Gustavo Teodoro Laureano

**Criação de mapas de disparidades empregando análise
multi-resolução e agrupamento perceptual**

Dissertação apresentado à Escola de Engenharia de São Carlos
da Universidade de São Paulo, como parte dos requisitos para
obtenção do título de Mestre em Engenharia Elétrica

Área de Concentração: Processamento de Sinais e Instru-
mentação

Orientadora: Prof^a. Dr^a. Maria Stela Veludo de Paiva

Aos meus pais, José Valoir Laureano e Laurice Teodoro Laureano, com muito amor e admiração dedico esse trabalho pelo exemplo, pelo cuidado e pela incansável luta para que eu e meus irmãos sempre tivéssemos honra, coragem e respeito, e por se entregarem completamente aos interesses comuns da nossa família.

Agradecimentos

Ao Grande Arquiteto do Universo, Senhor de todos os Senhores e Conhecedor de todos os mistério que buscamos desvendar.

À Dr^a Maria Stela Veludo de Paiva, por todo apoio científico no desenvolvimento desse trabalho e pela confiança em mim depositada.

À minha família e à minha namorada, Regiane, que sempre me deram amor, cuidado e apoio no decorrer desse trabalho.

Ao Dr. Clarimar José Coelho, sem o qual não teria aberto os olhos para a pesquisa científica.

Aos vários amigos que tive o prazer de conhecer nos Laboratórios da EESC/USP, na república e em especial, no Laboratório de Visão Computacional (LAVI).

À Capes pelo apoio financeiro, que certamente foi de grande valia para a conclusão desse trabalho.

A todos aqueles que de alguma forma estiveram presentes na minha jornada pela busca do conhecimento.

"Non nobis, Domine, non nobis, sed nomini Tuo da gloriam"

Resumo

O trabalho apresentado por essa dissertação busca contribuir com a atenuação do Problema da Correspondência em Visão Estéreo a partir de uma abordagem local de soluções. São usadas duas estratégias como solução às ambigüidades e às oclusões da cena: a análise multi-resolução das imagens empregando a estrutura piramidal, e a força de agrupamento perceptual, conhecida como *Gestalt Theory* na psicologia. Inspirado no sistema visual humano, a Visão Estéreo é uma área de grande interesse em Visão Computacional, e está relacionada à recuperação de informações tridimensionais de uma cena a partir de imagens da mesma. Para isso, as imagens são capturadas em posições diferentes para o futuro relacionamento das várias projeções de um mesmo ponto 3D. Apesar de ser estudada há quase quatro décadas, ela ainda apresenta problemas de difícil solução devido às dificuldades relacionadas às distorções produzidas pela mudança da perspectiva de visualização. Dentre esses problemas destacam-se os relacionados à oclusão de pontos e também à ambigüidade gerada pela repetição ou ausência de textura nas imagens. Esses por sua vez compõem a base do problema estéreo, chamado de Problema da Correspondência. Os resultados obtidos são equivalentes aos obtidos por técnicas globais, com a vantagem de requerer menor complexidade computacional. O uso da teoria de agrupamento perceptual faz desse trabalho um método moderno de estimação de disparidades, visto que essa técnica é alvo de atenção especial em recentes estudos na área de visão computacional.

Palavras-chave: Visão Estéreo, Análise Multi-Resolução, Disparidade, *Gestalt Grouping*.

Abstract

This work aims to give a contribution to the Correspondence Problem using a local approach. Two strategies are used as solution to ambiguities and occlusions: the multi-resolution analysis with images pyramids and the other is the perceptual grouping weight, called Gestalt Theory in the psychology. Inspired by human vision system, the Stereo Vision is an very important area in Computer Vision. It is related with the 3D information recovery from a pair of images. The images are captured from different positions to hereafter association of the 3D point projections. Although it has being studied for quite a long time, stereo vision presents some difficult problems, related to the change of visualisation perspective. Among the different problems originated from point of view changes, occlusions and ambiguities have special attention and compose the foundation of stereo problem, named Correspondence Problem. The results obtained were closer to the ones generated by global techniques, with the advantage of requiring less computational complexity. The use of Gestalt Theory makes this a modern disparity estimation method, as this theory has been received special attention in computer vision researchs.

Key words: *Stereo Vision, Multi-Resolution Analysis, Disparity, Gestalt Grouping.*

Lista de Figuras

1.1	Esquema geral de um sistema de visão estéreo.	3
2.1	Modelo geométrico de uma câmera.	8
2.2	Geometria estéreo binocular.	10
2.3	Ambigüidade de regiões em imagens estéreo.	13
2.4	Exemplo de regiões oclusas em imagens estéreo.	14
2.5	Repetição de padrões de textura.	14
3.1	Geometria epipolar.	18
3.2	Sistema estéreo binocular coplanar.	20
3.3	Retificação de imagens estéreo.	21
3.4	Situação evitada pela restrição de unicidade.	22
3.5	Falha ao considerar a restrição de unicidade.	23
3.6	Ordenação de correspondências.	24
3.7	Problema em considerar a restrição de ordenação.	24
4.1	Exemplo do processo de correspondência.	28
4.2	Representação do espaço de disparidade da imagem.	29
4.3	Representações para a aplicação de programação dinâmica.	36
5.1	Janelas deslocadas reduzindo o problema de oclusão.	41
5.2	Janelas deslocadas.	42

6.1	Esquema geral da metodologia adotada.	45
6.2	Gráfico de similaridade.	47
6.3	Agrupamento por similaridade de cor.	47
6.4	Agrupamento por proximidade espacial.	48
6.5	Agregação dos pesos de Similaridade e Proximidade: Janela suporte.	49
6.6	Estrutura piramidal.	50
6.7	Decomposição em pirâmides.	51
6.8	Busca por regiões com e sem a decomposição piramidal.	52
6.9	Redução da ambigüidade com análise multi-resolução.	53
6.10	Piramide de disparidades.	54
6.11	Número de chamadas à função de custo de similaridade considerando $t_d = 15$. . .	55
6.12	Fluxograma do algoritmo proposto.	58
7.1	Par estéreo <i>Tsukuba</i>	62
7.2	Par estéreo <i>Venus</i>	62
7.3	Par estéreo <i>Teddy</i>	63
7.4	Par estéreo <i>Cones</i>	63
7.5	Melhor resultado para o par <i>Tsukuba</i> . Obtido com $n_v = 17$, $\gamma_s = 9$ e $B = 2.37$. . .	65
7.6	Melhor resultado para o par <i>Venus</i> . Obtido com $n_v = 17$, $\gamma_s = 9$ e $B = 1.47$	65
7.7	Melhor resultado para o par <i>Teddy</i> . Obtido com $n_v = 15$, $\gamma_s = 18$ e $B = 0.23$. . .	66
7.8	Melhor resultado para o par <i>Cones</i> . Obtido com $n_v = 13$, $\gamma_s = 11$ e $B = 0.44$. . .	66
7.9	Resultados do par <i>Tsukuba</i> para outras estratégias de correspondência.	67
7.10	Resultados do par <i>Venus</i> para outras estratégias de correspondência.	68
7.11	Resultados do par <i>Teddy</i> para a técnica <i>Shiftable Windows</i>	69
7.12	Resultados do par <i>Cones</i> para a técnica <i>Shiftable Windows</i>	70
7.13	Gráfico de erro para o par <i>Tsukuba</i>	72
7.14	Gráfico de erro para o par <i>Venus</i>	73

7.15	Gráfico de erro para o par <i>Teddy</i>	74
7.16	Gráfico de erro para o par <i>Cones</i>	75

Lista de Tabelas

4.1	Pseudo-código de criação do EDI	29
4.2	Métricas de correspondência utilizadas no casamento de regiões.	33
7.1	Parâmetros característicos dos pares estéreo.	64
7.2	Valores de B para os pares <i>Tsukuba</i> e <i>Venus</i>	69
7.3	Valores de B para a técnica <i>Shifttable Windows</i>	69
7.4	Melhores resultados para o par <i>Tsukuba</i>	76
7.5	Melhores resultados para o par <i>Venus</i>	76
7.6	Melhores resultados para o par <i>Teddy</i>	76
7.7	Melhores resultados para o par <i>Cones</i>	77
9.1	Resultados do par <i>Tsukuba</i>	81
9.2	Resultados do par <i>Venus</i>	82
9.3	Resultados do par <i>Teddy</i>	82
9.4	Resultados do par <i>Cones</i>	83

Lista de Abreviaturas e Siglas

ARPA *Advanced Research Projects Agency*

IU *Image Understanding Community*

CCZN Correlação Cruzada Zero Normalizada

CCN Correlação Cruzada Normalizada

CCNR Correlação Cruzada Normalizada Rápida

DSI *Disparity Space Image*

EDI Espaço de Disparidade da Imagem

SQD Soma dos Quadrados das Diferenças

SQDN Soma dos Quadrados das Diferenças Normalizada

SDA Soma das Diferenças Absolutas

EQM Erro Quadrático Médio

TR Transformação *Rank*

TC Transformação *Celsius*

Lista de Símbolos

Convenções geométricas

Ω, Π	Planos geométricos são designados com letras gregas maiúsculas em negrito.
$\mathbf{l}$	Retas são designadas com letras minúsculas em negrito.
P, O	Pontos no mundo tridimensional são designados com letras maiúsculas em itálico.
p, q	Projeções bidimensionais são designadas com letras minúsculas em itálico.

Convenções matemáticas

a, x, N, R	Escalaes são escritos com caracteres normais em itálico.
$\mathbf{M}, \mathbf{E}, \mathbf{F}$	Matrizes são escritas com letras maiúsculas em negrito.
$\mathbf{d}, \mathbf{v}$	Vetores são escritos com letras minúsculas em negrito.
$\alpha, \beta, \gamma, \lambda, \psi$	Constantes são escritas com letras gregas minúsculas.
$\hat{}$	O chapéu é usado para indicar que uma variável é estimada.
$\overline{\mathbf{W}}, \overline{\mathbf{d}}$	A barra superior indica a média de uma matriz ou um vetor.
T	Indica que a matriz ou vetor está transposto.

Símbolos

P	Ponto no espaço tridimensional.
$[X \ Y \ Z]^T$	Coordenadas do ponto P .
λ	Constante de modelagem da câmera.
$\mathbf{M}$	Matriz de modelagem da câmera.

$\mathbb{R}^*$	Conjunto dos números reais, excluindo o número 0.
O	Centro focal da câmera.
z	Eixo focal da câmera.
f	Distância focal da câmera.
Π	Plano de imagem.
p	Ponto no plano de imagem. Projeção de P .
$[u \ v]^T$	Coordenadas da projeção p em Π .
b	Tamanho da linha de base do sistema estéreo.
O_e	Centro focal da câmera esquerda.
z_e	Eixo focal da câmera esquerda.
f_e	Distância focal da câmera esquerda.
Π_e	Plano de imagem esquerdo.
p_e	Projeção esquerda do ponto P .
$[u_e \ v_e]^T$	Coordenadas da projeção p_e .
$\mathbf{c}_e$	Vetor de coordenadas de p_e .
O_d	Centro focal da câmera direita.
z_d	Eixo focal da câmera direita.
f_d	Distância focal da câmera direita.
Π_d	Plano de imagem direito.
p_d	Projeção direita do ponto P .
$[u_d \ v_d]^T$	Coordenadas da projeção p_d .
$\mathbf{c}_d$	Vetor de coordenadas de p_d .
$\mathbf{d}$	Vetor de disparidades.
d_r	Valor da disparidade na posição r .
$[d_u \ d_v]^T$	Disparidade vertical e horizontal.
Ω	Plano epipolar.
H_Π	Transformação entre os planos de imagem.
$\mathbf{l}_e$	Linha epipolar esquerda.
e_e	Epipolo esquerdo.
$\mathbf{l}_d$	Linha epipolar direita.
e_d	Epipolo direito.

$[e_d]_{\times}$	Matriz anti-simétrica.
$[e_d^x \ e_d^y \ e_d^z]$	Coordenadas do epipolo direito.
$p_e^{i,j}$	Projeção esquerda nas coordenadas $[i \ j]^T$.
$p_d^{i,j}$	Projeção direita nas coordenadas $[i \ j]^T$.
$disp()$	Função de disparidade.
$\mathbf{I}_r$	Imagem de referência.
$\mathbf{I}_a$	Imagem alvo.
$\mathbf{I}_r(y, x)$	<i>Pixel</i> da imagem de referência na linha y e coluna x .
n_v	Número de vizinhos.
$\mathbf{W}_r$	Janela de vizinhança da imagem de referência.
$\mathbf{W}_a$	Janela de vizinhança da imagem alvo.
EDI	Espaço de disparidades da imagem.
$\mathbf{EDI}(y, x, d_r)$	Custo de similaridade do <i>pixel</i> com coordenadas $[x \ y]^T$ para a disparidade d_r .
$\mathbf{EDI}(y, x)$	Vetor de custos de similaridade para o <i>pixel</i> com coordenadas $[x \ y]^T$.
$[d_{min}, d_{max}]$	Intervalo de disparidades.
<i>Similaridade</i>	Função genérica de similaridade entre duas regiões.
D	Mapa de disparidades.
$\hat{\mathbf{D}}$	Mapa de disparidades estimado.
E	Função de energia.
E_{smooth}	Função de suavidade.
$\mathbf{c}_p$	Cor do <i>pixel</i> p .
$\mathbf{c}_q$	Cor do <i>pixel</i> q .
$\Delta_{c_{pq}}$	Distância entre as cores $\mathbf{c}_p$ e $\mathbf{c}_q$.
γ_s	Constante de agrupamento por similaridade.
f_s	Força de agrupamento perceptual por similaridade de cor.
Δd_{pq}	Distância espacial entre os <i>pixels</i> p e q .
γ_p	Constante de agrupamento por proximidade espacial.
f_p	Força de agrupamento perceptual por proximidade espacial.
f_t	Força de agrupamento perceptual total.

$\mathbf{W}_s$	Matriz suporte.
R	Número de linhas da imagem.
C	Número de colunas da imagem.
N	Quantidade de níveis da pirâmide.
$\mathbf{I}_n$	Imagem no nível n da pirâmide.
$\mathbf{W}_c$	Matriz de convolução.
t_d	Quantidade de disparidades considerada.
N_r	Número de requisições à função de similaridade.
$\mathbf{W}_e$	Matriz das diferenças absolutas.
B	Porcentagem de disparidades erradas.

Sumário

Resumo	ix
Abstract	xi
Lista de Figuras	xiii
Lista de Tabelas	xvii
Lista de Abreviaturas e Siglas	xix
Lista de Símbolos	xxi
1 Introdução	1
1.1 Motivação	2
1.2 Objetivo do trabalho	5
1.3 Organização do trabalho	5
2 Princípios Básicos de Visão Estéreo	7
2.1 Considerações iniciais	7
2.2 Modelo Geométrico de uma Câmera	8
2.3 Geometria Estéreo Binocular	9
2.4 Aspectos iniciais de visão estéreo	11
2.4.1 Calibração das câmeras	11
2.4.2 Correspondência	12
2.4.3 Reconstrução	15
2.5 Considerações finais	16
3 Problema da Correspondência	17
3.1 Considerações iniciais	17
3.2 Restrições Estéreo	18
3.2.1 Geometria epipolar e retificação das imagens	18

3.2.2	Unicidade	22
3.2.3	Ordenação	23
3.2.4	Simetria	24
3.3	Considerações finais	25
4	Métodos de Computação Estéreo e suas Classificações	27
4.1	Considerações iniciais	27
4.2	Modelo genérico de computação estereo	27
4.3	Classificação dos métodos estereo	30
4.3.1	Métodos Locais de Correspondência	32
4.3.2	Métodos Globais de Correspondência	35
4.4	Considerações finais	37
5	Trabalhos Relacionados	39
5.1	Considerações iniciais	39
5.2	Revisão de técnicas relacionadas	39
5.3	Considerações finais	42
6	Algoritmo de Correspondência	43
6.1	Considerações iniciais	43
6.2	Metodologia	43
6.3	<i>Gestalt Grouping</i>	45
6.3.1	Agrupamento por similaridade	46
6.3.2	Agrupamento por proximidade	47
6.3.3	Suporte de agrupamento	49
6.4	Pirâmides de Imagens	50
6.5	Vantagens de pirâmides de imagens no problema da correspondência	53
6.5.1	Ambigüidades	53
6.5.2	Custo computacional	54
6.6	Descrição do algoritmo	55
6.7	Considerações finais	59
7	Resultados	61
7.1	Considerações iniciais	61
7.2	Base de dados utilizada	61
7.3	Considerações sobre a avaliação dos resultados	64
7.4	Apresentação dos resultados	65

7.5	Discussão sobre os parâmetros do algoritmo	70
7.6	Considerações finais	77
8	Conclusões	79
9	Apêndice A - Resultados Detalhados	81
	Referências Bibliográficas	85

Capítulo 1

Introdução

Desde o início de sua existência, os seres humanos vêm buscando cada vez mais construir ferramentas que tornam sua vida mais fácil e, apesar disso, até hoje não existe nenhuma máquina que possa substituir o homem definitivamente (Shapiro e Stockman, 2001). Isso porque, com o passar do tempo, o homem desenvolveu habilidades tão complexamente eficientes que extrapolam os próprios limites de sua compreensão. Desde a publicação das idéias de Charles Darwin, pode-se entender que a interação e o conhecimento do meio ambiente permitiu a evolução de organismos simples até os mais complexos, os quais desenvolveram técnicas muito bem elaboradas que parecem ser ao mesmo tempo triviais e enigmáticas quando analisadas mais de perto.

A compreensão total de como os seres interagem com o ambiente e como conseguem aprender com experiências, sendo essas vivenciadas ou não, ainda está muito longe de ser atingida. Sobre um aspecto natural onde a capacidade de se locomover com segurança, detectar alvos, estabelecer melhores caminhos, reconhecer objetos, entre outros, são essenciais para o conhecimento e a sobrevivência no meio, os sentidos biológicos passam a ter fundamental importância no que diz respeito à observação e interação com o mundo.

Seguindo essa linha de raciocínio, fica fácil de entender a razão do crescente interesse em pesquisas sobre a melhor compreensão dos sentidos biológicos, em especial o sentido da visão que consegue fornecer grande quantidade de informações sobre o ambiente. Vários sistemas de visão artificial são desenvolvidos em todo o mundo com a expectativa não somente de resolver problemas específicos, mas também de contribuir com o esclarecimento de como o sistema visual natural pode funcionar.

A área de Visão Computacional é uma grande promessa de que é possível facilitar ainda mais

o dia-a-dia das pessoas e resolver problemas antes nunca imagináveis. Mas, como em qualquer outro ramo da ciência, o amadurecimento dessa área exige um extenso tempo de pesquisa e muitos obstáculos a serem superados. Contudo, avanços consideráveis já foram atingidos, mas ainda há muito o que desvendar (Brown et al., 2003).

1.1 Motivação

Dentre as várias aplicações de Visão Computacional, o uso de imagens bidimensionais são suficientes para produzir resultados plausíveis. Entretanto, em diversas aplicações, faz-se necessário o conhecimento de informações adicionais, tais como a distância de objetos, geometria, formas, orientação de superfícies ou a completa disposição tridimensional da cena. Informações como essas não podem ser recuperadas facilmente de uma simples imagem bidimensional. Pesquisas em navegação autônoma, fotogrametria, detecção de obstáculos, reconstrução e renderização de objetos tridimensionais são exemplos de problemas que exigem tais informações adicionais.

Existe vários métodos de determinação das características 3D de uma cena, tais como os que utilizam dispositivos ultra-sônicos, sensores de profundidade a *laser*, métodos fotométricos de iluminação controlada e também dispositivos multi-oculares, que usam várias câmeras para a captura de imagens da cena. Por outro lado, cada método apresenta um conjunto de vantagens e uma vasta lista de problemas relacionados, em especial o custo e a qualidade das informações obtidas, como o caso dos dispositivos ultra-sônicos que, além de ter elevado custo de aquisição, conseguem fornecer apenas informações esparsas sobre o ambiente e apresentam problemas com os ângulos de deflexão. Os métodos que envolvem *laser* são invasivos, ou seja, apresentam risco às pessoas que possam estar na cena, e não trabalham muito bem com superfícies brilhantes e reflexivas. Já as técnicas fotométricas necessitam de um conhecimento detalhado sobre a iluminação e sobre as propriedades reflexivas dos objetos que compõem a cena. Por último, os sistemas multi-oculares são de baixo custo, não invasivos, fazem uso de sensores passivos e não necessitam de conhecimento prévio das características da cena. Porém, apesar desses aspectos positivos, trabalhar com dispositivos multi-oculares significa enfrentar algumas dificuldades já bem conhecidas na área de Visão Computacional, tais como lidar com ambigüidades, ruídos, problemas de iluminação, etc.

A área da Visão Computacional relacionada à identificação das informações tridimensionais de uma cena, bem como a superação dos problemas relacionados, é conhecida como Visão Tridi-

mensional. Quando presente mais de um dispositivo de captura de imagem é chamada de Visão Estéreo ou Visão Estereoscópica (Hartley e Zisserman, 2003). Ela busca inferir as informações tridimensionais utilizando imagens bidimensionais de uma cena capturadas a partir de diferentes pontos de vista (Trucco e Verri, 1998).

Inspirada no sistema de visão humana, a visão estereo é uma das áreas de Visão Computacional mais pesquisada nas últimas décadas. Os primeiros trabalhos sobre esse tema surgiram por volta dos anos 70 e foram introduzidos pela *Image Understanding Community* (IU) fundada pela *Advanced Research Projects Agency* (ARPA) (Barnard e Fischler, 1982; Brown et al., 2003; Thompson, 1975). Desse período até hoje, pesquisas nessa área amadureceram consideravelmente, muitos avanços foram feitos quanto à solução de parte dos problemas apresentados em visão estereo e até hoje surgem algoritmos inéditos. Apesar dessa intensa investigação, extrair informações tridimensionais de imagens estereoscópicas ainda não é uma tarefa fácil. Alguns problemas clássicos como iluminação, ambigüidade, deformação e oclusão de objetos tornam o problema estereo um problema mal condicionado e de difícil solução (Trucco e Verri, 1998).

A Figura 1.1 apresenta um esquema genérico de visão estereo. Ele é formado basicamente por duas ou mais câmeras colocadas em posições diferentes, que capturam imagens da cena no mesmo instante de tempo.

Figura 1.1: Esquema geral de um sistema de visão estereo.

A configuração ilustrada pela Figura 1.1 permite que os pontos da cena, no caso ideal, possuam duas representações, uma em cada câmera. Dessa forma, as câmeras 1 e 2 coletam imagens com diferenças provocadas pela mudança de perspectiva. Essas imagens são analisadas por um programa de computador que relaciona *pixels* que representam os mesmos pontos físicos na cena. As diferenças de localização desses *pixels* são chamadas de disparidades e formam o mapa de disparidades da cena, sendo este disposto como uma imagem bidimensional, onde cada *pixel* guarda a diferença entre as projeções de um mesmo ponto no ambiente tridimensional.

Existe três problemas fundamentais em Visão Estéreo (Brown et al., 2003): a calibração de câmeras, que se preocupa em determinar as características geométricas do sistema de aquisição de imagens, a correspondência de características, que consiste em relacionar as projeções de um mesmo ponto físico pertencente ao ambiente tridimensional e por último a reconstrução da cena e determinação de modelos geométricos encontrados.

O segundo item mencionado pelo parágrafo anterior constitui um dos problemas mais investigados em Visão Estéreo. A determinação de correspondências é influenciada por diversos fatores estruturais e de luminosidade no mundo tridimensional, o que a torna um dos problemas mais difíceis de se resolver dentro de todo o contexto de visão estereo.

Devido à grande complexidade presente na correspondência de características, vários trabalhos propuseram-se a analisar o desempenho dos algoritmos existentes. É possível fazer um acompanhamento histórico da evolução de visão estereo analisando alguns artigos que tratam da revisão, crítica e taxonomia das técnicas existentes. No trabalho de Barnard e Fischler (1982) por exemplo, os autores fazem uma análise geral sobre visão estereo enfatizando algumas etapas indispensáveis no processamento das imagens e além disso, eles introduzem algumas métricas como critério de avaliação dos algoritmos da época. Outros trabalhos semelhantes surgiram à medida que os algoritmos estereo evoluíram, diferenciando-se pela revisão de problemas na análise de imagens estereoscópicas. Em Dhond e Aggarwal (1989) os autores fazem uma análise das técnicas que surgiram na década e revisam o conceito de multi-resolução aplicada à visão estereo. O trabalho de Koschan (1993b) oferece um tutorial sobre esta área, revisando cerca de oitenta e seis artigos relacionados. De uma maneira geral, de década em década surge um trabalho com o objetivo de inspecionar os algoritmos publicados. Críticas mais recentes podem ser encontradas em Brown et al. (2003) e Scharstein e Szeliski (2002) enfatizando a taxonomia dos métodos existentes. A leitura desses cinco trabalhos citados anteriormente é recomendada para qualquer pesquisa relacionada à visão estereo.

1.2 Objetivo do trabalho

A principal dificuldade da maioria das pesquisas na área de visão estéreo é a construção de algoritmos que unam menor esforço computacional e a fidelidade dos resultados com a configuração tridimensional real da cena. Técnicas locais têm a vantagem de produzir resultados rápidos mas deixam a desejar quanto sua veracidade. Normalmente, devido a existência de ruídos, distorções, oclusões e ambigüidades, as disparidades não condizem com a realidade da cena observada. Como alternativa a isso, alguns autores empregam técnicas globais com o objetivo de evitar a influência dessas dificuldades. Por outro lado, as técnicas globais exigem maior tempo de processamento, inviabilizando aplicações que necessitam de menor tempo de resposta.

O principal objetivo desse trabalho é elaborar um algoritmo que una as vantagens das técnicas locais e globais na construção de mapas densos de disparidades. A metodologia adotada funde duas estratégias conhecidas na área de visão computacional e na psicologia: a análise multi-resolução e a teoria *Gestalt* de agrupamento perceptual (Chang et al., 2007; Forsyth e Ponce, 2002). A união dessas duas técnicas permite atenuar os problemas envolvidos na construção de mapas de disparidade que normalmente são de alta complexidade, contribuindo para uma implementação mais simples que a de vários trabalhos da área.

O algoritmo foi elaborado utilizando algumas técnicas já bem estabelecidas na área de visão estéreo, diferenciando-se pela aplicação de um raciocínio baseado em agrupamento perceptual na determinação dos *pixels* que possuem a mesma disparidade. Além disso, é investigada uma abordagem usando pirâmides de imagens para a redução de ambigüidades e computações desnecessárias.

Considerando a taxonomia da área, esse trabalho pode ser classificado no grupo dos métodos locais de correspondência. Essa taxonomia bem como os termos e os pontos chave para a compreensão de visão estéreo são discutidos nos Capítulos posteriores.

1.3 Organização do trabalho

A organização dos tópicos dessa dissertação visa primeiramente o embasamento teórico sobre a área de visão estéreo seguido pela descrição do trabalho proposto e a apresentação dos resultados.

No Segundo Capítulo é apresentada uma visão geral sobre a computação estéreo, incluindo aspectos geométricos, e também uma breve introdução do que é um sistema estéreo e os conceitos

da geometria de uma câmera, e de um sistema estéreo binocular. Ainda no mesmo Capítulo, é apresentada uma visão geral de como é construído um sistema estereoscópico, e quais os principais módulos que o compõe.

Como complemento, no Terceiro Capítulo, são discutidos os problemas clássicos encontrados em visão estéreo, como a dificuldade de se lidar com oclusões e ambigüidades contidas nas imagens. São mencionadas as restrições e suposições feitas sobre as cenas em busca da simplificação do problema da correspondência.

O Quarto Capítulo permite a compreensão de como é implementado um simples algoritmo de correspondência estéreo. Nele também é apresentada a estrutura geral e a classificação dos métodos existentes. O Quinto Capítulo é dedicado exclusivamente à explanação de alguns trabalhos relacionados a essa dissertação, dando maior interesse às técnicas locais de correspondência. No Sexto Capítulo encontra-se a teoria suficiente para o entendimento da proposta do trabalho, e também a estrutura do algoritmo desenvolvido.

No Sétimo Capítulo são apresentados os resultados alcançados. Nele é descrito a metodologia de avaliação utilizada para a comparação com trabalhos relacionados. Ainda nesse Capítulo, é realizada uma discussão sobre os parâmetros do algoritmo e seus efeitos sobre o resultado final.

Por fim, o Capítulo Oito apresenta as conclusões da dissertação, seguido do Capítulo Nove, um Apêndice que contém resultados detalhados obtidos nesse trabalho.

Capítulo 2

Princípios Básicos de Visão Estéreo

2.1 Considerações iniciais

Visão estereo emprega a análise de várias imagens da mesma cena obtidas a partir de locais diferentes. A noção da terceira dimensão é calculada considerando as diferenças entre essas imagens. Normalmente, os sistemas artificiais de visão estereo baseiam-se nos princípios biológicos da visão estereo, e um bom exemplo é o sistema visual humano. Cada pessoa possui dois olhos alinhados e separados por uma distância fixa, o que permite a visualização de uma cena a partir de dois pontos de vista diferentes. A posição de um objeto situado na imagem observada pelo olho esquerdo é ligeiramente diferente da posição do mesmo objeto na imagem capturada pelo olho direito. Essa diferença varia de acordo com o afastamento ou aproximação do objeto em relação ao observador.

Análogo ao sistema natural, um sistema artificial estereo utiliza câmeras de aquisição, sendo essas localizadas em posições distintas e conhecidas. Cada câmera captura uma imagem da cena no mesmo instante de tempo. Dessa maneira pode-se estimar a distância relativa ao sistema estereo de um determinado ponto no mundo tridimensional que pode ser observado a partir de pelo menos duas posições diferentes. As seções seguintes introduzem conceitos fundamentais para o entendimento geométrico da visão estereo.

2.2 Modelo Geométrico de uma Câmera

De uma maneira geral, uma câmera é constituída por um plano de imagem chamado plano Π , sendo esse perpendicular a seu eixo focal z , como ilustrado na Figura 2.1. Em relação ao mundo tridimensional, uma câmera possui como referência seu próprio centro óptico O . A distância focal f é a distância entre o plano de imagem e o centro óptico da câmera.

Ainda na Figura 2.1, um ponto P localizado no mundo tridimensional, cujas coordenadas são: $[X \ Y \ Z]^T$, projeta sua imagem na intersecção da reta PO com o plano de imagem Π como um ponto p associado às coordenadas $[u \ v]^T$. A projeção de P é dada pela seguinte equação (Chambon, 2005):

$$\lambda \cdot \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{M} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (2.1)$$

Onde $\lambda \in \mathbb{R}^*$ e $\mathbf{M}$ é a matriz que contém os parâmetros de modelagem da câmera. Essa matriz é estimada através da calibração da câmera (ver Seção 2.4.1).

Figura 2.1: Modelo geométrico de uma câmera.

O modelo exemplificado pela Figura 2.1 é conhecido como modelo de uma câmera *pinhole* ou "buraco-de-alfinete" (Hartley e Zisserman, 2003). Para um sistema composto por uma única câmera, ou seja, mono-ocular, a distância em que o ponto P se encontra não interfere na localização de sua projeção, ou seja, P pode ocupar qualquer posição na reta PO e p sempre terá as mesmas coordenadas, o que faz das projeções de P e P^1 equivalentes.

2.3 Geometria Estéreo Binocular

Os sistemas multi-oculares incorporam várias câmeras para o mapeamento da cena. Idealmente, para a reconstrução perfeita de uma cena, as câmeras devem cobrir todos os ângulos possíveis de visualização do meio, o que é naturalmente inviável. Quanto maior o número de pontos de visualização, maior pode ser o nível de detalhamento do ambiente, mas, por outro lado, maior é a complexidade envolvida no sistema.

De uma forma geral, nem sempre é necessário o mapeamento tridimensional de todo o ambiente e, por motivos de simplificação, a configuração mais difundida usa somente duas câmeras de aquisição, sendo esta chamada de sistema estéreo binocular (Hartley e Zisserman, 2003). Esse sistema possui duas câmeras, seus eixos ópticos z_e e z_d são convergentes e os centros focais O_e e O_d são espaçados por uma distância de base b . A Figura 2.2 ilustra um sistema desse tipo.

As câmeras são representadas pelo modelo *pinhole*, cada uma com seus eixos de coordenadas. Em um sistema binocular é comum nomear os planos de imagem como plano esquerdo, Π_e , e direito, Π_d . Por hora, será usada essa nomenclatura para facilitar a compreensão dos sistemas desse tipo.

Na Figura 2.2, o ponto P possui duas projeções, p_e e p_d e, diferentemente da configuração mono-ocular, qualquer mudança de posição do ponto P representa uma alteração das coordenadas de suas projeções em pelo menos um plano de imagem. Dessa forma P e P^1 possuem a mesma projeção p_e no plano de imagem Π_e , enquanto que no plano Π_d , eles possuem projeções diferentes p_d e p_d^1 .

A diferença entre as coordenadas das projeções de um mesmo ponto é conhecida como *disparidade*. Sendo $\mathbf{c}_e = [u_e \ v_e]^T$ e $\mathbf{c}_d = [u_d \ v_d]^T$ as coordenadas das projeções esquerda e direita respectivamente, a disparidade é dada pela simples diferença $\mathbf{d}(p_e) = [d_u \ d_v]^T = \mathbf{c}_e - \mathbf{c}_d$. O conjunto de disparidades referente a todos os *pixels* de um plano de imagem é chamado de *mapa de disparidades* (Scharstein e Szeliski, 2002).

Figura 2.2: Geometria estereó binocular.

O termo conhecido como *profundidade* é o significado físico do valor de disparidade, ou seja, a distância real que o ponto se encontra no mundo 3D tomando como referência o centro da linha de base do sistema estereó. A disparidade é inversamente proporcional à profundidade e quanto mais próximo o ponto P estiver do sistema de observação, maior a diferença de posicionamento de suas projeções.

Na verdade, nem sempre é possível calcular a disparidade de todos os pontos contidos na cena. Pelo fato das câmeras estarem em posições diferentes, a perspectiva de observação muda e, conseqüentemente, alguns pontos são projetados em apenas uma das câmeras, o que torna impossível o cálculo direto de suas disparidades. Dessa forma, cada plano de imagem contém um conjunto de pontos que não possuem projeções na outra câmera. Esses pontos físicos onde não é possível calcular suas disparidades são ditos *pontos meio oclusos* e o conjunto desses pontos correspondem ao *mapa de oclusão* do plano de imagem em questão (Scharstein e Szeliski, 2002).

2.4 Aspectos iniciais de visão estéreo

O princípio básico da computação estéreo consiste em, dado duas imagens da mesma cena, caso seja possível localizar regiões nas imagens que correspondem ao mesmo ponto físico, então é possível determinar a localização tridimensional desse ponto (Brown et al., 2003).

Barnard e Fischler (1982) definiram seis passos fundamentais para a computação estéreo: aquisição das imagens, modelagem das câmeras, aquisição de características, casamento das imagens, determinação das profundidades e interpolação. Mas, de uma forma mais abrangente, existem três problemas básicos no que diz respeito à computação estéreo.

1. Calibração das câmeras.
2. Correspondência de características.
3. Reconstrução.

Cada uma dessas etapas pode ser considerada como uma ciência à parte dentro do contexto de visão estéreo. Como esse trabalho tem como objetivo principal a criação de mapas de disparidades, o item que é discutido com maior riqueza de detalhes é relacionado à correspondência de características. Quanto aos outros dois itens, apesar de serem de suma importância para toda a computação estéreo, são apresentados em uma breve introdução conceitual. Para suprir esses outros assuntos, são citados alguns trabalhos importantes da área para uma futura consulta caso seja de interesse do leitor.

2.4.1 Calibração das câmeras

O cálculo das profundidades é inteiramente dependente dos parâmetros geométricos do sistema estéreo. Os valores de f , b , O e o centro real das imagens nem sempre são conhecidos com a precisão necessária. Em aplicações como renderização e reconstrução de objetos, onde valores quantitativos são cruciais, essa etapa é de fundamental importância e a calibração das câmeras provê o conhecimento detalhado sobre toda a geometria do sistema estéreo.

A idéia básica por trás desse processo é encontrar as equações de projeções das câmeras relacionando um conjunto de pontos tridimensionais conhecidos com suas respectivas projeções. Para isso são usadas imagens chamadas de *padrões de calibração*. Elas contêm as projeções de objetos geométricos com dimensões e localização definidas, normalmente munidas de características de

fácil identificação (cantos, bordas). Uma vez calculada a posição desses pontos, é possível determinar os parâmetros extrínsecos (orientação e localização das câmeras) e intrínsecos (distância focal, centro focal, distorções das lentes) do sistema estéreo.

Quanto mais fina for a calibração, maior a precisão durante a conversão dos valores de disparidades (distância entre *pixels*) para medidas reais de profundidade da cena. A teoria necessária para compreender como é realizada a calibração de câmeras pode ser encontrada em Forsyth e Ponce (2002) e Trucco e Verri (1998).

2.4.2 Correspondência

A etapa de correspondência é considerada a mais problemática de toda a computação estéreo (Zhou et al., 2003). Capturando duas imagens, como ilustrado na Figura 2.2, e definindo uma delas como sendo a imagem de referência e a outra como imagem alvo, o processo de correspondência consiste em encontrar para cada *pixel* da imagem de referência o *pixel* correspondente ao mesmo ponto físico dentro da imagem alvo. Esse processo é conhecido como o *Problema da Correspondência* (Scharstein e Szeliski, 2002).

Essa etapa de processamento é computacionalmente custosa. Para uma configuração de câmeras não calibradas, o *pixel* de referência pode estar em qualquer região da imagem alvo ou simplesmente estar ocluso. É praticamente inviável fazer a varredura completa da imagem alvo para cada *pixel* da imagem de referência (Scharstein e Szeliski, 2002).

Teoricamente, a correspondência de pontos pode ser definida por uma simples comparação entre os níveis de cinza das imagens. Um mesmo ponto na cena deveria projetar sua imagem da mesma maneira em ambas as câmeras, mas essa é uma consideração falsa. Tanto pela mudança de perspectiva quanto pelas variações de luminosidade, os objetos estão sujeitos a deformações, alterações de cor e oclusões. Mesmo assim, por mais que as projeções possam ser influenciadas, as projeções correspondentes tendem a ser o mais semelhante possível.

Considerando somente o nível de cinza do *pixel* de referência, a busca pelo *pixel* correspondente na imagem alvo gera muita ambigüidade. Apesar de uma imagem fornecer toda a informação presente em uma cena, a análise individual de cada *pixel* não é a melhor estratégia. Como uma imagem chega a possuir muito mais que centenas de *pixels*, é muito pouco provável que exista somente um *pixel* na imagem alvo com o mesmo nível de cinza do *pixel* da imagem de referência. Faz sentido usar, além do *pixel* de interesse, uma certa porção de seus vizinhos,

supondo que boa parte deles seja o mais semelhante possível aos vizinhos do *pixel* alvo.

Essa consideração transforma a busca de um *pixel* em uma busca por regiões da imagem. É fácil observar que, partindo dessa abordagem, o número de falsas correspondências diminui. Em contrapartida, a probabilidade de considerar *pixels* oclusos aumenta com o tamanho da janela de vizinhança, o que dificulta bastante todo o processo.

A comparação entre os *pixels* pode ser feita de várias formas e varia desde uma simples diferença entre suas intensidades, até técnicas complexas de otimização de funções de energia. A metodologia mais utilizada baseia-se na Correlação Cruzada Zero Normalizada (CCZN) devido à sua robustez no tratamento da influência de diferentes luminosidades (Sutton et al., 2000). Esse processo de comparação é melhor discutido no Capítulo 3.

A Figura 2.3 mostra um simples exemplo de ambigüidade. Mesmo considerando a vizinhança do *pixel*, não é possível afirmar qual o verdadeiro canto correspondente ao canto do cubo da imagem esquerda sem uma análise global da cena.

Figura 2.3: Ambigüidade de regiões em imagens estéreo.

A Figura 2.4 ilustra o problema de oclusão em imagens estereoscópicas. O *pixel* de interesse está localizado no centro do quadrado na Figura 2.4(a) e seu correspondente no centro do quadrado na Figura 2.4(b).

Observando somente a área delimitada pelos quadrados observa-se que a vizinhança dos *pixels* correspondentes nas duas imagens são muito diferentes. Na Figura 2.4(a) a região selecionada considera muitos pixels de fundo que estão oclusos na outra imagem. Apesar dos pontos serem correspondentes ao mesmo ponto físico, a informação disponível por uma boa parcela dos níveis de cinza destacados não convergem para essa afirmação.

Figura 2.4: Exemplo de regiões oclusas em imagens estéreo.

Outro problema a ser considerado é a própria disposição dos padrões presentes na imagem. A quantidade de ambigüidade é aumentada em regiões onde existe repetição ou falta de textura, como na Figura 2.5. Além disso, existem também problemas relacionados à especularidade das superfícies e à forma dos objetos, que podem ser projetados de maneiras diferentes nas câmeras. Devido a essa grande complexidade, é exatamente nessa área que se concentram a maior parte das pesquisas relacionadas à Visão Estéreo.

Figura 2.5: Repetição de padrões de textura.

Existem várias maneiras de se calcular um mapa de disparidades (Brown et al., 2003), e de uma forma bem ampla, a computação estéreo tem dois objetivos principais: calcular as disparidades com a maior precisão possível e criar um mapa denso de disparidades (Brown et al., 2003).

A união dessas duas características é um dos maiores desafios encarados pelas pesquisas mais recentes nessa área. Além disso, atualmente há uma grande preocupação com o fator de desempenho dos algoritmos na solução das disparidades estéreo.

Visando resultados em tempo real, alguns autores como Hariyama et al. (2005) e Stefano et al. (2004b) sugerem a implementação em *hardware* desses métodos, mas o nível de complexidade de

alguns algoritmos acaba exigindo muita capacidade de processamento ou torna muito complicada sua implementação.

Outros trabalhos como Chung (1997), Stefano et al. (2004a) e Sun (1997) propõem métodos rápidos com uma abordagem bastante simplificada e de fácil implementação prática, mas sem grande preocupação com o grau de precisão das profundidades. Um estudo comparativo desses algoritmos de tempo real foi realizado por Sunyoto et al. (2004), discutindo as potencialidades dos algoritmos rápidos de visão estéreo aplicados à navegação robótica.

Sobre outra perspectiva, Bleyer e Gelautz (2007) e Lee et al. (2005) já apresentam soluções mais robustas, mas com um alto nível de complexidade em seus algoritmos. De fato, um método genérico, que une velocidade de processamento e a construção de um mapa de disparidade preciso ainda não foi alcançado (Brown et al., 2003).

2.4.3 Reconstrução

Com o mapa de disparidade devidamente calculado, a próxima fase trata da reconstrução da cena. Essa etapa preocupa-se em determinar a estrutura tridimensional através do mapa de disparidade baseado na geometria do sistema estéreo (Seitz et al., 2006).

Na figura 2.2 as projeções do ponto P nas coordenadas $[X \ Y \ Z]^T$ são representadas por p_e e p_d nas coordenadas $[u_e \ v_e]^T$ e $[u_d \ v_d]^T$, respectivamente. Usando um simples cálculo geométrico, pode-se chegar às equações das coordenadas do ponto. Através de triangularização pode-se escrever que (Ahmeda, 2006):

$$\frac{v_e}{f} = \frac{x + \frac{b}{2}}{z_e} \quad (2.2)$$

$$\frac{v_d}{f} = \frac{x - \frac{b}{2}}{z_e} \quad (2.3)$$

$$\frac{u_e}{f} = \frac{u_d}{f} = \frac{y}{z} \quad (2.4)$$

$$\frac{v_e - v_d}{f} = \frac{b}{z} \quad (2.5)$$

Rearranjando as equações anteriores as coordenadas X , Y e Z podem ser escritas como:

$$X = b \cdot \frac{(v_e + v_d)/2}{v_e - v_d} \quad (2.6)$$

$$Y = b \cdot \frac{(v_e - v_d)/2}{v_e - v_d} \quad (2.7)$$

$$Z = b \cdot \frac{f}{v_e - v_d} \quad (2.8)$$

As coordenadas X , Y e Z têm como referência um sistema de coordenadas com eixos paralelos aos eixos dos planos de imagem com origem no centro da linha de base b .

2.5 Considerações finais

Tanto a calibração das câmeras, a etapa de correspondência e o módulo de reconstrução, constituem cada um deles uma ciência a parte. No entanto, a correspondência é considerada a fase mais complicada de toda a computação estéreo. Diante disso, nesse trabalho são utilizadas imagens previamente retificadas, para a construção dos mapas de disparidade. A etapa de reconstrução não é abordada, já que as estruturas geométricas dos objetos da cena não estão postos em discussão, mas somente a disposição geral da mesma.

Capítulo 3

Problema da Correspondência

3.1 Considerações iniciais

Como mencionado anteriormente, a correspondência estéreo tem como objetivo encontrar as projeções que correspondem ao mesmo ponto físico da cena. O problema da correspondência resalta as várias dificuldades encontradas nesse processo. Infelizmente, a mudança da perspectiva de visualização favorece o surgimento de vários problemas que não são trivialmente tratados, tais como distorções, oclusões e ambigüidades. Sem sombra de dúvidas esses são os principais obstáculos a serem encarados por qualquer aplicação estéreo. Diante disso, é comum usar algumas restrições e suposições sobre a cena a fim de atenuar esses problemas.

Em busca de uma simplificação do algoritmo de correspondência, normalmente pressupõem-se que a cena seja composta por estruturas mais próximas à superfícies lambertianas, ou seja, que refletem a luz incidente da mesma forma para todas as direções (difusor ideal). Além disso, a disparidade de um determinado *pixel* tem grande chance de ser igual à disparidade de sua vizinhança, o que permite afirmar que as disparidades se alteram suavemente, com exceção nas bordas dos objetos (Boykov et al., 1998; Kanade e Okutomi, 1994).

Na verdade, essas considerações apresentam somente uma direção para o desenvolvimento de algoritmos de correspondência, e algumas flutuações desses parâmetros são perfeitamente aceitáveis. Inevitavelmente, a única informação a ser considerada em uma imagem é a intensidade dos *pixels* que a compõem. Sendo esses parâmetros influenciados ou não, cabe aos métodos suportarem as variações dessas intensidades de forma a maximizar as correspondências verdadeiras.

3.2 Restrições Estéreo

3.2.1 Geometria epipolar e retificação das imagens

Realizar a busca por correspondências em toda a imagem alvo para cada *pixel* da imagem de referência não é a tarefa mais inteligente. Esse processo é custoso computacionalmente e torna-se inviável para aplicações que exigem alto desempenho. Voltando à definição do problema da correspondência, dadas as projeções p_e e p_d de um ponto P no mundo tridimensional, qual seria a relação entre eles?

A geometria epipolar busca relacionar esses pontos, de maneira a diminuir o espaço de busca por correspondências. Supondo um ponto P no espaço 3D, com projeções p_e e p_d em um sistema binocular estéreo, como ilustrado na Figura 3.1. Os pontos P , p_e , p_d , O_e e O_d definem o plano Ω chamado de plano epipolar. As projeções e_e e e_d são conhecidos como epipolos das câmeras, daí a origem do termo "epipolar". Esses pontos nada mais são que as projeções dos centros focais das câmeras umas nas outras. Pode-se perceber facilmente que as linhas l_e e l_d , resultantes da intersecção entre os três planos, são coplanares a Ω . Essa propriedade é de grande importância na busca por correspondências. Supondo agora que somente p_e seja conhecido, o mesmo plano Ω pode ser encontrado considerando apenas p_e , O_e e a linha de base b . A intersecção entre Ω e Π_d define a linha epipolar l_d . Dessa forma, pelo fato de l_e e l_d serem coplanares, pode-se afirmar que p_d só poderá existir na região definida pela linha l_d e não é necessária a busca por correspondências em toda a imagem (Zhang et al., 2004).

Figura 3.1: Geometria epipolar.

O ponto chave para a determinação dos pares de linhas epipolares é o conhecimento dos epipolos das câmeras. Dado uma projeção p_d , a linha epipolar l_d que passa pelos pontos e_d e p_d pode ser escrita como

$$\mathbf{l}_d = e_d \cdot p_d = [e_d]_{\times} \cdot p_d \quad (3.1)$$

Desde que p_d seja escrito como $p_d = H_{\Pi} \cdot p_e$ tem-se:

$$\mathbf{l}_d = [e_d]_{\times} \cdot H_{\Pi} \cdot p_e = \mathbf{F} \cdot p_e \quad (3.2)$$

sendo H_{Π} a transformação que leva p_e a p_d e a notação $[e_d]_{\times}$ equivalente a uma matriz anti-simétrica do tipo

$$[e_d]_{\times} = \begin{bmatrix} 0 & -e_d^z & e_d^y \\ e_d^z & 0 & -e_d^x \\ -e_d^y & e_d^x & 0 \end{bmatrix} \quad (3.3)$$

onde $[e_d^x \ e_d^y \ e_d^z]^T$ corresponde às coordenadas do epipolo e_d no mundo 3D. A matriz $\mathbf{F}$ é chamada de Matriz Fundamental. Ela representa o mapeamento do plano de projeção bidimensional da primeira imagem através do epipolo e_e , ou seja, identifica o mapeamento do espaço de projeção bidimensional para o unidimensional (Hartley e Zisserman, 2003).

Vários trabalhos propuseram-se a estudar os métodos de estimação da matriz fundamental, e podem ser divididos em métodos lineares, iterativos e robustos (Armangué e Salvi, 2003). Um caso particular dessa matriz é quando os parâmetros intrínsecos das câmeras são previamente conhecidos, o que facilita todo o processo de estimação de $\mathbf{F}$. Nesse caso a matriz é chamada de Matriz Essencial, $\mathbf{E}$. O trabalho de Armangué e Salvi (2003) apresenta um bom estudo sobre os principais métodos encontrados na literatura. Os autores apresentam uma breve introdução sobre a taxonomia dos métodos, incluindo resultados experimentais e o detalhamento dos algoritmos utilizados.

Então, considerando um sistema previamente calibrado, a geometria epipolar pode ser determinada de qualquer configuração de câmeras, e a busca por correspondências pode ser simplificada. As linhas epipolares definem o espaço em que seguramente o ponto P projetará sua imagem, tendo sido dado pelo menos uma de suas projeções. O espaço de busca, que antes era bidimensional (realizado em toda imagem), agora é limitado a uma simples busca unidimensional (Zhang et al., 2004), o que reduz drasticamente a complexidade do algoritmo de busca e,

conseqüentemente, a quantidade de ambigüidade.

Agora, supondo que os eixos ópticos z_e e z_d sejam paralelos e os planos de imagem Π_e e Π_d coplaneres como ilustrado na Figura 3.2.

Figura 3.2: Sistema estéreo binocular coplanar.

Para uma configuração desse tipo, as linhas epipolares coincidem com as próprias linhas das imagens. Ela apresenta a vantagem de desconsiderar a busca por correspondências no eixo vertical, e também a de não ser necessário o cálculo das linhas epipolares para cada projeção de referência. De fato, a construção desse tipo de sistema estéreo exige equipamentos de alta precisão no posicionamento das câmeras. Sobre um aspecto construtivo, é muito difícil chegar ao perfeito alinhamento das linhas epipolares com as linhas das imagens (Mallon e Whelan, 2005).

Diante dessa dificuldade, as imagens são previamente processadas com técnicas de retificação de imagens estéreo. É bom citar que existe basicamente dois tipos de retificação de imagens tratados em visão computacional. O primeiro trata da retificação das imagens capturadas por uma única câmera, buscando eliminar as distorções produzidas pela lente (distorção radial) ou pela própria perspectiva de visualização. O outro tipo, é a retificação de imagens estéreo, que consiste em transformar um par de imagens, normalmente originadas de um sistema de câmeras

convergentes, em imagens alinhadas, como se tivessem sido capturadas por um sistema de câmeras paralelas como ilustrado pela Figura 3.3. A retificação de imagens estéreo é conseguida através da aplicação de um conjunto de transformações geométricas sobre os planos de imagens, de forma que seus conjugados de linhas epipolares sejam colineares e paralelas a um dos eixos da imagem, normalmente o eixo horizontal.

Figura 3.3: Retificação de imagens estéreo.

Existe vasto material sobre o processo de retificação de imagens estéreo. Atualmente trabalhos como o de Mallon e Whelan (2005) estão sendo realizados com o propósito de eliminar as distorções provocadas pela retificação, e também tentando maximizar a similaridade com as imagens originais. Outro foco é abordado por Fusiello et al. (2000), que apresentam uma forma bastante compacta e de simples implementação do processo de retificação, o que é bastante interessante para aplicações que exigem menor tempo de processamento.

Há muito o que se falar sobre retificação de imagens estéreo, no entanto a compreensão das técnicas de retificação não é fundamental para a implementação de algoritmos de correspondência estéreo. O ponto de maior interesse nesse caso pode ser restringido somente à propriedade da redução do espaço de busca. A teoria sobre retificação de imagens é melhor discutida em Dankers et al. (2004) e Hartley e Zisserman (2003).

3.2.2 Unicidade

A restrição de unicidade é amplamente utilizada em visão estéreo. Ela considera que um dado ponto no mundo tridimensional possui projeções únicas em cada um dos planos de imagem e que cada região corresponde a um ponto diferente na cena. Sendo assim, caso uma região da imagem alvo já tenha sido relacionada a uma região de referência, a mesma não pode ser relacionada a outro ponto. Essa restrição é definida por (Chambon, 2005):

$$p_e^{i,j_1} + disp(p_e^{i,j_1}) = p_d^{i,r}, \quad \forall \quad j_1 \neq j_2, \quad p_e^{i,j_2} + disp(p_e^{i,j_2}) \neq p_d^{i,r} \quad (3.4)$$

onde p_e^{i,j_1} e p_e^{i,j_2} são *pixels* localizados no plano de imagem esquerdo, nas posições $[i \ j_1]$ e $[i \ j_2]$, respectivamente. A variável $p_d^{i,r}$ representa um *pixel* localizado na posição $[i \ r]$ dentro do plano de imagem direito, e a função $disp()$, é a função que devolve o valor de disparidade de um dado *pixel*.

A Equação 3.4 estabelece que, caso um *pixel* da imagem direita já tenha sido relacionado a um *pixel* da imagem esquerda, esses não podem ser relacionados a nenhum outro ponto. Um exemplo de uma situação inadmissível pela restrição de unicidade é ilustrado pela Figura 3.4.

Figura 3.4: Situação evitada pela restrição de unicidade.

Esse tipo de raciocínio favorece superfícies fronto paralelas, ou seja, aquelas cujo plano é paralelo à linha de base do sistema estéreo. Nesse caso, todos os pontos que compõem essas superfícies terão suas próprias projeções e a unicidade é válida.

No caso de superfícies não paralelas à linha de base, suas projeções são diferentes em ambas as imagens, vários pontos da cena podem possuir as mesmas projeções em uma das imagens. Uma simples configuração como ilustrado na Figura 3.5 já revela problemas ao considerar a restrição de unicidade. Nela, todos os pontos que estão entre p_e^{i,j_1} e p_e^{i,j_2} da imagem esquerda resumem-se em um único ponto, $p_d^{i,r}$, na imagem direita.

Figura 3.5: Falha ao considerar a restrição de unicidade.

3.2.3 Ordenação

Essa restrição considera que as projeções obedecem uma ordem, normalmente crescente, de suas posições sobre a linha epipolar. Da mesma forma que a restrição de unicidade, a restrição de ordenação funciona muito bem em casos específicos da cena. Infelizmente essa consideração não pode ser generalizada para qualquer tipo de ambiente. O problema apresentado por essa restrição ocorre normalmente quando objetos menores sobrepõem objetos maiores. Formalmente, a restrição de ordenação pode ser escrita como (Chambon, 2005):

$$p_e^{i,j_1} + \text{disp}(p_e^{i,j_1}) = p_d^{i,r_1} \text{ e } p_e^{i,j_2} + \text{disp}(p_e^{i,j_2}) = p_d^{i,r_2}, \forall (j_1 - j_2) \cdot (r_1 - r_2) \neq 0 \quad (3.5)$$

A Equação 3.5 descreve a restrição de ordenação. Ela determina que os valores de disparidades para os *pixels* p_e^{i,j_1} e p_e^{i,j_2} devem possuir o mesmo sinal, em relação aos seus correspondentes p_d^{i,r_1} e p_d^{i,r_2} . A Figura 3.6 mostra o que deve acontecer quando considerada a restrição de ordenação. As correspondências conservam a ordem das características da imagem de referência.

Figura 3.6: Ordenação de correspondências.

A Figura 3.7 apresenta um caso onde essa restrição falha. A presença de objetos pequenos entre o sistema de visualização e objetos maiores favorecem o surgimento de correspondências que não obedecem a restrição de ordenação.

Figura 3.7: Problema em considerar a restrição de ordenação.

3.2.4 Simetria

Como o nome já diz, essa restrição define que o algoritmo de procura por correspondência seja simétrico. Ou seja, sendo um par de imagens estéreo, o algoritmo de correspondência deve produzir o mesmo resultado, não importando qual das imagens foi eleita alvo ou referência. Formalmente pode-se escrever (Chambon, 2005):

$$p_e^{i,j} + \text{disp}(p_e^{i,j}) = p_d^{i,r} \quad (3.6)$$

$$p_d^{i,r} + \text{disp}(p_d^{i,r}) = p_e^{i,j} \quad (3.7)$$

Normalmente essa restrição é bastante útil no processo de detecção de oclusões na cena. Podem ser considerados pontos oclusos aqueles que possuem disparidades diferentes quando a imagem de referência é tomada como a imagem alvo.

3.3 Considerações finais

Nesse Capítulo foram abordadas as restrições consideradas pela maioria dos algoritmos de visão estéreo. No entanto, dentre elas, as que possuem significativa importância no desenvolvimento de novas metodologias é a consideração de que a cena é constituída por superfícies próximas à superfícies Lambertianas e a restrição da geometria epipolar. As outras restrições citadas no texto constituem casos específicos, que normalmente não estão presentes em grande parte das cenas reais, induzindo erros no processo de correspondência.

Capítulo 4

Métodos de Computação Estéreo e suas Classificações

4.1 Considerações iniciais

Até agora foram apresentadas as considerações e suposições feitas para facilitar a geração de mapas de disparidades mais precisos, mas sem imergir diretamente nos aspectos mais técnicos de visão estéreo. Nesse Capítulo é apresentado um panorama de como um algoritmo de visão estéreo é implementado, bem como a classificação dos métodos mais recentes.

4.2 Modelo genérico de computação estéreo

Considerando a restrição epipolar e imagens devidamente retificadas, como mencionado no Capítulo anterior, a computação das disparidades pode ser feita somente nas linhas das imagens.

A Figura 4.1 ilustra um par de imagens estéreo retificadas. Para cada *pixel* da imagem de referência procura-se o *pixel* correspondente na imagem alvo e, considerando uma janela de *pixels* vizinhos, o processo de casamento é dado sobre regiões das imagens.

Fixando um *pixel* de interesse na imagem de referência, a busca dá-se transladando uma janela de mesma dimensão na mesma linha dentro da imagem alvo. Para cada incremento de deslocamento é calculado um custo que resume o quanto a região é similar à região de referência, ou seja, o custo de correspondência.

Uma simples métrica desse custo que pode ser usada, é a distância entre as intensidades

Figura 4.1: Exemplo do processo de correspondência.

dos *pixels* das janelas. Considerando n_v *pixels* vizinhos, fazendo $\mathbf{W}_r$ a janela de referência centralizada no *pixel* com coordenadas $[y, x]^T$ e $\mathbf{W}_a$ a janela alvo centralizada no *pixel* com coordenadas $[y, x + d]^T$, ambas com dimensões $[2 \cdot n_v + 1 \times 2 \cdot n_v + 1]$, a equação da distância entre as regiões é dada por

$$\mathbf{EDI}(y, x, d_r) = \sqrt{\sum_{i,j=-n_v}^{n_v} [\mathbf{W}_r(y+i, x+j) - \mathbf{W}_a(y+i, x+j+d_r)]^2} \quad (4.1)$$

Esse processo resulta em uma matriz tridimensional chamada de *Disparity Space Image* (DSI) ou Espaço de Disparidade da Imagem (EDI). Ela mapeia o custo de similaridade de cada disparidade para o *pixel* de referência no espaço $[y, x, d_r]$, onde d_r é o valor de disparidade do índice r no vetor de deslocamento $\mathbf{d}$, e varia de d_{min} à d_{max} (ver Figura 4.2).

A Tabela 4.1 apresenta o pseudo-código desenvolvido nesse trabalho para a construção do EDI. O primeiro passo consiste na captura e posterior retificação das imagens. Como parâmetros, são passados o tamanho da janela de vizinhança e o vetor de possíveis disparidades. Nas linhas de 05 à 11, o cálculo de similaridade é realizado e armazenado para cada *pixel* da imagem de referência considerando todas as d_r disparidades dentro da imagem alvo.

Adotando que o valor ótimo seja o melhor de todos os valores, o mapa de disparidade $\mathbf{D}$ pode ser montado pela Equação 4.2. Teoricamente, a melhor disparidade é aquela que apresenta custo ótimo da métrica utilizada. Para esse exemplo, a melhor disparidade é o argumento do menor valor de distância encontrado pela Equação 4.1, mas nem sempre isso é verdade. A análise

Figura 4.2: Representação do espaço de disparidade da imagem.

Tabela 4.1: Pseudo-código de criação do EDI

Pseudo-código para a construção do EDI	
01:	Capturar as imagens Π_r e Π_a
02:	Aplicar método de retificação das imagens
03:	Definir tamanho da janela de vizinhança
04:	Definir espaço de busca: $\mathbf{d} = [d_{min} : d_{max}]$
05:	Para cada <i>pixel</i> $p_r^{y,x} \in \Pi_r$ fazer:
06:	Definir $\mathbf{W}_r$ referente à $p_r^{y,x} \in \Pi_r$
07:	Para cada deslocamento $d_r \in \mathbf{d}$ fazer:
08:	Definir $\mathbf{W}_a$ referente à $p_a^{y,x+d_r} \in \Pi_a$
09:	$\mathbf{EDI}(y, x, d_r) = Similaridade(\mathbf{W}_r, \mathbf{W}_a)$
10:	Fim Para
11:	Fim Para

puramente local das disparidades dentro do EDI não leva ao mapa de disparidade esperado, já que as imagens estão sujeitas aos problemas de oclusão, ambigüidade e variação de iluminação e os casamentos das projeções nem sempre são verdadeiros.

$$\mathbf{D}(y, x) = argmin(\mathbf{EDI}(y, x)) \quad (4.2)$$

Técnicas baseadas no comportamento global das disparidades conseguem tratar melhor esses impasses, mas por outro lado possuem maior custo computacional. Na próxima seção são citadas algumas dessas técnicas bem como técnicas locais e suas classificações.

4.3 Classificação dos métodos estéreo

Existem várias maneiras de se calcular o mapa de disparidades, mas em todas elas a correspondência de características é feita relacionando as projeções do mesmo ponto físico, diferenciando-se apenas pelas estratégias adotadas para a associação dos custos.

Em um senso de classificação mais amplo, pode-se dizer que essas estratégias são divididas em primitivas baseadas em área e baseadas em características (Dhond e Aggarwal, 1989). A primeira consiste em correlacionar as intensidades da vizinhança de um dado *pixel* de referência com a vizinhança de outro *pixel* na imagem alvo. Essa estratégia é justificada pela suposição de continuidade das disparidades, que considera que a vizinhança de um ponto possui disparidades semelhantes. Geralmente, *pixels* adjacentes em uma imagem muito provavelmente representam pontos contíguos no mundo tridimensional, evidenciando alguma superfície, o que torna essa estratégia factível. Dessa forma, é possível calcular a correspondência de cada *pixel* e montar um mapa denso de disparidades.

O problema ligado a essa estratégia é o fato de usar diretamente as intensidades dos *pixels*. Isso a torna sensível às variações de perspectiva, bem como às mudanças de intensidades e iluminação. A ambigüidade presente na cena, junto às oclusões geradas por essas distorções, prejudicam a construção fiel do mapa de disparidade.

Já em uma abordagem baseada em características, a correspondência é feita correlacionando padrões derivados das intensidades da vizinhança do *pixel* de interesse. A busca por esses padrões exige uma etapa adicional de pré-processamento que consiste em identificar as regiões que possuem variações consideráveis dos seus valores de intensidades, o que pode evidenciar a presença de bordas, segmentos, vértices e outros padrões de fácil identificação. Dessa forma, a computação torna-se mais estável quanto às variações de luminosidade do ambiente e resulta em casamentos mais precisos.

A correspondência baseada em características, apesar de apresentar robustez quanto à variação de luminosidade, não representa uma solução completa. A mudança de perspectiva tem influência direta nas características geométricas dos padrões encontrados. Uma reta horizontal pode facilmente corresponder a uma reta inclinada na imagem alvo ou o mesmo vértice pode ser composto de mais segmentos na outra imagem.

Além disso, é fácil notar que o uso de correspondência baseado em características não provê um mapa denso de disparidades. Isso porque nem sempre é possível relacionar um padrão à

vizinhança de cada *pixel*, como é o caso de regiões com falta de textura. Um exemplo claro disso é o pré-processamento baseado em detectores de bordas. Após o processamento, os pontos de borda representam uma pequena parcela da área total da imagem. Por outro, lado métodos baseados em área, fornecem um mapa denso, mas também cheio de ambigüidades. Quando usadas as duas estratégias em conjunto, as características podem ser usadas como uma orientação para os métodos baseados em área.

Dentro do escopo dos métodos que fornecem mapas densos de disparidade, Scharstein e Szeliski (2002) elaboraram uma taxonomia bastante coerente junto a uma metodologia de avaliação dos métodos com esse propósito. Eles dividiram a avaliação dos métodos em quatro passos:

1. **Computação do custo de correspondência:** faz uma análise das métricas utilizadas no cálculo da semelhança entre as regiões alvo e de referência. A métrica normalmente utilizada é a Correlação Cruzada Zero Normalizada (CCZN) que é menos sensível às variações de luminosidade. No entanto, alguns trabalhos tem apresentado bons resultados com o uso da Soma dos Quadrados das Diferenças (SQD) e Soma das Diferenças Absolutas (SDA).
2. **Custo de agregação:** avalia como os custos de correspondência são relacionados. Frequentemente, métodos agregam os custos pela soma ou pela média dentro do EDI. Normalmente os custos considerados pertencem a uma janela retangular ou janelas adaptativas, e também a aplicação de matrizes de pesos são usadas para ponderar a importância dos valores de custo.
3. **Computação e otimização das disparidades:** trata a forma de eleger a melhor disparidade, considerando parâmetros locais ou globais da imagem. Em uma abordagem simplesmente local, a computação da disparidade é trivial, sendo eleita aquela que possui o melhor custo de agregação. Por outro lado, técnicas globais relacionam o cálculo das disparidades com a solução de um problema de otimização de uma função de energia, considerando características comuns à regiões vizinhas.
4. **Refinamento das disparidades:** aumenta a resolução dos resultados. Na maioria dos algoritmos as disparidades são calculadas dentro de um espaço discreto, fato que não ocorre com as profundidades no mundo real. Então, como alternativa são empregadas técnicas de estimação de *sub-pixels* para o aumento da resolução do algoritmo e refinamento do resultado.

Brown et al. (2003) apresentam uma taxonomia mais simples e também mais recente. Eles

dividiram os métodos em algoritmos baseados em parâmetros locais e globais. Nesse contexto, os métodos locais consideram somente uma certa vizinhança do *pixel* de interesse. Já os métodos globais utilizam restrições que consideram toda a linha de varredura ou toda a imagem.

4.3.1 Métodos Locais de Correspondência

Casamento de regiões

Também conhecidos como métodos baseados em área, os métodos de casamento de regiões são montados sobre a estratégia de comparação de porções das imagens. A disparidade é calculada por uma função de custo que mede o quanto as regiões são correspondentes.

Um exemplo simples de como é desenvolvido um algoritmo baseado em regiões foi sumarizado na seção 4.2. Além da métrica exemplificada existem várias outras citadas na literatura e algumas delas são apresentadas na Tabela 4.2 (Brown et al., 2003).

Como dito anteriormente, métodos locais têm a vantagem de fornecer um mapa denso de disparidades, mas por outro lado é uma estratégia bastante influenciada por distorções locais. A Correlação Cruzada Normalizada (CCN) é a técnica padrão de comparação na área de processamento de sinais, e é uma das métricas mais utilizadas devido à sua robustez no tratamento das variações de luminosidade, mas à medida que o tamanho da janela aumenta, o esforço computacional torna-se muito alto (Fitch et al., 2005).

Alguns trabalhos propõem métodos rápidos para esse cálculo, tais como Sun (1997) e Sun (2002), que utilizam a Correlação Cruzada Normalizada Rápida (CCNR), calculando a variância e covariância das janelas com técnicas *Box-Filtering*¹ que têm a vantagem de serem mais velozes como discutido em McDonnell (1981). Outra alternativa é a aplicação da Soma dos Quadrados das Diferenças (SQD) e sua variação normalizada, Soma dos Quadrados das Diferenças Normalizada (SQDN), que também pode atenuar as distorções de luminosidade. No trabalho de Roma et al. (2000) é possível encontrar um estudo comparativo entre essas e outras métricas de similaridade usadas em visão estéreo.

As métricas onde o tratamento de distorções locais está mais expícito são a Transformação *Rank* (TR) e Transformação *Celsus* (TC) (Zabih e Woodfill, 1994). A primeira consiste na trans-

¹Procedimentos de filtragem de bloco apresentam a vantagem de serem mais velozes porque normalmente aproveitam o mesmo laço de interação para o cálculo de outras variáveis. No exemplo da variância, ela é calculada junto ao cálculo da média dos dados.

Tabela 4.2: Métricas de correspondência utilizadas no casamento de regiões.

Sigla	Definição da métrica de similaridade
CCN	$\frac{\sum_{i,j=-n}^n (\mathbf{W}_r(y+i, x+j) - \overline{\mathbf{W}}_r) \cdot (\mathbf{W}_a(y+i, x+j+d_r) - \overline{\mathbf{W}}_a)}{\sqrt{\sum_{i,j=-n}^n (\mathbf{W}_r(y+i, x+j) - \overline{\mathbf{W}}_r)^2 \cdot (\mathbf{W}_a(y+i, x+j+d_r) - \overline{\mathbf{W}}_a)^2}}$
SQD	$\sum_{i,j=-n}^n (\mathbf{W}_r(y+i, x+j) - \mathbf{W}_a(y+i, x+j+d_r))^2$
SQDN	$\sum_{i,j=-n}^n \left[\frac{\mathbf{W}_r(y+i, x+j) - \overline{\mathbf{W}}_r}{\sqrt{\sum_{i,j=-n}^n (\mathbf{W}_r(y+i, x+j) - \overline{\mathbf{W}}_r)^2}} - \frac{\mathbf{W}_a(y+i, x+j+d_r) - \overline{\mathbf{W}}_a}{\sqrt{\sum_{i,j=-n}^n (\mathbf{W}_a(y+i, x+j+d_r) - \overline{\mathbf{W}}_a)^2}} \right]^2$
SDA	$\sum_{i,j=-n}^n \mathbf{W}_r(y+i, x+j) - \mathbf{W}_a(y+i, x+j+d_r) $
TR	$\sum_{i,j=-n}^n \mathbf{W}'_r(y+i, x+j) - \mathbf{W}'_a(y+i, x+j+d_r), \text{ onde}$ $\mathbf{W}'_k(x, y) = \sum_{n, m \in \mathbf{W}_k} \mathbf{W}_k(m, n) < \mathbf{W}_k(y, x)$
TC	$\sum_{i,j=-n}^n \text{HAMMING}(\mathbf{W}'_r(y+i, x+j), \mathbf{W}'_a(y+i, x+j+d_r)), \text{ onde}$ $\mathbf{W}'_k(x, y) = \text{BITSTRING}(\mathbf{W}_k(m, n) < \mathbf{W}_k(y, x))$

formação das imagens originais em imagens cujos níveis representam a quantidade de vizinhos com intensidade menor que a intensidade do *pixel* central. Essa transformação reduz a sensibilidade às variações de ganho de iluminação, mas reduz a eficiência discriminatória da métrica utilizada na correspondência, já que os níveis de cinza dos *pixels* são muito comprimidos. Ainda no mesmo trabalho Zabih e Woodfill (1994) propuseram a Transformação *Celsus* que transforma os *pixels* em uma seqüência de caracteres binários preservando a distribuição espacial do *Rank* calculado. A métrica de similaridade é então feita utilizando a distância de *HAMMING*² entre as seqüências.

²Distância de *HAMMING* é uma medida de distância que tem como resultado a quantidade de bits diferentes entre duas seqüências. Por exemplo: *HAMMING*({1001}, {1000}) = 1 e *HAMMING*({1101}, {1010}) = 3.

Casamento de características

Se cada *pixel* fosse único perante toda a imagem, então o processo de correspondência seria realizado naturalmente sem a preocupação com erros de casamento, mas na prática isso não acontece. Mesmo adotando a restrição epipolar para a diminuição do espaço de busca, ainda existe alta probabilidade de encontrar regiões semelhantes.

Superfícies com falta de textura presentes na imagem favorecem o surgimento de regiões onde as intensidades dos *pixels* são homogêneas e não podem ser casadas com confiança. A falta ou mesmo a repetição de padrões discriminatórios na imagem é um problema que desafia as técnicas baseadas em área, e são as responsáveis pela ambigüidade durante a criação do EDI.

Em técnicas que aplicam casamento de regiões, a identificação de características pode ser usada como um pré-processamento, que identifica as regiões da imagem as quais proporcionam melhor confiança de correspondência, sem a preocupação com o tipo de padrão encontrado. Usualmente, o operador de Moravec³ (Moravec, 1977) e suas variações realizam bem esse trabalho.

Esse procedimento seleciona as regiões que proporcionam maior precisão durante a correspondência, e evita a busca por correspondências nas áreas mais homogêneas. Mesmo assim, o problema da ambigüidade ainda está presente, já que é realizado o casamento baseado em área. Diante disso, vários autores propõem métodos de identificação de padrões pré definidos, com o objetivo de realçar a unicidade de cada ponto. Com essa finalidade, são identificados segmentos, bordas, vértices que possam ser diferenciados, ou seja, bons candidatos para o casamento.

Koschan (1993a) realizou um extenso estudo sobre essas duas estratégias de correspondência em visão estéreo evidenciando as diferenças entre elas. Um trabalho mais recente com o mesmo propósito pode ser encontrado em Étienne Vincent e Laganière (2002). Concluiu-se nesses trabalhos que algoritmos baseados em características são realmente mais precisos, mas não podem ser usados quando há a necessidade de mapas mais densos de disparidades. Alguns outros estudos vão ainda mais longe tentando introduzir as vantagens dos métodos baseado em áreas, como Venkateswar e Chellappa (1995) que, durante a detecção das características, as organiza em uma árvore hierárquica de relacionamento considerando suas conectividades e proximidades, a fim de

³O Operador de Moravec é um operador de detecção de cantos bastante utilizado em aplicações de visão computacional. Ele foi desenvolvido por Hans Moravec em 1977, e tem como princípio básico medir a variação das intensidades dos *pixels* de uma janela centrada no ponto de interesse, quando transladada nas direções horizontais, verticais e nos quatro sentidos das diagonais.

detectar relacionamentos que definem superfícies com mesma disparidade.

4.3.2 Métodos Globais de Correspondência

Em uma abordagem local a escolha da melhor disparidade pode ser feita selecionando o melhor de todos os valores de similaridade como na Equação 4.2. Mas considerar somente a disparidade de um ponto individual sem considerar sua vizinhança significa deixar o algoritmo sensível à ambigüidade, já que um único ponto de referência pode ter vários candidatos a um bom casamento. Além disso, a presença de *outliers*⁴ pode levar a correspondências errôneas ou falsas verdades, que fogem do controle caso não sejam considerados aspectos globais da imagem.

Normalmente, estratégias globais realizam quase todo o trabalho de correspondência e determinação do melhor casamento, anexando uma função de suavidade à variação das disparidades. Muitos algoritmos globais caem no problema de minimização de uma função de energia, semelhante à Equação 4.3.

$$E(d_r) = \sum_{y,x} \mathbf{EDI}(y, x, d_r) + \alpha \cdot E_{smooth}(d_r) \quad (4.3)$$

O segundo termo da Equação 4.3 descreve a função de suavidade adotada e α o grau de importância da função (Scharstein e Szeliski, 2002).

As estratégias globais têm o objetivo de eleger a melhor disparidade considerando o comportamento global das correspondências geradas por técnicas locais. Naturalmente, considerar características globais torna essa estratégia mais complexa do que os métodos convencionais, além de exigir maior tempo de processamento (Brown et al., 2003).

Algoritmos globais usam diferentes métodos para obter o valor ótimo da função de energia. Pode-se citar como exemplos algoritmos baseados em *Dynamic Programming* (Bensrhair et al., 1996; Gong e Yang, 2003; Kim et al., 2005; Yip e Ho, 1998; Torr e Criminisi, 2004), *Intrinsic Curves* (Tomasi e Manduchi, 1998), *Belief Propagation* (Sun et al., 2003), *Nonlinear Diffusion* (Shah, 1993; Scharstein e Szeliski, 1998) e *Graphs Cuts* (Boykov et al., 1999).

⁴ *Outliers* assemelham-se com a definição de ruído. *Outliers* correspondem a dados muito diferentes do esperado, que fogem totalmente da tendência apresentada pelo conjunto.

Programação Dinâmica

Programação dinâmica consiste em reduzir a complexidade de um problema de otimização quebrando-o em sub-problemas de menor complexidade (Brown et al., 2003). Sua aplicação em visão estéreo dá-se considerando aspectos globais pertencentes à linha de varredura das imagens. Ela é formulada sobre uma abordagem de otimização de caminhos sobre uma fatia do EDI que possam ser mais contínuos e de menor custo global.

Existem basicamente duas formas de visualizar o problema de correspondência como um problema de programação dinâmica, ilustradas na Figura 4.3.

Figura 4.3: Representações para a aplicação de programação dinâmica.

Em ambas as imagens dessa figura, os *pixels* correspondem aos nós pelos quais deve passar algum caminho. Os de maior intensidade, representados pela cor branca, são os melhores custos calculados por alguma técnica baseada em regiões. A representação ilustrada pela Figura 4.3(a) é a representação do espaço que relaciona cada *pixel* da linha de referência com cada *pixel* da linha alvo. A matriz gerada contém todos os custos de similaridade possíveis no par de linhas estéreo. Dessa forma, o melhor caminho é aquele que parte do canto inferior esquerdo até o superior direito com a menor soma de custos. Passar por um *pixel* escuro significa considerar uma correspondência não confiável. A disparidade é dada pela diferença entre os índices por onde o caminho passa.

A segunda representação, ilustrada pela Figura 4.3(b), é semelhante à primeira, porém com a diferença de considerar somente o deslocamento $[d_{min}, d_{max}]$ na etapa de busca por correspondências. Essa representação é idêntica a uma fatia do EDI. Nesse caso, a otimização é realizada sobre o caminho que vai do lado esquerdo para o direito da imagem.

É fácil concluir que é praticamente inviável o teste de todos os caminhos possíveis dentro da matriz. Para a primeira forma de representação, por exemplo, numa imagem de dimensão N , o nível de complexidade utilizando programação dinâmica é de $O(N^4)$ sem considerar a etapa de construção do EDI. Pode-se reduzir drasticamente essa complexidade adotando um deslocamento máximo, como ilustrado na Figura 4.3(b). Dessa forma, a complexidade é de apenas $O(N^2)$ contra $O(N^N)$ requisitados pela otimização global da imagem completa, que não usa programação dinâmica (Scharstein e Szeliski, 2002).

A principal vantagem de se utilizar programação dinâmica é a capacidade de considerar o comportamento global da linha de varredura. Mesmo em regiões com falta de textura, pode-se ter uma boa estimativa da disparidade real considerando que o caminho seja suave. Mas, por outro lado, essa abordagem é altamente dependente de resultados passados. Um erro anterior no caminho pode ser propagado ao longo de toda linha. Além disso, essa técnica trabalha melhor quando pode ser considerada a restrição de ordenação, o que não é um caso comum em cenas reais.

Para a compreensão mais detalhada sobre o desenvolvimento de algoritmos baseados em programação dinâmica pode ser consultado o livro Forsyth e Ponce (2002), e também Duarte (2002) que traz um precioso anexo contendo um exemplo simples e didático sobre esse assunto.

4.4 Considerações finais

Nesse Capítulo foi apresentado um exemplo simples de criação de mapas de disparidade, juntamente à taxonomia mais recente dos métodos da área.

A compreensão da seção 4.2 é de grande importância para o entendimento da metodologia utilizada no algoritmo proposto por essa dissertação. A leitura da seção 4.3 proporciona a visualização de onde o trabalho percorrido se encaixa perante as várias características classificatórias dos algoritmos de visão estéreo. Foi dado ênfase também à programação dinâmica, devido sua extensa utilização nos trabalhos da área.

Capítulo 5

Trabalhos Relacionados

5.1 Considerações iniciais

Esse Capítulo é dedicado à apresentação de trabalhos relacionados com essa dissertação.

5.2 Revisão de técnicas relacionadas

O estado da arte de visão estéreo com o foco na criação de mapas de disparidades divide-se basicamente em duas principais vertentes, o uso de técnicas locais e globais. Vários autores defendem o uso de estratégias globais para a solução do problema da correspondência baseando-se no argumento de melhor qualidade dos resultados finais. Por outro lado, estratégias locais são ditas como soluções relativamente boas pelo tempo reduzido de computação que elas apresentam. O ponto comum dessa discussão dá-se no equilíbrio entre a fidelidade dos resultados e do menor tempo de processamento dos algoritmos. Não se pode dizer que uma técnica é melhor que a outra pois essa afirmação depende inteiramente do tipo da aplicação final.

Na perspectiva de soluções globais, destacam-se os métodos baseados em *Dynamic Programming*, *Intrinsic Curves*, *Belief Propagation*, *Nonlinear Diffusion* e *Graphs Cuts*. Na abordagem local, os métodos mais conhecidos são *Adaptive Windows*, *Shiftable Windows* e também *Adaptive Weights*.

De uma maneira geral, entre os vários algoritmos de correspondência, sejam eles locais ou globais, o aspecto decisivo acaba caindo sobre como é realizada a correspondência entre as regiões das imagens. Afinal, a qualidade dos valores apresentados pelo EDI é a base para as etapas de

pós-processamento.

A suposição feita ao considerar que todos os *pixels* pertencentes às janelas de casamento possuem a mesma disparidade, por mais que seja amplamente utilizada nos trabalhos da área, pode não ser a melhor estratégia de correspondência, principalmente em presença de descontinuidades dentro do mapa real de disparidades. Enquanto que em superfícies mais suaves a maioria dos *pixels* possuem disparidades semelhantes, as regiões que correspondem às extremidades dos objetos apresentam uma grande diferença entre as disparidades vizinhas (descontinuidades). Nessas áreas, a probabilidade da janela de vizinhança considerar mais pontos oclusos ao arranjo estéreo é muito maior que nas outras. Por outro lado, em regiões mais suaves, a homogeneidade dos *pixels* implica em ambigüidades que são diretamente influenciadas pelo tamanho da janela de vizinhança. Quanto menor a janela, mais chances se tem de encontrar regiões homogêneas, e quanto maior a janela, maior é a probabilidade de considerar variações significativas de níveis de cinza e também pontos oclusos. Sendo a oclusão e as regiões homogêneas características inteiramente dependentes da natureza da cena, o tamanho e a forma da janela de vizinhança é um problema de extrema importância para a obtenção de uma solução correta, podendo variar tanto de cena para cena quanto no próprio espaço de cada imagem.

Kanade e Okutomi (1994) tentaram solucionar esse impasse utilizando janelas adaptativas com dimensões apropriadas em cada ponto da imagem. Com o objetivo de definir o melhor tamanho da janela de vizinhança, eles aplicaram operadores estatísticos que modelam a distribuição de intensidades e disparidades dos *pixels* vizinhos. Dessa forma, as dimensões das janelas são ajustadas para que as estimativas de disparidades sejam geradas com o menor grau de incertezas em cada ponto da imagem.

Um trabalho mais recente, publicado por Veksler (2003), descreve como principal idéia, um algoritmo que procura uma faixa de dimensões de janelas, que exploram da melhor forma a vizinhança do *pixel* de interesse e também a elaboração de um custo de comparação entre janelas de diferentes tamanhos. O problema enfrentado é o tempo de computação exigido pelo algoritmo. Para um determinado *pixel* de interesse dentro do espaço de imagem, várias janelas podem contê-lo e a computação acaba tornando-se redundante em suas adjacências. A solução apresentada é o uso de Imagens Integrais¹ a fim de evitar computações desnecessárias. Outros trabalhos com idéias semelhantes encontram-se em Boykov et al. (1998) e Veksler (2002).

Sobre uma nova linha de raciocínio, o problema relacionado aos pontos oclusos pode ser

¹ Imagens Integrais também são conhecidas como *Summed Area Tables*.

reduzido quando são usadas Janelas Deslocadas ou *Shiftable Windows*. Supondo que uma certa região da imagem seja semelhante à Figura 5.1, onde a imagem do centro representa a imagem de referência. As regiões marcadas com *B*, *C*, *D* e *E* estão parcialmente ocluídas, enquanto que a marcada com *F* representa um quadrado que sobrepõe todas as demais.

(a) Janela fixa

(b) Janelas deslocadas

Figura 5.1: Janelas deslocadas reduzindo o problema de oclusão.

No primeiro caso, Figura 5.1(a), a janela é fixa e centralizada no *pixel* de interesse destacado. Um algoritmo que usa esse tipo de janela certamente irá cometer erros ao realizar a correspondência dessas regiões. O uso de janelas deslocadas ajuda a evitar esse tipo de situação. Na segunda imagem por exemplo, Figura 5.1(b), as janelas não estão centralizadas no *pixel* de interesse e sobrepõem exatamente a região que o contém, o que favorece o casamento correto das regiões por não considerar *pixels* ocluídos. Bobick e Intille (1999) e Kang et al. (2001) utilizaram nove janelas, cada uma com o *pixel* de interesse deslocado. A Figura 5.2 ilustra a estrutura das janelas usada. O quadrado preto indica o *pixel* de interesse de cada uma.

Como pode ser observado, cada janela é responsável pelo casamento correto em nove posições vizinhas ao *pixel* em destaque. A correspondência é realizada em cada uma das estruturas e a identificação da disparidade é feita considerando a janela que possui maior grau de semelhança, descartando-se as outras. Assim, a possibilidade de se considerar pontos que não possuem disparidades semelhantes ao *pixel* de referência é reduzida. No entanto, existe um acréscimo no

Figura 5.2: Janelas deslocadas.

tempo computacional, já que para cada ponto é necessário correlacionar mais oito janelas diferentes. Além disso ainda permanece a dificuldade de determinar o tamanho ideal da janela de casamento.

Quanto às técnicas baseadas em *Adaptive Weights* destaca-se o trabalho de Xu et al. (2002). Nesse artigo, os autores definem uma janela de suporte através de computações radiais, cujos pesos são definidos considerando as disparidades vizinhas a um certo raio. O raio e o grau de certeza das disparidades vizinhas contribuem para a relevância do pixel analisado. Essa metodologia consegue melhorar as correspondências em presença de ambigüidades mas não consegue lidar com regiões oclusas. Quanto aos problemas envolvidos, os próprios autores sugerem uma maior investigação no tratamento de oclusões e um refinamento no algoritmo, que, em presença de regiões muito homogêneas, apresenta tempo de computação alto por selecionar janelas de suporte muito grandes.

5.3 Considerações finais

Existem vários métodos em busca de soluções para o problema da correspondência. Os trabalhos discutidos nesse Capítulo correspondem a uma pequena parcela do real montante de publicações existentes, sendo que a maioria dos estudos convergem para a aplicação de alguma estratégia de otimização global sobre o EDI. O fato desse trabalho tratar o problema da correspondência buscando utilizar somente técnicas locais e de simples implementação, é um dos fatores que o diferencia dos demais trabalhos.

Capítulo 6

Algoritmo de Correspondência

6.1 Considerações iniciais

No Capítulo anterior foram mencionados alguns trabalhos que visam definir a melhor forma de se construir o EDI para se chegar em um mapa de disparidades mais satisfatório. Nesse Capítulo é detalhado o raciocínio utilizado na construção do algoritmo de correspondência elaborado nesse trabalho. Os métodos aplicados são discutidos, citando algumas vantagens de se usar pirâmides e agrupamento perceptual para o tratamento de correspondência de regiões.

6.2 Metodologia

De fato, a aplicação de técnicas que evitam a inclusão dos pontos oclusos nos cálculos das correspondências e também a redução de ambigüidades, implica em melhorias quando comparadas à forma convencional, definida pelo uso de janelas de tamanho fixo e centralizadas no *pixel* de interesse. Entretanto, dificilmente é elaborado um método que consegue lidar com esses dois problemas ao mesmo tempo.

O uso de janelas de tamanho adaptativo tem como principal vantagem reduzir os erros causados pelas ambigüidades, enquanto que os algoritmos que usam janelas deslocadas concentram esforços para evitar que as regiões oclusas atrapalhem o processo de correspondência. Em ambos os casos não existe o tratamento simultâneo desses dois problemas. Para suprir essa deficiência, a maioria dos trabalhos apresenta uma etapa de pós-processamento construída sobre uma perspectiva de otimização global da imagem. Esse procedimento visa eliminar as falsas correspondências

apresentadas pelo EDI, para assim chegar no mapa de disparidades final. Além disso, a forma da janela de vizinhança não é genérica para todas as regiões da imagem, isso porque formas retangulares não trabalham muito bem em presença de descontinuidades. É evidente que, apesar da melhoria apresentada pelas técnicas discutidas, elas não conseguem resolver completamente os problemas de oclusões e ambigüidades que dificultam a construção fiel do mapa de disparidades.

Em virtude disso, o trabalho abordado por essa dissertação busca uma contribuição para a melhoria da construção do EDI. Para isso, é indispensável a redução dos erros causados por oclusões e ambigüidades na etapa do casamento das regiões. A metodologia adotada busca evitar os pontos oclusos, e ao mesmo tempo considera uma região de vizinhança que reduza a influência de ambigüidades. Com essa finalidade, um modelo de busca por correspondências é proposto baseado sobre a afirmação de que cada *pixel* da janela de vizinhança possui uma contribuição diferente para o casamento correto das regiões.

Nos algoritmos tradicionais, as tentativas de melhorar os resultados dão-se variando o tamanho e a forma das janelas de correspondência, enquanto que os *pixels* que a compõem possuem o mesmo grau de importância nos cálculos. Essa metodologia leva a janelas que não se ajustam às variadas regiões das imagens. Nesse trabalho, tanto a forma quanto o tamanho da janela de vizinhança são definidos pela inclusão de uma matriz de agregação, aqui chamada de janela suporte, que contém os pesos referentes a cada *pixel*. Ela é construída com base em uma teoria bastante conhecida na psicologia, a Teoria *Gestalt* de agrupamento perceptual, que busca agrupar elementos visuais através da similaridade de cores e proximidade geométrica dos *pixels*.

O problema da oclusão é atenuado quando é aplicada a janela suporte para seleção das regiões com disparidades semelhantes, enquanto que a ambigüidade é reduzida considerando uma análise multi-resolução das imagens.

A Figura 6.1 apresenta um diagrama que explica a metodologia adotada nesse trabalho. Cada imagem do par estéreo passa por um processo de decomposição, chamado de decomposição piramidal, fornecendo cópias em resoluções diferentes. Com o auxílio da Teoria *Gestalt*, um mapa de disparidades é gerado considerando as imagens do topo da pirâmide. Esse mapa dá uma estimativa inicial das disparidades do nível seguinte, funcionando como ponto de partida para o refinamento dos valores de disparidades. Para cada nível da pirâmide, um mapa de disparidade é gerado considerando o mapa do nível superior. O resultado final é obtido realizado a busca por correspondências de cima para baixo até que a base da pirâmide seja alcançada. As seções posteriores descrevem a teoria e o algoritmo de correspondência mais detalhadamente.

Figura 6.1: Esquema geral da metodologia adotada.

6.3 *Gestalt Grouping*

A teoria *Gestalt* é uma das várias teorias sobre organização perceptual, e é amplamente utilizada na compreensão de como se dá o agrupamento dos estímulos sensitivos humanos. Ela tenta explicar como humanos conseguem organizar elementos individuais em grupos, e como percebem e reconhecem diferentes padrões (Chang et al., 2007).

Cada elemento perceptual possui natureza e características próprias, mas somente a natureza desses elementos não permite dizer como um grupo deve ser percebido. Devido a isso, essa teoria estabelece alguns princípios de agrupamento baseados em Similaridade, Proximidade, Conexidade, Continuidade, Forma, Alinhamento, Paralelismo, entre outros, com o objetivo de auxiliar o processo de formação de conjuntos maiores, mais complexos e com algum significado perceptível.

Apesar de ter sido descrita por volta de 1910, a teoria *Gestalt* ultimamente é alvo de atenção especial na área de visão computacional. Recentemente, trabalhos como Chang e Nesbitt (2006) e Shi e Malik (2000) que utilizam esses princípios para resolver problemas de visão computacional com soluções mais próximas à percepção humana, estão se tornando cada vez mais freqüentes, sendo que a maioria deles com o foco principal no uso das propriedades de similaridade e proximidade.

Uma das principais aplicações dessa teoria está na explicação de como o cérebro humano consegue agrupar e organizar os diversos estímulos visuais. Os pontos que de alguma forma possuem uma ou mais características em comum são agrupados e identificados como um novo e maior objeto visual, ou seja, um *Gestalt*.

6.3.1 Agrupamento por similaridade

De uma forma geral, o princípio de similaridade resume a capacidade das pessoas agruparem elementos visuais baseados em atributos como cores, formas, área, etc. Para o caso de agrupamento de *pixels*, são considerados semelhantes aqueles que possuem cores parecidas. A medida de similaridade entre as cores $\mathbf{c}_p = [R_p \ G_p \ B_p]$ e $\mathbf{c}_q = [R_q \ G_q \ B_q]$ dos *pixels* p e q respectivamente, pode ser obtida pela distância euclidiana entre elas,

$$\Delta_{c_{pq}} = \sqrt{(R_p - R_q)^2 + (G_p - G_q)^2 + (B_p - B_q)^2} \quad (6.1)$$

onde as letras R , G e B correspondem às intensidades dos canais RGB . A força de agrupamento perceptual gerada pela similaridade das cores dos elementos é então formulada da seguinte forma (Yoon e Kweon, 2006):

$$f_s(\Delta_{c_{pq}}) = \exp\left(-\frac{\Delta_{c_{pq}}}{\gamma_s}\right) \quad (6.2)$$

onde a função $\exp(x)$ é uma função elementar do tipo e^x , sendo e o número de Euler, e γ_s , uma constante que normalmente possui valor 14, definido pela teoria *Gestalt*.

Fazendo uma rápida análise da Equação 6.2, é fácil perceber que quanto menor a diferença entre as cores, maior é a probabilidade dos dois *pixels* serem considerados pontos pertencentes ao mesmo grupo. Por outro lado, uma diferença maior implica em afirmar que os dois *pixels* analisados possuem menor grau de similaridade. Essas afirmações são verdadeiras se considerado que os pontos que compõem a mesma superfície refletem a luz de forma semelhante e, caso exista alguma variação de cor, esta se dá de maneira suave. A Figura 6.2 ilustra os gráficos de similaridade para alguns valores de γ_s .

Agora, supondo uma janela de vizinhança como ilustrado pela Figura 6.3(a), e o *pixel* p definido pelo quadrado no centro da imagem, a função de similaridade é calculada para cada vizinho q . O cálculo de similaridades resulta em uma imagem como a 6.3(b), onde as regiões mais claras correspondem aos pontos com maior grau de similaridade.

Figura 6.2: Gráfico de similaridade.

Figura 6.3: Agrupamento por similaridade de cor.

6.3.2 Agrupamento por proximidade

Os pontos que possuem cores semelhantes ainda dependem de outro fator de agrupamento para formarem uma superfície. A presença de *pixels* iguais mas isolados não evidenciam a presença de regiões contínuas, então, além de serem semelhantes, os pontos que formam alguma superfície devem estar próximos uns dos outros de maneira a serem categorizados como um grupo maior. A distância entre cada *pixel* é obtida pela distância geométrica entre eles. A equação de

proximidade do *pixel* p em relação ao *pixel* q é dada por:

$$\Delta d_{pq} = \sqrt{(u_p - u_q)^2 + (v_p - v_q)^2} \quad (6.3)$$

sendo $[u_p \ v_p]$ e $[u_q \ v_q]$ as coordenadas dos *pixels* p e q respectivamente.

A força de agrupamento por proximidade está descrita pela Equação 6.4, e é obtida da mesma forma que a de similaridade. Pela teoria *Gestalt*, o peso de agrupamento de um *pixel* decresce proporcionalmente ao afastamento espacial do *pixel* dado como âncora. Assim, somente pequenas distâncias proporcionam maior percepção de agrupamento (Yoon e Kweon, 2006).

$$f_p(\Delta d_{pq}) = \exp\left(-\frac{\Delta d_{pq}}{\gamma_p}\right) \quad (6.4)$$

Na Equação 6.4, γ_p é uma constante e normalmente possui valor igual à metade da dimensão da janela de vizinhança. A Figura 6.4(a) ilustra o agrupamento de proximidade para as dimensões da janela representada pela Figura 6.3(a).

(a) Máscara de proximidade

(b) Pesos de proximidade aplicados à imagem original 6.3

Figura 6.4: Agrupamento por proximidade espacial.

A janela resultante da combinação entre a imagem original e os pesos de proximidade espacial, ilustrada pela Figura 6.4(b), mostra a variação do efeito de proximidade dos *pixels*. Quanto mais próximos à região de interesse, mais claro é o tom de cinza, enquanto que os *pixels* das extremidades da janela, são praticamente desconsiderados devido a grande distância em relação ao ponto considerado.

6.3.3 Suporte de agrupamento

A janela suporte completa, que reúne as informações da força de agrupamento por similaridade e proximidade, pode ser escrita como:

$$f_t\left(\frac{\Delta c_{pq}}{\gamma_s}, \frac{\Delta d_{pq}}{\gamma_p}\right) = f_s(\Delta c_{pq}) \cdot f_p(\Delta d_{pq}) \quad (6.5)$$

ou então,

$$\mathbf{W}_s(p, q) = \exp\left(-\left(\frac{\Delta c_{pq}}{\gamma_s} + \frac{\Delta d_{pq}}{\gamma_p}\right)\right) \quad (6.6)$$

onde $\mathbf{W}_s$ é a janela suporte final que guarda os pesos de todos os vizinhos de p . A Figura 6.5 apresenta a visualização da janela suporte final e sua visualização tridimensional.

(a) Janela suporte

(b) Janela suporte 3D

Figura 6.5: Agregação dos pesos de Similaridade e Proximidade: Janela suporte.

O uso dessa metodologia permite a construção de janelas totalmente adaptativas a qualquer região da imagem. Ela consegue modelar a região com maior probabilidade de possuir disparidades semelhantes baseada somente nas informações contidas na janela de vizinhança. Além disso, as dificuldades impostas pela forma retangular da janela não é mais problema, visto que a força de agrupamento é calculada para cada *pixel* de vizinhança.

6.4 Pirâmides de Imagens

O tratamento de ambigüidades exige uma análise eficiente das regiões homogêneas. A alternativa normalmente adotada é variar o tamanho das janelas de vizinhança até que seja detectada uma variação considerável dos níveis de cinza que a compõem. Essa metodologia, em muitos casos, pode levar à geração de janelas muito grandes e prejudicar o desempenho total da busca por correspondências. Há problemas também com a determinação do limiar correto de variação das intensidades que, na maioria dos casos, não é o mesmo para todas as regiões da imagem.

Muitas vezes, as regiões ambíguas só são reveladas quando os *pixels* são analisados de muito perto, ou seja, em alta resolução e podem ser desmascaradas quando observadas a partir de uma resolução mais baixa. Ou seja, uma região totalmente homogênea em uma imagem de alta resolução pode representar um pequeno detalhe quando analisada em uma resolução mais baixa.

Nesse caso, a aplicação de pirâmides de imagens pode facilitar o processo de busca por correspondências, possibilitando a análise de imagens em diversas resoluções.

O nome "*Pirâmide*" é derivado da estrutura utilizada para organizar as imagens de diferentes escalas. Ela é constituída por uma seqüência de cópias da imagem original onde as resoluções são reduzidas em passos regulares. As Figuras 6.6 e 6.7 mostram um exemplo de estrutura piramidal e o resultado da decomposição respectivamente.

Figura 6.6: Estrutura piramidal.

O nível 0 corresponde à imagem original. O próximo nível é obtido através da aplicação de um filtro passa-baixa junto a um fator de amostragem que é normalmente 2. Nos níveis maiores que 0 cada ponto resume informações sobre uma quantidade fixa de pontos da imagem de nível menor. Esse procedimento implica em imagens menores e também mais suaves, o que é carac-

Figura 6.7: Decomposição em pirâmides.

terística da presença de baixas frequências. Esse processo de decomposição da imagem original em imagens de menor resolução é conhecido como Análise da Imagem. É comum encontrar também a nomenclatura "Redução" para esse processo. Cada nível da pirâmide é obtido pela redução do nível logo abaixo. O processo inverso é chamado de Síntese ou "Expansão". Ele se dá, recuperando a imagem de alta resolução a partir de sua cópia de resolução mais baixa, utilizando algoritmos de interpolação para calcular os novos pontos criados. A maneira mais simples consiste na simples repetição dos valores calculados.

A vantagem acrescentada pelo uso dessa estrutura é o fato de se ter acesso simultâneo à tendência global das estruturas que compõem a imagem, e também aos detalhes discriminatórios de cada uma. Quanto mais alto o nível da pirâmide menor é a quantidade de detalhes, mas também melhor é a representação das características globais da mesma. Além disso, existe um acréscimo na eficiência computacional. Essa estrutura permite trabalhar com imagens de dimensões menores sem perda considerável de informações. Um simples exemplo pode ser visto na Figura 6.8, onde a busca pela região escura na imagem original é feita comparando 16 *pixels*, enquanto que a busca realizada na estrutura piramidal de cima para baixo usa somente 8 comparações.

Esse exemplo ilustra uma decomposição unidimensional de pirâmides. Cada ponto carrega consigo a informação das intensidades dos dois pontos imediatamente abaixo dele. A informação contida na alta resolução é transportada até o topo da pirâmide.

O cálculo das dimensões da janela de vizinhança para cada *pixel*, na busca por correspondências, pode ser extremamente custoso caso seja feito na imagem original. Essa estrutura permite utilizar uma janela de tamanho fixo para toda a imagem, e ao mesmo tempo evitar as regiões

Figura 6.8: Busca por regiões com e sem a decomposição piramidal.

ambíguas. Uma janela relativamente pequena, se comparada à imagem original, consegue considerar informações de muito mais *pixels* quando aplicada em um nível de decomposição mais alto, ou seja, nas baixas resoluções da pirâmide. Dessa forma uma região que antes era ambígua pode muito bem ser uma região fácil de ser diferenciada no topo da pirâmide.

De uma maneira mais formal, cada nível da pirâmide pode ser obtido por (Adelson et al., 1984):

$$I_{n+1}(y, x) = \sum_i \sum_j \mathbf{W}_c(i, j) \cdot I_n(2y + i, 2x + j), \quad \text{para } 0 \leq n < N \quad (6.7)$$

onde N é a quantidade de níveis de decomposição e $\mathbf{W}_c$ uma matriz de convolução aplicada sobre a imagem $\mathbf{I}_n$. A matriz $\mathbf{W}_c$ modela a ação de um filtro passa-baixa no domínio espacial. Uma das formas mais simples de Análise emprega $\mathbf{W}_c$ como um extrator da média de 4 *pixels* do nível abaixo. Nesse caso, $\mathbf{W}_c$ é uma matriz 2×2 como descrito a seguir.

$$\mathbf{W}_c = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix} \quad (6.8)$$

Existem várias metodologias sobre a análise multi-resolução de imagens. Vários trabalhos tem sido publicados com o propósito de se obter as decomposições que melhor representam as características globais dos níveis de maior resolução, motivados principalmente por esse tipo de estrutura estar naturalmente embutida no sistema visual humano. Bister (2004) realiza uma análise de três métodos bem conhecidos, as Pirâmides Laplacianas, o Espaço de Escala Gaussiano e a Transformada *Wavelet*. Ressaltando uma análise a partir de uma perspectiva baseada na transformada de *Fourier*, o autor apresenta fortes semelhanças entre essas técnicas e cita algumas

vantagens relacionadas a cada uma. A decomposição Laplaciana corresponde à decomposição menos complexa entre as mencionadas e permite uma boa análise multi-resolução.

6.5 Vantagens de pirâmides de imagens no problema da correspondência

Para o caso do problema de correspondência, a análise de pirâmides de imagens favorece dois pontos principais: um como solução para o problema das ambigüidades e o outro como metodologia de redução do tempo de processamento.

6.5.1 Ambigüidades

Considerando as imagens ilustradas pela Figura 6.9 que expressam claramente o problema de ambigüidade. A Figura 6.9(a) corresponde à imagem da base da pirâmide. Ela destaca uma porção da imagem onde o casamento de regiões dá-se sobre uma área homogênea, onde grande parte dos métodos estéreo pode falhar. Nessa região, o custo de similaridade é semelhante para qualquer janela dentro do espaço de deslocamento selecionado.

(a) Imagem no nível 0 da pirâmide (alta resolução)

(b) Imagem no nível 1 da pirâmide (baixa resolução)

Figura 6.9: Redução da ambigüidade com análise multi-resolução.

Já a Figura 6.9(b) equivale ao primeiro nível de decomposição da pirâmide. Nesse caso, a mesma janela de vizinhança considera uma quantidade maior de informações, aumentando a probabilidade de sobrepor variações de intensidade de *pixel*. À medida em que a análise é feita nos níveis superiores da pirâmide, menor é a chance de se considerar regiões ambíguas.

6.5.2 Custo computacional

A busca por correspondências é um processo que exige muito tempo computacional. Apesar de minimizado com a restrição epipolar, cada ponto da imagem de referência precisa ser comparado com toda ou parte da linha epipolar correspondente dentro da imagem alvo. Pela forma tradicional, uma imagem com R linhas e C colunas exige N_r chamadas à função de custo de similaridade. No pior caso, cada ponto de referência é comparado com toda a linha da imagem alvo. Para reduzir computações desnecessárias, define-se uma faixa de disparidades, sendo d_{max} a maior e d_{min} a menor disparidade que pode ser encontrada no par estéreo. Ainda assim o número de requisições à função de similaridade é exagerado. As Equações 6.9 e 6.10 mostram o número de requisições para a busca padrão e por faixa de disparidades, respectivamente.

$$N_r = R \cdot C \cdot C \quad (6.9)$$

$$N_r = R \cdot C \cdot t_d \quad (6.10)$$

onde $t_d = d_{max} - d_{min}$, a quantidade de disparidades considerada.

Com a análise multi-resolução, os níveis de baixa resolução possuem muito menos *pixels* que a imagem original. No topo da pirâmide, tanto as dimensões das imagens como a faixa de disparidade estabelecida são reduzidas por um fator de 2^n . Essa propriedade implica em menos requisições à função de similaridade e também menor tempo de processamento. Por outro lado, resulta em um mapa de disparidade precário, já que muita informação de detalhes é perdida na construção dos níveis da pirâmide. Como os níveis de baixa resolução são formados principalmente por baixas frequências, o mapa calculado evidencia somente uma estimativa das profundidades para o próximo nível da pirâmide. A Figura 6.10 mostra as disparidades em cada nível da pirâmide.

Figura 6.10: Pirâmide de disparidades.

Nesse trabalho, as estimativas de disparidades são utilizadas como parâmetro de redução do número de chamadas à função de similaridade. A partir de um valor de disparidade estimado, a busca por correspondências no próximo nível é realizada somente na vizinhança do ponto alvo estimado pelo nível anterior, o que funciona como processo de refinamento dos valores de disparidade. Dessa forma, adotando que o erro máximo no nível 0 seja dado por t_{max} vizinhos, o número de requisições à função custo é descrito como

$$N_r = \frac{R \cdot C \cdot t_d}{2^{3 \cdot (N-1)}} + \sum_{n=N-2}^0 \frac{R \cdot C}{2^n} \cdot \left(2 \cdot \frac{t_{max}}{2^n} + 1\right) \quad (6.11)$$

sendo $t_{max}/2^n$ a tolerância adotada no nível n . O primeiro termo da Equação 6.11 computa a varredura completa das imagens no topo da pirâmide e o segundo descreve o número de requisições dos níveis abaixo em função de n e do erro de estimação considerado t_{max} . A Figura 6.11 mostra a relação entre os valores de N_r para as estratégias de busca mencionadas. Observa-se claramente que o uso de pirâmides é uma boa estratégia para a melhoria do desempenho computacional, superando as duas formas tradicionais.

Figura 6.11: Número de chamadas à função de custo de similaridade considerando $t_d = 15$.

6.6 Descrição do algoritmo

A implementação do algoritmo foi baseada na teoria explicada pelas seções 6.3 e 6.4. A estrutura geral e as etapas percorridas pelo código estão ilustradas na Figura 6.12. A estra-

tégia adotada pode ser dividida em quatro etapas principais que são diferentes dos métodos tradicionais. São elas:

1. Decomposição piramidal da imagem de referência e alvo;
2. Geração do mapa de disparidades inicial;
3. Estimação das disparidades do próximo nível;
4. Refinamento das disparidades.

Na primeira etapa é realizada a decomposição piramidal das imagens alvo e de referência. Cada nível da decomposição é armazenado separadamente para serem utilizadas futuramente. A segunda consiste na determinação das tendências das profundidades do ambiente. Nessa etapa a cena é totalmente desconhecida pelo algoritmo e os objetos que a compõem podem estar dispostos das mais variadas formas. Por esse motivo, a busca por correspondências é feita no topo da pirâmide, e permite ter uma noção global da disposição das estruturas mais simples, devido à própria natureza da análise multi-resolução. A busca por correspondências acaba por percorrer e quantificar toda a faixa de disparidades proporcional ao topo da pirâmide, no caso $d/2^{N-1}$. Embora essa estratégia seja computacionalmente custosa, no topo da decomposição, tanto a faixa de disparidades como as dimensões das imagens são reduzidos por um fator de $2^{(N-1)}$, o que torna factível a aplicação do método convencional de buscas por correspondências. A métrica de similaridade utilizada é a Soma das Diferenças Absolutas (SDA) associada à janela suporte definida por *Gestalt Grouping*.

Sendo $\mathbf{W}_e$ a matriz das diferenças absolutas entre os vizinhos do *pixel* de referência e do *pixel* alvo, ela pode ser escrita como

$$\mathbf{W}_e(q_r, q_a) = \sum_{c \in R, G, B} [I_c(q_r) - I_c(q_a)]^2 \quad (6.12)$$

onde q_r e q_a são todos os *pixels* que pertencem à janela de referência e à janela alvo, respectivamente. O custo total de similaridade é obtido pela Equação 6.13,

$$e(p_r, p_a) = \frac{\sum_i \sum_j \mathbf{W}_e(q_r, q_a) \cdot \mathbf{W}_s(p_r, q_r) \cdot \mathbf{W}_s(p_a, q_a)}{\sum_i \sum_j \mathbf{W}_s(p_r, q_r) \cdot \mathbf{W}_s(p_a, q_a)} \quad (6.13)$$

sendo $\mathbf{W}_s(p_r, q_r)$ e $\mathbf{W}_s(p_a, q_a)$ as janelas suportes para as janelas de vizinhança de referência e alvo, obtidas pela Equação 6.6.

Com o EDI calculado, resta então a seleção das melhores disparidades. Pelo motivo desse trabalho basear-se em uma técnica local de correspondência, a alternativa de seleção acaba por ser definida pelo método característico das técnicas locais, ou seja, o argumento do valor mínimo do custo de similaridade, expresso pela Equação 4.2. A aplicação dessa equação resulta no mapa de disparidades para o par estéreo do topo da pirâmide. A partir desse estágio, é possível ter uma estimativa das disparidades globais da cena, as quais permitem guiar a busca pelas regiões mais prováveis à correspondência e, evidentemente, a redução de computações desnecessárias.

Na terceira etapa, a dimensão do mapa e as disparidades disponibilizados pela análise do topo da pirâmide é duas vezes menor que as do próximo nível. Por isso, o valor de disparidade é duplicado e a estimativa da profundidade é calculada pela função de expansão da própria pirâmide. O processo adotado foi a interpolação linear para suprir os novos *pixels* adicionais. Nesse ponto, cada ponto possui uma estimativa de onde deve estar sua região correspondente.

Na quarta etapa, o algoritmo possui um mapa de disparidades estimado e o usa para direcionar os cálculos de similaridade para iniciar ao refinamento das disparidades em um nível de resolução mais alta. A busca por correspondência é realizada somente nas regiões vizinhas à estimada, sendo que a quantidade de chamadas à função de similaridade passa a ser igual a $2 \cdot (t_{max}/2^n) + 1$ para cada ponto. O mesmo procedimento é repetido até que o nível 0 da pirâmide seja alcançado. O fluxograma do algoritmo é ilustrado pela Figura 6.12.

Figura 6.12: Fluxograma do algoritmo proposto.

6.7 Considerações finais

Nesse Capítulo foi descrita a metodologia adotada para a criação de mapas densos de disparidades. O uso de pirâmides de imagens permite atenuar a influência negativa causada pelas regiões homogêneas, enquanto que a aplicação de cálculos de agrupamento perceptual, teoria *Gestalt*, modela a melhor forma da janela de correspondência para o *pixel* de referência e evita os problemas causados pelas discontinuidades.

Capítulo 7

Resultados

7.1 Considerações iniciais

Nesse Capítulo são apresentados os primeiros resultados desse trabalho. O código de visão estéreo foi implementado sobre a plataforma MATLAB R2007a pela facilidade com que a linguagem lida com operações matriciais e operações com imagens.

7.2 Base de dados utilizada

Os resultados foram obtidos pela execução do Algoritmo de Correspondência descrito no Capítulo anterior, utilizando as imagens disponibilizadas pelo Colégio Universitário de Middlebury no estado de Vermont (Middlebury Stereo Vision Page, 2007).

O banco de imagens é formado principalmente por quatro pares estéreo devidamente retificados, constituindo a base dos testes para vários métodos da área. São eles: os pares de *Tsukuba*, *Venus*, *Teddy* e *Cones* representadas pelas Figuras 7.1, 7.2, 7.3 e 7.4 respectivamente. Para efeito de avaliação, estão disponíveis também o mapa de disparidade real, chamado de *Ground Truth*, e o mapa de pontos oclusos de cada par.

O uso dessas imagens é justificado pela natureza de cada uma. O par *Venus* é composto sinteticamente, enquanto que os outros são imagens de cenas reais. Em todos os casos existem pontos oclusos, regiões texturizadas e com falta de textura, indicando que cada um dos pares exige robustez dos métodos estéreo no tratamento de ambigüidades e oclusões.

(a) Imagem direita

(b) Imagem esquerda

(c) *Ground truth*

(d) Mapa de oclusão

Figura 7.1: Par estéreo *Tsukuba*.

(a) Imagem direita

(b) Imagem esquerda

(c) *Ground truth*

(d) Mapa de oclusão

Figura 7.2: Par estéreo *Venus*.

(a) Imagem direita

(b) Imagem esquerda

(c) *Ground truth*

(d) Mapa de oclusão

Figura 7.3: Par estéreo *Teddy*.

(a) Imagem direita

(b) Imagem esquerda

(c) *Ground truth*

(d) Mapa de oclusão

Figura 7.4: Par estéreo *Cones*.

7.3 Considerações sobre a avaliação dos resultados

A avaliação do mapa de disparidades é feita comparando o mapa real com o mapa produzido pelo algoritmo. Cada par estéreo possui algumas características particulares definidas pelo próprio sistema de captura, como o tamanho das imagens, a disparidade máxima encontrada e a escala utilizada na avaliação dos resultados.

Embora no mundo real as profundidades assumam infinitos valores, na forma digital elas são truncadas pela própria aquisição dos dados. Como a forma de se calcular a distância entre duas regiões dentro de uma imagem se dá contando a quantidade de *pixels* que as separam, as disparidades são formadas por números inteiros já que a busca é feita somente nas linhas. Sendo assim, a quantidade de profundidades é inteiramente dependente da resolução das câmeras de aquisição e implica em imagens muito grandes para uma melhor resolução. A Tabela 7.1 apresenta os parâmetros característicos dos pares estéreo utilizados. A coluna chamada de Faixa de Disparidades, apresenta a faixa de valores de disparidades que são explorados pelo algoritmo. Ela indica o valor mínimo e máximo do deslocamento feito pela janela de correspondência, sendo eles determinados pela geometria do sistema estéreo. A coluna chamada de Tamanho indica as dimensões das imagens que compõem o par estéreo. Por último, a coluna nomeada Escala apresenta os valores de escala para cada par estéreo usado na avaliação do mapa de disparidade criado.

Tabela 7.1: Parâmetros característicos dos pares estéreo.

Par estéreo	Escala	Faixa de Disparidades	Tamanho
<i>Tsukuba</i>	16	$0 \dots 15$	$[288 \times 384]$
<i>Venus</i>	8	$0 \dots 19$	$[383 \times 434]$
<i>Teddy</i>	4	$0 \dots 59$	$[375 \times 450]$
<i>Cones</i>	4	$0 \dots 59$	$[375 \times 450]$

A identificação dos *pixels* corretos é feita comparando cada *pixel* do mapa de disparidades criado com o *ground truth* do par estéreo. A quantificação do erro apresentado pelos resultados baseia-se na porcentagem de disparidades errôneas, onde são contabilizadas como erro aquelas cujos valores diferenciam-se por mais de 1. As outras que permanecerem dentro dessa margem de erro são consideradas como disparidades corretas. A eficiência, B , do algoritmo é calculada pela porcentagem de disparidades erradas, descrita pela equação 7.1.

$$B = \frac{100}{R \cdot C} \cdot \sum_{(y \in R, x \in C)} |\mathbf{D}(y, x) - \hat{\mathbf{D}}(y, x)| > 1 \quad (7.1)$$

7.4 Apresentação dos resultados

A metodologia usada na geração dos resultados foi idealizada com o propósito de permitir avaliações posteriores dos parâmetros de entrada do algoritmo, como mostrado na seção 7.5. Para efeito de testes, o tamanho da janela de vizinhança, ou seja, a quantidade de vizinhos n_v considerados, e a constante de similaridade γ_s foram variados. A quantidade de níveis da pirâmide foi fixada em $N = 3$, pois, a partir do quarto nível de decomposição, o topo da pirâmide passa a ter dimensões muito pequenas quando comparadas às dimensões das janelas de vizinhança utilizadas nesse trabalho.

As Figuras 7.5, 7.6, 7.7 e 7.8 mostram os melhores resultados obtidos para cada par estéreo. Aquelas identificadas com as letras (a), (b) e (c) apresentam o mapa de disparidades gerado em cada nível da pirâmide, sendo (a) correspondente ao topo da pirâmide, (b) ao nível intermediário de decomposição, e (c) com a base da pirâmide. Os mapas reais de disparidades estão apresentados pelas Figuras identificadas com a letra (d), buscando uma melhor comparação com o resultado final.

Figura 7.5: Melhor resultado para o par *Tsukuba*. Obtido com $n_v = 17$, $\gamma_s = 9$ e $B = 2.37$.

Figura 7.6: Melhor resultado para o par *Venus*. Obtido com $n_v = 17$, $\gamma_s = 9$ e $B = 1.47$.

Figura 7.7: Melhor resultado para o par *Teddy*. Obtido com $n_v = 15$, $\gamma_s = 18$ e $B = 0.23$.

Figura 7.8: Melhor resultado para o par *Cones*. Obtido com $n_v = 13$, $\gamma_s = 11$ e $B = 0.44$.

A análise das Figuras 7.5, 7.6, 7.7 e 7.8 permite acompanhar a evolução do refinamento das disparidades. No nível mais alto da pirâmide, as imagens utilizadas pelo algoritmo são pobres em detalhes e fornecem a tendência geral das disparidades da cena. O processamento das camadas intermediárias da pirâmide revela características não observadas pelos níveis superiores até que se alcance a base da pirâmide, permitindo considerar as características de alta resolução.

Em todos os casos, o mapa obtido condiz com o mapa ideal, respeitando a distribuição das intensidades de cinza por ele definida. Nas quatro figuras, os pontos mais claros correspondem àqueles que apresentam maior disparidade e conseqüentemente menor profundidade. As regiões mais escuras indicam os pontos que estão mais distantes do sistema estéreo.

As Figuras 7.9 e 7.10 permitem fazer uma comparação visual entre os resultados produzidos e os de alguns algoritmos da área. A Figura 7.9(a) mostra o resultado do par *Tsukuba* com o uso de *Shiftable Windows* (Scharstein e Szeliski, 2002) e a Figura 7.9(b) ilustra o resultado de Black e Rangarajan (1996) com a técnica *Fast Correlation*. Ambas constituem técnicas locais, enquanto que as Figuras 7.9(c) e 7.9(d) apresentam os resultados de técnicas globais utilizando *Belief Propagation* (Sun et al., 2003) e *Graph Cuts* (Belhumeur e Mumford, 1992), respectivamente. A Figura 7.10 segue o mesmo padrão de exibição.

(a) *Shiftable Windows*

(b) *Fast Correlation*

(c) *Belief Propagation*

(d) *Graph Cuts*

(e) Algoritmo desenvolvido

Figura 7.9: Resultados do par *Tsukuba* para outras estratégias de correspondência.

(a) *Shiftable Windows*

(b) *Fast Correlation*

(c) *Belief Propagation*

(d) *Graph Cuts*

(e) Algoritmo desenvolvido

Figura 7.10: Resultados do par *Venus* para outras estratégias de correspondência.

O algoritmo desenvolvido, apresentado pelas Figuras 7.9(e) e 7.12(b), apesar de ser essencialmente local, aproxima-se dos resultados obtidos pelas técnicas globais. A Tabela 7.2 mostra os valores de erro das técnicas citadas para os pares *Tsukuba* e *Venus* considerando 17 vizinhos na janela de correspondência, $n_v = 17$.

Tabela 7.2: Valores de B para os pares *Tsukuba* e *Venus*.

	<i>Shiftable Windows</i>	<i>Fast Corr.</i>	<i>Belief Propagation</i>	<i>Graph Cuts</i>	Desenvolvido
<i>Tsukuba</i>	5.23	9.76	1.15	1.86	2.37
<i>Venus</i>	3.74	6.48	1.00	1.69	1.47

Os mapas de disparidades gerados pelo algoritmo desenvolvido apresentam erro inferior aos das técnicas locais, aproximando-se ou superando os resultados obtidos pelas técnicas globais. Os algoritmos citados não apresentam resultados para os pares *Teddy* e *Cones*, com exceção da técnica *Shiftable Windows*, cujos erros estão disponíveis na Tabela 7.3.

Tabela 7.3: Valores de B para a técnica *Shiftable Windows*.

	<i>Shiftable Windows</i>	Desenvolvido	n_v
<i>Tsukuba</i>	5.23	2.37	17
<i>Venus</i>	3.74	1.47	17
<i>Teddy</i>	16.5	0.23	15
<i>Cones</i>	10.6	0.44	13

As Figuras 7.11 e 7.12 mostram os mapas de disparidades gerados com a técnica *Shiftable Windows* para os pares *Teddy* e *Cones*, respectivamente.

(a) *Shiftable Windows*

(b) Algoritmo desenvolvido

Figura 7.11: Resultados do par *Teddy* para a técnica *Shiftable Windows*.

(a) *Shiftable Windows*

(b) Algoritmo desenvolvido

Figura 7.12: Resultados do par *Cones* para a técnica *Shiftable Windows*.

7.5 Discussão sobre os parâmetros do algoritmo

Além do par de imagens estéreo, a implementação sugerida por esse trabalho necessita de alguns parâmetros adicionais, como a faixa de disparidades (d), o número de vizinhos considerados para formar a janela de vizinhança de cada *pixel* (n_v), a quantidade de níveis de decomposição da Pirâmide (N) e também as constantes γ_s e γ_p .

O valor de γ_p foi definido como sendo $(n_v \cdot 2 + 1)/2$, o que corresponde à metade do tamanho da janela de vizinhança. Apesar da teoria *Gestalt* definir um valor específico para a constante de similaridade, $\gamma_s = 14$, é interessante observar como essa variável interfere no processo de agrupamento perceptual das regiões.

A variação do tamanho da janela de vizinhança tem influência direta no desempenho total da busca por correspondências. Quanto maior o número de vizinhos maior é a quantidade de dados a processar, mas também melhor pode ser a diferenciação das regiões, principalmente na base da pirâmide, onde é necessário maior robustez do algoritmo. Atribuiu-se a n_v os valores de [4 5 7 9 11 13 15 17] o que resultou respectivamente em janelas de tamanho [9×9 11×11 15×15 19×19 23×23 27×27 31×31 35×35].

As Figuras 7.13, 7.14, 7.15, 7.16 mostram os gráficos que relacionam a porcentagem de erro e a constante de similaridade para cada tamanho da janela de vizinhança aplicada sobre os pares *Tsukuba*, *Venus*, *Teddy* e *Cones*, respectivamente.

A análise desses gráficos mostra que o tamanho da janela e a constante de similaridade

têm grande influência no resultado final do algoritmo, mas se estabilizam quando ultrapassam determinados valores. Pode-se observar que existe uma tendência de redução da porcentagem de erro quando o tamanho da janela de vizinhança aumenta. As janelas com dimensões maiores apresentam vantagens por considerar uma maior quantidade de informações sobre a vizinhança do *pixel* de interesse, favorecendo a aplicação da teoria de agrupamento perceptual. Por outro lado, janelas menores não conseguem abranger *pixels* suficientes para a detecção de grupos maiores com mesma disparidade, dificultando assim o casamento correto das regiões.

Pode-se notar que a partir 13 vizinhos considerados, o tamanho da janela deixa de ser muito importante para o resultado final. Além disso, os valores se estabilizam quando γ_s ultrapassa o valor 6.

Figura 7.13: Gráfico de erro para o par *Tsukuba*.

Figura 7.14: Gráfico de erro para o par *Venus*.

Figura 7.15: Gráfico de erro para o par *Teddy*.

Figura 7.16: Gráfico de erro para o par *Cones*.

As Tabelas 7.4, 7.5, 7.6 e 7.7 apresentam os valores da constante de similaridade que proporcionam melhores resultados para cada par estéreo, considerando todas as dimensões das janelas de vizinhança testadas.

Tabela 7.4: Melhores resultados para o par *Tsukuba*.

Tamnho da Janela	Erro B (%)	Constante de similaridade (γ_s)
09×09	6.04	20
11×11	5.23	17
15×15	4.16	12
19×19	3.46	10
23×23	2.96	10
27×27	2.64	10
31×31	2.49	10
35×35	2.37	09

A Tabela 7.4 apresenta os melhores resultados do par *Tsukuba*. Para janelas com dimensões maiores que $[23 \times 23]$, o erro B apresenta uma pequena variação se comparado com as janelas menores. Além disso, nota-se uma estabilidade de γ_s a partir da janela $[15 \times 15]$.

Tabela 7.5: Melhores resultados para o par *Venus*.

Tamnho da Janela	Erro B (%)	Constante de similaridade (γ_s)
09×09	8.63	20
11×11	7.08	20
15×15	5.33	13
19×19	4.08	11
23×23	2.99	10
27×27	2.28	09
31×31	1.79	10
35×35	1.47	09

A Tabela 7.5 relaciona os resultados para o par *Venus*. Nela, o comportamento dos resultados se assemelha ao produzido pelo par *Tsukuba*, onde γ_s permanece próximo do valor 9.

Tabela 7.6: Melhores resultados para o par *Teddy*.

Tamnho da Janela	Erro B (%)	Constante de similaridade (γ_s)
09×09	2.73	20
11×11	2.17	20
15×15	1.11	20
19×19	0.43	20
23×23	0.23	20
27×27	0.23	17
31×31	0.23	18
35×35	0.2556	15

Os resultados do par *Teddy*, mostrados na Tabela 7.6, diferem-se dos outros pares quanto ao

valor de γ_s . Nesse caso, a constante de similaridade que gerou melhores resultados corresponde ao dobro dos valores apresentados nas Tabelas 7.4 e 7.5.

Tabela 7.7: Melhores resultados para o par *Cones*.

Tamnhho da Janela	Erro B (%)	Constante de similaridade (γ_s)
09 $\times$ 09	0.88	20
11 $\times$ 11	0.96	20
15 $\times$ 15	1.02	20
19 $\times$ 19	0.55	16
23 $\times$ 23	0.48	14
27 $\times$ 27	0.44	11
31 $\times$ 31	0.46	10
35 $\times$ 35	0.45	09

Os resultados do par *Cones*, (Tabela 7.7), se dão de forma semelhante aos pares *Tsukuba* e *Venus*.

Pela análise dessas tabelas, é possível notar que o tamanho da janela de vizinhança passa a ter pouca influência quando ultrapassa o tamanho $[23 \times 23]$, o que deixa claro que é possível chegar a resultados semelhantes usando janelas menores. Na prática, a redução da janela de vizinhança implica na redução do tempo de execução, já que o algoritmo passa a considerar uma quantidade menor de *pixels* em seus cálculos. A relação completa dos resultados de cada par estéreo são apresentados no Apêndice A, onde os resultados podem ser observados mais detalhadamente.

7.6 Considerações finais

A maioria dos trabalhos que prometem boa qualidade do mapa de disparidades recorre a técnicas globais que, normalmente, exige muito tempo computacional. Nesse trabalho, a seleção das disparidades dentro do EDI é feita considerando somente o custo individual de similaridade de cada *pixel* em relação ao deslocamento da janela de correspondência. Essa abordagem faz desse trabalho uma técnica local de correspondências, sendo de fácil implementação e exigindo menor número de requisições à função de similaridade em relação a outros algoritmos também considerados locais, como mostrado na Figura 6.11. Os resultados apresentados nesse Capítulo permitem afirmar que a abordagem adotada possibilita a construção de mapas de disparidades densos e próximos àqueles produzidos por algumas técnicas globais, com a vantagem de usar métodos de menor complexidade em relação aos discutidas no Capítulo 5.

Capítulo 8

Conclusões

Essa dissertação aborda um tema de grande interesse na área de visão computacional. Ela descreve uma estratégia computacional para a recuperação de informações tridimensionais de uma cena a partir de imagens estereoscópicas. A metodologia base resume-se na criação de mapas densos de disparidades, que permitem compreender como estão dispostas as diversas estruturas que compõem uma cena real, evidenciando assim sua distribuição 3D.

Para esse fim, foi desenvolvido um algoritmo com o objetivo de reduzir a complexidade dos métodos abordados pela literatura e, ao mesmo tempo, preservar a qualidade dos resultados por eles encontrada. Em sua grande maioria, os métodos globais apresentam bons resultados mas perdem pela complexidade de implementação e pelo elevado custo computacional. Em contrapartida, os métodos locais exigem menor custo computacional, mas deixam a desejar na qualidade de seus resultados.

Uma das metas foi a de desenvolver um algoritmo rápido, para isso, a aplicação da análise multi-resolução apresentou-se vantajosa quanto à redução da quantidade de chamadas à função de custo de similaridade, o que é extremamente importante quando se espera um algoritmo mais veloz. Outro objetivo atingido foi a atenuação dos efeitos negativos produzidos pelas ambigüidades e oclusões dentro do par estéreo. Para o caso de regiões ambíguas, a própria natureza da análise multi-resolução dá condições de evitar tais regiões quando processadas em baixa resolução. Os pontos oclusos foram evitados pela aplicação da teoria de agrupamento perceptual, selecionando pontos que compõem grupos maiores e com grandes chances de possuírem a mesma disparidade.

Os resultados apresentados permitem realizar uma investigação detalhada da influência da

constante de similaridade e do tamanho da janela de vizinhança para todos os pares de imagens mencionados. Quando extrapolados o valor 6 para γ_s e 13 para n_v (número de vizinhos), tanto o tamanho da janela quanto a constante de similaridade passam a ter pouca influência no erro apresentado pelos mapas de disparidades.

Por fim, o trabalho desenvolvido alcançou seus objetivos, gerando mapas densos de disparidades, reduzindo a quantidade de chamadas à função de similaridade e aproximando-se dos resultados obtidos por técnicas globais. O uso da teoria de agrupamento perceptual faz desse trabalho um método moderno de estimação de disparidades, visto que essa técnica é alvo de atenção especial em recentes estudos na área de visão computacional.

Como trabalhos futuros, pode ser proposto uma investigação maior na análise multi-resolução das imagens, junto à aplicação e comparação de outras técnicas de multi-resolução. Outro aspecto a ser observado é a identificação e tratamento de pontos oclusos na cena para facilitar o processo de reconstrução tridimensional do mapa de disparidades. Além disso, pode ser sugerido a calibração do sistema de visão estéreo com o objetivo de converter disparidades em profundidades, ou seja, em medidas de distâncias reais. E por último, considerando a simplicidade do algoritmo, a implementação em *hardware*, o que possibilitaria a aplicação e teste de sistemas estéreo embarcados.

Capítulo 9

Apêndice A - Resultados Detalhados

As Tabelas 9.1, 9.2, 9.3 e 9.4 apresentam os resultados para os pares *Tsukuba*, *Venus*, *Teddy* e *Cones* respectivamente. A primeira coluna indica o valor da constante de similaridade, γ_s , enquanto que as outras indicam o tamanho da janela de vizinhança. As linhas contém o valor de erro B , obtido pela equação 7.1, em função da dimensão da janela e do valor de γ_s .

Tabela 9.1: Resultados do par *Tsukuba*.

γ_s	Dimensão da janela							
	09 × 09	11 × 11	15 × 15	19 × 19	23 × 23	27 × 27	31 × 31	35 × 35
1	27.87	26.78	25.55	24.59	23.86	23.26	22.78	22.34
2	21.02	19.44	17.35	15.72	14.73	13.91	13.26	12.64
3	16.26	14.55	12.10	10.62	9.61	8.82	8.16	7.69
4	12.99	11.30	9.06	7.89	6.90	6.21	5.63	5.09
5	10.93	9.25	7.31	6.24	5.22	4.49	3.90	3.53
6	9.52	8.04	6.09	4.87	4.09	3.52	3.17	2.81
7	8.46	7.07	5.24	4.22	3.51	3.06	2.69	2.46
8	7.77	6.41	4.72	3.88	3.20	2.82	2.53	2.37
9	7.18	5.92	4.46	3.61	3.04	2.67	2.50	2.37
10	6.78	5.64	4.27	3.46	2.96	2.64	2.49	2.42
11	6.52	5.45	4.20	3.51	2.96	2.67	2.54	2.48
12	6.36	5.37	4.16	3.54	3.02	2.74	2.63	2.57
13	6.27	5.29	4.19	3.56	3.10	2.85	2.78	2.71
14	6.17	5.27	4.23	3.59	3.17	2.96	2.88	2.82
15	6.13	5.24	4.25	3.62	3.23	3.05	2.99	2.95
16	6.08	5.24	4.27	3.68	3.29	3.18	3.12	3.09
17	6.05	5.23	4.28	3.71	3.39	3.29	3.23	3.23
18	6.05	5.24	4.31	3.75	3.47	3.39	3.33	3.35
19	6.04	5.26	4.36	3.79	3.56	3.51	3.45	3.46
20	6.04	5.32	4.39	3.83	3.62	3.60	3.56	3.58

Tabela 9.2: Resultados do par *Venus*.

γ_s	Dimensão da janela							
	09 × 09	11 × 11	15 × 15	19 × 19	23 × 23	27 × 27	31 × 31	35 × 35
1	37.87	36.78	35.14	33.58	32.33	31.27	30.47	29.70
2	28.23	25.74	22.22	19.59	17.46	15.82	14.55	13.59
3	21.49	18.62	14.66	12.03	10.15	8.73	7.68	6.93
4	17.28	14.37	10.61	8.38	6.73	5.51	4.67	4.04
5	14.61	11.83	8.48	6.52	5.10	4.02	3.22	2.74
6	12.83	10.21	7.25	5.49	4.18	3.19	2.52	2.05
7	11.60	9.25	6.47	4.88	3.59	2.68	2.09	1.70
8	10.79	8.59	5.98	4.48	3.26	2.41	1.88	1.52
9	10.22	8.12	5.71	4.23	3.07	2.28	1.80	1.47
10	9.85	7.85	5.52	4.11	2.99	2.29	1.79	1.55
11	9.53	7.61	5.44	4.08	3.03	2.32	1.87	1.64
12	9.30	7.47	5.37	4.08	3.08	2.41	1.98	1.74
13	9.17	7.38	5.33	4.12	3.14	2.48	2.07	1.88
14	9.07	7.29	5.35	4.16	3.20	2.59	2.21	2.04
15	8.92	7.23	5.35	4.20	3.28	2.69	2.35	2.19
16	8.84	7.18	5.34	4.24	3.35	2.78	2.45	2.32
17	8.77	7.14	5.38	4.27	3.41	2.86	2.57	2.48
18	8.71	7.12	5.37	4.29	3.47	2.93	2.67	2.62
19	8.67	7.12	5.40	4.31	3.53	3.00	2.76	2.73
20	8.63	7.08	5.40	4.35	3.59	3.07	2.85	2.85

Tabela 9.3: Resultados do par *Teddy*.

γ_c	Dimensão da janela							
	09 × 09	11 × 11	15 × 15	19 × 19	23 × 23	27 × 27	31 × 31	35 × 35
1	14.37	14.18	12.96	12.34	11.83	11.48	11.43	11.35
2	11.74	11.42	9.66	8.81	8.27	7.81	7.54	7.45
3	9.3	8.65	6.85	5.83	4.95	4.41	4.11	3.81
4	7.47	6.93	5.06	3.96	3.22	2.94	2.74	2.58
5	6.36	5.84	3.92	3.03	2.61	2.48	2.29	2.20
6	5.48	5.03	3.41	2.70	2.37	2.22	2.06	2.06
7	4.97	4.50	3.05	2.53	2.21	2.02	1.92	1.98
8	4.59	4.10	2.90	2.43	2.06	1.91	1.82	1.83
9	4.29	3.8	2.77	2.32	1.65	1.45	1.31	1.10
10	4.02	3.64	2.58	1.68	1.52	1.28	1.00	0.99
11	3.81	3.49	2.16	1.36	1.22	1.03	0.93	0.72
12	3.63	3.33	1.84	1.26	1.09	0.83	0.68	0.54
13	3.43	2.95	1.70	1.09	0.88	0.61	0.43	0.34
14	3.35	2.85	1.60	0.94	0.71	0.44	0.27	0.25
15	3.10	2.43	1.52	0.80	0.53	0.29	0.26	0.25
16	2.95	2.37	1.40	0.68	0.42	0.24	0.25	0.25
17	2.88	2.31	1.32	0.62	0.35	0.23	0.25	0.40
18	2.81	2.26	1.27	0.54	0.27	0.24	0.23	0.40
19	2.76	2.19	1.14	0.49	0.25	0.51	0.46	0.41
20	2.73	2.17	1.11	0.43	0.23	0.55	0.48	0.44

Tabela 9.4: Resultados do par *Cones*.

γ_c	Dimensão da janela							
	09 × 09	11 × 11	15 × 15	19 × 19	23 × 23	27 × 27	31 × 31	35 × 35
1	13.25	13.32	13.34	12.86	12.13	11.96	11.89	11.84
2	10.50	10.05	8.95	7.52	6.26	5.78	5.44	5.23
3	6.95	6.20	5.07	3.94	3.06	2.73	2.53	2.40
4	4.71	4.12	3.54	2.58	1.83	1.57	1.48	1.36
5	3.35	3.17	2.65	1.67	1.19	1.00	0.91	0.82
6	2.68	2.47	2.06	1.27	0.86	0.70	0.67	0.63
7	2.11	1.99	1.75	0.98	0.67	0.55	0.52	0.51
8	1.69	1.73	1.51	0.84	0.56	0.47	0.47	0.48
9	1.47	1.51	1.41	0.73	0.52	0.47	0.46	0.45
10	1.31	1.37	1.32	0.68	0.53	0.44	0.46	0.48
11	1.19	1.26	1.30	0.69	0.49	0.44	0.47	0.47
12	1.14	1.21	1.23	0.66	0.48	0.44	0.53	0.48
13	1.06	1.12	1.21	0.61	0.49	0.47	0.56	0.48
14	1.03	1.08	1.19	0.58	0.48	0.55	0.58	0.50
15	0.97	1.05	1.18	0.57	0.59	0.55	0.58	0.50
16	0.91	1.02	1.17	0.55	0.57	0.58	0.60	0.51
17	0.89	0.98	1.09	0.61	0.58	0.59	0.60	0.59
18	0.90	1.00	1.10	0.59	0.59	0.60	0.60	0.59
19	0.89	0.97	1.05	0.59	0.6	0.60	0.60	0.59
20	0.88	0.96	1.02	0.58	0.61	0.60	0.62	0.59

Referências Bibliográficas

- Adelson, E. H., C. H. Anderson, J. R. Bergen, P. J. Burt, e J. M. Ogden (1984). Pyramid methods in image processing. *RCA Engineer* 29(6).
- Ahmeda, M. F. (2006). Development of stereo vision system for outdoor mobile robots. Dissertação de Mestrado, University of Florida.
- Armangué, X. e J. Salvi (2003). Overall view regarding fundamental matrix estimation. *International Journal of Image and Vision Computing, IVC* 21(2), 205–220.
- Barnard, S. T. e M. A. Fischler (1982). Computational stereo. *ACM Computing Surveys (CSUR)* 14(4), 553–572.
- Belhumeur, P. N. e D. Mumford (1992). A bayesian treatment of the stereo correspondence problem using half-occluded regions. *IEEE Conference on Computer Vision and Pattern Recognition*, 506–512.
- Bensrhair, A., P. Miché, e R. Debrie (1996). Fast and automatic stereo vision matching algorithm based on dynamic programming method. *Pattern Recogn. Lett.* 17(5), 457–466.
- Bister, M. (2004). A new perspective on multiresolution image processing. In *CGIV '04: Proceedings of the International Conference on Computer Graphics, Imaging and Visualization (CGIV'04)*, Washington, DC, USA, pp. 267–272. IEEE Computer Society.
- Black, M. J. e A. Rangarajan (1996). On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. *International Journal of Computer Vision* 19(1), 57–91.
- Bleyer, M. e M. Gelautz (2007). Graph-cut-based stereo matching using image segmentation with symmetrical treatment of occlusions. *Image Commun.* 22(2), 127–143.

- Bobick, A. F. e S. S. Intille (1999). Large occlusion stereo. *Int. J. Comput. Vision* 33(3), 181–200.
- Boykov, Y., O. Veksler, e R. Zabih (1999). Fast approximate energy minimization via graph cuts. In *ICCV (1)*, pp. 377–384.
- Boykov, Y., O. Veksler, e R. Zabith (1998). A variable window approach to early vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20(12), 1283–1294.
- Brown, M. Z., D. Burschka, e G. D. Hager (2003). Advances in computational stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25(8), 993–1008.
- Chambon, S. (2005). *Mise en correspondance stéréoscopique d’images couleur en présence d’occultations*. Tese de Doutorado, Université Toulouse III - École Doctorale Informatique et Télécommunications.
- Chang, D. e K. V. Nesbitt (2006). Developing gestalt-based design guidelines for multi-sensory displays. In *MMUI ’05: Proceedings of the 2005 NICTA-HCSNet Multimodal User Interaction Workshop*, Darlinghurst, Australia, Australia, pp. 9–16. Australian Computer Society, Inc.
- Chang, D., K. V. Nesbitt, e K. Wilkins (2007). The gestalt principles of similarity and proximity apply to both the haptic and visual grouping of elements. In *AUIC ’07: Proceedings of the eight Australasian conference on User interface*, Darlinghurst, Australia, Australia, pp. 79–86. Australian Computer Society, Inc.
- Chung, K. (1997). A fast algorithm for stereo matching. *Information Processing Letters* 63(2), 57–61.
- Dankers, A., N. Barnes, , e A. Zelinsky (2004). Active vision - rectification and depth mapping. *Proceedings ACRA04: The Australasian Conference on Robotics and Automation*.
- Dhond, U. R. e J. K. Aggarwal (1989). Structure from stereo - a review. *IEEE Transactions on Systems, Man and Cybernetics* 19(6), 1489–1510.
- Duarte, M. H. V. (2002). *Codificação de Imagens Estéreo Usando Recorrência de Padrões*. Tese de Doutorado, Universidade Federal do Rio de Janeiro.
- Fitch, A. J., A. Kadryov, W. Christmas, e J. Kittler (2005). Fast robust correlation. In *IEEE Transactions on Image Processing*, Volume 14, pp. 1063–1073.

-
- Forsyth, D. A. e J. Ponce (2002). *Computer Vision: A Modern Approach*. Prentice Hall; US Ed edition.
- Fusiello, A., E. Trucco, e A. Verri (2000). A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications* 12(1), 16–22.
- Gong, M. e Y.-H. Yang (2003). Fast stereo matching using reliability-based dynamic programming and consistency constraints. In *ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision*, Washington, DC, USA, pp. 610. IEEE Computer Society.
- Hariyama, M., Y. Kobayashi, H. Sasaki, e M. Kameyama (2005). Fpga implementation of a stereo matching processor based on window-parallel-and-pixel-parallel architecture. In *Proceedings of 48th Midwest Symposium on Circuits and Systems, 2005.*, pp. 1219–1222. IEEE.
- Hartley, R. e A. Zisserman (2003). *Multiple View Geometry in Computer Vision* (Second ed.). 40 West 20th Street, New York, NY 10011-4211, USA: Cambridge University Press.
- Kanade, T. e M. Okutomi (1994). A stereo matching algorithm with an adaptive window: theory and experiment. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16(9), 920 – 932.
- Kang, S. B., R. Szeliski, e J. Chai (2001). Handling occlusions in dense multi-view stereo. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. (CVPR 2001)* 1(1), 103–110.
- Kim, J. C., K. M. Lee, B. T. Choi, e S. U. Lee (2005). A dense stereo matching using two-pass dynamic programming with generalized ground control points. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 05)*, Volume 2, Washington, DC, USA, pp. 1075–1082. IEEE Computer Society.
- Koschan, A. (1993a). A framework for area-based and feature-based stereo vision. In *Machine Graphics and Vision*, Volume 2, pp. 285–308.
- Koschan, A. (1993b). What is new in computational stereo since 1989: A survey on current stereo papers. Technical Report 93-22, Technische Universität Berlin.
- Lee, S., J. Yi, e J. Kim (2005). Real-time stereo vision on a reconfigurable system. In *Proceedings of SAMOS'2005 - Systems, Architectures, Modeling, and Simulation*, Volume 3553/2005 of *Lecture Notes in Computer Science*, pp. 299–307. Springer-Verlag Berlin Heidelberg.

- Mallon, J. e P. F. Whelan (2005). Projective rectification from the fundamental matrix. *Image and Vision Computing* 23(7), 643–650.
- McDonnell, M. J. (1981). Box-filtering techniques. In *Computer Graphics and Image Processing*, Volume 17, pp. 65–70.
- Middlebury Stereo Vision Page (2007). Acessado pela última vez no mês de outubro de 2007. In <http://vision.middlebury.edu/stereo/>, Vermont, USA.
- Moravec, H. (1977). Towards automatic visual obstacle avoidance. In *Proceedings of the 5th International Joint Conference on Artificial Intelligence*, MIT, Cambridge, Mass., pp. 584.
- Roma, N., J. Santos-Victor, e J. Tomé (2000). A comparative analysis of cross-correlation matching algorithms using a pyramidal resolution approach. In *ECCV2000 - 2nd Workshop on Empirical Evaluation Methods in Computer Vision*, Dublin, Ireland.
- Scharstein, D. e R. Szeliski (1998). Stereo matching with non-linear diffusion. *IJCV* 28(2), 155–174.
- Scharstein, D. e R. Szeliski (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision* 47(1-3), 7–42.
- Seitz, S. M., B. Curless, J. Diebel, D. Scharstein, e R. Szeliski (2006). A comparison and evaluation of multi-view stereo reconstruction algorithms. *CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 519–528.
- Shah, J. (1993). A nonlinear diffusion model for discontinuous disparity and half-occlusions in stereo. *Conference on Computer Vision and Pattern Recognition, Proceedings CVPR '93.*, 15–17.
- Shapiro, L. e G. Stockman (2001). *Computer Vision*. Upper Saddle River: Prentice-Hall.
- Shi, J. e J. Malik (2000). Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 22(8), 888–905.
- Stefano, L. D., M. Marchionni, e S. Mattoccia (2004a). A fast area-based stereo matching algorithm. *Proceedings from the 15th IAPR/CIPRS International Conference on Vision Interface* 22(12), 983–1005.
- Stefano, L. D., M. Marchionni, e S. Mattoccia (2004b). A pc-based real-time stereo vision system. In *Machine GRAPHIC and VISION*, Volume 13, pp. 197–220.

-
- Sun, C. (1997). A fast stereo matching method. *Digital Image and Vision Computing: Techniques and Applications (DICTA 1997)*, 95–100.
- Sun, C. (2002). Fast stereo matching using rectangular subregioning and 3d maximum-surface techniques. In *International Journal of Computer Vision*, Volume 47, Hingham, MA, USA, pp. 99–117. Kluwer Academic Publishers.
- Sun, J., N.-N. Zheng, e H.-Y. Shum (2003). Stereo matching using belief propagation. *IEEE Trans. Pattern Anal. Mach. Intell.* 25(7), 787–800.
- Sunyoto, H., W. van der Mark, e D. M. Gravila (2004). A comparative study of fast dense stereo vision algorithms. In *2004 IEEE Intelligent Vehicles Symposium*, Volume 14, pp. 319–324. IEEE.
- Sutton, M. A., S. R. McNeill, J. D. Helm, e Y. J. Chao (2000). *Advances in Two-Dimensional and Three-Dimensional Computer Vision*, Volume 77 of *Topics in Applied Physics, Photomechanics*, Chapter 26, pp. 323–372. Verlag Berlin Heidelberg: Springer.
- Thompson, C. (1975). Depth preception in stereo computer vision. Technical Report STAN-CS-7 5-521, ARPA - Advanced Research Project Agency.
- Étienne Vincent e R. Laganière (2002). An empirical study of feature point matching strategies. In *Proceedings of Conference Vision Interface*, Calgary, Canada, pp. 139–145.
- Tomasi, C. e R. Manduchi (1998). Stereo matching as a nearest-neighbor problem. *IEEE Trans. Pattern Anal. Mach. Intell.* 20(3), 333–340.
- Torr, P. e A. Criminisi (2004). Dense stereo using pivoted dynamic programming. In *Image and Vision Computing*, Volume 22, pp. 795–806. Elsevier B.V.
- Trucco, E. e A. Verri (1998). *Introductory Techniques for 3-D Computer Vision*. Upper Saddle River, NJ, USA: Prentice Hall PTR.
- Veksler, O. (2002). Stereo correspondence with compact windows via minimum ratio cycle. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(12), 1654–1660.
- Veksler, O. (2003). Fast variable window for stereo correspondence using integral images. *IEEE Proceedings of Computer Society Conference on Computer Vision and Pattern Recognition*. 01, 556–561.

- Venkateswar, V. e R. Chellappa (1995). Hierarchical stereo and motion correspondence using feature groupings. In *International Journal of Computer Vision*, Volume 15, Hingham, MA, USA, pp. 245–269. Kluwer Academic Publishers.
- Xu, Y., D. Wang, T. Feng, e H.-Y. Shum (2002). Stereo computation using radial adaptive windows. In *ICPR '02: Proceedings of the 16 th International Conference on Pattern Recognition (ICPR'02) Volume 3*, Washington, DC, USA, pp. 30595. IEEE Computer Society.
- Yip, R. K. K. e W. P. Ho (1998). A multi-level dynamic programming method for stereo line matching. *Pattern Recogn. Lett.* 19(9), 839–855.
- Yoon, K. e I. S. Kweon (2006). Adaptive support-weight approach for correspondence search. *IEEE Trans. Pattern Anal. Mach. Intell.* 28(4), 650.
- Zabih, R. e J. Woodfill (1994). Non-parametric local transforms for computing visual correspondence. In *Proceedings of the Third European Conference-Volume II on Computer Vision*, Volume 2, Secaucus, NJ, USA, pp. 151–158. Springer-Verlag.
- Zhang, W., Q. Zhang, L. Qu, e S. Wei (2004). A stereo matching algorithm based on multiresolution and epipolar constraint. *ICIG '04: Proceedings of the Third International Conference on Image and Graphics (ICIG'04)*, 180–183.
- Zhou, D., Q. Wu, Y. Liu, e X. Cai (2003). A stereo matching algorithm based on fuzzy identification. *Proceedings of 2003 IEEE International Conference on Robotics. Intelligent Systems and Signal Processing 2*, 1108–1113.