

ELAINE GRACINDO ALVES

**Desenvolvimento de Algoritmos de Calibração de
Câmeras para Aplicações em Robótica**

LONDRINA

2007

UNIVERSIDADE ESTADUAL DE LONDRINA
CURSO DE PÓS-GRADUAÇÃO EM ENGENHARIA
ELÉTRICA

Desenvolvimento de Algoritmos de Calibração de Câmeras para
Aplicações em Robótica

Dissertação de mestrado submetido à
Universidade Estadual de Londrina
como parte dos requisitos para a obtenção
do grau de mestre em Engenharia Elétrica.

ELAINE GRACINDO ALVES

Londrina, Abril de 2007.

Desenvolvimento de Algoritmos de Calibração de Câmeras para Aplicações em Robótica

Elaine Gracindo Alves

‘Este trabalho foi julgado adequado para a obtenção do título de mestre em engenharia elétrica e aprovado em sua forma final pela Coordenação do Curso de Pós-Graduação em Engenharia Elétrica da Universidade Estadual de Londrina.’

José Alexandre de França
Orientador

Taufik Abrão
Coordenador do Programa de Pós-Graduação em Engenharia Elétrica

Banca Examinadora:

Prof. Dr.-Ing. Marcelo Ricardo Stemmer

Prof. Dr. Robinson Samuel Vieira Hoto

A Deus, que sempre esteve comigo, dos momentos bons aos difíceis. Para aqueles que amo, meus pais, José e Eunice, meu noivo Diógenes, minha irmã Erci, meus irmãos, cunhadas e sobrinhos, pelo incentivo nas horas de desânimo, pela compreensão, confiança e carinho.

AGRADECIMENTOS

A Deus, pela vida, proteção, ajuda, oportunidade e condições de estar concluindo este curso.

Aos meus pais, José e Eunice, que sempre acreditaram em mim, me dando muito amor e carinho.

Ao meu noivo, Diógenes, pela compreensão, paciência, incentivo e apoio sempre dado.

À minha irmã Erci, meus irmãos Ednir, Ednilson, Ederson, cunhadas Marisa, Mara e Valquíria, cunhado Eliseu, sobrinhos Moisés, Maellin, Monara, Taini e Camilly, por acreditarem em mim, e me incentivarem em todo momento.

Ao Prof^o Dr. José Alexandre de França, por me atender sempre com simpatia, incentivando, apoiando e ajudando no desenvolvimento deste trabalho.

Aos professores do Departamento de Engenharia Elétrica, que sempre colaboraram para o meu aperfeiçoamento.

Enfim, a todos aqueles que contribuíram direta ou indiretamente para a conclusão deste trabalho, meus sinceros agradecimentos e respeito.

Resumo da dissertação apresentada à UEL como parte dos requisitos necessários para obtenção do grau de mestre em Engenharia Elétrica.

Desenvolvimento de Algoritmos de Calibração de Câmeras para Aplicações em Robótica

Elaine Gracindo Alves

ABRIL/2007

Orientador: José Alexandre de França

Área de Concentração: Simulação de Sistemas

Palavras-chave: Visão Computacional; Calibração de Câmeras; Gabarito 1D; Calibração Automática

Nos dias de hoje, a visão computacional é muito utilizada em diversas áreas do conhecimento humano como, por exemplo, medicina, diversão, automação e robótica. Em aplicações como metrologia e robótica, é necessário entender como a câmera utilizada forma a imagem. Isto pode ser entendido conhecendo-se os parâmetros intrínsecos e extrínsecos da câmera, através de um processo chamado calibração.

Neste trabalho, avaliou-se melhor a técnica proposta recentemente de calibração de câmeras com auxílio de um gabarito de uma única dimensão. Foi proposta uma normalização nas projeções dos pontos do gabarito que torna o método mais robusto a ruídos. Além disso, realizou-se testes mais detalhados e confiáveis, quando comparado aos disponíveis na literatura, para avaliar o desempenho do método quanto ao número de imagens utilizadas, o número de pontos do gabarito e o nível de ruídos presente nos dados. Isto foi necessário para provar que o método pode ser realmente utilizado na prática e que a normalização proposta produz de fato aumento na exatidão dos resultados.

No caso da visão binocular, a calibração pode ser realizada de forma automática, pois nesta configuração, restrições extras tornam linear a solução do problema de calibração. Em vista disso, está sendo construído um sistema de visão binocular que possa ser calibrado automaticamente. Para isto, o conjunto de câmeras deve realizar uma série de deslocamentos enquanto a distância relativa das câmeras permanecerá constante. Neste trabalho, apresenta-se o algoritmo que permitirá calibrar o sistema construído de forma automática.

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

Camera Calibration Algorithmics Development to Robotics Applications

Elaine Gracindo Alves

APRIL/2007

Advisor: José Alexandre de França

Area of Concentration: Systems Simulation

Keywords: Computer Vision; Camera Calibration; 1D Pattern; Selfcalibration

Nowadays, the computational vision is very used in many areas in the human knowledge as, medicine, entertainment, automation and robotics. In applications as 3D reconstruction and robotics, it is necessary to understand how the camera forms the image. This can be understood knowing the intrinsic and extrinsic parameters of the camera, through a process called calibration.

In this work, the camera calibration's technique was evaluated with an one dimension known pattern's help. It's proposed a normalization in the projections of the known pattern's points answer that becomes the method most robust in presence of noise. Besides, it was realized more detailed and trustworthy tests, when compared with the available of the ones in literature, to evaluate the performance of the method how much to the number of used images, the number of points of the answer sheet and the present level of noises in the data. This was necessary to prove that the method really can be used in practical, and that the normalization proposed produces increase in the exactness of the results.

In the stereo vision's case, the calibration can be realized through of automatic form, therefore in this configuration, extra restrictions become linear the solution of the calibration problem. For this reason, it is being constructed a stereo vision's system that can be calibrated automatically. For this, the set of cameras must realize a lot of displacement while the distance relative of the cameras will remain constant. In this work, is presented an algorithm that could calibrate the constructed system in the automatic form.

Conteúdo

Lista de Figuras	ix
Lista de Tabelas	xi
Lista de Símbolos e Abreviações	xii
1 Introdução	1
1.1 Organização do trabalho	2
1.2 Notações	2
2 Fundamentação Teórica	4
2.1 Modelo da Câmera	4
2.2 Visão Binocular	7
2.2.1 Geometria Epipolar	7
2.2.2 Matriz Fundamental	8
2.3 Correspondência de Pontos	12
2.3.1 Métodos locais	13
2.3.2 Métodos Globais	15
2.3.3 Algoritmo utilizado	15
2.4 Calibração da Câmera	18
2.4.1 Método Clássico	18
2.4.2 Método de Tsai	19
2.4.3 Método Não-Linear	19
2.4.4 Calibração Monocular	20
2.5 Calibração Automática	23
2.5.1 Calibração Projetiva	24
2.5.2 Autocalibração por Etapas	26

3	Contribuições	33
3.1	Calibração com gabarito 1D	33
3.1.1	Normalização dos Dados	33
3.1.2	Considerando a Distorção Radial	34
3.2	Calibração Automática	35
4	Avaliação dos Algoritmos	37
4.1	Calibração 1D	37
4.1.1	Dados Sintéticos	37
4.1.2	Dados Reais	41
4.2	Calibração Automática	43
4.2.1	Deslocamento Planar	44
4.2.2	Deslocamento Euclideano	46
5	Conclusões e Trabalhos Futuros	47
	Bibliografia	49

Lista de Figuras

2.1	Modelo de câmera <i>pinhole</i>	4
2.2	Modelo <i>pinhole</i> e seus sistemas de coordenadas	5
2.3	Determinação da ordenada do ponto m em função das coordenadas do ponto M	5
2.4	Geometria Epipolar.	7
2.5	Exemplo de oclusão. Os pontos $P2$ são visíveis por ambas as câmeras, sendo que os pontos $P1$ são visíveis somente para as câmeras $C1$	13
2.6	Exemplo de transformadas posto e censo. Nota-se que todas as imagens tem a mesma transformada posto, e diferentes transformada censo.	14
2.7	Correspondência de pontos por correlação.	16
2.8	Um par (m_{1i}, m_{2j}) candidato à correspondência com vizinhos dentro de uma janela de lado $2l_n$. O par (n_{11}, n_{21}) aumenta a confiança do par (m_{1i}, m_{2j})	17
2.9	Gabarito utilizado na calibração de câmeras.	20
2.10	Esboço do gabarito 1D utilizado na calibração.	21
2.11	Deslocamento euclidiano desconhecido de um conjunto binocular.	27
3.1	Ilustração da “cabeça binocular” que está sendo desenvolvida para calibração do conjunto binocular.	36
4.1	Esboço do movimento realizando pelo gabarito 1D de três pontos durante uma simulação.	38
4.2	Erro relativo obtido na estimação dos parâmetros intrínsecos da câmera pelo método linear com dados normalizados.	39
4.3	Erro relativo obtido na estimação dos parâmetros intrínsecos da câmera pelo método linear com dados sem normalização.	39
4.4	Erro relativo obtido na estimação dos parâmetros intrínsecos da câmera pelo método não-linear inicializado com parâmetros obtidos com dados sem normalização.	40
4.5	Erro relativo obtido na estimação dos parâmetros intrínsecos da câmera pelo método linear com dados normalizados.	41

4.6	Erro relativo obtido na estimação dos parâmetros intrínsecos da câmera pelo método linear com dados normalizados.	42
4.7	Gabarito 1D utilizado na avaliação do método proposto.	43
4.8	Gabarito plano utilizado na comparação do método proposto.	43
4.9	Três exemplares da seqüência de imagens-teste utilizada na avaliação do algoritmo de calibração automática com deslocamentos planares.	44
4.10	Esboço da reconstrução do gabarito realizada após a calibração automática do conjunto binocular.	45
4.11	Três exemplares da seqüência de imagens-teste utilizada na avaliação do algoritmo de calibração automática com deslocamentos euclidianos.	46

Lista de Tabelas

4.1	Parâmetros intrínsecos estimados pelo algoritmo linear.	42
4.2	Parâmetros intrínsecos estimados pelo algoritmo não-linear proposto.	43
4.3	Resultados experimentais da calibração do conjunto binocular que realizou deslocamentos planares. Observe que apenas os parâmetros ZHANG são valores absolutos. Os outros resultados são dados na forma do erro relativo (em percentagem) aos parâmetros ZHANG	45
4.4	Resultados experimentais da calibração do conjunto binocular que realizou deslocamentos euclidianos. Observe que apenas os parâmetros ZHANG são valores absolutos. Os outros resultados são dados na forma do erro relativo (em percentagem) aos parâmetros ZHANG	45

Lista de Símbolos e Abreviações

$\text{diag}(\dots)$	$\text{diag}(a_1, a_2, \dots, a_n)$ representa uma matriz $n \times n$, cuja diagonal principal é dada por a_1, a_2, \dots, a_n e todos os outros elementos são iguais a zero.
$(\mathbf{x} \leftrightarrow \mathbf{y})$	Indica correspondência entre os pontos \mathbf{x} e \mathbf{y} ou entre as retas \mathbf{x} e \mathbf{y} .
$\mathbb{R}^{n \times m}$	Representa o conjunto de todas as matrizes de dimensão $n \times m$.
\mathbb{R}^n	Representa o conjunto de todos os vetores $n \times 1$.
$[\mathbf{x}]_{\times}$	É a matriz antissimétrica (3×3) do vetor \mathbf{x} , ou seja, $[\mathbf{x}]_{\times} \mathbf{v} = \mathbf{x} \times \mathbf{v}$, onde $\mathbf{v} \in \mathbb{R}^3$.
θ_{sk}	Ângulo que corrige a ortogonalidade da matriz de sensores da câmera. $\theta_{sk} \approx \pi/2$. (Veja a equação (2.4).)
$\mathbf{0}_n$	Vetor nulo de dimensão $n \times 1$.
$\mathbf{\Pi}_{\infty}$	Vetor que representa a equação do plano no infinito.
$d(\mathbf{x}, \mathbf{y})$	Indica distância Euclidiana entre os pontos \mathbf{x} e \mathbf{y} ou entre o ponto \mathbf{x} e a reta \mathbf{y} .
f	Distância focal da câmera. (Veja a equação (2.4).)
$n_{x/y}$	Parâmetro intrínseco da câmera, inserido para levar em consideração que os sensores utilizados pela câmera não são um quadrado perfeito. $n_{x/y} \approx 1$. (Veja a equação (2.4).)
\mathbf{A}	Matriz 3×3 que contém os parâmetros intrínsecos da câmera 1. A matriz \mathbf{A} é dada pela equação (2.4). (Veja também o símbolo \mathbf{M}' .)
\mathbf{B}	Matriz 3×3 que representa a equação de ω . Esse matriz relaciona-se com os parâmetros intrínsecos da câmera por $\mathbf{B} = \mathbf{A}^{-T} \mathbf{A}^{-1}$. (Veja também o símbolo \mathbf{M}' .)
\mathbf{C}_x	Representa o centro da câmera x .
\mathbf{F}	Matriz fundamental (3×3) que relaciona duas imagens estéreo.
\mathbf{H}_{∞}	Matriz 3×3 que representa a homografia do infinito.
\mathbf{I}_n	Matriz identidade de dimensão $n \times n$.
I_x	O plano de imagem da “câmera x ”. A “câmera 1” refere-se à câmera que está mais a esquerda em um conjunto estéreo. A “câmera 2” é a que está à direita.
\mathbf{l}_{∞}	Vetor que representa a equação da reta no infinito, no plano da imagem.
\mathbf{l}_{xi}	Vetor que representa a i -ésima reta epipolar que está no plano I_x .

\mathbf{M}'	Refere-se a uma matriz relacionada a câmara 2, ou seja, \mathbf{M}' é para a câmara 2 o que \mathbf{M} é para a câmara 1.
$\widetilde{\mathbf{M}}_i$	Representa, em coordenadas homogêneas, um ponto 3D. Geralmente, $\widetilde{\mathbf{M}}_i = [X_i, Y_i, Z_i, T_i]^T$ e $\mathbf{M}_i = [X_i, Y_i, Z_i]^T$.
$\widetilde{\mathbf{m}}_{xi}$	Representa, em coordenadas homogêneas, a projeção de ponto 3D no plano de imagem I_x . Geralmente, $\widetilde{\mathbf{m}}_{xi} = [u_{xi}, v_{xi}, t_{xi}]^T$ e $\mathbf{m}_{xi} = [u_{xi}, v_{xi}]^T$.
\mathbf{m}_0	Ponto central da câmara. Geralmente, $\mathbf{m}_0 = [u_0, v_0]^T$. (Veja a equação (2.4).)
\mathcal{M}	Coordenadas de um ponto \mathbf{M} no espaço projetivo tridimensional.
$\widetilde{\mathbf{e}}_x$	Representa, em coordenadas homogêneas, o epipolo da imagem I_x . $\widetilde{\mathbf{e}}_1 \in I_1$ e, geralmente, é representado por $\widetilde{\mathbf{e}}_1 = [e_u, e_v, e_t]^T$, e; $\widetilde{\mathbf{e}}_2 \in I_2$ e, geralmente, é visto como $point_{he_2} = [e'_u, e'_v, e'_t]^T$.
\mathbf{P}	Matriz de projeção (3×4) da câmara 1. (Veja também o símbolo \mathbf{M}' .)
\mathcal{P}^2	Espaço projetivo bidimensional.
\mathcal{P}^3	Espaço projetivo tridimensional.
\mathcal{P}^1	Espaço projetivo unidimensional.
\mathcal{P}	Na calibração projetiva, representa a matriz de projeção (3×4) da câmara 1. (Veja também o símbolo \mathbf{M}' .)
SVD	Decomposição do autovalor. (Do inglês: <i>Singular Value Decomposition</i> .)
\mathbf{T}_{PE}	Matriz 4×4 que representa a homografia que atualiza uma matriz de projeção Euclidiana em uma projetiva.
$\text{trace}(\mathbf{M})$	Somatório de todos os elementos da diagonal principal da matriz quadrada \mathbf{M} .

INTRODUÇÃO

De todos os sentidos do ser humano, a visão é um dos mais complexos e interessantes, pois é possível fazer extrações de uma representação do mundo real com base em dados visuais. Em especial, é possível recuperar informações geométricas como, por exemplo, ângulos, distâncias, paralelismo. Contudo, para que isto seja possível, o sistema deve ser calibrado (Hartley e Zisserman, 2000).

Neste trabalho, discute-se dois métodos distintos de calibração de câmeras. Ambos foram implementados para uso em aplicações de robótica e visam aumentar a flexibilidade do processo de calibração.

O processo de extrair informações de uma dada imagem em visão computacional, é semelhante ao funcionamento do sistema de visual natural, contudo, enquanto no sistema natural usamos a retina, em visão computacional usa-se a câmera.

Uma câmera é modelada como um dispositivo que projeta pontos do espaço 3D em pontos em um plano. Esse processo não é reversível, ou seja, dado um ponto em uma imagem não é possível determinar as coordenadas do ponto 3D correspondente de forma unívoca, pois neste processo, toda a informação 3D é perdida.

Uma possível abordagem para recuperar a informação 3D é através da visão binocular, sendo necessário o uso de uma segunda câmera para encontrar unicamente as coordenadas do ponto no espaço 3D.

Com relação à visão monocular, o uso da visão binocular apresenta algumas vantagens dentre as quais pode-se citar o fato de ser possível perceber onde e como os objetos estão dispostos em uma cena, a partir da noção de profundidade. Isto permite fazer conclusões sobre a posição e orientação de objetos no espaço.

Um exemplo de informação que pode ser extraída de uma cena tridimensional é a obtenção de coordenadas de pontos no espaço, dada suas respectivas projeções nos planos de imagens de um

conjunto binocular. Contudo, isso só é possível se o processo pelo qual as câmeras projetam pontos no espaço em pontos nos planos for bem conhecido, ou seja, se os parâmetros intrínsecos e extrínsecos do conjunto de câmeras forem conhecidos, sendo necessário para isso um processo chamado calibração.

O processo de calibração pode ser realizado com auxílio de um gabarito ou de forma automática, durante a operação normal do sistema de visão. Dentro deste contexto, dois algoritmos de calibração foram implementados. O primeiro usa um objeto 1D (pontos sobre uma reta) para extrair os parâmetros da câmera. O segundo é aplicado à visão binocular e realiza a calibração de forma automática.

1.1 Organização do trabalho

Além desta introdução, há neste trabalho mais quatro capítulos, que tratam dos assuntos descritos a seguir.

No capítulo dois, é apresentado um estudo sobre visão binocular. Aborda-se todos os assuntos que são pré-requisitos para entender o problema da dissertação, tais como: modelo da câmera utilizado, geometria epipolar, matriz fundamental, e as diversas técnicas existentes para obter as correspondências de pontos e a calibração das câmeras.

No capítulo três, as contribuições do trabalho para a área são apresentadas.

No capítulo quatro, apresenta-se a metodologia e os resultados dos testes utilizados para avaliar os algoritmos implementados.

Por último, no capítulo cinco, uma discussão geral do trabalho é apresentada.

1.2 Notações

Em todo o texto, matrizes e vetores são representados por letras, números ou símbolos em negrito. Já constantes são representadas por letras, números ou símbolos em itálicos. Para acessar os elementos de uma matriz, uma notação específica é adotada, por exemplo, $\mathbf{A}(i, j)$ representa o elemento da matriz \mathbf{A} na i -ésima linha e j -ésima coluna. O caracter “:” é utilizado para destacar mais de um elemento da matriz. Assim, $\mathbf{A}(i : j, n : m)$ representa todos os elementos de \mathbf{A} que pertencem a uma das linhas entre i e j e uma das colunas entre n e m . Outra notação adotada é considerar $A^{-T} = (A^{-1})^T = (A^T)^{-1}$ para toda matriz quadrada e inversível.

As câmeras pertencentes ao conjunto binocular são chamadas de “câmera 1” e “câmera 2”. Normalmente, a câmera 1 é tomada como referência. Assim, os planos de imagem das câmeras são expressos por I_1 e I_2 para, respectivamente, a câmera 1 e a câmera 2. No caso de matrizes, um apóstrofo (ou a ausência dele) é utilizado para indicar a qual câmera esta se refere. Por exemplo, \mathbf{A} está relacionada a “câmera 1” e \mathbf{A}' a “câmera 2”.

As coordenadas de um ponto 3D no sistema de coordenadas do ambiente de uma câmera, considerando o modelo de câmera *pinhole* (Faugeras e Luong, 2001) são apresentados como $\mathbf{M} = [x, y, z]^T$ e a projeção correspondente no plano de imagem I , como $\mathbf{m} = [u, v]^T$. Além disso, as coordenadas homogêneas de um ponto $\mathbf{m} = [x, y, \dots]^T$ são representadas por $\tilde{\mathbf{m}}$, isto é, $\tilde{\mathbf{m}} = [x, y, \dots, 1]^T$. De uma forma mais geral, um ponto qualquer em coordenadas homogêneas é representado por $\tilde{\mathbf{m}} = [x, y, \dots, t]^T$. Um índice, se houver, indica a posição do ponto em um conjunto de pontos.

Finalmente, uma reta l_α , no plano de imagem I_α e que passa pelo ponto $\mathbf{m}_\alpha = [u_\alpha, v_\alpha]^T$, deve satisfazer $a_\alpha u_\alpha + b_\alpha v_\alpha + c = 0$. Essa mesma reta é representada no texto como sendo $\mathbf{l}_\alpha = [a_\alpha, b_\alpha, c_\alpha]^T$. Assim, temos uma prática notação para a equação da mesma, ou seja, $\mathbf{l}_\alpha^T \mathbf{m}_\alpha = 0$ ou $\mathbf{m}_\alpha^T \mathbf{l}_\alpha = 0$. Novamente, um segundo índice, se houver, indica a posição da reta em um conjunto de retas.

FUNDAMENTAÇÃO TEORICA

Neste capítulo são apresentados alguns conceitos da área de visão computacional relacionados ao tema do trabalho. Em especial, o modelo utilizado para a câmera, a geometria epipolar e métodos de calibração de câmeras são discutidos em detalhes.

2.1 Modelo da Câmera

Neste trabalho, foi utilizado o modelo *pinhole* (Faugeras e Luong, 2001). Em tal modelo, basicamente, um ponto M no espaço 3D é projetado em um ponto m , pertencente a um plano chamado “plano de imagem” (como ilustrado na figura 2.1). Além disso, o raio de projeção do ponto M passa, necessariamente, por um ponto C no espaço 3D, chamado “centro da câmera”. A distância do plano de imagem ao ponto C é conhecida como “distância focal” e representada por f . A figura 2.2 mostra a câmera inserida num sistema de coordenadas, conhecido como “sistema de coordenadas da câmera”, cuja origem é o ponto C . Além disso, é representado no plano de imagem um outro sistema de coordenadas, com origem em m_0 , chamado “sistema de coordenadas da imagem”. O ponto m_0 , chamado “centro da imagem”, é o ponto no qual o eixo Z corta o plano de imagem.

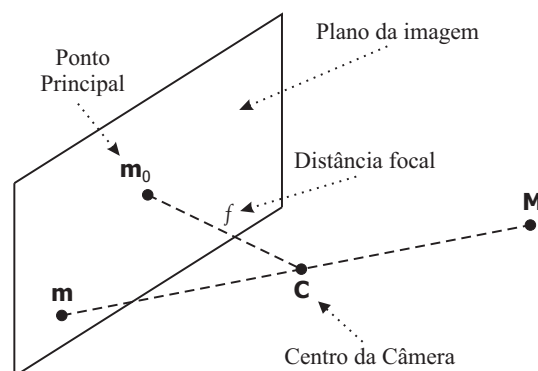


Figura 2.1 – Modelo de câmera *pinhole*.

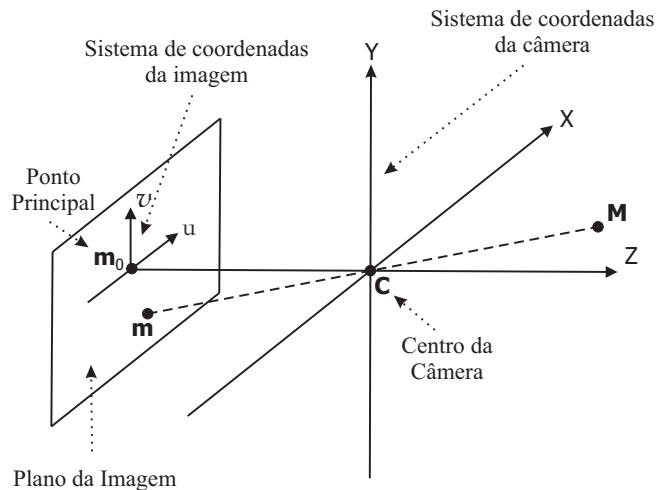


Figura 2.2 – Modelo *pinhole* e seus sistemas de coordenadas

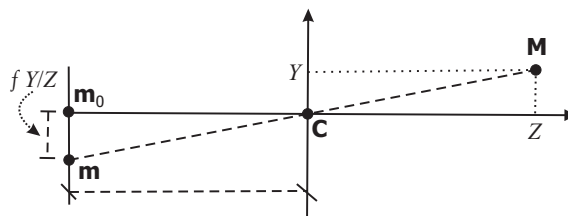


Figura 2.3 – Determinação da ordenada do ponto m em função das coordenadas do ponto M

A figura 2.3 mostra a relação entre o ponto M no espaço, no sistema de coordenadas da câmera, e sua projeção m , no sistema de coordenadas da imagem, obtida através de semelhança de triângulos, onde observa-se que a ordenada de m é $u = fY/Z$. Por analogia, a abscissa de m é $v = fX/Z$. Logo:

$$\mathbf{m} = [fX/Z, fY/Z]^T.$$

Pode-se utilizar uma equação homogênea para representar a equação anterior, uma vez que a mesma relaciona as coordenadas de \mathbf{m} e \mathbf{M} de forma não linear, ou seja,

$$\begin{bmatrix} fX \\ fY \\ Z \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \quad (2.1)$$

A equação (2.1) é reescrita frequentemente como

$$\tilde{\mathbf{m}} \simeq \mathbf{A}[\mathbf{I} \ \mathbf{0}]\tilde{\mathbf{M}} \quad (2.2)$$

onde

$$\mathbf{A} = \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2.3)$$

concentra os “parâmetros intrínsecos da câmera”. O símbolo \simeq é utilizado para explicitar que os dois lados de uma igualdade diferem-se por uma constante desconhecida e não-nula, ou seja, trata-se de uma equação homogênea, como por exemplo, a equação (2.2). Somente em um modelo muito simplificado de uma câmera é utilizada a matriz \mathbf{A} dada em (2.3). Na prática, o modelo mais geral utiliza \mathbf{A} dada por

$$\mathbf{A} = \begin{bmatrix} \alpha & \alpha \cot \theta_{sk} & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (2.4)$$

Os sensores utilizados pela câmera não são um quadrado perfeito, ou seja, a altura pode diferir da largura. Dessa forma, α é feito diferente de β para levar em consideração que os sensores utilizados pela câmera não são um quadrado perfeito. Já o parâmetro $\cot \theta_{sk}$ está relacionado ao ângulo θ_{sk} , que corrige a ortogonalidade da matriz destes sensores. O ponto $\mathbf{m}_0 = [u_0, v_0]^T$ (coordenadas do ponto central da câmera) é inserido para levar em consideração que, normalmente, a origem do sistema de coordenadas da imagem está no canto superior esquerdo, e não no ponto \mathbf{m}_0 .

A equação (2.4) aproxima-se cada vez mais da equação (2.3) à medida que a tecnologia evolui. Por exemplo, geralmente, $|\theta_{sk} - \pi/2|$ é da ordem de 10^{-6} ou seja, $sk \approx 0$ e α/β é muito próximo da unidade. Por isto, neste trabalho, considera-se $sk = 0$. Dessa forma, há apenas quatro parâmetros intrínsecos a serem estimados.

Outro fato a ser considerado é a estimação do ponto \mathbf{m}_0 . Esta é muito sensível a ruídos, levando muitos autores assumir que \mathbf{m}_0 localiza-se exatamente no centro da imagem (Zhang et al., 1996). Felizmente, esta aproximação tem pouca influência na recuperação das informações 3D. Uma outra observação é que as coordenadas do ponto \mathbf{M} podem ser fornecidas em relação a outro referencial diferente do sistema de coordenadas da câmera. Quando isto ocorre, o referencial em questão é dito ser o “sistema de coordenadas do ambiente”, e a equação (2.2) deve ser reescrita como

$$\tilde{\mathbf{m}} \simeq \mathbf{A}[\mathbf{R} \ \mathbf{t}]\tilde{\mathbf{M}} \quad (2.5)$$

onde $[\mathbf{R} \ \mathbf{t}]$, rotação e translação, são os “parâmetros extrínsecos da câmera”, sendo \mathbf{R} uma matriz 3x3 que representa a rotação da câmera em relação ao sistema de coordenadas do ambiente e \mathbf{t} um vetor que equivale à translação do centro da câmera com relação ao sistema de coordenadas do ambiente (Hartley e Zisserman, 2000, capítulo 5, página 142)

Freqüentemente, define-se a matriz de projeção \mathbf{P} de uma câmera como sendo

$$\mathbf{P} \simeq \mathbf{A}[\mathbf{R} \ \mathbf{t}]. \quad (2.6)$$

Logo, a equação (2.2) pode ser reescrita como

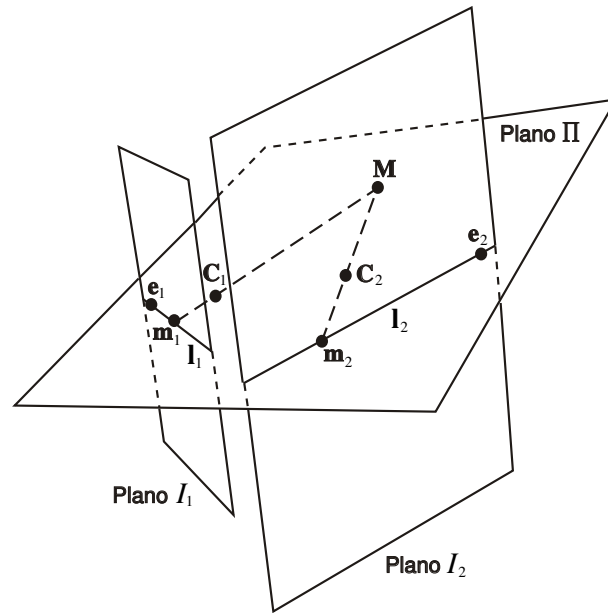


Figura 2.4 – Geometria Epipolar.

$$\tilde{m} \simeq \widetilde{PM}. \quad (2.7)$$

2.2 Visão Binocular

A equação (2.7) permite, dadas as coordenadas de um ponto M , encontrar a sua projeção, m , no plano de imagem. Contudo, tal processo não é reversível, ou seja, dado o ponto m , é impossível encontrar as coordenadas do ponto M unicamente. Isto ocorre porque qualquer ponto sobre a reta \overline{CM} (figura 2.1) tem a projeção no mesmo ponto m . Nesta seção, apresenta-se uma discussão sobre visão binocular que permite contornar este problema.

2.2.1 Geometria Epipolar

Duas imagens de uma única cena são relacionadas pela geometria epipolar. Esta geometria já foi descrita em numerosos artigos como, por exemplo, (Armangué e Salvi, 2003; Torr e Murray, 1997; Zhang, 1998). Contudo, basicamente esta geometria pode ser entendida observando-se a figura 2.4. Nesta, os pontos C_1 e C_2 são os centros ópticos das, respectivamente, 1ª e 2ª câmeras. Então, os pontos m_1 e m_2 são as projeções do ponto M nos planos I_1 e I_2 respectivamente. Além disso, o correspondente do ponto m_1 na 1ª imagem está restrito à uma reta, chamada “reta epipolar” de m_1 , denotada por l_2 . Este fato é normalmente referenciado como “restrição epipolar”.

A reta l_2 é a intersecção do plano Π (definido por M , C_1 e C_2) com o plano da 2ª imagem I_2 . Como já discutido, isto ocorre porque o ponto m_1 pode ser a projeção de qualquer ponto sobre a reta

$\overline{C_1M}$, cuja projeção em I_2 é a reta l_2 . Além disso, é possível notar que todas as retas epipolares dos pontos de I_1 passam num ponto comum e_2 , chamado “epipolo” de I_2 , que é a intersecção do plano de imagem I_2 com a reta $\overline{C_1C_2}$.

Pode-se também observar a simetria da geometria epipolar. O correspondente em I_1 de cada ponto m_2 , sobre a reta l_2 , precisa pertencer a reta epipolar l_1 , que é a intersecção do mesmo plano Π com o plano I_1 . Todas as retas epipolares formam um conjunto contendo o epipolo e_1 , que é a intersecção da reta $\overline{C_1C_2}$ com o plano I_1 .

A restrição epipolar pode ser imposta algebricamente através da seguinte equação

$$\tilde{m}_2^T \mathbf{F} \tilde{m}_1 = 0. \quad (2.8)$$

Na equação anterior, o termo $\mathbf{F} \tilde{m}_1$ é equivalente à reta l_2 . Assim, a equação (2.8) simplesmente garante que o ponto \tilde{m}_2 pertence a reta l_2 .

2.2.2 Matriz Fundamental

Como discutido na próxima seção, a restrição epipolar é útil quando deseja-se realizar correspondência de pontos entre duas imagens, ou seja, dado um ponto $m_1 \in I_1$, sabe-se que o seu correspondente $m_2 \in I_2$ está sobre a reta $l_2 = \mathbf{F} \tilde{m}_1$. Isto reduz o problema de duas para apenas uma única dimensão.

Da figura 2.4, observa-se que a reta l_2 pode ser completamente definida pelo epipolo e_2 e qualquer projeção em I_2 de um ponto sobre $\overline{C_1M}$. Em particular, um ponto sobre $\overline{C_1M}$ pode ser expresso por $\mathbf{P}^+ \tilde{m}_1$, cuja projeção em I_2 é dada por $\mathbf{P}' \mathbf{P}^+ \tilde{m}_1$, onde \mathbf{P}^+ é a pseudoinversa da matriz \mathbf{P} , ou seja, $\mathbf{P}^+ = (\mathbf{P}^T \mathbf{P})^{-1} \mathbf{P}^T$. Dessa forma,

$$\begin{aligned} l_2 &= \tilde{e}_2 \times (\mathbf{P}' \mathbf{P}^+ \tilde{m}_1) \\ &= [\tilde{e}_2]_{\times} (\mathbf{P}' \mathbf{P}^+) \tilde{m}_1, \end{aligned} \quad (2.9)$$

onde $[\tilde{e}_2]_{\times}$ é matriz antissimétrica do vetor \tilde{e}_2 .

A equação anterior relaciona m_1 e l_2 matematicamente. Normalmente, a matriz

$$\mathbf{F} = [\tilde{e}_2]_{\times} (\mathbf{P}' \mathbf{P}^+) \quad (2.10)$$

é chamada de “matriz fundamental”.

A equação (2.9) mostra que, para cada ponto $m_1 \in I_1$, existe uma reta epipolar, l_2 , na imagem I_2 , ou seja, há uma correspondência entre pontos na primeira imagem e retas epipolares na segunda imagem.

Partindo da equação (2.10), a matriz fundamental também pode ser expressa em termos dos parâmetros intrínsecos e extrínsecos do conjunto binocular. Este fato é resumido no teorema a seguir.

Teorema 2.2.1 *A matriz fundamental de um conjunto binocular, cujas matrizes de projeção são $\mathbf{P} = \mathbf{A}[\mathbf{I} \ 0]$ e $\mathbf{P}' = \mathbf{A}'[\mathbf{R} \ \mathbf{t}]$, é dada por*

$$\mathbf{F} = \mathbf{A}'^{-T}[\mathbf{t}]_{\times} \mathbf{R} \mathbf{A}^{-1}. \quad (2.11)$$

Prova: O epipolo \mathbf{e}_2 é a projeção do ponto \mathbf{C}_1 no plano de imagem I_2 , ou seja,

$$\begin{aligned} \tilde{\mathbf{e}}_2 &= \mathbf{P}'\tilde{\mathbf{C}}_1 = \mathbf{A}'[\mathbf{R} \ \mathbf{t}] \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix} = [\mathbf{A}'\mathbf{R} \ \mathbf{A}'\mathbf{t}] \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix} \\ &= \mathbf{A}'\mathbf{t}. \end{aligned} \quad (2.12)$$

Além disso, como já discutido, $\mathbf{P}^+ = (\mathbf{P}^T \mathbf{P})^{-1} \mathbf{P}^T$, ou seja,

$$\begin{aligned} \mathbf{P}^+ &= \left(\begin{bmatrix} \mathbf{A}^T \\ \mathbf{0}^T \end{bmatrix} [\mathbf{A} \ 0] \right)^{-1} \begin{bmatrix} \mathbf{A}^T \\ \mathbf{0}^T \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{A}^{-1} \mathbf{A}^{-T} & \mathbf{0} \\ \mathbf{0}^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{A}^T \\ \mathbf{0}^T \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{A}^{-1} \\ \mathbf{0}^T \end{bmatrix}. \end{aligned}$$

Dessa forma,

$$\begin{aligned} \mathbf{P}'\mathbf{P}^+ &= \mathbf{A}'[\mathbf{R} \ \mathbf{t}] \begin{bmatrix} \mathbf{A}^{-1} \\ \mathbf{0}^T \end{bmatrix} = [\mathbf{A}'\mathbf{R} \ \mathbf{A}'\mathbf{t}] \begin{bmatrix} \mathbf{A}^{-1} \\ \mathbf{0}^T \end{bmatrix} \\ &= \mathbf{A}'\mathbf{R}\mathbf{A}^{-1}. \end{aligned} \quad (2.13)$$

Agora, substituindo (2.12) e (2.13) em (2.10), tem-se¹

$$\begin{aligned} \mathbf{F} &= [\mathbf{A}'\mathbf{t}]_{\times} \mathbf{A}'\mathbf{R}\mathbf{A}^{-1} \\ &= \mathbf{A}'^{-T}[\mathbf{t}]_{\times} \mathbf{A}'^{-1} \mathbf{A}'\mathbf{R}\mathbf{A}^{-1} \\ &= \mathbf{A}'^{-T}[\mathbf{t}]_{\times} \mathbf{R}\mathbf{A}^{-1} \end{aligned}$$

que conclui a demonstração do teorema. □

Contudo, antes de ser utilizada, a matriz fundamental deve ser estimada. Se o conjunto binocular está calibrado, a equação (2.11) pode ser utilizada para este fim. Em sistemas não-calibrados, a matriz fundamental ainda pode ser estimada a partir de um conjunto de correspondência de pontos.

¹Nesta demonstração, utilizou-se o fato que, dada uma matriz não-singular \mathbf{M} e um vetor \mathbf{v} , $[\mathbf{M}\mathbf{v}]_{\times} = \mathbf{M}^{-T}[\mathbf{v}]_{\times} \mathbf{M}^{-1}$. Veja-se, por exemplo, (Hartley e Zisserman, 2000, página 555) para a prova.

Encontram-se na literatura diversos métodos para estimar a matriz fundamental. Veja-se, por exemplo, os trabalhos de (Armangué e Salvi, 2003; Torr e Murray, 1997; Zhang, 1998) para três análises críticas diferentes de tais métodos. Contudo, o método mais conhecido é o algoritmo de 8 pontos (Hartley, 1997a). Tal método, dado um conjunto com $n \geq 8$ correspondências, estima a matriz fundamental de forma linear, minimizando

$$\sum_{i=1}^n \|\tilde{\mathbf{m}}_{2i}^T \mathbf{F} \tilde{\mathbf{m}}_{1i}\|^2, \quad \text{sujeito a } \|\mathbf{F}\| = 1. \quad (2.14)$$

O algoritmo de 8 pontos já pressupõe a existência de um conjunto de correspondência de pontos. Normalmente, tais correspondências, por serem estabelecidas sem auxílio da geometria epipolar, não são confiáveis. Isto pode fazer com que o algoritmo de 8 pontos forneça resultados inexatos. A solução para este problema é estimar a matriz fundamental de forma robusta, ou seja, eliminando-se as falsas correspondências enquanto se realiza a estimação.

Um dos métodos robustos mais conhecidos na literatura é o RANSAC², introduzido na visão computacional por Fischler e R.Bolles (1981). Basicamente, o RANSAC pode ser resumido nos passos a seguir.

Dado um conjunto de N correspondências $c_i = (\mathbf{m}_{1i}, \mathbf{m}_{2i})$, escolhe-se, aleatoriamente, N_{sc} subconjuntos de N_{nc} correspondências. Para cada subconjunto (indexado por j), estima-se a matriz fundamental, \mathbf{F}_j , e calcula-se o resíduo, $r_{ji}^2(\mathbf{F}_j, c_i)$, de todas as N correspondências. Cada resíduo é comparado com um limiar r_{th}^2 . Se $r_{ji}^2 < r_{th}^2$, a correspondência c_i é considerada uma boa correspondência. Após N_{sc} iterações, retém-se a matriz fundamental, \mathbf{F}_{win} , que ajustou-se ao maior número de boas correspondências. Por último, estima-se \mathbf{F} com apenas as boas correspondências (identificadas por $r_{(win)i}^2 < r_{th}^2$).

O número de subconjuntos, N_{sc} , utilizado no RANSAC deve ser tal que, supondo, dentre todas as N correspondências, uma porcentagem ϵ de falsas correspondências, exista uma probabilidade p de que (ao menos) um subconjunto j seja composto apenas por boas correspondências. Neste caso, N_{sc} é igual a

$$N_{sc} = \log(1 - p) / \log(1 - (1 - \epsilon)^{N_{nc}}). \quad (2.15)$$

Neste trabalho, a tática utilizada é iniciar N_{sc} usando a equação anterior e atualizá-lo a cada iteração j , ou seja, uma vez que a porcentagem, ϵ_j , de falsas correspondências tenha sido determinada, N_{sc} é atualizado por (2.15).

Da equação anterior, vê-se que N_{sc} aumenta exponencialmente com N_{nc} e ϵ . Por exemplo, considerando uma probabilidade $p = 99\%$, se $\epsilon = 25\%$ e $N_{nc} = 7$, então $N_{sc} = 33$. Contudo, se $\epsilon = 40\%$ e $N_{nc} = 8$, temos $N_{sc} = 272$. Assim, desde que quanto maior N_{sc} , maior o custo computacional, o ideal é ter-se N_{nc} e ϵ tão pequenos quanto possível.

²Do inglês: “Random Sample Consensus”.

O valor de ϵ depende de como o conjunto total de correspondências foi estabelecido, ou seja, ele depende do grau de confiança do algoritmo de correspondência de pontos.

Como \mathbf{F} tem sete graus de liberdade, o valor mínimo para N_{nc} é sete. Contudo, a solução com apenas sete correspondências não é estável. Por isso, neste trabalho, é utilizado o método de 8 pontos com $N_{nc} = 8$.

No RANSAC, o parâmetro mais crítico a ser escolhido é o limiar r_{th}^2 , pois dele depende o critério que diz se uma correspondência é boa ou ruim. Se r_{th}^2 é muito pequeno, boas correspondências podem ser consideradas ruins. Por outro lado, um r_{th}^2 grande faz com que algumas falsas correspondências não sejam detectadas.

Normalmente, o resíduo r_{ji}^2 é dado por

$$r_{ji}^2 = d^2(\tilde{\mathbf{m}}_{2i}, \mathbf{F}_j \tilde{\mathbf{m}}_{1i}) + d^2(\tilde{\mathbf{m}}_{1i}, \mathbf{F}_j^T \tilde{\mathbf{m}}_{2i}), \quad (2.16)$$

onde $d^2(\cdot)$ é o quadrado da distância euclidiana.

Além disso, se for considerado um ruído com uma distribuição gaussiana de média zero e desvio padrão σ_d , o resíduo r_{th}^2 é definido como

$$r_{th}^2 = d_{th}^2 \sigma_d^2, \quad (2.17)$$

onde d_{th}^2 deve ser escolhido tal que exista uma probabilidade p_d de uma boa correspondência ser erroneamente considerada uma falsa correspondência. Contudo, muitas vezes, d_{th}^2 é escolhido empiricamente. Por exemplo, Hartley e Zisserman (2000) utilizaram $d_{th}^2 = 3,84$, Zhang (1998) utilizou $d_{th}^2 = 2,5$ e Torr e Murray (1997) consideraram $d_{th}^2 = 1,99$.

O método RANSAC pode obter uma boa estimação de \mathbf{F} , mesmo que mais de 50% das correspondências sejam falsas. A desvantagem evidente é que ele necessita de uma estimação do desvio padrão do ruído, σ_d .

Outra característica importante a ser observada é que, de forma geral, o algoritmo RANSAC procura a matriz \mathbf{F}_j que minimiza a função a seguir

$$\min_{\mathbf{F}_j} \sum_{i=1}^N \mathcal{J}(r_{ji}^2),$$

onde

$$\mathcal{J}(r_{ji}^2) = \begin{cases} 0, & \text{se } r_{ji}^2 \leq r_{th}^2 \\ 1, & \text{se } r_{ji}^2 > r_{th}^2 \end{cases}.$$

Na equação anterior, é evidente que, se r_{th}^2 tiver um valor muito elevado, todas as correspondências serão consideradas boas. Nesse caso, qualquer matriz \mathbf{F}_j teria a mesma pontuação, ou seja, $\sum_{i=1}^N \mathcal{J}(r_{ji}^2)$ seria sempre igual a N . Por isso, Torr e Zisserman (1998) sugeriram uma discreta

alteração na função $\mathcal{J}(r_{ji}^2)$, ou seja,

$$\mathcal{J}(r_{ji}^2) = \begin{cases} r_{ji}^2, & \text{se } r_{ji}^2 \leq r_{th}^2 \\ r_{th}^2, & \text{se } r_{ji}^2 > r_{th}^2 \end{cases}.$$

Agora, cada boa correspondência contribui com um valor diferente e proporcional ao seu grau de ajuste a \mathbf{F}_j . Assim, mesmo considerando um número igual de boas correspondências, a função $\mathcal{J}(r_{ji}^2)$ deve ter valores diferentes para matrizes fundamentais diferentes.

Torr e Zisserman (1998) demonstraram que essa pequena modificação produz uma sensível melhoria no desempenho do algoritmo RANSAC. Assim, desde que o custo computacional adicionado é desprezível, este algoritmo (conhecido como MSAC³) é utilizado no presente trabalho.

2.3 Correspondência de Pontos

Na seção anterior, discute-se o problema de estimação da matriz fundamental a partir de um conjunto de correspondência de pontos. Contudo, quando o sistema não está calibrado e nem a geometria epipolar é conhecida, o processo de correspondência de pontos é uma tarefa muito complexa. Tal complexidade é devida principalmente ao espaço de busca, ou seja, cada ponto em uma imagem pode corresponder a todos os pontos da outra imagem. Felizmente, para o caso da estimação da matriz fundamental, apenas um conjunto de correspondências esparsas é necessário. Isto facilita a solução do problema.

Nesta seção, discute-se algumas das técnicas encontradas na literatura para resolver o problema da correspondência de pontos. Além disso, a técnica utilizada neste trabalho é descrita em detalhes ao final da seção.

O problema de correspondência de pontos consiste em, dado um ponto no espaço 3D, determinar as suas projeções correspondentes em duas ou mais imagens. Para isto, não há uma solução geral, devido a alguns problemas como, por exemplo, a ambigüidade (inerente ao problema, requer a introdução de restrições físicas e geométricas); oclusões, i.e., pontos em uma imagem sem um correspondente na outra (figura 2.5); distorções radiométricas, que fazem a projeção 3D de um ponto ter tons de cinza distintos em imagens diferentes; texturas uniformes, que acentuam o problema da ambigüidade, e; distorções projetivas que tornam diferente a forma de um objeto quando capturadas de outro ponto de vista. Dessa forma, para que tal problema torne-se tratável, é necessário o uso de algumas restrições. Normalmente, tais restrições são classificadas em dois grupos: locais e globais. Restrições locais aplicam-se a um pequeno número de pixels próximos a um pixel de interesse. Por outro lado, restrições globais aplicam-se à imagem completa.

Métodos baseados em restrições locais são mais rápidos e simples de serem implementados. Contudo, são ineficientes para realizar correspondência de pontos em regiões que apresentam oclusões ou

³Do inglês: “*M-Estimator Sample Consensus*”.

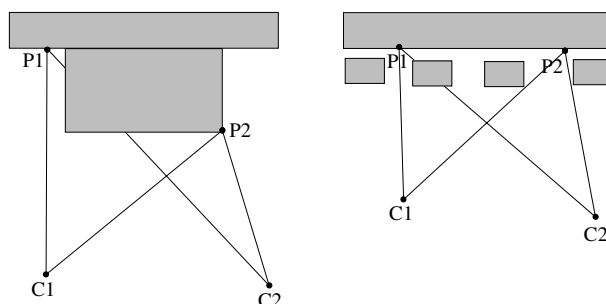


Figura 2.5 – Exemplo de oclusão. Os pontos P2 são visíveis por ambas as câmeras, sendo que os pontos P1 são visíveis somente para as câmeras C1.

textura uniforme (Brown et al., 2003). Por outro lado, métodos baseados em restrições globais analisam todas as regiões da imagem em conjunto. Por isso, são mais robustos. Contudo, normalmente apresentam um custo computacional maior.

Nas seções a seguir, apresentam-se alguns exemplos de métodos encontrados na literatura.

2.3.1 Métodos locais

A seguir, é dada uma visão geral dos métodos locais de correspondência de pontos.

Emparelhamento de Regiões

Métodos baseados no “emparelhamento de regiões⁴” procuram resolver o problema da correspondência de pontos comparando uma pequena região da imagem, em torno do ponto de interesse, com regiões de outra imagem. Por exemplo, dado um ponto \mathbf{m}_1 na imagem I_1 , define-se uma janela (ou região), W_1 , centrada em \mathbf{m}_1 . Em seguida, busca-se a janela W_2 centrada em \mathbf{m}_2 que, segundo algum critério pré-definido, mais se correlaciona com W_1 . Neste caso, \mathbf{m}_2 é dito o correspondente de \mathbf{m}_1 . A restrição epipolar é útil, pois reduz o espaço de busca de duas para apenas uma única dimensão.

Normalmente, algoritmos de Block Matching utilizam a restrição da semelhança para emparelhar regiões em imagens diferentes. Essa restrição obriga o correspondente de um ponto ser semelhante a ele. Para quantificar a semelhança entre pontos, podem ser utilizadas diversas medidas como, por exemplo, correlação cruzada normalizada⁵ (NCC) e soma dos quadrados das diferenças⁶ (SSD).

A correlação cruzada normalizada é o método estatístico padrão para determinar similaridade. A soma dos quadrados das diferenças (SSD) é computacionalmente tão simples quanto a correlação cruzada normalizada.

⁴Do inglês: “Block Matching”.

⁵Do inglês: “Normalized Cross Correlation”.

⁶Do inglês: “Sum of Squared Differences”.

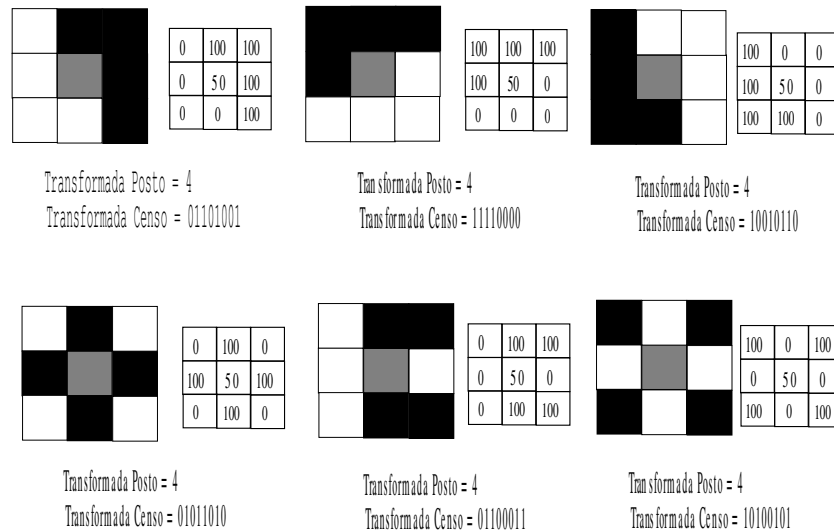


Figura 2.6 – Exemplo de transformadas posto e censo. Nota-se que todas as imagens tem a mesma transformada posto, e diferentes transformada censo.

Tanto a correlação por NCC quanto a por SSD falham na presença de distorções radiométricas⁷. Em vista disso, medidas mais robustas quanto a este tipo de distorção foram propostas por Zabih e Woodfill (1994), a saber, a “transformada posto⁸” e a “transformada censo⁹”. A transformada posto troca o valor da intensidade, $I(u, v)$, de cada um dos pixels de uma imagem, I , pelo número de pixels dentro de uma janela, W_n centrada em $\mathbf{m} = [u, v]^T$, cujas intensidades são maiores que $I(u, v)$. A transformada censo é uma variação da transformada posto. Tem a vantagem de conservar a distribuição espacial dos pixels. Isto é possível porque $I(u, v)$ é transformado em uma seqüência de bits, ou seja, se a intensidade de um pixel dentro de W_n é maior do que $I(u, v)$, o bit correspondente da seqüência é feito igual a 1, caso contrário, este torna-se 0. Na figura 2.6 pode-se entender melhor as transformadas acima mencionadas.

Recentemente, Banks et al. (1999) realizaram uma excelente análise crítica das transformadas posto e censo. Além disso, outras medidas de semelhança podem ser encontradas na literatura. Em especial, P.Aschwanden e W.Guggenbuhl (1993) realizaram uma comparação detalhada de tais medidas.

Emparelhamento de características

Métodos de emparelhamento de regiões são conhecidos por falharem em regiões com descontinuidade na disparidade ou com texturas uniformes. A classe de algoritmos de “emparelhamento de características¹⁰”, ao invés de tentar emparelhar pontos da imagem, realiza a correspondência entre características da imagem como, por exemplo, cantos, retas, retângulos, etc. Esses métodos buscam

⁷Distorções radiométricas fazem com que os tons de cinza de pontos correspondentes difiram por um ganho e um desvio.

⁸Do inglês: “rank”.

⁹Do inglês: “census”.

¹⁰Do inglês “Feature Matching”.

superar as limitações das regiões de apoio utilizadas no emparelhamento de regiões usando características mais invariantes. Maiores detalhes sobre os métodos podem ser encontrados em Dhond e Aggarwal (1989).

2.3.2 Métodos Globais

Esses métodos exploram restrições globais para aumentar as chances de sucesso em regiões problemáticas devido a oclusões, texturas uniformes, etc. Infelizmente, o uso de restrições globais torna o custo computacional maior em comparação aos métodos locais. Por isso, métodos globais normalmente assumem que pontos correspondentes possuem disparidade vertical nula. Isto limita a região de busca para apenas uma linha da imagem, chamada de “linha de busca¹¹”.

Geralmente, métodos globais são baseados em programação dinâmica¹², que usa restrições da imagem como ordem e suavidade para otimizar correspondência em cada linha de busca.

Devido ao custo computacional elevado e as considerações impostas para reduzi-lo, métodos de correspondência de pontos baseados em restrições globais são normalmente utilizados apenas para estabelecer um conjunto de correspondências densas para, por exemplo, construção de mapas de disparidades. Em vista disto, neste trabalho apenas métodos locais foram utilizados.

Para maiores detalhes sobre métodos globais, sugere-se os trabalhos de Scharstein e Szeliski (1996, 2002); Shah (1993).

2.3.3 Algoritmo utilizado

Neste trabalho, o algoritmo de correspondência de pontos utilizado é o proposto por de França et al. (2005). Tal algoritmo pressupõe a existência de um conjunto de cantos segmentados em cada imagem do conjunto binocular. Em seguida, um conjunto de pares de cantos candidatos à correspondência é formado baseado na correlação entre tais cantos. Logo após, a ambigüidade é eliminada atribuindo-se um grau de confiança a cada par candidato e eliminando-se correspondências pouco confiáveis. Por último, a matriz fundamental é estimada robustamente e a restrição epipolar utilizada para eliminar as falsas correspondências.

Para formação do conjunto de candidatos à correspondência, é utilizada a restrição da semelhança. Inicialmente, aplica-se a transformada censo às duas imagens. Em seguida, para cada canto segmentado, \mathbf{m}_{1i} , de I_1 , calcula-se a semelhança entre este e todos os cantos segmentados, \mathbf{m}_{2j} , de I_2 que estão dentro de uma janela de dimensão $2l_s \times 2l_s$, centrada em \mathbf{m}_{1i} (figura 2.7). Então, o par $c_{ij} = (\mathbf{m}_{1i}, \mathbf{m}_{2j})$ que apresentar a maior semelhança, é considerado um candidato à correspondência.

¹¹Do inglês “scanline”

¹²Método de Programação Dinâmica é um método matemático que reduz a complexidade computacional de problemas de otimização, subdividindo-os em problemas menores e mais simples.

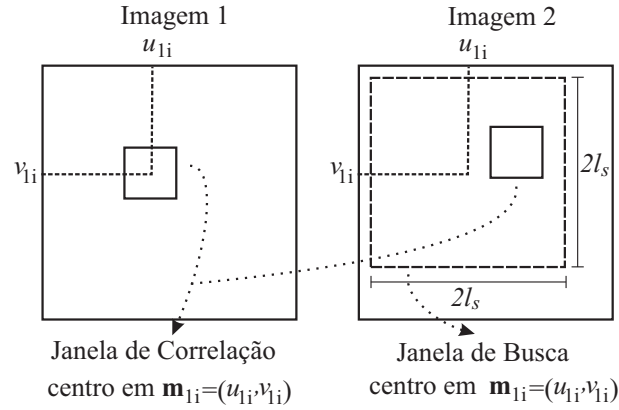


Figura 2.7 – Correspondência de pontos por correlação.

Devido a instabilidade dos algoritmos disponíveis para segmentar os cantos em ambas as imagens, na prática, diversos cantos de uma imagem não correspondem a nenhum canto na outra imagem. Isso dificulta o processo de seleção dos candidatos à correspondência e pode fazer com que exista ambigüidade entre o conjunto de candidatos à correspondência, ou seja, um canto de I_2 forme um par com mais de um ponto de I_1 . Assim, uma outra etapa para eliminar a ambigüidade é necessária.

Considerando o par $c_{ij} = (\mathbf{m}_{1i}, \mathbf{m}_{2j})$ um candidato à correspondência, representa-se o conjunto de vizinhos de \mathbf{m}_{1i} e de \mathbf{m}_{2j} dentro de uma janela de dimensão $2l_n \times 2l_n$ (figura 2.8) por, respectivamente, $N(\mathbf{m}_{1i})$ e $N(\mathbf{m}_{2j})$. Assim, se c_{ij} for um bom candidato à correspondência, espera-se que existam muitos candidatos à correspondência $V_{kl} = (\mathbf{n}_{1k}, \mathbf{n}_{2l})$, onde $\mathbf{n}_{1k} \in N(\mathbf{m}_{1i})$ e $\mathbf{n}_{2l} \in N(\mathbf{m}_{2j})$, tal que a “posição relativa” entre \mathbf{n}_{1k} e \mathbf{m}_{1i} é semelhante à posição relativa entre \mathbf{n}_{2l} e \mathbf{m}_{2j} . Por outro lado, se c_{ij} não é um bom candidato, não se espera encontrar a mesma relação entre esse pontos. Baseado nesta propriedade, definiu-se uma medida da confiança de um candidato à correspondência. Tal medida foi proposta tendo como principal critério a simplicidade.

Formalmente, de França et al. (2005) definiram a confiança, $Reab$, de um par $c_{ij} = (\mathbf{m}_{1i}, \mathbf{m}_{2j})$ candidato à correspondência, baseada nos candidatos à correspondência vizinhos, pela equação

$$Reab(c_{ij}) = \sum_{\mathbf{n}_{1k} \in N(\mathbf{m}_{1i})} \left(\sum_{\mathbf{n}_{2l} \in N(\mathbf{m}_{2j})} \Phi(c_{ij}, v_{kl}) \right), \quad (2.18)$$

onde $\Phi(c_{ij}, v_{kl})$ é igual a 1 se $(\mathbf{n}_{1k}, \mathbf{n}_{2l})$ é um candidato à correspondência e $r(c_{ij}, v_{kl}) < \varepsilon_r$; caso contrário, é igual a zero. Aqui,

$$r(c_{ij}, v_{kl}) = \frac{|d(\mathbf{m}_{1i}, \mathbf{n}_{1k}) - d(\mathbf{m}_{2j}, \mathbf{n}_{2l})|}{[d(\mathbf{m}_{1i}, \mathbf{n}_{1k}) + d(\mathbf{m}_{2j}, \mathbf{n}_{2l})]/2} \quad (2.19)$$

é uma medida do erro das posições relativas entre os pares $(\mathbf{m}_{1i}, \mathbf{n}_{1k})$ e $(\mathbf{m}_{2j}, \mathbf{n}_{2l})$, e ε_r é um limiar para esta medida.

Agora, alguns comentários devem ser feitos:

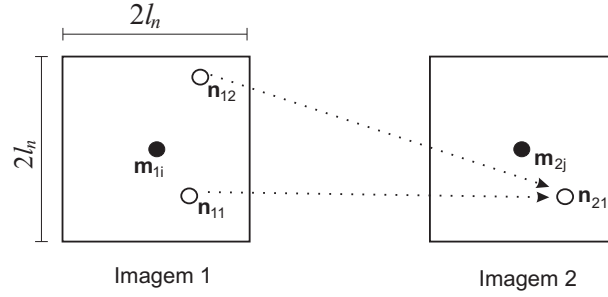


Figura 2.8 – Um par (m_{1i}, m_{2j}) candidato à correspondência com vizinhos dentro de uma janela de lado $2l_n$. O par (n_{11}, n_{21}) aumenta a confiança do par (m_{1i}, m_{2j}) .

1. A equação (2.18) conta o número de candidatos à correspondências que são vizinhos do candidato $(\mathbf{m}_{1i}, \mathbf{m}_{2j})$ e possuem posições relativas semelhantes.
2. Para que um par v_{kl} aumente a confiança de c_{ij} , pode-se impor que o ângulo, θ , entre $\overrightarrow{\mathbf{m}_{1i}\mathbf{n}_{1k}}$ e $\overrightarrow{\mathbf{m}_{2j}\mathbf{n}_{2l}}$ seja menor que um limiar θ_{th} . Isso aumenta o custo computacional, mas reduz o número de falsas correspondências.
3. Se mais de um ponto $\mathbf{n}_{1k} \in N(\mathbf{m}_{1i})$ forma um par candidato com um mesmo ponto $\mathbf{n}_{2l} \in N(\mathbf{m}_{2j})$ (figura 2.8), o somatório da equação (2.18) deve contar todos os pares ambíguos como se fossem apenas um único par.
4. Na equação (2.19), pode-se fazer $d(\mathbf{m}_x, \mathbf{n}_y)$ igual à distância euclidiana. Contudo, neste trabalho utilizou-se

$$d(\mathbf{m}_x, \mathbf{n}_y) = |u_m - u_n| + |v_m - v_n|, \quad (2.20)$$

onde considera-se $\mathbf{m}_x = [u_m, v_m]^T$ e $\mathbf{n}_y = [u_n, v_n]^T$. A equação (2.20) tem um custo computacional menor que o da distância euclidiana e o seu uso na equação (2.19) não prejudica o cálculo da confiança de $(\mathbf{m}_{1i}, \mathbf{m}_{2j})$.

Para eliminar a ambigüidade, deve-se aplicar a equação (2.18) em todos os candidatos à correspondência e:

1. Candidatos com $Reab(c_{ij}) \leq rb_{th}$ são descartados.
2. Caso haja ambigüidade, prevalece o candidato mais confiável.
3. Se dois pares ambíguos possuem a mesma confiança, prevalece aquele que tem a maior semelhança.

Seguindo as orientações acima, a ambigüidade é eliminada em apenas um único passo. Contudo, espera-se que algumas falsas correspondências persistam. Por isso, a matriz fundamental deve ser estimada e a restrição epipolar utilizada para eliminar as falsas correspondências.

A etapa de eliminação das falsas correspondências é a única fase iterativa do algoritmo proposto. Contudo, o uso do algoritmo MSAC aumenta significativamente a robustez do método.

2.4 Calibração da Câmera

Como discutido na seção 2.2,7, dado duas projeções correspondentes é possível estimar o ponto 3D equivalente. Isto é possível apenas se o raio óptico de cada projeção possa ser determinado. Por sua vez, tal determinação depende das matrizes de projeção do conjunto binocular. Sendo assim, essas matrizes devem ser estimadas através de um processo conhecido como calibração. Nesta seção, alguns dos métodos mais importantes de calibração de câmeras são apresentados.

2.4.1 Método Clássico

No método clássico, a calibração é realizada em duas etapas, ou seja,

- ✓ estimar a matriz de projeção \mathbf{P} .
- ✓ decompor a matriz de projeção em matrizes de parâmetros intrínsecos \mathbf{A} e extrínsecos \mathbf{R} e \mathbf{t} .

Nesta abordagem, as câmeras são calibradas uma de cada vez, e as restrições da geometria epipolar não são utilizadas. O processo ocorre como segue. Dados n pontos, $\mathbf{M}_i = [x_i, y_i, z_i]^T$, cujas posições são conhecidas e suas projeções no plano de imagem correspondentes, $\mathbf{m}_i = [u_i, v_i]^T$, a matriz de projeção \mathbf{P} pode ser reescrita como (Benevenuti, 2001)

$$\begin{bmatrix} \lambda u_i \\ \lambda v_i \\ \lambda \end{bmatrix} = \mathbf{P} \begin{bmatrix} x_i \\ y_i \\ z_i \\ 1 \end{bmatrix} \quad \text{ou} \quad \begin{bmatrix} \lambda u_i \\ \lambda v_i \\ \lambda \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ z_i \\ 1 \end{bmatrix}. \quad (2.21)$$

Como λ 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. Assim, restarão apenas 11 parâmetros a serem determinados. Dessa forma, a equação (2.21) pode ser expandida em

$$\begin{cases} p_{11}x_i + p_{12}y_i + p_{13}z_i + p_{14} = \lambda u_i \\ p_{21}x_i + p_{22}y_i + p_{23}z_i + p_{24} = \lambda v_i \\ p_{31}x_i + p_{32}y_i + p_{33}z_i + 1 = \lambda \end{cases} \quad (2.22)$$

Substituindo a equação acima em 2.21 e eliminando λ , tem-se:

$$\begin{cases} p_{11}x_i + p_{12}y_i + p_{13}z_i + p_{14} - p_{31}x_i u_i - p_{32}y_i u_i - p_{33}z_i u_i = u_i \\ p_{21}x_i + p_{22}y_i + p_{23}z_i + p_{24} - p_{31}x_i u_i - p_{32}y_i u_i - p_{33}z_i u_i = v_i. \end{cases} \quad (2.23)$$

Agora, considerando os n pares de pontos $\mathbf{M} \leftrightarrow \mathbf{m}$, o problema pode ser organizado da seguinte forma

$$\begin{bmatrix} x_1 & y_1 & z_1 & 1 & 0 & 0 & 0 & 0 & -x_1u_1 & -y_1u_1 & -z_1u_1 \\ 0 & 0 & 0 & 0 & x_1 & y_1 & z_1 & 1 & -x_1u_1 & -y_1u_1 & -z_1u_1 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ x_1 & y_1 & z_1 & 1 & 0 & 0 & 0 & 0 & -x_1u_1 & -y_1u_1 & -z_1u_1 \\ 0 & 0 & 0 & 0 & x_1 & y_1 & z_1 & 1 & -x_1u_1 & -y_1u_1 & -z_1u_1 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ x_1 & y_1 & z_1 & 1 & 0 & 0 & 0 & 0 & -x_1u_1 & -y_1u_1 & -z_1u_1 \\ 0 & 0 & 0 & 0 & x_1 & y_1 & z_1 & 1 & -x_1u_1 & -y_1u_1 & -z_1u_1 \end{bmatrix} \begin{bmatrix} p_{11} \\ p_{12} \\ p_{13} \\ p_{14} \\ p_{21} \\ p_{22} \\ \vdots \\ p_{33} \end{bmatrix} = \begin{bmatrix} u_1 \\ v_1 \\ \vdots \\ u_i \\ v_i \\ \vdots \\ u_n \\ v_n \end{bmatrix}. \quad (2.24)$$

A equação (2.24) é um sistema de equações lineares na forma

$$\mathbf{U}\mathbf{x} = \mathbf{b}, \quad (2.25)$$

onde \mathbf{U} é uma matriz $2n \times 11$, \mathbf{x} é a matriz de projeção representada como um vetor e \mathbf{b} é um vetor $2n$. A solução de \mathbf{x} e, logo, \mathbf{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 pares $\mathbf{M} \leftrightarrow \mathbf{m}$ para calcular \mathbf{P} .

2.4.2 Método de Tsai

O método de Tsai (1987) calcula a informação de calibração utilizando o modelo de câmera *pinhole* e projeção em perspectiva, porém adicionando um parâmetro k que é o coeficiente de distorção radial da lente. Além disso, o método de Tsai utiliza ainda um sistema de coordenadas auxiliar para tratamento da distorção radial e desvio do centro de projeção das lentes. 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 ponto, também deve ser conhecida a posição de sua projeção no sistema de coordenadas da imagem. Inicialmente, o método de Tsai calcula os parâmetros intrínsecos da câmera sem tratar a distorção radial, utilizando uma técnica de otimização por mínimos quadrados. A solução exata para os parâmetros é então calculada em uma nova etapa de otimização. Finalmente o coeficiente de distorção é calculado, também utilizando otimização por mínimos quadrados. Uma descrição mais detalhada pode ser encontrada em Tsai (1987).

2.4.3 Método Não-Linear

Os métodos apresentados anteriormente não são ideais, pois minimizam um critério sem significado físico. Para contornar esse problema, deve-se levar em consideração que as coordenadas dos

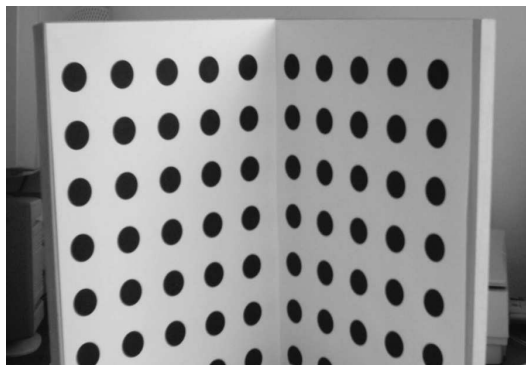


Figura 2.9 – Gabarito utilizado na calibração de câmeras.

pontos 3D normalmente são obtidas com a ajuda de um gabarito de calibração, como o da figura 2.9, onde dois ou três planos ortogonais, cujas formas e dimensões são bem conhecidos, são utilizados para formar um sistema de coordenadas “imaginário”. Logo, a exatidão das coordenadas dos pontos M_i dependem apenas do processo de manufatura do gabarito de calibração e geralmente, são conhecidas com extrema precisão. Assim, pode-se supor que a fonte de erros está nos pontos m_i e que uma alternativa para estimar P é resolver

$$\min \sum_{i=1}^n d(\tilde{m}_i, P\tilde{M}_i)^2. \quad (2.26)$$

A equação (2.26) consiste em minimizar a distância Euclidiana do ponto medido, m_i ao ponto “real” $P\tilde{M}_i$, com relação à todo o conjunto de n pontos. Entretanto, esse método exige o uso de um algoritmo de otimização não-linear que, normalmente, exige uma estimativa inicial de P , que por sua vez, pode ser obtida com o método não-linear.

O modelo da câmera utilizado neste trabalho pode ser considerado linear. Entretanto, no passado, a maioria das câmeras apresentavam distorção radial, característica que necessitava ser modelada através de um modelo não-linear. Por isso, é comum encontrar na literatura trabalhos, por exemplo Shih et al. (1996), Salvi et al. (2002), que propõem soluções para estimar também a distorção radial. Todos os métodos clássicos têm em comum a necessidade de um gabarito de calibração. Em vista disso, foram desenvolvidas novas técnicas que não necessitam de gabaritos, e que fornecem maior flexibilidade ao sistema.

2.4.4 Calibração Monocular

O gabarito 1D utilizado na calibração deve ser algo como um bastão (ou mesmo um cordão) com diversas esferas (indexadas por j) ao longo de sua extensão. A distância entre esferas consecutivas deve ser conhecida com exatidão. Então, no processo de calibração, o gabarito deve deslocar-se no campo visual da câmera. Enquanto isso, uma seqüência de imagens (indexada por i) deve ser capturada. Zhang (2004) demonstrou que a calibração monocular com este tipo de gabarito só é

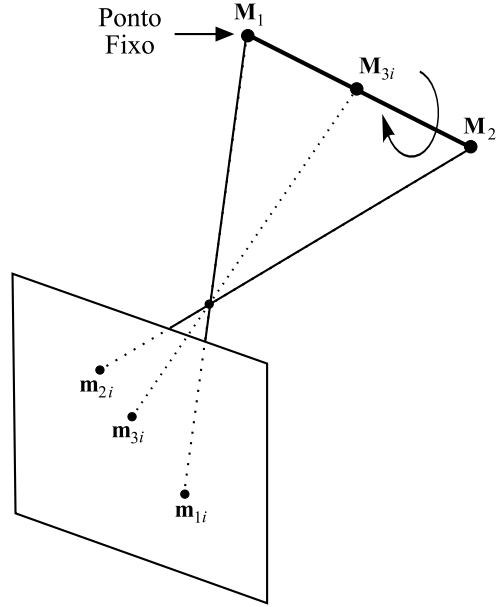


Figura 2.10 – Esboço do gabarito 1D utilizado na calibração.

possível se duas condições forem satisfeitas: o gabarito conter três ou mais pontos e um desses pontos, M_1 , deve ser fixo, como mostra o esboço da figura 2.10. Dessa forma, a projeção deste ponto, m_1 , é a mesma em todas as imagens da seqüência.

No esquema da figura 2.10, o comprimento, L , do gabarito pode ser expresso em função das coordenadas dos pontos extremos M_1 e M_{2i} , ou seja,

$$\|M_{2i} - M_1\| = L. \quad (2.27)$$

Além disso, desde que as posições relativas das esferas são conhecidas, um ponto do gabarito localizado entre M_1 e M_{2i} pode ser expresso por

$$M_{ji} = \lambda_{1j}M_1 + \lambda_{2j}M_{2i}, \quad (2.28)$$

onde λ_{1j} e λ_{2j} são conhecidos.

Da equação (2.2), considerando a profundidade (desconhecida) do ponto M_{ji} igual a z_{ji} , tem-se

$$M_1 = z_1 \mathbf{A}^{-1} \tilde{\mathbf{m}}_1, \quad (2.29)$$

$$M_{2i} = z_{2i} \mathbf{A}^{-1} \tilde{\mathbf{m}}_{2i}, \quad (2.30)$$

$$M_{ji} = z_{ji} \mathbf{A}^{-1} \tilde{\mathbf{m}}_{ji}. \quad (2.31)$$

Dessa forma, substituindo as equações anteriores em (2.28), obtém-se

$$z_{ji} \tilde{\mathbf{m}}_{ji} = z_1 \lambda_{1j} \tilde{\mathbf{m}}_1 + z_{2i} \lambda_{2j} \tilde{\mathbf{m}}_{2i} \quad (2.32)$$

após eliminar-se \mathbf{A}^{-1} de ambos os lados.

Agora, aplicando-se o produto cruzado em ambos os lados da equação anterior com $\tilde{\mathbf{m}}_{ji}$, tem-se

$$z_1 \lambda_{1j} (\tilde{\mathbf{m}}_{1i} \times \tilde{\mathbf{m}}_{ji}) + z_{2i} \lambda_{2j} (\tilde{\mathbf{m}}_{2i} \times \tilde{\mathbf{m}}_{ji}) = \mathbf{0} \quad (2.33)$$

que pode ser escrita como

$$z_{2i} = -z_1 \frac{\lambda_{1j} (\tilde{\mathbf{m}}_{1i} \times \tilde{\mathbf{m}}_{ji}) \cdot (\tilde{\mathbf{m}}_{2i} \times \tilde{\mathbf{m}}_{ji})}{\lambda_{2j} (\tilde{\mathbf{m}}_{2i} \times \tilde{\mathbf{m}}_{ji}) \cdot (\tilde{\mathbf{m}}_{2i} \times \tilde{\mathbf{m}}_{ji})}. \quad (2.34)$$

Substituindo (2.29) e (2.30) em (2.27) e considerando z_{2i} dado por (2.34), obtém-se

$$z_1 \|\mathbf{A}^{-1} \mathbf{h}_{ji}\| = L, \quad (2.35)$$

onde

$$\mathbf{h}_{ji} = \tilde{\mathbf{m}}_1 + \frac{\lambda_{1j} (\tilde{\mathbf{m}}_1 \times \tilde{\mathbf{m}}_{ji}) \cdot (\tilde{\mathbf{m}}_{2i} \times \tilde{\mathbf{m}}_{ji})}{\lambda_{2j} (\tilde{\mathbf{m}}_{2i} \times \tilde{\mathbf{m}}_{ji}) \cdot (\tilde{\mathbf{m}}_{2i} \times \tilde{\mathbf{m}}_{ji})} \tilde{\mathbf{m}}_{2i}. \quad (2.36)$$

A equação (2.35) é equivalente a

$$\mathbf{h}_{ji}^T \mathbf{B} \mathbf{h}_{ji} = L^2, \quad (2.37)$$

onde, desde que $s_k = 0$,

$$\mathbf{B} = z_1^2 \mathbf{A}^{-T} \mathbf{A}^{-1} = \begin{bmatrix} B_{11} & 0 & B_{13} \\ 0 & B_{22} & B_{23} \\ B_{13} & B_{23} & B_{33} \end{bmatrix}. \quad (2.38)$$

Levando em consideração que

$$\mathbf{h}_{ji}^T \mathbf{B} \mathbf{h}_{ji} = [a_{ji}^2, b_{ji}^2, 2a_{ji}c_{ji}, 2b_{ji}c_{ji}, c_{ji}^2]^T \mathbf{b} = \mathbf{u}_{ji} \mathbf{b}, \quad (2.39)$$

com $\mathbf{h}_{ji} = [a_{ji}, b_{ji}, c_{ji}]^T$ e $\mathbf{b} = [B_{11}, B_{22}, B_{13}, B_{23}, B_{33}]^T$, a equação (2.37) pode ser reescrita como

$$\mathbf{u}_{ji}^T \mathbf{b} = L^2. \quad (2.40)$$

Com n imagens, tem-se $\mathbf{U}_n = [\mathbf{u}_{j1}, \mathbf{u}_{j2}, \dots, \mathbf{u}_{jn}]^T$. Dessa forma, considerando $\mathbf{L}^2 = [L^2, \dots, L^2]^T$, pode-se encontrar \mathbf{b} resolvendo-se

$$\mathbf{U}_n \mathbf{b} = \mathbf{L}^2, \quad (2.41)$$

ou seja,

$$\mathbf{b} = (\mathbf{U}_n^T \mathbf{U}_n)^{-1} \mathbf{U}_n^T \mathbf{L}^2. \quad (2.42)$$

Dado um gabarito com mais de três pontos, tem-se um número maior de equações. Contudo, para cada imagem, apenas uma equação é linearmente independente. Assim, desde que existem 5 incógnitas (os 4 parâmetros de \mathbf{A} e z_1), são necessários, no mínimo, 5 deslocamentos do gabarito para solucionar o problema.

Uma vez que \mathbf{B} for conhecida, a matriz $z_1 \mathbf{A}^{-1}$ pode ser obtida de \mathbf{B} através da decomposição de Cholesky Golub e Van Loan (1996). Por sua vez, com z_1 e \mathbf{A} conhecidos, o ponto \mathbf{M}_1 pode ser obtido a partir de (2.29) e os pontos \mathbf{M}_{2i} a partir de (2.34) e (2.30). Por último, pode-se usar (2.28) para obter-se \mathbf{M}_{ji} .

O método descrito nesta seção minimiza um critério sem significado geométrico. Este é um dos motivos que contribuem para que o mesmo não forneça bons resultados na presença de ruído. Assim, Zhang (2004) definiu um critério de otimização que envolve as projeções dos pontos do bastão, \mathbf{m}_1 , \mathbf{m}_{2i} e \mathbf{m}_{ji} , ou seja, minimizando

$$\sum_{i=1}^n \left(\|\mathbf{m}_{1i} - \hat{\mathbf{m}}_1(\mathbf{A}, \mathbf{M}_1)\|^2 + \sum_{j=2}^q \|\mathbf{m}_{ji} - \hat{\mathbf{m}}_{ji}(\mathbf{A}, \mathbf{M}_{ji})\|^2 \right), \quad (2.43)$$

onde n é o número de imagens capturadas do gabarito, q é o número de pontos do mesmo e $\hat{\mathbf{m}}_{ji}(\mathbf{A}, \mathbf{M}_{ji})$ é a estimativa da projeção de \mathbf{M}_{ji} de acordo com (2.2).

Na equação anterior, as incógnitas a serem estimadas são: os 4 parâmetros intrínsecos de \mathbf{A} , o ponto \mathbf{M}_1 e $2n$ parâmetros, θ_i e ϕ_i , que definem as coordenadas esféricas dos pontos \mathbf{M}_{ij} de acordo com

$$\mathbf{M}_{2i} = \mathbf{M}_1 + L[\text{sen } \theta_i \cos \phi_i, \text{sen } \theta_i \text{sen } \phi_i, \cos \theta_i]^T, \quad (2.44)$$

onde θ_i e ϕ_i são ângulos que definem a orientação do gabarito.

Como discutido, a minimização da equação (2.43) envolve a solução de um problema de minimização não-linear. Normalmente, faz-se necessário uma estimativa inicial de todas as incógnitas envolvidas no problema. Evidentemente, para isto, o método linear apresentado pode ser utilizado.

2.5 Calibração Automática de um Conjunto Binocular

Nos métodos de calibração descritos na seção anterior é necessário a utilização de um gabarito. Em algumas aplicações, a utilização de tal gabarito é inconveniente, pois a calibração não pode ser realizada durante a operação normal do sistema. Por exemplo, caso haja uma mudança na distância focal das câmeras (um *zoom*), a inspeção visual deve ser interrompida e o gabarito posto a frente das câmeras. Em vista disso, Maybank e Faugeras (1992) propuseram uma nova teoria que permite que a calibração seja realizada de forma automática, durante a operação normal do sistema. Infelizmente, o processo de calibração automática possui um custo computacional bastante elevado. Por isso, tal processo ainda é pouco utilizado na prática.

Outro fato importante a ser observado é que, na calibração clássica, as câmeras que compõem o conjunto binocular são calibradas individualmente, ou seja, as matrizes \mathbf{P} e \mathbf{P}' são estimadas independentemente. Em seguida, através de manipulações algébricas (Hartley e Zisserman, 2000), as matrizes \mathbf{A} , \mathbf{A}' , \mathbf{R} e o vetor \mathbf{t} são calculados. Esta abordagem não é boa, pois é sabido que a matriz

fundamental derivada a partir de matrizes de projeções estimadas dessa forma satisfazem a restrição fundamental fracamente (Zhang et al., 1997).

Felizmente, a visão binocular acrescenta um conjunto de restrições que reduz a complexidade do processo de calibração automática. De fato, a calibração pode ser realizada apenas com equações lineares. Isto torna a técnica mais atrativa para ser utilizada na prática.

Nas seções a seguir, a teoria de calibração automática de um conjunto binocular é descrita em mais detalhes.

2.5.1 Calibração Projetiva

Do trabalho de Hartley et al. (1992), sabe-se que, se apenas a matriz fundamental do sistema for conhecida, é possível obtermos matrizes de projeção, \mathcal{P} e \mathcal{P}' , que permitem realizar uma calibração projetiva¹³, isto é, dada uma correspondência ($\mathbf{m}_1 \leftrightarrow \mathbf{m}_2$), o ponto 3D obtido com as matrizes \mathcal{P} e \mathcal{P}' está no espaço \mathcal{P}^3 e não no Euclidiano.

Na calibração projetiva, as matrizes de projeção das câmeras são as seguintes

$$\mathcal{P} \simeq [\mathbf{I} \quad \mathbf{0}] \quad \text{e} \quad \mathcal{P}' \simeq [\mathbf{H} \quad w\mathbf{e}_2], \quad (2.45)$$

onde (Luong e Viéville, 1996)

$$\mathbf{H} = \mathbf{H}_\infty + \mathbf{e}_2 \mathbf{w}^T, \quad (2.46)$$

com w igual a uma constante qualquer não-nula e \mathbf{w} é um vetor 3D qualquer, ou seja, \mathbf{H} tem infinitas soluções. Contudo, uma solução possível pode ser obtida com o teorema a seguir.

Teorema 2.5.1 *Conhecendo a matriz fundamental, \mathbf{F} , de um conjunto binocular não-calibrado, uma possível homografia, \mathbf{H} , que pode ser utilizada na calibração projetiva é*

$$\mathbf{H} = \left(-\frac{[\tilde{\mathbf{e}}_2]_\times}{\|\tilde{\mathbf{e}}_2\|^2} \right) \mathbf{F}, \quad (2.47)$$

onde $\tilde{\mathbf{e}}_2$ é o epipolo da segunda imagem.

Prova: Para obter a matriz \mathbf{H} a partir da matriz \mathbf{F} , podemos utilizar a seguinte propriedade válida para todo vetor \mathbf{v} (de França et al., 2004)

$$\|\mathbf{v}\|^2 \mathbf{I} = \mathbf{v}\mathbf{v}^T - [\mathbf{v}]_\times^2.$$

Assim, substituindo \mathbf{v} por $\tilde{\mathbf{e}}_2$ na equação anterior e multiplicando-a em ambos os lados por \mathbf{F} , tem-

¹³Do inglês: “weak calibration”

se

$$\begin{aligned}\mathbf{F} &= \frac{1}{\|\tilde{\mathbf{e}}_2\|^2}(\tilde{\mathbf{e}}_2\tilde{\mathbf{e}}_2^T - [\tilde{\mathbf{e}}_2]_{\times})\mathbf{F} \\ &= \frac{1}{\|\tilde{\mathbf{e}}_2\|^2}\tilde{\mathbf{e}}_2\tilde{\mathbf{e}}_2^T\mathbf{F} + [\tilde{\mathbf{e}}_2]_{\times}\left(-\frac{[\tilde{\mathbf{e}}_2]_{\times}}{\|\tilde{\mathbf{e}}_2\|^2}\mathbf{F}\right).\end{aligned}$$

Contudo, desde que $\mathbf{F}^T\tilde{\mathbf{e}}_2 = 0$, a equação anterior resume-se a

$$\mathbf{F} = [\tilde{\mathbf{e}}_2]_{\times}\left(-\frac{[\tilde{\mathbf{e}}_2]_{\times}}{\|\tilde{\mathbf{e}}_2\|^2}\mathbf{F}\right).$$

Finalmente, comparando a equação anterior com (2.10), tem-se que $\mathbf{P}'\mathbf{P}^+ = \left(-\frac{[\tilde{\mathbf{e}}_2]_{\times}}{\|\tilde{\mathbf{e}}_2\|^2}\mathbf{F}\right)$. Além disso, de (2.13) e considerando $\mathbf{H}_{\infty} = \mathbf{A}'\mathbf{R}\mathbf{A}^{-1}$ (Hartley e Zisserman, 2000), vê-se que, fazendo \mathbf{w}^T igual ao vetor nulo, na equação (2.46), $\mathbf{H} = \left(-\frac{[\tilde{\mathbf{e}}_2]_{\times}}{\|\tilde{\mathbf{e}}_2\|^2}\mathbf{F}\right)$ \square

Apesar de muitos conceitos importantes do espaço Euclidiano não serem válidos em \mathcal{P}^3 , diversos autores, como, por exemplo, (Beardsley et al., 1995; Hespanha et al., 1999; Ruf e Horaud, 2000; Zeller e Faugeras, 1994), desenvolveram inúmeras aplicações baseadas na calibração projetiva.

Hartley et al. (1992) e Faugeras (1992) demonstraram que as matrizes \mathcal{P} e \mathcal{P}' estão relacionadas com as matrizes de projeção \mathbf{P} e \mathbf{P}' por uma transformação projetiva, \mathbf{T}_{PE} , ou seja,

$$\mathcal{P} = \mathbf{P}\mathbf{T}_{PE} \quad \text{e} \quad \mathcal{P}' = \mathbf{P}'\mathbf{T}_{PE}. \quad (2.48)$$

Além disso, o teorema a seguir fornece um resultado importante e bastante utilizado neste trabalho.

Teorema 2.5.2 *Dado um conjunto binocular, a transformação \mathbf{T}_{PE} , que relaciona as matrizes de projeção, \mathbf{P} e \mathbf{P}' , com as matrizes \mathcal{P} e \mathcal{P}' da calibração projetiva, é dada por*

$$\mathbf{T}_{PE} = \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{w}^T & w \end{bmatrix}, \quad (2.49)$$

onde $\mathbf{W}_{\infty} = [\mathbf{w}^T, w]^T$ é um vetor 4×1 .

Prova: A equação (2.7) (pág. 7), pode ser reescrita como

$$\tilde{\mathbf{m}}_1 \simeq \mathcal{P}\mathbf{T}_{PE}^{-1}\mathbf{T}_{PE}\tilde{\mathcal{M}}, \quad (2.50)$$

onde \mathcal{M} representa um ponto 3D no espaço \mathcal{P}^3 .

Agora, reescrevendo \mathcal{P} na equação (2.45) como sendo

$$\mathcal{P} \simeq [\mathbf{I} \ \mathbf{0}] \simeq [\mathbf{A} \ \mathbf{0}] \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{w}^T & w \end{bmatrix}. \quad (2.51)$$

De forma semelhante, substituindo (2.46) em (2.45), tem-se

$$\begin{aligned} \mathbf{P}' &\simeq [\mathbf{H}_\infty + \tilde{\mathbf{e}}_2 \mathbf{w}^T \quad w \tilde{\mathbf{e}}_2] \\ &\simeq [\mathbf{A}'\mathbf{R}\mathbf{A}^{-1} + \mathbf{A}'\mathbf{t}\mathbf{w}^T \quad w\mathbf{A}'\mathbf{t}] \\ &\simeq [\mathbf{A}'\mathbf{R} \quad \mathbf{A}'\mathbf{t}] \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{w}^T & w \end{bmatrix}, \end{aligned} \quad (2.52)$$

onde $\tilde{\mathbf{e}}_2 = \mathbf{A}'\mathbf{t}$ e $\mathbf{H}_\infty = \mathbf{A}'\mathbf{R}\mathbf{A}^{-1}$.

Agora, substituindo (2.51) e (2.52) em (2.50), tem-se

$$\tilde{\mathbf{m}}_1 \simeq \mathbf{P} \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{w}^T & w \end{bmatrix} \mathbf{T}_{PE}^{-1} \mathbf{T}_{PE} \tilde{\mathbf{M}} \quad (2.53)$$

e

$$\tilde{\mathbf{m}}_2 \simeq \mathbf{P}' \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{w}^T & w \end{bmatrix} \mathbf{T}_{PE}^{-1} \mathbf{T}_{PE} \tilde{\mathbf{M}}. \quad (2.54)$$

Desde que $\tilde{\mathbf{M}} \simeq \mathbf{T}_{PE} \tilde{\mathbf{M}}$,

$$\begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{w}^T & w \end{bmatrix} \mathbf{T}_{PE}^{-1} \simeq \mathbf{I},$$

de onde obtem-se (2.49). □

Apesar de \mathbf{T}_{PE} ser, normalmente, desconhecida, ela pode ser estimada se as coordenadas Euclidianas de alguns pontos do ambiente forem conhecidas ou de forma automática (Horaud et al., 2000).

2.5.2 Autocalibração por Etapas

Na figura 2.11, um conjunto binocular é movido da posição 1 para a posição 2 por um deslocamento euclidiano \mathbf{D}_{12} . Além disso, o ponto \mathbf{M} no espaço está relacionado às correspondências $(\mathbf{m}_{11} \leftrightarrow \mathbf{m}_{21})$ e $(\mathbf{m}_{12} \leftrightarrow \mathbf{m}_{22})$, respectivamente, nas posições 1 e 2. Assim, dadas as matrizes de projeção do conjunto binocular, é possível encontrar duas reconstruções 3D diferentes para o mesmo ponto \mathbf{M} , por exemplo, $\tilde{\mathbf{X}}_1$ e $\tilde{\mathbf{X}}_2$ para as posições 1 e 2 respectivamente. Evidentemente, essas reconstruções estão relacionadas por

$$\tilde{\mathbf{X}}_2 \simeq \mathbf{D}_{12} \tilde{\mathbf{X}}_1, \quad (2.55)$$

onde

$$\mathbf{D}_{12} = \begin{bmatrix} \mathbf{R}_{12} & \mathbf{t}_{12} \\ \mathbf{0}^T & 1 \end{bmatrix} \quad (2.56)$$

é a matriz que representa a rotação (\mathbf{R}_{12}) e a translação (\mathbf{t}_{12}) do conjunto binocular na posição 2 com relação a este mesmo conjunto na posição 1.

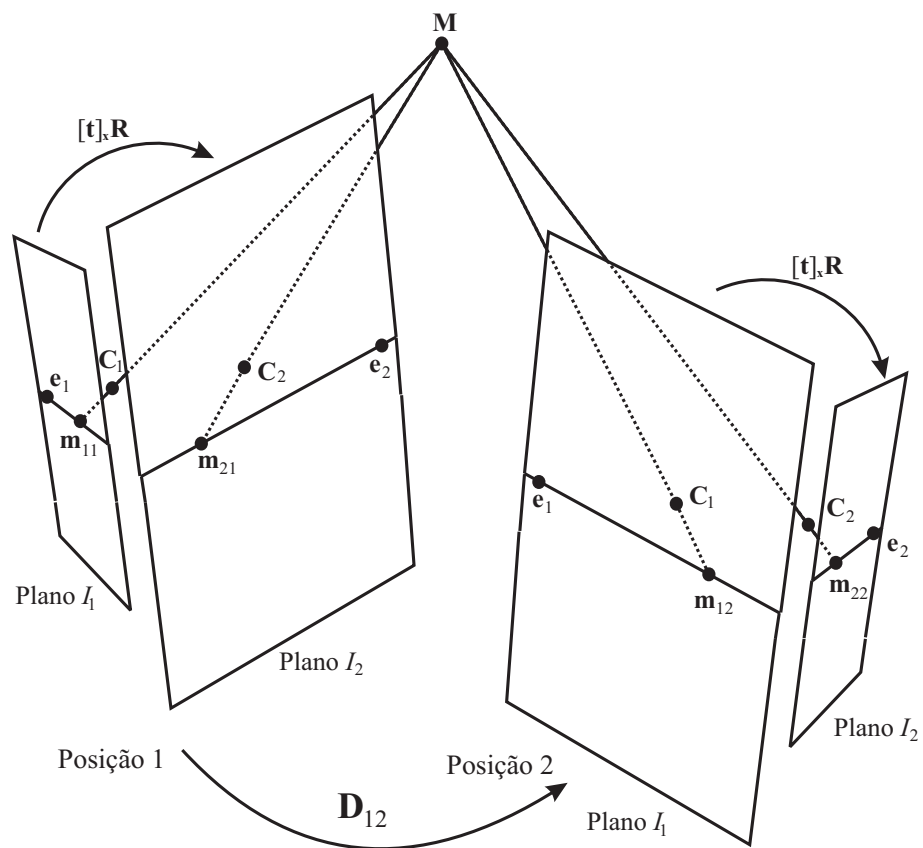


Figura 2.11 – Deslocamento euclidiano desconhecido de um conjunto binocular.

Se o sistema não estiver calibrado e o deslocamento é desconhecido, não é possível realizar a reconstrução euclidiana. Contudo, utilizando o teorema 2.5.1, é possível encontrar as matrizes \mathcal{P} e \mathcal{P}' e realizar uma reconstrução projetiva (\mathcal{X}_1 e \mathcal{X}_2). Além disso, pelo teorema 2.5.2, sabe-se que tal reconstrução relaciona-se com a reconstrução euclidiana por uma transformação projetiva, \mathbf{T}_{PE} , ou seja,

$$\tilde{\mathcal{X}}_1 \simeq \mathbf{T}_{PE} \tilde{\mathcal{X}}_1 \quad \text{e} \quad \tilde{\mathcal{X}}_2 \simeq \mathbf{T}_{PE} \tilde{\mathcal{X}}_2, \quad (2.57)$$

onde $\tilde{\mathcal{X}}_1$ é devido a ($\mathbf{m}_{11} \leftrightarrow \mathbf{m}_{21}$) e $\tilde{\mathcal{X}}_2$ a ($\mathbf{m}_{12} \leftrightarrow \mathbf{m}_{22}$).

De (2.55) e (2.57), segue que \mathcal{X}_1 e \mathcal{X}_2 relacionam-se por

$$\tilde{\mathcal{X}}_2 \simeq \mathbf{H}_{12} \tilde{\mathcal{X}}_1, \quad (2.58)$$

onde

$$\mathbf{H}_{12} = s_\alpha \mathbf{T}_{PE}^{-1} \mathbf{D}_{12} \mathbf{T}_{PE}, \quad (2.59)$$

na qual s_α é uma constante desconhecida e não-nula.

A equação (2.58) possui três equações independentes. Assim, desde que a matriz $\mathbf{H}_{12} \in \mathbb{R}^{4 \times 4}$ tem quinze graus de liberdade, dado um conjunto de $n \geq 5$ correspondências ($\tilde{\mathcal{X}}_1 \leftrightarrow \tilde{\mathcal{X}}_2$), é possível determinar \mathbf{H}_{12} univocamente. Veja, por exemplo, o trabalho de Csurka et al. (1999) para uma análise crítica de diversos métodos que realizam tal determinação.

Zisserman et al. (1995) observaram que \mathbf{H}_{12} concentra diversas propriedades do conjunto binocular. De fato, se apenas \mathbf{H}_{12} e a matriz fundamental do conjunto binocular forem conhecidas, é possível determinar \mathbf{T}_{PE} univocamente e, dessa forma, calibrar o sistema. As etapas que compõem o método são descritas em detalhes a seguir.

Normalização

Sempre é possível encontrar uma transformação de similaridade, \mathbf{T}_S , que transforme um deslocamento euclidiano qualquer (conhecido como movimento genérico¹⁴) em um movimento plano¹⁵ (equivalente a uma rotação em torno do eixo do movimento¹⁶ e uma translação perpendicular a esse mesmo eixo (Hartley e Zisserman, 2000, Sec.2.4.1)). Assim, considerando θ_{12} como o ângulo de rotação em torno do eixo do movimento, \mathbf{D}_{12} é dada por

$$\mathbf{D}_{12} = \begin{bmatrix} \cos \theta_{12} & -\text{sen} \theta_{12} & 0 & 0 \\ \text{sen} \theta_{12} & \cos \theta_{12} & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (2.60)$$

a menos de uma transformação de similaridade.

¹⁴Do inglês: “*geral motion*”.

¹⁵Do inglês: “*planar motion*”.

¹⁶Do inglês: “*screw axis*”.

De (2.59), observa-se que \mathbf{H}_{12} e a matriz definida em (2.60) estão relacionadas por duas transformações de similaridade, ou seja, \mathbf{T}_{PE} e \mathbf{T}_S . Assim, desde que determinantes e traços são invariantes a tais transformações, a constante s_α em (2.59) pode ser conhecida de (2.60) fazendo

$$\text{trace}(\mathbf{H}_{12}) = \text{trace}(s_\alpha \mathbf{D}_{12}) = 2s_\alpha(1 + \cos \theta_{12}) \quad (2.61)$$

$$\det(\mathbf{H}_{12}) = \det(s_\alpha \mathbf{D}_{12}) = s_\alpha^4. \quad (2.62)$$

De (2.62), $|s_\alpha| = \sqrt[4]{\det(\mathbf{H}_{12})}$. Assim, desde que $(1 + \cos \theta) \geq 0$ em (2.61), tem-se

$$s_\alpha = \text{sign}(\text{trace}(\mathbf{H}_{12})) \sqrt[4]{\det(\mathbf{H}_{12})}. \quad (2.63)$$

Deste ponto em diante, com s_α conhecida, \mathbf{H}_{12} será considerada uma transformação normalizada, ou seja,

$$\mathbf{H}_{12} \simeq \frac{1}{s_\alpha} \mathbf{H}_{12} = \mathbf{T}_{PE}^{-1} \mathbf{D}_{12} \mathbf{T}_{PE}. \quad (2.64)$$

Estimação do plano no infinito

Na autocalibração por etapas, é necessário converter uma estrutura projetiva 3D, representada pelas matrizes \mathcal{P} e \mathcal{P}' , em uma estrutura Euclidiana (representada por \mathbf{P} e \mathbf{P}'). Essa conversão é um mapeamento do espaço \mathcal{P}^3 para o seu subespaço Euclidiano. Além disso, do teorema 2.5.2, tal mapeamento é realizado por \mathbf{T}_{PE} .

Para encontrar \mathbf{T}_{PE} , após a normalização de \mathbf{H}_{12} , o próximo passo é determinar \mathbf{W}_∞ . Isto é feito baseado no teorema a seguir.

Teorema 2.5.3 *Em um conjunto binocular, o vetor \mathbf{W}_∞ que faz parte de \mathbf{T}_{PE} , a transformação que converte uma calibração projetiva (representada por \mathcal{P} e \mathcal{P}') em uma Euclidiana, é dado pelo autovetor de \mathbf{H}_{12}^{-T} que corresponde ao autovalor unitário.*

A prova do teorema anterior requer a demonstração dos lemas a seguir.

Lema 2.5.3.1 *Se \mathbf{D}_{12} descreve um deslocamento de um conjunto binocular, o autovetor de \mathbf{D}_{12}^{-T} que corresponde ao autovalor unitário é o plano no infinito.*

Prova: Desde que \mathbf{D}_{12} é dado por (2.56), seus autovalores são $\lambda \in \{e^{i\theta_{12}}, e^{-i\theta_{12}}, 1, 1\}$. Portanto, os autovalores de \mathbf{D}_{12}^{-T} são $\lambda_{-T} \in \{e^{-i\theta_{12}}, e^{i\theta_{12}}, 1, 1\}$. Isto prova que \mathbf{D}_{12}^{-T} possui, pelo menos um, autovalor unitário.

Agora, pela definição de autovetor, um autovetor, \mathbf{V}_{Di} , que corresponde ao autovalor unitário é

$$\mathbf{D}_{12}^{-T} \mathbf{V}_{Di} = \mathbf{V}_{Di}. \quad (2.65)$$

Como, de (2.60),

$$\mathbf{D}_{12}^{-T} = \begin{bmatrix} \mathbf{R}_{12} & \mathbf{0} \\ -\mathbf{t}_{12}^T \mathbf{R}_{12} & 1 \end{bmatrix},$$

e considerando $\mathbf{V}_{Di} = [\mathbf{v}^T, v]^T$, tem-se

$$\mathbf{R}_{12}\mathbf{v} = \mathbf{v} \quad (2.66)$$

$$-\mathbf{t}_{12}^T \mathbf{R}_{12}\mathbf{v} + v = v. \quad (2.67)$$

A equação (2.67) leva a seguinte equação

$$\mathbf{t}_{12}^T \mathbf{R}_{12}\mathbf{v} = 0. \quad (2.68)$$

Como, da definição de autovetor, $\mathbf{V}_{Di} \neq \mathbf{0}$, o sistema de equações formado por (2.66) e (2.68) tem duas soluções, ou seja, $\mathbf{V}_{D1} = [\mathbf{0}^T \ 1]^T$ e $\mathbf{V}_{D2} = [\mathbf{n}^T \ 1]^T$, onde \mathbf{n} é um vetor paralelo ao eixo de rotação ($\mathbf{R}_{12}\mathbf{n} = \mathbf{n}$) e perpendicular ao plano de translação ($\mathbf{t}_{12}^T \mathbf{n} = 0$). Como $\mathbf{V}_{D1} = [\mathbf{0}^T \ 1]^T$ é a equação do plano no infinito, isto conclui a demonstração. \square

Lema 2.5.3.2 Se \mathbf{V}_{Ti} é um autovetor que corresponde ao autovalor unitário de \mathbf{H}_{12}^{-T} e \mathbf{V}_{Di} é o um autovetor que corresponde ao autovalor unitário de \mathbf{D}_{12}^{-T} , tem-se

$$\mathbf{V}_{Ti} = \mathbf{T}_{PE}^T \mathbf{V}_{Di}. \quad (2.69)$$

Prova: Da equação (2.64), tem-se

$$\mathbf{H}_{12}^{-T} = \mathbf{T}_{PE}^T \mathbf{D}_{12}^{-T} \mathbf{T}_{PE}^{-T}.$$

Explicitando \mathbf{D}_{12}^{-T} na equação anterior e substituindo em (2.65), tem-se

$$\begin{aligned} \mathbf{T}_{PE}^{-T} \mathbf{H}_{12}^{-T} \mathbf{T}_{PE}^T \mathbf{V}_{Di} &= \mathbf{V}_{Di} \\ \mathbf{H}_{12}^{-T} (\mathbf{T}_{PE}^T \mathbf{V}_{Di}) &= (\mathbf{T}_{PE}^T \mathbf{V}_{Di}). \end{aligned}$$

\square

O lema anterior mostra que, dado \mathbf{V}_{D1} ,

$$\mathbf{V}_{T1} = \begin{bmatrix} \mathbf{A}^{-T} & \mathbf{w} \\ \mathbf{0} & w \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = [\mathbf{w}^T, w]^T = \mathbf{W}_{\infty}^T$$

ou seja, \mathbf{W}_{∞} é o autovetor de \mathbf{H}_{12}^{-T} que corresponde ao autovalor unitário. Isto conclui a prova do teorema 2.5.3.

Se \mathbf{D}_{12} corresponde a um movimento genérico, \mathbf{H}_{12}^{-T} possui apenas um autovetor unitário. Assim, como \mathbf{H}_{12}^{-T} e \mathbf{H}_{12}^T possuem os mesmo autovetores (Horaud et al., 2000), \mathbf{W}_{∞} é igual ao autovetor de \mathbf{H}_{12}^T que corresponde ao único autovalor unitário e pode ser conhecido resolvendo-se ($\mathbf{H}_{12}^T -$

$\mathbf{I})\mathbf{W}_\infty = \mathbf{0}$. Além disso, dado n movimentos genéricos, $\mathbf{D}_{i(i+1)}$, a cada um desses é possível associar uma matriz normalizada, ou seja, $\mathbf{H}_{12}^T, \mathbf{H}_{23}^T, \dots, \mathbf{H}_{n(n+1)}^T$. Como \mathbf{T}_{PE} é independente do deslocamento $\mathbf{D}_{i(i+1)}$, fica claro que o autovalor que corresponde ao autovetor unitário de todas essas matrizes normalizadas são idênticos. Então, \mathbf{W}_∞ pode ser estimado resolvendo o seguinte sistema de equações

$$\begin{bmatrix} \mathbf{H}_{12}^T - \mathbf{I} \\ \mathbf{H}_{23}^T - \mathbf{I} \\ \vdots \\ \mathbf{H}_{n(n+1)}^T - \mathbf{I} \end{bmatrix} \mathbf{W}_\infty = \mathbf{0}. \quad (2.70)$$

Se \mathbf{D}_{12} corresponde a um movimento planar, pela demonstração do lema 2.5.3.1, \mathbf{D}_{12}^{-T} possui o autovalor unitário com multiplicidade dois. Assim, \mathbf{H}_{12}^T possui dois autovalores que correspondem ao autovalor unitário, ou seja, não é possível estimar \mathbf{W}_∞ univocamente com apenas um movimento planar. Contudo, se o conjunto binocular realiza $n \geq 2$ movimentos genéricos, tem-se $\mathbf{H}_{12}^T, \mathbf{H}_{23}^T, \dots, \mathbf{H}_{n(n+1)}^T$. Além disso, cada uma dessas matrizes têm dois autovetores associados ao autovalor unitário. Um desses autovetores é comum as duas matrizes e define o plano no infinito. O outro é totalmente desconhecido e depende da direção do eixo do movimento. Assim, desde que os movimentos são distintos, pois o eixo de rotação são diferentes, \mathbf{W}_∞ é o único autovetor comum as matrizes e pode ser conhecido resolvendo-se o sistema (2.70) com $n \geq 2$.

Estimando a homografia do infinito

É possível estabelecer uma homografia do infinito¹⁷, representada por $\mathbf{H}_{\infty 12}$, entre o plano de imagem I_1 na posição 1 e I_1 na posição 2 (figura 2.11). Se $\mathbf{H}_{\infty 12}$ for conhecida, é possível estimar a matriz \mathbf{A} (de França, 2005, seção A.4).

Horaud et al. (2000) propuseram um método, resumido na proposição a seguir, que permite estimar $\mathbf{H}_{\infty 12}$ diretamente de \mathbf{H}_{12} .

Proposição 2.5.1 Considerando $\mathbf{W}_\infty = [\mathbf{w}^T w]^T$ e a homografia \mathbf{H}_{12} dada por

$$\mathbf{H}_{12} = \begin{bmatrix} \mathbf{H}_{12}(1:3, 1:3) & \mathbf{H}_{12}(1:3, 4) \\ \mathbf{H}_{12}(4, 1:3) & \mathbf{H}_{12}(4, 4) \end{bmatrix}, \quad (2.71)$$

conhecidos, tem-se que

$$\mathbf{H}_{\infty 12} = \mathbf{H}_{12}(1:3, 1:3) - \frac{1}{w} \mathbf{H}_{12}(1:3, 4) \mathbf{w}^T.$$

Prova: Sabendo que as matrizes \mathbf{T}_{PE} e \mathbf{D}_{12} têm formato dado por (2.49) e (2.56) respectivamente,

¹⁷Considerando um plano, Π , no espaço projetivo 3D, se um ponto, \mathbf{M} , pertence a Π e \mathbf{m}_{1i} e \mathbf{m}_{2i} são projeções deste ponto nos planos I_1 e I_2 respectivamente, então, $\tilde{\mathbf{m}}_{2i} \simeq \mathbf{H}_\pi \tilde{\mathbf{m}}_{1i}$, onde \mathbf{H}_π é uma matriz 3×3 e é chamada de "homografia induzida pelo plano Π ".

pode-se reescrever (2.64) como sendo

$$\begin{aligned}
\mathbf{H}_{12} &= \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \frac{-\mathbf{w}^T \mathbf{A}}{w} & \frac{-1}{w} \end{bmatrix} \begin{bmatrix} \mathbf{R}_{12} & \mathbf{t}_{12} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{w}^T & w \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \frac{-\mathbf{w}^T \mathbf{A}}{w} & \frac{-1}{w} \end{bmatrix} \begin{bmatrix} \mathbf{R}_{12} \mathbf{A}^{-1} + \mathbf{t}_{12} \mathbf{w}^T & w \mathbf{t}_{12} \\ \mathbf{w}^T & w \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{A} \mathbf{R}_{12} \mathbf{A}^{-1} + \mathbf{A} \mathbf{t}_{12} \mathbf{w}^T & w \mathbf{A} \mathbf{t}_{12} \\ \frac{-\mathbf{w}^T \mathbf{A} \mathbf{R}_{12} \mathbf{A}^{-1} - \mathbf{A} \mathbf{T} \mathbf{w}^T + \mathbf{w}^T}{w} & -\mathbf{A} \mathbf{T} + 1 \end{bmatrix},
\end{aligned}$$

onde $\mathbf{A} \mathbf{T} = \mathbf{w}^T \mathbf{A} \mathbf{t}_{12}$.

Comparando a equação anterior com (2.71) e sabendo que $\mathbf{H}_{\infty 12} = \mathbf{A} \mathbf{R}_{12} \mathbf{A}^{-1}$, conclui-se a demonstração. \square

Se \mathbf{A} é obtida de uma única homografia do infinito, o resultado é muito susceptível a ruído e é extremamente dependente da exatidão da estimação de $\mathbf{H}_{\infty 12}$ (Hartley, 1997b; Zisserman et al., 1995). Por isso, na prática, deve-se realizar “diversos” deslocamentos, $\mathbf{D}_{n(n+1)}$, e usar a redundância para melhorar a estimação de \mathbf{A} .

Uma vez que \mathbf{W}_{∞} e \mathbf{A} são conhecidos, o teorema 2.5.2, deve ser utilizado para atualizar as matrizes \mathcal{P} e \mathcal{P}' .

Apesar da calibração por etapas ser constituída apenas de equações lineares, como todo método automático, a técnica é sensível a ruídos. Do trabalho de Horaud et al. (2000), sabe-se que a estimação do vetor \mathbf{W}_{∞} é a etapa mais crítica do processo, pois tal estimação é instável e constitui a principal fonte de erros. De fato, a técnica só fornece bons resultados se \mathbf{W}_{∞} for estimado na presença de um ruído gaussiano com desvio padrão de até 0,5 pixels. Contudo, desde que as transformações $\mathbf{H}_{i(i+1)}$ são estimadas usando correspondências de pontos obtidas com o sistema descalibrado, essa suposição nem sempre é verdadeira.

CONTRIBUIÇÕES

Como o título evidencia, o foco do presente trabalho concentra-se apenas na calibração de câmeras. Apesar disso, todos os conceitos e algoritmos discutidos no capítulo anterior estão relacionados ao problema de calibração. Neste capítulo, apresenta-se as contribuições para a área de calibração de câmeras que advêm deste trabalho. Estas, concentram-se na recente abordagem de calibração monocular com auxílio de um gabarito 1D e na calibração automática de um conjunto binocular.

3.1 Calibração com gabarito 1D

O método de calibração com um gabarito 1D proposto por Zhang (2004) é flexível, pois impõe apenas uma restrição ao movimento que o gabarito de calibração pode realizar, ou seja, uma de suas extremidades deve ser fixa. Contudo, a solução de (2.43) pode convergir para um mínimo local na falta de uma boa estimativa inicial das incógnitas envolvidas no problema. Tal estimativa deve ser obtida mesmo na presença de ruído. Por isso, nesta seção são propostas algumas melhorias ao método linear que o torna mais robusto.

3.1.1 Normalização dos Dados

Na visão computacional, as coordenadas das projeções são dadas em relação ao canto superior-esquerdo da imagem. Aplicando uma transformação projetiva, \mathbf{T}_2 , aos pontos, $\tilde{\mathbf{m}}$, da imagem, tem-se $\hat{\mathbf{m}} = \mathbf{T}_2\tilde{\mathbf{m}}$. Dessa forma, da equação (2.2), obtém-se $\hat{\mathbf{m}} \simeq \mathbf{T}_2\mathbf{A}[\mathbf{I} \ \mathbf{0}]\tilde{\mathbf{M}}$. Esta relação implica que $\hat{\mathbf{A}} = \mathbf{T}_2\mathbf{A}$ representa os parâmetros intrínsecos da câmara que correspondem as projeções $\hat{\mathbf{m}}$.

Como discutido no capítulo anterior, para o caso da calibração com um gabarito 1D, tem-se a equação (2.35) que leva a (2.41). De forma semelhante, considerando os pontos $\hat{\mathbf{m}}$ e a matriz $\hat{\mathbf{A}}$, obtém-se um sistema de equações da forma $\hat{\mathbf{U}}_n\hat{\mathbf{b}} = \mathbf{L}^2$. Além disso, é fácil demonstrar que $\hat{\mathbf{U}}_n = \mathbf{U}_n\mathbf{S}$, onde \mathbf{S} é uma matriz 4×4 que depende apenas dos elementos de \mathbf{T}_2 .

Com o resultado anterior, seria de se esperar que o problema de calibração com um gabarito 1D chegasse a exatamente a mesma solução caso fossem realizados os passos a seguir.

1. Transformar os pontos fazendo $\tilde{\mathbf{m}}_{ji} = \mathbf{T}_2 \tilde{\mathbf{m}}_{ji}$.
2. Encontrar $\hat{\mathbf{b}}$, resolvendo (2.42) e utilizando os pontos $\tilde{\mathbf{m}}_{ji}$.
3. Fazer $\mathbf{b} = \mathbf{S}\hat{\mathbf{b}}$.

Contudo, isto seria verdadeiro se e somente se $\hat{\mathbf{b}} = \mathbf{S}^{-1}\mathbf{b}$ fosse solução do problema

$$\min_{\mathbf{x}} \|(\mathbf{U}_n \mathbf{S})\mathbf{x} - \mathbf{L}^2\|.$$

No entanto, baseado nos resultados experimentais da próxima seção, isto não é verdadeiro. Dessa forma, a solução do problema de calibração com um gabarito 1D é extremamente dependente da matriz \mathbf{T}_2 . Sendo assim, é necessário encontrar a transformação \mathbf{T}_2 que forneça a melhor estimativa do vetor \mathbf{b} e, por tanto, da matriz $\hat{\mathbf{A}}$. Diversos autores já propuseram soluções para este problema Izquierdo e Guerra (2003); Torr (1995). Em particular, neste trabalho utilizou-se a transformação proposta por Hartley (1997a), ou seja, \mathbf{T}_2 deve ser tal que o centróide do conjunto de pontos \mathbf{m}_{ji} esteja na origem e a distância média de um ponto à origem seja $\sqrt{2}$. Esta mudança produz melhora significativa na exatidão da matriz \mathbf{A} . Além disso, o custo computacional acrescentado ao problema devido a esta transformação é desprezível.

Uma vez que os parâmetros intrínsecos da câmera tenham sido estimados no sistema de coordenadas definido por \mathbf{T}_2 , a matriz \mathbf{A} (do sistema de coordenadas original) pode ser obtida fazendo

$$\mathbf{A} = \mathbf{T}_2^{-1} \hat{\mathbf{A}}. \quad (3.1)$$

3.1.2 Considerando a Distorção Radial

No equacionamento das seções anteriores, a distorção presente nas lentes do conjunto binocular foi desprezada. Contudo, as câmeras de baixo custo costumam apresentar uma distorção radial significativa. Neste caso, o método proposto não pode ignorar tal problema. Felizmente, a distorção radial pode ser facilmente inserida no modelo da câmera bastando apenas considerar o ponto estimado como sendo

$$\tilde{\mathbf{m}}_{xji} = [\hat{u} + d\hat{u}, \hat{v} + d\hat{v}, 1]^T, \quad (3.2)$$

com (Salvi et al., 2002)

$$\begin{aligned} d\hat{u} &= (\hat{u} - u_0)k_1(\hat{x}^2 + \hat{y}^2) \\ d\hat{v} &= (\hat{v} - v_0)k_1(\hat{x}^2 + \hat{y}^2), \end{aligned}$$

onde $\hat{x} = (\hat{u} - u_0)$ e $\hat{y} = (\hat{v} - v_0)$.

Agora, o problema de calibração 1D pode ser resolvido minimizando a equação a seguir

$$\sum_{i=1}^n \left(\|\mathbf{m}_1 - \hat{\mathbf{m}}_1(\mathbf{A}, \mathbf{M}_1, k_1)\|^2 + \sum_{j=2}^q \|\mathbf{m}_{ji} - \hat{\mathbf{m}}_{ji}(\mathbf{A}, \mathbf{M}_{ji}, k_1)\|^2 \right), \quad (3.3)$$

onde $\hat{\mathbf{m}}_1(\mathbf{A}, \mathbf{M}_1, k_1)$ é a projeção de \mathbf{M}_{j_i} após a correção (3.2) e de acordo com (2.2).

Novamente, a solução do problema (3.3) requer um método de minimização não-linear que necessita de um valor inicial para k_1 . Neste caso, pode-se fazer, inicialmente, $k_1 = 0$.

3.2 Calibração Automática

Apesar de extremamente versátil, a calibração automática ainda não é amplamente utilizada na prática. Isto se deve principalmente a dois motivos. Primeiro, os algoritmos de calibração automática possuem um custo computacional extremamente elevado (Hayman e Murray, 2003). Isto limita o uso de tais algoritmos a plataformas com alto poder de processamento. Além disso, normalmente, a calibração automática nos leva a resultados muito menos exatos quando comparados a calibração clássica, que utiliza algum tipo de gabarito (Habed e Boufama, 2002).

Como discutido na seção 2.5.2, no caso de um conjunto binocular, o problema de calibração automática tem solução linear. Evidentemente, isto elimina o problema do custo computacional. Contudo, o problema da pouca exatidão ainda persiste (Horaud et al., 2000). No entanto, a metodologia de calibração discutida na seção 2.5.2 pode ser utilizada na robótica móvel com sucesso. Isto é possível porque, em tal aplicação, normalmente, a visão binocular é utilizada para identificação e localização de obstáculos presentes no campo de visão do robô e formação de um “mapa de obstáculos” (Burschka e Hager, 2002). Em tais aplicações, mesmo com um sistema binocular calibrado com auxílio de um gabarito, devido a complexidade inerente ao processo de correspondência de pontos, a localização dos obstáculos é conhecida com pouca exatidão (veja-se, por exemplo, os trabalho de Zisserman et al. (1995)). Dessa forma, o mapa de obstáculos é construído através de um tratamento estatístico dos dados. Tal tratamento pode lidar também com os erros oriundos do processo de calibração automática. Sendo assim, a visão binocular em conjunto com a calibração automática pode dotar o robô com um sistema de guiagem bastante flexível, pois, por exemplo, permite o uso de câmeras com *zoom*.

O processo de calibração binocular descrito na seção 2.5.2 necessita que o conjunto binocular realize um deslocamento euclidiano. Para um conjunto binocular fixo em um robô móvel, tal deslocamento seria apenas um deslocamento planar, ou seja, composto apenas por uma translação e uma rotação em torno do eixo y . Esta limitação faz com que seja necessária uma quantidade maior de deslocamentos, logo uma quantidade maior de pares de imagens e um conseqüente aumento no custo computacional do processo de calibração, e pode também levar a dados mal-condicionados que impossibilitem a solução do problema (Horaud et al., 2000). Em vista disto, como parte deste trabalho, está sendo construída uma “cabeça binocular” semelhante a ilustrada na figura 3.1, adaptada de (Bernardino, 2004). Tal dispositivo permitirá mover o conjunto binocular em torno dos eixos x e y , além

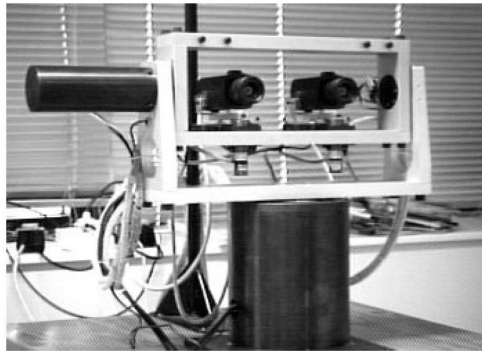


Figura 3.1 – Ilustração da “cabeça binocular” que está sendo desenvolvida para calibração do conjunto binocular.

de permitir também uma translação na direção y . Estes movimentos adicionais devem permitir que o sistema possa ser calibrado de forma automática com uma maior robustez.

Neste trabalho, através de simulação, o processo de calibração automática será realizado com dados proveniente de duas situações: a) um conjunto binocular que realiza um deslocamento euclidiano completo, ou seja, com rotação e translação na três direções, e; b) uma plataforma semelhante a da figura 3.1, que realiza um deslocamento limitado. O objetivo é comparar os resultados obtidos nas duas situações (quanto a exatidão dos parâmetros) para prover subsídios para a construção de um sistema de visão para um robô móvel.

AVALIAÇÃO DOS ALGORITMOS IMPLEMENTADOS

Neste capítulo apresenta-se a metodologia de testes proposta para avaliar-se o desempenho dos algoritmos implementados. Os testes visam principalmente avaliar os resultados dos métodos quanto ao ruído.

4.1 Calibração com gabarito 1D

Na avaliação deste algoritmo, serão utilizados dados sintéticos e dados adquiridos de imagens reais. O algoritmo foi implementado no matlab. Os dados sintéticos foram obtidos a partir de um bastão simulado também no matlab, como esboçado na figura 4.1.

4.1.1 Dados Sintéticos

Os dados sintéticos foram obtidos supondo uma câmera com os seguintes parâmetros: $\alpha = 842$; $\beta = 879$; $s_k = 0$; $u_0 = 358$; $v_0 = 207$. A resolução da imagem é feita igual a 640×512 pixels. Além disso, foi simulado um bastão de 30 cm de comprimento e três pontos co-lineares e equidistantes. A distância entre um ponto e outro no bastão variou de acordo com o número de pontos simulados (o espaçamento entre dois pontos consecutivos foram sempre iguais).

Na geração dos dados sintéticos, os ângulos da equação (2.44), $\theta_i \in [-\pi/2, \pi/2]$ e $\phi_i \in [-\pi/2, \pi/2]$, variaram aleatoriamente mas de acordo com uma distribuição uniforme. Tais dados foram utilizados para avaliar o desempenho do método com respeito aos seguintes parâmetros: nível de ruído presente nos dados, o número de pares de imagens utilizados e o número de pontos no gabarito.

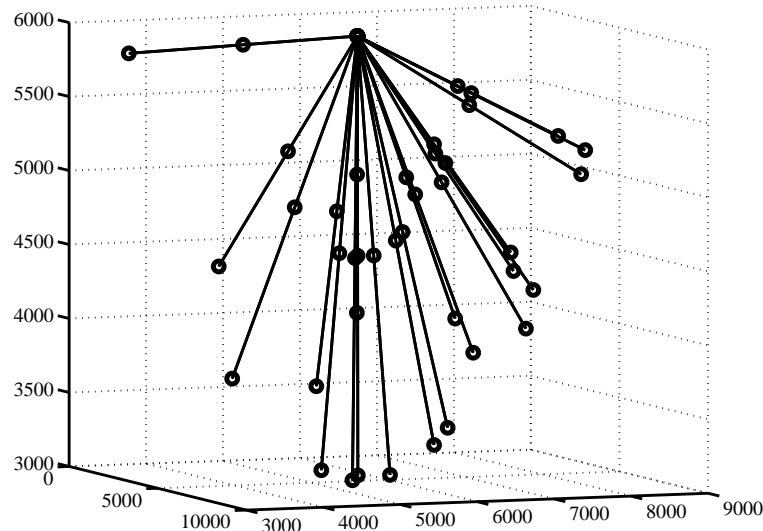


Figura 4.1 – Esboço do movimento realizado pelo gabarito 1D de três pontos durante uma simulação.

Quanto ao nível de ruído

Este experimento foi utilizado para avaliar o desempenho do método com respeito ao nível de ruído presente nos dados. Para isto, ruído gaussiano de média zero e desvio padrão σ foi acrescentado aos pontos projetados nas imagens sintéticas. Tal ruído variou de 0,1 a 1,0 pixel. Para cada nível de ruído foram realizadas 250 simulações e a mediana de cada parâmetro intrínseco foi retido. Essas medianas foram comparadas com os parâmetros da câmera simulada. Em cada simulação, os parâmetros intrínsecos da câmera foram estimados pelo método de Zhang (2004) e pelo proposto no presente trabalho. Os resultados da comparação são apresentados nas figuras 4.2 e 4.3. Nestas, os erros são apresentados de forma relativa (em porcentagem).

Analisando os resultados observa-se que a exatidão do método que utiliza a normalização proposta degrada-se com o aumento do nível de ruído presente nos dados. Por outro lado, a exatidão do método de Zhang varia mais rapidamente. Além disso, a exatidão do método proposto é da ordem de 10 vezes maior que a do método de Zhang. Em especial, para $\sigma = 0,5$ (valor maior do que os encontrados na prática), tem-se um erro para α e β em torno de 0,1 %. De fato, com a normalização proposta os resultados obtidos são melhores do que, até mesmo, os alcançados com o método não-linear proposto por Zhang e resumido na equação (2.43). Isto pode ser observado através da figura 4.4. Esta mostra o resultado da minimização não-linear que utilizou um valor inicial obtido sem auxílio da normalização proposta. Observa-se que a solução não-linear reduz pela metade o erro obtido na solução linear sem normalização (figura 4.3). Contudo, a solução não-linear apresenta um erro bem maior que o obtido pelo método linear, cujos dados foram normalizados (figura 4.2). Evidentemente, o algoritmo de minimização não-linear ficou preso em um mínimo local devido a inicialização ruim dos parâmetros a serem estimados.

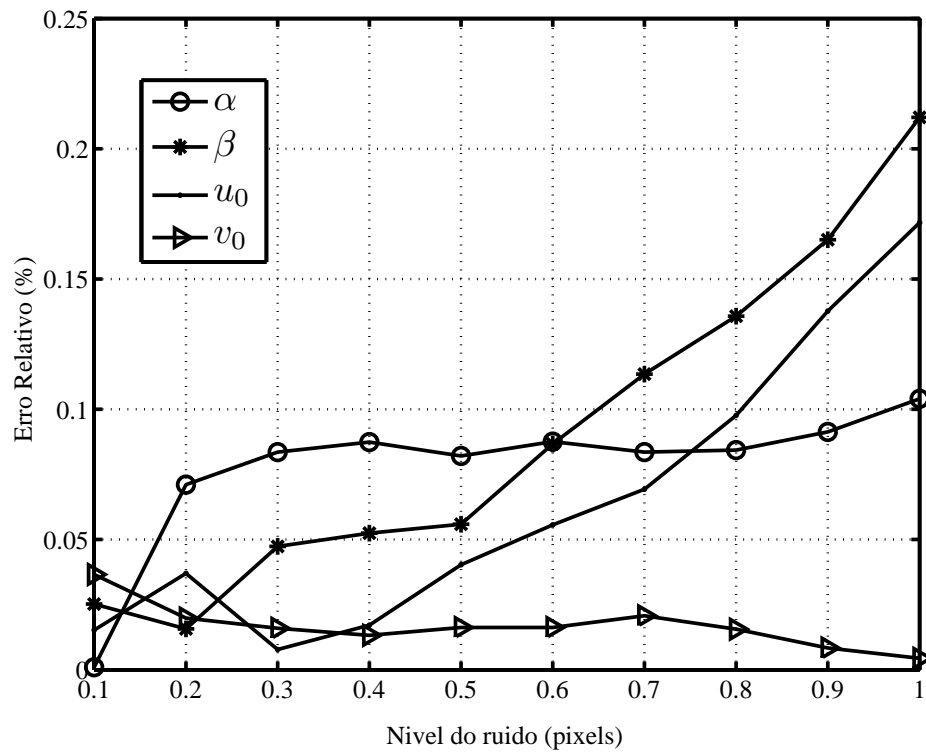


Figura 4.2 – Erro relativo obtido na estimativa dos parâmetros intrínsecos da câmera pelo método linear com dados normalizados.

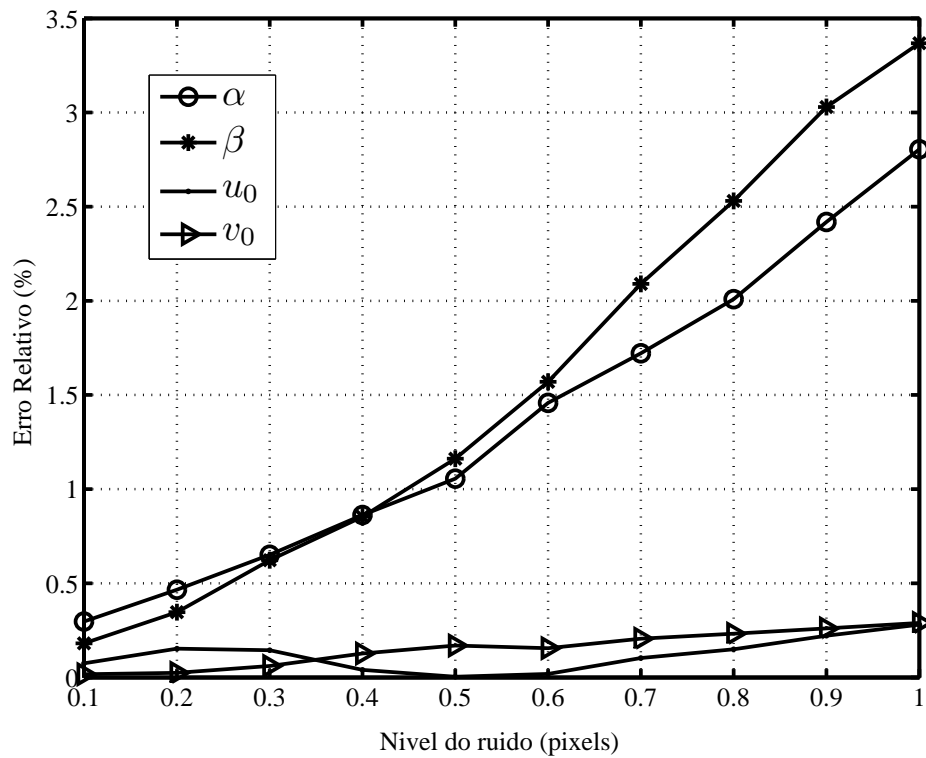


Figura 4.3 – Erro relativo obtido na estimativa dos parâmetros intrínsecos da câmera pelo método linear com dados sem normalização.

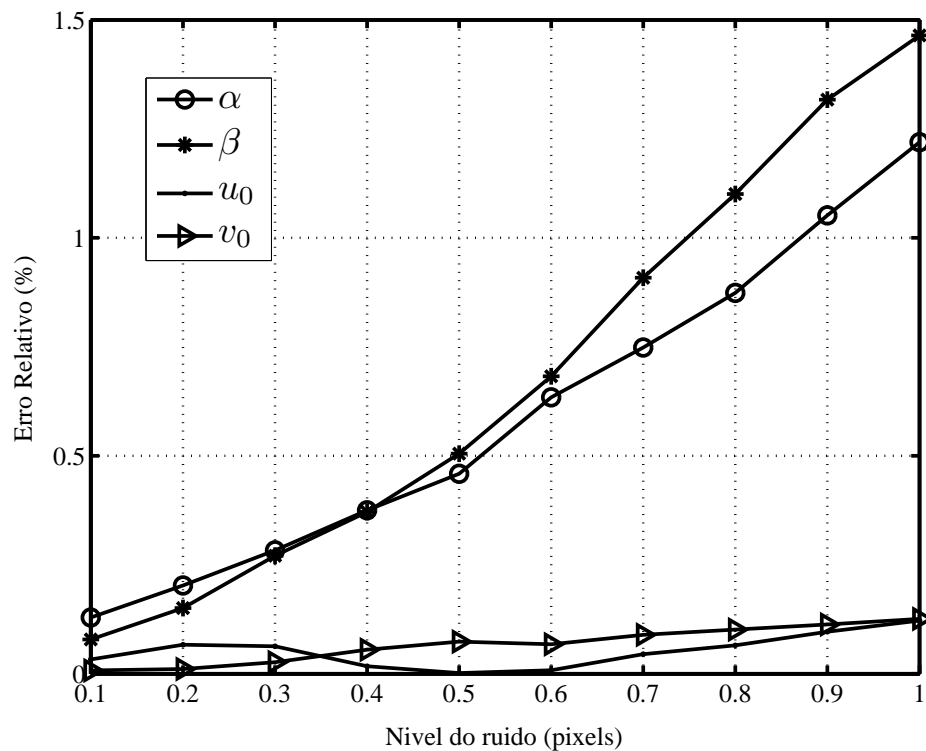


Figura 4.4 – Erro relativo obtido na estimação dos parâmetros intrínsecos da câmera pelo método não-linear inicializado com parâmetros obtidos com dados sem normalização.

Quanto ao número de imagens

Este experimento investiga o desempenho do algoritmo proposto com relação ao número de imagens utilizado na calibração. Na simulação, o número de imagens variou de 6 até 100, enquanto que o nível de ruído permaneceu constante, com $\sigma = 0,4$. Novamente, para uma mesma quantidade de imagens, os parâmetros retidos são a mediana de 250 simulações. Como observa-se na figura 4.5, o erro decresce quando mais imagens são utilizados. Em especial, quando esse número varia de 6 para 30 ocorre uma grande redução no erro. Contudo, com 50 pares de imagens ou mais, o erro torna-se praticamente constante.

Quanto ao número de pontos do gabarito

Com um gabarito com mais de três pontos, não se acrescenta mais equações independentes ao problema de calibração 1D. Contudo, teoricamente, o uso de mais pontos deve aumentar a redundância e, portanto, melhorar a robustez do algoritmo. Este experimento visa avaliar a quantidade de pontos ideal para o gabarito 1D. A quantidade de pontos variou de três (o mínimo possível) até dez, enquanto que o nível de ruído permanecerá constante, com $\sigma = 0,4$. Como nos outros casos, para uma mesma quantidade de pontos, os parâmetros retidos serão a mediana de 250 simulações. Os resultados são apresentados na figura 4.6. Observa-se que o erro permanece quase constante, independente

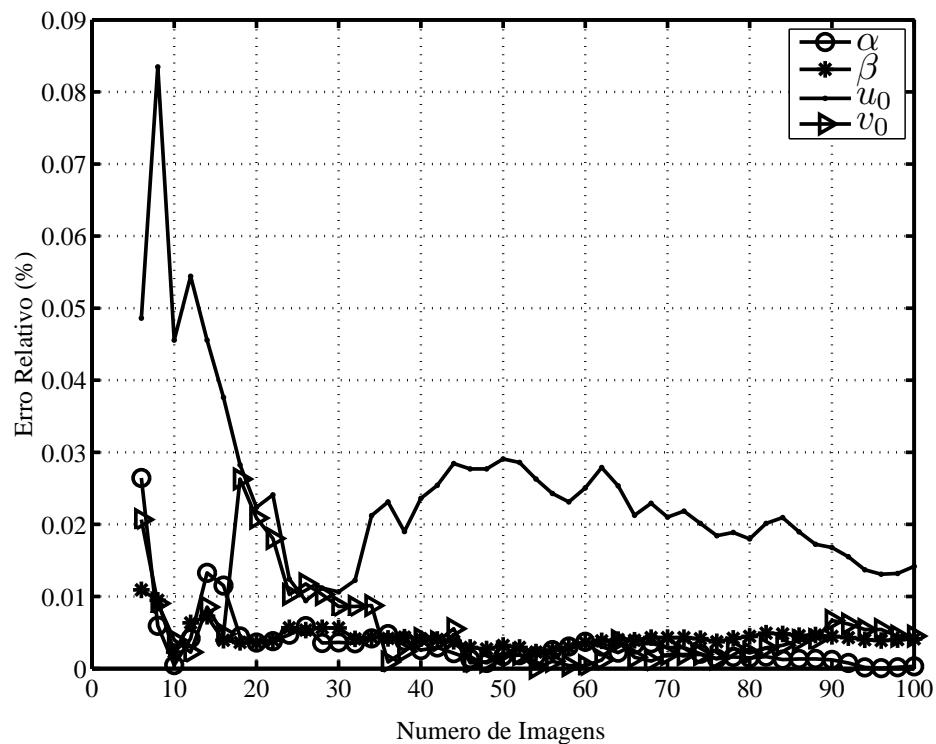


Figura 4.5 – Erro relativo obtido na estimação dos parâmetros intrínsecos da câmera pelo método linear com dados normalizados.

do número de pontos. Assim, desde que mais pontos implicam em um aumento na complexidade do problema, o uso de um gabarito com apenas três pontos é aconselhável.

4.1.2 Dados Reais

Para adquirir os dados reais, foi utilizado um gabarito com três esferas de plástico branco, com 2 cm de diâmetro cada uma. Tais esferas foram fixadas em um bastão de metal. A distância entre os centróides de esferas vizinhas foi feita, com auxílio de uma régua comum, igual 15 cm. Manualmente, o gabarito foi movido em diferentes posições a uma distância de aproximadamente 50 cm da câmera. Durante o movimento, 20 imagens foram adquiridas (uma dessas é apresentada na figura 4.7). Devido as imagens serem consecutivas, desde que a localização das esferas seja marcada manualmente na primeira imagem e os respectivos centróides conhecidos, é simples identificá-las nas outras imagens com o auxílio de um algoritmo de rastreamento¹ baseado em um algoritmo *blob* gaussiano (Lindeberg, 1994, Cap. 7). A localização do centróide de cada esfera foi utilizada como dado de entrada para o algoritmo proposto. O resultado deste experimento é apresentado tabela 4.1. O resultado referenciado como **NORMALIZADO** foi obtido com o método linear e a normalização proposta. O resultado **ZHANG** é a solução não-linear do problema de calibração 1D sem normalização.

A título de comparação, os parâmetros intrínsecos da câmera utilizada nos testes também foram estimados pelo método proposto por Zhang (2000). O gabarito utilizado neste método é um plano

¹Do inglês: *tracking*.

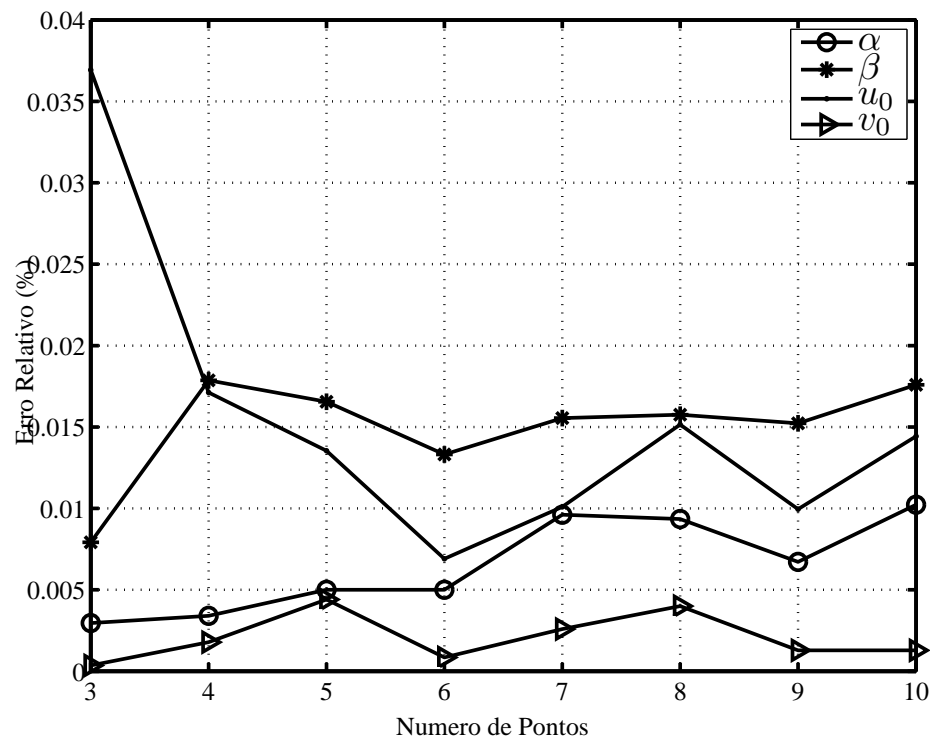


Figura 4.6 – Erro relativo obtido na estimação dos parâmetros intrínsecos da câmera pelo método linear com dados normalizados.

Tabela 4.1 – Parâmetros intrínsecos estimados pelo algoritmo linear.

Método	Parâmetros			
	α	β	u_0	v_0
NORMALIZADO	844,9312	876,9088	355,5348	208,4789
ZHANG	852,1759	894,4997	365,9896	203,3809
BOUGUET	842,4040	879,5474	358,0411	207,2355

com um padrão quadriculado. Neste trabalho, utilizou-se um gabarito planar com 121 pontos (figura 4.8). Durante a calibração, foram realizados 10 deslocamentos e uma seqüência com 10 imagens foram capturadas. Em seguida, a implementação de Bouguet (2004) do algoritmo de Zhang (2000) foi utilizada para estimar os parâmetros da câmera. O resultado desta calibração está apresentado na terceira linha da tabela 4.1 e é referenciado por **BOUGUET**. Observa-se que o resultado **ZHANG** está bastante diferente do obtido pelo algoritmo Bouguet (2004). Por outro lado, mesmo sendo obtido de forma linear, o resultado **NORMALIZADO** é muito próximo do **BOUGUET**. Por último, deve-se observar que o critério não-linear proposto na equação (3.3) foi suficiente para modelar a distorção radial da câmera. Isto pode ser comprovado observado os valores muito próximos obtidos para a constante k_1 em ambos os métodos **NORMALIZADO** e **ZHANG** (tabela 4.2).

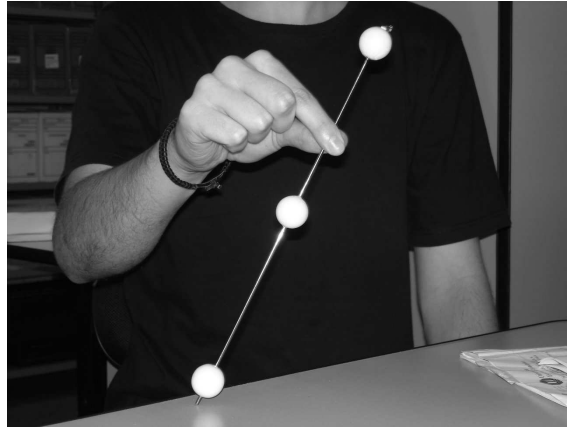


Figura 4.7 – Gabarito 1D utilizado na avaliação do método proposto.

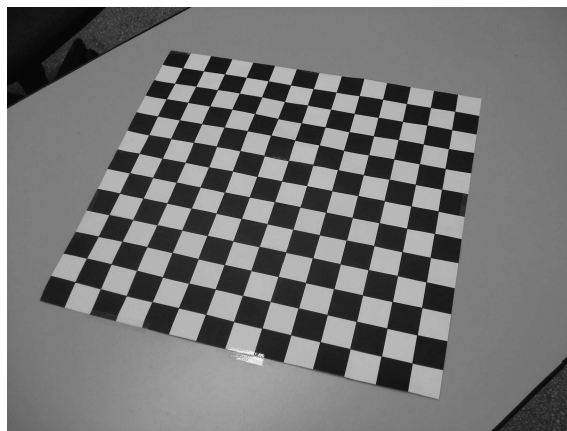


Figura 4.8 – Gabarito plano utilizado na comparação do método proposto.

4.2 Calibração Automática

A avaliação das rotinas de calibração automática foi realizada com auxílio de um gabarito de calibração (figura 4.9). Contudo, nenhuma informação prévia sobre as dimensões do gabarito foi utilizada no processo de calibração. O gabarito foi necessário apenas para facilitar o rastreamento de pontos entre imagens consecutivas. No caso, os pontos rastreados foram os vértices dos quadrados que compõem o gabarito. As localizações desses vértices foram obtidas com exatidão de 0,1 pixel, com uma adaptação na rotina de Bouguet (2004) para extração de cantos em uma plano.

Tabela 4.2 – Parâmetros intrínsecos estimados pelo algoritmo não-linear proposto.

Método	Parâmetros				
	α	β	u_0	v_0	k_1
NORMALIZADO	843,7335	880,0086	354,1828	208,1790	-0,1138
ZHANG	8472,5945	884,6538	360,4297	203,1875	-0,1165
BOUGUET	842,4040	879,5474	358,0411	207,2355	-0,1568

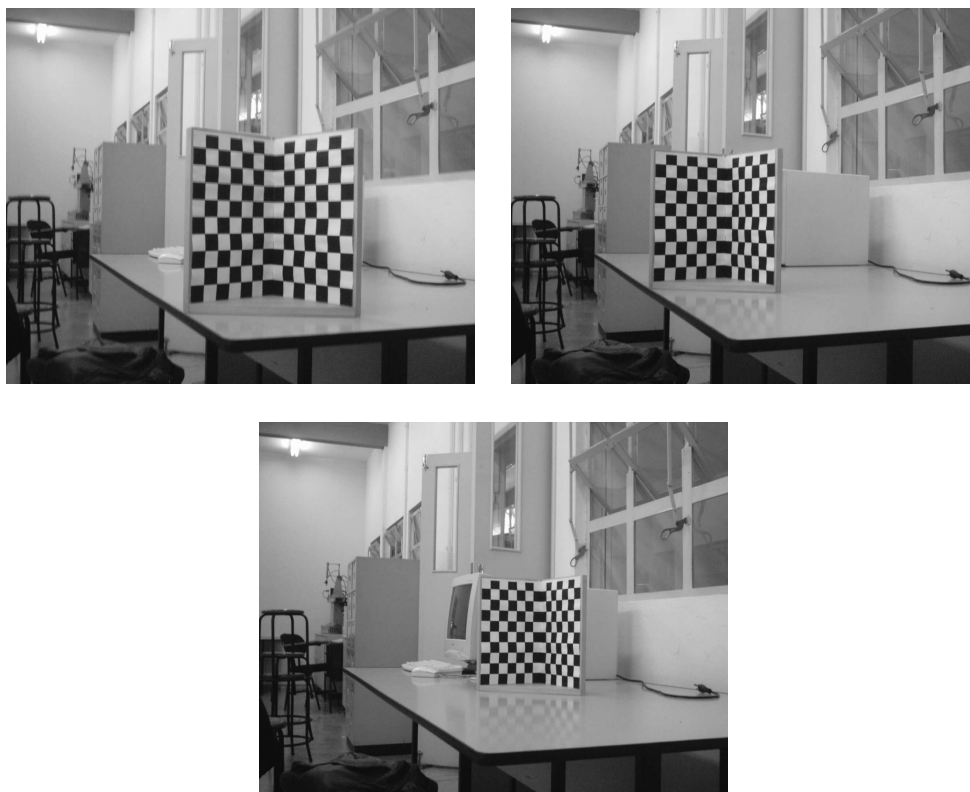


Figura 4.9 – Três exemplares da seqüência de imagens-teste utilizada na avaliação do algoritmo de calibração automática com deslocamentos planares.

Os testes foram realizados com o conjunto binocular fazendo dois tipos de deslocamentos: planar e euclidiano (translação perpendicular a nenhum eixo). Em ambos os casos, os cantos extraídos do gabarito foram utilizados para calcular a matriz fundamental. Dessa forma, tal matriz foi estimada com extrema exatidão. Em seguida, o algoritmo utilizou o conjunto de cantos extraídos dos pares de imagens para efetuar a calibração.

Novamente, a título de comparação, os parâmetros intrínsecos do conjunto binocular foram estimados com auxílio da implementação de Bouguet (2004) do algoritmo de Zhang (2000). Tais parâmetros são apresentados na tabela 4.3 e são referenciados por **ZHANG**.

Para cada caso, os resultados dos testes são apresentados a seguir.

4.2.1 Deslocamento Planar

Com as câmeras realizando um deslocamento planar foi capturada uma seqüência de 20 pares de imagens binoculares. (A figura 4.9 apresenta três das imagens utilizadas nos testes.) Após a identificação semi-automática dos pontos do gabarito, o conjunto de pontos foi processado pelo algoritmo implementado em duas execuções diferentes. Inicialmente, utilizou-se apenas dados provenientes de 10 pares de imagens. Em seguida, todos os dados foram processados de uma vez, em outra execução. Isto foi feito dessa forma para avaliar a influencia do número de deslocamentos na

Tabela 4.3 – Resultados experimentais da calibração do conjunto binocular que realizou deslocamentos planares. Observe que apenas os parâmetros **ZHANG** são valores absolutos. Os outros resultados são dados na forma do erro relativo (em percentagem) aos parâmetros **ZHANG**.

Método	Câmera 1				Câmera 2			
	α	β	u_0	v_0	α	β	u_0	v_0
ZHANG	842,40	879,55	358,04	207,24	832,58	829,54	357,54	206,54
PLANAR	4,90	5,30	5,01	4,45	4,79	4,19	5,45	5,15
EUCLIDEANO	3,08	2,98	3,45	3,30	2,78	3,50	3,27	3,00

Tabela 4.4 – Resultados experimentais da calibração do conjunto binocular que realizou deslocamentos euclidianos. Observe que apenas os parâmetros **ZHANG** são valores absolutos. Os outros resultados são dados na forma do erro relativo (em percentagem) aos parâmetros **ZHANG**.

Método	Câmera 1				Câmera 2			
	α	β	u_0	v_0	α	β	u_0	v_0
ZHANG	842,40	879,55	358,04	207,24	832,58	829,54	357,54	206,54
PLANAR	4,56	3,69	4,76	3,89	3,86	4,58	4,51	4,01
EUCLIDEANO	2,89	3,30	2,74	3,07	3,64	3,27	2,89	3,01

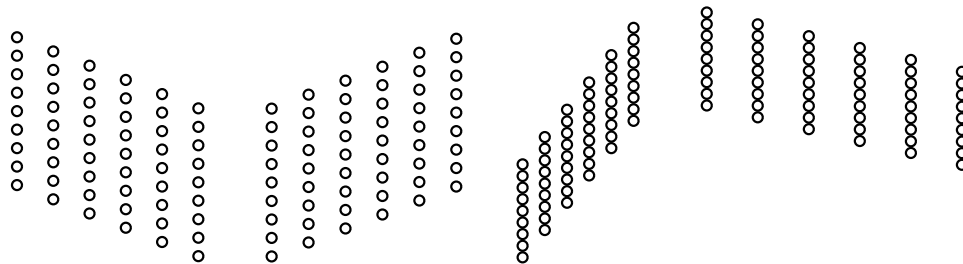


Figura 4.10 – Esboço da reconstrução do gabarito realizada após a calibração automática do conjunto binocular.

exatidão dos resultados. Tais resultados, de ambas as execuções, são referenciados como **PLANAR** e são apresentados nas tabelas 4.3 e 4.4 para, respectivamente, 10 e 20 pares de imagens.

Analisando os resultados observa-se que os parâmetros intrínsecos da câmera 1 são estimados com uma exatidão maior que os da câmera 2. Iste fato já foi comprovado empiricamente por Zisserman et al. (1995) e demonstrado matematicamente por Horaud et al. (2000). Isto se deve ao fato da estimação do plano **W** (seção 2.5.2) ser muito susceptível a ruídos.

Comparando com o resultado obtido pelo método **ZHANG**, o resultado **PLANAR** apresenta um erro da ordem de 5 e 4% para, respectivamente, o resultado obtido com dados de 10 (tabela 4.3) e 20 (tabela 4.4) imagens binoculares. Apesar da baixa exatidão, a estimação permite a recuperação da informação 3D. Por exemplo, na figura 4.10 apresenta-se a reconstrução do gabarito obtida com auxílio das matrizes de projeção estimadas e o algoritmo de Hartley e Sturm (1997).

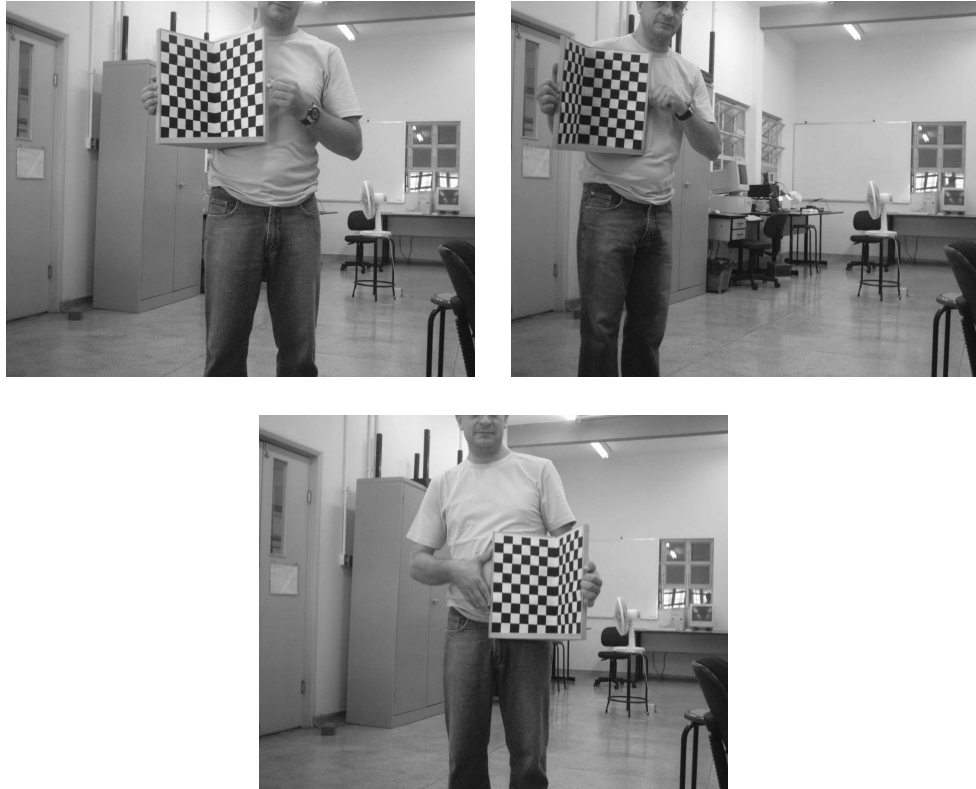


Figura 4.11 – Três exemplares da seqüência de imagens-teste utilizada na avaliação do algoritmo de calibração automática com deslocamentos euclidianos.

4.2.2 Deslocamento Euclidiano

Neste experimento, o conjunto binocular realizou um deslocamento euclidiano completo. Novamente, foi capturada uma seqüência de 20 pares de imagens e a calibração foi realizada em duas execuções diferentes. (Dentre essas imagens, três delas estão apresentadas na figura 4.11.) Na primeira execução utilizou-se dados provenientes de apenas 10 pares de imagens. Na segunda, todos os dados foram utilizados. Os resultados são referenciados por **EUCLIDEANO** e apresentados na tabela 4.3 (10 pares) e na tabela 4.4 (20 pares). Analisando tais resultados observa-se que a exatidão obtida está em torno de 3%. Também, não houve grande diferença na exatidão para o caso de utilizar-se 10 ou 20 pares de imagens.

CONCLUSÕES E TRABALHOS FUTUROS

É possível extrair uma infinidade de informações de uma seqüência de imagens. Em particular, é possível extrair informações como, por exemplo, distâncias, ângulos, velocidade, etc. Em vista disto, a visão computacional já está sendo muito utilizada em algumas aplicações de metrologia e robótica. Contudo, em tais aplicações, é necessário entender como a câmera utilizada forma a imagem. Isto pode ser entendido conhecendo-se os parâmetros intrínsecos e extrínsecos da câmera, através de um processo chamado calibração.

O processo de calibração pode ser dividido em dois tipos: calibração com gabarito e calibração automática. A principal vantagem da calibração com gabarito é a sua exatidão. Por isso, é a mais utilizada para aplicações de metrologia. Contudo, o processo de calibração com gabarito deve ocorrer antes da operação normal da câmera. Isto torna o método bastante inconveniente se a aplicação exigir que os parâmetros da câmera mudem constantemente. Por outro lado, a calibração automática torna a operação de calibração um processo bastante flexível, pois este pode acontecer sem que a inspeção visual seja interrompida. No entanto, trabalhos anteriores revelaram que a calibração automática leva a resultados pouco exatos. Em vista disso, esta ainda é pouco utilizada na prática.

Neste trabalho, avaliou-se em detalhes a técnica proposta recentemente de calibração monocular com auxílio de um gabarito de uma única dimensão. Foi proposta uma normalização nas projeções dos pontos do gabarito que torna o método mais robusto a ruídos. Além disso, realizou-se testes mais detalhados e confiáveis (quando comparado aos disponíveis na literatura). Tais testes avaliaram o desempenho do método quanto ao número de imagens utilizadas, o número de pontos do gabarito e o nível de ruídos presente nos dados. Isto mostrou que o método pode ser realmente utilizado na prática e que a normalização proposta produz de fato um grande aumento na exatidão dos parâmetros estimados.

Um tipo de aplicação que pode usufruir das vantagens da calibração automática é a robótica móvel. Isto porque, normalmente, os robôs utilizados neste tipo de aplicação são ditos autônomos. Evidentemente, tais robôs perdem muito de sua autonomia se o processo de calibração é manual, com a intervenção de um operador. Este fato é agravado ainda mais se os parâmetros da câmera

estiverem sujeitos a alterações no decorrer do percurso do robô. Além do caso óbvio de uma mudança intencional da distância focal da câmera, na robótica móvel, tais alterações também podem ocorrer involuntariamente, por exemplo, por razões mecânicas, devido a trepidação do sistema de visão.

No caso da visão binocular, a calibração é ainda mais atraente, pois nesta configuração, restrições extras tornam linear a solução do problema de calibração. Em vista disso, está sendo construído um sistema de visão binocular que possa ser calibrado automaticamente. Para isto, o conjunto de câmeras deve realizar uma série de deslocamentos enquanto a distância relativa das câmeras permanecerá constante.

Apesar da estrutura mecânica deste sistema de visão não ter sido ainda concluída, o algoritmo de calibração automática já foi implementado. Testes em imagens reais revelaram que a calibração efetuada de forma automática pode ser utilizada em aplicações que não exigem muito exatidão como, por exemplo, em aplicações de detecção e localização de obstáculos para a robótica móvel.

O sistema de calibração automática ainda está incompleto, pois falta implementar um algoritmo de rastreamento automático de pontos. Isto é necessário para que o processo de calibração possa dispensar o gabarito e, realmente, tornar-se um procedimento automático.

Bibliografia

- Armangué, X. e Salvi, J. (2003). Overall view regarding fundamental matrix estimation. *Image and Vision Computing*, 21:205–220.
- Banks, J., Bennamoun, M., e Corke, P. (1999). A hybrid stereo matching algorithm incorporating the rank constraint. Em *7th International Conference on Image Processing And Its Applications*, pp. 33–37, Manchester, United Kingdom.
- Beardsley, P. A., Reid, I. D., Zisserman, A., e Murray, D. W. (1995). Active visual navigation using non-metric structure. Em *International Conference on Computer Vision*, pp. 58–64. [online] <http://citeseer.ist.psu.edu/beardsley95active.html>.
- Benevenuti, F. (2001). Calibração de um conjunto binocular para uso em estereoscopia e robótica móvel. Dissertação de Mestrado, Universidade Federal de Santa Catarina.
- Bernardino, A. J. M. (2004). *Binocular Head Control with Foveal Vision : Methods and Applications*. Tese de Doutorado, UNIVERSIDADE TÉCNICA DE LISBOA, Lisboa, Portugal.
- Bouguet, J. T. (2004). *Camera Calibration Toolbox for Matlab*. MRL - Intel Corp. [online] http://www.vision.caltech.edu/bouguetj/calib_doc/index.html.
- Brown, M., Burschka, D., e Hager, G. (2003). Advances in computational stereo. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(8):993–1008. [online] <http://jobber.cs.jhu.edu/~burschka/>.
- Burschka, D. e Hager, G. (2002). Stereo-Based Obstacle Avoidance in Indoor Environments with Active Sensor Re-Calibration. Em *International Conference on Robotics and Automation*, pp. 2066–2072.
- Csurka, G., Demirdjian, D., e Horaud, R. (1999). Finding the collineation between two projective reconstructions. *Computer Vision and Image Understanding*, 75(3):260–268. [online] <http://citeseer.ist.psu.edu/article/csurka98finding.html>.
- de França, J. A., Stemmer, M. R., e de M. França, M. B. (2005). Algoritmo robusto para correspondência de pontos em imagens estereoscópicas na ausência de calibração. *Semina: Ciências Exatas/Tecnológicas*, 26(2):1–14. <http://research.eeol.org/publications/>.
- de França, J. A. (2005). *Calibração de Um Conjunto de Visão Estéreo: de Correspondências de Pontos à Reconstrução Euclideana*. Tese de Doutorado, Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina, Brasil. [online] <http://research.eeol.org>.
- de França, J. A., França, M. B. M., Hoto, R., e Stemmer, M. R. (2004). A simple linear and iterative

- LMS algorithm for fundamental matrix estimating. Em *I Encontro de Robótica Inteligente (EnRI) in Proceedings of the XXIV Congresso da Sociedade Brasileira de Computação*, Bahia, Brasil. [online] <http://research.eeol.org/publications/index.htm>.
- Dhond, U. R. e Aggarwal, J. K. (1989). Structure from stereo—a review. Em *IEEE Conference on Computer Vision and Pattern Recognition*, volume 19, pp. 1489–1510.
- Faugeras, O. e Luong, Q.-T. (2001). *The Geometry of Multiple Images: The Laws That Govern the Formation of Multiple Images of a Scene and Some of Their Applications*. MIT Press, Cambridge, MA.
- Faugeras, O. D. (1992). What can be seen in three dimensions with an uncalibrated stereo rig? Em *II European Conference on Computer Vision*, pp. 563–578.
- Fischler, M. e R. Bolles (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of ACM*, 24(6):381–385.
- Golub, G. H. e Van Loan, C. F. (1996). *Matrix Computations*. The Johns University Press.
- Habed, A. e Boufama, B. S. (2002). Camera self-calibration from two views. Em *IEEE International Conference on Systems, Man and Cybernetics*, volume 4, página 5 pp.
- Hartley, R. (1997a). In defence of the eight point algorithm. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(6):580–593. [online] <http://rsise.anu.edu.au/~hartley/My-Papers.html>.
- Hartley, R. (1997b). Self-calibration from stationary cameras. *International Journal of Computer Vision*, 22(1):5–23. [online] <http://rsise.anu.edu.au/~hartley/My-Papers.html>.
- Hartley, R., Gupta, R., e Chang, T. (1992). Stereo from uncalibrated cameras. Em *II European Conference on Computer Vision*, pp. 579–587. [online] <http://www.syseng.anu.edu.au/~hartley/My-Papers.html>.
- Hartley, R. e Zisserman, A. (2000). *Multiple View Geometry in Computer Vision*. Cambridge University Press.
- Hartley, R. I. e Sturm, P. (1997). Triangulation. *Computer Vision and Image Understanding*, 68(2):146–157. [online] <http://citeseer.ist.psu.edu/hartley94triangulation.html>.
- Hayman, E. e Murray, D. (2003). The effects of translational misalignment in the self-calibration of rotating and zooming cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8):1015–1020. [online] <http://citeseer.ist.psu.edu/hayman02effects.html>.
- Hespanha, J. P., Dodds, Z., Hager, G. D., e Morse, A. S. (1999). What tasks can be performed with an uncalibrated stereo vision system? *International Journal of Computer Vision*, 35(1):65–85. [online] <http://citeseer.ist.psu.edu/hespanha97what.html>.
- Horau, R., Csurka, G., e Demirdijian, D. (2000). Stereo calibration from rigid motions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1446–1452. [online] <http://citeseer.ist.psu.edu/horau98stereo.html>.
- Izquierdo, E. e Guerra, V. (2003). Estimating the essential matrix by efficient linear techniques. *IEEE Trans. on Circuits and Systems for Video Technology*, 13(9):925–934.
- Lindeberg, T. (1994). *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers.
- Luong, Q.-T. e Viéville, T. (1996). Canonical representations for the geometries of multiple projective

- views. *Computer Vision and Image Understanding*, 64(2):193–229. [online] <http://citeseer.ist.psu.edu/luong94canonic.html>.
- Maybank, S. e Faugeras, O. (1992). A theory of self-calibration of a moving camera. *The International Journal of Computer Vision*, 8(2):123–151.
- P.Aschwanden e W.Guggenbuhl (1993). *Experimental Results from a Comparative Study on Correlation-Type Registration Algorithms*. Forstner and Ruwiedel, Wickmann.
- Ruf, A. e Horaud, R. (2000). Vision-based guidance and control of robots in projective space. Em *European Conference on Computer Vision*, pp. 50–66. [online] <http://citeseer.ist.psu.edu/ruf00visionbased.html>.
- Salvi, J., Armangué, X., e Batlle, J. (2002). A comparative review of camera calibrating methods with accuracy evaluation. *Pattern Recognition*, 35:1617–1635.
- Scharstein, D. e Szeliski, R. (1996). Stereo matching with non-linear diffusion. Relatório Técnico TR96-1575, International Journal of Computer Vision. [online] <http://citeseer.ist.psu.edu/scharstein98stereo.html>.
- Scharstein, D. e Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1):7–42. [online] <http://citeseer.ist.psu.edu/scharstein01taxonomy.html>.
- Shah, J. (1993). A nonlinear diffusion model for discontinuous disparity and half-occlusions in stereo. Em *Computer Vision and Pattern Recognition*, pp. 34–40.
- Shih, S., Hung, Y., e Lin, W. (1996). Accuracy analysis on the estimation of camera parameters for active vision systems. Relatório Técnico TR-IIS-96-006, Institute of Information Science, Taipei, Taiwan. [online] <http://citeseer.ist.psu.edu/shih96accuracy.html>.
- Torr, P. e Murray, D. (1997). The development and comparison of robust methods for estimating the fundamental matrix. *International Journal of Computer Vision*, 24(3):271–300. [online] <http://citeseer.ist.psu.edu/torr97development.html>.
- Torr, P. H. S. (1995). *Outlier Detection and Motion Segmentation*. Tese de Doutorado, University of Oxford, Engineering Dept, Inglaterra. [online] <http://wwwcms.brookes.ac.uk/~philiptorr/papers.htm>.
- Torr, P. H. S. e Zisserman, A. (1998). Robust computation and parametrization of multiple view relations. Em Desai, U., editor, *ICCV6*, pp. 727–732. Narosa Publishing House. [online] <http://wwwcms.brookes.ac.uk/~philiptorr/>.
- Tsai, R. Y. (1987). A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf cameras and lenses. *IEEE Journal Robotics and Automation*, RA-3(4):323–344.
- Zabih, R. e Woodfill, J. (1994). Non-parametric local transforms for computing visual correspondence. Em *3rd European Conference on Computer Vision*, volume 2, pp. 151–158. [online] <http://citeseer.ist.psu.edu/article/zabih94nonparametric.html>.
- Zeller, C. e Faugeras, O. (1994). Applications of non-metric vision to some visual guided tasks. Em *International Conference on Pattern Recognition*, pp. 132–136, Jerusalem, Israel. [online]

- <http://citeseer.ist.psu.edu/zeller94applications.html>.
- Zhang, Z. (1998). Determining the epipolar geometry and its uncertainty: A review. *The International Journal of Computer Vision*, 27(2):161–195. [online] <http://citeseer.ist.psu.edu/zhang98determining.html>.
- Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334. [online] <http://citeseer.ist.psu.edu/zhang98flexible.html>.
- Zhang, Z. (2004). Camera calibration with one-dimensional objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:892–899. [online] <http://citeseer.ist.psu.edu/zhang02camera.htm>.
- Zhang, Z., Faugeras, O., e Deriche, R. (1997). 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*, 1(1):58–68. [online] <http://citeseer.ist.psu.edu/zhang97effective.html>.
- Zhang, Z., Luong, Q. T., e Faugeras, O. (1996). Motion of an uncalibrated stereo rig: Self-calibration and metric reconstruction. *IEEE Transactions on Robotics and Automation*, 12(1):103–113. [online] citeseer.ist.psu.edu/article/zhang93motion.html.
- Zisserman, A., Beardsley, P., e Reid, I. (1995). Metric calibration of a stereo rig. Em *IEEE Workshop on Representations of Visual Scenes*, pp. 93–100, Boston, EUA. [online] <http://citeseer.ist.psu.edu/zisserman95metric.html>.