

**MAURICIO EDGAR STIVANELLO**

**DESENVOLVIMENTO DE UMA BIBLIOTECA PARA  
SISTEMAS DE VISÃO ESTEREOOSCÓPICA PARA  
ROBÓTICA MÓVEL**

**FLORIANÓPOLIS**

**2008**

**UNIVERSIDADE FEDERAL DE SANTA CATARINA**

**PROGRAMA DE PÓS-GRADUAÇÃO  
EM ENGENHARIA ELÉTRICA**

**DESENVOLVIMENTO DE UMA BIBLIOTECA PARA  
SISTEMAS DE VISÃO ESTEREOOSCÓPICA PARA  
ROBÓTICA MÓVEL**

Dissertação submetida à  
Universidade Federal de Santa Catarina  
como parte dos requisitos para a  
obtenção do grau de Mestre em Engenharia Elétrica.

**MAURICIO EDGAR STIVANELLO**

Florianópolis, fevereiro de 2008.

# **DESENVOLVIMENTO DE UMA BIBLIOTECA PARA SISTEMAS DE VISÃO ESTEREOOSCÓPICA PARA ROBÓTICA MÓVEL**

Mauricio Edgar Stivanello

‘Esta Dissertação foi julgada adequada para obtenção do título de Mestre em Engenharia Elétrica, Área de Concentração em Controle, Automação e Informática Industrial, e aprovada em sua forma final pelo curso de Pós-Graduação.’

---

Prof. Marcelo Ricardo Stemmer, Dr. Ing.  
Orientador

---

Prof<sup>a</sup>. Kátia Campos de Almeida  
Coordenador do Programa de Pós-Graduação em Engenharia Elétrica

Banca Examinadora:

---

Prof. Marcelo Ricardo Stemmer, Dr. Ing.  
Presidente

---

Prof<sup>a</sup>. Analucia Vieira Fantin, Dra.

---

Prof. Armando Albertazzi Gonçalves Jr., Dr. Ing.

---

Prof. Celso Peres Fernandes, Dr. Ing.

---

Prof. Joceli Mayer, Ph.D.

*À Leila Daiana Klock,  
Pelo Amor, Paciência e Companheirismo.*

## **Agradecimentos**

Ao Professor Marcelo Ricardo Stemmer, pela orientação, confiança e apoio na realização deste e de outros trabalhos.

Aos colegas do Grupo S2I, em especial ao Eduardo S. Leal e ao Nicolas Palluat, pelo companheirismo e pela ajuda prestada durante o desenvolvimento deste projeto.

Aos colegas e professores dos cursos PGEEL e PGEAS da UFSC, pela troca de conhecimentos e experiências.

Ao CNPq, pela bolsa de estudos que possibilitou a realização deste trabalho.

Resumo da Dissertação apresentada à UFSC como parte dos requisitos necessários para obtenção do grau de Mestre em Engenharia Elétrica.

# **DESENVOLVIMENTO DE UMA BIBLIOTECA PARA SISTEMAS DE VISÃO ESTEREOOSCÓPICA PARA ROBÓTICA MÓVEL**

**Mauricio Edgar Stivanello**

Fevereiro/2008

Orientador: Marcelo Ricardo Stemmer, Dr. Ing.

Área de Concentração: Controle, Automação e Informática Industrial.

Palavras-chave: Visão estéreo, Estereoscopia, Robótica móvel, Visão computacional, Calibração, Reconstrução, Mapas de disparidade, Correspondência.

Número de Páginas: 116

A demanda por aplicações de robótica móvel vem crescendo consideravelmente nos últimos anos. Independente da natureza ou fim, robôs autônomos móveis devem interagir com o mundo para alcançar seus objetivos. Para isto, de alguma maneira precisam obter informações a respeito do ambiente. Dentre diferentes abordagens existentes, bons resultados têm sido alcançados pelo emprego de sistemas de visão estereoscópica. Através de um sistema como este é possível extrair informação tridimensional do meio em que o robô está inserido. A informação tridimensional pode então ser usada para orientar as ações do robô, seja para navegação, reconhecimento ou manipulação. No presente trabalho é apresentado o desenvolvimento de uma biblioteca para sistemas de visão estereoscópica para robótica móvel. Para tal, foram tratados diferentes problemas da estereoscopia, procurando prover mapas de profundidade com detalhes do ambiente suficientes para a operação geral de um robô móvel que tenha um sistema de visão estereoscópica como principal fonte de informação. Neste contexto, são avaliados e propostos modelos, métodos e soluções para diferentes problemas como calibração de câmeras, retificação de imagens, reconstrução e, principalmente, geração de mapas de profundidade densos. Os resultados obtidos demonstram a efetividade da utilização da infra-estrutura disponibilizada e dos métodos propostos no desenvolvimento de aplicações para robótica móvel.

Abstract of Dissertation presented to UFSC as a partial fulfillment of the requirements for the degree of Master in Electrical Engineering.

## **DEVELOPMENT OF A LIBRARY FOR STEREOSCOPIC VISION SYSTEMS FOR MOBILE ROBOTICS**

**Mauricio Edgar Stivanello**

February/2008

Advisor: Marcelo Ricardo Stemmer, Dr. Ing.

Area of Concentration: Control, Automation and Industrial Computing.

Keywords: Stereo vision, Stereoscopy, Mobile robots, Computer vision, Calibration, Reconstruction, Disparity maps, Stereo correspondence.

Number of Pages: 116.

The demand for mobile robotics applications has grown considerably in recent years. Regardless of the nature or purpose, autonomous mobile robots must interact with the world to achieve their goals. To do that, information about the environment must be retrieved. Among different approaches, good results have been achieved using stereoscopic vision systems. These systems allow extracting three-dimensional information of the environment in which the robot is inserted. Then, the three-dimensional information can be used to guide the robot actions. In this paper, the development of a library that includes basic functions used on configuration and operation of stereoscopic vision systems for mobile robotics is described. In this context, methods and solutions related to stereoscopy are evaluated and proposed. Results demonstrate the library effectiveness on the development of mobile robotics applications, and the benefits provided by the proposed methods.

# Sumário

<b>1 INTRODUÇÃO .....</b>	<b>1</b>
1.1 Motivação.....	2
1.2 Objetivos e Contribuições.....	3
1.3 Estrutura do trabalho.....	3
<b>2 ESTEREOSCOPIA.....</b>	<b>5</b>
2.1 Modelo de Câmera.....	6
2.1.1 Parâmetros Intrínsecos .....	8
2.1.2 Parâmetros Extrínsecos.....	9
2.1.3 Projeção Mundo-Imagem .....	11
2.1.4 Calibração Monocular .....	12
2.2 Sistema de Visão Estereoscópica .....	15
2.2.1 Parâmetros do Sistema de Visão Estéreo .....	18
2.2.2 Geometria Epipolar.....	19
2.3 Calibração Estereoscópica .....	24
2.3.1 Calibração Clássica.....	25
2.3.2 Calibração Projetiva.....	25
2.4 Retificação.....	27
2.5 Correspondência.....	30
2.5.1 Correspondência Densa .....	31
2.5.1.1 Abordagem Local .....	32
2.5.1.2 Abordagem Global .....	38
2.5.1.3 Outras Abordagens .....	39
2.5.2 Correspondência Esparsa.....	41
2.6 Reconstrução .....	43
2.7 Considerações Finais .....	46
<b>3 DESENVOLVIMENTO DE UMA BIBLIOTECA PARA SISTEMAS DE VISÃO ESTEREOSCÓPICA PARA ROBÓTICA MÓVEL.....</b>	<b>47</b>

3.1	Especificação e Análise dos Requisitos .....	47
3.2	Aparato Experimental e Ambiente de Desenvolvimento .....	50
3.3	Funcionalidades Principais .....	55
3.3.1	Modelo de Câmera e de Sistema de Visão Estéreo .....	55
3.3.2	Calibração do Sistema de Visão Estéreo.....	55
3.3.3	Retificação do Par de Imagens Estéreo .....	61
3.3.4	Geração de Mapas de Disparidade .....	62
3.3.4.1	Algoritmos Baseados na Literatura.....	63
3.3.4.2	Método Proposto: Correspondência Local com Suporte Regional.....	66
3.3.4.3	Resultados Experimentais .....	71
3.3.5	Reconstrução.....	80
3.4	Especificação do Sistema.....	82
3.5	Utilização da Biblioteca.....	85
<b>4</b>	<b>CONCLUSÕES E PERSPECTIVAS .....</b>	<b>89</b>
4.1	Conclusões .....	89
4.2	Perspectivas Futuras.....	90
<b>ANEXO 1 – ARTIGOS PRODUZIDOS.....</b>		<b>92</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>		<b>113</b>

# Listas de Figuras

<b>Figura 1:</b> Geometria presente no modelo de câmera <i>pinhole</i> .....	6
<b>Figura 2:</b> Similaridade de triângulos na projeção de coordenadas.....	7
<b>Figura 3:</b> Transformação entre sistemas de coordenadas do mundo e da câmera.....	10
<b>Figura 4:</b> Modelo simplificado de sistema de visão estéreo.....	16
<b>Figura 5:</b> Parâmetros extrínsecos e intrínsecos do sistema de visão estéreo .....	18
<b>Figura 6:</b> Geometria epipolar presente em sistema de visão estéreo .....	20
<b>Figura 7:</b> Transformação da configuração estéreo pela retificação .....	28
<b>Figura 8:</b> Mapa de disparidade calculado por correspondência densa.....	32
<b>Figura 9:</b> Processo de correspondência por correlação de janelas .....	34
<b>Figura 10:</b> Normalização por LOG .....	35
<b>Figura 11:</b> Descrição baseada em intensidade .....	36
<b>Figura 12:</b> Imagens processadas por filtros extratores de características .....	42
<b>Figura 13:</b> Correspondência esparsa por SIFT .....	43
<b>Figura 14:</b> Reconstrução projetiva e métrica ( Fonte: HARTLEY e ZISSEMAN [14] ) ..	44
<b>Figura 15:</b> Reconstrução de ponto por triangulação .....	45
<b>Figura 16:</b> Sistema binocular de aquisição de imagens .....	50
<b>Figura 17:</b> Imagens da base de referência e respectivo mapa de disparidade real .....	53
<b>Figura 18:</b> Ambiente de simulação para robótica móvel.....	54
<b>Figura 19:</b> Seqüência de amostras do gabarito de calibração.....	57
<b>Figura 20:</b> Cálculo das distâncias entre linhas epipolares e pontos .....	60
<b>Figura 21:</b> Retificação do par de imagens estéreo .....	61
<b>Figura 22:</b> Verificação de consistência left-right check .....	65
<b>Figura 23:</b> Descrição baseada em intensidade com suporte regional .....	67
<b>Figura 24:</b> Cálculo dos descritores regionais .....	69
<b>Figura 25:</b> Diferença entre mapas calculados e mapas de referência .....	73
<b>Figura 26:</b> Correspondência em regiões homogêneas .....	75
<b>Figura 27:</b> Diferença entre os mapas calculados por janelas simples e múltiplas .....	76
<b>Figura 28:</b> Pares estéreo próprios capturados com sistema de aquisição montado .....	77
<b>Figura 29:</b> Mapa de disparidade obtido com par estéreo não normalizado .....	78
<b>Figura 30:</b> Mapas de disparidade calculados com pares próprios.....	79
<b>Figura 31:</b> Reconstrução do cenário de calibração.....	80
<b>Figura 32:</b> Reconstrução de mapa de disparidade .....	81
<b>Figura 33:</b> Diagrama de classes.....	83
<b>Figura 34:</b> Diagrama de seqüência.....	84
<b>Figura 35:</b> Detecção de obstáculos.....	86
<b>Figura 36:</b> Detecção de obstáculos e pavimento em simulação .....	87

# Listas de Tabelas

<b>Tabela 1:</b> Especificação das câmeras utilizadas no sistema binocular .....	51
<b>Tabela 2:</b> Especificação das lentes utilizadas no sistema binocular .....	52
<b>Tabela 3:</b> Especificação da placa de aquisição utilizada no sistema binocular .....	52
<b>Tabela 4:</b> Parâmetros dos pares estéreo de referência .....	54
<b>Tabela 5:</b> Cenários de calibração.....	59
<b>Tabela 6:</b> Parâmetros do sistema de visão estereoscópica obtidos nos diferentes cenários	59
<b>Tabela 7:</b> Resultado da avaliação da qualidade da calibração.....	60
<b>Tabela 8:</b> Tempos de processamento (s) obtidos com algoritmos de janela.....	64
<b>Tabela 9:</b> Algoritmos de correspondência baseados na literatura .....	66
<b>Tabela 10:</b> Variações de algoritmos propostos .....	71
<b>Tabela 11:</b> Algoritmos avaliados.....	72
<b>Tabela 12:</b> Erros obtidos no cálculo das disparidades .....	74
<b>Tabela 13:</b> Tempos de processamento para geração dos mapas de disparidade.....	76
<b>Tabela 14:</b> Resultado da avaliação da qualidade da reconstrução.....	82

# List of Symbols and Abbreviations

$\pi$	Plano de imagem da câmera.
$C$	Centro de projeção da câmera.
$f$	<i>Distância focal</i> , dada pela distância entre $\pi$ e $C$ .
$c$	<i>Ponto principal</i> , dado pela intersecção entre o <i>eixo óptico</i> e $\pi$ .
$M$	Coordenadas de um ponto 3-D no <i>sistema de coordenadas da câmera</i> .
$W$	Coordenadas de um ponto 3-D no <i>sistema de coordenadas do mundo</i> .
$\tilde{W}$	Coordenadas homogêneas de um ponto 3-D no <i>sistema de coordenadas do mundo</i> .
$m$	Coordenadas da projeção de um ponto 3-D no plano $\pi$ .
$\tilde{m}$	Coordenadas homogêneas da projeção de um ponto 3-D no plano $\pi$ .
$(s_x, s_y)$	Tamanho efetivo do <i>pixel</i> em milímetros na horizontal e na vertical.
$(k_1, k_2)$	Coeficientes de distorção radial.
$t$	Vetor de translação que descreve a posição relativa entre as origens de dois sistemas de coordenadas.
$R$	Matriz de rotação $3 \times 3$ que descreve a orientação relativa entre os eixos de dois sistemas de coordenadas.
$A$	Matriz $3 \times 3$ que contém os parâmetros intrínsecos de uma câmera.
$P$	Matriz de projeção $3 \times 4$ que contém tanto parâmetros intrínsecos como parâmetros extrínsecos.
$e$	Epipolo.
$E$	Matriz Essencial.
$F$	Matriz Fundamental
$SSD$	Soma das diferenças quadráticas, do inglês <i>Sum of squared differences</i> .
$SAD$	Soma das diferenças absolutas, do inglês <i>Sum of absolute differences</i> .
$WTA$	Abordagem de seleção empregada em algoritmos de correspondência, do inglês <i>Winner-Takes-All</i> .
$SIFT$	Transformada de característica invariante à escala, do inglês <i>Scale-Invariant Feature Transform</i> .

# Capítulo 1

## Introdução

Já se passou o tempo em que a palavra robô lembrava manipuladores mecânicos inseridos em linhas de produção, realizando tarefas muito específicas e repetitivas. A robótica móvel quebrou este paradigma e hoje temos as aplicações mais diversas, com robôs das mais variadas formas e com diferentes finalidades. Esta realidade foi possível graças aos resultados obtidos por um grande número de pesquisas na área [1,2,3], pelo surgimento de ferramentas de desenvolvimento específicas [4] e pela popularização das técnicas e barateamento dos dispositivos que a viabilizam.

Automóveis autônomos começam a circular tanto em desertos como em cidades [5]. A indústria bélica investe pesado em projetos de pesquisa para criar exércitos de robôs [6]. Inúmeros robôs humanóides têm sido desenvolvidos [7,8], com funções que vão de serviços domésticos [9] ao entretenimento. Os primeiros produtos comercializáveis também começam a aparecer, impulsionando o interesse pela área e financiando novos projetos.

Apesar dos consideráveis avanços, temos que a capacidade destes exemplares ainda está distante das descrições fantásticas feitas pelos primeiros visionários a considerar sua existência [10]. Estes robôs ainda se limitam a realizar certo número de tarefas bem determinadas. A flexibilidade e autonomia continua sendo um dos objetivos mais almejados pelos pesquisadores da área.

Independente da natureza ou fim, assim como nós, robôs autônomos móveis devem interagir com o mundo para alcançar seus objetivos. Para isto, de alguma maneira devem extrair informações a respeito do ambiente onde estão inseridos. Diferentes abordagens são utilizadas para este fim. Sensores como sonares, lasers e câmeras são empregados com o objetivo de extrair uma descrição do ambiente e permitir a interação do robô com elementos presentes no mesmo [11].

Seguindo o sistema visual humano, muitos sistemas robóticos utilizam um conjunto binocular composto por câmeras, caracterizando um sistema de visão estereoscópica [12]. Através de um sistema como este é possível extrair informação tridimensional do meio em que o robô está inserido. A informação tridimensional obtida pode ser utilizada para guiar as ações do robô, necessárias para alcançar o objetivo específico do sistema.

Os sistemas de visão estereoscópica podem ser vistos como sendo dispositivos de captura cuja saída é um mapa de profundidade da cena observada. Este mapa, que pode ser denso ou esparsos, é formado pela avaliação do deslocamento relativo na posição de elementos da cena, observados sob os pontos de vista de cada câmera que compõe o sistema.

A geração dos mapas de profundidade, por sua vez, nos leva a considerar problemas da estereoscopia como calibração, correspondência e reconstrução. A calibração refere-se à determinação dos parâmetros que descrevem o sistema binocular utilizado. A correspondência, considerada por muitos como o principal problema da estereoscopia, está relacionada ao processo de corresponder elementos presentes nas imagens capturadas pelo sistema binocular. Finalmente, através da reconstrução recupera-se a informação de profundidade com base nos parâmetros obtidos na etapa de calibração e nos pares de pontos correspondentes obtidos na etapa de correspondência. Existem diferentes soluções para cada um destes problemas, baseadas em diferentes modelos de sistema de visão estereoscópica e propostas para diferentes aplicações.

No presente trabalho foram pesquisadas, avaliadas e implementadas diferentes técnicas no sentido de tratar cada um destes problemas e criar uma infra-estrutura básica para o desenvolvimento de sistemas de visão estereoscópica para robótica móvel.

## 1.1 Motivação

A demanda por aplicações de robótica móvel tem apresentado um grande crescimento nos últimos tempos. Por sua vez, a utilização de sistemas de visão estereoscópica como fonte de informação primária tem sido muito empregada e torna-se cada vez mais viável na medida em que novos trabalhos são desenvolvidos na área.

Entretanto, o estudo preliminar e a implementação das etapas presentes na estereoscopia continuam a representar um alto custo para os projetos que a empregam.

Os requisitos impostos aos sistemas de visão estereoscópica em aplicações de robótica móvel são muitas vezes parecidos, visto que tratam problemas comuns e apresentam objetivos similares. O compartilhamento de uma infra-estrutura básica pode colaborar para o desenvolvimento destes sistemas, visto que cada projeto poderia focar o trabalho nas funções específicas do sistema, seja de navegação, reconhecimento ou manipulação.

Os benefícios a serem obtidos com a disponibilização de uma biblioteca que encapsule as funções básicas para configuração e operação de um sistema de visão estereoscópica para robótica móvel motivaram e justificaram a realização do presente trabalho.

## 1.2 Objetivos e Contribuições

O objetivo geral desta dissertação de mestrado foi estudar cientificamente, avaliar e desenvolver técnicas e algoritmos relacionados à estereoscopia. O objetivo específico do trabalho foi disponibilizar uma biblioteca encapsulando as funcionalidades básicas necessárias à configuração e operação de um sistema de visão estereoscópica para robótica móvel.

Neste contexto, foram levantados métodos e soluções para diferentes problemas pertinentes à utilização de um sistema de visão estereoscópica. Adicionalmente, foi proposto um novo método de correspondência densa, que resultou em uma considerável melhoria nos resultados obtidos por algoritmos clássicos de correspondência local. O conjunto de técnicas implementadas foi disponibilizado na forma de uma biblioteca.

## 1.3 Estrutura do trabalho

No capítulo 2 são abordados os principais tópicos relacionados à estereoscopia, onde são tratados os modelos, técnicas, problemas e soluções empregadas no desenvolvimento deste trabalho. No Capítulo 3 é descrito o desenvolvimento de uma biblioteca para sistemas de visão estereoscópica, onde são apresentadas as soluções

empregadas e os resultados obtidos para cada um dos problemas tratados. No capítulo 4 são expostos os resultados gerais e conclusões obtidas pelo desenvolvimento do presente trabalho, assim como sugestões para trabalhos futuros.

## Capítulo 2

# Estereoscopia

A extração de informação tridimensional de uma cena através do processamento de imagens pode ser baseada em diferentes características como sombreamento, textura, contornos ou movimento, dentre outras. A estereoscopia, que recupera a profundidade avaliando as projeções de um mesmo ponto no espaço obtidas sob diferentes pontos de vista, tem apresentado resultados superiores em relação a outras abordagens dentro da robótica móvel [13].

De maneira geral, um sistema de visão estereoscópica precisa tratar três problemas principais: calibração, correspondência e reconstrução. Na calibração procura-se determinar os parâmetros que descrevem o sistema de aquisição utilizado. O problema da correspondência consiste em determinar qual elemento na imagem capturada sob um ponto de vista corresponde a um dado elemento na imagem capturada sob outro ponto de vista. Por sua vez, através da reconstrução procura-se recuperar a informação de profundidade com base nos parâmetros obtidos na etapa de calibração e nos pares de pontos correspondentes obtidos na etapa de correspondência.

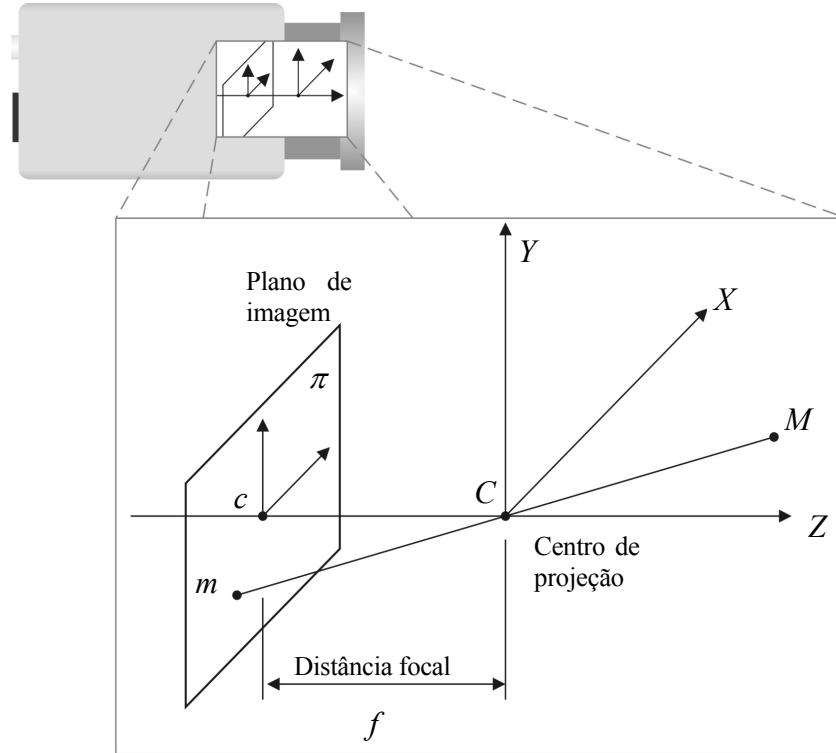
No presente trabalho, é considerado um sistema de visão estereoscópica baseado em um sistema de aquisição binocular. Entretanto, alguns sistemas utilizam um sistema de aquisição monocular, onde são capturadas seqüências de imagens em diferentes instantes durante o deslocamento do robô para prover projeções da cena sob diferentes pontos de vista. Outros sistemas empregam ainda espelhos em conjunto com um sistema de aquisição monocular, de maneira a obterem diferentes projeções em uma única captura de imagem.

O presente capítulo apresenta modelos, técnicas e soluções relacionadas aos problemas da estereoscopia. Inicialmente serão apresentados os modelos de câmera e de sistema de visão estereoscópica utilizados. Em seguida serão apresentadas considerações e soluções relacionadas aos problemas levantados.

## 2.1 Modelo de Câmera

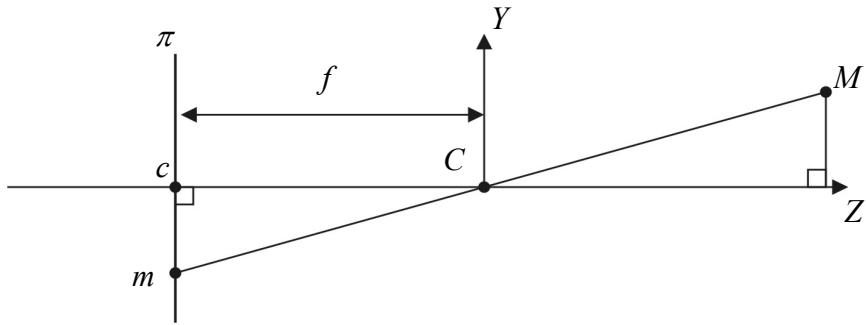
No presente trabalho é utilizado um modelo de sistema de visão estereoscópica composto por duas câmeras *pinhole* [12,14]. Conforme ilustrado na Figura 1, o modelo de câmera *pinhole* é composto por um plano de imagem  $\pi$  e por um ponto no espaço  $C$  que corresponde ao centro de projeção da câmera. A distância entre  $\pi$  e  $C$  é denominada *distância focal* ou simplesmente  $f$ . A reta que passa por  $C$  e é perpendicular ao plano  $\pi$  é denominada como *eixo óptico*, e a intersecção  $c$  entre  $\pi$  e o *eixo óptico* é denominada como *ponto principal*. O sistema de coordenadas 3-D cuja origem é  $C$  e onde o plano  $\pi$  é ortogonal ao eixo  $Z$  é conhecido como *sistema de coordenadas da câmera*.

Conforme ilustrado, um ponto no espaço  $M = [X, Y, Z]^T$  é mapeado ao ponto  $m$  no plano  $\pi$  onde a reta através de  $M$  e  $C$  intersecta o plano  $\pi$ .



**Figura 1:** Geometria presente no modelo de câmera *pinhole*

Pela similaridade de triângulos exibida na Figura 2, podemos definir o mapeamento  $[fX/Z, fY/Z, f]^T$ , que descreve a projeção central de coordenadas do mundo para coordenadas no plano de imagem.



**Figura 2:** Similaridade de triângulos na projeção de coordenadas

Ignorando a última coordenada, podemos descrever o mapeamento do espaço euclidiano 3-D para o espaço euclidiano 2-D como:

$$m = \begin{bmatrix} f \frac{X}{Z} & f \frac{Y}{Z} \end{bmatrix}^T. \quad (1)$$

Em diferentes aplicações, algoritmos de visão computacional utilizados na recuperação de informação da cena observada precisam relacionar as coordenadas dos pontos 3-D no ambiente com as coordenadas de sua projeção no plano da imagem. As equações acima apresentadas descrevem estas relações com base no *sistema de coordenadas da câmera*.

Muitas vezes, entretanto, o sistema de coordenadas de interesse é diferente ao da câmera. O sistema de coordenadas de interesse pode ter a origem definida por um marco do ambiente e apresentar qualquer orientação, sendo assim denominado como *sistema de coordenadas do mundo*.

Além disso, nas equações da projeção de perspectiva apresentadas os pontos da imagem são expressos em unidades métricas. É essencial que as coordenadas dos pontos da imagem no *sistema de coordenadas da câmera* possam ser recuperadas a partir das coordenadas inteiras em *pixel*, já que são as únicas diretamente disponíveis a partir da

imagem. Considerando as coordenadas em *pixel* disponíveis na imagem podemos definir um novo sistema de coordenadas denominado *sistema de coordenadas da imagem*.

Por estas considerações, faz-se necessário conhecer algumas das características e relacionamentos das câmeras utilizadas, denominadas como parâmetros intrínsecos e parâmetros extrínsecos.

Nas seções 2.1.1, 2.1.2 e 2.1.3 são apresentadas as equações básicas que descrevem os parâmetros intrínsecos e extrínsecos e a transformação de coordenadas entre os diferentes sistemas de coordenadas. O problema de estimar os valores destes parâmetros é denominado calibração de câmera, e será tratado na seção 2.1.4.

### 2.1.1 Parâmetros Intrínsecos

Os parâmetros intrínsecos relacionam as coordenadas em *pixel* de um ponto no *sistema de coordenadas da imagem* com as correspondentes coordenadas no *sistema de coordenadas da câmera*. Estes parâmetros descrevem as características ópticas, geométricas e digitais da câmera utilizada. Para o modelo de câmera *pinhole*, podemos considerar três conjuntos de parâmetros intrínsecos principais.

O primeiro conjunto é formado apenas pelo parâmetro  $f$ , ou *distância focal*, introduzida na descrição da projeção de perspectiva.

O segundo conjunto de parâmetros é diretamente empregado na transformação entre coordenadas do *sistema de coordenadas da câmera* e coordenadas em *pixel* do *sistema de coordenadas da imagem*. Esta transformação é descrita pelas equações ( 2 ) e ( 3 ), onde são ignoradas possíveis distorções geométricas introduzidas pelo sistema óptico. Nestas equações,  $(c_x, c_y)$  representam as coordenadas do *ponto principal* no *sistema de coordenadas da imagem*,  $(s_x, s_y)$  representam o tamanho efetivo do *pixel* em milímetros na horizontal e na vertical e  $(x_{img}, y_{img})$  são as coordenadas de um ponto no *sistema de coordenadas da imagem* correspondendo às coordenadas  $(x_{cam}, y_{cam})$  de um ponto no *sistema de coordenadas da câmera*.

$$x_{cam} = -(x_{img} - c_x)s_x \quad (2)$$

$$y_{cam} = -(y_{img} - c_y)s_y \quad (3)$$

A mudança de sinal presente nas equações acima se deve ao fato de que os eixos horizontal e vertical do *sistema de coordenadas da câmera* e do *sistema de coordenadas da imagem* possuem orientação oposta.

O terceiro conjunto de parâmetros intrínsecos está relacionado com as distorções geométricas introduzidas pelo conjunto óptico utilizado. Em função da qualidade das lentes utilizadas, o conjunto óptico pode introduzir distorções que geralmente são evidenciadas na região periférica da imagem. Estas distorções podem ser modeladas com certa precisão como simples distorções radiais, de acordo com as relações descritas pelas equações (4) e (5), onde  $(x_d, y_d)$  são as coordenadas dos pontos distorcidos, onde  $r^2 = x_d^2 + y_d^2$  e onde  $k_1$  e  $k_2$  correspondem aos coeficientes de distorção.

$$x_{cam} = x_d(1 + k_1r^2 + k_2r^4) \quad (4)$$

$$y_{cam} = y_d(1 + k_1r^2 + k_2r^4) \quad (5)$$

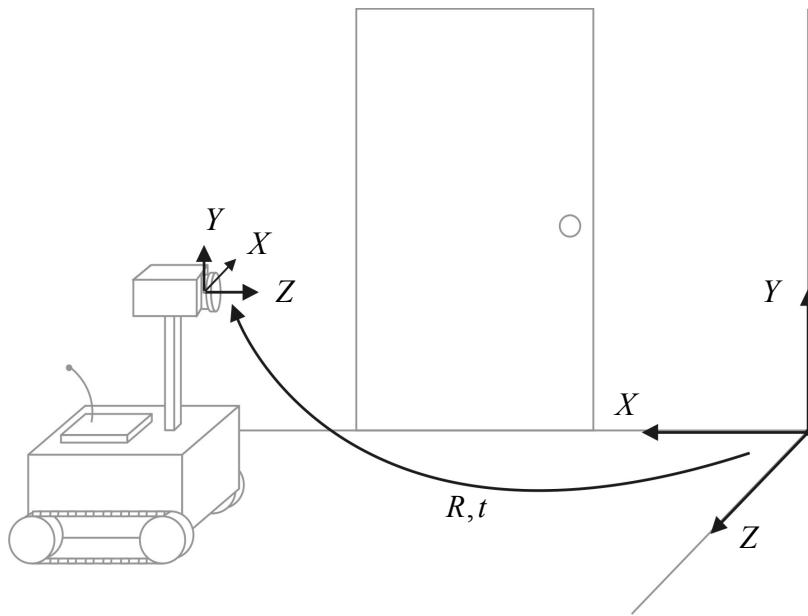
Nestas equações, a distorção consiste em um deslocamento radial nos pontos da imagem. Este deslocamento é nulo no centro da imagem, e aumenta com o distanciamento do centro da imagem. Observa-se que como os valores de  $k_1$  e  $k_2$  são normalmente muito pequenos, a distorção radial pode ser ignorada quando alta precisão não é necessária em todas as regiões da imagem.

Finalmente, temos que o conjunto de parâmetros intrínsecos é formado por  $f$ ,  $c_x$ ,  $c_y$ ,  $s_x$ ,  $s_y$ ,  $k_1$  e  $k_2$ .

### 2.1.2 Parâmetros Extrínsecos

Considerando apenas o *sistema de coordenadas da câmera* podemos escrever as equações fundamentais da projeção perspectiva de maneira simplificada, conforme apresentado na equação (1). Porém, o *sistema de coordenadas da câmera* é muitas vezes

desconhecido, e um problema comum é determinar a sua localização e orientação em relação a outro sistema de coordenadas utilizando somente informações da imagem. Os parâmetros extrínsecos são definidos como o conjunto de parâmetros geométricos que descrevem a transformação entre o *sistema de coordenadas da câmera* e o *sistema de coordenadas do mundo* conhecido.



**Figura 3:** Transformação entre sistemas de coordenadas do mundo e da câmera

Esta transformação é especificada através de um vetor de translação  $t$  e uma matriz de rotação  $R$ . O vetor de translação  $t$  descreve a posição relativa entre as origens dos dois sistemas de coordenadas. A matriz  $3 \times 3$  de rotação  $R$  descreve a orientação relativa entre os eixos dos dois sistemas de coordenadas. Esta transformação é ilustrada na Figura 3.

Assim sendo, a relação entre as coordenadas de um ponto no *sistema de coordenadas do mundo* e no *sistema de coordenadas da câmera*,  $W$  e  $M$  respectivamente, é dada por:

$$M = R(W - t). \quad (6)$$

### 2.1.3 Projeção Mundo-Imagem

A partir dos parâmetros e relacionamentos do modelo de câmera *pinhole* apresentados podemos descrever as relações que permitem o mapeamento direto as coordenadas de um ponto 3-D no ambiente com as coordenadas em *pixel* da projeção deste ponto na imagem, sem explicitamente referenciar o *sistema de coordenadas da câmera*.

Já temos definidos os dois passos necessários para a projeção mundo-imagem. Com os parâmetros extrínsecos podemos realizar a transformação entre o *sistema de coordenadas do mundo* e o *sistema de coordenadas da câmera*. Com os parâmetros intrínsecos, podemos realizar a transformação entre o *sistema de coordenadas da câmera* e o *sistema de coordenadas da imagem*. Ignorando a ocorrência de possíveis distorções radiais e substituindo as equações (2), (3) e (6) na equação (1) obtemos as equações (7) e (8), que relacionam as coordenadas de um ponto 3-D no mundo com as coordenadas do ponto correspondente na imagem, onde  $R_n, n = 1, 2, 3$  é um vetor 3-D formado pela enésima linha da matriz  $R$ .

$$-(x_{im} - c_x)s_x = f \frac{R_1^T(W-t)}{R_3^T(W-t)} \quad (7)$$

$$-(y_{im} - c_y)s_y = f \frac{R_2^T(W-t)}{R_3^T(W-t)} \quad (8)$$

Embora possamos usar as equações acima apresentadas diretamente, é conveniente expressá-las na forma de produto entre matrizes lineares. Para isso, podemos definir as matrizes apresentadas em (9) e (10), onde  $A$  depende somente de parâmetros intrínsecos e  $[R | t]$  depende somente de parâmetros extrínsecos.

$$A = \begin{bmatrix} -f/s_x & 0 & c_x \\ 0 & -f/s_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (9)$$

$$[R | t] = \begin{bmatrix} r_{11} & r_{12} & r_{13} & -R_1^T t \\ r_{21} & r_{22} & r_{23} & -R_2^T t \\ r_{31} & r_{32} & r_{33} & -R_3^T t \end{bmatrix} \quad (10)$$

A equação matricial linear que descreve a projeção de perspectiva empregando as matrizes lineares definidas acima é dada por:

$$\tilde{m} \cong A[R | t]\tilde{W}, \quad (11)$$

onde  $\tilde{W}$  corresponde a  $W$  empregando 1 como a quarta coordenada,  $\tilde{m}$  representa em coordenadas homogêneas a projeção de  $W$  no plano projetivo  $\pi$  e onde  $\cong$  é empregado para indicar que os dois lados da igualdade podem diferir por uma constante não-nula desconhecida, já que se trata de uma equação homogênea.

Finalmente, sendo  $\tilde{m} = [u, v, w]^T$ , as coordenadas do ponto  $W$  no *sistema de coordenadas da imagem* podem ser obtidas por:

$$m = [u/w \quad v/w]^T, \quad (12)$$

completando desta forma a descrição do processo de formação da imagem.

É importante observar que frequentemente o mapeamento acima descrito é representado empregando a *matriz de projeção* [12] dada por:

$$P \cong A[R | t]. \quad (13)$$

Neste caso, a equação ( 11 ) deve ser reescrita como:

$$\tilde{m} \cong P\tilde{W}. \quad (14)$$

#### 2.1.4 Calibração Monocular

A calibração de uma câmera consiste na determinação dos parâmetros intrínsecos e extrínsecos que, conforme discutido nas seções 2.1.1 e 2.1.2, correspondem às

características geométricas e ópticas internas da câmera e da sua orientação e posicionamento em relação ao *sistema de coordenadas do mundo*.

Diversos métodos de calibração encontram-se disponíveis. Enquanto alguns destes seguem a abordagem de auto-calibração, empregando elementos estruturais da cena como segmentos de reta e vértices diretamente, a maioria realiza a calibração empregando um gabarito cuja geometria é conhecida [15,16,17]. O grande benefício ao empregar gabaritos de geometria conhecida é que as projeções de seus pontos característicos nas imagens capturadas pela câmera são mais facilmente detectadas.

Seguindo esta abordagem [18], são necessárias ao procedimento de calibração as coordenadas no *sistema de coordenadas do mundo* de pontos característicos de um gabarito e uma imagem contendo o gabarito, capturada pela câmera a ser calibrada. Inicialmente são empregadas técnicas para a extração de características da imagem, como vértices e contornos, a fim de identificar os pontos candidatos às projeções de pontos do gabarito. Tomando o conjunto de pontos característicos extraídos, faz-se uso de restrições particulares ao gabarito utilizado para corresponder cada ponto 3-D do gabarito com o respectivo ponto 2-D da imagem. Como resultado desta etapa obtém-se  $n$  pares formados por pontos do mundo  $\{(X_i, Y_i, Z_i) | i = 1, \dots, n\}$  e suas respectivas projeções no plano de imagem  $\{(x_i, y_i, z_i) | i = 1, \dots, n\}$ .

Conforme introduzido pela equação ( 14 ), a transformação das coordenadas do *sistema de coordenadas do mundo* em coordenadas do *sistema de coordenadas da imagem* pode ser escrita empregando a *matriz de projeção*  $P$  como:

$$\begin{bmatrix} \lambda x_i \\ \lambda y_i \\ \lambda \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix}. \quad (15)$$

Sendo  $\lambda$  um fator de escala de um dos elementos da matriz de projeção, podemos igualar  $p_{34}$ , por exemplo, a 1 sem perda de generalidade. Desta forma, restam somente onze parâmetros a serem determinados no processo de calibração.

Expandindo a equação ( 15 ) temos:

$$\begin{cases} p_{11}X_i + p_{12}Y_i + p_{13}Z_i + p_{14} = \lambda x_i \\ p_{21}X_i + p_{22}Y_i + p_{23}Z_i + p_{24} = \lambda y_i \\ p_{31}X_i + p_{32}Y_i + p_{33}Z_i + 1 = \lambda \end{cases}, \quad (16)$$

onde substituindo a última equação nas duas primeiras e eliminando  $\lambda$  obtemos:

$$\begin{cases} p_{11}X_i + p_{12}Y_i + p_{13}Z_i + p_{14} - p_{31}X_i x_i - p_{32}Y_i x_i - p_{33}Z_i x_i = x_i \\ p_{21}X_i + p_{22}Y_i + p_{23}Z_i + p_{24} - p_{31}X_i y_i - p_{32}Y_i y_i - p_{33}Z_i y_i = y_i \end{cases}. \quad (17)$$

Podemos reformular o problema como um sistema de equações lineares na forma  $A\varphi = b$  (18), onde  $A$  é uma matriz  $2n \times 11$ ,  $\varphi$  é um vetor contendo as 11 incógnitas da *matriz de projeção*  $P$  e onde  $b$  é um vetor  $2n$ , sendo  $A$  e  $b$  obtidos a partir dos pontos de calibração.

$$\left[ \begin{array}{ccccccccc|ccc} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -X_1x_1 & -Y_1x_1 & -Z_1x_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -X_1y_1 & -Y_1y_1 & -Z_1y_1 \\ \dots & \dots \\ X_i & Y_i & Z_i & 1 & 0 & 0 & 0 & 0 & -X_ix_i & -Y_ix_i & -Z_ix_i \\ 0 & 0 & 0 & 0 & X_i & Y_i & Z_i & 1 & -X_iy_i & -Y_iy_i & -Z_iy_i \\ \dots & \dots \\ X_n & Y_n & Z_n & 1 & 0 & 0 & 0 & 0 & -X_nx_n & -Y_nx_n & -Z_nx_n \\ 0 & 0 & 0 & 0 & X_n & Y_n & Z_n & 0 & -X_ny_n & -Y_ny_n & -Z_ny_n \end{array} \right] \begin{bmatrix} p_{11} \\ p_{12} \\ \vdots \\ p_{31} \\ p_{32} \\ \vdots \\ p_{33} \end{bmatrix} = \begin{bmatrix} x_1 \\ y_1 \\ \vdots \\ x_i \\ y_i \\ \vdots \\ x_n \\ y_n \end{bmatrix}. \quad (18)$$

Com onze incógnitas e cada ponto conhecido fornecendo duas equações ao sistema, são necessários pelo menos seis pontos para calcular  $P$ . A estimativa  $\hat{\varphi}$  dos valores de  $\varphi$ , e logo de  $P$ , pode ser obtida através de otimização pelo método dos mínimos quadrados:

$$\hat{\varphi} = \min_{\varphi} \|A\varphi - b\|^2, \quad (19)$$

ou, na forma matricial:

$$\hat{\varphi} = (A^T A)^{-1} A^T b. \quad (20)$$

Caso seja necessário, os parâmetros extrínsecos e intrínsecos podem ainda ser isolados a partir da matriz  $P$  [12].

O método descrito ignora possível influência de distorção radial. Muitos métodos consideram esta influência empregando modelos de câmera não lineares, diferentes do empregado neste trabalho. Entretanto, o estado da arte atual permite produzir câmeras cuja distorção radial pode ser desprezada quando não é necessária alta precisão [12, 19]. Uma avaliação do efeito da distorção radial no processo de calibração e reconstrução é apresentada por BENEVENUTI [19].

O método acima apresenta o inconveniente de que a informação de calibração somente será válida no volume em torno da posição dos pontos pertencentes ao gabarito de calibração utilizado. Por esta razão, um método de calibração onde são incluídos pontos correspondentes do padrão tomados em diferentes instantes é proposto por ZHANG [16]. O método consiste nas seguintes etapas:

- 1) Encontrar a matriz de transformação de perspectiva entre o plano do gabarito de calibração e o plano da visão da câmera, para um conjunto de amostras capturadas em diferentes instantes.
- 2) Inicializar os parâmetros intrínsecos.
- 3) Encontrar os parâmetros extrínsecos.
- 4) Realizar a otimização pela minimização do erro da projeção dos pontos com todos os parâmetros.

Este método apresenta resultados superiores aos obtidos pelo método anterior, sendo que a qualidade da calibração não se mostra degradada para pontos fora do padrão de calibração. Este método apresenta ainda a vantagem de empregar um padrão de calibração planar simplificado e de fácil construção.

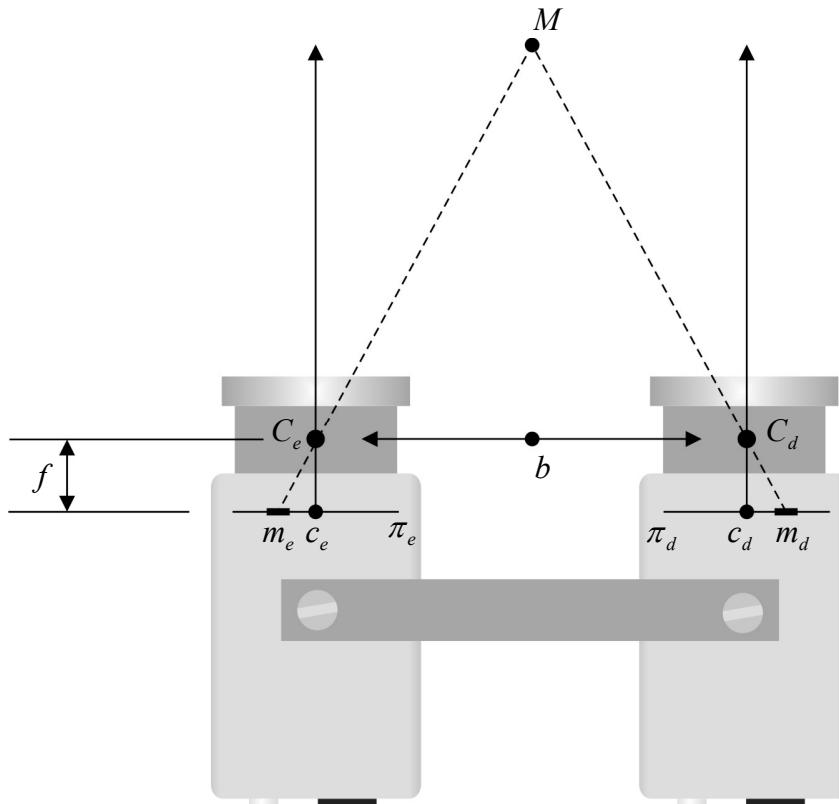
## 2.2 Sistema de Visão Estereoscópica

Na seção anterior descrevemos o modelo de câmera *pinhole* como o dispositivo que projeta pontos do espaço 3-D em pontos em um plano de imagem. Este processo não é reversível, visto que um ponto  $m$  no plano de imagem pode corresponder à projeção de

qualquer ponto sobre uma determinada reta no espaço. Por este motivo, os sistemas de visão estereoscópica fazem uso de uma segunda câmera para introduzir a restrição adicional necessária para que seja possível realizar o processo de reconstrução, recuperando assim as coordenadas de um ponto no espaço de maneira unívoca.

Através de um sistema de visão estereoscópica podemos recuperar a informação de profundidade de um dado ponto no espaço a partir da distância relativa entre dois pontos que o representam em imagens capturadas sob diferentes pontos de vista. Esta distância é denominada *disparidade*, e foi originalmente utilizada para descrever a diferença na localização de pontos característicos observados pelo olho esquerdo e pelo olho direito do sistema visual humano [20].

O modelo exibido na Figura 4 apresenta a geometria básica presente em um sistema de visão estereoscópica composto por duas câmeras do modelo *pinhole*.



**Figura 4:** Modelo simplificado de sistema de visão estéreo

Neste modelo simplificado [12], os planos de imagem das câmeras da esquerda  $\pi_e$  e da direita  $\pi_d$  são coplanares e seus eixos ópticos são paralelos, o que significa que a intersecção dos mesmos encontra-se infinitamente distante das câmeras. Os centros de projeção  $C_e$  e  $C_d$  são separados por uma distância denominada *linha de base*, ou  $b$ , perpendicular aos eixos ópticos.

Podemos determinar a posição de  $M$  no espaço por triangulação, ou seja, pela intersecção dos raios definidos pelos centros de projeção e as imagens de  $M$ ,  $m_e$  e  $m_d$ .

Sejam  $x_e$  e  $x_d$  as coordenadas de  $m_e$  e  $m_d$  em relação aos pontos principais  $c_e$  e  $c_d$ ,  $f$  a *distância focal* e tomando o ponto equidistante aos centros de projeção das duas câmeras como a origem do sistema de coordenadas a qual  $M$  pertence. Por similaridade de triângulos  $(m_e, M, m_d)$  é possível observar que a profundidade do ponto  $M$  pode ser calculada por:

$$Z = f \frac{b}{d} \quad (21)$$

onde  $d$  corresponde à *disparidade*, que por sua vez é dada por:

$$d = x_d - x_e. \quad (22)$$

Assim sendo, temos que a profundidade é inversamente proporcional à disparidade. É importante notar que esta característica tem efeito sobre a precisão da profundidade recuperada, onde a posição recuperada para objetos mais próximos será mais precisa do que a posição recuperada para objetos mais distantes [18].

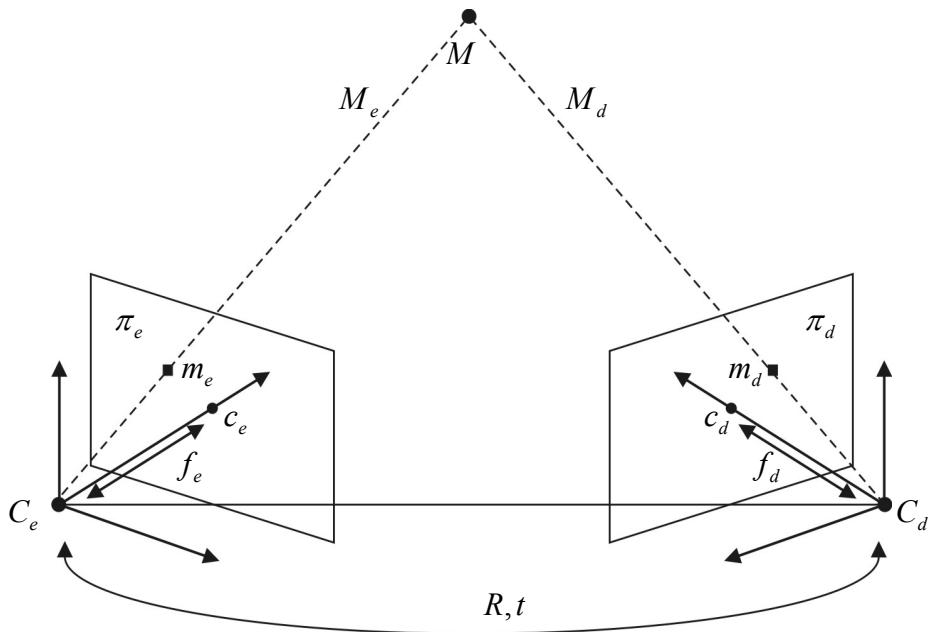
É possível observar ainda que, como a dimensão  $x$  é definida como sendo ao longo da linha que liga os centros dos pontos principais das imagens, temos que qualquer ponto projetado sobre ambas as imagens aparecerá na mesma altura, ou na mesma coordenada para a dimensão  $y$ . Desta forma podemos ignorar a dimensão  $y$  na recuperação da profundidade por triangulação.

A configuração de sistema de visão estereoscópica assumida neste modelo simplificado dificilmente é alcançada na prática, uma vez que é muito difícil construir um sistema binocular onde as câmeras estejam exatamente alinhadas conforme apresentado. Por este motivo são empregados os recursos de calibração do conjunto estéreo, descrito na seção 2.3, de maneira a compensar as diferenças obtidas.

Na seção 2.2.1 serão apresentados os parâmetros do sistema de visão estéreo, e na seção 2.2.2 serão apresentadas as importantes relações geométricas existentes entre os pontos  $m_e$  e  $m_d$ , descritas pela *geometria epipolar*.

### 2.2.1 Parâmetros do Sistema de Visão Estéreo

Assim como no modelo de câmera *pinhole*, os parâmetros de um sistema de visão estereoscópica podem ser divididos como *intrínsecos* e os *extrínsecos* [12]. Estes parâmetros são apresentados na Figura 5. Observe que na ilustração, por simplicidade, os planos de imagens estão representados entre o ponto observado e os centros de projeção. Apesar de esta representação estar distante da realidade, as relações matemáticas são preservadas e os modelos são formalmente equivalentes.



**Figura 5:** Parâmetros extrínsecos e intrínsecos do sistema de visão estéreo

Neste caso, os parâmetros intrínsecos correspondem aos mesmos apresentados na seção 2.1.1 para cada uma das câmeras que compõe o sistema, e são utilizados para mapear as coordenadas de pontos do *sistema de coordenadas da imagem* para o *sistema de coordenadas da câmera*, e vice-versa. Um conjunto mínimo de parâmetros intrínsecos inclui, para cada câmera, as coordenadas do *ponto principal* ( $c_e$  e  $c_d$ ) e a *distância focal* ( $f_e$  e  $f_d$ ).

Os parâmetros extrínsecos diferem ligeiramente dos parâmetros presentes no modelo de câmera *pinhole*. Para um sistema de visão estéreo, eles descrevem a transformação que relaciona os sistemas de coordenadas das duas câmeras um ao outro, definida por um vetor de translação,  $t = (C_d - C_e)$ , e uma matriz de rotação  $R$ .

A relação entre os vetores  $M_e = [X_e, Y_e, Z_e]^T$  e  $M_d = [X_d, Y_d, Z_d]^T$  que se referem ao mesmo ponto  $M$  no espaço, representado nos sistemas de coordenadas das câmeras esquerda e direita é dada por:

$$M_d = R(M_e - t). \quad (23)$$

Seja  $t_e$ ,  $R_e$ , e  $t_d$ ,  $R_d$  os parâmetros extrínsecos das duas câmeras que compõe o sistema, temos que os parâmetros extrínsecos do sistema de visão estereoscópica podem ser obtidos pelas equações (24) e (25).

$$R = R_d R_e^T \quad (24)$$

$$t = t_e - R^T t_d \quad (25)$$

### 2.2.2 Geometria Epipolar

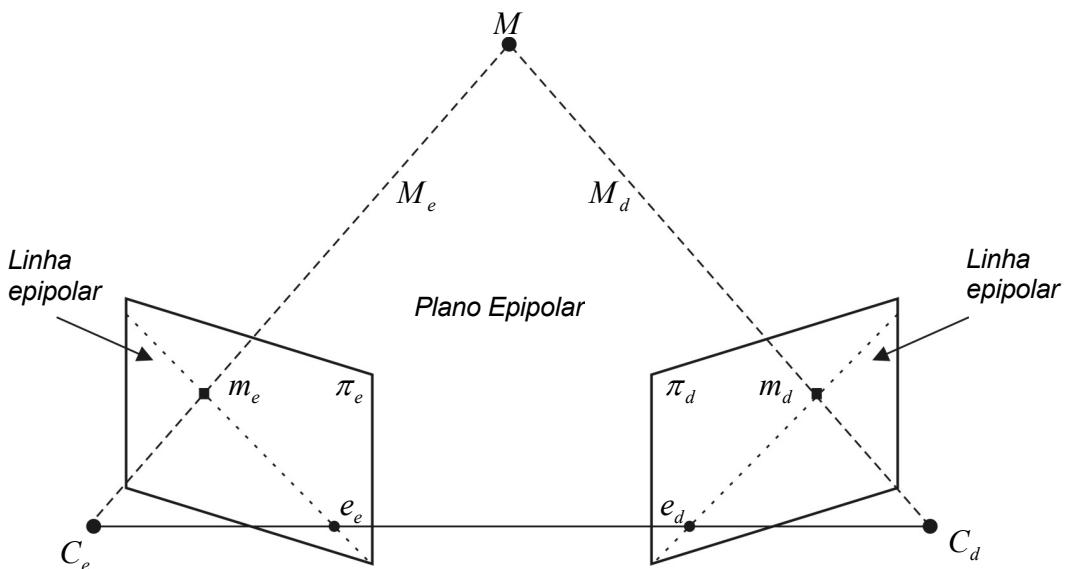
O relacionamento geométrico existente entre duas imagens de uma mesma cena tridimensional é descrito pela *geometria epipolar*. Segundo a *geometria epipolar*, os pontos correspondentes nas imagens estéreo devem aparecer em linhas particulares das imagens. Dado um ponto em uma imagem de um par estéreo, podemos encontrar seu ponto correspondente na outra imagem sobre uma linha específica, e não sobre toda imagem. Isto

reduz significativamente a complexidade do problema de correspondência, a ser descrito na seção 2.5.

A Figura 6 apresenta os elementos da *geometria epipolar* presente em um sistema de visão estereoscópica composto por duas câmeras *pinhole*, representadas pelos seus centros de projeção  $C_e$  e  $C_d$ , e seus planos de imagem  $\pi_e$  e  $\pi_d$  [14]. Cada câmera possui um sistema de coordenadas 3-D cuja origem coincide com seu centro de projeção e cujo eixo Z coincide com o *eixo óptico*. Os vetores  $m_e = [x_e, y_e, z_e]^T$  e  $m_d = [x_d, y_d, z_d]^T$  referem-se às projeções de  $M$  nos planos de imagem esquerdo e direito, respectivamente.

Os pontos em que a linha através dos centros de projeção intersecta os planos de imagem das câmeras são chamados epipolos, representados na Figura 6 por  $e_e$  e  $e_d$ . Sendo assim, o epipolo da esquerda é a imagem do centro de projeção da câmera direita, e vice-versa.

O plano definido por  $M, C_e$  e  $C_d$  é denominado plano epipolar, e a linha reta correspondendo à intersecção do plano epipolar com o plano da imagem é denominado linha epipolar. Considerando uma linha que passa pelo centro de projeção de uma câmera e um ponto no plano de imagem, temos que a linha epipolar é a imagem desta linha no plano de imagem da outra câmera. Adicionalmente, temos que todas as linhas epipolares intersectam no epipolo.



**Figura 6:** Geometria epipolar presente em sistema de visão estéreo

Dado  $m_e$ ,  $M$  está em qualquer posição no raio através de  $C_e$  e de  $m_e$ . Sendo que a imagem deste raio na imagem direita é a linha epipolar através do ponto correspondente  $m_d$ , o casamento correto precisa estar sobre a linha epipolar. Esta importante característica é denominada *restrição epipolar*. Ela estabelece um mapeamento entre pontos na imagem esquerda e linhas na imagem direita, e vice-versa.

Se, por exemplo, fizermos o mapeamento entre pontos na imagem esquerda e as correspondentes linhas epipolares na imagem direita, podemos restringir a busca pelos casamentos de cada ponto  $m_e$  sobre a correspondente linha epipolar. Assim sendo, a busca por correspondências é reduzida para um problema unidimensional.

A *geometria epipolar* pode ser descrita e empregada de diferentes maneiras, dependendo da quantidade de informação que se tem a respeito do sistema de visão estéreo. Quando tanto os parâmetros intrínsecos como extrínsecos são conhecidos, podemos expressar as restrições da geometria diretamente por:

$$m_d^T(t \times Rm_e) = 0, \quad (26)$$

como a linha homogênea através de dois pontos expressa por seu produto vetorial [12], onde  $t$  e  $R$  referem-se aos parâmetros do sistema de visão estereoscópica introduzidos na seção 2.2.1.

Quando somente os parâmetros intrínsecos são conhecidos, a *geometria epipolar* pode ser descrita pela *matriz essencial* [14]. Tomando a câmera da esquerda como origem do *sistema de coordenadas do mundo*, podemos reescrever as matrizes de projeção das câmeras conforme apresentado nas equações ( 27 ) e ( 28 ), onde o desconhecimento dos parâmetros extrínsecos é tornado explícito.

$$P_e = [I \mid 0] \quad (27)$$

$$P_d = [R \mid t] \quad (28)$$

O produto vetorial pode ser escrito como o produto de uma matriz anti-simétrica e um vetor. Sendo assim, a equação ( 26 ) pode ser equivalentemente escrita como:

$$\mathbf{m}_d^T [t] \times R \mathbf{m}_e = 0, \quad (29)$$

onde  $[t]$  é a matriz anti-simétrica do vetor  $t$ . Se multiplicarmos as matrizes nesta equação, obteremos uma matriz que descreve o relacionamento entre os pontos correspondentes na imagem  $\mathbf{m}_e$  e  $\mathbf{m}_d$  em coordenadas normalizadas.

Esta matriz é denominada *matriz essencial*, sendo dada por:

$$E = [t] \times R. \quad (30)$$

Desta forma, o relacionamento entre pontos correspondentes na imagem em coordenadas da câmera pode ser expressa empregando a *matriz essencial* por:

$$\mathbf{m}_d^T E \mathbf{m}_e = 0. \quad (31)$$

A *matriz essencial* tem cinco graus de liberdade, correspondendo a uma rotação e uma translação 3-D.

Quando não existe informação de calibração disponível, a *geometria epipolar* pode ser descrita pelo cálculo da *matriz fundamental* [14]. Tomando a câmera da esquerda como origem do *sistema de coordenadas do mundo*, podemos reescrever as matrizes de projeção das câmeras conforme apresentado pelas equações (32) e (33), onde é assumido o desconhecimento de todos os parâmetros do sistema de visão estéreo.

$$P_e = A_e [I \mid 0] \quad (32)$$

$$P_d = A_d [R \mid t] \quad (33)$$

Como no caso da *matriz essencial*, podemos reescrever a equação (26) neste caso como:

$$\mathbf{m}_d^T A_d [t] \times R A_e^{-1} \mathbf{m}_e = 0. \quad (34)$$

A matriz:

$$F = A_d^{-T} [t] \times R A_e^{-1} \quad (35)$$

é denominada *matriz fundamental*, e descreve o relacionamento entre pontos correspondentes na imagem em coordenada de *pixel*.

Assim sendo, o relacionamento entre pontos correspondentes da imagem em coordenadas de *pixel* pode ser expressa empregando a *matriz fundamental* por:

$$m_d^T F m_e = 0. \quad (36)$$

A *matriz fundamental* é uma representação algébrica da *geometria epipolar*, e apresenta sete graus de liberdade, visto que é definida sobre um fator de escala e seu determinante é zero.

Para qualquer ponto  $m_e$  na imagem da esquerda, a linha epipolar correspondente  $l_d$  na imagem da direita pode ser expressa como:

$$l_d = F m_e. \quad (37)$$

De maneira similar, a linha epipolar  $l_e$  na imagem esquerda para um ponto  $m_d$  na imagem direita pode ser expresso como:

$$l_e = F^T m_d. \quad (38)$$

Já a relação entre os epipolos e a matriz fundamental é dada pelas equações (39) e (40), onde o epipolo  $e_e$  é o vetor nulo direito da matriz fundamental e o epipolo  $e_d$  é o vetor nulo esquerdo da matriz fundamental.

$$F e_e = 0 \quad (39)$$

$$e_d^T F = 0 \quad (40)$$

A estimativa das matrizes *essencial* e *fundamental* acima apresentadas é tratada como um procedimento de calibração [12].

## 2.3 Calibração Estereoscópica

A calibração de um sistema de visão estereoscópica consiste na determinação dos parâmetros internos, orientação e posicionamento das câmeras que o compõe.

Quando o sistema de visão estereoscópica é rígido, cujos parâmetros não são alterados durante a operação do sistema, normalmente a calibração é realizada como uma etapa preliminar à operação do sistema. Já para os casos onde durante a operação do sistema podem ocorrer mudança de parâmetros intrínsecos ou extrínsecos como resultado de operações como foco e zoom (visão ativa), é comum empregar uma técnica denominada auto-calibração [19], onde não existe uma etapa inicial para determinação dos parâmetros do conjunto. Ao invés disso, a estimativa dos parâmetros necessários é realizada durante a operação do sistema. Este tipo de calibração é bastante flexível, porém nem sempre são obtidos resultados confiáveis.

De modo geral, podemos classificar os métodos de calibração como fortes (calibração clássica) ou fracos (calibração projetiva) [14]. A calibração forte refere-se à determinação tanto dos parâmetros intrínsecos como dos parâmetros extrínsecos, possibilitando assim uma reconstrução métrica dos pontos do mundo. Já na calibração fraca, apenas a *geometria epipolar* do sistema é determinada, sendo que desta forma apenas uma reconstrução projetiva é possível.

Pela complexidade e falta de precisão existente na calibração forte em sistemas de visão estéreo, é muitas vezes utilizada uma abordagem de visão qualitativa, onde não é utilizada uma métrica precisa [13]. Nestes casos, apenas a calibração fraca é empregada. Apesar de não ser utilizada informação métrica, ainda é possível obter informação tridimensional da cena e, empregando descritores invariantes, realizar tarefas que envolvam detecção, navegação, etc. Por outro lado, existe uma motivação biológica para realizar investigação seguindo o paradigma da visão qualitativa. O afastamento da abordagem métrica está de acordo com a nossa sensibilidade, com a qual podemos facilmente distinguir o mais próximo entre dois objetos, mas com a qual não podemos determinar precisamente a distância dos mesmos.

Na seqüência são descritas abordagens para realização das calibrações forte e fraca, de interesse ao modelo de sistema de visão considerado neste trabalho.

### 2.3.1 Calibração Clássica

Na calibração clássica [18], cada câmera que compõe o sistema binocular é calibrada separadamente, seguindo o procedimento descrito na seção 2.1.4. A determinação dos parâmetros do sistema de visão estereoscópica é, em seguida, realizada através do procedimento descrito na seção 2.2.1.

Seguindo esta abordagem, temos que os parâmetros intrínsecos do sistema de visão estereoscópica são formados pelos parâmetros intrínsecos de cada câmera que compõe o sistema, e cujo conjunto mínimo inclui as coordenadas dos pontos principais e as distâncias focais. Os parâmetros extrínsecos, por sua vez, são calculados diretamente pelas equações (24) e (25) empregando os parâmetros extrínsecos obtidos para cada câmera do sistema.

### 2.3.2 Calibração Projetiva

A calibração projetiva procura determinar os parâmetros da *geometria epipolar* presente no sistema de visão estéreo. Os métodos que seguem esta abordagem baseiam-se geralmente na estimativa da *matriz fundamental*. Para isso, é empregado um conjunto de pontos correspondentes identificados no par de imagens estéreo. De maneira similar à calibração clássica, pode-se empregar um gabarito de calibração para facilitar a identificação dos pontos correspondentes.

Pela equação (36) podemos observar que cada par de pontos correspondentes entre as duas imagens estéreo provê uma restrição linear à *matriz fundamental*. Isto permite que a *matriz fundamental* seja estimada linearmente por um mínimo de oito correspondências independentes [12].

Tomando-se  $n$  pontos correspondentes nas imagens do par estéreo, cada correspondência  $m_{ei} \leftrightarrow m_{di}$  gera uma restrição aos coeficientes de  $F$ , como apresentado pela equação (41).

$$[x_{di}, y_{di}, 1] \begin{bmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{bmatrix} \begin{bmatrix} x_{ei} \\ y_{ei} \\ 1 \end{bmatrix} = 0 \quad (41)$$

Combinando as equações obtidas para cada correspondência obtemos um sistema linear que pode ser reescrito como  $A\varphi = 0$ , conforme apresentado pela equação ( 42 ). Nesta equação  $\varphi$  corresponde ao vetor contendo as nove incógnitas de  $F$  e  $A$  corresponde a uma matriz  $N \times 9$  em que cada linha é construída a partir das coordenadas de  $m_d$  e  $m_e$  referentes a uma simples correspondência.

$$\begin{bmatrix} x_{d1}x_{e1} & x_{d1}y_{e1} & x_{d1} & y_{d1}x_{e1} & y_{d1}y_{e1} & y_{d1} & x_{e1} & y_{e1} & 1 \\ \dots & \dots \\ x_{di}x_{ei} & x_{di}y_{ei} & x_{di} & y_{di}x_{ei} & y_{di}y_{ei} & y_{di} & x_{ei} & y_{ei} & 1 \\ \dots & \dots \\ x_{dn}x_{en} & x_{dn}y_{en} & x_{dn} & y_{dn}x_{en} & y_{dn}y_{en} & y_{dn} & x_{en} & y_{en} & 1 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ \vdots \\ f_9 \end{bmatrix} = 0 \quad (42)$$

A partir de oito pontos correspondentes é possível determinar a estimativa  $\hat{\varphi}$  dos valores de  $\varphi$ , logo de  $F$ , sujeito a um fator de escala, pela solução de mínimos quadrados ( 43 ).

$$\hat{\varphi} = \min_{\varphi} \| A\varphi \|^2 \text{ sujeito a } \| \varphi \| = 1 \quad (43)$$

BENEVENUTI [18] observa que, como a origem do sistema de coordenadas do *sistema de coordenadas da imagem* está posicionada no canto superior da imagem, tem-se um efeito sobre a estabilidade do sistema apresentado pela matriz da equação ( 42 ) introduzido pela homogeneidade das coordenadas dos pontos. Adicionalmente, apresenta um método onde são utilizados pontos conhecidos no mundo para contornar a limitação da *matriz fundamental* no que diz respeito à reconstrução projetiva. Segundo este método, conhecendo-se as coordenadas de alguns pontos no espaço e as coordenadas de alguns pontos correspondentes nas duas imagens estéreo, pode-se contornar esta limitação e calcular a matriz de transformação que permita uma reconstrução no sistema de coordenadas do ambiente. Este método apresenta cinco etapas principais:

- 1) Estimação da *geometria epipolar* a partir de todos os pares de pontos correspondentes disponíveis.
- 2) Cálculo das matrizes de projeção das câmeras compatíveis com a *geometria epipolar*.

- 3) Calculo da reconstrução dos pontos cuja posição no sistema de coordenadas do ambiente é conhecida.
- 4) Estimação da matriz de transformação entre a estrutura projetiva e a estrutura euclidiana.
- 5) Cálculo das matrizes de projeção das câmeras em relação ao sistema de coordenadas do ambiente.

Seguindo este procedimento, são obtidas matrizes de projeção que permitem a reconstrução métrica de pontos do ambiente.

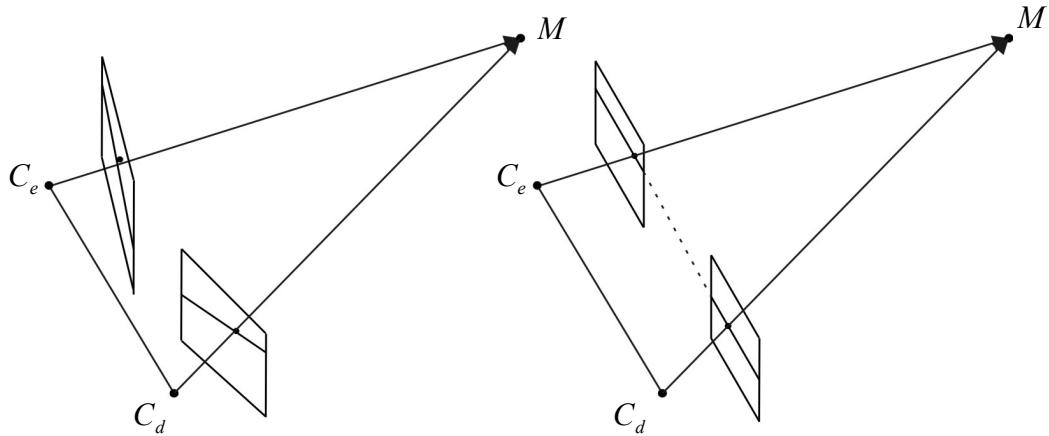
## 2.4 Retificação

Conforme discutido na seção 2.2.2, observando a restrição epipolar temos que dado um ponto em uma imagem, podemos encontrar seu ponto correspondente na outra imagem sobre uma linha específica. Isto reduz significativamente a complexidade do problema de correspondência entre pontos. Neste caso, a *geometria epipolar* permite reduzir uma busca essencialmente 2-D, onde seria necessário procurar por um ponto correspondente em toda a imagem, a uma busca 1-D, onde a busca é limitada a uma linha específica.

Convenientemente, pode-se fazer uso do recurso de retificação para simplificar ainda mais o problema de correspondência. A retificação de um par de imagens estéreo consiste na transformação de cada imagem de modo que os pares de linhas epipolares fiquem colineares e paralelas ao eixo  $x$  da imagem. Nas novas projeções obtidas pela retificação do par de imagens estéreo, as linhas epipolares são paralelas ao eixo  $x$  em ambas as imagens, e consequentemente as disparidades entre as imagens são apenas na direção  $x$ , não existindo disparidade no eixo  $y$ . Desta forma, através da retificação de imagens, para encontrar um ponto correspondente a  $(x_l, y_l)$  na imagem da esquerda, simplesmente procuramos através da linha  $y_r = y_l$  na imagem da direita. Pelos benefícios obtidos, muitos algoritmos de correspondência assumem a retificação como um passo preliminar.

Conforme observado por FUSIELLO et al. [21], a transformação do par de imagens estéreo por transformações de projeção apropriadas reduzem o problema de correspondência à *geometria epipolar* encontrada em um par de câmeras idênticas colocadas

lado a lado com seus eixos principais alinhados, conforme ilustrado pela Figura 7. Em outras palavras, a retificação transforma o par de imagens estéreo de modo que se tornem equivalentes a um par de imagens capturadas por uma nova configuração de sistema de visão estéreo, em conformidade com o modelo simplificado apresentado na seção 2.2.



**Figura 7:** Transformação da configuração estéreo pela retificação

Diferentes métodos podem ser empregados para realizar a retificação, dependendo do tipo de informação disponível a respeito do sistema de visão estéreo. TRUCCO et al. [12] descrevem um procedimento de retificação onde são utilizados os parâmetros intrínsecos e extrínsecos. O método emprega uma matriz ortogonal definida por vetores unitários mutuamente ortogonais  $e_1$ ,  $e_2$  e  $e_3$ , apresentada pela equação (44).

$$R_{rect} = \begin{pmatrix} e_1^T \\ e_2^T \\ e_3^T \end{pmatrix} \quad (44)$$

O vetor  $e_1$  é dado pelo epipolo. Sendo que o centro da imagem está na origem,  $e_1$  coincide com a direção da translação e é dado por:

$$e_1 = \frac{T}{\|T\|}. \quad (45)$$

O vetor  $e_2$  deve ser ortogonal a  $e_1$ , e para isso podemos calcular e normalizar o produto vetorial com o vetor de direção do *eixo óptico*. Desta forma,  $e_2$  é dado por:

$$e_2 = \frac{1}{\sqrt{T_x^2 + T_y^2}} [-T_y, T_x, 0]^T. \quad (46)$$

Finalmente,  $e_3$  é determinado inequivocamente por:

$$e_3 = e_1 \times e_2. \quad (47)$$

A matriz  $R_{rect}$  rotaciona a câmera da esquerda em torno do centro de projeção de maneira a tornar as linhas epipolares paralelas ao eixo horizontal.

O algoritmo completo de retificação consiste nos seguintes passos:

- 1) Construir a matriz  $R_{rect}$  conforme foi descrito anteriormente.
- 2) Definir  $R_e = R_{rect}$  e  $R_d = RR_{rect}$ .
- 3) Para cada ponto  $m_e = [x, y, f]^T$  calcule  $[x', y', z'] = R_e p_e$ , e então calcule as coordenadas do ponto retificado como  $p'_e = \frac{f}{z'} [x', y', z']$ .
- 4) Repita os passos anteriores para a câmera da direita, empregando  $R_d$  e  $p_d$ .

Como saída do algoritmo temos dois conjuntos correspondendo aos pontos retificados das imagens da direita e da esquerda.

Da mesma maneira, é possível realizar a retificação quando somente a *matriz fundamental* é conhecida, fazendo uso de transformações projetivas denominadas homografias [14]. A entrada é um par de imagens estéreo e a *matriz fundamental* do sistema de visão estéreo. Os passos gerais do algoritmo são:

- 1) Identifique os epipolos  $e_e$  e  $e_d$  nas duas imagens.
- 2) Selecione a transformação projetiva  $H_d$  que mapeia o epipolo  $e_d$  ao ponto no infinito  $[1, 0, 0]^T$ .

- 
- 3) Encontre a transformação projetiva de correspondência  $H_e$  que minimize a distância quadrática  $\sum d(H_e x_{ei}, H_d x_{di})^2$ .
- 4) Transforme a primeira imagem de acordo com a transformação projetiva  $H_e$  e a segunda imagem de acordo com a transformação projetiva  $H_d$ .

A saída é um par de imagens transformadas de maneira que as linhas epipolares nas duas imagens sejam horizontais, e que os pontos correspondentes nas duas imagens estejam tão próximas uma das outras quanto possível.

## 2.5 Correspondência

Correspondência em estereoscopia refere-se ao procedimento de relacionar elementos presentes nas imagens capturadas pelo sistema de visão estéreo, identificando quais pontos da imagem da esquerda e da direita são projeções do mesmo ponto na cena tridimensional observada. Este é um dos tópicos mais pesquisados em visão computacional, e é uma das principais questões a serem tratadas na estereoscopia.

Oclusões, diferenças na iluminação, distorções radiométricas, distorções projetivas e texturas pobres são apenas alguns dos fatores que tornam o problema de correspondência uma tarefa complexa [18,19]. Por este motivo, a maioria dos métodos de correspondência existentes faz uso de uma série de restrições e pré-processamentos, com o objetivo de restringir o espaço de busca ou de normalizar as imagens do par estéreo.

Supondo que todos os pontos da cena a serem correspondidos são visíveis a ambos os pontos de vista e que os pontos correspondentes são similares, podemos considerar o problema de correspondência como sendo um problema de busca [12]. Desta forma, para um dado elemento na imagem da esquerda deve-se encontrar o elemento correspondente na imagem da direita. Seguindo esta abordagem, devem ser tratadas duas questões principais para proceder à correspondência. A primeira diz respeito à escolha dos elementos da imagem a serem correspondidos. A segunda refere-se à escolha da medida de similaridade que será utilizada para realizar a correspondência.

A escolha referente à quais pontos serão correspondidos ajuda a classificar convenientemente os muitos algoritmos de correspondência existentes em duas grandes

classes, os métodos de correspondência densa e os métodos de correspondência esparsa. Os métodos de correspondência densa procuram corresponder todos os pontos pertencentes à imagem. Já os métodos de correspondência esparsa correspondem um conjunto reduzido de pontos relevantes da imagem.

Na seqüência serão descritas as diferentes classes de algoritmos de correspondência.

### 2.5.1 Correspondência Densa

Os métodos desta classe realizam a correspondência para cada um dos pontos pertencentes à imagem. Para os algoritmos que implementam tal abordagem, a entrada é um par de imagens capturadas pelo sistema binocular, normalmente retificadas. Tomando uma das imagens como referência, para cada ponto desta imagem é realizada uma busca pelo ponto correspondente na outra imagem, segundo algum critério de similaridade. Para cada par de pontos correspondidos será calculado o valor da disparidade com base na distância relativa entre as coordenadas dos mesmos. O valor da disparidade será armazenado como o valor de intensidade na posição do ponto da imagem de referência em uma terceira imagem, conhecida como mapa de disparidades.

Os valores de intensidade para cada ponto do mapa de disparidade armazenam uma informação de profundidade da cena observada, uma vez que representam o valor da disparidade entre as imagens para cada ponto, e que a disparidade é inversamente proporcional à profundidade (vide equação ( 21 )). Assim, no mapa de disparidades, as regiões mais claras ou com valores de intensidade mais altos representam objetos mais próximos e as regiões mais escuras ou com valores de intensidades mais baixos representam objetos mais distantes.

Na Figura 8 é apresentado um par estéreo composto por imagens retificadas e o respectivo mapa de disparidade calculado utilizando correspondência densa.



**Figura 8:** Mapa de disparidade calculado por correspondência densa

Podemos ainda classificar os algoritmos de correspondência densa como sendo locais, globais ou híbridos, conforme a abordagem utilizada para corresponder os pontos pertencentes às imagens [20]. As diferentes abordagens são detalhadas nas seções seguintes.

### 2.5.1.1 Abordagem Local

Os algoritmos de correspondência local assumem que os pontos da imagem são envoltos por uma janela de pontos vizinhos onde a disparidade é a mesma. Desta forma, a correspondência para cada um dos pontos da imagem é realizada com base na correspondência destas janelas empregando funções de custo baseadas em similaridade.

O método básico utilizado para correspondência segundo esta abordagem é a correlação [12]. Vários algoritmos foram propostos utilizando variações desta abordagem. Estes algoritmos apresentam, de maneira geral, um comportamento comum. Dadas duas imagens capturadas pelo sistema binocular, escolha uma das imagens como referência e defina o tamanho  $2n+1$  para uma janela quadrada a ser utilizada na correspondência. Para cada ponto da imagem de referência, faça:

- 1) Centralize a janela no ponto de interesse na imagem de referência.
- 2) Correlacione a janela posicionada na imagem de referência com uma segunda janela, movendo-a sobre todas as posições pertencentes a um espaço de busca na segunda imagem.
- 3) Determine o ponto na segunda imagem que corresponde ao ponto de interesse na primeira imagem, selecionando a posição onde a correlação minimiza a função de custo empregada.

4) Calcule a disparidade pela diferença relativa da posição central das janelas.

Armazene este valor numa terceira imagem, na posição do ponto central da janela na imagem de referência.

A saída é um mapa de disparidades denso, contendo a disparidade para cada ponto da imagem de referência.

Os métodos de correspondência por correlação utilizam funções de custo que levam em conta a similaridade dos valores de intensidade de cada ponto da imagem, como a soma das diferenças quadráticas (*SSD*) ou soma das diferenças absolutas (*SAD*). Uma função de custo empregando a similaridade *SAD* comumente empregada é dada por:

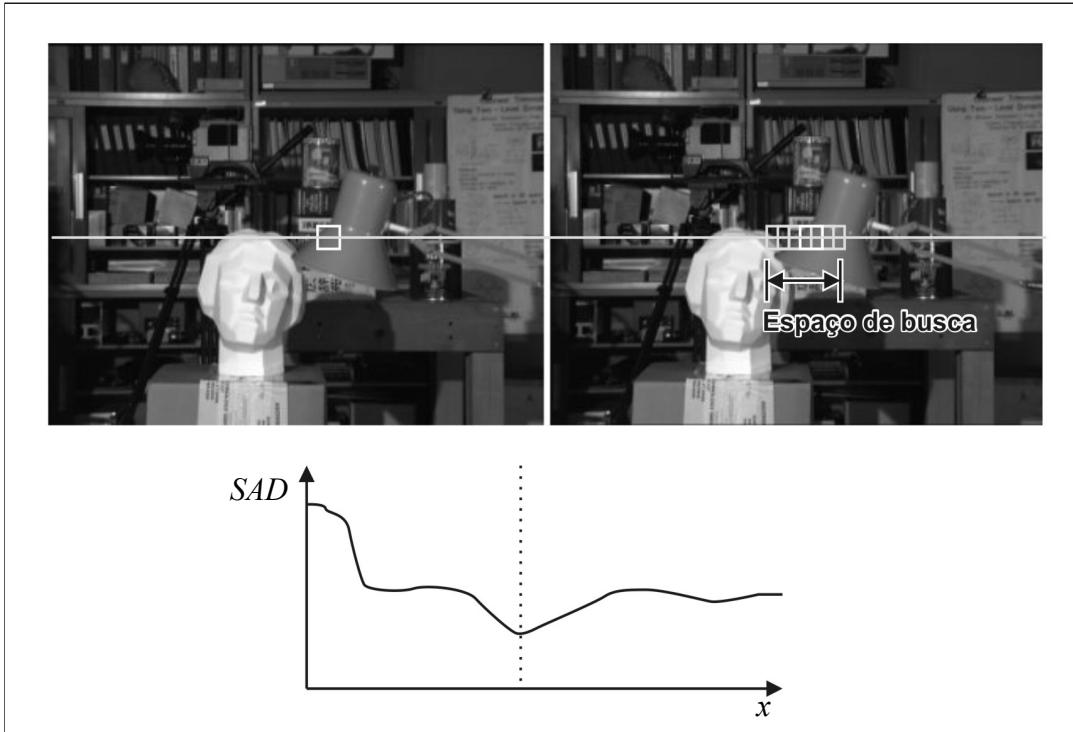
$$C_{sad}(x, y, d) = \sum_{i,j=-n}^n |(D_{le}(x+j, y+i) - D_{ld}(x+d+j, y+i))|. \quad (48)$$

onde  $d$  é uma dada disparidade,  $D_{le}(x, y)$  e  $D_{ld}(x, y)$  correspondem a uma função de descrição baseada em intensidade referente às imagens esquerda e direita, respectivamente. Uma janela de tamanho  $2n + 1$  é utilizada.

O espaço de busca é definido pelas disparidades máxima e mínima possíveis para uma dada configuração. Estes limites de disparidade, ou espaço de busca, podem ser, por exemplo, definidos pela mínima distância permitida entre o conjunto binocular e um objeto da cena, o que descreve uma disparidade máxima. Já a definição da disparidade mínima nos permite limitar o espaço de busca, por exemplo, para uma dada profundidade máxima de interesse.

A seleção da janela de correlação dentro do espaço de busca da segunda imagem que corresponde um ponto de interesse na primeira imagem é muitas vezes denominada abordagem *Winner-Takes-All (WTA)* [20].

A Figura 9 ilustra o processo de correspondência por correlação de janelas descrito anteriormente, em que para um dado ponto na primeira imagem é definido um espaço de busca na segunda imagem onde é selecionada a janela de correlação que minimiza a função de custo utilizada.



**Figura 9:** Processo de correspondência por correlação de janelas

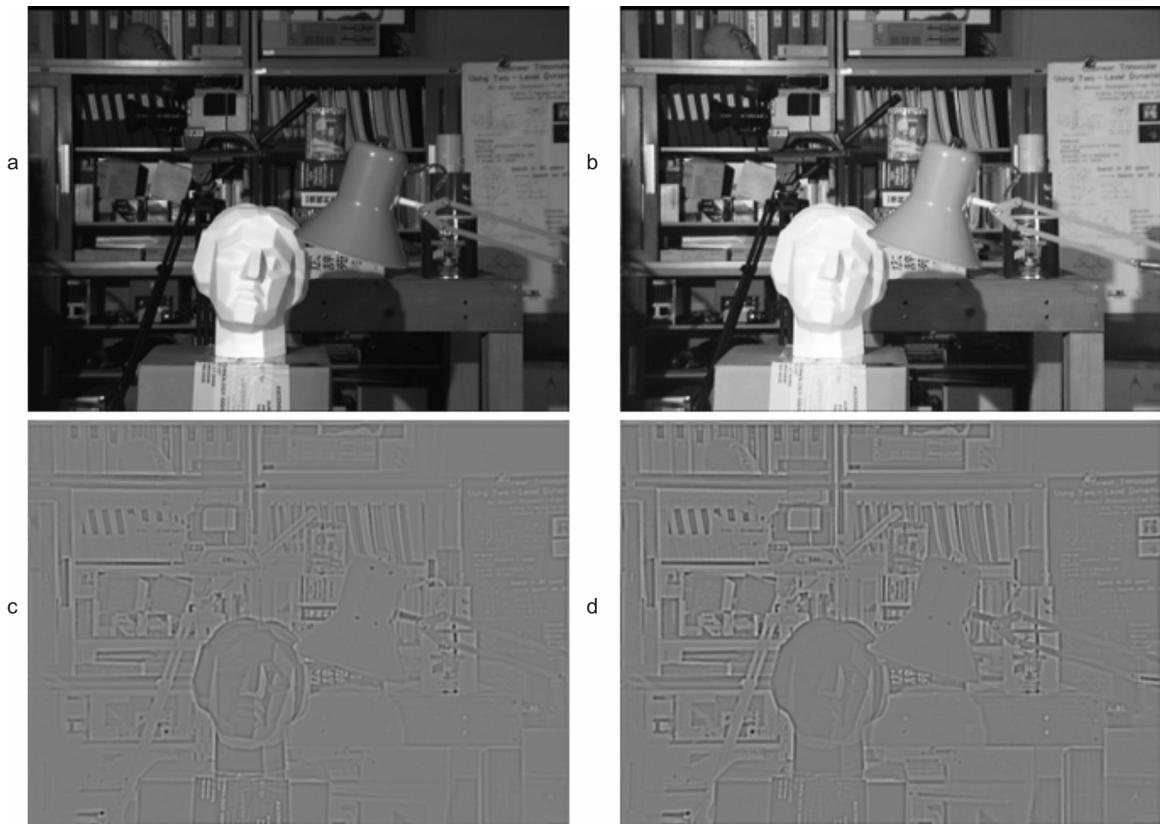
A correlação de uma simples janela está sujeita a erros nas descontinuidades de profundidade, devido à presença de oclusões nestas regiões. Com o objetivo de reduzir este problema, HIRSCHMULLER et al. [22] propõe uma extensão desta abordagem local onde faz o uso de múltiplas janelas para corresponder os pontos das imagens. Segundo esta proposta, o erro em regiões de descontinuidade pode ser reduzido se somente forem consideradas no cálculo da função de custo as partes da janela de correlação que não introduza este erro.

Para tal, é utilizada uma janela  $C$  formada por um número selecionado de janelas menores  $C_i$ . Estas janelas menores são dispostas de maneira a circularem uma janela central  $C_0$ , sobrepondo-a em alguns pontos. A correlação para cada uma destas janelas  $C_i$  é feita individualmente. O valor de correlação para a janela  $C$  é calculado adicionando ao valor de correlação da janela  $C_0$  somente os melhores valores de correlação obtidos pelas janelas que a rodeiam. O número de janelas a contribuírem com a janela central é proporcional ao

número de janelas definidas na configuração. Em outras palavras, utiliza-se uma pequena janela central onde a correlação tem o suporte das janelas próximas.

No processo de correspondência pelos métodos por correlação acima descritos a intensidade assume um papel de descritor para cada ponto das imagens. Esta descrição mostra-se pouco robusta visto que o valor de intensidade de cada *pixel* é sensível a mudanças na intensidade absoluta, introduzidos por influências como diferenças no ganho dos sensores, variações radiométricas, entre outros.

Uma solução largamente empregada é a utilização da normalização das imagens pelo filtro LOG [24]. O filtro LOG é composto por uma filtragem gaussiana e uma filtragem laplaciana. O filtro laplaciano reduz a influência da iluminação, já que a resposta ao mesmo é zero nas áreas com intensidade constante e é negativa ou positiva próximo às bordas com gradiente de intensidade alta. O filtro gaussiano é aplicado antes do filtro laplaciano para reduzir a influência de ruído, visto que o segundo é altamente sensível a este.

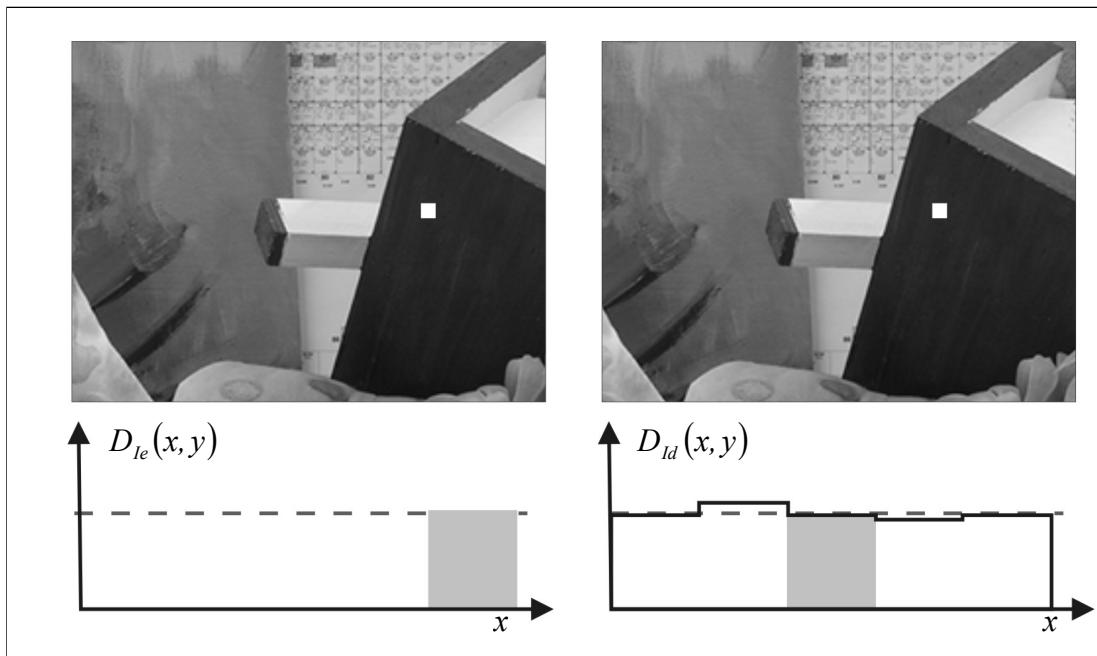


**Figura 10:** Normalização por LOG

A Figura 10 ilustra a normalização de um par de imagens estéreo empregando o filtro LOG. A imagem original direita apresenta uma diferença significativa na intensidade luminosa em relação à imagem da esquerda (a e b). Porém, ao empregarmos o filtro LOG nas imagens originais obtemos uma resposta muito semelhante (c e d).

Apesar de que a utilização do LOG na normalização das imagens aumenta a qualidade dos mapas de disparidade gerados a partir de imagens reais, são obtidas grandes regiões homogêneas, resultante de áreas onde a derivada de intensidade é baixa. Nestes casos, os critérios de similaridade clássicos continuam a apresentar um resultado não satisfatório.

A Figura 11 ilustra o problema existente na descrição de pontos baseada em intensidade durante o processo correspondência. Nesta ilustração, somente é empregado o descriptor baseado em intensidade  $D_I(x, y)$ .



**Figura 11:** Descrição baseada em intensidade

Considere a janela de correlação representada pelo ponto branco nas imagens. É possível observar como os valores que descrevem cada ponto podem ser próximos num dado espaço de busca. O emprego de uma função de custo baseada unicamente neste descriptor pode resultar na ocorrência de correspondências arbitrárias.

O uso de janelas grandes mostra-se uma opção, visto que reduz o número de janelas completamente sobrepostas a uma região homogênea. Janelas maiores aumentam a confiabilidade pela média sobre grandes regiões, além de reduzirem o efeito de ruído [23]. Porém prejudicam o resultado nas regiões de descontinuidade de profundidade. A escolha da janela de correlação corresponde, desta forma, a um compromisso entre aumentar a confiabilidade nas áreas com intensidade homogênea e diminuir o erro nas regiões onde a profundidade muda.

Para eliminar as inevitáveis correspondências arbitrárias, muitos métodos incluem a verificação de consistência denominada *left-right check* [25,26]. Esta consistência é realizada para cada ponto para verificar se as correspondências calculadas para um dado ponto são coerentes quando é alterada a imagem de referência que guia a correspondência [2]. O processo da verificação pode ser descrito como segue. Num primeiro momento, para cada ponto da imagem da esquerda é calculada a melhor correspondência com os pontos da imagem da direita. Num segundo momento, a imagem de referência é invertida e para cada ponto da imagem da direita é calculada a melhor correspondência com os pontos da imagem da esquerda. Finalmente, somente são consideradas as correspondências coerentes tanto na correspondência esquerda para direita quanto na correspondência direita para esquerda. Caso os valores de disparidade mostrem-se divergentes, estes pontos são marcados. Numa etapa seguinte, os pontos marcados podem ser removidos através de interpolação.

Um algoritmo de correlação implementado na sua forma original pode apresentar um alto custo computacional. Observa-se que a tarefa mais cara computacionalmente é o cálculo dos valores de custo de correspondência. Uma característica da função de custo baseada em similaridade *SAD* ou *SSD* é de que ela pode ser calculada eficientemente explorando o fato de que janelas vizinhas se sobrepõem [25].

Temos que em janelas vizinhas com a mesma disparidade os pontos sobrepostos terão os mesmos valores de diferença absoluta, tanto para o descriptor de intensidade como para o descriptor regional. Desta forma, um novo valor de custo pode ser calculado para a janela corrente utilizando o valor calculado para uma janela antiga subtraindo os valores dos pontos que fazem parte apenas da janela antiga, e adicionando os valores dos pontos que fazem parte da nova janela. Esta otimização apresenta ainda a vantagem de tornar o custo computacional praticamente independente do tamanho da janela de correlação empregada.

Adicionalmente, algoritmos baseados em correlação podem ser otimizados através da utilização de recursos de processamento paralelo SIMD disponíveis em diversas arquiteturas [22,24,25].

### 2.5.1.2 Abordagem Global

Os métodos que seguem a abordagem de correspondência densa global, em oposição à abordagem local, procuram propagar a informação de disparidade de um ponto da imagem para seus vizinhos. Para isso, comumente empregam a minimização de alguma função de energia sobre o mapa de disparidade inteiro.

O algoritmo de programação dinâmica é um dos algoritmos que calcula a disparidade segundo esta abordagem [27]. Inicialmente é definida uma função de custo que leva em conta o número de oclusões, correspondências e uma medida de dissimilaridade. Para cada linha da imagem é realizada uma busca exaustiva pela melhor solução de disparidade dentre as soluções possíveis, segundo esta função de custo.

A partir de duas imagens de entrada, o algoritmo realiza um processo de emparelhamento denso para cada linha. É realizada uma medição seguindo um critério de dissimilaridade entre os pontos, atribuindo a cada seqüência de correspondências  $M$  um custo, dado por:

$$\gamma(M) = N_{occ}k_{occ} - N_m k_r + \sum_{i=1}^{N_m} d(m_{ei}, m_{di}). \quad (49)$$

O custo  $\gamma(M)$  mede a probabilidade de que essa seqüência seja uma correspondência correta. Os parâmetros  $N_{occ}$  e  $N_m$  representam o número de oclusões e de emparelhamentos, sendo  $k_{occ}$  e  $N_r$  as respectivas constantes de penalidade e recompensa, e  $d(m_{ei}, m_{di})$  uma função de dissimilaridade baseada em intensidade entre os pontos da imagem esquerda e da imagem direita. A seqüência de emparelhamento ótima é realizada empregando uma busca exaustiva através de um algoritmo de programação dinâmica.

Como as intensidades entre linhas de uma imagem não são independentes, é realizada uma análise coluna a coluna no mapa de disparidades. Primeiro, a disparidade de

um *pixel*, cujos vizinhos verticais tenham disparidade igual, mas diferente dele próprio, torna-se igual à dos seus vizinhos. A seguir, para ambas as direções horizontal e vertical, utiliza-se o gradiente na direção considerada na imagem original para propagar regiões de confiança no mapa de disparidade. A confiança, para cada *pixel*, define-se como o número de pontos contíguos na direção a considerar com disparidade igual à sua. Depois, o mapa de disparidade é filtrado com um filtro de média nas duas direções seqüencialmente, de forma a preservar os vértices dos objetos.

As principais vantagens apresentadas pelo algoritmo é o tratamento de regiões homogêneas, a utilização de uma medida de similaridade insensível à amostragem e um algoritmo de programação dinâmica mais eficiente.

No geral, os métodos desta classe apresentam um melhor resultado em regiões homogêneas, mas eles freqüentemente tem parâmetros que são difíceis de definir e geralmente demandam um maior tempo de processamento do que nas técnicas locais [20].

### 2.5.1.3 Outras Abordagens

Alguns algoritmos de correspondência seguem abordagens que não permitem enquadrá-los nas classes de métodos locais ou globais. Dentre estes temos a correspondência baseada em segmentação [28, 29], redes neurais [30], algoritmos combinando informação estéreo e monocular [31], algoritmos cooperativos [23,32], dentre outros.

Os algoritmos denominados cooperativos seguem uma abordagem para o cálculo da disparidade inspirada em modelos computacionais da visão estereoscópica humana. Assim como nos métodos locais, estes algoritmos realizam operações locais para efetuar a correspondência entre os pontos da imagem. Porém, são iterativamente guiados a realizarem a minimização de funções globais, resultam em um comportamento geral equivalente aos algoritmos de otimização.

De maneira geral estes algoritmos fazem uso de regras globais que descrevem características envolvendo a observação de uma cena através de um sistema de visão estéreo. MARR et al. [23] introduziram duas suposições baseadas em restrições físicas num dos primeiros algoritmos pertencentes a esta classe de algoritmos. A primeira suposição, denominada unicidade, descreve que a um ponto de cada imagem somente é possível atribuir

um único valor de disparidade, devido ao fato de que um ponto corresponde a somente uma posição física. A segunda suposição, denominada continuidade, descreve que a disparidade varia suavemente em quase toda região da imagem, devido ao fato de que somente em uma pequena parte da imagem temos fronteiras que caracterizam descontinuidade da profundidade. Baseados nessas suposições eles propõe um algoritmo simples para difundir suporte entre estimativas de disparidade.

O algoritmo cooperativo proposto por ZITNICK et al. [32] utiliza-se destas mesmas regras globais para construir uma função de atualização  $L_{n+1}(x, y, d)$  para os valores de correspondência, dada por:

$$L_{n+1}(x, y, d) = L_0(x, y, d) * R_n(x, y, d), \quad (50)$$

onde  $L_0(x, y, d)$  corresponde ao valor inicial calculado através de uma função de similaridade convencional e  $R_n(x, y, d)$  corresponde a quantidade de inibição recebida por uma área de inibição para uma região de correspondência na iteração anterior, segundo a suposição de unicidade. Esta função de atualização gera valores contínuos e únicos pela difusão do suporte entre valores de correspondência vizinhos.

Fazendo uso de uma matriz 3D( $x, y, d$ ) correspondendo a cada ponto na imagem de referência e ao intervalo de disparidade possível, o algoritmo pode ser descrito de forma simplificada através dos seguintes passos:

- 1) Atribuir valores iniciais  $L_0$  de correspondência utilizando uma função de custo convencional baseada em intensidade.
- 2) Atualizar iterativamente os valores  $L_n$  através da função de atualização definida pela equação (50), até que os valores de correspondência tenham收敛ido.
- 3) Para cada ponto  $(x, y)$  encontre o ponto  $(x, y, d)$  com o maior valor de correspondência.

- 4) Caso o maior valor de correspondência seja maior que um limiar, considere o valor de disparidade  $d$ , caso contrário classifique o ponto como uma oclusão.

Este método gera mapas de disparidade com valores confiáveis para as regiões de descontinuidade e para regiões com intensidade homogênea, entretanto apresenta um custo computacional elevado.

### 2.5.2 Correspondência Esparsa

Os métodos desta classe restringem a busca por correspondência a um conjunto reduzido de pontos. Segundo esta abordagem, inicialmente o par de imagens estéreo é pré-processado por algum operador que extraia pontos característicos das imagens. A correspondência é então realizada com base na distância entre os atributos associados aos pontos característicos detectados.

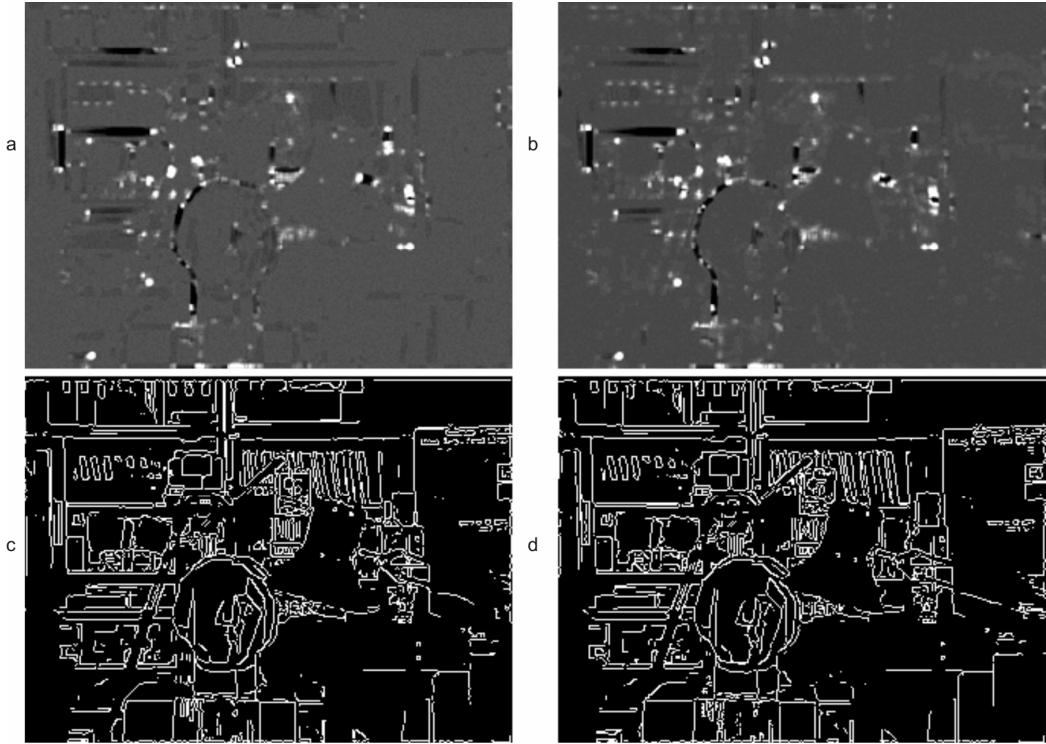
Contornos, linhas, segmentos de curva e vértices são características de imagens amplamente utilizadas em trabalhos de visão estereoscópica [33, 34]. Contornos e vértices são facilmente detectados, porém são menos robustos à oclusão. Segmentos de linha e curva são mais robustos à oclusão, entretanto demandam um tempo computacional adicional.

A escolha da característica a ser utilizada depende de vários fatores, como o tipo de objeto observado, as condições de iluminação e a média de contraste da imagem. Por exemplo, características de mais alto nível como círculos, polígonos ou elipses são empregados com sucesso em cenas estruturadas. Outras vezes, um conjunto de características é empregado em um único algoritmo de correspondência estéreo [33].

De maneira geral, um requisito básico imposto à característica utilizada é que a mesma seja invariante à mudança de perspectiva. A Figura 12 exibe o resultado do processamento de um par estéreo por diferentes operadores que atendem a este requisito. As imagens a e b apresentam o resultado do processamento do par estéreo pelo detector de vértices Harris [35]. As imagens c e d apresentam o resultado do processamento do par estéreo pelo detector de bordas Canny [36].

Ao conjunto de características extraídas, reduz-se ainda mais o número de correspondências possíveis utilizando diferentes restrições. Pode-se, por exemplo, considerar

as restrições geométricas impostas pela *geometria epipolar* e ainda a restrição imposta pela definição de um espaço de busca de interesse.



**Figura 12:** Imagens processadas por filtros extratores de características

Os passos de algoritmos desta classe são summarizados por TRUCCO [12]. A entrada é um par de imagens do sistema de visão estéreo,  $I_e$  e  $I_d$ , e dois conjuntos correspondentes de descritores de características. Seja  $R(f_e)$  a região de busca na imagem da direita associada com um descritor de características  $f_e$ , e  $d(f_e, f_d)$  a disparidade entre duas características correspondentes,  $f_e$  e  $f_d$ . Para cada  $f_e$  no conjunto da imagem esquerda faça:

- 1) Calcule a medida de similaridade entre  $f_e$  e cada característica de imagem em  $R(f_e)$ .
- 2) Selecione a característica da imagem direita,  $f_d$ , que maximize a medida de similaridade.
- 3) Salve a correspondência e a disparidade de  $f_e$ .

A saída é formada pela lista de características correspondentes e um mapa de disparidade esparso.

Na Figura 13 são ilustradas por linhas (abaixo) as correspondências de pontos característicos detectados sobre um par de imagens estéreo (acima), empregando o método SIFT [37].



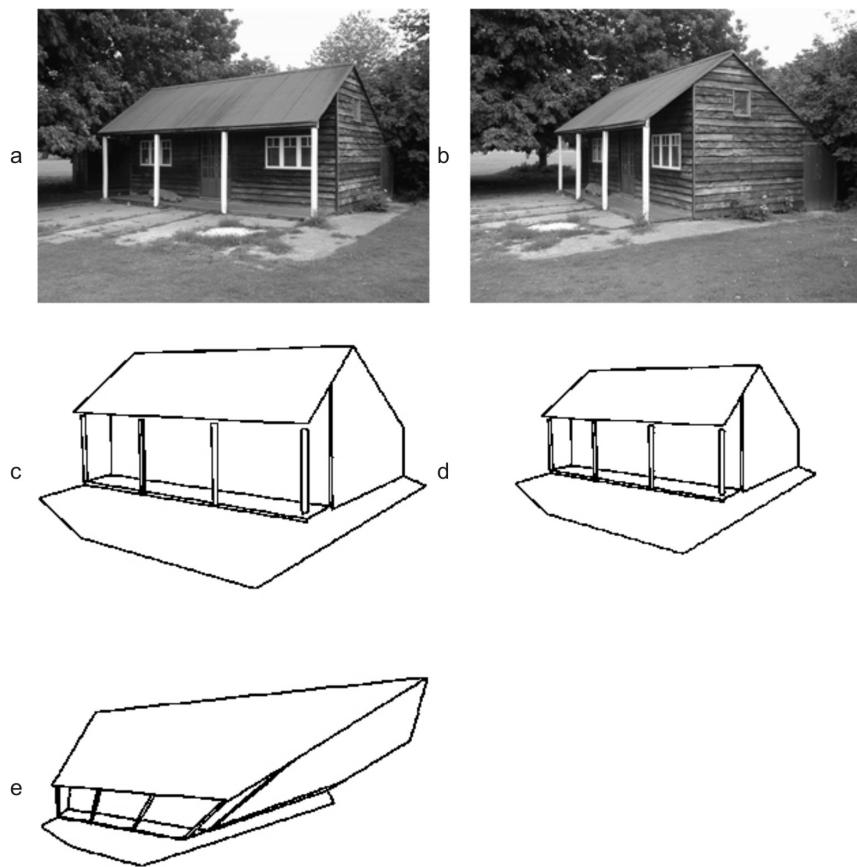
**Figura 13:** Correspondência esparsa por SIFT

Sendo que métodos por características empregam comparações simples entre os atributos das características, eles são geralmente rápidos e mais precisos do que métodos de correspondência densa. Porém, algumas vezes faz-se necessário empregar um método de interpolação para as áreas sem características, especialmente em casos onde a reconstrução de superfície é desejada [31].

## 2.6 Reconstrução

A reconstrução está relacionada à recuperação de informação de profundidade a partir de pontos correspondentes identificados nas duas imagens estéreo e dos parâmetros

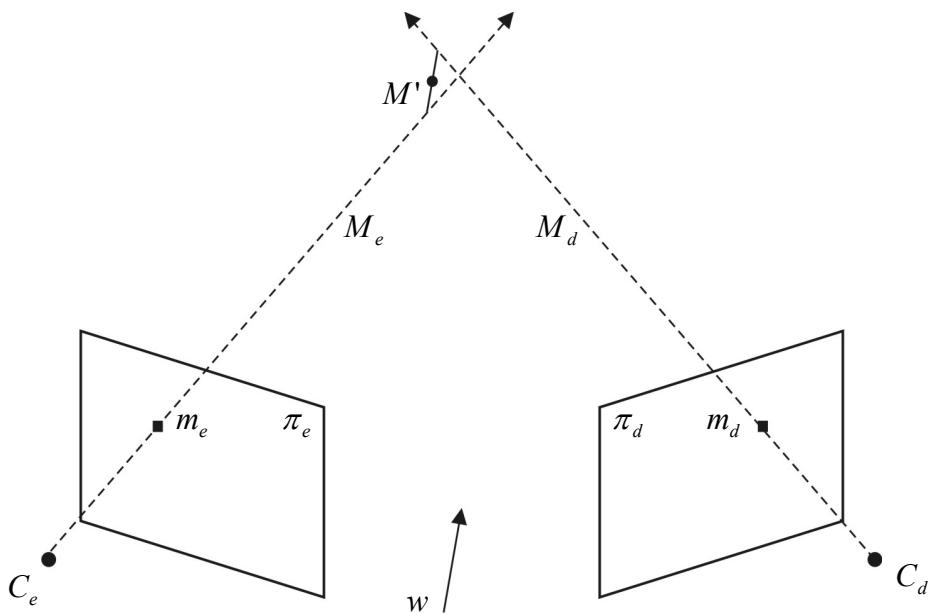
que descrevem o sistema de visão estéreo. O que pode ser reconstruído depende diretamente do que é conhecido a respeito da cena e do sistema de visão estereoscópica [14]. A Figura 14 ilustra a reconstrução da estrutura tridimensional de uma cena observada sob diferentes pontos de vista (a,b) para os diferentes cenários possíveis. Se tanto parâmetros intrínsecos como extrínsecos são conhecidos, podemos realizar a reconstrução métrica (c). Se somente parâmetros intrínsecos são conhecidos, podemos realizar a reconstrução sujeita a um fator de escala (d). No caso onde nem parâmetros intrínsecos e nem parâmetros extrínsecos são conhecidos e somente temos a informação de pontos correspondentes, podemos somente realizar a reconstrução sujeita a uma transformação projetiva global desconhecida (e).



**Figura 14:** Reconstrução projetiva e métrica ( Fonte: HARTLEY e ZISSERMAN [14] )

TRUCCO et al. [12] descrevem o processo de reconstrução por triangulação para o caso onde tanto parâmetros intrínsecos como extrínsecos são conhecidos. Conforme exibido na Figura 6, o ponto  $M$ , projetado no par de pontos correspondentes  $m_e$  e  $m_d$  encontra-se na intersecção dos raios  $M_e$  e  $M_d$ , representando os raios que intersectam a

imagem nos pontos  $m_e$  e  $m_d$  e intersectam os centros de projeção  $C_e$  e  $C_d$ , respectivamente. Estes raios são conhecidos e sua intersecção pode ser calculada. Entretanto, sendo que os parâmetros e posição das imagens são conhecidos apenas aproximadamente, os dois raios não intersectam realmente. Desta forma, a intersecção deve ser estimada como o ponto de mínima distância entre os dois raios.



**Figura 15:** Reconstrução de ponto por triangulação

Seja  $am_e$  ( $a \in \Re$ ) o raio  $M_e$ , e seja  $t + bR^T m_d$  ( $b \in \Re$ ) o raio  $M_d$ , expressos no sistema de coordenadas esquerdo. Seja  $w = m_e \times R^T m_d$  um vetor ortogonal tanto a  $M_e$  quanto a  $M_d$ . O problema encontra-se em determinar o ponto médio,  $M'$  do segmento  $s$  paralelo a  $w$  que liga  $M_e$  e  $M_d$ , conforme ilustrado pela Figura 15.

Para isso, podemos determinar os pontos finais do segmento  $s$ ,  $a_0m_e$  e  $t + b_0R^T m_d$ , resolvendo o sistema de equações lineares dado por:

$$am_e - bR^T m_d + c(m_e \times R^T m_d) = t, \quad (51)$$

para  $a_0$ ,  $b_0$  e  $c_0$ . O ponto triangulado  $M'$  pode então ser calculado diretamente como o ponto médio do segmento  $s$ .

Observe que a reconstrução descrita considera a triangulação a partir de pontos com coordenadas referentes à geometria original do sistema de visão estéreo. FUSIELO et al. [21] descrevem como é possível realizar a reconstrução diretamente a partir de coordenadas retificadas.

## 2.7 Considerações Finais

Neste capítulo foram discutidos modelos e técnicas relacionadas à estereoscopia. Seu conteúdo corresponde à fundamentação teórica que orientou o desenvolvimento do presente trabalho, fornecendo o ferramental necessário para tratar problemas como calibração, correspondência e reconstrução.

No capítulo seguinte será descrito o emprego destes modelos e técnicas em uma aplicação prática, no contexto do desenvolvimento de uma infra-estrutura básica para sistemas de visão estereoscópica para robótica móvel. Neste contexto, serão discutidas a escolha de métodos para cada problema, os detalhes de sua implementação, assim como os resultados obtidos por sua aplicação.

## Capítulo 3

# Desenvolvimento de uma Biblioteca para Sistemas de Visão Estereoscópica para Robótica Móvel

Neste capítulo é descrito o desenvolvimento de uma biblioteca de software para sistemas de visão estereoscópica que emprega um conjunto de modelos e métodos apresentados no capítulo anterior com o objetivo de disponibilizar uma infra-estrutura básica para o desenvolvimento de sistemas de visão para robótica móvel.

Inicialmente serão descritos os requisitos levantados segundo a ótica de robótica móvel, que guiaram o desenvolvimento da biblioteca. O aparato utilizado para o desenvolvimento e para a realização das experimentações será posteriormente detalhado. Em seguida serão descritas as principais funcionalidades disponibilizadas a fim de atender os requisitos levantados. Será exibida a organização das funcionalidades, através dos principais modelos que especificam a arquitetura da biblioteca. Finalmente, será discutida a utilização da biblioteca por aplicações cliente.

### 3.1 Especificação e Análise dos Requisitos

O presente trabalho procurou disponibilizar uma infra-estrutura básica para desenvolvimento de sistemas de visão estereoscópica para robótica móvel. Embora o foco em robótica móvel reduza o número de aplicações a serem atendidas pela biblioteca a ser desenvolvida, a gama de requisitos continua variada.

Procurou-se criar uma biblioteca independente da aplicação de robótica móvel cliente. Para delinear os limites dos requisitos a serem atendidos considerou-se que a biblioteca, em conjunto com o hardware necessário, deve agir como um *dispositivo* de

captura cuja saída seja informação tridimensional suficiente para operações gerais de robótica móvel. Desta forma, as aplicações cliente ficam responsáveis por tratar a informação fornecida pelo *dispositivo* a fim de atender seus objetivos específicos.

Convencionou-se durante o levantamento dos requisitos e seleção das técnicas a serem utilizadas que a informação gerada pelo *dispositivo* acima descrito deveria ser suficiente para a realização do ciclo proposto por KRAGIC et al. [9], que sumariza a operação geral de robôs móveis como uma combinação das etapas de *detecção, aproximação e manipulação*.

Com base nas considerações descritas, foram selecionados os seguintes requisitos funcionais a serem atendidos pela biblioteca:

- 1) O sistema deve permitir a calibração métrica e a persistência dos parâmetros do sistema de visão estéreo.
- 2) O sistema deve gerar mapas de disparidade densos.
- 3) O sistema deve permitir a reconstrução métrica da cena observada.
- 4) O sistema deve permitir a operação empregando diferentes fontes de dados.

Sabe-se que algumas aplicações de robótica móvel, como as que assumem uma abordagem quantitativa [13], não necessitam calibração ou reconstrução métrica. Entretanto, um grande número de aplicações é baseado em abordagens métricas. Desta forma, os requisitos de calibração e reconstrução métricas basearam-se na consideração de que ao se introduzir a informação métrica são atendidos os dois grupos de aplicações.

O requisito da geração de mapas de disparidade densos foi baseado na suposição de que, ao contrário de mapas esparsos, mapas densos fornecem informação suficiente para a realização de operações gerais de robótica móvel, descritas pelo ciclo *detectar, aproximar e manipular*. Além disso, considerou-se que o mapa de disparidade esparso é subconjunto de um mapa de disparidade denso gerado pelo mesmo par de imagens estéreo, e que desta forma, a abordagem densa deve atender inclusive às aplicações que eventualmente empregariam pontos característicos.

Além disso, considerou-se que robôs autônomos móveis precisam de um mínimo de detalhes da cena para realizar um ciclo de operação completo. Informação sobre volume, tamanho e forma são necessários para tarefas de reconhecimento, para orientar a navegação através do ambiente e para manipular objetos. Apesar de apresentarem um custo computacional mais elevado em relação à correspondência esparsa, a correspondência densa mostra-se uma opção cada vez mais viável devido ao avanço das arquiteturas de hardware e propostas de algoritmos mais eficientes.

O requisito relacionado à operação com diferentes fontes de dados foi levantado considerando que o desenvolvimento de um grande número de algoritmos para robótica móvel é realizado de maneira simulada. Este requisito prevê a disponibilização de meios para que a *aplicação cliente* possa ser desenvolvida utilizando bases de imagens de referência, simuladores, ou mesmo dispositivos reais de captura.

Adicionalmente, foram levantados os seguintes requisitos não funcionais a serem atendidos pela biblioteca:

- 1) A operação do sistema deve atender a requisitos de tempo real.
- 2) O sistema deve ser disponibilizado para Windows e para GNU Linux.
- 3) Empregar somente componentes ou ferramentas livres.
- 4) O sistema deve permitir extensão funcional.

O requisito de tempo real foi levantado em função de estar presente na grande maioria das aplicações de robótica móvel. Pela disponibilização para diferentes sistemas operacionais e pelo emprego de ferramentas livres procura-se atender às exigências de um maior número de usuários, principalmente da comunidade acadêmica. Finalmente, pelo requisito de extensão funcional pretende-se permitir uma maior flexibilidade para que os usuários da biblioteca possam customizar e estender as funcionalidades já providas pela biblioteca.

### 3.2 Aparato Experimental e Ambiente de Desenvolvimento

Durante a realização do presente trabalho foi utilizado um conjunto de ferramentas, construídas ou selecionadas com o objetivo de oferecer suporte ao desenvolvimento e experimentação.

Para auxiliar o desenvolvimento da biblioteca, foi montado um sistema de aquisição de imagens binocular, composto por um par de câmeras e lentes, placa de aquisição e um computador. O conjunto binocular é rígido, no sentido de que não apresenta qualquer controle motorizado sobre a posição e orientação das câmeras ou sobre as lentes. Este modelo de conjunto binocular foi selecionado pela facilidade de montagem e pela disponibilidade dos componentes empregados na construção do mesmo. Adicionalmente, visto que este modelo simplificado representa o tipo de conjunto binocular utilizado em muitos trabalhos da literatura, sua utilização no presente trabalho colabora para com a compatibilidade da biblioteca para novos trabalhos.

A Figura 16 apresenta o sistema de aquisição binocular construído.



**Figura 16:** Sistema binocular de aquisição de imagens

As especificações das câmeras, lentes, e placa de aquisição utilizados na construção do sistema de aquisição de imagens binocular são apresentadas respectivamente

pela Tabela 1, Tabela 2 e Tabela 3. Este sistema foi conectado a um computador de arquitetura x86 contendo um processador Pentium 4 de 3.0GHz e 1GB de memória RAM.

Os requisitos não funcionais que descrevem a necessidade de disponibilização da biblioteca para o ambiente Windows e GNU Linux, e que somente deveriam ser empregados componentes livres orientou a composição de um ambiente de desenvolvimento através da seleção um conjunto de ferramentas abertas. Os requisitos não funcionais que descrevem a necessidade de operação em tempo-real e possibilidade de extensão funcional orientaram a seleção da linguagem de programação orientada a objetos C++. O ambiente de desenvolvimento utilizado foi composto, principalmente, pelo IDE Eclipse<sup>1</sup>, Plug-in CDT<sup>2</sup> para fornecer suporte à linguagem C++ e a biblioteca de visão computacional OpenCV<sup>3</sup> para disponibilização das estruturas e funções básicas de manipulação de imagem. A modelagem, por sua vez, foi feita utilizando a linguagem UML através da ferramenta StarUML<sup>4</sup>.

<i>Parâmetro</i>	<i>Valor</i>
Fabricante	Watec
Modelo	WAT-508
Tipo de Sinal	Composto
Sensor	1/3 in / CCD
Número total de <i>pixels</i>	510 (H) x 492 (V)
Tamanho unitário do <i>pixel</i>	9.6 µm (H) x 7.5 µm (V)
Iluminação mínima	0.03 lx
Dimensões	42.0 mm x 44.0 mm x 53.0 mm

**Tabela 1:** Especificação das câmeras utilizadas no sistema binocular

---

<sup>1</sup> <http://www.eclipse.org>

<sup>2</sup> <http://www.eclipse.org/CDT>

<sup>3</sup> <http://sourceforge.net/projects/opencvlibrary>

<sup>4</sup> <http://www.staruml.org>

<i>Parâmetro</i>	<i>Valor</i>
Fabricante	Computar
Modelo	M1614-MP
Distância focal	16 mm
Campo de Visão	2/3" (38.0°, 30.8°, 23.4)
Distorção	-0.3 %
Foco/Íris	Manual
Dimensões	f33.5 mm x 28.2 mm

**Tabela 2:** Especificação das lentes utilizadas no sistema binocular

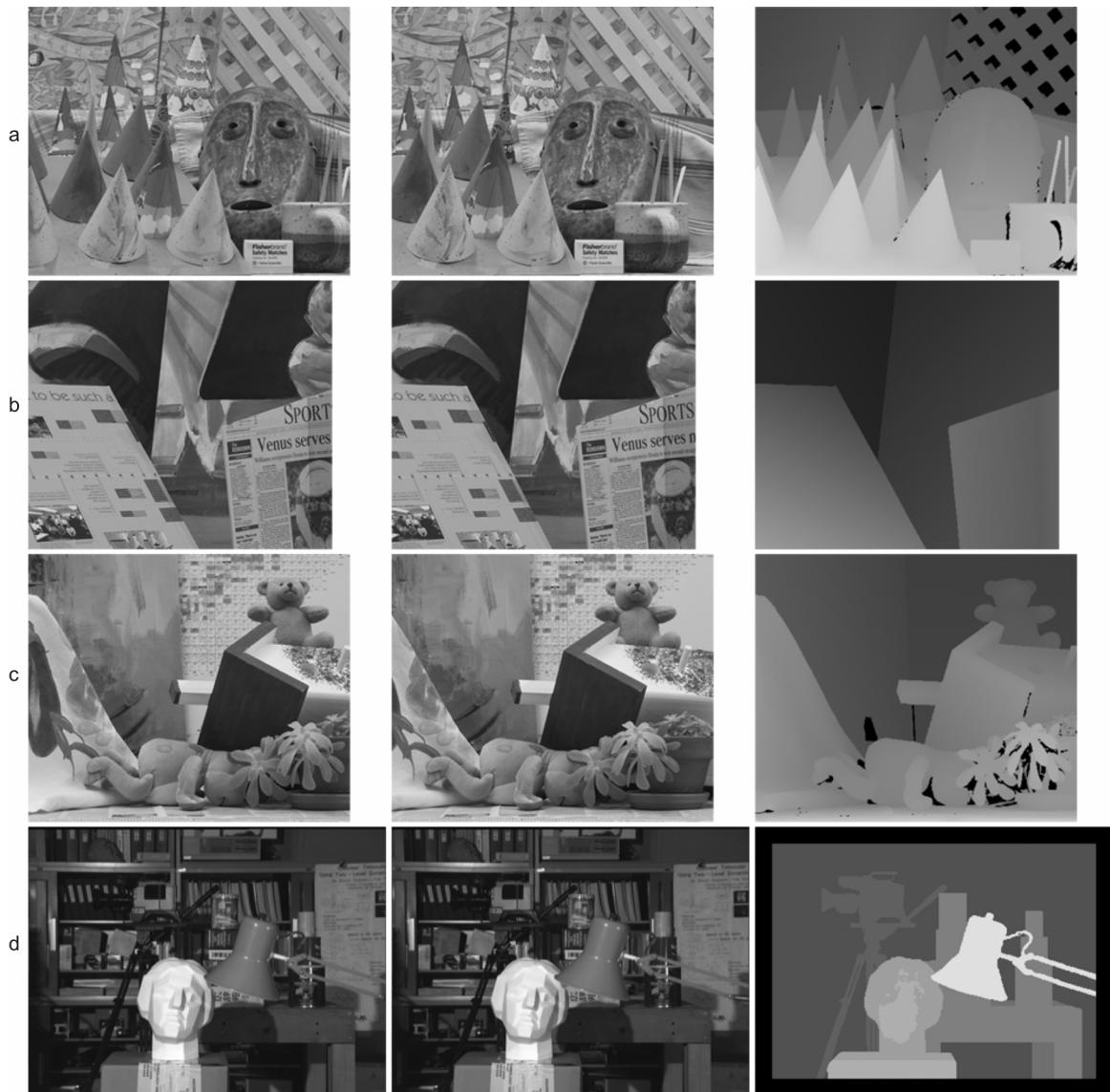
<i>Parâmetro</i>	<i>Valor</i>
Fabricante	EureSys
Modelo	Picolo Pro 2
Sinal	Composto (4 entradas BNC)
Resolução	640 x 480
Formato	PAL, NTSC, CCIR, EIA
Microprocessador	BT848
Taxa de Aquisição Max.	30 fps
Conexão	PCI (32 bits / 33 Mhz / 5V)

**Tabela 3:** Especificação da placa de aquisição utilizada no sistema binocular

Durante o desenvolvimento do trabalho, principalmente no que diz respeito aos algoritmos de correspondência, é preciso utilizar pares de imagens estéreo devidamente retificados. A aquisição do pares de imagens através do sistema de aquisição binocular durante o desenvolvimento ou realização de testes demanda um tempo adicional relativo à calibração e posterior retificação dos mesmos. A criação de uma base própria de imagens estéreo mostrou-se apropriada, a princípio. Entretanto, a falta dos mapas de disparidade reais

tornou a avaliação dos resultados muito subjetiva. A construção de mapas de disparidade de referência mostrou-se inviável.

Por este motivo, foi selecionada para o desenvolvimento do trabalho uma base de imagens estéreo de referência vastamente utilizada na literatura [20], composta de 4 pares de imagens estéreo. Os pares de imagens estéreos são livres de ruído, são devidamente normalizados e retificados, além de possuírem um mapa de disparidade real associado. Isso tornou a avaliação dos algoritmos de correspondência densa menos subjetiva, e permitiu o confrontamento dos resultados obtidos com outros trabalhos da literatura que utilizam a mesma base de referência.



**Figura 17:** Imagens da base de referência e respectivo mapa de disparidade real

Os pares de imagens estéreo de referência utilizados são exibidos com seu respectivo mapa de disparidade real na Figura 17, e os mesmos são denominados como Cones (a), Vênus (b), Teddy (c) e Tsukuba (d).

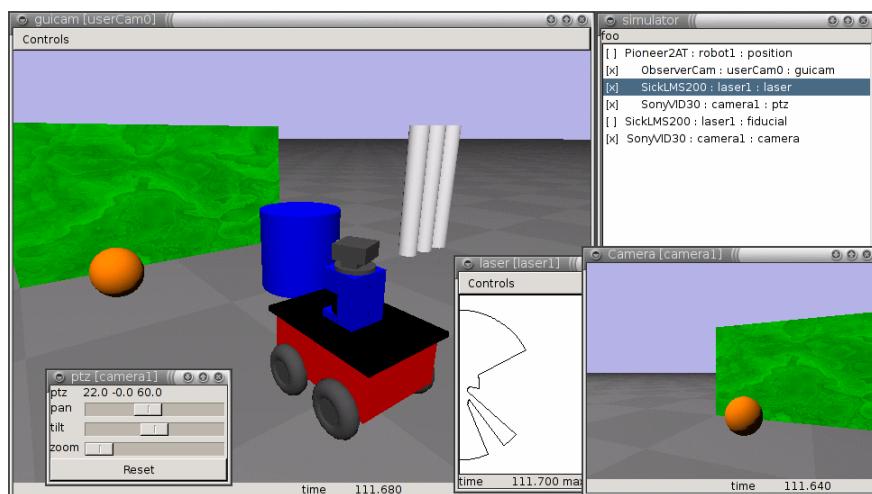
A Tabela 4, por sua vez, apresenta os parâmetros associados a cada par estéreo da base de referência.

<i>Par</i>	<i>Altura</i>	<i>Largura</i>	<i>Disp. Mínima</i>	<i>Disp. Máxima</i>
Cones	375	450	0	59
Venus	383	434	0	19
Teddy	375	450	0	59
Tsukuba	288	384	0	20

**Tabela 4:** Parâmetros dos pares estéreo de referência

As imagens dos pares estéreo de referência apresentam diferentes características como presença de objetos complexos, objetos em diferentes profundidades gerando variadas oclusões, regiões sem textura ou com intensidade homogênea, entre outras. Desta forma, torna-se ideal para a avaliação de algoritmos de correspondência sob diferentes critérios.

Para alguns dos experimentos relacionados à operação de robôs móveis orientada pelo sistema de visão estereoscópica utilizou-se o ambiente de simulação robótica Gazebo<sup>5</sup>.



**Figura 18:** Ambiente de simulação para robótica móvel

<sup>5</sup> <http://playerstage.sourceforge.net>

A Figura 18 apresenta algumas janelas da interface do simulador, que vem sendo utilizado em diversos trabalhos na área de robótica móvel.

Através desta ferramenta é possível realizar simulações realísticas em diferentes ambientes tridimensionais. O simulador utiliza elaboradas *engines* gráficas e físicas, gerando ambientes realísticos e permitindo interações físicas plausíveis com os objetos da cena. Nas simulações podem ser utilizados diferentes robôs, e estes podem por sua vez ser configurados para utilizarem diferentes tipos de sensores como câmeras, sonares, etc. A configuração de um robô móvel com duas câmeras permitiu a simulação de um sistema de visão estereoscópica de maneira satisfatória.

### **3.3 Funcionalidades Principais**

Nesta seção são descritas as principais funcionalidades implementadas para atender aos requisitos levantados para a biblioteca. Para cada funcionalidade será descrita a técnica empregada, assim como os resultados obtidos pelo emprego da mesma.

#### **3.3.1 Modelo de Câmera e de Sistema de Visão Estéreo**

De modo a fornecer as funções e estruturas básicas necessárias ao desenvolvimento de todas as outras funcionalidades da biblioteca, foram implementados os modelos de câmera e sistema de visão estereoscópica descritos na seção 2.1 e na seção 2.2, respectivamente.

Isto abrange todas as estruturas necessárias para manutenção dos parâmetros intrínsecos e extrínsecos, funções implementando as equações que descrevem a projeção de perspectiva e conversão de coordenadas entre os sistemas de coordenadas de imagem, câmera e ambiente.

#### **3.3.2 Calibração do Sistema de Visão Estéreo**

Conforme descrito na seção 2.1.4, através da calibração estereoscópica são determinados os parâmetros do conjunto binocular necessários para a operação do sistema de visão estéreo. Dentre os métodos de calibração estereoscópica disponíveis, foram

considerados tanto os enquadrados na abordagem clássica como na projetiva. Considerando que o conjunto binocular utilizado não altera seus parâmetros durante a operação, e observando o requisito de reconstrução métrica, optou-se por empregar a abordagem clássica para o presente trabalho.

A calibração estereoscópica clássica, por sua vez, exige que cada câmera seja calibrada independentemente. Dentre os diferentes métodos de calibração de câmeras avaliados foi selecionado o método proposto por ZHANG [16] descrito na seção 2.1.4. O gabarito de calibração empregado é simplificado e de fácil construção, consistindo em uma impressão em alta resolução em um papel de alta qualidade fixado sobre uma superfície plana e rígida. O padrão presente no gabarito é formado por um conjunto de quadrados pretos, organizados de maneira que os vértices formados pela junção dos mesmos são facilmente detectáveis. Outra vantagem apresentada pelo método é que os resultados obtidos são superiores a outros métodos [15], visto que durante a calibração são apresentadas várias amostras do gabarito obtidas em diferentes instantes e posições do ambiente. A calibração individual das câmeras por este método foi implementada empregando funções existentes na biblioteca OpenCV específicas para calibração monocular de câmeras, e segue a descrição de calibração clássica apresentada na seção 2.3.1.

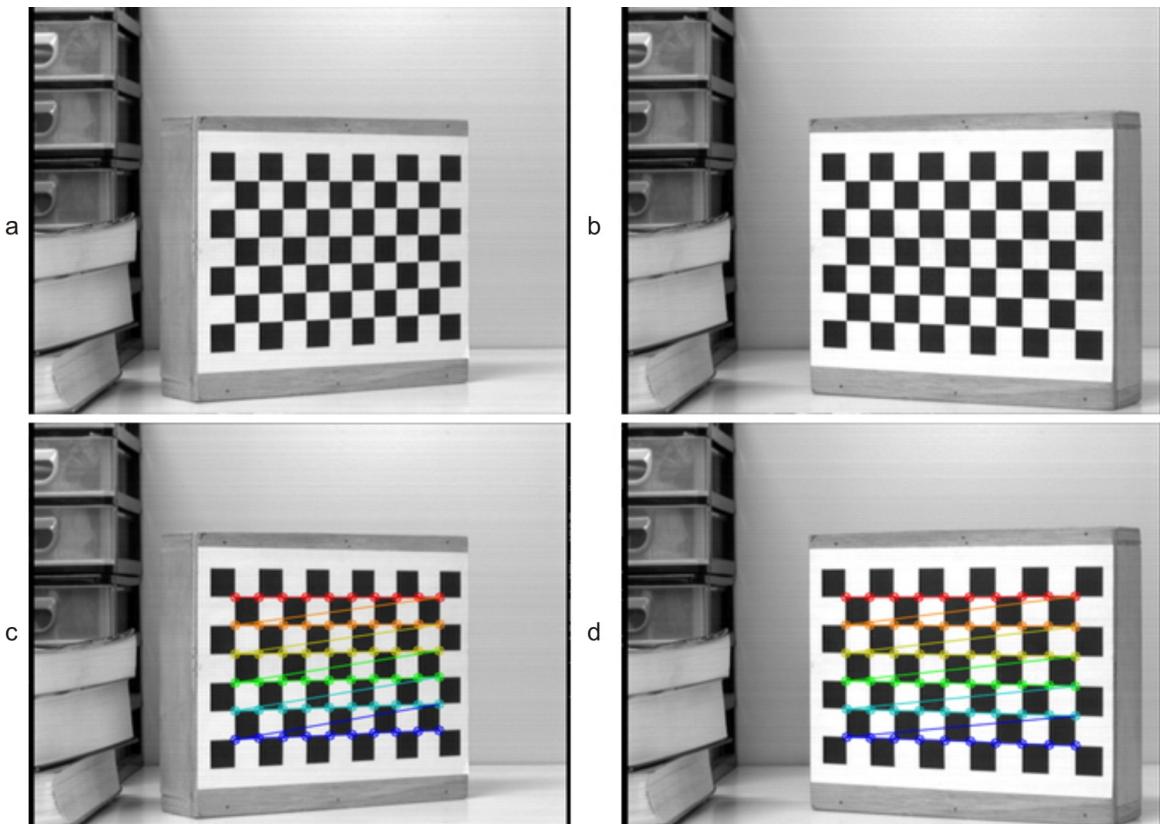
A rotina de calibração estéreo foi implementada neste trabalho como um passo inicial, que antecede a operação do sistema de visão. Sabe-se, entretanto, de que mesmo sendo o sistema binocular fixo, em aplicações de robótica móvel eventualmente será necessária uma nova calibração devido a possível ocorrência de alterações na configuração do sistema resultante de colisões ou vibrações.

O primeiro passo da rotina de calibração estéreo consiste na calibração de cada câmera do conjunto binocular de forma independente. Nesta etapa, são capturadas várias amostras do gabarito de calibração cuja geometria é conhecida. Cada amostra é capturada em diferentes instantes, e para cada uma delas o gabarito é posicionado em diferentes posições do ambiente. Conforme descrito por BENEVENUTI [18], deve-se apresentar o gabarito em diferentes profundidades de maneira a representar toda a faixa nas quais as câmeras irão operar. Este procedimento evita que os parâmetros de calibração sejam somente válidos em torno de uma dada posição. Observa-se ainda que na utilização de padrões coplanares, podem ser obtidos resultados discrepantes quando o gabarito está posicionado

aproximadamente paralelo ao plano de imagem. Por este motivo esta configuração é evitada, sendo sempre empregada uma angulação relativa entre o gabarito e o plano de imagem superior a 30 graus. A reconstrução de um cenário de calibração que segue estas recomendações é apresentada na seção 3.3.5.

Em cada amostra capturada são identificados os vértices internos do padrão presente no gabarito, correspondendo aos pontos característicos de interesse. Para cada ponto são associadas coordenadas do ambiente. A origem do ambiente é definida como sendo o primeiro vértice interno partindo do canto superior esquerdo do gabarito. As coordenadas  $X$  e  $Y$  dos outros vértices internos refletem a distância em milímetros em relação à origem. A coordenada  $Z$  é definida como 0.

Na Figura 19 são exibidas duas amostras do gabarito empregadas na calibração (a e b) e as respectivas imagens utilizadas para ilustrar a identificação dos vértices do padrão durante a execução da rotina (c e d).



**Figura 19:** Seqüência de amostras do gabarito de calibração

Para cada amostra, o conjunto de pares formados pelos pontos na imagem e suas respectivas coordenadas do ambiente são apresentados para a função que implementa a estimativa dos parâmetros intrínsecos e extrínsecos da câmera, que segue a descrição realizada na seção 2.1.4. Dentre os parâmetros intrínsecos obtidos temos a *distância focal* em unidade de *pixel*, as coordenadas da posição central da imagem em coordenadas de *pixel* e os coeficientes de distorção radial das lentes. Já dentre os parâmetros extrínsecos obtidos, temos a matriz de rotação e a matriz de translação.

Após a calibração individual de cada câmera, são calculados os parâmetros de calibração do sistema de visão estéreo. Conforme descrito na seção 2.2.1, o conjunto de parâmetros intrínsecos do conjunto estéreo corresponde aos parâmetros intrínsecos das duas câmeras que o compõe. Os parâmetros extrínsecos, por sua vez, são calculados diretamente a partir dos parâmetros extrínsecos das duas câmeras, empregando as equações ( 24 ) ( 25 ) no cálculo da matriz de rotação e da matriz de translação.

Finalmente, os parâmetros extrínsecos de cada câmera são alterados em função dos parâmetros extrínsecos do sistema estéreo, de modo que a origem do sistema de coordenadas do mundo esteja fixada na primeira câmera, ou seja:

$$R_e = [I], t_e = [0], R_d = R \text{ e } t_d = t. \quad ( 52 )$$

Conforme descrito na seção 2.2.2, para um dado ponto na imagem da esquerda temos associada uma reta epipolar na imagem da direita que deve conter o ponto correspondente, e vice-versa. Esta relação é estabelecida pela geometria epipolar, que para um dado sistema de visão estereoscópica encontra-se descrita pelos parâmetros obtidos na calibração. Sendo assim, podemos utilizar a distância existente entre os pontos e as retas epipoares onde estes deveriam aparecer como medida da qualidade obtida para uma dada calibração [14].

Este critério foi empregado para avaliação da qualidade obtida pelo método implementado. O sistema de visão estereoscópica foi calibrado três vezes, e para cada cenário de calibração foram utilizadas um mínimo de 5 amostras do gabarito de calibração. A Tabela 5 apresenta os parâmetros que descrevem cada um dos cenários de calibração.

Cenário	Distância de operação	Linha de base	Tamanho do padrão
1	1300mm ~ 1600mm	70mm	10 x 6
2	1600mm ~ 1900mm	70mm	10 x 6
3	1300mm ~ 2000mm	70mm	10 x 6

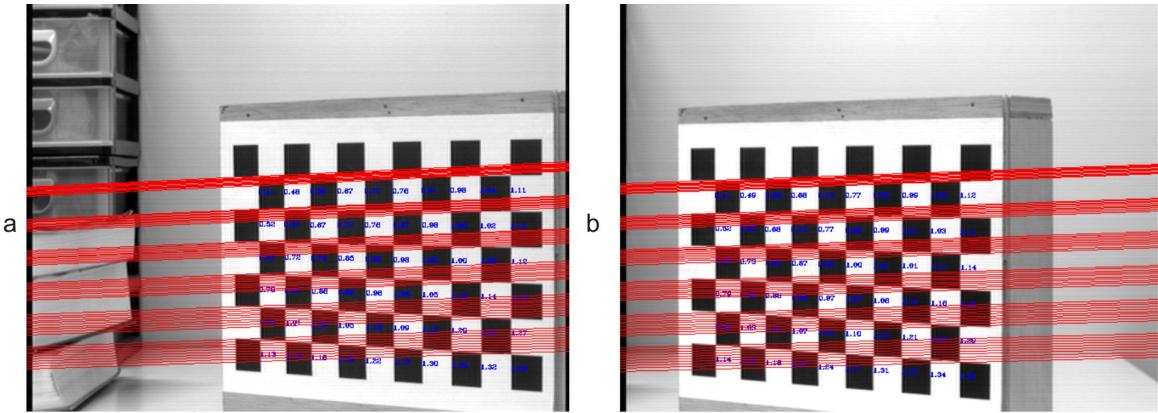
**Tabela 5:** Cenários de calibração

A Tabela 6 apresenta os parâmetros de calibração obtidos nos diferentes cenários, onde  $f$  e  $c$  estão expressos em unidades de *pixel* e onde  $R$  está expresso como um vetor de rotação.

Parâmetro	Cenário 1		Cenário 2		Cenário 3	
	E	D	E	D	E	D
$f_x$	2167,26	2219,61	2190,36	2203,52	2188,25	2232,02
$f_y$	2223,61	2274,98	2245,69	2256,97	2241,70	2282,49
$c_x$	279,96	254,56	407,39	362,54	317,94	278,72
$c_y$	179,92	151,67	259,90	270,27	219,74	196,47
$R$	$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0,01 \\ -0,01 \\ 8404,86 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1999,85 \\ -2044,88 \\ 8923,72 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} -0,01 \\ -6919,59 \\ 8480,14 \end{bmatrix}$
$t$	$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 71,92 \\ -4,07 \\ -14,88 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 74,23 \\ -0,92 \\ 11.11 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 70,44 \\ -2,59 \\ -8,22 \end{bmatrix}$

**Tabela 6:** Parâmetros do sistema de visão estereoscópica obtidos nos diferentes cenários

Para cada cenário, foram capturados vários pares de imagens estéreo do gabarito de calibração. Tomando a imagem da esquerda de cada par estéreo, para cada ponto presente no padrão foi calculada a linha epipolar sobre a imagem da direita onde o ponto correspondente deveria estar localizado. De maneira similar, foram calculadas as linhas epipolares para os pontos da imagem da direita, conforme apresentado pela Figura 20.



**Figura 20:** Cálculo das distâncias entre linhas epipoares e pontos

A distância entre cada ponto correspondente e a linha epipolar calculada foi então computada através da equação ( 53 ). Este procedimento foi repetido para todos os pares capturados.

$$d = \frac{|ax + by + c|}{\sqrt{a^2 + b^2}}. \quad ( 53 )$$

A Tabela 7 apresenta os resultados obtidos seguindo o critério de qualidade estabelecido, para todos os cenários de calibração. Os valores estão apresentados em unidade de *pixel*.

Parâmetro	Cenário 1	Cenário 2	Cenário 3
Distância Média	0,5209	0,8106	0,5254
Desvio Padrão	0,3907	0,0873	0,3478
Mínimo	0,00152	0,5404	0,0021
Máximo	1,7360	1,0366	1,3408

**Tabela 7:** Resultado da avaliação da qualidade da calibração

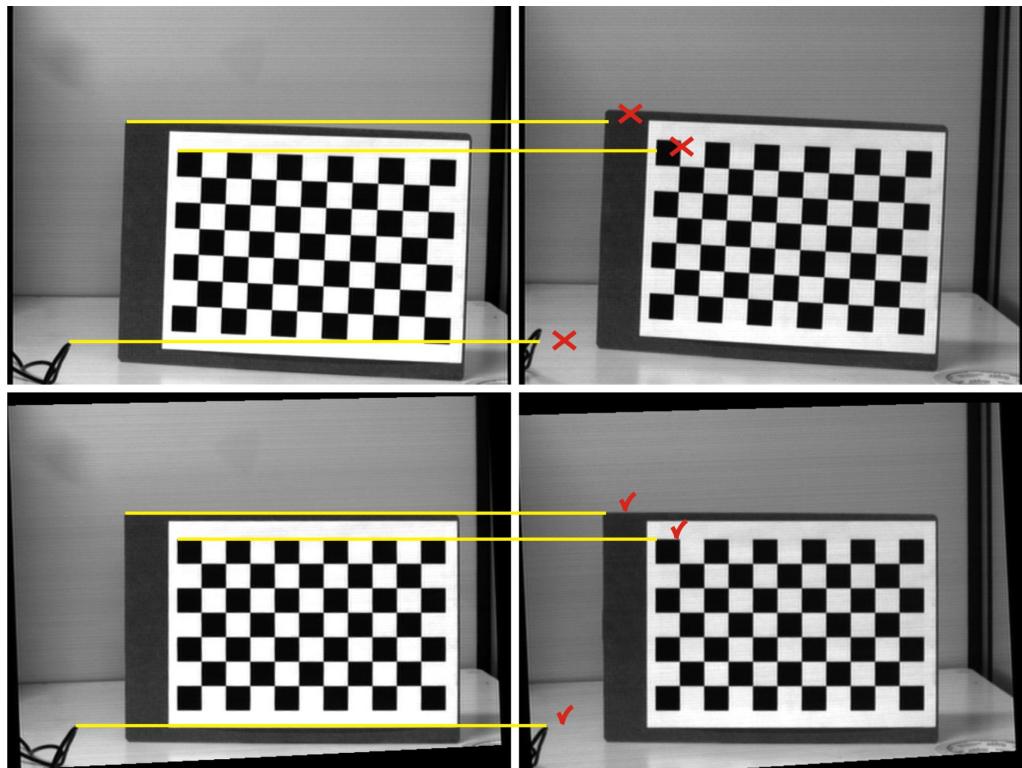
Os resultados obtidos na avaliação de qualidade da calibração foram condizentes aos resultados obtidos por outros trabalhos [18], e foram considerados satisfatórios.

### 3.3.3 Retificação do Par de Imagens Estéreo

Conforme descrito na seção 2.4, a retificação simplifica consideravelmente a complexidade do problema de correspondência ao alinhar os pontos correspondentes nas duas imagens de um par estéreo. Na verdade, a retificação é considerada como um passo preliminar para muitos algoritmos de correspondência [20]. Por este motivo, esta funcionalidade foi implementada para fornecer o suporte necessário à etapa de correspondência densa.

O algoritmo implementado emprega os parâmetros obtidos na calibração e segue os passos descritos na seção 2.4, alinhando os pontos correspondentes do par estéreo de imagens de forma a limitar a busca por pontos correspondentes entre as duas imagens ao longo da linha horizontal.

A Figura 21 apresenta um par de imagens não retificadas do gabinete de calibração (acima), e seu respectivo par de imagens retificadas (abaixo). As linhas que ligam os pares de imagens ilustram para diferentes pontos do par retificado como as correspondências podem ser feitas ao longo de um único eixo, o que não ocorre nas imagens não retificadas.



**Figura 21:** Retificação do par de imagens estéreo

A partir das correspondências feitas sobre as imagens retificadas, pode-se aplicar a transformação inversa à empregada para realizar a retificação de modo a restaurar as coordenadas originais dos pontos. Entretanto, é possível realizar a reconstrução diretamente a partir de pontos correspondentes retificados [21].

A retificação corresponde a uma transformação sobre as imagens que altera as coordenadas originais de cada um dos pontos presentes nas mesmas, e por este motivo pode modificar os resultados obtidos na reconstrução. Para mensurar esta influência, foram comparadas coordenadas de pontos correspondentes reconstruídos a partir de imagens não retificadas e suas respectivas versões retificadas. A diferença média obtida para 14 imagens do gabarito de calibração, sendo que para cada uma delas foram reconstruídos todos os vértices do padrão, foi de 0.6%. Tendo em vista a magnitude da diferença, considera-se que para aplicações de robótica móvel o compromisso entre desempenho e precisão deve normalmente levar a optar-se pela retificação, visto o ganho obtido pelo emprego da mesma para a etapa de correspondência.

### 3.3.4 Geração de Mapas de Disparidade

A geração de mapas de disparidade densos está relacionada ao processo de correspondência, considerado por muitos como o problema principal da estereoscopia. Por este motivo, representa o assunto para o qual maior esforço de pesquisa e desenvolvimento foi despendido neste trabalho.

Inicialmente, pela vasta utilização na literatura, foi selecionado o algoritmo de correlação por janela simples [12] como o método de correspondência densa a ser implementado e disponibilizado na biblioteca. Este algoritmo foi estudado e implementado. Entretanto, os primeiros resultados obtidos não foram satisfatórios. Os requisitos de tempo real não foram atendidos e os mapas de disparidade gerados não apresentaram a qualidade esperada.

Por este motivo, optou-se por realizar uma avaliação mais criteriosa sobre um conjunto de métodos, a fim de buscar a opção que melhor atendesse aos requisitos levantados.

Na próxima seção será descrito o conjunto de algoritmos implementados com base nos métodos encontrados na literatura. Em seguida, será descrita a proposta de um novo algoritmo de correspondência baseado em suporte regional. Finalmente, serão apresentados os resultados da comparação entre estes algoritmos.

### **3.3.4.1 Algoritmos Baseados na Literatura**

Foram selecionados métodos encontrados na literatura seguindo as abordagens global (2.5.1.2) e local (2.5.1.1), representando as duas grandes classes de algoritmo de correspondência densa. Estes métodos foram escolhidos com base nos resultados divulgados por avaliações recentes encontradas na literatura [20].

Para representar a abordagem local foram selecionados e implementados os algoritmos de correlação por janela simples [12] e por múltiplas janelas [22]. Estes métodos têm sido amplamente utilizados por sua simplicidade e bons resultados obtidos [13,28].

Verificou-se nos primeiros experimentos realizados que os algoritmos implementados apresentavam um custo computacional elevado, e por consequência proibitivo para aplicações de robótica móvel. Este resultado mostrou-se acentuado quando janelas grandes foram utilizadas. Para resolver o problema realizou-se uma pesquisa em busca de possíveis otimizações dos algoritmos (vide seção 2.5.1.1).

Com base nas técnicas de otimização levantadas novas versões dos algoritmos foram implementadas. A inclusão do cálculo otimizado da função de custo que leva em consideração a sobreposição de janelas vizinhas foi suficiente para fazer com que as novas versões atendessem aos requisitos de tempo real, sem prejudicar os resultados qualitativos obtidos nas versões originais dos algoritmos. Por este motivo, as otimizações possíveis através de instruções específicas da arquitetura de processadores não foram implementadas. Sabe-se, porém, que é possível diminuir o tempo de processamento em até duas vezes empregando estas técnicas [25].

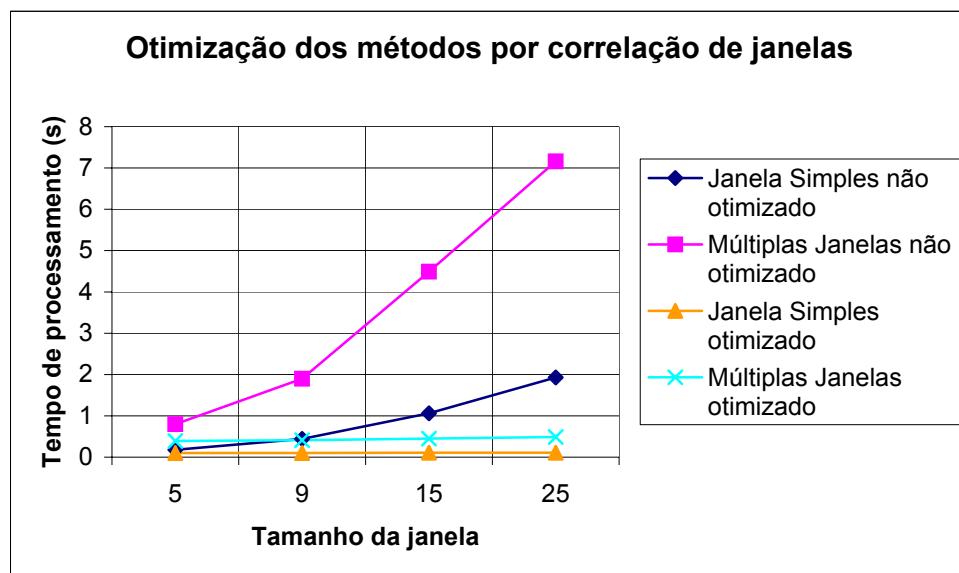
Na Tabela 8 são exibidos os tempos em segundos obtidos no processamento do par estéreo Tsukuba empregando os algoritmos otimizados e não otimizados. A melhoria obtida pela otimização mostra-se significativa quando comparados os tempos de processamento

apresentados pela versão não otimizados dos algoritmos e suas respectivas versões otimizadas.

Método	Tamanho da janela			
	5	9	15	25
Janela Simples não otimizado	0,18	0,44	1,06	1,93
Janela Simples otimizado	0,11	0,11	0,12	0,12
Múltiplas Janelas não otimizado	0,8	1,9	4,49	7,16
Múltiplas Janelas otimizado	0,39	0,41	0,45	0,49

**Tabela 8:** Tempos de processamento (s) obtidos com algoritmos de janela

Pelo Gráfico 1 pode-se observar que a otimização que considera a sobreposição das janelas apresenta ainda a vantagem de tornar o tempo de computação consumido pelas técnicas de correlação por janelas muito menos dependente do tamanho das mesmas. Isto porque, independente do tamanho das janelas, um número fixo de *pixels* será utilizado no cálculo da função de custo. Isso é muito importante, visto que muitas vezes janelas grandes podem ser mais interessantes, para aplicações onde uma grande precisão nos detalhes das bordas não é necessária.



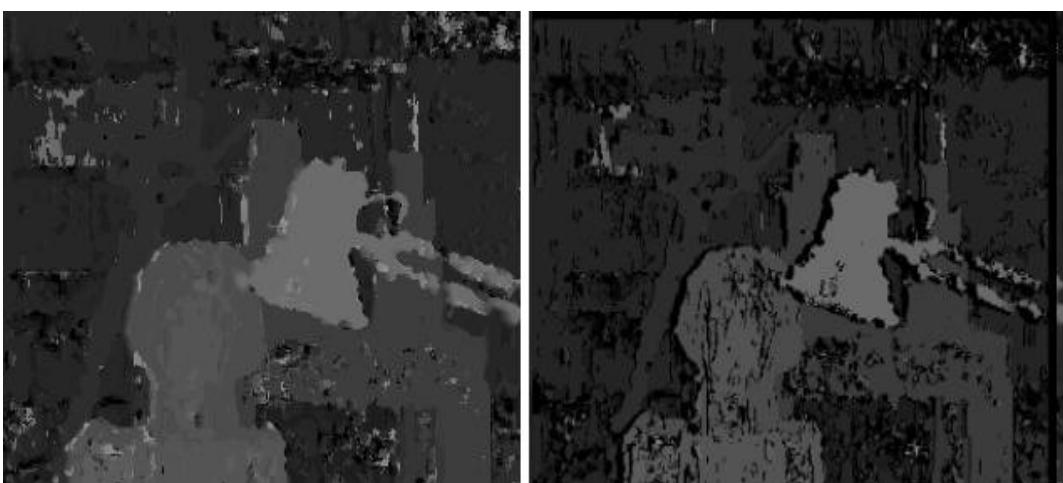
**Gráfico 1:** Tempos de processamento de algoritmos por correlação

Para representar a abordagem global foram selecionados os métodos Programação Dinâmica [27] e Cooperativo [32]. Para o método Programação Dinâmica foi utilizada uma implementação incorporada na biblioteca OpenCV. Este algoritmo apresentou nos testes iniciais um ótimo desempenho e mapas de disparidade satisfatórios. Além disso, a implementação já incorpora otimização e apresenta um baixo custo computacional.

Adicionalmente, foi avaliado o algoritmo Cooperativo, utilizando uma implementação disponibilizada pelo autor. Este algoritmo gera mapas de disparidade com um nível de detalhes satisfatório. Entretanto, o alto custo computacional obtido nas experimentações torna proibitiva a sua utilização para robótica móvel. Este resultado é condizente com os divulgados em [32], e como não foram encontrados meios de otimizar a implementação, o algoritmo foi desconsiderado.

Tanto algoritmos locais quanto globais apresentaram correspondências arbitrárias. Para minimizar o problema, foi incorporado a todos os algoritmos o método de verificação de consistência *left-right check*, descrito na seção 2.5.1.1. Esta consistência elimina um grande número de correspondências incorretas, e a sua utilização é valida considerando-se que é melhor não possuir informação de profundidade de um ponto do que ter a informação incorreta.

A Figura 22 mostra um mapa de disparidade gerado por abordagem local e o respectivo mapa de disparidade obtido empregando a verificação de consistência *left-right check*.



**Figura 22:** Verificação de consistência *left-right check*

Como resultados do trabalho descrito, foram disponibilizados na biblioteca diferentes algoritmos de correspondência densa incorporando métodos encontrados na literatura. Os algoritmos e suas respectivas descrições são apresentados na Tabela 9.

<i>Algoritmo</i>	<i>Técnica</i>	<i>Abordagem</i>
WTA SAD SW	Correlação de janela simples com função de custo SAD, seleção WTA, consistência left-right check e otimização.	Local
WTA SAD MW	Correlação de múltiplas janelas com função de custo SAD, seleção WTA, consistência left-right check e otimização.	Local
DP	Programação dinâmica com função de custo baseada em recompensas, penalidades e dissimilaridade e com consistência left-right check e otimização.	Global

**Tabela 9:** Algoritmos de correspondência baseados na literatura

Na seção seguinte é proposto um novo método de correspondência local desenvolvido durante a realização deste trabalho.

### 3.3.4.2 Método Proposto: Correspondência Local com Suporte Regional

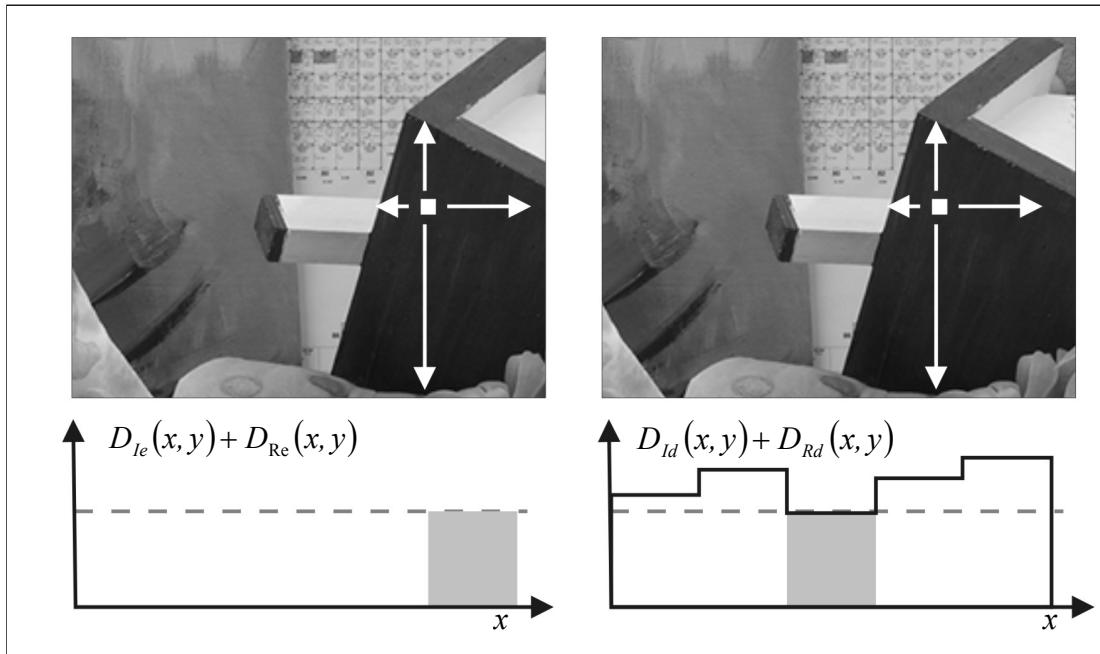
No processo de correspondência baseado em correlação de janelas apresentado na seção 2.5.1.1, temos que a intensidade assume o papel de descritor para cada ponto da imagem. Porém, que uma descrição baseada em intensidade não é suficiente para identificarmos univocamente um par de pontos correspondentes. O mesmo ocorre com o sistema visual humano quando observamos uma superfície extremamente lisa ou em um ambiente pouco iluminado. Nestas condições, onde temos a falta de detalhes finos para que possamos orientar a correspondência entre pontos observados pelos olhos, naturalmente recorremos à informação fornecida por elementos de alta freqüência presentes na região periférica para determinar a profundidade.

O método proposto procura modelar este comportamento. A idéia principal é descrever cada ponto das imagens tanto por um descritor de intensidade como por um descritor regional. O descritor regional deve agregar informação periférica proveniente dos componentes de alta freqüência existentes próximos ao ponto observado. Desta forma, podemos definir a função de custo com suporte regional por:

$$C_R(x, y, d) = S_I(x, y, d) + \alpha S_R(x, y, d). \quad (54)$$

onde  $S_I(x, y, d)$  é uma função de similaridade baseada em um descritor de intensidade  $D_I(x, y)$ ,  $S_R(x, y, d)$  correspondente a uma função de similaridade baseada em um descritor regional  $D_R(x, y)$  e  $\alpha$  correspondendo ao peso empregado à função de suporte regional.

A Figura 23 ilustra o efeito da descrição de pontos com suporte regional durante o processo correspondência. Diferente da situação ilustrada pela Figura 11, neste caso são empregados tanto o descritor baseado em intensidade  $D_I(x, y)$  como o descritor regional  $D_R(x, y)$ . É possível observar como uma descrição baseada nesta combinação de descritores permite uma melhor distinção entre os pontos. Uma função de custo baseada nestes descritores permite uma melhor seleção das correspondências corretas.



**Figura 23:**Descrição baseada em intensidade com suporte regional

Uma função de similaridade  $S_R(x, y, d)$  baseada em descritores regionais é dada por:

$$S_R(x, y, d) = \sum_{i,j=-n}^n \sqrt{D_v(x, y, d)^2 + D_h(x, y, d)^2}. \quad (55)$$

Por sua vez,  $D_v(x, y, d)$  e  $D_h(x, y, d)$  são dados respectivamente por:

$$D_v(x, y) = \frac{(D_t(x, y))}{(D_t(x, y) + D_b(x, y))} - \frac{(D_t(x + d, y))}{(D_t(x + d, y) + D_b(x + d, y))} \quad (56)$$

e

$$D_h(x, y) = \frac{(D_l(x, y))}{(D_l(x, y) + D_r(x, y))} - \frac{(D_l(x + d, y))}{(D_l(x + d, y) + D_r(x + d, y))} \quad (57)$$

onde  $D_t(x, y)$ ,  $D_b(x, y)$ ,  $D_l(x, y)$  e  $D_r(x, y)$  são descritores regionais que representam a distância entre o ponto observado e o componente de alta freqüência mais próximo acima, abaixo, à esquerda e à direita. Observe que  $D_v(x, y, d)$  e  $D_h(x, y, d)$  são adimensionais, a fim de minimizar possíveis diferenças dimensionais introduzidas pela perspectiva na geração das imagens.

Os itens apresentados na seqüência apresentam as etapas gerais de um algoritmo de correspondência local com suporte regional. Estes passos foram utilizados na implementação de algoritmos incluídos na biblioteca.

#### *Cálculo dos descritores regionais*

Em função da inserção do termo de suporte regional na função de custo de correspondência, precisamos calcular os descritores regionais para cada uma das imagens. Estes descritores podem ser baseados na distância entre cada ponto e os componentes de alta freqüência que delimitam a região homogênea onde o mesmo está inserido. Para identificar estes componentes podemos fazer uso de filtros de borda. Porém, o resultado obtido por estes filtros é normalmente ruidoso.

É importante observar que podemos limitar o suporte regional às regiões onde a intersecção entre as janelas de correlação e componentes de alta freqüência é igual à zero, visto que o desempenho das funções de similaridade baseadas em intensidade é satisfatório nas regiões diferentes destas. Sendo assim, podemos empregar um método de segmentação baseado na variância das intensidades para isolar as regiões homogêneas desejadas.

A segmentação baseada no mapa de variância de intensidades mostra-se interessante pelo seu custo computacional reduzido quando comparado a outras técnicas, e pode ser calculado por:

$$\sigma_L^2(x, y) = \frac{1}{N^2} \sum_{i,j=-n}^n (L(x + j, y + i) - \mu_L(x, y))^2 \quad (58)$$

onde  $N = (2n + 1)^2$  e  $\mu L(x, y)$  é a média dos valores de intensidade [25].

A partir do mapa de variância calculado podemos aplicar uma limiarização global para obter uma imagem binária contendo as regiões homogêneas isoladas. A partir desta imagem binária podemos calcular mapas de distância correspondentes aos descritores regionais  $D_t(x, y)$ ,  $D_b(x, y)$ ,  $D_l(x, y)$  e  $D_r(x, y)$ .

A Figura 24 ilustra o processo de cálculo de descritores regionais apresentados acima, onde é apresentada a imagem original (a), o mapa de variância calculado sobre a imagem original (b), o resultado da binarização do mapa de variância (c) e o mapa de distância correspondente ao descritor  $D_t(x, y)$  (d), calculado sobre a imagem binária.



**Figura 24:** Cálculo dos descritores regionais

A partir dos descritores regionais de distância, podemos calcular os mapas dos descritores adimensionais  $D_v(x, y, d)$  e  $D_h(x, y, d)$  diretamente, conforme apresentado nas equações ( 56 ) e ( 57 ).

### *Correspondência*

A partir da equação ( 54 ) podemos criar várias funções de custo através da combinação entre diferentes funções de similaridade baseadas em intensidade e região. Uma observação importante é que, sendo a função de similaridade baseada em região apresentada pela equação ( 55 ) essencialmente uma função de correlação, temos que a mesma ajusta-se muito bem à funções de similaridade baseadas em intensidade da mesma natureza.

Desta forma, empregando a função de similaridade baseada em intensidade apresentada pela equação ( 48 ) e a equação de similaridade baseada em região apresentada pela equação ( 55 ), podemos reescrever a função de custo com suporte regional dada pela equação ( 54 ) como:

$$C_{sadr}(x, y, d) = \sum_{i,j=-n}^n |(D_{le}(x+j, y+i) - D_{ld}(x+d+j, y+i))| + \alpha \sqrt{D_v(x, y, d)^2 + D_h(x, y, d)^2} \quad (59)$$

Observe que o cálculo do termo de suporte regional é realizado utilizando os mapas de descritores regionais previamente calculados, conforme descrito na seção anterior.

Uma vez calculados os valores de custo no espaço de busca de disparidade, a correspondência correta é selecionada segundo a abordagem *WTA*.

### *Verificações de Consistência e Otimização*

Nos algoritmos implementados seguindo a abordagem de suporte regional pode ser incluída a otimização baseada no cálculo de funções de custo considerando a sobreposição de janelas. Adicionalmente, é importante observar que como a inclusão de suporte regional proposto neste trabalho não modifica o comportamento baseado em correlação nos algoritmos a que for aplicado, estes também podem se beneficiar das otimizações através de instruções específicas da arquitetura de processadores citadas na seção 2.5.1.1.

### *Variações do algoritmo*

A função de custo com suporte regional apresentada permite uma diversidade de combinações, resultantes da utilização de diferentes funções de similaridade baseadas em intensidade e de diferentes funções de suporte regional. Utilizando os componentes apresentados nesta seção, foram incluídos na biblioteca dois novos algoritmos empregando o método de suporte regional proposto, apresentados na Tabela 10.

<i>Algoritmo</i>	<i>Técnica</i>	<i>Abordagem</i>
WTA SADR SW	Correlação de janela simples com função de custo SAD com Suporte Regional, seleção WTA, consistência left-right check e otimização.	Local/Regional
WTA SADR MW	Correlação de múltiplas janelas com função de custo SAD com Suporte Regional, seleção WTA, consistência left-right check e otimização.	Local/Regional

**Tabela 10:** Variações de algoritmos propostos

#### 3.3.4.3 Resultados Experimentais

Nesta seção são apresentados os resultados experimentais obtidos pela comparação dos algoritmos propostos com os diferentes algoritmos de correspondência implementados com base nos métodos encontrados na literatura. Uma relação dos algoritmos utilizados na comparação é apresentada na Tabela 11.

Na avaliação foram utilizados os critérios de custo computacional e qualidade do mapa de disparidade. A realização de uma avaliação objetiva sobre imagens reais torna-se inviável pela dificuldade de criação de um mapa de disparidade de referência contendo a disparidade real para cada ponto. Por este motivo, ao invés de utilizar pares estéreos capturados pelo sistema de visão estereoscópica próprio, optou-se por utilizar a base de referência apresentada na seção 3.2.

Cada algoritmo foi utilizado para gerar mapas de disparidade tendo como entrada cada um dos pares estéreo de referência. Para todos os algoritmos foram utilizados os espaços de busca definidos pelo autor da base de referência. Para os algoritmos WTA SAD SW, WTA SADR SW, WTA SAD MW e WTA SADR MW foram utilizadas janelas de

tamanho 5. Para o algoritmo DP foram empregados os mesmos parâmetros para penalidades e recompensas recomendados pelo autor.

<i>Algoritmo</i>	<i>Técnica</i>	<i>Abordagem</i>
WTA SAD SW	Correlação de janela simples com função de custo SAD, seleção WTA, consistência left-right check e otimização.	Local
WTA SADR SW	Correlação de janela simples com função de custo SAD com Suporte Regional, seleção WTA, consistência left-right check e otimização.	Local/Regional
WTA SAD MW	Correlação de múltiplas janelas com função de custo SAD, seleção WTA, consistência left-right check e otimização.	Local
WTA SADR MW	Correlação de múltiplas janelas com função de custo SAD com Suporte Regional, seleção WTA, consistência left-right check e otimização.	Local/Regional
DP	Programação dinâmica com função de custo baseada em recompensas, penalidades e dissimilaridade, consistência left-right check e otimização.	Global

**Tabela 11:** Algoritmos avaliados

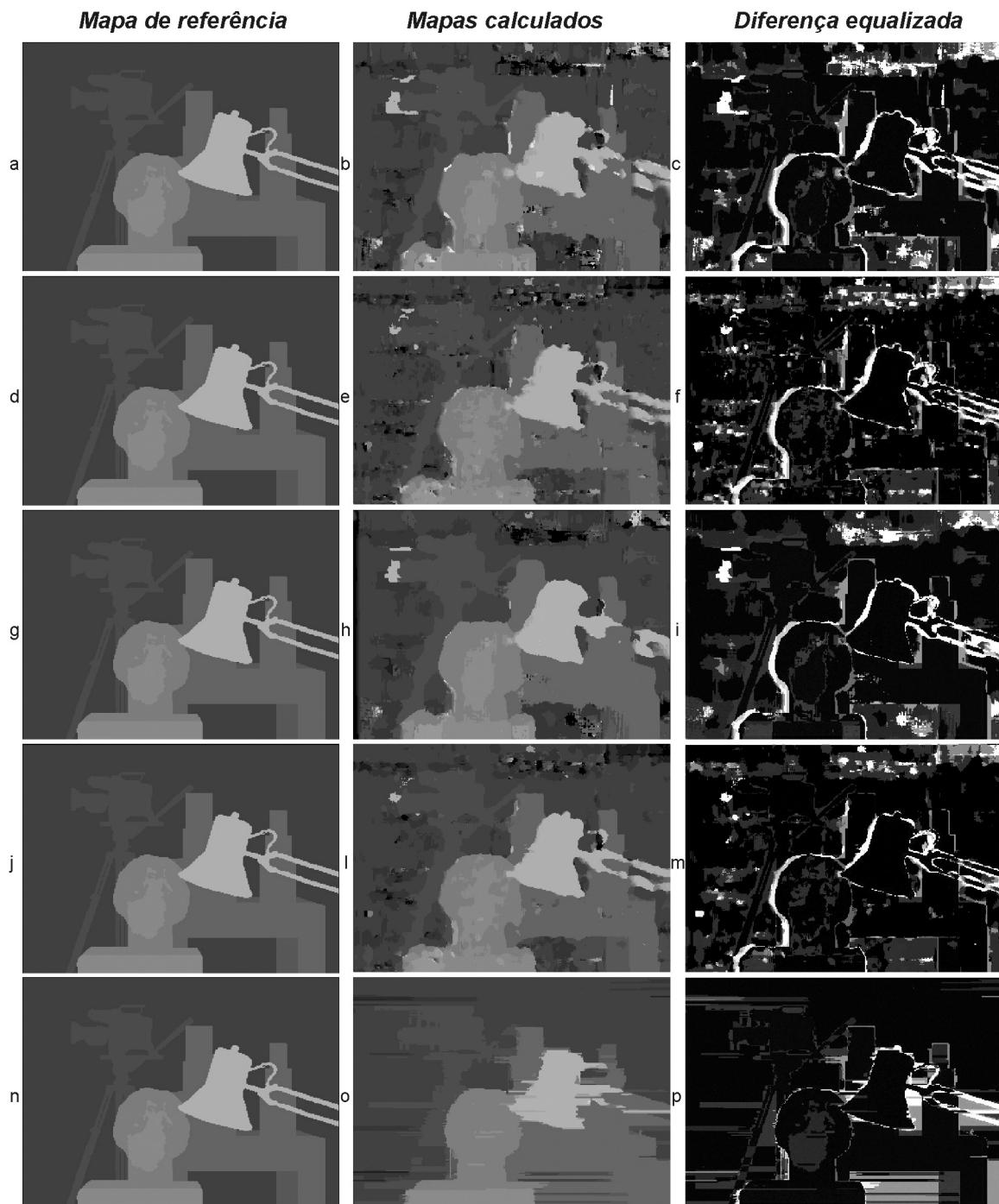
A Figura 25 exibe os resultados obtidos para o par de imagens Tsukuba, onde (a), (d), (g), (j) e (n) correspondem ao mapa de disparidade de referência, (b), (e), (h), (l) e (o) correspondem ao mapa de disparidade obtido por cada algoritmo, e onde (c), (f), (i), (m) e (p) correspondem à diferença equalizada entre o mapa de disparidade real e o mapa de disparidade obtido.

A comparação de desempenho dos métodos avaliados foi realizada empregando como medida de qualidade a porcentagem de correspondências incorretas [20], dada por:

$$E = \frac{1}{N} \sum_{x,y} |(d_c(x,y) - d_T(x,y)) > \delta_d| \quad (60)$$

onde  $\delta_d$  é um valor de tolerância ao erro da disparidade.

Na Tabela 12 são exibidos os erros gerais obtido no cálculo das disparidades para os diferentes algoritmos avaliados. Para o cálculo do erro geral foi empregado um valor de  $\delta_d$  igual a 2.



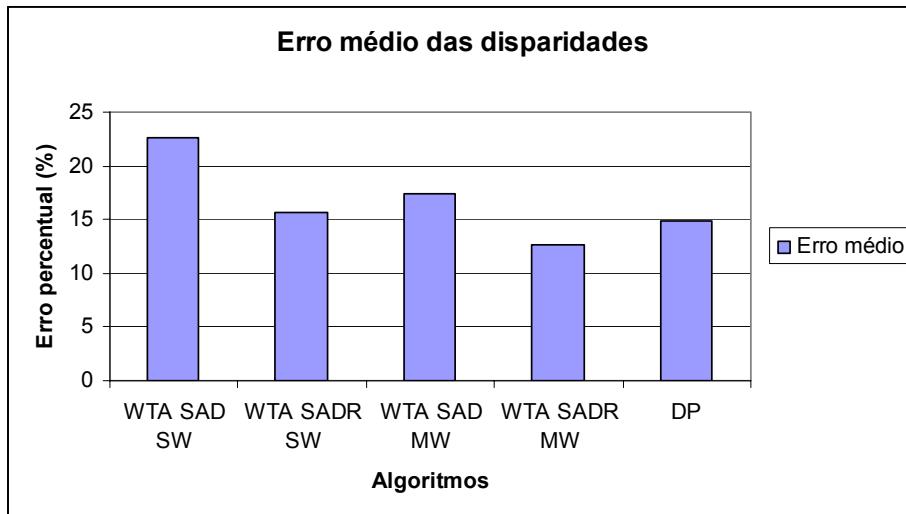
**Figura 25:** Diferença entre mapas calculados e mapas de referência

Avaliando os resultados obtidos entre os métodos WTA SAD SW e WTA SAD MW e as suas variações com suporte regional WTA SADR SW e WTA SADR MW, claramente é observado o ganho introduzido pelo método de suporte regional proposto no presente trabalho.

<i>Algoritmo</i>	<i>Tsukuba</i>	<i>Cones</i>	<i>Teddy</i>	<i>Vênus</i>
WTA SAD SW	14.6	27.5	30.9	17.6
WTA SADR SW	8.45	17.6	26.9	9.45
WTA SAD MW	9.51	22.8	26.7	10.7
WTA SADR MW	<b>5.96</b>	<b>15.0</b>	22.8	<b>7.12</b>
DP	7.98	20.9	<b>20.2</b>	10.7

**Tabela 12:** Erros obtidos no cálculo das disparidades

Como pode ser observado no Gráfico 2, o algoritmo WTA SADR MW que implementa o método proposto apresentou o menor erro médio, seguido pelo algoritmo DP.

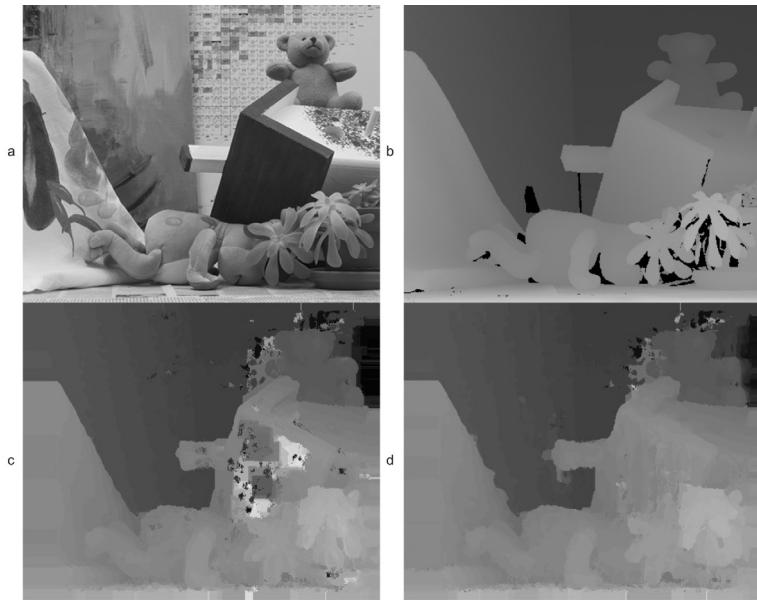


**Gráfico 2:** Erro médio das disparidades

Os algoritmos de correlação por janela simples ou múltipla apresentaram bons resultados nas regiões ricas em textura e bordas das imagens. Entretanto, ambas apresentam correspondências arbitrárias em regiões homogêneas ou com padrões de texturas repetitivas.

Os algoritmos de correlação por janelas simples e múltiplas com suporte regional melhoraram o desempenho dos algoritmos locais, principalmente em imagens com regiões homogêneas. Esta característica torna-se claramente visível na comparação dos mapas de disparidade Teddy. A Figura 26 apresenta a imagem esquerda do par estéreo Teddy (a), o respectivo mapa de disparidade real (b), o mapa de disparidade gerado pelo algoritmo WTA

SAD MW (c) e o mapa de disparidade gerado pelo algoritmo WTA SADR MW (d). Enquanto o algoritmo local falha ao corresponder os pontos da região homogênea correspondente ao telhado da pequena casa presente na imagem, o método com suporte regional conseguiu sucesso nesta região.



**Figura 26:** Correspondência em regiões homogêneas

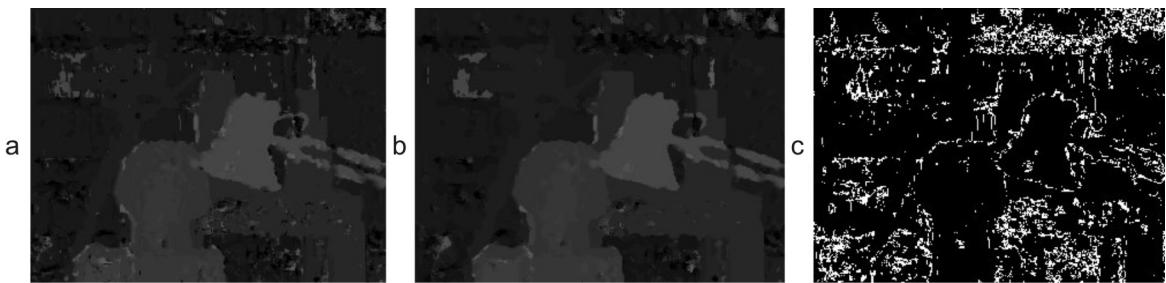
O algoritmo global DP apresentou um bom resultado médio, entretanto apresenta a característica de introduzir padrões formados por linhas que muitas vezes descaracterizam as bordas das descontinuidades presentes nas imagens. Isto poderia dificultar operações de robótica móvel como reconhecimento ou manipulação.

A Tabela 13 apresenta os tempos de processamento em segundos obtido para cada método na geração dos mapas de disparidade para cada par estéreo da base de referência. Como pode ser observado pelos tempos exibidos, o método global DP apresentou o melhor tempo para todos os pares, seguido do algoritmo do método local WTA SAD SW. Pode-se considerar pelos tempos apresentados por estes algoritmos que ambos atendem ao requisito de tempo real. O mesmo não ocorre com o método local WTA SAD MW e com os métodos locais com suporte regional WTA SADR SW e WTA SADR MW, principalmente nas imagens Teddy e Cones onde o espaço de busca é maior. Para atender aos requisitos de tempo real seriam necessários maiores esforços, empregando as técnicas citadas na seção 2.5.1.1, no sentido de otimizar ainda mais a implementação dos algoritmos.

Método	<i>Tsukuba</i>	<i>Cones</i>	<i>Teddy</i>	<i>Vênus</i>
WTA SAD SW	0.11	0.68	0.56	0.19
WTA SADR SW	0.52	2.25	2.23	1.05
WTA SAD MW	0.39	2.45	2.44	0.97
WTA SADR MW	1.66	9.32	9.32	3.35
DP	<b>0.09</b>	<b>0.35</b>	<b>0.25</b>	<b>0.16</b>

**Tabela 13:** Tempos de processamento para geração dos mapas de disparidade

Verificou-se ainda que para aplicações com alto requisito de tempo real cuja operação não exija muita precisão pode ser interessante utilizar algoritmos que implementem técnica empregando janela simples, ao invés da técnica implementando múltiplas janelas. Na Figura 27 são exibidas imagens dos mapas de disparidade obtidos com as técnicas de janela simples (a) e múltiplas janelas (b), e uma imagem da diferença nos mapas de disparidade obtidos pelas mesmas (c). Apesar de que com técnicas de múltipla janela foram obtidas melhorias nas regiões de descontinuidade onde existem oclusões, a diferença pode ser considerada desprezível para tarefas como, por exemplo, navegação de um robô. Já a diferença entre os tempos obtidos por algoritmos de janela simples e múltiplas janelas correspondentes não são desprezíveis.



**Figura 27:** Diferença entre os mapas calculados por janelas simples e múltiplas

Visto que o algoritmo DP apresentou o menor custo computacional e o segundo menor erro na geração dos mapas de disparidade para todos os pares da base de referência, considera-se que o algoritmo representa o melhor compromisso entre precisão e tempo de processamento para aplicações de robótica móvel.

Dado que foram obtidos resultados satisfatórios na obtenção de mapas de disparidade sobre a base de referência, e que o custo computacional obtido já atendia um requisito de tempo real, foram iniciadas experimentações sobre imagens reais capturadas pelo sistema de visão estereoscópica montado descrito na seção 3.2.

Inicialmente foram capturados pares estéreos de cenas que configurassem situações reais de operação de robôs móveis e de modo a compreender os cenários presentes no ciclo *detectar, aproximar e manipular*. Imagens com campo de visão mais amplo foram selecionadas para caracterizar um passo de reconhecimento e navegação. Já imagens com campo de visão mais limitado contendo objetos mais próximos ao sistema binocular foram capturadas para representar um passo de manipulação. Os três principais pares estéreo utilizados nas experimentações são apresentados na Figura 28, onde (a), (c) e (e) correspondem às imagens da esquerda e (b), (d) e (f) correspondem às imagens da direita.



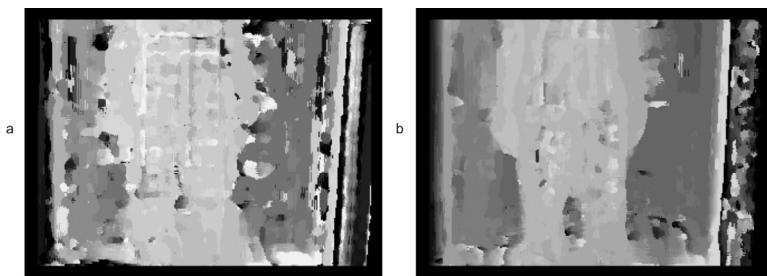
**Figura 28:** Pares estéreo próprios capturados com sistema de aquisição montado

Os primeiros resultados obtidos mostraram-se bastante inferiores quando comparados aos obtidos com os pares estéreo da base de referência, principalmente para os

algoritmos locais e locais com suporte regional. Para descobrir o motivo do problema, um experimento simples foi realizado capturando-se duas imagens por uma mesma câmera, em diferentes posições. Em seguida, o par estéreo capturado pela mesma câmera foi utilizado para geração do mapa de disparidade empregando os algoritmos implementados. O resultado foi superior ao obtido empregando as duas câmeras do sistema de visão estéreo, apesar destas se tratarem de modelos de câmeras e lentes idênticas.

Ficou claro que o problema foi introduzido pela sensibilidade das funções de similaridade baseadas em intensidade à diferença no ganho das câmeras e configuração das lentes. Este problema mostra-se agravado nos algoritmos de abordagem local. Foram realizadas experiências no sentido de buscar uma alternativa à função *SAD* empregada. Entretanto, resultados obtidos com as funções *SSD* e *SSD* normalizado foram similares.

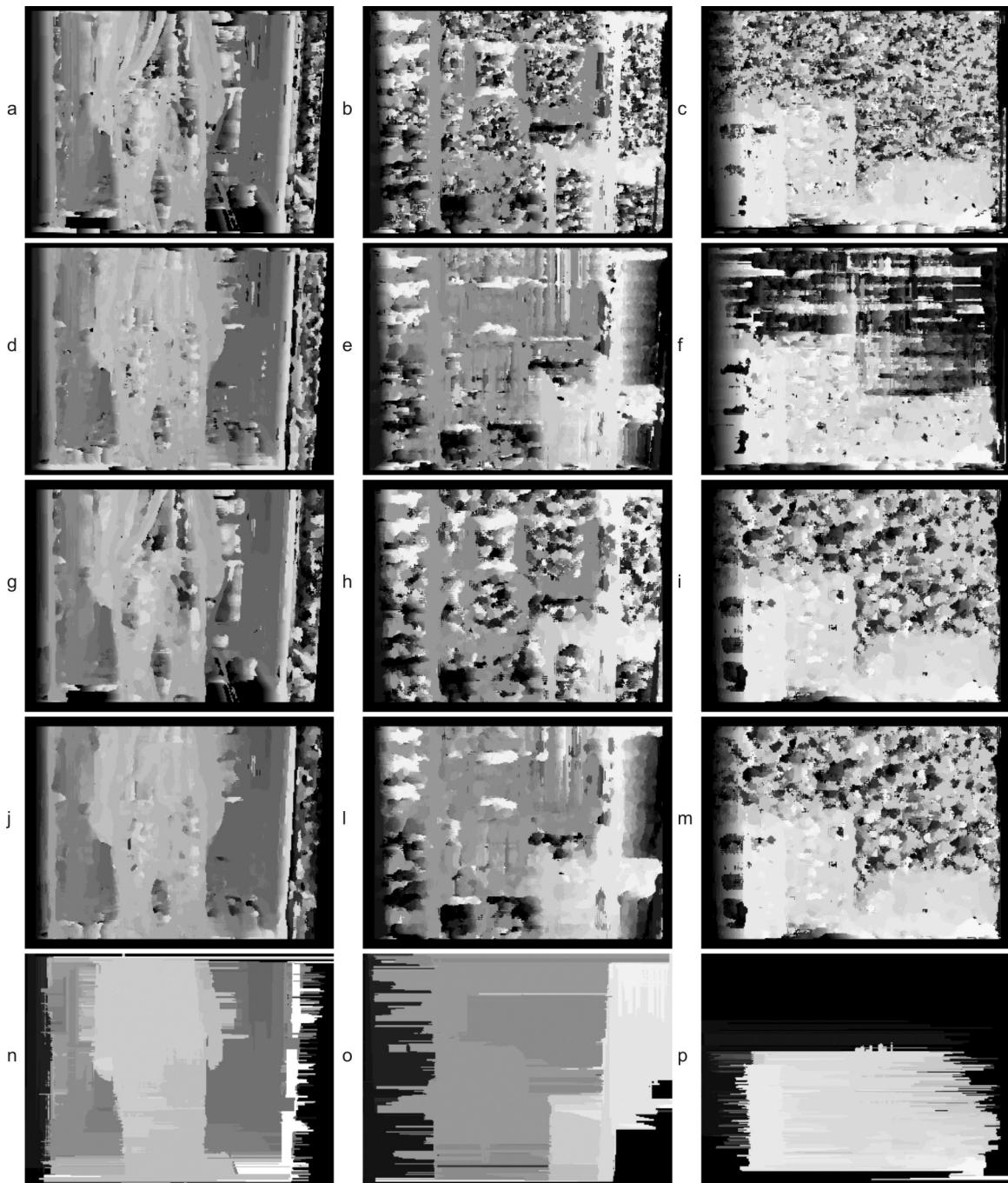
A alternativa encontrada na literatura que apresentou melhores resultados refere-se à utilização da normalização das imagens pelo filtro LOG, descrita na seção 2.5.1.1. Na Figura 29 são apresentados os mapas de disparidade calculados pelo algoritmo WTA SADR MW para um par estéreo não normalizado (a) e normalizado (b).



**Figura 29:** Mapa de disparidade obtido com par estéreo não normalizado

A utilização do LOG na normalização das imagens introduziu um aumento na qualidade dos mapas de disparidade gerados a partir de imagens reais. Porém, alguns dos algoritmos continuaram a apresentar um resultado não satisfatório nas regiões com intensidade homogêneas. Isto ocorre por que nestas regiões os valores calculados para diferentes janelas pelas funções de similaridade utilizadas apresentam diferenças pequenas entre si, introduzindo disparidades arbitrárias. O uso de janelas grandes mostrou-se uma opção, visto que reduzem o número de janelas sobre uma região contínua. Porém, ao utilizar este recurso a precisão de disparidade nas bordas mostrou-se muito prejudicada.

Na Figura 30 são exibidos os mapas de disparidade gerados para os pares estéreo exibidos na Figura 28 pelos algoritmos WTA SAD SW (a, b, c), WTA SADR SW (d, e, f), WTA SAD MW (g, h, i), WTA SADR MW (j, l, m) e DP (n, o, p). Para a geração destes mapas foram incluídas a normalização pelo filtro LOG e a verificação *left-right check*.



**Figura 30:** Mapas de disparidade calculados com pares próprios

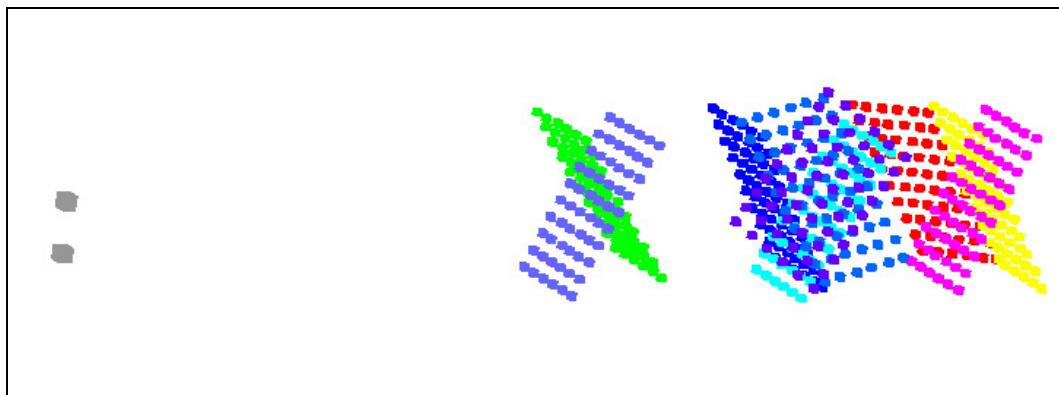
Apesar do emprego da normalização em imagens reais pelo filtro LOG, os algoritmos continuaram a introduzir um grande número de correspondências falsas aos mapas de disparidade. A eliminação destas correspondências pela utilização da verificação de consistência *left-right check* resultou em mapas de disparidade bastante incompletos.

Os algoritmos locais extraíram mapas de disparidade onde as únicas regiões com disparidades confiáveis encontram-se nas bordas existentes na imagem. Em regiões com intensidade homogênea não foram obtidos resultados satisfatórios. O algoritmo global DP extraíu um mapa de disparidade onde de maneira geral foram capturados os diferentes planos presentes na cena. Porém, para cada plano foram geradas faixas contínuas sem respeitar os limites dos objetos. Os algoritmos locais com suporte regional, principalmente empregando múltiplas janelas, apresentam resultados intermediários entre os algoritmos locais e globais.

### 3.3.5 Reconstrução

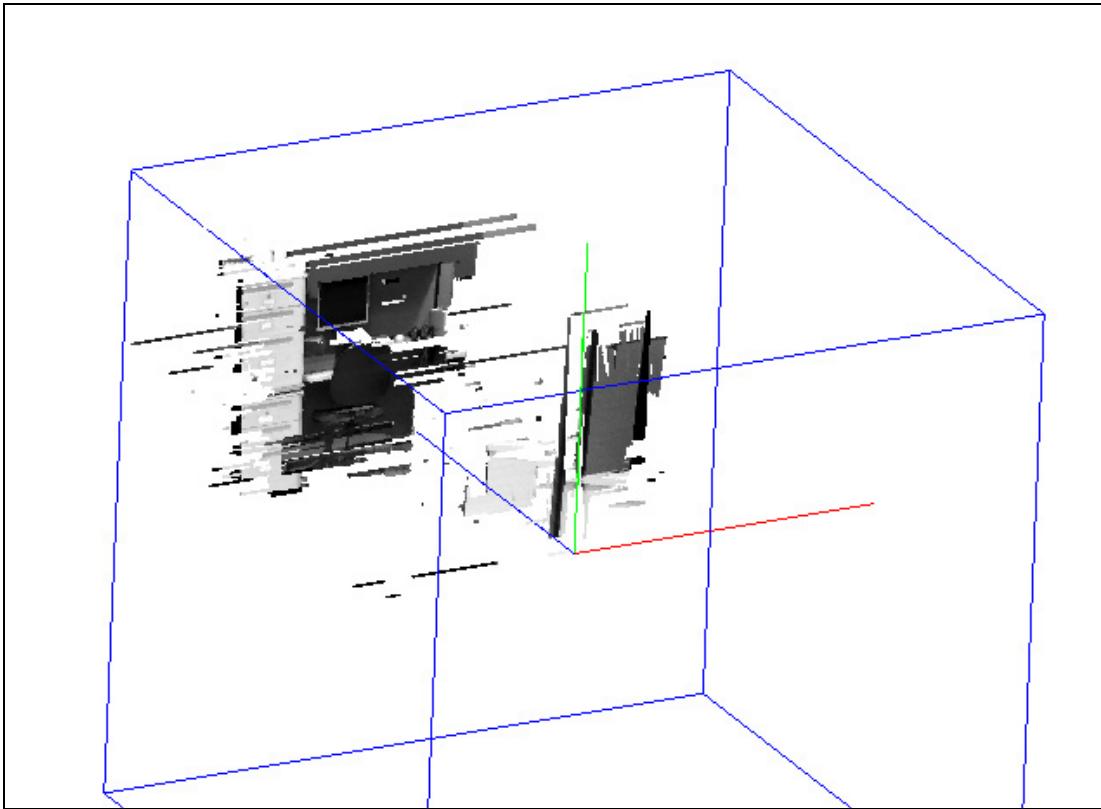
Empregando os parâmetros estimados no momento da calibração, e utilizando os pares de pontos correspondentes obtidos na etapa de correspondência, podemos realizar a reconstrução. Dado o tipo de sistema binocular utilizado e os parâmetros de calibração conhecidos, podemos convenientemente realizar a reconstrução métrica através de triangulação. O algoritmo implementado na biblioteca segue a descrição feita na seção 2.6.

A Figura 31 exibe a plotagem referente à reconstrução dos pontos do gabinete em uma seqüência de amostras utilizadas na calibração do conjunto binocular. Os centros de projeção das câmeras também são reconstruídos.



**Figura 31:** Reconstrução do cenário de calibração

A Figura 32 exibe a reconstrução realizada sobre um mapa de disparidade denso.



**Figura 32:** Reconstrução de mapa de disparidade

Para avaliação da qualidade da reconstrução foi adaptado o procedimento empregado por BENEVENUTI [18], onde as coordenadas 3-D conhecidas do gabarito de calibração empregado são confrontados com as coordenadas reconstruídas. Como no presente trabalho é empregado um padrão coplanar, não é possível utilizar diretamente as coordenadas 3-D do gabarito. Ao invés disso, utilizamos a distância euclidiana 3-D da linha diagonal entre o vértice superior esquerdo e o vértice inferior direito do padrão como medida, visto que a mesma é influenciada pelo erro de reconstrução para todas as dimensões.

A Tabela 14 apresenta os resultados obtidos seguindo o critério de avaliação descrito, para cada um dos cenários de calibração definidos na seção 3.3.2.

Pela avaliação do erro obtido, considerou-se que a precisão obtida nos experimentos é suficiente para um grande número de aplicações de robótica móvel, onde uma precisão apurada não é necessária.

<i>Parâmetro</i>	<i>Cenário 1</i>	<i>Cenário 2</i>	<i>Cenário 3</i>
Distância de Referência (mm)	257,39	257,39	257,39
Distância Média (mm)	254,04	257,78	260,89
Desvio Padrão (mm)	6,7841	3,1231	9,1025
Mínimo (mm)	242,14	253,07	249,65
Máximo (mm)	258,911	260,577	269,96
Erro Médio (%)	1,5362	1,0077	3,2902

**Tabela 14:** Resultado da avaliação da qualidade da reconstrução

### 3.4 Especificação do Sistema

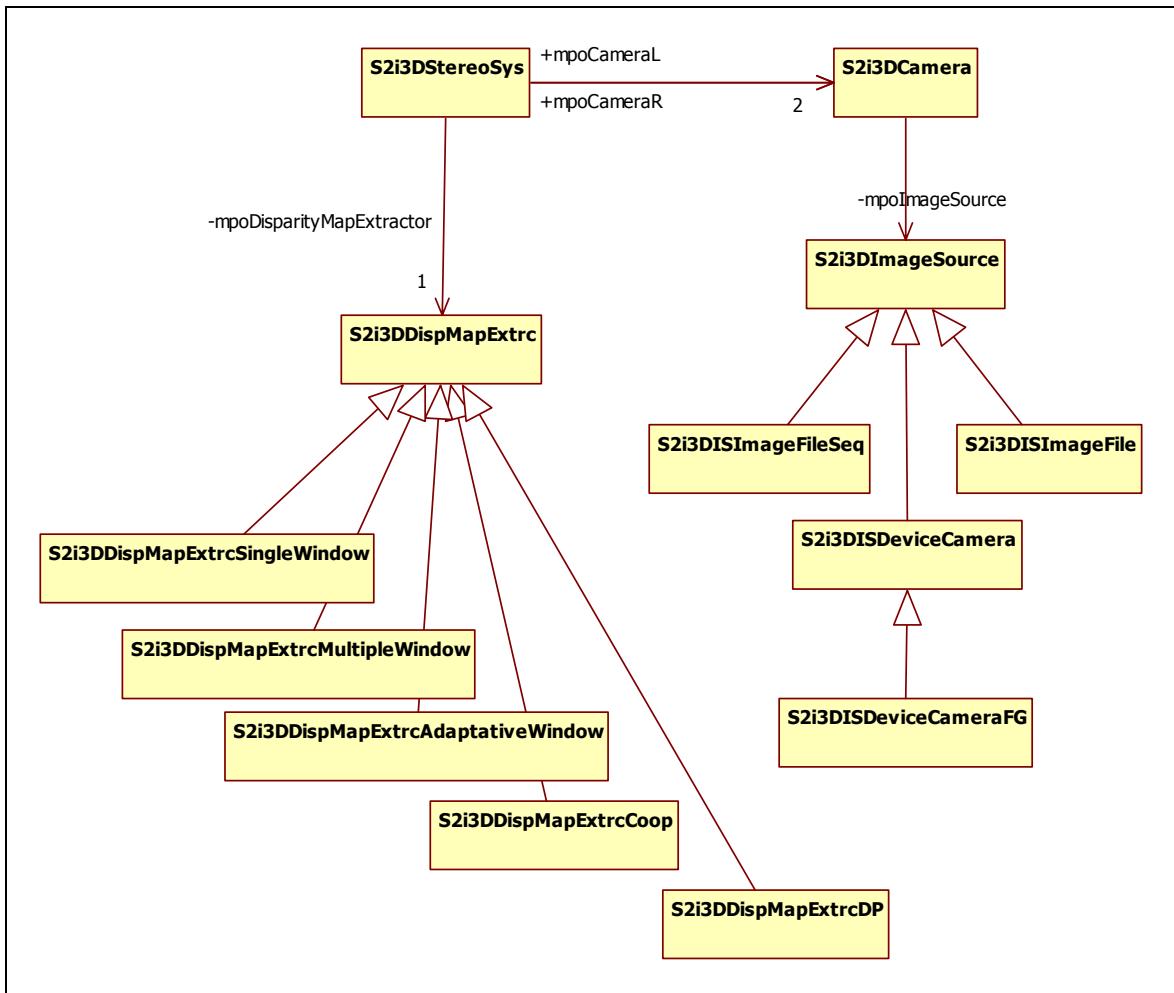
Nesta seção são apresentados os diagramas mais relevantes que descrevem a arquitetura e funcionamento da biblioteca desenvolvida.

O diagrama de classes exibido na Figura 33 apresenta as classes e relacionamentos mais relevantes na biblioteca.

A classe *S2i3DCamera* implementa o modelo de câmera *pinhole* e as equações da projeção de perspectiva descritos na seção 2.1. Dentre as funcionalidades principais disponibilizadas pela classe temos a calibração monocular, a transformação de coordenadas entre sistemas de coordenadas e a obtenção das imagens a partir de uma fonte de dados associada.

A classe abstrata *S2i3DImageSource* define o comportamento a ser implementado por uma fonte de dados responsável pelo fornecimento de imagens à câmera. As subclasses que a estendem representam diferentes fontes de imagens, que podem variar de dispositivos de captura à seqüências de imagens estáticas.

A classe *S2i3DStereoSys* implementa o modelo de visão estereoscópica apresentado na seção 2.2. Esta classe representa a interface principal entre a biblioteca e a *aplicação cliente* que a utiliza. Dentre as principais funcionalidades disponibilizadas por esta classe temos a configuração e inicialização do sistema de visão através da calibração, geração de mapas de disparidade densos e a reconstrução.

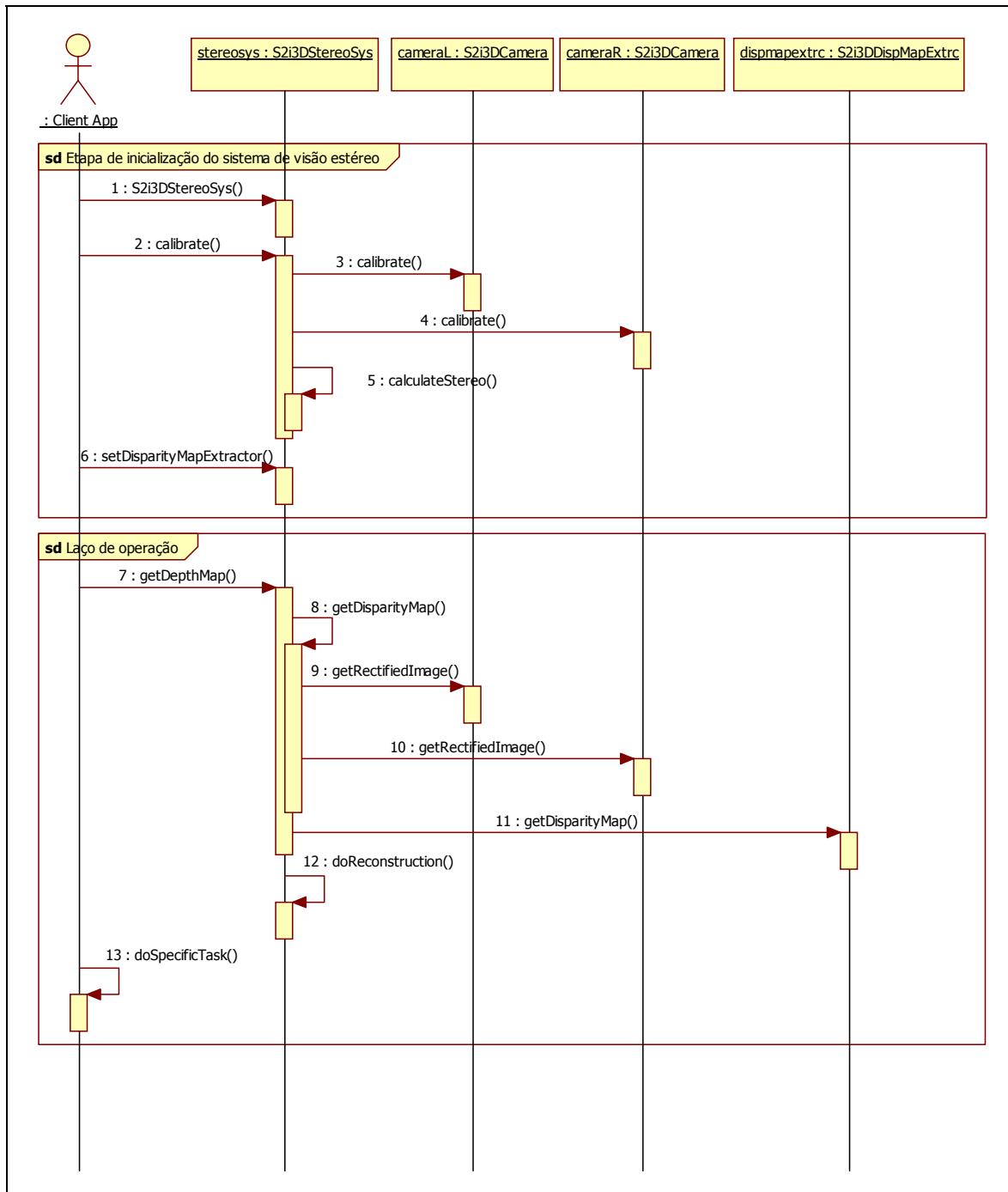


**Figura 33:** Diagrama de classes

A classe abstrata *S2i3DDispMapExtrc* define o comportamento a ser implementado por um extrator de mapas de disparidade denso, que corresponde à implementação de um método de correspondência específico. Além disso, compartilha funções comuns às classes especializadas, como por exemplo, consistências sobre os mapas de disparidade gerados.

O requisito de extensão funcional imposto pelos requisitos não funcionais são atendidos por esta arquitetura, visto que novas funcionalidades podem ser incluídas pela especialização das classes já existentes.

O diagrama de seqüência apresentado pela Figura 34 ilustra as interações principais entre o cliente e entre as classes que compõe a biblioteca durante a operação do sistema.



**Figura 34:** Diagrama de seqüência

É possível observar como, através da biblioteca, um mínimo de esforço é necessário por parte da *aplicação cliente* para configurar, inicializar e utilizar um sistema de visão estereoscópica.

### 3.5 Utilização da Biblioteca

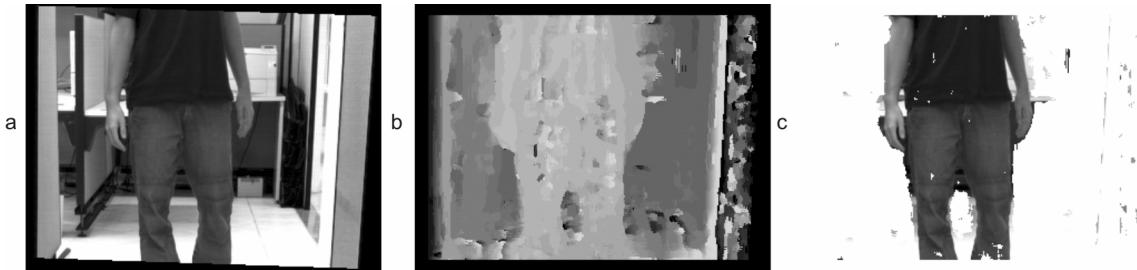
Após implementadas as funcionalidades principais da biblioteca, foram realizadas algumas experiências para validar a utilização da biblioteca no desenvolvimento de aplicações para robótica móvel. Estas experiências procuraram ilustrar o emprego dos mapas de profundidade para a realização de operações presentes no ciclo de operação geral de um robô móvel.

Considerando inicialmente a tarefa de navegação, procurou-se realizar a detecção de obstáculos presentes em uma cena observada, a partir dos mapas de profundidade gerados pela biblioteca. Duas experiências foram realizadas neste sentido, sendo que na primeira foram utilizadas imagens reais capturadas pelo sistema de visão estereoscópica montado, e na segunda foram utilizadas imagens sintéticas geradas pelo simulador Gazebo.

A técnica de detecção de obstáculos descrita por FRANÇA [38], baseada na segmentação dos mapas de disparidade, foi implementada na forma de uma *aplicação cliente*. A aplicação foi testada utilizando um par estéreo de imagens reais. Empregando um limiar global sobre os mapas de disparidade gerados pela biblioteca, foi possível isolar os agrupamentos de *pixels* que se encontravam a uma dada proximidade do sistema binocular. Sobre a imagem resultante da limiarização global foram empregadas operações morfológicas, que se mostraram eficientes na eliminação de ruídos relativos a pequenos segmentos resultantes de correspondências arbitrárias presentes no mapa de disparidade.

A Figura 35 apresenta os resultados obtidos na realização deste experimento, onde (a) representa uma imagem do par estéreo utilizada como referência, (b) representa o respectivo mapa de disparidade sobre o qual foi aplicado um limiar global, e (c) representa o segmento isolado pela limiarização global para um dado valor de limiar e cujos *pixels* foram coloridos com os valores correspondentes obtidos na imagem de referência.

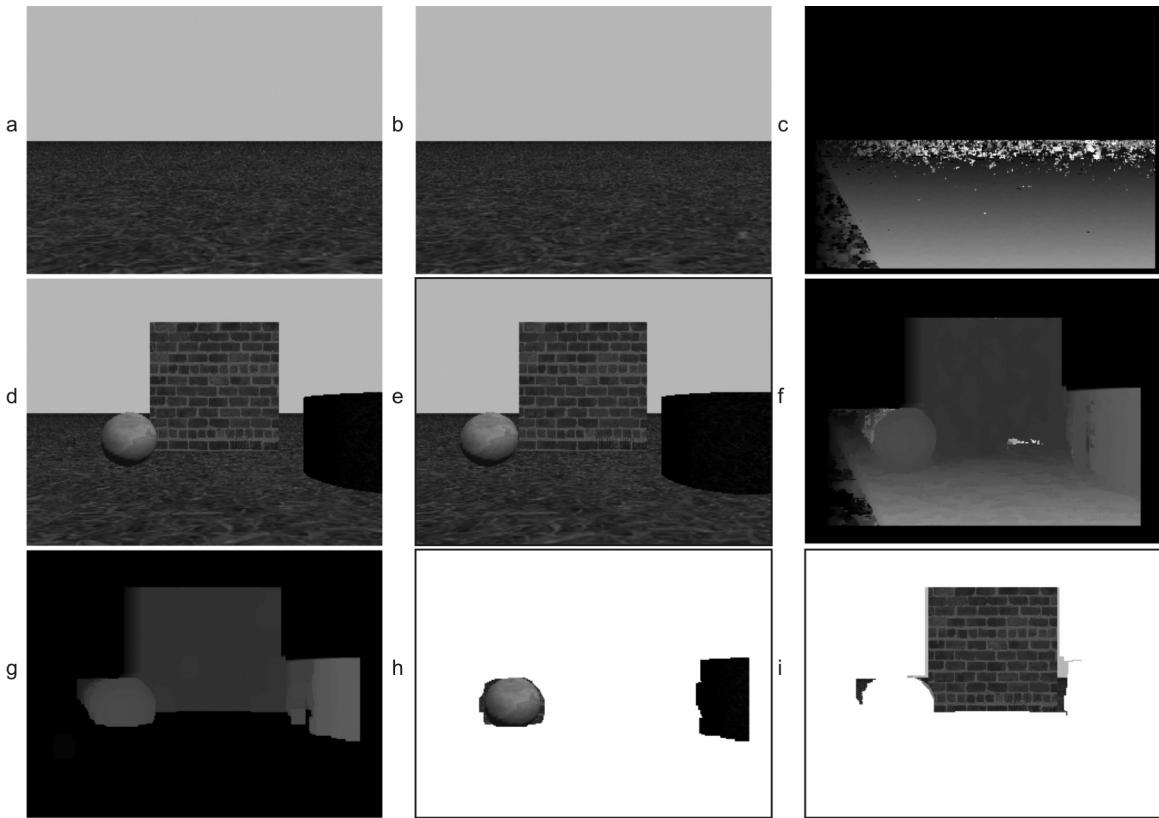
Observou-se durante a realização do experimento que pode ser interessante realizar a segmentação empregando dois limiares, de modo a definir os extremos de uma faixa de valores de disparidade representando uma dada faixa de distância de interesse.



**Figura 35:** Detecção de obstáculos

Num segundo experimento procurou-se incluir a detecção de planos, que muitas vezes é empregada para isolar o chão, teto ou paredes, principalmente em situações de navegação em ambientes estruturados. Seguindo a técnica de extração de pavimento descrita por GASPAR [13], podemos determinar, em etapa preliminar à operação, os parâmetros que definem o plano sobre o qual se desloca o robô. Durante a operação do robô, podemos utilizar os parâmetros deste plano para extrair os pontos correspondentes ao chão nos mapas de disparidade gerados durante a operação, pela simples comparação de intensidades observando uma margem de erro. Todos os pontos presentes no mapa de disparidade cujo plano de chão foi removido são considerados obstáculos. A abordagem de limiarização descrita no experimento anterior pode então ser empregada para segmentar os obstáculos com base na sua distância. Apesar da simplicidade do método, que considera que toda a operação do robô se dará sobre o mesmo plano, o mesmo mostra-se suficiente para diversos cenários.

Para realizar este experimento foi implementada mais uma *aplicação cliente*, e foram utilizadas as imagens sintéticas geradas pelo simulador Gazebo. A Figura 36 exibe as imagens obtidas durante o processo descrito anteriormente, tanto para detecção de obstáculos como para detecção do plano de chão. As imagens (a), (b) e (c) correspondem ao par estéreo e ao respectivo mapa de disparidade utilizado no momento da determinação dos parâmetros do plano de chão. As imagens (d), (e) e (f) correspondem a um par estéreo capturado no momento de operação e o seu respectivo mapa de disparidade. A imagem (g) corresponde ao mapa de disparidade resultante da extração do plano de chão sobre o mapa de disparidade da imagem (f). Finalmente, as imagens (h) e (i) apresentam obstáculos detectados em diferentes profundidades.



**Figura 36:** Detecção de obstáculos e pavimento em simulação

Apesar de que nos experimentos descritos os obstáculos detectados não apresentaram grande precisão nas bordas, considera-se que os resultados obtidos são suficientes para navegação em um grande número de aplicações. Tomando como base os mapas de obstáculos obtidos, é possível empregar técnicas para o cálculo de trajetória [39], identificando os possíveis caminhos a serem seguidos dentre os obstáculos pelo robô móvel.

Durante a realização deste trabalho não houve tempo hábil para a realização de experiências objetivas a fim de verificar a eficácia do emprego da biblioteca desenvolvida para aplicações de reconhecimento e manipulação. De qualquer maneira, algumas considerações podem ser feitas com base nos resultados obtidos durante a realização dos trabalhos descritos.

Assim como ocorre no processamento de imagens 2-D [40], considera-se que a interpretação de uma cena descrita por um mapa de profundidade exija a realização das etapas de segmentação, descrição e reconhecimento. Neste sentido, inicialmente os objetos presentes na imagem ou mapa de profundidade devem ser agrupados em conjuntos de pontos

relacionados segundo algum critério. Em seguida, os agrupamentos são descritos de alguma forma de modo a simplificar a etapa seguinte que trata do reconhecimento.

Por uma avaliação subjetiva, sugere-se que ainda não é possível obter resultados apurados para segmentação, descrição e reconhecimento baseados apenas nos mapas de profundidade gerados pela biblioteca. A falta de suavidade na descontinuidade de profundidade obtida e a falta de precisão nas bordas são os maiores problemas encontrados. Além disso, todos os algoritmos de correspondência apresentam um grande número de correspondências falsas.

Uma opção possível é a utilização combinada da informação fornecida pelo mapa de profundidade e pelas imagens do par estéreo. Por exemplo, conforme demonstrado na descrição de detecção de obstáculos realizada anteriormente, a descontinuidade de profundidade pode ser utilizada como importante fonte de informação para segmentação dos componentes que constituem a cena. Sugere-se que a partir dos agrupamentos de componentes conexos segmentados pela descontinuidade de profundidade poderiam ser empregadas técnicas 2-D de descrição e reconhecimento.

## Capítulo 4

# Conclusões e Perspectivas

Neste capítulo são apresentadas conclusões resultantes do desenvolvimento deste trabalho, assim como sugestões a trabalhos futuros.

### 4.1 Conclusões

No presente trabalho foi apresentado o desenvolvimento de uma biblioteca para sistemas de visão estereoscópica para robótica móvel. Para tal, foram tratados diferentes problemas da estereoscopia, procurando prover mapas de profundidade com detalhes do ambiente suficientes para a operação geral de um robô móvel que tenha um sistema de visão estereoscópica como principal fonte de informação.

O modelo de sistema de visão estereoscópica utilizado, em conjunto com as técnicas de calibração, retificação e reconstrução apresentou resultados satisfatórios na recuperação da informação de profundidade quando a correspondência de pontos é correta.

O problema de correspondência mostrou-se o maior desafio encontrado, estando de acordo com relatos encontrado na literatura. Foram avaliados diversos métodos, seguindo tanto abordagens locais como globais. No presente trabalho foi ainda proposto um novo método de correspondência estéreo, baseada na adição de um descritor regional na função de custo empregada por métodos de correspondência local [41]. Este método melhorou o desempenho de algoritmos de correspondência locais já conhecidos. Além disso, elevou os resultados destes algoritmos a ponto de se tornarem superiores aos obtidos com técnicas de correspondência global.

Entretanto, pelo elo fraco que a correspondência continua a representar no ciclo de processamento empregado para a geração dos mapas de profundidade, conclui-se que no momento apenas operações de navegação possam ser realizadas tomando estes como única

fonte de informação. Sugere-se que operações mais apuradas de reconhecimento e manipulação devam ser realizadas combinando a informação tridimensional provida pelos mapas de profundidade e a informação bidimensional provida pelo par de imagens estéreo.

A biblioteca em si atendeu aos requisitos levantados, representando uma ferramenta útil no desenvolvimento de sistemas de visão para robótica móvel. O esforço para configuração, inicialização e operação de um sistema de visão estereoscópica é reduzido significativamente pela utilização da biblioteca. Desta forma, o desenvolvedor poderá abstrair as funcionalidades básicas de um sistema de visão estereoscópica e poderá manter o foco na resolução das questões específicas à sua aplicação.

Finalmente, como principais contribuições do trabalho temos a disponibilização de uma infra-estrutura básica para auxiliar o desenvolvimento de sistemas robóticos que empreguem visão estereoscópica, a avaliação de diferentes métodos de geração de mapas de disparidade tendo como foco a utilização para robótica móvel [42], e a proposta de um novo método de correspondência local com suporte regional [41].

## 4.2 Perspectivas Futuras

Durante o desenvolvimento deste trabalho foram levantadas algumas idéias para trabalhos futuros, que podem ser realizadas no sentido de melhorar ou estender as capacidades da biblioteca disponibilizada.

O problema de correspondência pode ainda ser atacado no sentido de melhorar a qualidade dos mapas de profundidade gerados, principalmente para as situações de operação em ambientes estruturados onde existe uma grande incidência de regiões homogêneas. Uma possível solução seria utilizar a dimensão do tempo como uma nova fonte de informação, no sentido de combinar a informação obtida pelas imagens obtidas em diferentes instantes para calcular os mapas de disparidade. Esta abordagem possui uma motivação no comportamento humano, onde procuramos observar um objeto ou cena sob diferentes pontos de vista quando os mesmos não nos são claros.

Além disso, observou-se que alguns algoritmos de correspondência densa deixam de atender os requisitos de tempo-real quando são incluídas verificações de consistência. A utilização de recursos de paralelismo disponíveis na arquitetura de processadores atuais, ou

de processamento empregando a unidade de processamento gráfico (GPU), são soluções a serem consideradas.

A biblioteca disponibilizada foi concebida para atuar, em conjunto com o conjunto binocular, como um dispositivo de captura 3-D independente de aplicação. Uma extensão certamente útil seria criar uma camada de abstração no sentido de disponibilizar soluções para tarefas de robótica móvel de mais alto nível, como navegação, reconhecimento e manipulação.

## **ANEXO 1 – ARTIGOS PRODUZIDOS**

# Stereo vision based on local correspondence with regional support

Mauricio E. Stivanello, Eduardo S. Leal, Nicolas Palluat,  
Marcelo R. Stemmer

*DAS, Federal University of Santa Catarina, Brazil*

---

## Abstract

In this paper, a new method of stereo correspondence based on regional support is proposed. Using the information provided by high frequency components, a new regional point descriptor is introduced to improve matching accuracy. This approach fits well to area-based techniques and it is suitable to real time applications. Results obtained by extending known area-based techniques to the proposed method demonstrate its effectiveness.

*Key words:* Stereo vision, Disparity map, Stereo correspondence, Regional support, Stereo matching

---

## 1 Introduction

Stereo vision systems retrieve three-dimensional information of a scene through the identification of corresponding points in captured images. Problems as occlusions, areas without texture, non uniform illumination, among others introduce errors in correspondence and this task becomes complex and with a high computational cost.

A large number of algorithms has been proposed to solve this problem. In general, cost functions based on similarity are defined to carry out the correspondence between the stereo pair images. Based on the strategy used to perform the correspondence, they can be classified as global or local algorithms [1].

---

*Email addresses:* `maustiva@das.ufsc.br` (Mauricio E. Stivanello),  
`sleal@das.ufsc.br` (Eduardo S. Leal), `nicolas@das.ufsc.br` (Nicolas Palluat),  
`marcelo@das.ufsc.br` (Marcelo R. Stemmer).

Global algorithms rely on iterative schemes that calculate disparities through the minimization of global cost functions, based on rewards and penalties [2,3]. Although they carry out accurate correspondences, these algorithms present a high computational cost.

Local algorithms calculate disparities for each point based on the intensity of neighbouring points [4,5]. When compared to global algorithms, they present a significantly low precision of the calculated disparity maps. However, they present a reduced computational cost and have been successfully employed in many systems with real time requirements [6,7].

In this work, improvements to local correspondence methods are proposed through the introduction of a regional support to the cost functions traditionally used. In section 2, the approach of basic local correspondence is described, and its problems are discussed. In section 3, the proposed regional support for local correspondence is introduced. The general steps of a local algorithm with regional support are presented in section 4. Finally, in section 5 local algorithms modified by the introduction of regional support are compared to its original versions and other local and global algorithms.

## 2 Local correspondence based on intensity

In local correspondence algorithms it is assumed that the image points are involved by a window of neighbouring elements where the disparity is the same. This way, the correspondence for each point is performed based on the correspondence of these windows.

The basic method used for correspondence in this approach is the correlation [5]. Given a stereo image pair, one of the images is chosen as reference and the size of a square window to be used in the correspondence is defined. For each point in the reference image, the following steps are done:

- (1) Centralize the window in the point of interest in the reference image;
- (2) Do the correlation between the window located in the reference image with the second window, moving it on all positions through the search space in the second image;
- (3) Select the point in the second image that corresponds to the element of interest in the first image, choosing the position where the correlation minimizes the employed cost function;
- (4) Disparity is given by the relative difference between the central position of the windows. Store disparity values in one third image, at the same coordinates of the central point of the window in the reference image.

The result is a dense map of disparities, composed by the calculate disparity for each element in the reference image.

The correspondence algorithms require that the input images are rectified, in order to align the epipolar lines of the images and thus to limit the search space [8]. Moreover, search space is also limited by the possible maximum and minimum disparities for a given stereo configuration.

The correspondence methods by correlation use cost functions considering the similarity of the intensity values of each image point, such as the sum of the squared differences (*SSD*) and the sum of absolute differences (*SAD*). A cost function using similarity of the intensity values *SAD* is given by

$$C_{sad}(x, y, d) = \sum_{i,j=-n}^n |D_{il}(x + j, y + i) - D_{ir}(x + d + j, y + i)| \quad (1)$$

where  $d$  is a given disparity,  $D_{il}(x, y)$  and  $D_{ir}(x, y)$  correspond to a description function based on intensity referring to the left and right images. A square window of size  $2n + 1$  is used.

In the correspondence algorithm described above, the intensity assumes the role of descriptor for each image points. This description doesn't show robustness, since the intensity value of each point is sensible to influences such as differences in the gain of sensors, radiometric variations, among others. that the intensity assume a role as a descriptor for each image points'

In an effort to solve this problem, *LOG* filter [6] is widely used to normalize the images. *LOG* filter is composed by a Gaussian filter and a Laplacian filter. The Laplacian filter reduces illumination influences, since its result is zero in areas with constant intensity and it is negative or positive next to edges with high intensity gradient. Laplacian filter is highly sensible to noise influence. For this reason, the Gaussian filter is applied before it to minimize this problem.

Fig. 1 illustrates a stereo image pair normalized using the *LOG* filter. The original right image presents a significant difference in the luminous intensity when compared to the left image (a and b). However, applying *LOG* filter in both images we achieve a very similar result (c and d).

The use of *LOG* normalization of the images increases the quality in the generated disparity maps. However, large homogeneous regions, resultant of areas where the intensity derivative is low are obtained. In these cases, classical similarity criteria still don't present satisfactory results.

Fig. 2 illustrates the existing problem in the description of points based in intensity during the correspondence process. In the illustrated correspondence,

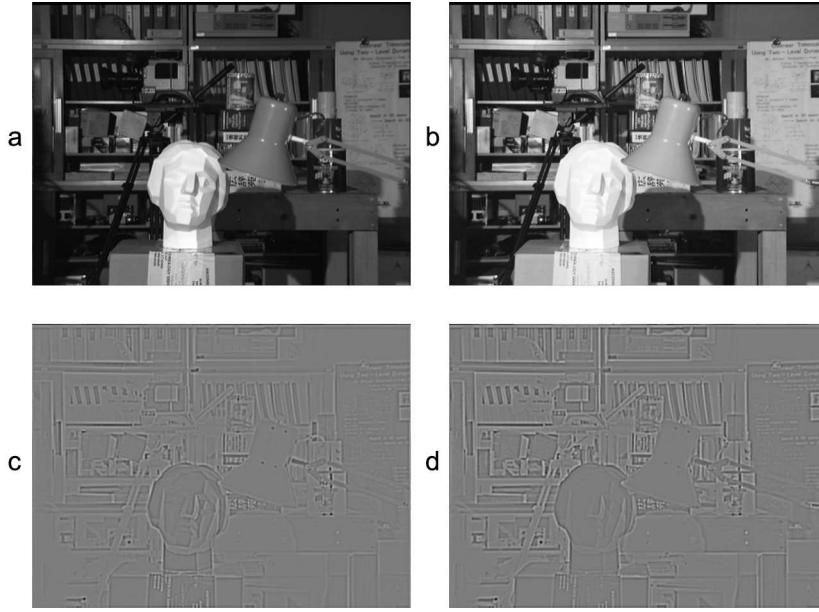


Figure 1. LOG normalization - a,b)Original c,d)Filtered

only the descriptor based on intensity  $D_i(x, y)$  is used. It is possible to observe how the values that describe each point can be similar. The use of a cost function only based on this descriptor can yield in the occurrence of arbitrary correspondences.

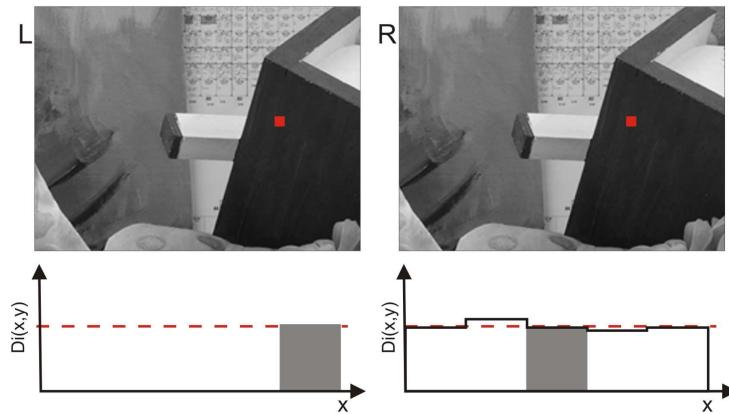


Figure 2. Description using intensity

The use of wide windows reveals an option, since it reduces the number of windows completely overlapped in a homogeneous region. However, when this resource is used, the precision of disparity in the edges is very poor.

### 3 Proposed method

In the correspondence process described previously, the intensity assumes the role of descriptor for each image points. It is observed, however, that a description based on intensity is not enough to identify a pair of corresponding points uniquely. The same occurs with the human visual system when we observe a surface extremely smooth or in poorly illuminated environment. In these conditions, in the lack of fine details that let us guide correspondence between points observed by eyes, in a natural way we appeal to the information supplied by high frequency elements located in the peripheral region to determine the depth.

The proposed method in this work attempt to model this behavior. The main idea is to describe each image points with an intensity descriptor and a regional descriptor. The regional descriptor must aggregate peripherical information provided by components of high frequency located next to the observed point. This way, we can define a cost function with regional support as

$$C_r(x, y, d) = S_i(x, y, d) + \alpha S_r(x, y, d) \quad (2)$$

where  $S_i(x, y, d)$  is a similarity function based on an intensity descriptor  $D_i(x, y)$ ,  $S_r(x, y, d)$  correspond to a similarity function based on a regional descriptor  $D_r(x, y)$  and  $\alpha$  correspond to the employed weight to the function of regional support.

Fig. 3 illustrates the effect of the description of points with regional support during the correspondence process. Different from the situation illustrated in Fig. 2, in this case both the intensity descriptor  $D_i(x, y)$  and the regional descriptor  $D_r(x, y)$  are used. It is possible to observe how a description based on this combination of descriptors allows a better distinction between points. A cost function based on these descriptors permits a better selection of the correct correspondences.

A similarity function  $S_r(x, y, d)$  based on regional descriptors is given by

$$S_r(x, y, d) = \sum_{i,j=-n}^n \sqrt{D_v(x, y, d)^2 + D_h(x, y, d)^2}. \quad (3)$$

The terms  $D_v(x, y, d)$  and  $D_h(x, y, d)$  in eq. 3 are given respectively by

$$D_v(x, y, d) = \frac{D_t(x, y)}{D_t(x, y) + D_b(x, y)} - \frac{D_t(x + d, y)}{D_t(x + d, y) + D_b(x + d, y)} \quad (4)$$

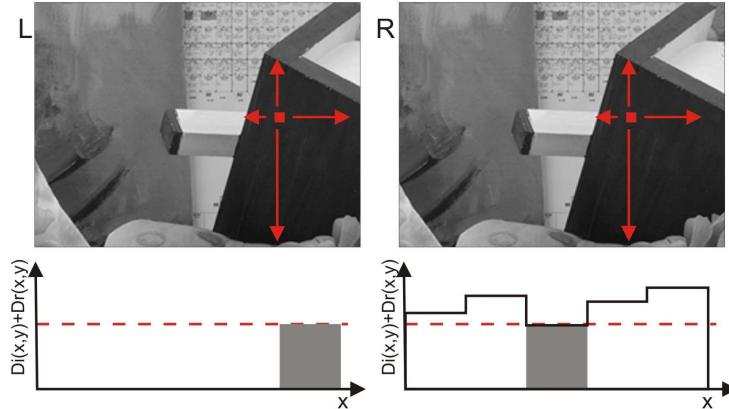


Figure 3. Description using intensity and regional support

and

$$D_h(x, y, d) = \frac{D_l(x, y)}{D_l(x, y) + D_r(x, y)} - \frac{D_l(x + d, y)}{D_l(x + d, y) + D_r(x + d, y)}, \quad (5)$$

where  $D_t(x, y)$ ,  $D_b(x, y)$ ,  $D_l(x, y)$  and  $D_r(x, y)$  are regional descriptors that represent the distance between the observed point and the next component of high frequency above, below, to the left and to the right, respectively. Both  $D_v(x, y, d)$  and  $D_h(x, y, d)$  must be adimensional, due the possible dimensional differences introduced by perspective in the images capture.

## 4 Algorithm summary and optimizations

This section presents the general steps of a local correspondence algorithm with regional support.

### 4.1 Calculation of Regional Descriptors

Due to insertion of the regional support term in the correspondence cost function, we need to calculate the regional descriptors for each image. These descriptors can be based on the distance between each point and the components of high frequency that delimit the homogeneous region where the point is located. To identify these components we can make use of edge filters. However, the result gotten with these filters is normally noisy.

An important consideration is the non obligation to apply the regional support to the whole image. When the intersection between the correlation win-

dow and components of high frequency is different from zero, we don't need it because the performance of similarity functions based on intensity is satisfactory. Hence, we can use a segmentation method based on the variance of the intensities to isolate the desired homogeneous regions.

A segmentation based on the intensities variance map shows reduced computational cost when compared to other techniques, e.g. Mean Shift [9]. The variance map can be calculated by

$$\sigma_L^2(x, y) = \frac{1}{N^2} \sum_{i,j=-n}^n L(x + j, y + j)^2 - \mu_L(x, y)^2, \quad (6)$$

where  $N = (2n + 1)^2$  and  $\mu_L(x, y)$  is the average of the intensity values [10].

We can isolate the homogeneous regions of the image applying a global linearization over the calculated map of variances. From the resultant binary image, we can calculate the distance maps corresponding to the regional descriptors  $D_t(x, y)$ ,  $D_b(x, y)$ ,  $D_l(x, y)$  and  $D_r(x, y)$ .

Fig. 4 Illustrates the process described above. Fig. 4-a presents the original image. Fig. 4-b presents the variance map calculated using the original image. Fig. 4-c presents the result of the binarization of the variance map. Furthermore, Fig. 4-d presents the distance map corresponding to the descriptor  $D_t(x, y)$ , calculated over the binary image.

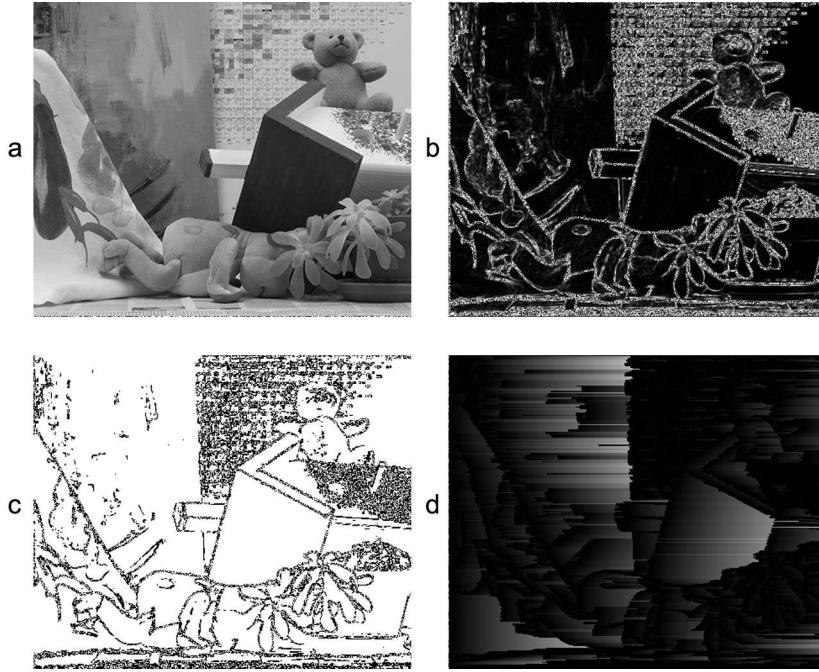


Figure 4. Calculation of regional description

With the distance regional descriptors, we can calculate the adimensional descriptors maps  $D_v(x, y, d)$  and  $D_h(x, y, d)$  according to eq. 4 and eq. 5.

#### 4.2 Correspondence

From eq. 2 we can define many cost functions from the combination between different similarity functions based on intensity and region. Since the regional similarity function presented in eq. 5 is essentially a correlation function, it fits very well to intensity similarity function of the same nature.

Therefore, using intensity similarity function presented in eq. 1 and the regional similarity function presented in eq. 3, we can define a new cost function based on *SAD* and with regional support rewriting the cost function with regional support given by eq. 2 as

$$C_{sadr}(x, y, d) = \sum_{i,j=-n}^n |D_{il}(x + j, y + i) - D_{ir}(x + d + j, y + i)| + \alpha \sqrt{D_v(x, y, d)^2 + D_h(x, y, d)^2}. \quad (7)$$

Here, the calculation of the term of regional support is performed using the maps of descriptors previously calculated described in the previous section.

Finally, once we have computed the cost values over all disparity space search, the corrected correspondence is selected according to Winner-Takes-All approach (WTA) [1].

#### 4.3 Error elimination

In order to eliminate false correspondences, the left-right consistency verification method is employed. This consistency is carried out for each point to verify if the correspondences calculated are coherent when the reference image that guides the correspondence is modified [4].

The process of verification can be described as follows. First, for each point in the left image the best correspondence with the points in the right image is calculated. Then, the reference image is inverted and, for each point in the right image, the best correspondence with the points in the left image is calculated. At last, only are considered the points where the correspondence from left to right and the correspondence from right to left are coherent. In

case that the values of disparity diverge, the points are marked attributing value zero.

In a following step, the points with value of disparity equal to zero are removed through interpolation.

#### 4.4 Optimizations

In a correspondence algorithm through correlation of windows, the most expensive task is the calculation of the correspondence cost values. A cost function based on similarity  $SAD$  can be calculated efficiently exploring the fact that neighbouring windows overlap [11] each other. This same characteristic can be extended for the cost function presented in eq. 7.

In neighbouring windows with the same disparity, overlapping points will have the same values of absolute difference both for the intensity descriptor as for regional descriptor. Therefore, a new value of cost can be calculated for the current window using the value calculated for an old window deducting the values from the points that are part only of the old window, and adding the values of the points that are part of the new window. This optimization still presents the advantage to make the computational cost practically independent of the window size used in correlation.

Additionally, algorithms based on correlation can be optimized through the use of resources of parallel processing SIMD available on diverse architectures [11,7]. Since the inclusion of regional support does not modify the correlation based behavior of the algorithms where it is applied, it can also take benefits of these advantages.

#### 4.5 Algorithm variations

The cost function with regional support presented in eq. 2 allows a diversity of combinations, due to the use of the different intensity similarity functions and regional similarity functions. Using the components presented in the previous sections, two algorithms were created based on the proposed method of regional support:

- (1) **WTA SADR SW** - Uses a single window correlation, a cost function based on  $SAD$  with regional support, WTA approach for selection of the correct correspondence and consistency of disparities by left-right check [10].

- (2) **WTA SADR MW5** - Presents the same characteristics of the algorithm 1, but employ Multiple Windows correlation [4].

## 5 Experimental results

In this section, the experimental results obtained in the comparison between the proposed algorithms and different known correspondence algorithms are presented. The selected algorithms, representing both local, global, and the proposed local/regional approaches, are presented in the table 1.

Table 1  
Correspondence methods compared

Abbreviation	Description	Approach
WTA SAD SW	Single window (SAD/WTA)	Local
WTA SADR SW	Single window with regional support (SADR/WTA)	Local/Regional
WTA SAD MW5	Multiple windows (SAD/WTA)	Local
WTA SADR MW5	Multiple windows with regional support (SADR/WTA)	Local/Regional
SSD MF	Shiftable windows (SSD/WTA)	Local
DP	Dynamic programming	Global
SO	Scanline optimization	Global

The experiments were carried out using a base of reference stereo images with available ground true [1]. We used the reference stereo image pairs of Tsukuba, Teddy, Cones and Venus. Fig. 5 presents the left image and the ground true of the pairs of reference Tsukuba and Teddy.

The reference stereo image pairs present different characteristics such as presence of complex objects, objects in different depths generating varied occlusions, regions without texture or with homogeneous intensity, among others. Hence, they are ideal for evaluation of correspondence algorithms under different criteria.

Fig. 6 presents the results obtained in the calculation of the disparity for the stereo pair Tsukuba, using the methods WTA SADR SW (a) and WTA SADR MW5 (b).

Fig. 7 presents the results of disparity calculation obtained for Terry stereo

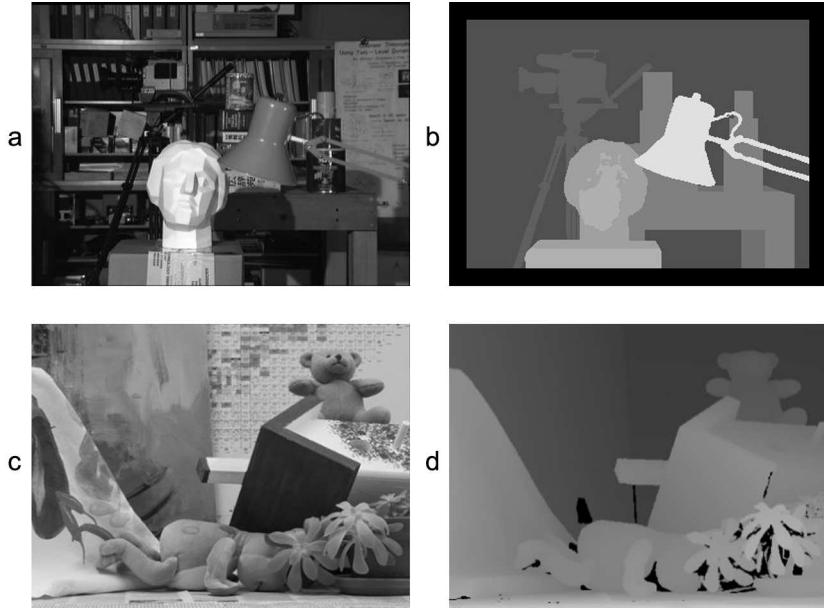


Figure 5. Reference images - a,b)Tsukuba c,d)Teddy

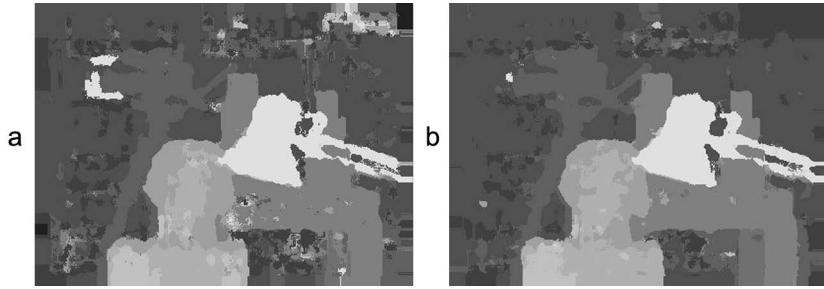


Figure 6. Calculate disparity (tsukuba) - a)WTA SAD MW5 b)WTA SADR MW5 pair, using the methods WTA SADR SW (a) and WTA SADR MW5 (b).

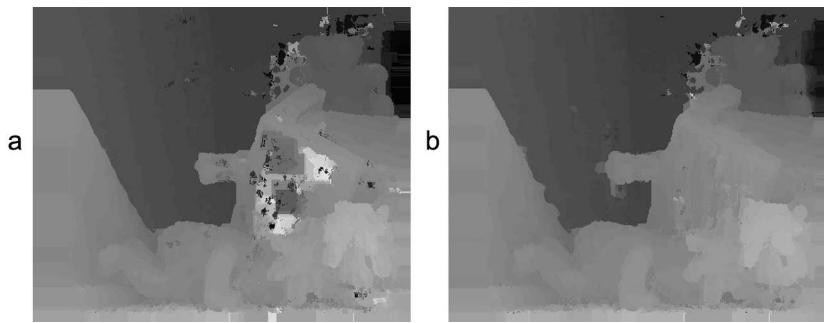


Figure 7. Calculate disparity (Teddy) - a)WTA SAD MW5 b)WTA SADR MW5  
The comparison of the evaluated methods was performed using the percentage of incorrect correspondences as a quality measure. The percentage of incorrect correspondences is given by

$$E = \frac{1}{N} \sum_{x,x} |d_c(x,y) - d_t(x,y)| > \delta_d \quad (8)$$

where  $d_c(x, y)$  is the calculated disparity,  $d_t(x, y)$  is the true disparity and  $\delta_d$  is the value of tolerance to the disparity error [1].

Table 2 shows the general error presented in the calculation of the disparities with the different evaluated methods. The employed search space was the same suggested by the reference base, and all methods run with constant parameters on all 4 stereo pairs. For the calculation of the general error a value to  $\delta_d$  equal to 1 was used.

Table 2  
General error of disparities

Method	Tsukuba	Venus	Teddy	Cones
WTA SAD SW	14.6	17.6	30.9	27.5
WTA SADR SW	8.45	9.45	26.9	17.6
WTA SAD MW5	9.51	10.7	26.7	22.8
WTA SADR MW5	<b>5.96</b>	7.12	22.8	<b>15.0</b>
SSD MF	7.07	<b>5.16</b>	24.8	19.8
DP	7.98	10.7	<b>20.2</b>	20.9
SO	7.22	10.9	28.2	22.8

We can clearly observe the improvements provided by the proposed method, when comparing the results obtained from the methods where the regional support was included (WTA SADR SW / WTA SADR MW5) as opposed to its variants without regional support (WTA SAD SW / WTA SAD MW5).

Fig. 8 shows a chart of the average values obtained over the reference stereo image pairs for each one of the different evaluated methods.

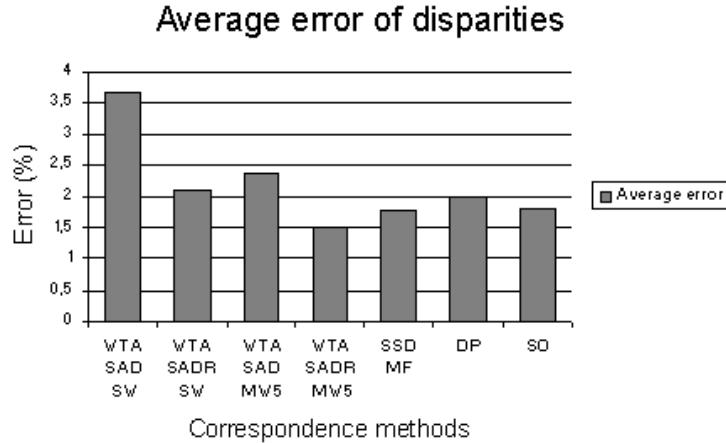


Figure 8. Average error of disparities

The regional support method proposed in this work raised the results precision

for local correspondence algorithms where it was applied. Additionally, the multiple windows algorithm with regional support presented superior results in relation to the obtained using algorithms with global approaches.

## 6 Conclusions

In this work a new method of stereo correspondence with regional support is proposed, based on addition of a similarity function using regional descriptors to the commonly used cost functions. This method improved the performance of already known local correspondence algorithms. Moreover, it obtains better results than the techniques using global correspondence.

Being simple and presenting a superior performance, it can be employed to calculate the initial disparity estimative in more elaborated correspondence methods [12], substituting the currently used.

Finally, similar results to the presented here have been obtained through the on-line evaluation in the web page Middlebury Stereo Vision [1], confirming the validity of the proposed method.

## 7 Acknowledgments

This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ).

## References

- [1] D. Scharstein, R. Szeliski, A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms, *International Journal of Computer Vision* 47 (1-3) (2002) 7–42.
- [2] S. Birchfield, C. Tomasi, Depth Discontinuities by Pixel-to-Pixel Stereo, in: *ICCV '98: Proceedings of the Sixth International Conference on Computer Vision*, IEEE Computer Society, Washington, DC, USA, 1998, pp. 1073–1080.
- [3] S. Mattoccia, F. Tombari, L. Di Stefano, Stereo Vision Enabling Precise Border Localization within a Scanline Optimization Framework, in: S. B. K. H. Z. Yasushi Yagi, In So Kweon (Ed.), *8th Asian Conference on Computer Vision*, Tokyo, Japan, 2007.

- [4] H. Hirschmüller, P. R. Innocent, J. Garibaldi, Real-Time Correlation-Based Stereo Vision with Reduced Border Errors, *International Journal of Computer Vision* 47 (1-3) (2002) 229–246.
- [5] E. Trucco, A. Verri, *Introductory Techniques for 3-D Computer Vision*, Prentice Hall PTR, Upper Saddle River, NJ, USA, 1998.
- [6] P. Foggia, J.-M. Jolion, A. Limongiello, M. Vento, A New Approach for Stereo Matching in Autonomous Mobile Robot Applications, in: IJCAI-07: International Joint Conference on Artificial Intelligence, Hyderabad, India, 2007, pp. 2103 – 2108.
- [7] W. van der Mark, D. M. Gavrila, Real-Time Dense Stereo for Intelligent Vehicles, *IEEE Transactions on Intelligent Transportation Systems* 7 (1) (2006) 38 – 50.
- [8] R. Hartley, A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, New York, NY, USA, 2003.
- [9] D. Comaniciu, P. Meer, Mean Shift: A Robust Approach Toward Feature Space Analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (5) (2002) 603–619.
- [10] L. Di Stefano, M. Marchionni, S. Mattoccia, A Fast Area-Based Stereo Matching Algorithm, *Image and Vision Computing* 22 (12) (2004) 983–1005.
- [11] L. Di Stefano, S. Mattoccia, Fast Stereo Matching for the VIDET System using a General Purpose Processor with Multimedia Extensions, in: CAMP '00: Proceedings of the Fifth IEEE International Workshop on Computer Architectures for Machine Perception (CAMP'00), IEEE Computer Society, Washington, DC, USA, 2000, pp. 356–362.
- [12] A. Klaus, M. Sormann, K. Karner, Segment-Based Stereo Matching Using Belief Propagation and a Self-Adapting Dissimilarity Measure, in: ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition, Vol. 3, IEEE Computer Society, Washington, DC, USA, 2006, pp. 15–18.

# CORRESPONDÊNCIA DENSA PARA SISTEMAS DE VISÃO ESTEREOOSCÓPICA PARA ROBÓTICA MÓVEL

MAURICIO E. STIVANELLO, EDUARDO S. LEAL, NICOLAS PALLUAT, MARCELO R. STEMMER

S2I, Departamento de Automação e Sistemas, Universidade Federal de Santa Catarina  
DAS / CTC / UFSC

88040-900, Florianópolis, SC, BRASIL

E-mails: maustiva@das.ufsc.br, sleal@das.ufsc.br, nicolas@das.ufsc.br,  
marcelo@das.ufsc.br

**Abstract**— The use of a stereo vision system as main source of information in autonomous mobile robots is becoming common. However, stereo matching remains the biggest challenge for these systems. In this paper, different methods of stereo correspondence widely cited in the literature are described. Additionally, a new method of stereo correspondence based on regional support is proposed. Results obtained with algorithms implementing the different methods are then compared, observing the requirements imposed by mobile robotics applications to the stereoscopic vision systems.

**Keywords**— Stereo vision, Stereoscropy, Mobile robots, Computer vision, Stereo matching, Correspondence, Disparity map.

**Resumo**— A utilização de um sistema de visão estereoscópica em robôs autônomos móveis como fonte principal de informação torna-se cada vez mais comum. Entretanto, a correspondência continua a ser o maior desafio para este tipo de sistema. Neste trabalho são descritos diferentes métodos de correspondência densa vastamente citados na literatura. Adicionalmente, um novo método de correspondência densa baseada em suporte regional é proposto. Os resultados obtidos pelos diferentes métodos são comparados, observando os requisitos impostos a sistemas de visão estereoscópica por aplicações de robótica móvel.

**Palavras-chave**— Estereoscopia, Robótica móvel, Visão computacional, Correspondência, Mapas de disparidade.

## 1 Introdução

Robôs autônomos móveis estão cada vez mais presentes na vida cotidiana, graças aos significativos avanços na área [1]. Independente da natureza ou fim, robôs devem interagir com o mundo para alcançar seus objetivos. Para isso, precisam extraír informações a respeito do ambiente onde estão inseridos de maneira a orientar suas ações.

Diferentes abordagens são utilizadas para este fim, baseadas em sensores como sonares, lasers e câmeras. Seguindo o sistema visual humano, muitos sistemas robóticos utilizam um conjunto binocular composto por câmeras, caracterizando um sistema de visão estereoscópica [2]. Através de um sistema como este é possível extraír informação tridimensional da cena observada pelo tratamento de problemas como calibração, correspondência e reconstrução [3].

A correspondência é considerada por muitos como o maior desafio da estereoscopia, o que justifica o grande número de métodos para tratá-la [4]. No presente trabalho<sup>1</sup>, diferentes métodos de correspondência densa são avaliados, tendo como foco a sua utilização em robótica móvel. Adicionalmente, um novo método de correspondência densa é proposto.

Na seção 2 são apresentadas as principais considerações a respeito do processo de correspondência. As diferentes abordagens utilizadas no tratamento do problema são discutidas na seção 3. Na seção 4, um novo método de correspondência densa baseada em

suporte regional é proposto. Os resultados obtidos com algoritmos implementando diferentes métodos são comparado na seção 5. Finalmente, na seção 6 são apresentadas as conclusões do trabalho.

## 2 Correspondência Densa

Através de um sistema de visão estereoscópica é possível recuperar a informação de profundidade de um ponto no espaço a partir da distância relativa entre dois pontos que o representam em imagens capturadas sob diferentes pontos de vista. Esta distância é denominada *disparidade*. A disparidade em um sistema de visão estéreo é obtida pela diferença das coordenadas dos pontos que correspondem às projeções de um ponto no espaço sobre os planos de imagem das câmeras que compõe o sistema.

Sendo assim, para determinar a disparidade é preciso relacionar elementos presentes nas imagens capturadas, identificando quais pontos da imagem da esquerda e da direita são projeções do mesmo ponto na cena tridimensional observada. Este procedimento, denominado correspondência, é um dos tópicos mais pesquisados em visão computacional.

Para realizar a correspondência de pontos, duas questões principais devem ser consideradas: quais elementos serão correspondidos e qual a medida de similaridade será utilizada. Com base em quais pontos serão correspondidos podemos classificar os muitos algoritmos de correspondência existentes em duas grandes classes: os métodos de correspondência *densa* e os métodos de correspondência *esparsa*. Os métodos de correspondência densa procuram corres-

<sup>1</sup> Este trabalho foi financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ).

ponder todos os pontos das imagens. Já os métodos de correspondência esparsa correspondem um conjunto reduzido de pontos relevantes das imagens.

Os métodos de correspondência densa são particularmente interessantes para robótica móvel, visto que fornecem informação de profundidade suficiente para a realização de operações gerais de robótica móvel [5]. Para os algoritmos que implementam tal abordagem, a entrada é um par de imagens capturadas pelo sistema binocular, normalmente retificadas [3]. Tomando uma das imagens como referência, para cada ponto desta imagem é realizada uma busca pelo ponto correspondente na outra imagem, segundo algum critério de similaridade. Para cada par de pontos correspondidos será calculado o valor da disparidade com base na distância relativa entre as coordenadas dos mesmos. O valor da disparidade será armazenado como o valor de intensidade na posição do ponto da imagem de referência em uma terceira imagem, conhecida como *mapa de disparidades*.

Os valores de intensidade para cada ponto do mapa de disparidade armazenam informação de profundidade da cena, uma vez que a disparidade é inversamente proporcional à profundidade [1].

### 3 Métodos de Correspondência Densa de Referência

Oclusões, diferenças na iluminação, distorções radiométricas, distorções projetivas são apenas alguns dos fatores que tornam o problema de correspondência uma tarefa complexa. Por este motivo, diversos métodos têm sido propostos para tratar o problema.

Podemos classificar os algoritmos de correspondência densa como sendo locais ou globais, conforme a estratégia utilizada para corresponder os pontos pertencentes às imagens [4]. As diferentes abordagens são detalhadas nas seções seguintes

#### 2.1 Abordagem Local

Os algoritmos desta classe assumem que os pontos da imagem são envoltos por uma janela de pontos vizinhos onde a disparidade é a mesma. Desta forma, a correspondência para cada um dos pontos é realizada com base na correspondência destas janelas empregando funções de custo baseadas em similaridade.

O método básico utilizado para correspondência segundo esta abordagem é a correlação [3]. Vários algoritmos foram propostos com base neste método. Estes algoritmos apresentam, de maneira geral, um comportamento comum. Dado um par de imagens estéreo, escolha uma das imagens como referência e defina o tamanho  $2n+1$  para uma janela quadrada. Para cada ponto da imagem de referência, faça:

- 1) Centralize a janela no ponto de interesse na imagem de referência.
- 2) Correlacione a janela posicionada na imagem de referência com uma segunda janela, moven-

do-a sobre todas as posições pertencentes a um espaço de busca na segunda imagem.

- 3) Determine o ponto na segunda imagem que corresponde ao ponto de interesse na primeira imagem, selecionando a posição onde a correlação minimiza a função de custo empregada.
- 4) Calcule a disparidade pela diferença da posição central das janelas. Armazene este valor numa terceira imagem, na posição do ponto central da janela na imagem de referência.

A saída é um mapa contendo a disparidade para cada ponto da imagem de referência.

Estes métodos utilizam funções de custo que levam em conta a similaridade dos valores de intensidade de cada ponto. Uma função de custo comumente empregada é a soma das diferenças absolutas (*SAD*), dada por:

$$C_{sad}(x, y, d) = \sum_{i,j=-n}^n \left| D_{le}(x+j, y+i) - D_{ld}(x+d+j, y+i) \right| \quad (1)$$

onde  $d$  é uma dada disparidade,  $D_{le}(x, y)$  e  $D_{ld}(x, y)$  correspondem a uma função de descrição baseada em intensidade referente às imagens esquerda e direita, respectivamente. Uma janela de tamanho  $2n+1$  é utilizada.

A abordagem de seleção de pontos correspondentes baseada unicamente na minimização da função de custo é denominada *Winner-Takes-All* (*WTA*) [4].

A correlação de uma janela simples está sujeita a erros nas descontinuidades de profundidade, devido à presença de oclusões nestas regiões. Para reduzir este problema, é possível empregar uma extensão desta abordagem onde se faz o uso de múltiplas janelas para corresponder os pontos das imagens [6]. Segundo esta proposta, o erro em regiões de descontinuidade pode ser reduzido se somente forem consideradas no cálculo da função de custo as partes da janela de correlação que não introduza este erro.

No processo de correspondência pelos métodos acima descritos a intensidade assume um papel de descritor para cada ponto das imagens. Esta descrição mostra-se pouco robusta, principalmente em regiões de intensidade homogênea. A Figura 1 ilustra o problema existente quando somente é empregado um descritor baseado em intensidade,  $D_t(x, y)$ .

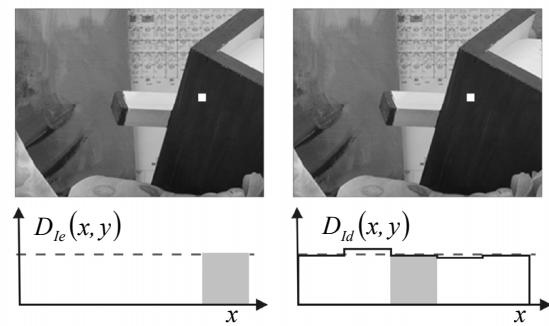


Figura 1. Descrição baseada em intensidade.

Considere a janela de correlação representada pelo ponto branco nas imagens. É possível observar como os valores que descrevem cada ponto podem ser próximos num dado espaço de busca. O emprego de uma função de custo baseada unicamente neste descritor pode resultar na ocorrência de correspondências arbitrárias.

O uso de janelas grandes é uma opção, já que reduz o número de janelas completamente sobrepostas a uma região homogênea. Porém, prejudica o resultado nas descontinuidades de profundidade.

## 2.2 Abordagem Global

Os métodos que seguem a abordagem de correspondência densa global procuram propagar a informação de disparidade de um ponto para seus vizinhos. Para isso, empregam a minimização de alguma função de energia sobre uma grande região da imagem.

O algoritmo de programação dinâmica é um dos algoritmos que calcula a disparidade segundo esta abordagem [7]. Neste caso é empregada uma função de custo que considera o número de oclusões, correspondências e uma medida de dissimilaridade. Para cada linha da imagem é realizada uma busca exaustiva pela melhor solução de disparidade dentre as soluções possíveis, segundo esta função de custo.

A partir das imagens de entrada, é realizado um processo de emparelhamento denso para cada linha. É feita uma medição seguindo um critério de dissimilaridade entre os pontos, atribuindo a cada seqüência de correspondências  $M$  um custo, dado por:

$$\gamma(M) = N_{occ}k_{occ} - N_m k_r + \sum_{i=1}^{N_m} d(m_{ei}, m_{di}) \quad (2)$$

O custo  $\gamma(M)$  mede a probabilidade de que essa seqüência seja uma correspondência correta. Os parâmetros  $N_{occ}$  e  $N_m$  representam o número de oclusões e de emparelhamentos, sendo  $k_{occ}$  e  $N_r$  as respectivas constantes de penalidade e recompensa, e  $d(m_{ei}, m_{di})$  uma função de dissimilaridade baseada em intensidade entre os pontos da imagem esquerda e da imagem direita. A seqüência de emparelhamento ótima é realizada empregando uma busca exaustiva através de um algoritmo de programação dinâmica.

A principal vantagem dos métodos desta classe é o tratamento de regiões homogêneas. Entretanto, geralmente utilizam parâmetros difíceis de definir.

## 3 Método Proposto: Correspondência Regional com Suporte Regional

No processo de correspondência baseado em correlação de janelas descrito na seção 2.1, temos que a intensidade assume o papel de descritor para cada ponto. Porém, que uma descrição baseada em intensidade não é suficiente para identificarmos univocamente um par de pontos correspondentes. O mesmo

ocorre com o sistema visual humano quando observamos uma superfície extremamente lisa ou em um ambiente pouco iluminado. Nestas condições, onde temos a falta de detalhes finos para que possamos orientar a correspondência entre pontos observados pelos olhos, naturalmente recorremos à informação fornecida por elementos de alta freqüência presentes na região periférica para determinar a profundidade.

O método proposto procura modelar este comportamento. A idéia principal é descrever cada ponto das imagens tanto por um descritor de intensidade como por um descritor regional. O descritor regional deve agregar informação periférica proveniente dos componentes de alta freqüência existentes próximos ao ponto observado. Desta forma, podemos definir a função de custo com suporte regional por:

$$C_R(x, y, d) = S_I(x, y, d) + \alpha S_R(x, y, d) \quad (3)$$

onde  $S_I(x, y, d)$  é uma função de similaridade baseada em um descritor de intensidade  $D_I(x, y)$ ,  $S_R(x, y, d)$  correspondente a uma função de similaridade baseada em um descritor regional  $D_R(x, y)$  e  $\alpha$  correspondendo à contribuição do suporte regional.

A Figura 2 ilustra o efeito da descrição de pontos com suporte regional durante a correspondência.

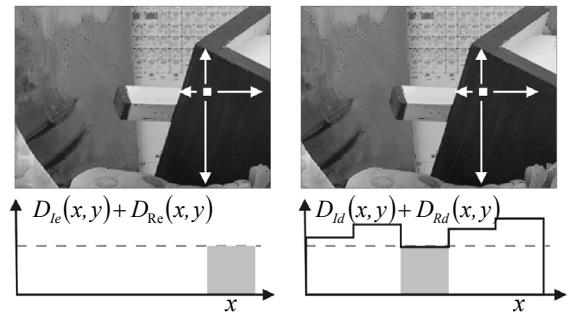


Figura 2. Descrição baseada em intensidade com suporte regional.

Diferente da situação ilustrada pela Figura 1, neste caso são empregados tanto o descritor baseado em intensidade  $D_I(x, y)$  como o descritor regional  $D_R(x, y)$ . É possível observar como uma descrição baseada nesta combinação de descritores permite uma melhor distinção entre os pontos. Uma função de custo baseada nestes descritores permite uma melhor seleção das correspondências corretas.

Uma função de similaridade  $S_R(x, y, d)$  baseada em descritores regionais é dada por:

$$S_R(x, y, d) = \sum_{i,j=-n}^n \sqrt{D_v(x, y, d)^2 + D_h(x, y, d)^2} \quad (4)$$

Por sua vez,  $D_v(x, y, d)$  e  $D_h(x, y, d)$  são dados respectivamente por:

$$D_v(x, y) = \frac{(D_t(x, y))}{(D_t(x, y) + D_b(x, y))} - \frac{(D_t(x+d, y))}{(D_t(x+d, y) + D_b(x+d, y))} \quad (5)$$

e

$$D_h(x, y) = \frac{(D_l(x, y))}{(D_l(x, y) + D_r(x, y))} - \frac{(D_l(x+d, y))}{(D_l(x+d, y) + D_r(x+d, y))} \quad (6)$$

onde  $D_t(x, y)$ ,  $D_b(x, y)$ ,  $D_l(x, y)$  e  $D_r(x, y)$  são descritores regionais que representam a distância entre o ponto observado e o componente de alta freqüência mais próximo acima, abaixo, à esquerda e à direita. Note que  $D_v(x, y, d)$  e  $D_h(x, y, d)$  são adimensionais, a fim de minimizar as diferenças dimensionais introduzidas pela diferença de perspectiva das imagens.

Em função da inserção do termo de suporte regional na função de custo de correspondência, é preciso calcular os descritores regionais para cada uma das imagens. Estes descritores podem ser baseados na distância entre cada ponto e os componentes de alta freqüência que delimitam a região homogênea onde o mesmo está inserido. É importante observar que podemos limitar o suporte regional às regiões onde a intersecção entre as janelas de correlação e componentes de alta freqüência é igual à zero, visto que o desempenho das funções de similaridade baseadas em intensidade é satisfatório nas regiões diferentes destas. Sendo assim, podemos empregar um método de segmentação baseado na variância das intensidades para isolar as regiões homogêneas desejadas.

A segmentação baseada no mapa de variância de intensidades [8] mostra-se interessante pelo seu custo computacional reduzido quando comparado a outras técnicas, e pode ser calculado por:

$$\sigma_L^2(x, y) = \frac{1}{N^2} \sum_{i,j=-n}^n L(x+j, y+i)^2 - \mu_L(x, y)^2 \quad (7)$$

onde  $N = (2n+1)^2$  e  $\mu L(x, y)$  é a média dos valores de intensidade.

A partir do mapa de variância calculado podemos aplicar uma limiarização global para obter uma imagem binária contendo as regiões homogêneas isoladas. A partir desta imagem podemos calcular mapas de distância correspondentes aos descritores regionais  $D_t(x, y)$ ,  $D_b(x, y)$ ,  $D_l(x, y)$  e  $D_r(x, y)$ . Através destes descritores de distância, podemos calcular os mapas dos descritores adimensionais  $D_v(x, y, d)$  e  $D_h(x, y, d)$  diretamente, conforme apresentado nas Equações (5) e (6).

A Figura 3 ilustra o processo de cálculo de descritores regionais apresentados acima, onde é apresentada a imagem original (a), o mapa de variância calculado sobre a imagem original (b), o resultado da binarização do mapa de variância (c) e o mapa de distância correspondente ao descritor  $D_t(x, y)$  (d), calculado sobre a imagem binária.

A partir da equação (3) podemos criar várias funções de custo através da combinação entre dife-

rentes funções de similaridade baseadas em intensidade e região. Empregando a função de similaridade baseada em intensidade apresentada pela equação (1) e a equação de similaridade baseada em região apresentada pela equação (4), podemos reescrever a função de custo com suporte regional dada pela equação (3) como:

$$C_{sadr}(x, y, d) = \sum_{i,j=-n}^n \frac{\left| D_{le}(x+j, y+i) - D_{ld}(x+d+j, y+i) \right|_+}{\alpha \sqrt{D_v(x, y, d)^2 + D_h(x, y, d)^2}} \quad (8)$$

Observe que o cálculo do termo de suporte regional é realizado utilizando os mapas de descritores regionais previamente calculados.

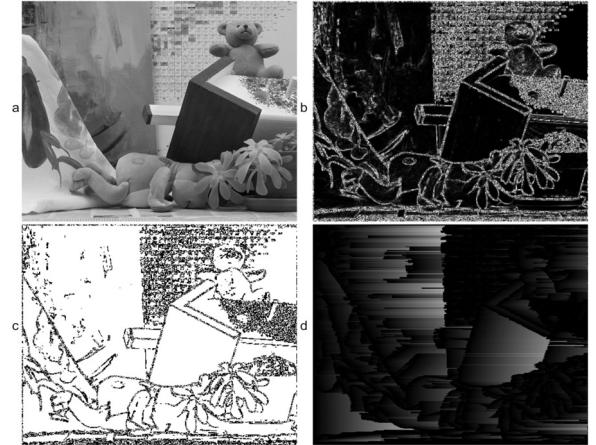


Figura 3. Cálculo dos descritores regionais.

Uma vez calculados os valores de custo no espaço de busca de disparidade, a correspondência correta é selecionada segundo a abordagem *WTA*.

#### 4 Resultados Experimentais

Os resultados obtidos por algoritmos implementando diferentes métodos de correspondência densa foram comparados observando os requisitos existentes em aplicações gerais de robótica móvel. Para isso, foram selecionados métodos representando tanto a abordagem global como a abordagem local, que correspondem às duas grandes classes de métodos de correspondência densa.

Para representar a abordagem local foram implementados algoritmos baseados nos métodos de correlação por janela simples [3] e por múltiplas janelas [6]. Para representar a abordagem global foi selecionado o método de Programação Dinâmica [7], para o qual foi utilizada uma implementação incorporada na biblioteca OpenCV<sup>2</sup>. Adicionalmente, foram implementados dois novos algoritmos pela modificação dos algoritmos locais por correlação de janelas simples e múltiplas pela inclusão de suporte regional descrita na seção 3.

<sup>2</sup> <http://sourceforge.net/projects/opencvlibrary>

Sobre os algoritmos locais, foram incluídas otimizações relacionadas ao cálculo da função de custo [8], com o objetivo de que os mesmos atendessem aos requisitos de tempo real. Para todos os algoritmos foi incorporado ainda o método de verificação de consistência *left-right check* [3], que elimina um grande número de correspondências incorretas.

Uma relação dos algoritmos utilizados na comparação é apresentada na Tabela 1.

Tabela 1. Algoritmos avaliados.

Algoritmo	Técnica	Classe
WTA SAD SW	Correlação de janela simples, função de custo SAD, seleção WTA e consistência left-right check.	Local
WTA SADR SW	Correlação de janela simples, função de custo SAD com Suporte Regional, seleção WTA, consistência left-right check.	Local / Regional
WTA SAD MW	Correlação de múltiplas janelas, função de custo SAD, seleção WTA e consistência left-right check.	Local
WTA SADR MW	Correlação de múltiplas janelas, função de custo SAD com Suporte Regional, seleção WTA e consistência left-right check.	Local / Regional
DP	Programação dinâmica com função de custo baseada em recompensas, penalidades e dissimilaridade e consistência left-right.	Global

Para a avaliação foram utilizados os critérios de custo computacional e qualidade do mapa de disparidade. Os experimentos foram realizados utilizando uma base de imagens estéreo de referência vastamente utilizada na literatura [4], composta de 4 pares de imagens estéreo. Os pares de imagens estéreos são devidamente normalizados e retificados, além de possuírem um mapa de disparidade real associado.

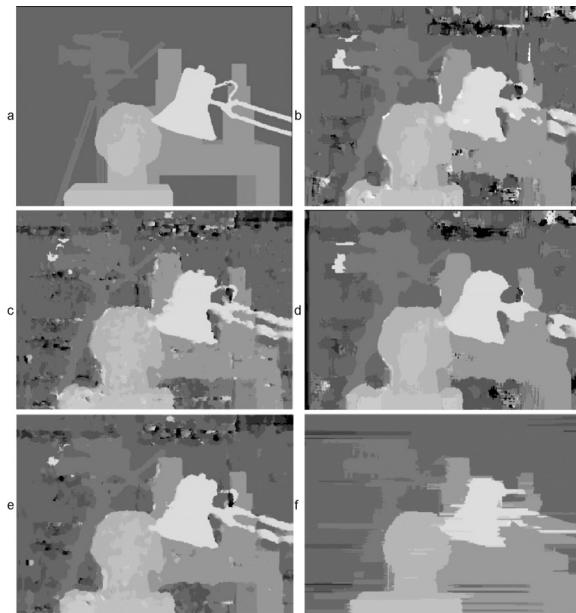


Figura 4. Mapas de disparidade calculados para o par Tsukuba.

Foram gerados mapas de disparidade utilizando cada algoritmo implementado para cada par estéreo da base de referência. A Figura 4 exibe os resultados obtidos para o par de imagens Tsukuba, onde (a)

corresponde ao mapa de disparidade de referência, (b), (c), (d), (e) e (f) correspondem aos mapas calculados pelos algoritmos WTA SAD SW, WTA SADR SW, STA SAD MW, STA SADR MW e DP, respectivamente.

Os algoritmos de correlação por janela simples e múltipla apresentaram bons resultados nas regiões ricas em textura de bordas. Entretanto, ambas apresentam correspondências arbitrárias em regiões homogêneas ou com padrões de texturas repetitivas.

A introdução de suporte regional melhorou o desempenho dos algoritmos locais, principalmente em imagens com regiões homogêneas. Esta característica é claramente visível na comparação dos resultados para o par estéreo Teddy. A Figura 5 apresenta a imagem esquerda (a), o respectivo mapa de disparidade real (b), o mapa de disparidade gerado pelo algoritmo WTA SAD MW (c) e o mapa de disparidade gerado pelo algoritmo WTA SADR MW (d). Enquanto o algoritmo local falha ao corresponder os pontos da região homogênea correspondente ao telhado da pequena casa presente na imagem, sua versão com suporte regional obtém sucesso.

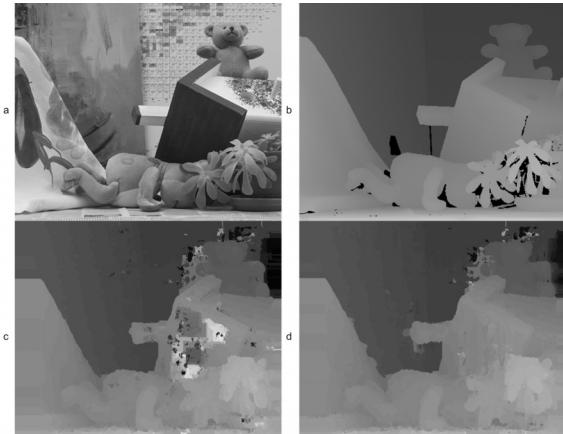


Figura 5. Correspondência em regiões de intensidade homogênea.

O algoritmo global DP apresentou um bom resultado médio, entretanto apresenta a característica de introduzir padrões formados por linhas que muitas vezes descaracterizam as bordas das descontinuidades presentes nas imagens.

Para uma avaliação objetiva, foi utilizada como medida de qualidade a porcentagem de correspondências incorretas [4]. Na Tabela 2 são exibidos os erros gerais obtido no cálculo das disparidades para os diferentes algoritmos avaliados.

Tabela 2. Erros obtidos no cálculo das disparidades.

Algoritmo	Tsukuba	Cones	Teddy	Vênus
WTA SAD SW	14.6	27.5	30.9	17.6
WTA SADR SW	8.45	17.6	26.9	9.45
WTA SAD MW	9.51	22.8	26.7	10.7
WTA SADR MW	<b>5.96</b>	<b>15.0</b>	22.8	<b>7.12</b>
DP	7.98	20.9	<b>20.2</b>	10.7

Como pode ser observado, o algoritmo WTA SADR MW que implementa o método proposto apresentou o menor erro médio, seguido pelo algorit-

mo DP. Avaliando os resultados obtidos entre os métodos WTA SAD SW e WTA SAD MW e as suas variações com suporte regional WTA SADR SW e WTA SADR MW, claramente é observado o ganho introduzido pelo método de suporte regional proposto.

A Tabela 3 apresenta os tempos de processamento em segundos obtido para cada algoritmo na geração dos mapas de disparidade para cada par estéreo.

Tabela 3. Tempos de processamento consumido pelos algoritmos.

Método	Tsukuba	Cones	Teddy	Vénus
WTA SAD SW	0.11	0.68	0.56	0.19
WTA SADR SW	0.52	2.25	2.23	1.05
WTA SAD MW	0.39	2.45	2.44	0.97
WTA SADR MW	1.66	9.32	9.32	3.35
DP	<b>0.09</b>	<b>0.35</b>	<b>0.25</b>	<b>0.16</b>

Como pode ser observado pelos tempos exibidos, o método global DP apresentou o melhor tempo para todos os pares, seguido do algoritmo do método local WTA SAD SW. Pode-se considerar pelos tempos apresentados por estes algoritmos que ambos atendem ao requisito de tempo real. O mesmo não ocorre com o método local WTA SAD MW e com os métodos locais com suporte regional WTA SADR SW e WTA SADR MW, principalmente nas imagens Teddy e Cones onde o espaço de busca é maior.

Verificou-se ainda que para aplicações com alto requisito de tempo real cuja operação não exija muita precisão pode ser interessante utilizar algoritmos que implementem técnica empregando janela simples, ao invés da técnica implementando múltiplas janelas. Apesar de que com técnicas de múltipla janela foram obtidas melhorias nas regiões de descontinuidade onde existem oclusões, a diferença pode ser considerada desprezível para tarefas como, por exemplo, navegação de um robô. Já a diferença entre os tempos obtidos por algoritmos de janela simples e múltiplas janelas correspondentes não são desprezíveis.

#### 4 Conclusão

No presente trabalho diferentes métodos de correspondência densa vastamente citados na literatura, assim como uma nova proposta de correspondência baseada em suporte regional, são avaliados no contexto da utilização em sistemas de visão estereoscópica para robótica móvel.

Os métodos locais mostram-se muito pouco robustos em regiões de intensidade homogênea. A grande ocorrência de regiões como estas, principalmente em cenas estruturadas, dificulta a sua utilização efetiva. O algoritmo global mostrou-se mais robusto a estas situações, entretanto descaracteriza a forma geral dos objetos presentes nas imagens. Isto pode dificultar operações de robótica móvel como reconhecimento ou manipulação. O algoritmo de correspondência local com suporte regional proposto apresentou resultados intermediários.

O método proposto obteve o menor erro médio no cálculo dos mapas de disparidade. Entretanto, para seu uso efetivo em aplicações de robótica móvel faz-se necessário um maior esforço para a otimização dos algoritmos que o implementam, visto o forte requisito de tempo real. Sugere-se que, para isso, sejam explorados os recursos disponibilizados pelas arquiteturas dos processadores atuais ou mesmo pelas unidades de processamento gráfico [8].

Visto que o algoritmo DP apresentou o menor custo computacional e o segundo menor erro na geração dos mapas de disparidade para todos os pares da base de referência, considera-se que o algoritmo representa o melhor compromisso entre precisão e tempo de processamento para aplicações de robótica móvel.

#### Referências Bibliográficas

- SIEGWART, R.; NOURBAKSH, I. R. Introduction to Autonomous Mobile Robots. 1 ed. MIT Press, 2004.
- STIVANELLO, M. E. Desenvolvimento de uma Biblioteca para Sistemas de Visão Estereoscópica. Florianópolis, 2008. Dissertação (Mestrado em Engenharia Elétrica) – Centro Tecnológico, Universidade Federal de Santa Catarina.
- TRUCCO, E.; VERRI, A. Introductory Techniques for 3-D Computer Vision, 1 ed. Prentice Hall, 2003.
- SCHARSTEIN, D.; SZELINSKI, R. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *International Journal on Computer Vision*, v. 47, p. 7-42, June 2002.
- IWASHITA, S.; MURASE, Y.; YASUKAWA, Y. et al. Developing a Service Robot. *IEEE International Conference on Mechatronics and Automation*, v. 2, n. 2, p. 1057-1062, Aug. 2005.
- HIRSCHMULLER, H.; INNOCENT, P. R.; GARIBALDI, J. M. Real-Time Correlation-Based Stereo Vision with Reduced Border Errors. *International Journal on Computer Vision*, v. 47, n. 1, p. 229-246, Jan. 2002.
- BIRCHFIELD, S.; TOMASI, C. Depth Discontinuities by Pixel-to-Pixel Stereo. In: *IEEE International Conference on Computer Vision* (1998: Bombay), Proceedings. India. p. 1073-1080.
- STEFANO, L. D.; MATTOCCIA, S. Fast Stereo Matching for the VIDET System using a General Purpose Processor with Multimedia Extensions. In: *IEEE CAMP'00* (2000: Padova), Proceedings. Padova, 2000. p. 356-362.

# Referências Bibliográficas

- [1] IWASHITA, S.; MURASE, Y.; YASUKAWA, Y. et al. Developing a Service Robot. *IEEE International Conference on Mechatronics and Automation*, v. 2, n. 2, p. 1057-1062, Aug. 2005.
- [2] BIEGELBAUER, G.; VINCZE, M. 3-D Vision-Guided Bore Inspection System. *IEEE International Conference on Computer Vision Systems*, v. 4, n. 7, p. 22-22, Jan. 2006.
- [3] DEVARAKOTA, P. R. R.; MIRBACH, B.; CASTILLO-FRANCO, M. et al. 3-D Vision Technology for Occupant Detection and Classification. *3DIM Fifth International Conference on 3-D Digital Imaging and Modeling*, v. 13, n. 16, p. 72-79, June 2005.
- [4] VAUGHAN, R. T. Stage: A Múltiple Robot Simulator. Technical Report IRIS-00-394, Institute for Robotics and Intelligent Systems, School of Engineering, University of Southem, California, 2000.
- [5] DARPA. Darpa Chalenge, 2007. Disponível em: <http://www.darpa.mil/grandchallenge/index.asp>. Acesso em 01/11/2007.
- [6] DARPA. Darpa Chalenge, 2007. Disponível em: <http://www.darpa.mil/grandchallenge/index.asp>. Acesso em 01/11/2007.
- [7] DARPA. Tactical Technologic Office, 2006. Disponível em: <http://www.darpa.mil/TTO/programs.htm>. Acesso em 01/11/2007.
- [8] MIT. Humanoid Robotics Group, 2007. Disponível em: <http://www.ai.mit.edu/projects/humanoid-robotics-group/>. Acesso em 01/11/2007.
- [9] KRAGIC, D.; BJORKMAN, M.; CHRISTENSEN, H. I. et al. Vision for Robotic Object Manipulation in Domestic Settings. *Robotics and Autonomous Systems*, v. 52, n. 1, p. 85-100, July 2005.
- [10] ASIMOV, I. Eu, Robo. 2. ed. Edibolso, 1976.
- [11] SIEGWART, R.; NOURBAKHSH, I. R. Introduction to Autonomous Mobile Robots. 1 ed. MIT Press, 2004.

- [12] TRUCCO, E.; VERRI, A. Introductory Techniques for 3-D Computer Vision, 1 ed. Prentice Hall, 2003.
- [13] GASPAR, J. A. C. P. *Visão para Robótica Móvel: Detecção de Obstáculos sobre Pavimento Plano*. Lisboa, 1994. Dissertação (Mestrado em Engenharia Eletrotécnica e de Computadores) - Universidade Técnica de Lisboa.
- [14] HARTLEY, R.; ZISSERMAN, A. Multiple View Geometry in Computer Vision. 2. ed. Cambridge University Press, 2003.
- [15] TSAI, R. Y. A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-The-Shelf TV Cameras and Lenses. *IEEE Journal of Robotics and Automation*, v.3, n. 4, p. 323-344, Aug. 1987.
- [16] ZHANG, Z. A Flexible New Technique for Camera Calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 22, n. 11, p. 1330-1334 Nov. 2000.
- [17] FRANÇA, J. A.; STEMMER, M. R.; FRANÇA, M. B. M. Calibração Monocular com Gabaritos 1D: Uma Ferramenta Prática com o Matlab. In: Simpósio Brasileiro de Automação Inteligente (Oct. 2007: Florianópolis). *Anais*. Florianópolis, 2007.
- [18] BENEVENUTI, F. *Calibração de um Conjunto Binocular para Uso em Estereoscopia e Robótica Móvel*. Florianópolis, 2001. Dissertação (Mestrado em Engenharia Elétrica) – Centro Tecnológico, Universidade Federal de Santa Catarina.
- [19] FRANÇA, J. A. *Calibração de um Sistema de Visão Estéreo: de Correspondência de Pontos à Reconstrução Euclidiana*. Florianópolis, 2005. Tese (Doutorado em Engenharia Elétrica) – Centro Tecnológico, Universidade Federal de Santa Catarina.
- [20] SCHARSTEIN, D.; SZELINSKI, R. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *International Journal on Computer Vision*, v. 47, p. 7-42, June 2002.
- [21] FUSIELLO, A.; TRUCCO, E.; VERRI, A. A Compact Algorithm for Rectification of Stereo Pairs. *Machine Vision and Applications*, v. 12, n. 1, p.16-22, Jan. 2000.
- [22] HIRSCHMULLER, H.; INNOCENT, P. R.; GARIBALDI, J. M. Real-Time Correlation-Based Stereo Vision with Reduced Border Errors. *International Journal on Computer Vision*, v. 47, n. 1, p. 229-246, Jan. 2002.
- [23] MARR, D.; POGGIO, T. Cooperative Computation of Stereo Disparity, *Science*, v. 194, p. 209-236, 1976.

- [24] MARK, V. D.; GAVRILLA, D. M. Real-Time Dense Stereo for Intelligent Vehicles. *IEEE Transactions on Intelligent Transportation Systems*, v. 7, n. 1, p. 38-50, Mar. 2006.
- [25] STEFANO, L. D.; MATTOCCIA, S. Fast Stereo Matching for the VIDET System using a General Purpose Processor with Multimedia Extensions. In: IEEE CAMP'00 (2000: Padova), *Proceedings*. Padova, 2000. p. 356-362.
- [26] STEFANO, L. D.; MARCHIONNI, M.; MATTOCCIA, S. A PC-based Real-Time Stereo Vision System. *Machine Graphics & Vision International Journal*, v. 13, n. 3, p. 197-220, Jan. 2004.
- [27] BIRCHFIELD, S.; TOMASI, C. Depth Discontinuities by Pixel-to-Pixel Stereo. In: IEEE International Conference on Computer Vision (1998: Bombay), *Proceedings*. India. p. 1073-1080.
- [28] FOGLIA, P.; JOLION, J. M. A New Approach for Stereo Matching in Autonomous Mobile Robot Applications. In: International Joint Conference on Artificial Intelligence (Jan. 2007: India), *Proceedings*, India, 2007. p. 2103-2108.
- [29] KLAUS, A.; SORMANN, M.; KARNER, K. Segment-Based Stereo Matching Using Belief Propagation and a Self-Adapting Dissimilarity Measure. *IEEE International Conference on Pattern Recognition*, v. 3, n. 1, p. 15-18, 2006.
- [30] ACHOUR, K.; MAHIDDINE, L. Hopfield Neural Network Based Stereo Matching Algorithm. *Journal of Mathematical Imaging and Vision*, v.16, n. 1, p. 17-29, Jan. 2002.
- [31] FUA, P. V. Combining Stereo and Monocular Information to Compute Dense Depth Maps. International Joint Conferences on Artificial Intelligence (1991), *Proceedings*, 1991. p. 1292-1298.
- [32] ZITNICK, C. L.; KANADE, T. A Cooperative Algorithm for Stereo Matching and Occlusion Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 22, n. 7, p. 675-684, July 2000.
- [33] LIM, H. S.; BINFORD, T. O. Stereo correspondence: A hierarchical approach. In: IEEE Image Understanding Workshop (1987), *Proceedings*, 1987. p. 234-241.
- [34] BHALLA, K.; DURDLE, N. G. et al. Automatic Feature Detection and Correspondence in a Stereo-Vision Application. In: IEEE International Conference on Systems, Man and Cybernetics (1995), *Proceedings*. 1995. p. 3537-3542.

- [35] HARRIS, C.; STEPHENS, M. A Combined Corner and Edge Detection. In: Fourth Alvey Vision Conference (1988), *Proceedings*, 1988. p. 147-151.
- [36] CANNY, J. F. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 8, n. 6, p. 679-698, Nov. 1986.
- [37] LOWE, D. G. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, v.60, n. 2, p. 91-110, 2004.
- [38] FRANÇA, J. A.; FRANÇA, M. B.; STEMMER, M. R. Stereo-Based Detection and Localization of Obstacles in Indoor Environments. In: VI Induscon (2004: Joinville), Brazil, 2004.
- [39] JENNINGS, C.; MURRAY, D. Stereo Vision Based Mapping and Navigation for Mobile Robots. In: IEEE International Conference on Robotics and Automation (1997), *Proceedings*. 1997. p. 1694-1699.
- [40] GONZALES, R. C.; WOODS, R. E. Processamento de Imagens Digitais, 1 ed. Edgard Blucher, 2000.
- [41] STIVANELLO, M. E.; LEAL, E. S.; PALLUAT, N. et al. Stereo Vision Based on Local Correspondence with Regional Support. *Elsevier - Image and Vision Computing Journal*, Ms. No. IMAVIS-D-07-00359, Submetido em 09/11/2007.
- [42] STIVANELLO, M. E.; LEAL, E. S.; PALLUAT, N. et al. Correspondência Densa para Sistemas de Visão Estereoscópica para Robótica Móvel. In: XVII Congresso Brasileiro de Automática (2008: Juiz de Fora), Brazil, Submetido em 16/03/2008.