

Universidade Estadual de Campinas
Faculdade de Engenharia Elétrica e de Computação
Departamento de Engenharia de Computação e Automação Industrial

Reconstrução Tridimensional por Visão Estéreo e Detecção de Oclusões

Autor: Fernando Marques de Almeida Nogueira

Orientador: Clésio Luis Tozzi

Tese apresentada à Faculdade de Engenharia Elétrica e de Computação, FEEC - UNICAMP, como parte dos requisitos para obtenção do título de Doutor em Engenharia Elétrica. Área de Concentração: **Engenharia de Computação.**

Banca Examinadora

Prof. Dr. Antônio Maria Garcia Tommaselli.....FCT/UNESP
Prof. Dr. Edson Aparecido Mitishita.....UFPR
Prof. Dr. José Mario De Martino.....DCA/FEEC/UNICAMP
Prof. Dr. Léo Pini Magalhães.....DCA/FEEC/UNICAMP
Prof. Dr. Roberto Alencar Lotufo.....DCA/FEEC/UNICAMP

Campinas, SP
Dezembro de 2007

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA ÁREA DE ENGENHARIA E ARQUITETURA - BAE -
UNICAMP

N689r Nogueira, Fernando Marques de Almeida
Reconstrução tridimensional por visão estéreo e detecção
de oclusões / Fernando Marques de Almeida Nogueira. --
Campinas, SP: [s.n.], 2007.

Orientador: Clésio Luis Tozzi.
Tese (doutorado) - Universidade Estadual de Campinas,
Faculdade de Engenharia Elétrica e de Computação.

1. Visão binocular. 2. Visão de robô. 3. Visão por
computador. I. Tozzi, Clésio Luis. II. Universidade
Estadual de Campinas. Faculdade de Engenharia Elétrica e
de Computação. III. Título.

Título em Inglês: Tridimensional Reconstruction through Stereo Vision
and Occlusion Detection

Palavras-chave em Inglês: Binocular vision, Robot vision, Computer vision

Área de concentração: Automação

Titulação: Doutor em Engenharia Elétrica

Banca examinadora: Edson Aparecido Mitishita, Léo Pini Magalhães,
Antônio Maria Garcia Tommaselli, José Mario de
Martino, Roberto Alencar Lotufo

Data da defesa: 19/12/2007

Programa de Pós-Graduação: Engenharia Elétrica

COMISSÃO JULGADORA - TESE DE DOUTORADO

Candidato: Fernando Marques de Almeida Nogueira

Data da Defesa: 19 de dezembro de 2007

Título da Tese: "Reconstrução Tridimensional por Visão Estéreo e Detecção de Oclusão"

Prof. Dr. Clésio Luiz Tozzi (Presidente): _____

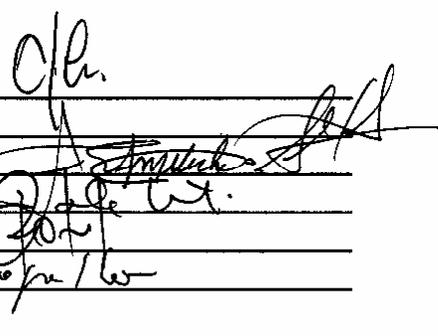
Prof. Dr. Antônio Maria Garcia Tommaselli: _____

Prof. Dr. Edson Aparecido Mitishita: _____

Prof. Dr. José Mario De Martino: _____

Prof. Dr. Roberto de Alencar Lotufo: _____

Prof. Dr. Léo Pini Magalhães: _____



The image shows handwritten signatures in black ink over the signature lines. The signatures are: 1. Clésio Luiz Tozzi (President), 2. Antônio Maria Garcia Tommaselli, 3. Edson Aparecido Mitishita, 4. José Mario De Martino, 5. Roberto de Alencar Lotufo, and 6. Léo Pini Magalhães. The signatures are written in a cursive style.

Agradecimentos

Ao Prof. Dr. Clésio Luis Tozzi pela orientação, amizade e paciência dedicada a mim durante a realização deste trabalho;

Aos demais professores do curso de Pós-Graduação da FEEC pelos ensinamentos;

Aos membros da banca examinadora: prof. Dr. Clésio Luis Tozzi, Prof. Dr. Antônio Maria Garcia Tommaselli, Prof. Dr. Edson Aparecido Mitishita, Prof. Dr. José Mario De Martino, Prof. Dr. Léo Pini Magalhães, Prof. Dr. Roberto Alencar Lotufo pela disponibilidade.

Aos funcionários da FEEC pelo apoio.

A Xuxu pelo incentivo, dedicação e paciência.

Ao meu amigo Maurício Galo pelas proveitosas discussões e sugestões.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq pela concessão de bolsa, durante o primeiro ano e meio de doutorado.

Dedico este trabalho as pessoas mais importantes na minha vida: minha esposa Lilian e minha filha Mariana e meus pais Marco Aurélio e Lúcia que sempre me apoiaram com amor, carinho e compreensão.

Resumo

Este trabalho apresenta uma nova abordagem iterativa e hierárquica para a obtenção de mapas densos de disparidade, a partir de um par de imagens estereoscópicas retificadas, na qual oclusões são explicitamente identificadas. Esta abordagem se fundamenta num processo de busca hierárquica onde são determinados pares homólogos prioritários sobre os quais é aplicada a Injunção de Ordem. Pares homólogos prioritários consistem nos pontos que apresentam a propriedade de Correspondência Bidirecional para o maior número de janelas de busca organizadas em ordem crescente de tamanho. A aplicação da Injunção de Ordem aos pares homólogos prioritários elimina, já nas primeiras iterações, uma grande quantidade de correspondências ambíguas aumentando o índice de acertos e reduzindo o tempo de processamento. A abordagem proposta destaca-se pela utilização de apenas dois parâmetros que devem ser fornecidos a priori de maneira empírica: o tamanho inicial da janela e o número de níveis hierárquicos. A eficiência do algoritmo proposto foi avaliada a partir de imagens reais e simuladas e os resultados comparados com outras abordagens descritas na literatura.

Palavras Chaves: Correspondência Estéreo, Oclusão, Visão Estéreo, Disparidade.

Abstract

This work presents a new hierarchical and iterative approach to obtain dense disparity maps from a pair of rectified stereoscopic images in which occlusions are explicitly identified.

This approach is based on a process of hierarchical searching where priority matches are determined and upon which is applied the Order Constraint. Priority matches consist of points that present a property of Bidirectional Correspondence for the greatest number of searching windows, these ones organized in a size ascending order. The Order Constraint application on the priority matches eliminates, usually after the first iterations, a lot of ambiguous matches, increasing the rate of true matches and reducing the execution time.

The proposed approach stands out by using only two parameters that must be provided a priori in an empirical way: the initial window size and the number of hierarchical levels.

The efficiency of the proposed algorithm was evaluated using real and simulated images and the results were compared to other approaches described in the literature.

Key Words: Correspondence Stereo, Occlusion, Stereo Vision, Disparity.

Sumário

AGRADECIMENTOS	v
RESUMO	ix
ABSTRACT	x
SUMÁRIO	xi
LISTA DE TABELAS	xiv
LISTA DE QUADROS	xv
LISTA DE FIGURAS	xvi
1 Introdução	1
1.1 O Problema de Reconstrução e de Correspondência.....	1
1.2 Dificuldades para a Solução do Problema de Correspondência.....	2
1.3 Abordagens Existentes e Categorização.....	3
1.4 Abordagem Proposta.....	9
1.5 Objetivo.....	13
1.6 Organização.....	13
2 Revisão	15
2.1 O Processo de Reconstrução Estéreo.....	15
2.1.1 Retificação.....	16
2.1.2 Estabelecimento da Correspondência.....	18
2.2 Fatores que Dificultam o Estabelecimento da Correspondência.....	25
2.2.1 Fatores Geométricos.....	26

2.2.2 Fatores Radiométricos.....	32
2.2.3 Condições de Textura e de Iluminação.....	34
2.2.4 Ruído.....	35
2.3 Matriz de Correspondência e Injunções.....	37
2.3.1 Matriz de Correspondência.....	37
2.3.2 Injunções e Seus Efeitos na Matriz de Correspondência.....	38
3 Trabalhos Relacionados	52
3.1 Trabalhos Disponíveis na Literatura.....	52
4 Abordagem Proposta para Solução do Problema de Correspondência	59
4.1 Detecção de Oclusão - Correspondência Bidirecional e Injunção de Ordem.....	60
4.2 Tamanho da Janela de Correlação - Correspondência Bidirecional Hierárquica.	68
4.3 Abordagem Proposta.....	71
4.3.1 Matriz de Prioridades e Processo Iterativo.....	71
4.3.2 Regras Complementares.....	72
5 Algoritmo e Implementação	79
5.1 Medida de Correspondência.....	79
5.2 Tabela para Implementação da Injunção de Ordem.....	81
5.3 Algoritmo.....	83
5.4 Exemplo.....	86
5.5 Implementação.....	89
6. Testes e Resultados	92
6.1 Metodologia de Testes.....	92
6.2 Resultados Quantitativos.....	94
6.2.1 Imagens Map.....	96
6.2.2 Imagens Sawtooth.....	99
6.2.3 Imagens Tsukuba.....	102

6.2.4	Imagens Vênus.....	105
6.3	Comparação dos Resultados com outras Abordagens.....	110
6.4	Resultados para Estereogramas de Pontos Randômicos.....	112
6.4.1	Imagens Retângulo.....	113
6.4.2	Imagens Redondo.....	115
6.4.3	Imagens Xadrez.....	117
6.5	Resultados Qualitativos.....	119
6.5.1	Imagens Corredor.....	119
6.5.2	Imagens Manequim.....	120
6.6	Discussão e Síntese dos Resultados.....	121
7	Conclusões e Recomendações	123
7.1	Síntese.....	123
7.2	Conclusões.....	124
7.3	Contribuições.....	125
7.4	Trabalhos Futuros.....	126
Anexo		127
A.1	Aquisição.....	127
A.2	Calibração.....	129
A.3	Triangulação.....	130
Referências bibliográficas		133

Lista de tabelas

Tabela 1 – Resultados para Imagens Map.....	98
Tabela 2 – Resultados para Imagens Sawtooth.....	101
Tabela 3 – Resultados para Imagens Tsukuba.....	104
Tabela 4 – Resultados para Imagens Venus.....	107
Tabela 5 – Resultados disponibilizados no <i>site</i>	111

Lista de Quadros

5.1	Exemplo de tabela para imposição da Injunção de Ordem.....	82
5.2	Efeito da imposição da Injunção de Ordem sobre (8,4) na Tabela e na Matriz de Prioridade.....	83

Lista de Figuras

2.1	Etapas do Processo Estéreo.....	16
2.2	Imagens Estéreo não retificadas.....	16
2.3	Imagens Estéreo retificadas.	17
2.4	Elementos da Geometria Epipolar.	18
2.5	Problema de Correspondência em Visão Estéreo.	20
2.6	Estabelecimento da Correspondência baseado em Feições.....	21
2.7	Estabelecimento da Correspondência baseado em Áreas, com janela de referência na imagem esquerda e janelas de pesquisa na imagem da direita.....	22
2.8.a	Imagem Estéreo Esquerda.....	25
2.8.b	Mapa de Disparidade Esquerdo.....	25
2.9	Superfície Fronto Planar: pontos são deslocados relativamente entre as imagens de mesmo valor de disparidade, no caso, $d = 3$	26
2.10	Superfície Não Fronto Planar: pontos são deslocados relativamente entre as imagens de valores distintos de disparidade.....	27
2.11	Exemplo de feição reta em uma imagem homóloga a uma feição curva na outra imagem.....	28
2.12	Janelas de correlação sobre regiões homólogas.	28
2.13	Janelas de correlação sobre regiões parcialmente homólogas.	29
2.14	Imagens esquerda e direita ilustrando efeitos das distorções perspectivas.....	29
2.15	Imagens esquerda e direita.	31
2.16	Imagens esquerda e direita com regiões de oclusão representada pelas sombras.	31
2.17	Modelo de Iluminação de Phong.....	33
2.18	Imagens estéreo com padrões periódicos.....	34
2.19	Imagens estéreo com texturas homogêneas.....	35
2.20	Imagem estéreo esquerda sem ruído e seu respectivo mapa de disparidade associado.	36
2.21	Imagem estéreo esquerda com 10 dB de ruído e seu respectivo mapa de	

	disparidade associado.....	36
2.22	Imagem estéreo esquerda com 20 dB de ruído e seu respectivo mapa de disparidade associado.....	37
2.23	Câmeras com $W = 7$	38
2.24	Matriz de Correspondência para um par de linhas epipolares homólogas.....	38
2.25	Matriz de Correspondência para $H = 3$	38
2.26	Efeito da Injunção de Disparidade Não Negativa.....	40
2.27	Injunção de Unicidade.....	41
2.28	Efeito conjunto da Injunção de Unicidade e de Disparidade Não Negativa na Matriz de Correspondência.....	41
2.29	Injunção de Ordem.....	43
2.30	Efeito conjunto da Injunção de Ordem e de Disparidade Não Negativa na Matriz de Correspondência.....	43
2.31	Sistema de coordenadas do espaço ciclopeano na Matriz de Correspondência..	45
2.32	Injunção do Limite de Gradiente de Disparidade com módulo = 2.....	47
2.33	Efeito conjunto da Injunção de Limite de Gradiente de Disparidade com módulo = 2 e de Disparidade Não Negativa.....	48
2.34	Efeito conjunto da Injunção de Limite de Gradiente de Disparidade com módulo = 1 e de Disparidade Não Negativa.....	49
2.35	Efeito conjunto da Injunção de Limite de Gradiente de Disparidade com módulo = 1 e de Disparidade Não Negativa, para uma Matriz de Correspondência 100 x 100.....	50
2.36	Efeito conjunto da Injunção de Limite de Gradiente de Disparidade com módulo = 2 e de Disparidade Não Negativa, para uma Matriz de Correspondência 100 x 100.....	50
2.37	Imagem estéreo esquerda com 20 dB de ruído e seu respectivo mapa de disparidade associado.....	51
4.1	Exemplo em que o par (6,2) é Correspondência Bidirecional na representação esquemática das câmeras e na respectiva Matriz de Correspondências.....	61
4.2	Exemplo em que o par (6,2) não é Correspondência Bidirecional na representação esquemática das câmeras e na respectiva Matriz de	

Correspondências.....	62
4.3 Oclusão na imagem da esquerda na representação esquemática das câmeras e efeito conjunto na Matriz de Correspondências da Injunção de Ordem e de Disparidade Não Negativa ocasionando oclusão na imagem da esquerda.....	66
4.4 Oclusão na imagem da direita na representação esquemática das câmeras e efeito conjunto na Matriz de Correspondências da Injunção de Ordem e de Disparidade Não Negativa ocasionando oclusão na imagem da direita.....	67
4.5 Medidas de correspondência para tamanho de janela = 1 (nível 1).....	74
4.6 Medidas de correspondência para tamanho de janela = 3 (nível 2).....	74
4.7 Matriz de Prioridades para as Matrizes de Correspondências das Figuras 4.5 e 4.6.....	74
4.8 Medidas de correspondências para tamanho de janela = 1 após Injunção de Ordem para pares com prioridade = 2.....	75
4.9 Medidas de correspondências para tamanho de janela = 3 após Injunção de Ordem para pares com prioridade = 2.....	75
4.10 Matriz de Prioridades para as Matrizes de Correspondências das Figuras 4.8 e 4.9.....	75
4.11 Medidas de correspondências para tamanho de janela = 1.....	76
4.12 Medidas de correspondências para tamanho de janela = 3.....	76
4.13 Matriz de Prioridades para as Matrizes de Correspondências das Figuras 4.11 e 4.12.....	77
4.14 Região de pares impossíveis determinada pela Injunção de Ordem sobre (6,2).	77
4.15 Região de pares impossíveis determinada pela Injunção de Ordem sobre (4,3) e (5,4).....	77
5.1 Fluxograma para <i>Main</i>	85
5.2 Fluxograma para <i>SetPrioridade</i>	86
5.3 Exemplo de Matriz de Correspondências para nível hierárquico $n = 1$ para imagem real.....	87
5.4 Exemplo de Matriz de Prioridade real.....	88
5.5 Matriz de Correspondências da Figura 5.3 após primeira iteração.....	88
5.6 Interface do programa computacional implementado.....	89

5.7	<i>Pixel</i> central e bordas vertical e lateral para janela de correlação 5 x 5.....	91
6.1	Conjunto de imagens Map: Imagens Estereoscópicas, Mapa de Disparidade Verdadeiro, Mapa de Oclusões, Mapa de Descontinuidade e Mapa de regiões sem Texturas.....	97
6.2	Resultados para imagens Map: Mapas de Disparidade reconstruídos com e sem oclusão e Mapa de pontos errados.....	98
6.3	Gáficos 3D para o mapa de disparidade reconstruído com $S_i^1 \times S_j^1 = 7 \times 7$ - imagens Map.....	99
6.4	Conjunto de imagens Sawtooth: Imagens Estereoscópicas, Mapa de Disparidade Verdadeiro, Mapa de Oclusões, Mapa de Descontinuidade e Mapa de regiões sem Texturas.....	100
6.5	Resultados para imagens Sawtooth: Mapas de Disparidade reconstruídos com e sem oclusão e Mapa de pontos errados.....	101
6.6	Gáficos 3D para o mapa de disparidade reconstruído com $S_i^1 \times S_j^1 = 7 \times 7$ - imagens Sawtooth.....	102
6.7	Conjunto de imagens Tsukuba: Imagens Estereoscópicas, Mapa de Disparidade Verdadeiro, Mapa de Oclusões, Mapa de Descontinuidade e Mapa de regiões sem Texturas.....	103
6.8	Resultados para imagens Tsukuba: Mapas de Disparidade reconstruídos com e sem oclusão e Mapa de pontos errados.....	104
6.9	Gáficos 3D para o mapa de disparidade reconstruído com $S_i^1 \times S_j^1 = 11 \times 11$ - imagens Tsukuba.....	105
6.10	Conjunto de imagens Venus: Imagens Estereoscópicas, Mapa de Disparidade Verdadeiro, Mapa de Oclusões, Mapa de Descontinuidade e Mapa de regiões sem Texturas.....	106
6.11	Resultados para imagens Venus: Mapas de Disparidade reconstruídos com e sem oclusão e Mapa de pontos errados.....	107
6.12	Gáficos 3D para o mapa de disparidade reconstruído com $S_i^1 \times S_j^1 = 13 \times 13$ - imagens Venus.....	108
6.13	Erro E.....	108

6.14	Erro ET.....	108
6.15	Erro ED.....	108
6.16	Erro EO.....	108
6.17	Tempo de Processamento T.....	109
6.18	Gráfico das abordagens existentes versus a somatória dos erros.....	112
6.19	Geração de Estereogramas de Pontos Randômicos, com textura A diferente de textura B, embora A e B sejam randômicas.....	113
6.20	Imagens Retângulo: Imagens Estereoscópicas, Mapas de Disparidade reconstruídos, Mapas de Confiança e gráficos tridimensionais para $s_i \times s_j = 1 \times 1$	114
6.21	Imagens Retângulo: Imagens Estereoscópicas, Mapas de Disparidade reconstruídos, Mapas de Confiança e gráficos tridimensionais para $s_i \times s_j = 41 \times 41$	115
6.22	Imagens Redondo: Imagens Estereoscópicas, Mapas de Disparidade reconstruídos, Mapas de Confiança e gráficos tridimensionais para $s_i \times s_j = 1 \times 1$	116
6.23	Imagens Redondo: Imagens Estereoscópicas, Mapas de Disparidade reconstruídos, Mapas de Confiança e gráficos tridimensionais para $s_i \times s_j = 39 \times 39$	116
6.24	Imagens Xadrez: Imagens Estereoscópicas, Mapas de Disparidade reconstruídos, Mapas de Confiança e gráficos tridimensionais para $s_i \times s_j = 1 \times 1$	117
6.25	Resultados para Imagens Xadrez em função do tamanho inicial da janela.....	118
6.26	Imagens Corredor : Imagens Estereoscópicas, Mapas de Disparidade reconstruídos, Mapas de Confiança e gráficos tridimensionais para $s_i \times s_j = 1 \times 1$	120
6.27	Imagens Manequim: Imagens Estereoscópicas, Mapas de Disparidade reconstruídos, Mapas de Confiança e gráficos tridimensionais para $s_i \times s_j = 9 \times 9$	121

A-1	Triângulo de reconstrução.....	131
-----	--------------------------------	-----

Capítulo 1

1. Introdução

1.1 O Problema de Reconstrução e de Correspondência

O Problema de Reconstrução consiste em determinar a estrutura tridimensional de uma cena a partir de uma ou mais imagens. Quando duas ou mais imagens da mesma cena tomadas a partir de posições distintas são utilizadas para a obtenção da estrutura tridimensional denomina-se Visão Estéreo.

A Visão Estéreo utiliza a disparidade entre dois pontos correspondentes situados em imagens distintas, denominados homólogos, como fonte de informação da estrutura tridimensional da cena. Dessa forma, a principal etapa a ser realizada para solucionar o Problema de Reconstrução por Visão Estéreo consiste em, dado um ponto em uma imagem, determinar o seu homólogo na outra imagem. Esta etapa é conhecida por Problema de Correspondência.

A consideração fundamental utilizada na Visão Estéreo é que a projeção de qualquer região suficientemente pequena da cena é pouco alterada relativamente entre as imagens tomadas em posições distintas. Com isso, em termos gerais, a solução do Problema de Correspondência é realizada através do estabelecimento da correspondência por áreas ou por feições, as quais utilizam como primitivas para correlação uma área (janela) ou atributos, respectivamente.

No processo de correspondência baseado em feições é possível gerar um mapa de disparidade esparsos, que permite apenas a reconstrução de pontos no referencial do espaço objeto associado às feições homólogas, enquanto a correspondência baseada em áreas possibilita gerar um mapa denso de disparidade. A obtenção de um mapa denso de disparidade a partir de um processo de correspondência baseado em feições é possível através de interpolação dos valores esparsos de disparidade.

A requisição de um mapa denso de disparidade é motivada por aplicações da Visão Estéreo, tais como inspeção industrial, robótica, cartografia, imagens médicas, entre outras, as quais requerem estimativas dos valores de disparidade para todas as regiões da imagem.

Independente do processo de correspondência utilizado existe nas imagens pontos de oclusão e para estes os valores de disparidade não podem ser determinados. Pontos de oclusão ocorrem quando um ponto da cena é visível apenas em uma das imagens e são decorrentes das posições distintas das câmeras e das variações de profundidade da cena. Assim, um mapa de disparidade denso deve possuir valores de disparidades para todos os seus pontos, exceto para as regiões do mapa associadas aos pontos de oclusão, que devem ser devidamente rotuladas como tais.

A identificação dos pontos de oclusão se faz importante a fim de minimizar a ocorrência de falsas correspondências entre as imagens estereoscópicas (tais pontos não devem ser considerados no Problema de Correspondência), além de serem importantes fontes de informação para a identificação das regiões do mapa de disparidade onde se observam descontinuidades.

No passado, as pesquisas sobre reconstrução de mapas densos de disparidade tinham por objetivo geral a determinação precisa dos valores de disparidades, embora não inteiramente para todos os pontos do mapa. No entanto, enquanto valores de disparidades imprecisos podem ainda produzir imagens sintetizadas aceitáveis para as regiões de pontos homólogos, regiões de oclusão inconsistentes em geral derivam erros inaceitáveis.

1.2 Dificuldades para a Solução do Problema de Correspondência

O Problema de Reconstrução Estéreo é um problema de visão difícil de ser resolvido devido a vários fatores que prejudicam a resolução do Problema de Correspondência. Em termos gerais, pode-se destacar os seguintes fatores:

- Distorções geométricas;
- Distorções radiométricas;

- Ruídos;
- Superfícies com texturas transparentes ou homogêneas; e
- Descontinuidades entre valores de profundidade da cena.

As distorções geométricas e radiométricas que ocorrem de maneira relativa entre as imagens estereoscópicas, de tal maneira que regiões homólogas podem apresentar formas e valores de intensidades diferentes entre si, dificultam o estabelecimento da correspondência entre essas regiões.

Os processos ruidosos ocorrem de maneira independente para cada imagem devido a variações nas condições de iluminação (para imagens tomadas em instantes distintos), degradações por borramento e características inerentes dos sensores. Dessa forma, uma região ou feição em uma imagem pode apresentar valores de intensidade distintos dos valores de intensidade da região ou feição homóloga, comprometendo os resultados das correspondências.

A presença na cena de objetos com superfícies transparentes torna impossível a reconstrução da geometria dos mesmos, uma vez que esta informação não é disponível nas imagens. Já a presença na cena de objetos com texturas homogêneas (texturas com variância nula) resulta em soluções ambíguas para o Problema de Correspondência.

A presença na cena de descontinuidades dos valores de profundidade, tais como regiões entre dois objetos distintos, ocasionam regiões de oclusão, isto é, regiões que são visíveis em apenas uma das imagens estereoscópicas. Para os pontos das regiões de oclusão a correspondência não pode ser estabelecida, sendo que, se tais pontos não forem desprezados podem originar falsas correspondências.

1.3 Abordagens Existentes e Categorização

O Problema de Correspondência pode ser abordado de várias formas. As diferentes abordagens existentes utilizam-se de uma ou mais técnicas para superar as dificuldades citadas no item 1.2.

Em geral, as abordagens existentes podem ser classificadas em métodos **locais** ou **globais**.

Os métodos locais utilizam apenas a correspondência baseada em áreas (ou feições) para determinar o valor de disparidade associado a cada *pixel* da imagem de referência. A principal vantagem dos métodos locais é que estes são relativamente simples de serem implementados. Outra vantagem é o baixo tempo de processamento que pode ser obtido utilizando janelas de correlação pequenas ou feições que são facilmente segmentadas. A principal desvantagem desses métodos é a baixa precisão que em geral é observada. No entanto, em ambientes controlados é possível manter as condições de iluminação e posicionamento das câmeras em condições estáveis, bem como, cenas simples com poucos objetos, os resultados podem apresentar boa precisão.

Os métodos globais utilizam em geral uma função-objetivo, que deve ser otimizada, cujo domínio é o mapa de disparidade sendo que todos os valores de disparidades são determinados simultaneamente. A principal vantagem dos métodos globais é a robustez da solução, porém o tempo de processamento é geralmente superior em relação aos métodos locais. Outra grande dificuldade desses métodos está em modelar a própria função-objetivo, uma vez que não existe um modelo matemático já consagrado para este problema. Em geral, os modelos matemáticos utilizados são não-convexos, e, portanto, apresentam mínimos locais, o que torna difícil a obtenção da solução ótima pelos métodos de otimização existentes.

Em termos mais específicos, algumas abordagens classificadas como métodos globais são os métodos baseados em Regularização, Processos Cooperativos, Programação Dinâmica, Fluxo Máximo em Grafos, *Layered* e Segmentos.

Regularização (Poggio & Kock, 1985; Poggio et al., 1985; Horn, 1986; Terzopoulos, 1986; Bertero et. al., 1988) é uma técnica matemática utilizada geralmente para estabilização de problemas inversos mal-colocados (*ill-posed*) através da quantificação de suavidade aliada a demais objetivos em um processo de otimização. Em Visão Estéreo, esta técnica é normalmente empregada através da formulação de um funcional que deve ser minimizado. Uma dificuldade encontrada em utilizar técnicas de Regularização para resolver o Problema de Correspondência é a representação de oclusões, uma vez que todo ponto da superfície regularizada possui um valor de disparidade associado. Outra dificuldade destes métodos consiste em determinar valores adequados para os parâmetros

que estabelecem as relações entre os termos de suavidade e demais termos modelados no funcional. Tais parâmetros são determinados geralmente de maneira empírica.

Os métodos baseados em **Processos Cooperativos** (Marr & Poggio, 1976 e Zitnick & Kanade, 2000; Mayer, 2003) são inspirados em sistemas nervosos biológicos e implementam diretamente as Injunções de Continuidade e Unicidade em um sistema iterativo e localmente conectado. Estes métodos operam diretamente em um espaço de correspondência, ao invés do espaço-imagem. O espaço de correspondência é uma matriz 3D de pesos com valores contínuos parametrizado de várias maneiras. Tipicamente, um sistema de coordenadas (j_e, i_e, d) é utilizado, com (j_e, i_e) representando a posição (coluna, linha) na imagem da esquerda (adotada como referência arbitrariamente sem perda de generalidade) e d representando a disparidade. Considerando imagens retificadas, uma parametrização possível deste espaço é (j_e, j_d, i) com (j_e, j_d) representando as coordenadas de um ponto na imagem esquerda e de seu candidato a homólogo na imagem da direita na direção das colunas das imagens respectivamente, e i a coordenada destes pontos na direção das linhas da imagem. Qualitativamente, o peso em (j_e, j_d, i) representa a verossimilhança com que o ponto de coordenadas (j_e, i) na imagem esquerda e o ponto (j_d, i) na imagem da direita correspondem um com o outro. Inicialmente este espaço 3D recebe medidas de similaridades locais, geralmente oriundas de processos de correlação com pequenas janelas quadradas. Subseqüentemente, estas medidas são alteradas iterativamente: se a medida presente em (j_e, j_d, i) possui um alto valor, os pontos para região inibitória são subtraídos por um valor proporcional ao valor presente em (j_e, j_d, i) e os pontos para região excitatória são adicionados por este mesmo valor. A região inibitória é geralmente definida como a região de pares impossíveis para a Injunção de Unicidade e a região excitatória é geralmente definida como a região de pares que recebem reforço para a Injunção de Continuidade. No entanto, outras formas para estas regiões podem ser definidas. Este processo de alterar as medidas de correspondência das regiões excitatórias e inibitórias é realizada de maneira iterativa até que tais alterações sejam desprezíveis. Após a convergência (estabilização) do processo iterativo, um limiar é imposto sobre os pesos com valores reais a fim de binarizá-los para a obtenção do mapa de disparidade. Se todos os pesos associados a um *pixel* em uma imagem possuem valores inferiores ao limiar imposto, este *pixel* é classificado como oclusão.

A principal vantagem destes métodos está em sua simplicidade de implementação, uma vez que o algoritmo atua de maneira auto-organizável. Por outro lado, a principal desvantagem destes métodos é a suavidade em demasia geralmente observada nos mapas de disparidades reconstruídos, principalmente nas regiões de descontinuidades, além da necessidade de definição do limiar citado.

Programação Dinâmica (Ohta & Kanade, 1985 e Belhumeur, 1993; Forstmann et al., 2004; Kim et al., 2005; Veksler, 2005) é um método matemático que reduz a complexidade computacional de problemas de otimização através da decomposição do problema original em subproblemas menores e mais simples. Uma função de custo é computada em estágios, com a transição entre estágios definida por um conjunto de restrições. Para o Problema de Correspondência, a Injunção de Ordem permite que a função de custo global seja determinada como o caminho de custo mínimo no espaço de correspondência (como definido nos métodos baseados em Processos Cooperativos). O custo do caminho ótimo é a soma dos custos dos caminhos parciais obtidos recursivamente. A exemplo dos métodos Cooperativos, os métodos baseados em Programação Dinâmica também operam em um espaço de disparidade discretizado favorecendo pares de pontos homólogos que respeitem as injunções de Ordem e Continuidade. Entretanto, enquanto os métodos Cooperativos são iterativos e encontram um conjunto ótimo localmente de pesos com valores reais que precisam então ser limiarizados a fim de definir o mapa de disparidade, a Programação Dinâmica é não iterativa e encontra um conjunto ótimo globalmente de pesos com valores binários que diretamente codificam a presença ou ausência de cada par de pontos homólogos. De maneira análoga aos métodos baseados em Processos Cooperativos, se todos os pesos associados a um *pixel* em uma imagem possuem valores nulos (pesos binários), este *pixel* é classificado como oclusão.

A principal dificuldade desta técnica está em determinar a função-objetivo e conseqüentemente seus parâmetros a fim de ser utilizada na Programação Dinâmica.

Assim como nos métodos baseados em Programação Dinâmica, os métodos baseados em **Fluxo Máximo em Grafos** (Roy & Cox, 1998; Ishikawa & Geiger, 1998; Boykov et al., 2001; Kolmogorov & Zabih, 2001; Kolmogorov & Zabih, 2002) utilizam técnicas de otimização combinatória, porém permitindo interações entre linhas epipolares. Estes métodos determinam a superfície de disparidade como o corte de custo mínimo do

problema de Fluxo Máximo (Ford & Fulkerson, 1962)) em um grafo cujos arcos possuem capacidade finita. Em geral, há duas classes de métodos baseados em Grafos. Os métodos que computam o mínimo global de um funcional de energia convexo com um único corte de custo mínimo e os métodos que computam apenas um mínimo local de um funcional de energia não-convexo com corte de custo mínimo iterativo. A determinação dos *pixels* homólogos e de oclusões é realizada da mesma maneira que os métodos baseados em Programação Dinâmica.

A principal dificuldade destes métodos está em definir o funcional de energia, uma vez que ainda não existe um modelo consagrado de funcional para o Problema de Correspondência.

Os métodos *Layered* (Darrell & Pentland, 1995; Lin & Tomasi, 2002) estimam a estrutura da cena como uma coleção de superfícies suaves. As disparidades contidas em cada dessas superfícies são modeladas por *splines* com valores contínuos, enquanto a extensão de cada superfície é representada por uma segmentação (rotulação) das imagens estereoscópicas. As disparidades e extensões das superfícies são alternadamente estimadas por ajustamento de superfícies e corte de grafos, respectivamente, em um processo de minimização de energia. As imagens estereoscópicas são tratadas simetricamente e oclusões são explicitamente modeladas. A exemplo dos demais métodos globais, a principal dificuldade destes métodos está em definir a função de energia a ser minimizada, bem como designar valores adequados para os parâmetros utilizado no processo de segmentação.

Recentemente os métodos baseados em **Segmentos** (Hong & Chen, 2004; Bleyer & Gelautz, 2005; Deng et. al., 2005) tem despertado atenção devido aos bons resultados obtidos. Estes consideram que a estrutura da cena pode ser aproximada por um conjunto de planos não superpostos no espaço de disparidade e que cada plano está associado com no mínimo uma região com cor ou intensidade homogênea segmentada na imagem de referência. Estes métodos geralmente realizam quatro etapas distintas. Na primeira etapa, regiões de cores ou intensidades homogêneas são localizadas na imagem de referência através de algum método de segmentação. Na segunda etapa, um método de correspondência baseado em áreas é utilizado para determinar disparidades. Na terceira etapa é utilizada alguma técnica de ajustamento de plano no espaço de disparidades.

Somente medidas de disparidades confiáveis (sob alguma consideração) associadas a cada região segmentada na imagem de referência são utilizadas. Dessa maneira, é associado um plano no espaço de disparidade a cada região segmentada na imagem de referência. Na quarta etapa, um processo de rotulação ótima é aproximado através de alguma técnica de otimização local a fim de agrupar regiões vizinhas com planos associados semelhantes. Regiões da imagem que possuem apenas medidas de disparidades não confiáveis são classificadas como oclusões.

A principal dificuldade destes métodos está em designar valores adequados para uma grande quantidade de parâmetros requeridos nos processos de segmentação das regiões homogêneas, ajustamento dos planos e rotulação ótima.

Discussão

De maneira cronológica, os métodos locais foram inicialmente propostos seguidos dos métodos globais. Naturalmente, os métodos mais recentes têm apresentado melhores resultados em comparação aos métodos propostos anteriormente.

No entanto, percebe-se que paralelamente a esta evolução, é cada vez maior o número de parâmetros utilizados nas abordagens propostas. Em geral, tais parâmetros são determinados de maneira empírica além de serem cada vez mais dependentes das características das imagens e/ou das cenas. Com isso, um conjunto de valores designados para os parâmetros utilizados em um par de imagens estereoscópicas que resultaram em um mapa de disparidade com ótima reconstrução podem resultar em um péssimo mapa de disparidade reconstruído apenas considerando um outro par de imagens estereoscópicas para a mesma abordagem.

O fato citado acima coloca em questão a utilização das abordagens citadas que possuem uma grande quantidade de parâmetros determinados empiricamente e extramente sensíveis as condições e características das imagens.

Entende-se que uma boa abordagem é aquela que obtém bons resultados para um grande número de imagens estereoscópicas com características bastante distintas entre elas e que utilizam valores relativamente constantes para os parâmetros empregados na abordagem, independente do número de parâmetros que possua.

Ainda, em relação à identificação de oclusões, esta é diretamente dependente de um limiar (parâmetro) nos métodos citados acima, o qual em geral, é determinado de maneira também empírica.

1.4 Abordagem Proposta

Considerando as dificuldades dos métodos globais para resolução do Problema de Correspondência, este trabalho propõe uma abordagem baseada em um processo hierárquico e número reduzido de parâmetros de controle e que, mesmo considerando um tratamento local, obtém resultados equiparáveis aos obtidos por complexos sistemas globais, possibilitando adicionalmente a determinação de oclusões.

Inicialmente, utiliza-se a correspondência baseada em áreas para obter as medidas de correlação entre os *pixels* de um par de linhas epipolares homólogas. Estas medidas são realizadas para vários tamanhos de janelas de correlação e armazenadas em uma Matriz de Correspondências. Embora na implementação realizada tenha sido utilizada a Somatória das Diferenças Absolutas (SDA) como função de correlação, é importante destacar a independência da abordagem proposta em relação a função de correlação.

Um procedimento denominado Correspondência Bidirecional Hierárquica é utilizado para definir prioridades para todas as correspondências possíveis. Um par de *pixels* possui a propriedade de Correspondência Bidirecional se a melhor medida de correspondência para um *pixel* na imagem da esquerda é obtida para um *pixel* na imagem na direita e vice-versa. A prioridade de um par de *pixels* que possui a propriedade de Correspondência Bidirecional para um dado tamanho de janela de correlação é aumentada somente se este mesmo par possui esta mesma prioridade para um tamanho de janela de correlação imediatamente superior em uma lista de tamanhos de janelas organizadas em ordem crescente.

Os pares de *pixels* com maiores prioridades são classificados como homólogos e então a Injunção de Ordem é imposta sob estes, excluindo com isso, todos os pares de *pixels* que não satisfazem esta injunção. As prioridades são então redefinidas para os pares ainda possíveis em um processo iterativo até que não exista mais pares possíveis.

Este procedimento é então realizado novamente para outro par de linhas epipolares homólogos até que todos os pares de linhas epipolares sejam resolvidos.

Categorização da Abordagem Proposta

A abordagem proposta opera de maneira **local** entre os pares de linhas epipolares homólogos, uma vez que os pares de linhas epipolares homólogos são tratados de maneira independente entre si e de maneira **global** entre os *pixels* sobre um mesmo par de linhas epipolares homólogas. Dessa forma, esta abordagem não pode ser classificada formalmente com um método local ou global. Entende-se que o correto é classificar a abordagem proposta em uma nova categoria, proposta neste trabalho, como **métodos híbridos**.

A abordagem proposta neste trabalho pode ser considerada também um método baseado em Processos Cooperativos, uma vez que a Injunção de Ordem utilizada define uma região inibitória e uma região excitatória, porém de maneira booleana, excluindo ou não pares de *pixels* candidatos a pares homólogos.

Justificativas

Como já citado, a abordagem proposta utiliza-se de injunções para determinação dos pontos homólogos e exclusão dos pontos que não se enquadram nesta condição.

A Injunção de Ordem é também conhecida como Injunção de Monotonicidade sendo que para *pixels* ao longo de uma linha epipolar em uma imagem, os *pixels* homólogos têm que ocorrer na mesma ordem sobre a linha epipolar homóloga na outra imagem do par de imagens estereoscópicas.

Sob esta injunção, oclusões ocorrem em uma imagem entre dois *pixels* cujos homólogos são vizinhos na outra imagem e que apresentem valores de disparidade diferentes. O tamanho da região de oclusão é igual à variação da disparidade naquela região.

A título de esclarecimento, cabe observar que neste trabalho considera-se imagens estereoscópicas retificadas, de tal maneira que as linhas epipolares homólogas correspondam as mesmas linhas de rasterização das imagens, fazendo com que exista disparidade apenas nesta direção. Apesar de não ser comum, é possível retificar as imagens estereoscópicas de tal maneira que as linhas epipolares homólogas estejam paralelas às

colunas das imagens. Nesse caso, a relação geométrica existente entre regiões de oclusão e variação da disparidade também se verifica, porém os valores de disparidade ocorrem na direção das colunas das imagens.

A abordagem proposta considera como informação para a resolução do Problema de Correspondência apenas o par de imagens estereoscópicas. Nenhuma outra informação, tal como, suavidade das superfícies constituintes da cena, posição das fontes de iluminação, entre outras é considerada. Sob estas condições, a utilização de algum procedimento que realize interações entre pares de linhas epipolares torna-se meramente "especulativo", uma vez que não existe nenhuma informação nas imagens estereoscópicas que possibilite inferir a maneira que tais interações devam ser utilizadas. Em termos práticos, o valor de disparidade associado a um *pixel* sobre uma linha epipolar em uma imagem é totalmente independente (não é função) do valor de disparidade associado ao *pixel* vizinho imediatamente acima ou abaixo deste.

Com isso, entende-se que os pares de linhas epipolares podem ser resolvidos de maneira independente. As abordagens que realizam interações entre pares de linhas epipolares homólogas comumente consideram que os valores de disparidade entre estas linhas não devem variar abruptamente, o que em geral, pode acarretar mapas de disparidade com suavidade em demasia.

A propriedade de Correspondência Bidirecional desempenha papel fundamental na abordagem proposta, uma vez que auxilia no processo de determinação do ponto homólogo através da redução do número de possíveis soluções ambíguas, bem como na identificação dos pontos de oclusão, uma vez que, para estes pontos os valores de prioridades associados tendem a serem mínimos.

A justificativa para a utilização de um procedimento hierárquico é que se visa reduzir a dependência do processo de correlação em relação ao tamanho da janela empregada tornando o procedimento de correlação invariante com relação ao aumento do tamanho da janela.

O procedimento hierárquico é do tipo "*bottom-up*" (janelas pequenas para janelas grandes) ao contrário do tipo "*Top-Down*" muitas vezes utilizado em abordagens mais antigas. A utilização deste tipo de hierarquia visa reconstruir melhor as possíveis regiões delgadas (estreitas ou finas) e de descontinuidades no mapa de disparidade, uma vez que as

informações oriundas da correspondência com janelas grandes tendem a ser ruins para estas regiões.

Para as regiões suaves do mapa de disparidade, as informações oriundas da correlação com grandes janelas são boas e principalmente necessárias nas regiões das imagens estereoscópicas com texturas homogêneas. Dessa forma, o procedimento iterativo designa inicialmente valores de prioridade grande para os pares de *pixels* homólogos associados a pontos do mapa de disparidade situados em regiões suaves, deixando para posteriori a determinação dos pontos do mapa de disparidade em regiões delgadas ou de descontinuidade.

O fato citado acima é uma característica importante da abordagem proposta porque a ocorrência de um par de *pixels* homólogos determina a exclusão de vários outros pares de *pixels* até então candidatos a homólogos, facilitando a desambiguação do Problema de Correspondência, além de diminuir consideravelmente o tempo de processamento.

Discussão

Ao contrário da maioria das abordagens globais existentes, o algoritmo proposto não utiliza uma função-objetivo explícita que deve ser minimizada a fim de obter uma solução global, mas sim, um Processo Cooperativo restrito a um par de linhas epipolares homólogas. Este fato é importante porque a utilização de uma função-objetivo que pode ser apropriada para um tipo de imagem/cena pode não ser para outro tipo de imagem/cena.

Na abordagem proposta existem apenas dois parâmetros que devem ser definidos pelo usuário. Um parâmetro é o tamanho inicial da janela de correlação utilizado. Os demais tamanhos da janela de correlação são definidos em função deste valor inicial. O outro parâmetro é o número de tamanhos distintos de janelas de correlação.

Para fins práticos, este último parâmetro citado pode ser fixado como um valor constante grande (acima de 5 níveis) na abordagem não sendo então necessário a sua determinação por parte do usuário. O único prejuízo desta consideração é um possível aumento desnecessário no tempo de processamento sem influência na qualidade da solução obtida.

Embora ainda exista a necessidade da definição do parâmetro que determina o tamanho inicial da janela de correlação, este é bem mais simples de ser determinado e

entendido pelo usuário mesmo leigo no assunto, do que os demais parâmetros comumente utilizados em outras abordagens, tais como, parâmetros que definem a "quantidade de suavidade" na função-objetivo, o tamanho mínimo e/ou máximo de uma região com textura homogênea em um processo de segmentação/rotulação, limiares para segmentação de regiões com texturas homogêneas, entre outros.

1.5 Objetivo

O objetivo deste trabalho consiste em desenvolver, avaliar e implementar na forma de um algoritmo computacional a abordagem introduzida na seção 1.4 para a resolução do Problema de Correspondência Estéreo, considerando um par de imagens estereoscópicas retificadas como única fonte de informação.

O mapa de disparidade resultante deve ser denso de tal modo que exista um valor de disparidade associado a cada *pixel* das imagens estereoscópicas, com exceção daqueles que são visíveis apenas em uma das imagens estereoscópicas. Estes *pixels* devem ser classificados como oclusões.

1.6 Organização

Este trabalho é composto de sete capítulos, incluindo esta introdução.

No capítulo 2 é apresentado a revisão de temas referentes a formação da imagem, estabelecimento da correspondência, injunções, entre outros.

O capítulo 3 apresenta de maneira sucinta, os trabalhos relacionados com o tema principal deste trabalho.

O capítulo 4 trata da solução do problema de correspondência preservando discontinuidades e identificando regiões de oclusão através da abordagem proposta.

No capítulo 5 são apresentados detalhes do algoritmo proposto e da implementação.

No capítulo 6 são apresentados os resultados dos testes realizados de maneira quantitativa e qualitativa para imagens reais e sintéticas. A metodologia de avaliação utilizada para os resultados quantitativos é também apresentada e discutida.

O capítulo 7 traz as conclusões sobre este trabalho.

O Anexo apresenta alguns assuntos de interesse para a Visão Estéreo, mas não essenciais para a compreensão deste trabalho.

Capítulo 2

2. Revisão

O objetivo deste capítulo é apresentar a fundamentação sobre os principais conceitos envolvidos em Visão Estéreo, mesmo que alguns destes conceitos não estejam diretamente envolvidos na abordagem proposta.

Inicialmente são apresentados conceitos básicos de acordo com as etapas envolvidas na recuperação da informação tridimensional da cena a partir de imagens estereoscópicas. Os conceitos sobre cada uma dessas etapas são descritos de maneira resumida. O leitor já familiarizado com o tema pode, sem perda de seqüenciamento, prosseguir a leitura a partir da seção 2.3.

Após a descrição das etapas, a revisão segue com temas mais diretamente relacionados com o problema tratado neste trabalho.

2.1 O Processo de Reconstrução Estéreo

A Visão Estéreo é uma técnica capaz de recuperar a informação tridimensional da cena a partir de duas ou mais imagens tomadas em posições distintas. No entanto, para isto, esta técnica necessita que várias etapas sejam realizadas. Neste trabalho a Visão Estéreo foi dividida em cinco etapas:

- Aquisição
- Calibração
- Retificação
- Estabelecimento das Correspondências
- Triangulação

As etapas aqui propostas estão ordenadas de maneira seqüencial, de tal forma que, a realização de uma etapa anterior é precedente para a realização da etapa posterior, salvo em algumas aplicações onde, por exemplo, utiliza-se servo-motores para controle do estado da câmera no instante da tomada, permitindo com isso a não realização da etapa de Calibração e até mesmo a de Retificação. A Figura 2.1 mostra estas etapas.

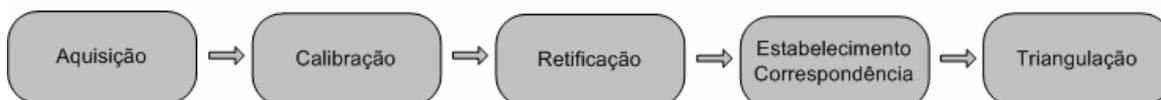


Fig. 2.1 - Etapas do Processo Estéreo.

Uma vez que este trabalho considera apenas imagens estereoscópicas retificadas, este capítulo aborda, em caráter de revisão, apenas a etapa de Retificação e Estabelecimento da Correspondência. As demais etapas são descritas em Anexo, para fins de complementação do assunto.

2.1.1 Retificação

O espaço de busca do ponto homólogo é, a princípio, toda a imagem, ou seja, um espaço bidimensional. A Figura 2.2 mostra um exemplo no qual, um ponto A e B na imagem da esquerda possuem homólogos em A' e B' na imagem da direita.

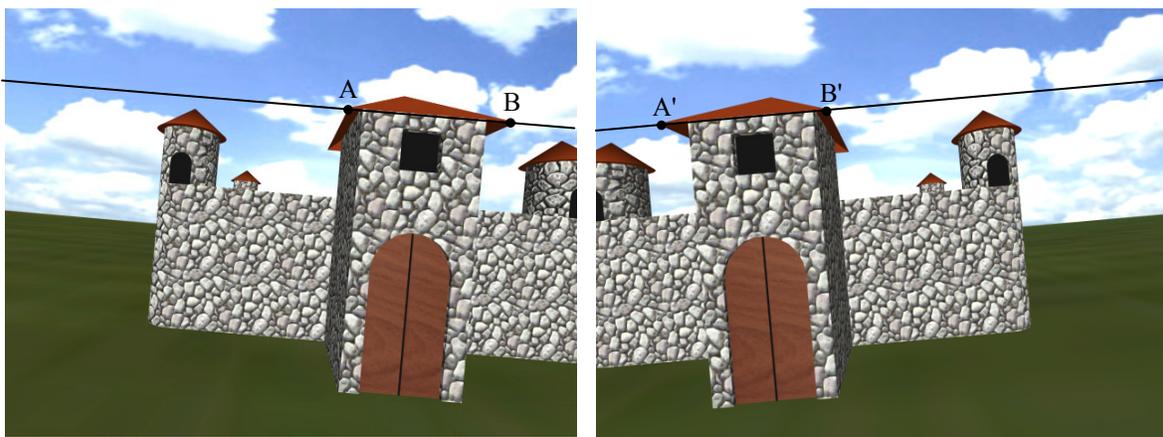


Fig. 2.2 – Imagens Estéreo não retificadas.

Pode-se perceber que A não possui a mesma coordenada na direção y (linha) de que A'. O mesmo se verifica para o par B e B'. Assim, para o ponto A, por exemplo, faz-se necessário pesquisar em toda a imagem da direita a posição de A'. O mesmo é válido para B e B'.

No entanto, de posse dos parâmetros de orientação das câmeras (coordenadas X, Y e Z do centro perspectivo, ângulos de rotação em torno dos eixos X, Y e Z, distância focal, entre outros) é sempre possível realizar o processo de busca do ponto homólogo em um espaço unidimensional aplicando-se os conceitos de Geometria Epipolar e Retificação Estéreo.

O processo de Retificação também conhecido como Normalização de Imagens (*Image Normalization*) Estereoscópicas ou Retificação entre Pares de Imagens (*Pair-Wise Retification*) é a próxima etapa a ser realizada após o processo de Calibração. No entanto, a não realização desta etapa não compromete a realização das demais etapas. A função do processo de Retificação é a de apenas facilitar a próxima etapa denominada Estabelecimento da Correspondência, que por sua vez, é a etapa mais difícil de todo o processo. A Figura 2.3 mostra as mesmas imagens da Figura 2.2, porém retificadas.

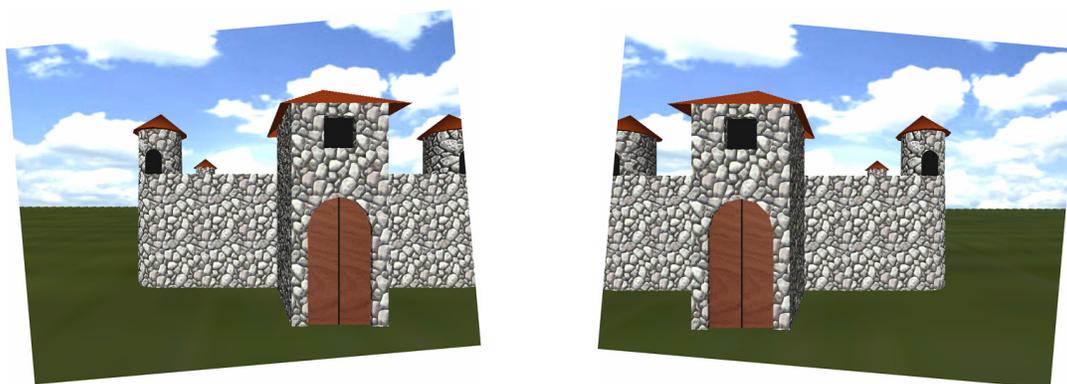


Fig. 2.3 – Imagens Estéreo retificadas.

O processo de Retificação das imagens estéreo utiliza-se da Geometria Epipolar existente no sistema estéreo para tornar as linhas das imagens estereoscópicas paralelas as linhas epipolares homólogas.

A geometria epipolar é baseada no fato de que um ponto (P) no espaço objeto, a projeção deste ponto (p_E e p_D) nos planos imagem (π_{IE} e π_{ID}) da esquerda e da direita,

respectivamente, bem como os centros perspectivos de ambas as câmeras (CP_E e CP_D) são coplanares, como mostra a Figura 2.4.

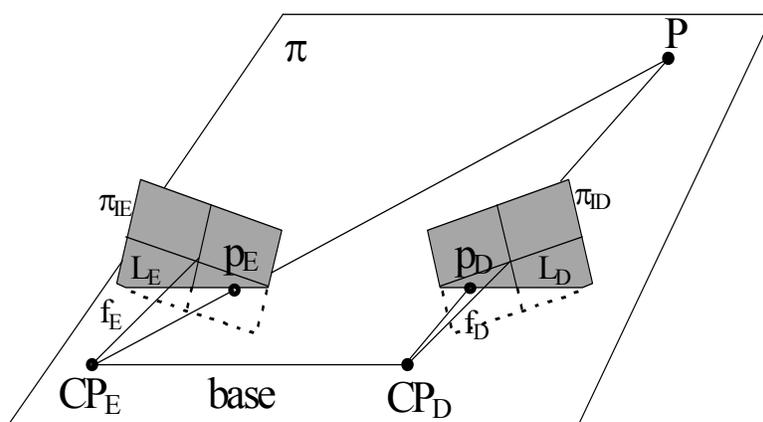


Fig. 2.4 – Elementos da Geometria Epipolar.

O plano definido por estes pontos é denominado “plano epipolar” (π) e sua interseção com os planos imagens (π_{IE} e π_{ID}) gera as “linhas epipolares homologas” (L_E e L_D). Desta maneira, qualquer ponto sobre uma linha epipolar em uma imagem está sobre a linha epipolar homóloga na outra, o que torna o problema de correspondência unidimensional.

A Retificação consiste então em reamostrar o par de imagens estereoscópicas de modo que as linhas epipolares se tornem coincidentes às linhas das imagens. O processo de reamostragem consiste em utilizar algum critério de interpolação nos níveis de cinza das imagens, como por exemplo, interpolação bilinear, splines, entre outras.

A Retificação pode ser modelada por uma transformação que por sua vez é função dos parâmetros obtidos na etapa de Calibração. A condição principal utilizada nesta transformação é que o plano sobre o qual serão geradas as imagens retificadas seja paralelo à base (linha entre os centros perspectivos das câmeras).

2.1.2 Estabelecimento da Correspondência

O Problema de Correspondência ou *Matching* (termo em inglês, porém muito utilizado também na literatura especializada em português) consiste em associar um objeto de referência (padrão, dado, atributo) a um ou mais objeto(s) de pesquisa(s) sob algum

critério de correspondência. Ainda é possível não associar um objeto de referência a algum objeto de pesquisa, caso o objeto de referência não possua homólogo.

O critério de correspondência pode ser entendido como uma regra, sistema ou função que representa o conhecimento necessário para estabelecer as associações entre os objetos de referência e de pesquisa com sucesso.

O Problema de Correspondência pode ser encontrado em inúmeras aplicações como, por exemplo, em:

- Testes de DNA: a correspondência é estabelecida entre duas ou mais seqüências de DNA de indivíduos distintos para verificação de parentesco ou doenças hereditárias.
- Buscas em Banco de Dados: a correspondência é estabelecida entre a informação solicitada pelo usuário e os dados armazenados no banco.
- Análise de Voz: a correspondência é estabelecida entre dois ou mais conjuntos de dados de vozes para identificação.
- Análise de Dados Sísmicos: a correspondência é estabelecida entre dados capturados de sensores distintos para a recuperação da informação estrutural do solo.
- Controle de Qualidade: a correspondência é estabelecida entre um conjunto de medidas oriundas de um produto e o conjunto de medidas padrão.

Em Processamento de Imagens, o Problema de Correspondência é também encontrado em diversas aplicações, tais como:

- Reconhecimento de padrões: a correspondência é estabelecida entre uma imagem de pesquisa e uma imagem de referência, por exemplo, para reconhecimento de face e impressão digital.
- Compressão: a correspondência é estabelecida entre cada parte de uma imagem e um conjunto de imagens padrão codificados.
- Inspeção Industrial: a correspondência é estabelecida entre a imagem de um produto a ser inspecionado e a imagem de um produto padrão.

- Registro: a correspondência é estabelecida entre duas imagens tomadas em tempos distintos do mesmo ponto de vista, como por exemplo, em imagens de satélite para fins meteorológicos e monitoração de crescimento da malha urbana.
- *Shape from Motion*: a correspondência é estabelecida entre partes de duas ou mais imagens tomadas em tempos distintos do mesmo ponto de vista para recuperação da informação tridimensional da cena.
- Visão Estéreo: a correspondência é estabelecida entre partes de duas ou mais imagens tomadas em posições distintas para recuperação da informação tridimensional da cena.

De maneira bastante simples, o Problema de Correspondência para fins de Visão Estéreo consiste em, dado um ponto em uma imagem, determinar o seu homólogo na outra imagem. A Figura 2.5 ilustra o problema.

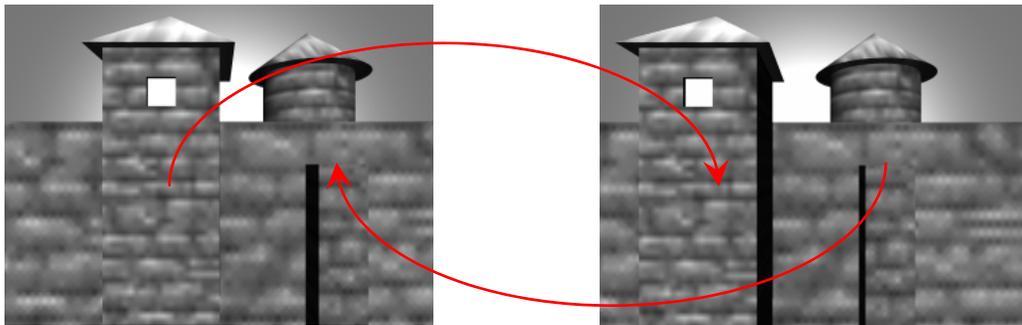


Fig. 2.5 – Problema de Correspondência em Visão Estéreo.

Atualmente é comum encontrar na literatura especializada duas classes em que se pode separar as técnicas para mensuração das correspondências em função da primitiva utilizada: estabelecimento das correspondências baseada em Feições e em Áreas.

O estabelecimento das correspondências **baseado em Feições** é realizado no espaço de atributos dos objetos segmentados nas imagens, como por exemplo, comprimento e orientação para objetos lineares; área e raio para objetos circulares, momentos de várias ordens para objetos regulares ou mesmo irregulares, entre outras. A Figura 2.6 ilustra este tipo de correspondência, onde a parte inferior mostra os objetos segmentados das imagens.

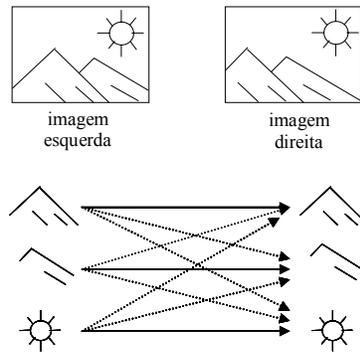


Fig. 2.6 – Estabelecimento da Correspondência baseado em Feições.

As principais vantagens desta técnica são a precisão obtida e, geralmente, o baixo tempo de processamento (devido à dimensão do espaço de atributos ser geralmente pequena).

A principal desvantagem desta técnica está em apenas mensurar a correspondência para os pontos pertencentes aos objetos segmentados nas imagens estereoscópicas. Esta desvantagem fez com que esta técnica tenha recebido pouca atenção ultimamente segundo Brown et. al. (2003).

No estabelecimento das correspondências **baseado em Áreas**, as medidas de correlação são realizadas entre uma janela (área) na imagem esquerda e várias janelas situadas em diferentes posições na imagem direita.

As principais vantagens desta técnica são as possibilidades da determinação de medidas de correlação para todos os *pixels* das imagens estereoscópicas, com exceção das regiões de oclusão, obviamente, e a não necessidade de pré-processamento monocular, tais como, extração de feições e rotulações.

Por outro lado, pode-se considerar como principal desvantagem desta técnica a sua grande exposição aos problemas referentes às distorções geométricas e radiométricas entre as imagens.

Geralmente, associa-se a posição da janela ao *pixel* central. A Figura 2.7 mostra este processo.

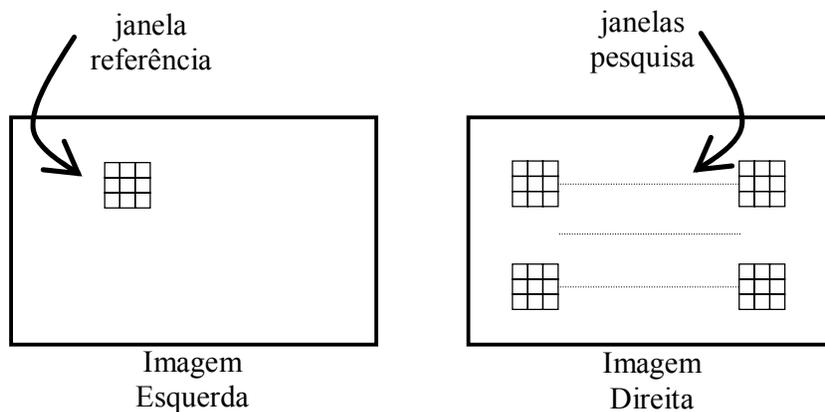


Fig. 2.7 – Estabelecimento da Correspondência baseado em Áreas, com janela de referência na imagem esquerda e janelas de pesquisa na imagem da direita.

Existem várias funções específicas de correlação, entre elas, as mais utilizadas são o Coeficiente de Correlação (CC), o qual é invariante à translação radiométrica, e a Somatória das Diferenças Absolutas (SDA) (Nogueira, 1998).

O Coeficiente de Correlação (CC) é dado por:

$$CC(a,b) = \frac{\sum_{i=0}^{r_y-1} \sum_{j=0}^{r_x-1} [g_r(i,j) - \bar{g}_r] \cdot [g_p(i+a, j+b) - \bar{g}_{p_{a,b}}]}{\sqrt{\sum_{i=0}^{r_y-1} \sum_{j=0}^{r_x-1} [g_r(i,j) - \bar{g}_r]^2 \cdot \sum_{i=0}^{r_y-1} \sum_{j=0}^{r_x-1} [g_p(i+a, j+b) - \bar{g}_{p_{a,b}}]^2}} \quad (2.1)$$

onde:

$CC(a,b)$ é o coeficiente de correlação da função covariância cruzada para o par (a,b) .

$$0 < a < p_x - r_x + 1 \quad (2.2)$$

$$0 < b < p_y - r_y + 1 \quad (2.3)$$

$$\bar{g}_r = \frac{1}{r_x \cdot r_y} \sum_{i=0}^{r_y-1} \sum_{j=0}^{r_x-1} g_r(i,j) \quad (2.4)$$

$$\bar{g}_{p_{a,b}} = \frac{1}{r_x \cdot r_y} \sum_{i=0}^{r_y-1} \sum_{j=0}^{r_x-1} g_p(i+a, j+b) \quad (2.5)$$

p_x é a dimensão da janela de pesquisa na direção x;

r_x é a dimensão da janela de referência na direção x;

p_y é a dimensão da janela de pesquisa na direção y;

r_y é a dimensão da janela de referência na direção y;

$g_r(i,j)$ é o valor numérico da imagem (função de tom de cinza) na posição (i,j) no sistema da janela de referência;

$g_p(i,j)$ é o valor numérico da imagem (função de tom de cinza) na posição (i,j) no sistema da janela de pesquisa;

A Somatória das Diferenças Absoluta (SDA) é dada por:

$$SDA(a,b) = \sum_{i=0}^{r_y-1} \sum_{j=0}^{r_x-1} |g_r(i,j) - g_p(i+a,j+b)| \quad (2.6)$$

Independentemente da maneira como as medidas de correlação entre *pixels* das imagens estereoscópicas foram obtidas, utiliza-se algum critério a fim de definir, dado um *pixel* em uma imagem, qual o seu homólogo na outra imagem. Um simples critério seria escolher como homólogo a um *pixel* em uma imagem, aquele na outra imagem que apresente maior medida de correlação.

Uma vez definido os pares de *pixels* homólogos, calcula-se então a diferença entre suas coordenadas. Esta diferença é denominada **disparidade** (paralaxe).

Seja $P(X, Y, Z)$ um ponto referenciado no sistema de coordenadas do espaço-objeto e $p_E(x_E, y_E)$ e $p_D(x_D, y_D)$ suas respectivas imagens nos sistemas de coordenadas da imagem da esquerda e da direita, respectivamente, tem-se que:

$$d_H = x_E - x_D \quad (2.7)$$

e

$$d_V = y_E - y_D \quad (2.8)$$

onde:

d_H é a disparidade horizontal (na direção x); e

d_V é a disparidade vertical (na direção y).

Desconsiderando erros residuais, em imagens retificadas (seção 2.1.1), a disparidade vertical é nula para todos os pontos das imagens.

A importância em medir disparidades é que esta grandeza é inversamente proporcional à profundidade. Para imagens retificadas, a profundidade de um ponto da cena pode ser calculada por:

$$Z = \frac{b \cdot f}{d_H} \quad (2.9)$$

onde:

Z é a profundidade de um ponto da cena;

b é a base (distância entre os centros perspectivos das câmeras da esquerda e da direita);

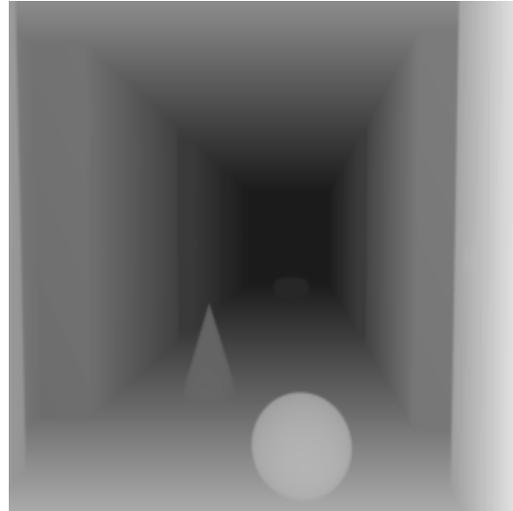
f é a distância focal.

O Mapa de Disparidade é uma imagem (matriz) geralmente de mesmas dimensões das imagens estereoscópicas cujos valores associados a cada *pixel* da imagem representam os valores de disparidades. A Figura 2.8.a mostra uma imagem estereoscópica e a Figura 2.8.b seu respectivo mapa de disparidade.



a

Fig. 2.8.a – Imagem Estéreo Esquerda.



b

Fig. 2.8.b – Mapa de Disparidade Esquerdo.

Na Figura 2.8.b os tons de cinza mais claros representam maiores valores de disparidade (menores valores de profundidade) enquanto os mais escuros representam menores valores de disparidade (maiores valores de profundidade).

2.2 Fatores que Dificultam o Estabelecimento da Correspondência

O objetivo desta seção é discriminar e descrever isoladamente os vários fatores que dificultam o estabelecimento da Correspondência. A análise destes fatores e seus respectivos efeitos sobre as imagens estereoscópicas são importantes para a construção de um critério de correspondência adequado que seja robusto às várias distorções existentes entre as imagens estereoscópicas. Sem perda de generalidade, estas análises consideram imagens estereoscópicas retificadas (disparidade na direção y nula).

Estes fatores podem ser classificados em quatro classes: Fatores Geométricos, Radiométricos, Condições de Textura e Iluminação e Ruído.

2.2.1 Fatores Geométricos

O principal fator que ocasiona diferenças entre as imagens estereoscópicas é a diferença de posição das câmeras. No entanto, para pontos da cena de igual profundidade em relação às câmeras (superfícies paralelas aos planos sensores das câmeras), a diferença que ocorre entre as imagens destes pontos é apenas uma translação (*shift*) de mesmo valor para todos os pontos, sendo esta proporcional ao comprimento da base (distância entre as câmeras) e da profundidade. A Figura 2.9 mostra esta translação para quatro pontos de mesma profundidade, onde o ponto na posição 4 na imagem da esquerda é homólogo ao ponto na posição 1 na imagem direita, resultando em uma disparidade $d = 3$. Todos os demais pontos também possuem homólogos deslocados de mesmo valor de disparidade $d = 3$.

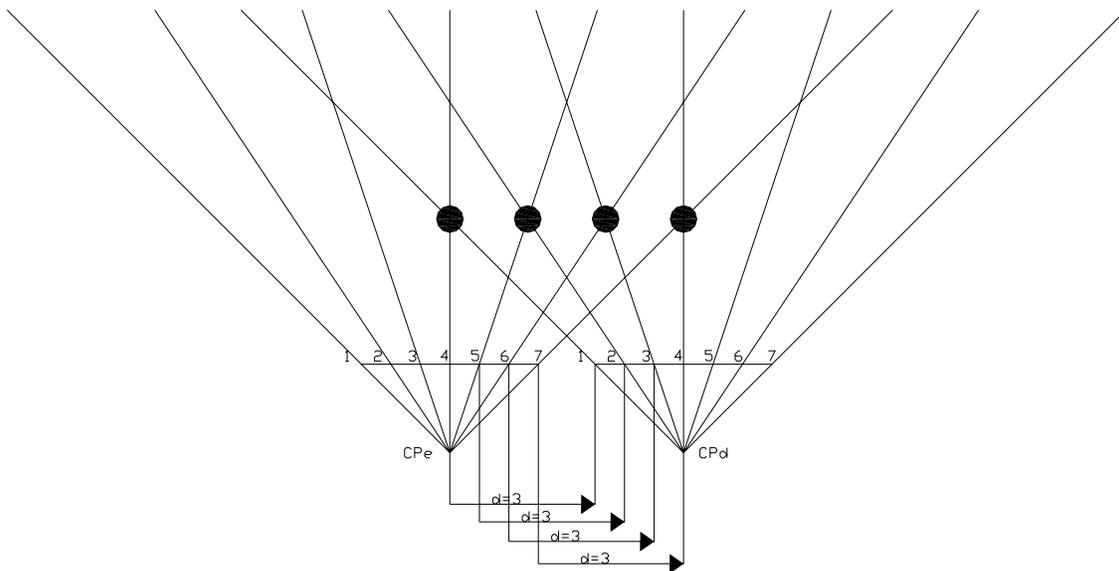


Fig. 2.9 – Superfície Fronto Planar: pontos são deslocados relativamente entre as imagens de mesmo valor de disparidade, no caso, $d = 3$.

Quando as superfícies da cena não são paralelas em relação aos planos sensores das câmeras cada par de pontos homólogos possui um valor de disparidade. Assim, só terão o mesmo valor de disparidade, aqueles pares de pontos homólogos que tiverem à mesma profundidade em relação às câmeras. A Figura 2.10 ilustra estas translações relativas, onde

o ponto na posição 3 na imagem da esquerda é homólogo ao ponto na posição 1 na imagem direita, resultando em uma disparidade $d = 2$. Já o ponto na posição 4 na imagem da esquerda é homólogo também ao ponto na posição 1 na imagem direita, resultando em uma disparidade $d = 3$. Para os demais pontos, o raciocínio é análogo.

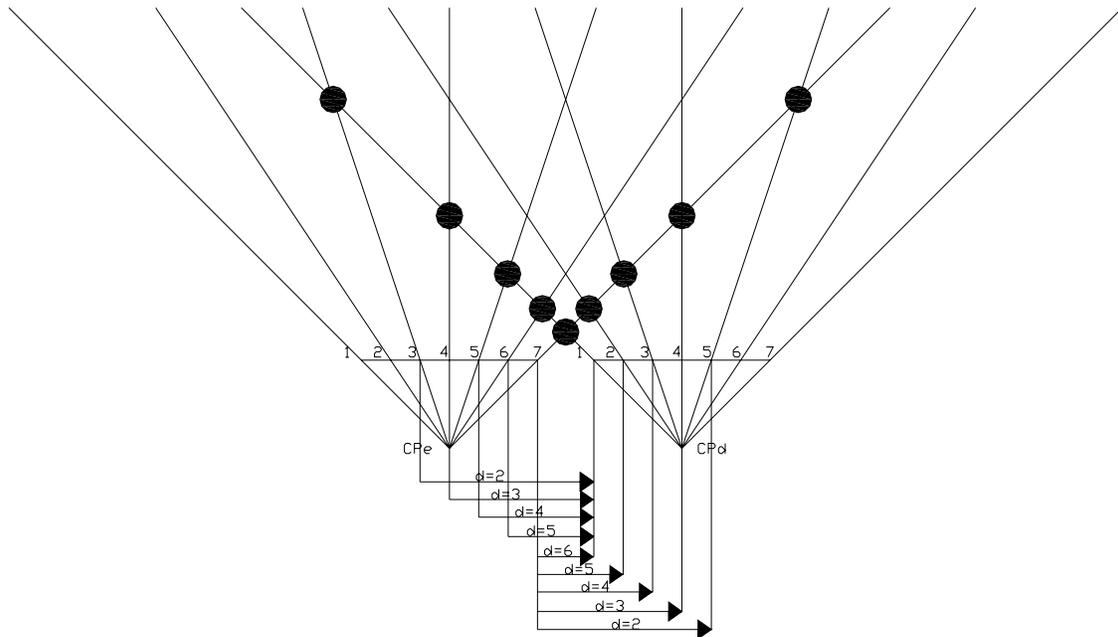


Fig. 2.10 – Superfície Não Fronto Planar: pontos são deslocados relativamente entre as imagens de valores distintos de disparidade.

Estas diferenças entre as imagens estereoscópicas podem dificultar significativamente a imposição de restrições ao Problema de Correspondência. A Figura 2.11 mostra um exemplo em que o limite entre duas texturas distintas (cinza escuro e cinza claro) sobre uma esfera forma uma linha reta na imagem da esquerda, enquanto na imagem da direita está mesma linha homóloga aparece com uma curvatura bastante acentuada. Com isso, impor, por exemplo, uma restrição ao Problema de Correspondência em que feições retas na imagem da esquerda devam aparecer como feições retas na imagem da direita pode falhar.



Fig. 2.11 – Exemplo de feição reta em uma imagem homóloga a uma feição curva na outra imagem.

Um dos problemas críticos para o estabelecimento da correspondência baseado em Áreas é selecionar o **tamanho da janela de correlação**.

A escolha do tamanho ótimo da janela depende da magnitude da variação local da disparidade e da textura. As imagens da Figura 2.12 mostram um par de imagens estereoscópicas com janelas de correlação situadas em posições homólogas. Neste exemplo, todos os *pixels* em uma janela possuem correspondentes na outra janela devido a estas janelas estarem situadas sobre imagens de superfícies plano-frontais. Pode-se reparar cuidadosamente nas características da textura para comprovar esta afirmação.

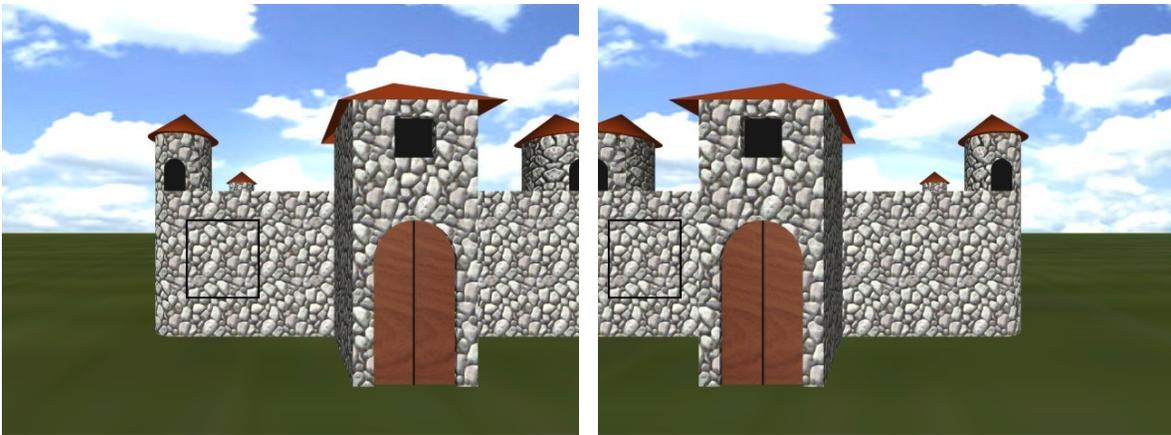


Fig. 2.12 - Janelas de correlação sobre regiões homólogas.

As imagens da Figura 2.13 são análogas às imagens da Figura 2.12, porém a janela de correlação na imagem da esquerda além de conter imagens de superfícies fronto-planares, contém também imagens de uma superfície não fronto-planar em região de oclusão. Neste exemplo é evidente que a medida de correspondência será prejudicada devido a estes efeitos.

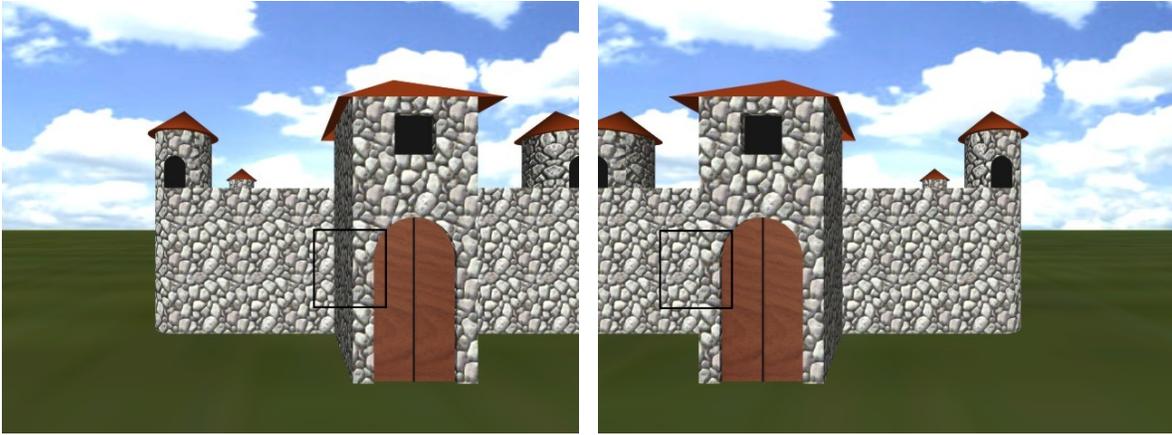


Fig. 2.13 - Janelas de correlação sobre regiões parcialmente homogêneas.

Nas imagens da Figura 2.14, as janelas de correlação não estão sobre nenhuma região de oclusão, mas sim, sobre pontos que correspondem a superfícies não fronto-planares na cena, o que também prejudica a medida de correspondência, uma vez que existem poucos pontos que são homólogos em ambas as janelas.

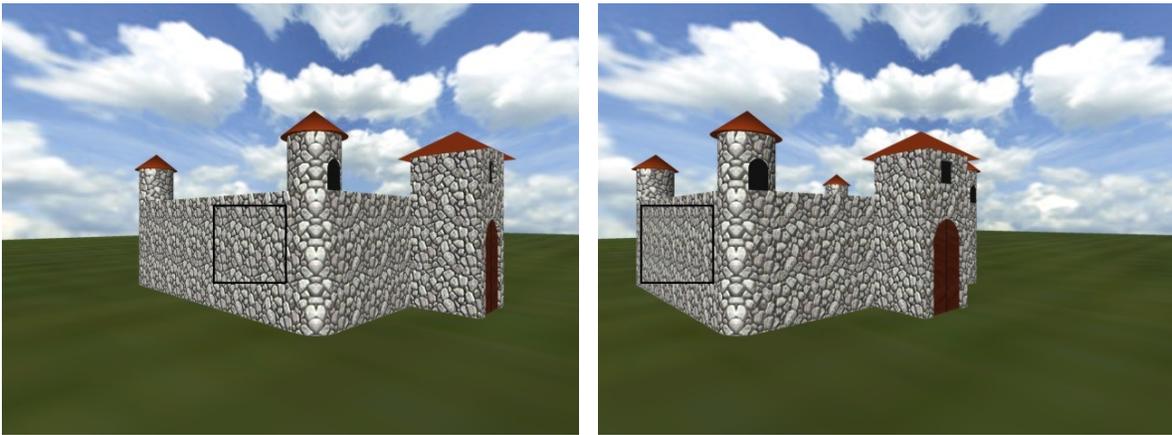


Fig. 2.14 - Imagens esquerda e direita ilustrando efeitos das distorções perspectivas.

Em geral, uma pequena janela é desejável para evitar suavidade em demasia no mapa de disparidade resultante. No entanto, em áreas de textura lisa (variância pequena), uma grande janela é necessária de tal forma que contenha bastante variação de intensidade para minimizar a quantidade de medidas ambíguas. Este fato cria o fenômeno de extensão da região com alta variação de intensidade local sobre regiões com texturas mais homogêneas. Por outro lado, quando a disparidade varia dentro da área delimitada pela

janela, ou seja, a superfície correspondente não é fronto-paralela, os valores de intensidade não se correspondem devido à distorção projetiva (Zitnick & Kanade, 2000).

Nos casos das Figuras 2.13 e 2.14, as janelas de correlação deveriam ser menores a fim de minimizarem a quantidade de *pixels* não homólogos dentro das janelas. Porém, como já citado, janelas muito pequenas tendem a aumentar as ambigüidades devido à baixa quantidade de amostras (pontos) da intensidade.

Alguns trabalhos [Hariyama et. al. (2001); Kanade & Okutomi (1994); Lotti & Giraudon (1994); Woo & Ortega (2000); Veksler (2001); Veksler (2003)] são citados como referência sobre técnicas de janelas adaptativas para o processo de correspondência.

Outro fenômeno que ocorre devido a fatores geométricos e que por sua vez, dificultam o estabelecimento da correspondência é a **Oclusão**.

A Visão Estéreo é geralmente tratada (erroneamente) como uma técnica capaz de recuperar a estrutura tridimensional da cena visível como um todo, porém as posições distintas das câmeras e os diferentes valores de profundidades da cena são os fatores causadores de pontos em uma imagem que não possuem homólogos na outra imagem. Estes pontos são denominados "oclusões", e para estes, a determinação das coordenadas tridimensionais não é possível, uma vez que não estabelecem pares homólogos no par de imagens estereoscópicas.

A Figura 2.15 mostra com maior detalhe as imagens da câmera da esquerda e da direita. Nestas imagens percebe-se facilmente que muitos pontos em uma imagem não possuem homólogos na outra imagem. Estes pontos de oclusão não podem ser reconstruídos através de triangulação, portanto, os algoritmos desenvolvidos para recuperarem a informação tridimensional de maneira autônoma devem estar habilitados em detectá-los a fim de evitar falsas coordenadas tridimensionais da cena.

A Figura 2.16 mostra as mesmas imagens da Figura 2.15, porém as regiões sombreadas indicam todos os *pixels* na imagem da esquerda que não possuem homólogos na imagem da direita e vice-versa.

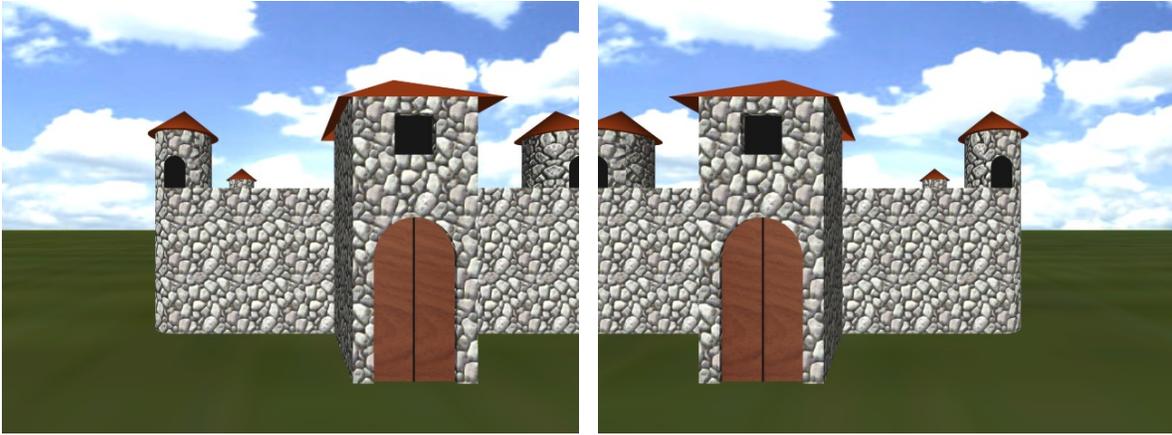


Fig. 2.15 - Imagens esquerda e direita.

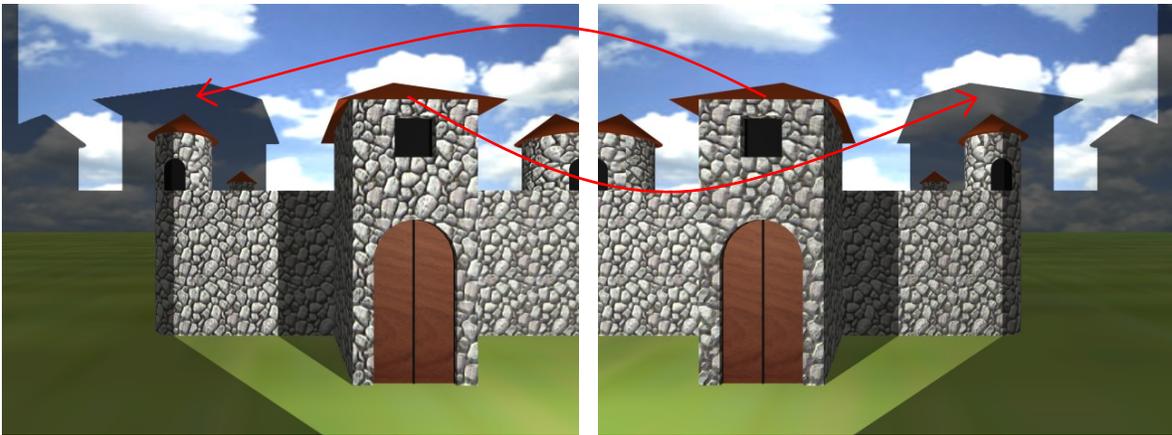


Fig. 2.16 - Imagens esquerda e direita com regiões de oclusão representada pelas sombras.

Estas regiões de oclusão na imagem da esquerda da Figura 2.16 foram criadas adicionando-se uma fonte de luz pontual coincidente com a posição do centro perspectivo da câmera da direita e vice-versa para criar as regiões de oclusão na imagem da direita. Para comprovar esta afirmação experimentalmente, pode-se, por exemplo, comparar as posições das sombras projetadas em função da torre frontal (que possui um portão) sobre o plano que contém a textura representando o céu nas imagens esquerda e direita. Nota-se facilmente, que a sombra desta torre em uma imagem é exatamente a posição desta mesma torre na outra imagem. As setas indicam esta afirmação. Para as demais feições, o raciocínio é análogo.

2.2.2 Fatores Radiométricos

A principal heurística que geralmente é utilizada para a resolução do Problema de Correspondência é a consideração que um par de pontos homólogos possui a mesma intensidade. No entanto, a rigor e teoricamente, esta heurística ocorre apenas em situações/configurações bastante específicas que restringem muito as aplicações práticas.

Existem vários modelos de iluminação que descrevem a formação radiométrica da imagem. Independente do modelo de iluminação utilizado e desconsiderando a ocorrência de ruídos, pode-se deduzir que apenas a componente especular é o fator responsável para um par de *pixels* homólogos não apresentarem a mesma intensidade (ou cor). Por sua vez, a componente especular é função da posição da câmera e uma vez que as imagens estereoscópicas são capturadas em posições distintas, um par de *pixels* homólogos pode não apresentar a mesma intensidade (cor) fazendo que falhe a heurística que *pixels* homólogos possuem a mesma intensidade (cor).

Um simples modelo de iluminação que pode ser utilizado para verificar esta dedução é o modelo de Phong:

$$I = I_a k_a + I_L (K_d \cos(\theta) + K_e \cos^n(\alpha)) \quad (2.10)$$

com:

$$\cos(\theta) = \frac{N \cdot L}{\|N\| \|L\|} \quad (2.11)$$

$$\cos(\alpha) = \frac{R \cdot O}{\|R\| \|O\|} \quad (2.12)$$

onde:

I é a intensidade luminosa em um ponto P de uma superfície S ;

I_a é a intensidade da luz ambiente

K_a é o coeficiente de reflexão ambiente;

I_L é a intensidade da fonte de luz;

K_d é o coeficiente de reflexão difusa;

K_e é o coeficiente de reflexão especular;

$n \in \mathfrak{R}$;

N é o vetor normal de S em P ;

L é o vetor direção da fonte de luz, dado por:

$$L = \overrightarrow{P \text{ Luz}} = \text{Luz} - P; \quad (2.13)$$

Luz é a posição $(X_{\text{Luz}}, Y_{\text{Luz}}, Z_{\text{Luz}})$ da fonte de luz; e

$\|N\|$ e $\|L\|$ é a norma dos vetores normal e direção da fonte de luz, respectivamente.

R é o vetor de reflexão, dado por:

$$R = 2\langle L, N \rangle N - L; \quad (2.14)$$

O é o vetor direção da câmera, dado por:

$$O = \overrightarrow{P \text{ CP}} = \text{CP} - P; \quad (2.15)$$

CP é o centro perspectivo da câmera.

A Figura 2.17 mostra os elementos do modelo de Phong.

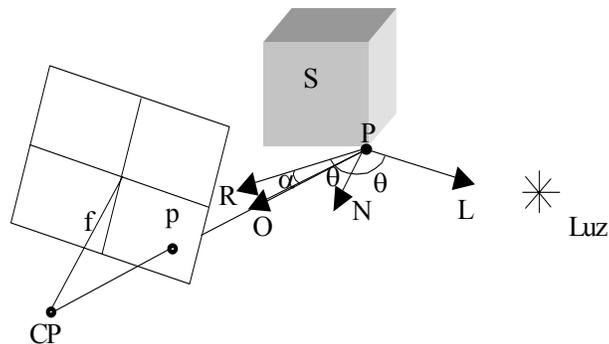


Fig. 2.17 - Modelo de Iluminação de Phong

Analisando o modelo de iluminação de Phong (expressão 2.10), nota-se que apenas o ângulo α , o qual está relacionado com a componente especular, é função da posição da câmera (CP). Assim, pode-se concluir que as posições distintas das câmeras em um sistema estéreo é o fator responsável pela presença de pontos homólogos com intensidades

luminosas distintas. Este fato não se verifica apenas para os pontos homólogos cujo ângulo α associado com o ponto na imagem da esquerda é igual ao ângulo α associado com o ponto na imagem da direita ou para superfícies absolutamente difusas (não especulares, com $K_e = 0$).

Desta forma, qualquer técnica para realizar o estabelecimento da correspondência que utiliza a heurística de que pontos homólogos possuem a mesma intensidade luminosa está considerando que a cena é constituída apenas por superfícies difusas.

Esta discrepância radiométrica entre pontos homólogos poderia ser utilizada como uma informação relevante para a resolução do Problema de Correspondência, no entanto, seria necessário conhecer os demais parâmetros do modelo de iluminação de Phong, tais como, coeficientes de reflexão difusa e especular, posição da(s) fonte(s) de luz, entre outros, o que na prática, raramente se verifica.

2.2.3 Condições de Textura e de Iluminação

A presença de texturas com padrões periódicos ou homogêneos nas superfícies constituintes da cena pode originar padrões ambíguos nas imagens estereoscópicas, dificultando a resolução do Problema de Correspondência. As imagens na Figura 2.18 ilustram a ocorrência de padrões periódicos e a Figura 2.19 mostra a ocorrência de padrões homogêneos. Cabe observar que o caso de textura homogênea também pode ser entendido como um caso de padrão periódico com frequência zero e, portanto, período infinito.

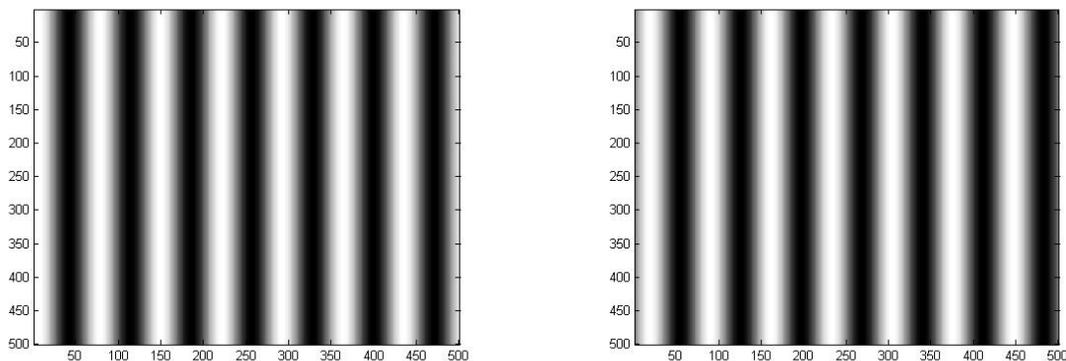


Fig. 2.18 – Imagens estéreo com padrões periódicos.

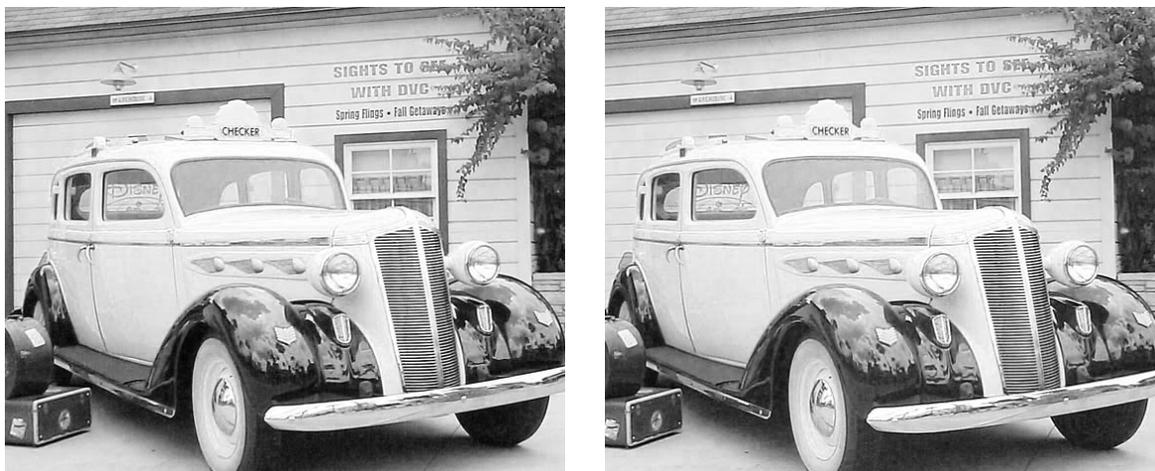


Fig. 2.19 – Imagens estéreo com texturas homogêneas.

As condições de iluminação da cena podem influenciar substancialmente na resolução do Problema de Correspondência, uma vez que cenas pouco ou muito iluminadas podem acarretar imagens com baixo contraste, caracterizando, de certo modo, texturas homogêneas.

Outro problema que pode ocorrer ainda relacionado às condições de iluminação é o caso quando as imagens estereoscópicas são capturadas por uma única câmera que então se desloca para outra posição a fim de capturar duas ou mais imagens da cena a partir de posições distintas. Neste caso, as imagens são tomadas em tempos distintos inevitavelmente (devido, no mínimo, ao tempo gasto durante o deslocamento da câmera para outra posição). Durante o intervalo de tempo entre as tomadas das imagens estereoscópicas as condições de iluminação podem ser alteradas.

2.2.4 Ruído

O processo ruidoso que ocorre em uma imagem é provavelmente independente do processo ruidoso que ocorre na outra imagem se as câmeras em um sistema estéreo são independentes e completamente desacopladas. Neste caso, os processos ruidosos “contaminam” de maneira independente as imagens estereoscópicas, dificultando a resolução do Problema de Correspondência. A Figura 2.20 mostra a imagem estéreo esquerda sem ruído e seu respectivo mapa de disparidade obtido através do estabelecimento

da correspondência baseado em Áreas utilizando uma janela de 7×7 pixels. A Figura 2.21 mostra a imagem estéreo esquerda acrescida de ruído branco com 10 dB e seu respectivo mapa de disparidade obtido pelo mesmo processo anteriormente descrito. A Figura 2.22 difere da Figura 2.21 apenas na potência do processo ruidoso, que no caso, é de 20 dB.

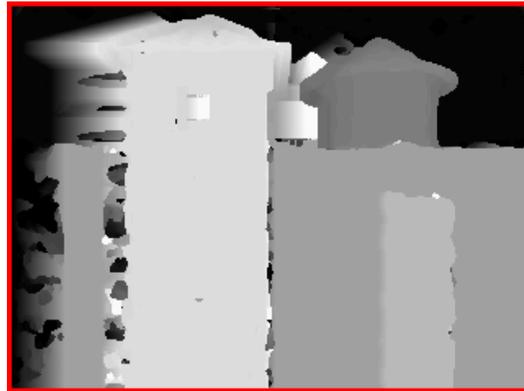
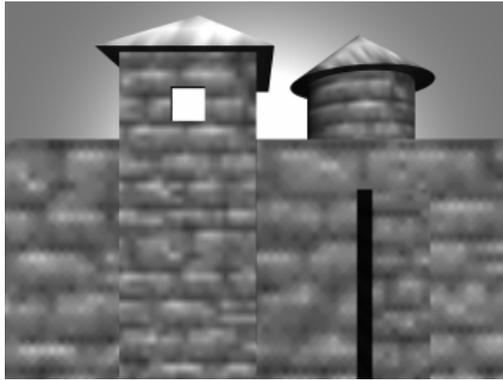


Fig. 2.20 – Imagem estéreo esquerda sem ruído e seu respectivo mapa de disparidade associado.

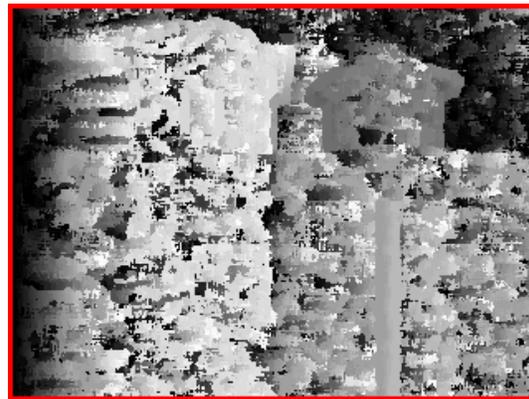
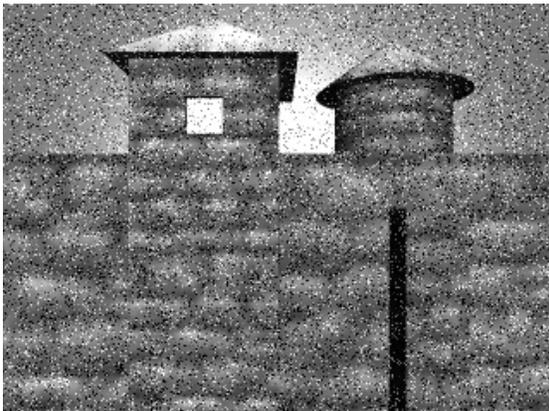


Fig. 2.21 – Imagem estéreo esquerda com 10 dB de ruído e seu respectivo mapa de disparidade associado.

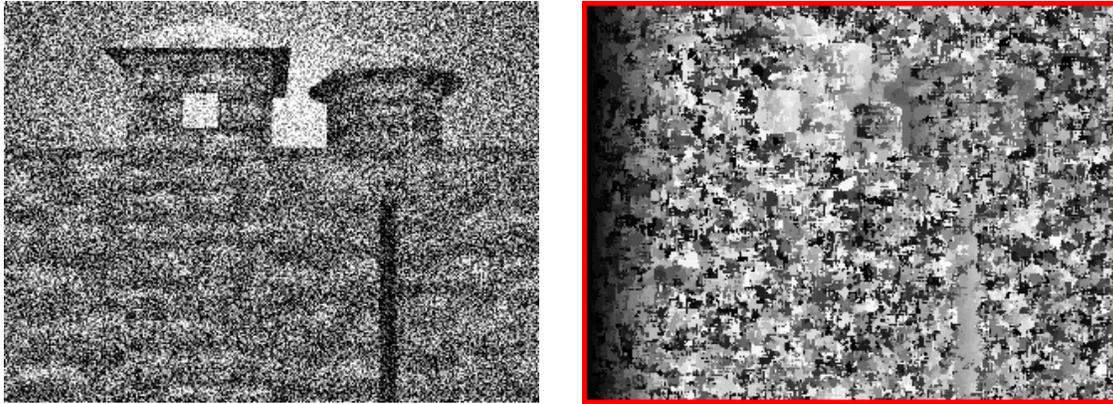


Fig. 2.22 – Imagem estereo esquerda com 20 dB de ruído e seu respectivo mapa de disparidade associado.

Como se pode perceber nas Figuras 2.20, 2.21 e 2.22, a qualidade do mapa de disparidade reconstruído é fortemente influenciada pela potência do processo ruidoso sobre as imagens estereoscópicas.

2.3 Matriz de Correspondência e Injunções

Esta seção apresenta as principais injunções utilizadas no estabelecimento da correspondência. Os efeitos dessas injunções são melhores compreendidos através da Matriz de Correspondência.

2.3.1 Matriz de Correspondência

Seja um par de imagens estereoscópicas de dimensões $H \times W$, onde H é o número de linhas e W é o número de colunas, uma matriz M com dimensões $H \times W \times H \times W$ pode ser montada, na qual o valor do elemento $M(x_E, y_E, x_D, y_D)$ representa uma medida de correspondência entre um ponto $p_E(x_E, y_E)$ na imagem esquerda e um ponto $p_D(x_D, y_D)$ na imagem direita.

Sem perda de generalidade, o par de imagens estereoscópicas pode ser retificado através da reamostragem das linhas das imagens ao longo das linhas epipolares. Este processo é denominado Retificação Estéreo, como já citado na seção 2.1.1, o qual permite reduzir uma dimensão da Matriz de Correspondência.

Desta forma para cada linha epipolar i tem-se uma matriz $W \times W$ e, portanto, para todas as linhas epipolares tem-se H matrizes $W \times W$. O valor associado a cada posição da Matriz de Correspondência $M(y, x_E, x_D)$ com dimensão $H \times W \times W$, representa uma medida de correspondência entre um ponto $p_E(x_E, y)$ na imagem esquerda retificada e um ponto $p_D(x_D, y)$ na imagem direita retificada.

A Figura 2.23 mostra duas câmeras com $W = 7$ para uma linha epipolar i , a Figura 2.24 mostra sua respectiva Matriz de Correspondência e a Figura 2.25 mostra a Matriz de Correspondência para $H = 3$.

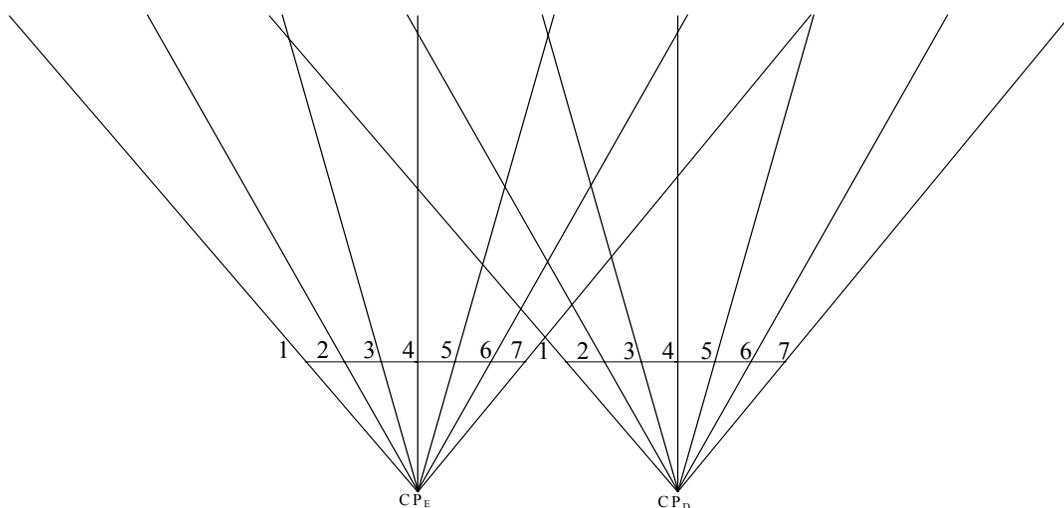


Fig. 2.23 – Câmeras com $W = 7$.

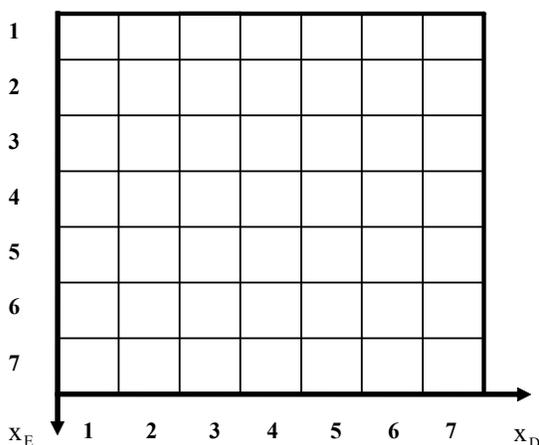


Fig. 2.24 – Matriz de Correspondência para um par de linhas epipolares homólogas.

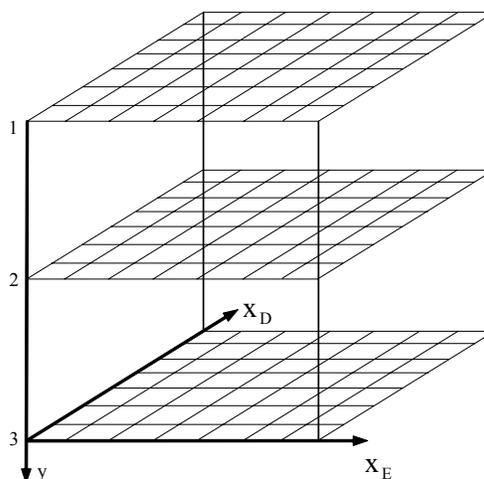


Fig. 2.25 – Matriz de Correspondência para $H = 3$.

2.3.2 Injunções e Seus Efeitos na Matriz de Correspondência

Todas as injunções apresentadas nesta seção são impostas independentemente para cada par de linhas epipolares homólogas i . Assim, o efeito de cada injunção sobre a Matriz de Correspondência pode ser visualizado em uma matriz com dimensões $W \times W$ (semelhantes à matriz da Figura 2.24). O trabalho de Dhond & Aggarwal, 1989 é sugerido como referência sobre injunções em Visão Estéreo.

Disparidade Não Negativa

Considerando que as imagens estereoscópicas estão retificadas, a profundidade Z de um ponto da cena se relaciona de maneira inversamente proporcional a disparidade associada aos *pixels* homólogos nas imagens estereoscópicas, de acordo com a expressão 2.9.

Para profundidades no intervalo $[0, +\infty]$ e considerando, sem perda de generalidade, que os parâmetros b e f na expressão 2.9 são sempre positivos, os valores de disparidade d_H também estão no intervalo $[0, +\infty]$ (não negativo). Valores de disparidades negativas resultam em profundidades negativas, portanto atrás das câmeras, o que é impossível. Assim, pode-se concluir, de acordo com a expressão 2.7, que:

$$x_E \geq x_D \tag{2.16}$$

A Figura 2.26 mostra o efeito dessa injunção na Matriz de Correspondência M .

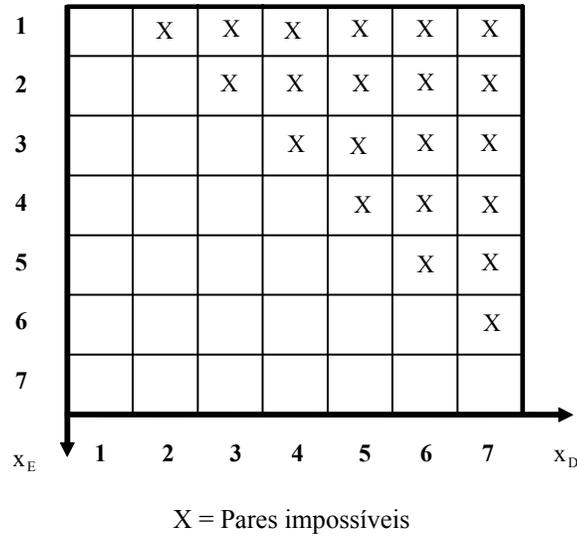


Fig. 2.26 – Efeito da Injunção de Disparidade Não Negativa.

A utilização desta injunção reduz significativamente o custo computacional, além de evitar a ocorrência de pontos homólogos impossíveis. Devido à inexistência de restrições em utilizar esta injunção, as próximas Figuras que demonstram os efeitos das demais injunções estarão acrescidas desta.

Unicidade

A injunção de unicidade estabelece que um ponto $p_E(x_E, y)$ na imagem da esquerda é homólogo a um único ponto $p_D(x_D, y)$ na imagem da direita e vice-versa. Considerando $x_E = k_E$ e $x_D = k_D$, coordenadas de um par de pontos homólogos, então:

$$p_E(k_E, y) \text{ não pode ser homólogo a qualquer outro } p_D(q, y), \text{ com } q \neq k_D \quad (2.17)$$

e

$$p_D(k_D, y) \text{ não pode ser homólogo a qualquer outro } p_E(q, y), \text{ com } q \neq k_E \quad (2.18)$$

A Figura 2.27 mostra o efeito dessa injunção na representação esquemática das câmeras (como exemplo, $(k_E, k_D) = (6, 2)$). O ponto preto representa o par de pontos

homólogos escolhido, e os pontos cinza representam os pares de pontos impossíveis decorrente desta escolha e da injeção de Unicidade.

Esta injeção não é válida para regiões da cena onde exista superfícies transparentes.

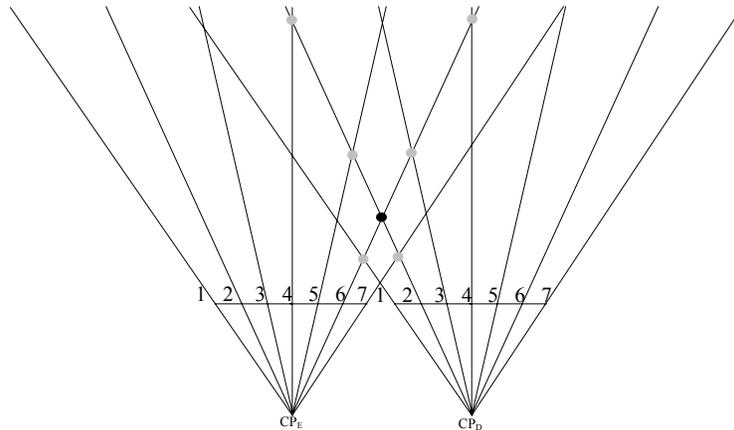


Fig. 2.27 – Injeção de Unicidade.

A Figura 2.28 mostra o efeito conjunto dessa injeção na Matriz de Correspondência M e a da injeção de disparidade não negativa.

1		X	X	X	X	X	X		
2		U	X	X	X	X	X		
3		U		X	X	X	X		
4		U			X	X	X		
5		U				X	X		
6	U	O	U	U	U	U	X		
7		U							
j_E	\downarrow	1	2	3	4	5	6	7	j_D

X = Pares impossíveis pela Injeção de Disparidade Não-Negativa

O = Par Escolhido

U = Pares impossíveis pela Injeção de Unicidade

Fig. 2.28 – Efeito conjunto da Injeção de Unicidade e de Disparidade Não Negativa na Matriz de Correspondência.

Ordem

A Injunção de Ordem impõe que a ordem dos pontos homólogos não seja alterada. Seja $p_E(k_E, y)$ e $p_D(k_D, y)$ um par de pontos homólogos então:

$$p_E(k_E + L, y) \text{ não pode ser homólogo a qualquer outro } p_D(q, y), \text{ com } q \leq k_D \quad (2.19)$$

$$p_E(k_E - L, y) \text{ não pode ser homólogo a qualquer outro } p_D(q, y), \text{ com } q \geq k_D \quad (2.20)$$

$$p_D(k_D + L, y) \text{ não pode ser homólogo a qualquer outro } p_E(q, y), \text{ com } q \leq k_E \quad (2.21)$$

$$p_D(k_D - L, y) \text{ não pode ser homólogo a qualquer outro } p_E(q, y), \text{ com } q \geq k_E \quad (2.22)$$

onde:

$$L \in \mathfrak{R}^+ \quad (2.23)$$

A Figura 2.29 mostra o efeito dessa injunção na representação esquemática das câmeras (como exemplo, $(k_E, k_D) = (6, 2)$). O ponto preto representa o par de pontos homólogos escolhido, e os pontos cinzas representam os pares impossíveis decorrentes desta escolha e da Injunção de Ordem.

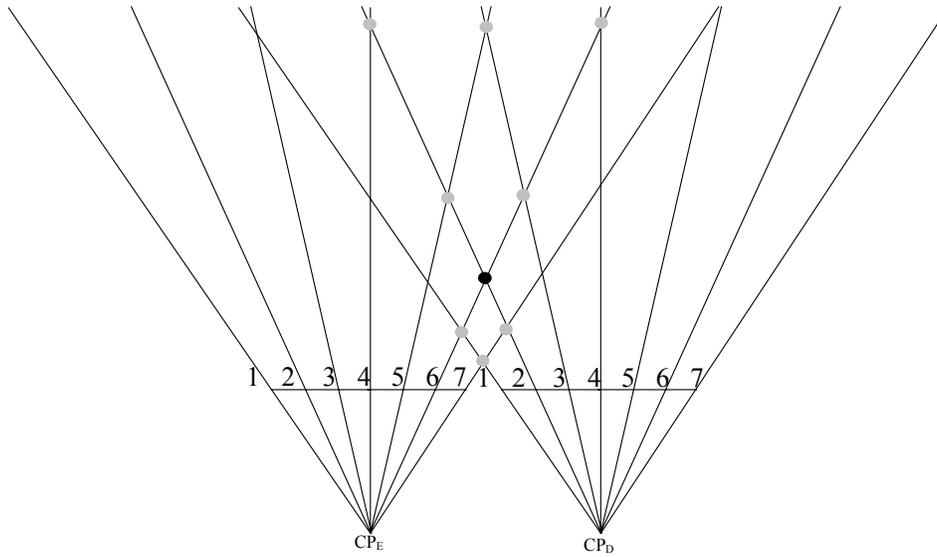


Fig. 2.29 – Injunção de Ordem.

A Figura 2.30 mostra o efeito conjunto dessa injunção na Matriz de Correspondência M e da injunção de Unicidade e de Disparidade Não Negativa.

1		X	X	X	X	X	X
2		W/U	X	X	X	X	X
3		W/U	W	X	X	X	X
4		W/U	W	W	X	X	X
5		W/U	W	W	W	X	X
6	W/U	O	W/U	W/U	W/U	W/U	X
7	W	W/U					

X = Pares impossíveis pela Injunção de Disparidade Não-Negativa

O = Par Escolhido

W/U = Pares impossíveis pela Injunção de Unicidade e de Ordem

W = Pares impossíveis pela Injunção de Ordem

Fig. 2.30 – Efeito conjunto da Injunção de Ordem e de Disparidade Não Negativa na Matriz de Correspondência.

Esta injunção não é válida para regiões da cena com superfícies transparentes e/ou regiões onde o módulo do gradiente de disparidade é maior ou igual a 2 (ver seção seguinte).

Limite do Módulo do Gradiente de Disparidade

O Módulo do Gradiente de Disparidade $|\gamma|$ pode ser definido como a diferença entre as disparidades associadas a dois pares de pontos homólogos próximos (1 e 2), divididos por sua separação (distância) no sistema de coordenadas do espaço ciclopeano*. Analiticamente:

$$|\gamma| = \frac{|d_2 - d_1|}{|D_{12}|} \quad (2.24)$$

onde:

d_1 e d_2 são valores de disparidade associados aos pontos 1 e 2;

D_{12} é a distância entre 1 e 2 no espaço ciclopeano.

Considerando imagens retificadas, a transformação do sistema de coordenadas do espaço das imagens estereoscópicas para o sistema de coordenadas do espaço ciclopeano é dada por:

$$x_C = \frac{x_E + x_D}{2} \quad (2.25)$$

$$y_C = y_E = y_D \quad (2.26)$$

$$d(x_C, y_C) = x_E - x_D \quad (2.27)$$

* ver a Figura 2.31 e expressões 2.25 a 2.30 para definição do espaço ciclopeano

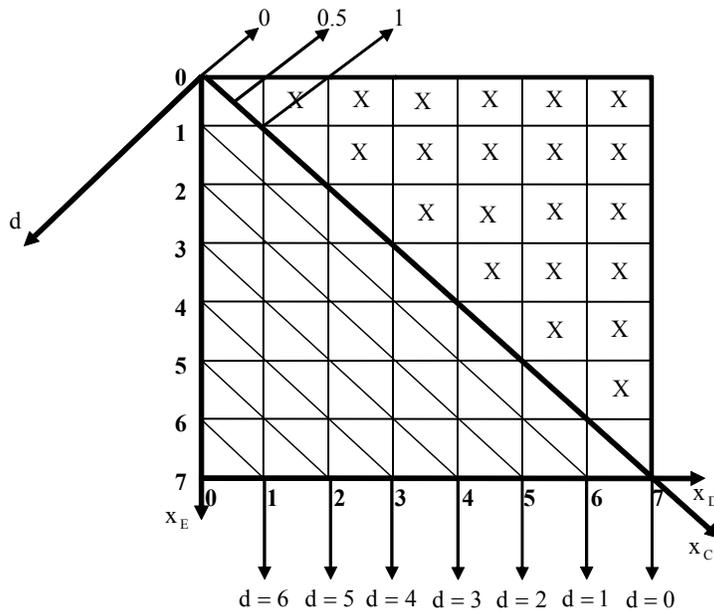
A transformação do sistema de coordenadas do espaço ciclopeano para o sistema de coordenadas do espaço das imagens estereoscópicas é dada por:

$$x_E = x_C + \frac{d(x_C, y_C)}{2} \quad (2.28)$$

$$x_D = x_C - \frac{d(x_C, y_C)}{2} \quad (2.29)$$

$$y_E = y_D = y_C \quad (2.30)$$

A Figura 2.31, mostra os eixos x_C e d representados na Matriz de Correspondência.



X = Pares impossíveis pela Injunção de Disparidade Não-Negativa

Fig. 2.31 – Sistema de coordenadas do espaço ciclopeano na Matriz de Correspondência.

Pode-se notar na Figura 2.31 que a unidade no sistema de coordenadas (x_C, d) equivale a $\sqrt{2}$ unidades no sistema de coordenadas (x_E, x_D) .

A Injunção do Limite do Módulo do Gradiente de Disparidade implica, obviamente, na imposição de um limite para o módulo deste gradiente.

De maneira analítica, pode-se verificar a equivalência entre a Injunção de Ordem e o Limite do Módulo do Gradiente de Disparidade igual a 2: seja $x_E = x_{E1}$ a coordenada na direção x de um ponto 1 na imagem esquerda e $x_D = x_{D1}$ a coordenada na direção x deste mesmo ponto 1, porém na imagem direita e $x_E = x_{E2}$ e $x_D = x_{D2}$ de maneira análoga ao ponto 1, porém para o ponto 2. Substituindo estas coordenadas na expressão 2.24, fica:

$$|\gamma| = \frac{|d_2 - d_1|}{|D_{12}|} = \frac{|(x_{E2} - x_{D2}) - (x_{E1} - x_{D1})|}{\left| \frac{(x_{E1} + x_{D1})}{2} - \frac{(x_{E2} + x_{D2})}{2} \right|} \quad (2.31)$$

fazendo simples manipulações algébricas tem-se que:

$$|\gamma| = \frac{|x_{E2} - x_{D2} - x_{E1} + x_{D1}|}{\left| \frac{(x_{E1} + x_{D1}) - (x_{E2} + x_{D2})}{2} \right|} = 2 \frac{|x_{E2} - x_{D2} - x_{E1} + x_{D1}|}{|x_{E1} + x_{D1} - x_{E2} - x_{D2}|} \quad (2.32)$$

De acordo com Pollard et. al., 1986, a expressão 2.32 pode ser rearranjada para melhor compreensão como:

$$|\gamma| = 2 \frac{|(x_{D1} - x_{D2}) - (x_{E1} - x_{E2})|}{|(x_{D1} - x_{D2}) + (x_{E1} - x_{E2})|} \quad (2.33)$$

As coordenadas do ponto 2 podem ser definidas em função das coordenadas do ponto 1 por:

$$\begin{aligned} x_{E2} &= x_{E1} + M \\ x_{D2} &= x_{D1} + N \end{aligned} \quad (2.34)$$

onde:

$$M \in \mathfrak{R}^+ \quad (2.35)$$

$$N \in \mathfrak{R}^+$$

Substituindo 2.33 em 2.32, fica:

$$|\gamma| = 2 \frac{|(x_{Dl} - (x_{Dl} + N)) - (x_{El} - (x_{El} + M))|}{|(x_{Dl} - (x_{Dl} + N)) + (x_{El} - (x_{El} + M))|} = 2 \frac{|M - N|}{|-M - N|} \quad (2.36)$$

A inversão de ordem ocorre quando:

$$M > 0 \quad \text{e} \quad N < 0 \quad (2.37)$$

ou

$$M < 0 \quad \text{e} \quad N > 0 \quad (2.38)$$

Analisando os intervalos dados em 2.35 e 2.36 em 2.33, verifica-se que para estes o quociente $\frac{|M - N|}{|-M - N|} > 1$, o que implica que quando $|\gamma| > 2$ ocorre a inversão de ordem.

A Figura 2.32 mostra o efeito dessa injunção na representação esquemática das câmeras (como exemplo, $(x_E, x_D) = (6, 2)$). O ponto preto representa o par de pontos escolhido como homólogos e as regiões hachuradas representam os pares impossíveis para esta injunção (região onde o módulo do gradiente de disparidade é maior ou igual a 2).

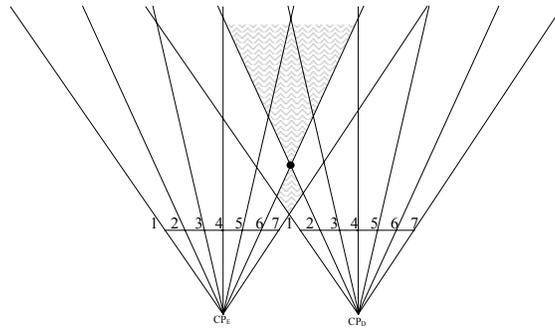


Fig. 2.32 – Injunção do Limite de Gradiente de Disparidade com módulo = 2.

A Figura 2.33 mostra o efeito conjunto dessa injunção na Matriz de Correspondência M e a injunção de disparidade não negativa.

1		X	X	X	X	X	X		
2		G	X	X	X	X	X		
3		G	G	X	X	X	X		
4		G	G	G	X	X	X		
5		G	G	G	G	X	X		
6	G	O	G	G	G	G	X		
7	G	G							
	x_E	1	2	3	4	5	6	7	x_D

X = Pares impossíveis pela Injunção de Disparidade Não-Negativa

O = Par Correto

G = Pares impossíveis pela Injunção do Limite do Gradiente de Disparidade com módulo = 2.

Fig. 2.33 – Efeito conjunto da Injunção de Limite de Gradiente de Disparidade com módulo = 2 e de Disparidade Não Negativa.

A imposição de um limite para o Módulo do Gradiente de Disparidade igual a 2 é equivalente à imposição da Injunção de Ordem. Analisando as Figuras 2.30 e 2.33, pode-se observar que as regiões de pares impossíveis para ambas são idênticas.

Embora este trabalho não seja voltado para a Visão Humana, é interessante relatar que através de testes psicofísicos, Burt & Julesz (1980), determinaram que a fusão das imagens estereoscópicas em seres humanos ocorre para regiões onde o Módulo do Gradiente de Disparidade é menor que 1. Para as regiões onde este limite é ultrapassado ocorre o fenômeno de diplopia (enxergar duplo um mesmo objeto).

A Figura 2.34 mostra o efeito dessa injunção, para módulo = 1, na Matriz de Correspondência M e da injunção de Disparidade Não Negativa, considerando, como exemplo, $(x_{E1}, x_{D1}) = (6, 2)$.

1	G	X	X	X	X	X	X
2	G	G	X	X	X	X	X
3	G	G	G	X	X	X	X
4		G	G	G	X	X	X
5		G	G	G	G	X	X
6	G	O	G	G	G	G	X
7	G	G			G	G	G

x_E ↓ 1 2 3 4 5 6 7 x_D →

X = Pares impossíveis pela Injunção de Disparidade Não-Negativa

O = Par Correto

G = Pares impossíveis pela Injunção do Limite do Gradiente de Disparidade com módulo = 1.

Fig. 2.34 – Efeito conjunto da Injunção de Limite de Gradiente de Disparidade com módulo = 1 e de Disparidade Não Negativa.

A título de melhor visualização, a Figura 2.35 mostra em branco a região de pares possíveis e em preto a região de pares impossíveis em uma Matriz de Correspondência de 100 x 100 elementos, considerando, como exemplo, $(x_{E1}, x_{D1}) = (50, 20)$ para a Injunção do Limite do Gradiente de Disparidade com módulo = 1 acrescida da Injunção de Disparidade Não Negativa.

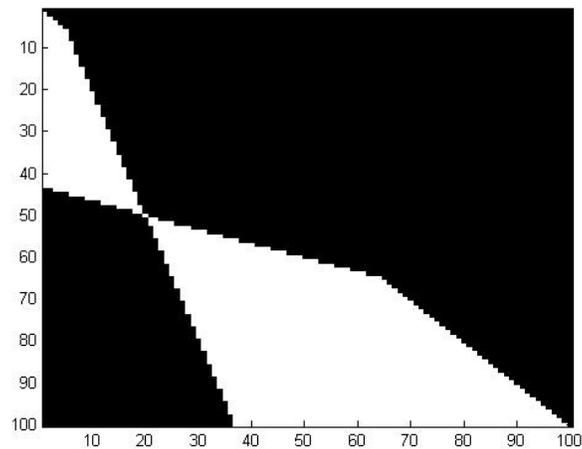


Fig. 2.35 – Efeito conjunto da Injunção de Limite de Gradiente de Disparidade com módulo = 1 e de Disparidade Não Negativa, para uma Matriz de Correspondência 100 x 100.

A Figura 2.36 é análoga à Figura 2.35, exceto que o módulo do Gradiente de Disparidade é igual a 2.

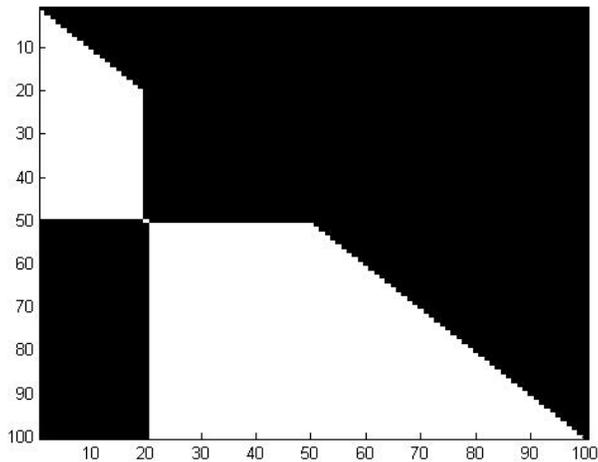


Fig.2.36 – Efeito conjunto da Injunção de Limite de Gradiente de Disparidade com módulo = 2 e de Disparidade Não Negativa, para uma Matriz de Correspondência 100 x 100.

O gradiente de disparidade pode ser também determinado em função das coordenadas no espaço-objeto de dois pontos. Por simplificação, considera-se que o sistema de coordenadas do espaço-imagem da câmera da esquerda é paralelo é coincidente com o sistema de coordenadas do espaço-objeto e o sistema de coordenadas do espaço-imagem da

câmera da direita difere do sistema de coordenadas do espaço-imagem da câmera da esquerda apenas por uma distância b na direção X das coordenadas do centro perspectivo da câmera da direita e que ainda ambas as câmeras possuam a mesma distância focal.

Seja (X_1, Z_1) as coordenadas na direção X e Z , respectivamente, do ponto 1 no espaço-objeto e (X_2, Z_2) o mesmo que (X_1, Z_1) , porém para o ponto 2 e f a distância focal; as coordenadas na direção x_E no sistema de coordenadas do espaço-imagem da câmera da esquerda e da câmera da direita x_D para os pontos 1 e 2 ficam:

$$x_{E1} = f \frac{X_1}{Z_1} \quad (2.39)$$

$$x_{E2} = f \frac{X_2}{Z_2} \quad (2.40)$$

$$x_{D1} = f \frac{(X_1 - b)}{Z_1} \quad (2.41)$$

$$x_{D2} = f \frac{(X_2 - b)}{Z_2} \quad (2.42)$$

Substituindo 2.39, 2.40, 2.41 e 2.42 em 2.33, fica:

$$|\gamma| = 2 \frac{|b(Z_1 - Z_2)|}{|2(X_1 Z_2 - X_2 Z_1) + b(Z_1 - Z_2)|} \quad (2.43)$$

A expressão 2.43 pode ser determinada em função de demais parâmetros de calibração. Para isto as expressões 2.39, 2.40, 2.41 e 2.42 podem ser generalizadas, tais como as expressões A.2, apresentadas no capítulo de Anexos.

Capítulo 3

3. Trabalhos Relacionados

O objetivo deste capítulo é apresentar, de maneira resumida uma revisão de trabalhos apresentados na literatura especializada, os quais, de alguma forma, estão relacionadas com a abordagem a ser proposta. Estas revisões são apresentadas em “caixas de texto (molduras)” sendo, logo abaixo a cada uma dessas caixas, apresentadas as contribuições deste trabalho em relação às supra citadas.

3.1 Trabalhos Disponíveis na Literatura

Uma quantidade significativamente grande de artigos sobre o Problema de Correspondência Estéreo pode ser encontrada na literatura especializada. No entanto, a maioria desses artigos não trata o problema de oclusão como uma característica inerente ao processo estéreo. Muitas vezes, as regiões de oclusão são negligenciadas, assumindo-se que estas são desprezíveis ou que simplesmente não existem. Na seqüência são apresentados alguns trabalhos que consideram as regiões de oclusão.

Chung & Nevatia (1991) denominaram as regiões de oclusão como regiões NVFOV (*Not Visible From Other View*). Através de simples análise, formularam a seguinte injunção de ocorrência de regiões NVFOV: "Em qualquer limite de oclusão entre duas superfícies, a região NVFOV tem que aparecer na imagem esquerda se a superfície oclusa está à esquerda da superfície oclusora, ou à direita, caso contrário". As regiões de oclusão são caracterizadas através da análise de uma descrição estrutural de mais alto nível da cena obtida por técnicas de agrupamento em um processo hierárquico.

A abordagem proposta não utiliza uma descrição estrutural de mais alto nível da cena para identificar as regiões de oclusão, mas sim, características geométricas de formação das imagens estereoscópicas em função da cena. A vantagem em utilizar apenas características geométricas é que estas são independentes da complexidade da cena, enquanto descrições estruturais de mais alto nível baseadas em técnicas de agrupamento não obtém bons resultados em cenas complexas.

Nishikawa et. al. (1995), descreveram um algoritmo para a detecção das regiões de oclusão utilizando um sistema de visão estéreo dinâmico. A técnica consiste basicamente em estimar disparidades através do fluxo óptico formado quando o par de câmeras desloca-se conjuntamente na direção paralela à base. Uma vez que as imagens dos pontos mais profundos da cena possuem menores magnitudes de suas velocidades instantâneas em relação às imagens dos pontos menos profundos, as disparidades (deslocamento) são obtidas pela somatória (integração) dos vários valores das velocidades instantâneas medidas em um determinado intervalo de tempo. Os pontos cujas velocidades não podem ser estimadas são classificados como oclusões.

A abordagem proposta não utiliza nenhuma informação obtida através do deslocamento, quer seja do par de câmeras, quer seja de objetos constituintes da cena, para estimar disparidades e/ou oclusões. Caso haja interesse ou necessidade de reconstruir a estrutura tridimensional da cena considerando deslocamentos, a abordagem proposta pode ser utilizada sem qualquer alteração para cada par de imagens (*frames*) da seqüência de imagens (vídeo) tomadas.

Zitnick & Kanade (2000) descreveram um algoritmo cooperativo, inspirado no clássico algoritmo de Marr & Poggio (1976), no qual um processo de relaxação aumenta uma medida de confiabilidade para as disparidades que obedecem as injunções de continuidade e unicidade dentro de uma vizinhança excitatória e diminui esta medida de confiabilidade para as demais disparidades dentro de uma vizinhança inibitória. As oclusões são explicitamente detectadas selecionando-se um limiar (*threshold*) para as medidas de confiabilidade, uma vez que as disparidades corretas devem possuir altas medidas de confiabilidade e as oclusões o oposto. Neste trabalho não foi relatado nenhuma técnica para a seleção adequada do limiar citado, o qual nos testes realizados foi determinado de maneira empírica.

A abordagem proposta também utiliza um processo cooperativo para designar medidas de confiabilidade (denominadas neste trabalho por “prioridades”) para cada par de pontos candidatos a pontos homólogos, no entanto, não se utiliza de um limiar (designado geralmente de maneira empírica) para determinar as regiões de oclusão, o que torna a abordagem proposta praticamente independente da designação de parâmetros.

Vários trabalhos (Sato & Ohta, 1996; Nakamura et. al., 1996; Ohta et. al. 1998; Sugaya & Ohta, 2000 e Gong & Yang, 2002) utilizaram uma configuração de câmeras denominada SEA (*Stereo by Eye Array*) composta por nove (ou quinze) câmeras dispostas de maneira matricial com três linhas e três colunas (ou 5 linhas e 5 colunas). Apesar dos algoritmos desenvolvidos nestes artigos apresentarem algumas diferenças entre si, todos foram baseados em um mesmo princípio: “tirar proveito da redundância de informação proporcionada pela abundância de câmeras com o intuito de verificar se uma medida de similaridade ruim foi ocasionada por ruído ou oclusão, uma vez que oclusões são explicitamente dependentes das posições das câmeras, enquanto ruídos não”.

A abordagem proposta utiliza apenas um par de câmeras, que é a configuração mínima para um sistema de visão estéreo. A principal desvantagem em utilizar mais de duas câmeras está no aumento dos dados para processamento, o que implica diretamente no aumento do tempo de processamento e na necessidade de maior capacidade de

armazenamento. Outra desvantagem é obviamente no custo de aquisição e utilização de um sistema com mais de duas câmeras em relação ao sistema convencional com apenas duas câmeras utilizado neste trabalho.

Triantafyllidis et. al., 2000 e Triantafyllidis et. al., 2001, formularam o problema de detecção de oclusões através de um método de decisão Bayesiana. No entanto, os valores de disparidades devem ser pequenos e possuem distribuição gaussiana, além do modelo do ruído utilizado ser branco gaussiano aditivo de média zero. Tais considerações não condizem necessariamente com o observado em situações reais, limitando assim, a aplicação desta metodologia apenas para casos específicos.

A abordagem proposta não faz qualquer consideração a respeito da distribuição de probabilidade dos valores de disparidade, nem tampouco quanto a sua magnitude (pequenos ou grandes valores).

Uma outra abordagem utilizada para a detecção das regiões de oclusão consiste em tratar o Problema de Correspondência como um problema de otimização (Shah, 1993; Belhumeur, 1993; Kang et. al., 2001) no qual a solução ótima é a de menor energia de um funcional. Nesta abordagem, uma das variáveis deste funcional é uma função binária que descreve as regiões de oclusão. As demais variáveis geralmente envolvidas são uma função de disparidade e uma função de descontinuidade também binária, que estabelece as regiões de descontinuidade da cena. O artigo de Redert et al. (1999) traz uma ótima revisão sobre o problema de correspondência, inclusive sobre a modelagem de funcionais de energia.

A abordagem proposta não utiliza nenhum funcional de energia, devido entre outras, a própria dificuldade em descrever um modelo matemático adequado para o problema, bem como, a definição de seus parâmetros.

Ainda sobre minimização de energia, por exemplo, Chang et. al. (1991) utilizaram uma técnica denominada *Matching Dual Paralelo* para descrever as regiões de oclusão. Nesta técnica, um processo estabelece a correspondência da imagem da esquerda para a direita, enquanto outro estabelece a correspondência da direita para a esquerda em paralelo. Os mapas de disparidades com menores diferenças na vizinhança dos limites das regiões de oclusão são utilizados como entrada no processo de minimização de energia de um funcional que estabelece uma relação entre a suavidade do mapa de disparidade e as discontinuidades modeladas por um Processo Binário. O processo de minimização é realizado por *Simulated Annealing* com amostrador de Gibbs.

A abordagem proposta também considera as medidas de correspondência da imagem da esquerda em relação a imagem da direita e vice-versa (correspondência bidirecional), porém tais medidas não são empregadas em um funcional de energia para posterior minimização através de *Simulated Annealing*, o qual não garante a obtenção de mínimo global e em geral, possui alto custo computacional.

Kanade & Okutomi (1994) desenvolveram um algoritmo para selecionar de maneira autônoma um tamanho apropriado da janela de correlação em função da variação local da intensidade luminosa e da disparidade, uma vez que regiões com texturas homogêneas (baixa variação da intensidade luminosa) necessitam de janelas maiores para diminuir a redundância de informação e regiões com grande variação de seus respectivos valores de disparidades necessitam de janelas pequenas para reduzir os efeitos da distorção projetiva. Os autores empregaram um modelo estatístico para a distribuição da disparidade dentro da janela. Este modelo permitiu estimar quanto à variação da disparidade, bem como à variação da intensidade dentro da janela afeta a incerteza na estimativa da disparidade para o ponto central da janela. Assim, um método de busca para uma janela que produz uma estimativa da disparidade com a menor incerteza para cada ponto da imagem foi desenvolvido. Este método controlou não somente o tamanho, mas também a forma da janela (quadrada ou retangular). O maior problema deste algoritmo é que as distribuições da disparidade e do ruído foram modeladas por distribuições gaussianas com média zero, o que, a rigor, não pode ser generalizado para situações práticas.

A abordagem proposta utiliza um esquema hierárquico a fim de controlar o tamanho da janela de correspondência no qual as medidas de correspondência oriundas dos menores tamanhos de janela possuem prioridade em relação as medidas oriundas de janelas maiores.

Hariyama et. al. 2001 desenvolveram um algoritmo no qual uma medida de similaridade é confiável se um gráfico de todas as medidas de similaridade de um ponto em uma imagem em relação a todos os seus possíveis homólogos possui um único mínimo. Para isto uma medida de confiança foi implementada baseada na diferença entre os dois melhores valores de similaridade divididos pelo quadrado do comprimento da janela. A medida de similaridade adotada é a que possui melhor medida de confiança.

A abordagem proposta considera inicialmente os pontos correspondentes não ambíguos (pontos que possuem um único mínimo) e impõe sobre estes a Injunção de Ordem. Esta imposição é realizada em um processo iterativo, no qual, pontos ambíguos em

uma iteração podem ser não ambíguos na próxima iteração devido a Injunção de Ordem imposta sobre outros pontos nas iterações anteriores.

Capítulo 4

4. Abordagem Proposta para Solução do Problema de Correspondência

Este capítulo apresenta a abordagem proposta para solucionar o Problema de Correspondência em Visão Estéreo. Em termos gerais, esta abordagem foi desenvolvida para atender os requisitos de reconstrução da estrutura tridimensional da cena conjuntamente à identificação de oclusões sob as condições descritas abaixo:

Condições de Contorno para a Abordagem Proposta

A abordagem proposta considera com a fonte de informação apenas um par de imagens estereoscópicas. Sem perda de generalidade, as imagens devem estar retificadas, de tal forma que não apresentem disparidade na direção y .

Outra condição considerada nesta abordagem é que a cena é composta por objetos opacos (não transparentes), o que impossibilita um *pixel* em uma imagem ser homólogo a dois ou mais *pixels* na outra imagem. No caso de existirem superfícies transparentes na cena, faz-se necessário utilizar sensores ativos, tal como sensores baseados em *laser* ou sonar, por exemplo, a fim de reconstruir também as superfícies transparentes, que por sua vez, está fora do escopo deste trabalho.

Ainda, outra condição a ser considerada é que a cena é formada por uma função, de tal modo que exista apenas um ponto sobre a superfície da cena para cada par de coordenadas (X, Y) no sistema de coordenadas da cena. A coordenada Z é a profundidade.

Por fim, a última condição necessária para a aplicação da abordagem proposta é que a cena possua textura. Esta consideração não descarta a aplicação da abordagem para cenas que apresentem regiões com texturas homogêneas, desde que essas regiões não ocupem todas as áreas das imagens.

Requisitos para a Abordagem Proposta

A abordagem proposta deve considerar a presença de regiões de oclusão no par de imagens estereoscópicas. Para essas regiões, os valores de disparidade associados não podem ser determinados. Um atributo deve então ser utilizado a fim de identificar de maneira única os *pixels* pertencentes às regiões de oclusão.

Outro requisito é que a abordagem possua poucos parâmetros cujos valores sejam determinados de maneira empírica. Este requisito se faz importante para tornar a abordagem aplicável para vários tipos de cenas e diferentes condições de captura das imagens.

4.1 Detecção de Oclusão - Correspondência Bidirecional e Injunção de Ordem

O problema de detecção de oclusão é resolvido através da utilização conjunta da Correspondência Bidirecional e da imposição da Injunção de Ordem.

A Correspondência Bidirecional consiste em analisar as medidas de correspondência da imagem da esquerda em relação à imagem da direita e vice-versa.

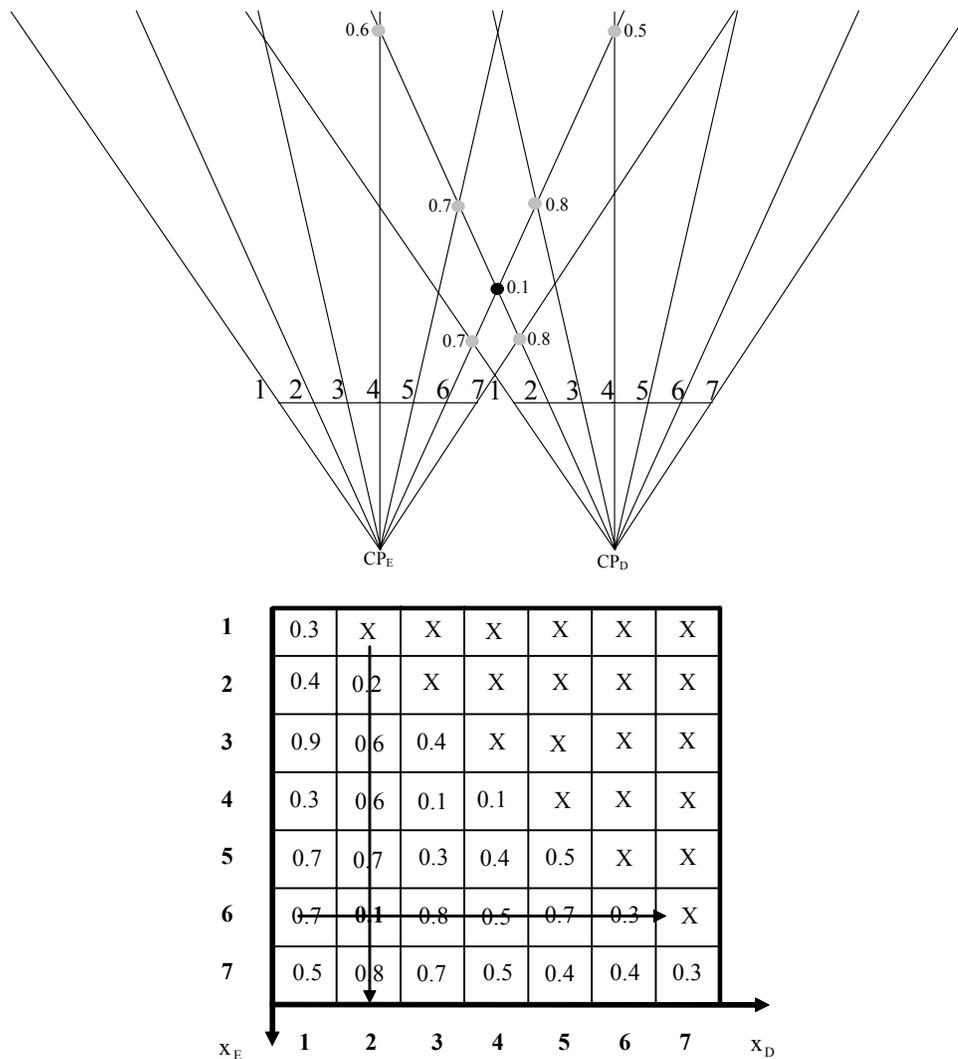
O princípio que sustenta a utilização da Correspondência Bidirecional para a detecção de oclusão é que um par de *pixels* homólogos p e p' deve apresentar Correspondência Bidirecional, ou seja, a melhor medida de correspondência para um *pixel* p na imagem da esquerda é com um *pixel* p' na imagem da direita, que por sua vez, possui melhor medida de correspondência para p .

Desconsiderando a existência de superfícies transparentes na cena e uma vez que oclusões são *pixels* na imagem da esquerda ou da direita que não possuem homólogos, as linhas ou colunas na Matriz de Correspondência que não possuem pares correspondentes bidirecionais são "fortes candidatos" a serem classificados como **oclusões**.

Este princípio é melhor compreendido através de um exemplo, no qual, considera-se o *pixel* 6 na imagem da esquerda homólogo ao *pixel* 2 na imagem da direita. Na Figura 4.1 este par está representado em preto e em cinza os demais pares que não são homólogos em uma representação esquemática das câmeras. Alguns valores das medidas de correspondência estão dispostos ao lado dos pares possíveis. Ainda, esta Figura ilustra

abaixo a respectiva Matriz de Correspondências, na qual a seta horizontal dentro da matriz representa as medidas de correspondência com relação ao *pixel 6* na imagem da esquerda como referência e a seta vertical mostra as medidas de correspondência com relação ao *pixel 2* na imagem da direita como referência.

As medidas de Correspondência estão codificadas com 0 a melhor medida possível e 1 a pior medida possível.



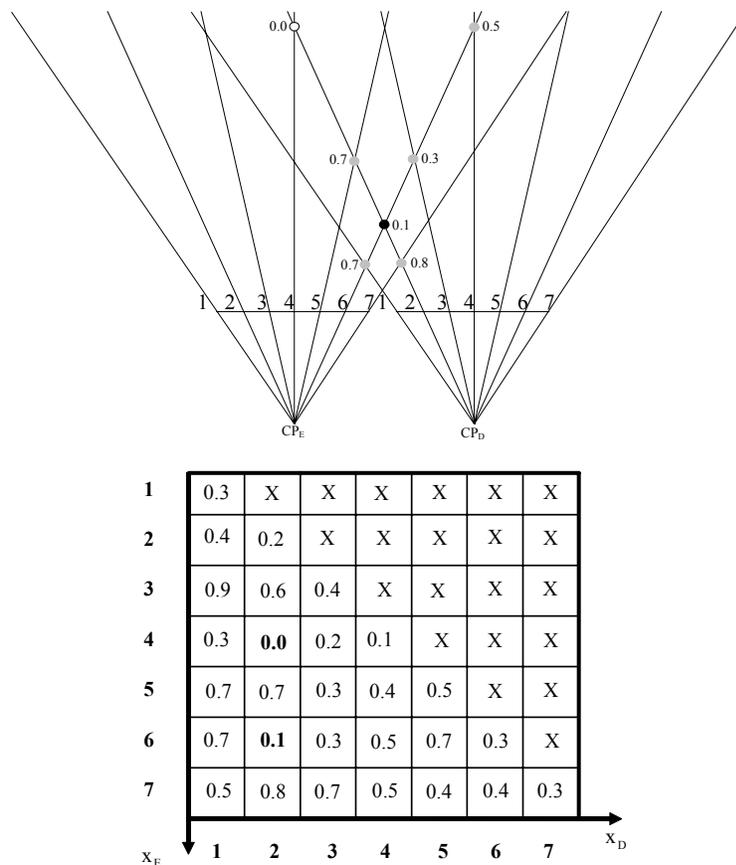
X = Pares impossíveis

Fig. 4.1 – Exemplo em que o par (6,2) é Correspondência Bidirecional na representação esquemática das câmeras e na respectiva Matriz de Correspondências.

Utilizando a Matriz de Correspondência, a Correspondência Bidirecional é implementada realizando a busca das melhores medidas de Correspondência ao longo das

linhas e colunas da Matriz de Correspondência. A posição (6,2) é uma Correspondência Bidirecional porque ao longo da linha 6 a melhor medida é com a coluna 2 e ao longo da coluna 2 a melhor medida é com a linha 6. Nenhum outro valor de correspondência ao longo da linha 6 e da coluna 2 é menor que 0.1. Assim a melhor correspondência para o *pixel* 6 na imagem da esquerda é com o *pixel* 2 na imagem da direita e para o *pixel* 2 na imagem da direita, a melhor correspondência é com o *pixel* 6 na imagem da esquerda.

Já a Figura 4.2, mostra um exemplo semelhante ao da Figura 4.1, em que entre os pares possíveis para o *pixel* 6 ((6,1), (6,2), (6,3),...) na imagem da esquerda, a melhor correspondência é com o *pixel* 2 na imagem da direita, representado em preto. Porém, para o *pixel* 2 na imagem da direita, entre os pares possíveis para este ponto ((7,2), (6,2), (5,2),...) a melhor correspondência se dá com o *pixel* 4 na imagem da esquerda, representado em branco.



X = Pares impossíveis

Fig. 4.2 – Exemplo em que o par (6,2) não é Correspondência Bidirecional na representação esquemática das câmeras e na respectiva Matriz de Correspondências.

A posição (6,2) não é uma Correspondência Bidirecional porque ao longo da linha 6 a melhor medida é com a coluna 2 (o valor 0.1), mas ao longo da coluna 2 a melhor medida é com a linha 4 (o valor 0.0). Assim, tomando a imagem da esquerda como referência, o *pixel* 6 na imagem da esquerda deveria ser homólogo ao *pixel* 2 na imagem da direita, porém adotando a imagem da direita como referência, o *pixel* 2 nesta imagem deveria ser homólogo ao *pixel* 4 na imagem da esquerda.

Este último exemplo mostra uma situação que se verifica em situações reais apenas se a superfície da cena na posição (6,2) é transparente ou semi-transparente (o par (4,2) é visível a partir do *pixel* 2 na imagem da direita somente se o par (6,2) é transparente). Este exemplo mostra o único caso em que a Injunção de Unicidade falha. A Injunção de Ordem também falha para superfícies transparentes.

Apesar de a Correspondência Bidirecional possibilitar a identificação de oclusões, esta heurística é utilizada neste trabalho com o intuito de determinar com melhor precisão os pares homólogos, e não os *pixels* que não possuem homólogos (oclusões).

A utilização da Correspondência Bidirecional para identificação de oclusões implica no emprego de um limiar, sob o qual, pares de *pixels* que possuam valores de correspondência acima deste limiar são classificados como homólogos. O *pixel* em uma imagem que possua valores de correspondência abaixo deste limiar para todos os possíveis *pixels* candidatos a homólogos na outra imagem é classificado como oclusão.

A dificuldade em utilizar a Correspondência Bidirecional para identificação de oclusões está na própria determinação deste limiar, o que é em geral, determinado de maneira empírica, tornando a qualidade da reconstrução obtida pela abordagem extremamente dependente deste limiar (parâmetro).

Uma alternativa mais robusta para o problema descrito acima é identificar oclusões a partir de relações entre *pixels* vizinhos que se verificam no mapa de disparidade.

A utilização da Injunção de Ordem, ou de maneira análoga, a utilização de um limite para o módulo do Gradiente de Disparidade igual a 2, como já citado no capítulo 2, impõe que oclusões ocorram sempre que houver variação dos valores de disparidade associados à *pixels* vizinhos.

Utilizando a notação empregada na seção 2.2.2, onde $x_E = x_{E1}$ e $x_D = x_{D1}$ são as coordenadas na direção x do par 1 na imagem da esquerda e da direita, respectivamente e $x_E = x_{E2}$ e $x_D = x_{D2}$ de maneira análoga, porém para o par 2, as coordenadas deste par 2 podem ser definidas em função das coordenadas do par 1 por:

$$\begin{aligned} k_{E2} &= x_{E1} + M \\ k_{D2} &= x_{D1} + N \end{aligned} \quad (4.1)$$

onde:

$$\begin{aligned} M &\in \mathfrak{R}^+ \\ N &\in \mathfrak{R}^+ \end{aligned} \quad (4.2)$$

Assim, para um par de *pixels* vizinhos 1 e 2, com disparidades d_1 e d_2 , respectivamente, a diferença $\Delta d = d_1 - d_2$ fica:

$$\Delta d = (x_{E1} - x_{D1}) - (x_{E2} - x_{D2}) \quad (4.3)$$

Aplicando 4.1 em 4.3 tem-se que:

$$\Delta d = (x_{E1} - x_{D1}) - (x_{E1} + M - (x_{D1} + N)) \quad (4.4)$$

que implica em:

$$\Delta d = -M + N \quad (4.5)$$

Para imagens discretas e considerando dois *pixels* vizinhos na imagem da esquerda, M assume o valor 1, assim:

$$\Delta d = -1 + N \quad (4.6)$$

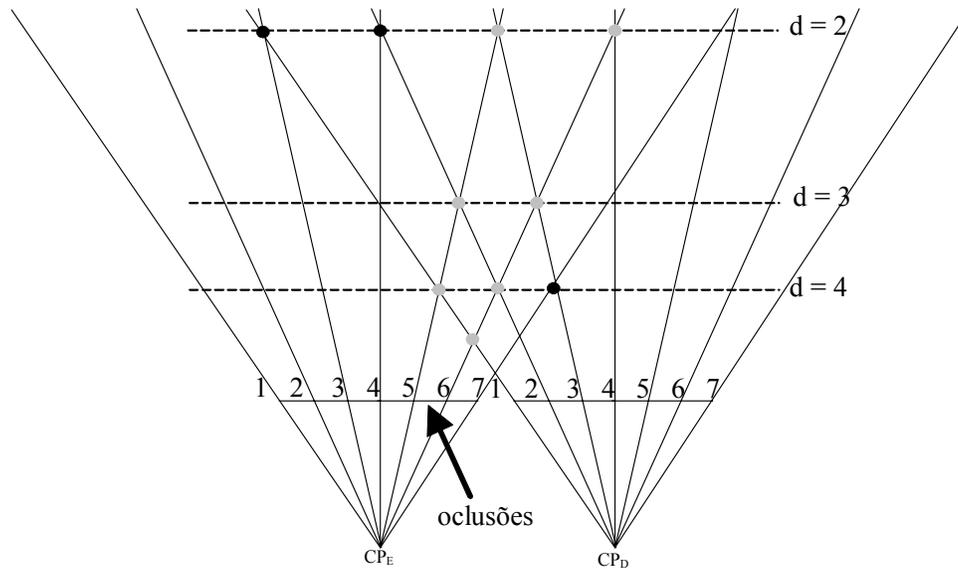
A expressão 4.6 mostra que respeitando a Injunção de Ordem, faz-se necessário existir uma região com $N-1$ *pixels* de oclusão na imagem da direita se a diferença das disparidades associadas a estes *pixels* vizinhos é Δd . Com isso, pode-se deduzir um importante resultado: "a diferença de disparidade $\Delta d = 1$ na imagem da esquerda entre dois *pixels* vizinhos, implica em $N = 2$ para satisfazer a igualdade em 4.6 e, portanto, existe $N-1 = 2-1 = 1$ *pixel* na imagem da direita que é oclusão". Para outros valores de Δd o raciocínio é análogo.

Considerando dois *pixels* vizinhos na imagem da direita (e imagens discretas), N também assume o valor 1 e assim a expressão 4.5 fica:

$$\Delta d = -M + 1 \tag{4.7}$$

A expressão 4.7 mostra que respeitando a Injunção de Ordem, faz-se necessário existir uma região com $M-1$ *pixels* de oclusão na imagem da esquerda se a diferença das disparidades associadas a estes *pixels* vizinhos é Δd .

A Figura 4.3 ilustra a ocorrência de oclusões e mostra em preto pares corretos e em cinza os pontos de oclusão na representação esquemática das câmeras e a respectiva Matriz de Correspondências com os pares corretos e o efeito da Injunção de Ordem para estes pares.



1	W	X	X	X	X	X	X		
2	W	W	X	X	X	X	X		
3	O	W	W	X	X	X	X		
4	W	O	W	W	X	X	X		
5	W	W	W	W	W	X	X		
6	W	W	W	W	W	W	X		
7	W	W	O	W	W	W	W		
	x_E	1	2	3	4	5	6	7	x_D

O = Par Correto

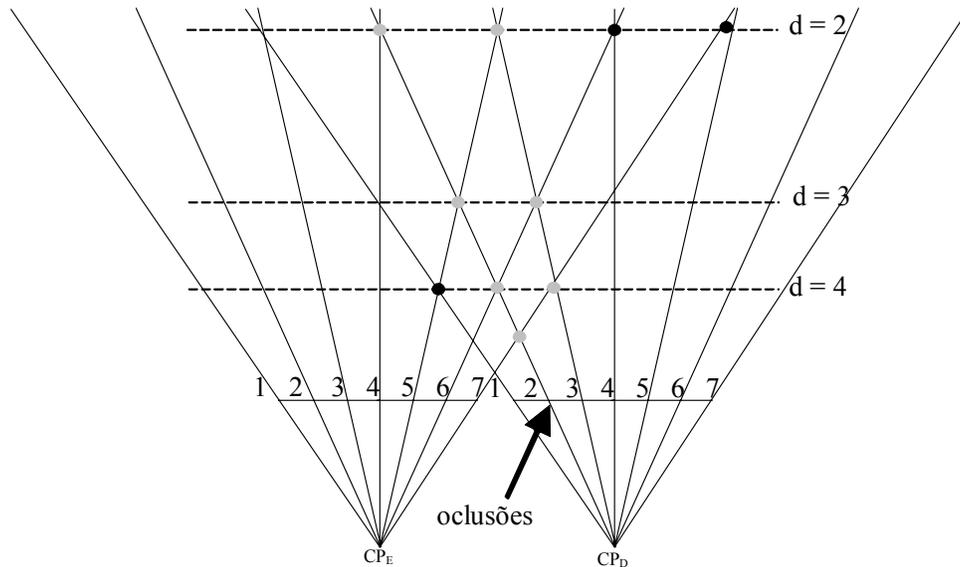
W = Pares impossíveis para Injunção de Ordem

X = Pares impossíveis

Fig. 4.3 – Oclusão na imagem da esquerda na representação esquemática das câmeras e efeito conjunto na Matriz de Correspondências da Injunção de Ordem e de Disparidade Não Negativa ocasionando oclusão na imagem da esquerda.

Na Figura 4.3, os *pixels* 3 e 4 na imagem da esquerda são homólogos aos *pixels* 1 e 2 na imagem da direita, respectivamente, resultando em uma disparidade $d = 2$ para ambos. Já os *pixels* 5 e 6 na imagem da esquerda não podem ter *pixels* homólogos na imagem da direita, uma vez que o *pixel* 7 na imagem da esquerda é homólogo ao *pixel* 3 na imagem da direita, o que resulta em uma disparidade $d = 4$ associada a estes *pixels*. Com isso, existe uma variação de disparidade de $\Delta d = 2 - 4 = -2$, determinando uma região de oclusão na imagem da esquerda de 2 *pixels* (os *pixels* 5 e 6 na imagem da esquerda).

Dando seqüência ao exemplo, a Figura 4.4 mostra um caso de oclusão na imagem da direita em que os *pixels* em preto representam pares corretos e em cinza os *pixels* de oclusão na representação esquemática das câmeras e a respectiva Matriz de Correspondências.



1	W	X	X	X	X	X	X
2	W	W	X	X	X	X	X
3	W	W	W	X	X	X	X
4	W	W	W	W	X	X	X
5	O	W	W	W	W	X	X
6	W	W	W	O	W	W	X
7	W	W	W	W	O	W	W

x_E 1 2 3 4 5 6 7 x_D
2 3
 Oclusões

X = Pares impossíveis

O = Par Correto

W = Pares impossíveis para Injunção de Ordem

Fig. 4.4 – Oclusão na imagem da direita na representação esquemática das câmeras e efeito conjunto na Matriz de Correspondências da Injunção de Ordem e de Disparidade Não Negativa ocasionando oclusão na imagem da direita.

Na Figura 4.4, o *pixel 5* na imagem da esquerda é homólogo ao *pixel 1* na imagem da direita, resultando em uma disparidade $d = 4$. Os *pixels 6* e *7* na imagem da esquerda são homólogos aos *pixels 4* e *5* na imagem da direita, resultando em uma disparidade $d = 2$ para ambos. Assim, os *pixels 2* e *3* na imagem da direita não possuem homólogos na imagem esquerda. Existe uma variação de disparidade de $\Delta d = 4 - 2 = 2$, determinando uma região de oclusão na imagem da esquerda de 2 *pixels* (os *pixels 2* e *3* na imagem da direita).

A partir das considerações tratadas nesta seção, conclui-se que a determinação dos *pixels* homólogos com imposição da Injunção de Ordem ocasiona, conseqüentemente, a detecção das regiões de oclusão. Esta é a heurística que fundamenta a utilização da Injunção de Ordem na abordagem proposta.

No entanto, faz-se necessário determinar um critério para quantificar sobre quais pares (os melhores) deve-se impor a Injunção de Ordem.

4.2 Tamanho da Janela de Correlação - Correspondência Bidirecional Hierárquica

Existem várias possibilidades para se determinar um critério que defina sobre quais pares deve-se impor a Injunção de Ordem. Um critério simples é escolher qual a melhor medida de correspondência associada a cada *pixel*, por exemplo, na imagem da esquerda e então impor a Injunção de Ordem para estes *pixels*. Utilizando a Matriz de Correspondência, este critério se dá analisando qual a melhor medida de correspondência ao longo das linhas dessa matriz.

Embora simples este critério não é robusto, uma vez que os resultados obtidos para os *pixels* na imagem da esquerda (analisando a Matriz de Correspondência ao longo das linhas dessa matriz) podem diferir dos resultados obtidos para os *pixels* na imagem da direita (analisando a Matriz de Correspondência ao longo das colunas dessa matriz).

Um critério mais robusto é escolher como *pixels* homólogos aqueles que apresentam a propriedade de Correspondência Bidirecional (como citado na seção anterior) e então impor a Injunção de Ordem para estes pares de *pixels*.

Utilizando a técnica de Correspondência baseada em Áreas (capítulo 2), as medidas de Correspondência são, entre outras, função do tamanho da janela de correspondência utilizada. Assim, o problema do tamanho da janela torna-se evidente.

Como já citado no capítulo 2 (seção 2.3.1), janelas pequenas são necessárias em regiões onde a superfície da cena não é fronto-planar, entretanto, a pequena quantidade de *pixels* (amostras) contidas na janela tende a aumentar a ambigüidade das medidas de correspondência. Janelas grandes possuem uma maior quantidade de *pixels* (amostras), o que tende a minimizar as ambigüidades devido às medidas de correlação serem mais precisas, porém, a posição de melhor medida de correspondência pode não resultar em um valor de disparidade correto devido à janela cobrir uma região referente a superfícies não necessariamente fronto-planares.

A maneira proposta neste trabalho de minimizar o problema do tamanho da janela é quantificar os pares que possuem a propriedade de Correspondência Bidirecional a partir de medidas de correspondências oriundas de vários tamanhos de janela.

A heurística desta abordagem é que um par que possui a propriedade de Correspondência Bidirecional a partir de medidas oriundas de vários tamanhos de janela deve ser provavelmente formado por *pixels* homólogos de que aqueles que não são.

De maneira resumida, a heurística citada acima diz que um par de *pixels* são homólogos se estes possuem a propriedade de Correspondência Bidirecional independente do tamanho da janela de correlação. No entanto, esta heurística não considera o problema de preservar as descontinuidades da cena.

Para tentar preservar as descontinuidades da cena, pode-se quantificar quais os pares que são Correspondências Bidirecionais primeiramente para as medidas oriundas do menor tamanho de janela utilizado e então, posteriormente, para os demais tamanhos de janelas organizadas de maneira hierárquica em ordem crescente, de tal forma que um par apenas poderá ser Correspondência Bidirecional se este mesmo par é também Correspondência Bidirecional para as medidas oriundas das janelas menores. Esta organização hierárquica impõe o conceito de níveis de informações, onde os níveis inferiores correspondem às medidas oriundas das janelas menores e os níveis superiores às medidas oriundas das janelas maiores.

Esta hierarquia é do tipo "*bottom - up*" e esta foi adotada (e não ao contrário, do tipo "*top-down*") porque nas regiões de descontinuidades as janelas pequenas sofrem muito menos o efeito da variação local da disparidade. A desvantagem em utilizar janelas pequenas é o aumento do número de ambigüidades, dificultando a resolução do Problema de Correspondência, como já citado. A fim de minimizar esta dificuldade, ou seja, reduzir o número de ambigüidades preservando as descontinuidades, é que se utilizam janelas maiores de maneira hierárquica.

É importante ressaltar que esta abordagem não se aplica apenas a medidas de correspondências oriundas da Correspondência baseada em Áreas, mas sim a qualquer outra técnica que quantifique, sob algum critério ou métrica, a correspondência estéreo.

É de se esperar que os pares que possuem a propriedade de Correspondência Bidirecional em um número maior de níveis hierárquicos tenham maiores chances de estarem corretos (ser realmente formado por *pixels* homólogos) do que aqueles que são Correspondência Bidirecional em apenas poucos níveis hierárquicos. Com isso, deve-se então aplicar a Injunção de Ordem sobre os pares com as maiores chances entre todos os demais.

Ao impor a Injunção de Ordem para os pares com maiores prioridades, restarão provavelmente ainda alguns pares com prioridades menores e que não foram classificados como pares impossíveis em decorrência da imposição da Injunção de Ordem para os pares com maiores prioridades. No entanto, a imposição da Injunção de Ordem para os pares com maiores prioridades pode fazer com que a prioridade de outros pares com prioridades mais baixas aumente em uma nova verificação das prioridades. Isto ocorre porque a imposição da Injunção de Ordem sobre um par classifica outros pares como impossíveis.

Após a imposição da Injunção de Ordem sobre os pares com maiores prioridades faz-se necessário verificar se ainda existem pares não resolvidos. Caso ainda existam pares não resolvidos, faz-se necessário reavaliar a Correspondência Bidirecional para estes pares e quantificar as suas prioridades para então aplicar a Injunção de Ordem novamente. Este fato impõe a necessidade de um processo iterativo.

O fato citado pode ser visto também como uma restrição do espaço de busca do *pixel* homólogo devido à ocorrência de outro par, o que torna a abordagem proposta um Processo Cooperativo, uma vez que a ocorrência de um par de *pixels* homólogos favorece a

ocorrência de outros pares na região que a Injunção de Ordem é respeitada enquanto elimina a ocorrência dos pares na região que esta injunção é violada.

Com isso, o processo iterativo termina quando não houver mais pares não resolvidos (ainda ambíguos) para um par de linhas epipolares homólogas.

4.3 Abordagem Proposta

O objetivo deste item é descrever a abordagem proposta embasada nos fundamentos citados em 4.1 e 4.2.

4.3.1 Matriz de Prioridades e Processo Iterativo

A fim de organizar as prioridades de cada par, uma Matriz de Prioridades P , com a mesma estrutura de uma Matriz de Correspondências, porém com uma dimensão a menos, é montada, onde cada elemento desta matriz possui uma medida de prioridade para se impor a Injunção de Ordem.

A hierarquia do tipo "*bottom - up*" é descrita da seguinte maneira: “um elemento $P(x_E, x_D)$ na Matriz de Prioridades, cuja medida de correspondência entre x_E e x_D para um nível hierárquico $n + 1$ resulte em uma Correspondência Bidirecional, somente terá prioridade $n + 1$ se a medida de correspondência para estes mesmos *pixels* para o nível hierárquico n também resulte em uma Correspondência Bidirecional”.

Cabe ressaltar que nesta abordagem também se considera como Correspondência Bidirecional quando há empate dos valores de melhor medida em uma mesma linha ou mesma coluna. Esta consideração é devida à discretização dos valores de intensidade (níveis de cinza) das imagens digitais (fato que possibilita haver o “empate” citado) e principalmente devido à abordagem ser baseada em um processo hierárquico, no qual os níveis hierárquicos mais altos tendem a ser menos ambíguos, possibilitando haver o “desempate” nestes níveis.

O processo hierárquico descrito necessita ser realizado de maneira iterativa, no qual, a cada iteração são selecionados os pares de *pixels* com maiores prioridades e então imposta

a Injunção de Ordem sobre estes até que não existam mais pares ambíguos em um par de linhas epipolares homólogas.

É importante ressaltar também que a correspondência baseada em áreas, geralmente um processo de alto custo computacional, é realizada apenas uma vez, embora para vários tamanhos de janela. A cada iteração as Matrizes de Correspondência são apenas atualizadas com rótulos que representam pares impossíveis devido à imposição da Injunção de Ordem.

A principal idéia que suporta este processo iterativo é que, para cada par verdadeiro existem vários outros pares falsos, de tal forma que a ocorrência de um par verdadeiro auxilia de maneira cooperativa os demais pares verdadeiros ainda ambíguos, enquanto que ao mesmo tempo desfavorece os pares falsos, porém ainda ambíguos e que poderiam vir a ser classificados erroneamente como pares verdadeiros.

Com o intuito de formalização, pode-se definir a seguinte regra para o evento descrito:

Regra 1

Selecionar pares correspondentes e impor a Injunção de Ordem para os pares com maiores prioridades em uma iteração.

4.3.2 Regras Complementares

A regra 1, embora básica, é insuficiente para estabelecer completamente a abordagem proposta. Em vista deste fato, algumas regras complementares devem ser estabelecidas.

O principal problema que ocorre é quando um par com máxima prioridade em uma iteração está localizado na região de pares impossíveis determinada por outro par também com máxima prioridade na mesma iteração. Neste caso a imposição da Injunção de Ordem sobre um par implica na classificação do outro como par impossível. Este fato ocorre porque as prioridades são grandezas discretas, o que possibilita haver empate de dois ou mais pares localizados em regiões conflitantes com a mesma prioridade máxima em uma iteração.

A fim de sanar este problema, a seguinte regra pode ser utilizada:

Regra 2

No caso em que dois ou mais pares com prioridade máxima ocorram em regiões conflitantes em uma mesma iteração deve-se diminuir a prioridade de -1 daquele que, para ocorrer, seja necessário classificar como par impossível o maior número de pares com prioridade máxima.

Em caso de empate do número de pares que ocorrem na região conflitante, a seguinte regra é utilizada:

Regra 3

Diminuir a prioridade de -1 daquele par que tiver maior disparidade (menor profundidade).

A regra 3 não é suportada por nenhuma idéia de caráter lógico e deve ser entendida apenas como uma decisão que tem que ser tomada quando os demais critérios de decisão empataram.

O exemplo a seguir ilustra a maneira que se dá a imposição da Injunção de Ordem sob os pares de *pixels* homólogos com maiores prioridades.

EXEMPLO 1 – UTILIZAÇÃO DA REGRA 1

A Figura 4.5 mostra uma Matriz de Correspondência cujos valores foram obtidos realizando a Correspondência baseada em áreas com tamanho de janela igual a 1 (nível hierárquico 1) e função de correlação SDA. Os elementos da matriz com “X” representam pares impossíveis e os valores em negrito representam os pares que possuem a propriedade de Correspondência Bidirecional.

A Figura 4.6 mostra o mesmo que a Figura 4.5, porém as medidas de correspondência foram obtidas com tamanho de janela igual a 3 (nível hierárquico 2).

1	0.3	X	X	X	X	X	X
2	0.7	0.2	X	X	X	X	X
3	0.1	0.6	0.4	X	X	X	X
4	0.3	0.6	0.1	0.5	X	X	X
5	0.7	0.7	0.3	0.1	0.5	X	X
6	0.7	0.1	0.8	0.5	0.2	0.3	X
7	0.5	0.8	0.7	0.5	0.4	0.1	0.3

x_E ↓ 1 2 3 4 5 6 7 x_D

Fig. 4.5 – Medidas de correspondência para tamanho de janela = 1 (nível 1).

1	0.3	X	X	X	X	X	X
2	0.7	0.2	X	X	X	X	X
3	0.1	0.6	0.4	X	X	X	X
4	0.3	0.6	0.1	0.5	X	X	X
5	0.7	0.7	0.3	0.1	0.5	X	X
6	0.7	0.6	0.8	0.5	0.4	0.3	X
7	0.5	0.8	0.7	0.5	0.4	0.1	0.3

x_E ↓ 1 2 3 4 5 6 7 x_D

Fig. 4.6 – Medidas de correspondência para tamanho de janela = 3 (nível 2).

A Matriz de Prioridades para as Matrizes de Correspondências das Figuras 4.5 e 4.6 é apresentada na Figura 4.7, onde -1 codifica os pares impossíveis. Os elementos desta matriz com valor 0 (zero) indicam que estes pares não possuem a propriedade de Correspondência Bidirecional já para o primeiro nível hierárquico e, portanto, continuaram rotulados com “prioridade = 0” mesmo se algum destes apresentar a propriedade de Correspondência Bidirecional para níveis superiores do processo hierárquico.

O elemento com valor 1, indica que este par apresentou a propriedade de Correspondência Bidirecional apenas para o nível hierárquico 1.

Os elementos com valor 2, indicam que estes pares apresentaram a propriedade de Correspondência Bidirecional para os níveis 1 e 2 do processo hierárquico.

1	0	-1	-1	-1	-1	-1	-1
2	0	0	-1	-1	-1	-1	-1
3	2	0	0	-1	-1	-1	-1
4	0	0	2	0	-1	-1	-1
5	0	0	0	2	0	-1	-1
6	0	1	0	0	0	0	-1
7	0	0	0	0	0	2	0

x_E ↓ 1 2 3 4 5 6 7 x_D

Fig. 4.7 – Matriz de Prioridades para as Matrizes de Correspondências das Figuras 4.5 e 4.6.

Ao selecionar os pares correspondentes e impor a Injunção de Ordem sobre estes pares com “prioridade = 2”, as Matrizes de Correspondências são alteradas. As Figuras 4.8 e 4.9 mostram essas alterações.

1	X	X	X	X	X	X	X	
2	X	X	X	X	X	X	X	
3	0.1	X	X	X	X	X	X	
4	X	X	0.1	X	X	X	X	
5	X	X	X	0.1	X	X	X	
6	X	X	X	X	0.2	X	X	
7	X	X	X	X	X	0.1	X	
x_E	1	2	3	4	5	6	7	x_D

Fig. 4.8 – Medidas de correspondências para tamanho de janela = 1 após Injunção de Ordem para pares com prioridade = 2.

1	X	X	X	X	X	X	X	
2	X	X	X	X	X	X	X	
3	0.1	X	X	X	X	X	X	
4	X	X	0.1	X	X	X	X	
5	X	X	X	0.1	X	X	X	
6	X	X	X	X	0.4	X	X	
7	X	X	X	X	X	0.1	X	
x_E	1	2	3	4	5	6	7	x_D

Fig. 4.9 – Medidas de correspondências para tamanho de janela = 3 após Injunção de Ordem para pares com prioridade = 2.

Como se pode notar nas Figuras 4.8 e 4.9, o elemento (6,5) que anteriormente possuía prioridade = 0 (ver Figura 4.7) agora irá apresentar prioridade = 2, pois possui a propriedade de Correspondência Bidirecional para o nível 1 e nível 2 do processo hierárquico. A Figura 4.10 mostra a Matriz de Prioridades para as Matrizes de Correspondência das Figuras 4.8 e 4.9.

1	0	-1	-1	-1	-1	-1	-1	
2	0	0	-1	-1	-1	-1	-1	
3	2	0	0	-1	-1	-1	-1	
4	0	0	2	0	-1	-1	-1	
5	0	0	0	2	0	-1	-1	
6	0	0	0	0	2	0	-1	
7	0	0	0	0	0	2	0	
x_E	1	2	3	4	5	6	7	x_D

Fig. 4.10 – Matriz de Prioridades para as Matrizes de Correspondências das Figuras 4.8 e 4.9.

Ainda neste exemplo destaca-se que o *pixel* 2 na imagem da direita (2^o coluna) corresponde a uma região de oclusão. É importante ressaltar que esta região de oclusão foi naturalmente classificada devido a imposição da Injunção de Ordem, independente de qualquer outro procedimento exclusivo para detecção de oclusões.

O processo iterativo, neste exemplo se encerra porque não existem mais pares não resolvidos. Porém caso outras iterações houvessem que ser realizadas, a imposição da Injunção de Ordem poderia se dar sobre pares com prioridades mais baixas (no caso, prioridade = 1), desde que não houvesse mais nenhum par com prioridade = 2 ainda sem ser imposta a Injunção de Ordem.

EXEMPLO 2 – UTILIZAÇÃO DA REGRA 1 E 2

Este exemplo é idêntico ao exemplo 1, com exceção apenas do valor da medida de correspondência do elemento (6,2) da Figura 4.6, o qual foi alterado de 0.6 para 0.1.

1	0.3	X	X	X	X	X	X		
2	0.7	0.2	X	X	X	X	X		
3	0.1	0.6	0.4	X	X	X	X		
4	0.3	0.6	0.1	0.5	X	X	X		
5	0.7	0.7	0.3	0.1	0.5	X	X		
6	0.7	0.1	0.8	0.5	0.2	0.3	X		
7	0.5	0.8	0.7	0.5	0.4	0.1	0.3		
	x_E	1	2	3	4	5	6	7	x_D

Fig. 4.11 – Medidas de correspondências para tamanho de janela = 1.

1	0.3	X	X	X	X	X	X		
2	0.7	0.2	X	X	X	X	X		
3	0.1	0.6	0.4	X	X	X	X		
4	0.3	0.6	0.1	0.5	X	X	X		
5	0.7	0.7	0.3	0.1	0.5	X	X		
6	0.7	0.1	0.8	0.5	0.4	0.3	X		
7	0.5	0.8	0.7	0.5	0.4	0.1	0.3		
	x_E	1	2	3	4	5	6	7	x_D

Fig. 4.12 – Medidas de correspondências para tamanho de janela = 3.

A Matriz de Prioridades para as Matrizes de Correspondências das Figuras 4.11 e 4.12 é apresentada na Figura 4.13.

1	0	-1	-1	-1	-1	-1	-1
2	0	0	-1	-1	-1	-1	-1
3	2	0	0	-1	-1	-1	-1
4	0	0	2	0	-1	-1	-1
5	0	0	0	2	0	-1	-1
6	0	2	0	0	0	0	-1
7	0	0	0	0	0	2	0

x_E ↓ 1 2 3 4 5 6 7 x_D

Fig. 4.13 – Matriz de Prioridades para as Matrizes de Correspondências das Figuras 4.11 e 4.12.

Neste caso, se a Injunção de Ordem for imposta sobre o elemento (6,2) da Matriz de Prioridade da Figura 4.13, os elementos (4,3) e (5,4) desta mesma Matriz serão classificados como pares impossíveis, ou seja, a ocorrência de (6,2) implica na eliminação de dois pares com prioridade máxima. A Figura 4.14 mostra em cinza a região de pares impossíveis determinada pela Injunção de Ordem sobre (6,2). Já para o elemento (4,3) ocorrer, implica apenas na eliminação de (6,2), ou seja, a ocorrência de (4,3) implica na eliminação de um par com prioridade máxima. O mesmo é válido para o elemento (5,4). A Figura 4.15 mostra em cinza a região de pares impossíveis determinada pela Injunção de Ordem sobre (4,3) e (5,4).

1	0	-1	-1	-1	-1	-1	-1
2	0	0	-1	-1	-1	-1	-1
3	2	0	0	-1	-1	-1	-1
4	0	0	2	0	-1	-1	-1
5	0	0	0	2	0	-1	-1
6	0	2	0	0	0	0	-1
7	0	0	0	0	0	2	0

x_E ↓ 1 2 3 4 5 6 7 x_D

Fig. 4.14 – Região de pares impossíveis determinada pela Injunção de Ordem sobre (6,2).

1	0	-1	-1	-1	-1	-1	-1
2	0	0	-1	-1	-1	-1	-1
3	2	0	0	-1	-1	-1	-1
4	0	0	2	0	-1	-1	-1
5	0	0	0	2	0	-1	-1
6	0	2	0	0	0	0	-1
7	0	0	0	0	0	2	0

x_E ↓ 1 2 3 4 5 6 7 x_D

Fig. 4.15 – Região de pares impossíveis determinada pela Injunção de Ordem sobre (4,3) e (5,4).

Discussão

A abordagem proposta não utiliza uma medida de suavidade como, em geral, ocorre nas abordagens que fazem uso de uma função-objetivo explícita, embora o procedimento para cada par de linhas epipolares homólogas seja realizado de maneira global.

Outra característica importante desta abordagem é a não exigência de segmentação das imagens estereoscópicas, que de modo geral dependem de uma grande quantidade de parâmetros, usualmente determinados de maneira empírica.

Capítulo 5

5. Algoritmo e Implementação

Este capítulo apresenta, inicialmente, as formulações para a Medida de Correspondência utilizada, bem como, os detalhes de implementação computacional que contribuem para a execução do algoritmo de maneira mais rápida e eficiente da tabela para implementação da Injunção de Ordem. Posteriormente é apresentado o algoritmo para a abordagem proposta no capítulo 4.

5.1 Medida de Correspondência

A medida de correspondência empregada nesta abordagem é a Somatória das Diferenças Absolutas (SDA) entre uma janela na imagem esquerda e uma janela na imagem direita. Esta medida de correspondência, a qual formalmente é uma medida de correlação, foi adotada por ser bastante simples e apresentar um menor custo computacional em relação a várias outras medidas de correlação, tais como, Somatória das Diferenças Quadráticas ou Coeficiente de Correlação. A SDA é dada por:

$$SDA(y, x_E, x_D) = \sum_{i=-\frac{S_i}{2}}^{\frac{S_i}{2}} \sum_{j=-\frac{S_j}{2}}^{\frac{S_j}{2}} |p_E(y+i, x_E+j) - p_D(y+i, x_D+j)| \quad (5.1)$$

onde:

S_i é o tamanho da janela na direção i ;

S_j é o tamanho da janela na direção j ;

$p_E(.,.)$ e $p_D(.,.)$ são os valores de intensidade dos *pixels* na imagem da esquerda e da direita, respectivamente.

Utilizando imagens com níveis de cinza representados com 8 *bits* e $S_i = S_j = 1$, o maior valor para SDA é 255 e o menor é 0 (zero). Cabe ressaltar que a medida 0 (zero) implica em melhor medida possível e 255 em pior medida possível.

Tais medidas $SDA(y, x_E, x_D)$ podem ser normalizadas para o intervalo [0,1] fazendo:

$$SDA^n(y, x_E, x_D) = \frac{SDA(y, x_E, x_D)}{255 \times S_i \times S_j} \quad (5.2)$$

onde:

$SDA^n(y, x_E, x_D)$ é a medida $SDA(y, x_E, x_D)$ normalizada em [0,1].

Dado um tamanho de janela inicial $S_i^1 \times S_j^1$, e considerando níveis hierárquicos, a correlação deve ser calculada utilizando janelas com tamanhos $S_i^1 \times S_j^1$, $S_i^1+2 \times S_j^1+2$, $S_i^1+4 \times S_j^1+4, \dots, S_i^1+2(N-1) \times S_j^1+2(N-1)$, onde N é o nível máximo do processo hierárquico. A SDA fica:

$$SDA(n, y, x_E, x_D) = \sum_{i=-\frac{S_i^n}{2}}^{\frac{S_i^n}{2}} \sum_{j=-\frac{S_j^n}{2}}^{\frac{S_j^n}{2}} |p_E(y+i, x_E+j) - p_D(y+i, x_D+j)| \quad (5.3)$$

onde:

$$S_i^n = S_i^1 + 2(n-1); \quad (5.4)$$

$$S_j^n = S_j^1 + 2(n-1); \text{ e} \quad (5.5)$$

$$n \in \mathbb{Z} \quad | \quad 1 \leq n \leq N \quad (5.6)$$

5.2 Tabela para Implementação da Injunção de Ordem

A maneira mais simples para implementar a imposição da Injunção de Ordem é para cada par classificado como verdadeiro, varrer a Matriz de Prioridade inteira verificando se cada elemento dessa matriz pertence ou não à região de pares impossíveis, determinada por este par verdadeiro.

Embora simples, esta metodologia impõe um alto custo computacional. Supondo por exemplo, que um par de imagens estereoscópicas possua 480 linhas por 640 colunas e que exista 100 pares verdadeiros em cada par de linhas epolares homólogas, faz-se necessário então realizar 480 vezes o procedimento de varrer 100 vezes uma Matriz de Prioridade com dimensão 640 x 640, verificando se cada elemento dessa matriz pertence ou não a região de pares impossíveis, determinada por cada par verdadeiro.

Nesta seção é apresentada uma maneira de implementar a imposição da Injunção de Ordem utilizando uma tabela, o que reduz drasticamente o custo computacional citado acima.

Uma vez que a medida de Correspondência utilizada na metodologia proposta é associada ao *pixel* central da janela de Correspondência, não é possível calcular esta medida nas regiões limítrofes das imagens estereoscópicas. A exceção é apenas quando o tamanho da janela de correspondências é $s_i \times s_j = 1 \times 1$. No caso de $s_i \times s_j = 3 \times 3$, por exemplo, não é possível calcular a Medida de Correspondência para todos os *pixels* da primeira e última linha e da primeira e última coluna.

A tabela criada nesta implementação possui na primeira coluna representação tanto para as coordenadas x_E quanto para as coordenadas x_D , ou seja, o valor nesta coluna serve tanto para indicar a coordenada x na imagem da esquerda quanto para indicar a coordenada x na imagem da direita.

As colunas x_{E1} e x_{E2} contém as coordenadas inicial e final, respectivamente, na imagem da esquerda dos pares possíveis para cada x_D na coluna “ $x_E \& x_D$ ”. Já, as colunas x_{D1} e x_{D2} contém as coordenadas inicial e final, respectivamente, na imagem da direita dos pares possíveis para cada x_E na coluna “ $x_E \& x_D$ ”. Para regiões de pares impossíveis, utiliza-se o rótulo “-1”.

O quadro 5.1 mostra à esquerda a tabela criada para um par de imagens estereoscópicas com largura (w) = 15 e a direita a Matriz de Prioridade, onde x representa pares impossíveis. Nesta tabela tanto a linha = 15 e a coluna = 0 estão rotuladas com x em decorrência do tamanho da janela utilizada ao se realizar a Correspondência baseada em Áreas, que neste exemplo é $s_i \times s_j = 3 \times 3$.

Cabe ressaltar que na Matriz de Prioridade, as linhas são as coordenadas x_E e as colunas são as coordenadas x_D .

O valor 8, por exemplo, na coluna “ $x_E \& x_D$ ”, indica que $x_E = 8$, pode ser homólogo a $x_{D1} = 1$ até $x_{D2} = 8$ sendo que a Injunção de Ordem não é violada. Estes possíveis homólogos são representados em cinza claro no quadro 5.1.

O valor 8, na coluna “ $x_E \& x_D$ ”, indica também que $x_D = 8$, pode ser homólogo a $x_{E1} = 8$ até $x_{E2} = 14$ sendo que a Injunção de Ordem não é violada. Estes possíveis homólogos são representados em cinza escuro no quadro 5.1. O elemento (8,8) na Matriz de Prioridade está em preto devido a este poder ser homólogo a ambos $x_E = 8$ e $x_D = 8$.

Tabela					Matriz de Prioridade																	
$x_E \& x_D$	x_{D1}	x_{D2}	x_{E1}	x_{E2}		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
0	-1	-1	-1	-1		0	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	0
1	1	1	1	14		1	x		x	x	x	x	x	x	x	x	x	x	x	x	x	1
2	1	2	2	14		2	x			x	x	x	x	x	x	x	x	x	x	x	x	2
3	1	3	3	14		3	x				x	x	x	x	x	x	x	x	x	x	x	3
4	1	4	4	14		4	x					x	x	x	x	x	x	x	x	x	x	4
5	1	5	5	14		5	x						x	x	x	x	x	x	x	x	x	5
6	1	6	6	14		6	x							x	x	x	x	x	x	x	x	6
7	1	7	7	14		7	x								x	x	x	x	x	x	x	7
8	1	8	8	14		8	x									x	x	x	x	x	x	8
9	1	9	9	14		9	x										x	x	x	x	x	9
10	1	10	10	14		10	x											x	x	x	x	10
11	1	11	11	14		11	x												x	x	x	11
12	1	12	12	14		12	x													x	x	12
13	1	13	13	14		13	x														x	13
14	1	14	14	14		14	x															14
15	-1	-1	-1	-1		15	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	15
$x_E \& x_D$	x_{D1}	x_{D2}	x_{E1}	x_{E2}			0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Quadro 5.1 – Exemplo de tabela para imposição da Injunção de Ordem.

O quadro 5.2 mostra o efeito da imposição da Injunção de Ordem para o par com coordenada $(x_E, x_D) = (8, 4)$, representado na Matriz de Prioridade pelo símbolo 0.

Tabela					Matriz de Prioridade																		
x_E & x_D	x_{D1}	x_{D2}	x_{E1}	x_{E2}		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
0	-1	-1	-1	-1	0	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	0
1	1	1	1	7	1	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	1
2	1	2	2	7	2	x			x	x	x	x	x	x	x	x	x	x	x	x	x	x	2
3	1	3	3	7	3	x				x	x	x	x	x	x	x	x	x	x	x	x	x	3
4	1	3	8	8	4	x				x	x	x	x	x	x	x	x	x	x	x	x	x	4
5	1	3	9	14	5	x				x	x	x	x	x	x	x	x	x	x	x	x	x	5
6	1	3	9	14	6	x				x	x	x	x	x	x	x	x	x	x	x	x	x	6
7	1	3	9	14	7	x				x	x	x	x	x	x	x	x	x	x	x	x	x	7
8	4	4	9	14	8	x	x	x	x	O	x	x	x	x	x	x	x	x	x	x	x	x	8
9	5	9	9	14	9	x	x	x	x	x						x	x	x	x	x	x	x	9
10	5	10	10	14	10	x	x	x	x	x							x	x	x	x	x	x	10
11	5	11	11	14	11	x	x	x	x	x								x	x	x	x	x	11
12	5	12	12	14	12	x	x	x	x	x									x	x	x	x	12
13	5	13	13	14	13	x	x	x	x	x										x	x	x	13
14	5	14	14	14	14	x	x	x	x	x												x	14
15	-1	-1	-1	-1	15	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	15
x_E & x_D	x_{D1}	x_{D2}	x_{E1}	x_{E2}		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		

Quadro 5.2 – Efeito da imposição da Injunção de Ordem sobre (8,4) na tabela e na Matriz de Prioridade.

Na tabela do quadro 5.2 pode-se perceber que na linha 8, as colunas $x_{D1} = x_{D2} = 4$ indicam que $x_E = 8$ só pode ser homólogo a $x_D = 4$, da mesma forma que na linha 4, as colunas $x_{E1} = x_{E2} = 8$ indicam que $x_D = 4$ só pode ser homólogo a $x_E = 8$.

A maioria das demais linhas da tabela também sofreu alterações restringindo os intervalos de homólogos possíveis.

A utilização desta tabela diminui drasticamente o custo computacional para impor a Injunção de Ordem, uma vez que apenas os intervalos dos *pixels* homólogos possíveis são alterados em uma tabela e não na própria Matriz de Prioridade.

5.3 Algoritmo

Esta seção apresenta o algoritmo da abordagem proposta acompanhado de Figuras explicativas (quando necessário), os fluxogramas e alguns detalhes empregados na implementação computacional. As palavras sublinhadas no algoritmo são os nomes dos principais procedimentos implementados na versão computacional. A seguinte simbologia é utilizada:

- n → número do nível hierárquico;
- N → número de níveis hierárquicos;

$$S_i^n \rightarrow S_i^1 + 2(n-1)$$

$$S_j^n \rightarrow S_j^1 + 2(n-1)$$

$h \rightarrow$ altura das imagens estereoscópicas;

$w \rightarrow$ largura das imagens estereoscópicas;

$y \rightarrow$ número do par de linhas epipolares;

$$b \rightarrow \frac{S_i^N - 1}{2};$$

$x_E \rightarrow$ número da coluna na imagem da esquerda de um *pixel* no intervalo $[b-1, w-b-1]$;

$x_D \rightarrow$ número da coluna na imagem da direita de um *pixel* no intervalo $[b-1, w-b-1]$;

$x_E^\Psi \rightarrow$ número da coluna na imagem da esquerda de um *pixel* contido na região de pares impossíveis determinada por (x_E, x_D) ;

$x_D^\Psi \rightarrow$ número da coluna na imagem da direita de um *pixel* contido na região de pares impossíveis determinada por (x_E, x_D) ;

$p_E(\cdot, \cdot)$ e $p_D(\cdot, \cdot) \rightarrow$ valores de intensidade dos *pixels* na imagem da esquerda e da direita, respectivamente.

Ω e $\Psi \rightarrow$ conjunto de pares (x_E, x_D) ;

$SDA[\cdot, \cdot, \cdot, \cdot] \rightarrow$ matriz com medidas de correlação;

$P[\cdot, \cdot] \rightarrow$ matriz com valores de prioridades.

A Figura 5.1 mostra o fluxograma para o Algoritmo Principal (*Main*).

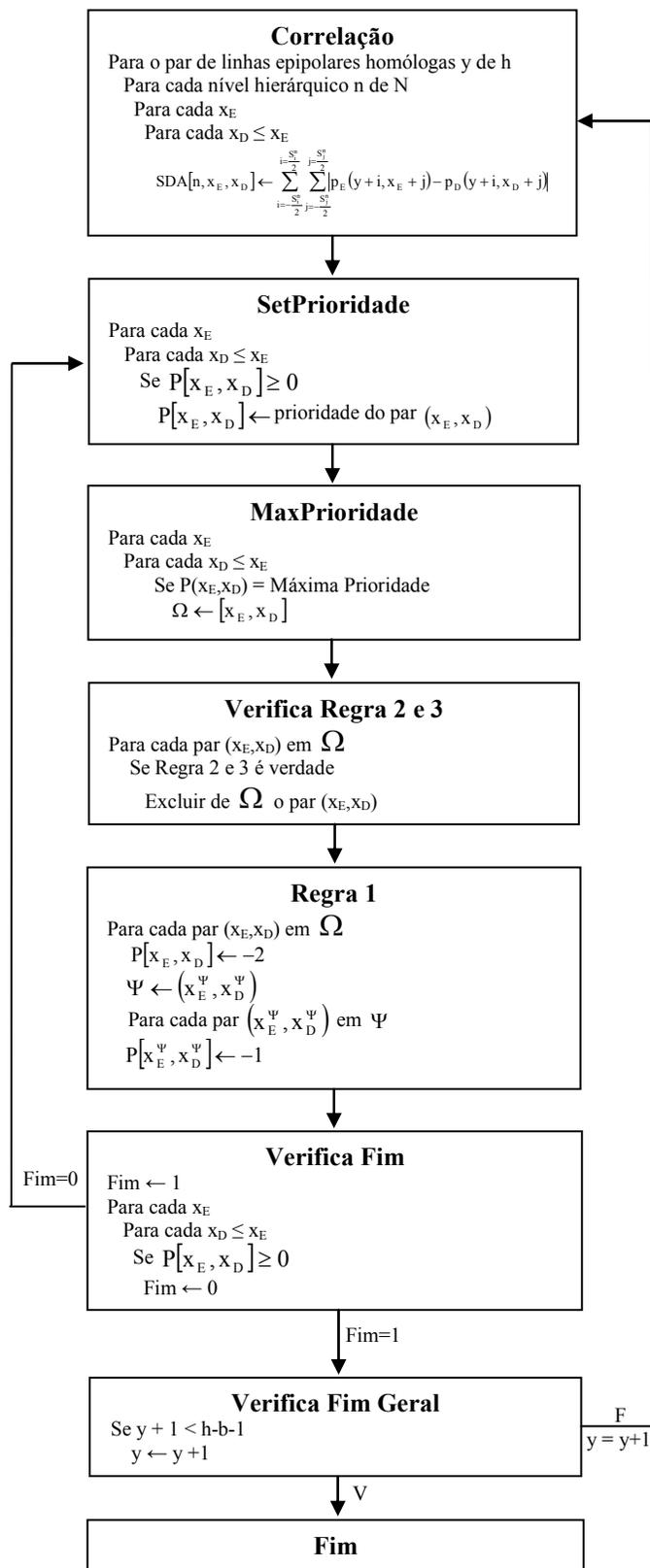


Fig. 5.1 – Fluxograma para *Main*.

A Figura 5.2 mostra o fluxograma para Setprioridade.

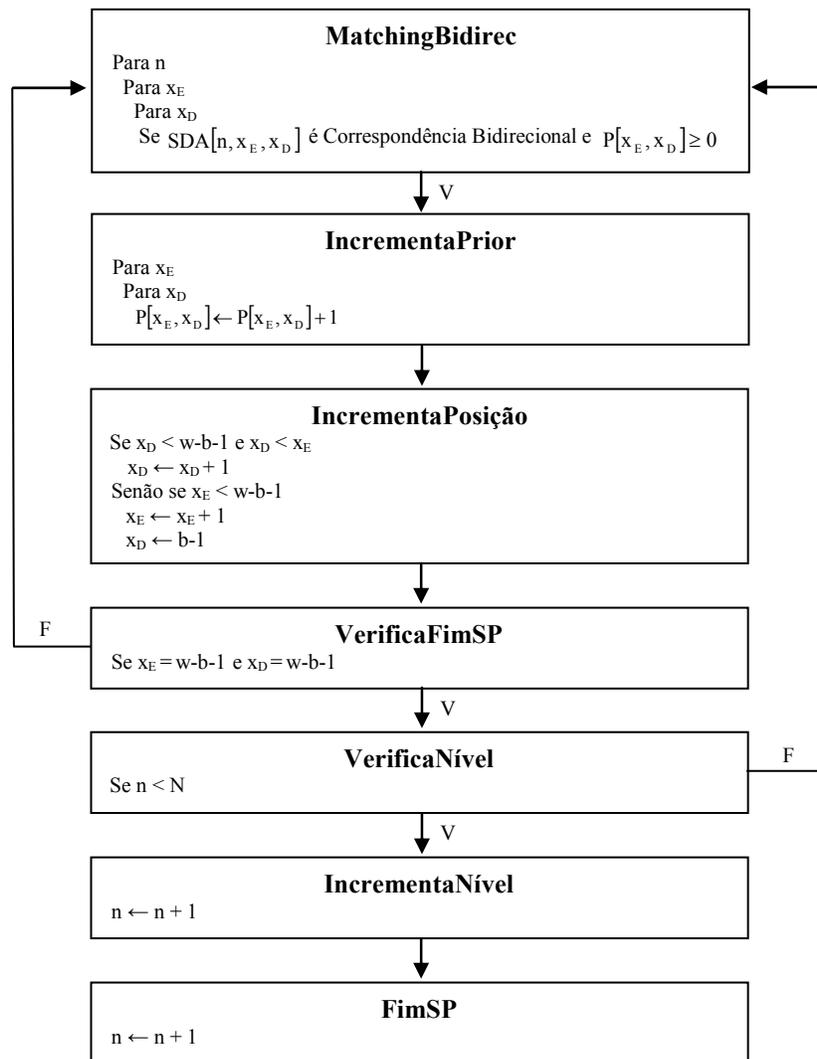


Fig. 5.2 – Fluxograma para SetPrioridade.

5.4 Exemplo

A Figura 5.3 mostra um exemplo de uma Matriz de Correspondências computada para um par de imagens estereoscópicas com $w = 80$, $N = 1$ e $s_i^j x s_j^i = 1 \times 1$. A escala de cinza utilizada é logarítmica de base 2. Os níveis de cinza mais escuros representam as melhores medidas de correspondências.

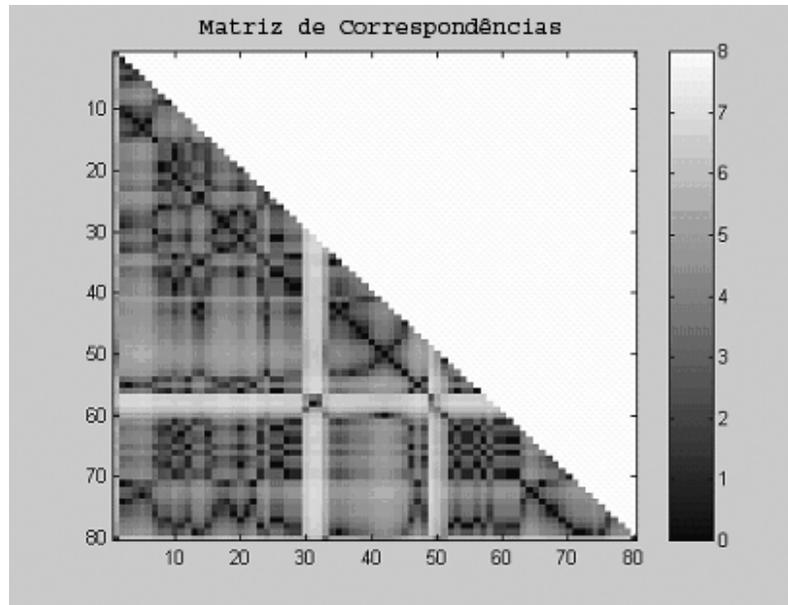


Fig. 5.3 – Exemplo de Matriz de Correspondências para nível hierárquico $n = 1$ para imagem real.

A Figura 5.4 mostra a Matriz de Prioridades em que a grande área cinza médio (codificada com valor -1) representa os pares impossíveis logo após o término da primeira iteração. As regiões brancas (codificadas com valor 0) representam os pontos ainda ambíguos e os pontos pretos (codificadas com valor -2) representam os pares assumidos como correspondentes pelo algoritmo. Pode-se notar nesta Figura que a imposição da Injunção de Ordem para alguns pares classificados como correspondentes reduziu drasticamente a quantidade de pontos candidatos a homólogos.

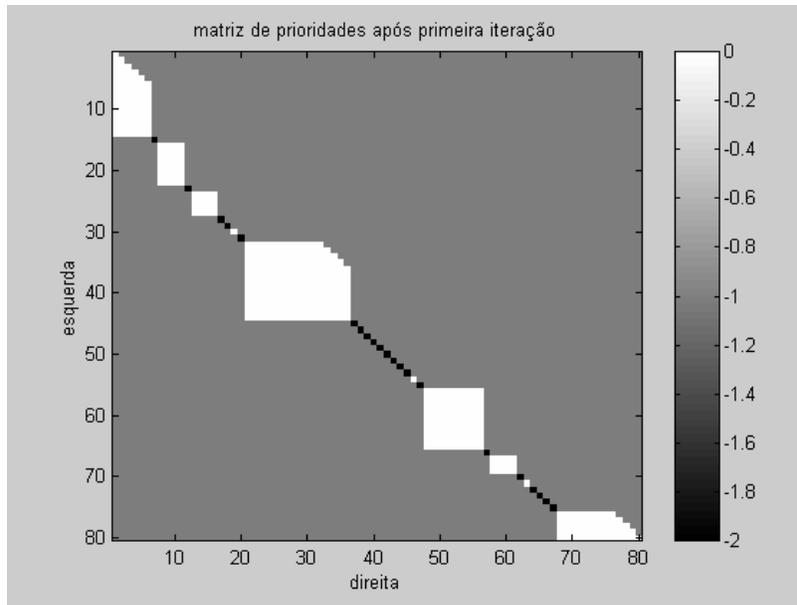


Fig. 5.4 – Exemplo de Matriz de Prioridade real.

A Figura 5.5 mostra a Matriz de Correspondências após a primeira iteração.

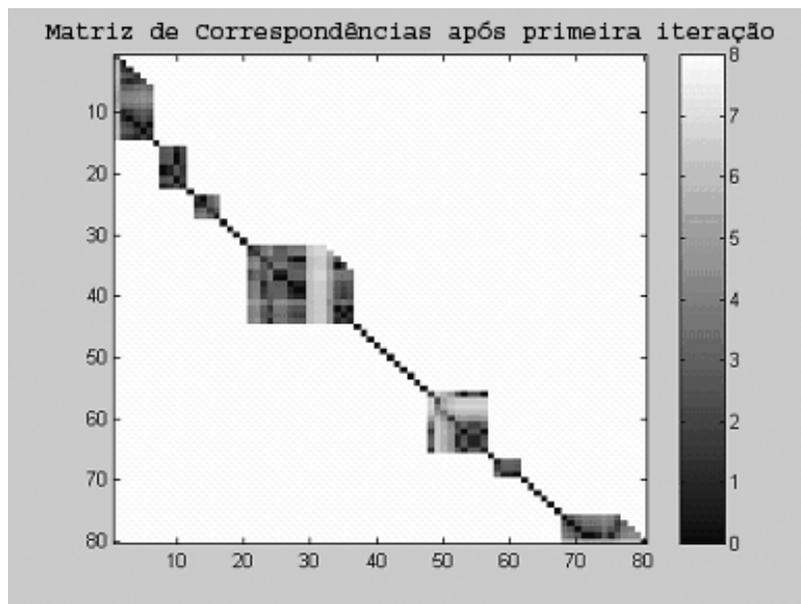


Fig. 5.5 – Matriz de Correspondências da Figura 5.3 após primeira iteração.

Cabe ressaltar que os mapas de disparidade associados às imagens da esquerda e da direita possuem seus valores inicialmente todos designados para -1, representando oclusões. A resolução completa da Matriz de Correspondências da figura 5.5 irá designar os *pixels*

homólogos de apenas uma linha dos mapas de disparidades associados às imagens da esquerda e da direita. Para as demais linhas dos mapas de disparidade o procedimento é análogo. As regiões de oclusão são identificadas pelos pontos dos mapas de disparidades que não receberam valores de disparidade e que, portanto, após todo o processo continuaram com valores -1.

5.5 Implementação

O programa computacional com a abordagem proposta foi implementado em linguagem C++ Builder 5.0. A Figura 5.6 abaixo mostra a interface do programa.

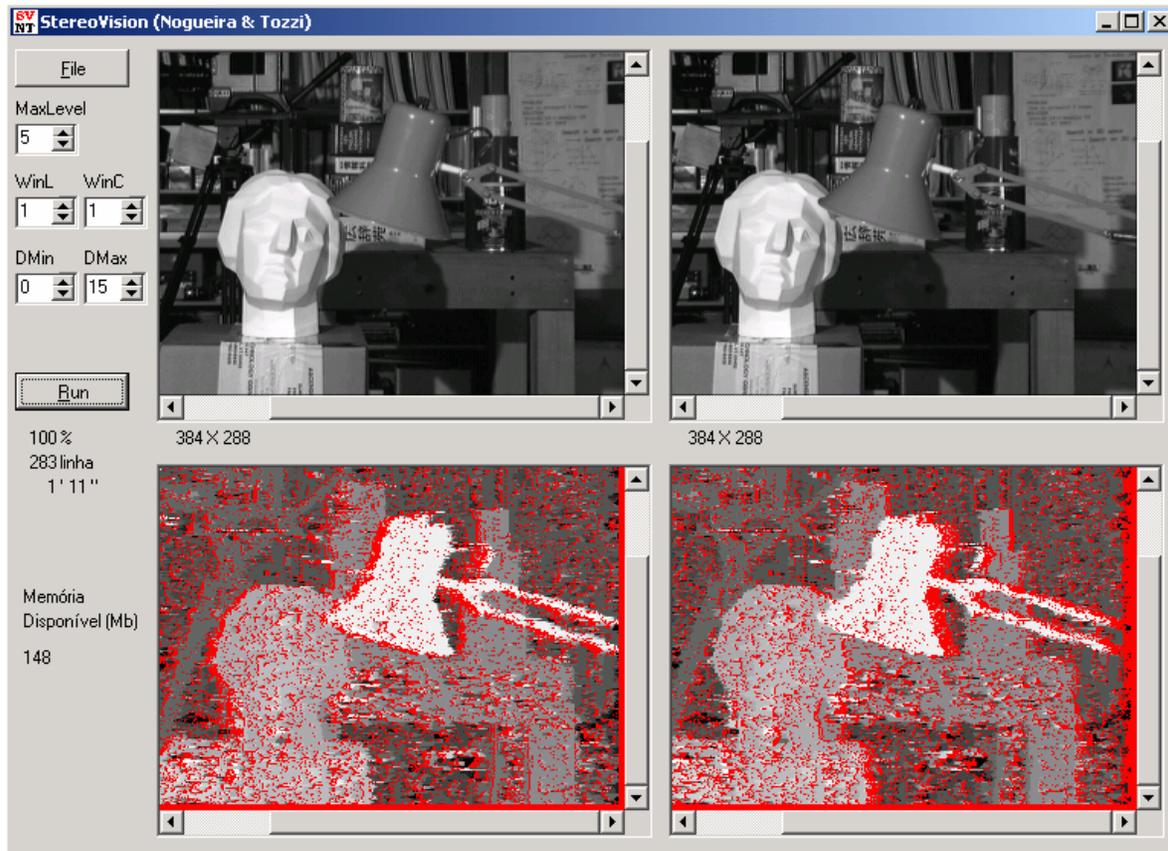


Fig. 5.6 – Interface do programa computacional implementado.

Esta linguagem de programação foi adotada devido ao autor deste trabalho já possuir experiência em linguagem C além das facilidades disponíveis para a construção de interface gráfica no compilador Builder 5.0.

O programa permite que o usuário selecione um par de imagens estereoscópicas no formato BMP (*Bitmap*). As imagens podem ser do tipo *grayscale* (8 *bits* sem *palette* de cores), indexada (8 *bits* com *palette* de cores) ou *true color* (24 *bits*). As imagens *true color* ou indexada são transformadas do modelo de cor RGB para o modelo de cor YUV, sendo então utilizado apenas o canal Y (também em 8 *bits*) do modelo YUV para a realização da correspondência baseada em áreas. Com isso a correlação é realizada para valores dos *pixels* no intervalo [0, 255]. A transformação utilizada do modelo de cor RGB para o modelo de cor YUV é descrita em detalhes em Foley et. al., 1982.

Ainda na interface gráfica, o usuário pode selecionar o número máximo de níveis (*MaxLevel*) a serem realizados no processo hierárquico, as dimensões nas direções y (*WinL*) e x (*WinC*) iniciais da janela de correlação, bem como o valor de disparidade mínima (*DMin*) e disparidade máxima (*Dmax*).

Os valores de *Dmin* e *Dmax* são determinados pelo programa com base no tamanho das imagens estereoscópicas. Estes valores podem ser alterados pelo usuário. O valor *Dmin* estabelecido pelo programa é 0 (zero) e o valor *Dmax* é:

$$D_{max} = W - 2N - S_j^1 + 2 \quad (5.7)$$

onde:

W é largura (comprimento) das imagens;

N é o nível máximo do processo hierárquico; e

S_j^1 é o tamanho inicial da janela de correlação na direção j.

A limitação de *Dmax* de acordo com a expressão 5.7 é necessária dada a impossibilidade de determinação da correlação para os *pixels* situados na borda lateral e vertical das imagens em função do tamanho da janela de correlação.

A Figura 5.7 mostra um exemplo para uma janela de correlação 5 x 5, resultando em uma borda vertical e lateral (em cinza) com 2 *pixels* de altura e largura, respectivamente.

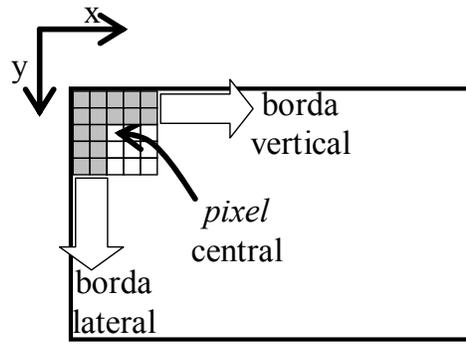


Fig. 5.7 – *Pixel* central e bordas vertical e lateral para janela de correlação 5 x 5.

Capítulo 6

6. Testes e Resultados

As imagens utilizadas nos testes realizados são retificadas e não contém objetos transparentes. No entanto, muitas delas apresentam regiões com texturas homogêneas e variações abruptas (não suave) dos valores de disparidade, o que ocasionam grandes regiões de oclusão, dificultando substancialmente a resolução do Problema de Correspondência.

Com o intuito de quantificar os resultados obtidos, utilizou-se um conjunto de imagens estereoscópicas disponíveis no *site* <http://www.middlebury.edu/stereo>. Este conjunto contém, além de imagens estereoscópicas reais (Map e Tsukuba) e sintéticas (Sawtooth e Venus), os respectivos mapas de disparidade verdadeiros, mapas das regiões de descontinuidades, mapas de regiões de oclusão e mapas de regiões com texturas homogêneas, sendo todos estes associados às imagens da esquerda.

Um outro conjunto de imagens estereoscópicas foi também utilizado para testar a abordagem proposta neste trabalho. No entanto, as imagens reais deste conjunto não possuem mapas de disparidade associados, o que impossibilita quantificar os erros. Estes resultados são avaliados apenas por inspeção visual. As imagens sintéticas deste conjunto possuem mapas de disparidade associados e a quantificação dos erros foi realizada.

6.1 Metodologia de Testes

Com o intuito de quantificar os resultados obtidos e facilitar a comparação dos resultados com outras abordagens, faz-se necessário adotar alguma métrica. Neste trabalho utiliza-se a mesma métrica elaborada por Scharstein & Szeliski, 2002, embora se discorde de algumas considerações feitas por estes.

Um problema que ocorre com os mapas de disparidade verdadeiros disponibilizados no citado *site* é que as disparidades são determinadas a partir da estrutura tridimensional da

cena e não a partir das correspondências dos *pixels* homólogos (Lin & Tomasi, 2002). Esta técnica resulta em valores de disparidades mesmo para regiões de oclusão, o que não se verifica quando as disparidades são obtidas a partir das correspondências dos *pixels* homólogos.

Para Scharstein & Szeliski (2002), um par é considerado errado se o módulo da diferença entre o valor de disparidade calculado e o valor de disparidade obtido do mapa de disparidade verdadeiro é maior que 1.

O resultado total é contabilizado como a porcentagem de pares errados, em relação ao total de *pixels* nas imagens estereoscópicas descontando uma “borda” nas regiões limítrofes das imagens estereoscópicas. O tamanho da “borda” é fornecido para cada par de imagens estereoscópicas e apresentado na seqüência do texto.

O *site* citado além de disponibilizar o conjunto de imagens estereoscópicas utilizadas no item 6.2, fornece também uma tabela com os resultados obtidos por várias abordagens. Estes resultados foram enviados por seus respectivos autores e tem sido intensivamente utilizado nos artigos da área como principal material para comparação de resultados entre metodologias.

Esta tabela contém os erros que cada metodologia cometeu para os quatro pares de imagens estereoscópicas (Map, Sawtooth, Tsukuba e Venus) utilizadas na seção 6.2 em três categorias que são: *all*, *textureless* (*untex*) e *discontinuity* (*disc*). O seguinte parágrafo foi extraído e traduzido do *site* citado:

Os números representam a porcentagem de "bad" pixels, ou seja, pixels cujo módulo do erro da disparidade associada é maior que um. Para cada par de imagem nós reportamos porcentagem para todos os pixels (all), pixels em áreas sem texturas (untex) e pixels próximos as descontinuidades de disparidade (disc). Somente pixels não oclusos são considerados em todos os três casos, e nós ignoramos uma borda de dez pixels (dezoito para Tsukuba) quando computamos estas estatísticas.

É importante ressaltar a opção em não computar os erros para as regiões de oclusão dificulta a realização de comparações dos resultados obtidos neste trabalho com os demais, uma vez que a metodologia proposta está fundamentada em respeitar e identificar as regiões de oclusão.

De maneira prática, os valores de disparidade determinados erroneamente para as regiões de oclusão não são contabilizados para a determinação da porcentagem citada. Assim, para as regiões de oclusão qualquer valor de disparidade pode ser determinado, independente da magnitude do erro, porque estes *pixels* não são computados para a porcentagem total citada.

Scharstein & Szeliski (2002), justificam a exclusão das regiões de oclusão do cálculo da porcentagem total de pares errados, como:

Nós excluímos as regiões de oclusão do cálculo da porcentagem já citada porque poucos algoritmos modelam oclusões, sendo que, aqueles que consideram fornecem resultados bastante ruins ainda.

O artigo de Scharstein & Szeliski (2002), que aborda detalhadamente a metodologia de avaliação e computação das estatísticas citada acima, sugere extrapolar ou interpolar valores de disparidade para as regiões de oclusão para as metodologias que identificam regiões de oclusão. O método para extrapolação ou interpolação não é definido.

Nesta tese são apresentados os resultados com valores de disparidade para as regiões de oclusão, aqui designados por "mapa de disparidade com oclusões preenchidas", como citado na seção 6.2, sendo estes mapas os utilizados para computar as estatísticas citadas. As regiões de oclusão são preenchidas por extrapolação do valor de disparidade vizinho.

6.2 Resultados Quantitativos

Uma vez que os mapas de disparidades verdadeiros originais não apresentam regiões de oclusão, foram gerados novos mapas de disparidades verdadeiros com oclusão, apenas para fins de comparação visual com os mapas de disparidades reconstruídos. A geração destes mapas foi realizada impondo-se a Injunção de Ordem nos mapas de disparidades verdadeiros originais. As regiões de oclusão foram rotuladas com valores -1 e estão representadas em vermelho. As regiões em níveis de cinza codificam os valores de

disparidade, onde as regiões claras representam altos valores de disparidade e as regiões escuras representam baixos valores de disparidade.

Os mapas de disparidade calculados são também apresentados com as regiões de oclusão “preenchidas” com valores de disparidade dos pares mais próximos à esquerda das regiões de oclusão. Este método de interpolação é adequado para superfícies plano frontais, pois propaga os menores valores de disparidade (superfícies mais profundas) de maneira constante. Estes mapas são utilizados para comparação com os mapas de disparidades verdadeiros.

O programa implementado fornece também um “mapa de confiança”, o qual contém o nível em que a Injunção de Ordem foi imposta para cada posição. Assim, nas Figuras, as regiões mais claras deste mapa representam maiores níveis enquanto as regiões mais escuras representam menores níveis. Como é de se esperar, as regiões de oclusão estão representadas por preto, uma vez que a Injunção de Ordem não foi imposta para os pontos contidos nestas regiões.

Todos os testes nesta seção foram realizados com 5 níveis hierárquicos, porém para 7 valores distintos do tamanho inicial da janela de correlação (exemplificando, se $s_i \times s_j = 3 \times 3$, tem-se $S_i \times S_j = 11 \times 11$).

As tabelas 1 (Map), 2 (Sawtooth), 3 (Tsukuba) e 4 (Venus) apresentam resultados quantitativos das reconstruções em função dos valores do tamanho inicial da janela de correlação, onde E é a porcentagem de pares, em relação ao total de pares, cujos valores dos mapas de disparidades reconstruídos diferem, em módulo, de valores maiores que 1 dos respectivos valores dos mapas de disparidades verdadeiros; ET é análogo à E, porém calculado apenas para as regiões classificadas como “sem texturas”, ED é análogo à E, porém calculado apenas para as regiões classificadas como “descontinuidade”, EO é análogo à E, porém calculado apenas para as regiões de oclusão e T é o tempo de processamento em segundos utilizado apenas como referência para estimativa de esforço computacional decorrente do aumento do tamanho inicial da janela e determinados para o programa desenvolvido e executado em um processador Pentium 4, 2.8GHz e 512 Mb RAM. Ainda para estas tabelas, os valores em negrito representam os melhores resultados.

Os mapas de pares errados ilustram a ocorrência espacial destes pares.

As Figuras 6.1, 6.4, 6.7 e 6.10 mostram, da esquerda para direita e de cima para baixo, a imagem estereoscópica esquerda, imagem estereoscópica direita, mapa de disparidade verdadeiro, mapa de oclusão (oclusão em preto), mapa de descontinuidade (descontinuidade em preto) e mapa de regiões sem texturas (regiões sem textura em preto).

As imagens na Figuras 6.2, 6.5, 6.8 e 6.11 mostram os resultados obtidos em função do tamanho da janela inicial. Cada linha desta Figura mostra da esquerda para a direita, o mapa de disparidade com oclusões em vermelho, o mapa de disparidade “preenchidoa” as regiões de oclusão e o mapa de erros entre o mapa de disparidade calculado e o mapa de disparidade verdadeiro, onde a cor branca representa os *pixels* cuja disparidade associada difere, em módulo, por mais de 1 unidade do valor verdadeiro.

As Figuras 6.3, 6.6, 6.9 e 6.12 mostram dois gráficos tridimensionais (no da esquerda as cores codificam os valores de disparidade, no da direita a informação tridimensional é apresentada de maneira texturizada) das disparidades para o teste utilizando tamanho de janela $S_x \times S_y$ que forneceu o melhor resultado quando comparado com outros tamanhos de janela para o erro total E.

6.2.1 Imagens Map

Este par de imagens estereoscópicas apresentado na Figura 6.1 mostra uma cena bastante simples, composta por 2 planos bem distintos e altamente texturizados. Os dois planos estão separados por uma grande diferença de disparidade, resultando em uma significativa região de oclusão. A abordagem proposta corretamente identifica a grande região de oclusão.

Uma vez que a textura está presente em praticamente todas as regiões das imagens e a estrutura da cena é simples, é relativamente fácil determinar disparidades corretas para os *pixels* das imagens que não pertencem as regiões de oclusão. Entretanto, a grande região de oclusão ocasiona uma grande quantidade de erros para as abordagens que não consideram a existência de oclusões. Este fato é observado em muitas abordagens disponíveis no *site* <http://www.middlebury.edu/stereo>, porém, como já citado, os valores de disparidades associados aos *pixels* que pertencem às regiões de oclusão (verdadeira) não são

considerados para a porcentagem de erros, fazendo com que estas obtenham boas classificações no referido *site*.

As imagens estereoscópicas possuem 216 linhas x 284 colunas. A região de borda possui 10 *pixels* e a disparidade varia de 0 a 29.

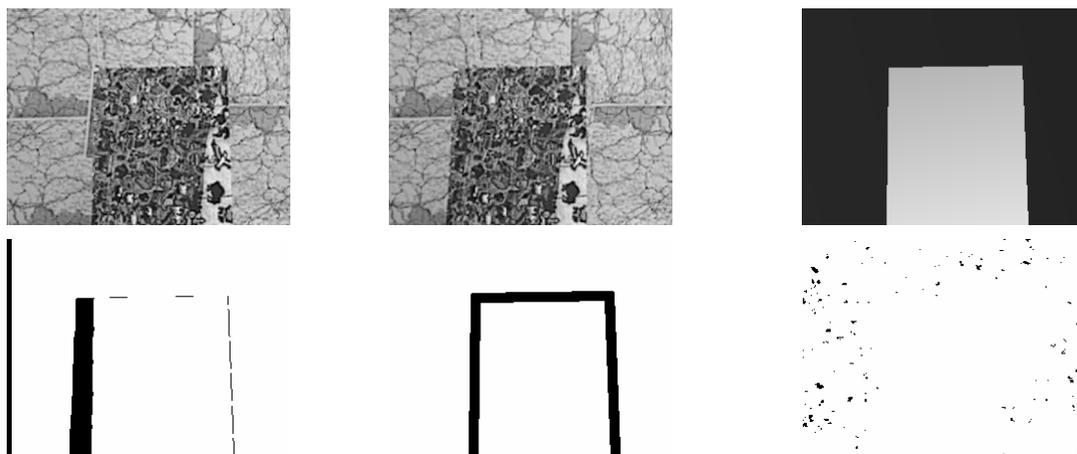
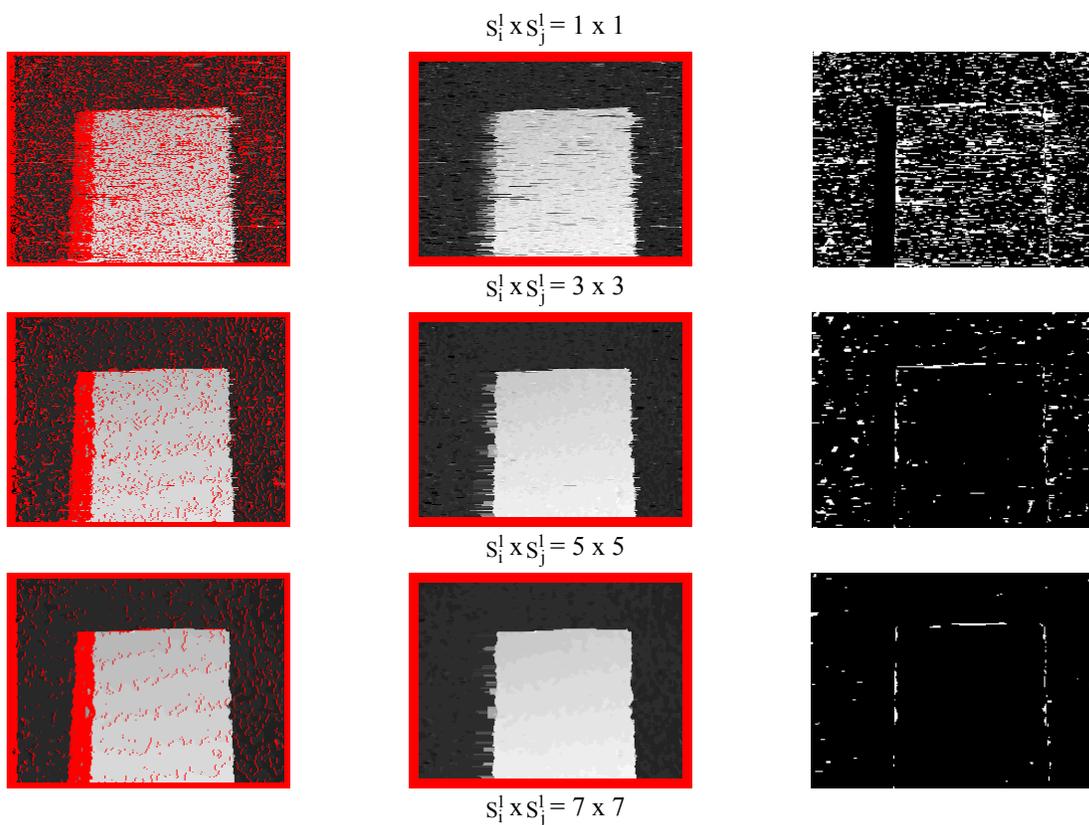


Fig. 6.1 – Conjunto de imagens Map: Imagens Estereoscópicas, Mapa de Disparidade Verdadeiro, Mapa de Oclusões, Mapa de Descontinuidade e Mapa de regiões sem Texturas.



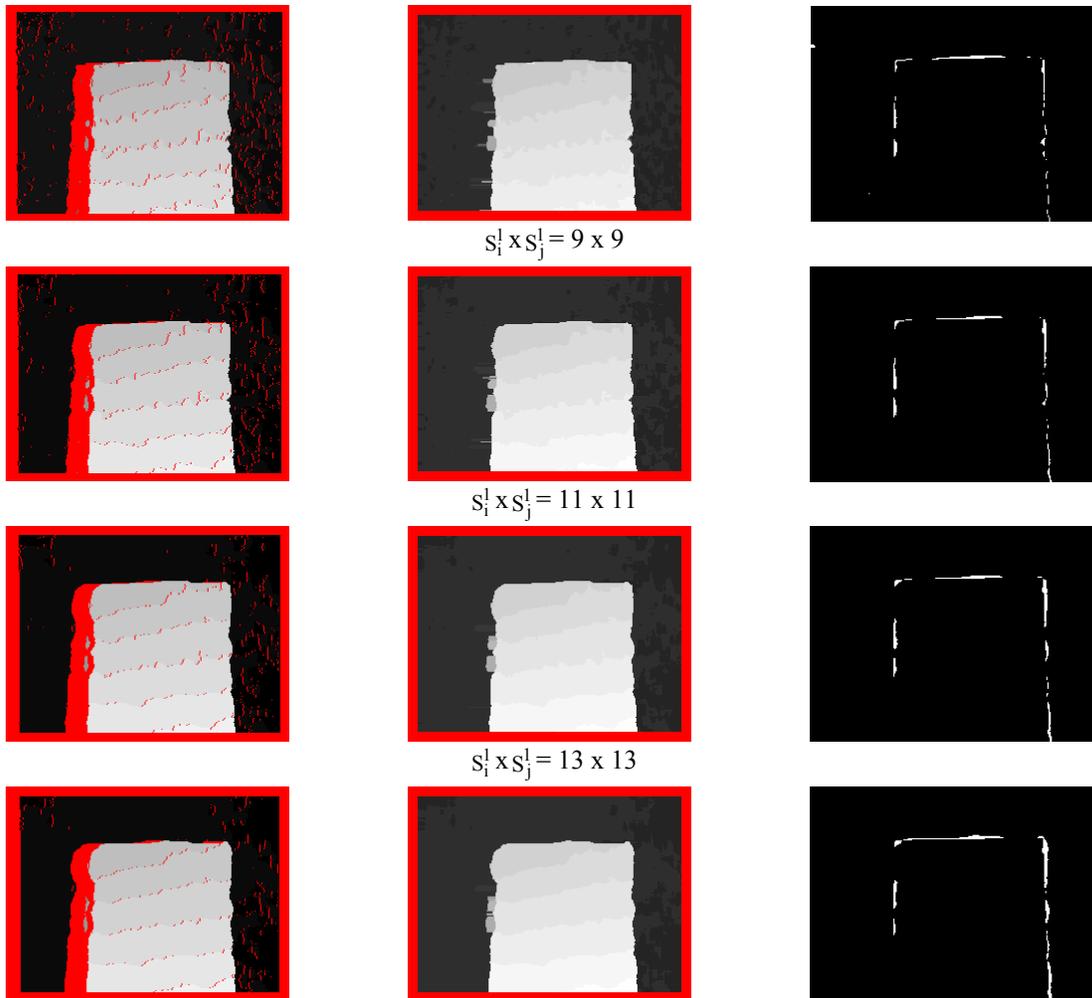


Fig. 6.2 – Resultados para imagens Map: Mapas de Disparidade reconstruídos com e sem oclusão e Mapa de pontos errados.

Tabela 1 – Resultados para imagens Map.

	Tamanho da janela inicial						
	1x1	3x3	5x5	7x7	9x9	11x11	13x13
E(%)	18.6	3.4	0.9	0.7	0.8	0.9	1.2
ET(%)	34.8	32.9	5.9	0.7	0.2	0.5	0.7
ED(%)	27.7	10.0	7.0	8.5	9.2	11.0	13.6
EO(%)	15.7	9.9	8.5	9.6	10.7	12.2	13.3
T(s)	22	6	7	8	10	12	14

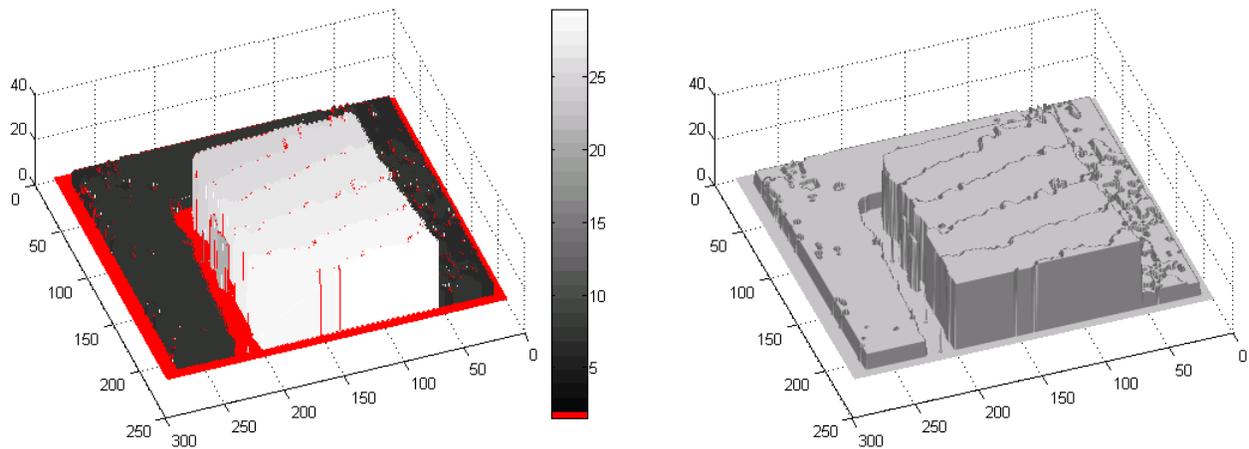


Fig. 6.3 – Gáficos 3D para o mapa de disparidade reconstruído com $S_1^1 \times S_2^1 = 7 \times 7$ - imagens Map.

6.2.2 Imagens Sawtooth

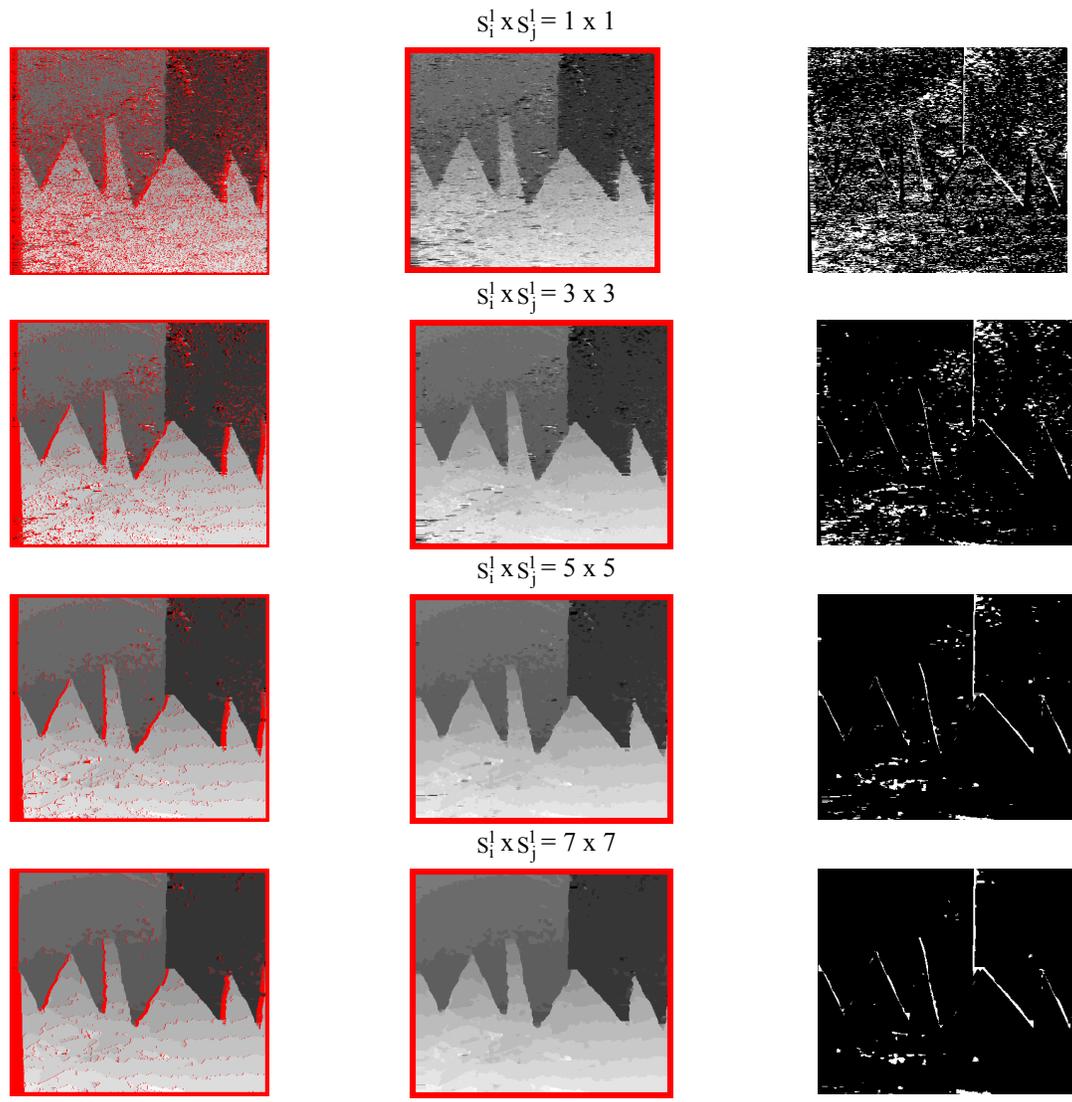
Este par de imagens estereoscópicas registra uma cena composta por 3 planos bem distintos e quantidade variante de textura. A principal dificuldade desta cena consiste na reconstrução das regiões de mesma disparidade, porém com formas "estreitas" e pontiagudas.

A abordagem proposta determina valores corretos de disparidades para a grande maioria dos *pixels* das imagens. A grande maioria dos erros ocorre nas regiões de descontinuidade da cena, principalmente entre regiões "estreitas" e com baixa textura.

As imagens estereoscópicas possuem 380 linhas x 434 colunas. A região de borda possui 10 *pixels* e a disparidade varia de 0 a 19 *pixels*.



Fig. 6.4 – Conjunto de imagens Sawtooth: Imagens Estereoscópicas, Mapa de Disparidade Verdadeiro, Mapa de Oclusões, Mapa de Descontinuidade e Mapa de regiões sem Texturas.



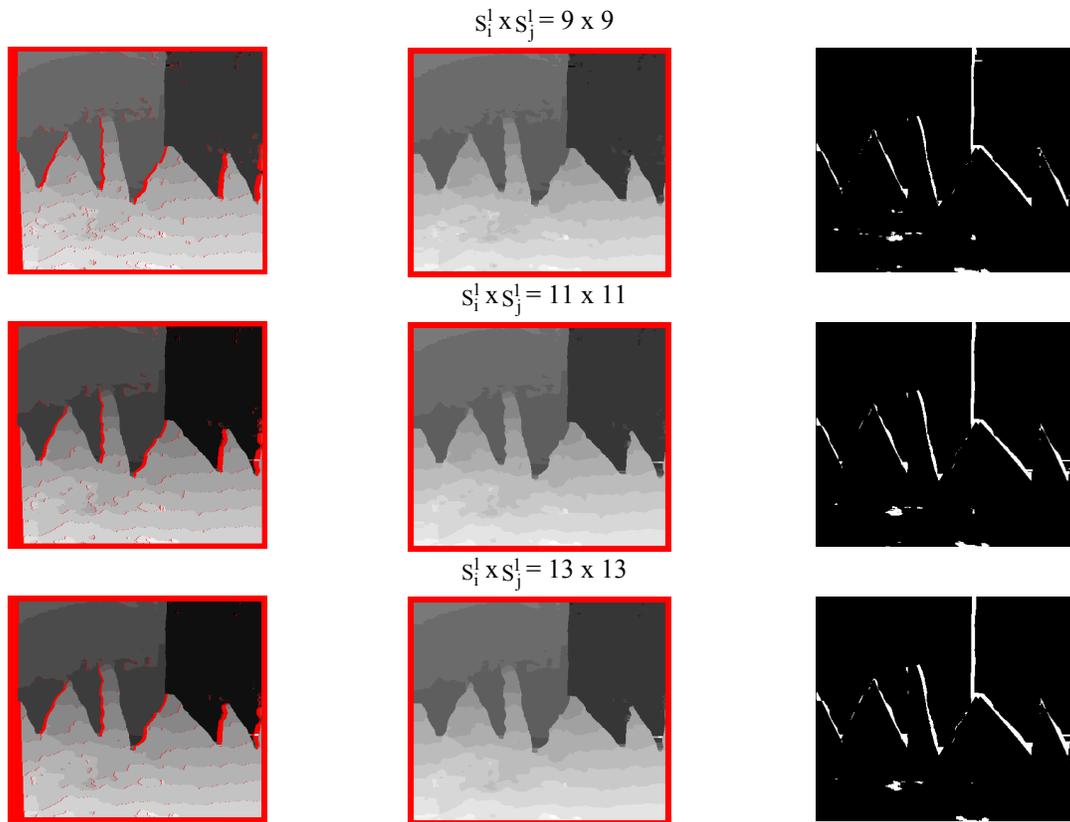


Fig. 6.5 – Resultados para imagens Sawtooth: Mapas de Disparidade reconstruídos com e sem oclusão e Mapa de pontos errados.

Tabela 2 – Resultados para imagens Sawtooth.

	Tamanho da janela inicial						
	1x1	3x3	5x5	7x7	9x9	11x11	13x13
E(%)	20.7	5.1	2.6	2.2	2.3	2.7	3.0
ET(%)	26.2	11.6	2.7	0.9	0.8	1.5	2.3
ED(%)	23.5	9.1	10.7	13.8	17.1	20.3	22.4
EO(%)	20.8	15.3	16.6	19.9	23.1	25.8	27.9
T(s)	177	17	16	18	21	24	28

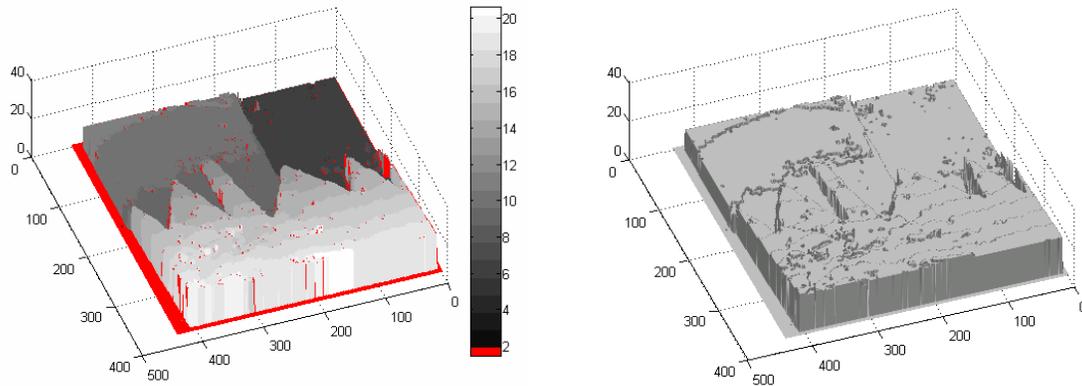


Fig. 6.6 – Gáficos 3D para o mapa de disparidade reconstruído com $S_1^1 \times S_1^1 = 7 \times 7$ - imagens Sawtooth.

6.2.3 Imagens Tsukuba

Este par de imagens estereoscópicas, cortesia de Y. Ohta e Y. Nakamura da University of Tsukuba, apresenta uma cena composta por vários planos, superfícies suavemente curvadas e objetos não suaves. Existem regiões com pouca textura e objetos “estreitos”, o que dificulta a resolução do Problema de Correspondência.

Uma grande dificuldade deste par de imagens estereoscópicas está em reconstruir a estrutura que suporta a luminária, uma vez que esta é bastante estreita. A abordagem proposta reconstrói adequadamente esta estrutura apenas para pequenos tamanhos da janela inicial da Correspondência Bidirecional Hierárquica. Para os tamanhos da janela inicial maiores esta estrutura não é adequadamente reconstruída devido ao grande suporte da janela de correlação em relação a largura dessa estrutura.

As imagens estereoscópicas possuem 288 linhas x 384 colunas. A região de borda possui 18 *pixels* e a disparidade varia de 0 a 15.

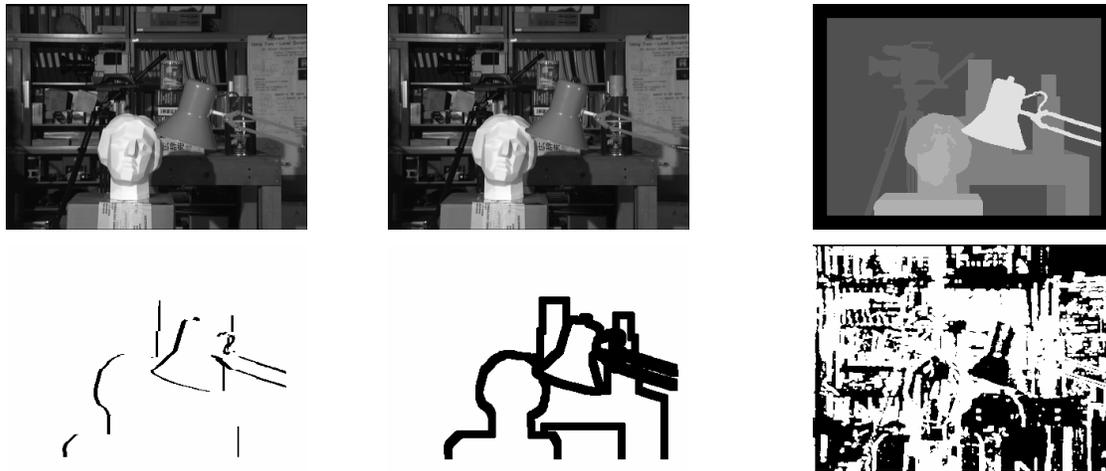
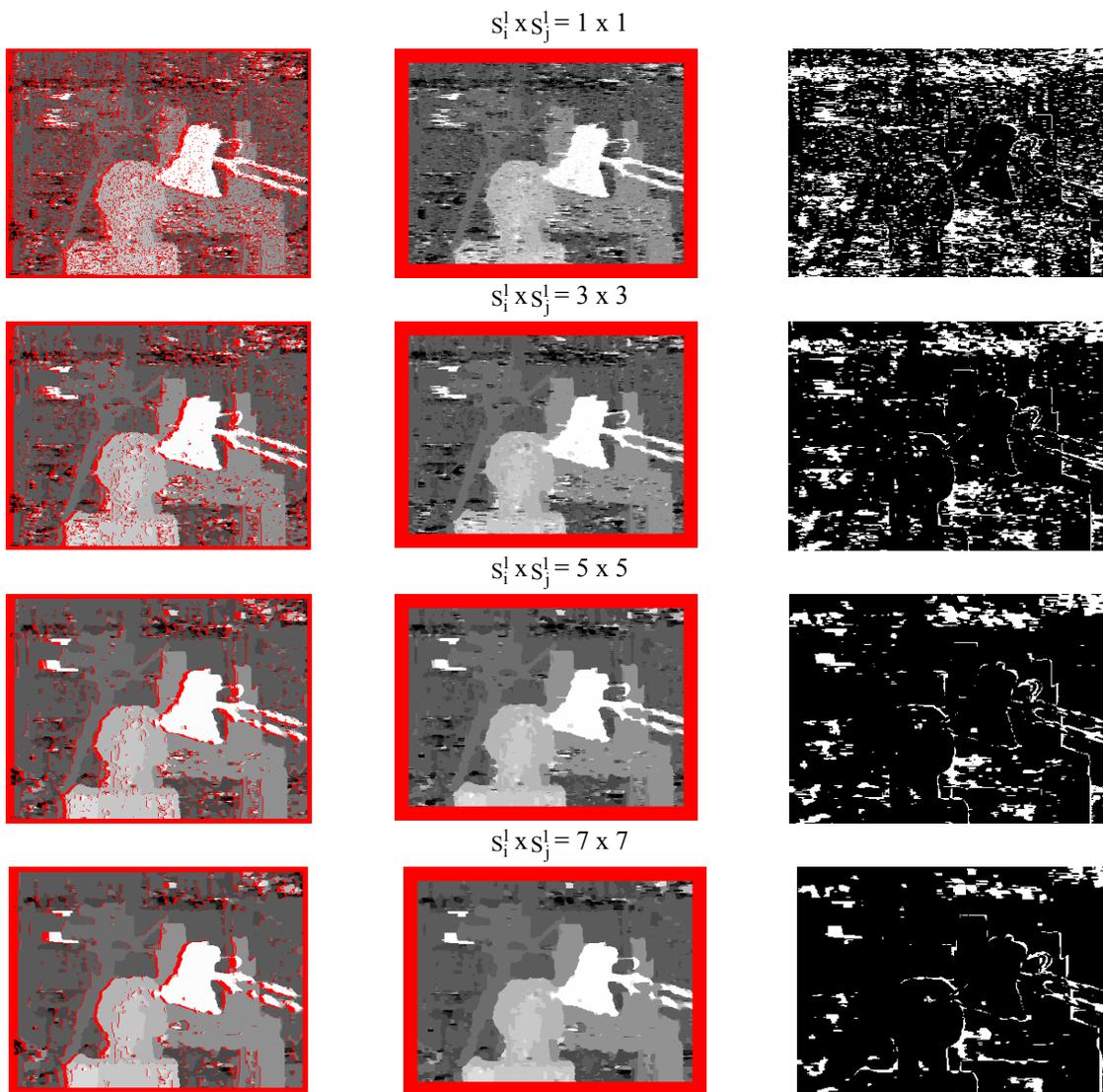


Fig. 6.7 – Conjunto de imagens Tsukuba: Imagens Estereoscópicas, Mapa de Disparidade Verdadeiro, Mapa de Oclusões, Mapa de Descontinuidade e Mapa de regiões sem Texturas.



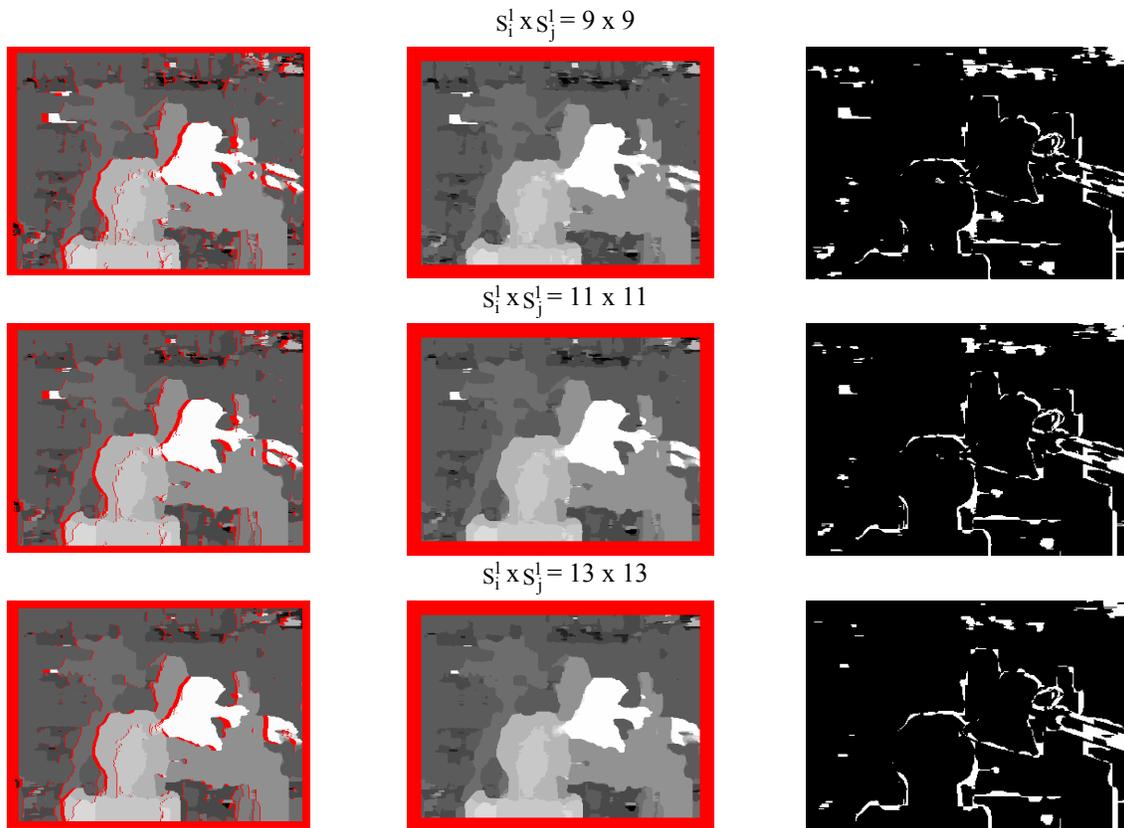


Fig. 6.8 – Resultados para imagens Tsukuba: Mapas de Disparidade reconstruídos com e sem oclusão e Mapa de pontos errados.

Tabela 3 – Resultados para imagens Tsukuba.

	Tamanho da janela inicial						
	1x1	3x3	5x5	7x7	9x9	11x11	13x13
E(%)	23.1	14.0	10.1	7.9	6.8	6.4	6.7
ET(%)	33.9	25.1	17.3	12.2	9.1	7.6	7.1
ED(%)	19.6	12.9	13.2	15.3	18.1	21.7	25.3
EO(%)	37.6	39.8	47.8	56.8	65.9	72.8	77.8
T(s)	66	14	10	11	12	13	15

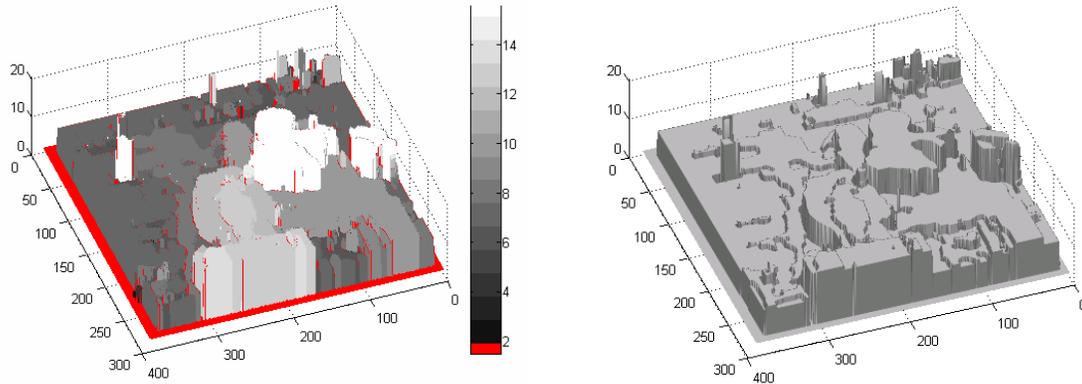


Fig. 6.9 – Gáficos 3D para o mapa de disparidade reconstruído com $S_1^1 \times S_1^1 = 11 \times 11$ - imagens Tsukuba.

6.2.4 Imagens Venus

Este par de imagens estereoscópicas apresenta uma cena composta por 4 planos bem distintos, mas com regiões com poucas texturas.

A abordagem proposta apresenta os melhores resultados para os maiores tamanhos da janela inicial devido à baixa textura das regiões das imagens, no entanto, este fato ocasiona uma má reconstrução nas regiões de descontinuidade da cena. Para os tamanhos da janela inicial menores observa-se uma grande quantidade de ruídos nos mapas de disparidades reconstruídos mesmo para as regiões suaves da cena, tais como, os planos inclinados cujas texturas são folhas de jornal. Nota-se, que existem grandes regiões nestes jornais que são homogêneas, sendo que para estas, os valores de disparidades são mal determinados.

As imagens estereoscópicas possuem 383 linhas x 434 colunas. A região de borda possui 10 *pixels* e a disparidade varia de 0 a 19.

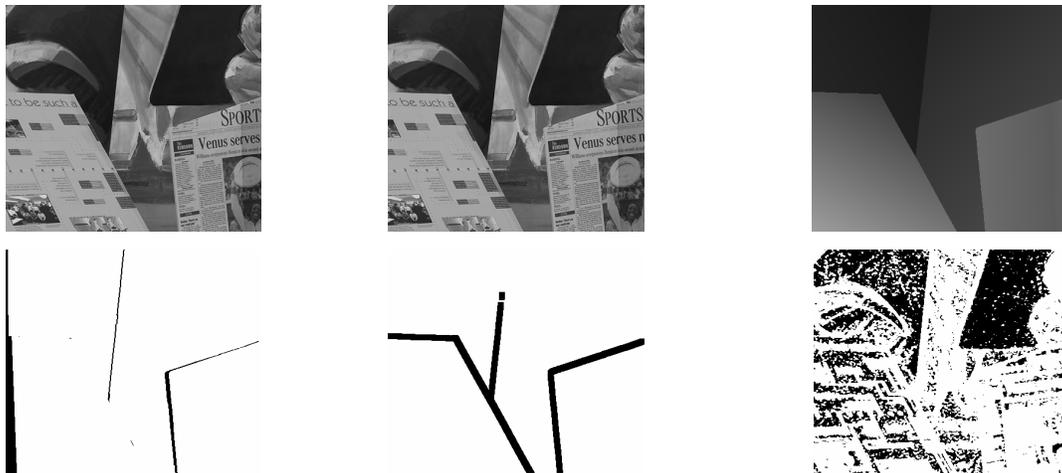
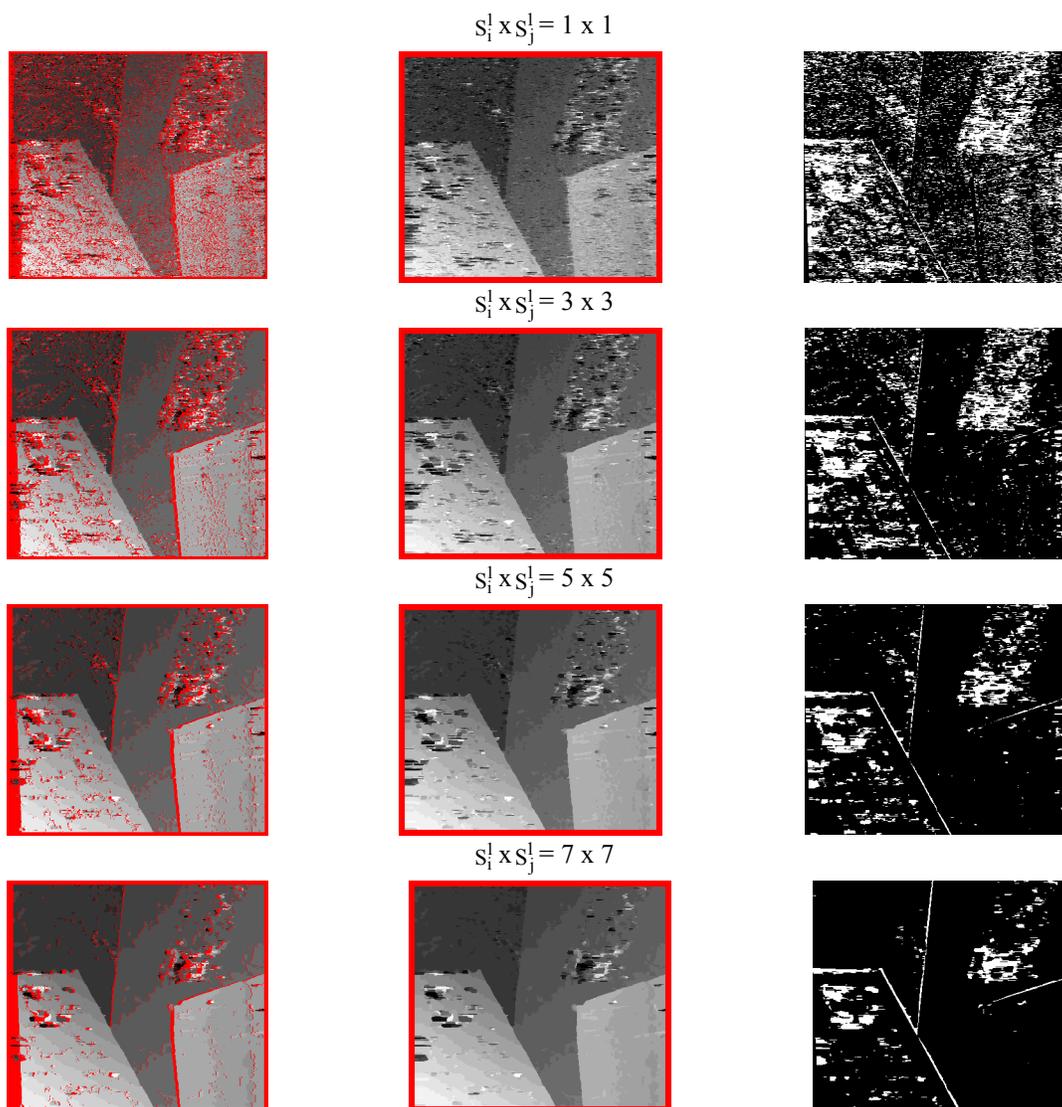


Fig. 6.10 – Conjunto de imagens Venus: Imagens Estereoscópicas, Mapa de Disparidade Verdadeiro, Mapa de Oclusões, Mapa de Descontinuidade e Mapa de regiões sem Texturas.



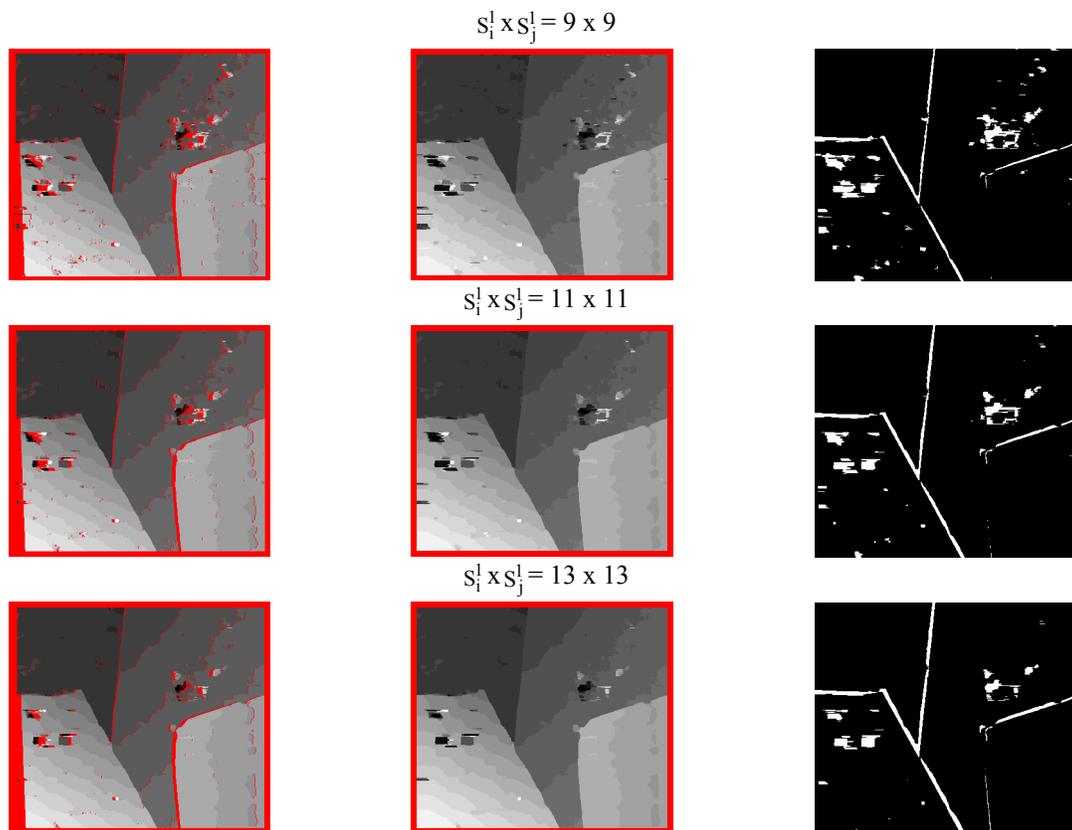


Fig. 6.11 – Resultados para imagens Venus: Mapas de Disparidade reconstruídos com e sem oclusão e Mapa de pontos errados.

Tabela 4 – Resultados para imagens Vênus.

	Tamanho da janela inicial						
	1x1	3x3	5x5	7x7	9x9	11x11	13x13
E(%)	27.1	14.6	8.8	5.9	4.7	4.2	4.0
ET(%)	40.2	30.2	18.1	11.1	7.8	6.5	5.7
ED(%)	23.5	14.5	14.9	19.4	24.5	29.1	31.8
EO(%)	23.9	21.2	22.2	23.0	24.8	25.6	25.5
T(s)	170	34	21	20	22	26	28

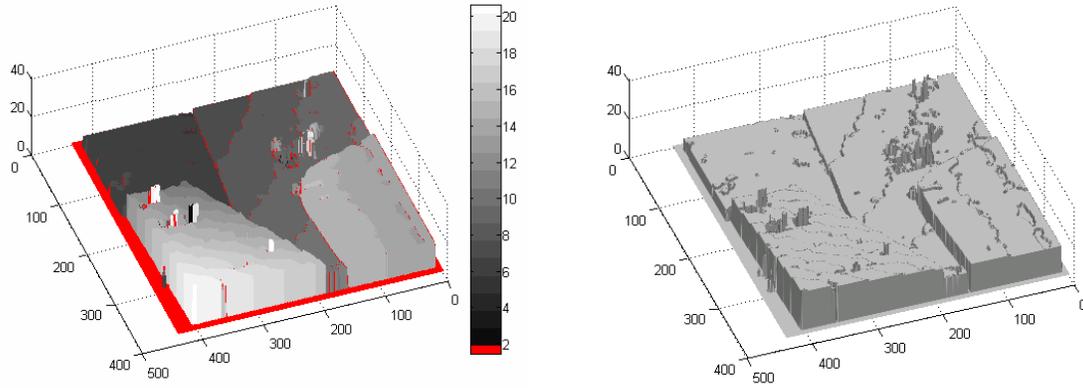


Fig. 6.12 – Gáficos 3D para o mapa de disparidade reconstruído com $S_i^1 \times S_j^1 = 13 \times 13$ - imagens Venus.

Os gráficos das Figuras 6.13, 6.14, 6.15 e 6.16 mostram os resultados para os erros E, ET, ED e EO, respectivamente, em função do tamanho inicial da janela. O gráfico 6.17 mostra o tempo de processamento também em função do tamanho inicial da janela.

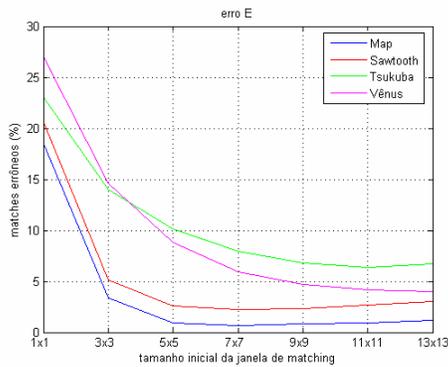


Fig. 6.13 – Erro E.

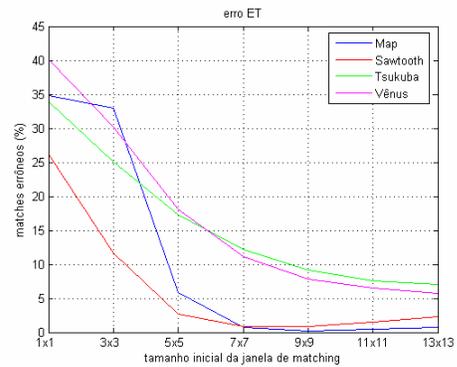


Fig. 6.14 – Erro ET.

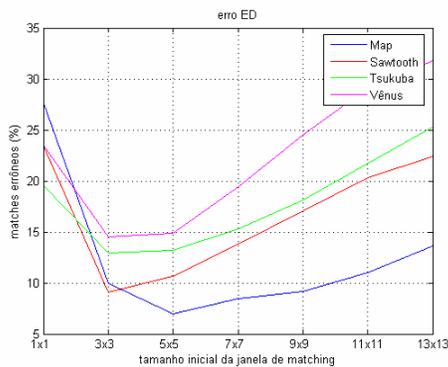


Fig. 6.15 – Erro ED.

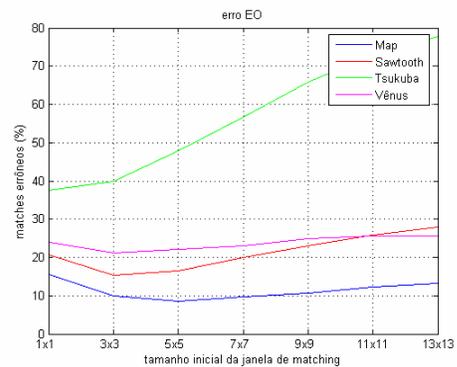


Fig. 6.16 – Erro EO.

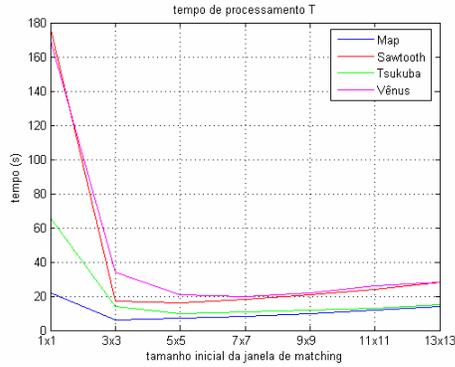


Fig. 6.17 – Tempo de Processamento T.

Sobre estes gráficos, é interessante observar o comportamento das curvas no gráfico da Figura 6.15, que retrata o erro cometido nas regiões de descontinuidade. Os menores erros são alcançados, na sua maioria, para o tamanho inicial da janela $s_i \times s_j = 3 \times 3$, evidenciando que nestas regiões as janelas pequenas fornecem melhores resultados. Seguindo a mesma linha de raciocínio, as curvas do gráfico da Figura 6.16, que retrata o erro cometido nas regiões de oclusão, também alcança os menores valores, em sua maioria, para tamanho inicial da janela $s_i \times s_j = 3 \times 3$. Isto se deve ao fato que uma boa determinação das oclusões é devida a uma boa determinação das descontinuidades.

Já o gráfico da Figura 6.14 apresenta menores erros para tamanhos iniciais da janela maiores, mostrando que em regiões com pouca textura, faz-se necessário um maior número de amostras (*pixels*) dos níveis de cinza das imagens.

O gráfico da Figura 6.13, que retrata o erro total, apresenta menores erros para tamanhos iniciais “médios” da janela, mostrando um equilíbrio ponderado entre regiões suaves e regiões com baixa textura versus regiões de oclusão e de descontinuidade. Assim, para imagens com grande quantidade de regiões suaves e/ou regiões com pouca texturas, tamanhos iniciais maiores da janela devem fornecer melhores resultados, enquanto para imagens com grande quantidade de regiões de descontinuidade e de regiões de oclusão, tamanhos iniciais menores das janelas tendem a fornecer melhores resultados.

O gráfico da Figura 6.17, que representa o tempo de processamento em função do tamanho inicial da janela, apresenta melhores resultados também para tamanhos iniciais da janela menores. Isto se deve a janelas grandes aumentam o tempo dispendido para efetuar os cálculos da correlação, porém para o tamanho inicial da janela $s_i \times s_j = 1 \times 1$, o tempo é

enorme em relação aos demais tamanhos iniciais. Isto se deve a que o tamanho da janela $s_i \times s_j = 1 \times 1$ aumenta significativamente o número de ambigüidades, aumentando com isso, o tempo gasto para resolver as ambigüidades.

6.3 Comparação dos Resultados com outras Abordagens

A tabela 5 mostra os resultados disponíveis no *site* para as abordagens que não utilizam segmentação das imagens estereoscópicas e/ou dos mapas de disparidade a fim de determinar disparidades para as regiões segmentadas e que se comparam com a abordagem proposta que utiliza apenas os valores de intensidade das imagens estereoscópicas como fonte de informação para o processo estéreo.

Como já citado no capítulo 1, as abordagens baseadas em Segmentação alcançam ótimos resultados, no entanto, faz-se necessário designar valores para vários parâmetros envolvidos na etapa de segmentação das imagens e/ou dos mapas de disparidade. Estes valores são estabelecidos de maneira empírica, tal como, "tamanho mínimo e máximo para as regiões segmentadas", além de outros de difícil interpretação, tal como "quantidade de textura homogênea".

Outras abordagens que também não foram utilizadas para efeito de comparação são aquelas baseadas na otimização de funcionais que envolvem termos de suavidade, uma vez que estes termos são ponderados por valores estabelecidos de maneira também empírica nestas abordagens.

Já a abordagem proposta utiliza-se apenas da definição do tamanho inicial da janela e o número de níveis do processo hierárquico como parâmetros que devem ser definidos pelo usuário, podendo este último ser negligenciado, uma vez que sempre é possível utilizar um alto valor (constante) para o número de níveis hierárquicos, afetando apenas o tempo de processamento. Esta é a justificativa de utilizar em todos os testes, sempre 5 níveis hierárquicos, uma vez que com 6 ou mais níveis, os resultados não apresentam mudanças significativas.

As abordagens estão ordenadas em ordem decrescente em relação a uma medida de desempenho geral computada como a somatória de todos os resultados de cada abordagem.

Este valor de somatórias encontra-se na penúltima coluna. Na última coluna tem-se a classificação ordinária.

Os erros E, ET e ED, apresentados nas tabelas 1, 2, 3 e 4 são análogos aos, apresentados na tabela 5.

A abordagem proposta neste trabalho, segundo esta classificação, ocupou a 1^o posição entre as demais.

Tabela 5 – Resultados disponibilizados no <i>site</i> .													
Abordagem	Tsukuba			Sawtooth			Venus			Map		Σ	
	<u>all</u>	untex	disc	<u>all</u>	untex	disc	<u>all</u>	untex	disc	<u>all</u>	disc		
Abordagem Proposta	6.40	7.10	12.9	2.20	0.80	9.10	4.00	5.70	14.50	0.70	7.00	70.40	1
Lee et. al., 2002	3.95	4.08	15.49	2.45	0.90	10.58	2.45	2.41	21.84	1.31	7.79	73.25	2
Scharstein & Szeliski, 2002 (Bay. diff.)	6.49	11.62	12.29	1.45	0.72	9.29	4.00	7.21	18.39	0.20	2.49	74.15	3
Zitnick & Kanade, 2000	3.49	3.65	14.77	2.03	2.29	13.41	2.57	3.52	26.38	0.22	2.37	74.70	4
Gong & Yang, 2002	2.96	2.66	14.97	2.21	2.76	13.96	2.49	2.89	23.04	1.04	10.91	79.89	5
Brockers et. al., 2004	6.33	6.63	22.93	1.51	0.29	15.06	1.44	1.24	19.11	0.43	5.99	80.96	6
Scharstein & Szeliski, 2002 (SSD+MF)	5.23	3.80	24.66	2.21	0.72	13.97	3.74	6.82	12.94	0.66	9.35	84.10	7
Birchfield & Tomasi, 1998	5.12	7.06	14.62	2.31	1.79	14.93	6.30	11.37	14.57	0.50	6.83	85.40	8
Roy & Cox, 1998	2.98	2.00	15.10	3.47	3.00	14.19	2.16	2.24	21.73	3.13	15.98	85.98	9
Mühlmann, et. al., 2002	9.76	13.85	24.39	4.76	1.87	22.49	6.48	10.36	31.29	8.42	12.68	146.35	10
Shao, 2002	9.67	7.04	35.63	4.25	3.19	30.14	6.01	6.70	43.91	2.36	33.01	181.91	11
Sun, 2002	11.1	10.70	41.99	5.51	5.56	27.39	4.36	4.78	41.13	4.17	27.88	184.57	12

A Figura 6.18 mostra um gráfico das abordagens existentes versus a somatória dos erros (Σ) apresentado na tabela 5.

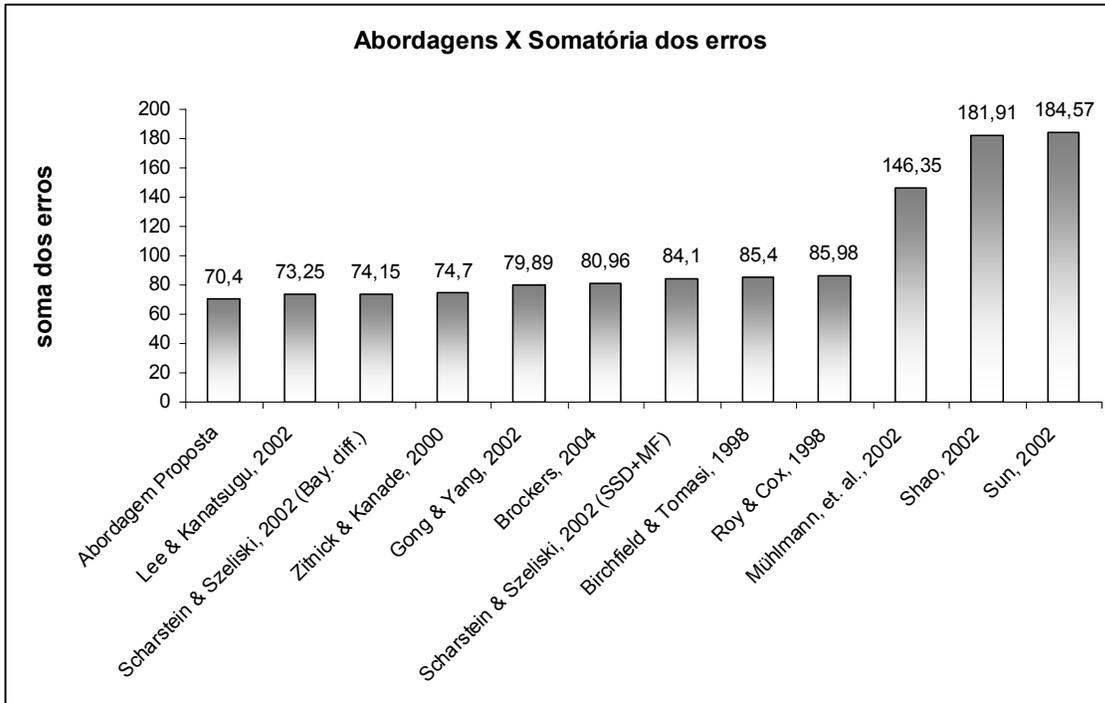


Fig. 6.18 – Gráfico das abordagens existentes versus a somatória dos erros.

6.4 Resultados para Estereogramas de Pontos Randômicos

A abordagem proposta foi testada também para Estereogramas de Pontos Randômicos.

Estereogramas de Pontos Randômicos podem ser gerados através do deslocamento relativo (disparidade) de regiões entre um par de imagens. Estas regiões homólogas recebem texturas aleatórias, porém iguais, de tal forma que a textura de uma região na imagem da esquerda é copiada para a região homóloga na imagem da direita.

A Figura 6.19 mostra um estereograma com uma textura do fundo diferente da textura do retângulo a fim de ilustrar a maneira que é gerado um Estereograma de Pontos Randômicos. Nesta Figura as texturas não são randômicas. Quando estas texturas são randômicas torna-se impossível identificar qualquer informação a respeito de formas ou feições de mais alto nível.

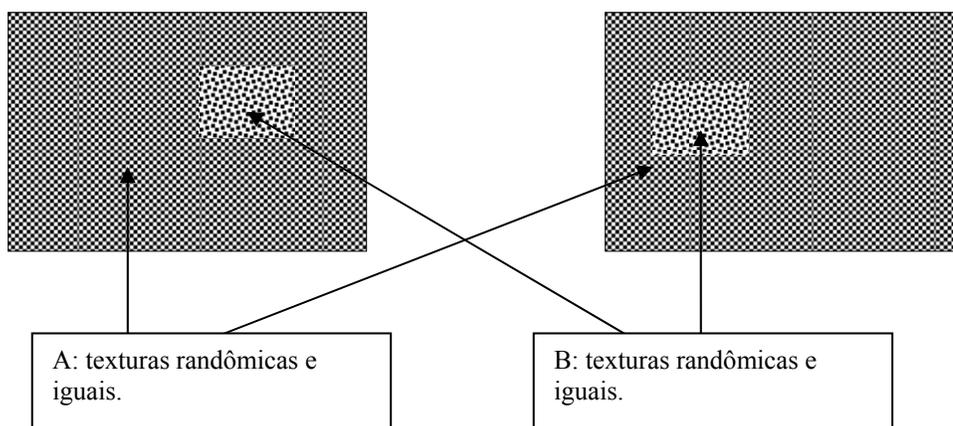


Fig. 6.19 – Geração de Estereogramas de Pontos Randômicos, com textura A diferente de textura B, embora A e B sejam randômicas.

Estereogramas de Pontos Randômicos são importantes para avaliar a capacidade do algoritmo em resolver o Problema de Correspondência isento de considerações sobre intensidades, texturas e ruídos, de tal forma que a solução é derivada apenas em função da estrutura tridimensional da cena.

Os algoritmos que realizam algum tipo de segmentação a priori das imagens estereoscópicas, para posteriormente resolver o Problema de Correspondência, não conseguem determinar a estrutura tridimensional da cena para este tipo de imagens, uma vez que não existe nenhuma informação a respeito de formas, objetos ou outras feições de mais alto nível.

6.4.1 Imagens Retângulo

As imagens Retângulo possuem 240 linhas por 320 colunas e a cena é composta por um plano ao fundo sobreposto por um outro plano limitado por um quadrado. A Figura 6.20 mostra da esquerda para direita e de cima para baixo, o par de imagens estereoscópicas, os mapas de disparidade reconstruídos para a imagem da esquerda e da direita, os mapas de confiança associados às imagens da esquerda e da direita e dois gráficos tridimensionais do mapa de disparidade esquerdo. Os mapas de confiança mostram para cada *pixel* qual o nível hierárquico do processo hierárquico foi estabelecida à correspondência. Os tons mais claros codificam os níveis superiores e os tons escuros os níveis inferiores.

O tamanho inicial da janela utilizado é $s_i \times s_j = 1 \times 1$ e 5 níveis hierárquicos.

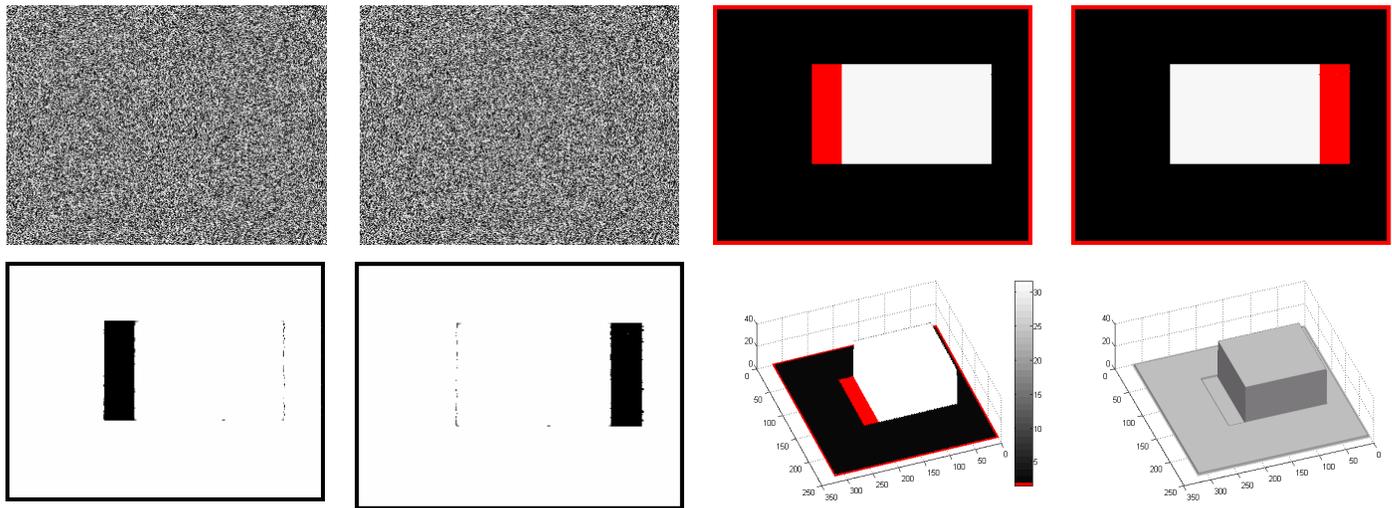


Fig. 6.20 – Imagens Retângulo: Imagens Estereoscópicas, Mapas de Disparidade reconstruídos, Mapas de Confiança e gráficos tridimensionais para $s_i \times s_j = 1 \times 1$.

Este par de imagens estereoscópicas é interessante para entender o fenômeno da oclusão, que neste caso, ocorre à esquerda do retângulo na imagem da esquerda e a direita do retângulo na imagem da direita. Isto porque, no sentido da esquerda para a direita, a oclusão ocorre na imagem da esquerda quando a disparidade varia de um valor menor para um valor maior e a oclusão ocorre na imagem da direita quando a disparidade varia de um valor maior para um valor menor.

Existem apenas dois *pixels* (por inspeção visual dos mapas de disparidades reconstruídos) nestes mapas de disparidade que tiveram valores errôneos de disparidade associados, o que resulta em aproximadamente 0.0026% de pontos errados. Estes *pixels* errados ocorrem na região de oclusão no mapa de disparidade da direita.

A Figura 6.21 mostra o mesmo que a Figura 6.20, porém para tamanho inicial da janela $s_i \times s_j = 41 \times 41$.

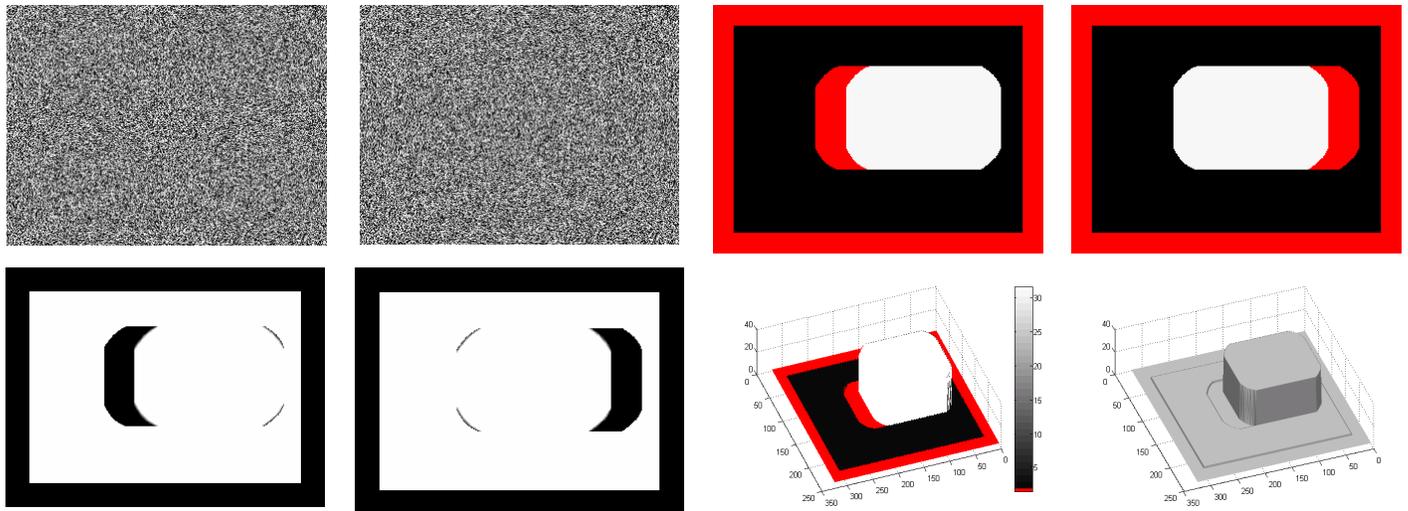


Fig. 6.21 – Imagens Retângulo: Imagens Estereoscópicas, Mapas de Disparidade reconstruídos, Mapas de Confiança e gráficos tridimensionais para $s_i^1 \times s_j^1 = 41 \times 41$.

Como é de se esperar, os cantos (canto "vivo") do retângulo não foram reconstruídos adequadamente devido ao tamanho da janela ser exageradamente grande, já para o primeiro nível do processo hierárquico. Neste teste pode-se perceber que, apesar da regularidade das feições nos mapas de disparidade, a quantidade de pares errados é maior em comparação a quantidade de pares errados para a Figura 6.20.

6.4.2 Imagens Redondo

Estas imagens possuem 240 linhas por 320 colunas, são sintéticas e a cena é composta por um plano ao fundo sobreposto por uma semi-esfera.

A Figura 6.22 mostra da esquerda para direita e de cima para baixo, o par de imagens estereoscópicas, os mapas de disparidade reconstruídos para a imagem da esquerda e da direita, os mapas de confiança associados às imagens da esquerda e da direita e dois gráficos tridimensionais do mapa de disparidade esquerdo.

O tamanho inicial da janela utilizado é $s_i^1 \times s_j^1 = 1 \times 1$ e 5 níveis hierárquicos.

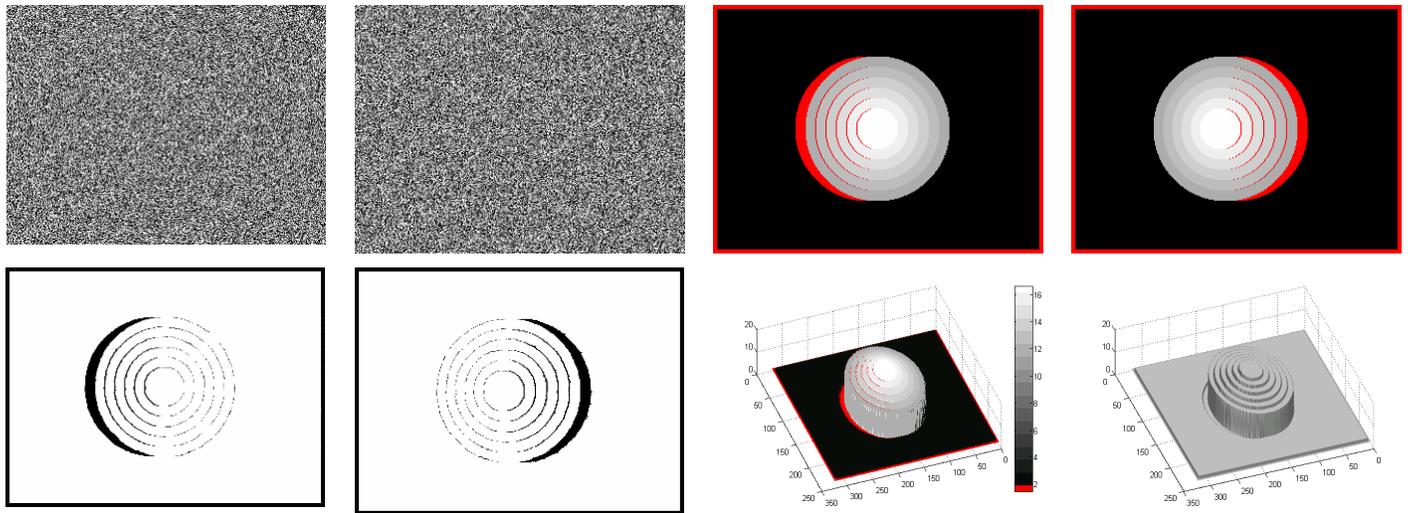


Fig. 6.22 – Imagens Redondo: Imagens Estereoscópicas, Mapas de Disparidade reconstruídos, Mapas de Confiança e gráficos tridimensionais para $s_i^1 \times s_i^1 = 1 \times 1$.

Neste exemplo, se percebe facilmente nos mapas de disparidade as regiões de oclusão que ocorrem paralelas às curvas de nível, no lado esquerdo no mapa de disparidade da esquerda e no lado direito no mapa de disparidade da direita.

Não existe nenhuma *pixel* (por inspeção visual) nestes mapas de disparidade que tiveram valores de disparidade associados errados.

A Figura 6.23 utiliza a mesma organização das imagens na Figura 6.22. Os resultados são para tamanho inicial da janela $s_i^1 \times s_i^1 = 39 \times 39$.

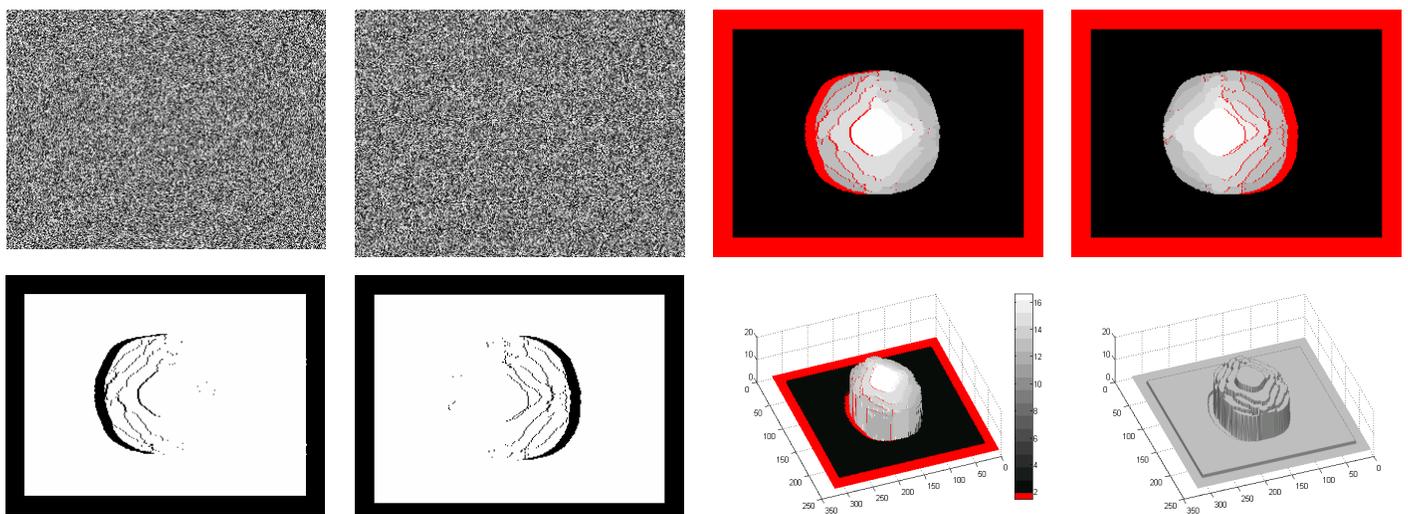


Fig. 6.23 – Imagens Redondo: Imagens Estereoscópicas, Mapas de Disparidade reconstruídos, Mapas de Confiança e gráficos tridimensionais para $s_i^1 \times s_i^1 = 39 \times 39$.

Neste, o tamanho inicial da janela extremamente grande prejudicou sensivelmente o mapa reconstruído.

6.4.3 Imagens Xadrez

Estas imagens possuem 240 linhas por 320 colunas, são sintéticas e a cena é composta por um quadriculado no estilo de um tabuleiro de xadrez, onde metade dos quadrados do tabuleiro possui um valor de disparidade e a outra metade possui outro valor de disparidade.

A Figura 6.24 mostra da esquerda para direita e de cima para baixo, o par de imagens estereoscópicas, os mapas de disparidade reconstruídos para a imagem da esquerda e da direita, os mapas de confiança associados às imagens da esquerda e da direita e dois gráficos tridimensionais do mapa de disparidade esquerdo.

O tamanho inicial da janela utilizado é $s_i^1 \times s_j^1 = 1 \times 1$ e 5 níveis hierárquicos.

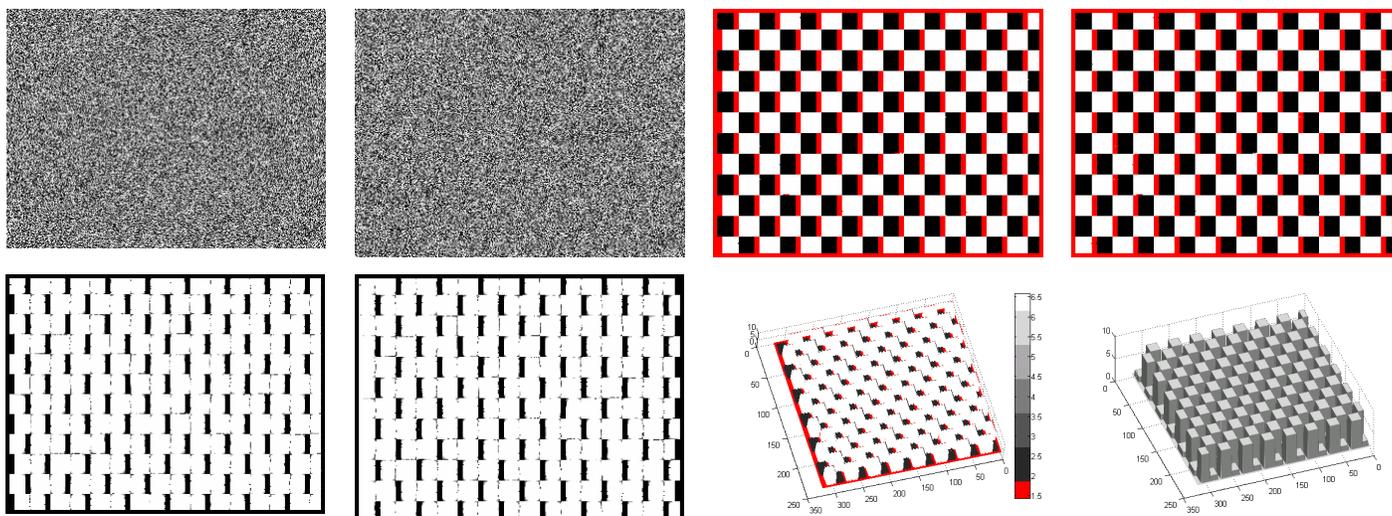


Fig. 6.24 – Imagens Xadrez: Imagens Estereoscópicas, Mapas de Disparidade reconstruídos, Mapas de Confiança e gráficos tridimensionais para $s_i^1 \times s_j^1 = 1 \times 1$.

Neste exemplo, pode-se perceber que mesmo para regiões pequenas (cada quadrado do "tabuleiro de xadrez") a reconstrução da superfície é praticamente perfeita, mesmo em regiões de descontinuidade da superfície.

Foram contabilizados 19 pares errados (por inspeção visual), o que resulta em 0.024% de pares errados.

Um dos principais parâmetros geralmente empregados nos algoritmos baseados em segmentação é o tamanho da área mínima de cada região rotulada. Como neste exemplo, a cena possui "pedaços" de planos pequenos, os resultados destas metodologias podem resultar em erros bastante grosseiros, uma vez que o classificador pode rotular duas regiões com valores de disparidade distintos com o mesmo rotulo devido a pequena quantidade de *pixels* em cada uma dessas regiões.

A Figura 6.25 mostra os mapas de disparidade e o gráfico tridimensional apenas associados a imagem da esquerda gerados para distintos tamanhos iniciais da janela e 5 níveis hierárquicos.

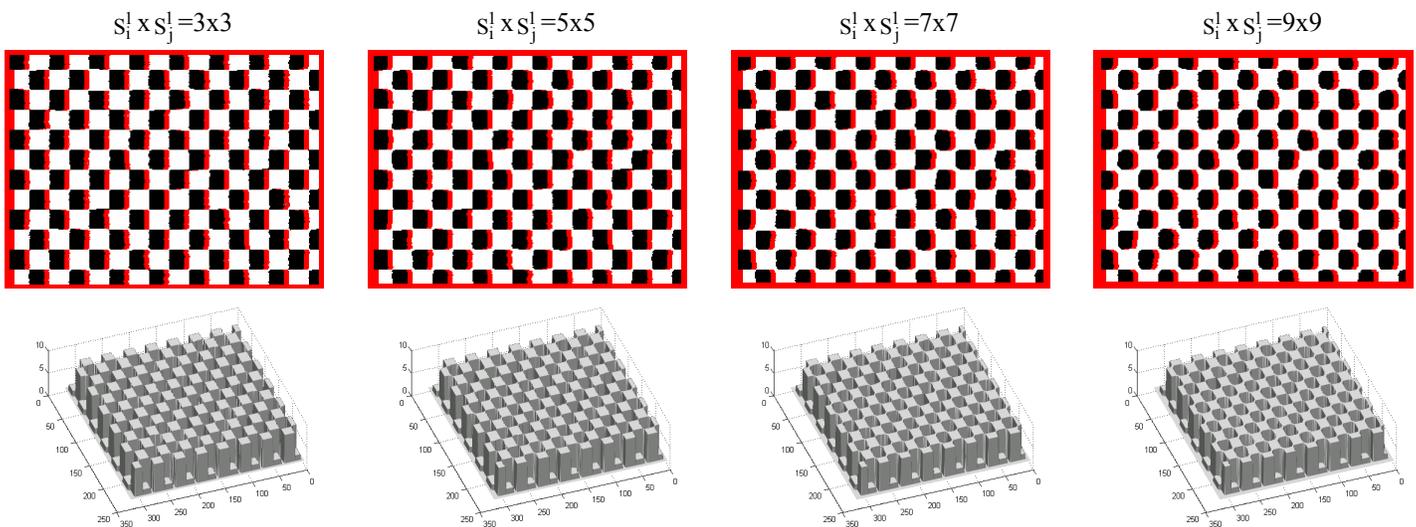


Fig. 6.25 – Resultados para Imagens Xadrez em função do tamanho inicial da janela.

Pode-se perceber que o tamanho da janela inicial está diretamente relacionado com a degradação das regiões de descontinuidade da cena. Uma vez que sempre que ocorre uma descontinuidade entre valores de disparidade ocorre oclusão, estas por sua vez, também são degradadas de maneira diretamente proporcional ao tamanho da janela de correlação.

6.5 Resultados Qualitativos

Os resultados apresentados nesta seção são apenas para inspeção visual, fazendo-se necessário um bom conhecimento das dificuldades existentes nas cenas e das fragilidades dos algoritmos para reconstrução tridimensional baseado em Visão Estéreo, a fim de inferir algum julgamento de caráter subjetivo.

6.5.1 Imagens Corredor

Estas imagens possuem 256 linhas por 256 colunas, são sintéticas e a cena é composta por um corredor, uma bola e um cone. A principal dificuldade em gerar os mapas de disparidade para este par de imagens é devido a grande quantidade de regiões com texturas homogêneas.

A Figura 6.26 mostra da esquerda para direita e de cima para baixo as imagens estereoscópicas, os mapas de disparidade reconstruídos para a imagem da esquerda e da direita, os mapas de confiança associados às imagens da esquerda e da direita e dois gráficos tridimensionais do mapa de disparidade esquerdo.

O tamanho inicial da janela utilizado é $s_i \times s_j = 1 \times 1$ e 5 níveis hierárquicos.

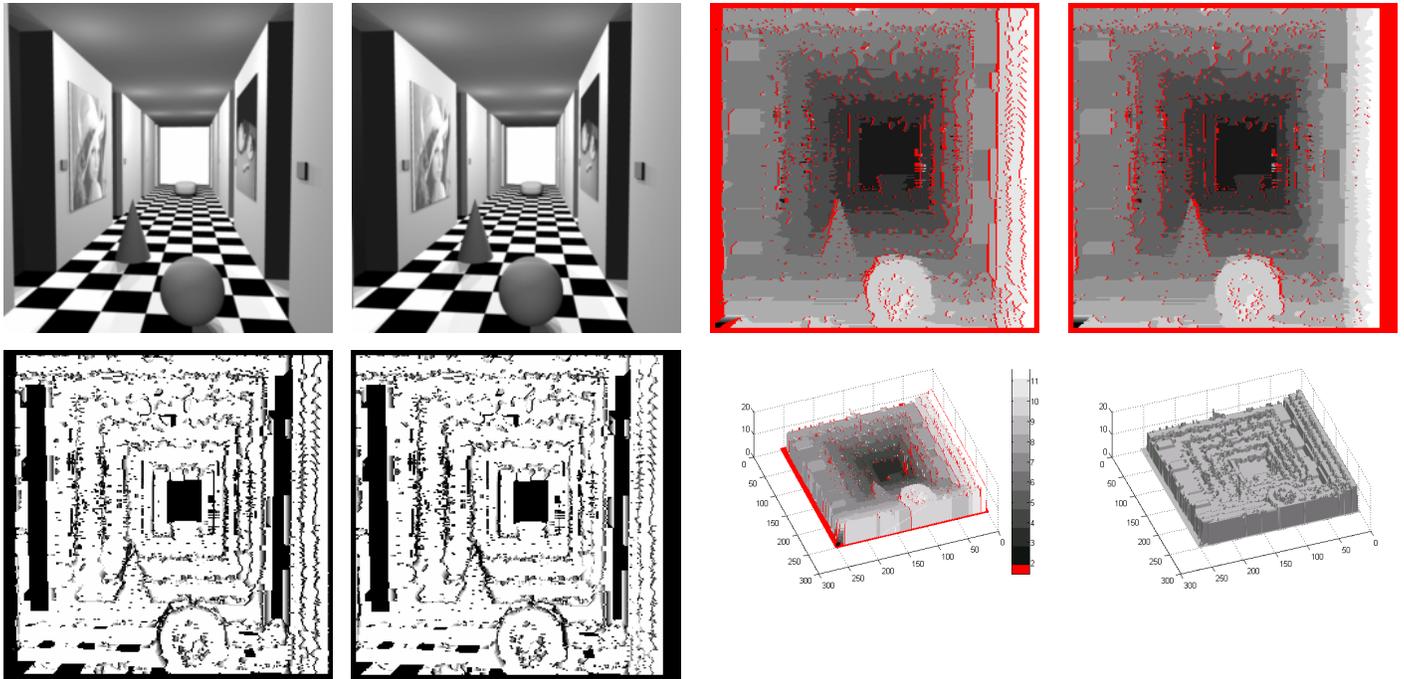


Fig. 6.26 – Imagens Corredor : Imagens Estereoscópicas, Mapas de Disparidade reconstruídos, Mapas de Confiança e gráficos tridimensionais para $s_i^1 \times s_j^1 = 1 \times 1$.

Neste exemplo, pode-se perceber nos mapas de confiança que as regiões de textura homogêneas aparecem representadas pela cor preta, indicando que os valores de disparidade associados a estas regiões possuem baixa confiança.

6.5.2 Imagens Manequim

Estas imagens possuem 320 linhas por 240 colunas, são reais e a cena é composta por o rosto de um manequim e um plano de fundo.

A Figura 6.27 mostra da esquerda para direita e de cima para baixo o par de imagens estereoscópicas, os mapas de disparidade reconstruídos para a imagem da esquerda e da direita, os mapas de confiança associados às imagens da esquerda e da direita e na quarta linha e dois gráficos tridimensionais do mapa de disparidade esquerdo.

O tamanho inicial da janela utilizado é $s_i^1 \times s_j^1 = 9 \times 9$ e 5 níveis hierárquicos.

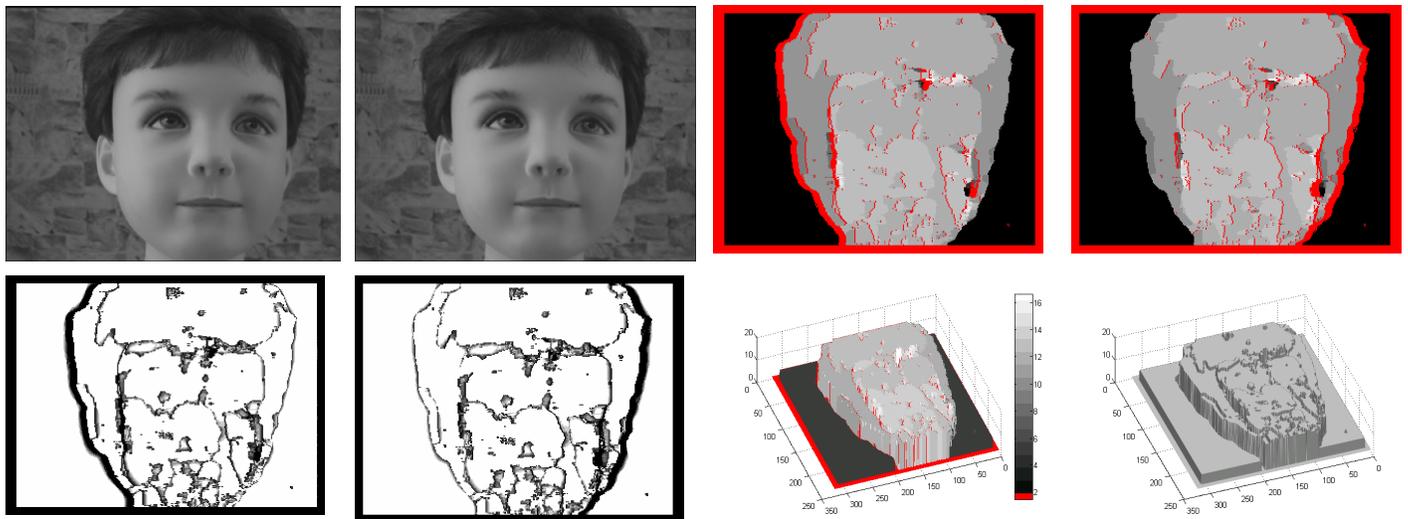


Fig. 6.27 – Imagens Manequim: Imagens Estereoscópicas, Mapas de Disparidade reconstruídos, Mapas de Confiança e gráficos tridimensionais para $s_i^1 \times s_j^1 = 9 \times 9$.

Uma vez que o par de imagens estereoscópicas apresenta regiões com texturas bastante homogêneas fez-se necessário utilizar o tamanho inicial da janela grande. No entanto, este fato não acarretou maiores problemas para a cena reconstruída uma vez que apenas nos limites do rosto é que se verificam regiões de descontinuidade de superfícies ao passo que, no restante da cena (fundo e o próprio rosto) não existem regiões de descontinuidade.

6.6 Discussão e Síntese dos Resultados

De maneira geral, a principal característica da metodologia proposta que se observa nos resultados apresentados, é que, para tamanhos iniciais de janelas pequenas resultam em mapas de disparidade ruidosos, porém preservando de certo modo as descontinuidades presentes nas cenas. Já quando o tamanho inicial da janela é grande, os mapas de disparidade são menos ruidosos, mais suaves, não preservam as descontinuidades e representam mal os objetos “estritos” ou finos da cena.

A abordagem proposta foi a que apresentou o melhor resultado geral dentre as demais abordagens apresentadas na tabela 5. Sobre esse resultado, cabe ainda ressaltar que os resultados obtidos pela abordagem proposta poderiam ser ainda melhores em relação aos resultados das demais abordagens se a metodologia de testes considerasse os *pixels*

pertencentes às regiões de oclusão para efetivar as estatísticas da metodologia utilizada nos testes. Dessa forma, as abordagens que não consideram oclusões determinam valores de disparidades para todos os *pixels* das imagens e então para as regiões de oclusão esses *pixels* estariam errados.

Entende-se que a metodologia de testes deveria considerar como corretos os *pixels* pertencentes às regiões de oclusão se estes foram devidamente rotulados como oclusão (e não simplesmente desconsiderar estes *pixels*).

Pode-se dizer que a abordagem proposta neste trabalho opera em um baixo nível de informação. O trabalho de Marr & Poggio, 1976, mostrou que a fusão das imagens estereoscópicas pelo cérebro humano é possível mesmo para imagens estereoscópicas que não possuem nenhuma informação a respeito de formas, objetos ou outras feições de mais alto nível, tal como Estereogramas de Pontos Randômicos, quando observados monocularmente. No entanto é possível perceber a estrutura tridimensional da cena mesmo a olho nu. Este fato mostra que a segmentação das imagens estereoscópicas não é uma atividade necessária para a fusão das imagens estereoscópicas e conseqüente reconstrução da cena pelo sistema visual humano.

Os algoritmos que realizam algum tipo de segmentação a priori das imagens estereoscópicas, para posteriormente resolver o Problema de Correspondência, não conseguem determinar a estrutura tridimensional da cena para Estereogramas de Pontos Randômicos, uma vez que não existe nenhuma informação a respeito de formas, objetos ou outras feições de mais alto nível.

A importância em utilizar Estereogramas de Pontos Randômicos na avaliação de uma abordagem não está no fato direto da abordagem estar habilitada em resolver ou não Estereogramas de Pontos Randômicos, uma vez que este tipo de imagem não ocorre em situações reais, mas sim em avaliar a independência da abordagem em relação aos parâmetros utilizados nos processos de segmentação. Com isso, entende-se que Estereogramas de Pontos Randômicos deveriam ser considerados também para avaliar as abordagens para resolução do Problema de Correspondência.

A abordagem proposta resolveu, de maneira bastante precisa a reconstrução dos Estereogramas de Pontos Randômicos apresentados neste capítulo. Este fato comprova a robustez da abordagem proposta em relação ao tipo de imagem/cena utilizada.

Capítulo 7

7. Conclusões e Recomendações

Este capítulo apresenta a síntese deste trabalho, conclusões, contribuições e sugestões para trabalhos futuros baseadas nos resultados obtidos através dos testes realizados em vários pares de imagens estereoscópicas.

7.1 Síntese

Este trabalho apresentou uma nova abordagem para reconstrução da estrutura tridimensional de uma cena a partir de um par de imagens estereoscópicas retificadas, na qual oclusões são explicitamente identificadas. A abordagem proposta consiste basicamente na imposição da Injunção de Ordem e demais regras auxiliares em um processo hierárquico e iterativo no qual são atribuídas prioridades aos pontos do mapa de disparidade que possuem a propriedade de Correspondência Bidirecional para o maior número de janelas de correspondência organizadas em ordem crescente de tamanho.

A abordagem proposta enfatiza o respeito e a identificação das regiões de oclusão presentes nos mapas de disparidades reconstruídos. A identificação das regiões de oclusão é realizada conjuntamente com a resolução (desambiguação) do mapa de disparidade.

Uma característica importante da abordagem proposta é a utilização de apenas dois parâmetros que devem ser fornecidos a priori de maneira empírica que são: tamanho inicial da janela e número de níveis hierárquicos.

De maneira prática, desconsiderando o tempo de processamento, o número de níveis hierárquicos pode ser sempre super-dimensionado, uma vez que o aumento deste número não acarreta em resultados piores. Portanto, pode-se considerar que a abordagem proposta utiliza-se apenas do parâmetro tamanho inicial da janela de maneira efetiva. Este parâmetro por sua vez pode ser entendido como o agente responsável pela qualidade da reconstrução, principalmente, nas regiões de descontinuidade da estrutura tridimensional da cena.

Dessa forma, o tamanho inicial pequeno de janela ocasiona, em geral, uma boa reconstrução das regiões de descontinuidades e/ou delgadas (finas ou "estreitas") presentes no mapa de disparidade, porém, ocorrendo ruídos para as demais regiões, principalmente para aquelas associadas às regiões com baixa textura nas imagens estereoscópicas.

Já para o tamanho inicial grande de janela as regiões de descontinuidades e/ou delgadas não são bem definidas no mapa de disparidade reconstruído. No entanto, as regiões suaves e/ou com texturas homogêneas associadas nas imagens estereoscópicas apresentam boa reconstrução, sem ocorrência de ruídos.

Cabe ressaltar ainda que a abordagem proposta não utiliza uma função-objetivo explícita que deve ser minimizada a fim de obter uma solução global para todos os pontos do mapa de disparidade. A solução global é realizada apenas para cada linha do mapa de disparidade (que por sua vez está associado a um par de linhas epipolares homólogas) através de um Processo Cooperativo fundamentado em aspectos geométricos do processo de formação das imagens definido pela Injunção de Ordem.

Este fato é importante porque a utilização de uma função-objetivo que pode ser apropriada para um tipo de imagem/cena pode não ser para outro tipo de imagem/cena, restringindo a aplicabilidade da abordagem.

7.2 Conclusões

De acordo com os resultados apresentados no capítulo 6 percebe-se que os piores resultados ocorrem para o tamanho da janela inicial 1x1 e imagens com ruídos e texturas homogêneas devido ao grande número de pares ambíguos gerados para este tamanho de janela. Com isso, a porcentagem de pares errôneos e o tempo de processamento são muito maiores em relação aos demais tamanhos iniciais de janela de correlação. No entanto, quando esses fatores (ruídos e texturas homogêneas) não são presentes, como nos Estereogramas de Pontos Randômicos, o tamanho da janela inicial 1x1 apresenta os melhores resultados.

Como esperado, pode-se afirmar que o principal fator responsável em preservar descontinuidades e, portanto, conseqüentemente oclusões, é a Correspondência Bidirecional Hierárquica, iniciando com tamanho inicial da janela pequena. No entanto,

deste modo as regiões de superfícies suaves da cena não são bem reconstruídas devido a grande presença de pares ambíguos presentes nestas regiões com texturas homogêneas e ruidos.

Com isso pode-se concluir que o tamanho inicial ótimo da janela deve ser obtido em função de medidas que quantifiquem coerentemente a energia do processo ruidoso e das regiões com texturas homogêneas.

A imposição da Injunção de Ordem mostrou-se fundamental para a resolução do problema de Correspondência, excluindo nas primeiras iterações uma grande quantidade de pares errôneos. Este fato, conseqüentemente reduziu significativamente o tempo de processamento.

A identificação das regiões de oclusão do par de imagens estereoscópicas é de igual importância à determinação dos valores de disparidade e, portanto, devem ser consideradas em qualquer algoritmo estéreo. Neste trabalho foi apresentada uma abordagem consistente para a detecção das regiões de oclusões através da imposição da Injunção de Ordem.

Quanto à comparação dos resultados com outras abordagens pode-se concluir que a abordagem proposta foi superior as demais abordagens que não são baseadas em Segmentação e/ou se utilizam de otimização de funcionais cujos parâmetros são determinados de maneira empírica.

Os resultados alcançados com a abordagem proposta são, em geral, inferiores aos resultados alcançados pelas abordagens recentes baseadas em Segmentação, porém, como já citado, tais abordagens utilizam um grande número de parâmetros definidos de maneira empírica, os quais dependem das condições de iluminação e principalmente da própria estrutura tridimensional da cena.

7.3 Contribuições

A principal contribuição deste trabalho é a definição da Correspondência Bidirecional Hierárquica.

Apesar da Correspondência Bidirecional Hierárquica ser praticamente invariante ao aumento do tamanho da janela de correlação para grandes valores de números de níveis hierárquicos (os resultados são praticamente idênticos para 5 ou 6 níveis hierárquicos, por

exemplo), esta não é invariante ao tamanho inicial da janela de correlação. Isto se deve ao fato da abordagem estabelecer prioridades de maneira hierárquica, de tal forma que se um par correto não possuir a propriedade de Correspondência Bidirecional para um dado nível n do processo hierárquico, sua prioridade não poderá mais aumentar para o nível $n + 1$.

Outra contribuição deste trabalho é a utilização da tabela para implementação da Injunção de Ordem, reduzindo consideravelmente o custo computacional para a imposição desta injunção, uma vez que apenas os intervalos dos *pixels* homólogos possíveis são alterados em uma tabela e não na Matriz de Prioridade. A utilização desta tabela reduz significativamente o custo computacional na abordagem proposta.

7.4 Trabalhos Futuros

Com o intuito de tornar suave o mapa de disparidades preservando discontinuidades, propõe-se como trabalho futuro o emprego da Correspondência Bidirecional Hierárquica com formas adaptativas das janelas de correlação ou mesmo operações de filtragem diretamente sobre o mapa de disparidade reconstruído.

Ainda como trabalho futuro, sugere-se a extensão da abordagem proposta para três ou mais imagens estereoscópicas, o que, em princípio, deve melhorar os resultados obtidos, uma vez que um *pixel* que não possua homólogo em uma segunda imagem e, portanto, identificado como oclusão, pode possuir um *pixel* homólogo em uma terceira imagem, tornando possível reconstruir uma quantidade maior de regiões (área) da cena.

Uma configuração interessante das imagens estereoscópicas seria utilizar quatro imagens dispostas em bloco 2×2 , com um par de imagens estereoscópicas superior e outro inferior. Nessa situação, iterações entre linhas epipolares poderiam ser estabelecidas fundamentadas em aspectos geométricos do processo de formação das imagens definido pela Injunção de Ordem, porém entre as imagens superiores e inferiores.

Outro aspecto da abordagem proposta que pode ser tratado em trabalhos futuros é a determinação do tamanho da janela inicial através de técnicas de otimização considerando alguma métrica que quantifique a qualidade do mapa de disparidade reconstruído em função de medidas que quantifiquem coerentemente a energia do processo ruidoso e das regiões com texturas homogêneas.

Anexo

A.1 Aquisição

A aquisição das imagens é a etapa inicial de todo o processo estéreo. São estas as fontes de informação utilizada para a recuperação da geometria da cena e, portanto, conseqüentemente, a qualidade das imagens influencia diretamente a qualidade do resultado final.

A condição fundamental para adquirir imagens estéreo é que o par (ou mais) de imagens sejam capturas a partir de posições distintas. Assim, duas maneiras de se obter as imagens podem ser consideradas nesta etapa em relação ao número de sensores (câmeras) utilizado:

- Aquisição com uma câmera; e
- Aquisição com duas ou mais câmeras.

Na abordagem que utiliza apenas uma câmera, as imagens são tomadas deslocando a mesma câmera para posições distintas. Devido ao inevitável tempo despendido no deslocamento da posição da câmera, as mesmas condições de iluminação e geometria da cena não podem ser garantidas para todas as imagens capturadas.

Um exemplo típico desta abordagem é quando a câmera é transportada em um veículo aéreo e as imagens são capturadas durante o deslocamento desta aeronave. Esta técnica é denominada de Aerofotogrametria sendo geralmente utilizada para fins cartográficos. Nesta situação é bastante comum que as condições de iluminação sejam alteradas entre as imagens devido ao próprio deslocamento natural da fonte de luz (sol) ou devido à ocorrência de nuvens altas (acima da aeronave) diminuindo consideravelmente a quantidade de luz transmitida (sombas).

Embora, ainda sobre a Aerofotogrametria, as alterações da geometria da cena ocorram entre as tomadas, tais como deslocamentos de veículos e pessoas, estas são geralmente irrelevantes, uma vez que as feições de interesse cartográfico como, por exemplo, edificações e relevo, permaneçam estáticos.

A principal vantagem em utilizar apenas uma câmera é que os parâmetros de calibração interior da câmera, tais como distância focal (no caso de câmera com distância focal fixa), deslocamento do ponto principal, coeficientes de distorções do sistema de lentes, entre outros, podem ser considerados constantes para todas as imagens capturadas na maioria das aplicações.

Na abordagem que utiliza duas ou mais câmeras, as imagens podem ser tomadas simultaneamente de posições distintas, bastando para isso, sincronizar o instante de captura dos sensores. Nesta situação pode-se considerar inalteradas as condições de iluminação e geometria da cena entre as imagens. A desvantagem desta abordagem em relação à abordagem que considera apenas uma câmera é que os parâmetros de calibração interior das câmeras não podem ser considerados iguais para todas as imagens, necessitando, portanto, de um processo preciso de calibração interior para todas as câmeras. Outra desvantagem é obviamente o maior custo de compra/manutenção de duas ou mais câmeras.

Formação Geométrica da Imagem

O modelo de câmera geralmente utilizado em Visão Estéreo é o modelo de câmera de orifício. Sem demonstrações, o modelo matemático para este problema pode ser dado por:

$$\begin{aligned} x &= f \frac{[r_{11}(X - X_{CP}) + r_{12}(Y - Y_{CP}) + r_{13}(Z - Z_{CP})]}{[r_{31}(X - X_{CP}) + r_{32}(Y - Y_{CP}) + r_{33}(Z - Z_{CP})]} \\ y &= f \frac{[r_{21}(X - X_{CP}) + r_{22}(Y - Y_{CP}) + r_{23}(Z - Z_{CP})]}{[r_{31}(X - X_{CP}) + r_{32}(Y - Y_{CP}) + r_{33}(Z - Z_{CP})]} \end{aligned} \quad (\text{A.1})$$

onde:

(X, Y, Z) são coordenadas de um ponto no sistema referencial do espaço-objeto;

(x, y) são coordenadas de um ponto no sistema referencial do espaço-imagem;

(X_{CP}, Y_{CP}, Z_{CP}) são coordenadas do centro perspectivo da câmera;

f é a distância focal da câmera;

r_{ij} são os elementos da matriz de rotação entre o sistema referencial dos espaço-imagem em relação ao sistema referencial do espaço-objeto, para $i, j = 1, 2$ e 3 .

Este modelo pode ser acrescido de outras variáveis, tais como, distorção radial do sistema de lentes, deslocamento do ponto principal e fator de escala entre as dimensões horizontal e vertical do *pixel*. Considerando estas variáveis o modelo pode ser escrito como:

$$x_R - c_x + (x_R - c_x)(k_1 r^2) + (x_R - c_x) ds_x = f \frac{[r_{11}(X - X_{CP}) + r_{12}(Y - Y_{CP}) + r_{13}(Z - Z_{CP})]}{[r_{31}(X - X_{CP}) + r_{32}(Y - Y_{CP}) + r_{33}(Z - Z_{CP})]} \quad (A.2)$$

$$y_R - c_y + (y_R - c_y)(k_1 r^2) = f \frac{[r_{21}(X - X_{CP}) + r_{22}(Y - Y_{CP}) + r_{23}(Z - Z_{CP})]}{[r_{31}(X - X_{CP}) + r_{32}(Y - Y_{CP}) + r_{33}(Z - Z_{CP})]}$$

onde:

(x_R, y_R) são as coordenadas do ponto principal;

ds_x é o fator de escala entre as dimensões horizontais e verticais do *pixel*;

(c_x, c_y) são as coordenadas do sistema de lentes; e

k_1 é o coeficiente de distorção radial do sistema de lentes.

A.2 Calibração

A etapa de Calibração consiste em determinar os parâmetros de algum modelo matemático (por exemplo, o modelo dado pela expressão A.2) que estabelece a correspondência entre os pontos 3D no espaço-objeto constituintes da cena e suas respectivas imagens 2D no espaço-imagem.

O processo de calibração é realizado através da resolução de um sistema de equações, geralmente sobre-determinado, no qual as variáveis do sistema são os parâmetros do modelo empregado e as observações são as coordenadas no espaço-objeto e suas respectivas coordenadas no espaço-imagem.

Os parâmetros empregados nos modelos são geralmente classificados como intrínsecos e extrínsecos.

Os parâmetros intrínsecos estão relacionados com os aspectos físicos de construção da câmera, tais como distância focal, distorções do sistema lentes, deslocamentos do ponto principal, tamanho do *pixel*, entre outros.

Os parâmetros extrínsecos estão relacionados com o estado da câmera no momento da tomada da imagem. Estes parâmetros são os três ângulos de rotação e os três deslocamentos do sistema referencial do espaço-imagem em relação ao sistema referencial do espaço-objeto.

Uma grande dificuldade encontrada no processo de calibração é a mensuração das coordenadas 3-D (espaço-objeto) com precisão suficiente para garantir que a precisão almejada para os parâmetros seja obtida no processo de calibração. Para isto faz-se necessário o emprego de equipamentos que possibilitam realizar tais medidas, como, por exemplo, teodolitos, distanciômetros ou mesmo GPS (*Global Position System*).

A quantidade mínima de pontos 3-D necessários neste processo é função do número de parâmetros envolvidos no modelo matemático. Um dos modelos mais utilizados foi desenvolvido por Tsai, 1987. Uma consideração relevante sobre este trabalho é o desenvolvimento de uma metodologia para alvos (pontos com coordenadas conhecidas no espaço-objeto) coplanares, os quais são mais fáceis de serem construídos (e medidos) em relação aos alvos não coplanares.

A.3 Triangulação

A etapa de Triangulação é a última do processo estéreo. É nesta etapa que são determinadas coordenadas tridimensionais (X,Y,Z) referentes a um sistema de coordenadas do espaço-objeto para cada ponto da imagem esquerda e/ou direita do par estereoscópico. A Figura A.1 mostra o problema de triangulação.

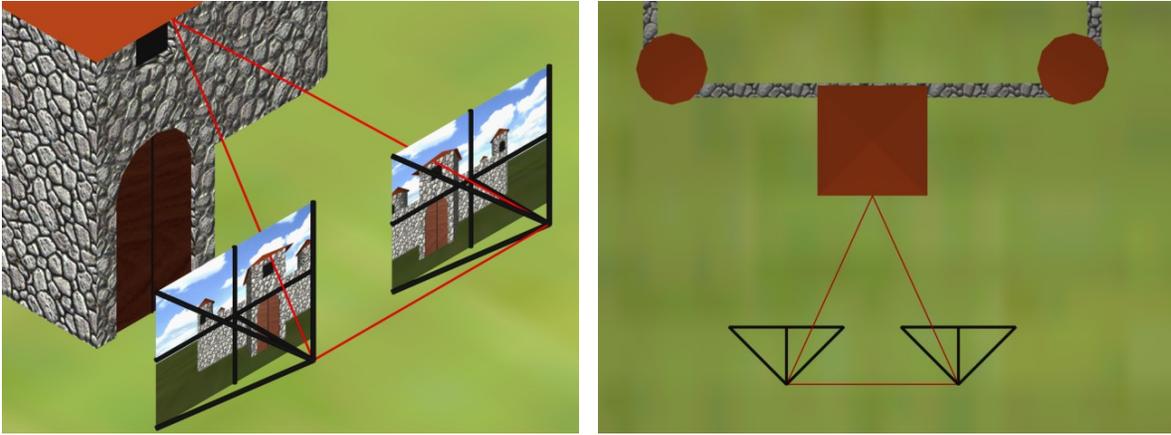


Fig. A.1 – Triângulo de reconstrução.

O triângulo de reconstrução pode ser resolvido matematicamente pela resolução do seguinte sistema de equações:

$$\left\{ \begin{array}{l} x_e = f_e \frac{[r_{11}^e(X - X_{CP}^e) + r_{12}^e(Y - Y_{CP}^e) + r_{13}^e(Z - Z_{CP}^e)]}{[r_{31}^e(X - X_{CP}^e) + r_{32}^e(Y - Y_{CP}^e) + r_{33}^e(Z - Z_{CP}^e)]} \\ y_e = f_e \frac{[r_{21}^e(X - X_{CP}^e) + r_{22}^e(Y - Y_{CP}^e) + r_{23}^e(Z - Z_{CP}^e)]}{[r_{31}^e(X - X_{CP}^e) + r_{32}^e(Y - Y_{CP}^e) + r_{33}^e(Z - Z_{CP}^e)]} \\ x_d = f_d \frac{[r_{11}^d(X - X_{CP}^d) + r_{12}^d(Y - Y_{CP}^d) + r_{13}^d(Z - Z_{CP}^d)]}{[r_{31}^d(X - X_{CP}^d) + r_{32}^d(Y - Y_{CP}^d) + r_{33}^d(Z - Z_{CP}^d)]} \\ y_d = f_d \frac{[r_{21}^d(X - X_{CP}^d) + r_{22}^d(Y - Y_{CP}^d) + r_{23}^d(Z - Z_{CP}^d)]}{[r_{31}^d(X - X_{CP}^d) + r_{32}^d(Y - Y_{CP}^d) + r_{33}^d(Z - Z_{CP}^d)]} \end{array} \right. \quad (A.3)$$

onde:

(X, Y, Z) são coordenadas de um ponto no sistema referencial do espaço-objeto;

(x_e, y_e) e (x_d, y_d) são coordenadas de um ponto no sistema referencial do espaço-

imagem da câmera da esquerda e da direita, respectivamente;

$(X_{CP}^e, Y_{CP}^e, Z_{CP}^e)$ e $(X_{CP}^d, Y_{CP}^d, Z_{CP}^d)$ são coordenadas dos centros perspectivos das câmeras da esquerda e da direita, respectivamente.

f_e e f_d são as distâncias focais das câmeras da esquerda e da direita, respectivamente.

r_{ij}^e e r_{ij}^d são os elementos das matrizes de rotação entre o sistema referencial dos espaços-imagem da esquerda e da direita em relação ao sistema referencial do espaço-objeto, respectivamente, para $i, j = 1, 2$ e 3 .

Este sistema, como pode ser visto, é composto por quatro equações e três incógnitas, resultando, portanto, em um grau de liberdade. Esta redundância possibilita a determinação também de alguns parâmetros estatísticos, os quais podem ser utilizados para estimativas da precisão alcançada.

É importante ressaltar ainda que este sistema torna-se mal-condicionado para os pontos com grandes profundidades. A interpretação geométrica deste fenômeno é explicada facilmente devido à dificuldade em determinar o ponto de interseção das duas retas formadas pelos pontos homólogos e os seus respectivos centros perspectivos das duas câmeras, uma vez que para grandes profundidades, estas retas são quase paralelas.

Como alternativa para o problema de mal-condicionamento citado acima, utiliza-se geralmente grandes bases (distância entre os centros perspectivos das câmeras), melhorando assim, a geometria da triangulação. No entanto, o emprego de grandes bases dificulta consideravelmente a resolução do Problema de Correspondência de maneira autônoma devido as maiores distorções relativas entre as imagens para tais configurações de câmeras.

Referências Bibliográficas

P.N. BELHUMEUR. A binocular stereo algorithm for reconstructing sloping, creased, and broken surfaces in the presence of half-occlusion. *Computer Vision*, 1993. Proceedings, Fourth International Conference on, April 1993. Page(s): 431 -438.

M. BERTERO, T. POGGIO and V. TORRE. Ill-posed Problems in Early Vision. *Proceedings of the IEEE*, 76, 869-889, 1988.

S. BIRCHFIELD & C. TOMASI. Depth discontinuities by pixel-to-pixel stereo. *ICCV* 1998.

M. BLEYER & M. GELAUTZ. Graph-based surface reconstruction from stereo pairs using image segmentation. *Proc. SPIE*, vol. 5665, January 2005.

Y. BOYKOV, O. VEKSLER and R. ZABIH. Fast approximate energy minimization via graph cuts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2001.

R. BROCKERS, M. HUND and B. MERTSCHING. A fast cost relaxation stereo algorithm with occlusion detection for mobile robot applications. *VMV* 2004.

M. Z. BROWN; D. BURSCHKA and G. D. HAGER. Advances in Computational Stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, august 2003, pp. 993 – 1007.

P. BURT & B. JULESZ. A Disparity Gradient Limit for Binocular Fusion. *Science*, 208:615-617, 1980.

C. CHANG, S. CHATTERJEE and P.R. KUBE. On an analysis of static occlusion in stereo vision. *Computer Vision and Pattern Recognition*, 1991. Proceedings CVPR 91, IEEE Computer Society Conference on, 1991 Page(s): 722 -723.

R.C. CHUNG & R. NEVATIA. Use of monocular groupings and occlusion analysis in a hierarchical stereo system. Computer Vision and Pattern Recognition, 1991. Proceedings CVPR 91, IEEE Computer Society Conference on, 1991 Page(s): 50 -56.

T. DARRELL and A. PENTLAND. Cooperative robust estimation using layers of support. IEEE Transactions on Pattern Analysis and Machine Intelligence, 17(5):474-487, 1995.

Y. DENG, Q. YANG, X. LIN and X. TANG. A symmetric patch-based correspondence model for occlusion handling. ICCV 2005.

U. R. DHOND and J. K. AGGARWAL. Structure from Stereo - A Review. IEEE Transactions on Systems, Man and Cybernetics, 19(6):1489-1510, 1989.

J. D. FOLEY, A. V. DAM., S. K. FEINER and, J. F. HUGHES. Computer Graphics: principles and practice. Addison Wesley Publishing Company, Reading, 1982.

L. R. FORD & D. R. FULKERSON. Flows in Networks. Princeton University Press, Princeton, NJ, 1962.

S. FORSTMANN, J. OHYA, Y. KANOU, A. SCHMITT and S. THUERING. Real-time stereo by using dynamic programming. Computer Vision and Pattern Recognition, 2004 Workshop on real-time 3D sensors and their use.

M. GONG and Y.-H. YANG. Multi-baseline stereo matching using genetic algorithm. Computer Vision and Pattern Recognition, 2001. Stereo Workshop / IJCV 2002.

M. HARIYAMA, T. TAKEUCHI and M. KAMEYAMA. VLSI processor for reliable stereo matching based on adaptive window-size selection. Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference on, Volume: 2, 2001 Page(s): 1168 -1173 vol.2.

L. HONG and G. CHEN. Segment-based stereo matching using graph cuts. *Computer Vision and Pattern Recognition*, 2004.

B.K.P. HORN. *Robot Vision*. MIT Press, Cambridge, MA, 1986.

H. ISHIKAWA and D. GEIGER. Occlusions, discontinuities, and epipolar lines in stereo. In *Proceedings of the European Conference on Computer Vision*, volume 1, pg. 232-249, 1998.

T. KANADE and M. OKUTOMI. A stereo matching algorithm with an adaptive window: theory and experiment. *Pattern Analysis and Machine Intelligence*, *IEEE Transactions on*, Volume: 16 Issue: 9, Sept. 1994. Page(s): 920 -932

S. B. KANG, R. SZELISKI and J. CHAI. Handling occlusions in dense multi-view stereo. *Computer Vision and Pattern Recognition*, 2001. *Computer Vision and Pattern Recognition*, 2001. *Proceedings of the 2001 IEEE Computer Society Conference on*, Volume: 1, 2001. Page(s): 103 -110

C. KIM, K.M. LEE, B.T. CHOI and S.U. LEE. A dense stereo matching using two-pass dynamic programming with generalized ground control points. *Computer Vision and Pattern Recognition*, 2005.

V. KOLMOGOROV and R. ZABIH. Computing visual correspondence with occlusions using graph cuts. *ICCV 2001*.

V. KOLMOGOROV and R. ZABIH. Multi-camera scene reconstruction via graph cuts. *ECCV 2002*.

S. H. LEE, Y. KANATSUGU and J.-I. PARK. Hierarchical stochastic diffusion for disparity estimation. Computer Vision and Pattern Recognition, 2001. Stereo Workshop / IJCV 2002.

M. LIN and C. TOMASI. Surfaces with occlusions from layered stereo. Ph.D. thesis, Stanford University, 2002.

J.-L. LOTTI and G. GIRAUDON. Adaptive window algorithm for aerial image stereo. Pattern Recognition, 1994. Vol. 1 - Conference A: Computer Vision & Image Processing., Proceedings of the 12th IAPR International Conference on , Volume: 1 , 1994. Page(s): 701 -703 vol.1.

D. MARR & T. POGGIO. Cooperative computation of stereo disparity. Science, 194:301-328, 1976.

H. MAYER. Analysis of means to improve cooperative disparity estimation. ISPRS Conf. on Photogrammetric Image Analysis, 2003.

K. MÜHLMANN, D. MAIER, J. HESSER and R. MÄNNER. Calculating dense disparity maps from color stereo images, an efficient implementation. CVPR 2001 Stereo Workshop / IJCV 2002.

Y. NAKAMURA, T. MATSUURA, K. SATOH and Y. OHTA. Occlusion detectable stereo-occlusion patterns in camera matrix. Computer Vision and Pattern Recognition, 1996. Proceedings CVPR '96, 1996 IEEE. Computer Society Conference on, 1996. Page(s): 371 -378.

A. NISHIKAWA, S. OGAWA, N. MARU and F. MIYAZAKI. Detecting object surfaces by using occlusion information from active binocular stereo. Robotics and Automation, 1995. Proceedings, 1995 IEEE International Conference on, Volume: 3, 1995 Page(s): 2974 -2981 vol.3

F.M.A. NOGUEIRA. Geração Automática de Mapas de Disparidade em Visão Estéreo. Dissertação de Mestrado. FEEC/UNICAMP, Campinas, 1998.

Y. OHTA and T. KANADE. Stereo by intra- and inter-scanline search using dynamic programming. IEEE Transactions on Pattern Analysis and Machine Intelligence, 7(2):139-154, 1985.

Y. OHTA, K. SATOH, I. KITAHARA and T. MATSUURA. Displaying motion parallax by occlusion detectable stereo. Computer Vision for Virtual Reality Based Human Communications, 1998. Proceedings, 1998 IEEE and ATR Workshop on, 1998 Page(s): 35 -42.

T. POGGIO & C. KOCH. Ill-posed Problems in Early Vision: From Computational Theory to Analog Networks. Proceedings of the Royal Society London B, 226, 303-323, 1985.

T. POGGIO, V. TORRE and C. KOCH. Computational vision and regularization theory. Nature, 317:314-319, 1985.

S. B. POLLARD, J. PORRILL, J. E.W. MAYHEW, AND J. P. FRISBY. Disparity gradient, Lipschitz continuity and computing binocular correspondences. in Proc. 3rd Int. Symp. Robotics Research, 1986, pp. 19–26.

A. REDERT, E. HENDRIKS and J. BIEMOND. Correspondence estimation in image pairs. IEEE Signal Processing Magazine, Volume: 16 Issue: 3, May 1999 Page(s): 29-46.

S. ROY & I. J. COX. A maximum-flow formulation of the N-camera stereo correspondence problem. ICCV 1998.

K. SATOH, Y. OHTA. Occlusion detectable stereo-systematic comparison of detection algorithms. Pattern Recognition, 1996. Proceedings of the 13th International Conference on, Volume: 1 , 1996. Page(s): 280-286 vol.1

D. SCHARSTEIN & R. SZELISKI. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms, IJCV, 2002.

J. SHAH. A nonlinear diffusion model for discontinuous disparity and half-occlusions in stereo. Computer Vision and Pattern Recognition, 1993. Proceedings on Computer Vision and Pattern Recognition, 1993. Page(s): 34 -40.

J. SHAO. Combination of stereo, motion and rendering for 3D footage display. Computer Vision and Pattern Recognition, 2001 Stereo Workshop / IJCV 2002.

Y. SUGAYA and Y. OHTA. Stereo by integration of two algorithms with/without occlusion handling. Pattern Recognition, 2000. Proceedings. 15th International Conference on, Volume: 1, Page(s): 109 -113 vol.1

C. SUN. Fast stereo matching using rectangular subregioning and 3D maximum-surface techniques. Computer Vision and Pattern Recognition, 2002. Stereo Workshop / IJCV 2002.

D. TERZOPOULOS. Regularization of inverse visual problems involving discontinuities. IEEE Transactions on Pattern Analysis and Machine Intelligence, 8(4):413-424, 1986.

R. Y. TSAI. An efficient and Accurate Camera Calibration Technique for 3-D Machine Vision. Proceedings, IEEE conference on Computer Vision and Pattern Recognition, 1:364-374, 1986.

G.A.TRIANTAFYLLIDIS, D. TZOVARAS and M.G. STRINTZIS. Occlusion and visible background and foreground areas in stereo: a Bayesian approach. Circuits and Systems for

Video Technology, IEEE Transactions on, Volume: 10 Issue: 4, June 2000. Page(s): 563 - 575.

G.A.TRIANTAFYLLIDIS, D. TZOVARAS and M.G. STRINTZIS. A compound Bayes test for detection of occlusion and visible background and foreground areas in stereo. Digital and Computational Video, 2001. Proceedings. Second International Workshop on, 2001. Page(s): 102 -109.

O. VEKSLER. Stereo matching by compact windows via minimum ratio cycle. ICCV 2001.

O. VEKSLER. Fast variable window for stereo correspondence using integral images. Computer Vision and Pattern Recognition, 2003.

O. VEKSLER. Stereo correspondence by dynamic programming on a tree. Computer Vision and Pattern Recognition, 2005.

W. WOO & A. ORTEGA. Overlapped block disparity compensation with adaptive windows for stereo image coding. Circuits and Systems for Video Technology, IEEE Transactions on, Volume: 10 Issue: 2, March 2000. Page(s): 194 -200.

L. ZITNICK & T. KANADE. A cooperative algorithm for stereo matching and occlusion detection. Pattern Analysis and Machine Intelligence, IEEE Transactions on Vol. 22 Issue: 7, July 2000. Page(s): 675 -684.