

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

RODRIGO DA SILVA GUERRA

**Calibração Automática de Sistemas de
Visão Estéreo a Partir de Movimentos
Desconhecidos**

Porto Alegre
2004

RODRIGO DA SILVA GUERRA

**Calibração Automática de Sistemas de
Visão Estéreo a Partir de Movimentos
Desconhecidos**

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Rio Grande do Sul como parte dos requisitos para a obtenção do título de Mestre em Engenharia Elétrica.

Área de concentração: Automação e Instrumentação Eletro-Eletrônica

ORIENTADOR: Prof. Dr. Walter Fetter Lages

Porto Alegre
2004

RODRIGO DA SILVA GUERRA

**Calibração Automática de Sistemas de
Visão Estéreo a Partir de Movimentos
Desconhecidos**

Esta dissertação foi julgada adequada para a obtenção do título de Mestre em Engenharia Elétrica e aprovada em sua forma final pelo Orientador e pela Banca Examinadora.

Orientador: _____
Prof. Dr. Walter Fetter Lages, UFRGS
Doutor pelo Instituto Tecnológico de Aeronáutica – São José dos Campos – Brasil

Banca Examinadora:

Prof. Dr. Altamiro Amadeu Susin, UFRGS
Doutor pelo Institut National Polytechnique de Grenoble – Grenoble, França

Prof^a. Dr^a. Letícia Vieira Guimarães, UFRGS
Doutora pelo Muroran Institute of Technology – Muroran, Japão

Prof. Dr. Teodiano Freire Filho Bastos, UFES
Doutor pela Universidad Complutense de Madrid – Madrid, Espanha

Coordenador do PPGEE: _____
Prof. Dr. Carlos Eduardo Pereira

DEDICATÓRIA

Dedico este trabalho à minha esposa Helena e à minha mãe Solange.

AGRADECIMENTOS

Agradeço aos meus professores e colegas do Programa de Pós-Graduação em Engenharia Elétrica da UFRGS por todo apoio e auxílio prestados destacando principalmente a CAPES pelo apoio financeiro, o professor Walter pela constante disponibilidade, e o colega Kühne pelo auxílio na formatação deste documento.

RESUMO

Esta dissertação apresenta um método para calibração automática de um par de câmeras que realiza um movimento desconhecido. O processo de calibração aqui descrito é baseado inteiramente na rigidez da cena observada e na invariabilidade dos parâmetros intrínsecos das câmeras durante o movimento. Não há a necessidade de nenhum artefato de calibração especial. Este trabalho mostra uma abordagem completa, desde o processo de casamento de pontos até a estimação dos parâmetros da câmera. Não é feita nenhuma suposição sobre o movimento das câmeras, tampouco sobre a disposição da cena 3D.

Palavras-chave: Calibração de câmeras, visão estéreo, estimação de parâmetros.

ABSTRACT

This dissertation presents a method for automatic calibration of a pair of cameras undergoing an unknown motion. The calibration process here described is entirely based on the scene rigidity and on the invariance of the cameras intrinsic parameters during the movement. There is no necessity for any special calibration artifact. This work shows a complete approach, from the point matching process to the estimation of camera parameters. No assumption is done about the knowledge of cameras motion or the scene 3D layout.

Keywords: Camera calibration, stereo vision, parameter estimation.

SUMÁRIO

LISTA DE ILUSTRAÇÕES	17
LISTA DE TABELAS	19
LISTA DE ABREVIATURAS	21
LISTA DE SÍMBOLOS	23
1 INTRODUÇÃO	27
1.1 Motivação	27
1.2 Calibração de câmeras	28
1.3 Problema abordado	30
1.4 Organização do documento	31
2 MODELAGEM	33
2.1 Matriz de projeção de perspectiva	34
2.2 Parâmetros intrínsecos	36
2.2.1 Relação entre os parâmetros intrínsecos e a cônica absoluta	38
2.3 Parâmetros extrínsecos	39
2.4 Modelo completo	40
2.5 Sistemas Binoculares	41
3 METODOLOGIA	45
4 PAREAMENTO DE PONTOS	49
4.1 Detector de pontos de interesse	50
4.2 Correlação cruzada	51
4.3 Força de um pareamento	54
4.4 Processo de relaxamento	54
4.5 Resultados no pareamento	55
4.6 Reincidência	59
5 ESTIMAÇÃO DA GEOMETRIA EPIPOLAR	63
5.1 Estimação linear	63
5.2 Estimação não-linear	64
5.3 Resultados para dados simulados	65
5.4 Resultados para dados reais	65

6	RECONSTRUÇÃO NO ESPAÇO PROJETIVO	69
6.1	Triangulação	69
6.2	Reconstrução	71
7	MOVIMENTO PROJETIVO	75
7.1	Estimação do movimento projetivo	75
7.1.1	Estimação linear	76
7.1.2	Estimação não-linear	76
7.2	Resultados	77
8	COMPUTAÇÃO DOS PARÂMETROS DA CÂMERA	79
8.1	Resultados para dados reais	82
9	CONCLUSÃO	85
	REFERÊNCIAS	87
	APÊNDICE A CONCEITOS DE GEOMETRIA PROJETIVA	91
A.1	Introdução	91
A.2	Espaços projetivos n-dimensionais	92
A.2.1	Base projetiva	92
A.2.2	Colineação	93
A.2.3	A reta	93
A.2.4	Pontos fixos	94
A.2.5	Mudança de base projetiva	94
A.3	O plano projetivo: \mathbb{P}^2	95
A.3.1	Coordenadas homogêneas	95
A.3.2	Espaço de raios	95
A.3.3	A esfera unitária	96
A.3.4	Plano \mathbb{R}^2 aumentado	97
A.3.5	A reta no \mathbb{P}^2	97
A.3.6	Relações entre \mathbb{R}^2 e \mathbb{P}^2	98
A.3.7	O ponto e a reta no infinito	98
A.3.8	Base projetiva do \mathbb{P}^2	99
A.3.9	Relações entre retas e pontos	99
A.3.10	Dualidade entre ponto e reta no \mathbb{P}^2	100
A.3.11	Lápis de retas	101
A.3.12	Razão cruzada	101
A.3.13	Razão cruzada de quatro retas que se interceptam em um ponto	102
A.3.14	Colineação no \mathbb{P}^2	102
A.3.15	Cônicas	103
A.3.16	Pontos absolutos	104
A.3.17	Transformação de similaridade	105
A.4	O espaço projetivo: \mathbb{P}^3	106
A.4.1	Coordenadas homogêneas	106
A.4.2	O plano no \mathbb{P}^3	107
A.4.3	O plano no infinito	107
A.4.4	Razão cruzada entre quatro planos	108
A.4.5	Lápis de planos	108

A.4.6	Colineação no \mathbb{P}^3	108
A.4.7	Transformação de similaridade	108
A.4.8	Quádricas	110
A.4.9	Cônica absoluta	110

LISTA DE ILUSTRAÇÕES

Figura 1.1:	Exemplo de objeto para calibragem baseada em referência 3D.	29
Figura 2.1:	Formação da imagem numa câmera.	33
Figura 2.2:	Modelo simplificado.	34
Figura 2.3:	Projeção de um ponto.	34
Figura 2.4:	Projeção de um ponto próximo ao plano focal.	35
Figura 2.5:	Mudança do sistema de coordenadas no plano da imagem.	36
Figura 2.6:	Parâmetros intrínsecos.	38
Figura 2.7:	Parâmetros extrínsecos.	39
Figura 2.8:	Sistema de visão estéreo.	41
Figura 2.9:	Parâmetros intrínsecos de um sistema binocular.	42
Figura 2.10:	Restrição epipolar.	43
Figura 4.1:	Efeito do tamanho da janela na detecção de pontos de interesse.	52
Figura 4.2:	Efeito do valor do limiar na detecção de pontos de interesse.	53
Figura 4.3:	Detecção de pontos de interesse pelo operador de <i>Moravec</i>	53
Figura 4.4:	Semelhança de posição entre n e n' em relação a m e m' respectivamente.	54
Figura 4.5:	Efeito da variação do tamanho da janela de correlação cruzada no pareamento de pontos de duas imagens.	56
Figura 4.6:	Efeito do valor limiar de correlação cruzada no pareamento de pontos de duas imagens.	57
Figura 4.7:	Efeito do valor limiar de de força de um pareamento no casamento de pontos de duas imagens.	58
Figura 4.8:	Dois pares de imagens, um antes e outro após o movimento.	59
Figura 4.9:	Exemplo de pontos encontrados nas quatro imagens através da constatação das reincidências.	61
Figura 5.1:	Geometria epipolar recuperada através da estimação da matriz fundamental.	67
Figura 6.1:	Inconsistências na geometria epipolar: raios podem não se cruzar.	70
Figura 6.2:	Pontos retificados para caírem com exatidão sobre suas retas epipolares.	72
Figura 6.3:	Reconstrução numa base projetiva arbitrária.	73
Figura 7.1:	Deslocamento no espaço Euclideano e correspondente movimento no espaço projetivo.	75
Figura 8.1:	Reconstrução de pontos no \mathbb{R}^3	83

LISTA DE TABELAS

Tabela 4.1:	Número de pontos de interesse para diferentes valores de w_i	51
Tabela 4.2:	Número de pontos de interesse para diferentes valores de l_i	51
Tabela 4.3:	Número de pontos pareados para diferentes valores de w_μ	55
Tabela 4.4:	Número de pontos pareados para diferentes valores de l_μ	59
Tabela 4.5:	Número de pontos pareados para diferentes valores de ε_r	59
Tabela 5.1:	Minimização do critério da expressão (5.5) para diferentes valores de A em (5.9).	66
Tabela 7.1:	Minimização do critério da expressão (7.8) para diferentes valores de A em (5.9).	77

LISTA DE ABREVIATURAS

2D	bidimensional, planar ou de duas dimensões
3D	tridimensional, espacial ou de três dimensões
IA	inteligência artificial
imagem-da	imagem da câmera da direita no instante anterior ao movimento
imagem-dd	imagem da câmera da direita no instante posterior ao movimento
imagem-ea	imagem da câmera da esquerda no instante anterior ao movimento
imagem-ed	imagem da câmera da esquerda no instante posterior ao movimento
pares-ad-d	conjunto dos pares de pontos encontrados entre imagem-da e imagem-dd
pares-ad-e	conjunto dos pares de pontos encontrados entre imagem-ea e imagem-ed
pares-ed-a	conjunto dos pares de pontos encontrados entre imagem-ea e imagem-da
pares-ed-d	conjunto dos pares de pontos encontrados entre imagem-ed e imagem-dd

LISTA DE SÍMBOLOS

$a \simeq b$	significa que $\lambda a = b$ para algum λ arbitrário
\overline{ab}	é o seguimento de reta que une os pontos a e b
a'	parâmetros com o sinal “ ’ ” correspondem à câmera da direita
$[a]_{\times}$	é a matriz anti-simétrica que aplica o produto vetorial
\hat{A}	o sinal “ ^ ” representa a operação de <i>smoothing</i> , i.e. retorna o valor médio.
$\mathbf{0}_n$	vetor $[0 \ 0 \ \cdots \ 0]^T$ de n elementos
\mathbf{c}	imagem do centro ótico da câmera da esquerda sobre o plano retinal da mesma
$\tilde{\mathbf{c}}$	centro ótico da câmera da esquerda
$\hat{\mathbf{C}}$	matriz das diretivas direcionais
$d(\cdot)$	função que aplica um movimento qualquer (rotação e translação)
$d(\mathbf{a}, \mathbf{b})$	distância entre os pontos ou retas \mathbf{a} e \mathbf{b}
\mathbf{D}	transformação do \mathbb{P}^3 que preserva π_{∞} e Ω , uma colineação (veja a seção 2.3)
\mathbf{e}	epipólo
f	distância focal da câmera da esquerda
f_{ij}	elemento da linha i e coluna j da matriz \mathbf{F}
\mathbf{F}	matriz fundamental do sistema de visão binocular
$\hat{\mathbf{F}}$	matriz \mathbf{F} transformada
h_{ij}	elemento da linha i e coluna j da matriz \mathbf{H}
$I(x, y)$	intensidade do píxel nas coordenadas (x, y)
\bar{I}_W	valor médio de $I(x, y)$ na janela W
\mathbf{I}_n	matriz identidade $n \times n$
$\mathbf{I}_{\vec{x}}$	derivativa horizontal
$\mathbf{I}_{\vec{y}}$	derivativa vertical
k	razão entre os fatores de escala vertical e horizontal
\mathbf{K}	matriz de parâmetros intrínsecos da câmera da esquerda
k_c	constante de ajuste do detector de cantos de <i>Plessey</i> , geralmente igual a 0,04

k_u	parâmetro intrínseco: $1/k_u$ representa a altura de um píxel
k_v	parâmetro intrínseco: $1/k_v$ representa a largura de um píxel
L	matriz que aplica uma translação
m	vetor de coordenadas de um ponto na imagem da câmera da esquerda
$\tilde{\mathbf{m}}$	vetor de coordenadas de um ponto no \mathbb{P}^3
$\overline{\mathbf{m}}$	vetor composto pelas primeiras 3 coordenadas de $\tilde{\mathbf{m}}$
$\hat{\mathbf{m}}$	ponto da imagem, próximo a m , que satisfaz com exatidão a geometria epipolar
$\tilde{\mathbf{m}}_n$	vetor de coordenadas de um ponto tridimensional no sistema de coordenadas n
n	ponto vizinho de m
$\mathcal{N}(\mathbf{m})$	vizinhança do ponto m dentro de um determinado raio
o	coordenadas da origem no plano retinal para base em píxeis
P	matriz de projeção da câmera da esquerda, numa base do \mathbb{P}^3 , ou seja, sem parâmetros intrínsecos
\mathbf{P}_1	matriz de projeção da câmera da esquerda levando em conta apenas o parâmetro intrínseco f
\mathbf{P}_4	matriz de projeção da câmera de 4 parâmetros intrínsecos
\mathbf{P}_5	matriz de projeção da câmera com todos os 5 parâmetros intrínsecos
P	matriz de projeção da câmera da esquerda normalizada
$\overline{\mathbf{P}}$	matriz composta pelo bloco 3×3 esquerdo de P
$\hat{\mathbf{P}}$	matriz que representa a projeção numa base projetiva que em geral não corresponde ao espaço Euclidiano
$\hat{\mathbf{p}}_i$	i -ésima linha de $\hat{\mathbf{P}}$
\mathbf{P}_E	matriz de projeção da câmera da esquerda calibrada no \mathbb{R}^3
\mathbb{P}^2	espaço projetivo 2D, ou plano projetivo
\mathbb{P}^3	espaço projetivo 3D
Q	matriz que aplica uma rotação
r	$r\alpha$ representa a inclinação da imagem
R	raio que determina uma vizinhança
R	matriz de rotação
\mathbb{R}^3	espaço Euclidiano 3D
$s(t)$	polinômio a ser minimizado para triangulação
S	matriz que dá a mudança de escala $\mathbf{i} = \mathbf{S}\mathbf{i}_n$ e $\mathbf{j} = \mathbf{S}\mathbf{j}_n$
t	vetor de translação
U	matriz utilizada na estimação linear de F
u_0	coordenada horizontal da imagem do centro de projeção em píxeis

v_0	coordenada vertical da imagem do centro de projeção em píxeis
α	parâmetro intrínseco: fator de escala horizontal
α_u	parâmetro intrínseco: distância focal expressa em píxeis horizontais
α_v	parâmetro intrínseco: distância focal expressa em píxeis verticais
φ	plano focal
γ	reta epipolar
$\gamma(t)$	reta epipolar parametrizada em função de t
λ	fator de escala arbitrário diferente de zero
$\mu_{\mathbf{m},\mathbf{m}'}$	correlação cruzada entre \mathbf{m} e \mathbf{m}'
Ω	cônica absoluta
ω_ρ	imagem de Ω sobre o plano ρ
ρ	plano retinal
$\sigma_{\mathbf{m},\mathbf{m}'}$	medida de força de um pareamento
τ_{xy}	custo que determina a existência de um ponto de interesse
θ	parâmetro intrínseco: representa o ângulo entre \mathbf{i}_n e \mathbf{j}_n

1 INTRODUÇÃO

O barateamento das câmeras e aumento da velocidade dos microprocessadores têm possibilitado a utilização de técnicas de visão computacional em escalas que eram antes inconcebíveis. Apesar da utilização da visão artificial para fins de extração de informação 3D (a chamada *visão estereoscópica*) já vir sendo estudada há anos, apenas nas mais recentes décadas esta tecnologia tornou-se viável para aplicações práticas. Exemplos deste avanço já podem ser vistos em diversas áreas, onde destacam-se principalmente a robótica móvel (LAGES, 1998) e os sistemas de *visual servoing*.

Visão computacional engloba o processamento de imagens porém não resume-se a isto. Em processamento de imagens estudam-se formas de transformar a imagem no seu estado bruto em outra imagem mais adequada a algum objetivo em questão. Em visão computacional estudam-se formas de utilizar a visão artificial de uma ou mais câmeras para extrair-se algum tipo de informação útil, comumente 3D.

A visão computacional e a computação gráfica são assuntos que se complementam. Se de um lado busca-se obter informações a partir de imagens, do outro tenta-se gerar imagens a partir de informações. Em ambos os casos são necessários modelos que descrevam de forma adequada o comportamento geométrico do sistema de projeção das câmeras e do ambiente.

1.1 Motivação

Em diversos trabalhos na área de veículos autônomos, informações 3D sobre o ambiente baseiam-se em sensores como sonares (LEONARD; DURRANT-WHYTE, 1992) e lasers para determinar as posições e orientações do robô, do seu alvo ou trajetória e de seus obstáculos. A visão estereoscópica é uma ferramenta que tende a ser eficiente para este fim, pois oferece uma maior densidade de dados do que os sonares com mesma rapidez e menor consumo do que scanners a laser. Além disso, possibilita também a integração de diversas aplicações intimamente associadas à visão, como por exemplo reconhecimento de padrões (DRAPER, 1996) e reconhecimento facial (SUNG; POGGIO, 1998).

Na área de IA existem pesquisas onde estuda-se o aprendizado realimentado por visão e o aprendizado por imitação baseada em visão. Projetos ousados, como o *RoboCup Humanoid Challenge* (KITANO; ASADA, 1998), têm tirado vantagem de avanços na área da visão 3D para navegação e interação entre diversos agentes, inclusive na construção de robôs autônomos para auxílio no resgate de sobreviventes em escombros (KITANO, 2000). Também já existem diversos robôs humanóides, onde destacam-se o Asimo da Honda e o QRIO da Sony, os quais fazem uso de sistemas de visão estereoscópica para interpretação de gestos, identificação de falantes, localização de objetos e auxílio à navegação.

Existem estudos de visão estereoscópica com diversos arranjos de câmeras, como por exemplo através do uso de duas ou três câmeras fixas entre si, ou simplesmente utilizando uma única câmera em movimento. Inspirados na visão humana, os sistemas binoculares parecem ser os mais comumente utilizados para este fim. Em qualquer um destes casos, é preciso ter-se um modelo fiel destas câmeras para viabilizar a extração de informação métrica 3D a partir de imagens 2D.

Todos estes estudos de visão estereoscópica baseam-se em modelos que representam de forma adequada o funcionamento do sistema de visão em questão. Estes modelos descrevem através de expressões o comportamento geométrico do processo de aquisição das imagens. Seja qual for o modelo escolhido, é imprescindível que se conheça todas as constantes que diferenciam uma câmera de outra. O processo de levantamento destas constantes é portanto crucial para uma correta representação do um sistema de visão estéreo segundo qualquer modelo.

1.2 Calibração de câmeras

A calibração de câmeras é o processo através do qual estimam-se as diversas constantes que definem as propriedades geométricas associadas a uma ou várias câmeras segundo um determinado modelo. Mais recentemente alguns estudos têm explorado também a utilização de câmeras não calibradas para aplicações de visão estéreo, mas isto impõe algumas restrições que limitam as informações que podem ser obtidas a apenas algumas propriedades básicas associadas a espaços projetivos.

O assunto de calibração de câmeras começou a ser estudado na fotogrametria (BROWN, 1971; FAIG, 1975), onde buscava-se determinar posições, dimensões e orientações de objetos ou pessoas em determinadas cenas fotografadas, muitas vezes para aplicações na ciência forense.

A maioria dos conceitos básicos que fundamentam estes trabalhos são baseados em teorias de geometria projetiva e visão computacional (FAUGERAS, 1993). Conceitos de geometria projetiva relevantes ao entendimento deste trabalho são revisados no apêndice A.

As técnicas de calibração de câmeras podem ser classificadas a grosso modo em duas categorias: calibração baseada em objeto de referência 3D e auto-calibração.

- **Calibração baseada em objeto de referência 3D.** É realizada através da observação de um objeto cuja geometria no espaço 3D é conhecida com grande precisão. A calibração pode ser realizada muito eficientemente. O objeto de calibração comumente consiste em dois ou três planos ortogonais entre si como mostrado na figura 1.1 (ZHANG; FAUGERAS; DERICHE, 1997). Em alguns casos também é possível utilizar-se um plano que sofre deslocamentos. Estes tipos de abordagem requerem aparatos de calibração dispendiosos e arranjos elaborados.
- **Auto-calibração.** Não se utiliza nenhum objeto específico como referência para calibração. A calibração é realizada apenas pelo movimento da câmera em uma cena estática. A partir de um deslocamento, a rigidez da cena fornece restrições aos parâmetros das câmeras, unicamente através da informação das imagens capturadas. Este tipo de calibração tem a vantagem de ser mais versátil e não exigir conhecimentos especializados, facilitando o uso para usuários leigos (ZHANG, 2000).

Com o barateamento e a popularização da tecnologia das câmeras é também interessante o aperfeiçoamento de técnicas simples e baratas para calibração destas câmeras,

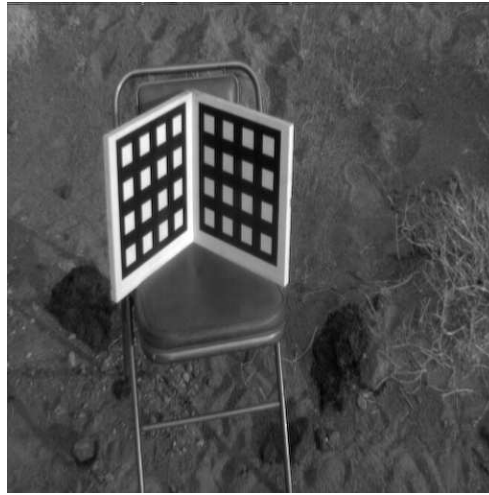


Figura 1.1: Exemplo de objeto para calibragem baseada em referência 3D.

possibilitando assim também a popularização de aplicações baseadas em visão computacional 3D de baixo custo.

Para calibração de uma única câmera destacam-se trabalhos que abordam auto-calibração, utilizando as imagens de uma superfície plana em pelo menos duas orientações diferentes como referência. Estes métodos tendem a ser consideravelmente eficazes e bastante simples e baratos, já que a superfície pode ser uma impressão de boa qualidade sobre a capa rígida de um livro. Em (ZHANG, 2000) qualquer superfície pode ser utilizada, mas requer-se casamento de pontos entre as imagens. Em (MENG; HU, 2003) é utilizada a impressão de um círculo com algumas retas passando pelo seu centro, facilitando assim a localização automática dos pontos de interesse e evitando assim o casamento de pontos entre imagens.

No caso de sistemas de visão binoculares destacam-se mais recentemente os trabalhos de auto-calibração baseada na aquisição de imagens de uma cena em diferentes pontos de vista. Inicialmente, em (FAUGERAS; LUONG; MAYBANK, 1992) sugeriu-se um método que baseava-se na solução das chamadas *expressões de Kuppria*, mas isto envolvia métodos de resolução de expressões não-lineares, dificultando o processo. Em (MOONS et al., 1996) é descrito um método baseado na detecção de pontos de fuga através de movimentos de translação pura, no entanto isto restringe e dificulta o processo de calibração. Em (BEARDSLEY et al., 1995; ZISSERMAN; BEARDSLEY; REID, 1995) é mostrado que movimentos no espaço Euclidiano e respectivos movimentos no espaço projetivo são conjugados. Com isto, estes autores deram também uma grande contribuição ao mostrar que a calibração para o espaço afim é uma propriedade algébrica, definida pelos auto-vetores da transformação 3D projetiva. Vide apêndice A para uma discussão de conceitos de geometria projetiva. A partir daí, (HORAUD; CSURKA; DEMIRDJIAN, 2000) desenvolveram soluções baseadas totalmente em álgebra linear, onde é feita uma abordagem estratificada, calibrando primeiramente no espaço afim (recuperando o plano no infinito) que é seguida pela calibração Euclideana. Estas soluções estratificadas apresentam a vantagem de permitir resoluções inteiramente baseadas em álgebra linear. Em (CSURKA; DEMIRDJIAN; HORAUD, 1998) são mostradas soluções fechadas, onde não é necessário computar explicitamente o plano no infinito, já que se sabe que esta é uma etapa bastante crítica no processo. Existem outras técnicas como calibração através de pontos de fuga para direções ortogonais e calibração a partir de rotações puras (HARTLEY,

1994), mas estas soluções exigem montagens muito específicas e movimentos complexos e controlados. Apesar da auto-calibração ser uma abordagem bastante flexível, estudos neste sentido ainda estão em fase de amadurecimento (ZHANG, 2000). Este trabalho aborda o estado da arte neste tema.

1.3 Problema abordado

Esta dissertação aborda o problema da auto-calibração de um sistema de visão estéreo binocular, ou seja, a estimação dos parâmetros que modelam geometricamente o sistema de projeção de um par de câmeras, sem a utilização de um artefato especial de calibração. Este tipo de abordagem é vantajosa por permitir a calibração totalmente automática das câmeras, sem a necessidade de um arranjo particular. Não é preciso construir-se nenhum tipo de estrutura física especial, pois a calibração baseia-se unicamente nas imagens da cena-alvo. O uso da própria imagem da cena-alvo para a calibração permite que esta seja realizada baseada exatamente o que se deseja medir. Além disso, a interferência humana é mínima ou nenhuma. Todas estas características tornam este tipo de método de calibração adequado para sistemas de visão estéreo de baixo custo, permitindo a fácil calibração por um usuário não especialista. Este tipo de calibração também é adequado para situações onde a interferência humana não é viável, como para calibração de câmeras de robôs de exploração planetária, ou em robôs nas profundezas de oceanos, já que estas máquinas possivelmente sofrem alterações no comportamento de seus sistemas de visão devido às adversidades enfrentadas durante o percurso até seu destino, i.e. trepidação, mudança de temperatura, humidade, gravidade, pressão.

O processo de calibração aqui utilizado consiste em aplicar um movimento genérico ao sistema de visão enquanto este observa uma cena 3D estruturada. Para isto supõe-se que as duas câmeras permanecem fixas entre si e que seus parâmetros internos são invariantes.

Não se faz nenhum tipo de suposição a respeito da estrutura geométrica 3D da cena observada. Entretanto, requer-se que esta cena permaneça invariante, ou pelo menos sofra alterações desprezíveis entre os instantes antes e depois do movimento do sistema de visão. A quantidade de relevos e contrastes combinados com a opacidade, iluminação, sombra e textura, e a distribuição espacial dos objetos no enquadramento das imagens são fatores que influenciam a eficiência do algoritmo.

A partir de pelo menos dois pares de imagens desta cena, um par capturado antes e o outro depois do movimento, realiza-se o casamento de pontos através de um algoritmo heurístico (ZHANG et al., 1995). Estas correspondências fornecem dados que permitem então a calibração.

O processo de calibração é composto de diversas etapas:

- Estimação da geometria epipolar (ZHANG, 1998);
- Triangulação (HARTLEY; STURM, 1997);
- Estimação do movimento projetivo (CSURKA; DEMIRDJIAN; HORAUD, 1999);
- Calibração do sistema de forma fechada (CSURKA; DEMIRDJIAN; HORAUD, 1998).

Métodos de auto-calibração semelhantes a este já foram publicados (ZISSERMAN; BEARDSLEY; REID, 1995; CSURKA; DEMIRDJIAN; HORAUD, 1998; HORAUD;

CSURKA; DEMIRDIJIAN, 2000; DERMIDJIAN; CSURKA; HORAUD, 1998). Entretanto não se tem notícia de nenhum trabalho que descreva uma solução completa, incluindo todas as etapas, desde o casamento de pontos até a estimação dos parâmetros, sendo portanto esta a principal contribuição desta dissertação.

1.4 Organização do documento

O capítulo 2 explica a modelagem utilizada e revisa os conceitos básicos mais relevantes ao entendimento do restante do texto. Recomenda-se ao leitor não familiar com os conceitos de geometria projetiva uma breve leitura do apêndice A antes de iniciar a leitura deste capítulo. Referências às seções do apêndice A são feitas quando necessário. No capítulo 3 o método de calibração é sumarizado em poucas palavras, oferecendo uma visão geral do conteúdo. O leitor familiarizado com o assunto pode começar a ler deste capítulo. O capítulo 4 trata do pareamento de pontos entre duas imagens, que é um passo necessário e crucial na calibração. Entretanto este assunto pouco interfere no entendimento dos capítulos subsequêntes. O capítulo 5 explica a estimação da geometria epipolar. O capítulo 6 trata da reconstrução dos pontos pareados no espaço projetivo. O capítulo 7 aborda o assunto da estimação do movimento projetivo. Para facilitar o entendimento dos conceitos, sugere-se a leitura seqüencial dos capítulos 5, 6 e 7. O capítulo 8 fala da estimação dos parâmetros das câmeras a partir da matriz que dá o movimento projetivo. Este capítulo pode ser lido separadamente pelo leitor familiarizado com o assunto. Por fim, apresenta-se a conclusão.

2 MODELAGEM

Neste capítulo será apresentada a modelagem de sistemas de visão baseadas no chamado *modelo pinhole*. Segundo este modelo desprezam-se as distorções provocadas pelo formato geralmente esférico das lentes.

O sistema da figura 2.1 consiste em duas telas. Um pequeno furo foi feito na primeira tela, e por este furo passam alguns dos raios de luz emitidos ou refletidos pelo objeto, formando na segunda tela uma imagem invertida do mesmo. Um modelo deste sistema é ilustrado na figura 2.2. Este modelo consiste em um plano ρ , chamado *plano retinal*, no qual a imagem é formada através de uma operação chamada *projeção de perspectiva*.

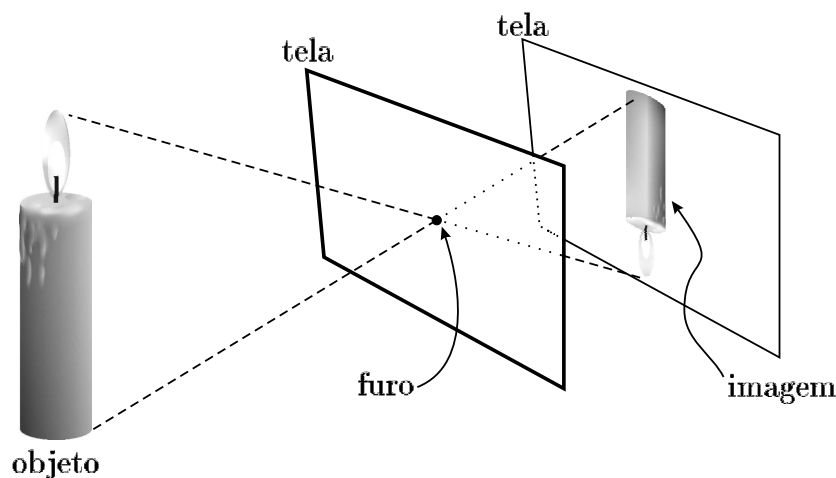


Figura 2.1: Formação da imagem numa câmera.

O ponto \tilde{c} , chamado *centro ótico*, que fica a uma distância f do plano retinal, chamada *distância focal*, é usado para formar a imagem m de um ponto 3D \tilde{m} como sendo a intersecção entre este plano ρ e a reta $\overline{\tilde{c}m}$ (FAUGERAS, 1993).

O chamado *eixo ótico* é a reta $\overline{\tilde{c}c}$ que passa pelo centro ótico \tilde{c} e é perpendicular ao plano retinal ρ . O chamado *plano focal* é aquele plano que passa por \tilde{c} e é paralelo a ρ , ilustrado como plano φ na mesma figura (FAUGERAS, 1993).

Se um dado ponto \tilde{m} do espaço projetivo 3D (de agora em diante referido como IP^3) estiver sobre o plano φ então este ponto não possuirá imagem, pois a reta $\overline{\tilde{c}m}$ é paralela ao plano ρ . Em geometria projetiva isto equivale a dizer que a imagem deste ponto está no infinito.

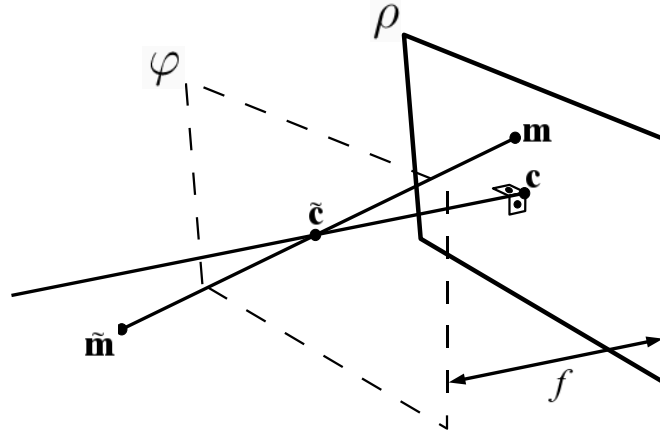


Figura 2.2: Modelo simplificado.

2.1 Matriz de projeção de perspectiva

Um ponto do espaço 3D Euclidiano (de agora em diante referido como \mathbb{R}^3) localizado nas coordenadas (x, y, z) é representado no \mathbb{P}^3 pelo vetor $\tilde{\mathbf{m}} = [\lambda_s x \ \lambda_s y \ \lambda_s z \ \lambda_s]^T$ onde λ_s é um fator de escala arbitrário diferente de zero. A imagem deste ponto, localizada nas coordenadas (u, v) do plano Euclidiano (de agora em diante referido como \mathbb{R}^2), é representada no plano projetivo (de agora em diante referido como \mathbb{P}^2) pelo vetor $\mathbf{m} = [\lambda_p u \ \lambda_p v \ \lambda_p]^T$ sendo λ_p também um fator de escala diferente de zero. Na forma canônica, estes pontos são $\tilde{\mathbf{m}} \simeq [x \ y \ z \ 1]^T$ e $\mathbf{m} \simeq [u \ v \ 1]^T$, sendo que o sinal “ \simeq ” representa igualdade em função de um fator de escala arbitrário.

Por simples relações trigonométricas, observando a figura 2.3 se obtém:

$$-\frac{f}{z} = \frac{u}{x} = \frac{v}{y} \quad (2.1)$$

Em coordenadas projetivas $[-fx/z \ -fy/z \ 1]^T \simeq [-fx \ -fy \ z]^T$, portanto a

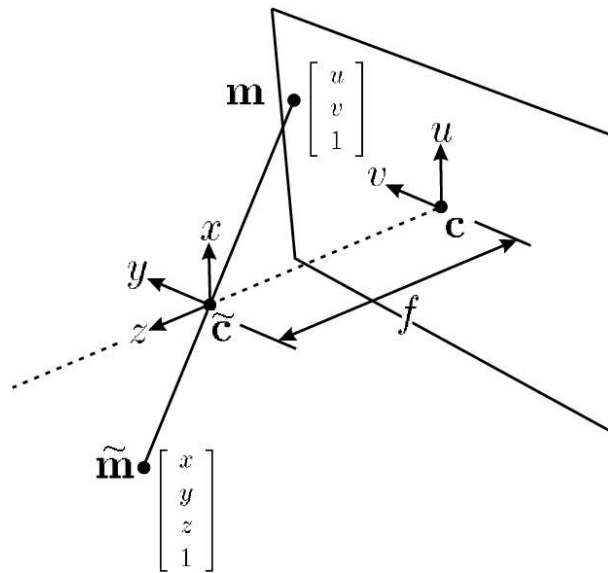


Figura 2.3: Projeção de um ponto.

expressão (2.1) na sua forma matricial, e em coordenadas projetivas, é:

$$\mathbf{m} = \begin{bmatrix} \bar{u} \\ \bar{v} \\ \bar{s} \end{bmatrix} = \begin{bmatrix} -f & 0 & 0 & 0 \\ 0 & -f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (2.2)$$

Se $\bar{s} \neq 0$ tem-se que $u = \bar{u}/\bar{s}$ e $v = \bar{v}/\bar{s}$, sendo u e v as coordenadas da imagem no plano \mathbb{R}^2 .

Quando $\bar{s} = 0$, significa que $z = 0$, ou seja, o ponto $\tilde{\mathbf{m}}$ está sobre o plano focal da câmera, portanto as coordenadas u e v de sua imagem não estão definidas no \mathbb{R}^2 . Apesar disso, a imagem deste ponto existe no \mathbb{P}^2 , onde este ponto é apenas um ponto como outro qualquer. Se $\bar{s} = 0$ diz-se que o ponto \mathbf{m} está no infinito, e portanto as coordenadas \bar{u} e \bar{v} definem apenas uma direção. Isto é facilmente notado ao imaginar-se um ponto se aproximando gradativamente do plano focal conforme mostra a figura 2.4. Todos os pontos que têm $\bar{s} = 0$ são os chamados *pontos no infinito* do plano retinal, e juntos formam a *reta no infinito* do plano retinal. Esta reta é a “imagem” do plano focal.

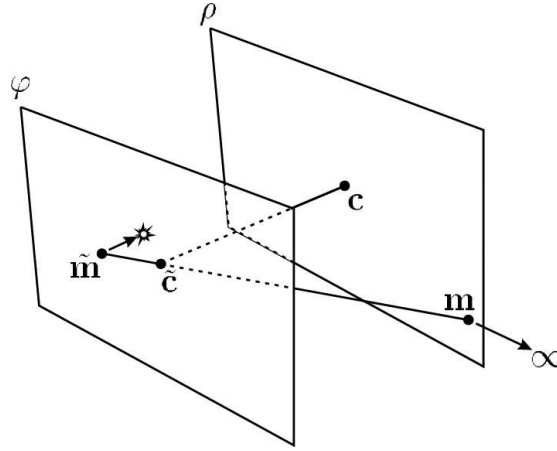


Figura 2.4: Projeção de um ponto próximo ao plano focal.

A expressão (2.2) pode ser escrita de forma mais genérica utilizando-se as coordenadas projetivas $\bar{x} = \bar{t}x$, $\bar{y} = \bar{t}y$ e $\bar{z} = \bar{t}z$ para um $\bar{t} \neq 0$ arbitrário como sendo

$$\mathbf{m} \simeq \mathbf{P}_1 \tilde{\mathbf{m}} \quad (2.3)$$

onde

$$\mathbf{P}_1 = \begin{bmatrix} -f & 0 & 0 & 0 \\ 0 & -f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{m} = \begin{bmatrix} \bar{u} \\ \bar{v} \\ \bar{s} \end{bmatrix} \in \mathbb{P}^2 \quad \text{e} \quad \tilde{\mathbf{m}} = \begin{bmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \\ \bar{t} \end{bmatrix} \in \mathbb{P}^3 \quad (2.4)$$

Uma câmera é, portanto, um sistema que realiza uma transformação projetiva de $\mathbb{P}^3 \rightarrow \mathbb{P}^2$. Conforme demonstrado em (2.3), em coordenadas projetivas esta transformação é representável linearmente.

2.2 Parâmetros intrínsecos

Os chamados *parâmetros intrínsecos* de uma câmera são as constantes que definem como se dá a transformação das coordenadas de um ponto em píxeis para suas coordenadas métricas no plano retinal.

O vetor que representa um ponto no plano pode ser decomposto na soma de dois vetores não colineares e unitários, multiplicados por escalares que são suas coordenadas naquele sistema. Estes dois vetores representam a unidade e a direção vertical e horizontal para um dado sistema de coordenadas. Dependendo da altura, da largura e do formato de cada píxel na retina da câmera, se definirão as coordenadas de um ponto na imagem. Sendo assim, recuperar as coordenadas métricas de um ponto da imagem é equivalente a recuperar as constantes que definem as dimensões dos píxeis da retina da câmera em questão.

Considere um ponto $\mathbf{m} = [\bar{u} \ \bar{v} \ \bar{s}]^T$, onde $u = \bar{u}/\bar{s}$ e $v = \bar{v}/\bar{s}$ são as coordenadas canônicas deste ponto no plano retinal em píxeis. Observe a ilustração da figura 2.5.

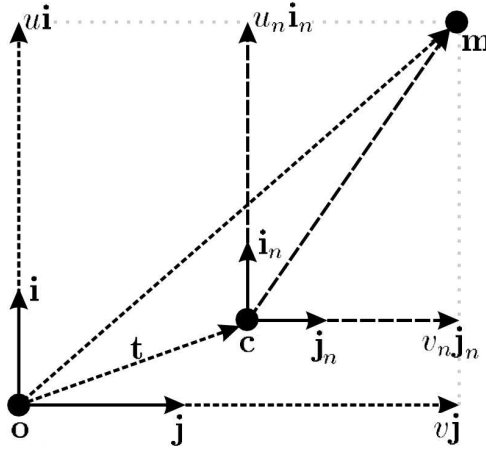


Figura 2.5: Mudança do sistema de coordenadas no plano da imagem.

Tem-se que $\overline{om} = u\mathbf{i} + v\mathbf{j}$ onde \mathbf{i} e \mathbf{j} são dois vetores representando as unidades em píxeis. Estes dois vetores são unitários e ortogonais para o sistema de coordenadas que representam, em píxeis, mas não necessariamente unitários e ortogonais para o sistema de coordenadas em unidades métricas devido ao formato normalmente retangular de cada píxel. Ambos os vetores \mathbf{i} e \mathbf{j} partem do ponto o do plano retinal (geralmente um dos cantos da imagem). Este ponto é a origem do sistema de coordenadas em píxeis.

A representação equivalente deste mesmo ponto \mathbf{m} no sistema de coordenadas do mundo real é dada pelos vetores \mathbf{i}_n e \mathbf{j}_n , representando as unidades métricas. Estes dois vetores são unitários e ortogonais entre si, centrados num ponto \mathbf{c} que é a imagem do centro óptico da câmera.

Assim, tem-se que $\overline{om} = \overline{oc} + \overline{cm}$. No sistema de coordenadas $\mathbf{i}_n \times \mathbf{j}_n$ o ponto é dado pelo vetor $\overline{cm} = u_n\mathbf{i}_n + v_n\mathbf{j}_n$.

Perceba que a mudança do sistema de coordenadas $\mathbf{i} \times \mathbf{j}$ para o sistema de coordenadas $\mathbf{i}_n \times \mathbf{j}_n$ envolve uma mudança de escala e uma translação. A mudança de escala pode ser dada por uma matriz \mathbf{S} tal que $\mathbf{i}_n = \mathbf{S}\mathbf{i}$ e $\mathbf{j}_n = \mathbf{S}\mathbf{j}$, ou seja

$$\mathbf{S} = \begin{bmatrix} k_u & 0 \\ 0 & k_v \end{bmatrix} \quad (2.5)$$

onde k_u é o fator de escala horizontal e k_v é o fator de escala vertical. A translação pode ser denotada por um vetor $t = \overline{oc}$ representado no sistema $\mathbf{i} \times \mathbf{j}$.

Desta forma pode-se escrever a transformação de um sistema de coordenadas para outro em coordenadas projetivas $\mathbf{m} = \mathbf{K}\mathbf{m}_n$ onde $\mathbf{m}_n = [u_n \ v_n \ 1]^T$, $\mathbf{m} = [u \ v \ 1]^T$, e

$$\mathbf{K} = \begin{bmatrix} \mathbf{S} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \quad (2.6)$$

Representando o ponto \mathbf{m} do plano retinal como sendo a projeção de um ponto $\tilde{\mathbf{m}}$ do \mathbb{P}^3 através de (2.3), tem-se

$$\mathbf{m} = \mathbf{K}\mathbf{P}_1\tilde{\mathbf{m}} \quad (2.7)$$

ou seja, a matriz de projeção, que era dada por \mathbf{P}_1 em (2.4) se torna agora $\mathbf{P}_4 = \mathbf{K}\mathbf{P}_1$, sendo

$$\mathbf{P}_4 = \begin{bmatrix} -f\mathbf{S} & \mathbf{t} & \mathbf{0} \\ \mathbf{0} & 1 & 0 \end{bmatrix} = \begin{bmatrix} -fk_u & 0 & u_0 & 0 \\ 0 & -fk_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (2.8)$$

onde $\mathbf{t} = [u_0 \ v_0]^T$.

Os parâmetros $\alpha_u = -fk_u$, $\alpha_v = -fk_v$, u_0 e v_0 não dependem da posição nem da orientação da câmera e por isso são chamados de *parâmetros intrínsecos* da câmera.

Na expressão (2.8) as grandezas x, y, z e f são expressas em unidades de comprimento (metros por exemplo) enquanto que as grandezas u e v estão em píxeis. A partir de (2.7) obtém-se as expressões

$$u = -fk_u \frac{x}{z} + u_0 \quad (2.9)$$

$$v = -fk_v \frac{y}{z} + v_0 \quad (2.10)$$

que mostram que k_u e k_v são expressos em píxeis $\times \text{m}^{-1}$ e α_u e α_v em píxeis. As grandezas $1/k_u$ e $1/k_v$ podem ser interpretadas como sendo a altura e a largura respectivamente de um píxel e os parâmetros α_u e α_v podem ser interpretados como sendo a distância focal expressa em píxeis horizontais e em píxeis verticais respectivamente (vide figura 2.6).

Numa modelagem mais completa deve-se levar em consideração ainda que a grade de píxeis pode não ser exatamente ortogonal, por isso é preciso introduzir o parâmetro θ que representa o ângulo entre \mathbf{i}_n e \mathbf{j}_n , ou seja, o ângulo entre dois lados de um píxel (vide figura 2.6). Assumindo, sem perda de generalidade, que os vetores \mathbf{i} e \mathbf{i}_n ainda são paralelos, e mantendo a relação $\mathbf{i} = \mathbf{S}\mathbf{i}_n$ e $\mathbf{j} = \mathbf{S}\mathbf{j}_n$, tem-se que o novo \mathbf{S} é dado por:

$$\mathbf{S} = \begin{bmatrix} k_u & -\frac{k_u}{\tan \theta} \\ 0 & \frac{k_v}{\sin \theta} \end{bmatrix} \quad (2.11)$$

Agora a matriz de projeção, em (2.8) representada por \mathbf{P}_4 pode ser reescrita de uma forma ainda mais genérica como sendo:

$$\mathbf{P}_5 = \begin{bmatrix} -f\mathbf{S} & \mathbf{t} & \mathbf{0} \\ \mathbf{0} & 1 & 0 \end{bmatrix} = \begin{bmatrix} -fk_u & fk_u \cot \theta & u_0 & 0 \\ 0 & -\frac{fk_v}{\sin \theta} & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (2.12)$$

Na prática é comum ignorar a inclinação da grade de píxeis, fazendo $\theta = 0$. Normalmente esta é uma aproximação bastante razoável para a maioria das câmeras modernas (HORAUD; CSURKA; DEMIRDIJIAN, 2000). Uma câmera modelada de tal forma é conhecida como sendo uma *câmera de quatro parâmetros*. Na sequência deste trabalho será considerado um modelo de câmera de quatro parâmetros.

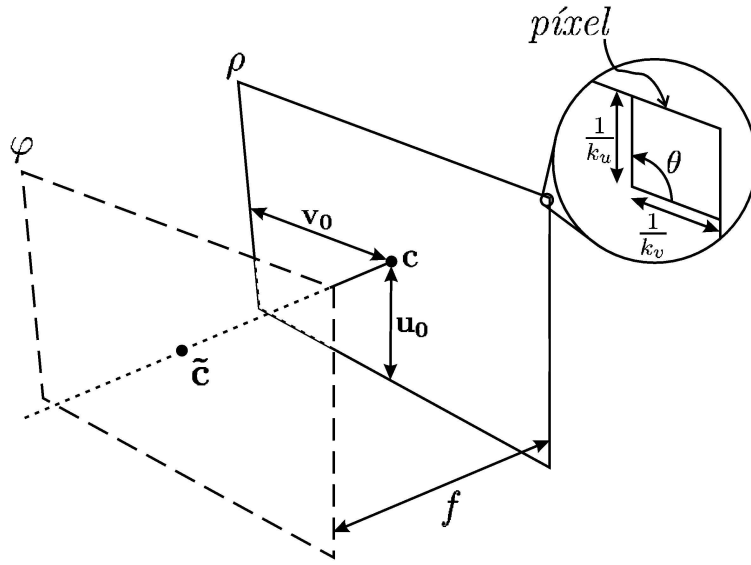


Figura 2.6: Parâmetros intrínsecos.

2.2.1 Relação entre os parâmetros intrínsecos e a cônica absoluta

A cônica absoluta é uma entidade que permanece fixa mediante transformações de similaridade, i.e. translações e rotações (vide seção A.4.9).

Suponha por exemplo uma câmera composta de um plano retinal ρ e de um centro ótico c . Esta câmera, representada pela matriz de projeção P , produzirá sobre ρ a imagem $\omega_\rho = P\Omega$ da cônica absoluta Ω . Ao aplicar-se um movimento qualquer (rotação e translação) $d(\cdot)$ sobre esta câmera, seu novo plano retinal será $d(\rho)$ e seu novo centro ótico será $d(c)$, mas a imagem da cônica será a mesma, pois $\Omega = d(\Omega)$ e $\omega_\rho = Pd(\Omega)$ permanecerá inalterada. A imagem da cônica absoluta depende apenas dos parâmetros intrínsecos da câmera.

Seja $\tilde{m} = [\bar{x} \ \bar{y} \ \bar{z} \ \bar{t}]^T$ um ponto sobre Ω . Este ponto satisfaz (vide seção A.4.9) a expressão:

$$\bar{x}^2 + \bar{y}^2 + \bar{z}^2 = 0 = \bar{t} \quad (2.13)$$

A imagem m do ponto \tilde{m} satisfaz a expressão $m = P\tilde{m}$. Definindo-se $\overline{\tilde{m}}$ e \overline{P} tais que

$$\tilde{m} = \begin{bmatrix} \overline{\tilde{m}} \\ 0 \end{bmatrix} \quad \text{e} \quad P = \begin{bmatrix} \overline{P} & 0_3 \end{bmatrix} \quad (2.14)$$

sendo $\overline{\tilde{m}}$ o vetor composto pelos primeiros três elementos de \tilde{m} , e \overline{P} o bloco 3×3 da esquerda da matriz de projeção P .

A partir daí tem-se que

$$\overline{\tilde{m}} = \overline{P}^{-1} m \quad (2.15)$$

$$\overline{\tilde{m}} = \begin{bmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} m \quad (2.16)$$

Como o ponto \tilde{m} satisfaz (2.13) e (2.15) então

$$\overline{\tilde{m}}^T \overline{\tilde{m}} = 0 \quad (2.17)$$

$$m^T \overline{P}^{-T} \overline{P}^{-1} m = 0 \quad (2.18)$$

Supondo que $\theta = 0$, (2.18) é equivalente a:

$$\mathbf{m}^T \begin{bmatrix} \frac{1}{\alpha_u^2} & 0 & -\frac{u_0}{\alpha_u^2} \\ 0 & \frac{1}{\alpha_v^2} & -\frac{v_0}{\alpha_v^2} \\ -\frac{u_0}{\alpha_u^2} & -\frac{v_0}{\alpha_v^2} & 1 + \frac{u_0^2}{\alpha_u^2} + \frac{v_0^2}{\alpha_v^2} \end{bmatrix} \mathbf{m} = 0 \quad (2.19)$$

Com o ponto \mathbf{m} na sua forma canônica, ou seja $\mathbf{m} = [u \ v \ 1]^T$, pode-se reescrever (2.19) da forma:

$$\left(\frac{u - u_0}{\alpha_u} \right)^2 + \left(\frac{v - v_0}{\alpha_v} \right)^2 + 1 = 0 \quad (2.20)$$

A expressão (2.20) mostra que as coordenadas dos pontos da imagem ω_ρ da cônica absoluta Ω sobre o plano retinal ρ dependem apenas dos parâmetros intrínsecos da câmera e *vice-versa*.

2.3 Parâmetros extrínsecos

Os chamados *parâmetros extrínsecos* descrevem uma mudança de base, da representação do sistema de coordenadas centrado em $\tilde{\mathbf{c}}$ para um novo sistema de coordenadas centrado em um ponto $\tilde{\mathbf{o}}$ qualquer. Esta transformação consiste em uma rotação dada por uma matriz de rotação \mathbf{R} e de uma translação dada por um vetor \mathbf{t} . Vide figura 2.7.

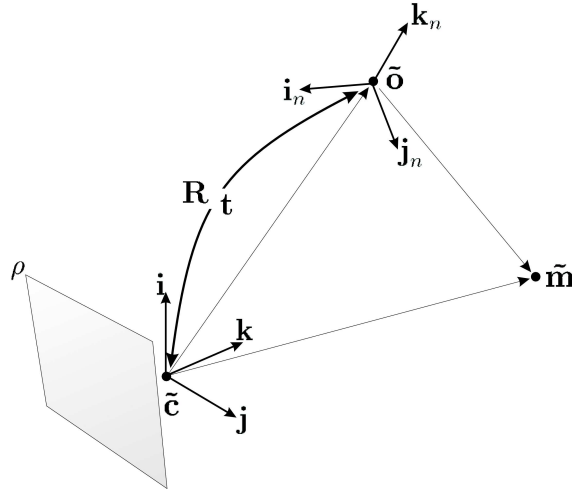


Figura 2.7: Parâmetros extrínsecos.

Considere um ponto $\tilde{\mathbf{m}} = [\bar{x} \ \bar{y} \ \bar{z} \ \bar{t}]^T$ cujas coordenadas euclidianas são $x = \bar{x}/\bar{t}$, $y = \bar{y}/\bar{t}$ e $z = \bar{z}/\bar{t}$. Tem-se que $\overline{\mathbf{c}\tilde{\mathbf{m}}} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ onde \mathbf{i} , \mathbf{j} e \mathbf{k} são versores que formam a base canônica do espaço euclidiano, centrada no ponto $\tilde{\mathbf{c}}$. Deseja-se encontrar a representação deste mesmo ponto num novo sistema de coordenadas $\mathbf{i}_n \times \mathbf{j}_n \times \mathbf{k}_n$, que se encontra rotacionado e transladado para o ponto $\tilde{\mathbf{o}}$.

Observa-se através da figura 2.7 que $\overline{\mathbf{c}\tilde{\mathbf{m}}} = \overline{\mathbf{c}\tilde{\mathbf{o}}} + \overline{\tilde{\mathbf{o}}\tilde{\mathbf{m}}}$. Definindo um vetor \mathbf{t} como sendo a representação de $\overline{\mathbf{c}\tilde{\mathbf{o}}}$ no sistema de coordenadas $\mathbf{i} \times \mathbf{j} \times \mathbf{k}$ e a matriz \mathbf{R} como sendo uma matriz de rotação, então

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{R} \begin{bmatrix} x_n \\ y_n \\ z_n \end{bmatrix} + \mathbf{t} \quad (2.21)$$

Ou seja, no sistema de coordenadas $\mathbf{i}_n \times \mathbf{j}_n \times \mathbf{k}_n$ o ponto $\tilde{\mathbf{m}}$ é

$$\overline{\tilde{\mathbf{om}}} = x_n \mathbf{i}_n + y_n \mathbf{j}_n + z_n \mathbf{k}_n \quad (2.22)$$

O sistema de coordenadas $\mathbf{i}_n \times \mathbf{j}_n \times \mathbf{k}_n$ está apenas rotacionado e transladado em relação ao sistema de coordenadas $\mathbf{i} \times \mathbf{j} \times \mathbf{k}$. Esta operação de rotação seguida de translação pode ser representada no \mathbb{P}^3 por um operador matricial \mathbf{D} tal que

$$\tilde{\mathbf{m}}_n = \mathbf{D} \tilde{\mathbf{m}} \quad (2.23)$$

onde

$$\mathbf{D} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}_3^T & 1 \end{bmatrix} \quad (2.24)$$

A matriz \mathbf{D} é uma transformação do \mathbb{P}^3 que preserva π_∞ e Ω (uma colineação), ou seja, é uma transformação euclídeana (vide seções A.2.2 e A.3.17). A matriz \mathbf{R} e o vetor \mathbf{t} descrevem a posição e a orientação da câmera com respeito ao novo sistema de coordenadas. Estes são os chamados *parâmetros extrínsecos* da câmera.

2.4 Modelo completo

Conforme mostrado nas seções anteriores, percebe-se que os parâmetros intrínsecos e os parâmetros extrínsecos descrevem de forma completa uma câmera segundo o modelo *pinhole*. Sendo assim, uma câmera pode ser considerada como um sistema que depende de parâmetros intrínsecos e extrínsecos. Com base em (2.7) e (2.23) tem-se que

$$\mathbf{m} = \mathbf{KPD} \tilde{\mathbf{m}} \quad (2.25)$$

onde \mathbf{m} é o vetor de coordenadas da imagem, em píxeis, do ponto cujas coordenadas no \mathbb{P}^3 são representadas pelo vetor $\tilde{\mathbf{m}}$. Sendo assim pode-se construir uma matriz de projeção universal

$$\mathbf{P}_4 = \mathbf{KPD} \quad (2.26)$$

onde \mathbf{K} é a matriz definida em (2.6) que contém os parâmetros intrínsecos ($\alpha_u, \alpha_v, u_0, v_0$ e eventualmente θ) e \mathbf{D} é a matriz que contém os seis parâmetros extrínsecos (três para definir as rotações da matriz \mathbf{R} e três para o vetor de translação \mathbf{t}).

Esta matriz \mathbf{P}_4 pode portanto ser escrita (para $\theta = 0$), como sendo:

$$\mathbf{P}_4 = \begin{bmatrix} \alpha_u \mathbf{r}_1 + u_0 \mathbf{r}_3 & \alpha_u t_x + u_0 t_z \\ \alpha_v \mathbf{r}_2 + v_0 \mathbf{r}_3 & \alpha_v t_y + v_0 t_z \\ \mathbf{r}_3 & t_z \end{bmatrix} \quad \text{onde} \quad \mathbf{R} = \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \mathbf{r}_3 \end{bmatrix} \quad \text{e} \quad \mathbf{t} = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \quad (2.27)$$

Esta matriz tem posto completo para $\alpha_u \neq 0$ e $\alpha_v \neq 0$, o que sempre vai acontecer pois o píxel têm dimensão definida. Incluindo o parâmetro θ , temos então a forma mais genérica possível, com os cinco parâmetros intrínsecos da matriz de projeção para uma câmera segundo o modelo *pinhole*:

$$\mathbf{P}_5 = \begin{bmatrix} \alpha_u \mathbf{r}_1 - \frac{\alpha_u}{\tan \theta} \mathbf{r}_2 + u_0 \mathbf{r}_3 & \alpha_u t_x - \frac{\alpha_u}{\tan \theta} t_y + u_0 t_z \\ \frac{\alpha_v}{\sin \theta} \mathbf{r}_2 + v_0 \mathbf{r}_3 & \frac{\alpha_v}{\sin \theta} t_y + v_0 t_z \\ \mathbf{r}_3 & t_z \end{bmatrix} \quad (2.28)$$

Resultando o modelo completo:

$$\mathbf{m} = \mathbf{P}_5 \tilde{\mathbf{m}} \quad (2.29)$$

Assim, uma câmera é representada por uma matriz \mathbf{P} de tamanho 3×4 e posto igual a 3. Um ponto $\tilde{\mathbf{m}} = [\bar{x} \ \bar{y} \ \bar{z} \ \bar{t}]^T$ do \mathbb{P}^3 é relacionado com sua projeção $\mathbf{m} = [\bar{u} \ \bar{v} \ \bar{s}]^T$ no \mathbb{P}^2 por $\mathbf{m} \simeq \mathbf{P}\tilde{\mathbf{m}}$, onde o símbolo “ \simeq ” representa igualdade para um fator de escala arbitrário.

Apesar de qualquer base do \mathbb{P}^3 ser adequada para modelar de forma linear a operação de projeção de perspectiva que uma câmera realiza, é preciso manter-se uma relação entre esta base e o \mathbb{R}^3 , que representa o mundo real e nada mais é do que o \mathbb{P}^3 para uma dada base do em particular.

Para restringir o espaço projetivo a ser o \mathbb{R}^3 , basta reescrever \mathbf{P} em na forma calibrada, como sendo $\mathbf{P}_E \simeq \mathbf{K}[\mathbf{R}_o \ \mathbf{t}_o]$ onde \mathbf{R}_o e \mathbf{t}_o são respectivamente uma rotação e uma translação representando a posição e orientação em relação à origem do sistema de coordenadas no espaço (parâmetros extrínsecos), e \mathbf{K} é a matriz triangular superior com os parâmetros intrínsecos da câmera. A matriz \mathbf{K} costuma também ser escrita de forma simplificada da seguinte maneira

$$\mathbf{K} = \begin{bmatrix} \alpha & r\alpha & u_0 \\ 0 & k\alpha & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2.30)$$

onde α é o fator de escala horizontal, k é a razão entre os fatores de escala vertical e horizontal, $r\alpha$ é a inclinação da imagem e u_0 e v_0 representam as coordenadas em píxeis do centro de projeção.

2.5 Sistemas Binoculares

Este trabalho aborda o problema da auto-calibração de um sistema binocular, ou seja, a estimação dos parâmetros intrínsecos e extrínsecos que descrevem um par de câmeras. Este sistema é composto de duas câmeras rigidamente fixas uma em relação à outra, como pode ser visto na figura 2.8. É portanto adequado considerar o sistema como um todo e não apenas cada câmera individualmente.

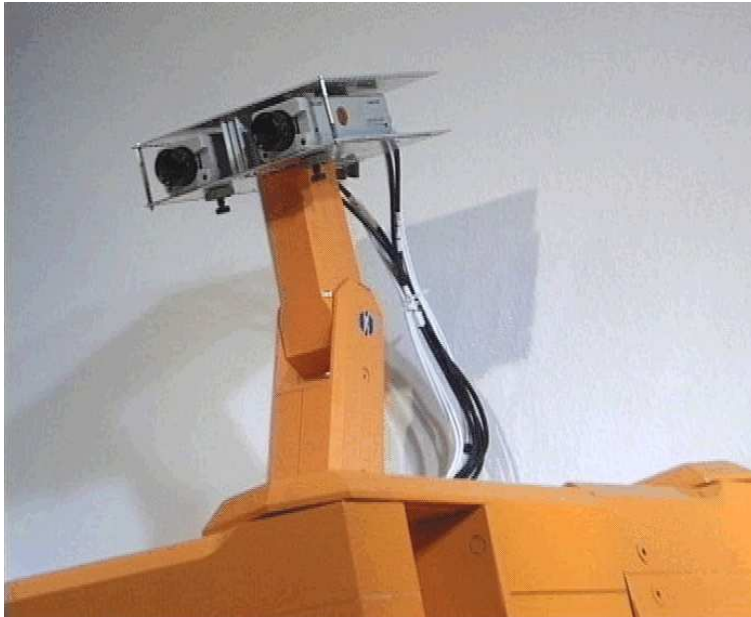


Figura 2.8: Sistema de visão estéreo.

Entende-se por sistema de visão estéreo aquele sistema que busca reconstruir informação espacial a partir de imagens. É possível obter-se visão estéreo a partir de uma sequência de imagens de uma mesma câmera. Apesar disso, a configuração binocular é o arranjo mais simples possível capaz de viabilizar visão estéreo independente de sequenciamento de imagens (FAUGERAS, 1993).

No caso de um sistema binocular, também pode-se considerar parâmetros intrínsecos do sistema de visão a rotação \mathbf{R} e a translação \mathbf{t} que descrevem a diferença de posição entre uma câmera e outra, conhecida por *disparidade*. Para facilitar o equacionamento é comum mover-se a origem do sistema de coordenadas do \mathbb{R}^3 para o centro ótico da câmera da esquerda, representado pelo vetor $\tilde{\mathbf{c}}$. Neste caso pode-se dizer que a disparidade representada por \mathbf{R} e \mathbf{t} descreve a diferença de posição entre $\tilde{\mathbf{c}}$ e $\tilde{\mathbf{c}}'$ que é o centro ótico da câmera da direita. Além disso, introduz-se agora também a matriz \mathbf{K}' que é a matriz dos parâmetros intrínsecos da câmera da direita, cuja imagem do centro ótico é \mathbf{c}' (vide figura 2.9). Já os parâmetros extrínsecos passam a ser os parâmetros que descrevem a posição e orientação do sistema de visão como um todo, no espaço 3D.

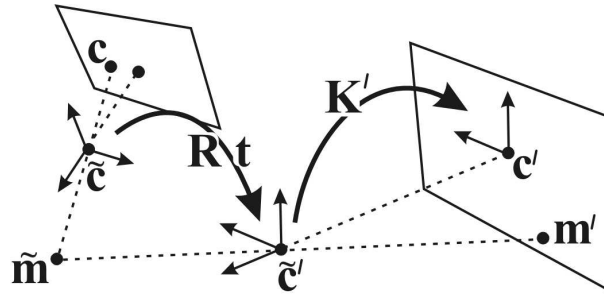


Figura 2.9: Parâmetros intrínsecos de um sistema binocular.

Movendo-se a origem do sistema de coordenadas do \mathbb{R}^3 para o ponto $\tilde{\mathbf{c}}$ e alinhando o mesmo com a reta $\overline{\tilde{\mathbf{c}}\mathbf{c}}$ obtém-se

$$\lambda \mathbf{P}_{\mathbf{E}} = \mathbf{K} \begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_3 \end{bmatrix} \quad (2.31)$$

$$\lambda' \mathbf{P}_{\mathbf{E}'} = \mathbf{K}' \begin{bmatrix} \mathbf{R} & \mathbf{t} \end{bmatrix} \quad (2.32)$$

onde λ e λ' são fatores de escala.

A chamada *restrição epipolar* é uma restrição geométrica muito poderosa associada ao *stereo rig* que transforma um problema bidimensional em um problema unidimensional.

Considere a figura 2.10. Dado um ponto \mathbf{m} na imagem da câmera da esquerda, todos os pontos do \mathbb{R}^3 que podem ter gerado esta imagem estão contidos na semi-reta $\overline{\tilde{\mathbf{m}}\mathbf{c}}$, assim como em particular o próprio ponto \mathbf{m} . Como consequência todos as possíveis imagens deste ponto na câmera da direita, estarão contidas na imagem da semi-reta $\overline{\tilde{\mathbf{m}}\mathbf{c}}$ na retina da câmera da direita, correspondendo à projeção do conjunto de todos os possíveis pontos que teriam condições de ter gerado \mathbf{m} . Esta imagem da semi-reta $\overline{\tilde{\mathbf{m}}\mathbf{c}}$ na retina da câmera da direita é uma outra semi-reta, chamada *reta epipolar*. Obviamente, o ponto \mathbf{m}' , projeção do ponto $\tilde{\mathbf{m}}$ sobre a câmera direita está sobre a reta epipolar. Semelhantemente, tem-se a reta epipolar da esquerda e o epípólo da esquerda. A reta epipolar inclui o ponto \mathbf{e}' , que é a projeção do ponto $\tilde{\mathbf{c}}$ na câmera da direita. O ponto \mathbf{e}' é o chamado *epipólo* da câmera da direita em relação à da esquerda. Todas as retas epipolares de pontos \mathbf{m} na câmera da direita passam pelo epípólo desta câmera.

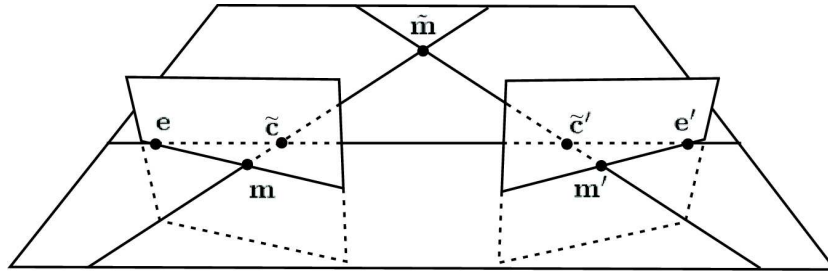


Figura 2.10: Restrição epipolar.

Como mostrado na figura 2.10, ambas as projeções \mathbf{m} e \mathbf{m}' de um ponto tridimensional $\tilde{\mathbf{m}}$, ambos os centros óticos $\tilde{\mathbf{c}}$ e $\tilde{\mathbf{c}}'$, e ambos os epipólos \mathbf{e} e \mathbf{e}' caem sobre um mesmo plano. Esta coplanaridade é expressa por

$$\mathbf{m}'^T \mathbf{F} \mathbf{m} = 0 \quad (2.33)$$

onde $\mathbf{F} = \mathbf{K}'^{-T} \mathbf{t} \times \mathbf{R} \mathbf{K}$ é a chamada *matriz fundamental* do sistema de visão binocular (FAUGERAS, 1993).

Conhecendo a matriz fundamental, o processo de pareamento de pontos é facilitado, pois um ponto \mathbf{m}' na imagem da câmera da direita corresponderá a algum ponto sobre a reta $\gamma \simeq \mathbf{F} \mathbf{m}$ na imagem da câmera da esquerda e correspondentemente um ponto \mathbf{m} na imagem da câmera da esquerda corresponderá a algum ponto sobre a reta $\gamma' \simeq \mathbf{F}^T \mathbf{m}'$ na imagem da câmera da direita. A partir da matriz \mathbf{F} é possível definir-se uma matriz 3×3 que transforma qualquer reta epipolar γ na correspondente reta γ' e vice-versa (FAUGERAS, 1993).

Não é possível recuperar as matrizes de projeção das câmeras calibradas para o \mathbb{R}^3 apenas a partir da matriz fundamental. No entanto, a matriz \mathbf{F} permite a recuperação de duas matrizes de projeção \mathbf{P} e \mathbf{P}' válidas para uma base arbitrária do \mathbb{P}^3 . Estas matrizes de projeção permitem a reconstrução 3D dos pontos nesta base, criando uma representação projetivamente distorcida da estrutura original da cena no \mathbb{R}^3 .

É o acompanhamento das coordenadas tridimensionais dos pontos reconstruídos nesta base arbitrária do \mathbb{P}^3 que permite a calibração do sistema de visão estéreo, e por isso a importância da estimação da matriz fundamental para o processo de auto-calibração.

3 METODOLOGIA

Este trabalho utiliza o sistema de visão binocular de um robô humanóide (vide figura 2.8). O sistema de duas câmeras é montado sobre uma cabeça articulada com duas juntas rotacionais. O sistema de aquisição de imagens é composto por duas placas de captura baseadas no chip BT878 conectadas a um computador Athlon XP 2700 com 512Mb de memória. O software é desenvolvido usando C/C++ sobre o sistema operacional Linux.

Supõe-se que as duas câmeras estão fixas entre si e que seus parâmetros intrínsecos mantêm-se invariantes no tempo. Tal sistema de visão é conhecido como *stereo rig*. O *stereo rig* sofre um movimento genérico e não planar (rotação seguida de translação em diferentes planos). Os cálculos realizados para a estimação dos parâmetros do sistema de visão baseiam-se em dois pares de imagens de uma mesma cena 3D, sendo um par capturado antes e o outro depois do movimento.

O método consiste no cálculo dos parâmetros intrínsecos da câmera da esquerda através da estimação da imagem da cônica absoluta. A imagem da cônica absoluta é recuperada por meio da estimação da colineação que aplica o movimento nos pontos representados em uma base arbitrária do \mathbb{P}^3 . Esta colineação é estimada através das coordenadas 3D dos pontos reconstruídos nesta base. Para viabilizar esta reconstrução faz-se a estimação da geometria epipolar do sistema (ZHANG; FAUGERAS; DERICHE, 1997). Desta forma é possível estabelecer as diversas etapas do processo como sendo:

- Estimação da geometria epipolar;
- Reconstrução dos pontos em uma base do \mathbb{P}^3 ;
- Estimação da colineação que aplica o movimento nesta base;
- Estimação da imagem da cônica absoluta;
- Determinação dos parâmetros intrínsecos da câmera da esquerda.

Todo o método fundamenta-se no acompanhamento das coordenadas da projeção de um mesmo conjunto de pontos da cena observada nas quatro imagens capturadas. A fim de reduzir a carga computacional deste processo seleciona-se apenas alguns pontos das imagens. Por ser necessária a identificação de pontos correspondentes em diferentes imagens, é extremamente interessante que estes pontos sejam bem característicos, facilitando o estabelecimento de correspondências e diminuindo a chance de ocorrência de ambigüidades. Para isto utiliza-se um algoritmo que seleciona alguns pontos de cada imagem, chamados *pontos de interesse*. Estes pontos são o espaço de busca para uma técnica robusta que estabelece correspondências entre pontos de duas imagens diferentes (ZHANG et al., 1995).

O algoritmo de pareamento é responsável pelo estabelecimento de correspondências entre diversos pontos de duas imagens diferentes, formando assim um conjunto de pares. Ao todo estimam-se quatro conjunto de pares de pontos, cada conjunto correspondente respectivamente a diferentes combinações de pares de imagens. Dois destes conjuntos de pares são obtidos respectivamente dos dois pares de imagens obtidas antes e depois do movimento, enquanto que os outros dois conjuntos de pares são obtidos cada um de um dos pares de imagens de uma mesma câmera em diferentes instantes de tempo, como detalhado na seção 4.6.

Sendo invariante a disparidade entre as câmeras, a geometria epipolar também mantém-se inalterada durante o movimento. Dessa forma, pontos dos dois conjuntos de pareamentos casados entre as imagens das câmeras antes e após o movimento podem ser utilizados para estimar a geometria epipolar (ZHANG, 1998). É esperado que o algoritmo de pareamento encontre um pequeno percentual de falsos pares, já que este ainda não faz uso de nenhum tipo de informação sobre a disparidade entre as câmeras. Por isso utiliza-se uma técnica baseada em estatística que permite estimar-se de forma robusta a geometria epipolar mesmo que alguns pareamentos fornecidos sejam incorretos (ZHANG, 1998). Esta técnica baseia-se na seleção, entre muitos grupos de alguns pares, aquele grupo que melhor estima a matriz fundamental. Após esta etapa os pares discrepantes são considerados falsos e descartados. Com os pares restantes utiliza-se agora um algoritmo não-linear de minimização para permitir realizar a refinação do resultado a partir da avaliação de uma função de custo de significado físico bem definido.

De posse da matriz fundamental pode-se encontrar as matrizes de projeção para uma base projetiva arbitrária do \mathbb{P}^3 (ZHANG; FAUGERAS; DERICHE, 1997). Com estas matrizes é possível reconstruir-se os pontos neste espaço projetivo (HARTLEY; STURM, 1997).

Normalmente tem-se um número grande de pares remanescentes em sua maioria com coordenadas perturbadas por ruído e incertezas inerentes ao processo de aquisição. Estes pontos não satisfazem com exatidão a geometria epipolar impossibilitando uma correta reconstrução espacial. Por este motivo as coordenadas de todos os pares remanescentes são retificadas, forçando-os a satisfazer com exatidão a geometria epipolar computada, para a seguir então estimar a reconstrução destes pontos no espaço projetivo por triangulação (HARTLEY; STURM, 1997).

Com as coordenadas dos quatro conjuntos de pareamentos, um algoritmo heurístico encontra pontos 3D correspondentes entre as reconstruções antes e depois do movimento, conforme será detalhado no capítulo 4. Estas correspondências projetivas permitem a estimação da colineação que estabelece o movimento no \mathbb{P}^3 conforme os passos que serão detalhados no texto até seção 7.1. Como acontece na estimação robusta da geometria epipolar, novamente nesta etapa uma técnica estatística é utilizada para estimar a colineação de forma robusta, mesmo na presença de falsos pares. A colineação é então estimada repetidas vezes, iterativamente, e de forma linear para permitir uma massiva computação da mesma para um grande número de amostras de pares de pontos 3D. Assim, escolhem-se aqueles pares 3D que melhor estimam a colineação que aplica o movimento no espaço projetivo, segundo uma função de custo que representa uma grandeza física bem definida. Depois desta primeira estimação da colineação que representa o movimento projetivo, pares estabelecendo correspondências não conformes com a colineação estimada são descartados e uma minimização não-linear com base neste custo representando uma grandeza de significado físico refina o resultado a partir dos pares 3D restantes (CSURKA; DEMIRDJIAN; HORAUD, 1999).

Com esta colinação é possível estimar-se a dual da cônica absoluta e assim os parâmetros intrínsecos da câmera podem ser recuperados através de solução fechada (CSURKA; DEMIRDJIAN; HORAUD, 1998).

4 PAREAMENTO DE PONTOS

A calibração do sistema de visão passa pela estimação da colineação que descreve o movimento do sistema, que por sua vez depende da reconstrução de pontos a partir da geometria epipolar. As técnicas clássicas e robustas conhecidas para a extração da geometria epipolar dependem do encontro de coordenadas de pontos correspondentes nas imagens de duas câmeras (ZHANG, 1998).

Encontrar correspondências entre duas imagens de uma mesma cena é uma das tarefas mais difíceis em visão computacional (ZHANG et al., 1995). A única restrição que se conhece entre duas imagens é a restrição epipolar. Entretanto, em câmeras ainda não calibradas, a própria estimação da geometria epipolar depende do pareamento entre duas imagens. Os métodos encontrados na literatura sempre exploram alguma forma de heurística (ZHANG et al., 1995). Os algoritmos de pareamento conhecidos podem ser classificados em duas categorias principais:

- **Pareamento de moldes:** Nesta categoria os algoritmos buscam correlacionar de alguma forma os níveis de intensidade de retalhos das imagens nas diferentes vistas (BALLARD; BROWN, 1982; GOSHTASBY; GAGE; BARTHOLIC, 1984; HANNAH, 1989; CHOU; CHEN, 1990; FUA, 1993). Este tipo de método possui um argumento válido para áreas relativamente bem texturizadas e para pares de imagens com pequenas diferenças. Entretanto pode retornar correspondências equivocadas nos limites de oclusões ou em regiões com poucas variações de intensidade.
- **Pareamento de características:** Nesta categoria os algoritmos primeiramente salientam primitivas nas imagens, como segmentos ou contornos, e depois encontram correspondências entre as primitivas de duas ou mais diferentes vistas (ULLMAN, 1979; SHAPIRO; MARALICK, 1981; BARNARD; THOMPSON, 1980; CHENG; HUANG, 1984; RADIG, 1984; MAITRE; WU, 1987; HORAUD; SKORDAS, 1989; WENG; AHUJA; HUANG, 1992). Estes métodos são rápidos, pois apenas uma pequena amostra dos píxeis das imagens é utilizada, mas podem falhar se as primitivas não puderem ser detectadas de forma confiável.

Cada método tem suas vantagens e desvantagens de acordo com o objetivo a ser atingido, que pode ser, por exemplo, a reconstrução de volumes 3D, reconhecimento de padrões, etc.

Para o objetivo específico em questão neste trabalho, que é a estimação dos parâmetros das câmeras, é interessante a obtenção de correspondências localizadas, dispersas de forma homogênea pelas imagens, diminuindo assim as chances de mal condicionamento numérico. O método utilizado neste trabalho consiste no pareamento de pontos a partir de técnicas clássicas, passando pela extração de pontos de alta curvatura, como cantos,

a fim de reduzir a quantidade de píxeis processada (ZHANG et al., 1995) sem dar muita margem a ambigüidades.

4.1 Detector de pontos de interesse

Ao tentar-se fazer uma busca exaustiva por correspondências corretas entre todos os píxeis das imagens, a complexidade computacional torna-se proibitivamente alta. Isto se deve ao fato de que cada píxel de uma imagem precisaria ser comparado com cada píxel da outra imagem. Isto se agrava ainda mais ao considerar-se que os algoritmos que realizam pareamento entre pontos exigem uma massiva computação de correlações cruzadas entre janelas centradas nos respectivos pontos candidatos. O algoritmo detector de pontos de interesse tem a função de restringir este espaço de busca.

O detector de pontos de interesse escolhe nas imagens apenas um conjunto de alguns pontos bem caracterizados. Um ponto é entendido como ponto de interesse se uma dada janela no entorno deste ponto apresentar pouca probabilidade de se correlacionar com qualquer outra janela no entorno de outro ponto que não seja ele mesmo. Estes pontos são geralmente aqueles cuja janela no entorno projeta uma imagem de textura rica e contraste forte. Pontos nesta categoria são comumente cantos e bordas. No entanto pontos em bordas não são tão adequados para este fim, já que para pontos ao longo de uma mesma borda há uma tendência de haver alta correlação entre as janelas em seus entornos. Por este motivo é comum utilizar-se detectores de cantos, reduzindo assim a possibilidade de ocorrerem ambigüidades no processo de pareamento. Avaliações de diversos detectores de pontos de interesse podem ser encontradas em (SCHMID; MOHR; BAUCKHAGE, 2000), onde diversos critérios são testados, e onde se destaca o chamado *detector de cantos de Plessey*.

O detector de cantos de Plessey é detalhado em (NOBLE, 1988), e baseia-se na função de custo $\tau_{xy} = \det(\hat{\mathbf{C}}) - k_c \text{trace}^2(\hat{\mathbf{C}})$, onde o autor sugere $k_c = 0,04$ e onde $\hat{\mathbf{C}}$ é uma matriz 2×2 dada por

$$\hat{\mathbf{C}} = \begin{bmatrix} \widehat{\mathbf{I}_{\vec{x}}^2} & \widehat{\mathbf{I}_{\vec{x}}\mathbf{I}_{\vec{y}}} \\ \widehat{\mathbf{I}_{\vec{x}}\mathbf{I}_{\vec{y}}} & \widehat{\mathbf{I}_{\vec{y}}^2} \end{bmatrix} \quad (4.1)$$

onde os elementos de $\mathbf{I}_{\vec{x}}$ e $\mathbf{I}_{\vec{y}}$ são respectivamente as derivadas discretas horizontal e vertical de cada quadrado 3×3 centrado em cada coordenada de uma janela de tamanho $(2w_R + 1) \times (2w_R + 1)$, e “ $\widehat{}$ ” representa a operação de *smoothing*, i.e. retorna o valor médio dos elementos da matriz de derivadas correspondente.

Esta operação é realizada em cada coordenada de ainda outra janela de tamanho $(2w_i + 1) \times (2w_i + 1)$. Se o ponto de máximo τ_{xy} dentro da janela fornece um resultado maior do que um dado limiar, $\tau_{xy} > l_i$, então este ponto é selecionado. Esta janela desliza sobre toda a imagem com passos de w_i píxeis, selecionando no máximo um ponto de interesse de cada vez. Se a distância entre dois pontos selecionados é menor que w_i o ponto para o qual τ_{xy} retorna menor valor é descartado.

As figuras 4.1 e 4.2 mostram respectivamente o efeito do tamanho da janela w_i e do valor do limiar l_i na detecção de pontos de interesse em experimentos com imagens reais. As tabelas 4.1 e 4.2 mostram a quantidade de pontos de interesse encontrados nestas mesmas imagens.

O detector de pontos de interesse de MORAVEC (1977), computacionalmente menos dispendioso, também foi testado, mas apresentou o problema de encontrar muitos pontos ao longo de bordas, aumentando a ambigüidade entre estes. A figura 4.3 mostra os

resultados para este detector, conhecido como operador de *Moravec*.

Tabela 4.1: Número de pontos de interesse para diferentes valores de w_i .

w_i	Número de pares
6	569
9	343
12	252
15	163

Tabela 4.2: Número de pontos de interesse para diferentes valores de l_i .

l_i	Número de pares
10^{-4}	476
10^{-3}	359
10^{-2}	182
10^{-1}	28

4.2 Correlação cruzada

Uma vez definidos os pontos de interesse, o passo seguinte é estabelecer um conjunto de pares candidatos entre os pontos de interesse de duas imagens diferentes. Nesta etapa encontram-se todas as combinações possíveis de pares de pontos que possuem alguma chance de corresponderem entre si. O critério utilizado para definir se um par é ou não candidato é a correlação cruzada entre as janelas no entorno de cada ponto.

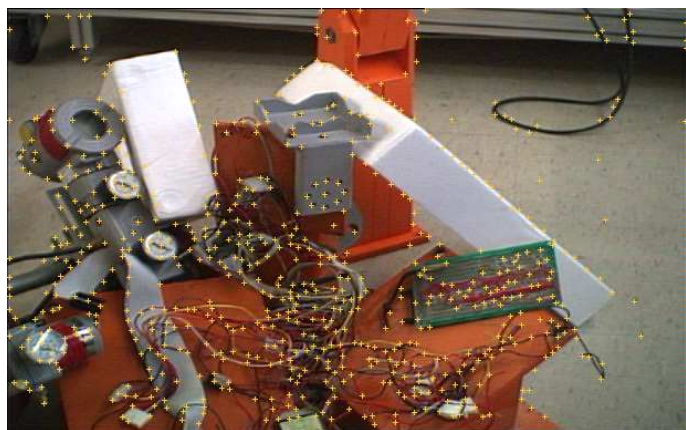
Dados dois pontos de interesse \mathbf{m} e \mathbf{m}' de duas imagens, a correlação cruzada $\mu_{\mathbf{m},\mathbf{m}'}$ é computada entre janelas W de tamanho $(2w_\mu + 1) \times (2w_\mu + 1)$ centradas respectivamente em \mathbf{m} e \mathbf{m}' , usando

$$\mu_{\mathbf{m},\mathbf{m}'} = \frac{\sum_{x,y \in W} [I(x,y) - \bar{I}_W] [I'(x,y) - \bar{I}'_W]}{\sum_{x,y \in W} [I(x,y) - \bar{I}_W]^2 \sum_{x,y \in W} [I'(x,y) - \bar{I}'_W]^2} \quad (4.2)$$

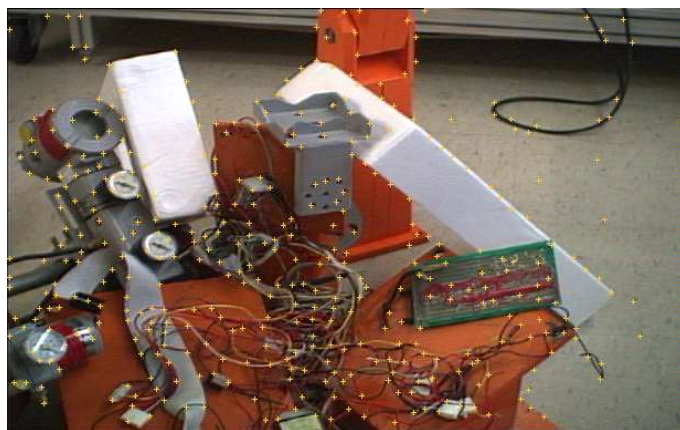
onde $I(x, y)$ é a intensidade do píxel da janela e \bar{I}_W é seu valor médio. Para evitar excesso de computação, pares de pontos cujas distâncias entre as coordenadas são maiores do que um dado raio não são computados. Este raio é ajustado de forma a não ser tão pequeno a ponto de não impor restrição maior do que a disparidade entre as imagens das câmeras e nem tão grande a ponto de candidatar pares cuja distância entre os pontos seja muito maior do que a disparidade poderia ter imposto.

Aqueles pares para os quais $\mu_{\mathbf{m},\mathbf{m}'} > l_\mu$ são considerados candidatos, onde l_μ é um limiar. Este limiar l_μ não deve ser tão baixo a ponto de selecionar pares praticamente não correlacionados e nem tão alto a ponto de não tolerar pequenas diferenças causadas pela disparidade entre as imagens.

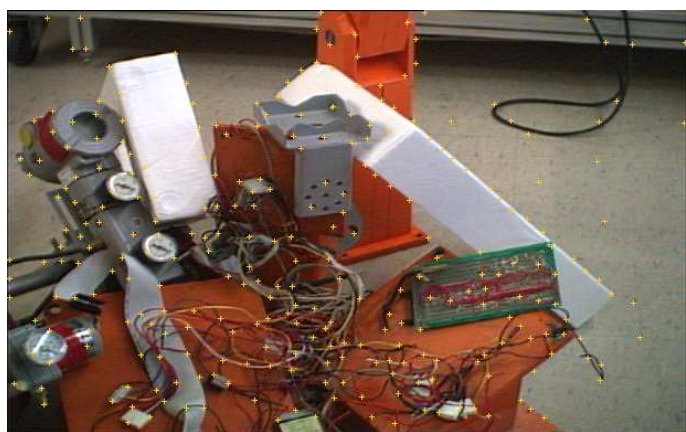
Neste processo um mesmo ponto numa imagem muito provavelmente será candidato várias vezes, formando pares com diferentes pontos na outra imagem. Não se busca nesta etapa estabelecer-se definitivamente os pares correspondentes, mas sim formar um



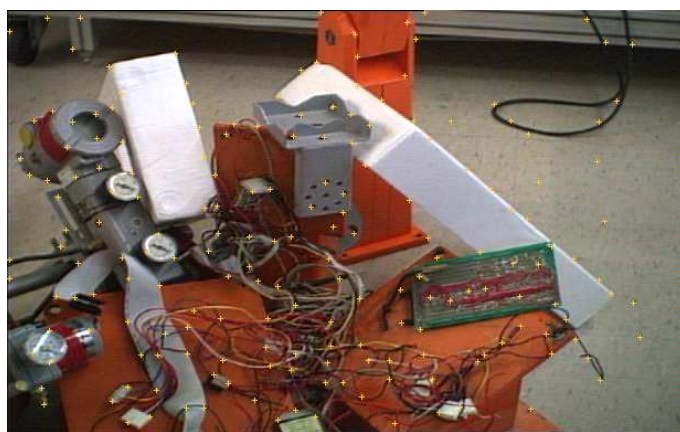
(a) $w_i = 6$.



(b) $w_i = 9$.



(c) $w_i = 12$.

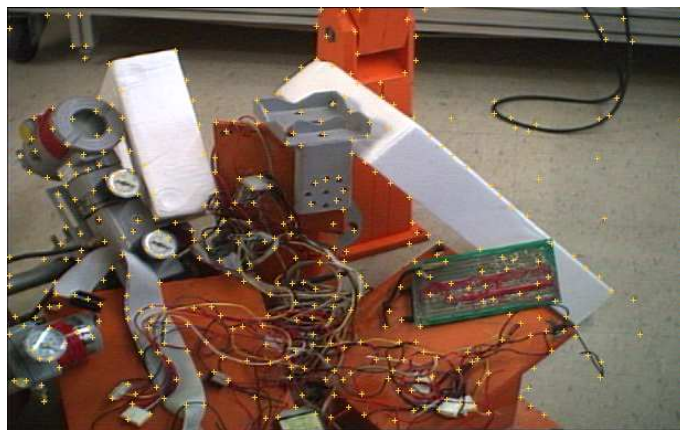


(d) $w_i = 15$.

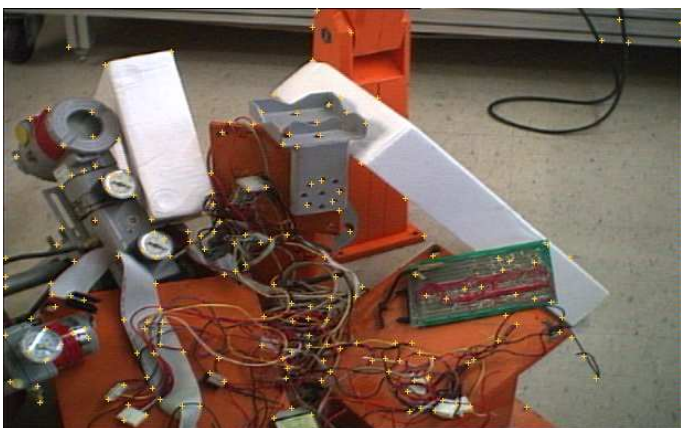
Figura 4.1: Efeito do tamanho da janela na detecção de pontos de interesse.



(a) $l_i = 10^{-4}$.



(b) $l_i = 10^{-3}$.

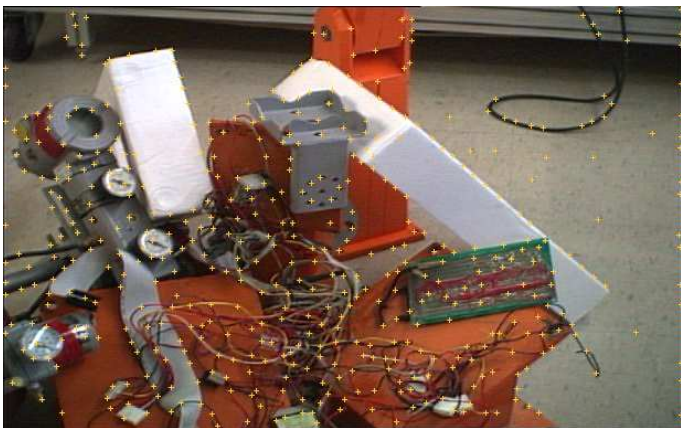


(c) $l_i = 10^{-2}$.



(d) $l_i = 10^{-1}$.

Figura 4.2: Efeito do valor do limiar na detecção de pontos de interesse.



(a) $w_i = 9$.



(b) $w_i = 6$.

Figura 4.3: Detecção de pontos de interesse pelo operador de *Moravec*.

conjunto de pares candidatos como espaço de busca para um método heurístico mais rebuscado.

4.3 Força de um pareamento

Numa mesma cena há uma grande probabilidade de existirem diferentes pontos cujas imagens demonstrem alta correlação entre janelas em seus entornos. Para encontrar-se pares correspondentes de forma robusta é necessário mais do que apenas um critério baseado em correlação cruzada (ZHANG et al., 1995).

É utilizado aqui um critério, chamado de *força do pareamento* (ZHANG et al., 1995), que estabelece uma função de custo que busca favorecer pareamentos cuja localização geométrica dos pontos se assemelhe à localização geométrica de outros pareamentos na vizinhança.

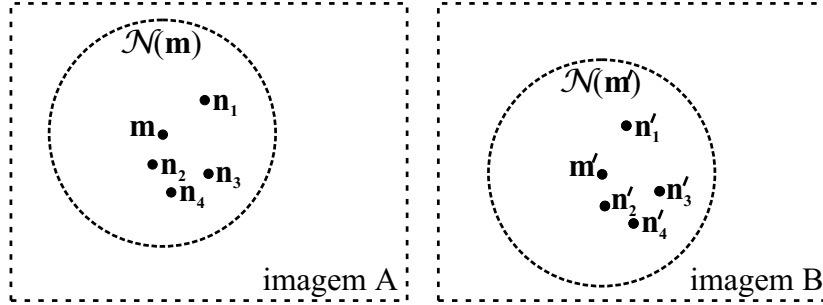


Figura 4.4: Semelhança de posição entre n e n' em relação a m e m' respectivamente.

Sejam $\mathcal{N}(m)$ e $\mathcal{N}(m')$, os conjuntos de pontos vizinhos de m e m' , respectivamente, dentro de um disco de raio R . Se os pontos m e m' formam um bom par, espera-se existir muitos pontos $n \in \mathcal{N}(m)$ e $n' \in \mathcal{N}(m')$ tais que a posição de n em relação a m é semelhante a de n' em relação a m' . Esta medida pode ser dada por (ZHANG et al., 1995)

$$\sigma_{m,m'} = \sum_{\substack{n \in \mathcal{N}(m) \\ n' \in \mathcal{N}(m')}} s \quad (4.3)$$

$$\text{onde } s = \frac{2 \mu_{n,n'} e^{-r/\varepsilon_r}}{2 + \|m - n\| + \|m' - n'\|} \quad \text{e } r = 2 \frac{\left| (\|m - n\| - \|m' - n'\|) \right|}{\|m - n\| + \|m' - n'\|}$$

A soma é realizada sobre todos os pares possíveis para os quais $r > \varepsilon_r$, sendo ε_r um limiar. Se um vizinho de m (ou respectivamente m') é candidato com mais de um vizinho de m' (ou respectivamente m) apenas o par para o qual s for máximo é considerado.

4.4 Processo de relaxamento

Sabe-se que um mesmo ponto muito provavelmente será candidato várias vezes, formando pares com diferentes outros pontos. Assim, estabelecido que dois pontos são correspondentes, eliminam-se todos os outros pares candidatos que também incluam estes pontos. Estes pares eliminados seriam portanto falsos pares e conseqüentemente não deveriam contribuir para a medida de força de outros pareamentos nas vizinhanças. Por este

motivo a força de cada pareamento deve ser recalculada a medida que vão se eliminando falsos pares.

O chamado *processo de relaxamento* é o processo que, de forma iterativa, estabelece quais pontos são correspondentes e elimina as ambigüidades do conjunto de pares. Existem diferentes propostas para se abordar este problema, algumas mais velozes e outras mais robustas (ZHANG et al., 1995). A abordagem *some-winners-take-all* aqui implementada é flexível através do ajuste de um parâmetro q que faz esta troca entre velocidade e robustez (ZHANG et al., 1995). Este processo de relaxamento corresponde ao seguinte procedimento:

1. calcula-se $\sigma_{m,m'}$ para cada par candidato;
2. cria-se uma lista de todos os pares ordenada a partir do mais forte;
3. cria-se outra lista ordenada por $1 - \sigma_{m,m'}/\sigma_{n,n'}$ onde n, n' é o par mais forte para o qual ou $m = n$ ou $m' = n'$;
4. consideram-se pares corretos apenas aqueles pares que estiverem entre os primeiros q por cento de ambas as listas;
5. eliminam-se do processo pares que tenham ambigüidade com pares agora considerados pares corretos;
6. repete-se a partir do passo 1 até que toda a ambigüidade se acabe.

4.5 Resultados no pareamento

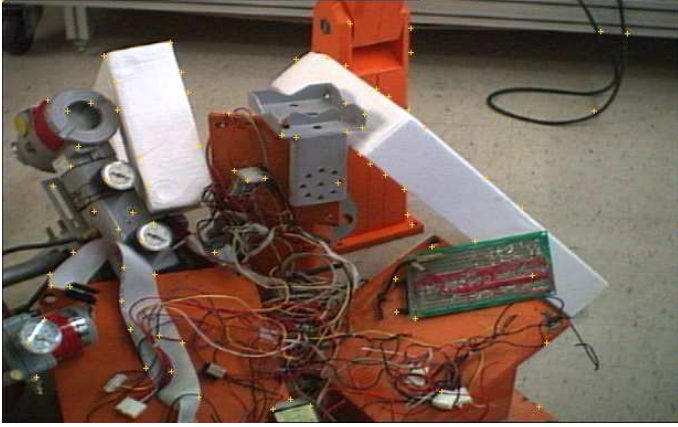
Os resultados dependem muito do ajuste dos muitos parâmetros, limiares e tamanhos de janelas. Tais parâmetros interagem entre si de forma não-linear, dificultando o ajuste fino. É comum ocorrerem alguns pareamentos incorretos que acabam sendo descartados após a estimação da geometria epipolar.

As figuras 4.5, 4.6 e 4.7 mostram respectivamente o efeito dos parâmetros w_μ , l_μ e ε_r no pareamento de pontos de duas imagens. Experimentos demonstram que um w_μ razoavelmente alto combinado com um limiar l_μ entre 0,70 e 0,85 diminui as chances de equívocos no casamento de pontos. As tabelas 4.3, 4.4 e 4.5 mostram o número de pontos pareados em cada par de imagens (nenhum ponto descartado).

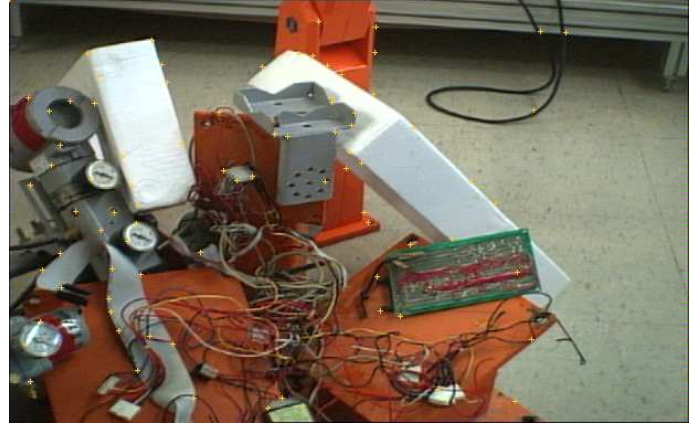
Tabela 4.3: Número de pontos pareados para diferentes valores de w_μ .

w_μ	Número de pares
9	68
12	70
15	56

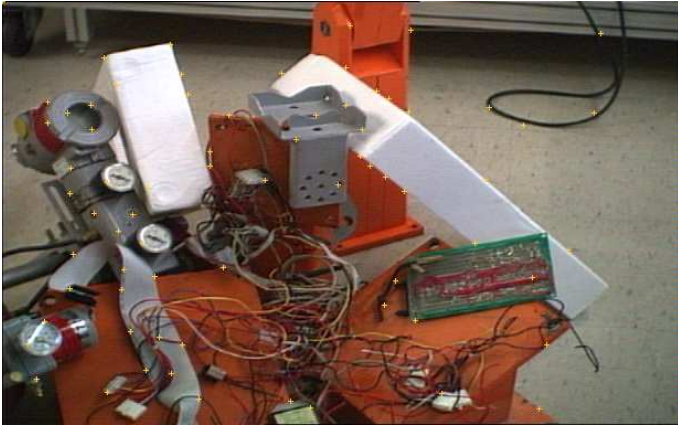
O algoritmo de pareamento de pontos aqui implementado não é capaz de fornecer precisão de sub-píxel para as coordenadas de localização dos pontos nas imagens. É possível explorar esta possibilidade através da estimação da correlação cruzada em posições deslocadas nos dois eixos e utilizando os resultados para a interpolação de uma função normal 2D. Maximizando a função interpolada obtém-se coordenadas em frações de píxeis (ZHANG et al., 1995).



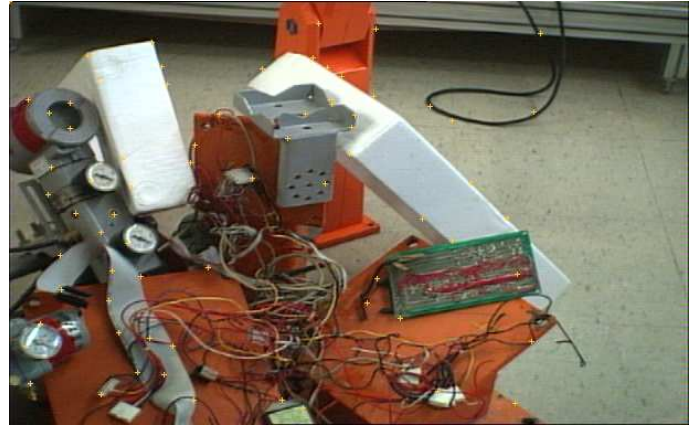
(a) Imagem esquerda, $w_\mu = 9$.



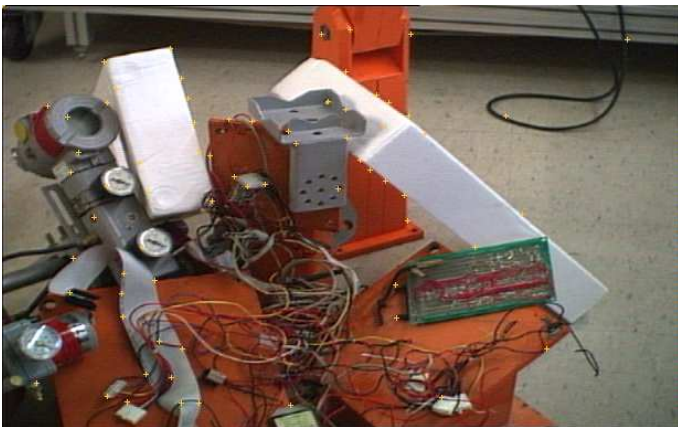
(b) Imagem direita, $w_\mu = 9$.



(c) Imagem esquerda, $w_\mu = 12$.



(d) Imagem direita, $w_\mu = 12$.

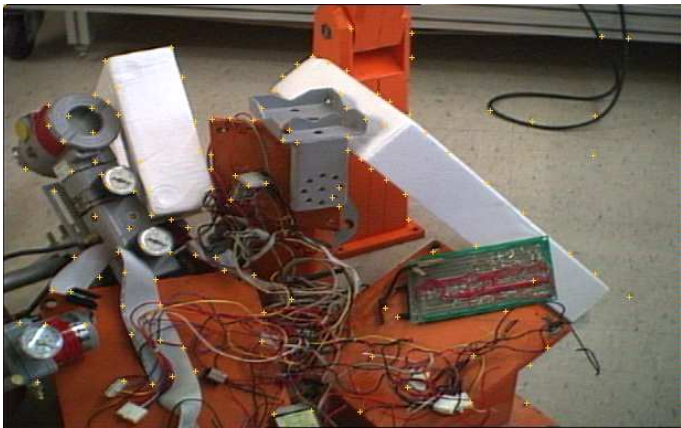


(e) Imagem esquerda, $w_\mu = 15$.

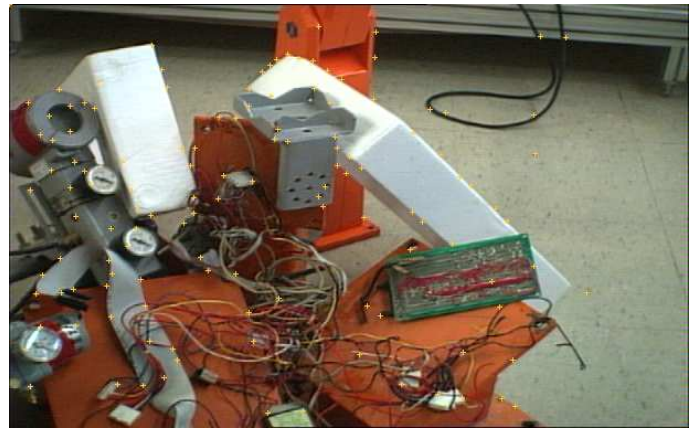


(f) Imagem esquerda, $w_\mu = 15$.

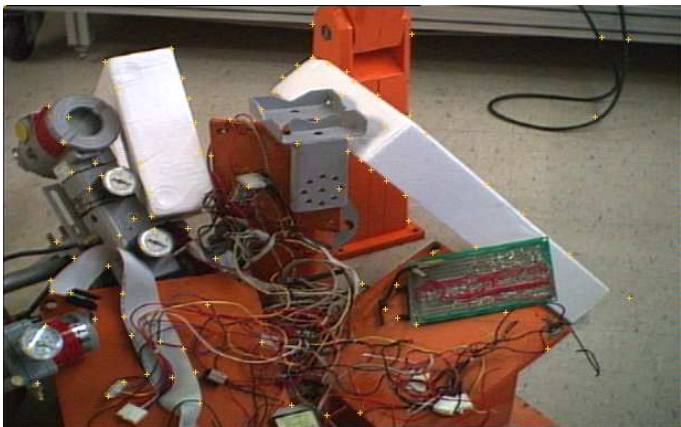
Figura 4.5: Efeito da variação do tamanho da janela de correlação cruzada no pareamento de pontos de duas imagens.



(a) Imagem esquerda, $l_\mu = 70$.



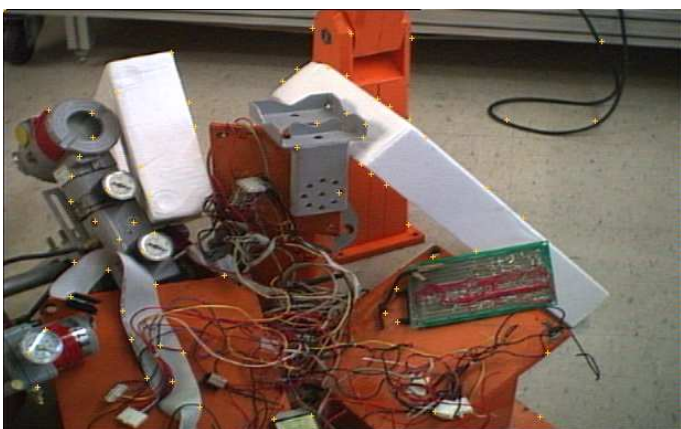
(b) Imagem direita, $l_\mu = 70$.



(c) Imagem esquerda, $l_\mu = 80$.



(d) Imagem direita, $l_\mu = 80$.

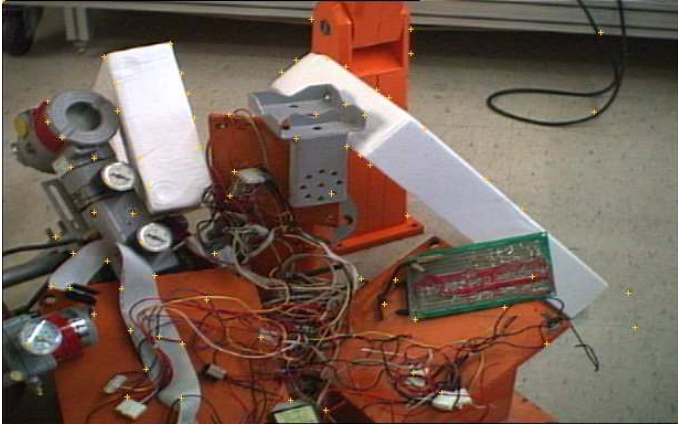


(e) Imagem esquerda, $l_\mu = 90$.

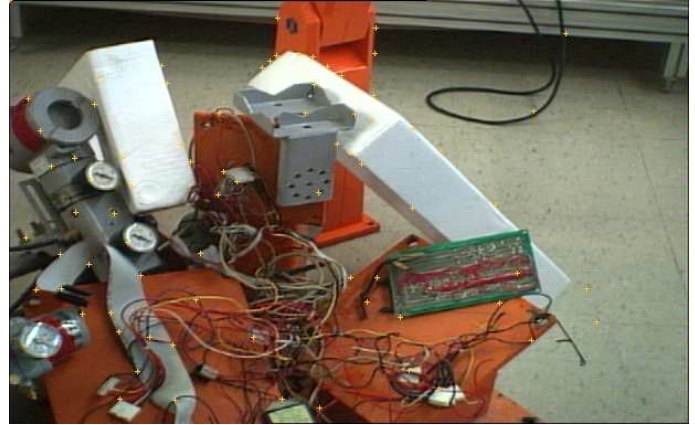


(f) Imagem esquerda, $l_\mu = 90$.

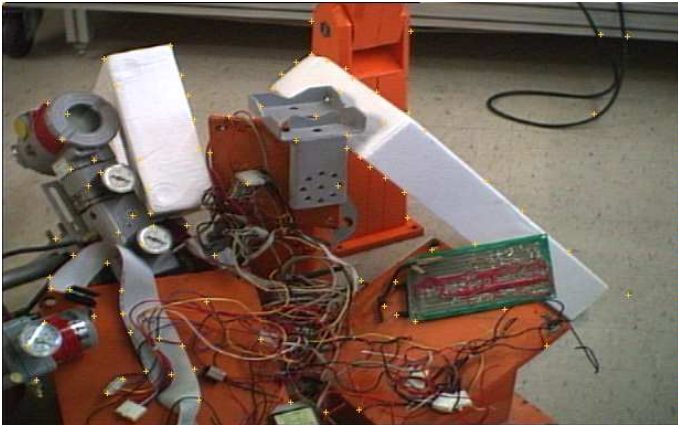
Figura 4.6: Efeito do valor limiar de correlação cruzada no pareamento de pontos de duas imagens.



(a) Imagem esquerda, $\varepsilon_r = 0,05$.



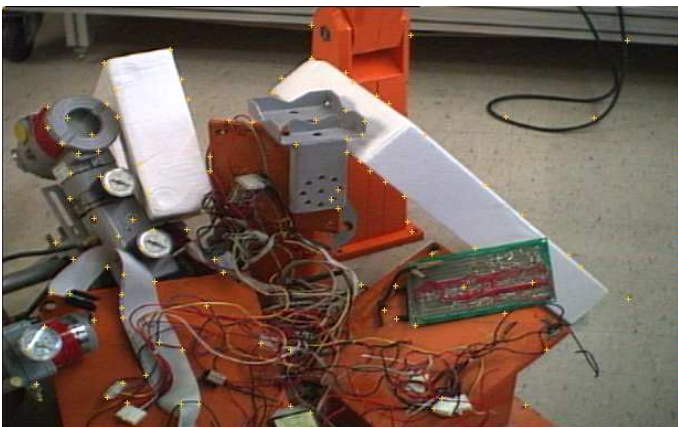
(b) Imagem direita, $\varepsilon_r = 0,05$.



(c) Imagem esquerda, $\varepsilon_r = 0,2$.



(d) Imagem direita, $\varepsilon_r = 0,2$.



(e) Imagem esquerda, $\varepsilon_r = 0,8$.



(f) Imagem esquerda, $\varepsilon_r = 0,8$.

Figura 4.7: Efeito do valor limiar de de força de um pareamento no casamento de pontos de duas imagens.

Tabela 4.4: Número de pontos pareados para diferentes valores de l_μ .

l_μ	Número de pares
70	137
80	89
90	60

Tabela 4.5: Número de pontos pareados para diferentes valores de ε_r .

ε_r	Número de pares
0,05	75
0,2	102
0,8	116

4.6 Reincidência

O processo de casamento descrito nas seções precedentes tem função de estabelecer coincidências que permitem a estimação da geometria epipolar. A geometria epipolar permite a reconstrução destes pontos num espaço projetivo. Entretanto, para estimar os parâmetros intrínsecos é preciso estimar-se a colineação (vide definição da seção A.2.2) que aplica o movimento no \mathbb{P}^3 , conforme será mostrado na seção 7.1. Para tornar isto possível é necessário também estabelecer correspondências entre os momentos antes e depois do movimento.

Por isso os pontos são casados para as quatro combinações de pares de imagens: entre imagens das câmeras da direita e da esquerda, obtidas no mesmo momento, e entre imagens de uma mesma câmera, obtidas antes e depois do movimento. A repetição de pontos entre os diferentes conjuntos de pares não é muito provável, entretanto é possível aumentar-se as chances disto ocorrer realizando pareamentos iterativamente.

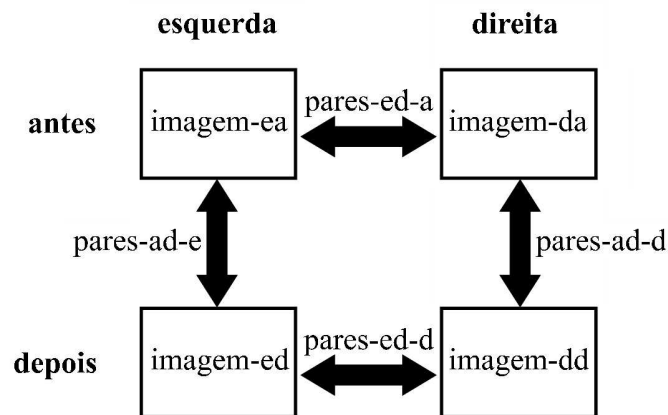


Figura 4.8: Dois pares de imagens, um antes e outro após o movimento.

O método consiste em refazer-se o processo de pareamento sucessivamente, substituindo os pontos de interesse de uma imagem por pontos desta imagem previamente pareados com outra imagem. O procedimento é detalhado a seguir (vide figura 4.8).

1. Primeiramente todos os pareamentos são realizados utilizando-se os respectivos

pontos de interesse de cada imagem, obtendo-se os conjuntos de pareamentos:

pares-ed-a conjunto de pareamentos obtido pareando os pontos de interesse de **imagem-ea** e de **imagem-da**;

pares-ed-d conjunto de pareamentos obtido pareando os pontos de interesse de **imagem-ed** e de **imagem-dd**;

pares-ad-e conjunto de pareamentos obtido pareando os pontos de interesse de **imagem-ea** e de **imagem-ed**;

pares-ad-d conjunto de pareamentos obtido pareando os pontos de interesse de **imagem-da** e de **imagem-dd**.

2. Em seguida cada pareamento é atualizado utilizando como pontos de interesse de uma das imagens as coordenadas de pontos pareados anteriormente em outro pareamento, obtendo-se novamente os conjuntos de pareamento:

pares-ed-a conjunto de pareamentos obtido pareando os pontos de **imagem-ea** pretendentes a **pares-ad-e** e os pontos de interesse de **imagem-da**;

pares-ed-a conjunto de pareamentos obtido pareando os pontos de interesse de **imagem-ea** e os pontos de **imagem-da** pretendentes a **pares-ad-d**;

pares-ed-d conjunto de pareamentos obtido pareando os pontos de **imagem-ed** pretendentes a **pares-ad-e** e os pontos de interesse de **imagem-dd**;

pares-ed-d conjunto de pareamentos obtido pareando os pontos de interesse de **imagem-ed** e os pontos de **imagem-dd** pretendentes a **pares-ad-d**;

pares-ad-e conjunto de pareamentos obtido pareando os pontos de **imagem-ea** pretendentes a **pares-ed-a** e os pontos de interesse de **imagem-ed**;

pares-ad-e conjunto de pareamentos obtido pareando os pontos de interesse de **imagem-ea** e os pontos de **imagem-ed** pretendentes a **pares-ed-d**;

pares-ad-d conjunto de pareamentos obtido pareando os pontos de **imagem-da** pretendentes a **pares-ed-a** e os pontos de interesse de **imagem-dd**;

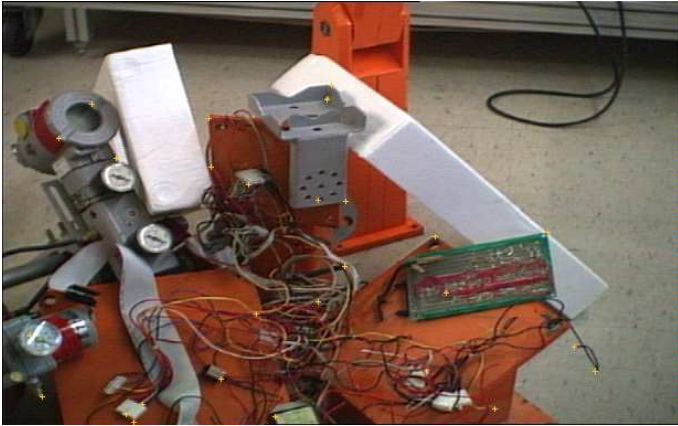
pares-ad-d conjunto de pareamentos obtido pareando os pontos de interesse de **imagem-ed** e os pontos de **imagem-dd** pretendentes a **pares-ed-d**.

3. Os pareamentos do item 2 são repetidos até que mais nenhum novo par seja encontrado.

Aqueles pontos que já estão pareados não são mais usados como pontos de interesse para novos pareamentos. A figura 4.9 mostra alguns pontos encontrados correspondentemente nas quatro imagens segundo o processo descrito acima.

Não há necessidade da correlação cruzada ser recalculada para cada pareamento que se repete entre duas imagens, pois quaisquer que sejam os pontos utilizados como pontos de interesse para o pareamento, eles estarão sempre contidos no conjunto de pontos de interesse originalmente detectados para aquelas imagens. Por isso as correlações cruzadas entre pontos de interesse de duas imagens são armazenadas e referenciadas a cada pareamento para calcular-se a força do pareamento.

Este método interativo é eficiente para forçar a localização de pontos correspondentes nas quatro imagens, sem a necessidade de um acompanhamento das coordenadas dos pontos durante o movimento.



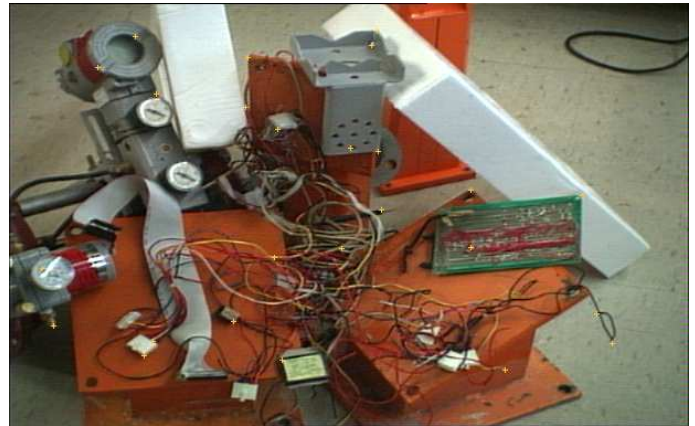
(a) **imagem-ea.**



(b) **imagem-da.**



(c) **imagem-ed.**



(d) **imagem-dd.**

Figura 4.9: Exemplo de pontos encontrados nas quatro imagens através da constatação das reincidências.

5 ESTIMAÇÃO DA GEOMETRIA EPIPOLAR

Dadas as coordenadas de pelo menos oito pontos correspondentes entre as duas imagens, é possível estimar-se linearmente a matriz fundamental de forma exata (HARTLEY, 1995). No entanto, geralmente tem-se muito mais do que oito correspondências, todas corrompidas por ruído e incertezas. Neste caso o algoritmo de estimação linear, apesar de simples e rápido, não é tão robusto (HARTLEY, 1995). Por outro lado, é possível estabelecer-se uma função de custo com um significado físico bem definido e assim realizar uma minimização não-linear (ZHANG, 1998). Entretanto esta minimização deve ser realizada com cuidado pois a função de custo tende a apresentar diversos mínimos locais devido aos efeitos do ruído e incertezas nas coordenadas dos pontos.

A minimização linear, baseada em mínimos quadrados, serve para de forma rápida efetuar repetidas vezes o cálculo da matriz F para diferentes amostras, enquanto que a minimização não-linear, inicializada com a melhor F estimada até então, apresenta a vantagem de minimizar um custo de significado físico bem definido, permitindo assim um ajuste fino da F estimada.

O método de minimização multi-variável de *Powell* (PRESS et al., 1992) foi utilizado para estimar a matriz fundamental F minimizando este custo de significado físico definido. Para evitar mínimos locais é necessário realizar uma busca por uma boa estimativa inicial. Esta busca é realizada através da repetição iterativa da estimação linear da matriz fundamental. A matriz é estimada linearmente numerosas vezes, cada vez com um conjunto diferente de pontos. Dentre muitas amostras de conjuntos de $n \geq 8$ pares de pontos cada, escolhe-se o conjunto para o qual a minimização linear (vide seção 5.1) retorna o menor valor para um critério de significado físico, dado por (5.5). Os pares de pontos são amostrados utilizando-se as chamadas técnicas de *bucketing* (ZHANG et al., 1995) que asseguram uma boa distribuição da localização destes pontos por toda a imagem. Aquele conjunto de pontos através do qual melhor se estimou a matriz fundamental é tido como um bom conjunto de pontos e esta matriz é então utilizada como estimativa inicial para a minimização não linear. Para um melhor desempenho do algoritmo de minimização não-linear eliminam-se aqueles pares que já não estão de acordo com a geometria epipolar linearmente estimada. Pares que retornam um valor maior do que 2,5 desvios padrões para o custo (5.5) são descartados como falsos pares. Só então, a matriz F inicial é finalmente mapeada para uma minimização em 7 direções utilizando o algoritmo de *Powell* (vide seção 5.2).

5.1 Estimação linear

Sendo $\mathbf{m} = [u \ v \ 1]^T$ e $\mathbf{m}' = [u' \ v' \ 1]^T$ os vetores de coordenadas que descrevem dois pontos em duas imagens distintas, e sendo f_{ij} os elementos que compõem a matriz

\mathbf{F} pode-se reescrever (2.33) na forma vetorial $\mathbf{u}^T \mathbf{f} = 0$, onde

$$\mathbf{u} = [uu' \quad vu' \quad u' \quad uv' \quad vv' \quad v' \quad u \quad v \quad 1]^T \quad (5.1)$$

$$\mathbf{f} = [f_{11} \quad f_{12} \quad f_{13} \quad f_{21} \quad f_{22} \quad f_{23} \quad f_{31} \quad f_{32} \quad f_{33}]^T \quad (5.2)$$

Dados $n \geq 8$ pares correspondentes, é possível fazer-se $\mathbf{U} = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_n]$ e resolver-se $\mathbf{U}^T \mathbf{f} = \mathbf{0}_9$ linearmente por mínimos quadrados, onde procura-se o vetor \mathbf{f} tal que $\mathbf{f}^T \mathbf{U} \mathbf{U}^T \mathbf{f} = 0$. A solução é o auto-vetor associado ao menor auto-valor de $\mathbf{U} \mathbf{U}^T$.

O posto da \mathbf{F} resultante deveria ser igual a 2, mas dados corrompidos por ruído são inevitáveis e portanto o posto da matriz fundamental estimada dificilmente será singular. Para impor esta restrição sobre o posto de \mathbf{F} uma decomposição por valores singulares é realizada fazendo $\mathbf{F} = \mathbf{Q} \mathbf{R} \mathbf{S}^T$, onde \mathbf{R} é diagonal e \mathbf{Q} e \mathbf{S} são ortogonais. O menor elemento diagonal de \mathbf{R} é substituído por zero e \mathbf{F} é reconstruída, agora com posto igual a 2.

Para evitar erros de tendências na localização dos epipólos, antes de efetuar o processo de minimização linear acima descrito, as coordenadas dos pontos devem ser normalizadas (ZHANG et al., 1995). Isto se faz através da imposição de transformações $\hat{\mathbf{m}} = \mathbf{T} \mathbf{m}$ e $\hat{\mathbf{m}}' = \mathbf{T}' \mathbf{m}'$ que movem o centróide dos pontos para a origem e escalonam suas coordenadas para que a distância média até a origem seja igual a $\sqrt{2}$ (HARTLEY, 1995). Este tipo de transformação é dado por

$$\mathbf{T} = \begin{bmatrix} 1/s & 0 & -\bar{u}/s \\ 0 & 1/s & -\bar{v}/s \\ 0 & 0 & 1 \end{bmatrix} \quad (5.3)$$

onde $s = \sum_i \sqrt{u_i^2 + v_i^2}$ é uma mudança de escala e \bar{u} e \bar{v} são respectivamente as médias de u_i e v_i para os pontos $\mathbf{m}_i = [u_i \quad v_i \quad 1]^T$.

Desta forma a minimização não retorna a matriz fundamental correta, mas sim uma matriz fundamental $\hat{\mathbf{F}}$ satisfazendo a restrição epipolar para os pontos transformados. Depois de concluída a minimização, a matriz fundamental original é recuperada por:

$$\mathbf{F} = \mathbf{T}'^T \hat{\mathbf{F}} \mathbf{T} \quad (5.4)$$

5.2 Estimação não-linear

A idéia da estimação não-linear é minimizar uma quantidade que possua um significado físico. Esta quantidade é a soma do quadrado das distâncias entre os pontos e suas respectivas retas epipolares medidas no plano da imagem. O critério minimizado é dado por

$$J = \frac{1}{N} \sum_{i=0}^N (d^2(\mathbf{m}', \mathbf{F} \mathbf{m}) + d^2(\mathbf{m}, \mathbf{F}^T \mathbf{m}')) \quad (5.5)$$

onde $d(\cdot, \cdot)$ representa a distância entre um ponto e uma reta, calculada como sendo

$$d(\mathbf{u}, \mathbf{v}) = \frac{\mathbf{u}^T \mathbf{v}}{\sqrt{v_1^2 + v_2^2}} \quad (5.6)$$

com $\mathbf{v} = [v_1 \quad v_2 \quad v_3]^T$.

Já que \mathbf{F} tem 8 elementos mas apenas 7 graus de liberdade, a minimização não-linear deve realizada num espaço parametrizado (ZHANG, 1998). Para isto a matriz fundamental é decomposta de forma que uma de suas colunas e uma de suas linhas sejam expressas

como combinação linear das outras duas: i_0 e j_0 . A parametrização é formada pelas coordenadas dos epipólos (x, y, x', y') , e quatro elementos (a, b, c, d) representando a homografia entre os dois lápis de retas epipolares¹. É fácil observar que isto pode resultar em 9 mapas diferentes, cada um correspondendo respectivamente à coluna e à linha escolhidas. Por exemplo, se a linha e coluna escolhidas são $i_0 = 3$ e $j_0 = 3$, então:

$$\mathbf{F} = \begin{bmatrix} a & b & -ax - by \\ c & d & -cx - dy \\ -ax' - cy' & -bx' - dy' & (ax + by)x' + (cx + dy)y' \end{bmatrix} \quad (5.7)$$

Dada uma matriz \mathbf{F} e seus epipólos, ou uma aproximação disto, precisa-se escolher, dentre diferentes mapeamentos de parametrização, o mais adequado. Denotando $\mathbf{f}_{i_0 j_0}$ como sendo o vetor dos elementos de \mathbf{F} decompostos da mesma forma que em (5.2), i_0 e j_0 são escolhidos de forma a maximizar o posto da matriz jacobiano 9×8 :

$$\mathbf{J} = \frac{d\mathbf{f}_{i_0 j_0}}{d\mathbf{p}} \quad \text{onde} \quad \mathbf{p} = [x \ y \ x' \ y' \ a \ b \ c \ d]^T \quad (5.8)$$

Isto é feito maximizando-se a norma do vetor cujas coordenadas são os determinantes das nove submatrizes 8×8 de \mathbf{J} . A literatura (ZHANG, 1998) mostra que esta norma é dada pela expressão $(ad - bc)^2 \sqrt{x^2 + y^2 + 1} \sqrt{x'^2 + y'^2}$.

5.3 Resultados para dados simulados

Para avaliar o funcionamento dos métodos de estimação em suas várias etapas é interessante saber com exatidão os parâmetros que estão sendo estimados a fim de avaliar o comportamento do erro de estimação em função do ruído. Com este intuito desenvolveu-se um algoritmo gera coordenadas tridimensionais de pontos aleatoriamente distribuídos dentro de um prisma retangular, com as coordenadas geradas segundo uma distribuição uniforme. Estes pontos são então projetados nos planos das imagens através das matrizes de projeção de duas câmeras simuladas com parâmetros arbitrados e conhecidos. Estas coordenadas projetadas são então alimentadas no algoritmo e os vários parâmetros são estimados.

A fim de avaliar o comportamento do critério (5.5) as coordenadas dos pontos na imagem foram ambas perturbadas por um ruído seguindo a distribuição uniforme cuja função densidade probabilidade é:

$$f(x) = \frac{1}{2A} \quad \text{para} \quad -A \leq x \leq A \quad (5.9)$$

Para a estimação da geometria epipolar obtiveram-se os resultados mostrados na tabela 5.1. Resultados $J < 1$ para em (5.5) demonstram erro de frações de píxel para o significado físico do critério.

5.4 Resultados para dados reais

Avaliou-se o funcionamento do método de estimação da matriz fundamental em imagens reais através da utilização dos pares de pontos obtidos do processo de pareamento

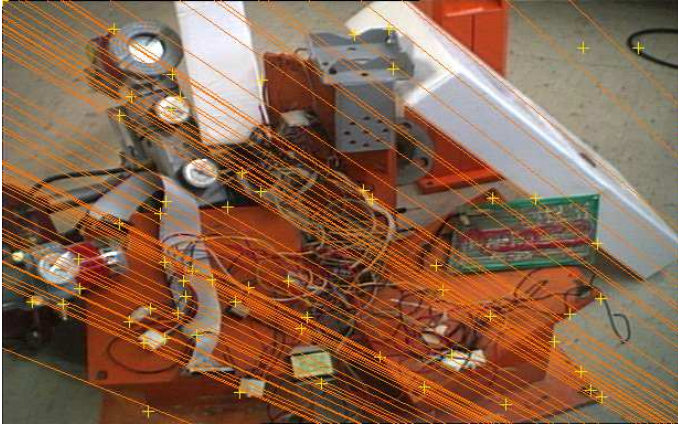
¹Um lápis de retas é um espaço projetivo unidimensional formado por todas as retas que se cruzam num determinado ponto (FAUGERAS, 1993).

Tabela 5.1: Minimização do critério da expressão (5.5) para diferentes valores de A em (5.9).

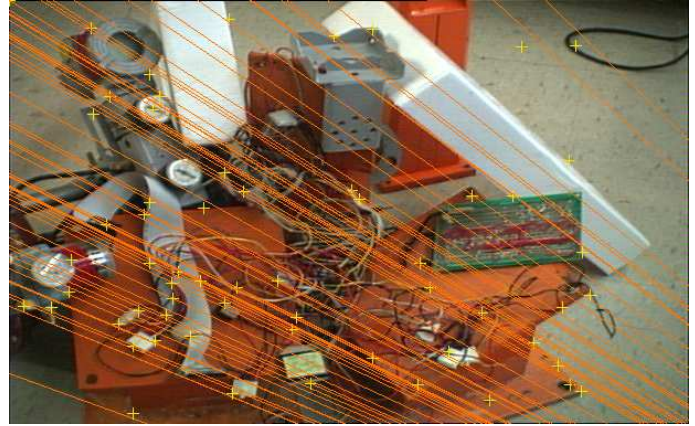
A	Média do critério minimizado
10^{-5}	$5,27 \times 10^{-8}$
10^{-4}	$4,50 \times 10^{-7}$
10^{-3}	$3,48 \times 10^{-5}$
10^{-2}	$3,23 \times 10^{-3}$
10^{-1}	$3,01 \times 10^{-1}$

descrito no capítulo 4. As coordenadas dos pares de pontos foram alimentadas no algoritmo e os resultados foram observados nas imagens e no critério (5.5).

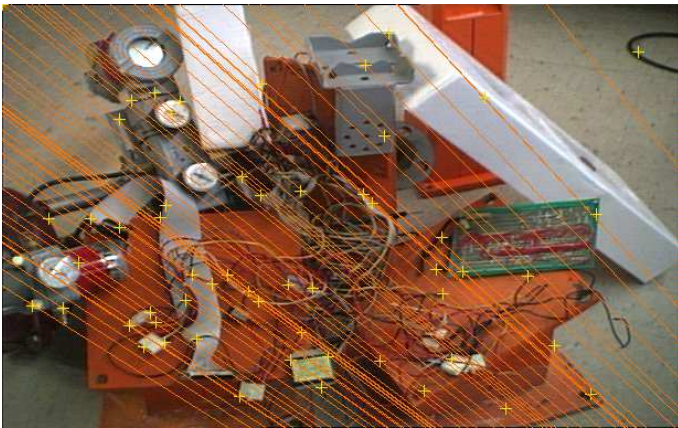
A figura 5.1 mostra dois pares de imagens onde são plotados os pontos pareados e suas respectivas retas epipolares calculadas a partir da matriz fundamental. O par de imagens das figuras 5.1(a) e 5.1(b) mostra os resultados para uma \mathbf{F} estimada utilizando apenas o método linear da seção 5.1. Para este par, o critério da expressão (5.5) minimizou em $J = 3,546 \times 10^2$. No par de imagens das figuras 5.1(c) e 5.1(d) mostra-se os resultados para uma \mathbf{F} estimada utilizando o método linear seguido da minimização não-linear conforme descrito na seção 5.2. Neste par, o critério da expressão (5.5) minimizou em $J = 2,456 \times 10^2$.



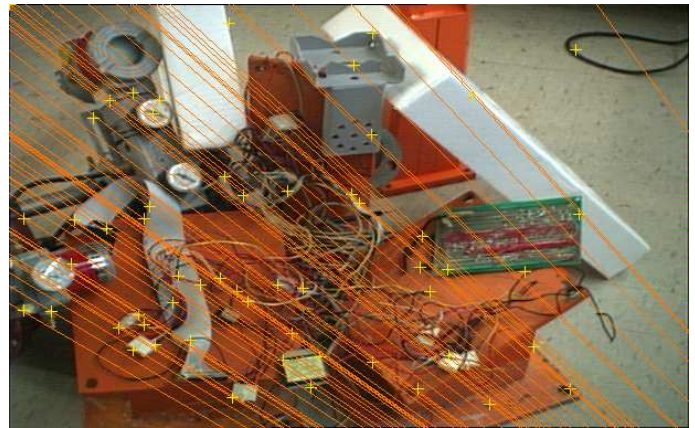
(a) Imagem esquerda, \mathbf{F} estimada linearmente.



(b) Imagem direita, \mathbf{F} estimada linearmente.



(c) Imagem esquerda, \mathbf{F} estimada com *Powell*.



(d) Imagem direita, \mathbf{F} estimada com *Powell*.

Figura 5.1: Geometria epipolar recuperada através da estimação da matriz fundamental.

6 RECONSTRUÇÃO NO ESPAÇO PROJETIVO

Sejam duas matrizes de projeção \hat{P} e \hat{P}' e um par de pontos correspondentes \mathbf{m} e \mathbf{m}' no plano da imagem de duas câmeras. Com elas é possível reconstruir-se o ponto $\tilde{\mathbf{m}}$ do \mathbb{P}^3 que está na intersecção dos dois raios de projeção no espaço, que são duas retas que partem uma de cada ponto, e passam cada uma respectivamente no centro ótico de cada câmera.

Através da matriz fundamental que descreve a geometria epipolar de um par de câmeras é possível determinar este par de matrizes de projeção numa base arbitrária do \mathbb{P}^3 . Dada uma determinada matriz fundamental, existe uma família de pares de matrizes de projeção em diferentes bases projetivas que satisfazem esta geometria epipolar. Indiferente da escolha da base projetiva, uma vez determinadas as matrizes de projeção, pode-se reconstruir cada ponto neste espaço projetivo. Com estes pontos reconstruídos no \mathbb{P}^3 é possível estimar a colineação que aplica o movimento do sistema de visão neste espaço e assim estimar os parâmetros intrínsecos.

O método aqui utilizado para encontrar uma base projetiva em particular consiste em procurar uma matriz \mathbf{M} para a qual $\mathbf{F} = \mathbf{e}' \times \mathbf{M}$ e então definir-se $\hat{P} = [\mathbf{I} \ 0]$ e $\hat{P}' = [\mathbf{M} \ \mathbf{e}']$, sendo $\mathbf{e}' = [x' \ y' \ 1]^T$ o vetor de coordenadas do epipólo, que já é conhecido pelo mapeamento da matriz fundamental descrito na seção 5.2. Existem diferentes soluções para este problema, cada uma correspondendo a uma diferente base do \mathbb{P}^3 . Neste trabalho a matriz \mathbf{M} é encontrada como sendo (ZHANG; FAUGERAS; DERICHE, 1997):

$$\mathbf{M} = \frac{-\mathbf{e}' \times \mathbf{F}}{\|\mathbf{e}'\|} \quad (6.1)$$

6.1 Triangulação

Devido ao ruído, pares de pontos podem não satisfazer com exatidão a geometria epipolar e portanto na reconstrução destes pontos os dois raios geralmente não vão se interceptar no espaço, conforme a figura 6.1.

Encontrar o ponto médio da perpendicular comum entre os dois raios não é adequado para a reconstrução projetiva, já que conceitos como distância e perpendicularidade não são válidos no contexto de geometria projetiva (HARTLEY; STURM, 1997). Neste trabalho utilizou-se o método *polinomial* que é a solução ótima para o problema de triangulação no espaço projetivo (HARTLEY; STURM, 1997). Neste método minimiza-se uma função de custo baseada em distâncias medidas no plano retinal.

Procuram-se dois pontos da imagem, $\hat{\mathbf{m}}$ e $\hat{\mathbf{m}}'$, que satisfazem com exatidão (2.33) e minimizam o critério

$$J = d^2(\mathbf{m}, \hat{\mathbf{m}}) + d^2(\mathbf{m}', \hat{\mathbf{m}}') \quad (6.2)$$

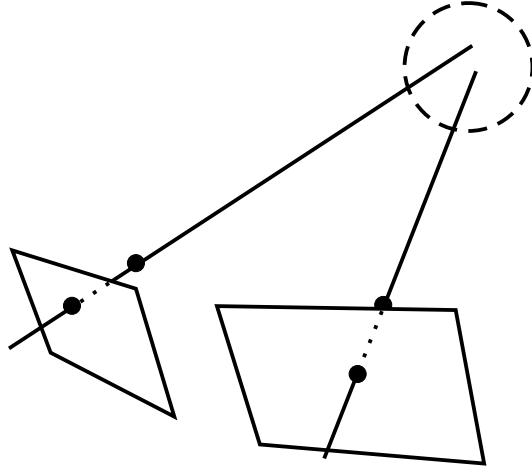


Figura 6.1: Inconsistências na geometria epipolar: raios podem não se cruzar.

onde \mathbf{m} e \mathbf{m}' são pontos pareados com ruído, e $d(\cdot, \cdot)$ representa a distância entre eles. Como qualquer par de pontos satisfazendo (2.33) cai sobre um respectivo par de retas epipolares, γ e γ' , é possível reescrever o critério (6.2) como sendo

$$J = d^2(\mathbf{m}, \gamma) + d^2(\mathbf{m}', \gamma') \quad (6.3)$$

onde $d(\cdot, \cdot)$ é agora a distância entre um ponto e uma reta.

Para solucionar este problema, primeiramente ambos os lápis de retas epipolares são parametrizados em função de t , respectivamente $\gamma(t)$ e $\gamma'(t)$. Assim, (6.3) transforma-se em um polinômio em t . Uma vez encontrado o t que minimiza este polinômio, $\hat{\mathbf{m}}$ e $\hat{\mathbf{m}}'$ são encontrados como sendo as projeções ortogonais de \mathbf{m} e \mathbf{m}' respectivamente em $\gamma(t)$ e $\gamma'(t)$.

Para facilitar o processo, a matriz fundamental é escrita como

$$\hat{\mathbf{F}} = \mathbf{Q}'\mathbf{L}'^{-T}\mathbf{F}\mathbf{L}^{-1}\mathbf{Q}^T \quad (6.4)$$

sendo

$$\mathbf{Q} = \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{e} \quad \mathbf{L} = \begin{bmatrix} 1 & 0 & -u \\ 0 & 1 & -v \\ 0 & 0 & 1 \end{bmatrix} \quad (6.5)$$

onde $\mathbf{m} = [u \ v \ 1]^T$, $\mathbf{e} = [x \ y \ 1]^T$ e $\theta = -\tan^{-1} \frac{y-v}{x-u}$ e respectivas definições para \mathbf{Q}' e \mathbf{L}' . Estas operações não mudam o valor do custo (6.3) mas transformam \mathbf{F} na forma:

$$\hat{\mathbf{F}} = \begin{bmatrix} ff'd & -f'c & -f'd \\ -fb & a & b \\ -fd & c & d \end{bmatrix} \quad (6.6)$$

As operações acima descritas são transformações rígidas e têm a função de mover os pontos para a origem do sistema de coordenadas, ou seja $\mathbf{m} = \mathbf{m}' = [0 \ 0 \ 1]^T$. Além disto, a rotação \mathbf{Q} move os epipólos para $[1 \ 0 \ f]^T$ e $[1 \ 0 \ f']^T$, alterando a matriz \mathbf{F} de forma correspondente.

Procurar uma reta epipolar próxima aos pontos traduz-se em procurar por uma reta epipolar passando o mais próximo possível da origem. Supõe-se, sem perda de generalidade, que esta reta passa sobre o ponto $[0 \ t \ 1]^T$ próximo da origem. Estas retas são

então dadas por $[0 \ t \ 1]^T \times [1 \ 0 \ f]^T$ ou seja $\gamma(t) = [tf \ 1 \ -t]^T$ e conseqüentemente $\gamma'(t) = [-f'(ct + d) \ at + b \ ct + d]^T$. Com isto é possível reescrever como (6.3):

$$s(t) = \frac{2t}{(1 + (tf)^2)^2} - \frac{2(ad - bc)(ac + b)(ct + d)}{((at + b)^2 + f'^2(ct + d)^2)^2} \quad (6.7)$$

Uma vez encontrado t que minimiza $s(t)$, é possível computar $\hat{\mathbf{m}}$ e $\hat{\mathbf{m}}'$ como sendo respectivamente os pontos de intersecção entre as retas perpendiculares a $\gamma(t)$ e $\gamma'(t)$, passando por \mathbf{m} e \mathbf{m}' . A partir daí é possível reconstruir $\tilde{\mathbf{m}}$ neste espaço projetivo, como é mostrado na seção 6.2.

Para isto desenvolveu-se a derivada de (6.7) e utilizou-se técnicas numéricas para obtenção das raízes (PRESS et al., 1992).

A figura 6.2 mostra dois pares de imagens para duas cenas diferentes. O par de imagens das figuras 6.2(a) e 6.2(b) foi capturado antes do movimento, e o par de imagens das figuras 6.2(c) e 6.2(d) foi capturado após o movimento. Destacam-se os pontos pareados em comum em todas as quatro imagens. As geometrias epipolares mostradas foram computadas a partir dos pontos pareados entre as imagens das câmeras da direita e da esquerda. Para esta estimação utilizam-se todos os pares obtidos entre as imagens das câmeras, tanto antes quanto depois do movimento. Cada reta traçada numa imagem corresponde a todas as possíveis localizações de um ponto da outra imagem. As retas traçadas são aquelas correspondentes aos pontos pareados. As coordenadas destes pontos foram retificadas para satisfazerem com exatidão a geometria epipolar computada segundo o método descrito na seção 6.1.

6.2 Reconstrução

Denotando $\hat{\mathbf{p}}_i$ e $\hat{\mathbf{p}}'_i$ como sendo as i -ésimas linhas de $\hat{\mathbf{P}}$ e $\hat{\mathbf{P}}'$, respectivamente, e isolando os fatores de escala em

$$\hat{\mathbf{m}} = \lambda \hat{\mathbf{P}} \tilde{\mathbf{m}} \quad (6.8)$$

$$\hat{\mathbf{m}}' = \lambda' \hat{\mathbf{P}}' \tilde{\mathbf{m}} \quad (6.9)$$

para serem $\lambda = \hat{\mathbf{p}}_3^T \tilde{\mathbf{m}}$ e $\lambda' = \hat{\mathbf{p}}_3'^T \tilde{\mathbf{m}}$ é possível, considerando $\hat{\mathbf{m}} = [\hat{u} \ \hat{v} \ 1]$ e $\hat{\mathbf{m}}' = [\hat{u}' \ \hat{v}' \ 1]$, escrever $\mathbf{A} \tilde{\mathbf{m}} = \mathbf{0}_4$ onde

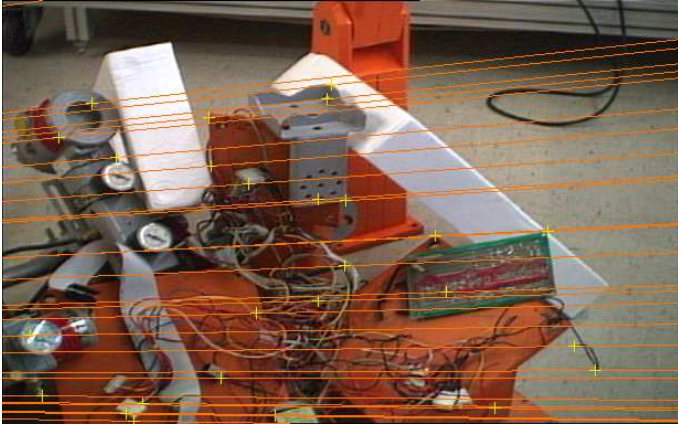
$$\mathbf{A} = \begin{bmatrix} \hat{\mathbf{p}}_1^T - \hat{u} \hat{\mathbf{p}}_3^T \\ \hat{\mathbf{p}}_2^T - \hat{v} \hat{\mathbf{p}}_3^T \\ \hat{\mathbf{p}}_1'^T - \hat{u}' \hat{\mathbf{p}}_3'^T \\ \hat{\mathbf{p}}_2'^T - \hat{v}' \hat{\mathbf{p}}_3'^T \end{bmatrix} \quad (6.10)$$

Então $\tilde{\mathbf{m}}$ é o auto-vetor associado ao menor auto-valor de $\mathbf{A}^T \mathbf{A}$.

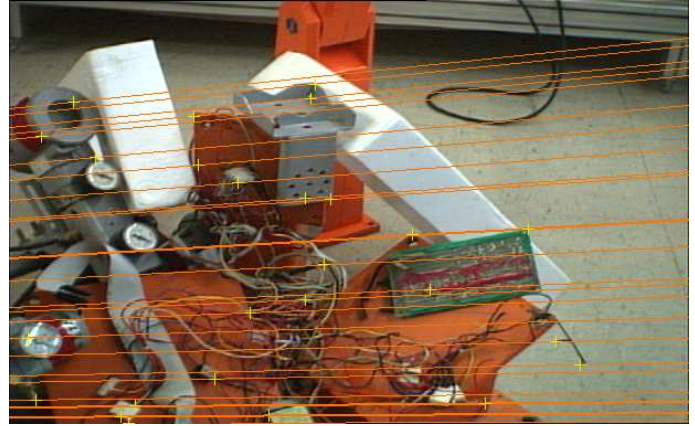
Note que $\hat{\mathbf{P}}$ e $\hat{\mathbf{P}}'$ são matrizes que representam a projeção numa base projetiva que em geral não corresponderá ao espaço Euclidiano. Para um espaço projetivo qualquer nem todas as propriedades da estrutura 3D original do \mathbb{R}^3 são preservadas¹.

Na figura 6.3(c), gerada utilizando o software *Geomview* (PHILIPS; LEVY; MUNZNER, 1993) mostra-se a reconstrução projetiva calculada a partir dos pontos das imagens que estão nas figuras 6.3(a) e 6.3(b). O prisma retangular tridimensional representado na figura 6.3(c) tem suas arestas alinhadas com os eixos do sistema de coordenadas e

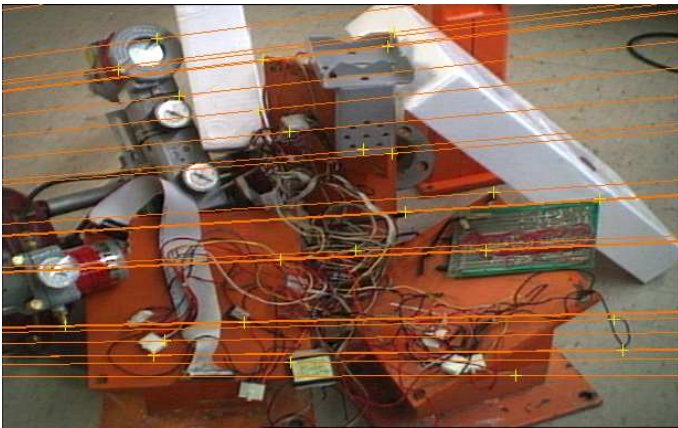
¹Detalhes sobre propriedades preservadas no espaço projetivo são revisados no apêndice A.



(a) **imagem-ea.**



(b) **imagem-da.**



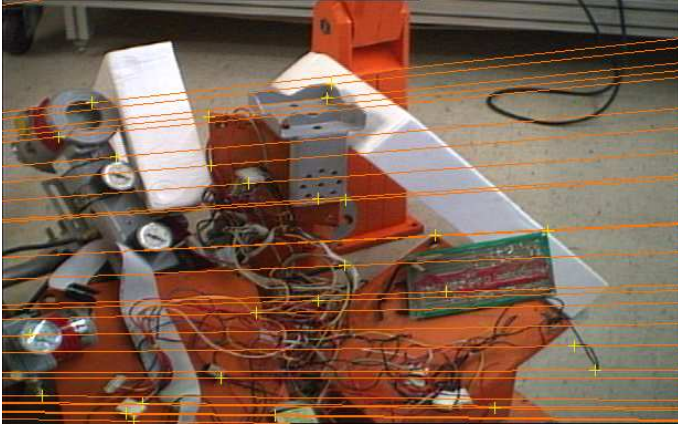
(c) **imagem-ed.**



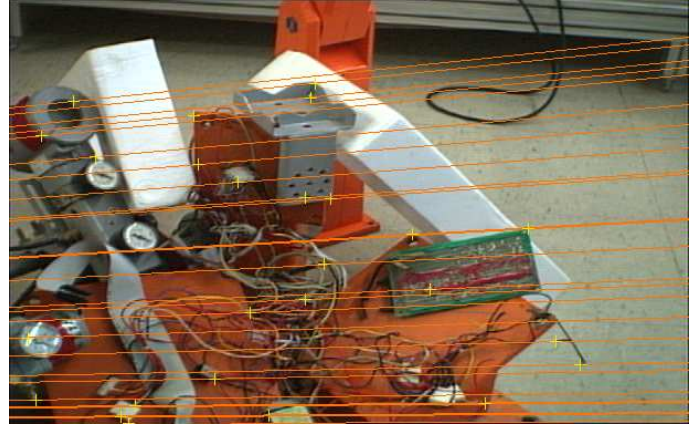
(d) **imagem-dd.**

Figura 6.2: Pontos retificados para caírem com exatidão sobre suas retas epipolares.

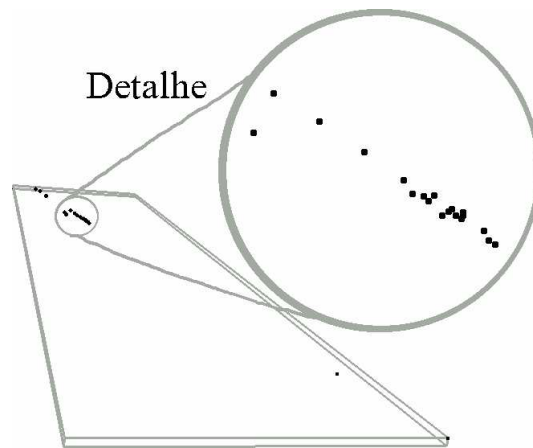
seus lados representam os limites dentro de onde se encontram os pontos reconstruídos. Observa-se a deformação na distribuição dos pontos reconstruídos neste espaço, que em nada relembra a distribuição observada nas imagens. Apesar da aparente irregularidade estes pontos representam de forma fiel uma cena projetiva tridimensional. Cada um destes pontos do \mathbb{P}^3 , se projetados pelas expressões (6.8) e (6.9), gera os pontos das imagens originais.



(a) Imagem esquerda.



(b) Imagem direita.



(c) Reconstrução do \mathbb{P}^3 .

Figura 6.3: Reconstrução numa base projetiva arbitrária.

7 MOVIMENTO PROJETIVO

Sejam A_P e B_P dois conjuntos de pontos reconstruídos no \mathbb{P}^3 , como mostra a figura 7.1, ambos representando um mesmo grupo de pontos do \mathbb{R}^3 , A_E antes do movimento e B_E depois do movimento do sistema de visão. Existe uma matriz H_{AB} 4×4 que transforma as coordenadas de cada ponto \tilde{m}_A da reconstrução A_P , antes do movimento, em suas respectivas coordenadas \tilde{m}_B da reconstrução B_P , após o movimento:

$$\tilde{m}_B \simeq H_{AB} \tilde{m}_A \quad (7.1)$$

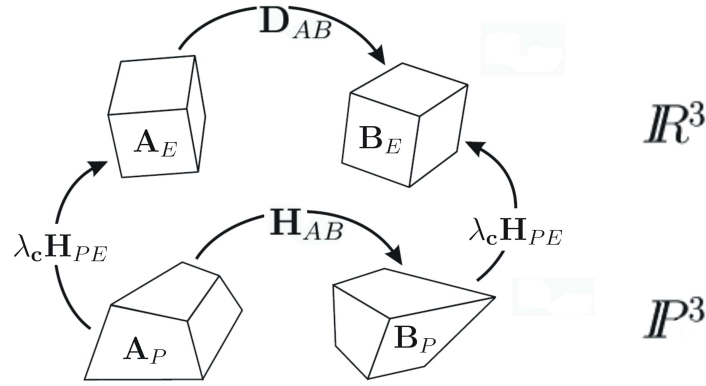


Figura 7.1: Deslocamento no espaço Euclidiano e correspondente movimento no espaço projetivo.

Esta matriz H_{AB} é portanto uma *colineação*, ou seja, uma transformação no \mathbb{P}^3 . Multiplicar o vetor de coordenadas de um ponto do \mathbb{P}^3 por H_{AB} é a operação que descreve o movimento realizado pelo sistema de visão no \mathbb{P}^3 . Esta matriz do movimento projetivo relaciona-se com o deslocamento euclidiano D_{AB} pela expressão

$$H_{AB} = \lambda_c H_{PE}^{-1} D_{AB} H_{PE} \quad (7.2)$$

onde H_{PE} é outra colineação que transforma coordenadas do \mathbb{P}^3 para coordenadas do \mathbb{R}^3 , e λ_c é um fator de escala. Percebe-se que H_{PE} representa uma mudança de base projetiva. As coordenadas dos pontos no espaço Euclidiano podem ser encontradas como sendo $\tilde{n}_A = H_{PE} \tilde{m}_A$ e $\tilde{n}_B = H_{PE} \tilde{m}_B$ onde $\tilde{n}_B = D_{AB} \tilde{n}_A$.

7.1 Estimação do movimento projetivo

O processo de estimação da matriz H_{AB} se assemelha bastante ao processo de estimação da matriz fundamental, assunto abordado no capítulo 5. Aqui mais uma vez surge a

maioria dos problemas já comentados naquele capítulo: é possível estimar-se esta colineação linearmente e de forma exata, mas não de forma robusta, enquanto que a estimação não-linear está sujeita a falsos mínimos, mas permite a minimização de uma quantidade de significado físico bem definido CSURKA; DEMIRDJIAN; HORAUD (1999).

Fazendo uso da mesma idéia adotada na estimação da matriz fundamental, aplicou-se uma técnica baseada em estatísticas onde estima-se a colineação linearmente numerosas vezes de forma interativa, cada vez com um diferente conjunto de $n \geq 5$ pares de pontos 3D correspondentes. Escolhe-se, dentre muitos grupos de pares de pontos 3D, aquele grupo que melhor estima \mathbf{H}_{AB} segundo um determinado critério medido no plano da imagem. Esta colineação estimada linearmente será utilizada como estimativa inicial para a minimização não-linear. Os pares não conformes com a estimativa inicial de \mathbf{H}_{AB} são descartados e o algoritmo de minimização *Powell* é utilizado com os pares restantes para refinação do resultado. Novamente, antes das minimizações, as coordenadas de todos os pontos são normalizadas. Este método de estimação de colineações é detalhado em (CSURKA; DEMIRDJIAN; HORAUD, 1999).

7.1.1 Estimação linear

Para um determinado par de pontos reconstruídos $\tilde{\mathbf{m}}_A$ e $\tilde{\mathbf{m}}_B = [x \ y \ z \ w]$ procura-se \mathbf{H}_{AB} tal que

$$\lambda_H \tilde{\mathbf{m}}_B = \mathbf{H}_{AB} \tilde{\mathbf{m}}_A \quad (7.3)$$

A maneira clássica de estimar esta matriz é eliminando o fator de escala λ_H de (7.3). Desta forma encontra-se

$$wa - xd = 0 \quad (7.4)$$

$$wb - yd = 0 \quad (7.5)$$

$$wc - zd = 0 \quad (7.6)$$

onde considera-se $\mathbf{H}_{AB} \tilde{\mathbf{m}}_A = [a \ b \ c \ d]^T$. Normalmente, se $\tilde{\mathbf{m}}_B$ não está no infinito, pode-se fazer $w = 1$ dividindo-se os elementos do vetor $\tilde{\mathbf{m}}_B$ por sua quarta coordenada. Considerando \mathbf{H}_{AB} numa forma vetorial, $\mathbf{H}_{AB} = [h_{11} \ h_{12} \ \dots \ h_{44}]^T$, pode-se reescrever (7.3) como sendo $\mathbf{B} \mathbf{H}_{AB} = \mathbf{0}_{16}$, onde

$$\mathbf{B} = \begin{bmatrix} w\tilde{\mathbf{m}}_A^T & \mathbf{0}_4^T & \mathbf{0}_4^T & x\tilde{\mathbf{m}}_A^T \\ \mathbf{0}_4^T & w\tilde{\mathbf{m}}_A^T & \mathbf{0}_4^T & y\tilde{\mathbf{m}}_A^T \\ \mathbf{0}_4^T & \mathbf{0}_4^T & w\tilde{\mathbf{m}}_A^T & z\tilde{\mathbf{m}}_A^T \end{bmatrix} \quad (7.7)$$

Para $n \geq 5$ pares de pontos reconstruídos, é possível encontrar-se a colineação computando-se \mathbf{H}_{AB} , que é o auto-vetor associado ao menor auto-valor de $\sum_{i=1}^n \mathbf{B}_i^T \mathbf{B}_i$.

7.1.2 Estimação não-linear

A colineação \mathbf{H}_{AB} linearmente estimada é então utilizada como uma estimativa inicial para a minimização *Powell* em 7 direções (a colineação é uma matriz de 8 elementos, mas é definida em função de um fator de escala arbitrário). Aqui o critério a ser minimizado é a soma do quadrado das distâncias entre os pontos projetados nos planos das imagens. Medem-se respectivamente as distâncias dos dois pontos originais na imagem do instante depois do movimento ($\hat{\mathbf{m}}_B$ e $\hat{\mathbf{m}}'_B$) até as projeções computadas a partir da matriz que estima o movimento ($\hat{\mathbf{P}} \mathbf{H}_{AB} \tilde{\mathbf{m}}_A$ e $\hat{\mathbf{P}}' \mathbf{H}_{AB} \tilde{\mathbf{m}}_A$), como segue:

$$J = \sum_{i=1}^n d^2(\hat{\mathbf{m}}_{iB}, \hat{\mathbf{P}} \mathbf{H}_{AB} \tilde{\mathbf{m}}_{iA}) + d^2(\hat{\mathbf{m}}'_{iB}, \hat{\mathbf{P}}' \mathbf{H}_{AB} \tilde{\mathbf{m}}_{iA}) \quad (7.8)$$

7.2 Resultados

Com pontos gerados da mesma maneira descrita na seção 5.3 a matriz que aplica a colineação no espaço projetivo foi estimada e o critério (7.8) foi avaliado em função do valor de A que dá o intervalo $(-A, A)$ na distribuição uniforme, que segue a distribuição de densidade de probabilidade (5.9). Os resultados são mostrados na tabela 7.1

Tabela 7.1: Minimização do critério da expressão (7.8) para diferentes valores de A em (5.9).

A	Média do critério minimizado
10^{-5}	1.08×10^{-5}
10^{-4}	1.73×10^{-5}
10^{-3}	3.64×10^{-5}
10^{-2}	3.72×10^{-3}
10^{-1}	2.84×10^{-1}

Em dados reais, para as mesmas imagens mostradas nas figuras 6.2 e 6.3, observou-se a minimização de $J = 1.646 \times 10^1$ para o critério (7.8).

8 COMPUTAÇÃO DOS PARÂMETROS DA CÂMERA

A partir da matriz \mathbf{H}_{AB} é possível desenvolver-se uma solução de forma fechada para os parâmetros da câmera da esquerda. O método descrito neste capítulo não depende explicitamente da calibração no espaço afim como acontece em abordagens estratificadas. Além disso, todos os cálculos necessários baseiam-se em técnicas de álgebra linear tais como decomposição em valores singulares (CSURKA; DEMIRDJIAN; HORAUD, 1998).

Sabe-se (FAUGERAS, 1993) que a relação entre a matriz dos parâmetros intrínsecos \mathbf{K} e a imagem da cônica absoluta Ω é dada por $\Omega \simeq \mathbf{K}^{-T} \mathbf{K}^{-1}$. Nesta abordagem a calibração se baseia na estimação da dual $\mathbf{A} = \Omega^{-T}$ desta cônica. Para uma câmera de 4 parâmetros \mathbf{A} é:

$$\mathbf{A} \simeq \mathbf{K} \mathbf{K}^T = \begin{bmatrix} \alpha^2 + u_0^2 & u_0 v_0 & u_0 \\ u_0 v_0 & k^2 \alpha^2 + v_0^2 & v_0 \\ u_0 & v_0 & 1 \end{bmatrix} \quad (8.1)$$

com \mathbf{K} definido como na expressão (2.30), sendo $r = 0$.

A matriz \mathbf{A} é simétrica e o elemento $a_{33} = 1$ determina o fator de escala de forma que $\mathbf{A} = \mathbf{K} \mathbf{K}^T$. Para a câmera de 4 parâmetros, como pode ser notado em (8.1), existe ainda a restrição $a_{12} - a_{13} a_{23} = 0$.

O fato de que traços e determinantes permanecem inalterados mediante transformações de similaridade, permite computar-se o fator de escala da expressão (7.2) que é

$$\lambda_c = \text{sign}(\text{trace}(\mathbf{H}_{AB})) \sqrt[4]{\det(\mathbf{H}_{AB})} \quad (8.2)$$

Eliminando λ_c pode-se normalizar o movimento projetivo:

$$\overline{\mathbf{H}}_{AB} = \frac{\mathbf{H}_{AB}}{\lambda_c} = \mathbf{H}_{PE}^{-1} \mathbf{D}_{AB} \mathbf{H}_{PE} \quad (8.3)$$

onde agora $\det(\overline{\mathbf{H}}_{AB}) = 1$.

A matriz \mathbf{D}_{AB} que descreve o deslocamento do sistema de visão no \mathbb{R}^3 possui uma estrutura bem específica:

$$\mathbf{D}_{AB} = \begin{bmatrix} \mathbf{R}_{AB} & \mathbf{t}_{AB} \\ \mathbf{0}_3^T & 1 \end{bmatrix} \quad (8.4)$$

As matrizes \mathbf{D}_{AB} e \mathbf{H}_{AB} são similares e portanto compartilham o mesmo conjunto de auto-valores, que são $\{e^{i\psi}, e^{-i\psi}, 1, 1\}$, onde ψ é o ângulo associado à rotação.

Dada qualquer matriz Φ com tais auto-valores, existe uma matriz não singular Λ tal

que

$$\Phi = \Lambda \begin{bmatrix} \cos \psi & -\sin \psi & 0 & 0 \\ \sin \psi & \cos \psi & 0 & 0 \\ 0 & 0 & 1 & \epsilon \\ 0 & 0 & 0 & 1 \end{bmatrix} \Lambda^{-1} \quad (8.5)$$

onde $\epsilon = 1$ se a multiplicidade geométrica¹ do auto-valor unitário duplo é 1, e $\epsilon = 0$ se sua multiplicidade geométrica é 2 (CSURKA; DEMIRDJIAN; HORAUD, 1998). A fatorização do tipo acima é conhecida como *fatorização Jordan real*.

A multiplicidade geométrica do auto-valor duplo é 2 no caso de movimentos planares, ou seja, movimentos onde tanto o deslocamento quanto a rotação estão restritos a um mesmo plano (CSURKA; DEMIRDJIAN; HORAUD, 1998). Para um movimento genérico, que é a hipótese aqui estudada, a multiplicidade geométrica deste auto-valor unitário é 1 (CSURKA; DEMIRDJIAN; HORAUD, 1998).

A fatorização Jordan real de uma matriz Φ não é única. Tanto $\Phi = \Lambda J \Lambda$ quanto $\Phi = \Lambda' J \Lambda'$ são fatorizações Jordan reais se e apenas se existe uma matriz M tal que $MJ = JM$ onde $\Lambda' = \Lambda M$.

A partir da igualdade $MJ = JM$ deriva-se a estrutura de M , que para um movimento genérico tem 4 graus de liberdade e se apresenta da seguinte forma:

$$M = \begin{bmatrix} \alpha & -\beta & 0 & 0 \\ \beta & \alpha & 0 & 0 \\ 0 & 0 & \gamma & \omega \\ 0 & 0 & 0 & \gamma \end{bmatrix} \quad (8.6)$$

Se $\Phi = D$ é um movimento no espaço euclidiano, então $\Lambda = \Sigma$, ou seja, $D = \Sigma J \Sigma^{-1}$ onde Σ será da forma

$$\Sigma = \begin{bmatrix} Q & t \\ 0_3^T & 1 \end{bmatrix} \quad (8.7)$$

sendo Q uma matriz ortogonal, ou seja Σ é uma rotação seguida de uma translação. Com mais esta restrição, M torna-se

$$M = \begin{bmatrix} \cos \psi & \sin \psi & 0 & 0 \\ -\sin \psi & \cos \psi & 0 & 0 \\ 0 & 0 & \pm 1 & \omega \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (8.8)$$

A fatorização é estimada na forma

$$H_{AB} = [u_1 \ u_2 \ u_3 \ w] J [u_1 \ u_2 \ u_3 \ w]^{-1} = \Lambda J \Lambda^{-1} \quad (8.9)$$

Em (CSURKA; DEMIRDJIAN; HORAUD, 1998) é mostrado que os vetores u_1 e u_2 podem ser encontrados sabendo que $\text{trace}(H_{AB}) = 2 + 2 \cos \psi$, de onde $\cos \psi = \frac{\text{trace}(H_{AB}) - 2}{2}$ e $\sin \psi = \sqrt{1 - \cos \psi}$ e fazendo-se

$$U \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = 0_8 \quad (8.10)$$

onde

$$U = \begin{bmatrix} H_{AB} - \cos \psi I_4 & -\sin \psi I_4 \\ \sin \psi I_4 & H_{AB} - \cos \psi I_4 \end{bmatrix} \quad (8.11)$$

¹Multiplicidade geométrica de um auto-valor é a dimensão do auto-espaço associado ao mesmo.

Assim $[\mathbf{u}_1^T \ \mathbf{u}_2^T]^T$ é encontrado como sendo o auto-vetor associado ao menor auto-valor de $\mathbf{U}^T \mathbf{U}$.

Na prática percebeu-se que este auto-vetor precisa ser estimado com cuidado pois os elementos que compõem os auto-vetores de $\mathbf{U}^T \mathbf{U}$ tendem a ser muito discrepantes em amplitude. A maioria dos algoritmos convencionais comumente retornavam um auto-vetor incorreto, incluindo a rotina `eig()` do *Matlab*TM e as rotinas `jacobi()` e `tqli()` publicadas em (PRESS et al., 1992). Nesta implementação foi necessário o uso da função `dsyevr_()` da versão em C da biblioteca LAPACK (ANDERSON et al., 1999).

O vetor \mathbf{u}_3 é o auto-vetor associado ao auto-valor unitário (CSURKA; DEMIRDJIAN; HORAUD, 1998), encontrado resolvendo-se

$$[\mathbf{H}_{AB} - \mathbf{I}_4] \mathbf{u}_3 = \mathbf{0}_4 \quad (8.12)$$

onde novamente a solução é o auto-vetor associado ao menor auto-valor de $\mathbf{U}^T \mathbf{U}$ para agora sendo $\mathbf{U} = [\mathbf{H}_{AB} - \mathbf{I}_4]$.

Por fim, o vetor \mathbf{w} é encontrado como sendo (CSURKA; DEMIRDJIAN; HORAUD, 1998)

$$\mathbf{w} = \begin{bmatrix} [\mathbf{H}_{AB} - \mathbf{I}_3]^{-1} \bar{\mathbf{u}}_3 \\ w_4 \end{bmatrix} \quad (8.13)$$

onde $\mathbf{u}_3 = [\bar{\mathbf{u}}_3^T \ u_3]^T$ e w_4 é escolhido arbitrariamente desde que $\det([\mathbf{u}_1 \ \mathbf{u}_2 \ \mathbf{u}_3 \ \mathbf{w}]) \neq 0$

Considerando $\overline{\Lambda \mathbf{M}}$ como sendo o bloco superior esquerdo 3×3 da matriz $\Lambda \mathbf{M}$ existe a seguinte relação (CSURKA; DEMIRDJIAN; HORAUD, 1998)

$$\mathbf{K} \mathbf{K}^T = (\overline{\Lambda \mathbf{M}}) (\overline{\Lambda \mathbf{M}})^T \quad (8.14)$$

Assim, levando-se em conta a estrutura de \mathbf{M} ,

$$\Lambda \mathbf{M} = \begin{bmatrix} \alpha \mathbf{u}_1 + \beta \mathbf{u}_2 & \alpha \mathbf{u}_2 - \beta \mathbf{u}_1 & \gamma \mathbf{u}_3 & \omega \mathbf{u}_3 + \gamma \mathbf{w} \end{bmatrix} \quad (8.15)$$

deduz-se que

$$\mathbf{K} \mathbf{K}^T = (\overline{\Lambda \mathbf{M}}) (\overline{\Lambda \mathbf{M}})^T = (\alpha^2 + \beta^2)(\bar{\mathbf{u}}_1 \bar{\mathbf{u}}_1^T + \bar{\mathbf{u}}_2 \bar{\mathbf{u}}_2^T) + \gamma^2 \bar{\mathbf{u}}_3 \bar{\mathbf{u}}_3^T \quad (8.16)$$

onde $\mathbf{u}_i = [\bar{\mathbf{u}}_i^T \ u_i]^T$. Dadas as restrições que se aplicam a uma câmera de quatro parâmetros, tem-se que $a_{33} = 1$ e que $a_{12} - a_{13}a_{23} = 0$ onde a_{ij} é o elemento de $\mathbf{A} = \mathbf{K} \mathbf{K}^T$ na linha i com a coluna j . A partir destas restrições e com (8.15), tem-se

$$\gamma^2 = - \frac{\prod_{j=1}^2 (u_1^{(j)} u_2^{(3)} - u_1^{(3)} u_2^{(j)})}{\prod_{j=1}^2 (u_3^{(3)} (u_1^{(j)} u_1^{(3)} + u_2^{(j)} u_2^{(3)}) - u_3^{(j)} ((u_1^{(3)})^2 + (u_2^{(3)})^2))} \quad (8.17)$$

$$(\alpha^2 + \beta^2) = \frac{1 - \gamma^2 (u_3^{(3)})^3}{((u_1^{(3)})^2 + (u_2^{(3)})^2)} \quad (8.18)$$

onde $u_i^{(j)}$ é o j -ésimo elemento do vetor \mathbf{u}_i .

Os parâmetros da câmera da esquerda são então calculados encontrando-se $\mathbf{K} \mathbf{K}^T$ das expressões (8.10~8.13) e (8.16~8.18).

A partir daí é possível recuperar ambas as matrizes de projeção que definem as câmeras no espaço euclidiano. Em (HORAUD; CSURKA; DEMIRDJIAN, 2000) tem-se que

$$\begin{bmatrix} \mathbf{H}_{AB}^T - \mathbf{I}_4 \end{bmatrix} \pi_\infty = \mathbf{0}_4 \quad (8.19)$$

com

$$\pi_{\infty} = \begin{bmatrix} \mathbf{a} \\ a_4 \end{bmatrix} \quad (8.20)$$

sendo π_{∞} o vetor que representa o plano no infinito.

Sabe-se que uma matriz tem os mesmos auto-valores que sua transposta. Portanto, se \mathbf{H}_{AB} tem um auto-valor duplo unitário, \mathbf{H}_{AB}^T também terá, e π_{∞} , é o auto-vetor associado ao mesmo, como mostra a expressão (8.19).

Em (CSURKA; DEMIRDJIAN; HORAUD, 1998) mostra-se que

$$\mathbf{H}_{PE} = \begin{bmatrix} \mathbf{K}^{-1} & \mathbf{0}_3 \\ \mathbf{a}^T & a_4 \end{bmatrix} \quad (8.21)$$

Para que para a base projetiva arbitrária estipulada tem-se

$$\mathbf{m} = \mathbf{P} \tilde{\mathbf{m}} \quad (8.22)$$

$$\mathbf{m}' = \mathbf{P}' \tilde{\mathbf{m}} \quad (8.23)$$

Sabe-se ainda que $\tilde{\mathbf{n}} = \mathbf{H}_{PE} \tilde{\mathbf{m}}$ é o vetor de coordenadas que representa o ponto no espaço euclideano. Desta forma tem-se que

$$\mathbf{m} = \mathbf{P} \mathbf{H}_{PE}^{-1} \mathbf{H}_{PE} \tilde{\mathbf{m}} = \mathbf{P} \mathbf{H}_{PE}^{-1} \tilde{\mathbf{n}} \quad (8.24)$$

ou seja, as matrizes de projeção das câmeras da esquerda e da direita do sistema de visão no espaço euclideano são dadas respectivamente por

$$\mathbf{P}_E = \mathbf{P} \mathbf{H}_{PE}^{-1} \quad (8.25)$$

$$\mathbf{P}'_E = \mathbf{P}' \mathbf{H}_{PE}^{-1} \quad (8.26)$$

Computando-se a inversa de \mathbf{H}_{PE} , tem-se que:

$$\mathbf{H}_{PE}^{-1} = \begin{bmatrix} \mathbf{K} & \mathbf{0}_3 \\ -\frac{\mathbf{a}^T}{a_4} \mathbf{K} & -\frac{1}{a_4} \end{bmatrix} \quad (8.27)$$

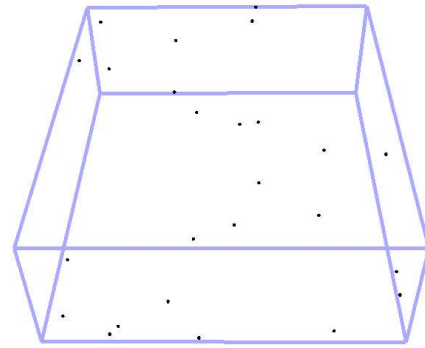
Esta última expressão permite recuperar-se as matrizes de projeção das câmeras calibradas através das expressões (8.25) e (8.26).

8.1 Resultados para dados reais

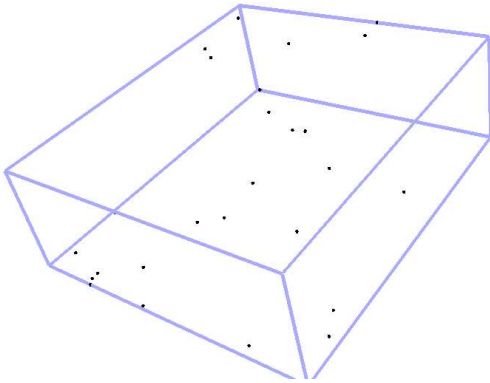
A figura 8.1 mostra a reconstrução tridimensional euclideana de alguns pontos remanescentes das etapas de pareamento, estimação da matriz fundamental e estimação da colineação. A figura 8.1(a) mostra a imagem original da câmera da esquerda antes do movimento, com os pontos em destaque. As figuras 8.1(b), 8.1(c) e 8.1(d) mostram a reconstrução destes pontos no \mathbb{R}^3 em diferentes vistas. No processo completo até a obtenção destas imagens obteve-se a matriz fundamental retornando $J = 1,121 \times 10^0$ para (5.5) e a matriz da colineação projetiva retornando $J = 1,646 \times 10^1$ para (7.8).



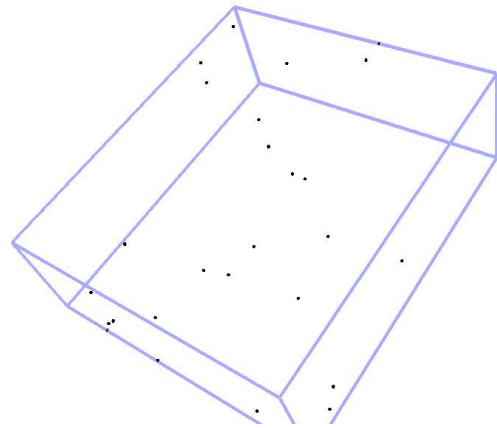
(a) Imagem original.



(b) Reconstrução no \mathbb{R}^3 : vista I.



(c) Reconstrução no \mathbb{R}^3 : vista II.



(d) Reconstrução no \mathbb{R}^3 : vista III.

Figura 8.1: Reconstrução de pontos no \mathbb{R}^3 .

9 CONCLUSÃO

Este trabalho apresentou um método totalmente automático para a calibração de um par de câmeras que realiza um movimento desconhecido. O método de calibração foi descrito e equacionado de forma completa, desde as imagens até a recuperação das matrizes de projeção das câmeras. O processo foi apresentado em diferentes etapas, partindo do pareamento de pontos entre as imagens, estimação da geometria epipolar, triangulação, estimação do movimento projetivo até por fim realizar o levantamento dos parâmetros das câmeras.

A etapa de casamento de pontos é adequada para identificação e pareamento de pequenas amostras de pontos bem localizados nas imagens, porém certamente não é o método mais adequado para a densa reconstituição tridimensional de cenas, já que isso exigiria diversas outras etapas, entre elas a retificação das imagens e identificação de planos, bordas entre outros (FRANCA, 2003). Além disso, apesar do processo não fazer nenhum tipo de suposição sobre a estrutura tridimensional da cena observada, a etapa de pareamento de pontos deixa claro que certas características são desejáveis nas imagens, tais como textura, sombra, cantos, etc.

Quanto à estimação da geometria epipolar, experimentos mostraram que esta importante característica do sistema binocular já pode ser estimada eficientemente e com razoável precisão. Ainda assim melhoramentos podem ser atingidos com o aumento tanto na quantidade de pontos pareados, quanto na precisão da localização dos mesmos.

No que se refere à triangulação e reconstrução tridimensional projetiva, estas etapas se mostraram complementares. A reconstrução tridimensional projetiva é uma etapa intermediária importante que inevitavelmente acumula erros provenientes da estimação da matriz fundamental.

O processo de estimação da colineação se assemelha muito ao processo de estimação da matriz fundamental. A qualidade da colineação resultante, assim como na estimação da matriz fundamental, depende da quantidade de amostras, mas também da qualidade da reconstrução das mesmas.

Fora do escopo deste trabalho, a robustez do método em suas diversas etapas deve ser cuidadosamente avaliada. Outros trabalhos mostram estudos de robustez na calibração automática em função do ruído na localização de pontos (HORAUD; CSURKA; DEMIRDJIAN, 2000) e em alguns movimentos críticos (CSURKA; DEMIRDJIAN; HORAUD, 1998).

Este método apresenta atrativos interessantes tais como a simplicidade e a versatilidade que a ausência de um objeto de calibração 3D oferece. Estes atrativos têm se demonstrado suficientes para justificar o estudo de métodos como este por diversos pesquisadores, como vem acontecendo nos últimos anos (vide referências da seção 1.2).

Para trabalhos futuros é possível aprimorar o método aqui apresentado no intuito de

torná-lo mais robusto e rápido. O casamento de pontos pode retornar coordenadas em frações de píxel através da estimação da correlação cruzada em posições deslocadas nos dois eixos e utilizando os resultados para a interpolação de uma função normal 2D. Maximizando a função interpolada obtém-se coordenadas em frações de píxeis (ZHANG et al., 1995).

Também podem-se realizar estudos no sentido de buscar o aproveitamento diversos pares de imagens adquiridos seqüencialmente durante o movimento do sistema de visão, através de um sistema em tempo-real, para refinação dos resultados.

REFERÊNCIAS

ANDERSON, E.; BAI, Z.; BISCHOF, C.; BLACKFORD, S.; DEMMEL, J.; DONGARRA, J.; DU CROZ, J.; GREENBAUM, A.; HAMMARLING, S.; MCKENNEY, A.; SORENSEN, D. **LAPACK Users' Guide**. 3.ed. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1999.

BALLARD, D. H.; BROWN, C. M. **Computer Vision**. Englewood Cliffs, New Jersey: Prentice-Hall Inc., 1982.

BARNARD, S.; THOMPSON, W. Disparity analysis of images. **IEEE Trans. Pattern Analysis and Machine Intelligence**, Washington, DC, v.2, n.4, p.333–340, July 1980.

BEARDSLEY, P. A.; REID, I. D.; ZISSERMAN, A.; MURRAY, D. W. Active Visual Navigation Using Non-Metric Structure. In: ICCV, 1995, Boston, MA. **Anais...** [S.l.: s.n.], 1995. p.58–.

BIRCHFIELD, S. **An Introduction to Projective Geometry**. Stanford, CA: ISPRS, 1998.

BROWN, D. C. Close-range camera calibration. **Photogrammetric Engineering**, Bethesda, Maryland, v.37, n.8, p.855–866, 1971.

CHENG, J. K.; HUANG, T. Image Registration by Matching Relational Structures. **Pattern Recognition**, New York, v.17, n.1, p.149–159, 1984.

CHOU, C.-H.; CHEN, Y.-C. Moment-preserving pattern matching. **Pattern Recognition**, New York, v.23, n.5, p.461–474, 1990.

CSURKA, G.; DEMIRDJIAN, D.; HORAUD, R. Closed form solutions for the euclidean calibration of a stereo rig. **Proceedings of the 5th European Conference on Computer Vision**, Freiburg, Germany, p.426–444, 1998.

CSURKA, G.; DEMIRDJIAN, D.; HORAUD, R. Finding the collineation between two projective reconstructions. **Computer Vision and Image Understanding: CVIU**, San Diego, CA, v.75, n.3, p.260–268, 1999.

DERMIDJIAN, D.; CSURKA, G.; HORAUD, R. Autocalibration in the presence of critical motions. **British Machine Vision Conference**, Southampton, Hampshire, UK, p.751–759, september 1998.

DRAPER, B. Learning grouping strategies for 2d and 3d object recognition. **Proceedings ARPA Image Understanding Workshop**, Palm Springs, CA, p.1447–1454, 1996.

FAIG, W. Calibration of close-range photogrammetric systems: mathematical formulation. **Photogrammetric Eng. and Remote Sensing**, Bethesda, Maryland, v.14, n.12, p.1479–1486, 1975.

FAUGERAS, O. **Three-Dimensional Computer Vision**. London: MIT Press, 1993.

FAUGERAS, O.; LUONG, Q.-T.; MAYBANK, S. J. Camera self-calibration: theory and experiments. **Proceedings of the 2nd European conference on computer vision**, Santa Margherita Ligure, Italy, p.321–334, may 1992.

FRANCA, J. A. de. **Desenvolvimento de Algoritmos de Visão Estereoscópica para Aplicações em Robótica Móvel**. [S.l.]: UFSC, Brasil, 2003.

FUA, P. A parallel stereo algorithm that produces dense depth maps and preserves image features. **Machine Vision and Applications**, New York, v.6, n.1, p.35–49, 1993.

GOSHTASBY, A.; GAGE, S.; BARTHOLIC, J. A Two-Stage Cross-Correlation Approach to Template Matching. **IEEE Trans. Pattern Analysis and Machine Intelligence**, Washington, DC, v.6, n.3, p.374–378, 1984.

HANNAH, M. A System for Digital Stereo Image Matching. **Photogrammetric Engineering & Remote Sensing**, Bethesda, Maryland, v.55, p.1765–1770, 1989.

HARTLEY, R. In defence of the 8-point algorithm. **Proceedings of the 5th International Conference on Computer Vision**, Boston, USA, p.1064–1070, june 1995.

HARTLEY, R. I. Self-calibration from multiple views with a rotating camera. In: **THIRD EUROPEAN CONFERENCE ON COMPUTER VISION**, 1994, Stockholm, Sweden. **Proceedings...** [S.l.: s.n.], 1994. p.471–478.

HARTLEY, R. I.; STURM, P. Triangulation. **Computer Vision and Image Understanding**, San Diego, CA, v.68, n.2, p.146–157, november 1997.

HORAUD, R.; CSURKA, G.; DEMIRDIJIAN, D. Stereo Calibration from Rigid Motions. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Washington, DC, v.22, n.12, p.1446–1452, 2000.

HORAUD, R.; SKORDAS, T. Stereo Correspondence Through Feature Grouping and Maximal Cliques. **IEEE Trans. Pattern Analysis and Machine Intelligence**, Washington, DC, v.11, n.11, p.1168–1180, November 1989.

KITANO, H. RoboCup Rescue: a grand challenge for multi-agent systems. **Fourth International Conference on MultiAgent Systems (ICMAS-2000)**, Boston, Massachusetts, 2000. Invited Talk.

KITANO, H.; ASADA, M. RoboCup Humanoid Challenge: that's one small step for a robot, one giant leap for mankind. **Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems**, Victoria, British Columbia, Canada, p.419–424, 1998.

LAGES, W. F. **Controle e Estimação de Posição e Orientação de Robôs Móveis**. 1998. Tese (Doutorado em Engenharia Eletrônica e Computação) — Instituto Tecnológico de Aeronáutica, São José dos Campos, SP, Brasil.

- LEONARD, J. J.; DURRANT-WHYTE, H. F. **Directed sonar sensing for mobile robot navigation**. Norwell, MA, USA: Kluwer Academic Publishers, 1992.
- MAITRE, H.; WU, Y. Improving dynamic programming to solve image registration. **Pattern Recognition**, New York, v.20, n.4, p.443–462, 1987.
- MENG, X.; HU, Z. A new easy camera calibration technique based on circular points. **The Journal of the pattern recognition society**, Vienna, v.36, p.1155–1164, 2003.
- MOHR, R.; TRIGGS, B. Projective Geometry for Image Analysis. **International Society for Photogrammetry and Remote Sensing**, Vienna, July 1996.
- MOONS, T.; GOOL, L. V.; PROESMANS, M.; PAUWELS, E. Affine reconstruction from perspective image pairs with a relative object camera translation in between. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Washington, DC, v.18, n.1, p.77–83, 1996.
- MORAVEC, H. P. Techniques Towards Automatic Visual Obstacle Avoidance. **Proceedings of the 5th International Joint Conference on Artificial Intelligence**, Cambridge, USA, August 1977.
- NOBLE, J. Finding corners. **Image and vision computing**, Amsterdam, v.6, p.121–128, May 1988.
- PHILIPS, M.; LEVY, S.; MUNZNER, T. Geomview - An interactive geometry viewer. **Notices of the American Mathematical Society**, Providence, RI, p.985–988, October 1993.
- POLLEFEYS, M. **Tutorial on 3D modeling from images**. Dublin, Ireland: ECCV, 2000.
- PRESS, W. H.; TEUKOLSKY, S. A.; VETTERLING, W. T.; FLANNERY, B. P. **Numerical Recipes in C**. Cambridge, MA: Cambridge University Press, 1992.
- RADIG, B. Image sequence analysis using relational structures. **Pattern Recognition**, New York, v.17, n.1, p.161–167, 1984.
- SCHMID, C.; MOHR, R.; BAUCKHAGE, C. Evaluation of Interest Point Detectors. **International Journal of Computer Vision**, Norwell, MA, USA, v.37, n.2, p.151–172, 2000.
- SHAPIRO, L. G.; MARALICK, R. M. Structural description and inexact matching. **IEEE Trans. Pattern Analysis and Machine Intelligence**, Washington, DC, v.3, p.504–519, September 1981.
- SUNG, K. K.; POGGIO, T. Example-Based Learning for View-Based Human Face Detection. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Washington, DC, v.20, n.1, p.39–51, 1998.
- ULLMAN, S. **The Interpretation of Visual Motion**. Cambridge, MA: The MIT Press, 1979.
- WENG, J.; AHUJA, N.; HUANG, T. Matching Two Perspective Views. **IEEE Trans. Pattern Analysis and Machine Intelligence**, Washington, DC, v.14, n.8, p.806–825, August 1992.

ZHANG, Z. Determining the Epipolar Geometry and its Uncertainty: a review. **International Journal of Computer Vision**, Boston, USA, v.27, n.2, p.161–195, 1998.

ZHANG, Z. A flexible new technique for camera calibration. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Washington, DC, v.22, n.11, p.1330–1334, 2000.

ZHANG, Z.; DERICHE, R.; FAUGERAS, O. D.; LUONG, Q.-T. A Robust Technique for Matching two Uncalibrated Images Through the Recovery of the Unknown Epipolar Geometry. **Artificial Intelligence**, Essex, UK, v.78, n.1-2, p.87–119, 1995.

ZHANG, Z.; FAUGERAS, O.; DERICHE, R. An Effective Technique for Calibrating a Binocular Stereo Through Projective Reconstruction Using Both a Calibration Object and the Environment. **Videre: Journal of Computer Vision Research**, Boston, USA, v.1, n.1, 1997.

ZISSERMAN, A.; BEARDSLEY, P. A.; REID, I. Metric Calibration of a Stereo Rig. **Proceedings of the IEEE Workshop on Representation of Visual Scenes**, Cambridge, MA, p.93–100, June 1995.

APÊNDICE A CONCEITOS DE GEOMETRIA PROJETIVA

A.1 Introdução

Quando considera-se o processo de captura de imagens de uma câmera, fica claro que a geometria Euclideana é insuficiente, pois comprimentos e ângulos não são mais preservados e retas paralelas podem se interceptar.

O estudo da chamada *geometria projetiva* foi iniciado pelos pintores renascentistas italianos que queriam produzir imagens de forma a produzir a ilusão de profundidade tridimensional em suas pinturas arquitetônicas (MOHR; TRIGGS, 1996). A geometria projetiva como é hoje foi inventada pelo matemático francês Desargues (1591-1661) (FAUGERAS, 1993).

Observa-se o exemplo da figura A.1. Apesar das retas de ambas as bordas da estrada estarem em paralelo no \mathbb{R}^3 , na imagem estas retas parecem convergir conforme se aproximam do horizonte. Qualquer par de linhas paralelas horizontais parece se encontrar num ponto do horizonte correspondente à sua direção em comum. Mais ainda, quaisquer planos horizontais parecem se encontrar e se interceptar na linha do horizonte, ou linha no infinito.

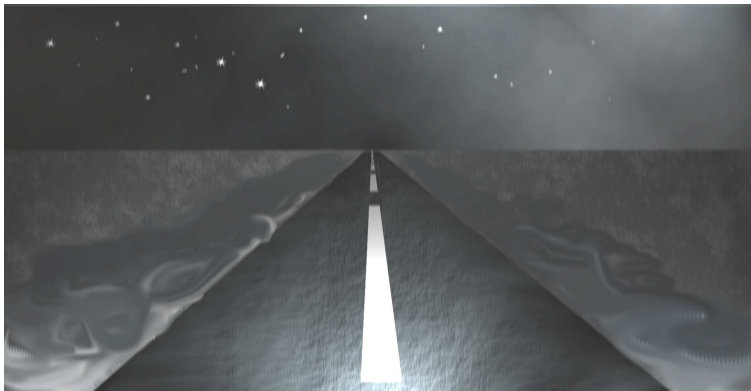


Figura A.1: Imagem de uma estrada.

Todas estas intersecções no infinito permanecem constantes conforme o observador se move. A estrada parece sempre desaparecer no mesmo ponto do horizonte e, à noite, as estrelas permanecem fixas enquanto se caminha.

Estes simples exemplos mostram que conceitos convencionais de geometria finita precisam se estender para lidar com os fenômenos “no infinito”.

A geometria projetiva modela adequadamente o processo de captura da imagem de uma câmera porque permite uma classe muito maior de transformações do que apenas

translações e rotações, uma classe que inclui projeções de perspectiva. Em contraponto, na geometria projetiva menos medidas são preservadas (BIRCHFIELD, 1998).

As transformações projetivas preservam:

- Tipo: pontos continuam pontos e retas continuam retas;
- Incidência: se um ponto cai sobre uma reta, ele continuará caindo sobre esta reta após a transformação; e
- Razão cruzada: que é uma medida que será explicada na seção A.3.12

A.2 Espaços projetivos n -dimensionais

O chamado *espaço projetivo n -dimensional* é denotado \mathbb{P}^n . Um ponto \mathbf{p} neste espaço é representado por um vetor com $n + 1$ coordenadas (números reais ou complexos), onde pelo menos uma é diferente de zero:

$$\mathbf{p} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \\ x_{n+1} \end{bmatrix} \in \mathbb{P}^n \quad (\text{A.1})$$

Os números x_i são as chamadas *coordenadas homogêneas* deste ponto. Em geometria projetiva um ponto é definido **independentemente de escala**, isto é:

$$\mathbf{p} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \\ x_{n+1} \end{bmatrix} = \begin{bmatrix} \lambda x_1 \\ \vdots \\ \lambda x_n \\ \lambda x_{n+1} \end{bmatrix} \quad \forall \quad \lambda \neq 0 \quad (\text{A.2})$$

A.2.1 Base projetiva

Uma *base projetiva* do espaço \mathbb{P}^n é um conjunto de $n + 2$ pontos tais que $n + 1$ são linearmente independentes. A base projetiva padrão é o conjunto formado por:

$$\mathbf{e}_i = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \quad \forall \quad i = 1, \dots, n + 1 \quad \mathbf{e}_{n+2} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \quad (\text{A.3})$$

Na expressão acima, o valor 1 está sempre na i -ésima posição para $i = 1, \dots, n + 1$. Por exemplo, a base projetiva padrão para o \mathbb{P}^1 ($n = 1$) é:

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \mathbf{e}_3 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (\text{A.4})$$

A.2.2 Colineação

Uma *colineação* é uma transformação linear de $\mathbb{P}^n \rightarrow \mathbb{P}^n$, definida por uma matriz \mathbf{A} de dimensão $(n+1) \times (n+1)$ tal que $\det \mathbf{A} \neq 0$ (FAUGERAS, 1993). Qualquer colineação sempre preserva a colinearidade dos pontos nos quais é aplicada e é definida em função de um fator de escala não nulo.

$$\mathbf{p}' \simeq \mathbf{A}\mathbf{p} \quad (\text{A.5})$$

onde a notação $a \simeq b$ significa que $\lambda a = b$ para algum λ arbitrário

Proposição 1 *Sejam $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n, \mathbf{x}_{n+1}, \mathbf{x}_{n+2}$, $n+2$ vetores representando pontos no \mathbb{P}^n , onde $n+1$ deles são linearmente independentes, ou seja, formam uma base projetiva. Sejam ainda $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n, \mathbf{e}_{n+1}, \mathbf{e}_{n+2}$ a base projetiva padrão, então existem matrizes não singulares \mathbf{A} tais que*

$$\mathbf{A}\mathbf{e}_i = \lambda_i \mathbf{x}_i \quad (\text{A.6})$$

onde os λ_i são escalares diferentes de zero; quaisquer duas matrizes que satisfaçam esta propriedade diferem no máximo por um fator escalar (FAUGERAS, 1993).

Prova: A matriz \mathbf{A} satisfaz as primeiras $n+1$ condições

$$\mathbf{A}\mathbf{e}_i = \lambda_i \mathbf{x}_i \quad \forall \quad i = 1, \dots, n+1 \quad (\text{A.7})$$

Observe que cada multiplicação $\mathbf{A}\mathbf{e}_i$ acima resulta sempre na i -ésima coluna da matriz \mathbf{A} , ou seja, a igualdade (A.7) implica que

$$\mathbf{A} = \begin{bmatrix} \lambda_1 \mathbf{x}_1 & \cdots & \lambda_{n+1} \mathbf{x}_{n+1} \end{bmatrix} \quad (\text{A.8})$$

Desta forma resta que

$$\mathbf{A}\mathbf{e}_{n+2} = \lambda_{n+2} \mathbf{x}_{n+2} \quad (\text{A.9})$$

é equivalente a

$$\begin{bmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_{n+1} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_{n+1} \end{bmatrix} = \lambda_{n+2} \mathbf{x}_{n+2} \quad (\text{A.10})$$

Pela hipótese de independência linear dos vetores \mathbf{x}_i , a matriz à esquerda na expressão acima tem posto completo. Portanto as razões entre os λ_i são unicamente determinadas e nenhum λ_i é zero. A matriz \mathbf{A} é unicamente determinada em função de um fator escalar e é não singular. \square

A.2.3 A reta

Uma reta \mathbf{u} num espaço projetivo n -dimensional qualquer pode ser definida como o conjunto de todos os pontos \mathbf{p}_u que podem ser formados a partir da combinação linear de dois dados pontos \mathbf{p}_1 e \mathbf{p}_2 , ou seja:

$$\mathbf{p}_u = \mathbf{p}_1 + \lambda \mathbf{p}_2 \quad \forall \quad \lambda \in (-\infty, +\infty) \quad (\text{A.11})$$

Repare que a definição permanece a mesma independente de escala, ou seja, $\lambda_1 \mathbf{p}_1 + \lambda_2 \mathbf{p}_2$ com $\lambda \in (-\infty, +\infty)$ define a mesma reta independentemente dos valores de λ_1 e λ_2 , desde que ambos sejam diferentes de zero.

A.2.4 Pontos fixos

Os chamados *pontos fixos* de uma colinação são pontos que permanecem invariantes à mesma (BIRCHFIELD, 1998). Uma colinação A qualquer no \mathbb{P}^n possui $n + 1$ pontos fixos (reais ou complexos). Em outras palavras, estes pontos satisfazem a expressão:

$$A\mathbf{p} = \lambda\mathbf{p} \quad (\text{A.12})$$

Isto é, os pontos fixos são os pontos cujo vetor de coordenadas é proporcional aos auto-vetores à direita de A .

A.2.5 Mudança de base projetiva

A seguir provar-se-á que, considerando dois conjuntos de $n + 2$ pontos $\mathbf{x}_1, \dots, \mathbf{x}_{n+2}$ e $\mathbf{y}_1, \dots, \mathbf{y}_{n+2}$, todos em posições arbitrárias, existe uma única colinação que mapeia o primeiro conjunto de pontos no segundo conjunto de pontos.

Proposição 2 *Se $\mathbf{x}_1, \dots, \mathbf{x}_{n+2}$ e $\mathbf{y}_1, \dots, \mathbf{y}_{n+2}$ são dois conjuntos de $n + 2$ pontos dispostos de forma que em cada conjunto $n + 1$ vetores de coordenadas sejam linearmente independentes, i.e., formam duas bases projetivas no \mathbb{P}^n , então existe uma matriz não singular G tal que $G\mathbf{x}_i = \rho_i\mathbf{y}_i$, $i = 1, \dots, n + 2$, onde os ρ_i são escalares e a matriz G é unicamente determinada, exceto por um fator de escala (FAUGERAS, 1993).*

Prova: Como foi mostrado na proposição 1, pode-se escolher uma matriz não singular A e escalares $\lambda_1, \dots, \lambda_{n+2}$ tais que

$$A\mathbf{e}_i = \lambda_i\mathbf{x}_i \quad \forall \quad i = 1, \dots, n + 2 \quad (\text{A.13})$$

e, similarmente, pode-se escolher uma matriz não singular B e escalares μ_1, \dots, μ_{n+2} tais que

$$B\mathbf{e}_i = \mu_i\mathbf{y}_i \quad \forall \quad i = 1, \dots, n + 2 \quad (\text{A.14})$$

Então, isolando \mathbf{e}_i na expressão (A.13),

$$\mathbf{e}_i = A^{-1}\lambda_i\mathbf{x}_i \quad (\text{A.15})$$

e substituindo na expressão (A.14) tem-se:

$$BA^{-1}\lambda_i\mathbf{x}_i = \mu_i\mathbf{y}_i \quad (\text{A.16})$$

que equivale a

$$BA^{-1}\mathbf{x}_i = \frac{\mu_i}{\lambda_i}\mathbf{y}_i \quad (\text{A.17})$$

Assim, podemos definir $G = BA^{-1}$ e $\rho_i = \mu_i/\lambda_i$, de forma que

$$G\mathbf{x}_i = \rho_i\mathbf{y}_i \quad (\text{A.18})$$

□

A.3 O plano projetivo: \mathbb{P}^2

A principal importância do *plano projetivo*, \mathbb{P}^2 está no fato deste ser útil na modelagem do plano da imagem. Existem quatro formas de se pensar a respeito do plano projetivo (BIRCHFIELD, 1998):

1. Coordenadas Homogêneas;
2. Espaço de Raios;
3. Esfera Unitária;
4. Plano Afim Aumentado.

Todas estas formas de representação são equivalentes, sendo que cada uma salienta diferentes propriedades intrínsecas ao \mathbb{P}^2 .

A.3.1 Coordenadas homogêneas

Supondo que temos um ponto $[u \ v]^T$ no plano Euclidiano, representa-se este mesmo ponto no plano projetivo simplesmente adicionando-se uma terceira coordenada de valor 1 ao final:

$$\begin{bmatrix} u \\ v \end{bmatrix} \in \mathbb{R}^2 \Leftrightarrow \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \in \mathbb{P}^2 \quad (\text{A.19})$$

Fator de escala não é relevante, portanto o ponto $[u \ v \ 1]^T$ é o mesmo que o ponto $[\lambda u \ \lambda v \ \lambda]^T$, para qualquer $\lambda \neq 0$. Isto também significa que o ponto $[0 \ 0 \ 0]^T$ não é permitido. Em outras palavras:

$$\begin{bmatrix} \bar{u} \\ \bar{v} \\ \bar{s} \end{bmatrix} = \lambda \begin{bmatrix} \bar{u} \\ \bar{v} \\ \bar{s} \end{bmatrix} \quad \forall \ \lambda \neq 0 \quad (\text{A.20})$$

Esta é a representação de um ponto do plano projetivo em suas chamadas *coordenadas homogêneas*.

A.3.2 Espaço de raios

Foi mostrado que vindo do plano Euclidiano para o plano projetivo, aquilo que representava um ponto do \mathbb{R}^2 passa a ter três coordenadas. Pode-se considerar que um ponto no \mathbb{R}^2 , ao passar para o plano projetivo, transforma-se num ponto dentro de um conjunto específico de pontos no \mathbb{R}^3 , unidos por um fator de escala não nulo.

Portanto um ponto $\mathbf{p} = [\bar{u} \ \bar{v} \ \bar{s}]^T$ do \mathbb{P}^2 pode ser visto como uma reta no \mathbb{R}^3 que passa pela origem e pelo ponto de coordenadas \mathbf{p} no \mathbb{R}^3 (na verdade sem incluir a origem). Este espaço tridimensional é conhecido como *Espaço dos Raios* (BIRCHFIELD, 1998). O espaço de raios é ilustrado na figura A.2. De forma similar, uma reta $\mathbf{u} = [a \ b \ c]^T$ do \mathbb{P}^2 pode ser vista no espaço de raios como sendo um plano passando pela origem e perpendicular ao vetor \mathbf{u} . A reta ideal do \mathbb{P}^2 no espaço de raios seria o plano horizontal $\bar{s} = 0$ e os pontos ideais do \mathbb{P}^2 no espaço de raios seriam retas contidas naquele plano.

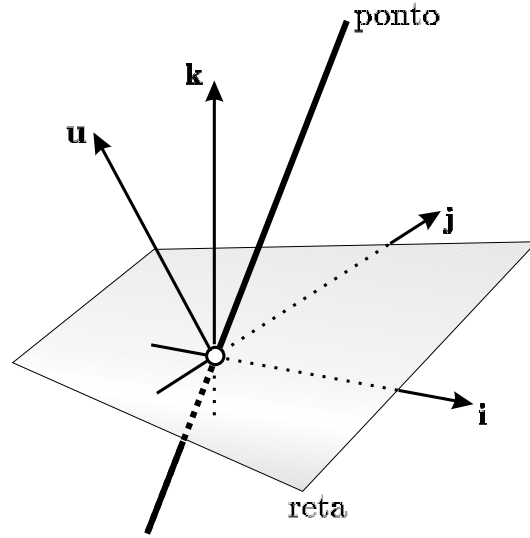


Figura A.2: Espaço de raios.

A.3.3 A esfera unitária

Como as coordenadas não são afetadas por multiplicação por escalar, \mathbb{P}^2 é bi-dimensional, mesmo que seus pontos contêm três coordenadas. Cada ponto $\mathbf{p} = (\bar{u}, \bar{v}, \bar{s})$ que é representado por uma reta no espaço de raios, pode ser projetado na esfera de raio unitário (BIRCHFIELD, 1998) centrada na origem, bastando para isso ajustar a escala de forma que $\bar{u}^2 + \bar{v}^2 + \bar{s}^2 = 1$, obtendo-se o ponto:

$$\mathbf{p} = \frac{1}{\sqrt{\bar{u}^2 + \bar{v}^2 + \bar{s}^2}} \begin{bmatrix} \bar{u} \\ \bar{v} \\ \bar{s} \end{bmatrix} \quad (\text{A.21})$$

O denominador $\sqrt{\bar{u}^2 + \bar{v}^2 + \bar{s}^2}$ nunca é zero já que o ponto $[0 \ 0 \ 0]^T$ não é permitido. Pontos no plano projetivo podem ser vistos como pontos sobre a esfera de raio unitário, como ilustra a figura A.3.

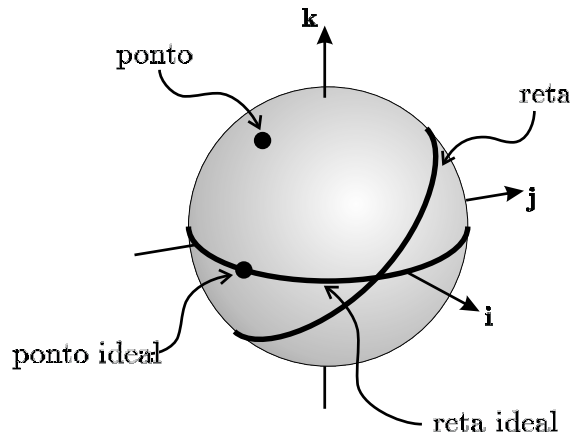


Figura A.3: Esfera de raio unitário.

Como cada reta do espaço de raios corta a esfera duas vezes, estes dois pontos da superfície da esfera representam o mesmo ponto do plano projetivo. De forma similar

cada plano do espaço de raios intercepta a esfera unitária em circunferências maiores (de raio máximo). Assim uma reta do plano projetivo é vista como uma circunferência de raio unitário, resultado da intersecção entre a esfera e o plano perpendicular a \mathbf{u} .

A reta ideal é a circunferência da seção média da esfera e os pontos ideais são os pontos sobre esta circunferência.

A.3.4 Plano \mathbb{R}^2 aumentado

Os pontos e retas do espaço de raios serão projetados, ou seja, representados respectivamente por retas e planos, sobre o plano $\bar{s} = 1$. Cada ponto $[\bar{u} \ \bar{v} \ \bar{s}]^T$ é portanto mapeado no ponto $[\frac{\bar{u}}{\bar{s}} \ \frac{\bar{v}}{\bar{s}} \ 1]^T$, que está na intersecção do plano $\bar{s} = 1$ com a reta do espaço de raios que representa o ponto. De forma similar cada reta do \mathbb{P}^2 é mapeada na reta que resulta da intersecção do plano do espaço de raios (que representa a reta do \mathbb{P}^2) com o plano $\bar{s} = 1$.

Pontos ideais e a reta ideal são projetados como pontos no infinito e a reta no infinito respectivamente, conforme ilustra a figura A.4. Assim retorna-se a uma representação onde pontos são pontos e retas são retas.

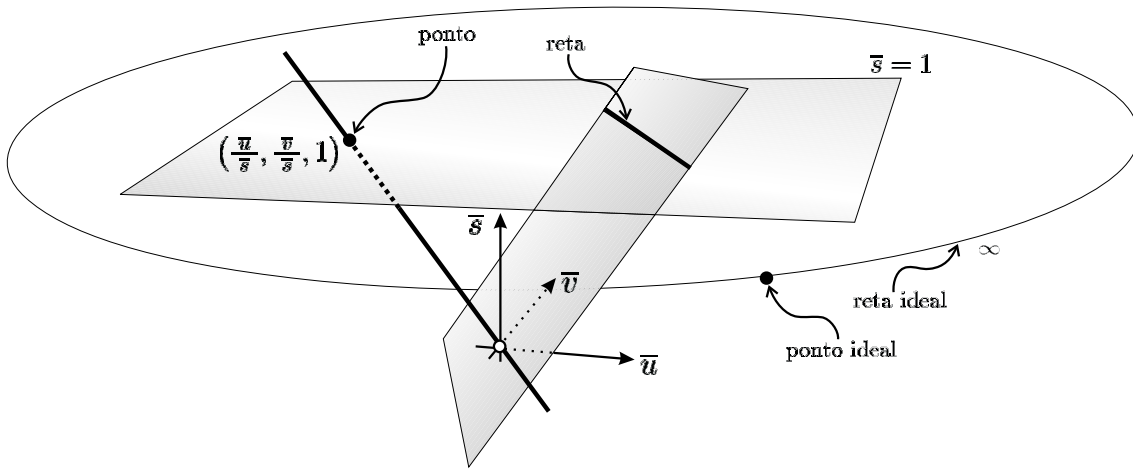


Figura A.4: Plano \mathbb{R}^2 aumentado.

Uma definição concisa do plano projetivo pode ser dada agora.

Definição 1 O plano projetivo, \mathbb{P}^2 , é o plano afim aumentado por uma única reta ideal e um conjunto de pontos ideais, um para cada direção, onde a reta ideal e os pontos ideais não são diferenciados de pontos e retas regulares.

A.3.5 A reta no \mathbb{P}^2

Para representar uma reta no plano projetivo, começaremos com a fórmula Euclidiana padrão para uma reta no \mathbb{R}^2 :

$$ax + by + c = 0 \quad (\text{A.22})$$

A expressão (A.22) representa uma reta qualquer, de declividade $-a/b$, como ilustra a figura A.5. A expressão (A.22) também pode ser escrita na forma vetorial como sendo:

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix}^T \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = 0 \quad (\text{A.23})$$

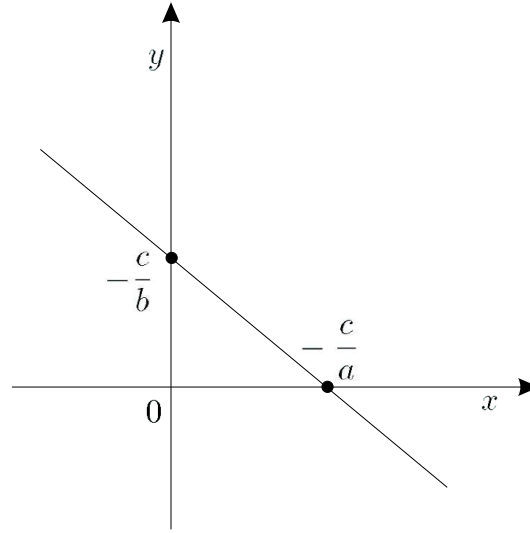


Figura A.5: Reta Euclideana.

Como no plano projetivo pontos não dependem de escala, pode-se definir $\bar{u} = \bar{s} x$ e $\bar{v} = \bar{s} y$. Assim obtém-se a expressão da reta:

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix}^T \begin{bmatrix} \bar{u} \\ \bar{v} \\ \bar{s} \end{bmatrix} = 0 \quad (\text{A.24})$$

Ou ainda, definindo-se a reta $\mathbf{u} = [a \ b \ c]^T$ e o ponto $\mathbf{p} = [\bar{u} \ \bar{v} \ \bar{s}]^T$:

$$\mathbf{u}^T \mathbf{p} = 0 \quad (\text{A.25})$$

A expressão (A.25) pode ser vista como expressão das retas \mathbf{u} que passam pelo ponto \mathbf{p} ou ainda como a expressão dos pontos \mathbf{p} que estão sobre a reta \mathbf{u} (BIRCHFIELD, 1998). Percebe-se que pontos e retas possuem a mesma representação no plano projetivo.

A.3.6 Relações entre \mathbb{R}^2 e \mathbb{P}^2

Para transformar um ponto no plano projetivo de volta para as correspondentes coordenadas Euclidianas, simplesmente divide-se as duas primeiras coordenadas pela terceira, fazendo-se $[u \ v]^T = [\bar{u}/\bar{s} \ \bar{v}/\bar{s}]^T$, ou na forma vetorial:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \bar{u}/\bar{s} \\ \bar{v}/\bar{s} \\ 1 \end{bmatrix} \quad (\text{A.26})$$

A.3.7 O ponto e a reta no infinito

O plano projetivo contém mais pontos que o plano Euclidiano: pontos cuja terceira coordenada é zero. Estes pontos são os chamados *pontos ideais* ou *pontos no infinito*:

$$\begin{bmatrix} \bar{u} \\ \bar{v} \\ 0 \end{bmatrix} \quad (\text{A.27})$$

Observe que a medida que faz-se $\bar{s} \rightarrow 0$, as coordenadas Euclidianas de um ponto qualquer $[\bar{u}/\bar{s} \ \bar{v}/\bar{s}]^T$ aumentam. O ponto vai para o infinito percorrendo a reta que o une à

origem, como mostra a figura A.6. Cada ponto ideal está associado a uma direção no plano Euclidiano. Por exemplo, os pontos $[1 \ 0 \ 0]^T$ e $[0 \ 1 \ 0]^T$ estão associados às direções horizontal e vertical respectivamente (BIRCHFIELD, 1998).

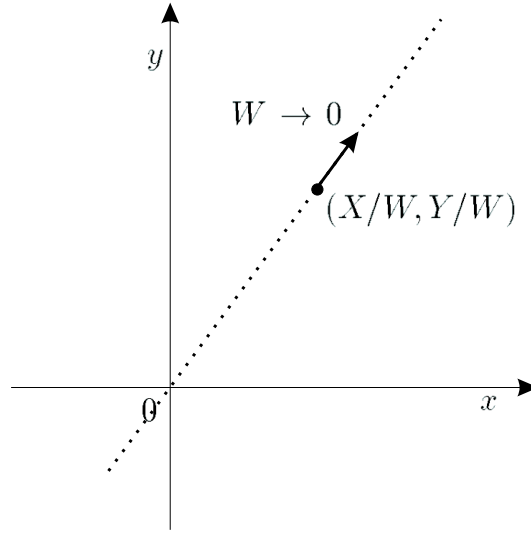


Figura A.6: Pontos ideais ou pontos no infinito.

No plano projetivo os pontos ideais não recebem nenhum tipo de tratamento especial, são como qualquer outro ponto no \mathbb{P}^2 .

Todos os pontos ideais caem sobre uma mesma reta chamada *reta ideal* ou *reta no infinito* (BIRCHFIELD, 1998), que por sua vez também é tratada como qualquer outra reta. A reta ideal é representada em sua forma canônica como $[0 \ 0 \ 1]^T$:

$$\begin{bmatrix} 0 \\ 0 \\ \bar{s} \end{bmatrix} \quad (\text{A.28})$$

Observando a expressão (A.24) percebe-se que toda reta $\mathbf{u} = [a \ b \ c]^T$ do plano projetivo intercepta a reta ideal no ponto $\mathbf{p}_\infty = [-b \ a \ 0]^T$ que é o ponto ideal associado com a direção daquela reta.

A.3.8 Base projetiva do \mathbb{P}^2

Para o plano projetivo, \mathbb{P}^2 , a base projetiva padrão é:

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{e}_4 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad (\text{A.29})$$

A.3.9 Relações entre retas e pontos

A seguir serão dadas algumas relações úteis entre pontos e retas no plano projetivo (BIRCHFIELD, 1998; MOHR; TRIGGS, 1996).

A.3.9.1 Ponto de intersecção entre duas retas

Para encontrar o ponto de intersecção entre duas retas basta fazer como na álgebra elementar:

$$\mathbf{p} = \mathbf{u}_1 \times \mathbf{u}_2 \quad (\text{A.30})$$

Sendo $\mathbf{u}_1 = [a_1 \ b_1 \ c_1]^T$ e $\mathbf{u}_2 = [a_2 \ b_2 \ c_2]^T$.

Daí tira-se $\mathbf{p} = [b_1c_2 - b_2c_1 \ a_2c_1 - a_1c_2 \ a_1b_2 - a_2b_1]^T$. Se as duas retas são paralelas, i.e., suas declividades são $\gamma = -a_1/b_1 = -a_2/b_2$, o ponto de intersecção é simplesmente $[b_1c_2 - b_2c_1 \ a_2c_1 - a_1c_2 \ 0]^T$ que é o ponto ideal associado à direção da reta de declividade γ .

Observação: Perceba que dados dois vetores \mathbf{m} e \mathbf{n} :

$$\mathbf{m} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \quad e \quad \mathbf{n} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (\text{A.31})$$

Pode-se definir uma matriz, aqui notada como $[\mathbf{m}]_{\times}$ (MOHR; TRIGGS, 1996), tal que:

$$[\mathbf{m}]_{\times} \mathbf{n} = \mathbf{m} \times \mathbf{n} \quad (\text{A.32})$$

Esta matriz é o operador linear equivalente que resulta no produto escalar dos vetores, e é definida como sendo:

$$[\mathbf{m}]_{\times} = \begin{bmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{bmatrix} \quad (\text{A.33})$$

□

A.3.9.2 Reta que passa por dois pontos

De forma similar, dados dois pontos \mathbf{p}_1 e \mathbf{p}_2 a expressão da reta que os une é dada por:

$$\mathbf{u} = \mathbf{p}_1 \times \mathbf{p}_2 \quad (\text{A.34})$$

A.3.9.3 Três pontos colineares

Agora suponha que deseja-se determinar se os pontos \mathbf{p}_1 , \mathbf{p}_2 e \mathbf{p}_3 caem sobre uma mesma reta. A reta que une os primeiros dois pontos é $\mathbf{p}_1 \times \mathbf{p}_2$. Portanto o terceiro ponto cai sobre esta reta se $\mathbf{p}_3^T(\mathbf{p}_1 \times \mathbf{p}_2) = 0$, ou, de forma mais sucinta, se o determinante da matriz 3×3 formada pelas coordenadas dos pontos é zero:

$$\begin{vmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 \end{vmatrix} = 0 \quad (\text{A.35})$$

A.3.9.4 Três retas concorrentes

Similarmente, três retas são concorrentes (interceptam-se num mesmo ponto) se:

$$\begin{vmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{vmatrix} = 0 \quad (\text{A.36})$$

A.3.10 Dualidade entre ponto e reta no \mathbb{P}^2

Existem diversas similaridades entre retas e pontos. Suas representações são idênticas. A fórmula para encontrar o ponto de intersecção de duas retas é a mesma que para encontrar a reta que une dois pontos. Estas observações são fruto da dualidade que existe entre pontos e retas no plano projetivo (FAUGERAS, 1993). Em outras palavras, qualquer teorema ou afirmação que é verdadeira para o plano projetivo pode ser reescrita palavra por palavra trocando pontos por retas e retas por pontos, e a afirmação resultante também será verdadeira.

A.3.11 Lápis de retas

Um conjunto de infinitas retas concorrentes (retas que passam por um mesmo ponto) no plano projetivo, \mathbb{P}^2 , é por sua vez um espaço projetivo uni-dimensional chamado *Lápis de retas* (FAUGERAS, 1993). Pode verificar-se que é um espaço uni-dimensional aplicando-se o princípio da dualidade: um conjunto de retas concorrentes é dual a um conjunto de pontos colineares.

Esta estrutura tem numerosas aplicações, principalmente em visão estéreo e movimento.

A.3.12 Razão cruzada

Dados quatro pontos, \mathbf{p}_1 , \mathbf{p}_2 , \mathbf{p}_3 e \mathbf{p}_4 , e supondo que os quatro sejam colineares, podemos escrever cada um deles na forma (vide subsecção A.2.3):

$$\mathbf{p}_i = \mathbf{p}_a + \lambda_i \mathbf{p}_b \quad (\text{A.37})$$

Assume-se aqui que nenhum dos pontos \mathbf{p}_i seja coincidente com \mathbf{p}_b .

Assim, a razão cruzada entre os quatro pontos pode ser definida como sendo (POLLEFEYS, 2000):

$$\{\mathbf{p}_1, \mathbf{p}_2; \mathbf{p}_3, \mathbf{p}_4\} = \frac{\lambda_1 - \lambda_3}{\lambda_1 - \lambda_4} \div \frac{\lambda_2 - \lambda_3}{\lambda_2 - \lambda_4} = \frac{(\lambda_1 - \lambda_3)(\lambda_2 - \lambda_4)}{(\lambda_1 - \lambda_4)(\lambda_2 - \lambda_3)} \quad (\text{A.38})$$

Dados dois pontos $\mathbf{p}_i = [\bar{u}_i \ \bar{v}_i \ \bar{s}_i]^T$ e $\mathbf{p}_j = [\bar{u}_j \ \bar{v}_j \ \bar{s}_j]^T$, a distância Euclideana entre estes, que aqui será denotada como Δ_{ij} , é definida como sendo:

$$\Delta_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (\text{A.39})$$

onde $x_k = \frac{\bar{u}_k}{\bar{s}_k}$ e $y_k = \frac{\bar{v}_k}{\bar{s}_k}$. Então a razão cruzada também pode ser calculada como sendo:

$$\{\mathbf{p}_1, \mathbf{p}_2; \mathbf{p}_3, \mathbf{p}_4\} = \frac{\Delta_{13}}{\Delta_{14}} \div \frac{\Delta_{23}}{\Delta_{24}} = \frac{\Delta_{13}\Delta_{24}}{\Delta_{14}\Delta_{23}} \quad (\text{A.40})$$

Em outras palavras, escolhe-se um ponto, por exemplo \mathbf{p}_1 . Calcula-se a razão das distâncias deste ponto até dois outros pontos, digamos \mathbf{p}_3 e \mathbf{p}_4 . Calcula-se, então, a razão das distâncias do ponto que resta, neste caso \mathbf{p}_2 , até aqueles mesmos dois pontos. A razão destas razões é a razão cruzada, e permanece invariante sob transformações projetivas (ou colineações). Em outras palavras, dada uma colineação \mathbf{A} :

$$\{\mathbf{p}_1, \mathbf{p}_2; \mathbf{p}_3, \mathbf{p}_4\} = \{\mathbf{A}\mathbf{p}_1, \mathbf{A}\mathbf{p}_2; \mathbf{A}\mathbf{p}_3, \mathbf{A}\mathbf{p}_4\} \quad (\text{A.41})$$

Na verdade a razão cruzada é a mesma, não importando que coordenada é utilizada como divisor em (A.39) (desde que seja usada sempre a mesma), permitindo assim os pontos ideais onde $\bar{s} = 0$ (BIRCHFIELD, 1998).

Apesar da razão cruzada ser invariante, seu valor muda conforme a ordem dos pontos escolhidos. Quatro pontos podem ser, portanto escolhidos em $4! = 24$ ordens diferentes, mas de fato apenas seis valores distintos são produzidos, que se relacionam da seguinte forma (BIRCHFIELD, 1998):

$$\left\{ \lambda, \frac{1}{\lambda}, 1 - \lambda, \frac{1}{1 - \lambda}, \frac{\lambda - 1}{\lambda}, \frac{\lambda}{\lambda - 1} \right\} \quad (\text{A.42})$$

A.3.13 Razão cruzada de quatro retas que se interceptam em um ponto

Dadas quatro retas do plano projetivo, u_1, u_2, u_3, u_4 que se interceptam em um mesmo ponto, sua razão cruzada, denotada como $\{u_1, u_2; u_3, u_4\}$, é a razão cruzada de seus respectivos quatro pontos de intersecção com qualquer outra reta u (FAUGERAS, 1993) (vide figura A.7):

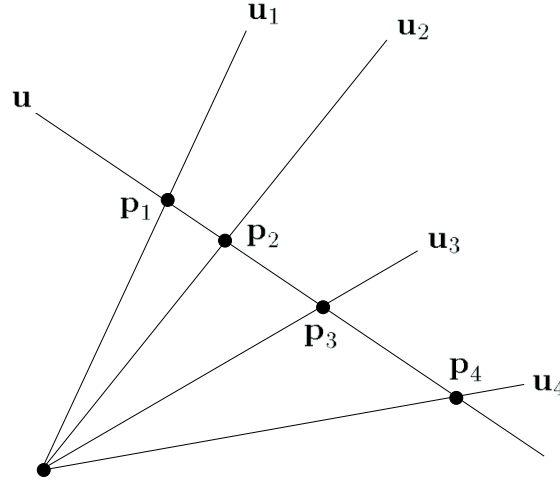


Figura A.7: Razão cruzada de entre concorrentes.

$$\{u_1, u_2; u_3, u_4\} = \{p_1, p_2; p_3, p_4\} = \frac{\Delta_{p_1 p_3} \Delta_{p_2 p_4}}{\Delta_{p_1 p_4} \Delta_{p_2 p_3}} \quad (A.43)$$

Este valor é independente da escolha de u .

A.3.14 Colineação no \mathbb{P}^2

Uma colineação no \mathbb{P}^2 é definida como um mapeamento do plano para si próprio, $\mathbb{P}^2 \rightarrow \mathbb{P}^2$, tal que a colinearidade de qualquer conjunto de pontos é preservada. Tal mapeamento é feito pela multiplicação por uma matriz 3×3 . Cada ponto p é transformado num ponto \bar{p} :

$$\bar{p} = Ap \quad (A.44)$$

Como a escala não é importante, a matriz A possui apenas 8 graus de liberdade. Portanto, como cada ponto possui dois graus de liberdade, são necessários 4 pontos para definir uma colineação.

Um fato importante sobre colineações é que, se um conjunto de pontos era colinear, após a colineação esta colinearidade é preservada.

Para aplicar uma colineação à uma reta observe que a colinearidade precisa ser preservada, isto é, se um ponto p cai sobre a linha u , então o correspondente ponto \bar{p} precisa cair na correspondente linha \bar{u} . Portanto, supondo que:

$$p^T u = 0 \quad (A.45)$$

Isola-se p em (A.44):

$$p = A^{-1} \bar{p} \quad (A.46)$$

Substitui-se (A.46) em (A.45):

$$(A^{-1} \bar{p})^T u = (\bar{p})^T (A^{-T} u) \quad (A.47)$$

Isto mostra que a colineação equivalente para as retas é:

$$\bar{\mathbf{u}} = (\mathbf{A}^{-1})^T \mathbf{u} \quad (\text{A.48})$$

A matriz $(\mathbf{A}^{-1})^T$ é chamada *matriz dual* da matriz \mathbf{A} .

Uma translação no plano de $(x, y, 1)$ para $(x + \Delta x, y + \Delta y, 1)$ pode ser escrita como:

$$\begin{bmatrix} 1 & 0 & \Delta x \\ 0 & 1 & \Delta y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (\text{A.49})$$

No \mathbb{P}^2 assim como no \mathbb{P}^1 (e em geral no \mathbb{P}^n), translações não movem os pontos no infinito.

A.3.14.1 Pontos fixos

Uma colineação \mathbf{A} qualquer no \mathbb{P}^2 possui 3 pontos fixos (reais ou complexos). Estes são os pontos cujo vetor de coordenadas é proporcional aos auto-vetores à direita de \mathbf{A} .

No \mathbb{P}^2 , as retas fixas de uma colineação \mathbf{A} são as retas que unem quaisquer dois pontos fixos. A reta fixa permanece invariante mediante a sua colineação.

A.3.15 Cônicas

Na geometria projetiva, elipses, parábolas e hipérboles são equivalentes: qualquer forma pode se projetar em qualquer outra forma. Estas curvas são chamadas de *cônicas*, sem distinções entre as diferentes formas.

Definição 2 *Uma cônica na geometria projetiva é definida como sendo o lugar geométrico dos pontos que possuem uma razão cruzada sempre constante em relação a quatro pontos fixos, onde pelo menos três destes não são colineares (BIRCHFIELD, 1998).*

Para definirmos uma cônica, primeiramente definiremos o operador do \mathbb{R}^3 para o \mathbb{R}^1 chamado $S(\cdot)$ como sendo:

$$S(\mathbf{v}) = \mathbf{v}^T \mathbf{A} \mathbf{v} \quad \text{onde } \mathbf{A} = \mathbf{A}^T \in \mathbb{R}^{3 \times 3} \text{ e } \mathbf{v} \in \mathbb{R}^3 \quad (\text{A.50})$$

A matriz \mathbf{A} é uma matriz simétrica, o vetor \mathbf{v} do \mathbb{R}^3 pode representar um ponto do \mathbb{P}^2 , e o resultado da operação é um número real.

A expressão de uma cônica é dada por:

$$S(\mathbf{p}) = \mathbf{p}^T \mathbf{C} \mathbf{p} = 0 \quad \forall \quad \mathbf{p} \in \mathbb{P}^2 \quad \text{onde } \mathbf{C} = \mathbf{C}^T \in \mathbb{R}^{3 \times 3} \quad (\text{A.51})$$

Ou seja:

$$\begin{bmatrix} \bar{u} \\ \bar{v} \\ \bar{s} \end{bmatrix}^T \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{12} & c_{22} & c_{23} \\ c_{13} & c_{23} & c_{33} \end{bmatrix} \begin{bmatrix} \bar{u} \\ \bar{v} \\ \bar{s} \end{bmatrix} = 0 \quad (\text{A.52})$$

$$c_{11}\bar{u}^2 + c_{22}\bar{v}^2 + c_{33}\bar{s}^2 + 2c_{12}\bar{u}\bar{v} + 2c_{13}\bar{u}\bar{s} + 2c_{23}\bar{v}\bar{s} = 0 \quad (\text{A.53})$$

Sendo que a matriz \mathbf{C} é simétrica e é definida independente de fator de escala, ou seja, possui 5 variáveis independentes.

A.3.15.1 Cônica dual

De forma dual, a cônica pode ser considerada como um envelope de retas tangentes àqueles pontos (BIRCHFIELD, 1998), onde a expressão equivalente é:

$$\mathbf{u}^T |\mathbf{C}| \mathbf{C}^{-1} \mathbf{u} = 0 \quad (\text{A.54})$$

Substituindo \mathbf{p} da expressão (A.46) em (A.51) tem-se que uma colinação \mathbf{A} em uma cônica é dada por:

$$\mathbf{C}' = \mathbf{A}^{-T} \mathbf{C} \mathbf{A}^{-1} \quad (\text{A.55})$$

Ou ainda, aplicando a mesma cônica à uma reta:

$$|\mathbf{C}'| \mathbf{C}'^{-1} = \mathbf{A} |\mathbf{C}| \mathbf{C}^{-1} \mathbf{A}^T \quad (\text{A.56})$$

A.3.15.2 Intersecção entre uma cônica e uma reta

Como foi visto na subseção A.2.3, uma reta que passa por dois pontos pode ser expressa como o conjunto de todos os pontos formados a partir da combinação linear daqueles dois pontos:

$$\mathbf{p}_u = \mathbf{p}_1 + \lambda \mathbf{p}_2 \quad (\text{A.57})$$

Os pontos \mathbf{p}_u estão sobre a reta \mathbf{u} e interceptam uma dada cônica se:

$$S(\mathbf{p}_u) = 0 \quad \rightarrow \quad S(\mathbf{p}_1 + \lambda \mathbf{p}_2) = 0 \quad (\text{A.58})$$

Esta última expressão pode ser aberta da seguinte forma (FAUGERAS, 1993):

$$S(\mathbf{p}_1) + 2\lambda S(\mathbf{p}_1, \mathbf{p}_2) + \lambda^2 S(\mathbf{p}_2) = 0 \quad (\text{A.59})$$

sendo que $S(\mathbf{p}_1, \mathbf{p}_2) = \mathbf{p}_1^T \mathbf{C} \mathbf{p}_2$. Em (A.59), dados \mathbf{p}_1 e \mathbf{p}_2 , e fazendo a expressão em função de λ , em geral, encontrar-se-á dois pontos que a satisfazem. Perceba que pode-se definir, por exemplo $k_1 = S(\mathbf{p}_2)$, $k_2 = 2S(\mathbf{p}_1, \mathbf{p}_2)$ e $k_3 = S(\mathbf{p}_1)$, então a expressão fica:

$$k_1 \lambda^2 + k_2 \lambda + k_3 = 0 \quad (\text{A.60})$$

que é apenas uma simples expressão quadrática em λ .

A.3.15.3 Reta tangente à cônica

Ainda olhando a expressão (A.60), percebe-se que se:

$$k_2^2 - 4k_1 k_3 = 0 \quad \rightarrow \quad S^2(\mathbf{p}_1, \mathbf{p}_2) - S(\mathbf{p}_1) S(\mathbf{p}_2) = 0 \quad (\text{A.61})$$

então (A.59) possui uma única raiz de multiplicidade dupla. Portanto a reta \mathbf{u} que passa por \mathbf{p}_1 e \mathbf{p}_2 é tangente à cônica.

A.3.16 Pontos absolutos

Todo o círculo intercepta a reta ideal $\bar{s} = 0$ em dois pontos fixos. Para perceber isto note que um círculo é uma cônica com todos os elementos fora da diagonal principal zerados e os elementos da diagonal principal iguais:

$$\mathbf{p}^T \mathbf{C} \mathbf{p} = \begin{bmatrix} \bar{u} \\ \bar{v} \\ \bar{s} \end{bmatrix}^T \begin{bmatrix} c_{11} & 0 & 0 \\ 0 & c_{11} & 0 \\ 0 & 0 & c_{11} \end{bmatrix} \begin{bmatrix} \bar{u} \\ \bar{v} \\ \bar{s} \end{bmatrix} = 0 \quad (\text{A.62})$$

Como escala não é importante, dividindo por c_{11} obtemos:

$$\bar{u}^2 + \bar{v}^2 + \bar{s}^2 = 0 \quad (\text{A.63})$$

Que portanto intercepta a reta ideal (pontos de $\bar{s} = 0$) em:

$$\bar{u}^2 + \bar{v}^2 = 0 \quad (\text{A.64})$$

Esta expressão possui duas raízes complexas chamadas *pontos absolutos* (BIRCHFIELD, 1998):

$$\mathbf{i} = \begin{bmatrix} 1 \\ i \\ 0 \end{bmatrix} \quad \text{e} \quad \mathbf{j} = \begin{bmatrix} 1 \\ -i \\ 0 \end{bmatrix} \quad (\text{A.65})$$

A.3.17 Transformação de similaridade

Transformações no plano Euclidiano conservam ângulos e razões de distâncias. Em geral, uma transformação no \mathbb{R}^2 pode ser descrita como:

$$\begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \quad (\text{A.66})$$

No plano projetivo, \mathbb{P}^2 , uma transformação com tais restrições é chamada *transformação de similaridade* (FAUGERAS, 1993). Este tipo de transformação é dada pela colineação:

$$\mathbf{T} = \begin{bmatrix} \cos \theta & \sin \theta & \Delta \bar{u} \\ -\sin \theta & \cos \theta & \Delta \bar{v} \\ 0 & 0 & \Delta \bar{s} \end{bmatrix} \quad (\text{A.67})$$

Proposição 3 *Toda a transformação que preserva os pontos absolutos*

$$\begin{bmatrix} 1 \\ +i \\ 0 \end{bmatrix} \quad \text{e} \quad \begin{bmatrix} 1 \\ -i \\ 0 \end{bmatrix} \quad (\text{A.68})$$

inalterados é uma transformação de similaridade (FAUGERAS, 1993)

Prova: *Suponha, inicialmente, uma transformação qualquer \mathbf{T} , sem restrições:*

$$\mathbf{T} = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{bmatrix} \quad (\text{A.69})$$

O fato de que $[1 \ i \ 0]^T$ é preservado implica

$$\begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{bmatrix} \begin{bmatrix} 1 \\ i \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ i \\ 0 \end{bmatrix} \quad (\text{A.70})$$

Já que escala não é importante, isto resulta nas restrições:

$$\frac{t_{11} + it_{12}}{t_{21} + it_{22}} = \frac{1}{i} \quad (\text{A.71})$$

$$t_{31} + it_{32} = 0 \quad (\text{A.72})$$

Como os elementos de \mathbf{T} são restritos aos números reais, isto leva às três seguintes restrições:

$$t_{11} = t_{22} \quad (\text{A.73})$$

$$t_{12} = -t_{21} \quad (\text{A.74})$$

$$t_{31} = t_{32} = 0 \quad (\text{A.75})$$

Assim a matriz \mathbf{T} se restringe a:

$$\mathbf{T} = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ -t_{12} & t_{11} & t_{23} \\ 0 & 0 & t_{33} \end{bmatrix} \quad (\text{A.76})$$

Dados dois números reais t_{12} e t_{13} quaisquer, sempre é possível parametrizá-los na forma polar:

$$t_{11} = k \cos \theta \quad (\text{A.77})$$

$$t_{12} = k \sin \theta \quad (\text{A.78})$$

Assim, dividindo a matriz \mathbf{T} por k (já que escala não é relevante) tem-se:

$$\mathbf{T} = \begin{bmatrix} \cos \theta & \sin \theta & t_{13}/k \\ -\sin \theta & \cos \theta & t_{23}/k \\ 0 & 0 & t_{33}/k \end{bmatrix} \quad (\text{A.79})$$

Esta última expressão está na forma (A.67). \square

A.4 O espaço projetivo: \mathbb{P}^3

Todos os conceitos discutidos até o momento têm analogias no espaço projetivo tri-dimensional, \mathbb{P}^3 : Existe dualidade entre ponto e plano, um lápis de planos é um espaço projetivo bi-dimensional, a razão cruzada entre planos é invariante, quádricas fazem o papel das cônicas, como poderá ser visto nas próximas seções.

A.4.1 Coordenadas homogêneas

Um ponto \mathbf{p} do \mathbb{R}^3 no \mathbb{P}^3 é representado em coordenadas homogêneas, na forma canônica como:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{R}^3 \Leftrightarrow \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \in \mathbb{P}^3 \quad (\text{A.80})$$

Assim como foi visto anteriormente, os pontos são definidos em função de um fator de escala, portanto pode-se fazer $\bar{x} = \bar{t}x$, $\bar{y} = \bar{t}y$, $\bar{z} = \bar{t}z$:

$$\mathbf{p} = \begin{bmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \\ \bar{t} \end{bmatrix} \in \mathbb{P}^3 \quad (\text{A.81})$$

A.4.2 O plano no \mathbb{P}^3

Um plano π no \mathbb{P}^3 é representado por um vetor de coordenadas:

$$\pi = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} \quad (\text{A.82})$$

A dualidade que existe no \mathbb{P}^2 entre a reta e o ponto existe no \mathbb{P}^3 entre o plano e o ponto. Por exemplo, a expressão do plano que passa pelo ponto é:

$$\pi^T \mathbf{p} = 0 \quad (\text{A.83})$$

De forma dual, (A.83) também pode ser vista como a expressão do ponto que está sobre o plano.

A.4.3 O plano no infinito

O plano formado pelos pontos de coordenada $\bar{t} = 0$ é chamado *plano no infinito* ou *plano ideal* (BIRCHFIELD, 1998). Ele é representado por:

$$\pi_\infty = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \bar{t} \end{bmatrix} \quad (\text{A.84})$$

De forma análoga ao que acontece no \mathbb{P}^1 e no \mathbb{P}^2 , esta terminologia vem do fato de que no espaço Euclidiano, os pontos correspondentes (conforme (A.80)) para $\bar{t} \rightarrow 0$ vão para o infinito seguindo a direção da reta que os une à origem.

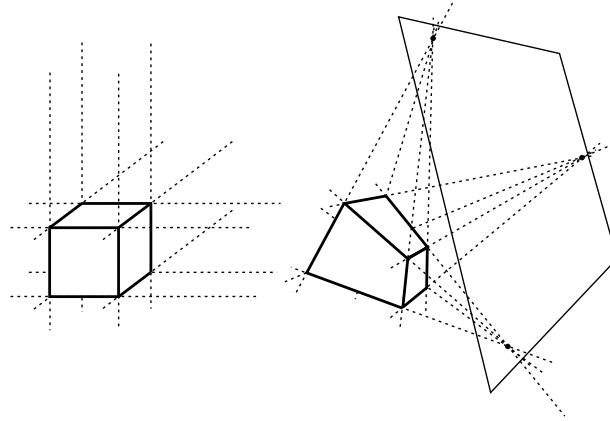


Figura A.8: No espaço projetivo algumas propriedades geométricas são perdidas.

Ou seja, pode-se pensar nos pontos que estão no π_∞ como direções no espaço \mathbb{R}^3 (nossa noção de “ponto de fuga”). Uma colineação qualquer, na sua forma mais genérica, pode eventualmente mudar a posição do plano no infinito (POLLEFEYS, 2000) (como mostrado na figura A.8).

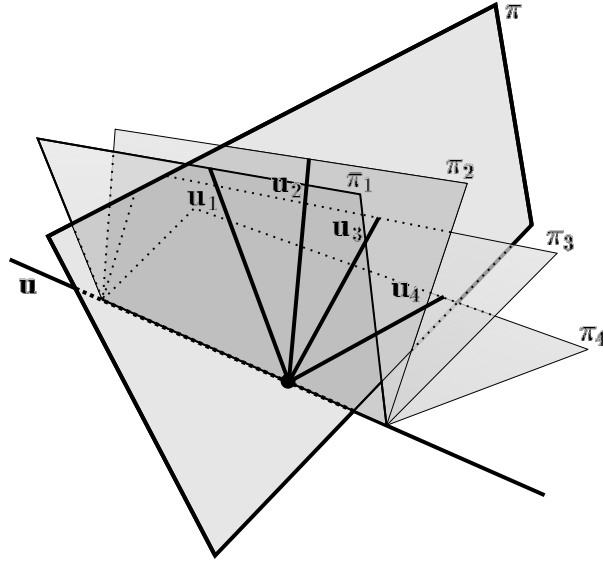


Figura A.9: razão cruzada entre quatro planos.

A.4.4 Razão cruzada entre quatro planos

Dados quatro planos π_1, π_2, π_3 e π_4 do \mathbb{P}^3 que se interceptam na reta \mathbf{u} , sua razão cruzada é definida como sendo a razão cruzada de suas quatro respectivas retas de intersecção com qualquer plano π (MOHR; TRIGGS, 1996) (vide figura A.9):

$$\{\pi_1, \pi_2; \pi_3, \pi_4\} = \{\mathbf{u}_1, \mathbf{u}_2; \mathbf{u}_3, \mathbf{u}_4\} \quad (\text{A.85})$$

A.4.5 Lápis de planos

Um *lápiz de planos* é uma estrutura análoga ao lápis de retas do \mathbb{P}^2 : é o conjunto de todos os planos que se interceptam em uma dada reta. Esta estrutura também é um espaço projetivo unidimensional, análogo a \mathbb{P}^1 .

A.4.6 Colineação no \mathbb{P}^3

Colineações no \mathbb{P}^3 são definidas por matrizes 4×4 inversíveis e sempre em função de um fator de escala. Colineações no espaço projetivo tridimensional transformam pontos em pontos, retas em retas, planos em planos e lápis de planos em lápis de planos, sempre preservando as razões cruzadas.

A.4.7 Transformação de similaridade

Uma transformação de similaridade no espaço \mathbb{P}^3 é uma transformação que aplica uma rotação, uma escala e um deslocamento. Na representação Euclideana:

$$\begin{bmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \end{bmatrix} = \sigma \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} \quad (\text{A.86})$$

Na representação em coordenadas homogêneas a expressão fica:

$$\mathbf{p}' = \begin{bmatrix} \sigma \mathbf{R}_{(3 \times 3)} & \mathbf{t}_{(3 \times 1)} \\ \mathbf{0}_{(1 \times 3)} & 1 \end{bmatrix} \mathbf{p} = \begin{bmatrix} \sigma r_{11} & \sigma r_{12} & \sigma r_{13} & t_1 \\ \sigma r_{21} & \sigma r_{22} & \sigma r_{23} & t_2 \\ \sigma r_{31} & \sigma r_{32} & \sigma r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{p} \quad (\text{A.87})$$

Os elementos r_{ij} formam a matriz de rotação. Uma matriz de rotação genérica pode ser escrita como combinação das rotações nos três eixos, ou seja:

$$\mathbf{R}_1 = \begin{bmatrix} \sin \theta_1 & \cos \theta_1 & 0 \\ \cos \theta_1 & \sin \theta_1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (\text{A.88})$$

$$\mathbf{R}_2 = \begin{bmatrix} \sin \theta_2 & 0 & \cos \theta_2 \\ 0 & 1 & 0 \\ \cos \theta_2 & 0 & \sin \theta_2 \end{bmatrix} \quad (\text{A.89})$$

$$\mathbf{R}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sin \theta_3 & \cos \theta_3 \\ 0 & \cos \theta_3 & \sin \theta_3 \end{bmatrix} \quad (\text{A.90})$$

$$\mathbf{R} = \mathbf{R}_1 \cdot \mathbf{R}_2 \cdot \mathbf{R}_3 \quad (\text{A.91})$$

$$= \begin{bmatrix} \sin \theta_1 \cos \theta_2 & \sin \theta_1 \sin \theta_2 \sin \theta_3 + \cos \theta_1 \cos \theta_3 & \sin \theta_1 \sin \theta_2 \cos \theta_3 - \cos \theta_1 \sin \theta_3 \\ \cos \theta_1 \cos \theta_2 & \cos \theta_1 \sin \theta_2 \sin \theta_3 - \sin \theta_1 \cos \theta_3 & \cos \theta_1 \sin \theta_2 \cos \theta_3 + \sin \theta_1 \sin \theta_3 \\ -\sin \theta_2 & \cos \theta_2 \sin \theta_3 & \cos \theta_2 \cos \theta_3 \end{bmatrix} \quad (\text{A.92})$$

usaremos aqui a notação mais enxuta s_n para $\sin \theta_n$ e c_n para $\cos \theta_n$. De (A.92) tem-se portanto que:

$$\mathbf{r}_1 = \begin{bmatrix} r_{11} & r_{12} & r_{13} \end{bmatrix} \quad (\text{A.93})$$

$$= \begin{bmatrix} s_1 s_2 & s_1 s_2 s_3 + c_1 c_3 & s_1 s_2 c_3 - c_1 s_3 \end{bmatrix} \quad (\text{A.94})$$

$$\mathbf{r}_2 = \begin{bmatrix} r_{21} & r_{22} & r_{23} \end{bmatrix} \quad (\text{A.95})$$

$$= \begin{bmatrix} c_1 c_2 & c_1 s_2 s_3 + s_1 c_3 & c_1 s_2 c_3 - s_1 s_3 \end{bmatrix} \quad (\text{A.96})$$

$$\mathbf{r}_3 = \begin{bmatrix} r_{31} & r_{32} & r_{33} \end{bmatrix} \quad (\text{A.97})$$

$$= \begin{bmatrix} -s_2 & c_2 s_3 & c_2 c_3 \end{bmatrix} \quad (\text{A.98})$$

A soma do quadrado dos elementos em uma determinada linha ou coluna sempre é unitária, isto é:

$$\sum_{k=1}^3 r_{ki}^2 = \sum_{k=1}^3 r_{ik}^2 = 1 \quad \text{onde } i \in \{1, 2, 3\} \quad (\text{A.99})$$

e o produto interno de uma linha com a terceira linha, ou de uma coluna com a terceira coluna é

$$\sum_{k=1}^3 r_{ki} r_{k3} = \sum_{k=1}^3 r_{ik} r_{3k} = 0 \quad \text{para } i \neq 3 \in \{1, 2\} \quad (\text{A.100})$$

A matriz \mathbf{R} cujos elementos são os r_{ij} é uma matriz de rotação se e somente se:

$$\mathbf{R}^T \mathbf{R} = \mathbf{R} \mathbf{R}^T = \mathbf{I}_3 \quad \text{e} \quad |\mathbf{R}| = 1 \quad (\text{A.101})$$

sendo \mathbf{I}_3 a matriz identidade 3×3 .

Este tipo de transformação conta com 7 graus de liberdade: 3 para rotação, 3 para translação e 1 para escala. Esta transformação mantém invariantes os comprimentos relativos e os ângulos.

Além disso, uma transformação de similaridade também mantém inalterada uma entidade chamada *cônica absoluta* que será definida na seção A.4.9.

A.4.8 Quádricas

No espaço projetivo tridimensional \mathbb{P}^3 as chamadas *quádricas* fazem o mesmo papel que as cônicas no \mathbb{P}^2 (FAUGERAS, 1993).

Uma quádrica é uma superfície do \mathbb{P}^3 definida pelo lugar dos pontos que satisfazem:

$$\mathbf{p}\mathbf{Q}^T\mathbf{p} = 0 \quad \forall \quad \mathbf{Q} = \mathbf{Q}^T \in \mathbb{R}^{4 \times 4} \quad (\text{A.102})$$

Ou seja:

$$\begin{bmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \\ \bar{t} \end{bmatrix}^T \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{12} & c_{22} & c_{23} & c_{24} \\ c_{13} & c_{23} & c_{33} & c_{34} \\ c_{14} & c_{24} & c_{34} & c_{44} \end{bmatrix} \begin{bmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \\ \bar{t} \end{bmatrix} = 0 \quad (\text{A.103})$$

A.4.9 Cônica absoluta

A *cônica absoluta*, denotada Ω , é a intersecção da quádrica

$$\mathbf{p}^T \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{p} = 0 \quad (\text{A.104})$$

com o plano π_∞ . A cônica Ω é uma entidade que permanece inalterada mediante uma transformação de similaridade. O conceito de cônica absoluta é abstrato. Esta cônica pode ser vista como um círculo imaginário de raio $i = \sqrt{-1}$ sobre o plano π_∞ (POLLEFEYS, 2000).

Um ponto $\mathbf{p} = [\bar{x} \ \bar{y} \ \bar{z} \ \bar{t}]^T$ é um ponto sobre a cônica se suas coordenadas homogêneas satisfazem:

$$\bar{x}^2 + \bar{y}^2 + \bar{z}^2 = 0 \quad \text{e} \quad \bar{t} = 0 \quad (\text{A.105})$$

Isto implica a existência de pontos imaginários. Parametrizando-se um ponto $\tilde{\mathbf{p}}_\infty = [\bar{x} \ \bar{y} \ \bar{z} \ 0]^T$ localizado sobre o plano π_∞ como sendo $\mathbf{p} = [\bar{x} \ \bar{y} \ \bar{z}]^T$, isto é, vendo o plano no infinito π_∞ como um espaço projetivo \mathbb{P}^2 , então a cônica absoluta, agora notada como ω_∞ pode ser dada pela expressão:

$$\mathbf{p}^T \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{p} = 0 \quad (\text{A.106})$$