forma, a imagem esquerda só é vista pelo olho esquerdo. Quando é exibida a imagem direita do par estéreo a opacidade dos óculos se alterna da lente direita para a esquerda.

- Dispositivos HMD

Dispositivos HMD (*Head Mounted Display*) possuem duas telas distintas para que cada uma das imagens do par estéreo seja vista de exclusivamente por seu respectivo olho. Dessa forma, o par estéreo pode ser visualizado sem perdas de cor nem de intensidade luminosa. Esses dispositivos têm crescido em disponibilidade nos últimos anos devido a miniaturização da eletrônica e a popularização das iniciativas de desenvolvimento de óculos de realidade virtual e aumentada. Existem soluções baratas como o Google Cardboard (Fig. 9a) que permitem transformar um smartphone comum em um HMD para a visualização estéreo, e também dispositivos específicos dedicados a essa tarefa como o Hololens, da Microsoft (Fig. 9b).

Figura 9 – Dispositivos HMD

(a) Google Cardboard

(b) Microsoft HoloLens

- Métodos auto estereoscópicos

Os métodos apresentados até aqui têm um ponto em comum: todos necessitam de óculos especiais para a visualização das imagens estereoscópicas. O uso destes óculos por longos períodos pode causar desconforto e até mesmo fadiga. Assim, os métodos auto estereoscópicos tentam propiciar a visão estereoscópica sem a necessidade de óculos, deixando para a tela a tarefa de apresentar as imagens do par estéreo de maneira individualizada ao espectador. A Figura 10a mostra o método de barreira de paralaxe. Nesse método é adicionada uma barreira composta por fendas estreitas intercaladas com placas opacas

que bloqueiam a luz em certas direções. Assim, cada olho tem uma visão de diferentes conjuntos de pixels. O método apresentado na Figura 10b utiliza uma película especial composta por pequenas lentes que são capazes de direcionar a luz de cada pixel para um dos olhos do observador. Devido as suas características, a produção de monitores auto estereoscópicos ainda é cara. Outro problema com esse tipo de tecnologia é que para assistir o conteúdo nesse tipo de tela o observador precisa estar posicionado a uma distância específica do monitor.

Figura 10 – Métodos de visualização auto estereoscópicos

Fonte – Yugoshi (2018)

2.2.2 Problemas de visualização

Os métodos de visualização apresentados na Seção 2.2.1 dependem da utilização de dispositivos que muitas vezes apresentam imperfeições. Como exemplo, em um sistema de visualização anaglífica *ideal* os óculos deveriam ser capazes de filtrar perfeitamente o anáglifo, como mostra a Figura 11a. Acontece que em sistemas reais, muitas vezes por imperfeições nos filtros dos óculos ou por diferenças de calibração dos projetores e monitores, a filtragem do anáglifo é incompleta, permitindo a passagem de alguns componentes da visão oposta (Figura 11b). Isto causa o problema da *interferência (crosstalk)* e faz com que o olho esquerdo veja uma imagem com componentes da visão direita e vice-versa, como mostra a Figura 12.

Figura 11 – Problema de visualização *interferência*

Fonte: Elaborada pelo autor.

Na Figura 12 é possível notar que a visão esquerda não foi completamente filtrada, sendo possível ver uma "sombra" da imagem direita. Essa sombra é chamada de *ghosting* e pode causar desconforto durante a visualização de uma imagem estereoscópica. Esse efeito também pode ocorrer com outras técnicas e não é exclusiva do método anaglífico. Quando o *ghosting* é intenso pode levar um problema chamado *rivalidade de retina*.

Figura 12 – Problema de visualização *ghosting*

Fonte – Adaptado de (SANFTMANN; WEISKOPF, 2011)

A rivalidade de retina é um efeito da visão que acontece quando imagens diferentes são apresentadas a cada um dos olhos. Quando isso ocorre o cérebro não é capaz de unir as imagens corretamente e isso faz com que essas imagens sejam vistas de maneira independente e intercalada. A Figura 13 exemplifica o problema da rivalidade de retina. Quando apresentadas imagens diferentes aos olhos, a imagem percebida não é devidamente fundida pelo cérebro, proporcionando a visão independente de cada uma delas e de forma aleatória (SANFTMANN; WEISKOPF, 2011).

Figura 13 – Problema de visualização *rivalidade de retina*

Fonte – Adaptado de (DIETER; TADIN, 2011)

2.2.3 Método anaglífico

Cada imagem do par estéreo, quando representadas no espaço de cores RGB, é composta por 3 canais de cores: o vermelho (*Red*), o verde (*Green*) e o azul (*Blue*). Como já mencionado na Seção 2.2.1, durante a codificação anaglífica alguns desses canais são descartados das imagens do par estéreo e os canais restantes são unidos em uma única imagem, o *anáglofo*. A Figura

14 mostra o processo básico de codificação anaglífica. Nota-se que no caso exemplificado foi escolhido eliminar os canais verde e azul da visão esquerda e o canal vermelho da visão direita. Assim, união do canal vermelho da imagem esquerda aos canais verde e azul da imagem direita resulta em um anáglifo denominado *vermelho-ciano*.

Figura 14 – Processo de codificação anaglífica vermelho-ciano

Fonte: Elaborada pelo autor.

É possível criar anáglifos com outras combinações dos canais do par estéreo: a união do canal verde da visão esquerda com os canais vermelho e azul da visão direita dá origem a um anáglifo *verde-magenta* e quando se utiliza o canal azul da visão esquerda com o vermelho e o verde da visão direita o resultado é um anáglifo *azul-amarelo* ou *azul-âmbar*. Apesar de não haver um esquema de cores padrão e [Andrade e Goularte \(2009\)](#) indicar que o melhor esquema para a percepção de profundidade é o verde-magenta, é possível perceber a predominância de conteúdo codificados no esquema vermelho-ciano. Assim, neste trabalho daremos ênfase a esse tipo de anáglifo.

Figura 15 – Esquemas de cores X Tipos de óculos

Fonte: Elaborada pelo autor.

Para a visualização das imagens utilizando o método anaglífico é necessário utilizar um óculos com filtros de cores como os apresentados na Figura 15, que permitem a filtragem do anáglifo e a visualização individual das imagens do par estéreo por seu respectivo olho. O tipo de óculos a ser utilizado depende do esquema de cores usado para codificar o anáglifo. As Figuras 15a, 15b e 15c mostram qual tipo de óculos usar para cada esquema de cores.

Apesar de outros métodos, como os de multiplexação por luz polarizada e multiplexação temporal, apresentarem melhores resultados por reproduzirem totalmente as cores do par estéreo, o método anaglífico é o único que permite a visualização de imagens estereoscópicas em monitores e televisões tradicionais. Por outro lado, anáglifos são mais suscetíveis aos problemas de visualização apresentados na Seção 2.2.2, como a rivalidade de retina e o efeito de *ghosting*. A fim de minimizar esses problemas, diversos *modos* de codificação podem ser utilizados.

2.2.3.1 Anáglifos coloridos (*Color anaglyphs*)

Esse é o modo de codificação que utilizamos até o momento e que está exemplificado na Figura 14. Nesse modo, para se obter um anáglifo vermelho-ciano, são descartados os canais verde e azul da imagem esquerda e o canal vermelho da imagem direita. Os canais restantes são sobrepostos e resulta no anáglifo. Essa codificação é capaz de reproduzir cores, mas dependendo da cena poderá ocorrer a rivalidade de retina ao observador. A Figura 16 mostra um anáglifo vermelho-ciano colorido.

As Equações 2.1 a 2.4 mostram como obter os valores RGB ($r_a g_a b_a$), de um pixel do anáglifo em cada um dos modos apresentados. A entrada das equações são os valores ($r_l g_l b_l$) da imagem esquerda e ($r_r g_r b_r$) da imagem direita do par estéreo. As equações devem ser aplicadas a todos os pixels da imagem.

$$\begin{bmatrix} r_a \\ g_a \\ b_a \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} r_l \\ g_l \\ b_l \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} r_r \\ g_r \\ b_r \end{bmatrix} \quad (2.1)$$

Figura 16 – Anáglifo vermelho-ciano codificado no modo *colorido*

Fonte: Elaborada pelo autor.

2.2.3.2 Anáglifos 'verdadeiros' (True anaglyphs)

Nesse modo descarta-se toda a informação de cor das imagens esquerda e direita e são utilizadas apenas a informação de luminosidade nelas contida. Na Equação 2.2 é possível notar que os valores (0,299;0,587;0,114) são as proporções dos canais RGB definidos pela ITU-T que devem ser utilizados para a conversão de imagens RGB em escala de cinza (RICHARDSON, 2002). Para o caso do anáglifo vermelho-ciano, a imagem esquerda convertida em escala de cinza é atribuída ao canal vermelho do anáglifo e a imagem direita, também em escala de cinza, ao canal azul ou ao canal verde do anáglifo. Nesse modo de codificação a imagem resultante é escura e sem reprodução de cores. Entretanto, esse modo traz a vantagem de apresentar pouco *ghosting*. A Figura 17 mostra um anáglifo vermelho-ciano verdadeiro.

$$\begin{bmatrix} r_a \\ g_a \\ b_a \end{bmatrix} = \begin{bmatrix} 0,299 & 0,587 & 0,114 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} r_l \\ g_l \\ b_l \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0,299 & 0,587 & 0,114 \end{bmatrix} \times \begin{bmatrix} r_r \\ g_r \\ b_r \end{bmatrix} \quad (2.2)$$

Figura 17 – Anáglifo vermelho-ciano codificado no modo 'verdadeiro'

Fonte: Elaborada pelo autor.

2.2.3.3 Anáglifos cinzas (Gray anaglyphs)

Esse modo também descarta a informação de cor do par estéreo e utiliza apenas a informação de luminosidade nelas contida. Entretanto, para o caso do anáglifo vermelho-ciano, a imagem esquerda é atribuída ao canal vermelho do anáglifo e a imagem direita aos canais azul e verde. Assim como as imagens codificadas no modo *True Anaglyph*, imagens codificadas nessa

modalidade também não reproduzem cores e por serem mais claras, apresentam mais *ghosting*. A Figura 18 mostra um anáglifo vermelho-ciano no modo cinza.

$$\begin{bmatrix} r_a \\ g_a \\ b_a \end{bmatrix} = \begin{bmatrix} 0,299 & 0,587 & 0,114 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} r_l \\ g_l \\ b_l \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0,299 & 0,587 & 0,114 \\ 0,299 & 0,587 & 0,114 \end{bmatrix} \times \begin{bmatrix} r_r \\ g_r \\ b_r \end{bmatrix} \quad (2.3)$$

Figura 18 – Anáglifo vermelho-ciano codificado no modo *cinza*

Fonte: Elaborada pelo autor.

2.2.3.4 Anáglifos semicoloridos (*Half color anaglyphs*)

Esse modo é um meio termo entre *Gray Anaglyphs* e *Color Anaglyphs*. A imagem esquerda é convertida em escala de cinza e atribuída ao canal vermelho do anáglifo e os canais azul e verde da imagem direita são copiados para se formar o anáglifo vermelho-ciano. A reprodução de cores é menor que nos *Color Anaglyphs*, mas a ocorrência de rivalidade de retina também é menor. A Figura 16 mostra um anáglifo vermelho-ciano semicolorido.

$$\begin{bmatrix} r_a \\ g_a \\ b_a \end{bmatrix} = \begin{bmatrix} 0,299 & 0,587 & 0,114 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} r_l \\ g_l \\ b_l \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} r_r \\ g_r \\ b_r \end{bmatrix} \quad (2.4)$$

Figura 19 – Anáglifo vermelho-ciano codificado no modo *semicolorido*

Fonte: Elaborada pelo autor.

2.3 Correspondência estéreo x Correspondência anaglífica

Dentro da comunidade de *visão computacional*, a *correspondência estéreo* é uma das áreas mais ativas atualmente, com um grande número de trabalhos sendo realizados. Por outro lado, a *correspondência anaglífica* é uma área relativamente nova, com poucos trabalhos dedicados a essa tarefa. Tanto a correspondência estéreo quanto a correspondência anaglífica têm a finalidade de obter o deslocamento (*disparidade*) entre pontos correspondentes das visões esquerda e direita de imagens estereoscópicas. O que torna essas técnicas diferentes é sua entrada: Enquanto que as técnicas de correspondência estéreo dispõem do *par estéreo* para realizar as comparações, as técnicas de correspondência anaglífica tem como entrada o *anáglifo*. Essa é uma grande diferença pois um par estéreo possui as visões esquerda e direita com canais vermelho, verde e azul, totalizando 6 canais, enquanto anáglifos possuem apenas um canal da visão esquerda e dois da visão direita, ou seja, anáglifos possuem 50% da informação contida em um par estéreo, o que torna a correspondência anaglífica mais difícil que a correspondência estéreo.

As disparidades encontradas por ambas as técnicas frequentemente são dispostas em um *mapa de disparidades*, como o apresentado na Figura 20b. O mapa de disparidades contém os valores da disparidade para cada pixel do par estéreo ou do anáglifo. Valores pequenos de disparidade geram um ponto escuro no mapa e valores grandes geram pontos claros. Nota-se que na Figura 20b os cones no primeiro plano apresentam tonalidade mais clara. Isto significa que a disparidade desse cone no par estéreo é mais alta que os cones em segundo plano. O principal uso dos mapas de disparidades obtidos pelas técnicas de correspondência estéreo é o cálculo da

Figura 20 – Mapa de disparidades

Fonte: Elaborada pelo autor.

distância entre a câmera e objetos, o que permite a carros autônomos e a robôs se deslocarem em um espaço tridimensional. Já no caso da correspondência anaglífica, o principal uso do mapa de disparidade é a reversão anaglífica. Com as disparidades entre os pixels dos canais do anáglifo, a recuperação de grande parte das imagens esquerda e direita do par estéreo pode ser realizada simplesmente copiando as cores faltantes na posição correspondente na imagem da visão oposta.

Segundo [Scharstein, Szeliski e Zabih \(2001\)](#), os algoritmos de correspondência estéreo podem ser locais ou globais. Algoritmos locais tendem a ser mais eficientes e buscam as correspondências dentro de janelas predeterminadas, mas também são mais sensíveis a problemas como oclusões - regiões que são visíveis em uma das imagens, mas não na outra. Algoritmos globais são menos eficientes que algoritmos locais, mas são mais precisos, buscando correspondências na imagem toda. Problemas enfrentados pelas técnicas de reversão anaglífica muitas vezes tem a mesma natureza dos problemas enfrentados por técnicas de correspondência estéreo. [Kumari e Kaur \(2016\)](#) listam os principais desafios que a área de correspondência estéreo trata atualmente:

- **Variações fotométricas e radiométricas** - É um dos principais desafios na correspondência estéreo atualmente. Esse problema é causado pela falta de consistência entre as imagens que estão sendo comparadas, sendo essas inconsistências causadas pela variação de intensidade luminosa e de cor entre as imagens. Tais variações decorrem de diversos fatores, como o uso de sensores com características diferentes ou por diferentes parâmetros (abertura, tempo de exposição, etc.) durante a captura das imagens. É possível notar que o par estéreo apresentado nas Figuras [21a](#) e [21b](#) estão com diferentes exposições. Essa diferença atrapalha a busca por correspondências. No caso da correspondência anaglífica esse é o principal problema encontrado, porque as comparações devem ser feitas entre canais de cores diferentes das imagens direita e esquerda. As imagens [21c](#) e [21d](#) mostram os canais vermelho e verde de um anáglifo vermelho-ciano. É possível notar a diferença

Figura 21 – Par estéreo com diferentes intensidades

(a) Imagem esquerda

(b) Imagem direita

(c) Canal vermelho do anáglifo

(d) Canal verde do anáglifo

(e) Região de baixa textura

Fonte: Elaborada pelo autor.

entre as intensidades dos pixels que compõem os cones em ambas as visões.

- **Regiões com baixa e sem textura** - É importante que o processo de correspondência estéreo obtenha uma correspondência única entre os pontos correspondentes nas imagens. Regiões com iluminação constante podem apresentar baixa textura e, assim, um ponto pode ter mais de uma correspondência possível. A Figura 21e mostra um exemplo de

região de baixa textura. No caso específico da reversão anaglífica, regiões com cores primárias puras (verde, vermelho e azul) tem esse problema potencializado.

Scharstein, Szeliski e Zabih (2001) realizam uma comparação entre diversas técnicas de correspondência estéreo e propõe uma taxonomia básica, na qual identificaram que tais técnicas podem ser divididas em 4 etapas:

1. Cálculo do custo de correspondência
2. Agregação dos custos
3. Cálculo das disparidades
4. Refinamento da disparidade

Essas etapas são como blocos básicos das técnicas, sendo que as técnicas não necessariamente implementam todas as etapas. Por exemplo, algumas técnicas locais podem realizar o cálculo do custo de correspondência e a agregação em apenas uma etapa. Existem também técnicas globais que não realizam a agregação dos custos de correspondência.

A etapa *cálculo do custo de correspondência* tem a finalidade de quantificar a similaridade entre dois pixels em um valor chamando de *custo de correspondência*, que é usado para determinar qual a melhor correspondência para um pixel. Essa quantificação pode ser feita utilizando-se diversas métricas que usam diretamente a intensidade dos pixels ou até mesmo características da região próxima ao pixel, como bordas e gradientes. A Tabela 1 lista as principais dessas métricas.

A segunda etapa, *agregação dos custos*, é muito aplicada a técnicas *locais* e tem a finalidade de agregar/juntar os custos de correspondência em regiões predeterminadas, normalmente em janelas quadradas ou em regiões segmentadas das imagens para aumentar a área das comparações. A agregação é feita por meio da soma ou média dos valores de custo de correspondência, podendo ser feita linearmente ou com médias ponderadas cujos pesos variam em função das distâncias entre os pixels. Uma abordagem comum é realizar a ponderação utilizando a distribuição gaussiana. Essa etapa ajuda a diminuir a ocorrência de correspondências erradas, principalmente em regiões com baixa textura e com textura repetitiva.

O *cálculo das disparidades* é uma etapa cuja complexidade varia em função do tipo de técnica aplicada. Em técnicas *locais*, após o cálculo e agregação dos custos de correspondência, a decisão sobre qual a melhor correspondência para um pixel é, na maioria das vezes, feita através do menor ou maior custo de correspondência obtido para aquele pixel, dependendo da métrica escolhida. Já em técnicas *globais*, essa é a etapa que leva maior parte do processamento pois o cálculo das disparidades é feito para a imagem toda simultaneamente, visando minimizar uma função global.

Tabela 1 – Métricas para correspondência

Métrica de Correspondência	Definição
Correlação Cruzada Normalizada (NCC)	$\frac{\sum_{u,v} (I_1(u,v) - \bar{I}_1) \cdot (I_2(u+d,v) - \bar{I}_2)}{\sqrt{\sum_{u,v} (I_1(u,v) - \bar{I}_1)^2 \cdot (I_2(u+d,v) - \bar{I}_2)^2}}$
Soma das Diferenças Quadradas (SSD)	$\sum_{u,v} (I_1(u,v) - I_2(u+d,v))^2$
SSD Normalizada	$\sum_{u,v} \left(\frac{(I_1(u,v) - \bar{I}_1)}{\sqrt{\sum_{u,v} (I_1(u,v) - \bar{I}_1)^2}} - \frac{(I_2(u+d,v) - \bar{I}_2)}{\sqrt{\sum_{u,v} (I_2(u+d,v) - \bar{I}_2)^2}} \right)^2$
Soma das Diferenças Absolutas (SAD)	$\sum_{u,v} I_1(u,v) - I_2(u+d,v) $
Rank	$\sum_{u,v} (I'_1(u,v) - I'_2(u+d,v))$ $I'_k(u,v) = \sum_{m,n} I_k(m,n) < I_k(u,v)$
Censo	$\sum_{u,v} \text{HAMMING}(I'_1(u,v), I'_2(u+d,v))$ $I'_k(u,v) = \text{BITSTRING}_{m,n}(I_k(m,n) < I_k(u,v))$

Fonte – Adaptado de [Brown, Burschka e Hager \(2003\)](#)

2.4 Colorização de imagens

A correspondência anaglífica permite a recuperação da maior parte do par estéreo a partir do anáglifo realizando-se a cópia direta das cores em pixels correspondentes. Entretanto, em regiões oclusas, regiões de grandes diferenças radiométricas e em regiões de baixa textura, a correspondência anaglífica não pode ser realizada com exatidão. Nessas regiões as técnicas de reversão anaglífica utilizam adaptações de técnicas de *colorização* para estimar valores e completar a recuperação do par estéreo.

Segundo ([YATZIV; SAPIRO, 2006](#)), a primeira técnica computacional de colorização foi criada por Wilson Markle em 1970 e adicionava cor a filmes em preto e branco. Desde então diversas técnicas foram desenvolvidas e frequentemente utilizam 2 abordagens:

- Transferência de cor

Técnicas baseadas em transferência de cor têm duas entradas: uma imagem em escala de cinza, que será colorizada, e uma imagem colorida que será utilizada como referência. A técnica busca na imagem de referência as cores que vai atribuir à imagem monocromática. A Figura 22 exemplifica uma imagem colorizada a partir de uma imagem de referência. À esquerda tem-se a imagem de referência que terá suas cores transferidas para a imagem a ser colorizada, ao centro. Por fim, à direita tem-se a imagem colorizada com as cores da imagem de referência.

Figura 22 – Colorização de imagens a partir da transferência de cor

Fonte – Welsh, Ashikhmin e Mueller (2002)

- Anotações

Técnicas baseadas em anotações possuem apenas uma entrada: a imagem a ser colorizada contendo anotações sobre quais cores cada região deverá possuir. A Figura 23 mostra um exemplo de uma imagem colorizada a partir de anotações. A imagem à esquerda, que é a entrada para a técnica de colorização, é uma imagem monocromática à qual foram adicionadas anotações com as cores que se deseja colorizar. A imagem à direita é a saída da técnica.

Figura 23 – Colorização a partir de anotações

Fonte – Levin, Lischinski e Weiss (2004)

As técnicas de colorização normalmente trabalham em espaços de cor que contêm informações sobre a luminância das imagens, como o YCrCb ou YUV. Como o anáglifo é composto por apenas um canal de cor da visão esquerda e dois da visão direita, não é possível obter a luminância para cada visão (esquerda e direita), que é função dos seus canais RGB. Assim, as técnicas de colorização de imagens monocromáticas não podem ser aplicadas diretamente para recuperar canais de um anáglifo, necessitando ser adaptadas para esse caso. Tais técnicas adaptadas não apresentam resultados bons como as técnicas originais, pois trabalham com menos informações.

2.5 Qualidade em imagens e vídeos digitais

Imagens e vídeos digitais que passaram por técnicas de processamento, codificação e compressão podem apresentar diversas distorções e ruídos que influenciam negativamente sua qualidade. Esses ruídos podem causar desconforto e dificultar a sua visualização. Para garantir a qualidade desse tipo de conteúdo após o processamento digital e permitir a comparação entre diferentes técnicas, é preciso realizar a avaliação da qualidade das imagens e vídeos por elas gerados. Essa avaliação pode ser feita utilizando métricas objetivas e subjetivas.

Métricas objetivas são baseadas em modelos matemáticos que calculam a qualidade da imagem. São rápidas de serem obtidas e facilmente reproduzíveis, mas não levam em conta as especificidades do sistema visual humano, que é complexo. Por outro lado, métricas subjetivas envolvem a avaliação das imagens por pessoas, o que as torna mais demoradas e mais difíceis de serem aplicadas.

As próximas Seções, 2.5.2, e 2.5.1 mostram as principais características desses dois tipos de avaliação:

2.5.1 Avaliação subjetiva de qualidade

A avaliação subjetiva da qualidade de imagens e vídeos é feita com a participação de avaliadores. Diversos procedimentos para esse tipo de avaliação foram definidos na recomendação BT.500-11 da ITU-R (BT *et al.*, 2002), sendo que um dos procedimentos mais utilizados é o método *Double Stimulus Continuous Quality Scale (DSCQS)*.

No DSCQS uma imagem ou vídeo que passou por algum tipo de processamento é exibido sequencialmente com o conteúdo original ao avaliador, daí vem o nome "duplo estímulo" (*Double Stimulus*). A exibição pode ser feita de duas formas: na primeira, a avaliação é feita por apenas um avaliador e esse pode alternar livremente entre as imagens original e processada para ter uma ideia da qualidade de cada uma delas dentro de um período de tempo predeterminado; na segunda, que realiza a avaliação com diversos observadores simultaneamente, a exibição é feita conforme a Figura 24, onde os conteúdos original e processado (A e B) são exibidos duas ou mais vezes por um tempo de exibição (x), com tempo de transição (y) entre eles. Em ambas as variações os avaliadores não podem saber qual é a imagem original nem qual a processada. Para isso, a ordem de exibição dessas imagens deve ser aleatória.

Após a visualização os avaliadores devem dar uma nota a cada uma das imagens. Isto é feito inserindo uma marca em uma escala vertical como a mostrada na Figura 25. As escalas de avaliação do par de imagens devem ser impressas aos pares para facilitar o processo de avaliação. Cada escala possui um sistema de notas contínuo para evitar erros de quantização, mas é dividida em cinco seções com os valores de qualidade associadas a cada Seção (BT *et al.*, 2002).

Os resultados obtidos pela aplicação do DSCQS utilizando vários avaliadores podem ser

Figura 24 – Esquema de avaliação do DSCQ

Fonte – Adaptado de (BT *et al.*, 2002)

Figura 25 – Escala de avaliação do DSCQS

Fonte – Adaptado de (BT *et al.*, 2002)

interpretados utilizando-se a métrica *Mean Opinion Score (MOS)*, que representa a média das pontuações dadas por esses avaliadores. Quando o MOS é utilizado para a avaliação, as notas obtidas no DSCQS são representadas em uma escala que varia de 1 a 5. Nessa escala a pontuação 5 é equivalente ao conceito *excelente* e significa que nenhum defeito foi percebido na imagem ou vídeo pelo avaliador. A pontuação 4 equivale ao conceito *bom*, onde algum defeito foi percebido, mas não causou desconforto. A pontuação 3 corresponde ao conceito *aceitável* e os defeitos percebidos causaram desconforto. Já na pontuação 2, *ruim*, apesar da grande degradação na imagem, o avaliador conseguiu visualizar alguma informação. Por fim, a pontuação 1 equivale ao conceito *péssimo* e o avaliador não foi capaz de entender a imagem Yugoshi (2018).

O DSCQS geralmente é aceito como uma medida realística da qualidade visual subjetiva, mas ele sofre com alguns problemas: os resultados obtidos podem variar significativamente dependendo dos avaliadores e das imagens sendo testadas. Essa variação pode ser compensada com a utilização muitos avaliadores. Avaliadores com experiência tendem a dar respostas enviesadas, o que torna preferível a avaliação por pessoas sem experiência. Entretanto avaliadores sem experiência tendem a aprender rapidamente os problemas das imagens, passando a dar respostas

enviesadas. A busca por novos avaliadores pode tornar o processo de avaliação demorado e trabalhoso.

2.5.2 Avaliação objetiva de qualidade

Por causa dos problemas que os métodos de medição subjetivos podem apresentar, muitas vezes é necessário a utilização de métricas de avaliação objetivas. As avaliações objetivas não substituem as subjetivas, mas são muito mais simples de serem aplicadas e muito úteis principalmente nas etapas de desenvolvimento para a comparação entre técnicas. As métricas de qualidade objetivas não necessitam da interação humana e são obtidas por algoritmos e fórmulas que são aplicadas diretamente nos vídeos e imagens avaliadas. Esse tipo de medida traz resultados precisos sobre a quantidade de ruído presente nesses vídeos e imagens. As métricas de qualidade objetivas podem ser classificadas quanto a disponibilidade do vídeo ou imagem original para comparação. Quando a medida utiliza o conteúdo original para comparação, temos uma medida de *Referência Total* (FR - Full-Reference). Quando não utiliza o conteúdo original, a medida é *Sem Referência* (NR - No-Reference). Em alguns casos apenas algumas informações do conteúdo original estão disponíveis e métricas de *Referência Reduzida* (RR - Reduced Reference) podem ser utilizadas. (Zhou Wang; SHEIKH; BOVIK, 2003).

O PSNR (*Peak Signal-to-Noise Ratio*) é uma medida de referência total. É uma das técnicas mais utilizadas na avaliação objetiva de imagens e vídeos. O cálculo dessa métrica é feito pela comparação pixel a pixel entre duas imagens: a imagem original e a imagem que passou por algum tipo de processamento. O resultado do PSNR é expresso em decibéis (dB) em uma escala que vai de 0 a 100. Nessa métrica, valores altos significam uma alta similaridade entre as imagens original e processada.

Seu cálculo é realizado utilizando as Equações 2.5 e 2.6. Primeiramente é feito o cálculo da *soma média dos erros* (MSE - Mean Squared Error) entre as duas imagens, que é obtido pela Equação 2.5, onde (I) é imagem original e (K) a imagem processada, (m) é a largura da imagem e (n) a altura. Após o cálculo do MSE, aplica-se a Equação 2.6, onde MAX é o valor máximo de intensidade de pixel da imagem. Quando os pixels são representados utilizando 8 bits, MAX vale 255.

$$MSE = \frac{1}{mn} \cdot \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2 \quad (2.5)$$

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right) \quad (2.6)$$

Apesar de ser uma métrica muito utilizada e boa para analisar a adição de ruídos em imagens, o PSNR não é capaz de avaliar as imagens levando em consideração todos os fatores

que envolvem a percepção visual humana. Assim, não é possível afirmar que quanto maior o PSNR melhor a qualidade da percepção visual da imagem.

2.6 Considerações finais

A visão humana leva em conta diversos fatores para a percepção e sensação de profundidade. Dentre esses fatores, a estereopsia é um dos mais importantes. A estereoscopia estuda métodos para a criação e visualização de imagens que reproduzem a sensação de profundidade utilizando a estereopsia. Dentre os métodos disponíveis, o método anaglífico se destaca por ser barato e facilmente utilizado. Apesar de não existir um padrão definido para a codificação anaglífica, é possível perceber a predominância de anáglifos codificados no esquema de cores vermelho-ciano e no modo colorido. Assim, este trabalho irá utilizar esse tipo de anáglifo.

O processo de reversão anaglífica pode ser dividido em duas etapas. A primeira recupera informações utilizando-se a correspondência anaglífica, demonstrada na Seção 2.3. A segunda busca colorizar as regiões do par estéreo recuperado onde não foi possível realizar a correspondência anaglífica. Técnicas de colorização foram apresentadas na Seção 2.4.

A Seção 2.5 mostrou dois tipos de métodos de avaliação de qualidade em imagens e vídeos: métodos objetivos e subjetivos. Como visto, os métodos objetivos são práticos pela facilidade de aplicação, mas, por não levar em conta avaliadores humanos, seus resultados muitas vezes não são precisos. Por outro lado, testes subjetivos descrevem melhor a qualidade percebida, sendo mais confiáveis, mas esses testes são de difícil aplicação. Durante a execução desse trabalho ambos os tipos de métodos de avaliação serão utilizados em momentos distintos. Durante a fase de desenvolvimento serão utilizados testes objetivos e durante a fase de validação da técnica serão realizados testes subjetivos.

REVISÃO DE LITERATURA

Neste Capítulo serão apresentados os principais trabalhos encontrados na literatura que abordam a *reversão anaglífica*. Esses trabalhos apresentam dois tipos de abordagens: técnicas que usam informações externas aos anáglifos e técnicas que utilizam apenas as informações internas aos anáglifos, que serão apresentadas na Seção 3.1. As *técnicas baseadas em informações externas ao anáglifo* modificam o processo de codificação anaglífica, gerando informações adicionais sobre o par estéreo original que são armazenadas em estruturas *externas* ao anáglifo e são utilizadas juntamente com o anáglifo no momento da reversão. Por dependerem dessas estruturas adicionais, essas técnicas não são capazes de reverter conteúdo legado, mas apresentam qualidade superior às que não utilizam tais estruturas.

A *técnicas baseadas em informações internas ao anáglifo* utilizam apenas as informações disponíveis no próprio anáglifo para o processo de reversão. Isso faz com que elas possam ser utilizadas em conteúdos anaglíficos legados que foram criados por métodos de codificação tradicionais. Entretanto, por contar com menos informações para realizar a reversão anaglífica, as imagens recuperadas por essas técnicas tendem a apresentar qualidade inferior quando comparadas com imagens recuperadas por técnicas que utilizam informações externas.

3.1 Técnicas baseadas em informações internas ao anáglifo

Joulin e Kang (2013) propuseram uma técnica de reversão que se baseia apenas no anáglifo. Essa técnica utiliza uma versão modificada do método SIFT-Flow (LIU *et al.*, 2008) (denominada ASIFT-Flow) para realizar a correspondência anaglífica e obter um conjunto inicial de correspondências entre as duas visões que compõem o anáglifo. Essa correspondência inicial então é refinada aplicando-se o algoritmo de fluxo ótico (LUCAS; KANADE, 1981), gerando novas correspondências que são avaliadas para verificar a consistência de textura e

a reciprocidade de correspondência. Em regiões com pouca textura e em regiões oclusas a técnica sofre dificuldade em encontrar boas correspondências e utiliza uma adaptação do método proposto por [Levin, Lischinski e Weiss \(2004\)](#) para colorizar essas regiões.

Figura 26 – Técnica proposta por [Joulin e Kang \(2013\)](#)

Fonte – Adaptado de [Joulin e Kang \(2013\)](#)

A técnica é resumida na Figura 26. A entrada é um anáglifo colorido vermelho-ciano. Os canais vermelho, referente à visão esquerda, e ciano, referente à visão direita, são separados e é realizada a busca pelas correspondências considerando as consistências estéreo, de textura e ausência de textura. Então utiliza-se uma adaptação da técnica de [Levin, Lischinski e Weiss \(2004\)](#) para o preenchimento de cores que não compatíveis e o processo é repetido.

Apesar de apresentar bons resultados visuais, essa técnica acaba trocando cores nas regiões que foram colorizadas. É possível perceber o surgimento de cores que, apesar de causarem pouco desconforto visual, são muito diferentes das cores das imagens originais. Isto faz com que essa técnica tenha bom desempenho em testes subjetivos, mas resultados ruins em testes objetivos. Outra limitação provém da utilização da técnica de fluxo óptico, que possui a complexidade $O(n^3)$ ([YUGOSHI, 2018](#)), com n representando o número de pixels da imagem, tornando a técnica computacionalmente cara.

[Williem, Raskar e Park \(2015\)](#) desenvolveram uma técnica iterativa que obtém um mapa de disparidade preciso da cena a partir do anáglifo e, a partir das disparidades, recupera o par estéreo. Essa técnica utiliza dois novos métodos baseados em *limitadores por cores locais* e na *distribuição reversa de intensidade* para realizar a correspondência anaglífica e gerar o mapa de disparidades. Após encontrar as disparidades, a técnica realiza a cópia das cores conhecidas de uma imagem para outra e coloriza os pixels restantes que se encontram em regiões oclusas utilizando colorização baseada em difusão, com uma técnica adaptada de [Levin, Lischinski e Weiss \(2004\)](#).

A Figura 27 demonstra a estrutura geral da técnica. A entrada é um anáglifo colorido vermelho ciano. Os canais esquerdo e direito são separados e comparados utilizando-se os novos métodos em busca de pixels correspondentes. O mapa de disparidade é calculado em função das correspondências encontradas e os canais faltantes do par estéreo são recuperados com base nessas correspondências. As regiões em que não foram encontradas correspondências são preenchidos via colorização por difusão.

Figura 27 – Técnica proposta por [Williem, Raskar e Park \(2015\)](#)

Fonte – Adaptado de [Williem, Raskar e Park \(2015\)](#)

Essa técnica é capaz de obter uma boa aproximação do par estéreo original, obtendo bons resultados em testes objetivos. Entretanto, seu desempenho é intermediário em testes subjetivos. É possível notar que essa técnica não consegue recuperar bem regiões de bordas, sendo possível encontrar artefatos em torno de bordas de objetos com descontinuidade de profundidade. Os próprios autores reconhecem que esse é um método computacionalmente caro. Segundo [Yugoshi \(2018\)](#), essa técnica possui complexidade $O(n^4)$.

A técnica proposta por [Yugoshi \(2018\)](#) utiliza um algoritmo de correspondência entre blocos (*block matching*) adaptado para imagens estéreo ([HANNAH, 1974](#)) para realizar a reversão anaglífica. Essa adaptação leva em conta a característica do um par estéreo de possuir imagens de cada visão parecidas e levemente deslocadas horizontalmente, o que permite limitar a área de busca de blocos a uma janela ao redor de um bloco alvo. Isso torna essa técnica mais rápida. Como métrica para definir a melhor correspondência é utilizada a Soma Absoluta das Diferenças (SAD). O autor divide a técnica em quatro etapas, conforme a Figura 28.

Na primeira etapa a imagem é lida e os parâmetros do anáglifo necessários para as etapas posteriores, como a altura, largura e número de bits por pixel da imagem, são obtidos. Em seguida os canais da imagem são separados e é aplicado o detector de bordas de Canny. A busca das correspondências é feita através da comparação entre essas imagens de bordas. Na terceira etapa é realizada a correspondência de blocos com transferência de cor. A busca é feita em blocos de 8x8, 16x16, 32x32 ou 64x64 pixels. Os blocos são buscados em uma janela de busca de 4 pixels na altura e na largura. Quando a melhor correspondência de um bloco

Figura 28 – Técnica proposta por Yugoshi (2018)

Fonte – Yugoshi (2018)

é encontrada esse bloco é copiado para o canal a ser recuperado. Na quarta etapa os canais existentes são unidos aos canais recuperados e o par estéreo final é salvo em arquivo. Essa técnica tem a vantagem de ser computacionalmente mais barata que as técnicas anteriores, possuindo complexidade $O(n)$, mas apresenta problemas na quantidade de artefatos presentes nas imagens geradas, como o serrilhamento que surge nas bordas dos objetos. Por utilizar a imagem de bordas obtida pelo Canny para realizar a correspondência estéreo, essa técnica está mais sujeita a realizar correspondências erradas.

Por fim, a técnica proposta por Kunze (2019) denominada SIRA (*Stereo Images Retrieving from Anaglyphs*) é baseada em algoritmos de análise de séries temporais. Nessa técnica, sequências de pixels adjacentes do anáglifo são representadas por séries temporais compostas por valores diferenciais de intensidade de cor e essas séries são então comparadas por similaridade, com o objetivo de encontrar pixels correspondentes provenientes das visões distintas do par estéreo que permitem a reversão anaglífica.

A técnica SIRA é composta por 6 etapas, ilustradas na Figura 29. Nas duas primeiras etapas os canais do anáglifo são separados, segmentados e são geradas representações para cada pixel utilizando séries temporais. Essas séries são usadas na etapa *alinhamento de séries*, que utiliza técnicas de alinhamento de séries temporais nas representações dos canais para gerar matrizes contendo as disparidades entre os canais esquerdo e direito. Os resultados de disparidade obtidos na etapa anterior são refinados nas etapas *validação de reciprocidade* e *propagação de alinhamento*. Em regiões oclusas e em regiões que não foi possível obter disparidades confiáveis

Figura 29 – Técnica proposta por Kunze (2019)

Fonte – Kunze (2019)

a técnica SIRA utiliza um método de colorização baseado no algoritmo *k nearest neighbors*, buscando cores em regiões próximas. A técnica SIRA é consideravelmente mais eficiente que as técnicas de Joulin e Kang (2013) e Williem, Raskar e Park (2015), tendo custo computacional de $O(n^2)$. Entretanto, testes subjetivos e objetivos apontam que as imagens recuperadas possuem qualidade ligeiramente inferior.

O maior problema que as técnicas de reversão anaglífica que usam informações internas aos anáglifos enfrentam é que elas precisam comparar as visões esquerda e direita do anáglifo usando os canais de cores diferentes. Esses canais possuem diferenças radiométricas que dificultam busca por correspondências. Todos os trabalhos aqui citados desenvolveram seus próprios descritores de pixel de modo a indiretamente contornar esse problema: Joulin e Kang (2013) realizaram uma adaptação do algoritmo SIFT-Flow para o caso anaglífico, criando o ASIFT-flow; Williem, Raskar e Park (2015) criaram um descritor baseado em função de minimização de energia, que acaba reduzindo as diferenças radiométricas; Yugoshi (2018) realizou a conversão dos canais do anáglifo em imagens de borda (Canny) e Kunze (2019) trabalhou com as diferenças entre a intensidade de pixels adjacentes ao criar as séries temporais. Contudo, nenhum desses métodos foi especialmente projetado para minimizar os problemas decorrentes das diferenças radiométricas e das regiões com baixa textura. A tabela 2 resume as técnicas apresentadas.

Tabela 2 – Comparação entre métodos de reversão anaglífica

Técnica	Qualidade	Custo	Método Correspondência
Joulin	Média	Alto	ASIFT-Flow
Williem	Média	Alto	Stereo Matching
Yugoshi	Baixa	Baixo	Block Matching
Kunze	Baixa	Baixo	Séries Temporais

Fonte: Elaborada pelo autor.

3.2 Considerações finais

Neste Capítulo foram apresentados os principais trabalhos encontrados na literatura que tratam de alguma maneira o problema de reversão anaglífica. Os trabalhos que usam informações externas aos anáglifos são robustos e conseguem recuperar uma boa aproximação do par estéreo, gerando poucos artefatos nas imagens. Entretanto, a necessidade de modificação no processo de codificação anaglífica impede a utilização desse tipo de técnica em conteúdos legados e aumenta a necessidade de armazenamento de dados devido as informações adicionais que precisam ser mantidas. Por outro lado, as técnicas que realizam a reversão anaglífica utilizando informações internas do anáglifo somente utilizam os canais de cor do próprio anáglifo. Isto torna possível a reversão anaglífica de conteúdo legado e a obtenção de arquivos menores, haja visto que um anáglifo possui 50% menos informações de cor do que seu par estéreo correspondente. Entre as técnicas que realizam a reversão anaglífica utilizando informações internas do anáglifo, as propostas de [Joulin e Kang \(2013\)](#) e [Williem, Raskar e Park \(2015\)](#) mostram bons resultados, mas possuem alto custo computacional. Já as propostas de [Yugoshi \(2018\)](#) e [Kunze \(2019\)](#) conseguem ser mais rápidas do que as anteriores. Entretanto, suas imagens apresentam qualidade ligeiramente inferior às demais.

TÉCNICA PIXRADIO

Como foi apresentado no Capítulo 3, as técnicas de recuperação anaglífica podem ser divididas em duas categorias: *técnicas que usam informações externas aos anáglifos* e *técnicas que usam informações internas aos anáglifos*. A técnica proposta, denominada PixRadio¹ se encaixa na segunda categoria, buscando melhorar a qualidade da reversão anaglífica utilizando apenas as informações contidas nos anáglifos. Essa melhoria é obtida principalmente pela implementação de um vetor de representação para os pixels que torna a técnica menos sensível às diferenças radiométricas existentes entre os canais do anáglifo, o que melhora o processo de correspondência anaglífica.

Essa nova técnica é capaz de recuperar imagens codificadas em qualquer esquema de cores apresentado na seção 2.2.3 (*vermelho-ciano*, *verde-magenta* ou *azul-amarelo*). Entretanto, pela predominância do uso na literatura de anáglifos codificados no esquema de cores vermelho-ciano no modo colorido, esse esquema de cores foi escolhido para os testes e exemplos. Outro motivo para essa escolha, para efeitos de comparação, foi o fato que as técnicas no estado da arte também trabalham tais anáglifos.

A PixRadio pode ser dividida em 2 grandes etapas: a recuperação dos canais por correspondência direta e a colorização das regiões em que não foram encontradas correspondências. Essas etapas são descritas, respectivamente nas seções 4.1 e 4.2.

4.1 Recuperação por correspondência direta

Conforme descrito na Seção 2.2.3, anáglifos são gerados a partir da sobreposição de determinados canais das visões esquerda e direita do par estéreo, sendo esse formado por duas visões muito parecidas de uma mesma cena. Essa característica permite que parte da reversão anaglífica seja realizada por meio da comparação entre essas visões, encontrando-

¹ O código fonte da técnica está disponível em <<https://github.com/andreldc/pixradio>>.

se correspondências entre elas e copiando-se as informações de cor perdidas no processo de conversão análogica *diretamente* do canal de cor oposto.

Essa etapa da técnica busca recuperar o máximo de pixels utilizando as correspondências, deixando a menor quantidade possível de pixels sem cor definida. Ela pode ser dividida em 8 subetapas: Na primeira, chamada **Carregamento, separação e mistura de canais**, o anáglifo é carregado em memória e suas visões esquerda e direita são separadas. Também é realizada a mistura dos canais verde e azul (dando origem à cor ciano) que compõem a visão direita para facilitar o processamento pela fase posterior. A segunda subetapa é chamada **Obtenção de representações** e nela são extraídos vetores de representações para cada um dos pixels de cada visão. Esses vetores agregam as informações espaciais na região ao redor do pixel utilizando codificação diferencial que permite uma melhor comparação esse pixel e os pixels da visão oposta durante a próxima subetapa. *Essa subetapa é um dos diferenciais da técnica*, pois esses vetores de representação proporcionam grande parte do ganho final em qualidade.

Após a obtenção das representações é realizado o **Cálculo dos custos de correspondência** para todos os pixels de ambas as visões. Essa subetapa obtém valores que indicam o quão similares são um determinado par de pixels. Em sequência, para diminuir a quantidade de correspondências erradas e a influência de ruídos, na etapa **Agregação dos custos de correspondência** é realizada a soma ponderada dos custos de correspondência de cada um dos pixels e de seus vizinhos próximos. Então na subetapa **Obtenção do mapa de disparidades** os mapas de disparidade iniciais são obtidos. Entretanto, esses mapas contêm regiões com disparidades erradas, como por exemplo regiões de oclusão, onde não deveriam existir disparidades. Também pode ocorrer a falta de reciprocidade em regiões de baixa textura e de grandes diferenças radiométricas, onde a disparidade encontrada para a visão esquerda é diferente daquela encontrada para a visão direita. Essas regiões são identificadas através da subetapa **Verificação da reciprocidade**. Em seguida é realizado o **Refinamento do mapa de disparidades** com a finalidade de preencher, quando possível, regiões inválidas do mapa de disparidades com a cópia das disparidades de regiões válidas. Por fim é realizada a **Recuperação do par estéreo** através das disparidades obtidas.

4.1.1 Carregamento, separação e mistura de canais

Anáglifos *vermelho-ciano* são compostos por três canais, sendo o *canal vermelho* (An_{vm}) correspondente à *visão esquerda* (V_E) e os canais *verde* (An_{vd}) e *azul* (An_{az}) correspondentes à *visão direita* (V_D) do par estéreo. Essa etapa do algoritmo carrega o anáglifo em memória, separando seus canais de cor e, conseqüentemente, suas visões, além de mesclar os canais azul e verde que são originais da visão direita, gerando um *canal ciano* (An_c), por meio da Equação 4.1, na qual k é um valor entre 0 e 1 que define o quanto de cada canal será usado na mistura.

$$An_c = k.An_{vd} + (1 - k).An_{az} \quad (4.1)$$

Figura 30 – Separação e mistura dos canais do anáglifo

Fonte – Elaborado pelo autor

A mistura dos canais azul e verde é realizada para facilitar a execução das etapas futuras,

pois assim é possível trabalhar apenas com duas imagens: uma que representa a visão esquerda, que nesse caso é o canal vermelho e uma que representa a visão direita, que é o canal ciano. Algumas técnicas selecionam apenas um dos canais (verde ou azul) da visão direita, descartando o outro. Ao se misturar os canais, suas informações espaciais são mescladas e não ocorre a perda de informações. A Figura 30 demonstra esse processo de separação e mistura.

Essa etapa tem como entrada o anáglifo com dimensões $(a, l, 3)$ e tem como saída os canais An_{vm} e An_c , cada um com dimensões (a, l) , que serão utilizados na próxima etapa. Tanto a operação de separação quanto a de mistura dos canais apresentam complexidade $O(1)$.

Foram realizados testes empíricos para a otimização do parâmetro k . Nesses testes variou-se k entre 0 e 1, com incrementos de 0,05, mantendo-se todos os demais parâmetros da técnica constantes e foi medido o PSNR médio das imagens. O maior valor de PSNR foi obtido com o valor $k = 0.95$. Assim, percebe-se que uma maior parcela do canal verde traz melhores resultados para a PixRadio.

4.1.2 Obtenção de representações

A Seção 2.3 mostrou que as diferenças de intensidades dos pixels presentes entre as duas visões, chamadas diferenças radiométricas, dificultam o processo de correspondência anaglífica. Os trabalhos relacionados descritos no Capítulo 2 lidam indiretamente com esse problema. Em geral, ao construírem seus descritores de pixel com o objetivo de encontrar correspondências, as técnicas acabam, de algum modo, calculando diferenças espaciais e de intensidade, mas apenas no sentido horizontal. A redução dessas diferenças contribui para reduzir as diferenças radiométricas, mas seus efeitos ficam limitados.

Neste trabalho, para tratar desse problema, a todos os pixels (x, y) de cada uma das visões é atribuída uma representação $R(x, y)$ do bloco que tem esse pixel no centro. Essa representação é baseada nas diferenças de intensidades nos sentidos vertical e horizontal dos pixels vizinhos dentro do bloco e agrega em um vetor as informações espaciais desse bloco.

A Figura 31a apresenta um bloco de 3x3 pixels onde o pixel central possui coordenadas (x, y) . A Figura 31b contém os valores numéricos de intensidade de cada um dos pixels desse bloco.

Figura 31 – Bloco de tamanho 3x3

Fonte: Elaborada pelo autor.

Para a obtenção da representação do bloco ao redor pixel (x,y) são calculadas as diferenças entre as intensidades na direção horizontal e na vertical (Eq. 4.2) entre os pixels vizinhos.

$$Horizontal : \begin{bmatrix} -64 & 170 \\ -4 & -157 \\ -6 & 173 \end{bmatrix}, Vertical : \begin{bmatrix} 133 & 193 & -134 \\ -178 & -180 & 150 \end{bmatrix} \quad (4.2)$$

Esses valores são, então, associados ao pixel (x,y) como um único vetor (Eq. 4.3):

$$R(x,y) = \begin{bmatrix} -64 & 170 & -4 & -157 & \dots & -134 & -178 & 180 & 150 \end{bmatrix} \quad (4.3)$$

Essa codificação reduz a influência das diferenças radiométricas pois torna o vetor de representações relativo à diferença entre um pixel e seus vizinhos e não aos valores absolutos de intensidade dos pixels do bloco. As Figuras 32a e 32c mostram 2 blocos com diferenças radiométricas e as Figuras 32b e 32d mostram, respectivamente, seus valores de intensidade. Em uma análise visual é fácil notar que esses blocos são correspondentes, pois apresentam as mesmas características espaciais. Entretanto, a intensidade média do bloco B1 é de 121, enquanto a intensidade média do bloco B2 é de 152. Isso mostra que, apesar de os blocos serem parecidos, o segundo bloco é mais "claro". Assim é possível perceber que a intensidade absoluta dos pixels tem pouca influência sobre o fato de os blocos serem correspondentes ou não, e o que mais importa para determinar sua correspondência é a variação de intensidade entre os pixels do bloco.

Na Seção 4.1.3 será discutido como é realizado o cálculo dos custos de dados através da métrica SAD (*Soma das Diferenças Absolutas* - do Inglês *Sum of Absolute Differences*). Essa seção mostrará que valores baixos de SAD implicam em alta similaridade entre os blocos comparados. Se a comparação entre os blocos B1 e B2 for realizada utilizando somente seus valores de intensidade, o SAD seria de 286, o que poderia ser considerado baixa correspondência por ser um valor alto. Por outro lado, utilizando a representação proposta, o SAD cai para 13, indicando que os blocos podem ser correspondentes.

Figura 32 – Blocos com diferenças radiométricas

Fonte: Elaborada pelo autor.

É possível notar que o tamanho do bloco pode interferir na qualidade da representação, pois um bloco muito pequeno carrega poucas informações sobre a vizinhança do pixel em questão, o que torna a representação menos discriminatória. Por outro lado, um bloco muito grande implica em um vetor de representação grande, que pode tornar cálculo dos custos de correspondência uma tarefa lenta. O tamanho t_{rep} do vetor de representação $R(x,y)$ é dado pela Equação 4.4, onde j_{rep} é o tamanho da janela de representação. Nesse exemplo, no qual a janela tem tamanho $j_{rep} = 3$, o vetor tem tamanho $t_{rep} = 12$.

$$t_{rep} = 2(j_{rep}^2 - j_{rep}) \quad (4.4)$$

Foram realizados testes empíricos para a otimização do parâmetro j_{rep} . Nesses testes variou-se j_{rep} entre 3 e 31, com incrementos de 2 pixels, mantendo-se todos os demais parâmetros da técnica constantes e foi medido o PSNR médio das imagens. O maior valor de PSNR médio foi obtido com o valor $j_{rep} = 13$. Foi possível perceber que janelas com tamanho superior a 13 tornam a técnica lenta, sem ganhos significativos de PSNR.

Partindo-se da Equação 4.4 e considerando-se que o cálculo de cada posição do vetor de representação deve ser realizado para todos N pixels da imagem, para as 2 visões, temos que a complexidade dessa etapa é $O(2.N.t_{rep})$.

Após o cálculo do vetor de representação para cada um dos $a.l$ pixels dos canais An_{vm} e An_c , tem-se como resultado uma matriz contendo os valores das representações da visão esquerda R_E e outra da visão direita R_D , ambas com dimensões (a,l,t_{rep}) . Essa matriz é a saída dessa segunda etapa e será utilizada na próxima etapa para o *cálculo do custo de correspondência*.

4.1.3 Cálculo do custo de correspondência

A Seção 2.3 apresenta que técnicas de correspondência estéreo *locais* normalmente são mais eficientes que técnicas *globais*. Assim, a fim de tornar a técnica mais eficiente, a busca por correspondências é realizada localmente, isto é, a *melhor correspondência* para cada pixel é buscada dentro de uma janela próxima a esse pixel, não na imagem toda.

Para se encontrar a melhor correspondência entre dois pixels é necessário utilizar uma métrica que estabeleça o quão parecidos esses pixels são. Essa métrica é chamada na literatura de *custo de correspondência* (SCHARSTEIN; SZELISKI; ZABIH, 2001). O custo de correspondência utilizado na PixRadio é a métrica *SAD* - *Soma das Diferenças Absolutas*, que é dada pela Equação 4.5, onde A e B são vetores de tamanho t .

$$sad(A, B) = \sum_{n=1}^t |A(n) - B(n)| \quad (4.5)$$

A *SAD* é uma métrica muito utilizada para a comparação de similaridade entre dois blocos de imagem. Essa métrica foi escolhida por possuir menor custo computacional que as demais métricas para o mesmo fim (BROWN; BURSCHKA; HAGER, 2003). É possível verificar que na Eq. 4.5 quando A e B são vetores iguais, o valor da *SAD* é 0 e quando A e B são muito diferentes, o valor da *SAD* é alto.

O custo de correspondência c entre um pixel com coordenadas (x_1, y_1) da V_E e outro pixel com coordenadas (x_2, y_2) da V_D é computado obtendo-se seus respectivos vetores de representação $R_E(x_1, y_1)$ e $R_D(x_2, y_2)$ e os aplicando à equação 4.5. Obtém-se assim a equação 4.6, onde t_{rep} é o tamanho do vetor de representação, explicado na Seção 4.1.2.

$$c(x_1, y_1, x_2, y_2) = \sum_{n=1}^{t_{rep}} |R_E(x_1, y_1, n) - R_D(x_2, y_2, n)| \quad (4.6)$$

A equação 4.6 pode ser usada para encontrar o custo de correspondência para um par de pixels. Entretanto, para se encontrar a *melhor correspondência* para um pixel (x, y) da V_E na V_D , é necessário calcular os custos de correspondência desse pixel comparando-o em um intervalo de j_{busca} pixels da V_D , a fim de se determinar qual desses pixels tem o menor custo de correspondência. Para isso, o vetor de representação do pixel $R_E(x, y)$ é comparado com os vetores de representação dos pixels dentro da janela de busca utilizando a equação 4.6, obtendo-se j_{busca} valores de custo de correspondência. Esses valores são armazenados no vetor chamado de *vetor de custos de correspondência*. Assim, para um pixel (x, y) da V_E , o vetor de custos de correspondência é dado pela equação 4.7.

$$VC_E(x, y) = [c(x, x+1, y) \quad c(x, x+2, y) \quad c(x, x+3, y) \quad \dots \quad c(x, x+j_{busca}, y)] \quad (4.7)$$

No exemplo apresentado a busca é realizada apenas horizontalmente, ou seja, na equação 4.6 adota-se que $y_1 = y_2 = y$. Esse caso é possível em uma situação onde as imagens são retificadas e possuem apenas deslocamentos horizontais². Em um cenário onde as imagens não

² Idealmente as imagens de um par estéreo não deveriam possuir deslocamentos verticais dado que as lentes das 2 câmeras de captura deveriam estar alinhadas

são retificadas, podem existir deslocamentos verticais e essa limitação não é verdadeira. Nesses casos a busca deve ser feita também verticalmente, ou as imagens precisam ser retificadas.

O processo de busca da melhor correspondência é repetido para todos os $a.l$ pixels da visão esquerda, obtendo-se uma matriz tridimensional de tamanho (a, l, j_{busca}) . Essa matriz é chamada de *matriz de custo de correspondência da visão esquerda* C_E . O mesmo processo deve ser repetido em relação à visão direita: cada pixel da V_D deve ser buscado dentro de uma janela da V_E , obtendo-se a *matriz de custo de correspondência da visão direita* C_D .

As matrizes C_E e C_D contém os custos de correspondência para cada deslocamento d na janela de busca. Assim, para cada valor de d é possível obter uma seção da matriz $C_E(d)$, que possui dimensões (a, l) e contém apenas os custos de correspondência para o deslocamento d . A Figura 33 apresenta alguns exemplos de valores de seção da matriz de custos C_E para os valores de d iguais a 21, 25 e 49. Essas figuras foram geradas a partir da conversão dos valores de SAD para intervalo 0-255 para poderem ser exibidas como uma imagem. Essa conversão torna possível a inspeção visual das matrizes de custo de correspondência. Nessas figuras, regiões que possuem valores altos de SAD aparecem mais claras nas imagens, enquanto regiões que possuem valores baixos de SAD aparecem escuras. Como dito anteriormente, valores baixos de SAD indicam grande similaridade entre os vetores comparados. É possível notar que a região mais escura na Figura 33b corresponde à região ao fundo na Figura 33a (a região contém uma grade que aparenta ser de madeira). Isso indica que os pixels mais escuros dessa região possuem disparidade de 21 pixels. Do mesmo modo, as regiões mais escuras nas Figuras 33c e 33d correspondem, respectivamente, ao cone branco mais ao fundo e ao cone verde mais à direita e têm disparidades de 25 e 49 pixels, respectivamente.

Essa etapa pode ser implementada utilizando uma janela deslizante, o que permite o cálculo dos custos de dados de forma eficiente. Cada um dos N pixels da imagem possui um vetor de representação de tamanho t_{rep} , que deve ser comparado com os vetores de representação dos pixels dentro da janela de tamanho j_{busca} . Assim, a complexidade dessa etapa é $O(N \cdot t_{rep} \cdot j_{busca})$.

4.1.4 Agregação dos custos de correspondência

Após o cálculo dos custos de correspondência há informações suficientes para a obtenção do mapa de disparidades, determinando as correspondências entre os pixels da V_E e da V_D . Entretanto, a literatura mostra que a *agregação* desses custos resulta em melhores correspondências (SCHARSTEIN; SZELISKI; ZABIH, 2001). Essa melhoria ocorre porque são agregados ao custo de correspondência de um pixel (x, y) uma parcela dos custos de seus pixels vizinhos, o que diminui a influência de ruídos aumenta a confiabilidade das correspondências encontradas.

A PixRadio realiza essa agregação aplicando um filtro gaussiano (GONZALEZ; WOODS, 2007) nas seções das matrizes de custo $C_E(d)$ e $C_D(d)$ para cada nível de disparidade d . A aplicação do filtro gaussiano realiza a soma ponderada dos custos que estão próximos de si,

Figura 33 – Secionamento da matriz de custo de correspondência

(a) Visão esquerda do par estéreo

(b) $C_{ED}(21)$ - Seção com $d=21$ (c) $C_{ED}(25)$ - Seção com $d=25$ (d) $C_{ED}(49)$ - Seção com $d=49$

Fonte: Elaborada pelo autor.

agregando tais valores em um único valor. Assim, para cada disparidade $d < j_{busca}$ das matrizes de custos C_E e C_D é aplicado o filtro gaussiano, resultando nas *matrizes de custos agregados* C_{agE} e C_{agD} , que possuem dimensões (a, l, j_{busca}) . A Figura 34 exemplifica uma seção dos custos de dados antes e após o processo de agregação.

O filtro gaussiano possui um parâmetro σ que deve ser devidamente ajustado. Foram realizados testes empíricos para sua otimização onde variou-se σ entre 1 e 10, com incrementos de 0,5 pixels, mantendo-se todos os demais parâmetros da técnica constantes e foi medido o PSNR médio das imagens. O maior valor de PSNR médio foi obtido com o valor $\sigma = 4$.

Esse filtro pode ser implementado com complexidade $O(N)$ e deve ser aplicado às j_{busca} disparidades. Assim, a complexidade dessa etapa é $O(N \cdot j_{busca})$.

Figura 34 – Seção da matriz de correspondência com $d=25$

Fonte: Elaborada pelo autor.

4.1.5 Obtenção do mapa de disparidades

A Subseção 4.1.3 mostrou que os custos de correspondência para um pixel (x, y) da V_E em um intervalo da janela de busca da V_D são armazenados em um vetor chamado de *vetor de custos de correspondência* $V_{c_E}(x, y)$. Os vetores de custo de correspondência para todos os pixels da imagem compõem a *matriz de custos de correspondência* C_E . Da mesma maneira, a *matriz de custos agregados* C_{agE} obtida na subetapa anterior é composta por *vetores de custos agregados* para $V_{c_{agE}}(x, y)$ cada pixel (x, y) da imagem.

Para se determinar qual a *melhor correspondência* para um pixel de coordenadas (x, y) , deve-se analisar todos os custos do *vetor de custos agregados* $V_{c_{agE}}(x, y)$ associado a esse pixel e verificar em qual disparidade tem-se o menor custo. A Figura 35 plota o vetor de custo correspondência para um pixel da visão esquerda. No eixo vertical, tem-se o valor do custo de correspondência agregado $V_{c_{agE}}(d)$ (métrica SAD), enquanto o eixo horizontal representa o valor de deslocamento d da janela (disparidade). Aproximadamente no valor $d = 26$ há um vale no gráfico, onde o valor do custo de correspondência cai drasticamente. Nesse ponto é encontrada a melhor correspondência e assim, pode-se determinar que a disparidade para o pixel em questão é igual a 26. Em termos matemáticos, $D_E(x, y) = 26$.

Em geral, a disparidade $D_E(x, y)$ de um pixel da V_E pode ser obtida pela da Eq. 4.8, onde argmin é uma função que retorna o índice onde se encontra o menor valor do vetor $C_{agE}(x, y)$. Para se obter o *mapa de disparidades* da visão esquerda aplica-se essa equação para todos os

Figura 35 – Vetor de custos de correspondência agregados para o pixel $x=231, y=120$

Fonte: Elaborada pelo autor.

pixels dessa visão. O mesmo deve ser feito para a V_D .

$$D_E(x, y) = \operatorname{argmin}(C_{agE}(x, y)) \quad (4.8)$$

É comum a conversão do mapa de disparidades em uma imagem para possibilitar uma análise visual. As Figuras 36a e 36b apresentam os mapas de disparidades das visões esquerda e direita D_E e D_D . Como apresentado na Seção 2.3, valores pequenos de disparidade geram um ponto escuro no mapa e valores grandes geram pontos claros.

Figura 36 – Mapas de disparidades

(a) Visão Esquerda

(b) Visão Direita

Fonte: Elaborada pelo autor.

Os mapas de disparidade obtidos nessa etapa possuem regiões com disparidades atribuídas de maneira incorreta pois existem disparidades associadas até mesmo para pixels localizados em regiões de oclusão, onde não existem correspondências. Para detectar tais regiões, realiza-se o processo de verificação de reciprocidade, descrito na próxima Subseção.

Como essa etapa precisa buscar o menor valor de disparidade dentro dos vetores de custos de dados agregados, que possuem tamanho j_{busca} , para todos os N pixels da imagem, a complexidade desse subetapa é dada por $O(N \cdot j_{busca})$.

A entrada desta etapa são as matrizes de custos agregados C_{agE} e C_{agD} , que possuem dimensões (a, l, j_{busca}) , e a saída são os mapas de disparidades das visões esquerda D_E e direita D_D , ambos com dimensões (a, l) .

4.1.6 Verificação de reciprocidade

A busca por correspondências descrita na seção anterior é realizada em todos os pixels das visões esquerda e direita. Entretanto, pixels em regiões oclusas não possuem correspondências e a correspondência encontrada pode estar errada para pixels localizados em regiões de baixa textura ou de grandes diferenças radiométricas. A Figura 37 apresenta um vetor de custos de um pixel de uma dessas regiões. Verifica-se que nesse vetor não é claro qual a real disparidade pois nenhum custo se destaca.

Figura 37 – Vetor de custos de correspondência para o pixel $x=84, y=243$

Fonte: Elaborada pelo autor.

Assim, é necessário verificar em quais as regiões foram encontradas correspondências válidas. Isso pode ser realizado por meio da *verificação da reciprocidade de correspondência*, que é obtida pela comparação entre os mapas de disparidade das visões esquerda D_E e direita

D_D . A reciprocidade para um pixel (x, y) da visão esquerda pode ser verificada utilizando-se a equação 4.9 e para um pixel da visão direita utilizando-se a equação 4.10.

$$V_E(x, y) = \begin{cases} 1, & \text{se } |D_E(x, y) - D_D(x - D_E(x, y), y)| < \delta \\ 0, & \text{caso contrário.} \end{cases} \quad (4.9)$$

$$V_D(x, y) = \begin{cases} 1, & \text{se } |D_D(x, y) - D_E(x + D_D(x, y), y)| < \delta \\ 0, & \text{caso contrário.} \end{cases} \quad (4.10)$$

A equação 4.9 verifica se a disparidade encontrada para o pixel (x, y) a partir do mapa de disparidades da visão esquerda tem valor próximo ao valor de disparidade encontrada para o pixel correspondente ao pixel (x, y) na visão direita. Se a diferença entre as disparidades encontradas estiver dentro de um limiar δ o pixel é considerado válido. O mesmo ocorre para a equação 4.10. Foram realizados testes empíricos a otimização do limiar δ onde testou-se valores entre 0 e 5, com incrementos de 1 pixel, mantendo-se todos os demais parâmetros da técnica constantes e foi medido o PSNR médio das imagens. O maior valor de PSNR médio foi obtido com o valor $\delta = 1$.

As matrizes V_E e V_D indicam quais pixels são válidos no mapa de disparidades. Essas matrizes podem ser exibidas como uma máscara, que é apresentada nas Figuras 36a e 36b para as visões esquerda e direita. As regiões em preto nessas figuras indicam pixels inválidos, que são pixels que não passaram no teste de reciprocidade. As máscaras são aplicadas aos mapas de disparidade D_E e D_D e as regiões inválidas são setadas com o valor 0 nesses mapas, gerando os mapas de disparidades *válidos* da visão esquerda D_{vE} e da visão direita D_{vD} . As Figuras 38c e 38d apresentam esses mapas.

A verificação de reciprocidade é realizada para os N pixels da imagem. Assim, essa etapa apresenta complexidade $O(N)$.

4.1.7 Refinamento do mapa de disparidades

Como discutido na Seção 3.1, o processo de reversão anaglífica pode ser dividido em 2 grandes etapas: a recuperação de pixels por correspondência direta e a colorização dos pixels que não foram possíveis obter a correspondência direta. O *mapa de disparidades válido* é importante pois é ele que determina quais pixels são recuperados por cada uma dessas etapas: pixels que possuem valor válido no mapa (diferente de 0) são recuperados por correspondência direta, enquanto pixels inválidos (valores iguais a 0) serão recuperados por colorização.

A recuperação por correspondência direta, por copiar a cor diretamente da visão oposta do anáglifo, é mais confiável e traz resultados melhores que a colorização, além de normalmente ser mais rápida. Assim, é importante que o mapa de disparidades possua o máximo de pixels com valores válidos, o que melhora a qualidade obtida pela técnica e a torna mais eficiente.

Figura 38 – Máscaras de reciprocidade

Fonte: Elaborada pelo autor.

Com a finalidade de diminuir a quantidade de pixels inválidos, após a obtenção do mapa de disparidades válido é realizado o *refinamento do mapa de disparidades*, que é a propagação das disparidades de pixels válidos para pixels vizinhos inválidos. Isso é possível porque pixels próximos tendem a ter disparidades próximas. A Figura 39 apresenta o mapa de disparidades válido da V_E antes e depois do processo de refinamento. Essa figura mostra que alguns pixels antes inválidos foram preenchidos e se tornaram válidos após o refinamento.

A PixRadio realiza o refinamento do mapa de disparidades com a aplicação de um *operador morfológico de fechamento* (GONZALEZ; WOODS, 2007) sobre os mapas de disparidades válidos D_{VE} e D_{VD} utilizando um *kernel* elíptico de 35x35 pixels. Esse operador copia a disparidade de pixels válidos para seus vizinhos inválidos, diminuindo a quantidade desses pixels. Após a aplicação desse operador, o processo de cálculo de reciprocidade descrito na seção

Figura 39 – Refinamento do mapa de disparidade da V_E

Fonte: Elaborada pelo autor.

4.1.6 é repetido para se obter os *mapas de disparidades válidos e refinados das visões esquerda* D_{vrE} e *direita* D_{vrD} .

O operador de fechamento pode ser implementado com complexidade $O(N.S)$, onde N é o tamanho da imagem e S é o tamanho do kernel utilizado.

4.1.8 Recuperação do par estéreo

Após o refinamento do mapa de disparidades é possível realizar a reversão anaglífica nas regiões válidas desse mapa. Dado o anáglifo vermelho-ciano de entrada An , composto pelos canais de cor vermelho An_{vm} , verde An_{vd} e azul An_{az} , para se recuperar os três canais de um pixel de coordenadas (x, y) da visão esquerda, utiliza-se a equação 4.11, onde Cn_{vmE} , Cn_{vdE} e Cn_{azE} são, respectivamente, os canais de cor vermelho, verde e azul da *imagem recuperada esquerda* I_{rdE} e D_{vrE} é o mapa de disparidades válido e refinado da visão esquerda.

$$I_{rdE} = \begin{cases} Cn_{vmE} = An_{vm}(x, y) \\ Cn_{vdE} = An_{vd}(x - D_{vrE}(x, y), y) \\ Cn_{azE} = An_{az}(x - D_{vrE}(x, y), y) \end{cases} \quad (4.11)$$

O mesmo pode ser realizado para se recuperar os três canais para um pixel (x, y) da visão direita. Nesse caso utiliza-se a equação 4.12, onde Cn_{vmD} , Cn_{vdD} e Cn_{azD} são, respectivamente, os canais de cor vermelho, verde e azul da *imagem recuperada direita* I_{rdD} e D_{vrD} é o mapa de

disparidades válido e refinado da visão direita.

$$I_{rdD} = \begin{cases} Cn_{vmD} = An_{vm}(x + D_{vrD}(x, y), y) \\ Cn_{vdD} = An_{vd}(x, y) \\ Cn_{azD} = An_{az}(x, y) \end{cases} \quad (4.12)$$

Nas equações, os canais *vermelho da visão esquerda* e os canais *verde e azul da visão direita* são os canais originais do anáglifo com suas respectivas cores. As equações são aplicadas a todos os N pixels válidos dos mapas de disparidades D_{vrE} e D_{vrD} . Assim a complexidade dessa etapa é $O(N)$. As imagens recuperadas são apresentadas na Figura 40. Essas imagens já são bastante similares às imagens originais do par estéreo, com a exceção das regiões em preto que não foram possíveis recuperar via correspondência direta por serem regiões oclusas ou regiões cujas correspondências encontradas não são confiáveis. Essas regiões serão recuperadas por um processo de colorização, descrito na Seção 4.2.

Figura 40 – Recuperação por correspondência direta

Fonte: Elaborada pelo autor.

Como mostramos, a complexidade computacional de cada uma das subetapas apresentadas anteriormente é dada por:

Carregamento, separação e mistura de canais: $O(1)$

Obtenção de representações: $O(2.N.t_{rep})$

Cálculo do custo de correspondência: $O(N.t_{rep}.j_{busca})$

Agregação dos custos de correspondência: $O(N.j_{busca})$

Obtenção do mapa de disparidades: $O(N.j_{busca})$

Verificação de reciprocidade: $O(N)$

Refinamento do mapa de disparidades: $(N.S)$

Recuperação do par estéreo: $O(N)$

Assim, a complexidade total da etapa de recuperação por correspondência direta pode ser dada pela soma das complexidades individuais de cada etapa $O(1 + 2.N.t_{rep} + N.t_{rep}.j_{busca} + N.j_{busca} + N.j_{busca} + N + N.S + N)$. Considerando-se que os parâmetros de ajuste são constantes, temos que a complexidade final desta etapa é de $O(7.N + 1) = O(N)$.

4.2 Colorização

Após a etapa de recuperação direta sobram as regiões com pixels inválidos para serem recuperadas. Essas regiões são compostas em sua grande maioria por pixels oclusos e por uma quantidade menor de pixels cuja correspondência encontrada com a visão oposta não é confiável. Assim, para pixels nestas regiões não é possível realizar a recuperação pela cópia de cores. Nessas regiões, a PixRadio realiza o processo de colorização, que é a obtenção da cor de um pixel marcado como inválido através das cores de seus vizinhos marcados como válidos. O processo de colorização implementado pela PixRadio é baseado na técnica de colorização proposta por [Kunze \(2019\)](#), com o diferencial de realizar a colorização de forma incremental, o que impede o surgimento de cores indevidas nas regiões colorizadas e torna a transição de cores nessa região mais suave. Essa técnica foi escolhida por possuir menor custo computacional quando comparada às adaptações da técnica de [Levin, Lischinski e Weiss \(2004\)](#), usadas por [Joulin e Kang \(2013\)](#) e [Williem, Raskar e Park \(2015\)](#).

Este capítulo demonstra o processo de colorização desenvolvido na PixRadio. Para melhor entendimento, o processo de colorização de apenas um pixel é apresentado na Seção 4.2.1. Em seguida a Seção 4.2.2 apresenta como os pixels são escolhidos para serem colorizados iterativamente, a fim de minimizar o surgimento de cores espúrias e artefatos nas regiões colorizadas.

4.2.1 Processo de colorização de um pixel

Um pixel *inválido* da V_E pode ter sua componente de cor vermelha recuperada diretamente do anáglifo pela equação 4.11. Entretanto, por não existir uma correspondência para esse pixel na V_D , não é possível recuperar as componentes verde e azul da mesma maneira. Assim, essas componentes são recuperadas *indiretamente*: é realizada uma busca por pixels vizinhos *válidos* que são parecidos com o pixel a ser colorizado e as componentes verde e azul desses pixels são usadas para gerar as componentes faltantes.

A Figura 41a apresenta um pixel (x, y) da *imagem recuperada via reversão direta* I_{rdE} (vide seção 4.1.8) cujas cores não puderam ser obtidas pela recuperação direta, apresentada na

seção interior. A busca por pixels similares a (x,y) somente pode utilizar como referência o canal vermelho do anáglifo, pois esse é o único canal dessa visão. Então esse pixel é comparado com seus vizinhos dentro de um bloco quadrado de tamanho b_{col} . A Figura 41b apresenta um exemplo desse bloco.

Figura 41 – Processo de colorização de um pixel

Fonte: Elaborada pelo autor.

A intensidade de (x,y) é subtraída das intensidades de seus vizinhos dentro do bloco e o valor absoluto dessa diferença é usado como indicativo de similaridade, sendo armazenados em uma *matriz de pesos*, que é apresentada na Figura 41c. Nessa matriz, valores próximos de 0 indicam que o pixel em questão é similar ao pixel central, então regiões escuras na imagem

indicam pixels similares a (x, y) .

Após a obtenção da matriz de pesos, os pixels cujos valores de similaridade estão abaixo de um limiar γ são selecionados, gerando uma *máscara de similaridade* (apresentada na Figura 41d) que indica a localização desses pixels. Se ao menos 2 pixels forem selecionados o procedimento de colorização segue. Caso contrário, o tamanho do bloco b_{col} é incrementado por um valor i e a busca por pixels similares é repetida. O bloco de busca é incrementado até que a quantidade mínima de pixels similares seja obtida ou até um limite b_{max} . Se mesmo quando o bloco atingir o tamanho b_{max} não for possível obter a quantidade mínima de pixels, então passa-se a incrementar o valor de γ em 1 e o processo é repetido.

Quando a máscara de similaridade tem no mínimo 2 pixels, para a obtenção das componentes faltantes aplica-se essa máscara ao bloco centrado em (x, y) da *imagem recuperada via reversão direta* I_{rdE} e seleciona-se todos seus pixels *válidos*. Então separa-se os canais verde e azul desse bloco, que são apresentados nas Figuras 41e e 41f. Por fim é realizada a média das intensidades entre os valores selecionados e essas médias são associadas aos canais verde e azul, juntamente com sua componente vermelha original, para o pixel (x, y) . A Figura 41g apresenta o pixel colorizado.

Os diversos parâmetros dessa etapa foram otimizados de forma empírica para uma imagem de 450x375 pixels e seus valores são listados abaixo.

$$b_{col} = 43f \quad \gamma = 5 \quad i = 12f \quad b_{max} = (l/4)^2$$

Alguns desses parâmetros devem ser reajustados em função do tamanho da imagem a ser colorizada utilizando-se um fator de escala $f = la/450 \times 375$, onde l é a largura da imagem e a é a sua altura. Cada parâmetro foi otimizado de maneira individual, variando-se o parâmetro em questão dentro de uma faixa de valores, mantendo-se todos os demais parâmetros da técnica constantes e foi medido o PSNR das imagens.

A complexidade dessa etapa depende do tamanho do bloco usado para a busca das cores e de quantas iterações foram necessárias para se obter a quantidade mínima de pixels necessária para realizar a média. No pior dos casos serão realizadas $(b_{max} - b_{col})/i$ iterações, e como o tamanho médio do bloco é $(b_{col} + b_{max})/2$, a complexidade é dada por $O((b_{col}^2 + b_{max}^2)/2i)$.

4.2.2 Colorização incremental

Para que a colorização seja realizada de maneira uniforme e para diminuir a quantidade de artefatos gerados por essa etapa, a técnica executa a colorização *incrementalmente*. A cada iteração é realizada a colorização dos pixels inválidos que tenham ao menos um vizinho direto válido, até que todos os pixels inválidos da imagem sejam colorizados. A Figura 42a apresenta a máscara obtida a partir da matriz de pixels válidos V_E , introduzida na Seção 4.1.6. Nessa máscara regiões em preto são inválidas, devendo ser colorizadas. A partir dessa máscara são obtidos os pixels de borda apresentados na Figura 42b. Na primeira iteração a colorização será realizada

somente nos pixels marcados em branco nessa imagem.

Figura 42 – Bordas das regiões válidas

Após a colorização de cada pixel a máscara de pixels válidos é atualizada e o pixel que antes estava marcado como inválido passa a ser considerado válido. Então o processo de obtenção das bordas dessa máscara se repete e seus pixels selecionados são colorizados, até que não haja mais pixels marcados como inválidos. A Tabela 3 apresenta o processo de colorização da imagem da visão esquerda. A primeira coluna apresenta o número da iteração. A segunda coluna apresenta a máscara de pixels válidos no início da iteração e a terceira coluna suas bordas, que são os pixels que serão colorizados nessa iteração. Por fim, a quarta coluna apresenta a imagem com os pixels de borda colorizados. Veja que os pixels são colorizados de maneira gradual, isso impede o surgimento de cores espúrias e a quantidade de artefatos gerados por esse processo. Esse processo é repetido para a imagem da visão direita e após todos os pixels terem sido colorizados chega-se ao fim da técnica, obtendo-se o par estéreo recuperado, apresentado na Figura 43.

Figura 43 – Par estéreo recuperado

Fonte: Elaborada pelo autor.

Tabela 3 – Colorização incremental

I	Máscara	Bordas da Máscara	Imagem colorizada
0
2
5
9
30

Fonte: Elaborada pelo autor.

A complexidade dessa etapa depende da quantidade q de pixels inválidos e da complexidade para se colorizar um único pixel que, como vimos na seção anterior, é $O((b_{col}^2 + b_{max}^2)/2i)$. Desta forma, a complexidade final da etapa de colorização é $O(q(b_{col}^2 + b_{max}^2)/2i)$.

4.3 Considerações finais

Neste Capítulo foram apresentadas as etapas da técnica PixRadio. A Seção 4.1 mostrou como é feita a obtenção das disparidades e a reversão dos pixels do anáglifo com base nessas disparidades. Destaca-se a Seção 4.1.2 que apresentou o vetor de representações que a técnica utiliza para lidar com o problema das diferenças radiométricas presentes entre os canais do anáglifo. A Seção 4.2 mostrou como é feita a colorização dos pixels que não possuem correspondência definidas. A complexidade total da técnica é dada pela soma entre as complexidades dessas duas etapas, sendo: $O(N + q(b_{col}^2 + b_{max}^2)/2i)$. Levando se em consideração que os parâmetros b_{col} , b_{max} e i são constantes e que o valor de q é proporcional ao número de pixels da imagem N , temos que a complexidade total da técnica é $O(N)$.

AVALIAÇÃO EXPERIMENTAL

Este capítulo apresenta os experimentos realizados para avaliar a técnica PixRadio. A Seção 5.1 realiza uma análise sobre as imagens do *dataset* utilizado, buscando identificar e quantificar regiões que frequentemente são apontadas como problemáticas para as técnicas de reversão anaglífica: regiões oclusas, de baixa textura e de grandes diferenças radiométricas. A Seção 5.2 apresenta uma comparação objetiva entre a PixRadio e as técnicas do estado da arte descritas no Capítulo 3. Nessa comparação as técnicas também são avaliadas quanto à sua capacidade de realizar a reversão direta e a colorização. Já a Seção 5.3 apresenta os testes subjetivos realizados comparando as técnicas.

5.1 Base de Imagens

A avaliação da técnica desenvolvida foi realizada utilizando três conjuntos de imagens estereoscópicas fornecidos pelo dataset *Middlebury*¹ lançadas nos anos 2001, 2003 e 2005. O principal motivo que levou a adoção dessas imagens foi o fato de elas serem muito utilizadas na área de correspondência estéreo e possuírem *ground-truth* dos mapas de disparidades, o que possibilita uma análise mais detalhada da técnica desenvolvida. Além disso, algumas dessas imagens foram utilizadas por trabalhos relacionados, como os de Joulin (JOULIN; KANG, 2013), Williem (WILLIEM; RASKAR; PARK, 2015) e Kunze (KUNZE, 2019), o que padroniza e facilita o processo de comparação entre as técnicas.

Cada imagem do *dataset* da *Middlebury* pode variar em alguns aspectos como resolução, iluminação, exposição e posicionamento de câmera. A Tabela 4 especifica a base de imagens utilizada. Nesta tabela, as colunas *Esq* e *Dir* indicam o nome da variante da imagem selecionada para os canais esquerdo e direito. A coluna *Disp* indica se a imagem possui *ground-truth* dos mapas de disparidades. As imagens do dataset são disponibilizadas em diversos formatos e foram

¹ <https://vision.middlebury.edu/stereo/data/>

convertidas para o formato .bmp antes de serem utilizadas. Destaca-se, também, que as variantes selecionadas são as mesmas utilizadas por trabalhos relacionados.

Tabela 4 – Imagens selecionadas do *dataset Middlebury*

Nome	Ano	Resolução	Esq	Dir	Disp
Tsukuba	2001	384 x 288	col2	col3	Não
Venus	2001	434 x 383	im2	im6	Sim
Sawtooth	2001	434 x 380	im2	im6	Sim
Bull	2001	433 x 381	im2	im6	Sim
Poster	2001	435 x 383	im2	im6	Sim
BarnOne	2001	432 x 381	im2	im6	Sim
BarnTwo	2001	430 x 381	im2	im6	Sim
Cones	2003	450 x 375	im2	im6	Sim
Teddy	2003	450 x 375	im2	im6	Sim
Art	2005	695 x 555	im1	im5	Sim
Books	2005	695 x 555	im1	im5	Sim
Dolls	2005	695 x 555	im1	im5	Sim
Laundry	2005	671 x 555	im1	im5	Sim
Moebius	2005	695 x 555	im1	im5	Sim
Reindeer	2005	671 x 555	im1	im5	Sim
Computers	2005	665 x 555	im1	im5	Não
Drumsticks	2005	695 x 555	im1	im5	Não
Dwarves	2005	695 x 555	im1	im5	Não

Fonte: Elaborada pelo autor.

Como discutido na Seção 3.1 as técnicas de reversão anaglífica baseadas em informações internas ao anáglifo têm dificuldade em recuperar pixels em regiões *occlusas*, regiões de *baixa textura* e em regiões com altas *diferenças radiométricas*. Assim, foram realizadas análises para identificar tais regiões nas imagens do *dataset*. Essas análises permitem a criação de máscaras para as imagens que são usadas para comparar a qualidade de imagem entre a PixRadio e as técnicas relacionadas. Essas análises são descritas nas próximas subseções.

5.1.1 Identificação de oclusão

A *occlusão* ocorre quando uma região é visível em apenas uma das imagens do par estéreo. É possível determinar regiões oclusas nas imagens do *dataset* utilizando-se os mapas de disparidade *ground truth* pela verificação de reciprocidade (detalhado na Seção 4.1.6). Como resultado são obtidas máscaras binárias que determinam as regiões oclusas das imagens no par estéreo. As Figuras 44a e 44b apresentam os mapas de disparidades *ground truth* para as visões esquerda D_{gtE} e direita D_{gtD} de uma das imagens do *dataset*. A verificação de reciprocidade gera *máscaras de oclusão* para as visões esquerda M_{oE} e direita M_{oD} que indicam quais pixels são visíveis e quais estão oclusos. As Figuras 44c e 44d apresentam tais máscaras. Pixels oclusos são indicados com o valor 0, então todas as regiões em preto nessas figuras são consideradas oclusas.

Figura 44 – Mapas de disparidades e máscaras de oclusão

Fonte: Elaborada pelo autor.

As máscaras de oclusão são importantes pois elas dizem quais regiões possuem correspondência e quais regiões não as possuem. Isso permite determinar quais pixels das imagens são passíveis de recuperação via correspondência direta e quais somente poderão ser recuperados via colorização. As máscaras de oclusão podem ser aplicadas às suas respectivas imagens do par estéreo, permitindo a identificação das regiões oclusas e não-occlusas dessas imagens. As figuras 45a e 45b exemplificam a aplicação da máscara de oclusão em uma imagem da visão esquerda. Todas as regiões na Figura 45a possuem correspondência determinada na imagem da visão direita, enquanto todas as regiões que podem ser vistas na Figura 45b não possuem correspondência por estarem oclusas. Uma técnica de correspondência anaglífica *ideal* deveria ser capaz encontrar todas as correspondências entre as visões esquerda e direita, gerando uma imagem como essa. Entretanto, toda técnica enfrenta dificuldades ao buscar correspondências, principalmente em regiões de baixa textura e grandes diferenças radiométricas e, assim, as correspondências finais encontradas podem variar.

Figura 45 – Exemplo da identificação de regiões oclusas

Fonte: Elaborada pelo autor.

Para a análise geral do *dataset* foram obtidas as máscaras de oclusão de todas as imagens que possuem *ground truth* do mapa de disparidades e foi calculada a proporção de pixels oclusos e não-occlusos para cada uma delas. Estes dados são exibidos na Tabela 5. Essa análise mostra que pixels oclusos correspondem, em média, a 12% dos pixels dessas imagens. Assim, considerando-se apenas a oclusão, podemos estimar que cerca de 12% dos pixels não poderão ser recuperados por correspondência direta por não existir correspondência entre as visões do par estéreo. Estes pixels têm grande probabilidade de serem recuperados via colorização.

Tabela 5 – Identificação de oclusão

Imagem	Esquerda		Direita	
	Não-ocluso	Ocluso	Não-ocluso	Ocluso
Venus	96%	4%	97%	4%
Cones	85%	15%	85%	15%
Teddy	87%	13%	89%	13%
Sawtooth	95%	5%	95%	5%
Bull	98%	2%	98%	2%
Poster	95%	5%	96%	5%
BarnOne	96%	4%	96%	4%
BarnTwo	96%	4%	96%	4%
Art	73%	27%	73%	27%
Books	87%	13%	86%	13%
Dolls	84%	16%	84%	16%
Laundry	81%	19%	81%	19%
Moebius	85%	15%	85%	15%
Reindeer	80%	20%	78%	20%
Média	88,4%	11,6%	88,5%	11,5%

Fonte: Elaborada pelo autor.

5.1.2 Identificação de textura

A *textura* fornece informações sobre a disposição espacial de cores ou de intensidades em uma determinada região de uma imagem. Regiões com cores contínuas e com baixa variação de intensidade são consideradas regiões de baixa textura. Todas as técnicas de reversão anaglífica enfrentam dificuldades de realizar a reversão nessas regiões, pois a baixa textura dificulta o processo de comparação entre os pixels, atrapalhando a correspondência anaglífica. Assim torna-se importante identificar e analisar a ocorrência de tais regiões nas imagens do *dataset*.

A identificação de textura é realizada criando-se *mapas de texturas* para as visões esquerda T_E e direita T_D do par estéreo. Os valores desse mapa que estão abaixo de um limiar α são considerados pixels em regiões de baixa textura. Os resultados da comparação com o limiar são armazenados em matrizes denominadas *máscaras de textura* das visões esquerda M_{tE} e direita M_{tD} .

Figura 46 – Mapas de texturas e máscaras de baixas texturas

Fonte: Elaborada pelo autor.

No procedimento para a obtenção do mapa de texturas, converte-se as imagens do par

estéreo em escala de cinza. Em seguida, para cada pixel (x, y) é definido um bloco de tamanho b , contendo o pixel (x, y) ao centro. Então calcula-se o *desvio padrão* entre todos os valores de intensidade dos pixels dentro desse bloco e esse valor indica a textura da região do pixel (x, y) . Valores pequenos de desvio padrão indicam baixa textura e valores grandes indicam alta textura. As Figuras 46a e 46b apresentam, respectivamente, os mapas de textura das visões esquerda e direita do par estéreo de uma das imagens do *dataset* criados utilizando-se uma janela $b = 13$. Essas figuras possibilitam a inspeção visual das texturas da imagem e são obtidas pela conversão dos valores de desvio padrão contidos no mapa de textura para o intervalo de 0-255. Para essa conversão, adota-se que o menor valor encontrado no mapa de texturas equivale ao valor 0 e o maior valor o valor 255, ajustando-se proporcionalmente os valores intermediários. Regiões escuras indicam baixa textura e regiões claras indicam alta textura.

Após a obtenção dos mapas de textura, é possível obter as *máscaras* que indicam quais pixels possuem baixa textura utilizando-se a equação 5.1 para todos os pixels dos mapas de textura T_E da visão esquerda. O mesmo pode ser feito para a visão direita.

$$M_{tE}(x, y) = \begin{cases} 1, & \text{se } T_E(x, y) \leq \alpha \\ 0, & \text{caso contrário.} \end{cases} \quad (5.1)$$

As Figuras 44c e 44d apresentam tais máscaras. Pixels com baixa textura são indicados com o valor 0, então todas as regiões em preto nessas figuras são regiões que possuem textura abaixo do limiar α .

Assim como foi feito com as máscaras de oclusão, as máscaras de textura também podem ser aplicadas às suas respectivas imagens do par estéreo, permitindo a inspeção visual das regiões de baixa textura dessas imagens. As Figuras 47a e 47b exemplificam a aplicação da máscara de textura em uma imagem da visão esquerda. Todas as regiões na Figura 47a possuem textura acima do limiar α , enquanto todas as regiões que podem ser vistas na Figura 47b possuem textura abaixo do limiar α .

Para a análise geral do *dataset* foram obtidas as máscaras de textura para todas as imagens que possuem *ground truth* do mapa de disparidades e foi calculada a proporção entre pixels que possuem textura maior e menor que o limiar $\alpha = 1$, considerando-se um bloco de tamanho $b = 13$. Em outras palavras, serão considerados como sendo pixels de baixa textura aqueles que possuem desvio padrão menor que 1². Esses dados são exibidos na Tabela 6. Essa análise mostra que pixels com baixa textura correspondem em média a 1,5% dos pixels dessas imagens.

² Os valores de α e b foram obtidos empiricamente. Foram testadas diversas combinações para esses valores e foram selecionados os valores que encontram corretamente regiões de baixa textura.

Figura 47 – Exemplo da identificação de regiões de baixa textura ($\alpha = 3, b = 13$) Tabela 6 – Identificação de textura ($\alpha = 1, b=13$)

	Esquerda		Direita	
Imagem	$> \alpha$	$\leq \alpha$	$> \alpha$	$\leq \alpha$
Venus	98%	2%	98%	2%
Cones	100%	0%	100%	0%
Teddy	99%	1%	97%	3%
Sawtooth	100%	0%	100%	0%
Bull	100%	0%	100%	0%
Poster	99%	1%	99%	1%
BarnOne	100%	0%	100%	0%
BarnTwo	100%	0%	100%	0%
Art	99%	1%	99%	1%
Books	98%	2%	97%	3%
Dolls	99%	1%	99%	1%
Laundry	97%	3%	96%	4%
Moebius	96%	4%	97%	3%
Reindeer	100%	0%	97%	3%
Média	98,9%	1,1%	98,5%	1,5%

Fonte: Elaborada pelo autor.

5.1.3 Identificação de diferenças radiométricas

Parte do processo de reversão anaglífica é realizado via correspondência anaglífica, que é obtida a partir da comparação entre a intensidade dos pixels em diferentes canais de cor do anáglifo. As regiões de uma imagem que possuem cores primárias ou muito saturadas possuem distribuição de intensidades muito diferentes nos canais RGB.

A Figura 48 apresenta os canais vermelho da visão esquerda e verde da visão direita de um dos pares estéreos do *database*. É fácil notar que esses canais possuem diferentes intensidades em pixels correspondentes. Tomando como exemplo o urso de pelúcia apresentado no canto superior

Figura 48 – Exemplo de diferenças radiométricas

Fonte: Elaborada pelo autor.

direito dessas imagens: no canal vermelho (48a) a intensidade de seus pixels é visivelmente maior que no seu canal verde (48b). Essas *diferenças radiométricas* atrapalham o processo de correspondência estéreo e, conseqüentemente, a reversão anaglífica. Assim torna-se importante identificar e analisar tais regiões nas imagens do *dataset*.

Essa análise é feita por meio da criação de *mapas de diferenças radiométricas* para as visões esquerda D_{rE} e direita D_{rD} . Os valores desses mapas que estão acima de um limiar β são considerados pixels em regiões de grandes diferenças radiométricas. Os resultados da comparação com o limiar são armazenados em matrizes denominadas *máscaras de diferenças radiométricas* das visões esquerda M_{drE} e direita M_{rdD} .

O mapa de diferenças radiométricas da visão esquerda pode ser obtido pela aplicação da equação 5.2 para todos os pixels (x, y) da V_E . Nota-se que essa equação realiza a subtração do valor de intensidade do pixel (x, y) no canal vermelho da V_E da intensidade de seu pixel correspondente no canal verde da V_D . A disparidade desse pixel é obtida pelo mapa de disparidades D_E .

$$D_{rE}(x, y) = |Cn_{vmE}(x, y) - Cn_{cD}(x - D_E(x, y), y)| \quad (5.2)$$

As Figuras 49a e 49b apresentam, respectivamente, os mapas de diferenças radiométricas das visões esquerda e direita do par estéreo de uma das imagens do *dataset*. Essas figuras possibilitam a inspeção visual das diferenças entre os canais da imagem e são obtidas convertendo-se os valores contidos no mapa de diferenças para o intervalo de 0-255. Regiões claras indicam alta diferença radiométrica.

Após a obtenção dos mapas, são criadas *máscaras* que indicam quais pixels possuem alta diferença radiométrica aplicando-se a equação 5.3 para todos os valores do mapa da imagem

Figura 49 – Mapas e máscaras de diferença radiométrica

esquerda. O mesmo pode ser realizado para o mapa da imagem direita.

$$M_{drE}(x,y) = \begin{cases} 1, & \text{se } D_{rE}(x,y) \geq \beta \\ 0, & \text{caso contrário.} \end{cases} \quad (5.3)$$

As Figuras 49c e 49d apresentam as máscaras obtidas. Pixels com altas diferenças radiométricas são indicados com o valor 0, então todas as regiões em preto nessas figuras são regiões que possuem diferenças radiométricas acima do limiar β .

Assim como foi feito com as máscaras de oclusão e de textura, as máscaras de diferenças radiométricas também podem ser aplicadas às suas respectivas imagens do par estéreo, permitindo a inspeção visual das regiões afetadas. As Figuras 50a e 50b exemplificam a aplicação da máscara de diferenças radiométricas em uma imagem da visão esquerda. Todas as regiões na Figura 50b possuem diferenças radiométricas acima do limiar β , enquanto todas as regiões que podem ser vistas na Figura 50a possuem diferenças radiométricas abaixo do limiar β .

Para a análise geral do *dataset* foram obtidas as máscaras para todas as imagens que

Figura 50 – Exemplo da identificação de regiões de grandes diferenças radiométricas

Fonte: Elaborada pelo autor.

possuem *ground truth* do mapa de disparidades e foi calculada a proporção entre pixels que possuem diferenças maiores e menores que o limiar $\beta = 64^3$. Esses dados são exibidos na Tabela 7. Pixels com altas diferenças radiométricas correspondem em média a 8% dessas imagens.

Tabela 7 – Identificação de diferenças radiométricas ($\beta = 64$)

Nome	Esquerda		Direita	
	$< \beta$	$\geq \beta$	$< \beta$	$\geq \beta$
Venus	95%	5%	95%	5%
Cones	81%	19%	81%	19%
Teddy	82%	18%	82%	18%
Sawtooth	99%	1%	99%	1%
Bull	98%	2%	98%	2%
Poster	80%	20%	80%	20%
BarnOne	92%	8%	92%	8%
BarnTwo	97%	3%	97%	3%
Art	97%	3%	97%	3%
Books	97%	3%	97%	3%
Dolls	87%	13%	87%	13%
Laundry	97%	3%	96%	4%
Moebius	90%	10%	90%	10%
Reindeer	96%	4%	96%	4%
Média	92%	8%	92%	8%

Fonte: Elaborada pelo autor.

³ O valor de β foi obtido empiricamente. Dentre diversos valores testados concluiu-se que $\beta = 64$ identifica corretamente regiões de grandes diferenças radiométricas.

5.2 Avaliação Objetiva

Esta seção descreve a comparação da técnica PixRadio com as técnicas de [Joulin e Kang \(2013\)](#), [Williem, Raskar e Park \(2015\)](#) e [Kunze \(2019\)](#), que são as principais técnicas baseadas em informações internas ao anáglifo que se encontram no estado da arte. Como uma simplificação, para facilitar a leitura, nesta Seção tais trabalhos serão citados, respectivamente, como Joulin, Williem e Kunze. Tais técnicas foram devidamente descritas na Seção 3.1. Para a execução dos testes foram utilizadas as imagens obtidas diretamente pela execução das implementações que os autores forneceram de suas técnicas.

5.2.1 Par estéreo recuperado

O primeiro teste foi realizado para a comparação da qualidade objetiva das imagens geradas pelas técnicas por meio do cálculo do PSNR para as imagens esquerda e direita do par estéreo recuperado por cada uma das técnicas. O PSNR foi adotado por ter sido utilizado por todos os trabalhos relacionados. Assim, sua adoção facilita o processo de comparação entre as técnicas. É importante ressaltar diversas outras métricas objetivas podem ser encontradas na literatura. Mais informações sobre o PSNR podem ser encontradas na Seção 2.5.2,

Tabela 8 – PSNR dos canais esquerdo e direito recuperados

Imagem	Esquerda				Direita			
	Joulin	Williem	Kunze	Costa	Joulin	Williem	Kunze	Costa
Art	23,37	22,92	22,65	23,56	29,32	29,84	26,49	29,35
BarnOne	26,99	30,61	29,68	30,59	27,91	33,08	30,32	32,62
BarnTwo	30,33	30,87	31,36	32,43	32,33	35,50	33,34	36,17
Books	25,87	22,14	27,47	26,23	32,40	30,23	29,06	31,70
Bull	31,54	32,29	33,64	33,54	33,68	37,03	37,48	37,22
Computer	26,03	25,61	26,42	26,45	32,06	29,46	28,39	30,21
Cones	21,54	24,26	23,84	22,74	24,54	26,84	25,72	25,78
Dolls	27,28	24,11	24,27	27,28	29,62	27,57	28,28	25,29
Drumsticks	25,92	22,88	22,46	26,30	28,28	25,78	24,14	28,47
Dwarves	29,02	25,88	23,48	28,05	29,89	30,06	26,00	29,29
Laundry	20,89	20,40	22,62	22,41	27,04	28,63	28,06	29,68
Moebius	26,26	23,34	26,01	26,83	27,94	26,95	28,02	28,42
Poster	24,85	28,48	27,86	28,06	26,40	32,03	30,86	30,07
Reindeer	28,06	27,84	22,44	29,15	32,24	26,93	24,63	28,59
Sawtooth	29,53	29,44	29,57	29,70	32,49	34,37	32,89	33,15
Teddy	21,27	23,35	24,65	26,81	24,70	30,71	29,81	30,08
Tsukuba	30,83	32,83	30,51	30,30	32,89	33,98	30,55	31,35
Venus	29,80	31,48	31,94	32,30	32,58	35,02	34,09	34,86
Média	26,63	26,60	26,72	27,93	29,80	30,78	29,34	30,68

Fonte: Elaborada pelo autor.

A Tabela 8 apresenta os resultados obtidos. Nessa tabela a primeira coluna indica o nome da imagem do *dataset*. As demais colunas indicam o valor de PSNR para cada uma das técnicas avaliadas em *decibéis* (dB). As cores das células variam em função do valor do PSNR dentro de uma visão (esquerda ou direita). A cor branca é atribuída ao menor valor de PSNR e a cor verde escura ao maior valor de PSNR. Tons intermediários de verde indicam, proporcionalmente, valores intermediários. Isso implica que cores próximas indicam valores próximos. Também é marcada como **negrito** a entrada de valor máximo. Como exemplo, veja os resultados para a *visão esquerda* da imagem *Art*. O melhor resultado de PSNR foi obtido pela técnica PixRadio (Costa - 23,56), seguido pela técnica de Joulin (23,37), Williem (22,92) e, por fim, Kunze (22,65).

A codificação por cores na tabela permite perceber a predominância de bons resultados para a PixRadio na visão esquerda, o que é confirmado pelo valor da média obtida (27,93). Essa é a maior das médias, seguida pelas técnicas de Kunze (26,72), Joulin (26,63) e Williem (26,72). Na visão direita, a técnica com maior média é a de Williem (30,78), seguida pela PixRadio (30,68), Joulin (29,80) e Kunze (29,34). A diferença entre a PixRadio e o segundo lugar na visão esquerda é de 1,21 pontos e na visão direita é de -0,10 pontos.

É importante destacar que a qualidade da imagem recuperada para a visão esquerda naturalmente tende a ser menor que a qualidade obtida pela imagem da visão direita. Isso ocorre porque as técnicas têm que recuperar dois canais da visão esquerda (verde e azul) e apenas um da visão direita (vermelho).

Os resultados apresentados na Tabela 8 permitem a avaliação de cada técnica para as duas visões do par estéreo de modo independente. Entretanto, essa análise dificulta determinar o desempenho geral das técnicas: para a visão esquerda a tabela indica que a melhor técnica foi a proposta, enquanto para a visão direita a tabela indica que a melhor técnica é a de Williem. Para avaliar as visões dos pares estéreo recuperados de uma maneira unificada, a Tabela 9 apresenta os valores de PSNR médios⁴ entre as duas visões. Nessa tabela é possível perceber que a PixRadio obteve melhores resultados de qualidade objetiva (média de 28,98), seguida pelas técnicas de Williem (28,08), Joulin (27,85) e Kunze (27,78). Percebe-se, ainda, que a técnica PixRadio obteve o melhor valor de PSNR em 7 das 18 imagens (contra 4/18 de Joulin, 5/18 de Williem e 2/18 de Kunze). Percebe-se também que em nenhuma imagem a técnica PixRadio obteve o pior valor de PSNR.

Para analisar a relevância dos resultados obtidos foi realizado o *teste t de Student*. Esse teste verifica se as médias dos valores de PSNR obtidos por cada uma das técnicas concorrentes são de fato diferentes da média obtida pela PixRadio. A aplicação desse teste requer que 2 premissas sejam verdadeiras:

⁴ O PSNR médio entre duas imagens é obtido calculando-se o MSE médio entre as imagens originais e recuperadas utilizando-se a Equação 2.5 e em seguida aplicando-se a Equação 2.6, ambas apresentadas na seção 2.5.2

Tabela 9 – PSNR médio entre as visões esquerda e direita

	Joulin	Williem	Kunze	Costa
Art	25,39	25,12	24,16	25,56
BarnOne	27,43	31,67	29,99	31,48
BarnTwo	31,21	32,60	32,23	33,91
Books	28,01	24,52	28,19	28,16
Bull	32,48	34,04	35,15	35,00
Computer	28,08	27,12	27,30	27,94
Cones	22,79	25,36	24,68	24,00
Dolls	28,29	25,51	25,83	26,18
Drumsticks	26,94	24,09	23,22	27,25
Dwarves	29,43	27,48	24,56	28,63
Laundry	22,96	22,80	24,54	24,67
Moebius	27,02	24,78	26,90	27,55
Poster	25,55	29,90	29,10	28,95
Reindeer	29,67	27,36	23,40	28,86
Sawtooth	30,76	31,24	30,92	31,09
Teddy	22,65	25,62	26,51	28,14
Tsukuba	31,74	33,37	30,53	30,79
Venus	30,97	32,90	32,88	33,39
Média	27,85	28,08	27,78	28,98

Fonte: Elaborada pelo autor.

1. As duas amostras comparadas devem ter distribuição normal;
2. As duas amostras devem ter mesma variância;

A primeira premissa pode ser verificada pela aplicação do teste de [SHAPIRO e WILK \(1965\)](#). Nesse teste adota-se como hipótese nula que os dados possuem distribuição amostral normal. Caso o *p-value* obtido pelo teste seja menor ou igual a 0,05, rejeita-se a hipótese nula e conclui-se que a distribuição amostral não é normal. A Tabela 10 mostra que todas as amostras possuem *p-value* acima de 0,05. Assim, pode-se considerar que as amostras passaram nesse teste e que as distribuições amostrais possuem distribuição normal.

Tabela 10 – Teste de Shapiro-Wilk

Amostra	<i>p-value</i>
Joulin	0,324
Williem	0,061
Kunze	0,318
Costa	0,631

A segunda premissa pode ser avaliada com o teste de [Levene \(1960\)](#). Nesse teste as amostras são comparadas aos pares com a hipótese nula de que elas foram retiradas de populações

com variâncias iguais. Caso o p -value obtido pelo teste seja menor ou igual a 0,05, rejeita-se a hipótese nula e conclui-se que as amostras não foram retiradas de populações com variâncias iguais. Veja na Tabela 11 que todos os testes possuem p -value acima de 0,05. Assim, pode-se considerar as amostras passaram nesse teste e que a variância entre a amostra da PixRadio e as demais é igual.

Tabela 11 – Teste de Levene

Amostra	p-value
Costa x Joulin	0,958
Costa x Williem	0,166
Costa x Kunze	0,378

Aplicou-se o teste t de Student para amostras pareadas. Nesse teste adota-se como hipótese nula que as amostras possuem médias iguais. Caso o p -value obtido seja menor que 0,05 rejeita-se a hipótese nula e conclui-se que as amostras possuem médias distintas.

Tabela 12 – Teste de t de Student

Amostra	p-value
Costa x Joulin	0,028
Costa x Williem	0,030
Costa x Kunze	0,008

A Tabela 12 mostra que todos os p -value obtidos estão abaixo do limiar de 0,05, então em todos os casos podemos rejeitar a hipótese nula e concluir que a PixRadio é, de fato, superior em qualidade objetiva que todas as técnicas concorrentes. Entretanto, essa análise leva em conta a qualidade da imagem *toda* e não nos permite concluir que a melhoria na qualidade ocorreu devido à melhorias implementadas na etapa de correspondência estéreo, com a adoção do descritor de pixels, ou se essa melhoria ocorreu devido à técnica de colorização. Para realizar essa análise de forma distinta é necessário remover a influência da etapa de colorização e analisar apenas as regiões que foram recuperadas via correspondência direta.

5.2.2 Recuperação por correspondência direta

Para a análise da capacidade das técnicas de realizar a reversão anaglífica via *correspondência direta* podemos considerar apenas as regiões não-occlusas das máscaras de oclusão M_{oE} e M_{oD} , apresentadas na Seção 5.1.1. Essas máscaras permitem determinar quais regiões são passíveis de serem recuperadas via recuperação direta e em quais regiões isso é impossível. Isso elimina a influência da etapa de colorização e permite avaliar melhor a influência da técnica de correspondência anaglífica, foco da técnica PixRadio. Assim, foi realizado o cálculo do PSNR aplicando-se as máscaras de oclusão nas imagens obtidas por cada uma das técnicas. Esse cálculo considera apenas os pixels cujos valores correspondentes na máscara são iguais a 1, ou seja, regiões não-occlusas. A Tabela 13 apresenta os resultados obtidos nessa análise. Nota-se que a

PixRadio é melhor em realizar a reversão por correspondência direta tanto no canal esquerdo quanto no direito do par estéreo.

Assim como na seção anterior, também foram obtidos os valores de PSNR médio entre as duas visões. Estes valores são apresentados na Tabela 14. Nessa tabela a PixRadio também apresenta melhores resultados que as concorrentes no que se trata da capacidade de recuperar regiões por correspondência direta. A relevância desses resultados é confirmada ao se realizar o teste *t de student* detalhado na Tabela 15.

Essa análise mostra que a técnica PixRadio apresenta resultados superiores em qualidade objetiva nas regiões recuperadas por recuperação direta, o que prova, indiretamente, que o método de recuperação proposto proposto leva a melhores correspondências.

Tabela 13 – PSNR das visões esquerda e direita recuperadas por recuperação direta

Imagem	PSNR (db)							
	Esquerda				Direita			
	Joulin	Wiliem	Kunze	Costa	Joulin	Wiliem	Kunze	Costa
Art	31,64	30,39	30,28	31,77	34,61	36,24	32,98	36,65
BarnOne	31,98	36,92	34,93	36,30	32,72	38,60	36,01	38,15
BarnTwo	35,39	36,54	36,96	38,01	37,08	40,58	38,31	41,45
Books	32,59	31,56	33,60	33,70	38,23	37,23	36,99	37,89
Bull	36,60	38,16	39,09	38,89	38,43	41,99	42,85	42,08
Cones	27,62	30,64	30,82	30,82	29,99	34,07	32,82	32,75
Dolls	34,17	31,64	31,38	34,93	37,89	35,20	35,82	37,45
Laundry	29,93	30,69	30,69	33,41	31,70	35,09	35,03	37,11
Moebius	32,19	31,85	33,90	34,07	34,49	36,37	34,88	34,99
Poster	29,85	34,34	33,58	33,61	31,10	37,60	35,78	35,00
Reindeer	35,94	34,33	29,20	37,12	38,51	33,50	31,89	38,74
Sawtooth	34,89	36,01	35,46	35,35	37,47	39,94	38,09	38,67
Teddy	27,10	32,04	31,55	33,63	29,18	36,53	35,42	36,26
Venus	34,89	37,45	37,64	37,71	37,40	40,51	39,22	39,85
Média	32,48	33,75	33,51	34,95	34,91	37,39	36,15	37,65

Fonte: Elaborada pelo autor.

Tabela 14 – PSNR médio entre as visões esquerda e direita recuperadas por recuperação direta

PSNR (dB)				
Imagem	Joulin	Williem	Kunze	Costa
Art	32,88	32,40	31,42	33,56
BarnOne	32,34	37,68	35,44	37,13
BarnTwo	36,15	38,11	37,59	39,40
Books	34,54	33,51	34,96	35,29
Bull	37,42	39,66	40,58	40,20
Cones	28,64	32,03	31,71	31,68
Dolls	35,64	33,07	33,06	36,01
Laundry	30,73	32,36	32,35	34,88
Moebius	33,19	33,55	34,36	34,51
Poster	30,43	35,68	34,55	34,25
Reindeer	37,02	33,90	30,32	37,85
Sawtooth	35,99	37,55	36,58	36,70
Teddy	28,03	33,74	33,08	34,76
Venus	35,97	38,72	38,36	38,65
Média	33,50	35,14	34,60	36,06

Fonte: Elaborada pelo autor.

Tabela 15 – p-value para os testes de relevância

Teste de relevância				
p-value	Joulin	Williem	Kunze	Costa
Shapiro-Wilk	0,242	0,054	0,868	0,962
Levene (x Costa)	0,222	0,345	0,560	-
t-Student (x Costa)	0,000	0,042	0,020	-

5.2.3 Recuperação por colorização

A análise apresentada na seção anterior permitiu a comparação do desempenho da PixRadio com as demais técnicas do estado da arte nas regiões das imagens em que a *recuperação direta* é possível. Nesta seção essa comparação será repetida, entretanto dessa vez serão consideradas as regiões dos *mapas de oclusão* que possuem valor igual a 0. Em outras palavras, serão analisadas as regiões que possuem tendência de serem recuperadas via *colorização*, onde a recuperação direta não é possível, que são regiões onde ocorre a *occlusão*.

A Tabela 16 apresenta os valores médios de PSNR entre as imagens esquerda e direita. Nessa tabela a técnica de Joulin aparenta apresentar melhores resultados, seguida pela técnica PixRadio, Kunze e Williem. Entretanto, ao se realizar o teste de *t de student* (apresentado na Tabela 17) para verificar a relevância dos resultados, verifica-se que a diferença entre a qualidade obtida pela técnica de colorização proposta e a técnica de colorização utilizada por Joulin não

são estatisticamente relevantes ($p\text{-value} = 0,169$). O mesmo pode ser concluído comparando-se a PixRadio com Kunze ($p\text{-value} = 0,089$). Assim, somente podemos afirmar que a PixRadio é melhor que a técnica de Williem no que se refere à colorização ($p\text{-value} = 1\%$) e que a PixRadio possui médias iguais às médias obtidas por Joulin e Kunze.

Tabela 16 – PSNR médio entre as visões esquerda e direita recuperadas por colorização

PSNR (dB)				
Imagem	Joulin	Williem	Kunze	Costa
Art	26,62	26,49	25,53	26,48
BarnOne	29,83	28,09	28,31	29,01
BarnTwo	33,01	30,39	30,79	31,80
Books	27,92	22,43	27,79	27,37
Bull	32,88	29,17	31,17	32,50
Cones	24,06	25,38	24,31	23,05
Dolls	27,86	24,90	25,50	24,36
Laundry	22,79	21,73	24,34	23,38
Moebius	27,77	23,20	26,17	27,30
Poster	28,36	27,05	27,51	28,10
Reindeer	30,10	28,55	24,20	28,35
Sawtooth	30,92	27,82	29,18	29,53
Teddy	24,55	23,51	25,89	27,47
Venus	31,97	29,27	30,58	32,25
Média	28,47	26,28	27,23	27,93

Fonte: Elaborada pelo autor.

Tabela 17 – p-value para os testes de relevância

Teste de relevância - Colorização				
p-value	Joulin	Williem	Kunze	Costa
Shapiro-Wilk	0,645	0,512	0,168	0,459
Levene (x Costa)	0,715	0,909	0,742	-
t-Student (x Costa)	0,169	0,010	0,089	-

Fonte: Elaborada pelo autor.

5.3 Avaliação Subjetiva

Como foi apresentado na Seção 2.5, medidas objetivas são baseadas em modelos matemáticos que calculam a qualidade da imagem. Apesar de serem rápidas de serem obtidas e facilmente reprodutíveis, as medidas objetivas não levam em conta as especificidades do

sistema visual humano, que é complexo. A fim de se avaliar a qualidade visual dos pares estéreo recuperados foi realizada uma análise subjetiva aplicando o método DSCQS (*Double Stimulus Continuous Quality Scale*), apresentado na Seção 2.5.1.

Devido às condições de distanciamento social impostas pela pandemia de Covid-19 não foi possível realizar a aplicação do DSCQS presencialmente, como é previsto na recomendação (ITU, 2002). Assim, desenvolveu-se uma plataforma web para execução dos testes DSCQS remotamente⁵. Essa plataforma implementa todos os passos contidos no teste presencial, o que inclui um questionário sobre o perfil do avaliador, testes de Daltonismo (teste de Ishihara (ISHIHARA; E.J., 1918)), uma etapa de treinamento para o teste e, por fim, o teste DSCQS. A plataforma guia os avaliadores durante todo o procedimento de teste, dando explicações e instruções para a correta utilização da ferramenta. A fim de limitar e controlar as condições de aplicação, a plataforma solicita aos avaliadores que se assegurem de algumas condições do ambiente de aplicação do teste, como o excesso de luz ambiente e se há reflexos em seu monitor. O teste também solicita que a sua execução seja feita em um ambiente confortável e livre de distrações, além de que o avaliador verifique o brilho, contraste e calibração de cores do seu monitor.

Além disso, o avaliador foi instruído a não aumentar o zoom do navegador em nenhum momento, a executar o teste em tela maximizada e a não realizar o teste em telas pequenas como as de smartphones e tablets. Para garantir tais condições, foram salvas informações sobre a tela do avaliador, o que permitiu eliminar respostas em que tais eventos ocorreram.

Uma preocupação sobre a realização do teste remotamente foi a confiabilidade dos resultados devido à diversidade de dispositivos usados na avaliação. Para analisar se essa variedade de dispositivos realmente tem efeito sobre os resultados e possibilitar a comparação do modo de aplicação remoto com o presencial, as imagens a serem testadas foram as mesmas escolhidas por Kunze (2019), que realizou o DSCQS de presencialmente no ano de 2019 para a sua técnica e as técnicas de Joulin e Williem. Assim, para o experimento foram escolhidas 9 imagens do lado esquerdo e 9 imagens do lado direito do *dataset* (listadas na Tabela 4). O teste foi aplicado apenas em metade das imagens recuperadas, pois a realização com ambas as imagens do par estéreo tornaria o teste muito longo e estaria fora das especificações do DSCQS (ITU, 2002).

A sessão de teste foi elaborada de modo que um determinado avaliador analisasse todas as 18 imagens de apenas uma das técnicas testadas. Essa técnica foi escolhida automaticamente e aleatoriamente pelo sistema. A plataforma ficou disponível para a execução durante aproximadamente 3 semanas e o link para a realização dos testes foi amplamente divulgado em redes sociais. Nesse período foram obtidas 135 respostas, das quais 30 foram eliminadas por estarem incompletas (casos em que os avaliadores desistiram do teste). Das 105 respostas restantes, avaliou-se o teste de daltonismo e realizou-se uma verificação quanto ao tamanho do display

⁵ O sistema pode ser acessado em <<http://dscqs.andreldc.com/>>

utilizado, assim como a ocorrência de eventos como redimensionamento da tela e zoom e mais 25 testes foram eliminados. Por fim restaram 80 testes válidos, sendo 20 testes de cada uma das 4 técnicas testadas.

Tabela 18 – MOS das imagens originais e recuperadas

Imagem	Visão	Joulin		Williem		Kunze		Costa	
		Orig	Rec	Orig	Rec	Orig	Rec	Orig	Rec
Books	Esquerda	4,26	3,88	4,69	3,48	4,64	3,37	4,45	2,94
Computer	Esquerda	4,28	3,83	4,50	2,60	4,46	1,83	4,55	2,65
Drumsticks	Esquerda	4,30	3,11	4,13	1,94	4,54	1,44	4,30	2,58
Moebius	Esquerda	3,71	3,27	4,26	2,72	3,96	2,04	4,16	2,46
Poster	Esquerda	4,29	4,19	4,56	3,96	4,67	3,04	4,30	3,75
Reindeer	Esquerda	4,11	3,12	4,42	2,52	4,30	2,09	4,38	3,06
Sawtooth	Esquerda	4,29	4,20	4,16	3,69	4,25	3,04	4,14	3,82
Tsukuba	Esquerda	3,83	3,59	4,40	4,04	4,25	2,73	4,15	3,65
Venus	Esquerda	4,41	4,29	4,58	4,29	4,37	3,58	4,49	4,29
Art	Direita	4,13	3,95	4,38	3,44	4,38	3,37	4,38	3,95
BarnOne	Direita	3,90	3,94	4,07	3,73	4,12	3,77	3,75	3,72
BarnTwo	Direita	3,62	3,82	3,69	3,42	3,46	3,65	3,56	3,54
Bull	Direita	4,37	4,26	4,21	4,19	4,18	4,25	4,21	4,37
Cones	Direita	4,32	4,19	4,51	4,26	4,70	4,00	4,39	4,16
Dolls	Direita	4,29	4,22	4,68	3,28	4,43	3,54	4,42	4,29
Dwarves	Direita	4,22	3,40	4,42	3,83	4,34	2,98	4,46	3,90
Laundry	Direita	4,02	3,65	3,96	3,56	4,34	2,94	4,29	3,38
Teddy	Direita	4,58	4,41	4,66	4,80	4,60	4,47	4,53	4,17
Média		4,16	3,85	4,35	3,54	4,33	3,12	4,27	3,59
Média Normalizada		5,00	4,63	5,00	4,07	5,00	3,60	5,00	4,20

Fonte: Elaborada pelo autor.

O MOS obtido pelas respostas é apresentado na Tabela 18. Nessa tabela, a primeira coluna identifica a imagem avaliada, a segunda coluna (V) identifica se a imagem é proveniente da visão esquerda (E) ou direita (D), enquanto as demais indicam os resultados da avaliação para as quatro técnicas, indicando o MOS para a imagem original (O) e para a imagem recuperada (R) pela técnica. Os resultados são resumidos nas linhas contendo a média geral para ambas as visões tanto para as imagens originais como para as recuperadas. As médias das imagens recuperadas mostram que as técnicas apresentam resultados que variam entre 3,85 (Joulin) e 3,12 (Kunze), estando todas entre os conceitos 3 - *regular* e 4 - *bom*.

Entretanto, também é possível notar que os avaliadores têm a tendência de subavaliar as imagens, pois mesmo imagens originais foram avaliadas com notas próximas ao conceito 4 - *bom*, sendo que poucas foram avaliadas próximas ao conceito 5 - *ótimo*.

Assim, a fim de se eliminar essa tendência de subavaliação foi realizado o cálculo de *médias normalizadas*. Para a obtenção dessa média, o valor da média da *imagem original* é considerado como sendo 5 - *ótimo* e o valor da média da imagem recuperada corrigido em função

desse valor. A Equação 5.4 descreve como obter a *média normalizada* M_n a partir das médias recuperada M_r e original M_o .

$$M_n = 5 * M_o / M_r \quad (5.4)$$

Analisando-se essas médias normalizadas percebe-se que técnica com melhor desempenho na avaliação subjetiva é a de Joulin e Kang (JOULIN; KANG, 2013) (4,63) seguida pela técnica proposta (4,20), Williem e Park (Williem; RASKAR; PARK, 2015) (4,07) e por fim Kunze et al. (KUNZE; GOULARTE; SOUSA, 2020) (3,60). Para verificar a relevância desses resultados, foi realizado o teste *t de student* entre os valores de MOS das imagens recuperadas, obtendo-se a Tabela 19, que aponta a superioridade da técnica de Joulin sobre a técnica proposta. Além disso, o teste de relevância não permite afirmar se a técnica proposta é superior à técnica de Williem e Park e, por fim, aponta a superioridade da técnica proposta sobre a técnica de Kunze et al.

É importante notar dois pontos. Primeiro, apesar de a técnica proposta estar numericamente empatada em segundo lugar com a técnica de Williem e Park, do ponto de vista da análise subjetiva as técnicas nas 3 primeiras posições possuem o mesmo conceito: 4 - *bom*. Segundo, apesar de a técnica de Joulin e Kang ter alcançado melhores valores de MOS, essa técnica não é fiel às cores originais, colorizando algumas regiões reconstruídas com a cor errada.

Tabela 19 – p-value para os testes de relevância - DSCQS/MOS

Teste de relevância - DSCQS/MOS				
p-value	Joulin	Williem	Kunze	Costa
Shapiro-Wilk	0,112	0,624	0,629	0,103
Levene (x Costa)	0,086	0,747	0,282	-
t-Student (x Costa)	0,017	0,608	0,000	-

Fonte: Elaborada pelo autor.

A fim de entender as diferenças entre a técnica PixRadio e a técnica que apresenta melhores resultados na avaliação subjetiva, que é a de Joulin, foi realizada uma análise visual em algumas imagens recuperadas por essas técnicas. A Figura 51 apresenta a visão direita da imagem *Cones* em sua versão original e nas versões recuperadas pelas técnicas comparadas. Comparado-se com a imagem original, é possível perceber que a técnica de Joulin gera uma distribuição de cores mais uniforme nas regiões colorizadas, causando menos desconforto visual. A PixRadio gera pequenas regiões com ligeiras alterações de cores que são perceptíveis. Por outro lado, é possível notar que a técnica de Joulin recupera algumas regiões com cores incorretas, como por exemplo as flores no canto superior esquerdo da imagem, que na imagem original são laranjas/amarelas e passaram a ser roxas na imagem recuperada por Joulin. Essa alteração na cor,

por ser uniforme, pode não ser percebida pelo observador, não afetando negativamente os testes subjetivos, mas é percebida em testes objetivos.

Figura 51 – Análise subjetiva detalhada da visão direita da imagem *Cones*

O mesmo ocorre na Figura 52, que apresenta a visão esquerda da imagem *Teddy* em sua versão original e nas versões recuperadas pela técnica PixRadio e de Joulin. Nota-se, no canto esquerdo, que parte do tecido que era branco com uma estampa de beringela na imagem original passou a ter cor predominantemente alaranjada na técnica de Joulin. Essa troca de cor também ocorre no quadro branco ao fundo. Novamente, pelo fato de as mudanças de cor nessa imagem acontecerem "suavemente" é possível que tais problemas não aparentem ser erros ao observador. Por outro lado, apesar de a imagem recuperada pela PixRadio não apresentar grandes regiões com cores inconsistentes, é possível notar algumas regiões com pequenos artefatos.

Figura 52 – Análise subjetiva detalhada da visão direita da imagem *Teddy*

5.4 Considerações finais

Nesse Capítulo foram apresentadas os testes executados para a avaliação da técnica PixRadio. A Seção 5.1 realizou uma análise sobre as imagens do *dataset* utilizado que identificou e quantificou regiões que frequentemente são apontadas como problemáticas para as técnicas de reversão anaglífica: regiões oclusas, de baixa textura e de grandes diferenças radiométricas.

Já a Seção 5.2 apresentou uma comparação objetiva entre a PixRadio e as técnicas do estado da arte descritas no Capítulo 3. Nessa comparação as técnicas foram avaliadas quanto à sua capacidade de realizar a reversão direta e a colorização e foi possível perceber que a PixRadio apresenta melhores resultados de PSNR que as demais técnicas. Já a Seção 5.3 apresentou os testes subjetivos realizados comparando as técnicas. Nessa avaliação a técnica ficou em segundo lugar.

CONCLUSÕES

Neste trabalho de mestrado foi proposta e desenvolvida uma nova técnica de reversão anaglífica que recupera uma aproximação do par estéreo utilizado para a criação do anáglifo.

A utilização do método anaglífico tem como vantagem o baixo custo para ser utilizado e a redução do volume de dados, o que facilita o armazenamento e a transmissão de conteúdo estereoscópico. Já a reversão anaglífica possibilita a visualização de imagens e vídeos anaglíficos por meio de outros métodos, permitindo a independência de métodos de visualização.

Considerando que um dos principais obstáculos enfrentados pelos métodos de reversão existentes são as diferenças radiométricas entre os canais de cor do anáglifo, que atrapalham a correspondência anaglífica, esse trabalho desenvolveu um método que é menos sensível a estas diferenças, baseado em um novo descritor de pixels que torna a técnica mais robusta aos problemas de baixa textura e grandes diferenças radiométricas.

Pode-se concluir que os objetivos do trabalho foram alcançados analisando-se os resultados da avaliação experimental conduzida utilizando-se o dataset *Middlebury*. Os resultados obtidos indicaram melhoria na qualidade objetiva (PSNR) quando comparada com trabalhos relacionados. Em adição, os resultados indicam que a melhoria observada se deve método de recuperação direta proposto, baseado em um descritor de pixels projetado para tratar diferenças radiométricas. Já a avaliação subjetiva indica que a PixRadio produz qualidade de imagem similar à dos trabalhos relacionados, ficando em segundo lugar, mas reproduzindo cores com mais fidelidade.

6.1 Contribuições

A principal contribuição desse trabalho é a técnica proposta, chamada PixRadio, para a reversão anaglífica utilizando apenas informações contidas no anáglifo, que utiliza um novo descritor baseado em codificação diferencial para representar a informação espacial na vizinhança

de um pixel. A técnica de colorização utilizada também possui características inovadoras, como o processo de colorização incremental, que ajuda na redução de ruídos. Também podem ser consideradas contribuições a implementação da técnica (disponível em <https://github.com/andreldc/anaglyphreversion>), que foi feita em módulos, permitindo que novas funcionalidades sejam adicionadas e testadas. Por fim, o sistema utilizado para a realização dos testes DSCQs será disponibilizado e poderá ser utilizado em pesquisas futuras.

Ainda como contribuições destaca-se: a formação de recurso humano qualificado em nível de mestrado; o aprofundamento do conhecimento do grupo de pesquisas sobre técnicas de reversão anaglíficas e stereo matching; a submissão de um artigo completo a evento da área: - Costa, A. L. D.; Rodrigues, F. M.; Goularte, R. 2021. PixRadio - a new anaglyphic reversion technique based on radiometric differences reduction. 8p. [Submetido para Web-Media 2021 em 24/05/2021]

6.2 Limitações e Trabalhos Futuros

Durante a realização deste trabalho de mestrado foi possível verificar algumas possibilidades para trabalhos futuros para melhorar a qualidade final do par estéreo e também para expandir as aplicações da técnica.

Conforme apresentado na Seção 5.2 a técnica de colorização adotada não foi capaz de ser melhor que as técnicas usadas pelos demais trabalhos em estado da arte. Assim um trabalho futuro poderia avaliar outras alternativas para as técnicas de colorização visando reduzir tanto o tempo de processamento quanto a qualidade da colorização, aumentando, consequentemente, a qualidade do par estéreo final.

Outro problema que poderá ser abordado em um trabalho futuro é a redução de artefatos presentes nas imagens recuperadas.

Este trabalho focou na reversão anaglífica de imagens. Outro trabalho futuro poderia aplicar a técnica proposta para a reversão de vídeos anaglíficos utilizando, além das relações entre a visão direita e esquerda, as relações existentes entre os quadros subsequentes de vídeo.

As técnicas de reversão anaglífica têm sido avaliadas utilizando-se PSNR e DSCQS. Essas técnicas de avaliação apenas comparam se a imagem recuperada é similar à imagem original, não levando em conta a percepção da sensação de tridimensionalidade obtida pelos observadores através das imagens recuperadas. Assim, trabalhos futuros poderiam realizar a avaliação das técnicas utilizando os métodos de visualização estereoscópicos citados na Seção 2.2.

REFERÊNCIAS

ANDRADE, L. A. D.; GOULARTE, R. Percepção Estereoscópica Anaglífica em Vídeos Digitais Comprimidos com Perda. **Proceedings of the XV Brazilian Symposium on Multimedia and the Web - WebMedia '09**, ACM Press, New York, New York, USA, p. 1–8, 2009. Disponível em: <http://portal.acm.org/citation.cfm?doid=1858477.1858506>. Citado na página 34.

BEURDEN, M. H. van; IJSSELSTEIJN, W. A.; JUOLA, J. F. Effectiveness of stereoscopic displays in medicine: A review. **3D Research**, v. 3, n. 1, p. 1–13, 1 2012. ISSN 20926731. Disponível em: [https://doi.org/10.1007/3DRes.01\(2012\)3](https://doi.org/10.1007/3DRes.01(2012)3). Citado na página 19.

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, 2003. ISSN 01628828. Citado nas páginas 42 e 61.

BT, I.-r.; BT, I.-r.; ITU-R, Q.; ITU-R, Q.; ITU, T.; ITU, T.; ASSEMBLY, R.; ASSEMBLY, R. RECOMMENDATION ITU-R BT.500-11 Methodology for the subjective assessment of the quality of television pictures. **Methodology**, 2002. Citado nas páginas 44 e 45.

Cisco. **Cisco Visual Networking Index: Forecast and Trends, 2017–2022**. 2018. 38 p. Citado na página 19.

DIETER, K. C.; TADIN, D. Understanding attentional modulation of binocular rivalry: A framework based on biased competition. **Frontiers in Human Neuroscience**, 2011. ISSN 16625161. Citado na página 33.

FENG, X.; JIANG, Y.; YANG, X.; DU, M.; LI, X. **Computer vision algorithms and hardware implementations: A survey**. 2019. Citado na página 20.

GOLDBERG, S. B.; MAIMONE, M. W.; MATTHIES, L. Stereo vision and rover navigation software for planetary exploration. In: **IEEE Aerospace Conference Proceedings**. [S.l.: s.n.], 2002. v. 5, p. 2025–2036. ISBN 078037231X. ISSN 1095323X. Citado na página 19.

GOLDSTEIN, B. **Sensation and Perception**. 8. ed. Belmont, CA: Wadsworth Cengage Learning, 2013. 1–490 p. ISBN 9780495760504. Citado nas páginas 19, 23, 24, 25, 26, 27 e 28.

GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing (3rd Edition)**. [S.l.: s.n.], 2007. ISBN 013168728X. Citado nas páginas 62 e 68.

HANNAH, M. J. **Computer Matching of Areas in Stereo Images**. 134 p. Tese (Doutorado) — Stanford, CA, 1974. Disponível em: <http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=AD0786720%5Chttp://dl.acm.org/citation.cfm?id=907372>. Citado na página 51.

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. ISSN 01636804. Citado na página 19.

ISHIHARA, S.; E.J. Tests for Color Blindness. **American Journal of Ophthalmology**, v. 1, n. 5, 1918. ISSN 00029394. Citado na página 94.

ITU. RECOMMENDATION ITU-R BT . 500-11 Methodology for the subjective assessment of the quality of television pictures ANNEX 1 Description of assessment methods Common features. **Methodology**, v. 211, n. BT.500-11, p. 1–48, 2002. Disponível em: <<http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Methodology+for+the+subjective+assessment+of+the+quality+of+television+pictures#2>>. Citado na página 94.

JACQUESSON, T.; SIMON, E.; DAULEAC, C.; MARGUERON, L.; ROBINSON, P.; MERTENS, P. Stereoscopic three-dimensional visualization: interest for neuroanatomy teaching in medical school. **Surgical and Radiologic Anatomy**, v. 42, n. 6, 2020. ISSN 12798517. Citado na página 20.

JOULIN, A.; KANG, S. B. Recovering stereo pairs from anaglyphs. In: **Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition**. IEEE Computer Society Washington, DC, USA: 2013 IEEE Conference on Computer Vision and Pattern Recognition, 2013. p. 289–296. ISBN 978-0-7695-4989-7. ISSN 10636919. Citado nas páginas 13, 49, 50, 53, 54, 71, 77, 87 e 96.

KUMARI, D.; KAUR, K. A Survey on Stereo Matching Techniques for 3D Vision in Image Processing. **International Journal of Engineering and Manufacturing**, v. 6, n. 4, p. 40–49, 2016. ISSN 23053631. Citado nas páginas 21 e 39.

KUNZE, L. SIRA - Reversão Anaglífica Utilizando Técnicas de Análise de Séries Temporais. p. 137, 2019. Citado nas páginas 13, 52, 53, 54, 71, 77, 87 e 94.

KUNZE, L. F.; GOULARTE, R.; SOUSA, E. P. M. d. SIRA - An efficient method for retrieving stereo images from anaglyphs. **Signal Processing: Image Communication**, v. 85, 2020. ISSN 09235965. Citado na página 96.

LEVENE, H. Robust tests for equality of variances. **Contributions to probability and statistics: Essays in ...**, 1960. Citado na página 89.

LEVIN, A.; LISCHINSKI, D.; WEISS, Y. Colorization using optimization. **ACM Transactions on Graphics**, v. 23, n. 3, p. 689, 2004. ISSN 07300301. Disponível em: <<http://portal.acm.org/citation.cfm?doid=1186562.1015780>>. Citado nas páginas 43, 50 e 71.

LIPTON, L. **Foundations of the stereoscopic cinema: a study in depth**. 1. ed. 135 West 50th Street New York, NY 10020: Van Nostrand Reinhold Company, 1982. 325 p. ISBN 0442247249. Citado nas páginas 19 e 29.

_____. **Stereographics Developers' Handbook**. 1. ed. 2171 East Francisco Boulevard San Rafael, CA 94901 United States: StereoGraphics Corporation, 1997. 66 p. ISBN 9781604270488. Citado nas páginas 19 e 29.

LIU, C.; YUEN, J.; MEMBER, S.; TORRALBA, A. SIFT Flow : Dense Correspondence across Scenes and its Applications. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 33, n. 5, p. 1–17, 2008. Citado na página 49.

LUCAS, B. D.; KANADE, T. Iterative Image Registration Technique With an Application To Stereo Vision. v. 2, n. April 1981, p. 674–679, 1981. Citado na página 49.

MURRAY, D.; LITTLE, J. J. Using real-time stereo vision for mobile robot navigation. **Autonomous Robots**, v. 8, n. 2, p. 161–171, 2000. ISSN 09295593. Disponível em: <<https://doi.org/10.1023/A:1008987612352>>. Citado na página 19.

NAM, K. W.; PARK, J.; KIM, I. Y.; KIM, K. G. Application of stereo-imaging technology to medical field. **Healthcare informatics research**, Korean Society of Medical Informatics, v. 18, n. 3, p. 158–163, 9 2012. ISSN 2093-369X. Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/23115737https://www.ncbi.nlm.nih.gov/pmc/PMC3483472/>>. Citado na página 19.

RICHARDSON, I. E. G. **Video codec design : developing image and video compression systems**. New York, NY, USA: John Wiley & Sons, Inc., 2002. 303 p. ISBN 9780471485537. Citado na página 36.

SANFTMANN, H.; WEISKOPF, D. Anaglyph stereo without ghosting. **Computer Graphics Forum**, v. 30, n. 4, p. 1251–1259, 2011. ISSN 14678659. Citado na página 33.

SCHARSTEIN, D.; SZELISKI, R.; ZABIH, R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. **Proceedings - IEEE Workshop on Stereo and Multi-Baseline Vision, SMBV 2001**, n. 1, p. 131–140, 2001. Citado nas páginas 20, 21, 39, 41, 61 e 62.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). **Biometrika**, Oxford University Press (OUP), v. 52, n. 3-4, p. 591–611, 12 1965. ISSN 0006-3444. Disponível em: <<https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/52.3-4.591>>. Citado na página 89.

ULHAQ, A.; BORN, J.; KHAN, A.; GOMES, D. P. S.; CHAKRABORTY, S.; PAUL, M. **COVID-19 control by computer vision approaches: A survey**. 2020. Citado na página 20.

WELSH, T.; ASHIKHMIN, M.; MUELLER, K. Transferring color to greyscale images. **Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '02**, n. July 2002, p. 277–280, 2002. ISSN 0730-0301. Citado na página 43.

WHEATSTONE, C.; WHEATSTONE, C. Contributions to the Physiology of Vision.—Part the First. On some remarkable, and hitherto unobserved, Phenomena of Binocular Vision. **The Scientific Papers of Sir Charles Wheatstone**, p. 225–283, 1838. ISSN 0261-0523. Citado nas páginas 28 e 29.

Williem; RASKAR, R.; PARK, I. K. Depth map estimation and colorization of anaglyph images using local color prior and reverse intensity distribution. **Proceedings of the IEEE International Conference on Computer Vision**, v. 2015 Inter, p. 3460–3468, 2015. ISSN 15505499. Citado nas páginas 13, 21, 50, 51, 53, 54, 71, 77, 87 e 96.

YATZIV, L.; SAPIRO, G. Fast image and video colorization using chrominance blending. **IEEE Transactions on Image Processing**, v. 15, n. 5, p. 1120–1129, 2006. ISSN 10577149. Citado na página 42.

YUGOSHI, J. K. **Reversão anaglífica baseada em busca local rápida**. 2018. 75 p. Citado nas páginas 13, 32, 45, 50, 51, 52, 53 e 54.

ZHANG, L.; VÁZQUEZ, C.; KNORR, S. 3D-TV content creation: Automatic 2D-to-3D video conversion. **IEEE Transactions on Broadcasting**, v. 57, n. 2 PART 2, p. 372–383, 2011. ISSN 00189316. Citado na página 20.

Zhou Wang; SHEIKH, H. R.; BOVIK, A. C. Objective video quality assessment. **Optical Engineering**, v. 45, n. 1, p. 017004, 2003. ISSN 0091-3286. Disponível em: <<http://opticalengineering.spiedigitallibrary.org/article.aspx?doi=10.1117/1.2160515>>. Citado na página 46.

GLOSSÁRIO

Anáglifo: Imagem composta pela sobreposição das visões esquerda e direita de um par estéreo após a eliminação de canais de cores complementares.

Codificação Anaglífica: Processo de geração de um anáglifo a partir de um par estéreo.

Colorização: Técnica que tem por finalidade colorizar imagens monocromáticas.

Correspondência Estéreo Anaglífica: Ou *Correspondência Anaglífica*. Busca por pontos correspondentes nos canais das visões esquerda e direita de um anáglifo.

Correspondência Estéreo: Busca por pontos correspondentes em imagens estereoscópicas.

Disparidade de Retina: Distância entre as projeções de determinado ponto de uma imagem nas retinas dos olhos esquerdo e direito.

DSCQS: *Double Stimulus Continuous Quality Scale* (Escala de qualidade contínua de duplo estímulo). Método que utiliza avaliadores para a medição subjetiva de qualidade em imagens e vídeos que foram submetidas a técnicas de compressão ou processamento.

Estereopsia: Sensação de profundidade proporcionado pelo cérebro ao receber o estímulo das imagens obtidas por cada um dos olhos.

Estereoscopia: Técnica com métodos para a criação e visualização de imagens binoculares que reproduzem a sensação de profundidade utilizando a estereopsia.

MOS: *Mean Opinion Score*. Valor numérico, na faixa de 1 a 5, que representa a média das notas dadas por observadores em um processo de avaliação subjetiva, onde 1 = Péssimo, 2 = Ruim, 3 = Aceitável, 4 = Bom e 5 = Excelente.

MSE: *Mean Squared Error* (Erro Quadrático Médio). Métrica que mede similaridade entre sinais discretos.

Oclusão: Região visível em apenas uma das imagens do par estéreo devido à diferença de ângulo entre as duas visões.

Par Estéreo: Par de imagens que, quando visualizadas individualmente por cada um dos olhos de um observador, permitem a percepção de profundidade a partir da estereopsia.

PSNR: *Peak Signal-to-Noise Ratio* (Relação sinal-ruído de pico) Métrica objetiva usada para medir a qualidade de imagens que foram submetidas a técnicas de compressão ou processamento.

Radiometria: Medição da intensidade de radiação eletromagnética, o que inclui o espectro de luz visível.

Reversão Anaglífica: Processo oposto à codificação anaglífica, onde se busca recuperar o par estéreo de uma imagem estereoscópica a partir do anáglifo.

SAD: *Sum of Absolute Differences* (Soma das Diferenças Absolutas). Métrica que mede similaridade entre sinais discretos.

Streaming: Método de transmissão contínuo de dados sobre redes de computador muito usado para transmissão de conteúdo multimídia. Uma característica importante do Streaming é permitir a reprodução do conteúdo enquanto o restante dos dados ainda estão sendo recebidos.

Vídeo sob Demanda: Ou *Video on Demand*. Tipo de serviço que permite ao telespectador assistir a qualquer momento ao conteúdo que escolher, diferentemente do sistema de televisão tradicional onde o conteúdo é programado.