

UNIVERSIDADE DE COIMBRA
FACULDADE DE CIÊNCIAS E TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELECTROTÉCNICA

Reconstrução Tridimensional Utilizando Visão Dinâmica

JORGE MANUEL MIRANDA DIAS

COIMBRA - 1994

**Reconstrução Tridimensional
Utilizando Visão Dinâmica**

Jorge Manuel Miranda Dias

*Departamento de Engenharia Electrotécnica
Faculdade de Ciências e Tecnologia
Universidade de Coimbra
1994*

Resumo

A construção de "robots" com autonomia e com possibilidade de operarem em ambientes não estruturados, é um grande desafio científico e exige o desenvolvimento de sistemas artificiais de percepção. A tese aqui apresentada, insere-se neste domínio e mais em concreto, no domínio da visão por computador. Nela são descritos métodos para determinar a estrutura tridimensional do universo circundante de um "robot" utilizando imagens fornecidas por câmaras.

Neste trabalho foram desenvolvidas e aperfeiçoadas técnicas para reconstrução da estrutura tridimensional de uma cena. A estrutura tridimensional é representada através de um conjunto de primitivas geométricas da qual fazem parte pontos, rectas ou planos. Para a reconstrução dessas primitivas geométricas são captadas imagens em diferentes posições no espaço ou com diferentes focagens. Estas técnicas diferem das técnicas "tradicionais" que utilizam imagens captadas com as câmaras sempre estáticas. Com o controlo da posição das câmaras no espaço ou dos parâmetros ópticos das lentes é possível estabelecer um conjunto de técnicas de reconstrução tridimensional.

As técnicas apresentadas determinam os valores dos parâmetros das primitivas de forma absoluta ou relativa utilizando o controlo da focagem ou a relação geométrica entre as posições das câmaras. As experiências realizadas mostraram a viabilidade das técnicas, tendo sido desenvolvido um sistema de visão em torno destes princípios. Esse sistema, designado VARMA, é apresentado na parte final do trabalho e é uma base experimental para o desenvolvimento de algoritmos que envolvem a movimentação das câmaras na "percepção" do espaço tridimensional.

**3D Reconstruction
Using
Dynamic Computer Vision**

Jorge Manuel Miranda Dias

*Departamento de Engenharia Electrotécnica
Faculdade de Ciências e Tecnologia
Universidade de Coimbra
1994*

Abstract

The development of a new class of robots, with high-levels of autonomy and possibility of operating in unstructured environment, is a great scientific challenge. This development requires new techniques for perception using the information provided by sensors. This thesis describes techniques to recover the three-dimensional structure of a scene by using images. The system acquires images around the robot that are used to recover the three-dimensional structure of the scene.

Points, lines and planes are considered the basic primitives for the representation of the three-dimensional structure. In this work different computer vision techniques to estimate the mathematical parameters of these primitives were developed and enhanced. All of these techniques explore the principle of "adaptation" by capturing the images in different points of the scene and/or with different focus. These techniques are different from the "classical" methods that use images captured with cameras always in a static position. The experimental results obtained with these techniques are promising and show that "action" is also important for the process of "perception."

The developed techniques acquire images with the cameras in different configurations and positions. The techniques explore the stereo configuration, the acquisition of images during an execution of a movement trajectory and techniques based on lenses focusing. These techniques encouraged us to develop a new vision system where different parameters, such as the camera's position and the lens parameters, can be controlled. The last chapter of the text describes this vision system, called VARMA, developed for testing new algorithms using the principles of "action" associated with "perception."

*Dissertação submetida à Universidade de Coimbra
para obtenção do grau de doutor em Engenharia
Electrotécnica, especialidade de Instrumentação e
Controlo*

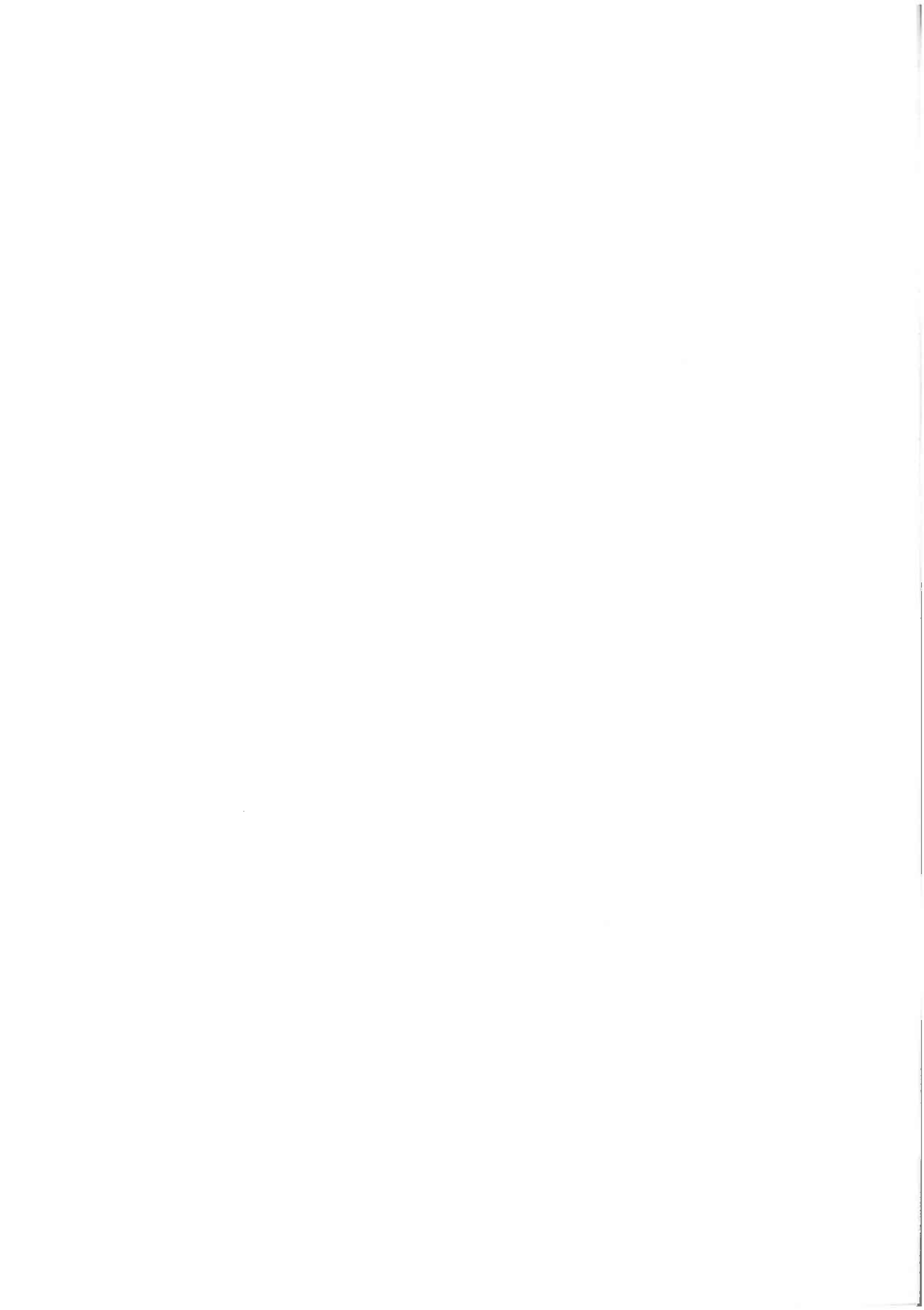
ÍNDICE

1. Introdução.....	1
1.1 A Visão por Computador e a Determinação da Profundidade	3
1.2 Trabalhos Anteriores	5
1.3 A Visão por Computador e a sua Integração na Robótica.....	8
1.4 Descrição Resumida do Trabalho	10
2. Modelos para a Formação da Imagem e Representação das	
 Primitivas	13
2.1 Introdução	13
2.2. Modelos da formação da imagem	13
2.2.1. Projecção em perspectiva.....	14
2.2.2. Modelos de Projecção Alternativos	16
2.2.3. Lentes e Focagem	19
2.2.4 Formulação matemática do modelo	23
2.3 Imagens de cor.....	26
2.4. Representação de Primitivas Bidimensionais	30
2.4.1 Pontos	30
2.4.2 Rectas.....	30
2.5 Representação de Primitivas Tridimensionais.....	32
2.5.1 Pontos	32
2.5.2 Rectas.....	32
2.5.3 Planos.....	34
2.5.4 Equação Geral de Superfícies Quádricas	36
2.6. Invariantes das Representações Tridimensionais	37
3. Geometria do sistema de visão	41
3.1 Introdução	41
3.2 Geometria do Movimento do Sistema de Visão	41
Introdução	41
Transformações utilizadas no sistema experimental	43
Cinemática do sistema de visão	45
Velocidade do sistema de visão.....	46

3.3	Calibração do Sistema de Visão.....	51
3.4	Efeito do Movimento das Câmaras	55
3.4.1	Pontos.....	55
3.4.2	Rectas	55
3.4.3	Planos	56
4.	Detecção de primitivas em imagens - O caso de segmentos de recta.....	57
4.1.	Introdução.....	57
4.2.	Detecção de Pontos Característicos	57
4.3	Detecção de Segmentos de Recta 2D.....	59
4.3.1.	Detecção de Arestas.....	61
4.3.2	Contornos	68
4.3.3	Segmentação dos Contornos	70
	Teste do Desvio Angular	70
	Teste da orientação média do contorno.....	72
	Transformada de Hough	72
4.3.4	Parâmetros Geométricos dos Segmentos de Recta.....	72
4.4	Grupos de Segmentos.....	76
5.	Segmentação de imagens utilizando focagem.....	81
5.1.	Introdução.....	81
5.2.	Medida da Qualidade da Focagem	82
5.3.	Estudo de um Algoritmo para Controlo da Focagem	88
5.4.	Calibração da Focagem.....	89
5.5	Segmentação da Imagem Utilizando Focagem	91
5.5.1.	Cálculo da profundidade	92
5.5.2.	Segmentação	92
5.6.	Segmentação de Imagens Utilizando Cooperação entre Focagem e Cor	97
5.7	Pormenores a Investigar	105
5.8	Relação com as técnicas de medida de desfocagem	106
6.	Estereovisão Dinâmica - Estudo de um processo multiocular.....	109
6.1	Introdução.....	109
6.2	Geometria da Estereovisão	112
6.2.1	Pontos Epipolares e Rectas Epipolares.....	112
6.2.2	Rectificação de Imagens.....	116
6.2.3	Profundidade Através da Visão Estereo	120
6.3	Correspondência.....	126
6.3.1	Técnicas Anteriormente Propostas	126
	Restrição da Recta Epipolar	126
	Restrições Locais	127
	Restrição da Correspondência Única	128
	Restrição de Ordem	129
	Restrição de Continuidade.....	130
	Disparidade máxima e mínima	130
	Gradiente de disparidade	132
6.3.2	Novas Restrições pela Exploração dos Graus de Liberdade do Sistema de Visão	133
	Utilização de um Processo Multiocular.....	133
	Cruzamento das Listas de Hipóteses.....	134

Restrição local global.....	135
Restrição geométrica qualitativa.....	135
Restrição baseada na focagem.....	137
6.3.3 Validação de Hipóteses.....	137
Validação com base na vizinhança.....	137
Validação com base na correlação e na restrição de ordem.....	138
Correlação.....	138
6.4 Detalhes.....	140
6.4.1 Efeitos Devido ao Movimento das Câmaras.....	140
O caso de pontos.....	140
O caso de rectas.....	140
O caso de planos.....	141
Restrições ao movimento.....	142
6.4.2 Quantificação do erro na correspondência.....	143
6.5 Resultados Experimentais.....	145
6.5.1 Algoritmos.....	145
Algoritmo de visão estereo biocular.....	145
Algoritmo estereo multiocular.....	146
Pormenores práticos.....	148
6.5.2 Resultados com Imagens Reais.....	149
Resultados da aplicação do algoritmo de estereovisão.....	151
6.5.3 Resultados Comparativos.....	154
6.5.4 Programas Desenvolvidos.....	158
7. Mapas Volumétricos - Experiências com base na focagem.....	161
7.1 Introdução.....	161
7.2 Definição de Mapa Volumétrico.....	162
Operações de Actualização do Mapa.....	163
Zonas do mapa onde se deverá actualizar a informação.....	166
7.3 Resultados Experimentais com Base na Focagem.....	167
8. Desenvolvimento de Um Sistema de Visão Activa.....	175
8.1 Introdução.....	175
8.2 Desenvolvimento de um Sistema de Visão Activa.....	177
8.2.1 Introdução.....	177
8.2.2 Vantagens de um sistema de Visão Activa.....	179
8.2.3 Uma Base Experimental para Visão Activa.....	180
8.2.4 Estrutura Mecânica.....	181
8.2.5 "Hardware" de Controlo do Sistema Mecânico.....	182
8.2.6 "Software" de Controlo.....	185
8.2.7 Sistemas de coordenadas.....	187
8.3 Fixação de Objectos Estacionários.....	188
A fixação.....	188
8.3.1 Tempo até colisão através de vergência.....	190
8.3.2 Distância ao ponto de fixação.....	191
8.4 Explorando as Imagens de Movimento.....	193
8.4.1 Introdução.....	194
Campo de Movimento.....	195
8.4.2 Efeito do Movimento da Câmara.....	197
Efeito de Rotação.....	198

Efeito de Translação	200
Invariância do Ponto de Fuga de Uma Recta.....	202
8.4.3 Determinação da Profundidade	203
Utilizando o FDC/FDE.....	203
Utilizando fluxo óptico normal ("normal flow")	205
Mapa de profundidade relativo	208
Experiências.....	210
8.5 Outros aspectos da visão activa	213
9. Conclusões e Pesquisa futura	215
Anexo A. Estimação de parâmetros.....	A.1
Anexo B. Testes estatísticos	B.1
Anexo C. Combinação de dados	C.1
Bibliografia	



CAPÍTULO 1

Introdução

Este primeiro capítulo introduz um trabalho na área da visão por computador descrito ao longo dos capítulos seguintes. O trabalho centra-se em torno do desenvolvimento de técnicas para determinação da tridimensionalidade de cenas a partir das suas imagens. Uma área, entre muitas outras, na qual estas técnicas poderão ser exploradas com êxito é a robótica.

A construção de "robots" com autonomia e com possibilidade de operarem em ambientes não estruturados, é um grande desafio científico e exige o desenvolvimento de sistemas artificiais de percepção. Com a possibilidade de percepção do local onde se movimenta, o "robot" deixará de estar limitado à utilização em locais de dimensões conhecidas e nos quais a sequência de acontecimentos está previamente estabelecida. Para alargar as aplicações da robótica além destes locais é necessário o desenvolvimento de técnicas que dotem os "robots" de capacidade de movimentação e actuação em universos cujas características físicas não são conhecidas "a priori". Para além disso, a boa percepção do universo que rodeia o "robot" aumenta a sua capacidade de executar reacções adequadas no caso de situações inesperadas. Os sistemas de visão são uns dos sistemas sensoriais actualmente disponíveis para aplicação em robótica. Estes, baseando-se em imagens, poderão ser uma fonte de informação acerca das características físicas do universo em torno do "robot", pois as imagens dão muita informação sobre o local onde foram adquiridas.

A visão tem um papel relevante na percepção humana e por isso tem tomado a atenção de muitos investigadores e estudiosos ao longo da História. Neste século e com o desenvolvimento actual da tecnologia dos sensores e dos computadores tem sido investigada a possibilidade de simular uma capacidade de percepção idêntica através de máquinas. Todavia, o processamento de informação visual apresenta particularidades difíceis de contornar devido à dificuldade em explicar de forma objectiva e através de modelos, a interpretação das imagens. Grande parte desta dificuldade reside no facto das imagens serem representações bidimensionais de cenas tridimensionais, dando uma dificuldade acrescida ao

problema da percepção. É relativamente fácil conseguir exemplos da complexidade da percepção e interpretação das imagens. Um desses exemplos é a figura 1.1 que mostra duas fotografias de uma réplica de uma das fachadas da Universidade de Coimbra.

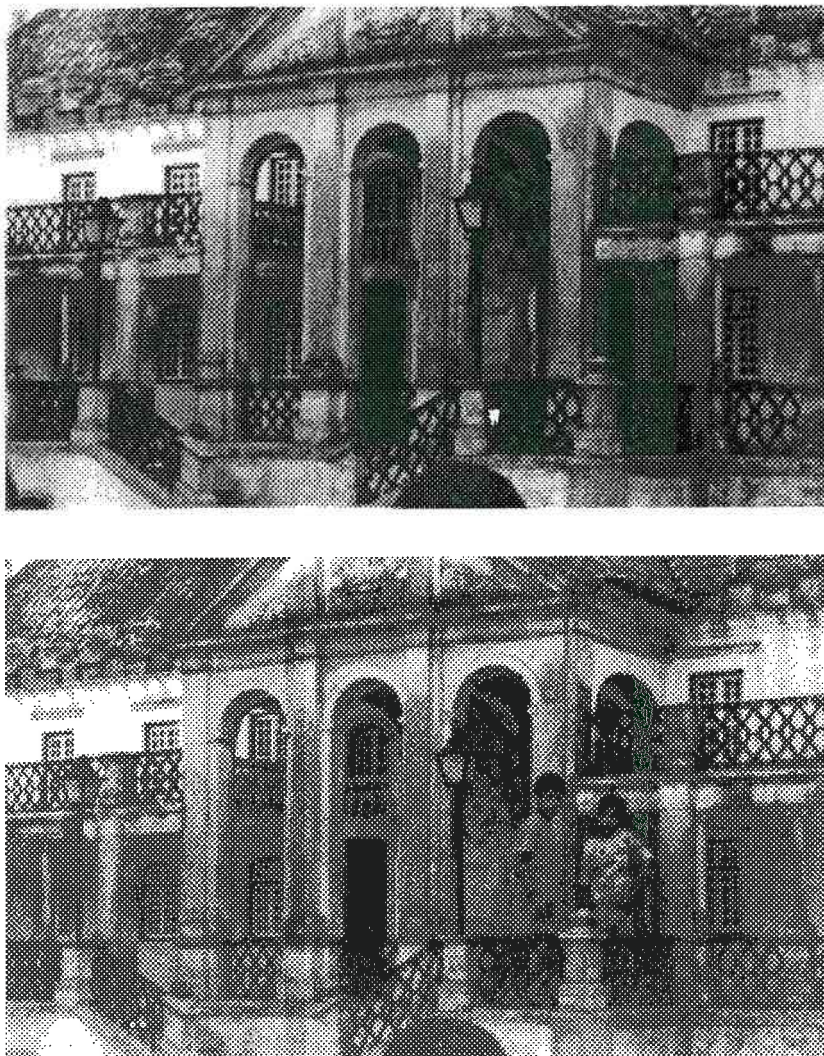


Figura 1.1 - As fotografias exemplificam que a noção de dimensionalidade retirada a partir de uma imagem é dependente dos conhecimentos que temos sobre os elementos da cena.

Pela primeira fotografia e com base nos conhecimentos sobre a dimensionalidade real do edifício é possível estabelecer estimativas para as suas dimensões. No entanto se tentarmos realizar as mesmas estimativas utilizando a segunda fotografia é possível que essas estimativas sejam completamente diferentes. Este exemplo mostra que a inferência sobre a tridimensionalidade dos objectos a partir de imagens pode depender dos conhecimentos que temos sobre o mundo que nos rodeia. De uma forma mais geral, podemos concluir que a percepção da tridimensionalidade é inerentemente um processo complexo e a figura 1.1 é disso um exemplo.

1.1 A Visão por Computador e a Determinação da Profundidade

O sistema visual humano utiliza diversas formas de processamento das imagens da retina para distinguir a posição relativa dos objectos [Gibson 50]. Essas maneiras são por vezes designadas por "cues" na literatura sobre visão por computador e podem também envolver os efeitos gerados pela movimentação do sistema visual. A terminologia "cue" exprime a forma como se realiza a percepção visual, segundo a abordagem computacional da teoria da percepção [Cutting 86,pag40] [Ullman 79] [McAllister 92] [Marr 82]. Segundo esta teoria as "cues" exprimem entidades que são responsáveis por indicar ao sistema visual quais os processamentos a executar sobre as imagens. As "cues" para percepção tridimensional têm por objectivo dar uma estimativa sobre a forma dos objectos, fazendo inferências sistemáticas e utilizando alguma forma de computação ou lembrando alguma relação através de experiências adquiridas. Existem várias "cues" que permitem aos humanos ultrapassar a dificuldade da dedução da estrutura tridimensional do espaço e dos objectos com base na informação das imagens. Algumas destas são: o movimento relativo, o movimento projectado na imagem, a disparidade binocular provocada pelas imagens estereo, o efeito gerado pelas sombras, os contornos dos objectos, a orientação de textura, as reflexões e a acomodação visual. Marr, no seu livro, sugere que estas "cues" podem ser processadas independentemente e correspondem a módulos que, a partir das imagens, fornecem representações da profundidade a que os objectos se encontram [Marr 82]. Estas representações são por vezes designadas por mapas de profundidade. Cada um destes mapas poderá ser posteriormente combinado numa representação que Marr designou por "2½ sketch".

Diferentes descritores podem ser utilizados na representação "2½ sketch". O descritor pode representar a distância, orientação da superfície visível, curvatura ou forma dos objectos na cena. A escolha do descritor tridimensional depende da acção ou o fim para o qual se destina a informação tridimensional. Por exemplo, um mapa de profundidade correspondente a pontos ou arestas pode ser suficiente para a manipulação de objectos ou navegação, enquanto que a curvatura de uma superfície de um objecto pode ser útil para o seu reconhecimento. Na figura 1.2 é dado um diagrama que mostra como a estrutura tridimensional de uma cena pode ser representada com diferentes níveis de abstracção, a partir da informação fornecida por imagens. Cada um das "cues" que permite a percepção da tridimensionalidade, gera diferentes representações tridimensionais. Estas representações são classificadas em representações de baixa ordem. As representações de ordem mais alta podem ser obtidos através da combinação e interacção de várias representações de ordem inferior.

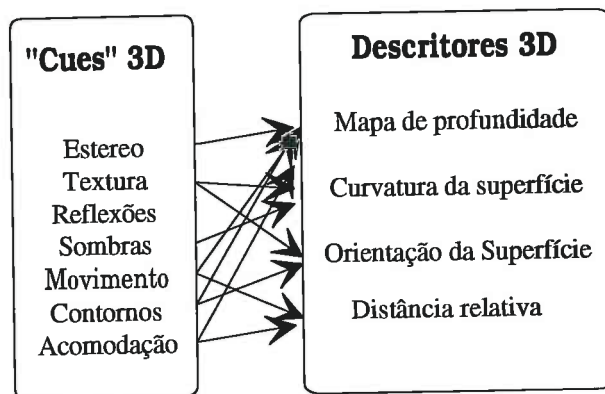


Figura 1.2 - A estrutura tridimensional de uma cena a partir de imagens pode ser representada com diferentes níveis de abstracção. Cada "cue", individualmente, gera uma representação tridimensional designada por representação de baixa ordem. A partir destas e da sua adequada combinação, podemos obter representações designadas por representações de ordem superior.

A abordagem computacional do problema da percepção visual humana, tem influenciado os estudos mais recentes sobre a psicofísica e visão por computador. Nos estudos de percepção da tridimensionalidade realizados na área da psicofísica e da visão por computador têm sido explorados os diferentes tipos de "cues". Grande parte deles fazem-no de forma isolada e sem interacção entre si. Entre estes estudos, existem trabalhos relevantes na utilização de imagens estereo [Julesz 71][Marr 76, 79][Mayhew 81][Prazdny 85], na análise do efeito das sombras [Black 85][Ikeuchi 81][Mingolla 86][Pentland 84], na percepção da profundidade a partir de textura [Aloimonos 85][Bajcsy 76][Pentland 86][Witkin 81], na percepção da profundidade através de arestas e contornos [Barrow 81][Stevens 81][Stevens 87], na determinação da estrutura tridimensional a partir de movimento [Longuet-Higgins 81][Koenderink 86][Landy 87][Ullman 79,84] e da utilização da acomodação visual ou focagem [Pentland 85][Krotkov 88,89]. Na grande maioria destes estudos propõem-se algoritmos que apresentam bons resultados quando são aplicados a imagens sintéticas mas são menos estáveis quando são aplicados a imagens naturais [Aloimonos 89]. Este facto, aliado a uma ausência total de uma teoria geral que uniformize a abordagem computacional ao problema da percepção visual, justifica uma contínua investigação neste domínio. Entre alguns dos temas importantes para o desenvolvimento do conhecimento neste domínio, encontram-se a detecção de mudanças de intensidade nas imagens, a detecção e determinação de movimento nas imagens, a determinação da forma e da estrutura tridimensional. Mais recentemente, tem sido sugerido o desenvolvimento de técnicas que envolvam a cooperação entre vários processos visuais e o movimento dos sensores.

1.2 Trabalhos Anteriores

As variações fortes de intensidade numa imagem têm um papel fundamental na percepção da cena, tal como foi demonstrado por Marr [Marr 78]. A detecção destas variações é normalmente feita através de filtros que implementam diferenças espaciais de intensidade. Grande parte desses filtros são representados por máscaras que são convolucionadas com a imagem. Estas máscaras representam operadores que, de forma aproximada, geram resultados correspondentes às primeiras e segundas derivadas direccionais [Ballard 82] [Nalwa 86] [Haralick 84] [Marr 80]. Estes operadores só apresentam bons resultados para cenas em que os objectos apresentam contornos ou gamas de variação da intensidade luminosa com baixa curvatura. No entanto estes operadores podem ser estudados de modo a darem uma resposta óptima para determinadas condições da imagem [Canny 85][Derich 89]. Os coeficientes destes filtros são determinados para fornecerem uma resposta óptima com base em critérios previamente definidos, de localização das arestas, da amplitude da variação e da resposta múltipla do filtro. Recentemente já é possível realizar este tipo de filtros com um baixo custo computacional [Derich 89][Monga 89].

Os vértices de um objecto são também locais de grande interesse para a percepção visual pois representam locais de intersecção das suas arestas. No entanto estes pontos são bastante difíceis de obter na imagem de forma estável e independente das condições em que a imagem é adquirida, tal como se pode depreender dos trabalhos de Moravec, Gong e Harris [Moravec 77] [Canny 85] [Gong 88] [Canny 85][Haralick 84] [Harris 88].

A estereovisão permite a obtenção da tridimensionalidade de uma forma imediata se o problema da correspondência entre zonas nas duas imagens estereo for resolvido. Nos estudos realizados por Marr com imagens estereo este problema ficou perfeitamente definido [Marr 76][Marr 79] [Grimson 81]. Têm sido propostas várias soluções para este problema que se distinguem pelo modo como realizam o teste de semelhança entre as zonas nas duas imagens e determinam zonas que poderão ter correspondência. Deste modo, a correspondência pode ser realizada com base em zonas de grande variação da intensidade luminosa e correspondentes a arestas ou contornos [Pollard 85], ou utilizando regiões com características idênticas [Wrobel-Dautcourt 88]. O problema da correspondência pode também ser facilitado pela utilização de três imagens captadas com câmaras posicionadas no espaço, segundo uma geometria predefinida tal como o fazem Ayache, Ohta, Peitikäimen ou Yachida [Ayache 87a] [Ohta 85] [Peitikäimen 86] [Yachida 86]. Com uma geometria predefinida das câmaras é possível construir restrições geométricas muito fortes, aos quais devem obedecer as posições das zonas correspondentes nas imagens.

A informação sobre a tridimensionalidade de uma cena pode ser obtida a partir do movimento aparente numa sequência de imagens. A este movimento aparente dá-se o nome de "optical flow", sendo a sua determinação um processamento anterior à determinação da estrutura tridimensional. Alguns dos algoritmos mais relevantes para a sua determinação foram propostos por Nagel, Horn, Hildreth e Buxton [Nagel 83] [Horn 81] [Hildreth 84] [Buxton 83]. Martin, Harris e Navab propõem algoritmos para determinação de estrutura tridimensional pressupondo que o "optical flow" foi previamente determinado [Martin 88] [Harris 90] [Navab 90]. A determinação das formas através da análise do movimento e da textura foi estudada por Kanatani e Aloimonos [Kanatani 84] [Aloimonos 85]. Posteriormente Aloimonos apresentou resultados da combinação das duas técnicas para a obtenção da forma, demonstrando que a aplicação simultânea das duas técnicas pode conduzir a maior estabilidade dos resultados [Aloimonos 89].

Analisando os trabalhos publicados em visão por computador, verificamos que o esforço de investigação se concentra em tópicos muito bem definidos e resultando em diferentes tipos de descritores tridimensionais. Entre esses tópicos constam o estudo da detecção das mudanças de intensidade nas imagens em diferentes tipos de cenas, a estereovisão, a focagem, a análise cromática das imagens, a detecção e determinação do movimento, etc... . Muitos destes tópicos relacionam-se directamente com a descrição feita anteriormente sobre os diferentes tipos de "cues" tridimensionais e correspondendo a diferentes métodos de calcular os descritores tridimensionais. Esses métodos realizam módulos computacionais que geram uma saída *Y* a partir de uma entrada *X*. Nem todos estes módulos operam directamente na imagem, mas sim em representações de informação intermédias e previamente calculadas por outros módulos.

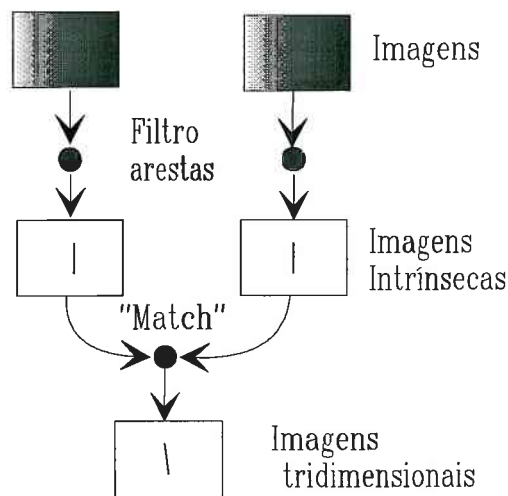


Figura 1.3 - Hierarquia do processamento visual no caso de imagens estereo [Marr 78] [Barrow 78].

Uma característica dos resultados gerados por alguns destes módulos é representarem, por vezes, campos muito densos de informação no domínio da imagem bidimensional, sendo designados imagens intrínsecas. Esta informação densa tem que ser posteriormente segmentada e agrupada em superfícies coerentes ou em partes de/ou objectos que serão mais tarde processados por outros estágios. Esta abordagem computacional sugere uma hierarquia de processamento idêntica à que foi proposta por Marr [Marr 78] [Barrow 78] e ilustrada na figura 1.3.

As imagens intrínsecas formam uma representação intermédia útil e facilitam a tarefa de processamento de alto nível. A sua descrição fornece uma forma mais estável do que as imagens de intensidades, visto serem menos sensíveis a factores de distúrbio na cena, como por exemplo a mudanças de iluminação. Estas representações intermédias permitem também estabelecer um esquema para integração da informação oriundas de múltiplos módulos que processam a imagem.

A combinação de diversos módulos permite uma melhor estabilidade nos resultados representados pelas imagens intrínsecas face ao ruído nas imagens de intensidade. As combinações possíveis entre módulos são variadas tal como se mostra na figura 1.4, existindo alguma experiência sobre combinação. Assim, Waxman, Richards, Huang e Milenkovic combinam estereovisão e movimento, Grimson combina o efeito das sombras e a estereovisão e Abbot combina focagem e estereovisão [Waxman 87] [Richards 86] [Huang 85] [Milenkovic 85] [Grimson 84] [Abbot 88].

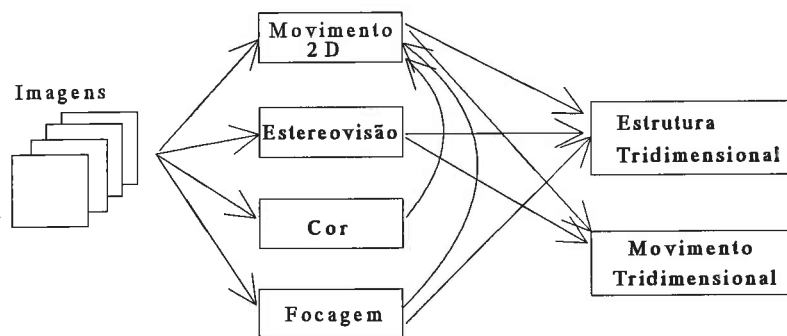


Figura 1.4 - Algumas formas de combinação de informação fornecida por diferentes módulos visuais.

Na visão por computador assume-se que o universo tridimensional pode ser representado através de pontos, contornos ou superfícies. Os pontos são primitivas que têm a vantagem de serem comuns em cenas de universo estruturados ou em cenas naturais. Para além disso são as primitivas mais fáceis de integrar em representações globais e que envolvem processos dinâmicos. Por exemplo Brady considera que os pontos são representações fundamentais nos mapas tridimensionais utilizados em robótica [Brady 89]. Os contornos são formados pelos conjuntos de "pixels" que representam mudanças bruscas

de intensidade na imagem. Os contornos são primitivas menos gerais do que pontos mas definem melhor o objecto. Alternativamente, os contornos podem ser reunidos em primitivas mais completas, como os segmentos de rectas. Este tipo de primitivas foram utilizadas nos trabalhos de Ayache e Lustman [Ayache 89] [Lustman 87]. As superfícies também são primitivas importantes na representação tridimensional pois demarcam a dimensão de um sólido. No entanto, a recuperação das superfícies pode ser simplificada considerando que uma superfície pode ser representada através de pequenas pedaços de superfícies planas ou de superfícies exibindo curvatura mas representáveis por quádricas [Black 87] [Terzopoulos 86] [Grossmann 89].

Apesar do esforço de investigação que tem sido realizado na área da visão por computador os resultados não geraram numa teoria geral para a visão artificial. Os problemas relacionados com a visão artificial estão longe de estar solucionados e, com a sua aplicação em robótica, o conjunto de problemas em aberto ainda é alargado.

Com aplicações no domínio da robótica surge também o problema do tempo de computação para execução dos algoritmos. A necessidade de reacção do "robot" a situações inesperadas exige tempos de decisão curtos. Se a decisão depende dos dados do sistema de visão então o tempo de processamento do sistema de visão deverá ser ainda menor. O tempo de processamento poderá ser diminuído através de novas arquitecturas de "hardware" para processamento das imagens [Dias 88] [Sá 88].

1.3 A Visão por Computador e a sua Integração na Robótica

A robótica é um domínio científico que envolve vários tipos de problemas tais como, o processamento de informação sensorial, a manipulação ou o planeamento das trajectórias. Estes problemas têm sido estudadas noutros ramos do conhecimento o que, por sua vez, tem mantendo a esperança de uma interacção e integração destes conhecimentos permita, um dia, construir máquinas autónomas e com capacidade de reacção a situações não previamente estudadas. Esta forma de reacção exige uma grande capacidade de coordenação da percepção no meio circundante.

Dos vários trabalhos que têm vindo a ser realizados em robótica, transparece o consenso de que uma maior evolução da robótica, depende de uma melhor utilização da informação sensorial. Esta situação leva a que o actual estado dos estudos em robótica, esteja principalmente centrado na utilização correcta dos sensores, englobando problemas ligados ao seu aperfeiçoamento e sua utilização no controlo, ao planeamento do movimento, na manipulação de peças e na integração dos diversos sistemas que fazem parte do "robot".

Em robótica, o processamento associado aos mecanismos de reacção com base na informação sensorial tem vindo a ser dividido em diferentes níveis de complexidade [Harmon

87] [Durrant-Whyte 88] [Luo 89] . Brady sugere que o nível baixo desta forma de reacção deve corresponder à capacidade de resposta controlada às mudanças existentes no ambiente que o rodeia [Brady 89]. Esta capacidade deverá assegurar a integridade do "robot" e uma resposta rápida. Neste caso a capacidade de obter dados do meio que rodeia o "robot", só envolve descrições simples do universo de trabalho. A níveis mais altos, deverá permitir a movimentação do "robot" em universos de trabalho mais complexos, incertos e eventualmente sempre em mudança.

Em robótica, também é comum classificar os sensores como internos e externos ao "robot". Os sensores internos formam o conjunto de sensores que permitem o controlo e supervisionamento do "robot". Nesta classe estão incluídos os codificadores, sensores de posição e potenciómetros que podem ser encontrados nos manipuladores e nos "robots" móveis. Os sensores externos podem ser classificados quanto à forma como realizam a medida da grandeza: os que exigem um contacto directo com o meio físico e os que não o exigem. Do conjunto de sensores externos que exigem contacto directo com o meio físico incluem-se os interruptores de contacto, sensores de tacto e de força. Os sensores de tacto ainda apresentam uma fraca resolução espacial e uma gama dinâmica muito limitada [Dario 89] e [Hollerbach 87]. Dos sensores externos também fazem parte os sensores que não envolvem o contacto directo com o meio físico. Os sensores sem contacto mais utilizados actualmente englobam os sensores de ultra-sons, infravermelhos, laser e câmaras vídeo. De toda a investigação que tem sido realizada utilizando estes sensores, a que apresenta maior volume de trabalhos é a visão por computador que utiliza a informação obtida por câmaras vídeo. No entanto os ultra-sons têm sido também muito utilizados, sobretudo devido ao seu baixo custo, robustez e pequena largura de banda dos sinais - o que permite uma electrónica de baixo custo. Actualmente existe uma grande variedade de "robots" que utilizam ultra-sons para determinar o universo envolvente e para detecção de obstáculos [Giralt 84] [Brown 86] [Elfes 87] [Richardson 88] [Cox 89].

Um problema comum aos algoritmos de interpretação dos dados sensoriais é a instabilidade do seu funcionamento nas diferentes circunstâncias em que os dados são adquiridos. Estes problemas levaram ao desenvolvimento de técnicas que utilizam vários sensores simultaneamente e de modo que a informação seja redundante ou complementar. Isto tem conduzido ao problema da fusão de dados multi-sensoriais [Durrant-Whyte 88] [Luo 89]. Durrant-Whyte explora esta ideia e inclui também o problema da integração da informação dos sensores obtida em diferentes posições [Durrant-Whyte 88].

Em robótica existem inúmeros problemas cujas soluções envolvem a inspecção, manipulação ou medida de objectos tridimensionais. Soluções para estes problemas requerem técnicas que possibilitem o fornecimento de descrições tridimensionais dos objectos. A representação, para a forma e a estrutura tridimensional, depende muito da aplicação ou do objectivo a alcançar. Assim por exemplo as formas de representação que têm vindo a ser propostas para interpretação, para manipulação de objectos ou para

navegação entre objectos estacionários, diferem a nível do detalhe e da estrutura de dados utilizada para a representação. As representações que têm vindo a ser propostas são, na sua grande parte, orientadas para uma aplicação e normalmente não asseguram mecanismos de mudança entre outros tipos de representação mais adequados para outras aplicações ou objectivos. Por este facto o estudo de representações para os objectos é um tópico ainda em aberto. No entanto, na sua forma mais comum essas representações passam por descrições através de mapas de profundidade ou de distância, que servirão depois como dados de entrada para outros módulos que se encarregarão de obter outras formas mais elaboradas de representação dos dados.

A obtenção do mapa tridimensional do universo que rodeia o "robot" ou manipulador, poderá ser feita utilizando sensores que fazem a medida directa da distância ou através de técnicas ópticas [Besl 88]. Os sensores laser e os sensores de ultra-sons são sensores caracterizados por emitirem energia, cuja interacção com os objectos da cena, é posteriormente medida por detectores. A energia é emitida de forma a ter uma distribuição predefinida no domínio espacial, espectral ou temporal e que permite obter o resultado desejado. No caso da medição de distâncias, pode ser também utilizada a diferença de tempo entre emissão e recepção de energia para obter uma medida da distância. Para ter um mapa tridimensional local, a energia emitida pelo sensor é também distribuída espacialmente. Estes sensores são considerados como *ativos* por emitirem energia que interacciona com os objectos na cena.

1.4 Descrição Resumida do Trabalho

Presentemente, o desenvolvimento tecnológico dos "robots" está fortemente condicionado ao desenvolvimento de mecanismos de percepção dos locais em que os "robots" se movimentam. Neste trabalho são apresentados métodos para obter informação acerca da estrutura tridimensional do universo circundante de um "robot" utilizando imagens. A determinação dessa informação consiste essencialmente, na determinação da distância a que se encontram os objectos nesse universo. No trabalho são exploradas vias para a determinação da distância relativa entre os objectos (profundidade) e a determinação da distância absoluta (distância) medida relativamente a um referencial estático com origem no "robot". Ao longo do trabalho são propostas algumas técnicas que se baseiam na aquisição de imagens em diferentes posições do universo circundante utilizando um sistema visão colocado na extremidade de um manipulador ou num "robot" móvel.

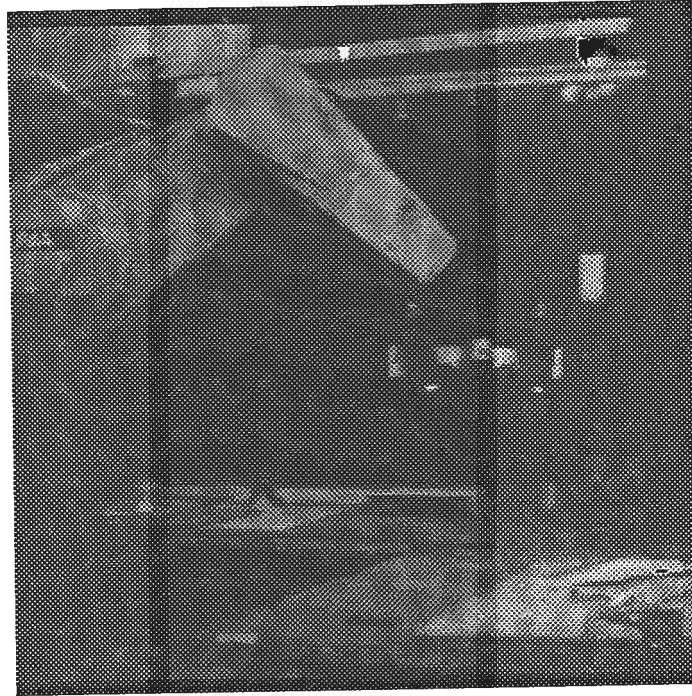


Figura 1.6 - Imagem do sistema de visão utilizado para adquirir as imagens.

Com um sistema de visão em que as câmaras são móveis, ou os seus parâmetros ópticos podem ser controlados, as imagens poderão ser amostradas nos locais que melhor favorecem a resolução do problema da determinação da estrutura tridimensional. Ballard, Aloimonos, Bajcsy, Abbot e outros têm publicado trabalhos que provam a existência de grandes vantagens na utilização desta forma de realizar a visão por computador, comparadas com as obtidas com câmaras fixas [Ballard 88] [Aloimonos 88] [Aloimonos 89] [Bajcsy 88] [Abbot 88]. Neste trabalho o princípio é utilizado para obter informação tridimensional sobre a cena. Visto este conceito estar ligado à noção de observador activo, ocupado com alguma forma de actividade, as restrições geométricas que regem o fenómeno de percepção podem ser manipuladas. Esse facto poderá ser aproveitado para a determinação da profundidade através de focagem, movimento ou estereovisão dinâmica. Neste trabalho faz-se uso de princípios semelhantes, para obter a informação tridimensional da cena.

Nos capítulos 2 e 3 são apresentados os modelos de formação das imagens assim como os modelos geométricos utilizados para representarem as primitivas nas imagens. É também descrita a geometria do sistema experimental e a forma como se realiza a calibração do sistema de visão.

No capítulo 4 são apresentados métodos de detecção de primitivas das imagens. Neste capítulo é descrito em pormenor o algoritmo que foi desenvolvido para extracção de segmentos de rectas em imagens. Neste capítulo são também desenvolvidas técnicas para agrupamento de segmentos de recta com base nas suas características geométricas.

No capítulo 5 descrevem-se métodos de determinação da profundidade através de focagem. Os métodos permitem a extracção de um mapa de profundidade em todo o domínio da imagem. No mesmo capítulo é estudada a sua aplicação a imagens de cor.

No capítulo 6 é descrito um algoritmo para estereovisão baseado num processo multiocular para determinar mapas de profundidade. No mesmo capítulo são propostas algumas novas restrições para o estabelecimento da correspondência em estereovisão.

O capítulo 7 descreve técnicas de combinação de mapas de profundidade obtidos por focagem. Os mapas são obtidos com as câmaras em diferentes pontos do espaço e ao longo de uma trajectória do sistema de visão. Os mapas são posteriormente transformados e combinados de modo a ser descritos num único referencial.

No capítulo 8 é descrito um sistema de visão com vários graus liberdade e cuja realização vem no seguimento dos estudos realizados no nosso laboratório. O sistema permite o controlo da orientação das câmaras no espaço assim como o controlo dos parâmetros ópticos das lentes. O sistema tem como objectivo a realização de estudos sobre visão activa e nesse capítulo são apresentadas algumas das vantagens da visão activa. Os sistemas de visão activa envolvem movimento das câmaras. Com base nas imagens de movimento e utilizando "normal optical flow" foi desenvolvida uma técnica para determinação da profundidade relativa.

CAPÍTULO 2

Modelos para a Formação da Imagem e Representação das Primitivas

2.1 Introdução

Nos primeiros dois pontos deste capítulo são abordados os princípios da formação da imagem e os modelos geométricos utilizados para representar matematicamente essa formação. No ponto 2.4 são descritas as representações para as primitivas geométricas bidimensionais utilizadas neste trabalho. No ponto seguinte são descritas as formas de representação utilizadas para as primitivas geométricas tridimensionais. No último ponto são descritos alguns dos invariantes projectivos para segmentos de recta.

2.2. Modelos da formação da imagem

Num trabalho que envolva a análise de imagens para determinar a sua estrutura tridimensional, é necessário estabelecer um modelo que descreva a sua formação. Uma imagem é um registo em duas dimensões das amostras do fluxo de energia luminosa emitida pelos objectos. O modo como os objectos reflectem a luz incidente e o modelo geométrico que explica a formação da imagem são aspectos importantes no estudo da formação das imagens. Diferentes pontos de um objecto em frente de uma câmara apresentarão diferentes radiações que dependem da forma como são iluminados e do modo como reflectem a energia luminosa. Neste ponto são apresentados vários modelos geométricos que descrevem a formação da imagem numa câmara e, dentro destes, o adoptado neste trabalho.

Uma propriedade fundamental na formação de uma imagem é que a *qualquer ponto do universo tridimensional da cena corresponde um só ponto no plano da imagem*. Qualquer modelo de formação da imagem deverá obedecer a esta propriedade.

2.2.1. Projecção em perspectiva

A projecção em perspectiva é um dos modelos geométricos que explica a formação da imagem com rigor e obedece a esta propriedade.

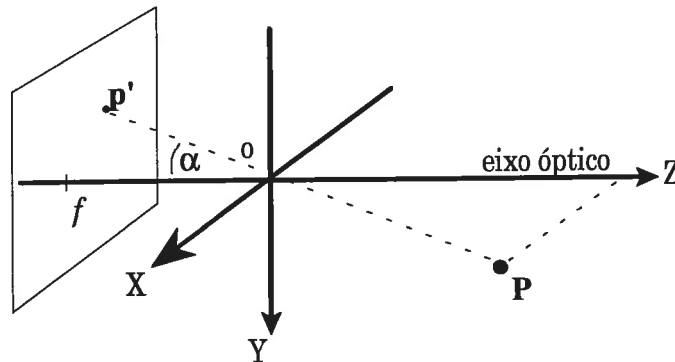


Figura 2.1 - Projecção em perspectiva com centro de perspectiva no ponto **O** e com um plano da imagem a uma distância f de **O**. A imagem p' de um ponto tridimensional **P** corresponde à intersecção do raio projectante com o plano imagem e que passa pelo ponto e pelo centro de perspectiva.

Na projecção em perspectiva, um ponto **P** tridimensional tem uma imagem p' definida pela intersecção de um raio projectante que passa pelo centro de perspectiva e pelo ponto tridimensional, com um plano projectante conhecido por *plano imagem*. Colocando a origem de um sistema de coordenadas tridimensional no centro de perspectiva **O** e o eixo-z numa direcção perpendicular ao plano de imagem (eixo óptico) então o plano de imagem será paralelo ao plano-xy do sistema de coordenadas. A distância f segundo o eixo óptico entre o plano-xy e a origem do sistema de coordenadas é designada *distância* ou *comprimento focal*.

Um ponto **P** dum objecto projecta-se na imagem num ponto p' que obedece a uma relação geométrica entre as distâncias de **P** e de p' ao centro de perspectiva. Se tomarmos um vector $\mathbf{v}=(X, Y, Z)$ e um vector $\mathbf{v}'=(x', y', f)$ que representam, no sistema de coordenadas, o ponto **P** e o ponto p' na imagem e se α é o ângulo que o raio projectante faz com o eixo óptico então podemos estabelecer as seguintes relações geométricas entre os vectores

$$Z = |\mathbf{v}| \cos \alpha$$

$$f = |\mathbf{v}'| \cos \alpha$$

sendo a relação entre os módulos dos dois vectores dada por

$$\frac{|\mathbf{v}'|}{|\mathbf{v}|} = \frac{f}{Z} = k.$$

A partir da representação vectorial para os dois vectores \mathbf{v} e \mathbf{v}' e atendendo ao sentido diferente dos vectores, podemos escrever a relação

$$\mathbf{v}' = -k\mathbf{v} = -\frac{f}{Z}\mathbf{v}.$$

Por vezes, assume-se que a formação da imagem não é num plano colocado a uma cota negativa do eixo óptico, mas sim numa cota positiva desse mesmo eixo. Esta disposição permite uma melhor compreensão da formação da imagem evitando a sua inversão e abolindo o sinal na equação anterior.

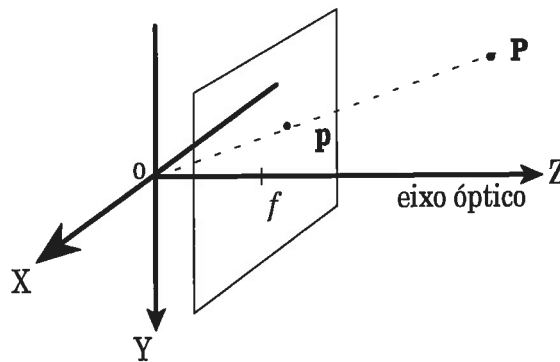


Figura 2.2 - Projecção em perspectiva com centro no ponto O e com um plano da imagem colocado na parte positiva do eixo óptico. Com o plano imagem nesta posição a imagem não é invertida mas toda a geometria de projecção mantém-se.

Neste caso, se o ponto P é expresso em coordenadas tridimensionais por $\mathbf{v} = (X\hat{i} + Y\hat{j} + Z\hat{k})$ e a sua imagem p' por $\mathbf{v}' = (x'\hat{i} + y'\hat{j} + f\hat{k})$. As relações entre as coordenadas tridimensionais e as coordenadas na imagem para as componentes segundo o eixo-x e o eixo-y são dadas pelas expressões

$$x' = u = f \frac{X}{Z} \quad y' = v = f \frac{Y}{Z} \quad (2.1)$$

o qual estabelece uma relação geométrica entre pontos tridimensionais dos objectos e as suas imagens $(x', y') = (u, v)$. Esta relação não representa a dimensão z que corresponde à distância a que se encontra o ponto P do centro de perspectiva. Este modelo de projecção capta o efeito da perspectiva na formação das imagens, embora muitos autores optem por utilizar modelos mais simples. No ponto seguinte abordar-se-ão alguns destes modelos matemáticos que poderão ser utilizados em visão por computador.

Os dispositivos de captação de imagens utilizam lentes para produzir uma imagem da cena. Nesse caso e assumindo que a lente é fina e se comporta de modo ideal, o centro da lente corresponderá ao centro de perspectiva O . No caso de o sistema óptico ser composto por várias lentes ele poderá ser reduzido a uma lente deste tipo [Williams 86].

2.2.2. Modelos de Projecção Alternativos

Projecção ortográfica

Um dos modelos mais simples de projecção é a projecção ortográfica. Esta forma de projecção está adequada para análise de imagens onde o tamanho dos objectos, na cena tridimensional, é muito pequeno comparado com a distância dos objectos ao plano imagem. Um exemplo deste tipo de imagens é a fotografia aérea.

Este tipo de projecção constitui uma variante à projecção em perspectiva, e é equivalente a colocar o factor de amplificação $m = f/Z$ igual a 1. Neste caso utilizando as equações (2.1) obtemos

$$x' = x \quad y' = y \quad (2.2).$$

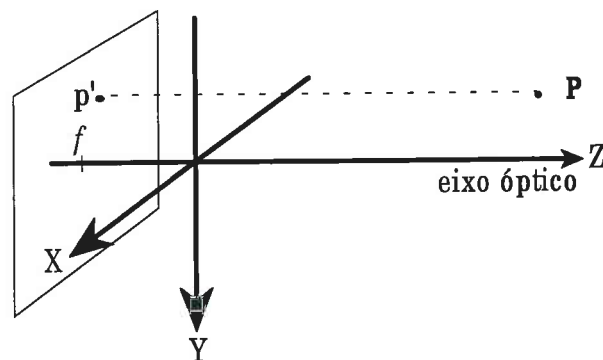


Figura 2.3 - A projecção ortográfica é um modelo de formação da imagem adequado para análise de imagens quando o tamanho dos objectos na cena é muito pequeno, comparado com a distância dos objectos ao plano da imagem.

Nesta forma de projecção um ponto tridimensional P terá uma imagem p' que é obtida através de uma projecção paralela ao eixo óptico. Assim na projecção ortográfica e ao contrário da projecção em perspectiva, as distâncias em planos paralelos ao plano de imagem são preservadas. Este modelo de projecção conduz a expressões mais simples nas expressões utilizadas na análise das imagens e por vezes com vantagens significativas sobre as equações com base no modelo de projecção em perspectiva.

Paraperspectiva

A projecção ortográfica é um modelo simples para a formação das imagens e menos rigoroso do que a projecção em perspectiva. Por outro lado a projecção em perspectiva produz equações difíceis de manipular que tornam a análise da imagem mais complexa. Uma outra forma de projecção é a paraperspectiva, que consiste numa modificação da projecção em perspectiva e que, de certo modo, se aproxima à projecção ortográfica. Aloimonos, utilizou esta forma de projecção e descreveu algumas das suas propriedades comparando-as com a projecção em perspectiva e a ortográfica [Aloimonos 86].

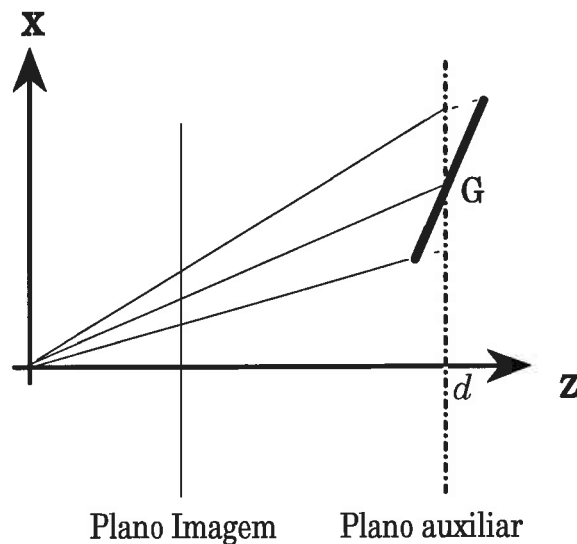


Figura 2.4 - Vista lateral da projecção paraperspectiva. A superfície S que se pretende projectar tem um centro de massa em G cuja cota segundo o eixo- z é d . A projecção faz-se primeiro para o centro de massa G e segundo uma direcção que passa pela origem do sistema de coordenadas. Após isso todos os pontos da superfície são projectados no plano auxiliar segundo essa mesma direcção. Após essa projecção todos os pontos no plano auxiliar serão de novo projectados no plano imagem segundo a projecção em perspectiva.

Para uma breve descrição deste tipo de projecção, considere um sistema de coordenadas com o plano de imagem colocado na parte positiva do eixo- z de modo a não inverter a imagem como se mostra na figura 2.4. Colocando um plano paralelo à imagem na cota $z=d$ do eixo óptico, a projecção de uma superfície com centro de massa G à distância d do eixo- z do sistema de coordenadas é feita como se mostra na figura. A projecção faz-se para o centro de massa G e segundo uma direcção que passa pela origem do centro de coordenadas O . Após isso, todos os pontos da superfície são projectados no plano auxiliar, segundo essa mesma direcção. Após essa projecção todos os pontos no plano auxiliar serão projectados no plano imagem, segundo a projecção em perspectiva.

Proiecção Esférica

Se um objecto é projectado no plano imagem segundo a projecção em perspectiva, a sua imagem é alterada quando se verifica uma rotação da câmara segundo um dos eixo-x ou y do referencial da câmara. No caso de o ângulo de rotação ser grande existe a possibilidade de perder a imagem do objecto. Existem formas de projecção que são invariantes a este tipo de movimento. Uma destas formas é a projecção esférica que faz a projecção segundo uma esfera de raio f , centrada num ponto O e obtendo uma imagem esférica [Kanatani 90]. As dimensões da imagem de um objecto projectado nessa esfera não sofrem alteração com a rotação segundo quaisquer dos três eixos x , y ou z .

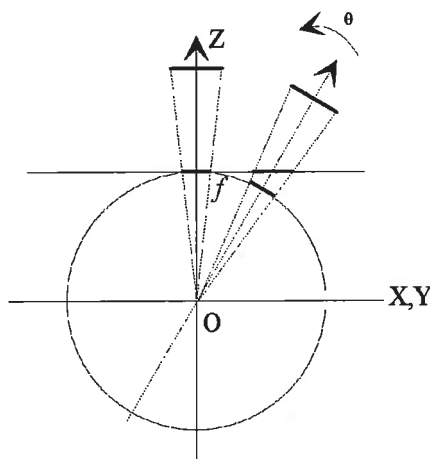


Figura 2.5 - A figura representa num plano o efeito da rotação de uma câmara. Quando é utilizada a projecção em perspectiva, a imagem do objecto é alterada quando câmara sofre uma rotação de um ângulo θ segundo um dos seus eixo-x ou y.

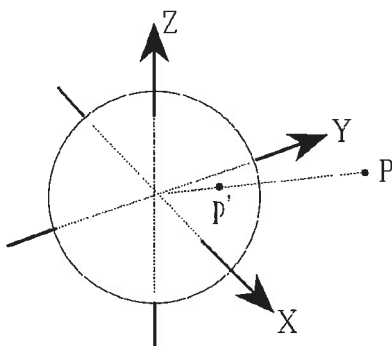


Figura 2.6 - Em projecção esférica um ponto $P=(x,y,z)$ é projectado no ponto P' de intersecção da recta com a esfera $X^2 + Y^2 + Z^2 = f^2$ que passa pelo ponto e pela origem O .

Em perspectiva esférica qualquer ponto $P=(x,y,z)$ da cena é projectado num ponto p' . Esse ponto corresponde à intersecção da recta que se apoia no ponto P e no centro O da esfera. Com este de modelo a rotação da câmara não provoca o efeito de distorção dos objectos visualizados, reduzindo a complexidade dos problemas da determinação de áreas

projectadas. Se o movimento da câmara relativamente ao objecto for só de rotação, os comprimentos e ângulos entre arestas dos objectos serão invariantes com o movimento.

Este modelo de projecção pode facilitar a análise da imagem no caso do movimento de rotação, mas na prática os sensores utilizados nas câmaras são planares e os sinais por eles gerados são amostrados de forma a fazerem o seu "display" também em dispositivos planos. Teoricamente a utilização de uma imagem esférica permite um ângulo de visualização que pode ir até a um hemisfério da esfera da figura 2.6. No entanto o problema da utilização desta projecção é a forma como a esfera deverá ser repartida de forma periódica, em locais que representem amostras do sinal vídeo e o mecanismo de indexação desses pedaços tanto para escrita como para leitura.

2.2.3. Lentes e Focagem

Um ponto de um objecto numa cena, quando iluminado, reflecte a luz incidente em várias direcções. Os raios de luz que são reflectidos na direcção do sensor, atravessam as lentes que fazem parte do dispositivo e formam uma imagem. Os dispositivos que se baseiam em lentes só captam uma quantidade finita de radiação luminosa. Em muitas lentes é possível controlar, através de um dispositivo designado por *íris*, a quantidade de radiação luminosa que as atravessam. A *íris*, pode ser constituída por um só diafragma ou por um conjunto de dispositivos desenhados de modo a controlarem a passagem da radiação luminosa [Baker 80]. Quanto maior for a *abertura* da íris mais energia luminosa irradiada pelos objectos é captada. Outro parâmetro importante nas lentes é a *distância focal*. A *distância focal*, tal como mostra a figura 2.7, limita o valor máximo do ângulo entre um raio de luz e o eixo óptico, definindo o *campo de visão* do dispositivo.

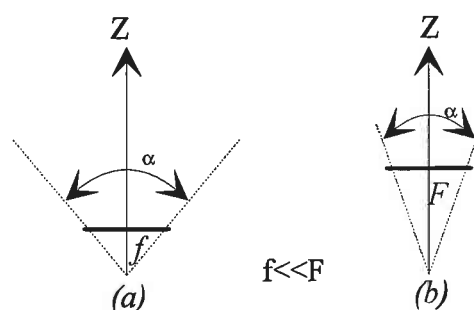


Figura 2.7 - O campo de visão de uma câmara depende da distância focal da lente e do diâmetro da sua abertura. Para distâncias focais pequenas comparadas com as dimensões da imagem (grande angular), o campo de visão é maior - figura (a). Para distâncias focais grandes comparadas com as dimensões da imagem (teleobjectivas) o campo de visão é menor - figura (b).

No caso de teleobjectivas o campo de visão é bastante pequeno porque normalmente a sua distância focal é grande comparada com as dimensões da imagem. No caso das lentes

do tipo grande-angular passa-se o inverso, isto é, permitem um grande ângulo de visão. Neste caso as lentes apresentam uma pequena distância focal relativamente às dimensões da imagem.

Na análise da formação da imagem pode ser utilizada a geometria óptica clássica ou a óptica física e teoria da difracção. A geometria óptica clássica utiliza o conceito de raio luminoso e a forma como ele se comporta quando atravessa uma lente para explicar a formação de uma imagem. Esta abordagem é uma aproximação de primeira ordem pois despreza o efeito da distorção da lente. A teoria da óptica física e a teoria da difracção envolvem equações não lineares para explicar o comportamento da energia luminosa quando atravessa a lente e dá origem à imagem. Como as câmaras que são utilizadas têm suficiente resolução espacial, esta última aproximação torna a análise demasiado complexa e sem ganhos significativos na explicação do processo, podendo ser utilizada a óptica geométrica.

Para deduzir as expressões seguintes vamos assumir que as lentes utilizadas no sistema de aquisição da imagem não provocam um desvio significativo nos raios luminosos ao refractarem a radiação luminosa. Nesse caso as lentes podem ser representadas por uma lente delgada e biconvexa. Como a lente é delgada podemos desprezar a sua espessura e definir um plano, chamado plano principal, que passa pelo seu centro e é perpendicular ao eixo óptico, intersectando-o num ponto, denominado ponto principal C . Neste ponto C será colocado o centro de perspectiva do modelo em perspectiva. Relativamente a esse plano definem-se dois outros planos paralelos, colocados à distância f para ambos os lados da lente e denominados planos focais. A figura 2.8 representa uma lente delgada, biconvexa, com distância focal f , em frente da qual se encontra um ponto P à distância d_{out} do plano principal. Esse ponto tem como imagem o ponto p' à distância d_{in} do plano principal.

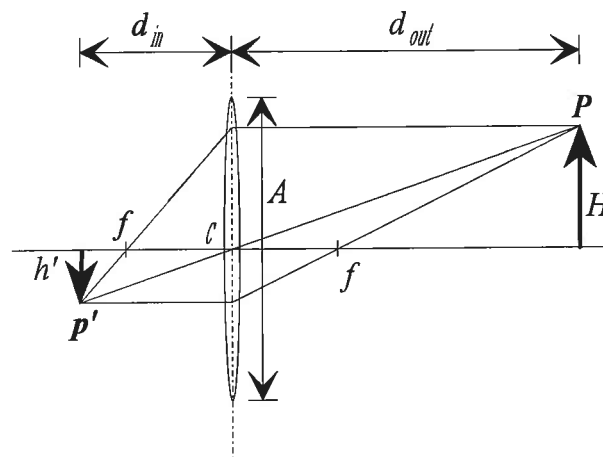


Figura 2.8 - Processo de formação de uma imagem com uma lente delgada.

Utilizando a figura 2.8 obtemos por semelhança de triângulos as seguintes expressões

$$\frac{H + h'}{d_{in}} = \frac{H}{f} \quad \frac{H + h'}{d_{out}} = \frac{h'}{f} \quad (2.3)$$

que somadas dão a expressão

$$\frac{1}{d_{in}} + \frac{1}{d_{out}} = \frac{1}{f} \quad (2.4)$$

que relaciona a distâncias entre o objecto, a sua imagem e a distância focal para o caso de lentes finas.

A partir do processo de formação das imagens, podemos descrever o processo de desfocagem. O processo de desfocagem verifica-se quando a convergência de toda energia luminosa irradiada por um ponto no mundo tridimensional não se encontra num ponto do plano imagem mas sim, numa determinada área do plano. O fenómeno está ilustrado na figura 2.9 onde se mostra o efeito do deslocamento do plano imagem de uma distância δ segundo o eixo óptico. Um ponto tridimensional P (teoricamente) no infinito tem uma imagem num ponto p' no plano imagem colocado à distância focal f . Se o plano do sensor é deslocado de uma distância δ , sendo a abertura da lente circular, a imagem de P deixa de ser um ponto p' para ser um círculo de diâmetro $2r_c$. Pela figura 2.9 e utilizando uma relação de triângulos obtemos a seguinte expressão para o raio do círculo

$$\frac{r_l}{d_{in}} = \frac{r_c}{\delta} \Rightarrow r_c = \frac{\delta r_l}{d_{in}} \quad (2.5).$$

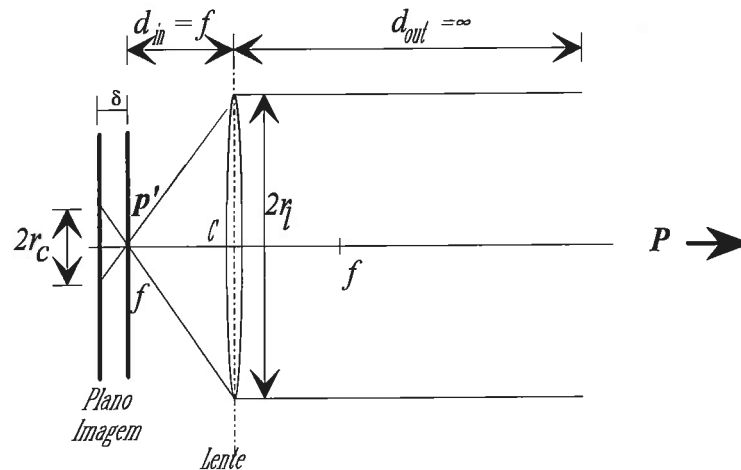


Figura 2.9 - Processo de formação de uma imagem focada de um ponto P colocado no infinito, utilizando uma lente delgada. Deslocando o plano de imagem de uma distância δ , a imagem que se forma é desfocada.

Normalmente associada à focagem está associada a noção de *profundidade de campo*. A profundidade de campo é o intervalo de distâncias sobre os quais os objectos podem ser considerados como bem focados. Este parâmetro está relacionado com a resolução espacial do sensor de imagem, elementos restantes do dispositivo óptico e a distância ao objecto. Devido à resolução espacial finita do sensor existe a possibilidade do

círculo de desfocagem $2r_c$ não poder ser detectado no caso de ter um diâmetro menor do que a resolução do sensor de imagem. A essa zona de incerteza na definição da imagem, designada *profundidade de campo*, corresponde uma zona na qual os objectos se poderão localizar mantendo uma imagem sempre focada. A expressão para a *profundidade de campo* D pode ser obtida com o auxílio da figura 2.10.

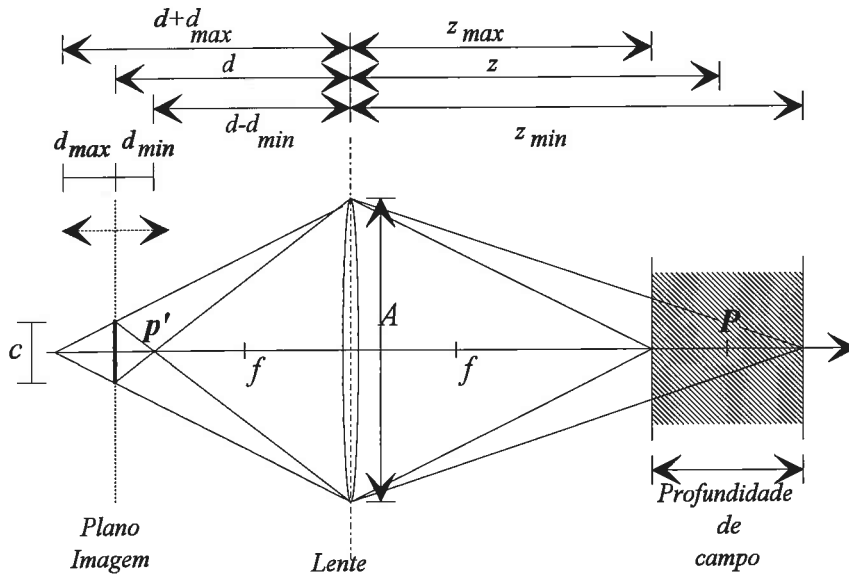


Figura 2.10 - A resolução finita do sensor permite que se defina uma zona na qual os objectos no seu interior têm imagens sempre focadas. A distância, segundo o eixo óptico, correspondente a essa zona é conhecida como profundidade de campo.

Com base na semelhança de triângulos e na equação (2.4) podemos obter expressões para as distâncias de deslocamento d_{\min} e d_{\max} e para a distância z . As expressões são

$$d_{\min} = \frac{cd}{A+c} \quad d_{\max} = \frac{cd}{A-c} \quad z = \frac{df}{d-f} \quad (2.6).$$

Pela expressão (2.4) para as lentes e transformando e relacionando as distâncias do lado do plano imagem e do lado dos objectos, obtemos as seguintes expressões para z_{\max} e z_{\min}

$$z_{\max} = \frac{(d+d_{\max})f}{(d+d_{\max})-f} = \frac{zAf}{fA+c(z-f)} \quad z_{\min} = \frac{(d-d_{\min})f}{(d-d_{\min})-f} = \frac{zAf}{fA-c(z-f)} \quad (2.7)$$

sendo a profundidade de campo dada por

$$D = z_{\min} - z_{\max} = \frac{2zAcf(z-f)}{f^2A^2 - c^2(z-f)^2} \quad (2.8).$$

A expressão (2.8) mostra que a profundidade de campo D aumenta de forma quadrática com o aumento da distância z do objecto ao sensor, tal como se mostra na figura 2.11.

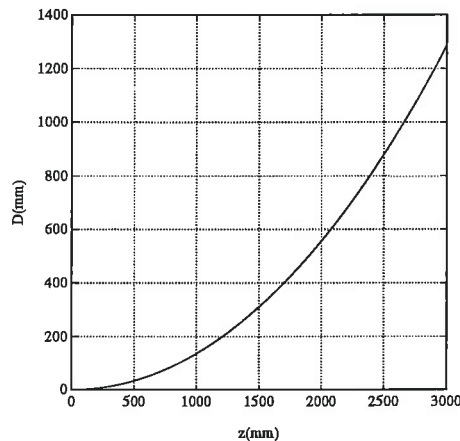


Figura 2.11- A figura mostra a variação da profundidade de campo D em função da distância z do objecto ao sensor. Para o cálculo teórico da função utilizou-se $f=16\text{mm}$ e $A=10\text{mm}$ e $c=0.011\text{mm}$.

2.2.4 Formulação matemática do modelo

No modelo em perspectiva a imagem obtida pela câmara é representada por um plano I e o centro de perspectiva representado por O - veja-se a figura 2.12. Por definição, a projecção de um ponto P no espaço corresponde à intersecção da recta que passa por P e está assente em O . Essa intersecção corresponde a um ponto na imagem I e designado por p . Se $c = (u_0, v_0)$ é a projecção de O sobre a imagem, então a distância focal f é definida pela distância de c ao ponto O .

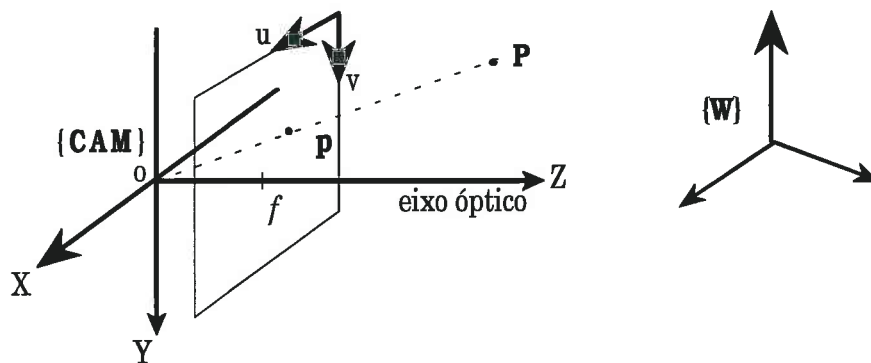


Figura 2.12 - Modelo geométrico da câmara

Para descrever este processo matematicamente, assumimos que as coordenadas de um ponto $P = (X, Y, Z)$ no referencial $\{W\}$ podem ser expressas no referencial $\{CAM\}$

utilizando uma transformação que relaciona os dois referenciais. A expressão dessa transformação é dada por (2.9) e pode ser escrita na forma matricial (2.10) utilizando coordenadas homogêneas. A transformação T da expressão (2.10) é, por definição, uma matriz não singular e inclui valores da rotação e da translação que relaciona os dois referenciais.

$$\begin{aligned} x &= \frac{c_{11}X + c_{12}Y + c_{13}Z + c_{14}}{c_{41}X + c_{42}Y + c_{43}Z + c_{44}} \\ y &= \frac{c_{21}X + c_{22}Y + c_{23}Z + c_{24}}{c_{41}X + c_{42}Y + c_{43}Z + c_{44}} \\ z &= \frac{c_{31}X + c_{32}Y + c_{33}Z + c_{34}}{c_{41}X + c_{42}Y + c_{43}Z + c_{44}} \end{aligned} \quad (2.9)$$

$$\begin{bmatrix} sx \\ sy \\ sz \\ s \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \\ c_{31} & c_{32} & c_{33} & c_{34} \\ c_{41} & c_{42} & c_{43} & c_{44} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \Rightarrow {}^{CAM}\mathbf{P} = \mathbf{T} {}^W\mathbf{P} \quad (2.10)$$

Como a imagem corresponde ao lugar geométrico no qual se encontram as projecções \mathbf{p} dos pontos tridimensionais \mathbf{P} então a expressão

$$sx A + sy B + sz C + s D = 0$$

deve ser satisfeita. Esta equação representa um plano que, para o caso da figura 2.12, deverá ter os coeficientes $A = B = 0$. Nesse caso o plano será expresso pela equação

$$z C + D = 0 \quad (2.11)$$

com $C \neq 0$. Se as expressões (2.9) e (2.11) forem combinadas obtém-se a equação

$$Cz + D = C \frac{c_{31}X + c_{32}Y + c_{33}Z + c_{34}}{c_{41}X + c_{42}Y + c_{43}Z + c_{44}} + D = 0 \quad (2.12).$$

Esta condição é uma restrição que é necessário satisfazer para todos os pontos $\mathbf{P} = (X, Y, Z)^T$ expressos no referencial $\{W\}$. Para satisfazer a expressão (2.12) é necessário que, pelo menos se mantenha uma relação constante entre os coeficientes $c_{3i}/c_{4i} = k$, sendo k uma constante diferente de zero. Note-se que para todos os valores possíveis de k e diferentes de zero a restrição (2.12) é sempre satisfeita. Por exemplo definindo $c_{44} = 1$

equivale a definir implicitamente a relação c_{3i}/c_{4i} . Representando as coordenadas da imagem por vectores e levando em conta que o valor de z é sempre constante teremos

$$\begin{bmatrix} sx \\ sy \\ s \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \\ c_{31} & c_{32} & c_{33} & c_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \Rightarrow \mathbf{p} = \mathbf{CALIB} \mathbf{P} \quad (2.13).$$

Esta expressão representa um operador de projecção em que (X, Y, Z) representam, de modo genérico, as coordenadas tridimensionais de um ponto na cena. A equação de projecção representada por (2.13) pode ser considerada como sendo composta de três aplicações :

- uma aplicação que prolonga o espaço euclidiano \mathcal{R}^3 ao espaço projectivo \mathbf{P}^3
 $(x, y, z)^T \rightarrow (x, y, z, 1)^T$;
- uma aplicação linear projectiva **CALIB** de \mathbf{P}^3 no plano projectivo \mathbf{P}^2 ;
- uma aplicação de \mathbf{P}^2 no plano euclidiano \mathcal{R}^2 e correspondente imagem
 $(a, b, s)^T \rightarrow (a/s, b/s)^T$.

Esta decomposição da operação descrita por (2.13) é essencial para exemplificar que algumas propriedades da geometria projectiva se verificam também por este modelo para as câmaras.

Centro de perspectiva

O centro de perspectiva **O** é o local geométrico no plano \mathcal{R}^3 que verifica a equação

$$\mathbf{CALIB} \begin{bmatrix} \mathbf{O} \\ 1 \end{bmatrix} = 0 \quad (2.14)$$

Decompondo a matriz $\mathbf{CALIB} = [\mathbf{C}_{(3 \times 3)} \quad \mathbf{a}_{(3 \times 1)}]$, corresponderá à solução

$$\mathbf{C} \mathbf{O} = -\mathbf{a} \quad (2.15).$$

Recta de projecção

Se **CALIB** representa uma matriz de calibração então as coordenadas dos pontos (u, v) na imagem são obtidos através da relação

$$\begin{bmatrix} U \\ V \\ S \end{bmatrix} = \mathbf{CALIB} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \mathbf{CALIB} \mathbf{P} \quad (2.16)$$

em que as coordenadas (u, v) são dadas por

$$u = \frac{U}{S} \quad \text{e} \quad v = \frac{V}{S} \quad (2.17).$$

Representando a matriz **CALIB** numa forma mais compacta

$$\mathbf{CALIB} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \\ c_{31} & c_{32} & c_{33} & c_{34} \end{bmatrix} = \begin{bmatrix} \mathbf{c}_1^T & c_{14} \\ \mathbf{c}_2^T & c_{24} \\ \mathbf{c}_3^T & c_{34} \end{bmatrix} \quad (2.18)$$

então tirando expressões explícitas para (u, v) através de (2.16) obtemos o sistema

$$\begin{cases} (\mathbf{c}_1 - u \mathbf{c}_3)^T \mathbf{P} + c_{14} - u c_{34} = 0 \\ (\mathbf{c}_2 - v \mathbf{c}_3)^T \mathbf{P} + c_{24} - v c_{34} = 0 \end{cases} \quad (2.19).$$

Estas duas equações representam dois planos que contêm o centro óptico **O** e se intersectam segundo uma recta que neste caso corresponderá à recta de projecção do ponto **P**.

Estas propriedades geométricas são algumas das propriedades projectivas deste modelo e que são exploradas nos capítulos seguintes e em especial no capítulo dedicado à estereovisão.

2.3 Imagens de cor

A cor é um atributo importante para percepção visual. Para a análise das imagens a cores é necessário ter formas de representação da cor. As formas de representação utilizadas durante a execução deste trabalho são aqui descritas. A sensação de cor é normalmente expressa através de palavras, as quais incluem termos como cor vermelha, cor azul ou amarela e alguns adjectivos tais como "brilhante", "suave" e "viva", que exprimem a saturação e a intensidade da cor. Esta forma de descrever a cor ajuda a comunicar a sua percepção, mas é desajustada quando se pretende uma especificação objectiva e rigorosa.

Para isso, é necessário definir grandezas de medida da cor e estabelecer mecanismos de as determinar fisicamente.

A sensação de cor corresponde à sensação visual de uma composição espectral da energia luminosa. Distinguir entre várias cores implica ter estímulos visuais diferentes para cada cor mas podemos ter a mesma sensação de cor com composições espectrais diferentes. No entanto a cor é uma característica física da maioria dos objectos que nos rodeiam e portanto tem interesse medi-la. Uma maneira de o fazer é por comparação visual com um conjunto de amostras "standard". Este metodo é muito subjectivo pois depende das classificações feitas pelos observadores. Uma maneira mais objectiva é a utilização da teoria electromagnética em que a luz visível é tratada como energia electromagnética com uma distribuição espectral na sua parte visível representando a sua composição espectral através das três grandezas: o *comprimento de onda dominante*, a *pureza* e a *intensidade*. O *comprimento de onda dominante* corresponde ao comprimento de onda mais frequente no espectro enquanto que a *pureza* indica em que medida a radiação luminosa é composta por mais do que um comprimento de onda. A *intensidade* pode ser descrita em termos de fotões ou de onda electromagnética. Em termos de fotões, corresponde à quantidade de fotões presentes e em termos de ondas electromagnéticas descreve a amplitude da onda dominante. Uma radiação luminosa com um único comprimento de onda ou com uma banda espectral muito estreita é monocromática e, por definição, 100% pura. Esta representação não evita que diferentes distribuições espectrais de energia produzam a mesma sensação de cor, pois a relação entre distribuição espectral e cor é de muitas para uma. Relacionando esta representação com a percepção visual podemos ter uma definição mais objectiva da sensação de cor. Assim, a cor corresponde ao *comprimento de onda dominante* da radiação, saturação da cor à *pureza* e o brilho à *intensidade* [Thorell 90].

A retina de um ser humano contém três receptores de cor diferentes, que apresentam sensibilidades diferentes em três zonas diferentes do espectro. Os receptores têm picos de sensibilidade a aproximadamente 445(azul), 535(verde) e 570(vermelho) nanometros. Cada tipo de receptor integra a energia da luz incidente em vários comprimentos de onda e em proporção à sua sensibilidade para aquele comprimento de onda. Os três valores resultantes são responsáveis pela sensação de cor pelos humanos [Netravali 88]. Esta é, de facto, a base para a teoria de que a cor é obtida pela combinação de três valores e não por uma função de vários comprimentos de onda sobre todo o espectro visível como se referiu atrás. Segundo este princípio, qualquer cor pode ser sintetizada por uma mistura de três primários escolhidos de modo apropriado e os desenvolvimentos da teoria da cor permitem aplicar princípios da associatividade e da linearidade à composição de cores [Thorell 90].

Em 1931 a CIE (Commission Internationale de L'Eclairage) definiu três dimensões (X,Y,Z) que combinadas de forma aditiva permitem definir todas as cores visíveis. Estas dimensões não são, na verdade, estímulos de cor reais, mas permitem definir uma determinada cor. O espaço CIE pode ser convertido em duas dimensões pois é possível

definir um triplete (x, y, z) que depende da quantidade total de energia da radiação e cujos elementos se relacionam pelas expressões

$$x = \frac{X}{X+Y+Z} \quad y = \frac{Y}{X+Y+Z} \quad z = \frac{Z}{X+Y+Z} \quad x+y+z=1 \quad (2.20).$$

Deste modo é possível desenhar uma representação para todas as cores num diagrama bidimensional. Desenhando os valores das coordenadas (x, y) para todas as cores possíveis obtemos o *diagrama de cromaticidade CIE*.

A comissão definiu, com base em testes de várias pessoas, valores para os estímulos R, G e B de um observador "standard". Esses valores permitiram a definição de padrões de comparação com base em três estímulos primários de cor e com comprimentos de onda de 700nm(R), 546,1nm(G) e 435,8nm(B) nm e correspondentes ao vermelho, verde e azul. Para além disto, e com base no observador "standard", a CIE definiu a relação entre as variáveis (X, Y, Z) e os estímulos (R, G, B) através da relação linear

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 2.365 & -0.897 & -0.468 \\ -0.515 & 1.426 & 0.089 \\ 0.005 & -0.014 & 1.009 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2.21).$$

É possível obter a relação inversa fazendo a inversão da matriz que relaciona as duas quantidades. Esta relação é de grande utilidade pois permite relacionar as componentes R, G e B utilizadas em dispositivos de visualização e em câmaras de aquisição de imagens a cores, com as dimensões utilizadas pelo sistema CIE, que é um "standard" internacional para comparação de cores.

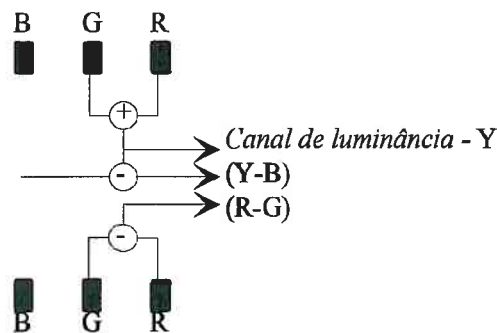


Figura 2.12 - Modelo dos canais opostos. Os estímulos tricromáticos R,G e B são transformados de modo a darem num canal a informação da luminância e noutros dois canais, os sinais cromáticos.

A utilização do princípio dos três estímulos não é a única teoria para explicar a sensação de cor. Pela nossa experiência, sabemos que temos uma percepção idêntica de uma

imagem quando olhamos para a sua versão "preto/branco" ou a sua versão a cores. Do ponto de vista humano os receptores que são utilizados são os mesmos da detecção da cor. Para explicar este fenómeno tem sido apontada a teoria de que o modelo tricromático é posteriormente recodificado em dois canais cromáticos e um canal correspondente à informação de luminância tal como mostra a figura 2-12.

Este modelo pressupõe que os canais cromáticos R, G e B são entrada para o modelo, fornecendo posteriormente as duas saídas cromáticas ($r-g$) e ($y-b$) chamadas *canais opostos* e uma saída de *luminância* Y.

Os modelos anteriores não têm em conta a nossa noção intuitiva de *cor*, *saturação* e *brilho ou intensidade*. Para levarem em conta esta noção outros modelos foram desenvolvidos e de entre vários um foi utilizado neste trabalho - o modelo HSI ("Hue Saturation Intensity") [Ballard 82] [Thorell 90]. Neste modelo o espaço dos valores dos vários componentes de cor estão restritos a um cone duplo tal como mostra a figura 2.13. O valor para a *cor* (Hue) é um ângulo que é medido em graus e na qual o valor 0 corresponde à cor vermelha. A *saturação* é medida radialmente a partir do eixo vertical com valores entre 0, junto ao eixo, até 1, na superfície do cone. Para o *brilho* ("Lightness") a sua medida é feita no eixo vertical e corresponde 0 para o preto na extremidade inferior do cone e 1 para o branco na extremidade superior do cone. Segundo Ballard [Ballard 90] a relação entre as componentes RGB e HSI são dadas pelas expressões

$$\begin{aligned}
 I &= r + g + b \\
 S &= 1 - \frac{3 \min(r, g, b)}{I} \\
 H &= \cos^{-1} \left\{ \frac{\frac{1}{2}[(r-g) + (r-b)]}{\sqrt{(r-g)^2 + (r-b)(g-b)}} \right\} \\
 \text{Se } (b > g) \text{ então } H &= 2\pi - H
 \end{aligned}
 \tag{2.22}.$$

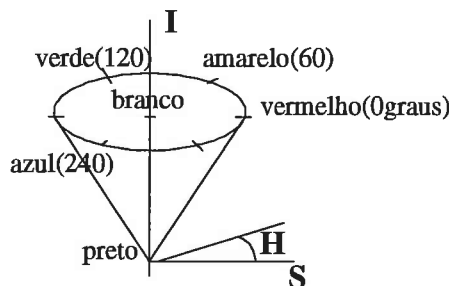


Figura 2.13 - O modelo HSI

2.4. Representação de Primitivas Bidimensionais

Segundo o modelo de projecção descrito pela expressão (2.13) as imagens representam, num plano 2D, as projecções de pontos existentes numa cena. Por vezes essas projecções apresentam características geométricas específicas que podem ser representadas por modelos geométricos conhecidos. Esses modelos são designados genericamente: *primitivas*. As primitivas bidimensionais mais básicas que se podem encontrar numa imagem são pontos e rectas. Neste trabalho são utilizadas extensivamente estas primitivas básicas e neste ponto são abordadas as formas utilizadas para a sua representação.

2.4.1 Pontos

Os pontos são descritos por duas coordenadas (u,v) que se relacionam com as coordenadas tridimensionais de um ponto no espaço (X,Y,Z) através das expressões do modelo em perspectiva.

$$u = f \frac{X}{Z} \quad v = f \frac{Y}{Z}$$

em que f representa a distância focal da lente que está a ser utilizada para adquirir as imagens.

2.4.2 Rectas

Uma representação mínima para uma recta em duas dimensões é um ponto e uma direcção. Como existem vários pares (*vectores, pontos*) que representam a mesma recta, esta definição não conduz a uma representação única. Uma maneira mais correcta e conduzindo a uma representação única é a utilização da representação paramétrica

$$x \cos \phi + y \sin \phi = \rho \quad (2.23)$$

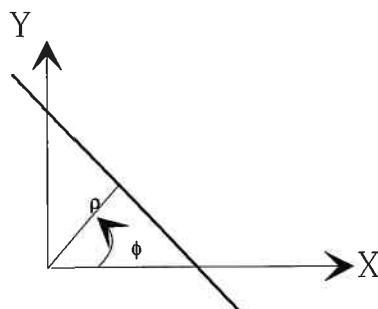


Figura 2.14 - Representação para uma recta em duas dimensões

e com $\rho \geq 0$ e $\phi \in [0..2\pi]$. Esta será a forma preferencial utilizada ao longo deste trabalho para representação duma recta bidimensional. Esta forma paramétrica pode ser expressa numa forma vectorial por

$$\begin{bmatrix} \cos \phi & \sin \phi & -\rho \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{v}^T \mathbf{p} = 0 \quad (2.24).$$

O vector \mathbf{v} tem a particularidade de representar, tridimensionalmente, um vector perpendicular à recta 2D existente num plano com cota $z=1$ pois o produto escalar representado por (2.24), entre \mathbf{v} e qualquer ponto \mathbf{p} da recta é sempre nulo. No caso desta recta 2D ser a projecção de uma recta 3D no espaço então este vector é também uma normal ao plano formado pela recta 3D e pelo centro da imagem e chamado plano projectante. Como qualquer ponto da recta 2D pertence ao plano projectante da recta 3D então se o vector \mathbf{v} é perpendicular a qualquer ponto da recta também o será ao plano projectante.

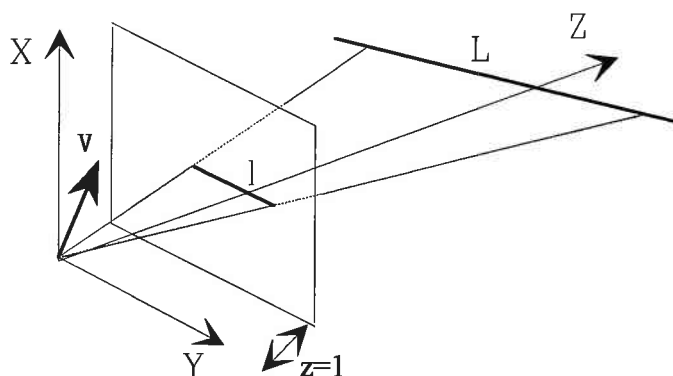


Figura 2.15 - O vector \mathbf{v} é perpendicular ao plano projectante da recta L

Outra forma alternativa para representação de uma recta e utilizada, por vezes, neste trabalho é a forma

$$Ax + By + C = 0 \quad (2.25).$$

Esta forma pode ser obtida pelo conhecimento de dois pontos $\mathbf{P}(x_1, y_1)$ e $\mathbf{Q}(x_2, y_2)$ pertencentes à recta. Utilizando estes dois pontos a expressão para a recta poderá ser obtida através do determinante

$$\begin{vmatrix} x & y & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{vmatrix} = 0$$

2.5 Representação de Primitivas Tridimensionais

Na análise e tratamento de imagens feita ao longo do trabalho partimos do princípio que todos os objectos que encontramos no mundo real podem ser formados pelo seguinte conjunto de primitivas tridimensionais: pontos, rectas e planos. No interior de edifícios estas primitivas permitem representar a grande maioria dos objectos existentes nos locais de trabalho de um manipulador ou "robot" móvel. Por outro lado, se existirem objectos que envolvam curvatura estes podem ser representados de forma aproximada, por estas primitivas simples.

2.5.1 Pontos

A representação que é utilizada para pontos tridimensionais é a do triplete (x, y, z) correspondente às três coordenadas segundas os três eixos ortonormados x , y e z . Em ocasiões em que tal seja favorável podemos utilizar a representação através de coordenadas homogéneas. Neste caso um ponto no espaço é representado por um vector $\mathbf{v}=[a,b,c,d]=[x, y, z, 1]$ em que

$$x = \frac{a}{d}; y = \frac{b}{d}; z = \frac{c}{d}; \quad (2.26)$$

ou seja um vector em que a 4ª coordenada é um factor de escala. No caso de este factor ser zero então estamos perante um vector $\mathbf{t}=[u, v, z, 0]$ que não representa um ponto mas sim uma direcção dada pelos valores (u, v, z) [Penna 86].

2.5.2 Rectas

As rectas são primitivas geométricas que permitem representar pontos da cena com uma relação geométrica entre si, bem definida. Para sermos mais rigorosos, uma recta corresponde a uma "linha" de pontos no espaço tridimensional Euclidiano desde $-\infty$ e $+\infty$. A representação matemática mais comum para uma recta é considerar um ponto dessa recta e uma orientação para essa recta. Estes componentes são representadas por um ponto da recta

$$\mathbf{P}=(p_x, p_y, p_z) \quad (2.27)$$

e uma orientação

$$\mathbf{o} = (o_x, o_y, o_z) \quad (2.28)$$

que representam os cossenos directores da recta relativos aos eixos x , y e z .

Uma outra forma comum de representação de uma recta com base nestes componentes é a equação paramétrica

$$\mathbf{r} = \mathbf{P} + t\mathbf{o}, \text{ com } t \in [-\infty, +\infty] \quad (2.29)$$

em que t é uma variável real. Resolvendo a equação vectorial (2.29), em ordem a t , para cada um dos seus elementos, obtemos uma outra expressão para a recta dada por

$$\frac{x - p_x}{o_x} = \frac{y - p_y}{o_y} = \frac{z - p_z}{o_z} \quad (2.30).$$

Apesar destas representações serem bastante utilizadas, apresentam deficiências por não conduzirem a uma solução única na representação de uma recta. Por exemplo, existe um conjunto infinito de pontos \mathbf{P} que podem ser escolhidos para a representação da mesma recta, assim como um conjunto infinito de representações para a orientação \mathbf{o} (incluindo aquelas com direcção contrária, isto é $-\mathbf{o}$). Sendo o nosso problema a análise da imagem com o fim de determinar rectas no espaço, a utilização de uma representação deste tipo conduz a um espaço infinito de soluções para o problema da determinação dos parâmetros de uma recta.

Como normalmente não estamos perante rectas infinitas mas sim segmentos de recta, as alternativas para a representação de rectas podem ser os pontos extremos do segmento de recta. Mas novamente essa situação não conduz à forma mínima e em termos de imagem as extremidades dos segmentos de recta não são primitivas estáveis pois um segmento de uma recta numa imagem pode aparecer completo enquanto em outra pode já aparecer "partido". Uma alternativa a esta forma de representação consiste em expressar a orientação \mathbf{o} por dois ângulos: respectivamente o azimute ϕ e a elevação α .

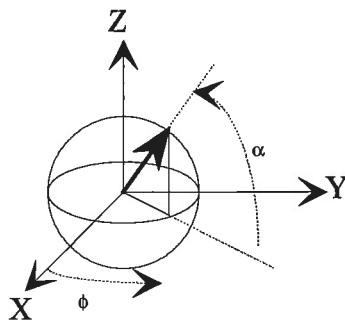


Figura 2.16 - Ângulos utilizados para descrever a orientação de uma recta em três dimensões

Neste caso a orientação da recta será dada por:

$$\mathbf{o} = (o_x, o_y, o_z) = (\cos(\phi) \cos(\alpha), \sin(\phi) \cos(\alpha), \sin(\alpha)) \quad (2.31)$$

mas novamente não estamos perante uma forma mínima pois são necessários cinco parâmetros: as 3 coordenadas de \mathbf{P} e mais 2 ângulos.

Outra forma alternativa explorada no sentido de minimizar o número de parâmetros foi proposta por Plücker na qual uma recta é representada pela intersecção de dois planos (em [Roberts 88]). Neste caso uma recta r é expressa por:

$$\{ (x, y, z) \mid x = \alpha z + l, y = \beta z + a \} \quad (2.32)$$

Os 4 parâmetros de uma recta podem ser representados pelo vector $v = (\alpha, \beta, l, a)$ em que (l, a) representa o ponto de intersecção da recta no plano-xy. Novamente esta representação não conduz a uma solução única pois não permite representar rectas perpendiculares ao eixo-z. Para esses casos é necessário utilizar outra representação alternativa que pode ser:

$$\{ (x, y, z) \mid y = \alpha x + l, z = \beta x + a \} \quad (2.33)$$

mas que também não permite representações para rectas perpendiculares ao eixo-x ou paralelas ao plano-yz, ou ainda a representação

$$\{ (x, y, z) \mid x = \alpha y + l, z = \beta y + a \} \quad (2.34)$$

mas que também não permite representações de rectas perpendiculares ao eixo-y. Este tipo de representações foram as utilizadas por Ayache nos seus trabalhos [Ayache 89]

Uma nota importante neste tipo de representações é que estas são apropriadas *para rectas* e não *para segmentos de recta*. Neste caso para representar um segmento de recta é necessário fornecer também um ponto inicial e final do segmento (+6 parâmetros) ou um ponto inicial e o seu comprimento (+4 parâmetros).

2.5.3 Planos

Um plano pode ser definido através de três parâmetros que constitui a forma mínima para a sua representação. Uma representação possível para um plano é através das componentes (n_x, n_y) da normal \mathbf{n} (com $\|\mathbf{n}\|=1$) ao plano e a distância d à origem do sistema de coordenadas, medida segundo essa normal. No entanto não é uma forma única pois $-\mathbf{n}$ também é normal ao plano. Para restringir o problema podemos convencionar que uma das componentes de \mathbf{n} é positiva. Se considerarmos n_z positivo e diferente de zero então não podemos representar planos paralelos ao eixo-z e neste caso teremos que escolher outra

convenção - por exemplo n_x ou n_y positivo. Sendo n_x ou n_y positivos e diferentes de zero, temos novamente o problema da representação de planos paralelos aos eixos-x e y. Deste modo podemos ter as seguintes representações para os planos

$$\begin{cases} (\mathbf{n}, d) & \text{com } n_z > 0 \Rightarrow \text{não representa planos paralelos ao eixo- z} \\ (\mathbf{n}, d) & \text{com } n_x > 0 \Rightarrow \text{não representa planos paralelos ao eixo- x} \\ (\mathbf{n}, d) & \text{com } n_y > 0 \Rightarrow \text{não representa planos paralelos ao eixo- y} \end{cases} \quad (2.35)$$

Qualquer plano no espaço pode também ser representado na forma

$$a_x x + a_y y + a_z z = c \quad (2.36).$$

Definindo $\mathbf{p} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ como um vector correspondente à posição de qualquer ponto do plano e $\mathbf{a} = a_x\mathbf{i} + a_y\mathbf{j} + a_z\mathbf{k}$ um vector segundo uma normal ao plano podemos exprimir (2.36) na seguinte forma vectorial

$$\mathbf{a}^T \mathbf{p} = c \quad (2.37).$$

Sendo $\|\mathbf{a}\| \neq 0$ então o vector \mathbf{n} unitário com direcção normal ao plano será dado por

$$\mathbf{n} = \frac{\mathbf{a}}{\|\mathbf{a}\|}.$$

Dividindo (2.37) por $\|\mathbf{a}\|$ obtemos a expressão

$$\mathbf{n}^T \mathbf{p} = d \quad \text{com } d = \frac{c}{\|\mathbf{a}\|} \quad (2.38)$$

na qual d representa a distância do plano à origem do sistema de coordenadas. Deste modo a expressão (2.36) equivale à representação para um plano utilizando a normal \mathbf{n} e a distância d mas sem a restrição $\|\mathbf{n}\| = 1$. Isso permite estabelecer uma alternativa à expressão (2.35) para representar um plano numa forma mínima através de um triplo (a, b, c) definindo o valor de uma das componentes do vector \mathbf{a} . Assim, por semelhança com a expressão (2.35) e utilizando a expressão (2.36), colocamos a coordenada que pretendemos positiva com valor 1, obtendo a expressão

$$\begin{cases} ax + by + z = c & \Rightarrow \text{não representa planos paralelos ao eixo- z} \\ x + ay + bz = c & \Rightarrow \text{não representa planos paralelos ao eixo- x} \\ ax + y + bz = c & \Rightarrow \text{não representa planos paralelos ao eixo- y} \end{cases} \quad (2.39)$$

2.5.4 Equação Geral de Superfícies Quádricas

Tal como foi mencionado no início deste capítulo, as primitivas geométricas envolvendo superfícies com curvatura não são utilizadas neste trabalho. No entanto, para dar um desenvolvimento mais completo do conjunto de representações, faz-se aqui uma referência para a expressão geral de uma superfície quádrlica que permite representar um conjunto vasto de superfícies.

Esta expressão geral para a representação de diferentes tipos de quádrlicas é conhecida pela equação geral das quádrlicas

$$F(x,y,z) = Ax^2 + By^2 + Cz^2 + 2Dxy + 2Exz + 2Gyz + 2Kx + 2Ly + 2Mz + N = 0 \quad (2.40)$$

que pode ser descrita na forma vectorial $\mathbf{p}^T \mathbf{Q} \mathbf{p} = 0$ em que

$$\mathbf{p} = \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad \mathbf{Q} = \begin{bmatrix} A & D & E & K \\ D & B & G & L \\ E & G & C & M \\ K & L & M & N \end{bmatrix}.$$

Esta equação pode representar diferentes tipos de superfícies nas quais se incluem esferas, cilindros, cones, elipsóides, parabolóides, hiperbolóides e planos. Nesta forma de representação, uma esfera $x^2 + y^2 + z^2 = r^2$ é representada por uma matriz \mathbf{Q} diagonal com $A=B=C=1$ e $N = -r^2$. Se definirmos os seguintes parâmetros

$$I_1 = A + B + C \quad I_2 = \begin{vmatrix} B & G \\ G & C \end{vmatrix} + \begin{vmatrix} A & E \\ E & C \end{vmatrix} + \begin{vmatrix} A & D \\ D & C \end{vmatrix}$$

$$I_3 = \begin{vmatrix} A & D & E \\ D & B & G \\ E & G & C \end{vmatrix} \quad I_4 = \begin{vmatrix} A & D & E & K \\ D & B & G & L \\ E & G & C & M \\ K & L & M & N \end{vmatrix}$$

teremos um elipsóide quando $I_4 < 0$ e $I_1, I_2, I_3 > 0$ e um cilindro quando I_4 e $I_3 = 0$ [Santos 70]. Apesar da elegância desta forma geral de representação de uma superfície, não é a mais fácil de manipular nem a mais útil na representação de um universo de imagens reais onde normalmente existem múltiplos objectos colocando um grau de complexidade muito grande na determinação dos parâmetros destas primitivas.

2.6. Invariantes das Representações Tridimensionais

À medida que uma câmara se movimenta no espaço tridimensional a imagem obtida pelo sensor sofre deformação contínua. Se o modelo utilizado para descrever a formação de imagem é a projecção em perspectiva, estas deformações podem ser analisadas utilizando transformações projectivas. Algumas medidas descritivas da relação entre primitivas tais como, a colinearidade, a coplanaridade, a convexidade, a posição relativa e a razão dupla entre 4 pontos colineares são medidas invariantes nesta forma de projecção. Em seguida descrevem-se quatro destes invariantes que podem ser utilizados na análise das imagens para determinar grupos de primitivas.

Grupos direccionais são conjuntos de segmentos de recta paralelos no espaço e que podem ser detectados na imagem. Estes grupos direccionais são constituídos por segmentos que têm a propriedade de convergirem todos no mesmo ponto de fuga (convergirem para o mesmo ponto na linha do horizonte). Em imagens reais de interiores de edifícios, normalmente podemos definir dois grandes grupos direccionais: grupos de segmentos perpendiculares entre si e grupos de segmentos horizontais visto que existe grande número deste tipo de segmentos.

Linhas ou conjunto colinear de pontos que podem pertencer a um grupo direccional. O ordenamento por grupos direccionais pode ser então feito testando a colinearidade entre dois quaisquer segmentos s e s' testando se

$$|\hat{\text{ângulo}}(s) - \hat{\text{ângulo}}(s')| < \text{limiar}.$$

Junções ou pares de segmentos de recta que partilham um ponto comum. Se s e s' forem dois segmentos e pretendermos verificar se estes se interceptam num ponto i então testamos

$$d + d' < \text{limiar}$$

sendo d e d' as distâncias mínimas de entre os segmentos.

No capítulo 4, dedicado à detecção de primitivas em imagens, é proposto um algoritmo de detecção de grupos baseados em medidas sobre os segmentos de recta de uma imagem.

A *razão cruzada entre quatro pontos colineares* é um invariante básico em geometria projectiva [Duda 73][Penna 86] a partir da qual se podem derivar outros invariantes projectivos [Mohr 92]. Esta descoberta é já bastante antiga e prova-se que qualquer

transformação homográfica preserva esta razão invariante [Mohr 92,pag 444]. Uma transformação homográfica é qualquer transformação linear expressa em coordenadas homogêneas, a qual inclui a projecção em perspectiva, mudança de escala, rotação, translação, Consideremos, tal como mostra a figura 2.17, um conjunto de rectas coplanares S_1, S_2, S_3, S_4 , com centro em $C=(-a, b, 0)$ e cruzando o eixo-x e eixo-y de um sistema de coordenadas ortonormadas. Esse conjunto de rectas cruza o eixo-x nos pontos $(x_i, 0, 0)$ com $i=1 \dots 4$ e o eixo-y nos pontos $(0, y_i, 0)$ com $i=1 \dots 4$. Por relação de triângulos podemos verificar que para qualquer das rectas S_i com $i=1 \dots 4$, se verifica a condição seguinte

$$1 = \frac{b}{y_i} + \frac{-a}{x_i}.$$

Subtraindo esta relação para uma recta i pela relação para a recta j obtemos

$$b \left(\frac{y_j - y_i}{y_j y_i} \right) = a \left(\frac{x_j - x_i}{x_j x_i} \right) \quad (2.41).$$

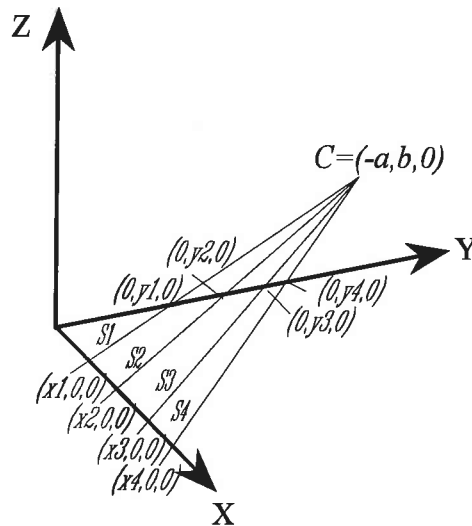


Figura 2.17 - Conjunto de rectas coplanares S_1, S_2, S_3, S_4 , com centro em $C=(-a, b)$ e cruzando o eixo-x e eixo-y de um sistema de coordenadas.

Consideremos o par (j, i) como designando o segmento j e o segmento i . Se dividirmos o produto das equações (2.41) para os segmentos (S_2, S_4) e (S_3, S_1) pelo produto das mesmas equações para os segmentos (S_1, S_2) e (S_3, S_4) , obtemos a razão cruzada

$$\frac{(x_3 - x_1)(x_2 - x_4)}{(x_2 - x_1)(x_3 - x_4)} = \frac{(y_3 - y_1)(y_2 - y_4)}{(y_2 - y_1)(y_3 - y_4)} \quad (2.42).$$

A expressão (2.42) relaciona as distâncias entre coordenadas de pontos mas não depende da posição do centro de perspectiva C . Para além disso, mostra que duas secções de um conjunto de rectas coplanares S_1, S_2, S_3, S_4 , com centro em C , têm sempre a mesma razão cruzada. Consideremos, tal como mostra a figura 2.18, uma recta R no espaço tridimensional. Essa recta é cruzada nos pontos (M, N, A, B) , por um conjunto de rectas coplanares S'_1, S'_2, S'_3, S'_4 , com centro em C' e por um conjunto de rectas coplanares $S''_1, S''_2, S''_3, S''_4$, com centro em C'' . Os pontos de cruzamento (M, N, A, B) de R têm imagens (M', N', A', B') e (M'', N'', A'', B'') correspondentes (em *correspondência projectiva*) as quais foram obtidas por centros de perspectiva C' e C'' colocados em posições diferentes no espaço. Neste caso a razão cruzada

$$\frac{\overline{N'B'}}{\overline{A'B'}} \frac{\overline{A'M'}}{\overline{N'M'}} = \frac{\overline{N''B''}}{\overline{A''B''}} \frac{\overline{A''M''}}{\overline{N''M''}}$$

é um invariante para qualquer dos locais onde o centro de perspectiva seja colocado [Mohr 92].

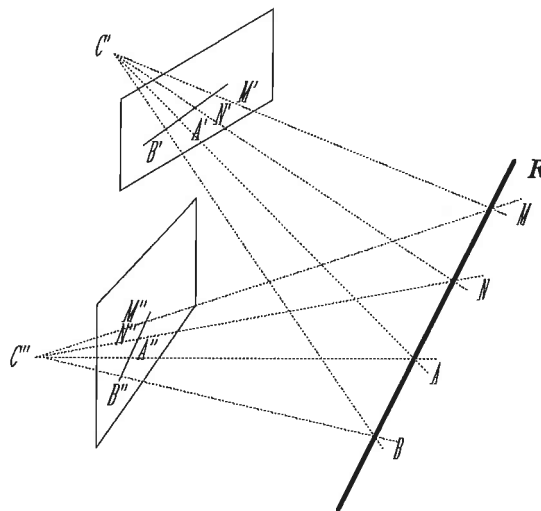


Figura 2.18 - A razão dupla é invariante qualquer que seja o local onde seja colocada o centro de perspectiva.

CAPÍTULO 3

Geometria do sistema de visão

3.1 Introdução

Em grande parte do trabalho experimental realizado foi utilizado um sistema de visão com duas câmaras vídeo colocadas na extremidade de um manipulador com 6 graus de liberdade. Através do controlo do manipulador é possível movimentar e posicionar o sistema de visão no universo de trabalho do manipulador. Como o movimento é controlado, é possível analisar o seu efeito nas imagens. Este capítulo aborda esta questão e ainda a da calibração do sistema. No primeiro ponto serão revistos os conceitos associados ao movimento do sistema de visão enquanto no segundo será abordado o problema da sua calibração. No último ponto serão apresentados o estudo analítico do efeito do movimento nas representações geométricas utilizadas.

3.2 Geometria do Movimento do Sistema de Visão

Introdução

Na representação geométrica do espaço tridimensional é muito comum utilizarem-se coordenadas homogéneas. Nesta representação os pontos e direcções no espaço, são representados em coordenadas homogéneas por vectores de dimensão (4×1) tal como foi referido anteriormente. As coordenadas homogéneas também permitem representar, de uma forma mais compacta, as transformações entre referenciais [Paul 81][Fu 87]. As transformações são aplicações que permitem mudança de um referencial para outro (mudança de espaço), e no caso de coordenadas homogéneas são representadas por matrizes

de dimensão (4x4). Uma transformação entre referenciais inclui as coordenadas da posição da origem de um referencial relativamente ao outro (translação) assim como a orientação dos eixos de um dos sistema de coordenadas relativamente a outro (rotação). Em coordenadas homogéneas, a matriz de transformação tem uma submatriz correspondente à rotação e um vector correspondente à translação. Uma transformação **T** entre referenciais é representada em coordenadas homogéneas por

$$\mathbf{T}_{(4 \times 4)} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_{13} \\ r_{21} & r_{22} & r_{23} & t_{23} \\ r_{31} & r_{32} & r_{33} & t_{33} \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{Rot}_{(3 \times 3)} & \mathbf{Trans}_{(3 \times 1)} \\ \mathbf{0}_{(1 \times 3)} & 1 \end{bmatrix} \quad (3.1).$$

A última linha da matriz **T** é representada por [0 0 0 1] caso a transformação não envolva mudança de escala ou não represente uma transformação de perspectiva.

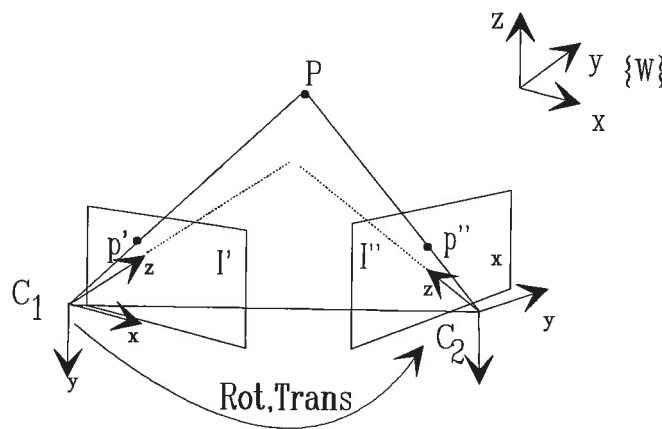


Figura 3.1 - Transformação de coordenadas correspondente à mudança da posição da câmara da posição C1 para C2.

A notação, utilizada ao longo do texto, para representar vectores e transformações em coordenadas homogéneas é a seguinte:

v - vector coluna;

T - matriz de transformação de coordenadas;

${}^0\mathbf{T}_1$ - matriz de transformação (4x4) que relaciona o sistema de coordenadas 1 com o sistema de coordenadas 0;

t_{ij} - elemento da linha *i* e coluna *j* da matriz de transformação **T**;

Trans - vector (3x1) que representa uma translação de uma quantidade *x*, *y* e *z* ao longo dos eixo-*x*, *y* e *z* respectivamente;

Rot(θ ,eixo) ou **Rot** - matriz de transformação que representa uma rotação de um ângulo θ em torno de um eixo (p. ex. x, y ou z). O sentido positivo da rotação é o sentido contrário ao movimento dos ponteiros do relógio.

Transformações utilizadas no sistema experimental

Grande parte das experiências foram realizadas com um manipulador PUMA 560 da UNIMATION de 6 graus de liberdade na extremidade do qual foi colocado o sistema de visão. Cada grau de liberdade é obtido por uma rotação segundo um eixo permitindo o posicionamento do sistema de coordenadas solidário com a extremidade do manipulador, em qualquer ponto do espaço de trabalho.



Figura 3.2 - Sistema de visão colocado na extremidade do manipulador.

Os principais parâmetros geométricos e transformações dos sistema são os seguintes:

- {BASE} - Referencial solidário com a base do manipulador. É fixo enquanto é executado um movimento e pode ser modificada a sua posição pelo utilizador;
- {W} - Referencial fixo colocado no local onde se encontra o manipulador e que representa o referencial global;
- {CAM_i} - Referencial que representa o sistema de coordenadas 3D colocado na câmara *i*, sendo *i*=ESQUERDA, DIREITA e com o eixo-z coincidente com o eixo óptico;
- ${}^{BASE}T_6$ - Transformação que relaciona a base do manipulador e a ponta da sua última ligação mecânica;

TOOL - Transformação que representa a geometria da ferramenta e que relaciona o sistema de coordenadas na ponta dessa ferramenta com a última ligação do manipulador;

${}^{TOOL}T_{CAM_i}$ - Transformação que relaciona o sistema de coordenadas na ponta da ferramenta e o sistema de coordenadas 3D colocado na câmara i , sendo $i=ES$ Querda, DIR eita;

${}^W T_{CAM_i}$ - Transformação que relaciona o sistema de coordenadas 3D colocado na câmara e o referencial global estático;

${}^W T_{BASE}$ - Transformação que relaciona o referencial solidário com a base do manipulador e o referencial global estático;

$CAM_{ESQ}CALIB$ - Transformação (3x4) que representa a mudança de perspectiva da imagem ES Querda correspondente à transformação de coordenadas 3D em coordenadas 2D na imagem ES Querda;

$CAM_{DIR}CALIB$ - Transformação (3x4) que representa a mudança de perspectiva da imagem DIR eita correspondente à transformação de coordenadas 3D em coordenadas 2D na imagem DIR eita;

$CAM_{ESQ}DELTA_CAM_{CAM_{DIR}}$ - Transformação que representa a relação entre o sistema de coordenadas 3D da câmara DIR eita e um sistema de coordenadas 3D da câmara ES Querda.

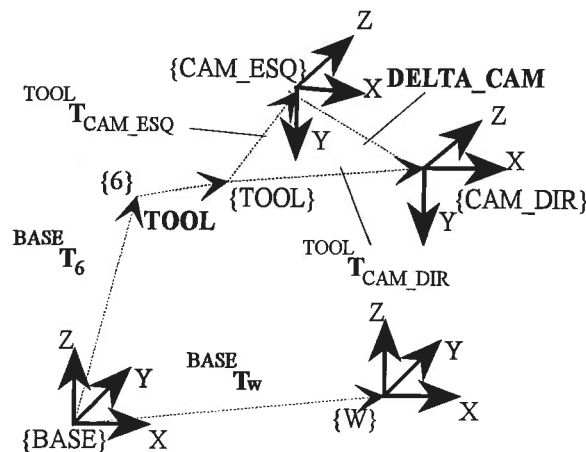


Figura 3.3 - Relação entre as várias transformações e os vários sistemas de coordenadas

As relações geométricas entre os vários referenciais estão representadas, de forma esquemática, na figura 3.3. É possível relacionar todas estas transformações pela seguinte equação

$${}^W T_{CAM_i} = {}^W T_{BASE} \cdot {}^{BASE} T_6 \cdot TOOL \cdot {}^{TOOL} T_{CAM_i} \quad (3.2)$$

com $i=ES$ ou DIR .

Para o nosso trabalho interessa saber que a transformação ${}^{BASE}T_6$ e transformação ${}^{BASE}T_{TOOL}$ são possíveis de obter directamente pela linguagem de controlo do manipulador (VAL II) através do comando WHERE.

Cinemática do sistema de visão

Cada uma das câmaras utilizadas no sistema comporta-se como um corpo rígido, sendo possível estudar o seu movimento como tal. Um corpo rígido é uma colecção de partículas cuja distância entre si é invariante. Na figura 3.4 está representado um corpo no qual foram escolhidos três pontos P_1 , P_2 e P_3 fixos e não-colineares.

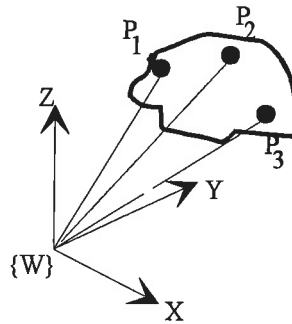


Figura 3.4 - Pontos P_i fixos num corpo rígido e descritos num referencial $\{W\}$.

Para definir a posição destes pontos no espaço é necessário atribuir valores para as coordenadas (x,y,z) de cada um dos três pontos correspondendo a nove graus de liberdade para atribuição dos valores. Se os pontos tiverem coordenadas fixas relativamente a um referencial $\{W\}$ estático, então o corpo não se moverá relativamente a $\{W\}$. Para o corpo se poder movimentar é necessário que o valor de alguma destas coordenadas varie dinamicamente quando medidas relativamente ao referencial $\{W\}$. Se um dos pontos no corpo se move sem restrição no espaço, então o corpo movimenta-se com os três graus de liberdade correspondentes aos valores das suas coordenadas (x,y,z) desse ponto. Juntando um segundo ponto, o movimento do corpo passará a ter mais dois graus de liberdade pois, sendo o corpo rígido, a distância entre os dois pontos é sempre constante. Um terceiro ponto adiciona unicamente mais um grau de liberdade porque a distância aos outros dois pontos é constante. A introdução de pontos adicionais para descrever o movimento não introduz graus de liberdade adicionais ao movimento porque, sendo o corpo rígido, as distâncias aos três primeiros pontos é sempre constante. Estas restrições podem ser descritas pelas expressões (3.3) em que d_{ij} expressa a distância entre os pontos P_i e P_j . Deste modo o movimento de três pontos não-colineares de um corpo rígido determina o seu movimento completamente e esse movimento tem 6 graus de liberdade.

$$\begin{aligned}
 (x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2 &= d_{21}^2 \\
 (x_3 - x_1)^2 + (y_3 - y_1)^2 + (z_3 - z_1)^2 &= d_{31}^2 \\
 (x_2 - x_3)^2 + (y_2 - y_3)^2 + (z_2 - z_3)^2 &= d_{23}^2
 \end{aligned}
 \tag{3.3}.$$

Suponha-se que se fixam, de uma forma estática, as coordenadas de um dos pontos P_i . Nesta situação os movimentos possíveis do corpo rígido são restritos a movimentos de rotação segundo eixos que passem pelo ponto escolhido. Este caso corresponde a fixar valores para três dos seis graus de liberdade anteriores, sendo os restantes responsáveis pela rotação do corpo. Como conclusão, podemos dizer que dos seis graus de liberdade anteriores, três correspondem à translação do corpo rígido e os restantes à sua rotação. Outra maneira de encarar o problema consiste em considerar um referencial ortogonal e solidário com o corpo rígido. Se $\{CAM\}$ representar o referencial solidário ao corpo rígido e $\{W\}$ representar um referencial estático então o comportamento dinâmico da origem do referencial $\{CAM\}$, medido relativamente ao referencial $\{W\}$, corresponde ao movimento de translação. O comportamento dinâmico das orientações dos eixos do referencial $\{CAM\}$, corresponde ao movimento de rotação.

Velocidade do sistema de visão

As experiências realizadas envolvem a movimentação das câmaras. Para obter expressões para sua velocidade e aceleração partimos do princípio que as câmaras se comportam como um corpo rígido.

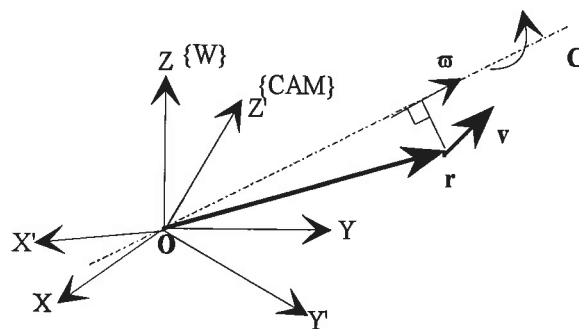


Figura 3.5 - Vector de velocidade angular de um corpo rígido

Primeiro vamos considerar que o corpo só está sujeito a um movimento de rotação. Consideremos, tal como se mostra na figura 3.5, dois referenciais ortogonais, $\{W\}$ e $\{CAM\}$, com origens coincidentes. O referencial $\{W\}$ é estático e o referencial $\{CAM\}$, de uma câmara, está sujeito a um movimento de rotação em torno de um eixo \overline{OC} . Esse eixo passa pela origem O e o movimento tem uma velocidade angular instantânea dada por um

vector $\boldsymbol{\omega}$ com suporte no eixo \overline{OC} . Se \mathbf{P} for um ponto fixo nesse corpo representado por um vector $\mathbf{r} = [x \ y \ z]^T$ então a velocidade \mathbf{v} desse ponto \mathbf{P} pode ser expressa por

$$\mathbf{v} = \boldsymbol{\omega} \wedge \mathbf{r} \quad (3.4).$$

Escrevendo os vectores $\boldsymbol{\omega} = [\omega_x, \omega_y, \omega_z]^T$ e $\mathbf{r} = [x \ y \ z]^T$ em coordenadas do referencial $\{\text{CAM}\}$ teremos a seguinte expressão para a velocidade

$$\mathbf{v} = \boldsymbol{\omega} \wedge \mathbf{r} = (\omega_y z - \omega_z y) \hat{\mathbf{i}} + (\omega_z x - \omega_x z) \hat{\mathbf{j}} + (\omega_x y - \omega_y x) \hat{\mathbf{k}} \quad (3.5)$$

que consiste num vector perpendicular ao plano formado por $\boldsymbol{\omega}$ e \mathbf{r} e com direcção dada pela definição de produto vectorial.

Estas rotações podem ser expressas também de uma forma matricial mas utilizando rotações infinitesimais. Esta forma de representação é compacta e simplifica muitas operações. A rotação ilustrada na figura 3.5 pode ser descrita em termos de ângulos de Euler através da representação RPY ("Roll, Pitch, Yaw") [Fu 87]. Esta rotação corresponde às seguintes rotações em sequência:

- rotação θ_x em torno do eixo-x ("yaw");
- rotação θ_y em torno do eixo-y ("pitch");
- rotação θ_z em torno do eixo-z ("roll").

Essa rotação equivale a uma rotação ${}^W\mathbf{Rot}_{\text{CAM}}$ que em termos matriciais é

$$\begin{aligned} {}^W\mathbf{Rot}_{\text{CAM}} &= \text{Rot}(z, \theta_z) \text{Rot}(y, \theta_y) \text{Rot}(x, \theta_x) \\ &= \begin{bmatrix} \cos \theta_z & -\sin \theta_z & 0 \\ \sin \theta_z & \cos \theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos \theta_y & 0 & \sin \theta_y \\ 0 & 1 & 0 \\ -\sin \theta_y & 0 & \cos \theta_y \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_x & -\sin \theta_x \\ 0 & \sin \theta_x & \cos \theta_x \end{bmatrix} \end{aligned} \quad (3.6).$$

Considerando que os ângulos de rotação são infinitesimais podemos aproximar $\sin \delta\theta \approx \delta\theta$ e $\cos \delta\theta \approx 1$, desprezando os termos de ordem mais alta. Aplicando este princípio à expressão (3.6) obtemos uma expressão matricial para a rotação ${}^W\mathbf{Rot}_{\text{CAM}}$ que será dada por

$${}^W\mathbf{R}_{\text{CAM}} = \begin{bmatrix} 1 & -\delta\theta_z & \delta\theta_y \\ \delta\theta_z & 1 & -\delta\theta_x \\ -\delta\theta_y & \delta\theta_x & 1 \end{bmatrix} = \begin{bmatrix} 1 & -\omega_z & \omega_y \\ \omega_z & 1 & -\omega_x \\ -\omega_y & \omega_x & 1 \end{bmatrix} \quad (3.7)$$

$$= \mathbf{I} + \Delta \mathbf{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix} \quad (3.8).$$

Deste modo um vector \mathbf{r} representando um ponto \mathbf{P} da câmara e expresso no referencial da câmara, irá ter um deslocamento incremental $\Delta \mathbf{r}$ da sua posição relativa ao referencial estático $\{W\}$. Essa mudança incremental será resultado da rotação incremental com componentes $\boldsymbol{\omega} = (\omega_x, \omega_y, \omega_z)$ do sistema $\{CAM\}$ em movimento. Se $\mathbf{r} = (x, y, z)^T$ então as suas componentes expressas no referencial $\{W\}$ serão dadas por

$$\begin{aligned} \mathbf{R} \cdot \mathbf{r} &= \begin{bmatrix} 1 & -\omega_z & \omega_y \\ \omega_z & 1 & -\omega_x \\ -\omega_y & \omega_x & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} z\omega_y - y\omega_z \\ x\omega_z - z\omega_x \\ y\omega_x - x\omega_y \end{bmatrix} \\ &= \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ \omega_x & \omega_y & \omega_z \\ x & y & z \end{vmatrix} = \mathbf{r} + \boldsymbol{\omega} \wedge \mathbf{r} \end{aligned} \quad (3.9).$$

Adicionando ao movimento de rotação do referencial $\{CAM\}$ um movimento de translação, podemos obter uma expressão para a sua velocidade. Essa expressão utiliza como referencial estático o referencial $\{W\}$. Um ponto \mathbf{P} , na câmara, é representado no referencial $\{CAM\}$ pelo vector \mathbf{r}' e no referencial $\{W\}$ pelo vector \mathbf{r} . A translação entre o referencial $\{CAM\}$ e o referencial $\{W\}$ é dado pelo vector \mathbf{s} que se relaciona com os vectores \mathbf{r}' e \mathbf{r} por (figura 3.6)

$$\mathbf{r} = \mathbf{r}' + \mathbf{s}.$$

A velocidade instantânea do ponto \mathbf{P} expressa no referencial $\{W\}$ é dada pela derivada temporal de \mathbf{r}

$$\mathbf{v} = \frac{d\mathbf{r}}{dt} = \frac{d\mathbf{r}'}{dt} + \frac{d\mathbf{s}}{dt}$$

e utilizando a expressão (3.4), pode ser expressa por

$$\mathbf{v} = \frac{d\mathbf{s}}{dt} + \boldsymbol{\omega} \wedge \mathbf{r}'$$

com $\mathbf{r}' = \frac{dx'}{dt} \hat{\mathbf{i}} + \frac{dy'}{dt} \hat{\mathbf{j}} + \frac{dz'}{dt} \hat{\mathbf{k}}$ [Fu 87].

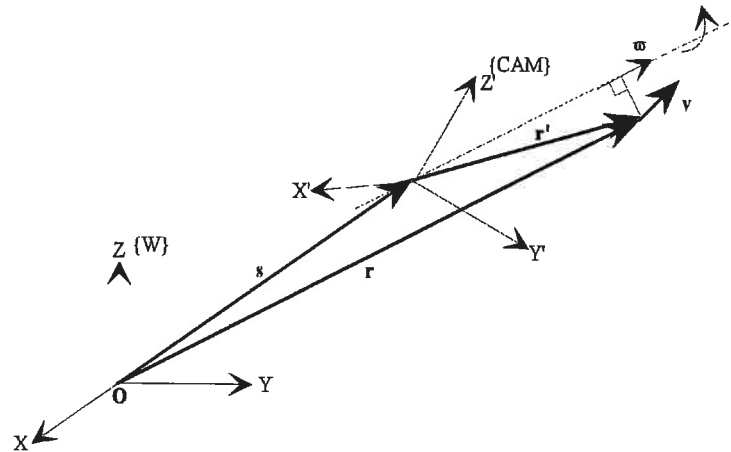


Figura 3.6 - Sistema de coordenadas sujeito aos movimentos de rotação e translação

Se o referencial {CAM} está solidário com o corpo sólido da câmara então o movimento pode ser descrito em termos da translação da origem de {CAM} e da sua rotação relativa ao sistema {W}. Supondo que a velocidade de translação do referencial {CAM} é dada pelo vector \mathbf{v}_{Trans} e a velocidade $\dot{\mathbf{r}}' = 0$ então a velocidade total do ponto P relativamente ao referencial fixo {W} é dada por

$${}^W \mathbf{v} = {}^W \mathbf{v}_{Trans} + {}^W \boldsymbol{\omega} \wedge \mathbf{r} \quad (3.10)$$

Esta equação pode ser também expressa em termos de uma matriz de transformação \mathbf{T} , assumindo translações e rotações infinitesimais e respectivamente $\mathbf{v} = (\delta x, \delta y, \delta z)$ e $\boldsymbol{\omega} = (\omega_x, \omega_y, \omega_z)$. Essa transformação \mathbf{T} terá a seguinte representação matricial em coordenadas homogêneas

$$\mathbf{T} = \begin{bmatrix} 1 & -\omega_z & \omega_y & \delta x \\ \omega_z & 1 & -\omega_x & \delta y \\ -\omega_y & \omega_x & 1 & \delta z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3.11).$$

Assim, se $\mathbf{r}_{(t)} = (x, y, z, 1)^T$ representa, em coordenadas homogêneas, um ponto na câmara e expresso no referencial {CAM} no tempo (t) então esse ponto terá uma posição no tempo ($t + \delta t$) dada pela expressão

$$\begin{aligned} \mathbf{r}(t + \delta t) = \mathbf{T} \cdot \mathbf{r}(t) &= \begin{bmatrix} 1 & -\omega_z & \omega_y & \delta x \\ \omega_z & 1 & -\omega_x & \delta y \\ -\omega_y & \omega_x & 1 & \delta z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \\ &= \mathbf{r}(t) + \begin{bmatrix} \delta x \\ \delta y \\ \delta z \end{bmatrix} + \begin{bmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ \omega_x & \omega_y & \omega_z \\ x & y & z \end{bmatrix} = \mathbf{r}(t) + \mathbf{v}(t) + \boldsymbol{\omega} \wedge \mathbf{r}(t) \end{aligned} \quad (3.12)$$

Se \mathbf{v} é a velocidade do sistema de coordenadas {CAM} e se \mathbf{x}_0 é a posição inicial do referencial então a sua posição final \mathbf{x} ao fim de um tempo t deverá ser uma transformação ${}^W\mathbf{T}_{CAM}$ que relaciona os dois referenciais. Essa transformação poderá ser decomposta em duas transformações distintas e correspondentes a uma translação **Trans** e numa rotação **Rot** [Craig 86][Paul 81][Fu 87].

A matriz ${}^W\mathbf{T}_{CAM}$, que representa a posição e orientação do referencial {CAM} no final do movimento, é função da translação e da rotação. Esta matriz é

$${}^W\mathbf{T}_{CAM} = \begin{bmatrix} {}^W\mathbf{Rot}_{CAM} & {}^W\mathbf{Trans}_{CAM} \\ \mathbf{0}_{(1 \times 3)} & 1 \end{bmatrix} \quad (3.13).$$

Se o referencial considerado como estático for o referencial {CAM}, então o sistema de coordenadas {W} tem um movimento num sentido contrário ao anterior. Neste caso a expressão para a velocidade virá

$$\mathbf{v} = -{}^W\mathbf{v}_{Trans} - {}^{CAM}\boldsymbol{\omega} \wedge \mathbf{r}' \quad (3.14).$$

No âmbito deste trabalho o manipulador serve para posicionar as câmaras em diferentes pontos do espaço. Com este sistema podemos simular a execução da trajectória das câmaras como se estivessem sujeitas a um movimento de velocidade instantânea conhecida. Para isso consideramos o referencial {W} como estático e solidário com o manipulador e o referencial {CAM} como móvel e solidário com as câmaras. A partir do conhecimento da velocidade angular $\boldsymbol{\omega}$ e de translação \mathbf{v}_{Trans} do movimento podemos, através da expressão (3.12) obter a matriz ${}^W\mathbf{T}_{CAM}$ (3.13) que corresponde à posição das câmaras no instante $(t + \Delta t)$.

3.3 Calibração do Sistema de Visão

O sistema de visão utilizado nesta parte do trabalho consiste num par de câmaras vídeo mecanicamente solidárias entre si e a "tool" do manipulador. O sistema não tem nenhuma geometria pré-definida para os referenciais das câmaras, sendo a sua vergência fixa desde o momento em que as câmaras são instaladas. O processo de determinar a relação entre referenciais (translação, orientação) e os parâmetros internos das câmaras (comprimento focal e mudança de escala) é conhecido por calibração. Neste ponto será abordado inicialmente a forma como é realizada a calibração do sistema numa posição fixa do espaço de trabalho do manipulador. Na realidade é com o processo de calibração que se determinam os valores dos parâmetros do modelo da câmara descrito no capítulo 2 e expresso pela matriz (2.13). Tal como foi referido, esse modelo explica o processo de projecção em perspectiva e dele deverão fazer parte as constantes responsáveis pela mudança de escala entre os sistemas de coordenadas "pixel" em 2D e metros em 3D. Para além disso um ponto na cena normalmente tem coordenadas conhecidas relativamente ao referencial {W} e não ao referencial {CAM}. Desse modo, este modelo deverá envolver todos esses parâmetros.

Transformação de perspectiva entre referencial {CAM} e o plano imagem ideal

A transformação de perspectiva dos pontos do referencial tridimensional para referencial bidimensional que contém o plano de imagem ideal, é dado pelas equações

$$\begin{aligned}x_u &= f \frac{x}{z} \\ y_u &= f \frac{y}{z}\end{aligned}\tag{3.15}$$

em que (x_u, y_u) são as coordenadas do ponto no plano imagem e f corresponde à *distância focal*.

Relação entre os pontos do plano imagem e os "pixels" da imagem amostrada

Esta transformação é de grande importância pois dá a relação entre a unidade de medida utilizada na imagem amostrada pelo computador (os "pixels") e a unidade de medida do referencial tridimensional {CAM}. Os parâmetros (x_u, y_u) representam o número da linha e coluna do "pixel" da imagem em computador, (u_0, v_0) o número da linha e coluna do

"pixel" do centro da imagem em computador e (S_x, S_y) os factores de escala segundo os eixos x e y . Os parâmetros f, S_x, S_y assim como o centro (u_0, v_0) são chamados *parâmetros intrínsecos* da câmara.

$$\begin{cases} u = S_x x_u + u_0 \\ v = S_y y_u + v_0 \end{cases} \quad (3.16)$$

Uma relação entre os parâmetros intrínsecos e extrínsecos da câmara podem ser obtidas pela manipulação das equações anteriores. A relação final entre os pontos 2D na matriz da imagem em computador e os pontos 3D referenciados ao sistema tridimensional colocado na câmara será então dada pela matriz

$$\begin{bmatrix} su \\ sv \\ s \end{bmatrix} = \begin{bmatrix} f.S_x & 0 & u_0 & 0 \\ 0 & f.S_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{Rot}_{(3 \times 3)} \\ \mathbf{0}_{(1 \times 3)} \\ 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \mathbf{CALIB} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (3.17).$$

Para obter a matriz de calibração é utilizada a seguinte metodologia: no local visualizado pela câmara é colocada uma grelha de pontos dos quais se conhecem as suas coordenadas medidas relativamente a um referencial 3D previamente escolhido - ver figura 3.7.

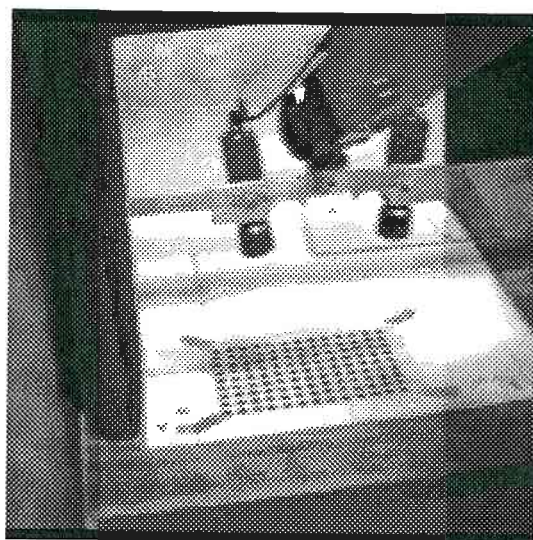


Figura 3.7 - Local de calibração do sistema de visão

As imagens desses pontos deverão ser bastante contrastadas relativamente ao "fundo" da grelha para que a segmentação da imagem e a detecção dos pontos seja fácil. Deste modo, ficam disponíveis pontos dos quais se conhecem as suas coordenadas

tridimensionais (em metros) e as suas coordenadas correspondentes na imagem (em "pixel"). Para obter os valores dos elementos da matriz **CALIB** com base nestes dados, foram ensaiados vários processos [Faugeras 87] [Tsai 88] [Dias 91] [Batista 92].

Matriz de calibração invariante

Como os parâmetros obtidos pela calibração das duas câmaras só são válidos para o local onde foi realizada, é necessário determinar os parâmetros que a tornam invariante e válida em qualquer ponto do espaço. A partir da equação (3.2) e para qualquer uma das câmaras $i=ESQ$ ou DIR , a matriz inversa de ${}^W T_{CAM_i}$ é dada pela expressão

$${}^{CAM_i} T_W = {}^{CAM_i} T_{TOOL} \cdot [(TOOL)^{-1} \cdot {}^6 T_{BASE} \cdot {}^{BASE} T_W]$$

A partir desta expressão obtemos a transformação entre o referencial $\{CAM_i\}$ solidário com a câmara $i=ESquerda$ ou $DIReita$ e o referencial $\{TOOL\}$ no qual as câmaras têm suporte. Esta expressão é dada por

$${}^{CAM_i} T_{TOOL} = {}^{CAM_i} T_W \cdot {}^W T_{BASE} \cdot {}^{BASE} T_6 \cdot TOOL \tag{3.18}$$

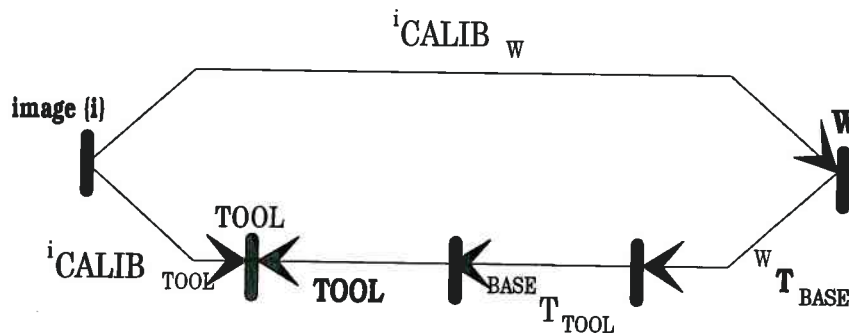


Figura 3.8 - Diagrama esquemático das relações geométricas expressas pela equação (3.11)

No caso de os pontos visualizados pelas câmaras serem conhecidos relativamente ao sistema de coordenadas $\{W\}$ é possível obter uma matriz de calibração ${}^i CALIB_W$ que relaciona pontos na imagem da câmara i e pontos no referencial $\{W\}$. Esta matriz é responsável pela mudança de perspectiva e pode ser conhecida para diferentes referenciais a partir da relação

$$\begin{aligned} {}^i CALIB_W &= {}^i CALIB_{TOOL} \cdot (TOOL)^{-1} \cdot {}^6 T_{BASE} \cdot {}^{BASE} T_W \\ &= {}^i CALIB_{TOOL} \cdot {}^{TOOL} T_W \end{aligned} \tag{3.19}$$

sendo

$${}^{\text{TOOL}}\mathbf{T}_W = (\mathbf{TOOL})^{-1} \cdot {}^6\mathbf{T}_{\text{BASE}} \cdot {}^{\text{BASE}}\mathbf{T}_W.$$

Note-se que ${}^i\text{CALIB}_{\text{TOOL}}$ é uma calibração que nunca varia com a posição ou movimento do manipulador podendo ser obtida através da relação

$${}^i\text{CALIB}_{\text{TOOL}} = {}^i\text{CALIB}_W \cdot {}^W\mathbf{T}_{\text{TOOL}} \quad (3.20).$$

Na calibração do sistema de visão pretende-se obter esta transformação para a câmara ESQUERDA e DIREITA com precisão elevada.

Um sistema com o objectivo de determinar as coordenadas 3D utilizando as duas câmaras, necessita da relação entre os dois sistemas de coordenadas 3D $\{\text{CAM_ESQ}\}$ e $\{\text{CAM_DIR}\}$. Esta relação, representada por ${}^{\text{CAM_ESQ}}\text{DELTA_CAM}_{\text{CAM_DIR}}$, pode ser obtida pela relação ${}^{\text{CAM}_i}\mathbf{T}_{\text{TOOL}}$ para a câmara ESQUERDA e DIREITA. Esta relação, pode ser obtida pela expressão

$${}^{\text{CAM_ESQ}}\text{DELTA_CAM}_{\text{CAM_DIR}} = ({}^{\text{TOOL}}\mathbf{T}_{\text{CAM_ESQ}})^{-1} \cdot {}^{\text{TOOL}}\mathbf{T}_{\text{CAM_DIR}} \quad (3.21)$$

A partir desta transformação e através da coluna correspondente à translação podemos obter o vector que une os dois sistemas de coordenadas CAM_ESQ e CAM_DIR .

A metodologia para obtenção da matriz de calibração ${}^i\text{CALIB}_W$ para cada uma das câmaras ESQUERDA e DIREITA é idêntica e, por isso, no seguimento do texto todas as deduções para uma câmara são válidas para ambas as câmaras.

3.4 Efeito do Movimento das Câmaras

3.4.1 Pontos

Se a transformação entre os referenciais das diferentes posições das câmaras for conhecido é possível saber a imagem de um ponto nessas diferentes posições. Sendo a equação de relação que descreve a passagem de coordenadas de um referencial $\{A\}$ para o referencial $\{B\}$ dada pela matriz ${}^B\mathbf{T}_A$ então o ponto ${}^A\mathbf{x}$ tem por expressão ${}^B\mathbf{x}$ no referencial $\{B\}$ obtida por

$${}^B\mathbf{x} = {}^B\mathbf{T}_A \cdot {}^A\mathbf{x} = \begin{bmatrix} \mathbf{Rot}_{(3 \times 3)} & \mathbf{Trans}_{(3 \times 1)} \\ \mathbf{0}_{(1 \times 3)} & 1 \end{bmatrix} {}^A\mathbf{x} \quad (3.22)$$

Se agora a matriz ${}^B\mathbf{T}_A$ relaciona as duas posições $\{A\}$ e $\{B\}$ da câmara, a imagem do ponto \mathbf{x} com a câmara colocada em $\{B\}$ terá uma projecção dada por

$$\begin{bmatrix} su \\ sv \\ s \end{bmatrix} = \text{CALIB} \begin{bmatrix} {}^B x \\ {}^B y \\ {}^B z \\ 1 \end{bmatrix} = \text{CALIB} {}^B \mathbf{x} \quad (3.23)$$

sendo a matriz **CALIB** a matriz de calibração da câmara em {B}.

3.4.2 Rectas

Uma recta tridimensional passando por um ponto **P** com direcção **o** pode ser expressa pela forma paramétrica (2.29)

$$\mathbf{r} = \mathbf{P} + t \mathbf{o} \quad \text{com } t \in [-\infty, +\infty]$$

Vamos supor que a recta é sujeita a um movimento que a coloca num local cuja relação com o anterior é dada através de uma transformação **T** composta por uma rotação **Rot** e uma translação **Trans**. Neste caso, qualquer ponto **r'** da recta na nova posição pode ser obtido a partir da expressão da recta na posição anterior e a matriz de transformação **T**, através de

$$\mathbf{r}' = (\text{Rot } \mathbf{P} + \text{Trans}) + t (\text{Rot } \mathbf{o}) \quad \text{com } t \in [-\infty, +\infty] \quad (3.24).$$

3.4.3 Planos

Se para a representação de um plano Π for utilizada a forma (2.35), em que $\mathbf{n} = (n_x \hat{\mathbf{i}} + n_y \hat{\mathbf{j}} + n_z \hat{\mathbf{k}})$ representa a normal ao plano e d a distância do plano à origem do sistema de coordenadas, então podemos utilizar uma forma compacta de representação através do vector

$$\Pi = [n_x \quad n_y \quad n_z \quad -d]^T \quad (3.25).$$

Se (3.25) for a representação de um plano então um ponto **P**(x,y,z) pertencente ao plano Π deverá obedecer à equação seguinte

$$\Pi \mathbf{P} = [n_x \quad n_y \quad n_z \quad -d]^T \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = 0 \quad (3.26).$$

Se ${}^B T_A$ for a transformação entre as posições {A} e {B} correspondente ao início e ao fim do movimento então o plano Π , representado no referencial {A}, terá uma representação Θ , no referencial {B}, dada por

$$\Theta = \Pi ({}^B T_A)^{-1} \quad (3.27).$$

Para provar (3.27) podemos considerar um ponto ${}^B Q$ no plano Θ e um ponto ${}^A P$ no plano Π . Ambos os pontos deverão obedecer à expressão (3.26) e portanto podemos estabelecer a igualdade

$$\Pi {}^A P = \Theta {}^B Q \quad (3.28).$$

Como ${}^B Q = {}^B T_A {}^A P$ então verifica-se que $\Pi = \Theta {}^B T_A$, o que permite demonstrar a relação (3.27).

CAPÍTULO 4

Detecção de primitivas em imagens- O caso de segmentos de recta

4.1. Introdução

Uma primitiva corresponde a um local na imagem e representa uma característica do sinal bidimensional. Neste capítulo são apresentados alguns processos para determinação de primitivas em imagens obtidas por câmaras. Mais particularmente são exploradas técnicas para determinação de segmentos de rectas nas imagens e determinação de grupos de primitivas que partilham características geométricas comuns. Essas características são a partilha de pontos comuns (vértices), grupos de rectas paralelas e conjuntos de rectas colineares. Estes segmentos de recta correspondem a arestas dos objectos ou a locais em que a intensidade luminosa varia bruscamente. Para nós estas características têm uma importância fulcral na percepção do ambiente que nos rodeia e além disso são bastante utilizadas para representar os objectos que nos rodeiam.

4.2. Detecção de Pontos Característicos

Na análise de uma imagem nem todos os seus pontos são importantes para a percepção dos objectos. Existem pontos com características e com inter-relações geométricas que permitem uma melhor percepção dos objectos. Os cantos ou intersecção de arestas dos objectos ou os pontos centrais de zonas com grande simetria são exemplos desses pontos. Para a detecção de primitivas nas imagens e correspondentes a estes "pontos chave" ou "pontos de interesse", têm sido exploradas duas aproximações: a utilização de operadores aplicados às imagens de níveis de cinzento e a detecção de pontos de máxima

curvatura dos contornos correspondentes às arestas dos objectos. Um processo de determinação dos "pontos de interesse" foi proposto por Moravec [Moravec 80]. Estes pontos podem ser obtidos por utilização de um operador que convolucionado com a imagem permite determinar pontos que apresentam um máximo da variação dos níveis de cinzento segundo várias direcções e na área correspondente a uma janela. Os pontos que não apresentam variação numa das direcções não devem ser tomados em conta pois poderão pertencer a arestas ou a regiões homogéneas na imagem. O processo proposto por Moravec tenta seleccionar regiões que apresentem um máximo numa medida de variação dos níveis de cinzento segundo várias direcções. Ao definir como "pontos de interesse" os pontos que apresentam um máximo nessa medida, é possível distinguir primitivas na imagem. A figura 4.1 ilustra a forma como essa medida é obtida. O somatório do quadrado da diferença entre "pixels" adjacentes é determinada para uma janela 5 por 5 e para cada uma das direcções sendo a medida de interesse da imagem para essa janela o menor dos quatro somatórios. Um ponto é classificado como de "interesse" se, para uma janela 5 por 5 centrada nesse ponto, ele apresenta um máximo local da medida de variação. Deste modo só são tidos em conta pontos com grande variação direccional. Para que o sistema seja imune a pequenas variações da medida, utiliza-se um limiar que define o menor valor que o máximo local deverá ter para se poder classificar o ponto como "ponto de interesse".

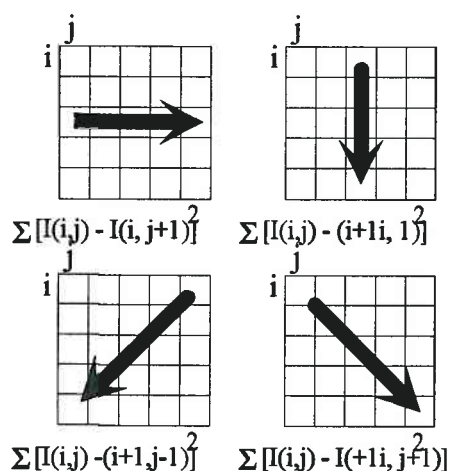


Figura 4.1- Janela típica do operador de interesse e direcções utilizadas para o cálculo da variação. A medida de interesse é a mínima das quatro somas.

A escolha do limiar para o valor do máximo local tem um efeito muito grande no resultado final do algoritmo. A figura 4.2 mostra os pontos de interesse da imagem (a) determinados com valores diferentes de limiar - (b) com limiar 200, (c) com limiar 500 e (d) com limiar 800. Apesar de ser um algoritmo muito conhecido a sua aplicação conduz por vezes à detecção de pontos que não são verdadeiros vértices dos objectos. Existem porém outros operadores, propostos por Deriche e Reisfeld, que exploram a noção de simetria [Deriche 90] [Reisfeld 90].

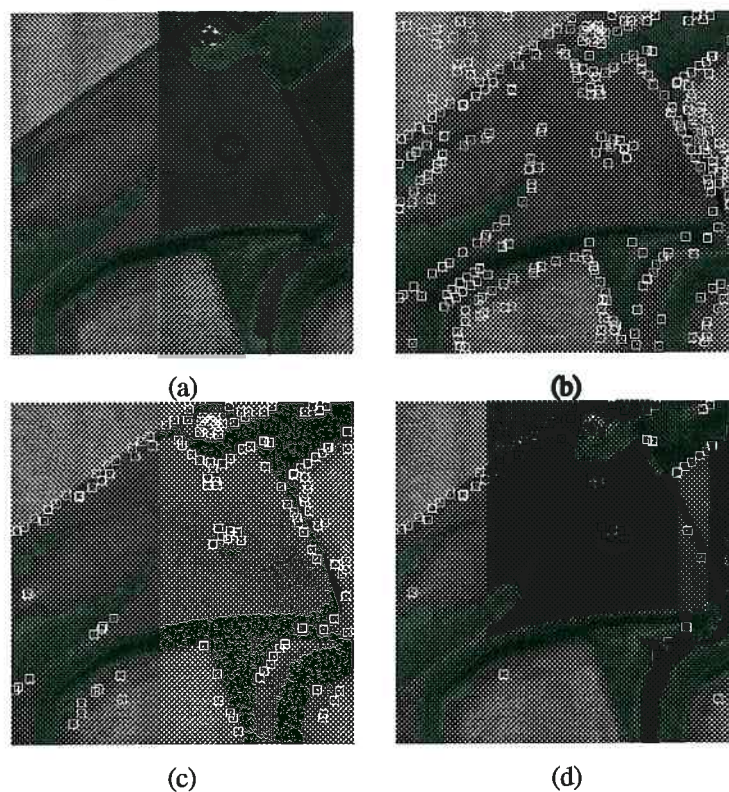


Figura 4.2- Imagens com aplicação do operador de pontos de interesse. A escolha do limiar mínimo tem muito efeito nos resultados - (a) imagem original; resultados da aplicação do operador de Moravec para (b) limiar 200, (c) limiar 500, (d) limiar 800.

Existem outros métodos que fazem a detecção dos "pontos de interesse" a partir dos contornos dos objectos. Estes métodos envolvem a detecção prévia das arestas e a aplicação sequencial de operações locais ao longo do seu contorno. Um exemplo consiste no algoritmo descrito em [Duda 73] que utiliza a distância de um ponto do contorno ao segmento de recta que une as duas extremidades do contorno [Duda 73] [Espelid 91]. Neste método é traçado um segmento de recta com início numa das extremidades do contorno e apoiado nos sucessivos pontos pertencentes ao longo do contorno. Para cada "pixel" do contorno é calculada a distância do pixel ao segmento de recta. A distância máxima atingida para cada "pixel" é retida num histograma de distâncias. Os pontos de interesse são associados aos "pixels" cujos valores de distância retidos correspondem a máximos locais excedendo um determinado valor de limiar. A aproximação realizada neste trabalho para a detecção de vértices é um pouco diferente desta técnica e baseia-se na intersecção de segmentos.

4.3 Detecção de Segmentos de Recta 2D

As arestas, contornos ou segmentos de recta são das primitivas mais importantes em visão por computador. As arestas são importantes porque permitem separar ou segmentar a

imagem em regiões uniformes ou em locais que servem de fronteira entre regiões e em que os valores de intensidade luminosa, cor ou gama de profundidade são uniformes. Um método de detectar as arestas ou segmentos de recta nas imagens consiste em detectar os pontos de mudança brusca da intensidade luminosa (ou outra grandeza intrínseca que a imagem represente) ligando-os posteriormente de forma a definirem contornos ou arestas. Este processo implica a utilização de um operador para a detecção dessa passagem abrupta e que deverá ser aplicado a cada um dos pontos da imagem. Outro método consiste na segmentação da imagem em regiões uniformes e posteriormente obter as linhas de fronteira dessas regiões definindo arestas. Neste ponto é utilizada a primeiro método embora uma técnica de segmentação por regiões utilizando o controlo da focagem seja também proposta neste trabalho.

Outra técnica de detectar segmentos de recta em imagens consiste na utilização da transformada de Hough [Ballard 82]. Neste método uma recta é representada pela expressão paramétrica

$$x \cos \theta + y \sin \theta = \rho$$

onde ρ representa a distância, segundo a normal à recta passando pela origem do sistema de coordenadas e θ representa o ângulo dessa normal com o eixo-x. A transformada de Hough de uma linha é um ponto de coordenadas (ρ, θ) no domínio polar. Na primeira fase do método o domínio (ρ, θ) é dividido em células e para cada célula que a curva atravessa, um acumulador associado à célula é incrementado. No fim de percorrer todos os "pixel" da aresta obtemos um histograma global em que as zonas que apresentam máximos locais representam conjuntos de pontos colineares e com valores das rectas dados pelo par (ρ, θ) . As células que apresentam pequenos valores acumulados correspondem a pontos isolados. Este método é muito popular para a determinação de rectas mas só com base no histograma não é possível detectar o início e o fim de segmentos de recta.

Neste ponto propomos um algoritmo para detecção de segmentos de recta baseado em três estágios: detecção de arestas, detecção de contorno e estimação dos parâmetros geométricos da recta. No primeiro estágio faz-se a detecção das arestas utilizando um operador de gradiente aplicado à imagem previamente filtrada. O resultado é uma imagem binária com contornos correspondentes às arestas. No segundo estágio é criada uma lista de contorno que é posteriormente processada. Neste processamento é feita a fusão de contornos partidos e feita a eliminação de contornos curtos. No último estágio, e a partir da lista de contornos, são determinados os segmentos de recta e os parâmetros geométricos que lhes estão associados.

4.3.1. Detecção de Arestas

A detecção das arestas dos objectos nas imagens implica a detecção dos locais de passagem abrupta dos níveis de intensidade luminosa e correspondentes à projecção dessas arestas na imagem. Uma forma de realizar essa detecção é através de operadores diferenciadores tais como o gradiente ou o laplaciano. Os locais correspondentes aos máximos do módulo do gradiente ou aos locais de passagem por zero do laplaciano correspondem a contornos que definem locais de mudança abrupta da intensidade luminosa na imagem. Sendo $I(x, y)$ uma imagem então o gradiente dessa imagem é dado por

$$\mathbf{G}[I(x, y)] = \left(\frac{\partial I(x, y)}{\partial x}, \frac{\partial I(x, y)}{\partial y} \right) \quad (4.1)$$

e o Laplaciano dado por

$$\mathbf{L}[I(x, y)] = \left(\frac{\partial^2 I(x, y)}{\partial x^2} + \frac{\partial^2 I(x, y)}{\partial y^2} \right) \quad (4.2).$$

Normalmente são utilizadas aproximações para $\mathbf{G}[I(x, y)]$ e para $\mathbf{L}[I(x, y)]$ que são aplicadas às imagens sob forma de máscaras. Estas máscaras são posteriormente convolucionadas com a imagem. Esses operadores são sensíveis ao ruído impulsional e por vezes é necessário pré-processar a imagem de modo a eliminar o ruído. Para obter melhores resultados é conveniente uma filtragem prévia da imagem por um filtro passa-baixo com o objectivo de eliminar altas frequências que esporadicamente aparecem na imagem e são consequência do processo de aquisição da imagem. Se $\mathbf{G}[\]$ e $\mathbf{L}[\]$ representarem, respectivamente, o operador gradiente e o operador laplaciano então a sua aplicação a uma imagem $I(x, y)$ previamente filtrada por um filtro $F(x, y)$ pode ser expressa por

$$\mathbf{G}[I(x, y) * F(x, y)] \quad \text{e} \quad \mathbf{L}[I(x, y) * F(x, y)] \quad (4.3).$$

Atendendo à propriedade da linearidade da convolução e da operação de derivação do gradiente e do laplaciano a expressão (4.3) pode ser expressa na forma

$$I(x, y) * \mathbf{G}[F(x, y)] \quad \text{e} \quad I(x, y) * \mathbf{L}[F(x, y)] \quad (4.4).$$

Para identificar as arestas, se for utilizado o gradiente é necessário pesquisar pelos máximos locais do módulo do gradiente enquanto se for utilizado o laplaciano é necessário pesquisar os locais da sua passagem por zero.

Marr e Hildreth [Marr 80] propuseram um método de determinação das arestas baseado num filtro $L(g(x, y))$ em que $g(x, y)$ representa uma distribuição gaussiana

$$g(x, y) = \frac{1}{2\pi\sigma} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

sendo σ o desvio padrão da distribuição. Segundo Marr existiram dois motivos para a escolha deste tipo de filtro. Este filtro é o que apresenta melhores características de localização no domínio espacial e espectral. No domínio espacial o filtro tende a reduzir a gama de intervalos de variação das mudanças de intensidade na imagem. O espectro em frequência tende a ser mais suave e de banda limitada. No domínio espacial tende a localizar correctamente o aumento das variações de intensidade na imagem [Marr 80]. A expressão para este filtro em duas dimensões é

$$F(x, y) = L(g(x, y)) = \frac{1}{\pi\sigma^4} \left[\frac{(x^2 + y^2)}{2\sigma^2} - 1 \right] e^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)}.$$

No entanto as máscaras bidimensionais correspondentes às aproximações do gradiente ou do laplaciano apresentam problemas graves de tempo de computação. Se uma imagem tem dimensões $(A_x \times A_y)$ e a máscara tem dimensões $(p \times p)$ então o tempo de computação é proporcional a $(p^2 \times A_x \times A_y)$. Para reduzir esse tempo de computação têm sido estudados filtros com resposta separada e direccionais ou seja filtros que permitam ter uma expressão do tipo

$$F(x, y) = F_x(x, y) * F_y(x, y).$$

A filtragem separada tem as vantagens de reduzir o tempo de computação da filtragem e de permitir ter filtros com características diferentes em direcções diferentes.

Canny desenvolveu um detector de arestas, baseando-se num modelo unidimensional [Canny 85]. Neste modelo a detecção é realizada através da convolução de uma função $f(x)$ e um sinal $I(x)$ contendo uma aresta. Ao sinal $I(x)$ foi adicionado ruído branco gaussiano de média nula e variância σ^2 . O ponto correspondente à aresta é o ponto x_0 que maximiza o resultado da convolução

$$\theta(x_0) = \int_{-\infty}^{+\infty} I(x) f(x_0 - x) dx.$$

A partir deste modelo, Canny definiu três critérios de otimização para determinar a função $f(x)$. Assim, um filtro deve ter uma baixa probabilidade de não detectar um verdadeiro ponto de uma aresta e de marcar falsos pontos do contorno. Para além disso deve permitir localizar correctamente os contornos, marcando-os o mais ao centro possível da verdadeira aresta. O filtro deverá assinalar um só contorno quando detecta uma aresta, ou seja, deverá ter uma única resposta na presença de uma aresta. Estes critérios podem ser expressos em termos matemáticos e combinados de modo a obter uma equação diferencial cuja solução corresponde à função $f(x)$ [Canny 85] [Monga 89]. Shen e Castan propõem a determinação do filtro utilizando um critério que conjuntamente inclui a detecção e a localização das arestas e um critério com base na resposta do filtro ao ruído [Shen 86]. Ambos os critérios são combinados para determinarem uma equação diferencial cuja solução corresponde à função $f(x)$. Este filtro corresponde à solução óptima para os dois primeiros critérios de Canny apresentando por isso o melhor compromisso detecção/localização, embora não esteja otimizado de modo a impedir múltiplas detecções das arestas.

No nosso caso foi utilizado um filtro cuja derivada é uma solução exacta da equação diferencial obtida pela combinação dos três critérios de Canny [Derich 88]. O filtro baseia-se na função unidimensional

$$s(x) = (c|x| + 1) e^{-\alpha|x|} \quad (4.5)$$

em que c é escolhido de maneira a obter um filtro normalizado tal que

$$\int_{-\infty}^{+\infty} s(x) dx = 1$$

e α define a largura do filtro.

Durante o trabalho foi utilizado o gradiente para a detecção das arestas. O filtro que apresenta o melhor compromisso dos critérios de Canny, tem a seguinte expressão [Derich 88]

$$s'(x) = f(x) = -c x e^{-\alpha|x|} \quad (4.6)$$

e pode ser realizado de forma recursiva [Monga 90]. Para um sinal, com pelo menos N amostras, representadas por x_m sendo $m = 1 \dots N$, a sequência de resposta y_m correspondente ao filtro (4.6) é dada pela expressão

$$\begin{cases} y_m = c_1 (y_m^+ + y_m^-) & \text{com } m = 1 \dots N \\ y_m^+ = a_2 x_{m-1} - b_1 y_{m-1}^+ - b_2 y_{m-2}^+ & \text{com } m = 1 \dots N \\ y_m^- = a_3 x_{m+1} - b_1 y_{m+1}^- - b_2 y_{m+2}^- & \text{com } m = N \dots 1 \end{cases} \quad (4.7)$$

com $c_1 = -(1 - e^{-\alpha})^2$; $a_2 = 1$; $a_3 = -1$; $b_1 = -2 e^{-\alpha}$; $b_2 = e^{-2\alpha}$. A sequência de resposta y_m correspondente ao filtro (4.5) e para o mesmo sinal x_m é dada pela expressão

$$\begin{cases} y_m = (y_m^+ + y_m^-) & \text{com } m = 1 \dots N \\ y_m^+ = a_0 x_m + a_1 x_{m-1} - b_1 y_{m-1}^+ - b_2 y_{m-2}^+ & \text{com } m = 1 \dots N \\ y_m^- = a_2 x_{m+1} + a_3 x_{m+2} - b_1 y_{m+1}^- - b_2 y_{m+2}^- & \text{com } m = N \dots 1 \end{cases} \quad (4.8)$$

com

$$\begin{cases} a_0 = k; a_1 = k(\alpha - 1)e^{-\alpha}; a_2 = k(\alpha + 1)e^{-\alpha}; a_3 = -ke^{-2\alpha}; \\ b_1 = -2 e^{-\alpha}; b_2 = e^{-2\alpha} \\ k = \frac{(1 - e^{-\alpha})^2}{(1 + 2 \alpha e^{-\alpha} - e^{-2\alpha})} \end{cases}$$

Os valores típicos para α situam-se entre [0.5 .. 1.1]. A expressão (4.7) corresponde a um filtro unidimensional que, ao ser aplicado na imagem segundo direcções diferentes e ortogonais (direcção u e direcção v), permite ter um filtro bidimensional.

Utilizando o gradiente, as arestas correspondem aos locais onde existem extremos da sua amplitude. Para a detecção das arestas será necessário detectar ao longo do contorno os pontos com a maior amplitude local do gradiente. Se \mathbf{p} é um ponto da imagem para o qual é conhecido o gradiente $G(\mathbf{p})$ então este ponto corresponde a um máximo local do gradiente se $|G(\mathbf{p})| > |G(\mathbf{p}_1)|$ e $|G(\mathbf{p})| > |G(\mathbf{p}_2)|$ sendo \mathbf{p}_1 e \mathbf{p}_2 pontos da imagem, simetricamente opostos, pertencendo a uma recta que passa por \mathbf{p} - figura 4.3.

Estes pontos são determinados por

$$\mathbf{p}_1 = \mathbf{p} + d \frac{G(\mathbf{p})}{|G(\mathbf{p})|} \quad \text{e} \quad \mathbf{p}_2 = \mathbf{p} - d \frac{G(\mathbf{p})}{|G(\mathbf{p})|} \quad (4.9)$$

sendo d uma distância pré-definida. Nas experiências realizadas, a distância d utilizada foi de 1 "pixel". Com este teste assegura-se que o máximo local é efectivamente determinado na aresta.

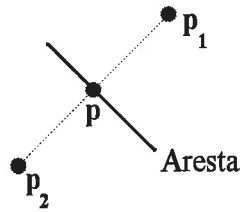


Figura 4.3 - Detecção dos extremos locais

Normalmente, em imagens reais e devido ao ruído das imagens, muitos dos extremos do gradiente não correspondem a verdadeiros pontos das arestas. Para evitar esse tipo de problemas é definido um limiar mínimo para a qual se pode classificar um ponto como sendo um ponto de uma aresta. No entanto seleccionar um limiar "a priori" pode levar à marcação de falsos pontos de aresta. Uma maneira de reduzir este problema é definir um limiar com histerese. Um limiar com histerese, permite detectar os pontos da aresta entre um grupo V de pontos vizinhos com gradiente superior a um limiar mínimo l_{min} , mas que têm uma amplitude do gradiente superior a um limiar máximo l_{max} . Os elementos do grupo de pontos devem ser adjacentes e, para efeitos de cálculo, podem ser determinados com base na conectividade entre "pixels" (8 ou 24). Os pontos p_i que pertencem ao conjunto V e fazem parte da aresta, devem conservar a orientação local da aresta no ponto p . Como temos uma estimação do gradiente é possível determinar a distância dos pontos à tangente que passa pelo ponto p - figura 4.4.

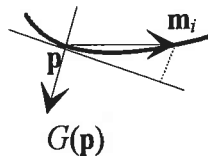


Figura 4.4 - Detecção local da aresta

Assim, se p for um ponto com amplitude do gradiente superior a l_{max} então, para a uma determinada vizinhança V de "pixels", os pontos $m_i \in V$ que obedecem á relação

$$\frac{|\overline{pm_i} \cdot G(p)|}{\|G(p)\|} < D \quad (4.10)$$

são classificados como pontos da aresta. A escolha de D está relacionada com a curvatura do contorno admissível e o seu valor típico estará entre 0.07 e 1, para uma vizinhança de 8 "pixels". Se a curvatura for muito acentuada devem ser escolhidos valores altos para D e dever-se-á fazer o contrário para a detecção de rectas.

Na experiência realizadas os valores para l_{min} e para l_{max} foram determinados automaticamente e com base no histograma de amplitudes do gradiente para todos pontos da

imagem. Nas experiências realizadas, o valor de l_{max} utilizado foi 75% do valor do módulo do gradiente com maior ocorrência no histograma e o valor de l_{min} foi 50% desse valor. O algoritmo utilizado para a filtragem das imagens está descrito no quadro abaixo.

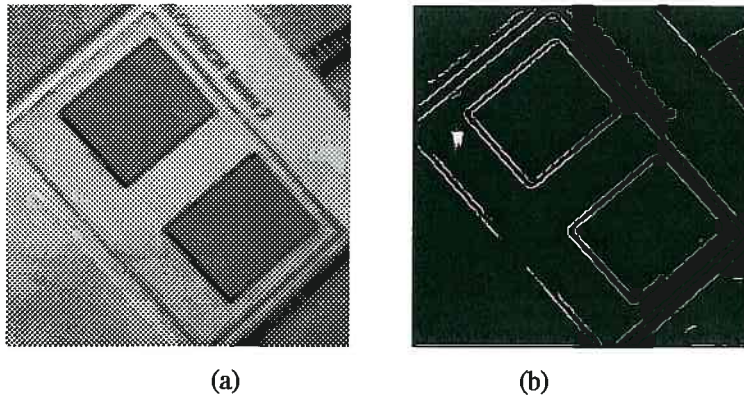


Figura 4.5a- Resultados da aplicação do detector de arestas. (a) imagem original; (b) $\alpha=1.5$; $d=1$; $D=0.7$.

Algoritmo 1 Detecção de arestas

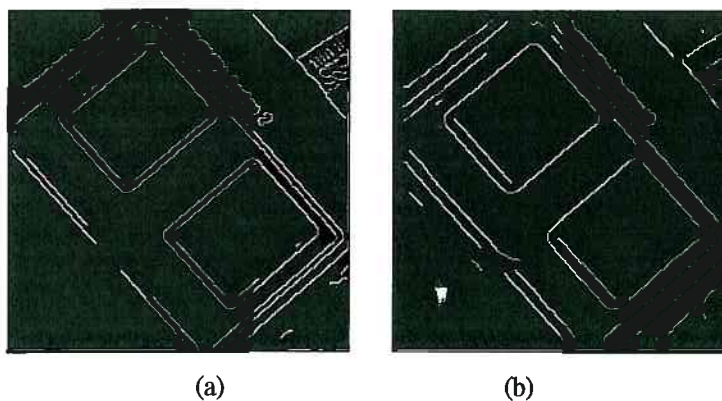
Definições

- Ii** - imagem de entrada em níveis de cinzento.
- Io** - imagem binária de saída.
- Iu** - imagem gradiente segundo a direcção u .
- Iv** - imagem gradiente segundo a direcção v .
- Su** - Filtro (4.8) aplicado segundo a direcção u .
- Sv** - Filtro (4.8) aplicado segundo a direcção v .
- Gu** - Filtro (4.7) aplicado segundo a direcção u .
- Gv** - Filtro (4.7) aplicado segundo a direcção v .

Início

- Convolve o filtro Su com imagem **Ii**;
- Convolve o filtro Gv com a imagem do resultado anterior e obtém **Iv**;
- Convolve o filtro Sv com imagem **Ii**;
- Convolve Gu com a imagem do resultado anterior e obtém **Iu**;
- Detecta os extremos do gradiente a partir de **Iv** e **Iu**;
- Detecta as arestas utilizando histerese;

fim Algoritmo



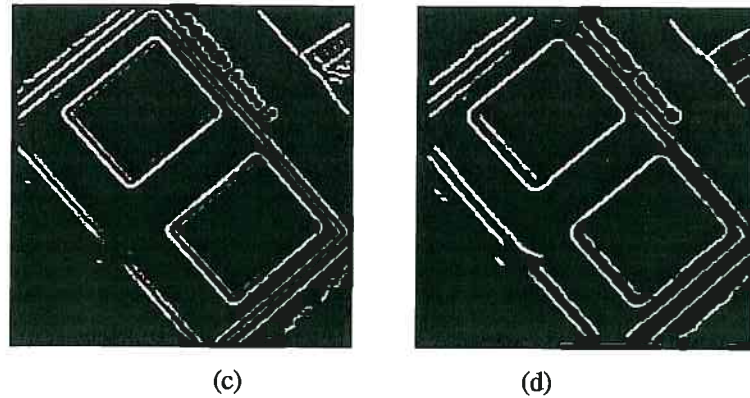


Figura 4.5b- Resultados da aplicação do detector de arestas com diferentes valores para α , d e D . (a) $\alpha=1.5$; $d=1$; $D=1$; (b) $\alpha=0.5$; $d=1$; $D=1$; (c) $\alpha=1.5$; $d=3$; $D=0.7$; (d) $\alpha=0.5$; $d=3$; $D=0.7$.

Aplicabilidade do filtro a imagens de cor

Algumas das propriedades físicas de objectos podem ser determinadas utilizando imagens de cor. Para a análise de imagens de cor utiliza-se, normalmente, a informação contida nas três bandas espectrais (R, G e B). Por exemplo, Brockelbank propõe a utilização de imagens de cor no seu algoritmo de recuperação de estrutura tridimensional utilizando estereovisão [Brockelbank 89]. No entanto, os algoritmos que se baseiam na detecção de arestas nas imagens de cor apresentam instabilidades nos seus resultados. O problema surge pelo facto de as arestas, nestas imagens, sofrerem de deslocamento espacial. Tal facto está representado nas figuras abaixo que representam as imagens da *diferença entre as imagens* que foram obtidas após a aplicação do detector de arestas.

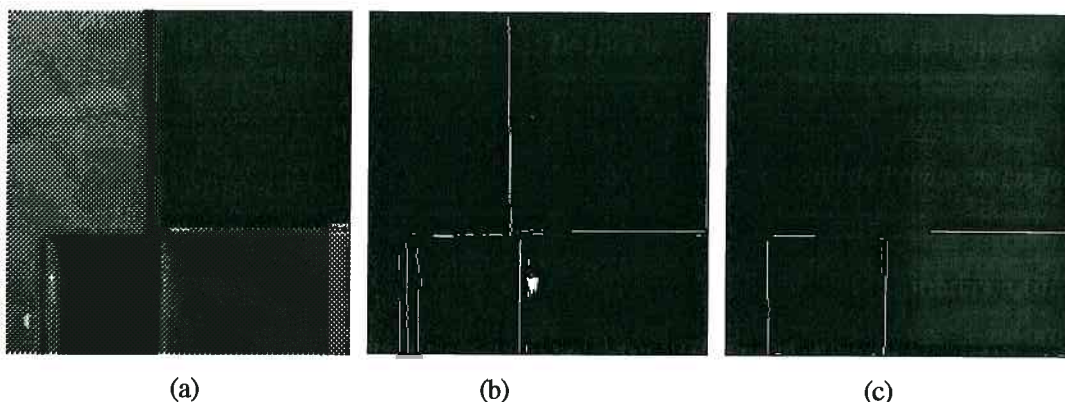


Figura 4.6- Resultados da aplicação do detector de arestas a imagens de cor. (a) Imagem de luminância resultante das três bandas. Diferença de imagens de arestas R-B (b) B-G (c). Se não existisse deslocamento espacial a diferença apresentada na figura (b) e (c) seria nula e por consequência a imagem seria totalmente escura.

Implicitamente assumimos que a informação existente num "pixel" específico, corresponde ao mesmo ponto no espaço objecto. Na realidade não é isto que se passa devido à influência do sistema de captação das imagens e devido à aberração cromática das lentes. Em muitos sistemas a aquisição baseia-se em câmaras com uma lente e um só sensor para captar as imagens de cor. Em tecnologia CCD ("Charge-Coupled Device"), a técnica mais utilizada para captar as imagens de cor consiste em utilizar filtros para cada banda (R, G, B) que, segundo uma linha, se repetem de três em três detectores do sensor. A utilização desta técnica, provoca uma deslocação espacial entre o ponto no sinal vídeo e o ponto no mundo objecto. Este deslocamento é segundo uma direcção paralela ao sensor sendo conhecido por efeito de deslocamento lateral [Willson 91]. A aberração cromática é outro efeito importante e intrínseco das lentes. A aberração cromática existe devido aos diferentes índices de refacção da lente, em função do comprimento de onda. Essa diferença dos índices de refacção faz com que as componentes da radiação luminosa em diferentes comprimentos de onda sejam refractados de maneira diferente. Na prática, e para lentes não corrigidas, a componente espectral correspondente à banda B gera uma imagem bem focada numa posição diferente da banda R. Esta diferença corresponde a uma diferença entre posições de focagem medida segundo o eixo óptico da lente. Esta diferença ou deslocamento denomina-se deslocamento longitudinal [Willson 91]. Para minimizar estes efeitos podemos utilizar mecanismos correctores e determinar os verdadeiros valores através de polinómios de interpolação. Para isso é necessário fazer uma calibração do sistema para determinar, para cada uma das bandas de cor, o polinómio interpolador que nos dá a aproximação mais correcta do valor do nível do sinal para cada "pixel". Esse valor será função da posição da focagem da lente e da posição (u, v) do "pixel" na imagem. Foram realizadas algumas experiências de calibração para a determinação destes polinómios de interpolação para as lentes disponíveis. No entanto, como os algoritmos de recuperação de estrutura tridimensional utilizam imagens de níveis de cinzento, os resultados não foram utilizados durante o trabalho.

4.3.2 Contornos

Após a detecção das arestas foi utilizado um seguidor de contornos que gera listas correspondentes aos contornos existentes na imagem. A imagem é previamente processada por um operador que, numa primeira passagem, assegura que os contornos correspondentes às arestas, tenham a espessura de um só "pixel". Após esse processo são geradas as listas de contornos. Cada lista de um contorno contém as coordenadas dos "pixels" pertencentes ao contorno. O primeiro elemento da lista contém o "pixel" inicial, a direcção que é necessária tomar para seguir o contorno e a quantidade de "pixels" segundo essa direcção. Se ao seguir o contorno surge uma nova direcção, então é criado um novo nó na lista que armazena a nova direcção e a quantidade de "pixels" nessa direcção. A partir de um "pixel" é possível

seguir em 8 direcções diferentes - ver figura 4.7. Fazendo corresponder códigos diferentes para as direcções diferentes é possível definir uma tabela que, para cada código, indica o "pixel" a percorrer. A tabela indica também se o "pixel" actual corresponde ao início ou ao fim do contorno.

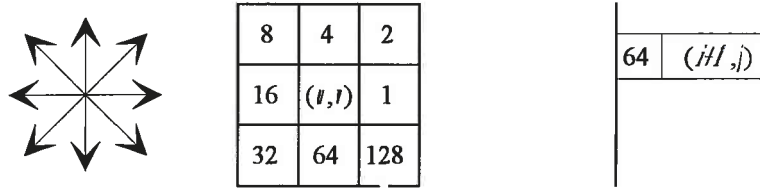


Figura 4.7 - As 8 direcções definem entradas para uma "look-up table" que indica qual o próximo "pixel" a seguir. A mesma tabela permite definir se estamos no início ou fim de um contorno.

Após a obtenção destas listas de contornos são eliminadas listas demasiado curtas e fundidas listas de contornos cujas extremidades sejam muito próximas. Na prática isso corresponde a uma filtragem de ruído e a uma melhor definição das primitivas da imagem.

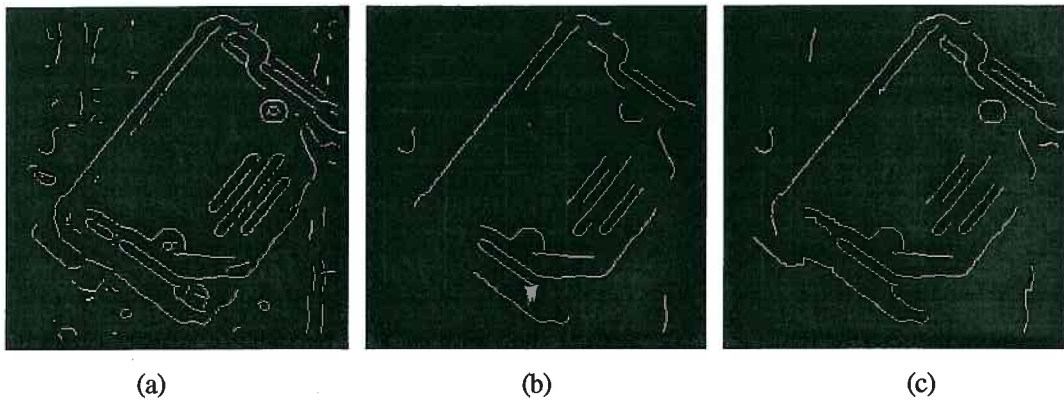


Figura 4.8- A fusão de contornos com extremidades muito próximas permite uma melhor definição das primitivas da imagem. Em (a) a imagem de contornos (b) a detecção de contornos é feita sem a fusão dos contornos, na imagem (c) a detecção é feita com a fusão dos contornos.

Após a determinação dos contornos poderíamos determinar pontos característicos dos contornos como por exemplo as suas extremidades. No entanto e como se pode ver pela figura abaixo estas extremidades dependem muito da qualidade da imagem e dos parâmetros do operador. A posição destas primitivas na imagem é bastante instável, não sendo viável a sua utilização em algoritmos para determinação de estrutura tridimensional.

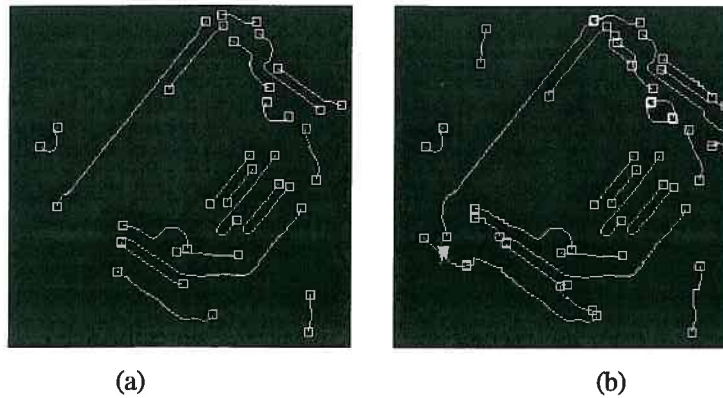


Figura 4.9 - Imagens das extremidades dos contornos (a) sem de fusão de contornos curtos e (b) com fusão de contornos curtos. As posições das extremidades dos contornos são fortemente dependentes da qualidade das imagens, da forma de detecção dos contornos.

4.3.3 Segmentação dos Contornos

Após a determinação da lista de contornos, segue-se a determinação dos segmentos de recta existentes nos contornos. Para cada contorno é necessário determinar os troços correspondentes a segmentos de recta. A partir destes troços será feita a estimação das características geométricas dos segmentos. A determinação dos troços corresponde a uma segmentação do contorno e envolve um teste de rectilinearidade de um contorno.

Teste do Desvio Angular

O teste de rectilinearidade baseia-se no teste de desvio angular [Lustman 87] [Ayache 88a]. Esta forma de segmentação do contorno apresentou melhores resultados na segmentação dos contornos do que os métodos baseados no máximo da curvatura ou baseados na distância até à recta que une as extremidades do contorno [Duda 73] [Espelid 91]. Isso permite um controlo mais fino dos parâmetros de teste da rectilinearidade. Esse teste foi utilizado na segmentação de um contorno sob uma forma de algoritmo recursivo.

O princípio do teste do desvio angular consiste em verificar se existe um desvio angular significativo entre a recta, que une o ponto inicial do contorno e o ponto p_i , e a recta que une o ponto inicial do contorno e o ponto p_{i+1} . Inicialmente o contorno é percorrido a partir de um ponto inicial P_0 e segundo os pontos P_k seguintes. Para cada ponto P_k é medida a distância desse ponto ao ponto P_0 e o teste de rectilinearidade inicia-se no momento em que a distância de P_0 ao ponto P_k é maior do que l - ver figura 4.10. Quando a distância de P_0 a P_k é superior ou igual a um valor l , é determinado um cone inicial formado pelas duas rectas que passam pelo ponto P_0 e são tangentes a uma circunferência de raio l , centrada no ponto P_k - tal como mostra a figura 4.10.

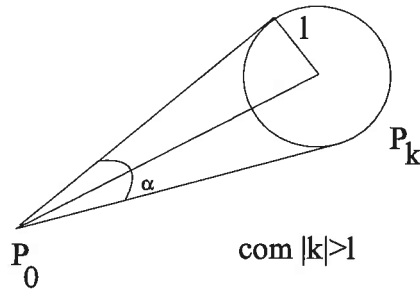


Figura 4.10- Cone admissível por cada um dos pontos do contorno

Se existe intersecção entre o cone formado pelo ponto P_0 e P_k e o cone formado pelos pontos P_0 e P_{k+1} então estamos a seguir um segmento de recta e o contorno vai sendo percorrido. À medida que os pontos do contorno vão sendo percorridos vai sendo calculado um cone de intersecção C_{int} com base na intersecção dos sucessivos cones $C_1 \dots C_j$. Quando esta intersecção é nula ou tem um ângulo mínimo (α_{min}), o contorno deixa de ser seguido e inicia-se o processo de teste do desvio angular. O teste consiste em recuar no contorno de modo que, partindo do ponto de ruptura, se determina o ponto que conduz à menor distância até ao eixo do cone C_{int} . Este teste é bastante importante para evitar o corte de cantos de arestas perfeitas e determinar o verdadeiro ponto de ruptura do contorno- veja-se figura 4.11. Após ter sido determinado este ponto de ruptura e se ainda existir contorno, então o processo é aplicado recursivamente ao restante dos pontos do contorno.

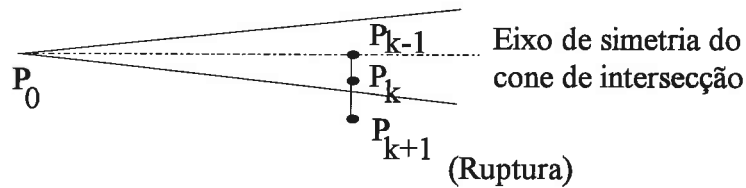


Figura 4.11 - Teste do desvio angular. O ponto P_{k-1} é escolhido apesar da ruptura se dar em

P_{k+1}

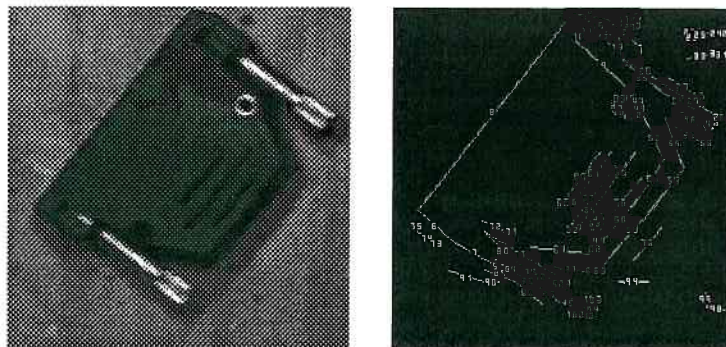


Figura 4.12 - Rectas determinadas através de um teste de rectilinearidade e utilizando o método do desvio angular.

Teste da orientação média do contorno

Uma alternativa ao teste do desvio angular é a detecção dos segmentos de recta seguindo o contorno por troços sendo o número de elementos do troço menor que o total de elementos do contorno. Por cada novo troço do contorno é determinado o ângulo médio do troço $\theta_{troço}$ sendo o ângulo $\theta_{troço}$ a média dos ângulos dos segmentos de recta que unem o ponto inicial do troço aos sucessivos pontos do troço. Se o ângulo médio do troço for muito diferente do ângulo médio do contorno até ao ponto anterior ao novo troço θ_i então estamos perante um ponto de ruptura do contorno. Sendo θ o ângulo médio até aquele ponto e $\theta_{troço}$ o ângulo médio do novo troço, então se a diferença entre os ângulos for maior do que $\Delta\theta$ então começou um novo segmento. Caso a diferença não seja significativa determina-se um novo valor médio θ e prossegue para o troço seguinte. Este método apresenta melhores resultados quando, após a detecção de um segmento de recta, é realizado o teste do desvio angular proposto anteriormente.

Transformada de Hough

Outra alternativa para a determinação do segmento de recta num contorno é utilização de um histograma de orientações tal como na transformada de Hough. Apesar de ser possível determinar se um contorno contém uma ou mais rectas não determina exactamente onde estas começam e onde estas acabam. Isto é, depois de uma primeira passagem por todo o contorno para gerar o histograma, seria necessário detectar os máximos locais do histograma e posteriormente os pontos de ruptura do contorno, percorrendo o contorno numa segunda passagem. Este procedimento não é desejável visto que seria necessário percorrer o contorno duas vezes.

4.3.4 Parâmetros Geométricos dos Segmentos de Recta

O resultado do processamento das listas de contornos é uma lista com o conjunto de contornos correspondentes às arestas rectilíneas dos objectos. Cada lista desse conjunto contém um conjunto de "pixels" (u_i, v_i) pertencentes a um contorno e que pretendemos aproximar por uma recta. O problema consiste em estimar os parâmetros geométricos da recta a partir dos pontos (u_i, v_i) .

A representação paramétrica de uma recta tem a seguinte expressão matemática

$$u n_u + v n_v = \rho \quad (4.11)$$

em que (u, v) é um ponto da recta, $\mathbf{n} = (n_u, n_v)$ é um vector e normal à recta com módulo unitário e ρ é a distância da origem do sistema de coordenadas à recta.

Esta representação pode ser expressa em função de cada ponto (u_i, v_i) do contorno, e pode ser transformada na seguinte expressão vectorial

$$\mathbf{p}_i^T \mathbf{n} = [u_i \quad v_i] \begin{bmatrix} n_u \\ n_v \end{bmatrix} = d_i \quad (4.12)$$

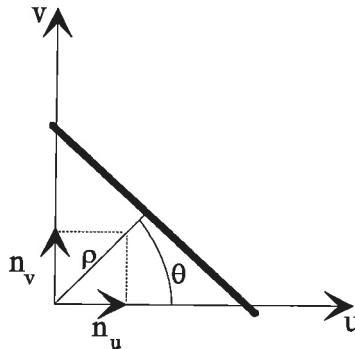


Figura 4.13- Representação paramétrica de um segmento de recta.

em que d_i representa a distância do ponto (u_i, v_i) à origem do sistema de coordenadas, mas projectada segundo a direcção da normal \mathbf{n} . O valor $|d_i - \rho|$ representa uma diferença entre distâncias medidas sobre a perpendicular à recta. Como pretendemos representar o conjunto de pontos (u_i, v_i) por uma recta, então a minimização desta diferença é um critério para determinar os parâmetros da recta.

Utilizando o critério do mínimo erro quadrático, corresponderá à minimização do valor

$$\xi = \sum_{i=1}^m (\mathbf{p}_i^T \mathbf{n} - \rho)^2 = \sum_{i=1}^m (u_i n_u + v_i n_v - \rho)^2 \quad (4.13)$$

sujeito à restrição $n_u^2 + n_v^2 = 1$ e correspondente ao valor do módulo de \mathbf{n} . Os valores para $\mathbf{h} = (n_u, n_v, \rho)^T$ que conduzem à minimização de (4.13) equivalem aos valores dos zeros da derivada de 1ª ordem da função ξ .

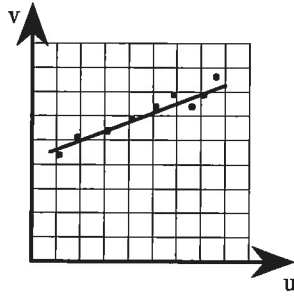


Figura 4.14- Representação dos pontos por uma da recta

Do ponto de vista matemático equivale à determinação do vector de parâmetros \mathbf{h} que satisfaçam a restrição

$$\frac{\partial \xi}{\partial \mathbf{h}} = 0 \quad (4.14)$$

que de forma equivalente conduz a que as seguintes restrições sejam satisfeitas

$$\begin{cases} pn_u + qn_v - s\rho = 0 \\ qn_u + wn_v - t\rho = 0 \\ sn_u + tn_v - m\rho = 0 \\ n_u^2 + n_v^2 = 1 \end{cases} \quad (4.15)$$

em que m representa o número total de pontos e

$$p = \sum_{i=1}^m u_i^2; \quad s = \sum_{i=1}^m u_i; \quad q = \sum_{i=1}^m u_i v_i;$$

$$t = \sum_{i=1}^m v_i; \quad w = \sum_{i=1}^m v_i^2;$$

Uma solução para as restrições (4.15) corresponde ao vector

$$\mathbf{h} = \frac{1}{\sqrt{(st - qm)^2 + (pm - s^2)^2}} \begin{bmatrix} st - mq \\ pm - s^2 \\ qs - pt \end{bmatrix} \quad (4.16)$$

Esta solução pode ser também obtida por mínimos quadrados exprimindo a equação (4.11) através da equação

$$u_i \alpha + b = v_i \quad (4.17)$$

com
$$\alpha = -\frac{n_u}{n_v} \quad \text{e} \quad b = \frac{\rho}{n_v}.$$

então se existirem α pontos (u_i, v_i) com $\alpha > 2$, então é possível exprimir a equação (4.17) através da forma matricial

$$\begin{bmatrix} \cdots \\ u_i & 1 \\ \cdots \end{bmatrix} \begin{bmatrix} \alpha \\ b \end{bmatrix} = \begin{bmatrix} \cdots \\ v_i \\ \cdots \end{bmatrix} \Rightarrow \mathbf{A} \mathbf{k} = \mathbf{b} \quad (4.18).$$

Pelo critério do mínimo dos erros quadráticos o valor de \mathbf{b} é dado pela pseudo-inversa (ver Apêndice A)

$$\mathbf{k} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b} \quad (4.19)$$

se $(\mathbf{A}^T \mathbf{A})$ for invertível. Os elementos de $(\mathbf{A}^T \mathbf{A})^{-1}$ são dados por

$$\begin{bmatrix} \sum_{i=1}^m u_i^2 & \sum_{i=1}^m u_i \\ \sum_{i=1}^m u_i & m \end{bmatrix}^{-1} = \begin{bmatrix} p & s \\ s & m \end{bmatrix}^{-1} = \frac{1}{pm - s^2} \begin{bmatrix} m & -s \\ -s & p \end{bmatrix} \quad (4.20).$$

Para que a matriz seja invertível, é necessário que $pm - s^2 \neq 0$, ou seja, que $u_i \neq u_j$ com $i, j = 1 \cdots m$. Assim para rectas verticais será necessário utilizar uma expressão alternativa do tipo $u_i = v_i \alpha' + b'$.

Utilizando as equações (4.17) e (4.11), podemos obter expressões mais desenvolvidas para os elementos do vector \mathbf{k}

$$\begin{aligned} \mathbf{k} = \begin{bmatrix} \alpha \\ b \end{bmatrix} &= \frac{1}{pm - s^2} \begin{bmatrix} m & -s \\ -s & p \end{bmatrix} \begin{bmatrix} \sum_{i=1}^m u_i v_i \\ \sum_{i=1}^m v_i \end{bmatrix} \\ &= \frac{1}{pm - s^2} \begin{bmatrix} m & -s \\ -s & p \end{bmatrix} \begin{bmatrix} q \\ t \end{bmatrix} = \frac{1}{pm - s^2} \begin{bmatrix} mq - st \\ pt - qs \end{bmatrix} \end{aligned} \quad (4.21).$$

As expressões para o cálculo dos parâmetros α e b são

$$\alpha = \frac{mq - st}{pm - s^2} \text{ e } b = \frac{pt - qs}{pm - s^2} \quad (4.22).$$

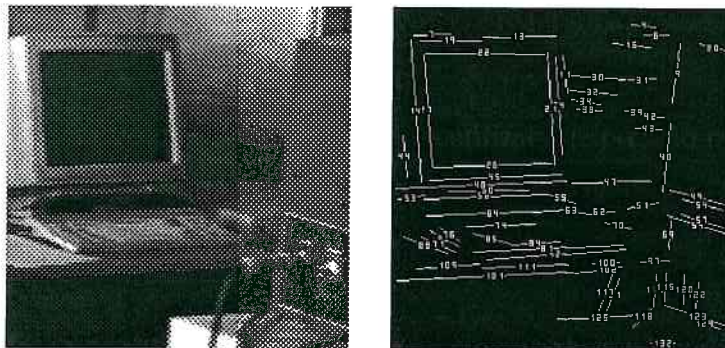
Para obter os valores para ρ e o vector \mathbf{n} podemos utilizar a expressão de uma distância de um ponto a uma recta. Se a expressão matemática da recta (4.17) está na forma $Au + Bv + C = 0$ então a distância de um ponto (u_0, v_0) à recta é dada por

$$d = \frac{Au_0 + Bv_0 + C}{\sqrt{A^2 + B^2}}.$$

Assim o valor para ρ é dado pela distância da recta ao ponto (0,0) ou seja

$$\rho = \frac{b}{\sqrt{\alpha^2 + 1}} = \frac{pt - qs}{\sqrt{(mq - st)^2 + (mp - s^2)^2}} = \frac{pt - qs}{L}$$

$$n_v = \frac{\rho}{b} = \frac{mp - s^2}{L} \text{ e } n_u = \frac{st - mq}{L} \quad (4.23).$$



(a)

(b)

Figura 4.15 - Rectas detectadas num conjunto de imagens estereo- (a) Imagem original.; (b) Imagem das rectas

4.4 Grupos de Segmentos

Ao analisarmos uma imagem ou uma cena existe uma forte tendência para favorecer grupos de primitivas que apresentam relações entre si de colinearidade, de paralelismo ou de repetição de formas. Este fenómeno é chamado organização perceptual e muitas teorias têm vindo a ser propostas tomando em conta a formação destes grupos [Lowe 90][Mohan 92]. Este facto favorece a ideia da utilização destas relações para melhorar o desempenho dos algoritmos que utilizam correspondência de primitivas entre imagens, como no caso da visão estereo ou movimento.

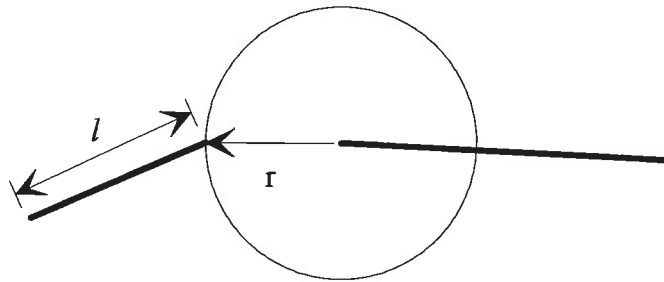
Neste ponto e com base neste princípio, propõe-se um método simples de detecção de grupos característicos no conjunto de segmentos de recta. Os mesmos argumentos

justificam a utilidade da detecção de outras formas de primitivas que incluam textura ou semelhança de cor.

Proximidade

A proximidade é uma propriedade que se reflecte da seguinte maneira nas imagens: se dois pontos tridimensionais estão próximos na cena então também as suas projecções nas imagens estão próximas. Todavia é também possível que dois pontos fortemente separados na cena possam estar muito juntos na imagem devido à orientação da câmara relativamente ao objecto. No entanto a medida de proximidade, conjuntamente com outras, permite estabelecer formas de detecção de grupos de segmentos.

Consideremos então dois segmentos numa imagem, tal como mostra a figura.



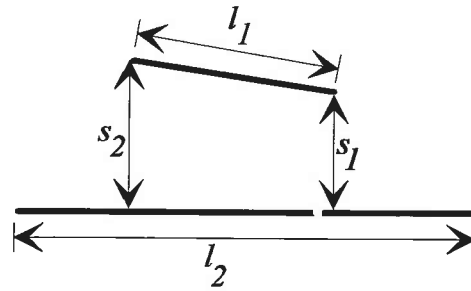
Se l é o comprimento do segmento de menor comprimento e r a distância da sua extremidade ao outro segmento, então uma medida da proximidade entre dois segmentos é

$$P = \left(1 - \frac{r}{\sqrt{N_u^2 + N_v^2}} \right) \times 100\% \quad (4.23)$$

em que N_u e N_v são as dimensões da imagem em "pixels" e r é a menor das distâncias entre extremos dos segmentos de recta. P toma o valor máximo quando os dois extremos dos segmentos se tocam e toma o valor mínimo quando a distância entre as extremidades é igual à diagonal da imagem.

Paralelismo

Uma medida semelhante pode ser estabelecida para o paralelismo entre segmentos.



Consideremos dois segmentos de recta com comprimentos l_1 e l_2 e com $l_1 < l_2$. Os dois segmentos estão separados pelas distâncias s_1 e s_2 medidas desde as extremidades do segmento l_1 até ao segmento l_2 e na perpendicular a este último. Se as medidas s_1 e s_2 existem e as rectas são representadas pelas expressões

$$\begin{cases} Ax + By + C = 0 \\ A'x + B'y + C' = 0 \end{cases}$$

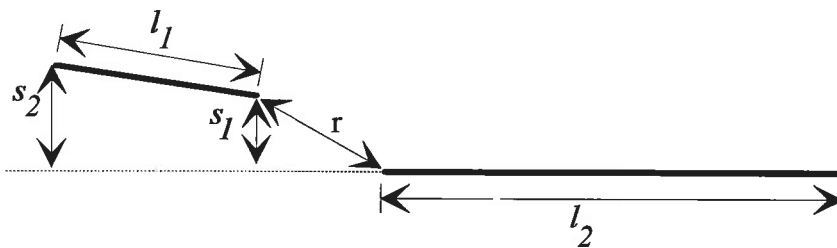
então uma medida do paralelismo entre segmentos é

$$PA = \left(1 - \frac{\left| \arctan\left(\frac{AB' - A'B}{AA' + BB'}\right) \right|}{\left(\frac{\pi}{2}\right)} \right) \times 100\% \quad (4.24).$$

em que $\left(\frac{AB' - A'B}{AA' + BB'}\right)$ representa a tangente do ângulo entre as duas rectas [Santos 70].

Colinearidade

A colinearidade tem muita semelhança com o paralelismo pois em ambos os casos a relação envolve dois segmentos de recta com a mesma orientação. No entanto, na colinearidade os segmentos são separados de uma distância r tal como se ilustra na figura seguinte.



Uma medida da colinearidade pode ser obtida à custa das medidas expressas por (4.23) e (4.24). Neste caso as distâncias s_1 e s_2 não existem mas poderão ser obtidas se o segmento l_2 for prolongado. Neste caso uma medida da colinearidade entre segmentos é

$$C = \left(\frac{P}{100} \times \frac{PA}{100} \right) \times 100\% \quad (4.25).$$

Grupos

Estas medidas poderão ser utilizadas de modo a detectarem grupos de segmentos que detêm propriedades tais como partilha de vértices ou paralelismo. Estas propriedades são testadas entre pares de segmentos formando grupos de segmentos

Para testar estas propriedades em grupos de primitivas é realizado o cálculo dos valores para P , PA e C . Na figura abaixo estão representadas imagens de exemplos de grupos de segmentos.

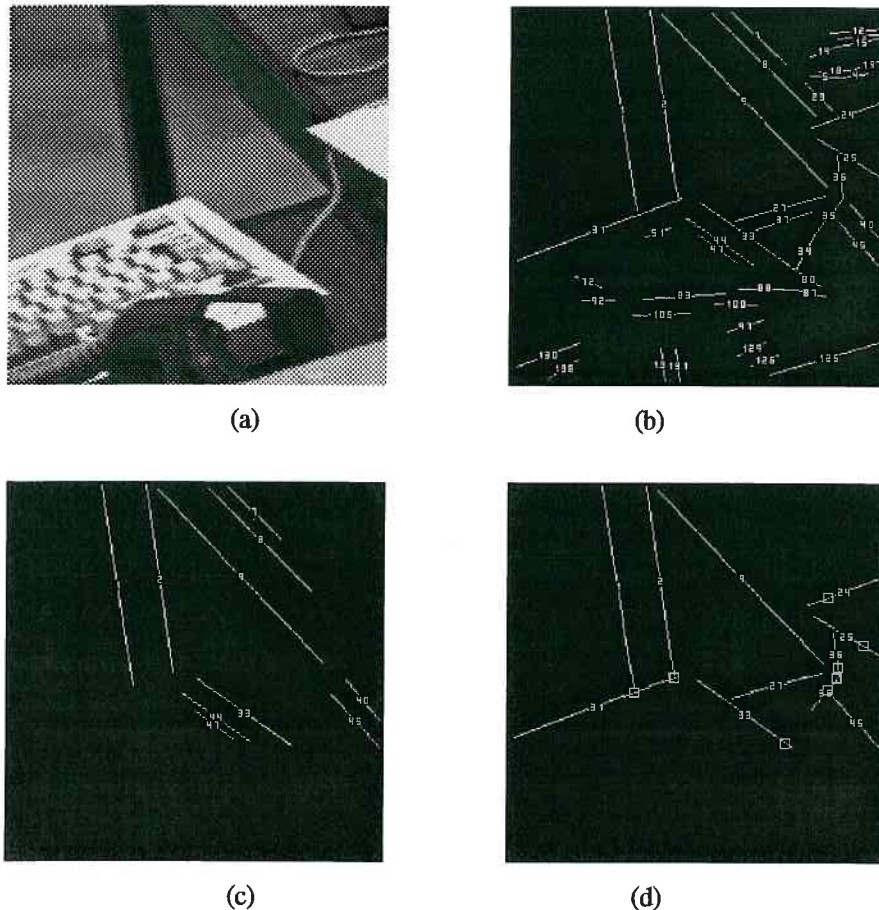


Figura 4.16- Grupos de segmentos obtidos através dos valores P , PA e C . (a) Imagens originais; (b) Segmentos de recta originais; (c) Segmentos de recta paralelos com ângulos próximos dos 45 e 90 graus para os intervalos de variação para $P=(50\%,90\%)$, $PA=(95\%,100\%)$, $C=(0\%,10\%)$. (d) Vértices resultantes do cruzamento de segmentos de recta para os intervalos de variação $P=(0\%,10\%)$, $PA=(0\%,15\%)$, $C=(0\%,10\%)$.

CAPÍTULO 5

Segmentação de imagens utilizando focagem

5.1. Introdução

Neste capítulo apresenta-se um método de segmentação de imagens e recuperação de profundidade utilizando focagem. Com base nos princípios físicos da formação da imagem e nos modelos geométricos da sua formação expostos no capítulo 2, são abordados, no ponto 5.2, alguns critérios para determinação de uma boa focagem. No ponto 5.3 é proposto um algoritmo de controlo de focagem com base em medidas realizadas na imagem. No ponto 5.4 é abordado o problema da calibração da focagem cujo resultado é posteriormente utilizado num algoritmo para segmentação de imagens, apresentado no ponto 5.5. O algoritmo permite a detecção de regiões na imagem, correspondentes a *superfícies planas* de objectos a diferentes distâncias da câmara e, através dos parâmetros determinados na calibração, permite a geração de um mapa tridimensional da cena. No ponto 5.6 é apresentado um algoritmo de segmentação de imagens de cor, utilizando um mecanismo de cooperação entre medidas sobre a cor e medidas de qualidade de focagem. Na realidade o algoritmo de focagem do ponto 5.5, é readaptado ao problema da segmentação de imagens de cor.

O objectivo da segmentação de uma imagem é o de conseguir a divisão da imagem num conjunto de regiões não sobrepostas e correspondentes à sua decomposição diferentes partes com idênticas propriedades locais. No entanto é difícil construir programas de computador para segmentação que determinem, de forma estável, zonas que consideramos "importantes" nas imagens. Os resultados obtidos pela aplicação destes programas devem obedecer a algumas regras:

- As regiões de uma imagem segmentada devem ter as suas características uniformes e homogêneas (nível de cinzento, cor, textura ou profundidade).

- As regiões adjacentes devem ter valores significativamente diferentes na característica para a qual ambas as regiões são diferentes.
- As fronteiras das regiões devem ser perfeitamente definidas.

5.2. Medida da Qualidade da Focagem

A medida da focagem permite a recuperação da profundidade porque o plano focado está relacionado directamente com a posição das lentes. Neste trabalho propõem-se algoritmos para determinar um mapa de profundidade relativo, explorando o controlo da distância da lente ao sensor para fazer a focagem das imagens.

O controlo da focagem através da transformada de Fourier e análise do espectro de frequência da imagem foi proposto por Horn [Horn 68]. Pentland propõe dois métodos para determinar o mapa de profundidade de uma cena [Pentland 87]. O primeiro método é baseado na medida da indefinição das arestas de uma imagem desfocada. Grossmann apresenta os resultados de algumas experiências baseadas neste princípio e cujo método requer a localização e conhecimento da amplitude da variação de nível de cinzento causada por uma aresta numa imagem [Grossmann 87]. O segundo método proposto por Pentland é baseado na comparação de duas imagens com diferentes diâmetros de abertura. Subbarao e Gurumoorthy fazem também a recuperação da profundidade utilizando as indefinições das arestas [Subbarao 87] [Subbarao 88]. Recentemente Hwang propôs um algoritmo de duas fases, onde o processo de desfocagem é modelado como uma função Gaussiana de dispersão bidimensional [Hwang 89]. Todos estes métodos recuperam a profundidade da cena directamente das imagens desfocadas.

É possível também estabelecer outros algoritmos para determinação da qualidade da profundidade relativa através da medida da focagem. Esses algoritmos utilizam funções que implementam um critério para a medida da desfocagem. Em Krotkov é analisado e comparada a "performance" de diferentes funções de critério de análise da focagem [Krotkov 89]. No mesmo trabalho, Krotkov compara o algoritmo proposto por Tenenbaum baseado na maximização da amplitude do gradiente e ainda os algoritmos baseados na minimização da entropia do histograma propostos por Jarvis [Jarvis 76] e Schlag [Schlag 83].

Operadores para medida da focagem

Para medir automaticamente o grau de focagem de uma região na imagem precisamos de um critério para medir o grau de nitidez da região. Analisando este fenómeno segundo a óptica física, a desfocagem começa a notar-se quando deixam de aparecer as componentes de alta-frequência da energia luminosa que chega ao dispositivo óptico. Se o

sistema óptico que estamos a utilizar tem uma abertura circular e considerarmos um ponto **P** na cena, então a imagem focada de **P** será um ponto **p'** (ver figura 2.8). Se o sensor de imagem é colocado noutra ponto do eixo óptico então o ponto **P** não estará focado e a sua imagem será a uma área circular no sensor, correspondente à intersecção do plano imagem com o cone de luz que atravessa a lente (ver figura 2.9). Essa área define um domínio que é circular, podendo ser descrito matematicamente o seu efeito através de uma função de gere um resultado idêntico. Dois tipos de funções são normalmente utilizadas [Horn 86][Subbarao 88] [Hwang 89]: a função de dispersão idêntica a uma função gaussiana bidimensional

$$h_1(x, y) = \frac{1}{2\pi\sigma^2} e^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)} \quad (5.1)$$

onde σ é o parâmetro de dispersão, que é proporcional ao raio do círculo de indefinição; a função circular

$$h_2(x, y) = \begin{cases} \frac{1}{\pi r_c^2} & \text{se } x^2 + y^2 \leq r_c^2 \\ 0 & \text{noutros casos} \end{cases} \quad (5.2)$$

sendo r_c definido pela expressão (2.5) e assumindo que a intensidade luminosa no interior é aproximadamente constante.

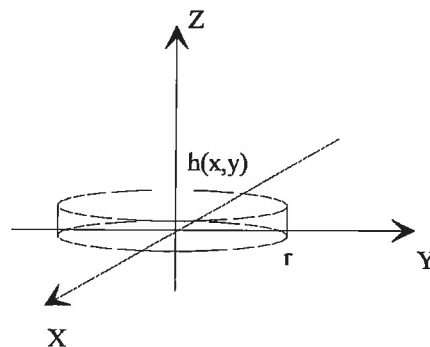


Figura 5.1- Forma da função $h_2(x, y)$.

Em ambos os casos, a imagem indefinida pode ser identificada com o resultado do processamento da convolução de uma imagem ideal $I(x, y)$ com uma função $h(x, y)$

$$I(x, y) * h(x, y) \quad (5.3).$$

Para verificar o efeito destas funções no domínio da frequência podemos analisar as suas transformadas de Fourier. Para o caso da função (5.1) a sua transformada de Fourier é dada por

$$\mathbf{H}_1(f_x, f_y) = \int \int_{-\infty}^{+\infty} \frac{1}{2\pi\sigma^2} e^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)} e^{-j(xf_x+yf_y)} dx dy = e^{-\frac{1}{2}(f_x^2+f_y^2)\sigma^2} \quad (5.4)$$

A equação (5.4) mostra que a transformada de Fourier é também uma função circular e simétrica no domínio da frequência. Esta função, quando aplicada à imagem ideal $\mathbf{I}(x,y)$, reduz fortemente a amplitude das componentes de alta frequência para frequências acima de $(1/\sigma)$. No caso da função circular a transformada de Fourier é dada por

$$F\{\mathbf{h}_2(x,y)\} = 2 \frac{J_1\{r_c \sqrt{f_x^2 + f_y^2}\}}{r_c \sqrt{f_x^2 + f_y^2}} = 2 \frac{J_1\{r_c R\}}{r_c R} = \mathbf{H}_2(f_x + f_y) \quad (5.3)$$

onde J_1 é a função de Bessel de primeiro tipo e de primeira ordem [Schreiber 86] [Lim 90].

Esta transformação é circular, simétrica e tem um pico central. A função $\frac{J_1\{r_c R\}}{r_c R}$ tem, para duas dimensões, um comportamento idêntico à função $\frac{\sin(x)}{x}$ nos sistemas unidimensionais. A convolução de uma imagem com esta função equivale a reduzir em amplitude as suas componentes de alta frequência. Como a função oscila em torno do zero, algumas das frequências serão totalmente anuladas nos locais onde a função é zero, não existindo possibilidade de as recuperar a partir da imagem desfocada. Estas funções de dispersão baseia-se em considerações geométricas mas no caso de um sistema óptico é necessário considerar vários outros efeitos. Uma fonte de distorção da imagem é a difracção devida à natureza ondulatória da radiação luminosa. As funções de transferência para diferentes formas de intensidade luminosa podem ter expressões mais elaboradas [Subbarao 87].

Pelas funções de transferência anteriores concluímos que o processo de desfocagem é equivalente a uma filtragem das frequências altas da radiação luminosa. Visto que o processo de desfocagem funciona como um filtro passa-baixo, a largura de banda da energia luminosa decresce com o aumento da desfocagem. Esta conclusão sugere a quantificação das componentes de alta-frequência existentes na imagem como critério de qualidade da focagem. A transformada de Fourier será um operador possível para obter uma quantificação das componentes de alta-frequência. No entanto, devido à quantidade de cálculos necessários realizar, esta transformação não é normalmente considerada face a outro tipo de operadores.

Como a desfocagem altera as características das arestas, fazendo com que se tornem menos acentuadas na imagem, é natural utilizar um detector de arestas como operador de uma medida de qualidade da focagem. O gradiente é um dos operadores que pode ser utilizado na detecção de arestas. O operador de Sobel realiza de forma aproximada este operador e consiste na convolução da imagem com as máscaras

$$i_x = \frac{1}{4} \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad i_y = \frac{1}{4} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}.$$

Para este operador a amplitude do gradiente é determinada por

$$S = \sqrt{(i_x * I(x, y))^2 + (i_y * I(x, y))^2} \quad (5.8).$$

Neste caso a função critério para a qualidade da focagem deverá procurar o máximo de S em diferentes imagens. Outro operador possível é o Laplaciano

$$\nabla^2 I = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}$$

simulado pela máscara

$$L = \frac{1}{6} \begin{bmatrix} 1 & 4 & 1 \\ 4 & -20 & 4 \\ 1 & 4 & 1 \end{bmatrix} \quad (5.9)$$

que deverá ser convolucionada com a imagem. Sendo o Laplaciano um operador diferenciador de segunda ordem dará resultado zero em todos os locais excepto nas zonas em que se iniciam mudanças abruptas da intensidade. Este o operador tem uma resposta diferente de zero nas zonas de transição da intensidade e poderá ser utilizado como indicador da quantidade de arestas existentes numa imagem. Nesse caso a função critério de qualidade da focagem consistirá em procurar o máximo do módulo de L aplicado a diferentes imagens.

Na maioria das vezes, uma grande variância do histograma de cinzentos de uma imagem está associada com uma imagem nítida enquanto que uma baixa variância está associada a uma imagem pouco nítida. Este facto permite estabelecer outro critério de qualidade da focagem. Uma imagem desfocada quando é comparada com uma imagem focada apresenta menor gama de variação dos níveis de intensidade. Devido a este facto uma possível medida da focagem poderá ser a maximização da expressão correspondente à definição de variância

$$\sigma^2 = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N (I(i, j) - \mu)^2 \quad (5.10)$$

onde μ é a média da distribuição dos níveis de cinzento e N é número de "pixels" numa das dimensões da imagem.

Nas figuras seguintes mostra-se a evolução de alguns dos critérios acima abordados, para uma sequência de imagens com diferentes focagens. Como se poderá observar existe um máximo global e máximos locais.

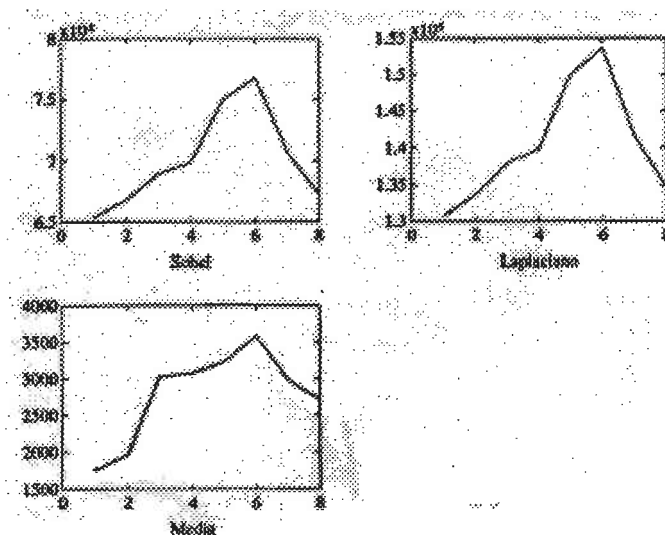


Figura 5.2- Comportamento dos operadores (5.8), (5.9) e (5.10) para uma sequência de imagens com diferentes focagens

Para minimizar, mas sem anular, a diferença entre o valor dos máximos locais relativamente ao máximo global podemos calcular o resultado da expressão

$$A = \frac{1}{N^2} \sum_{i=1}^{N^2} \sum_{j=1}^{N^2} |\nabla I(i, j)|^2 \quad (5.11)$$

aplicada a todo o domínio da imagem. Neste operador o módulo do gradiente é multiplicado de modo a salientar o valor do máximo global.

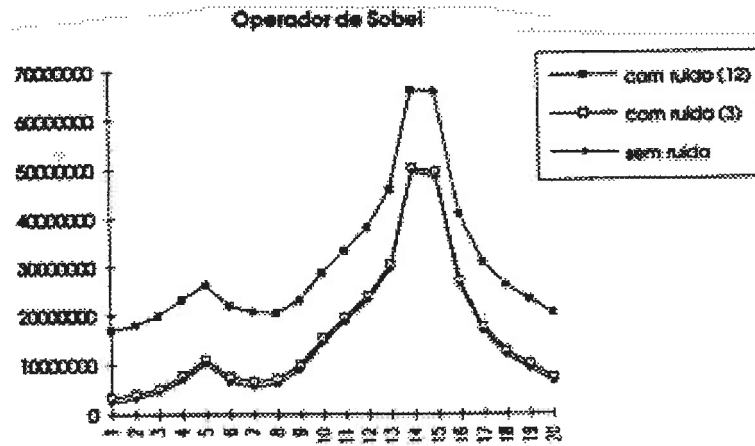


Figura 5.3- Comportamento do critério de Sobel para a mesma sequência de imagens com diferentes valores de ruído gaussiano, de média nula e com $\sigma^2=12$ e $\sigma^2=3$.

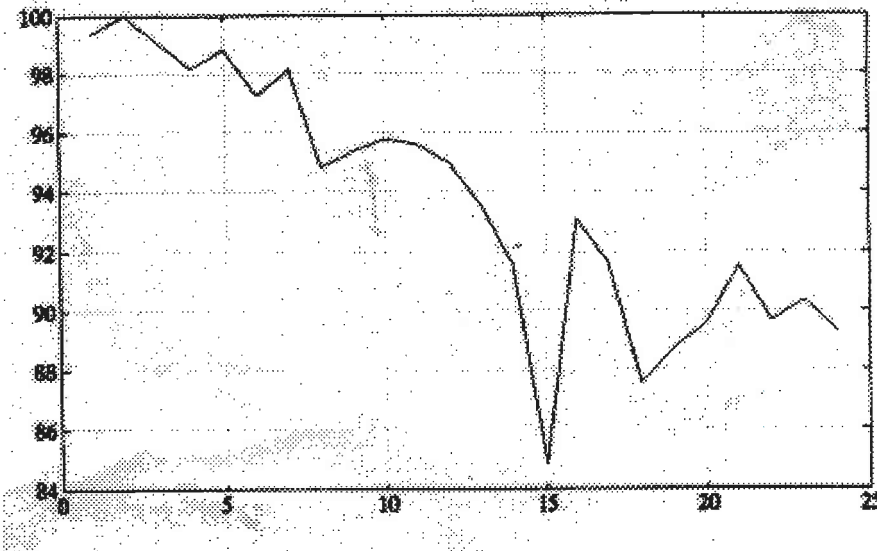


Figura 5.4- Comportamento dinâmico do operador (5.11) durante um intervalo de tempo correspondente a 24 minutos e para uma sequência de imagens de uma cena. Os valores são dados em percentagem do valor máximo.

Para o bom desempenho dos algoritmos de controlo da focagem, é necessário conhecer o comportamento da medida de qualidade de focagem ao longo do tempo e com condições constantes- iluminação e os objectos em cena. Seria de esperar que a curva fosse praticamente constante, mas tal facto não se verifica. Na realidade, para as câmaras CCD utilizadas no trabalho, verificou-se uma variação do valor máximo do critério de qualidade

de focagem. O valor dessa variação foi 20% para intervalos de tempo correspondentes a 15 minutos.

5.3. Estudo de um Algoritmo para Controlo da Focagem

Os algoritmos de focagem com base numa optimização global envolvem a análise do conjunto de imagens obtidas para todas as posições de focagem da lente. A mudança incremental da posição das lentes envolve um dispêndio grande de tempo. Uma maneira de minimizar esse dispêndio de tempo é realizando um controlo local.

Diversas estratégias tem sido propostas, entre elas a do teste do sinal correspondente à diferença entre os resultados da medida de qualidade da focagem, para duas posições sucessivas da lente. Vamos supor que o critério de qualidade consiste na pesquisa de um máximo do resultado de um dos operador descritos atrás. Se a diferença dos resultados entre duas posições sucessivas é negativa, então a movimentação das lentes está a provocar desfocagem e devemos fazer o movimento no sentido oposto [Horii 92]. Caso a diferença seja positiva podemos continuar a focar. Outro método consiste na detecção do máximo dos resultados obtidos por um dos operadores de medida de qualidade de focagem. A sua aplicação não é muito prática pois é necessário deslocar as lentes de modo a obter os resultados dos operadores nas várias posições possíveis das lentes.

Algoritmo Focagem

Define intervalo dt para cálculo da variação temporal ;
 Calcula a variação temporal dos resultados da aplicação do operador de medida da qualidade da focagem;
 Define um valor inicial para o resultado do operador de medida da qualidade da focagem;
Para sempre faz
 Adquire uma imagem e calcula o resultado do operador de qualidade de focagem;
 Se $\text{módulo}(\text{critério} - \text{critério inicial}) > \% \text{ variação temporal}$ **então**
 Muda para várias posições de focagem no sentido A.
 Determina o mínimo e o máximo dos resultados da aplicação do operador tendo em conta a variação temporal;
 Muda para várias posições de focagem no sentido B.
 Determina o mínimo e o máximo dos resultados da aplicação do operador tendo em conta a variação temporal;
 Define o maior dos máximos como sendo o valor inicial para o resultado do operador e define nova posição de imagem focada;
 Calcula a variação temporal do resultado durante um intervalo de tempo dt ;

fimse

fimpara

fimAlgoritmo

O algoritmo só actua quando a diferença entre o resultado actual e o anterior da medida de qualidade de focagem é superior ao valor correspondente à percentagem de erro devido à variação temporal do operador (ver ponto anterior).

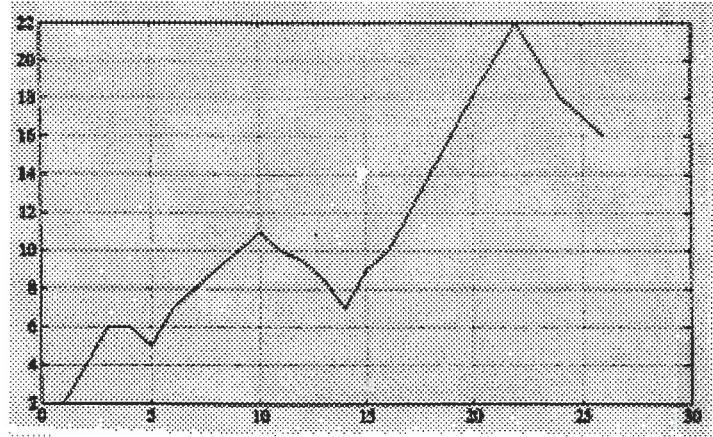


Figura 5.7 - A curva da figura corresponde a uma das curvas típicas dos valores do critério de focagem. O algoritmo anterior aplicado a esta situação pesquisa o máximo global de maneira eficaz para diferentes tolerâncias do valor do critério.

5.4. Calibração da Focagem

Para a efectiva utilização do sistema é necessário fazer uma calibração das posições angulares das lentes e as distância correspondentes ao plano de focagem (plano em que se dá o máximo do critério de focagem). Esta calibração utiliza uma grelha colocada a diferentes distâncias e consiste na determinação experimental de vários valores para os ângulos das lentes e respectivas distâncias da câmara ao plano de focagem.

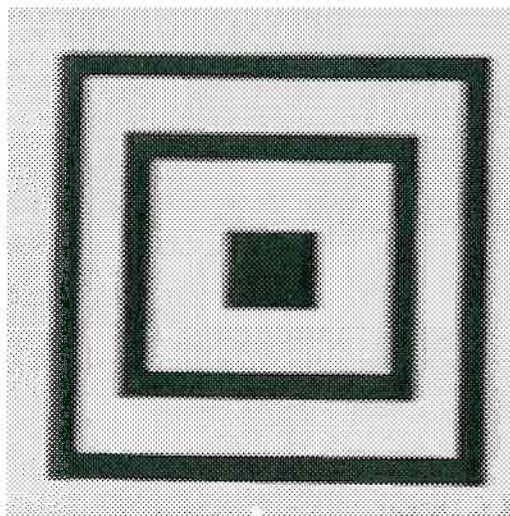


Figura 5.8 - Grelha utilizada na calibração da focagem

O resultado dessa calibração é uma tabela de valores (z_i, k_i) , com z_i a distância da câmara ao alvo e k_i a posição angular da lente. A distância d_{in} pode ser expressa por $d_{in} = \alpha k + f$, em que f é a distância focal, k a posição angular da lente e que obedece à restrição $(k = 0) \Rightarrow (d_{in} = f)$ e α o parâmetro a estimar. A distância z_i é medida a partir de um ponto que pode ser relacionado com d_{out} através de $d_{out} = z_i + X$ em que X é um parâmetro a estimar. A partir da expressão (2.4) obtemos a seguinte expressão para um polinómio que representa de forma aproximada os valores da tabela (z_i, k_i)

$$z_i = a + \frac{b}{k_i} \text{ com } a = f - X \text{ e } b = \frac{f^2}{\alpha} \quad (5.12)$$

Podemos transformar a expressão (5.12) numa expressão idêntica a uma recta do tipo $y = mx + b$ pela transformação $x = 1/k_i$. Aplicando uma técnica de regressão linear ao conjunto dos pontos da tabela 1 obtemos os valores para as constantes a e b do polinómio. Na figura abaixo apresentam-se os valores obtidos e respectivos gráficos para o polinómio interpolador dos valores da tabela 5.1.

DISTÂNCIA (m)	ÂNGULO (graus)	POSIÇÃO (0.4* graus)	ABERTURA (1.8)	DISTÂNCIA (m)	ÂNGULO (graus)	POSIÇÃO (0.4* graus)	ABERTURA (1.8)
2.5	82.5	33	337179930	1.75	122.5	49	338182332
2.45	85	34	346797072	1.70	125	50	389802674
2.4	87.5	35	343521500	1.65	130	52	307835180
2.35	89	35.6	351525182	1.60	132.5	53	309139020
2.3	90	36	355744976	1.55	137.5	55	306524512
2.25	92.5	37	343277740	1.5	145	58	304277866
2.2	95	38	330290648	1.45	150	60	297328424
2.15	97.5	39	337365392	1.4	157.5	63	291912674
2.1	100	40	351969916	1.35	162.5	65	287335702
2.05	102.5	41	338078444	1.3	170	68	282912998
2	105	42	338302780	1.25	177.5	71	285205302
1.95	107.5	43	353568520	1.20	187.5	75	280447982
1.9	110	44	355506634	1.15	197.5	79	297979804
1.85	115	46	363233074	1.1	207.5	83	369535732
1.8	120	48	339103238				

Tabela 5.1 - Tabela dos valores acumulados do operador de Sobel para imagens adquiridas para diferentes posições de focagem.

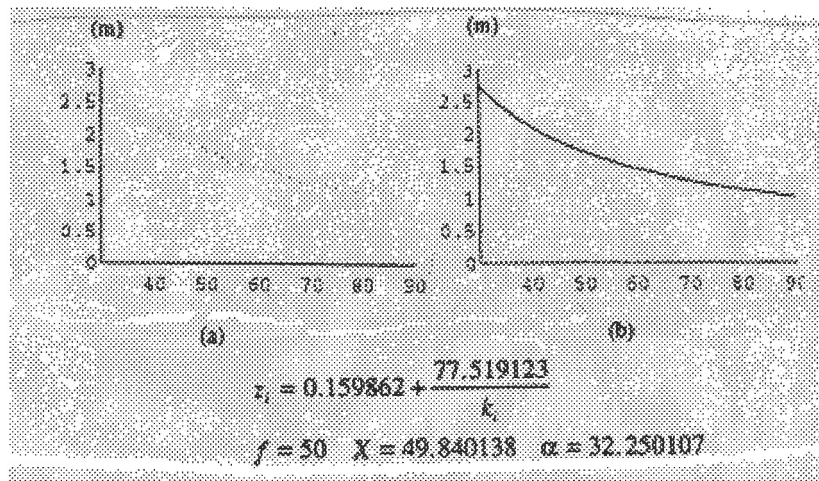


Figura 5.9 - Resultados da interpolação dos pontos da tabela 5.1. (a) Pontos medidos; (b) Representação da curva de interpolação e dos pontos medidos.

5.5 Segmentação da Imagem Utilizando Focagem

Se uma cena apresentar objectos a diferentes distâncias cujas áreas sejam relativamente iguais então as curvas do critério de focagem apresentam vários máximos. A figura 5.10 ilustra um desses exemplos.

A curva do critério de qualidade de focagem da imagem, ao apresentar mais do que um máximo local indica que existem objectos a distâncias diferentes relativamente à câmara. Este facto pode ser aproveitado para realizar uma segmentação por áreas de planos a diferentes profundidades gerando uma imagem segmentada por regiões. Simultaneamente esse processo de segmentação pode ser aproveitado para gerar um mapa de profundidade da cena.

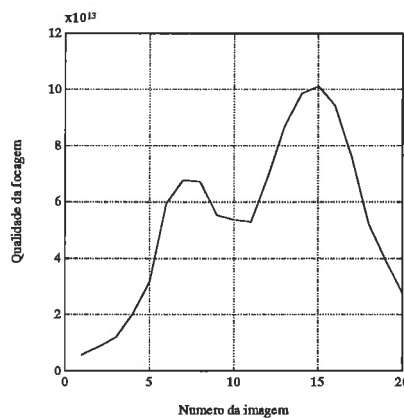


Figura 5.10 - Para uma cena com objectos com área sensivelmente igual a curva de critério de focagem apresenta mais do que um máximo local. O operador utilizado foi o operador expresso pela equação (5.8).

5.5.1. Cálculo da profundidade

Utilizando a equação das lentes delgadas e para imagens perfeitamente focadas obtemos a equação

$$d_{out} = \frac{f d_{in}}{d_{in} - f} \quad (5.13)$$

que relaciona a distância da lente à posição do sensor, que nas lentes mais vulgares equivale a um determinado deslocamento angular da lente. Mantendo a distância f constante e se a distância d_{out} entre o objecto e a lente for fixa poderemos focar uma imagem através da mudança da distância d_{in} entre o sensor e a lente. Se para as várias distâncias d_{out} for determinada a distância d_{in} equivalente através do processo de calibração (ver ponto 5.4), então podemos recuperar a profundidade da câmara ao objecto e por sua vez a um referencial mundo. Esta forma de determinação da profundidade é mais um processo de determinação da profundidade através de visão artificial.

5.5.2. Segmentação

O objectivo de uma segmentação é a partição de uma imagem em regiões homogêneas numa determinada propriedade. Normalmente a aproximação utilizada consiste na determinação das fronteiras entre regiões com base na diferença entre texturas ou intensidades. No nosso caso essa propriedade corresponde em obter regiões com distâncias relativamente idênticas utilizando os critérios de medida de qualidade da focagem. Se \mathbf{R} representa uma região na imagem então a aplicação do operador $\mathbf{P}(\mathbf{R})$ para a medida da qualidade de focagem terá que verificar $\mathbf{P}(\mathbf{R}) = \mathbf{P}(\mathbf{R}_i)$ para qualquer sub-região \mathbf{R}_i de \mathbf{R} .

Para realizar a segmentação começamos por verificar inicialmente se a curva obtida com os resultados da aplicação do operador de medida da focagem em diferentes posições de focagem apresenta diferentes máximos locais. Se apresenta então é porque existem zonas com diferentes distâncias e é possível separar as regiões responsáveis por esses máximos. O método de segmentação consiste na divisão sucessiva do conjunto de imagens por regiões de modo que as regiões finais contenham um único máximo. As figuras 5.11 e 5.12 ilustram a aplicação deste método num conjunto de imagens tiradas em diferentes focagens.

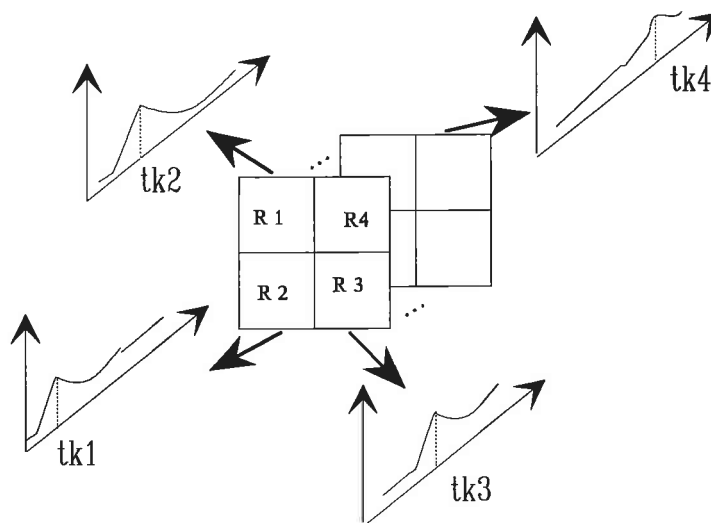


Figura 5.11 - Num conjunto de imagens com diferentes focagens, os máximos do critério de focagem são atingidos em diferentes locais (tk_i) e posições da lente.

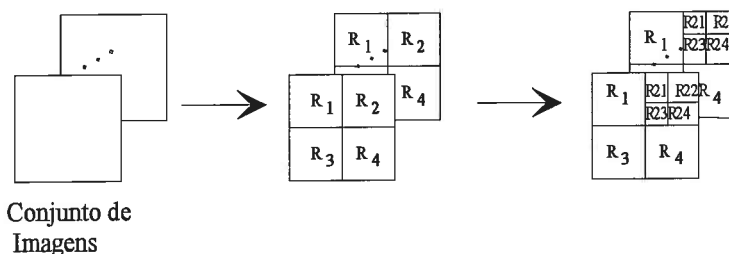


Figura 5.12 - Aproximações utilizadas para realizar a segmentação das imagens

Definições

ConjImag - Conjunto de imagens adquiridas em diferentes posições da lente.

ImageSegm - Imagem Segmentada.

DimImagem - Dimensões iniciais da imagem.

SupEsq - 1/4 superior esquerdo de uma imagem.

SupDir - 1/4 superior direito de uma imagem.

InfEsq - 1/4 inferior esquerdo de uma imagem.

InfDir - 1/4 inferior direito de uma imagem.

Algoritmo Segmentação_por_Focagem ("Top to down")

Adquire um conjunto de imagens em diferentes focagens (ConjImag);

Define uma imagem final como global (ImageSegm);

Procedimento ToptoDown(DimImagem);

Se pretende mapa de profundidade *então*

Com os índices de posição na imagem (ImageSegm) determina a imagem de profundidade utilizando a expressão (5.12);

fimse

fim Algoritmo


```
Procedimento ToptoDown(região)
Determina a curva de critério de focagem para a região
no conjunto de imagens(ConjImag);
Se a curva contém mais de um máximo então
  Divide a imagem de entrada em 4 regiões
  (SupEsq, SupDir, InfEsq, InfDir);
  Procedimento ToptoDown(SupEsq);
  Procedimento ToptoDown( SupDir);
  Procedimento ToptoDown(InfEsq);
  Procedimento ToptoDown( InfDir);
senão Preenche a região equivalente na imagem final (ImageSegm)
  com o índice da posição da imagem onde se atinge o
  máximo no conjunto(ConjImag);
fimse
fimProcedimento
```

Os resultados obtidos com o algoritmo e utilizando como critério de qualidade de focagem o módulo do operador de Sobel são

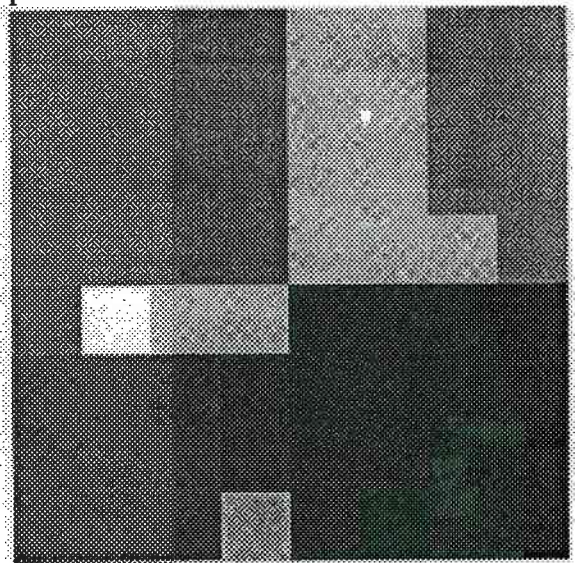


Figura 5.12a - Resultados da aplicação do algoritmo de focagem - Geração de uma imagem totalmente focada .

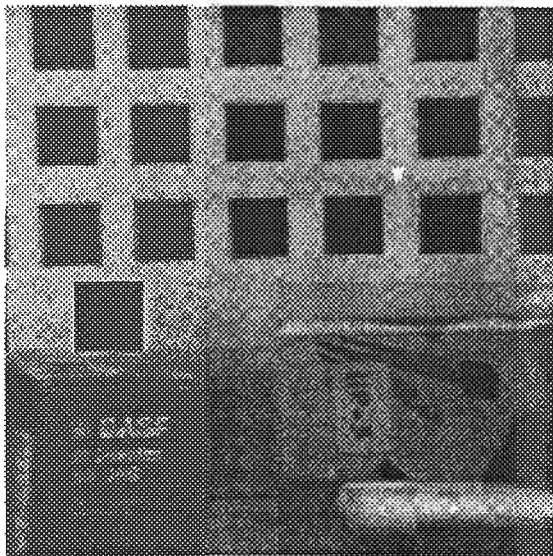


Figura 5.12b - Resultados da aplicação do algoritmo de focagem - Imagem da segmentação.

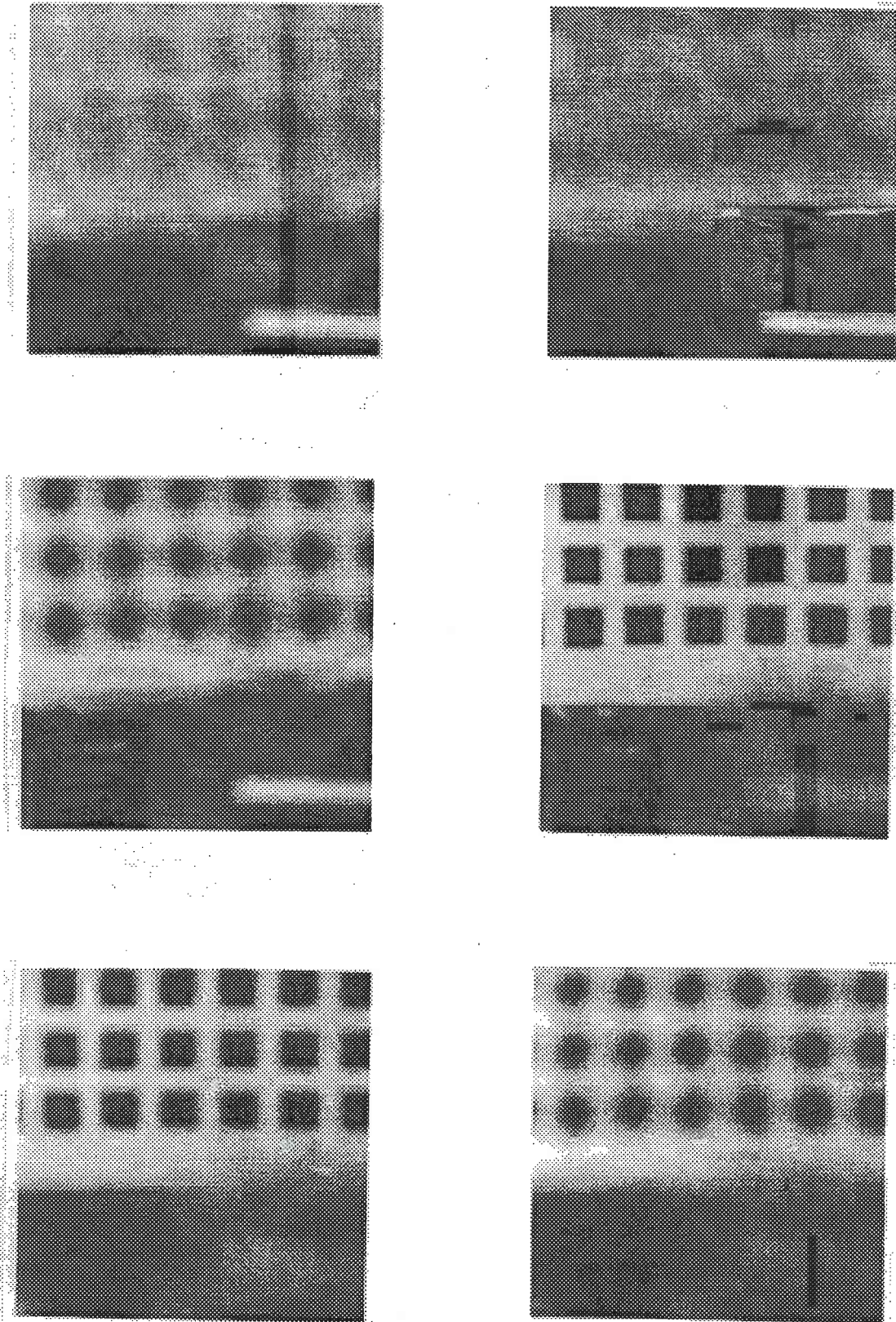


Figura 5.12c - Resultados da aplicação do algoritmo de focagem - Amostras do conjunto de imagens.

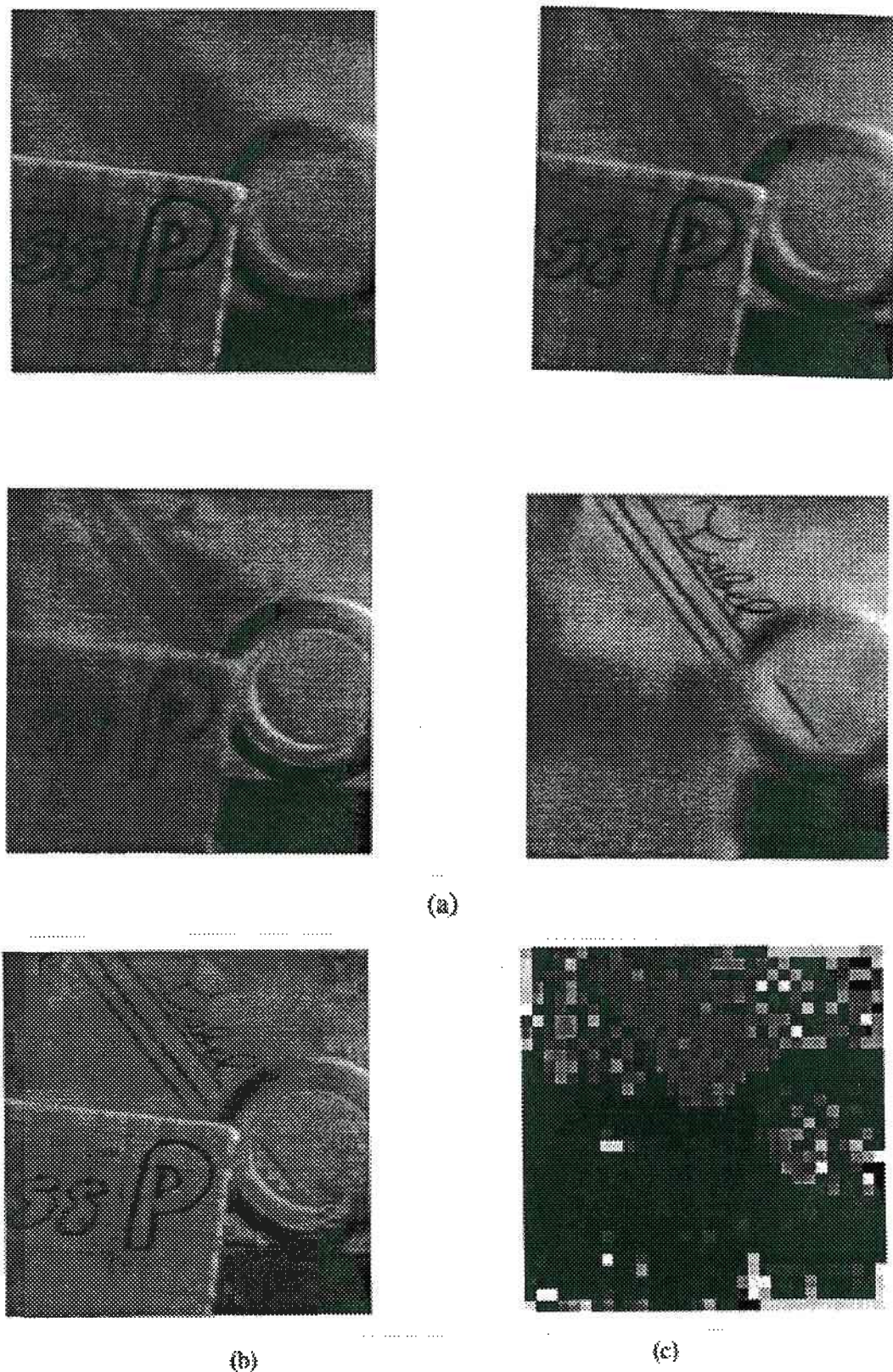


Figura 5.13 - Resultados da aplicação dos algoritmos de focagem. (a) Amostras do conjunto de imagens; (b) Geração de uma imagem totalmente focada; (c) Imagem do mapa de profundidade (mais escuro representa mais perto)

5.6. Segmentação de Imagens Utilizando Cooperação entre Focagem e Cor

O objectivo da segmentação é a identificação de objectos numa imagem delimitados pelas suas fronteiras materiais. Muitos dos processos de segmentação que têm sido propostos, baseiam-se nas propriedades cromáticas dos objectos para realizarem as interpretações das imagens [Klinker 88][Bajcsy 90][Healey 89]. Os resultados da aplicação deste princípio, conduzem a imagens segmentadas por regiões, que por vezes ultrapassam as fronteiras materiais dos objectos, dão origem a "buracos" nas regiões ou definem como regiões os locais que exibem fortes variações de cor e de intensidade [Klinker 90]. Neste ponto, propomos um algoritmo para segmentação de imagens de cor baseado nas regiões iniciais obtidas pelo processo de focagem descrito no ponto anterior. Com base nessas regiões iniciais é desenvolvido um algoritmo que realiza o crescimento dessas regiões de modo gerar uma imagem com as regiões correspondentes aos objectos de cores diferentes existentes na cena. Os objectos possuem uma única cor.

Crítérios de homogeneidade e de fusão de regiões

O algoritmo de base começa por aplicar o processo de segmentação por focagem, proposto no ponto anterior aos lotes de imagens obtidas para as componentes R, G e B. O resultado deste processo são três imagens segmentadas, cujas regiões apresentam os valores obtidos pelo critério de focagem e que definem a posição da lente para a qual aquela região e naquela banda está bem focada. Conceptualmente esperar-se-ia que as regiões nas três imagens coincidissem na imagem e no valor da posição da lente. Tal não se passa na realidade porque, os valores fornecidos pelo critério de focagem, como dependem fortemente do contraste, apresentam pouca diferença entre fronteiras de cor. Por exemplo a componente R ("Red") apresenta pouco contraste na fronteira de um objecto castanho com um objecto verde. Devido a isso foram ensaiadas outras opções para a determinação das regiões iniciais. A opção foi utilizar as componentes correspondentes às diferenças (R-G), (R-B) e (G-B). Estas diferenças (em módulo) tendem a diminuir os efeitos de baixo contraste entre objectos de cores diferentes. Outra opção foi utilizar a componente I derivada da transformação HIS (ver capítulo 2). Como esta componente equivale ao brilho ou intensidade da imagem será equivalente ao resultado da aplicação do algoritmo anterior para imagens preto/branco. O primeiro critério de definição da região homogénea é então o conjunto de locais na imagem, cujo valor da função critério de focagem para as três bandas se encontra dentro de um determinado intervalo. Essas regiões são designadas por *regiões iniciais* e formam parte dos dados de entrada para a fase seguinte do algoritmo. A região

inicial é uma região que engloba o conjunto de pontos para o qual o valor da posição de focagem óptima, verifica o teste estatístico T (ver apêndice B), para cada uma das três componentes.

Após a obtenção destas regiões iniciais o algoritmo tenta o crescimento destas regiões iniciais utilizando os valores de tonalidade das imagens, após conversão para o modelo HIS (ver capítulo 2). Para isso é determinado o valor médio da *cor* (tonalidade) das regiões iniciais e a sua variância estatística. O critério de fusão de uma pequena região a uma região inicial que lhe seja vizinha baseia-se também no teste estatístico T aplicado a uma distribuição F (ver apêndice B). Assim assume-se que a hipótese das duas regiões pertencerem ao mesmo objecto não é rejeitável, se o valor da distribuição estatística F com 2 e (N_1+N_2-3) graus de liberdade for menor do que o valor dado pela expressão (5.14) para um dado valor de significância α .

$$\hat{F} = \frac{N_1 N_2 (N_1 + N_2 - 3)}{(N_1 + N_2)^2} (\hat{\mu}_1 - \hat{\mu}_2)^T \mathbf{W}^{-1} (\hat{\mu}_1 - \hat{\mu}_2) \quad (5.14)$$

Na expressão (5.14) a variável N_1 corresponde ao número de amostras da região a fundir e N_2 ao número de amostras da região. Os símbolos $\hat{\mu}_1$ e $\hat{\mu}_2$ são as médias da região a fundir e da região já existente e \mathbf{W} a matriz de covariância dada por

$$\mathbf{W} = \mathbf{S}_1 + \mathbf{S}_2 \quad (5.15)$$

em que

$$\begin{aligned} \mathbf{S}_1 &= (\mathbf{x} - \hat{\mu}_1)^T (\mathbf{x} - \hat{\mu}_1) \\ \mathbf{S}_2 &= (\mathbf{x} - \hat{\mu}_2)^T (\mathbf{x} - \hat{\mu}_2) \end{aligned} \quad (5.16)$$

Assim o critério de fusão de duas regiões com o mesmo valor cromático e com índice de focagem idêntico consiste em fundir regiões em que o valor \hat{F} , dado pela equação (5.14), é menor que o valor da distribuição F com 2 e (N_1+N_2-3) graus de liberdade e para um dado valor de significância α .

Na parte final do algoritmo é conveniente classificar com a mesma etiqueta todas as regiões com propriedades cromáticas idênticas, mas que se encontram dispersas na imagem. Ou seja, existe a necessidade de decidir a qual das regiões R_i já existentes, é do tipo de uma outra região R_j com $i \neq j$. Para isso usa-se a expressão (5.14) para determinar a semelhança entre duas quaisquer regiões. Foram realizadas experiências com níveis de significância de $\alpha = 1\%$, 5% e 25% .

Como a fusão e expansão de regiões é um processo iterativo, utiliza dados determinados em iterações anteriores. Para actualizar as grandezas estatísticas de forma rápida, deduziram-se expressões para a determinação da média, variância e da matriz \mathbf{S} .

Para obtenção da média para n amostras, conhecendo a média para $n-1$ amostras, utilizámos a expressão

$$\hat{\mu}_n = \frac{N-1}{N} \hat{\mu}_{n-1} + \frac{x_n}{N} \quad (5.16).$$

A expressão equivalente para a variância

$$s_n^2 = \frac{N-1}{N} (s_{n-1}^2 + \hat{\mu}_{n-1}^2) + \frac{x_n}{N} - \hat{\mu}_n^2 \quad (5.17).$$

em que $\hat{\mu}_n$ representa a média para n amostras, N total número de amostras, x_n valor da amostra actual e s_n^2 a variância para n amostras. Para o cálculo do valor estatístico F precisamos de proceder à actualização da matriz de covariância \mathbf{W} (5.15). Como utilizámos dois parâmetros para caracterizar uma região ("hue" e posição de focagem) a matriz \mathbf{S} é de tamanho 2×2 , com a forma

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \quad (5.18)$$

Como esta matriz é simétrica os elementos $S_{12} = S_{21}$ da matriz \mathbf{S} são dados por

$$\begin{aligned} S_{ii_n} &= S_{ii_{n-1}} + (x_n^i - \hat{\mu}_n^i)^2 + 2 \frac{N-1}{N} \hat{\mu}_{n-1}^i (\hat{\mu}_{n-1}^i - x_n^i) \\ &+ (N-1) \left[\left(\frac{1}{N^2} - \frac{2}{N} \right) (\hat{\mu}_{n-1}^i)^2 + 2 \left(1 - \frac{1}{N} \right) \hat{\mu}_{n-1}^i \frac{x_n^i}{N} + \frac{(x_n^i)^2}{N^2} \right] \end{aligned} \quad (5.19)$$

e por

$$S_{ij} = \sum_{k=1}^{n-1} (x_k^i \cdot x_k^j) + x_n^i \cdot x_n^j - \hat{\mu}_n^i \cdot \hat{\mu}_n^j \cdot N \quad \text{com } i \neq j \quad (5.20).$$

Descrição do algoritmo

Na fase inicial do algoritmo, as imagens são processadas utilizando o algoritmo de segmentação em profundidade. Na fase seguinte procede-se à formação das regiões iniciais utilizando os resultados da segmentação em profundidade. Estando criadas as regiões iniciais, procede-se à sua expansão utilizando os valores cromáticos.

Definições

RegMin - Região mínima na imagem e com uma forma quadrada.
R(i) - Uma região na imagem com uma forma quadrada e identificada pelo índice *i*.
Viz(R(i)) - Região mínima adjacente à região R(i).
Os vizinhos de uma região quadrada são os 8 vizinhos adjacentes.

Algoritmo Segmentação_de_imagens_cor

Define uma região mínima - RegMin ;
Procedimento Segmentação_por_Focagem: R, G, B; (R-G), (R-B), (G-B); H, S, I;
Procedimento Forma_Regiões_Iniciais;
Para cada região inicial R(i) **faz**
 Procedimento Expansão_da_Região(R(i)); **fimpara**
Para cada região não marcada como região inicial ou marcada como ambígua
 no momento da expansão **faz**
 Procedimento Resolve_Regiões_Ambíguas; **fimpara**
 Procedimento Fusão_das_Regiões_Finais;

fim Algoritmo

Procedimento Forma_Regiões_Iniciais

Para cada região vizinha Viz(R(i)) da região R(i) **faz**
 Se a região R(i) e a região vizinha Viz(R(i)) cumprem a definição de região inicial
 então Considera R(i) e a região vizinha Viz(R(i)) como a mesma região **fimse**
fimpara

fim Procedimento

Procedimento Expansão_da_Região(R(i))

Determina a fronteira da região R(i);
Para cada região RegMin da fronteira **faz**
 Para cada uma das janelas vizinhas adjacentes e não pertencendo à região R(i) **faz**
 Se a janela vizinha adjacente verifica o critério de fusão
 então **Se** não foi ainda considerada como sendo de outra região
 então funde-a com a região R(i);
 senão classifica-a como ambígua **fimse**

fimse

fimpara

fimpara

fim Procedimento

Procedimento Resolve_Regiões_Ambíguas

Para cada região classificada como ambígua **faz**
 Determina as funções discriminantes para cada uma das regiões que provocam ambiguidade;
 Fundea a região ambígua à região que apresenta melhor aproximação segundo
 o critério de decisão;

fimpara

fim Procedimento

Procedimento Fusão_das_Regiões_Finais

Para cada uma das regiões finais R(i) **faz**
 Para cada uma das regiões finais R(j) com $j \neq i$ **faz**
 Se R(j) e R(i) verificam o critério de fusão
 então funde as regiões R(i) e R(j);

fimse

fimpara

fimpara

fim Procedimento

Resultados experimentais

Os resultados apresentados baseiam-se em imagens de duas cenas reais e designadas cena A e cena B e ainda um conjunto de imagens geradas por computador. Na figura 5.14 (a), (b) e (c) estão representadas algumas das imagens correspondentes a estas cenas. As imagens captadas foram processadas segundo três opções diferentes. A primeira opção utiliza as componentes RGB de um lote de imagens, cujo resultado da aplicação da segmentação de regiões com diferentes profundidades, produz três imagens de regiões (ver figura 5.15(a)). As outras duas opções seguintes utilizam os critérios de "diferenças" (R-G, R-B, G-B) e o sistema HIS (ver figuras 5.15(b) e 5.15(c) respectivamente).

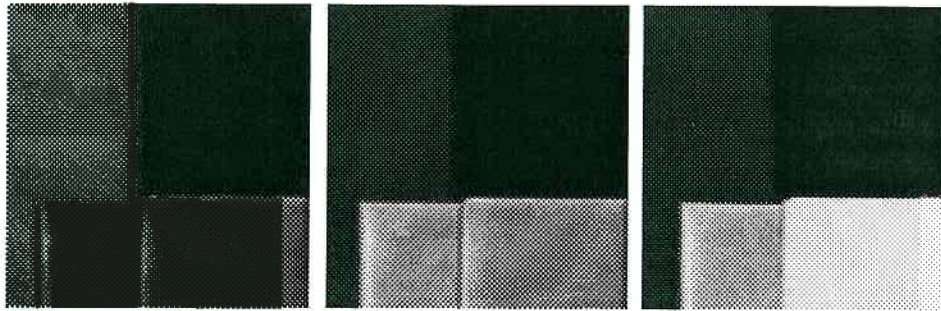


Figura 5.14a - Imagem da cena A e correspondente à "frame" 10 para componentes R, G e B respectivamente.

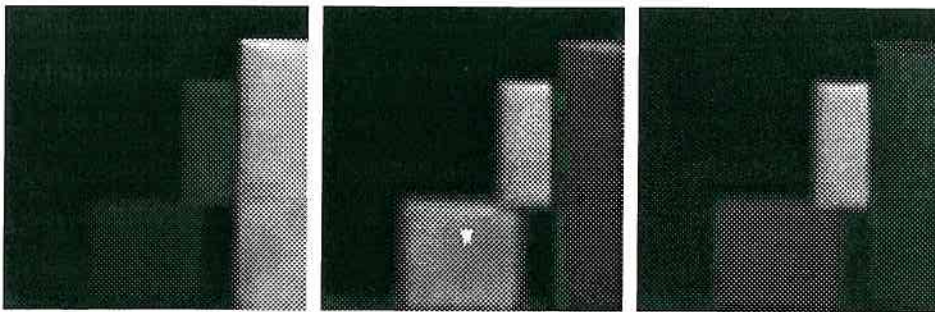


Figura 5.14b - Imagem da cena B e correspondente à "frame" 10 para componentes R, G e B respectivamente.

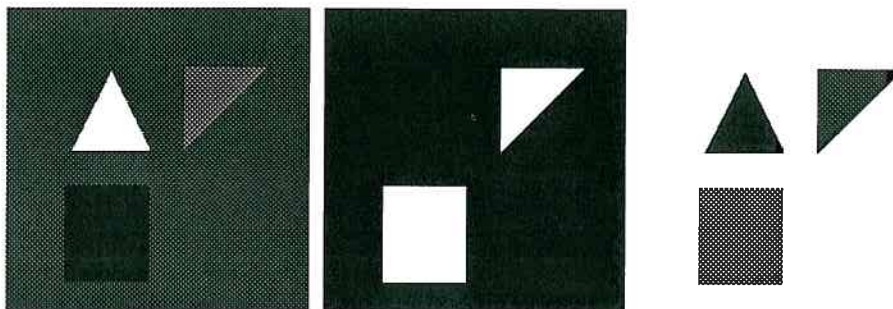
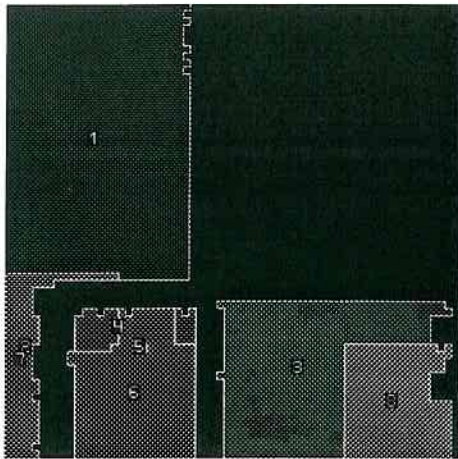
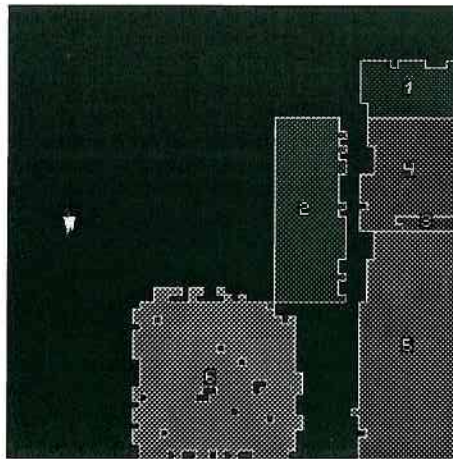


Figura 5.14c - Imagens de uma cena criada artificialmente para componentes R, G e B respectivamente

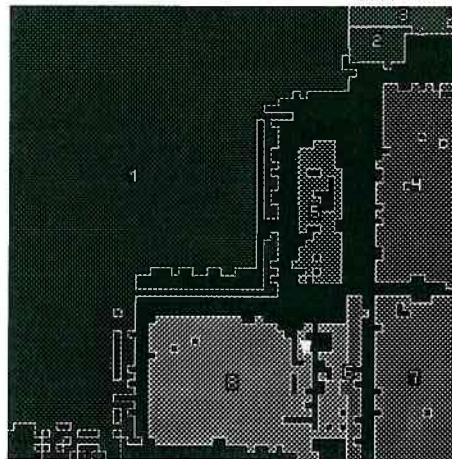


Cena A: $\alpha=5$, tratamento (R-G;R-B;G-B)



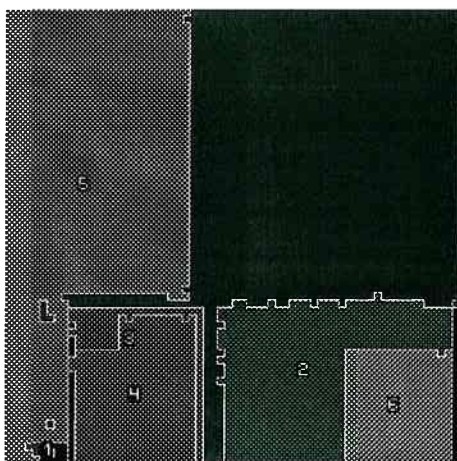
Cena B: $\alpha=5$, tratamento RGB

Figura 5.15a

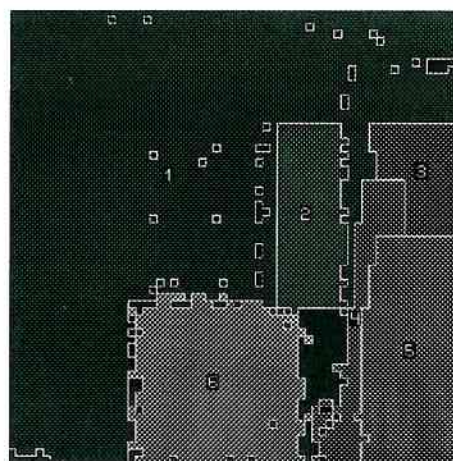


Cena B: $\alpha=5$, tratamento HIS

Figura 5.15a

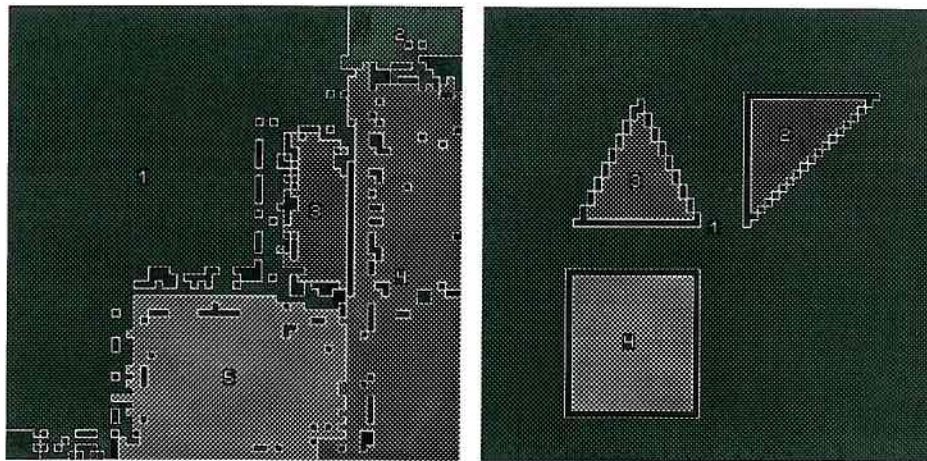


Cena A: $\alpha=5$, tratamento (R-G;R-B;G-B)

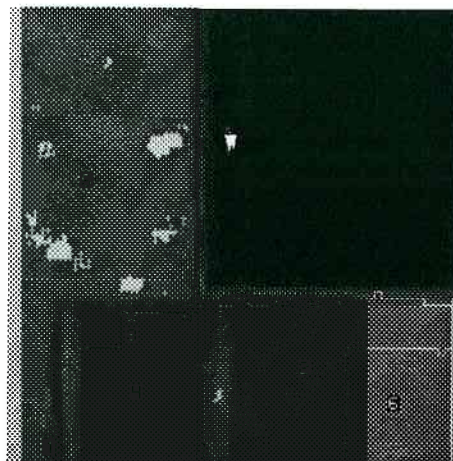


Cena B: $\alpha=5$, tratamento RGB

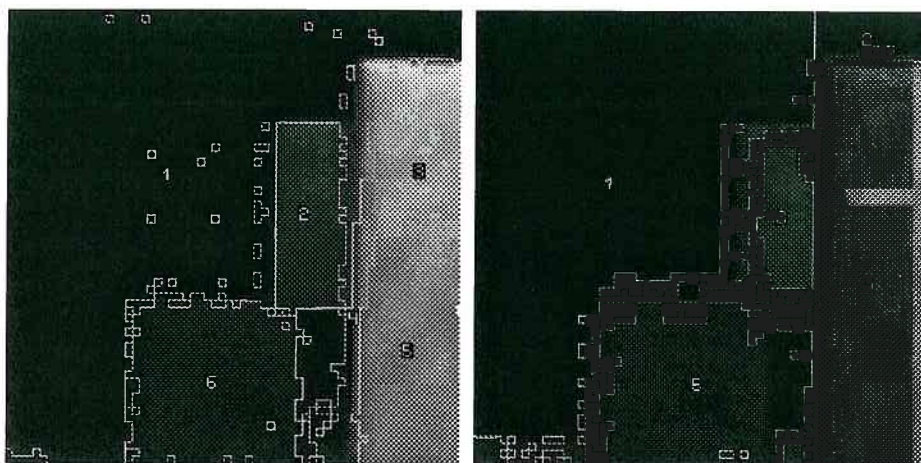
Figura 5.15b



Cena B: $\alpha=5$, tratamento HIS Imagem artificial $\alpha=5$
Figura 5.15b



Cena A: $\alpha=5$, tratamento (R-G;R-B;G-B)
Figura 5.15c



Cena B: $\alpha=5$, tratamento RGB Cena B: $\alpha=5$, tratamento HIS
Figura 5.15a

Figura 5.15 - Resultados da aplicação do algoritmo de segmentação de imagens de cor utilizando os resultados prévios da segmentação por focagem. (a) Conjunto de regiões iniciais; (b) Conjunto de regiões finais; (c) Imagens originais sobrepostas com as fronteiras das regiões segmentadas

Os resultados obtidos para diferentes valores do nível de significância α e para diferentes valores de limiar do operador de qualidade de focagem (Sobel), são apresentados nas tabelas seguintes.

limiar	Sobel		limiar de janelas (n° de janelas)	α	n° de regiões iniciais	n° de fusões	n° de regiões finais
	R	G					
20	25	30	25	1	109	169	8
20	25	30	25	5	110	169	8
20	25	30	25	25	114	203	7

Tabela 1- Imagens do lote A

limiar	Sobel			limiar de janelas (n° de janelas)	α	n° de regiões iniciais	n° de fusões	n° de regiões finais
	R-G	R-B	G-B					
30	30	20	25	1	94	227	6	
30	30	20	25	5	95	227	6	
30	30	20	25	25	98	235	6	

Tabela 2 - Imagens do lote A

limiar	Sobel			limiar de janelas (n° de janelas)	α	n° de regiões iniciais	n° de fusões	n° de regiões finais
	H	S	I					
30	30	30	25	1	21	100	6	
11	11	11	25	5	87	383	6	
15	15	15	25	25	59	206	6	

Tabela 3 - Imagens do lote A

limiar	Sobel			limiar de janelas (n° de janelas)	α	n° de regiões iniciais	n° de fusões	n° de regiões finais
	R	G	B					
20	25	30	25	1	265	2430	6	
20	25	30	25	5	272	2430	6	
20	25	30	25	25	285	2430	6	

Tabela 4 - Imagens do lote B

limiar	Sobel			limiar de janelas (n° de janelas)	α	n° de regiões iniciais	n° de fusões	n° de regiões finais
	R-G	R-B	G-B					
30	30	20	25	1	205	2546	4	
30	30	20	25	5	213	2546	4	
30	30	20	25	25	213	2546	4	

Tabela 5 - Imagens do lote B

limiar	Sobel		limiar de janelas (n° de janelas)	α	n° de regiões iniciais	n° de fusões	n° de regiões finais
	R-G	R-B					
14	14	14	25	1	150	584	5
14	14	14	25	5	159	620	5
14	14	14	25	25	174	681	5

Tabela 6 - Imagens do lote B

Conforme se pode verificar nas tabelas 1 a 6, o parâmetro α (valor de significância) é determinante na formação das regiões iniciais, regulando a selectividade das regiões com profundidade semelhante. No caso em que usarmos um valor de α menos restritivo (ex: $\alpha=1\%$) as regiões iniciais formadas são menos, mas de maiores dimensões, ao contrário de se usarmos um valor de α mais restritivo (ex: $\alpha=25\%$), em que existe um maior número das regiões iniciais. As regiões finais são dependentes das regiões iniciais e para evitar que existam também um número grande de regiões, é realizada uma filtragem de regiões demasiado pequenas, permitindo no final obter regiões mais coerentes.

As representações RGB e das "diferenças" (R-G) (R-B) e (G-B), apresentam melhores resultados pois a segmentação final tem regiões mais coerentes. Estes resultados devem-se sobretudo ao estágio de segmentação em profundidade em que existe uma definição mais rigorosa das áreas de profundidade igual.

5.7 Pormenores a Investigar

Nos algoritmos propostos nos pontos anteriores ignorou-se grande parte deste comportamento não ideal das lentes, como aliás tem sido ignorado na maioria dos algoritmos da Visão por Computador. No entanto tem interesse, em futuros refinamentos dos algoritmos, analisar o efeito da variação da ampliação com a focagem assim como o efeito da aberração cromática nos algoritmos propostos.

Tal como foi apresentado nos pontos anteriores, um método de recuperar a profundidade a que se encontra um objecto, consiste em realizar várias medidas de focagem sobre imagens com a lente em diferentes posições de focagem. Colocando a lente numa determinada distância focal (se for uma lente "Zoom") ou utilizando uma lente de distância focal fixa, a ampliação é dependente da distância do objecto e da distância ao sensor. Esta última distância tem influência na qualidade da focagem e portanto o controlo da focagem também intervém no valor da ampliação. Os operadores propostos no ponto 5.2 dependem do conteúdo da imagem, isto é, do tamanho das fronteiras dos objectos de maior contraste. Os valores fornecidos por estes operadores são fortemente influenciados por esse tamanho, o qual é influenciado pela ampliação. Os problemas causados pela mudança de ampliação podem ser discriminados em três tipos: deslocação espacial das primitivas, re-escalamento

dos perímetros das primitivas e re-escalamento do valor do gradiente de contraste. O primeiro efeito já foi salientado por Krotkov e outros [Krotkov 89] [Darrel 88] [Nayar 90]. Os outros dois efeitos poderão ser compensados através do controlo simultâneo do "zoom" e da focagem pois a ampliação é uma função da distância focal das lentes e da distância do objecto à lente (ponto 2.2.2). Note-se a ampliação da imagem é alterada pela variação da posição da focagem da lente embora o seu efeito seja muito menor do que a mudança produzida pela variação da distância focal, podendo ser desprezada.

Os aspectos anteriores referem-se sobretudo ao controlo da focagem de imagens preto/branco pois no caso das imagens de cor existem outros problemas adicionais. A análise de imagens de cor utiliza a informação contida nas três bandas espectrais para determinar as propriedades da cena que está a ser visualizada. Implicitamente, em qualquer análise de imagem de cor, assume-se que amostras com os mesmos índices (i, j) nas três bandas correspondem, no mundo objecto, ao mesmo ponto, região ou volume. No entanto este pressuposto não é verdadeiro em todos os casos pois o efeito da aberração cromática provoca diferença de ampliação entre bandas, diferença de focagem entre bandas e deslocalização dos objectos entre bandas [Willson 91]. Este efeito pode ser compensado por um processo de calibração, utilizando medidas sobre imagens a cor de um alvo e captadas com diferentes posições de focagem.

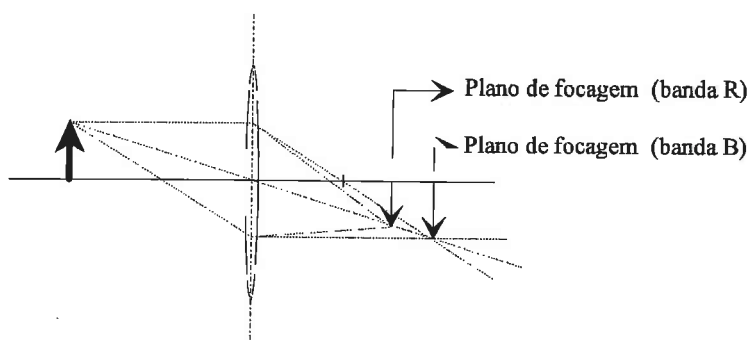


Figura 5.16 - Efeito da aberração cromática de uma lente delgada na captura de imagens RGB. O caso da focagem.

5.8 Relação com as técnicas de medida de desfocagem

A recuperação da profundidade utilizando a técnica proposta neste capítulo não é a única forma de recuperar a profundidade utilizando a focagem. A recuperação da profundidade através da medida da desfocagem é outra uma outra forma alternativa. A desfocagem pode ser considerada como o resultado da convolução de uma imagem focada, por um operador de desfocagem, tal como foi demonstrado no ponto 5.2. Determinando esse operador de desfocagem a partir de uma imagem desfocada é possível estabelecer a relação entre a desfocagem da imagem e a profundidade do objecto à câmara. Esta técnica

foi explorada principalmente por Pentland, Subbarao e por Ens, sendo uma alternativa à técnica proposta neste capítulo [Subbarao 87] [Subbarao 88] [Ens 93][Pentland 85][Pentland 87]. As soluções que têm sido apresentadas com base nesta técnica têm simplificado as soluções assumindo casos particulares de operadores de desfocagem. Pentland restringiu as suas soluções ao caso de operadores de desfocagem baseados na função Gaussiana. Todavia Subbarao argumenta que este modelo apresenta ainda algumas insuficiências [Subbarao 88].

Embora esta técnica possa ser encarada como forma alternativa à técnica proposta neste capítulo elas são distintas na forma como as imagens são adquiridas. Enquanto a nossa técnica é totalmente activa, requerendo várias imagens em diferentes posições de focagem e um pequeno processamento dos dados, a outra poderá requerer, no máximo, duas imagens com aberturas diferentes e um maior processamento dos dados para a determinação do operador de desfocagem [Pentland 87][Hwang 89].

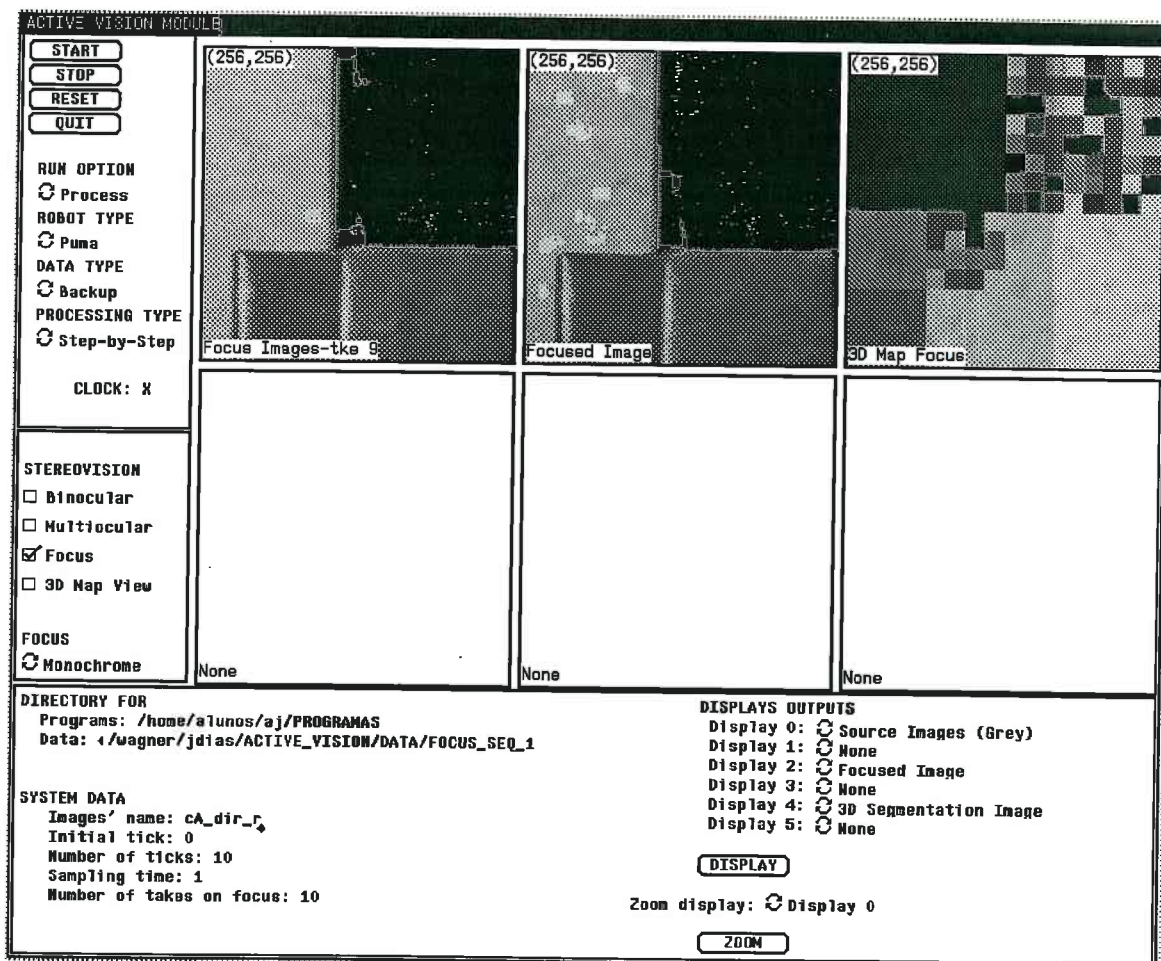


Figura 5.17 - A figura corresponde a uma imagem do programa utilizado para a realização das experiências. Devido à sua capacidade gráfica o utilizador pode alterar facilmente os parâmetros de entrada do programa e visualizar os resultados do processamento. A figura mostra a utilização deste programa para experiências de segmentação de imagens preto/branco utilizando focagem.

CAPÍTULO 6

ESTEREOVISÃO DINÂMICA

Estudo de um processo multiocular

6.1 Introdução

A recuperação da profundidade ou estrutura tridimensional de uma cena pode ser realizada através de duas imagens captadas com câmaras em configuração estereo. As duas imagens devem ser captadas em locais diferentes e de modo a causar um deslocamento relativo das primitivas visualizadas nas duas imagens. Esse deslocamento relativo é designado *disparidade* e permite, por intermédio de triangulação, a determinação da profundidade a que as primitivas se encontram das câmaras. Esta forma de recuperação da estrutura tridimensional tem sido bastante estudada durante os últimos anos, pois reflecte uma maneira natural de adquirir informação tridimensional sobre um objecto em cena. No entanto, a estereovisão está fortemente ligada à forma como é solucionado o problema da correspondência de primitivas ou regiões entre duas imagens estereo.

O processamento de imagens estereo com o objectivo de determinar a tridimensionalidade dos objectos pode ser decomposto nos seguintes passos: (1) calibração das câmaras; (2) detecção de primitivas nas imagens; (3) determinação da correspondência entre primitivas; (4) recuperação da tridimensionalidade. Estes passos correspondem a diferentes fases do processo algorítmico da visão estereo inicialmente proposto por David Marr [Marr 77a][Marr 77b][Marr 79][Marr 82][Grimson 85]. Marr propõe a utilização de pontos na imagem como primitivas e utiliza um método de relaxação para fazer a correspondência e garantir uma consistência global para o conjunto de correspondências. Moravec utiliza como primitivas, os pontos detectados através de um operador que determina zonas com simetria dos níveis de cinzento (ver Capítulo 4) [Moravec 77]. Para estabelecer a correspondência entre primitivas das duas imagens, Moravec propõe a utilização de um operador de correlação. Pollard propõe um método baseado numa restrição

com base no limite máximo do gradiente da disparidade. O gradiente da disparidade deve ser encarado como uma medida de continuidade das superfícies dos objectos existentes numa cena [Pollard 85]. Ayache propõe, no seu método de geração e verificação de hipóteses, a utilização de restrições baseadas nas relações de vizinhança entre primitivas [Ayache 87b]. Ayache e Medioni propõem a utilização de primitivas mais complexas como sejam os segmentos de recta [Ayache 87b][Medioni 85]. Embora a utilização de tais primitivas torne os algoritmos menos gerais, tem a grande vantagem de fornecer resultados mais consistentes.

Tradicionalmente, em estereovisão são utilizadas duas imagens binoculares para determinar informação tridimensional. Recentemente foram propostas técnicas trinoculares baseadas em três imagens adquiridas numa posição não-colinear [Ayache 87a][Ito 86][Pietikäinen 86][Yachida 86]. O objectivo destas técnicas é melhorar o desempenho de algoritmos no estabelecimento da correspondência entre primitivas. Dhoud, em [Dhoud 91], avalia o benefício e a complexidade adicional da utilização de uma terceira câmara. Estes algoritmos fornecem essencialmente uma restrição epipolar adicional que permite dar uma maior consistência às hipóteses de correspondência. Esta restrição é a principal diferença entre a estereovisão normal e triocular. Yachida propõe um algoritmo de estereovisão triocular utilizando imagens de contornos. O princípio baseia-se na utilização da restrição epipolar triocular para obter um conjunto de 3 pontos candidatos em cada uma das imagens e que simultaneamente satisfazem as restrições geométricas. Em caso de ambiguidade, devido à multiplicidade de pontos, é determinado o valor tridimensional do ponto e comparado com os valores dos que lhe estão em redor. Este procedimento é semelhante à restrição do gradiente de disparidade que limita o valor máximo do gradiente da disparidade, que se baseia no facto de não existirem grandes descontinuidades de profundidade nas superfícies dos objectos. Ito utiliza um processo idêntico ao anterior baseando-se também na restrição epipolar para estabelecer a correspondência local [Ito 86]. Neste caso as primitivas para a correspondência estereo são arestas que recebem um processamento posterior para realizar a ligação entre arestas. A partir disso a semelhança geométrica entre duas arestas é calculada utilizando a diferença de intensidade entre os seus vizinhos locais. Durante o processo não existe uma verificação acerca da consistência global.

Pietikäinen, em [Pietikäinen 86], utiliza um sistema triocular com uma geometria de eixos ortogonais. A geometria do sistema consiste numa câmara colocada na origem do sistema de coordenadas e mais outras duas câmaras deslocadas segundo a direcção vertical e horizontal respectivamente. As características locais das arestas, tais como a sua orientação e contraste de intensidade, são utilizadas como atributos para testar a semelhança local. A estratégia utilizada para estabelecer a correspondência consiste em testar inicialmente a multiplicidade de hipóteses na correspondência entre primitivas. Se não existe multiplicidade de hipóteses de correspondência então a correspondência é aceite, caso contrário, a

conectividade local dos contornos servirá para retirar a ambiguidade das correspondências e gerar uma única hipótese.

Ayache e Lustman desenvolveram uma técnica triocular baseada em segmentos de recta [Ayache 87a]. Para estabelecer a correspondência utilizaram a restrição epipolar e também restrições locais tais como, a orientação média da aresta, o valor médio do gradiente e o ponto médio do segmento. Um grafo de vizinhança é utilizado para armazenar a informação sobre a adjacência dos segmentos de recta em cada imagem. No método utilizado para fazer a correspondência entre primitivas de imagens diferentes, Ayache propõe um método de geração e verificação de hipóteses.

Neste capítulo propõe-se uma abordagem idêntica ao da visão triocular mas em que o conceito é alargado a multiocular. A possibilidade de movimentação das câmaras no espaço tridimensional e segundo uma trajetória conhecida permite a realização prática de um sistema deste tipo.

Analisando de forma informal, a estereovisão consiste numa técnica de determinar a distância a que os objectos se encontram numa cena, utilizando imagens. Essas imagens são adquiridas de tal maneira que entre cada imagem sucessiva pelo menos um parâmetro do sistema de captação das imagens é diferente. Depois de extrair valores numéricos apropriados usando uma operação da imagem, a distância ao objecto é obtida pela solução de um sistema de uma ou mais equações que relacionam esses valores numéricos com a distância das câmaras ao objecto. As equações intervenientes nesse sistema são resultado do modelo utilizado para a projecção das grandezas físicas na imagem. No caso tradicional e o mais simples as equações correspondem ao modelo de perspectiva que relaciona pontos da imagem 2D com pontos do mundo descritos em 3D.

Em resumo existem quatro quantidades distintas que caracterizam uma técnica de estereovisão. A primeira é a medida realizada para *caracterizar fisicamente o objecto*. A segunda é a *operação realizada nas imagens* para extrair os valores numéricos apropriados para resolver o sistema de equações. A terceira consiste no conjunto de *parâmetros de aquisição das imagens* em que pelo menos um dos parâmetros tem valores diferentes em posições diferentes. A quarta consiste nos *modelos utilizados para as câmaras* que relacionam os objectos com as suas imagens. Nesta fase de descrição do trabalho já parte destes parâmetros estão definidos. Neste capítulo a *característica física* a determinar é a medida de distância (profundidade absoluta). Os valores numéricos utilizados para o cálculo dessa distância baseiam-se nas primitivas extraídas nas imagens, pela aplicação dos *operadores* descritos no capítulo 4. O modelo utilizado nas câmaras foi já descrito no capítulo 2. Resta descrever mais em pormenor a forma como os parâmetros de aquisição variam e quais as equações que nesse caso, são utilizadas para determinar a distância absoluta. Esse assunto será abordado neste capítulo.

6.2 Geometria da Estereovisão

A visão estereo baseia-se em duas imagens adquiridas segundo a configuração ilustrada pela figura 6.1. Cada câmara tem um referencial ortonormado, com origem no centro da lente, C , e com o eixo-z coincidente com o eixo óptico da câmara. A câmara 1 tem centro C_1 e a câmara 2 tem centro C_2 . Para cada uma das câmaras são também conhecidas as matrizes de calibração que permitem localizar os seus referenciais relativamente ao referencial $\{W\}$. As imagens de cada câmara são designadas por I' e I'' e um ponto P tridimensional existente na cena, projecta-se nas imagens, segundo as coordenadas designadas por p' e p'' . Como ambos os pontos p' e p'' são uma projecção do mesmo ponto tridimensional, são designados pontos correspondentes. O principal problema do processamento de imagens estereo é a determinação dos pontos correspondentes nas duas imagens.

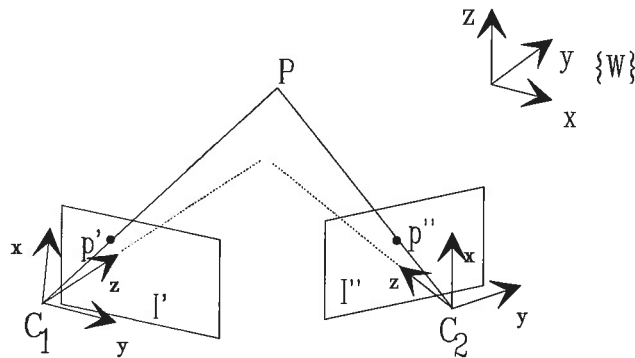


Figura 6.1 - Configuração de um sistema estereo. As câmaras têm os dois referenciais tridimensionais e ortonormados com origem nos centros das lentes C_1 e C_2 .

6.2.1 Pontos Epipolares e Rectas Epipolares

Se unirmos os centros das lentes C_1 e C_2 com um segmento de recta veremos que este intersecta os planos de imagem I' e I'' em dois pontos E_1 e E_2 . Estes dois pontos são conhecidos por *pontos epipolares*. Estes pontos podem ser determinados através da projecção dos centros das lentes nas imagens. Para determinar os pontos epipolares E_1 e E_2 é necessário calcular previamente os valores dos centros C_1 e C_2 . Estes valores são dados pelas soluções dos dois sistemas de três equações e três incógnitas:

$$\text{CALIB}_1 \begin{bmatrix} x_{C_1} \\ y_{C_1} \\ z_{C_1} \\ 1 \end{bmatrix} = \vec{0} \quad \text{e} \quad \text{CALIB}_2 \begin{bmatrix} x_{C_2} \\ y_{C_2} \\ z_{C_2} \\ 1 \end{bmatrix} = \vec{0} \quad (6.1)$$

Obtidas as coordenadas de cada centro $C_i = [x_{C_i} \ y_{C_i} \ z_{C_i}]^T$, $i=1,2$ as soluções para cada um dos pontos epipolares $E_i = [U_{E_i} \ V_{E_i} \ S_{E_i}]^T$ são obtidas por:

$$\text{CALIB}_1 \begin{bmatrix} x_{C_2} \\ y_{C_2} \\ z_{C_2} \\ 1 \end{bmatrix} = \begin{bmatrix} U_{E_1} \\ V_{E_1} \\ S_{E_1} \end{bmatrix} \quad \text{e} \quad \text{CALIB}_2 \begin{bmatrix} x_{C_1} \\ y_{C_1} \\ z_{C_1} \\ 1 \end{bmatrix} = \begin{bmatrix} U_{E_2} \\ V_{E_2} \\ S_{E_2} \end{bmatrix} \quad (6.2)$$

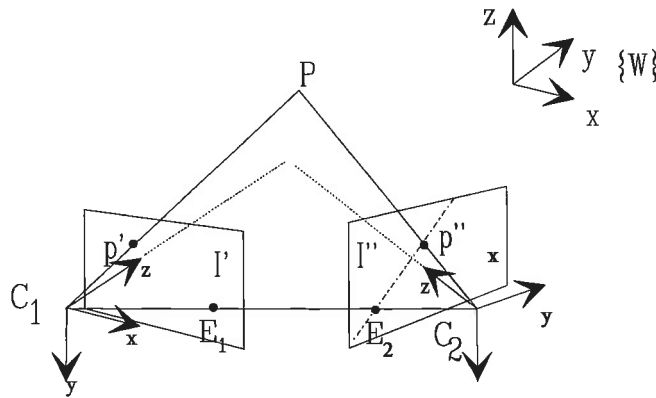


Figura 6.2 - As linhas epipolares passam pelos pontos epipolares E_1 e E_2 e pelo ponto da imagem correspondente.

Tal como ilustra a figura 6.2, o ponto P pertence à linha $C_1 p'$. Esta linha tem uma projecção no plano imagem da câmara 2 que passa pelo E_2 e contém o ponto p'' . Esta projecção é um segmento de recta na imagem e é designada por linha ou recta epipolar. Como esta recta é a projecção de uma recta que contém o ponto P , então conterá sempre a projecção correspondente do ponto na imagem. O mesmo princípio pode ser aplicado à câmara 1 relativamente ao ponto p' . Esta recta epipolar restringe a pesquisa dos pontos correspondentes à recta epipolar.

As duas matrizes de calibração **CALIB**, correspondentes às duas câmaras, relacionam as coordenadas (x, y, z) dos pontos no mundo e as coordenadas dos pontos (u, v) na imagem através da relação

$$\begin{bmatrix} U \\ V \\ S \end{bmatrix} = \mathbf{CALIB} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (6.3)$$

em que as coordenadas (u, v) são dadas por

$$u = \frac{U}{S} \quad e \quad v = \frac{V}{S} \quad (6.4).$$

Desenvolvendo a equação (6.3) obtemos a expressão

$$\begin{bmatrix} U \\ V \\ S \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \\ c_{31} & c_{32} & c_{33} & c_{34} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{c}_1^T \mathbf{P} + c_{14} \\ \mathbf{c}_2^T \mathbf{P} + c_{24} \\ \mathbf{c}_3^T \mathbf{P} + c_{34} \end{bmatrix}$$

e utilizando as equações (6.4) poderemos escrever a seguinte par de equações

$$\begin{cases} (\mathbf{c}_1 - u \mathbf{c}_3)^T \mathbf{P} + c_{14} - u c_{34} = 0 \\ (\mathbf{c}_2 - v \mathbf{c}_3)^T \mathbf{P} + c_{24} - v c_{34} = 0 \end{cases} \quad (6.5)$$

em que \mathbf{c}_i é um vector que representa as três primeiras colunas da linha i da matriz de calibração. Para cada uma das câmaras estas duas equações representam dois planos cuja intersecção é uma linha recta que passa pelo ponto \mathbf{P} do mundo, pela sua imagem \mathbf{p} e pelo centro da lente \mathbf{C} . Esta recta tem por direcção um vector que deverá ser perpendicular aos dois vectores normais aos planos. Se $(\mathbf{c}_1 - u \mathbf{c}_3)$ e $(\mathbf{c}_2 - v \mathbf{c}_3)$ são os dois vectores normais aos planos, então o vector director da recta projectante será dado pelo produto vectorial

$$\begin{aligned} \mathbf{n} &= (\mathbf{c}_1 - u \mathbf{c}_3) \wedge (\mathbf{c}_2 - v \mathbf{c}_3) = u(\mathbf{c}_2 \wedge \mathbf{c}_3) + v(\mathbf{c}_3 \wedge \mathbf{c}_1) + (\mathbf{c}_1 \wedge \mathbf{c}_2) \\ &= u \mathbf{g} + v \mathbf{h} + \mathbf{k} \end{aligned} \quad (6.6).$$

Como a recta passa pelo centro da lente \mathbf{C} , a expressão da recta projectante é

$$\mathbf{C} + \lambda \mathbf{n} \quad (6.7).$$

A recta projectante expressa no referencial da câmara 1 é dada por $\mathbf{C}_1 + \lambda \mathbf{n}$ e sua imagem na câmara 2 é obtida pela expressão:

$$\begin{bmatrix} U_2 \\ V_2 \\ S_2 \end{bmatrix} = \mathbf{CALIB}_2 \begin{bmatrix} C_1 + \lambda \mathbf{n}_1 \\ 1 \end{bmatrix} \quad (6.8).$$

Esta equação pode ser expressa através da seguinte forma vectorial

$$\begin{bmatrix} U_2 \\ V_2 \\ S_2 \end{bmatrix} = \mathbf{E}_2 + \lambda \mathbf{F}_2 = \begin{bmatrix} E_u \\ E_v \\ E_s \end{bmatrix} + \lambda \begin{bmatrix} F_u \\ F_v \\ F_s \end{bmatrix} \quad (6.9).$$

em que \mathbf{E}_2 é o ponto epipolar da câmara 2 e \mathbf{F}_2 é vector director da recta epipolar. Comparando (6.9) com (6.6) podemos obter uma expressão para o vector \mathbf{F}_2 através de

$$\mathbf{F}_2 = \begin{bmatrix} \mathbf{c}_1^T \\ \mathbf{c}_2^T \\ \mathbf{c}_3^T \end{bmatrix} [\mathbf{g} \quad \mathbf{h} \quad \mathbf{k}] \begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} \quad (6.10)$$

com

$$\mathbf{g} = (\mathbf{c}_2 \wedge \mathbf{c}_3), \mathbf{h} = (\mathbf{c}_3 \wedge \mathbf{c}_1), \mathbf{k} = (\mathbf{c}_1 \wedge \mathbf{c}_2) \quad (6.11)$$

em que (u_1, v_1) são as coordenadas do ponto \mathbf{p}' , resultado da projecção do ponto \mathbf{P} na imagem \mathbf{I}' . Utilizando a equação (6.9) e (6.4), podemos obter as coordenadas (u_2, v_2) de qualquer ponto da recta epipolar na imagem \mathbf{I}'' . A expressão para essas coordenadas é dada através de

$$\begin{cases} u_2 = \frac{E_u + \lambda F_u}{E_s + \lambda F_s} \\ v_2 = \frac{E_v + \lambda F_v}{E_s + \lambda F_s} \end{cases} \quad (6.12).$$

Resolvendo cada uma das expressões do sistema (6.12) em ordem a λ , obteremos a equação da recta na imagem na forma

$$u_2 \mathbf{A} + v_2 \mathbf{B} + \mathbf{C} = 0 \quad (6.13)$$

e equivalente ao determinante

$$\begin{vmatrix} u_2 & v_2 & 1 \\ E_u & E_v & E_s \\ F_u & F_v & F_s \end{vmatrix} = 0 \quad (6.14).$$

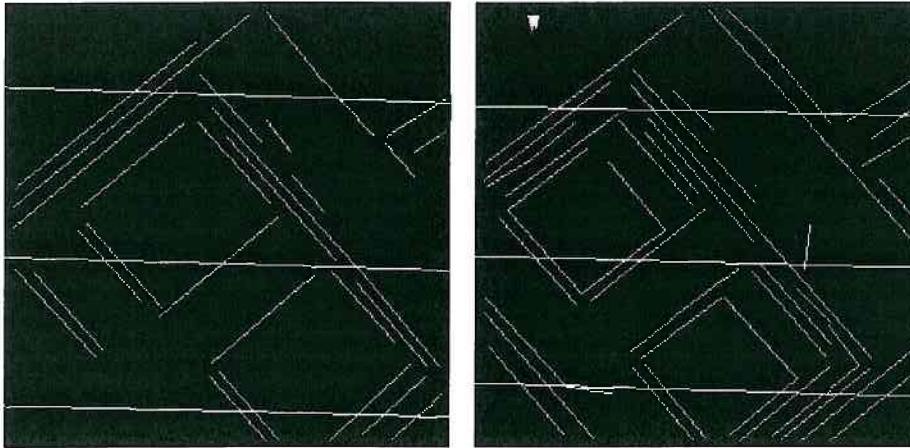


Figura 6.3 - Imagens estereo com rectas epipolares.

Problemas com as linhas epipolares

Existem casos em que as linhas epipolares não têm qualquer utilidade. São os casos em que:

- A recta que une o centro de perspectiva e o ponto no espaço tridimensional passa também pelo centro de perspectiva da outra câmara. Neste caso a recta epipolar traduz-se por um único ponto na imagem da outra câmara: ponto epipolar;
- O plano formado pelos dois centros de perspectiva e a recta de projecção de um ponto no espaço é paralelo ao plano imagem da outra câmara;
- A linha epipolar está fora da área representada pela imagem que foi adquirida.

6.2.2 Rectificação de Imagens

Quando os planos de imagem são coplanares e paralelos à recta que une os centros C_1 e C_2 das duas câmaras então os pontos epipolares E_1 e E_2 encontram-se no infinito e as rectas epipolares são um conjunto de linhas paralelas. Neste caso, para cada ponto da imagem $\mathbf{p}'=(u', v')$ é possível definir uma recta epipolar de modo que as coordenadas do ponto correspondente $\mathbf{p}''=(u'', v'')$ tem coordenadas $v''=v'$. Para isso é necessário definir cuidadosamente os referenciais das imagens. A transformação geométrica que, aplicada a cada uma das imagens, realiza esse objectivo designa-se *rectificação de imagem*.

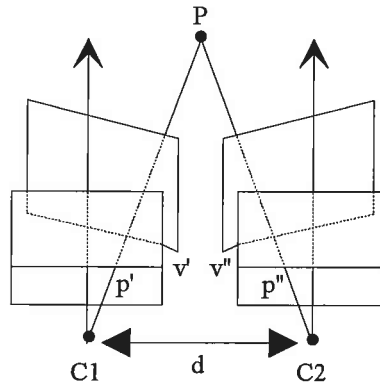


Figura 6.4- Colocando as imagens em planos coplanares e paralelos à recta que une os centros C_1 e C_2 então as rectas epipolares são paralelas e as projecções $\mathbf{p}'=(u', v')$ e $\mathbf{p}''=(u'', v'')$ têm coordenadas $v'=v''$.

As imagens rectificadas reduzem a procura de pontos correspondentes, a uma pesquisa segundo uma linha horizontal da imagem. Para além disso, o cálculo da disparidade entre pontos correspondentes reduz-se à diferença entre as coordenadas u' e u'' rectificadas.

Ayache propõe um método de rectificação de imagens para estereovisão binocular e triocular em que determina as novas matrizes de perspectiva com base num conjunto de restrições gerais que não tomam em conta as posições das câmaras [Ayache 88b]. O efeito prático desta aproximação é conduzir a situações de degeneração do cálculo das matrizes de perspectiva quando as câmaras se encontram em determinadas posições (por exemplo com um dos centros das câmaras no ponto $(0, 0, 0)$). O método que aqui é proposto assenta num processo diferente e que assegura que as distâncias focais são iguais para as duas câmaras.

O resultado da rectificação das duas imagens estereo é equivalente à projecção dos pontos tridimensionais segundo duas novas matrizes de projecção \mathbf{CALIB}_1^{rect} e \mathbf{CALIB}_2^{rect} . Essas matrizes definem um plano comum e paralelo à recta que passa pelos centros C_1 e C_2 e que contém as duas imagens rectificadas. Estas matrizes equivalem a duas novas câmaras virtuais cujos parâmetros devem obedecer às seguintes restrições:

- os centros de projecção dessas novas matrizes de projecção deverão ser ainda C_1 e C_2 ;
- a distância focal representada por essas duas matrizes \mathbf{CALIB}_1^{rect} e \mathbf{CALIB}_2^{rect} deverão ser iguais;
- para um ponto tridimensional $\mathbf{P}=(x, y, z)$ as suas imagens \mathbf{p}' e \mathbf{p}'' deverão ter coordenadas $v'=v''$.

A partir das matrizes de calibração das câmaras reais é possível determinar os valores das matrizes \mathbf{CALIB}_1^{rect} e \mathbf{CALIB}_2^{rect} que obedecem a estas condições.

A matriz de calibração de uma câmara é composta por duas matrizes- uma correspondente à perspectiva propriamente dita (parâmetros intrínsecos) e outra à posição da câmara no espaço (parâmetros extrínsecos). No caso das duas câmaras estereo

poderemos dizer que as matrizes de calibração para a câmara 1 e para a câmara 2, respectivamente $CALIB_1$ e $CALIB_2$, são dadas por

$$\begin{cases} CALIB_1 = \text{Persp} & {}^1T_w \\ CALIB_2 = \text{Persp} & {}^2T_w \end{cases} \quad (6.15).$$

Persp representa a matriz de mudança de perspectiva baseada nos parâmetros intrínsecos das câmaras e 1T_w e 2T_w representam as matrizes de parâmetros extrínsecos das câmaras e correspondentes à sua posição no espaço. O conhecimento destas matrizes permite determinar os centros C_1 e C_2 através da expressão (6.1). Conhecidos os centros das câmaras podemos definir um plano formado pelo versor \hat{j} da câmara 1 e o versor com suporte na recta que une C_1 e C_2 , com sentido de C_1 para C_2 e que tem por expressão

$$\hat{i} = \frac{\overline{C_1 C_2}}{\|C_1 C_2\|}.$$

Neste plano deverão estar as duas imagens rectificadas das câmaras virtuais que realizam o processo de rectificação. A orientação tridimensional das duas câmaras virtuais é dada pelos versores

$$\hat{j}, \hat{i} = \frac{\overline{C_1 C_2}}{\|C_1 C_2\|} \quad \text{e} \quad \hat{k} = \hat{j} \wedge \frac{\overline{C_1 C_2}}{\|C_1 C_2\|} \quad (6.16).$$

Estes versores são conhecidos relativamente a $\{W\}$ e definem as orientações das câmaras virtuais relativamente a $\{W\}$. Este processo permite determinar a rotação necessária para colocar as duas câmaras paralelas, mantendo as restrições acima. Assim, para obter as novas matrizes de projecção para as câmaras virtuais, basta alterar as matrizes 1T_w e 2T_w que representam as matrizes de parâmetros extrínsecos das câmaras reais. As duas novas matrizes de posição ${}^1T_w^{rect}$ e ${}^2T_w^{rect}$ serão dadas por

$$\begin{cases} {}^1T_w^{rect} = \left(\begin{bmatrix} \hat{i}_{(3x1)} & \hat{j}_{(3x1)} & \hat{k}_{(3x1)} & C_{1(3x1)} \\ 0 & 0 & 0 & 1 \end{bmatrix} \right)^{-1} \\ {}^2T_w^{rect} = \left(\begin{bmatrix} \hat{i}_{(3x1)} & \hat{j}_{(3x1)} & \hat{k}_{(3x1)} & C_{2(3x1)} \\ 0 & 0 & 0 & 1 \end{bmatrix} \right)^{-1} \end{cases} \quad (6.17).$$

Para obter as novas matrizes de projecção $CALIB_1^{rect}$ e $CALIB_2^{rect}$ teremos de aplicar a matriz de mudança de perspectiva **Persp**

$$\begin{cases} \mathbf{CALIB}_1^{rect} = \mathbf{Persp} & {}^1\mathbf{T}_W^{rect} \\ \mathbf{CALIB}_2^{rect} = \mathbf{Persp} & {}^2\mathbf{T}_W^{rect} \end{cases} \quad (6.18).$$

Estas matrizes de projecção têm a propriedade de gerarem imagens estereo com idêntica coordenada v para o mesmo ponto tridimensional \mathbf{P} .

No entanto, para rectificar as imagens, será mais interessante obter duas transformações que permitam gerar as imagens directamente a partir das imagens reais. Para isso tomemos por referência um ponto $\mathbf{p}=(u,v)$ na imagem da câmara 1 e correspondente à projecção de um ponto $\mathbf{P}=(x, y, z)$ no mundo. A equação paramétrica da recta projectante do ponto \mathbf{P} na imagem da câmara 1 é dada pela expressão (6.7)

$$\mathbf{P} = \mathbf{C}_1 + \lambda \mathbf{n}_1$$

em que o ponto \mathbf{C}_1 corresponde ao centro da câmara 1, cujos valores se deverão manter com o método de rectificação, e o vector \mathbf{n}_1 é dado pela expressão (6.6). A expressão para a projecção do ponto \mathbf{P} na imagem rectificada é dada por

$$\mathbf{p}'_{rect} = \mathbf{CALIB}_1^{rect} \mathbf{P} = \mathbf{CALIB}_1^{rect} \begin{bmatrix} \mathbf{C}_1 + \lambda \mathbf{n}_1 \\ 1 \end{bmatrix} = \mathbf{M} \mathbf{n}_1 \quad (6.19)$$

tendo em atenção que $\mathbf{CALIB}_1^{rect} \mathbf{C}_1 = 0$ e fazendo $\lambda=1$. A matriz \mathbf{M} é uma matriz (3x3) obtida a partir da matriz de calibração rectificada sem a 4ª coluna. Como o vector \mathbf{n}_1 é função de u e v então a matriz **Rect** que relaciona imagens rectificadas e imagens reais será dada por

$$\begin{bmatrix} U'_{rect} \\ V'_{rect} \\ S'_{rect} \end{bmatrix} = \mathbf{M} \begin{bmatrix} (\mathbf{c}_2 \wedge \mathbf{c}_3) & (\mathbf{c}_3 \wedge \mathbf{c}_1) & (\mathbf{c}_1 \wedge \mathbf{c}_2) \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{Rect} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \quad (6.20)$$

e pode ser determinada logo após o processo de calibração da câmara, não sendo necessário fazer o seu cálculo sempre que se pretende a rectificação das imagens. As expressões para o processo de rectificação das imagens da câmara 2 é idêntico ao desenvolvimento para a câmara 1.

As imagens abaixo apresentam exemplos de imagens antes e depois da rectificação.

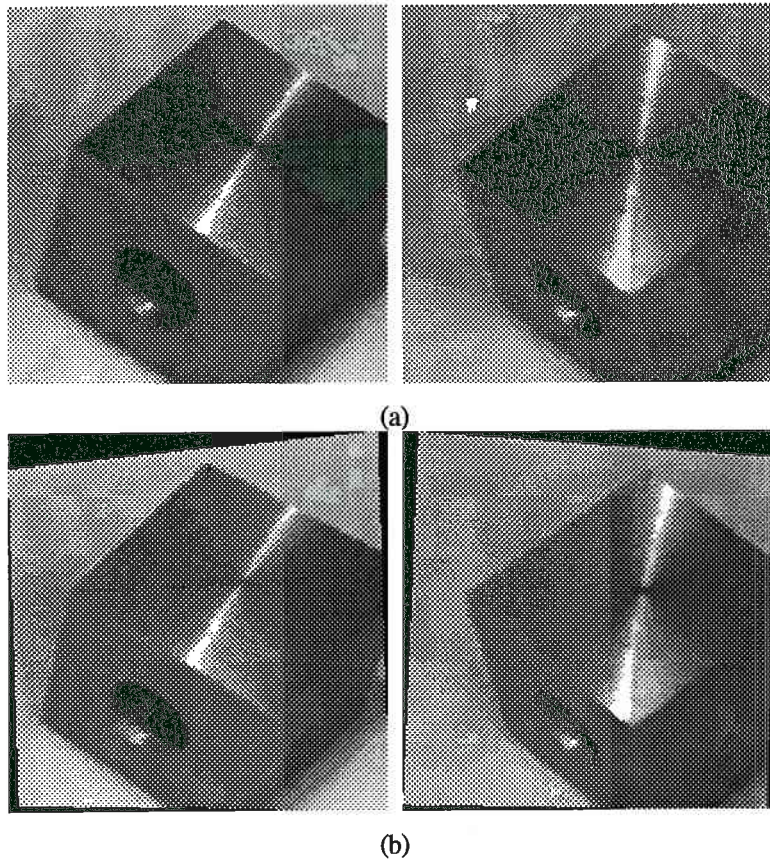


Figura 6.5 - As imagens representam imagens antes da rectificação (a) e após rectificação (b).

6.2.3 Profundidade Através da Visão Estereo

O caso de pontos

A profundidade de um ponto na cena, projectado nas duas imagens estereo, pode ser determinada se for conhecida a posição das suas projecções e a orientação e translação relativa entre as duas câmaras. Para isso é necessário conhecer a correspondência correcta entre as várias projecções existentes nas duas imagens.

Considerando o modelo para a formação da imagem já descrito anteriormente e a geometria da figura 6.1 a determinação das coordenadas tridimensionais de um ponto \mathbf{P} é feita com base nas suas imagens \mathbf{p}' e \mathbf{p}'' . As coordenadas (u, v) dos pontos \mathbf{p}' e \mathbf{p}'' na imagem e as coordenadas (x, y, z) do ponto \mathbf{P} no mundo real podem ser relacionadas através da expressão (6.5). Juntando as expressões para cada uma das câmaras, é possível escrever 4 equações com incógnitas (x, y, z) correspondentes às coordenadas do ponto \mathbf{P} que pretendemos obter. Estas equações são

$$\begin{cases} \left({}^1\mathbf{c}_1 - {}^1u \ {}^1\mathbf{c}_3 \right)^T \mathbf{P} + {}^1c_{14} - {}^1u \ {}^1c_{34} = 0 \\ \left({}^1\mathbf{c}_2 - {}^1u \ {}^1\mathbf{c}_3 \right)^T \mathbf{P} + {}^1c_{24} - {}^1v \ {}^1c_{34} = 0 \\ \left({}^2\mathbf{c}_1 - {}^2u \ {}^2\mathbf{c}_3 \right)^T \mathbf{P} + {}^2c_{14} - {}^2u \ {}^2c_{34} = 0 \\ \left({}^2\mathbf{c}_2 - {}^2u \ {}^2\mathbf{c}_3 \right)^T \mathbf{P} + {}^2c_{24} - {}^2v \ {}^2c_{34} = 0 \end{cases} \quad (6.21)$$

em que o índice 1 (ou 2) no canto superior esquerdo de uma variável ou vector, designa a câmara 1 (ou 2). Este sistema tem quatro equações e só três incógnitas. Sem um critério objectivo para eliminação de uma das equações, a solução terá que se basear num critério de optimização. Uma solução será a minimização do erro médio quadrático. O método da pseudo-inversa minimiza este erro e ainda permite a utilização de mais do que um par de imagens para determinar o ponto \mathbf{P} (ver Apêndice A).

Colocando o sistema (6.21) na forma matricial

$$\mathbf{A} \mathbf{P} = \mathbf{b}$$

com $\mathbf{P} = (x, y, z)$ e

$$\mathbf{A} = \begin{pmatrix} \dots \\ \left({}^i\mathbf{c}_1 - {}^iu \ {}^i\mathbf{c}_3 \right)^T \\ \left({}^i\mathbf{c}_2 - {}^iv \ {}^i\mathbf{c}_3 \right)^T \\ \dots \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} \dots \\ {}^iu \ {}^i c_{34} - {}^i c_{14} \\ {}^iv \ {}^i c_{34} - {}^i c_{24} \\ \dots \end{pmatrix}$$

em que o índice i designa a câmara (i) ou a posição tridimensional (i) de uma câmara. Utilizando esta forma matricial a solução pela pseudo-inversa é dada por

$$\mathbf{P} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b} = \mathbf{A}^+ \mathbf{b} \quad (6.22)$$

em que \mathbf{A}^+ é conhecida por matriz pseudo-inversa. Para existir uma solução através da expressão (6.22) é necessário que $(\mathbf{A}^T \mathbf{A})$ seja invertível.

O filtro de Kalman é outra possibilidade para estimação das coordenadas (x, y, z) de forma recursiva e que pode ser alargado à utilização de N imagens. No Apêndice A é fornecida alguma informação para a implementação deste processo.

O caso de rectas

Consideremos uma recta tridimensional representada pelas equações

$$\begin{cases} x = az + p \\ y = bz + q \end{cases} \quad (6.23)$$

mas que exclui as rectas perpendiculares ao eixo-z. Para uma representação deste tipo de rectas é necessário utilizar uma representação alternativa à expressão (6.23) e descrita no ponto 2.5.2. Definindo um vector de direcção \mathbf{n} e um deslocamento \mathbf{d} como sendo

$$\mathbf{n} = \begin{bmatrix} a \\ b \\ 1 \end{bmatrix} \quad \mathbf{d} = \begin{bmatrix} p \\ q \\ 0 \end{bmatrix} \tag{6.24}$$

é possível uma representação da recta (6.23) na seguinte forma vectorial

$$\mathbf{P} = \mathbf{zn} + \mathbf{d}.$$

A sua projecção nas imagens das câmaras são segmentos de recta 2D que podem ser expressos matematicamente através da forma

$$v = m u + l \tag{6.25}$$

sendo m o declive e l o valor de v quando $u=0$.

Estabelecida a correspondência entre os segmentos de recta é possível a reconstrução tridimensional pela determinação da intersecção entre os dois planos que passam pelos segmentos de recta nas imagens e pelos centros das câmaras. Na figura 6.6 ilustra-se o caso da reconstrução de uma recta, utilizando um par de imagens estereo.

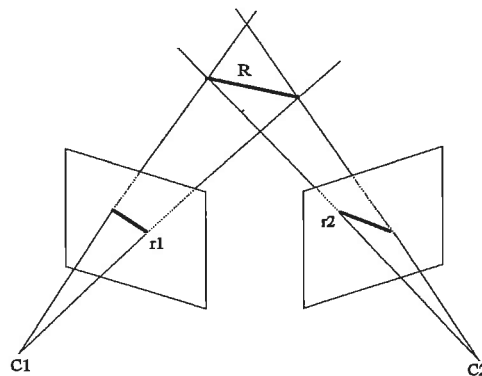


Figura 6.6 - Reconstrução de uma recta no caso de duas imagens estereo

No caso de serem utilizadas mais de duas imagens para a reconstrução tridimensional então o processo é mais complexo. As figuras abaixo representam o caso de imagens obtidas a partir de três planos.

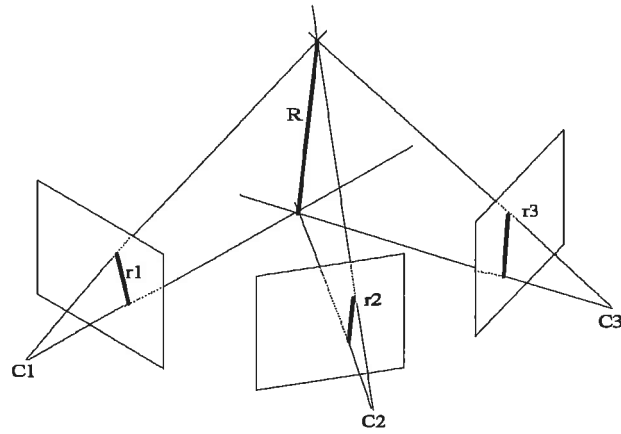


Figura 6.7 - Reconstrução de uma recta no caso de três imagens (caso ideal)

Normalmente, e devido ao ruído dos dados, a intersecção dos planos normalmente não se verifica no mesmo segmento de recta. Uma solução consiste em minimizar o erro de reconstrução. Para isso estabelecemos um sistema de equações para cada uma das imagens da recta com base nas equações dos planos (6.5). Esse sistema é posteriormente resolvido por um método que minimiza o erro médio quadrático, utilizando a matriz pseudo-inversa.

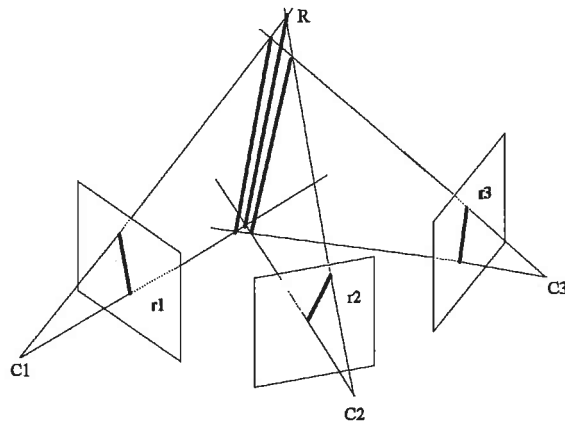


Figura 6.8 - Reconstrução de uma recta no caso de três imagens (caso real)

A determinação dos parâmetros de uma recta tridimensional é feita com base nos parâmetros dos segmentos de recta correspondentes à sua projecção nas imagens. Estes parâmetros são representados por um vector $\hat{\mathbf{a}}$

$$\hat{\mathbf{a}} = \begin{bmatrix} \hat{m} \\ \hat{l} \end{bmatrix}$$

cujos valores são estimados a partir de dados.

Consideremos uma recta \mathbf{R} , tridimensional, e que não é perpendicular ao eixo-z. Se $\mathbf{P}=(x,y,z)$ é um ponto genérico da recta \mathbf{R} então deverá verificar as equações

$$\begin{cases} x = az + p \\ y = bz + q \end{cases} \quad (6.26).$$

A imagem do ponto \mathbf{P} é dada pela equação de projecção

$$\begin{bmatrix} U \\ V \\ S \end{bmatrix} = \mathbf{CALIB} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & t_{13} & t_{14} \\ t_{21} & t_{22} & t_{23} & t_{24} \\ t_{31} & t_{32} & t_{33} & t_{34} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (6.27)$$

e as coordenadas (u, v) na imagem são dadas por

$$\begin{cases} u = \frac{(at_{11} + bt_{12} + t_{13})z + pt_{11} + qt_{12} + t_{14}}{(at_{31} + bt_{32} + t_{33})z + pt_{31} + qt_{32} + t_{34}} \\ v = \frac{(at_{21} + bt_{22} + t_{23})z + pt_{21} + qt_{22} + t_{24}}{(at_{31} + bt_{32} + t_{33})z + pt_{31} + qt_{32} + t_{34}} \end{cases} \quad (6.28)$$

Visto que o ponto \mathbf{P} pertence à recta \mathbf{R} então a sua projecção na imagem obedecerá à equação da recta 2D (6.25). Esta equação poderá ser escrita na forma

$$\hat{m}u + \hat{l} - v = 0 \quad (6.29)$$

podendo substituir as coordenadas u e v pelas expressões dadas em (6.28). A projecção na imagem dos pontos \mathbf{P} da recta \mathbf{R} deverão cumprir a relação (6.29) o que é o mesmo que cumprir as seguintes restrições

$$\begin{cases} \hat{m}(at_{11} + bt_{12} + t_{13}) - (at_{21} + bt_{22} + t_{23}) + \hat{l}(at_{31} + bt_{32} + t_{33}) = 0 \\ \hat{m}(pt_{11} + qt_{12} + t_{14}) - (pt_{21} + qt_{22} + t_{24}) + \hat{l}(pt_{31} + qt_{32} + t_{34}) = 0 \end{cases} \quad (6.30).$$

Rearranjando a equação (6.30) de modo a isolar os termos (a, b) e (p, q) é possível escrever uma equação linear do tipo

$$\begin{cases} a(\hat{m}_i t_{11} - t_{21} + \hat{l}_i t_{31}) - b(\hat{m}_i t_{12} - t_{22} + \hat{l}_i t_{32}) + (\hat{m}_i t_{13} - t_{23} + \hat{l}_i t_{33}) = 0 \\ p(\hat{m}_i t_{11} - t_{21} + \hat{l}_i t_{31}) - q(\hat{m}_i t_{12} - t_{22} + \hat{l}_i t_{32}) + (\hat{m}_i t_{14} - t_{24} + \hat{l}_i t_{34}) = 0 \end{cases} \quad (6.31).$$

A expressão (6.31) pode ser escrita numa forma mais compacta, utilizando matrizes e o índice (i) como identificador da imagem. Nesse caso teremos

$$\mathbf{n}_i = \mathbf{M}_i \mathbf{m} \quad (6.32)$$

com

$$\mathbf{n}_i = \begin{bmatrix} -(\hat{m}_i t_{13}^i - t_{23}^i + \hat{l}_i t_{33}^i) \\ -(\hat{m}_i t_{14}^i - t_{24}^i + \hat{l}_i t_{34}^i) \end{bmatrix} \quad (6.33)$$

e

$$\mathbf{M}_i = \begin{bmatrix} (\hat{m}_i t_{11}^i - t_{21}^i + \hat{l}_i t_{31}^i) & (\hat{m}_i t_{12}^i - t_{22}^i + \hat{l}_i t_{32}^i) & 0 & 0 \\ 0 & 0 & (\hat{m}_i t_{11}^i - t_{21}^i + \hat{l}_i t_{31}^i) & (\hat{m}_i t_{12}^i - t_{22}^i + \hat{l}_i t_{32}^i) \end{bmatrix} \quad (6.34)$$

e o vector de incógnitas a determinar

$$\mathbf{m} = [a \quad b \quad p \quad q]^T \quad (6.35).$$

Estabelecidos os valores para a equação (6.32) a solução para o vector \mathbf{m} pode ser obtida pela minimização do erro médio quadrático, utilizando a pseudo-inversa ou o filtro de Kalman.

As equações anteriores foram deduzidas para rectas não perpendiculares ao eixo-z. Para as aplicar a rectas deste tipo é necessário desenvolver expressões apropriadas. No caso da recta tridimensional ser representada pelas equações

$$\begin{cases} y = a x + p \\ z = b x + q \end{cases} \quad (6.36)$$

que não permitem a representação de rectas perpendiculares ao eixo-x, as equações (6.33) e (6.34) terão a forma

$$\mathbf{n}_i = \begin{bmatrix} -(\hat{m}_i t_{11}^i - t_{21}^i + \hat{l}_i t_{31}^i) \\ -(\hat{m}_i t_{14}^i - t_{24}^i + \hat{l}_i t_{34}^i) \end{bmatrix} \quad (6.37)$$

$$\mathbf{M}_i = \begin{bmatrix} (\hat{m}_i t_{12}^i - t_{22}^i + \hat{l}_i t_{32}^i) & (\hat{m}_i t_{13}^i - t_{23}^i + \hat{l}_i t_{33}^i) & 0 & 0 \\ 0 & 0 & (\hat{m}_i t_{12}^i - t_{22}^i + \hat{l}_i t_{32}^i) & (\hat{m}_i t_{13}^i - t_{23}^i + \hat{l}_i t_{33}^i) \end{bmatrix} \quad (6.38)$$

Para o caso da recta tridimensional ser representada pelas equações

$$\begin{cases} x = a y + p \\ z = b y + q \end{cases} \quad (6.39)$$

que não permitem a representação de rectas perpendiculares ao eixo-y, as matrizes \mathbf{n}_i e \mathbf{M}_i serão

$$\mathbf{n}_i = \begin{bmatrix} -(\hat{m}_i t_{12}^i - t_{22}^i + \hat{l}_i t_{32}^i) \\ -(\hat{m}_i t_{14}^i - t_{24}^i + \hat{l}_i t_{34}^i) \end{bmatrix} \quad (6.40)$$

$$\mathbf{M}_i = \begin{bmatrix} (\hat{m}_i t_{11}^i - t_{21}^i + \hat{l}_i t_{31}^i) & (\hat{m}_i t_{13}^i - t_{23}^i + \hat{l}_i t_{33}^i) & 0 & 0 \\ 0 & 0 & (\hat{m}_i t_{11}^i - t_{21}^i + \hat{l}_i t_{31}^i) & (\hat{m}_i t_{13}^i - t_{23}^i + \hat{l}_i t_{33}^i) \end{bmatrix} \quad (6.41)$$

A escolha das equações que representam a recta tridimensional pode ser feita com base no valor do determinante de \mathbf{M} . No caso de este ser nulo para um par de imagens então terá que ser escolhida outra representação alternativa. Outra forma prática de escolher a representação mais adequada é através da análise das projecções das rectas nas imagens.

6.3 Correspondência

Na reconstrução de uma cena tridimensional a partir de imagens estereo, o grande problema consiste na determinação da correspondência entre primitivas nas duas imagens. Partindo de uma primitiva numa imagem, o problema está em localizar a primitiva correspondente na outra imagem, assumindo que esta primitiva não está oculta. Este é um dos grandes problemas da visão estereo e que também se verifica noutros processos da visão computacional que envolvam a determinação da estrutura tridimensional.

No entanto é possível definir algumas restrições aos quais deverão obedecer as primitivas correspondentes. Pelos estudos biológicos sobre a visão humana, pela geometria da formação das imagens e pelas características físicas do mundo que nos rodeia, é possível estabelecer um conjunto de restrições para estabelecer correspondência [Marr 82], [Grimson 81], [Grimson 85]. Algumas destas restrições foram sendo propostas ao longo dos vários trabalhos sobre visão estereo.

6.3.1 Técnicas Anteriormente Propostas

Restrição da Recta Epipolar

A principal restrição do processo de estabelecimento de correspondência é de ordem geométrica e corresponde à restrição da recta epipolar. Esta restrição assegura que um ponto \mathbf{p}' na imagem esquerda \mathbf{I}' tem uma imagem correspondente na imagem direita \mathbf{I}'' num ponto da recta epipolar. Esta recta corresponde à intersecção do plano da imagem \mathbf{I}'' com o

plano formado pela recta que une os centros das duas imagens e pelo ponto \mathbf{p}' . Esta restrição reduz o problema da pesquisa bidimensional de um ponto correspondente a um problema unidimensional de pesquisa ao longo de uma recta.

No caso particular de os eixos ópticos serem paralelos, todas as rectas epipolares serão horizontais. Nesse caso o ponto $\mathbf{p}''=(u'', v'')$ da imagem \mathbf{I}'' , correspondente ao ponto $\mathbf{p}'=(u', v')$ de \mathbf{I}' , estará na linha $v''=v'$ da imagem \mathbf{I}'' . Este facto é de grande utilidade prática, sendo para isso utilizado o processo de rectificação de imagens abordado anteriormente. A rectificação das imagens transforma as imagens de modo a poder ser aplicado este princípio.

Como as primitivas que utilizamos são segmentos de recta, a recta epipolar na imagem \mathbf{I}'' é determinada para o ponto médio do segmento na imagem \mathbf{I}' , para a qual pretendemos obter segmentos correspondentes. A recta epipolar gerada a partir do ponto médio deverá intersectar o segmento de recta correspondente num ponto. Se a recta epipolar intersecta mais do que um segmento de recta todos eles serão candidatos para a correspondência. Uma forma de reduzir o número de candidatos consiste em testar se a intersecção ocorre para três pontos diferentes do segmento na imagem \mathbf{I}' . No nosso algoritmo o teste é realizado com três pontos do segmento de recta: ponto médio do segmento e dois pontos do segmento situados a 1/4 das suas extremidades.

A restrição geométrica é insuficiente para determinar, sem ambiguidade, uma só primitiva correspondente. Para isso é necessário definir restrições adicionais que reduzam a lista de candidatos a um único.

Restrições Locais

É possível definir restrições de carácter local com base nos valores das características das primitivas. Estas restrições poderão estar relacionadas com a orientação das primitivas na imagem, com o seu gradiente médio de intensidade, com a sua textura, com o seu comprimento ou com as suas características cromáticas. Tal como foi referido no capítulo 4, as arestas de imagens cromáticas apresentam, na maioria das vezes, uma deslocalização lateral devido à aberração cromática das lentes. Por isso restrições com base em características cromáticas deverão ser utilizadas cautelosamente quando as primitivas correspondem a detecções de arestas. Como as primitivas utilizadas são segmentos de recta, definiram-se as restrições locais só com base na orientação dos segmentos de recta e no seu comprimento.

Intervalo de variação do ângulo

Os ângulos de dois segmentos de recta correspondentes deverão ter valores absolutos semelhantes. Definindo um intervalo admissível para a diferença de ângulos dos dois segmentos de recta estabelecemos um método rápido de rejeitar emparelhamentos não

correspondentes. Esta restrição tem forte capacidade de rejeição de hipóteses de correspondência que estão incorrectas.

Intervalo de variação do comprimento

Um raciocínio idêntico ao feito para a orientação pode ser utilizado para o comprimento de um segmento. Na prática é definido um intervalo de variação $[l_{\min}, l_{\max}]$ para o comprimento de um segmento de recta correspondente. Como a estabilidade do valor para o comprimento é fortemente afectada pelo processo de segmentação, esta restrição serve mais para reforçar a probabilidade da correspondência, no caso de existirem vários candidatos.

Restrição da Correspondência Única

Como se pode ver pela figura 6.9, só com base na restrição epipolar e nas restrições locais, é possível estabelecer várias correspondências para dois pontos numa imagem. No entanto a uma primitiva na imagem I' deverá corresponder uma única primitiva na imagem I'' . Isso corresponde à restrição de correspondência única que define que uma primitiva numa imagem possui, no máximo, uma primitiva correspondente na outra imagem. Esta restrição é, só por si, insuficiente pois poderá ser violada se as primitivas forem pontos ou segmentos de recta. Se os pontos 3D estiverem alinhados com o centro óptico de uma das câmaras, tal como mostra a figura 6.10, então os pontos 3D alinhados com o centro óptico projectar-se-ão num único ponto.

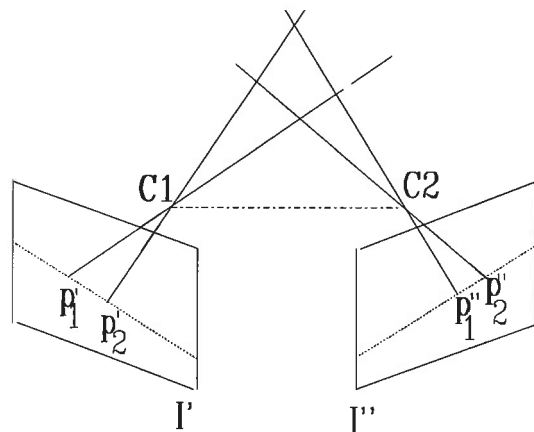


Figura 6.9- A figura apresenta o problema da correspondência mesmo que a restrição epipolar seja satisfeita

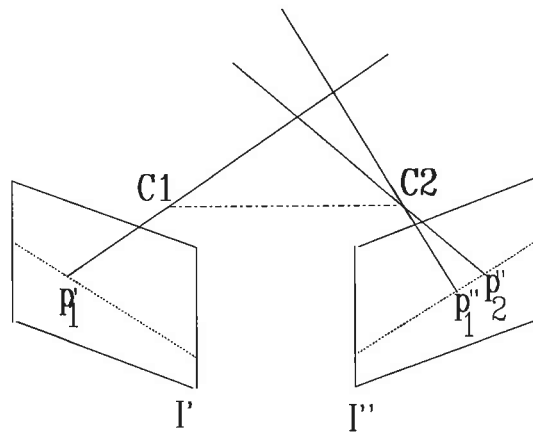


Figura 6.10- A restrição de correspondência única não permite a existência de pontos alinhados com o centro óptico

Para o caso de segmentos de recta a situação é similar. A figura seguinte mostra uma situação em que os segmentos S_1 e S_2 têm a mesma projecção na imagem I' e projecções diferentes na imagem I'' .

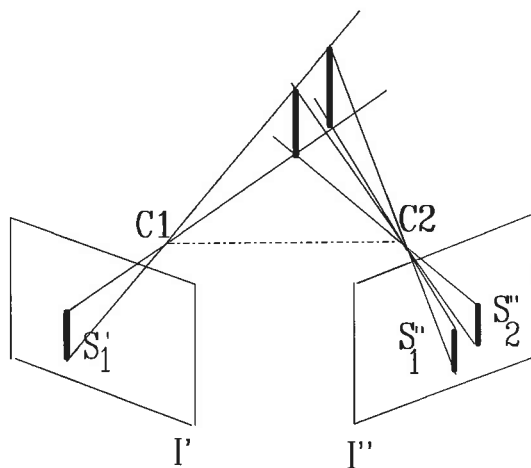


Figura 6.11 - Violação da restrição de correspondência única para segmentos de recta

Restrição de Ordem

Como normalmente os objectos são compostos de superfícies opacas é possível estabelecer uma restrição de ordem que se aplica a pontos ou primitivas nessas superfícies. A restrição de ordem impõe que pontos da mesma linha epipolar estejam segundo a mesma ordem para ambas as imagens.

Na figura 6.12 mostra-se um exemplo da restrição de ordem. Projectando os pontos M , P e Q da superfície nas imagens I' e I'' , a ordem da sua projecção mantêm-se para as duas imagens. Ou seja as projecções (m', p', q') têm projecções correspondentes em (m'', p'', q'') . Esta restrição é ainda insuficiente para resolver as ambiguidades de correspondência pois nem sempre as duas imagens estereo cobrem a mesma zona da cena. Para além disso, no

caso de objectos transparentes ou muito finos (fios por exemplo) esta restrição falha redondamente devendo, por isso, ser usada com precaução.

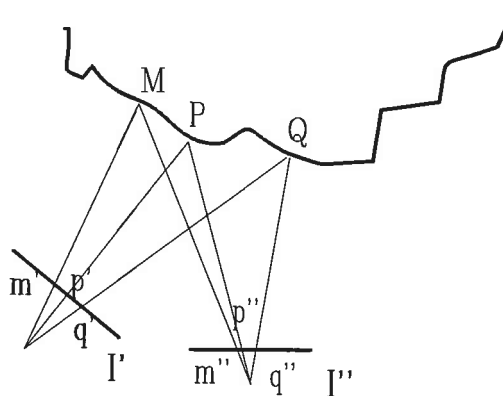


Figura 6.12 - Exemplificação da restrição de ordem

Restrição de Continuidade

A restrição de continuidade exprime o facto de que, numa maneira geral, a profundidade relativa entre pontos vizinhos dos objectos não variar de forma abrupta. Este facto conduz a uma restrição na profundidade e que pode ser definida com base na noção de *disparidade*. Se $P = (x, y, z)$ é um ponto no espaço então a diferença entre as coordenadas (u, v) das projecções p' e p'' de estão relacionadas com a posição tridimensional de P . Esta posição poderá ser determinada através de um único parâmetro chamado *disparidade*. A *disparidade* é um parâmetro que pode ser obtido a partir da posição relativa de pontos correspondentes que obedecem à restrição epipolar.

Disparidade máxima e mínima

A *disparidade* é um parâmetro que se relaciona com a profundidade a que se encontra o objecto e pode ser obtida a partir da posição das suas projecções nas imagens estereo. Para obter uma relação entre disparidade e profundidade vamos supor, que temos duas câmaras idênticas, tal como se ilustra na figura 6.4. As câmaras têm os eixos ópticos paralelos, com planos de imagem coincidentes e visualizando um ponto P no espaço. Esta situação é idêntica à que se obtêm quando é feita a rectificação das imagens. Neste caso, as rectas epipolares são horizontais e para dois pontos correspondentes, $p'=(u', v')$ e $p''=(u'', v'')$, o valor de $v'=v''$ será igual para ambos. Nesse caso a *disparidade* é definida pela diferença

$$\delta = u'' - u' \tag{6.42}.$$

É possível relacionar a disparidade com a distância a que um ponto \mathbf{P} se encontra da recta que une os dois centros ópticos das câmaras. Consideremos um ponto \mathbf{P} no espaço tridimensional que pode ser expresso em coordenadas tridimensionais dos referenciais das câmaras e colocados nos centros \mathbf{C}_1 e \mathbf{C}_2 . Nesse caso \mathbf{P} terá coordenadas $\mathbf{P}'=(x', y', z')$ para a câmara em \mathbf{C}_1 e $\mathbf{P}''=(x'', y'', z'')$ para a câmara em \mathbf{C}_2 , sendo $y'=y''=y$ e $z'=z''=z$ devido à geometria das câmaras. Pela mesma razão a relação entre as coordenadas x'' e x' depende só do afastamento entre os sistemas de coordenadas que se admite como sendo d . Neste caso, supondo que \mathbf{C}_2 se encontra na parte positiva do eixo- x do referencial colocado em \mathbf{C}_1 , a relação entre as coordenadas x'' e x' é dada por:

$$x''=x'+d$$

em que d é a distância entre os centros das câmaras. Projectando o ponto \mathbf{P} nos dois referenciais obtemos os valores para u' e u'' , que são dados por

$$u' = f \frac{x'}{z} \quad e \quad u'' = f \frac{x''}{z}$$

Utilizando a relação entre referenciais das câmaras e manipulando estas equações obtemos as seguintes expressões

$$x' = u' \frac{z}{f} \quad e \quad x'' = x' + d = u'' \frac{z}{f} \quad (6.43).$$

Subtraindo as duas equações obtemos uma expressão para a disparidade δ dada por

$$\delta = (u'' - u') = \frac{f d}{z} \quad (6.44)$$

e que relaciona a disparidade com a distância a que o ponto se encontra.

Pela expressão (6.44) conseguimos uma relação entre os valores das coordenadas da projecção de \mathbf{P} e a profundidade z a que \mathbf{P} se encontra das câmaras. Esta expressão mostra que a disparidade tende a ser nula para pontos muito longe da câmara ($z \rightarrow \infty$) e vai aumentando para pontos mais próximos das câmaras. Na prática os valores admissíveis para a disparidade são limitados pois as dimensões das cenas observadas são limitadas e os pontos observados estão circunscritos a um espaço tridimensional bem definido. Deste modo o valor da disparidade deverá ter valores dentro de um intervalo $[\delta_{\min}, \delta_{\max}]$. A definição deste intervalo é equivalente à definição de um intervalo de distâncias possíveis para o ponto \mathbf{P} .

Definir um intervalo para a disparidade equivale a uma nova restrição para a correspondência entre primitivas.

Gradiente de disparidade

Grande parte dos objectos existentes no universo que nos rodeia são opacos e apresentam superfícies sem grandes descontinuidades. Com base nesse pressuposto é possível definir uma restrição de continuidade para dois pontos correspondentes. Essa restrição pode ser traduzida por uma medida do *gradiente de disparidade*. O *gradiente de disparidade* dá uma noção da variação da disparidade numa vizinhança de um ponto e, quando as superfícies observadas são opacas e não apresentam grandes descontinuidades na profundidade, as disparidades de pontos vizinhos na imagem são muito semelhantes. Sendo assim, o gradiente de disparidade assemelha-se a uma medida de continuidade das superfícies. Esta restrição é especialmente útil na eliminação de hipóteses múltiplas de correspondência e foi explorada por Pollard e os seus colegas em algoritmos de visão estereo [Pollard 85].

Para definirmos uma medida de continuidade com base no gradiente de disparidade consideremos duas imagens estereo I' e I'' de uma cena. Se (p', p'') e (q', q'') são dois pares de pontos correspondentes, então o gradiente de disparidade é dado pela expressão

$$GD = \frac{\text{diferença na disparidade}}{\text{separação "ciclópica "}} \tag{6.45}$$

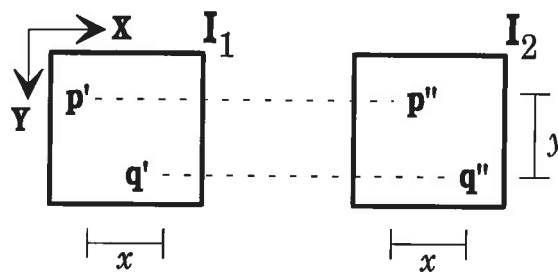


Figura 6.13 - Cálculo do gradiente de disparidade

O noção de separação "ciclópica" ("cyclopean separation") foi introduzida por [Burt 80] e corresponde à distância entre os pontos médios dos pares de pontos correspondentes. Visto que os pontos intermédios são dados por $(p'+p'')/2$ e $(q'+q'')/2$ e os vectores disparidade associados são $(p''-p')$ e $(q''-q')$ então

$$\frac{GD}{2} = \frac{\|(p''-q'')-(p'-q')\|}{\|(p''-q'')+(p'-q')\|} = \frac{(x''-x')}{\sqrt{(x''+x')^2 + 4y^2}} \tag{6.46}$$

em que $\|\cdot\|$ representa a norma $\left\| \begin{bmatrix} x \\ y \end{bmatrix} \right\|_2 = \sqrt{x^2 + y^2}$.

Criando um limite máximo para o gradiente da disparidade definimos uma restrição de continuidade das superfícies. Limitando o GD para valores dentro do intervalo $[0..2]$ é assegurada essa continuidade dos objectos [Pollard 89]. Se $GD \approx 0$ a restrição é muito forte permitindo só superfícies fronto-paralelas, mas com $GD \approx 2$ as discontinuidades suaves são possíveis. Os valores de GD podem tomar valores intermédios sendo os valores entre 0.5 e 1 que conduzem a melhor desambiguidade.

6.3.2 Novas Restrições pela Exploração dos Graus de Liberdade do Sistema de Visão

Utilização de um Processo Multiocular.

Se for utilizado um processo multiocular, isto é, um processo em que sejam utilizados mais de um par de imagens é possível estabelecer restrições geométricas suplementares para a geração das hipóteses. Consideremos então dois pares de imagens estereo tiradas em tempos diferentes (t) e ($t+\delta t$). A trajectória entre esses dois pontos é conhecida e é representada pela matriz de transformação T .

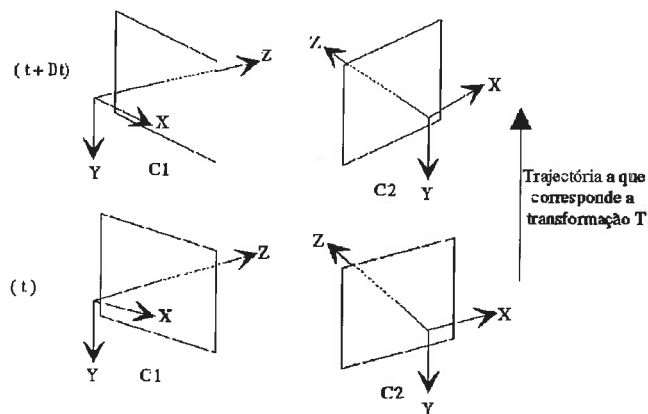


Figura 6.14 - Processo multiocular.

Neste trabalho são consideradas duas formas de geração de hipóteses de emparelhamento. A primeira consiste em gerar hipóteses de correspondência com base nas imagens estereo obtidas no tempo $(t+\delta t)$ e as imagens estereo obtidas no tempo (t) .

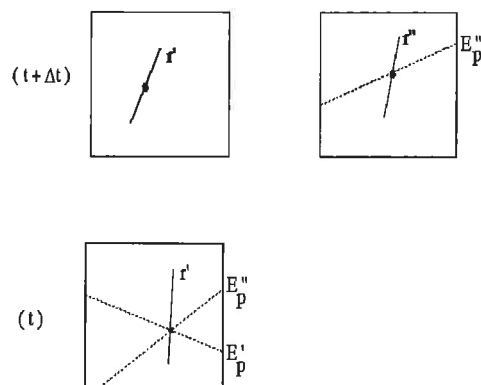


Figura 6.15 - Geração das rectas epipolares no processo de estereovisão multiocular. A recta r' dá origem à epipolar $E''_p(t+\delta t)$ para a imagem direita no tempo $(t+\delta t)$ e à epipolar $E''_p(t)$ na imagem esquerda no instante (t) . Se a epipolar $E''_p(t+\delta t)$ no tempo $(t+\delta t)$ intersecta a recta r'' da imagem direita então r'' dá origem à epipolar $E'_p(t)$ no tempo (t) . Se r'' é a recta correspondente a r' então na intersecção das duas epipolares $E''_p(t)$ com $E'_p(t)$ no tempo (t) deverá estar uma recta.

Inicialmente e a partir do segmento de recta $r'_{t+\delta t}$, na imagem esquerda no tempo $(t+\delta t)$, é determinada a epipolar $E''_p(t+\delta t)$ na imagem direita no tempo $(t+\delta t)$ e a epipolar $E''_p(t)$ na imagem esquerda no tempo (t) . Notar que para cada segmento de recta $r''_{t+\delta t}$, intersectado pela epipolar na imagem direita corresponde uma epipolar $E'_p(t)$ na imagem esquerda obtida no tempo (t) - ver figura 6.15. As duas epipolares na imagem esquerda obtida no tempo (t) devem-se intersectar. Se no ponto de intersecção das duas epipolares, ou em redor, existir um segmento r'_t , que obedece às restrições locais então o tripleto $(r'_{t+\delta t}, r''_{t+\delta t}, r'_t)$ é uma hipótese de correspondência. Este procedimento, aplicado a todos os segmentos da imagem esquerda obtida no tempo $(t+\delta t)$, gera uma lista de hipóteses de correspondência para a imagem esquerda. Aplicando este procedimento de igual forma para a imagem direita obtida no tempo $(t+\delta t)$, geramos uma lista de hipóteses de correspondência para a imagem direita. Os elementos de ambas as listas são hipóteses de correspondência para os segmentos. Esta restrição é geométrica e estabelece uma *condição necessária* para a correspondência entre dois segmentos *mas não suficiente* para assegurar uma correspondência única.

Cruzamento das Listas de Hipóteses

A partir das listas para as imagens esquerda e direita estas poderão ser cruzadas de modo a gerar uma lista final em que só existem correspondências que se verificam simultaneamente nas duas imagens. Este mecanismo é equivalente a uma filtragem que elimina hipóteses com fraca possibilidade de serem correctas e pode ser aplicado na

estereovisão binocular ou multiocular. A prática demonstra que este mecanismo é muito útil para eliminar grande parte das hipóteses múltiplas.

Restrição local global

As restrições locais assentam basicamente no teste da diferença entre comprimentos e na diferença entre ângulos entre pares de segmentos candidatos à correspondência. Estes dois testes podem ser realizados em separado mas com o problema inerente da adaptação dos limites máximos das diferenças. Uma forma mais conveniente é criar um mecanismo que os envolva conjuntamente. Uma solução consiste em normalizar a diferença relativamente a um limite máximo e exprimir a diferença em termos percentuais. No nosso caso a diferença de comprimento e a diferença do ângulo são normalizados relativamente a valores máximos pré-estabelecidos e expressos em valor percentual. Após esta operação os dois valores são somados não podendo o resultado ultrapassar um determinado limite (tipicamente 50%). Esta forma de comparação soluciona também o problema da execução sequencial dos dois testes.

Restrição geométrica qualitativa

Na maioria dos casos a configuração geométrica do sistema de visão exibe vergência, sendo diferente da configuração "fronto paralela" descrita acima. Nesse caso é possível estabelecer uma restrição geométrica *qualitativa* no momento da geração de hipóteses no processo multiocular. Para isso consideremos a figura 6.16 na qual se assume que os dois sistemas de coordenados de cada câmara têm os seus eixos-z no mesmo plano que passa pelos centros de projecção das duas câmaras. O ponto de intersecção correspondente à intersecção dos dois eixos-z designa-se por ponto de fixação **Pf**.

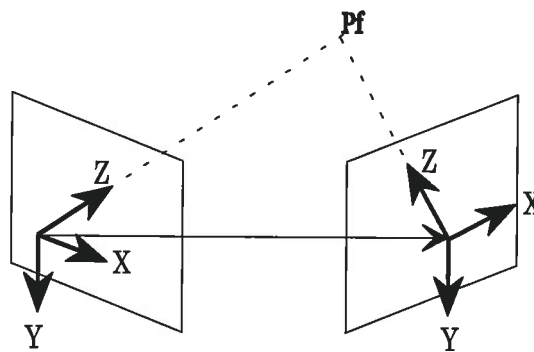


Figura 6.16 -Geometria do sistema de visão estereo exibindo vergência fixa. O ponto **Pf** designa-se por ponto de fixação e corresponde ao ponto de intersecção dos eixos-z das duas câmaras e para a qual a disparidade é zero. A projecção desse ponto na imagem esquerda e direita faz-se precisamente no centro das imagens. O eixos-z encontram-se no mesmo plano que passa pelo centro das imagens.

O ponto de fixação **Pf** tem as suas projecções no centro da imagem podendo as suas coordenadas tridimensionais serem determinadas com base neste pressuposto. Vamos supor que utilizamos um processo multiocular com imagens estereo obtidas em posições diferentes no tempo (t) e $(t+\delta t)$. Nesse caso as coordenadas do ponto **Pf** poderão ser determinadas para a posição $(t+\delta t)$ com base no pressuposto de que as suas projecções **pf'** e **pf''** coincidem com o centro da imagem. No entanto se o ponto **Pf** for reprojectado nas imagens da posição (t) as suas projecções na imagem esquerda (**pf'**) e direita (**pf''**) serão em local diferente - ver figura 6.17.

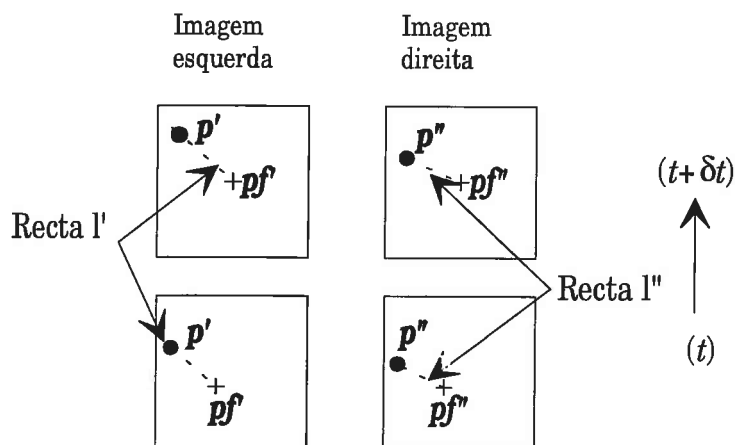


Figura 6.17 - O ponto de fixação **Pf** tem projecções no centro da imagem para as imagens obtidas no tempo $(t+\delta t)$. Esse mesmo ponto terá projecções diferentes nas imagens obtidas no tempo (t) . É possível definir uma recta l' (ou l'') passando por uma dessas projecções **pf'** (ou **pf''**) e um outro qualquer ponto na imagem. Essas rectas terão inclinações idênticas para as imagens da esquerda (ou direita).

Consideremos, para facilidade de interpretação, só as projecções nas imagens da esquerda. Nessas imagens vamos considerar um ponto **p'** pertencente ao segmento de recta para o qual procuramos uma correspondência. Para ambas as imagens é possível unir com uma recta l' , a projecção **pf'** e ponto **p'**. Em ambos os casos obteremos uma recta para a posição (t) e para a posição $(t+\delta t)$. Se a transformação que relaciona as duas posições não envolver translações ou rotações de grande amplitude, então a inclinação da recta l' para as posições em (t) e $(t+\delta t)$ deverá ser idêntica. O mesmo princípio poderá também ser aplicado às imagens da direita e permite restringir o numero de hipóteses de correspondência. A esta restrição aplicada às imagens da esquerda e da direita designamos por *restrição geométrica qualitativa*. A restrição é considerada qualitativa por reflectir, de modo qualitativo, a posição da primitiva relativamente a um ponto fixo no espaço (ponto de fixação).

Restrição baseada na focagem

No capítulo 5 foi utilizada a focagem como um método de determinar um mapa tridimensional da cena. Este mapa tem a vantagem de ser um mapa tridimensional com tamanho igual à imagem e a sua precisão depende da profundidade de campo. Os valores de profundidade descritos nesse mapa podem ser relativos à posição de focagem ou absolutos no caso das lentes serem calibradas. Em ambos os casos essa profundidade permite definir o intervalo admissível para a profundidade em qualquer ponto da imagem.

Para a utilização desta forma de cooperação entre visão estereo e focagem são determinados mapas de profundidade para as imagens da esquerda e da direita. Com base nestes mapas é possível estabelecer uma restrição para o valor da profundidade de primitivas correspondentes. *Deste modo um segmento de recta na imagem esquerda deverá ter um segmento de recta correspondente na imagem direita com uma profundidade relativa ou absoluta idêntica.* Esta é uma *condição necessária* para que os segmentos de recta sejam correspondentes.

Na prática esta restrição permite realizar um mecanismo eficaz de eliminar hipóteses de correspondência concorrentes.

6.3.3 Validação de Hipóteses

Grande parte das restrições acima descritas são úteis na fase de geração das hipóteses e na qual eliminam hipóteses correspondentes. Todavia em visão estereo o processo de geração de hipóteses conduz muitas vezes a múltiplas hipóteses de correspondência. Estas situações surgem porque o conjunto de restrições impostas são *condições necessárias* para a correspondência e não conseguem eliminar, na totalidade, as diferentes possibilidades das primitivas corresponderem entre si.

Validação com base na vizinhança

Uma maneira de obter a verdadeira correspondência quando existem hipóteses múltiplas é através do cálculo de um valor de suporte para cada uma das hipóteses [Murray 90], [Sherman 90]. O mesmo princípio foi ensaiado nos algoritmos aqui propostos.

Se r_i' for a primitiva da imagem esquerda, então o suporte inicial para a sua correspondência com r_j'' da imagem direita poderá ser feita com base no gradiente de disparidade, através da expressão

$$\text{Suporte}_{ij} = e^{-10GD} \quad (6.47)$$

em que GD é o gradiente de disparidade dado pela expressão (6.45).

Este suporte inicial é posteriormente actualizado através dos vizinhos da primitiva r_i' existentes no interior de um círculo de raio W em torno centro de massa de r_i' . A actualização é feita através da soma do valor inicial de suporte com o valor do suporte para cada um dos vizinhos. Para além do suporte existe um acumulador A que armazena o número de vizinhos que suportam a correspondência.

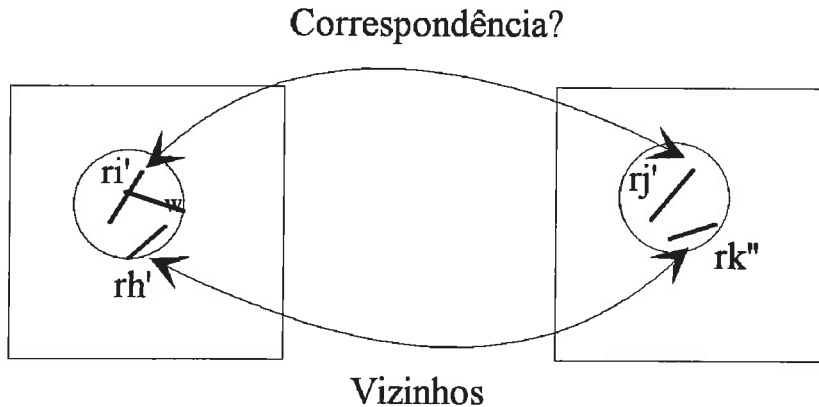


Figura 6.18- A validação das hipóteses baseia-se no cálculo do suporte com base no apoio dos vizinhos da primitiva, em torno de um círculo de raio W .

Este procedimento é repetido para todos os pares de correspondências que têm multiplicidade e após ordenamento é escolhida a correspondência que apresenta maior valor na relação entre o valor de suporte e o número armazenado no acumulador A .

Validação com base na correlação e na restrição de ordem

Uma alternativa para a eliminação de hipóteses múltiplas é a utilização de um operador de medida de semelhança. Este operador mede a semelhança entre a região em torno da primitiva com as regiões de cada uma das primitivas candidatas para correspondência. A primitiva escolhida como a mais semelhante é posteriormente testada para verificar se cumpre a restrição de ordem.

Correlação

Uma das medidas mais populares para o teste da semelhança entre dois sinais é a correlação. A correlação dá um resultado que indica o grau de semelhança dos sinais que estão a ser correlacionados. Se um sinal padrão for correlacionado com outros sinais então o sinal para o qual o correlacionador atinge o valor máximo indica que tem grande semelhança com o sinal padrão. Este mesmo princípio é também aplicado no nosso algoritmo para determinar qual é a melhor das hipóteses de correspondência. A correlação faz-se entre

regiões (janela) dos segmentos que pertencem à hipótese de correspondência. A região (janela) é centrada em torno de um ponto do segmento, o qual é escolhido de acordo com a fase de execução do algoritmo. Assim, para o segmento padrão o ponto será igual ao ponto escolhido para gerar epipolar, para os segmentos candidatos à correspondência será o ponto de intersecção da epipolar com o segmento de recta, ou o ponto correspondente à intersecção das duas epipolares no caso do processo multiocular. A correlação utilizada foi a correlação normalizada $C(x, y)$ cuja expressão é

$$C(x, y) = \frac{\sigma_{12}^2}{\sqrt{\sigma_1^2 \sigma_2^2}} \quad (6.48)$$

em que σ_1^2 e σ_2^2 representam a variância dos valores de intensidade dos "pixel" da janela na imagem I_1 e I_2 respectivamente. Essa variância é dada por

$$\sigma_l^2 = \frac{1}{(2m+1)^2} \sum_{i=-m}^m \sum_{j=-m}^m [I_l(x+i, y+j) - \mu_l]^2 \quad \text{com } l=1,2 \quad (6.49)$$

em que μ_l representa a média dos valores de intensidade na janela $(2m+1) \times (2m+1)$. e σ_{12}^2 é dada por

$$\sigma_{12}^2 = \frac{1}{(2m+1)^2} \sum_{i=-m}^m \sum_{j=-m}^m [I_1(x+i, y+j) - \mu_1][I_2(x+i, y+j) - \mu_2] \quad (6.50).$$

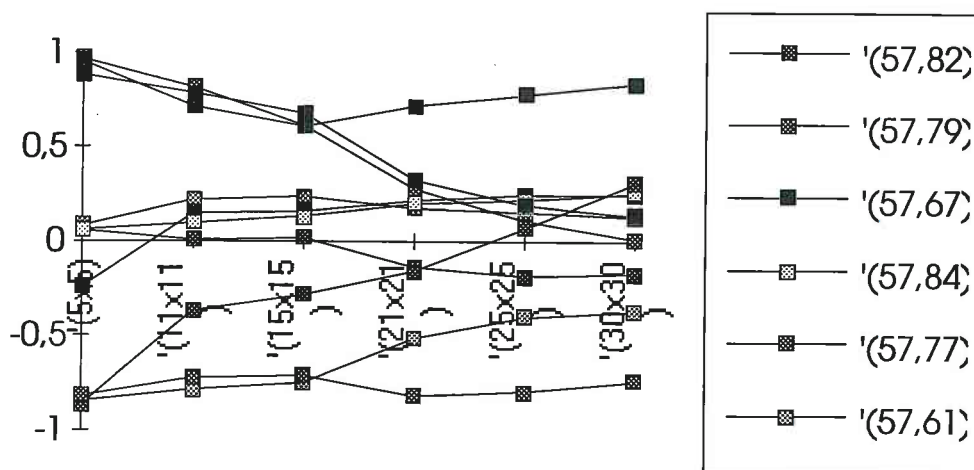


Figura 6.19 - Curvas do valor da correlação para janelas de vários tamanhos para uma lista de candidatos à correspondência. Neste exemplo pretende-se determinar a correspondência para o segmento número 57.

Quando esta medida de semelhança é utilizada para distinguir entre várias hipóteses de correspondência aquela que apresenta um máximo será escolhida como melhor hipótese.

Esta medida é fortemente dependente do tamanho da janela que é utilizada na correlação. Como se pode verificar pelo gráfico da figura 6.19 os valores da correlação variam bastante com o tamanho da janela.

6.4 Detalhes

6.4.1 Efeitos Devido ao Movimento das Câmaras

Como o sistema é dinâmico, tem interesse estudar como se comportam as projecções das primitivas com o movimento do sistema de aquisição de imagem. O movimento é normalmente referenciado a um sistema de coordenadas comum à das câmaras. As primitivas para a qual o efeito são pontos, rectas e planos.

O caso de pontos

Se a transformação que representa o movimento for conhecida é possível prever o efeito da projecção de um ponto, com coordenadas tridimensionais conhecidas, na imagem. Se a transformação entre os referências das diferentes posições das câmaras for conhecida é possível saber a imagem de um ponto nessas diferentes posições.

Se a matriz ${}^B\mathbf{T}_A$ representa a relação que descreve a passagem de coordenadas de um referencial $\{A\}$ para o referencial $\{B\}$ então o ponto ${}^A\mathbf{x}$ pode ser expresso no referencial $\{B\}$ por

$${}^B\mathbf{x} = {}^B\mathbf{T}_A {}^A\mathbf{x} = \begin{bmatrix} {}^B\mathbf{Rot}_{A(3x3)} & {}^B\mathbf{Trans}_{A(3x1)} \\ \mathbf{0}_{(3x3)} & 1 \end{bmatrix} {}^A\mathbf{x} \quad (6.51)$$

Se a matriz ${}^B\mathbf{T}_A$ relaciona duas posições $\{A\}$ e $\{B\}$ de uma câmara, a imagem do ponto \mathbf{x} com a câmara colocada em $\{B\}$ terá uma projecção na imagem dada por

$$\begin{bmatrix} su \\ sv \\ s \end{bmatrix} = \mathbf{CALIB} \begin{bmatrix} {}^Bx \\ {}^By \\ {}^Bz \\ 1 \end{bmatrix} = \mathbf{CALIB} {}^B\mathbf{x} \quad (6.52)$$

sendo a matriz \mathbf{CALIB} a matriz de calibração da câmara em $\{B\}$. O ponto ${}^B\mathbf{x}$ pode ser relacionado com o ponto na posição $\{A\}$ através da expressão (6.51).

O caso de rectas

Consideremos uma recta tridimensional passando por um ponto \mathbf{x}_p , com direcção \mathbf{l} . Esta recta pode ser expressa pela forma paramétrica

$$\mathbf{x} = \mathbf{x}_p + \alpha \mathbf{l}$$

sendo α um parâmetro que pertence aos números reais. Se a recta sofre um movimento para um local cuja relação com o anterior é dada através de uma transformação \mathbf{T} , composta por uma rotação **Rot** e translação **Trans**, então qualquer ponto da recta na nova posição, será obtido por

$$\mathbf{x}' = (\mathbf{Rot} \mathbf{x}_p + \mathbf{Trans}) + \alpha \mathbf{Rot} \mathbf{l} \quad (6.53)$$

(notar que \mathbf{l} é uma direcção e como tal fica só sujeita à rotação).

Se for utilizada a representação em coordenadas homogéneas, então a recta na nova posição é representada por

$$\begin{bmatrix} x' \\ y' \\ z' \\ s' \end{bmatrix} = \begin{bmatrix} \mathbf{Rot}_{(3 \times 3)} & \mathbf{Trans}_{(3 \times 1)} \\ \mathbf{0}_{(3 \times 3)} & 1 \end{bmatrix} \left(\begin{bmatrix} x_p \\ y_p \\ z_p \\ 1 \end{bmatrix} + \alpha \begin{bmatrix} l_x \\ l_y \\ l_z \\ 0 \end{bmatrix} \right) \quad (6.54)$$

O caso de planos

Se para a representação de um plano for utilizada um vector $\mathbf{n} = n_x \mathbf{i} + n_y \mathbf{j} + n_z \mathbf{k}$, normal ao plano e a distância d do plano à origem do sistema de coordenadas então o plano pode ser representado pelo vector

$$\Pi = [n_x \quad n_y \quad n_z \quad d] \quad (6.55).$$

Qualquer ponto $\mathbf{P}(x,y,z)$ pertencente ao plano Π obedecerá à equação

$$\Pi \mathbf{P} = [n_x \quad n_y \quad n_z \quad d] \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = 0 \quad (6.56).$$

em que o produto escalar de \mathbf{n} por \mathbf{P} corresponde à projecção da distância segundo \mathbf{n} , a que se encontra o ponto \mathbf{P} pertencente ao plano.

Se ${}^B\mathbf{T}_A$ for a transformação entre duas posições $\{A\}$ e $\{B\}$ para câmara, então as transformações nos parâmetros do plano serão dados por

$$\Theta = \Pi ({}^B\mathbf{T}_A)^{-1} \quad (6.57)$$

Sendo Θ a representação do plano Π no referencial $\{B\}$. Tomando um ponto ${}^B\mathbf{P}$ do plano Θ então a sua representação no referencial $\{A\}$ será ${}^A\mathbf{P} = ({}^B\mathbf{T}_A)^{-1} {}^B\mathbf{P}$. Como o ponto pertence ao plano então deverá verificar a equação

$$\Pi {}^A\mathbf{P} = \Theta {}^B\mathbf{P} \quad (6.58)$$

que permite demonstrar a relação (6.57).

Restrições ao movimento

Poderão existir situações nas quais o sistema de visão estereo é sujeito a movimento para obter diferentes perspectivas da mesma cena. Exemplos dessas situações são a visão estereo dinâmica e visão estereo multiocular. Nessas situações esperamos que o movimento ao qual as câmaras são sujeitas tenha influência na projecção dos pontos tridimensionais na imagem. Ou seja, é necessário que os valores das coordenadas (u, v) da projecção do ponto ${}^A\mathbf{x}$ variem para valores (u', v') . Considerando o ruído na imagem nulo, a variação tem que ser tal que permita existir, no mínimo, uma diferença de uma unidade de pixel entre coordenadas u e v e que por outro lado não cause o desaparecimento da projecção do ponto na área captada pelo sensor. Consideremos um caso em que executamos um movimento ${}^B\mathbf{T}_A$ que muda a câmara da posição $\{A\}$ para a posição $\{B\}$. Neste caso o movimento deverá estar sujeito às seguintes restrições

$$\begin{cases} 1 \leq |u - u'| \leq N^\circ_{\text{max_pixel_eixo_u}} \\ 1 \leq |v - v'| \leq N^\circ_{\text{max_pixel_eixo_v}} \end{cases}$$

para os valores das coordenadas da imagem do ponto \mathbf{x} . Tendo em atenção as expressões (6.3) e (6.4) é possível estabelecer restrições para o movimento em termos de coordenadas projectivas. As restrições para a qual o movimento deverá estar sujeito são dadas por

$$\begin{cases} 1 \leq \left| \frac{\mathbf{l}_1^T \mathbf{A} \mathbf{X}}{\mathbf{l}_3^T \mathbf{A} \mathbf{X}} - \frac{\mathbf{l}_1^T \mathbf{B} \mathbf{T}_A \mathbf{A} \mathbf{X}}{\mathbf{l}_3^T \mathbf{B} \mathbf{T}_A \mathbf{A} \mathbf{X}} \right| \leq N^\circ \text{ _max_pixel_eixo_u} \\ 1 \leq \left| \frac{\mathbf{l}_2^T \mathbf{A} \mathbf{X}}{\mathbf{l}_3^T \mathbf{A} \mathbf{X}} - \frac{\mathbf{l}_2^T \mathbf{B} \mathbf{T}_A \mathbf{A} \mathbf{X}}{\mathbf{l}_3^T \mathbf{B} \mathbf{T}_A \mathbf{A} \mathbf{X}} \right| \leq N^\circ \text{ _max_pixel_eixo_v} \end{cases} \quad (6.59)$$

em que \mathbf{l}_i e \mathbf{l}_i' é o vector coluna correspondente à linha i das matrizes de calibração **CALIB** e **CALIB'** para o referencial $\{A\}$ e $\{B\}$ respectivamente.

6.4.2 Quantificação do erro na correspondência

Se os valores das coordenadas (u, v) de uma primitiva estiverem sujeitos a um erro, então esse erro irá repercutir-se na estimação do valor da profundidade. Este erro pode ser determinado, utilizando o valor da *disparidade* para as primitivas correspondentes.

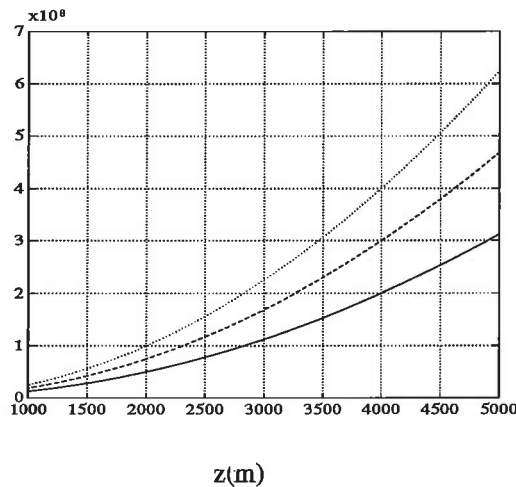


Figura 6.20- O diagrama apresenta os valores da relação entre o erro em profundidade e o erro na disparidade, $|\Delta z|/|\Delta \delta|$, em função da distância $z(m)$. Os valores foram calculados para a distância focal de 16 mm e para $d=200\text{mm}$ (pontilhado), $d=300\text{mm}$ (tracejado), $d=400\text{mm}$ (contínuo) sendo d a distância entre câmaras. Para a mesma distância z o erro é maior para menores distâncias entre as câmaras.

Se o valor de uma das coordenadas u' ou u'' é determinado com erro então o valor da disparidade δ virá também com erro. Em valor desse erro, em valor absoluto $|d\delta|$, pode ser relacionado com o valor da profundidade através da primeira derivada da expressão (6.44). A expressão em módulo é

$$|\Delta z| = \frac{z^2}{f d} |\Delta \delta| \quad (6.60)$$

Como podemos verificar, o erro $|\Delta\delta|$ é amplificado por um valor proporcional à distância z . Quanto maior a distância maior será a influência do erro da disparidade¹.

Como se pode verificar, o erro é maior para distâncias maiores do objecto. No entanto *este erro pode ser controlado pela distância que separa as duas câmaras ou pelo comprimento focal da lente*. Daqui podemos concluir que uma das maneiras de controlar o valor máximo do erro é através do controlo do comprimento focal f , quando se utiliza uma lente "zoom". Estes factos são importantes num sistema visual que permita o controlo destes parâmetros do sistema, como no caso de um sistema de visão activa como o descrito capítulo 9.

¹ O valor $d\delta$ e as expressões anteriores pressupõem medidas em metros, embora se possa realizar a conversão da disparidade de milímetros para pixels.

6.5 Resultados Experimentais

As experiências realizadas basearam-se inicialmente em imagens sintéticas geradas com câmaras realizando um movimento de translação. Posteriormente foram utilizadas imagens reais obtidas no local experimental descrito no capítulo 3.

6.5.1 Algoritmos

Neste ponto são descritos os diferentes algoritmos utilizados no sistema de visão estereo. Inicialmente é proposto um algoritmo de estereovisão biocular ao qual é posteriormente é integrado a restrição multiocular. O algoritmo de base determina um conjunto de hipóteses de correspondência baseando-se em restrições geométricas locais. Posteriormente realiza a intersecção dos dois conjuntos de hipóteses da câmara esquerda e para a câmara direita. Na fase seguinte é realizada a validação das hipóteses utilizando algumas das restrições globais descritas atrás. De seguida é determinada a estrutura geométrica tridimensional da primitiva.

Algoritmo de visão estereo biocular

Definições

lesq e ldir - Listas de primitivas detectadas nas imagens estereo

{W} - referencial da cena

Mp - Mapa de profundidade relativa obtido pela focagem

Algoritmo Base_para_visão_estereo

Lê os parâmetros relativos às câmaras;

Procedimento *Filtra_elementos_da_lista*;

Se Utiliza a restrição de focagem **então**

Determina mapa de profundidade relativa *Mp* **fim se**;

Para cada elemento *lesq* **faz**

Determina recta epipolar com base no segmento esquerdo;

Para cada elemento da imagem direita que intersecta a linha epipolar **faz**

Procedimento *Verifica as restrições locais*;

Se *Verifica as restrições locais* **então**

Gera um elemento na lista de hipóteses de correspondência **fim se**;

fim faz; **fim faz**;

Para cada elemento *ldir* **faz**

Determina recta epipolar com base no segmento direito;

Para cada elemento da imagem esquerda que intersecta a linha epipolar **faz**

Procedimento *Verifica as restrições locais*;

Se *Verifica as restrições locais* **então**

Gera um elemento na lista de hipóteses de correspondência **fim se**;

fim faz; **fim faz**;

Procedimento *Cruzamento_de_listas*;

Procedimento *Validação_das_hipóteses*;

Procedimento *Determina_estrutura_tridimensional* ;

fim Algoritmo *Base_para_visão_estereo*

Algoritmo estereo multiocular

Definições

$lesq_{(t)}$ e $ldir_{(t)}$ - Listas de primitivas detectadas nas imagens estereo no tempo (t) .
 $lesq_{(t+\delta t)}$ e $ldir_{(t+\delta t)}$ - Listas de primitivas detectadas nas imagens estereo no tempo $(t+\delta t)$.
 $Mp(t)$ - Mapa de profundidade relativa obtido pela focagem no tempo $Mp(t)$.
 $Mp(t+\delta t)$ - Mapa de profundidade relativa obtido pela focagem $Mp(t+\delta t)$.

Algoritmo Visão_estereo_multiocular

Lê os parâmetros relativos às câmaras;
Procedimento Filtra_elementos_da_lista;
Se Utiliza a restrição de focagem **então**
 Determina mapa de profundidade relativa $Mp(t+\delta t)$ para o tempo $(t+\delta t)$ e
 Determina mapa de profundidade relativa $Mp(t)$ para o tempo (t) **fim se**;
Para cada elemento $lesq_{(t+\delta t)}$ **faz**
 Determina recta epipolar com base no segmento esquerdo;
 Para cada elemento da imagem direita que intersecta a linha epipolar **faz**
 Procedimento Verifica as restrições locais;
 Se Verifica as restrições locais **então**
 Gera um elemento na lista de hipóteses de correspondência **fim se**;
 fim faz;
 fim faz;
 Para cada elemento $ldir_{(t+\delta t)}$ **faz**
 Determina recta epipolar com base no segmento direito;
 Para cada elemento da imagem esquerda que intersecta a linha epipolar **faz**
 Procedimento Verifica as restrições locais;
 Se Verifica as restrições locais **então**
 gera um elemento na lista de hipóteses de correspondência **fim se**;
 fim faz;
 fim faz;
 Procedimento Cruzamento_de_listas;
 Para Cada elemento da lista de hipóteses de correspondência **faz**
 Calcula rectas epipolares na imagem esquerda e direita para o tempo (t) com
 base na lista de hipóteses de correspondência para o tempo $(t+\delta t)$;
 Procura em torno da intersecção das duas epipolares a existência de um segmento;
 Se Verifica as restrições locais para o segmento no tempo (t) e $(t+\delta t)$ **então**
 Se Verifica a restrição qualitativa **então**
 Gera um elemento na lista de hipóteses de correspondência **fim se**;
 fim se;
 fim faz;
 Procedimento Validação_das_hipóteses;
 Procedimento Determina_estrutura_tridimensional ;

Fim Algoritmo Visão_estereo_multiocular

Procedimento Filtra_elementos_da_lista

Para cada elemento da lista $lesq$ e $ldir$ **faz**
 Elimina segmentos demasiado curtos;
 fim faz

Fim Procedimento

Procedimento Matrizes_de_rectificação*Para* cada câmara *faz*

Calcula coordenadas do centro da lente;

*fim faz*Determina a rotação em relação ao referencial $\{W\}$;

Determina as novas matrizes de orientação dos referenciais das câmaras;

Para cada câmara *faz*

Determina a nova matriz de perspectiva com base na matriz de

mudança de perspectiva e na matriz de posição;

Determina a nova matriz de perspectiva em relação à imagem adquirida;

*fimfaz***Fim Procedimento****Procedimento** Verifica_as_restrições_locais*Se* Utiliza a restrição de focagem*então* Verifica a restrição de focagem;*Senão* Verifica a restrição de disparidade vertical e horizontal *fim se*;*Se* Verifica a restrição *então***Procedimento** Verifica_restrição_local_global;*Senão* rejeita a hipótese *fim se*;**Fim Procedimento** Verifica_as_restrições_locais**Procedimento** Verifica_restrição_local_global

Determina as diferenças de ângulo e comprimento entre segmentos;

Normaliza as diferenças relativamente a valores máximos pré-estabelecidos;

Se Soma das diferenças normalizadas é superior a valor pré-estabelecido *então*Rejeita a hipótese de correspondência *fim se*;**Fim Procedimento** Verifica_as_restrições_locais**Procedimento** Cruzamento_de_listas*Para* cada elemento da lista de emparelhamento direita *faz*

Verifica se existe elemento correspondente na lista de emparelhamento esquerda;

Se existe *então* insere elemento na lista de hipóteses de correspondências *fim se*;*fim faz***Fim Procedimento** Verifica_restrição_local_global**Procedimento** Validação_das_hipóteses (* Versão com a correlação *)*Para* cada elemento da lista de hipóteses de correspondências *faz*

Calcula correlação para cada elemento da lista de hipóteses de correspondências;

Escolhe a melhor hipótese com base no critério de correlação

para a opção de "pixels" ou do vector das características;

Testa a ordem para essa hipótese;

Se não violou a ordem *então*Calcula o suporte da hipótese S ;Escolhe a hipótese com maior suporte S e insere-a na lista

de correspondências finais;

Elimina a sua ocorrência na lista de hipóteses de correspondências;

*senão*Retira o elemento da lista de hipóteses de correspondência; *fimse**fim faz***Fim Procedimento** Validação_das_hipóteses

```

Procedimento Validação_das_hipóteses (* Versão com o gradiente de disparidade *)
  Para cada elemento da lista de hipóteses de correspondências faz
    Verifica a vizinhança na imagem esquerda e na direita
      entre os dois elementos seleccionados;
    Se forem vizinhos então
      Testa a ordem;
      Se não violou a ordem então
        Calcula o suporte da hipótese S;
        Escolhe a hipótese com maior suporte S e insere-a na
          lista de correspondências finais;
        Elimina a sua ocorrência na lista de hipóteses
          de correspondências;
      senão
        Retira o elemento da lista de hipóteses
          de correspondência; fimse
    fimse
  fim faz
Fim Procedimento Validação_das_hipóteses

Procedimento Determina_estrutura_tridimensional
  Para cada elemento da lista de correspondências faz
    Determina recta tridimensional;
  fim faz
Fim Procedimento

```

Pormenores práticos

Na programação destes algoritmos é necessário alertar para alguns pormenores que são de importância vital para o seu correcto funcionamento. Assim, no algoritmo de base foi introduzida uma pequena variante na restrição de ângulo que permite uma melhoria significativa do desempenho do algoritmo. Esta variante tem por objectivo resolver os casos em que a recta epipolar tem um ângulo muito próximo do ângulo do segmento. Normalmente os algoritmos de estereovisão rejeitam esses segmentos como hipóteses de emparelhamento pois geram grande multiplicidade de intersecções do segmento com a epipolar. No algoritmo, e para estes casos, é só considerado o ponto correspondente do ponto médio da intersecção da epipolar com o segmento de recta. Deste modo, e com este refinamento, consegue-se recuperar várias hipóteses de emparelhamento.

Outro pormenor importante é a forma como é gerada a recta epipolar. No algoritmo proposto por Ayache a recta epipolar é gerada com base no ponto médio do segmento. No entanto a utilização deste princípio conduz a geração de muitas hipóteses de correspondência falsas pois a epipolar poderá intersectar os segmentos em vários pontos ao longo do seu comprimento. Uma forma mais restritiva é exigir que a restrição epipolar se verifique para mais pontos do segmento. Nos algoritmos propostos acima a restrição epipolar dever-se-á cumprir para três pontos do segmento e correspondentes a $1/3, 1/2, 2/3$ do seu comprimento.

6.5.2 Resultados com Imagens Reais

Para testar os algoritmos utilizaram-se as seguintes de imagens:

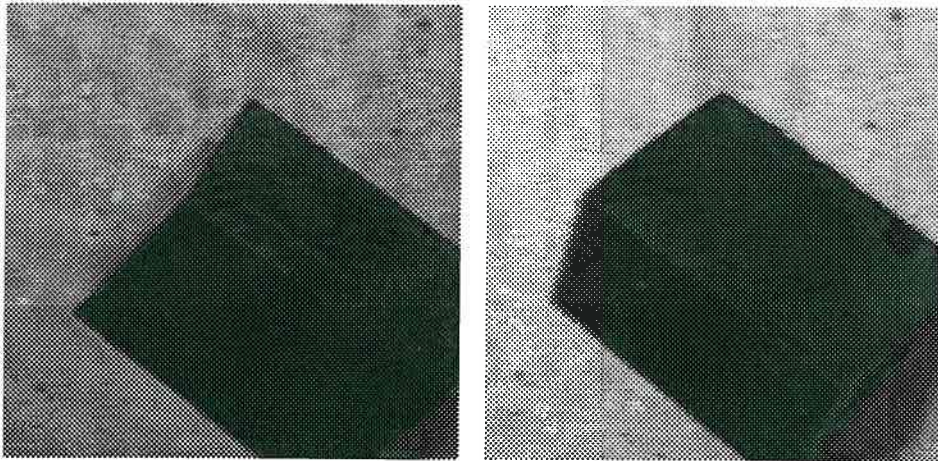


Figura 6.39- Imagens da cena Prisma

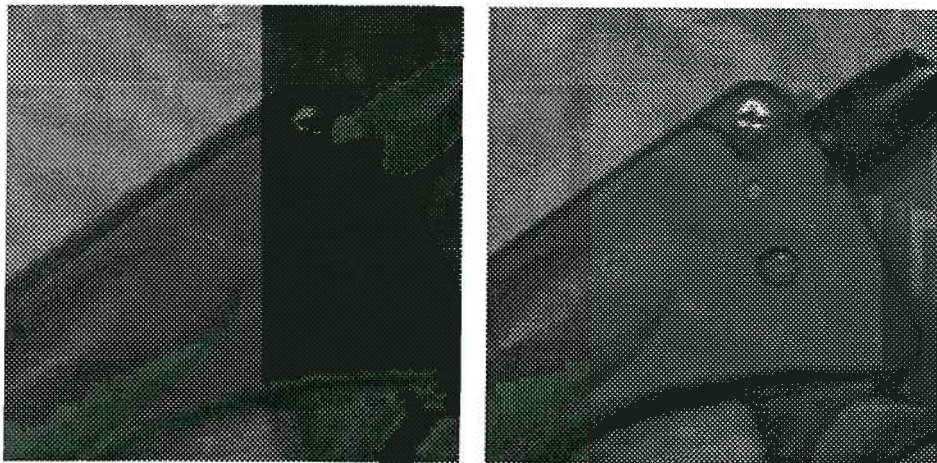


Figura 6.40- Imagens da cena Pistola

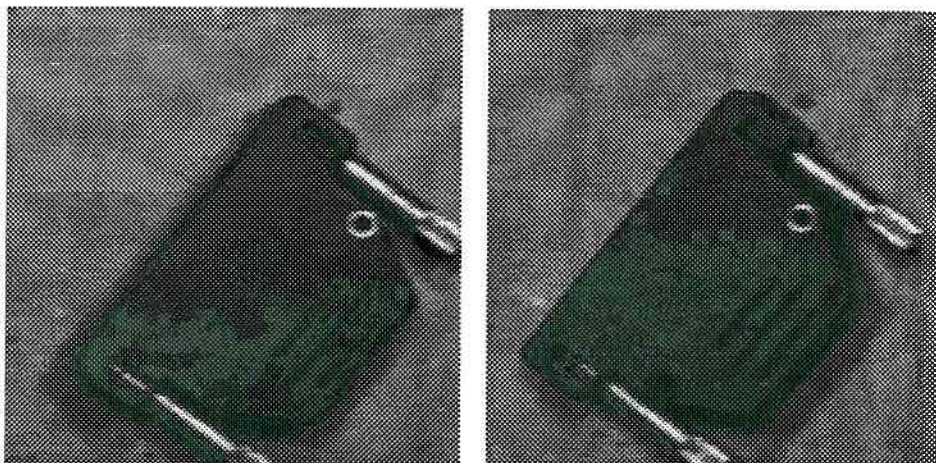


Figura 6.41- Imagens da cena RS232

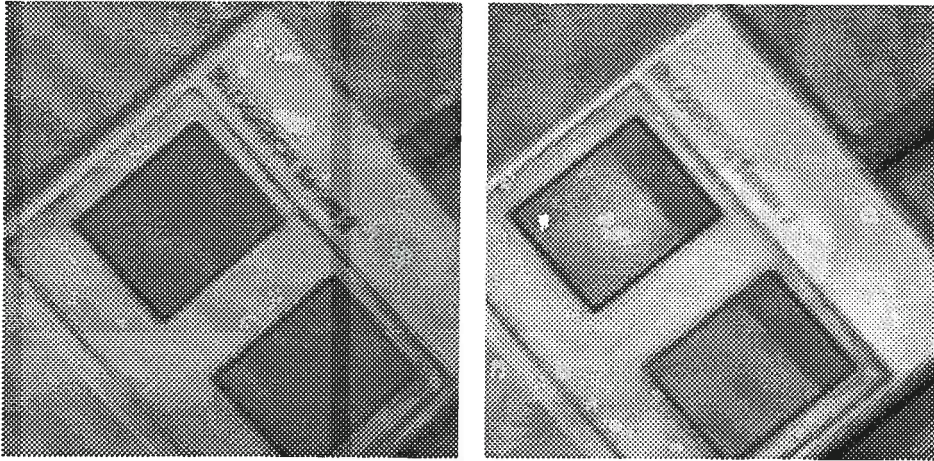


Figura 6.42- Imagens da cena Rato

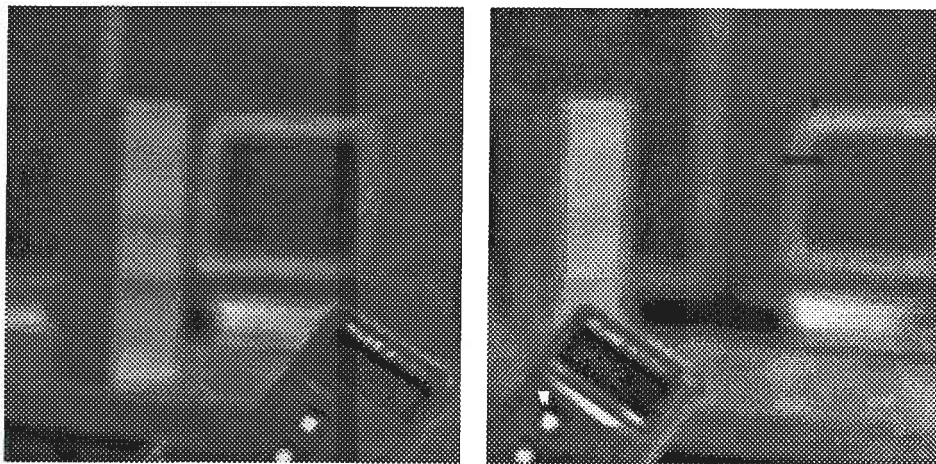


Figura 6.43- Imagens da cena World 0 e correspondente à primeira imagem da sequência de imagens de focagem adquiridas para esta cena.

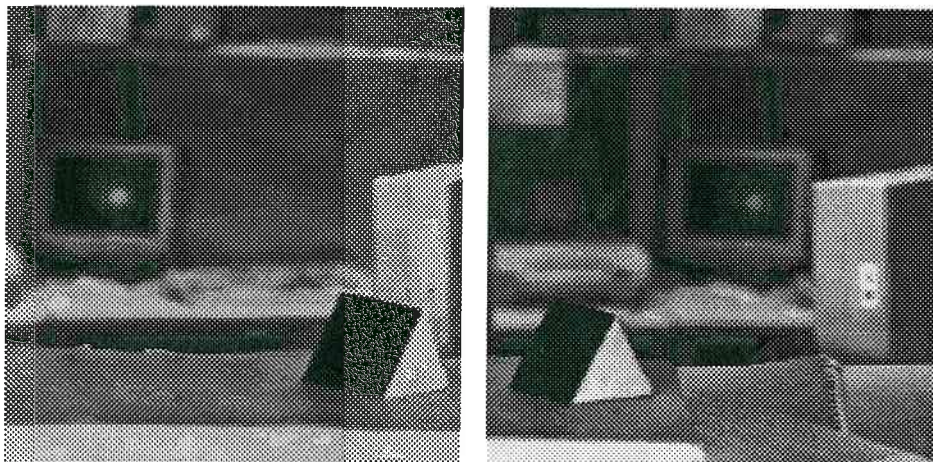


Figura 6.44- Imagens da cena World 1 e correspondente à décima imagem da sequência de imagens de focagem adquiridas para esta cena.

Resultados da aplicação do algoritmo de estereovisão

Cena Rato

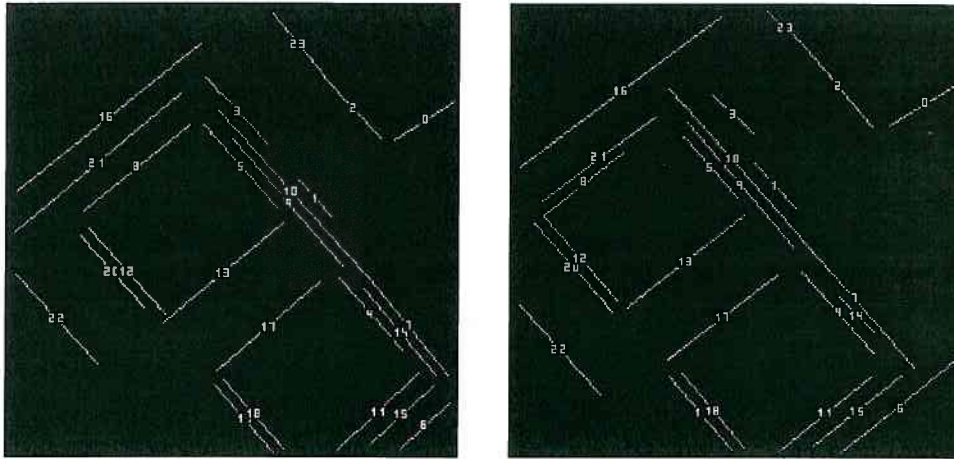


Figura 6.45- Correspondências obtidas para a cena Rato para estereovisão multi-ocular

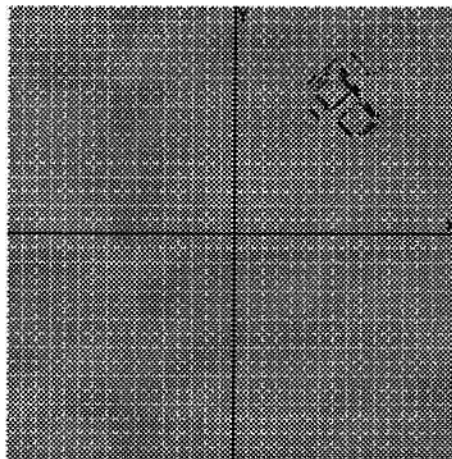


Figura 6.46- Reconstrução tridimensional - Mapa XY para a cena Rato (cada quadrícula representa 10cm x 10cm).

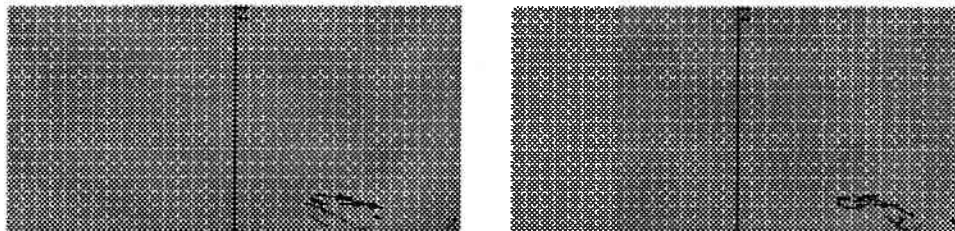


Figura 6.47- Reconstrução tridimensional - Mapas ZX e ZY para a cena Rato (cada quadrícula representa 10cm x 10cm).

Cena Laboratório

Resultados para a cena Laboratório com afectação directa das distâncias à origem do sistema de coordenadas invariante do sistema de visão.

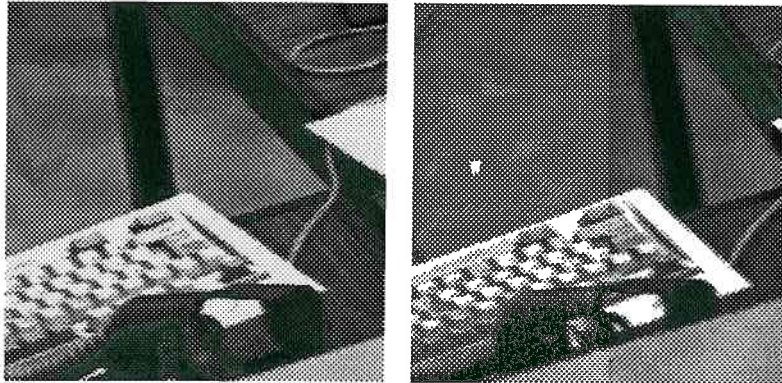


Figura 6.48- Imagens estereo da cena Laboratório.

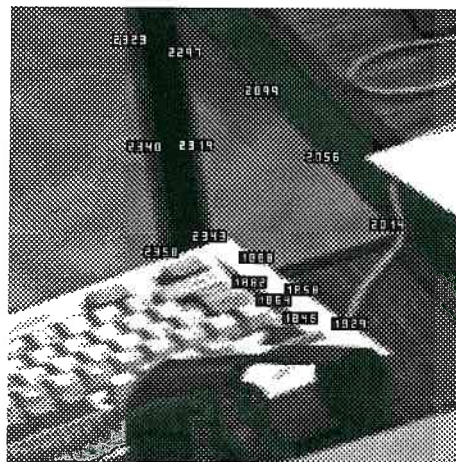


Figura 6.49- Reconstrução tridimensional da cena anterior mas com afectação directa das distâncias à imagem e utilizando estereovisão multi-ocular.

Cena World 0

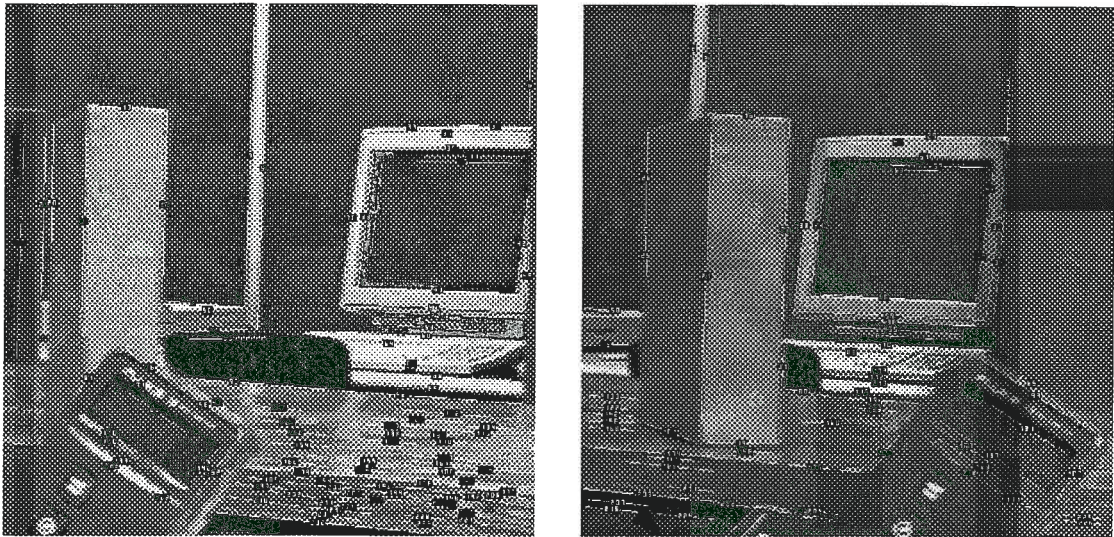


Figura 6.50- Imagem estereo de segmentos para utilização com a restrição de focagem. Ambas as imagens foram obtidas após recuperação das zonas melhor focadas

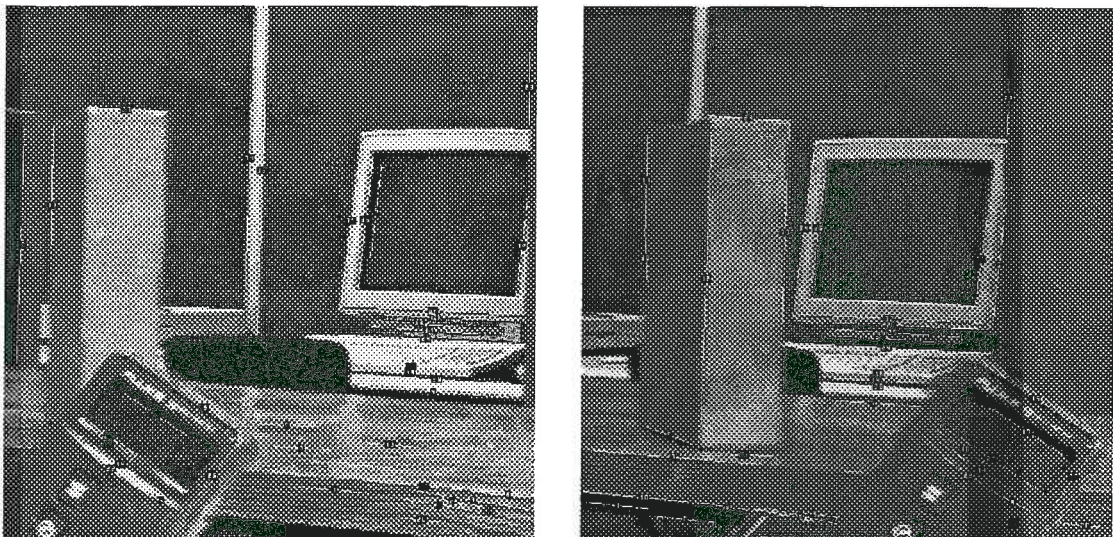


Figura 6.51- Imagens estereo dos segmentos correspondentes para o caso da cena anterior e com restrição de focagem

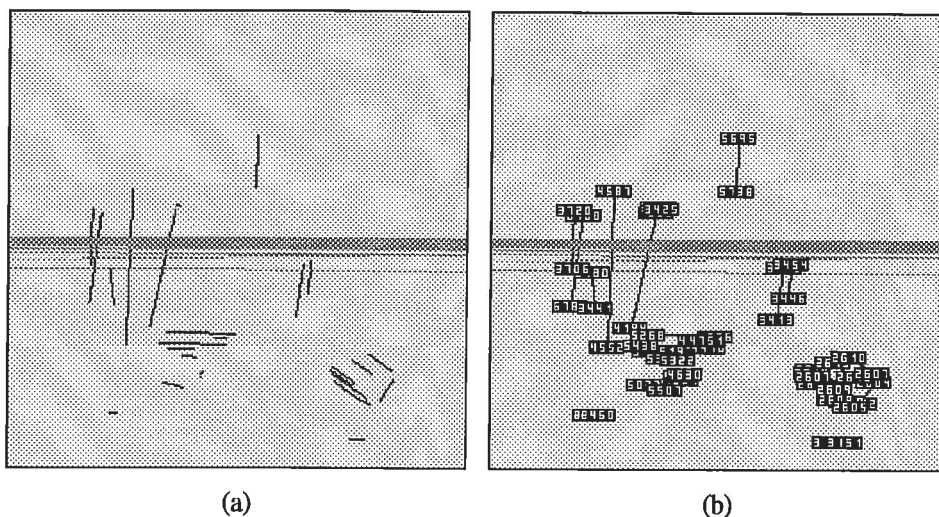


Figura 6.52- Reconstrução tridimensional. (a) mapa reconstruído com a grelha designando o plano $z=0$, (b) o mesmo mapa com as distâncias dos extremos dos segmentos.

6.5.3 Resultados Comparativos

Um vasto conjunto de restrições foi descrito anteriormente. Estas restrições podem ser utilizadas nos algoritmos de estereovisão, mas não implica que todas se utilizem simultaneamente. Aliás, a sua utilização conjunta poderá provocar maior tempo de computação sem acréscimo de qualidade dos resultados. Deste modo, torna-se conveniente um estudo comparativo dos resultados dos algoritmos com diferentes restrições. Para se realizar esse estudo comparativo definiram-se alguns índices comparativos percentuais baseados nos resultados em diferentes fases do algoritmo. Os índices são baseados em medidas realizadas sobre as imagens e sobre os resultados obtidos pela aplicação dos algoritmos. As medidas que foram definidas foram:

- número de hipóteses de correspondência geradas,
- número de hipóteses de correspondência que estão correctas após a fase de geração de hipóteses,
- número de correspondências correctas e falsas,
- número de primitivas geométricas (rectas) reconstruídas,
- número de primitivas que foram detectadas e que tinham possibilidade de serem reconstruídas.

Os índices foram:

$$Sucesso_relativo = \frac{n^\circ_correspondencias_correctas}{n^\circ_hipóteses_correctas_após_geração_de_hipóteses} \quad (I1)$$

$$\text{Erro_relativo} = \frac{n^\circ_correspondencias_falsas}{n^\circ_hipóteses_correctas_apos_geração_hipóteses} \quad (I2)$$

$$\text{Desempenho} = \frac{n^\circ_primitivas_reconstruídas}{n^\circ_primitivas_que_poderiam_ter_reconstrução} \quad (I3)$$

$$\text{Taxa_de_falha} = \frac{n^\circ_primitivas_reconstruídas_que_estão_erradas}{n^\circ_primitivas_que_poderiam_ter_reconstrução} \quad (I4)$$

$$\text{Geração} = \frac{n^\circ_hipótese_correctas_geradas}{n^\circ_primitivas_que_poderiam_ter_reconstrução} \quad (I5)$$

$$\text{Sucesso_real} = \frac{n^\circ_correspondências_correctas}{n^\circ_total_de_hipóteses_geradas} \quad (I6)$$

$$\text{Taxa_de_falha_real} = \frac{n^\circ_correspondências_erradas}{n^\circ_total_de_hipóteses_geradas} \quad (I7)$$

É conveniente notar que embora algumas das medidas pareçam ser complementares, como no caso de (I6) e (I7), na verdade não o são. De facto, por vezes existem hipóteses que, na fase da sua validação, são excluídas. Ou seja, as hipóteses excluídas nunca serão consideradas no universo dos resultados finais.

Do conjunto de resultados, obtidos pelas experiências realizadas, foram resumidos os mais relevantes nos quadros seguintes. Os quadros apresentam os resultados percentuais relativos aos índices de *Sucesso Relativo*, *Desempenho*, *Sucesso Real* e *Geração* por serem considerados os mais relevantes. *Todas as experiências realizadas utilizam a restrição de cruzamento de listas e a restrição epipolar.*

CENAS	<i>Sucesso Relativo</i>	<i>Desempenho</i>	<i>Sucesso real</i>	<i>Geração</i>
PRISMA	100	75	100	75
RATO	100	60	30,77	60
RS232	66,6	44,4	12,9	66,7
PISTOLA	100	50	50	50

Tabela 6.1 - Resultados com as restrições *locais, multiocular, qualitativa, correlação e ordem.*

CENAS	<i>Sucesso Relativo</i>	<i>Desempenho</i>	<i>Sucesso real</i>	<i>Geração</i>
PRISMA	0	0	0	100
RATO	50	40	11,6	80
RS232	14,29	11,11	2,94	77,77
PISTOLA	66,67	40	33,33	60

Tabela 6.2 - Resultados com as restrições *locais, gradiente de disparidade e ordem.*

CENAS	<i>Sucesso Relativo</i>	<i>Desempenho</i>	<i>Sucesso real</i>	<i>Geração</i>
PRISMA	100	75	100	75
RATO	92,3	60	30,77	65
RS232	66,66	44,44	12,9	66,67
PISTOLA	100	50	50	50

Tabela 6.3 - Resultados com as restrições *locais, multiocular, qualitativa e correlação*.

CENAS	<i>Sucesso Relativo</i>	<i>Desempenho</i>	<i>Sucesso real</i>	<i>Geração</i>
PRISMA	0	0	0	60
RATO	100	75	20,55	75
RS232	30	23,08	6,52	76,92
PISTOLA	57,14	44,44	30,77	77,78

Tabela 6.4 - Resultados com as restrições *locais e gradiente de disparidade*.

CENAS	<i>Sucesso Relativo</i>	<i>Desempenho</i>	<i>Sucesso real</i>	<i>Geração</i>
PRISMA	100	75	100	75
RATO	100	60	30,77	60
RS232	66,66	44,44	12,9	66,67
PISTOLA	100	50	50	50

Tabela 6.5 - Resultados com as restrições *multiocular, qualitativa, correlação e ordem*.

CENAS	<i>Sucesso Relativo</i>	<i>Desempenho</i>	<i>Sucesso real</i>	<i>Geração</i>
PRISMA	0	0	0	100
RATO	44,4	40	5,93	90
RS232	16,67	11,11	2,86	66,67
PISTOLA	66,67	40	33,33	60

Tabela 6.6 - Resultados com as restrições *gradiente disparidade e ordem*.

CENAS	<i>Sucesso Relativo</i>	<i>Desempenho</i>	<i>Sucesso real</i>	<i>Geração</i>
PRISMA	100	75	100	75
RATO	100	60	30,77	60
RS232	50	33,33	9,68	66,67
PISTOLA	100	50	50	50

Tabela 6.7 - Resultados com as restrições *locais, multiocular, qualitativa e correlação*.

CENAS	<i>Sucesso Relativo</i>	<i>Desempenho</i>	<i>Sucesso real</i>	<i>Geração</i>
PRISMA	100	75	100	75
RATO	92,3	60	30,77	60
RS232	66,67	44,44	12,9	66,67
PISTOLA	80	40	40	50

Tabela 6.8 - Resultados com as restrições *multi-ocular e correlação*.

CENAS	<i>Sucesso Relativo</i>	<i>Desempenho</i>	<i>Sucesso real</i>	<i>Geração</i>
PRISMA	100	75	100	75
RATO	100	60	30,77	60
RS232	66,66	44,44	12,9	66,67
PISTOLA	100	50	50	50

Tabela 6.9 - Resultados com as restrições *multiocular, qualitativa e correlação*.

A partir das tabelas e utilizando o índice de *sucesso real*, verificamos que a associação de restrições que gera bons resultados é a *associação da restrição multi-ocular com a restrição de correlação e a restrição qualitativa*. Mesmo para o índice *desempenho* ou *sucesso real* as conclusões são as mesmas. A influência da restrição de ordem é muito relativa e pelas experiências realizadas tem maior influência na diminuição do número de correspondências erradas, como seria de esperar.

Nas tabelas seguintes apresentam-se resultados relativos aos testes com a restrição de *focagem*.

CENAS	<i>Sucesso Relativo</i>	<i>Desempenho</i>	<i>Sucesso real</i>	<i>Geração</i>
World 0	100	22,58	30,43	22,58
World 1	100	18,37	11,84	18,37

Tabela 6.10 - Resultados com as restrições *focagem, multiocular, qualitativa e correlação*.

CENAS	<i>Sucesso Relativo</i>	<i>Desempenho</i>	<i>Sucesso real</i>	<i>Geração</i>
World 0	61,9	41,94	7,88	67,74
World 1	29,41	10,21	2,57	34,7

Tabela 6.11 - Resultados com as restrições *locais, focagem e gradiente de disparidade*.

CENAS	<i>Sucesso Relativo</i>	<i>Desempenho</i>	<i>Sucesso real</i>	<i>Geração</i>
World 0	100	22,58	30,44	22,58
World 1	100	20,41	13,15	20,41

Tabela 6.12 - Resultados com as restrições *locais, focagem, multiocular, correlação e qualitativa*.

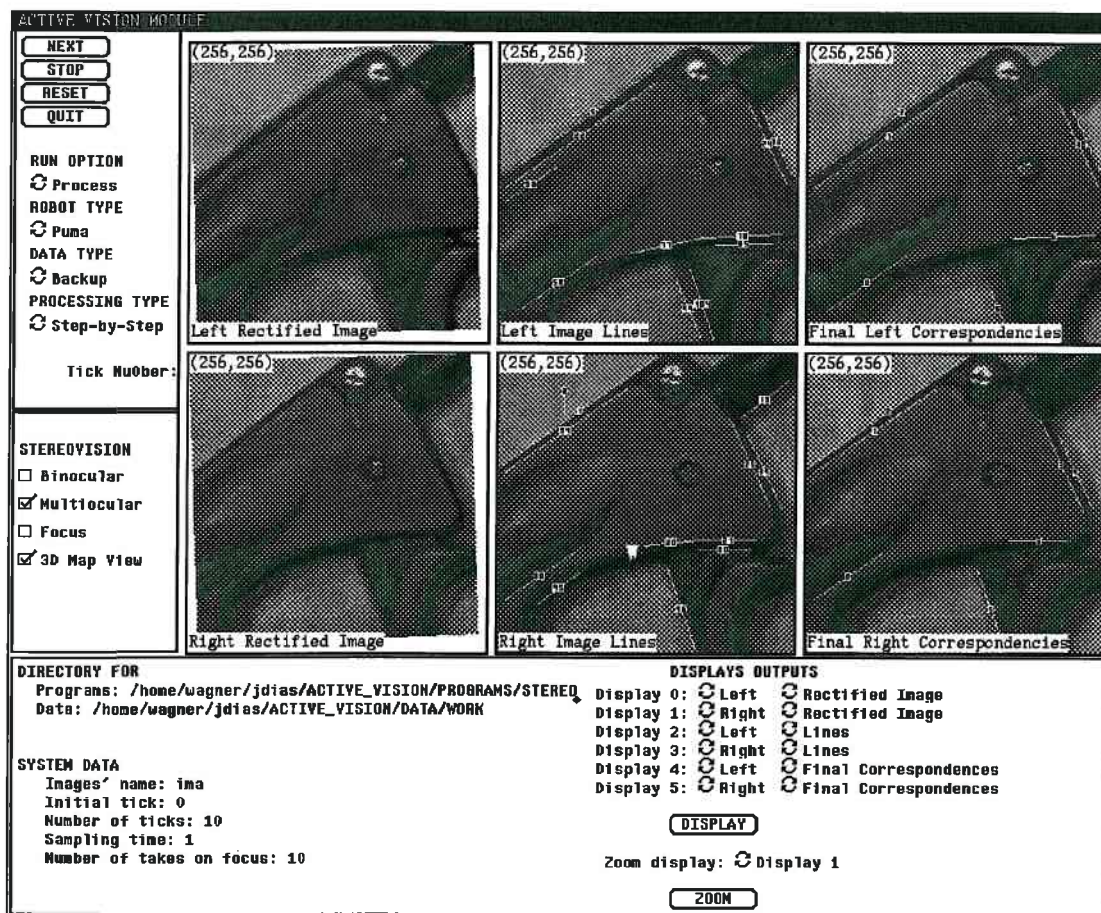
CENAS	<i>Sucesso Relativo</i>	<i>Desempenho</i>	<i>Sucesso real</i>	<i>Geração</i>
World 0	83,33	16,13	21,74	19,35
World 1	10	2,04	1,3157	20,41

Tabela 6.13 - Resultados com as restrições *locais, focagem, multiocular e gradiente de disparidade*.

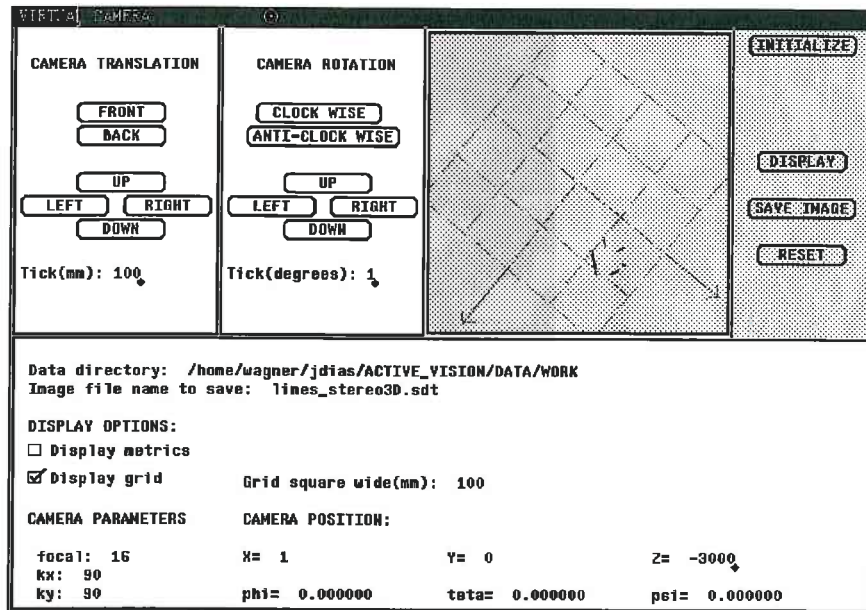
A partir das tabela e utilizando o índice de *sucesso real*, verificamos que a associação de restrições que gera bons resultados é a *associação da restrição multiocular com a restrição de correlação, focagem e a restrição qualitativa*. Mesmo para o índice *desempenho* ou *sucesso real* as conclusões são as mesmas. A utilização da restrição de *focagem* não gera, só por si, resultados suficientemente satisfatórios. É necessário associar esta restrição com outra.

6.5.4 Programas Desenvolvidos

Para a realização das experiências foram desenvolvidos programas que facilitam a execução e a análise de resultados dos vários algoritmos. As figuras em baixo ilustram as possibilidades gráficas dos programas.



(a)



(b)

Figura 6.53- As figuras são imagens do programa desenvolvido para a realização das experiências com estereovisão. Devido à sua capacidade gráfica o utilizador pode alterar facilmente os parâmetros de entrada do programa (a) e visualizar os resultados do processamento (b).



Figura 6.54 - Os programas permite a utilização de imagens de dimensão variável podendo ser utilizada a opção de zoom. A figura ilustra a execução do programa na opção de focagem e com imagens (512x512).

CAPÍTULO 7

Mapas Volumétricos - Experiências com base na focagem

7.1 Introdução

Os resultados obtidos por algoritmos que determinam as características geométricas do mundo tridimensional a partir das imagens poderão ter maior utilidade se forem englobados numa representação geométrica que represente o local em torno do sistema de visão e do "robot". Num "robot", a existência de um sistema de representação deste tipo é importante para alcançar maior grau de autonomia na movimentação e acção sobre o meio onde se encontra. Esta representação do universo é normalmente designada *modelo do mundo* ("World model") e representa fundamentalmente informação com características estáticas. Esta representação engloba toda a informação disponível sobre os locais onde o "robot" se movimenta. No entanto para um "robot" autónomo realizar movimentos sem risco de colisão com os objectos existentes no seu local de trabalho é necessário ter informação acerca dos locais onde se encontra. Esta informação poderá ser armazenada em modelos ou representações que contêm informação acerca dos objectos existentes dentro do espaço de trabalho, incluindo as propriedades dos objectos e as relações geométricas entre si. Para além disso poderão conter informação adicional acerca de acontecimentos que podem ocorrer e informação importante para as tarefas que o "robot" deve desempenhar. Da informação global que o "robot" recebe, só uma pequena parte será necessária para realizar um movimento local. Essa parte da informação integra-se no designado *modelo local*. Este modelo contém a informação sobre as características físicas do local e depende das características específicas da tarefa a realizar. Essa pequena porção do universo, designada por *modelo local*, poderá ter uma informação mais detalhada sobre essa zona do universo. Na realidade os modelos ideais deveriam permitir decomposições hierárquicas da

informação, com várias escalas de resolução e de abstracção. As várias escalas de resolução permitiriam fornecer uma representação detalhada duma zona assim como, dar uma representação global do universo. Na prática a construção destes sistemas envolve recursos computacionais elevados. Diferentes modelos de representação têm sido propostos para estas representações. Chien, Srivastava e Potmesil propõem modelos com base nas representações tridimensionais de objectos e a partir de silhuetas de diferentes lados [Chien 86] [Srivastava 87] [Potmesil 87]. Elfes constrói um mapa tridimensional do ambiente utilizando leituras dos sensores e uma aproximação probabilística para combinar a informação dos sensores em diferentes posições [Elfes 87].

Na realização de um sistema autónomo com sensores com possibilidade de gerarem um *modelo do mundo*, são importantes as relações entre as posições dos sensores e dos objectos que os circundam. Estas relações podem ser dinâmicas ou estáticas, o que permite ter as seguintes possibilidades:

- 1ª - Os sensores são estáticos e o ambiente é estático;
- 2ª - Os sensores são estáticos e o ambiente é dinâmico;
- 3ª - Os sensores são dinâmicos e o ambiente é estático
- 4ª - O sensor é dinâmico e o ambiente é dinâmico.

No nosso caso só trataremos problemas que se inserem na terceira possibilidade, embora a quarta possibilidade seja um objectivo para trabalhos futuros.

Neste capítulo propomos uma estrutura de representação do modelo local, que permite uma representação dos dados tridimensionais fornecidos pelo sistema de visão. Esta estrutura possibilita actualizações da informação acerca dos objectos, combinando a informação fornecida pelos sensores. No nosso caso essa informação corresponde a medidas numéricas feitas pelos sensores de visão e obtidas pelo processo da determinação da profundidade através da focagem. Esta estrutura permite a utilização de técnicas estatísticas para a combinação das medidas obtidas em diferentes pontos do espaço de trabalho do "robot". O método adoptado neste trabalho para combinar a informação utiliza a teoria da probabilidade. As razões desta opção devem-se ao facto de a teoria da probabilidade estar suficientemente desenvolvida para a sua aplicação nesta área, permitindo uma fácil especificação, manipulação e propagação da incerteza das medidas. Algumas das técnicas da teoria da probabilidade utilizadas neste trabalho são descritas nos apêndices ao texto.

7.2 Definição de Mapa Volumétrico

O modelo de representação que propomos é volumétrico (3D) e formado por elementos cúbicos de volume pré-definido. Esse elemento cúbico é conhecido por "voxel" e corresponde à designação também utilizada por Tabak [Tabak 88]. Como cada "voxel" corresponde a um volume no espaço, a informação sobre esse volume pode ser conhecida ou

desconhecida. Inicialmente a informação volumétrica correspondente a cada "voxel" do *modelo local* é desconhecida. No nosso caso o "voxel" corresponde a um volume de 5cmx5cmx5cm.

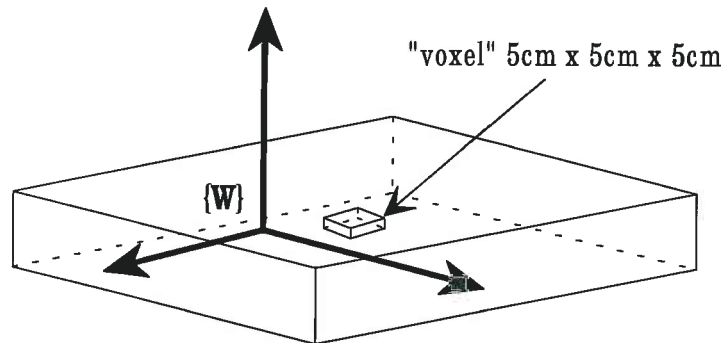


Figura 7.1 - Representação volumétrica do mapa local. O cubo central corresponde a um "voxel".

A estrutura de dados associada a um "voxel" engloba o seu estado (*conhecido, desconhecido*) e uma lista de características que envolve o valor da distância ao referencial associado ao modelo local e a classificação de espaço *ocupado* ou *livre*. Essa distância corresponderá ao valor da profundidade obtido pelo processo de focagem. A cada distância está associado um nível de incerteza que no nosso caso é representado pela soma da incerteza resultante do cálculo do seu valor e da profundidade de campo.

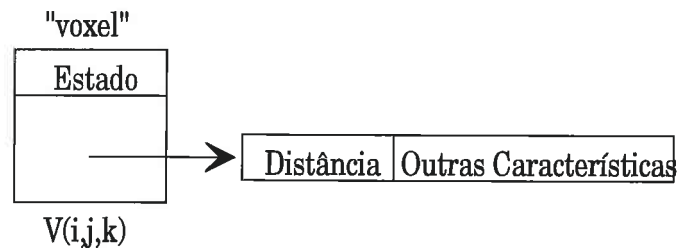


Figura 7.2 - Estrutura de dados associada a um "voxel"

Operações de Atualização do Mapa

A primeira operação a realizar no mapa é a sua iniciação. Inicialmente o estado do mapa local é desconhecido, ou seja, a informação correspondente ao volume representado por cada "voxel" $v(i, j, k)$ é desconhecida. Após a iniciação do mapa começa a sua actualização com a informação fornecida pelo sistema de visão e através do processo da determinação da profundidade por focagem. Para isso é necessário relacionar a distância fornecida pelo sistema de visão, conhecida relativamente ao referencial da câmara, com o referencial solidário com o mapa. No caso de um manipulador o sistema de coordenadas do

modelo local e o sistema de coordenadas do *modelo global* coincidem sendo ambos representados por um único referencial designado $\{W\}$. No caso de um "robot" móvel os sistemas poderão ser diferentes e a sua relação geométrica dependerá do caminho percorrido pelo veículo e deverá ter associada uma incerteza de localização da origem de coordenadas. Do ponto de vista geométrico, as medidas obtidas em diferentes posições no espaço, podem ser relacionadas através das transformações geométricas entre si. No entanto como as medidas são realizadas a nível do sensor, elas podem ser todas relacionadas com o sistema de coordenadas local. A figura 7.3 ilustra um sistema de coordenadas associado ao sensor, cujas medidas deverão ser relacionadas com o sistema de coordenadas local. Como o nosso sensor é uma câmara, essa transformação é determinada através da combinação da matriz de calibração invariante e a posição da câmara no espaço. Nas experiências realizadas, foram utilizadas duas câmaras na extremidade do manipulador numa configuração estereo, sendo a sua calibração realizada da forma descrita no capítulo 3.

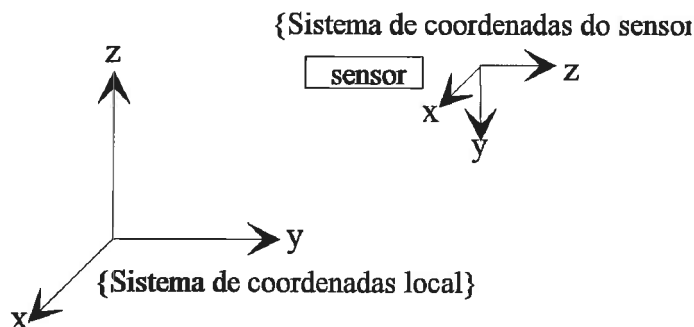


Figura 7.3 - Existe um sistema de coordenadas associado a cada sensor que deverá ser relacionado com o sistema de coordenadas local através de uma matriz de transformação.

Como o sistema de coordenadas do *modelo local* ao "robot" não se movimenta, o preenchimento do mapa só implica a sua actualização na zona fisicamente coberta pelos sensores. No caso de já existir informação acerca dessa zona e o sistema de visão fornecer de novo informação, será necessário "combinar" a nova informação com a existente. No caso de o sistema se movimentar relativamente a um sistema de coordenadas global, tal como acontece num "robot" móvel, é necessário reajustar o mapa local para a nova posição com base na informação do mapa anterior. Sendo o nosso sistema experimental baseado num manipulador, esta operação não existe pois o referencial local e global são coincidentes. O mesmo já não se passa num "robot" móvel em que é necessário utilizar as transformações entre os dois referenciais.

A informação fornecida pelo sistema de visão pode coincidir com a informação já existente no mapa e nesse caso é necessário definir uma técnica para a sua combinação. A informação fornecida pelo processo de focagem é uma informação sobre uma grandeza

geométrica dos objectos em torno do manipulador ou "robot" móvel. Mais concretamente a grandeza geométrica refere-se à posição tridimensional de um ponto na imagem. Seja um ponto ou outro tipo de grandeza geométrica a sua descrição geométrica pode ser considerada como uma variável aleatória \hat{x} , à qual corresponde uma função de densidade de probabilidade $p(\hat{x})$. Deste modo a combinação da informação dos dados fornecidos por estes sensores pode ser reduzida ao problema da combinação dos dados representados por variáveis estatísticas. As funções de densidade de probabilidade $p(\hat{x})$ dependem, evidentemente, das características físicas do sensor, podendo a forma da distribuição de probabilidade ser diferente de uma distribuição normal. Nesse caso o problema da fusão tornar-se-á mais complexo, dependendo a solução do tipo de função de probabilidade. Uma aproximação a este problema consiste em assumir a hipótese de a função de densidade de probabilidade seguir uma forma normal. Este pressuposto justifica-se no caso dos erros de medida serem de amplitude pequena e pouco frequentes ou serem considerados como soma de um número largo de erros independentes. Com base nesta hipótese a função de distribuição de probabilidade fica bem especificada pela média \bar{x} e pela covariância Λ . Nesse caso podemos utilizar um conjunto de técnicas da teoria da probabilidade para manipular, estimar e combinar informação. O processo de combinação utilizado nestas experiências baseia-se na utilização de técnicas de combinação da informação, admitindo a hipótese da distribuição de probabilidade seguir uma lei normal.

Como a grandeza física a combinar é correspondente à distância ao objecto obtida pela focagem, a sua caracterização em termos estatísticos é definida pela sua média \bar{z} e variância σ . O resultado da combinação de duas distâncias z_i e z_j terão como resultado uma distância final z_f que é dado pelas expressões (C.1) (C.8) e (C.9) discutidas no apêndice C. Estas expressões são

$$z_f = z_i + \alpha(z_j - z_i) = \frac{\sigma_i^2 z_j + \sigma_j^2 z_i}{\sigma_i^2 + \sigma_j^2}$$

$$\alpha = \frac{\sigma_j^2}{\sigma_i^2 + \sigma_j^2} \quad (7.1)$$

A variância associada à estimação dada pelas expressões (7.1) é dada pela expressão

$$\sigma_f^2 = \frac{\sigma_j^2 \sigma_i^2}{\sigma_i^2 + \sigma_j^2} \quad (7.2).$$

Zonas do mapa onde se deverá actualizar a informação

Para se determinar a zona onde deverá ser combinada a informação utiliza-se a informação da posição das câmaras e a sua geometria de projecção. Se por aplicação do processo de focagem uma determinada medida z é determinada para um ponto p da imagem então o ponto tridimensional P é conhecido no espaço 3D - veja figura 7.4.

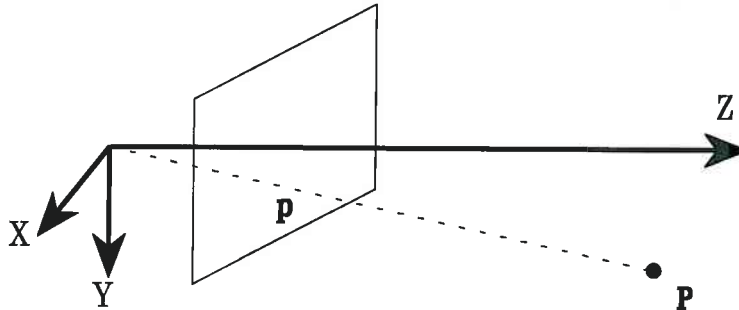


Figura 7.4 - Rectificação da medida dada pela focagem

No caso de a recta de projecção de P intersectar um dos "voxel" da estrutura volumétrica e esse "voxel" apresente o seu estado conhecido então teremos que combinar as duas grandezas segundo (7.1) e (7.2). Para realizar a pesquisa do espaço volumétrico são testados os pontos da recta projectante que intersectam os "voxel". Essa pesquisa é limitada a um intervalo em torno do valor da distância z . Para além disso a combinação das medidas está condicionada à realização de um teste estatístico da hipótese de correspondência das medidas. O teste estatístico baseia-se nos valores da média \bar{z} e da variância σ obtidos com sensores diferentes ou com as câmaras colocadas em locais diferentes do espaço. Se i e j forem duas medidas descritas pelo vector de média x e sua covariância P então é possível estabelecer um método para o teste da hipótese de correspondência, com base na estatística χ^2 da diferença entre os valores das médias para x_i e x_j [Bar-Shalom 81] [Ayache 88a]. Para se realizar este teste defina-se por d_{ij} a diferença entre os valores das medidas i e j através da expressão

$$d_{ij} = x_i - x_j \tag{7.3}$$

Assumindo que as medidas realizadas são independentes, então a matriz de covariância Σ_{ij} associada à diferença (7.3) é dada pela soma

$$\Sigma_{ij} = P_i + P_j \tag{7.4}$$

em que \mathbf{P}_i e \mathbf{P}_j correspondem respectivamente às matrizes de covariância da primitiva i e j . Como se admite, por hipótese, que o comportamento estatístico das amostras segue uma lei normal, assume-se que a diferença \mathbf{d}_{ij} segue também uma lei normal. Deste modo a quantidade

$$d^2 = \mathbf{d}_{ij}^T \Sigma_{ij}^{-1} \mathbf{d}_{ij} \quad (7.5)$$

irá ter uma distribuição χ^2 com um número de graus de liberdade p igual ao número de elementos do vector diferença \mathbf{d}_{ij} [Bar-Shalom 81]. Um teste da hipótese de as duas primitivas pertencerem à mesma grandeza física consiste em verificar se

- se $d^2 < T_s$ então as características pertencem à mesma grandeza física;
- se $d^2 \geq T_s$ então as características não pertencem à mesma grandeza física.

sendo T_s um limiar de semelhança. A escolha de T_s baseia-se nas propriedades χ^2 de d^2 e envolve o ajuste do seu valor para a aplicação em causa. Para além disso a sua escolha envolve o grau de confiança que pretendemos para o teste (95%, 98% ou 99%). Este teste foi utilizado nas experiências sendo o vector de medidas representado pela medida z ao qual está associada a variância σ .

Após este teste os valores das grandezas podem ser combinados e o "voxel" correspondente ao valor da distância determinada é actualizado. O "voxel" existente anteriormente no mapa é actualizado e classificado como espaço *ocupado* ou *livre* de acordo com a sua posição na recta de projecção.

7.3 Resultados Experimentais com Base na Focagem

Os valores de profundidade fornecidos pela segmentação por focagem correspondem a mapas densos, com valores aproximados para a profundidade em todos os pontos da imagem. Estes mapas são úteis para definir zonas de espaço livre para movimentação do "robot".

A informação de profundidade fornecida pela técnica da focagem corresponde à distância de um plano de focagem ao centro da lente da câmara que captou a imagem. No entanto essa informação, para ser utilizada no mapa local, deverá ser transformada de modo a converter a distância em coordenadas do referencial local. Para isso é necessário conhecer a posição da câmara no espaço, ou seja que a câmara esteja também calibrada. Para isso consideremos a matriz de calibração da câmara (6.3) e dada por

$$\begin{bmatrix} U \\ V \\ S \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \\ c_{31} & c_{32} & c_{33} & c_{34} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{c}_1^T \mathbf{P} + c_{14} \\ \mathbf{c}_2^T \mathbf{P} + c_{24} \\ \mathbf{c}_3^T \mathbf{P} + c_{34} \end{bmatrix} \quad (7.6)$$

em que \mathbf{c}_i é um vector que representa as três primeiras colunas da linha i da matriz de calibração. Pela equação (6.6) sabemos que o vector director da recta projectante do ponto \mathbf{P} no mundo é dado por

$$\mathbf{n} = u(\mathbf{C}_2 \wedge \mathbf{C}_3) + v(\mathbf{C}_3 \wedge \mathbf{C}_1) + (\mathbf{C}_1 \wedge \mathbf{C}_2) = u\mathbf{g} + v\mathbf{h} + \mathbf{k} \quad (7.7)$$

em que (u, v) representam as coordenadas da imagem do ponto.

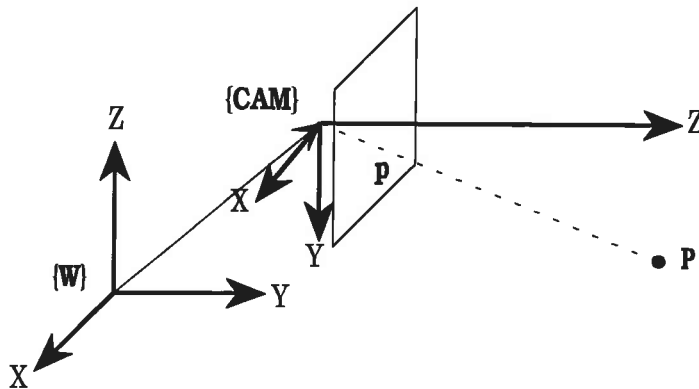


Figura 7.5 - Projecção do ponto $\mathbf{P} = (x, y, z)$ no ponto da imagem $\mathbf{p} = (u, v)$

A recta projectante, tal como mostra a figura 7.5, pode ser expressa através da expressão paramétrica

$$\mathbf{C} + \lambda \mathbf{n} \quad (7.8)$$

em que λ corresponde à distância a que o ponto \mathbf{P} se encontra do centro e é dado pelo valor da distância d' fornecida no mapa de profundidade obtido por focagem afectado de um valor de rectificação. Esta rectificação corresponde ao ajuste geométrico necessário realizar de modo a compensar o ângulo entre a recta de projecção e o eixo óptico e sobre o qual se descreve a distância d' .

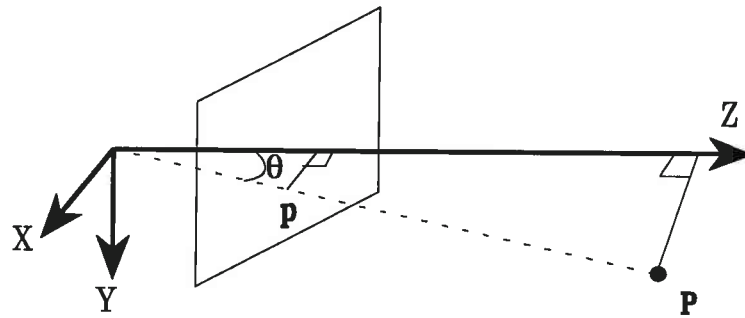


Figura 7.6 - Rectificação da medida dada pela focagem

Deste modo λ é dado por

$$\lambda = \frac{d'}{\cos\theta} = \frac{d'}{\hat{\mathbf{k}} \hat{\mathbf{n}}} \quad (7.9)$$

em que $\hat{\mathbf{n}}$ é o vector unitário obtido pela expressão (7.7) e $\hat{\mathbf{k}}$ é o resultado da mesma expressão fazendo $(u, v) = (0, 0)$.

A medida da distância z é obtida através de uma relação que é calculada durante o processo de calibração da lente e descrito no ponto 5.4. Esta expressão dá uma relação entre a posição angular da focagem da lente e a distância z . A medida de z obtida por este processo é afectada da incerteza devido à profundidade de campo e da incerteza estatística da curva de interpolação que exprime a relação entre a posição angular da focagem da lente e a distância. Para calcular a incerteza devida à profundidade de campo é utilizada a expressão (2.8) que permite obter uma ordem de grandeza para a profundidade de campo das lentes. Esta expressão é dependente da abertura, da distância focal e de um parâmetro c correspondente ao diâmetro do círculo de confusão. Como as câmaras utilizadas são do tipo CCD o diâmetro de c é dependente da área de cada sensor no "array" CCD. Pelas características técnicas das câmaras chegamos à conclusão que cada sensor tem uma área $17\mu\text{m} \times 13\mu\text{m}$ ou seja podemos admitir um diâmetro $c = 17\mu\text{m}$ para o círculo de confusão. Na figura 7.7 está ilustrada a curva expressa pela relação (2.8) para a lente de 50mm que foi utilizada nas experiências. Esta curva exprime a componente da incerteza devido à profundidade de campo para a abertura no máximo, tal como foram realizadas as experiências.

A esta incerteza da medida deverá ser somada a incerteza estatística da curva que exprime a posição angular da focagem da lente e a distância a que se encontra o objecto. Para a lente utilizada (50mm) a curva de interpolação tem a expressão dada na figura 5.9 e que foi obtida para $n=29$ amostras. Essa expressão é a seguinte

$$z_i = 0.159862 + \frac{77.519123}{k_i} \quad (7.10).$$

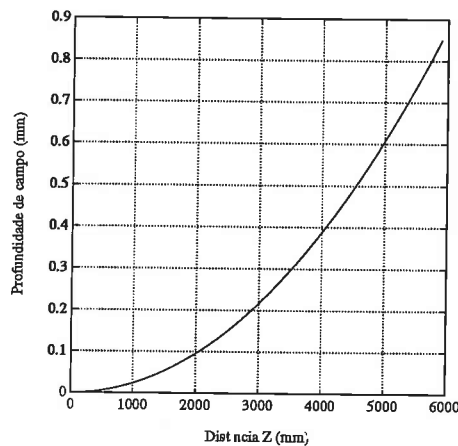


Figura 7.7 - Curva correspondente ao valor da profundidade de campo para as lentes de 50mm(1.8).

A curva expressa por (7.10) dá valores próximos dos valores obtidos para as amostras pelo processo da focagem. Nesse caso o valor para a dispersão é dado por

$$s^2 = \frac{\sum_{i=1}^n (z_i - \bar{z}_i)^2}{n - 2} = 0.011922 \quad (7.11)$$

sendo n o número de amostras (29 no nosso caso). Com base neste valor estimado para a dispersão é possível obter um valor máximo do erro de uma previsão para z_i com um nível de confiança pré-definido (por exemplo 95%) [Moore 93]. Nesse caso uma previsão para z_i poderá ser obtida pela expressão

$$\hat{z}_i \pm t^* s_{\hat{z}_i} \quad (7.12)$$

em que $s_{\hat{z}_i}$ é dada pela expressão

$$s_{\hat{z}_i} = s \sqrt{1 + \frac{1}{n} + \frac{(k - \bar{k})^2}{\sum_{i=1}^n (k_i - \bar{k})^2}} \quad (7.13)$$

e t^* é dado pelo valor tabelado da distribuição- t para o valor crítico $\frac{1-c}{2}$ para $(n - 2)$ graus de liberdade. O valor c é o nível de confiança desejado para o teste (por exemplo para $c=95\%$ e 7 graus de liberdade dá uma valor $t^* = 2.365$).

Nas experiências realizadas esta incerteza (profundidade de campo+incerteza estatística) foi calculada com base num nível de confiança de 95%. Na maioria dos casos os seus valores não conseguem ter expressão gráfica nas figuras. No caso das experiências

realizadas o valor da incerteza aos 6m atinge um valor de $\approx 15\%$, correspondendo a $\approx 6 \pm 0.45$ metros.

As figuras seguintes ilustram alguns dos resultados das experiências realizadas. O primeiro conjunto consiste na obtenção e combinação de mapas obtidos por um par de câmaras estereo (um mapa de focagem por câmara) e o segundo o mapa final obtido pela combinação dos mapas obtidos por uma câmara em diferentes posições de uma trajectória. Na figura 7.8 ilustra os resultado para o caso das câmaras em configuração estereo. A figura 7.8a representa as imagens estereo iniciais e as demarcação de zonas de grande variação da profundidade relativa obtida pelo processo da focagem. A figura 7.8b ilustra os diferentes planos de corte do mapa volumétrico que corresponde à combinação dos mapas de focagem individuais para cada câmara. Nas figuras 7.8c, 7.8d, 7.8e representam uma pequena zona do mapa individual de cada câmara e do mapa global e ilustram as zonas em que foram determinados os mapas de profundidade. Pela figura 7.8d podemos verificar as zonas nas quais foi utilizado o processo de fusão dos dois mapas.

A figura 7.9 ilustra a aplicação da mesma técnica numa maior extensão de imagens. A figura 7.9b representa planos dos mapas locais nos quais se representam as zonas de espaço vazio (zonas mais claras).



Figura 7.8 (a) - Imagem Esquerda



Figura 7.8 (a) Imagem Direita

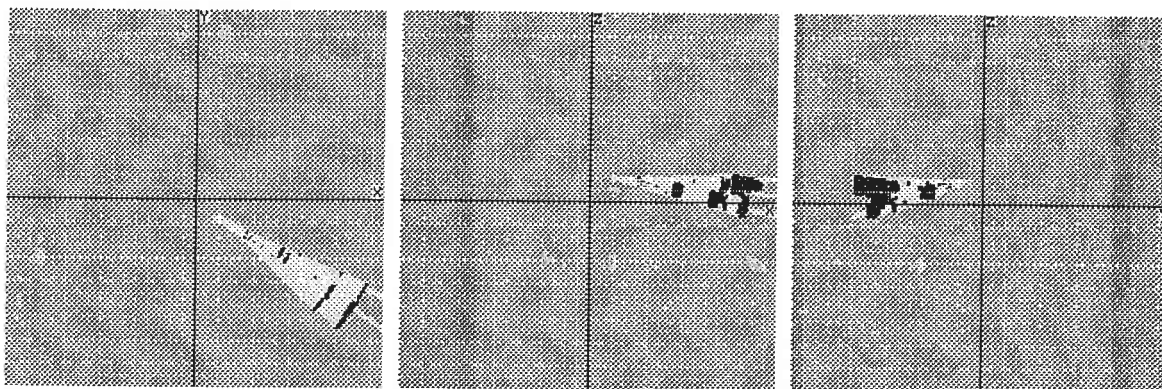


Figura 7.8 (b) Cortes correspondentes ao mapa volumétrico

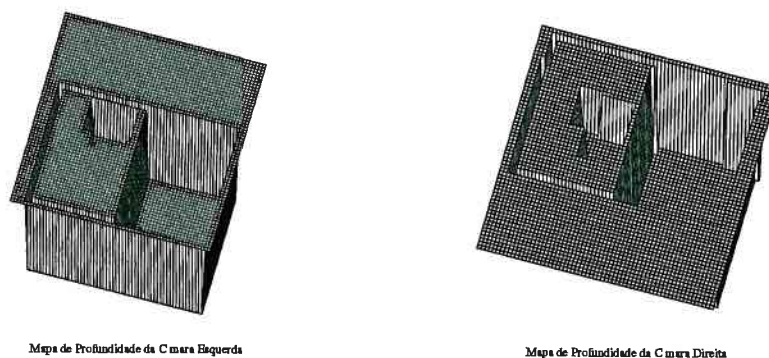


Figura 7.8 (c) Zona detalhada do mapa parcial de cada câmara

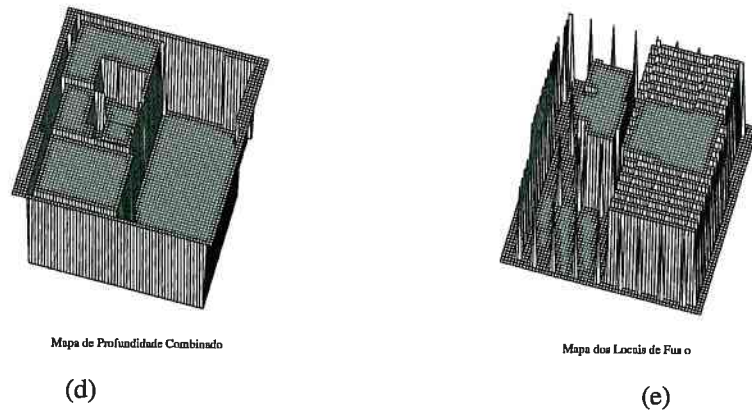


Figura 7.8 - Fusão de mapas de profundidade obtidos pelo método da focagem. Na figura (a) estão representadas as imagens estereo. Na figura (b) estão representados planos correspondentes aos mapas densos obtidos. Na figura (c) estão representados dois pequenos pedaços dos mapas da câmara esquerda e direita. A figura (d) representa o mapa combinado. Na figura (e) estão representas as zonas da fusão da informação.

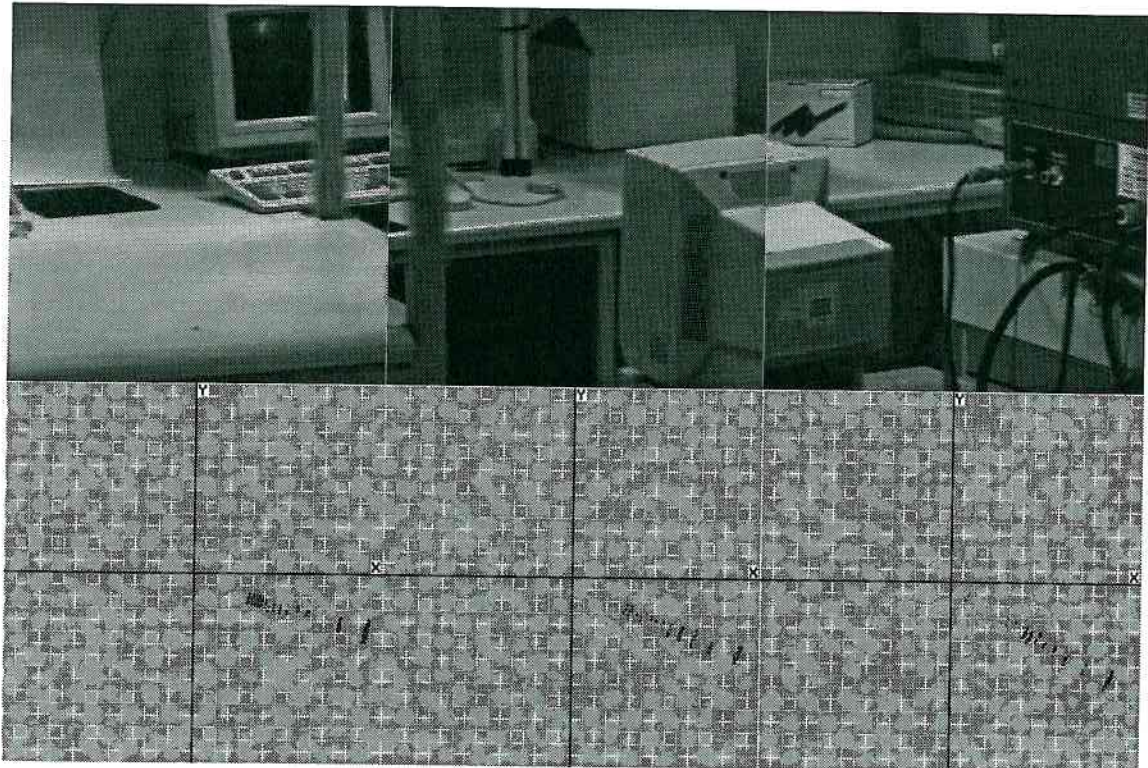


Figura 7.9 (b) Sequência de imagens

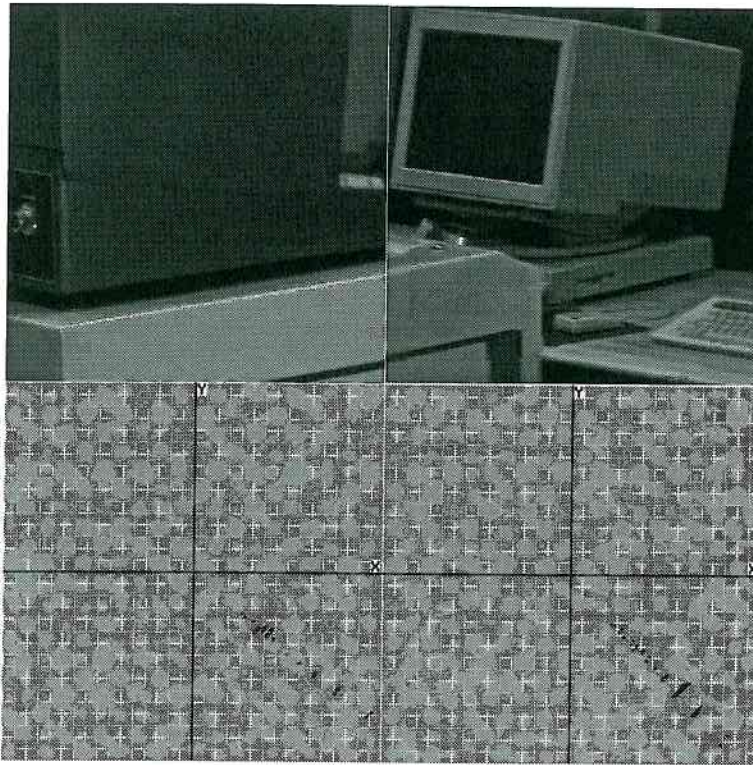
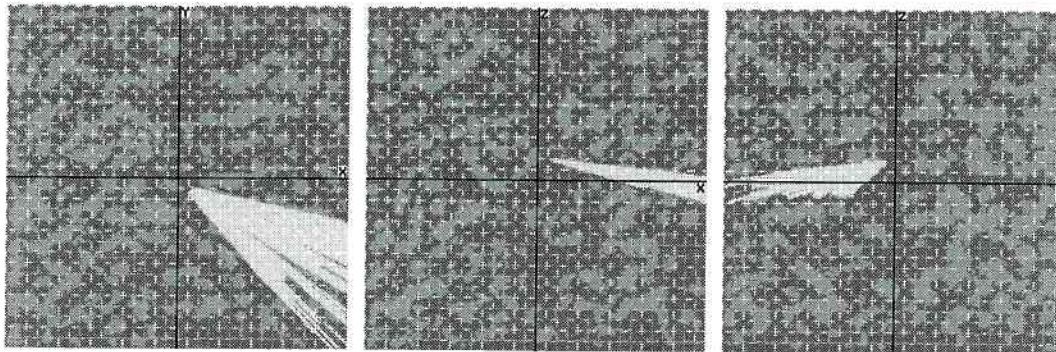


Figura 7.9 (b) Continuação



(b)

Figura 7.9 - Fusão de mapas de profundidade para uma zona extensa. Na figura (a) estão representados uma das imagens estereo dos vários locais da zona extensa. Por baixo das imagens estão representados cortes horizontais (x-y) dos mapas parciais correspondentes aos diferentes locais. Notar que cada ponto do mapa representa um área de 1cm x 1cm. Na figura (b) estão representados os cortes (x-y), (x-z) e (z-y) do mapa volumétrico final e correspondente aos mapas anteriores combinados. As zonas claras representam espaço vazio.

CAPÍTULO 8

Desenvolvimento de Um Sistema de Visão Activa

8.1 Introdução

Uma grande parte dos algoritmos que foram desenvolvidos na área da Visão por Computador não exploram a possibilidade de adaptar o sistema de visão às condições das cenas. No entanto essa adaptação pode ser necessária para obter os resultados desejados.

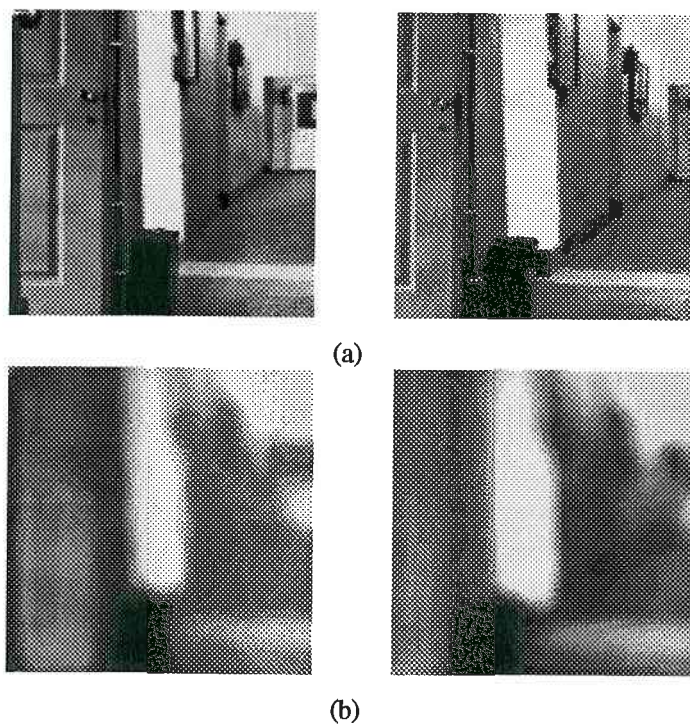


Figura 8.1- Par de imagens estereo. (a) Focadas perfeitamente. (b) Captadas com focagem desadequada.

Na figura 8.1 é apresentado um exemplo em que um par de imagens estereo não está correctamente adaptado para ser utilizado pelo sistema de estereovisão. Para adquirir as imagens correctamente seria necessário controlar a focagem e eventualmente a vergência entre as câmaras. Para além da adaptação do sistema às condições da cena existem situações em que é necessário adaptar o sistema de visão aos objectivos do processamento. Por exemplo na figura 8.2, são apresentadas duas imagens de um mesmo objecto obtidas com uma lente "zoom" em diferentes distâncias focais. Na imagem da figura 8.2 (a) o objecto é mais pequeno comparado com a imagem da figura 8.2 (b) onde o objecto já aparece mais pormenorizado na imagem. Se o objectivo for a análise pormenorizada do objecto central, então a imagem mais adequada é a da figura 8.2 (b). Estes exemplos ilustram que é necessário adaptar o sistema de visão às condições das cenas e por vezes aos objectivos pretendidos. Outro exemplo é a forma como a distância focal influencia o erro na determinação da profundidade utilizando estereovisão tal como demonstrado analiticamente pelas equações (6.60). Pela expressão verificamos que podemos diminuir o erro controlando a distância focal.

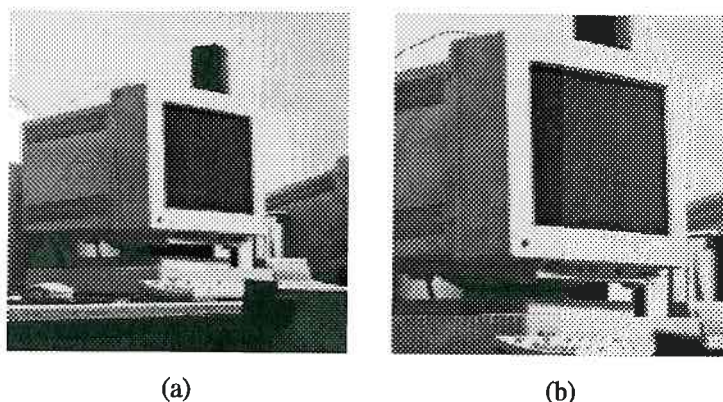


Figura 8.2 - (a) A imagem esquerda corresponde a uma imagem obtida com um comprimento focal de 16 mm; (b) Imagem idêntica mas para um comprimento focal de 25 mm.

A adaptação de um sistema de visão às condições da cena ou aos objectivos pretendidos com o processamento, pode ser realizada através do controlo de alguns parâmetros do sistema de visão. Nestes sistemas esses parâmetros são de dois tipos: ópticos e mecânicos. Os primeiros estão relacionados com as lentes e os segundos com o sistema de suporte e orientação das câmaras. Ambos funcionam como graus de liberdade do sistema e podem ser controlados por computador. Dos graus de liberdade ópticos fazem parte o controlo da focagem, "zoom" e a abertura e dos graus de liberdade mecânicos fazem parte todas as juntas de rotação e translação que permitem posicionar as câmaras no espaço.

Esta forma controlada de captação de imagens apresenta alguma semelhança com o sistema visual humano. Nos seres humanos a percepção visual é um processo activo, fortemente influenciado pelo meio e pelos objectivos a concretizar. Por exemplo, a percepção de um local ou o reconhecimento de uma face é executada pela verificação

sucessiva de diferentes pontos de interesse para o objectivo e não através da análise de uma única imagem [Yarbus 67]. Os movimentos realizados pelos olhos humanos, são também componente integral na percepção visual. Outro exemplo é quando o sistema visual humano é confrontado com uma nova cena. Neste caso o sistema visual organiza e executa uma sequência de movimentos dos olhos de modo a obter uma percepção da cena. Estes movimentos são caracterizados por breves períodos de fixação num ponto e de rápidos movimentos entre pontos de fixação [Yarbus 67] [Carpenter 88].

8.2 Desenvolvimento de um Sistema de Visão Activa

8.2.1 Introdução

A visão humana é uma referência para quem estuda visão computacional. Para isso é importante analisar o sistema visual humano, relacionando-o com os dispositivos actuais de captação e processamento de imagens. O olho humano pode ser comparado a um dispositivo óptico que recebe a energia radiada pelos objectos na cena. Essa energia forma uma imagem invertida na retina e excita um conjunto de células visuais que geram impulsos nervosos. Esses impulsos nervosos são transmitidos pelo nervo óptico até ao cérebro e estimulam os centros visuais do cérebro. Este é, de uma forma resumida, o procedimento que nos permite *ver*. A figura 8.3 representa, de maneira esquemática, um corte horizontal do olho humano.

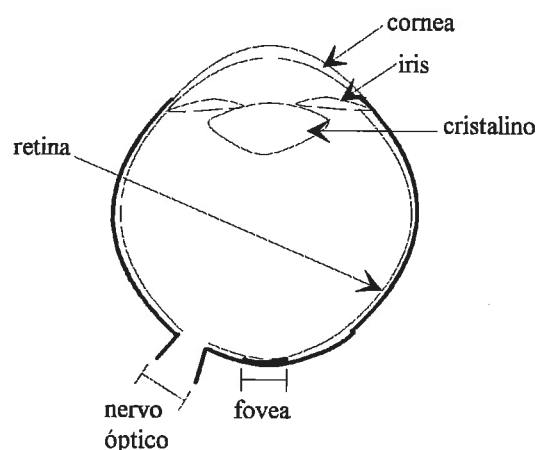


Figura 8.3- Representação grosseira de um corte horizontal do olho humano

A entrada de radiação é feita pela córnea, passando pela abertura da íris e o cristalino e alcança, na parte posterior, a retina. São os foto-receptores existentes na retina que são responsáveis pela transformação de energia luminosa em impulsos nervosos na forma de sequências de pulsos modulados em frequência. O olho humano tem em média 24mm de

diâmetro, é flexível e sabe-se que só cerca de 50% da energia luminosa recebida pela cornea chega à retina [Levine 85][LeGrand 80]. A pupila é uma abertura no centro da íris cuja pigmentação é a responsável pela cor dos nossos olhos. Esta abertura circular é similar à abertura existente numa lente, cujo diâmetro é controlado de modo a permitir a entrada de mais ou menos radiação no cristalino. A activação do controlo é influenciada por vários factores, tais como a quantidade de iluminação incidente e o estado emocional da pessoa. A variação do tamanho da abertura serve três objectivos[Levine 85]:

- O controlo da quantidade de radiação que passa o cristalino e chega à retina.
- O controlo da profundidade de campo para objectos muito próximos.
- A redução das aberrações da imagem de modo a reduzir as indefinições das imagens no caso de situações de iluminação muito brilhante.

O cristalino permite a adaptação para visão ao perto e visão ao longe através da mudança da sua forma e cujo comportamento se assemelha ao controlo da focagem de uma lente. O cristalino garante que a imagem está focada na zona de maior acuidade visual, chamada de fovea.

O olho humano pode ser encarado como um sistema óptico que transforma a radiação oriunda do espaço tridimensional numa imagem bidimensional desse mesmo espaço. Essa transformação é limitada pelas características físicas do olho e controlada pelos graus de liberdade ópticos tais como a íris e a acomodação do cristalino. Estes graus de liberdade podem ser *simulados* através de lentes motorizadas que podem controlar a focagem, "zoom" e a abertura.

O sistema visual humano movimenta os olhos durante a execução dos processos visuais como por exemplo a seguimento ou a fixação. Os movimentos realizados pelos olhos podem ser simulados pelos movimentos ilustrados pelas setas da figura 8.4. Estes movimentos podem ser simulados através de sistemas mecânicos cujo movimento é controlado por motores.

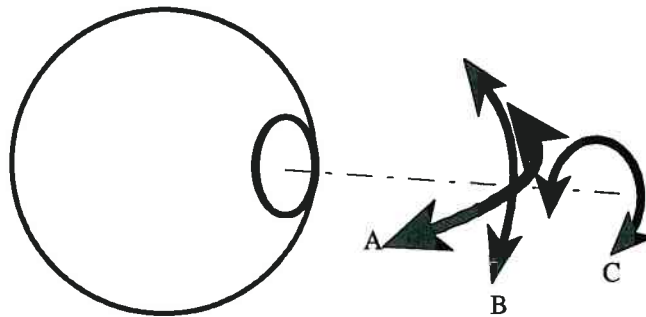


Figura 8.4 - Graus de liberdade mecânicos do olho humano. O movimento do olho pode ser decomposto nos movimentos A, B e C.

A simulação dos graus de liberdade do olho humano, através de lentes e dispositivos mecânicos controlados por computador, permitirá explorar novos algoritmos de visão por

computador através de técnicas de controlo da aquisição das imagens. Este assunto é abordado, em parte, neste capítulo através da descrição de uma plataforma experimental desenvolvida para realizar trabalhos de visão activa. Os algoritmos de visão computacional que envolvem este tipo de controlo são conhecidos como algoritmos sobre visão activa.

8.2.2 Vantagens de um sistema de Visão Activa

A possibilidade de controlar o modo de aquisição de imagens através de um sistema de Visão Activa permite a exploração de novos modos de processamento em visão por computador. Um sistema de visão deste tipo deixa de funcionar de maneira isolada para funcionar cooperativamente com o processo de controlo dos vários graus de liberdade: zoom, abertura, focagem e posicionamento no espaço das câmaras. Um exemplo dessa cooperação, é o controlo do sistema de modo a fixá-lo num determinado ponto existente na cena. Para conservar o sistema de visão fixo no ponto, num ambiente dinâmico, é necessário controlar diferentes graus de liberdade do sistema. Este controlo é bastante facilitado se durante o processo for utilizada a imagem da região circundante ao ponto de intersecção dos eixos ópticos das duas câmaras e designado por ponto de fixação. Outro exemplo é a necessidade de adaptar o sistema de visão às variações de iluminação da cena. Para que o sistema se acomode às diferentes condições de iluminação, o sistema deverá permitir variações no diâmetro de abertura da lente. Além disso, é necessário que o sistema se adapte à distância a que os objectos se encontram na cena para que as imagens captadas se encontrem perfeitamente focadas. Neste caso a posição relativa entre o sensor e a lente deverá ser controlada. O sistema óptico da lente deverá também permitir diferentes distâncias focais através do controlo do "zoom".

Ballard, Aloimonos, Bajcsy e Abbot têm realçado as vantagens da Visão Activa comparadas com as obtidas com câmaras fixas [Ballard 88] [Aloimonos 88] [Aloimonos 89] [Bajcsy 88] [Abbot 88]. Em termos de visão por computador o controlo dos diferentes graus de liberdade corresponde, em parte, a uma manipulação da forma como as imagens são adquiridas. O efeito é idêntico ao modo como o movimento influencia o fenómeno de percepção. O controlo da maneira como as imagens são adquiridas permite obter estabelecer novas formas para a determinação da profundidade através de focagem, movimento e vergência ou ainda, por estereovisão dinâmica tal como ficou demonstrado nos capítulos anteriores. Mas as vantagens não estão circunscritas à determinação da estrutura tridimensional. A localização das regiões de interesse no espaço tridimensional é outra das possibilidades de um sistema de visão activa [Bajcsy 88] [Nelson 88]. Em resumo, a visão activa apresenta diferentes tipos de vantagens:

- 1) *Permite pesquisar os objectos que se encontram na cena.* O sistema pode ser controlado de modo a movimentar as câmaras para se aproximar mais dos objectos, mudar a focagem ou mudar o ponto de vista no qual as imagens estão a ser captadas

[Pentland 85] [Krotkov 88] [Wixson 89]. Por vezes esta pesquisa visual do objecto tem custos menores do que a pesquisa algorítmica numa única imagem [Nelson 88].

- 2) *Permite executar processamentos com base em movimentos conhecidos do sistema de visão.* Se os movimentos são conhecidos, então fornecem restrições adicionais no processo de captação de imagens [Aloimonos 88]. Esse facto facilita o processo computacional. Por exemplo, propriedades que são difíceis de calcular com uma câmara fixa podem ser obtidas com facilidade com um sistema de câmaras moveis. Uma das primeiras demonstrações destas vantagens foi o trabalho de Bandopadhyay para determinação dos parâmetros de movimento do corpo [Bandopadhyay 87].
- 3) *Permite referenciar os objectos relativamente a um sistema de coordenadas externas.* A capacidade de controlo de um sistema de visão activa e sobretudo a capacidade de se "fixar" num objecto da cena permite ao sistema visual determinar estruturas tridimensionais, ou outras propriedades físicas, com base num sistema de coordenadas solidário com o objecto [Ballard 90].
- 4) *Permite utilizar algoritmos de controlo com base no posicionamento relativo.* Supondo um sistema de coordenadas solidário com um ponto do objecto sobre o qual o sistema de visão se encontra fixo, é possível estabelecer estratégias de controlo do sistema de visão baseadas nesse referencial. Estas estratégias de controlo são mais simples do que as utilizadas com o sistema de coordenadas no objecto [Ballard 90].
- 5) *Permite a segmentação das imagens entre objectos e o fundo.* Se um objecto em movimento está a ser seguido, então o fluxo óptico desse objecto será muito pequeno comparado com o restante da cena (o fundo). As diferenças nos campos de fluxo podem ser utilizadas para segmentar a imagem. Para seguir o objecto faz-se uso de graus de liberdade dos mecanismos de controlo do sistema activo [Brown 90a] [Brown 90b].
- 6) *Os sistemas activos podem explorar o universo circundante através do controlo selectivo da área da cena que está a ser processada.* O controlo da posição das câmaras e dos graus de liberdade ópticos de um sistema de visão activa, permite uma selecção da zona da cena que é mais relevante para o processamento em curso. Esta selecção permite reduzir e seleccionar a informação visual a processar.

Estas são algumas das muitas vantagens de um sistema de visão activa.

8.2.3 Uma Base Experimental para Visão Activa

Para o desenvolvimento de algoritmos com base nos conceitos de visão activa foi construída uma base experimental que permite o controlo da orientação das câmaras e das lentes através de um computador. Neste ponto descrevem-se os graus de liberdade do sistema mecânico e o "hardware" de controlo do sistema. Durante o desenvolvimento do

o sistema foi prevista a sua interligação com outros sistemas. A figura 8.5a ilustra com um diagrama funcional o sistema de visão activa.

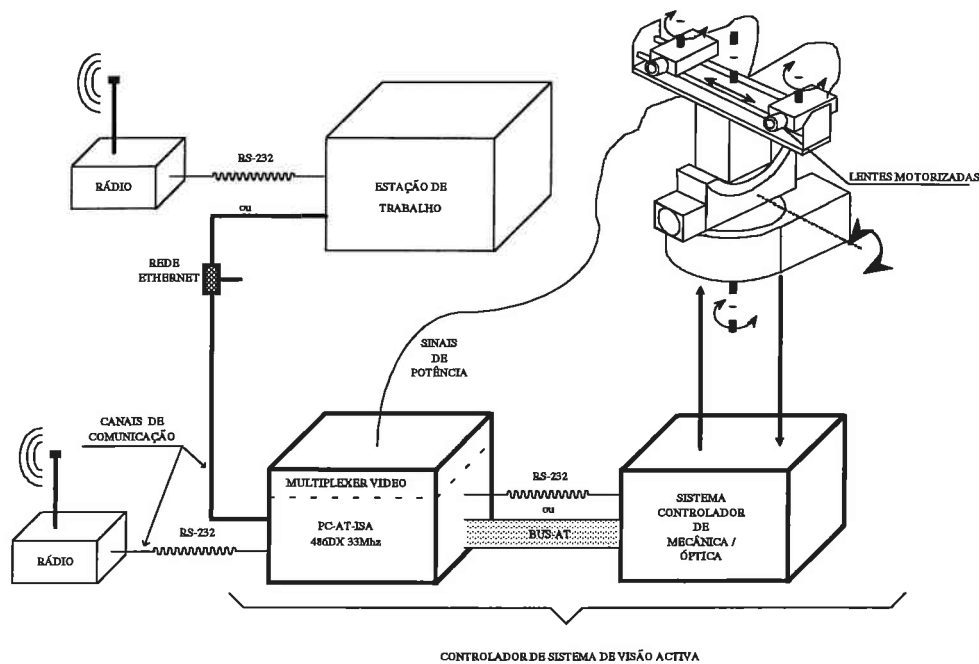


Figura 8.5 (a) - Diagrama funcional do sistema de Visão Activa

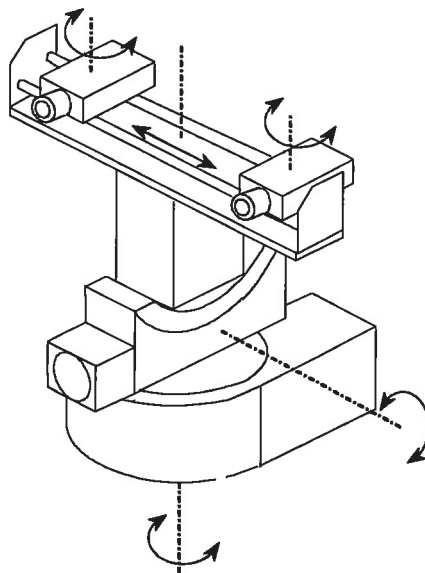


Figura 8.5 (b) - Graus de liberdade mecânicos sistema de visão activa

8.2.4 Estrutura Mecânica

A estrutura mecânica é composta por três partes: uma plataforma de suporte das câmaras, um sistema mecânico que orienta essa plataforma e um sistema mecânico que controla as lentes. A plataforma de suporte das câmaras possibilita a simulação do movimento *A* da figura 8.4 (*vergência*) e o controlo da distância intra-ocular isto é, a

distância entre as câmaras (*baseline*). O sistema mecânico que suporta a plataforma é responsável pela execução do movimento *B* da figura 8.4 (*tilt*) e do movimento de rotação do nosso pescoço segundo um eixo vertical (*neck pan*). Este último movimento é responsável pela orientação do suporte das câmaras. O sistema mecânico que controla as lentes é responsável pelo controlo dos parâmetros ópticos de cada uma das lentes, respectivamente a *abertura*, a zona de *focagem* e a *distância focal*. A figura 8.5b ilustra o os eixos correspondentes aos graus de liberdade sistema.

A plataforma de suporte das câmaras utiliza motores de passo para os movimentos da *vergência* e da *baseline*. O sistema permite o controlo separado da vergência de cada uma das câmaras. Para o movimento *tilt* e para o movimento *neck pan* utilizam-se também motores de passo. Os motores utilizados nestes graus de liberdade permitem uma velocidade máxima de 83° por segundo. Cada um dos motores pode ser controlado de modo a executar movimentos de meios-passo mas todos, à excepção dos motores da *vergência*, executam movimentos de passo inteiro. Associado a cada motor de passo existe um codificador óptico incremental, dois sensores de fim de curso e um sensor da posição de referência (*home*). A função dos codificadores incrementais é permitir a identificação da posição actual do sistema mecânico. Os sensores de fim de curso evitam que os motores realizem movimentos para posições proibidas e a posição de referência é útil na fase de iniciação do sistema.

As lentes motorizadas utilizam motores de corrente contínua para controlarem os graus de liberdade correspondentes à *íris*, *focagem* e ao *zoom*. Esses motores são de pequena potência e estão situados junto da lente. As lentes utilizadas permitem uma distância focal entre 12,5mm e 75mm e uma focagem de 1,3m a ∞ variável com a distância focal.

Tipo de movimento	Tipo de motor	Deslocamento máximo	Resolução	Pulsos por unidade de deslocamento
<i>Vergência</i>	Motor de passo	$\pm 34,2^\circ$	$\pm 0,18^\circ$	$\approx 6 \text{ passos / } ^\circ$
<i>Baseline</i>	Motor de passo	$2 \times 138 \text{ mm}$	$0,15 \text{ mm} / 100 \text{ passos}$	$\approx 666 \text{ passos} / \text{ mm}$
<i>Tilt</i>	Motor de passo	$\pm 25^\circ$	$\pm 0,01^\circ$	$\approx 100 \text{ passos} / ^\circ$
<i>Neck pan</i>	Motor de passo	$\pm 100^\circ$	$\pm 0,01^\circ$	$\approx 100 \text{ passos} / ^\circ$

Tabela 8.1 - Parâmetros dos motores utilizados no sistema de visão activa

8.2.5 "Hardware" de Controlo do Sistema Mecânico

Para efectuar um controlo dos motores de passo utilizados na estrutura mecânica foi construído uma placa de controlo para funcionar em computadores com "bus-PC AT" designado Cartão de Controlo de Motores de Passo (CC-MP). Cada placa CC-MP é responsável pelo controlo e monitorização dos motores de passo até ao máximo de seis motores por cartão. Para o controlo dos motores de corrente contínua foi construído uma

mecânico. A troca de informação entre o *controlador do sistema mecânico* e o exterior pode ser realizada por uma ligação série RS-232 ou através de rede "ethernet". No caso de o sistema ser utilizado num "robot" móvel poderá ser utilizada a transmissão série RS-232 e um "modem". O mesmo tipo de ligação assegura, através de um "software" especialmente projectado para o efeito, a transmissão de imagens sem erros. Esta característica é especialmente útil quando o sistema é utilizado num "robot" móvel.



Figura 8.7 - O Sistema de Visão Activa

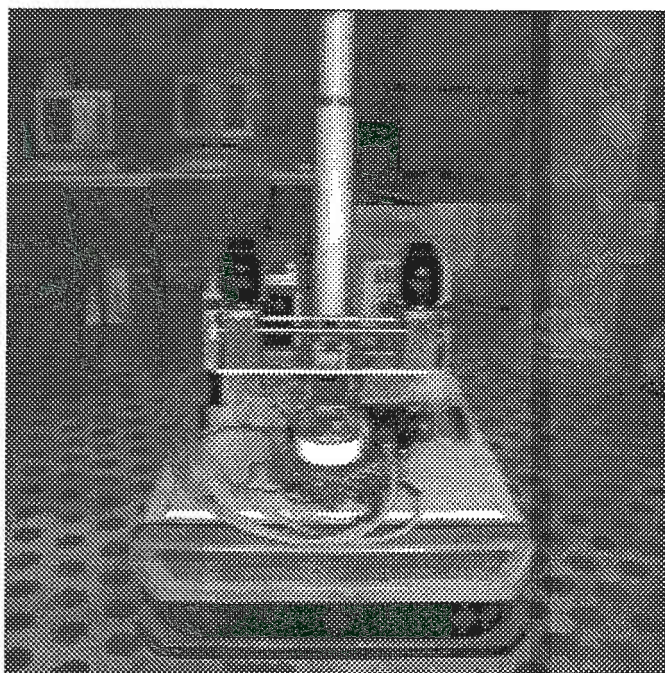


Figura 8.8 - O sistema de visão activa colocada no "robot" móvel ROBUTER

8.2.6 "Software" de Controlo

Existem duas possibilidades para realizar o controlo do sistema de visão activa. Uma consiste na realização de programas pelo utilizador do sistema e que recorram a rotinas existentes numa biblioteca de programas. As rotinas foram desenvolvidas em linguagem C para serem utilizadas com o sistema operativo *PC-DOS*. As rotinas existentes nessa biblioteca executam comandos de movimento e controlam a sua correcta execução. Para além disso o programador tem a possibilidade de ter acesso aos vários controladores de motores existentes no sistema, discriminadores incrementais, *reset* das placas e o estado do movimento de cada motor. Os comandos disponíveis são considerados de *baixo nível*, deixando à responsabilidade do programador os problemas de controlo de parâmetros de velocidade, estado geral do sistema e geração das trajectórias. Esta forma de operação é a mais adequada para aplicações de *tempo-real* e nos casos em que não existe necessidade de interactividade do utilizador com sistema.

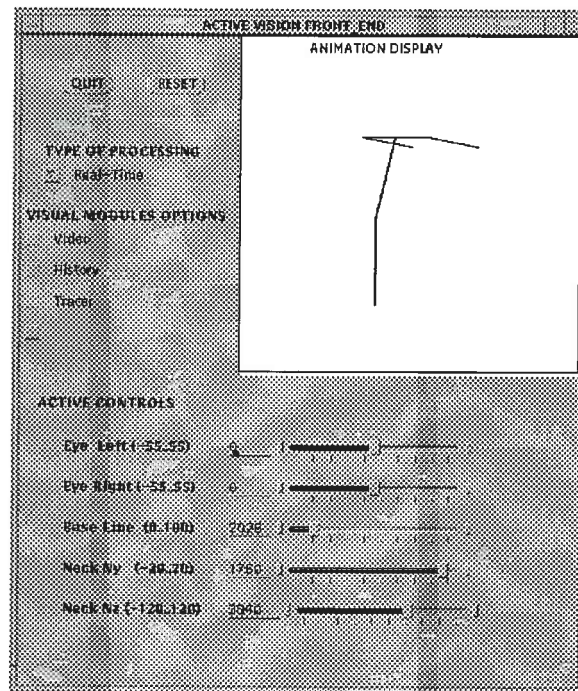


Figura 8.9a - "Interface" gráfico do programa que permite o controlo manual do sistema de visão activa.

Para o comando manual do sistema de visão activa pode ser utilizado um programa baseado em janelas e que funciona em ambiente *XWindows* para o sistema operativo *UNIX*. Este programa foi desenvolvido em *XView(Sun)* e permite realizar o controlo manual do sistema. O programa utiliza a *interface* gráfico ilustrado na figura 8.9a no qual existem vários botões de comando. O programador tem várias possibilidades de comandar o sistema como sejam o *reset* manual do sistema e o controlo manual dos vários graus de liberdade

outra placa de controlo também para funcionar em computadores com "bus-PC AT" e designado Cartão de Controlo de Motores DC (CC-DC). Os motores de passo recebem a potência necessária ao seu funcionamento através de uma outra placa designada Cartão de Fornecimento de Potência (CFP). A placa CFP é responsável pelo fornecimento de energia aos motores de passo até a um máximo de três motores por placa e é controlada pela placa CC-MP (ver figura 8.6). No caso dos motores de corrente contínua utilizados nas lentes a energia para o seu funcionamento é fornecida directamente pela placa CC-DC.

O controlo dos motores de corrente contínua é feito através da informação de posição absoluta fornecido por um potenciómetro. O controlo de posicionamento dos diferentes graus de liberdade é efectuado através da diferença entre o valor digitalizado da tensão fornecida pelo potenciómetro e o valor correspondente posição pretendida. A posição pretendida é programada através dos registos de controlo existentes na placa e acessíveis através do "bus-PC AT"

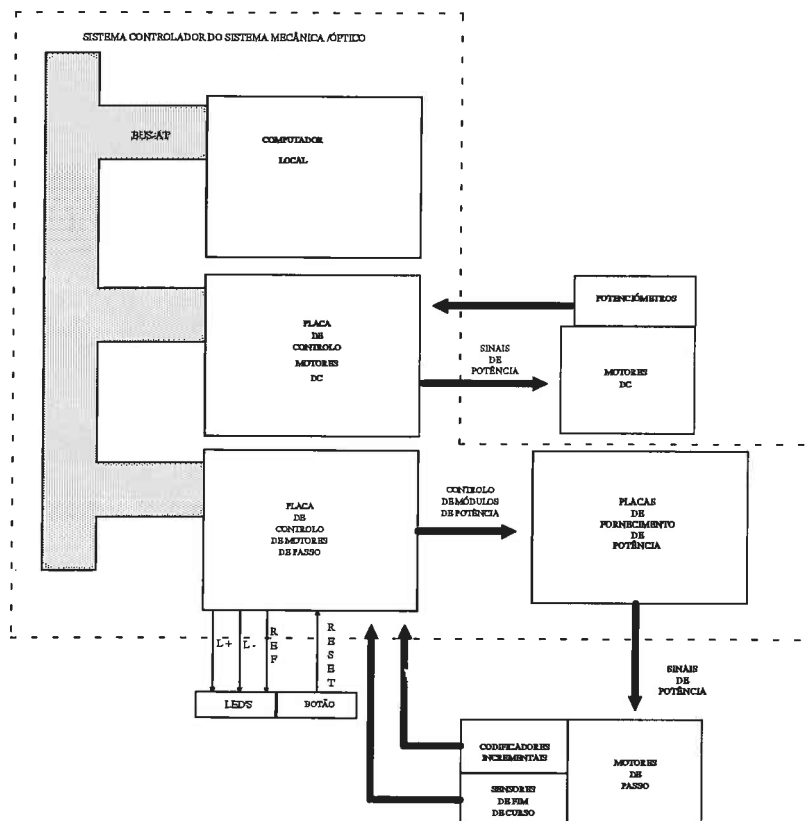


Figura 8.6 - Controlador do sistema mecânico

O conjunto do "hardware" responsável pelo controlo de todos os motores utilizados no sistema designa-se *controlador do sistema mecânico*. O sistema foi projectado de forma a funcionar modularmente, sendo possível introduzir vários cartões no sistema para controlar vários grupos de motores de passo ou de motores de corrente contínua de pequena potência. O estado dos vários graus de liberdade (*home*, *limit left*, *limit right*) pode ser visualizado através de um sinais luminosos existentes no painel frontal do *controlador do sistema*

como sejam o *reset* manual do sistema e o controlo manual dos vários graus de liberdade mecânicos. A geometria do sistema é continuamente actualizada num gráfico existente na janela. Para o controlo manual das lentes (*zoom*, *focus* e *iris*) e a aquisição das imagens o programador pode utilizar a opção *vídeo* que corresponde à figura 8.9b. Nessa opção o programador pode adquirir e armazenar imagens nas situações que desejar. Pela opção *history* o operador tem a possibilidade de visualizar, de uma forma temporal, os ângulos percorridos por cada um dos graus de liberdade - figura 8.9c. Para o caso de querer saber qual foi a troca de comandos entre o sistema e a *workstation* o programador poderá utilizar a opção *trace* que faz uma listagem dos comandos enviados - figura 8.9d.

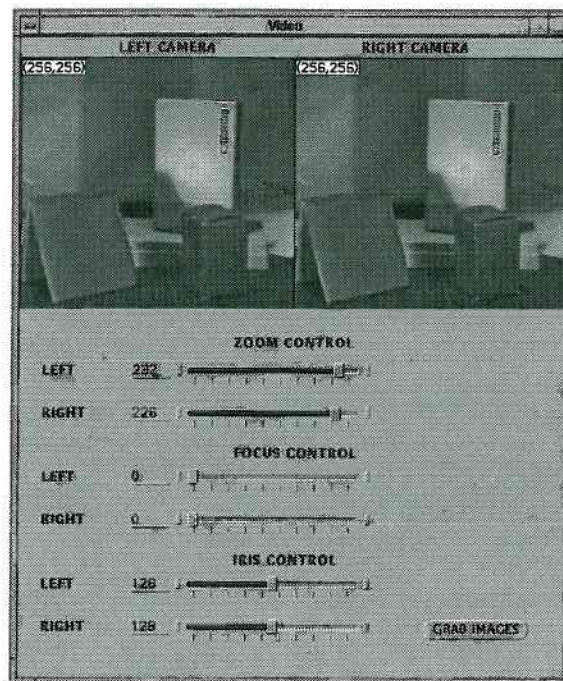


Figura 8.9b - A opção *vídeo* permite a visualização das imagens.

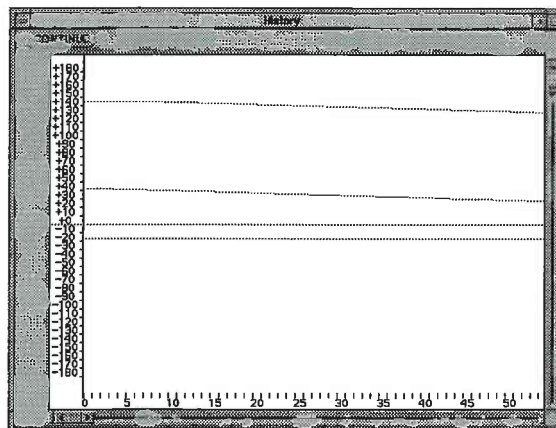


Figura 8.9c - Opção *history* permite visualizar, de uma forma temporal, os ângulos percorridos por cada um dos graus de liberdade.

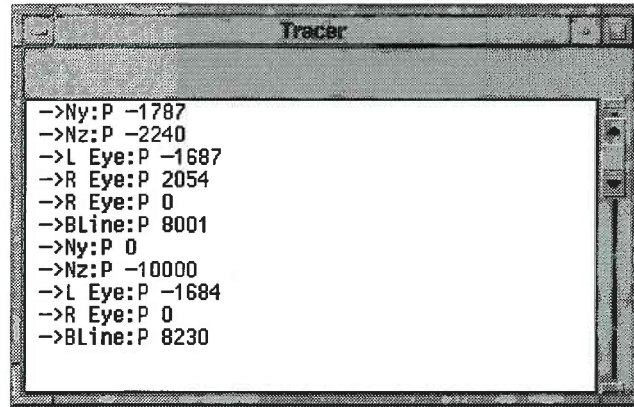


Figura 8.9d - Opção trace permite verificar quais os comando enviados para o controlador do sistema.

8.2.7 Sistemas de coordenadas

A cada um dos graus de liberdade do sistema de visão activa foi associado um sistema de coordenadas. Cada um desses referenciais pode ser relacionado com um referencial solidário com a base de suporte do sistema e designado {BASE}. A figura 8.9a ilustra quais os referenciais do sistema e quais os ângulos ou distâncias variáveis. Na tabela 8.2 estão representados os valores para cada um dos parâmetros variáveis dos sistemas de coordenadas.

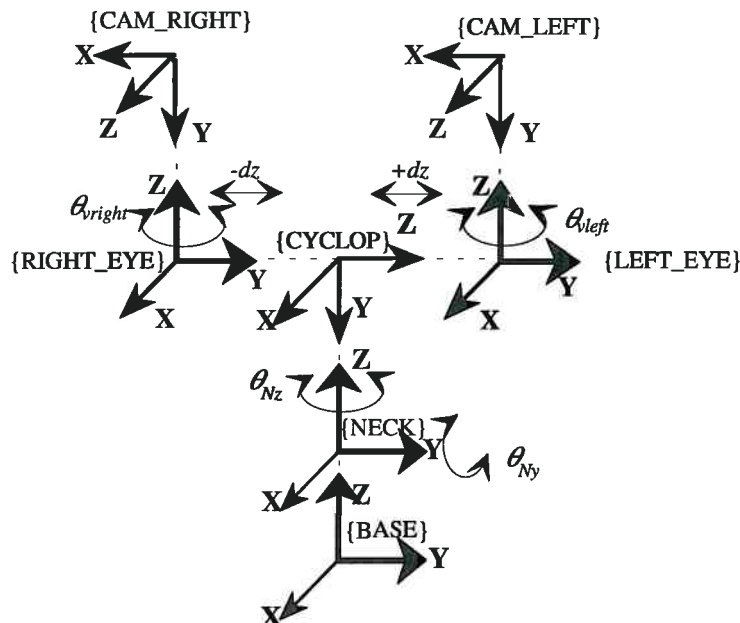


Figura 8.10 - Diagrama esquemáticos dos vários sistema de coordenadas definidos no sistema de visão activa.

Tipo de movimento	Referencial	Variável	Deslocamento máximo
<i>Vergência</i>	{RIGHT_EYE} ou {LEFT_EYE}	θ_{Vright} θ_{Vleft}	$\pm 34,2^\circ$
<i>Baseline</i>	{CYCLOP}	d_z	0–138mm
<i>Tilt</i>	{NECK}	θ_{Ny}	$\pm 25^\circ$
<i>Neck pan</i>	{NECK}	θ_{Nz}	$\pm 100^\circ$

Tabela 8.2 - Parâmetros dos motores utilizados no sistema de visão activa

8.3 Fixação de Objectos Estacionários

A fixação

O processo de controlar o sistema visual de modo a manter estável a projecção na retina de um ponto no espaço tridimensional, é conhecido como fixação. O ponto para o qual a imagem se mantém estável é designado ponto de fixação. As pesquisas realizadas neste domínio mostram que as localizações destes pontos de fixação não são escolhidos aleatoriamente nem são rigidamente planeadas com antecedência, mas são fruto do próprio processo de percepção [Yarbus 67] [Haustein 90] [Tweed 90].

A mudança de ponto de fixação é acompanhada por movimentos muito rápidos de ambos os olhos e conhecidos pelo nome de "sacadas". Este movimento de "sacada" tem uma amplitude uniforme para ambos os olhos e é do tipo "balístico" - isto é, não existe realimentação visual durante o movimento.

Quando o objecto de percepção é estacionário relativamente ao observador, o processo de percepção que tem lugar entre duas "sacadas" é conhecido por: "processo de fixação". Yarbus demonstra que durante a percepção de objectos estacionários o observador só se encontra entre dois estados: em processo de fixação ou em processo de mudança entre dois pontos de fixação [Yarbus 67]. Durante a fixação nos objectos os olhos não estão completamente parados mas realizam pequenos movimentos, mas de modo que o ponto de fixação se encontre, para cada olho, dentro da fovea. Durante este processo a abertura da pupila conserva-se praticamente constante, realizando pequenos ajustes com a mudança de intensidade luminosa ou para mudar a profundidade de campo.

Se os pontos de fixação se encontram a diferentes distâncias dos olhos do observador então o seu movimento é constituído por um movimento de "sacada" e também um movimento de convergência ou divergência dos eixos ópticos. Segundo Yarbus, um julgamento sobre as dimensões e as distâncias aos objectos é obtida através da posição relativa dos eixos ópticos e da acomodação dos olhos. A mudança entre pontos de fixação é iniciada com a convergência ou divergência e a preparação da "sacada". A convergência ou divergência ao ser iniciada antes da "sacada" permite a diminuição do tempo requerido para a mudança do sistema visual entre pontos de fixação, pois segundo Yarbus a duração dos

movimentos de convergência ou divergência têm uma velocidade dez vezes maior do que a "sacada". A figura 8.11 mostra esquematicamente o resultado do ensaio realizado por Yarbus para o estudo da vergência

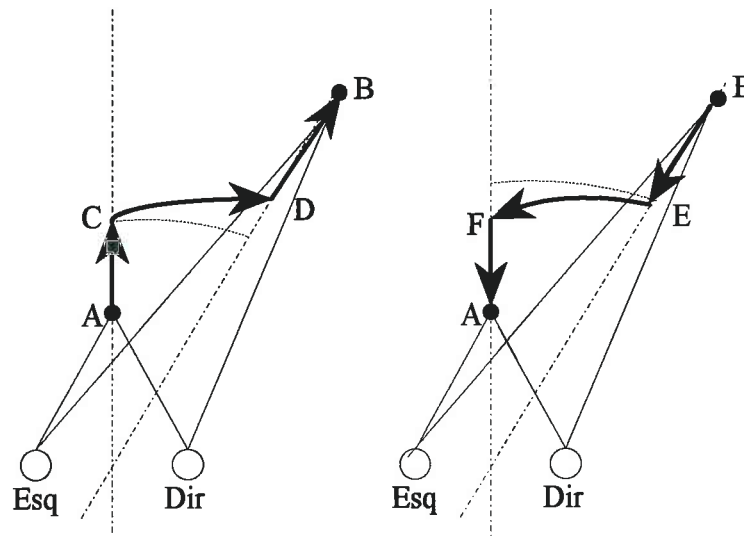


Figura 8.11 - Diagrama do movimento dos olhos durante a mudança de pontos de fixação.

Segundo estes estudos o movimento dos olhos durante a mudança do ponto de fixação A para B é composta de - ver figura 8.11: (a) divergência na porção AC; (b) "sacada" dos olhos segundo um arco que passa pelo C e D. O eixo de simetria é orientado para o ponto B; (c) divergência dos olhos até que o ponto de intersecção dos eixos ópticos se encontrem no ponto B. Durante a mudança do ponto de fixação B para A o movimento dos olhos é composto de: (a) convergência até ao ponto E; (b) "sacada" segundo um arco EF e de modo que o eixo de simetria passe por A; (c) convergência dos olhos até atingir o ponto A. Um pormenor importante nesta experiência é a execução paralela destes movimentos embora o movimento de divergência ou convergência seja iniciado primeiro e acabe em último.

No nosso trabalho consideramos o processo de fixação um assunto importante na investigação em visão activa. Este assunto foi já abordado por outros autores. É caso do controlo da vergência e focagem em imagens estereo utilizado por Krotkov [Krotkov 87] ou só da vergência, utilizado por Olson [Olson 90]. Li e Ishiguro utilizam a fixação como mecanismo de realimentação para navegação de um "robot" móvel [Li 91] [Ishiguro 90]. Nos pontos seguintes são desenvolvidas algumas expressões que demonstram as possibilidades de determinação do tempo até ao impacto e da distância do objecto ao sistema de visão utilizando fixação.

8.3.1 Tempo até colisão através de vergência

O tempo até colisão pode ser facilmente calculado através da variação do ângulo de vergência. Consideremos que o sistema visual se encontra fixo num ponto tal como mostra a figura 8.12. Na dedução das expressões pressupõe-se que o sistema é controlado de modo que o ângulo de "observação", θ_d seja 90° .

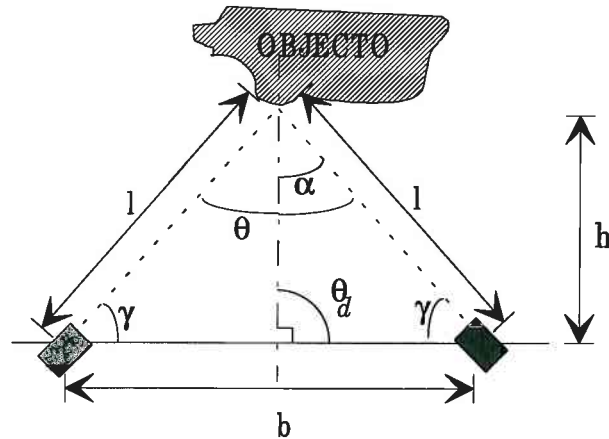


Figura 8.12 - Fixação de objecto em frente ao sistema

Pela figura obtemos

$$\begin{cases} \theta = 2\alpha \\ \alpha = 90^\circ - \gamma \end{cases} \Rightarrow \theta = 180^\circ - 2\gamma$$

e

$$\begin{cases} h = l \sin \gamma \\ \frac{b}{2} = l \cos \gamma \end{cases}$$

Estas relações permitem retirar a conclusão que

$$h = \frac{b}{2} \tan \gamma \tag{8.1}$$

Derivando a expressão acima obtemos

$$\dot{h} = \frac{b}{2} (\sec^2 \gamma) \dot{\gamma}$$

Deste modo o tempo até impacto- $t_{ai}(\gamma)$, é dado por

$$t_{ai}(\gamma) = \frac{h}{\dot{h}} = \frac{\tan \gamma}{(\sec^2 \gamma) \dot{\gamma}} = \frac{\cos \gamma \sqrt{1 - \cos^2 \gamma}}{\dot{\gamma}} \tag{8.2}$$

Conhecendo a velocidade angular $\dot{\gamma}$ de forma aproximada é possível ter uma estimativa do tempo até à colisão no objecto.

8.3.2 Distância ao ponto de fixação

Rodando as câmaras em torno de um eixo perpendicular ao plano definido pelos dois eixos ópticos é possível mudar a posição em que os dois eixos se intersectam - o ponto de fixação. Mudando a posição do ponto de fixação existem dois parâmetros envolvidos: o ângulo (de vergência) e o ângulo (de orientação) - ver figura 8.13. O ângulo de vergência é o ângulo entre os dois eixos ópticos. O ângulo de orientação é o menor ângulo entre a linha que une os dois centros ópticos ("baseline") e a linha que une o ponto de fixação e o ponto intermédio da "baseline".

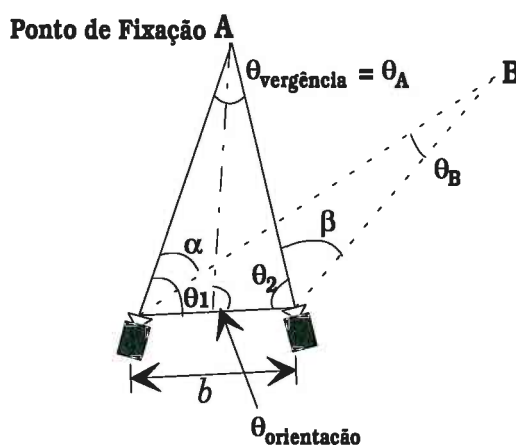


Figura 8.13 - Definição de ângulo de vergência e de versão

Vamos supor que, tal como se ilustra na figura 8.13, o sistema muda o ponto de fixação de A para B. Nesse caso é possível estabelecer uma relação simples entre os ângulos (θ_A, θ_B) e os ângulos (α, β) . Com base na geometria plana sabemos que

$$\begin{aligned} 180^\circ - \alpha - \theta_A &= 180^\circ - \beta - \theta_B \\ \alpha - \beta &= \theta_A - \theta_B \end{aligned} \quad (8.3).$$

Esta relação geométrica facilita o processo de controlo do sistema de visão para realizar uma mudança do local de fixação do sistema.

Neste tipo de configuração é possível determinar a distância a que o ponto de fixação se encontra da "baseline". A expressão para essa distância pode ser determinada pela geometria de triângulos e é dada por

$$z = b \frac{\tan \theta_1 \tan \theta_2}{\tan \theta_1 + \tan \theta_2} \quad (8.4).$$

Na figura 8.14 é ilustrada a forma como a distância z depende do valor dos ângulos $\theta_1 = \theta_2 = \theta$ (assumindo o caso simétrico). Como se pode verificar pelo gráfico, para ângulos próximos de 90° a distância z varia muito rapidamente com uma pequena variação do ângulo θ . Por outro lado, com a distância inter-ocular de 30cm ("baseline"), as distâncias entre os 2 metros e os 10 metros podem ser determinadas com ângulos próximo dos 90° . Para estas distâncias é necessário ter uma boa resolução angular nos motores para conseguir boa precisão no cálculo da distância z .

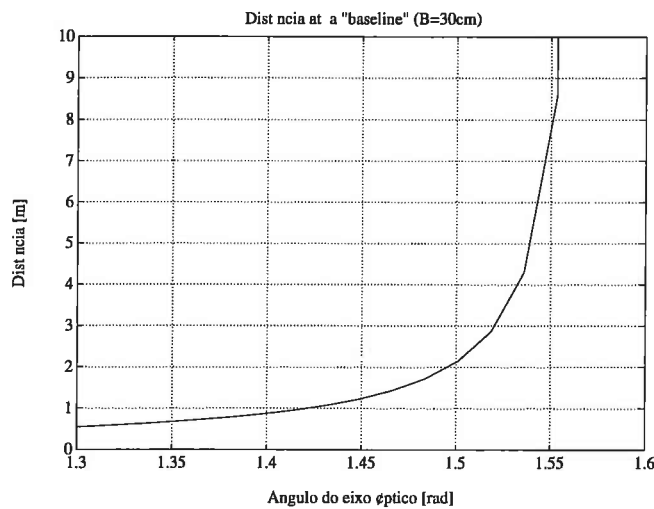


Figura 8.14 - Distância z em função do ângulo $\theta_1 = \theta_2 = \theta$ e para uma "baseline" de 30cm.

Para investigar a relação entre a resolução angular dos motores de vergência e a distância z , determinemos as derivadas parciais da equação anterior para $\theta_1 = \theta_2 = \theta$:

$$\frac{\partial z}{\partial \theta} = \frac{b}{2} (1 + \tan^2 \theta) \quad (8.5).$$

Na figura 8.15 está ilustrada a variação da incerteza relativa $\left(\frac{\partial z}{z}\right)$ em função do ângulo e da resolução angular dos motores para o caso de uma vergência simétrica. *Notar que esta incerteza relativa é independente da "baseline".*

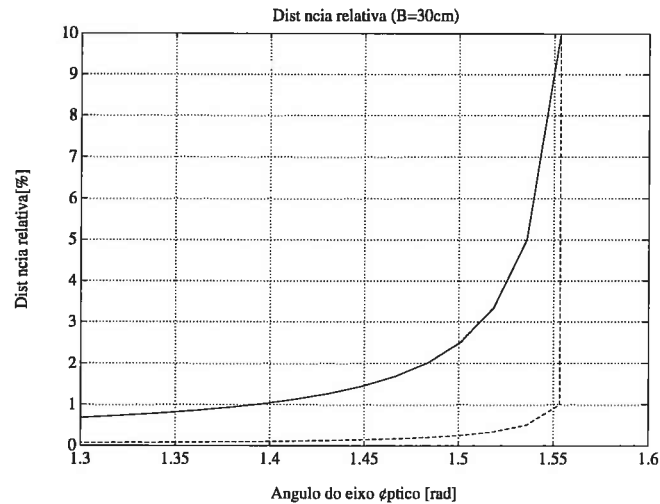


Figura 8.15 - Precisão relativa na determinação da profundidade como função da resolução angular dos motores de vergência. A linha a cheio corresponde à resolução angular de 0.1 graus enquanto a linha a tracejado corresponde à resolução angular de 0.01 graus.

8.4 Explorando as Imagens de Movimento

Como o sistema permite a movimentação das câmaras, é interessante analisar em que medida as imagens em movimento poderão ser exploradas para determinação da profundidade. Tal como foi referido no capítulo 1, o movimento nas imagens pertence ao conjunto de fontes de informação que nos permite ter uma percepção da profundidade dos objectos que nos cercam [Marr 82] [Ullman 79] [Nagel 83][Yarbus 67].

Na análise de imagens de movimento existem inúmeros trabalhos realizados e em muitos dos algoritmos propostos é possível distinguir diferentes estágios do processamento. O primeiro estágio consiste na detecção e localização de movimento. Este estágio baseia-se em métodos de detecção da variação do brilho na imagem ou, em alternativa, da correspondência de primitivas entre imagens sucessivas. O estágio seguinte consiste na extracção de medidas quantitativas sobre o movimento detectado na imagem e o terceiro estágio consiste na recuperação de informação final com base nessas medidas. Neste ponto será dada uma panorâmica breve sobre os princípios matemáticos que permitem recuperar essa informação, utilizando imagens de movimento. Para além disso, serão apresentados resultados experimentais para a recuperação de profundidade baseado no fluxo óptico normal utilizando imagens obtidas com um sistema de visão activa.

8.4.1 Introdução

O movimento das formas no plano de imagem pode ser devido ao movimento da câmara, ao movimento dos objectos que estão a ser visualizados ou a ambos. O movimento pode também ser aparente, se existe mudança do brilho da imagem que dá uma ilusão de movimento.

No estudo do movimento com imagens é muito comum encontrar dois tipos de designações: *campo de movimento* e *fluxo óptico* [Horn 87]. A projecção da velocidade/deslocamento tridimensional no plano imagem designa-se por *campo de movimento*. Se forem conhecidos os valores deste *campo de movimento* então, os parâmetros do movimento tridimensional que o geraram $\mathbf{v}_t = (v_x, v_y, v_z)$ e $\mathbf{\omega} = (\omega_x, \omega_y, \omega_z)$ ou a estrutura tridimensional da cena, podem vir a ser determinados [Horn 87] [Ullman 79].

Para introduzir a noção de *fluxo óptico*, consideremos imagens de objectos em movimento. Nessas imagens podem ser detectadas regiões correspondentes a objectos que se movem. Ao movimento aparente dessas regiões designa-se *fluxo óptico*. Idealmente o *fluxo óptico* deveria ser igual ao *campo de movimento*, mas nem sempre isso se verifica [Horn 87][Verri 87].

O *fluxo óptico* é a informação que pode ser retirada a partir das imagens de movimento e como tal, pode ser encarado como uma forma intermédia para a sua representação. Sendo o *fluxo óptico* igual ao *campo de movimento* então, toda a análise feita com base no *campo de movimento* poderá ser aplicada ao *fluxo óptico*, o qual poderá ser utilizado para:

- Segmentação da imagem nas regiões que pertencem a objectos movendo-se independentemente.
- Navegação passiva através da recuperação do movimento relativo entre imagens captadas de cada objecto em cena.
- Determinação da tridimensionalidade através da recuperação das coordenadas tridimensionais dos objectos ou dar uma noção de profundidade relativa dos objectos.

Na análise de imagens em movimento têm sido explorados métodos baseados na análise das diferenças de intensidade nas imagens obtidas em instantes de tempo diferentes [Horn 81] [Nagel 83][Nagel 86] [Horn 87] [Murray 90] e métodos baseados na determinação prévia de primitivas, aplicando posteriormente técnicas para estabelecer a sua correspondência entre imagens sucessivas ("tracking") [Faugeras 88] [Lustman 87] [Ayache 88a] [Navab 90].

O método baseado na análise das diferenças gera um resultado denso de informação (quase toda a imagem). Para a sua determinação têm sido propostos vários algoritmos que

se apoiam em diferentes pressupostos e na existência de um conjunto de relações entre o movimento real dos objectos e as suas projecções nas imagens. Ou seja, que se conhece alguma relação entre o *campo de velocidade* e o *fluxo óptico*. Quando estes pressupostos não se verificam, os algoritmos apresentam instabilidade nos resultados [Verri 87].

Como existem situações em que, o *campo de velocidade* é diferente do *fluxo óptico*, ou em que algum dos pressupostos utilizados não se verificam, o *fluxo óptico* servirá melhor como fonte de informação qualitativa, acerca da estrutura tridimensional.

O método com base na correspondência de primitivas entre imagens sucessivas ("tracking") não oferece campos tão densos de informação como no caso anterior. No entanto, após estabelecer a correspondência entre primitivas é relativamente fácil obter uma estimativa para a estrutura tridimensional da cena com base na geometria descrita no capítulo 6. Apesar desta vantagem, esta aproximação necessita da determinação prévia das primitivas na imagem, antes de iniciar o processo que determina a correspondência.

Ambas as aproximações têm algum suporte nos estudos de psicologia que indicam que existem dois processos de estimação de movimento no sistema visual humano: um processo global, que determina a correspondência entre primitivas de imagens sucessivas e um local, processando localmente as imagens, mas não exigindo um pré-processamento das imagens.

Campo de Movimento

O *campo de movimento* corresponde à projecção do movimento na imagem e idealmente seria equivalente ao *fluxo óptico* medido na imagem. Por isso vamos assumir que as nossas câmaras seguem a projecção em perspectiva ilustrada na figura 2.11. Neste caso um ponto tridimensional $\mathbf{P} = (x, y, z)$ terá uma imagem num ponto $\mathbf{p} = (u, v, f)$ que estão relacionados pela expressão

$$\mathbf{p} = \frac{f}{z} \mathbf{P} = \frac{f}{\mathbf{P} \hat{\mathbf{k}}} \mathbf{P} \quad (8.6)$$

onde $z = \mathbf{P} \hat{\mathbf{k}}$ é conhecida por *profundidade* do ponto \mathbf{P} .

A velocidade de um ponto \mathbf{p} na imagem pode ser obtida derivando a expressão acima em relação ao tempo para \mathbf{p} e \mathbf{P} .

$$\frac{d\mathbf{p}}{dt} = \frac{d}{dt} \left(\frac{f}{\mathbf{P} \hat{\mathbf{k}}} \mathbf{P} \right) \quad (8.7).$$

Desenvolvendo esta expressão mas utilizando a notação (') para designar a derivada, teremos

$$\mathbf{p}' = \frac{f}{\mathbf{P}\hat{\mathbf{k}}} \mathbf{P}' - \frac{f}{(\mathbf{P}\hat{\mathbf{k}})^2} (\mathbf{P}\hat{\mathbf{k}})' \mathbf{P} \quad (8.8)$$

que expressa numa forma mais explícita, corresponde à expressão

$$\begin{bmatrix} u' \\ v' \end{bmatrix} = \frac{f}{Z} \begin{bmatrix} X' \\ Y' \end{bmatrix} - \frac{f}{Z^2} Z' \begin{bmatrix} X \\ Y \end{bmatrix} \quad (8.9)$$

desprezando a última linha da expressão (8.8).

No caso das câmaras se moverem relativamente a uma cena fixa com velocidade de translação \mathbf{v}_{Trans} e velocidade de rotação $\boldsymbol{\omega}$, a velocidade de um ponto na cena relativamente ao observador será dado por

$$\mathbf{P}' = -\mathbf{v}_{Trans} - \boldsymbol{\omega} \wedge \mathbf{P}$$

ou numa forma matricial mais explícita corresponde à expressão

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = - \begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix} - \begin{bmatrix} \omega_y Z - \omega_z Y \\ \omega_z X - \omega_x Z \\ \omega_x Y - \omega_y X \end{bmatrix}$$

Substituindo X' e Y' em (8.9) obtemos a expressão

$$\begin{bmatrix} u' \\ v' \end{bmatrix} = \frac{f}{Z} \begin{bmatrix} -v_x + \omega_z Y - \omega_y Z \\ -v_y + \omega_x Z - \omega_z X \end{bmatrix} + \frac{f}{Z^2} (v_z + \omega_x Y - \omega_y X) \begin{bmatrix} X \\ Y \end{bmatrix}$$

ou seja, rearranjando os termos e utilizando as expressões que traduzem o efeito de perspectiva $u = f X/Z$ e $v = f Y/Z$ obtemos a expressão

$$\begin{bmatrix} u' \\ v' \end{bmatrix} = \frac{1}{Z} \begin{bmatrix} -f & 0 \\ 0 & -f \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} + \begin{bmatrix} \frac{vu}{f} & -\left(f + \frac{u^2}{f}\right) \\ \left(f + \frac{v^2}{f}\right) & -\frac{uv}{f} \end{bmatrix} \begin{bmatrix} v \\ -u \end{bmatrix} \begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix} \quad (8.10).$$

Nesta expressão torna-se mais evidente que o valor da profundidade (Z) a que se encontram os pontos, pode ser determinado pelo conhecimento da primeira derivada da projecção do ponto e do movimento das câmaras.

Como os valores instantâneos da velocidade na imagem podem ser aproximados por valores $(\Delta u, \Delta v)$ correspondentes à diferença entre pontos na imagem (u, v) para intervalos de tempo muito pequenos, passando-se o mesmo para a velocidade tridimensional $(v_x, v_y, v_z, \omega_x, \omega_y, \omega_z)$ que se pode aproximar pela variação da posição e orientação da câmara $(\Delta X, \Delta Y, \Delta Z, \Delta \theta_x, \Delta \theta_y, \Delta \theta_z)$. Nesse caso uma expressão baseada nessas variações é dada por

$$\begin{bmatrix} \Delta u \\ \Delta v \end{bmatrix} = \frac{1}{Z} \begin{bmatrix} -f & 0 & u \\ 0 & -f & v \end{bmatrix} \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix} + \begin{bmatrix} \frac{vu}{f} & -\left(f + \frac{u^2}{f}\right) & v \\ \left(f + \frac{v^2}{f}\right) & -\frac{vu}{f} & u \end{bmatrix} \begin{bmatrix} \Delta \theta_x \\ \Delta \theta_y \\ \Delta \theta_z \end{bmatrix} \quad (8.11)$$

se a translação e a rotação da câmara por instante de tempo for conhecida.

A figura 8.16 mostra o comportamento de um ponto tridimensional quando é sujeito a uma translação e a sua correspondente projecção na imagem

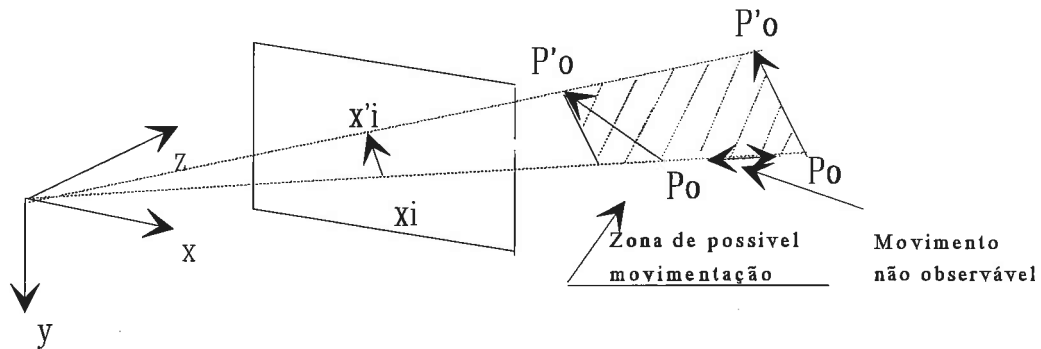


Figura 8.16 - A figura ilustra a forma como um ponto no espaço se projecta numa imagem quando a câmara está sujeita a um movimento de translação. O movimento projectado (x_i, x_i') poderá ser originado pelo movimento do ponto entre a recta projectante de x_i e a recta projectante de x_i' .

Como se pode concluir pela imagem, o conhecimento da projecção do movimento não é suficiente para a determinação da tridimensionalidade utilizando um só ponto.

8.4.2 Efeito do Movimento da Câmara

Como o nosso sistema de visão activa permite movimentar as câmaras, é importante analisar os seus efeitos na imagem. Se o movimento da câmara for conhecido então um

ponto numa cena descreve trajectórias que projectadas no plano imagem são locais perfeitamente conhecidos. Tomando como referência um ponto $\mathbf{P}=(x,y,z)$ que se projecta, segundo um modelo de perspectiva, num ponto (u,v) no plano imagem, é possível verificar o efeito do movimento da câmara nas imagens em sequência.

Efeito de Rotação

Quando uma câmara é rodada em torno do centro da sua lente existe mudança do local onde os pontos se encontram no plano de imagem. O efeito de rotação pode ser então descrito por uma função que transforma um plano noutro. No caso da rotação da imagem segundo o eixo óptico Z, qualquer ponto na imagem move-se segundo um círculo em torno do centro da imagem.

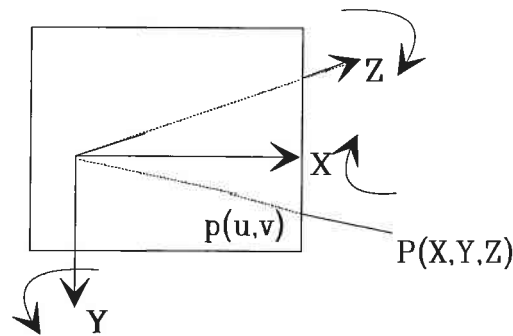


Figura 8.17 - A rotação da câmara é feita segundo os eixos X, Y e Z

Rodando a câmara segundo uma rotação pura \mathbf{Rot} é equivalente ao deslocamento de um ponto \mathbf{P} para um ponto \mathbf{P}^* , que estão relacionados pela expressão

$$\mathbf{P}^* = \mathbf{Rot} \mathbf{P} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} X^* \\ Y^* \\ Z^* \end{bmatrix} \tag{8.12}.$$

Projectando o ponto \mathbf{P}^* no plano imagem obtemos o ponto $\mathbf{p}^*=(u^*,v^*)$ que se relaciona com o ponto tridimensional através de

$$\begin{cases} u^* = f \frac{X^*}{Z^*} = f \frac{\mathbf{r}_1 [X \ Y \ Z]^T}{\mathbf{r}_3 [X \ Y \ Z]^T} \\ v^* = f \frac{Y^*}{Z^*} = f \frac{\mathbf{r}_2 [X \ Y \ Z]^T}{\mathbf{r}_3 [X \ Y \ Z]^T} \end{cases} \tag{8.13}$$

sendo r_i a linha i da matriz de rotação. Relacionando as coordenadas X,Y e Z do ponto **P** com a sua projecção $\mathbf{p}=(u,v)$ antes da rotação temos

$$\left\{ \begin{array}{l} u^* = f \frac{\mathbf{r}_1 \begin{bmatrix} \frac{uZ}{f} & \frac{vZ}{f} & Z \end{bmatrix}^T}{\mathbf{r}_3 \begin{bmatrix} \frac{uZ}{f} & \frac{vZ}{f} & Z \end{bmatrix}^T} = f \frac{\mathbf{r}_1 [u \ v \ f]^T}{\mathbf{r}_3 [u \ v \ f]^T} \\ v^* = f \frac{\mathbf{r}_2 \begin{bmatrix} \frac{uZ}{f} & \frac{vZ}{f} & Z \end{bmatrix}^T}{\mathbf{r}_3 \begin{bmatrix} \frac{uZ}{f} & \frac{vZ}{f} & Z \end{bmatrix}^T} = f \frac{\mathbf{r}_2 [u \ v \ f]^T}{\mathbf{r}_3 [u \ v \ f]^T} \end{array} \right. \quad (8.14)$$

A expressão (8.14) dá uma relação entre coordenadas antes e após o movimento de rotação. Se a rotação for segundo o eixo-x e de um ângulo θ , seguida de uma rotação segundo o eixo-y de um ângulo φ então

$$Rot(X, \theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{bmatrix} \quad Rot(Y, \varphi) = \begin{bmatrix} \cos \varphi & 0 & \sin \varphi \\ 0 & 1 & 0 \\ -\sin \varphi & 0 & \cos \varphi \end{bmatrix}$$

O resultado total das duas rotações será

$$\mathbf{Rot} = \begin{bmatrix} \cos \varphi & 0 & \sin \varphi \\ \sin \varphi \sin \theta & \cos \theta & -\sin \theta \cos \varphi \\ -\cos \theta \sin \varphi & \sin \theta & \cos \theta \cos \varphi \end{bmatrix}$$

e então os pontos obedecerão à expressão

$$\begin{bmatrix} u' \\ v' \end{bmatrix} = \frac{f}{-\cos \theta \sin \varphi u + \sin \theta v + f \cos \theta \cos \varphi} \begin{bmatrix} \cos \varphi u + f \sin \varphi \\ \sin \theta \sin \varphi u + \cos \theta v - \sin \theta \cos \varphi \end{bmatrix} \quad (8.15).$$

Para o caso de uma rotação ψ segundo o eixo Z (eixo óptico)

$$Rot(Z, \psi) = \begin{bmatrix} \cos \psi & \sin \psi & 0 \\ -\sin \psi & \cos \psi & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

obtemos a expressão

$$\begin{bmatrix} u' \\ v' \end{bmatrix} = \begin{bmatrix} u \cos \psi + v \sin \psi \\ -u \sin \psi + v \cos \psi \end{bmatrix} \quad (8.16)$$

que dá a relação antes e após rotação quando a câmara está sujeita ao movimento segundo o eixo-z. As figura seguintes dão alguns exemplos do efeito de rotação segundo o eixo X, Y e Z.

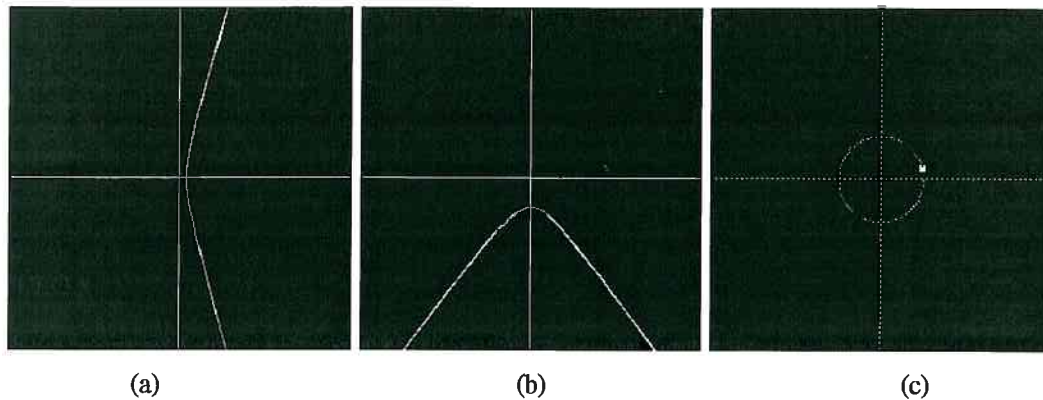


Figura 8.18 - A figura exemplifica o efeito na imagem de diferentes tipos de rotação na imagem. Em (a) está representado o efeito da rotação segundo X para um ponto, em (b) é para uma rotação do eixo Y e em (c) uma rotação segundo Z para 4 pontos.

Efeito de Translação

Quando a câmara se movimenta segundo uma translação pura **Trans** um ponto P_i na cena movimenta-se relativamente à câmara (ou veículo móvel) segundo uma translação com sentido contrário. Como cada ponto da cena se movimenta relativamente à câmara com a mesma translação, o efeito desse movimento são trajectórias paralelas que são projectadas na imagem. Pela geometria em perspectiva esse conjunto de trajectórias projectadas nas imagens passam por um único ponto no plano imagem chamado ponto de contracção. Quando a câmara se movimenta no sentido positivo do eixo-z e segundo uma trajectória de translação, os pontos nas imagens parecem divergir deste ponto de dispersão a que se chama "Foco De Expansão"-FDE ou convergir quando a câmara recua relativamente aos pontos e neste caso tem o nome de "Foco De Contracção"-FDC. A figura 8.19 ilustra o efeito da projecção das trajectórias quando a câmara realiza um movimento de translação.

Como se pode ver pela figura, a linha recta que passa pelo centro da lente C e o $FDE = (u_F, v_F)$ é paralelo aos vectores de deslocamento tridimensionais. Tendo a câmara calibrada, a direcção da translação da câmara pode ser determinada localizando o FDE na

imagem. Se F for a localização do FDE na imagem, então o vector \overline{CF} fornece a direcção da translação mas não a sua amplitude.

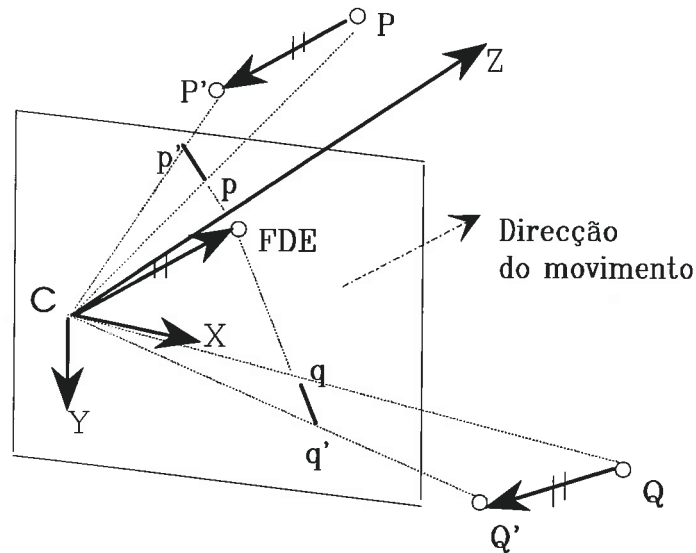


Figura 8.19 - Localização do foco de expansão (FDE). No caso de uma translação pura da câmara, os pontos P e Q da cena movem-se ao longo de paralelas e com sentido idêntico ao vector que une o centro do referencial tridimensional da câmara ao FDE no plano da imagem.

Assim uma expressão para a translação é

$$\mathbf{Trans} = \lambda \overline{CF} = \lambda \begin{bmatrix} u_F \\ v_F \\ f \end{bmatrix} \quad \lambda \in \mathbb{R} \quad (8.17)$$

em que λ é um parâmetro dependendo da distância deslocada após a normalização do vector \overline{CF} . O inverso deste processo pode ser aplicado e determinar o FDE (FDC). Neste caso a translação é conhecida, a rotação é nula e a equação (8.10) toma a forma

$$\begin{bmatrix} u' \\ v' \end{bmatrix} = \frac{1}{Z} \begin{bmatrix} -f & 0 & u \\ 0 & -f & v \end{bmatrix} \begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix} \quad (8.18)$$

Para este movimento as imagens do fluxo óptico serão totalmente radiais e com centro no FDE(FDC). Neste ponto os valores para o fluxo óptico serão nulos, ou seja as derivadas $(u',v')=(0,0)$ para o foco FDE(FDC). Pela equação (8.18) o local na imagem correspondente ao FDE será então dado por

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \frac{1}{Z} \begin{bmatrix} -f & 0 & u \\ 0 & -f & v \end{bmatrix} \begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix}$$

correspondendo à posição

$$\begin{cases} u_0 = f \frac{-v_x}{v_z} \\ v_0 = f \frac{-v_y}{v_z} \end{cases} \quad (8.19)$$

ou seja: conhecida a velocidade translacional da câmara, o FDE está à partida definido. Caso v_z seja nulo então o local do FDE será no infinito. A partir da expressão (8.19) também se conclui que a localização do FDE só se altera quando a direcção do movimento de translação também se altera.

Invariância do Ponto de Fuga de Uma Recta

Consideremos uma recta tridimensional que é visualizada por uma câmara em movimento de translação. Se a câmara tiver um movimento de translação dado por $-\mathbf{v}_t$ então a recta mover-se-á relativamente à imagem segundo trajectórias definidas tridimensionalmente através de

$$\mathbf{P}(t) = \mathbf{P}_0 + \lambda \mathbf{n} + \mathbf{v}_t \quad \text{com } \lambda \in]-\infty, +\infty[$$

em que \mathbf{P}_0 é um ponto da recta e \mathbf{n} o vector director da recta.

A imagem de perspectiva desta linha é dada por

$$\begin{bmatrix} u(\lambda) \\ v(\lambda) \end{bmatrix} = \frac{f}{z_0 + \lambda n_z + v_z t} \begin{bmatrix} x_0 + \lambda n_x + v_x t \\ y_0 + \lambda n_y + v_y t \end{bmatrix} \quad (8.20)$$

o que corresponde a um ponto de fuga na imagem dado pelo ponto (u, v) quando $\lambda \rightarrow \infty$

$$\begin{bmatrix} u(\infty) \\ v(\infty) \end{bmatrix} = \frac{f}{n_z} \begin{bmatrix} n_x \\ n_y \end{bmatrix} \quad (8.21)$$

Por esta equação verificamos que o ponto de fuga não varia com o tempo nem com a direcção e amplitude do movimento. Este ponto é *invariante* ao movimento de translação e pode ser utilizado como restrição para estabelecer correspondência entre rectas no caso de imagens de movimento em translação. Note-se que este não é o caso do FDE (FDC) que varia com a direcção do movimento.

8.4.3 Determinação da Profundidade

Se for dada uma estimaco, mesmo que aproximada, para o campo de movimento ou algum dos seus parâmetros ento, ser possvel determinar alguma informao final sobre algumas caractersticas da cena.

Utilizando o FDC/FDE

Como se pode ver pela expresso (8.10) o valor da profundidade Z a que se encontra um ponto no espao relaciona-se com a velocidade de translao. Se a velocidade de rotao da cmara for nula e se for conhecido o ponto de contraco/expanso da imagem (FDE/FDC), ento  possvel determinar a profundidade Z a que se encontra um ponto se for conhecido o campo de movimento nesse ponto.

Vamos supor que a cmara realiza uma translao pura. Nesse caso a trajectria da cmara provoca um efeito na imagem equivalente a uma linha recta que passa pela projeco do ponto original e o ponto correspondente ao FDE ou seja, ser o conjunto de pontos \mathbf{p}^* na imagem que obedecem  condio

$$\mathbf{p}^*_i = \mathbf{p}_i + \mu(\mathbf{p}_i - \overline{FDE}), \quad \text{com } \mu_i \in \mathbb{R} \text{ e } \mu_i \geq 0 \quad (8.22).$$

Para exemplificar como a determinao do FDE pode ajudar na determinao da estrutura tridimensional, tomemos as projeces de um ponto \mathbf{P} numa cmara da qual s utilizamos o plano X-Z (os resultados s idênticos para o plano Y-Z). Aps uma translao **Trans** da cmara o ponto estar relativamente  cmara, em \mathbf{P}^* . Vamos tambm assumir que o foco de expanso (FDE) se encontra nas coordenadas (u_0, v_0) .

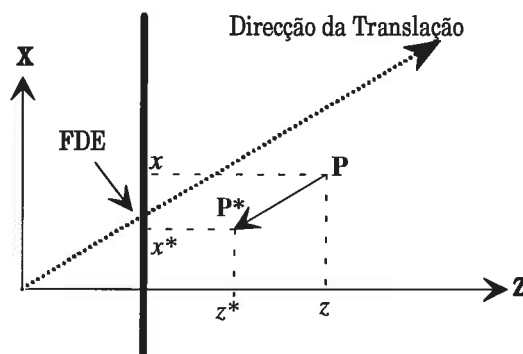


Figura 8.20 - Rebatimento no plano X-Z de um movimento de translao puro da cmara. O ponto correspondente ao foco de expanso (FDE) tem coordenadas (u_0, v_0) .

Utilizando as coordenadas tridimensionais para \mathbf{P} e \mathbf{P}^* mas desprezando a coordenada Y , obtemos uma relação entre os pontos antes e após a translação. Essa relação é dada por

$$\mathbf{P} = \begin{bmatrix} X \\ Y \end{bmatrix} \quad \mathbf{P}^* = \begin{bmatrix} X^* \\ Y^* \end{bmatrix} = \begin{bmatrix} X - \Delta X \\ Y - \Delta Y \end{bmatrix}$$

em que ΔX e ΔZ correspondem à translação **Trans**. Utilizando as equações de perspectiva teremos:

$$Z = \frac{f}{u} X \quad (8.23a)$$

e

$$Z^* = Z - \Delta Z = \frac{f}{u^*} X^* \quad (8.23b).$$

Por semelhança de triângulos obtemos uma relação para ΔX e ΔZ

$$\frac{\Delta X}{\Delta Z} = \frac{u_o}{f} \Rightarrow \frac{\Delta X}{u_o} = \frac{\Delta Z}{f}.$$

Substituindo esta última expressão na equação (8.23b) e após manipulação algébrica obtemos a seguinte relação para a profundidade Z em função da posição do FDE.

$$\frac{Z}{\Delta Z} = \frac{u^* - u_o}{u^* - u} \quad (8.24).$$

Deste modo a razão de expansão de pontos da imagem segundo o eixo X da imagem (ou eixo- y) a partir do FDE, contém informação directa acerca da distância dos pontos correspondentes 3D da câmara. Assim, se a câmara se deslocar em translação pura e se o FDE for localizado então a estrutura tridimensional pode ser determinada através da projecção dos pontos com a câmara em movimento. A distância Z de um ponto P pode ser obtido relativo a um factor de escala ΔZ , e correspondente à distância que a câmara avançou ao longo do eixo Z no intervalo de tempo entre duas imagens. Esse valor corresponde ao valor da coordenada Z , da transformação **Trans**. Deste modo é necessário que o nosso sistema de visão tenha um movimento translacional com uma componente na coordenada Z para que seja possível aplicar este método para o cálculo da profundidade. Com um sistema de visão activa tal movimento não é possível realizar a menos que o sistema esteja colocado no "robot" móvel. Uma outra alternativa que não envolve translação segundo o eixo- z , é a

técnica de determinação da profundidade utilizando o fluxo óptico normal que é descrita no ponto seguinte.

Utilizando fluxo óptico normal ("normal flow")

Através do fluxo óptico normal é possível ter uma noção de profundidade relativa. Para isso vamos supor que representamos por $I(u, v, t)$, o valor do nível de cinzento da imagem de um ponto P no tempo t . Vamos assumir também que, para pequenos intervalos de tempo Δt , o valor do nível de cinzento de P no tempo $t + \Delta t$ se mantém ou seja que

$$I(u + u' \Delta t, v + v' \Delta t, t + \Delta t) = I(u, v, t) \quad (8.25).$$

A expressão (8.25) traduz uma restrição que nos poderá ser muito útil para a determinação do *fluxo óptico* (u', v'). Assumindo que o campo de movimento é contínuo e que a função $I()$ tem derivadas no ponto $I(u, v, t)$, então é possível obter uma aproximação local para função através da série de Taylor da expressão (8.25)

$$I(u, v, t) = I(u, v, t) + \frac{\partial I}{\partial u} \Delta u + \frac{\partial I}{\partial v} \Delta v + \frac{\partial I}{\partial t} \Delta t + e \quad (8.26)$$

em que e representa os termos de ordem mais alta para Δt , Δu e Δv . Cancelando $I(u, v, t)$, dividindo a expressão por Δt e tomando o limite quando $\Delta t \rightarrow 0$ obtemos a seguinte igualdade

$$\frac{\partial I}{\partial u} \frac{du}{dt} + \frac{\partial I}{\partial v} \frac{dv}{dt} + \frac{\partial I}{\partial t} = 0 \quad (8.27).$$

que é correspondente à expansão da derivada total $dI/dt = 0$. Se utilizarmos as abreviações

$$u' = \frac{du}{dt}, v' = \frac{dv}{dt}, I_u = \frac{\partial I}{\partial u}, I_v = \frac{\partial I}{\partial v}$$

a expressão (8.27) pode ser escrita na forma

$$I_u u' + I_v v' + I_t = 0 \quad (8.28)$$

que representa a restrição para o qual o fluxo óptico deverá estar sujeito, se o pressuposto descrito acima e expresso por (8.25) for verdadeiro. Note-se que os valores I_u, I_v e I_t podem

ser medidos directamente na imagem e como tal, esta restrição poderá ser utilizada na construção de algoritmos para determinar o *fluxo óptico* [Horn 81].

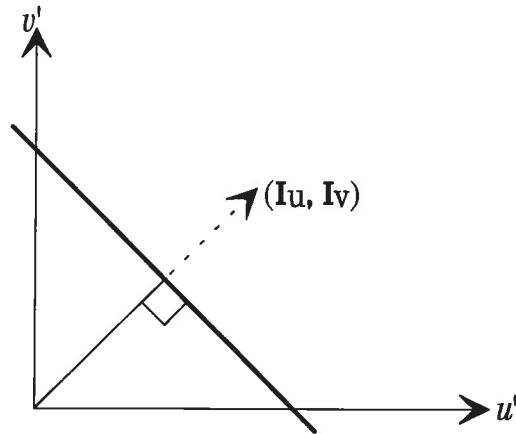


Figura 8.21 - Espaço velocidade. A recta representa o lugar geométrico correspondente à restrição (8.28).

No entanto o fluxo óptico calculado desta maneira apresenta o denominado *problema da abertura*. Para descrever esse problema considere-se um espaço a duas dimensões, no qual os dois eixos são as coordenadas u' e v' - figura 8.21. Este espaço poderá ser designado por *espaço velocidade* e nesse caso a expressão (8.28) é uma linha recta nesse espaço. Essa recta será o lugar geométrico de todos os valores possíveis para o fluxo óptico (u', v') e consiste na informação que podemos retirar através das medidas realizadas nas imagens.

No entanto, re-escrevendo a expressão (8.28) numa forma vectorial

$$[I_u \quad I_v]^T \begin{bmatrix} u' \\ v' \end{bmatrix} = -I_t \quad (8.29)$$

verificamos que utilizando unicamente a restrição (8.25), só conseguimos calcular a projecção do fluxo óptico segundo a direcção do gradiente $[I_u \quad I_v]^T$. Exemplificando graficamente através da figura 8.22, se um contorno na imagem (curva de irradiação constante) se move para uma nova posição, não sabemos exactamente para que locais se movem os pontos pertencentes a esse contorno. Esse é o designado *problema da abertura*. Na realidade, a única informação correcta disponível é a projecção do fluxo óptico na direcção do gradiente e que é normal à curva do contorno. Ou seja, utilizando só esta restrição, podemos calcular o módulo do *fluxo óptico normal* ("normal flow") que nesse caso será dado por

$$v_n = -\frac{I_t}{\sqrt{I_u^2 + I_v^2}} = -\frac{I_t}{\|\nabla \mathbf{I}\|} \quad (8.30).$$

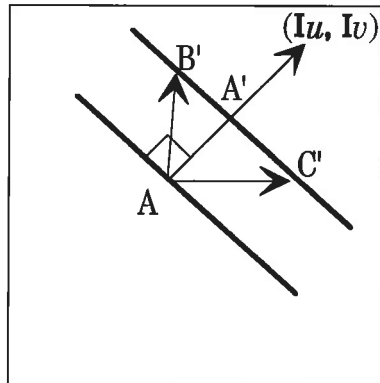


Figura 8.22 - O problema da abertura. O ponto A poder-se-á ter movido para A', B' ou C' mas seja qual for o movimento, a sua projecção segundo a direcção do gradiente em A será sempre AA'.

O cálculo do fluxo óptico normal é relativamente fácil de obter tal como se pode verificar pela equação (8.30). Mas qual é a relação do *fluxo óptico normal* com a projecção do movimento tridimensional na imagem - o *campo de movimento*? Para analisar esta questão vamos considerar o fluxo óptico $\mathbf{v} = (u', v')^T$ num ponto (u, v) em que o gradiente local normalizado é

$$\mathbf{n} = \frac{1}{\sqrt{I_u^2 + I_v^2}} \begin{bmatrix} I_u \\ I_v \end{bmatrix} = \frac{\nabla \mathbf{I}}{\|\nabla \mathbf{I}\|}.$$

O campo de movimento normal ("normal motion field") nesse ponto é

$$u_n = \mathbf{n}^T \mathbf{v} = \frac{\nabla \mathbf{I}}{\|\nabla \mathbf{I}\|} \begin{bmatrix} u' \\ v' \end{bmatrix} = \frac{1}{\|\nabla \mathbf{I}\|} (I_u u' + I_v v') \quad (8.31).$$

Se o campo de movimento normal for igual ao fluxo óptico normal então a subtracção das expressões (8.31) e (8.30) deverá dar um resultado nulo ou seja

$$u_n - v_n = \frac{1}{\|\nabla \mathbf{I}\|} (I_u u' + I_v v' + I_t) = \frac{1}{\|\nabla \mathbf{I}\|} \frac{d\mathbf{I}}{dt}.$$

Desta expressão podemos concluir que o *campo de movimento* e o *fluxo óptico normal* serão aproximadamente iguais para locais em que a variação da intensidade na imagem dI/dt é suficientemente pequena e o gradiente local tem uma grande amplitude. Ou seja, o *fluxo óptico normal* deverá ser medido em regiões com grande amplitude de gradiente. Isso acrescenta um pormenor prático ao cálculo da expressão (8.30).

Mapa de profundidade relativo

Como poderá o fluxo normal ser útil no cálculo da profundidade? Para isso vamos considerar duas imagens obtidas por uma câmara em movimento que realiza um processo de fixação, mas que conserva o seu ponto central fixo num ponto da cena, tal como mostra a figura 8.23. As imagens são obtidas com uma diferença de tempo Δt muito pequena e as velocidades de rotação segundo o eixo-x e o eixo-z são nulas ($\omega_x = 0, \omega_z = 0$) assim como a velocidade de translação segundo o eixo-y ($v_y = 0$).

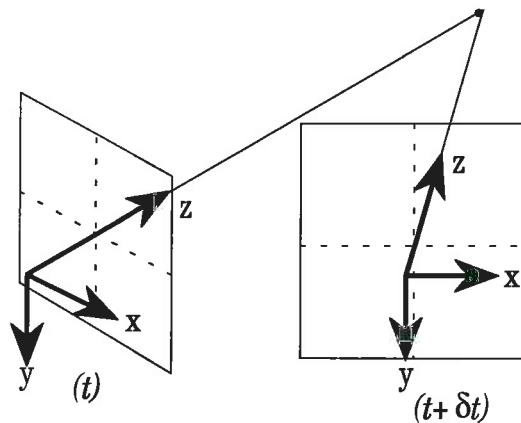


Figura 8.23 - Geometria do movimento das câmaras quando fixas num ponto no espaço. As imagens são obtidas com uma diferença de tempo Δt muito pequena e as velocidades de rotação segundo o eixo-x e o eixo-z são nulas assim como a velocidade de translação segundo o eixo-y.

Utilizando a expressão (8.10) e fazendo $\omega_z = \omega_x = v_y = 0$ obtemos a seguinte expressão para o movimento projectado

$$\begin{bmatrix} u' \\ v' \end{bmatrix} = \frac{1}{Z} \begin{bmatrix} -fv_x + uv_z \\ vv_z \end{bmatrix} + \omega_y \begin{bmatrix} -\left(f + \frac{u^2}{f}\right) \\ -\frac{uv}{f} \end{bmatrix} \tag{8.32}.$$

Esta expressão pode ser re-escrita na forma

$$\begin{aligned} \begin{bmatrix} u' \\ v' \end{bmatrix} &= \begin{bmatrix} -f \frac{v_x}{Z} + u \frac{v_z}{Z} - f\omega_y - \omega_y \frac{u^2}{f} \\ v \frac{v_z}{Z} - \omega_y \frac{uv}{f} \end{bmatrix} \\ &= \begin{bmatrix} -f \frac{v_x}{Z} - f\omega_y \\ 0 \end{bmatrix} + \left(\frac{v_z}{Z} - \omega_y \frac{u}{f} \right) \begin{bmatrix} u \\ v \end{bmatrix} \end{aligned} \quad (8.33).$$

Vamos agora considerar as medidas do fluxo óptico normal para estas imagens e que será

$$v_n = [I_u \quad I_v]^T \begin{bmatrix} u' \\ v' \end{bmatrix} = [I_u \quad I_v]^T \begin{bmatrix} -f \frac{v_x}{Z} - f\omega_y \\ 0 \end{bmatrix} + [I_u \quad I_v]^T \left(\frac{v_z}{Z} - \omega_y \frac{u}{f} \right) \begin{bmatrix} u \\ v \end{bmatrix} \quad (8.34).$$

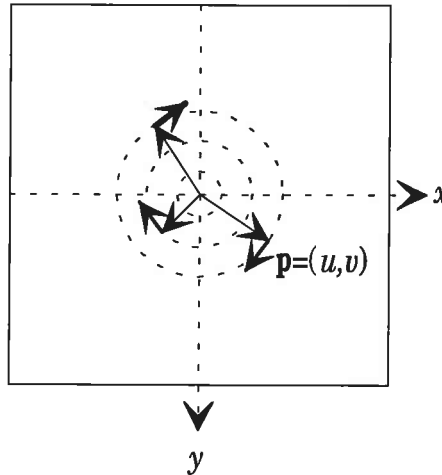


Figura 8.24 - Relação entre o vector de posição \mathbf{p} e o gradiente. Se ambos os vectores forem perpendiculares é possível encontrar uma relação entre o fluxo óptico normal e a profundidade.

Escolhendo os pontos do fluxo óptico normal em que os vectores de gradiente são perpendiculares aos vectores posicionais $\mathbf{p} = [u \quad v]^T$ - ver figura 8.24- então a expressão (8.34) simplifica-se e virá

$$v_n = fI_u \left(-\frac{v_x}{Z} - \omega_y \right) \Leftrightarrow \frac{v_n}{I_u} = -\frac{v_x}{Z} f - f\omega_y, \quad (8.35).$$

Esta equação dá uma relação directa entre o valor do fluxo óptico normal v_n para o ponto $\mathbf{p} = [u \quad v]^T$ e a sua profundidade Z .

Conhecendo o valor da expressão (8.35) para dois pontos i e j da imagem é possível estabelecer uma relação entre a profundidade relativa de ambos. Subtraindo os valores da expressão (8.35) para os dois pontos

$$\begin{cases} \frac{v_{ni}}{I_{ui}} = -\frac{v_x}{Z_i} f - f \omega_y \\ \frac{v_{nj}}{I_{uj}} = -\frac{v_x}{Z_j} f - f \omega_y \end{cases} \Rightarrow \frac{v_{ni}}{I_{ui}} - \frac{v_{nj}}{I_{uj}} = v_x \left(\frac{1}{Z_j} - \frac{1}{Z_i} \right) \quad (8.36).$$

Se o resultado da diferença (8.36) for negativo então $Z_i < Z_j$ e o contrário se a diferença for positiva. Aplicando o mesmo princípio a todos os pontos que obedecem a essas condições e ordenando-os, determinamos um *mapa de profundidade relativa* ao centro de fixação.

Experiências

Com base neste princípio foram realizadas algumas experiências com um sistema de visão activa em que as câmaras foram movimentadas de modo a que $\omega_z = \omega_x = v_y = 0^1$. As imagens foram adquiridas com o centro da imagem fixo num ponto do espaço e assegurando as condições de movimento acima descritas. A zona seleccionada para determinar o mapa de profundidade relativa foi uma zona circular central com raio igual a 100 "pixels".

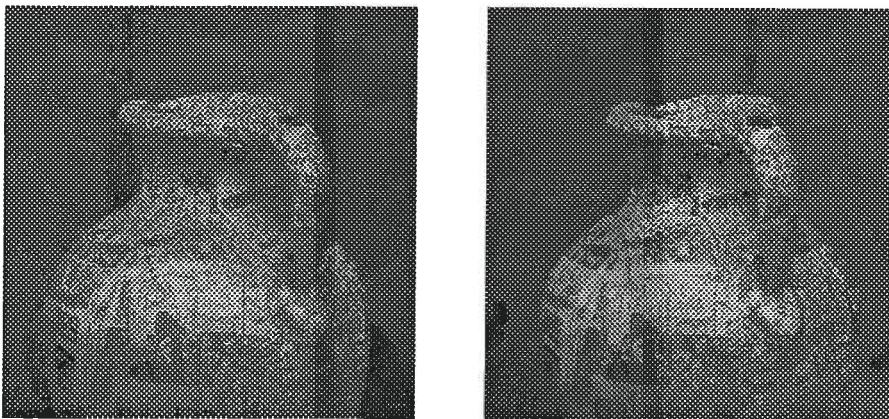


Figura 8.25 (a)

¹As imagens foram adquiridas com o sistema de visão activa (Charlie) do Computational Vision and Active Perception Laboratory (CVAP) -KTH, Stockolm University, Sweden

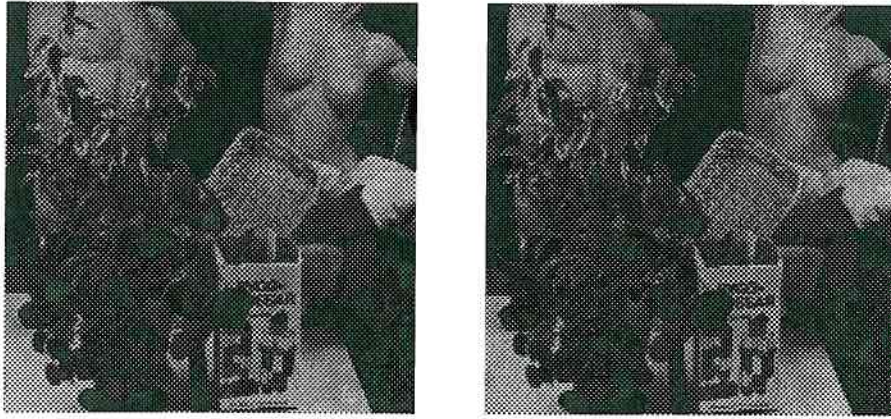


Figura 8.25 (b)

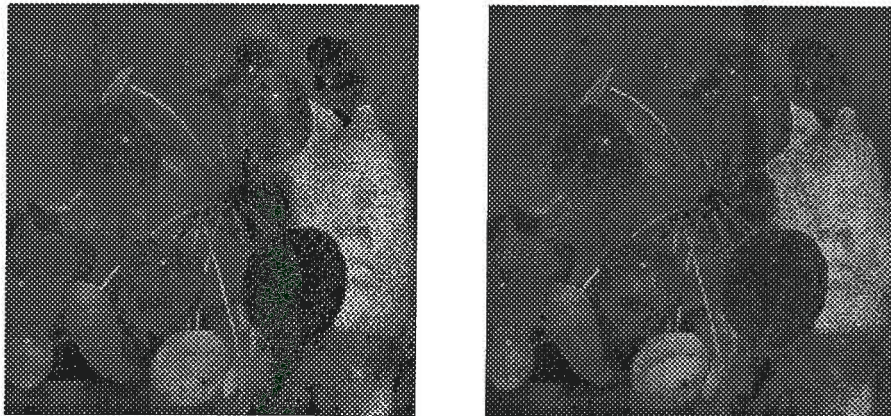


Figura 8.25(c)

Figura 8.25 - Algumas das imagens das sequências (t) e (t+dt) utilizadas para realizar as experiências.

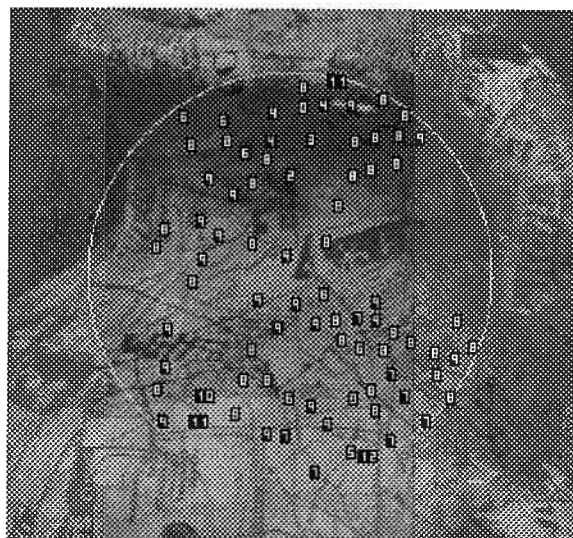


Figura 8.26(a)

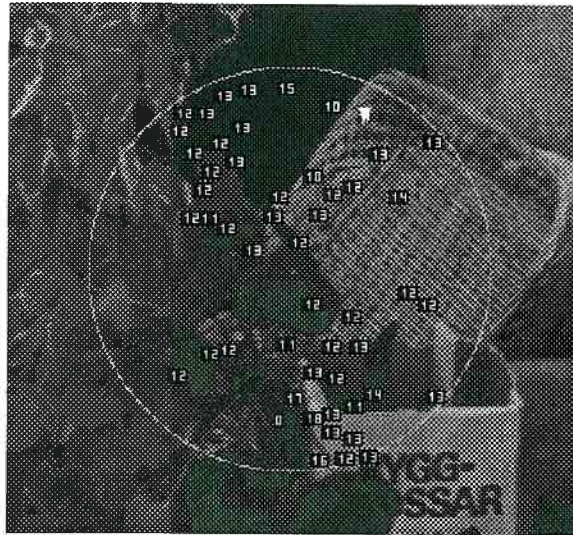


Figura 8.26(b)

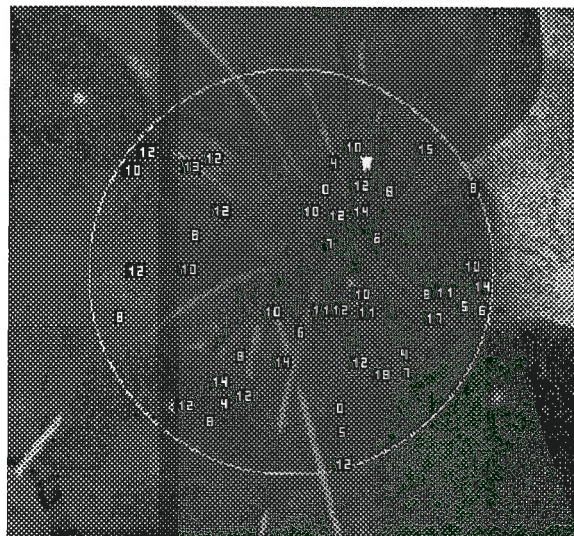


Figura 8.26(c)

Figura 8.26 - Resultados da aplicação da técnica da profundidade relativa com base no fluxo óptico normal. As etiquetas referem-se a pontos nas imagens correspondendo o valor 0 a zonas mais longínquas .

A aplicação desta técnica para determinar um mapa de profundidade relativa levanta duas questões importantes. A primeira é a dificuldade da integração de mapas obtidos em instantes de tempo diferentes pois a profundidade é relativa ao ponto de fixação. A interligação entre mapas de sequências levanta o problema da correspondência entre, pelo menos, um ponto na imagem para que o ordenamento dos pontos nos dois mapas seja possível. Outra questão está ligada com a escassa informação que se obtém por cada sequência devido à forte restrição imposta pela expressão (8.35).

8.5 Outros aspectos da visão activa

Um sistema de visão activa não se limita às vantagens geradas pelo controlo da fixação. Um sistema de visão activa também possibilita o seu controlo de acordo com os objectivos do processamento. Esse controlo pode englobar a forma como se realiza a aquisição do sinal vídeo (no espaço e em velocidade). Neste controlo estão englobados os procedimentos que permitem a concentração do sistema em determinadas zonas da cena. Esta concentração pode englobar a selecção de objectos em movimento com uma velocidade pré-determinada. Pode também englobar a concentração do sistema nos objectos que se encontram a distâncias pré-determinadas, utilizando focagem ou estereo. O conceito ligado àquilo que se designa por controlo da "atenção" ou "concentração" do sistema, engloba todos estes problemas.

A fixação, referida anteriormente, faz parte daquilo que se designa, em visão activa, por "controlo da observação". Este controlo engloba dois aspectos: estabilização da observação e mudança da observação. O primeiro aspecto inclui o conceito de fixação e consiste no controlo dos graus de liberdade de modo a manter o sistema concentrado num determinado local. No seguimento dos objectos o problema já é diferente e envolve a mudança da observação. O "controlo da observação" está ligado à selecção de novos pontos de observação. Este aspecto é importante para o desenvolvimento e aplicação destes sistemas no futuro. Será a este nível que estarão os mecanismos responsáveis pelo controlo global do sistema de forma a permitir a captação da informação necessária à percepção.

A coordenação entre acção visual e motora é um aspecto muito importante para a integração da visão activa em robótica. Sem o aspecto da integração motora e da manipulação, um sistema de visão activa não aproveitará todas as suas potencialidades.

O controlo dos graus de liberdade ópticos tal como a abertura e a focagem é essencial para assegurar uma boa qualidade das imagens e dos resultados dos algoritmos.

O "zoom" é outro dos graus de liberdade bastante útil para obter detalhe nas imagens e correspondente minimização do erro de reconstrução tridimensional. Esse último aspecto é especialmente evidente na expressão (6.44)

$$\delta = \frac{fd}{z} \Rightarrow z = \frac{fd}{\delta}$$

Desta expressão verificamos que uma imprecisão de 1% no comprimento focal irá introduzir, pelo menos, a mesma quantidade de incerteza de estimação da distância z . Este é mais um facto que indica a necessidade de uma correcta estimação do modelo de projecção e da cinemática e dinâmica do sistema. Este processo vulgarmente conhecido por calibração é um assunto em aberto em sistemas de visão activa.

CAPÍTULO 9

Conclusões e Pesquisa futura

A imagem é uma fonte rica de informação e a sua utilização tem permitido novos desenvolvimentos na robótica e em vários outros domínios. O volume de investigação realizada neste domínio levou ao aparecimento de uma nova disciplina designada por visão por computador ou visão artificial. Este trabalho insere-se nessa área e pretende contribuir para o aumento do conhecimento neste domínio.

No texto descritos métodos de visão por computador que envolvem dinâmica do sistema de visão e o controlo da focagem das lentes de forma a reconstruir a tridimensionalidade das cenas a partir das suas imagens. A introdução de actividade do sistema de visão permite definir novas técnicas de processamento e análise das imagens. No essencial, foi descrito um conjunto de técnicas para determinação da profundidade relativa e distância absoluta com base no controlo dinâmico de parâmetros do sistema de visão. As experiências realizadas levaram posteriormente ao desenvolvimento de um sistema de visão activa descrito no último capítulo.

Nos capítulos iniciais foi dado ênfase à calibração de câmaras e à detecção de primitivas nas imagens. A calibração é um assunto essencial em visão e é responsável pela estimação do modelo da formação das imagens. O modelo utilizado poderá ser aplicado em lentes em que a distorção é pequena. Com este modelo obtiveram-se bons resultados durante as experiências realizadas. No capítulo dedicado à geometria e calibração do sistema de visão, foi desenvolvida uma metodologia para a determinação do modelo da câmara invariante ao movimento. No caso de ser utilizado um modelo de projecção que envolva parâmetros de distorção, será necessário desenvolver uma nova geometria para a estereovisão e análise de imagens de movimento.

A detecção de primitivas é uma fase anterior aos algoritmos de determinação da estrutura tridimensional. Desse modo é importante que os resultados que fornece sejam correctos, e de modo a que se assegure o desempenho dos estágios seguintes do processamento. Para isso optou-se pela detecção de primitivas correspondentes às

projecções das arestas dos objectos. Para isso foram desenvolvidas técnicas para a sua detecção nas imagens e nas relações topológicas das suas projecções. As características topológicas mostram ser importantes para a correspondência entre primitivas. O mesmo se poderá argumentar relativamente à detecção de outro tipo de primitivas tais como, zonas com características físicas particulares (simetria, textura, cantos ou intersecções). Novas técnicas para a sua detecção podem ser propostas com base no controlo da focagem, zoom ou orientação das câmaras.

A focagem é um processo que permite fornecer uma medida da distância e da profundidade relativa embora de precisão dependente da profundidade de campo da lente que é utilizada. No capítulo dedicado à focagem é descrito um método de gerar um mapa de profundidade com dimensão igual à imagem. O mesmo método é explorado para um algoritmo de segmentação de imagens de cor. A focagem é também um processo essencial na visão humana e poderá influenciar o desempenho de outros processos visuais. A sua influência no processo de correspondência foi tratada no capítulo consagrado à estereovisão. A focagem fornece, de forma expedita, os volumes do espaço livre. Comparado com os actuais processos por laser ou sonar é um processo bastante atraente para robótica móvel. Este aspecto foi abordado no capítulo dedicado à construção de mapas volumétricos e à combinação de informação nesses mapas. O método proposto poderá ser modificado de modo a introduzir novas primitivas na construção do mapa.

A estereovisão é um processo clássico de determinação de distância e envolve a solução do problema da correspondência entre primitivas. No capítulo dedicado à estereovisão é proposto um processo multiocular para estabelecer uma nova restrição geométrica para a correspondência. Essa restrição é possível com base na mobilidade das câmaras. Nesse capítulo são comparados os resultados da aplicação deste método e da aplicação de restrições propostas por outros autores. A aplicação do método, associada à utilização da técnica da correlação, leva a resultados melhores quando comparadas com outras técnicas como seja o gradiente de disparidade. Pelas experiências realizadas, a restrição baseada na focagem é bastante útil para melhorar os resultados.

No capítulo final são descritas algumas das vantagens da construção de um sistema de visão com a possibilidade de controlar a orientação no espaço e os graus de liberdade ópticos (zoom, focagem e íris). No mesmo capítulo é descrito o sistema mecânico assim como os módulos de "hardware" e "software" do sistema de visão que foi desenvolvido. Para além disso, é descrito o processo visual correspondente à fixação e os resultados que podem ser obtidos através da simulação artificial deste tipo de actividade visual. As imagens de movimentos são outra fonte de informação sobre a profundidade. Esse tema foi abordado neste capítulo e na qual se fez principal referência à utilização do fluxo óptico normal para determinar um mapa de profundidade relativa.

A utilização da actividade nos sistemas artificiais de visão assim como a forma de cooperação entre os vários graus de liberdade do sistema permite criar novos mecanismos de

percepção visual artificial. Este princípio ficou demonstrado pelas técnicas de reconstrução tridimensional que foram desenvolvidas e pelos resultados práticos obtidos. Por exemplo, no caso da técnica de segmentação de imagens utilizando focagem é possível determinar mapas densos de profundidade. Estes mapas são de grande utilidade na robótica, pela informação que dão acerca do espaço vazio entre as câmaras e os objectos. A possibilidade de deslocar as câmaras no espaço permite desenvolver técnicas multioculares para estereovisão. Com base nas diferentes posições das câmaras é possível desenvolver um conjunto de novas restrições para a correspondência.

Existem diversas extensões possíveis às técnicas que foram utilizadas. Algumas delas estarão relacionadas com os modelos para a formação das imagens e que envolvam o efeito de todos os graus de liberdade ópticos ("zoom", abertura e focagem). Este é um dos aspectos mais importantes para o desenvolvimento de algoritmos de visão por computador com base nos sistemas de visão activa. Neste desenvolvimento está envolvida a detecção e localização de primitivas nas imagens (arestas, pontos de interesse, textura). A determinação de primitivas nas imagens, com a utilização da cooperação entre outras formas de processamento, graus de liberdade do sistema de visão e propriedades topológicas das primitivas nas imagens, permitirá a segmentação das imagens nas zonas de maior interesse. Por exemplo, com a utilização do "zoom" será possível detectar com maior rigor as arestas dos objectos. Com a cooperação entre diversos processos visuais é possível desenvolver novos processos de correspondência entre primitivas de imagens. No caso de imagens estereo é possível envolver no processo da correspondência diferentes graus de liberdade do sistema de visão (ângulo de vergência por exemplo). Esta forma de abordar o problema da visão está relacionada com as formas de processamento das imagens que envolvem o movimento das câmaras. Estudar formas de duplicar, em sistemas artificiais, alguns dos processos visuais humanos (fixação, observação, seguimento) permitirá realizar novos algoritmos de visão por computador.

Anexo A

A. Estimação de parâmetros

A.1. Estimação por mínimos quadrados

O exemplo mais conhecido da estimação de parâmetros é a determinação dos coeficientes de um polinómio (ou funcional) com um comportamento análogo ao comportamento de um conjunto de pontos ou medidas. O critério mais utilizado para medir a adaptação da curva do polinómio ao conjunto de pontos, é a soma dos quadrados das diferenças entre as medidas e os valores do polinómio (ou funcional) nos mesmos locais. Como essa soma representa o erro global da adaptação, o método mais vulgar consiste na minimização deste erro global. A estimação, baseada neste critério, é aplicável a sistemas com comportamentos lineares e a classe de estimadores baseados neste critério é normalmente conhecida por classe de estimadores lineares.

Para aplicar este tipo de estimação a um sistema linear do tipo $\mathbf{z}=f(\mathbf{h})$ é necessário ter um conjunto de medidas $\{(\mathbf{z}_i, \mathbf{h}_i), i = 1, n\}$ que corresponde às saídas e respectivas entradas do sistema. Representando vectorialmente o conjunto total das saídas por \mathbf{Z} , e por \mathbf{H} o respectivo conjunto das entradas, podemos representar o comportamento do sistema pela expressão vectorial

$$\mathbf{Z} = \mathbf{H} \mathbf{x} + \mathbf{v}$$

em que \mathbf{x} representa os parâmetros do sistema e \mathbf{v} é o vector dos erros das medidas originadas por um ruído branco, de média nula e Gaussiano. Neste caso o objectivo é a determinação de uma estimação para \mathbf{x} , normalmente representada por $\hat{\mathbf{x}}$, que minimize a soma dos quadrados da diferença vectorial

$$(\mathbf{Z} - \mathbf{H}\hat{\mathbf{x}})$$

e que equivale a minimizar o valor de um escalar J resultante do produto escalar $(\mathbf{Z} - \mathbf{H}\hat{\mathbf{x}})^T (\mathbf{Z} - \mathbf{H}\hat{\mathbf{x}})$ cuja expressão é

$$J = (\mathbf{Z} - \mathbf{H}\hat{\mathbf{x}})^T (\mathbf{Z} - \mathbf{H}\hat{\mathbf{x}})$$

A minimização do escalar J relativamente ao vector \mathbf{x} é obtida por

$$\frac{\partial J}{\partial \hat{\mathbf{x}}} = 0$$

o que após diferenciação e igualando a zero conduz à expressão vectorial

$$\mathbf{H}^T \mathbf{H} \hat{\mathbf{x}} = \mathbf{H}^T \mathbf{Z}$$

Quando $(\mathbf{H}^T \mathbf{H})$ possui inversa, o que equivale a dizer que o seu determinante não é nulo, a estimação por erro quadrático mínimo será então dada por

$$\hat{\mathbf{X}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{Z} \quad (\text{A.1})$$

sendo $(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T$ conhecida por matriz "pseudo-inversa".

A.2. Forma recursiva de estimação por mínimos quadrados

A forma anterior pressupõe a disponibilidade de todas as medidas simultaneamente. Supondo que ao longo do tempo podem ser obtidas mais medidas, então pode ser utilizada uma metodologia que permite uma estimação, por mínimos quadrados, que não implica o aumento sucessivo da dimensão das matrizes. Esse método consiste na utilização da forma recursiva da estimação por mínimos quadrados em que a estimação para o vector de parâmetros \mathbf{x} e representado por $\hat{\mathbf{x}}$, baseado em k medidas pode ser obtido com base na estimação de $\hat{\mathbf{x}}$ para $k-1$ medidas e na medida de índice k . Este método evita a utilização de grandes quantidades de memória e cálculos com matrizes de grandes dimensões. Consideremos a resposta de um sistema linear no momento $(k+1)$ dada pela medida \mathbf{z}_{k+1} e relacionada com entrada \mathbf{h}_{k+1} e com o ruído de medida \mathbf{v}_{k+1} através de

$$\mathbf{z}_{k+1} = \mathbf{h}_{k+1}^T \mathbf{x} + \mathbf{v}_{k+1}$$

Uma forma recursiva para a determinar $\hat{\mathbf{x}}$ é dada por

$$\hat{\mathbf{x}}_{k+1} = \hat{\mathbf{x}}_k + \mathbf{G}_{k+1} (\mathbf{z}_{k+1} - \mathbf{h}_{k+1}^T \hat{\mathbf{x}}_k) \quad (\text{A.2})$$

com

$$\mathbf{G}_{k+1} = \mathbf{P}_k \mathbf{h}_{k+1} (\mathbf{h}_{k+1}^T \mathbf{P}_k \mathbf{h}_{k+1} + 1)^{-1} \quad (\text{A.3})$$

e

$$\mathbf{P}_{k+1} = (\mathbf{I} - \mathbf{G}_{k+1} \mathbf{h}_{k+1}^T) \mathbf{P}_k \quad (\text{A.4})$$

em que \mathbf{P}_0 e $\hat{\mathbf{x}}_0$ podem ser obtidos por

$$\mathbf{P}_0^{-1} = \frac{1}{a} \mathbf{I}_n + \mathbf{h}_0 \mathbf{h}_0^T$$

$$\hat{\mathbf{x}}_0 = \mathbf{P}_0 \left(\frac{1}{a} \boldsymbol{\varepsilon} + \mathbf{h}_0 \mathbf{z}_0 \right)$$

sendo a um número muito grande e $\boldsymbol{\varepsilon} = (\varepsilon, \varepsilon, \dots, \varepsilon)$ um vector de números muito pequenos [Mendel 87]. Outra maneira de iniciar o sistema é através de uma estimação grosseira de $\hat{\mathbf{x}}$ a partir de medidas já realizadas.

A.3. Filtro de Kalman

Se o sistema for descrito pelas seguintes equações de estado

$$\begin{cases} \mathbf{x}_{j+1} = \mathbf{A}_{j+1} \mathbf{x}_j + \mathbf{B}_{j+1} \mathbf{u}_j \\ \mathbf{y}_j = \mathbf{C}_j \mathbf{x}_j \end{cases}$$

em que $\mathbf{u}(k)$ e $\mathbf{y}(k)$ são respectivamente, as entradas e saídas do sistema. Neste caso o problema centra-se na determinação de uma estimação para o estado do sistema $\hat{\mathbf{x}}$ que, no caso particular da visão por computador, consiste na optimização dos parâmetros de primitivas das imagens ou calculadas a partir das imagens.

Se o modelo do sistema for conhecido e se as condições forem as ideais (sem ruído) e se o processo for observável, os estados de um sistema podem ser determinados exactamente através de um modelo que o descreve. Suponhamos que $\hat{\mathbf{x}}_{k|k}$ é a estimação feita para o estado do sistema para o instante (k) e com base nas medidas realizadas até ao instante (k). Neste caso podemos fazer uma previsão do estado do sistema no instante ($k+1$) através de

$$\hat{\mathbf{x}}_{k+1|k} = \mathbf{A}_{k+1} \hat{\mathbf{x}}_{k|k} + \mathbf{B}_{k+1} \mathbf{u}_{k+1}$$

A saída do sistema pode ser igualmente prevista através do modelo do sistema

$$\hat{\mathbf{y}}_{k+1|k} = \mathbf{C}_{k+1} \hat{\mathbf{x}}_{k+1|k}$$

e no instante $(k+1)$ é possível obter o erro \mathbf{e}_{k+1} da nossa previsão através de uma medida directa na saída do sistema

$$\mathbf{e}_{k+1} = \mathbf{y}_{k+1} - \hat{\mathbf{y}}_{k+1|k}$$

e que mede a fiabilidade da previsão.

Se o erro \mathbf{e} não é nulo, então poderá ser utilizado para melhorar a estimação do vector de estado nesse momento, através de uma expressão do tipo

$$\hat{\mathbf{x}}_{k+1|k+1} = \hat{\mathbf{x}}_{k+1|k} + \alpha \mathbf{C}_{k+1}^{-1} \mathbf{e}_{k+1}$$

em que α é um coeficiente a escolher. Esta equação é a forma básica de um algoritmo de estimação recursivo que faz uso nas medidas presentes para melhorar a estimação do estado do sistema utilizando α como ganho. A escolha correcta do valor do ganho é difícil, pois deveria ter valores variáveis nos vários instantes (k) e com base nas propriedades estatísticas dos erros das medidas e dos erros próprios do modelo. Se estas propriedades estatísticas são conhecidas, então as técnicas derivadas dos estudos realizados por R.E. Kalman podem ser utilizadas para especificar α . Nesse caso o estimador passará a ser conhecido por filtro de Kalman.

Supondo que estamos perante um sistema linear estocástico e invariante no tempo (ver [Chui 87] e [Mendel 87] para outros tipos de sistemas) então as matrizes constantes \mathbf{A} e \mathbf{C} são invariantes e conhecidas. Se este sistema é perturbado por uma sequência de ruído branco gaussiano de média nula ξ e as medidas realizadas sobre o sistema perturbadas por ruído branco gaussiano de média nula η , então a sua descrição de estado é dada por

$$\begin{cases} \mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \xi_k \\ \mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \eta_k \end{cases} \quad (\text{A.5})$$

sendo \mathbf{Q} e \mathbf{R} as matrizes de covariância, independentes de (k) , para as sequências de ruído ξ e η . O filtro de Kalman para este caso especial pode ser descrito pelas seguintes equações recursivas:

$$\begin{cases} \hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{G}_k (\mathbf{y}_k - \mathbf{C}\hat{\mathbf{x}}_{k|k-1}) \\ \hat{\mathbf{x}}_{k|k-1} = \mathbf{A} \hat{\mathbf{x}}_{k-1|k-1} \\ \hat{\mathbf{x}}_{0|0} = E(\mathbf{X}_0) \end{cases} \quad (\text{A.6})$$

e

$$\begin{cases} \mathbf{P}_{0|0} = \text{Var}(\mathbf{X}_0) \\ \mathbf{P}_{k|k-1} = \mathbf{A}\mathbf{P}_{k|k-1}\mathbf{A}^T + \mathbf{Q} \\ \mathbf{G}_k = \mathbf{P}_{k|k-1}\mathbf{C}^T(\mathbf{C}\mathbf{P}_{k|k-1}\mathbf{C}^T + \mathbf{R})^{-1} \\ \mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{G}_k\mathbf{C})\mathbf{P}_{k|k-1} \end{cases} \quad (\text{A.7})$$

em que \mathbf{G} é conhecido por ganho do filtro e que depende do instante (k) e a matriz \mathbf{P} a covariância associada à estimação do estado \mathbf{X} .

Antes do filtro de Kalman poder ser utilizado é necessário especificar as propriedades estatísticas do ruído ξ e η , através das suas matrizes de covariância. Normalmente esta informação é desconhecida, ou conhecida sem precisão e sem garantir que o ruído é realmente branco e de média nula. Para ultrapassar este problema poder-se-á injectar uma fonte de ruído branco no processo e assegurar a aplicabilidade do filtro.

Utilizando a informação da sequência dos erros de estimação é possível modificar os valores de \mathbf{Q} e \mathbf{R} durante a operação do filtro. Neste caso o filtro de Kalman passa a ser conhecido por filtro de Kalman adaptativo [Leigh 85].

Muitos dos processos são por vezes não lineares. Neste caso para que seja possível aplicar o filtro de Kalman é necessário linearizar as equações (por exemplo através de série de Taylor) e adaptar as expressões às condições da aplicabilidade do filtro [Mendel 87].

O processo de estimação é iniciado a partir de um determinado valor inicial para a estimação do estado \mathbf{X} e com covariância conhecida \mathbf{P} . Se o número de medidas utilizadas foram (k) então ao fim do processo recursivo obtemos uma estimação $\hat{\mathbf{X}}_k$ com matriz de covariância \mathbf{P}_k , que minimiza o critério

$$\min \left\{ \left(\hat{\mathbf{X}}_0 - \mathbf{X} \right)^T \mathbf{P}_0^{-1} \left(\hat{\mathbf{X}}_0 - \mathbf{X} \right) + \sum_{i=1}^k \left(\mathbf{Y} - \mathbf{C}_i \mathbf{X} \right) \mathbf{R}_i^{-1} \left(\mathbf{Y} - \mathbf{C}_i \mathbf{X} \right) \right\} \quad (\text{A.8}).$$

O primeiro termo representa a distância de \mathbf{X} à estimação inicial $\hat{\mathbf{X}}_0$ ponderada pela matriz de covariância \mathbf{P}_0 e o segundo termo consiste no critério dos mínimos quadrados ponderado pela matriz de covariância das medidas. O filtro de Kalman fornece o melhor estimador linear no sentido da minimização da variância do erro dada por $E \left(\left(\mathbf{X} - \hat{\mathbf{X}}_k \right)^T \left(\mathbf{X} - \hat{\mathbf{X}}_k \right)^T \right)$. Como os erros do modelo são brancos, gaussianos e de média nula as estimações dadas pelos filtros de Kalman são também gaussianas.

Anexo B

B. Testes estatísticos

B.1. Teste de semelhança utilizando a função χ^2

A distribuição normal ou Gaussiana é uma função de distribuição estatística importante em estatística aplicada. No entanto existem outras funções de distribuição associadas à função de distribuição normal, que têm grande importância em estatística aplicada tal como o caso da função de distribuição χ^2 .

Considere-se uma função distribuição χ^2 , resultado da adição de n variáveis independentes z_1, z_2, \dots, z_n , cada uma com um comportamento idêntico a uma distribuição normal de média nula e variância unitária. Se todas estas variáveis forem adicionadas segundo a expressão

$$\chi^2 = z_1^2 + z_2^2 + \dots + z_n^2$$

então a variável χ_n^2 é uma variável aleatória conhecida por variável qui-quadrado com n graus de liberdade. O número de graus de liberdade representa o número de variáveis aleatórias existentes na expressão. Segundo Lehmann [Lehmann 86, pag 56] a função de densidade de probabilidade para χ^2 é dada por

$$p(\chi^2) = \left[2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right) \right]^{-1} (\chi^2)^{\left(\frac{n}{2}-1\right)} e^{-\left(\chi^2/2\right)} \quad (\text{B.1})$$

onde $\Gamma(\frac{n}{2})$ é uma função gama. Integrando a equação (B.1) desde $-\infty$ até um determinado valor de χ_n^2 obtemos uma distribuição qui-quadrado com n graus de liberdade cujos valores são normalmente disponíveis através de tabelas tendo como entradas: o valor da probabilidade dada pelo integral; o limite superior do integral χ_n^2 e o número de graus de liberdade n . Uma propriedade interessante desta distribuição é o facto de que o valor médio e a variância da variável χ_n^2 serem proporcionais ao número de graus de liberdade através das expressões

$$\mu_{\chi^2} = n \quad \text{e} \quad \sigma_{\chi^2}^2 = 2n$$

A distribuição de probabilidade da variável χ^2 é assimétrica à direita mas tende a ser simétrica com um eixo de simetria positivo, à medida que o valor de n aumenta.

Para valores de n superior a 30 a sua distribuição é muito idêntica a uma distribuição normal.

Uma das aplicações mais usuais desta distribuição é no teste de hipóteses. Suponhamos que dispomos de um estimador para o parâmetro ϕ que a partir de N observações independentes fornece uma estimação $\hat{\phi}$. Suponhamos que existem fortes razões para supor que o parâmetro tem uma média com valor ϕ_0 , mas a média estimada $\hat{\phi}$ não é exactamente igual a ϕ_0 devido à variabilidade da amostragem associada com $\hat{\phi}$. O problema consiste em testar a hipótese das duas médias serem iguais ($\phi_0 = \hat{\phi}$) e determinar o valor da diferença a partir da qual a hipótese passa a ser rejeitada. O problema pode ser solucionado em termos estatísticos, tendo em conta a probabilidade de ocorrência da diferença entre ϕ_0 e $\hat{\phi}$. Se a probabilidade de uma dada diferença é pequena, a diferença poderá ser considerada significativa e a hipótese $\phi_0 = \hat{\phi}$ deverá ser rejeitada. Se a probabilidade de uma dada diferença não for pequena, a diferença deverá ser aceite como uma variação estatística normal e a hipótese $\phi_0 = \hat{\phi}$ poderá ser aceite. Para clarificar esta técnica, considere-se o valor estimado $\hat{\phi}$ e a sua função de densidade de probabilidade $p(\hat{\phi})$. Se a hipótese $\phi_0 = \hat{\phi}$ é verdadeira então $p(\hat{\phi})$ deverá ter um valor médio ϕ_0 tal como mostra a figura B.1. A probabilidade do valor estimado $\hat{\phi}$ estar na zona de rejeição

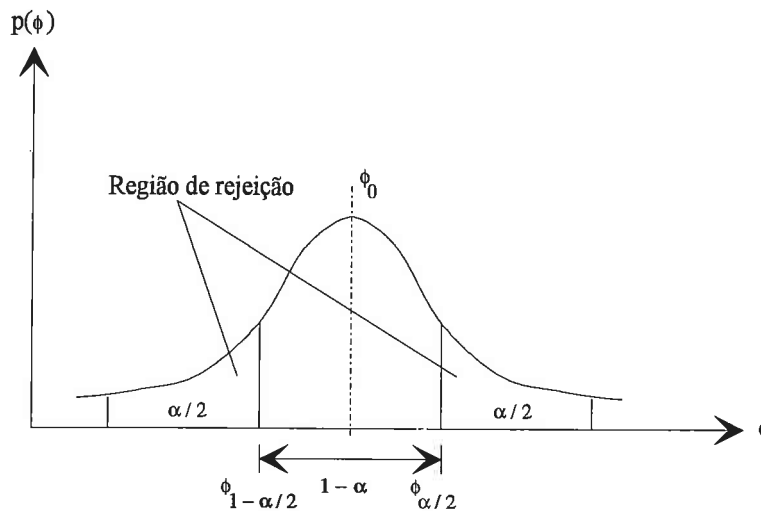


Figura B.1. - Regiões de aceitação e rejeição do teste de hipóteses

inferior é dada por

$$\Pr\left[\hat{\phi} \leq \phi_{1-\frac{\alpha}{2}}\right] = \int_{-\infty}^{\phi_{1-\frac{\alpha}{2}}} f(\hat{\phi}) d\hat{\phi} = \frac{\alpha}{2}.$$

A probabilidade do valor estimado $\hat{\phi}$ estar acima da região de aceitação da hipótese, é dada por

$$\Pr\left[\hat{\phi} > \phi_{\frac{\alpha}{2}}\right] = \int_{\phi_{\frac{\alpha}{2}}}^{\infty} p(\hat{\phi}) d\hat{\phi} = \frac{\alpha}{2}.$$

Sendo assim a probabilidade de $\hat{\phi}$ estar na região de rejeição é α . Se, pela amostragem realizada, o valor de $\hat{\phi}$ se encontra fora da região definida pelos limites $[\phi_{1-\frac{\alpha}{2}}, \phi_{\frac{\alpha}{2}}]$ então a hipótese $\phi_0 = \hat{\phi}$ deverá ser rejeitada. Por outro lado, se $\hat{\phi}$ está no intervalo $[\phi_{1-\frac{\alpha}{2}}, \phi_{\frac{\alpha}{2}}]$ então não existe uma razão forte para rejeitar a hipótese e esta deverá ser aceite. A probabilidade α é conhecida como *nível de significância* do teste e a probabilidade $(1-\alpha)$ é conhecido como *grau de confiança* do teste. Para testar a equivalência entre duas funções de densidade de probabilidade é normalmente utilizada a função distribuição χ^2 . O procedimento geral envolve a utilização duma distribuição χ^2 adequada, como medida da discrepância entre as duas funções de distribuição. Em muitos dos casos, são distribuições do tipo normal que estão em jogo, embora este raciocínio possa ser adaptado a outras funções de distribuição [Lehmann 86].

A função de probabilidade para uma distribuição normal e no caso multi-dimensional é dada pela expressão

$$\Pr(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{p}{2}} |\mathbf{S}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{S}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right] \quad (\text{B.2})$$

em que $(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{S}^{-1}(\mathbf{x}-\boldsymbol{\mu})$ segue uma lei χ^2 de p graus de liberdade (dimensão do vector de medidas \mathbf{x}). A partir da expressão (B.2) é possível definir fronteiras de regiões de densidade de probabilidade constante [Duda 73] para a qual o quadrado de $(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{S}^{-1}(\mathbf{x}-\boldsymbol{\mu})$ é constante. Estas fronteiras correspondem a elipses no caso bi-dimensional, a elipsóides no caso tri-dimensional e hiper-elipsóides no caso multi-dimensional. Estas fronteiras são especificadas definindo valores para χ^2 que, por sua vez, definem as probabilidades de uma determinada amostra \mathbf{x} da função normal se

encontrar dentro da região interior do hiper-elipsóide. Ao valor $d_{\chi^2} = (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{S}^{-1} (\mathbf{x} - \boldsymbol{\mu})$ é normalmente chamado de distância de Mahalanobis.

No caso bi-dimensional a região será uma elipse definida a partir de um valor específico de χ^2 . Um ponto (x, y) da amostra da distribuição normal $N(\boldsymbol{\mu}, \mathbf{S})$, tem uma probabilidade de se encontrar no interior dessa região do plano-xy. Esta probabilidade é igual à probabilidade que uma função χ^2 tem, com 2 graus de liberdade (no caso bi-dimensional), de não exceder um valor seleccionado. Por exemplo e ainda para o caso bi-dimensional, a partir de uma tabela da função χ^2 , a probabilidade de $\chi^2 \leq 4.605$ é $p(\chi^2 \leq 4.605) = 0.9$. Assim o ponto (x, y) tem uma probabilidade de 90% de se encontrar no interior da elipse $(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{S} (\mathbf{x} - \boldsymbol{\mu}) = 4.605$. Esta forma de interpretação é válida para um maior número de variáveis.

B.2. Determinação da região onde se encontra a média da amostra

A probabilidade de uma amostragem ter a sua média $\hat{\boldsymbol{\mu}}$ no interior de um hiper-elipsóide definido para um determinado valor de χ^2 é dada pela expressão [Tatsuoka 88; pg79]

$$p\left((\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^T \mathbf{S}^{-1} (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) \leq \frac{\chi^2}{N}\right) \quad (\text{B.3.})$$

Se conhecermos a matriz de covariância \mathbf{S} mas não o centróide $\boldsymbol{\mu}$ e se realizarmos sucessivas amostras de tamanho N então os sucessivos hiper-elipsóides com médias dadas por $\hat{\boldsymbol{\mu}}$ vão conter a média da população $\boldsymbol{\mu}$ no seu interior (ou fronteira) com probabilidade dada por (B.3). Mais concretamente, uma região de confiança de $100(1-\alpha)\%$ para a média $\boldsymbol{\mu}$ de uma população normal $N(\boldsymbol{\mu}, \mathbf{S})$ com matriz de covariância conhecida está circunscrita ao interior do hiper-elipsóide com centro em $\hat{\boldsymbol{\mu}}$ e definido pela equação

$$(\mathbf{y} - \hat{\boldsymbol{\mu}})^T \mathbf{S}^{-1} (\mathbf{y} - \hat{\boldsymbol{\mu}}) = \frac{\chi_{p,1-\alpha}^2}{N} \quad (\text{B.4.})$$

em que \mathbf{y} representa um ponto na superfície do hiper-elipsóide, N é o tamanho da amostra e χ^2 é a função qui-quadrado para $100(1-\alpha)\%$ e p graus de liberdade.

B.3. Teste baseado na distância de Mahalanobis

O estabelecimento de uma região de confiança é muito idêntico, em raciocínio, ao teste estatístico. O teste estatístico de um determinado valor considera a hipótese do valor ser rejeitável, a um nível de significado α , se e só se a região de confiança de $100(1-\alpha)\%$ não contiver o valor da hipótese.

Para testar, como hipótese, o valor μ_0 correspondente à média de uma variável normal multi-dimensional, conhecendo à partida o valor da matriz S basta verificar o cumprimento da desigualdade

$$(\mu_0 - \hat{\mu})^T S^{-1} (\mu_0 - \hat{\mu}) > \chi_{p,1-\alpha}^2 / N \quad (\text{B.5})$$

para um nível de significância α . No caso de a desigualdade se cumprir então a hipótese é rejeitada para aquele nível de significância.

O caso anterior assume que a matriz de covariância da população é conhecida. Por vezes a hipótese consiste em testar a média de uma população $\hat{\mu}$ e a variância $\hat{\sigma}^2$, ambas estimadas a partir da amostra. Se s^2 é a estimação da variância para a medida feita com base em N amostras através do estimador

$$s^2 = \hat{\sigma}^2 = \frac{1}{N-1} \sum_i (x_i - \hat{\mu})^2 \quad (\text{B.6})$$

então é possível definir um teste da hipótese de μ_0 [Tatsuoka 88; pg82] utilizando a seguinte estatística unidimensional

$$t^2 = \frac{(\hat{\mu} - \mu_0)^2}{s^2 / N} = (\hat{\mu} - \mu_0) (s^2)^{-1} (\hat{\mu} - \mu_0) N \quad (\text{B.7}).$$

Esta estatística, para o caso multi-dimensional, é obtida substituindo $(\hat{\mu} - \mu_0)$ por $(\hat{\mu} - \mu_0)$ e s^2 por $S / (N-1)$ sendo S dado por

$$S = (\mathbf{x} - \hat{\mathbf{x}})^T (\mathbf{x} - \hat{\mathbf{x}}) = \mathbf{x}^T \mathbf{x} - \hat{\mathbf{x}}^T \hat{\mathbf{x}} \quad (\text{B.8}).$$

Neste expressão a variável \mathbf{x} representa um vector de m medidas numa só amostra de N elementos

$$\mathbf{x} = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & & \vdots \\ x_{N1} & \cdots & x_{Nm} \end{bmatrix} \quad (\text{B.9})$$

e $\hat{\mathbf{x}}$ é um vector de médias dado por

$$\hat{\mathbf{x}} = \begin{bmatrix} \mu_{11} & \cdots & \mu_{1m} \\ \vdots & & \vdots \\ \mu_{N1} & \cdots & \mu_{Nm} \end{bmatrix} \quad (\text{B.10}).$$

A estatística representada por (B.7) é conhecida por estatística T^2 [Tatsuoka 88]. Como neste caso o teste é baseado só numa amostragem da população utiliza-se o índice 1 em T^2 (T_1^2). No caso teremos a expressão:

$$T_1^2 = N(N-1)(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0)^T \mathbf{S}^{-1}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0) \quad (\text{B.11}).$$

Demonstra-se que T_1^2 é proporcional a uma distribuição- F [Tatsuoka 88]. A expressão que as relaciona é a expressão

$$\frac{N-p}{(N-1)p} T_1^2 = F_{N-p}^p \quad (\text{B.12})$$

onde p é o número de variáveis correspondente ao número de graus de liberdade. No caso de $p=1$ então $F_{N-1}^1 = t_{N-1}^2$, o qual corresponde ao quadrado de uma distribuição de "student". No caso do teste da hipótese $\boldsymbol{\mu}_0$, esta será rejeitada se o valor de (B.11) exceder o valor tabelado para a distribuição- F , para o nível de significância pretendida.

Se pretendemos testar a significância da diferença entre duas médias obtidas por duas amostragens independentes, podemos utilizar a estatística T^2 para dois graus de liberdade ([Tatsuoka 88;pg 86]). Neste caso a representamos T^2 com um índice 2 (T_2^2). Esta estatística permite verificar se as duas amostras pertencem à mesma população sendo a expressão para o teste dada por

$$T_2^2 = \frac{N_1 N_2 (N_1 + N_2 - 2)}{(N_1 + N_2)} (\bar{\mathbf{X}}^{(1)} - \bar{\mathbf{X}}^{(2)})^T \mathbf{W}^{-1} (\bar{\mathbf{X}}^{(1)} - \bar{\mathbf{X}}^{(2)}) \quad (\text{B.13})$$

em que $(\bar{\mathbf{X}}^{(1)} - \bar{\mathbf{X}}^{(2)})$ corresponde à diferença entre os dois vectores média de cada uma das amostras. A matriz \mathbf{W} corresponde à matriz de covariância obtida à custa das matrizes \mathbf{S} (ver eq. B.8) para as amostras 1 e 2.

$$\mathbf{W} = \mathbf{S}_1 + \mathbf{S}_2$$

Os valores do tamanho da amostra são dados por N_1 e N_2 . Segundo Tatsuoka [Tatsuoka 88], a estatística T_2^2 pode ser relacionada com uma distribuição- F . Esta distribuição utiliza dois índices correspondentes a graus de liberdade possíveis para a distribuição. Neste caso os graus de liberdade são dados por p e $(N_1 + N_2 - p - 1)$ sendo a relação entre estatísticas dada através da expressão

$$\frac{N_1 + N_2 - p - 1}{(N_1 + N_2 - 2)} T_2^2 = F_{(N_1 + N_2 - p - 1)}^p \quad (\text{B.14}).$$

Neste caso a hipótese de as duas populações serem iguais então, para um determinado nível de significância α , será rejeitada se o valor da expressão (B.14) for maior do que o valor tabelado para a distribuição- F .

No caso de as amostras corresponderem a uma só variável então a expressão (B.14) poderá ser escrita como sendo

$$F_{(N_1 + N_2 - 2)}^1 = (\hat{\mu}_1 - \hat{\mu}_2)^T \left(s^2 \frac{N_1 + N_2}{N_1 N_2} \right) (\hat{\mu}_1 - \hat{\mu}_2)$$

com s^2 dado por

$$s^2 = \frac{\sum x_1^2 + \sum x_2^2}{N_1 + N_2 - 2}$$

Consideremos um processo de medida no qual as medidas deverão ter resultados com uma distribuição estatística do tipo normal. Por vezes é necessário testar se a média estimada $\hat{\mathbf{X}}$ da amostra, com matrix de covariância \mathbf{P} , corresponde a uma determinada média \mathbf{F} , com matrix de covariância \mathbf{Q} . Uma boa estimação desta compatibilidade é a distância de Mahalanobis [Duda 73]. Se não existe correlação entre as medidas que conduziram à média $\hat{\mathbf{X}}$ e à média \mathbf{F} , então a matrix de covariância \mathbf{S} da diferença entre $\hat{\mathbf{X}}$ e \mathbf{F} é soma de \mathbf{P} mais \mathbf{Q} . A distância de Mahalanobis é então definida por

$$d_{\chi^2} = (\mathbf{F} - \hat{\mathbf{X}})^T \mathbf{S}^{-1} (\mathbf{F} - \hat{\mathbf{X}}) \quad (\text{B.15})$$

que tem distribuição de probabilidade χ^2 com N graus de liberdade sendo N a dimensão da matriz \mathbf{S} . Consultando uma tabela de distribuição χ^2 é fácil de determinar um limiar máximo para esta distância assegurando um grau de confiança pré-determinado (por exemplo 95%).

Anexo C

C. Combinação de dados

C.1. Relações matemáticas necessárias à combinação de dados

Na combinação de informação é muito comum o problema da combinação de duas medidas obtidas de forma independente através de medidas da mesma grandeza em instantes de tempo diferentes ou realizadas por sensores diferentes. Na sua forma mais simples esta fusão consiste na fusão de dois escalares.

Consideremos então dois escalares x_1 e x_2 para a qual pretendemos realizar uma estimação

$$x_e = x_1 + \alpha(x_2 - x_1) \quad (\text{C.1})$$

onde α é um factor que deverá ser escolhido de modo que o erro médio quadrático de x_e seja mínimo. O erro em x_e pode ser relacionado com os erros em x_1 e x_2 através da expressão

$$\delta x_e = \delta x_1 + \alpha(\delta x_2 - \delta x_1) \quad (\text{C.2})$$

Assumindo erro de média nula a variância do erro (ϵ^2) é dada pelo valor da esperança matemática $E[\]$ aplicada a δx_e^2

$$\begin{aligned} \epsilon^2 &= E[\delta x_e^2] = E[(\delta x_1 + \alpha(\delta x_2 - \delta x_1))^2] \\ &= \sigma_1^2 + 2\alpha E[\delta x_2 \cdot \delta x_1] - 2\alpha \sigma_1^2 + \alpha^2 \sigma_\delta \end{aligned} \quad (\text{C.3})$$

com $\sigma_\delta = \sigma_1^2 + \sigma_2^2 - 2E[\delta x_1 \cdot \delta x_2]$. Como a correlação entre erros é, por definição, dada por

$$E[\delta x_1 \delta x_2] = R_{12} \quad (\text{C.4})$$

a equação (C.4) virá então igual a

$$\epsilon^2 = (1 - 2\alpha + \alpha^2)\sigma_1^2 + \alpha^2\sigma_2^2 + 2(\alpha - \alpha^2)R_{12} \quad (\text{C.5})$$

Como o objectivo é minimizar o erro médio quadrático, basta determinar o valor de α para o qual a primeira derivada de ϵ^2 a zero. Ou seja

$$\frac{\partial \epsilon^2}{\partial \alpha} = -2(1-\alpha)\sigma_1^2 + 2\alpha\sigma_2^2 + 2(1-2\alpha)R_{12} = 0 \quad (\text{C.6})$$

Equacionando (C.6) para α , obtemos a expressão que conduz uma combinação óptima dos escalares x_1 e x_2 do ponto de vista estatístico. A expressão à qual α deve obedecer é

$$\alpha = \frac{\sigma_1^2 - R_{12}}{\sigma_1^2 + \sigma_2^2 - 2R_{12}} \quad (\text{C.7}).$$

No caso de não existir correlação entre as medidas então $R_{12} = 0$ e a expressão para x_e é dada por

$$x_e = x_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} (x_2 - x_1) = \frac{x_2 \sigma_1^2 + x_1 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} \quad (\text{C.8})$$

Combinando as expressões (C.8) e (C.3) obtemos a expressão

$$\sigma_{x_e}^2 = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} \quad (\text{C.9})$$

que nos dá a variância do erro resultante para o caso de não existir correlação entre as medidas (caso mais usual). Como podemos ver pela expressão (C.8) a estimação final apresentará um valor dependente do valor da variância - isto é, se $\sigma_1^2 = \sigma_2^2$ então o resultado corresponde à média aritmética das medidas. No caso de uma das medidas ser perfeita (por exemplo $\sigma_1 = 0$) então a outra medida é rejeitada (rejeição de x_2).

No caso das estimações serem vectores ($\mathbf{x}_1, \mathbf{x}_2$) o raciocínio anterior ainda é válido e a variâncias passam a ser representadas por matrizes de covariância. Neste caso

$$\sigma_1^2 \rightarrow \mathbf{P}_1, \sigma_2^2 \rightarrow \mathbf{P}_2$$

e a correlação representada pela matriz de covariância cruzada que por definição é dada por [Gelb 88]

$$\begin{aligned} \mathbf{P}_{12} &= E\left[(\mathbf{x}_1 - E[\mathbf{x}_1])(\mathbf{x}_2 - E[\mathbf{x}_2])^T\right] \\ &= E[\mathbf{x}_1 \mathbf{x}_2^T] - E[\mathbf{x}_1]E[\mathbf{x}_2] \end{aligned} \quad (\text{C.10})$$

em que $E[\]$ indica o operador esperança matemática. Tendo em conta que $2\mathbf{P}_{12} = \mathbf{P}_{12} + \mathbf{P}_{21}^T$ a matriz equivalente ao factor α é dada por [Blackman 86]

$$\mathbf{S} = [\mathbf{P}_1 - \mathbf{P}_{12}][\mathbf{P}_1 + \mathbf{P}_2 - \mathbf{P}_{12} - \mathbf{P}_{12}^T]^{-1} \quad (\text{C.11})$$

e a matriz de covariância resultante é

$$\mathbf{P}_c = \mathbf{P}_1 - [\mathbf{P}_1 - \mathbf{P}_{12}][\mathbf{P}_1 + \mathbf{P}_2 - \mathbf{P}_{12} - \mathbf{P}_{12}^T]^{-1}[\mathbf{P}_1 - \mathbf{P}_{12}]^T \quad (\text{C.12}).$$



BIBLIOGRAFIA

- [Abbot 88] A. Abbot and N. Ahuja, "Surface reconstruction by dynamic integration of focus, camera vergence, and stereo", *International Conference Computer Vision - ICCV 88*, Tampa, Florida, 1988, pág.532-543
- [Aloimonos 85] J. Aloimonos, M. Swain, "Shape from texture", *Proceedings International Joint Conference on Artificial Intelligence - IJCAI 85*, Ago 1985, Los Angeles.
- [Aloimonos 86] J. Aloimonos, *Computing intrinsic images*, Ph.D. Thesis, TR -198, Department of Computer Science, University of Rochester, New York, Sept. 1986
- [Aloimonos 87] J. Aloimonos, M. Swain, *Paraperspective projection: between orthography and perspective*, CAR-TR -320, Center for Automation Research, University of Maryland, Maryland, May 1987
- [Aloimonos 88] J. Aloimonos, A. Bandopadhyay, and I. Weiss. "Active vision", In *Proceedings, 1st International Conference on Computer Vision - ICCV'88*, pp. 34-54, June 1987. Also in *International Journal of Computer Vision*, 1,4, pp. 333-356, 1988.
- [Aloimonos 89] J. Aloimonos and D. Shulman, *Integration of Visual Modules- An Extension of the Marr Paradigm*, Academic Press, 1989
- [Arnold 80] R. Arnold and T. Binford, "Geometric Constraints in Stereo Vision", *Society Photo-Optical Instrument Engineers*, 238, 1980, pg. 281-292
- [Ayache 87a] N. Ayache and F. Lustman, "Fast and reliable trinocular stereovision", *Proceedings 1st International Conference Computer Vision*, June 8-11, 1987, pág. 422-427
- [Ayache 87b] N. Ayache, and B. Fararja, "Efficient registration of stereo images by instaling graph descriptions of edge segments", *International Journal of Computer Vision*, N.1, Vol 2, 1987
- [Ayache 88a] N. Ayache, *Construction et Fusion de Représentations Visuelles 3D - Applications à la Robotique Mobile*, These de Docteur es Sciences, Université de Paris-Sud, Centre d'Orsay, Mai 1988
- [Ayache 88b] N. Ayache and C. Hansen, *Rectification of images for binocular and trinocular Stereo Vision*, RR-860, INRIA, Juin 1988

BIBLIOGRAFIA

- [Ayache 89] N. Ayache and O. Faugeras, "Maintaining Representations of the Environment of a Mobile Robot", *IEEE Transactions on Robotics and Automation*, Vol 5, No6, Dec 89.
- [Bajcsy 76] R. Bajcsy and L. Lieberman, "Texture gradient as a depth cue", *Computer Vision, Graphics, and Image Processing*, 5, pg.52-67, 1976
- [Bajcsy 88] R. Bajcsy, "Active perception", *Proceedings of IEEE*, 79:996-1005,1988
- [Bajcsy 90] R. Bajcsy, S. Lee, and A. Leonardis, "Color image segmentation with detection of highlights and local illumination induced by inter-reflections", *Proceedings of 10th International Conference on Pattern Recognition*, Atlantic City, June 1990
- [Baker 80] J. Baker, and others, "Elements of Photogrammetric Optics", in *Manual of Photogrammetry*, Chester C. Slama (Ed. in Chief), American Society of Photogrammetry, ISBN 0-937294-01-2, 1980
- [Baker 90] H. Baker, "Scene structure from a moving camera", *AI and the Eye*, Blake, A., and Troscianko, T. (Eds), John Wiley & Sons, 1990
- [Ballard 82] D. Ballard and C. Brown, *Computer Vision*, Englewood Cliffs, Prentice-Hall, 1982
- [Ballard 88] D. Ballard, "Eye fixation and early vision: kinetic depth", *International Conference Computer Vision - ICCV 88*, Tampa, Florida, 1988, pp 524-531
- [Ballard 90] D. Ballard, *Animate Vision*, TR-329 University of Rochester, USA, Feb., 1990
- [Bandopadhyay 87] A. Bandopadhyay, *A computational study of rigid motion*, Ph.D. thesis, Computer Science Dept., U.Rochester, 1987.
- [Barrow 78] H. Barrow and J. Tenenbaum, "Recovering intrinsic scene characteristics from images", em *Computer Vision Systems*, Ed. A. R. Hanson & E. M. Riseman, Academic Press, NY, 1978, pág. 3-26
- [Barrow 81] H. Barrow and J. Tenenbaum, "Interpreting line drawings as three-dimensional surfaces", *Artificial Intelligence*, 17, pg.75-116, 1981
- [Bar-Shalom 81] Y. Bar-Shalom, "On the Track-to-Track Correlation Problem", *IEEE Transactions on Automatic Control*, AC-26, April 1981
- [Batista 92] J. Batista, *Calibração de câmaras de video*, Provas de Aptidão de Capacidade Científica e Pedagógica, Faculdade de Ciências e Tecnologia da Universidade de Coimbra, 1992
- [Besl 88] P. J. Besl, "Active, Optical Range Imaging Sensors", *Machine Vision and Applications*, 1, 1988, pg.127-152
- [Blake 85] A. Blake, A. Zisserman and G. Knowles, "Surface description from stereo and shading", *Image and Vision Computing*, 3, pg.183-191, 1985
- [Black 87] A. Black and A. Zisserman, *Visual Reconstruction*, MIT Press, Cambridge MA, 1987
- [Blackman, 86] S. Blackman, *Multiple-Target Tracking with Radar Applications*, Artech House, Inc, 1986, ISBN 0-89006-179.
- [Brady 85] M. Brady, "Artificial Intelligence and Robotics", *Artificial Intelligence*, 26, pág. 79-121

- [Brady 89] M. Brady, "Introduction: The Problems of Robotics", em *Robotics Science*, ed. por Michael Brady, The MIT Press, Cambridge, Massachusetts, 1989, pág. 1 - 35
- [Brockelbank 89] D. Brockelbank and Y. Yang, "An Experimental Investigation in the Use of Color in Computational Stereopsis", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 19, No 6, Nov, Dec 1989
- [Brooks 85] R. Brooks, "Symbolic error analysis and robot planning", *International Journal of Robotics Research*, Vol. 1, No4, 1985, pág. 29-68
- [Brooks 86] R. Brooks, "A Robust Layered Control System For A Mobile Robot", *IEEE Journal of Robotics and Automation*, Vol. RA-2, No. 1, March 1986, pag 14-23
- [Brown 86] M. Brown, "The extraction of curved surface features with generic range sensors", *International Journal of Robotics Research*, Vol. 1, No. 5, 1986, pág. 3-18
- [Brown 90a] C. Brown, "Gaze controls with interactions and delays", *IEEE Transactions on Systems, Man, and Cybernetics*, 20, 2, March 1990.
- [Brown 90b] C. Brown, "Prediction and cooperation in gaze control", *Biological Cybernetics*, 1990b
- [Burger 90] W. Burger and B. Bhanu, "Estimating 3-D Egomotion from Perspective Image Sequences", *IEEE Transactions on Pattern Analysis Machine Intelligence*, Vol PAMI-12, No11, Nov. 1990
- [Burt 80] P. Burt and B. Julesz, "Modifications of the classical notion of Panum's fusional area", *Perception*, 9, 1980, pág. 671-682
- [Buxton 83] B. Buxton and H. Buxton, "Monocular depth perception from optical flow using space time signal processing", *Proceedings Royal Society of London, B*, Vol. 218, 1983, pág. 27-47
- [Carpenter 88] R.H.S. Carpenter, *Movements of the Eyes*, Pion Limited 207 Brondesbury Park, London, 1988, ISBN 0-85086-109-0
- [Canny 85] J. Canny, "A computational approach to edge detection", *IEEE Transactions Pattern Analysis Machine Intelligence*, Vol PAMI-8, No6, Jan. 1985, pág.679-698
- [Chien 86] C. Chien and J. Aggarwal, "Identification of 3D objects from Multiple Silhouettes Using Quatrees/Octrees", *Computer Vision, Graphics, and Image Processing*, 36, Pg 256-273, 1986
- [Chui 87] C. Chui and G. Chen, *Kalman Filtering-With real-time applications*, Springer-Verlag Series in Information Sciences, Vol 17, 1987
- [Cox 89] I. Cox, "Blanche: Position Estimation for an Autonomous Robot Vehicle", *Proceedings of International Workshop on Robots and Systems - IROS'89*, Tuksuba, Japan, Sept. 4-6, 1989
- [Craig 86] J. Craig, *Introduction to Robotics - Mechanics & Control*, Addison-Wesley Pub. Co., 1986.
- [Crowley 90] J. Crowley and P. Stelmaszyk, "Measurement and Integration of 3D Structures By Tracking Edge Lines", *Proceedings First European Conference on Computer Vision - ECCV 90*, Antibes, France, April 1990, Springer-Verlag.

BIBLIOGRAFIA

- [Cutting 86] James E. Cutting, "Perception with an eye for motion", The MIT Press, 1986, ISBN 0-262-03119-1
- [Dario 89] P. Dario, "Contact Sensing for Robot Active Touch", em *Robotics Science*, ed. por Michael Brady, MIT Press, 1989, pág. 138 - 163
- [Darrel 88] T. Darrel and K Wohn, "Pyramid based depth from focus", *Proceedings of Computer Vision and Pattern Recognition*, Ann Arbor, Michigan, June 1988, IEEE Computer Society.
- [Derich 88] R. Derich and J. Cocquerez, "An efficient method to build early image description", *International Conference on Pattern Recognition - ICCV 88*, Rome, 1988
- [Derich 89] R. Derich, " Using Canny's criteria to derive a recursively implement optimal edge detector", *International Journal of Computer Vision*, 1(2), May 1987.
- [Derich 90] R. Derich, G. Giraudon, "Accurate Corner Detection: An Analytical Study", *Third International Conference on Computer Vision - ICCV 90*, Dec 4-7, 1990
- [Dhoud 91] U. Dhoud and J. Aggarnal, "A Cost-Benefit Analysis of a third camera for stereo correspondance", *International Journal of Computer Vision*, Vol 6, Nº 1, 1991
- [Dias 88] J. Dias, L. de Sá, "Módulo de Processamento Digital de Imagem com Microprocessador de Sinal", *III Simpósio da Electrónica das Telecomunicações*, Porto, 4-6 Maio, 1988
- [Dias 91] Dias, J., de Almeida, A., Araújo, H., Batista, J., "Improving Camera Calibration by Using Multiple Frames in Hand-Eye Robotic Systems", *IROS 91*, Osaka (Japan), Nov.3-5, 1991
- [Duda 73] R. Duda, and P. Hart, *Pattern Classification and Scene Analysis*, Wiley, NY, 1973
- [Durrant-Whyte 88] H. Durrant-Whyte, *Integration, Coordination, and Control of Multisensor Robot Systems*, Kluwer Acad. Pub., 1988
- [Elfes 87] A. Elfes, "Sonar-Based Real-World Mapping and Navigation", *IEEE Journal of Robotics and Automation*, Vol RA-3, No 3, 1987
- [Ens 93] J. Ens and P. Lawrence, "An Investigation of Methods for Determining Depth from Focus", *IEEE PAMI*, Vol 15, No 2, February, 1993
- [Espelid 91] R. Espelid, I. Jonassen, "A comparison of splitting methods for the identification of corner-points", *Pattern Recognition Letters*, 12, 1991
- [Faugeras 87] O. Faugeras, and G. Toscani, "The Calibration Problem for Stereoscopic Vision", *Proceedings of the NATO Advanced Research Workshop on Sensor Devices and Systems for Robotics*, Catalonia (Spain)- Oct 1987, in NATO ASI Series, Vol F52, Springer-Verlag
- [Faugeras 88] O. Faugeras, *A Few Steps Toward Artificial 3D Vision*, RR-790 INRIA, 1988
- [Flynn 88] A. Flynn, "Combining Sonar and Infrared Sensors for Mobile Robot Navigation", *International Journal of Robotics Research*, Vol 7, No 6, Dec. 1988
- [Fu 87] K. Fu, R. Gonzalez, C. Lee, *Robotics - Control, Sensing, Vision and Intelligence*, McGrawHill, 1987

- [Gibson 50] J. Gibson, *The perception of the visual world*, Houghton Mifflin, Boston, 1950
- [Giralt 84] G. Giralt, R. Chatilla, and M. Vaisset, "An integrated navigation and motion control system for autonomous multisensory mobile robots", Ed. Brady and Paul, *First International Symposium of Robotics Research*, The MIT Press, Cambridge, Massachusetts, 1984, pág. 109-116
- [Gong 88] S. Gong, "Improved local flow", *Proceedings 4th Alvey Vision Conference*, Manchester, UK, 1988, pág. 147-152
- [Grimson 81] W. Grimson, "A computer implementation of a theory of human stereo vision", *Proceedings Royal Society of London*, Vol B 292, 1981, pág. 217-253
- [Grimson 84] W. Grimson, "Binocular shading and visual surface reconstruction", *Computer Vision Graphics and Image Proceedings*, 1984, pág. 18-44
- [Grimson 85] W. Grimson, "Computacional experiments with a feature based stereo algorithm", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-7, pág. 17-33, 1985
- [Grossmann 87] P. Grossamann, "Depth from focus", *Pattern recognition Letters*, Vol 5, No 1, 1987
- [Grossmann 89] P. Grossmann, "COMPACT - a surface representation scheme", *Image and Vision Computing*, Vol 7, No 2, pg. 115-121
- [Haralick 80] R. Haralick, "Using perspective transformations in scene analysis", *Computer Graphics and Image Processing*, Vol.13, 1980, pg. 191-221
- [Haralick 84] R. Haralick, "Digital steps edges from zero crossing of second directional derivatives", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-6, No.6, Jan1984, pág.58-68
- [Harmon 87] S. Harmon, "The ground surveillance robot (GSR): an autonomous vehicle designed to transit unknow terrain", *IEEE Journal Robotics and Automation*, Vol. RA-3, No. 3, pág. 266-279
- [Harris 88] C. Harris and M. Stephens, "A combined corner and edge detector", *Proceedings 4th Alvey Vision Conference*, Manchester, UK, 1988, pág. 147-152
- [Harris 90] C. Harris, "Structure from motion under orthographic projection", *Proceedings 1st European Conference on Computer Vision - ECCV 90*, ed. O.Faugeras, Springer-Verlag, 1990
- [Haustein 90] W. Haustein, "Considerations on Listing's Law and the Primary Position by Means of a Matrix Description of Eye Position Control", *Biological Cybernetics*, 60
- [Hayashi 90] B. Hayashi and S. Negandssipour, "Direct Motion Stereo: Recovery of Observer Motion and Scene Structure", 1990
- [Haynes 87] S. Haynes and R. Jain, "A qualitative approach for recovering depth in dynamic scenes" Proc. of The IEEE Workshop on Computer Vision, Miami Beach, pg. 66-71
- [Healey 89] G. Healey and T. Binford, "Using color for geometry-insensitive segmentation", *Journal of the Optical Society of America*, 6, 188
- [Hildreth 84] E. Hildreth, *The measurement of visual motion*, The MIT Press, Cambridge, Massachusetts, 1984.

BIBLIOGRAFIA

- [Hollerbach 87] J. Hollerbach, "Robot hands and tactile sensing", em *AI in the 1980s and Beyond*, ed. por Grimson and Patil, The MIT Press, Cambridge, Massachusetts, 1987, pág. 317-342.
- [Hong 92] J. Hong, X. Tan, B. Pinette, R. Weiss, E. Riseman, "Image-Based Homing", *IEEE Control Systems Magazine*, Vol 12, No1, Feb. 1992
- [Horn 68] B. Horn, *Focusing*, MIT Artificial Intelligence Laboratory, Memo Nº 160, May, 1968
- [Horn 75] B. Horn, "Obtaining shape from shading information", em *The Psychology of Computer Vision*, ed. P. H. Winston, McG.-Will, New York, 1975
- [Horn 81] B. Horn and B. Schunk, "Determining optical flow", *Artificial Intelligence*, 8, 1981, pág. 185-204
- [Horn 86] B. Horn, *Robot Vision*, The MIT Press, Cambridge, Massachusetts, 1986
- [Horii 92] A. Horii, *The Focusing Mechanism in the KTH Head Eye System*, T.R. - CVAP, Department of Numerical Analysis and Computing Science, Royal Inst. of Technology, Stockholm, Sweden
- [Horn 87] B. Horn, "Motion fields are hardly ever ambiguous", *International Journal of Computer Vision*, 1 (3), Oct. 1987
- [Huang 85] T. Huang and M. Blonstein, "Robust algorithms for computing three dimensional motion from image sequences", *Proceedings IEEE Computer Vision and Pattern Recognition*, 1985
- [Hwang 89] T. Hwang, J. Clark, and A. Yuille, *A depth recovery algorithm using defocus information*, TR.89-2, Harvard Robotics, Lab, USA, 1989
- [Ikeuchi 81] K. Ikeuchi and B. Horn, "Numerical shape from shading and occluding boundaries", *Artificial Intelligence*, 17, pg.141-184, 1981
- [Ishiguro 90] H. Ishiguro, P. Stelmazyk, and S. Tsuji, "Acquiring 3D Structure by Controlling visual attention of a mobile robot", *Proceedings International Conference Robotics and Automation*, 1990
- [Ito 86] M. Ito and A. Ishii, "Three view stereo analysis", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Nº 4, PAMI-8, 1986
- [Jarvis 76] R.A. Jarvis, "Focus Optimization Criteria for Computer Image Processing", *Microscope*, Vol24, No2, 1976
- [Julesz 71] B. Julesz, "Foundations of cyclopean perception", The University of Chicago Press, Ltd., 1971
- [Kanatani 84] K. I. Kanatani, "Detection of surface orientation and motion from texture by a stereological technique", *Artificial Intelligence*, 23, 1984, 213-237.
- [Kanatani 90] K. I. Kanatani, *Group-Theoretical Methods in Image Understanding*, Springer-Verlag, 1990
- [Klinker 88] G. Klinker, S. Shafer, T. Kanade, "Image segmentation and reflection analysis through color", *In Proceedings of DARPA Image Understanding Workshop*, Pittsburgh, PA, 1988

- [Klinker 90] G. Klinker, S. Shafer, T. Kanade, "A Physical Approach To Color Image Understanding", *International Journal of Computer Vision*, 4,7-38, 1990.
- [Krotkov 88] E. Krotkov, "Focusing", *International Journal of Computer Vision*, 1,3, pp. 223-238, 1988
- [Krotkov 89] E. Krotkov, *Active Computer vision by Cooperative Focus and Stereo*, Springer-Verlag Series in Perception Engineering, 1989
- [Koenderink 86] J. Koenderink, "Optic flow", *Vision Research*, 26, pg.161-180, 1986
- [Leigh 85] J.R. Leigh, *Applied Digital Control-Theory, Design and Implementation*, Prentice-Hall-UK, 1985, ISBN 0-13-040171-4
- [Levine 85] M. Levine, *Vision in Man and Machine*, McGraw-Hill, New York, 1985
- [Li 91] S. Li, I. Miyawaki, H. Ishigoro, S. Tsuji, "Realizing Active Vision by a Mobile Robot", *IEEE Workshop on Visual Motion*, Oct. 7-9, 1991
- [Lim 90] J. Lim, *Two-Dimensional Signal and Image Processing*, Prentice-Hall, 1990, pág. 29-30
- [Longuet-Higgins 81] H. Longuet-Higgins and K. Prazdny, "The interpretation of a moving retinal image", *Proc. of the royal Society London, B*, 208, pg.385-397, 1981
- [Lowe 90] D.G. Lowe, " Visual Recognition as Probabilistic Inference from Spatial Relations" em *AI and the eye*, ed. A Black & T. Troscianko, John Wiley & Sons, Ltd., 1990, ISBN 0 471 921947
- [Luo 89] R. C. Luo, M. G. Kay, "Multisensor Integration and Fusion in Intelligent Systems", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 9, No. 5, 1989
- [Lustman 87] F. Lustman, *Vision Stereoscopique et Perception du Mouvement en Vision Artificielle*, These Docteur en Science, University de Paris-Sud, Centre d'Orsay, Dec, 1987
- [Landy 87] M. Landy, "Parallel model of the kinetic depth effect using local computations", *Journal of the Optical Society of America, A*, 4, pg.864-877
- [Mayhew 81] J.E. Mayhew and J.P. Frisby, "Psychophysical and computational studies towards a theory of human stereopsis", *Artificial Intelligence*, 17, pg.376-379, 1981
- [Martin 88] W. Martin, J. Aggarwal, *Motion Understanding - Robot and Human Vision*, Kluwer Academic Publ., ISBN 0-89838-258-0
- [Marr 76] D. Marr and T. Poggio, "Cooperative computation of stereo disparity", *Science*, 194, pág 283-287, 1976
- [Marr 77a] D. Marr, "Analysis of occluding contour ", *Proceedings Royal Society of London, B* 197, pág 491-475
- [Marr 77b] D. Marr, "Artificial Intelligence- a personal view", *Artificial Intelligence*, 9, pág 37-48
- [Marr 78] D. Marr, "Representing visual information", em *Computer Vision Systems*, Ed. A. R. Hanson & E. M. Riseman , Academic Press, NY, 1978, pág. 61-80

BIBLIOGRAFIA

- [Marr 79] D. Marr and T. Poggio, "A Computational theory of human stereo vision", *Proceedings Royal Society of London*, B 204, 1979
- [Marr 80] D. Marr and E. Hildreth, "Theory of edge detection", *Proceedings Royal Society of London*, B 207, 1980, pág. 187-217
- [Marr 82] D. Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, Pub. W.H. Freeman, 1982
- [McAllister 92] D. McAllister, "3D Displays", *BYTE*, May 1992
- [Mclauchlan 92] P.F. Mclauchlan, I. Reid and D. W. Murray, "Coarse Image Motion for Saccade Control", *3rd British Machine Vision Conference*, Leeds, Sept. 1992
- [Medioni 85] G. Medioni and R. Nevatia, "Segment -based stereo matching", *Computer Vision, Graphics and Image Processing*, 32, pág. 2-18, 1985
- [Mendel 87] J. Mendel, "Lessons in Digital Estimation Theory", *Prentice-Hall Signal Processing Series*, New-Jersey, 1987, ISBN 0-13-53 0809-7 025
- [Milenkovich 85] V. Milenkovich and T.Kanade, "Trinocular stereo", *Proceedings Image Understanding Workshop*, Miami, Florida, 1985
- [Mingolla 86] E. Mingolla and J. Todd, "Perception of solid shape from shading", *Biological Cybernetics*, 53, pg.137-151, 1986
- [Mitiche 86] A. Mitiche, S. Seida, and J. Aggarwal, "Interpretation of Structure and Motion from Line Correspondences", *Proceedings 8th International Conference on Pattern Recognition*, 1986.
- [Moravec 77] H. P. Moravec, "Towards Automatics Visual Obstacle Avoidance ", *Proceedings of the 5th International Joint Conference on Artificial Intelligence - IJCAI*, Cambridge, Massachusetts, 1977
- [Moravec 80] H. P. Moravec, *Obstacle avoidance and navigation in the real world by a seeing robot rover*, PhD. Thesis, Report No.STAN-CS-80-813, Stanford University, 1980
- [Mohan 92] R. Mohan, and R. Nevatia, "Perceptual Organization for Scene Segmentation and Description", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, No.6, June 1992
- [Mohr 92] R. Mohr, L. Morin, L., and E. Grosso, "Relative Positioning with Uncalibrated Cameras", in *Geometric Invariance in Computer Vision*, editors Mundy, J.L., and Zisserman, A., , MIT Press, ISBN 0-262-13263-X
- [Monga 89] O. Monga, R. Deriche, G. Malandain, J.-P. Cocquerez, *Recursive Filtering and Edge Closing: Two Primary Tools for 3D Edge Detection*, RR-1103, INRIA, Oct. 1989
- [Monga 90] O. Monga, *Segmentation d'images: où en sommes nous ?*, RR-1216, INRIA, Avril 1990
- [Moore 93] D. S. Moore, G. P. McCabe, *Introduction to the Practice of Statistics*, W.H. Freeman and Company, ISBN 0-7167-2250-X
- [Murray 90] D.W. Murray and B.F. Buxton, *Experiments in the machine interpretation of visual motion*, MIT Press, ISBN 0-262-13263-X

- [Nagel 83] H.-H. Nagel, "Overview on image sequence analysis", *NATO-ASI on Image Sequence Processing and Dynamic Scene Analysis*, T.S. Huang ed. Braunlage, FR Germany, Springer-Verlag, 1982
- [Nagel 86] H.-H. Nagel and W. Enkelmann, "An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-8, No 5, 1986
- [Nalwa 86] V.S.Nalwa and T.O.Binford, "On detecting edges", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol PAMI-8, No.6, Nov. 1986, pág.699-714
- [Navab 90] N. Navab, R. Deriche, and O. Faugeras, "Recovering 3D Motion and Structure from Stereo and 2D Token Tracking Cooperation", *Third International Conference on Computer Vision - ICCV 90*, Dec. 4-7, 1990, Osaka, Japan
- [Nayar 90] S.K. Nayar and Y. Nakagawa, "Shape from focus: An effective approach for rough surfaces", *Proceedings of IEEE International Conference on Robotics and Automation*, Cincinnati, OH, May 1990
- [Nelson 88] R.C. Nelson, *Visual Navigation*, Computer Vision Laboratory, University of Maryland, Technical Report CS-TR-2087
- [Nelson 90] R.C. Nelson, *Qualitative Detection of Motion by a Moving Observer*, TR-341, University of Rochester, April 90
- [Netravali 88] A.N. Netravali, B.G. Haskell, *Digital Pictures*, Plenum Press, NY, ISBN 0-306-42791-5
- [Ohta 85] Y.Ohta and T. Kanade, "Stereo by intra- and inter-scanline search", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.PAMI-7, No.2, Mar. 1985, pág.139-154
- [Olson 90] T. Olson, D. Coombs, *Real-Time Vergence Control for Binocular Robot*, University Rochester TR-348, June 1990
- [Paul 81] R. P. Paul, *Robot Manipulators*, The MIT Press, Cambridge, Massachusetts, 1981
- [Pietikäinäm 86] M. Pietikäinäm and D. Harwood, "Depth from three camera stereo", *Proceedings IEEE Conference on Pattern recognition*, Miami Beach, FL, June 22-26, 1986, pág. 2-6
- [Penna 86] M.A. Penna, and R. Patterson, *Projective Geometry and its Applications to Computer Graphics*, Prentice-Hall, ISBN 0-13-730649-0, 1986.
- [Pentland 84] A. Pentland, "Local shading analysis", *IEEE Trans. on PAMI*, PAMI-6, pg. 170-187, 1984
- [Pentland 85] A. Pentland, "A new sense of depth of field", *Proceedings International Joint Conference on Artificial Intelligence - IJCAI*, pp. 988-994, August, 1985
- [Pentland 86] A. Pentland, "Shading into texture", *Artificial Intelligence*, 29, pg. 147-170, 1986
- [Pentland 87] A. Pentland, "A New Sense for Depth of Field", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, VOL PAMI-9, No4, July 1987
- [Poggio 85] T. Poggio V. Torre and C. Koch, "Computational Vision and Regularization Theory", *Nature*, 317, pg. 314-319

BIBLIOGRAFIA

- [Pollard 85] S.B. Pollard, J.E.W. Mayhew, and J.P. Frisby, "PMF: A Stereo Correspondence Algorithm Using a Disparity Gradient Limit", *Perception*, 1985, 14,pg.449-470
- [Pollard 89] S.B. Pollard, J. Porril, J.E.W. Mayhew and J.P. Frisby, "Disparity Gradient, Lipschitz Continuity, and Computing Binocular Correspondences", *Image Understanding 89*, ed. by S.Ullman & W.Richards, Ablex Pub. Corp., USA, ISBN 0-89391-547-5
- [Potmesil 87] M. Potmesil, "Generating Octree Models of 3D Objects from Their Silhouettes in 2 Sequence of Images", *Computer Vision, Graphics, and Image Processing*, 40, 1987
- [Prazdny 85] K. Prazdny, "Detection of binocular disparities", *Biological Cybernetics*, 52, pg.93-99, 1985
- [Reisfeld 90] D. Reisfeld, H. Wolfson, and Y. Yeshurun, "Detection of Interest Points Using Symmetry", *Third International Conference on Computer Vision*, Dec 4-7, 1990
- [Richardson 88] J.M. Richardson and K.A. Marsh, "Fusion of Multisensor Data", *International Journal of Robotics Research*, Vol7, No6, Dec. 1988
- [Richards 84] W. Richards, "Structure from stereo and motion", *Journal of Optical Society of America*, A2, 1984, pág. 343-349
- [Roberts 88] K. S. Roberts, "A new Representation for a line ", *Proceedings of International Conference on Computer Vision and Pattern Recognition*, Ann Arbor, MI, 1988.
- [Sá 88] L. de Sá, J. Dias, V. Silva, "A Modular Approach to Image Processing Through the VMEbus", *Conference ESONE-VMEbus in Research*, ETH, Zurich, Oct. 1988
- [Salisbury 82] J. K. Salisbury and J. J. Craig, "Articulated Hands: Force Control and Kinematic Issues", *International Journal of Robotics Research*, Vol. 1, No. 1, Springer 1982
- [Santos 70] E. Santos, *Formulário de Matemáticas Gerais*, Tipografia Minerva, Montijo, Portugal, 1970.
- [Shen 86] J. Shen, S. Castan, "An optimal linear operator for edge detection", *IEEE Conference on Vision and Pattern Recognition*, 1986
- [Sherman 90] D. Sherman and S. Peleg, "Stereo by Incremental Matching of Contours", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.PAMI-12, No. 11, Nov. 1990
- [Slama 80] C. Slama, C. Theurer and S.W. Henriksen, "Manual of Photogrammetry", *American Society of Photogrammetry*, Falls Church, VA, 1980
- [Shlag 83] J.F. Schlag, A.C. Sanderson, C.P. Neumann, and F.C. Wimberly, *Implementation of Automatic Focusing Algorithms for a Computer Vision System with Camera Control*, Tech. Report CMU-RI-TR-83-14, Carnegie Mellon University, Aug. 1983
- [Srivastava 87] S. Srivastava, N. Ahuja, "An Algorithm for Generating Octrees from Object Silhouettes in Perspective Views", *Proceedings of the IEEE Workshop on Computer Vision*, Nov 30 - Dec 2, 1987
- [Subbarao 87] M. Subbarao, "Direct Recovery of Depth-map I: Differential Methods", *Proceedings of the IEEE Computer Society- Workshop on Computer Vision*, Kiawah Island, SC, 1987

- [Subbarao 88] M. Subbarao and N. Gurumoorthy, "Depth Recovery from Blurred Edges", *Proceedings of the IEEE Computer Vision and Pattern Recognition*, 1988
- [Stevens 81] K. Stevens, "The visual interpretation of surface contours", *Artificial Intelligence*, 17, pg.47-73, 1981
- [Stevens 87] K. Stevens and A. Brooks, "Probing depth in monocular images", *Biological Cybernetics*, 56, pg.355-366, 1987
- [Tabak 88] Y.R.-Tabak and R. Jain, *Building an environment model using depth Information*, CSE-TR-07-88, University of Michigan, USA
- [Tatsuoka 88] M. Tatsuoka, *Multivariate Analysis - Techniques for Educational and Psychological Research*, John-Wiley & Sons, ISBN 0-02-919120-5, 1988
- [Terzopoulos 86] D. Terzopoulos, "Regularization of inverse visual problems involving discontinuities", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI- Vol8, pg. 413-424
- [Thorell 90] L. Thorell and W. Smith, *Using computer color effectively: an illustrated reference*, Prentice-Hall Inc., ISBN 0-13-939878-3, 1990
- [Tikhonov 77] A.N. Tikhonov and V.Y.Arsenin, *Solution of Ill-Posed Problems*, Winston, 1977
- [Tsai 85] R. Tsai, *A Versatile Camera Calibration Technique for High Accuracy 3D Machine Vision Metrology Using off-the-shelf TV Cameras and Lenses*, IBM RC 11413, October 1985
- [Tsai 88] R. Tsai, and R. Lenz, "Overview of a unified calibration trio for robot eye, eye-to-hand, and hand calibration using 3D machine vision", *Proceedings SPIE- Sensor Fusion: Spatial Reasoning and Scene Interpretation*, 7-9 Nov 1988, Cambridge, MA, USA
- [Tweed 90] D. Tweed and T. Vilis, "Geometric Relations of Eye Position Velocity Vectors During Saccades", *Vision Research*, Vol. 30, N° 1, 1990
- [Ullman 79] S. Ullman, *The interpretation of visual motion*, MIT Press, USA, 1979.
- [Ullman 84] S. Ullman, "Maximizing rigidity: The incremental recovery of 3D structure from rigid and non-rigid motion", *Perception*, 13, pg.255-274
- [Verri 87] A. Verri and T. Poggio, "Against quantitative optical flow", *Proceedings of the 1st International Conference on Computer Vision*, UK, pg. 171-180, 1987
- [Waxman 87] A.Waxman and J. Duncan, "Binocular image flows", *Workshop on Motion*, Kiawah Island, SC, 1987
- [Witkin 81] A. Witkin, "recovering surface shape and orientation from texture", *Artificial Intelligence*, 17, pg.17-47, 1981
- [Willson 91] R.G. Willson, and S.A. Shafer, "Active lens control for high precision computer imaging", *Proceedings of IEEE International Conference on Robotics and Automation*, Vol 3, Sacramento, CA, April, 1991
- [Williams 86] C.S. Williams and O.A. Becklund, "OPTICS: A Short Course for Engineers and Scientists", John Wiley & Sons, Inc. USA, ISBN 0-89874-617-5

BIBLIOGRAFIA

- [Wixson 89] L. Wixson and D. Ballard, "Color Histograms for real-time object search", *Proceedings SPIE Sensor Fusion II: Human and Machine Strategies Workshop*, Philadelphia, PA, November, 1989
- [Wrobel-Dautcourt 88] B. Wrobel-Dautcourt, *Perception de la distance par mise en correspondance de régions entre des images stéréoscopiques*, Thèse du Doctorat INPL, INRIA Lorraine, 1988
- [Yachida 86] M. Yachida, Y. Kitamura, and M. Kimachi, "Trinocular Vision: New Approach for Correspondence Problem", *Proceedings 8th International Conference Pattern Recognition*, Paris, Oct. 27-31, 1986, pág. 1041-1044
- [Yarbus 67] A.L. Yarbus, *Eye Movements and Vision*, New York, NY: Plenum, 1967