UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

DEPARTAMENTO ACADÊMICO DE INFORMÁTICA

BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

WILLIAM TAKESHI OMOTO

GERAÇÃO DE MAPA DE PONTOS 3D UTILIZANDO IMAGENS ESTÉREO

TRABALHO DE CONCLUSÃO DE CURSO

PONTA GROSSA

2019

WILLIAM TAKESHI OMOTO

GERAÇÃO DE MAPA DE PONTOS UTILIZANDO IMAGENS ESTÉREO

Trabalho de Conclusão de Curso apresentada como requisito parcial à obtenção do título de Bacharel em Ciência da Computação, do Departamento Acadêmico de Informática, da Universidade Tecnológica Federal do Paraná.

Orientador: Prof. Dr. Erikson Freitas de Morais

PONTA GROSSA

2019

Dedico esse trabalho à minha família e amigos. Em especial à minha mãe Claudia, que me deu apoio nos momentos mais difíceis, e sempre me incentivou para que esse sonho fosse possível.

AGRADECIMENTOS

Em primeiro lugar, gostaria de agradecer à minha família, que proveu todo o suporte e apoio necessário para que eu pudesse chegar até aqui. Em especial a minha mãe Claudia, que sempre foi minha conselheira, amiga, e incentivadora, para que esse momento fosse possível.

Agradeço também aos meus amigos, que sempre estiveram do meu lado, sejam eles em momentos bons ou ruins. Que muitas vezes disponibilizaram seu tempo para me ajudar nos mais diversos momentos durante o curso. Sou muito grato pelo companheirismo e amizade, e sem dúvidas, toda essa jornada não seria a mesma sem vocês por perto.

À Luise, que nos momentos finais e mais estressantes dessa caminhada, sempre esteve ao meu lado, me ouvindo, dando apoio e conselhos. Muito obrigado por ser a melhor pessoa que eu poderia ter do meu lado, e sempre me apoiar nos momentos difíceis, sou eternamente grato por isso.

Aos professores que fizeram parte da minha vida acadêmica, sou grato por todo o conhecimento repassado, e que me tornará um profissional melhor. Em especial ao meu orientador Prof. Dr. Erikson Freitas de Morais, que mesmo nos momentos difíceis desse trabalho, me deu todo o suporte necessário para a sua conclusão.

Por fim, gostaria de agradecer a todos que de alguma forma contribuíram para que esse momento pudesse ser alcançado. Não foram poucos os momentos de dificuldade e incertezas, porém com o apoio de todos, hoje esse sonho se torna realidade.

Ministério da Educação
Universidade Tecnológica Federal do Paraná
Câmpus Ponta Grossa

Diretoria de Graduação e Educação Profissional
Departamento Acadêmico de Informática
Bacharelado em Ciência da Computação

TERMO DE APROVAÇÃO

GERAÇÃO DE MAPA DE PONTOS 3D UTILIZANDO IMAGENS ESTÉREO

por

WILLIAM TAKESHI OMOTO

Este Trabalho de Conclusão de Curso (TCC) foi apresentado em 13 de novembro de 2019 como requisito parcial para a obtenção do título de Bacharel em Ciência da Computação. O candidato foi arguido pela Banca Examinadora composta pelos professores abaixo assinados. Após deliberação, a Banca Examinadora considerou o trabalho aprovado.

_____
Prof. Dr. Erikson Freitas de Morais
Orientador

_____
Profa. Dra. Mauren Louise Sguario Coelho de Andrade
Membro titular

_____
Profa. Dra. Simone Bello Kaminski Aires
Membro titular

_____
Prof. MSc. Geraldo Ranthum
Responsável pelo Trabalho de Conclusão de Curso

_____
Profa Dra. Mauren Louise Sguario Coelho de Andrade
Coordenadora do curso

RESUMO

OMOTO, W. T. **Geração de Mapa de Pontos Utilizando Imagens Estéreo.** 2019. 57p. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) - Universidade Tecnológica Federal do Paraná. Ponta Grossa, 2019.

Com a evolução da tecnologia, componentes eletrônicos estão cada vez mais baratos, permitindo uma disseminação de dispositivos eletrônicos cada vez maior entre o público em geral. Surgiram áreas como a Robótica e a Realidade Aumentada, em que um dos desafios encontrados é a transformação do plano bidimensional captado por uma câmera, em um cenário tridimensional, em que objetos possuem profundidades diferentes na cena. Diversos métodos foram desenvolvidos para solucionar esse desafio, desde métodos que utilizam uma ou mais câmeras, até os que fazem uso de diversos sensores, como o infravermelho para obterem dados mais precisos. Porém a estereoscopia ainda é o método mais conhecido e utilizado. A estereoscopia consiste em utilizar duas ou mais câmeras iguais para captar imagens ao mesmo tempo de uma cena, porém com ângulos diferentes. A diferença de posição dos objetos entre essas imagens é chamada de disparidade. A disparidade pode ser obtida fazendo uso da geometria epipolar, que possui métodos para resolver problemas como a correspondência e reconstrução. Com a disparidade em mãos, é possível realizar o cálculo da profundidade, e assim gerar o mapa de pontos.

Palavras-chave: Mapa de Pontos. Mapa de Profundidade. Estereoscopia. Mapa de Disparidade. Geometria Epipolar.

ABSTRACT

OMOTO,W. T. **Point Cloud Generation Using Stereo Images**. 2019. 57p. Work of Conclusion Course (Graduation in Bachelor in Computer Science) - Federal Technology University - Paraná. Ponta Grossa, 2019.

With the evolution of technology, electronic components are becoming cheaper and cheaper, allowing an increasing dissemination of electronic devices among the general public. Areas such as Robotics and Augmented Reality have arisen, in which one of the challenges is the transformation of the two-dimensional plane captured by a camera in a three-dimensional scene in which objects have different depths in the scene. Several methods have been developed to solve this challenge, from methods that use one or more cameras, to those using a variety of sensors, such as infrared, to obtain more accurate data. But stereoscopy is still the best known and most widely used method. Stereoscopy consists of using two or more equal cameras to capture images at the same time of a scene, but with different angles. The difference in position of objects between these images is called disparity. The disparity can be obtained by using epipolar geometry, which has methods to solve problems such as matching and reconstruction. With the calculated disparity, it is possible to perform the depth calculation, and thus generate the point map.

Keywords: Point Map. Depth Map. Stereoscopy. Disparity Map. Epipolar Geometry

LISTA DE ILUSTRAÇÕES

LISTA DE ABREVIATURAS, SIGLAS E ACRÔNIMOS

2D	2 Dimensões
3D	3 Dimensões
ADAS	Advanced Driver Assistance Systems
HSI	Hue Saturation Intensity
LASER	Light Amplification by Stimulated Emission of Radiation
LIDAR	Light Detection And Ranging
RANSAC	Random Sample Consensus
RV	Realidade Virtual

SUMÁRIO

1 INTRODUÇÃO

Pode-se argumentar que o sentido mais importante do ser humano é a visão. É através dele que são realizadas atividades rotineiras, como a identificação de pessoas e objetos, além da sua extrema importância para a locomoção em um espaço (LENNON, 2015).

A locomoção é o ato de deslocar o corpo de um lugar para outro, seja ele correndo, andando, utilizando uma bicicleta ou automóvel. Para isso a visão desempenha um importante papel ao guiar processos de reação, predição e antecipação. Enquanto os dois primeiros utilizam outros mecanismos do corpo humano para realizarem seu papel, o último funciona utilizando somente a visão. Isso porque a visão fornece informações espaço-temporais de uma forma muito precisa, de forma que é muito útil para por exemplo, desviar de um objeto que se encontra em rota de colisão. (HIGUCHI, 2013)

Ao captar impulsos luminosos através do olho, e transformá-los em impulsos elétricos, o cérebro consegue interpretar e formar uma imagem do espaço ao seu redor. Uma das informações mais importantes obtidos pelo cérebro, é a profundidade. O cérebro humano utiliza várias técnicas para estimar a profundidade de objetos em uma cena, como a estimação por foco, defoco, perspectiva, oclusão e estereopsia. Com a posição dos objetos na cena, e a profundidade estimada, o cérebro consegue recriar a cena tridimensionalmente e utilizá-la para a locomoção (ACHARYYA et al, 2013).

Visão estéreo é a capacidade de se inferir dados sobre uma estrutura tridimensional de acordo com duas perspectivas diferentes (TRUCCO; VERRI, 1998). Seres humanos e grande parte dos animais existentes possuem pelo menos um par de olhos, é através da diferença entre as imagens captadas por eles, que podemos explorar a sensação de profundidade encontrada em ambientes tridimensionais (FORSYTH; PONCE, 2011).

A diferença de posição dos objetos entre as imagens capturadas pelos olhos, é chamada de disparidade. Caso a geometria dos pares estéreo for conhecida, pode-se utilizar a estereoscopia, e duas imagens capturadas ao mesmo tempo para calcular a profundidade de objetos da cena, e gerar um mapa de pontos (FORSYTH; PONCE, 2011).

Estereoscopia é a técnica que recria a percepção de profundidade em uma imagem. A maior parte dos métodos de estereoscopia utilizam duas imagens, que correspondem ao olho direito e esquerdo de uma pessoa. A estereoscopia assim produz a ilusão de profundidade em uma imagem combinando informações das duas imagens em uma (TRUCCO; VERRI, 1998).

Estereopsia é derivado das palavras gregas *stereo*, que significa sólido, e *opsia*, que significa visão. É um termo utilizado para se referir a percepção de profundidade em uma estrutura tridimensional através da informação visual proveniente de dois olhos de indivíduos que possuem visão binocular. Em um ser humano, os olhos mudam seu ângulo de acordo com a distância do objeto observado. Um computador obtém a profundidade através de cálculos utilizando a geometria epipolar (HOWARD; ROGERS, 1995).

A geometria epipolar, também chamada de geometria da visão estéreo, estabelece uma correlação entre pontos de duas imagens de uma cena capturada ao mesmo tempo a partir de dois ângulos diferentes. O caso mais simples, e geralmente mais utilizado por simplificar os cálculos necessários é quando as duas imagens estão no mesmo plano. Para isso, é realizado o processo de retificação das imagens, que é uma transformação linear para tornar as imagens coplanares (XU; ZHANG, 1996).

Este trabalho propõe a geração de um mapa de pontos de uma determinada cena utilizando um par de imagens estéreo. Será utilizada a técnica de estereoscopia, juntamente com a geometria epipolar para calcular a disparidade entre as imagens. De posse dos parâmetros da câmera, e a disparidade calculada, é possível calcular a profundidade dos objetos para a câmera, e assim gerar um mapa de pontos para ser reconstruído tridimensionalmente em um software de modelagem tridimensional.

1.1 PROBLEMA

Existem diversas áreas que podem se beneficiar com o aprimoramento da estereoscopia. Na área médica, ao ensinar anatomia humana a estudantes, professores utilizam cadáveres, livros, ilustrações entre outros meios. Um deles, e mais recente, é a realidade virtual (BENBASSAT; POLAK; JAVITT, 2012).

No cenário da saúde, a Realidade Virtual - RV é utilizada pois permite uma interação e imersão tridimensional muito maior. Dispostos de óculos RV, pode-se simular diversas situações do mundo real de forma muito mais realista (BENBASSAT; POLAK; JAVITT, 2012).

Já na área da robótica é indispensável um sistema para o cálculo de distâncias, para que o robô possa desviar de obstáculos. De todos os meios para se estimar profundidade, a estereoscopia é a forma mais utilizada, e estimar a profundidade de objetos utilizando sistemas de visão computacional, vem se tornando cada vez mais importante (ACHARYYA et al, 2016). Atualmente o robô *Mars Exploration Rover* possui um conjunto de 6 câmeras para auxiliar a navegação do robô em solo marciano. Esses pares estéreos de câmera são utilizados para o cálculo de profundidade de objetos, e elevação do terreno para a locomoção segura do robô (NASA, 2019).

Carros com sistemas ADAS (*Advanced Driver Assistance Systems*), ou Sistemas Avançados de Assistência ao Motorista, mais avançados utilizam sensores LIDAR (*Light Detection and Ranging*) para construir um mapa de pontos para auxílio na direção. Esse mapa de pontos é utilizado para o auxílio da navegação do carro, desde alertas de colisão, até mesmo ações para evitar acidentes.

Apesar de muito úteis, mapas de pontos possuem um custo elevado, sensores LIDAR e equipamentos de medição a laser não são baratos, elevando o custo final do produto. Porém com a melhoria na qualidade das câmeras, e seu preço baixo, estas juntamente com a estereoscopia se tornam grandes candidatas a substituírem equipamentos mais caros (YUQUAN et al, 2017).

1.2 PROPOSTA

Este trabalho propõe a geração de um mapa de pontos de uma determinada cena utilizando um par de imagens estéreo. Será utilizada a técnica de estereoscopia, juntamente com a geometria epipolar para calcular a disparidade entre as imagens. De posse dos parâmetros da câmera, e a disparidade calculada, é possível estimar a profundidade dos objetos para a câmera, e assim gerar um mapa de pontos para ser reconstruído tridimensionalmente em um software de modelagem tridimensional.

O presente trabalho realizou testes e aprimorou o algoritmo desenvolvido para que este apresente o máximo de acurácia. O resultado é um mapa de pontos 3D o qual é possível distinguir a distância e os objetos da cena de forma natural, e que pode ser utilizado em diversas aplicações.

1.3 OBJETIVOS

O objetivo geral deste trabalho é a geração de um mapa de pontos 3D de uma cena por meio de duas imagens estéreo obtidas por câmeras convencionais calibradas adequadamente.

Como objetivos específicos podemos listar:

- Determinar o par estéreo de entrada;
- Aplicar filtros para eliminar ruídos das imagens;
- Gerar o mapa de disparidade correspondente ao par estéreo;
- Calcular a profundidade correspondente ao mapa de disparidade;
- Visualizar os pontos 3D encontrados usando software adequado e mapa de profundidade obtido no passo anterior.

1.4 ORGANIZAÇÃO DO TRABALHO

Nas próximas seções serão abordados tópicos necessários ao entendimento do trabalho. O Capítulo 2 apresenta o Referencial Teórico, que aborda alguns conhecimentos úteis para melhor compreensão do trabalho. O Capítulo 3 demonstra o desenvolvimento realizado para atingir os objetivos descritos na Seção 1.3. O Capítulo 4 mostra os resultados obtidos com o algoritmo desenvolvido. O Capítulo 5 traz a conclusão e os trabalhos futuros.

2 REFERENCIAL TEÓRICO

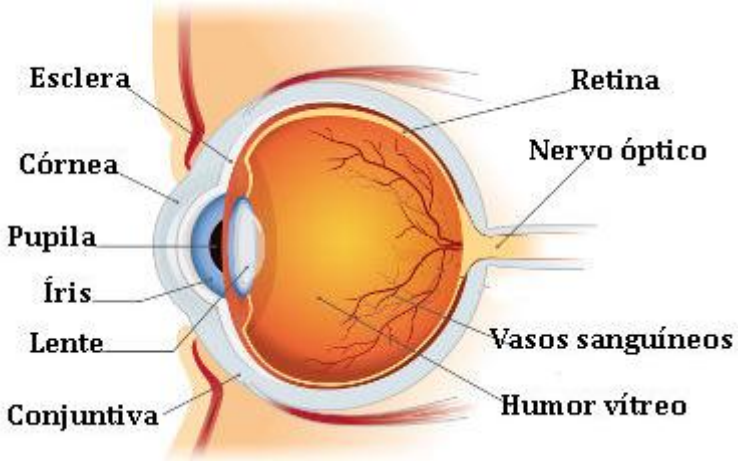
Nessa seção, serão apresentados conceitos essenciais para o entendimento da visão estéreo e seu funcionamento, tais como a estereopsia, geometria epipolar, e os principais problemas que um sistema estéreo deve resolver.

2.1 VISÃO HUMANA

A visão humana é o sentido responsável por capturar impulsos luminosos e transformá-los em imagens, para que assim informações à nossa volta possam ser processadas e decisões tomadas. Todo esse processo inicia-se com o olho humano, esse pequeno órgão consiste em estruturas transparentes para que a luz possa passar sem perturbações (RAMAMURTHY; LAKSHMINARAYANAN, 2015).

A Figura 1 ilustra um exemplo da estrutura de um olho humano, é possível visualizar alguns de seus componentes mais importantes, como a retina, íris, pupila, lente e córnea.

Figura 1 - Exemplo da estrutura de um olho humano



Fonte: (Brasil Escola)

A córnea é a primeira parte do olho humano que a luz encontra, e é responsável por concentrar os impulsos luminosos que o olho humano utilizará para a formação de imagens. A lente é a parte responsável pelo foco do olho humano, ao se contrair e distender, a lente aumenta ou diminui a distância focal do olho (RAMAMURTHY; LAKSHMINARAYANAN, 2015).

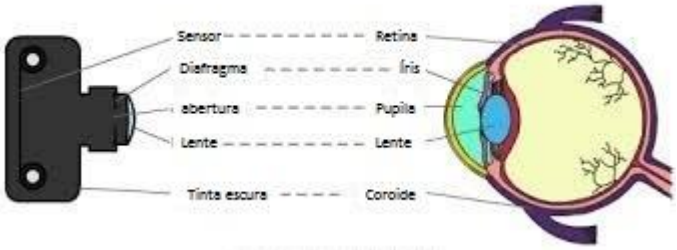
A íris e pupila são responsáveis por controlar a quantidade de luz que entra no olho humano. Quanto menor a quantidade de luz disponível, maior é a abertura

para que o sistema visual humano consiga mais informações durante a formação da imagem. Da mesma forma, quanto maior a quantidade de luz disponível, menor a abertura (RAMAMURTHY; LAKSHMINARAYANAN, 2015).

A retina fica na parte mais interna do olho humano, e consiste em terminações nervosas que passam os impulsos luminosos para o cérebro. É na retina também que se encontra a fóvea, uma região em que se encontram diversos outros componentes que auxiliam na formação da imagem (RAMAMURTHY; LAKSHMINARAYANAN, 2015).

Em uma câmera também encontramos componentes que realizam a mesma função do olho humano ao gerar uma imagem digital. A Figura 2 exemplifica quais componentes de uma câmera correspondem aos componentes do olho humano.

Figura 2 - Comparação entre o olho humano e uma câmera



Fonte: (Adaptado de Quora)

Como pode ser visualizado na Figura 2, é possível traçar um paralelo entre o olho humano e uma câmera, onde a íris e pupila seriam o diafragma da câmera, a lente possui a mesma função entre os dois, e a retina corresponde ao sensor da câmera.

2.2 VISÃO COMPUTACIONAL

Trucco e Verri disseram em 1998 que por mais de 30 anos a visão computacional vem sendo tratada como uma área interdisciplinar. Seus ramos se espalham pela inteligência artificial, robótica, processamento de sinais, neurociência, psicologia, reconhecimento de padrões entre outros (TRUCCO; VERRI, 1998).

A visão computacional inclui processos para a aquisição, processamento e análise de imagens. Seu resultado deve ser um ou mais conjunto de dados que

auxilie na tomada de decisões. Essa extração de dados para a tomada de decisões pode ser vista como uma interpretação de informações simbólicas dentro de uma imagem utilizando modelos construídos com a ajuda da geometria, teoria da aprendizagem, estatística e física (FORSYTH; PONCE, 2011).

A visão computacional é uma área de pesquisa muito vasta, pois em uma imagem existem diversas informações diferentes que podem ser utilizadas para alguma tomada de decisão, como por exemplo: distância, reconhecimento e rastreamento de objetos; reconhecimento de borda, entre outras. Por isso a visão computacional é dividida em subáreas de pesquisa, como por exemplo: detecção de características de imagem, segmentação baseada em característica, análise de movimento, detecção de objetos, visão estéreo, entre outras (TRUCCO; VERRI, 1998).

Com uma gama tão grande de subáreas de pesquisa, e a grande quantidade de informações que é possível obter de uma imagem, a visão computacional também possui várias áreas de aplicação, como por exemplo: inspeção industrial e controle de qualidade, vigilância e segurança, direção autônoma (ADAS), análise médica, aplicação nas indústrias militares e espaciais, entre muitas outras.

2.3 ESTEREOPSIA

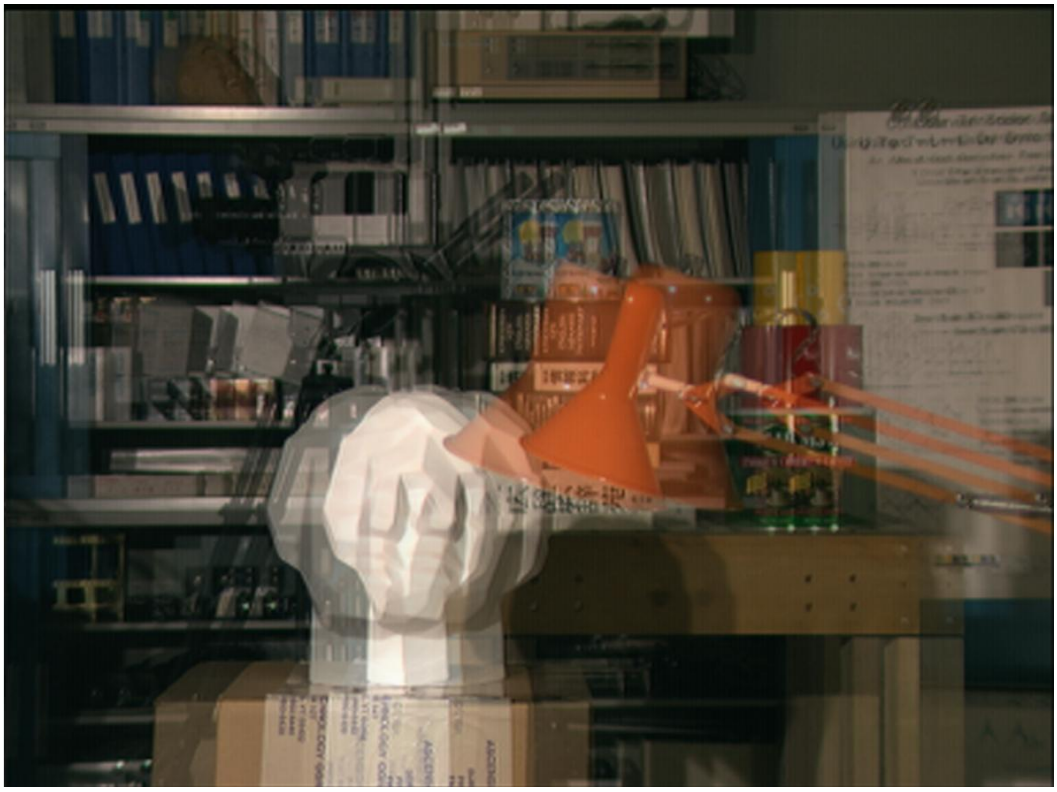
Estereopsia é um conceito geralmente utilizado para se referir a percepção tridimensional que um indivíduo obtém, baseado na informação recebida de duas imagens diferentes. Seres humanos e a grande maioria dos animais possuem dois olhos separados horizontalmente, resultando na projeção de duas imagens diferentes nas retinas dos olhos. Essa diferença entre as imagens é calculada pelo cérebro, e chamada de disparidade retinal (WANG; WU, 2016). Esse processo natural, pode ser simulado computacionalmente pelo método da estereoscopia.

A profundidade também pode ser estimada utilizando uma imagem única através de técnicas como, profundidade por movimento, perspectiva, tamanho, oclusão, foco, entre outras. O cérebro humano define qual método é o mais confiável no momento, e os alterna sem que o indivíduo perceba (LAKSHMANAN; SENTHILNATHAN, 2016). Embora seja possível estimar a profundidade utilizando

uma imagem única, eles não garantem uma percepção tão intensa quanto métodos que utilizam visão binocular (BARRY; SACKS, 2010).

A Figura 3 representa duas imagens sobrepostas de uma mesma cena, obtidas a partir de perspectivas diferentes. É possível observar que as duas imagens possuem diferenças no posicionamento entre os objetos, essa diferença é utilizada pela estereoscopia para estimar a profundidade, e é conhecida como disparidade (ACHARYYA et al, 2016).

Figura 3 – Sobreposição de imagens capturadas por perspectivas diferentes



Fonte: (Adaptado de StackOverflow)

2.4 ESTEREOSCOPIA

Estereoscopia é a capacidade de se inferir dados sobre uma estrutura tridimensional utilizando duas ou mais imagens com perspectivas diferentes (TRUCCO; VERRI, 1998). A visão estéreo, envolve dois processos distintos: a fusão binocular das duas imagens capturadas, e a reconstrução da cena tridimensionalmente (FORSYTH; PONCE, 2011).

A fusão binocular das imagens, consiste em achar a correspondência de milhões de pixels entre as imagens. Existem ainda fatores que dificultam essa tarefa, como padrões repetitivos, e ruídos contidos em imagens, o que podem levar a erros, prejudicando o resultado final do trabalho (FORSYTH; PONCE, 2011).

Com a popularização de áreas que utilizam a informação da profundidade, diversas técnicas de estereoscopia foram desenvolvidas, duas delas estão listadas a seguir:

Anaglifos: Duas imagens de uma cena são produzidas, a partir de duas câmeras espaçadas horizontalmente. Uma das imagens recebe um tom azul, e a outra vermelha. Ao utilizar um par de óculo com lente azul e vermelha, cada olho recebe somente uma das imagens, permitindo assim a percepção de profundidade da cena captada. Essa é uma das técnicas mais comuns, pois exige um custo baixo para implementação (SILVA; MARQUES, 2015).

Polarização da Luz: A luz é polarizada para que passe a vibrar somente em um plano, horizontal ou vertical. Assim duas imagens podem ser polarizadas, para que o usuário ao utilizar um óculo com uma lente polarizada horizontalmente, e outra verticalmente, receba uma imagem diferente em cada olho (SILVA; MARQUES, 2015).

As duas técnicas apresentadas acima utilizam as diferenças entre imagens para dar a sensação de profundidade ao usuário. A estereoscopia utiliza essa diferença para estimar a profundidade de objetos na cena. Ao utilizar imagens obtidas de câmeras calibradas, pode-se utilizar a geometria epipolar para restringir a busca de pontos correspondentes entre as imagens. Conhecendo a geometria do sistema estéreo, e em posse das disparidades, é possível então calcular a profundidade da cena (TRUCCO; VERRI, 1998)

2.5 DISPARIDADE

Disparidade, no contexto da visão computacional, se refere a diferença de posição de objetos entre imagens estéreas, e é inversamente proporcional à distância do objeto para a câmera. A disparidade pode ser calculada para cada pixel na forma de um mapa de disparidade, o qual é utilizado para obter informações de profundidade de uma cena (RAAJAN et al, 2012).

A partir do cálculo da disparidade individual de cada *pixel* da imagem é possível calcular também a distância desse *pixel* até a câmera. Esse cálculo é realizado utilizando a equação dada por:

$$Z = \frac{baseline * f}{d + doffs}$$

Onde *Z* é a distância do *pixel* até a câmera, *baseline* é a distância entre os pontos ópticos do par de câmeras estéreo, *f* é a distância focal das câmeras, *d* é a disparidade calculada do pixel, e *doffs* é a diferença de posição horizontal entre os pontos da imagem esquerda e direita.

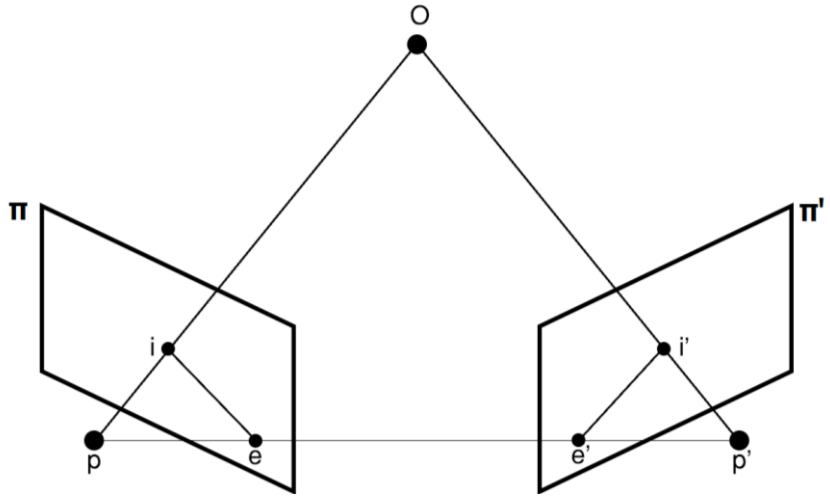
2.6 GEOMETRIA EPIPOLAR

Geometria epipolar é a geometria da visão estéreo. Quando duas câmeras veem uma cena 3D a partir de posições distintas, há um certo número de relações geométricas entre os pontos 3D e suas projeções para as imagens 2D que podem ser utilizadas para calcular a distância (XU; ZHANG, 1996).

A geometria epipolar geralmente é o primeiro passo para muitas tarefas da visão computacional, como reconstrução tridimensional de uma cena, auto-calibração, e navegação robótica. Sua estimação entre duas imagens é um problema fundamental da visão computacional, e possui muitas aplicações (KUKELOVA et al, 2015).

A Figura 4 representa uma simplificação da geometria epipolar. Considere uma cena composta por um objeto *O*, observada por duas câmeras de centro óptico *p* e *p'*. Cada câmera captura duas imagens nos planos *π* e *π'*, e o objeto intercepta os planos nos pontos *i* e *i'*, respectivamente. Os pontos *e, e'* são os epipolos das duas câmeras, ou seja, as imagens dos centros ópticos definidos por *p* e *p'*, respectivamente. Os vetores *ie* e *i'e'*, são definidos como restrição epipolar, e o *baseline* do sistema é o vetor que une os dois pontos ópticos das câmeras, aqui definido como *pp'*.

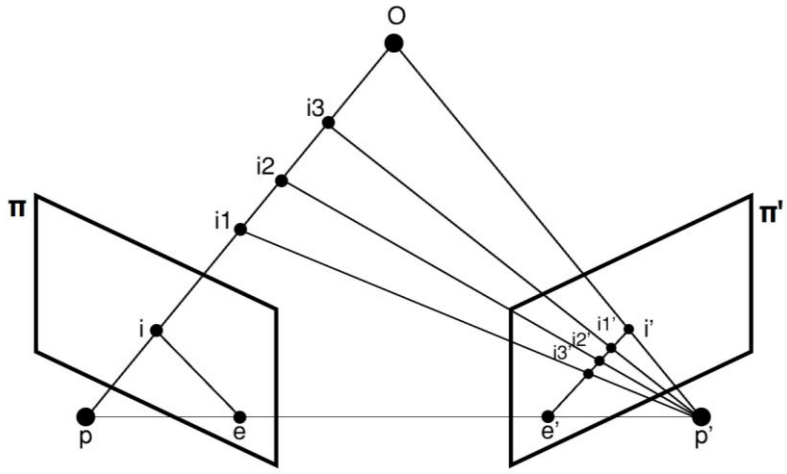
Figura 4 – Geometria Epipolar



Fonte: (Adaptado de Trucco e Verri, 1998)

A profundidade do objeto até a câmera pode ser estimada como algum ponto no vetor *pO*. Esses pontos, representados na Figura 5 como *i1*, *i2* e *i3*, são as profundidades do objeto que podem ser estimadas ao realizar a correspondência de pontos entre as imagens. O erro entre a profundidade real do objeto, e a estimada, depende da solução do problema da correspondência. Porém ao realizar a correspondência dos pontos entre todos os pixels das imagens, a chance da ocorrência de erros é muito grande, devido à grande quantidade de pixels que serão comparados, prejudicando o resultado final.

Figura 5 – Linha Epipolar



Fonte: (Adaptado de Trucco e Verri, 1998)

A Figura 5 apresenta um objeto O, cuja projeção intercepta o plano π no ponto i, pode estar em qualquer ponto do vetor definido por pO. Os pontos i1, i2 e i3, podem ser uma possível localização do objeto O, em um cenário tridimensional. Porém com uma única imagem, é impossível dizer com certeza qual deles é o ponto mais próximo da profundidade real do objeto. Utilizando uma segunda imagem, definida aqui pelo plano π', é possível perceber que as projeções desses pontos interceptam a imagem através da restrição epipolar, definida por i'e'. Assim é seguro assumir que o objeto no plano π, representado pelo ponto i, se encontra na restrição epipolar i'e' do plano π', limitando a busca do objeto entre os pixels das imagens.

2.7 PROBLEMAS DE UM SISTEMA ESTÉREO

Um sistema estéreo deve resolver dois problemas básicos: O primeiro é determinar corretamente quais objetos na imagem da esquerda correspondem aos objetos na imagem da direita. Esse problema é conhecido como Correspondência. O segundo diz respeito à localização da estrutura tridimensional, e o que se pode dizer sobre ela, dado um conjunto de pontos correspondentes e a geometria do sistema estéreo. Esse problema é conhecido como Reconstrução (TRUCCO; VERRI, 1998).

2.7.1 Problema da Correspondência

Podemos dividir o problema da correspondência em duas classes. Baseados em correlação e baseado em características. Apesar de parecerem iguais em um ponto de vista conceitual, eles possuem diferentes formas de implementação (TRUCCO; VERRI, 1998).

A Figura 6 exemplifica o problema da correspondência, onde é possível observar que na Figura 6 diversos pontos estão destacados. Cada ponto na Figura 6-a possui um ponto correspondente na Figura 6-b, que deve ser encontrado para que a disparidade do ponto possa ser calculada.

Figura 6 – Problema da Correspondência

Fonte: (Trucco e Verri, 1998)

Em métodos de correlação, os pontos são combinados com outros definidos em uma janela de tamanho fixo. O ponto escolhido para a correlação, é o que melhor se encaixa dentro do espaço definido pela janela escolhida (TRUCCO; VERRI, 1998).

Métodos baseados em características, restringem a busca por correspondência em um esparso número de características, ao invés de janelas. Muitos dos métodos reduzem o número de possíveis correlações utilizando restrições como as geométricas, para encontrar as possíveis correspondências (TRUCCO; VERRI, 1998).

2.7.2 Problema da Reconstrução

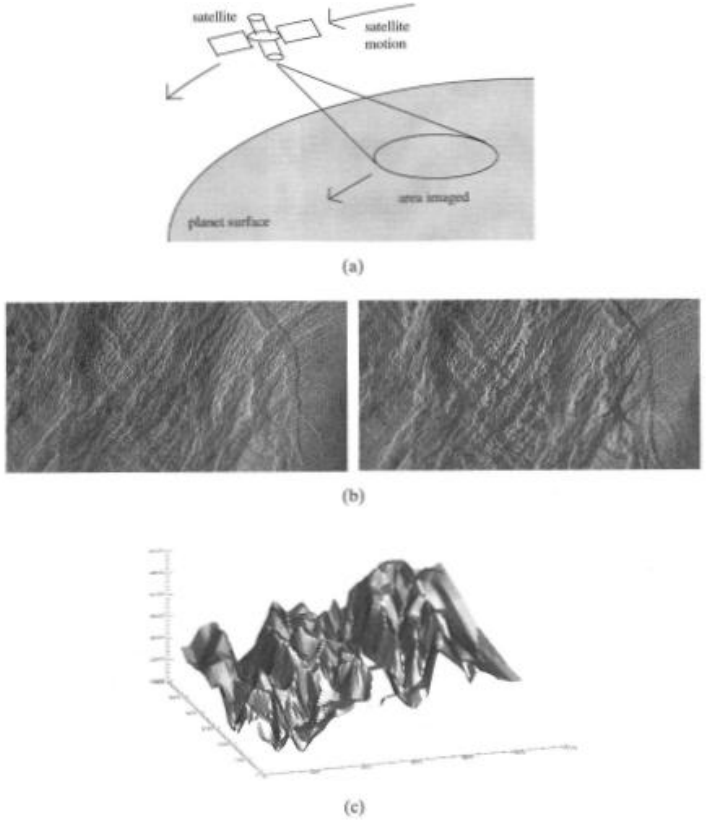
Após resolver o problema da correspondência, a reconstrução tridimensional da cena depende da quantidade de parâmetros intrínsecos e extrínsecos conhecidos. Apesar de ser possível reconstruir uma cena apenas com os parâmetros intrínsecos[1] do sistema, e estimar os parâmetros extrínsecos[2], a confiabilidade do sistema gerado pode ser prejudicada. Caso todos os parâmetros sejam conhecidos, o processo de reconstrução é realizado calculando a triangulação do sistema, utilizando a geometria epipolar (FORSYTH; PONCE, 2011).

---

[1] Parâmetros que fornecem características ópticas internas da câmera (LENNON, 2015)
[2] Parâmetros que fornecem informações da posição e orientação da câmera em relação ao sistema de coordenadas do mundo 3D (LENNON, 2015)

A Figura 7 ilustra o problema da reconstrução, onde um satélite tira diversas fotos da superfície de um planeta (Figura 7-a), e após encontrar a correspondência de cada ponto da imagem (Figura 7-b), precisa reconstruí-la (Figura 7-c).

Figura 7 – Problema da Reconstrução



Fonte: (Trucco e Verri, 1998)

2.8 IMAGENS

Uma imagem digital é composta por uma matriz finita de duas dimensões, onde cada ponto é conhecido como pixel. É adquirida através de um dispositivo digitalizador, como uma câmera digital, na qual uma cena tridimensional é discretizada. Esse processo determina a resolução da imagem, e a quantização dos pixels, ou seja, a quantidade máxima de cores distintas que cada pixel pode assumir (NOGUEIRA, 2016).

2.8.1 Sistema de Cores

Existem diversos modelos para se representar cor em imagens digitais. Entre eles o RGB, talvez o mais conhecido, que utiliza as cores vermelho, verde e azul para representar todas as outras cores (AZAD; HASAN; NASEER, 2017).

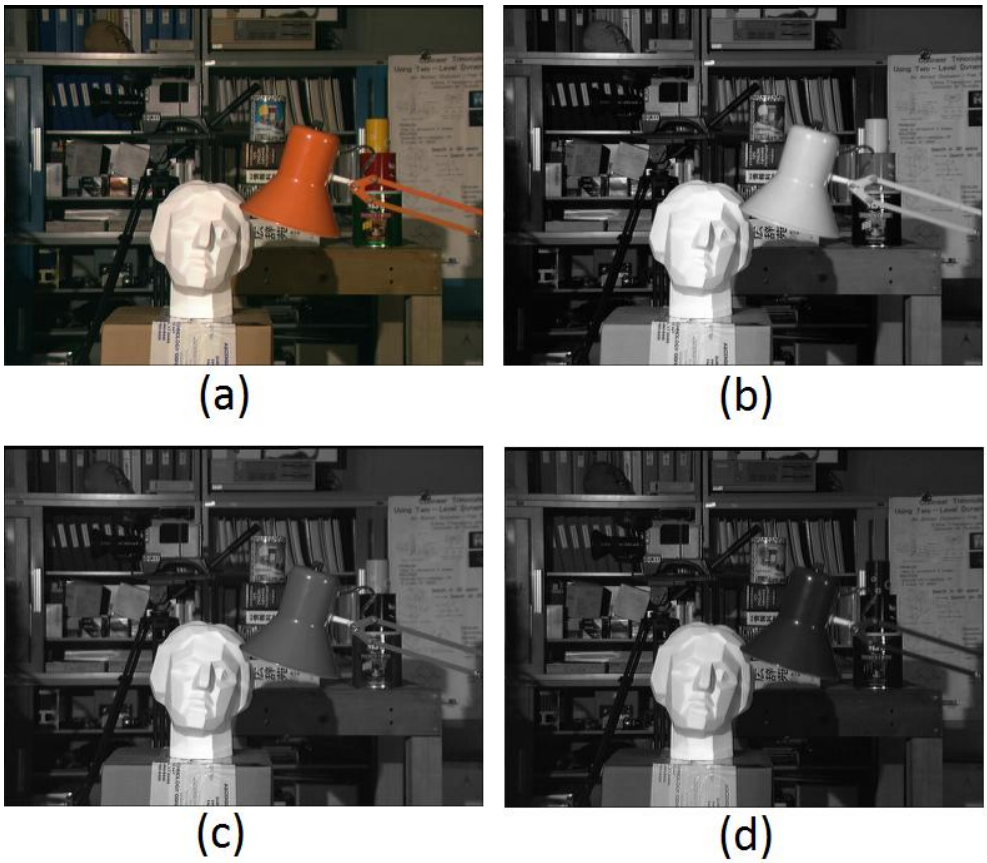
O modelo HSI utiliza as matrizes de matiz, saturação e intensidade para representar imagens. E o modelo mais simples, conhecido como tons de cinza, utiliza somente uma matriz para representar a imagem (YOSHINARI; HOSHI; TAGUCHI, 2014).

2.8.1.1 Espaço de cores RGB

O formato RGB, do ingês *red, green, blue* (vermelho, verde, azul) é um dos diversos formatos de cores na qual uma imagem pode ser representada. Através das diversas intensidades que cada matriz de cor pode assumir, outras cores são percebidas pelo olho humano. É o principal modelo de cor utilizado em dispositivos eletrônicos, como celulares, televisões, monitores entre outros (NOGUEIRA, 2016).

A Figura 8 apresenta a decomposição de uma imagem RGB em seus 3 canais diferentes, vermelho, verde e azul. A Figura 8-a é a imagem original, com os 3 canais simultâneos, a Figura 8-b é o canal vermelho da imagem, a Figura 8-c é o canal verde da imagem e a Figura 8-d é o canal azul da imagem. Quanto mais perceptível a cor de uma das matrizes na imagem original, mais intenso é a matriz correspondente. Pode-se perceber isso na luminária, que visivelmente possui um tom acentuado de vermelho, assim em sua matriz vermelha, ela aparece destacada na imagem (NOGUEIRA, 2016).

Figura 8 – Decomposição de uma imagem RGB em 3 canais



Fonte: (Autoria própria, 2019)

2.8.1.2Tons de Cinza

Uma imagem em tons de cinza, ou *grayscale*, no inglês, é uma imagem na qual somente uma matriz é utilizada para representar a intensidade dos pixels. Portanto a imagem varia entre o branco e o preto, passando por diversos tons de cinza, de onde vem o nome tons de cinza (ALRUBAIE; HAMEED, 2018).

Existem diversas equações que buscam adaptar uma imagem colorida para tons de cinza. A mais simples dela é utilizando a média de todos os canais (ALRUBAIE; HAMEED, 2018).

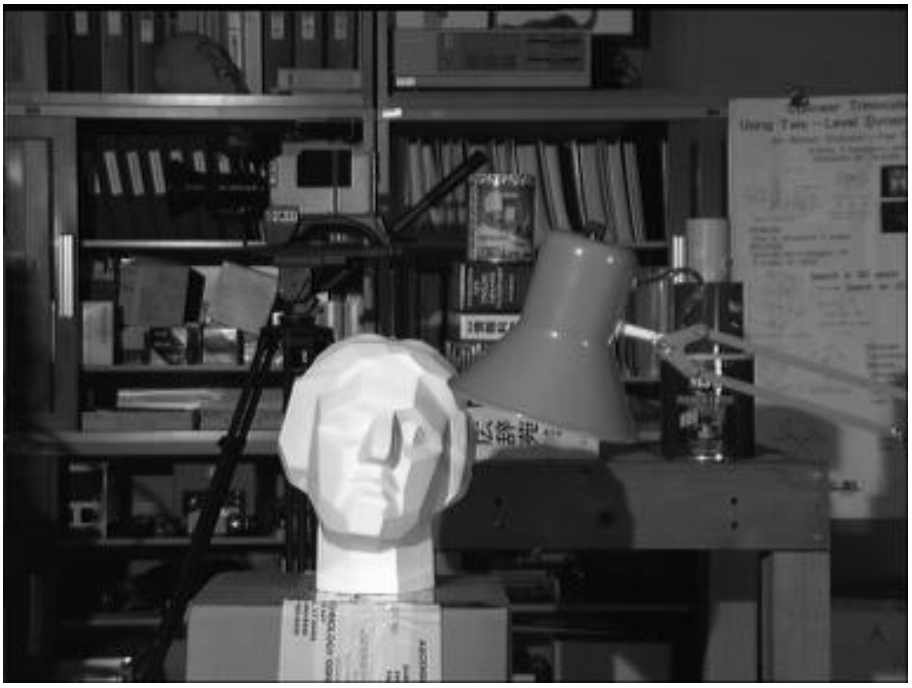
$$CINZA = \frac{VERMELHO + VERDE + AZUL}{3}$$

Outros métodos envolvem multiplicar as matrizes individualmente por pesos, os quais visam manter a fidelidade de cores na imagem em tons de cinza. A mais

utilizada visa manter a luminosidade da imagem e é definida pela fórmula a seguir (ALRUBAIE; HAMEED, 2018).

$$CINZA = 0{,}299 * VERMELHO + 0{,}587 * VERDE + 0{,}114 * AZUL$$

Figura 9 – Imagem em tons de cinza



Fonte: (Autoria própria, 2019)

A Figura 9 representa uma imagem em tons de cinza, na qual os seus 3 canais de cores foram submetidos à um cálculo utilizando a equação acima, para determinar seus novos valores em tons de cinza.

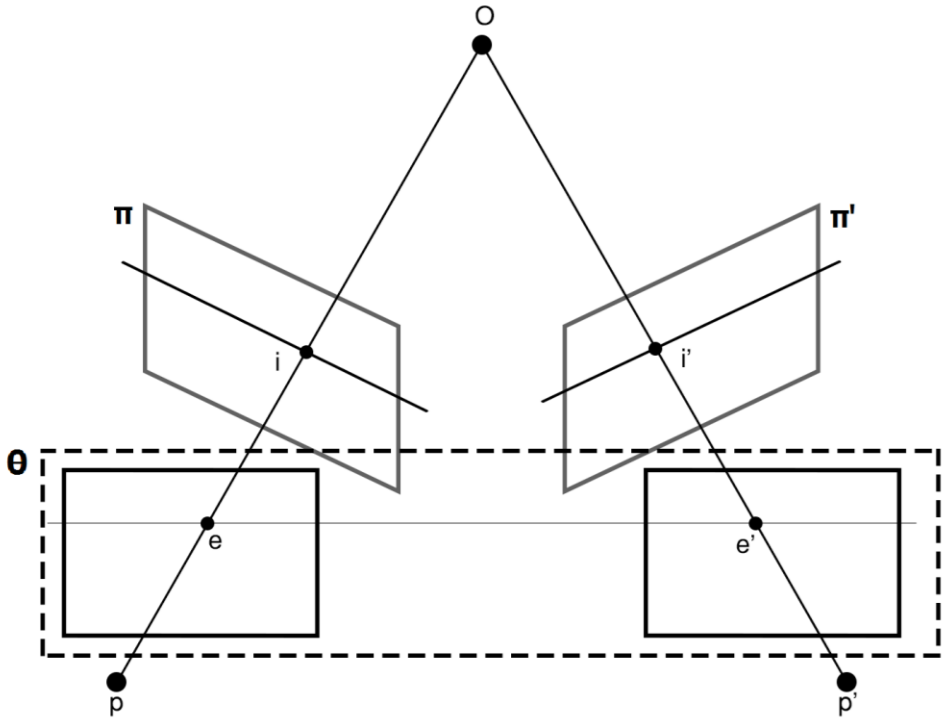
2.8.2 Retificação de Imagens

Retificação de imagens estéreo é o processo que calcula as transformações necessárias em cada imagem, para que suas linhas epipolares tornem-se colineares e paralelas com o eixo horizontal das imagens (SHETE; SARODE; BOSE, 2014). Os cálculos relacionados com a geometria epipolar tornam-se muito mais simples, uma vez que as imagens passam pelo processo de retificação (FORSYTH; PONCE, 2011).

Ao retificar duas imagens, o processo de correspondência torna-se muito mais confiável, e computacionalmente seu desempenho torna-se muito melhor. Como as linhas epipolares das imagens são colineares, o processo de correspondência é limitado a busca por uma linha. Esse limite de busca, proporciona uma chance maior de se encontrar a correspondência correta, e uma velocidade maior do que se a busca ocorresse na imagem inteira (SHETE; SARODE; BOSE, 2014).

A Figura 10 representa duas imagens da mesma cena, pertencentes a dois planos diferentes, chamados π e π'. Aplicando transformações geométricas em cada uma das imagens, estas passam a pertencer ao plano Θ. Assim a linha epipolar, das imagens acaba paralela ao *baseline* do sistema estéreo e ao eixo horizontal do sistema.

Figura 10 – Processo de Retificação de Imagens Estéreo



Fonte: (Adaptado de Trucco e Verri, 1998)

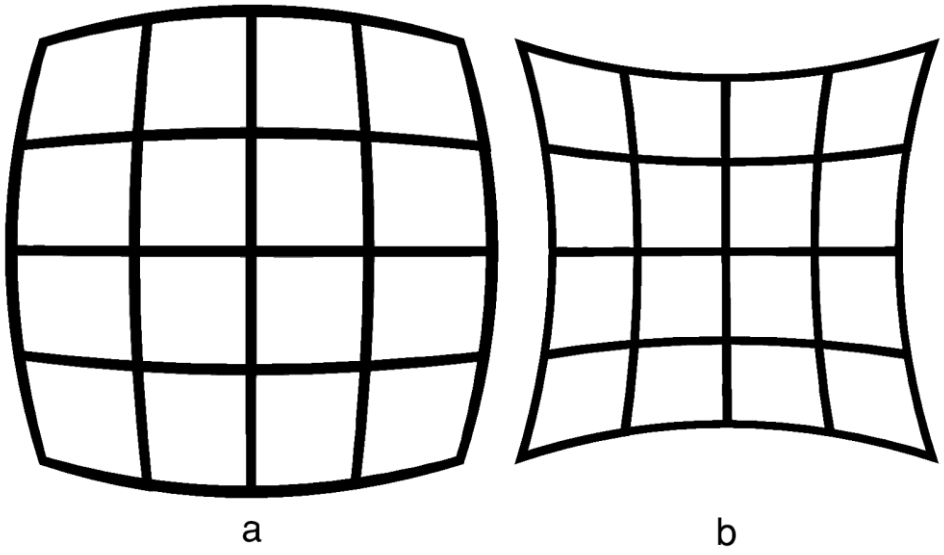
2.8.3 Correção da Distorção Radial

Distorções são a forma mais comum de erros nas lentes, e consiste na localização incorreta de objetos nas imagens, e não da formação incorreta da

imagem (RAHMAN; KROUGLICOF, 2012). Ou seja, os objetos são capturados corretamente, sem que partes da cena sejam perdidas. Porém, a aparência de objetos pode parecer distorcida nas imagens, em comparação com sua aparência no mundo real.

A distorção pelas lentes, pode ser considerada uma perturbação das coordenadas de imagens obtidas através do modelo de câmera *pinhole*[3]. Linhas que são retas na cena real, aparecem distorcidas nas imagens, como representado pela Figura 11 (RAHMAN; KROUGLICOF, 2012).

Figura 11 – Distorções de Lentes



Fonte: (Autoria própria, 2019)

A Figura 11 representa dois tipos de distorções comuns que ocorrem devido a lente das câmeras. Linhas que deveriam ser retas, aparecem com uma distorção. Uma das lentes mais comuns, que introduzem esse efeito de forma proposital, é a chamada lente *fisheye.* Estas lentes permitem uma amplitude de visão maior, devido a grande distorção radial que possuem. Porém objetos acabam distorcidos, e diferentes de sua aparência real.

[3] Pode-se encontrar mais informações sobre o modelo de câmera pinhole em ESTUDO E ANÁLISE DE DIFERENTES MÉTODOS DE CALIBRAÇÃO DE CÂMERAS (LENNON, 2015)

A Figura 11-a apresenta a distorção de tipo barril, geralmente causada por lentes grandes angulares, é caracterizada pelo centro da imagem magnificado em relação as bordas.

A Figura 11-b apresenta a distorção do tipo almofada, geralmente causada por lentes telefoto, é caracterizada pelas bordas da imagem magnificadas em relação ao centro.

2.8.4 Ruído

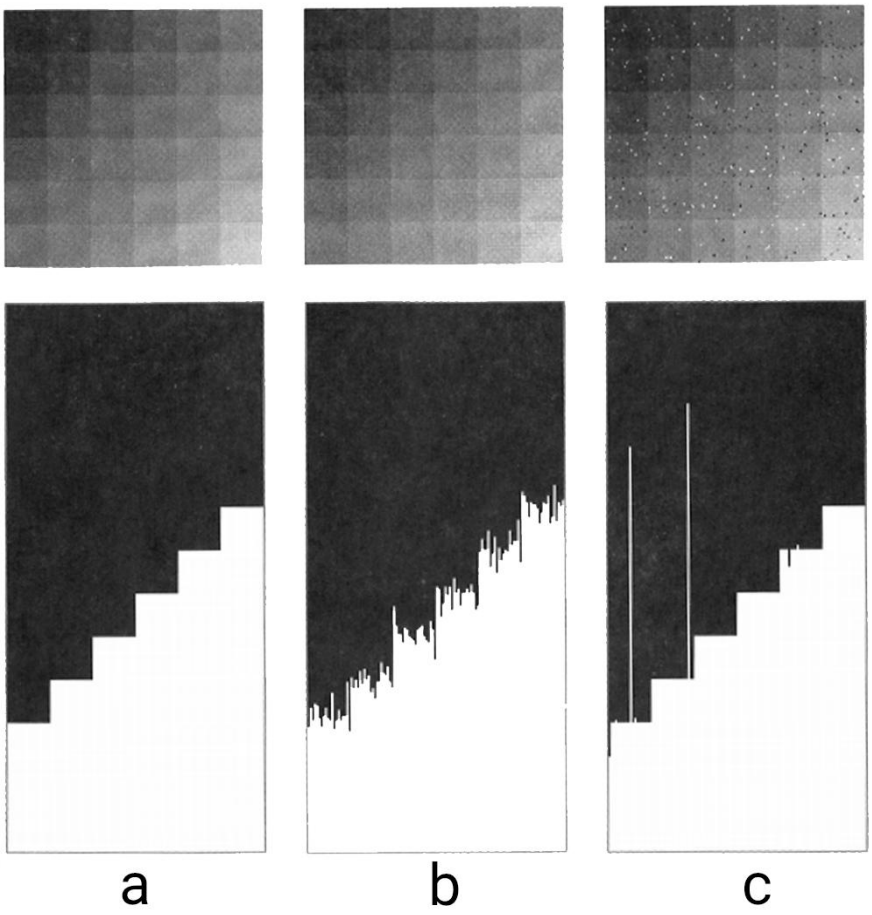
Na visão computacional, ruído se refere a qualquer dado ou resultado que não é alvo de interesse para a computação do trabalho. Por isso em um mesmo conjunto de dados, o que pode ser considerado ruído para um algoritmo específico, pode ser algum dado útil para outro algoritmo (TRUCCO; VERRI, 1998).

Dois tipos de ruídos bem conhecidos são o ruído gaussiano e o ruído impulsivo. O primeiro é um ruído estatístico, o qual prevê que o ruído será distribuído simetricamente. Essa distribuição é esperada de sistemas de bons sistemas de aquisição de imagens, o qual deve garantir baixos níveis de ruído.

O segundo é um ruído proveniente de alterações randômicas de pixels, o qual pode assumir um valor muito maior ou menor que seu valor original. Isso resulta em uma imagem com pontos brancos e pretos espalhados (TRUCCO; VERRI, 1998). A Figura 12 mostra uma imagem sem ruídos, com ruído gaussiano e com ruído impulsivo e seus perfis de cinza logo abaixo.

A Figura 12-a representa a imagem normal, e seu perfil de cinza, logo abaixo, é possível notar que todos os valores possuem distribuição semelhante. A Figura 12-b possui ruído gaussiano, pode-se perceber no perfil de cinza que a distribuição de valores segue um mesmo padrão, porém não é tão suave quanto a Figura 12-a. Já a Figura 12-c apresenta o ruído impulsivo, é possível notar no perfil de cinza, valores que devem ser considerados outliers, pois seus valores são muito discrepantes dos pixels ao seu redor.

Figura 12 – Exemplos de Ruídos



Fonte: (Trucco e Verri, 1998)

2.8.5 Filtros

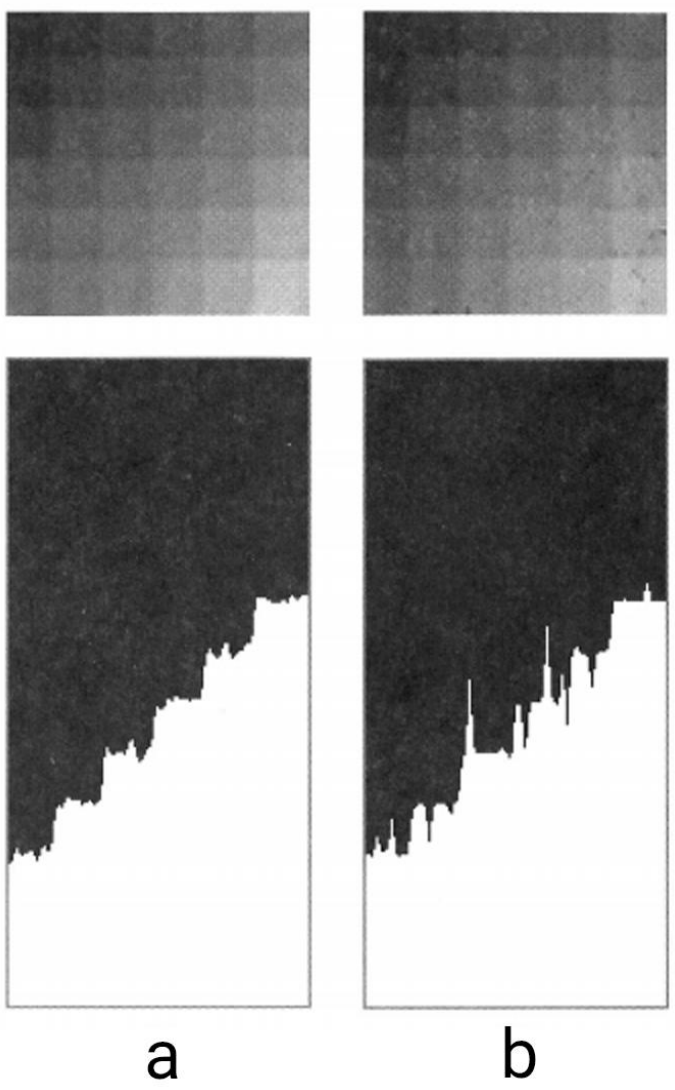
Como discutido na seção anterior, imagens são suscetíveis a diferentes tipos de ruído, os quais podem interferir na aquisição de dados úteis de uma ou mais imagens. Filtros desempenham um importante papel na visão computacional, são eles que realizam a função de atenuar ou eliminar ruídos de imagens (TRUCCO; VERRI, 1998).

Uma técnica comum para atenuar ruídos, é chamado de filtro linear, o qual consiste em utilizar uma matriz constante, chamada de máscara ou kernel, para convoluir a imagem (TRUCCO; VERRI, 1998).

Alguns dos filtros mais comuns que utilizam essa técnica são: filtro gaussiano, filtro da média e filtro da mediana. O filtro gaussiano funciona convoluindo uma função gaussiana na imagem. Essa função possui uma janela de

dimensão variável, e um desvio padrão sigma. Quanto maior o valor do sigma, maior é a influência dos pixels ao redor do pixel de origem, aumentando assim sua capacidade de suavização de ruídos (TRUCCO; VERRI, 1998).

Figura 13 – Exemplo Filtro Gaussiano

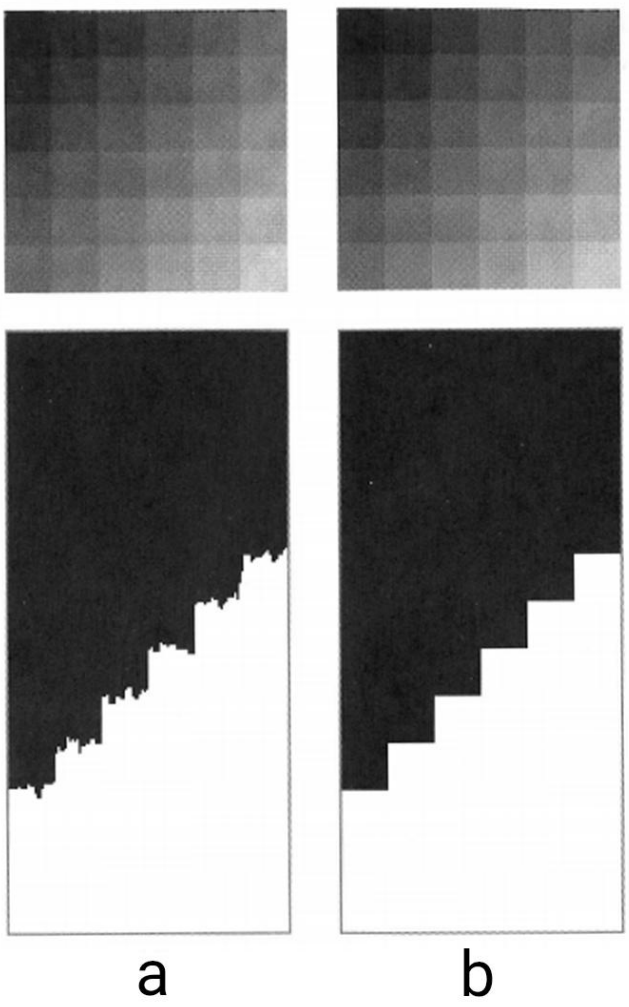


Fonte: (Trucco e Verri, 1998)

A Figura 13-a mostra o resultado da aplicação do filtro gaussiano sobre uma imagem com ruído gaussiano, já a Figura 13-b mostra o resultado da aplicação do filtro gaussiano sobre uma imagem com ruído impulsivo. Como podemos observar pelos perfis de cinza, em ambas as imagens os ruídos foram amenizados, porém não foram removidos. É possível observar que na imagem com ruído impulsivo, o ruído foi espalhado entre seus vizinhos, porém como seus valores normalmente são muito divergentes da sua vizinhança, ainda é possível perceber o ruído na imagem.

O filtro da média funciona convoluindo uma máscara sobre a imagem, porém o pixel central assume como valor a média de valores de todos os pixels abrangidos pela região da máscara. Já o filtro da mediana ordena todos os pixels da região abrangida pela máscara, e atribui sua mediana para o pixel central (TRUCCO; VERRI, 1998).

Figura 14 – Exemplo Filtro Mediana



Fonte: (Trucco e Verri, 1998)

A Figura 14-a apresenta o resultado do filtro da mediana sobre uma imagem com ruído gaussiano, já a Figura 14-b apresenta o resultado do filtro da mediana sobre uma imagem com ruído impulsivo. É possível perceber pelo perfil de cinza que no caso da imagem com ruído gaussiano, obteve-se uma melhora considerável. No caso da imagem com ruído impulsivo, esse foi totalmente eliminado. Isso ocorre pela maneira como o filtro funciona, ao ordenar os pixels da vizinhança, e selecionar sua mediana, o ruído acaba sendo desconsiderado, e assim eliminado no resultado final.

3 DESENVOLVIMENTO

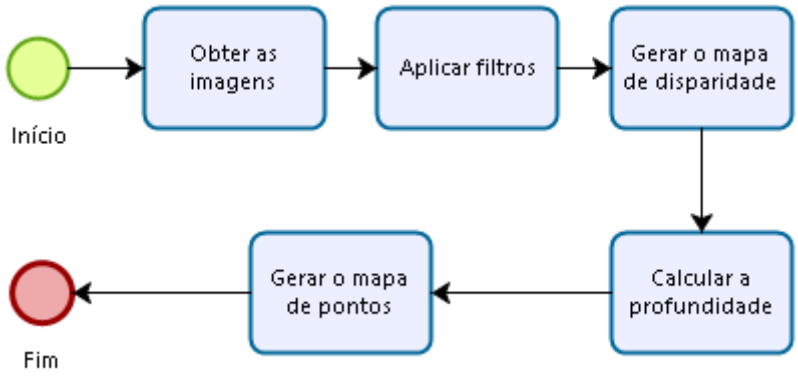
A partir de todo o conhecimento adquirido na Seção 2 deste trabalho, uma metodologia foi proposta para orientar o desenvolvimento do projeto.

A seção 3.1 trata da metodologia utilizada para o desenvolvimento do projeto. Na seção 3.2 é detalhado o processo de escolha dos pares estéreo para o teste do algoritmo. O pré-processamento realizado nas imagens é explicado na seção 3.3. Durante a seção 3.4 é detalhado como o mapa de disparidade é obtido. E por fim, na seção 3.5 é explicado o processo de geração do mapa de pontos 3D.

3.1 METODOLOGIA

A Figura 15 apresenta um fluxograma com os passos a serem seguidos para a realização do trabalho.

Figura 15 – Fluxograma de Atividades



Fonte: (Autoria Própria, 2019)

Toda a codificação será realizada na linguagem de programação C++, juntamente com a biblioteca destinada a visão computacional OpenCV na versão 3.1. Foi escolhida essa versão em detrimento da versão 2.4, pois a versão 3.1 possui um suporte maior a operações estéreo, necessárias para a realização deste trabalho. O ambiente de desenvolvimento integrado utilizado será o Netbeans na versão 8.2.

O primeiro passo é obter as imagens que serão utilizadas. Nesse trabalho as imagens utilizadas serão retiradas de um *dataset* de imagens estéreo disponibilizado pela Universidade de Middlebury (MIDDLEBURY, 2014). Além das imagens da

câmera esquerda e direita, o autor também disponibiliza imagens com uma exposição e iluminação diferente. Também é informado todas as configurações e parâmetros das câmeras utilizadas para a captura das imagens, e um *ground truth* da cena capturada.

O segundo passo é aplicar filtros nas imagens para que ruídos sejam atenuados. Os filtros à serem utilizados são o filtro gaussiano, que possui uma boa performance em ruídos diversos, e o filtro da mediana, que possui uma boa performance em ruídos impulsivos, como observado na Seção 2.8.4.

O terceiro passo é gerar o mapa de disparidade utilizando a geometria epipolar. Para isso a biblioteca de visão computacional OpenCV fornece ferramentas que ajudam a resolver o problema da correspondência e reconstrução, possibilitando assim a geração do mapa de disparidade.

Com os parâmetros das câmeras utilizadas para capturar as imagens calculados, o próximo passo é calcular a profundidade. Como mencionado na Seção 2.6, existem uma série de relações entre as imagens e seus pontos 3D, que podem ser relacionados para calcular a profundidade de objetos.

Após a profundidade da cena ser calculada, o último passo é gerar o mapa de pontos 3D. O mapa é gerado no formato PLY que é aceito pela maior parte dos *softwares* de manipulação de objetos tridimensionais. Neste trabalho será utilizado o *software* MeshLab, pois este é uma ferramenta *open source*, não necessitando de uma licença paga para uso (Meshlab, 2019).

3.2 ESCOLHA DO PAR ESTÉREO

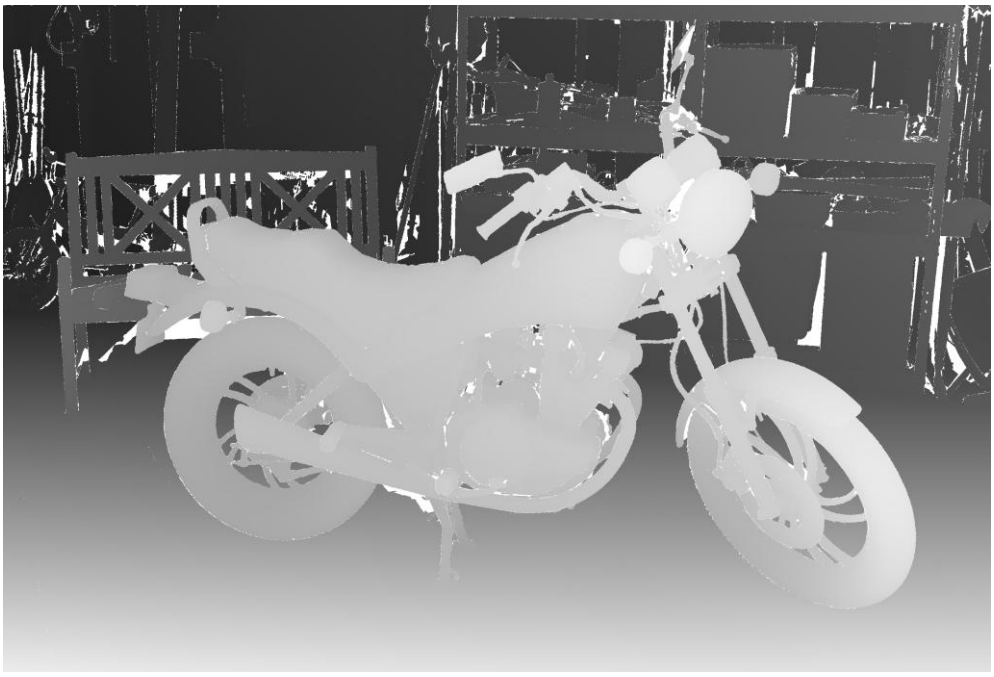
As imagens utilizadas nesse trabalho, provém de um *dataset* de imagens estéreo disponibilizado pela Universidade de Middlebury. O autor das imagens fornece mais de 30 pares de imagens estéreo obtidas em câmeras devidamente calibradas.

Figura 16 – Exemplo de par estéreo do dataset

Fonte: (Middlebury, 2014)

A Figura 16 mostra um exemplo de par estéreo disponibilizado pela Universidade de Middlebury. A Figura 16-a representa a imagem capturada com a câmera esquerda, e a Figura 16-b representa a imagem capturada com a câmera direita. Além das imagens estéreo, o autor disponibiliza também o *ground truth* das disparidades da cena, e as informações da câmera utilizada para a captura das imagens.

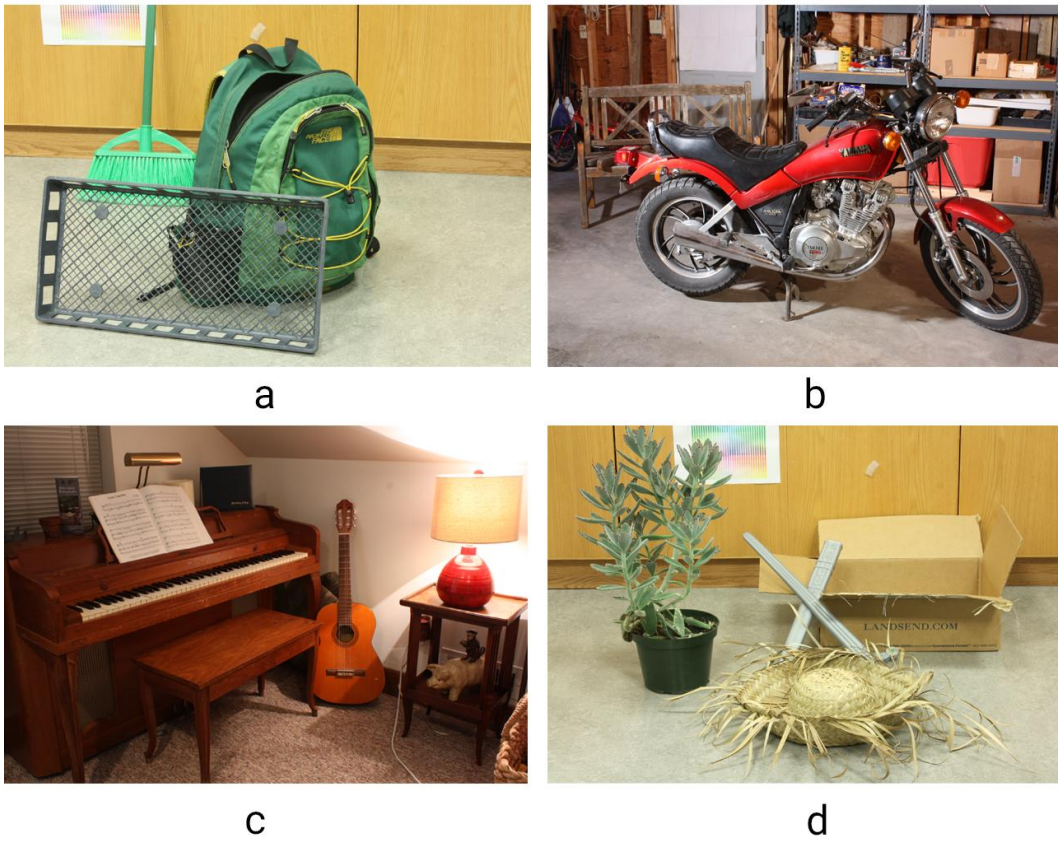
Figura 17 – Exemplo de ground truth de uma imagem do dataset



Fonte: (Middlebury, 2014)

A Figura 17 mostra um exemplo do *ground truth* de uma imagem do *dataset*. A partir do *dataset*, foram selecionadas 4 imagens de forma aleatória para compor o cenário de testes do algoritmo, que são representadas na Figura 18.

Figura 18 – Imagens para teste do algoritmo



Fonte: (Middlebury, 2014)

3.3 PRÉ-PROCESSAMENTO

Com as imagens de teste definidas, será realizado o pré-processamento, que será dividido em duas etapas: a aplicação de filtros, e a conversão de cores.

Os filtros definidos para o pré-processamento das imagens são o filtro gaussiano e o filtro da mediana. O primeiro foi escolhido pela grande capacidade de atenuação de diversos tipos de ruído, pois atua de forma estatística na imagem. O segundo foi adotado pela grande capacidade de atenuar ruídos do tipo impulsivo, que podem ocorrer durante a captura das imagens.

Figura 19 – Trecho de código onde são aplicados os filtros do pré-processamento

```
// Declaração de matrizes
Mat gaussianLeft, gaussianRight, medianLeft, medianRight;

// Leitura das imagens
imgL = imread("MotorcycleLeft.png");
imgR = imread("MotorcycleRight.png");

// Uso do filtro gaussiano
GaussianBlur(imgL, gaussianLeft, Size(sizeKernel, sizeKernel), 0, 0, 0);
GaussianBlur(imgR, gaussianRight, Size(sizeKernel, sizeKernel), 0, 0, 0);

// Uso do filtro da mediana
medianBlur(gaussianLeft, medianLeft, sizeKernel);
medianBlur(gaussianRight, medianRight, sizeKernel);
```

Fonte: (Autoria própria, 2019)

A Figura 19 ilustra o trecho de código onde ocorre a leitura das imagens, e a utilização dos filtros da mediana e gaussiano nas imagens. A Figura 20 exemplifica uma das imagens do conjunto de teste após a utilização dos filtros gaussiano e da mediana para a atenuação dos ruídos.

Figura 20 – Figura de teste após utilização de filtro gaussiano e da mediana

Fonte: (Autoria própria, 2019)

A conversão de cores será utilizada para diminuir a quantidade de cálculos necessários para se obter um mapa de profundidade. Ao invés de calcular a disparidade em 3 matrizes de cores RGB, será realizada a conversão para tons de cinza, reduzindo os cálculos necessários somente para uma matriz.

Figura 21 – Trecho de código onde ocorre a conversão de cores da imagem

```
// Conversão de cores de RGB para tons de cinza
cv::cvtColor(gaussianLeft, imgGrayL, CV_RGB2GRAY);
cv::cvtColor(gaussianRight, imgGrayR, CV_RGB2GRAY);
```

Fonte: (Autoria própria, 2019)

A Figura 21 mostra o trecho de código onde é realizada a conversão de cores da imagem. O resultado é apresentado pela Figura 22, que mostra o resultado da conversão da imagem de RGB para tons de cinza.

Figura 22 – Imagem de teste convertida de RGB para tons de cinza

Fonte: (Autoria própria, 2019)

3.4 OBTENDO O MAPA DE DISPARIDADE

Para a obtenção do mapa de disparidade, é preciso resolver os dois problemas básicos, o problema da correspondência e o problema da reconstrução. Para isso o OpenCV disponibiliza algumas funções que ajudam nesses problemas.

São definidos 3 passos principais para calcular o mapa de disparidade.

- O primeiro é utilizado para estabelecer a correspondência entre as imagens analisadas, e define o alcance máximo da busca pela disparidade. O algoritmo buscará a melhor disparidade entre um intervalo, que deverá ser definido através de testes.
- O segundo passo é computar a disparidade do par estéreo analisado, como resultado é gerado uma matriz de 16 bits.
- O terceiro passo é normalizar a matriz resultante, para exibição.

Figura 23 – Trecho de código onde é calculado o mapa de disparidade do par estéreo

```
// Função utilizada para definir a área de busca de pixels correspondentes
Ptr<StereoBM> stereo = cv::StereoBM::create(256, 11);
// Função utilizada para computar as disparidades entre as imagens
stereo->compute(imgGrayL, imgGrayR, disp);
// Função utilizada para normalizar o resultado
cv::normalize(disp, imgOut, 0, 255, CV_MINMAX, CV_8U);
```

Fonte: (Autoria própria, 2019)

A Figura 23 exemplifica o trecho de código onde é calculado o mapa de disparidade da função. A primeira função define a área de busca entre as duas imagens em que será realizada a busca de correspondência. Ela recebe dois parâmetros, o primeiro é a quantidade de disparidade existente entre as imagens, nesse caso é 256 pois em uma imagem de tons de cinza 8bits existem 256 valores possíveis. O segundo é a janela em que a disparidade deve ser buscada entre as duas imagens. A segunda função utiliza os parâmetros definidos na primeira função para realizar a busca pelas disparidades da imagem. Por último, a terceira função normaliza os valores da disparidade para visualização.

3.5 GERAR O MAPA DE PONTOS

Com o mapa de disparidade calculado, através da relação entre disparidade, distância focal e *baseline*, é possível calcular a distância do pixel para a câmera da imagem inteira, gerando assim um mapa de profundidade da imagem.

Após calcular a profundidade da cena, é gerado um arquivo com a extensão PLY, que é um formato de arquivo reconhecido por diversos softwares de

modelagem 3D. Esse arquivo contém um cabeçalho descrevendo as informações contidas no corpo do arquivo, e que serão utilizadas para reconstruir a cena tridimensionalmente. Para esse trabalho, foi incluído no cabeçalho, além das informações de localização tridimensional, as informações de cores de cada pixel na imagem, para assim facilitar a visualização do resultado final.

Figura 24 – Exemplo da estrutura de um arquivo .PLY

```
1   ply
2   format ascii 1.0
3   element vertex 5928000
4   property float32 x
5   property float32 y
6   property float32 z
7   property uchar red
8   property uchar green
9   property uchar blue
10  element face 0
11  property list uchar int vertex_indices
12  end_header
13  0 0 5884.76 124 77 52
14  0 1 5884.76 122 75 49
15  0 2 5884.76 123 75 48
16  0 3 5884.76 124 76 48
17  0 4 5884.76 124 78 47
18  0 5 5884.76 123 78 47
19  0 6 5884.76 124 76 46
20  0 7 5884.76 122 75 47
```

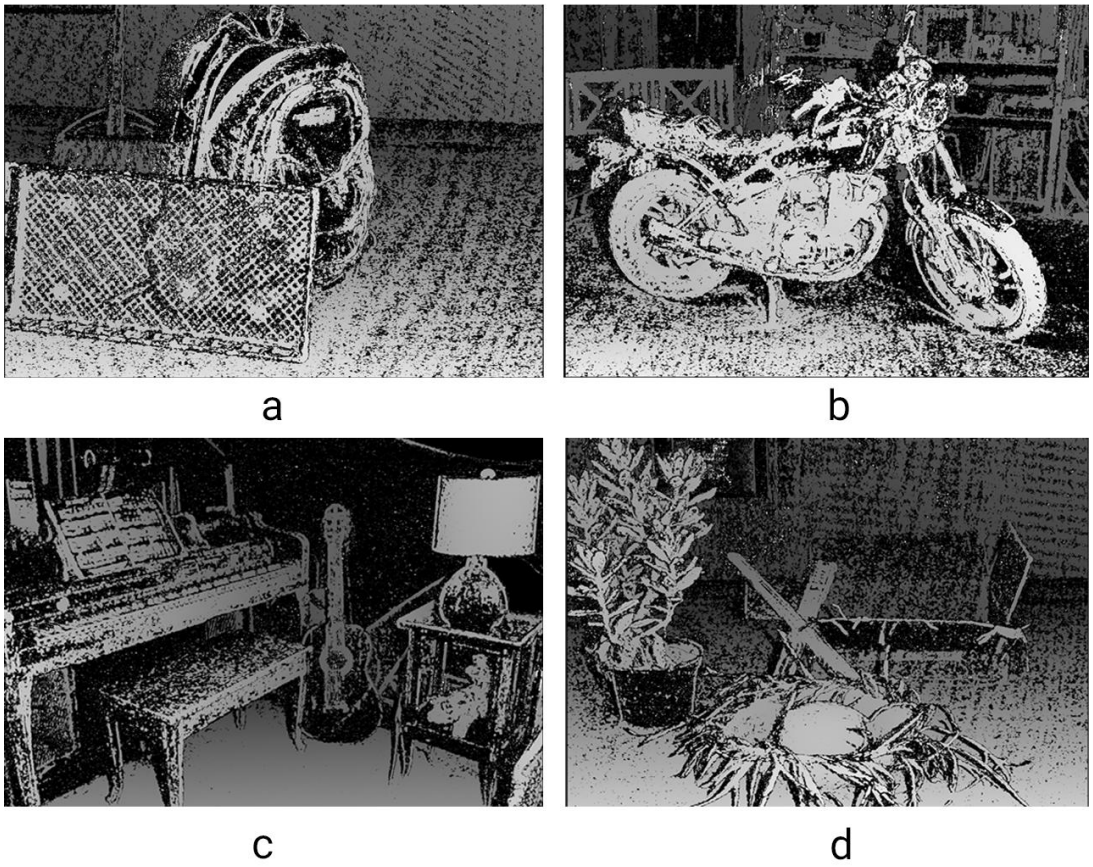
Fonte: (Autoria própria, 2019)

A Figura 24 mostra um exemplo da estrutura de um arquivo com a extensão PLY. Entre as linhas 1 e 12, é definido o cabeçalho do arquivo, que contém o formato da informação contida no arquivo, e informações para a recriação tridimensional da cena, como por exemplo, a quantidade de faces, que nesse caso é zero, pois, um mapa de pontos não possui faces. A partir da linha 13 é definido o corpo do arquivo, onde o primeiro valor representa a coordenada X, o segundo valor representa a coordenada Y, o terceiro valor representa a coordenada Z, o quarto valor representa o valor da matriz *Red*, o quinto valor representa o valor da matriz *Green* e o sexto e último valor representa o valor da matriz *Blue*.

4 RESULTADOS

As imagens selecionadas para o teste, representadas pela Figura 18 foram submetidas ao algoritmo de disparidade, e normalizadas para exibição. A Figura 25 mostra os mapas de disparidade obtidos com o algoritmo desenvolvido.

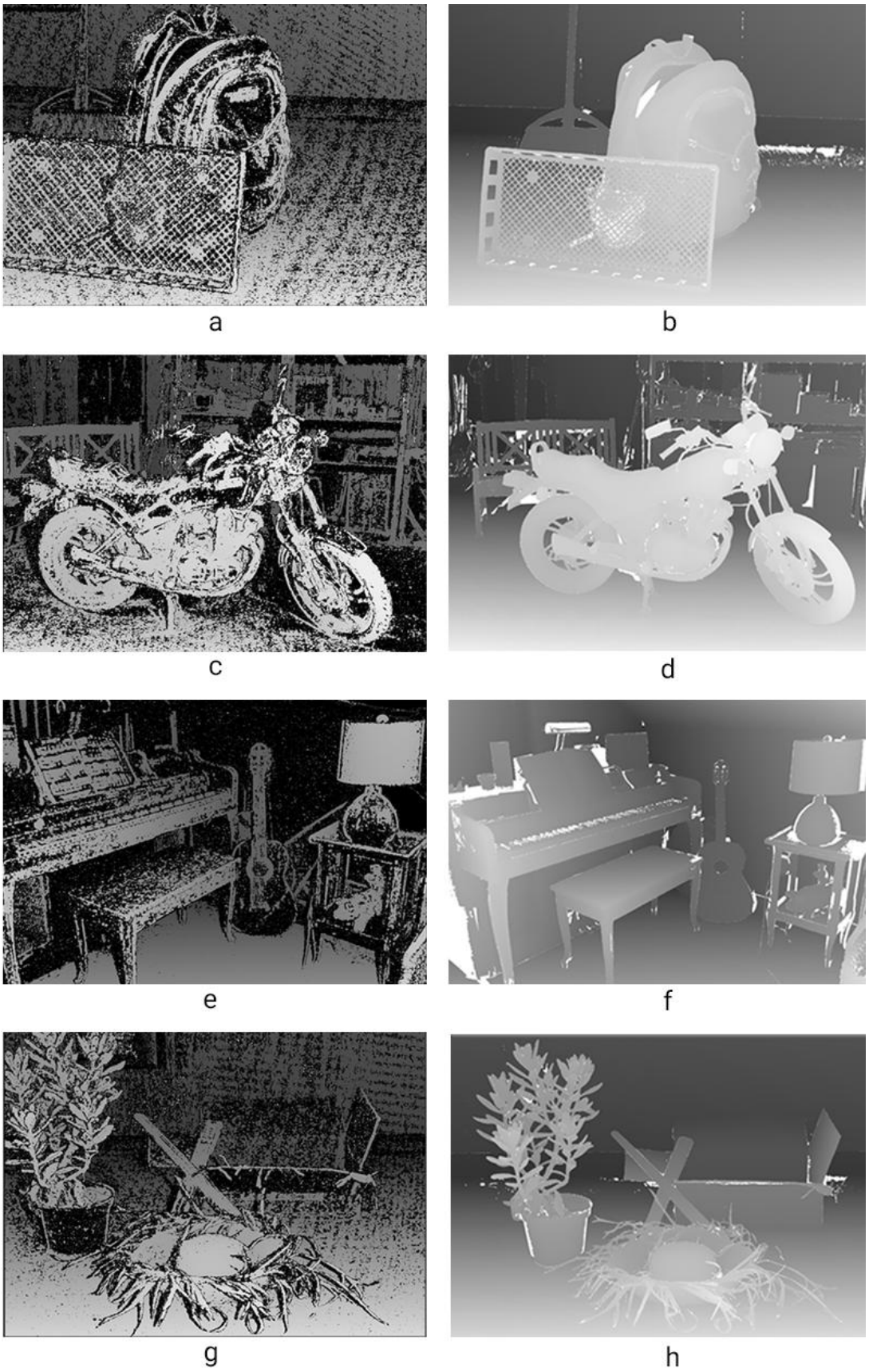
Figura 25 – Mapas de Disparidade das Imagens



Fonte (Autoria Própria, 2019)

Nos mapas de disparidade gerados, representado pela Figura 25, quanto mais claro os pixels, mais perto da câmera ele está. Além disso é possível identificar os diferentes objetos da cena. Também é possível notar que em regiões muito uniformes, como a parede da Figura 25-c o algoritmo possui dificuldade para encontrar os pixels correspondentes, resultando em um fundo preto.

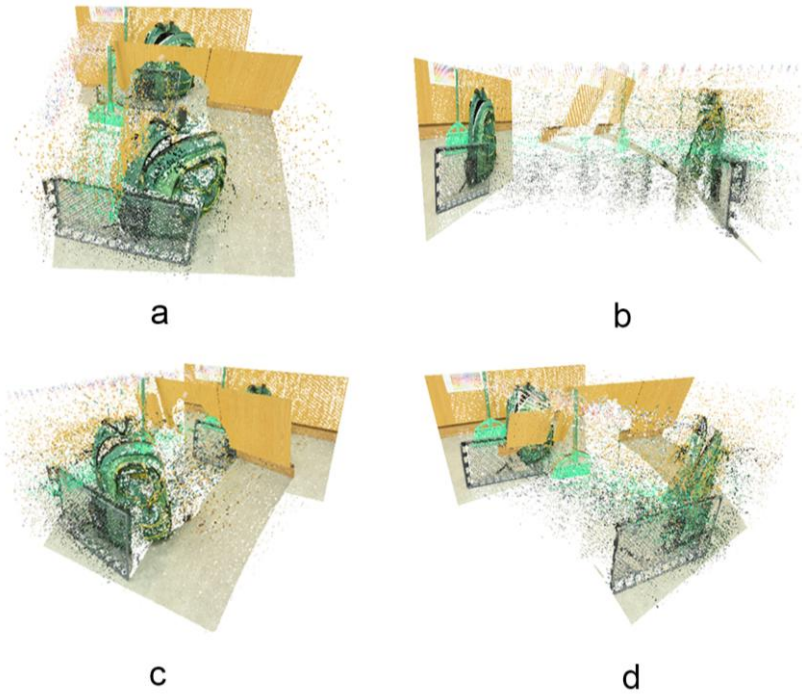
Figura 26 – Comparação Entre Mapas de Disparidade obtidos pelo algoritmo e o ground truth do autor das imagens



Fonte (Autoria Própria, 2019)

A Figura 26(a, c, e, g) ilustra a disparidade obtida com o algoritmo desenvolvido. A Figura 26(b, d, f, h) exemplifica o *ground truth* da disparidade disponibilizado pelo autor das imagens. Figura 25Apesar de não obter a disparidade de todos os pixels, visualmente é possível perceber todos os objetos da cena. Com a disparidade determinada, foi calculada a profundidade dos objetos, e esta utilizada para gerar o mapa de pontos, representado pela Figura 27.

Figura 27 – Primeiro mapa de ponto gerado pelo algoritmo desenvolvido

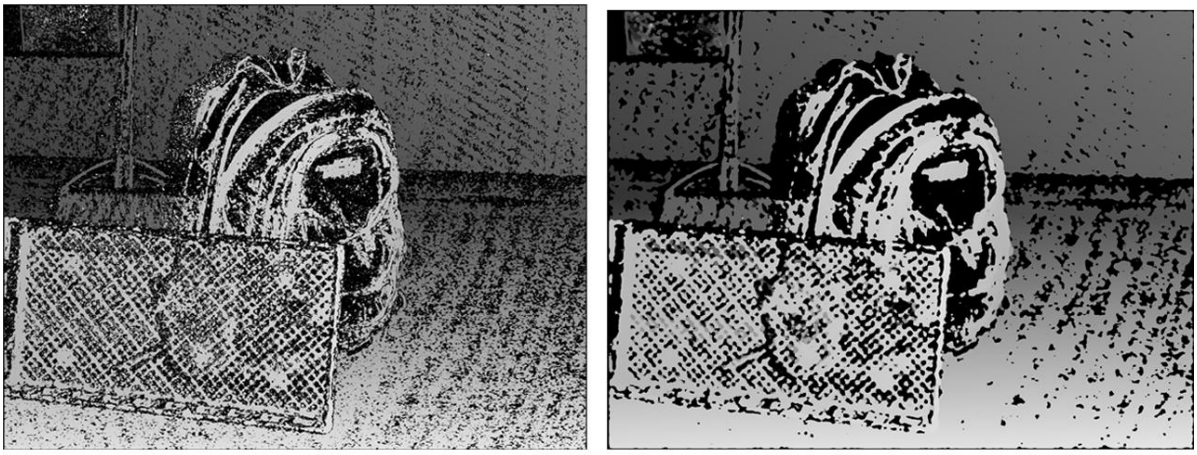


Fonte (Autoria Própria, 2019)

A Figura 27-a representa a visão superior, a Figura 27-b representa a visão lateral esquerda, a Figura 27-c representa a visão diagonal esquerda, e a Figura 27-d representa a visão diagonal direita do mapa de pontos gerado pelo algoritmo desenvolvido. Pode-se notar que existem muitos pontos que podem ser considerados como ruído, pois estes não contribuem para a reconstrução tridimensional da imagem. Como esses ruídos possuem características semelhantes à ruídos do tipo impulsivo, foi utilizado o filtro da mediana no mapa de disparidade gerado pelo algoritmo para eliminá-los. Como observado na Seção 2.8.5, o filtro da mediana funciona bem para esse tipo de ruído, pois ao ordenar o valor dos pixels da vizinhança, os valores muito altos ou muito baixos acabam sendo descartados para

o valor final. A Figura 28 mostra um comparativo entre a imagem antes e depois da utilização do filtro.

Figura 28 – Comparativo entre o mapa de disparidade antes e depois da utilização do filtro da mediana

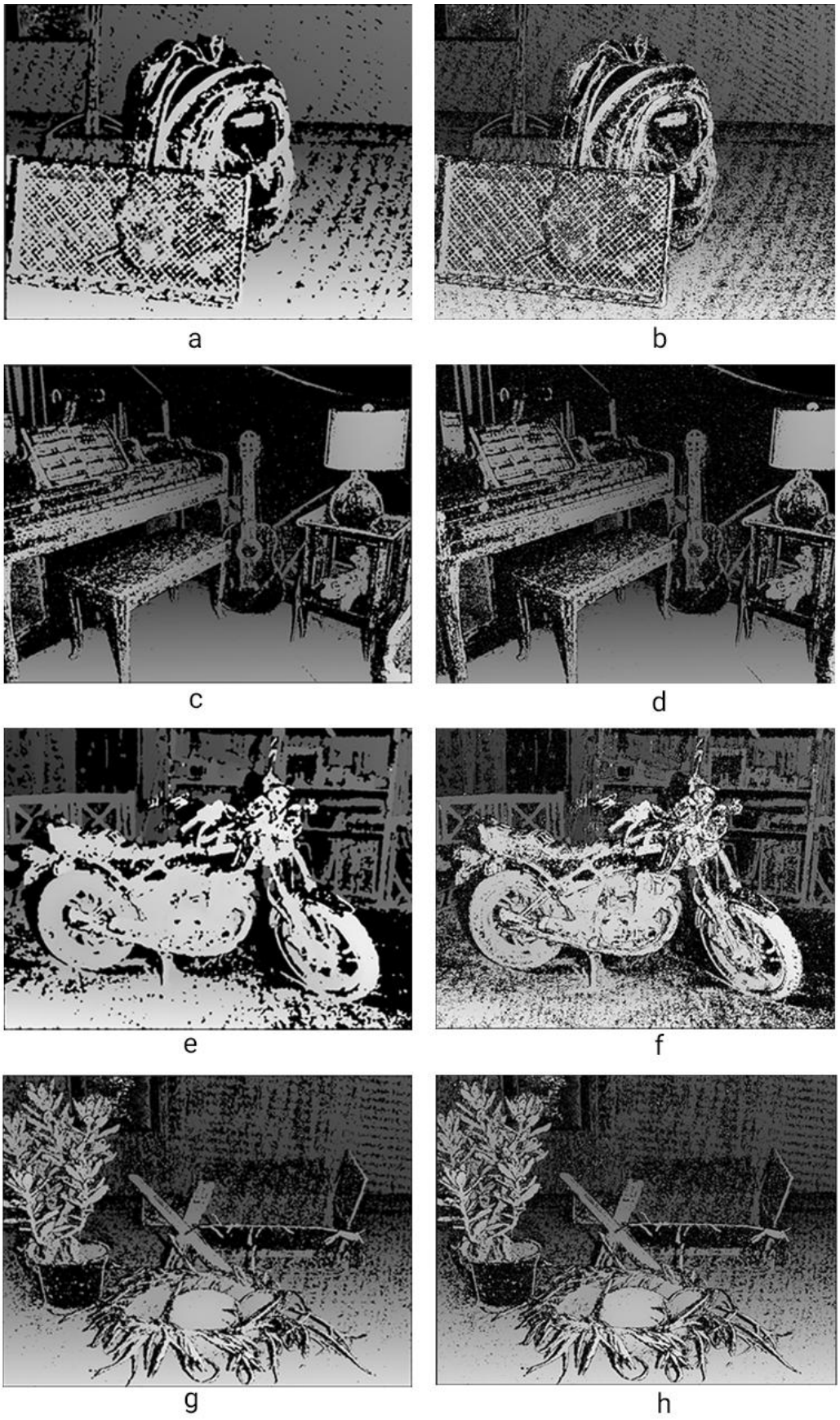


Fonte (Autoria Própria, 2019)

Como é possível notar, o mapa de disparidade ficou mais uniforme, e apresenta uma quantidade menor de ruído na imagem resultante. Com o resultado positivo em uma das imagens, as demais imagens do conjunto de teste também passaram pelo processo do filtro da mediana.

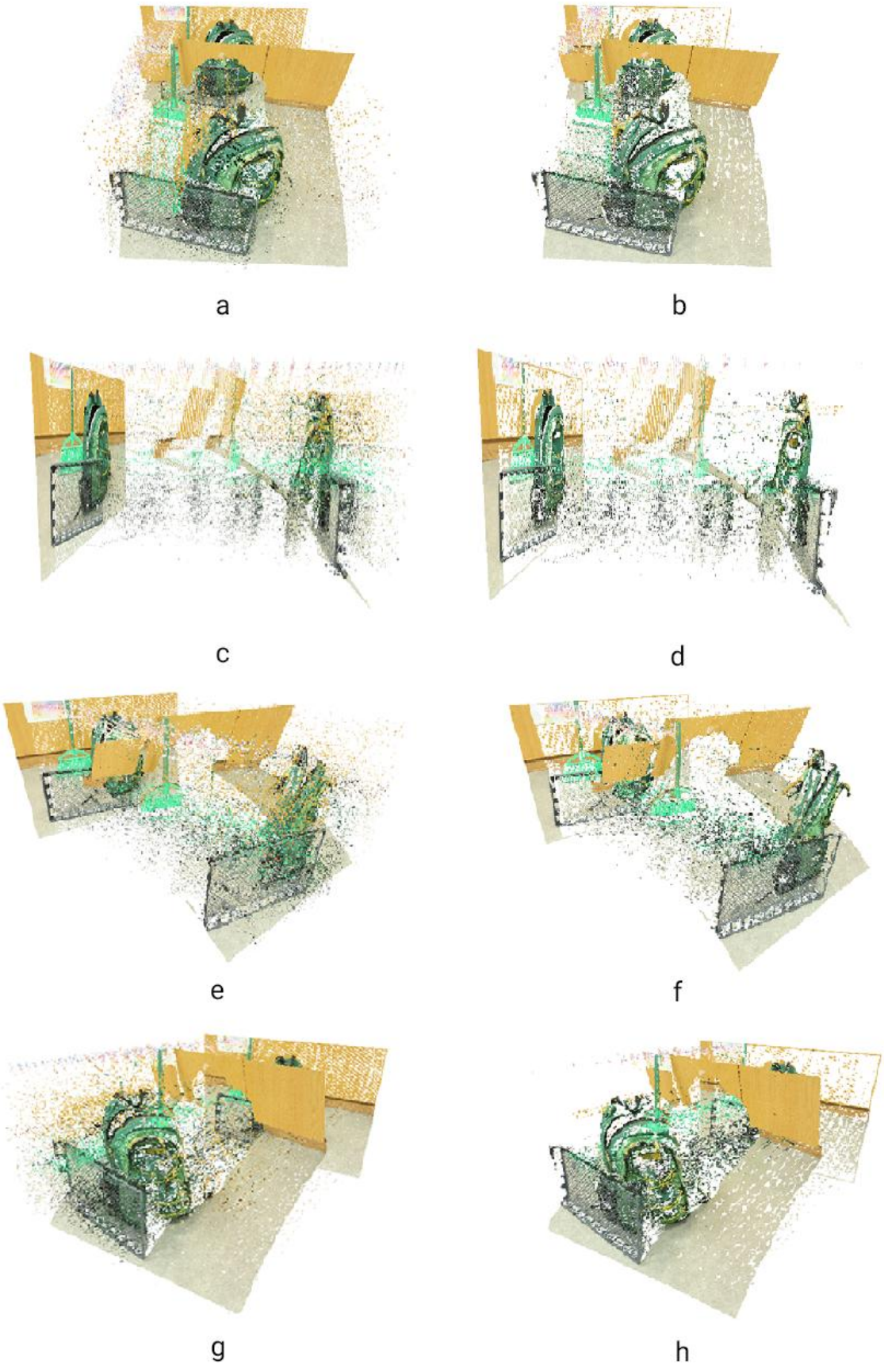
A Figura 29 mostra a comparação entre os mapas de disparidade após a utilização do filtro de mediana com um *kernel* de 17x17 e os mapas de disparidade sem a utilização do filtro. A Figura 29-a, Figura 29-c, Figura 29-e e Figura 29-g mostram o mapa de disparidade após a utilização do filtro. A Figura 29-b, Figura 29-d, Figura 29-f, Figura 29-h mostram o mapa de disparidade sem a utilização do filtro. É possível observar que o resultado é um mapa de disparidade mais suave, e com uma grande parte do ruído atenuado ou removido. Com esse mapa de disparidade em mãos, foi gerado outro mapa de pontos.

Figura 29 – Mapa de Disparidade Filtrado



Fonte (Autoria Própria, 2019)

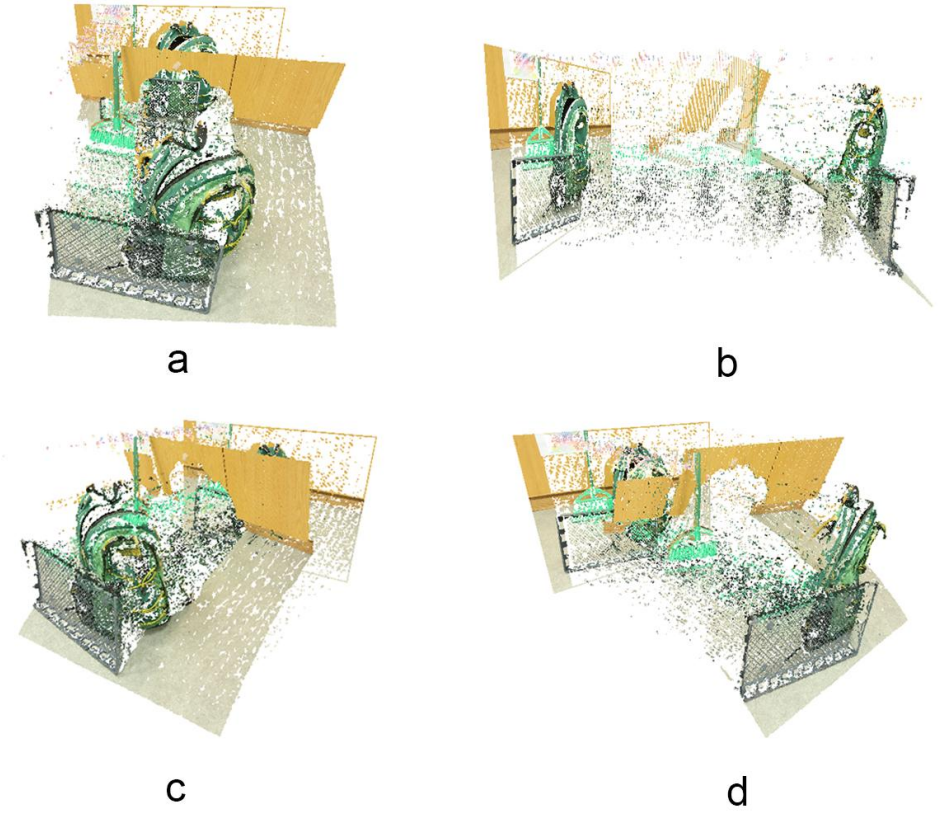
Figura 30 – Comparação Entre Mapa de Pontos



Fonte (Autoria Própria, 2019)

A Figura 30 apresenta um comparativo entre os mapas de pontos gerados. A Figura 30-a, Figura 30-c, Figura 30-e, Figura 30-g representam a visão superior, lateral esquerda, diagonal esquerda, e diagonal direita, respectivamente do mapa de pontos gerado sem a aplicação do filtro. A Figura 30-b, Figura 30-d, Figura 30-f, Figura 30-g representam a visão superior, lateral esquerda, diagonal esquerda, diagonal direita, respectivamente do mapa de pontos gerado a partir do mapa de disparidade em que foi utilizado o filtro da mediana para a eliminação de ruídos. Pode-se observar que em comparação com o mapa de pontos anterior, o resultado final possui grande parte do ruído atenuado.

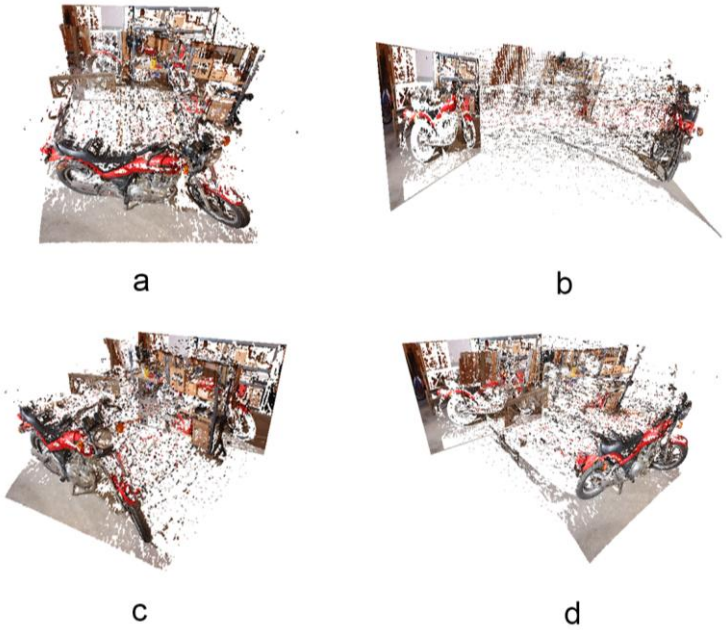
Figura 31 – Mapa de pontos final gerado da imagem de teste 1



Fonte (Autoria Própria, 2019)

A Figura 31 apresenta 4 perspectivas diferentes do mapa de pontos gerado de uma das imagens de teste. Nesse cenário de teste, grande parte do ruído foi atenuado, e assim é possível perceber que a posição da mochila e grade estão bem à frente do plano de fundo.

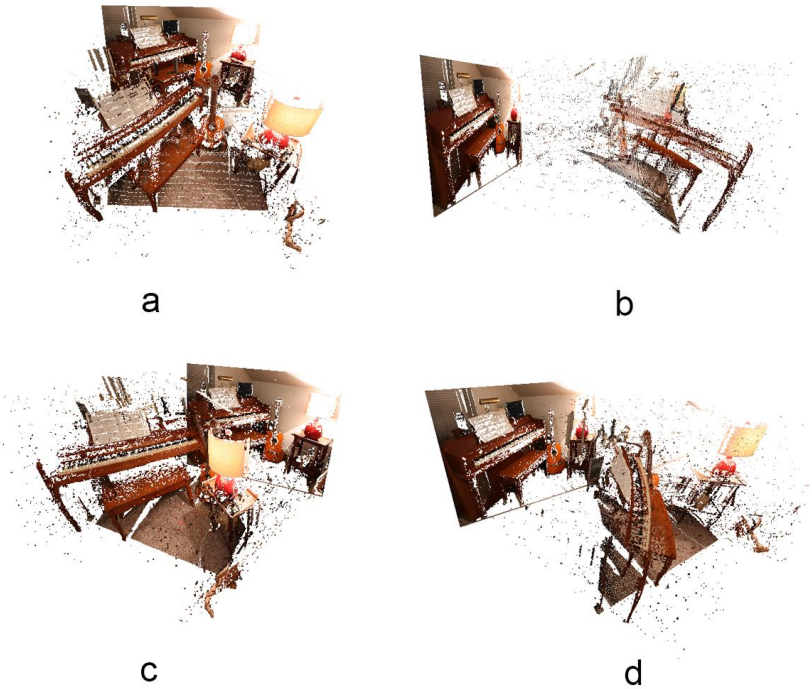
Figura 32 - Mapa de pontos final gerado da imagem de teste 2



Fonte (Autoria Própria, 2019)

A Figura 32 ilustra os resultados obtidos a partir da segunda imagem de teste. Nesse caso existe somente um objeto com maior destaque, sendo ele a motocicleta, e é possível distingui-la do plano de fundo apesar do ruído apresentado no resultado final.

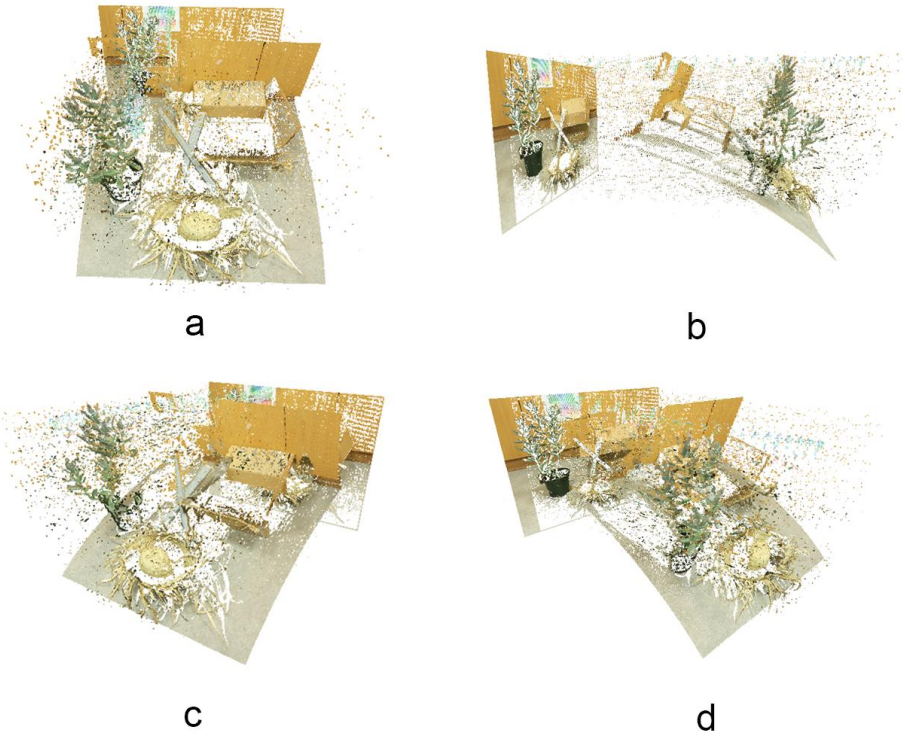
Figura 33 - Mapa de pontos final gerado da imagem de teste 3



Fonte (Autoria Própria, 2019)

A Figura 33 apresenta o mapa de pontos gerado a partir da terceira imagem de testes. Nessa imagem é possível notar dois objetos principais, o piano e abajur, que estão bem destacados do plano de fundo. É possível observar também que esse mapa de pontos possui uma quantidade de ruído menor que os outros resultados.

Figura 34 - Mapa de pontos final gerado da imagem de teste 4



Fonte (Autoria Própria, 2019)

A Figura 34 ilustra o resultado obtido a partir da última imagem de teste. Nessa figura não é possível reconhecer com facilidade os objetos da cena, e existe uma quantidade de ruído grande que interfere na reconstrução da imagem.

5 CONCLUSÃO

A utilização de dados de profundidade está cada vez mais presente em diversas áreas, desde a área médica, automotiva, robótica entre outras, incentivando cada vez mais a criação de métodos que possibilitem obter esse dado. A estereoscopia se mostrou um método confiável, e que possui um custo reduzido, se comparado a outros métodos como o LASER, por exemplo. Por isso a necessidade de se aprimorar cada vez mais o método, de forma a se obter resultados mais confiáveis e robustos.

Este trabalho propôs a geração de um mapa de pontos 3D a partir de imagens estéreo. Para isso foi definido uma metodologia a ser seguida, com objetivos específicos a serem cumpridos, sendo estes: determinar o par estéreo de entrada; aplicar filtros para eliminar ruídos das imagens; gerar o mapa de disparidade correspondente ao par estéreo; calcular a profundidade correspondente ao mapa de disparidade, e por último visualizar os pontos 3D encontrados usando *software* adequado.

O par estéreo foi determinado a partir de um *dataset* disponibilizado pela Universidade de Middlebury, com mais de 30 pares estéreo disponíveis. Foram aplicados dois filtros para atenuar os ruídos que possam interferir no resultado final, sendo estes o filtro da mediana, e o filtro gaussiano. Para gerar o mapa de disparidade, foi utilizado a biblioteca de visão computacional OpenCV, que possui algumas funções para facilitar a manipulação de imagens. A profundidade foi calculada utilizando uma função que engloba a disparidade calculada a partir do par estéreo, e os parâmetros intrínsecos da câmera, que foram disponibilizados pelo autor do *dataset.* A visualização dos pontos se deu no Meshlab, um *software open source,* que possibilita a visualização de arquivos com a extensão PLY, utilizada no desenvolvimento deste trabalho.

Durante o andamento do trabalho, foi necessário realizar algumas mudanças em relação à metodologia definida inicialmente, como a conversão de cores RGB para tons de cinza após a utilização dos filtros e aplicar filtros no mapa de disparidade gerado pelo algoritmo, além do par estéreo de entrada. Estas mudanças foram necessárias para atingir o objetivo geral deste trabalho.

Como resultado, foi apresentado os mapas de pontos obtidos pelo algoritmo a partir dos pares estéreo de entrada. Enquanto alguns possuíram resultados o qual é possível distinguir os objetos da cena e a sua profundidade em relação à câmera, em outras imagens o resultado não permite visualizar os objetos na cena de forma clara, deixando espaço para melhorias no algoritmo.

Em um cenário no qual a profundidade se torna cada vez mais importante para diversas áreas, a estereoscopia surge como um método promissor. A partir de duas imagens, que podem ser obtidas utilizando duas ou mais câmeras, é possível obter a profundidade de objetos na cena, algo antes possível somente com equipamentos específicos, como o LASER. Portanto é necessário cada vez mais aprimorar esta técnica, de modo que esta possua cada vez mais um grau de precisão maior, e possa ser utilizada nas mais diversas áreas em que se julgar necessária.

5.1 TRABALHOS FUTUROS

Apesar do mapa de pontos 3D gerado ser possível identificar os diversos objetos na imagem, ainda existem muitos pontos em que há necessidade de melhoria do trabalho. Como trabalho futuro, é possível apontar:

- Melhorar o processo de correspondência entre os pixels, para que assim o mapa de disparidade obtido seja mais coeso, e com menos inconsistências. Como pode-se observar na Figura 25, partes da imagem que possuem uniformidade elevada, a disparidade não é calculada de forma correta;
- O processo de filtragem do mapa de disparidade também pode ser melhorado, pois apesar do filtro da mediana apresentar um resultado mais uniforme e visivelmente uma quantidade menor de ruído, o filtro não remove todo o ruído.
- Existem algoritmos iterativos como o RANSAC (*Random Sample Consensus*), que utiliza estatística e para estimar a posição de pontos, baseado em um modelo matemático. Ou seja, dado um número X de

pontos, o RANSAC consegue estimar onde cada ponto deve se encaixar na imagem, isso reduziria a quantidade de ruído na imagem final, sem abrir mão dos detalhes da imagem

- Desenvolver uma configuração própria para a captura das imagens estéreo, assim haverá a possibilidade de regular outros parâmetros da câmera, para que a imagem capturada seja a melhor possível;
- Interpolar os pontos do mapa 3D para que estes possuam arestas e faces, assim seria possível recriar um objeto tridimensionalmente somente com imagens obtidas de câmeras estéreo.

REFERÊNCIAS

ACHARYYA, A.; et al. Depth estimation from focus and disparity. In: IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING (ICIP). 2016, Phoenix. **Anais...** Phoenix: University of California San Diego, 2016.

ALRUBAIE, S; HAMEED, A. Dynamic Weights Equations for Converting Grayscale Image to RGB Image. **Journal of University of Babylon for Pure and Applied Sciences,** v. 26, n. 8, p. 122-129, jul. 2018.

AZAD, M; HASAN, M; NASEER, M. Color Image Processing in Digital Image. **International Journal of New Technology and Research,** v. 3, n. 3, p. 56-62, mar. 2017.

BARRY, S. R.; SACKS, O. **Fixing My Gaze:** A Scientist's Journey Into Seeing in Three Dimensions. 1. ed. Nova Iorque: Basic Books, 2010.

BENBASSAT, J.; POLAK, B.C.; JAVITT J.C. **Objectives of Teaching Direct Ophthalmoscopy to Medical Students**. Acta Ophthalmologica, Vol.90, no. 6, p. 503-507, 2012.

BRASIL ESCOLA. Biologia. **Olhos humanos.** Disponível em: < https://brasilescola.uol.com.br/biologia/olhos-humanos.htm>. Acesso em: 29 nov. 2019.

DHOLE, P.; et al. Depth map estimation using SIMULINK tool. In: INTERNATIONAL CONFERENCE ON SIGNAL PROCESSING AND INTEGRATED NETWORKS (SPIN). 2016, Noida. **Anais...** Noida: Maharashtra Institute of Technology, 2016.

FORSYTH, D. A; PONCE, J. **Computer Vision:** A Modern Approach. 2. ed. Nova Jersey: Pearson, 2011.

HARTLEY, R.; KANG, S. B. Parameter-free radial distortion correction with center of distortion estimation. **IEEE Transactions on Pattern Analysis and Machine Intelligence,** v. 29, n. 9, p. 1309-1321, jun. 2007.

HIGUCHI, T. Visuomotor controlo f human adaptative locomotion: Understanding the antecipatory nature. **Frontiers in psycology,** v. 4, p. 277, mar. 2013.

HOWARD, I. P.; ROGERS, B. J. **Binocular vision and stereopsis**. 29. ed. New York: Oxford University Press, 1995.

KOSCHAN, A. Using perceptual atributes to obtain dense depth maps. In: IEEE SOUTHWEST SYMPOSIUM ON IMAGE ANALYSIS AND INTERPRETATION. 1996, San Antonio. **Anais...** San Antonio: Technical University of Berlin, 1996.

KUKELOVA, et al. Efficient Solution to the Epipolar Geometry for Radially Distorted Cameras. In: IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION (ICCV). 2015, Santiago. **Anais...** Santiago: Microsoft Research Ltd., 2016.

LAKSHMANAN, R.; SENTHILNATHAN, R. Depth map based reactive planning to aid in navigation for visually challenged. In: IEEE INTERNATIONAL CONFERENCE ON ENGINEERING AND TECHNOLOGY (ICETECH). 2016, Coimbatore. **Anais...** Coimbatore: SRM University, 2016.

LENNON, V. A. D. **Estudo e análise de diferentes métodos de calibração de câmeras**. 2015. 61 f. Monografia – Departamento acadêmico de informática, Universidade Tecnológica Federal do Paraná. Ponta Grossa, 2015.

MESHLAB. MeshLab. **MeshLab.** 12 out. 2019. Disponível em: <http://www.meshlab.net/>. Acesso em: 12 out. 2019.

MIDDLEBURY. Stereo datasets with ground truth. **Stereo Datasets.** 31 dez. 2014. Disponível em: <http://vision.middlebury.edu/stereo/data/scenes2014/>. Acesso em: 12 out. 2019.

NASA. MARS Exploration Rovers. **Eyes and Senses.** Disponível em: <https://mars.nasa.gov/mer/mission/rover/eyes-and-senses/>. Acesso em: 12 out. 2019.

NOGUEIRA, R. **Análise de Conversão de Imagem Colorida para Tons de Cinza Via Contraste Percebido**. 2016. 46 f. Monografia – Centro de informática, Universidade Federal de Pernambuco. Recife, 2016.

QUORA. **What is the comparison between the human eye and a digital camera.** Disponível em: <https://www.quora.com/What-is-the-comparison-between-the-human-eye-and-a-digital-camera>. Acesso em: 29 nov. 2019.

RAAJAN, N; et al. Disparity Estimation from Stereo Images. **International Conference on Modeling Optimization and Computing,** v. 29, n. 9, p. 1309-1321, jun. 2007.

RAHMAN, N.; KROUGLICOF, N. An Efficient Camera Calibration Technique Offering Robustness and Accuracy Over a Wide Range of Lens Distortion. **IEEE Transactions on Image Processing,** v. 38, p. 462-472, dez. 2012.

RAMAMURTHY, M.; LAKSHMINARAYANAN, V. **Human Vision and Perception**. Springer, 2015.

SHETE, P. P.; SARODE, D. M.; BOSE, S. K. A real-time stereo rectification of high definition image stream using GPU. In: INTERNATIONAL CONFERENCE ON ADVANCES IN COMPUTING, COMMUNICATIONS AND INFORMATICS (ICACCI). 2015, Nova Delhi. **Anais...** Nova Delhi: Bhabha Atomic Research Centre, 2014.

SILVA, S. K. G; MARQUES, F. L. S. N. Depth Perception Evaluation with Different Stereoscopic Techiniques: A Case Study. In: XVII SYMPOSIUM ON VIRTUAL AND AUGMENTED REALITY (SVR). 2015, São Paulo. **Anais...** São Paulo: Universidade de São Paulo, 2015.

STACK OVERFLOW. Computing a Depth Map from Stereo Images. **OpenCV.** 9 jul. 2015. Disponível em: < https://stackoverflow.com/questions/27726306/python-opencv-computing-a-depth-map-from-stereo-images>. Acesso em: 03 jun. 2017.

TRUCCO, E; VERRI, A. **Introductory Techiniques for 3-D Computer Vision**. 1. ed. Nova Jersey: Prentice Hall, 1998.

WANG, Y.; WU, Y. A Study on the Different Neural Mechanisms of Stereopsis between Fine Crossed and Uncrossed Disparity. In: INTERNATIONAL CONFERENCE ON ROBOTS & INTELLIGENT SYSTEM (ICRIS). 2016, Zhangjiajie. **Anais...** Zhangjiajie: Changchun University of Science and Technology, 2016.

XU, G; ZHANG, Z. **Epipolar Geometry in Stereo, Motion and Object Recognition:** A Unified Approach. 1. ed.: Springer, 1996.

YOSHINARI, K.; HOSHI, Y.; TAGUCHI, A. Color Image Enhancement in HIS Color Space Without Gamut Problem. **International Symposium on Comunications, Control and Signal Processing,** may. 2014, Athens.

YUQUAN, X.; VIJAY, J.; SEIICHI, M.; HOSSEIN T.; KAZUHISA, I.; SAKIKO, N . 3D point cloud map based vehicle localization using stereo camera. **IEEE Intelligent Vehicles Symposium,** june. 2017, Los Angeles.

ZHENG, Y.; SILONG, P. A practical roadside câmera calibration method based on least squares optimization. **IEEE Transactions on Intelligent Transportations Systems,** v. 15, n. 2, p. 831-843, nov. 2013.