PODER EXECUTIVO
MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO AMAZONAS
INSTITUTO DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

FELIPE GOMES DE OLIVEIRA

Estimação da Profundidade por meio da Fusão de Dados de Energia Visual de Múltiplas Câmeras

Manaus
2011

Felipe Gomes de Oliveira

Estimação da Profundidade por meio da Fusão de Dados de Energia Visual de Múltiplas Câmeras

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Informática do Departamento de Ciência da Computação da Universidade Federal do Amazonas, como parte dos requisitos necessários para a obtenção do título de Mestre em Informática.

Orientador: Prof. Dr. José Luiz de Souza Pio

Manaus

2011

Felipe Gomes de Oliveira

Estimação da Profundidade por meio da Fusão de Dados de Energia Visual de Múltiplas Câmeras

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Informática do Departamento de Ciência da Computação da Universidade Federal do Amazonas, como parte dos requisitos necessários para a obtenção do título de Mestre em Informática.

Banca Examinadora

Prof. Dr. José Luiz de Souza Pio (Orientador)
Universidade Federal do Amazonas

Prof. Dr. José Reginaldo Hughes de Carvalho
Universidade Federal do Amazonas

Prof. Dr. Cícero Augusto Mota Cavalcante
Universidade Federal do Amazonas - Itacoatiara

Manaus – 2011

À Deus por permitir alcançar tudo o que conquistei em minha vida.

Agradecimentos

A minha filha Dalila por me permitir a emoção da paternidade.

A minha esposa Daniela por estar sempre ao meu lado em todos os momentos.

Aos meus pais Pedro e Josemeire pelo amor, carinho, dedicação e principalmente pela educação proporcionada.

As minhas irmãs Rochelle e Micaelle pelo amor e compreensão.

Ao meu orientador José Luiz de Souza Pio por todo o suporte dado durante o mestrado.

A todos os meus amigos por me alegrar nos momentos de dificuldade.

A FAPEAM, pelo apoio financeiro.

Ao CNPq pelo apoio na implementação do laboratório e uso dos equipamentos do projeto...

A todos aqueles que tiveram contribuição direta ou indireta para a realização deste trabalho.

Sumário

Lista de Abreviaturas e Siglas **8**

Lista de Figuras **9**

Lista de Tabelas **15**

Resumo **17**

Abstract **18**

1 Introdução **19**

1.1 Considerações Iniciais 19

1.2 Motivação 21

1.3 Justificativa 22

1.4 Objetivos 23

1.4.1 Objetivo Geral 23

1.4.2 Objetivos Específicos 24

1.5 Organização do trabalho 24

2 Trabalhos Relacionados **25**

2.1 Fusão de Dados Visuais 25

3 Fusão de Dados de Energia Visual Para Estimação de Profundidade **32**

3.1 Definição do Problema 32
3.2 Metodologia 33
3.2.1 Aquisição de Imagens 35
3.2.2 Estimação da Energia de Foco 37
3.2.3 Fusão de Dados por Minimização de Energia 40

4 Resultados Experimentais 45
4.1 Descrição dos Experimentos e do Aparato Experimental 45
4.2 Apresentação dos Resultados 50
4.2.1 Experimento 1 - Avaliação comparativa entre a abordagem proposta e as técnicas clássicas estéreo 50
4.2.2 Experimento 2 - Validação da robustez do método frente a variação de luminosidade 56

5 Conclusões e Trabalhos Futuros 68
5.1 Conclusões 68
5.2 Limitações do Trabalho 69
5.3 Trabalhos Futuros 69
5.4 Considerações Finais 70

Referências bibliográficas 71

Apêndices 72

A Reconstrução da forma pelo Estéreo Fotométrico 73

B Reconstrução da forma a partir do foco 77

C Estado da arte: Estimações de Profundidade 81

Lista de Abreviaturas e Siglas

CCD	*Charge Coupled Device* (Dispositivo de carga acoplado)
SFF	*Shape from Focus* (Forma a partir do foco)
SML	*Sum Modified Laplancian* (Somatório do Laplaciano Modificado)
SSD	*Sum of Squared Differences* (Soma das Diferenças Quadráticas)
SAD	*Sum of Absolute Differences* (Soma das Diferenças Absolutas)
NCC	*Normalized Cross Correlation* (Correlação Cruzada Normalizada)
RMSE	*Root Mean Square Error* (Raiz do Erro Médio Quadrático)
BMP	*Bad Matching Pixel* (Casamento de pixels Ineficiente)
FOV	*Field of View* (Campo de Visão)
GHz	*Gigahertz*
GB	*Gigabyte*
RAM	*Randon Acess Memory* (Memória de Acesso Randômico)

Lista de Figuras

2.1 Sistema de câmeras combinado com um robô industruial. Conjunto de câmeras utilizado no processo de aquisição de imagens para reconstrução da estrutura tridimensional da cena. 29

2.2 Processo de estimação de profundidade e restauração de imagens. Este procedimento utiliza imagens com variação de foco em parte das imagens possibilitando determinar a profundidade na cena e determinar a melhor condição focal na imagem. 31

3.1 Duas câmeras do sistema de monitoramento que são convenientemente posicionadas obtém duas sequências de imagens da cena. Destas sequências são obtidos dados de profundidade pela variação de foco. Na determinação das correspondências a relação de energia envolvida no processo de aquisição de imagens ponderada pelas distâncias obtidas a partir do foco é usada em um modelo de fusão baseado em Minimização de energia. 34

3.2 Sistema distribuído de câmeras configurado como uma par estéreo para a aquisição de duas sequências de imagens com variação de foco. 36

3.3 Processo de captura de imagens com variação de foco. A partir de mudanças na distância focal da câmera é possível obter um conjunto de imagens com diferentes situações focais. 37

3.4 Espalhamento gaussiano da energia luminosa para variação de foco da sequência de imagens. 38

4.1 Modelo de imagens disponíveis na base. Nas Figuras (a) e (b) são apresentadas as imagens capturadas pelas câmeras esquerda e direita, respectivamente. Enquanto que nas Figuras (c) e (d) são mostrados os mapas de profundidade referente às imagens da esquerda e direita, respectivamente. 47

4.2 Exemplo de um conjunto de imagens com variação do parâmetro intrínseco, distância focal. A série de imagens apresenta informação de profundidade a medida que o foco varia controladamente pelas imagens. 48

4.3 Resultados obtidos na etapa de experimentação. As imagens (a) e (b) representam respectivamente os mapas de profundidade reais obtidos a partir das imagens capturadas pelas câmeras da esquerda e direita. Na imagem (c) é apresentado o mapa de profundidade obtido a partir da técnica proposta. Enquanto que as imagens (d), (e) e (f) consistem nos resultados obtidos através das técnicas de estéreo clássicas, SSD, SAD e NCC, respectivamente. 51

4.4 Resultados obtidos na etapa de experimentação. As imagens (a) e (b) representam respectivamente os mapas de profundidade com informações reais de profundidade obtidos a partir das imagens capturadas pelas câmeras da esquerda e direita. Na imagem (c) é apresentado o mapa de profundidade obtido a partir da técnica minimização de energia. Enquanto que as imagens (d), (e) e (f) consistem nos mapas obtidos através do SSD, SAD e NCC, respectivamente. . . 52

4.5 Resultados obtidos a partir do processo de fusão proposto. As imagens (a) e (b) consistem nos mapas de profundidade reais capturados pelas câmeras da esquerda e direita. A imagens (c) representa o mapa de profundidade resultante do processo de fusão por minimização de energia. Já nas imagens (d), (e) e (f) são mostrados os mapas providos pelas técnicas estéreo SSD, SAD e NCC. 53

4.6 Comparação dos resultados usando RMSE. Gráfico comparando técnicas clássicas do estéreo e a estimação baseada em Funções de Energia com o mapa de profunidade real da imagem esquerda. 55

4.7 Comparação dos resultados usando BMP. Gráfico comparando técnicas clássicas do estéreo e a estimação baseada em Funções de Energia com o mapa de profunidade real da imagem esquerda. 55

4.8 Modelo de imagens com variação de iluminação. As imagems mostradas acima apresentam mudanças na luminosidade da cena de modo a permitir uma análise mais aprofundada dos impactos da iluminação na estimação de profundidade. Na imagem (a) de cada pixel foi subtraído 15, na imagem (b) foi subtraído 25, a imagem (c) apresenta brilho normal, na imagem (d) foi adicionado 15 e na imagem (e) foi adicionado 25. 57

4.9 Resultados obtidos na etapa de experimentação com diminuição da iluminação (25). As imagens (a) e (b) representam respectivamente os mapas de profundidade reais obtidos a partir das imagens capturadas pelas câmeras da esquerda e direita. Na imagem (c) é apresentado o mapa de profundidade obtido a partir da técnica proposta. Enquanto que as imagens (d), (e) e (f) consistem nos resultados obtidos através das técnicas de estéreo clássicas, SSD, SAD e NCC, respectivamente. 58

4.10 Resultados obtidos na etapa de experimentação considerando diminuição de luminosidade na cena. As imagens (a) e (b) representam respectivamente os mapas de profundidade com informações reais de profundidade obtidos a partir das imagens capturadas pelas câmeras da esquerda e direita. Na imagem (c) é apresentado o mapa de profundidade obtido a partir da técnica minimização de energia. Enquanto que as imagens (d), (e) e (f) consistem nos mapas obtidos através do SSD, SAD e NCC, respectivamente. 59

4.11 Resultados obtidos a partir do processo de fusão proposto levando em conta baixa iluminação do ambiente monitorado. As imagens (a) e (b) consistem nos mapas de profundidade reais capturados pelas câmeras da esquerda e direita. A imagens (c) representa o mapa de profundidade resultante do processo de fusão por minimização de energia. Já nas imagens (d), (e) e (f) são mostrados os mapas providos pelas técnicas estéreo SSD, SAD e NCC. 60

4.12 Comparação dos resultados usando RMSE. Gráfico comparando técnicas clássicas do estéreo e a estimação baseada em Funções de Energia com o mapa de profunidade real da imagem esquerda. 61

4.13 Comparação dos resultados usando BMP. Gráfico comparando técnicas clássicas do estéreo e a estimação baseada em Funções de Energia com o mapa de profunidade real da imagem esquerda. 61

4.14 Resultados obtidos na etapa de experimentação com variação da iluminação (25). As imagens (a) e (b) representam respectivamente os mapas de profundidade reais obtidos a partir das imagens capturadas pelas câmeras da esquerda e direita. Na imagem (c) é apresentado o mapa de profundidade obtido a partir da técnica proposta. Enquanto que as imagens (d), (e) e (f) consistem nos resultados obtidos através das técnicas de estéreo clássicas, SSD, SAD e NCC, respectivamente. 63

4.15 Resultados obtidos na etapa de experimentação considerando o aumento de luminosidade na cena. As imagens (a) e (b) representam respectivamente os mapas de profundidade com informações reais de profundidade obtidos a partir das imagens capturadas pelas câmeras da esquerda e direita. Na imagem (c) é apresentado o mapa de profundidade obtido a partir da técnica minimização de energia. Enquanto que as imagens (d), (e) e (f) consistem nos mapas obtidos através do SSD, SAD e NCC, respectivamente. 64

4.16 Resultados obtidos a partir do processo de fusão proposto levando em conta alta iluminação do ambiente monitorado. As imagens (a) e (b) consistem nos mapas de profundidade reais capturados pelas câmeras da esquerda e direita. A imagens (c) representa o mapa de profundidade resultante do processo de fusão por minimização de energia. Já nas imagens (d), (e) e (f) são mostrados os mapas providos pelas técnicas estéreo SSD, SAD e NCC. 65

4.17 Comparação dos resultados usando RMSE. Gráfico comparando técnicas clássicas do estéreo e a estimação baseada em Funções de Energia com o mapa de profunidade real da imagem esquerda. 66

4.18 Comparação dos resultados usando BMP. Gráfico comparando técnicas clássicas do estéreo e a estimação baseada em Funções de Energia com o mapa de profunidade real da imagem esquerda. 67

A.1 Imagens digitais capturadas por um sistema estéreo de visão. A Figura (a) representa a imagem capturada pela câmera esquerda e a Figura (b) representa a imagem capturada pela câmera direita. . . . 74

A.2 Sistema de visão stereo composto por duas câmeras separadas por uma distância B denominada *Baseline*. 74

A.3 Etapa do processo de Visão estéreo. Inicialmente, o sistema de câmeras captura um par de imagens. Segundo, é executado o processo de retificação de imagens para alinhamentos das coordenadas. Então, são determinados os pontos correspondentes dentre as duas imagens. Despois, é computado o cálculo do mapa de disparidade da cena monitorada. Finalmente, é encontrado o mapa de profundidade contendo informações a respeito da estrutura tridimensional do ambiente. . . . 75

B.1 Sistema de visão usado na captura de uma série de imagens. Os parâmetros intrínsecos da câmera são variados a cada aquisição obtendo assim imagens com variação de foco. 78

B.2 Esquematização do processo de determinação da forma tridimensional de uma cena a partir da informação de foco nas imagens. No primeiro instante uma sequência de imagens é capturada com variação do foco. Posteriormente é realizada a medição do foco dentre as imagens, determinando para cada ponto do conjunto de imagens qual apresenta melhor foco. Em seguida é realizado o calculo de distância com base nos pixels mais bem focados da etapa anterior e por fim é contruído o mapa de profundidade da cena. 79

B.3 Modelo de convolução realizado para determinar o pixel melhor focado dentre o conjunto de imagens. Neste modelo o foco de cada ponto em cada imagem é medido visando determinar em qual das imagens o ponto encontra melhor situação focal. 80

Lista de Tabelas

C.1 Técnicas de Reconstrução da forma de cenas com características e breve descrição das abordagens. 82

Resumo

Este trabalho propõe uma abordagem de Fusão de Dados Visuais para estimar a estrutura tridimensional de uma cena a partir de sequências de imagens obtidas por meio de duas ou mais câmeras. Os métodos convencionais para estimar mapas de profundidade apresentam desvantagens relacionadas a mundanças na iluminação do ambiente e posicionamento de câmeras. Por essa razão, foi proposta uma estratégia de Fusão de Dados baseada em minimização de energia para aprimorar as medições proporcionadas pela disparidade entre pixels de uma imagem e pela variação de foco. A abordagem proposta faz uso de uma rede distribuída de sensores visuais utilizando um par de câmeras estéreo sem restrições de oclusão ou iluminação no processo de captura de imagens. A função de energia foi usada para integrar múltiplos frames e inferir a coerência geométrica contida na cena.

Para avaliar os resultados obtidos foram utilizadas métricas da literatura através de medições de similaridade entre técnicas de estéreo tradicionais e a estratégia desenvolvida. Os experimentos foram conduzidos a partir de imagens de cenas reais, e as informações de profundidade estimadas foram qualitativamente superior que os resultados obtidos pelos métodos tradicionais. Tais informações demonstram a qualidade dos resultados alcançados pela técnica proposta.

Palavras-Chave: Fusão de Dados Visuais, Função de Energia, Shape from Focus, Shape from Stereo

Abstract

This research presents a visual data fusion approach to recover dense depth map from sequences of images. The conventional methods to estimate depth map have many drawbacks with respect to environment illumination changes and camera positioning. We propose a Global optimization data fusion strategy to improve the measurements from stereo and focus depth maps. Different from typical stereo and focus fusion techniques, we use a single pair of stereo cameras to acquire series of images scenes without occlusion and illumination constraints. Then, we use Energy Functional fusion to associate the geometric coherence with multiple frames. In order to evaluate the results we defined a metric using similarity measurements between traditional stereo and the proposed approach. The experiments are performed in real scene images, and the estimated mapping was superior than those found using traditional stereo methods, which demonstrates the good performance and robustness of our approach.

KEYWORDS: Visual Data Fusion, Energy Functional, Shape from Focus, Shape from Stereo

Capítulo 1

Introdução

1.1 Considerações Iniciais

Este trabalho trata da utilização de técnicas de fusão de dados para a combinação de dados visuais de múltiplas câmeras dispostas no ambiente, aplicadas na estimação da estrutura tridimensional de uma cena. Ambientes dessa natureza são, em geral, formados por circuitos fechados de TV, que capturam imagens da cena deixando a encargo do operador humano qualquer forma de análise ou reconhecimento.

O aumento da produção de sensores e de câmeras a um baixo custo tem impulsionado a pesquisa para a integração das informações providas por redes de dispositivos, visando aprimorar os resultados obtidos, proporcionando maior área de cobertura e auxiliando o operador humano no exaustivo trabalho de análise ou reconhecimento de padrões visuais (Salustiano, 2002).

Este trabalho se insere na área Visão Computacional. Essa é uma área científica que busca mimicar a cognição humana e a habilidade do ser humano em tomar decisões de acordo com as informações contidas na imagem (Pedrini and Schwartz, 2007). A percepção da forma de objetos em uma cena é um dos problemas mais recorrentes encontrados na Visão Computacional, sendo uma tarefa extremamente trivial para

os seres humanos, todavia é um grande desafio para a Computação.

Para resolver problemas desta natureza, existem vários métodos clássicos na literatura de Visão Computacional muitos dos quais encontram-se concentrados em um arcabouço denominado *Shape-from-X* (Trucco and Verri, 1998). Entretanto, cada método aborda o problema considerando características distintas (Trucco and Verri, 1998). Além disso, eles podem apresentar desvantagens relacionadas a superfície, visibilidade, periodicidade da estrutura e iluminação, levando a estimações imprecisas (Gheta et al., 2007).

As técnicas tradicionais de estimação aplicadas separadamente apresentam deficiências referentes às desvantagens de cada estratégia. Com isso, a Fusão de Dados surge como uma eficiente alternativa para aumentar a precisão nas medições (Abidi and Gonzalez, 1992).

O processo de combinação de dados provenientes de múltiplos sensores de mesma natureza ou de naturezas distintas é denominado Fusão de Dados (Abidi and Gonzalez, 1992). O principal objetivo da fusão é o fornecimento de informação com maior qualidade, permitindo assim a redução de falhas no processo de estimação (Abidi and Gonzalez, 1992). Quando sensores de mesma natureza são combinados de maneira adequada, agrega-se maior exatidão, confiabilidade e o aumento da precisão do valor final (Abidi and Gonzalez, 1992). A utilização de sensores de natureza distintas, que trocam informações entre si em um esquema de cooperação, além da redução dos dados incorretos amplia a capacidade de obtenção de informações do sistema (Abidi and Gonzalez, 1992) (Baltzakis et al., 2003).

Neste trabalho é proposta uma abordagem para combinar as informações de energia do foco e do par de imagens capturado, aprimorando a informação de profundidade obtida a partir dos conjuntos de imagens do ambiente monitorado por meio de múltiplas câmeras (Horn, 1986). A etapa de fusão é conduzida por meio de um processo de minimização de energia o qual associa medições de energia proporcionadas

pelas disparidades e pelas variações de foco para a combinação dos dados visuais. Os resultados alcançados pela fusão de dados são avaliados usando como medida de similaridade a Raiz do Erro Médio Quadrático (RMSE - *Root Mean Square Error*) e Análise de casamento ineficiente (BMP - *Bad Matching Pixel*) (Pedrini and Schwartz, 2007) através da comparação com os resultados providos pelo estéreo.
As contribuições deste trabalho estão relacionadas à estratégia de fusão de dados por meio da minimização de energia e na forma de avaliação dos resultados obtidos no processo de Fusão Multisensorial. Especialmente na combinação das informações de energia proporcionadas pela determinação das correspondências e pelas informações da forma a partir do foco e na aplicação das métricas de similaridade RMSE e BMP para a avaliação dos resultados obtidos, possibilitando a validação da abordagem proposta. Considerando o problema da oclusão como restrição do método desenvolvido, os resultados alcançados mostraram-se pouco impactados frente aos problemas encontrados, tais como iluminação, possibilitando obter resultados melhores que os métodos clássicos.

1.2 Motivação

A principal motivação para a realização deste trabalho é o desenvolvimento de uma abordagem que visa integrar informações visuais para a obtenção de profundidade da cena. Essas informações são obtidas a partir de imagens capturadas por um par de câmeras fazendo uso da cooperação dentre as energias envolvidas na estimação para obter informações tridimensionais mais próximas da realidade. O desenvolvimento dessa aplicação amplia a confiabilidade da informação e eficácia do sistema em diversas aplicações reais, tais como o monitoramento de grandes regiões, Navegação Robótica e Inspeção Industrial.

Atualmente diversas aplicações utilizam múltiplos sensores combinados como meio de propiciar informações mais acuradas. Esses sistemas utilizam os dados obtidos para proporcionar ao usuário aplicações inteligentes que o poupa de tarefas enfadonhas. O Sensoriamento Remoto é uma técnica que utiliza métodos de extração de informação tridimensional a partir de imagens capturadas por satélites. Essa aplicação usa tais informações para mapear a área sensoriada e com isso monitorar a região de interesse com grande grau de precisão quanto à informação extraída. Um dos grandes desafios contidos na área de Robótica consiste em possibilitar que um robô navegue por um dado ambiente sem colidir com os obstáculos dispostos na cena (Baltzakis et al., 2003). A aplicação dessas técnicas em Robótica Móvel facilita o processo de tolerância a falhas, garantindo que o sistema continue a obter informação 3-D caso ocorra a perca de um dos dispositivos visuais de captura de imagens. Desse modo um algoritmo que estime profundidade com precisão auxiliaria de maneira efetiva a execução desta tarefa.

No setor industrial, por sua vez, uma das tarefas mais importantes consiste em assegurar a qualidade de seus produtos. Para isso, um processo de inspeção que garanta um controle de qualidade rigoroso da produção é fundamental. Isso é possível através de algoritmos de medição de distância, que podem solucionar problemas como a deteção da ausência ou presença de componentes, dentre outros problemas (Martins and Dim, 2008).

1.3 Justificativa

Neste trabalho é proposta uma abordagem para integrar as informações providas por mútilplos sensores dispostos em uma rede distribuída, considerando uma par de câmeras com configuração de um sistema de visão estéreo. Essa aplicação combina os dados visuais obtidos pelo brilho dos pixels do par de imagens e pela variação de

foco visando alcançar uma melhoria representativa a partir das informações de energia luminosa do ambiente monitorado. As abordagens convencionais encontradas na literatura que tratam o problema de estimação de profundidade a partir de dados visuais fazem uso de técnicas isoladas que visam determinar as informações tridimensionais de uma cena. No entanto, a informação computada por meio de apenas uma técnica não é suficiente para uma estimação de profundidade com boa acuidade (Nayar and Nakagawa, 1994) (Subbarao et al., 1997). Nesse sentido, alguns trabalhos da literatura buscaram realizar a fusão das técnicas isoladas visando o uso de informações redundantes, pois segundo Abidi (1992) a combinação de dados de mesma natureza agrega maior exatidão e confiabilidade. Dentre os trabalhos relacionados que realizam a Fusão de Dados Visuais é possível observar várias restrições quanto a iluminação, textura dos objetos e posição dos objetos em relação a câmera. Outras condições são impostas para o bom funcionamento das abordagens encontradas na literatura, tais como necessidade de parâmetros e grande quantidade de câmeras (V. Michael Bove, 1990) (Subbarao et al., 1997) (Krotkov and Bajcsy, 1993) (Gheta et al., 2007).

1.4 Objetivos

Para a melhor compreensão dos objetivos deste trabalho, os mesmos foram divididos em Geral e Específicos, como a seguir.

1.4.1 Objetivo Geral

O objetivo geral deste trabalho é propor uma abordagem para a fusão de dados integrando informações de energia do foco e do par de imagens fazendo uso de um ambiente de múltiplas câmeras.

1.4.2 Objetivos Específicos

Os objetivos específicos são apresentados abaixo:

- Garantir a qualidade da informação de profundidade estimada, provendo maior confiabilidade ao sistema de reconstrução da estrutura tridimensional da cena;
- Analisar técnicas de fusão baseadas em Funções de Energia considerando um ambiente com múltiplas câmeras;
- Montar protótipo experimental para obter resultados em cenários reais e simulados;
- Avaliar os resultados por meio da comparação entre técnicas de estimação de profundidade convencionais.

1.5 Organização do trabalho

Este texto está organizado em seis seções. A seção seguinte oferece uma visão geral do estado da arte por meio dos principais trabalhos relacionados com fusão de dados visuais e suas aplicações no contexto deste trabalho. A metodologia elaborada para a abordagem proposta é mostrada na Seção 3. A Seção 4 mostra os resultados experimentais. Na Seção 5 são apresentadas as conclusões e as expectativas para trabalhos futuros. Em seguida as referências bibliográficas consideradas para essa pesquisa são mostradas. E por fim, os Apêndices A, B e C com informações adicionais para complementar a compreensão do trabalho proposto são evidenciados.

Capítulo 2

Trabalhos Relacionados

Este trabalho integra as informações de energia providas pela variação de foco e pela disparidade encontrada entre um par de imagens. Neste capítulo serão apresentos os trabalhos relacionados, destacando o estado da arte e considerando a Fusão de dados visuais. É importante mencionar que os trabalhos relacionados às técnicas de estimação de profundidade (SFS e SFF) são descritos no Apêndice C.

2.1 Fusão de Dados Visuais

O aumento da produção de dispositivos de sensoriamento tem favorecido a aplicação de sensores mais baratos, substituindo o uso de sensores mais caros com grande precisão, por sensores de baixo custo atuando em rede e compartilhando as informações do ambiente ou fenômeno monitorado (Salustiano, 2002). Esse aspecto importante tem aguçado o interesse científico na fusão das informações proporcionadas por múltiplos sensores. Estas técnicas de integração de dados podem ser empregadas em uma grande variedade de aplicações, tais como: na reconstrução da forma de uma determinada cena ou objeto (Frese and Gheta, 2006), na determinação da trajetória de objetos de interesse (Dhar et. al., 1996), na navegação de robôs (DeSouza and

Kak, 2002) e em sistemas de monitoramento e segurança (Salustiano, 2002), com aumento considerável do desempenho do sistema e a melhora da qualidade dos resultados. Muitas aplicações hoje em dia fazem uso de sensores para monitorar regiões com objetivos distintos. Desse modo os sensores visuais têm se apresentado como uma alternativa viável para a constituição de redes distribuídas de câmeras para o monitoramento de ambientes. Tais sistemas de câmeras agregam mais confiança e precisão, proporcionando uma ampliação das informações extraídas das cenas sensoriadas (Salustiano, 2002).

Para que os dados obtidos através dos dispositivos visuais possam ser utilizados com um propósito bem definido é necessário que seja realizada uma combinação dessas informações. A integração destes dados gerados por múltiplos sensores de mesma natureza ou de naturezas distintas é denominada Fusão de Dados Multisensoriais (*Multisensor Data Fusion*). Possibilitando maior qualidade da informação, reduzindo a taxa de falhas no processo de estimação (Abidi and Gonzalez, 1992). As informações geradas pelos sensores podem ser combinadas fazendo uso de uma série de níveis de representação, dependendo das necessidades do sistema e do grau de similaridade dos sensores envolvidos. Os níveis de representação existentes consideram: sinal, pixel, característica ou símbolo. Tais níveis são descritos posteriormente.

Os níveis de Fusão Multisensorial podem ser diferenciados a partir das características encontradas no ambiente monitorado, características da técnica utilizada e dos dispositivos. De onde podemos destacar os seguintes aspectos considerados no processo de fusão: Tipo de informação sensoriada, Representação do nível de informação, Modelo da informação sensoriada, Grau de registro, Maneiras de realizar o registro, Método de Fusão e Melhoria para a Fusão.

A Fusão ao nível de sinal inicia seu processamento a partir de informações providas por sensores de uma determinada natureza. Então, as informações são combinadas resultando em uma informação do mesmo tipo, porém, de melhor qualidade. Uma

técnica utilizada no nível de sinal é a média ponderada, considerando a variância estimada como o peso de cada informação envolvida no procedimento de integração (Abidi and Gonzalez, 1992).

A Fusão ao nível de pixel consiste na combinação de informações exclusivamente visuais (imagens) possibilitando melhorias em tarefas que envolvem processamento de imagens, tais como: segmentação, extração de características e restauração de imagens. Algumas métodos de fusão consolidados deste nível são: Filtros lógicos, Morfologia matemática, Álgebra de imagens e Simulated Annealing (Abidi and Gonzalez, 1992).

A Fusão ao nível de característica baseia-se em duas funcionalidades principais. Incrementar a probabilidade de a característica extraída a partir da informação fornecida por um sensor corresponder a um aspecto importante do ambiente. Por outro lado pode ser utilizada como meio de gerar características compostas adicionais para posterior utilização pelo sistema. Algumas técnicas de fusão para este nível são Estatística Tie, Estimação Gauss-Markov e Filtro de Kalman extendido (Abidi and Gonzalez, 1992).

Por último temos a Fusão ao nível de símbolo que trata as informações proporcionadas por vários sensores com alto nível de abstração. Este nível apresenta como restrição o fato de poder somente combinar informações quando os sensores são bastante diferentes ou quando percebem características distintas do ambiente monitorado. As técnicas normalmente apresentam um grau de crença que representa o peso que a informação sensoriada tem no modelo. Métodos que correspondem ao nível de símbolo são: Estimação Bayesiana, Dempster-Shafer (baseados em confiança) e Lógica Fuzzy (Abidi and Gonzalez, 1992).

A Fusão Sensorial realiza a integração de diversos sensores possibilitando a combinação das informações em um ou mais destes níveis, podendo envolver diversas técnicas. Este trabalho consiste na utilização de um conjunto de métodos e fer-

ramentas com o objetivo de obter informações mais acuradas através de dados de baixa qualidade. Os trabalhos que fazem uso de técnicas de Fusão de Dados no contexto desta pesquisa são: Bove (1990) que define modelos probabilísticos considerando uma técnica baeada no foco e o estéreo como medidas de estimação de profundidade. A fusão das informações computadas é realizada por meio da média ponderada, onde são usadas as estimativas de variância local como os pesos da média ponderada. A relação matemática encontrada por Bove para realizar o cálculo citado acima é:

$$E = \frac{\sum_{x=1}^{n} \frac{1}{\sigma_i^2} E_i}{\sum_{x=1}^{n} \frac{1}{\sigma_i^2}}. \tag{2.1}$$

Estas estimativas de variância são calculadas a partir do limite inferior Cramer-Rao. No entanto, Bove utilizou um algoritmo simples de correlação para Estéreo que inviabiliza a realização de uma análise estatística similar para algoritmos de correspondência mais sofisticados.

Na abordagem desenvolvida por Krotkov (1993), também é proposta a integração das técnicas estéreo e de determinação da forma a partir do foco. Em sua pesquisa é argumentado que ambas técnicas (SFF e SFS) são inconsistentes na medição e, portanto, a média ponderada como métrica de fusão não se justifica. Para isso, Krotkov propôs uma abordagem para mensurar a consistência de cada medição utilizada na combinação das evidências, como abaixo:

$$x = \frac{Z_1 - Z_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}, \tag{2.2}$$

onde Z_1 e Z_2 representam medições independentes dentre os candidatos e $\sigma_1 + \sigma_2$ as variâncias de cada medição. Krotkov propôs um método de verificação, onde a técnica baseada em foco visa eliminar as falsas correspondências da técnica estéreo e de maneira semelhante a técnica estéreo auxilia na verifação da técnica baseada

em foco. Krotkov implementa um algoritmo de estéreo baseado em característica que não produz mapa de profundidade.

Gheta et. al. (2006) apresenta uma abordagem para realizar a reconstrução 3-D de cenas. As técnicas de estimação de profundidade utilizadas na fusão de dados são: SFS e SFF. A Fusão de Dados realizada nesse trabalho consiste na combinação de uma técnica de medição de foco, com a técnica estéreo de determinação dos pontos correspondentes. A combinação visa reduzir a determinação de pontos correspondentes incorretos utilizando uma rede de câmeras disposta em uma matriz 3x3, como ilustrado na Figura 2.1.

Figura 2.1: Sistema de câmeras combinado com um robô industruial. Conjunto de câmeras utilizado no processo de aquisição de imagens para reconstrução da estrutura tridimensional da cena.

A integração das informações é obtida através do funcional de energia com um corte de grafo mínimo. Porém, para que a abordagem apresente bom desempenho é necessário uma grande quantidade de câmeras, onde cada câmera apresenta posição focal fixa. Como consequência dessa configuração, temos a ocorrência de dois problemas: O número de posições focais é limitado pelo número de câmeras e as imagens são capturadas de diferentes pontos de visão. É importante mencionar a ausência de

uma métrica de avaliação dos resultados, sendo possível apenas contemplar visualmente o resultado obtido em uma única cena. O autor não formaliza nem caracteriza os conceitos de energia utilizados na pesquisa.

Em Rajagopalan et. al. (2004), foi proposta uma abordagem baseada no Campo Aleatório de Markov. Esse método Markoviano realiza a Fusão dos dados provenientes das técnicas baseadas na relação projetiva entre um par de imagens estéreo (SFS) e a variação de foco em uma sequência de imagens (SFF). As informações correspondentes à profundidade e suavização são integradas visando minimizar o funcional de energia através de *simulated annealing*, representando a informação estimada. Rajagopalan visa nesse trabalho a estimação de profundidade e restauração de imagens. Para isso são usadas imagens com variação de foco em parte da imagem, como meio para utilização das técnicas baseadas em foco e posteriormente a combinação com a técnica estéreo para a extração da profundidade, como pode ser contemplado na Figura 2.2. No trabalho de Schechner et. al. (1998), é realizado um estudo comparativo entre as técnicas de reconstrução da estrutura tridimensional de uma cena a partir da mudança de foco em um conjunto de imagens e da técnica baseada na relação entre pares de imagens apresentando as características principais de cada técnica. Segundo o autor as técnicas são similares, destacando como maiores diferenças o fato da técnica baseada em foco não tratar o problema da oclusão, que é resolvido com estéreo e a estrutura física entre os aparatos experimentais.

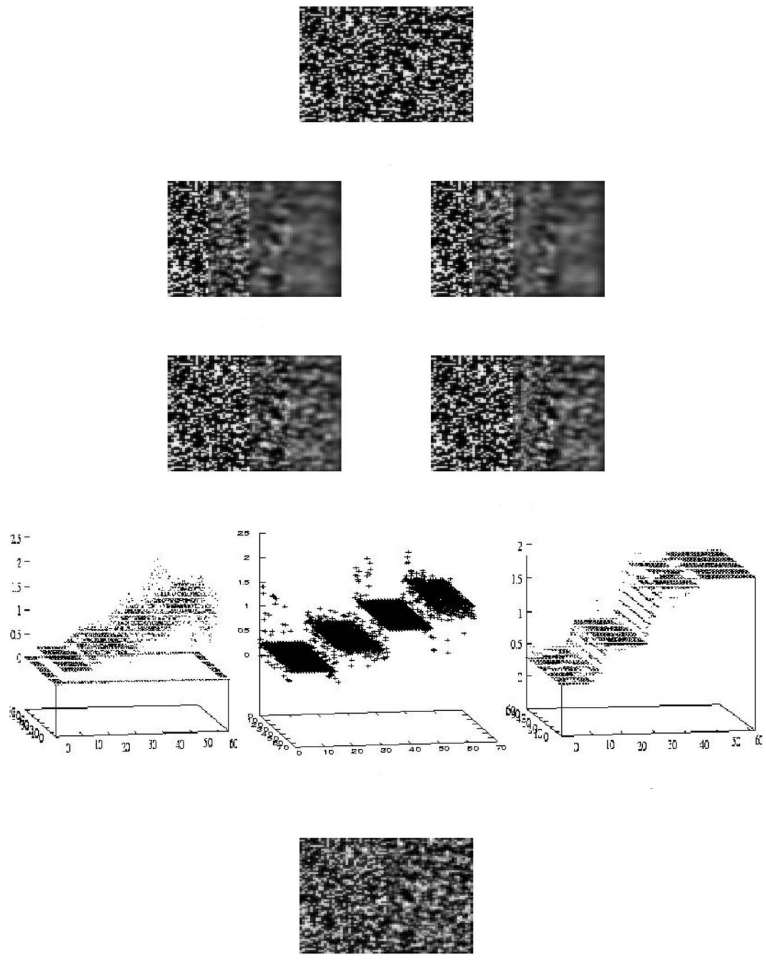


Figura 2.2: Processo de estimação de profundidade e restauração de imagens. Este procedimento utiliza imagens com variação de foco em parte das imagens possibilitando determinar a profundidade na cena e determinar a melhor condição focal na imagem.

Capítulo 3

Fusão de Dados de Energia Visual Para Estimação de Profundidade

Neste capítulo é desenvolvida uma abordagem de fusão de dados visuais para a obtenção de informações de profundidade de maneira mais acurada. Nas subseções seguintes serão apresentadas uma definição detalhada do problema e a descrição do procedimento metodológico adotado.

3.1 Definição do Problema

Considere que um sistema distribuído de sensores visuais $C = \{C_1, C_2, ..., C_N\}$ a partir do qual duas câmeras C_R e $C_L \in C$ separadas por uma distância B possuem condições de sobreposição dos seus campos de visão (FOV) de determinadas regiões de interesse L, configurado em um par estéreo de câmeras.

Considere ainda que se obtém de cada câmera (C_R e C_L) uma sequência de imagens $I^L = \{I_1^L, I_2^L \dots I_n^L\}$ e $I^R = \{I_1^R, I_2^R \dots I_n^R\}$ com variação de foco previamente definida, respectivamente. Onde cada variação de foco corresponde a uma medida de energia da cena contida em uma vizinhança definida no plano da imagem. Para

cada conjunto de imagens I^L e I^R é produzido um mapa de profundidade F^L e F^R levando em conta a informação de profundidade contida na mudança do foco presente nos conjuntos de imagens.

Em seguida escolhe-se convenientemente um par de imagens I_k^R e I_k^L para um valor k fixo. Esse par de imagens fornece uma medida de energia provida pela disparidade entre os pixels das duas imagens. O processo de fusão encontra as relações entre as medidas de energia do foco e do par de imagens. O problema consiste em definir o processo de fusão das informações de energia de foco e da disparidade entre pixels. Na seção seguinte as etapas realizadas neste trabalho são detalhadas.

3.2 Metodologia

A idéia principal dessa metodologia é combinar as medições de energia do foco provenientes dos mapas F^L e F^R e as medições de energia provenientes do par de imagens I_k^R e I_k^L.

Esta combinação ocorre através da energia envolvida no processo de captura de imagens e na formação de mapas a partir do foco. A fusão de dados é realizada por meio da Minimização das energias de foco e de disparidade, resultando em um mapa de profundidade M mais acurado da cena.

Em Visão Computacional e Processamento Digital de Imagens o conceito de energia é definido como uma medida de distância pseudo-métrica entre pixels. Trucco e Verri (1993), por exemplo, consideram que, no caso discreto, a energia de uma imagem pode ser obtida por meio das diferenças absolutas das intesidades de pixels.

$$E = ||p_i - p_j||, \tag{3.1}$$

onde p_i e p_j são pontos contidos nas imagens primária (esquerda) e secundária (direita), respectivamente. Por analogia aos sistemas físicos o processo de fusão compreende a fusão de pequenos núcleos de forma a construir um núcleo maior. A estratégia proposta é dividida em três etapas, que podem ser contempladas na Figura 3.1.

As etapas que compõem esta metodologia são:

1. Aquisição de Imagens;
2. Estimação da Energia de Foco;
3. Fusão de Dados por Minimização de Energia.

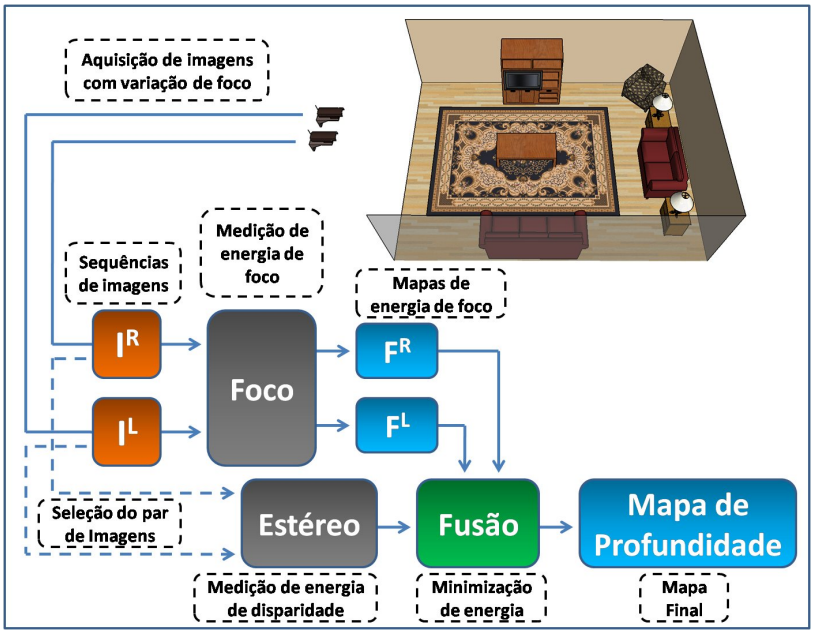


Figura 3.1: Duas câmeras do sistema de monitoramento que são convenientemente posicionadas obtém duas sequências de imagens da cena. Destas sequências são obtidos dados de profundidade pela variação de foco. Na determinação das correspondências a relação de energia envolvida no processo de aquisição de imagens ponderada pelas distâncias obtidas a partir do foco é usada em um modelo de fusão baseado em Minimização de energia.

A Figura 3.1 representa as etapas do processo metodológico. Inicialmente dois conjuntos de imagens são capturados por uma rede de câmeras configuradas em

par estéreo. Em seguida são computados dois mapas de profundidade através das variações de foco dentre as sequências de imagens. Na etapa seguinte um par de imagens é convenientemente escolhido, e a partir deste par é realizada a determinação das correspondências, que envolvem a relação de energia projetiva do par de imagens e as distâncias de foco previamente calculadas. A combinação das informações é realizada minimizando as relações de energia entre as informações de foco e de disparidade.

Para a melhor compreensão do processo metodológico, cada etapa será detalhada a seguir.

3.2.1 Aquisição de Imagens

O procedimento de aquisição de imagens é realizado por uma rede de câmeras $C = \{C_1, C_2, ..., C_N\}$ de modo a considerar para a captura duas câmeras C_R e $C_L \in C$ que apresentam configuração estéreo. Cada dispositivo de aquisição de imagens é uma câmera CCD, que consiste em uma matriz de células semicondutoras fotossensíveis, que atuam como capacitores, armazenando carga elétrica proporcional à energia luminosa incidente (Pedrini and Schwartz, 2007). As imagens são adquiridas de uma mesma cena, porém com características distintas que atendem a formação de dois mapas de profundidade iniciais (F^R e F^L) e uma mapa de profundidade final (M). Os dois primeiros mapas são gerados a partir da variação de foco, ou seja, a partir da variação do parâmetro intrínseco da câmera, distância focal, como definida na geometria de formação de imagem para o procedimento de desfocagem pela fórmula das lentes,

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v}, \tag{3.2}$$

onde f é a distância focal, u é a distância entre o objeto e o plano da lente e v é a distância entre a imagem focada e o plano da lente.

As sequências de imagens obtidas na captura baseada em foco são alcançadas através de sucessivas variações de f, como na série $\{f_1, f_2 \dots f_n\}$. Dessa forma é possível com base na lei da lentes delgadas adquirir duas sequências de imagens $I^R = \{I_1^R, I_2^R \dots I_n^R\}$ e $I^L = \{I_1^L, I_2^L \dots I_n^L\}$, uma para a câmera esquerda e outra para a câmera direita, por meio de sucessivas alterações do foco da imagem.

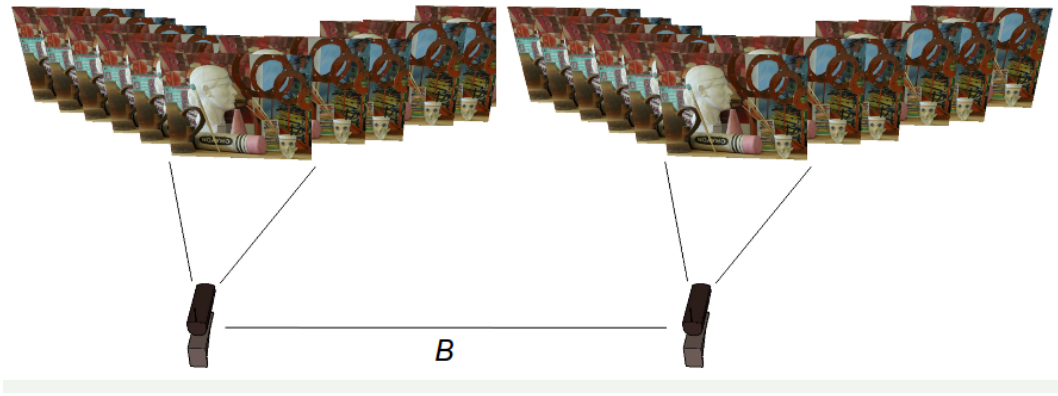


Figura 3.2: Sistema distribuído de câmeras configurado como uma par estéreo para a aquisição de duas sequências de imagens com variação de foco.

A Figura 3.2 apresenta uma rede distribuída de dispositivos visuais no formato de um sistema estéreo. As câmeras são separadas por uma distância B, cabendo a cada uma capturar uma sequência de imagens com variação de foco.

As câmeras que compõem a rede devem atender as restrições de um sistema de visão estéreo. Então duas imagens dentre as sequências de imagens (I^R e I^L), uma proveniente da câmera esquerda e outra proveniente da câmera direita, são selecionadas com a finalidade de determinar a estrutura tridimensional da cena. As imagens devem apresentar as mesmas variações de foco, tais como: I_K^R e I_K^L considerando um valor fixo de k que representa o nível de variação de foco. Essas imagens selecionadas são tidas como o par de imagens envolvido no processo de fusão.

A técnica de reconstrução da forma da cena a partir do foco será detalhada na seção

seguinte.

3.2.2 Estimação da Energia de Foco

A estimação de profundidade a partir do foco consiste na medição do espalhamento da energia luminosa que incide sobre as lentes a cada variação de foco. A técnica de determinação da estrutura 3-D da cena a partir do foco recupera a informação geométrica contida na cena a partir de uma sequência de imagens $(I_1, I_1 \ldots I_N)$, com variação da distância focal (f), monitoradas por uma câmera (C), como ilustrado na Figura 3.4.

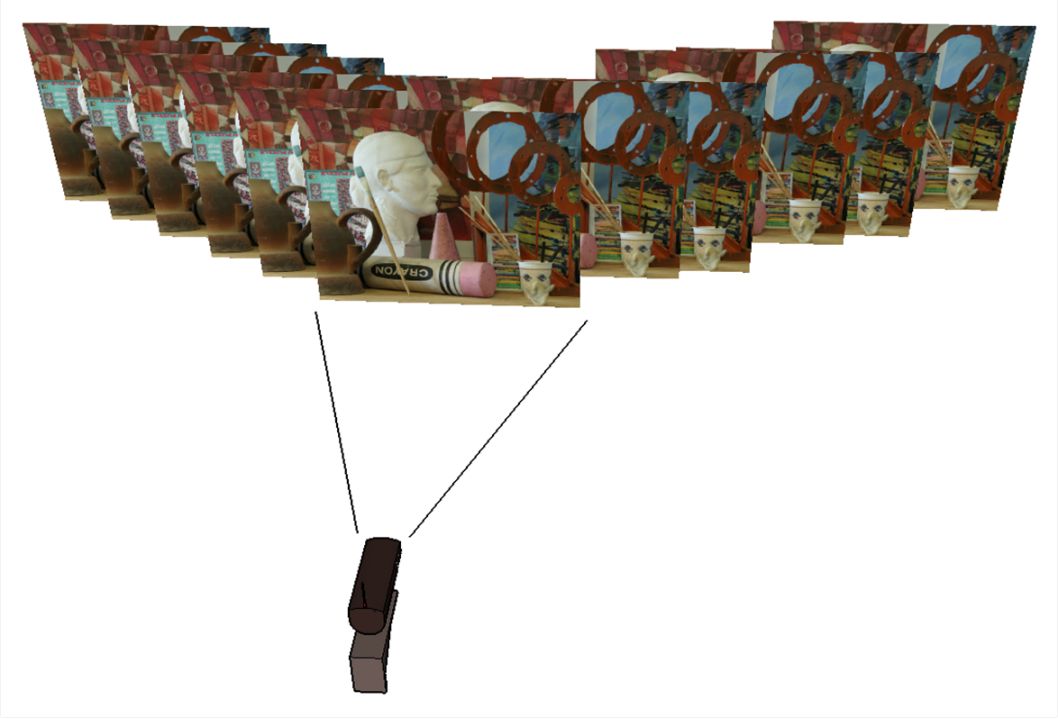


Figura 3.3: Processo de captura de imagens com variação de foco. A partir de mudanças na distância focal da câmera é possível obter um conjunto de imagens com diferentes situações focais.

Inicialmente, na variação de foco cada ponto do plano do objeto é projetado no plano da imagem, formando imagens em foco ou focadas. Entretanto, o plano do sensor (CCD) pode não coincidir com o plano da imagem e por isso é deslocado

uma distância δ, resultando em imagens fora de foco. A energia luminosa que incide sobre as lentes é uniformemente distribuída em um fragmento circular no plano do sensor. Dessa forma, variar o foco das imagens representa variar a distância entre o plano da imagem e o plano do sensor, apresentando comportamento gaussiano.

O espalhamento da energia luminosa é conduzido como na Figura 3.4,

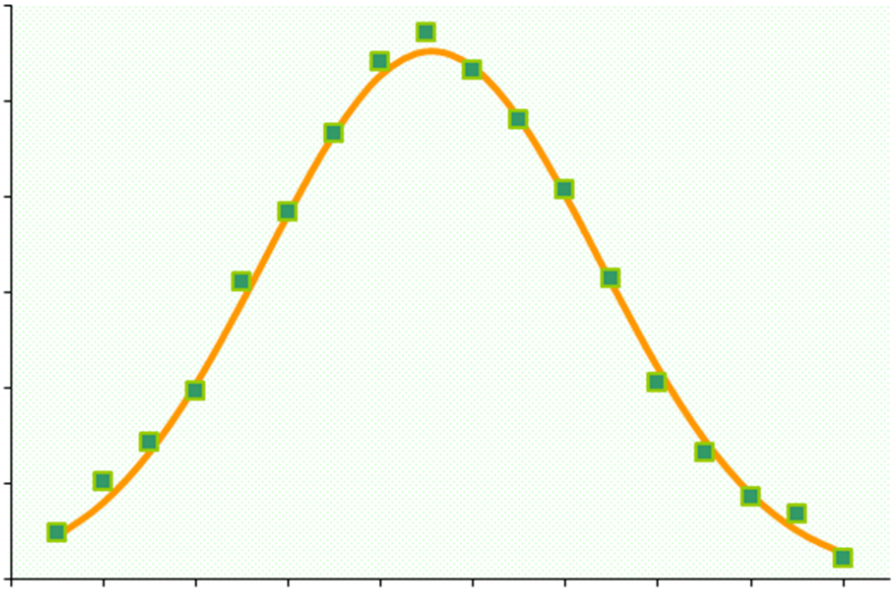


Figura 3.4: Espalhamento gaussiano da energia luminosa para variação de foco da sequência de imagens.

onde a medida que a distância entre os planos aumenta o raio de desfoque aumenta. Em seguida, o método baseado em foco calcula, dentre a série de imagens, os pontos que apresentam melhor foco, por meio do método denominado medição de foco. Na abordagem proposta é aplicado o método de medição de foco Soma do Laplaciano Modificado (SML), pois SML alcança bons resultados quando comparado com medições alternativas (Nayar and Nakagawa, 1994).

A técnica SML calcula inicialmente o Laplaciano Modificado, que refere-se à coordenada (x, y) por meio de uma janela local definida pela variável *step*, como pode

ser visto na Equação 3.3.

$$ML(x,y) = |2I(x,y) - I(x-step,y) - I(x+step,y)| + |2I(x,y) - I(x,y-step) - I(x,y+step)|. \tag{3.3}$$

Então, a medição de foco SML,

$$F(i,j) = \sum_{x=i-N}^{i+N} \sum_{y=j-N}^{j+N} ML(x,y) \text{ para } ML(x,y) \geq T, \tag{3.4}$$

é computada para cada pixel $ML(x,y)$ usando uma janela em volta da coordenada, sendo o tamanho determinado por N e T um limiar que viabiliza a escolha do pixel em foco, como ilustrado na Equação 3.4.

Com a informação de foco previamente calculada para cada imagem do conjunto é possível inferir a energia do ambiente monitorado, por meio da estratégia de estimação de energia baseada na Interpolação Gaussiana, como é possível visualizar na Equação 3.5. A interpolação gaussiana seleciona as três melhores medições de foco (F_m: melhor, F_{m-1}: segunda melhor e F_{m+1}: terceira melhor) para calcular a energia e a posição na sequência de imagens onde se encontram tais medições (d_m, d_{m-1} e d_{m+1}), a inferência de energia pode ser contemplada a seguir:

$$E_f = \frac{(lnF_m - lnF_{m+1}).(d_m^2 - d_{m+1}^2) - (lnF_m - lnF_{m-1}).(d_m^2 - d_{m-1}^2)}{2\delta\{(lnF_m - lnF_{m-1}) + (lnF_m - lnF_{m+1})\}} \tag{3.5}$$

onde F_m, F_{m+1} e F_{m-1} representam os melhores valores computados na medição de foco. Enquanto d_m, d_{m+1} e d_{m-1} consistem nos deslocamentos de cada medição, ou seja, em quais imagens do conjunto foram encontradas as melhores medições.

Em geral, medidas de energia em imagens estão relacionadas com distância entre o brilho de pixels. Por essa razão a Interpolação Gaussiana foi considerada para determinar energia de foco. O resultado final consiste em medições de energia de foco proporcionadas a partir das variações de foco nas imagens capturadas.

3.2.3 Fusão de Dados por Minimização de Energia

A fusão das informações é realizada por meio dos dados providos por uma rede distribuída de câmeras C em uma configuração estéreo C_R e C_L. Inicialmente duas sequências de imagens I^R e I^L com variação de foco são capturadas, uma pela câmera esquerda e outra pela câmera direita, para a realização das estimações de profundidade iniciais baseadas em foco. O resultado desta operação é a concepção de dois mapas de profundidade F^R e F^L. Os mapas de foco computados serão utilizados no processo de combinação, como medidas de energia auxiliando na decisão de pontos correspondentes mais precisos na estimação das disparidades. Durante a etapa de correspondência as medições de energia do foco são combinadas com a energia do par de imagens. Tais medições consistem na relação de energia envolvida no processo de aquisição de imagens. Na aquisição de imagens a energia luminosa incide sobre o *CCD*, que absorve esta energia armazenando uma carga elétrica proporcional a energia luminosa incidente. Para integrar os dados alcançados, uma abordagem baseada em minimização de energia foi adotada. A energia é definida como uma medida de distância pseudo-métrica entre pixels. Então minimizar a energia representa selecionar os pixels mais próximos, ou seja, pixels com maior relação de correspondência baseado nas energias.

Esta abordagem considera o uso das informações de energia computadas a partir do foco e da disparidade do par de imagens, como abaixo:

$$E(f) = argmin\{\alpha E_s(f) + \beta E_f(f)\}, \tag{3.6}$$

onde α e β são coeficiente de proporcionalidade que indicam o grau de atuação na estimação, $E_s(f)$ representa uma função que mede o nível de energia luminosa incidente sobre o CCD para um determinada disparidade f a partir do par de imagens. Enquanto $E_f(f)$ consiste na função de energia computada a partir da condição focal dentre o conjunto de imagens capturado, considerando os mapas de profundidade

F_K^L e F_K^R obtidos no passo de estimação anterior. O processo de minimização de energia consiste na seleção do rótulo de disparidade (f) que minimiza a função de energia E, ou seja, considera as energias encontradas anteriormente que minimizam a função de energia resultante.

Os termos que compõem a equação 3.6 correspondem a uma relação linear para aprimorar o problema de correspondência. Onde cada termo compreende medições de energia obtidas a partir das variações de foco e das relações projetivas do par de imagens.

A formulação para a energia luminosa incidente pode ser definida como sendo P um conjunto de pixels na imagem primária e $I_K^L = \{I_K^L(p)|p \in P\}$ e $I_K^R = \{I_K^R(p)|p \in P\}$ sendo as intensidades nas imagem primária e secundária, para um valor k fixo, respectivamente. Para cada ponto p é atribuído um rótulo que representa a disparidade entre pontos $f_p = \{f_p|p \in P\}$ na imagem primária. Cada rotulação significa uma correspondência entre um ponto p na imagem primária e um ponto $p + f_p$ na imagem secundária. A Minimização de Energia define a melhor configuração de disparidade f que minimiza a energia, ou seja, consiste em selecionar o melhor candidato a correspondente que apresente a energia mínima.

A função de energia para o processo de estimação de correspondências considera as intensidades resultantes da energia luminosa incidente, como pode ser definida a seguir:

$$E_s(f) = E_{d_s}(f) + E_{s_s}(f), \tag{3.7}$$

o termo E_{d_s} impõe penalizações para configurações que são inconsistentes considerando os dados observados I_K^L e I_K^R, ou seja penaliza casamentos inconsistentes. O termo E_{d_s} padrão é definido como:

$$E_{d_s}(f) = \sum_p D_p(f_p), \tag{3.8}$$

onde $D_p(f_p)$ é alguma medida de similaridade entre $I_K^L(p)$ e $I_K^R(p+f_p)$. O termo para

medida de similaridade baseado nos dados dos pixels do processo de correspondência, considera $D_p(f_p)$ como:

$$D_p(f_p) = (I_K^L(p) - I_K^R(p + f_p))^2, \tag{3.9}$$

onde é realizada uma medição de distância para inferir quão próximo o pixel da imagem primária ($I_K^L(p)$) encontra-se do pixel presente na imagem secundária ($I_K^R(p + f_p)$).

O segundo termo da função de energia mensura a suavização presente nas imagens. O termo examina a suavização dos pixels das imagens a partir da vizinhança de cada pixel, sendo possível estimar a taxa de discontinuidade na cena. O termo E_{s_s} é definido a seguir:

$$E_{s_s}(f) = \sum_{L,R} V_{L,R}(I_R^L, I_K^R), \tag{3.10}$$

onde $V_{L,R}(I_K^L, I_K^R)$ representa o custo dos pixels adjacentes às intensidades I_K^L e I_K^R. O termo $V_{L,R}(I_K^L, I_K^R)$ é calculado através da métrica de similaridade intervalar. Esta técnica encontra a similaridade entre pontos das imagens $I_k^L(x, y)$ e $I_k^R(x, y + d)$ nas sequências de imagens para um valor k fixo. A similaridade intevalar consiste em determinar pontos que pertençam a um determinado intervalo. Caso o ponto pertença ao intervalo, significa semelhança. O intervalo é determinado em função dos intervalos de confiança computados para a janela referência (Imagem esquerda) e para janela corrente (Imagem direita) (Bussab and Morettin, 2010). Para o cálculo do intervalo de confiança faz-se necessário calcular primeiramente o desvio padrão da amostra, como abaixo:

$$\sigma(x) = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}(x_{i,j} - \bar{x})^2}{n}}, \tag{3.11}$$

onde n é o tamanho da janela, $x_{i,j}$ é um ponto pertencente à janela, $\bar{x}$ representa a média dos elementos da janela.. Para computar o intervalo de confiança é preciso definir o nível de significância α, definido internacionalmente 0,05 (Bussab and

Morettin, 2010), que representa a probabilidade do valor procurado está fora do intervalo. O intervalo de confiança é definido abaixo:

$$(x - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, x + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}), \tag{3.12}$$

onde x representa o ponto central da janela, $z_{\frac{\alpha}{2}}$ é o valor de z da distribuição normal padrão (Bussab and Morettin, 2010), σ o desvio padrão da amostra e n o tamanho da amostra. O valor da variável $z_{\frac{\alpha}{2}}$ é determinado a partir de:

$$z = 0,5 - \frac{\alpha}{2}, \tag{3.13}$$

onde α representa o nível de significância. Para encontrar o valor de $z_{\frac{\alpha}{2}}$, é necessário buscar na tabela de distribuição normal o valor correspondente.

O precesso de correspondência consiste em dado um ponto (x, y) na imagem I_L, encontrar o correspondente na imagem I_R, examinando uma determinada quantidade de candidatos para determinar o vencedor. Na abordagem proposta é computado o intervalo de confiança da vizinhança da coordenada (x, y) para a imagem esquerda (I_L) e para os candidatos na imagem direita (I_R). O intervalo da métrica de similaridade proposta para determinação da correspondência estéreo é definido pela diferença entre os intervalos de confiança do ponto da imagem esqueda e o candidato. Pois dada uma vizinhança V_1 e uma vizinhança V_2, caso as mesmas sejam semelhantes seus intervalos de confiança serão valores próximos. Tendo em vista que valores semelhantes geram intervalos de confiança semelhantes considera-se a diferença dentre os intervalos de confiança computados como sendo o intervalo da medida de similaridade, como mostrado abaixo:

$$int = abs((z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}})_{im_{esq}} - (z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}})_{candidato}), \tag{3.14}$$

onde int corresponde ao intervalo utilizado pela medida de similaridade intervalar, $(z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}})_{im_{esq}}$ e $(z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}})_{candidato}$ representam os intervalos de confiança das vizinhanças contidas na imagem esquerda e na imagem direita (candidato), respectivamente.

Com o intervalo definido é possível verificar o pontos que são semelhantes e assim mensurar a proximidade dentre o pixel da imagem esquerda e os candidatos a correspondente.

A minimização de energia considera as somas das energias computadas a partir do foco e da disparidade dos pixels obtendo um conjunto de medições $E(f) = \{e_1, e_2, e_3, ..., e_n\}$. A medida usada como resultado da função será o $min\{E(f)\}$. Na próxima seção é mostrado experimentalmente que os resultados obtidos usando a Minimização de Energia como método de fusão obtém melhores resultados comparado com estratégias clássicas.

Capítulo 4

Resultados Experimentais

Neste capítulo são apresentadas as configurações utilizadas para os experimentos e os resultados obtidos pela técnica de minimização de energia visual de foco e da disparidade entre pixels proposta. O capítulo também contempla métodos de avaliação dos resultados obtidos visando validar a estratégia desenvolvida.

4.1 Descrição dos Experimentos e do Aparato Experimental

O conjunto de experimentos foi realizado utilizando o software matemático *Matlab 7.0* com ferramentas para processamento de imagens e executados em um Computador *Pentium Dual-Core* 2.18 GHz com 1 GB de memória RAM. O software Photoshop CS2 também foi utilizado, para simular as variações de foco através do filtro de lentes e para variação de iluminação através de mudanças na taxa de brilho nas imagens.

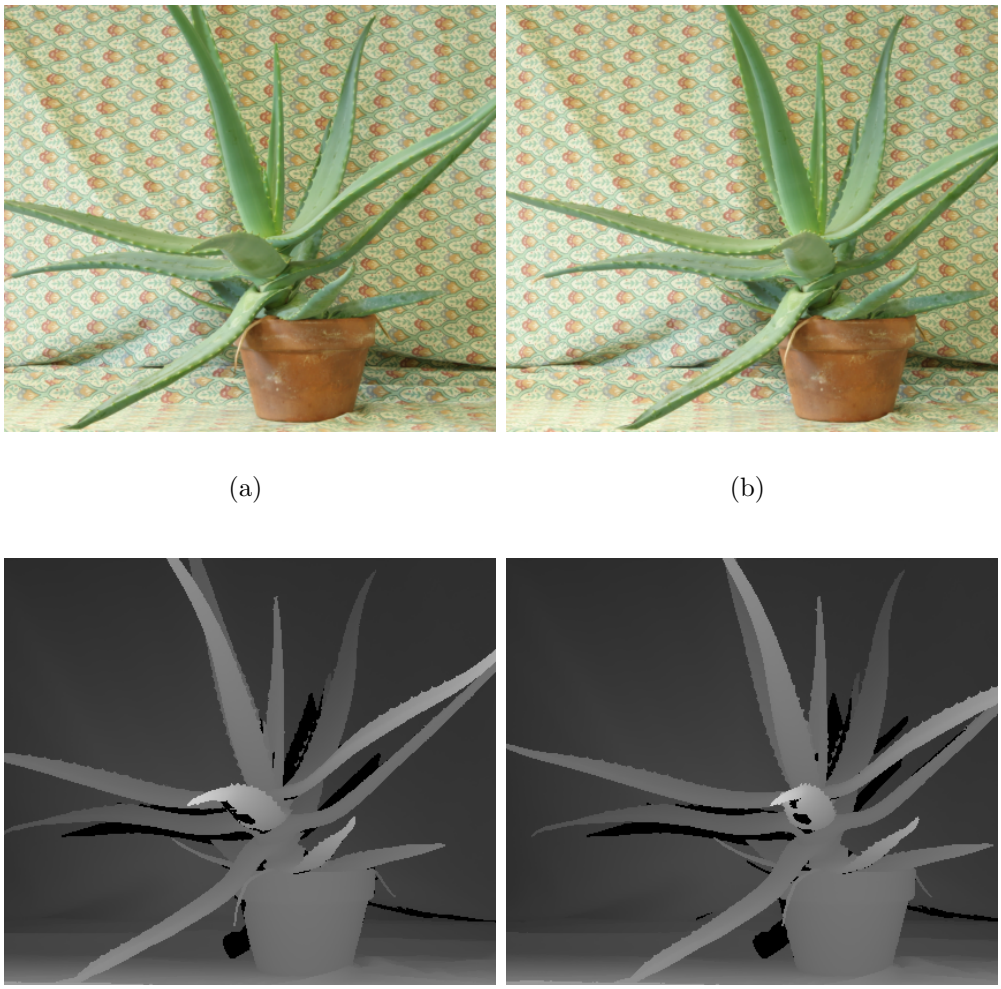
A base de imagens adotada para os experimentos consiste em um conjunto de imagens consolidado na literatura. Tal base de imagens é composta por 25 pares estéreo de imagens capturados a partir de diferentes ambientes reais. A base também contém

os mapas de profundidade reais de cada imagem do par estéreo, como ilustrado na Figura 4.1 (Scharstein and Szeliski, 2001). As Figuras 4.1 (a) e (b) representam o par estéreo de imagens capturado por um sistema de visão. Enquanto que as Figuras 4.1 (c) e (d) ilustram os mapas de profundidade referentes a cada imagem do par.
Um procedimento artificial foi realizado para alterar a distância focal de cada imagem do par estéreo. O resultado desta operação consiste na obtenção de duas sequências de imagens com variação de foco controlada por *software*, como podemos visualizar na Figura 4.2 abaixo.
Na Figura 4.2 são apresentadas imagens com variação de foco realizada por software. Nas imagens é possível observar objetos com melhor condição focal em uma imagem e posteriormente desfocadas. Esse efeito é executado utilizando o filtro de lentes do *Photoshop*, controlando a distância focal e dessa maneira alterando a posição do foco na imagem (Incorporated, 2009).
Os experimentos realizados para validação dos resultados obtidos são:

1. Experimento 1 - Avaliação comparativa entre a abordagem proposta e as técnicas clássicas estéreo;

2. Experimento 2 - Validação da robustez do método frente a variação de luminosidade;

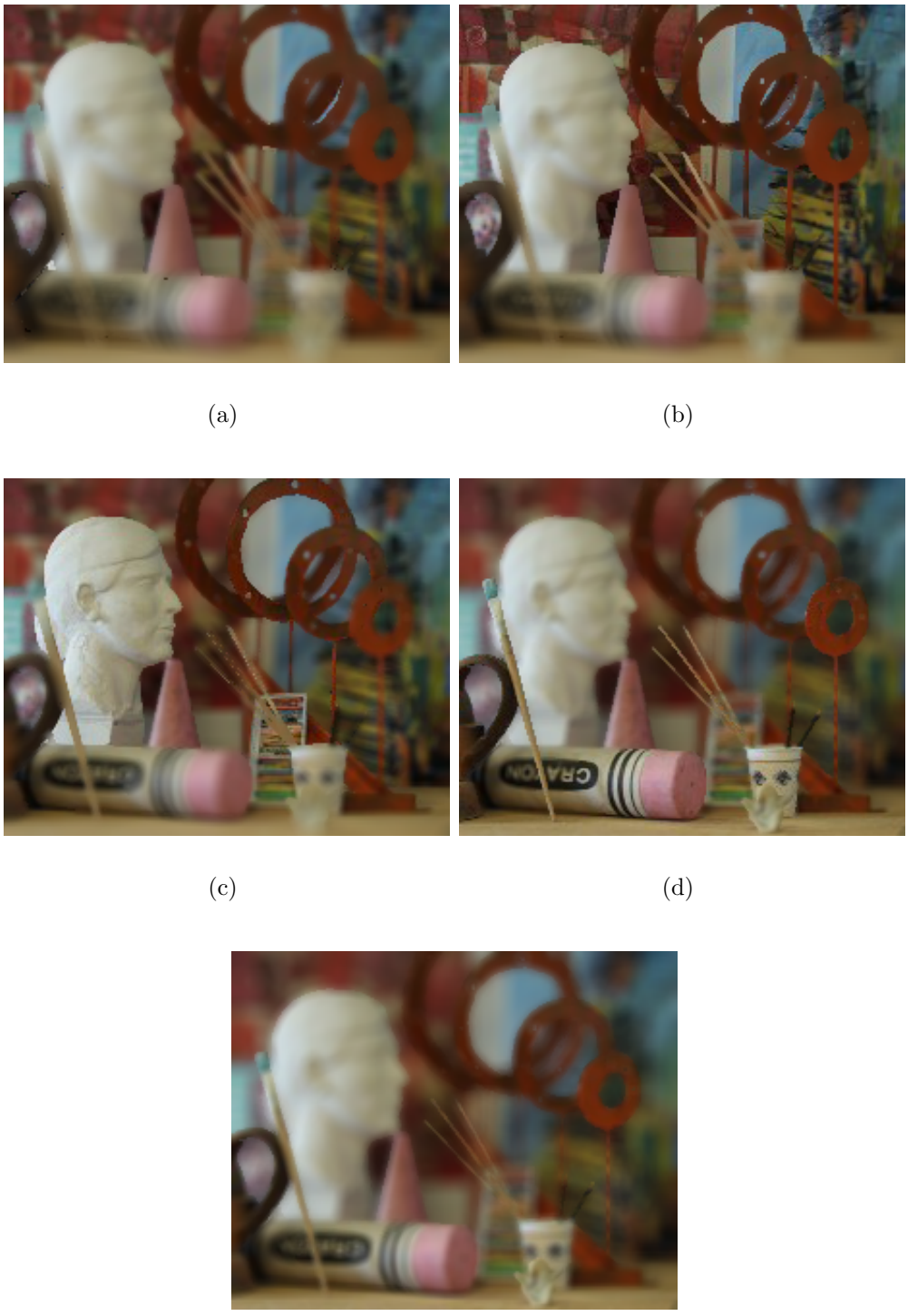
O Experimento 1 compara os resultados obtidos pelas técnicas de estimação de profundidade tradicionais estéreo com os resultados obtidos pela abordagem desenvolvida. Para o estéreo tradicional foram consideradas as consolidadas técnicas de determinação de correspondência SSD, SAD e NCC. Para avaliar o desempenho das técnicas em comparação foram utilizadas duas medidas para inferir quão próximo do resultado correto foi alcançado.
As medidas RMSE (Root Mean Squared Error) (Pedrini and Schwartz, 2007) e BMP (Bad Matching Pixel) (Scharstein and Szeliski, 2001) foram utilizadas para mensu-



(a) (b)

(c) (d)

Figura 4.1: Modelo de imagens disponíveis na base. Nas Figuras (a) e (b) são apresentadas as imagens capturadas pelas câmeras esquerda e direita, respectivamente. Enquanto que nas Figuras (c) e (d) são mostrados os mapas de profundidade referente às imagens da esquerda e direita, respectivamente.



(a) (b)

(c) (d)

(e)

Figura 4.2: Exemplo de um conjunto de imagens com variação do parâmetro intrínseco, distância focal. A série de imagens apresenta informação de profundidade a medida que o foco varia controladamente pelas imagens.

rar os mapas de profundidade alcançados e inferir a qualidade da reconstrução. O RMSE é calculado como a abaixo:

$$RMSE = \sqrt{\frac{1}{N}\sum_{(x,y)} |d_c(x,y) - d_t(x,y)|^2}, \tag{4.1}$$

onde N é o número total de pixels, d_c representa o mapa de disparidade computado e d_t representa o mapa verdadeiro.

Enquanto o BMP procede da seguinte forma:

$$BMP = \frac{1}{N}\sum_{(x,y)} (|d_c(x,y) - d_t(x,y)| > S_d), \tag{4.2}$$

onde N é o número total de pixels, d_c representa o mapa de disparidade computado, d_t representa o mapa verdadeiro e S_d significa uma constante de tolerância a erros, $S_d = 1$ como adotado no estudo publicado (Scharstein and Szeliski, 2001).

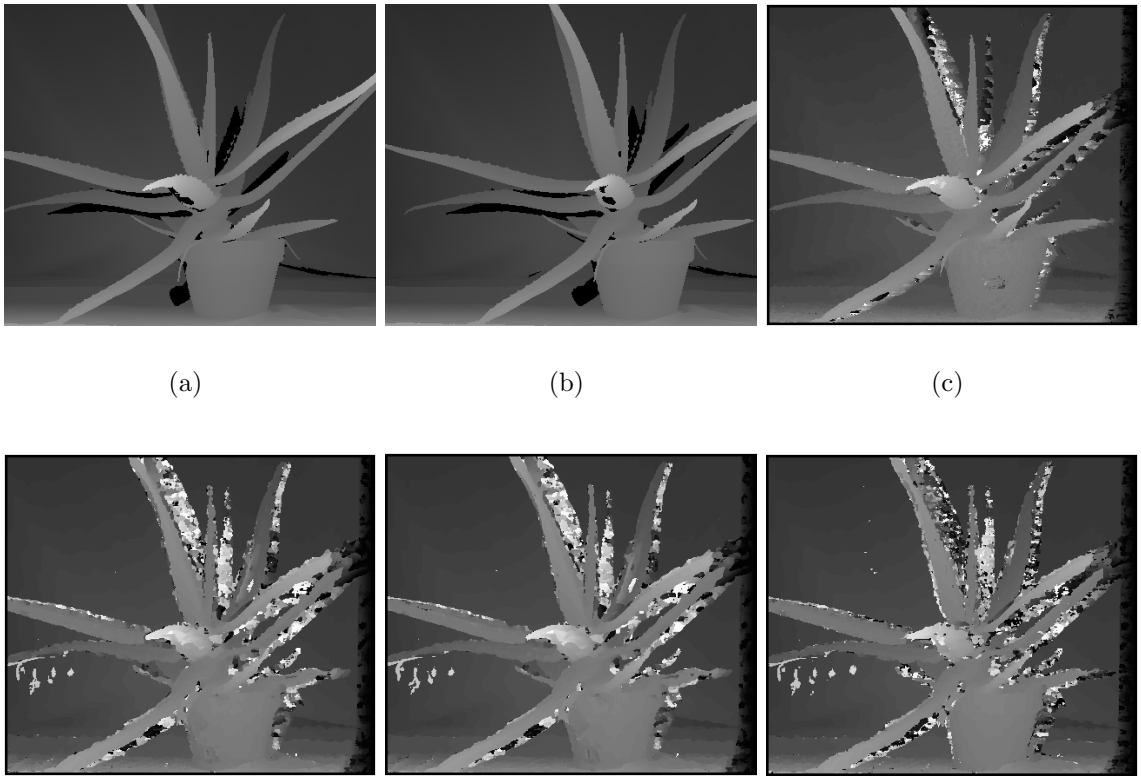
Dessa maneira é possível inferir quantitativamente a qualidade das reconstruções possibilitando a comparação dos resultados, de modo visual e por meio de gráficos. O Experimento 2 avalia a robustez da abordagem proposta através da variação de luminosidade das imagens. Os testes comparam a abordagem proposta com as técnicas tradicionais (SSD, SAD e NCC) frente às mesmas variações de luminosidade. Para a comparação são usadas as medidas RMSE e BMP, resultando em medições de proximidade da informação real da cena monitorada. Os resultados são apresentados para análise visual e por meio de gráficos

4.2 Apresentação dos Resultados

Nesta etapa os resultados obtidos nos processos de experimentação são expostos para validação da abordagem proposta. Os resultados são apresentados de maneira visual para análise dos mapas gerados pelas técnicas clássicas e pela estratégia proposta. Os resultados também são expostos através de medidas quantitativas fazendo uso de medidas de similaridade para mensurar a qualidade da profundidade estimada. Portanto gráficos evidenciam o desempenho alcaçado pela estratégia desenvolvida.

4.2.1 Experimento 1 - Avaliação comparativa entre a abordagem proposta e as técnicas clássicas estéreo

Neste experimento foram utilizados 25 conjuntos de imagens, dos quais podemos observar alguns resultados nas Figuras 4.3, 4.4 e 4.5:

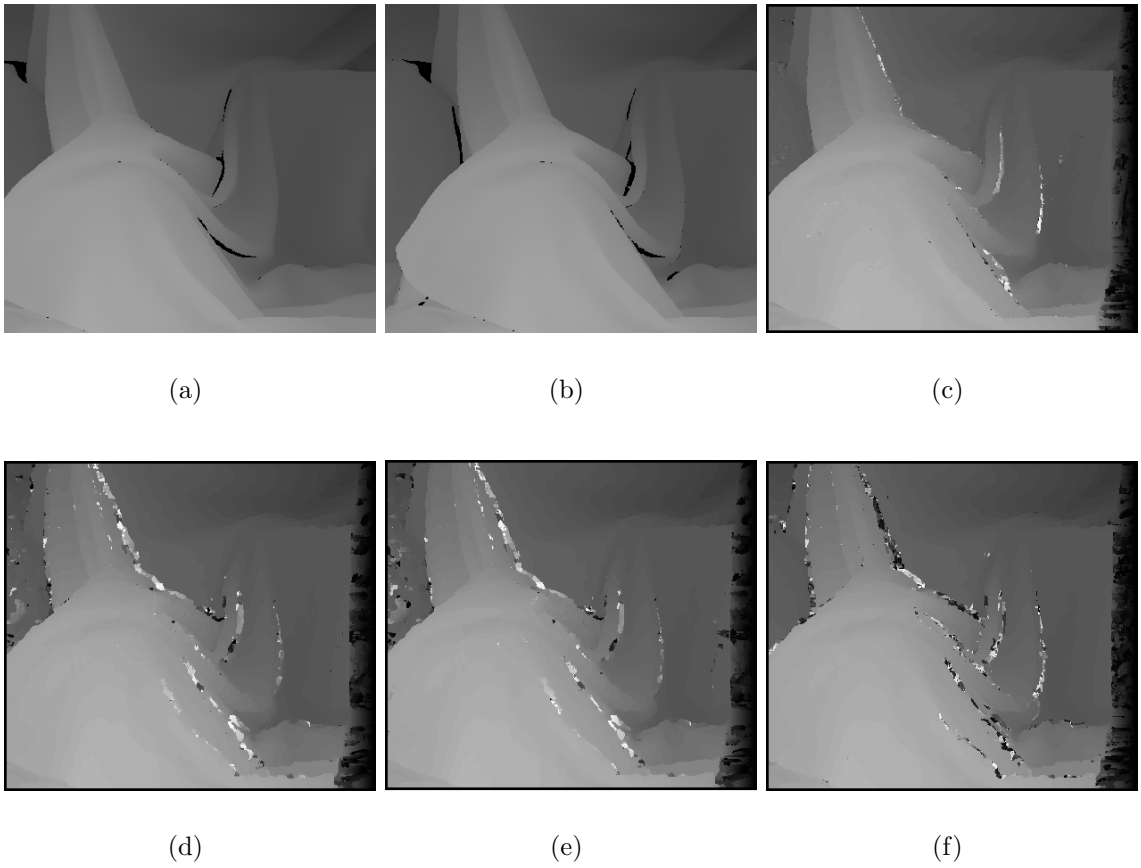


(a) (b) (c)

(d) (e) (f)

Figura 4.3: Resultados obtidos na etapa de experimentação. As imagens (a) e (b) representam respectivamente os mapas de profundidade reais obtidos a partir das imagens capturadas pelas câmeras da esquerda e direita. Na imagem (c) é apresentado o mapa de profundidade obtido a partir da técnica proposta. Enquanto que as imagens (d), (e) e (f) consistem nos resultados obtidos através das técnicas de estéreo clássicas, SSD, SAD e NCC, respectivamente.

Na Figura 4.3 é possível observar a melhor definição do plano de fundo da cena assim como do vaso e folhas da planta sensoriada. É válido destacar a redução dos ruídos na obtenção da profundidade do ambiente monitorado.

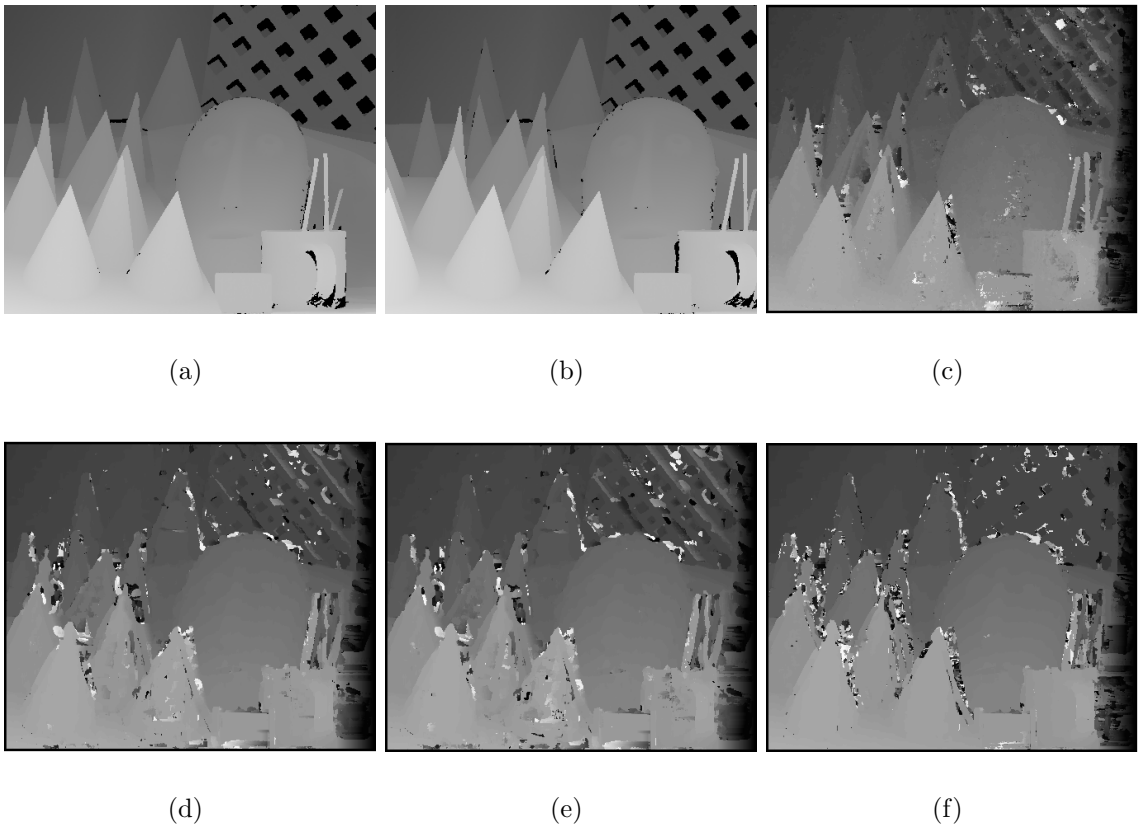


(a) (b) (c)

(d) (e) (f)

Figura 4.4: Resultados obtidos na etapa de experimentação. As imagens (a) e (b) representam respectivamente os mapas de profundidade com informações reais de profundidade obtidos a partir das imagens capturadas pelas câmeras da esquerda e direita. Na imagem (c) é apresentado o mapa de profundidade obtido a partir da técnica minimização de energia. Enquanto que as imagens (d), (e) e (f) consistem nos mapas obtidos através do SSD, SAD e NCC, respectivamente.

Na Figura 4.4 é ilustrado o ganho qualitativo na reconstrução da estrutura tridimensional da cena. Destacando a melhor estimação de distância para o ambiente com superfícies com alto grau de textura, juntamente com a eliminação de boa parte do ruído associado ao processo de determinação da forma da cena.



(a) (b) (c)

(d) (e) (f)

Figura 4.5: Resultados obtidos a partir do processo de fusão proposto. As imagens (a) e (b) consistem nos mapas de profundidade reais capturados pelas câmeras da esquerda e direita. A imagens (c) representa o mapa de profundidade resultante do processo de fusão por minimização de energia. Já nas imagens (d), (e) e (f) são mostrados os mapas providos pelas técnicas estéreo SSD, SAD e NCC.

Na Figura 4.5 é possível visualizar a melhoria obtida na estimação da forma da cena em ambientes compostos por superfícies com transições abruptas. Reduzindo significativamente a taxa de falhas no reconstrução 3-D do ambiente. Nas Figuras 4.3, 4.4 e 4.5 são exibidos os mapa de profundidade disponíveis na base de imagens adotada (Scharstein and Szeliski, 2001). O mapa alcançado pela técnica de minimização de energia proposta e os mapas obtidos por algumas técnicas consolidadas na literatura considerando o problema de correspondência estéreo, que são detalhados no Apêndice A.

Analisando as imagens resultantes podemos visualmente constatar a acurácia da

estratégia baseada em Fusão de Dados de Energia Visual desenvolvida e a redução de falhas no procedimento de estimação de profundidade.

Para validar os resultados alcançados é necessário uma medida que possa assegurar as melhorias atingidas, ao invés de comparar apenas visualmente os resultados como é largamente encontrado na literatura (V. Michael Bove, 1990) (Rajagopalan et al., 2004) (Frese and Gheta, 2006) (Gheta et al., 2007).

A avaliação dos resultados obtidos pela técnica de combinação proposta faz uso dos mapas de profundidade reais disponíveis na base. Os mapas de profundidade computados na abordagem proposta e nas técnicas tradicionais SSD, SAD e NCC são avaliados pelas medidas de similaridade RMSE e BMP para determinar quem mais aproxima-se da reconstrução real. As medições são conduzidas por meio das técnicas de similaridade (RMSE e BMP), que analisam a relação de proximidade envolvendo o mapa real da imagem esquerda e os mapas a serem avaliados.

Contemplando os gráficos 4.6 e 4.7 é possível visualizar o comportamento qualitativo das técnicas avaliadas, onde quanto menor a barra mais próxima da realidade foi alcançado. Ainda considerando os gráficos é possível observar com clareza a qualidade na estimação da estrutura tridimensional da cena realizada pela técnica de miniminação de energia proposta.

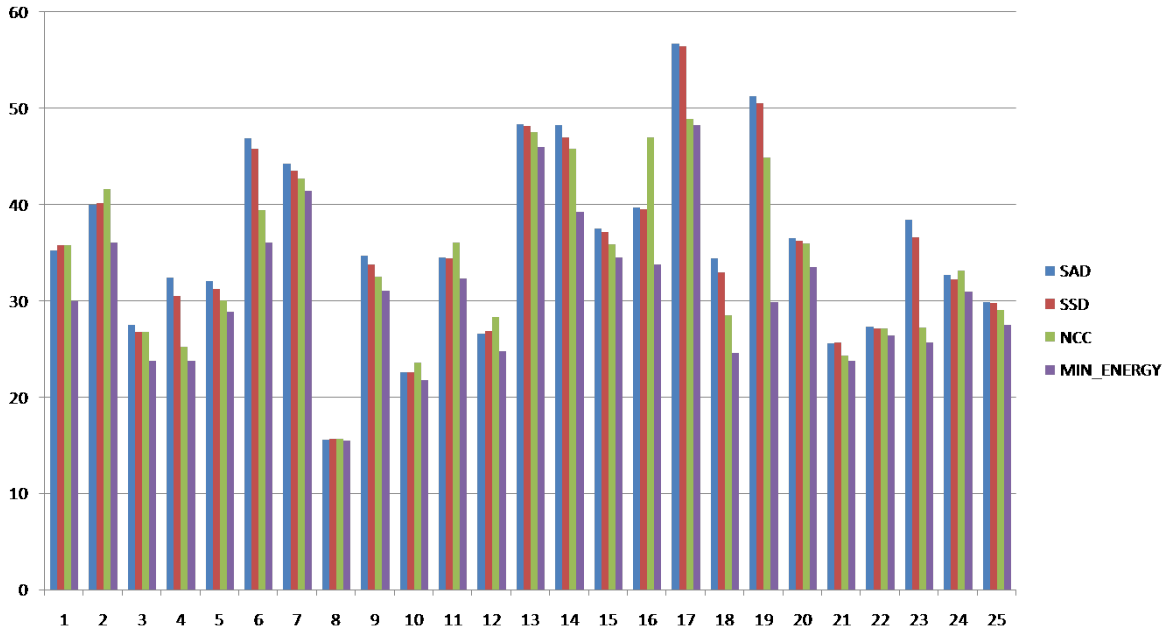


Figura 4.6: Comparação dos resultados usando RMSE. Gráfico comparando técnicas clássicas do estéreo e a estimação baseada em Funções de Energia com o mapa de profunidade real da imagem esquerda.

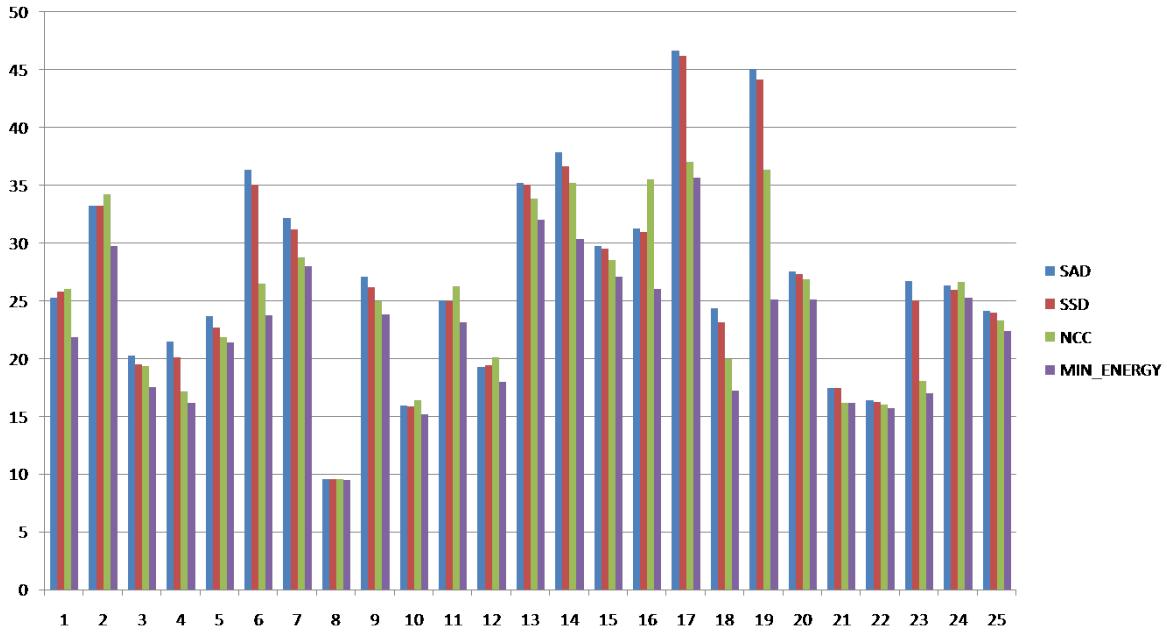


Figura 4.7: Comparação dos resultados usando BMP. Gráfico comparando técnicas clássicas do estéreo e a estimação baseada em Funções de Energia com o mapa de profunidade real da imagem esquerda.

4.2.2 Experimento 2 - Validação da robustez do método frente a variação de luminosidade

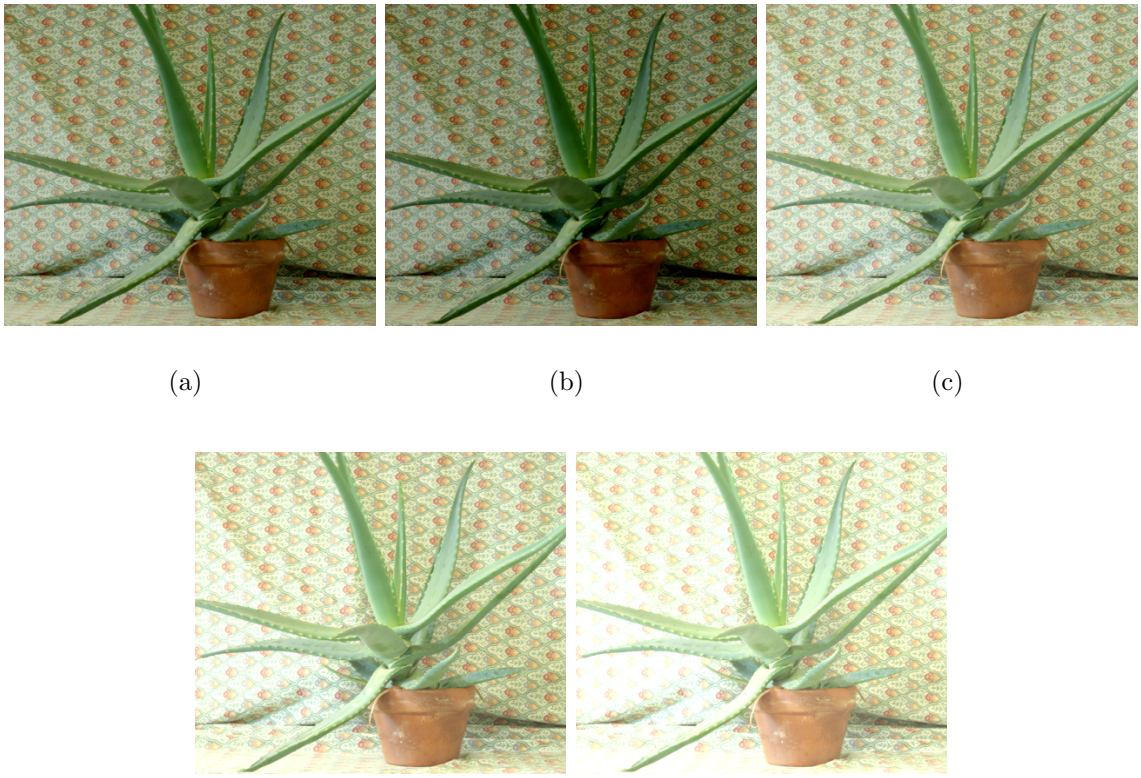
Para avaliar a robustez da abordagem desenvolvida foi conduzida uma etapa de experimentação, que consiste em variar o nível de iluminação da cena monitorada. Tal variação de iluminação visa avaliar os impactos ocasionados pela diminuição do nível de energia luminosa na cena, tornando a imagem mais escura, e pelo aumento do nível de luz na cena, tornando a imagem mais clara. Sendo possível determinar qual estratégia de estimação de profundidade resulta em mapas de profundidade mais confiáveis. Na Figura 4.8 são apresentadas algumas imagens com variação de iluminação.

A Figura 4.8 consiste nas mudanças de luminosidade na cena monitorada. Tais mudanças de luminosidade foram promovidas por um processo artificial controlado por *software* de variação do brilho da imagem, realizado no *Photoshop* (Incorporated, 2009).

Para esta etapa de avaliação dos impactos da iluminação serão consideradas apenas as extremidades de luminosidade, ou seja, as imagens com maior luminosidade e com menor luminosidade geradas. Este procedimento foi realizado para os mesmos 25 conjuntos de imagens.

4.2.2.1 Variação de iluminação com baixa luminosidade

Inicialmente serão realizadas medições em imagens com brilho diminuído, onde em cada pixel da imagem é subtraído 25 de seu valor de intesidade original. As Figuras 4.9, 4.10 e 4.11 contemplam os resultados obtidos no processo determinação das distâncias da cena.



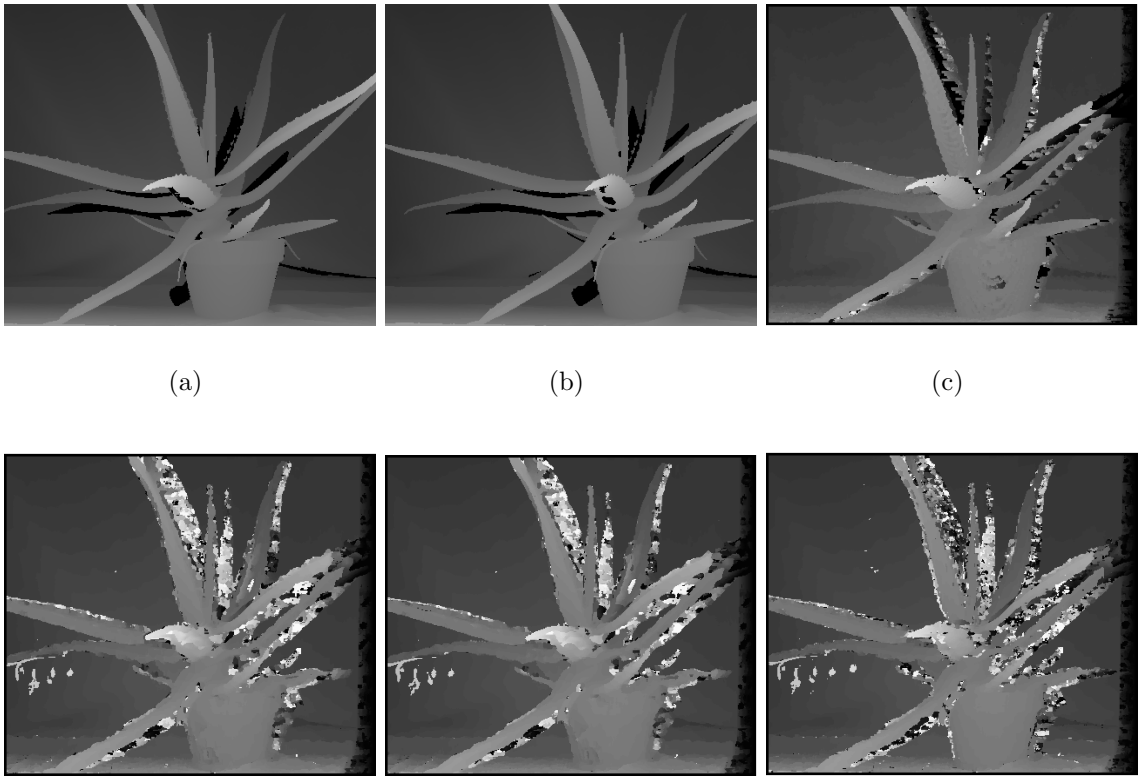
(a) (b) (c)

(d) (e)

Figura 4.8: Modelo de imagens com variação de iluminação. As imagems mostradas acima apresentam mudanças na luminosidade da cena de modo a permitir uma análise mais aprofundada dos impactos da iluminação na estimação de profundidade. Na imagem (a) de cada pixel foi subtraído 15, na imagem (b) foi subtraído 25, a imagem (c) apresenta brilho normal, na imagem (d) foi adicionado 15 e na imagem (e) foi adicionado 25.

As Figuras 4.9, 4.10 e 4.11 representam os mapas de profundidade obtidos pela estratégia de miniminação de energia proposta e pelas técnicas tradicionais de estimação de profundidade. Nas imagens ainda é possível constatar melhoria nas informações de profundidade alcançadas pela abordagem baseada em Função de Energia em detrimento das demais técnicas.

Para avaliar a qualidade dos mapas obtidos pelas técnicas de estimação de profundidade previamente citadas, foram usadas as técnicas de similaridade RMSE e BMP para determinar quantitativamente a melhor reconstrução.

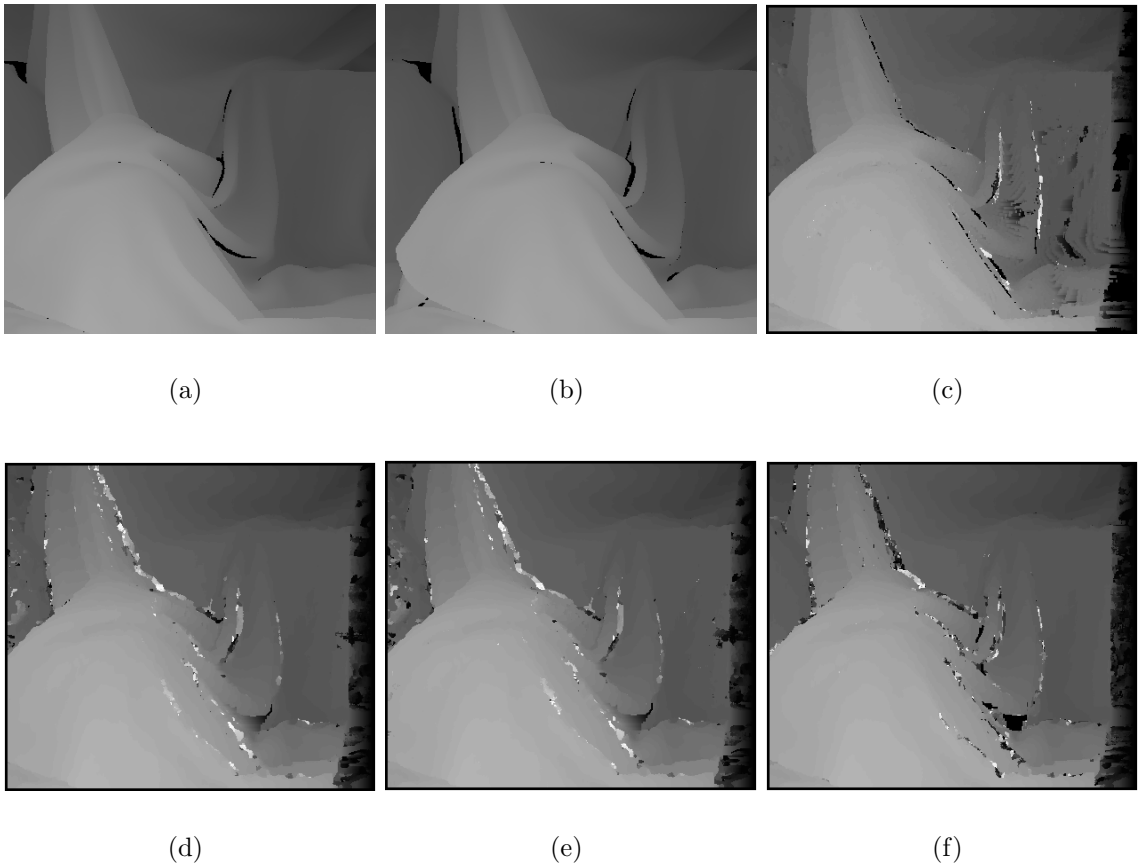


(a) (b) (c)

(d) (e) (f)

Figura 4.9: Resultados obtidos na etapa de experimentação com diminuição da iluminação (25). As imagens (a) e (b) representam respectivamente os mapas de profundidade reais obtidos a partir das imagens capturadas pelas câmeras da esquerda e direita. Na imagem (c) é apresentado o mapa de profundidade obtido a partir da técnica proposta. Enquanto que as imagens (d), (e) e (f) consistem nos resultados obtidos através das técnicas de estéreo clássicas, SSD, SAD e NCC, respectivamente.

Nos gráficos 4.12 e 4.13 é possível observar a eficiência da abordagem proposta, onde, mesmo com condições de iluminação variadas os mapas gerados pela estratégia de miniminação de energia proposta mostraram-se pouco impactados, ou seja, resultaram em mapas de profundidade com mais qualidade. Para uma melhor compreensão dos gráficos é importante mencionar que quanto menor a altura da coluna mostrada nos gráficos, melhor a qualidade da reconstrução.



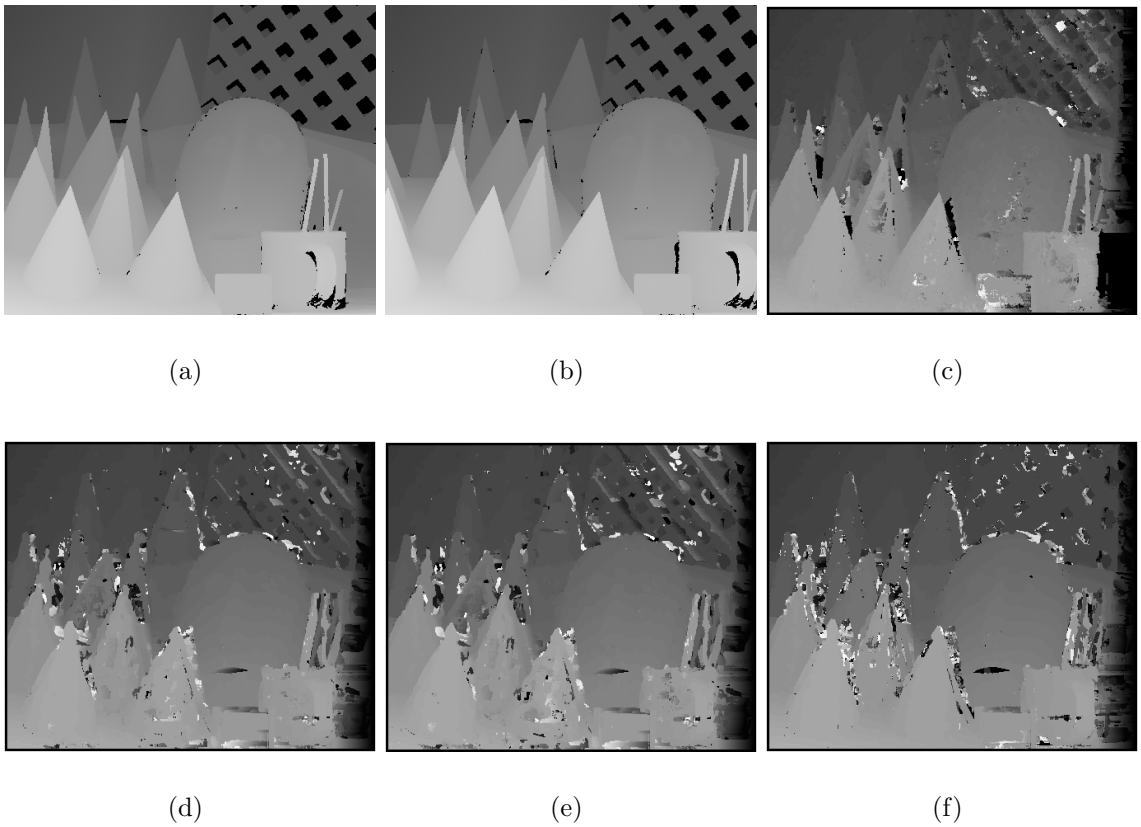
(a) (b) (c)

(d) (e) (f)

Figura 4.10: Resultados obtidos na etapa de experimentação considerando diminuição de luminosidade na cena. As imagens (a) e (b) representam respectivamente os mapas de profundidade com informações reais de profundidade obtidos a partir das imagens capturadas pelas câmeras da esquerda e direita. Na imagem (c) é apresentado o mapa de profundidade obtido a partir da técnica minimização de energia. Enquanto que as imagens (d), (e) e (f) consistem nos mapas obtidos através do SSD, SAD e NCC, respectivamente.

4.2.2.2 Variação de iluminação com alta luminosidade

Imagens com aumento no brilho serão consideradas para representar imagens com alta luminosidade, nessas tais imagens serão realizadas medições em imagens com brilho aumentado. Para reproduzir uma imagem com maior nível de iluminação é necessário que em cada pixel da imagem seja adicionado um valor que corresponda ao nível de luminosidade, para a presente pesquisa foi adicionado 25 a cada pixel. As Figuras 4.14, 4.15 e 4.16 apresentam os resultados obtidos na reconstrução da



(a) (b) (c)

(d) (e) (f)

Figura 4.11: Resultados obtidos a partir do processo de fusão proposto levando em conta baixa iluminação do ambiente monitorado. As imagens (a) e (b) consistem nos mapas de profundidade reais capturados pelas câmeras da esquerda e direita. A imagens (c) representa o mapa de profundidade resultante do processo de fusão por minimização de energia. Já nas imagens (d), (e) e (f) são mostrados os mapas providos pelas técnicas estéreo SSD, SAD e NCC.

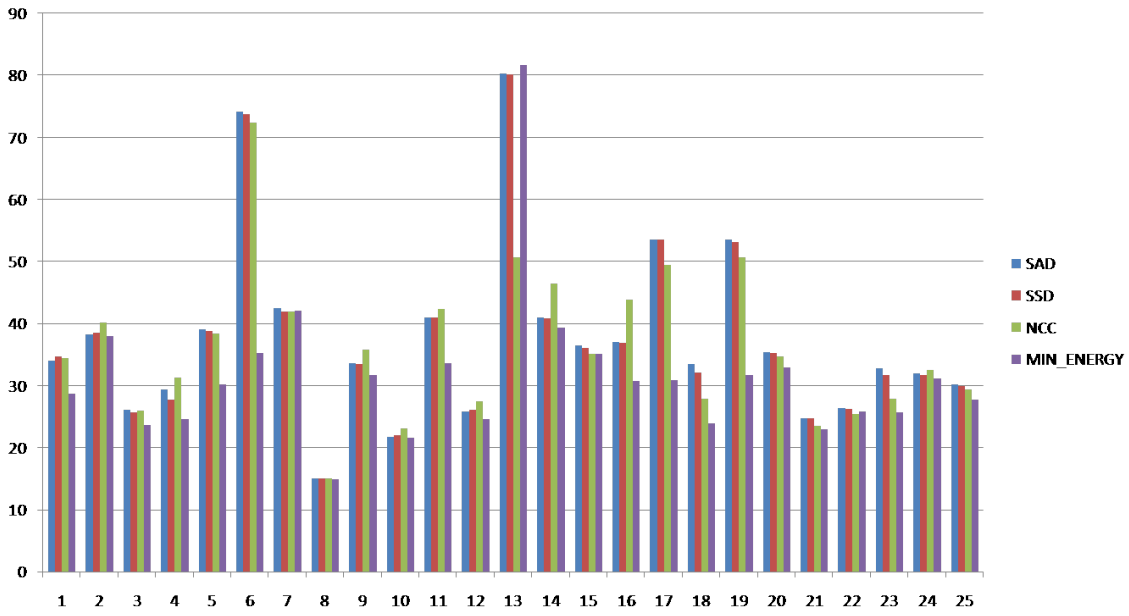


Figura 4.12: Comparação dos resultados usando RMSE. Gráfico comparando técnicas clássicas do estéreo e a estimação baseada em Funções de Energia com o mapa de profunidade real da imagem esquerda.

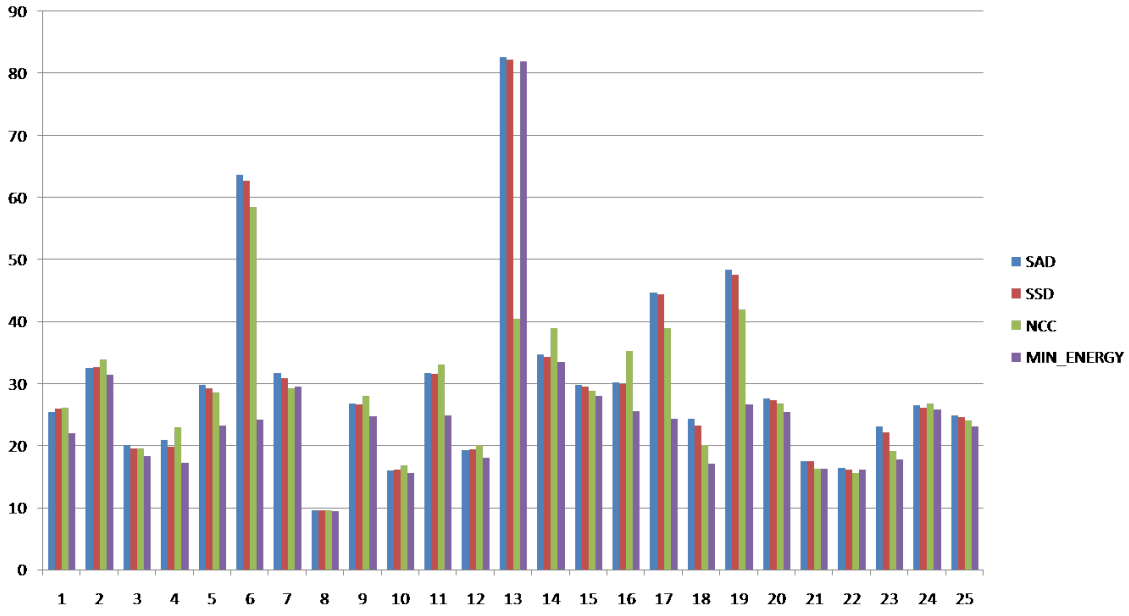
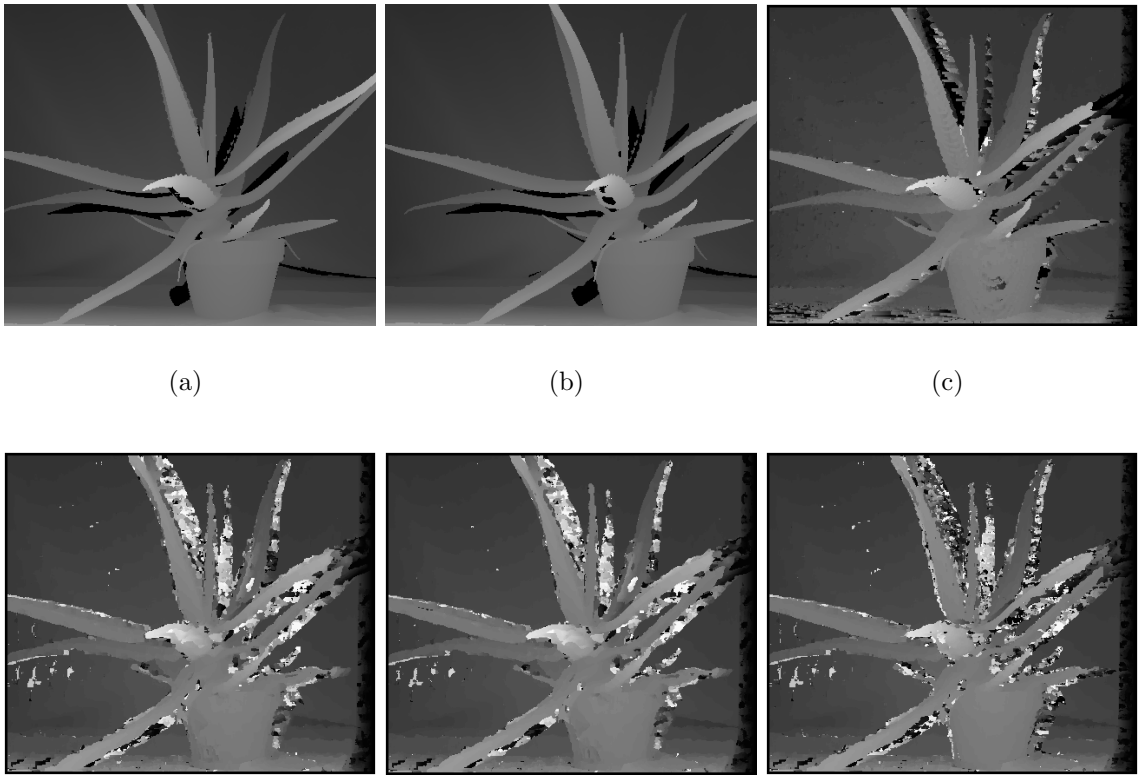


Figura 4.13: Comparação dos resultados usando BMP. Gráfico comparando técnicas clássicas do estéreo e a estimação baseada em Funções de Energia com o mapa de profunidade real da imagem esquerda.

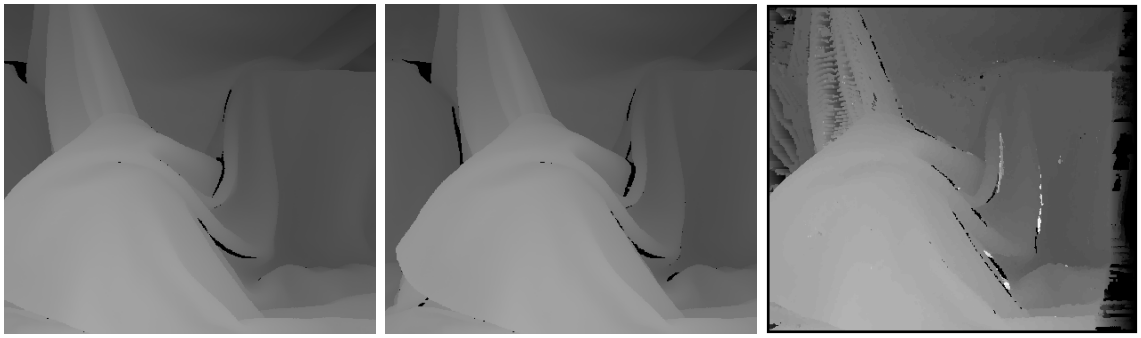
forma.



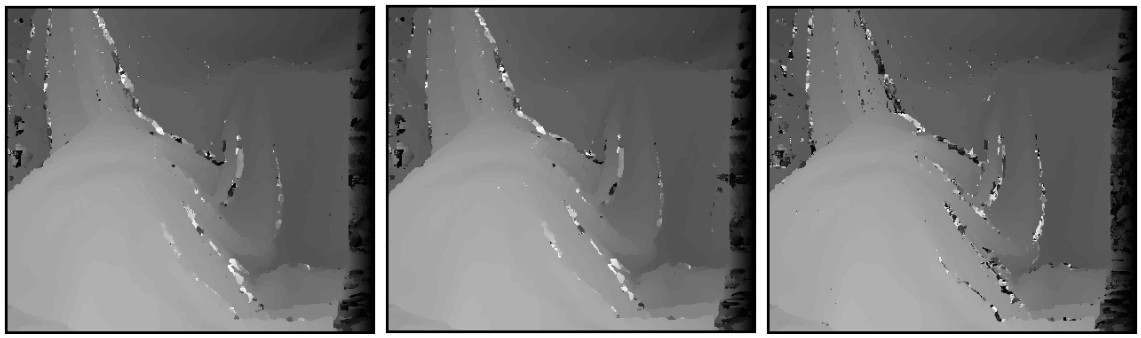
(a) (b) (c)

(d) (e) (f)

Figura 4.14: Resultados obtidos na etapa de experimentação com variação da iluminação (25). As imagens (a) e (b) representam respectivamente os mapas de profundidade reais obtidos a partir das imagens capturadas pelas câmeras da esquerda e direita. Na imagem (c) é apresentado o mapa de profundidade obtido a partir da técnica proposta. Enquanto que as imagens (d), (e) e (f) consistem nos resultados obtidos através das técnicas de estéreo clássicas, SSD, SAD e NCC, respectivamente.

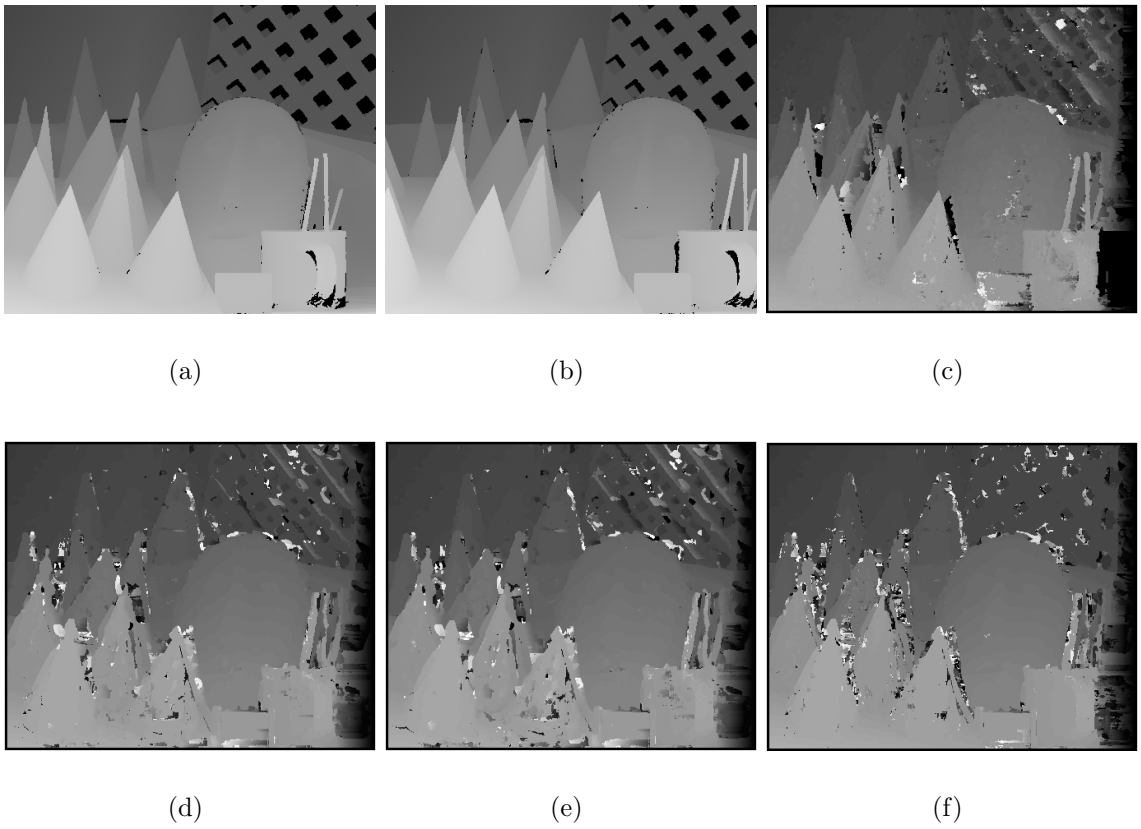


(a) (b) (c)



(d) (e) (f)

Figura 4.15: Resultados obtidos na etapa de experimentação considerando o aumento de luminosidade na cena. As imagens (a) e (b) representam respectivamente os mapas de profundidade com informações reais de profundidade obtidos a partir das imagens capturadas pelas câmeras da esquerda e direita. Na imagem (c) é apresentado o mapa de profundidade obtido a partir da técnica minimização de energia. Enquanto que as imagens (d), (e) e (f) consistem nos mapas obtidos através do SSD, SAD e NCC, respectivamente.



(a) (b) (c)

(d) (e) (f)

Figura 4.16: Resultados obtidos a partir do processo de fusão proposto levando em conta alta iluminação do ambiente monitorado. As imagens (a) e (b) consistem nos mapas de profundidade reais capturados pelas câmeras da esquerda e direita. A imagens (c) representa o mapa de profundidade resultante do processo de fusão por minimização de energia. Já nas imagens (d), (e) e (f) são mostrados os mapas providos pelas técnicas estéreo SSD, SAD e NCC.

As Figuras 4.14, 4.15 e 4.16 representam os mapas de profundidade obtidos pela estratégia de miniminação de energia proposta e pelas técnicas tradicionais de estimação de profundidade. Nas imagens ainda é possível constatar melhoria nas informações de profundidade alcançadas pela abordagem baseada em Função de Energia em detrimento das demais técnicas.

Para avaliar a qualidade dos mapas obtidos pelas técnicas de estimação de profundidade previamente citadas, foram usadas as técnicas de similaridade RMSE e BMP para determinar quantitativamente a melhor reconstrução.

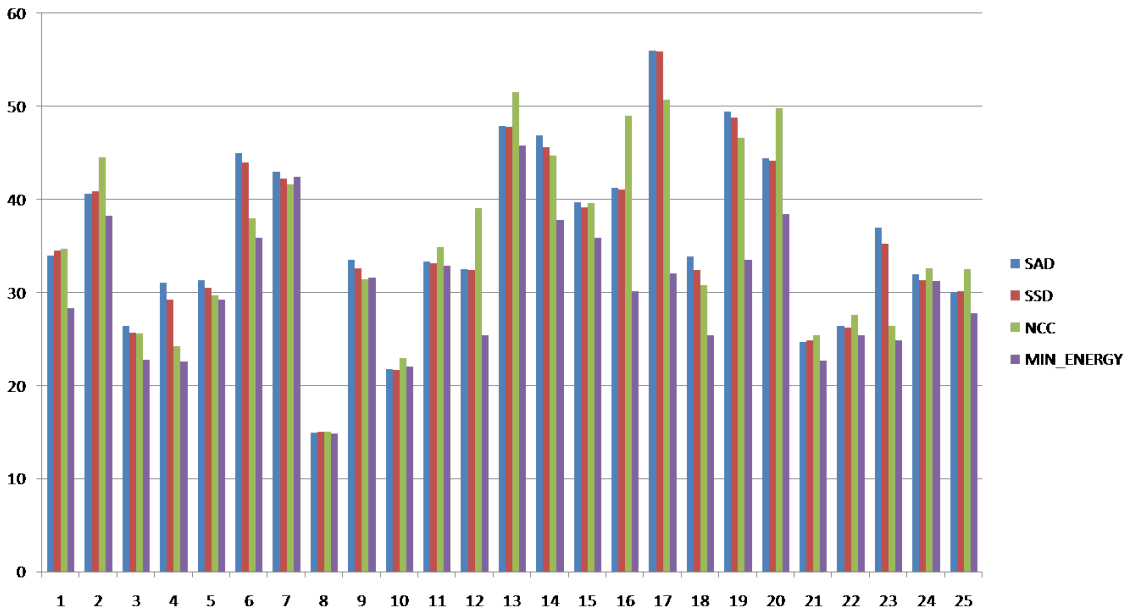


Figura 4.17: Comparação dos resultados usando RMSE. Gráfico comparando técnicas clássicas do estéreo e a estimação baseada em Funções de Energia com o mapa de profunidade real da imagem esquerda.

Nos gráficos 4.17 e 4.18 é possível observar a eficiência da abordagem proposta, onde, mesmo com condições de iluminação variadas os mapas gerados pela estratégia de miniminação de energia proposta apresentaram-se pouco impactados pelas mudanças na luminosidade, ou seja, resultaram em mapas de profundidade com mais qualidade.

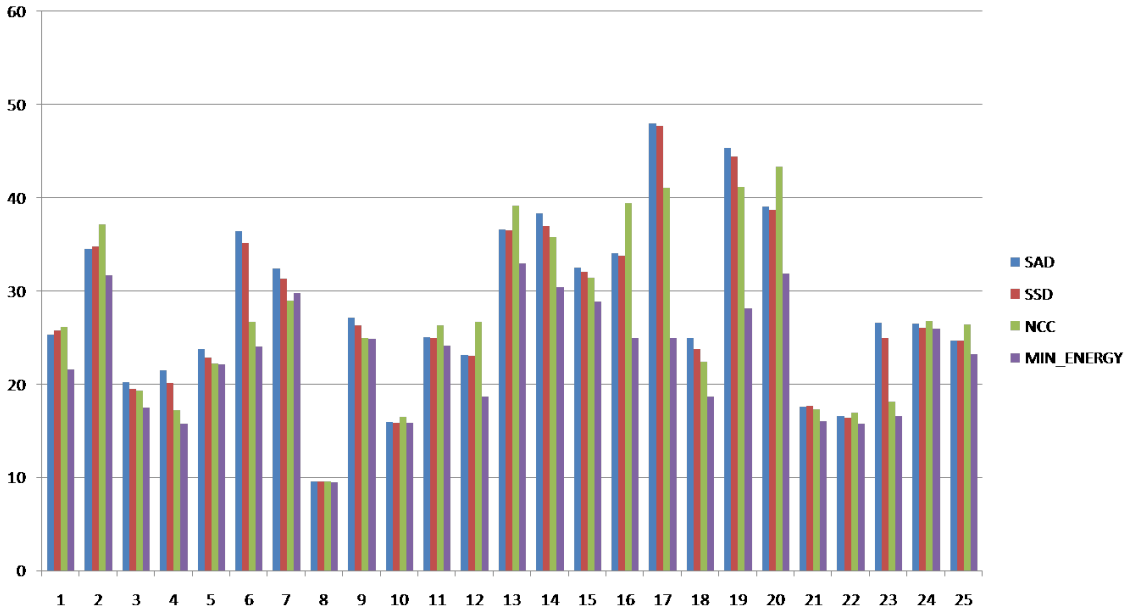


Figura 4.18: Comparação dos resultados usando BMP. Gráfico comparando técnicas clássicas do estéreo e a estimação baseada em Funções de Energia com o mapa de profunidade real da imagem esquerda.

Capítulo 5

Conclusões e Trabalhos Futuros

5.1 Conclusões

Este trabalho propõe uma abordagem para integrar as informações de energia de foco e de disparidade entre pixels considerando um ambiente de múltiplas câmeras. Essa abordagem é desenvolvida por meio de um processo que usa a Informação da energia de foco (obtida a partir das variações de foco dentre o conjunto de imagens) e Informação da energia de disparidade (energia a partir do brilho dos pixels dentre o par de imagens).

A Fusão baseia-se na minimização de energia visando evitar as falsas correspondências no processo de determinação dos pixels correspondentes. Para isso Funções de Energia foram utilizadas considerando medições energia referentes aos candidatos a correspondente, seus respectivos pixels adjacentes e medições de energia a partir da configuração focal dentre os conjuntos de imagens.

A estratégia proposta foi testada usando sequências de imagens adquiridas a partir de ambientes reais e seus resultados foram comparados com os resultados obtidos pelas técnicas tradicionais estéreo (SSD, SAD e NCC). Para isso foi utilizada a técnica de medição da similaridade Raiz do Erro Médio Quadrático (RMSE) que

visa determinar quão próxima a informação de profundidade obtida encontra-se da realidade. E a técnica de determinação de casamentos incorretos (BMP) que busca quantificar os casamentos realizados no processo de correspondência para inferir a qualidade do processo de estimação. Os experimentos demonstraram que a abordagem desenvolvida apresenta mapas de profundidade mais acurados que as abordagens clássicas estéreo (SSD, SAD e NCC), garantindo a qualidade e confiabilidade das informações de profundidade estimadas. Sendo importante mencionar que a abordagem proposta, como foi possível comprovar experimentalmente, não é impactada por variações na iluminação da cena. Isto nos permite inferir que a abordagem proposta se mostrará mais efetiva para aplicações práticas, tais como: automação industrial, robótica móvel dentre outras.

5.2 Limitações do Trabalho

A principal limitação é a falta de experimentação operacional, em ambiente real e não controlado. Isso deu-se devido a falta de infraestrutura necessária necessária para a captura de duas sequências de imagens em configuração estéreo com variação de foco. Tal procedimento de variação controlada de foco tornou-se inviável devido a falta de equipamentos para o desenvolvimento desta pesquisa. Outra limitação está na formulação mais rigorosa do modelo de energia utilizado.

5.3 Trabalhos Futuros

O prosseguimento da linha de pesquisa trilhada neste trabalho contempla algumas melhorias a serem adicionadas ao trabalho, tais como:

1. Realizar experimentos operacionais fazendo uso da rede de câmeras criada no laboratório. Com o uso de equipamentos adequados, experimentos controlados

no laboratório seriam conduzidos de modo a validar a abordagem proposta nesse contexto.

2. Aprimorar os mapas de profundidade gerados pela técnica de reconstrução a partir do foco. Existem algumas técnicas para a extração da forma a partir do foco, tais técnicas apresentam maior eficiência para determinado tipo de cena monitorado. Desse modo uma abordagem de Fusão de técnicas de foco poderia possibilitar resultados mais acurados.

3. Integrar a abordagem desenvolvida com Robótica Móvel de modo permitir a navegação de robôs. Este modelo deve apresentar tolerância a falhas, caso uma câmera apresente problemas a outra deve ser capaz de extrair a profundidade da cena.

4. Modelar o problema para ambientes com larga linha de base (*wide baseline*). O estudo dos impactos apresentados pela variação da linha de base poderia atribuir maior qualidade à informação estimada.

5.4 Considerações Finais

O trabalho contribuiu para consolidação da proposta da pesquisa em Visão Computacional distribuída, promovendo o aprimoramento de novas técnicas de obtenção de mapas de profundidade. Os conhecimentos para a realização da pesquisa foram adiquiridos com a colaboração do grupo de pesquisa ao qual o trabalho está contido, estigando a busca por novos desafios e soluções. O interesse pelo conteúdo estimula a busca por um aprofundamento dos conhecimentos inerentes à área de concentração e a contribuição com a comunidade científica, por meio de publicações dos resultados obtidos. O trabalho contou com o apoio do CNPq e da FAPEAM, sem os quais não seria possível sua realização.

Referências Bibliográficas

Abidi, M. A. and Gonzalez, R. C. (1992). *Data fusion in robotics and machine intelligence*. Academic Press Professional, Inc., San Diego, CA, USA.

Baltzakis, H., Argyros, A., and Trahanias, P. (2003). Fusion of laser and visual data for robot motion planning and collision avoidance. *Mach. Vision Appl.*, 15:92–100.

Bussab, W. and Morettin, P. (2010). *Estatística Básica*. Editora Saraiva.

DeSouza, G. N. and Kak, A. C. (2002). Vision for mobile robot navigation: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24:237–267.

Frese, C. and Gheta, I. (2006). Robust depth estimation by fusion of stereo and focus series acquired with a camera array. In *2006 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pages 243–248.

Gheta, I., Frese, C., Heizmann, M., and Beyerer, J. (2007). A new approach for estimating depth by fusing stereo and defocus information. In *GI Jahrestagung (1)*, pages 26–31.

Horn, B. K. (1986). *Robot Vision*. McGraw-Hill Higher Education, 1st edition.

Incorporated, A. S. (2009). Adobe photoshop cs5.

Krotkov, E. and Bajcsy, R. (1993). Active vision for reliable ranging: Cooperating

focus, stereo, and vergence. *International Journal of Computer Vision*, 11:187–203. 10.1007/BF01469228.

Martins, N. and Dim, J. (2008). Visual inspection and 3d reconstruction based on image-t o-mirror planes.

Nayar, S. K. and Nakagawa, Y. (1994). Shape from focus. *IEEE TPAMI*, 16:824–831.

Pedrini, H. and Schwartz, W. R. (2007). *Análise de Imagens Digitais: Princípios, Algoritmos e Aplicações*. Cengage Learning.

Rajagopalan, A. N., Chaudhuri, S., and Mudenagudi, U. (2004). Depth estimation and image restoration using defocused stereo pairs. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26:1521–1525.

Salustiano, R. E. (2002). *Aplicação de técnicas de fusão de sensores no monitoramento de ambientes*. Universidade Estadual de Campinas, Campinas, Brasil.

Scharstein, D. and Szeliski, R. (2001). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *INTERNATIONAL JOURNAL OF COMPUTER VISION*, 47:7–42.

Subbarao, M., Yuan, T., and Tyan, J.-K. (1997). Integration of defocus and focus analysis with stereo for 3d shape recovery. In *Shape Recovery, Three-Dimensional Imaging and Laser-based Systems for Metrology and Inspection III*, pages 14–15.

Trucco, E. and Verri, A. (1998). *Introductory Techniques for 3-D Computer Vision*. Prentice Hall PTR, Upper Saddle River, NJ, USA.

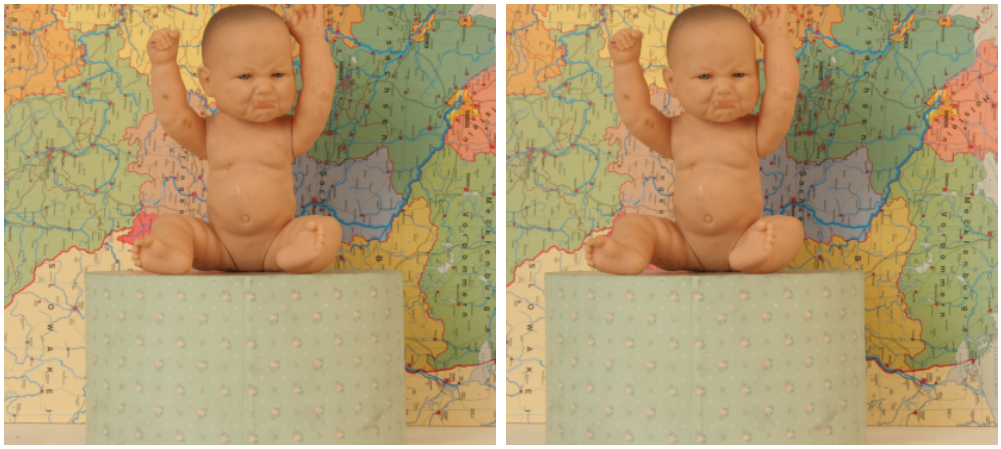
V. Michael Bove, J. (1990). Probabilistic method for integrating multiple sources of range data. *J. Opt. Soc. Am. A*, 7(12):2193–2198.

Apêndice A

Reconstrução da forma pelo Estéreo Fotométrico

Shape from Stereo é o processo de combinação e análise de duas ou mais imagens de uma cena que visa estimar as coordenadas tridimensionais do ambiente monitorado. A sensação de profundidade permitida pela visão estéreo possibilita que o ser humano possa estimar a forma de objetos e com isso avaliar o grau de proximidade ou afastamento de objetos presentes na cena. Na Visão estéreo assim como na visão humana, o olho esquerdo e o olho direito vêem imagens diferentes embora muito parecidas. E essa diferença é usada pelo cérebro para constituir uma imagem com as informações de profundidade.
A captura de imagens realizada pela técnica *Shape From Stereo* consiste na aquisição de um par de imagens providas por uma par de câmeras, como observado no exemplo da Figura A.1.

As câmeras encontram-se alinhadas verticalmente, deslocadas por uma distância horizontal B, denominada Linha de Base (*Baseline*), conforme a Figura A.2.



(a) (b)

Figura A.1: Imagens digitais capturadas por um sistema estéreo de visão. A Figura (a) representa a imagem capturada pela câmera esquerda e a Figura (b) representa a imagem capturada pela câmera direita.

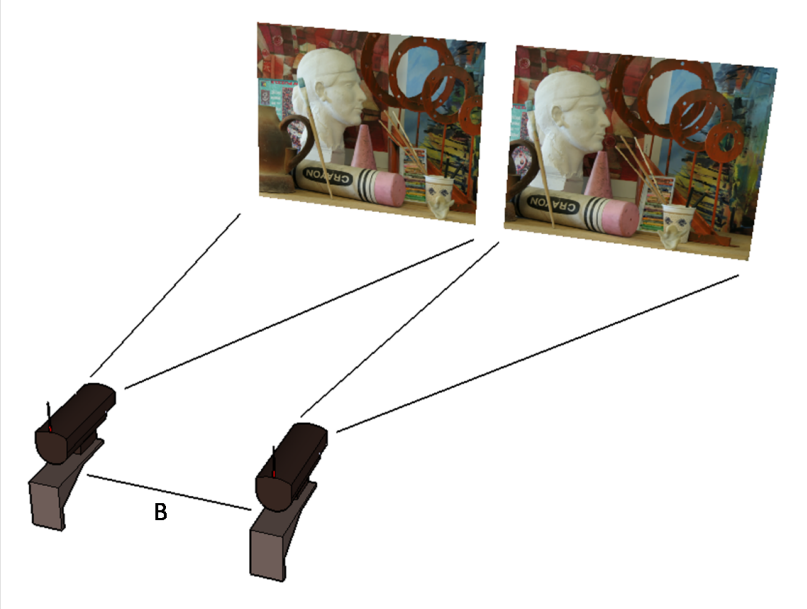


Figura A.2: Sistema de visão stereo composto por duas câmeras separadas por uma distância B denominada *Baseline*.

Para realizar a computação do procedimento de reconstrução estéreo são necessárias 4 etapas, como observado na Figura A.3. As etapas são:

- Retificação das imagens;

- Correspondência entre Pontos;
- Cálculo do Mapa de Disparidades;
- Criação do Mapa de Profundidade.

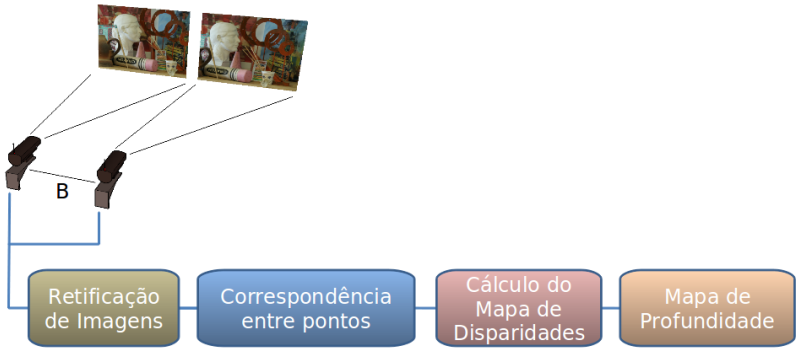


Figura A.3: Etapa do processo de Visão estéreo. Inicialmente, o sistema de câmeras captura um par de imagens. Segundo, é executado o processo de retificação de imagens para alinhamentos das coordenadas. Então, são determinados os pontos correspondentes dentre as duas imagens. Despois, é computado o cálculo do mapa de disparidade da cena monitorada. Finalmente, é encontrado o mapa de profundidade contendo informações a respeito da estrutura tridimensional do ambiente.

As etapas apresentadas contemplam os procedimentos necessários para a execução da técnica de estimação de profundidade estéreo. Após a aquisição do par de imagens que inicia o processo é preciso registrar as imagens para que as coordenadas das duas imagens estejam alinhadas, esse processo de alinhamento é denominado Retificação de Imagens. A etapa de Retificação não será detalhada, pois as imagens utilizadas no trabalho já encontram-se alinhadas.

O próximo passo é considerado como um dos principais problemas encontrados na técnica estéreo. A Correspondência entre Pontos consiste em determinar quais partes das imagens da esquerda e direita são projeções do mesmo elemento da cena. Para essa finalidade são apresentadas a seguir algumas técnicas consolidadas para determinação da correspondência.

Soma das Diferenças quadráticas (SSD):

$$SSD(d) = \sum_{x,y \in W} (I_k^L(x,y) - I_k^R(x,y+d))^2, \tag{A.1}$$

onde I_k^L representa a imagem primária e I_k^R representa a imagem secundária considerando uma janela de tamanho W. Soma das Diferenças Absolutas (SAD):

$$SAD(d) = \sum_{x,y \in W} |(I_k^L(x,y) - I_k^R(x,y+d))|, \tag{A.2}$$

onde I_k^L representa a imagem primária e I_k^R representa a imagem secundária considerando uma janela de tamanho W. Correlação Cruzada Normalizada (NCC):

$$NCC(d) = \frac{\sum_{x,y \in W} (I_k^L(x,y) \times I_k^R(x,y+d))}{\sum_{x,y \in W} I_k^{L^2}(x,y) \times \sum_{x,y \in W} I_k^{R^2}(x,y+d)}, \tag{A.3}$$

onde I_k^L representa a imagem primária e I_k^R representa a imagem secundária considerando uma janela de tamanho W.

O Cálculo do Mapa de Disparidades representa a diferença entre os pontos correspondentes nas imagens, como ilustrado na Equação A.4 , sendo fundamental para a Criação do Mapa de Profundidade que utiliza a disparidade e valores referentes aos parâmetros instrínsecos das câmeras para a reconstrução da estrutura.

$$Disparidade(i,j) = (I_k^L(x,y) - I_k^R(x,y+d)), \tag{A.4}$$

Apêndice B

Reconstrução da forma a partir do foco

Shape From Focus é uma técnica de estimação da forma de objetos ou cenas a partir de uma série de imagens. A aquisição das imagens consiste em um processo de variação do foco durante a captura das mesmas, ou seja, o foco da imagem se desloca em uma quantidade finita de passos, conforme a Figura B.1.

Para computar as informações de profundidade referentes à cena usando *Shape From Focus* são necessárias 4 etapas, como pode ser visualizado na Figura B.2. As etapas são:

- Cálculo da Medida de Foco das Imagens;
- Estimação da Profundidade;
- Criação do Mapa de Profundidade.

A técnica de estimação de profundidade baseada em foco decomposta nas etapas acima é apresentada em um esquema contemplando seu funcionamento, como pode ser visto na Figura B.2. Primeiramente, uma sequência de imagens com variação de foco é capturada a partir de uma única câmera.

Figura B.1: Sistema de visão usado na captura de uma série de imagens. Os parâmetros intrínsecos da câmera são variados a cada aquisição obtendo assim imagens com variação de foco.

Em seguida, o foco referente a cada ponto da série de imagens é mensurado por meio de técnicas de medição do foco. Na literatura existem várias estratégias para determinar qual ponto apresenta melhor foco, tais como: *Tenengrad*, variância, Soma do Laplaciano e Soma do Laplaciano Modificado. Sendo as mesmas avaliadas, determinando a técnica SML como a melhor dentre elas (Krotkov)(Nayar). Para essa finalidade as técnicas realizam operações que envolvem uma janela de tamannho definido pela variável *step* das Equações B.1 e B.2 e informações referentes à vizinhança dentre a série de imagens, como ilustrado na Figura B.3.

As Equações abaixo descrevem a técnica de medição de foco SML,

$$ML(x,y) = |2I(x,y) - I(x - step, y) - I(x + step, y)| + \tag{B.1}$$

$$|2I(x,y) - I(x, y - step) - I(x, y + step)|. \tag{B.2}$$

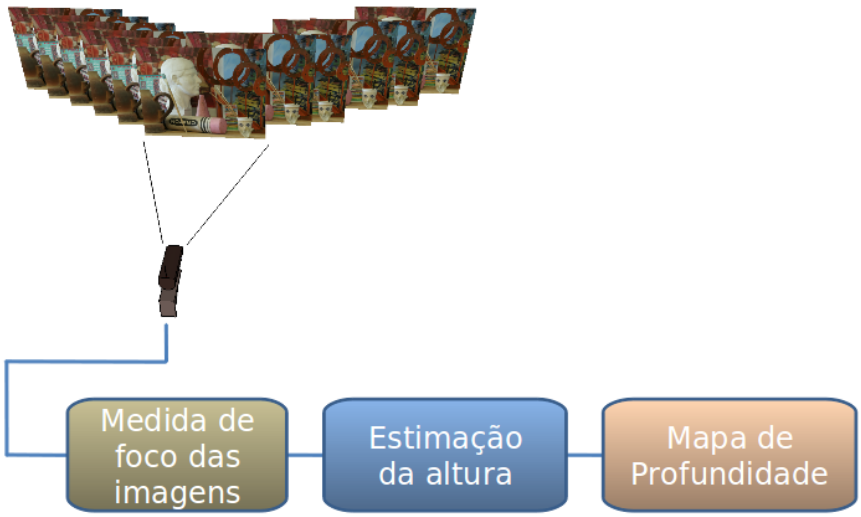


Figura B.2: Esquematização do processo de determinação da forma tridimensional de uma cena a partir da informação de foco nas imagens. No primeiro instante uma sequência de imagens é capturada com variação do foco. Posteriormente é realizada a medição do foco dentre as imagens, determinando para cada ponto do conjunto de imagens qual apresenta melhor foco. Em seguida é realizado o calculo de distância com base nos pixels mais bem focados da etapa anterior e por fim é contruído o mapa de profundidade da cena.

Então, a medição de foco SML,

$$SML(i,j) = \sum_{x=i-N}^{i+N} \sum_{y=j-N}^{j+N} ML(x,y) \text{ for} ML(x,y) \geq T, \tag{B.3}$$

Com a informação de foco previamente calculada para cada imagem do conjunto é possível inferir a distância no ambiente monitorado, por meio da estratégia de estimação de distância baseada na Interpolação Gaussiana, como é possível visualizar na Equação B.4. A interpolação gaussiana seleciona as três melhores medições de foco (F_m: melhor, F_{m-1}: segunda melhor e F_{m+1}: terceira melhor) para calcular a profundidade e a posição na sequência de imagens onde se encontram tais medições (d_m, d_{m-1} e d_{m+1}).

$$D = \frac{(lnF_m - lnF_{m+1}).(d_m^2 - d_{m+1}^2) - (lnF_m - lnF_{m-1}).(d_m^2 - d_{m-1}^2)}{2\delta\{(lnF_m - lnF_{m-1}) + (lnF_m - lnF_{m+1})\}} \tag{B.4}$$

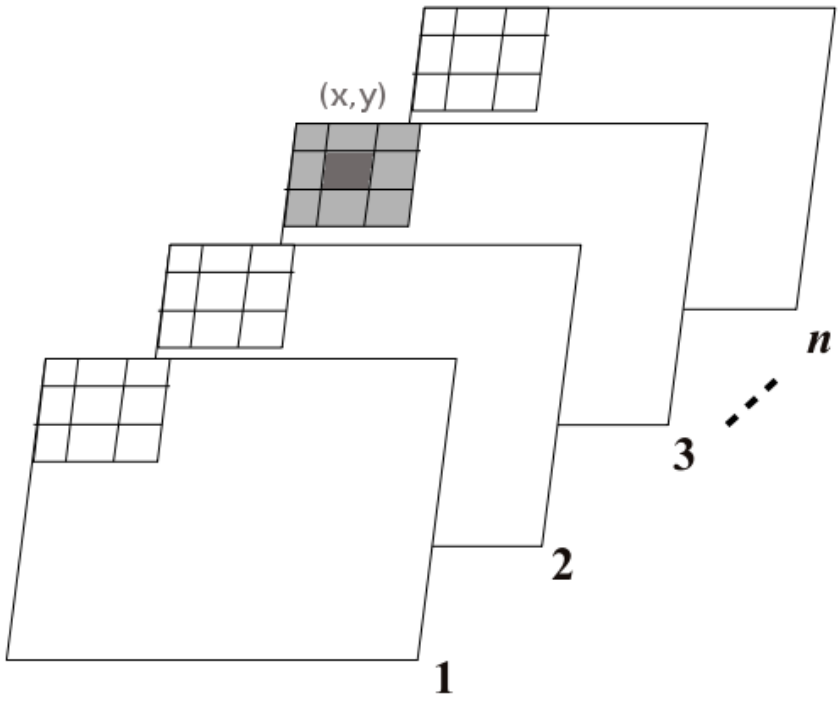


Figura B.3: Modelo de convolução realizado para determinar o pixel melhor focado dentre o conjunto de imagens. Neste modelo o foco de cada ponto em cada imagem é medido visando determinar em qual das imagens o ponto encontra melhor situação focal.

onde F_m, F_{m+1} e F_{m-1} representam os melhores valores computados na medição de foco. Enquanto d_m, d_{m+1} e d_{m-1} consistem nos deslocamentos de cada medição, ou seja, em quais imagens do conjunto foram encontradas as melhores medições. Finalmente, é criado um mapa de profundidade a partir da estimação realizada, resultando em um mapa com a estrutura tridimensional da cena.

Apêndice C

Estado da arte: Estimações de Profundidade

Com o passar do tempo a Reconstrução da estrutura 3-D de uma cena tem se mostrado uma tarefa complexa para a Visão Computacional, devido a problemas como: irregularidades na cena, posicionamento da(s) câmera(s), tipos variados de superfícies e problemas de iluminação. Por essa razão várias técnicas foram proposta para solucionar o problema de Estimação das distâncias de um ambiente, que coletivamente são conhecidas como *Shape-from-X* Trucco and Verri (1998).
Cada técnica de reconstrução aborda características diferentes e apresenta uma metodologia própria para tratar o problema, como ilustrado na Tabela C.1 abaixo:

A Tabela C.1 apresenta algumas das técnicas de estimação da estrutura tridimensional de um ambiente mais consolidadas na literatura, contemplando características e uma breve descrição de cada técnica. Para um conhecimento mais aprofundado sobre as técnicas *Shape-from-X* uma consulta às referências é recomendada.

Shape from	Quantidade de Imagens	Descrição
Stereo	2 (duas) ou mais	Técnica que utiliza a disparidade entre o par de imagens para determinar a profundidade na cena.
Focus	2 (duas) ou mais	Estratégia baseada na informação de foco presente na sequência de imagens para determinar a distância.
Shading	1 (uma)	Método que consiste no uso de sombras contidas na cena para inferir a estrutura do ambiente.
Texture	1 (uma)	Abordagem que trata a reconstrução da forma a partir da textura encontrada nos objetos da cena.

Tabela C.1: Técnicas de Reconstrução da forma de cenas com características e breve descrição das abordagens.

Neste trabalho serão abordadas duas técnicas de estimação de profundidade, a primeira baseada na informação de foco contida em um conjunto de imagens (*Shape from Focus*) e a segunda baseada na projeção dentre um par de imagens estéreo (*Shape from Stereo*).

Dentre as técnicas baseadas em foco destacam-se os trabalhos de Nayar et. al. (1994), onde é proposta uma nova técnica de medida de foco baseada na filtragem das altas frequências, sendo usada uma versão modificada do filtro Laplaciano (SML). Segundo o autor a medida de foco SML mostra-se a melhor alternativa para estimação de profundidade baseada no foco. No trabalho de Subbarao et. al. (1993), é apresentado um conjunto de técnicas de medida de foco classificando-as como: técnica de maximização de energia não filtrada, filtro passa-baixa, filtro passa-alta, filtro passa-banda. Dentre as técnicas testadas destaca-se o método baseado na Energia do gradiente da imagem com filtro passa-baixa. Enquanto que na abordagem proposta por Pradeep et. al. (2007), é apresentado um método para estimação da forma de objetos 3-D. O método apresenta melhorias nas técnicas tradicionais de depth from focus, usando o borramento de desfoque relativo das imagens como informação auxiliar para obter a forma tridimensional do objeto com melhores resultados que as abordagens tradicionais, considerando os casos testados.

Na literatura dentre os trabalhos baseados na técnica Stereo que mais se destaca está o trabalho de Truco et. al. (1998), que apresenta todo o problema da técnica Shape From Stereo enfatizando cada etapa do processo. O problema de determinação dos pontos correspondentes é uma das etapas mais críticas para o desenvolvimento do Shape From Stereo e Trucco propôs duas técnicas para a resolução desse problema. A primeira baseada na correlação entre os pontos. Enquanto a segunda técnica realiza a busca por propriedades de algumas características contidas na imagem. Para que seja possível realizar a busca por pontos correspondentes é fundamental que as imagens estejam Retificadas, ou seja, alinhadas. Nesse sentido Trucco apresenta as propriedades da Geometria Epipolar visando a computação da matriz essencial E e da matriz fundamental F que são utilizadas na computação do algoritmo de Oito Pontos, sendo posteriormente calculada a imagem Retificada. Por fim é realizada a Recontrução 3-D que necessita de conhecimento a priori, tendo como algumas técnicas a Reconstrução por Triangulação, Reconstrução pelo Fator de Escala e Reconstrução por Transformação Projetiva. A Reconstrução por Triangulação assume que os parâmetros internos e externos são conhecidos e calcula a localização 3-D dos pontos a partir das projeções dos pontos. Na Reconstrução baseada no Fator de Escala é assumido que apenas os parâmetros internos são conhecidos e n correspondentes são dados e com isso é possível calcular a posição 3-D dos pontos a partir de suas projeções. A Reconstrução por Transformação Projetiva assume n pontos correspondentes conhecidos e a localizacão dos epipolos, com isso é possível calcular as coordenadas 3-D dos pontos da cena.