

JOSÉ ALEXANDRE DE FRANÇA

**CALIBRAÇÃO DE UM SISTEMA DE VISÃO
ESTÉREO: DE CORRESPONDÊNCIAS DE PONTOS
À RECONSTRUÇÃO EUCLIDEANA**

FLORIANÓPOLIS

2005

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
CURSO DE PÓS-GRADUAÇÃO EM ENGENHARIA
ELÉTRICA**

**CALIBRAÇÃO DE UM SISTEMA DE VISÃO
ESTÉREO: DE CORRESPONDÊNCIAS DE PONTOS
À RECONSTRUÇÃO EUCLIDEANA**

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

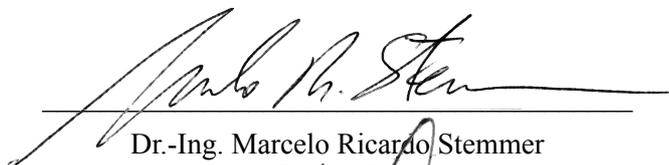
JOSÉ ALEXANDRE DE FRANÇA

Florianópolis, Agosto de 2005.

CALIBRAÇÃO DE UM SISTEMA DE VISÃO ESTÉREO: DE CORRESPONDÊNCIAS DE PONTOS À RECONSTRUÇÃO EUCLIDEANA

José Alexandre de França

'Esta tese foi julgada adequada para a obtenção do título de Doutor em Engenharia Elétrica, Área de Concentração em *Controle, Automação e Informática Industrial*, e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Santa Catarina.'

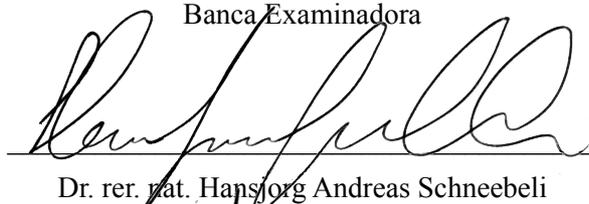


Dr.-Ing. Marcelo Ricardo Stemmer
Orientador

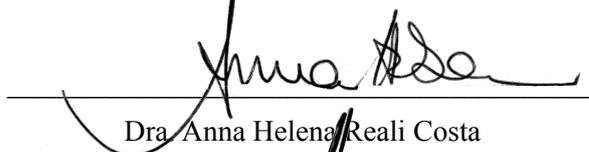


Dr. Alexandre Trofino Neto
Coordenador do Programa de Pós-Graduação em Engenharia Elétrica

Banca Examinadora



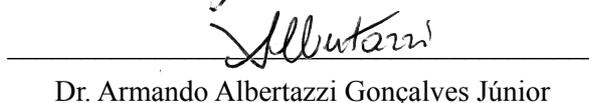
Dr. rer. nat. Hansjörg Andreas Schneebeli



Dra. Anna Helena Reali Costa



Dr. rer. nat. Aldo von Wangenheim



Dr. Armando Albertazzi Gonçalves Júnior

*Ao meu filho Tiago, que
suportou inúmeros pedidos de
silêncio enquanto este trabalho
era desenvolvido.*

AGRADECIMENTOS

À minha esposa por ser minha esposa e co-autora dos meus artigos.

Ao meu orientador pela liberdade e confiança. (Eu trabalho bem melhor assim.)

Aos professores Robinson Hoto (MAT/UEL) e Eduardo Camponogara (DAS/UFSC) pelo apoio técnico.

Ao professor Hans Schneebeli (DEE/UFES) pelo apoio e simpatia.

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

CALIBRAÇÃO DE UM SISTEMA DE VISÃO ESTÉREO: DE CORRESPONDÊNCIAS DE PONTOS À RECONSTRUÇÃO EUCLIDEANA

José Alexandre de França

Agosto/2005

Orientador: Dr.-Ing. Marcelo Ricardo Stemmer

Área de Concentração: Controle, Automação e Informática Industrial

Palavras-chave: Visão Estéreo, Matriz Fundamental, Calibração de Câmeras, Correspondência de Pontos

Número de Páginas: 152

Uma imagem fornece uma representação em duas dimensões de uma cena. Durante o processo de formação desta representação, todas as informações 3D são perdidas. Contudo, tais informações podem ser recuperadas, através da visão estéreo, se duas imagens de uma mesma cena, mas de perspectivas diferentes, estão disponíveis. No entanto, isto só é possível se o sistema estiver calibrado. Por isso, a calibração de câmeras é indispensável em diversas aplicações.

Neste trabalho, realiza-se uma análise crítica das diversas técnicas de calibração existentes na literatura. Além disso, novos métodos de calibração de um sistema de visão binocular são propostos. Basicamente, as técnicas propostas consistem em, inicialmente, recuperar a geometria epipolar e realizar uma calibração projetiva. Em seguida, os pontos de um gabarito de uma única dimensão são utilizados para atualizar a calibração projetiva para a euclidiana. Como é utilizado um gabarito de dimensões reduzidas e de fácil confecção, os métodos de calibração propostos são bastante flexíveis e de fácil execução.

Para auxiliar a tarefa de calibração, novos algoritmos de correspondência de pontos e estimação da matriz fundamental também são propostos. Em especial, o algoritmo de estimação da matriz fundamental utiliza um conjunto de correspondências virtuais que definem um plano também virtual. Evidentemente, tal plano induz uma homografia entre as imagens estéreo. O paralaxe de correspondências que não satisfazem tal homografia permite estimar um dos epipolos do sistema e, portanto, caracterizar a matriz fundamental. A principal vantagem do método proposto é obter uma boa estimação utilizando uma parametrização com apenas cinco incógnitas.

Experimentos em imagens reais e sintéticas validam as técnicas propostas, mostrando como seu desempenho varia quanto a quantidade de ruído e número de correspondências existentes.

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

CALIBRAÇÃO DE UM SISTEMA DE VISÃO ESTÉREO: DE CORRESPONDÊNCIAS DE PONTOS À RECONSTRUÇÃO EUCLIDEANA

José Alexandre de França

August/2005

Advisor: Dr.-Ing. Marcelo Ricardo Stemmer

Area of Concentration: Control, Automation and Industrial Computing

Key words: Stereo Vision, Fundamental Matrix, Camera Calibration, Point Match

Number of Pages: 152

An image could offer a bidimensional representation of a scene. During the image formation process every 3D scene information are lost. However, these informations could be recovered by an stereo vision system. But it is possible only if the system is calibrated. Therefore, the camera calibration process is essential in many applications.

In this work, an analysis of many different calibration techniques which are presenting in technical literature, is done. Behind this, new calibration methods of stereo vision systems are suggest. The methods consists of upgrading a 3D projective representation to Euclidean using a free-moving 1D calibration apparatus (points aligned on a line). For higher accuracy, a nonlinear technique based on the maximum likelihood criterion is used to estimate the stereo rig parameters (the intrinsic and extrinsic parameters).

We propose also new algorithms of stereo match and fundamental matrix estimation that helps the calibration process. The fundamental matrix estimation algorithm uses a set of virtual correspondences that define a virtual plane. Obviously, these plane induces an homography across the stereo images. For points not on the plane, they do not verify the homography. So the parallax error could be used to estimate an epipole and therefore could characterized the fundamental matrix. The manly advantage of proposed method is offer a good fundamental matrix estimation using only five unknowns.

Experiments in real and synthetic images validate the new methods and show that the results of the techniques are comparable to the performance of the most complex methods.

Conteúdo

Lista de Figuras	x
Lista de Tabelas	xiii
Lista de Algoritmos	xiv
Lista de Símbolos e Abreviações	xv
1 Introdução	1
1.1 Organização do trabalho	4
1.2 Notação	5
2 Fundamentação Teórica	7
2.1 Modelo da Câmera	7
2.2 Correspondência de Pontos	10
2.2.1 Caracteres Interessantes	12
2.2.2 Restrições de Correspondência	13
2.2.3 Método de Correspondência Básico	18
2.3 Geometria Epipolar	20
2.3.1 Matriz Fundamental	21
2.3.2 Matriz Essencial	25
2.4 Estimação da Matriz Fundamental	26
2.4.1 Derivação Algébrica	26
2.4.2 Solução com Sete Correspondências	27
2.4.3 Métodos Lineares	28
2.4.4 Estimação Não-Linear	29
2.4.5 Estimação Robusta	34
2.5 Calibração da Câmera	39
2.5.1 Método Clássico	39
2.5.2 Calibração Projetiva	42
2.5.3 Calibração Automática	44
2.6 Calibração Automática de um Conjunto Estéreo	51

2.6.1	Autocalibração por Etapas	51
2.6.2	Estimação de Todos os Parâmetros Euclidianos	57
2.7	Reconstrução	59
2.7.1	Método Linear	60
2.7.2	Método Ótimo	60
3	Contribuições à Calibração de Um Conjunto Binocular	64
3.1	Correspondência de pontos na ausência de calibração	64
3.1.1	Conjunto de candidatos à correspondência	65
3.1.2	Eliminação da Ambigüidade	67
3.1.3	Eliminando as falsas correspondências	69
3.2	Melhorando a Estimação da Matriz Fundamental	69
3.3	Estimação da Matriz Fundamental: Um Novo Método Não-Linear	70
3.3.1	Estimando Epipolos Através de Planos	71
3.3.2	Paralaxe Virtual	71
3.3.3	Paralaxe Virtual: Método Não-Linear	74
3.3.4	Detalhes da Minimização	76
3.3.5	Aperfeiçoamento do Método	77
3.4	Calibração estéreo com um bastão graduado	79
3.4.1	Exequibilidade Matemática	79
3.4.2	Bastão com movimento irrestrito: Solução linear	80
3.4.3	Bastão com movimento irrestrito: Solução não-linear	84
3.4.4	Bastão com uma extremidade fixa	86
3.4.5	Detalhes da minimização não-linear	89
4	Resultados Experimentais	92
4.1	Correspondência de pontos	92
4.2	Estimação da Matriz Fundamental	95
4.2.1	Métodos Avaliados	97
4.2.2	Avaliação do método linear proposto	97
4.2.3	Avaliação da solução linear do paralaxe virtual	100
4.2.4	Avaliação da solução não-linear do paralaxe virtual	100
4.3	Calibração Binocular	105
4.3.1	Testes com imagens sintéticas	106
4.3.2	Testes em imagens reais	108
5	Comentários, conclusões e sugestões de trabalhos	114

A Geometria Projetiva	116
A.1 Razão Cruzada	116
A.2 Intersecção de uma reta com a reta no infinito	116
A.3 Projeção ortogonal de um ponto em uma reta	117
A.4 Homografia induzida por um plano	117
B Ferramentas Matemáticas	120
B.1 Fatoração de Cholesky	120
B.2 Solução dos Quadrados-Mínimos para $\mathbf{Ax} = \mathbf{0}$	121
B.2.1 Solução para posto igual a $(n - 1)$	121
B.2.2 Solução para posto igual a $(n - p)$	122
B.3 Solução dos Quadrados-Mínimos para $\mathbf{Ax} = \mathbf{b}$	122
B.4 Algoritmo Levenberg-Marquardt	122
C Artigos Publicados	124
C.1 Artigos publicados em periódicos	124
C.2 Artigos publicados em congressos	124
Bibliografia	125

Lista de Figuras

1.1 Diagrama de blocos de um sistema de visão computacional típico. A representação da cena feita pelo sistema de visão pode ou não ser utilizada para atuar no ambiente.	2
2.1 Esboço de uma câmera <i>pinhole</i> .	8
2.2 (a) modelo <i>pinhole</i> e seus sistemas de coordenadas, (b) determinação da ordenada do ponto \mathbf{m} em função das coordenadas do ponto \mathbf{M} .	8
2.3 Raios ópticos associados a projeções de um mesmo ponto 3D se interceptam no espaço.	11
2.4 A falta de um correspondente para os pontos \mathbf{m}_{12} e \mathbf{m}_{23} pode fazer com que o algoritmo de correspondência de pontos forme a falsa correspondência ($\mathbf{m}_{12} \leftrightarrow \mathbf{m}_{23}$).	13
2.5 O ponto \mathbf{m}_1 é o correspondente dos pontos \mathbf{m}_{21} , \mathbf{m}_{22} e \mathbf{m}_{23} .	15
2.6 Coerência esquerda-direita usada para eliminar ambigüidade causada por oclusão.	15
2.7 O processo de correspondência de pontos minimiza o somatório das distâncias entre pontos correspondentes.	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 Correspondência de pontos por correlação.	19
2.10 Geometria epipolar.	21
2.11 Família de planos da geometria epipolar.	24
2.12 Gabarito de calibração com sistema de coordenadas “imaginário” esboçado.	41
2.13 Cone absoluto, Ω , no plano no infinito, Π_∞ , e sua projeção, ω , no plano da imagem.	46
2.14 Cone absoluto e a geometria epipolar.	48
2.15 Cone absoluto, Ω , no plano no infinito, Π_∞ , e sua projeção, ω , no plano da imagem.	52
2.16 Esboço de um deslocamento (indexado por j) de um conjunto estéreo.	58
3.1 Duas janelas de correlação 3×3 , cujos tons de cinza são diferentes, mas possuem distância Hamming nula.	66
3.2 (a) influência das oclusões na correlação com apenas uma janela de correlação. (b) correlação com 5 janelas. Aqui, em pelo menos uma janela, a semelhança é conservada.	67
3.3 Paralaxe induzido pelo plano Π .	72
3.4 Três pontos 3D e não co-lineares formam um plano no espaço. As projeções desses pontos também são não co-lineares.	73
3.5 Esquemas de calibração estéreo com uma haste, utilizando: (a) um bastão de comprimento conhecido e (b) um bastão com três pontos, cujas posições relativas são conhecidas.	81

4.1	Imagens utilizadas na avaliação dos métodos de correspondência de pontos: (a) é o par chamado MESA ; (b) é o par PLANTA , e; (c) é referenciado como DESKTOP . . .	93
4.2	Imagens-teste utilizadas na estimação da matriz fundamental: (a) e (b) formam o par GABARITO , que possui 1.373 correspondências marcadas e epipolos próximo ao centro das imagens; (c) e (d) compõem o par COMPUTADOR , com 495 correspondências e epipolos próximo a borda das imagens, e; (e) e (f) compõem o par QUARTO , possuem 403 correspondências e os epipolos tendendo para o infinito. . .	98
4.3	Resíduos obtidos com os métodos 8PONTOS , GOLDEN , CSURKA e GRADIEN-TE : (a) e (b) para o par GABARITO , e; (c) e (d) para o par QUARTO	99
4.4	Número típico de iterações necessárias por estes métodos GOLDEN , CSURKA e GRADIEN-TE	99
4.5	Gráficos com os resíduos obtidos com os métodos 8PONTOS , FRANÇA e CSURKA : (a) e (b) par GABARITO ; (c) e (d) par COMPUTADOR , e; (e) e (f) par QUARTO	101
4.6	Número de iterações requerido pelos métodos FRANÇA e CSURKA para os exper-imentos com os pares (a) GABARITO e (b) QUARTO	102
4.7	Comparação entre o método dos 8 pontos e o método do paralaxe virtual. (a) é o resíduo obtido para as correspondências do par GABARITO e (b) é o resíduo para o par QUARTO	102
4.8	Gráficos (a) e (b) mostram o resíduo obtido com os métodos FRANÇA1 e FRANÇA2 para os pares (a) GABARITO e (b) QUARTO , e; os gráficos do resíduo obtido com os métodos FRANÇA1 e FRANÇA3 são: (c) par GABARITO e (d) par QUARTO	103
4.9	Comparação entre os métodos 8PONTOS , FRANÇA3 , BARTOLI e ZHANG para: (a) e (b) o par GABARITO ; (c) e (d) o par COMPUTADOR , e; (e) e (f) o par QUARTO	104
4.10	Comparação entre os métodos ZHANG e CSURKA para: (a) o par de imagens GABARITO , e; (b) o par QUARTO	105
4.11	Influência do ruído na eficiência do algoritmo de calibração com um bastão com uma extremidade fixa. (a) mostra o resultado do método linear para a câmera 1 e (b) para a câmera 2. Os resultados do algoritmo não-linear são apresentados em (a) para a câmera 1 e em (b) para a câmera 2.	107
4.12	Influência do ruído na eficiência do algoritmo de calibração com um bastão com movimento irrestrito. (a) mostra o resultado do método não-linear para a câmera 1 e (b) para a câmera 2.	107
4.13	Influência do número de pares de imagens na eficiência do algoritmo de calibração com um bastão com uma extremidade fixa. (a) mostra o resultado do método linear para a câmera 1 e (b) para a câmera 2. Os resultados do algoritmo não-linear são apresentados em (a) para a câmera 1 e em (b) para a câmera 2.	109
4.14	Influência do número de pares de imagens na eficiência do algoritmo de calibração com um bastão com movimento irrestrito. (a) mostra o resultado do método não-linear para a câmera 1 e (b) para a câmera 2.	110
4.15	Conjunto binocular utilizado nos experimentos.	110
4.16	Gabarito de calibração utilizados nos testes.	111
4.17	Exemplos da sequência de imagens utilizada na calibração do conjunto binocular com um gabarito linear.	111

4.18	Exemplos da sequência de imagens utilizada na calibração do conjunto binocular com um gabarito planar.	112
A.1	Podemos utilizar o fato que retas paralelas encontram-se no infinito para obtermos projeções de pontos que estão em Π_∞ . As projeções de pontos neste plano estão relacionadas por \mathbf{H}_∞	118
A.2	O correspondente do ponto \mathbf{m}_{11} está sobre o segmento de reta $\overline{\mathbf{m}_\infty \mathbf{e}_2}$	119

Lista de Tabelas

4.1	Quantidade de correspondências encontradas em cada etapa do algoritmo pelos métodos analisados. Na primeira coluna, a quantidade de cantos encontrada em cada imagem do par é apresentada entre parênteses.	95
4.2	Tempo (em segundos) necessário para os métodos analisados concluírem cada um das etapas dos algoritmos.	95
4.3	Parâmetros intrínsecos do conjunto binocular utilizado para gerar os dados da simulação.	106
4.4	Resultados experimentais da calibração do conjunto binocular (θ_{sk} é dado em graus). Observe que apenas os parâmetros REAIS são valores absolutos. Os resultados de todos os outros métodos são dados na forma do erro relativo (em percentagem) aos parâmetros REAIS	113

Lista de Algoritmos

2.1	Método básico de correspondência de pontos.	20
2.2	Algoritmo dos 8 pontos normalizado.	29
2.3	Algoritmo RANSAC para estimação da matriz fundamental.	36
2.4	Algoritmo LMedS para estimação da matriz fundamental.	38
2.5	Calibração linear clássica de câmeras.	40
2.6	Calibração não-linear clássica de câmeras.	41
2.7	Calibração automática baseada nas restrições da matriz essencial.	45
2.8	Método ótimo de reconstrução 3D.	63
3.1	Método linear e iterativo para estimação da matriz fundamental	70
3.2	Novo método não-linear, baseado no paralaxe virtual, para estimação da matriz fundamental.	78
3.3	Calibração de um conjunto binocular com um bastão com movimento irrestrito: Solução linear.	85
3.4	Calibração de um conjunto binocular com um bastão com movimento restrito: Solução linear.	88
3.5	Algoritmo de Newton personalizado para o problema de calibração com uma haste graduada.	91

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} .
ω	Matriz 3×3 que representa a equação do cone absoluto.
ω^*	Matriz 3×3 que representa a equação do dual de ω .
$\mathbb{R}^{n \times m}$	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 $n \times 1$.
Π_{∞}	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{E}	Matriz essencial (3×3) que relaciona duas imagens estéreo.
\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 a câmera que está mais a esquerda em um conjunto estéreo. A “câmera 2” é a que está a direita.

\mathbf{K}	Matriz 3×3 que representa a equação de ω^* . Esse matriz relaciona-se com os parâmetros intrínsecos da câmera por $\mathbf{K} = \mathbf{A}\mathbf{A}^T$. (Veja também o símbolo \mathbf{M}' .)
\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 câmera 2, ou seja, \mathbf{M}' é para a câmera 2 o que \mathbf{M} é para a câmera 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 câmera. 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 $\text{pointhe}_2 = [e'_u, e'_v, e'_t]^T$.
\mathbf{P}	Matriz de projeção (3×4) da câmera 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 câmera 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} .
\mathbf{Z}	Matriz 3×3 igual a $\text{diag}(1, 1, 0)$.

INTRODUÇÃO

A visão é o sentido mais importante dos seres humanos. Sem nenhum contato físico, ela nos fornece uma quantidade extraordinária de informação sobre o ambiente que nos rodeia, permitindo-nos interagir com os mais diferentes objetos. Dessa forma, é natural que muitos esforços tenham sido devotados a dotar algumas máquinas com o sentido de visão. Contudo, a visão é o nosso sentido mais complexo. De fato, ainda hoje não se possui um entendimento completo de como o sistema de visão biológico atua (GRANLUND, 1999). Por isso, infelizmente, não é surpresa que o estado da arte da visão computacional esteja muito distante do “computador que vê”. Os sistemas existentes atualmente são construídos para aplicações específicas, supõem condições especiais de funcionamento (iluminação, tipo de ambiente, faixa de cor etc) e aplicam-se apenas a um conjunto limitado de situações (fornecendo resultados ruins para todas as outras).

Evidentemente, em um sistema de visão computacional, o item mais importante é a câmera, pois é ela que fornece os dados que o sistema deve processar. Este trabalho tem como objetivo: o estudo dos diversos métodos existentes para calibração de câmeras para uso em visão computacional. Em especial, analisam-se técnicas para calibração de sistemas binoculares. São também propostos novos algoritmos para estimação da geometria epipolar e para calibração de câmeras propriamente dita.

Na figura 1.1, apresenta-se o diagrama de blocos de um sistema típico que utiliza visão computacional. Sob condições controladas, um sistema de visão captura imagens de uma câmera e produz uma representação simbólica da cena. Possivelmente, tal representação descreve a cena apenas parcialmente. Contudo, esta deve ser suficiente para que o estágio de processamento obtenha a informação para a qual o sistema foi projetado. Em seguida, se necessário, alguma ação de controle pode ser produzida para interagir com o ambiente.

O diagrama da figura 1.1 é muito parecido com o esboço de um sistema geral de instrumentação eletrônica e controle. Neste caso, os dispositivos sensíveis à luz (ou a outro tipo de radiação), utilizados para formar a imagem, têm a mesma função dos sensores usados na instrumentação. Neste mesmo contexto, a câmera funciona como um transdutor que fornece uma matriz de números relacionada à cena capturada.

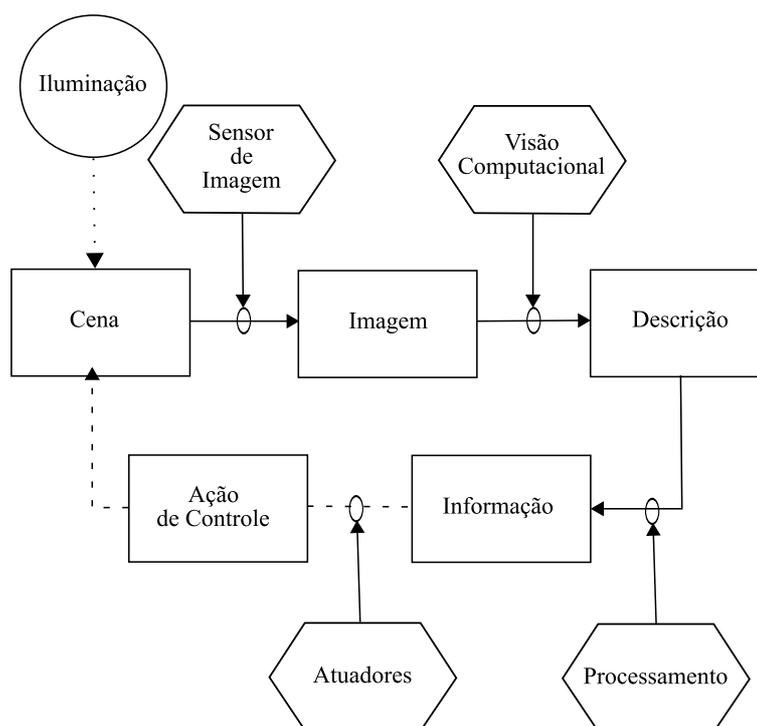


Figura 1.1: Diagrama de blocos de um sistema de visão computacional típico. A representação da cena feita pelo sistema de visão pode ou não ser utilizada para atuar no ambiente.

Assim como na instrumentação eletrônica, onde os transdutores devem ser calibrados antes de sua utilização no sistema de medição, em diversas aplicações da visão computacional, a câmera utilizada pelo sistema deve também estar calibrada. Tal calibração consiste em estimar os parâmetros que melhor ajustam o modelo matemático escolhido à câmera utilizada.

Normalmente, a calibração é realizada antes da operação normal do sistema. Neste caso, insere-se um objeto 3D (chamado gabarito) de forma e tamanho conhecidos no campo visual da câmera. Em seguida, o algoritmo de calibração retira as informações necessárias e estima os parâmetros da câmera. Apesar de extremamente exata, a calibração com gabarito é pouco flexível, pois, além de requerer um objeto de calibração difícil de ser confeccionado, exige que a operação do sistema seja interrompida sempre que uma nova calibração for necessária.

Para tornar o processo de calibração mais flexível, Faugeras (1992) e Hartley et al. (1992) desenvolveram a técnica de calibração projetiva. Esta não necessita de gabarito de calibração. Contudo, o modelo utilizado para a câmera não permite a recuperação de informações importantes da cena como, por exemplo, distâncias, ângulos e paralelismo (FAUGERAS; LUONG, 2001). Dessa forma, a técnica não pode ser utilizada em todas as aplicações.

Uma alternativa mais geral e que serve a maioria das aplicações é a calibração automática. Este conceito foi inicialmente proposto por Maybank e Faugeras (1992) e dispensa o uso de um gabarito de calibração. Ao invés disso, a câmera deve se movimentar com relação a uma cena estática. Dessa forma, a imagem do “cone absoluto”¹ (FAUGERAS; MAYBANK, 1990) fornece restrições que possibili-

¹Veja a definição na página 45.

tam a recuperação de todos os parâmetros do modelo da câmera. Por outro lado, a teoria de calibração automática, apesar de possuir um custo computacional extremamente elevado, leva-nos a resultados pouco exatos. Estes fatos contribuem para que essa técnica ainda não seja largamente utilizada na prática.

No processo de formação das imagens pela câmera, todas as informações 3D são perdidas. Assim, se tais informações forem necessárias, pode ser utilizado um sistema de visão binocular. Neste caso, duas câmeras capturam imagens de uma mesma cena. Além disso, a posição de uma câmera em relação a outra é sempre fixa. Como observado inicialmente por Zisserman et al. (1995), este fato insere mais restrições ao problema de calibração. Estas restrições tornam a calibração automática mais simples. Contudo, trabalhos recentes mostram que tais técnicas são muito sensíveis a ruído mesmo para a visão estereó (HORAUD et al., 2000). Por isso, os resultados obtidos são inexatos, quando comparados aos de métodos de calibração que utilizam gabaritos.

A principal restrição obtida com o uso de sistemas binoculares é a restrição epipolar (HARTLEY; ZISSERMAN, 2000). Esta relaciona pontos na imagem capturada por uma das câmeras a pontos na imagem da outra câmera do mesmo par estereó. Quando o sistema não está calibrado, a restrição epipolar é completamente caracterizada pela matriz fundamental. Tal matriz concentra uma série de informações sobre o sistema estereó e torna todas as operações realizadas menos complexas. Contudo, antes de ser útil, a matriz fundamental deve ser estimada a partir de um conjunto inicial de correspondência de pontos. Para obter tal conjunto, deve-se utilizar um algoritmo que consiga estabelecer correspondências com o sistema não-calibrado. Contudo, tais algoritmos não são muito numerosos na literatura. Além disso, na maioria das vezes, estes são pouco eficientes, pois utilizam técnicas complexas e de custo computacional elevado. Em vista disto, neste trabalho, propõem-se um novo método de correspondência de pontos para sistemas não-calibrados. Objetivou-se o desenvolvimento de um algoritmo rápido e que, se desejado, pudesse ser fácil implementado via hardware, através da linguagem VHDL (ASHENDEN, 2002).

A estimação da matriz fundamental é objeto de estudo há muitas décadas. Até o início da década de 90, os esforços concentravam-se na exatidão da estimação (LUONG; FAUGERAS, 1996; LUONG et al., 1993; HARTLEY, 1993). A um custo computacional elevado, a matriz fundamental era obtida através de parametrizações complexas e métodos não-lineares que minimizam algum critério específico. Algum tempo depois, a partir de meados dos anos 90, foram propostos diversos métodos robustos que identificam e eliminam falsas correspondências² que degradam a qualidade da matriz fundamental obtida (ZHANG, 1998; TORR; ZISSERMAN, 1998; TORR; MURRAY, 1997). Estes métodos tornaram possível estimar a matriz fundamental de forma automática. Atualmente, surge a tendência de tentar reduzir o custo computacional dos algoritmos não-lineares utilizando parametrizações mais simples (BARTOLI; STURM, 2004) ou espaços de busca reduzidos (ZHANG; LOOP, 2001; BARTOLI et al., 2001).

Neste trabalho, um novo método de estimação da matriz fundamental é proposto. Tal método utiliza um conjunto de correspondências virtuais que definem um plano também virtual. Evidentemente, tal plano induz uma homografia entre as imagens estereó. O paralaxe de correspondências de pontos que não satisfazem tal homografia permite estimar um dos epipolos do sistema e, portanto,

²Do inglês: *outliers*.

caracterizar a matriz fundamental. Em oposição à maioria dos trabalhos encontrados na literatura que, ao custo de utilizar algoritmos de implementação confusa e de custo computacional elevado, obtêm uma excelente estimativa da matriz fundamental, optou-se pela simplicidade. Assim, durante o desenvolvimento do método proposto, o pré-requisito principal foi obter uma boa estimativa da matriz fundamental a um baixo custo computacional.

Como proposto por Hartley et al. (1992), uma vez que a matriz fundamental é conhecida, esta permite realizar uma calibração projetiva. Tal calibração pode ser atualizada para uma calibração euclidiana se uma transformação projetiva, que tem características já bem conhecidas (HORAUD et al., 2000), for estimada. Neste trabalho, também se propõe um novo método de calibração de um conjunto binocular que estima tal transformação usando, como gabarito, um bastão graduado. Basicamente, o bastão funciona com um gabarito de uma única dimensão. Por utilizar um objeto de calibração de simples confecção e manuseio, o método proposto acrescenta flexibilidade ao problema. Além disso, os resultados obtidos mostram que o método obtém resultados bem mais exatos, em comparação às técnicas automáticas.

O método proposto é uma extensão do recente trabalho de Zhang (2004), que propôs uma técnica de calibração de uma única câmera com um gabarito 1D. Por sua vez, tal técnica surgiu da tendência atual, que pode ser comprovada examinando artigos recentes como, por exemplo, os trabalhos de Agrawal e Davis (2003), Zhang (2000) e de Sturm e Maybank (1999), de aumentar a flexibilidade da tarefa de calibração reduzindo as dimensões do gabarito.

§1.1 Organização do trabalho

Além desta introdução, este trabalho tem mais quatro capítulos e dois apêndices, que tratam dos assuntos descritos a seguir.

No capítulo dois é apresentado, de maneira sucinta, um estudo sobre visão binocular. Aborda-se todos os assuntos que são pré-requisitos para um bom entendimento sobre o problema da tese, tais como: o modelo de câmera utilizado, geometria epipolar, estimativa da matriz fundamental e as diversas técnicas existentes para calibração de câmeras.

No capítulo três apresentam-se as contribuições do trabalho para a área de visão estéreo. Descrevem-se os métodos propostos para correspondência de pontos e estimativa da matriz fundamental, além de diversos métodos lineares e não-lineares para calibração de um sistema binocular através de um bastão graduado.

Os resultados experimentais são apresentados no capítulo quatro. Os algoritmos propostos são testados em imagens “sintéticas” e imagens reais. Além disso, os resultados são comparados com os de outros métodos bem conhecidos na literatura.

Conclusões e comentários são apresentados no capítulo cinco. Além disso, discutem-se algumas perspectivas de projetos futuros derivados do presente trabalho.

Após o capítulo cinco, foram incluídos dois apêndices. O conteúdo destes é citado constantemente ao longo do texto do trabalho. No apêndice A, apresentam-se alguns conceitos importantes da geometria projetiva. Tal geometria vem sendo muito utilizada nas últimas décadas para modelar diversas características das câmeras. No apêndice B, alguns conceitos matemáticos importantes são apresentados.

Finalmente, o apêndice C contém a lista de artigos publicados pelo autor durante o desenvolvimento do trabalho e que têm relação com este trabalho.

§1.2 Notação

No decorrer do texto, matrizes e vetores são representados por letras, números ou símbolos em negrito. Constantes são expressas por letras, números ou símbolos em itálico. Além disso, uma notação específica para acessar os elementos de uma matriz é adotada. Por exemplo, $\mathbf{A}(i, j)$ representa o elemento da matriz \mathbf{A} na i -ésima linha e j -ésima coluna. Se necessário, o caracter “:” pode ser 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 . Uma outra prática convenção adotada é considerar $A^{-T} = (A^{-1})^T = (A^T)^{-1}$ para toda matriz quadrada e inversível. Finalmente, em algumas casos, o produto cruzado entre dois vetores 3D, \mathbf{v} e \mathbf{x} , é expresso através da matriz antissimétrica, ou seja, $\mathbf{v} \times \mathbf{x} = [\mathbf{v}]_{\times} \mathbf{x}$, com

$$[\mathbf{v}]_{\times} = \begin{bmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{bmatrix},$$

onde considerou-se $\mathbf{v} = [v_1, v_2, v_3]^T$.

As câmaras que pertencem a um conjunto binocular são chamadas de “câmera 1” e “câmera 2”. Normalmente, a câmera 1 é tomada como referência. Além disso, 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 à “câmera 1” e \mathbf{A}' é sempre associada à “câmera 2”.

Considerando o modelo de câmera *pinhole* (FAUGERAS; LUONG, 2001), as coordenadas de um ponto 3D no sistema de coordenadas do ambiente de uma câmera são apresentadas como $\mathbf{M} = [x, y, z]^T$ e a projeção correspondente no plano de imagem I_{α} , $\alpha \in \{1, 2\}$, como $\mathbf{m}_{\alpha} = [u_{\alpha}, v_{\alpha}]^T$. Além disso, as coordenadas homogêneas de um ponto $\mathbf{m}_{\alpha} = [x_{\alpha}, y_{\alpha}, \dots]^T$ são representadas por $\tilde{\mathbf{m}}_{\alpha}$, isto é, $\tilde{\mathbf{m}}_{\alpha} = [x_{\alpha}, y_{\alpha}, \dots, 1]^T$. De uma forma mais geral, um ponto qualquer em coordenadas homogêneas é representado por $\tilde{\mathbf{m}}_{\alpha} = [x_{\alpha}, y_{\alpha}, \dots, t_{\alpha}]^T$. Um segundo índice, se houver, indica a posição do ponto em um conjunto de pontos.

Por último, uma reta \mathbf{l}_{α} , no plano de imagem I_{α} e que passa pelo ponto $\mathbf{m}_{\alpha} = [u_{\alpha}, v_{\alpha}]^T$, deve satisfazer $a_{\alpha}u_{\alpha} + b_{\alpha}v_{\alpha} + c_{\alpha} = 0$. 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 TEÓRICA

Primeiro eles o ignoram. Depois riem de você. Então eles o combatem. Aí você vence.

Mohandas Gandhi

Neste capítulo, apresenta-se um resumo da teoria necessária para a compreensão do problema de calibração de um conjunto binocular. Em especial, descreve-se o modelo matemático de câmera adotado neste trabalho e analisam-se criticamente alguns dos principais métodos de estimação da matriz fundamental e de calibração de câmeras. Os conceitos descritos aqui servem como base a toda discussão dos capítulos posteriores.

§2.1 Modelo da Câmera

O modelo de câmera utilizado neste trabalho é o *pinhole*. Neste, um raio de luz, proveniente de um ponto M no espaço, passa através de um orifício (localizado no ponto C) e incide em um plano no ponto m (veja a figura 2.1 na página seguinte). O plano de incidência dos raios é chamado de “plano de imagem” e o ponto C é chamado de “centro da câmera”. A distância do plano de imagem ao ponto C é um parâmetro intrínseco da câmera e é conhecido como “distância focal” e, geralmente, é representado por f .

Na figura 2.2(a) na próxima página, a câmera é inserida em um sistema de coordenadas. Este é conhecido como “sistema de coordenadas da câmera” e tem sua origem no ponto C . Além disso, um outro sistema de coordenadas, com origem em m_0 , é representado no plano da imagem. Este é chamado “sistema de coordenadas da imagem”. Por sua vez, o ponto m_0 , chamado “centro da imagem”, é a projeção ortogonal do ponto C , ou seja, o ponto no qual o eixo Z corta o plano de imagem.

É importante conhecer a relação entre um ponto M no sistema de coordenadas da câmera e sua projeção, m , no sistema de coordenadas da imagem. Isto pode ser conseguido facilmente através de semelhanças de triângulos. Por exemplo, através da figura 2.2(b) vê-se que a ordenada de m é $u = fY/Z$. De maneira semelhante, pode-se estender o raciocínio para a abscissa de m , logo $m = (fX/Z, fY/Z)$. A equação anterior relaciona as coordenadas dos pontos M e m de forma

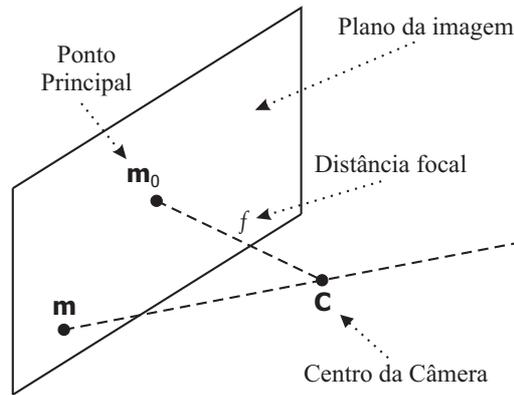
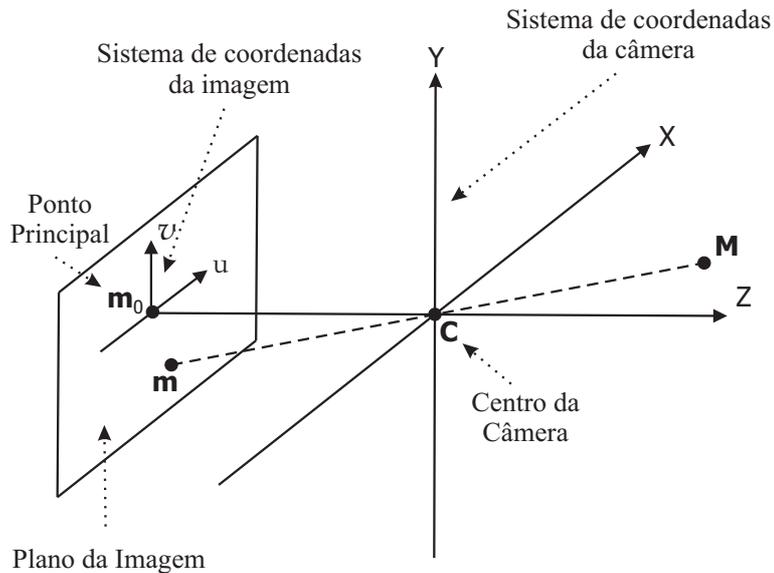
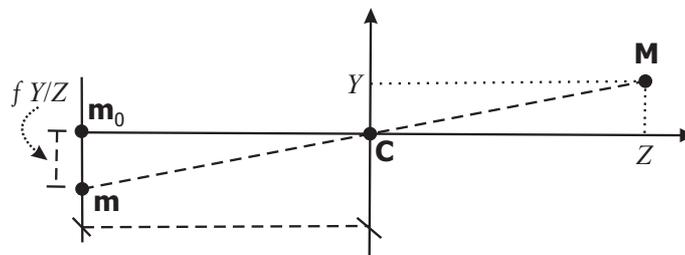


Figura 2.1: Esboço de uma câmera *pinhole*.



(a)



(b)

Figura 2.2: (a) modelo *pinhole* e seus sistemas de coordenadas, (b) determinação da ordenada do ponto m em função das coordenadas do ponto M .

não-linear. Para contornar esse problema, podemos utilizar uma equação homogênea, 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 freqüentemente como

$$\tilde{\mathbf{m}} \simeq \mathbf{A} [\mathbf{I}_3 \mathbf{0}_3] \tilde{\mathbf{M}}, \quad (2.2)$$

onde

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

concentra os “parâmetros intrínsecos da câmera”.

Como na equação (2.2), ao longo do texto, o símbolo “ \simeq ” é utilizado para explicitar que os dois lados de uma igualdade podem diferir por uma constante desconhecida e não-nula, ou seja, trata-se de uma equação homogênea.

A matriz \mathbf{A} dada em (2.3) é um modelo muito simplificado de uma câmera. Na prática, utiliza-se um modelo mais geral definido por

$$\mathbf{A} = \begin{bmatrix} f & f \cot \theta_{sk} & u_0 \\ 0 & f n_{x/y} & v_0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (2.4)$$

O parâmetro $n_{x/y}$ é inserido para levar em consideração que os sensores utilizados pela câmera não são um quadrado perfeito, ou seja, a altura pode diferir da largura. Já o parâmetro $s_k = f \cot \theta_{sk}$ está relacionado ao ângulo, θ_{sk} , que corrige a ortogonalidade da matriz destes sensores. O ponto $\mathbf{m}_0 = [u_0, v_0]^T$ são as coordenadas do ponto central da câmera. Essas coordenadas são inseridas para levar em consideração que, geralmente, a origem do sistema de coordenadas da imagem está no canto superior esquerdo e não no ponto \mathbf{m}_0 .

À medida que a tecnologia evolui, a equação (2.4) aproxima-se cada vez mais de (2.3). Assim, geralmente, $|\theta_{sk} - \pi/2|$ é da ordem de 10^{-6} , ou seja, $s_k \approx 0$. Além disso, $n_{x/y}$ é muito próximo da unidade. Por outro lado, ainda hoje, a estimação do ponto \mathbf{m}_0 é muito sensível a ruídos. Assim, muitos autores assumem que \mathbf{m}_0 localiza-se exatamente no centro da imagem. Felizmente, esta aproximação tem pouca influência na recuperação das informações 3D.

As coordenadas do ponto \mathbf{M} podem ser fornecidas em relação a um referencial diferente do sistema de coordenadas da câmera. Quando isto acontece, o referencial em questão é dito ser o “sistema de coordenadas do ambiente” e a equação (2.2) deve ser reescrita como sendo

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

onde \mathbf{R} é uma matriz 3×3 que representa a rotação da câmera em relação ao sistema de coordenadas do ambiente, \mathbf{t} é um vetor que equivale a translação do centro da câmera com relação ao sistema de coordenadas do ambiente. A matriz $[\mathbf{t}]_{\times} \mathbf{R}$ possui toda informação a respeito dos “parâmetros extrínsecos da câmera”.

Freqüentemente, define-se a “matriz de projeção” de uma câmera como sendo

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

Neste caso, a equação (2.5) torna-se

$$\tilde{\mathbf{m}} \simeq \mathbf{P} \tilde{\mathbf{M}}. \quad (2.7)$$

A seguir, dois resultados importantes relacionados à matriz de projeção da câmera são sumarizados nas proposições a seguir.

Proposição 2.1 *Dada uma câmera, cuja matriz de projeção, \mathbf{P} , é conhecida, o centro da câmera, \mathbf{C} , é dado por $\mathbf{P}\tilde{\mathbf{C}} = \mathbf{0}$.*

Prova: Como \mathbf{P} tem posto três, seu auto-espaço é gerado por um único vetor. Além disso, como $\mathbf{0}$ não pertence ao espaço \mathcal{P}^3 , $\mathbf{0}$ não é solução de $\mathbf{P}\tilde{\mathbf{C}} = \mathbf{0}$. Veja, por exemplo, Hartley e Zisserman (2000, pág. 144) para prova de que $\tilde{\mathbf{C}}$ é a única solução. \square

Proposição 2.2 *Se \mathbf{R} é uma matriz rotacional, $\mathbf{R}\mathbf{R}^T = \mathbf{0}$.*

Prova: Isto segue da definição de matriz rotacional. Veja, por exemplo, (HARTLEY; ZISSERMAN, 2000, pág. 11). \square

§2.2 Correspondência de Pontos

Pelo discutido na seção anterior, dado a projeção, \mathbf{m} , de um ponto 3D, \mathbf{M} , não é possível determinar as coordenadas de \mathbf{M} univocamente. Isto ocorre porque qualquer ponto 3D pertencente a reta $\overline{\mathbf{C}\mathbf{M}}$ (figura 2.1) tem sua projeção em \mathbf{m} .

Agora, considere a figura 2.3. Nesta, o ponto \mathbf{M} é “visto” por duas câmeras. Assim, tal ponto tem uma projeção no plano de imagem de cada câmera, ou seja, $\mathbf{m}_1 \in I_1$ e $\mathbf{m}_2 \in I_2$. Além disso, o raio óptico associado a essas projeções se interceptam no ponto \mathbf{M} . Isto significa que, se o sistema está calibrado, dada uma correspondência $(\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2i})$, é possível determinar as coordenadas do ponto 3D correspondente através de um processo chamado “reconstrução 3D” (discutido na seção 2.7). Contudo, o processo de correspondência de pontos é uma tarefa extremamente complexa. Diversos fatores contribuem para isto. Alguns dos quais são apresentados a seguir.

- ✓ **Oclusões.** Algumas vezes, o correspondente de um ponto na outra imagem pode não estar disponível. Isto ocorre quando o ponto correspondente está fora da parte visível da imagem ou ocluído por um objeto mais próximo a câmera. Oclusão é a maior fonte de erros no processo de correspondência (SCHARSTEIN; SZELISKI, 2002).

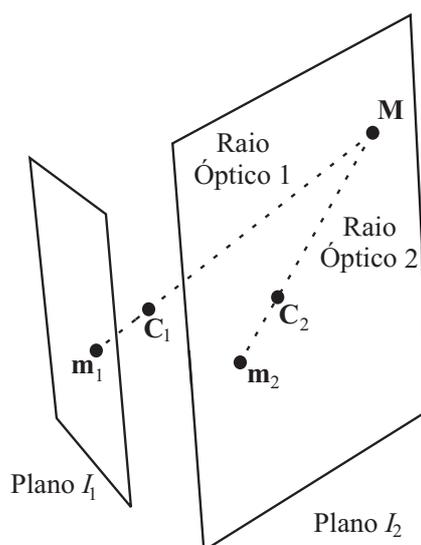


Figura 2.3: Raios ópticos associados a projeções de um mesmo ponto 3D se interceptam no espaço.

- ✓ **Diferença de Iluminação.** Se as imagens estéreo forem capturadas em instantes distintos, elas podem diferir entre si quanto a luminosidade por causa de diferenças no sombreamento da cena.
- ✓ **Distorções Radiométricas.** Se as imagens forem capturadas por câmeras diferentes, mesmo desprezando (ou compensando) diferenças de iluminação, as projeções de um mesmo ponto 3D podem ser representadas por tons de cinza diferentes. Geralmente, pixels que apresentam esse tipo de distorção estão relacionados por um ganho e um desvio.
- ✓ **Distorções Projetivas.** Devido ao seu processo de formação, uma imagem estéreo pode parecer “distorcida” em relação a outra, mesmo que ambas sejam capturadas pela mesma câmera. Essa distorção aumenta à medida que o deslocamento relativo e/ou a diferença na distância focal das câmeras também aumentam.
- ✓ **Texturas “Pobres”.** Mesmo desconsiderando todos os problemas anteriores, o processo de correspondência de pontos ainda é uma tarefa complexa em cenas com pouca textura, pois isto reduz a relação sinal ruído e aumenta a incerteza.

Oclusões e pouca textura dificultam o processo de correspondência por inserir ambigüidade. Já a diferença de iluminação e as distorções, por tornarem as imagens ligeiramente diferentes entre si.

Felizmente, em aplicações como, por exemplo, estimação da geometria epipolar (seção 2.3), calibração de câmeras (seções 2.5 e 2.6) e retificação de imagens (CHEN et al., 2003; LOOP; ZHANG, 1999), é necessário estabelecer apenas algumas poucas correspondências. Neste contexto, quanto à quantidade e densidade de correspondências, os algoritmos de correspondência de pontos podem ser classificados em dois grupos que são formalmente definidos a seguir.

Definição 2.1 (Correspondências Densas) *Como o nome sugere, é um conjunto de correspondências de pontos caracterizado por uma grande quantidade de pontos muito próximos e que ocupam toda (ou quase toda) a área da imagem. Geralmente, tais correspondências são estabelecidas após o sistema estar calibrado.*

Definição 2.2 (Correspondências Esparsas) *É um conjunto de correspondências composto por poucos pontos (em comparação ao número total de pixels das imagens). Idealmente, estes pontos estão espalhados por quase todo o plano da imagem, mas, na prática, também podem estar concentrados apenas em uma região.*

A seguir, discute-se em detalhes o problema de estabelecer um conjunto de correspondências esparsas.

2.2.1 Caracteres Interessantes

Como discutido na seção 2.3 (pág. 20), se a geometria epipolar que relaciona duas imagens estéreo não é conhecida, não existe nenhuma restrição geométrica na localização do correspondente de um ponto. Assim, dado um ponto $\mathbf{m}_1 \in I_1$, o seu correspondente, \mathbf{m}_2 , pode estar em qualquer posição do plano de imagem I_2 , ou seja, o espaço de busca envolvido no processo é muito grande. Para tornar o problema menos complexo, uma abordagem consiste em segmentar alguns “caracteres interessantes” em ambas as imagens e, em seguida, efetuar correspondência de pontos apenas dentre esses caracteres. Evidentemente, essa abordagem pode ser utilizada apenas para estabelecer correspondências esparsas.

A definição formal de caracteres interessantes utilizada neste trabalho é dada a seguir.

Definição 2.3 (Caracteres Interessantes) *São características que estão presentes em (quase) todas as cenas, são fáceis de segmentar e cujos atributos são pouco afetados por distorções projetivas, deslocamentos Euclidianos e diferença de iluminação. Exemplos de caracteres interessantes são: cantos, vértices, retas, planos etc.*

Para que a correspondência de caracteres possa ser realizada, eles devem ser extraídos das imagens. Essa é uma das etapas mais importantes do processo e é discutida a seguir.

§ Segmentação de Caracteres

Diversos trabalhos encontrados na literatura utilizam cantos¹ da imagem como caracteres interessantes. O principal motivo é que cantos são (quase) invariantes a deslocamentos Euclidianos, enquanto que, por exemplo, planos e retas podem até desaparecer, dependendo do deslocamento. O algoritmo de segmentação de cantos mais popular é o proposto por Harris e Stephens (1988).

Supondo um canto na imagem I_1 , na posição $\mathbf{m}_1 = [u_1, v_1]^T$, e o seu correspondente, $\mathbf{m}_2 = [u_2, v_2]^T$, na imagem I_2 , o algoritmo de segmentação fornece uma aproximação das coordenadas desses pontos, por exemplo, $\hat{\mathbf{m}}_1 = [u_1 + \delta_{x_1}, v_1 + \delta_{y_1}]^T$ e $\hat{\mathbf{m}}_2 = [u_2 + \delta_{x_2}, v_2 + \delta_{y_2}]^T$. Evidentemente, os deslocamentos δ_{x_1} , δ_{y_1} , δ_{x_2} e δ_{y_2} acrescentam incerteza ao processo. Essa incerteza, normalmente, é modelada na forma de um ruído Gaussiano de média zero. Tais deslocamentos devem ser tão pequenos quanto possível. De fato, eles devem ser de menos de um pixel. Isto pode ser conseguido utilizando detectores de cantos com precisão de sub-pixels.

¹De uma forma geral, cantos são as posições dos picos de variação local da imagem.

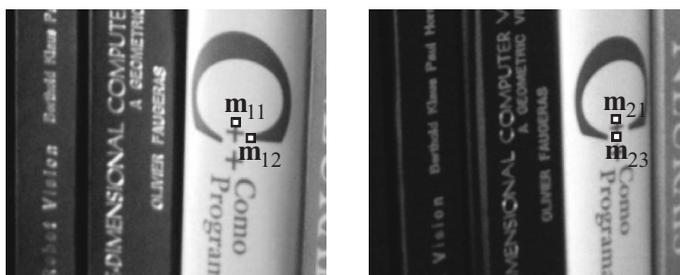


Figura 2.4: A falta de um correspondente para os pontos m_{12} e m_{23} pode fazer com que o algoritmo de correspondência de pontos forme a falsa correspondência ($m_{12} \leftrightarrow m_{23}$).

Ziou e Tabbone (1998) realizaram uma excelente análise crítica dos principais algoritmos de segmentação de cantos com precisão sub-pixel. Contudo, dois dos mais populares são o Canny (CANNY, 1986) e o SUSAN (SMITH, 1992). Implementações desses algoritmos estão disponíveis livremente na internet².

Um dos maiores problemas dos detectores de cantos é ilustrado na figura 2.4. Nesta, dois cantos estão segmentados em cada uma das imagens estéreo. Está evidente que os pontos m_{11} e m_{21} formam uma correspondência. Contudo, o algoritmo não segmentou o correspondente do ponto m_{12} e nem o correspondente de m_{23} . Isso gera ambigüidade e causa problemas semelhantes aos causados por oclusões, ou seja, pode fazer com que o algoritmo de correspondência de pontos, erroneamente, forme a falsa correspondência ($m_{12} \leftrightarrow m_{23}$).

A seguir, algumas restrições utilizadas para reduzir a incerteza inerente ao processo de correspondência de pontos são discutidas.

2.2.2 Restrições de Correspondência

Uma vez que os pontos interessantes tenham sido extraídos, um conjunto de características pode ser associado a eles. Normalmente, o conjunto de características de um ponto interessante é chamado de assinatura. Assim, o processo de correspondência de pontos consiste em, dado um ponto com uma determinada assinatura, restringir o seu correspondente ao ponto na outra imagem que possui uma assinatura semelhante.

A seguir, descreve-se algumas restrições que podem ser utilizadas no processo de correspondência de pontos.

§ Semelhança

Essa restrição é baseada no valor de intensidade da imagem (tons de cinza) na posição do ponto de interesse e obriga o seu correspondente ser similar a ele.

²Canny → <http://vxl.sourceforge.net/>, e; SUSAN → <http://www.fmrib.ox.ac.uk/~steve/susan/susan/susan.html>.

Considerando dois pontos, $\mathbf{m}_{1i} \in I_1$ e $\mathbf{m}_{2j} \in I_2$, a forma mais simples de expressar a semelhança entre eles é através do “somatório das diferenças absolutas” (SAD), dado pela seguinte equação

$$\text{SAD}(c_{ij}) = \sum_{(u,v) \in W} |I_1(u, v) - I_2(u + \delta_x, v + \delta_y)|, \quad (2.8)$$

onde $c_{ij} = (\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2j})$, $I_z(k, l)$ é o valor da função de intensidade da imagem I_z na posição (k, l) , W é uma janela de correlação centrada em \mathbf{m}_{1i} , δ_x e δ_y representam o deslocamento de \mathbf{m}_{2j} em relação a \mathbf{m}_{1i} , e $(u, v) \in W$ representa todos os pontos dentro de W .

A equação (2.8) é baseada na suposição de que os tons de cinza em torno da projeção de um mesmo ponto 3D são os mesmos nas duas imagens. Ela é utilizada em sistema com restrições de tempo por ter baixo custo computacional (BANKS et al., 1999; HIRSCHMÜLLER, 2001). Contudo, é muito sensível a distorções e diferenças de iluminação. Por isso, geralmente, a “correlação cruzada, normalizada e de média zero” é uma medida que fornece melhores resultados e é definida por

$$\text{ZNCC}(c_{ij}) = \frac{\sum_{(u,v) \in W} \delta_1(u, v) \delta_2(u + \delta_x, v + \delta_y)}{\sqrt{\sigma_1^2(u, v) \sigma_2^2(u + \delta_x, v + \delta_y)}}, \quad (2.9)$$

onde $\delta_z(k, l) = I_z(k, l) - \bar{I}_z$ e

$$\sigma_z^2(k, l) = \sum_{(u,v) \in W} [\delta_z(k, l)]^2.$$

Apesar de ser mais robusta, a equação (2.9) aumenta o custo computacional do algoritmo de correspondência.

Inúmeras outras medidas de semelhança podem ser encontradas na literatura. Recentemente, algumas delas foram analisadas e comparadas criticamente por Banks e Corke (2001). Contudo, nenhuma dessas medidas consegue eliminar totalmente a influência da variação dos tons de cinza entre as imagens.

§ Unicidade

Na figura 2.5 está ilustrado um caso típico, no qual um único ponto, $\mathbf{m}_1 \in I_1$, é o correspondente de três pontos, \mathbf{m}_{21} , \mathbf{m}_{22} e \mathbf{m}_{23} , na imagem I_2 . Apesar de casos como este serem possíveis, geralmente, assume-se que todo ponto só pode ter um único correspondente.

Como discutido, interferências como, por exemplo, oclusões e distorções, inserem ambigüidade ao processo de correspondência de pontos. Essa ambigüidade pode fazer com que um ponto tenha dois ou mais candidatos igualmente plausíveis de serem o seu correspondente. Quando isso ocorre, os critérios usados na análise dos candidatos ambíguos devem ser alterados na tentativa de reduzir a incerteza. Outra alternativa é, simplesmente, descartar todos os candidatos ambíguos e considerar o ponto em questão sem correspondente.

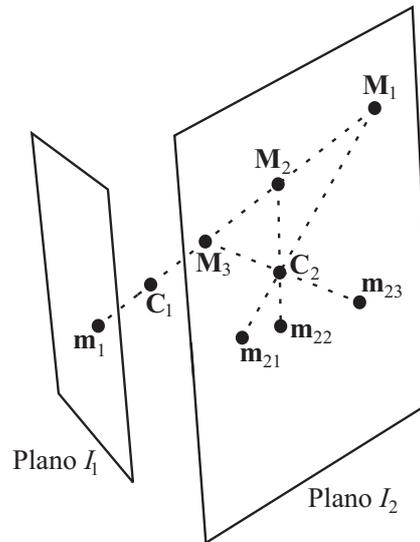


Figura 2.5: O ponto m_1 é o correspondente dos pontos m_{21} , m_{22} e m_{23} .

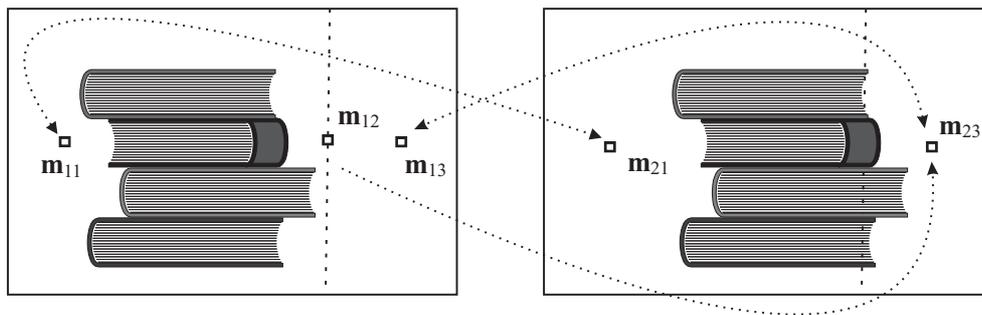


Figura 2.6: Coerência esquerda-direita usada para eliminar ambigüidade causada por oclusão.

§ Invariância Esquerda-Direita

Nos algoritmos de correspondência que utilizam uma abordagem local, uma imagem é tomada como referência. Então, dado um ponto na imagem-referência, o seu correspondente é procurado na outra imagem. Neste caso, pode ser desejável que o algoritmo de correspondência seja invariante à escolha da imagem de referência. Esta restrição também pode ser utilizada para eliminar ambigüidade. Este fato é ilustrado na figura 2.6. Nesta figura, tomando a imagem da esquerda como referência, os pontos m_{12} e m_{13} formam candidatos a correspondência com o mesmo ponto, m_{23} . Contudo, tomando a imagem da direita como referência, o ponto m_{23} não forma um candidato com m_{12} . Isso dá mais suporte ao candidato $(m_{13} \leftrightarrow m_{23})$ e enfraquece o $(m_{12} \leftrightarrow m_{23})$. Assim, $(m_{13} \leftrightarrow m_{23})$ é considerada a correspondência verdadeira.

Essa técnica constitui uma forma simples de, em alguns casos, eliminar a ambigüidade. Contudo, tomar as duas imagens como referência tem a desvantagem evidente de elevar o custo computacional do algoritmo. Assim, uma abordagem mais eficiente é procurar o correspondente de um ponto utilizando um critério que seja invariante a escolha da imagem de referência.

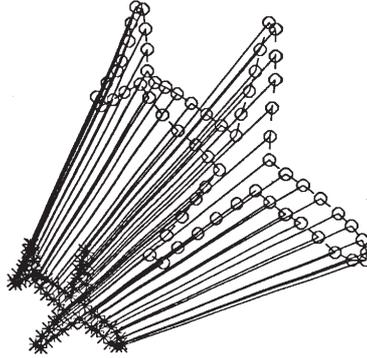


Figura 2.7: O processo de correspondência de pontos minimiza o somatório das distâncias entre pontos correspondentes.

§ Rigidez Global

Essa restrição é derivada do trabalho de Ullman (1984) e, basicamente, estabelece que quanto menor a distância $d_{ij}(\mathbf{m}_{1i}, \mathbf{m}_{2j})$, maior é a confiança da correspondência ($\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2j}$). Contudo, isto não significa, simplesmente, estabelecer correspondência com o ponto mais próximo. (De fato, uma abordagem desse tipo geraria muita ambigüidade.) Ao invés disso, o princípio da proximidade deve ser aplicado ao conjunto de possíveis correspondências como um todo. Isto é ilustrado na figura 2.7. Neste esboço, um segmento de reta liga cada ponto ao seu correspondente. Neste caso, dado n correspondências, a restrição de rigidez global garante o mínimo da seguinte equação

$$\mathcal{J} = \sum_{i=1}^n \sum_{j=1}^n c_{ij} d_{ij}^2(\mathbf{m}_{1i}, \mathbf{m}_{2j}),$$

onde $c_{ij} = 1$, se \mathbf{m}_{1i} e \mathbf{m}_{2j} formam uma correspondência, ou $c_{ij} = 0$, caso contrário.

Essa restrição foi usada com sucesso, por exemplo, por Pilu (1997) e por Grzywacz e Yuille (1988). Contudo, Scott e Longuet-Higgins (1991) demonstraram que essa restrição é eficiente apenas quando a rotação de uma câmera em relação a outra é “pequena”.

§ Disparidade Limitada e de Variação Suave

Considere uma correspondência ($\mathbf{m}_1 \leftrightarrow \mathbf{m}_2$), onde $\mathbf{m}_1 = [u_1, v_1]^T$ e $\mathbf{m}_2 = [u_2, v_2]^T$. A diferença $\Delta x = u_1 - u_2$ é chamada disparidade horizontal e $\Delta y = v_1 - v_2$ é a disparidade vertical.

A disparidade vertical pode ser compensada através de um processo chamado retificação de imagens (CHEN et al., 2003; LOOP; ZHANG, 1999). Por isso, a disparidade horizontal é um parâmetro mais importante. De fato, muitas vezes ela é referenciada, simplesmente, por “disparidade”.

Considerando fixo o deslocamento relativo das câmeras, quanto mais próximo um ponto 3D estiver das câmeras, maior será a disparidade de suas projeções. De fato, a relação é dada por (TRUCCO;

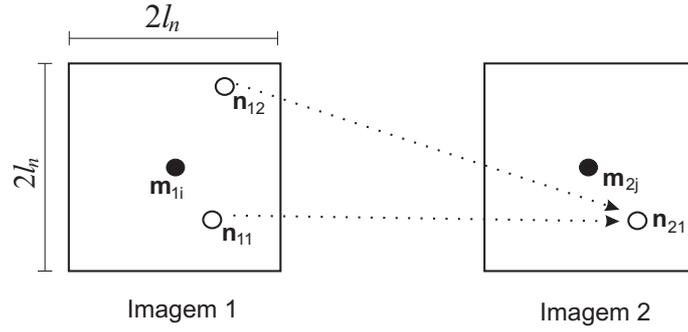


Figura 2.8: Um par (m_{1i}, m_{2j}) candidato à correspondência com vizinhos dentro de uma janela de lado $2l_n$. O par (n_{11}, n_{21}) aumenta a confiança do par (m_{1i}, m_{2j}) .

VERRI, 1998, Sec. 7.1.2)

$$Dz_i = f \frac{|d(\mathbf{C}_1, \mathbf{C}_2)|}{\Delta x_i}, \quad (2.10)$$

onde Dz_i é a distância perpendicular do ponto 3D \mathbf{M}_i ao plano que contém as imagens retificadas. Normalmente, Dz_i é chamado de profundidade.

Podemos considerar que a disparidade em uma cena seja limitada ao intervalo $\Delta_{\min} \leq \Delta x \leq \Delta_{\max}$. Isto é equivalente a supor que não existem objetos muito próximos da câmera e nem muito distantes também. O limite Δ_{\min} é válido, sobretudo, para ambientes fechados.

Outra suposição implícita à maioria dos algoritmos de correspondência é que a disparidade varia suavemente pela imagem. Isto é válido porque a maioria dos objetos tem uma superfície suave. Evidentemente, essa suposição falha na intersecção de regiões que são projeções de objetos que estão a profundidades diferentes, pois tais intersecções são caracterizadas por grandes saltos na disparidade da cena. Contudo, são esses “saltos” que tornam possível a segmentação dos objetos.

§ Vizinhaça

Considerando o par $c_{ij} = (\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2j})$ um candidato à correspondência, podemos representar 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 a correspondência $V_{kl} = (\mathbf{n}_{1k} \leftrightarrow \mathbf{n}_{2l})$, onde $\mathbf{n}_{1k} \in N(\mathbf{m}_{1i})$ e $\mathbf{n}_{2l} \in N(\mathbf{m}_{2j})$, tal que “uma determinada relação” entre \mathbf{n}_{1k} e \mathbf{m}_{1i} é semelhante a essa mesma relação 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.

Essa “determinada relação” pode ser algo simples como a correlação entre os pontos \mathbf{n}_{1k} e \mathbf{m}_{1i} e entre \mathbf{n}_{2l} e \mathbf{m}_{2j} . Contudo, Zhang et al. (1995) utilizaram a “posição relativa entre pontos”, através da seguinte equação

$$\mathcal{F}(\mathbf{m}_{1i}, \mathbf{m}_{2j}) = \sum_{\mathbf{n}_{1k} \in N(\mathbf{m}_{1i})} \left[\max_{\mathbf{n}_{2l} \in N(\mathbf{m}_{2j})} \left(\frac{\alpha_{kl} \delta(\mathbf{m}_{1i}, \mathbf{m}_{2j}; \mathbf{n}_{1k}, \mathbf{n}_{2l})}{1 + D(\mathbf{m}_{1i}, \mathbf{m}_{2j}; \mathbf{n}_{1k}, \mathbf{n}_{2l})} \right) \right], \quad (2.11)$$

onde α_{kl} são constantes que dependem do grau de confiança das correspondências ($\mathbf{n}_{1k} \leftrightarrow \mathbf{n}_{2l}$), e;

$$D(\mathbf{m}_{1i}, \mathbf{m}_{2j}; \mathbf{n}_{1k}, \mathbf{n}_{2l}) = \frac{d(\mathbf{m}_{1i}, \mathbf{m}_{2j}) + d(\mathbf{n}_{1k}, \mathbf{n}_{2l})}{2},$$

com

$$\delta(\mathbf{m}_{1i}, \mathbf{m}_{2j}; \mathbf{n}_{1k}, \mathbf{n}_{2l}) = \begin{cases} e^{-r/\varepsilon_r}, & \text{se é um candidato à correspondência e } r < \varepsilon_r \\ 0, & \text{caso contrário.} \end{cases},$$

onde

$$r = \frac{|d(\mathbf{m}_{1i}, \mathbf{n}_{1k}) - d(\mathbf{m}_{2j}, \mathbf{n}_{2l})|}{D(\mathbf{m}_{1i}, \mathbf{m}_{2j}; \mathbf{n}_{1k}, \mathbf{n}_{2l})}$$

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.

É importante observar que, caso um ponto \mathbf{m}_{1i} tenha vários candidatos à correspondência, o operador “máx” da equação (2.11) garante que apenas o que tem a posição relativa mais semelhante contribui para aumentar a confiança do par ($\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2j}$).

A equação (2.11) é bastante robusta. Contudo, o custo computacional envolvido na formação dos conjuntos de vizinhos, $N(\mathbf{m}_{1i})$ e $N(\mathbf{m}_{2j})$, e na avaliação de (2.11) é muito elevado.

2.2.3 Método de Correspondência Básico

Nesta seção, discute-se como as restrições apresentadas na seção anterior podem ser utilizadas para criar uma técnica que estabelece um conjunto de correspondências esparsas. Para isto, um algoritmo básico de correspondência que usa a abordagem local e tenta associar cantos em uma imagem com cantos na outra imagem é apresentado.

Inicialmente, um algoritmo de detecção de cantos deve ser utilizado para segmentar n_1 cantos na imagem I_1 e n_2 cantos em I_2 . Em seguida, a restrição de semelhança, através da equação (2.9), é utilizada para estabelecer um conjunto inicial de candidatos a correspondência. A tática mais básica é comparar todos os cantos de I_1 com todos os cantos de I_2 . Contudo, desde que o custo computacional de (2.9) é muito elevado, usaremos a restrição de disparidade para limitar o espaço de busca. Assim, como ilustrado na figura 2.9, dado um ponto $\mathbf{m}_{1i} = [u_{1i}, v_{1i}]^T \in I_1$, a restrição de semelhança é aplicada apenas aos cantos da imagem I_2 que pertencem a uma janela de busca quadrada, de lado Δx_{\min} e com centro em (u_{1i}, v_{1i}) .

O resultado da equação (2.9) varia entre -1 (sem nenhuma semelhança) e 1 (muito semelhante). Normalmente, considera-se um candidato à correspondência apenas os pares que obtenham um grau de semelhança maior que um limiar S_{th} . Um valor muito utilizado é $S_{th} = 0,8$ (ZHANG et al., 1995).

Após essa fase inicial de formação de um conjunto de candidatos à correspondência, é comum existir muita ambigüidade entre esses candidatos. Evidentemente, essa ambigüidade deve ser eliminada. Para isto, podemos utilizar a equação (2.11) para medir a confiabilidade dos candidatos e

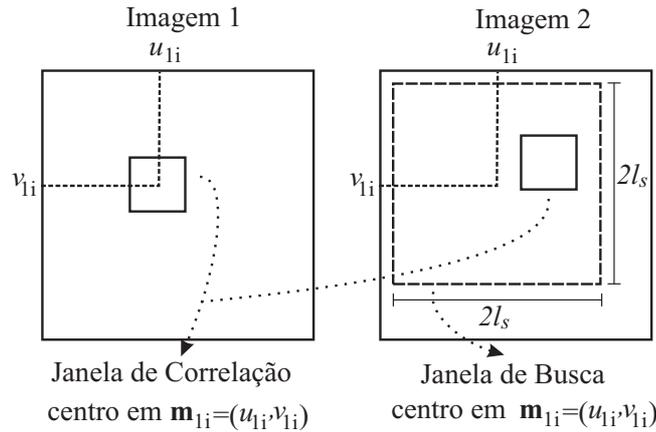


Figura 2.9: Correspondência de pontos por correlação.

eliminar os pouco confiáveis minimizando a equação a seguir

$$\mathcal{J} = \sum_{i=1}^n \sum_{j=1}^{n_i} \mathcal{F}(\mathbf{m}_{1i}, \mathbf{m}_{2j}),$$

onde n é o número de pontos \mathbf{m}_{1i} que possuem n_i pontos, \mathbf{m}_{2j} , candidatos à correspondência.

Geralmente, melhores resultados são obtidos se essa minimização for feita de forma iterativa. Isso pode ser realizado através de alguma técnica de relaxação. Por exemplo, Pollard et al. (1985) utilizaram a abordagem “o-vencedor-ganha-tudo”, ou seja, dado um ponto \mathbf{m}_{1i} com diversos pares ($\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2j}$) associados, considera-se o par vencedor aquele que obtiver a maior confiança $\mathcal{F}(\mathbf{m}_{1i}, \mathbf{m}_{2j})$. Nessa abordagem, a eliminação dos candidatos pouco confiáveis é realizada em um único passo. Contudo, essa técnica não produz bons resultados, pois as correspondências são tratadas de forma independente.

Uma abordagem que produz melhores resultados, a um custo computacional maior, é “o-perdedor-não-ganha-nada”. Nesta, a eliminação dos candidatos “pouco confiáveis³” é realizada em várias iterações. Em cada uma delas, dado um ponto \mathbf{m}_{1i} , apenas o par associado, ($\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2j}$), de menor suporte é eliminado. Em seguida, a equação (2.11) é aplicada novamente aos pares remanescentes. Então, novamente os pares menos confiáveis são eliminados até que não haja mais ambigüidade.

Um resumo da técnica discutida é apresentado no algoritmo 2.1. Contudo, o algoritmo tem muitas falhas. Uma delas é não satisfazer a restrição de invariância esquerda-direta, pois o sistema de correspondência por semelhança que usa uma janela de busca não é simétrico. Evidentemente, isso favorece a ambigüidade.

³Com uma alta probabilidade de ser uma falsa correspondência.

Entrada: n_1 cantos da imagem I_1 e n_2 de I_2 .

Saída: Um conjunto de correspondências ($\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2j}$).

- Para cada ponto $\mathbf{m}_{1i} = [u_{1i}, v_{1i}]^T \in I_1$, calcular a semelhança aplicando a equação (2.9) a todos os cantos de I_2 que pertencem a uma janela de busca quadrada, de lado Δx_{\min} e com centro em (u_{1i}, v_{1i}) .
- Rejeite as correspondências que possuem grau de semelhança menor que S_{th} .
- Usando uma técnica de relaxação, minimize a função a seguir

$$\mathcal{J} = \sum_{i=1}^n \sum_{j=1}^{n_i} \mathcal{F}(\mathbf{m}_{1i}, \mathbf{m}_{2j}).$$

Algoritmo 2.1: Método básico de correspondência de pontos.

§2.3 Geometria Epipolar

No caso de um sistema estéreo, se os parâmetros intrínsecos e extrínsecos das câmeras são conhecidos, o sistema é dito calibrado e, estabelecendo-se correspondências de pontos e utilizando triangulação, a informação 3D pode ser completamente recuperada. Contudo, em um sistema não-calibrado, a única informação disponível é a geometria epipolar. Esta geometria já foi descrita em numerosos artigos, por exemplo, em Torr e Murray (1997), Zhang (1998), Armangué e Salvi (2003). Contudo, basicamente, a geometria epipolar pode ser entendida se for considerado o caso de duas câmeras como apresentado na figura 2.10. Nesta, \mathbf{C}_1 e \mathbf{C}_2 são, respectivamente, os centros ópticos da primeira e segunda câmeras. Então, dado um ponto \mathbf{m}_1 na primeira imagem, I_1 , o ponto correspondente, \mathbf{m}_2 , na segunda imagem, I_2 , está restrito a uma reta chamada “reta epipolar” de \mathbf{m}_1 , representada na figura por l_2 . A reta l_2 é a intersecção do plano Π , definido por \mathbf{M} , \mathbf{C}_1 e \mathbf{C}_2 (chamado de plano epipolar), com o plano I_2 . Isto acontece porque o ponto \mathbf{m}_1 pode corresponder a qualquer ponto da reta $\overline{\mathbf{C}_1\mathbf{M}}$ e a projeção de $\overline{\mathbf{C}_1\mathbf{M}}$ em I_2 é a reta l_2 . Além disso, observa-se que todas as retas epipolares dos pontos de I_1 passam através de um ponto comum, \mathbf{e}_2 , em I_2 . Este ponto é conhecido como “epipolo”. O ponto \mathbf{e}_1 é a intersecção da reta $\overline{\mathbf{C}_1\mathbf{C}_2}$ com o plano I_2 . Isto pode ser entendido como segue. Para cada ponto \mathbf{m}_{1i} na imagem I_1 , a sua reta epipolar, l_{2i} , no plano I_2 , é a intersecção do plano Π_i , definido por \mathbf{C}_1 , \mathbf{C}_2 e o ponto 3D \mathbf{M}_i (cuja projeção é \mathbf{m}_{1i}), com o plano I_2 . Dessa forma, todos os planos epipolares Π_i formam um conjunto de planos contendo a reta $\overline{\mathbf{C}_1\mathbf{C}_2}$. Logo, estes planos precisam intersectar I_2 em um ponto comum, ou seja, no epipolo \mathbf{e}_2 . Finalmente, pode-se facilmente observar a simetria da geometria epipolar. O correspondente em I_1 de cada ponto \mathbf{m}_{2i} , sobre a reta l_{2i} , precisa pertencer à reta epipolar l_{1i} , que é a intersecção do mesmo plano Π_i com o plano I_1 . Todas as retas epipolares formam um conjunto contendo o epipolo \mathbf{e}_1 , que é a intersecção da reta $\overline{\mathbf{C}_1\mathbf{C}_2}$ com o plano I_1 . A simetria também leva à seguinte observação: se \mathbf{m}_1 (um ponto em I_1) e \mathbf{m}_2 (um ponto em I_2) correspondem a um único ponto \mathbf{M} no espaço, então \mathbf{m}_1 e \mathbf{m}_2 , \mathbf{C}_1 e \mathbf{C}_2 precisam pertencer a um único plano. Isto é conhecido como restrição co-planar ou restrição epipolar.

Um resultado importante, que é muito utilizado nas próximas seções, é dado pela proposição a seguir.

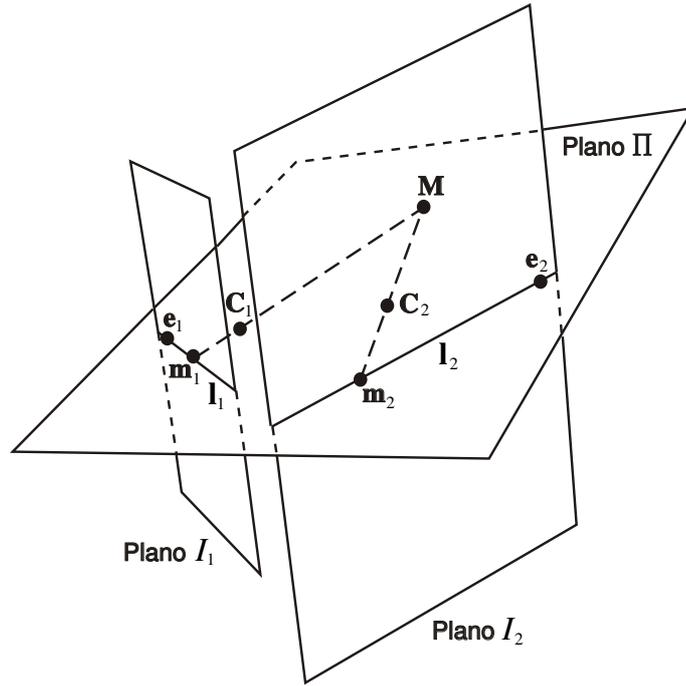


Figura 2.10: Geometria epipolar.

Proposição 2.3 Dado um par estéreo, cujas matrizes de projeção, \mathbf{P} e \mathbf{P}' , são conhecidas, os respectivos epipolos são dados por $\tilde{\mathbf{e}}_1 = \mathbf{P}\tilde{\mathbf{C}}_2$ e $\tilde{\mathbf{e}}_2 = \mathbf{P}'\tilde{\mathbf{C}}_1$.

Prova: A equação $\tilde{\mathbf{e}}_1 = \mathbf{P}\tilde{\mathbf{C}}_2$ é equivalente a afirmar que $\tilde{\mathbf{e}}_1$ é a projeção de $\tilde{\mathbf{C}}_2$. Contudo, isto é evidente, desde que $\tilde{\mathbf{e}}_1$ é a intersecção de $\overline{\mathbf{C}_1\mathbf{C}_2}$ com o plano I_1 . Isto prova a primeira parte da proposição. A segunda parte pode ser provada pela simetria da geometria epipolar. \square

2.3.1 Matriz Fundamental

Agora, sem perda de generalidade, assumimos que a câmera da esquerda (figura 2.10) está na origem do sistema de coordenadas do ambiente. Então, considerando a figura 2.10 e as equações (2.2) e (2.5), temos

$$\tilde{\mathbf{m}}_1 \simeq \mathbf{A}[\mathbf{I} \ \mathbf{0}]\tilde{\mathbf{M}} \quad \text{e} \quad \tilde{\mathbf{m}}_2 \simeq \mathbf{A}'[\mathbf{R} \ \mathbf{t}]\tilde{\mathbf{M}}. \quad (2.12)$$

A equação (2.12) pode ser reescrita em função das matrizes de projeção das câmeras, \mathbf{P} e \mathbf{P}' , ou seja,

$$\tilde{\mathbf{m}}_1 \simeq \mathbf{P}\tilde{\mathbf{M}} \quad \text{e} \quad \tilde{\mathbf{m}}_2 \simeq \mathbf{P}'\tilde{\mathbf{M}}, \quad (2.13)$$

com,

$$\mathbf{P} \simeq \mathbf{A}[\mathbf{I} \ \mathbf{0}] \quad \text{e} \quad \mathbf{P}' \simeq \mathbf{A}'[\mathbf{R} \ \mathbf{t}]. \quad (2.14)$$

Eliminando $\tilde{\mathbf{M}}$ das equações em (2.12), obtemos a seguinte equação fundamental

$$\tilde{\mathbf{m}}_2^T \mathbf{A}'^{-T}[\mathbf{t}]_{\times} \mathbf{R} \mathbf{A}^{-1} \tilde{\mathbf{m}}_1 = 0. \quad (2.15)$$

A equação (2.15) é uma restrição por trás de quaisquer duas imagens se estas são projeções em perspectiva de uma mesma cena.

Fazendo

$$\mathbf{F} = \mathbf{A}'^{-T}[\mathbf{t}]_{\times} \mathbf{R} \mathbf{A}^{-1}, \quad (2.16)$$

a equação (2.15) pode ser reescrita como

$$\tilde{\mathbf{m}}_2^T \mathbf{F} \tilde{\mathbf{m}}_1 = 0, \quad (2.17)$$

onde \mathbf{F} é conhecida como a “matriz fundamental” das duas imagens. Geometricamente, $\mathbf{F} \tilde{\mathbf{m}}_1$ define a reta epipolar do ponto \mathbf{m}_1 no plano I_2 . A equação (2.17) não diz nada além de que o ponto correspondente de \mathbf{m}_1 (em I_2), ou seja, \mathbf{m}_2 , está sobre a sua reta epipolar $\mathbf{F} \tilde{\mathbf{m}}_1$.

Transpondo a equação (2.17), encontramos a relação simétrica que relaciona I_2 com I_1 , ou seja,

$$\tilde{\mathbf{m}}_1^T \mathbf{F}^T \tilde{\mathbf{m}}_2 = 0.$$

A proposição a seguir é utilizada para derivar a principal propriedade da matriz fundamental.

Proposição 2.4 *Se \mathbf{F} é a matriz fundamental que relaciona uma imagem, I_1 , com outra imagem, I_2 , os epipolos, \mathbf{e}_1 e \mathbf{e}_2 , respectivamente, das imagens I_1 e I_2 , são dados por $\mathbf{F} \tilde{\mathbf{e}}_1 = \mathbf{0}$ e $\mathbf{F}^T \tilde{\mathbf{e}}_2 = \mathbf{0}$.*

Prova: Pelo exposto no início desta seção sobre retas epipolares e epipolos, a reta $\mathbf{F} \tilde{\mathbf{e}}_1$ é a projeção da reta 3D $\overline{\mathbf{C}_1 \mathbf{C}_2}$ no plano de imagem I_2 . Contudo, essa projeção é um único ponto, $\mathbf{e}_2 \in I_2$. Assim, $\mathbf{F} \tilde{\mathbf{e}}_1 = \mathbf{0}$. Isto prova a primeira parte da proposição. Por simetria, a segunda parte pode ser demonstrada de forma semelhante. \square

Uma conseqüência evidente da proposição anterior é a seguinte propriedade da matriz fundamental.

Propriedade 2.1 *Se \mathbf{F} é uma matriz fundamental o posto de \mathbf{F} é dois, ou seja,*

$$\det(\mathbf{F}) = 0. \quad (2.18)$$

Com esse resultado, desde que a equação (2.17) é homogênea, a matriz fundamental possui sete graus de liberdade.

A proposição 2.4 também nos leva a um resultado extremamente útil que pode ser resumido na proposição a seguir.

Proposição 2.5 *Se \mathbf{F} é uma matriz fundamental que relaciona a imagem I_1 com a imagem I_2 , o epipolo $\mathbf{e}_2 \in I_2$ é dado por $\tilde{\mathbf{e}}_2 = \mathbf{A}' \mathbf{t}$.*

Prova: De (2.16), $\mathbf{F}^T \mathbf{A}' \mathbf{t} = \mathbf{A}^{-T} \mathbf{R}^T [\mathbf{t}]_{\times}^T \mathbf{A}'^{-1} \mathbf{A}' \mathbf{t} = \mathbf{A}^{-T} \mathbf{R}^T [\mathbf{t}]_{\times}^T \mathbf{t} = \mathbf{A}^{-T} \mathbf{R}^T (-\mathbf{t} \times \mathbf{t}) = \mathbf{0}$. Então, $\mathbf{A}' \mathbf{t}$ é o autovetor de \mathbf{F}^T que corresponde ao autovalor nulo. Assim, pela proposição 2.4, $\tilde{\mathbf{e}}_2 = \mathbf{A}' \mathbf{t}$. \square

No restante do texto, os elementos da matriz fundamental serão referenciados individualmente considerando a equação a seguir.

$$\mathbf{F} = \begin{bmatrix} F_1 & F_2 & F_3 \\ F_4 & F_5 & F_6 \\ F_7 & F_8 & F_9 \end{bmatrix}. \quad (2.19)$$

§ Homografia Epipolar

Como discutido no início desta seção, a cada ponto \mathbf{M} no espaço está associado um plano epipolar definido por \mathbf{M} , \mathbf{C}_1 e \mathbf{C}_2 . Como ilustrado na figura 2.11, se \mathbf{M} gira em torno do segmento de reta $\overline{\mathbf{C}_1\mathbf{C}_2}$, define um conjunto de planos⁴. A cada um desses planos, está associada uma reta epipolar em ambas as imagens. Em cada imagem, essas retas encontram-se em um ponto comum, o epipolo. Então, quando \mathbf{M} , gira também define um conjunto de retas epipolares⁵. Além disso, considerando duas retas epipolares correspondentes, $\mathbf{l}_1 \in I_1$ e $\mathbf{l}_2 \in I_2$, e uma reta, $\mathbf{r} \in I_1$, que não corresponda a nenhuma reta epipolar, é fácil notar que $\mathbf{l}_2 \simeq \mathbf{F}[\mathbf{r}]_{\times}\mathbf{l}_1$, pois $[\mathbf{r}]_{\times}\mathbf{l}_1 = \mathbf{r} \times \mathbf{l}_1 \simeq \tilde{\mathbf{m}}_1$ é o ponto de intersecção dessas duas retas e $\mathbf{l}_2 \simeq \mathbf{F}\tilde{\mathbf{m}}_1$. Logo, podemos considerar $\mathbf{l}_2 \simeq \mathbf{H}\mathbf{l}_1$, onde $\mathbf{H} = \mathbf{F}[\mathbf{r}]_{\times}$. Isto quer dizer que retas epipolares correspondentes estão relacionadas por uma homografia. Esta homografia é chamada “homografia epipolar”. Além disso, desde que tal homografia relaciona retas que passam por \mathbf{e}_1 a retas que passam por \mathbf{e}_2 , a homografia epipolar é uma homografia do espaço \mathcal{P}^1 , representada neste trabalho por $\mathbf{h}_{ep} \in \mathbb{R}^{2 \times 2}$ e que tem apenas três graus de liberdade.

As retas epipolares \mathbf{l}_1 e \mathbf{l}_2 podem ser representadas por $\mathbf{l}_1 \simeq (\tilde{\mathbf{m}}_1 \times \tilde{\mathbf{e}}_1)$ e $\mathbf{l}_2 \simeq (\tilde{\mathbf{m}}_2 \times \tilde{\mathbf{e}}_2)$, onde $\tilde{\mathbf{m}}_1$ é um ponto qualquer sobre a reta \mathbf{l}_1 e $\tilde{\mathbf{m}}_2$ é um ponto sobre \mathbf{l}_2 . Nessas representações, o ponto $\tilde{\mathbf{m}}_1 = [u_1, v_1, t_1]^T$ pode ser substituído pelo ponto, $\tilde{\mathbf{m}}_{\infty 1} = [1, \tau_1, 0]^T$, que é a intersecção da reta \mathbf{l}_1 com a reta no infinito em I_1 . Da proposição A.1 (pág. 116), temos que

$$\tau_1 = \frac{e_t v_1 - e_v t_1}{e_t u_1 - e_u t_1}. \quad (2.20)$$

Seguindo o mesmo raciocínio, o ponto $\tilde{\mathbf{m}}_2 = [u_2, v_2, t_2]^T$ pode ser substituído por, $\tilde{\mathbf{m}}_{\infty 2} = [1, \tau_2, 0]^T$, onde

$$\tau_2 = \frac{e'_t v_2 - e'_v t_2}{e'_t u_2 - e'_u t_2}. \quad (2.21)$$

Agora, fica evidente que $\mathbf{l}_1 = (\tilde{\mathbf{m}}_{\infty 1} \times \tilde{\mathbf{e}}_1)$ e $\mathbf{l}_2 = (\tilde{\mathbf{m}}_{\infty 2} \times \tilde{\mathbf{e}}_2)$ são correspondentes se e somente se

$$\begin{bmatrix} 1 \\ \tau_2 \end{bmatrix} = \begin{bmatrix} h_1 & h_2 \\ h_3 & h_4 \end{bmatrix} \begin{bmatrix} 1 \\ \tau_1 \end{bmatrix},$$

onde $\mathbf{h}_{ep} = \begin{bmatrix} h_1 & h_2 \\ h_3 & h_4 \end{bmatrix}$. Essa equação nos leva ao seguinte resultado

$$\tau_2 = \frac{\tau_1 h_4 + h_3}{\tau_1 h_2 + h_1}. \quad (2.22)$$

⁴Do inglês: *Pencil of planes*

⁵Do inglês: *Pencil of epipolar lines*

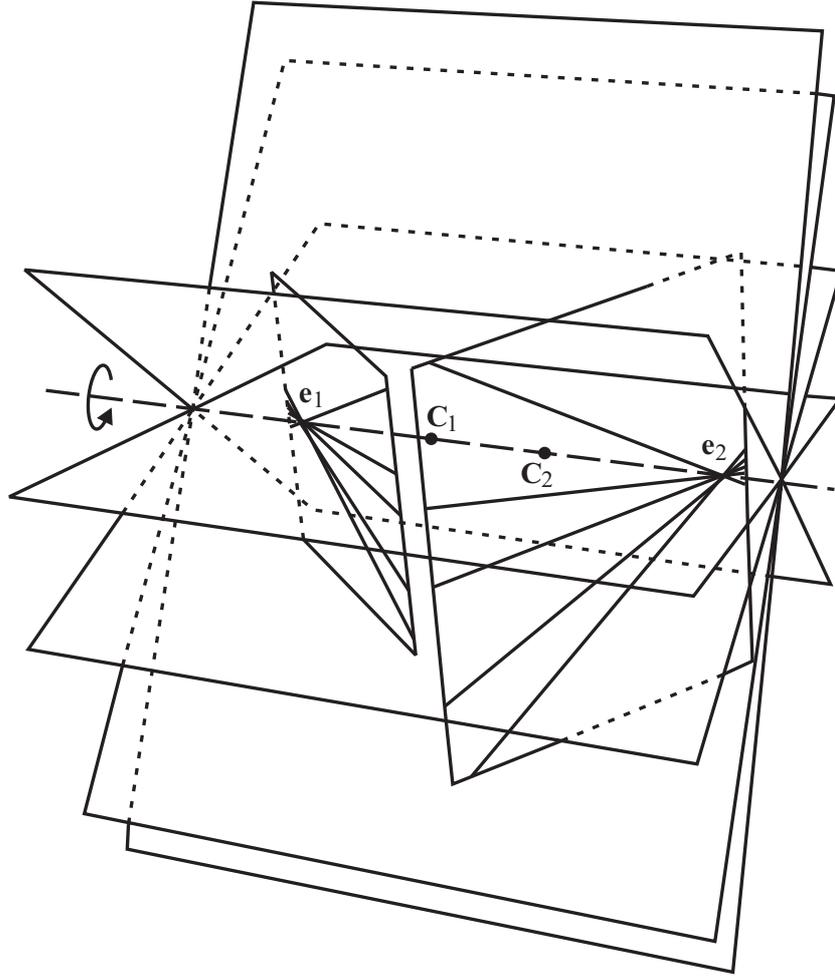


Figura 2.11: Família de planos da geometria epipolar..

A equação (2.22) é equivalente à equação fundamental (2.17). Assim, substituindo (2.20) e (2.21) em (2.22) e comparando a equação resultante com (2.17), temos que

$$\begin{aligned}
 F_1 &= -e_t e'_t h_3 \\
 F_2 &= -e_t e'_t h_4 \\
 F_3 &= e_u e'_t h_3 + e_v e'_t h_4 \\
 F_4 &= e_t e'_t h_1 \\
 F_5 &= e_t e'_t h_2 \\
 F_6 &= -e_u e'_t h_1 - e_v e'_t h_2 \\
 F_7 &= -e_t e'_v h_1 + e_t e'_u h_3 \\
 F_8 &= -e_t e'_v h_2 + e_t e'_u h_4 \\
 F_9 &= e_u e'_v h_1 + e_v e'_v h_2 - e_u e'_u h_3 - e_v e'_u h_4.
 \end{aligned} \tag{2.23}$$

Essa equação mostra que a matriz fundamental pode ser expressa em função das coordenadas dos seus epípolos e da homografia epipolar.

2.3.2 Matriz Essencial

Em um sistema de visão estéreo, a matriz essencial, \mathbf{E} , concentra toda a informação a respeito do deslocamento relativo das câmeras que compõem o sistema. Considerando a notação da figura 2.10,

$$\mathbf{E} = [\mathbf{t}]_{\times} \mathbf{R}. \quad (2.24)$$

Assim, (2.16) pode ser reescrita como

$$\mathbf{F} = \mathbf{A}'^{-T} \mathbf{E} \mathbf{A}^{-1}, \quad (2.25)$$

ou seja,

$$\mathbf{E} = \mathbf{A}'^T \mathbf{F} \mathbf{A}. \quad (2.26)$$

As principais propriedades da matriz essencial foram formuladas por Huang e Faugeras (1989) e estão descritas no teorema a seguir.

Teorema 2.1 *Se $\mathbf{E} \in \mathbb{R}^{3 \times 3}$ é uma matriz essencial, \mathbf{E} tem posto dois e dois autovalores não-nulos e iguais. Essas são condições necessárias e suficientes.*

A prova deste teorema depende do lema a seguir que é utilizado na seção 2.5.3 (pág. 44).

Lema 2.1.1 *A matriz essencial, \mathbf{E} , satisfaz $\mathbf{E}^T \mathbf{t} = \mathbf{0}$.*

Prova: De (2.24), $\mathbf{E}^T \mathbf{t} = ([\mathbf{t}]_{\times} \mathbf{R})^T \mathbf{t} = -\mathbf{R}^T [\mathbf{t}]_{\times} \mathbf{t} = -\mathbf{R}(\mathbf{t} \times \mathbf{t}) = \mathbf{0}$. □

Corolário 2.1.1 *O vetor \mathbf{t} pode ser obtido a partir da matriz \mathbf{E}^T , através da proposição B.2 (pág. 121).*

Do lema anterior, é evidente que o posto da matriz essencial, \mathbf{E} , é dois, ou seja, $\det(\mathbf{E}) = 0$. Isto prova a primeira parte do teorema 2.1. A prova da segunda parte deste teorema é mais complexa e está demonstrada em (FAUGERAS, 1993, Sec. 7.3.1.1).

O teorema 2.1 também nos leva aos dois importantes lemas a seguir, cujas respectivas provas foram originalmente apresentadas por Huang e Faugeras (1989).

Lema 2.1.2 *Se $\mathbf{E} \in \mathbb{R}^{3 \times 3}$ é uma matriz essencial, \mathbf{E} satisfaz*

$$\text{trace}^2(\mathbf{E}\mathbf{E}^T) = 2 \text{trace}((\mathbf{E}\mathbf{E}^T)^2). \quad (2.27)$$

Lema 2.1.3 *Se $\mathbf{E} \in \mathbb{R}^{3 \times 3}$ é uma matriz essencial, \mathbf{E} satisfaz*

$$(\mathbf{I}_1^2 + \mathbf{I}_2^2 + \mathbf{I}_3^2)^2 = 4(\|\mathbf{l}_1 \times \mathbf{l}_2\|^2 + \|\mathbf{l}_1 \times \mathbf{l}_3\|^2 + \|\mathbf{l}_3 \times \mathbf{l}_2\|^2), \quad (2.28)$$

onde \mathbf{l}_i é o vetor linha que corresponde a i -ésima linha de \mathbf{E} .

Apesar das equações (2.27) e (2.28) corresponderem à mesma restrição, Luong e Faugeras (1997) demonstraram que (2.27) pode ser decomposta em duas equações independentes, conhecidas como

equações de Trivedi (1988). Assim, levando em consideração a restrição de posto 2 e o fato de \mathbf{E} ser homogênea, a matriz essencial tem apenas cinco graus de liberdade, ou seja, três para a matriz \mathbf{R} e dois para o vetor \mathbf{t} . Por isso, geralmente, um vetor \mathbf{t}_i é normalizado pelo vetor unitário, \mathbf{u}_i , paralelo a \mathbf{t}_i , dado por

$$\mathbf{u}_i = \frac{\mathbf{t}_i}{\|\mathbf{t}_i\|}. \quad (2.29)$$

Assim, na maioria dos casos, apenas a direção da translação é conhecida.

§2.4 Estimação da Matriz Fundamental

Como discutido a seguir, se o sistema está calibrado, é fácil obter-se a matriz fundamental. Contudo, mesmo na ausência de calibração, tal matriz facilita-nos uma série de tarefas como, por exemplo, correspondência binocular, retificação de imagens e até mesmo calibração de câmeras. Neste caso, a matriz fundamental deve ser estimada a partir de um conjunto de correspondências de pontos.

Dada uma correspondência ($\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2i}$), podemos associar uma reta epipolar, l_{2i} , dada por $l_{2i} = \mathbf{F}\tilde{\mathbf{m}}_{1i}$, ao ponto \mathbf{m}_{2i} . Idealmente, \mathbf{m}_{2i} está sobre l_{2i} , ou seja, $d(\mathbf{m}_{2i}, l_{2i}) = 0$. Contudo, na prática, devido a ruídos nas coordenadas dos pontos ou a uma estimação “ruim” da matriz fundamental, $d(\mathbf{m}_{2i}, l_{2i}) \neq 0$. Além disso, quanto maior for essa distância, pior será o ajuste de \mathbf{F} a correspondência ($\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2i}$). Assim, dado n correspondências ($\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2i}$) e uma matriz fundamental \mathbf{F}_j , podemos usar

$$r^2(\mathbf{F}_j) = \frac{1}{2n} \sum_{i=0}^n [d^2(\tilde{\mathbf{m}}_{2i}, \mathbf{F}_j \tilde{\mathbf{m}}_{1i}) + d^2(\tilde{\mathbf{m}}_{1i}, \mathbf{F}_j^T \tilde{\mathbf{m}}_{2i})] \quad (2.30)$$

para medir o ajuste de \mathbf{F}_j ao conjunto de correspondências, ou seja, $r^2(\mathbf{F}_j)$ é uma medida da qualidade de \mathbf{F}_j com relação ao conjunto de correspondências ($\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2i}$).

A seguir, diversos métodos para estimação da matriz fundamental são analisados e comparados criticamente, em termos do custo computacional e da sua exatidão (baseada no critério da equação (2.30)).

2.4.1 Derivação Algébrica

Se um par estéreo está calibrado, é possível encontrar a matriz fundamental a partir das matrizes de projeção, \mathbf{P} e \mathbf{P}' , das câmeras. Neste trabalho, utiliza-se a metodologia proposta por Zhang e Xu (1997) e apresentada no teorema a seguir.

Teorema 2.2 *Considerando um par estéreo, cujas matrizes de projeção são \mathbf{P} e \mathbf{P}' , a matriz fundamental do sistema é dada por*

$$\mathbf{F} = [\tilde{\mathbf{e}}_2]_{\times} \mathbf{H}, \quad (2.31)$$

onde $\tilde{\mathbf{e}}_2$ é dado pela proposição 2.3 (pág. 20) e $\mathbf{H} = \mathbf{P}'\mathbf{P}^+$.

Prova: Dado um ponto \mathbf{m}_1 no plano da imagem I_1 , o raio óptico associado é $\widetilde{\mathbf{M}}(\alpha) = \mathbf{P}^+ \widetilde{\mathbf{m}}_1 + \alpha \widetilde{\mathbf{C}}_1$. Além disso, os pontos $\widetilde{\mathbf{M}} = \mathbf{P}^+ \widetilde{\mathbf{m}}_1$ e \mathbf{C}_1 (o centro da câmera) pertencem a este mesmo raio óptico (veja figura 2.10 (pág. 21)) e são projetados na imagem I_2 , respectivamente, em $\widetilde{\mathbf{m}}_2 = \mathbf{P}' \mathbf{P}^+ \widetilde{\mathbf{m}}_1$ e $\widetilde{\mathbf{e}}_2 = \mathbf{P}' \widetilde{\mathbf{C}}_1$ (o epípolo da imagem I_2). Assim, a reta epipolar de \mathbf{m}_1 pode ser expressa como sendo

$$\begin{aligned} \mathbf{l}_2 &= (\mathbf{P}' \widetilde{\mathbf{C}}_1) \times (\mathbf{P}' \mathbf{P}^+ \widetilde{\mathbf{m}}_1) \\ &= \widetilde{\mathbf{e}}_2 \times (\mathbf{P}' \mathbf{P}^+ \widetilde{\mathbf{m}}_1) \\ &= [\widetilde{\mathbf{e}}_2]_{\times} (\mathbf{P}' \mathbf{P}^+) \widetilde{\mathbf{m}}_1. \end{aligned}$$

Contudo, $\mathbf{l}_2 = \mathbf{F} \widetilde{\mathbf{m}}_1$. Então, da equação anterior, vemos que $\mathbf{F} = [\widetilde{\mathbf{e}}_2]_{\times} (\mathbf{P}' \mathbf{P}^+)$. \square

2.4.2 Solução com Sete Correspondências

Desde que a matriz \mathbf{F} tem sete graus de liberdade, são necessárias, no mínimo, sete correspondências para estimá-la univocamente. O primeiro algoritmo a realizar tal estimação a partir de um conjunto de apenas sete correspondências foi proposto por Hesse (1863) e, alguns anos depois, aperfeiçoado por Sturm (1869). Este envolve a busca das raízes de um polinômio de grau três e pode ser resumido como segue.

Dado um ponto $\mathbf{m}_{1i} = [u_{1i}, v_{1i}]^T \in I_1$ e seu correspondente $\mathbf{m}_{2i} = [u_{2i}, v_{2i}] \in I_2$, estes devem satisfazer a equação (2.17). Assim, (2.17) pode ser reescrita como sendo

$$\mathbf{u}_i^T \mathbf{f} = 0, \quad (2.32)$$

na qual $\mathbf{u}_i = [u_{1i}u_{2i}, v_{1i}u_{2i}, u_{2i}, u_{1i}v_{2i}, v_{1i}v_{2i}, v_{2i}, u_{1i}, v_{1i}, 1]^T$ e $\mathbf{f} = [\mathbf{f}_1^T, \mathbf{f}_2^T, \mathbf{f}_3^T]^T$, com \mathbf{f}_j^T igual a j -ésima linha de \mathbf{F} .

Com 7 correspondências de pontos, e considerando a equação anterior, temos

$$\mathbf{U}_7 \mathbf{f} = \mathbf{0},$$

com $\mathbf{U}_7 = [\mathbf{u}_1, \dots, \mathbf{u}_7]^T$. Este sistema pode ser resolvido com o método discutido na seção B.2.2 (pág. 122) e utilizando a propriedade 2.1 (pág. 22) para suprir a restrição exigida pelo método. Esta resolução envolve encontrar as raízes de um polinômio de grau três. Como, pelo menos, uma dessas raízes é real, o problema tem sempre solução. Se três raízes reais forem obtidas, três matrizes são candidatas à solução do problema. Evidentemente, a solução vencedora deve ser a melhor segundo o critério da equação (2.30).

A resolução com apenas sete correspondências é muito susceptível a ruídos. Felizmente, os algoritmos discutidos a seguir usam redundância para tentar solucionar esse problema.

2.4.3 Métodos Lineares

A principal vantagem dos métodos lineares para cálculo da matriz fundamental é a simplicidade. Estes permitem utilizar redundância para reduzir a influência de ruídos e, se alguns cuidados forem tomados, fornecem uma boa estimação de \mathbf{F} a um custo computacional bastante reduzido.

A seguir, esse tipo de técnica é discutida em detalhes.

§ Algoritmo dos 8 Pontos

Com n correspondências de pontos, e considerando a equação (2.32), temos o sistema

$$\mathbf{U}_n \mathbf{f} = \mathbf{0} \quad (2.33)$$

que pode ser resolvido facilmente com a proposição B.2 (pág. 121).

O método baseado na resolução da equação anterior é conhecido como “método dos 8 pontos” e foi derivado do trabalho de Longuet-Higgins (1981) sobre estimação da matriz essencial. Tal método é a maneira mais simples de estimar a matriz fundamental a partir de um conjunto de correspondências de pontos. A desvantagem evidente do método é que este não garante a restrição de posto 2. Assim, esta tem que ser imposta *a posteriori*. A forma mais conveniente de fazer isto é substituir a matriz estimada, $\hat{\mathbf{F}}$, pela matriz \mathbf{F} que minimiza $\|\hat{\mathbf{F}} - \mathbf{F}\|$, sujeito a $\det \mathbf{F} = 0$. Isto é feito fazendo $\mathbf{F} = \hat{\mathbf{F}}$ e substituindo o menor autovalor de \mathbf{F} por zero (FAUGERAS; LUONG, 2001, Sec. 6.1.3).

O resultado do algoritmo de 8 pontos depende do sistema de coordenadas utilizado. Normalmente, as coordenadas dos pontos utilizados na solução do problema são dadas em relação ao canto superior esquerdo da imagem. Contudo, uma simples mudança neste referencial torna o problema melhor condicionado e, portanto, conduz a um resultado mais estável. Esta mudança pode ser realizada através de transformações, \mathbf{T}_2 e \mathbf{T}'_2 , aplicadas ao conjunto de correspondências, ou seja, $(\mathbf{T}_2 \tilde{\mathbf{m}}_{1i} \leftrightarrow \mathbf{T}'_2 \tilde{\mathbf{m}}_{2i})$. Hartley (1997) observou que essas transformações devem ser tais que, em cada uma das imagens, independentemente, o centróide do conjunto de pontos esteja na origem e a distância média de um ponto à origem seja $\sqrt{2}$. Esta mudança produz uma grande melhora na exatidão do algoritmo. Além disso, o custo computacional acrescentado ao problema devido a esta normalização é desprezível.

Um resumo do método dos 8 pontos (incluindo o processo de normalização) está ilustrado no algoritmo 2.2 na página seguinte.

Resultado semelhante à normalização de Hartley já havia sido obtido alguns anos antes por Torr (1995). Neste caso, o problema (2.33) foi resolvido considerando os pontos como sendo $\tilde{\mathbf{x}}_{1i} = [\mathbf{m}_{1i}^T, \bar{w}_d]^T$ e $\tilde{\mathbf{x}}_{2i} = [\mathbf{m}_{2i}^T, \bar{w}_d]^T$, onde \bar{w}_d é a metade da largura da imagem. Contudo, a técnica de Hartley é mais popular e fornece melhores resultados.

Entrada: $n > 8$ correspondências ($\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2i}$).

Saída: A matriz fundamental de posto 2, \mathbf{F} .

- a) Transforme todo o conjunto de correspondências fazendo ($\mathbf{T}_2 \tilde{\mathbf{m}}_{1i} \leftrightarrow \mathbf{T}'_2 \tilde{\mathbf{m}}_{2i}$), onde \mathbf{T}_2 e \mathbf{T}'_2 são matrizes que fazem com que, em cada imagem, o centróide do conjunto de pontos esteja na origem e a distância média de um ponto à origem seja $\sqrt{2}$.
- b) Estime \mathbf{F} resolvendo o problema (2.33) usando a proposição B.2 (pág. 121).
- c) Usando a decomposição do autovalor, decomponha \mathbf{F} , tal que $\mathbf{F} = \mathbf{U} \text{diag}(\sigma_1, \sigma_2, \sigma_3) \mathbf{V}^T$, onde σ_3 é o menor autovalor de \mathbf{F} .
- d) Faça $\mathbf{F} = \mathbf{U} \text{diag}(\sigma_1, \sigma_2, 0) \mathbf{V}^T$.
- e) Faça $\mathbf{F} = \mathbf{T}'_2{}^T \mathbf{F} \mathbf{T}_2$.

Algoritmo 2.2: Algoritmo dos 8 pontos normalizado.

§ Outros Métodos Lineares

De uma forma geral, todos os métodos lineares propostos para estimar a matriz fundamental são baseados na equação (2.33). Contudo, na tentativa de melhorar ainda mais o desempenho do método dos 8 pontos, muitos outros autores propuseram modificações ao algoritmo.

Liu e Männer (2003) propuseram um método iterativo que minimiza o “Erro de Sampson” (SAMPSON, 1982) da distância dos pontos às retas epipolares correspondentes. Infelizmente, o método funciona bem apenas quando os epipolos estão próximos ao centro da imagem.

O problema (2.33) pode ser encarado como uma função a ser minimizada. Por isso, diversos autores (CHOJNACKI et al., 2003; TORR; FITZGIBBON, 2003) analisaram tal função e propuseram métodos semelhantes, baseados numa análise estatística do problema. Contudo, como observado por Torr e Fitzgibbon (2003, Sec. 4), os métodos possuem desempenho muito similares a do algoritmo de 8 pontos normalizado. Contudo, o método de Hartley é bem mais simples e direto.

2.4.4 Estimação Não-Linear

A restrição de posto dois não pode ser embutida a métodos lineares. Assim, como discutido na seção anterior, esta deve ser imposta após uma estimação inicial da matriz fundamental. Além disso, geralmente, tais métodos minimizam critérios sem significado físico. Para contornar essas limitações, algoritmos não-lineares foram propostos. Geralmente, a resolução de tais algoritmos exige uma estimação inicial da matriz fundamental. Devido a seu bom desempenho, o método dos 8 pontos normalizado é bastante utilizado para esta tarefa. Sua principal função é fornecer uma solução inicial bem próxima de um mínimo local. Dessa forma, a solução dos problemas não-lineares é atingida após poucas iterações.

Nas próximas seções, alguns dos métodos não-lineares mais importantes são analisados criticamente.

§ Impondo o Posto da Matriz por Parametrização

Como discutido na seção 2.3.1 (pág. 23), a matriz fundamental pode ser escrita em função dos epipolos e da homografia epipolar. Então, uma abordagem para impor a restrição de posto dois a matriz fundamental é parametrizá-la baseado na equação (2.23).

Supondo que $e_t \neq 0$ e $e'_t \neq 0$, ou seja, os epipolos não estão no infinito, podemos dividir a equação (2.23) por $-F_1 = e_t e'_t h_3$. Assim, considerando $\mathbf{e}_1 = [\alpha, \beta]^T$ e $\mathbf{e}_2 = [\lambda, \mu]^T$, podemos expressar a matriz fundamental em termos dos epipolos e da homografia epipolar, ou seja,

$$\mathbf{F} = \begin{bmatrix} 1 & a & -\alpha - a\beta \\ b & c & -b\alpha - c\beta \\ -\lambda - b\mu & -a\lambda - c\mu & F_9 \end{bmatrix}, \quad (2.34)$$

onde $a = -h_4/h_3$, $b = h_1/h_3$, $c = h_2/h_3$, $F_9 = \alpha(\lambda + b\mu) + \beta(a\lambda + c\mu)$.

A equação anterior expressa a matriz fundamental com apenas sete incógnitas. Este fato está de acordo com o discutido na seção 2.3.1 sobre a matriz fundamental possuir sete graus de liberdade. Além disso, considerando $\mathbf{F} = [\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3]$, a equação anterior mostra que $\mathbf{c}_3 = -\alpha\mathbf{c}_1 - \beta\mathbf{c}_2$, isto é, $\alpha\mathbf{c}_1 + \beta\mathbf{c}_2 + \mathbf{c}_3 = \mathbf{0}$. Isto está de acordo com a proposição 2.4 (pág. 22) e mostra que a parametrização escolhida expressa uma matriz de posto dois.

A equação (2.34) foi utilizada como parametrização pela primeira vez por Luong et al. (1993). Contudo, esta torna-se numericamente instável a medida que os epipolos tendem para o infinito. Além disso, ela não consegue expressar uma matriz fundamental que tenha uma das três formas a seguir: $[\mathbf{c}_1, \mathbf{c}_2, \alpha\mathbf{c}_2]$, $[\mathbf{c}_1, \mathbf{0}, \alpha\mathbf{c}_3]$ e $[\mathbf{c}_1, \mathbf{c}_2, \mathbf{0}]$. Assim, devemos estender a parametrização considerando outras linhas e outras colunas como linearmente independentes.

De uma forma geral, uma parametrização é definida escolhendo-se, como independentes, um par de colunas $(\mathbf{c}_{i_c}, \mathbf{c}_{j_c})$ e um par de linhas $(\mathbf{r}_{i_r}, \mathbf{r}_{j_r})$, onde $\{i_c, j_c, i_r, j_r\} \in [1, 3]$. Dessa forma, existe $\{\alpha_{i_c}, \beta_{j_c}, \lambda_{i_r}, \mu_{j_r}\} \in \mathbb{R}^4$, tal que a outra coluna, $\mathbf{c}_{j_c}^{i_c}$, e a outra linha, $\mathbf{r}_{j_r}^{i_r}$, possam ser representadas como $\mathbf{c}_{j_c}^{i_c} = -\alpha_{i_c}\mathbf{c}_{i_c} - \beta_{j_c}\mathbf{c}_{j_c}$ e $\mathbf{r}_{j_r}^{i_r} = -\lambda_{i_r}\mathbf{r}_{i_r} - \mu_{j_r}\mathbf{r}_{j_r}$ respectivamente. Desde que existem três combinações de pares para as linhas e três para as colunas, temos um total de nove parametrizações distintas.

Como na equação (2.34), as colunas $(\mathbf{c}_{i_c}, \mathbf{c}_{j_c})$ e as linhas $(\mathbf{r}_{i_r}, \mathbf{r}_{j_r})$ vão ser formadas por quatro elementos independentes, representados aqui por a, b, c e d . Como no caso anterior, para levar em consideração o fato de \mathbf{F} ser homogênea, um desses elementos deve ser escolhido para dividir os outros três. Luong et al. (1993) comprovou experimentalmente que melhores resultados são obtidos quando o elemento escolhido é aquele de maior valor absoluto.

Em resumo, a escolha de (i_c, j_c) e de (i_r, j_r) nos leva ao seguinte vetor de oito elementos

$$\mathbf{f}_8(i_c, j_c, i_r, j_r, \gamma) = [\alpha_{i_c}, \beta_{j_c}, \lambda_{i_r}, \mu_{j_r}, \frac{a}{\gamma}, \frac{b}{\gamma}, \frac{c}{\gamma}, \frac{d}{\gamma}]^T, \quad (2.35)$$

onde $\gamma \in \{a, b, c, d\}$. Assim, após a escolha de γ , eliminando o elemento unitário de \mathbf{f}_8 , tem-

se uma parametrizao de \mathbf{F} com apenas sete parmetros. Assim, dado $n \geq 8$ correspondncias $(\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2i})$, tal parametrizao pode ser utilizada para resolver

$$\min_{\mathbf{F}} \sum_{i=1}^n \tilde{\mathbf{m}}_{2i}^T \mathbf{F} \tilde{\mathbf{m}}_{1i}. \quad (2.36)$$

Evidentemente, para resolver tal problema utilizando uma parametrizao baseada em (2.35), um algoritmo de minimizao no-linear deve ser utilizado.

Pelo discutido, (2.35) nos leva a $3 \times 3 \times 4 = 36$ parametrizaes distintas. Analisando o jacobiano da funo que, dado um vetor $\mathbf{f}_8(i_c, j_c, i_r, j_r, \gamma)$, computa \mathbf{F} com apenas sete parmetros, Csurka et al. (1997) demonstraram que a melhor parametrizao  aquela que minimiza a norma dada por

$$(ad - bc)^2 \sqrt{(\alpha_{i_c}^2 + \beta_{j_c}^2 + 1)(\lambda_{i_r}^2 + \mu_{j_r}^2 + 1)}. \quad (2.37)$$

Para resolver (2.35), uma estimaco inicial, \mathbf{F}_0 , do problema  necessria. Para isso, geralmente,  utilizado o algoritmo dos 8 pontos. Com \mathbf{F}_0 , a equao anterior  utilizada para escolher-se a parametrizao a ser utilizada. Contudo, a parametrizao utilizada para representar \mathbf{F}_0 no , necessariamente, a melhor para representar a soluo final do problema, ou seja, em um determinado passo do processo de minimizao, a parametrizao utilizada pode mudar para uma que melhor se ajuste (segundo o critrio (2.37))  matriz fundamental daquele passo. Evidentemente, isto nos conduz a um algoritmo inconveniente, propenso a erros e de custo computacional elevado.

Apesar de impor a restrio de posto dois por parametrizao, a soluo de (2.36) tambm minimiza um critrio sem significado fsico. Uma soluo para este problema  discutida a seguir.

§ Minimizando as Distncias dos Pontos s Retas Epipolares

Se o sistema no est calibrado, durante o processo de estimaco da matriz fundamental, a nica informao disponvel so as correspondncias $(\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2i})$. Assim, o critrio a ser minimizado deve ser algo medido no plano de imagem. Uma possibilidade  minimizar a distncia dos pontos s retas epipolares correspondentes, ou seja,

$$\min_{\mathbf{F}} \sum_{i=0}^n [d^2(\tilde{\mathbf{m}}_{2i}, \mathbf{F}\tilde{\mathbf{m}}_{2i}) + d^2(\tilde{\mathbf{m}}_{1i}, \mathbf{F}^T \tilde{\mathbf{m}}_{2i})].$$

Essa equao pode ser reescrita como

$$\sum_{i=0}^n w_i (\tilde{\mathbf{m}}_{2i}^T \mathbf{F} \tilde{\mathbf{m}}_{1i})^2 = 0, \quad (2.38)$$

onde

$$w_i = \left(\frac{1}{\mathbf{l}_{1i}^T \mathbf{Z} \mathbf{l}_{1i}} + \frac{1}{\mathbf{l}_{2i}^T \mathbf{Z} \mathbf{l}_{2i}} \right), \quad (2.39)$$

com $\mathbf{Z} = \text{diag}(1, 1, 0)$, $\mathbf{l}_{2i} = \mathbf{F} \tilde{\mathbf{m}}_{1i}$ e $\mathbf{l}_{1i} = \mathbf{F}^T \tilde{\mathbf{m}}_{2i}$.

O solução do problema (2.38) também requer uma parametrização para impor a restrição de posto dois. Neste caso, a mesma parametrização descrita na seção anterior pode ser utilizada.

A abordagem descrita nesta seção foi proposta originalmente por Luong e Faugeras (1996) e tem provado ser um dos métodos mais exatos de estimação da matriz fundamental.

§ Minimizando a Distância Sampson

A matriz fundamental também pode ser obtida encontrando \mathbf{F} que minimiza o erro geométrico de

$$\sum_{i=1}^n \tilde{\mathbf{m}}_{2i}^T \mathbf{F} \tilde{\mathbf{m}}_{1i} = 0.$$

Uma aproximação de primeira-ordem desse erro pode ser obtida a partir do trabalho de Sampson (1982), ou seja,

$$\sum_{i=0}^n g_i (\tilde{\mathbf{m}}_{2i}^T \mathbf{F} \tilde{\mathbf{m}}_{1i})^2 = 0, \quad (2.40)$$

onde

$$g_i = \left(\frac{1}{\mathbf{1}_{1i}^T \mathbf{Z} \mathbf{1}_{1i} + \mathbf{1}_{2i}^T \mathbf{Z} \mathbf{1}_{2i}} \right). \quad (2.41)$$

A solução para o problema (2.40) também exige uma parametrização para impôr a restrição de posto dois. Evidentemente, a parametrização com os 36 mapas de busca pode ser utilizada. Nesta caso, devido as semelhanças entre as equações (2.38) e (2.40) e entre (2.39) e (2.41), este método fornece resultados quase idênticos ao método discutido na seção anterior. Este fato será demonstrado empiricamente no capítulo 4.

A equação (2.40) também pode ser obtida dividindo-se cada termo da equação (2.36) pela sua variância. Neste caso, o método é, normalmente, referenciado por Método Baseado no Gradiente⁶ (FAUGERAS; LUONG, 2001; ZHANG; LOOP, 2001; ZHANG, 1998).

§ Método Ótimo

Como discutido na seção 2.4.1, a matriz fundamental está diretamente relacionada às matrizes de projeção das câmeras. Além disso, como apresentado na seção 2.5.2, Hartley et al. (1992) demonstraram como recuperar um par de matrizes de projeção, \mathcal{P} e \mathcal{P}' , pertencente ao espaço projetivo, mesmo quando as câmeras não estão calibradas. Este fato permite-nos, dado um conjunto de n correspondências ($\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2i}$), estimar a matriz fundamental minimizando

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

Tal método baseia-se no fato de que as correspondências ($\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2i}$) estão “poluídas” por ruído. Então, devemos encontrar os pares ($\mathbf{x}_{1i} \leftrightarrow \mathbf{x}_{2i}$) livres de ruído e a matriz fundamental \mathbf{F} que satisfa-

⁶Do inglês: *Gradient-Based Method*

zem a equação fundamental (2.17) exatamente.

O problema anterior é equivalente ao seguinte problema irrestrito

$$\sum_{i=1}^n d^2(\tilde{\mathbf{m}}_{1i}, \mathcal{P}\tilde{\boldsymbol{\chi}}_i) + d^2(\mathbf{m}_{2i}, \mathcal{P}'\tilde{\boldsymbol{\chi}}_i),$$

onde $\tilde{\boldsymbol{\chi}}_i$ é o ponto em \mathcal{P}^3 que corresponde a correspondência $(\mathbf{x}_{1i} \leftrightarrow \mathbf{x}_{2i})$.

A solução do problema anterior requer uma estimativa inicial dos pontos $\tilde{\boldsymbol{\chi}}_i$ e das matrizes \mathcal{P} e \mathcal{P}' . Tal estimativa pode ser obtida pelos métodos descritos nas seções 2.5.2 e 2.7. Infelizmente, para um conjunto de n correspondências de pontos, o número de variáveis que necessitam ser estimadas é igual a $3n + 12$. Evidentemente, à medida que n cresce, a complexidade do problema aumenta significativamente.

§ Outros Métodos Não-Lineares

Zhang e Loop (2001) propuseram um método no qual transformações de \mathcal{P}^2 são aplicadas às correspondências $(\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2i})$. Essas transformações são tais que a matriz fundamental que relaciona os pontos transformados pode sempre ser parametrizada pela equação (2.34). Em seguida, a geometria epipolar é obtida resolvendo-se o problema (2.40). Evidentemente, reduzir o número de parametrizações de 36 para apenas uma também reduz a complexidade do problema. A desvantagem desta técnica é que as transformações utilizadas são determinadas a partir de uma estimativa inicial, \mathbf{F}_0 , da matriz fundamental. Contudo, como o próprio Zhang (1998) sugeriu, a parametrização utilizada para representar \mathbf{F}_0 não é, necessariamente, a melhor para representar a solução final do problema. Dessa forma, em alguns casos, o método com 36 parametrizações deve fornecer resultados mais exatos que os obtidos com o método de Zhang e Loop (2001).

Recentemente, Bartoli e Sturm (2004) propuseram um interessante método. Neste, através de SVD, a matriz fundamental é decomposta em $\mathbf{F} = \mathbf{U} \text{diag}(1, \sigma, 0) \mathbf{V}^T$, onde σ é a razão entre os dois autovalores de \mathbf{F} . Dessa forma, a matriz fundamental pode ser atualizada através das seguintes equações

$$\mathbf{U} \leftarrow \mathbf{UR}(\boldsymbol{\theta}_u), \quad \mathbf{V} \leftarrow \mathbf{VR}(\boldsymbol{\theta}_v), \quad \sigma \leftarrow \sigma + \delta_\sigma,$$

onde $\mathbf{R}(\boldsymbol{\theta}_i)$ representa matrizes de rotação baseadas nos ângulos de Euler⁷. Dessa forma, usando um algoritmo não-linear de otimização, é possível ajustar \mathbf{F} , iterativamente, a um conjunto de correspondências. A vantagem do método é que \mathbf{F} é representada com o mínimo de parâmetros possível (três para o vetor $\boldsymbol{\theta}_u$, três para $\boldsymbol{\theta}_v$ e um para δ_σ). Contudo, os resultados mostram que o algoritmo tende a ficar “preso” em mínimos locais e, nestes casos, oferece uma estimativa “ruim” de \mathbf{F} .

⁷Do inglês: *Euler angles*. É uma maneira de representar rotações no espaço 3D como um produto de 3 rotações sucessivas em torno dos eixos x , y e z .

2.4.5 Estimação Robusta

Devido a dificuldade de se estabelecer correspondência de pontos antes da geometria epipolar ser conhecida, espera-se que, em um conjunto de correspondências ($\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2i}$), existam erros e ruído nas coordenadas dos pontos \mathbf{m}_{1i} e \mathbf{m}_{2i} . Isto nos permite classificar as correspondências em boas ou falsas. Esses conceitos são definidos formalmente a seguir.

Definição 2.4 (Boas Correspondências) *Dada um conjunto de correspondências ($\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2i}$), se o ruído presente nas coordenadas dos pontos \mathbf{m}_{1i} e \mathbf{m}_{2i} for de até dois pixels, estas correspondências são chamadas “boas correspondências”.*

Definição 2.5 (Falsas Correspondências) *Uma correspondência ($\mathbf{m}_1 \leftrightarrow \mathbf{m}_2$) é dita uma “falsa correspondência” ou uma “correspondência ruim” quando, de fato, o ponto \mathbf{m}_2 não está sobre a reta epipolar de \mathbf{m}_1 .*

Normalmente, supõe-se que o ruído presente nas coordenadas dos pontos, \mathbf{m}_{1i} e \mathbf{m}_{2i} , das boas correspondências possui uma distribuição Gaussiana. Isso é uma suposição razoável, pois os erros nas coordenadas são de apenas poucos pixels.

Os algoritmos para cálculo da matriz fundamental são capazes de lidar com o ruído presente nas boas correspondências. Contudo, apenas algumas poucas falsas correspondências presentes no conjunto são capazes de degradar totalmente a estimação de \mathbf{F} . Neste caso, algoritmos robustos que conseguem tratar com tais correspondências devem ser utilizados.

Os algoritmos robustos mais bem sucedidos tentam detectar as falsas correspondências e retirá-las do conjunto. Desse modo, ao final do algoritmo, a matriz fundamental pode ser estimada por um dos algoritmos anteriores, pois haverá apenas boas correspondências.

A seguir, alguns dos algoritmos robustos mais utilizados para estimação da matriz fundamental são descritos em detalhes.

§ RANSAC

Um dos métodos robustos mais conhecidos na literatura é o RANSAC⁸. Este foi introduzido na visão computacional por Fischler e Bolles (1981). Basicamente, o RANSAC pode ser resumido nos passos a seguir.

Dado um conjunto de N correspondências $c_i = (\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2i})$, 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 $r_{(win)i}^2 < r_{th}^2$).

⁸Do inglês: *Random Sample Consensus*.

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, existe uma probabilidade p de que (ao menos) um subconjunto j seja composto apenas por boas correspondências. Neste caso, N_{sc} é igual a

$$N_{sc} = \log(1 - p) / \log(1 - (1 - \epsilon)^{N_{nc}}). \quad (2.42)$$

Talvez, uma tática inteligente seja iniciar N_{sc} usando a equação anterior e atualizá-lo a cada iteração j , ou seja, uma vez que a percentagem, ϵ_j , de falsas correspondências tenha sido determinada, N_{sc} pode ser atualizado por (2.42).

Da equação anterior, vemos 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 termos N_{nc} e ϵ tão pequenos quanto possível.

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, como visto na seção 2.4.2, a solução com apenas sete correspondências não é estável. Por isso, podemos utilizar N_{nc} igual ou um pouco maior que oito.

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, como o critério utilizado para medir a qualidade da matriz fundamental é dado por (2.30), o resíduo r_{ji}^2 é dado por

$$r_{ji}^2 = d^2(\tilde{\mathbf{m}}_{2i}, \mathbf{F}_j \tilde{\mathbf{m}}_{1i}) + d^2(\tilde{\mathbf{m}}_{1i}, \mathbf{F}_j^T \tilde{\mathbf{m}}_{2i}). \quad (2.43)$$

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

$$r_{th}^2 = d_{th}^2 \sigma_d^2, \quad (2.44)$$

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 (resumido no algoritmo 2.3) pode obter uma boa estimação de \mathbf{F} mesmo que mais de 50% das correspondências forem ruins. A desvantagem evidente é que este necessita de uma estimação do desvio padrão do ruído, σ_d .

Entrada: Um conjunto, C , de $N \geq 7$ correspondências $c_i = (\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2i})$.

Saída: A matriz fundamental, \mathbf{F} .

```

1  /* Inicialização das variáveis. */
2   $N_{sc} = \log(1 - p) / \log(1 - (1 - \epsilon)^{N_{nc}})$ ;    /* Calcula o número de iterações. */
3   $S = \text{FUNC\_ESCOLHE\_SUBCONJUNTOS}(C, N_{sc})$ ;    /* Escolhe os  $N_{sc}$  subconjuntos,  $S_j$ . */
4   $j = 1$ ;
5   $InliersCount = 0$ ;    /* Inicializa contador de boas correspondências. */
6
7  /* Iteração. */
8  while  $j \leq N_{sc}$  do
9   $\mathbf{F}_j = \text{FUNC\_ESTIMA\_F}(S_j)$ ;    /* Estima  $\mathbf{F}_j$  com  $N_{nc}$  correspondências. */
10 for  $i = 1$  to  $N$  do    /* Determina correspondências que ajustam-se a  $\mathbf{F}_j$ . */
11  $r_{ji}^2 = d^2(\tilde{\mathbf{m}}_{2i}, \mathbf{F}_j \tilde{\mathbf{m}}_{1i}) + d^2(\tilde{\mathbf{m}}_{1i}, \mathbf{F}_j^T \tilde{\mathbf{m}}_{2i})$ ;
12 if  $r_{ji}^2 \leq (d_{th} \sigma_d)^2$  then
13  $inliers = inliers + 1$ ;
14 endif
15 endfor
16 if  $InliersCount < inliers$  then    /* Se  $\mathbf{F}_j$  tem o melhor ajuste... */
17  $\mathbf{F}_{win} = \mathbf{F}_j$ ;    /* ...ela é a matriz vencedora */
18  $InliersCount = inliers$ ;
19 endif
20  $j = j + 1$ ;
21 endwhile
22
23 /* Usa  $\mathbf{F}_{win}$  para calcular  $r_{(win)i}^2$ . */
24 for  $i = 1$  to  $N$  do
25 if  $r_{(win)i}^2 \leq (d_{th} \sigma_d)^2$  then    /* Se  $r_{ji}^2$  é menor que o limiar... */
26  $Inliers \leftarrow c_i$ ;    /* ... $c_i$  é uma boa correspondências. */
27 endif
28 endfor
29
30  $\mathbf{F} = \text{FUNC\_ESTIMA\_F}(Inliers)$ ;    /* Estima  $\mathbf{F}$  apenas com as boas correspondências. */

```

Algoritmo 2.3: Algoritmo RANSAC para estimação da matriz fundamental.

§ MSAC

De uma 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. Neste 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 MSAC⁹) é preferível em comparação ao RANSAC.

§ MLESAC

Dada uma correspondência ($\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2i}$), na prática, \mathbf{m}_{1i} e \mathbf{m}_{2i} não satisfazem a equação fundamental (2.17) exatamente, pois as coordenadas desses pontos estão corrompidas por ruído. Contudo, dada uma matriz fundamental, \mathbf{F}_j , podemos encontrar a correspondência, ($\hat{\mathbf{m}}_{1i} \leftrightarrow \hat{\mathbf{m}}_{2i}$), mais próxima de ($\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2i}$) e que satisfaz (2.17), ou seja, podemos encontrar ($\hat{\mathbf{m}}_{1i} \leftrightarrow \hat{\mathbf{m}}_{2i}$) que minimize

$$r_{ji}^2 = d^2(\mathbf{m}_{1i}, \hat{\mathbf{m}}_{1i}) + d^2(\mathbf{m}_{2i}, \hat{\mathbf{m}}_{2i}), \quad \text{sujeito a } \tilde{\mathbf{m}}_{2i}^T \mathbf{F}_j \tilde{\mathbf{m}}_{1i}. \quad (2.45)$$

Evidentemente, na equação anterior, quanto menor r_{ji}^2 , melhor o ajuste de \mathbf{F}_j a correspondência ($\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2i}$). Por isso, Torr e Zisserman (2000) propuseram um novo método robusto, chamado MLESAC¹⁰, baseado na minimização da equação anterior pelo método MSAC.

Torr e Zisserman (2000) demonstraram que o algoritmo MLESAC produz uma melhor estimação da matriz fundamental (em comparação ao MSAC e ao RANSAC). Contudo, do trabalho de Hartley e Sturm (1997), sabe-se que a correspondência ($\hat{\mathbf{m}}_{1i} \leftrightarrow \hat{\mathbf{m}}_{2i}$), na equação (2.45), é encontrada determinando-se as raízes de um polinômio de grau seis. Por isso, apesar de Torr e Murray (1997) terem derivado uma aproximação linear para tal polinômio, o MLESAC possui um custo computacional muito elevado e nunca tornou-se muito popular.

§ LMedS

O algoritmo RANSAC só pode ser utilizado quando temos informação sobre a distribuição do ruído inserido no conjunto de correspondências. Como discutido na seção anterior, essa informação é utilizada no critério que define se uma correspondência é boa ou não. Caso tal informação não esteja disponível, podemos estimar \mathbf{F} utilizando o método LMedS¹¹ (resumido no algoritmo 2.4). Este é muito semelhante ao RANSAC. A principal diferença é que uma matriz \mathbf{F}_j é considerada boa quando a mediana do resíduo, dado pela equação (2.43), é pequena. Assim, após \mathbf{F}_j ser estimada, calcula-se

⁹Do inglês: *M-Estimator Sample Consensus*.

¹⁰Do inglês: *Maximum Likelihood Estimation Sample Consensus*.

¹¹Do inglês: *Least Median of Squares*.

Entrada: Um conjunto, C , de $N \geq 7$ correspondências $c_i = (\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{1i})$.

Saída: A matriz fundamental, \mathbf{F} .

```

1  /* Inicialização das variáveis. */
2   $N_{sc} = \log(1 - p) / \log(1 - (1 - \epsilon)^{N_{nc}})$ ;    /* Calcula o número de iterações. */
3   $S = \text{FUNC\_ESCOLHE\_SUBCONJUNTOS}(C, N_{sc})$ ;    /* Escolhe os  $N_{sc}$  subconjuntos,  $S_j$ . */
4   $j = 1$ ;
5   $MinMed = \infty$ ;    /* Inicializa a mediana do resíduo. */
6
7  /* Iteração. */
8  while  $j \leq N_{sc}$  do
9   $\mathbf{F}_j = \text{FUNC\_ESTIMA\_F}(S_j)$ ;    /* Estima  $\mathbf{F}_j$  com  $N_{nc}$  correspondências. */
10 for  $i = 1$  to  $N$  do    /* Determina o ajuste das correspondências a  $\mathbf{F}_j$ . */
11  $r_{ji}^2 = d^2(\tilde{\mathbf{m}}_{2i}, \mathbf{F}_j \tilde{\mathbf{m}}_{1i}) + d^2(\tilde{\mathbf{m}}_{1i}, \mathbf{F}_j^T \tilde{\mathbf{m}}_{2i})$ ;
12 endfor
13  $D_j = \text{MEDIAN}(r_{j1}^2, r_{j2}^2, \dots, r_{jN}^2)$ ;    /* Calcula mediana da distância. */
14 if  $MinMed > D_j$  then    /* Se  $\mathbf{F}_j$  tem o melhor ajuste... */
15  $\mathbf{F}_{win} = \mathbf{F}_j$ ;    /* ...ela é a matriz vencedora */
16  $MinMed = D_j$ ;
17 endif
18  $j = j + 1$ ;
19 endwhile
20
21  $\hat{\sigma}_d = 1,4826[1 + 5/(N - N_{nc})]\sqrt{MinMed}$ ;    /* Calcula o desvio padrão robusto. */
22
23 /* Usa  $\mathbf{F}_{win}$  para calcular  $r_{(win)i}^2$ . */
24 for  $i = 1$  to  $N$  do
25 if  $r_{(win)i}^2 \leq (d_{th}^2 \hat{\sigma}_d^2)$  then    /* Se  $r_{ji}^2$  é menor que o limiar... */
26  $Inliers \leftarrow c_i$ ;    /* ... $c_i$  é uma boa correspondências. */
27 endif
28 endfor
29
30  $\mathbf{F} = \text{FUNC\_ESTIMA\_F}(Inliers)$ ;    /* Estima  $\mathbf{F}$  apenas com as boas correspondências. */
31

```

Algoritmo 2.4: Algoritmo LMedS para estimação da matriz fundamental.

o resíduo, r_{ji}^2 , de todas as N correspondências. Após N_{sc} iterações, retém-se a matriz fundamental, \mathbf{F}_{win} , que obteve a menor mediana.

Após \mathbf{F}_{win} ter sido encontrada, um interessante resultado pode ser utilizado para eliminar as falsas correspondências. Rousseeuw (1987), através de análises estatísticas, determinou que uma boa estimação para σ_d pode ser dada por

$$\hat{\sigma}_d^2 = 1,4826[1 + 5/(N - N_{nc})]\sqrt{d_{med}^2}, \quad (2.46)$$

onde d_{med}^2 é a menor mediana de r_{ji}^2 dentre todos os N_{sc} subconjuntos j .

O algoritmo LMedS tem duas grandes desvantagens. Ele não consegue estimar a matriz fundamental corretamente se existirem mais de 50% de falsas correspondências. Além disso, desde que nenhuma informação sobre o ruído presente nos dados está disponível, N_{sc} não pode ser atualizado dinamicamente. Por isso, o custo computacional do LMedS tende a ser maior que o do RANSAC. Mesmo assim, diversos autores (ARMANGUÉ; SALVI, 2003; ZHANG, 1998; TORR; MURRAY,

1997) apontam o LMedS como o algoritmo robusto que obtém a melhor estimativa de \mathbf{F} .

§2.5 Calibração da Câmera

Como discutido na seção 2.1 (pág. 7), uma câmera é modelada como um dispositivo que projeta pontos do espaço 3D em pontos em um plano. Este processo não é reversível, ou seja, dado um ponto numa imagem não é possível determinar as coordenadas do ponto 3D correspondente. Isto deve-se ao fato de que um ponto, \mathbf{m} , no plano de imagem pode corresponder a qualquer ponto sobre uma determinada reta (associada a \mathbf{m}) no espaço. Sendo assim, é necessário uso de redundância, na forma de uma segunda câmera, para que, em um processo chamado reconstrução 3D, as coordenadas de um ponto no espaço possam ser conhecidas univocamente. Contudo, isto só é possível se o processo pelo qual as câmeras projetam pontos no espaço em pontos nos planos for bem conhecido, ou seja, os parâmetros intrínsecos e extrínsecos do conjunto de câmeras forem conhecidos. Para isto, é utilizado um processo chamado calibração. Por isso, geralmente, a calibração é a primeira tarefa realizada por um sistema de visão estéreo. Nessa seção, alguns dos métodos mais importantes de calibração de câmeras são analisados criticamente.

2.5.1 Método Clássico

Dado um ponto 3D, \mathbf{M}_i , e sua projeção no plano da imagem, \mathbf{m}_i , a equação (2.7) (pág. 10) pode ser expressa em termos do produto cruzado, ou seja, $\tilde{\mathbf{m}}_i \times \mathbf{P}\tilde{\mathbf{M}}_i = \mathbf{0}$. Além disso, se a j -ésima linha da matriz \mathbf{P} for denotada por \mathbf{p}_j^T , podemos escrever

$$\mathbf{P}\tilde{\mathbf{M}}_i = [\mathbf{p}_1^T \tilde{\mathbf{M}}_i, \mathbf{p}_2^T \tilde{\mathbf{M}}_i, \mathbf{p}_3^T \tilde{\mathbf{M}}_i]^T.$$

Assim, considerando, $\tilde{\mathbf{m}}_i = [u_i, v_i, t_i]^T$, podemos escrever

$$\tilde{\mathbf{m}}_i \times \mathbf{P}\tilde{\mathbf{M}}_i = \begin{bmatrix} v_i \mathbf{p}_3^T \tilde{\mathbf{M}}_i - t_i \mathbf{p}_2^T \tilde{\mathbf{M}}_i \\ t_i \mathbf{p}_1^T \tilde{\mathbf{M}}_i - u_i \mathbf{p}_3^T \tilde{\mathbf{M}}_i \\ u_i \mathbf{p}_2^T \tilde{\mathbf{M}}_i - v_i \mathbf{p}_1^T \tilde{\mathbf{M}}_i \end{bmatrix} = \mathbf{0}.$$

Contudo, como $\mathbf{p}_j^T \tilde{\mathbf{M}}_i = \tilde{\mathbf{M}}_i^T \mathbf{p}_j$, $j = 1, \dots, 3$, a equação anterior pode ser escrita como

$$\mathbf{U}_i \mathbf{p} = \mathbf{0},$$

onde $\mathbf{p} = [\mathbf{p}_1^T, \mathbf{p}_2^T, \mathbf{p}_3^T]^T$ e

$$\mathbf{U}_i = \begin{bmatrix} \mathbf{0}^T & -t_i \tilde{\mathbf{M}}_i^T & v_i \tilde{\mathbf{M}}_i^T \\ t_i \tilde{\mathbf{M}}_i^T & \mathbf{0}^T & -u_i \tilde{\mathbf{M}}_i^T \\ -v_i \tilde{\mathbf{M}}_i^T & u_i \tilde{\mathbf{M}}_i^T & \mathbf{0}^T \end{bmatrix}.$$

Entrada: $n > 6$ pares $(\mathbf{M}_i, \mathbf{m}_i)$.

Saída: A matriz de projeção da câmera, \mathbf{P} .

- a) Com os pontos \mathbf{M}_i e \mathbf{m}_i , compute as matrizes \mathbf{T}_3 e \mathbf{T}_2 respectivamente.
- b) Considerando os pares $(\mathbf{T}_3\mathbf{M}_i, \mathbf{T}_2\mathbf{m}_i)$, estime a matriz, \mathbf{P} , resolvendo o sistema (2.47) utilizando a proposição B.2 (pág. 121).
- c) Faça $\mathbf{P} = \mathbf{T}_2^{-T}\mathbf{P}\mathbf{T}_3^{-1}$.

Algoritmo 2.5: Calibração linear clássica de câmeras.

A equação anterior possui apenas duas equações independentes. Além disso, \mathbf{P} tem onze graus de liberdade. Assim, são necessários, no mínimo, seis pares $(\mathbf{M}_i, \mathbf{m}_i)$ para determiná-la univocamente. Além disso, se $n > 6$ desses pares estão disponíveis, podemos resolver o sistema de equações

$$\begin{bmatrix} \mathbf{U}_1 \\ \mathbf{U}_2 \\ \vdots \\ \mathbf{U}_n \end{bmatrix} \mathbf{p} = \mathbf{0} \quad (2.47)$$

utilizando a proposição B.2 (pág. 121).

Como no algoritmo dos 8 pontos, o resultado do método da equação anterior depende das coordenadas dos pares $(\mathbf{M}_i, \mathbf{m}_i)$. Assim, usando a mesma técnica da seção 2.4.3 (pág. 28), os pontos \mathbf{m}_i devem ser normalizados utilizando-se a transformação \mathbf{T}_2 . Os pontos \mathbf{M}_i também devem ser transformados por uma homografia, \mathbf{T}_3 . Contudo, neste caso, o centróide do conjunto de pontos $\mathbf{T}_3\widetilde{\mathbf{M}}_i$ deve estar na origem e a distância média desses pontos a origem igual, a $\sqrt{3}$. Após a resolução do problema (2.47), como descrito no algoritmo 2.5, a matriz de projeção estimada deve ser transformada para o sistema de coordenadas original.

Em alguns casos especiais, por exemplo, quando um ponto \mathbf{m}_i é o encontro de duas retas paralelas no plano da imagem, um ou mais pontos \mathbf{M}_i estão no infinito. Assim, não é possível encontrar uma homografia que leve o centróide desses pontos à origem. Com isso, a homografia \mathbf{T}_3 não pode ser utilizada.

O método do algoritmo 2.5, apesar de linear e não-iterativo, não é ideal, pois minimiza um critério sem significado físico. Para contornar esse problema, devemos levar em consideração a seguinte observação. Normalmente, as coordenadas dos pontos 3D, \mathbf{M}_i , são obtidas com a ajuda de um gabarito de calibração como o da figura 2.12 na próxima página. Neste caso, dois ou três planos ortogonais, com padrões, cuja forma e dimensão são bem conhecidos, são utilizados para formar um sistema de coordenadas “imaginário” (figura 2.12). Assim, a exatidão das coordenadas dos pontos \mathbf{M}_i depende apenas do processo de manufatura do gabarito de calibração e, de uma forma geral, são conhecidas com extrema precisão. Dessa forma, podemos supor que nossa fonte de erros está nos

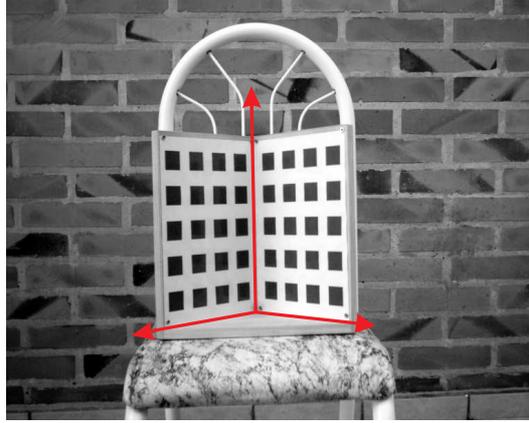


Figura 2.12: Gabarito de calibração com sistema de coordenadas “imaginário” esboçado.

Entrada: $n > 6$ pares $(\mathbf{M}_i, \mathbf{m}_i)$.

Saída: A matriz de projeção da câmera, \mathbf{P} .

- Com os pontos \mathbf{M}_i e \mathbf{m}_i , compute as matrizes \mathbf{T}_3 e \mathbf{T}_2 respectivamente.
- Considerando os pares $(\mathbf{T}_3\mathbf{M}_i, \mathbf{T}_2\mathbf{m}_i)$, estime a matriz, $\hat{\mathbf{P}}$, resolvendo o sistema (2.47) utilizando a proposição B.2 (pág. 121).
- Utilizando $\hat{\mathbf{P}}$ como estimativa inicial, encontre a matriz de projeção, \mathbf{P} , resolvendo o problema

$$\min \sum_{i=1}^n d(\mathbf{T}_2\mathbf{m}_i, \hat{\mathbf{m}}_i)^2,$$

onde $\hat{\mathbf{m}}_i = \mathbf{P}\tilde{\mathbf{X}}_i$ e $\mathbf{X}_i = \mathbf{T}_3\mathbf{M}_i$.

- Faça $\mathbf{P} = \mathbf{T}_2^{-T}\hat{\mathbf{P}}\mathbf{T}_3^{-1}$.

Algoritmo 2.6: Calibração não-linear clássica de câmeras.

pontos \mathbf{m}_i e que uma melhor alternativa para estimar \mathbf{P} é resolver

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

O problema da equação anterior consiste em minimizar a distância Euclidiana do ponto medido, \mathbf{m}_i , ao ponto “real”, $\mathbf{P}\tilde{\mathbf{M}}_i$, com relação a todo o conjunto de n pontos. Evidentemente, esta nova técnica (resumida no algoritmo 2.6) exige o uso de um algoritmo de otimização não-linear que, normalmente, exige uma estimativa inicial de \mathbf{P} que pode ser obtida com o algoritmo 2.5.

Como visto na seção 2.1 (pág. 7), o modelo da câmera utilizado neste trabalho pode ser considerado linear. Contudo, no passado, a maioria das câmeras apresentava distorção radial. Assim, essa característica necessitava ser modelada através de um modelo não-linear. Por isso, é comum encontrar na literatura trabalhos clássicos (veja, por exemplo, o método de Tsai (1987) e os trabalhos de Shih et al. (1996) e Salvi et al. (2002) para duas análises críticas) ou até mesmo alguns poucos trabalhos atuais (como o trabalho de Heikkilä (2000)), que propõem soluções para estimar também

a distorção radial das câmeras. Tais soluções são um pouco diferentes dos métodos discutidos nessa seção. Contudo, o estado da arte atual permite manufaturar câmeras, cujo modelo é muito próximo do linear. De fato, o modelo das câmeras atuais está muito próximo da câmera ideal da equação (2.3) (pág. 9). Assim, a distorção radial é desprezada neste trabalho.

Em algumas aplicações especiais, são utilizadas câmeras cuja principal característica é a distorção radial. Uma dessas aplicações é a vigilância civil, pois, nestes casos, são utilizadas câmeras fixas, mas que necessitam possuir um ângulo de visão elevado. Exemplos de trabalhos com esse tipo de câmeras são os trabalhos de Okamoto Jr e seus associados (CORRÊA et al., 2003; GASPAR et al., 2002; JR. et al., 2001). Contudo, este tipo de aplicação está fora do contexto desse trabalho.

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 gabarito. Assim, tais técnicas fornecem maior flexibilidade ao sistema, pois não necessitam interromper a inspeção visual (para que o gabarito seja posto à frente das câmeras novamente) sempre que os parâmetros da câmera sejam alterados, por exemplo, após a realização de um zoom.

2.5.2 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 projetiva¹², isto é, dada uma correspondência $(\mathbf{m}_1, \mathbf{m}_2)$, 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

$$\mathcal{P} \simeq [\mathbf{I} \quad \mathbf{0}] \quad \text{e} \quad \mathcal{P}' \simeq [\mathbf{H} \quad w\mathbf{e}_2], \quad (2.48)$$

onde (LUONG; VIÉVILLE, 1996)

$$\mathbf{H} = \mathbf{H}_\infty + \mathbf{e}_2\mathbf{w}^T, \quad (2.49)$$

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

$$\mathbf{H} = \left(-\frac{[\tilde{\mathbf{e}}_2]_\times}{\|\tilde{\mathbf{e}}_2\|^2} \right) \mathbf{F}, \quad (2.50)$$

onde $\tilde{\mathbf{e}}_2$ é dado pela proposição 2.4 (pág. 22).

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

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

¹²Do inglês: *weak calibration*

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

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

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

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

Finalmente, comparando a equação anterior com (2.31) completa-se a prova. \square

Apesar de muitos conceitos importantes do espaço Euclidiana não serem válidos em \mathcal{P}^3 , diversos autores, como, por exemplo, (RUF; HORAUD, 2000; HESPANHA et al., 1999; BEARDSLEY et al., 1995; ZELLER; 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,

$$\mathcal{P} = \mathbf{P} \mathbf{T}_{PE} \quad \text{e} \quad \mathcal{P}' = \mathbf{P}' \mathbf{T}_{PE}. \quad (2.51)$$

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

Teorema 2.4 *Dado um conjunto estéreo, 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*

$$\mathbf{T}_{PE} = \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{w}^T & w \end{bmatrix}, \quad (2.52)$$

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

Prova: As equações em (2.13) (pág. 21) e (2.51) podem ser reescritas como

$$\tilde{\mathbf{m}}_1 \simeq \mathcal{P} \mathbf{T}_{PE}^{-1} \mathbf{T}_{PE} \tilde{\mathcal{M}} \quad \text{e} \quad \tilde{\mathbf{m}}_2 \simeq \mathcal{P}' \mathbf{T}_{PE}^{-1} \mathbf{T}_{PE} \tilde{\mathcal{M}}, \quad (2.53)$$

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

Agora, reescrevemos \mathcal{P} na equação (2.48) como sendo

$$\mathcal{P} \simeq [\mathbf{I} \ \mathbf{0}] \simeq [\mathbf{A} \ \mathbf{0}] \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{w}^T & w \end{bmatrix}. \quad (2.54)$$

De forma semelhante, substituindo (2.49) em (2.48), temos

$$\begin{aligned} \mathcal{P}' &\simeq [\mathbf{H}_{\infty} + \tilde{\mathbf{e}}_2 \mathbf{w}^T \quad w \tilde{\mathbf{e}}_2] \\ &\simeq [\mathbf{A}' \mathbf{R} \mathbf{A}^{-1} + \mathbf{A}' \mathbf{t} \mathbf{w}^T \quad w \mathbf{A}' \mathbf{t}] \\ &\simeq [\mathbf{A}' \mathbf{R} \quad \mathbf{A}' \mathbf{t}] \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{w}^T & w \end{bmatrix}, \end{aligned} \quad (2.55)$$

onde utilizou-se a proposição 2.5 (pág. 22) e a equação (A.1) (pág. 117) para substituir \mathbf{H}_∞ .

Agora, substituindo (2.54) e (2.55) em (2.53), temos

$$\tilde{\mathbf{m}}_1 \simeq \mathbf{P} \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{w}^T & w \end{bmatrix} \mathbf{T}_{PE}^{-1} \mathbf{T}_{PE} \tilde{\mathbf{M}} \quad \text{e} \quad \tilde{\mathbf{m}}_2 \simeq \mathbf{P}' \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{w}^T & w \end{bmatrix} \mathbf{T}_{PE}^{-1} \mathbf{T}_{PE} \tilde{\mathbf{M}}. \quad (2.56)$$

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

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

donde obtemos (2.52). □

Apesar de \mathbf{T}_{PE} ser, normalmente, desconhecida, ela pode ser estimada se as coordenadas Euclidianas de alguns pontos do ambiente forem conhecidas ou, como discutido na seção 2.6 (pág. 51), de forma automática.

2.5.3 Calibração Automática

A técnica descrita na seção 2.5.1 (pág. 39), apesar de extremamente exata, pode apenas ser utilizada antes do processo de inspeção visual. Em outras palavras, a técnica não é aplicável em situações nas quais os parâmetros intrínsecos da câmera estão sob constantes mudanças devido a operações de foco e zoom. Esta situação é, particularmente, comum no contexto da “visão ativa”, onde os parâmetros ópticos da câmera sofrem mudanças propositais e constantes. Com isso, surgiu a necessidade de técnicas de calibração que não utilizem gabarito de calibração. Tais técnicas são referenciadas como técnicas de “calibração automática” ou “autocalibração” e são discutidas em detalhes a seguir.

§ Métodos baseados na equação de Faugeras e Huang

A geometria epipolar esboçada na figura 2.10 (pág. 21) não se aplica apenas ao caso de um par estéreo, ou seja, todas as restrições da seção 2.3 (pág. 20) aplicam-se também quando uma única câmera realiza um deslocamento desconhecido e “observa” uma cena estática. Se os parâmetros da câmera antes e após o deslocamento permanecem os mesmos, evidentemente, $\mathbf{A} = \mathbf{A}'$. Então, se imagens forem obtidas antes e após um deslocamento da câmera, é possível estimar a matriz fundamental que relaciona tais imagens. Neste caso, a equação (2.26) depende apenas dos parâmetros intrínsecos da câmera e, se esta for substituída na equação (2.27), temos uma equação independente que é função apenas de tais parâmetros. Desde que \mathbf{A} tem cinco graus de liberdade e (2.27) (pág. 25) fornece apenas duas equações independentes, são necessários três deslocamentos para estimar \mathbf{A} univocamente.

O método descrito no algoritmo 2.7 na próxima página é puramente teórico, pois os polinômios P_i^8 (algoritmo 2.7) têm grau oito. Assim, desde que cada um deles pode ser dividido em duas equações de Trivedi (1988), teoricamente, teríamos um total de seis equações polinomiais de grau oito independentes. Como é extremamente complexo resolver um sistema polinomial com uma equação

Entrada: Não são necessários dados de entrada.

Saída: A matriz \mathbf{A} que reúne os parâmetros intrínsecos da câmera.

- a) Realize três deslocamentos Euclidianos e, em cada uma das quatro posições, capture uma imagem I_i , onde $i = 1, 2, 3$ e 4.
- b) Para cada um dos deslocamentos, calcule \mathbf{F}_{ij} , onde \mathbf{F}_{ij} é a matriz fundamental que relaciona I_i com I_j .
- c) Para cada uma das três \mathbf{F}_{ij} , substitua (2.26) em (2.27) e forme o polinômio $P_i^8(f, \theta_{sk}, n_{x/y}, u_0, v_0)$.
- d) Resolva o sistema de equações formado por $P_1^8 = 0$, $P_2^8 = 0$ e $P_3^8 = 0$.

Algoritmo 2.7: Calibração automática baseada nas restrições da matriz essencial.

redundante, o sistema teria que ser resolvido com apenas cinco equações. Para cada um dos cinco parâmetros intrínsecos, cada equação fornece oito raízes diferentes. Assim, teríamos uma combinação de $8^5 = 32.768$ soluções distintas. Dentre essas, apenas uma é a verdadeira, por fornecer uma matriz \mathbf{A} , fisicamente, plausível.

Devido as dificuldades inerentes ao método, apenas recentemente Mendonça e Cipolla (1999) desenvolveram uma criativa técnica de calibração automática baseada no teorema 2.1 (pág. 25). Nesta, dada uma matriz fundamental, \mathbf{F}_{ij} , que relaciona a imagem I_i (antes do deslocamento) com a imagem I_j (após o deslocamento), baseado na equação (2.26), os autovalores não-nulos de \mathbf{E}_{ij} (σ_{1i} e σ_{2i}), são expressos em função dos elementos de \mathbf{F}_{ij} e dos parâmetros intrínsecos da câmera. Assim, dado $n \geq 3$ deslocamentos, Mendonça e Cipolla (1999) propuseram a seguinte função

$$\mathcal{J}(f, \theta_{sk}, n_{x/y}, u_0, v_0) = \sum_{i=1}^n \left(\frac{\sigma_{1i} - \sigma_{2i}}{\sigma_{2i}} \right)$$

que deve ser minimizada para encontrar-se a matriz \mathbf{A} .

O método proposto tem as vantagens de ter um custo computacional bem menor que o do algoritmo 2.7 e permitir, facilmente, explorar a redundância (utilizando $n > 3$ deslocamentos) para obter-se um resultado melhor. Contudo, os resultados experimentais de Mendonça e Cipolla (1999) mostram que a técnica fornece parâmetros com erros da ordem de 10%.

§ Métodos baseados nas equações de Kruppa

Um conceito bastante utilizado na visão computacional é o cone absoluto, Ω . Este é um cone localizado no plano no infinito, Π_∞ , e corresponde a um círculo de raio $i = \sqrt{-1}$. Além disso, a projeção de Ω no plano de imagem também é um cone e é representado neste trabalho por ω . Assim, um ponto \mathbf{m} pertencente a ω (veja figura 2.13) satisfaz a seguinte equação

$$\tilde{\mathbf{m}}^T \mathbf{B} \tilde{\mathbf{m}} = 0, \quad (2.57)$$

onde \mathbf{B} é uma matriz homogênea 3×3 que possui cinco graus de liberdade.

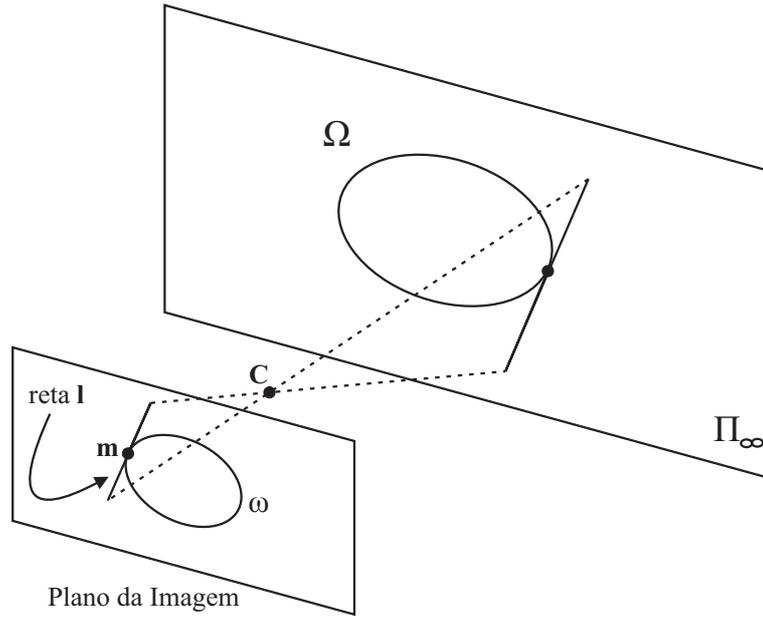


Figura 2.13: Cone absoluto, Ω , no plano no infinito, Π_∞ , e sua projeção, ω , no plano da imagem.

O dual de ω é representado por ω^* . O cone ω^* está relacionado as retas que são tangentes a ω . A reta l tangente a ω no ponto m é dada por $l = B\tilde{m}$. Assim, $\tilde{m} = B^{-1}l$ e $\tilde{m}^T = l^T B^{-T}$. Logo, podemos escrever a equação de ω^* em função de (2.57) como sendo

$$l^T K l = 0, \quad (2.58)$$

onde considerou-se $K = B^{-T}$.

A principal propriedade do cone absoluto é que sua imagem, ω , é invariante a transformações projetivas. Assim, esta depende apenas dos parâmetros intrínsecos da câmera. De fato, pode ser demonstrado (FAUGERAS; MAYBANK, 1990) que

$$B = A^{-T} A^{-1}, \quad \text{ou seja,} \quad K = A A^T. \quad (2.59)$$

A equação anterior nos leva a um resultado interessante que pode ser resumido na proposição a seguir.

Proposição 2.6 *Considerando a matriz do cone dual como sendo*

$$K = \begin{bmatrix} k_1 & k_2 & k_3 \\ k_2 & k_4 & k_5 \\ k_3 & k_5 & k_6 \end{bmatrix}, \quad (2.60)$$

se o parâmetro θ_{sk} da matriz A for nulo, tem-se

$$k_3 k_5 - k_6 k_2 = 0.$$

Prova: Substituindo (2.4) (pág. 9) em (2.59) e fazendo $\theta_{sk} = 0$, o resultado anterior é evidente. \square

A equação (2.59) deu origem a primeira teoria sobre calibração automática de uma câmera em movimento. Esta foi desenvolvida como segue.

No par estéreo da figura 2.14, as projeções do cone absoluto nos planos de imagem I_1 e I_2 são ω_1 e ω_2 respectivamente. Agora, considere dois planos Π_1 e Π_2 tangentes a Ω , ω_1 e ω_2 . Estes planos definem as retas l_{11} e l_{12} na imagem I_1 e as retas l_{21} e l_{22} na imagem I_2 (ver figura 2.14). Estas retas estão relacionadas pela geometria epipolar, ou seja, $l_{11} \leftrightarrow l_{21}$ e $l_{12} \leftrightarrow l_{22}$. Este mesmo raciocínio pode ser aplicado a qualquer plano tangente a Ω , ω_1 e ω_2 . Assim, da equação (2.58) e considerando a reta epipolar $l_{1i} = \tilde{\mathbf{e}}_1 \times \tilde{\mathbf{m}}_{1i}$ definida pelo epipolo \mathbf{e}_1 e por um ponto, \mathbf{m}_{1i} , tangente a ω_1 , temos

$$(\tilde{\mathbf{e}}_1 \times \tilde{\mathbf{m}}_{1i})^T \mathbf{K} (\tilde{\mathbf{e}}_1 \times \tilde{\mathbf{m}}_{1i}) = 0. \quad (2.61)$$

Contudo, $(\tilde{\mathbf{e}}_1 \times \tilde{\mathbf{m}}_{1i}) \leftrightarrow \mathbf{F} \tilde{\mathbf{m}}_{1i}$. Assim, a seguinte equação pode ser encontrada para a imagem I_2

$$\tilde{\mathbf{m}}_{1i}^T \mathbf{F}^T \mathbf{K}' \mathbf{F} \tilde{\mathbf{m}}_{1i} = 0. \quad (2.62)$$

O primeiro passo para resolver o sistema de equações formado pelas equações (2.61) e (2.62) é parametrizar as retas l_{1i} e $l_{2i} = \mathbf{F} \tilde{\mathbf{m}}_{1i}$ com um único parâmetro “ τ_{1i} ”. Isto pode ser feito facilmente se substituirmos \mathbf{m}_{1i} por um ponto, \mathbf{p}_{1i} , que pertence à reta no infinito e à l_{1i} .

Utilizando a proposição A.1 (pág. 116), temos $\tilde{\mathbf{p}}_{1i} = [1, \tau_{1i}, 0]^T$, onde

$$\tau_{1i} = \frac{e_t v_{1i} - e_v t_{1i}}{e_t u_{1i} - e_u t_{1i}}.$$

Agora, considerando (2.60) e substituindo \mathbf{p}_{1i} em (2.61), temos

$$\mathcal{K}_2 \tau_{1i}^2 + \mathcal{K}_1 \tau_{1i} + \mathcal{K}_0 = 0, \quad (2.63)$$

onde

$$\begin{aligned} \mathcal{K}_0 &= e_t^2 k_4 - 2e_t e_v k_5 + e_v^2 k_6 \\ \mathcal{K}_1 &= -2e_t^2 k_2 + 2e_t e_v k_3 + 2e_t e_u k_5 - 2e_u e_v k_6 \\ \mathcal{K}_2 &= e_t^2 k_1 - 2e_t e_u k_3 + e_u^2 k_6. \end{aligned} \quad (2.64)$$

De forma semelhante, considerando

$$\mathbf{K}' = \begin{bmatrix} k'_1 & k'_2 & k'_3 \\ k'_2 & k'_4 & k'_5 \\ k'_3 & k'_5 & k'_6 \end{bmatrix} \quad (2.65)$$

e substituindo \mathbf{p}_{1i} em (2.62), temos

$$\mathcal{K}'_2 \tau_{1i}^2 + \mathcal{K}'_1 \tau_{1i} + \mathcal{K}'_0 = 0, \quad (2.66)$$

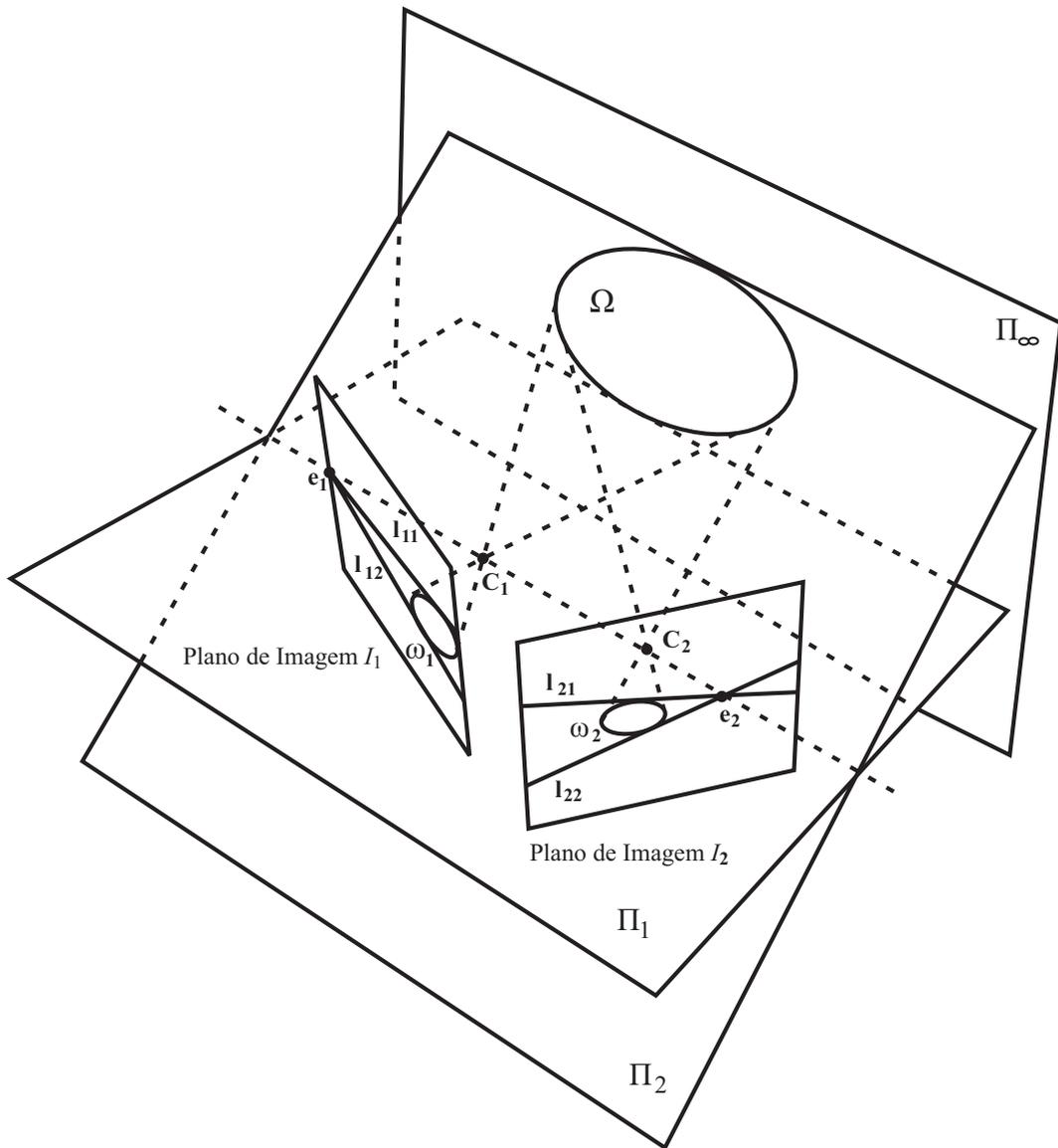


Figura 2.14: Cone absoluto e a geometria epipolar.

onde

$$\begin{aligned}
\mathcal{K}'_0 &= F_1^2 k'_1 + 2F_1 F_4 k'_2 + 2F_1 F_7 k'_3 + F_4^2 k'_4 + 2F_4 F_7 k'_5 + F_7^2 k'_6 \\
\mathcal{K}'_1 &= 2[F_1 F_2 k'_1 + (F_2 F_4 + F_1 F_5) k'_2 + (F_2 F_7 + F_1 F_8) k'_3 + F_4 F_5 k'_4 + \dots \\
&\quad \dots + F_4 F_8 k'_5 + F_7 F_8 k'_6] \\
\mathcal{K}'_2 &= F_2^2 k'_1 + 2F_2 F_5 k'_2 + 2F_2 F_8 k'_3 + F_5^2 k'_4 + 2F_5 F_8 k'_5 + F_8^2 k'_6.
\end{aligned} \tag{2.67}$$

Desde que $(\tilde{\mathbf{e}}_1 \times \tilde{\mathbf{p}}_{1i}) \leftrightarrow \mathbf{F}\tilde{\mathbf{m}}_{1i}$, ou seja, $(\tilde{\mathbf{e}}_1 \times \tilde{\mathbf{p}}_{1i})$ é tangente a ω_1 quando $\mathbf{F}\tilde{\mathbf{m}}_{1i}$ é tangente a ω_2 , as equações (2.63) e (2.66) possuem as mesmas raízes. De fato, (2.63) e (2.66) são iguais a menos de um fator de escala não nulo. Assim,

$$\begin{aligned}
\mathcal{K}_1 \mathcal{K}'_2 - \mathcal{K}_2 \mathcal{K}'_1 &= 0 \\
\mathcal{K}_0 \mathcal{K}'_1 - \mathcal{K}_1 \mathcal{K}'_0 &= 0
\end{aligned} \tag{2.68}$$

As equações em (2.68) são chamadas “equações de Kruppa” (KRUPPA, 1913) e foram introduzidas na visão computacional por Faugeras e Maybank (1990). Elas foram a base para Maybank e Faugeras (1992) criarem a primeira teoria de calibração automática.

As equações em (2.68) são polinômios de grau dois, cujas incógnitas são os elementos da matriz \mathbf{K} e \mathbf{K}' . Por sua vez, estas estão relacionadas, respectivamente, a \mathbf{A} e \mathbf{A}' pela equação (2.59). Novamente, se considerarmos uma única câmera, teremos $\mathbf{A} = \mathbf{A}'$, ou seja, $\mathbf{K} = \mathbf{K}'$, como (2.68) nos fornece apenas duas equações independentes, são necessários três deslocamentos para determinarmos \mathbf{K} univocamente.

Como no caso da seção anterior, um sistema polinomial com uma equação redundante deve ser resolvido. Contudo, agora, se desprezarmos uma equação, teremos cinco equações de grau dois. Assim, existe um total de $2^5 = 32$ soluções distintas, mas apenas uma delas é plausível. Após encontramos a solução para \mathbf{K} , a proposição B.1 (pág. 120) deve ser utilizada para obtermos a matriz \mathbf{A} .

O sistema de equações necessário à solução do problema de Maybank e Faugeras (1992) é bem menos complexo que o exigido pelo algoritmo 2.7. Apesar disso, a resolução de tal sistema continua sendo a principal dificuldade do método. Na verdade, só cinco anos depois do método ser proposto, com o trabalho de Luong e Faugeras (1997), surgiu uma solução adequada ao problema. Mesmo assim, o método não tem muita aplicação prática por vários problemas, alguns dos quais comentados a seguir.

- ✓ O custo computacional do método é extremamente elevado. De fato, muitos autores que utilizam a técnica como, por exemplo, Bougnoux (1998) e Habed e Boufama (2002), supõem modelos de câmeras mais simplificados para reduzir o número de equações necessárias e tornar o problema mais tratável. Contudo, tais simplificações tendem a resultados menos exatos.
- ✓ O método é muito sensível a ruídos e sua exatidão depende extremamente da exatidão da matriz fundamental estimada.

- ✓ Em contraste com o item anterior, a técnica de Luong e Faugeras (1997) não permite inserir redundância para melhorar os resultados.
- ✓ Ao contrário de (2.27), as equações em (2.68) dependem explicitamente das coordenadas do epipolo da matriz fundamental. Infelizmente, principalmente quando o epipolo tende para o infinito, tais coordenadas são difíceis de serem estimadas com exatidão. Esta característica degrada significativamente a exatidão dos parâmetros estimados (LOURAKIS; DERICHE, 1999).

Recentemente, baseado no trabalho de Hartley (1997), Lourakis e Deriche (2000) propuseram uma simplificação as equações (2.68) que aumentaram a estabilidade numérica da técnica por permitir o uso de redundância para lidar com incertezas. Mesmo assim, o problema continua apresentando um custo computacional extremamente elevado e a qualidade dos resultados obtidos não torna o método atraente para o uso prático.

§ Métodos baseados em deslocamentos restritos

Para realizar a calibração, os métodos anteriores assumem que a câmera realiza um movimento genérico. Com o objetivo de reduzir a complexidade do problema, Hartley (1997) propôs uma nova técnica, na qual a câmera realiza um movimento de rotação pura. Neste caso, se duas imagens, I_1 e I_2 , são capturadas em posições distintas da rotação, pode ser demonstrado (HARTLEY, 1997) que correspondências de pontos, $(\mathbf{m}_{1i}, \mathbf{m}_{2i})$, entre tais imagens estão relacionadas pela homografia do infinito (seção A.4), ou seja, $\tilde{\mathbf{m}}_2 \simeq \mathbf{H}_\infty \tilde{\mathbf{m}}_1$. Assim, dadas quatro correspondências de pontos, \mathbf{H}_∞ pode ser determinada univocamente. Em seguida, a matriz \mathbf{A} pode ser obtida de \mathbf{H}_∞ utilizando a proposição A.3 (pág. 118).

O método de Hartley (1997) tem a grande vantagem de utilizar apenas equações lineares. Contudo, para que a câmera realize um movimento de rotação pura, o eixo de rotação deve passar pelo seu centro. Na prática, este centro só é conhecido com exatidão após a calibração. Com isso, o método possui uma fonte de erros inerente e difícil de eliminar.

Existem poucos trabalhos que analisam o impacto de uma translação espúria ao movimento de rotação da câmera durante a fase de calibração. Contudo, os recentes trabalhos de Wang et al. (2004) e Hayman e Murray (2003) fornecem algumas indicações. Segundo estes estudos, uma translação espúria tem um efeito dramático na exatidão da estimação. Wang et al. (2004) sugerem que sejam utilizadas correspondências de pontos que correspondam a projeções de pontos 3D “distantes” da câmera. Além disso, os ângulos do movimento de rotação utilizado na calibração devem ser elevados.

Outro trabalho recente e interessante é o de Ji e Dai (2004). Neste, os autores modificam o método original para levar em consideração uma “pequena” translação no movimento da câmera. Contudo, o método requer que um conjunto de pontos 3D seja rastreado ao longo de três ou mais imagens capturadas de perspectivas bem distintas. Evidentemente, isto dificulta o processo de calibração e cria uma nova fonte de erros.

Na prática, apesar dos métodos de calibração automática terem surgidos para aumentar a flexibilidade da calibração, a calibração automática de uma câmera em rotação reduz tal flexibilidade,

pois limita os movimentos da câmera e restringem as correspondências que podem ser utilizadas no processo. Além disso, seus resultados conservam a inexatidão típica das técnicas da calibração automática.

§2.6 Calibração Automática de um Conjunto Estéreo

Evidentemente, as técnicas de calibração automática discutidas na seção anterior podem ser aplicadas à calibração de um par estéreo. Neste caso, as câmeras devem ser calibradas independentemente. Contudo, em um conjunto estéreo, as imagens dos planos I_1 e I_2 sempre diferem entre si por um deslocamento constante, apesar de possivelmente desconhecido. Isto insere mais restrições ao problema de calibração que, se exploradas, simplificam extremamente as equações. De fato, o problema pode ser formulado usando apenas equações lineares. Neste contexto, duas técnicas mais indicadas à calibração de um conjunto binocular são discutidas em detalhes a seguir.

2.6.1 Autocalibração por Etapas

Na figura 2.15, um par estéreo é movido da posição 1 para a posição 2 por um deslocamento euclidiano. Além disso, o ponto 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 par estéreo, é possível encontrar duas reconstruções 3D para o mesmo ponto M , por exemplo, \mathbf{X}_1 e \mathbf{X}_2 para as posições 1 e 2 respectivamente. Evidentemente, essas reconstruções estão relacionadas por

$$\tilde{\mathbf{X}}_2 \simeq \mathbf{D}_{12} \tilde{\mathbf{X}}_1, \quad (2.69)$$

onde

$$\mathbf{D}_{12} = \begin{bmatrix} \mathbf{R}_{12} & \mathbf{t}_{12} \\ \mathbf{0}^T & 1 \end{bmatrix} \quad (2.70)$$

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

Se o sistema não estiver calibrado e o deslocamento é desconhecido, não é possível realizarmos a reconstrução euclidiana. Contudo, utilizando o teorema 2.3 (pág. 42), é possível encontrarmos as matrizes \mathcal{P} e \mathcal{P}' e realizarmos uma reconstrução projetiva (\mathcal{X}_1 e \mathcal{X}_2). Além disso, pelo teorema 2.4 (pág. 43), sabemos que tal reconstrução relaciona-se com a reconstrução euclidiana por uma transformação projetiva, \mathbf{T}_{PE} , ou seja,

$$\tilde{\mathbf{X}}_1 \simeq \mathbf{T}_{PE} \tilde{\mathcal{X}}_1 \quad \text{e} \quad \tilde{\mathbf{X}}_2 \simeq \mathbf{T}_{PE} \tilde{\mathcal{X}}_2, \quad (2.71)$$

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

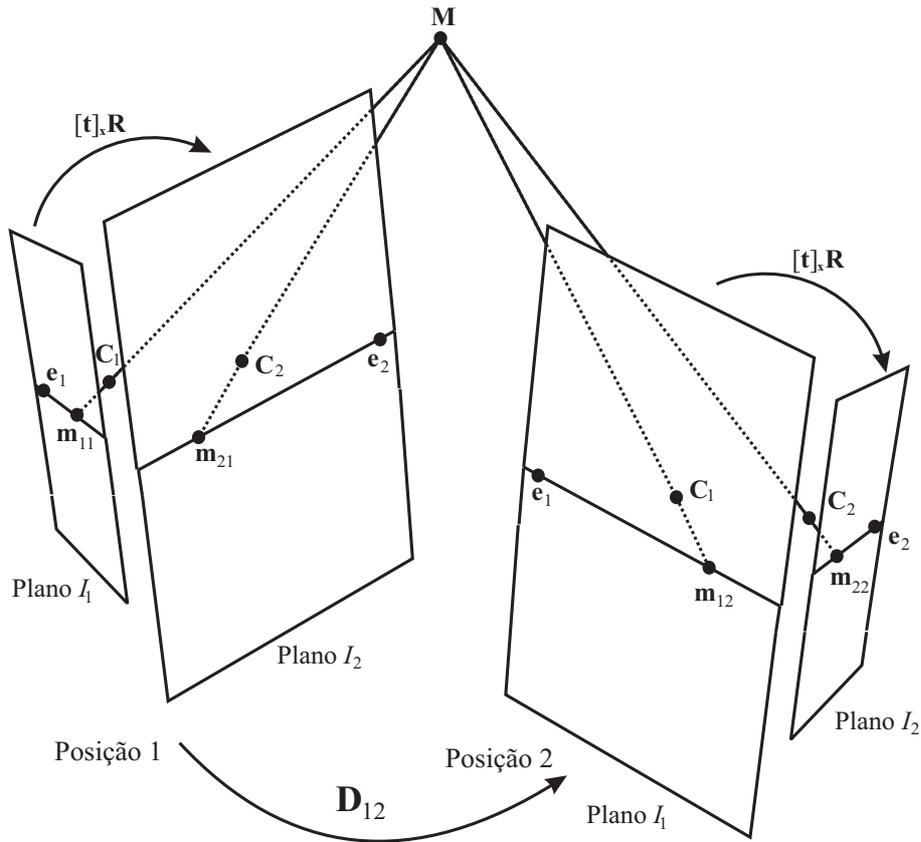


Figura 2.15: Cone absoluto, Ω , no plano no infinito, Π_∞ , e sua projeção, ω , no plano da imagem.

De (2.69) e (2.71), segue que \mathcal{X}_1 e \mathcal{X}_2 relacionam-se por

$$\tilde{\mathcal{X}}_2 \simeq \mathbf{H}_{12} \tilde{\mathcal{X}}_1, \quad (2.72)$$

onde

$$\mathbf{H}_{12} = s_\alpha \mathbf{T}_{PE}^{-1} \mathbf{D}_{12} \mathbf{T}_{PE}, \quad (2.73)$$

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

A equação (2.72) 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 par estéreo. De fato, se apenas \mathbf{H}_{12} e a matriz fundamental do conjunto estéreo 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

A constante s_α em (2.73) pode ser conhecida usando-se o fato que determinantes e traços são invariantes a transformações de similaridade como é o caso da transformação \mathbf{T}_{PE} . Antes, porém, o formato da matriz \mathbf{D}_{12} deve ser conhecido. Isso pode ser feito considerando que sempre é possível encontrar uma transformação de similaridade que transforme um deslocamento euclidiano qualquer (conhecido como movimento genérico¹³) em um movimento plano¹⁴ (equivalente a uma rotação em torno do eixo do movimento¹⁵ e uma translação perpendicular a esse mesmo eixo (HARTLEY; ZISSERMAN, 2000, Sec. 2.4.1)). Assim, sem perda de generalidade, podemos considerar

$$\mathbf{D}_{12} = \begin{bmatrix} \cos \theta_{12} & -\text{sen } \theta_{12} & 0 & 0 \\ \text{sen } \theta_{12} & \cos \theta_{12} & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (2.74)$$

onde θ_{12} é o ângulo de rotação em torno do eixo do movimento.

Agora, de (2.73) e (2.74), temos

$$\text{trace}(\mathbf{H}_{12}) = \text{trace}(s_\alpha \mathbf{D}_{12}) = 2s_\alpha(1 + \cos \theta_{12}) \quad (2.75)$$

$$\det(\mathbf{H}_{12}) = \det(s_\alpha \mathbf{D}_{12}) = s_\alpha^4. \quad (2.76)$$

De (2.76), $|s_\alpha| = \sqrt[4]{\det(\mathbf{H}_{12})}$. Assim, desde que $(1 + \cos \theta) \geq 0$ em (2.75), temos

$$s_\alpha = \text{sign}(\text{trace}(\mathbf{H}_{12})) \sqrt[4]{\det(\mathbf{H}_{12})}. \quad (2.77)$$

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

$$\mathbf{H}_{12} \simeq \frac{1}{s_\alpha} \mathbf{H}_{12} = \mathbf{T}_{PE}^{-1} \mathbf{D}_{12} \mathbf{T}_{PE}. \quad (2.78)$$

§ Encontrando o plano no infinito

Na autocalibração por etapas, precisamos 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 2.4 (pág. 43), 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 *Em um conjunto estéreo, 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.*

¹³Do inglês: *geral motion*.

¹⁴Do inglês: *planar motion*.

¹⁵Do inglês: *screw axis*.

A prova do teorema anterior é um pouco extensa e requer a definição dos lemas a seguir.

Lema 2.5.1 *Se \mathbf{T} é uma transformação de pontos entre dois sistemas de coordenadas, \mathbf{T}^{-T} é uma transformação de planos.*

Prova: Considere \mathbf{q} um plano do espaço \mathcal{P}^3 . Logo, um ponto \mathbf{M} pertencente a este plano satisfaz $\mathbf{q}^T \widetilde{\mathbf{M}} = 0$. Aplicando uma transformação, \mathbf{T} , a \mathbf{M} , temos $\widetilde{\mathbf{M}}' \simeq \mathbf{T} \widetilde{\mathbf{M}}$, ou seja, \mathbf{T} leva \mathbf{M} a outro sistema de coordenadas. Nesta nova referência, se considerarmos a equação do plano ao qual \mathbf{M}' pertence como sendo \mathbf{q}' , teremos $\mathbf{q}'^T \widetilde{\mathbf{M}}' = 0$. A equação anterior pode ser reescrita como sendo $\mathbf{q}'^T \mathbf{T} \widetilde{\mathbf{M}} = 0$, ou seja, $\mathbf{q}' \simeq \mathbf{T}^{-T} \mathbf{q}$. Isto conclui a demonstração. \square

O lema anterior permite provar o lema a seguir.

Lema 2.5.2 *Se \mathbf{D}_{12} descreve um deslocamento de um conjunto estéreo, o autovetor de \mathbf{D}_{12}^{-T} que corresponde ao autovalor unitário é o plano no infinito.*

Prova: Do lema 2.5.1 e da equação (2.78), temos que a transformação de planos entre a referência da posição 1 e a da posição 2 (figura 2.15) é dada por

$$\mathbf{H}_{12}^{-T} = \mathbf{T}_{PE}^T \mathbf{D}_{12}^{-T} \mathbf{T}_{PE}^{-T}. \quad (2.79)$$

Desde que \mathbf{D}_{12} é dado por (2.70), 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

$$\mathbf{D}_{12}^{-T} \mathbf{V}_{Di} = \mathbf{V}_{Di}. \quad (2.80)$$

Como, de (2.74),

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

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

$$\mathbf{R}_{12} \mathbf{v} = \mathbf{v} \quad (2.81)$$

$$-\mathbf{t}_{12}^T \mathbf{R}_{12} \mathbf{v} + v = v. \quad (2.82)$$

A equação (2.82) nos leva a seguinte restrição

$$\mathbf{t}_{12}^T \mathbf{R}_{12} \mathbf{v} = 0. \quad (2.83)$$

Como, da definição de autovetor, $\mathbf{V}_{Di} \neq \mathbf{0}$, o sistema de equações formado por (2.81) e (2.83) 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 Se \mathbf{V}_{T_i} é um autovetor que corresponde ao autovalor unitário de \mathbf{H}_{12}^{-T} e \mathbf{V}_{D_i} é o um autovetor que corresponde ao autovalor unitário de \mathbf{D}_{12}^{-T} , temos

$$\mathbf{V}_{T_i} = \mathbf{T}_{PE}^T \mathbf{V}_{D_i}.$$

Prova: Explicitando \mathbf{D}_{12}^{-T} na equação (2.79) e substituindo em (2.80), temos

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

□

A prova do lema a seguir conclui a prova do teorema 2.5.

Lema 2.5.4 Em um conjunto estéreo, 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 a equação do plano no infinito, Π_∞ , no sistema de coordenadas definido por \mathcal{P} e \mathcal{P}' .

Prova: Do lema 2.5.2, sabe-se que $\mathbf{V}_{D_1} = [\mathbf{0}^T \ 1]^T$ é um autovetor que corresponde ao autovalor unitário de \mathbf{D}_{12}^{-T} . Além disso, do lema anterior, \mathbf{V}_{D_1} no sistema de coordenadas de \mathbf{H}_{12}^{-T} é

$$\mathbf{V}_{T_1} = \begin{bmatrix} \mathbf{A}^{-T} & \mathbf{w} \\ \mathbf{0} & w \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = [\mathbf{w}^T, w]^T = \mathbf{W}_\infty^T.$$

□

Como demonstrado, 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

$$\begin{bmatrix} \mathbf{H}_{12}^T - \mathbf{I} \\ \mathbf{H}_{23}^T - \mathbf{I} \\ \vdots \\ \mathbf{H}_{n(n+1)}^T - \mathbf{I} \end{bmatrix} \mathbf{W}_\infty = \mathbf{0} \quad (2.84)$$

usando a proposição B.2 (pág. 121).

Se \mathbf{D}_{12} corresponde a um movimento planar, pela demonstração do lema 2.5.2, \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 par estéreo realiza $n \geq 2$ movimentos genéricos, temos $\mathbf{H}_{12}^T, \mathbf{H}_{23}^T, \dots, \mathbf{H}_{n(n+1)}^T$. 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 (2.84) com $n \geq 2$.

§ Estimando a homografia do infinito

É possível estabelecer uma homografia do infinito (pág. 117), representada por $\mathbf{H}_{\infty 12}$, entre o plano de imagem I_1 na posição 1 e I_1 na posição 2 (figura 2.15). Se $\mathbf{H}_{\infty 12}$ for conhecida, é possível estimar a matriz \mathbf{A} (pág. 9) com a proposição A.3 (pág. 118).

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

Proposição 2.7 Considerando $\mathbf{W}_\infty = [\mathbf{w}^T \ w]^T$ e a homografia \mathbf{H}_{12} dada por

$$\mathbf{H}_{12} = \begin{bmatrix} \mathbf{H}_{12}(1:3, 1:3) & \mathbf{H}_{12}(1:3, 4) \\ \mathbf{H}_{12}(4, 1:3) & \mathbf{H}_{12}(4, 4) \end{bmatrix}, \quad (2.85)$$

conhecidos, temos 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 pelas (2.52) e (2.70) respectivamente, podemos re-escrever (2.78) como sendo

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

Comparando a equação anterior com (2.85) 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, 1997; ZISSERMAN et

al., 1995). Por isso, na prática, devemos realizar “diversos” deslocamentos, $\mathbf{D}_{n(n+1)}$, e usar a redundância para melhorar a estimação de \mathbf{A} . Neste caso, a proposição A.4 (pág. 118) deve ser utilizado nesta estimação.

Uma vez que \mathbf{W}_∞ e \mathbf{A} são conhecidos, o teorema 2.4 (pág. 43) 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.

2.6.2 Estimação de Todos os Parâmetros Euclidianos

O método anterior fornece apenas as matrizes de projeção das câmeras. Assim, se todos os parâmetros Euclidianos que compõem o par estereo forem necessários, mais alguma computação é requerida. Geralmente, seria preciso estimar a homografia do infinito que relaciona as imagens referentes ao plano I_2 (figura 2.15). Assim, a matriz \mathbf{A}' seria conhecida e a relação $\mathbf{E} = \mathbf{A}'\mathbf{F}\mathbf{A}$ utilizada para conhecermos \mathbf{R} e \mathbf{t} , após a decomposição de \mathbf{E} . Contudo, neste caso, os parâmetros obtidos são pouco confiáveis, pois a matriz \mathbf{E} obtida dessa forma, geralmente, satisfaz fracamente o lema 2.1 (pág. 25).

Um método alternativo é estimar todos os parâmetros Euclidianos ao mesmo tempo. A idéia básica é considerar a equação (2.38) (pág. 31), mas utilizar (2.16) (pág. 22) para parametrizar a matriz fundamental. Novamente, como ilustrado na figura 2.16, o conjunto estereo deve realizar deslocamentos de uma posição j para outra $(j + 1)$. A cada deslocamento (indexado por j), associamos n_j correspondências ($\mathbf{m}_{1(j+1)i} \leftrightarrow \mathbf{m}_{2(j+1)i}$) entre as imagens I_1 e I_2 (ambas na posição $(j + 1)$), p_j correspondências ($\mathbf{m}_{1ji} \leftrightarrow \mathbf{m}_{1(j+1)i}$) entre as imagens I_1 na posição j e I_1 na posição $(j + 1)$, e q_j correspondências ($\mathbf{m}_{2ji} \leftrightarrow \mathbf{m}_{2(j+1)i}$) entre as imagens I_2 na posição j e I_2 na posição $(j + 1)$. Dessa forma, para cada deslocamento podemos definir uma função \mathcal{F}_j como sendo

$$\begin{aligned} \mathcal{F}_j = & \sum_{i=1}^{n_j} [d^2(\tilde{\mathbf{m}}_{2(j+1)i}, \mathbf{F}_j \tilde{\mathbf{m}}_{1(j+1)i}) + d^2(\tilde{\mathbf{m}}_{1(j+1)i}, \mathbf{F}_j^T \tilde{\mathbf{m}}_{2(j+1)i})] + \\ & \sum_{i=1}^{p_j} [d^2(\tilde{\mathbf{m}}_{1(j+1)i}, \mathbf{F}_{1j} \tilde{\mathbf{m}}_{1ji}) + d^2(\tilde{\mathbf{m}}_{1ji}, \mathbf{F}_{1j}^T \tilde{\mathbf{m}}_{1(j+1)i})] + \\ & \sum_{i=1}^{q_j} [d^2(\tilde{\mathbf{m}}_{2(j+1)i}, \mathbf{F}_{2j} \tilde{\mathbf{m}}_{2ji}) + d^2(\tilde{\mathbf{m}}_{2ji}, \mathbf{F}_{2j}^T \tilde{\mathbf{m}}_{2(j+1)i})], \end{aligned} \quad (2.86)$$

onde $\mathbf{F}_j = \mathbf{A}'^{-T}[\mathbf{t}_j]_{\times} \mathbf{R}_j \mathbf{A}^{-1}$, $\mathbf{F}_{1j} = \mathbf{A}^{-T}[\mathbf{t}_{1j}]_{\times} \mathbf{R}_{1j} \mathbf{A}^{-1}$ e $\mathbf{F}_{2j} = \mathbf{A}'^{-T}[\mathbf{t}_{2j}]_{\times} \mathbf{R}_{2j} \mathbf{A}'^{-1}$.

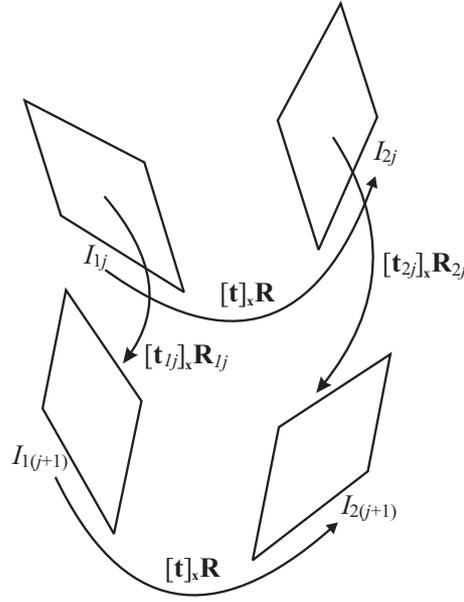


Figura 2.16: Esboço de um deslocamento (indexado por j) de um conjunto estereó.

Desde que cada matriz fundamental possui sete graus de liberdade, dado N deslocamentos j , a função \mathcal{F}_j fornece um total de $7 + 14N$ equações independentes. Por outro lado, a cada deslocamento o número de parâmetros desconhecidos também aumenta. Contudo, deve-se notar que $\mathbf{E}_{1j} = [t_{1j}]_x \mathbf{R}_{1j}$ e $\mathbf{E}_{2j} = [t_{2j}]_x \mathbf{R}_{2j}$ estão relacionados por $\mathbf{E} = [t]_x \mathbf{R}$. De fato, observando a figura 2.16 e utilizando álgebra simples, temos

$$\begin{aligned} \mathbf{R}_{2j} &= \mathbf{R} \mathbf{R}_{1j} \mathbf{R}^T \\ \mathbf{t}_{2j} &= \mathbf{t} + \mathbf{R} \mathbf{t}_{1j} - \mathbf{R}_{1j} \mathbf{t}. \end{aligned} \quad (2.87)$$

Assim, temos um total de $15 + 6N$ incógnitas para N deslocamentos. Este resultado mostra que, com N deslocamentos, é possível estimar todos os parâmetros resolvendo

$$\min \sum_{j=1}^N \mathcal{F}_j. \quad (2.88)$$

Com um único deslocamento, o número de equações independentes é igual ao número de incógnitas. Apesar de, teoricamente, ser possível resolver o problema (2.88) nesta condição, na prática, o problema torna-se instável e difícil de ser tratado (ZHANG et al., 1996). Dessa forma, $N \geq 2$ deslocamentos são exigidos.

Apesar de exigir poucos deslocamentos, o método possui algumas desvantagens apresentadas a seguir.

- ✓ A resolução do problema (2.88) exige uma estimativa inicial de todos os parâmetros Euclidianos. Estes precisam ser obtidos através de outro método de calibração automática, por exemplo, o método de calibração por etapas.

- ✓ O método exige que sejam estabelecidas correspondências ao longo de $3N$ pares de imagens. A técnica utilizada para isto deve ser robusta, pois os métodos utilizados para resolver (2.88), geralmente, não podem lidar com falsas correspondências.
- ✓ Os itens anteriores, somados ao grande número de incógnitas envolvidas na resolução do problema, torna o seu custo computacional muito elevado.
- ✓ Na prática, a restrição embutida na equação (2.87) é difícil de ser incluída na resolução do problema. Evidentemente, isto pode tornar incoerentes alguns dos parâmetros estimados.

Apesar dos problemas apresentados, com exceção do ponto central, o trabalho de Faugeras e Luong (2001) indica que os parâmetros intrínsecos das câmeras são estimados com exatidão.

§ Outros Métodos

De uma forma ou de outra, a grande maioria dos métodos de calibração de um conjunto estéreo são baseadas nos dois métodos anteriores. Por exemplo, Matabosch et al. (2003) substituíram a distância dos pontos às retas epipolares de seus correspondentes na equação (2.86) por outro critério baseado na simplificação de Hartley (1997) das equações de Kruppa. Contudo, devido a complexidade embutida nas equações de Kruppa, um modelo de câmera mais simples teve que ser adotado. Abordagem semelhante foi utilizada por Enciso e Vieville (1997) e Zhang et al. (1996), mas, nestes casos, a equação fundamental (2.17) foi o critério a ser minimizado.

Diversos autores também já propuseram outras versões para a calibração automática por etapas. Devernay e Faugeras (1996) propuseram um método não-linear para estimar os parâmetros Euclidianos de \mathbf{T}_{PE} . Além disso, considerando o fato de que \mathbf{H}_{12} não pode ser estimada quando só estão disponíveis projeções de pontos 3D que pertencem a um mesmo plano, Demirdjian et al. (2000) utilizaram a imagem do cone absoluto, ω , para estimar a equação do plano no infinito e a matriz \mathbf{K} neste caso específico. Já Csurka et al. (1998) propuseram um interessante método de estimar \mathbf{A} a partir de \mathbf{H}_{∞} . Tal método também permite detectar deslocamentos críticos, \mathbf{D}_{12} , que podem tornar a estimação de \mathbf{T}_{PE} inexata.

§2.7 Reconstrução

Uma vez que o par estéreo está calibrado, é possível conhecermos as coordenadas de um ponto \mathbf{M} , dado a correspondência $(\mathbf{m}_1, \mathbf{m}_2)$ equivalente. Este processo é conhecido como reconstrução 3D.

Como a geometria epipolar pode ser conhecida facilmente a partir das matrizes de projeção das duas câmeras, a busca pelo correspondente de um ponto pode ser realizada sobre a sua reta epipolar. Contudo, apesar do espaço de busca reduzido, o problema é extremamente complexo. Por isso, é

comum obter-se um par $(\mathbf{m}_1, \mathbf{m}_2)$ com ruído ou até mesmo uma falsa correspondência. Neste caso, os raios ópticos devidos a \mathbf{m}_1 e \mathbf{m}_2 não se interceptam em nenhum ponto do espaço. Com isso, como estes raios estão no espaço \mathcal{P}^3 (que não conserva nenhuma informação sobre comprimentos ou ângulos), a estimação do ponto \mathbf{M} nestas condições não é trivial. Contudo, nas próximas seções, são discutidos dois métodos que realizam tal estimação.

2.7.1 Método Linear

Dadas duas matrizes de projeção, \mathbf{P} e \mathbf{P}' , e uma correspondência $(\mathbf{m}_1, \mathbf{m}_2)$, podemos reescrever a equação (2.12) (pág. 21) como

$$s_1 \tilde{\mathbf{m}}_1 = \mathbf{P}\tilde{\mathbf{M}} \quad \text{e} \quad s_2 \tilde{\mathbf{m}}_2 = \mathbf{P}'\tilde{\mathbf{M}}, \quad (2.89)$$

onde s_1 e s_2 são constantes desconhecidas e não-nulas. Agora, considerando $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3]^T$ e $\mathbf{P}' = [\mathbf{p}'_1, \mathbf{p}'_2, \mathbf{p}'_3]^T$, podemos eliminar as constantes s_1 e s_2 , pois

$$s_1 = \mathbf{p}_3^T \tilde{\mathbf{M}} \quad \text{e} \quad s_2 = \mathbf{p}'_3^T \tilde{\mathbf{M}}. \quad (2.90)$$

Logo, substituindo (2.90) em (2.89), podemos escrever

$$\mathbf{U}\tilde{\mathbf{M}} = \mathbf{0}, \quad (2.91)$$

onde $\mathbf{U} \in \mathbb{R}^{4 \times 4}$ e é dada por

$$\mathbf{U} = [\mathbf{p}_1 - u_1 \mathbf{p}_3, \mathbf{p}_2 - v_1 \mathbf{p}_3, \mathbf{p}'_1 - u_2 \mathbf{p}'_3, \mathbf{p}'_2 - v_2 \mathbf{p}'_3]^T,$$

onde consideramos $\tilde{\mathbf{m}} = [u_1, v_1, 1]^T$ e $\tilde{\mathbf{m}}' = [u_2, v_2, 1]^T$ na equação (2.89).

A solução de (2.91) é dada pela proposição B.2 (pág. 121). Contudo, (2.91) fornece um resultado ótimo apenas se a correspondência $(\mathbf{m}_1, \mathbf{m}_2)$ satisfaz a equação (2.15) (pág. 21) exatamente. Contudo, ruídos nas coordenadas dos pontos da correspondência podem comprometer a exatidão da estimação de \mathbf{M} através da equação (2.91). Neste caso, o método da próxima seção é mais indicado.

2.7.2 Método Ótimo

O método da seção anterior não é ideal, pois minimiza um critério sem significado físico. Contudo, o critério a ser minimizado não pode pertencer ao espaço \mathcal{P}^3 , pois conceitos como distâncias e perpendicularidade não são válidos no contexto da geometria projetiva. Além disso, o ruído existente no processo ocorre, originalmente, nas coordenadas dos pontos da correspondência $(\mathbf{m}_1, \mathbf{m}_2)$ utilizada na reconstrução, pois, como discutido no início da seção, os raios ópticos só se encontram em \mathcal{P}^3 se a restrição fundamental for satisfeita. Isto significa que devemos substituir a correspondência original pelo par $(\hat{\mathbf{m}}_1, \hat{\mathbf{m}}_2)$, composto pelos pontos mais próximos de \mathbf{m}_1 e \mathbf{m}_2 , mas que satisfazem

a restrição fundamental exatamente. Assim, dada uma correspondência $(\mathbf{m}_1, \mathbf{m}_2)$ e a matriz fundamental, \mathbf{F} , o processo de reconstrução deve consistir, inicialmente, da busca pelos pontos $\hat{\mathbf{m}}_1$ e $\hat{\mathbf{m}}_2$ que minimizam

$$d^2(\mathbf{m}_1, \hat{\mathbf{m}}_1) + d^2(\mathbf{m}_2, \hat{\mathbf{m}}_2), \quad \text{sujeito a } \tilde{\mathbf{m}}_2^T \mathbf{F} \tilde{\mathbf{m}}_1. \quad (2.92)$$

Dada uma reta epipolar $\hat{\mathbf{l}}_1$, o ponto $\hat{\mathbf{m}}_1$, sobre a reta $\hat{\mathbf{l}}_1$, mais próximo de \mathbf{m}_1 é a projeção ortogonal de \mathbf{m}_1 em $\hat{\mathbf{l}}_1$. Assim, o problema da equação anterior pode ser reformulado no seguinte problema irrestrito: buscar $\hat{\mathbf{l}}_1$ e $\hat{\mathbf{l}}_2$ que minimizam

$$d(\mathbf{m}_1, \hat{\mathbf{l}}_1)^2 + d(\mathbf{m}_2, \hat{\mathbf{l}}_2)^2. \quad (2.93)$$

Para resolver este problema, temos que parametrizar as retas $\hat{\mathbf{l}}_1$ e $\hat{\mathbf{l}}_2$ com um único parâmetro “ λ ”. Para isto, o primeiro passo é encontrar duas transformações, \mathbf{H}_{1t} e \mathbf{H}_{2t} , tal que

$$\begin{aligned} \tilde{\mathbf{m}}_{1t} = \tilde{\mathbf{m}}_{2t} &= [0, 0, 1]^T \simeq \mathbf{H}_{1t} \tilde{\mathbf{m}}_1 \simeq \mathbf{H}_{2t} \tilde{\mathbf{m}}_2, \\ [1, 0, \alpha_1]^T &\simeq \mathbf{H}_{1t} \tilde{\mathbf{e}}_1, \mathbf{e}; \\ [1, 0, \alpha_2]^T &\simeq \mathbf{H}_{2t} \tilde{\mathbf{e}}_2. \end{aligned}$$

Dado $\tilde{\mathbf{m}}_x = [u_x, v_x, 1]^T$ e $\tilde{\mathbf{e}}_x = [e_{xu}, e_{xv}, e_{xt}]^T$, a transformação \mathbf{H}_{xt} correspondente pode ser fatorada em uma translação, \mathbf{T}_{xt} , e uma rotação, \mathbf{R}_{xt} , ou seja,

$$\mathbf{H}_{xt} \simeq \mathbf{R}_{xt} \mathbf{T}_{xt}. \quad (2.94)$$

Assim, é fácil ver que

$$\mathbf{T}_{xt} = \begin{bmatrix} 1 & 0 & -u_x \\ 0 & 1 & -v_x \\ 0 & 0 & 1 \end{bmatrix}$$

translada o ponto \mathbf{m}_x para a origem. Além disso,

$$\mathbf{R}_{xt} = \begin{bmatrix} \cos \theta_{xt} & -\text{sen } \theta_{xt} & 0 \\ \text{sen } \theta_{xt} & \cos \theta_{xt} & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

onde θ_{xt} é obtido considerando $\mathbf{R}_{xt} \mathbf{T}_{xt} \tilde{\mathbf{e}}_{xt} = [1, 0, \alpha_x]^T$, que nos leva a

$$\theta_{xt} = \arctan \left(\frac{e_{xt} v_x - e_{xv}}{e_{xu} - e_{xt} u_x} \right). \quad (2.95)$$

Uma vez que os pontos \mathbf{m}_1 e \mathbf{m}_2 tenham sido transformados, a matriz fundamental, \mathbf{F} , deve ser adaptada. Isto é realizado considerando uma nova matriz fundamental, \mathbf{F}_t , tal que

$$\mathbf{F}_t = \mathbf{H}_{2t}^{-T} \mathbf{F} \mathbf{H}_{1t}^{-1}, \quad (2.96)$$

onde a matriz \mathbf{F}_t tem uma forma especial dada por

$$\mathbf{F}_t = \begin{bmatrix} \alpha_1 \alpha_2 d & \alpha_2 c & -\alpha_2 d \\ -\alpha_1 b & a & b \\ -\alpha_1 d & c & d \end{bmatrix}.$$

Agora, considere $\hat{\mathbf{I}}_1(\lambda)$ igual a reta que passa pelo epipolo $[1, 0, \alpha_1]^T$ e pelo ponto $[0, \lambda, 1]^T$. Assim, $\hat{\mathbf{I}}_1(\lambda) = [1, 0, \alpha_1]^T \times [0, \lambda, 1]^T = [\lambda \alpha_1, 1, -\lambda]^T$. Além disso,

$$d(\mathbf{m}_{1t}, \hat{\mathbf{I}}_1(\lambda))^2 = \frac{\lambda^2}{1 + (\lambda \alpha_1)^2}$$

A matriz fundamental, \mathbf{F}_t , pode ser utilizada para obter-se $\hat{\mathbf{I}}_2(\lambda)$, ou seja, $\hat{\mathbf{I}}_2(\lambda) = \mathbf{F}_t [0, \lambda, 1]^T = [-\alpha_2(c\lambda + d), a\lambda + b, c\lambda + d]^T$. Logo,

$$d(\mathbf{m}_{2t}, \hat{\mathbf{I}}_2(\lambda))^2 = \frac{(c\lambda + d)^2}{(a\lambda + b)^2 + \alpha_2^2(c\lambda + d)^2}.$$

Com isso, o problema (2.93) pode ser resolvido encontrando o mínimo do polinômio

$$p(\lambda) = \frac{\lambda^2}{1 + (\lambda \alpha_1)^2} + \frac{(c\lambda + d)^2}{(a\lambda + b)^2 + \alpha_2^2(c\lambda + d)^2}. \quad (2.97)$$

Este mínimo ocorre em uma das raízes do polinômio

$$d(\lambda) = \lambda[(a\lambda + b)^2 + \alpha_2^2(c\lambda + d)^2]^2 - (ad - bc)(1 + (\alpha_1 \lambda)^2)^2(a\lambda + b)(c\lambda + d). \quad (2.98)$$

O polinômio $d(\lambda)$ tem grau seis. Assim, podemos ter até seis possíveis soluções reais para o problema. Dentre essas, evidentemente, a escolhida deve ser aquela que minimiza $p(\lambda)$.

Como resumido no algoritmo 2.8 na página seguinte, após encontrar as raízes do polinômio, o algoritmo da seção anterior deve ser utilizado para realizar a reconstrução.

A desvantagem do método descrito é que este deposita extrema confiança nos epipolos extraídos da matriz fundamental. Apesar de serem difíceis de estimar com exatidão, pois são muito susceptíveis a ruído, estes são utilizados em diversos estágios do método. Além disso, apesar de não-iterativo, o método necessita encontrar todas as raízes de um polinômio de grau seis sempre que a reconstrução de um único ponto é necessária. Evidentemente, isto onera significativamente o processo. Apesar disso, o método é bastante utilizado na literatura, principalmente devido as características do espaço \mathcal{P}^3 que restringem os critérios que podem ser minimizados.

Entrada: Uma correspondência $(\mathbf{m}_1, \mathbf{m}_2)$ e um par de matrizes de projeção \mathbf{P} e \mathbf{P}' .

Saída: O ponto 3D M que corresponde ao par $(\mathbf{m}_1, \mathbf{m}_2)$.

- a) Compute a matriz fundamental, \mathbf{F} , usando o teorema 2.2 (pág. 26).
- b) Estime os epipolos, \mathbf{e}_1 e \mathbf{e}_2 , usando a proposição 2.4 (pág. 22).
- c) Compute \mathbf{H}_{1t} e \mathbf{H}_{2t} usando a equação (2.94).
- d) Faça $\mathbf{F}_t = \mathbf{H}_{2t}^{-T} \mathbf{F} \mathbf{H}_{1t}^{-1}$.
- e) Encontre a raiz, λ_{\min} , de (2.98) que minimiza (2.97).
- f) Compute os pontos $\hat{\mathbf{m}}_1$ e $\hat{\mathbf{m}}_2$ que pertencem, respectivamente, às retas $\hat{\mathbf{l}}_1 = [\lambda_{\min} \alpha_1, 1, -\lambda_{\min}]^T$ e $\hat{\mathbf{l}}_2 = [-\alpha_2(c\lambda_{\min} + d), a\lambda_{\min} + b, c\lambda_{\min} + d]^T$ e estão mais próximos da origem.
- g) Use a correspondência $(\mathbf{H}_{1t}^{-1} \hat{\mathbf{m}}_1, \mathbf{H}_{2t}^{-1} \hat{\mathbf{m}}_2)$ e o método da seção 2.7.1 (pág. 60) para estimar o ponto M .

Algoritmo 2.8: Método ótimo de reconstrução 3D.

CONTRIBUIÇÕES À CALIBRAÇÃO DE UM CONJUNTO BINOCULAR

Neste capítulo são apresentadas as principais contribuições do trabalho. Em especial, são descritos em detalhes um novo método não-linear para cálculo da matriz fundamental, além de diversos métodos de calibração de um conjunto binocular com um bastão graduado. No próximo capítulo, o desempenho desses métodos é comparado aos de outros algoritmos bem conhecidos.

§3.1 Correspondência de pontos na ausência de calibração

Devido ao custo computacional elevado e a pouca qualidade da calibração obtida com os métodos automáticos, a calibração clássica ainda é muito mais utilizada. Assim, na maioria dos trabalhos, a geometria epipolar pode ser facilmente recuperada a partir das matrizes de projeção das câmeras. Por essa razão, pode-se encontrar diversas técnicas de correspondência de pontos que supõem que o sistema já esteja calibrado. De fato, recentemente, duas excelentes análises críticas sobre o assunto foram realizadas por Brown et al. (2003) e Scharstein e Szeliski (2002). Por outro lado, a correspondência de pontos na ausência de calibração é um tópico de pesquisa menos ativo. Este fato deve mudar apenas quando os métodos de calibração automática tornarem-se mais populares.

Como discutido na seção 3.4, o conhecimento da matriz fundamental é imprescindível para a técnica de calibração proposta. Assim, neste trabalho, o objetivo inicial é estimar tal matriz. Para isto, um conjunto de correspondência de pontos é necessário. Por isso, um método que estabelece correspondências de pontos na ausência de calibração é proposto. Este é dividido em três etapas. Na primeira, é usada correlação para formar um conjunto de pares de pontos candidatos à correspondência. Em seguida, a ambigüidade é eliminada atribuindo um grau de confiança para cada par candidato e eliminando correspondências pouco confiáveis. Por último, a matriz fundamental é estimada robustamente e a restrição epipolar utilizada para eliminar as falsas correspondências.

Nas próximas seções, o algoritmo é descrito em detalhes.

3.1.1 Conjunto de candidatos à correspondência

A restrição de semelhança é utilizada para criar um conjunto inicial de candidatos à correspondência. Contudo, tal restrição é muito sensível a distorções e oclusões. Por isso, algumas técnicas são utilizadas para reduzir a influência desses fenômenos.

§ Técnicas não-paramétricas

Como discutido na seção 2.2.2 (pág. 13), várias medidas de semelhança entre pontos estão disponíveis na literatura. Contudo, nenhuma delas consegue eliminar totalmente a influência das distorções radiométricas. Além disso, quanto mais a medida de semelhança é robusta a esse tipo de distorção, maior é o seu custo computacional. Em vista disso, neste trabalho, aplicaram-se técnicas não-paramétricas ao problema de correspondência de pontos. Estas técnicas, ao invés de considerar a intensidade de um pixel em si, baseiam-se na ordem relativa das intensidades dos pixels dentro de uma janela W_n .

Uma das técnicas não-paramétricas conhecidas é a transformada posto (ZABIH; WOODFILL, 1994). Nesta, um pixel na posição (u, v) , é substituído pelo número de pixels dentro de uma janela W_n , centrada em (u, v) , que possuem intensidade menor que $I(u, v)$. Assim, a imagem original é transformada em uma matriz de números. Além disso, o número de bits necessários para armazenar a imagem após a transformação é, geralmente, menor. Por exemplo, para W_n de dimensão 5×5 , são necessários apenas 3 bits para representar cada elemento da imagem transformada.

Já que reduz o efeito das distorções radiométricas, a transformada posto permite o uso de correlação por *SAD* (seção 2.2.2, pág. 14) para expressar a semelhança entre pontos. Contudo, esta transformada não retém a localização dos pixels dentro da janela W_n . Para isto, pode-se utilizar a transformada censo (ZABIH; WOODFILL, 1994). Na transformada censo, os pixels dentro da janela W_n , centrada em (u, v) , são mapeados em uma seqüência de bits. Assim, se um pixel tem intensidade maior que $I(u, v)$, o bit correspondente na seqüência é feito igual a “1”, caso contrário é igual a “0”. Com isso, para W_n de dimensão 5×5 , são necessários 24 bits para representar cada elemento da imagem transformada. Um aumento significativo.

Após a transformada censo, para comparar a semelhança entre elementos de duas imagens, deve-se utilizar distância a Hamming através do somatório das distâncias Hamming dentro de uma janela W , ou seja,

$$SHD(c_{ij}) = \sum_{(u,v) \in W} I'_1(u, v) \oplus I'_2(u + x, y + v), \quad (3.1)$$

onde \oplus é a distância Hamming entre I'_1 e I'_2 . Aqui, I'_1 e I'_2 representam, respectivamente, a transformada censo das imagens 1 e 2. Assim, quanto menor o *SHD*, maior a semelhança entre os pontos que compõem o par c_{ij} .

A robustez da transformada censo pode ser demonstrada observando a figura 3.1. Nesta, considere que os pixels centrais das janelas 3×3 são correspondentes. Contudo, as intensidades desses pixels são diferentes entre si. Obviamente, isto dificulta a correspondência por correlação. No entanto,

255	43	78
89	123	199
200	23	12

240	25	85
100	150	199
200	35	18

10001100 10001100

Figura 3.1: Duas janelas de correlação 3×3 , cujos tons de cinza são diferentes, mas possuem distância Hamming nula.

aplicando-se a transformada censo, os pixels em questão são transformados em duas seqüências de bits idênticas, logo a distância Hamming entre eles é nula. Pelos critérios descritos neste seção, isto equivale a dizer que os pixels têm a mesma assinatura. Além disso, aplicando a transformada censo a todos os pixels dentro de W e calculando a distância Hamming entre eles, deve revelar que tais pontos são correspondentes com uma maior segurança.

Como as transformadas posto e censo envolvem apenas comparações entre número inteiros, elas podem ser aplicadas a um baixo custo computacional. Contudo, como envolve a manipulação de uma quantidade menor de dados, a transformada posto tende a ser de implementação mais fácil.

§ Múltiplas janelas de correlação

Oclusões e distorções projetivas fazem com que a vizinhança de um ponto interessante seja diferente nas duas imagens. Na figura 3.2(a) ilustra-se um exemplo típico de oclusão. Nesta, objetos a distâncias diferentes das câmeras parecem se mover quando são observados de perspectivas diferentes. Assim, a imagem em torno do ponto de interesse é diferente nas duas imagens. Evidentemente, isto dificulta a correspondência por semelhança. Uma maneira de reduzir este problema é diminuir a largura, l_w , da janela de correlação. Contudo, isto aumenta a influência do ruído e tende a reduzir o número de boas correspondências. Em vista disso, Kanade e Okutomi (1994) propuseram um método no qual o tamanho e a forma da janela de correlação adaptam-se ao local do ponto interessante. Contudo, seu método tem um custo computacional muito elevado. Por isso, vários autores como, por exemplo, (HIRSCHMÜLLER et al., 2002; FUSIELLO et al., 1997), propuseram uma alternativa à técnica de Kanade e Okutomi (1994). Tal alternativa é baseada em múltiplas janelas de correlação. Por exemplo, Hirschmüller et al. (2002) utilizaram um sistema com cinco janelas de correlação, como ilustrado na figura 3.2(b). Assim, a correlação entre pontos utilizada era um valor baseado na correlação devido a cada uma das cinco janelas. Pode-se dizer que esse método é uma simplificação da técnica de Kanade e Okutomi (1994).

Técnicas de correlação com janelas melhor elaboradas aumentam significativamente o custo computacional do algoritmo. Por isso, neste trabalho, utilizou-se a configuração padrão, ou seja, uma única janela com tamanho fixo. Dessa forma, extrema confiança é depositada na transformada censo.

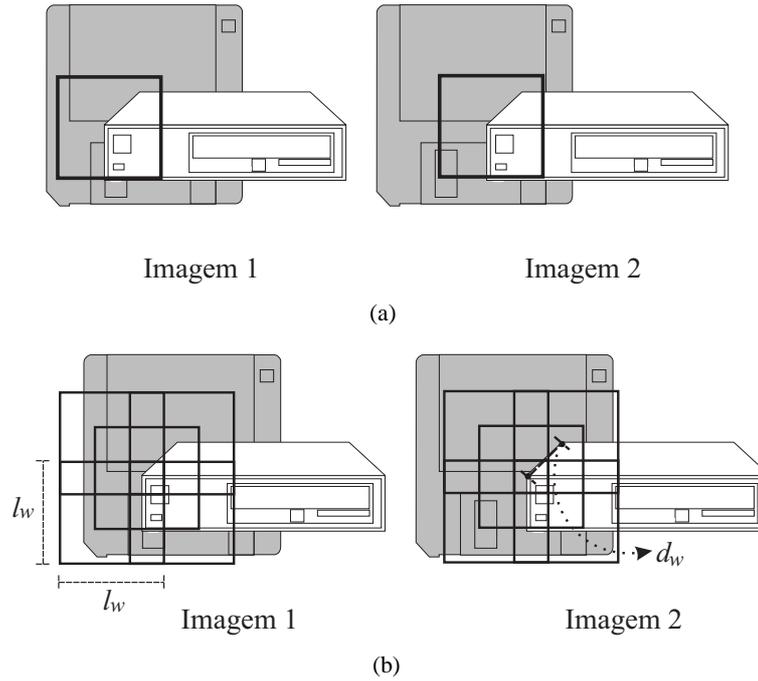


Figura 3.2: (a) influência das oclusões na correlação com apenas uma janela de correlação. (b) correlação com 5 janelas. Aqui, em pelo menos uma janela, a semelhança é conservada.

§ Abordagem proposta

Para escolher os pares candidatos à correspondência, na técnica proposta começa-se aplicando a transformada ceno às duas imagens. Em seguida, para cada ponto interessante, \mathbf{m}_{1i} , da imagem 1, calcula-se a semelhança entre este e todos os pontos interessantes, \mathbf{m}_{2j} , da imagem 2 que estão dentro de uma janela de dimensão $2l_s \times 2l_s$, centrada em \mathbf{m}_{1i} (figura 2.9, na página 19). Então, o par $c_{ij} = (\mathbf{m}_{1i}, \mathbf{m}_{2j})$ que apresentar a maior semelhança, é considerado um candidato à correspondência.

Devido a instabilidade dos algoritmos disponíveis para segmentar os pontos interessantes nas duas imagens, na prática, diversos pontos interessantes de uma imagem não correspondem a nenhum ponto interessante 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 ponto interessante da imagem 2 forme um par com mais de um ponto da image 1. Assim, uma outra etapa para eliminar a ambigüidade é necessária e discutida na próxima seção.

3.1.2 Eliminação da Ambigüidade

A restrição de vizinhança é utilizada para eliminar a ambigüidade do conjunto de candidatos à correspondência. Para isto, baseado na posição relativa dos candidatos vizinhos (seção 2.2.2, pág. 17), definiu-se uma medida da confiança de um candidato à correspondência. Tal medida foi proposta tendo como principal critério a simplicidade.

Formalmente, definimos a confiança, $Reab$, de um par $c_{ij} = (\mathbf{m}_{1i}, \mathbf{m}_{2j})$ candidato a corres-

pondência, baseado nos candidatos a correspondência vizinhos, pela equação

$$Reab(c_{ij}) = \sum_{\mathbf{n}_{1k} \in N(\mathbf{m}_{1i})} \left(\sum_{\mathbf{n}_{2l} \in N(\mathbf{m}_{2j})} \Phi(c_{ij}, v_{kl}) \right), \quad (3.2)$$

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,

$$r(c_{ij}, v_{kl}) = \frac{|d(\mathbf{m}_{1i}, \mathbf{n}_{1k}) - d(\mathbf{m}_{2j}, \mathbf{n}_{2l})|}{[d(\mathbf{m}_{1i}, \mathbf{n}_{1k}) + d(\mathbf{m}_{2j}, \mathbf{n}_{2l})]/2} \quad (3.3)$$

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:

1. A equação (3.2) conta o número de candidatos a 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} . Isto 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, pág. 17), o somatório da equação (3.2) deve contar todos os pares ambíguos como se fossem apenas um único par;
4. Na equação (3.3), pode-se fazer $d(\mathbf{m}_x, \mathbf{n}_y)$ igual a distância euclidiana. Contudo, neste trabalho utilizou-se

$$d(\mathbf{m}_x, \mathbf{n}_y) = |u_m - u_n| + |v_m - v_n|, \quad (3.4)$$

onde consideramos $\mathbf{m}_x = [u_m, v_m]^T$ e $\mathbf{n}_y = [u_n, v_n]^T$. A equação (3.4) tem um custo computacional menor que o da distância euclidiana e o seu uso na equação (3.3) não prejudica o cálculo da confiança de $(\mathbf{m}_{1i}, \mathbf{m}_{2j})$.

Para eliminar a ambigüidade, deve-se aplicar a equação (3.2) em todos os candidatos a correspondência e:

1. Candidatos com $Reab(c_{ij}) \leq rb_{th}$ são descartados;
2. Caso haja ambigüidade, prevalece o candidato mais confiável, e;
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.

3.1.3 Eliminando as falsas correspondências

Desde que, neste trabalho, busca-se um conjunto de boas correspondências para estimar a matriz fundamental. A melhor abordagem para rejeitar falsas correspondências é utilizar algoritmos que estimem a matriz fundamental de forma robusta, pois, ao mesmo tempo que as falsas correspondências são eliminadas, uma estimação inicial da matriz fundamental é conhecida. Para isto, neste trabalho, é utilizado o algoritmo MSAC. Como discutido na seção 2.4.5 (pág. 36), tal algoritmo é uma atualização do RANSAC e funciona bem mesmo quando a quantidade de falsas correspondências é superior a 50 %.

No algoritmo proposto, a eliminação das falsas correspondências é o único estágio iterativo. Isto fornece uma boa relação entre exatidão e velocidade. Contudo, infelizmente, o método não reconhece uma falsa correspondência se os pontos que a compõem estiverem muito próximos às linhas epipolares respectivas. Contudo, tais falsas correspondências não deterioram a estimação da matriz fundamental.

No próximo capítulo, o algoritmo proposto é avaliado e sua eficiência é comparada com a de outro algoritmo bastante conhecido na literatura.

§3.2 Melhorando a Estimação da Matriz Fundamental

Nesta seção, um novo método iterativo, mas linear, para cálculo da matriz fundamental é proposto. De fato, o método apenas melhora uma estimação já existente. Mais especificamente, dada uma matriz fundamental, o objetivo é melhorar a estimação dos epipolos, parâmetros que são muito susceptíveis a ruído. Tal objetivo é alcançado reestimando o epipolo $\tilde{\mathbf{e}}_2$ que minimiza o erro de Sampson (1982) dado pela equação (2.40), na página 32.

Inicialmente, uma estimação inicial, \mathbf{F}_0 , da matriz fundamental é decomposta pelo teorema 2.2 (pág. 26). Então, de (2.31) e (2.32), pode-se reescrever (2.40) como sendo

$$\sum_{i=1}^n G_i^2 (\boldsymbol{\alpha}_i^T \tilde{\mathbf{e}}_2)^2 = 0, \quad \text{sujeito a } \|\tilde{\mathbf{e}}_2\| = 1, \quad (3.5)$$

onde $G_i = \sqrt{g_i}$, com g_i dado por (2.41), e

$$\boldsymbol{\alpha}_i = \begin{bmatrix} H_{23} + H_{21}u_{1i} - H_{33}v_{2i} - H_{31}u_{1i}v_{2i} + H_{22}v_{1i} - H_{32}v_{2i}v_{1i} \\ -H_{13} + H_{33}u_{2i} - H_{11}u_{1i} + H_{31}u_{2i}u_{1i} - H_{12}v_{1i} + H_{32}u_{2i}v_{1i} \\ -H_{23}u_{2i} - H_{21}u_{2i}u_{1i} + H_{13}v_{2i} + H_{11}u_{1i}v_{2i} - H_{22}u_{2i}v_{1i} + H_{12}v_{2i}v_{1i} \end{bmatrix}, \quad (3.6)$$

com H_{ij} igual ao elemento da i -ésima linha e j -ésima coluna da matriz \mathbf{H} (veja o teorema 2.2), $\mathbf{m}_{1i} = [u_{1i}, v_{1i}]^T$ e $\mathbf{m}_{2i} = [u_{2i}, v_{2i}]^T$.

Considerando $\boldsymbol{\beta} = [G_1\alpha_1, G_2\alpha_2, \dots, G_n\alpha_n]^T$, podemos reescrever o problema (3.5), simples-

Entrada: n correspondências ($\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2i}$) e uma estimação, \mathbf{F}_0 , da matriz fundamental.

Saída: matriz fundamental, \mathbf{F} , de posto 2 que satisfaz a equação (2.40).

- a) Transforme as correspondências de pontos de acordo com $\tilde{\mathbf{m}}_{1i} = \mathbf{T}_2 \tilde{\mathbf{m}}_{1i}$ e $\tilde{\mathbf{m}}_{2i} = \mathbf{T}'_2 \tilde{\mathbf{m}}_{2i}$, onde \mathbf{T}_2 e \mathbf{T}'_2 são matrizes 3×3 que impõem a normalização proposta por (HARTLEY, 1997);
- b) Dada a matriz \mathbf{F}_0 , calcule \mathbf{H} pelo teorema 2.2;
- c) Com as correspondências ($\hat{\mathbf{m}}_{1i} \leftrightarrow \hat{\mathbf{m}}_{2i}$), fazendo $G_i = 1$ e α_i dado pela equação (3.6), inicialize a matriz β ;
- d) Compute o epipolo $\tilde{\mathbf{e}}_2$ resolvendo o problema (3.7), através da proposição B.2;
- e) Usando a equação $\mathbf{F} = [\tilde{\mathbf{e}}_2]_{\times} \mathbf{H}$, atualize a estimação de $\hat{\mathbf{F}}$;
- f) Usando a matriz $\hat{\mathbf{F}}$ calculada no passo anterior, atualize a matriz β considerando $G_i^{-2} = \hat{\mathbf{l}}_i^T \mathbf{T}_2 \mathbf{Z} \mathbf{T}_2^T \hat{\mathbf{l}}_i + \hat{\mathbf{l}}_i'^T \mathbf{T}'_2 \mathbf{Z} \mathbf{T}'_2{}^T \hat{\mathbf{l}}_i'$, onde $\hat{\mathbf{l}}_i = \hat{\mathbf{F}} \tilde{\mathbf{m}}_{1i}$ e $\hat{\mathbf{l}}_i' = \hat{\mathbf{F}}^T \tilde{\mathbf{m}}_{2i}'$;
- g) Compute o resíduo dado por $\|\beta \tilde{\mathbf{e}}_2\|$;
- h) Repita os passos de (d) a (g) até a convergência do resíduo, e;
- i) Faça $\mathbf{F} = \mathbf{T}'_2 [\tilde{\mathbf{e}}_2]_{\times} \mathbf{H} \mathbf{T}_2$.

Algoritmo 3.1: Método linear e iterativo para estimação da matriz fundamental

mente, como sendo

$$\|\beta \tilde{\mathbf{e}}_2\|^2 = 0, \quad \text{sujeito a } \|\tilde{\mathbf{e}}_2\| = 1. \quad (3.7)$$

Uma solução não-iterativa para o problema anterior não é possível, pois β depende da matriz fundamental. Contudo, uma solução iterativa, sem o uso de métodos não-lineares baseados no algoritmo de Newton, é factível. Tal solução é resumida no algoritmo 3.1.

A vantagem do método proposto é que a parte iterativa do algoritmo consiste de um problema de minimização extremamente simples. Este envolve a estimação de apenas três parâmetros (as coordenadas homogêneas do epipolo $\tilde{\mathbf{e}}_2$). Apesar disso, o algoritmo encontra a matriz fundamental que minimiza uma aproximação de primeira ordem do erro geométrico para todo o conjunto de correspondências de pontos.

§3.3 Estimação da Matriz Fundamental: Um Novo Método Não-Linear

Durante a estimação da matriz fundamental, a restrição de posto dois deve ser atendida. Caso contrário, a matriz estimada não se ajustará bem ao conjunto de correspondências, ou seja, a matriz obtida não representará a geometria epipolar de forma satisfatória.

Como discutido na seção 2.4.3, utilizando métodos lineares a restrição de posto dois só pode ser imposta após a matriz ter sido estimada. Por isso, métodos não-lineares que limitam o posto da matriz estimada obtêm resultados mais exatos (segundo o critério da equação (2.30)). Contudo, tais métodos

utilizam parametrizaes que fazem uso das coordenadas euclidianas dos epipolos. Assim, como discutido na seo 2.4.4, dependendo da localizao dos epipolos, a parametrizao torna-se instvel.

Uma parametrizao alternativa  dada pela equao (2.31) (pg. 26). Esta depende das coordenadas homogneas do epipolo, logo  mais estvel.

Como discutido na seo 2.5.2 (pg. 42), a matriz \mathbf{H} em (2.31) est associada a um plano no espao \mathcal{P}^3 e pertence a uma famlia de homografias. Assim, considerando uma homografia \mathbf{H} induzida por um plano π , para pontos \mathbf{m}_{1i} e \mathbf{m}_{2i} que no pertencem a π , temos $\tilde{\mathbf{m}}_{1i} \neq \lambda \mathbf{H} \mathbf{m}_{2i}$, onde λ um  fator de escala desconhecido. Como discutido a seguir, este fato (que representa o paralaxe) permite estimar um epipolo (se \mathbf{H} for conhecida) e, portanto, determinar a matriz fundamental.

3.3.1 Estimando Epipolos Atravs de Planos

Dada a homografia \mathbf{H} induzida por um plano π , considerando uma correspondncia ($\mathbf{m}_{11} \leftrightarrow \mathbf{m}_{21}$) formada por projees de pontos no pertencentes a π , temos

$$\tilde{\mathbf{p}}_{21} \simeq \mathbf{H} \tilde{\mathbf{m}}_{11}. \quad (3.8)$$

Como ilustrado na figura 3.3, o ponto \mathbf{p}_{21} est sobre a reta epipolar, \mathbf{l}_{21} , de \mathbf{m}_{11} . Assim, $\mathbf{l}_{21} \simeq (\tilde{\mathbf{m}}_{21} \times \tilde{\mathbf{p}}_{21})$. Dessa forma, dadas duas retas epipolares, \mathbf{l}_{21} e \mathbf{l}_{22} , \mathbf{e}_2 pode ser obtido pela interseo de tais retas, ou seja, $\mathbf{e}_2 \simeq (\mathbf{l}_{21} \times \mathbf{l}_{22})$. Uma vez que \mathbf{e}_2  conhecido, \mathbf{F} pode ser calculada utilizando (2.31).

O paralaxe entre imagens estreo j foi utilizado, por exemplo, por Bartoli et al. (2001) e Luong e Faugeras (1993), para clculo da matriz fundamental. Contudo, esta tcnica requer a identificao de um plano na imagem e a estimaco da homografia associada a tal plano. Infelizmente, um plano nem sempre est visvel. Alm disso, os objetos planos mais comuns, por exemplo, paredes e portas, so, geralmente, pobres em textura. Logo, so regies difceis de se estabelecer correspondncias de pontos. Felizmente, na prxima seo, discute-se um mtodo que elimina a necessidade do conhecimento prvio de um plano.

3.3.2 Paralaxe Virtual

Trs pontos no co-lineares no espao \mathcal{P}^3 definem um plano π (figura 3.4). Alm disso, desde que co-linearidade  uma propriedade invariante a transformaes projetivas, trs correspondncias ($\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2i}$), onde $i = 1, 2$ e 3 , e nenhum dos pontos \mathbf{m}_{1i} so co-lineares, tambm definem π (veja a figura 3.4). Alm disso, ($\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2i}$) satisfaz

$$\tilde{\mathbf{m}}_{1i} \simeq \mathbf{H} \tilde{\mathbf{m}}_{2i}.$$

Para cada correspondncia ($\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2i}$), a equao anterior fornece duas equaes independentes. Contudo, \mathbf{H} tem nove graus de liberdade. Dessa forma,  impossvel determin-la unicamente.

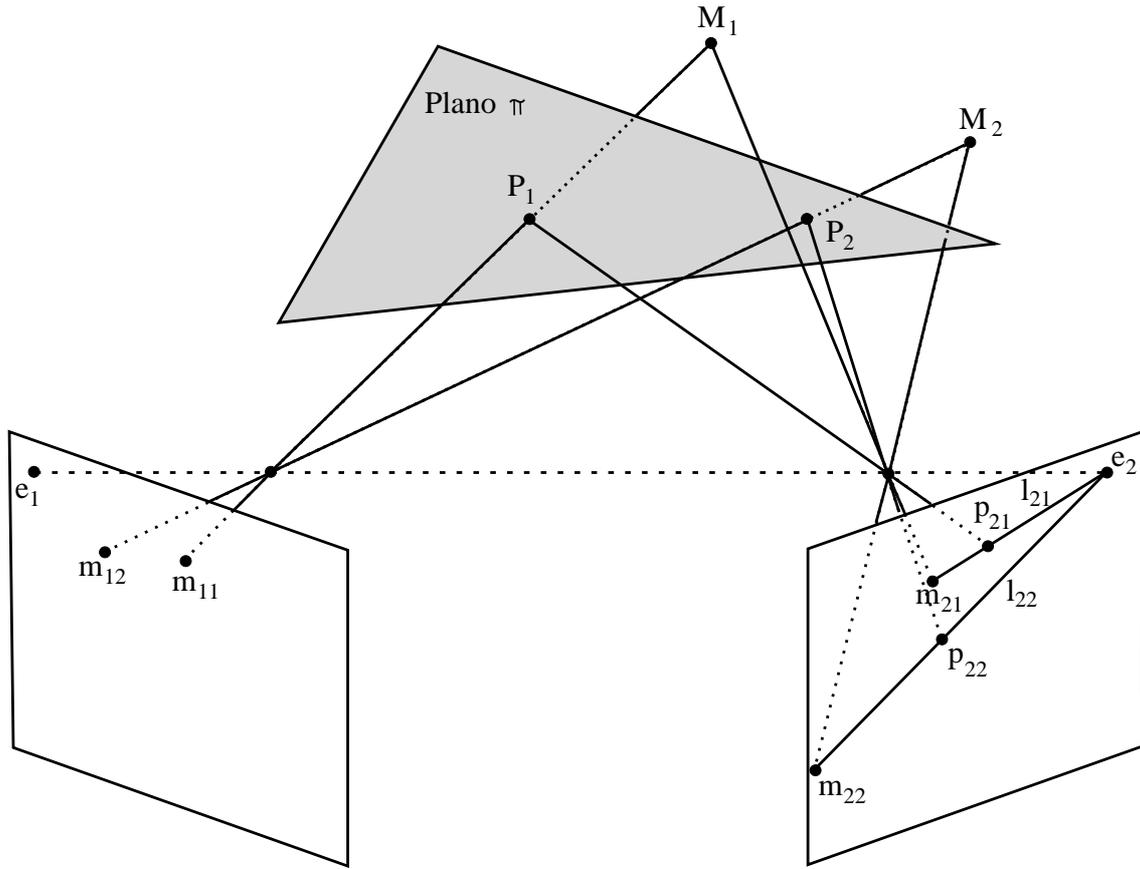


Figura 3.3: Paralaxe induzido pelo plano II.

Entretanto, é sempre possível determinar duas transformações de \mathcal{P}^2 , \mathbf{T}_p e \mathbf{T}'_p , tal que,

$$\begin{aligned} \mathbf{T}_p \tilde{\mathbf{m}}_{11} &= [1, 0, 0]^T, & \mathbf{T}'_p \tilde{\mathbf{m}}_{21} &= [1, 0, 0]^T \\ \mathbf{T}_p \tilde{\mathbf{m}}_{12} &= [0, 1, 0]^T, & \mathbf{T}'_p \tilde{\mathbf{m}}_{22} &= [0, 1, 0]^T \\ \mathbf{T}_p \tilde{\mathbf{m}}_{13} &= [0, 0, 1]^T, & \mathbf{T}'_p \tilde{\mathbf{m}}_{23} &= [0, 0, 1]^T. \end{aligned} \quad (3.9)$$

Agora, no sistema de coordenadas definido por \mathbf{T}_p e \mathbf{T}'_p , \mathbf{H} é representada por

$$\mathbf{H} = \begin{bmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \gamma \end{bmatrix}.$$

Além disso, desde que, necessariamente, \mathbf{H} é não-singular, α , β e γ são sempre diferentes de zero. Assim, \mathbf{H} pode ser simplificada como sendo

$$\mathbf{H} = \begin{bmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (3.10)$$

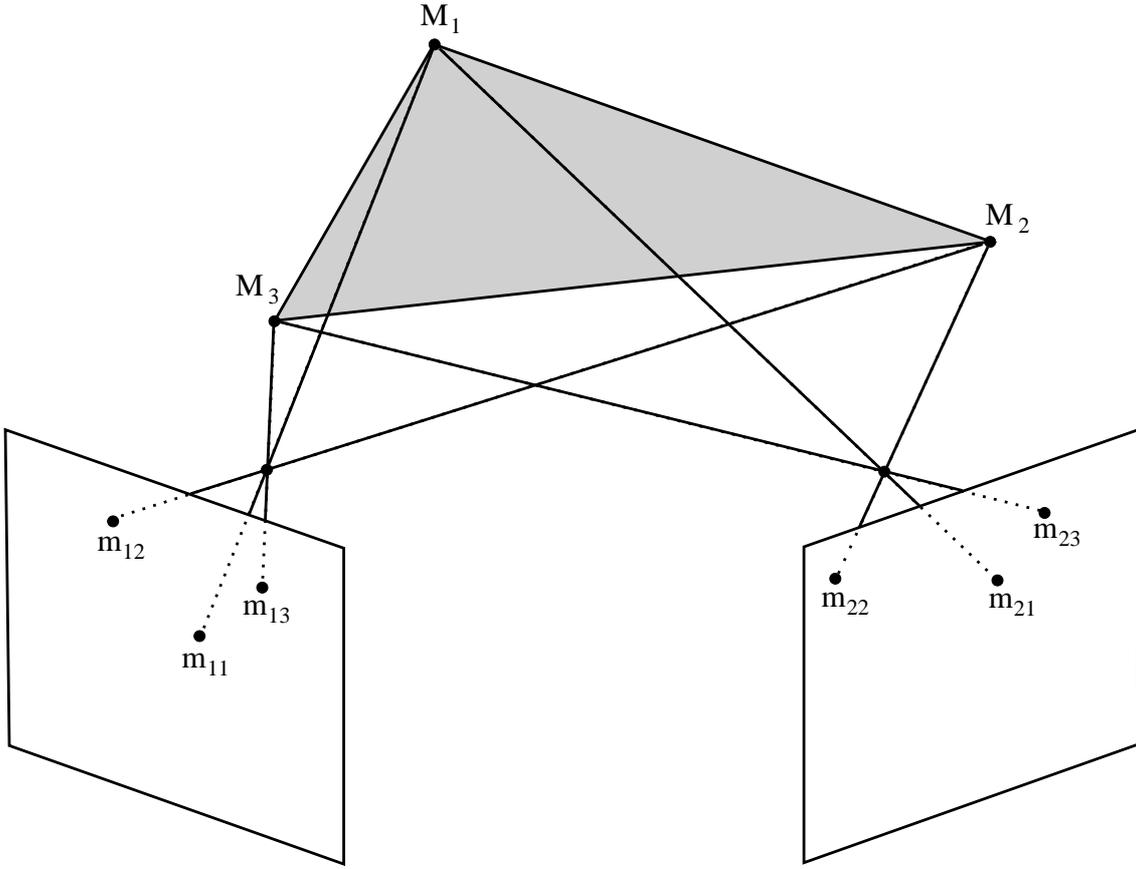


Figura 3.4: Três pontos 3D e não co-lineares formam um plano no espaço. As projeções desses pontos também são não co-lineares.

Logo, considerando $\tilde{\mathbf{m}}_{1j} = [u_{1j}, v_{1j}, t_{1j}]^T$ e substituindo (3.10) em (3.8), temos

$$\tilde{\mathbf{p}}_{2j} \simeq [\alpha u_{1j}, \beta v_{1j}, t_{1j}]^T. \quad (3.11)$$

Além disso, como discutido anteriormente, \mathbf{p}_{2j} pertence a reta epipolar $\mathbf{l}_{2j} \simeq (\tilde{\mathbf{m}}_{2j} \times \tilde{\mathbf{e}}_2)$, ou seja,

$$(\tilde{\mathbf{m}}_{2j} \times \tilde{\mathbf{e}}_2) \cdot \tilde{\mathbf{p}}_{2j} = 0.$$

Substituindo (3.11) na equação anterior e considerando $\tilde{\mathbf{m}}_{2j} = [u_{2j}, v_{2j}, t_{2j}]^T$ e $\tilde{\mathbf{e}}_2 = [e_{2u}, e_{2v}, e_{2t}]^T$, após um tratamento algébrico, temos

$$\mathbf{a}_j^T \mathbf{V} = 0, \quad (3.12)$$

na qual $\mathbf{a}_j = [-u_{1j}t_{2j}, t_{1j}u_{2j}, v_{1j}t_{2j}, -v_{1j}u_{2j}, -t_{1j}v_{2j}, u_{1j}v_{2j}]^T$ e $\mathbf{V} = [V_1, V_2, V_3, V_4, V_5, V_6]^T$, com

$$V_1 = \alpha e_{2v}, V_2 = e_{2v}, V_3 = \beta e_{2u}, V_4 = \beta e_{2t}, V_5 = e_{2u} \text{ e } V_6 = \alpha e_{2t}. \quad (3.13)$$

Da equação (3.12), com n correspondências de pontos ($\mathbf{m}_{1j} \leftrightarrow \mathbf{m}_{2j}$), temos o seguinte sistema de equações lineares

$$\mathbf{A}_n \mathbf{V} = \mathbf{0}, \quad (3.14)$$

onde $\mathbf{A}_n = [\mathbf{a}_1, \dots, \mathbf{a}_n]^T$.

Na equação anterior, \mathbf{A}_n tem posto cinco. Logo, (3.14) pode ser resolvido utilizando a proposição B.2. Além disso, substituindo (3.10) em (2.31) e considerando (3.13), podemos escrever \mathbf{F} em função do vetor \mathbf{V} , ou seja,

$$\begin{aligned} \mathbf{F} &= \begin{bmatrix} 0 & -e_{2t} & e_{2v} \\ e_{2t} & 0 & -e_{2u} \\ -e_{2v} & e_{2u} & 0 \end{bmatrix} \begin{bmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 0 & -\beta e_{2t} & e_{2v} \\ \alpha e_{2t} & 0 & -e_{2u} \\ -\alpha e_{2v} & \beta e_{2u} & 0 \end{bmatrix} \end{aligned} \quad (3.15)$$

$$= \begin{bmatrix} 0 & -V_4 & V_2 \\ V_6 & 0 & -V_5 \\ -V_1 & V_3 & 0 \end{bmatrix}. \quad (3.16)$$

A equação (3.14) é linear e possui apenas cinco parâmetros independentes. Considerando os três pontos ($\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2i}$), utilizados para calcular \mathbf{T}_p e \mathbf{T}'_p , são necessários oito correspondências para estimar a matriz fundamental a partir de (3.14).

Apesar de dispensar a identificação de um plano, o método do paralaxe virtual possui alguns problemas muito graves, listados a seguir.

- ✓ Devido a substituição de variáveis em (3.13), utilizada para tornar o problema linear, a matriz obtida em (3.16), geralmente, não tem posto dois. Por isso, essa restrição tem que ser imposta *a posteriori*.
- ✓ Como no caso do algoritmo dos 8 pontos, a solução de (3.14) é muito dependente do sistema de coordenadas utilizado. Contudo, devido a restrição (3.9), o conjunto de pontos não podem ser normalizados como sugerido por Hartley (1997) (seção 2.4.3).
- ✓ As transformações \mathbf{T}_p e \mathbf{T}'_p podem fazer com que o problema (3.14) torne-se muito mal-condicionado. Geralmente, isto torna a matriz fundamental estimada inutilizável.

Infelizmente, o método do paralaxe virtual não tem aplicação prática, pois a matriz fundamental obtida pelo método é muito dependente das transformações \mathbf{T}_p e \mathbf{T}'_p . Assim, como demonstrado empiricamente no próximo capítulo, dependendo da escolha das correspondências ($\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2i}$), o resultado pode variar entre o “excelente” e o “inútil”. Contudo, na próxima seção, apresenta-se uma metodologia que fornece um excelente método não-linear baseado no paralaxe virtual.

3.3.3 Paralaxe Virtual: Método Não-Linear

Para que a restrição de posto dois seja satisfeita, (3.13) não pode ser utilizada. Neste caso, o problema do paralaxe virtual torna-se não-linear. Contudo, a vantagem evidente é que (3.15) fornece

uma excelente parametrizao de posto dois, com apenas cinco parmetros, para a matriz fundamental. Como demonstrado empiricamente no prximo captulo, tal parametrizao funciona muito bem quando os epipolos esto no infinito ou prximo ao centro da imagem.

Outro problema do mtodo do paralaxe virtual  a escolha das transformaes \mathbf{T}_p e \mathbf{T}'_p . Hartley (1997) demonstrou que o problema de estimar a matriz fundamental torna-se bem melhor condicionado se forem aplicadas transformaes \mathbf{T}_2 e \mathbf{T}'_2 (seo 2.4.3), respectivamente, s imagens I_1 e I_2 . Por outro lado, escolhidas trs correspondncias ($\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2i}$), as matrizes \mathbf{T}_p e \mathbf{T}'_p tornam o problema mal-condicionado. Por isso, mesmo usando uma abordagem no-linear, a matriz fundamental obtida  inexata (segundo o critrio da equao (2.30)).

A soluo para este problema no  a alternativa bvia, ou seja, fazer $\mathbf{T}_p = \mathbf{T}_2$ e $\mathbf{T}'_p = \mathbf{T}'_2$, pois, dessa forma, as seguintes correspondncias no so satisfeitas

$$\begin{aligned} (\mathbf{T}_2 \tilde{\mathbf{m}}_{11} = [1, 0, 0]^T \leftrightarrow \mathbf{T}'_p \tilde{\mathbf{m}}_{21} = [1, 0, 0]^T), \\ (\mathbf{T}_2 \tilde{\mathbf{m}}_{12} = [0, 1, 0]^T \leftrightarrow \mathbf{T}'_p \tilde{\mathbf{m}}_{22} = [0, 1, 0]^T), \\ (\mathbf{T}_2 \tilde{\mathbf{m}}_{13} = [0, 0, 1]^T \leftrightarrow \mathbf{T}'_p \tilde{\mathbf{m}}_{23} = [0, 0, 1]^T). \end{aligned} \quad (3.17)$$

Assim, \mathbf{H} no tem a forma (3.10) e o algoritmo do paralaxe virtual no pode ser aplicado.

A soluo correta  escolher as transformaes \mathbf{T}_p e \mathbf{T}'_p que mais se aproximam de \mathbf{T}_2 e \mathbf{T}'_2 e satisfaa (3.17) exatamente. Isto pode ser feito como segue.

- i) Calculam-se \mathbf{T}_2 e \mathbf{T}'_2 como discutido na seo 2.4.3.
- ii) Calculam-se os pontos

$$\begin{aligned} \tilde{\mathbf{m}}_{11} &= \mathbf{T}_2^{-1}[1, 0, 0]^T, & \tilde{\mathbf{m}}_{21} &= \mathbf{T}'_2^{-1}[1, 0, 0]^T \\ \tilde{\mathbf{m}}_{12} &= \mathbf{T}_2^{-1}[0, 1, 0]^T, & \tilde{\mathbf{m}}_{22} &= \mathbf{T}'_2^{-1}[0, 1, 0]^T \\ \tilde{\mathbf{m}}_{13} &= \mathbf{T}_2^{-1}[0, 0, 1]^T, & \tilde{\mathbf{m}}_{23} &= \mathbf{T}'_2^{-1}[0, 0, 1]^T. \end{aligned} \quad (3.18)$$

- iii) Para $i = 1, 2$ e 3 , minimiza-se

$$d^2(\tilde{\mathbf{m}}_{1i}, \tilde{\mathbf{m}}_{1i}) + d^2(\tilde{\mathbf{m}}_{2i}, \tilde{\mathbf{m}}_{2i}), \quad \text{sujeito a } \tilde{\mathbf{m}}_{2i}^T \mathbf{F}_0 \tilde{\mathbf{m}}_{1i}, \quad (3.19)$$

onde \mathbf{F}_0  uma estimaco inicial da matriz fundamental.

- iv) Com as correspondncias, ($\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2i}$), obtidas, calculam-se \mathbf{T}_p e \mathbf{T}'_p que satisfazem (3.9).

Agora, as observaes a seguir devem ser consideradas.

- ✓ Na busca pelas transformaes, \mathbf{T}_p e \mathbf{T}'_p , ideais, o mtodo anterior necessita de uma estimaco inicial, \mathbf{F}_0 , da matriz fundamental. Contudo, desde que, na prtica, a soluo de problemas no-lineares necessitam do conhecimento prvio de uma soluo aproximada para o problema, este requerimento  perfeitamente vivel.

- ✓ O problema em (3.19) é semelhante ao problema em (2.92). Dessa forma, ele pode ser resolvido de forma idêntica.
- ✓ Pela metodologia proposta, as correspondências $(\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2i})$, $i = 1, 2$, e 3 , são “correspondências virtuais”, ou seja, não fazem parte, de fato, do conjunto de correspondência de pontos utilizado na estimação.
- ✓ Se \mathbf{F}_0 for uma estimação “ruim” da matriz fundamental, as correspondências virtuais, $(\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2i})$, funcionarão como falsas correspondências. Contudo, como discutido no próximo capítulo, o algoritmo proposto é bastante robusto com relação a uma estimação “pobre” da matriz fundamental.

Como discutido na seção 2.7.2, para resolver (3.19), é necessário obter-se todas as raízes de um polinômio de grau seis. Contudo, (3.19) pode ser reformulado de uma forma linear como segue.

- i) Para cada ponto $\hat{\mathbf{m}}_{2i}$, $i = 1, 2$, e 3 , em (3.18), encontra-se o ponto \mathbf{m}_{2i} que minimiza

$$d^2(\mathbf{m}_{2i}, \hat{\mathbf{m}}_{2i}), \quad \text{sujeito a } (\mathbf{F}_0 \tilde{\mathbf{m}}_{1i}) \cdot \tilde{\mathbf{m}}_{2i} = 0. \quad (3.20)$$

A solução do problema anterior corresponde a encontrar o ponto \mathbf{m}_{2i} mais próximo de $\hat{\mathbf{m}}_{2i}$ e que está sobre a reta $\hat{\mathbf{l}}_{2i} = \mathbf{F}_0 \tilde{\mathbf{m}}_{1i}$, ou seja, \mathbf{m}_{2i} é a projeção ortogonal de $\hat{\mathbf{m}}_{2i}$ em $\hat{\mathbf{l}}_{2i}$. Logo, \mathbf{m}_{2i} pode ser encontrado, de forma linear, através da proposição A.2.

- ii) Considera-se $\mathbf{T}_p = \mathbf{T}_2$ e, com os pontos \mathbf{m}_{2i} , $i = 1, 2$, e 3 , calculados no item anterior, estima-se \mathbf{T}'_p que satisfaz (3.9).

Com a abordagem anterior, todo o erro é transferido para a segunda imagem. Contudo, agora, as transformações \mathbf{T}_p e \mathbf{T}'_p são obtidas de forma extremamente simplificada.

Uma vez que \mathbf{T}_p e \mathbf{T}'_p tenham sido determinadas, a minimização não-linear pode ser efetuada para que \mathbf{F} seja estimada. Mais detalhes sobre tal minimização são dados a seguir.

3.3.4 Detalhes da Minimização

Para estimar a matriz fundamental, a parametrização (3.15) deve ser utilizada em conjunto com um critério a ser minimizado. Para derivar tal critério, deve ser observado que as correspondências $(\mathbf{m}_{1j} \leftrightarrow \mathbf{m}_{2j})$ relacionam-se com a matriz fundamental, $\hat{\mathbf{F}}$, do novo sistema de coordenadas (definido por \mathbf{T}_p e \mathbf{T}'_p), por

$$\mathcal{F}_j = \tilde{\mathbf{m}}_{2j}^T \mathbf{T}'_p{}^T \hat{\mathbf{F}} \mathbf{T}_p \tilde{\mathbf{m}}_{1j} = 0. \quad (3.21)$$

Além disso, como usual, considere as correspondências $(\mathbf{m}_{1j} \leftrightarrow \mathbf{m}_{2j})$ corrompidas, independentemente, por um ruído gaussiano de média zero e variância σ . Neste caso, as matrizes de covariância são dadas por

$$\Lambda_{\mathbf{m}_{1j}} = \Lambda_{\mathbf{m}_{2j}} = \sigma^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \quad (3.22)$$

Contudo, após as transformações \mathbf{T}_p e \mathbf{T}'_p serem aplicadas às correspondências ($\mathbf{m}_{1j} \leftrightarrow \mathbf{m}_{2j}$), a equação anterior deixa de ser uma suposição razoável. Por conta disso, um critério que levará a melhores resultados é um que preserve a distribuição do ruído do sistema de coordenadas original. Um critério com tais características é

$$\min_{\hat{\mathbf{F}}} \sum_j g_j (\tilde{\mathbf{m}}_{2j}^T \mathbf{T}'_p{}^T \hat{\mathbf{F}} \mathbf{T}_p \tilde{\mathbf{m}}_{1j})^2, \quad (3.23)$$

onde $g_j = \sigma_j^{-2}$ e σ_j é a variância de (3.21).

Considerando (3.22), a matriz de covariância dos pontos em coordenadas homogêneas é

$$\mathbf{\Lambda}_{\tilde{\mathbf{m}}_{1j}} = \mathbf{\Lambda}_{\tilde{\mathbf{m}}_{2j}} = \sigma^2 \text{diag}(\mathbf{1}, \mathbf{1}, \mathbf{0}).$$

Agora, podemos derivar uma aproximação de primeira ordem para σ_j , ou seja,

$$\begin{aligned} \sigma_j^2 &= \left(\frac{\partial \mathcal{F}_j}{\partial \tilde{\mathbf{m}}_{1j}} \right)^T \mathbf{\Lambda}_{\tilde{\mathbf{m}}_{1j}} \left(\frac{\partial \mathcal{F}_j}{\partial \tilde{\mathbf{m}}_{1j}} \right) + \left(\frac{\partial \mathcal{F}_j}{\partial \tilde{\mathbf{m}}_{2j}} \right)^T \mathbf{\Lambda}_{\tilde{\mathbf{m}}_{2j}} \left(\frac{\partial \mathcal{F}_j}{\partial \tilde{\mathbf{m}}_{2j}} \right) \\ &= \sigma^2 \left(\hat{\mathbf{l}}_{1j}^T \mathbf{T}_p \mathbf{Z} \mathbf{T}_p^T \hat{\mathbf{l}}_{1j} + \hat{\mathbf{l}}_{2j}^T \mathbf{T}'_p \mathbf{Z} \mathbf{T}'_p{}^T \hat{\mathbf{l}}_{2j} \right), \end{aligned}$$

onde $\mathbf{Z} = \text{diag}(1, 1, 0)$, $\hat{\mathbf{l}}_{1j} \simeq \hat{\mathbf{F}}^T \tilde{\mathbf{m}}_{2j}$ e $\hat{\mathbf{l}}_{2j} \simeq \hat{\mathbf{F}} \tilde{\mathbf{m}}_{1j}$.

Da equação anterior, desde que multiplicar \mathcal{F}_j por σ^2 não altera o resultado de (3.23), podemos fazer

$$g_j = \left(\frac{1}{\hat{\mathbf{l}}_{1j}^T \mathbf{T}_p \mathbf{Z} \mathbf{T}_p^T \hat{\mathbf{l}}_{1j} + \hat{\mathbf{l}}_{2j}^T \mathbf{T}'_p \mathbf{Z} \mathbf{T}'_p{}^T \hat{\mathbf{l}}_{2j}} \right). \quad (3.24)$$

O método não-linear proposto (resumido no algoritmo 3.2) tem a vantagem de estimar a matriz fundamental com baixo custo computacional (comparado aos métodos que utilizam a parametrização discutida na seção 2.4.4), pois constitui um problema simples de otimização com apenas cinco incógnitas e um único caminho de busca. De fato, não existe nenhum outro método geral (que não se restrinja a conjuntos binoculares com alguma geometria particular) na literatura que utilize tão poucas incógnitas.

3.3.5 Aperfeiçoamento do Método

Apesar da parametrização (3.16) utilizar as coordenadas homogêneas do epipolo, a exatidão do algoritmo proposto varia com a localização dos epipolos. Como sugerem os resultados experimentais do próximo capítulo, melhores resultados são obtidos quando os epipolos estão próximos ao centro das imagens. Infelizmente, esta não é a configuração mais comum. Para contornar este problema, podemos aplicar transformações projetivas, \mathbf{T}_t e \mathbf{T}'_t , ao conjunto de correspondências, de modo que, no novo sistema de coordenadas, os epipolos não estão no infinito. Isto pode ser realizado como segue.

1. Faça \mathbf{T}_t e \mathbf{T}'_t iguais a matriz identidade;

Entrada: n correspondências ($\mathbf{m}_{1j} \leftrightarrow \mathbf{m}_{2j}$) e uma estimação inicial, \mathbf{F}_0 , da matriz fundamental.

Saída: Uma matriz fundamental, \mathbf{F} , de posto dois.

a) Calculam-se \mathbf{T}_2 e \mathbf{T}'_2 como discutido na seção 2.4.3.

b) Calculam-se os pontos

$$\begin{aligned}\tilde{\mathbf{m}}_{11} &= \mathbf{T}_2^{-1}[1, 0, 0]^T, & \tilde{\mathbf{m}}_{21} &= \mathbf{T}'_2^{-1}[1, 0, 0]^T \\ \tilde{\mathbf{m}}_{12} &= \mathbf{T}_2^{-1}[0, 1, 0]^T, & \tilde{\mathbf{m}}_{22} &= \mathbf{T}'_2^{-1}[0, 1, 0]^T \\ \tilde{\mathbf{m}}_{13} &= \mathbf{T}_2^{-1}[0, 0, 1]^T, & \tilde{\mathbf{m}}_{23} &= \mathbf{T}'_2^{-1}[0, 0, 1]^T.\end{aligned}$$

c) Estima-se ($\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2i}$), $i = 1, 2$, e 3 , resolvendo o problema (3.19) (método ótimo) ou o (3.20) (método linear).

d) Com as correspondências, ($\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2i}$), obtidas, calcula-se \mathbf{T}_p e \mathbf{T}'_p que satisfaçam (3.9).

e) Estima-se \mathbf{F} resolvendo

$$\min_{\hat{\mathbf{F}}} \sum_j g_j (\tilde{\mathbf{m}}_{2j}^T \mathbf{T}'_p{}^T \hat{\mathbf{F}} \mathbf{T}_p \tilde{\mathbf{m}}_{1j})^2,$$

onde

$$g_j = \left(\frac{1}{\hat{\mathbf{I}}_{1j}^T \mathbf{T}_p \mathbf{Z} \mathbf{T}_p^T \hat{\mathbf{I}}_{1j} + \hat{\mathbf{I}}_{2j}^T \mathbf{T}'_p \mathbf{Z} \mathbf{T}'_p{}^T \hat{\mathbf{I}}_{2j}} \right).$$

f) Transfere-se a matriz fundamental estimada no item anterior para o sistema de coordenadas original, ou seja, $\mathbf{F} = \mathbf{T}'_p{}^T \hat{\mathbf{F}} \mathbf{T}_p$.

Algoritmo 3.2: Novo método não-linear, baseado no paralaxe virtual, para estimação da matriz fundamental.

2. Com uma estimação inicial, \mathbf{F}_0 , da matriz fundamental, calcule os epipolos $\tilde{\mathbf{e}}_1 = [e_{1u}, e_{1v}, e_{1t}]^T$ e $\tilde{\mathbf{e}}_2 = [e_{2u}, e_{2v}, e_{2t}]^T$;
3. Se $|e_{1v}| > |e_{1t}|$, permuta as linhas 2 e 3 de matriz \mathbf{T}_t ;
4. Se $|e_{2v}| > |e_{2t}|$, permuta os elementos e_{2v} e e_{2t} , e as linhas 2 e 3 de \mathbf{T}'_t ;
5. Faça $\mathbf{F}_0 = \mathbf{T}'_t \mathbf{F}_0 \mathbf{T}_t^T$, e;
6. Dado o conjunto de correspondências ($\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2i}$), aplique \mathbf{T}_t e \mathbf{T}'_t à \mathbf{m}_{1i} e \mathbf{m}_{2i} respectivamente.

Os passos anteriores levam a matriz \mathbf{F}_0 e as correspondências ($\mathbf{m}_{1i} \leftrightarrow \mathbf{m}_{2i}$) a um novo sistema de coordenadas, no qual os epipolos de ambas as imagens estão próximos ao centro das imagens. Além disso, como \mathbf{T}_t e \mathbf{T}'_t são transformações de permutação, o custo computacional inserido no problema é muito baixo.

Para conservar a distribuição do ruído, no algoritmo 3.2, a equação (3.24) deve ser substituída por

$$g_j = \left(\frac{1}{\hat{\mathbf{I}}_{1j}^T \mathbf{T}_p \mathbf{T}_t \mathbf{Z} \mathbf{T}_t^T \mathbf{T}_p^T \hat{\mathbf{I}}_{1j} + \hat{\mathbf{I}}_{2j}^T \mathbf{T}'_p \mathbf{T}'_t \mathbf{Z} \mathbf{T}'_t{}^T \mathbf{T}'_p{}^T \hat{\mathbf{I}}_{2j}} \right). \quad (3.25)$$

§3.4 Calibração estéreo com um bastão graduado

As teorias de calibração automática de câmeras surgiram da busca por uma maior flexibilidade no processo. Como visto no capítulo anterior, a um custo computacional extremamente elevado, tais teorias permitem que a calibração seja realizada durante a operação normal da câmera. No caso de um conjunto binocular, as restrições extras tornam menor o custo computacional do processo de calibração automática. Contudo, mesmo na visão estéreo, a autocalibração nos leva a resultados pouco exatos. Por conta disso, na prática, a calibração clássica é muito mais utilizada.

Na calibração clássica, os parâmetros da câmera são obtidos com extrema exatidão. Contudo, neste caso, um objeto (o gabarito de calibração) deve ser posto no campo de visão da câmera sempre que a calibração for necessária. Tais objetos, além de exigir que a operação da câmera seja interrompida, são difíceis de manufacturar, pois exigem que suas dimensões e forma sejam conhecidas com extrema exatidão.

Nas seções seguintes, um novo método de calibração de um conjunto binocular é proposto. Este utiliza, como gabarito de calibração, um bastão graduado, ou seja, um objeto de uma única dimensão. Por usar um gabarito de confecção mais simples e dimensões reduzidas, o método proposto permite efetuar a calibração de forma flexível, mas com exatidão.

3.4.1 Exeqüibilidade Matemática

Uma vez que a matriz fundamental de um sistema binocular tenha sido estimada, uma calibração projetiva pode ser realizada facilmente (seção 2.5.2, pág. 42). Além disso, tal calibração pode ser atualizada para uma calibração euclidiana se uma matriz, \mathbf{T}_{PE} (teorema 2.4, pág. 43), for conhecida. O método de calibração proposto estima \mathbf{T}_{PE} utilizando, como gabarito, um bastão de comprimento conhecido que se desloca livremente na frente das câmeras.

A transformada \mathbf{T}_{PE} possui nove graus de liberdade. Por isso, inicialmente, deve ser verificado se um gabarito na forma de um bastão fornece restrições suficientes para que todos os parâmetros de \mathbf{T}_{PE} sejam conhecidos unicamente.

§ Solução com Apenas Dois Pontos

Utilizando, simplesmente, uma haste de tamanho conhecido como gabarito, tem-se apenas dois pontos 3D, \mathbf{M}_{1i} e \mathbf{M}_{2i} (figura 3.5(a)), para auxiliar na calibração. Como as projeções desses pontos pertencem, necessariamente, ao mesmo plano (figura 2.10, pág. 21), as equações em (2.13) (pág. 21) fornecem apenas três restrições para cada ponto 3D. Assim, se a haste deslocar-se livremente, para n deslocamentos, temos $6n$ equações independentes. Por outro lado, \mathbf{M}_{2i} pode ser expresso em função das três coordenadas de \mathbf{M}_{1i} e duas incógnitas, θ_i e ϕ_i , ou seja,

$$\mathbf{M}_{2i} = \mathbf{M}_{1i} + L[\text{sen } \theta_i \cos \phi_i, \text{sen } \theta_i \text{sen } \phi_i, \cos \theta_i]^T, \quad (3.26)$$

onde L é o comprimento da haste de calibração e θ_i e ϕ_i são ângulos que definem a orientação da haste. Dessa forma, para n deslocamentos livres da haste de calibração e considerando os graus de liberdade de \mathbf{T}_{PE} , tem-se um total de $9+5n$ incógnitas. Assim, com $n = 9$ deslocamentos, é possível estimar \mathbf{T}_{PE} unicamente.

Se um ou mais pontos, \mathbf{M}_{ji} , $j \geq 3$, forem acrescentados à haste de calibração (figura 3.5(b)), considerando que as distâncias $d^2(\mathbf{M}_{1i}, \mathbf{M}_{ji})$ e $d^2(\mathbf{M}_{2i}, \mathbf{M}_{ji})$ são conhecidas, cada ponto \mathbf{M}_{ji} pode ser escrito em função das coordenadas dos pontos extremos, \mathbf{M}_{1i} e \mathbf{M}_{2i} , ou seja,

$$\mathbf{M}_{ji} = \lambda_{1j}\mathbf{M}_{1i} + \lambda_{2j}\mathbf{M}_{2i}, \quad (3.27)$$

onde λ_{1j} e λ_{2j} são constantes conhecidas que dependem de $d^2(\mathbf{M}_{1i}, \mathbf{M}_{ji})$ e $d^2(\mathbf{M}_{2i}, \mathbf{M}_{ji})$. Assim, a inserção de um ou mais pontos à haste de calibração não acrescenta nenhuma incógnita ao sistema. Contudo, como discutido a seguir, esses pontos extras fornecem mais restrições que reduzem o número de deslocamentos necessários à calibração.

§ Solução com Três Pontos

Considere um bastão com uma graduação de três pontos, \mathbf{M}_{1i} , \mathbf{M}_{2i} e \mathbf{M}_{3i} (figura 3.5(b)), com posições relativas conhecidas. Para n deslocamentos do bastão, os pontos \mathbf{M}_{1i} e \mathbf{M}_{2i} fornecem $6n$ equações independentes. Contudo, desde que o terceiro ponto deve ser co-linear a \mathbf{M}_{1i} e \mathbf{M}_{2i} , o ponto \mathbf{M}_{3i} fornece apenas n equações independentes. Dessa forma, tem-se um total de $6n + n = 7n$ equações e, como antes, apenas $9 + 5n$ incógnitas. Assim, a calibração é possível com apenas cinco deslocamentos e cinco pares de imagens estéreo.

§ Solução com Quatro ou Mais Pontos

Considere um bastão com quatro pontos, \mathbf{M}_{ji} , $j = 1, 2, 3$ e 4 , com posições relativas conhecidas. O ponto \mathbf{M}_{4i} , entre os pontos extremos da haste, deve satisfazer a restrição de razão cruzada (seção A.1, pág. 116) com os pontos \mathbf{M}_{1i} , \mathbf{M}_{2i} e \mathbf{M}_{3i} . Dessa forma, \mathbf{M}_{4i} não fornece equações independentes, pois uma vez que as projeções de \mathbf{M}_{1i} , \mathbf{M}_{2i} e \mathbf{M}_{3i} tenham sido obtidas, a projeção de \mathbf{M}_{4i} é completamente determinada. Com isso, uma quantidade de pontos superior a três não acrescenta nenhuma restrição ao problema. Contudo, na prática, o uso de mais pontos e mais de cinco deslocamentos é aconselhável, pois reduz a influência do ruído.

3.4.2 Bastão com movimento irrestrito: Solução linear

Em um deslocamento (indexado por i) de um bastão de calibração com q pontos marcados, temos uma correspondência ($\mathbf{m}_{1ji} \leftrightarrow \mathbf{m}_{2ji}$) para cada ponto \mathbf{M}_{ji} , onde $j = 1, 2, \dots, q$, pertencente ao gabarito. Como tais correspondências podem ser obtidas com exatidão, após n deslocamentos, a matriz fundamental do sistema pode ser estimada com excelente precisão usando $q \times n$ correspondências e, por exemplo, o método da seção 3.3.

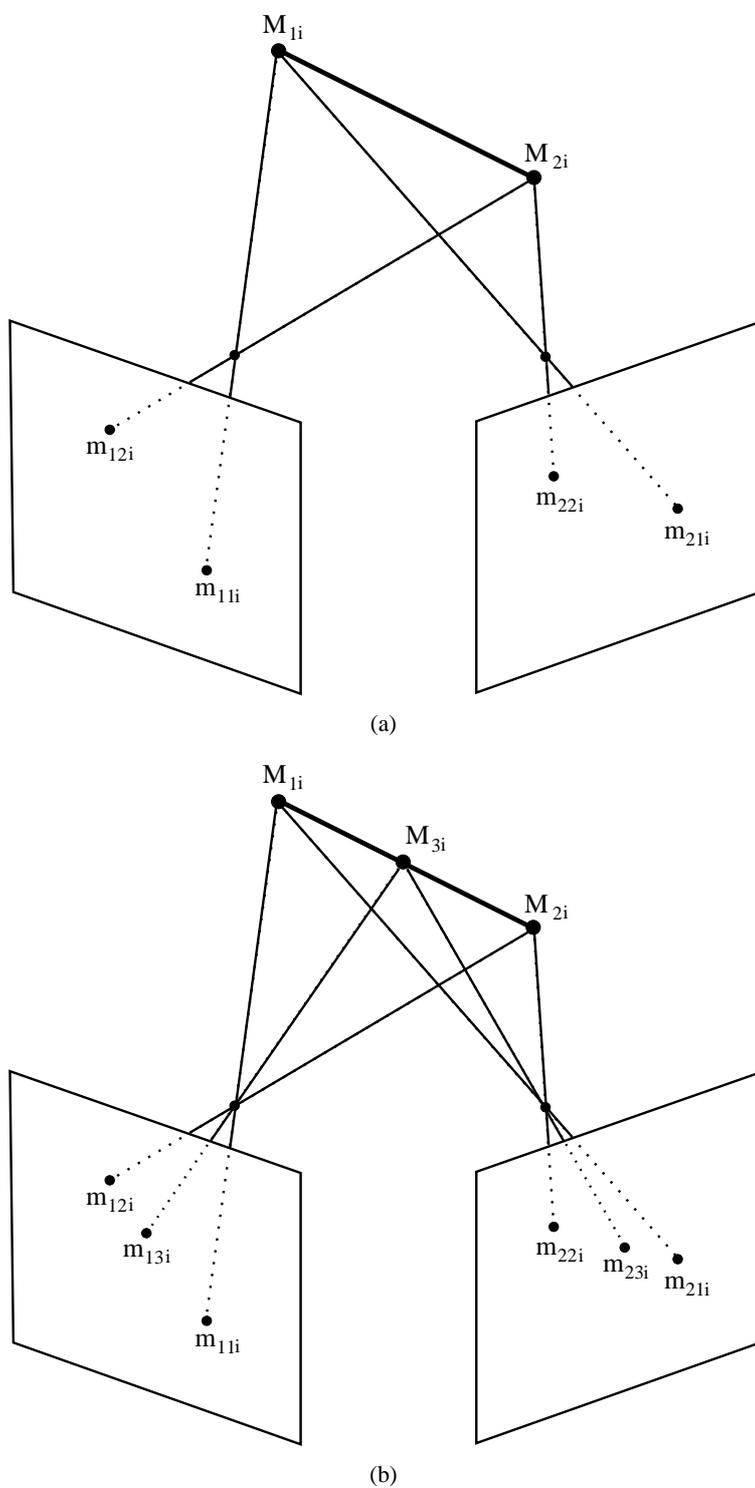


Figura 3.5: Esquemas de calibração estéreo com uma haste, utilizando: (a) um bastão de comprimento conhecido e (b) um bastão com três pontos, cujas posições relativas são conhecidas.

Uma vez que a matriz fundamental for conhecida, ela pode ser utilizada para obtermos \mathcal{P} e \mathcal{P}' . Dessa forma, uma reconstrução projetiva, \mathcal{M}_{ji} , de cada ponto \mathbf{M}_{ji} pode ser obtida. Além disso, da equação (2.51), temos

$$\widetilde{\mathbf{M}}_{ji} \simeq \mathbf{T}_{PE} \widetilde{\mathcal{M}}_{ji}.$$

A equação anterior pode ser transformada em uma igualdade se considerarmos que \mathbf{T}_{PE} é dada por (2.52), ou seja,

$$\mathbf{M}_{ji} = \frac{1}{\widetilde{\mathcal{M}}_{ji}^T \mathbf{W}} \mathbf{A}^{-1} \mathcal{M}_{ji}. \quad (3.28)$$

Como o comprimento, L , e a graduação (distância entre um ponto e outro) do bastão são conhecidos, as coordenadas dos pontos \mathbf{M}_{ji} , para $j = 3, 4 \dots q$, podem ser expressas como função dos pontos extremos, ou seja,

$$\mathbf{M}_{ji} = \lambda_{1j} \mathbf{M}_{1i} + \lambda_{2j} \mathbf{M}_{2i}, \quad (3.29)$$

onde λ_{1j} e λ_{2j} dependem da graduação do bastão, mas são conhecidos.

Substituindo (3.28) na equação anterior e multiplicando ambos os lados da equação por \mathbf{A} , temos

$$\frac{1}{\widetilde{\mathcal{M}}_{ji}^T \mathbf{W}} \mathcal{M}_{ji} = \lambda_{1j} \frac{1}{\widetilde{\mathcal{M}}_{1i}^T \mathbf{W}} \mathcal{M}_{1i} + \lambda_{2j} \frac{1}{\widetilde{\mathcal{M}}_{2i}^T \mathbf{W}} \mathcal{M}_{2i}.$$

Após alguma manipulação algébrica, a equação anterior torna-se

$$\mathbf{u}_{ji}^T \mathbf{W} = 0, \quad (3.30)$$

onde

$$\mathbf{u}_{ji}^T = \widetilde{\mathcal{M}}_{1i}^T + \frac{\lambda_{1j} (\mathcal{M}_{1i} \times \mathcal{M}_{ji}) \cdot (\mathcal{M}_{2i} \times \mathcal{M}_{ji})}{\lambda_{2j} (\mathcal{M}_{2i} \times \mathcal{M}_{ji}) \cdot (\mathcal{M}_{2i} \times \mathcal{M}_{ji})} \widetilde{\mathcal{M}}_{2i}^T. \quad (3.31)$$

Considerando um bastão com $q > 2$ pontos, a equação (3.30) nos leva a

$$\mathbf{U}_{qi} \mathbf{W} = \mathbf{0}, \quad (3.32)$$

onde

$$\mathbf{U}_{qi} = [\mathbf{u}_{3i}, \mathbf{u}_{4i} \dots \mathbf{u}_{qi}]^T.$$

Para $q \geq 4$, a equação (3.32) possui apenas duas equações independentes. Por isso, desde que \mathbf{W} possui três graus de liberdade, no mínimo, $n \geq 2$ deslocamentos são necessários para estimar \mathbf{W} unicamente. Além disso, com mais deslocamentos podemos usar redundância para reduzir a influência do ruído resolvendo

$$\begin{bmatrix} \mathbf{U}_{q1} \\ \mathbf{U}_{q2} \\ \vdots \\ \mathbf{U}_{qn} \end{bmatrix} \mathbf{W} = \mathbf{0}. \quad (3.33)$$

O sistema de equações anterior pode ser resolvido com a proposição B.2. Além disso, os pontos

\mathbf{M}_{ji} podem ser normalizados pela transformação \mathbf{T}_3 (seção 2.5.1, pág. 39). Neste caso, o vetor \mathbf{W} no sistema de coordenadas original pode ser obtido utilizando-se o lema 2.5.1 (pág. 54).

Uma vez que \mathbf{W} foi estimado, considerando apenas a câmera que está na origem do sistema de coordenadas, temos as seguintes relações

$$\begin{aligned}\mathbf{M}_{1i} &= Z_{1i} \mathbf{A}^{-1} \tilde{\mathbf{m}}_{11i}, \\ \mathbf{M}_{2i} &= Z_{2i} \mathbf{A}^{-1} \tilde{\mathbf{m}}_{12i}, \\ \mathbf{M}_{ji} &= Z_{ji} \mathbf{A}^{-1} \tilde{\mathbf{m}}_{1ji},\end{aligned}\tag{3.34}$$

onde Z_{xi} é a profundidade, ou seja, a coordenada Z de \mathbf{M}_{xi} , e $\tilde{\mathbf{m}}_{1ji}$ é a projeção de \mathbf{M}_{xi} na imagem I_1 .

Desde que \mathbf{W} já é conhecido e considerando (3.28), temos

$$\begin{aligned}Z_{xi} &= S_w \frac{\mathcal{Z}_{xi}}{\widetilde{\mathcal{M}}_{xi}^T \mathbf{W}} \\ &= S_w \beta_{xi},\end{aligned}\tag{3.35}$$

onde \mathcal{Z}_{xi} é a coordenada Z de $\widetilde{\mathcal{M}}_{xi}$ e S_w é uma constante inserida para levar em consideração que \mathbf{W} é um vetor homogêneo. Observe que S_w depende apenas de \mathbf{W} e não dos pontos \mathcal{M}_{xi} .

Agora, como o comprimento, L , e a gradação do bastão são conhecidos, para determinar \mathbf{A}^{-1} , podemos utilizar as equações a seguir

$$\begin{aligned}\|\mathbf{M}_{2i} - \mathbf{M}_{1i}\| &= L, \\ \alpha_{2j} \|\mathbf{M}_{2i} - \mathbf{M}_{ji}\| &= L, \\ \alpha_{1j} \|\mathbf{M}_{ji} - \mathbf{M}_{1i}\| &= L,\end{aligned}$$

onde $\alpha_{xi} = 1/(1 - \lambda_{xj})$.

Se substituirmos (3.34) nas equações anteriores, temos

$$\begin{aligned}\|S_w \mathbf{A}^{-1} (\beta_{2i} \tilde{\mathbf{m}}_{12i} - \beta_{1i} \tilde{\mathbf{m}}_{11i})\| &= L, \\ \alpha_{2j} \|S_w \mathbf{A}^{-1} (\beta_{2i} \tilde{\mathbf{m}}_{12i} - \beta_{ji} \tilde{\mathbf{m}}_{1ji})\| &= L, \\ \alpha_{1j} \|S_w \mathbf{A}^{-1} (\beta_{ji} \tilde{\mathbf{m}}_{1ji} - \beta_{1i} \tilde{\mathbf{m}}_{11i})\| &= L.\end{aligned}$$

A equação anterior é equivalente a

$$\begin{aligned}\mathbf{M}_{21i}^T \mathbf{B} \mathbf{M}_{21i} &= L^2, \\ (\alpha_{2j})^2 \mathbf{M}_{2ji}^T \mathbf{B} \mathbf{M}_{2ji} &= L^2, \\ (\alpha_{1j})^2 \mathbf{M}_{j1i}^T \mathbf{B} \mathbf{M}_{j1i} &= L^2,\end{aligned}\tag{3.36}$$

onde $\mathbf{M}_{xyi} = (\beta_{xi} \tilde{\mathbf{m}}_{1xi} - \beta_{yi} \tilde{\mathbf{m}}_{1yi})$ e $\mathbf{B} = S_w^2 \mathbf{A}^{-T} \mathbf{A}^{-1}$ é a matriz do cone absoluto (seção 2.5.3, página 46).

Levando em consideração que

$$\mathbf{a}^T \mathbf{B} \mathbf{a} = [a_1^2, 2a_1a_2, a_2^2, 2a_1a_3, 2a_2a_3, a_3^2]^T \mathbf{b}, \quad (3.37)$$

onde $\mathbf{a} = [a_1, a_2, a_3]^T$ e $\mathbf{b} = [b_{11}, b_{12}, b_{22}, b_{13}, b_{23}, b_{33}]^T$, com b_{ij} igual ao elemento da i -ésima linha e j -ésima coluna de \mathbf{B} , o sistema de equações (3.36) pode ser representado por

$$\mathbf{V}_i \mathbf{b} = \mathbf{L}^T \mathbf{L}, \quad (3.38)$$

onde \mathbf{V}_i é uma matriz que depende dos coeficientes dos vetores \mathbf{M}_{xyi} da equação (3.36) como indicado por (3.37) e $\mathbf{L} = [L, L, \dots, L]^T$.

Independente do número de pontos existentes no bastão de calibração, apenas uma das equações em (3.36) é linearmente independente. Então, como \mathbf{B} tem seis graus de liberdade (um de S_w e cinco de \mathbf{A}^{-1}) são necessários $n \geq 6$ deslocamentos para estimar \mathbf{B} unicamente. Assim, com n deslocamentos, podemos usar (B.3) para resolver

$$\begin{bmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \\ \vdots \\ \mathbf{V}_n \end{bmatrix} \mathbf{b} = \mathbf{L}^T \mathbf{L}. \quad (3.39)$$

Uma vez que \mathbf{B} for conhecida, a matriz $S_w \mathbf{A}^{-1}$ pode ser obtida de \mathbf{B} através da decomposição de Cholesky (pág. 120). Em seguida, se necessário, os pontos \mathbf{M}_{1i} e \mathbf{M}_{2i} podem ser obtidos de (3.28) e (3.26) respectivamente.

As projeções, na imagem I_1 , dos pontos do bastão podem ser normalizadas através da transformada \mathbf{T}_2 (seção 2.4.3, pág. 28). Neste caso, os parâmetros intrínsecos no sistema de coordenadas original podem ser obtidos pela equação a seguir

$$\mathbf{A} = \mathbf{T}_2^{-1} \mathbf{A}_n, \quad (3.40)$$

onde \mathbf{A}_n é a matriz da câmera 1 obtida com os pontos normalizados.

O método proposto é resumido no algoritmo 3.3. Infelizmente, tal método é muito instável. De fato, apesar do vetor \mathbf{W} ser estimado de forma satisfatória, na presença de ruído, a matriz \mathbf{B} estimada pelo método proposto não é positivamente definida e, portanto, a decomposição de Cholesky não pode ser aplicada. Por isso, na próxima seção uma solução não-linear para o problema é proposta.

3.4.3 Bastão com movimento irrestrito: Solução não-linear

O método da seção anterior 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. Contudo, desde que o conceito de distância não é válido no espaço \mathcal{P}^3 , os pontos 3D, \mathcal{M}_{ji} , no espaço

Entrada: Todas as projeções dos pontos de um bastão de calibração, ou seja, n correspondências ($\mathbf{m}_{11i} \leftrightarrow \mathbf{m}_{21i}$), n correspondências ($\mathbf{m}_{12i} \leftrightarrow \mathbf{m}_{22i}$) e $n(q-2)$ correspondências ($\mathbf{m}_{1ji} \leftrightarrow \mathbf{m}_{2ji}$).

Saída: \mathbf{P} e \mathbf{P}' , as matrizes de projeções das duas câmeras.

- a) Estime a matriz fundamental, \mathbf{F} , usando todas as correspondências disponíveis.
- b) Usando a equação (2.48) e o teorema 2.3, calcule as matrizes \mathcal{P} e \mathcal{P}' .
- c) Usando o método da seção 2.7 e as matrizes \mathcal{P} e \mathcal{P}' , realize uma reconstrução projetiva e obtenha n pontos 3D \mathcal{M}_{1i} , n pontos \mathcal{M}_{2i} e $n(q-2)$ pontos \mathcal{M}_{ji} .
- d) Normalize os pontos 3D com a transformação \mathbf{T}_3 (seção 2.5.1).
- e) Utilizando a proposição B.2, resolvendo o sistema de equações (3.33) e estime a equação, \mathbf{W} , do plano no infinito.
- f) Faça $\mathbf{W} = \mathbf{T}_3^T \mathbf{W}$.
- g) Normalize as projeções sobre I_1 usando a transformação \mathbf{T}_2 (seção 2.4.3).
- h) Encontre \mathbf{B} resolvendo (3.39).
- i) Encontre \mathbf{A}^{-1} e S_w decompondo \mathbf{B} , usando a proposição B.1.
- j) Faça $\mathbf{A}^{-1} = \mathbf{A}^{-1} \mathbf{T}_2$ e $\mathbf{W} = \mathbf{W} / S_w$.
- k) Monte \mathbf{T}_{PE} usando a equação (2.52), na página 43.
- l) Faça $\mathbf{P} = \mathcal{P} \mathbf{T}_{PE}^{-1}$ e $\mathbf{P}' = \mathcal{P}' \mathbf{T}_{PE}^{-1}$.
- m) Se necessário, encontre os pontos \mathbf{M}_{1i} usando \mathbf{T}_{PE} e (3.28), ou usando \mathbf{A} e (3.34), ou ainda usando \mathbf{P} , \mathbf{P}' e o método da seção 2.7.
- n) Se necessário, encontre os ângulos θ_i e ϕ_i de (3.26).

Algoritmo 3.3: Calibração de um conjunto binocular com um bastão com movimento irrestrito: Solução linear.

projetivo, não podem ser utilizados para definição de um critério de otimização. Assim, tal critério deve envolver as projeções dos pontos do bastão, \mathbf{m}_{1ji} e \mathbf{m}_{2ji} , nas imagens I_1 e I_2 . Uma solução possível, que age nas projeções dos pontos em uma imagem I_x , é

$$\mathcal{J}_{xi}(\mathcal{P}_x) = \left\| \tilde{\mathbf{m}}_{x1i} - \mathcal{P}_x \mathbf{T}_{PE}^{-1} \tilde{\mathbf{M}}_{1i} \right\| + \left\| \tilde{\mathbf{m}}_{x2i} - \mathcal{P}_x \mathbf{T}_{PE}^{-1} \tilde{\mathbf{M}}_{2i} \right\| + \sum_{j=3}^q \left\| \tilde{\mathbf{m}}_{xji} - \mathcal{P}_x \mathbf{T}_{PE}^{-1} \tilde{\mathbf{M}}_{ji} \right\|, \quad (3.41)$$

onde \mathcal{P}_x é uma matriz de projeção canônica [equação (2.48), pág. 42], $\tilde{\mathbf{M}}_{2i}$ é dado por (3.26) e $\tilde{\mathbf{M}}_{ji}$ é dado por (3.29).

Assim, considerando um bastão com q pontos marcados e n pares de imagens estéreo, \mathbf{T}_{PE} pode ser obtida minimizando-se o critério simétrico a seguir

$$\sum_{i=1}^n [\mathcal{J}_{1i}(\mathcal{P}) + \mathcal{J}_{2i}(\mathcal{P}')]. \quad (3.42)$$

A solução da equação anterior envolve a solução de um problema de minimização não-linear. Normalmente, faz-se necessário uma estimativa inicial de todas as incógnitas envolvidas no problema,

ou seja, da matriz \mathbf{A} , do plano \mathbf{W} , dos n pontos \mathbf{M}_{1i} e dos n ângulos θ_i e ϕ_i .

Uma estimação inicial de \mathbf{A} e \mathbf{W} pode ser encontrada pelo método discutido na seção anterior. Contudo, se a matriz \mathbf{B} , a matriz do cone absoluto, não for positivamente definida, \mathbf{A} pode ainda ser obtida considerando-se $\theta_{sk} = \pi/2$, $\mathbf{p}_0 = [u_0, v_0]^T$ e igual ao ponto central de I_1 , $n_{x/y} = 1$, e obtendo-se f , por exemplo, das especificações da câmera dadas pelo fabricante.

Uma vez que uma estimação inicial de \mathbf{A} e \mathbf{W} esteja disponível, os pontos \mathbf{M}_{1i} podem ser obtidos de (3.28) ou (3.34), e os ângulos θ_i e ϕ_i de (3.26).

O método discutido nesta seção é flexível, pois não impõe restrição ao movimento que o bastão de calibração pode realizar. Contudo, na falta de um método estável para obtenção de uma estimação inicial das incógnitas envolvidas no problema, o algoritmo pode convergir para um mínimo local. Além disso, para n pares de imagens, o problema envolve uma otimização não-linear de uma função com $9 + 5n$ incógnitas. Por isso, na próxima seção, propõe-se um método alternativo que insere novas restrições ao problema, mas aumenta a robustez e reduz o número de incógnitas.

3.4.4 Bastão com uma extremidade fixa

Para simplificar o problema, propõe-se uma restrição ao movimento do bastão, ou seja, que uma das extremidades do bastão esteja sempre em um mesmo ponto. Por exemplo, podemos considerar que o ponto \mathbf{M}_{1i} seja fixo, mas que a outra extremidade do bastão, o ponto \mathbf{M}_{2i} (veja a figura 3.5(b)), pode deslocar-se em todas as direções. Isto é equivalente a fazer $\mathbf{M}_{1i} = \mathbf{M}_1$ em todas as equações. Essa consideração simplifica as equações e insere mais robustez ao método.

Em uma seqüência de n imagens estéreo, apesar do ponto \mathbf{M}_1 estar fixo, na presença de ruído, é esperado obter-se n pares de correspondências ($\mathbf{m}_{11i} \leftrightarrow \mathbf{m}_{21i}$), todos ligeiramente diferentes entre si. Então, o primeiro passo para efetuar a calibração é estimar os pontos \mathbf{m}_{11} e \mathbf{m}_{21} , as verdadeiras projeções do ponto 3D \mathbf{M}_1 , respectivamente, nas imagens I_1 e I_2 . Tal estimação pode ser realizada pela média, ou seja,

$$\bar{\mathbf{m}}_{11} = \frac{1}{n} \sum_{i=1}^n \mathbf{m}_{11i} \quad \text{e} \quad \bar{\mathbf{m}}_{21} = \frac{1}{n} \sum_{i=1}^n \mathbf{m}_{21i}. \quad (3.43)$$

A correspondência ($\bar{\mathbf{m}}_{11} \leftrightarrow \bar{\mathbf{m}}_{21}$) não satisfaz, necessariamente, a equação fundamental (2.17). Por isso, devemos encontrar a correspondência ($\mathbf{m}_{11} \leftrightarrow \mathbf{m}_{21}$) que minimize a função a seguir

$$d^2(\bar{\mathbf{m}}_{11}, \mathbf{m}_{11}) + d^2(\bar{\mathbf{m}}_{21}, \mathbf{m}_{21}), \quad \text{sujeito a } \tilde{\mathbf{m}}_{21}^T \mathbf{F} \tilde{\mathbf{m}}_{11}. \quad (3.44)$$

A solução do problema anterior pode ser obtida com o procedimento discutido na seção 2.7.2, na página 60.

Uma vez que as projeções de \mathbf{M}_1 tenham sido estimadas, podemos realizar uma reconstrução projetiva utilizando as correspondências ($\mathbf{m}_{11} \leftrightarrow \mathbf{m}_{21}$), ($\mathbf{m}_{12i} \leftrightarrow \mathbf{m}_{22i}$) e ($\mathbf{m}_{1ji} \leftrightarrow \mathbf{m}_{2ji}$). Em seguida, o

plano \mathbf{W} pode ser estimado resolvendo-se o sistema de equações (3.33), considerando $\mathbf{M}_{1i} = \mathbf{M}_1$ na equação (3.31).

Agora, substituindo as equações em (3.34) na equação (3.29), após uma manipulação algébrica, temos

$$Z_{2i} = -Z_1 \frac{\lambda_{1j}(\tilde{\mathbf{m}}_{11} \times \tilde{\mathbf{m}}_{1ji}) \cdot (\tilde{\mathbf{m}}_{12i} \times \tilde{\mathbf{m}}_{1ji})}{\lambda_{2j}(\tilde{\mathbf{m}}_{12i} \times \tilde{\mathbf{m}}_{1ji}) \cdot (\tilde{\mathbf{m}}_{12i} \times \tilde{\mathbf{m}}_{1ji})}, \quad (3.45)$$

onde Z_{2i} e Z_1 são, respectivamente, as profundidades de \mathbf{M}_{2i} e \mathbf{M}_1 .

A matriz \mathbf{A} pode ser obtida considerando a equação a seguir

$$\|\mathbf{M}_{2i} - \mathbf{M}_{1i}\| = L,$$

ou seja, substituindo (3.34) na equação anterior, temos

$$\|\mathbf{A}^{-1}(Z_{2i}\tilde{\mathbf{m}}_{12i} - Z_1\tilde{\mathbf{m}}_{11})\| = L.$$

Por sua vez, substituindo (3.45) na equação anterior e considerando Z_1 dado por (3.35), temos

$$\|S_w \mathbf{A}^{-1} \mathbf{Y}_{ji}\| = L, \quad (3.46)$$

onde

$$\mathbf{Y}_{ji} = \frac{Z_1}{\tilde{\mathbf{M}}_1^T \mathbf{W}} \left(\tilde{\mathbf{m}}_{11} + \frac{\lambda_{1j}(\tilde{\mathbf{m}}_{11} \times \tilde{\mathbf{m}}_{1ji}) \cdot (\tilde{\mathbf{m}}_{12i} \times \tilde{\mathbf{m}}_{1ji})}{\lambda_{2j}(\tilde{\mathbf{m}}_{12i} \times \tilde{\mathbf{m}}_{1ji}) \cdot (\tilde{\mathbf{m}}_{12i} \times \tilde{\mathbf{m}}_{1ji})} \tilde{\mathbf{m}}_{12i} \right).$$

A equação (3.46) é equivalente a

$$\mathbf{Y}_{ji}^T \mathbf{B} \mathbf{Y}_{ji} = L^2, \quad (3.47)$$

onde, novamente, $\mathbf{B} = S_w^2 \mathbf{A}^{-T} \mathbf{A}^{-1}$.

Para um bastão de calibração com $q \geq 3$ pontos, temos $(q - 2)$ equações iguais a (3.47). Além disso, o resultado em (3.37) permite-nos representar tais equações em um sistema como em (3.38). Em seguida, considerando n pares de imagens, a matriz \mathbf{B} pode ser estimada resolvendo (3.39), onde as n matrizes \mathbf{V}_i dependem de (3.47).

Considerando um único par de imagens, para um bastão com $q \geq 3$ pontos, (3.47) fornece apenas uma equação independente. Novamente, semelhante ao discutido na seção 3.4.2, a matriz \mathbf{B} não é homogênea e tem seis graus de liberdade. Dessa forma, são necessários $n \geq 6$ pares de imagens para estimá-la completamente.

Se as projeções dos pontos do bastão forem normalizadas, o método proposto provê uma ótima solução mesmo na presença de ruído. De fato, de uma forma geral, a solução de (3.39) para este método fornece uma matriz \mathbf{B} positivamente definida. Sendo assim, a matriz \mathbf{A} , com os parâmetros intrínsecos da primeira câmera, é sempre obtida de forma satisfatória. Contudo, se uma exatidão maior for necessária, pode-se minimizar um critério semelhante ao da equação (3.41), ou seja,

$$\mathcal{J}_{xi}(\mathcal{P}_x) = \left\| \tilde{\mathbf{m}}_{11} - \mathcal{P}_x \mathbf{T}_{PE}^{-1} \tilde{\mathbf{M}}_1 \right\| + \left\| \tilde{\mathbf{m}}_{x2i} - \mathcal{P}_x \mathbf{T}_{PE}^{-1} \tilde{\mathbf{M}}_{2i} \right\| + \sum_{j=3}^q \left\| \tilde{\mathbf{m}}_{xji} - \mathcal{P}_x \mathbf{T}_{PE}^{-1} \tilde{\mathbf{M}}_{ji} \right\|. \quad (3.48)$$

Entrada: Todas as projeções dos pontos de um bastão de calibração, ou seja, n correspondências ($\mathbf{m}_{11i} \leftrightarrow \mathbf{m}_{21i}$), n correspondências ($\mathbf{m}_{12i} \leftrightarrow \mathbf{m}_{22i}$) e $n(q-2)$ correspondências ($\mathbf{m}_{1ji} \leftrightarrow \mathbf{m}_{2ji}$)

Saída: \mathbf{P} e \mathbf{P}' , as matrizes de projeções das duas câmeras.

- a) Estime a matriz fundamental, \mathbf{F} , usando todas as correspondências disponíveis.
- b) Usando a equação (2.48) e o teorema 2.3, calcule as matrizes \mathbf{P} e \mathbf{P}' .
- c) Calcule ($\bar{\mathbf{m}}_{11} \leftrightarrow \bar{\mathbf{m}}_{21}$) usando (3.43).
- d) Calcule ($\mathbf{m}_{11} \leftrightarrow \mathbf{m}_{21}$) resolvendo o problema (3.44).
- e) Usando o método da seção 2.7 e as matrizes \mathbf{P} e \mathbf{P}' , realize uma reconstrução projetiva e obtenha o ponto \mathcal{M}_1 , n pontos \mathcal{M}_{2i} e $n(q-2)$ pontos \mathcal{M}_{ji} .
- f) Normalize os pontos 3D com a transformação \mathbf{T}_3 (seção 2.5.1).
- g) Considerando $\mathcal{M}_1 = \mathcal{M}_{1i}$ e utilizando a proposição B.2, resolva o sistema de equações (3.33) e estime a equação, \mathbf{W} , do plano no infinito.
- h) Faça $\mathbf{W} = \mathbf{T}_3^T \mathbf{W}$.
 - i) Normalize as projeções sobre I_1 usando a transformação \mathbf{T}_2 (seção 2.4.3).
 - j) Encontre \mathbf{B} resolvendo (3.39), considerando as n matrizes \mathbf{V}_i dependentes de (3.47).
 - k) Encontre \mathbf{A}^{-1} e S_w decompondo \mathbf{B} , usando a proposição B.1.
 - l) Faça $\mathbf{A}^{-1} = \mathbf{A}^{-1} \mathbf{T}_2$ e $\mathbf{W} = \mathbf{W}/S_w$.
- m) Monte \mathbf{T}_{PE} usando a equação (2.52), na página 43.
- n) Faça $\mathbf{P} = \mathbf{P} \mathbf{T}_{PE}^{-1}$ e $\mathbf{P}' = \mathbf{P}' \mathbf{T}_{PE}^{-1}$.
- o) Se necessário, encontre os pontos \mathbf{M}_{1i} usando \mathbf{T}_{PE} e (3.28), ou usando \mathbf{A} e (3.34), ou ainda usando \mathbf{P} , \mathbf{P}' e o método da seção 2.7.
- p) Se necessário, encontre os ângulos θ_i e ϕ_i de (3.26).

Algoritmo 3.4: Calibração de um conjunto binocular com um bastão com movimento restrito: Solução linear.

Evidentemente, para obter-se uma solução que se ajuste bem aos pontos das duas imagens, o critério anterior deve ser utilizado para minimizar (3.42).

A solução inicial necessária à minimização não-linear pode ser obtida com o método linear proposto no início desta seção e resumido no algoritmo 3.4.

Considerando uma das extremidades do bastão sempre fixa, as coordenadas do ponto \mathbf{M}_1 são as mesmas em cada imagem. Assim, para uma série de n pares de imagens, o número total de incógnitas do problema são 9 parâmetros referentes à transformação \mathbf{T}_{PE} , as 3 coordenadas de \mathbf{M}_1 , os n ângulos θ_i e n ângulos ϕ_i que definem os pontos \mathbf{M}_{2i} segundo (3.26).

§ Estimção do ponto fixo

Na prática, a projeção do ponto fixo do gabarito 1D não precisa estar presente nas imagens, pois esta pode ser estimada a partir dos outros pontos visíveis do gabarito. Isto é possível porque, em uma

dada imagem, as projeções do gabarito definem uma reta l_i . Assim, dadas n imagens, a projeção da extremidade fixa do gabarito é o ponto, \mathbf{m} , que é comum a todas as retas. Assim, tal ponto pode ser estimado resolvendo

$$\mathbf{L}\tilde{\mathbf{m}} = \mathbf{0}, \quad (3.49)$$

onde $\mathbf{L} = [l_1, l_2, \dots, l_n]^T$.

Uma vez que a projeção do ponto fixo das imagens de cada uma das câmeras tenha sido estimada, o resultado pode ser refinado resolvendo-se o problema (3.44).

3.4.5 Detalhes da minimização não-linear

A complexidade da maioria dos métodos de otimização aumenta com o número de variáveis. Por isso, o número elevado de variáveis que precisam ser estimadas é uma grande desvantagem do método proposto. Contudo, como discute-se a seguir, algumas características do problema podem ser utilizadas para reduzir a complexidade significativamente.

Como discutido na seção B.4, na página 122, os algoritmos de otimização consideram a função a ser minimizada igual a $\mathcal{F}(\mathbf{Y}) = \mathbf{X}$, onde \mathbf{Y} são os parâmetros a serem estimados e \mathbf{X} valores medidos. Inicialmente, usando a mesma notação da seção B.4, considere que \mathbf{Y} possa ser particionado, ou seja,

$$\mathbf{Y} = [\mathbf{c}^T, \mathbf{d}^T]^T. \quad (3.50)$$

Assim, a matriz jacobiana, \mathbf{J} , da equação (B.5), na página 123, pode ser dividida em dois blocos, ou seja, $\mathbf{J} = [\mathbf{C}|\mathbf{D}]$, onde $\mathbf{C} = [\partial\mathbf{X}/\partial\mathbf{c}]$ e $\mathbf{D} = [\partial\mathbf{X}/\partial\mathbf{d}]$. Logo, a equação (B.6), na página 123, torna-se

$$\begin{bmatrix} \mathbf{C}^T\mathbf{C} & \mathbf{C}^T\mathbf{D} \\ \mathbf{D}^T\mathbf{C} & \mathbf{D}^T\mathbf{D} \end{bmatrix} \begin{bmatrix} \delta_c \\ \delta_d \end{bmatrix} = \begin{bmatrix} \mathbf{C}^T\epsilon \\ \mathbf{D}^T\epsilon \end{bmatrix}, \quad (3.51)$$

onde $\Delta = [\delta_c^T, \delta_d^T]^T$ (veja a seção B.4).

Após alguma manipulação algébrica, a equação anterior pode ser reescrita como sendo

$$\begin{bmatrix} \mathbf{U} - \mathbf{G}\mathbf{V}^{-1} & \mathbf{0} \\ \mathbf{G}^T & \mathbf{V} \end{bmatrix} \begin{bmatrix} \delta_c \\ \delta_d \end{bmatrix} = \begin{bmatrix} \epsilon_C - \mathbf{G}\mathbf{V}^{-1}\epsilon_D \\ \epsilon_D \end{bmatrix}, \quad (3.52)$$

onde $\mathbf{U} = \mathbf{C}^T\mathbf{C}$, $\mathbf{G} = \mathbf{C}^T\mathbf{D}$, $\mathbf{V} = \mathbf{D}^T\mathbf{D}$, $\epsilon_C = \mathbf{C}^T\epsilon$ e $\epsilon_D = \mathbf{D}^T\epsilon$.

A solução do problema anterior pode ser dividida em duas etapas. A primeira consiste em encontrar δ_c resolvendo

$$(\mathbf{U} - \mathbf{G}\mathbf{V}^{-1}\mathbf{G}^T)\delta_c = \epsilon_C - \mathbf{G}\mathbf{V}^{-1}\epsilon_D \quad (3.53)$$

e, em seguida, δ_d pode ser encontrado resolvendo

$$\mathbf{V}\delta_d = \epsilon_D - \mathbf{G}^T\delta_c. \quad (3.54)$$

Agora, observe a estrutura do problema da seção 3.4.3. A equação (3.41) pode ser vista como tendo a mesma estrutura de (B.4). Neste caso, o vetor \mathbf{X} é composto pelas coordenadas dos pontos que compõem as correspondências ($\mathbf{m}_{1ji} \leftrightarrow \mathbf{m}_{2ji}$). Assim, os elementos de \mathbf{Y} são os parâmetros intrínsecos da câmera 1, a equação do plano no infinito, \mathbf{W} , e as coordenadas dos pontos \mathbf{M}_{ji} .

Observe também que, por exemplo, o parâmetro \mathbf{M}_{23} influencia apenas as coordenadas dos pontos da correspondência ($\mathbf{m}_{123} \leftrightarrow \mathbf{m}_{223}$). Por outro lado, os parâmetros intrínsecos e o vetor \mathbf{W} têm influência em todos os elementos de \mathbf{X} , ou seja, em todas as correspondências de pontos. Então, considerando a equação (3.50), o vetor \mathbf{c} pode ser formado pelos elementos da matriz \mathbf{A} e do vetor \mathbf{W} , e; o vetor \mathbf{d} pode ser feito igual as coordenadas de todos os pontos \mathbf{M}_{ji} . Neste caso, levando em consideração que cada ponto \mathbf{M}_{ji} só influencia quatro elementos do vetor \mathbf{X} , a matriz jacobiana é esparsa. Assim, a equação (B.5) pode ser particionada na seguinte forma

$$\left\| \left[\begin{array}{c} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{array} \right] - \left[\begin{array}{cccc} \mathbf{C}_1 & \mathbf{D}_1 & & \\ & \mathbf{C}_2 & \mathbf{D}_2 & \\ & \vdots & & \ddots \\ & \mathbf{C}_n & & \mathbf{D}_n \end{array} \right] \left[\begin{array}{c} \delta_c \\ \delta_{d_1} \\ \vdots \\ \delta_{d_n} \end{array} \right] \right\|. \quad (3.55)$$

Da equação anterior, se considerarmos $\mathbf{X} = [\mathbf{X}_1^T, \mathbf{X}_2^T, \dots, \mathbf{X}_n^T]^T$, onde cada \mathbf{X}_l refere-se a uma única correspondência ($\mathbf{m}_{1ji} \leftrightarrow \mathbf{m}_{2ji}$), é fácil dividir a resolução dos problemas (3.53) e (3.54) em vários problemas menores. Tal resolução está resumida no algoritmo 3.5.

O resultado mais importante do algoritmo 3.5 é que este tem complexidade $O(n)$. Por outro lado, sem considerar as particularidades do problema de calibração com o gabarito 1D, a complexidade do problema é da ordem de $O(n^3)$. Evidentemente, estas mudanças reduzem o método de calibração proposto a um problema de ordem $O(n)$ também.

Para o problema da seção 3.4.4, a minimização da equação (3.48) também pode ser realizada com auxílio do algoritmo 3.5. Contudo, neste caso, o ponto fixo \mathbf{M}_1 tem influência sobre todas as correspondências, ou seja, sobre todos os elementos do vetor \mathbf{X} . Por isso, as coordenadas desse ponto também devem fazer parte do vetor \mathbf{c} , na equação 3.50. Exceto por esse detalhe, o processo de minimização é idêntico ao descrito para o caso do bastão com movimento irrestrito. .

Entrada: Uma estimaco inicial, particionada em um vetor $\widehat{\mathbf{Y}} = [\mathbf{c}^T, \mathbf{d}_1^T, \dots, \mathbf{d}_n^T]^T$, de todos os parâmetros a serem refinados e um conjunto de correspondências particionadas em um vetor $\mathbf{X} = [\mathbf{X}_1^T, \dots, \mathbf{X}_n^T]^T$.

Saída: O vetor \mathbf{Y}_r que minimiza (B.5).

a) Substituindo $\widehat{\mathbf{Y}}$ em (B.4), encontre $\widehat{\mathbf{X}} = [\widehat{\mathbf{X}}_1^T, \dots, \widehat{\mathbf{X}}_n^T]^T$ e calcule todas as matrizes $\mathbf{C}_l = [\partial \widehat{\mathbf{X}}_l / \partial \mathbf{c}]$ e $\mathbf{D}_l = [\partial \widehat{\mathbf{X}}_l / \partial \mathbf{d}_l]$.

b) Calcule:

$$\mathbf{U} = \sum_{l=1}^n \mathbf{C}_l^T \mathbf{C}_l;$$

$$\mathbf{V} = \text{diag}(\mathbf{V}_1, \dots, \mathbf{V}_n), \text{ onde } \mathbf{V}_l = \mathbf{D}_l^T \mathbf{D}_l;$$

$$\mathbf{G} = [\mathbf{G}_1, \dots, \mathbf{G}_n], \text{ onde } \mathbf{G}_l = \mathbf{C}_l^T \mathbf{D}_l;$$

$$\boldsymbol{\epsilon}_C = \sum_{l=1}^n \mathbf{C}_l^T \boldsymbol{\epsilon}_l, \mathbf{e};$$

$$\boldsymbol{\epsilon}_D = [\boldsymbol{\epsilon}_{D_1}^T, \dots, \boldsymbol{\epsilon}_{D_n}^T]^T, \text{ onde } \boldsymbol{\epsilon}_{D_l} = \mathbf{D}_l^T \boldsymbol{\epsilon}_l.$$

c) Calcule $\boldsymbol{\delta}_c$ resolvendo

$$\left(\mathbf{U} - \sum_{l=1}^n \mathbf{G}_l \mathbf{V}_l^{-1} \mathbf{G}_l^T \right) \boldsymbol{\delta}_c = \boldsymbol{\epsilon}_C - \sum_{l=1}^n \mathbf{G}_l \mathbf{V}_l^{-1} \boldsymbol{\epsilon}_{D_l}.$$

d) Calcule cada $\boldsymbol{\delta}_{d_l}$ da equaco

$$\boldsymbol{\delta}_{d_l} = \mathbf{V}_l^{-1} (\boldsymbol{\epsilon}_{D_l} - \mathbf{G}_l^T \boldsymbol{\delta}_c).$$

e) Compute os parâmetros refinados, $\widehat{\mathbf{Y}}_r = \widehat{\mathbf{Y}} + [\boldsymbol{\delta}_c^T, \boldsymbol{\delta}_{d_1}^T, \dots, \boldsymbol{\delta}_{d_n}^T]^T$.

f) Repita todos os passos até a convergência do resíduo.

Algoritmo 3.5: Algoritmo de Newton personalizado para o problema de calibração com uma haste graduada.

RESULTADOS EXPERIMENTAIS

Neste capítulo, os algoritmos propostos são avaliados. O objetivo é conhecer a influência que alguns fatores como, por exemplo, a quantidade das correspondências de pontos utilizadas nos cálculos, exercem sobre os resultados. Além disso, é feita uma comparação dos métodos propostos com outros algoritmos bem conhecidos na literatura. Tal comparação é feita, sobretudo, com relação a exatidão.

§4.1 Correspondência de pontos

Nesta seção, o método de correspondência de pontos proposto é avaliado. Para isto, são utilizados os três pares de imagens da figura 4.1. Cada imagem tem dimensão de 640×480 pixels. Por comodidade, o par de imagens da figura 4.1(a) é referenciado no texto por **MESA** e os das figuras 4.1(b) e 4.1(c), respectivamente, por **PLANTA** e **DESKTOP**. Em particular, o par **DESKTOP** apresenta uma distorção projetiva elevada. Isto ocorreu porque as câmeras que formavam o conjunto binocular possuíam distâncias focais bem distintas. Este fato dificulta ainda mais o processo de correspondência de pontos.

O algoritmo SUSAN (SMITH, 1992) foi utilizado para extrair cantos de cada uma das imagens. São esses cantos que os métodos avaliados devem tentar estabelecer correspondência de pontos.

O desempenho do algoritmo proposto é comparado ao do método de Zhang et al. (1995). Tal método é um dos mais citados na literatura e é tido como um dos mais robustos. Outro fator importante é que o método de Zhang et al. (1995) também é dividido em três etapas: formação do conjunto de candidatos a correspondência; eliminação da ambigüidade, e; identificação das falsas correspondências. Contudo, ao contrário do método proposto, o método de Zhang et al. (1995) não utiliza transformações não-paramétricas. Por isso, a correlação dada pela equação (2.9), na pág. 14, tem que ser usada para formar o conjunto de candidatos a correspondência. Outro diferencial é o uso de técnicas iterativas de relaxação para eliminar a ambigüidade. Tal técnica utiliza uma equação bastante complexa (equação (2.11), na pág. 17) para calcular o grau de confiança dos candidatos



(a)



(b)



(c)

Figura 4.1: Imagens utilizadas na avaliação dos métodos de correspondência de pontos: (a) é o par chamado **MESA**; (b) é o par **PLANTA**, e; (c) é referenciado como **DESKTOP**.

à correspondência. Isto contribui para que esta seja a fase do algoritmo que requer mais tempo de processamento.

Ambos os métodos testados foram implementados na forma de um programa para o MATLAB. Ao contrário do proposto por Zhang et al. (1995), a implementação do algoritmo **ZHANG** utilizada neste trabalho usa o método MSAC para identificar as falsas correspondências. Isto foi necessário por dois motivos: (a) o método LMedS (usado no artigo original de Zhang et al. (1995)) não fornece bons resultados se a quantidade de falsas correspondências é maior que 50 %, e; (b) já que o método proposto também usa o MSAC, isto facilita a comparação dos resultados.

O algoritmo **ZHANG** foi implementado com os parâmetros recomendados no artigo de Zhang et al. (1995). Para o método proposto, a transformada ceno foi aplicada às imagens com uma janela W_n de 5×5 pixels. Com isso, cada pixel da imagem é transformado em uma seqüência de 24 bits. Para formar o conjunto de candidados à correspondência, utilizou-se uma janela de correlação de 11×11 pixels. Além disso, dado um determinado ponto, $\mathbf{m}_{1i} \in I_1$, a correlação deste é calculada com todos os pontos pertencentes a I_2 e que estão dentro de uma janela quadrada de busca centrada em \mathbf{m}_{1i} , com largura igual a $1/4$ da largura da imagem. Estes são valores empíricos, mas que funcionaram bem para todos os casos testados.

Para a fase de eliminação da ambigüidade, o método proposto foi implementado considerando $\varepsilon_r = 0,04$, $\theta_{th} = \pi/2$ e $rb_{th} = 2,5\%$ do número de cantos extraídos da imagem 1. O parâmetro ε_r depende da “semelhança” entre as câmeras que compõem o conjunto binocular, ou seja, quanto mais distintos (entre si) forem os parâmetros das duas câmeras, maior deve ser o valor de ε_r . Por exemplo, no par **DESKTOP**, as câmeras possuem distâncias focais bem distintas. Por isso, para este par, ε_r foi feito igual a 0,10. Outro importante fator que influencia a escolha de ε_r é a separação entre as câmeras. Uma separação maior implica em um ε_r também maior. Por isso, esse parâmetro deve ser escolhido experimentalmente para cada conjunto binocular. Contudo, uma vez que as câmeras permanecem fixas durante toda a operação normal do conjunto binocular, um ε_r bem escolhido servirá adequadamente durante todo o tempo que o sistema for utilizado.

Na tabela 4.1 são apresentados, para cada par de imagens, a quantidade de correspondências estabelecidas em cada fase pelos métodos testados. Além disso, o ajuste da matriz fundamental estimada pelo algoritmo MSAC ao conjunto final de correspondências (dado pela equação (2.30), na página 26) também é dado. Este ajuste fornece uma indicação da qualidade do conjunto de correspondências encontrado. Um ajuste pequeno (menor que um) indica um bom conjunto de correspondências.

Observando a tabela 4.1, exceto na primeira etapa, nota-se que os resultados de ambos os métodos são bem semelhantes. Em particular, ambos têm dificuldade em estabelecer correspondências de pontos com o par **DESKTOP**. Contudo, como os métodos utilizam semelhança para formar o conjunto inicial de correspondências, isto já é esperado. Mesmo assim, os métodos conseguem estabelecer, em todos os casos, um conjunto de correspondências que se ajusta bem à geometria epipolar do conjunto binocular e que tem mais que 100 elementos. Como mostram os resultados das próximas seções, este número é suficiente para, por exemplo, estimar a matriz fundamental com exatidão.

A tabela 4.2 mostra o tempo de execução dos algoritmos. Observa-se que o algoritmo **ZHANG**

Par de imagens	Método utilizado	Número de correspondências			Ajuste
		Etapa 1	Etapa 2	Etapa 3	
MESA (2.079 × 1.982)	ZHANG	32.579	1.130	572	0,48
	PROPOSTO	2.079	1.083	524	0,53
PLANTA (2.731 × 2.833)	ZHANG	102.972	1.543	777	0,41
	PROPOSTO	2.731	1.625	770	0,36
DESKTOP (3.000 × 2.817)	ZHANG	158.532	1.499	230	0,57
	PROPOSTO	2.290	1.034	141	0,48

Tabela 4.1: Quantidade de correspondências encontradas em cada etapa do algoritmo pelos métodos analisados. Na primeira coluna, a quantidade de cantos encontrada em cada imagem do par é apresentada entre parenteses.

Par de imagens	Método utilizado	Tempo de execução (s)			
		Etapa 1	Etapa 2	Etapa 3	Total
MESA	ZHANG	18,03	79,01	3,96	107,02
	PROPOSTO	7,44	2,56	3,90	13,90
PLANTA	ZHANG	24,94	481,75	3,39	515,08
	PROPOSTO	11,35	15,03	8,75	35,13
DESKTOP	ZHANG	29,34	1.644,21	194,98	1.868,53
	PROPOSTO	16,48	21,18	167,85	205,51

Tabela 4.2: Tempo (em segundos) necessário para os métodos analisados concluírem cada um das etapas dos algoritmos.

necessita de um tempo muito maior para concluir as etapas 1 e 2, respectivamente, formação do conjunto de correspondências e eliminação da ambigüidade. Por isso, o algoritmo proposto é executado, em média, em um tempo total dez vezes menor. Isto acontece, principalmente, devido ao grande número de candidatos à correspondência encontrados na primeira etapa do processo. Evidentemente, isto sobrecarrega a etapa de eliminação da ambigüidade. Por outro lado, no método proposto, este processo é feito de forma bastante simples, não-iterativa e utilizando apenas operações com números inteiros.

§4.2 Estimação da Matriz Fundamental

Comparar o desempenho de dois ou mais métodos para cálculo da matriz fundamental não é uma tarefa óbvia. Isto deve-se a dois fatores. Primeiro, a matriz fundamental pode ser decomposta em seus epipolos e na homografia epipolar (seção 2.3.1, pág. 23). Contudo, a homografia epipolar pertence ao espaço \mathcal{P}^2 , que não possui nenhuma métrica (o que torna difícil definir uma medida de ajuste). Já os epipolos, apesar de poderem ser considerados parâmetros euclidianos, são muito susceptíveis a ruído e, por isso, muito difíceis de serem estimados com exatidão (sobretudo quando tendem ao infinito) (LUONG; FAUGERAS, 1998). Além disso, o desempenho dos métodos é muito dependente de fatores, tais como, quantidade de correspondências utilizadas na estimação, distribuição de tais correspondências no plano de imagem e localização dos epipolos. Por isso, contrariando diversos trabalhos encontrados na literatura (ARMANGUÉ; SALVI, 2003; TORR; FITZGIBBON, 2003; ZHANG; LOOP,

2001; ZHANG, 1998), os métodos avaliados não podem ser comparados baseados em um único par de imagens-teste ou em um único experimento. Em vista disso, neste trabalho será utilizada uma metodologia de testes semelhante à realizada por Hartley (1997), ou seja,

- a) Em cada experimento, são utilizadas N correspondências de pontos. Em todos os casos, considera-se que o conjunto de correspondências possua ruído com distribuição gaussiana de média zero e variância σ .
- b) Dentre as N correspondências, 50 subconjuntos (indexados por j) são selecionados aleatoriamente. Todos os subconjuntos possuem a mesma quantidade, n_l , de correspondências.
- c) A cada iteração, cada método avaliado utiliza todos os 50 subconjuntos escolhidos (um por vez) e estima 50 matrizes, \mathbf{F}_j , diferentes. Além disso, para cada método, é calculado o seguinte resíduo

$$\bar{r}_j^2 = \frac{1}{N} \sum_{i=1}^N d^2(\tilde{\mathbf{m}}_{2i}, \mathbf{F}_j \tilde{\mathbf{m}}_{1i}) + d^2(\tilde{\mathbf{m}}_{1i}, \mathbf{F}_j^T \tilde{\mathbf{m}}_{2i}).$$

- d) Após os 50 subconjuntos serem aplicados, a mediana \bar{R}_l^2 é calculada para cada método, onde

$$\bar{R}_l^2 = \text{mediana}\{\bar{r}_1^2, \dots, \bar{r}_{50}^2\}. \quad (4.1)$$

- e) Em seguida, o número de correspondências, n_l , é aumentado e outros 50 subconjuntos são escolhidos aleatoriamente. No total, n_l assume 43 valores diferentes (dentro do intervalo de 8 a 100). Dessa forma, para cada par de imagens, cada método é testado $43 \times 50 = 2.150$ vezes, com diferentes números de correspondências.

Com essa metodologia, comparando-se os resíduos \bar{R}_l^2 obtidos por métodos diferentes, tem-se uma boa medida de qual deles estimou a matriz fundamental que melhor ajustou-se ao conjunto total de correspondências, ou seja, o método com menor resíduo é o que obteve o melhor resultado.

Em todos os casos, o método de Levenberg-Marquardt (LEVENBERG, 1944; MARQUARDT, 1963) foi utilizado sempre que um algoritmo de otimização era necessário. A implementação deste método utilizada neste trabalho está detalhada na seção B.4 (pág. 122). Em todos os casos, a tolerância utilizada nos cálculos foi sempre igual a 1×10^{-8} .

Por último, o algoritmo de Levenberg-Marquardt exige uma estimativa inicial dos parâmetros procurados. Para isto, neste trabalho, o algoritmo dos oito pontos (HARTLEY, 1997) foi utilizado em todos os casos.

Diversos testes foram realizados para avaliar, na prática, a exatidão dos métodos propostos e compará-la às de outros métodos existentes na literatura. Os métodos escolhidos para comparação são listados a seguir.

4.2.1 Métodos Avaliados

Além dos métodos propostos e apresentados no capítulo anterior, foram também avaliados alguns métodos clássicos da literatura e outros mais recentes, ou seja, o de Bartoli e Sturm (2004) (referenciando neste trabalho como **BARTOLI**); o de Zhang e Loop (2001) (referenciado como **ZHANG**); o método de Csurka et al. (1997) (chamado aqui de **CSURKA**); o método referenciado por **GRADIENT**, que minimiza o Erro de Sampson (SAMPSON, 1982) e utiliza a parametrização proposta por Csurka et al. (1997), e, finalmente; o método **GOLD** (apresentado na seção 2.4.4, pág. 32).

Os métodos avaliados foram testados nas imagens apresentadas na figura 4.2. Como pode-se observar pelo esboço das linhas epipolares, os epipolos de cada par de imagens estão localizados em regiões bem distintas, ou seja, próximos ao centro das imagens (par referenciado por **GABARITO**), próximo a borda das imagens (par **COMPUTADOR**) e tendendo ao infinito (par **QUARTO**). Isto é necessário, pois, como comprovam os resultados das próximas seções, alguns métodos são muito influenciados pela localização dos epipolos.

Inicialmente, comparou-se o desempenho dos métodos **CSURKA**, **GRADIENT** e **GOLD**. Como discutido no capítulo 2, estes métodos são tidos na literatura como os que fornecem os melhores resultados. Contudo, a matriz fundamental é obtida a um custo computacional extremamente elevado. Nos gráficos da figura 4.3, apresentam-se os resultados dos testes. Nesses gráficos, assim como em todos os gráficos desta e das próximas seções, a ordenada é apresentada em escala logarítmica.

As figuras 4.3(a) e 4.3(b) mostram os resultados para os testes realizados com as correspondências extraídas do par **GABARITO**. Já os gráficos das figuras 4.3(c) e 4.3(d), são os resultados para o par de imagens **QUARTO**. Observando tais gráficos, percebe-se que os três métodos têm desempenho equivalentes. Em especial, os métodos **CSURKA** e **GRADIENT** fornecem resultados quase idênticos. De fato, é impossível distinguir as curvas desses dois métodos uma da outra. Outro fato importante é que só se obtém uma boa estimação da matriz fundamental com um conjunto de 20 ou mais correspondências. Quanto ao número de iterações (figura 4.4), como é de se esperar (devido aos 36 espaços de busca diferentes), os métodos **CSURKA** e **GRADIENT** necessitam de muito mais iterações que o **GOLD**.

Devido a semelhança dos resultados obtidos pelos métodos **CSURKA**, **GRADIENT** e **GOLD**, torna-se dispensável compará-los todos aos outros métodos avaliados. Dessa forma, para facilitar a visualização dos próximos resultados apresentados, apenas o método **CSURKA** é considerado nas próximas seções.

4.2.2 Avaliação do método linear proposto

O método iterativo, mas linear, proposto (referenciado aqui como **FRANÇA**) foi avaliado utilizando os três pares de imagens da figura 4.2. Os resultados apresentados nas figuras 4.5(a) e 4.5(b) correspondem as imagens **GABARITO**; os gráficos das figuras 4.5(c) e 4.5(d) correspondem ao par

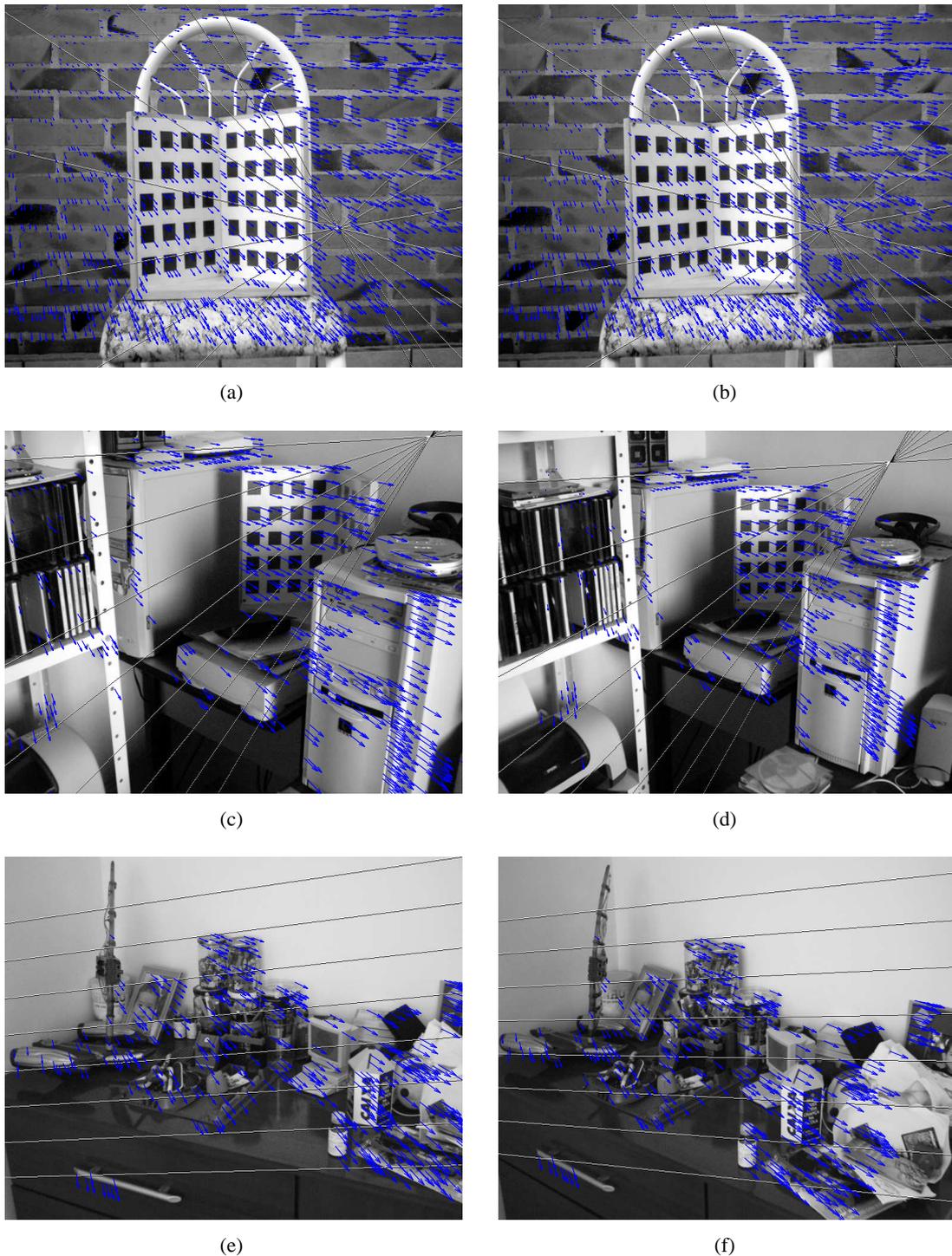


Figura 4.2: Imagens-teste utilizadas na estimação da matriz fundamental: (a) e (b) formam o par **GABARITO**, que possui 1.373 correspondências marcadas e epipolos próximo ao centro das imagens; (c) e (d) compõem o par **COMPUTADOR**, com 495 correspondências e epipolos próximo a borda das imagens, e; (e) e (f) compõem o par **QUARTO**, possuem 403 correspondências e os epipolos tendendo para o infinito.

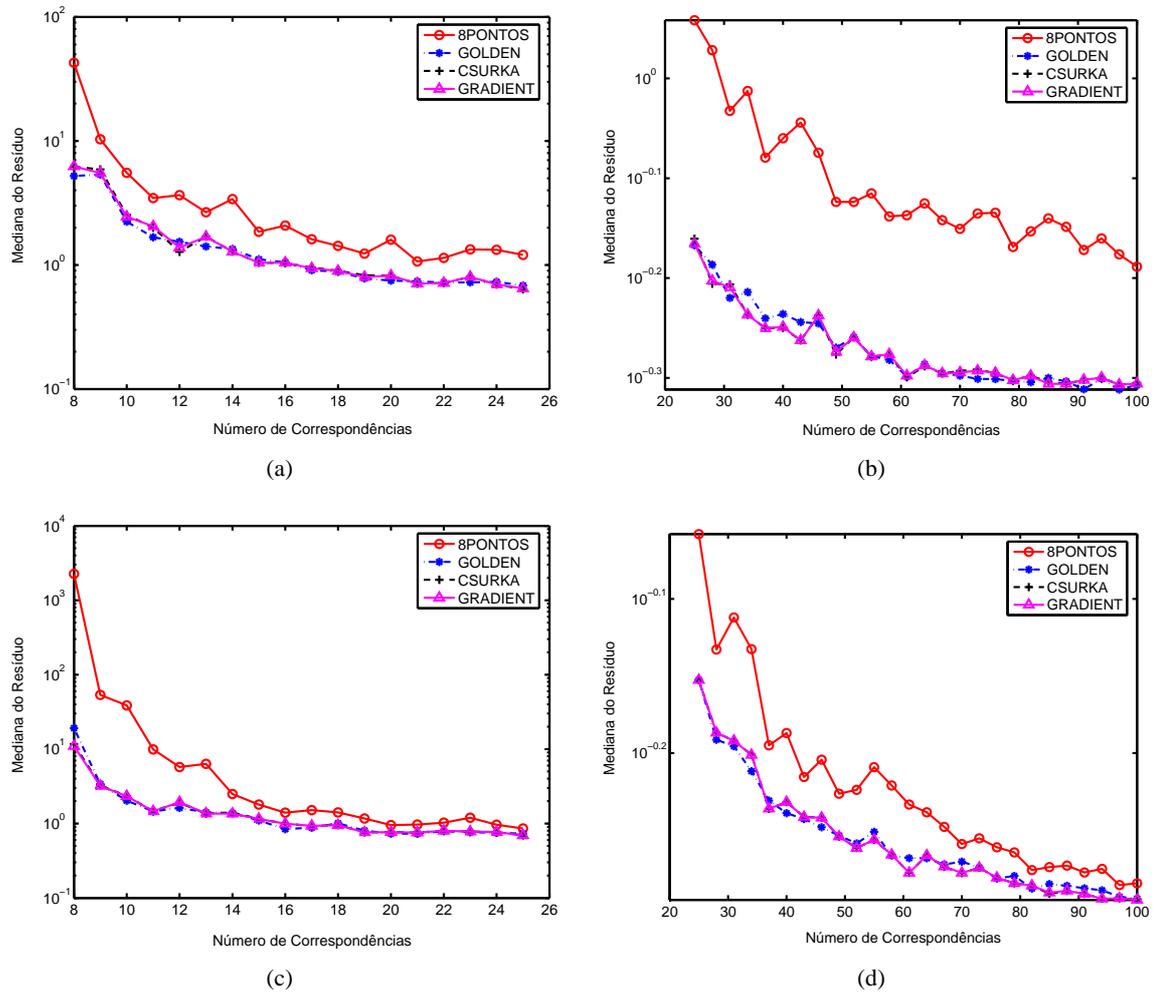


Figura 4.3: Resíduos obtidos com os métodos **8PONTOS**, **GOLDEN**, **CSURKA** e **GRADIENTE**: (a) e (b) para o par **GABARITO**, e; (c) e (d) para o par **QUARTO**.

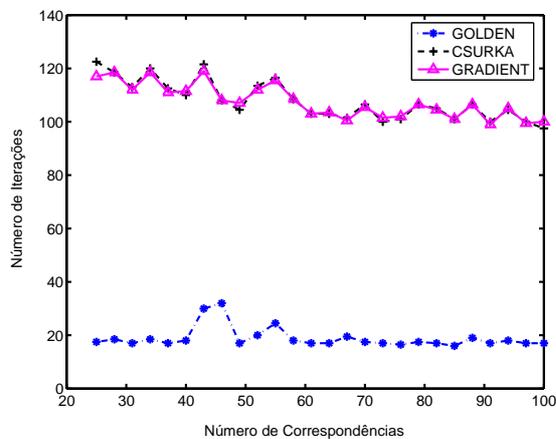


Figura 4.4: Número típico de iterações necessárias por estes métodos **GOLDEN**, **CSURKA** e **GRADIENTE**.

COMPUTADOR, e; os resultados para o par **QUARTO** são apresentados nas figuras 4.5(e) e 4.5(f). Analisando esses gráficos, pode-se observar que o método fornece uma sensível melhoria na matriz fundamental estimada pelo método dos 8 pontos. Contudo, a exatidão obtida com o método **FRANÇA** é inferior daquela obtida com o algoritmo **CSURKA**. Tal diferença torna-se maior à medida que os epípolos tendem para o centro das imagens (veja as figuras 4.5(a) e 4.5(b)). Contudo, desde que o método **FRANÇA** não utiliza métodos de otimização, por exemplo, como o Levenberg-Marquardt, e, normalmente, necessita de apenas quatro iterações para convergir (veja a figura 4.6), o método proposto é válido e demonstra ser uma ótima opção para obter-se resultados com exatidão intermediária.

4.2.3 Avaliação da solução linear do paralaxe virtual

Como discutido no capítulo anterior, a solução linear do problema do paralaxe virtual (referenciada aqui como **PARALAXE1**) não fornece um bom resultado. Isto pode ser comprovado observando os gráficos da figura 4.7. Tais gráficos correspondem ao seguinte experimento. Dado um conjunto de N correspondências, um subconjunto com 100 elementos foi escolhido aleatoriamente. Em seguida, o método do paralaxe virtual foi utilizado para estimar a matriz fundamental com tal subconjunto. Conservando-se o mesmo subconjunto de correspondências, a matriz fundamental foi estimada 30 vezes. Contudo, a cada iteração, bases virtuais diferentes foram escolhidas aleatoriamente. As figuras 4.7(a) e 4.7(b) mostram os resultados para, respectivamente, o par **GABARITO** e o par **QUARTO**. Nessas figuras, a curva **PARALAXE2** refere-se ao resíduo obtido impondo-se a propriedade 2.1 (pág. 22) à matriz fundamental obtida pelo método **PARALAXE1**. Analisando os gráficos da figura 4.7, nota-se que a exatidão do método é bastante afetada pela escolha das bases virtuais. Além disso, tal exatidão é bastante inferior à obtida com o algoritmo dos 8 pontos (também apresentado na figura 4.7). Contudo, com o experimento a seguir, vê-se que, quando resolvido de forma não-linear, o método do paralaxe virtual fornece ótimos resultados.

4.2.4 Avaliação da solução não-linear do paralaxe virtual

Como discutido na seção 4.8, o método não-linear do paralaxe virtual exige três correspondências de pontos virtuais. Para obter tais correspondências, dois métodos foram propostos. O primeiro (referenciado neste texto por **FRANÇA1**) é uma solução linear. Já o segundo (referenciado por **FRANÇA2**), requer que todas as raízes de um polinômio de grau seis sejam encontradas.

Nas figuras 4.8(a) e 4.8(b), a exatidão desses dois métodos são comparadas entre si e ao método **CSURKA**. Analisando os gráficos, observa-se que o resultado não é influenciado pelo modo como as correspondências virtuais são obtidas. De fato, os resíduos obtidos pelos métodos **FRANÇA1** e **FRANÇA2** diferem por um valor da ordem de 10^{-2} . Dessa forma, desde que o método **FRANÇA1** é mais direto, ele deve ser preferido ao método **FRANÇA2**.

Apesar de melhorar sensivelmente a estimação inicial da matriz fundamental (dada pelo algoritmo dos 8 pontos), o método **FRANÇA1** fornece um resultado de qualidade ainda inferior à obtida com o

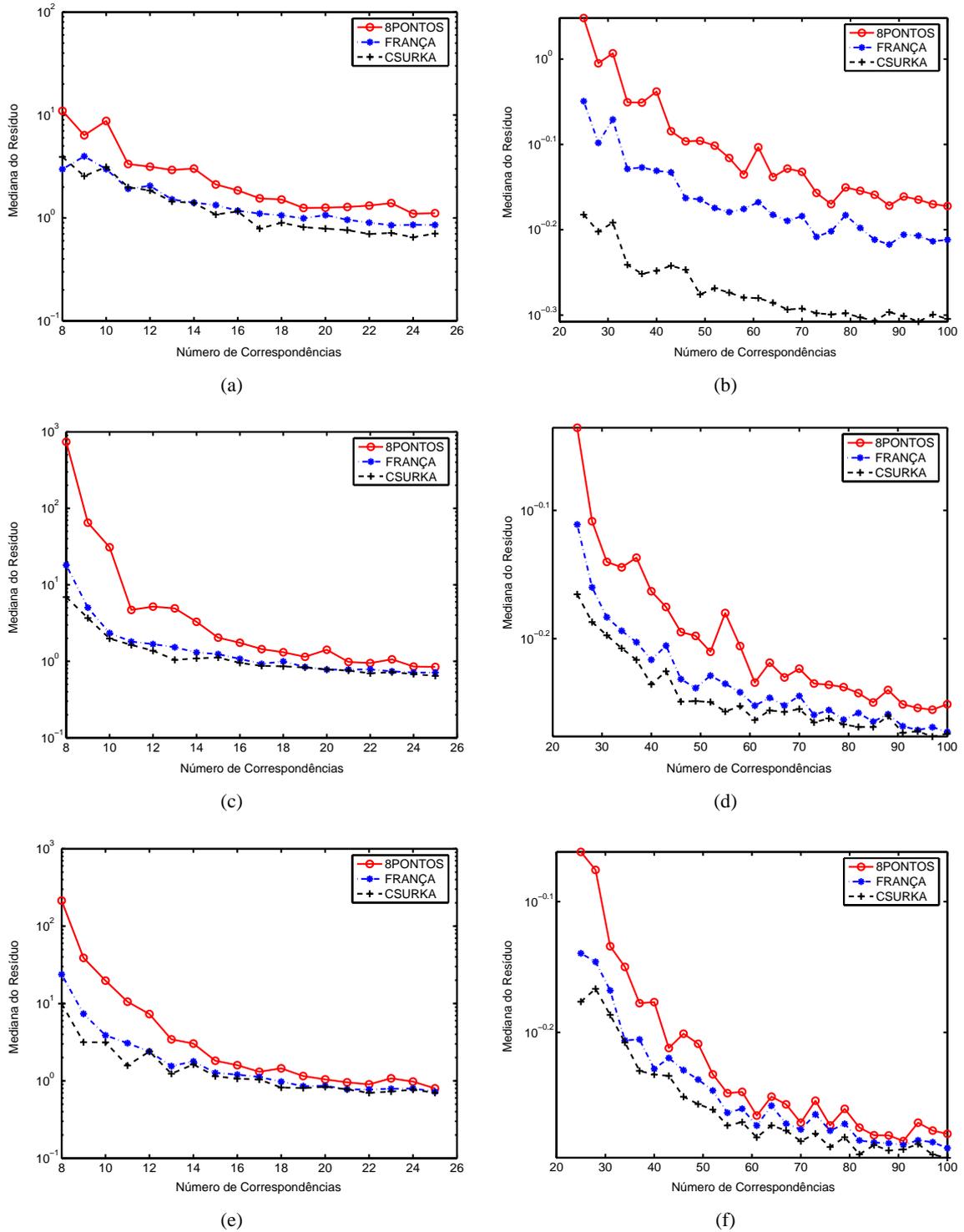


Figura 4.5: Gráficos com os resíduos obtidos com os métodos **8PONTOS**, **FRANÇA** e **CSURKA**: (a) e (b) par **GABARITO**; (c) e (d) par **COMPUTADOR**, e; (e) e (f) par **QUARTO**.

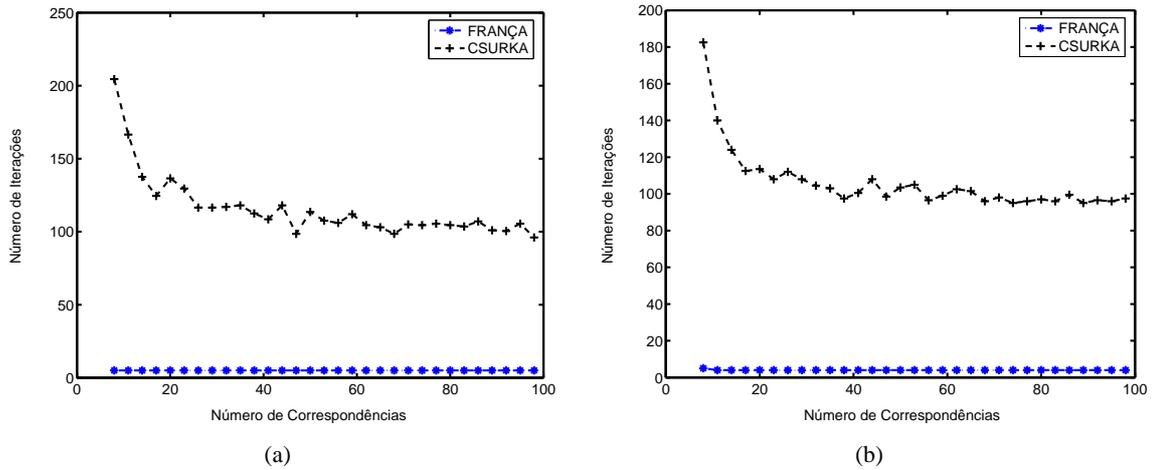


Figura 4.6: Número de iterações requerido pelos métodos **FRANÇA** e **CSURKA** para os experimentos com os pares (a) **GABARITO** e (b) **QUARTO**.

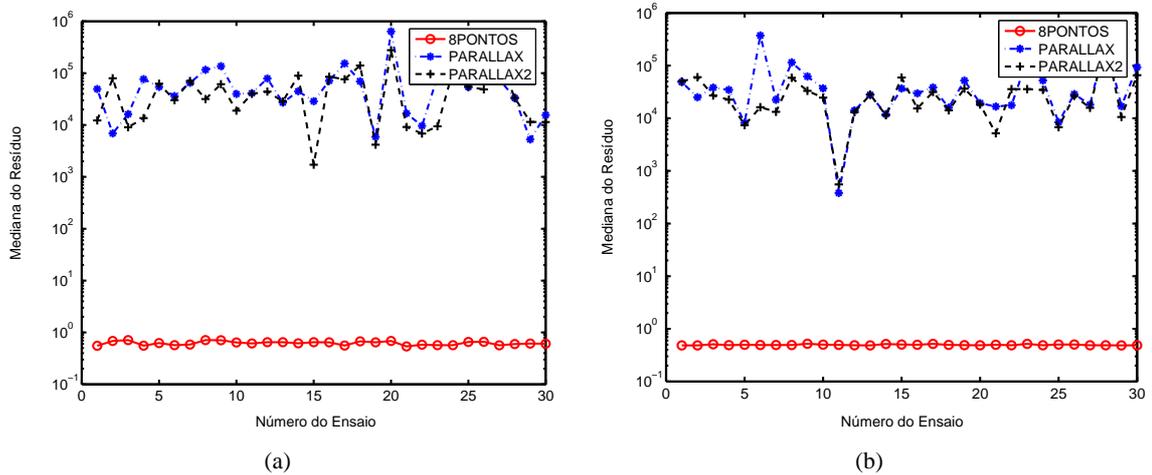


Figura 4.7: Comparação entre o método dos 8 pontos e o método do paralaxe virtual. (a) é o resíduo obtido para as correspondências do par **GABARITO** e (b) é o resíduo para o par **QUARTO**.

método **CSURKA**. Isto ocorre tanto com os epipolos próximos ao centro das imagens (figura 4.8(a)) como quando eles tendem para infinito (figura 4.8(b)).

Outro fato a ser observado é que a qualidade da estimação obtida pelo método **FRANÇA1** é reduzida à medida que os epipolos tendem ao infinito. Como discutido na seção 4.8, este problema pode ser contornado atualizando-se o método **FRANÇA1** para o método **FRANÇA3** (seção 3.3.5, pág. 77). Isto pode ser comprovado observando-se o gráfico da figura 4.8(c) (resultados referentes ao par de imagens **GABARITO**). Além disso, como mostra a figura 4.8(d) (referente ao par **QUARTO**), aplicando-se as transformações T_t e T'_t do método **FRANÇA3**, obtém-se uma “pequena” redução do resíduo mesmo quando os epipolos já tendem ao infinito.

Na figura 4.9, o método **FRANÇA3** é comparado aos métodos **BARTOLI** e **ZHANG**. Os gráficos referem-se aos pares **GABARITO** (figuras 4.9(a) e 4.9(b)), **COMPUTADOR** (figuras 4.9(c) e 4.9(d)) e **QUARTO** (figuras 4.9(e) e 4.9(f)). Como pode-se observar, os resultados obtidos com os três métodos são muito semelhantes. Contudo, como a complexidade do método **FRANÇA3** é menor,

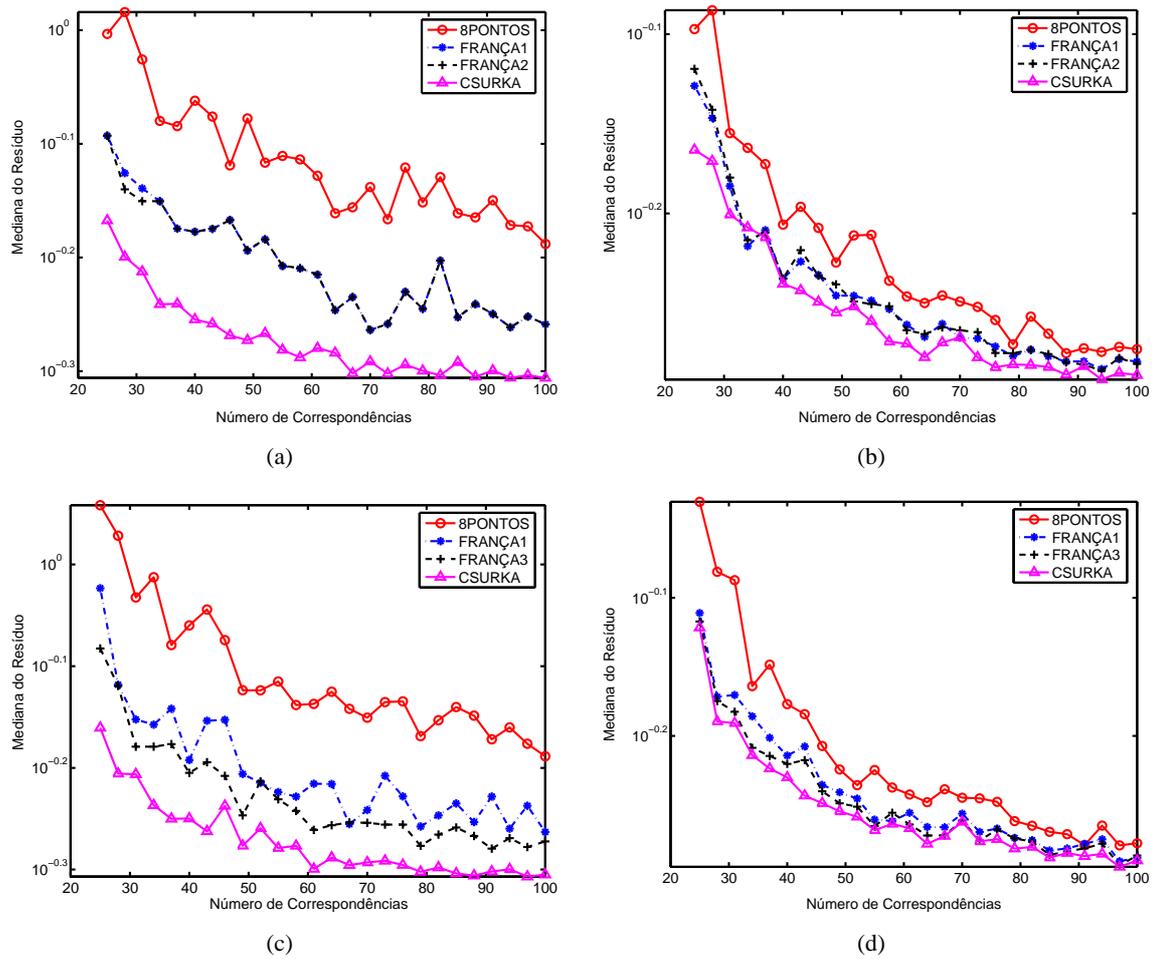


Figura 4.8: Gráficos (a) e (b) mostram o resíduo obtido com os métodos **FRANÇA1** e **FRANÇA2** para os pares (a) **GABARITO** e (b) **QUARTO**, e; os gráficos do resíduo obtido com os métodos **FRANÇA1** e **FRANÇA3** são: (c) par **GABARITO** e (d) par **QUARTO**

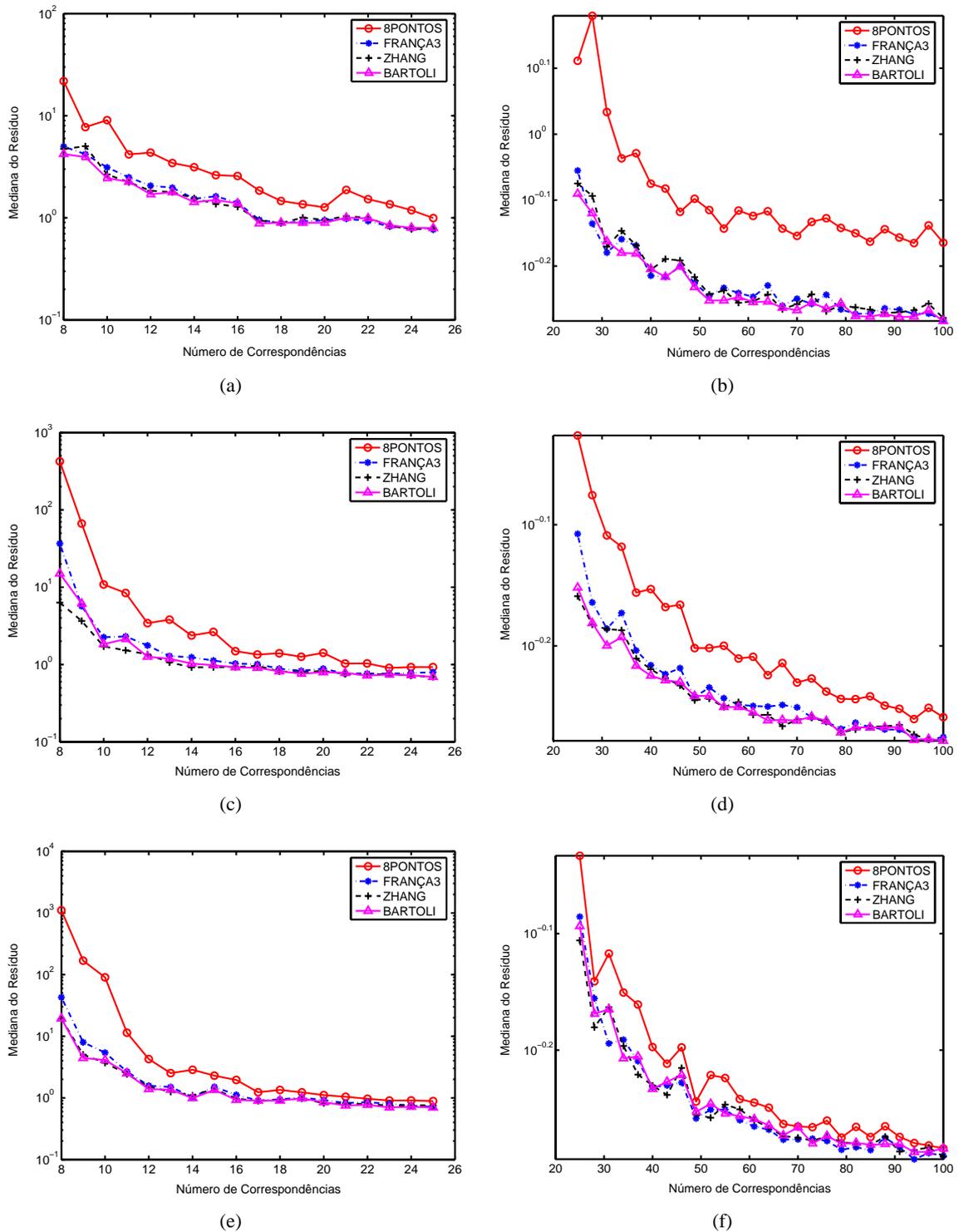


Figura 4.9: Comparação entre os métodos **8PONTOS**, **FRANÇA3**, **BARTOLI** e **ZHANG** para: (a) e (b) o par **GABARITO**; (c) e (d) o par **COMPUTADOR**, e; (e) e (f) o par **QUARTO**.

este merece destaque maior.

Um comentário importante deve ser feito com relação aos resultados do método **ZHANG**. Quando Zhang e Loop (2001) apresentaram esse método, seus resultados experimentais o apontaram como sendo equivalente ao **CSURKA**. Contudo, suas conclusões foram baseadas em apenas um

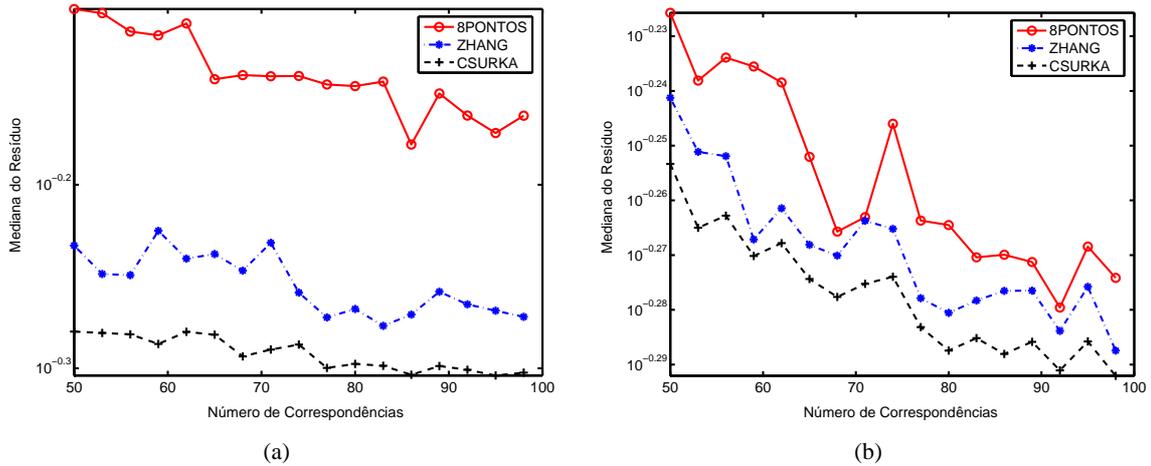


Figura 4.10: Comparação entre os métodos **ZHANG** e **CSURKA** para: (a) o par de imagens **GABARITO**, e; (b) o par **QUARTO**.

par de imagens (com epípolos no infinito) e um único subconjunto de correspondências. Agora, pelo resultado apresentado na figura 4.9, vê-se que suas conclusões foram equivocadas. Tal fato pode ser comprovado ainda pelos gráficos da figura 4.10. Essa figura mostra uma comparação entre os métodos **CSURKA** e **ZHANG**. O gráfico da figura 4.10(a) refere-se ao par de imagens **GABARITO**. Já o da figura 4.10(b), ao par **QUARTO**.

§4.3 Calibração Binocular

Como discutido na seção 3.4, os métodos de calibração propostos atualizam as matrizes de projeção canônicas para as matrizes de projeção euclidianas. Isto é feito estimando-se os parâmetros intrínsecos da câmera 1 e a equação do plano no infinito (o vetor \mathbf{W}) no sistema de coordenadas do conjunto binocular. Evidentemente, a eficiência do método de calibração depende da qualidade da estimação desses parâmetros. Contudo, a exatidão da estimação do vetor \mathbf{W} é difícil de ser avaliada, pois este é um parâmetro projetivo. Por isso, neste trabalho, tal exatidão é examinada de forma indireta, examinando-se os parâmetros intrínsecos da câmera 2, ou seja, a matriz \mathbf{A}' . Isto é válido pois a matriz de projeção da câmera 1 não sofre nenhuma influência de \mathbf{W} .

O algoritmo linear para o caso do gabarito com movimento irrestrito não foi avaliado, pois, como discutido no capítulo anterior, na presença de ruído, este método não nos leva a resultados úteis. Assim, para este tipo de gabarito, apenas o método não-linear foi considerado. Neste caso, para cada câmera, a estimativa inicial (necessária pela rotina de Levenberg-Marquardt) é tomada como $\theta_{sk} = 90^\circ$, $\mathbf{p}_0 = [320, 240]^T$ (o centro das imagens), $n_{x/y} = 1$ e f é suprido pelo método linear que utiliza gabarito com uma extremidade fixa. Contudo, na prática, esse valor poderia ser obtido, por exemplo, das especificações do fabricante das câmeras.

Para avaliar os algoritmos propostos, foram utilizados dados simulados e dados reais. Nas seções a seguir, apresenta-se os resultados de ambos os testes.

Câmera 1					Câmera 2				
f	$n_{x/y}$	$ \pi/2 - \theta_{sk} $	u_0	v_0	f	$n_{x/y}$	$ \pi/2 - \theta_{sk} $	u_0	v_0
150	0,9740	$6,67 \times 10^{-6}$	330	270	200	0,9524	$6,67 \times 10^{-6}$	336	262

Tabela 4.3: Parâmetros intrínsecos do conjunto binocular utilizado para gerar os dados da simulação.

4.3.1 Testes com imagens sintéticas

Os testes com imagens sintéticas foram realizados supondo duas câmaras cujos parâmetros intrínsecos são apresentados na tabela 4.3. Tais câmeras foram simuladas em um conjunto binocular com 15 cm de separação entre as mesmas. Além disso, uma câmera possui, nas três direções, um pequeno ângulo de rotação em relação a outra. Tal ângulo varia entre 5 e 15 graus.

Um bastão de calibração de 80 cm de comprimento e com três pontos foi simulado. Os pontos 3D deste bastão foram projetados no plano de imagem das câmeras sintéticas. Em seguida, tais projeções foram utilizadas para avaliar os algoritmos propostos. Na geração dos pontos 3D, o ângulo do bastão em relação a vertical ficou sempre dentro do intervalo $[\pi/5, \pi/2]$. Além disso, para a simulação de um bastão com uma extremidade fixa, o ponto fixo foi feito igual a $\mathbf{M}_1 = [0, -40, 100]^T$.

§ Sensibilidade a ruído

O primeiro teste efetuado com os dados simulados visa avaliar a influência do ruído sobre os parâmetros estimados. Para isto, as projeções dos pontos 3D foram corrompidas, independentemente, com ruído gaussiano de média zero e variância σ . A cada ensaio, σ variou entre 0,1 e 1 pixel. Além disso, para cada valor de σ , os parâmetros das câmeras são obtidos da mediana de 50 simulações diferentes. Os resultados desses testes são apresentados na figura 4.11 (para o bastão com uma extremidade fixa) e na figura 4.12 (bastão com movimento irrestrito). Analisando os gráficos, observa-se que, exceto pelo parâmetro s_k , a exatidão dos parâmetros varia quase que linearmente com o ruído. Além disso, os parâmetros referentes a câmera 2 são estimados com uma exatidão um pouco menor. Este fato está de acordo com o trabalho de Horaud et al. (2000) que demonstrou que, no caso da calibração binocular por etapas, a parcela dominante do erro é devida a equação do plano no infinito.

Os métodos não-lineares propostos produzem uma melhora significativa na estimação inicial dos parâmetros. Como mostram as figuras 4.11(c) e 4.11(d), o método não-linear para o bastão com uma extremidade fixa reduz os erros quase que pela metade. Enquanto que o método não-linear com movimento irrestrito reduz o erro por um fator menor (figura 4.12).

Infelizmente, pelos métodos propostos, a exatidão da estimação do parâmetro s_k da câmera 2 é muito influenciada pelo nível de ruído presente nos dados. Contudo, como discutido na seção 2.1, na prática, este parâmetro é muito próximo de zero, pois $\theta_{sk} \simeq \pi/2$. De fato, como argumentado por Luong e Faugeras (1997), tal parâmetro é considerado diferente de zero apenas quando é necessário inserir mais restrições ao problema de calibração.

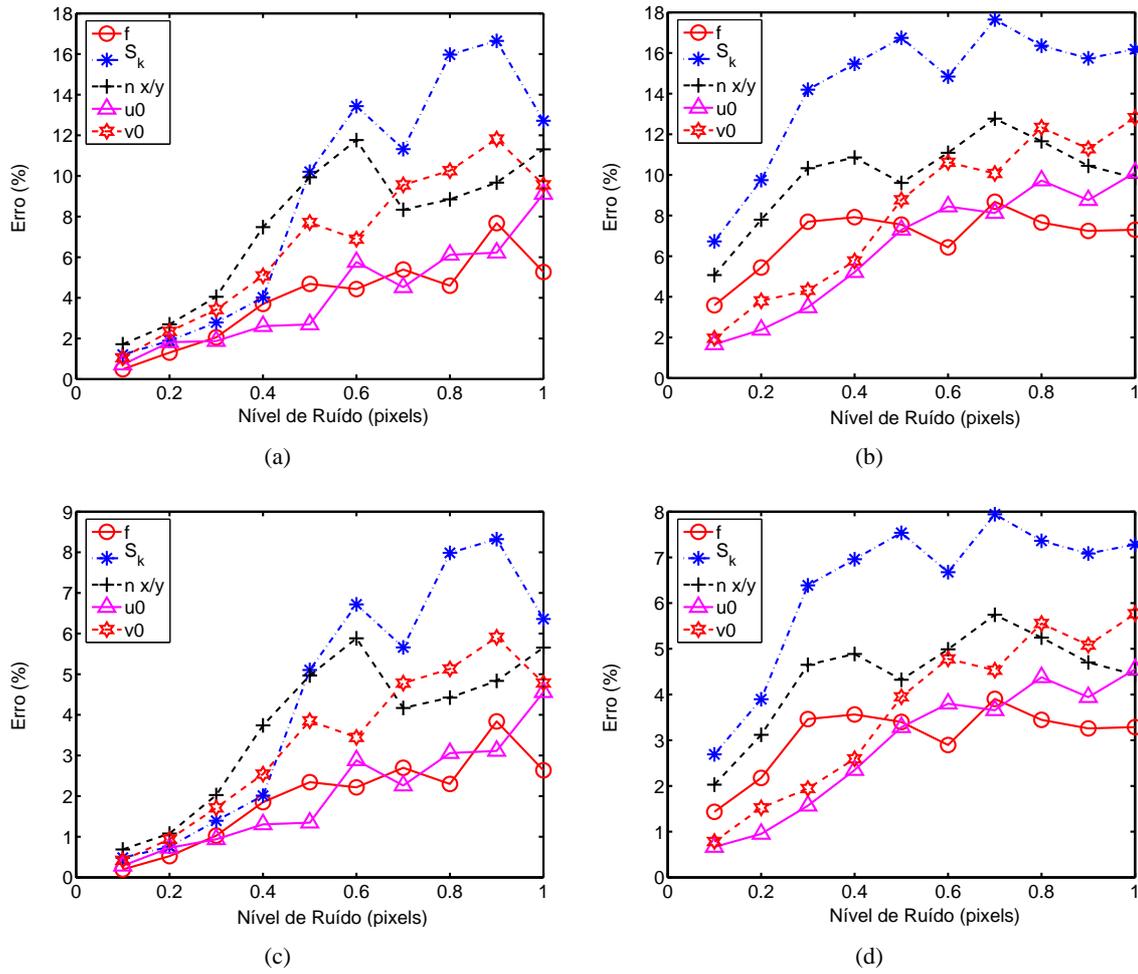


Figura 4.11: Influência do ruído na eficiência do algoritmo de calibração com um bastão com uma extremidade fixa. (a) mostra o resultado do método linear para a câmera 1 e (b) para a câmera 2. Os resultados do algoritmo não-linear são apresentados em (a) para a câmera 1 e em (b) para a câmera 2.

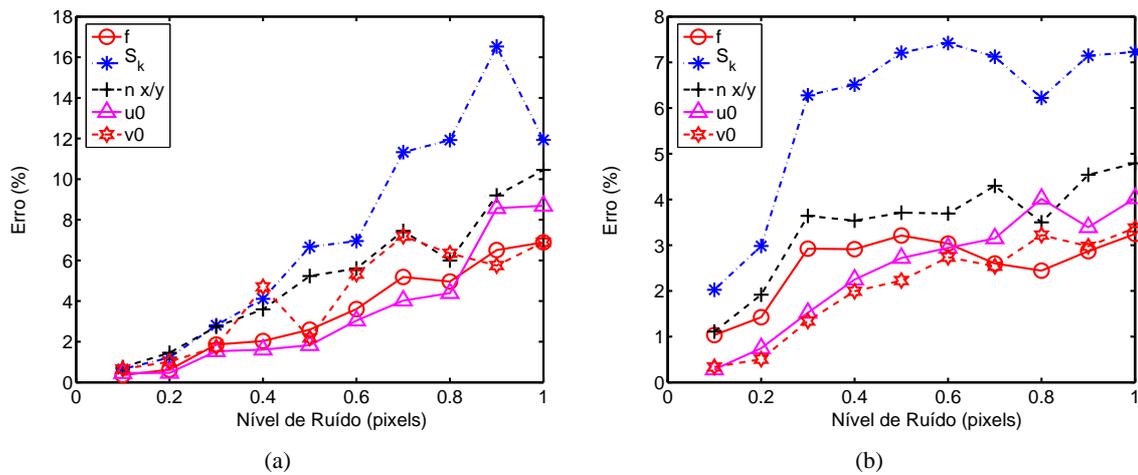


Figura 4.12: Influência do ruído na eficiência do algoritmo de calibração com um bastão com movimento irrestrito. (a) mostra o resultado do método não-linear para a câmera 1 e (b) para a câmera 2.

§ Influência do número de pontos 3D

Outra informação importante é saber como a quantidade de correspondências influencia nos resultados. Há duas formas de obter-se mais correspondências: aumentando o número de pontos do bastão ou o número de imagens capturadas. Um aumento excessivo do número de pontos do bastão reduz a distância entre pontos vizinhos. Isto pode reduzir a relação sinal ruído e comprometer o resultado. Por isso, a melhor opção é aumentar o número de imagens que fazem parte da seqüência de calibração.

Para aferir a influência do número de imagens na exatidão dos parâmetros estimados, como na seção anterior, um bastão com três pontos foi simulado. O número de pares de imagens utilizados na simulação variou entre seis e duzentos. Novamente, foram realizadas 50 simulações diferentes para uma mesma quantidade de imagens. As informações coletadas neste experimento são apresentadas na figura 4.13 (para o bastão com uma extremidade fixa) e na figura 4.14 (bastão com movimento irrestrito). Com todos os métodos, observa-se que, para uma quantidade de pares de imagens entre seis e quarenta, o erro diminui sensivelmente com o aumento do número de imagens. Contudo, para 40 ou mais pares, o ganho na exatidão é muito modesto. De fato, dependendo da aplicação, a exatidão obtida com um número grande de imagens pode não compensar o aumento do custo computacional do algoritmo.

4.3.2 Testes em imagens reais

O conjunto binocular utilizado nos testes (figura 4.15) é formado por duas câmeras USB, modelo Aiptek DV3100. Tais câmeras possuem um sensor CCD que permite capturar imagens nítidas, com dimensões de 640×480 pixels. Inicialmente, os parâmetros do par estéreo foram conhecidos através de um experimento de extrema exatidão. Isto é importante para avaliar a eficiência dos métodos propostos. Além disso, os resultados obtidos são comparados a de outro método não-linear, o método de Zhang (2000).

Nas seções a seguir, os experimentos são descritos em detalhes e os resultados são apresentados.

§ Parâmetros absolutos

Para avaliar os métodos de calibração propostos, é necessário conhecer os parâmetros do conjunto binocular utilizado nos testes. Para isso, um gabarito 3D foi confeccionado e, com auxílio de um micrômetro de rosca com uma haste de 75 mm, suas dimensões foram medidas com uma incerteza de 0,01 mm. Além disso, como mostra a figura 4.16, as faces do gabarito são de cor branca, sendo que duas delas possuem 100 padrões circulares da cor preta. Do trabalho de Heikkilä (2000), o centróide de tais padrões pode ser conhecido com uma exatidão de 0,05 pixels. Uma vez que todos os dados foram extraídos, a calibração foi efetuada utilizando o algoritmo 2.6, na página 41.

Devido a exatidão dos dados 3D e 2D utilizados, os parâmetros do conjunto binocular obtidos neste experimento e disponíveis na tabela 4.4 são considerados “parâmetros reais ou absolutos”. Por isso, estes foram utilizados para avaliar a precisão dos métodos testados.

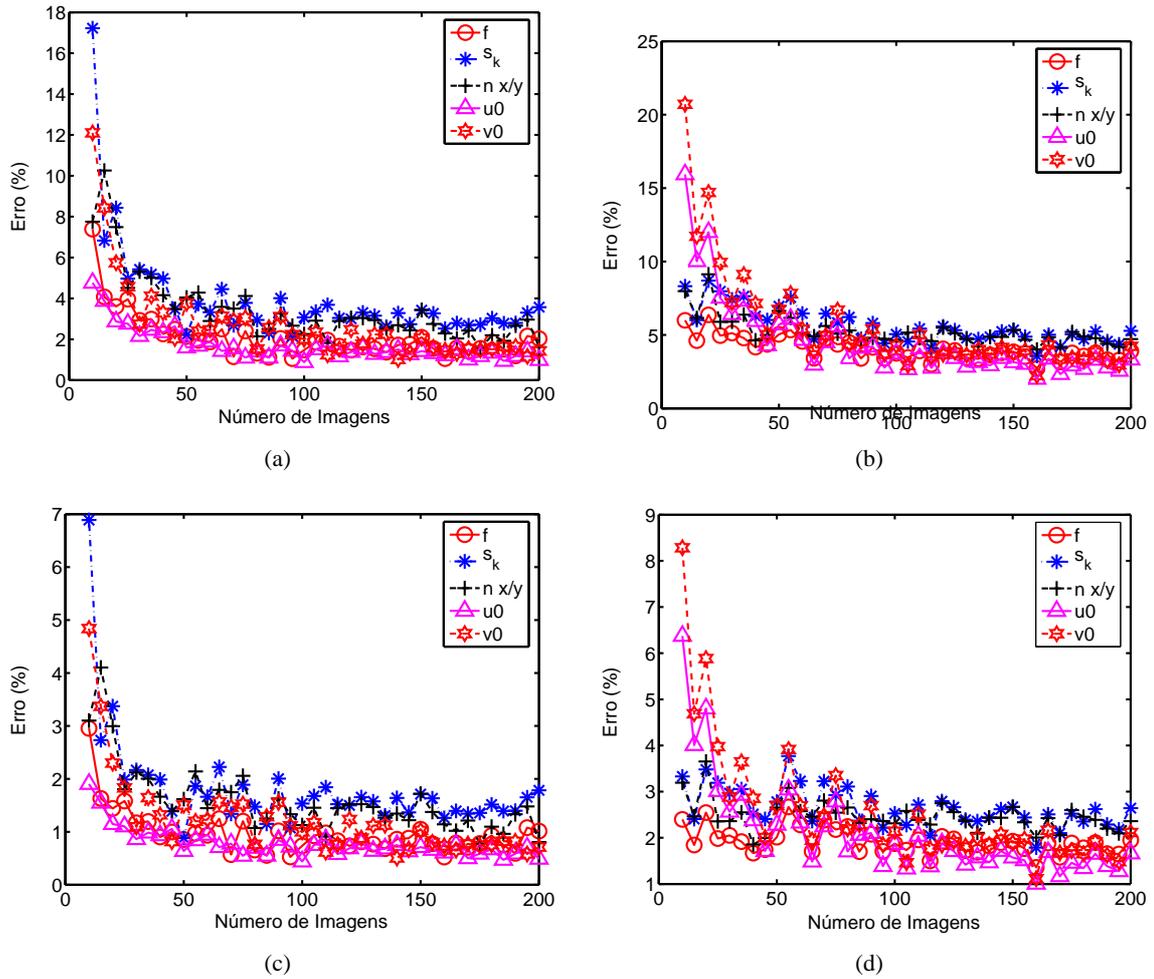


Figura 4.13: Influência do número de pares de imagens na eficiência do algoritmo de calibração com um bastão com uma extremidade fixa. (a) mostra o resultado do método linear para a câmera 1 e (b) para a câmera 2. Os resultados do algoritmo não-linear são apresentados em (a) para a câmera 1 e em (b) para a câmera 2.

§ Calibração com um plano

A exatidão dos métodos de calibração propostos foram comparadas a do método de Zhang (2000). O gabarito utilizado no método é um plano com um padrão quadriculado, logo sua confecção é mais simples que a do gabarito utilizado no método da seção anterior.

Na calibração, o gabarito deve fazer movimentos genéricos. Então, além dos parâmetros intrínsecos das câmeras, a cada deslocamento, as coordenadas 3D dos pontos do gabarito, a matriz rotação, \mathbf{R}_i , e o vetor translação, \mathbf{t}_i , do plano em relação a câmera também devem ser estimados. Para tal estimação, dado n deslocamentos de um gabarito com m pontos, Zhang (2000) propôs a seguinte função a ser minimizada

$$\sum_{i=1}^n \sum_{j=1}^m \|\mathbf{m}_{ji} - \hat{\mathbf{m}}(\mathbf{A}, \mathbf{R}_i, \mathbf{t}_i, \mathbf{M}_{ji})\|^2, \quad (4.2)$$

onde \mathbf{m}_{ji} é a projeção do ponto \mathbf{M}_{ji} , durante o i -ésimo deslocamento, e $\hat{\mathbf{m}}$ é o ponto estimado.

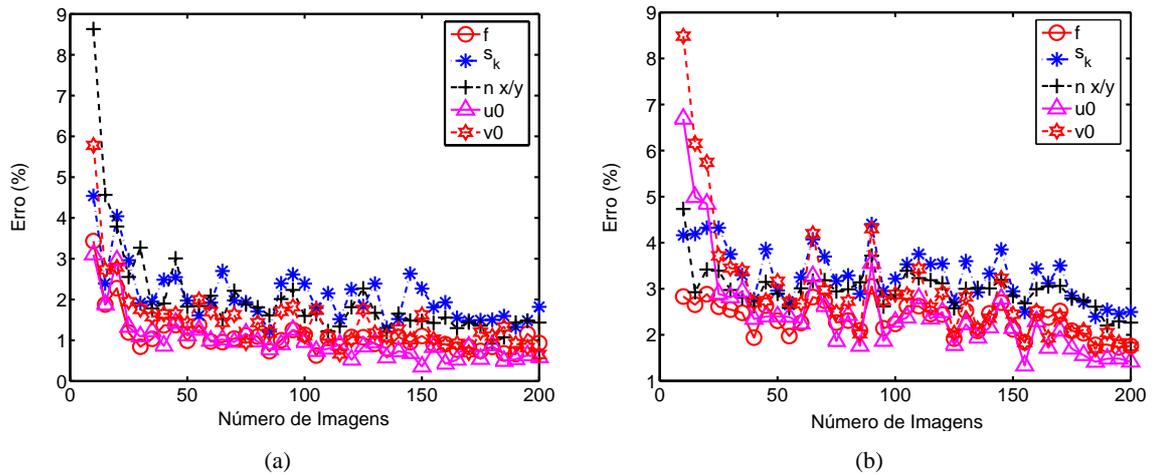


Figura 4.14: Influência do número de pares de imagens na eficiência do algoritmo de calibração com um bastão com movimento irrestrito. (a) mostra o resultado do método não-linear para a câmera 1 e (b) para a câmera 2.



Figura 4.15: Conjunto binocular utilizado nos experimentos.

Para n deslocamentos e um gabarito com m pontos, o número de incógnitas que devem ser estimadas é igual a $5 + 5n + 3mn$. Além disso, apesar do problema também poder ser particionado e adaptado ao algoritmo 3.5, a sua complexidade ainda é bem mais elevada que as dos algoritmos propostos neste trabalho. Isto é devido ao número maior de incógnitas e, principalmente, pelo fato da equação 4.2 estimar os parâmetros intrínsecos de apenas uma câmera do par estéreo, ou seja, são necessárias duas minimizações para determinar todos os parâmetros do par estéreo.

Neste trabalho, utilizou-se um gabarito planar com 121 pontos. Durante a calibração, foram realizados 10 deslocamentos e uma seqüência com 10 imagens estéreo, como as da figura 4.17, foram capturadas. Em seguida, a implementação de Bouguet (2004) do algoritmo de Zhang (2000), disponível livremente na internet, foi utilizada para estimar os parâmetros do conjunto binocular. Os resultados são apresentados na tabela 4.4.

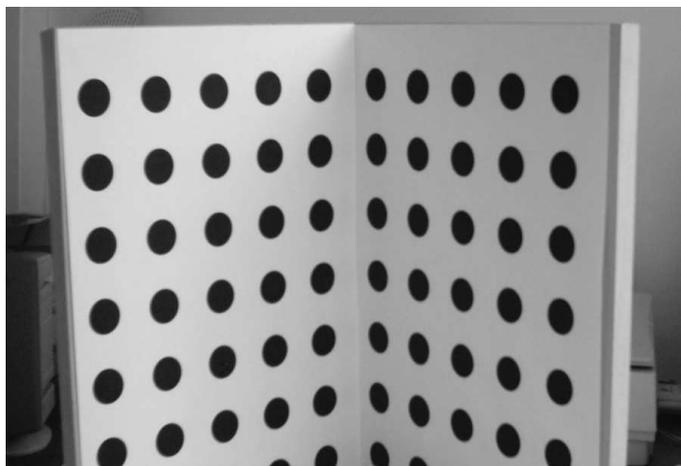


Figura 4.16: Gabarito de calibração utilizados nos testes.

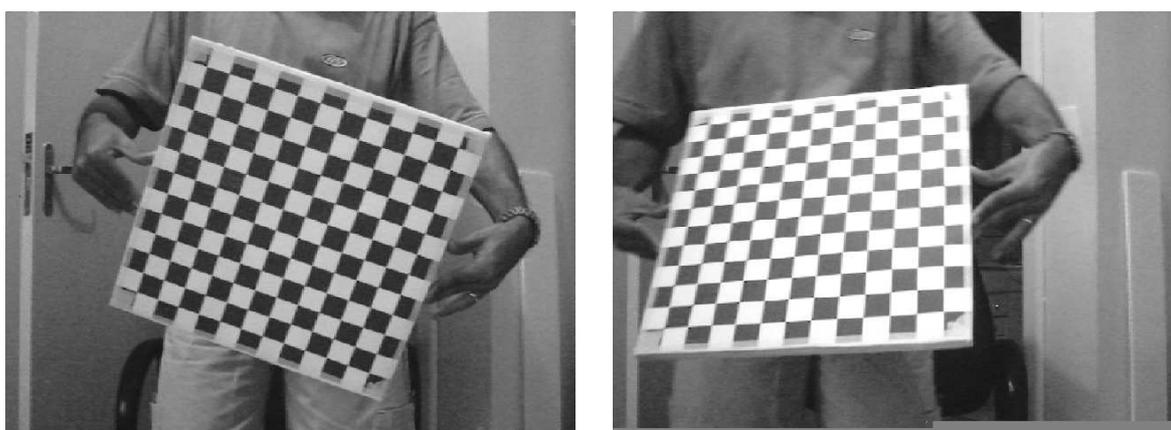


Figura 4.17: Exemplos da sequência de imagens utilizada na calibração do conjunto binocular com um gabarito linear.



Figura 4.18: Exemplos da sequência de imagens utilizada na calibração do conjunto binocular com um gabarito planar.

§ Calibração com uma haste

Para os testes com uma haste graduada, o gabarito da figura 4.18 foi confeccionado. Este é composto por um fio de cobre com 6 esferas brancas. Cada esfera tem cerca de 1,5 cm de diâmetro. Após o gabarito ser confeccionado, a distância entre uma esfera e outra foi medida com auxílio de uma régua comum. Quando necessário, uma das extremidades do fio de cobre foi fixa em um suporte, enquanto que a outra deslocava-se livremente. Neste caso, a projeção do ponto fixo foi estimada como discutido na seção 3.4.4.

Para capturar as imagens utilizadas nos testes, o gabarito foi movido a frente das câmeras e uma sequência com 60 imagens estéreo foi capturada por cada câmera do conjunto binocular. Devido as imagens serem consecutivas, desde que a localização das esferas seja marcada manualmente no primeiro par de imagens 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 (LINDBERG, 1994, Cap. 7).

Para fácil visualização, os resultados de todos os métodos avaliados são apresentados na tabela 4.4. Observe que apenas os parâmetros obtidos com o gabarito 3D (referenciado na tabela como **REAIS**) são dados de forma absoluta. Todos os outros métodos são valores dados em percentagem (referente ao erro relativo aos parâmetros **REAIS**).

Analisando os dados, nota-se que, como revelado pela simulação, o parâmetro θ_{sk} é estimado com um erro elevado (mais de 5 %). Contudo, se θ_{sk} for considerado igual a $\pi/2$, ou seja, este parâmetro não for incluído na estimação, o erro cometido é de apenas 0,01 %.

Quando comparado ao método com gabarito irrestrito (referenciado na tabela 4.4 por **LIVRE**), o método de calibração que utiliza um gabarito com uma extremidade fixa (referenciado por **FIXO**) produz uma melhor estimação dos parâmetros das câmeras. Este fato, associado com a menor complexidade, torna este o método mais indicado. Além disso, o fato de ter uma extremidade fixa

¹Do inglês: *tracking*.

Método	Câmera 1					Câmera 2				
	f	$n_{x/y}$	θ_{sk}	u_0	v_0	f	$n_{x/y}$	θ_{sk}	u_0	v_0
REAIS	205,97	1,02	90,01	330,56	270,67	201,89	1,02	90,01	336,67	262,23
ZHANG	1,56	1,69	0,50	3,76	2,89	1,86	1,58	0,53	3,51	3,01
LIVRE	3,08	2,98	6,04	4,01	3,56	2,49	2,56	5,76	4,01	4,38
FIXO	2,19	2,30	5,41	3,53	3,60	2,33	2,27	5,78	2,76	3,69

Tabela 4.4: Resultados experimentais da calibração do conjunto binocular (θ_{sk} é dado em graus). Observe que apenas os parâmetros **REAIS** são valores absolutos. Os resultados de todos os outros métodos são dados na forma do erro relativo (em percentagem) aos parâmetros **REAIS**.

não é, necessariamente, uma desvantagem, pois, mesmo quando o gabarito efetua um movimento irrestrito, este precisa ter algum tipo de suporte para efetuar o movimento. Assim, para que uma das extreminades seja fixa, basta limitar o movimento que o gabarito faz em seu suporte.

De uma forma geral, a exatidão obtida pelo método de Zhang (2000) (referenciado na tabela por **ZHANG**) foi um pouco melhor que a dos métodos propostos. Contudo, este fato foi devido principalmente à confecção do gabarito de calibração. No método de Zhang (2000), o gabarito foi confeccionado com ajuda do computador, logo as dimensões do gabarito são conhecidas com exatidão. No caso dos métodos propostos, o gabarito foi confeccionado à mão, com auxílio de uma régua ordinária. Levando este fato em consideração, os resultados da tabela 4.4 demonstram que a técnica é bastante promissora.

COMENTÁRIOS, CONCLUSÕES E SUGESTÕES DE TRABALHOS

Neste trabalho, apresentaram-se algumas contribuições à calibração de um conjunto binocular. Estas são relativas à correspondência de pontos em imagens estéreo, cálculo da matriz fundamental e calibração com um gabarito de uma única dimensão.

Talvez, devido à popularidade das técnicas de calibração clássica, maior ênfase é dada ao processo de correspondência de pontos em sistemas já calibrados. Para o caso não-calibrado, os algoritmos disponíveis são, na maioria das vezes, pouco eficientes, pois utilizam técnicas complexas e de custo computacional elevado ou, ao contrário, usam uma medida de semelhança simples, que forma um conjunto de correspondência inicial pouco confiável, e, em seguida, aplicam a restrição epipolar, através de algoritmos como o RANSAC ou LMedS, para identificar os candidatos confiáveis. Neste último caso, devido ao grande nível de ambigüidade, o algoritmo de correspondência requer muitas iterações. Em oposição, o algoritmo de correspondência proposto utiliza técnicas simples, mas muito eficientes, pois consegue estabelecer um conjunto de correspondência inicial reduzido, mas confiável. Tal confiança é aumentada ao longo de três fases distintas. O resultado é um algoritmo simples, eficaz e rápido.

A análise dos resultados da seção 4.1 mostra que a transformada censo, em conjunto com a distância Hamming, constitui uma excelente medida de semelhança entre pontos. Além disso, a equação (3.2) possibilita uma prática ferramenta para redução da ambigüidade no processo de correspondência. Contudo, o método ainda pode ser melhorado. Possivelmente, uma alteração que aumentaria significativamente a confiança das correspondências, é estabelecer um limiar máximo para a distância Hamming, ou seja, uma correspondência fora desse limiar seria desconsiderada. Entretanto, a escolha de tal limiar deve ser feita dinamicamente, de acordo com o conjunto total de possíveis correspondências. Um estudo deste tipo foi realizado recentemente por Kanatani e Kanazawa (2004), para o caso da equação (2.9). Um estudo semelhante para o caso da transformada censo reduziria o número de iterações necessárias à terceira fase do algoritmo proposto. Contudo, evidentemente, aumentaria o seu custo computacional.

O processo de estimação da matriz fundamental ainda carece de uma técnica que obtenha resultados exatos a um custo computacional aceitável. Os algoritmos propostos neste trabalho apontam nessa direção. O primeiro constitui um método linear. Contudo, apenas fornece um resultado cuja qualidade está entre a obtida pelo método de Hartley (1997) e a devida à maioria dos algoritmos não-lineares. No entanto, como tem uma solução linear, pode ser utilizado como uma alternativa aos métodos não-lineares mais exatos em sistemas que, por exemplo, tenham baixa capacidade de processamento. Por outro lado, o algoritmo não-linear proposto, baseado no paralaxe virtual, tem exatidão semelhante à maioria dos métodos não-lineares. Contudo, este usa uma parametrização com apenas cinco incógnitas e um único espaço de busca. Por isso, o seu custo computacional é menor.

Os métodos de calibração propostos foram desenvolvidos com o objetivo de flexibilizar o processo de calibração. Seu único pré-requisito é que o conjunto binocular “observe” um gabarito de uma única dimensão como, por exemplo, um bastão graduado ou um cordão de esferas, enquanto este realiza um movimento. Tal movimento não precisa ser conhecido, mas não deve ser uma translação pura ou, evidentemente, apenas rotacional.

Quando comparados a técnicas de calibração automática ou mesmo às que utilizam gabaritos mais complexos, os métodos propostos ganham em eficiência, pois possuem custo computacional menor que as técnicas automáticas e são bem mais flexíveis que os métodos clássicos. Além disso, como mostram os experimentos realizados em imagens reais e sintéticas, a exatidão obtida com as técnicas propostas é da mesma ordem de grandeza das obtidas com os métodos tradicionais.

A principal vantagem dos métodos propostos é utilizar um gabarito de simples confecção. Isto torna a calibração fácil de ser executada. Assim, no futuro, o método poderá ser utilizado, inclusive, por pessoas sem conhecimento técnico, em aplicações domésticas de entretenimento. Contudo, para que isto seja possível, o modelo da câmera deve ser alterado para levar em consideração a distorção radial (presentes em câmeras de baixo custo) e deve ser pesquisada alguma forma de detectar seqüências de movimentos que possam causar singularidades na minimização.

Na calibração com gabarito de movimento irrestrito, uma outra contribuição importante seria desenvolver um algoritmo linear robusto para inicializar a minimização não-linear. Isto tornaria o método mais geral, pois eliminaria a necessidade do conhecimento prévio de uma estimação da distância focal da câmera.

GEOMETRIA PROJETIVA

Uma das mais poderosas ferramentas matemáticas utilizadas na visão estéreo é a geometria projetiva. Tal geometria possui conceitos que fogem um pouco do senso comum. Por isso, alguns resultados, importantes que são utilizados para derivar grande parte da teoria dos capítulos anteriores, são descritos a seguir.

§A.1 Razão Cruzada

Ao contrário da geometria euclidiana, na geometria projetiva a razão de comprimentos sobre uma reta varia sobre transformações. Contudo, a razão de razões (ou razão cruzada) é invariante a transformações projetivas. Assim, se um ponto \mathbf{m}_i em uma reta for representado em coordenadas homogêneas por $\tilde{\mathbf{m}}_i = [u_i, t_i]^T$, dado quatro pontos $\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3$ e \mathbf{m}_4 , pertencentes a mesma reta, define-se a razão cruzada entre eles como sendo

$$Cross(\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3, \mathbf{m}_4) = \frac{|\tilde{\mathbf{m}}_1 \tilde{\mathbf{m}}_2| |\tilde{\mathbf{m}}_3 \tilde{\mathbf{m}}_4|}{|\tilde{\mathbf{m}}_1 \tilde{\mathbf{m}}_3| |\tilde{\mathbf{m}}_2 \tilde{\mathbf{m}}_4|},$$

onde

$$|\tilde{\mathbf{m}}_i \tilde{\mathbf{m}}_j| = \det \begin{bmatrix} u_i & u_j \\ v_i & v_j \end{bmatrix}.$$

Assim, dado o valor de $Cross(\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3, \mathbf{m}_4)$ e as coordenadas de três pontos dentre $\{\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3, \mathbf{m}_4\}$, é possível estimar as coordenadas do ponto restante.

§A.2 Intersecção de uma reta com a reta no infinito

Proposição A.1 Considere $l = [a, b, c]^T$ uma reta no espaço \mathcal{P}^2 . Esta reta intercepta a reta no infinito, $l_\infty = [0, 0, 1]^T$, no ponto $\tilde{\mathbf{m}}_\infty = [b, -a, 0]^T$.

Prova: Veja (HARTLEY; ZISSERMAN, 2000, Sec. 1.2) para prova. \square

§A.3 Projeção ortogonal de um ponto em uma reta

Freqüentemente, é necessário calcular a projeção ortogonal de um ponto em uma reta. Para isto, pode-se utilizar a proposição a seguir.

Proposição A.2 Considere um ponto $\tilde{\mathbf{m}} = [u, v, 1]^T$ e uma reta $\mathbf{l} = [a, b, c]^T$, com $b \neq 0$, no espaço \mathcal{P}^2 . A projeção ortogonal de \mathbf{m} em \mathbf{l} é dada por

$$\tilde{\mathbf{m}}' = \left[\lambda, -\frac{c + a\lambda}{b}, 1 \right]^T,$$

onde $\lambda = \frac{b^2u - abv - ac}{a^2 + b^2}$.

Prova: É fácil verificar que um ponto sobre \mathbf{l} tem a forma $\tilde{\mathbf{m}}' = [\lambda, -\frac{c+a\lambda}{b}, 1]^T$. Além disso,

$$d^2(\mathbf{m}, \mathbf{m}') = (u - \lambda)^2 + \left(v + \frac{c + a\lambda}{b} \right)^2.$$

Agora, \mathbf{m}' é a projeção ortogonal de \mathbf{m} em \mathbf{l} , se e somente se, $d^2(\mathbf{m}, \mathbf{m}')$ for igual a distância mínima entre \mathbf{m} e \mathbf{l} . Como, $d^2(\mathbf{m}, \mathbf{m}')$ é mínima quando $\lambda = \frac{b^2u - abv - ac}{a^2 + b^2}$, a prova está concluída. \square

A proposição A.2 exige que \mathbf{m} não esteja no infinito e $\mathbf{l} = [a, b \neq 0, c]^T$. Quando essas condições não forem satisfeitas, deve-se utilizar transformações que levem \mathbf{m} e \mathbf{l} a um referencial no qual tais condições sejam válidas.

§A.4 Homografia induzida por um plano

Considere 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 ”.

Um plano muito citado na literatura de visão estéreo é o plano no infinito, Π_∞ . Este plano é formado por todos os pontos que estão no infinito e é definido por $\Pi_\infty = [0, 0, 0, 1]^T$. A homografia induzida por este plano é chamada de “homografia do infinito”, \mathbf{H}_∞ . Ela mapeia projeções de pontos em Π_∞ entre as duas imagens (figura A.1). Além disso,

$$\mathbf{H}_\infty = \mathbf{A}'\mathbf{R}\mathbf{A}^{-1}. \quad (\text{A.1})$$

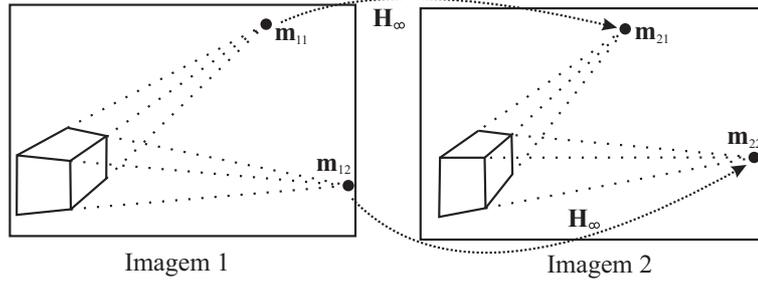


Figura A.1: Podemos utilizar o fato que retas paralelas encontram-se no infinito para obtermos projeções de pontos que estão em Π_∞ . As projeções de pontos neste plano estão relacionadas por \mathbf{H}_∞ .

Se $\mathbf{A} = \mathbf{A}'$, ou seja, as imagens forem capturadas por uma mesma câmera, mas de posições diferentes, a equação anterior conduz a um resultado importante que é resumido na proposição a seguir.

Proposição A.3 Se \mathbf{H}_∞ é conhecida, a matriz \mathbf{A} pode ser obtida resolvendo-se

$$\mathbf{h}_i \mathbf{k} = \mathbf{0}, \quad (\text{A.2})$$

onde $\mathbf{k} = [k_1, k_2, k_3, k_4, k_5, k_6]^T$ é composto pelos elementos da matriz do cone dual e \mathbf{h}_i é uma matriz 9×6 com os coeficientes de nove equações lineares obtidas de (A.1).

Prova: Da equação (A.1) é imediato que $\mathbf{R} = \mathbf{A}^{-1} \mathbf{H}_\infty \mathbf{A}$. Assim, da proposição 2.2 (pág. 10), temos que $\mathbf{A}^{-1} \mathbf{H}_\infty \mathbf{A} = \mathbf{A}^T \mathbf{H}_\infty^{-T} \mathbf{A}^{-T}$, ou seja,

$$\mathbf{H}_\infty \mathbf{K} = \mathbf{K} \mathbf{H}_\infty^{-T}, \quad (\text{A.3})$$

onde $\mathbf{K} = \mathbf{A} \mathbf{A}^T$ é a matriz do dual cone, ω^* .

A equação (A.3) nos fornece nove equações em função dos seis elementos de \mathbf{K} . Com isso, (A.3) pode ser re-escrita como (A.2). Contudo, na ausência de ruído, o posto de \mathbf{h}_i é quatro. Assim, (A.2) deve ser resolvido utilizando o procedimento da seção B.2.2 (pág. 122) e a proposição 2.6 (pág. 46) para obtermos uma restrição extra.

Uma vez que \mathbf{K} for conhecida, a proposição B.1 (pág. 120) pode ser utilizada para obter-se \mathbf{A} . □

Se a câmera realiza dois ou mais deslocamentos, a proposição a seguir pode ser utilizada para encontrar a matriz \mathbf{A} .

Proposição A.4 Se uma câmera realiza n deslocamentos de uma posição, i , para outra posição, $(i + 1)$, para cada deslocamento, é possível estimar uma homografia do infinito, $\mathbf{H}_{\infty i(i+i)}$. Além

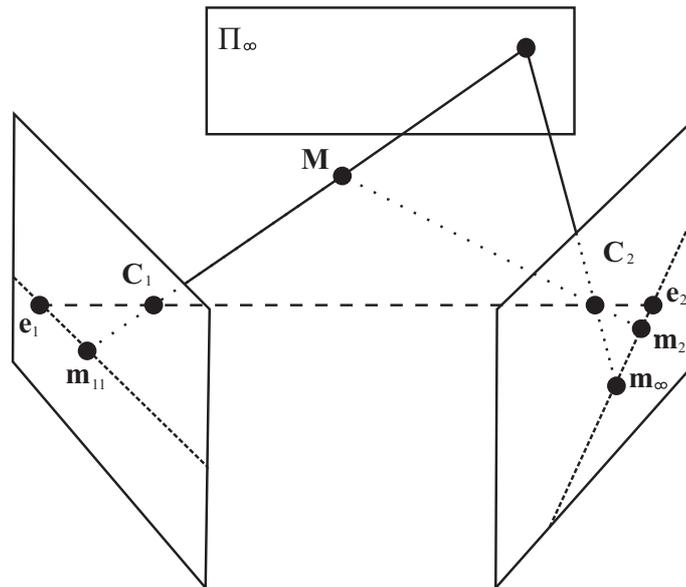


Figura A.2: O correspondente do ponto \mathbf{m}_{11} está sobre o segmento de reta $\overline{\mathbf{m}_{\infty}\mathbf{e}_2}$.

disso, a matriz \mathbf{A} da câmera pode ser estimada resolvendo

$$\begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \\ \vdots \\ \mathbf{h}_n \end{bmatrix} \mathbf{k} = \mathbf{0}.$$

Um observação importante é que, se \mathbf{H}_{∞} é conhecida, ela facilita o processo de correspondência de pontos. Pois, dado um ponto \mathbf{m}_{1i} em I_1 , o correspondente deste ponto em I_2 está sobre o segmento de reta $\overline{\mathbf{m}_{\infty}\mathbf{e}_2}$, onde $\mathbf{m}_{\infty} = \mathbf{H}_{\infty}\tilde{\mathbf{m}}_{1i}$ (figura A.2).

FERRAMENTAS MATEMÁTICAS

Além da fatoração de Cholesky, algoritmos de minimização baseados nos quadrados-mínimos são ferramentas muito utilizadas ao longo de todo o texto. Neste apêndice, tais ferramentas são descritas em detalhes e algumas proposições são apresentadas.

§B.1 Fatoração de Cholesky

Para determinar os parâmetros intrínsecos da câmera a partir da matriz, \mathbf{K} , do cone dual, podemos utilizar a fatoração de Cholesky, resumida na proposição a seguir.

Proposição B.1 *Se $\mathbf{K} \in \mathbb{R}^{n \times n}$ é simétrica e positivamente definida, então existe uma única matriz triangular superior, $\mathbf{A} \in \mathbb{R}^{n \times n}$, cujos elementos da diagonal principal são todos positivos e $\mathbf{K} = \mathbf{A}\mathbf{A}^T$.*

Prova: Veja (GOLUB; Van Loan, 1996, pág. 143) para prova. □

Com o suporte dado pela proposição anterior, os parâmetros intrínsecos da câmera podem ser determinados facilmente como segue.

Considere, \mathbf{A} dado por (2.4), temos

$$\mathbf{K} = \begin{bmatrix} k_1 & k_2 & k_3 \\ k_2 & k_4 & k_5 \\ k_3 & k_5 & 1 \end{bmatrix} = \begin{bmatrix} f^2 + (f \cot \theta_{sk})^2 + u_0 & f^2 n_{x/y} \cot \theta_{sk} + u_0 v_0 & u_0 \\ f^2 n_{x/y} \cot \theta_{sk} + u_0 v_0 & (f \cot \theta_{sk})^2 + v_0^2 & v_0 \\ u_0 & v_0 & 1 \end{bmatrix}$$

Logo, a fatoração de Cholesky da matriz acima pode ser efetuada se fizermos

$$\begin{aligned}
 u_0 &= k_3, \\
 v_0 &= k_5, \\
 f &= \sqrt{k_1 - u_0^2 - (k_2 - u_0 v_0)^2 / (k_4 - v_0^2)}, \\
 \theta_{sk} &= \cot^{-1} \left(\frac{\sqrt{k_1 - u_0 - f^2}}{f} \right) \quad \text{ou} \quad \theta_{sk} = \cot^{-1} \left(\frac{\sqrt{k_4 - v_0^2}}{f} \right) \\
 n_{x/y} &= (k_2 - u_0 v_0) / (f^2 \cot \theta_{sk}),
 \end{aligned} \tag{B.1}$$

§B.2 Solução dos Quadrados-Mínimos para $\mathbf{Ax} = 0$

Considerando $\mathbf{A} \in \mathbb{R}^{m \times n}$ e $\mathbf{x} \in \mathbb{R}^{m \times 1}$, onde $m \geq n$, devemos encontrar \mathbf{x} que satisfaça

$$\mathbf{Ax} = \mathbf{0}.$$

Na prática, na presença de ruído, devemos resolver o sistema de equações anterior como sendo

$$\min_{\mathbf{x}} \|\mathbf{Ax}\|^2 = 0, \quad \text{sujeito a } \|\mathbf{x}\|^2 = 1. \tag{B.2}$$

Para resolver o problema anterior, podemos encontrar duas situações: (a) o posto de \mathbf{A} é igual a $(n - 1)$ ou (b) o posto é igual a $(n - p)$, onde $p > 1$. Cada uma dessas situações é analisada a seguir.

B.2.1 Solução para posto igual a $(n - 1)$

O problema (B.2) pode ser reformulado, usando o multiplicador de Lagrange λ , como um problema irrestrito, ou seja,

$$\min_{\mathbf{x}} C(x, \lambda), \quad \text{onde } C(x, \lambda) = \|\mathbf{Ax}\|^2 + \lambda (1 - \|\mathbf{x}\|^2).$$

Como

$$\frac{\partial C}{\partial \mathbf{x}} = 2\mathbf{A}^T \mathbf{Ax} - 2\lambda \mathbf{x},$$

a condição $\partial C / \partial \mathbf{x} = 0$ nos leva a $\mathbf{A}^T \mathbf{Ax} = \lambda \mathbf{x}$, ou seja, \mathbf{x} precisa ser um autovetor de $\mathbf{A}^T \mathbf{A}$, associado ao autovalor λ . Como desejamos o mínimo da função, este resultado nos leva a seguinte proposição.

Proposição B.2 Dado um sistema $\mathbf{Ax} = \mathbf{0}$, onde $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{x} \in \mathbb{R}^{m \times 1}$ e $m \geq n$, se o posto de \mathbf{A} é igual a $(n - 1)$, a solução é o autovetor de $\mathbf{A}^T \mathbf{A}$ que corresponde ao menor autovalor.

B.2.2 Solução para posto igual a $(n - p)$

Como no caso anterior, a solução está nos autovetores de $\mathbf{A}^T \mathbf{A}$. Contudo, como o posto de \mathbf{A} é $(n - p)$, existe p autovetores, $\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_p$, que correspondem ao autovalor nulo. Assim, o autoespaço associado tem dimensão p . Além disso, os autovetores que correspondem ao autovalor nulo formam uma possível base desse autoespaço, ou seja, todas as possíveis soluções para o problema são representadas por

$$\mathbf{V}(\alpha_1, \alpha_2, \dots, \alpha_p) = \alpha_1 \mathbf{V}_1 + \alpha_2 \mathbf{V}_2 + \dots + \alpha_p \mathbf{V}_p.$$

Em particular, se $p = 2$, a equação anterior pode ser simplificada para

$$\mathbf{V}(\alpha) = \mathbf{V}_1 + (1 - \alpha) \mathbf{V}_2.$$

Neste caso, uma outra restrição deve ser usada para encontrar α e obtermos uma solução única para o problema (B.2).

§B.3 Solução dos Quadrados-Mínimos para $\mathbf{Ax} = \mathbf{b}$

Considerando $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{x} \in \mathbb{R}^{m \times 1}$ e $\mathbf{b} \in \mathbb{R}^{m \times 1}$, onde $m \geq n$, devemos encontrar \mathbf{x} que satisfaça

$$\mathbf{Ax} = \mathbf{b}.$$

Na prática, na presença de ruído, devemos resolver o sistema de equações anterior como sendo

$$\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|^2 = 0.$$

Se $\mathbf{A}^T \mathbf{A}$ é inversível, a solução do problema da equação anterior é dada por

$$\mathbf{x} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}. \quad (\text{B.3})$$

§B.4 Algoritmo Levenberg-Marquardt

Considere a seguinte função não-linear

$$\mathcal{F}(\mathbf{Y}) = \mathbf{X}, \quad (\text{B.4})$$

onde $\mathbf{X} \in \mathbb{R}^M$ e $\mathbf{Y} \in \mathbb{R}^N$ são vetores e $M \geq N$. Frequentemente, é necessário estimar o vetor $\hat{\mathbf{Y}}$ que melhor ajusta-se a um vetor \mathbf{X} medido. Tal problema pode ser reformulado como, dado o vetor \mathbf{X} , encontrar $\hat{\mathbf{Y}}$ que minimiza $\|\epsilon\|$, sujeito a $\mathbf{X} = \mathcal{F}(\hat{\mathbf{Y}}) + \epsilon$.

Dada uma estimação inicial $\hat{\mathbf{Y}}$, o método de Newton (PRESS et al., 1992) a melhora assumindo que $\mathcal{F}(\hat{\mathbf{Y}} + \mathbf{\Delta}) = \mathcal{F}(\hat{\mathbf{Y}}) + \mathbf{J}\mathbf{\Delta}$, onde \mathbf{J} é a matriz jacobiana, ou seja, $\mathbf{J} = \partial\mathbf{X}/\mathbf{Y}$. Dessa forma, minimizar $\|\epsilon\|$ equivale a minimizar

$$\|\epsilon - \mathbf{J}\mathbf{\Delta}\|, \quad (\text{B.5})$$

que é o mesmo que resolver

$$\mathbf{J}^T \mathbf{J} \mathbf{\Delta} = \mathbf{J}^T \epsilon. \quad (\text{B.6})$$

Além disso, a solução refinada, $\hat{\mathbf{Y}}_r$, é dada por

$$\hat{\mathbf{Y}}_r = \hat{\mathbf{Y}} + \mathbf{\Delta} \quad (\text{B.7})$$

e pode ser melhorada iterativamente.

Levenberg (1944) propôs alterações no método de Newton que aceleram, significativamente, a sua convergência. Tais alterações foram aperfeiçoadas por Marquardt (1963) e, atualmente, o algoritmo é chamado Levenberg-Marquardt.

No algoritmo Levenberg-Marquardt, a diagonal principal da matriz $\mathbf{J}^T \mathbf{J}$, da equação (B.6), é multiplicada por $(1 + \lambda)$. Tipicamente, λ é feito igual a 10^{-4} inicialmente. Contudo, a cada iteração o valor de λ é alterado. Se a solução de (B.6) conduzir a uma redução do resíduo, λ é dividido por 10. Caso contrário, este é multiplicado por 10.

A versão do algoritmo de Levenber-Marquardt utilizada neste trabalho é a contida na biblioteca VXL (VANROOSE, 2004). Tal implementação permite definir diversos aspectos da minimização e, baseado em uma estimação da matriz Hessiana, nos permite saber se a solução convergiu para um mínimo local ou não.

ARTIGOS PUBLICADOS

Neste apêndice, apresenta-se a lista de artigos publicados pelo autor durante o desenvolvimento do trabalho e que têm relação com o mesmo. Como outros artigos ainda estão sendo escritos, espera-se que a lista de publicações aumente após a versão final deste texto ser concluída.

§C.1 Artigos publicados em periódicos

José A. de França; Marcelo R. Stemmer e Maria B. de M. França, "Algoritmo robusto para correspondência de pontos em imagens estereoscópicas na ausência de calibração", *Semina: Ciências Exatas/Tecnológicas*, Londrina, vol. 26, n. 1, 2006. ISSN 1679-0375. In press.

§C.2 Artigos publicados em congressos

José A. de França and Marcelo R. Stemmer, "Um novo algoritmo não-linear para cálculo da matriz fundamental", XV Congresso Brasileiro de Automática, Gramado, Brazil, september 2004.

José A. de França, Maria B. de M. França and Marcelo R. Stemmer, "Stereo-Based Detection and Localization of Obstacles in Indoor Environments", VI Induscon, Joinville, Brazil, 2004.

José A. de França, Maria B. de M. França, Robinson Hoto and Marcelo R. Stemmer, "A Simple Linear and Iterative LMS Algorithm for Fundamental Matrix Estimating", II Encontro de Robótica Inteligente (EnRI) in Proceedings of the XXIV Congresso da Sociedade Brasileira de Computação (CSBC), Salvador, Brazil, 2004.

de França, J.A. e Stemmer, M.R. "Estimação robusta da matriz fundamental", In VI Simpósio Brasileiro de Automação Inteligente, Bauru, São Paulo, pp 253-258, 2003.

Bibliografia

GRANLUND, G. H. The complexity of vision. *Signal Processing*, v. 74, n. 1, p. 101–126, 1999. [online] <http://www.ecvision.info/education/summerschool03/GranlundComplexityofVision.pdf>.

FAUGERAS, O. D. What can be seen in three dimensions with an uncalibrated stereo rig? In: *II European Conference on Computer Vision*. [S.l.: s.n.], 1992. p. 563–578.

HARTLEY, R.; GUPTA, R.; CHANG, T. Stereo from uncalibrated cameras. In: *II European Conference on Computer Vision*. [S.l.: s.n.], 1992. p. 579–587. [online] <http://www.syseng.anu.edu.au/~hartley/My-Papers.html>.

FAUGERAS, O.; LUONG, Q.-T. *The Geometry of Multiple Images: The Laws That Govern the Formation of Multiple Images of a Scene and Some of Their Applications*. [S.l.]: MIT Press, Cambridge, MA, 2001.

MAYBANK, S.; FAUGERAS, O. A theory of self-calibration of a moving camera. *The International Journal of Computer Vision*, v. 8, n. 2, p. 123–151, 1992.

FAUGERAS, O.; MAYBANK, S. Motion from point matches: Multiplicity of solutions. *The International Journal of Computer Vision*, v. 4, n. 3, p. 225–246, 1990. [online] <http://www.inria.fr/rrrt/rr-1157.html>.

ZISSERMAN, A.; BEARDSLEY, P.; REID, I. Metric calibration of a stereo rig. In: *IEEE Workshop on Representations of Visual Scenes*. Boston, EUA: [s.n.], 1995. p. 93–100. [online] <http://citeseer.ist.psu.edu/zisserman95metric.html>.

HORAUD, R.; CSURKA, G.; DEMIRDIJIAN, D. Stereo calibration from rigid motions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 22, n. 12, p. 1446–1452, 2000. [online] <http://citeseer.ist.psu.edu/horaud98stereo.html>.

HARTLEY, R.; ZISSERMAN, A. *Multiple View Geometry in Computer Vision*. [S.l.]: Cambridge University Press, 2000.

ASHENDEN, P. J. *The Designer's Guide to VHDL*. [S.l.]: Morgan Kaufmann, 2002.

LUONG, Q.-T.; FAUGERAS, O. The fundamental matrix: theory, algorithms, and stability analysis. *The International Journal of Computer Vision*, v. 17, n. 1, p. 43–76, 1996. [online] <http://citeseer.ist.psu.edu/luong95fundamental.html>.

LUONG, Q. T. et al. *On determining the fundamental matrix : Analysis of different methods and experimental results*. [S.l.], 1993. 27 p. p. [online] <http://www.inria.fr/rrrt/rr-1894.html>.

HARTLEY, H. Euclidean reconstruction from uncalibrated views. In: *Proceedings II European-US Workshop on Invariance*. [S.l.: s.n.], 1993. p. 187–202. [online] <http://www.syseng.anu.edu.au/~hartley/My-Papers.html>.

ZHANG, Z. Determining the epipolar geometry and its uncertainty: A review. *The International Journal of Computer Vision*, v. 27, n. 2, p. 161–195, August 1998. [online] <http://citeseer.ist.psu.edu/zhang98determining.html>.

TORR, P. H. S.; ZISSERMAN, A. Robust computation and parametrization of multiple view relations. In: DESAI, U. (Ed.). *ICCV6*. [S.l.]: Narosa Publishing House, 1998. p. 727–732. [online] <http://www.cms.brookes.ac.uk/~philiptorr/>.

TORR, P.; MURRAY, D. The development and comparison of robust methods for estimating the fundamental matrix. *International Journal of Computer Vision*, v. 24(3), p. 271–300, 1997. [online] <http://citeseer.ist.psu.edu/torr97development.html>.

BARTOLI, A.; STURM, P. Nonlinear estimation of the fundamental matrix with minimal parameters. *IEEE Trans. on pattern analysis and machine intelligence*, v. 26, n. 3, p. 426–432, 2004. [online] <http://www.inrialpes.fr/movi/publi/Publications/2004/>.

ZHANG, Z.; LOOP, C. Estimating the fundamental matrix by transforming image points in projective space. *Computer Vision and Image Understanding*, v. 82, n. 2, p. 174–180, 2001. [online] <http://citeseer.ist.psu.edu/zhang01estimating.html>.

BARTOLI, A.; STURM, P.; HORAUD, R. Projective structure and motion from two views of a piecewise planar scene. In: *8th International Conference on Computer Vision*. Vancouver, Canada: [s.n.], 2001. v. 1, p. 593–598. [online] <http://citeseer.ist.psu.edu/bartoli01projective.html>.

ZHANG, Z. Camera calibration with one-dimensional objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 26, p. 892–899, 2004. [online] <http://citeseer.ist.psu.edu/zhang02camera.htm>.

AGRAWAL, M.; DAVIS, L. Camera calibration using spheres: A semidefinite programming approach. In: *IX IEEE International Conference on Computer Vision*. [S.l.: s.n.], 2003. p. 782–789. [online] <http://citeseer.ist.psu.edu/agrawal03camera.html>.

ZHANG, Z. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 22, n. 11, p. 1330–1334, 2000. [online] <http://citeseer.ist.psu.edu/zhang98flexible.html>.

STURM, P. F.; MAYBANK, S. J. On plane-based camera calibration: A general algorithm, singularities, applications. In: *IEEE Conf. Computer Vision and Pattern Recognition*. [S.l.: s.n.], 1999. p. 432–437. [online] <http://citeseer.ist.psu.edu/sturm99planebased.html>.

- SCHARSTEIN, D.; SZELISKI, R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, v. 47, n. 1, p. 7–42, 2002. [online] <http://citeseer.ist.psu.edu/scharstein01taxonomy.html>.
- CHEN, Z.; WU, C.; TSUI, H. T. A new image rectification algorithm. *Pattern Recognition Letters*, v. 24, n. 1-3, p. 251–260, 2003.
- LOOP, C.; ZHANG, Z. Computing rectifying homographies for stereo vision. In: *Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 1999. p. 125–131. [online] <http://citeseer.ist.psu.edu/loop99computing.html>.
- HARRIS, C. G.; STEPHENS, M. A combined corner and edge detector. In: *4th Alvey Vision Conference*. Manchester: [s.n.], 1988. p. 147–151.
- ZIOU, D.; TABBONE, S. Edge detection techniques: An overview. *International Journal of Pattern Recognition and Image Analysis*, v. 8, p. 537–559, 1998. [online] <http://citeseer.ist.psu.edu/ziou98edge.html>.
- CANNY, J. F. A computational approach to edge detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, v. 8, n. 6, p. 769–768, 1986.
- SMITH, S. A new class of corner finder. In: *3rd British Machine Vision Conference*. [S.l.: s.n.], 1992. p. 139–148. [online] <http://www.fmrib.ox.ac.uk/~steve/susan/>.
- BANKS, J.; BENNAMOUN, M.; CORKE, P. A hybrid stereo matching algorithm incorporating the rank constraint. In: *7th International Conference on Image Processing And Its Applications*. Manchester, United Kingdom: [s.n.], 1999. p. 33–37.
- HIRSCHMÜLLER, H. Improvements in real-time correlation-based stereo vision. In: *IEEE Workshop on Stereo and Multi-Baseline Vision*. [S.l.: s.n.], 2001. p. 141–148. [online] <http://www.cse.dmu.ac.uk/~hhm/publications.html>.
- BANKS, J.; CORKE, P. Quantitative evaluation of matching methods and validity measures for stereo vision. *The International Journal of Robotics Research*, v. 20, n. 7, p. 512–532, 2001.
- ULLMAN, S. Maximizing rigidity: the incremental recovery of 3D structure from rigid and rubbery motion. *Perception*, v. 13, p. 255–274, 1984.
- PILU, M. A direct method for stereo correspondence based on singular value decomposition. In: *Computer Vision and Pattern Recognition*. [S.l.: s.n.], 1997. p. 261–266. [online] <http://citeseer.ist.psu.edu/pilu97direct.html>.
- GRZYWACZ, N.; YUILLE, A. Massively parallel implementations of theories for apparent motion. *Spatial Vision*, v. 3, n. 1, p. 15–44, 1988. [online] vpl.usc.edu/papers/15-44.pdf.
- SCOTT, G. L.; LONGUET-HIGGINS, H. C. An algorithm for associating the features of two images. *Royal Society of London, B*, n. 244, p. 21–26, 1991.

TRUCCO, E.; VERRI, A. *Introductory Techniques for 3-D Computer Vision*. [S.l.]: Prentice Hall PTR, 1998. ISBN 0132611082.

ZHANG, Z. et al. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence*, v. 78, n. 1-2, p. 87–119, 1995. [online] <http://citeseer.ist.psu.edu/article/zhang94robust.html>.

POLLARD, S.; MAYHEW, J.; FRISBY, J. PFM: A stereo correspondence algorithm using a disparity gradient limit. *Perception*, v. 14, p. 449–470, 1985.

ARMANGUÉ, X.; SALVI, J. Overall view regarding fundamental matrix estimation. *Image and Vision Computing*, v. 21, p. 205–220, 2003.

HUANG, T. S.; FAUGERAS, O. D. Some properties of the e matrix in two view motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 11, n. 12, p. 1310–1312, 1989.

FAUGERAS, O. D. *Three-Dimensional Computer Vision: A Geometric Viewpoint*. [S.l.]: MIT Press, 1993.

LUONG, Q. T.; FAUGERAS, O. D. Self-calibration of a moving camera from point correspondences and fundamental matrices. *International Journal of Computer Vision*, v. 22, n. 3, p. 261–89, 1997. [online] <http://citeseer.ist.psu.edu/luong97selfcalibration.html>.

TRIVEDI, H. P. Can multiple views make up for lack of camera registration? *Image and Vision Computing*, v. 6, n. 1, p. 29–32, 1988.

ZHANG, Z.; XU, G. A general expression of the fundamental matrix for both perspective and affine cameras. In: *V International Joint Conference on Artificial Intelligence*. Nagoya, Japan: [s.n.], 1997. p. 1502–1510. [online] <http://citeseer.ist.psu.edu/zhang97general.html>.

HESSE, O. Die cubische gleichung, von welcher die lösung des problems der homographie von m. chasles abhängt. *J. reine angew. Math*, v. 13, n. 62, p. 188–192, 1863.

STURM, R. Das problem der projektivität und seine anwendung auf die flächen zweiten. In: *Math. Ann. 1*. [S.l.: s.n.], 1869. p. 533–574.

LONGUET-HIGGINS, H. A computer algorithm for reconstructing a scene from two projections. *Nature*, v. 293, n. 10, p. 133–135, September 1981.

HARTLEY, R. In defence of the eight point algorithm. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, v. 19, n. 6, p. 580–593, 1997. [online] <http://rsise.anu.edu.au/~hartley/My-Papers.html>.

TORR, P. H. S. *Outlier Detection and Motion Segmentation*. Tese (Doutorado) — University of Oxford, Engineering Dept, Inglaterra, 1995. [online] <http://wwwcms.brookes.ac.uk/~philiptorr/papers.htm>.

LIU, B.; MÄNNER, R. A linear iterative least-squares method for estimating the fundamental matrix. In: *IEEE International Symposium on Signal Processing and its Applications*. [S.l.: s.n.], 2003. p. 17–20. [online] <http://citeseer.ist.psu.edu/559963.html>.

SAMPSON, P. D. Fitting conics sections to ‘very scattered’ data: An iterative refinement of the Bookstein algorithm. *Computer Graphics and Image Processing*, v. 18, n. 1, p. 97–108, 1982.

CHOJNACKI, W. et al. Revisiting hartley’s normalized eight-point algorithm. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, v. 25, n. 9, p. 1172–1177, 2003. [online] http://www.cs.adelaide.edu.au/users/mjb/mjb_abstracts.html.

TORR, P.; FITZGIBBON, A. Invariant fitting of two view geometry or “In defiance of the eight point algorithm”. In: *British Machine Vision Conference*. [S.l.: s.n.], 2003. [online] <http://www.cms.brookes.ac.uk/~philiptorr/papers.htm>.

CSURKA, G. et al. Characterizing the uncertainty of the fundamental matrix. *Computer Vision and Image Understanding*, v. 68, n. 1, p. 18–36, 1997. [online] <http://citeseer.ist.psu.edu/csurka95characterizing.html>.

FISCHLER, M.; BOLLES, R. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of ACM*, v. 24, n. 6, p. 381–385, 1981.

TORR, P. H. S.; ZISSERMAN, A. MLESAC: A New Robust Estimator with Application to Estimating Image Geometry. *Computer Vision and Image Understanding*, v. 78, n. 1, p. 138–156, 2000. [online] <http://www.cms.brookes.ac.uk/~philiptorr/>.

HARTLEY, R. I.; STURM, P. Triangulation. *Computer Vision and Image Understanding*, v. 68, n. 2, p. 146–157, 1997. [online] <http://citeseer.ist.psu.edu/hartley94triangulation.html>.

ROUSSEEUW, P. J. *Robust Regression and Outlier Detection*. [S.l.]: Wiley, 1987.

TSAI, R. Y. 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, n. 4, p. 323–344, 1987.

SHIH, S.; HUNG, Y.; LIN, W. *Accuracy analysis on the estimation of camera parameters for active vision systems*. Taipei, Taiwan, 1996. 25 pp p. [online] <http://citeseer.ist.psu.edu/shih96accuracy.html>.

SALVI, J.; ARMANGUÉ, X.; BATLLE, J. A comparative review of camera calibrating methods with accuracy evaluation. *Pattern Recognition*, v. 35, p. 1617–1635, 2002. Disponível em: <[online] <http://citeseer.ist.psu.edu/salvi00comparative.html>>.

HEIKKILÄ, J. Geometric camera calibration using circular control points. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 22, n. 10, p. 1066–1076, 2000.

CORRÊA, F. R.; DECCÓ, C. C. G.; JR, J. O. Obtaining range information with an omnidirectional vision system. In: *XVII Congresso Brasileiro de Engenharia Mecânica*. São Paulo, Brasil: [s.n.], 2003. [online] <http://www.pmr.poli.usp.br/lpa/people/fabiano.asp>.

GASPAR, J. et al. Constant resolution omnidirectional cameras. In: *III Workshop on Omnidirectional Vision*. Copenhagen, Denmark: [s.n.], 2002. [online] <http://citeseer.ist.psu.edu/gaspar02constant.html>.

JR., V. G. et al. Development of an omnidirectional vision system. In: *XVI Congresso Brasileiro de Engenharia Mecânica*. Uberlândia, MG, Brasil: [s.n.], 2001. [online] <http://www.pmr.poli.usp.br/lpa/people/valdir.asp>.

LUONG, Q.-T.; VIÉVILLE, T. Canonic representations for the geometries of multiple projective views. *Computer Vision and Image Understanding*, v. 64, n. 2, p. 193–229, 1996. [online] <http://citeseer.ist.psu.edu/luong94canonic.html>.

FRANÇA, J. A. de et al. A simple linear and iterative LMS algorithm for fundamental matrix estimating. In: *I Encontro de Robótica Inteligente (EnRI) in Proceedings of the XXIV Congresso da Sociedade Brasileira de Computação*. Bahia, Brasil: [s.n.], 2004. [online] <http://research.eeol.org/publications/index.htm>.

RUF, A.; HORAUD, R. Vision-based guidance and control of robots in projective space. In: *European Conference on Computer Vision*. [S.l.: s.n.], 2000. p. 50–66. [online] <http://citeseer.ist.psu.edu/ruf00visionbased.html>.

HESPANHA, J. P. et al. What tasks can be performed with an uncalibrated stereo vision system? *International Journal of Computer Vision*, v. 35, n. 1, p. 65–85, 1999. [online] <http://citeseer.ist.psu.edu/hespanha97what.html>.

BEARDSLEY, P. A. et al. Active visual navigation using non-metric structure. In: *International Conference on Computer Vision*. [S.l.: s.n.], 1995. p. 58–64. [online] <http://citeseer.ist.psu.edu/beardsley95active.html>.

ZELLER, C.; FAUGERAS, O. Applications of non-metric vision to some visual guided tasks. In: *International Conference on Pattern Recognition*. Jerusalem, Israel: [s.n.], 1994. p. 132–136. [online] <http://citeseer.ist.psu.edu/zeller94applications.html>.

MENDONÇA, P.; CIPOLLA, R. A simple technique for self-calibration. In: *IEEE International Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 1999. p. 500–505. [online] citeseer.ist.psu.edu/mendonca99simple.html.

KRUPPA, E. Zur ermittlung eines objektes aus zwei perpesktiven mitinnerer orientierung. *Abhandlungen der Akademie der Wissenschaften, Wien*, v. 122, p. 1939–1948, 1913.

BOUGNOUX, S. From projective to euclidean space under any practical situation, a criticism of self calibration. In: *VI International Conference on Computer Vision*. Bombav, India: [s.n.], 1998. p. 790–796.

HABED, A.; BOUFAMA, B. S. Camera self-calibration from two views. In: *IEEE International Conference on Systems, Man and Cybernetics*. [S.l.: s.n.], 2002. v. 4, p. 5 pp.

LOURAKIS, M. I. A.; DERICHE, R. *Camera Self-Calibration Using the Singular Value Decomposition of the Fundamental Matrix: From Point Correspondences to 3D Measurements*. [S.l.], 1999. 45 p. p. [online] <http://citeseer.ist.psu.edu/lourakis99camera.html>.

- HARTLEY, R. I. Kruppa's equations derived from the fundamental matrix. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, v. 19, n. 2, p. 133–135, 1997. [online] <http://rsise.anu.edu.au/~hartley/My-Papers.html>.
- LOURAKIS, M. I. A.; DERICHE, R. Camera self-calibration using the singular value decomposition of the fundamental matrix. In: *Asian Conference on Computer Vision*. Taipei, Taiwan: [s.n.], 2000. I, p. 403–408.
- HARTLEY, R. Self-calibration from stationary cameras. *Internation Journal of Computer Vision*, v. 22, n. 1, p. 5–23, 1997. [online] <http://rsise.anu.edu.au/~hartley/My-Papers.html>.
- WANG, L. et al. Error analysis of pure rotation-based self-calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 26, n. 2, p. 275–280, 2004. [online] <http://citeseer.ist.psu.edu/article/wang01error.html>.
- HAYMAN, E.; MURRAY, D. The effects of translational misalignment in the self-calibration of rotating and zooming cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 25, n. 8, p. 1015–1020, 2003. [online] <http://citeseer.ist.psu.edu/hayman02effects.html>.
- JI, Q.; DAI, S. Self-calibration of a rotating camera with a translational offset. *IEEE Transactions on Robotics and Automation*, v. 20, n. 1, p. 1–14, 2004.
- CSURKA, G.; DEMIRDJIAN, D.; HORAUD, R. Finding the collineation between two projective reconstructions. *Computer Vision and Image Understanding*, v. 75, n. 3, p. 260–268, 1999. [online] <http://citeseer.ist.psu.edu/article/csurka98finding.html>.
- ZHANG, Z.; LUONG, Q. T.; FAUGERAS, O. Motion of an uncalibrated stereo rig: Self-calibration and metric reconstruction. *IEEE Transactions on Robotics and Automation*, v. 12, n. 1, p. 103–113, 1996. [online] citeseer.ist.psu.edu/article/zhang93motion.html.
- MATABOSCH, C.; SALVI, J.; FOREST, J. Stereo rig geometry determination by fundamental matrix decomposition. In: *Workshop on European Scientific and Industrial Collaboration*. Mickolj, Hungria: [s.n.], 2003. [online] <http://citeseer.ist.psu.edu/demirdjian98autocalibration.html>.
- ENCISO, R.; VIEVILLE, T. Self-calibration from four views with possibly varying intrinsic parameters. *IEEE Trans on Pattern Analysis and Machine Intelligence*, v. 15, p. 293–305, 1997. [online] <http://graphics.usc.edu/~renciso/publi.html>.
- DEVERNAY, F.; FAUGERAS, O. From projective to euclidean reconstruction. In: *IEEE International Computer Vision and Pattern Recognition*. San Francisco, California: [s.n.], 1996. p. 264–269. [online] <http://citeseer.ist.psu.edu/devernay96from.html>.
- DEMIRDJIAN, D.; ZISSERMAN, A.; HORAUD, R. Stereo autocalibration from one plane. In: *European Conference on Computer Vision*. Dublin: [s.n.], 2000. v. 2, p. 625–639. [online] <http://citeseer.ist.psu.edu/demirdjian00stereo.html>.
- CSURKA, G. et al. Closed-form solutions for the Euclidean calibration of a stereo rig. *Lecture Notes in Computer Science*, v. 1406, p. 426–443, 1998. [online] <http://citeseer.ist.psu.edu/article/csurka98closedform.html>.

BROWN, M.; BURSCHKA, D.; HAGER, G. Advances in computational stereo. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, v. 25, n. 8, p. 993–1008, 2003. [online] <http://jobber.cs.jhu.edu/~burschka/>.

ZABIH, R.; WOODFILL, J. Non-parametric local transforms for computing visual correspondence. In: *3rd European Conference on Computer Vision*. [S.l.: s.n.], 1994. v. 2, p. 151–158. [online] <http://citeseer.ist.psu.edu/article/zabih94nonparametric.html>.

KANADE, T.; OKUTOMI, M. A stereo matching algorithm with an adaptive window: Theory and experiment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 16, n. 9, p. 920–932, 1994.

HIRSCHMÜLLER, H.; INNOCENT, P.; GARIBALDI, J. Real-time correlation-based stereo vision with reduced border errors. *International Journal of Computer Vision*, v. 47, n. 1, p. 229–246, 2002. [online] <http://citeseer.ist.psu.edu/hirschmuller02realtime.html>.

FUSIELLO, A.; ROBERTO, V.; TRUCCO, E. Efficient stereo with multiple windowing. In: *IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 1997. p. 858–863. [online] <http://citeseer.ist.psu.edu/fusiello97efficient.html>.

LUONG, Q. T.; FAUGERAS, O. D. Determining the fundamental matrix with planes: Instability and new algorithms. In: *Conference on Computer Vision and Pattern Recognition*. New York, EUA: [s.n.], 1993. p. 489–494. [online] <http://citeseer.ist.psu.edu/luong93determining.html>.

LUONG, Q.-T.; FAUGERAS, O. D. On the determination of epipoles using cross-ratios. *Computer Vision and Image Understanding*, v. 71, n. 1, p. 1–18, 1998. [online] <http://citeseer.ist.psu.edu/luong98determination.html>.

LEVENBERG, K. A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Math.*, p. 164–168, 1944.

MARQUARDT, D. W. An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Indust. Appl. Math.*, v. 11, n. 2, p. 431–441, 1963.

BOUGUET, J. T. *Camera Calibration Toolbox for Matlab*. [S.l.]: MRL - Intel Corp., 2004. [online] http://www.vision.caltech.edu/bouguetj/calib_doc/index.html.

LINDBERG, T. *Scale-Space Theory in Computer Vision*. [S.l.]: Kluwer Academic Publishers, 1994.

KANATANI, K.; KANAZAWA, Y. Automatic thresholding for correspondence detection. *International Journal of Image and Graphics*, v. 4, n. 1, p. 21–33, 2004. [online] <http://www.suri.it.okayama-u.ac.jp/~kanatani/data/ejournal.html>.

GOLUB, G. H.; Van Loan, C. F. *Matrix Computations*. [S.l.]: The Johns University Press, 1996.

PRESS, W. H. et al. *Numerical Recipes in C: The Art of Scientific Computing*. [S.l.]: Cambridge University Press, 1992. [online] <http://lib-www.lanl.gov/numerical/bookcpdf.htm>.

VANROOSE, P. *VxL: Vision-something-Libraries*. [S.l.]: The TargetJr Consortium, 2004. [online]
<http://vxl.sourceforge.net/>.

Índice

- algoritmo
 - Levenberg-Marquardt, 89, 122
 - LMedS, 37
 - MLESAC, 37
 - MSAC, 36
 - RANSAC, 34
- cálculo
 - da decomposição de Cholesky, 120
 - da homografia do infinito, 56
 - da intersecção de uma reta com a reta no infinito, 116
 - da matriz fundamental, 26
 - algoritmo dos 8 pontos, 28
 - com um bastão graduado, 79
 - contribuições, 69, 70
 - critério de qualidade, 26
 - derivação algébrica, 26
 - impondo o posto, 30
 - método ótimo, 32
 - método LMedS, 37
 - método MLESAC, 37
 - método MSAC, 36
 - método RANSAC, 34
 - minimizando a Distância Sampson, 32
 - minimizando distâncias dos pontos as retas epipolares, 31
 - solução com 7 pontos, 27
 - da projeção ortogonal de um ponto em uma reta, 117
 - das retas epipolares, 23
 - do centro da câmera, 10
 - do plano no infinito, 53, 55
 - dos epipolos, 20, 22, 71
- calibração de câmeras, 39
 - automática, 44
 - calibração projetiva, 42
 - contribuições, 79
 - método clássico, 39
 - por etapas, 51
- caracteres interessantes
 - definição, 12
 - segmentação de, 12
- centro da câmera, 7
- centro da imagem, 7
- cone absoluto, 45
- cone dual, 46
- coordenadas homogêneas, 5
- correlação
 - pela relação cruzada, normalizada e de média zero, 14
 - pelo somatório das diferenças absolutas (SAD), 14
 - pelo somatório das distâncias Hamming, 65
- correspondência de pontos, 10
 - “boas correspondências”, 34
 - “falsas correspondências”, 34
- contribuições, 64
- densas, 12
- esparsas, 12
- fatores que dificultam a, 10
- restrições, 13
 - disparidade, 16
 - invariância esquerda-direita, 15
 - rigidez global, 16
 - semelhança, 14
 - unicidade, 14
 - vizinhança, 17
- técnicas não-paramétricas, 65

- desvio padrão robusto, 38
- disparidade, 16
- distância
 - Hamming, 65
 - Sampson, 32
- distorção
 - projetiva, 11
 - radiométrica, 11
- epipolo, 20
- equação
 - de Kruppa, 49
 - de projeção da câmera, 10
 - de projeção de um conjunto binocular, 21
 - do cone absoluto, 45, 46
 - fundamental, 21, 22, 25
 - homogênea, 9
- geometria epipolar, 20
- homografia, 71
 - T_{PE} , 43
 - da calibração projetiva, 42
 - do infinito, 50
 - epipolar, 23
 - induzida por um plano, 117
- matriz
 - antissimétrica, 5
 - de projeção da câmera, 10
 - de projeção de um conjunto binocular, 21
 - essencial, 25
 - fundamental, 21
- matriz fundamental, 22
- número de subconjuntos do RANSAC, 35
- normalização
 - da homografia H_{12} , 53
 - de coordenadas de pontos 2D, 28, 75
 - de coordenadas pontos 3D, 40
- oclusão, 11
- parâmetro extrínseco
 - matriz de rotação R , 10
 - vetor translação ou t , 10, 26
- parâmetro intrínseco
 - $n_{x/y}$, 9
 - s_k , 9
 - ângulo θ_{sk} , 9
 - distância focal, 7
 - ponto principal ou m_0 , 9
- parâmetros
 - extrínsecos da câmera, 10
 - intrínsecos
 - da câmera ideal, 9
 - da câmera não-ideal, 9
- paralaxe virtual, 71
- plano
 - de imagem, 7
 - epipolar, 20
 - no infinito, 53
- razão cruzada, 116
- reconstrução 3D, 59
- reta epipolar, 20
- sistema de coordenadas
 - da câmera, 7
 - da imagem, 7
 - do ambiente, 10
- transformada
 - censo, 65
 - posto, 65