

UNIVERSIDADE FEDERAL DE SANTA CATARINA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

**CALIBRAÇÃO DE UM CONJUNTO BINOCULAR PARA USO EM  
ESTEREOSCOPIA E ROBÓTICA MÓVEL**

Dissertação submetida à Universidade Federal de Santa Catarina  
como requisito parcial à obtenção do grau de

**Mestre em Engenharia Elétrica**

por

**Fabio Benevenuti**



03438286

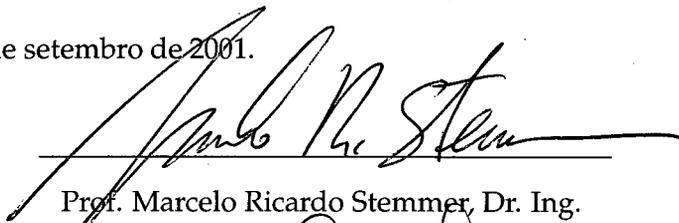
Florianópolis, 5 de setembro de 2001

**CALIBRAÇÃO DE UM CONJUNTO BINOCULAR PARA USO EM ESTEREOSCOPIA  
E ROBÓTICA MÓVEL**

**Fabio Benevenuti**

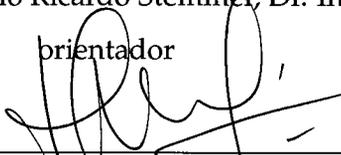
Esta dissertação foi julgada adequada para a obtenção do título de **Mestre em Engenharia** na especialidade **Engenharia Elétrica**, área de concentração **Controle, Automação e Informática Industrial**, e aprovada em sua forma final pelo curso de Pós-Graduação.

Florianópolis, 5 de setembro de 2001.



Prof. Marcelo Ricardo Stemmer, Dr. Ing.

orientador



Prof. Edson Roberto De Pieri, Dr.

Coordenador do Programa de Pós-Graduação em Engenharia Elétrica  
da Universidade Federal de Santa Catarina

**Banca Examinadora**



Prof. Marcelo Ricardo Stemmer, Dr. Ing.

orientador



Prof. Aldo von Wangenheim, Dr. Rer. Nat.



Prof. Guilherme Bittencourt, Dr. Rer. Nat.

*Copyright* ©2001 Fabio Benevenuti

## *Agradecimentos*

*Agradeço primeiramente a todos os amigos que conheci através deste programa de pós-graduação. Em especial merecem ser citados, sem desprezar os demais, os nomes de Alexandre Keller Albalustro e Marcos Banheti Rabello Vallim.*

*Também merece meu agradecimento Wojciech Buła pelo exemplo de perseverança em direção a um objetivo e pelo constante incentivo.*

*Finalmente gostaria de agradecer ao CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico, pela bolsa de estudo recebida entre os meses de junho de 1999 e fevereiro de 2000.*

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

## **CALIBRAÇÃO DE UM CONJUNTO BINOCULAR PARA USO EM ESTEREOSCOPIA E ROBÓTICA MÓVEL**

**Fabio Benevenuti**

setembro/2001

Orientador: Marcelo Ricardo Stemmer, Dr. Ing.

Área de Concentração: Controle, Automação e Informática Industrial

Palavras-chave: Calibração de câmeras, estereoscopia, geometria epipolar.

Número de Páginas: 112

Uma imagem obtida por uma câmera de vídeo é uma projeção em perspectiva do ambiente e somente com esta imagem não é possível determinar as coordenadas dos objetos no ambiente. Porém, se tomamos duas imagens obtidas por duas câmeras cujas configurações e posição e orientação relativas são conhecidas, torna-se possível determinar a posição dos objetos por triangulação. Estas configurações, a geometria do conjunto binocular, podem ser definidas *a priori* na construção e montagem das câmeras, porém sua manutenção ao longo do tempo não é possível em aplicações práticas de robótica móvel. Este trabalho relata algumas experiências de determinação da geometria do conjunto binocular a partir de imagens obtidas pelas câmeras, um processo que é chamado de calibração de câmeras. Estas imagens podem ser tomadas do ambiente em geral ou de algum objeto especial do ambiente cuja geometria é conhecida, o que chamamos de gabarito de calibração. Antes que se possa calibrar o conjunto binocular é necessário estabelecer correspondências entre os caracteres do gabarito e os caracteres detectados nas imagens ou entre os caracteres detectados em ambas as imagens. Dois métodos de correspondência foram testados e também dois métodos de calibração, em especial destacando a correspondência por relaxação e a calibração pela geometria epipolar. Os resultados dos experimentos demonstram que a precisão obtida com estes métodos é suficiente para o uso em robótica móvel no contexto da RoboCup, que é a aplicação aqui estudada.

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

**CALIBRATION OF A BINOCULAR HEAD FOR USE IN STEREOSCOPIC VISION  
AND MOBILE ROBOTICS**

**Fabio Benevenuti**

September/2001

Advisor: Marcelo Ricardo Stemmer, Dr. Ing.

Area: Control, Automation and Industrial Computing

Keywords: Câmera calibration, stereoscopy and epipolar geometry.

Number of Pages: 112

One image obtained from a single camera is a perspective projection of the world and this image solely is not enough to determine the objects coordinates in the world. If we take two images from two cameras whose configuration and relative position and orientation is known, it's possible to determinine the objects position through triangulation. These configurations, the binocular head geometry, can be defined a priori while fabricating and mounting the cameras, however their maintainence over time is not possible in practical mobile robotics applications. This work reports some experiments in determining the binocular head geometry using images obtained from the cameras, a process known as camera calibration. These images can be taken from the environment in general or from some special objects whose geometry is known, which is called calibration apparatus. Before we can do the calibration it is necessary to stablish correspondences, matches, between features known in the apparatus and features detected in the images or between the features detected in both images. Two methods of feature matching and two methods of camera calibration were evaluated, specially the feature matching through relaxation and camera calibration through epipolar geometry. The experiments results show that the precision obtained with these methods suffices to the use in mobile robotics in the context of the RoboCup, which is the application studied in this work.

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	O caso RoboCup . . . . .	1
1.2	Visão computacional e estereoscopia . . . . .	4
1.3	Localização em robótica móvel . . . . .	7
1.4	Motivação . . . . .	9
1.5	Objetivos do trabalho . . . . .	10
1.6	Estrutura do trabalho . . . . .	11
<b>2</b>	<b>Detecção e correspondência de caracteres</b>	<b>12</b>
2.1	Detecção de caracteres primitivos . . . . .	12
2.1.1	Detecção de vértices . . . . .	14
2.1.2	Detecção de blocos distinguíveis . . . . .	14
2.1.3	Expansão de regiões . . . . .	16
2.2	Correspondência de caracteres . . . . .	19
2.3	Correspondência por relaxação . . . . .	29
2.4	Resumo . . . . .	31
<b>3</b>	<b>Estereoscopia, calibração de câmeras e reconstrução</b>	<b>32</b>
3.1	Sensores de imagem . . . . .	32
3.1.1	Câmera <i>pinhole</i> e câmera com lentes . . . . .	32
3.1.2	Matriz de sensores e digitalização . . . . .	35
3.2	Visão estereoscópica . . . . .	38
3.2.1	Disparidade estereoscópica e profundidade . . . . .	38
3.2.2	Projeção em perspectiva e modelagem da câmera . . . . .	40
3.2.3	Geometria epipolar e a matriz fundamental . . . . .	44

3.3	Reconstrução . . . . .	48
3.4	Calibração de câmeras . . . . .	50
3.4.1	Distorção radial . . . . .	50
3.4.2	Calibração de uma câmera - método de Tsai . . . . .	51
3.4.3	Calibração de uma câmera - método alternativo . . . . .	53
3.4.4	Calibração de duas câmeras - geometria epipolar . . . . .	55
3.5	Resumo . . . . .	60
<b>4</b>	<b>Calibração de um conjunto binocular simplificado</b>	<b>62</b>
4.1	O conjunto binocular utilizado . . . . .	62
4.1.1	As câmeras . . . . .	62
4.1.2	O cartão de aquisição . . . . .	63
4.1.3	O conjunto binocular . . . . .	64
4.2	Deteção e correspondência de caracteres . . . . .	66
4.2.1	Deteção de regiões . . . . .	67
4.2.2	Deteção de vértices . . . . .	69
4.2.3	Correspondência de caracteres . . . . .	71
4.3	Fase 1 - avaliação da distorção radial . . . . .	76
4.4	Fase 2 - calibração com gabarito . . . . .	79
4.4.1	A escolha do gabarito . . . . .	79
4.4.2	Avaliação do erro de localização . . . . .	81
4.5	Fase 3 - calibração a partir de caracteres do ambiente . . . . .	84
4.6	Fase 4 - reconstrução métrica a partir da matriz fundamental . . . . .	90
4.7	Resumo . . . . .	96
<b>5</b>	<b>Conclusões e sugestões</b>	<b>97</b>
5.1	Conclusões . . . . .	97
5.2	Perspectivas futuras . . . . .	100
<b>A</b>	<b>Plataforma de software utilizada</b>	<b>102</b>
A.1	Ambiente de desenvolvimento . . . . .	102
A.2	Bibliotecas de funções . . . . .	102
A.3	Aplicações e utilitários desenvolvidos . . . . .	104

# Capítulo 1

## Introdução

*"Everything should be made as simple as possible, but not simpler."*

*Albert Einstein*

### 1.1 O caso RoboCup

Este trabalho foi iniciado dentro de um estudo maior envolvendo a construção de um time de robôs autônomos móveis pertencentes ao UFSCTeam [6], adaptados para competições de futebol, como é o caso da RoboCup [17].

A RoboCup tem como objeto promover pesquisas que integrem inteligência artificial e robótica. Para isto a RoboCup inclui competições e eventos para divulgação das pesquisas. Nas competições times de robôs móveis competem jogando futebol, em várias categorias. A categoria que se toma como referência neste trabalho é a categoria *small-size league*, onde cada time possui no máximo cinco robôs.

As regras da RoboCup determinam características específicas para a campo de jogo e para os robôs de acordo com a categoria. Na categoria *small-size league* as principais características são:

#### Campo de jogo

- Campo de jogo tem 274cm de comprimento por 152,5cm de largura, pintado de verde; as dimensões do campo de jogo e cores são as mesmas utilizadas em tênis de mesa;
- As paredes laterais do campo têm 10cm de altura, pintadas de branco;

- Gol tem 50cm de largura e 18cm de profundidade;
- As paredes do gol têm 15cm de altura, pintadas de um lado de amarelo e do outro de azul;
- Círculo central tem 25cm de diâmetro, delimitado com uma faixa de 1cm de largura, pintada de branco;
- A área penal tem 22,5cm de largura por 100cm de comprimento, delimitada por uma faixa de 1cm de largura, pintada de branco;
- A linha mediana é delimitada por uma faixa de 1cm de largura, pintada de branco;
- Cada canto do campo de jogo possui um anteparo para evitar que a bola fique presa, o qual é pintado de branco com duas faixas verticas paralelas de 1cm de largura, pintada de verde;
- A bola de jogo é uma bola de golf, pintada de alaranjado.

### Jogadores

- Os jogadores possuem altura máxima de 22,5cm;
- A área total de piso ocupada pelo robô não deve ultrapassar 180cm<sup>2</sup>;
- Todo robô deve se ajustar dentro de um cilindro de 18cm de diâmetro;
- Cada robô possui uma marcação no alto de seu corpo, que consiste em uma bola de tênis de mesa, pintada de amarelo ou de azul, dependendo do time do robô;
- O robô pode possuir quaisquer marcações adicionais, desde que não sejam amarelo, azul, verde ou alaranjado;
- É recomendado que o corpo do robô seja pintado de preto.

A figura 1.1 mostra uma vista sintética do campo de jogo e as marcações utilizadas pelos jogadores. A figura 1.2 mostra o mesmo ambiente do ponto de vista aproximado de um robô posicionado no ponto central do campo de jogo orientado para o gol da esquerda.

O presente trabalho trata de aspectos ligados ao sistema de visão destes robôs. Nas primeiras versões da RoboCup, e visão era fornecida por uma única câmera situada acima do campo de jogo. Versões futuras deverão utilizar câmeras embarcadas nos próprios robôs, o que torna o problema da visão computacional bem mais realista e ao mesmo tempo mais complexo. O presente trabalho já aborda questões ligadas à visão nesta nova situação.

Neste contexto, um mecanismo de visão computacional utilizando estereoscopia pode ser utilizado em cada jogador permitindo que estes reconheçam os objetos do jogo e deter-

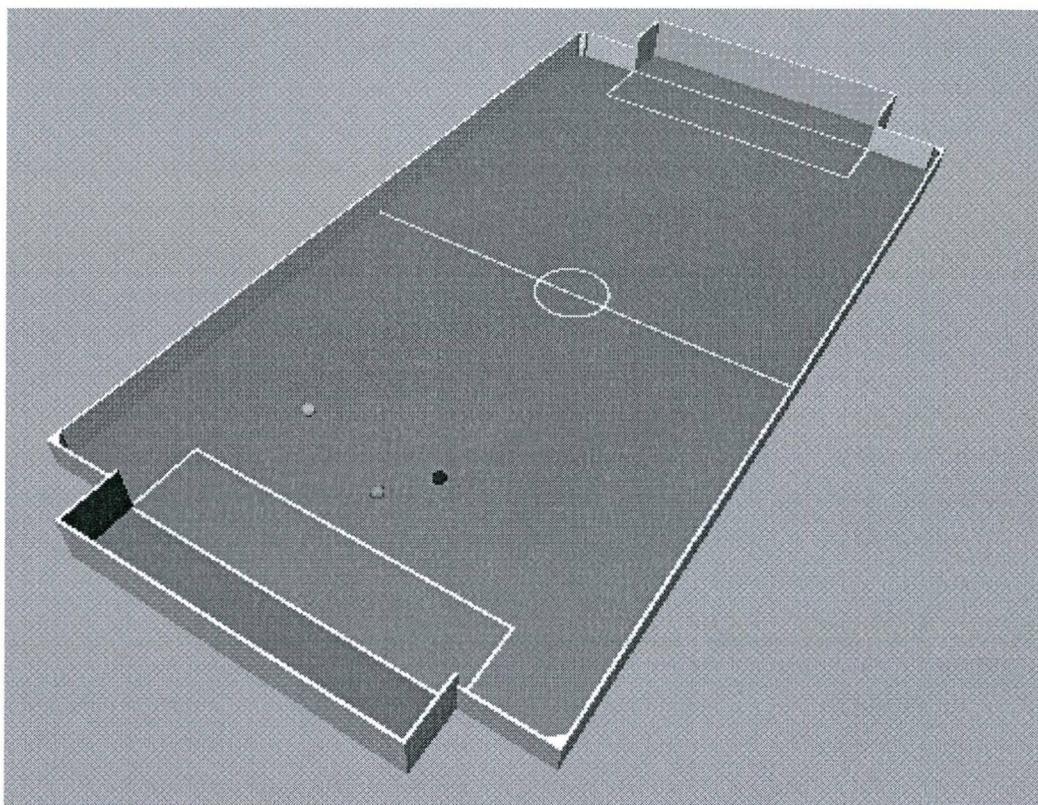


Figura 1.1: Vista sintética do campo de jogo

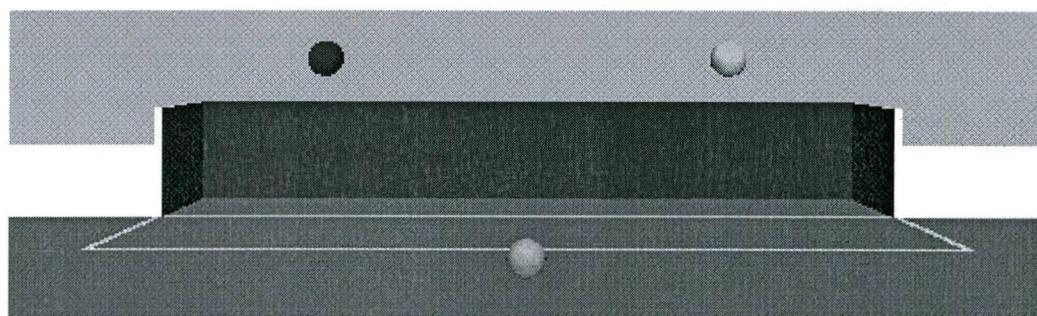


Figura 1.2: Vista sintética do campo de jogo a partir do ponto central

minem suas posições relativas ao jogador, no sistema de coordenadas de cada robô. Adicionalmente este mesmo mecanismo pode ser utilizado para determinar a posição absoluta do jogador no campo de jogo. As características do campo de jogo e do robô citadas acima nos fornecem a topologia e as dimensões da área de trabalho dos robôs e nos ajudam a dimensionar a precisão exigida do mecanismo de visão computacional.

## 1.2 Visão computacional e estereoscopia

Pode-se definir a visão computacional como o processo de extrair, caracterizar e interpretar informações de imagens do mundo tridimensional. Uma diferenciação entre a visão computacional e outras áreas correlatas é feita por Gomes e Velho [10] utilizando como critério a natureza da entrada e saída dos sistemas em cada área, o que está representado no diagrama da figura 1.3.

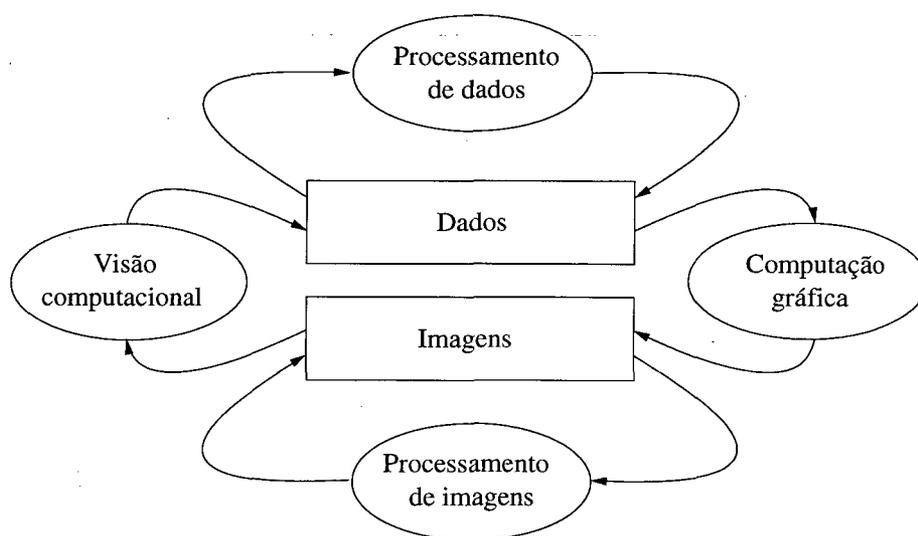


Figura 1.3: Visão computacional e áreas correlatas (extraída de [10])

De especial interesse para este trabalho são as áreas de visão computacional e de processamento de imagens. Segundo Gomes e Velho [10] a visão computacional tem por finalidade obter, a partir de uma imagem (entrada), informações geométricas, topológicas ou físicas sobre o ambiente que deu origem a essa imagem. No processamento de imagens os sistemas admitem como entrada uma imagem que, após processada, produz uma outra imagem de saída. Um exemplo deste tipo de processamento pode ser o realce de detalhes. Na visão computacional a imagem está presente na fase inicial do processo e o processamento de

imagens é dito pré-processamento.

Este processo de visão computacional pode ser dividido em três estágios, de acordo com o nível de sofisticação necessário em seu desenvolvimento: visão de baixo, médio e alto nível. Embora não existam fronteiras nítidas entre estas subdivisões, elas proporcionam uma forma de caracterizar os processos inerentes a um sistema de visão computacional [8, 12].

Fu *et al.* [8] tratam amostragem e pré-processamento como funções de visão de baixo nível, que vai desde o próprio processo de formação da imagem até compensações como redução de ruído e suavização, e finalmente à extração de caracteres primitivos da imagens tais como descontinuidades na intensidade. Associam à visão de médio nível aqueles processos que extraem, caracterizam e rotulam componentes de uma imagem resultante da visão de baixo nível. A segmentação, descrição e reconhecimento de objetos individuais são funções da visão de médio nível. A visão de alto nível se refere a processos que se propõem a emular o conhecimento [8].

A visão computacional pode ser classificada de várias formas, a tabela 1.1 resume algumas destas classificações.

Tabela 1.1: Classificações da visão computacional

Quanto a...	Tipo	Descrição
Interação com o ambiente	Ativa	Modificam as condições ambientais, geralmente a iluminação, como é o caso do uso de luz estruturada onde padrões de luz são projetados sobre os objetos e a forma do objeto é determinada a partir das distorções observadas
	Passiva	O ambiente não é modificado, somente a configuração dos sensores de imagem, como a posição das câmeras, a distância entre elas, o foco e a exposição
Controle da posição da câmeras	Propositiva	Câmeras possuem vários graus de liberdade, podendo modificar seus parâmetros de forma a alcançar uma posição ou configuração que permite observar com mais detalhe a área de interesse no ambiente
	Reconstrutivista	Câmera rigidamente acoplada ou com poucos graus de liberdade, observa o ambiente do ponto de vista para o qual a geometria da câmera é conhecida

Incorporar estereoscopia em um sistema de visão computacional tem a vantagem de fornecer informação de profundidade. Quando um sistema de visão computacional observa o mundo de pontos de vista ligeiramente diferentes pode ser percebido um deslocamento relativo na posição dos objetos, o que é chamado de *disparidade estereoscópica*. A quantidade de disparidade em um objeto indica sua distância em relação ao observador.

Merece ser citada aqui a diferença entre o estéreo fotométrico e o geométrico: no primeiro caso temos um mesmo ambiente sendo observado por uma mesma câmera, numa mesma posição, ou seja, todos os objetos do ambiente e a camera estão fixos, porém com uma mudança de iluminação entre uma imagem e outra, variando-se a posição ou intensidade da iluminação; no segundo caso o ambiente é observado de dois pontos de vista diferentes e, após detectados os caracteres primitivos ou objetos de mais alto nível presentes em ambas as imagens pode-se calcular a posição destes caracteres ou objetos no espaço tridimensional utilizando triangulação.

Existem algumas desvantagens no uso da visão estereoscópica. A principal é o custo adicional, uma vez que estereoscopia exige uma câmera adicional e capacidade computacional para processamento. O processamento adicional exigido também aumenta o tempo de resposta do sistema. A capacidade de um sistema de visão estereoscópica de determinar a distância absoluta de um objeto até a câmera também é limitada. Câmeras não alinhadas, resolução de imagem limitada, distorção das lentes e erros na determinação da correspondência são aspectos que levam à imprecisão no processo de cálculo da profundidade [24].

Para se aplicar estereoscopia a geometria do conjunto binocular precisa ser conhecida, para isso faz-se a calibração das câmeras. Calibração de câmera é o processo de determinação dos parâmetros internos e externos da câmera. Os parâmetros internos estão associados com a distância focal, razão de aspecto do píxel e o centro de projeção das câmeras, enquanto os parâmetros externos tratam da rotação e translação da câmera em relação a um sistema de coordenadas de referência.

Simplificadamente a calibração pode se dar em dois momentos:

- No momento estático: quando o conjunto binocular não está em operação e, geralmente, não se dispõe de qualquer informação de calibração anterior; neste momento pode-se utilizar gabaritos especiais de calibração, cuja geometria tridimensional é conhecida.
- No momento dinâmico: quando o conjunto binocular já está em operação, tendo sido

previamente calibrado, e procura-se manter esta calibração atualizada considerando possíveis mudanças na geometria impostas pelo aquecimento dos sensores, das lentes, impactos e vibração. Em alguns casos o próprio algoritmo de controle de conjunto binocular pode detectar perdas ou falhas na calibração. No caso de operação contínua não se pode utilizar um gabarito especial de calibração e logo toda a informação disponível para a calibração é a rigidez do ambiente e o conhecimento da estrutura tridimensional disponível dos últimos ciclos de processamento. Além disso, quando se trata de um sistema de visão propositivo, a geometria do conjunto binocular é modificada a todo momento pelo próprio algoritmo de controle ao buscar melhores condições de observação do ambiente.

### 1.3 Localização em robótica móvel

A determinação da posição e da orientação de um robô em seu ambiente de trabalho é um problema central em robótica móvel. Este conjunto de posição e orientação também é conhecido como localização.

Leonard e Durrant-Whyte citados por Borenstein *et al.* [4] dividem o problema geral de navegação de robôs móveis em três questões: "Onde estou?", "Onde estou indo?" e "Como posso chegar lá?".

Este trabalho procura fornecer ferramentas que auxiliem na resposta da primeira destas questões, que no caso específico da RoboCup pode ser enunciada de forma ligeiramente diferente: "Onde está o que vejo de onde estou?". Respondendo-se esta questão e dispondo-se de um mapa do ambiente pode-se responder a questão original "Onde estou?".

Borenstein *et al.* [4] agrupa as várias soluções para o problema de localização de robôs móveis em duas grandes categorias: medidas absolutas e medidas relativas de posição.

De especial interesse neste trabalho, o reconhecimento de marcações artificiais está nesta primeira categoria, ao lado do reconhecimento de marcações naturais, reconhecimento de modelos e o uso de sinalizadores ativos. Na segunda categoria pode-se citar as medidas de odometria e navegação inercial (giroscópios e acelerômetros). Comparado com o uso de odometria onde os erros de localização são cumulativos, a localização por marcações tem a vantagem de que os erros são limitados [4].

As marcações são caracteres facilmente distinguíveis que um robô possa reconhecer em sua entrada sensorial, podendo ser formas geométricas (como retângulos, linhas ou círcu-

los), e podendo incluir informações adicionais (como cores ou códigos de barras). As marcações devem ser cuidadosamente escolhidas para que sejam fáceis de identificar mesmo em condições ambientais adversas [4].

Em geral as marcações possuem uma posição fixa e conhecida no ambiente de trabalho, em relação à qual o robô pode se localizar. Há porém casos de marcações móveis (ou semipermanentes ou ainda dinâmicas) que não servem para a localização do robô no seu ambiente mas que permite a determinação da localização relativa entre o robô e os outros objetos.

Algumas características da navegação baseada em marcações apresentadas por Borenstein *et al.* [4] podem ser resumidas como:

- Marcações precisam estar disponíveis no ambiente de trabalho em torno do robô;
- A distância máxima entre o robô e a marcação é substancialmente menor que em sistemas de sinalizadores ativos;
- A precisão de posicionamento depende da distância e do ângulo entre o robô e a marcação; a navegação é menos precisa quando o robô está mais distante da marcação; um alto grau de precisão é obtido quando o robô está próximo da marcação;
- Substancialmente mais processamento é necessário que no caso de sistemas de sinalizadores ativos;
- Condições ambientais, como iluminação, podem ser problemáticas; com baixa visibilidade as marcações podem não ser reconhecidas ou outros objetos do ambiente com caracteres similares podem ser erroneamente reconhecidos como uma marcação;
- O sistema de navegação por marcações pode ser utilizado para corrigir periodicamente os erros acumulados de odometria do robô.

Vários métodos da visão computacional e processamento de imagens podem ser aplicados na localização. A localização está especialmente relacionada com a área de reconhecimento de objetos, que envolve a identificação da classe de um objeto, e a estimativa de sua posição e orientação, que pode ser feita, dentre outras formas, por estereoscopia, o tema deste trabalho.

O método de localização por marcações artificiais mostra-se adequado ao caso RoboCup e ao uso de estereoscopia, embora existam aí limitações na cobertura do sensor ótico e na

densidade de marcações no ambiente de trabalho do robô. Vários objetos e marcações do campo de jogo podem ser utilizados pelo mecanismo de visão, incluindo:

- Linha penal esquerda e direita ;
- Campo de jogo;
- Linha de fundo esquerda e direita;
- Linha lateral superior e inferior;
- Linha mediana;
- Ponto central;
- Meta esquerda e direita;
- Bandeira superior e inferior da linha penal esquerda e direita;
- Bola de jogo;
- Marcação de jogador do time da esquerda e da direita.

Os principais componentes do processo de localização utilizando visão computacional são:

- Representação do ambiente;
- Modelos dos sensores; e
- Algoritmos de localização.

Destes três componentes, o que é diretamente abordado neste trabalho, e que será tratado extensamente no capítulo 3, é a modelagem dos sensores, neste caso um conjunto de visão binocular que utiliza-se de estereoscopia para fornecer a localização. Observe-se todavia que a visão computacional não está estritamente vinculada à localização mas, ao mesmo tempo fornece outras informações úteis às demais tarefas do robô como é o caso do reconhecimento dos objetos.

## 1.4 Motivação

Em estereoscopia é necessário conhecer a geometria do conjunto binocular antes que se possa extrair quaisquer informações sobre a geometria no ambiente. Ou as câmeras e o conjunto binocular como um todo são montados precisamente dentro de especificações prévias ou a geometria precisa ser determinada após a montagem.

Este primeiro caso não é possível na maioria das aplicações práticas, e mesmo que o fosse, não se pode garantir facilmente que o sistema será mantido na sua configuração original quando sujeito a aquecimento das lentes e do sensor, impacto, vibrações e outros fatores inerentes à robótica móvel. Daí a importância da calibração, que trata o segundo caso, onde a geometria é determinada após a montagem.

Ainda, como esta calibração após a montagem não se utiliza de um modelo exato das câmeras mas sim de uma aproximação, é importante selecionar o método de calibração mais adequado. O método clássico de calibração trata de somente uma câmera de cada vez. Métodos alternativos têm sido investigados que permitem calibrar as duas câmeras em conjunto, como é o caso da calibração através da geometria epipolar que estará sendo aplicada neste trabalho. Esta estratégia de calibração tem importância pelo aumento da precisão da informação de calibração.

## 1.5 Objetivos do trabalho

Este trabalho tem como objetivo a determinação da geometria de um conjunto binocular de forma que se possa obter a localização de *caracteres primitivos* ou objetos detectados no ambiente em um sistema de coordenadas acoplado ao conjunto binocular. Como caracteres primitivos (do grego, *charaktér*) deve-se entender quaisquer marcas especiais que distingam as coisas entre si, o que no caso de visão computacional pode ser uma região com uma cor ou textura especial, uma forma geométrica, como linhas, vértices, curvas espiraladas, dentre outras.

Este trabalho não trata do reconhecimento de objetos, mas somente da determinação da localização dos caracteres ou objetos já conhecidas suas posições correspondentes nas duas imagens do sistema binocular.

Especificamente este trabalho trata do estéreo geométrico, um subprocesso da visão de baixo nível, inserido num sistema de visão passiva reconstrutivista, com calibração no momento estático.

Os objetivos específicos deste trabalho são:

- Detectar caracteres primitivos nas imagens fornecidas pelo conjunto binocular;
- Identificar caracteres primitivos presentes nas duas imagens e que são a projeção de um mesmo caractere do gabarito de calibração;

- Identificar caracteres primitivos presentes nas duas imagens e que são a projeção de um mesmo objeto no ambiente;
- Utilizar os caracteres tanto do ambiente quanto do gabarito para determinar a geometria do conjunto binocular;
- Utilizar a geometria do conjunto binocular para determinar a posição de um caractere no ambiente conhecida sua posição nas duas imagens.

## 1.6 Estrutura do trabalho

Este trabalho compõe-se de cinco capítulos, cujo conteúdo será descrito a seguir.

Neste primeiro capítulo procurou-se definir o escopo deste trabalho e posicioná-lo dentro do contexto onde está inserido. Foi exposta a necessidade de se estudar os métodos de calibração de câmeras e os objetivos específicos deste estudo realizado.

O segundo capítulo trata da detecção dos caracteres primitivos das imagens, especialmente vértices e regiões homogêneas, e da identificação de pares de caracteres correspondentes entre as duas imagens. São discutidas também as propriedades desejáveis dos caracteres primitivos para que a correspondência ocorra com maior confiabilidade.

No terceiro capítulo é apresentada inicialmente a formação de imagens pela projeção em perspectiva e o modelo de câmera que estará sendo utilizado. Em seguida discute-se a configuração do conjunto binocular, destacando a geometria epipolar. Na parte final descreve-se métodos clássicos de calibração de câmeras e o método da geometria epipolar.

O quarto capítulo inicia com a descrição da configuração do conjunto binocular simplificado que foi utilizado nos experimentos realizados. Em seguida são mostrados os resultados da correspondência de caracteres detectados em vários pares de imagens e é analisada a confiabilidade dos métodos aplicados. Em seguida são mostrados os dados experimentais da correção da distorção das lentes, da calibração por um método clássico e da calibração pela geometria epipolar.

Finalmente, o quinto capítulo apresenta as principais conclusões que puderam ser extraídas dos experimentos realizados, apontando as contribuições do método na redução do erro de calibração e as limitações observadas. Na parte final são feitas recomendações ou sugestões para novos estudos na área ou mesmo para a continuidade deste.

Cada capítulo inclui uma seção de revisão dos tópicos específicos discutidos no capítulo, a qual pode ser desprezada pelos leitores familiarizados com o tema.

## Capítulo 2

# Detecção e correspondência de caracteres

*“Automatic’ simply means that you can’t repair it yourself”*

*Mary H. Waldrip*

Introduzimos aqui alguns conceitos relacionados com a extração de caracteres primitivos e a correspondência destes caracteres entre duas imagens, conceitos estes que são relevantes para o processo de calibração de câmeras que será descrito no próximo capítulo.

Observe-se todavia que estes caracteres detectados não são estruturas de alto nível que possam ser utilizadas no processo de localização por marcações. Para tal, outros tipos de caracteres primitivos além dos que estão sendo utilizados na calibração também precisam ser extraídos das imagens, como bordas e regiões. Além disso processos de segmentação e rotulagem precisam ser aplicados sobre estes caracteres primitivos para identificar as estruturas de mais alto nível que caracterizam as marcações (citadas na seção 1.1) que finalmente seriam utilizadas na localização do robô (conforme descrito na seção 1.3).

### 2.1 Detecção de caracteres primitivos

Geralmente é desejável identificar e localizar certos caracteres em uma imagem. Numa situação de manufatura, por exemplo, marcas como furos ou ranhuras podem ser utilizadas para diferenciar componentes, ou ainda, durante a inspeção, pode-se identificar caracteres

associados a defeitos. No processo de calibração de câmeras estes caracteres são utilizados para estabelecer as relações geométricas entre a câmera e o ambiente ou entre duas ou mais câmeras.

O processo de identificar e isolar caracteres primitivos que representam a informação presente na imagem é geralmente chamado de *segmentação* [8, 12].

Existem duas diferentes abordagens para a extração de caracteres, uma que trabalha sobre a imagem original e outra que trabalha sobre as representações de bordas já detectadas [24].

Vértices, segmentos de linhas, junções e regiões podem ser detectados a partir das bordas e serem utilizados como caracteres primitivos. Vértices são geralmente definidos como pontos de alta curvatura ao longo de bordas. Segmentos de linha podem ser determinados analisando-se os píxeis de borda para determinar se grupos deles são colineares. Após extraídos os segmentos de linhas de uma imagem de bordas, pode-se identificar junções entre os segmentos. Junções são definidas como pontos onde dois ou mais segmentos de linha se encontram e podem representar vértices num objeto ou pontos onde objetos de sobrepõem [24].

Huynh [15] utiliza esta abordagem de detecção a partir de bordas num sistema de visão estereoscópica.

Um método simples de detecção de caracteres que utiliza as imagens originais é a correspondência de padrões. Nesta técnica a área da imagem é percorrida comparando-se com um padrão conhecido ao mesmo tempo que um grau de correspondência é calculado. Se a correspondência é suficiente então o padrão foi identificado. Este método funciona bem em situações controladas, onde a posição e tamanho dos caracteres na imagem são fixos ou conhecidos, mas não são suficientes em ambientes não estruturados ou onde ocorrem distorções de perspectiva [24].

Em estereoscopia pares de vértices correspondentes são tidos como mais eficientes no cálculo da posição tridimensional que segmentos de linhas, especialmente quando as posições dos pontos terminais dos segmentos de linhas não são conhecidas com precisão [15].

Para o procedimento de calibração realizado aqui estaremos interessados em três tipos de caracteres que serão extraídos diretamente da imagem e que serão tratados a seguir.

### 2.1.1 Detecção de vértices

Vértices são caracteres amplamente utilizados em visão computacional por serem bastante estáveis sob pequenas variações das condições de iluminação e orientação das câmeras [15].

Segundo Huynh [15], talvez o primeiro operador para detecção de vértices tenha sido um operador rotacionalmente invariante chamado DET, que é derivado da expansão de Taylor da intensidade da superfície  $I(x, y)$ :

$$DET = I_{xx}I_{yy} - I_{xy}^2 \quad (2.1)$$

onde  $I_{xx}, I_{yy}$  e  $I_{xy}$  são das derivadas parciais de segunda ordem de  $I(x, y)$ .

O operador utilizado neste trabalho é a função de resposta de vértice de Harris. Köthe [18] observa que este operador é bastante robusto embora desvie os vértices no sentido de uma das regiões, dependendo da escala.

Dada uma escala  $\sigma$ , o operador de Harris calcula a matriz suavizada do produto dos gradientes, em cada píxel, da seguinte forma:

$$\begin{pmatrix} G^\sigma * (g_x^\sigma g_x^\sigma) & G^\sigma * (g_x^\sigma g_y^\sigma) \\ G^\sigma * (g_y^\sigma g_x^\sigma) & G^\sigma * (g_y^\sigma g_y^\sigma) \end{pmatrix} = \begin{pmatrix} A & C \\ C & B \end{pmatrix} \quad (2.2)$$

onde  $G^\sigma *$  indica a convolução com um filtro Gaussiano de escala  $\sigma$ ,  $g_x^\sigma$  e  $g_y^\sigma$  são os filtros derivadas de primeira ordem do Gaussiano nas direções  $x$  e  $y$  na escala  $\sigma$ , e o produto indica multiplicação píxel e píxel. Então a função de resposta de vértice é calculada para cada píxel como:

$$V = AB - C^2 - 0,04(A + B)^2 \quad (2.3)$$

O máximo local da função de resposta de vértice indica os vértices. Além de aplicar o máximo local utilizamos um limiar de corte para selecionar somente os vértices cujas respostas forem maiores que o limiar.

### 2.1.2 Detecção de blocos distingúveis

Tomasi e Shi [31] definem um critério de seleção de caracteres distingúveis que está re-

acionado ao mecanismo de rastreamento de caracteres. Este mesmo critério é utilizado por Birchfield [3] para determinar as descontinuidades de profundidade em imagens obtidas por um conjunto binocular ou seqüencialmente durante um deslocamento de câmara.

A dedução do critério de seleção está intimamente relacionada com o mecanismo de rastreamento, que se concentra na determinação de um vetor deslocamento  $d$  que minimiza o indicador de *dissimilaridade* a seguir entre dois blocos, um na primeira imagem  $I$  e outro na segunda imagem  $J$ :

$$\epsilon = \int \int_W \left[ J\left(\mathbf{x} + \frac{\mathbf{d}}{2}\right) - I\left(\mathbf{x} - \frac{\mathbf{d}}{2}\right) \right]^2 w(\mathbf{x}) d(\mathbf{x}) \quad (2.4)$$

onde  $\mathbf{x} = [x, y]^T$ ,  $W$  é bloco (ou janela) da imagem, e  $w(\mathbf{x})$  é uma função de ponderação, geralmente  $w(\mathbf{x}) = 1$ . Se a  $w(\mathbf{x})$  for uma função tipo Gauss, por exemplo, pode-se enfatizar a região central da bloco.

Truncando a expansão de (2.4) por uma série de Taylor obtém-se a linearização do sistema, que após determinada a derivada, igualada a zero e aplicada uma iteração na minimização pelo método de Newton-Raphson, leva ao sistema

$$Zd = e \quad (2.5)$$

onde  $Z$  é a matriz  $2 \times 2$

$$Z = \int \int_W g(\mathbf{x}) g^T(\mathbf{x}) w(\mathbf{x}) d(\mathbf{x}) \quad (2.6)$$

e  $e$  é o vetor  $2 \times 1$  de erro

$$e = 2 \int \int_W \left[ I\left(\mathbf{x} + \frac{\mathbf{d}}{2}\right) - J\left(\mathbf{x} - \frac{\mathbf{d}}{2}\right) \right] g(\mathbf{x}) w(\mathbf{x}) d(\mathbf{x}) \quad (2.7)$$

onde

$$g = \begin{bmatrix} \frac{\partial}{\partial x}(I + J) \\ \frac{\partial}{\partial y}(I + J) \end{bmatrix} \quad (2.8)$$

e

$$J\left(\mathbf{x} + \frac{\mathbf{d}}{2}\right) \approx J(\mathbf{x}) + \frac{d_x}{2} \frac{\partial J}{\partial x}(\mathbf{x}) + \frac{d_y}{2} \frac{\partial J}{\partial y}(\mathbf{x}) \quad (2.9)$$

$$I\left(\mathbf{x} - \frac{\mathbf{d}}{2}\right) \approx I(\mathbf{x}) + \frac{d_x}{2} \frac{\partial I}{\partial x}(\mathbf{x}) - \frac{d_y}{2} \frac{\partial I}{\partial y}(\mathbf{x}) \quad (2.10)$$

A solução para o problema de rastreamento consiste em percorrer a imagem solucionando a equação (2.5) e deslocando  $I$  e  $J$  pelo deslocamento calculado.

Finalmente Tomasi e Shi [31] definem os caracteres distinguíveis como os blocos da imagem onde o sistema (2.5) pode ser facilmente resolvido e representa boas medições. Conseqüentemente pode-se tomar como parâmetro os autovalores da matriz simétrica  $Z$ ; dois autovalores baixos indicam intensidade constante ao longo do bloco, um autovalor alto e um baixo indica um padrão de textura unidirecional, dois autovalores altos indicam vértices ou qualquer outro padrão que pode ser rastreado facilmente.

Como as variações de intensidade no bloco são limitadas pelo valor máximo de intensidade do píxel, o maior autovalor não será arbitrariamente alto. Em conseqüência, se os dois autovalores de  $Z$  são  $\lambda_1$  e  $\lambda_2$ , o bloco é aceito como um caractere distinguível se

$$\min(\lambda_1, \lambda_2) > \lambda \quad (2.11)$$

onde  $\lambda$  é um limiar pré-definido.

Neste trabalho utilizamos a seguinte formulação: dada uma escala  $\sigma$ , calculamos a matriz do produto dos gradientes, em cada píxel, da seguinte forma:

$$\begin{pmatrix} g_x^\sigma g_x^\sigma & g_x^\sigma g_y^\sigma \\ g_y^\sigma g_x^\sigma & g_y^\sigma g_y^\sigma \end{pmatrix} = \begin{pmatrix} A & C \\ C & B \end{pmatrix} \quad (2.12)$$

onde  $g_x^\sigma$  e  $g_y^\sigma$  são os filtros derivadas de primeira ordem do Gaussiano nas direções  $x$  e  $y$  na escala  $\sigma$ , e produto indica multiplicação píxel a píxel. Então a função de resposta de bloco distinguível é calculada para cada píxel como:

$$K = \frac{A + B - \sqrt{(A - B)^2 + 4C^2}}{2} \quad (2.13)$$

### 2.1.3 Expansão de regiões

Uma região em uma imagem pode ser caracterizada por uma variação significativa na distribuição do nível de intensidade em uma direção.

Embora a presença de regiões com intensidades distintas nas imagens seja quase sempre

óbvia para a visão humana, o mesmo não ocorre em visão de máquina. Para um sistema de visão computacional uma imagem é simplesmente uma coleção retangular de elementos de imagens. O fato de que um grupo destes elementos poderem ser similares em intensidade ou ainda especialmente próximos é algo que precisa ser explicitamente determinado [24].

Uma região é geralmente definida como um grupo de elementos de imagem conectados e também similares em intensidade. A forma mais básica de expansão de região inicia por um elemento de imagem dito *semente* e segue examinando todos os elementos de imagem imediatamente conectados ao elemento semente. Aqueles que são suficientemente similares ao elemento semente são incluídos na região. O processo de construção de regiões continua investigando os elementos de imagem conectados a todos os novos elementos incluídos na região, e este processo continua iterativamente até que nenhum outro elemento de imagem similar seja encontrado [24].

Quase sempre é desejável identificar as regiões presentes na imagem associando a cada região um número único. A forma mais simples de fazer isto é simplesmente expandir as regiões serialmente, uma de cada vez. Pode ser atribuído um número à região inicial e escolhido um elemento como semente, dando prosseguimento ao processo de expansão de região. Cada elemento adicionado à região recebe como número de região o valor corrente. Quando a construção da primeira região está concluída a imagem pode ser pesquisada por qualquer elemento que ainda não esteja numa região e este elemento pode ser utilizado como nova semente. O número da região é então incrementado e o processo de expansão de região continua a partir do novo elemento semente. Este procedimento é repetido até que todos os elementos da imagem pertencem a alguma região [24].

Como este processo serial de identificação de regiões é demorado, métodos mais rápidos foram desenvolvidos. Um destes métodos é chamado de expansão simultânea de regiões. Neste método vários elementos semente são escolhidos e várias regiões são expandidas simultaneamente. Um passo adicional é incluído no processo de expansão de regiões para verificar se nenhum elemento de imagem similar já possui um número de região. Neste caso é estabelecida a equivalência entre estes dois números de região de forma que duas ou mais regiões possam ser agrupadas em uma única grande região tão logo o processo de expansão seja concluído. Os elementos semente podem ser escolhidos aleatoriamente ou distribuídos espacialmente equidistantes na imagem. A expansão simultânea de regiões é especialmente interessante em ambiente de processamento paralelo onde sua aplicação possui grande

impacto na redução do tempo de processamento [24].

Diversos outros métodos de segmentação de regiões, específicos para imagens coloridas, são descritos por Skarbek e Koschan [28].

Diversas medidas estatísticas podem ser tomadas sobre uma região com base nos valores dos píxeis e utilizadas na correspondência. Algumas das medidas utilizadas [24, 19, 26] são:

- Número de segmentos de linhas
- Número de vértices
- Número de curvas
- Simetria
- Área em píxeis ( $A$ )
- Perímetro em píxeis ( $P$ )
- Compactação ( $C$ )
- Coordenadas mínimas da região ( $x_{min}, y_{min}$ )
- Coordenadas máximas da região ( $x_{max}, y_{max}$ )
- Razão perímetro por área ( $R$ )
- Textura
- Dispersão ( $D$ )
- Centróide ( $\bar{x}, \bar{y}$ )
- Momento central de inércia ( $m_{ij}$ )
- Ângulo do eixo principal de inércia ( $\theta$ )
- Desvio do centróide em relação ao menor retângulo continente
- Dimensão máxima ( $L_{max}$ )
- Dimensão mínima ( $L_{min}$ )
- Excentricidade ( $E$ )
- Conectividade
- Número de Euler
- Média da intensidade das cores primárias ( $\bar{r}, \bar{g}, \bar{b}$ )
- Desvio padrão da intensidade das cores primárias ( $\sigma_r, \sigma_g, \sigma_b$ )

Algumas destas medidas possuem significado óbvio, outras precisam ser definidas:

$$\begin{aligned}
 C &= \frac{P^2}{A} & D &= 1 - \frac{2\sqrt{\pi A}}{P} & E &= \frac{L_{max}}{L_{min}} \\
 \bar{x} &= \frac{1}{A} \sum_x x & \bar{y} &= \frac{1}{A} \sum_y y & & \\
 m_{ij} &= \sum (x - \bar{x})^i (y - \bar{y})^j & & & & \\
 \theta &= \frac{1}{2} \tan^{-1} \left[ \frac{2m_{11}}{m_{20} - m_{02}} \right] & & & & 
 \end{aligned} \tag{2.14}$$

$$\tag{2.15}$$

Observe-se que a medida de dispersão é independente de tamanho, e seu valor é zero para um círculo e aproxima-se de um para contornos irregulares [24].

A partir do momento central pode-se definir diversos outros momentos que são invariantes em relação à posição, escala e orientação do objeto representado na região, porém estes momentos não são invariantes sob projeção em perspectiva. A excentricidade  $E$ , algumas vezes chamada de alongação, também pode ser definida a partir dos momentos [19].

Para uma região, o número de Euler é definido como um menos o número de lacunas na região. O número de Euler para um conjunto de regiões é o número de regiões menos o número de lacunas.

Definimos aqui uma métrica adicional que chamaremos de quadratura. Como estaremos interessados em regiões que são a projeção em perspectiva de quadrados do gabarito de calibração, utilizaremos como métrica a razão entre a área da região e a área do menor retângulo que contém a região, ambos calculados em píxeis.

$$Q = \frac{|(x_{max} - x_{min}) * (y_{max} - y_{min}) - A|}{A} \tag{2.16}$$

Observe-se que a medida de quadratura é zero quando a região for exatamente um quadrado com lados paralelos aos limites do plano de imagem.

## 2.2 Correspondência de caracteres

Na correspondência de caracteres (do inglês *feature matching*), caracteres como regiões, elementos de textura, bordas e vértices, obtidos em um estágio de pré-processamento das imagens, são as primitivas empregadas para a determinação dos pontos correspondentes entre imagens obtidas através de duas câmeras, com base no fato de que uma dada posição

no espaço da cena deve originar caracteres semelhantes em cada uma das imagens [32].

No próximo capítulo a estereoscopia será definida de forma detalhada, por ora precisamos definir somente três conceitos, disparidade, profundidade e linha epipolar (ver figura 2.1), pela importância destes conceitos no processo de correspondência.

Em estereoscopia duas câmeras são utilizadas na captura de um par de imagens do ambiente a partir de pontos de vista diferentes. Caracteres presentes nas duas imagens associados a um mesmo objeto no ambiente apresentam um deslocamento relativo, chamado de *disparidade*, que é tanto maior quanto mais próximo o objeto estiver das câmeras. A informação de disparidade pode ser utilizada para estimar a *profundidade* do objeto, que é a distância do objeto até o conjunto de câmeras.

No caso de câmeras onde os planos de imagem são coplanares e os eixos óticos estão alinhados, como no exemplo da figura 2.1, as distâncias focais  $f$  e tamanhos dos píxeis  $dx$  são iguais e conhecidos, e também conhecida a distância  $b$  entre os centros de projeção  $C_1$  e  $C_2$ , a relação entre a profundidade  $r$  e a disparidade  $d$  é dada por

$$r = \frac{b \times f}{d \times dx}. \quad (2.17)$$

Observe-se que esta é uma configuração bastante simplificada e que é bastante difícil de se obter em câmeras comuns. Um modelo mais completo, onde as câmeras assumem orientações arbitrárias será visto no próximo capítulo.

Considere um objeto pontual no ambiente e que está projetado em ambas as imagens das duas câmeras. Podemos definir o *plano epipolar* como sendo o plano que passa pelo ponto no ambiente e pelos dois centros de projeção das duas câmeras e, por conseqüência, também passa pelos dois pontos no plano de imagem onde o objeto está projetado (ver figura 2.1). Este plano corta os dois planos de imagem nas duas câmeras. A linha descrita onde o plano epipolar corta o plano de imagem é chamada de *linha epipolar* e as linhas epipolares das duas câmeras estão, geometricamente, relacionadas.

Dado o par de imagens, os caracteres ou objetos da imagem da esquerda precisam ser encontrados na imagem da direita antes que se possa fazer qualquer cálculo de posição. Embora a aparência dos objetos geralmente se mantenha aproximada em ambas as imagens, alguma distorção ocorrerá devido à diferença de ponto de vista. Além disso, devido à diferença de campo de visão e à sobreposição de objetos mais próximos por objetos mais distantes, alguns pontos da imagem da esquerda não aparecerão na imagem da direita e

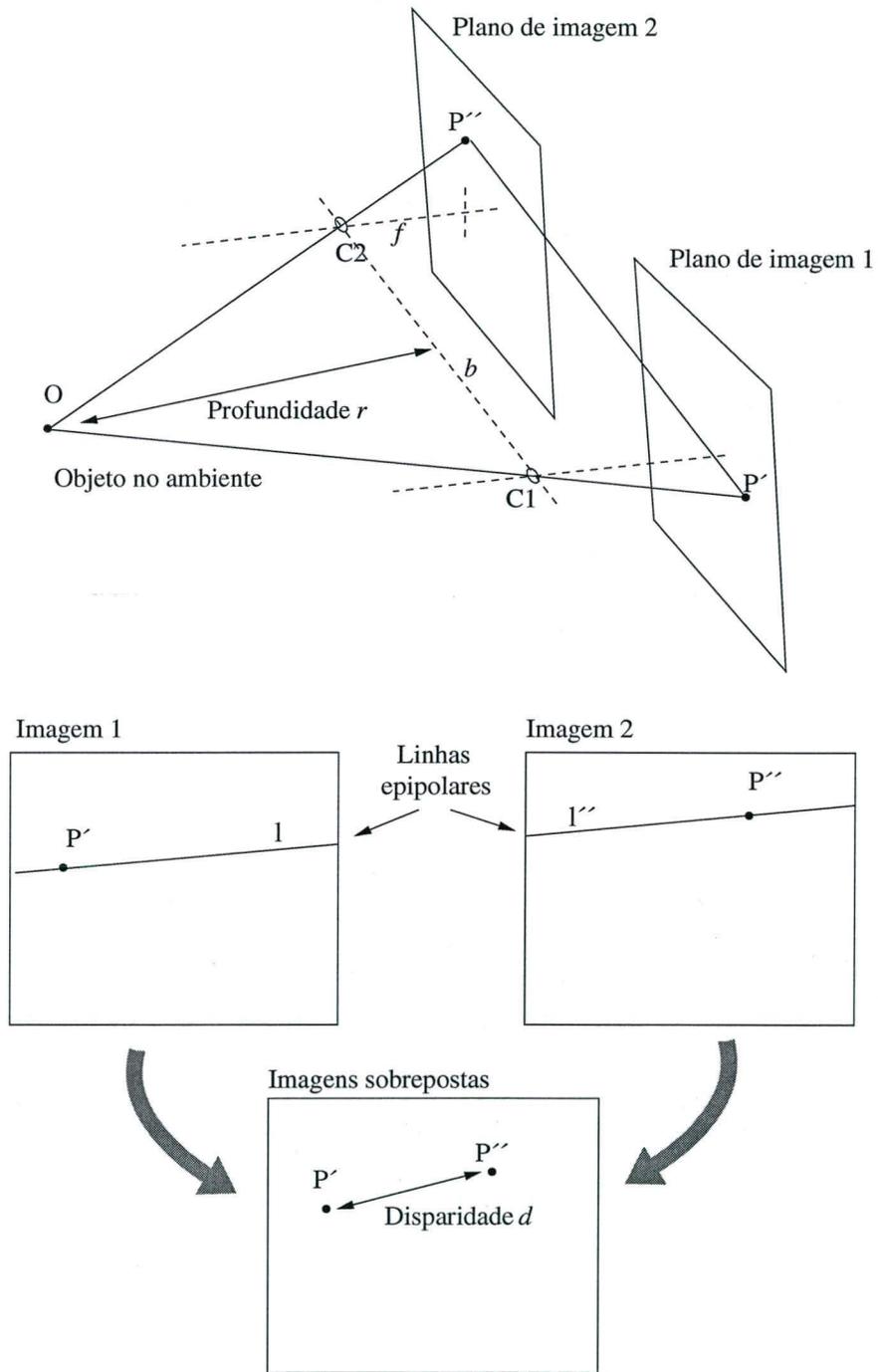


Figura 2.1: Disparidade, profundidade e linha epipolar

vice versa. Ainda, por definição, todo objeto em uma imagem estará em uma posição diferente na outra imagem, e objetos de diferentes profundidades mudarão de posição uns em relação aos outros. Todos estes fatos tornam difícil a identificação dos caracteres ou objetos correspondentes nas duas imagens [24, 32].

O processo de correspondência precisa manter sua robustez de forma a eliminar, tanto quanto possível, o número de correspondências inválidas e, conseqüentemente, o erro no cálculo da posição tridimensional dos objetos. Além disso os tipos de caracteres a serem utilizados também são importantes no processo de correspondência. Caracteres tais como segmentos de linhas, vértices e regiões poligonais são bastante estáveis diante do deslocamento relativo entre as câmeras e da variação das condições de iluminação [15].

Basicamente duas abordagens podem ser empregadas para resolver o problema da correspondência em visão estereoscópica:

- Baseadas em intensidade, onde um processo de correlação de intensidade é aplicado diretamente à informação de intensidade das imagens; e
- Baseadas em caracteres, onde o processo de correspondência é aplicado sobre caracteres detectados *a priori*.

Huynh [15] observa que o tipo de imagem disponível e a saída esperada de profundidade ou disparidade ocupam papéis importantes na decisão entre estas abordagens. Imagens com texturas complexas e pequena variação de profundidade entre os objetos no ambiente são boas candidatas ao uso da abordagem de intensidade uma vez que variações suaves de disparidade é um dos requisitos para que esta técnica seja eficiente.

Uma das técnicas utilizadas neste caso é o tratamento de cada uma das linhas epipolares em cada imagem como sendo uma forma de onda e a aplicação de um processo de correspondência entre os picos e vales nas formas de onda das linhas epipolares correspondentes em cada uma das imagens [15]. Observe-se todavia que esta técnica somente pode ser utilizada quando a geometria do conjunto binocular já é conhecida.

Uma técnica que alguns autores ([15]) classificam como correspondência baseada em intensidade é a correspondência de área, onde o processo de correspondência é também aplicado sobre a intensidade da imagem, mas somente onde blocos (ou janelas) que podem ser facilmente distinguíveis estão presentes. As regiões ou blocos de píxeis distinguíveis podem ser selecionados por vários métodos, podendo ser diretamente uma função da intensidade no bloco, como é o caso da equação de Kanade-Lucas-Tomasi [2, 3], ou uma descrição

geométrica de seu conteúdo, como é o caso dos vértices e circulares simétricos [13]. A determinação da forma e do tamanho ideal do bloco de correlação precisa ser considerada nesta abordagem.

Neste trabalho, estes blocos são tratados como caracteres e portanto os classificamos como pertencendo à segunda abordagem.

Ambas as abordagens tornam-se inadequadas quando as superfícies na cena apresentam estruturas periódicas de pequena escala ou algum outro tipo de redundância espacial que possa resultar em correspondências ambíguas. Num outro extremo, quando as imagens não apresentam uma estrutura local suficiente que permita a detecção de atributos adequados à correspondência, estas técnicas também não são indicadas.

Ambas as abordagens, baseadas em intensidade e em caracteres estão em uso atualmente e podem ser combinadas. O resultado da correspondência de caracteres pode ser utilizado na correspondência por intensidade [13].

Uma desvantagem da abordagem baseada em caracteres é o fato de sempre requerer alguma modalidade de interpolação, quando se deseja um mapa denso de disparidades, já que os caracteres têm geralmente uma distribuição esparsa nas imagens. Uma vantagem é que os blocos ou caracteres distinguíveis são em menor quantidade que o número total de píxeis na imagem, estas técnicas de correspondência são computacionalmente mais econômicas que a correlação de intensidade [32, 15].

Neste trabalho se dá mais ênfase à correspondência de caracteres.

Algumas propriedades são desejáveis nos caracteres que serão utilizados na correspondência [13]:

- Distinção: pontos devem ser distinguíveis em relação aos seus vizinhos; esta definição exclui pontos que pertençam a uma mesma borda detectada na imagem; caracteres distinguíveis podem ser vértices, regiões ou locais de alta rugosidade e regularidade; uma forma de medir a distinção é comparar os atributos do caracteres com sua vizinhança, se o caractere for o máximo local então conclui-se que ele é distinguível;
- Invariância: a seleção dos caracteres e posição do caractere selecionado deve ser invariante em relação às distorções geométricas e radiométricas esperadas;
- Estabilidade: a seleção dos caracteres e posição do caractere selecionado deve ser invariante em relação à diferença de orientação e deslocamento das câmeras, excluindo caracteres irrealis; esta propriedade garante que caracteres na imagem estejam relaci-

onados a objetos no ambiente; por exemplo, vértices são considerados estáveis por estarem associados aos objetos, enquanto junções em "T" são geralmente consideradas instáveis pois quase sempre resultam de oclusões e não dos objetos em si;

- Unicidade: Enquanto a distinção garante a separabilidade local, a unicidade procura a separabilidade global ao longo de toda a imagem; esta propriedade evita caracteres distinguíveis localmente mas que se repetem ao longo da imagem e que confundem ou atrasam o processo de correspondência;
- Interpretabilidade: além de serem úteis na correspondência e análise das imagens os caracteres devem ter algum significado que permite a interpretação das imagens; exemplos destes caracteres podem ser vértices, junções de linhas, centros de figuras circulares simétricas.

Segundo Haralick e Shapiro [13], um par arbitrário de caracteres ( $P'$ ,  $P''$ ) pode assumir dois estados: ou são correspondentes ou não. A solução da correspondência pode ser obtida num processo de duas etapas:

1. Selecionar caracteres apropriados nas duas imagens, gerando uma lista de caracteres para cada imagem;
2. Encontrar caracteres ou pares de pontos ( $P'$ ,  $P''$ ) correspondentes, com  $P'$  na lista de caracteres da primeira imagem e  $P''$  na lista da segunda, que satisfaçam aos critérios de similaridade e consistência; tanto similaridade quanto consistência são geralmente utilizados nos algoritmos que buscam a solução ótima, o que implica uma ponderação entre estes critérios.

Praticamente todos os métodos de correspondência seguem estes passos, mas utilizando tipos de caracteres, medidas de similaridade e consistência e algoritmos de busca diferentes. Em todos os casos os atributos dos caracteres são derivados da vizinhança dos caracteres [13].

Para garantir que o processo de correspondência funcione de forma confiável e robusta diversas restrições de correspondência podem ser utilizadas, dependendo se o conjunto binocular está ou não calibrado.

Algumas restrições, como as relacionadas ao acréscimo de informações provenientes de caracteres ou estruturas de mais alto nível detectadas nas imagens, como segmentos de linhas ou objetos, não serão tratadas neste trabalho pois não acrescentam informação útil durante o processo de calibração inicial do conjunto binocular, quando toda a geometria do

conjunto binocular é desconhecida e a informação de profundidade está ausente. Também a rotulagem de objetos pode estar ausente no momento da calibração das câmeras.

Uma outra restrição, de ordenamento, estabelece que pares de mesma profundidade deverão ocorrer na mesma ordem, por exemplo da esquerda para a direita, na primeira e na segunda imagem. Em poucos ambientes existe a ocorrência de pares de caracteres verdadeiramente correspondentes e que violam esta condição, o que a torna bastante útil no processo de correspondência. Todavia esta restrição também exige que as câmeras estejam alinhadas ou que as imagens sejam retificadas, diminuindo sua importância durante o processo de calibração.

Embora inicialmente não sejam importantes para o processo de calibração, estas restrições tornam-se úteis após a calibração das câmeras, durante o processo de cálculo das disparidades ou posição tridimensional dos objetos detectados, ou ainda durante o processo de atualização da calibração com o conjunto binocular já em operação.

As propriedades de coesão da matéria e a restrição de que qualquer ponto numa superfície pode ocupar somente uma posição no espaço em um dado momento levou ao desenvolvimento de duas restrições: *unicidade* e *vizinhança*. Outra restrição é a restrição de *similaridade* que determina que caracteres correspondentes possuem atributos similares [15].

Estas restrições, que se mostram interessantes no processo de calibração inicial do conjunto binocular, e mais a restrição da geometria epipolar serão discutidas a seguir.

### Unicidade

A restrição de unicidade indica que, quase sempre, um píxel ou caractere em uma das imagens corresponde a não mais que um píxel ou caractere na outra imagem.

Esta restrição pode não ser aplicável a segmentos de linhas e também falha quando objetos transparentes estão presentes no ambiente. Um exemplo de falha desta restrição é dado na figura 2.2, onde três pontos colineares no ambiente são projetados individualmente em uma das câmeras e como um só na outra.

Dado um caractere em uma das imagens seu caractere correspondente pode não estar presente na outra imagem por estar oculto ou fora dos limites da imagem. Neste caso nenhuma correspondência deve ser atribuída ao caractere.

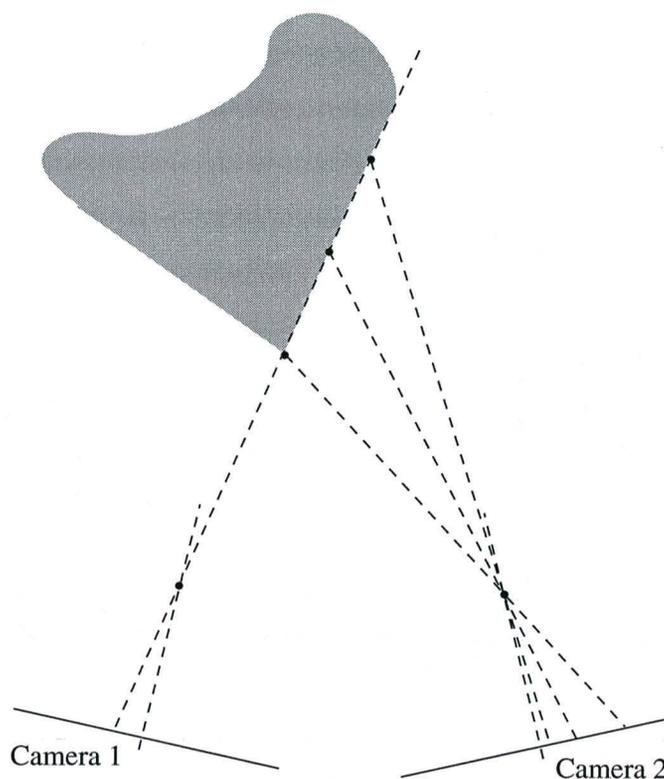


Figura 2.2: Falha da restrição de unicidade (adaptada de [29])

### Linha epipolar

A restrição epipolar é uma das mais fortes no processo de correspondência, embora exija o conhecimento da geometria do conjunto binocular.

Voltando à figura 2.1, dada a posição  $P'$  de um caractere numa das imagens e as coordenadas dos centros de projeção,  $C_1$  e  $C_2$ , do par de câmeras, o plano epipolar, que passa através destes três pontos, intersecta ambas as imagens em duas linhas,  $l'$  e  $l''$ , chamadas de linhas epipolares. Se o ponto  $P'$  pertence à linha  $l'$  então a posição  $P''$  do caractere correspondente, se existir na outra imagem, pertence à linha  $l''$ .

Sabendo-se a relação entre as linhas epipolares pode-se reduzir imensamente o espaço de busca por caracteres correspondentes.

Para um par de câmeras cujos eixos óticos são paralelos e os planos de imagem são coplanares e alinhados o uso da restrição epipolar é simplificada pela simples coincidência das linhas nas duas imagens. Todavia esta configuração é difícil de ser obtida e mantida em câmeras reais.

Uma simplificação grosseira desta restrição, e que só é aplicável se as câmeras forem

aproximadamente paralelas, é garantir que a distância vertical entre dois caracteres não seja superior a um limite especificado.

### Vizinhança

Quando se conhece aproximadamente a orientação relativa das câmeras e a faixa de profundidades do ambiente pode-se limitar a disparidade entre os caracteres correspondentes.

Novamente voltando à figura 2.1, no caso simplificado de um conjunto binocular onde os planos de imagem são coplanares, as distâncias focais  $f$  são iguais e os eixos óticos estão alinhados, conhecida a distância  $b$  entre os centros de projeção e a largura  $d_x$  dos píxeis das câmeras, e conhecida a faixa de profundidade dos objetos no ambiente, podemos determinar a disparidade mínima e máxima entre caracteres correspondentes na imagem utilizando a equação (2.17).

Outra restrição geralmente aplicada estabelece que a disparidade deve variar suavemente ao longo que quase toda a imagem, com descontinuidades ocorrendo onde os objetos se sobrepõem. Esta restrição falha em descontinuidades de profundidade uma vez que estas descontinuidades causam variação abrupta na disparidade.

Price *et al.* [23] introduz o conceito de gradiente de disparidade através do diagrama da figura 2.3 abaixo, onde vemos duas imagens e dois pontos do ambiente correspondentes nas duas imagens. A disparidade do par de caracteres  $P_1$  e  $P'_1$ , quando sobrepostos no mesmo plano de imagem, é  $d_1$ , e para o par  $P_2$  e  $P'_2$  é  $d_2$ .

A diferença de disparidade  $dd$ , para este caso de imagens retificadas ou de câmeras alinhadas e com linhas horizontais coincidentes nas duas imagens, é dado por ([23])

$$dd = |d_2 - d_1| \quad (2.18)$$

e a distância ciclopeana é

$$ds = \sqrt{\frac{(d_2 - d_1)^2}{4} + h^2} \quad (2.19)$$

Finalmente o gradiente de disparidade  $gd$  é dado por

$$gd = \frac{dd}{ds}. \quad (2.20)$$

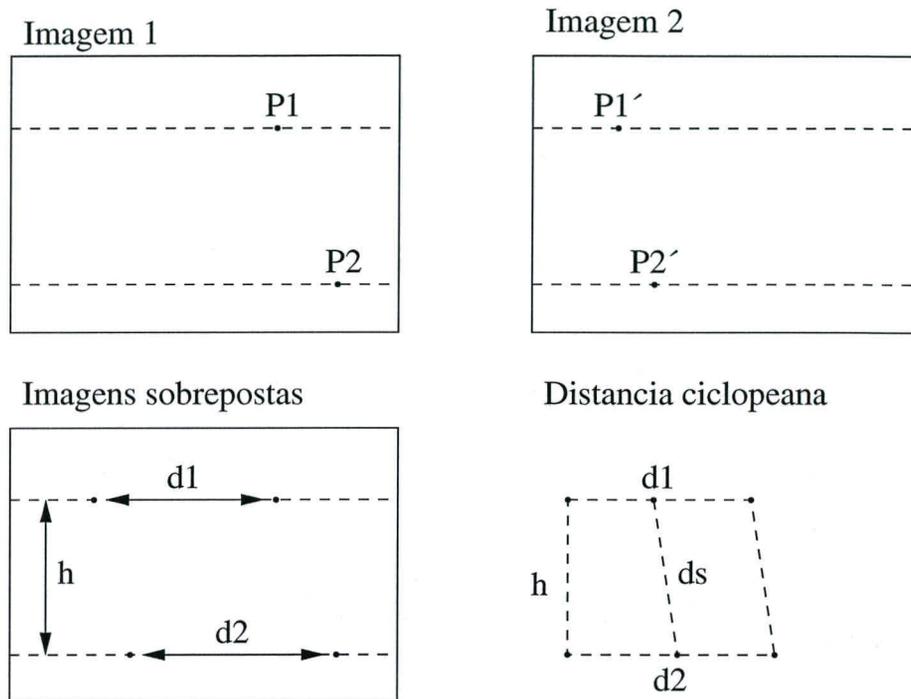


Figura 2.3: Distância ciclopeada (adaptada de [23])

No caso da visão humana, aumentando-se a disparidade enquanto se mantém a distância ciclopeana constante tende a não ocorrer a fusão estereoscópica e os dois pontos são vistos como pontos distintos (o limite em torno de 1). Também com disparidade constante e aumentando a distância ciclopeana a fusão tende a não ocorrer. Logo, a correspondência não é garantida simplesmente pela proximidade dos objetos [23].

Assim podemos restringir os pares correspondentes aplicando-se um limite 1 no gradiente de disparidade, onde 1 coincide com o valor relatado para o sistema de visão estereoscópica humana.

### Similaridade

A similaridade é baseada nos valores da função de intensidade da imagem na posição do caractere ou de atributos dos caracteres ou pontos calculados a partir da função intensidade na vizinhança da posição do caractere.

Esta restrição pode ser aplicada de várias formas: os caracteres devem possuir atributos similares, os píxeis na posição dos caracteres devem ter intensidades similares ou a vizinhança dos dois caracteres deve possuir uma alta correlação.

A medida mais básica da dissimilaridade entre píxeis pode ser definida como simplesmente a *diferença absoluta de intensidade* entre dois píxeis:

$$d = |I_e(x_e) - I_d(x_d)| \quad (2.21)$$

Birchfield [3] conclui que a medida de dissimilaridade (2.21) não é adequada para imagens digitais e sugere uma outra medida.

Dados  $x_e$  e  $x_d$  os píxeis onde se deseja calcular a dissimilaridade, defini-se  $\hat{I}_d$  como a função linearmente interpolada entre os pontos da imagem da direita. Então procura-se medir o quanto a intensidade em  $x_e$  se ajusta à região interpolada linearmente de  $x_d$ , ou seja, define-se a seguinte quantidade:

$$\bar{d}(x_e, x_d, I_e, I_d) = \max\{0, I_e(x_e) - I_{max}, I_{min} - I_e(x_e)\} \quad (2.22)$$

$$I_{max} = \max\{I_d^-, I_d^+, I_d(x_d)\} \quad (2.23)$$

$$I_{min} = \min\{I_d^-, I_d^+, I_d(x_d)\} \quad (2.24)$$

$$I_d^+ \equiv \hat{I}_d\left(x_d + \frac{1}{2}\right) = \frac{1}{2}(I_d(x_d) + I_d(x_d + 1)) \quad (2.25)$$

$$I_d^- \equiv \hat{I}_d\left(x_d - \frac{1}{2}\right) = \frac{1}{2}(I_d(x_d) + I_d(x_d - 1)) \quad (2.26)$$

### 2.3 Correspondência por relaxação

O procedimento de correspondência de caracteres por relaxação pode ser resumido como:

1. Selecionar pares de pontos candidatos a correspondentes;
2. Associar a cada par candidato um valor de probabilidade baseado na confiabilidade da correspondência;
3. Iterativamente atualizar esta probabilidade para cada par correspondente;

4. Finalmente eliminar pares onde o indicador de confiabilidade estiver abaixo de um limiar pré-definido.

Na seleção de pares candidatos todo tipo de conhecimento *a priori* e estratégias devem ser aplicados, o único requisito destas heurísticas e estratégias é que sejam conservadoras, ou seja, eliminem somente as correspondências realmente incorretas. Estas heurísticas e estratégias podem se basear nas restrições de correspondência já descritas na seção anterior. Deve-se observar ainda que os pares candidatos devem ter caracteres do mesmo tipo; não se pode tentar correspondência entre vértices e regiões, por exemplo. Em alguns casos um mesmo caractere em uma das imagens poderá ter mais de um caractere candidato a correspondente na outra imagem, logo poderá participar em mais de um par candidato.

Deve existir algum critério para distinguir os pares verdadeiros dos falsos. Neste caso pode-se utilizar o indicador de confirmação da vizinhança. Para calcular a confirmação da vizinhança utilizando o gradiente de disparidade pode-se utilizar o seguinte procedimento:

1. Selecionar o caractere na primeira imagem e todos os pares de candidatos com caracteres dentro da vizinhança do mesmo; a vizinhança é circular com centro nas posições dos caracteres do par candidato;
2. Considere a posição do primeiro caractere e a posição do caractere da vizinhança na primeira imagem como sendo os pontos terminais de um dipolo; o mesmo ocorre para obtenção de um dipolo na segunda imagem; calcular o gradiente de disparidade;
3. Se o gradiente de disparidade calculado usando estes pontos terminais não exceder um limiar pré-definido então este par da vizinhança confirma a correspondência do par no seu centro; a confirmação dada pela correspondência na vizinhança é inversamente proporcional à distância ciclopeana do par candidato do centro da vizinhança;
4. O indicador de confiabilidade do par do centro da vizinhança é a somatória da confirmação de todos os demais pares da vizinhança.

A seleção das correspondências verdadeiras é obtida utilizando um procedimento de relaxação [29]:

1. Detectar os caracteres nas duas imagens; estes caracteres podem ser, por exemplo, vértices;
2. Para cada caractere na primeira imagem selecionar os possíveis caracteres correspondentes na segunda imagem, utilizando as restrições disponíveis;

3. Para cada par candidato calcular o indicador de confiabilidade de acordo com os outros pares candidatos que não violem o limiar do gradiente de disparidade;
4. Para cada caractere da primeira imagem seleciona-se o par candidato com maior índice de confiabilidade e identifica-se o caractere correspondente na segunda imagem;
5. Se este caractere da segunda imagem não participa de nenhum outro par candidato com indicador maior que o par selecionado em (2) então aquele par é aceito;
6. Nenhum dos dois caracteres envolvidos no par aceito poderá participar de qualquer outro par candidato, garantindo a restrição de unicidade;
7. Terminar quando todos os pares possíveis forem selecionados; ou
8. Voltar ao passo (3) e recalculando os indicadores de confiabilidade.

A medida que as correspondências forem sendo aceitas, outros pares candidatos, então com indicadores menores, passarão a ser os de maior indicador e, logo, o procedimento deve ser iterativo para que se permita que todas as correspondências verdadeiras sejam consideradas e aceitas.

Pilu e Lorusso [22] sugerem um método alternativo, não iterativo, de correspondência de caracteres baseado na construção de uma matriz de correlação entre todos os pares possíveis de caracteres correspondentes e na fatorização desta matriz por valores singulares (SVD, do inglês, *singular value decomposition*).

Para uma descrição detalhada sobre a fatorização por valores singulares pode-se consultar Golub e van Loan [9] ou Tomasi [30].

## 2.4 Resumo

Neste capítulo foram discutidas técnicas de detecção de caracteres primitivos das imagens, destacando os tipos de caracteres que serão utilizados na calibração de câmeras.

Também foram discutidas técnicas de correspondência baseada em caracteres que podem ser aplicadas mesmo quando a geometria do conjunto binocular é desconhecida, o que é o caso da busca de correspondências que serão utilizadas como informação de calibração.

Tendo detectado os caracteres primitivos numa imagem e relacionado estes caracteres com coordenadas conhecidas no espaço tridimensional podemos determinar a geometria da câmera que forneceu esta imagem, o que será assunto do próximo capítulo.

## Capítulo 3

# Estereoscopia, calibração de câmeras e reconstrução

*"Ubi materia, ibi geometria."*

*Johannes Kepler*

Neste capítulo será descrito o processo de formação de imagens numa matriz de sensores através da projeção em perspectiva e serão descritas as relações geométricas existentes entre pares de pontos nas imagens obtidas de um conjunto binocular que são a projeção de um mesmo ponto no ambiente. Em seguida será discutido como estas relações podem ser utilizadas para determinar a geometria do conjunto binocular que forneceu aquelas imagens, processo este conhecido como calibração.

### 3.1 Sensores de imagem

#### 3.1.1 Câmera *pinhole* e câmera com lentes

Considere um orifício ideal circular infinitesimalmente pequeno a uma distância fixa a frente de um plano, dispondo de uma cobertura de forma que somente a luz vinda do orifício possa alcançar este plano. Cada ponto onde a luz incide neste plano corresponde a uma direção particular definida por um raio vindo daquele ponto através do orifício. Este tipo de configuração é conhecida como câmera *pinhole* (do inglês, orifício). O orifício em questão é chamado de *ponto focal* e o plano é chamado de *plano de imagem* da câmera e podemos

definir como *eixo ótico* a linha perpendicular indo do orifício  $O$  ao plano da imagem (figura 3.1). Uma câmera *pinhole* produz uma imagem que é uma *projeção em perspectiva* do ambiente [14, 3].

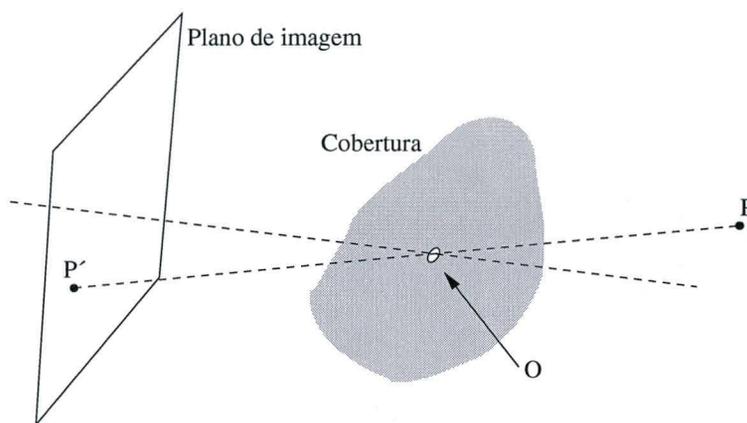


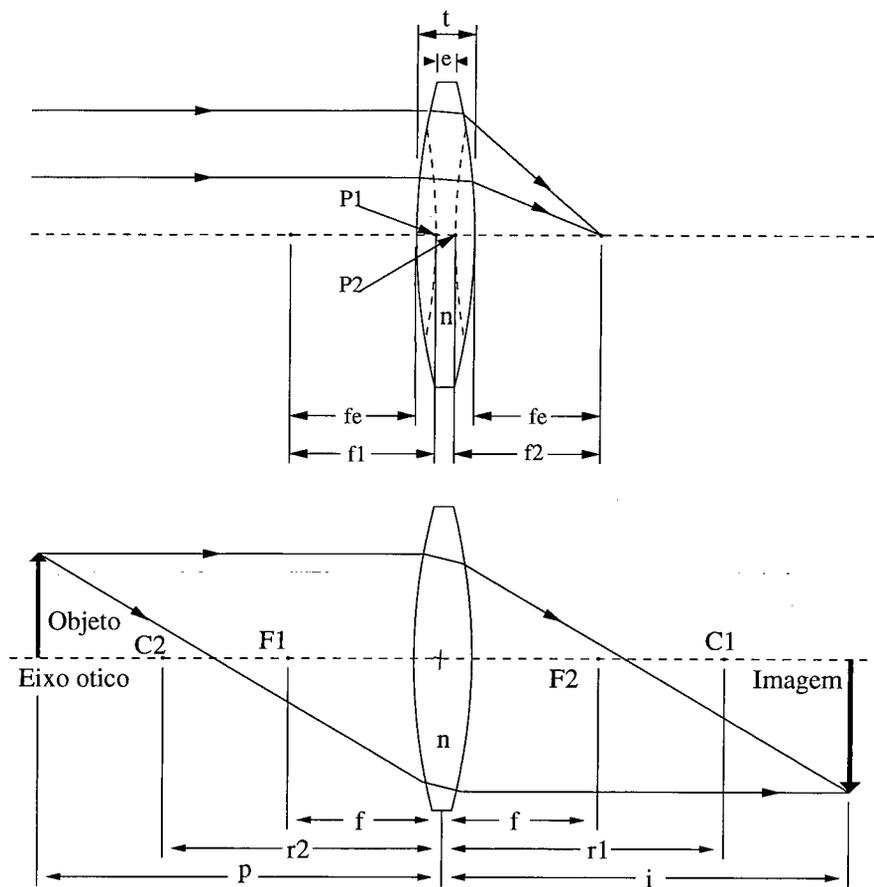
Figura 3.1: Modelo de uma câmera pinhole

Existem alguns problemas associados à câmera *pinhole*. Primeiro o orifício deve ter diâmetro finito diferente de zero para que uma quantidade finita de luz seja capturada no plano de imagem. Também o orifício não pode ser muito pequeno pois pode ocorrer difração e a luz seria dispersada por toda a imagem. Ao mesmo tempo, aumentando o diâmetro do orifício a projeção em perspectiva não se aplica mais, pois um ponto do ambiente será capturado no plano de imagem como um pequeno círculo [14].

Como pouca luz passa através do orifício da câmera *pinhole* ela tem pouca utilidade em aplicações práticas. Os sensores de imagem precisariam acumular fótons durante um longo período para capturar uma imagem com qualidade. Como resultado as câmeras utilizam lentes, que podem focalizar uma grande quantidade de luz. Uma lente simples é produzida de forma que suas duas superfícies refratoras possuam formatos esféricos. O eixo ótico é a linha que passa através dos centros destas duas esferas. Várias características das lentes simples estão representadas na figura 3.2.

Qualquer lente simples possuirá diversos defeitos e aberrações. Pode-se obter uma lente composta de melhores propriedades combinando-se várias lentes e alinhando seus eixos óticos. Os principais tipos de distorções que ocorrem em lentes reais são:

- Distorção radial: desvia o raios mais ou menos que nas lentes ideais (figura 3.3);
- Deslocamento do centro: o ponto principal é deslocado do eixo ótico.



- $i$ : distância da imagem
- $p$ : distância do objeto
- $f$ : distância focal
- $f_e$ : distância focal efetiva
- $f_1$  e  $f_2$ : distâncias focais das superfícies
- $P_1$  e  $P_2$ : pontos principais das superfícies
- $F_1$  e  $F_2$ : pontos focais das superfícies
- $r_1$  e  $r_2$ : raios de curvatura das superfícies
- $C_1$  e  $C_2$ : centros de curvatura das superfícies
- $n$ : índice de refração
- $t$ : espessura no centro da lente
- $e$ : espessura na borda da lente

Figura 3.2: Características das lentes (extraída de [7])

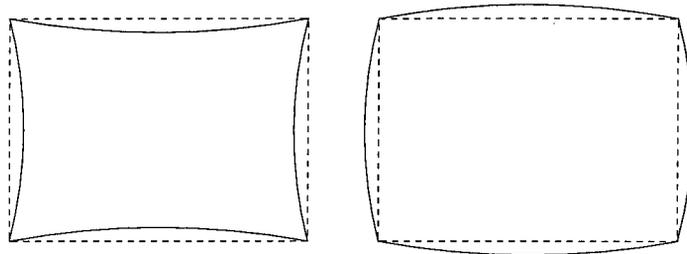


Figura 3.3: Efeitos das distorção radial da lente nas imagens

Uma lente ideal produz a mesma projeção em perspectiva do orifício ideal. Por este motivo toda a complexidade de modelagem de uma câmera com lentes é abstraída ao utilizar-se o modelo de câmera *pinhole* sobre imagens obtidas através de uma câmera com lentes, corrigindo-se apenas as distorções, especialmente a distorção radial. Deve-se, porém, observar que a distância focal nos dois modelos não será a mesma.

### 3.1.2 Matriz de sensores e digitalização

As câmeras de vídeo e os cartões de aquisição são os responsáveis pela amostragem e quantização das imagens. A imagem formada sobre o plano de imagem pode ser convertida em sinais elétricos por meio de sensores. Quase todos os sensores de imagem dependem da formação de pares elétron-lacuna quando fótons atingem uma camada de material semicondutor disposto na forma de um reticulado retangular no plano de imagem.

A principal tecnologia atualmente em uso é a matriz de dispositivos de acoplamento de carga (CCD, do inglês *charge-coupled device*). Dimensões típicas dos sensores são apresentadas na figura 3.4. Observe-se que o tamanho do sensor não é equivalente à área ativa do sensor pois existe um espaçamento entre os elementos sensores. Número de elementos sensores comuns em câmeras CCD são  $768(h) \times 494(v)$  e  $512(h) \times 494(v)$ .

Embora as câmeras baseadas em CCD sejam de estado sólido, geralmente elas são essencialmente analógicas na sua operação. Os valores de carga são armazenados como tensões que são transferidas utilizando os registradores de deslocamento para um amplificador que gera um sinal de vídeo analógico padronizado. Qualquer digitalização do sinal é feita depois que o sinal já foi transmitido em sua forma analógica. Devido a isto pode não haver uma correspondência um-para-um entre os dispositivos fotossensíveis da matriz e os elementos da imagem final. Por exemplo, uma câmera com uma matriz de  $512(h) \times 494(v)$  dispositivos pode ser utilizada para gerar um sinal de vídeo no padrão NTSC, onde temos 525 linhas

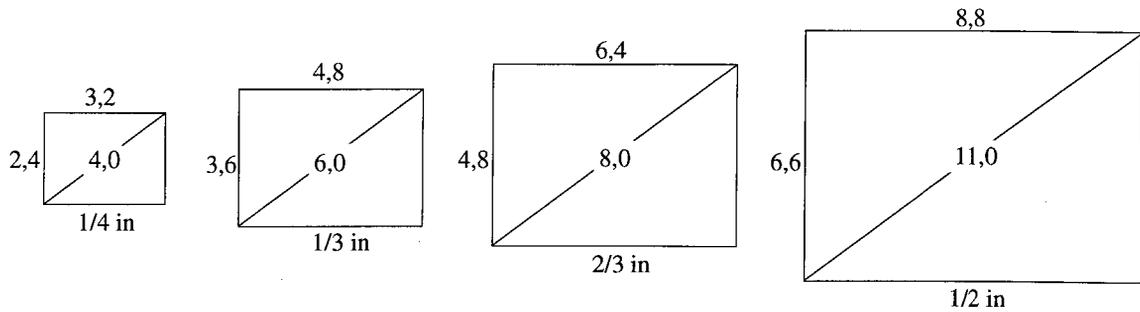


Figura 3.4: Formatos dos sensores CCD (medidas em *mm*, extraída de [7])

por imagem, que posteriormente poderá ser digitalizado gerando uma imagem  $640 \times 480$  elementos[24].

Características adicionais sobre a codificação do sinal de vídeo, sincronização e digitalização do sinal podem ser encontradas em Rosandish [24] e Jonker [16]. Gomes e Velho [10] traz também diversos conceitos relacionados à discretização e codificação da imagem, assim como conceitos sobre percepção de cor.

A discretização espacial da imagem, como feita pelas câmeras e cartões de aquisição, consiste em tomar o domínio como sendo um retângulo

$$U = [a, b] \times [c, d] = (x, y) \in \mathbb{R}^2; \quad a \leq x \leq b \text{ e } c \leq y \leq d \quad (3.1)$$

e discretizar esse retângulo usando pontos de um reticulado bidimensional  $\Delta = (\Delta_x, \Delta_y)$ ,

$$\Delta = \{(x_j, y_k) \in U; x_j = j \cdot \Delta_x; y_k = k \cdot \Delta_y; j, k \in \mathbb{Z}, \Delta_x, \Delta_y \in \mathbb{R}\} \quad (3.2)$$

conforme é mostrado na figura 3.5.

Como cada píxel,  $(x_j, y_k)$  da imagem pode ser representado por coordenadas inteiras  $(j, k)$ , a imagem pode ser convenientemente representada através de uma matriz  $A$  de ordem  $m \times n$ ,  $A = (a_{jk}) = (i(x_j, y_k))$ .

$$f(x, y) = \begin{bmatrix} f(0, 0) & f(0, 1) & \dots & f(0, M - 1) \\ f(1, 0) & f(1, 1) & \dots & f(1, M - 1) \\ \dots & \dots & \dots & \dots \\ f(N - 1, 0) & f(N - 1, 1) & \dots & f(N - 1, M - 1) \end{bmatrix} \quad (3.3)$$

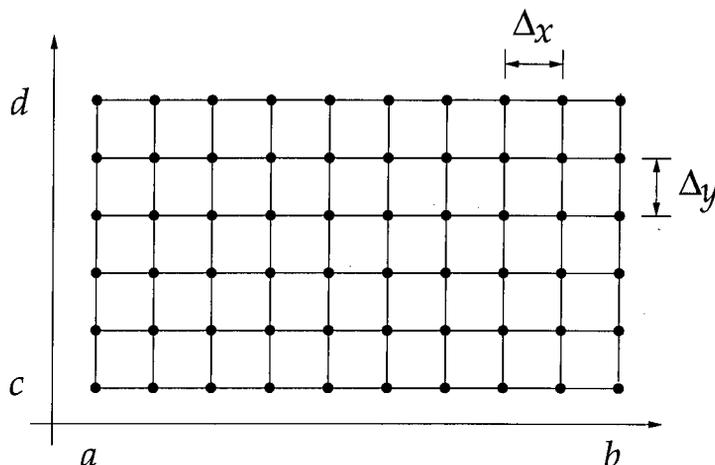


Figura 3.5: Reticulado uniforme da representação matricial da imagem (extraída de [10])

Cada elemento  $a_{jk}$ ,  $j = 1, \dots, m$  e  $k = 1, \dots, n$  da matriz representa o valor da função imagem  $i$  no ponto de coordenadas  $(x_j, y_k)$  do reticulado, sendo geralmente um vetor do espaço de cor, representando a cor do píxel de coordenadas  $(j, k)$ . Se a imagem for monocromática,  $(a_{jk})$  é uma matriz real, onde cada elemento é um número real que representa o valor de luminosidade do píxel [10].

O quociente  $(d - c)/(b - a)$  é chamado de *razão de aspecto da imagem* e imagens produzidas por câmeras de vídeo geralmente possuem razão de aspecto de 3/4. O quociente  $\Delta_y/\Delta_x$  é chamado de *razão de aspecto do píxel*. A razão de aspecto do píxel determina a forma geométrica do píxel, que pode ser retangular ou quadrada. É muito comum o uso de dispositivos com razão de aspecto do píxel igual a 1 (ou seja  $\Delta_x = \Delta_y$ ), no entanto existem diversos dispositivos onde essa razão não é 1

O número de linhas  $m$  da matriz  $A$  de píxeis é chamado de *resolução vertical* da imagem, e o número  $n$  de colunas é chamado de *resolução horizontal* da imagem. A ordem  $m \times n$  da matriz  $(a_{jk})$  que representa uma imagem, ou o produto  $m \times n$  da resolução vertical pela resolução horizontal é chamada de *resolução espacial* ou *resolução geométrica* da imagem. A resolução espacial estabelece a frequência de amostragem final da imagem. Dessa forma, quanto maior a resolução mais detalhes, isto é, altas frequências, da imagem podem ser captados na representação matricial [10].

Além da informação de cor, as outras componentes do contra domínio da imagem, estão freqüentemente associados a informações auxiliares, tais intensidade de borda no píxel, identificador de região do píxel e profundidade do ponto na cena [10].

O *gamute* de uma imagem digital é o conjunto de cores do espaço de cor quantizado da imagem. Uma imagem monocromática cujo gamute possui apenas duas cores é chamada de imagem binível, imagem binária, ou ainda imagem *bitmap*. Uma imagem monocromática cujo gamute possui mais de dois níveis é chamada de imagem de meio-tom, ou imagem com escala de cinza, (do inglês *halftone* ou *grayscale*) [10].

A maioria dos dispositivos sensores eletrônicos possui um espaço de cor tricromático tomando como base de cores primárias as cores vermelho (em inglês *red*), verde (em inglês *green*) e azul (do inglês *blue*), daí serem chamados de RGB.

Geralmente é necessário a conversão entre os diferentes formatos de representação de imagem. Imagens coloridas podem ser convertidas em imagens de escala de cinza utilizando a função de luminância do sistema de cor associado.

A luminância de uma cor  $C$  com coordenadas  $(R, G, B)$  no espaço de cor RGB é dada por:

$$L(C) = 0,299R + 0,587G + 0,114B \quad (3.4)$$

Finalmente, imagens em escala de cinza podem ser convertidas em imagens binárias aplicando-se um limiar de corte.

## 3.2 Visão estereoscópica

### 3.2.1 Disparidade estereoscópica e profundidade

Do mesmo modo que na visão humana, em um sistema de visão computacional a profundidade pode ser calculada a partir da disparidade se duas câmeras forem utilizadas para capturar imagens de pontos de vista diferentes. A geometria envolvida pode ser ilustrada com um exemplo simples, partindo do que foi introduzido na seção 2.2. Suponha que acoplemos rigidamente duas câmeras idênticas uma a outra de forma que seus planos de imagem sejam coplanares, seus eixos óticos sejam paralelos, perpendiculares ao plano de imagem, e separados por uma distância  $b$  (figura 3.6) [24, 14].

A linha conectando os centros das lentes é chamada de *linha de base* (do inglês *baseline*). Assuma que a linha de base seja perpendicular aos eixos óticos e oriente o eixo  $x$  de forma que ele seja paralelo à linha de base. As coordenadas  $(x, y, z)$  de um ponto  $P$  no ambiente são medidas em relação à origem equidistante aos centros de projeção das duas câmeras [14].

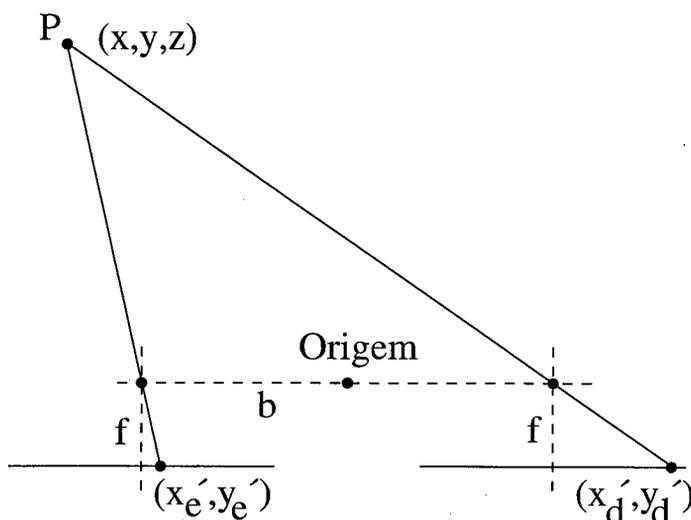


Figura 3.6: Geometria da câmera para estéreo geométrico (adaptada de [14])

Na figura 3.6, se a dimensão  $x$  é escolhida como sendo ao longo da linha que liga os centros dos planos de imagem de cada uma das câmeras, então qualquer ponto que apareça em ambas as imagens aparecerá na mesma altura (ou seja, mesma dimensão  $y$ ) e diferirá de posição somente na dimensão  $x$ .

Considere as coordenadas da projeção do ponto  $P$  nas imagens da esquerda e direita como sendo  $(x'_e, y'_e)$  e  $(x'_d, y'_d)$ , respectivamente. A diferença nas coordenadas  $x'_e - x'_d$  é a *disparidade* [14]:

$$d = x'_e - x'_d \quad (3.5)$$

Tomando a geometria dos triângulos equivalentes é simples mostrar que a profundidade  $z$ , ou distância da câmera, do ponto  $P$  pode ser calculada a partir da disparidade como:

$$z = f \frac{b}{d} \quad (3.6)$$

onde  $f$  é a distância focal das câmeras e  $b$  é a distância de linha de base entre as câmeras. Em outras palavras, a distância da câmera é inversamente proporcional à disparidade. Este fato possui efeito sobre a precisão. Objetos distantes possuem pouca disparidade de forma que determinar sua profundidade com precisão é impossível, enquanto objetos próximos possuem maior disparidade e suas profundidades podem ser determinadas com grande precisão [14, 24].

Na prática, as duas câmeras utilizadas para obter o par estéreo não estarão exatamente alinhadas, como foi assumido nesta análise simplificada. Pode ser muito difícil instalar o sistema de forma que os eixos óticos sejam exatamente paralelos e com a linha de base exatamente perpendicular aos eixos óticos. Daí a necessidade da calibração das câmeras após a montagem do conjunto binocular, e em alguns casos, até durante sua operação.

A calibração será assunto da seção 3.4 adiante, mas antes é conveniente detalhar alguns aspectos da projeção em perspectiva e o modelo de câmera que será utilizado neste trabalho.

### 3.2.2 Projeção em perspectiva e modelagem da câmera

A projeção em perspectiva descreve a formação de imagens por uma câmera *pinhole* ou com lentes, e está esquematizada na figura 3.7. O plano  $\pi$  mostrado na figura é o plano de imagem, a linha tracejada é o eixo ótico, e as lentes estão posicionadas perpendiculares ao eixo ótico no ponto focal  $C$  (também chamado de centro ótico ou centro de projeção). A distância focal  $f$  é um parâmetro da lente utilizada [29].

Embora o ponto focal esteja na realidade entre o objeto e o plano de imagem, geralmente é representado do lado contrário do plano de imagem, evitando o inconveniente de representar os objetos na posição invertida. É como se a câmera formasse uma imagem na frente da lente, e não atrás dela. Esta forma de modelar a câmera é mais distante da realidade física, porém as relações matemáticas são preservadas e os modelos são formalmente equivalentes [14, 3, 35].

Existem quatro sistemas de coordenadas envolvidos na figura 3.7: ambiente, câmera, sensor e imagem. Nos três primeiros sistemas de coordenadas geralmente utiliza-se milímetro como unidade de medida, enquanto no último utiliza-se píxel.

- Ambiente: sistema de coordenadas euclidiano, identificado pelo índice  $w$ , com origem no ponto  $O_w$ ;
- Câmera: sistema de coordenadas euclidiano, identificado pelo índice  $c$ , com origem no ponto focal  $C = O_c$ ;
- Imagem: sistema de coordenadas euclidiano, identificado pelo índice  $i$ , tem os eixos alinhados com o sistema de coordenadas da câmera, mas com origem no plano de imagem, no canto superior esquerdo;
- Sensor: sistema de coordenadas afim, identificado pelo índice  $a$ , com eixos  $u$ ,  $v$  e  $w$ , origem  $O_i$  coincidindo com a origem do sistema de coordenadas da imagem, com

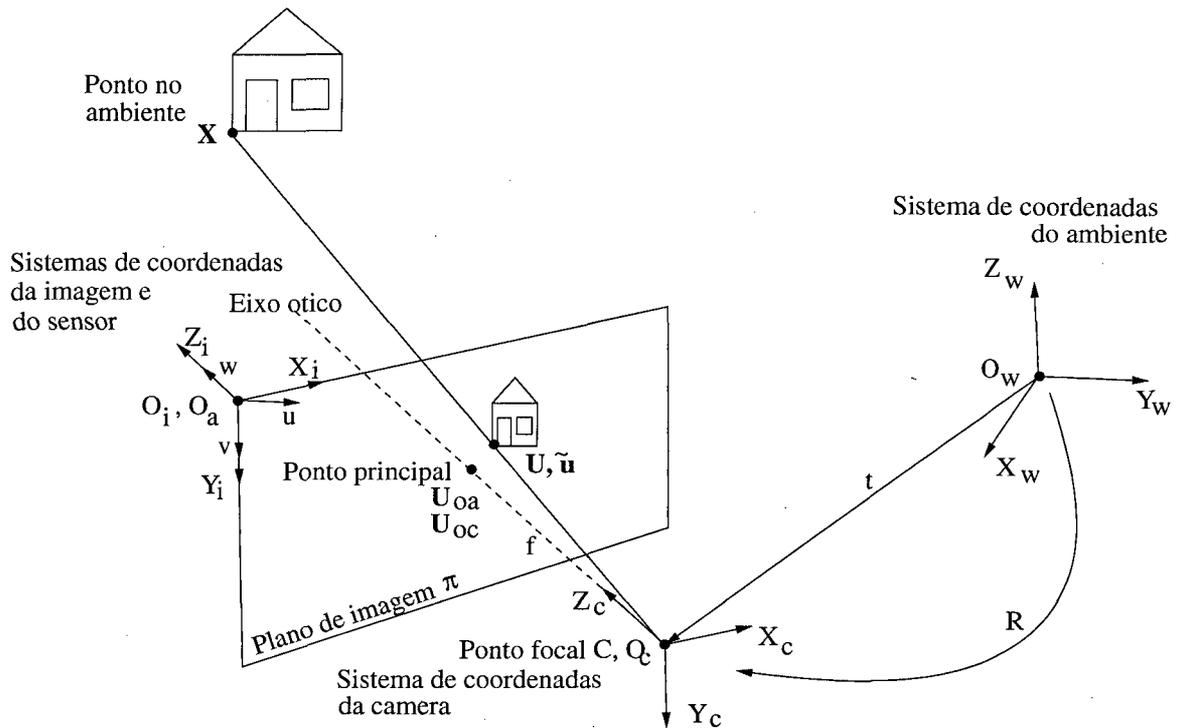


Figura 3.7: Geometria da projeção em perspectiva de uma câmera (adaptada de [29])

os eixos  $w$  e  $v$  alinhados com os eixos  $Z_i$  e  $Y_i$ , porém com o eixo  $u$  com orientação possivelmente diferente do eixo  $X_i$ .

A necessidade do sistema de coordenadas do sensor reside no fato de o reticulado de píxeis poder ser não perpendicular e poder possuir eixos com escalas diferentes.

Na figura 3.7, o ponto  $X$  no ambiente é mapeado para o sistema de coordenadas da câmera  $X_c$  por uma rotação  $R$  em torno da origem seguida de uma translação  $t$ :

$$X_c = RX_w + t \tag{3.7}$$

ou, em coordenadas homogêneas:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \tag{3.8}$$

Considerando que o plano de imagem seja realmente plano e que a distorção da lente

seja desprezível (e este modelo é geralmente próximo da realidade), a relação entre as coordenadas de um ponto  $\mathbf{X}_c$  no ambiente e sua projeção  $\mathbf{U}_c$  no plano de imagem, é bastante simples. As coordenadas  $x$  e  $y$  deste ponto podem ser calculadas por triângulos equivalentes como representado na figura 3.8 para  $x$  [29]:

$$\mathbf{U}_c = \left[ \frac{fx_c}{z_c}, \frac{fy_c}{z_c}, f \right]^T \quad (3.9)$$

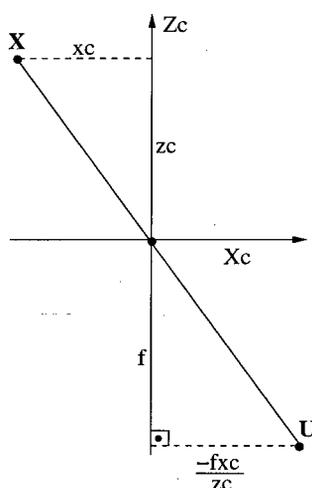


Figura 3.8: Cálculo das coordenadas do ponto projetado (extraída de [29])

Estas equações de projeção em perspectiva são inerentemente não lineares. Todavia elas podem ser escritas como um mapeamento linear em coordenadas homogêneas [1]:

$$\lambda \mathbf{U}_c = \lambda \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathbf{X}_c \quad (3.10)$$

onde  $\lambda \neq 0$  é um fator de escala .

O sistema de coordenadas afim do sensor representa o desvio e escala (razão de aspecto) do sistema de coordenadas da imagem. O ponto  $\mathbf{U}_0$ , que representa o centro da imagem, é a interseção do eixo óptico com o plano de imagem  $\pi$ . No sistema de coordenadas do sensor

o ponto  $U_0$  é expresso como:

$$\mathbf{U}_{0a} = [u_0, v_0, 0]^T \quad (3.11)$$

O ponto  $U$  é representado no plano de imagem  $\pi$  em coordenadas homogêneas como  $\tilde{\mathbf{u}} = [U, V, W]^T$ , ou, no plano euclidiano,  $\mathbf{u} = [u, v]^T = [U/W, V/W]^T$

O uso de coordenadas homogêneas nos permite expressar a transformação afim como a multiplicação por uma matriz  $3 \times 3$ :

$$\tilde{\mathbf{u}} = \begin{bmatrix} k_u & k_c & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{U}_c \quad (3.12)$$

onde as incógnitas  $k_u$ ,  $k_v$  e  $k_c$  representam fator de escala dos eixos e o desvio entre estes eixos, enquanto  $u_0$  e  $v_0$  representam as coordenadas afins do centro de projeção na imagem, a unidade de medida de  $k_u$ ,  $k_v$  e  $k_c$  é píxel por unidade de comprimento, usualmente píxel por milímetro.

$$k_u = \frac{f}{\Delta_x} \quad k_v = \frac{f}{\Delta_y} \quad (3.13)$$

O processo completo de formação de imagens é então:

$$\begin{aligned} \tilde{\mathbf{u}} &= \begin{bmatrix} k_u & k_c & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} \alpha_u & -\alpha_u \cot \theta & u_0 \\ 0 & \alpha_v / \sin \theta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \\ &= A[R|\mathbf{t}] \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \end{aligned} \quad (3.14)$$

o que define a *matriz de projeção*, ou ainda, *matriz da câmera*,  $\tilde{P}$ :

$$\tilde{\mathbf{u}} = \tilde{P} \begin{bmatrix} \mathbf{X}_w \\ 1 \end{bmatrix} \quad \tilde{P} = A[R|\mathbf{t}] \quad (3.15)$$

A matriz  $A$  contém todos os *parâmetros internos* da câmera, ou parâmetros intrínsecos, e é também conhecida como *matriz de calibração da câmera*. Estes parâmetros são independentes da posição e orientação da câmera. Os *parâmetros externos*, ou extrínsecos, são representados pela matriz  $R$  e pelo vetor  $\mathbf{t}$ . A matriz  $R$  representa três rotações elementares em torno dos eixos e a translação  $\mathbf{t}$  representa o deslocamento da câmera em relação ao sistema de coordenadas do ambiente. Os parâmetros  $\alpha_u$  e  $\alpha_v$  são os fatores de escala do plano de imagem (em unidades da distância focal  $f$ ),  $\theta$  é o ângulo entre os eixos da matriz de sensores, sendo  $\theta = \pi/2$  para a maioria das câmeras [1].

É conveniente decompor a matriz de projeção  $3 \times 4$   $\tilde{P}$  em uma matriz  $3 \times 4$   $P$  e um vetor  $3 \times 1$   $p$ :

$$\tilde{P} = [P \ p] \quad (3.16)$$

de forma que

$$P = A R \quad e \quad p = A t \quad (3.17)$$

Também se conhecemos  $\tilde{P}$  podemos obter  $A$ ,  $R$ , e  $t$  utilizando fatorização QR. Para detalhes sobre a fatorização QR pode-se consultar Golub e van Loan [9]. Em resumo, podemos obter  $A$ ,  $R$ , e  $t$  como:

$$\begin{aligned} [\hat{Q}, \hat{R}] &= QR(P^{-1}) \\ R &= \hat{Q}^{-1} \\ t &= \hat{R} p \end{aligned} \quad (3.18)$$

### 3.2.3 Geometria epipolar e a matriz fundamental

A geometria epipolar descreve as relações geométricas entre duas câmeras. Três conceitos podem ser definidos (ver figura 3.9):

- *Epipolo*: é o ponto de intersecção do plano de imagem com a linha unindo os dois pontos focais (a linha de base); o epipolo é a imagem do ponto focal de uma câmera no plano de imagem da outra ( $e_1$  e  $e_2$ );
- *Plano epipolar*: o plano definido por um ponto no espaço tridimensional e os dois pontos focais, ou, de forma equivalente, por um ponto no plano de imagem e os dois pontos focais ( $Mc_1c_2$ ,  $m_1c_1c_2$  ou  $m_2c_1c_2$ );
- *Linha epipolar*: é a linha reta de intersecção do plano epipolar com o plano de imagem; considere uma linha passando pelo ponto focal e um ponto no plano de imagem, a linha epipolar é a imagem desta linha no plano de imagem da outra câmera; todas as linhas epipolares intersectam no epipolo ( $e_1m_1$  ou  $e_2m_2$ ).

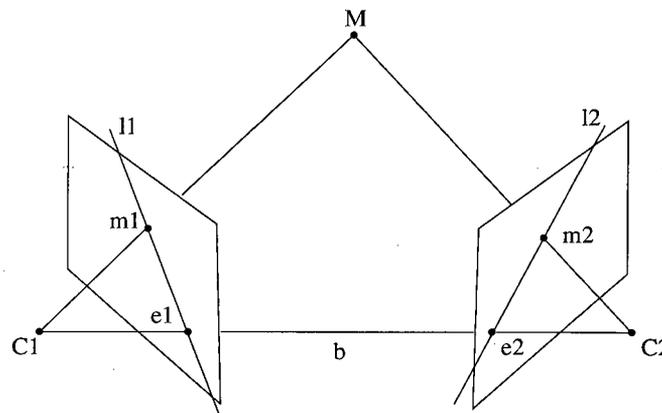


Figura 3.9: A geometria epipolar

Se tomarmos pontos diferentes no espaço observamos que os planos epipolares giram em torno da linha de base (figura 3.10), o que define uma família de planos.

O vetor do ponto focal da primeira câmera ao ponto no primeiro plano de imagem, o vetor do ponto focal na segunda câmera ao ponto no segundo plano de imagem, e o vetor do ponto focal da primeira câmera ao ponto focal da segunda são todos coplanares, o que é conhecido como a restrição epipolar [1].

Um caractere que apareça em uma imagem pode ou não possuir seu correspondente na outra imagem; mas se possuir, este deve estar em algum lugar ao longo da linha epipolar correspondente. Observe-se ainda que na geometria simplificada da figura 3.6 todas as linhas epipolares são paralelas ao eixo  $x$ , o que não é verdade no caso geral [14, 24].

Suponha que um conjunto binocular esteja vendo um ponto  $M$  no espaço, o qual é projetado nos dois planos de imagem nas posições  $\bar{m}_1$  e  $\bar{m}_2$ . Como estamos tratando de coordena-

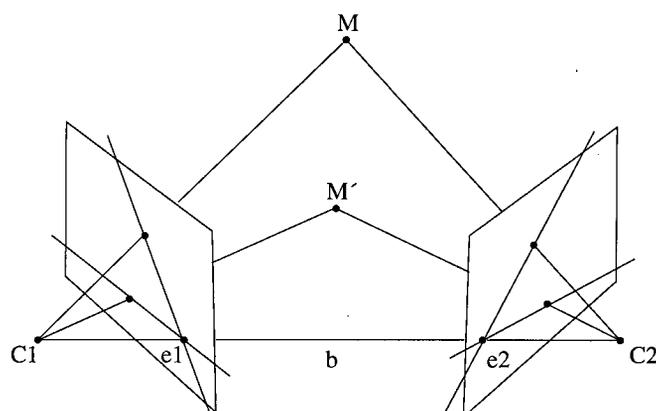


Figura 3.10: Família de planos epipolares

nadas homogêneas  $\mathbf{M}$  é um vetor  $4 \times 1$  enquanto  $\bar{\mathbf{m}}_1$  e  $\bar{\mathbf{m}}_2$  são vetores  $3 \times 1$ . Se considerarmos que a informação de calibração das câmeras está disponível então obtemos  $\bar{\mathbf{m}}_1$  e  $\bar{\mathbf{m}}_2$  no sistema e coordenadas da câmera com origem no ponto focal. No sistema de coordenadas da câmera estas restrições da geometria epipolar podem ser expressas simplesmente como [1]:

$$\bar{\mathbf{m}}_2^T (\mathbf{t} \times R\bar{\mathbf{m}}_1) = 0 \quad (3.19)$$

onde  $R$  e  $\mathbf{t}$  referem-se à rotação e translação entre os sistemas de coordenadas das duas câmeras.

Agora suponha que a informação de calibração não esteja disponível e só conheçamos as posições no sistema de coordenadas da imagem. Neste caso os efeitos dos parâmetros internos, representados pelas matrizes  $A_1$  e  $A_2$  (de (3.12)), estarão presentes, transformando as coordenadas da câmera em coordenadas da imagem:

$$\begin{aligned} \mathbf{m}_1 &= A_1 \bar{\mathbf{m}}_1 \\ \mathbf{m}_2 &= A_2 \bar{\mathbf{m}}_2 \end{aligned} \quad (3.20)$$

O que nos leva às seguintes equações [1]:

$$\begin{aligned} (A_2^{-1} \mathbf{m}_2)^T (\mathbf{t} \times R A_1^{-1} \mathbf{m}_1) &= 0 \\ \mathbf{m}_2^T A_2^{-T} (\mathbf{t} \times R A_1^{-1} \mathbf{m}_1) &= 0 \end{aligned} \quad (3.21)$$

Definindo-se  $[\mathbf{t}]_{\times}$  como a matriz antissimétrica tal que  $[\mathbf{t}]_{\times} \mathbf{y} = \mathbf{t} \times \mathbf{y}$ , para qualquer vetor  $\mathbf{y}$ , ou seja, se  $\mathbf{t} = [a \ b \ c]^T$ , então

$$[\mathbf{t}]_{\times} = \begin{bmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{bmatrix} \quad (3.22)$$

podemos agora escrever, de (3.21),

$$\begin{aligned} \mathbf{m}_2^T A_2^{-T} ([\mathbf{t}]_{\times} R A_1^{-1} \mathbf{m}_1) &= 0 \\ \mathbf{m}_2^T F \mathbf{m}_1 &= 0 \end{aligned} \quad (3.23)$$

onde  $F = A_2^{-T} [\mathbf{t}]_{\times} R A_1^{-1}$  é a *matriz fundamental*.

A matriz fundamental descreve plenamente a relação geométrica entre pontos correspondentes nos dois planos de imagem e depende somente dos parâmetros internos das câmeras e da rotação e translação relativa entre as câmeras. A matriz fundamental possui sete parâmetros, ou graus de liberdade, dois para cada epipolo e três para a relação entre as linhas epipolares.

Dado um ponto  $m_1$  com coordenadas  $(u_1, v_1)$  no plano de imagem  $\pi_1$  da primeira câmera, sabe-se então que seu ponto correspondente na outra câmera e a matriz fundamental estão relacionados por:

$$[u, v, 1] F \begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} = 0. \quad (3.24)$$

o que pode ser expandido para

$$(F_{11}u_1 + F_{12}v_1 + F_{13})u + (F_{21}u_1 + F_{22}v_1 + F_{23})v + (F_{31}u_1 + F_{32}v_1 + F_{33}) = 0 \quad (3.25)$$

onde  $u_1$  e  $v_1$  são conhecidos e  $u$  e  $v$  são variáveis. Esta equação é a equação da linha epipolar no plano de imagem  $\pi_2$  da segunda câmera a qual contém o ponto  $m_2$  correspondente.

Dado o ponto  $m_1$ , para calcular a linha epipolar  $l_{m_1}$  que contém  $m_1$  precisamos primeiro calcular o epipolo pertencente ao plano de imagem  $\pi_1$ . O epipolo é definido como a inter-

secção de todas as linhas epipolares em  $\pi_1$ , logo:

$$\begin{aligned} m_2^T F e &= 0 \quad \forall m_2^T \in \pi_2 \\ F e &= 0 \end{aligned} \quad (3.26)$$

O epipolo  $e_1$  em  $\pi_1$  é a projeção de  $C_1$  em  $\pi_1$ , enquanto o epipolo  $e_2$  em  $\pi_2$  é a projeção de  $C_2$  em  $\pi_2$ .

Agora as coordenadas de  $l_{m1}$  podem ser calculadas como o produto  $e_1 \times m_1$ .

### 3.3 Reconstrução

Dispondo das matrizes de projeção  $\tilde{P}_1$  e  $\tilde{P}_2$  das duas câmeras do conjunto binocular e das coordenadas  $\tilde{u}_1$  e  $\tilde{u}_2$  da projeção de um ponto  $P$  no espaço podemos calcular as coordenadas  $P_w$  deste ponto no sistema de coordenadas do ambiente. Seguindo o modelo de câmera *pinhole* temos:

$$\begin{aligned} \lambda_1 [u_{i1}, v_{i1}, 1]^T &= P_1 [x_w, y_w, z_w, t_w]^T \\ \lambda_2 [u_{i2}, v_{i2}, 1]^T &= P_2 [x_w, y_w, z_w, t_w]^T \end{aligned} \quad (3.27)$$

onde  $\lambda_1$  e  $\lambda_2$  são fatores de escala.

Indicando como  $p_{1i}$  o vetor correspondente à  $i$ -ésima linha de  $\tilde{P}_1$  e como  $p_{2i}$  o vetor correspondente à  $i$ -ésima linha de  $\tilde{P}_2$ , os fatores de escala podem ser calculados como [37]:

$$\begin{aligned} \lambda_1 &= p_{13}^T P_w \\ \lambda_2 &= p_{23}^T P_w \end{aligned} \quad (3.28)$$

Eliminando  $\lambda_1$  e  $\lambda_2$  da equação (3.27) temos a seguinte equação [37]:

$$B P_w = 0 \quad (3.29)$$

onde  $B$  é uma matriz  $4 \times 4$  dada por:

$$B = \begin{bmatrix} \mathbf{P}_{11} - u_{i1}\mathbf{P}_{13} \\ \mathbf{P}_{12} - v_{i1}\mathbf{P}_{13} \\ \mathbf{P}_{21} - u_{i2}\mathbf{P}_{23} \\ \mathbf{P}_{22} - v_{i2}\mathbf{P}_{23} \end{bmatrix} \quad (3.30)$$

A solução da equação (3.29) pode ser obtida por decomposição por valores singulares (SVD), tomando-se o autovetor da matriz  $B^T B$  associado ao menor autovalor. Supondo-se que não existam pontos no infinito ([1]) podemos determinar  $t = 1$  em  $\mathbf{P}_w$  e obter a reconstrução projetiva da mesma forma que a reconstrução euclidiana.

O método de solução por SVD tem a desvantagem de que o critério de minimização não possui uma interpretação física adequada. Como alternativa podemos proceder à minimização no plano de imagem, minimizando o seguinte critério [37]:

$$\left( u_{i1} - \frac{\mathbf{p}_{11}^T \mathbf{P}_w}{\mathbf{p}_{13}^T \mathbf{P}_w} \right)^2 + \left( v_{i1} - \frac{\mathbf{p}_{12}^T \mathbf{P}_w}{\mathbf{p}_{13}^T \mathbf{P}_w} \right)^2 + \left( u_{i2} - \frac{\mathbf{p}_{21}^T \mathbf{P}_w}{\mathbf{p}_{23}^T \mathbf{P}_w} \right)^2 + \left( v_{i2} - \frac{\mathbf{p}_{22}^T \mathbf{P}_w}{\mathbf{p}_{23}^T \mathbf{P}_w} \right)^2 \quad (3.31)$$

A solução calculada por SVD pode ser utilizada como estimativa inicial para a minimização iterativa utilizando o critério acima.

Em muitas aplicações, e em especial neste trabalho, o sistema de coordenadas do ambiente está acoplado inicialmente ao gabarito de calibração. Por conveniência pode-se definir um sistema de coordenadas com origem no conjunto binocular, a qual, geralmente, se faz coincidir com o centro de projeção de uma das câmeras ou com o ponto médio da linha de base.

Novamente, dispondo das matrizes de projeção  $\tilde{P}_1$  e  $\tilde{P}_2$  das duas câmeras do conjunto binocular pode-se obter novas matrizes de projeção  $\hat{P}_1$  e  $\hat{P}_2$  que fornecerão a posição de ponto  $P$  do espaço no sistema de coordenadas do conjunto binocular. Para isso é necessário primeiro determinar uma transformação  $T$  entre estes sistemas de coordenadas.

Tomando-se como origem do sistema de coordenadas do conjunto binocular o centro de projeção da câmera com matriz de projeção  $\tilde{P}_1$  temos:

$$\begin{aligned}
T &= \begin{bmatrix} R_1^{-1} & -R_1^{-1}\mathbf{t}_1 \\ \mathbf{0} & 1 \end{bmatrix} \\
\hat{P}_1 &= A_1[R_1 \ \mathbf{t}_1]T \\
\hat{P}_2 &= A_2[R_2 \ \mathbf{t}_2]T
\end{aligned} \tag{3.32}$$

onde  $A$ ,  $R$  e  $\mathbf{t}$  são obtidos por fatorização QR, conforme equação (3.18).

A reconstrução utilizando as matrizes  $\hat{P}_1$  e  $\hat{P}_2$  pode ser obtida utilizando-se os mesmos métodos discutidos anteriormente, observando-se somente que a posição no espaço tridimensional será dada em relação ao centro de projeção da câmera.

## 3.4 Calibração de câmeras

### 3.4.1 Distorção radial

A compensação da distorção radial permite que se obtenha imagens não distorcidas a partir das imagens fornecidas pela câmeras.

Como visto na seção 3.1.1 a distorção radial e o deslocamento do ponto principal são os componentes da distorção das lentes mais relevantes. Na maioria dos casos ambos podem ser tratados como rotacionalmente simétricos e são modelados através de polinômios.

Seja  $(u, v)$  as coordenadas de imagem corretas,  $(\tilde{u}, \tilde{v})$  as coordenadas de imagem medidas e distorcidas provenientes das coordenadas de sensor  $(x, y)$ , e a posição estimada do ponto principal  $(\hat{u}_0, \hat{v}_0)$ . Podemos então escrever:

$$\tilde{u} = x - \hat{u}_0 \quad \tilde{v} = y - \hat{v}_0 \tag{3.33}$$

As coordenadas de imagem corretas podem ser obtidas se as correções  $(\delta u, \delta v)$  são adicionadas às coordenadas medidas incorretas:

$$u = \tilde{u} + \delta u \quad v = \tilde{v} + \delta v \tag{3.34}$$

Os valores de correção  $(\delta u, \delta v)$  são geralmente modelados como polinômios em potências pares para garantir a propriedade de invariância de rotação, geralmente elementos até a

sexta ordem são considerados:

$$\delta u = (\tilde{u} - u_p)(k_1 r^2 + k_2 r^4 + k_3 r^6) \quad \delta v = (\tilde{v} - v_p)(k_1 r^2 + k_2 r^4 + k_3 r^6) \quad (3.35)$$

onde  $(u_p, v_p)$  é a correção da posição do ponto principal e  $r^2$  é o quadrado da distância radial do centro da imagem, dado por:

$$r^2 = (\tilde{u} - u_p)^2 + (\tilde{v} - v_p)^2 \quad (3.36)$$

Os valores  $(u_p, v_p)$  podem ser utilizados após a calibração para obter a posição correta do ponto principal:

$$u_0 = \hat{u}_0 + u_p \quad v_0 = \hat{v}_0 + v_p \quad (3.37)$$

Em casos típicos a distorção radial pode ser satisfatoriamente modelada pelo polinômio de segunda ordem e sem deslocamento do ponto principal

$$u = \tilde{u}(1 \pm k_1(\tilde{u}^2 + \tilde{v}^2)) \quad v = \tilde{v}(1 \pm k_1(\tilde{u}^2 + \tilde{v}^2)) \quad (3.38)$$

### 3.4.2 Calibração de uma câmera - método de Tsai

O método de calibração de câmera de Tsai [33] calcula a informação de calibração utilizando o modelo de câmera *pinhole* e projeção em perspectiva como mostrado na seção 3.2.2, porém adicionando um parâmetro  $\kappa$  que é o coeficiente de distorção radial da lente. Além dos sistemas de coordenadas já descritos, o método de Tsai utiliza ainda um sistema de coordenadas auxiliar para tratamento da distorção radial e desvio do centro de projeção das lentes.

A rigor, o método de Tsai trata não só a calibração da câmera mas do conjunto câmera e cartão de aquisição como um todo.

Mesmo no caso de um conjunto binocular o método de Tsai não faz uso das restrições da geometria epipolar pois as câmeras são calibradas uma de cada vez.

A informação de calibração é calculada partindo de pontos no ambiente para os quais são conhecidas as posições no sistema de coordenadas do ambiente. Para cada um dos pontos deve ser conhecida também a posição de sua projeção no sistema de coordenadas da imagem.

A calibração se dá em várias etapas. Assumindo-se que se disponha de uma estimativa da altura e largura do píxel e das coordenadas do centro da imagem o método de Tsai primeiro calcula os parâmetros internos da câmeras, sem tratar a distorção radial, utilizando uma técnica de otimização por mínimos quadrados. A solução exata para os parâmetros internos é então calculada em uma nova etapa de otimização. Finalmente é calculado o coeficiente de distorção radial e o centro de projeção, também utilizando otimização por mínimos quadrados. Uma descrição mais detalhada deste método de calibração pode ser encontrado em Tsai [33]

O resultado da calibração pelo método de Tsai é o valor para onze parâmetros, cinco internos e seis externos, que descrevem a geometria da câmera:

- $f$ : distância focal da câmera *pinhole*;
- $\kappa_1$ : coeficiente de distorção radial de primeira ordem;
- $u_0, v_0$ : coordenadas do centro de distorção radial das lentes e também a intersecção do eixo ótico com o plano de imagem;
- $s_x$ : fator de escala tratando quaisquer imprecisões causadas pela diferença de resolução horizontal entre a matriz de sensor e o cartão de aquisição;
- $R_x, R_y, R_z$ : ângulos de rotação para transformação entre os sistemas de coordenadas do ambiente e da câmera;
- $t_x, t_y, t_z$ : componentes das translação para transformação entre os sistemas de coordenadas do ambiente e da câmera.

Dispondo dos parâmetros de Tsai podemos calcular a matriz de projeção. Tomando  $k_u$  e  $k_v$  como:

$$\begin{aligned} k_u &= \frac{f}{s_x \Delta_x} \\ k_v &= \frac{f}{\Delta_y} \end{aligned} \quad (3.39)$$

podemos compor a matriz de parâmetros internos  $A$ :

$$A = \begin{bmatrix} k_u & 0 & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3.40)$$

Para calcular a matriz de rotação  $R$ , a partir dos ângulos  $R_x$ ,  $R_y$  e  $R_z$ , temos [8]:

$$R = \begin{bmatrix} \cos R_z & -\text{sen } R_z & 0 \\ \text{sen } R_z & \cos R_z & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos R_y & 0 & \text{sen } R_y \\ 0 & 1 & 0 \\ -\text{sen } R_y & 0 & \cos R_y \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos R_x & -\text{sen } R_x \\ 0 & \text{sen } R_x & \cos R_x \end{bmatrix} \quad (3.41)$$

Finalmente o vetor  $\mathbf{t}$  é simplesmente a translação  $\mathbf{t} = [t_x, t_y, t_z]^T$ .

Daí podemos voltar à equação (3.15)

$$\tilde{P} = A [R|\mathbf{t}] \quad (3.42)$$

### 3.4.3 Calibração de uma câmera - método alternativo

Outro método de calibração calcula a matriz de projeção diretamente a partir dos pontos no ambiente cujas coordenadas são conhecidas, no sistema de coordenadas do ambiente e no sistema de coordenadas da imagem. Este método pode ser tratado em duas etapas:

- Estimar a matriz de projeção  $\tilde{P}$ , e
- Decompor a matriz de projeção em matrizes de parâmetros internos e externos  $A$ ,  $R$  e  $\mathbf{t}$ .

Para a maioria das aplicações de estereoscopia esta segunda etapa não é necessária, a menos que seja necessária alguma transformação do sistema de coordenadas do ambiente após a calibração.

Da mesma forma que no método de Tsai, as câmeras são calibradas uma de cada vez e não são utilizadas aqui as restrições da geometria epipolar característica do conjunto binocular.

Dados  $n$  pontos cujas posições são conhecidas,  $\{(x_{wi}, y_{wi}, z_{wi}) | i = 1, \dots, n\}$  e suas projeções no plano de imagem correspondentes,  $\{(u_i, v_i) | i = 1, \dots, n\}$ , a matriz de projeção  $\tilde{P}$ , da equação (3.15), que transforma pontos do espaço tridimensional em coordenadas de imagem pode ser reescrita como [15]:

$$\begin{bmatrix} \lambda u_i \\ \lambda v_i \\ \lambda \end{bmatrix} = \tilde{P} \begin{bmatrix} x_{wi} \\ y_{wi} \\ z_{wi} \\ 1 \end{bmatrix} \quad \text{ou} \quad \begin{bmatrix} \lambda u_i \\ \lambda v_i \\ \lambda \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} x_{wi} \\ y_{wi} \\ z_{wi} \\ 1 \end{bmatrix} \quad (3.43)$$

Como  $\lambda$  é um fator de escala um dos elementos da matriz de projeção, por exemplo  $p_{34}$ , pode ser igualado a 1 sem perda de generalidade e, desta forma, restam somente onze parâmetros a serem determinados no processo de calibração.

A equação (3.43) pode ser expandida em:

$$\begin{cases} p_{11}x_{wi} + p_{12}y_{wi} + p_{13}z_{wi} + p_{14} = \lambda u_i \\ p_{21}x_{wi} + p_{22}y_{wi} + p_{23}z_{wi} + p_{24} = \lambda v_i \\ p_{31}x_{wi} + p_{32}y_{wi} + p_{33}z_{wi} + 1 = \lambda \end{cases} \quad (3.44)$$

Substituindo a última equação nas duas primeiras e eliminando  $\lambda$  temos:

$$\begin{cases} p_{11}x_{wi} + p_{12}y_{wi} + p_{13}z_{wi} + p_{14} - p_{31}x_{wi}u_i - p_{32}y_{wi}u_i - p_{33}z_{wi}u_i = u_i \\ p_{21}x_{wi} + p_{22}y_{wi} + p_{23}z_{wi} + p_{24} - p_{31}x_{wi}v_i - p_{32}y_{wi}v_i - p_{33}z_{wi}v_i = v_i \end{cases} \quad (3.45)$$

As incógnitas  $m_{ij}$  podem ser organizadas na forma de uma 11-tupla:

$$\begin{bmatrix} x_{w1} & y_{w1} & z_{w1} & 1 & 0 & 0 & 0 & 0 & -x_{w1}u_1 & -y_{w1}u_1 & -z_{w1}u_1 \\ 0 & 0 & 0 & 0 & x_{w1} & y_{w1} & z_{w1} & 1 & -x_{w1}v_1 & -y_{w1}v_1 & -z_{w1}v_1 \\ \dots & \dots \\ x_{wi} & y_{wi} & z_{wi} & 1 & 0 & 0 & 0 & 0 & -x_{wi}u_i & -y_{wi}u_i & -z_{wi}u_i \\ 0 & 0 & 0 & 0 & x_{wi} & y_{wi} & z_{wi} & 1 & -x_{wi}v_i & -y_{wi}v_i & -z_{wi}v_i \\ \dots & \dots \\ x_{wn} & y_{wn} & z_{wn} & 1 & 0 & 0 & 0 & 0 & -x_{wn}u_n & -y_{wn}u_n & -z_{wn}u_n \\ 0 & 0 & 0 & 0 & x_{wn} & y_{wn} & z_{wn} & 1 & -x_{wn}v_n & -y_{wn}v_n & -z_{wn}v_n \end{bmatrix} \begin{bmatrix} p_{11} \\ p_{12} \\ p_{13} \\ p_{14} \\ p_{21} \\ p_{22} \\ \vdots \\ p_{33} \end{bmatrix} = \begin{bmatrix} u_1 \\ v_1 \\ \vdots \\ u_i \\ v_i \\ \vdots \\ u_n \\ v_n \end{bmatrix} \quad (3.46)$$

A equação (3.46) é um sistema de equações lineares na forma

$$Ax = b \quad (3.47)$$

onde  $A$  é uma matriz  $2n \times 11$ ,  $x$  é a matriz de projeção representada como um vetor e  $b$  é um vetor  $2n$ . A solução de  $x$  e, logo,  $\tilde{P}$ , pode ser obtida por otimização por mínimos quadrados.

Com onze incógnitas e cada ponto conhecido fornecendo duas equações ao sistema são necessários pelo menos seis pontos para calcular  $\tilde{P}$ .

### 3.4.4 Calibração de duas câmeras - geometria epipolar

Segundo Zhang *et al.* [37], os métodos de calibração anteriores, quando aplicados a um conjunto binocular, apresentam dois inconvenientes:

- Não está sendo explorado o fato de que os pontos presentes em uma das imagens possuem correspondentes na outra imagem, o que reduz a precisão na estimativa da geometria epipolar; e
- A informação de calibração é válida somente no volume em torno da posição dos pontos conhecidos utilizados na calibração, geralmente pontos pertencentes a um gabarito de calibração; a qualidade da calibração mostra-se degradada quando utilizada fora da região então ocupada pelo gabarito de calibração.

Estes inconvenientes têm origem no fato de que se está trabalhando com o ajuste do modelo de câmera a um conjunto limitado de dados, o que nos leva a somente uma aproximação da geometria real da câmera. O ajuste obtido é válido onde existem dados de calibração mas não na região do espaço onde não há dados.

Zhang *et al.* [37] propõe que pontos correspondentes do ambiente sejam adicionados ao processo de calibração, além dos pontos conhecidos, provenientes do gabarito de calibração. Também propõe que, sendo rígida a geometria do conjunto binocular, sejam incluídos no processo de calibração pontos correspondentes tomados em vários instantes diferentes, ou seja, tomados de vários pares de imagens.

Dispondo-se da matriz fundamental pode-se obter duas matrizes de projeção compatíveis com esta matriz fundamental. E, ainda, conhecendo-se a posição no espaço e nas duas imagens de alguns pontos correspondentes, pode-se calcular a matriz de distorção que permitirá a reconstrução dos pontos no sistema de coordenadas do ambiente.

O método de Zhang *et al.* consiste em cinco etapas:

1. Estimar a geometria epipolar a partir de todos os pontos correspondentes disponíveis;
2. Calcular as matrizes de projeção das câmeras compatíveis com a geometria epipolar;
3. Calcular a reconstrução dos pontos cuja posição no sistema de coordenadas do ambiente é conhecida;
4. Estimar a matriz de distorção entre a estrutura projetiva e a estrutura euclidiana;

5. Calcular as matrizes de projeção das câmeras em relação ao sistema de coordenadas do ambiente.

Tomando a equação (3.23) e dispondo-se de  $n$  pontos correspondentes nas imagens das duas câmeras, cada correspondência  $U_{1i} \leftrightarrow U_{2i}$  gera uma restrição em  $F$ :

$$[u_{2i}, v_{2i}, 1] \begin{bmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{bmatrix} \begin{bmatrix} u_{1i} \\ v_{1i} \\ 1 \end{bmatrix} = 0 \quad (3.48)$$

A equação (3.48) pode ser reescrita como  $Af = \mathbf{0}$ , onde  $A$  é uma matriz  $n \times 9$ , e  $f$  é a matriz fundamental representada como um vetor [15]:

$$\begin{bmatrix} u_{21}u_{11} & u_{21}v_{11} & u_{21} & v_{21}u_{11} & v_{21}v_{11} & v_{21} & u_{11} & v_{11} & 1 \\ \dots & \dots \\ u_{2i}u_{1i} & u_{2i}v_{1i} & u_{2i} & v_{2i}u_{1i} & v_{2i}v_{1i} & v_{2i} & u_{1i} & v_{1i} & 1 \\ \dots & \dots \\ u_{2n}u_{1n} & u_{2n}v_{1n} & u_{2n} & v_{2n}u_{1n} & v_{2n}v_{1n} & v_{2n} & u_{1n} & v_{1n} & 1 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ \vdots \\ f_9 \end{bmatrix} = \mathbf{0} \quad (3.49)$$

Com oito ou mais pontos correspondentes pode-se determinar  $f$ , sujeita a um fator de escala.

Geralmente centenas de pontos correspondentes estão disponíveis para calibração, porém incluindo ruído e falsas correspondências. Nesta situação uma solução pode ser obtida por

$$\min_f \|Af\|^2 \quad \text{sujeito a} \quad \|f\| = 1 \quad (3.50)$$

Este é um problema de álgebra linear, cuja solução é o autovetor associado ao menor autovalor de  $A^T A$ . Esta solução pode ser facilmente obtida através da decomposição por valores singulares (SVD), o que é conveniente por não ser iterativo ao mesmo tempo que traz o inconveniente de ser extremamente sensível a ruído [11].

A origem do sistema de coordenadas da imagem está posicionada no canto superior esquerdo da imagem (3.7), o que favorece a homogeneidade das coordenadas dos pontos e tem efeitos sobre a estabilidade do sistema da matriz da equação (3.49). É importante,

então, que os dados de calibração disponíveis sejam normalizados, o que pode ser obtido pela aplicação de uma transformação nas coordenadas de calibração:

$$\begin{aligned}\tilde{\mathbf{u}}_{1i} &= H_1 \mathbf{u}_{1i} \\ \tilde{\mathbf{u}}_{2i} &= H_2 \mathbf{u}_{2i}\end{aligned}\quad (3.51)$$

onde  $H_1$  e  $H_2$  são matrizes  $3 \times 3$  de forma que as coordenadas transformadas tenham média zero e desvio padrão unitário. Dispondo da média  $\bar{u}$  e desvio padrão  $\sigma_u$  das coordenadas originais dos pontos podemos calcular  $H$  como:

$$H = \begin{bmatrix} 1/\sigma_u & 0 & -\bar{u}/\sigma_u \\ 0 & 1/\sigma_v & -\bar{v}/\sigma_v \\ 0 & 0 & 1 \end{bmatrix}\quad (3.52)$$

A matriz da equação (3.49) pode agora ser construída utilizando-se os valores normalizados  $\tilde{\mathbf{u}}_{1i}$  e  $\tilde{\mathbf{u}}_{2i}$  e levando à matriz fundamental  $\tilde{F}$ . A matriz fundamental  $F$  é obtida de

$$F = H_2^T \tilde{F} H_1\quad (3.53)$$

Além do método linear proposto acima podemos utilizar um procedimento de otimização não linear minimizando o quadrado (método LS, do inglês, *least squares*) da distância perpendicular média dos pontos de calibração até suas linhas epipolares (figura 3.11):

$$\min_f \sum_i (d_{\perp 1i}^2 + d_{\perp 2i}^2)\quad (3.54)$$

onde

$$(d_{\perp 1i}^2 + d_{\perp 2i}^2) = (\tilde{\mathbf{u}}_{2i}^T F \tilde{\mathbf{u}}_{1i})^2 \left( \frac{1}{(F \tilde{\mathbf{u}}_{1i})_1^2 + (F \tilde{\mathbf{u}}_{1i})_2^2} + \frac{1}{(F^T \tilde{\mathbf{u}}_{2i})_1^2 + (F^T \tilde{\mathbf{u}}_{2i})_2^2} \right)\quad (3.55)$$

No processo de minimização podemos utilizar como estimativa inicial a matriz fundamental calculada pelo linear.

Se os pontos correspondentes nas duas imagens foram determinados automaticamente é possível que existam falsas correspondências. Estas falsas correspondências influenciam a estimativa da geometria epipolar e, para reduzir esta influência, podemos utilizar um método mais robusto de estimativa.

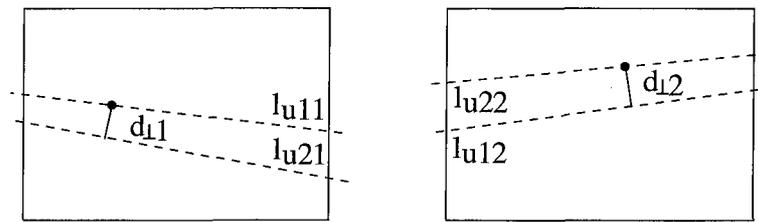


Figura 3.11: Distância perpendicular da linha epipolar

Gracias e Santos-Victor [11] citam o método LMedS (do inglês, *least median of squares*) como um dos mais eficientes para estimativa da matriz fundamental a partir de um conjunto de pontos correspondentes contendo falsas correspondências. Neste método os parâmetros são estimados minimizando-se a mediana do quadrado dos erros para o conjunto total de pontos correspondentes disponíveis. Como centenas de pontos podem estar sendo utilizados para estimar  $F$  é usual restringir a análise a um subconjunto aleatório de pares correspondentes.

Gracias e Santos-Victor [11] propõem ainda que, após uma primeira estimativa de  $F$  usando o método LMedS, o conjunto de pontos correspondentes seja reduzido, ignorando-se os pares menos compatíveis com a matriz  $F$  calculada. Então, após a redução, todos os pares restantes seriam utilizados para estimar uma nova matriz  $F$  utilizando novamente o método LS apresentado anteriormente.

A matriz fundamental  $F$ , representando a geometria epipolar de um conjunto binocular, pode ser fatorizada em uma matriz antissimétrica  $[e_2]_{\times}$  e uma matriz  $M$ , tal que:

$$F = [e_2]_{\times} M \quad (3.56)$$

onde  $e_2$  é o epipolo na segunda imagem.

A fatorização de  $F$  em  $[e_2]_{\times}$  e  $M$  geralmente não é única, pois se  $M$  é uma solução, então  $M + e_2 \mathbf{v}^T$  também é uma solução, para qualquer vetor  $\mathbf{v}$ .

Uma vez que  $F^T e_2 = \mathbf{0}$ , o epipolo na segunda imagem pode ser obtido decompondo-se  $FF^T$  por valores singulares (SVD), tomando-se o autovetor associado ao menor autovalor.

Utilizando-se a relação:

$$\|\mathbf{v}\|^2 I_3 = \mathbf{v}\mathbf{v}^T - [\mathbf{v}]_{\times}^2 \quad \text{para } \forall \mathbf{v} \quad (3.57)$$

podemos escrever [37]:

$$F = \frac{1}{\|\mathbf{e}_2\|^2} (\mathbf{e}_2 \mathbf{e}_2^T - [\mathbf{e}_2]_{\times}^2) F = \frac{1}{\|\mathbf{e}_2\|^2} \underbrace{\mathbf{e}_2 \mathbf{e}_2^T F}_0 + [\mathbf{e}_2]_{\times} \underbrace{\left( -\frac{[\mathbf{e}_2]_{\times}}{\|\mathbf{e}_2\|^2} F \right)}_0 \quad (3.58)$$

A matriz  $M$  pode ser tomada então como:

$$M = -\frac{[\mathbf{e}_2]_{\times}}{\|\mathbf{e}_2\|^2} F \quad (3.59)$$

Uma representação canônica pode agora ser utilizada para duas matrizes de projeção compatíveis com  $F$  [37]:

$$\begin{aligned} \mathbb{P}_1 &= [ I \quad \mathbf{0} ] \\ \mathbb{P}_2 &= [ M \quad \mathbf{e}_2 ] \end{aligned} \quad (3.60)$$

Este mesmo método é citado por Rothwell *et al.* [25] e Gracias e Santos-Victor [11], onde se pode obter uma descrição mais detalhada.

Utilizando as matrizes  $\mathbb{P}_1$  e  $\mathbb{P}_2$  podemos obter uma reconstrução projetiva  $\tilde{\mathbf{x}}_i = [x_i, y_i, z_i, t_i]^T$  ( $i = 1, \dots, n$ ) dos pontos correspondentes nas duas imagens do conjunto binocular. Uma reconstrução métrica não é possível pois os parâmetros internos das câmeras são desconhecidos.

Tomando-se somente os pontos correspondentes para os quais se conhece a posição no sistema de coordenadas do ambiente podemos determinar a matriz dos parâmetros internos. O procedimento consiste em calcular a matriz de distorção projetiva  $\mathbb{D}$  que transforma as coordenadas obtidas na reconstrução projetiva pelas matrizes  $\mathbb{P}_1$  e  $\mathbb{P}_2$  nas coordenadas de ambiente conhecidas:

$$\lambda_i \begin{bmatrix} x_i \\ y_i \\ z_i \\ t_i \end{bmatrix} = \mathbb{D} \begin{bmatrix} x_{wi} \\ y_{wi} \\ z_{wi} \\ 1 \end{bmatrix} \quad (3.61)$$

ou

$$\lambda_i \tilde{\mathbf{x}}_i = \mathbb{D} \mathbf{P}_{wi} \quad (3.62)$$

onde  $\lambda_i$  é um fator de escala.

Tomando  $\mathbf{v} = \mathbb{D} \mathbf{P}_{wi} = [v_1, v_2, v_3, v_4]^T$ , temos então três equações independentes [37]:

$$\begin{aligned} y_i v_1 - x_i v_2 &= 0 \\ z_i v_1 - x_i v_3 &= 0 \\ t_i v_1 - x_i v_4 &= 0 \end{aligned} \quad (3.63)$$

Representando a matriz de distorção  $\mathbb{D}$  como um vetor dos 16 parâmetros  $\mathbf{x} = [D_{11}, D_{12}, \dots, D_{44}]^T$  Zhang *et al.* [37] citam que a equação (3.63) é equivalente a

$$B_i \mathbf{x} = \mathbf{0} \quad (3.64)$$

onde

$$B_i = \begin{bmatrix} y_i \mathbf{P}_{wi} & -x_i \mathbf{P}_{wi} & \mathbf{0} & \mathbf{0} \\ z_i \mathbf{P}_{wi} & \mathbf{0} & -x_i \mathbf{P}_{wi} & \mathbf{0} \\ t_i \mathbf{P}_{wi} & \mathbf{0} & \mathbf{0} & -x_i \mathbf{P}_{wi} \end{bmatrix} \quad (3.65)$$

Dados  $m$  pontos correspondentes  $(\mathbf{P}_{wi}, \tilde{\mathbf{x}}_i)$ , temos  $m$  equações como a da equação (3.64), levando a uma matriz  $B$  de dimensões  $3m \times 16$ . O vetor  $\mathbf{x}$  pode ser determinado por decomposição por valores singulares (SVD), tomando-se o autovetor associado ao menor autovalor de  $B^T B$ .

Finalmente as matrizes de projeção das câmeras, seguindo o modelo *pinhole*, pode ser obtido por [37]:

$$\begin{aligned} \tilde{P}_1 &= P_1 \mathbb{D} \\ \tilde{P}_2 &= P_2 \mathbb{D} \end{aligned} \quad (3.66)$$

### 3.5 Resumo

Neste capítulo descrevemos o modelo de câmeras *pinhole* utilizado como uma aproximação das câmeras com lentes utilizadas em aplicações reais e toda a geometria da projeção em perspectiva associada. Também foi descrita a geometria epipolar que caracteriza as relações geométricas entre as duas câmeras do conjunto binocular.

Foram estudados métodos de reconstrução dos pontos do ambiente, dispondo-se das

coordenadas deste ponto nas imagens e da geometria das câmeras utilizadas. Para os casos onde inicialmente esta geometria não é conhecida estudou-se métodos para sua determinação, o que é chamado de calibração das câmeras.

Dois destes métodos de calibração estarão sendo avaliados no próximo capítulo, juntamente com a correção da distorção radial.

## Capítulo 4

# Calibração de um conjunto binocular simplificado

*"The least deviation from the truth is multiplied later."*

*Aristóteles*

Neste trabalho foram conduzidos vários tipos de experimentos, procurando identificar o método de calibração mais adequado ao problema, que estão separados em quatro fases descritas nas próximas seções, após uma descrição do conjunto binocular utilizado.

### 4.1 O conjunto binocular utilizado

#### 4.1.1 As câmeras

As câmeras utilizadas no conjunto binocular são minicâmeras coloridas Kalpyys modelo KAL-A1012N, utilizadas principalmente em circuitos de monitoração de segurança. Foram obtidas as especificações do fabricante para um modelo de câmera similar, modelo KAL-1012N. Além destas duas minicâmeras uma terceira câmera VHS compacta colorida Minolta modelo C-560C foi usada durante os testes de calibração. As especificações destes dois tipos de câmeras estão resumidas na tabela 4.1. A tabela 4.2 mostra a identificação mnemônica que será utilizada para cada uma das três câmeras ao longo deste trabalho.

Tabela 4.1: Especificações técnicas das câmeras utilizadas

Parâmetro	Modelo	
	KAL-1012N	C-560C
Fabricante	Kalpyys	Minolta
Tipo de sinal	Composto	Composto
Padrão	NTSC	NTSC
Campo de visão	—	—
Tipo de sensor	CCD	CCD
Formato do sensor de imagem (in)	1/3	1/4
Elementos sensores (pel)	542(h) x 492(v)	—
Distância focal (mm)	3,6 fixo	4,2 até 50,4 autom.
Balanço de branco	Automático	Automático
Potência de alimentação (V)	12	6
Consumo de potência (W)	3	5,9
Iluminação mínima (lux)	1,5	—
Dimensões externas (mm)	72.0 (l) x 75.0 (h) x 55.0 (p)	107.0 (l) x 112.0 (h) x 165.0 (p)

Tabela 4.2: Identificação das câmeras utilizadas

Identificação	Modelo	Número de série
KAL-233	Kalpyys KAL-A1012N	990802233
KAL-240	Kalpyys KAL-A1012N	990802240
MIN-077	Minolta C-550C	14852077

#### 4.1.2 O cartão de aquisição

O sinal de vídeo foi convertido nas imagens digitais através de um cartão de aquisição Prolink modelo PixelView PlayTVpro acoplado ao barramento PCI do microcomputador utilizado na calibração. Este cartão, de baixo custo, é geralmente utilizado por usuários domésticos. As especificações deste cartão estão resumidas na tabela 4.3.

Para referência algumas informações sobre os padrões de codificação do sinal de vídeo suportados pelo cartão de aquisição e também pelas câmeras utilizadas estão resumidas na tabela 4.4.

Na seção 3.1.2 foi introduzida a idéia de sincronização na aquisição das imagens. Um aspecto importante do cartão de aquisição utilizado é que, por ser projetado para uso doméstico e diferentemente do que costuma ocorrer com os cartões de aquisição para uso in-

Tabela 4.3: Especificações técnicas do cartão de aquisição

Parâmetro	
Modelo	PixelView PlayTVpro
Fabricante	Prolink
Tipo de sinal	Composto ou Y-C (S-Video)
Padrão	NTSC, PAL ou SECAM
Microprocessador	Brooktree/Conexant BT878
Taxa de aquisição máxima	30 quadros por segundo
Espaço de cor	YCrCb e RGB
Conexão ao hospedeiro	PCI rev 2.1

Tabela 4.4: Especificações dos padrões de codificação  
(extraída de Jonker [16])

Parâmetro	Padrão		
	NTSC	PAL	SECAM
Norma	RS-170A	—	—
Imagens por segundo	29,97	25	25
Tempo por imagem (ms)	33,37	40	40
Linhas por imagem	525	625	625
Razão de aspecto (h/v)	4:3	4:3	4:3
Entrelaçamento	2:1	2:1	2:1
Tempo por linha ( $\mu$ s)	63,56	64,00	64,00

dustrial, não há a possibilidade de sincronização externa, ou seja, a única informação de sincronização é a marcação de início de quadro presente no sinal de vídeo proveniente das câmeras. Além disso as imagens das duas câmeras não podem ser sincronizadas.

Outro aspecto importante, tanto das câmeras quanto do cartão de aquisição, é que ambos estão limitados à taxa de imagens por segundo determinada pelo padrão de codificação de vídeo utilizado, aproximadamente 30 imagens por segundo para o caso do NTSC.

#### 4.1.3 O conjunto binocular

Algumas lentes permitem ajustes na exposição, focalização e aproximação ao mesmo tempo que alguns conjuntos binoculares permitem movimentos independentes das câmeras podendo alterar a distância e rotação relativa entre as elas. No conjunto binocular utilizado neste trabalho nenhuma destas facilidades estão disponíveis, daí ele ter sido introduzido

como simplificado.

O conjunto binocular não possui qualquer controle motorizado sobre as lentes ou sobre a posição e orientação das câmeras, o que diminui, mas não elimina, a importância da calibração a partir do ambiente durante a operação normal do conjunto binocular. Assume-se que o conjunto binocular possa estar montado sobre um robô móvel e que possa ocorrer impacto, vibração ou variações de temperatura que venham a provocar pequenas alterações na configuração do conjunto.

A configuração exata do conjunto binocular, como distância entre os centros de projeção e a orientação de cada câmera, será conhecida após a calibração. Na figura 4.1 temos uma imagem do conjunto binocular obtida através da câmera MIN-077.

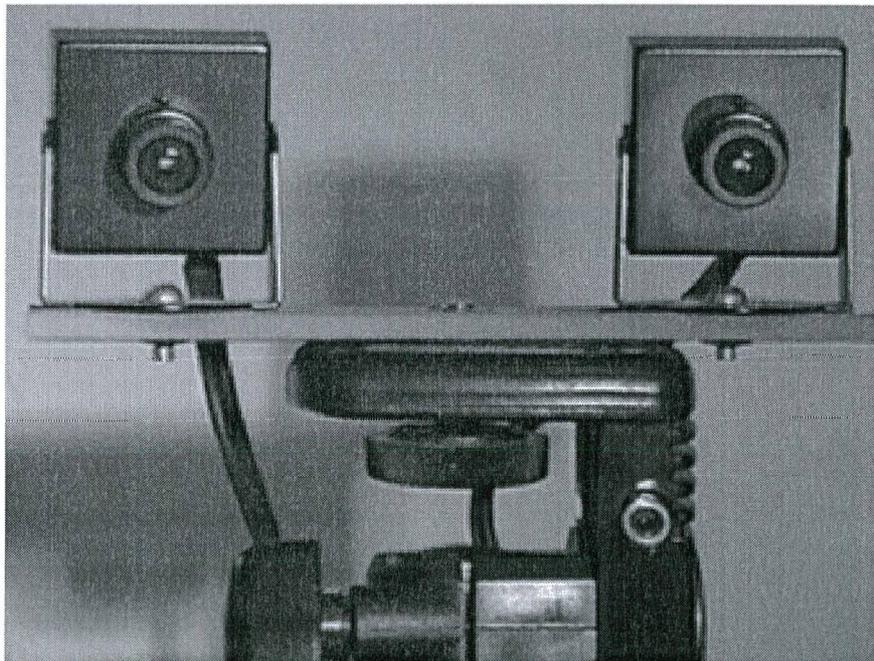


Figura 4.1: Conjunto binocular simplificado

Observando a figura 4.1 pode-se perceber a rigidez da configuração do conjunto. Considera-se que, se montado sobre um robô móvel, o conjunto será dotado de pelo menos o movimento de elevação, ou seja, rotação em torno do eixo horizontal, que lhe permitiria observar objetos mais próximos do robô aumentando o ângulo entre o solo e os eixos óticos e objetos mais distantes diminuindo este ângulo.

Como exemplos de outros conjuntos binoculares podemos citar:

- Huynh [15] que utiliza um par de câmeras Pulnix montadas sobre uma plataforma

que utiliza três motores de passo; dois motores controlam a rotação de cada uma das câmeras em torno do eixo vertical (vergência), e o terceiro aumenta ou diminui simetricamente a distância entre os centros de projeção; as imagens são obtidas através de um cartão de aquisição Imaging Technology modelo PCVisionplus projetado especificamente para uso em visão computacional que possui quatro entradas analógicas de vídeo monocromático.

- Santos-Victor *et al.* [27] utiliza um conjunto binocular com 4 graus de liberdade composto de duas câmeras monocromáticas onde se pode controlar a vergência de cada câmera independentemente; além do controle de vergência o conjunto dispõe de controle motorizado para elevação (acima e abaixo) e rotação (esquerda e direita) e controle manual da distância entre os centros de projeção; segundo Santos-Victor *et al.* [27] os movimentos disponíveis no conjunto binocular fornecem os principais graus de liberdade existentes no sistema ocular humano; além das câmeras compõem o conjunto um cartão de aquisição e um multiplexador de sinal de vídeo; as câmeras são controladas por motores de corrente contínua acoplados a *encoders*;
- Christensen [5] utiliza um conjunto binocular com 12 graus de liberdade que inclui duas lentes com ajuste controlável de aproximação, abertura e foco; o conjunto pode ainda controlar independentemente a rotação de cada câmera e a rotação e elevação de todo o conjunto.

## 4.2 Detecção e correspondência de caracteres

Neste estudo dois tipos de caracteres são utilizados: (1) regiões e (2) vértices. As regiões são utilizadas somente na calibração com gabarito, onde a posição de cada região no espaço é conhecida e, se todas as regiões aparecem na imagem, então a correspondência pode ser obtida imediatamente ordenando-se as regiões por linha e coluna da imagem. A correspondência entre os caracteres nas duas imagens é obtida a partir da correspondência com o gabarito. O mesmo procedimento se aplica se os vértices são utilizados na calibração com gabarito. Os vértices também são utilizados na calibração a partir das imagens do ambiente e a correspondência neste caso é menos simples pois não há informação sobre a posição dos caracteres no espaço e o ambiente é muito menos estruturado que o gabarito de calibração. Neste caso também é necessário obter diretamente a correspondência dos caracteres entre as imagens.

Nas próximas seções veremos a detecção de regiões, de vértices e a correspondência dos vértices entre duas imagens utilizando dois métodos diferentes.

#### 4.2.1 Detecção de regiões

A detecção de regiões é utilizada na calibração com gabarito, o qual consiste em um conjunto de quadrados posicionados regularmente ao longo de dois planos perpendiculares (ver imagem 4.2). O número de quadrados posicionados horizontal e verticalmente é conhecido para cada plano, assim como a distância entre os quadrados e todas as demais informações necessárias para se chegar à posição de cada região no espaço. O mais importante aqui é que os quadrados possuem alto contraste com o fundo do gabarito, o que facilita a segmentação das regiões pela aplicação de um limiar de corte na imagem.

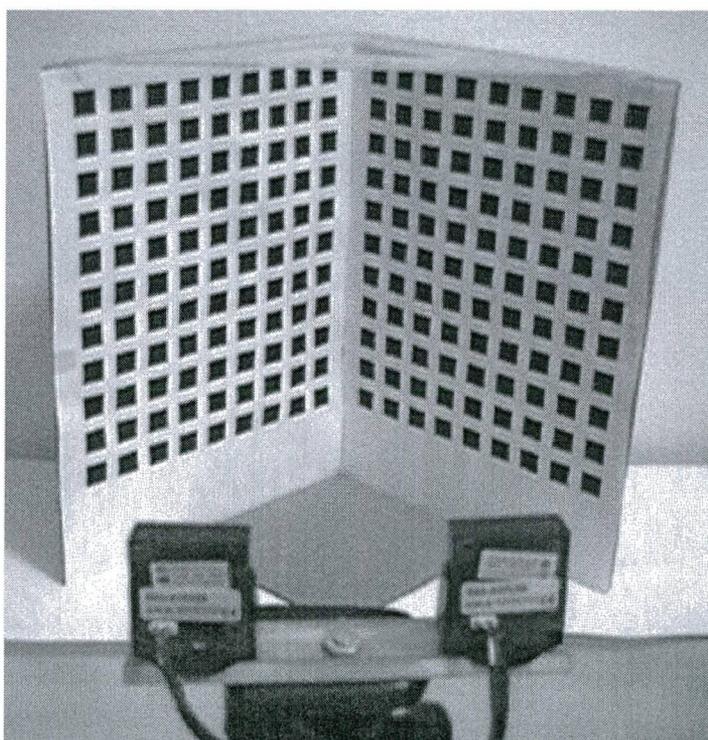


Figura 4.2: Gabarito não coplanar de calibração

Em especial estamos interessados em duas informações sobre a região detectada: (1) quadratura e (2) centróide. A primeira para que se possa selecionar somente as regiões que são os quadrados do gabarito e a segunda para relacionar as regiões na imagem com os quadrados no espaço.

O procedimento de busca de regiões do gabarito pode ser resumido como:

1. Converter imagens coloridas em imagem monocromática utilizando o operador de luminosidade RGB;
2. Aplicar limiar de corte na intensidade obtendo uma imagem binível; valores típicos de limiar de corte utilizados estão no intervalo 100-120;
3. Detectar regiões na imagem binível usando expansão de regiões;
4. Selecionar regiões usando quadratura; valores típicos de limiar de quadratura utilizados estão no intervalo 0,25-0,40;
5. Eliminar regiões excedentes;
6. Calcular centróide de cada região;
7. Ordenar regiões por linha e por coluna de acordo com a posição do centróide;
8. Relacionar centróides das regiões com as posições conhecidas do gabarito.

A figura 4.3 mostra um segmento de uma imagem do gabarito obtida através da câmera KAL-233. Para maior clareza deste exemplo, a imagem foi ampliada para o dobro de sua resolução espacial e foi tomado somente um segmento do canto inferior esquerdo de tamanho  $160 \times 120 \text{ pel}$  (figura 4.3a). A figura 4.3b mostra o resultado até após o passo 2 do procedimento acima e a figura 4.3c representa a informação final obtida da imagem, que é o centróide de cada região associada ao gabarito. A tabela 4.5 mostra os dados intermediários calculados para as regiões desta imagem.

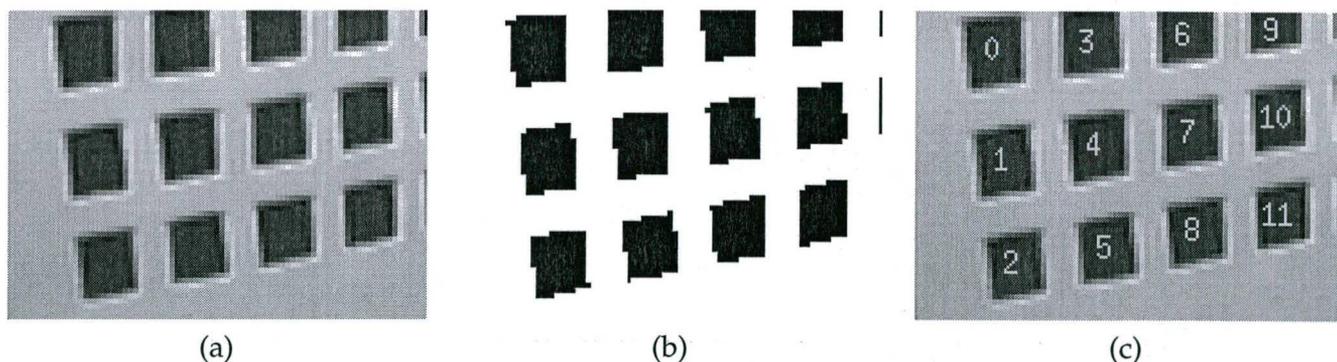


Figura 4.3: Segmento da imagem do gabarito (a), imagem após aplicação do limiar de corte (b) e após detecção das regiões(c)

Numa imagem completa do gabarito, o número de regiões detectadas varia de acordo com as condições de iluminação, do limiar de corte e da ocupação da imagem pelo gabarito. O gabarito mostrado na figura 4.2 possui 198 quadrados e este é o número final de regiões a serem selecionadas na imagem.

Tabela 4.5: Medidas obtidas sobre as regiões detectadas

Reg.	$x_{max}$	$y_{max}$	$x_{min}$	$y_{min}$	$A$	$Q$	$\sum_x x$	$\sum_y y$	$\bar{x}$	$\bar{y}$
0	39	29	17	2	550	0.0800	15952	8063	29	14
1	43	71	23	44	506	0.0672	16770	29137	33	57
2	49	111	27	86	480	0.1458	17956	46944	37	97
3	76	23	56	0	488	0.0574	32104	5436	65	11
4	78	65	58	40	490	0.0204	33508	25365	68	51
5	82	105	62	78	452	0.1947	32642	41322	72	91
6	111	19	91	0	372	0.0215	37692	3222	101	8
7	113	59	93	34	444	0.1261	45998	20638	103	46
8	115	97	95	72	458	0.0917	48418	38773	105	84
9	144	13	126	0	250	0.0640	33662	1529	134	6
10	146	53	128	28	420	0.0714	57322	17018	136	40
11	146	91	129	66	404	0.0520	55534	31538	137	78
Inválida	159	119	0	0	13854	0.3657	1093754	872567	78	62
Inválida	159	9	159	0	10	1.0000	1590	45	159	4
Inválida	159	47	159	26	22	1.0000	3498	803	159	36

#### 4.2.2 Detecção de vértices

A detecção de vértices é utilizada tanto na calibração com gabarito quanto na calibração a partir das imagens do ambiente. Neste trabalho utilizamos o operador de Harris para a detecção dos vértices, o que já foi introduzido na seção 2.1.

O procedimento de detecção de vértices pode ser resumido como:

1. Converter imagens coloridas em imagem monocromática utilizando o operador de luminosidade RGB;
2. Aplicar o operador de Harris sobre os píxeis da imagem obtendo uma imagem que é a função de resposta de vértice;
3. Determinar máximo local na 8-vizinhança de cada píxel da imagem da função de resposta de vértice;
4. Aplicar limiar de corte sobre a função de resposta de vértice desprezando valores abaixo do limiar; valor típico do limiar de corte utilizado neste trabalho está no intervalo de 400-500 para imagens do ambiente e 15000 para imagens do gabarito;
5. Relacionar vértices detectados e ordenar pelo valor da função resposta de vértice;
6. Percorrer relação de vértices em ordem decrescente, desprezando vértices com distâncias menores que um limiar pré-definido em relação a qualquer outro vértice de

maior valor; valor típico da distância mínima é  $10pel$ ; a distância mínima voltará a ser discutida na seção 4.2.3;

7. Finalmente seleciona-se somente os melhores vértices, ou seja aqueles que possuem maior valor da função de resposta de vértice; valor típico do número máximo de vértices selecionados utilizado neste trabalho é 300.

Na figura 4.4 são mostrados os resultados intermediários e o resultado final da aplicação do operador Harris na mesma imagem da figura 4.3a, enquanto a 4.5 mostra os mesmos resultados para uma imagem obtida do ambiente também através da câmeras KAL-233.

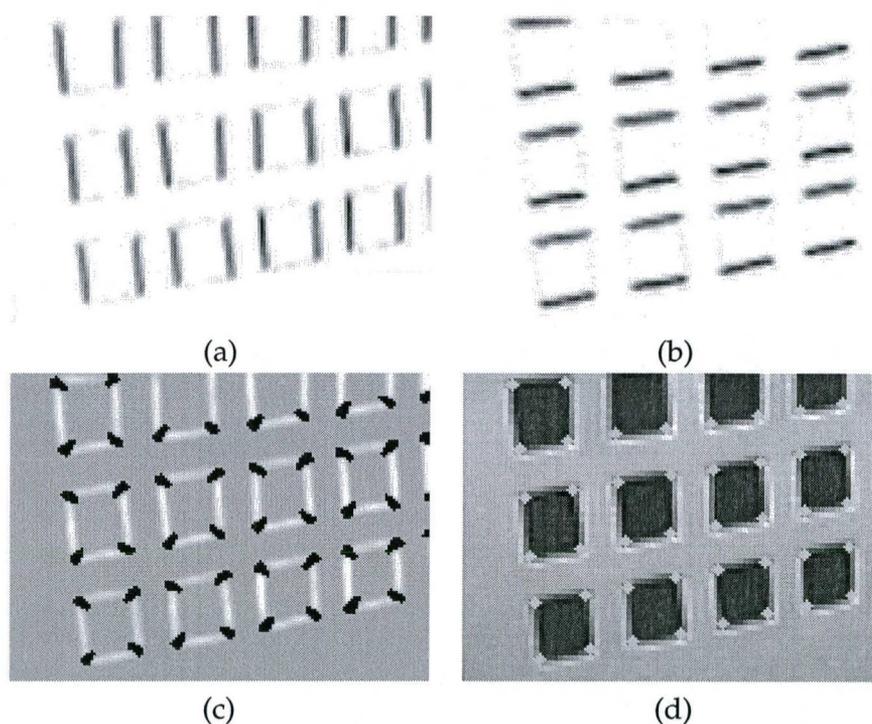


Figura 4.4: Gradiente em x do segmento da imagem do gabarito (a), gradiente em y (b), função de resposta de vértice (c) e vértices selecionados (d)

Como um complemento do vértice detectado, especialmente no caso de imagens do ambiente, podemos utilizar ainda um indicador da textura na vizinhança do vértice detectado, como é o caso do indicador de bloco distinguível também introduzido na seção 2.1 ou ainda somatória da intensidade na vizinhança do vértice suavizada por interpolação bilinear. Na tabela 4.6 estão relacionadas as posições de cada vértice detectado na figura 4.5, o valor da função de resposta de vértice  $V$ , o indicador de bloco distinguível  $K$  obtido com um tamanho de bloco de  $7 \times 7pel$  e a somatória da intensidade  $I$  no bloco de mesmo tamanho.

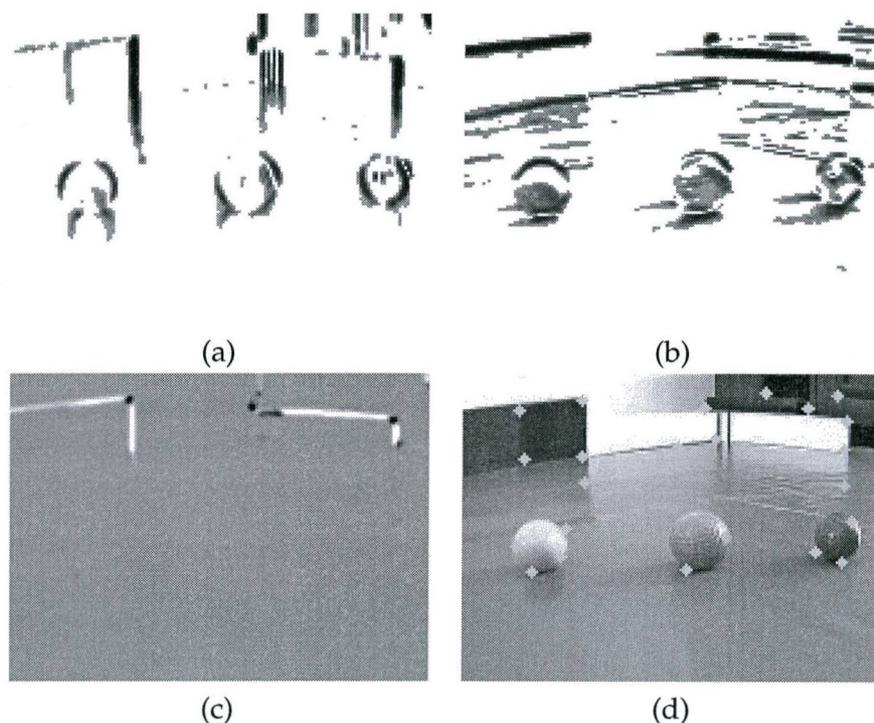


Figura 4.5: Gradiente em x da imagem do ambiente (a), gradiente em y (b), função de resposta de vértice (c) e vértices selecionados (d)

### 4.2.3 Correspondência de caracteres

Foram avaliados dois métodos de correspondência: correspondência por relaxação, discutidos na seção 2.3 e por valores singulares [22]. Como já citado no início desta seção, a correspondência entre os caracteres nas duas imagens só é utilizada na calibração a partir das imagens do ambiente, pois no caso da calibração com gabarito a correspondência já é conhecida *a priori*.

Em ambos os métodos de correspondência avaliados os caracteres utilizados são os vértices detectados na imagem. A correspondência é feita a partir de quatro valores obtidos de cada caractere durante a detecção: (1) posição  $P$  do caractere no sistema de coordenadas do sensor, (2) valor da função resposta de vértice ( $V$ ), (3) valor da função de caractere distinguível ( $K$ ) e (4) valor acumulado de intensidade ( $I$ ). Tanto na função de caractere distinguível quanto na somatória da intensidade foi utilizado um bloco de  $11 \times 11 pel$ .

Durante a detecção dos caracteres pode-se aplicar uma distância mínima entre os caracteres selecionados. Nos testes iniciais foi utilizada uma distância mínima de  $10 pel$  entre os caracteres, porém observou-se grande ocorrência de erros de correspondência, possivelmente proveniente do fato de que os caracteres correspondentes haviam sido desprezados

Tabela 4.6: Medidas obtidas sobre os vértices detectados

Vért.	$x$	$y$	$V$	$K$	$I$
0	146	18	1418406	34292	6332
1	92	13	1117474	26343	8006
2	45	10	1085258	35921	8099
3	96	25	94779	8316	9952
4	45	32	74665	3907	5520
5	131	13	55301	2831	2677
6	148	57	48564	2922	8255
7	145	29	29354	2872	9640
8	134	69	25754	3084	5366
9	26	77	24188	3030	5674
10	100	63	18113	2965	9634
11	21	14	13302	1394	4498
12	85	76	11219	3928	4590
13	144	73	10340	2451	4131
14	37	71	8630	3513	9847
15	115	7	4876	2376	1701
16	142	8	2357	765	2884
17	20	67	1898	1554	10502
18	46	42	1312	840	8497
19	39	60	1016	724	9683
20	82	28	769	176	10852
21	145	43	664	578	8837
22	23	33	657	833	2841

em uma ou outra imagem para satisfazer a distância mínima. Com base nesta observação passamos a utilizar distância mínima zero.

Diversos pares de imagens foram utilizados durante o ajuste dos parâmetros do mecanismo de detecção de vértices e de correspondência de caracteres. Para a avaliação final dos métodos de correspondência foram tomados quatro pares de imagens representativos de ambientes distintos:

1. Ambiente altamente estruturado com objetos dispostos esparsamente; este ambiente se aproxima de um campo de jogo RoboCup, como descrito na seção 1.1, porém sem as marcações da área penal, linha mediana e círculo central;
2. Ambiente com objetos dispostos aproximadamente ao longo de um plano bidimensional, com profundidade de até  $3m$ .
3. Ambiente com superfícies de alta especularidade, com profundidade de até  $6m$ .
4. Gabarito de calibração não coplanar posicionado a aproximadamente  $40cm$  do conjun-

to binocular.

O procedimento de correspondência foi dividido em três etapas: (1) normalização dos dados, (2) preparação da *matriz de correspondência cruzada* e (2) refinamento da correspondência. Na última etapa tem-se como objetivo principal alcançar uma condição de unicidade na correspondência dos caracteres e é onde os dois métodos diferem.

Na primeira etapa cada um dos conjuntos de valores dos caracteres de ambas as imagens é normalizado da seguinte forma:

$$\tilde{P} = \frac{P - \bar{P}}{\sigma_P} \quad \tilde{V} = \frac{V - \bar{V}}{\sigma_V} \quad \tilde{K} = \frac{K - \bar{K}}{\sigma_K} \quad \tilde{I} = \frac{I - \bar{I}}{\sigma_I} \quad (4.1)$$

onde  $\tilde{P}, \tilde{V}, \tilde{K}$  e  $\tilde{I}$  são os valores normalizados,  $\bar{P}, \bar{V}, \bar{K}$  e  $\bar{I}$  são as médias dos valores e  $\sigma_P, \sigma_V, \sigma_K$  e  $\sigma_I$  são os desvios padrão dos valores de  $P, V, K,$  e  $I,$  respectivamente, nas duas imagens.

Na segunda etapa é construída uma matriz, aqui chamada de *matriz de correspondência cruzada*, que representa o produto entre os caracteres nas duas imagens. Cada elemento da matriz contém um indicador de confiabilidade na correspondência daqueles dois caracteres. Se na detecção o número de caracteres selecionados é limitado, por exemplo, aos 300 melhores, o número máximo de elementos desta matriz será também limitado, neste caso a 90.000. Os pares de caracteres que possuem um alto valor de confiabilidade são candidatos a pares correspondentes.

O valor de confiabilidade  $C$  da correspondência, adaptado de Pilu e Lorusso [22], é a produto ponderado das diferenças dos quatro valores dos caracteres. Cada um destes componentes tem um fator de ponderação ( $\lambda$ ) associado e o valor de confiabilidade entre dois caracteres é calculado como:

$$C = e^{-\frac{(P_1 - P_2)^2}{2\lambda_P^2}} \times e^{-\frac{(V_1 - V_2)^2}{2\lambda_V^2}} \times e^{-\frac{(K_1 - K_2)^2}{2\lambda_K^2}} \times e^{-\frac{(I_1 - I_2)^2}{2\lambda_I^2}} \quad (4.2)$$

Nos testes realizados foi utilizado um valor de  $\lambda_P$  igual a 4, enquanto  $\lambda_V, \lambda_K$  e  $\lambda_I$  eram todos iguais a 0,7. Pilu e Lorusso [22] sugere utilizar  $\lambda_P$  igual a 1/8 da largura da imagem, o que em nossos testes significaria  $\lambda_P$  igual a 40. Este valor não foi utilizado pois levou a um número muito grande de pares candidatos a correspondentes.

Ao valor de confiabilidade da correspondência acima pode-se ainda introduzir um termo adicional, após conhecida a matriz fundamental, relativo à distância da linha epipolar de um

caractere projetado na outra imagem até o caractere desta outra imagem. Este termo não foi utilizado neste trabalho pois a correspondência de caracteres foi feita somente uma vez, antes de conhecida a matriz fundamental.

### Refinamento por relaxação

No refinamento por relaxação o primeiro passo é a aplicação de um limiar de corte no valor de confiabilidade, selecionando os pares com valor acima do limiar. O valor típico deste limiar está entre 0,5-0,8. Nestes testes foi utilizado o valor 0,5. Observe-se que quanto menor o limiar maior a quantidade de pares candidatos resultante e maior a necessidade de robustez do procedimento de relaxação. O passo seguinte é a representação dos pares candidatos através de uma relação.

A primeira conclusão obtida na avaliação do método de relaxação é que a natureza repetitiva e regular dos gabaritos de calibração impede o uso deste método com este tipo de imagem. Como exemplo, dispúnhamos de dois conjunto de 300 vértices detectados em um par de imagens obtidas de um gabarito de calibração. Na preparação dos pares candidatos obteve-se um total de 4.185 pares candidatos a correspondentes. Após a aplicação do procedimento de relaxação restaram 244 pares correspondentes, onde havia zero pares corretos.

Além da observação acima, deve-se considerar que o gabarito de calibração é posicionado muito perto das câmeras, se comparado com as imagens do ambiente, e que quanto maior a proximidade também maior é a disparidade que influencia a seleção dos pares candidatos através do parâmetro  $\lambda_P$  do valor de confiabilidade.

O procedimento de relaxação utiliza três parâmetros, o primeiro deles é o limite do número de iterações, o segundo é o raio da vizinhança em torno do caractere que está sendo analisado e o terceiro é o limiar do gradiente de disparidade. Este último parâmetro, como já citado na seção 2.2, possui valor típico unitário. O limite de iterações utilizado foi de 20 iterações, sendo que o procedimento implementado também é parado quando detecta a saturação da condição de unicidade das correspondências, o que geralmente ocorre antes da metade deste limite de iterações. Finalmente o raio da vizinhança utilizado é de  $20\sigma_P$  que, por conveniência, também é normalizado para ser consistente com  $\tilde{P}$ , o que é obtido através da norma do raio de vizinhança dividido por  $\sigma_P$ .

Uma característica observada na relaxação é que este método é eficiente na distinção entre pares de caracteres próximos nas duas imagens, o que é conveniente, pois erros de

Tabela 4.7: Dados experimentais da avaliação do refinamento por relaxação

Par	Vért.		Cand.	Primeira iteração		Iterações restantes		Total	
	Esq.	Dir.		Corr.	% Acerto	Corr.	% Acerto	Corr.	% Acerto
1	68	94	1014	47	83,0	18	16,7	65	64,6
2	300	300	6942	203	84,7	68	20,6	271	68,6
3	300	300	10.871	114	50,0	149	14,8	263	30,0

correspondência nestes caracteres não seriam detectados durante a calibração das câmeras. Por outro lado, além de aceitar as correspondências corretas, muitas correspondências incorretas são aceitas. Acredita-se que estas correspondências incorretas tenham sido aquelas restantes após algumas iterações do algoritmo e onde não haviam outras correspondências na vizinhança para serem utilizadas como critério de aceitação.

### Refinamento por valores singulares

No caso de ser aplicado refinamento da correspondência por valores singulares os pares candidatos ficam representados na própria matriz de correspondência cruzada e também não há a aplicação de um limiar de corte sobre os valores de confiabilidade antes do refinamento. Todavia, o limiar de corte pode ser aplicado durante a seleção final do pares correspondentes, e os valores típicos estão também entre 0,5-0,8.

Na tabela 4.8 estão resumidos os resultados obtidos a partir da mesma matriz de correspondência cruzada utilizada na avaliação do método de relaxação.

Tabela 4.8: Dados experimentais da avaliação do refinamento por valores singulares

Par	Vért.		Cand.	Limiar	Correspondências	
	Esq.	Dir.			Total	% Acerto
1	68	94	6.392	Não	59	49,2
				0,5	57	50,9
2	300	300	90.000	Não	225	41,8
				0,5	215	41,9
3	300	300	90.000	Não	213	23,9
				0,5	203	23,2

### 4.3 Fase 1 - avaliação da distorção radial

Na seção 3.4.1 foi introduzido o modelo típico para compensação da distorção radial das lentes. Neste experimento o objetivo foi selecionar o modelo de compensação da distorção radial e foram avaliados quatro modelos conforme tabela 4.9.

Tabela 4.9: Modelos de compensação da distorção radial avaliados

Modelo	Coordenadas de sensor	
1	$u = \tilde{u}[1 + k_1(\tilde{u}^2 + \tilde{v}^2)]$ $v = \tilde{v}[1 + k_1(\tilde{u}^2 + \tilde{v}^2)]$	Polinômio de segunda ordem com mesmo coeficiente para ambos os eixos
2	$u = \tilde{u}[1 + k_{u1}(\tilde{u}^2 + \tilde{v}^2)]$ $v = \tilde{v}[1 + k_{v1}(\tilde{u}^2 + \tilde{v}^2)]$	Polinômio de segunda ordem com coeficientes independentes para cada eixo
3	$u = \tilde{u}[1 + k_1(\tilde{u}^2 + \tilde{v}^2) + k_2(\tilde{u}^2 + \tilde{v}^2)^2]$ $v = \tilde{v}[1 + k_1(\tilde{u}^2 + \tilde{v}^2) + k_2(\tilde{u}^2 + \tilde{v}^2)^2]$	Polinômio de quarta ordem com mesmo coeficiente para ambos os eixos
4	$u = \tilde{u}[1 + k_{u1}(\tilde{u}^2 + \tilde{v}^2) + k_{u2}(\tilde{u}^2 + \tilde{v}^2)^2]$ $v = \tilde{v}[1 + k_{v1}(\tilde{u}^2 + \tilde{v}^2) + k_{v2}(\tilde{u}^2 + \tilde{v}^2)^2]$	Polinômio de quarta ordem com coeficientes independentes para cada eixo

Para avaliação dos modelos foi utilizado o procedimento de calibração com gabarito não coplanar de Tsai, conforme descrito na seção 3.4.2. A distância aproximada da câmera até o gabarito é de 40cm. As câmeras utilizadas são as já descritas na seção 4.1.1. A câmera Minolta C-560C, que não faz parte do conjunto binocular, possui distorção radial desprezível e participa da avaliação dos modelos principalmente como referência para análise comparativa dos erros de calibração.

O procedimento geral para a calibração das câmeras está resumido a seguir:

1. Posicionar gabarito próximo à câmera de forma que apareça integralmente na imagem;
2. Capturar imagem do gabarito;
3. Aplicar limiar de corte na intensidade obtendo imagem binível;
4. Detectar regiões na imagem binível usando expansão de regiões;
5. Selecionar regiões usando fator de forma;
6. Calcular centróide de cada região;
7. Ordenar regiões por linha e por coluna de acordo com a posição do centróide;
8. Relacionar centróides das regiões com as posições conhecidas do gabarito;

9. Calibrar câmera obtendo parâmetros internos e externos;
10. Avaliar erro de calibração.

As mesmas imagens de cada câmera foram utilizadas em todos os procedimentos de calibração variando somente o modelo de compensação da distorção radial. A resolução espacial de todas as imagens é  $320 \times 240$  *pel*. A figura 4.6 abaixo dá exemplos de imagens obtidas com as câmeras KAL-233 (a) e MIN-077 (b), onde os pontos marcados são os centróides das regiões detectadas. O efeito da distorção radial na câmera KAL-240 é da mesma ordem de grandeza do que ocorre na KAL-233.

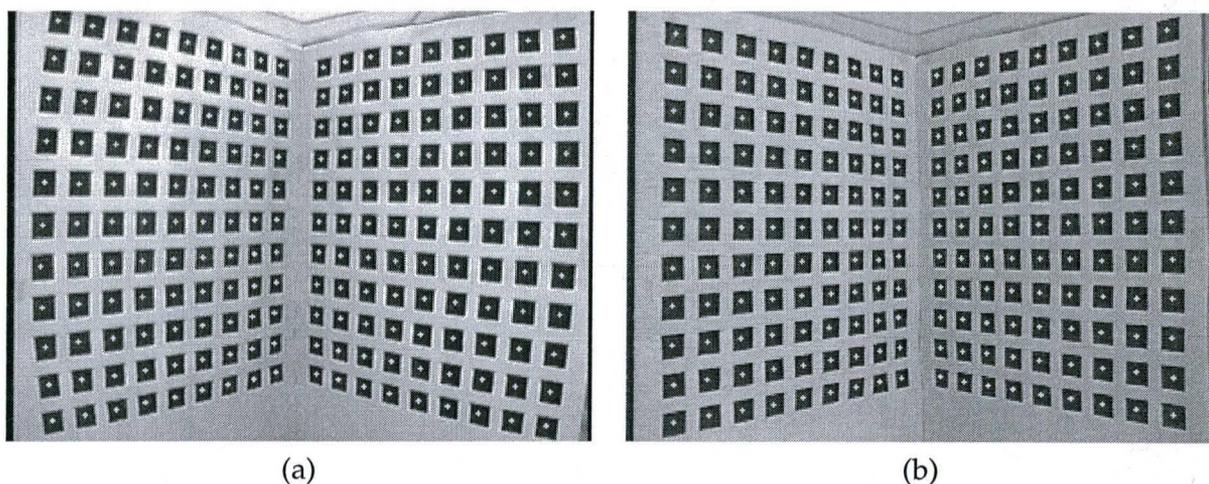


Figura 4.6: Exemplos de imagens do gabarito não coplanar

Para cada câmera o procedimento de calibração foi repetido pelo menos três vezes, com pequena variação da orientação ou posição da câmera em relação ao gabarito, e os melhores resultados de cada câmera em cada modelo estão resumidos na tabela 4.10.

Pelos dados da tabela 4.10 pode-se observar que mesmo utilizando o polinômio de quarta ordem nas câmeras KAL-233 e KAL-240 não foi possível alcançar os indicadores de erro da câmera MIN-077. De fato os dados são consistentes com observação feita por Zhang [36] de que o primeiro termo da distorção radial é o mais relevante e de que modelos mais elaborados não são de grande auxílio, além de introduzirem instabilidades numéricas.

A figura 4.7 mostra a mesma imagem da figura 4.6(a) seguida daquela imagem após aplicada a compensação da distorção radial pelo modelo 2 da tabela 4.9.

Tabela 4.10: Dados experimentais da avaliação do modelo de distorção radial

Mod.	Câmara	f (mm)	k (1/mm <sup>2</sup> )				u <sub>0</sub> (pel)	v <sub>0</sub> (pel)	UIPE (pel)	NCE (mm)
—	KAL-233	6.39	—				159.39	117.87	1.51	3.71
	KAL-240	6.32	—				153.68	110.86	1.38	3.39
	MIN-077	5.99	—				156.56	110.96	0.72	1.75
1	KAL-233	5.05	0.0157				154.59	107.34	0.96	2.35
	KAL-240	4.90	0.0183				149.01	121.67	0.89	2.19
	MIN-077	5.57	0.0045				156.64	110.59	0.65	1.59
2	KAL-233	4.99	0.0150	0.0190			156.04	110.60	0.91	2.25
	KAL-240	4.82	0.0178	0.0231			155.29	122.40	0.81	2.00
	MIN-077	5.50	0.0034	0.0085			156.58	109.51	0.55	1.34
3	KAL-233	5.04	0.0038	0.0015			167.10	115.93	0.89	2.18
	KAL-240	4.93	0.0063	0.0014			156.08	125.94	0.80	1.96
	MIN-077	5.61	0.0016	0.0003			156.74	109.56	0.65	1.58
4	KAL-233	4.99	0.0021	0.0089	0.0016	0.0013	166.91	116.41	0.82	2.02
	KAL-240	4.84	0.0055	0.0125	0.0014	0.0012	158.36	122.84	0.71	1.74
	MIN-077	5.50	0.0004	0.0100	0.0004	-0.0002	155.81	109.26	0.54	1.32

UIPE: erro no plano de imagem não normalizado. NCE: erro de calibração normalizado.

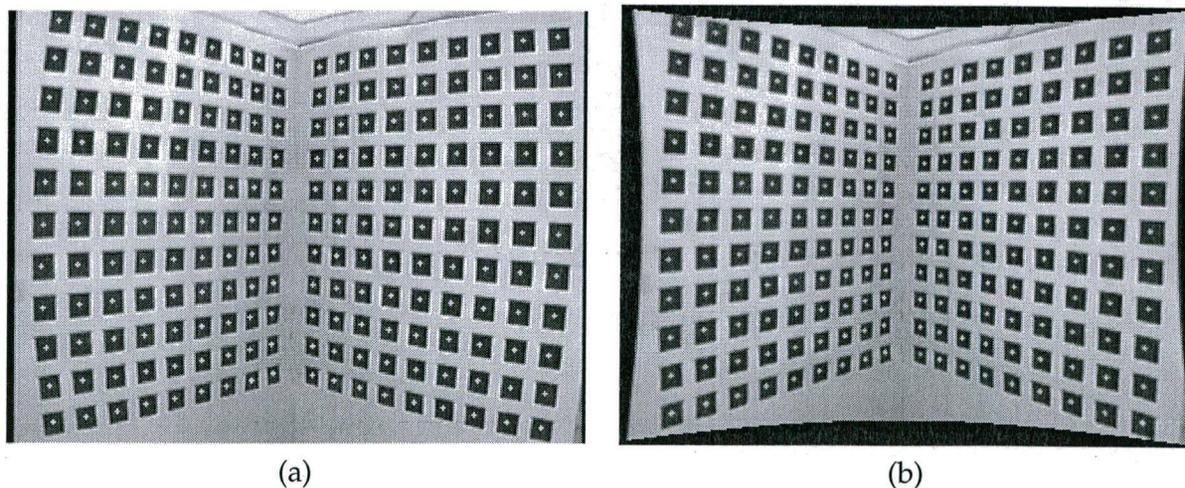


Figura 4.7: Compensação da distorção radial pelo polinômio de segunda ordem

## 4.4 Fase 2 - calibração com gabarito

### 4.4.1 A escolha do gabarito

Foram testados dois tipos de gabaritos de calibração. A figura abaixo mostra exemplos de pares de imagens obtidas do gabarito de calibração coplanar (4.8a câmera esquerda, 4.8b câmera direita) e não coplanar (4.8c câmera esquerda, 4.8d câmera direita).

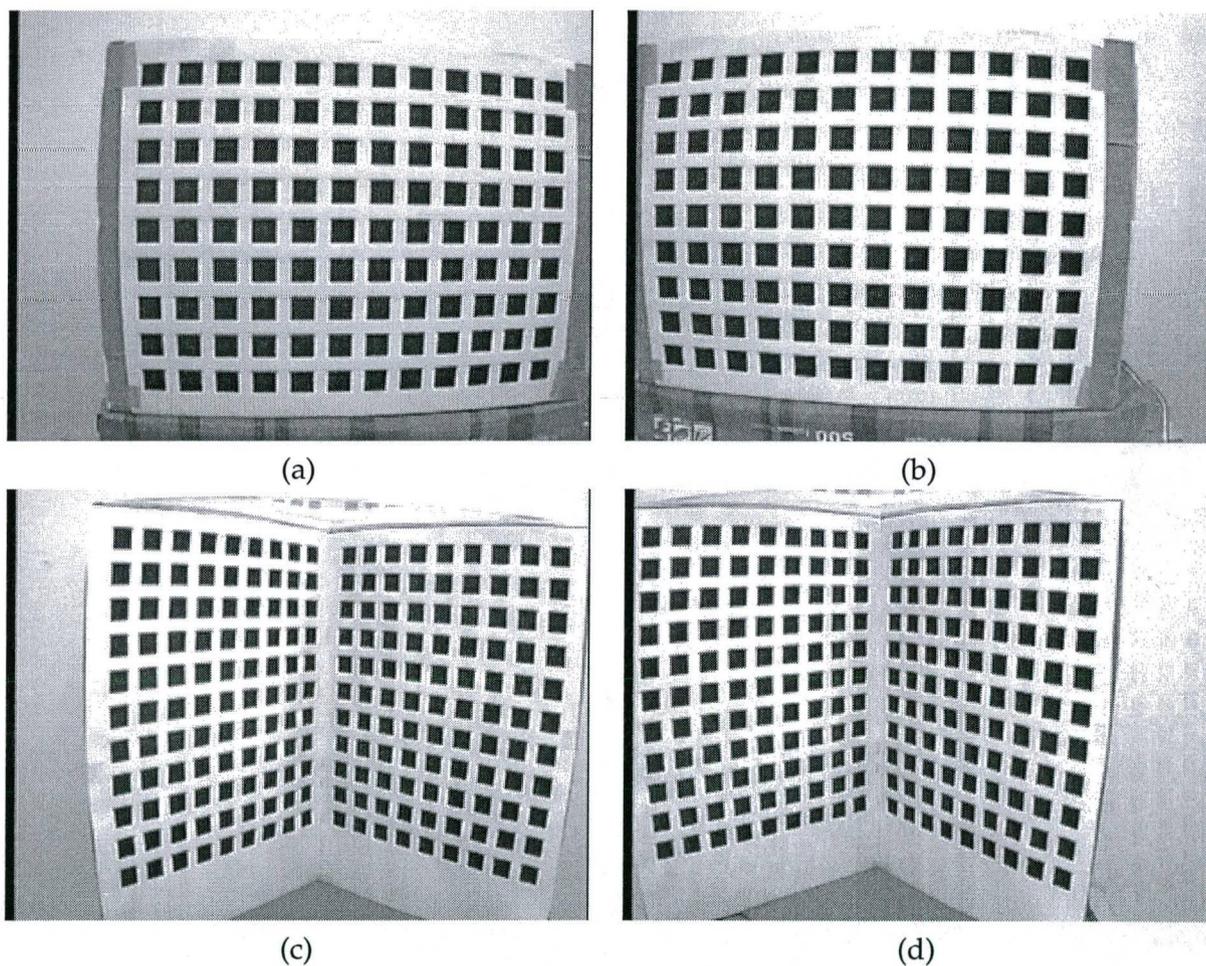


Figura 4.8: Gabaritos de calibração

Observou-se que o procedimento de calibração com gabarito coplanar leva, mais frequentemente, a resultados discrepantes, especialmente quando o gabarito está posicionado aproximadamente paralelo ao plano de imagem. Exemplos de imagens nesta situação podem ser vistos na figura 4.9 com os resultados da calibração na tabela 4.11 abaixo. Esta situação já havia sido prevista como uma situação indesejável por Willson [34] que recomenda um ângulo relativo entre o plano de imagem e o gabarito de  $30^\circ$  ou mais.

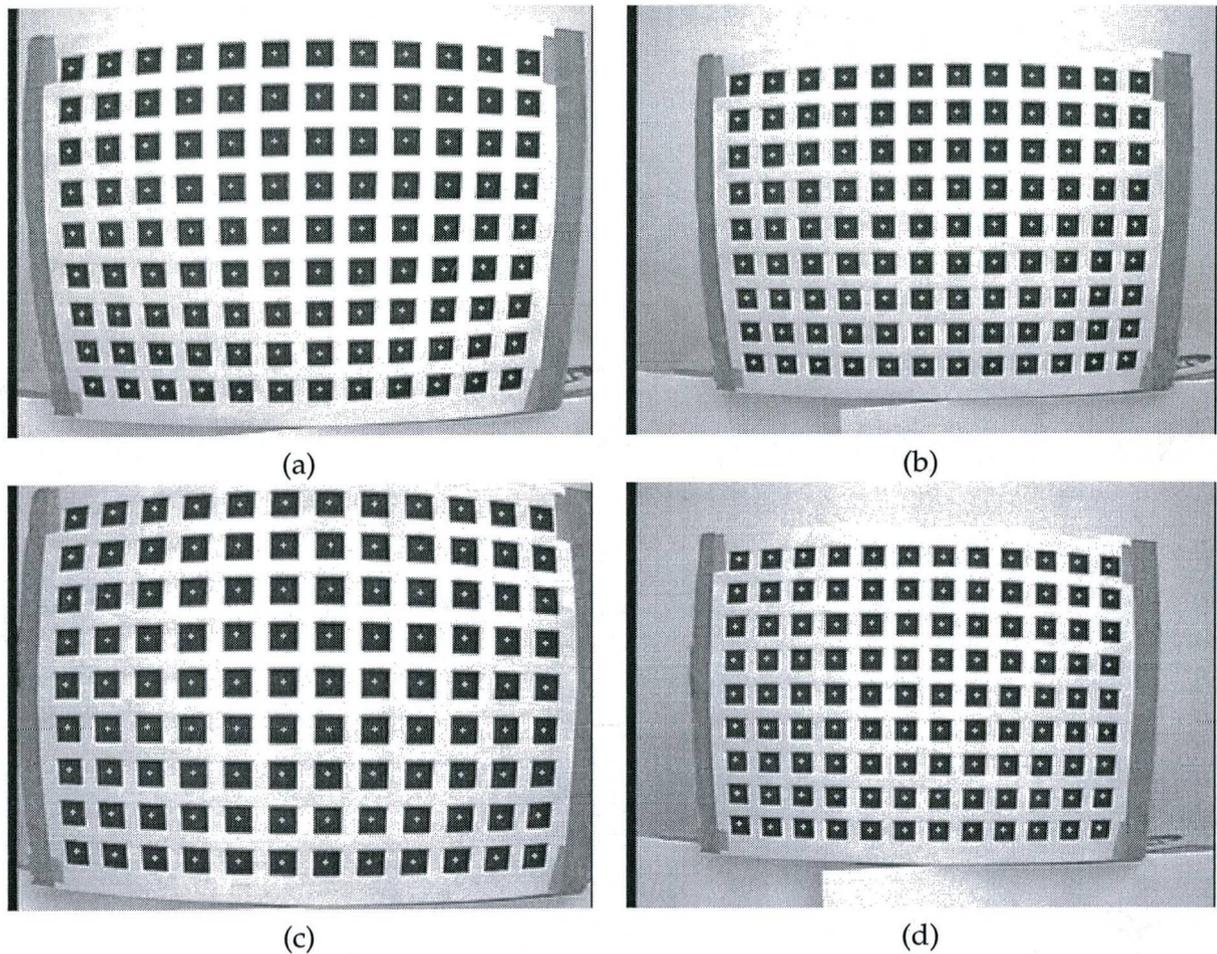


Figura 4.9: Gabarito posicionado aproximadamente paralelo ao plano de imagem

Tabela 4.11: Calibração com gabarito coplanar  
Gabarito aproximadamente paralelo ao plano de imagem

Parâmetro	Imagem (conforme figura 4.9)			
	a	b	c	d
Câmera	KAL-233 (E)	KAL-233 (E)	KAL-240 (D)	KAL-240 (D)
$f$ (mm)	-2927.34	-66.24	-4174.96	-15.06
$T_x$ (mm)	124.40	122.93	-118.05	-80.68
$T_y$ (mm)	36.58	44.32	-90.18	-77.41
$T_z$ (mm)	166763.91	4420.31	-220471.68	-1084.14
$R_x$ (°)	-181.39	-715.29	-179.53	178.69
$R_y$ (°)	5.60	1261.45	14.18	-14.84
$R_z$ (°)	179.38	-360.42	0.58	0.96
UIPE (pel)	1.22	0.89	0.97	0.73
NCE (mm)	2.99	2.17	2.38	1.79

UIPE: erro no plano de imagem não normalizado. NCE: erro de calibração normalizado

Também para o gabarito não coplanar Willson [34] prevê uma situação indesejável que leva a problemas de calibração. Neste caso problemas ocorrem quando o volume do gabarito não coplanar é muito menor que o volume da distância da câmera até o gabarito.

O que ocorre em ambos os tipos de gabarito é que o efeito da projeção em perspectiva é reduzido e a projeção se aproxima de uma projeção ortográfica. Deste modo não se pode distingüir entre o efeito da distância focal e o efeito da profundidade na variação de escala do gabarito, o que leva à instabilidade do algoritmo.

Uma vez que o conjunto binocular utilizado é convergente a ocorrência de resultados discrepantes na calibração a partir do gabarito coplanar é menos freqüente. Todavia, a partir deste momento passaremos a utilizar somente o gabarito não coplanar.

#### 4.4.2 Avaliação do erro de localização

Willson [34] observa que para estimar com precisão a distorção radial das lentes e o desvio do centro da imagem, os dados de calibração, em nosso caso os centróides das regiões do gabarito, devem estar distribuídos ao longo de toda a imagem. Willson [34] também observa que deve-se procurar fazer com que estes dados de calibração percorram toda a faixa de profundidade na qual as câmeras vão operar. O mesmo é observado por Zhang *et al.* [37] que cita que a calibração só é válida para o volume em torno da posição do gabarito e que os resultados degradam quando os objetos se afastam daquela posição.

Para contornar estes problemas pode-se utilizar vários gabaritos de calibração ao longo do ambiente ou fazer várias tomadas das imagens com o gabarito em posições diferentes; em todos os casos é necessário saber precisamente a posição do gabarito no ambiente em cada uma das imagens.

Para a avaliação da eficiência do procedimento de calibração com gabarito, em face das observações anteriores, o conjunto binocular foi calibrado duas vezes, sempre utilizando o gabarito não coplanar. Na primeira calibração o gabarito foi posicionado a uma distância aproximada de 450mm a frente do conjunto binocular e na segunda calibração foi posicionado a aproximadamente 700mm (ver tabela 4.12).

Após a calibração cada uma das duas configurações de calibração foi testada com todos os dois pares de imagens iniciais. Como a posição dos caracteres no gabarito é conhecida, assim como a correspondência exata entre estes caracteres, para cada caractere puderam ser calculadas três informações de referência: (1) erro de posição do caractere no espaço

Tabela 4.12: Calibração com gabarito não coplanar

Parâmetro	Distância do gabarito			
	450mm		700mm	
Câmera	KAL-233 (E)	KAL-240(D)	KAL-233 (E)	KAL-240 (D)
$f$ (mm)	4.78	4.71	4.37	4.63
$k_1$ ( $1/mm^2$ )	0.0146	0.0188	0.0136	0.0015
$k_2$ ( $1/mm^2$ )	0.0208	0.0254	0.0237	0.0255
$u_0$ (pel)	165.11	166.82	168.29	195.94
$v_0$ (pel)	111.16	129.15	104.78	120.24
$T_x$ (mm)	-112.11	-175.75	-57.72	-129.67
$T_y$ (mm)	-100.98	-121.76	-73.85	-103.64
$T_z$ (mm)	327.18	316.97	586.04	608.44
$R_x$ (°)	-89.52	-93.16	-87.51	-90.97
$R_y$ (°)	46.95	43.29	41.73	40.71
$R_z$ (°)	-177.78	-179.38	-176.77	-178.22
UIPE (pel)	0.65	0.69	0.56	0.52
NCE (mm)	1.61	1.70	1.36	1.16

UIPE: erro no plano de imagem não normalizado. NCE: erro de calibração normalizado.

tridimensional, (2) erro de distância do caractere no plano horizontal e (2) distância da linha epipolar. A medida da distância no plano horizontal é relevante para o caso RoboCup, onde a referência para a posição dos objetos é o plano do campo de jogo (ver seção 1.1). A distância da linha epipolar é importante pois pode ser usada como restrição na correspondência dos caracteres de imagens do ambiente. O procedimento para os testes é descrito abaixo:

1. Posicionar gabarito próximo ao conjunto binocular;
2. Capturar imagens das duas câmeras sem modificar a posição ou configuração do conjunto binocular;
3. Calibrar cada uma das câmeras individualmente:
  - (a) Detectar regiões nas imagens;
  - (b) Selecionar regiões usando fator de forma;
  - (c) Calcular centróide da região;
  - (d) Ordenar regiões por linha e por coluna de acordo com a posição do centróide;
  - (e) Relacionar centróides das regiões as com posições conhecidas do gabarito;
  - (f) Calibrar câmera obtendo matriz de projeção;
4. Calcular posição dos caracteres no espaço utilizando estereoscopia;

5. Comparar posição dos caracteres com as posições conhecidas do gabarito;
6. Comparar distância dos caracteres no plano  $OXZ$  com as distâncias conhecidas do gabarito;
7. Calcular matriz fundamental a partir das duas matrizes de projeção;
8. Comparar distância da linha epipolar  $l'$  para cada caractere:
  - (a) Selecionar caractere na primeira imagem;
  - (b) Projetar linha epipolar na segunda imagem;
  - (c) Calcular distância do caractere correspondente da segunda imagem até a linha epipolar.

Tabela 4.13: Calibração com gabarito não coplanar  
Posição dos caracteres do gabarito de calibração

Imagem	Medida	Erro		
		Posição (mm)	Dist. $OXZ$ (mm)	Dist. $l'$ (pel)
Calibrado com gabarito a 450mm				
450mm	Média	2.94	2.70	0.43
	Desvio padrão	2.08	2.17	0.31
	Máximo	10.62	10.33	1.54
700mm	Média	366.62	356.17	1.27
	Desvio padrão	15.40	23.36	0.66
	Máximo	406.30	397.59	2.82
Calibrado com gabarito a 700mm				
450mm	Média	296.81	285.26	1.86
	Desvio padrão	16.06	22.14	1.30
	Máximo	321.11	314.53	5.30
700mm	Média	7.12	6.89	0.39
	Desvio padrão	4.98	5.11	0.31
	Máximo	25.45	25.78	1.24

Na tabela acima se observa que aumentando a distância do gabarito também aumenta o erro no cálculo da posição dos caracteres. Isto não é uma consequência da distância, mas sim do tamanho do gabarito que, sendo fixo, ao se aumentar a distância do conjunto binocular ocupa uma área menor no plano de imagem. Observe-se ainda que em ambas as distâncias os erros de calibração foram aproximadamente os mesmos, conforme tabela 4.12.

Testes adicionais foram realizados com imagens obtidas dos quatro tipos de ambientes

descritos na seção 4.2.3 com o objetivo que avaliar a calibração do conjunto binocular em uma faixa de profundidade mais próxima da utilizada na RoboCup. Os resultados estão resumidos na tabela 4.14 abaixo.

Tabela 4.14: Calibração com gabarito não coplanar  
Posição dos caracteres do gabarito de calibração e do ambiente

Amb.	Par	$x_i$ (pel)	$y_i$ (pel)	$x'_i$ (pel)	$y'_i$ (pel)	Prof. <sup>(1)</sup> (mm)	Dist. OXZ (mm)	Dist. $l'$ (mm)
Calibrado com gabarito a 450mm								
1	3	109	95	114	85	2500	934.78	13.62
1	40	211	89	214	80	2200	884.07	13.93
2	57	142	63	145	60	1000	1146.79	7.32
2	137	220	38	233	35	2000	1147.52	11.17
3	53	254	158	248	159	2000	942.89	4.99
3	148	175	62	199	60	6000	619.91	8.80
4	0 <sup>(2)</sup>	72	28	28	34	350	359.27	0.38
4	99 <sup>(2)</sup>	188	50	153	50	450	446.29	0.90
4	0	143	73	127	73	600	669.75	2.58
4	99	208	80	198	81	700	805.94	2.09
Calibrado com gabarito a 700mm								
1	3	109	95	114	85	2500	1009.82	11.23
1	40	211	89	214	80	2200	915.69	10.58
2	57	142	63	145	60	1000	1209.61	3.33
2	137	220	38	233	35	2000	1595.65	3.18
3	53	254	158	248	159	2000	755.18	4.42
3	148	175	62	199	60	6000	4180.28	1.76
4	0	72	28	28	34	350	395.53	4.10
4	99	188	50	153	50	450	407.11	1.80
4	0 <sup>(2)</sup>	143	73	127	73	600	622.90	1.13
4	99 <sup>(2)</sup>	208	80	198	81	700	702.50	0.13

(1) Profundidade real aproximada estimada por medições no ambiente.

(2) Estes caracteres pertencem aos dados de calibração.

#### 4.5 Fase 3 - calibração a partir de caracteres do ambiente

Pôde ser observado na seção anterior que, como previsto por Zhang *et al.* [37], a informação de calibração obtida através do gabarito de calibração é válida somente no volume do gabarito. Todavia, a geometria do conjunto binocular pode ser obtida utilizando-se na calibração não só os caracteres identificados no gabarito mas também caracteres identificados

nas imagens obtidas do ambiente. A geometria assim obtida será válida em todo o ambiente e será inicialmente descrita através da matriz fundamental. Como visto na seção 3.2.3 a matriz fundamental incorpora informações sobre o conjunto binocular suficientes para a descrição da geometria epipolar.

O processo adotado para determinação da matriz fundamental consiste em várias etapas:

1. Detectar caracteres nas imagens do gabarito e do ambiente;
2. Identificar pares de caracteres correspondentes nas imagens do ambiente utilizando um mecanismo de relaxação;
3. Calcular matriz fundamental utilizando pares de caracteres do gabarito e pares, eventualmente incorretos, das imagens do ambiente utilizando um mecanismo que procure ser tolerante às falhas de correspondência;
4. Eliminar os pares inválidos do conjunto de pares de caracteres do ambiente utilizando como referência a matriz fundamental obtida;
5. Calcular nova matriz fundamental a partir dos pares de caracteres do gabarito e pares de caracteres restantes do ambiente.

Para a calibração do conjunto binocular em estudo foi tomado um par de imagens do gabarito de calibração e vários pares de imagens do ambiente procurando cobrir toda a faixa de profundidade do ambiente. Baseado neste critério quatro imagens do ambiente foram selecionadas para uso na calibração.

O resultado da identificação dos caracteres e da correspondência dos mesmos através do algoritmo de relaxação está resumido na tabela 4.15 a seguir.

A busca de robustez na determinação da matriz fundamental nos leva à utilização do algoritmo LMedS, já introduzido na seção 3.4.4. Para comparação o cálculo da matriz fundamental foi realizado tanto com o algoritmo LS quanto LMedS, em ambos os casos utilizando o mesmo conjunto de pares correspondentes das imagens, potencialmente incorretos. Para cada par de imagens foi obtida uma matriz fundamental utilizando os pares de caracteres correspondentes daquelas imagens. Também foi calculada uma matriz fundamental utilizando todos os pares de caracteres de todos os pares de imagens. Os erros de calibração em relação ao conjunto de calibração, medidos como a distância da linha epipolar, estão resumidos na tabela 4.16. Também foi calculado o erro de calibração das matrizes fundamentais obtidas em relação ao conjunto total de pares correspondentes em todas as imagens, o que está resumido na tabela 4.17.

Tabela 4.15: Resultado da correspondência dos caracteres a serem utilizados na calibração

Par de imagens	Caracteres		Candidatos	Selecionados	Acerto <sup>(1)</sup>	% Acerto
	Esquerda	Direita				
Gabarito	198	198	198	198 <sup>(2)</sup>	198	100,00%
Imagem 101	300	300	191	171	169	98,83%
Imagem 102	300	300	239	187	181	96,79%
Imagem 103	300	300	199	161	153	95,03%
Imagem 104	300	300	186	159	150	94,34%
Subtotal ambiente				678	653	96,31%
Total geral				876	851	97,15%

(1) Pares corretos foram identificados por inspeção visual para determinar taxa de acerto.

(2) Correspondência dos caracteres do gabarito é conhecida *a priori*.

Tabela 4.16: Erros de calibração em relação ao conjunto de calibração

Método	Par de de imagens	Mediana	Média	Desvio	RMS
LS	Gabarito	0,5278	0,6488	0,5367	0,8412
	Imagem 101	1,4037	1,7306	1,3946	2,2200
	Imagem 102	1,8658	2,5619	3,8234	4,5939
	Imagem 103	1,1281	1,5896	1,6083	2,2577
	Imagem 104	1,0376	1,4963	1,6857	2,2500
	Todos	1,0268	1,8072	5,1439	5,4494
LMedS	Gabarito	0,5243	0,7243	0,5850	0,9302
	Imagem 101	0,0000	1,5906	6,0660	6,2539
	Imagem 102	0,5605	1,5733	6,8445	7,0051
	Imagem 103	0,0000	1,4783	3,4316	3,7267
	Imagem 104	0,4148	1,9132	6,3050	6,5698
	Todos	0,8818	1,9657	5,4058	5,7492

Tabela 4.17: Erros de calibração em relação ao conjunto total de pares

Método	Par de imagens	Mediana	Média	Desvio	RMS
LS	Gabarito	1,7466	3,3192	5,7320	6,6209
	Imagem 101	4,1233	24,7751	38,7607	45,9835
	Imagem 102	4,4938	23,3588	36,0192	42,9131
	Imagem 103	7,6855	29,4797	43,1219	52,2152
	Imagem 104	3,5896	18,7801	30,2446	35,6170
LMedS	Gabarito	5,3811	5,9755	6,5806	8,8860
	Imagem 101	2,0000	2,7808	5,5650	6,2183
	Imagem 102	1,0154	1,9776	5,3175	5,6705
	Imagem 103	2,0000	2,7580	5,6025	6,2417
	Imagem 104	1,8370	3,2292	6,0306	6,8377

Podemos observar nas tabelas 4.16 e 4.17 a deficiência do algoritmo LS no tratamento do conjunto de pares de caracteres que contém correspondências inválidas se compararmos com o conjunto de pares de caracteres do gabarito e que não contém erros de correspondência. Porém nem mesmo o algoritmo LMedS é suficiente para obtenção de uma matriz fundamental consistente com todo o conjunto de pares se este conjunto de pares contém pares inválidos.

O que se deve fazer neste caso é utilizar a melhor matriz fundamental obtida, então por LMedS, como referência para reduzir o conjunto de pares de caracteres excluindo-se do conjunto os pares inválidos, como sugerido por Gracias e Santos-Victor [11]. Também foi avaliado aqui a utilização da matriz fundamental calculada diretamente a partir das matrizes de projeção das câmeras obtidas na calibração somente com gabarito mas os resultados foram insatisfatórios se comparados com a matriz fundamental obtida por LMedS.

Para a redução dos pares de caracteres correspondentes foram avaliadas quatro condições de exclusão:

- A - Quadrado da distância da linha epipolar maior que limiar proporcional ao quadrado do desvio padrão normalizado;
- B - Distância da linha epipolar maior que a mediana;
- C - Distância da linha epipolar maior que limiar proporcional ao valor RMS da distância;
- D - Quadrado da distância da linha epipolar maior que limiar proporcional ao desvio

padrão;

Os resultados da redução do conjunto de pares correspondentes utilizando-se cada uma destas condições podem ser comparados observando-se a tabela 4.18.

Tabela 4.18: Redução do conjunto de pares correspondentes

Par de imagens	Condição de exclusão	Limiar de seleção	Pares restantes	Acerto	% Acerto
Imagem 101	A	12,8632	166	165	99,40%
	B	0,8811	85	85	100,00%
	C	3,0740	166	165	99,40%
	D	35,5602	169	168	99,41%
Imagem 102	A	10,1970	181	179	98,90%
	B	0,7022	93	92	98,92%
	C	3,5193	182	180	98,90%
	D	47,1467	184	181	98,37%
Imagem 103	A	12,4021	155	152	98,06%
	B	0,8462	80	79	98,75%
	C	1,6609	128	126	98,44%
	D	8,8847	154	151	98,05%
Imagem 104	A	10,4515	154	148	96,10%
	B	0,7125	79	78	98,73%
	C	3,2674	154	148	96,10%
	D	40,0548	155	149	96,13%
Acumulado <sup>(1)</sup>	A	-	656	644	98,17%
	B	-	337	334	99,11%
	C	-	630	619	98,25%
	D	-	662	649	98,04%
Todos <sup>(2)</sup>	A	11,0914	657	645	98,17%
	B	0,7954	338	334	98,82%
	C	2,9930	654	642	98,17%
	D	33,3366	664	650	97,89%

(1) Totais acumulados da redução dos pares correspondentes de cada par imagem individualmente.

(2) Condição de exclusão aplicada a todos os pares correspondentes de todos os pares de imagens em conjunto.

Pode ser observado na tabela 4.18 que é mais conveniente a aplicar redução em separado para cada subconjunto de pares correspondentes de cada par de imagens e posteriormente agrupar os pares restantes formando o novo conjunto de calibração. Observa-se ainda que a condição de exclusão B levou à maior taxa de acerto, porém com a exclusão de metade dos pares correspondentes, enquanto a condição de exclusão C, embora não obtendo a mesma

taxa de acerto, mostra-se mais conservadora em relação ao número de pares correspondentes restantes.

Após a redução do conjunto de pares correspondentes, e supondo-se que os pares correspondentes inválidos tenham sido eliminados, podemos novamente calcular a matriz fundamental. Para o cálculo da nova matriz fundamental foram utilizados, em separado e para comparação, os dois conjuntos de pares correspondentes restantes da aplicação da condição de exclusão B e C comentadas anteriormente. A matriz fundamental foi calculada duas vezes, também para comparação, utilizando o algoritmo LS e o algoritmos LMedS, para cada um dos dois conjuntos de pares correspondentes disponíveis. Ainda como referência, foi calculada a matriz fundamental utilizando cada um dos dois algoritmos e tendo como base os pares correspondentes sabidamente corretos selecionados anteriormente por inspeção visual.

Os erros de calibração de cada uma das matrizes fundamentais obtidas em relação ao conjunto de calibração estão resumidos na tabela 4.19. Os erros de calibração destas mesmas matrizes fundamentais em relação ao total de pares correspondentes sabidamente corretos estão resumidos na tabela 4.20.

Tabela 4.19: Erros de calibração em relação ao conjunto de calibração  
Conjunto reduzido de pares correspondentes do ambiente mais pares correspondentes do gabarito de calibração

Par de de imagens	Pares corresp.	Método	Mediana	Média	Desvio	RMS
Condição B + Gabarito	535	LS	0,7982	1,1731	1,1748	1,6594
		LMedS	0,5074	1,1492	1,6622	2,0195
Condição C + Gabarito	828	LS	0,9349	1,2320	1,2103	1,7265
		LMedS	0,7897	1,2985	1,6177	2,0736
Referência	851	LS	0,9449	1,3123	1,3202	1,8606
		LMedS	0,8525	1,4591	1,9631	2,4450

Embora tenhamos obtido uma descrição da geometria do conjunto binocular na forma de uma matriz fundamental, o que nos permite utilizar com vantagem a geometria epipolar para redução da ambigüidade na correspondência de caracteres, não podemos, somente com a matriz fundamental, obter uma reconstrução métrica do ambiente, ou seja, ainda não podemos determinar a posição dos objetos, que deram origem aos caracteres identificados, no sistema de coordenadas do ambiente.

Tabela 4.20: Erros de calibração em relação ao conjunto pares corretos  
 Conjunto de pares correspondentes do ambiente conhecidamente  
 corretos mais pares correspondentes do gabarito de calibração

Par de imagens	Método	Pares de corretos	Mediana	Média	Desvio	RMS
Condição B + Gabarito	LS	Gabarito	1,1259	1,6374	1,4820	2,2060
		Ambiente	1,1662	1,4794	1,3070	1,9737
		Todos	1,1525	1,5162	1,3505	2,0299
	LMedS	Gabarito	1,5758	2,3024	2,2577	3,2206
		Ambiente	0,7646	1,0605	1,0326	1,4796
		Todos	0,9214	1,3494	1,5083	2,0232
Condição C + Gabarito	LS	Gabarito	1,4226	2,0203	1,8749	2,7530
		Ambiente	0,9239	1,1037	1,0067	1,4933
		Todos	0,9735	1,3169	1,3199	1,8640
	LMedS	Gabarito	1,6794	2,4630	2,4630	3,6129
		Ambiente	0,7365	1,0545	1,0723	1,5033
		Todos	0,8517	1,3822	1,6924	2,1843
Referência	LS	Gabarito	1,4294	2,0540	1,9056	2,7985
		Ambiente	0,8908	1,0873	0,9784	1,4622
	LMedS	Gabarito	1,8935	3,0480	3,1819	4,4004
		Ambiente	0,7055	0,9773	0,9827	1,3854

#### 4.6 Fase 4 - reconstrução métrica a partir da matriz fundamental

Como citado na seção anterior a obtenção de uma matriz fundamental não nos permite determinar a posição dos objetos no sistema de coordenadas do ambiente. Uma alternativa pode ser utilizar o método proposto por Zhang *et al.* [37] para obter as matrizes de projeção de cada uma das câmeras do conjunto binocular a partir da matriz fundamental e da geometria conhecida do gabarito de calibração. Dispondo das matrizes de projeção podemos então determinar a posição do objeto no sistema de coordenadas do ambiente.

Para a obtenção da matriz fundamental foi utilizado basicamente o mesmo método da seção anterior, porém tomando-se somente os pontos que puderam ser confirmados como correspondentes. Esta confirmação foi obtida analisando-se visualmente cada um dos pares correspondentes propostos e selecionando somente aqueles realmente correspondentes. O cálculo foi realizado com todos estes pontos confirmados ao mesmo tempo, e não tomando um subconjunto destes pontos como amostra de calibração. Outra diferença em relação ao cálculo da matriz fundamental da seção anterior é que, para se obter um maior número

de pontos de calibração do gabarito utilizou-se os vértices, e não os centróides, das regiões retangulares do gabarito.

É sabido que o operador de Harris, utilizado neste trabalho para detecção de vértices, tende a posicionar os vértices não exatamente na fronteira entre uma região e outra, mas sim predominantemente numa das regiões. Foi observado neste experimento que os vértices localizados se encontravam dentro dos retângulos, região escura, e não exatamente na borda. Esta anomalia do operador de Harris faz aumentar os indicadores de erro na calibração pelo método de Tsai (4.21) e no cálculo a matriz fundamental (4.22).

Por outro lado, quando se utiliza pontos do ambiente em muito maior quantidade que os pontos do gabarito a matriz fundamental fica como que incompatível com o gabarito, e observa-se o aumento dos erros de posição na reconstrução do gabarito. A reconstrução é mais consistente quando utiliza-se os vértices que quando utiliza-se os centróides, principalmente pelo melhor balanceamento da quantidade de pontos do gabarito e do ambiente. Em resumo, o erro proveniente do uso dos vértices na calibração é compensado pelo uso da maior quantidade de pontos fornecidos por estes vértices.

Para comparação dos resultados foram tomadas as matrizes de projeção calculadas pelo método de calibração de Tsai. Como, neste experimento, a correção da distorção radial ainda não foi adicionada ao procedimento de cálculo da matriz fundamental a calibração pelo método de Tsai utilizada na comparação também foi realizada sem a correção da distorção. Os indicadores de erro da calibração Tsai são mostrados na tabela 4.21. O caso (d), calibração a partir dos vértices sem correção da distorção radial, é a informação de calibração que será utilizada deste ponto em diante para comparação com reconstrução pela matriz fundamental.

Tabela 4.21: Erro de calibração pelo método de Tsai

Calib.	Dist. radial	Refer.	Câmera 1		Câmera 2	
			UIPE ( <i>pel</i> )	NCE ( <i>mm</i> )	UIPE ( <i>pel</i> )	NCE ( <i>mm</i> )
(a)	Sim	Centróides	0,705	1,737	0,784	1,937
(b)		Vértices	1,750	4,290	1,753	4,300
(c)	Não	Centróides	1,195	2,929	1,250	3,070
(d)		Vértices	1,912	4,685	1,897	4,649

UIPE: erro no plano de imagem não normalizado. NCE: erro de calibração normalizado.

A matriz fundamental foi calculada com três conjuntos de pontos correspondentes. No

primeiro caso somente com os pontos do gabarito de calibração. No segundo caso com os pontos do gabarito e mais uma pequena quantidade de pontos do ambiente. Finalmente a matriz foi calculada com os pontos do gabarito e todos os pontos correspondentes do ambiente em quantidade maior que a dos pontos do gabarito. Os resultados do cálculo nas três condições estão resumidos na tabela 4.22, indicados como (a), (b) e (c), respectivamente. Da mesma forma que na seção anterior, o indicadores de erro são calculados sobre a distância da linha epipolar. Na tabela 4.22, além da indicação do erro de calibração, podemos observar também a variação da distância da linha epipolar somente dos pontos do gabarito e somente dos pontos do ambiente conhecidamente correspondentes.

Tabela 4.22: Erro de calibração da matriz fundamental

Calib.	Pares		Erro de calibração		Erro no gabarito		Erro no ambiente	
	Gabarito	Ambiente	Média	$\sigma$	Média	$\sigma$	Média	$\sigma$
(a)	792	-	0,703	0,508	0,703	0,508	3,409	2,945
(b)	792	136	0,997	1,121	0,782	0,601	2,658	2,783
(c)	792	1085	1,915	1,998	1,678	1,479	1,817	1,894

Pode-se observar na tabela 4.22 que, embora tenha ocorrido um aumento do erro de calibração ao se utilizar os pontos correspondentes do ambiente, foi possível chegar a uma distribuição mais equilibrada entre os erros nos pontos do gabarito e nos pontos do ambiente. Ou seja, a matriz fundamental calculada no caso (c) da tabela é a de maior erro, porém é a única que é igualmente válida tanto na região do espaço ocupada pelo gabarito quanto no resto do ambiente.

Nas figuras 4.10(a) e (b) podem ser observadas as linhas epipolares da matriz fundamental obtida a partir do gabarito (caso (a) da tabela 4.22) e a partir do gabarito mais os pares correspondentes do ambiente (caso (c) da tabela 4.22), respectivamente. Em todas as imagens são exibidas as linhas epipolares que passam por pares de pontos do gabarito de calibração.

O procedimento de reconstrução pode ser resumido como:

1. Calcular matriz fundamental  $F$ ;
2. Calcular matrizes de projeção  $P_i$  compatíveis com  $F$ ;
3. Calcular reconstrução dos pontos do gabarito através das matrizes  $P_i$ ;
4. Calcular matriz de distorção  $D$  a partir das posições conhecidas dos pontos do gabarito

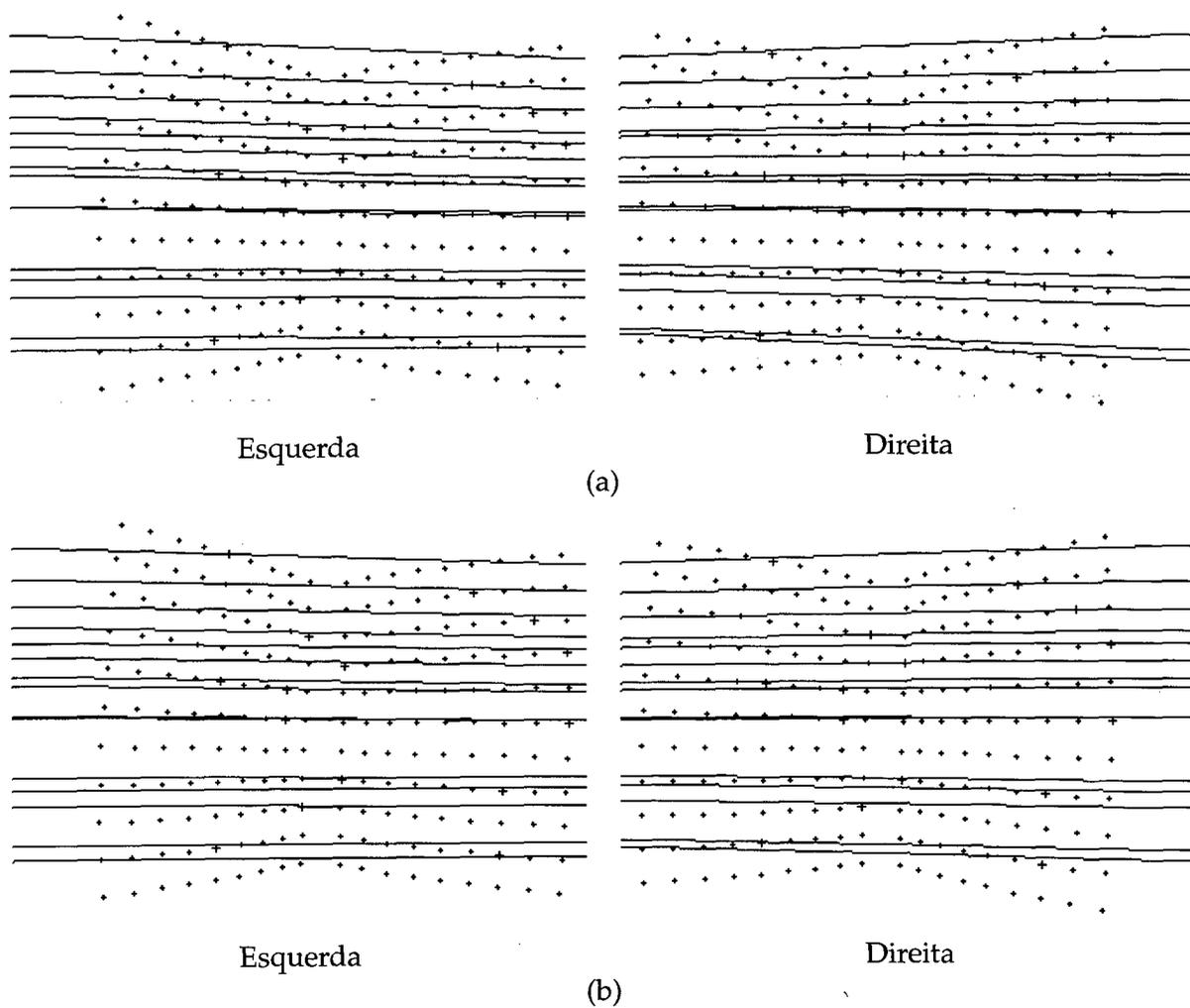


Figura 4.10: Linhas epipolares passando por pontos do gabarito

- e suas posições calculadas;
5. Calcular novas matrizes de projeção, euclidiana,  $M_i$ , para o sistema de coordenadas do gabarito;
  6. Decompor matrizes de projeção em parâmetros internos e externos,  $A_i$ ,  $R_i$  e  $t_i$ ;
  7. Calcular transformação do sistema de coordenadas,  $T$ , do gabarito para o sistema de coordenadas do conjunto binocular (centro de projeção da câmeras esquerda);
  8. Calcular novas matrizes de projeção,  $C_i$ , para o sistema de coordenadas do conjunto binocular.

Este procedimento foi realizado para cada uma das três matrizes fundamentais calculadas anteriormente. A tabela 4.23 mostra os resultados da reconstrução dos pontos do gabarito utilizando as matrizes de projeção  $M_i$  e utilizando as matrizes de projeção calculadas pelo método de Tsai. Os indicadores de erro foram calculados somando-se os erros de posição em cada uma das componentes (x,y,z) do sistema de coordenadas, dadas em milímetros.

Tabela 4.23: Erro de reconstrução do gabarito

Método	Calib. <sup>(1)</sup>	Média	$\sigma$
Tsai	(d)	3,751	1,726
$F$	(a)	4,216	1,675
$F$	(b)	4,392	1,811
$F$	(c)	5,527	3,722

(1) Mesma indicação utilizada nas tabelas 4.21 e 4.22.

O erro de reconstrução crescente que pode ser observado na tabela 4.23 é consistente com os erros observados no cálculo da matriz fundamental da tabela 4.22. Se a matriz fundamental leva em consideração também os pontos correspondentes do ambiente é mais difícil fazer com que a reconstrução se ajuste totalmente ao gabarito.

Foram tomados alguns pontos correspondentes pertencentes a objetos do ambiente e foram medidas, aproximadamente, as distâncias destes objetos até a origem do sistema de coordenadas do conjunto binocular, a exemplo do que é mostrado na tabela 4.14. Para cada um destes pontos foi calculada sua posição utilizando as matrizes de projeção  $C_i$  calculadas a partir da matriz fundamental e a matriz de projeção calculada pelo método de Tsai, também transformando o sistema de coordenadas para o sistema de coordenadas do conjunto binocular. Pode-se calcular então o erro da distância medida e calculada do objeto no am-

biente, o que está resumido na tabela 4.24 e é mostrado com mais detalhes na tabela 4.25. Observe-se que está sendo avaliado o erro da distância do objeto em relação ao conjunto binocular e não o erro de posição. Isto porque a distância pode ser facilmente medida, o que não ocorre com a posição do objeto no espaço em relação ao sistema de coordenadas da câmera.

Tabela 4.24: Erro de distância do ambiente

Método	Calib. <sup>(1)</sup>	Distância		Distância relativa	
		Média	$\sigma$	Média	$\sigma$
Tsai	(d)	405,027	345,993	0,227	0,138
<i>F</i>	(a)	446,434	349,846	0,261	0,125
<i>F</i>	(b)	409,043	321,976	0,242	0,118
<i>F</i>	(c)	264,265	162,144	0,184	0,087

(1) Mesma indicação utilizada nas tabelas 4.21 e 4.22.

Tabela 4.25: Detalhamento do erro distância do ambiente

Ponto	Distância medida	Tsai (d) <sup>(1)</sup>			F (c) <sup>(1)</sup>		
		Calc.	Erro	Erro rel.	Calc.	Erro	Erro rel.
1	600	639,468	39,468	0,066	742,904	142,904	0,238
2	730	854,712	124,712	0,171	964,356	234,356	0,321
3	1170	1133,576	36,424	0,031	1298,105	128,105	0,109
4	1220	1011,022	208,978	0,171	1158,067	61,933	0,051
5	1770	1225,873	544,127	0,307	1468,346	301,654	0,170
6	1900	1151,460	748,54	0,394	1446,773	453,227	0,239
7	2050	1437,831	612,169	0,299	1790,295	259,705	0,127
8	2440	1514,198	925,802	0,379	1907,762	532,238	0,218
Média			405,027	0,227		264,265	0,184
$\sigma$			345,993	0,138		162,144	0,087

(1) Mesma indicação utilizada nas tabelas 4.21 e 4.22.

Pode-se agora observar que a perda introduzida, pelo uso dos pontos correspondentes do ambiente, na reconstrução do gabarito é compensada pelo redução erro de reconstrução dos pontos fora da região em torno do gabarito de calibração.

## 4.7 Resumo

Neste capítulo foram documentados os resultados de experimentos realizados envolvendo correspondência de caracteres, calibração de câmeras utilizando os métodos clássicos e utilizando a geometria epipolar.

Na avaliação dos métodos de correspondência de caracteres observou-se que o método de correspondência por relaxação levou a resultados melhores, quando comparado com o método dos valores singulares.

Na calibração de câmeras pôde-se observar a influência que a correção da distorção radial pode ter na calibração das câmeras. Também foi comparado o resultado da calibração pelo método clássico com o resultado da calibração pela geometria epipolar, em ambos os casos sem correção da distorção radial das lentes.

O método de calibração pela geometria epipolar alcançou pelo menos a mesma qualidade de reconstrução alcançada pelo método clássico.

## Capítulo 5

# Conclusões e sugestões

*"A conclusion is the place where you got tired of thinking."*

*Arthur Block*

### 5.1 Conclusões

Foram utilizados dois métodos de calibração de câmeras, um método que se baseia somente nos pontos cujas coordenadas no espaço são conhecidas, e outro que se baseia em todos os pontos, com coordenadas conhecidas ou não. Os pontos conhecidos, em ambos os casos, pertencem a um gabarito de calibração.

Quando a calibração é feita somente com o gabarito, o erro médio e o desvio padrão da reconstrução do gabarito de calibração em ambos os métodos são aproximadamente iguais. No segundo método, porém, observa-se um ligeiro aumento do erro de reconstrução do gabarito quando os pontos do ambiente são adicionados à calibração. Considera-se que esta perda de precisão no gabarito é compensada por um ganho de precisão na reconstrução do resto do ambiente.

Para dimensionar este ganho de precisão no resto do ambiente, em ambos os métodos, foi avaliada a reconstrução de pontos que não pertenciam ao gabarito. Para estes pontos não são conhecidas as coordenadas no espaço, mas somente a distância dos objetos até o conjunto binocular, estimada a partir de medições no ambiente. Esta distância medida foi comparada com a distância calculada a partir das coordenadas fornecidas pelo conjunto binocular.

Quando utilizamos somente os pontos do gabarito na calibração o erro relativo médio na distância dos pontos do ambiente são menores no primeiro método. Porém, no segundo método podemos ainda acrescentar à calibração os pontos do ambiente e, fazendo assim, este segundo método de fato supera o primeiro obtendo-se erros menores.

Todavia, esperava-se que este acréscimo de informações do ambiente na calibração levasse a resultados ainda melhores que os obtidos. Para os pontos avaliados, variando numa faixa de 600mm a 2440mm, a média de erro na distância foi de 405mm no primeiro método e 264mm no segundo. O erro relativo médio caiu de 0,22 no primeiro método para 0,18 no segundo método.

Considera-se, mesmo assim, que os resultados obtidos são suficientes para a aplicação estudada pois na vizinhança do conjunto binocular os erros são menores, isto porque o erro é proporcional à distância. Os robôs móveis deverão interagir diretamente somente com outros robôs ou objetos mais próximos, bastando, para aqueles mais distantes, somente uma medida aproximada de posição e orientação relativa.

A figura 5.1 mostra um esboço do mecanismo de calibração implementado no segundo método.

Quanto aos objetivos descritos no início deste trabalho, considera-se que foram atingidos do seguinte modo:

- Os tipos de caracteres primitivos escolhidos, vértices e centróides de regiões, mostraram-se adequados à calibração das câmeras, merecendo somente ser observado que os vértices de fato foram deslocados para uma das regiões, como previsto, aumentando os erros de calibração, e que os centróides, apesar de considerados não invariantes quanto à projeção em perspectiva, puderam ser utilizados satisfatoriamente devido às características geométricas das regiões;
- A correspondência entre os pontos do gabarito de calibração e os caracteres detectados nas imagens pode ser feita pela ordenação dos caracteres conforme suas coordenadas na imagem, satisfazendo uma restrição adicional de ajuste à equação de reta;
- Os métodos de correspondência de caracteres entre imagens avaliados, incorporando medidas de vizinhança, de similaridade e, durante a relaxação, de gradiente de disparidade, levaram a erros de correspondência menores que o máximo de tolerância aceito pelo método robusto de calibração; em alguns experimentos o erro de correspondência

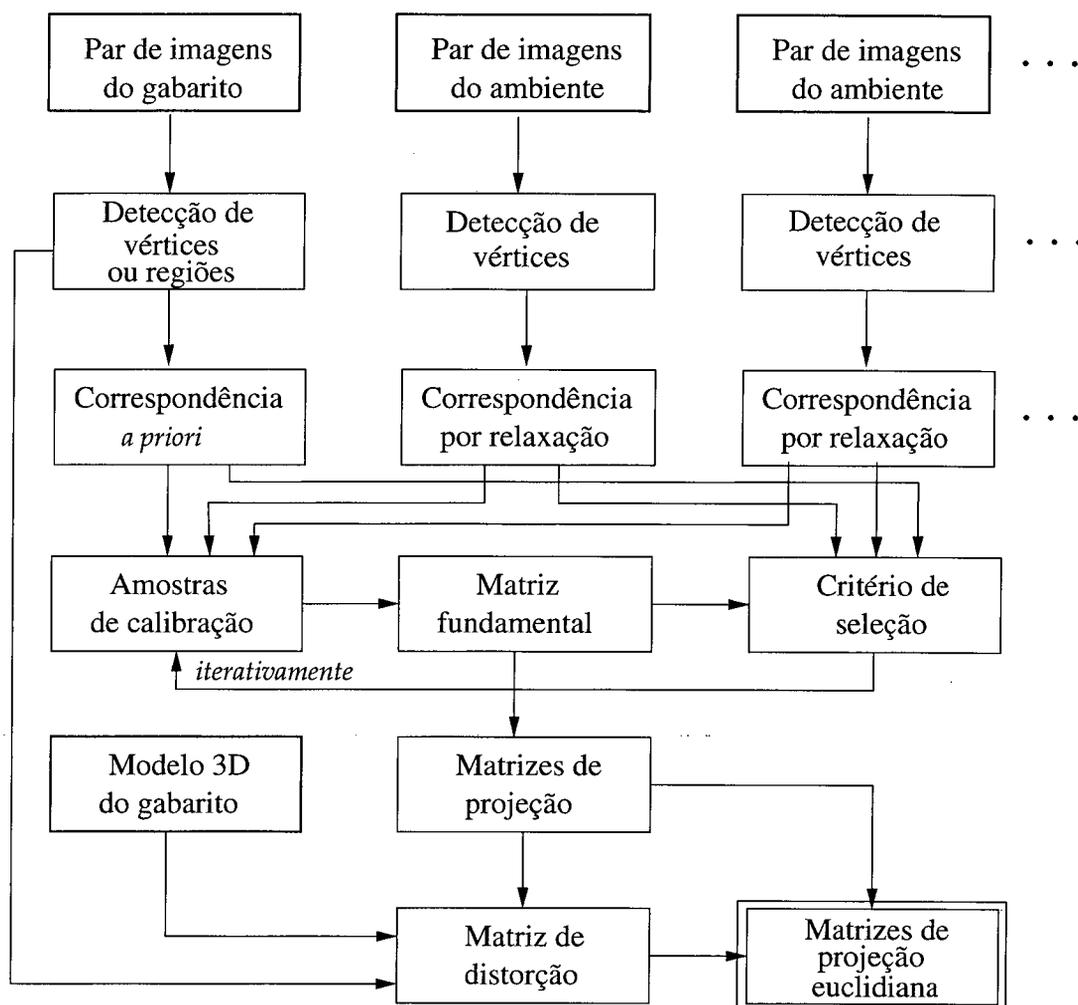


Figura 5.1: Mecanismo de calibração

foi menor que cinco por cento;

- Foi possível determinar a geometria do conjunto binocular sem conhecimento *a priori* de quaisquer características das câmeras ou do conjunto binocular, como distância focal ou distância da linha de base; para isso utilizou-se todas as informações disponíveis, tanto da correspondência entre o gabarito e as imagens quanto da correspondência entre as duas imagens, incluindo pontos não pertencentes ao gabarito;
- Com a informação de calibração obtida pode-se determinar as coordenadas dos pontos do gabarito com erro médio de 5,5mm; não estavam disponíveis medidas precisas de posição de pontos não pertencentes ao gabarito, para avaliação da calibração foi utilizada a distância medida dos objetos, obtendo-se erros menores que os obtidos

pelo método clássico de calibração.

É importante, porém, destacar como principal conclusão alcançada através da elaboração deste trabalho o fato de que a introdução de pontos do ambiente no processo de calibração não reduziu sensivelmente o erro na distância dos objetos fora da vizinhança do gabarito.

Considera-se como principal motivo os erros acumulados de posição no gabarito de calibração, construído de material sujeito a deformações, de posição dos caracteres na imagem que estavam sujeitos a deslocamento imposto pelo processo de detecção, e de distorções das lentes que não foram corrigidas no método utilizado.

## 5.2 Perspectivas futuras

Embora este trabalho, como citado em seu início, esteja inserido num estudo maior relacionado à robótica móvel e à RoboCup, os conceitos, estratégias e resultados apresentados aqui são de utilização geral e poderia ser utilizados diretamente em problemas de posicionamento de robôs manipuladores, reconhecimento de peças (como o caso RAP [21]) ajudando a resolver o problema da sobreposição.

No caso RoboCup, a estratégia de calibração e reconstrução estudada pode ser associada aos demais componentes do sistema de visão computacional para auxiliar tanto no reconhecimento quanto na localização dos objetos e do próprio robô móvel. Os principais componentes envolvidos são aqueles relacionados com a segmentação de regiões e construção de mapas locais, como mostrado na figura 5.2.

Neste trabalho, os métodos de calibração foram testados com a configuração do conjunto binocular constante. Ambos os métodos podem ser adaptados para casos onde a configuração do conjunto binocular é variável, ou seja, quando cada câmera possui controle motorizado de foco ou aproximação, ou controle sobre a geometria do conjunto binocular como, por exemplo, elevação e orientação das câmeras. Qualquer mudança na configuração do conjunto binocular pode invalidar todas as informações de calibração disponíveis, porém, se são registradas as coordenadas no espaço dos caracteres detectados na última configuração e se alguns destes caracteres puderem ser localizados nas imagens obtidas na nova configuração então é possível fazer uma nova calibração utilizando a posição daqueles caracteres como referência da mesma forma como foi utilizado o gabarito de calibração no primeiro instante.

De outro modo, se não estão disponíveis pontos com coordenadas no espaço conhecidas

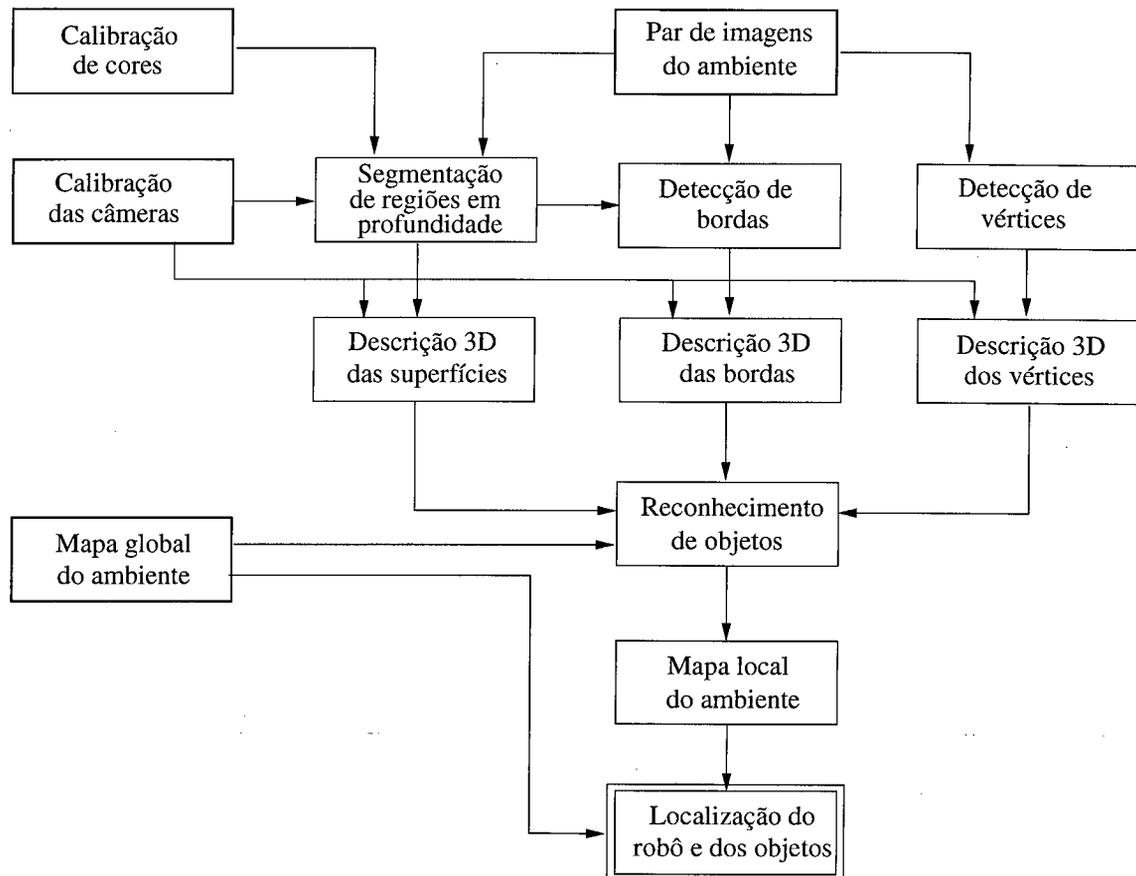


Figura 5.2: Integração da calibração ao sistema de visão

para calibração mas pode-se desprezar a reconstrução métrica do ambiente, então, mesmo assim, pode-se obter uma calibração parcial utilizando o segundo método e uma reconstrução do ambiente no espaço projetivo.

Ainda, mesmo quando não for interessante obter uma reconstrução do ambiente, pode-se utilizar o segundo método de calibração para determinar a geometria epipolar do conjunto binocular e utilizar esta informação para reduzir a ambigüidade de correspondência entre caracteres detectados nas duas imagens.

Para o aperfeiçoamento do método de calibração pela geometria epipolar, apresenta-se ainda como sugestão a introdução da correção da distorção radial das lentes que, no conjunto binocular estudado, possuem grande influência sobre os erros de reconstrução.

# Apêndice A

## Plataforma de software utilizada

### A.1 Ambiente de desenvolvimento

As aplicações e utilitários desenvolvidos como suporte a este trabalho têm como ambiente nativo o sistema operacional GNU/Linux. As principais ferramentas utilizadas no desenvolvimento estão enumeradas na tabela A.1.

### A.2 Bibliotecas de funções

Os procedimentos computacionais relacionados à visão computacional já vêm sendo estudados a vários anos e grande parte destes procedimentos estão disponíveis na forma de código fonte em linguagens de uso geral, como C/C++, ou linguagens de processamento numérico, como Fortran. Algumas das bibliotecas de funções utilizadas ao longo deste trabalho estão relacionadas na tabela A.2.

Tabela A.1: Principais ferramentas do ambiente de desenvolvimento

Produto	Versão	Autor
Linux	kernel 2.2.17	Linus Torvalds et. al.
GNU Make	3.79	Richard Stallman e Roland McGrath
GNU Compiler Collection	2.93.5	Free Software Foundation, Inc.
GNU Fortran	0.5.25	Free Software Foundation, Inc.
Qt GUI Toolkit	1.45	Troll Tech AS.
Qt Architect	1.4	Jeff Harris e Klaus Ebner
TMake	1.3	Troll Tech AS.

Tabela A.2: Principais bibliotecas de funções utilizadas

Produto	Versão	Descrição
LAPACK Linear Algebra Routines Package	3.0	Funções e estruturas de dados para álgebra linear, como minimização LS e decomposição QR e SVD. Univ. of Tennessee, Univ. of California Berkeley, NAG Ltd., Courant Institute, Argonne National Lab, e Rice University.
KLT Kanade-Lucas-Tomasi Feature Tracker	1.1.5	Funções para detecção e rastreamento de caracteres entre imagens obtidas a partir de uma mesma câmera. Stan Birchfield, Leland Stanford Junior Univeristy.
Tsai Camera Calibration	3.0b3	Funções e estruturas de dados para calibração de câmeras a partir de um conjunto de pontos conhecidos no ambiente. Reg Willson.
VIP Visual Image Processing	4.0	Funções para manipulação e transformação de imagens, incluindo operações aritméticas em imagens, estatísticas sobre imagens e operações em matrizes e vetores. Peter Kovesi et. al., University of Western Australia.
Neural	—	Funções e estruturas de dados para treinamento e ativação de redes neurais artificiais. Timothy Masters [20].
Video4Linux	—	Mecanismo que fornece uma interface padronizada, independente de dispositivo, para consultas ao hardware e captura de imagens, permitindo selecionar a origem do sinal de vídeo e o formato de cor. Alan Cox.

### A.3 Aplicações e utilitários desenvolvidos

Vários programas foram utilizados ao longo das pesquisas. O principal deles é um ambiente integrado para captura de imagens, configuração de calibração, detecção de vértices e calibração de câmeras utilizando o método de Tsai (conforme seção 3.4.2). Outros programas foram utilizados para detecção e correspondência de caracteres, e cálculo das matrizes fundamental, essencial e de projeção das câmeras. Estes programas estão relacionados na tabela A.3.

Tabela A.3: Principais programas desenvolvidos

Programa	Descrição
estereo	Permite a configuração da captura de vídeo, selecionando a origem como sendo uma imagem em arquivo ou uma placa de captura (figura A.1); pode-se configurar o tipo de caractere a ser detectado na imagem (vértices ou regiões aproximadamente retangulares) e a calibração das câmeras utilizando o método de Tsai (figura A.2); embora não diretamente relacionado com o objeto deste trabalho, este programa ainda permite a calibração e classificação de cores previamente selecionadas nas imagens e rotuladas utilizando redes neurais (figura A.3).
mkplano	Gera uma lista de coordenadas tridimensionais para pontos pertencentes ao gabarito de calibração coplanar; estes pontos podem ser os centróides ou os vértices das regiões do gabarito e são calculados a partir das dimensões conhecidas do gabarito .
mk3d	Gera uma lista de coordenadas tridimensionais para pontos pertencentes ao gabarito de calibração não coplanar; estes pontos podem ser os centróides ou os vértices das regiões do gabarito e são calculados a partir das dimensões conhecidas do gabarito .
undistort	Dada uma imagem e uma descrição da câmera através da qual a imagem foi obtida este programa calcula uma outra imagem corrigindo a distorção radial conforme os coeficientes de distorção da câmera.

extract	Detecta vértices ou regiões distinguíveis nas imagens, utilizando o operador de Harris ou o operador KLT, respectivamente; gera uma lista de coordenadas de cada caractere selecionado incluindo os valor do indicador de vértice, da intensidade e do indicador de região distinguível da coordenada.
mfromcd	Dadas duas listas de caracteres detectados em duas imagens na calibração pelo método de Tsai, este programa gera uma lista das correspondências entre as duas imagens; esta correspondência é determinada diretamente pela posição dos caracteres detectados.
singular	Dadas duas listas de caracteres detectados pelo programa <code>extract</code> , este programa seleciona os melhores candidatos a pares correspondentes utilizando o método de decomposição por valores singulares (conforme citado na seção 2.3).
relax2	Dadas duas listas de caracteres detectados pelo programa <code>extract</code> , este programa seleciona os melhores candidatos a pares correspondentes utilizando o método de relaxação baseado na confirmação da vizinhança (conforme seção 2.3).
matcheval	Dadas duas imagens e uma lista de pares de caracteres correspondentes gerada pelos programas <code>singular</code> ou <code>relax2</code> , este programa permite a confirmação visual da correspondência entre estes caracteres e o cálculo da taxa de acerto obtida (figura A.4)
ffromp	Dadas as descrições de duas câmeras, obtida pelo método de calibração de Tsai, por exemplo, este programa permite calcular a matriz fundamental do conjunto binocular.
ffromm	Este programa calcula a matriz fundamental do conjunto binocular pelo método LMedS a partir de uma lista de caracteres correspondentes gerada pelos programas <code>mfromcd</code> , <code>singular</code> ou <code>relax2</code> .
reduceby	Dada uma matriz fundamental e uma lista de caracteres correspondentes este programa remove da lista os pares de caracteres menos compatíveis com a matriz fundamental informada (ver seção 4.5).

ffromall	Este programa calcula a matriz fundamental do conjunto binocular utilizando todas os pares de caracteres fornecidos por uma lista de caracteres correspondentes gerada pelos programas mfromcd, singular ou relax2.
pfromf	Este programa calcula as matrizes de projeção de duas câmeras a partir de uma matriz fundamental e duas listas de caracteres detectados em duas imagens para obtidas com estas câmeras; estes caracteres devem ser correspondentes e devem ser conhecidas suas coordenadas tridimensionais, podendo ser obtidos a partir de um gabarito de calibração.

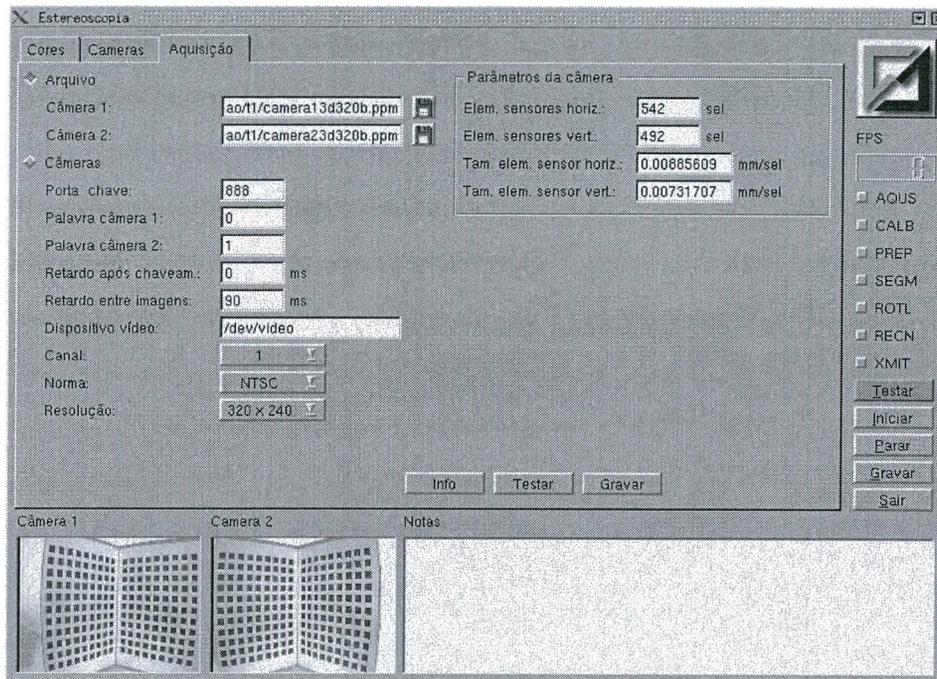


Figura A.1: Aquisição de imagem com o programa estereo

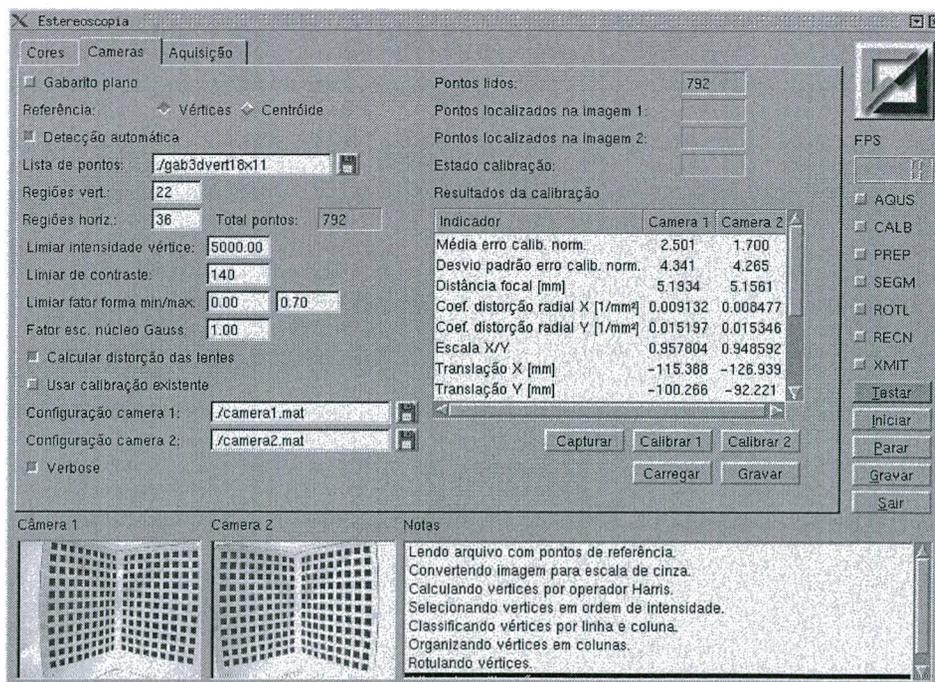


Figura A.2: Calibração de câmeras pelo método de Tsai usando o programa estereo

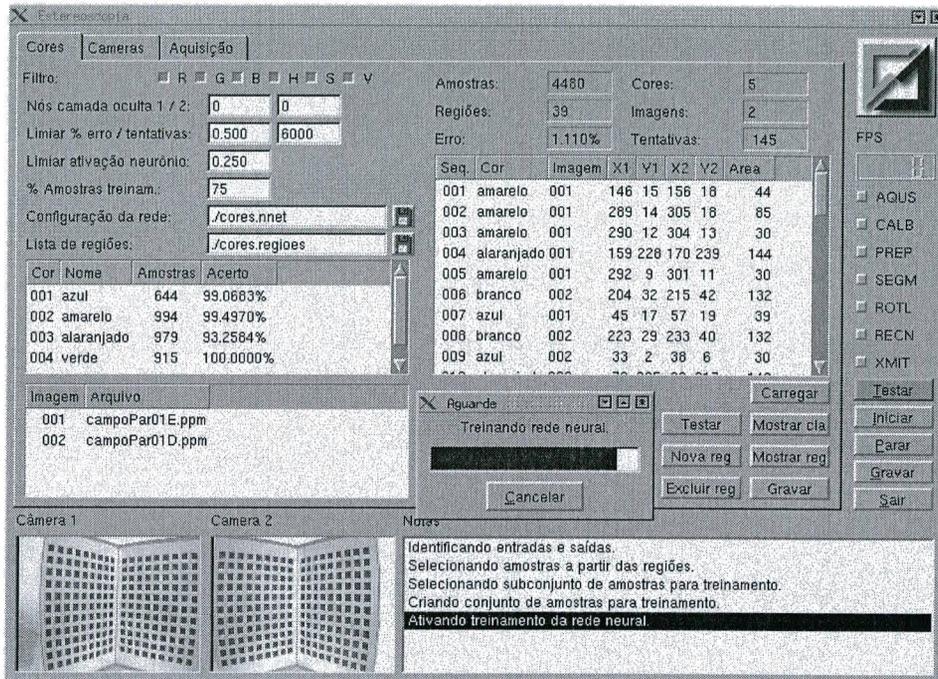


Figura A.3: Calibração de cores com o programa estereo

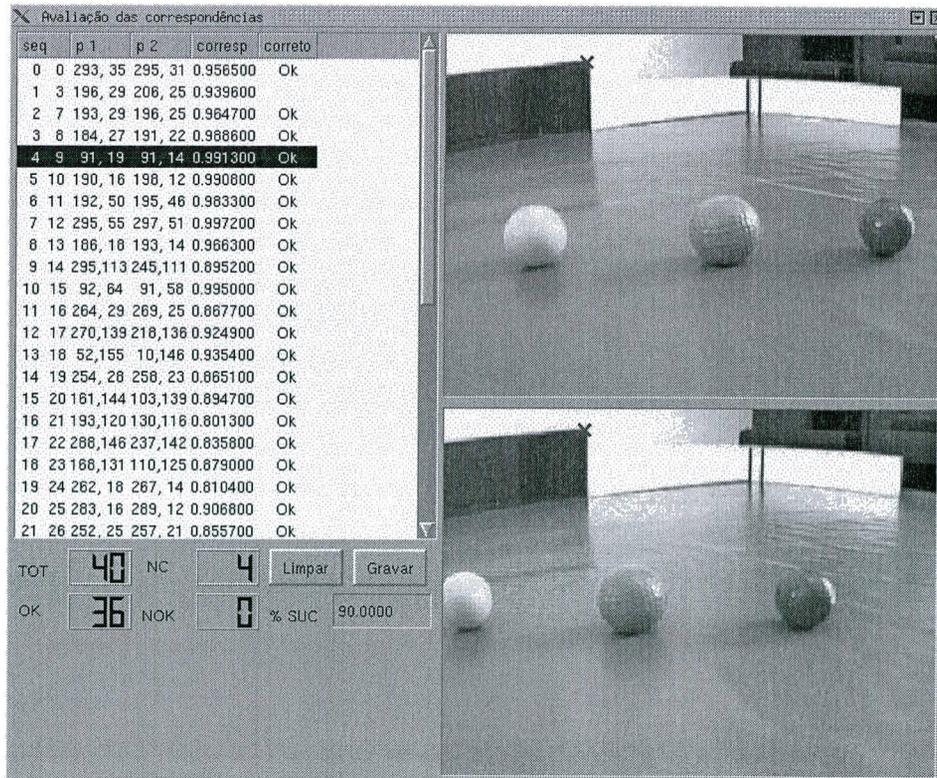


Figura A.4: Avaliação do mecanismo de correspondência de caracteres

# Referências Bibliográficas

- [1] BIRCHFIELD, Stanley Thomas. An introduction to projective geometry (for computer vision), 1998. URL <http://www.vision.stanford.edu/~birch/projective>. Documento Internet.
- [2] —. Klt an implementation of the kanade-lucas-tomasi feature tracker, 1998. URL <http://www.vision.stanford.edu/~birch/klt>. Documento Internet.
- [3] —. *Depth and motion discontinuities*. Tese de Doutorado, Department of Electrical Engineering - Stanford University, Stanford, 1999.
- [4] BORENSTEIN, Johann; *et al.*. *Where am I?: sensors and methods for mobile robot positioning*. Michigan: University of Michigan, 1996.
- [5] CHRISTENSEN, Henrik I. *The AUC robot camera head*. Rel. Téc. 9518, Laboratory of Image Analysis - Aalborg University, Aalborg, 1995. URL <http://www.vision.auc.dk/~hic/auc-head-rptp.ps>.
- [6] COSTA, Augusto Cesar Pinto Loureiro da. Ufsc-team, 1999. URL <http://www.lcmi.ufsc.br/ufsc-team>. Homepage oficial do UFSC-Team.
- [7] EDMUND INDUSTRIAL OPTICS, Inc. 2001 optics and optical instruments catalog, 2001.
- [8] FU, K. S. [King-Sun]; GONZÁLEZ, R. C.; LEE, C. S. G. *Robótica: control, detección, visión e inteligencia*. México: McGraw-Hill, 1988.
- [9] GOLUB, Gene H.; VAN LOAN, Charles F. *Matrix computations*. s.e? ed. Baltimore: Johns Hopkins University Press, s.d?
- [10] GOMES, Jonas; VELHO, Luiz. *Computação gráfica: imagem*. Rio de Janeiro: IMPA/SBM, 1994.

- [11] GRACIAS, Nuno; SANTOS-VICTOR, José. Robust estimation of the fundamental matrix and stereo correspondences, 1997.
- [12] GRAHAM, Deryn; BARRET, Anthony. *Knowledge-based image processing systems*. London: Springer-Verlag, 1996.
- [13] HARALICK, Robert M.; SHAPIRO, Linda G. *Computer and robot vision*, volume 2. [s.l.]: Addison-Wesley, 1993.
- [14] HORN, Berthold Klaus Paul. *Robot vision*. Cambridge: MIT Press, 1986.
- [15] HUYNH, Du Quan. *Feature-based stereo vision on a mobile platform*. Tese de Doutorado, Department of Computer Science and Software Engineering - University of Western Australia, Crawley, 1994.
- [16] JONKER, Pieter (coord.). *Image processing fundamentals*, 1998. URL <http://www.ph.tn.tudelft.nl/Courses/FIP>.
- [17] KITANO, Hiroaki; *et al.*. Robocup: the robot world cup initiative. In: *Autonomous Agents '97: Online Proceedings*, ed. MÜLLER, Jörg. New York: SIGART/ACM, 1997. URL <http://sigart.acm.org/proceedings/agents97/A172/A172.PDF>. Documento Internet.
- [18] KÖTHE, Ullrich. *VIGRA: vision with generic algorithms*. Fraunhofer Institut fuer Graphische Datenverarbeitung, 1998. URL <http://www.egd.igd.fhg.de/~ulli/vigra>.
- [19] MARSHALL, Dave. *Vision systems*, c1997. URL [http://www.cs.cf.ac.uk/Dave/Vision\\_lecture](http://www.cs.cf.ac.uk/Dave/Vision_lecture). Documento Internet.
- [20] MASTERS, Timothy. *Practical Neural Networks Recipes in C++*. Boston: Academic Press, 1993.
- [21] ORTH, Alexandre. Desenvolvimento e implementação de um sistema de reconhecimento automático de peças mecânicas em uma célula flexível de manufatura: projeto rap. In: *Congresso da Sociedade Brasileira de Computação (SBC'98)*. Belo Horizonte, 1998. URL <http://www.lcmi.ufsc.br/~ferdes/rap/Download/PaperSBC98.zip>.
- [22] PILU, Maurizio; LORUSSO, Adele. *Uncalibrated stereo correspondence by singular value decomposition*. Rel. Téc. HPL-97-96, Hewlett-Packard Research Labs, Bristol, 1997. URL <http://www.hpl.hp.com/techreports/97/HPL-97-96.html>.

- [23] PRICE, Sarah; FISHER, Bob; FITZGIBBON, Andrew; *et al.*. The marble project, 1996. URL [http://www.dai.ed.ac.uk/CVonline/LOCAL\\_COPIES/MARBLE](http://www.dai.ed.ac.uk/CVonline/LOCAL_COPIES/MARBLE). Documento Internet.
- [24] ROSANDISH, Ryan G. *Intelligent visual inspection: using artificial neural networks*. Duluth: Chapman & Hall, 1997.
- [25] ROTHWELL, Charlie; CSURKA, Gabriella; FAUGERAS, Oliver. *A comparison of projective reconstruction methods for pairs of views*. Rel. Téc. RR-2538, INRIA Sophia-Antipolis, Sophia-Antipolis, 1995.
- [26] ROWE, Neil C.; FREW, Brian. *Automatic classification of objects in captioned depictive photographs for retrieval*. Rel. Téc. s.n., U.S. Naval Postgraduate School, Monterey, s.d. URL <http://www.cs.nps.navy.mil/research/marie/roweshap.html>.
- [27] SANTOS-VICTOR, José; VAN TRIGT, Franc; SENTIEIRO, João. *MEDUSA: a stereo head for active vision*. Rel. Téc. s.n., Instituto Superior Técnico, Lisboa, 1994.
- [28] SKARBEEK, Władysław; KOSCHAN, Andreas. *Colour image segmentation: a survey*. Rel. Téc. 94-32, Technischer Universität Berlin, Berlin, 1994.
- [29] SONKA, Milan. 55:148 digital image processing, 2000. URL [css.engineering.uiowa.edu/~dip/LECTURE](http://css.engineering.uiowa.edu/~dip/LECTURE).
- [30] TOMASI, Carlo. CS205: Mathematical methods for robotics vision, 2000. URL <http://www.stanford.edu/class/cs205/notes/book.pdf>. Documento Internet.
- [31] TOMASI, Carlo; SHI, Jianbo. Good features to track. In: *CVPR94 IEEE Conference on Computer Vision and Pattern Recognition*. Seattle, 1994.
- [32] TORREÃO, José Ricardo de Almeida. *Métodos estocásticos em computação visual*. Porto Alegre: Instituto de Informática da UFRGS, 1992.
- [33] TSAI, Roger 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*, 1987. volume RA-3, nº 4, págs. 323–344.

- [34] WILLSON, Reg G. Camera calibration using tsai's method - revision 3.0b3, 1995. URL <http://www.ius.cs.cmu.edu/afs/cs.cmu.edu/user/rgw/www/TsaiCode.html>.
- [35] YOUNG, David. Stereoscopic vision and perspective projection, 1994. URL <http://www.cogs.susx.ac.uk/users/davidy/teachvision>.
- [36] ZHANG, Zhengyou. *A flexible new technique for camera calibration*. Rel. Téc. MSR-TR-98-71, Microsoft Corporation, Redmond, 1998. URL <http://research.microsoft.com/~zhang>.
- [37] ZHANG, Zhengyou; FAUGERAS, Olivier; DERICHE, Rachid. An effective technique for calibrating a binocular stereo through projective reconstruction using both a calibration object and the environment. *VIDERE: Journal of Computer Vision Research*, 1997. volume 1, nº 1. URL <http://mitpress.mit.edu/videre.html>.