

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

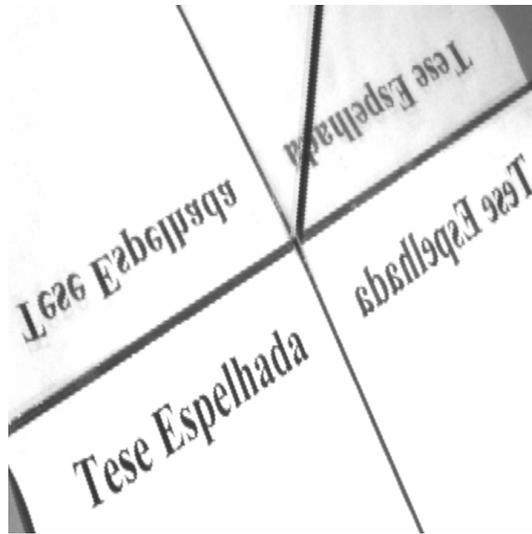
***RECONSTRUÇÃO TRIDIMENSIONAL
MULTI-OCULAR
UTILIZANDO SUPERFÍCIES REFLECTORAS PLANAS***

TESE DE MESTRADO

Nuno Alexandre Cid Martins
Licenciado em Engenharia Electrotécnica

Coimbra
Março de 2003

Dissertação submetida ao
Departamento de Engenharia Electrotécnica
Universidade de Coimbra
para a obtenção do grau de mestre



Trabalho desenvolvido sob orientação do Doutor Jorge Manuel Miranda Dias
Professor Associado do Departamento de Engenharia Electrotécnica da FCTUC

Agradecimentos

Os meus agradecimentos vão para

- o Doutor Jorge Dias, pela sua orientação na realização deste trabalho;
- a Rita, pelo amor demonstrado no seu constante incentivo, quer nas boas como nas más horas. À sua compreensão pela não realização de tantas coisas, o maior obrigada. Adoro-te;
- toda a minha família, pelo carinho e disponibilidade, especialmente à Gioconda pela ajuda nas traduções;
- o José Augusto, pela ajuda na construção do sistema de espelhos para os primeiros testes;
- o Inácio Fonseca, o Amâncio Santos, a Cristiana Areias, o João Alves, o João Ramos e o Eugénio Santos, pelo seu apoio e interesse na ajuda de alguns problemas;
- o Simão Paredes e o Fernando Silva, pela sua disponibilidade, a qual me permitiu ter mais algum tempo para este trabalho;
- o Luís Almeida e o Carlos Simplício, pela partilha de problemas que foram surgindo ao longo deste trabalho;
- todos os colegas do I.S.R. e do D.E.I.S., pela motivação;
- o I.S.R., o I.S.E.C. e o PRODEP, devido ao suporte logístico e financeiro, fundamental na execução deste trabalho;

Resumo

O posicionamento dos objectos no mundo é uma das informações necessárias ao nosso quotidiano. A área da visão por computador pode fornecer esta informação a partir de imagens. A dissertação aqui apresentada insere-se neste domínio.

Entre a multiplicidade de formas de recuperação daquela informação pretende-se, com este trabalho, estudar apenas o problema da reconstrução euclidiana tridimensional. Para atingir esse propósito, o trabalho parte de imagens captadas, sem qualquer outro conhecimento ou informação quanto à sua captação, e apoia-se no potencial das superfícies reflectoras planares. Posteriormente, para melhorar os resultados passa a usar imagens calibradas, captadas de múltiplos pontos de vista.

O desenrolar de todo o trabalho assenta no desenvolvimento de três metodologias de recuperação. Duas delas partem da aquisição de imagens com o auxílio de uma única câmara e de um conjunto de espelhos, verificando-se que com a ajuda dos espelhos se consegue simular qualquer configuração multi-ocular. O passo seguinte, em ambas essas metodologias, e que é, de resto, fundamental para a generalidade dos métodos de reconstrução, é o da correspondência entre imagens. Nesta fase as duas metodologias diferem pois uma delas executa este passo da forma "clássica", enquanto que a outra é inovadora no processo seguido. Nesta última, o processo parte do trabalho de Sameer [Sameer 98] e, com a ajuda dos espelhos, mostra que basta usar as projecções de apenas dois pontos tridimensionais, bem como das suas reflexões, para obter os epipolos nas imagens. A partir daí calcula a matriz fundamental e uma homografia, as quais relacionam as imagens. A fase seguinte é a da abordagem da determinação da métrica euclidiana. Aqui, as duas metodologias recorrem à intersecção dos raios projectivos (triangulação). A diferença está em que uma delas vai buscar

à matriz fundamental os parâmetros necessários a essa intersecção, enquanto que a outra metodologia simplifica o problema, recorrendo novamente aos espelhos. Depois da comprovação geométrica e matemática de ambas as metodologias, apresentam-se, finalmente, os resultados da sua aplicação. Da análise destes resultados conclui-se que uma das metodologias é sensível ao ruído. A outra metodologia consegue melhores resultados apresentando, porém, algumas limitações.

A terceira metodologia surge como forma de ultrapassar estas limitações. Para isso, na fase da determinação da métrica euclidiana, substitui a triangulação por um processo de escavação do espaço tridimensional (*space carving*). Embora prometedores, os resultados obtidos não reconstroem perfeitamente a cena, devido ao escasso número de imagens resultante da utilização do sistema de aquisição usado na metodologia anterior. Verificada esta situação, a terceira metodologia abandona o esquema de captação de imagens utilizada até então (uma câmara e um conjunto de espelhos), passando a usar múltiplas imagens, captadas de diferentes pontos de vista. Depois de calibradas as imagens, a sua utilização no processo de escavação do espaço tridimensional gera resultados extremamente satisfatórios.

Abstract

The positioning of the objects in the world is an information we need every day. The computer vision area can give us this information through the images. This essay can be included in this domain.

Among the multiplicity of processes to recover that kind of information, we selected to study only the problem of the Euclidean tridimensional reconstruction. To achieve this purpose we used planar reflecting surfaces and images without any other information concerning its acquisition. Subsequently, to improve the results, we use calibrated images, collected from several viewpoints.

The unfold of all the study is based on the development of three recovery methodologies. The images in two of that methodologies are acquired with, only, one camera and a set of mirrors. This acquisition system allows the simulation of any multi-ocular configuration. The next step for both methodologies was the correspondence between images, which is essential to many reconstruction methods. In this phase, these two methodologies differ because one of them carries out this step in a "classic" way and the other innovates. This innovation is based on Sameer's work [Sameer 98] and, with the help of the mirrors, it shows that it is enough to use the projections of only two tridimensional points and its reflection, in order to obtain the epipoles in the images. Then, computes the fundamental matrix and one homography, which relates the images. The next phase focus the Euclidean metric determination. Here, both methodologies use the intersection of projective rays (triangulation). The difference between them is that one obtain the necessary parameters from the fundamental matrix and the other uses the mirrors again to simplify the problem. After the geometric and mathematical confirmation of both methodologies, the results of its application

are finally presented. From the analysis of those results, we can conclude that one of the methodologies is noise sensitive. In spite of its limitations, the other methodology achieve better results.

To overcome those limitations we introduced the third methodology. In the Euclidean metric determination phase, it replaces the triangulation by the space carving process. Although promising, the obtained results don't reconstruct perfectly the scene, due to the short number of images resulting from the utilization of acquisition system used in the prior methodology. Being verified this situation, the third methodology leaves the acquisition system used so far (a camera and a set of mirrors), substituting the set of mirrors by multiple cameras. After the calibration of the images (captured from different viewpoints), its utilization in the space carving process generates most satisfactory results.

Conteúdo

	iii
	v
Agradecimentos	vii
Resumo	ix
Abstract	xi
1 Introdução	1
1.1 Enquadramento do trabalho e inserção do problema	1
1.2 Resolução do problema	3
1.2.1 Trabalhos relacionados	7
1.2.2 Objectivos e estrutura do trabalho realizado	12
1.3 Estrutura da dissertação	16
2 Bases para uma recuperação 3D	19
2.1 Introdução	19
2.2 Modelos de formação de imagens	20
2.2.1 Projecção em perspectiva	21
2.2.2 Lentes e focagens	23
2.2.3 Modelos de projecção alternativos	24
2.3 Determinação de informação tridimensional	26

2.3.1	Representação homogénea de coordenadas	27
2.3.2	Sistema de equações relativas ao uso de uma câmara	27
2.3.3	Determinação de coordenadas dos pontos 3D	31
2.4	Geometria epipolar e matriz fundamental	33
2.4.1	Matriz fundamental para configurações específicas	41
2.4.2	Estimação da matriz fundamental	44
2.4.3	Obtenção da distância focal	56
2.5	Transformações projectivas	56
2.5.1	Cálculo de transformações projectivas	57
2.5.2	Homografias infinitas	62
2.5.3	Relações entre a matriz fundamental e as homografias	63
2.6	Superfícies Reflectoras	64
2.6.1	Zona de cobertura dos sistemas de aquisição	66
2.6.2	A geometria epipolar e os espelhos	72
2.7	Métodos numéricos	75
2.7.1	Método dos mínimos quadrados	76
2.7.2	Consenso aleatório de amostras - RANSAC	80
2.7.3	Comportamento dos métodos numéricos apresentados	82
3	Recuperação de pontos 3D	89
3.1	Introdução	89
3.2	Reconstrução centrada na matriz fundamental	90
3.2.1	O sistema de aquisição	91
3.2.2	Correspondência	93
3.2.3	Auto-calibração	95
3.2.4	Reconstrução tridimensional dos pontos	96
3.2.5	Resultados experimentais	97
3.3	Reconstrução centrada nos espelhos	100
3.3.1	O sistema de aquisição	101
3.3.2	Correspondência	103
3.3.3	Reconstrução tridimensional dos pontos	115

3.3.4	Resultados experimentais	128
4	Reconstrução de superfícies	133
4.1	Introdução	133
4.2	Aplicações potenciais	135
4.3	Reconstrução por escavação do espaço 3D	137
4.3.1	Introdução	137
4.3.2	Representação de uma <i>Octree</i>	138
4.3.3	Algoritmo de escavação baseado em silhuetas	140
4.3.4	Extração das silhuetas e teste de intersecção	142
4.3.5	Recuperação das texturas	143
4.3.6	Resultados experimentais	144
5	Conclusões e trabalho futuro	167
5.1	Conclusões	167
5.2	Trabalho futuro	169
A	A qualidade da correspondência	171
A.1	Matriz fundamental	171
A.2	Matriz homográfica	172
B	Rotinas para MATLAB©	173
B.1	Condicionamentos de dados	173
B.2	Qualidade da correspondência	174
B.3	Resolução de sistemas lineares (mínimos quadrados)	175
B.4	Estimação da matriz fundamental	176
B.5	Estimação da matriz homográfica de grau 2	177
B.6	Eliminação de falsas correspondências	178
C	Propriedades das matrizes antisimétricas	179
	Bibliografia	181

Lista de Tabelas

2.1	<i>Diferentes maneiras que a geometria epipolar, na forma da matriz fundamental, \mathbf{F}, toma para configurações específicas (translação pura e movimentos planares).</i>	43
2.2	<i>Conjunto de dados de entrada base, para a comparação entre processos numéricos.</i>	82
2.3	<i>Novo conjunto de dados de entrada para a comparação entre processos numéricos, onde aos dados base, da tabela 2.2, foi acrescentada uma recta errada.</i>	83
2.4	<i>Resultado numérico da intersecção das rectas da tabela 2.4, para cada um dos métodos PSI, SVD e EIG.</i>	84
2.5	<i>Comportamento dos métodos de estimação da matriz fundamental. G refere-se às estimações que utilizam a forma geral método sem condicionamentos. Nas linhas $N+G$ e $G+C2$ acrescentou-se, respectivamente, um condicionamento dos dados de entrada e um dos dados de saída. Nas linhas $N+G+C2$ aplicaram-se ambos os condicionamentos.</i>	88
3.1	<i>Dados métricos fixos usados pelo sistema de aquisição de imagens criado.</i>	92
3.2	<i>Resultados do cálculo dos epipolos. Os primeiros três casos usam o método dos epipolos por construção. O quarto caso estima a matriz fundamental, obtendo, de seguida, o seu espaço nulo.</i>	107
3.3	<i>Resultados da estimação das homografias. A matriz \mathbf{H}_{ed} foi estimada com a primeira equação inserida em 3.7, enquanto que \mathbf{H}_{eda} foi obtida através da segunda equação inserida em 3.7.</i>	112

3.4	<i>Resultados da estimação das matrizes fundamentais. A qualidade de \mathbf{F} e as coordenadas dos epipolos estão em pixels. Os epipolos são obtidos através do cálculo do espaço nulo de \mathbf{F}.</i>	115
3.5	<i>Informação sobre os pontos recuperados do plano 3D. Os valores da sexta e sétima coluna referem-se à distância entre o ponto recuperado e os seus vizinhos de baixo e lateral, respectivamente.</i>	129

Lista de Figuras

1.1	<i>Diagrama das aproximações desenvolvidas com o objectivo de serem criados sistemas de reconstrução 3D multi-oculares. Nas aproximações (a) e (b) a reconstrução baseou-se na triangulação, enquanto que em (c1) e (c2) a base foi a escavação do espaço 3D. Além disso, a aproximação (a) apoiou-se na matriz fundamental, e (b) e (c1) apoiaram-se no potencial dos espelhos.</i>	13
2.1	<i>Projectão em perspectiva. Dado o ponto tridimensional P, no referencial $\{W\}$ e o centro óptico da câmara, O, o ponto p resulta da intersecção entre \overline{OP} com o plano imagem I. A distância focal da câmara é f. Os pontos B e b são, respectivamente, a projecção dos pontos P e p no plano X_cZ_c. Por seu lado, A e a são as respectivas coordenadas Z_c de cada um dos pontos P e p.</i>	22
2.2	<i>Sentido positivo das rotações sobre os eixos de coordenadas.</i>	29
2.3	<i>Transformação das coordenadas (u, v) em coordenadas (x, y).</i>	30
2.4	<i>Geometria estereoscópica. A determinação de P é conseguida através da intersecção dos segmentos de recta $\overline{Pp_e}$ e $\overline{Pp_d}$.</i>	32
2.5	<i>Geometria epipolar. As duas imagens do ponto P têm os centros ópticos nas posições Co_e e Co_d. O plano que contém Co_e, P e Co_d é chamado plano epipolar de P. Este plano intersecta as imagens ao longo das duas rectas epipolares r_e e r_d. Aos pontos e_e e e_d chamam-se epipolos.</i>	34

2.6	<i>(a) Efeito de uma matriz fundamental não singular no cálculo e respectiva representação das rectas epipolares. (b) Efeito de uma matriz fundamental singular no cálculo e respectiva representação das rectas epipolares. Esta matriz é obtida forçando a matriz da figura (a) a ter característica dois.</i>	46
2.7	<i>Modo de funcionamento dos espelhos. O $\triangle ABC_v$ é o maior corte lateral do volume de cobertura da câmara. C_v é o centro óptico da câmara virtual equivalente ao centro óptico C, da câmara real, que resulta da introdução do espelho. C_v fica à mesma distância, também medida na perpendicular, a que C_r está do espelho, do seu lado oposto.</i>	65
2.8	<i>Simetria criada pelo espelho. Um espelho inverte a direcção normal à superfície do espelho da cena 3D. Dependendo do posicionamento do espelho, a inversão pode ser feita da direita para a esquerda (parte esquerda da figura) ou do topo para o fundo (parte direita da figura). Quando ambas as inversões são feitas ao mesmo tempo obtém-se o mostrado na parte central da figura.</i>	67
2.9	<i>Zona de cobertura. Zona da cena 3D que a câmara consegue captar. Ao interior da figura geométrica, produzida pelos feixes de luz limítrofes de entrada no CCD, chama-se volume de cobertura. As intersecções dessa figura geométrica com os planos de medida e do chão designam-se por área de cobertura desses respectivos planos.</i>	68
2.10	<i>Vista perpendicular, em relação ao plano do chão, de uma configuração exemplificativa, a qual resulta de notáveis diferenças de distância de captação e imagem resultante desta configuração. A zona pintada, no esquema, é a área de cobertura segundo esse plano.</i>	69
2.11	<i>Geometria simétrica. Vista perpendicular ao plano do chão da configuração simétrica, escolhida para os espelhos e para a câmara, e imagem resultante do uso desta geometria. θ_e e θ_d transmitem a projecção do tamanho do CCD no plano do chão. T_1 e T_2 são as medidas das larguras dos espelhos pequenos e grandes, respectivamente. A zona pintada, A, é a área de cobertura no referido plano.</i>	70

2.12	<i>Geometria não simétrica. Vista perpendicular ao plano do chão da configuração não simétrica, escolhida para os espelhos e para a câmara, e imagem resultante do uso desta geometria. θ_e e θ_d transmitem o tamanho da projecção do CCD no plano do chão. T_e e T_d são as medidas das larguras dos espelhos. A zona pintada, A, é a área de cobertura no plano do chão.</i>	72
2.13	<i>Obtenção de informação epipolar através dos espelhos. M e M' são dois espelhos planares. Os pontos C_v e $C_{v'}$ são os centros ópticos virtuais, enquanto que C_r é o centro óptico real.</i>	73
2.14	<i>Resultado gráfico da intersecção das rectas, usando os métodos PSI, SVD e EIG. (a) Para a tabela 2.2;(b) Para a tabela 2.3.</i>	83
2.15	<i>Resultado gráfico da intersecção das rectas da tabela 2.4, para os métodos PSI, SVD e EIG, aplicando-lhes o RANSAC.</i>	84
2.16	<i>Resultado gráfico da intersecção de várias rectas, algumas totalmente correctas e outras com pequenos e grandes erros.</i>	85
2.17	<i>Resultado gráfico da intersecção de várias rectas, algumas totalmente correctas e outras com pequenos e grandes erros, aquando da aplicação do RANSAC. (a) Com uma tolerância que permite eliminar as rectas grosseiramente erradas;(b) Com uma tolerância muito pequena.</i>	85
3.1	<i>Imagem do sistema experimental para ser utilizada a aproximação (a).</i>	91
3.2	<i>Corte lateral da configuração seguida na aproximação (a).</i>	92
3.3	<i>Representação do sistema de coordenadas de duas câmaras virtuais e do mundo. O eixo óptico de cada câmara passa em P_f. Nesta configuração $Y_e = Y_d = -Y$ e $X_e Z_e$, $X_d Z_d$ e XZ estão no mesmo plano, ou seja, o sistema não tem ciclorsão.</i>	94
3.4	<i>Imagem resultante da aquisição feita pelo protótipo criado, com a representação dos pontos correspondentes usados para a reconstrução.</i>	97
3.5	<i>Resultado do método de reconstrução aplicado ao objecto representado na figura 3.4. Este resultado é obtido a partir da informação contida nas imagens e respectiva geometria das câmaras virtuais.</i>	98

3.6	<i>Dados sobre a estabilidade da aproximação (a).</i>	99
3.7	<i>Uma perspectiva do sistema utilizado.</i>	101
3.8	<i>Divisão da imagem captada pelo sistema de aquisição.</i>	102
3.9	<i>Construção geométrica do epipolo da imagem direita.</i>	105
3.10	<i>Visão tridimensional da utilização do espelho na obtenção dos pontos q_{4d} e q_d.</i>	105
3.11	<i>Rectas formadas para a obtenção dos epipolos, através da sua intersecção. (a) Caso onde se usam apenas dois pontos; (b) Caso onde se usam quatro pontos; (c) Caso onde se usam oito pontos; (d) Rectas epipolares obtidas através da matriz fundamental, estimada pelo algoritmo dos oito pontos.</i>	108
3.12	<i>Exemplos de obtenção do epipolo pelo método dos epipolos por construção.</i>	110
3.13	<i>Representação dos pontos usados e obtidos na estimação das homografias: (a) Caso dos quatro pontos normalizados; (b) Caso dos oito pontos não normalizados; (c) Caso dos oito pontos normalizados.</i> . .	111
3.14	<i>Rectas epipolares resultantes da estimação das matrizes fundamentais: (a) Estimação com dois pontos; (b) Estimação com quatro pontos; (c) Estimação com oito pontos.</i>	116
3.15	<i>Sistema de coordenadas do sistema de aquisição e reconstrução.</i> . . .	116
3.16	<i>Aspecto gráfico do método de intersecção entre duas rectas não paralelas e não coincidentes.</i>	123
3.17	<i>(a) Imagem não calibrada de um plano 3D, de onde se recuperaram os pontos assinalados com um \times; (b) Resultados da reconstrução planar.</i>	129
3.18	<i>(a) Imagem não calibrada, da qual se recuperaram pontos de diferentes superfícies planares; (b) Resultados da reconstrução, onde sobre os pontos recuperados se pôs o padrão original das superfícies da cena 3D.</i>	131
4.1	<i>Representação de um volume simples através de uma octree.</i>	139
4.2	<i>(a) Imagem original. (b) Silhueta extraída com a ajuda do filtro de canny. (c) Imagem no teste de consistência.</i>	142

4.3	<i>Resultado da recuperação da textura. (a) Uma das imagens usadas; (b) Uso da média da cor das projecções do centro do voxel nas imagens; (c) Uso da média da cor das projecções dos vértices do voxel nas imagens.</i>	143
4.4	<i>Imagens usadas na reconstrução baseada no método das homografias T_{xz} e T_{yz}.</i>	147
4.5	<i>Resultado da binarização das imagens da figura 4.4.</i>	149
4.6	<i>Diferentes vistas do resultado da reconstrução da cena 3D, usando as homografias T_{xz} e T_{yz}.</i>	149
4.7	<i>Cálculo das relações entre os sistemas métricos utilizados (metro e pixels).</i>	152
4.8	<i>Imagens usadas no processo de reconstrução de superfícies.</i>	156
4.9	<i>Posicionamento das câmaras que adquiriram as imagens da figura 4.8.</i>	157
4.10	<i>Imagens binárias usadas no processo de reconstrução de superfícies.</i>	158
4.11	<i>Resultados da escavação por níveis (visualização dos limites dos voxels): (a) Nível 2; (b) Nível 3; (c) Nível 4; (d) Nível 5; (e) Nível 6; (f) Nível 7; (g) Nível 8.</i>	160
4.12	<i>Diferentes vistas do resultados da escavação para nove níveis.</i>	161
4.13	<i>Imagens usadas no processo de reconstrução de superfícies.</i>	162
4.14	<i>Posicionamento das câmaras que adquiriram as imagens da figura 4.13.</i>	162
4.15	<i>Resultados da escavação por níveis: (a) Nível 1; (b) Nível 2; (c) Nível 3; (d) Nível 5; (e) Nível 6; (f) Nível 8.</i>	163
4.16	<i>Diferentes vistas do resultados da escavação para dez níveis.</i>	163
4.17	<i>Posicionamento das câmaras que adquiriram as imagens da figura 4.18.</i>	164
4.18	<i>Imagens usadas no processo de reconstrução de superfícies.</i>	164
4.19	<i>Resultados da escavação por níveis: (a) Nível 1; (b) Nível 2; (c) Nível 3; (d) Nível 4; (e) Nível 5; (f) Nível 6; (g) Nível 7; (h) Nível 8.</i>	165
4.20	<i>Diferentes vistas do resultados da escavação para nove níveis.</i>	166

Capítulo 1

Introdução

Este capítulo tem como objectivos a apresentação do trabalho a ser desenvolvido ao longo da presente dissertação e o seu enquadramento na área de investigação correspondente. Na primeira secção, a ênfase é colocada no enquadramento e na definição do problema que se procurará resolver. A segunda secção destina-se a uma revisão da literatura relativa aos principais contributos para a solução daquele problema e à apresentação do caminho seguido nesta dissertação. O capítulo termina com uma terceira, e última, secção, dedicada à estrutura da dissertação.

1.1 Enquadramento do trabalho e inserção do problema

A visão é de primordial importância na percepção humana e, por isso, tem sido estudada ao longo da história. Por volta de 1960, emergiu uma área de investigação autónoma, denominada visão por computador, cujo principal objectivo era a criação de máquinas com capacidades sensoriais para realizarem tarefas quase sem a intervenção humana. Desde então, esta área, composta por vários campos de pesquisa, nomeadamente o processamento e análise de imagens e o reconhecimento de formas, padrões e objectos, tem sofrido um progresso extraordinário. Algumas das justificações que podem ser apontadas para esse desenvolvimento são o aparecimento do

computador digital aplicado ao processamento de imagem, os avanços tecnológicos e funcionais dos sensores que captam a informação relativa à cena e, também, as técnicas e equipamentos especializados no processamento dessa informação.

Apesar de já se conhecerem os princípios nos quais se baseavam algumas dessas técnicas, só na década de 70 se reconheceu a importância desta área. A partir dessa altura, foram muitos os sistemas desenvolvidos nesta área, sendo as imagens da superfície da terra, tiradas a grande altitude, as imagens radiológicas e as imagens de documentos, disso exemplo. No entanto, estes sistemas não abordavam o aspecto da determinação de parâmetros tridimensionais. Assim, a visão por computador desenvolveu-se, nomeadamente, no sentido de solucionar esta limitação.

Com a abertura à tridimensionalidade apareceram novas técnicas que, directa ou indirectamente, permitiram retirar das imagens mais informação, como o conhecimento de posições tridimensionais relativas a um determinado referencial ou orientações locais das superfícies captadas nas imagens. Com estes avanços foram criados novos métodos para a execução de tarefas, onde a intervenção humana se foi minimizando. A indústria foi um dos sectores que mais beneficiou com estes avanços, já que tarefas como o transporte, a localização, a inspecção ou o reconhecimento de objectos, em geral, puderam ser executadas por máquinas, reduzindo a probabilidade de erro ou até diminuindo o perigo, por motivos como o cansaço ou a distração.

Resumidamente, pode dizer-se que um dos principais objectivos da visão por computador é a inferência de informação tridimensional de uma cena, pela análise de imagens dessa cena e da informação envolvida aquando da sua captação. Este procedimento de inferência, que se designa por reconstrução tridimensional, pode ser descrito de forma sintética como se segue. A informação é captada pelos sensores, tomando a forma física de imagens. Tendo-as como base, e conhecendo as características dos sensores, bem como a geometria envolvida no sistema sensores-cena 3D, torna-se possível obter a informação sobre a estrutura geométrica contida nas imagens, como por exemplo a distância de pontos tridimensionais a um determinado referencial (ou as coordenadas tridimensionais dos pontos) e as orientações locais das superfícies pertencentes à cena 3D. Quando, neste processo, não é executada nenhuma calibração aos sensores para obter as suas características internas (por exemplo

distância focal), diz-se que é feito com base em imagens não calibradas.

Na falta de qualquer tipo de informação *a priori*, um ponto da imagem corresponde a uma infinidade de pontos da cena do mundo 3D. Assim, a reconstrução tridimensional de uma cena do mundo 3D a partir de uma única imagem, sem mais nenhum outro tipo de informações, é teoricamente impossível. Portanto, são necessárias pelo menos duas imagens, obtidas de pontos de vista diferentes, para a obtenção de alguma informação tridimensional. No entanto, este facto só é totalmente correcto quando, entre esses dois pontos de vista existe, pelo menos, uma translação.

A reconstrução tridimensional de um conjunto de pontos de uma cena do mundo 3D, sem qualquer ligação ao modelo da câmara utilizada na captação das imagens, subdivide-se nos seguintes tipos: reconstrução projectiva, reconstrução afim e reconstrução euclidiana. A reconstrução projectiva é definida a menos de uma transformação projectiva e a informação obtida não contém o conhecimento sobre ângulos, distâncias ou paralelismos [Mundy 92][Rothwell 95]. A reconstrução afim é definida a menos de uma transformação afim e a informação obtida contém o conhecimento sobre as relações entre distâncias e paralelismos [Zhang 97]. A reconstrução euclidiana é definida a menos de uma semelhança e, em termos de informação obtida, é a mais rica de todas. Ela contém informações que o homem habitualmente manipula [Devernay 95].

Sem se especificar o tipo de reconstrução, a expressão reconstrução tridimensional é genérica e diz respeito às três reconstruções citadas anteriormente. Por abuso de linguagem, os termos reconstrução e estrutura são por vezes utilizados para designar a reconstrução tridimensional.

Ao longo desta dissertação, o problema que se procurará resolver é o da recuperação euclidiana de informação tridimensional, contida em imagens.

1.2 Resolução do problema

Os estudos sobre a reconstrução tridimensional têm dado origem a uma vasta série de técnicas de aquisição de informação tridimensional, nomeadamente técnicas

baseadas no princípio da triangulação ou em imagens monoculares, técnicas de Moiré, técnicas de detecção de ecos e técnicas baseadas no princípio da escavação do espaço tridimensional. Exemplos de algumas destas técnicas podem ser vistas, nomeadamente, em [Dias 94], [Silva 94], [Wong 01] e [Kutulakos 99].

Já que todas as técnicas têm vantagens e desvantagens, a sua escolha depende da aplicação, da precisão ou da resolução. Neste trabalho usar-se-ão apenas técnicas baseadas no princípio da triangulação, mais especificamente, as técnicas de estereoscopia passiva e técnicas baseadas na escavação do espaço tridimensional.

O princípio da triangulação [Hartley 95a] está presente sempre que se estabelece a correspondência entre pontos de diferentes imagens, que são a projecção do mesmo ponto da cena, e se determina o ponto de intersecção das rectas, formadas por cada uma das correspondências e pelos respectivos centros ópticos, a fim de se conseguir as coordenadas tridimensionais do ponto da cena. As técnicas baseadas neste princípio têm que adquirir pelo menos duas imagens da cena, a partir de posições diferentes, ou adquirir no mínimo uma imagem da cena, a qual capta também a projecção de um padrão de luz sobre a cena, cujo centro de projecção é conhecido.

Além da técnica de estereoscopia passiva, encaixam-se no grupo das técnicas baseadas no princípio da triangulação as técnicas estereoscópicas activas e as técnicas baseadas em sombras. A diferença entre as técnicas passivas e activas está no facto das primeiras utilizarem apenas a luz ambiente para iluminar a cena a captar, enquanto que as segundas são caracterizadas pela projecção directa de luz ou ultrasons sobre a cena.

A maior parte da investigação relacionada com a estereovisão tem sido dedicada ao problema da correspondência entre os pontos das imagens. O processo de obtenção de correspondências inicia-se com a detecção, para cada imagem, de pontos característicos, ou seja, pontos que, por alguma razão, se destacam dos outros pontos da imagem (como exemplo temos vértices e pontos pertencentes às arestas e às orlas dos objectos). Depois, efectua-se a correspondência local entre as imagens, isto é, tenta-se encontrar os pontos nas imagens que constituem a projecção do mesmo ponto tridimensional. Por fim tenta-se, retirar todas as falsas correspondências.

A técnica de estereoscopia passiva tem como vantagem a sua aplicabilidade a um

grande número de situações, não precisando de equipamentos especiais nem de uma elevada capacidade de processamento computacional.

Por seu lado, o princípio da escavação do espaço tridimensional baseia-se na ideia de que qualquer cena tridimensional pode ser representada volumetricamente através de agregações de primitivas elementares, vulgarmente conhecidas como *voxels* ou *pixels* 3D [Kutulakos 99] [Wong 01]. Cada *voxel* representa um pequeno volume do espaço tridimensional, podendo estar ocupado (ou não) e reter características correspondentes às propriedades físicas da cena nesse volume elementar.

A aplicação deste princípio começa por definir o espaço volumétrico que envolve toda a cena a reconstruir, e, em cada iteração do processo de escavação, esse espaço volumétrico vai sendo reduzido até se obter as superfícies pretendidas. A redução é feita através da eliminação de *voxels* que não são consistentes com as imagens captadas. Assim, no resultado final da reconstrução todos os pontos dentro do espaço volumétrico obtido são consistentes com essas imagens. Um ponto tridimensional, pertencente ao espaço volumétrico, é consistente com um ponto numa imagem se a sua cor resultar da radiação do ponto tridimensional [Kutulakos 99]. Isto só é válido para modelos de reflexão que são calculados localmente. Para evitar o cálculo dos modelos de luz envolvidos na captação das imagens, o teste de consistência pode ser feito verificando se a projecção do ponto 3D, do espaço volumétrico, está delimitada pelas silhuetas da cena em todas as imagens [Wong 01]. Optou-se, nesta dissertação, pelo teste de consistência dos *voxels* baseado nas silhuetas, sempre que se usar este princípio.

As silhuetas (ou contornos) são, muitas vezes, uma característica dominante nas imagens e podem ser extraídas com relativa facilidade. Fornecem informações importantes no que diz respeito à forma dos objecto e são, de facto, a única informação disponível no que toca a superfícies suaves quase sem textura [Wong 01]. No entanto, a correspondência entre pontos de várias silhuetas não é uma dessas informações, devido à sua dependência dos centros ópticos.

É possível configurar geometricamente o sistema de aquisição de diferentes maneiras. Uma das configurações geométricas mais populares consiste em colocar todas as câmaras, que se estão a usar, paralelas entre si, fazendo com que os seus respectivos

planos imagem pertençam a um mesmo plano (geometria fronto-paralela). Esta configuração facilita as correspondências entre pontos das imagens. Outra configuração coloca as câmaras em posições não coplanares, apesar de estarem a olhar para a mesma cena do mundo 3D. Esta última configuração é considerada como o caso geral para todas as configurações possíveis. Existe ainda uma configuração intermédia, na qual se mantêm as câmaras no mesmo plano, mas com os eixos ópticos a convergir para um mesmo ponto da cena 3D. A este ponto tridimensional, onde todos os eixos ópticos se cruzam, chama-se ponto de fixação, e a sua projecção deverá estar na mesma posição em ambas as imagens. Esta geometria pode vir a ser explorada com êxito em sistemas de visão activa, para reconstrução tridimensional [Francisco 94] [Almeida 99].

Tal como já foi referido, este trabalho irá usar o princípio da estereovisão passiva e o princípio da escavação do espaço tridimensional, baseado em silhuetas. No entanto, não se pretende que esta dissertação seja mais um estudo sobre estas técnicas, mas sim uma descrição sobre a criação de sistemas práticos, que usam estas técnicas para solucionar o problema da reconstrução tridimensional. Outra ideia que se quis explorar neste trabalho foi a dos sistemas catadióptricos [Hecht 74], capazes de simular multi-ocularidade.

O uso de sistemas de visão catadióptricos é um campo que está a ganhar cada vez mais importância na área de visão computacional. Estes sistemas resultam da utilização combinada de imagens refractadas (obtidas através do uso de lentes) e de imagens reflectidas (obtidas através do uso de espelhos), onde as lentes e os espelhos estão posicionados em cuidadas configurações, de forma a capturar o maior campo de visão possível.

Os sistemas catadióptricos ultrapassam a limitação do campo de visão dos sistemas convencionais de aquisição, permitindo colocar numa única imagem, sincronizada, toda a informação que só seria possível adquirir através do uso de várias câmaras. No que se refere ao aumento do campo de visão, fundamental nos processos de reconstrução tridimensional, são vários os autores que o demonstraram, como por exemplo [Rees 70] [Nayar 88] [Goshtasby 93]. Outro ponto importante a favor destes sistemas é o da facilidade com que se pode fazer com que eles possuam um único centro de

projecção [Baker 98] [Sameer 98]. Esta vantagem é bastante importante, pois permite capturar imagens com um modelo de perspectiva único. As imagens daqui resultantes podem ser processadas normalmente como se fossem, todas elas, obtidas por diversas câmaras com parâmetros iguais (tais como as lentes e sua distorção, o CCD, a focagem, a resposta espectral, a sincronização da captura, o controlo automático do ganho, o tamanho dos pontos, etc). Para além disso, estes sistemas permitem uma redução dos custos, através da substituição das câmaras por espelhos.

Como é óbvio, poderão ser utilizados em sistemas catadióptricos várias câmaras e espelhos, estes com formas diferentes. No entanto, este trabalho irá limitar-se, no que diz respeito aos sistemas catadióptricos, ao uso de uma única câmara e um conjunto de espelhos planares.

1.2.1 Trabalhos relacionados

O problema da reconstrução tridimensional pode ser resolvido, classicamente, em duas etapas: calibrando os sensores e reconstruindo a cena. A calibração de um sensor é um processo delicado, que precisa de utilizar pontos conhecidos no espaço tridimensional para obter as suas características [Faugeras 87] [Dias 91] [Faugeras 92] [Dias 94] [Silva 94] [Batista 99]. No entanto, este processo nem sempre é de possível realização, o que acontece, por exemplo, quando a câmara se encontra envolvida em qualquer tipo de movimento. Mais recentemente, a recuperação da estrutura tridimensional de uma cena passou a ser feita numa única etapa, evitando a calibração explícita dos sensores, ou seja, passou-se apenas a usar a informação contida nas imagens não calibradas e o conhecimento da geometria dos sensores-cena 3D.

Quando os únicos dados passíveis de utilização são as observações nas imagens não calibradas, o problema da reconstrução tridimensional torna-se complexo. No entanto, se a questão da correspondência entre zonas nas duas imagens estiver resolvida, é diminuída a complexidade da reconstrução.

Nos estudos realizados por Marr com imagens estéreo, a questão da correspondência ficou perfeitamente definida [Marr 76] [Marr 79]. As várias soluções, que têm sido propostas para obter uma boa correspondência entre imagens, distinguem-se

pelo modo como realizam o teste de semelhança entre as zonas nas duas imagens. Assim, a correspondência pode ser realizada com base em zonas de grande intensidade luminosa, correspondentes a arestas ou a contornos [Pollard 85], ou utilizando regiões com características idênticas [Wrobel 88]. A correspondência pode, no entanto, ser facilitada através de restrições geométricas, às quais devem obedecer as posições das zonas correspondentes nas imagens. Essas restrições poderão ser gerais, tal como a geometria epipolar [Zhang 96], ou então particulares, tais como a restrição de fixação [Martins 98] ou as provocadas pela configuração do sistema.

Em termos de resultados de reconstrução sem calibração da câmara, os trabalhos mais significativos que foram levados a cabo têm como ponto comum a utilização de uma base de referência.

Um dos primeiros trabalhos deste tipo foi desenvolvido por Koenderink e Doorn [Koenderink 89]. Estes autores mostraram que a reconstrução tridimensional afim de uma cena poderia ser calculada a partir de, pelo menos, duas imagens. Para isso foram utilizados quatro pontos arbitrários da cena (não coplanares) como base de referência afim, sendo a reconstrução de todos os outros pontos calculada relativamente a esta base.

Tomasi e Kanade [Tomasi 91] calcularam a estrutura tridimensional da cena e a rotação da câmara. As disparidades dos pontos nas imagens foram primeiramente colocadas sob a forma de uma matriz. Depois, utilizando a decomposição em valores singulares (SVD), definiram duas matrizes, uma contendo informação sobre a estrutura tridimensional e a outra contendo a informação sobre a rotação da câmara. Neste método, é necessário conhecer os parâmetros intrínsecos da câmara, de forma a obter uma reconstrução tridimensional euclidiana da cena.

O trabalho proposto por Weinshall [Weinshall 93] consistiu em calcular, em primeiro lugar, a reconstrução tridimensional afim, através do método de Koenderink e Doorn. De seguida, foi calculada a matriz de Gram (para quatro pontos da cena, não coplanares), a partir da qual foi extraída a reconstrução euclidiana desses pontos. Assim, este trabalho provou que se pode passar de uma estrutura tridimensional afim, já calculada, para uma euclidiana sem que seja necessário conhecer os parâmetros intrínsecos da câmara.

Poelman e Kanade [Poelman 94] propuseram uma variante ao método de Tomasi e Kanade, na qual melhoraram a qualidade das reconstruções. Tal foi conseguido devido ao facto de terem usado um melhor modelo para as câmaras CCD, pois este modelo permite registar as variações de profundidade, ou seja, a câmara (ou a cena 3D) pode efectuar translações ao longo do eixo óptico.

Até aqui, faltava a todos estes trabalhos uma maior proximidade com a câmara real, em termos de modelos matemáticos que a descrevem. Assim, o tratamento do problema da reconstrução de uma maneira mais correcta incitou vários investigadores a abandonar a geometria afim, em benefício da geometria projectiva.

Uma primeira aproximação realista neste sentido tinha já sido proposta por Sparr [Sparr 91]. Neste trabalho o autor desenvolveu um descritor para a estrutura tridimensional afim de um conjunto de pontos. Este descritor é independente da base afim utilizada e das propriedades intrínsecas da estrutura 3D, e permite obter a reconstrução afim relativa ao conjunto dos pontos. Para além disso, Sparr mostrou que o descritor está ligado a um modelo matemático que descreve a câmara com muita precisão. No entanto, o método apresentado por este trabalho só é útil quando é requerida apenas informação afim para uma dada aplicação. A resolução do sistema, altamente não linear, para se obter a reconstrução continua a ser o maior problema, limitando assim a sua utilização.

Mohr *et al* [Mohr 90] [Mohr 91] encontram-se entre os pioneiros na utilização das propriedades da geometria projectiva com vista a evitar uma calibração explícita da câmara. A sua aproximação baseou-se na reprojecção dos pontos das imagens em dois planos da cena, e na utilização da propriedade da invariância das coordenadas projectivas. Seis pontos conhecidos da cena foram utilizados como pontos de referência, sendo os outros reconstruídos em relação àqueles.

Mais tarde, Faugeras [Faugeras 92] desenvolveu um método linear para a reconstrução projectiva, unicamente a partir de pontos correspondentes nas duas imagens. Ele escolheu cinco pontos arbitrários da cena (quatro a quatro não coplanares) como base projectiva, fazendo-lhes corresponder as coordenadas homogêneas $(1, 0, 0, 0)$, $(0, 1, 0, 0)$, $(0, 0, 1, 0)$, $(0, 0, 0, 1)$ e $(1, 1, 1, 1)$. Graças a esta escolha e à correspondência entre as duas imagens, ficou só com duas, das vinte e duas, incógnitas das matrizes

de projecção para determinar. Para o fazer, o autor supôs que os epipolos eram conhecidos. Como estes estão intrinsecamente ligados às matrizes de projecção das imagens, pôde, assim, determinar inteiramente as duas matrizes de projecção e, por consequência, obteve uma reconstrução projectiva sem nenhuma informação métrica, nem mesmo afim.

Na mesma altura, Hartley *et al* [Hartley 92] obtiveram um resultado similar, uma reconstrução projectiva. Outros investigadores acabaram por propor métodos lineares, baseados também na geometria epipolar, para o cálculo da reconstrução projectiva de uma cena. Todos estes métodos podem ser considerados como similares porque utilizam as mesmas propriedades da geometria projectiva.

No que diz respeito à utilização de sistemas catadióptricos na reconstrução tridimensional, Nayar [Nayar 88] sugeriu um sistema estéreo com um campo de visão alargado (FOV), que consistia numa câmara convencional apontada para duas esferas especulares. O autor mostrou que, estabelecendo a correspondência entre as imagens da cena, que eram reflectidas pelas esferas, era possível, por triangulação, obter a profundidade. Mais tarde, generalizou esta ideia de forma a incluir n elementos reflectidos de formas arbitrárias. Foram também desenvolvidos outros sistemas similares onde as esferas foram substituídas por outros objectos. Estes sistemas tinham a grande desvantagem de não terem um único centro de projecção, aquando da aplicação de variadas técnicas.

Mais tarde, Baker e Nayar [Baker 98] deduziram as configurações dos sistemas catadióptricos estéreo que, quando vêm a cena em perspectiva, mantêm o centro de projecção. Baseada neste trabalho, foi então proposta uma variedade de sistemas que usam uma única câmara e dois ou mais espelhos [Sameer 98]. A partir do resultado de Sameer, foi conseguida, e usada nesta dissertação, a aquisição de duas ou mais projecções da cena na mesma imagem.

Mitsumoto *et al* [Mitsumoto 92] foram os primeiros a descrever um método de recuperação de profundidade, adquirindo na mesma imagem a projecção e a sua reflexão, usando espelhos planares. Estes espelhos foram usados para obter a imagem das partes ocultas da cena.

Por fim, destacam-se os trabalhos [Goshtasby 93] e [Mathieu 95], onde é usada

uma única câmara, apontada para um conjunto articulado de espelhos planares, para captar, de uma só vez, um par de imagens estéreo.

A quantidade de estudos feitos sobre a reconstrução baseada na calibração mostra que o problema que tem sido abundantemente tratado. Assim, devido ser uma das opções de trabalho, confinamos a revisão dos trabalhos sobre esta matéria apenas à reconstrução feita a partir de silhuetas.

A reconstrução de superfícies de objectos suaves a partir de silhuetas foi iniciada por Giblin e Weiss [Giblin 87]. Assumindo a projecção ortográfica, eles demonstraram que uma superfície pode ser reconstruída através de todos os seus planos tangentes, calculados directamente a partir da família de silhuetas da superfície, sob um movimento planar. Cipolla e Blake [Cipolla 92] alargaram o estudo de Giblin e Weiss à projecção em perspectiva, sob um movimento curvilíneo, e desenvolveram o método conhecido por *osculating circle*, introduzindo o conceito de parameterização epipolar. Vaillant e Faugeras [Vaillant 92] desenvolveram uma técnica similar, na qual a superfície é parametrizada pelas curvas radiais em vez de ser pelas curvas epipolares. Baseando-se no método *osculating circle*, Szeliski e Weiss [Szeliski 98] calcularam de forma linear as curvas epipolares em toda a superfície, juntamente com uma estimativa da incerteza, e os resultados da reconstrução melhoraram.

Em [Boyer 97], Boyer e Berger desenvolveram as derivadas da fórmula para a profundidade, a partir de uma aproximação local da superfície, até à ordem dois, para movimento discreto. Em [Wong 99], Wong *et al* desenvolveram uma técnica simples baseada na implementação de diferenças finitas de [Cipolla 92]. Apesar da sua simplicidade, o método desenvolvido em [Wong 99] obteve resultados comparáveis aos conseguidos em [Cipolla 92] e [Boyer 97].

As técnicas de intersecção de volume para a construção de modelos volumétricos de objectos, baseadas em visão multi-ocular, foram primeiramente propostas por Martin e Aggarwal [Martin 83], que introduziram a representação de segmentos de volume.

Em [Chien 86], Chien e Aggarwal apresentaram um algoritmo para a geração de uma *octree* de um objecto, a partir de três vistas ortogonais sob uma projecção ortográfica. O seu trabalho foi mais tarde desenvolvido por Ahuja e Veenstra [Ahuja 89], que conseguiram que o algoritmo lidasse com imagens de um qualquer subconjunto

de treze pontos de vista padrão.

Em [Hong 85], Hong e Shneier conceberam uma técnica para a criação de uma *octree* a partir de múltiplas vistas arbitrárias, sob uma projecção em perspectiva. A sua abordagem constrói primeiramente uma *octree* para cada imagem, projectando os cubos da *octree* em cada imagem e intersectando as suas projecções com a silhueta. A *octree* final do objecto é dada pela intersecção das *octrees* obtidas por todas as imagens. Em [Potmesil 87], Potmesil descreveu uma abordagem similar, na qual as imagens são representadas por *quadtrees* a fim de facilitarem as projecções do cubo com as silhuetas. Noutras abordagens similares, onde se incluem [Noborio 88] e [Srivastava 90], a *octree* para cada imagem é construída pela intersecção, num espaço tridimensional, dos cubos da *octree* com o cone de projecção das silhueta.

Em [Szeliski 93], Szeliski apresenta um algoritmo eficiente que constrói uma *octree* de uma forma hierárquica, indo da forma mais grosseira até atingir a perfeição. A sua abordagem é semelhante à de [Potmesil 87], com a excepção de ser construída apenas uma *octree* utilizando-se todas as imagens simultaneamente.

1.2.2 Objectivos e estrutura do trabalho realizado

Conforme já foi referido, o trabalho que vai ser desenvolvido nesta dissertação centra-se na resolução do problema da extracção de informação tridimensional euclidiana a partir de imagens, e tem como objectivos gerais:

- agrupar um conjunto de conhecimentos relacionados com este problema, mais especificamente, conhecimentos sobre os modelos de formação de imagens, a correspondência entre imagens e os processos de triangulação e escavação do espaço tridimensional;
- juntar toda a teoria inerente às superfícies reflectoras, de maneira a retirar o máximo de potencial do uso destas superfícies em sistemas de visão por computador;
- compilar os principais métodos numéricos lineares ligados à resolução do problema da determinação de informação 3D;

- mostrar o desenvolvimento do trabalho de criação de sistemas robustos, para a determinação de informação tridimensional a partir de imagens.

Centrando-nos um pouco mais pormenorizadamente no desenvolvimento do sistema de recuperação de informação tridimensional, passemos à sua descrição geral, ilustrada na figura 1.1. De maneira a se atingir a meta da criação de um sistema ro-

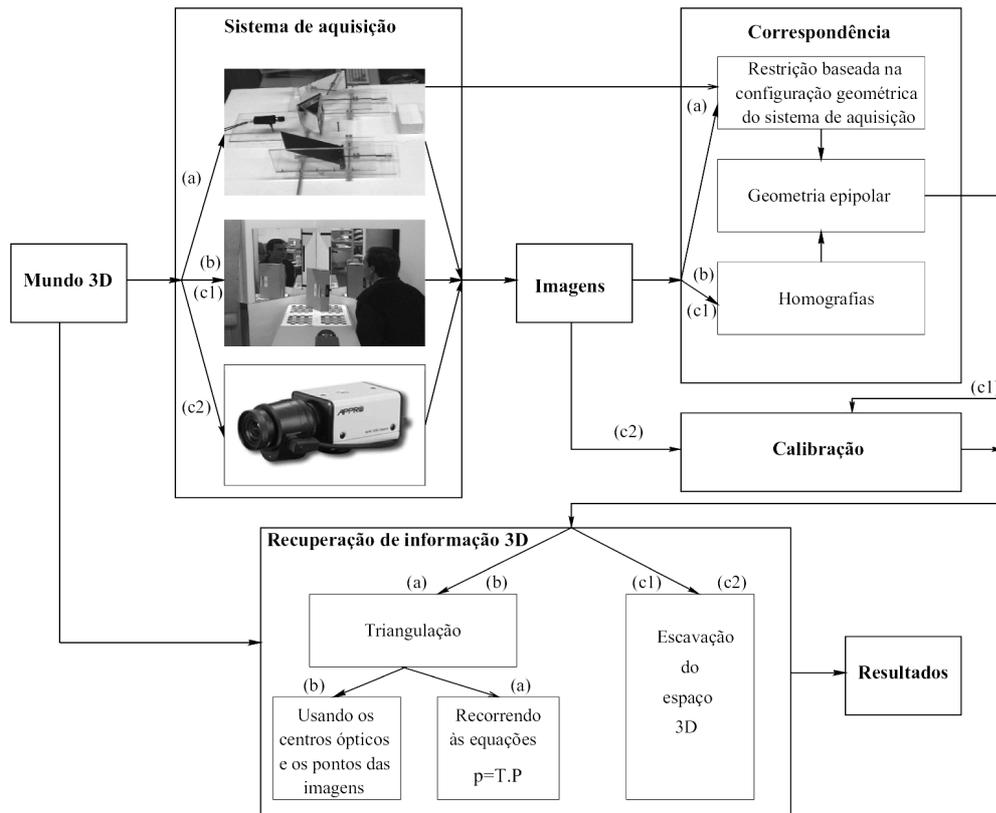


Figura 1.1: Diagrama das aproximações desenvolvidas com o objectivo de serem criados sistemas de reconstrução 3D multi-oculares. Nas aproximações (a) e (b) a reconstrução baseou-se na triangulação, enquanto que em (c1) e (c2) a base foi a escavação do espaço 3D. Além disso, a aproximação (a) apoiou-se na matriz fundamental, e (b) e (c1) apoiaram-se no potencial dos espelhos.

busto, começou-se por desenvolver a aproximação designada no diagrama da figura 1.1 por (a). Nesta aproximação fez-se uma recuperação de informação tridimensional essencialmente baseada na matriz fundamental. Depois de testada esta aproximação,

verificou-se que apesar da sua validade sofria de instabilidade. Assim, com o conhecimento e a experiência adquiridos, avançou-se para a aproximação (b), a qual provou ser bastante mais robusta, apresentando, porém, limitações na capacidade de reconstrução. De forma a ultrapassar essas limitações alterou-se, nesta aproximação, o módulo de recuperação de informação 3D, dando origem à aproximação (c1). Nestas duas últimas aproximações, o processo de reconstrução foi apoiado e facilitado pelo potencial dos espelhos. Embora prometedores os resultados obtidos pela aproximação (c1) não reconstruíam perfeitamente a cena devido, por um lado, ao escasso número de pontos 2D e 3D utilizados no processo de calibração, por outro, ao ruído existente na correspondência entre a cena 3D e a imagem e, por fim, ao limitado número de imagens da cena tridimensional que é possível obter. Verificada esta situação, surgiu a aproximação (c2), a qual abandonou o esquema de captação de imagens utilizado até então (uma câmara e um conjunto de espelhos), passando a usar múltiplas imagens captadas de diferentes pontos de vista. Os resultados obtidos por esta aproximação revelaram-se extremamente satisfatórios. Apesar das aproximações (c1) e (c2) passarem ambas por um processo de calibração das imagens, o método seguido na aproximação (c1) é inovador.

Como se pode verificar do diagrama da figura 1.1, as cenas do mundo real são captadas em imagens, em todas as aproximações, através de um sistema de aquisição. Para as aproximações (a), (b) e (c1), este sistema é composto apenas por uma câmara e um conjunto de espelhos, cuidadosamente posicionados de forma a cumprirem a sua missão. Esta missão passa pela redução do número de câmaras a usar na aquisição das imagens, pois a função dos espelhos é a de multiplicar a única câmara usada, simulando o uso de duas ou mais câmaras reais, como se comprovará mais à frente. No entanto, em fases posteriores, o potencial dos espelhos é novamente aproveitado para o cálculo da correspondência (aproximação (a)) ou mesmo para a obtenção da informação 3D, propriamente dita (aproximações (b) e (c1)). Para a aproximação (c2) o sistema de aquisição de imagens é composto apenas por um conjunto de câmaras.

Uma das diferenças entre as aproximações (a) e (b)/(c1), aparece no módulo da correspondência. A correspondência entre as imagens é uma das principais limitações dos sistemas de reconstrução de informação tridimensional, porque o custo computa-

cional para que essas correspondências se estabeleçam é geralmente elevado. Por vezes, surgem falsas correspondências, dando lugar a informação tridimensional errada. Estas falsas correspondências devem-se ao facto de não se encontrar os pontos correspondentes nas outras imagens, quer por impossível identificação, quer porque os pontos podem ser visíveis numas imagens mas não nas outras (problema das zonas ocultas). O problema da ocultação poderá ser minorado pela diminuição da distância entre as câmaras ou pela utilização de mais câmaras. No entanto, na primeira solução diminui-se a precisão da informação, enquanto que na segunda, embora melhore a precisão, a solução torna-se mais complexa e dispendiosa, com tempos de processamento mais elevados. A aproximação (c2) não necessita deste módulo.

O presente trabalho sugere algoritmos que, de uma forma rápida e robusta, minimizam o problema das correspondências. Estas são feitas com base na geometria epipolar, introduzindo conceitos como a relação entre esta geometria e o posicionamento das câmaras, para a aproximação (a), e a relação entre planos bidimensionais e a geometria epipolar para as aproximações (b) e (c1). A geometria epipolar é algo que está sempre presente quando existem duas imagens da mesma cena, captadas de posições diferentes. Com a restrição que é criada devido a esta geometria, o espaço de procura de um ponto correspondente de um dado ponto da imagem é drasticamente reduzido. Se o correspondente desse ponto escolhido estiver na outra imagem, ele terá que pertencer apenas a uma recta, com parâmetros conhecidos, à qual se chama recta epipolar.

Tal como se pode verificar na figura 1.1, o módulo da recuperação de informação 3D é feito recorrendo às técnicas da triangulação e da escavação do espaço tridimensional. Aqui, a diferença entre as aproximações (a) e (b) aparece no grau de simplificação que se lhe aplica. Na aproximação (a) a informação tridimensional é obtida utilizando a formulação geral da triangulação (sem qualquer simplificação), que geometricamente consiste na intersecção de três planos. Por outro lado, na aproximação (b), a triangulação fica reduzida a uma simples intersecção de rectas 3D. Esta redução deve-se ao uso do potencial dos espelhos no cálculo dos centros ópticos 3D das câmaras, real e virtuais. As rectas tridimensionais que se irão intersectar serão formadas pelos centros ópticos obtidos e pelos pontos correspondentes nas diferentes

imagens. Por outro lado, a diferença entre as aproximações (b) e (c1) está apenas no processo final de determinação da estrutura tridimensional, trocando o processo de triangulação pelo da escavação do espaço tridimensional. A aproximação (c2) utiliza o mesmo processo que a (c1), residindo a diferença nos dados fornecidos a esse processo. Assim, enquanto que (c1) faz uma calibração implícita na geometria dos espelhos e no processo de correspondência, (c2) faz uma calibração explícita, pelo modo clássico.

No desenrolar desta dissertação será possível verificar que o posicionamento do conjunto de espelhos, em relação à câmara, restringe fortemente o problema. Permite, assim, reduzir a complexidade das fórmulas matemáticas, inerentes a todo o processo de determinação da informação 3D, e aumentar a robustez dos sistemas que utilizam os espelhos.

Os sistemas, criados com as características referidas, combinam uma série de equipamentos e de programas capazes de produzir a informação 3D pretendida. Os equipamentos que integram o sistema são um computador, uma câmara, um conjunto de espelhos e um dispositivo de aquisição e processamento de imagens. Os programas que fazem a inter-acção com o utilizador são feitos em linguagem C++, a qual recorre às bibliotecas gráficas do OpenGL, e em MATLAB®.

Em termos futuros, os sistemas desenvolvidos poderão ser aproveitados na realização de aplicações de controlo de qualidade e de controlo dimensional no meio industrial, devido às suas características de baixo custo e precisão.

1.3 Estrutura da dissertação

Após este capítulo de introdução, esta dissertação prossegue com um capítulo de natureza teórica que servirá de base ao trabalho.

O capítulo 2 abordará conceitos gerais sobre os vários modelos de formação de imagens, o processo clássico para a recuperação de informação tridimensional, as restrições que facilitam o problema da correspondência, o uso de superfícies reflectoras em sistemas de aquisição de imagens e os métodos numéricos lineares, ligados à res-

olução dos sistemas de equações relativos ao problema da determinação de informação tridimensional.

No capítulo 3 é apresentado, pormenorizadamente, o desenvolvimento de duas aproximações à criação de sistemas para a determinação de informação 3D, baseados em espelhos, a partir de imagens não calibradas. Estes sistemas são suportados teoricamente pelos conceitos apresentados no capítulo anterior. A aproximação apresentada na secção 3.2 deu origem ao artigo [Martins 98]. Por seu lado, o sistema que se descreve na secção 3.3 deu origem aos artigos [Martins 99] e [Martins 00].

O capítulo 4 começa por descrever um algoritmo baseado no princípio da escavação do espaço tridimensional e nas silhuetas da cena 3D. De seguida faz a sua aplicação, em troca com a triangulação, à última aproximação vista no capítulo anterior. Por fim, abandonando o modo de aquisição de imagens baseado nos espelhos, apresenta-se um sistema prático e completo, no que diz respeito à reconstrução tridimensional de modelos de cenas do mundo 3D a partir de uma sequência de silhuetas bidimensionais. Estas silhuetas pertencem a imagens previamente calibradas. Em termos globais o assunto apresentado neste capítulo ainda proporcionou a execução do artigo [Martins 01].

Por fim, no capítulo 5, serão apresentadas as principais conclusões do trabalho e sugeridas linhas de aplicação futuras.

Capítulo 2

Bases para uma recuperação 3D

Neste capítulo passa-se em revista alguns dos princípios que, sendo a base de qualquer estudo sobre recuperação de informação tridimensional a partir de imagens, irão suportar também todo este trabalho.

2.1 Introdução

Os primeiros conceitos a definir deverão ser, obviamente, os que envolvem as imagens e a recuperação tridimensional a partir dessas imagens. Assim, considera-se que uma imagem é a representação bidimensional de uma cena de um mundo tridimensional. A informação tridimensional obtida a partir de imagens de uma cena, por seu lado, consiste num conjunto de coordenadas tridimensionais dos pontos ou de orientações das superfícies da cena, visíveis nas imagens. Esta informação é geralmente representada sob a forma de uma matriz designada por mapa de profundidades ou imagem $2D\frac{1}{2}$. Os elementos desta matriz representam a distância, dos pontos 3D da cena, a um determinado referencial, previamente definido.

A necessidade da determinação deste tipo de informação fez com que se desenvolvessem vários modelos matemáticos que representam a maneira como as imagens são formadas e como a informação tridimensional pretendida é extraída. Estes modelos apresentam diferentes níveis de elaboração e sofisticação conforme as necessidades de precisão das várias aplicações.

Na segunda secção deste capítulo será abordado o processo de formação de imagens, onde se focarão os modelos lineares que o descrevem. Na terceira secção far-se-á uma abordagem clássica do processo de determinação de informação tridimensional baseada em imagens. Este processo é modelado matematicamente através de um sistema de equações, podendo o sistema ser simplificado através de restrições ao problema inicial. A quarta secção tratará a restrição da geometria epipolar. Será também revista, nesta secção, a forma que esta restrição toma para algumas configurações específicas das câmara, nomeadamente, translação pura, fixação visual, câmaras afins e movimentos planares. Por sua vez, a quinta secção abordará alguns conceitos sobre transformações projectivas, baseando-se especificamente nas relações homográficas entre os planos imagem. Na sexta secção, a atenção incidirá sobre certos aspectos que advém da introdução de superfícies reflectoras no processo de aquisição das imagens. O capítulo termina com uma sétima secção onde serão revistos alguns métodos numéricos que são necessários para o desenvolvimento do trabalho apresentado nesta dissertação.

2.2 Modelos de formação de imagens

Antes de se abordar o processo de determinação de informação tridimensional contida numa imagem, há que conhecer os modelos matemáticos que representam a formação dessa imagem. No nosso caso, o modelo utilizado é linear, pois analisa a formação de uma imagem, com base no princípio da geometria óptica clássica. Esta baseia-se no conceito de raio luminoso rectilíneo e no modo como este atravessa a lente. A alternativa seria a utilização de teorias da óptica física e da difracção que utilizam equações não-lineares para chegarem à mesma explicação. Contudo, e apesar da primeira abordagem desprezar o efeito de distorção provocado pela lente, esta última forma de análise torna-se mais complexa e sem grandes vantagens práticas. Essa complexidade é fruto da resolução espacial das câmaras normalmente utilizadas e da possibilidade de colocação dos sensores em diferentes orientações e posições.

A imagem não é mais do que uma matriz, cujos elementos, normalmente conheci-

dos por *pixels*, são números correspondentes à quantificação da luz proveniente da cena tridimensional. Essa quantificação depende das características da câmara e da placa de aquisição de imagens. Segundo a teoria da geometria óptica clássica, a geometria dos raios incidentes no sensor e o modo de quantificação para cada *pixel*, constituem o modelo para a formação da imagem. Para descrever a verdadeira formação da imagem, este modelo deve respeitar a regra fundamental, de que qualquer ponto tridimensional, pertencente à cena captada, corresponde a um só *pixel* na imagem, embora o contrário não seja obrigatoriamente verdadeiro.

Ao longo desta secção serão apresentados: o modelo para o processo de formação de imagens, que será usado ao longo deste trabalho; o modo de funcionamento dos dispositivos de aquisição de imagens e outros modelos lineares alternativos ao escolhido.

2.2.1 Projecção em perspectiva

A projecção em perspectiva, também conhecida por modelo *Buraco de Alfinete* (“*pinhole*”, em inglês), é o modelo que explica com rigor o processo de formação de uma imagem. Tal como o nome indica, capta-se o efeito de perspectiva na formação da imagem, correspondente à representação bidimensional do mundo, perdendo-se, portanto, uma das dimensões do mundo representado através de três dimensões.

Considerando que $\{W\}$ é o referencial do mundo tridimensional e $\{CAM\}$ o referencial do sistema de coordenadas da câmara, cuja origem é o ponto O , pode-se verificar pela figura 2.1, que a qualquer ponto 3D, P , no referencial $\{W\}$, corresponde um ponto, p , pertencente a um plano I . Este plano é denominado como plano imagem e encontra-se a uma distância f do centro óptico da câmara, O . A distância f é conhecida por distância focal. O ponto p resulta da intersecção entre o plano I e a linha projectante \overline{OP} , que contém o centro de perspectiva (ou centro óptico) e o ponto P . Se o eixo Z_c , ou eixo óptico, for perpendicular ao plano imagem, então o plano I será paralelo ao plano X_cY_c do sistema de coordenadas $\{CAM\}$.

Tendo como base a figura 2.1, a modelação matemática do processo de formação da imagem parte da definição dos pontos no sistema de coordenadas $\{CAM\}$. Assim,

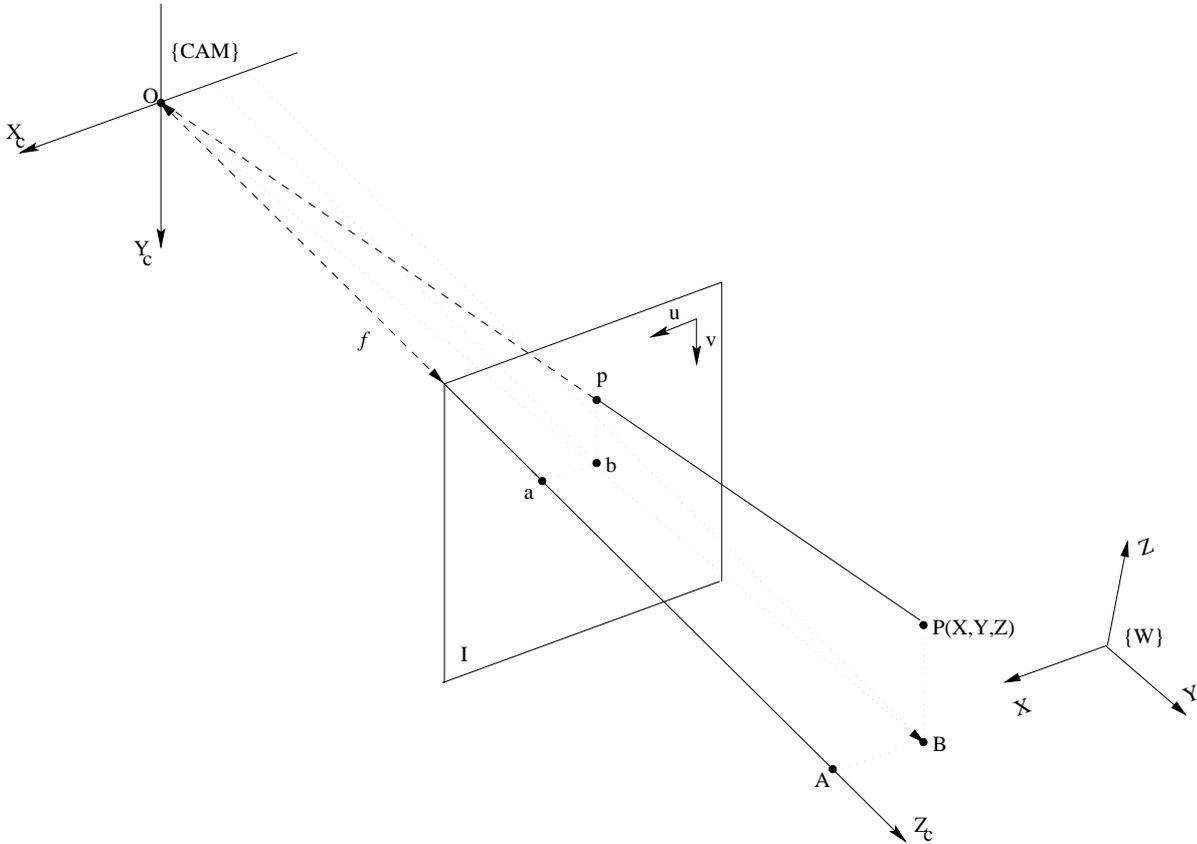


Figura 2.1: *Projecção em perspectiva. Dado o ponto tridimensional P , no referencial $\{W\}$ e o centro óptico da câmara, O , o ponto p resulta da intersecção entre \overline{OP} com o plano imagem I . A distância focal da câmara é f . Os pontos B e b são, respectivamente, a projecção dos pontos P e p no plano X_cZ_c . Por seu lado, A e a são as respectivas coordenadas Z_c de cada um dos pontos P e p .*

temos $P = (X_c, Y_c, Z_c)$ e $p = (u, v, f)$. Com as projecções dos pontos P e p no plano X_cZ_c pode-se considerar dois triângulos, o $\triangle OAB$ e o $\triangle Oab$. Através da semelhança entre estes triângulos obtém-se

$$\frac{\overline{b-a}}{\overline{Oa}} = \frac{\overline{B-A}}{\overline{OA}} \Leftrightarrow \frac{u}{f} = \frac{X_c}{Z_c}$$

ou, de outra forma,

$$u = \frac{fX_c}{Z_c} \quad (2.1)$$

pois B e b são, respectivamente, as projecções dos pontos P e p no plano X_cZ_c , e A

e a são, respectivamente, as coordenadas no eixo Z_c de cada um dos pontos P e p . Fazendo o mesmo raciocínio para uma projecção no plano Y_cZ_c chega-se a

$$v = \frac{fY_c}{Z_c} \quad (2.2)$$

Conhecendo as origens e orientações dos referenciais $\{W\}$ e $\{CAM\}$, através de uma simples transformação matemática linear poderão também ser usadas neste modelo as coordenadas do mundo 3D.

Contrariamente à suposição adoptada neste trabalho, quando se assume que a formação das imagens ocorre na parte negativa do eixo óptico, a relação entre as coordenadas do mundo e as coordenadas na imagem é idêntica, introduzindo-se somente o sinal de menos nas equações (2.1) e (2.2).

2.2.2 Lentes e focagens

Os dispositivos de aquisição de imagens utilizam lentes para produzir uma imagem da cena tridimensional. Na situação ideal considera-se o uso de uma só lente, fina e simétrica, que corresponde ao centro de perspectiva. Todo o sistema óptico composto por várias lentes, pode sempre ser reduzido ao caso ideal [Williams 86].

Quando os raios de luz incidentes num ponto da cena 3D, que se quer captar, se reflectem na direcção do dispositivo de aquisição de imagens, as lentes são atravessadas por eles, formando a imagem. Se os dispositivos se baseiam em lentes, existe então um limite à captação da luz e, em muitas situações, é mesmo possível haver um controlo das quantidades de luz recebidas, através de um dispositivo designado por íris. Na sua concepção, este dispositivo pode conter um só diafragma ou um conjunto de dispositivos especialmente concebidos para controlarem a passagem da radiação luminosa. Outro ponto importante a considerar é a distância focal das lentes, que impõe um limite ao valor máximo do ângulo entre um raio de luz e o eixo óptico, definindo assim o campo de visão do dispositivo.

Intimamente ligado às lentes está o processo de focagem. Este processo, que ocorre quando a convergência de toda a energia luminosa irradiada é concentrada num único ponto do plano imagem, pode ser descrito a partir do processo de formação

das imagens.

À noção de focagem vem, normalmente, associada a noção de profundidade de campo. A profundidade de campo é definida pelo intervalo de distância, sobre o qual as cenas se consideram bem focadas, e está relacionada com a resolução espacial do sensor da imagem, com o dispositivo óptico e com a distância à cena.

2.2.3 Modelos de projecção alternativos

O modelo descrito anteriormente pode ser simplificado, dando origem a outros modelos com grande aplicação prática.

De entre os modelos de projecção, o mais simples é o da projecção ortográfica. Este modelo adequa-se plenamente ao tratamento de imagens cujos objectos da cena tridimensional têm um tamanho reduzido, relativamente à distância entre esses objectos e o plano imagem. A única variante da projecção ortográfica em relação à projecção em perspectiva é que, naquela, o factor de ampliação $\frac{f}{Z_c} = 1$, fazendo com que um ponto no plano imagem seja obtido através de uma projecção paralela ao eixo óptico. O modelo da projecção ortográfica torna mais simples as equações (2.1) e (2.2), mas não capta o efeito de distância entre o sensor e o objecto, diminuindo a noção de profundidade normalmente contidas nas imagens.

Um outro modelo de projecção é o da perspectiva fraca, que consiste num misto de projecção em perspectiva e ortográfica. Para o perceber, considere-se um sistema de coordenadas, onde na parte positiva do eixo Z_c se coloca o plano imagem, na cota $Z_c = f$, sendo f a distância focal. Colocando um plano auxiliar, paralelo ao plano imagem, na cota $Z_c = Z_0$, com $Z_0 > f$, faz-se, primeiramente, a projecção do centro de massa de toda a cena que se quer captar segundo uma direcção perpendicular ao plano auxiliar. Depois, todos os pontos da cena são projectados, segundo a mesma direcção, nesse mesmo plano. O processo termina quando todos os pontos projectados no plano auxiliar forem projectados no plano imagem, segundo a projecção em perspectiva.

Relativamente à projecção ortográfica, o modelo da perspectiva fraca tem a vantagem de ter em conta a distância entre a cena observada e a câmara. Sendo assim, este modelo pode ser considerado como um modelo ortográfico melhorado. Dev-

ido a este facto, o modelo da perspectiva fraca é também conhecido por modelo de projecção ortográfica à escala. Este modelo poderia ser uma boa aproximação ao comportamento da câmara, se a relação de profundidade fosse fraca, isto é, se a estrutura da cena fosse relativamente pequena em relação à distância entre a câmara e a cena observada.

O modelo de projecção paraperspectiva, por seu lado, é uma variante do modelo de projecção da perspectiva fraca. A diferença entre eles está na direcção com que o centro de massa, de toda a cena que se quer captar, é projectado. No caso do modelo de projecção paraperspectiva, essa direcção é a que passa pela origem do sistema de coordenadas do mundo tridimensional. Relativamente à perspectiva fraca, este modelo tem a vantagem de ter em conta uma eventual translação lateral da cena em relação ao eixo óptico.

O modelo da câmara afim é uma generalização dos modelos de projecção ortográfica, de perspectiva fraca e de paraperspectiva. Este modelo foi introduzido por Mundy e Zisserman [Mundy 92]. O modelo da câmara afim, também pode ser visto como uma aproximação do modelo de projecção em perspectiva com o centro óptico localizado num plano no infinito. Tal como o modelo ortográfico, o modelo da câmara afim funciona bastante bem quando as medidas do objecto são pequenas, comparativamente à distância entre a câmara e o objecto.

A projecção em perspectiva altera a imagem de uma cena quando a câmara é sujeita a uma rotação, segundo um dos eixos X_c ou Y_c do referencial $\{CAM\}$ (ver figura 2.1). Se esse ângulo for grande, pode mesmo perder-se a imagem da cena pretendida. Na projecção esférica, qualquer ponto projectado no plano imagem é obtido através da intersecção da esfera de raio f (distância focal), medido a partir do centro óptico, e a recta que passa por esse ponto central e pelo ponto tridimensional. A imagem esférica da cena, resultante desta projecção, não sofre nenhuma distorção nos comprimentos e ângulos entre arestas dos objectos da cena, quando o movimento da câmara for só de rotação segundo qualquer um dos eixos X_c , Y_c ou Z_c do seu referencial. Isto reduz a complexidade dos problemas de determinação de áreas projectadas.

Apesar do modelo de projecção esférica facilitar a análise da imagem para este tipo

de movimentos, na prática surge um grande problema de utilização. Este deve-se aos sensores das câmaras serem planares e aos sinais por eles gerados serem amostrados em dispositivos planos. O problema de utilização surge na tentativa de repartição da esfera, de maneira periódica, em locais que representem amostras do sinal vídeo e no mecanismo de indexação desses bocados, tanto para escrita como para leitura.

2.3 Determinação de informação tridimensional

Tendo passado a fase da aquisição das imagens e a da escolha do modelo matemático a usar para descrever o processo de formação dessas imagens, está-se apto para iniciar a determinação de informação tridimensional baseada nas imagens adquiridas.

Quase todo o processo de determinação de informação tridimensional passa pelo cálculo das distâncias dos pontos da cena do mundo 3D a um determinado referencial, ou, de uma forma mais genérica, pelo cálculo das coordenadas desses pontos no sistema de eixos do mundo. Para se fazer este cálculo, deve-se, à partida, saber qual a posição e orientação relativa das câmaras envolvidas no processo. Posteriormente, agrega-se a informação obtida, referente a cada câmara, em diferentes sistemas de equações. Por fim, junta-se num único sistema matricial todos os sistemas de equações criados e envolvidos na aquisição das imagens. A resolução deste sistema matricial, usando todas as projecções do mesmo ponto 3D vistas nas várias imagens adquiridas, fornecerá as desejadas coordenadas tridimensionais desse ponto.

Assumindo daqui para a frente o uso da projecção em perspectiva, convém explicar melhor o que deve ser considerado no conhecimento da posição e orientação das câmaras envolvidas no processo de determinação de informação 3D. Assim, tendo por base a figura 2.1, devem ser considerados, para cada câmara, quatro sistemas de eixos. O sistema de eixos do mundo, $\{W\}$, onde são medidas as coordenadas 3D, no sistema métrico escolhido, normalmente em metros. O sistema de eixos da câmara, $\{CAM\}$, com a origem no centro óptico da câmara e com o eixo Z_c coincidente com o eixo óptico desta. O sistema de eixos associados ao plano imagem I , (u, v) , sendo este plano paralelo ao plano $X_c Y_c$ do sistema de eixos da câmara. E finalmente, o

sistema de eixos definidos na imagem, (x, y) , com origem no canto superior esquerdo da imagem. Neste sistema, a posição na coluna é representada por x enquanto que a posição na linha é representada por y . A unidade destas coordenadas é o *pixel*.

De seguida, mostrar-se-á que cada sistema de equações referente a cada câmara, explica a transformação matemática global dos pontos do sistema de eixos $\{W\}$ para o sistema de eixos definidos na imagem vista pela câmara. Será também provada a necessidade de existirem pelo menos duas câmaras, de forma a determinar, sem ambiguidade, a dimensionalidade dos pontos visualizados por ambas as câmaras.

2.3.1 Representação homogénea de coordenadas

Antes de definir quaisquer equações com vista à determinação da tridimensionalidade das cenas captadas, torna-se necessário rever o conceito de coordenada homogénea, frequentemente utilizado nesta dissertação.

Considere-se um espaço vectorial de n dimensões e um seu ponto genérico, $P = (x_1, x_2, \dots, x_n)$. Este ponto é representado em coordenadas homogéneas, através de um vector de $n + 1$ coordenadas, cuja forma matricial é

$$\mathbf{P} = \begin{bmatrix} kx_1 & kx_2 & \dots & kx_n & k \end{bmatrix}^T$$

O coeficiente k poderá ter um qualquer valor constante.

A transformação inversa, conhecida por normalização homogénea, é conseguida pela divisão das $n + 1$ coordenadas, por k , sendo as primeiras n coordenadas resultantes, consideradas as coordenadas cartesianas do ponto. No entanto, uma excepção a este cálculo surge quando $k = 0$. Nesta situação, interpreta-se o ponto como estando a uma distância infinita, obtido pela semi-recta que parte da origem e passa no ponto cartesiano $(kx_1, kx_2, \dots, kx_n)$.

2.3.2 Sistema de equações relativas ao uso de uma câmara

Para estabelecer a descrição matemática da transformação de um ponto tridimensional genérico, $P(X, Y, Z)$, na sua projecção na imagem, $p(x, y)$, em coordenadas *pixel*, consideremos P no sistema de coordenadas do mundo, $\{W\}$. A sua repre-

sentação no sistema de coordenadas da câmara, $\{CAM\}$, $P_c = (X_c, Y_c, Z_c)$, é dada, na forma matricial, pela transformação

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \Leftrightarrow P_c = \mathbf{R}P + \mathbf{t} \quad (2.3)$$

onde \mathbf{R} e \mathbf{t} são, respectivamente, a matriz de rotação e o vector de translação que relacionam os dois referenciais. O vector de translação é constituído pelas projecções da recta que une a origem do sistema de eixos $\{W\}$ à origem do sistema de eixos $\{CAM\}$, nos eixos X , Y e Z , por esta ordem [Fu 87]. Por seu lado, a matriz de rotação, na sua forma geral, é definida por

$$\mathbf{R} = \mathbf{R}_\theta \mathbf{R}_\beta \mathbf{R}_\alpha = \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{r}'_1 \\ \mathbf{r}'_2 \\ \mathbf{r}'_3 \end{bmatrix} \quad (2.4)$$

com

$$\mathbf{R}_\theta = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{R}_\beta = \begin{bmatrix} \cos(\beta) & 0 & \sin(\beta) \\ 0 & 1 & 0 \\ -\sin(\beta) & 0 & \cos(\beta) \end{bmatrix} \mathbf{R}_\alpha = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha) & -\sin(\alpha) \\ 0 & \sin(\alpha) & \cos(\alpha) \end{bmatrix}$$

Como se pode verificar, as normas das colunas, bem como das linhas, das matrizes de rotação \mathbf{R}_θ , \mathbf{R}_β e \mathbf{R}_α são unitárias. Além disso, estas matrizes têm a propriedade da sua inversa ser sempre igual à sua transposta. Estes factos fazem com que a matriz \mathbf{R} tenha, também, as mesmas características, isto é, $\mathbf{R}^{-1} = \mathbf{R}^T$ e

$$\| \mathbf{r}_1 \| = \| \mathbf{r}'_1 \| = \| \mathbf{r}_2 \| = \| \mathbf{r}'_2 \| = \| \mathbf{r}_3 \| = \| \mathbf{r}'_3 \| = 1 \quad (2.5)$$

Todas estas matrizes de rotação pertencem ao espaço vectorial $SO(3)$.

Como se pode verificar na figura 2.2, α , β e θ são ângulos de rotação sobre os eixos de coordenadas X , Y e Z , respectivamente. Estes ângulos são positivos quando medidos segundo o sentido inverso ao dos ponteiros de um relógio, ou seja, são medidos no sentido directo.

Tal como foi descrito atrás, a transformação das coordenadas no referencial da câmara em coordenadas do plano imagem, (u, v) , é feita segundo o modelo de pro-

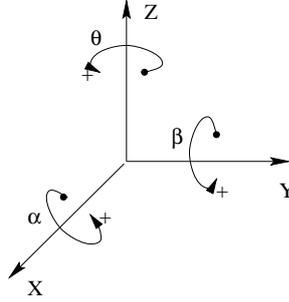


Figura 2.2: Sentido positivo das rotações sobre os eixos de coordenadas.

jecção em perspectiva, através das equações (2.1) e (2.2). Agrupando ambas as equações numa única equação vectorial linear, recorrendo à utilização de coordenadas homogéneas, obtém-se

$$k \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix}$$

As coordenadas na imagem, (x, y) , não correspondem às coordenadas físicas, (u, v) , no plano imagem. Com uma câmara CCD, a relação entre ambas depende do tamanho e forma dos *pixels* e da posição do CCD na câmara. A transformação final para a obtenção das coordenadas referentes aos eixos da imagem é então conseguida através de

$$x = k_x u + k_y \tan \varphi v + c_x = k_x u + \gamma v + c_x \quad y = k_y v + c_y$$

onde k_x e k_y são os factores de conversão, nos eixos x e y da imagem, respectivamente, do sistema métrico do mundo, normalmente em metros, para o sistema métrico da imagem, *pixels*. (c_x, c_y) são as coordenadas *pixel* do ponto principal, também referido como centro da imagem. φ é o ângulo de desvio indicado na figura 2.3. Por sua vez, γ é o factor que tem em conta o desvio que acontece quando os *pixels* não são rectangulares. Como foi visto na equação anterior, este factor contém φ . Para a maioria das câmaras, os *pixels* são quase rectângulos perfeitos, sendo muitas vezes assumido que o valor de γ é zero. Toda esta transformação é ilustrada na figura 2.3.

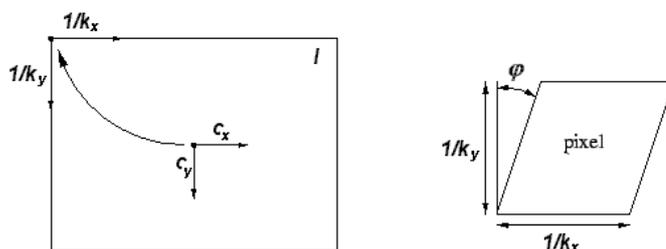


Figura 2.3: Transformação das coordenadas (u, v) em coordenadas (x, y) .

Este processo de relacionamento entre $P = (X, Y, Z)$ e $p = (x, y)$ pode ser descrito numa única transformação global, usando a representação homogênea, por

$$\begin{aligned} \begin{bmatrix} kx \\ ky \\ k \end{bmatrix} &= \begin{bmatrix} fk_x & \gamma & c_x & 0 \\ 0 & fk_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \Leftrightarrow \\ \Leftrightarrow k\mathbf{p} &= \begin{bmatrix} \mathbf{C} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \mathbf{P} \Leftrightarrow k\mathbf{p} = \mathbf{C} \begin{bmatrix} \mathbf{R} & \mathbf{t} \end{bmatrix} \mathbf{P} \end{aligned} \quad (2.6)$$

ou, de uma forma mais geral, por

$$\begin{bmatrix} kx \\ ky \\ k \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & t_{13} & t_{14} \\ t_{21} & t_{22} & t_{23} & t_{24} \\ t_{31} & t_{32} & t_{33} & t_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \Leftrightarrow k\mathbf{p} = \mathbf{TP} \quad (2.7)$$

onde k é um factor de escala arbitrário. Os vectores \mathbf{P} e \mathbf{p} representam os respectivos pontos P e p em coordenadas homogêneas. A matriz \mathbf{C} é conhecida por matriz de calibração. A matriz \mathbf{T} , por sua vez, é conhecida por matriz de transformação de projecção em perspectiva, ou, mais simplesmente, por matriz de projecção perspectiva.

Todos os parâmetros geométricos que entraram neste raciocínio estão agrupados em dois tipos, de denotação diferente: os extrínsecos e os intrínsecos. Os primeiros

indicam a posição da câmara em relação ao sistema de coordenadas do mundo. Estes parâmetros são todos os elementos da matriz \mathbf{R} e do vector \mathbf{t} . Os parâmetros intrínsecos, por seu lado, são todos os elementos que se relacionam com as propriedades inerentes à óptica da câmara, ou seja, a distância focal f , o centro da imagem (c_x, c_y) e os factores de escala k_x e k_y , que estão contidos na matriz de calibração.

A equação (2.7) pode ser representada de outra maneira, eliminando k de forma a se obter um sistema matricial de equações, em ordem às coordenadas X , Y e Z , do ponto P , e cujo resultado é

$$\begin{bmatrix} t_{11} - t_{31}x & t_{12} - t_{32}x & t_{13} - t_{33}x \\ t_{21} - t_{31}y & t_{22} - t_{32}y & t_{23} - t_{33}y \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} t_{34}x - t_{14} \\ t_{34}y - t_{24} \end{bmatrix}$$

ou, de uma forma mais clara ainda,

$$\begin{cases} (t_{11} - t_{31}x)X + (t_{12} - t_{32}x)Y + (t_{13} - t_{33}x)Z + (t_{14} - t_{34}x) = 0 \\ (t_{21} - t_{31}y)X + (t_{22} - t_{32}y)Y + (t_{23} - t_{33}y)Z + (t_{24} - t_{34}y) = 0 \end{cases} \quad (2.8)$$

Geometricamente, no sistema de equações (2.8) estão representados dois planos, cada um definido por uma das equações do sistema. Da sua intersecção resulta uma recta que contém o ponto tridimensional, P , o centro de perspectiva da câmara e o ponto na imagem, p . Esta recta é designada por recta projectiva de P .

2.3.3 Determinação de coordenadas dos pontos 3D

Na determinação de coordenadas tridimensionais de um ponto genérico $P = (X, Y, Z)$ a partir das suas coordenadas $p = (x, y)$, na imagem, é evidente a necessidade de utilizar uma câmara. No entanto, o mapeamento tridimensional da cena a partir de uma única imagem é uma transformação cuja solução é possível e indeterminada. Este facto prova-se, numericamente, recorrendo ao sistema de equações (2.8), que é composto por duas equações e três incógnitas, podendo um único ponto na imagem, ter uma infinidade de pontos do mundo. A explicação geométrica para a infinidade de soluções está em que, qualquer ponto 3D que pertencente à recta

definida em (2.8), tem como ponto de projecção o mesmo ponto p , no plano imagem. Para recuperar a informação de profundidade com a aquisição de imagens terá que se usar pelo menos outra câmara, criando-se assim um sistema com mais equações do que incógnitas.

Na visão humana, é a diferença entre a projecção dos pontos tridimensionais no olho esquerdo e no direito que dá a percepção de profundidade. Este facto, comprovado pelo que foi apresentado acima, levou ao estudo de uma geometria similar, conhecida por estéreo. A geometria estéreo, como se exemplifica na figura 2.4, é o modo mais simples para se recuperar as coordenadas dos pontos 3D da cena. A recu-

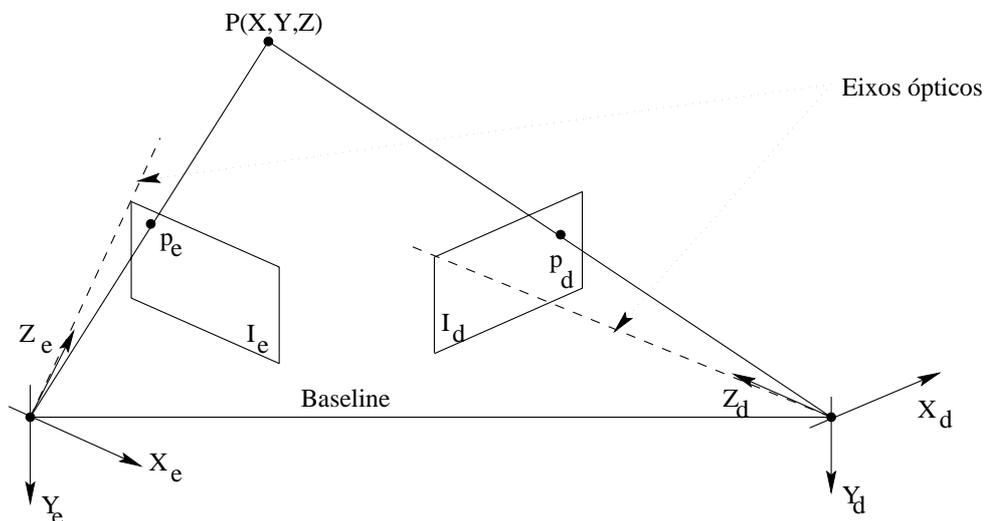


Figura 2.4: Geometria estereoscópica. A determinação de P é conseguida através da intersecção dos segmentos de recta $\overline{Pp_e}$ e $\overline{Pp_d}$.

peração tridimensional de pontos de uma cena, usando somente câmaras, envolve a obtenção de pelo menos duas imagens de uma mesma região de interesse, adquiridas em posições diferentes. Uma das maneiras para a aquisição das imagens da cena, consiste em separar as duas câmaras por uma distância conhecida, à qual se chama *baseline*, medida entre os seus centros ópticos.

Para o cálculo de coordenadas 3D da cena, usando duas câmaras, estabelece-se duas equações do mesmo tipo que (2.8), relativas a cada uma das imagens, esquerda

e direita, obtendo a equação matricial

$$\begin{bmatrix} t_{11_e} - t_{31_e}x_e & t_{12_e} - t_{32_e}x_e & t_{13_e} - t_{33_e}x_e & t_{14_e} - t_{34_e}x_e \\ t_{21_e} - t_{31_e}y_e & t_{22_e} - t_{32_e}y_e & t_{23_e} - t_{33_e}y_e & t_{24_e} - t_{34_e}y_e \\ t_{11_d} - t_{31_d}x_d & t_{12_d} - t_{32_d}x_d & t_{13_d} - t_{33_d}x_d & t_{14_d} - t_{34_d}x_d \\ t_{21_d} - t_{31_d}y_d & t_{22_d} - t_{32_d}y_d & t_{23_d} - t_{33_d}y_d & t_{24_d} - t_{34_d}y_d \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \Leftrightarrow \mathbf{Ax} = \mathbf{b} \quad (2.9)$$

com

$$\mathbf{A} = \begin{bmatrix} t_{11_e} - t_{31_e}x_e & t_{12_e} - t_{32_e}x_e & t_{13_e} - t_{33_e}x_e \\ t_{21_e} - t_{31_e}y_e & t_{22_e} - t_{32_e}y_e & t_{23_e} - t_{33_e}y_e \\ t_{11_d} - t_{31_d}x_d & t_{12_d} - t_{32_d}x_d & t_{13_d} - t_{33_d}x_d \\ t_{21_d} - t_{31_d}y_d & t_{22_d} - t_{32_d}y_d & t_{23_d} - t_{33_d}y_d \end{bmatrix}$$

$$\mathbf{x} = P = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} t_{34_e}x_e - t_{14_e} \\ t_{34_e}y_e - t_{24_e} \\ t_{34_d}x_d - t_{14_d} \\ t_{34_d}y_d - t_{24_d} \end{bmatrix}$$

e onde o índice e indica que pertence à imagem ou câmara esquerda, e o índice d que pertence à direita. Como se pode verificar, o sistema de equações que resulta de (2.9), tem quatro equações e três incógnitas. A sua solução óptima, ou por outras palavras, o ponto P , pode então ser obtido através de vários métodos numéricos, que serão alvo de uma melhor explanação na secção 2.7. Este raciocínio é também aplicado quando temos mais do que duas imagens, acrescentando ao sistema representado na equação (2.9) equações do tipo (2.8), por cada imagem adicional.

2.4 Geometria epipolar e matriz fundamental

Continuando na linha de pensamento sobre a apresentação de um conjunto de conceitos básicos que levam à determinação de informação 3D, avança-se para uma das suas fases mais importantes, a correspondência. É exactamente este processo de correspondência que suscita normalmente aquilo que se designa por geometria epipolar, e é com esse objectivo que se inicia a sua explicação.

Considere-se que um ponto genérico $P = (X, Y, Z)$ pode ser observado numa imagem esquerda e noutra direita, nas respectivas posições $p_e = (x_e, y_e)$ e $p_d = (x_d, y_d)$. O problema

da correspondência pode ser resumido na procura do ponto p_d em toda a imagem direita, que mais se aproxima do conhecido ponto p_e da imagem esquerda, pois não se sabe quais são as suas coordenadas. Se for suposto que p_d está restrito a uma recta na imagem direita, obtida com base no ponto conhecido p_e , e havendo um mecanismo para calcular essa recta, correspondente a p_e , o espaço de procura de p_d é drasticamente reduzido.

A restrição epipolar

$$\mathbf{p}_d^T \mathbf{F} \mathbf{p}_e = 0 \quad (2.10)$$

que emerge da existência de dois quaisquer pontos de vista diferentes, da mesma cena 3D, é a ideia chave para a obtenção do mecanismo para encontrar a recta correspondente ao conhecido ponto da imagem esquerda.

Na equação (2.10), \mathbf{F} , designada por matriz fundamental, é uma matriz 3×3 , porque os pontos \mathbf{p}_e e \mathbf{p}_d estão em coordenadas homogéneas. A matriz fundamental, \mathbf{F} , é a representação algébrica da geometria epipolar, pois ela encerra toda a informação geométrica que relaciona ambas as imagens. A geometria epipolar é intrínseca a toda a geometria projectiva entre duas vistas. Além disso, é independente da estrutura da cena do mundo 3D, mas depende dos parâmetros intrínsecos das câmaras envolvidas e da posição e orientação relativa entre elas.

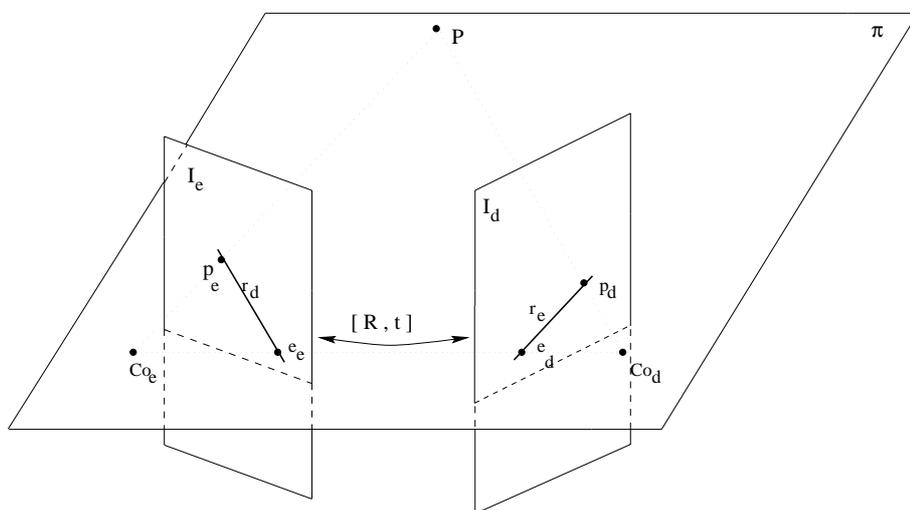


Figura 2.5: Geometria epipolar. As duas imagens do ponto P têm os centros ópticos nas posições Co_e e Co_d . O plano que contém Co_e , P e Co_d é chamado plano epipolar de P . Este plano intersecta as imagens ao longo das duas rectas epipolares r_e e r_d . Aos pontos e_e e e_d chamam-se epipolos.

Graficamente, a geometria epipolar pode ser explicada da maneira que se segue. Considere-se as imagens de uma cena 3D, representadas na figura 2.5, adquiridas segundo uma projecção em perspectiva. Estas duas imagens formam um par estéreo. Sejam, então, I_e e I_d , os planos imagem correspondentes à mesma cena e Co_e e Co_d , os centros ópticos relativos a cada imagem. A recta formada por Co_e e Co_d intersecta os planos I_e e I_d , dando origem aos pontos e_e e e_d , respectivamente. Estes pontos chamam-se pontos epipolares ou epipolos. Ao ponto p_e na imagem I_e , corresponde o ponto p_d da imagem I_d . O ponto p_d está condicionado a pertencer à recta correspondente a p_e , chamada de epipolar e denotada por r_e . Esta recta é obtida pela intersecção do plano epipolar, definido pelos pontos Co_e , Co_d e p_e , com o plano imagem I_d . O inverso também é válido, originando assim a recta epipolar correspondente a p_d , denotada por r_d .

A restrição, em que os pontos da imagem devem estar contidos nas rectas epipolares, acontece porque ambos os pontos p_e e p_d correspondem ao ponto tridimensional P . Assim, como P pertence ao plano formado por si próprio e pelos respectivos centros ópticos das imagens, as rectas epipolares resultam da intersecção desse plano com os planos imagem.

Todas as rectas epipolares da imagem I_e passam no ponto e_e enquanto que as de I_d contêm o ponto e_d . A justificação deve-se ao facto de todos os planos epipolares serem definidos pelo ponto P específico (varia para cada plano epipolar) e pelos centros ópticos Co_e e Co_d , formando em torno da recta que passa por Co_e e Co_d um feixe de planos que intersectam as imagens. Da intersecção destes planos com as imagens I_e e I_d resultam, em segunda análise, os pontos comuns a todas as intersecções, os respectivos epipolos, e_e e e_d . Um epipolo numa imagem é, portanto, a projecção nessa imagem do centro de projecção da outra imagem.

Toda esta relação geométrica entre as coordenadas das imagens e as suas correspondentes epipolares pode ser descrita, tal como foi dito acima, de forma matricial pela equação (2.10). Para lá chegar, assumamos que o sistema de coordenadas da câmara esquerda é o mesmo que o sistema de coordenadas do mundo. Desta forma, qualquer ponto no sistema da câmara direita pode ser obtido usando uma rotação \mathbf{R} e uma translação \mathbf{t} . Todos estes pressupostos não limitam a generalidade da dedução da equação (2.10).

Usando a equação (2.6), a expressão do ponto genérico P , representado na figura 2.5, em ambos os sistemas de coordenadas das imagens é dado por

$$k_e \mathbf{p}_e = \mathbf{C}_e \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} \mathbf{P} \quad k_d \mathbf{p}_d = \mathbf{C}_d \begin{bmatrix} \mathbf{R} & \mathbf{t} \end{bmatrix} \mathbf{P} \quad (2.11)$$

onde \mathbf{I} é a matriz identidade, 3×3 , e $\mathbf{0}$ o vector nulo, 3×1 . As matrizes \mathbf{C}_e e \mathbf{C}_d são as matrizes de calibração das câmaras esquerda e direita, respectivamente.

Definindo

$$\mathbf{P}_e = \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} \mathbf{P} \quad \mathbf{P}_d = \begin{bmatrix} \mathbf{R} & \mathbf{t} \end{bmatrix} \mathbf{P}$$

as equações de (2.11), assumindo que \mathbf{C}_e e \mathbf{C}_d são invertíveis, podem ser reformuladas passando a

$$\mathbf{P}_e = \mathbf{C}_e^{-1} k_e \mathbf{p}_e \quad \mathbf{P}_d = \mathbf{C}_d^{-1} k_d \mathbf{p}_d \quad (2.12)$$

O plano epipolar π , mostrado na figura 2.5, também pode ser definido pelos três vectores coplanares \mathbf{P}_e , \mathbf{P}_d e \mathbf{t} . O vector perpendicular a este plano é dado então por $(\mathbf{t} \wedge \mathbf{R}\mathbf{P}_e)$. Sendo \mathbf{P}_d um vector do plano π , o produto interno entre estes vectores é

$$\mathbf{P}_d^T (\mathbf{t} \wedge \mathbf{R}\mathbf{P}_e) = 0 \quad (2.13)$$

Considerando o vector de translação entre duas imagens $\mathbf{t} = \begin{bmatrix} t_x & t_y & t_z \end{bmatrix}^T$, a matriz antisimétrica $[\mathbf{t}]_\wedge$ é definida por

$$[\mathbf{t}]_\wedge = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \quad (2.14)$$

Esta matriz relaciona-se com o produto vectorial entre quaisquer vectores tridimensionais através de

$$\mathbf{t} \wedge \mathbf{v} = [\mathbf{t}]_\wedge \mathbf{v} \quad (2.15)$$

No apêndice C é deduzida esta expressão. A partir de (2.14), pode verificar-se que a matriz antisimétrica tem determinante nulo e que

$$[\mathbf{t}]_\wedge = -[\mathbf{t}]_\wedge^T$$

Notar que esta matriz tem sempre característica dois, pois, no caso de apenas um dos elementos do vector \mathbf{t} ser nulo ou no caso de dois dos componentes do vector \mathbf{t} serem nulos, existem sempre duas colunas linearmente independentes. Para o caso dos três elementos de \mathbf{t} serem nulos, não se pode falar em matriz antisimétrica, pois a matriz é nula.

Utilizando (2.15) na equação (2.13) obtém-se

$$\mathbf{P}_d^T [\mathbf{t}]_\wedge \mathbf{R}\mathbf{P}_e = 0 \quad (2.16)$$

de onde se retira a matriz

$$\mathbf{E} = [\mathbf{t}]_{\wedge} \mathbf{R} \quad (2.17)$$

designada por *matriz essencial* e introduzida por Longuet-Higgins [Longuet-Higgins 81]. Esta matriz relaciona os pontos 3D de cada referencial das câmaras.

Substituindo a equação de (2.12) em (2.16) obtém-se

$$\mathbf{p}_d^T \mathbf{C}_d^{-T} [\mathbf{t}]_{\wedge} \mathbf{R} \mathbf{C}_e^{-1} \mathbf{p}_e = 0 \quad (2.18)$$

sendo a matriz fundamental dada por

$$\mathbf{F} = \mathbf{C}_d^{-T} [\mathbf{t}]_{\wedge} \mathbf{R} \mathbf{C}_e^{-1} \quad (2.19)$$

Pela equação (2.10) pode-se concluir que quando se usa a matriz \mathbf{F} para relacionar as imagens, não é necessário serem conhecidos quaisquer parâmetros intrínsecos das câmaras.

Como $\det([\mathbf{t}]_{\wedge}) = 0$, pela propriedade dos determinantes para as matrizes quadradas,

$$\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$$

os determinantes de \mathbf{E} e \mathbf{F} são nulos (pag. 12 - [Datta 94]). Logo, as matrizes \mathbf{E} e \mathbf{F} são singulares.

Outra particularidade da matriz fundamental, \mathbf{F} , é que ela é definida a menos de um factor de escala, porque a equação (2.10) continua a verificar-se, se for multiplicada por um escalar arbitrário [Zhang 96].

A matriz \mathbf{F} tem característica dois porque considerando a equação (2.19) verifica-se que algumas das matrizes que compõem \mathbf{F} têm característica três e outras têm característica dois. As matrizes \mathbf{C}_e e \mathbf{C}_d têm característica três por serem matrizes 3×3 , invertíveis (ver equação 2.6). A matriz \mathbf{R} tem característica três por definição (ver equação 2.4) e a matriz $[\mathbf{t}]_{\wedge}$, também por definição, tem característica dois (ver equação 2.14). Assim, definindo a função que determina a característica de uma matriz \mathbf{M} por $C(\mathbf{M})$, pela propriedade da característica de uma matriz,

$$C(\mathbf{AB}) \leq \min(C(\mathbf{A}), C(\mathbf{B}))$$

\mathbf{F} só pode ter característica dois (pag. 13 - [Datta 94]).

Dado que \mathbf{F} é uma matriz 3×3 , definida a menos de um factor de escala, são necessárias oito equações linearmente independentes para que se obtenham os seus nove elementos. Se,

com base na equação (2.10), forem encontradas sete equações linearmente independentes, ficará a faltar uma equação para que o sistema de equações fique completo (ou seja, para que tenha solução). A aplicação do conhecimento de que o determinante de \mathbf{F} é nulo satisfaz a necessidade da oitava equação. Desta feita, pode-se dizer que \mathbf{F} têm sete graus de liberdade pois para ser definida apenas se precisou de encontrar sete equações linearmente independentes. Mais à frente será descrito com maior pormenor a aplicação deste conhecimento num processo de determinação de \mathbf{F} .

Uma recta genérica assume a forma

$$\mathbf{r} \equiv r_x x + r_y y + r_c = \begin{bmatrix} r_x & r_y & r_c \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{l}\mathbf{p} = 0$$

As representações projectivas das rectas epipolares para a câmara esquerda e direita, respectivamente \mathbf{r}_e e \mathbf{r}_d , podem ser obtidas através de

$$\mathbf{r}_e = \mathbf{F}\mathbf{p}_e \quad \mathbf{r}_d = \mathbf{F}^T \mathbf{p}_d$$

onde os coeficientes r_x, r_y e r_c de cada recta são obtidos, respectivamente, pelas multiplicações da primeira, segunda e terceira linhas de \mathbf{F} (caso de \mathbf{r}_e) e \mathbf{F}^T (caso de \mathbf{r}_d) com as coordenadas do ponto na imagem. O facto de existir um factor independente na definição algébrica destas rectas, vem confirmar a necessidade do uso de coordenadas homogéneas na utilização da matriz fundamental.

Outra informação importante é a expressão matemática que permite a obtenção dos epipolos. Por definição, e como se pode observar na figura 2.5, a projecção do centro óptico esquerdo na imagem direita é o epipolo direito, enquanto que a projecção do centro óptico direito na imagem esquerda é o epipolo esquerdo. Através da equação (2.6) têm-se

$$\mathbf{e}_e = \frac{1}{k_e} \mathbf{C}_e \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} \mathbf{C}_d \quad \mathbf{e}_d = \frac{1}{k_d} \mathbf{C}_d \begin{bmatrix} \mathbf{R} & \mathbf{t} \end{bmatrix} \mathbf{C}_e \quad (2.20)$$

Por outro lado, qualquer ponto do sistema de coordenadas da câmara esquerda pode ser representado no sistema de coordenadas da câmara direita, e vice-versa, com base na equação (2.3), através de

$$P_d = \mathbf{R}P_e + \mathbf{t} \Leftrightarrow P_e = \mathbf{R}^{-1}(P_d - \mathbf{t})$$

Este facto faz com que os centros ópticos das duas câmaras, devido a serem as origens nos respectivos sistemas de coordenadas, sejam vistos pelo outro sistema de coordenadas como

$$C_{O_e} = \mathbf{t} \quad C_{O_d} = -\mathbf{R}^{-1}\mathbf{t}$$

Juntando este conhecimento à equação (2.20), a expressão para cada epipolo passa a ser dada por

$$\mathbf{e}_e = -\frac{1}{k_e} \mathbf{C}_e \mathbf{R}^{-1} \mathbf{t} \quad \mathbf{e}_d = \frac{1}{k_d} \mathbf{C}_d \mathbf{t} \quad (2.21)$$

A equação (2.19) não é a única maneira de se definir a matriz fundamental. Assim, começa-se por apresentar a seguinte igualdade, deduzida no apêndice C,

$$[\mathbf{v}]_{\wedge} \mathbf{M} = \det(\mathbf{M}) \mathbf{M}^{-T} [\mathbf{M}^{-1} \mathbf{v}]_{\wedge} \quad (2.22)$$

onde \mathbf{v} é um qualquer vector tridimensional e \mathbf{M} uma qualquer matriz 3×3 não singular. Assumindo que $\mathbf{v} = \mathbf{C}_d \mathbf{t}$ e que $\mathbf{M} = \mathbf{C}_d$, a igualdade anterior passa a ser dada por

$$[\mathbf{C}_d \mathbf{t}]_{\wedge} \mathbf{C}_d = \det(\mathbf{C}_d) \mathbf{C}_d^{-T} [\mathbf{C}_d^{-1} \mathbf{C}_d \mathbf{t}]_{\wedge} \Leftrightarrow \mathbf{C}_d^{-T} [\mathbf{t}]_{\wedge} = \frac{1}{\det(\mathbf{C}_d)} [\mathbf{C}_d \mathbf{t}]_{\wedge} \mathbf{C}_d$$

Logo a equação (2.19) pode ser reformulada em

$$\mathbf{F} = \frac{1}{\det(\mathbf{C}_d)} [\mathbf{C}_d \mathbf{t}]_{\wedge} \mathbf{C}_d \mathbf{R} \mathbf{C}_e^{-1}$$

Aplicando à equação anterior a expressão (2.21), obtém-se

$$\mathbf{F} = \frac{1}{\det(\mathbf{C}_d)} [k_d \mathbf{e}_d]_{\wedge} \mathbf{C}_d \mathbf{R} \mathbf{C}_e^{-1} = \frac{k_d}{\det(\mathbf{C}_d)} [\mathbf{e}_d]_{\wedge} \mathbf{C}_d \mathbf{R} \mathbf{C}_e^{-1}$$

ou

$$\mathbf{F} = [\mathbf{e}_d]_{\wedge} \mathbf{C}_d \mathbf{R} \mathbf{C}_e^{-1} \quad (2.23)$$

pois \mathbf{F} é definida a menos de um factor de escala.

Por seu lado, se for assumido que $\mathbf{v} = \mathbf{t}$ e que $\mathbf{M} = \mathbf{R} \mathbf{C}_e^{-1}$, a igualdade (2.22) pode ser reformulada em

$$[\mathbf{t}]_{\wedge} \mathbf{R} \mathbf{C}_e^{-1} = \det(\mathbf{R} \mathbf{C}_e^{-1}) (\mathbf{R} \mathbf{C}_e^{-1})^{-T} [(\mathbf{R} \mathbf{C}_e^{-1})^{-1} \mathbf{t}]_{\wedge} = \det(\mathbf{R} \mathbf{C}_e^{-1}) \mathbf{R}^{-T} \mathbf{C}_e^T [\mathbf{C}_e \mathbf{R}^{-1} \mathbf{t}]_{\wedge}$$

Desta forma a equação (2.19) fica

$$\mathbf{F} = \det(\mathbf{R} \mathbf{C}_e^{-1}) \mathbf{C}_d^{-T} \mathbf{R}^{-T} \mathbf{C}_e^T [\mathbf{C}_e \mathbf{R}^{-1} \mathbf{t}]_{\wedge} = \det(\mathbf{R} \mathbf{C}_e^{-1}) (\mathbf{C}_d \mathbf{R} \mathbf{C}_e^{-1})^{-T} [\mathbf{C}_e \mathbf{R}^{-1} \mathbf{t}]_{\wedge}$$

Aplicando à equação anterior a expressão (2.21), obtém-se

$$\mathbf{F} = \det(\mathbf{R} \mathbf{C}_e^{-1}) (\mathbf{C}_d \mathbf{R} \mathbf{C}_e^{-1})^{-T} [-k_e \mathbf{e}_e]_{\wedge} = -k_e \det(\mathbf{R} \mathbf{C}_e^{-1}) (\mathbf{C}_d \mathbf{R} \mathbf{C}_e^{-1})^{-T} [\mathbf{e}_e]_{\wedge}$$

ou

$$\mathbf{F} = (\mathbf{C}_d \mathbf{R} \mathbf{C}_e^{-1})^{-T} [\mathbf{e}_e]_{\wedge} \quad (2.24)$$

pois \mathbf{F} é definida a menos de um factor de escala.

As equações (2.23) e (2.24) mostram que a matriz fundamental contém, entre outras, a informação sobre o posicionamento dos epipolos. Assim, passa-se a mostrar como eles são obtidos, utilizando apenas a matriz fundamental, \mathbf{F} . Para o demonstrar, começa-se por definir espaço nulo esquerdo e direito de uma matriz \mathbf{M} . O espaço nulo esquerdo \mathbf{x} , de \mathbf{M} , é definido quando $\mathbf{x}^T \mathbf{M} = 0$, com $\mathbf{x} \neq 0$, e é representado por $\mathcal{N}_e(\mathbf{M})$. Por seu lado, o espaço nulo direito \mathbf{x}' , de \mathbf{M} , é definido quando $\mathbf{M} \mathbf{x}' = 0$, com $\mathbf{x}' \neq 0$, e é representado por $\mathcal{N}_d(\mathbf{M})$. Além disso, aplicando a transposta a ambas as definições de espaço nulo, por exemplo ao espaço nulo direito,

$$(\mathbf{M} \mathbf{x}')^T = 0 \Leftrightarrow \mathbf{x}'^T \mathbf{M}^T = 0$$

conclui-se que $\mathcal{N}_d(\mathbf{M}) = \mathcal{N}_e(\mathbf{M}^T)$. Ainda no âmbito dos espaços nulos, sabe-se que estes relacionam-se com a característica da matriz \mathbf{M} , $m \times n$, (pag. 13 - [Datta 94]), através de

$$C(\mathbf{M}) + \mathcal{N}_e(\mathbf{M}) = C(\mathbf{M}) + \mathcal{N}_d(\mathbf{M}) = n \quad (2.25)$$

Com base nas equações (2.15) e (2.23), a multiplicação entre o epipolo direito, \mathbf{e}_d , e a matriz fundamental, \mathbf{F} , é dada por

$$\mathbf{e}_d^T \mathbf{F} = \mathbf{e}_d^T [\mathbf{e}_d]_{\wedge} \mathbf{C}_d \mathbf{R} \mathbf{C}_e^{-1} = -[\mathbf{e}_d]_{\wedge} \mathbf{e}_d^T \mathbf{C}_d \mathbf{R} \mathbf{C}_e^{-1} = -(\mathbf{e}_d \wedge \mathbf{e}_d)^T \mathbf{C}_d \mathbf{R} \mathbf{C}_e^{-1} = 0$$

pois o produto externo do mesmo vector é igual a zero. Assim, aplicando os conhecimentos, atrás referidos, sobre os espaços nulos, tem-se que $\mathbf{e}_d = \mathcal{N}_e(\mathbf{F}) = \mathcal{N}_d(\mathbf{F}^T)$. Fazendo o mesmo para o epipolo esquerdo, \mathbf{e}_e , usando agora a equação (2.24), obtém-se

$$\mathbf{e}_e^T \mathbf{F}^T = \mathbf{e}_e^T ((\mathbf{C}_d \mathbf{R} \mathbf{C}_e^{-1})^{-T} [\mathbf{e}_e]_{\wedge})^T = \mathbf{e}_e^T [\mathbf{e}_e]_{\wedge}^T (\mathbf{C}_d \mathbf{R} \mathbf{C}_e^{-1})^{-1} = (\mathbf{e}_e \wedge \mathbf{e}_e)^T (\mathbf{C}_d \mathbf{R} \mathbf{C}_e^{-1})^{-1} = 0$$

e, portanto, tem-se que $\mathbf{e}_e = \mathcal{N}_e(\mathbf{F}^T) = \mathcal{N}_d(\mathbf{F})$.

Como já foi mencionado, cada imagem tem apenas um epipolo, o qual está presente na definição da matriz fundamental, \mathbf{F} . Este facto reforça a ideia de \mathbf{F} ter sempre característica dois, pois pela relação (2.25) o número de colunas de uma matriz \mathbf{F} (que é 3) é igual à soma da sua característica (que é 2) com o espaço nulo que essa matriz tem (que é 1). Assim, se \mathbf{F} tivesse característica três não teria espaços nulos, e portanto, não teria epipolos. Se,

por outro lado, tivesse característica menor que dois teria dois ou mais espaços nulos, e portanto, dois ou mais epipolos. Obviamente, estas hipóteses são um absurdo, pois não se encaixam no conceito de geometria epipolar.

2.4.1 Matriz fundamental para configurações específicas

Para certas configurações das câmaras, a geometria epipolar, na forma da matriz fundamental, \mathbf{F} , pode depender de menos parâmetros relativamente ao que acontece no caso geral ou simplificar alguns processos relacionados com ela (obtenção de parâmetros intrínsecos ou extrínsecos das câmaras com base em \mathbf{F}). A redução do número de parâmetros de \mathbf{F} está relacionada com a eliminação de alguma rotação ou translação sobre um determinado eixo ou com alguma relação de igualdade entre alguns dos parâmetros intrínsecos das câmaras, como se pode verificar pelas equações (2.19), (2.23) e (2.24).

Nesta subsecção, querendo deixar-se apenas a ideia da facilidade com que se pode explorar outras configurações, serão brevemente descritas as configurações de translação pura entre as imagens e a de movimentos planares.

No caso da translação pura entre as imagens não existe rotação entre as câmaras. Logo a matriz de rotação, \mathbf{R} , representada na figura 2.5, reduz-se à matriz identidade. Este facto leva a que, com base nas equações (2.19), (2.23) e (2.24), a matriz \mathbf{F} seja dada por

$$\mathbf{F} = \mathbf{C}_d^{-T} [\mathbf{t}]_{\wedge} \mathbf{C}_e^{-1} = [\mathbf{e}_d]_{\wedge} \mathbf{C}_d \mathbf{C}_e^{-1} = (\mathbf{C}_d \mathbf{C}_e^{-1})^{-T} [\mathbf{e}_e]_{\wedge}$$

sendo o vector \mathbf{t} a referida translação entre as câmaras.

Quando os parâmetros intrínsecos, associados a cada imagem, são os mesmos para ambas, isto é, $\mathbf{C}_e = \mathbf{C}_d = \mathbf{C}$, então a equação anterior é simplificada, usando a propriedade das matrizes inversas ($\mathbf{M}\mathbf{M}^{-1} = \mathbf{I}$, com \mathbf{I} a matriz identidade), a igualdade (2.22) e o facto de \mathbf{F} ser definido a menos de um factor de escala, passando a ser dada pela seguinte matriz antisimétrica

$$\mathbf{F} = [\mathbf{C}\mathbf{t}]_{\wedge} = [\mathbf{e}_d]_{\wedge} = [\mathbf{e}_e]_{\wedge} \quad (2.26)$$

Esta equação torna evidente que, para esta situação específica, o epipolo \mathbf{e} , dado por $\mathbf{e} = \mathbf{e}_e = \mathbf{e}_d = \begin{bmatrix} ke_x & ke_y & k \end{bmatrix}^T$, é o epipolo de ambas as imagens. Deste facto sai uma das propriedades para apenas esta configuração, referida como *auto epipolar*, a qual diz que, se for assumido que ambas as imagens estão sobrepostas, os pontos correspondentes são

colineares [Hartley 00]. Recorrendo à equação (2.6), pode-se retirar da equação anterior, através da multiplicação da matriz \mathbf{C} com o vector \mathbf{t} , que

$$\begin{bmatrix} ke_x \\ ke_y \\ k \end{bmatrix} = \begin{bmatrix} fk_x t_x + c_x t_z \\ fk_y t_y + c_y t_z \\ t_z \end{bmatrix} \quad (2.27)$$

mostrando-se que quando não houver uma translacção sobre o eixo dos ZZ ($t_z = 0$) o epipolo \mathbf{e} situar-se-à no infinito. A razão para este facto reside na representação de \mathbf{e} ser homogénea.

Por fim, para se verificar a simplificação de processos relacionados com a matriz fundamental, veja-se o seguinte exemplo. Sendo $\mathbf{t} = [1 \ 1 \ 0]^T$, através da equação (2.26) e da igualdade (2.27), a matriz fundamental vem dada por

$$\mathbf{F} = \begin{bmatrix} 0 & 0 & fk_y \\ 0 & 0 & -fk_x \\ -fk_y & fk_x & 0 \end{bmatrix}$$

Da relação (2.10) entre os pontos correspondentes da imagem esquerda e direita, respectivamente $\mathbf{p}_e = [x_e \ y_e \ 1]^T$ e $\mathbf{p}_d = [x_d \ y_d \ 1]^T$, resulta

$$(x_d - x_e)fk_y + (y_e - y_d)fk_x = 0 \Leftrightarrow k_a = \frac{k_y}{k_x} = \frac{y_d - y_e}{x_d - x_e}$$

onde k_a é quociente de aparência das imagens estéreo. Este parâmetro é muito utilizado quando se pretende reconstruções sem recurso a processos explícitos de calibração.

Antes de entrar no caso dos movimentos planares, convém mencionar que qualquer translacção e rotação é equivalente a uma rotação sobre um eixo a (paralelo ao eixo de rotação) juntamente com a translacção ao longo desse eixo [Hartley 00]. Assim, o caso dos movimentos planares acontece quando, existindo uma translacção \mathbf{t} e uma rotação \mathbf{R} segundo um qualquer eixo, a translacção é ortogonal a esse eixo. Noutras palavras, a câmara desloca-se ao longo de um plano perpendicular ao eixo de rotação, efectuando também uma rotação em torno desse eixo.

Como o modo de raciocínio para os três casos, movimento da câmara no plano XZ , YZ e XY , é semelhante, apresenta-se apenas o movimento no plano XZ . Num exemplo de uma destas configurações, conhecido por elevação comum [Li 96], os parâmetros de \mathbf{F} estão relacionados de tal forma que só existem quatro elementos independentes.

O vector de translação da câmara, no plano XZ , é dado por $\mathbf{t} = \begin{bmatrix} t_x & 0 & t_z \end{bmatrix}^T$, enquanto que a matriz de rotação em torno do eixo Y é $\mathbf{R} = \mathbf{R}_\beta$, (ver equação (2.4) e figura 2.2). A matriz fundamental é conseguida, uma vez mais, substituindo o referido vector \mathbf{t} e a referida matriz \mathbf{R} na equação (2.19), na equação (2.23) ou na equação (2.24).

Para o plano YZ , a constituição de \mathbf{F} difere apenas em $\mathbf{t} = \begin{bmatrix} 0 & t_y & t_z \end{bmatrix}^T$ e $\mathbf{R} = \mathbf{R}_\alpha$ (definida em (2.4)). Para o movimento no plano XY , a matriz fundamental é a mesma que a do caso das câmaras afins. Este facto é devido à rotação ser sobre o eixo que passa pelo centro de projecção.

A tabela 2.1 resume as configurações específicas apresentadas, expandindo a sua formulação matricial.

<i>Tipos de configurações</i>	<i>Formas genéricas de \mathbf{F}</i>
Translação pura	<p>Forma geral</p> $\mathbf{F} = \begin{bmatrix} 0 & f_{12} & f_{13} \\ f_{21} & 0 & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}$ <p>Mesmas matrizes de calibração</p> $\mathbf{F} = \begin{bmatrix} 0 & f_{12} & -f_{13} \\ -f_{12} & 0 & f_{23} \\ f_{13} & -f_{23} & 0 \end{bmatrix}$
Movimentos planares	<p>No plano XZ ou YZ</p> $\mathbf{F} = \begin{bmatrix} 0 & f_{12} & f_{13} \\ f_{21} & 0 & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}$ <p>No plano XY</p> $\mathbf{F} = \begin{bmatrix} 0 & 0 & f_{13} \\ 0 & 0 & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}$

Tabela 2.1: *Diferentes maneiras que a geometria epipolar, na forma da matriz fundamental, \mathbf{F} , toma para configurações específicas (translação pura e movimentos planares).*

2.4.2 Estimação da matriz fundamental

O problema considerado de seguida é o da estimação da geometria epipolar na forma de matriz fundamental, \mathbf{F} . Sejam, então,

$$\mathbf{p}_e = \begin{bmatrix} x_e & y_e & 1 \end{bmatrix}^T \quad \mathbf{p}_d = \begin{bmatrix} x_d & y_d & 1 \end{bmatrix}^T$$

os pontos correspondentes na imagem esquerda e direita. Expandindo a equação (2.10) obtém-se

$$\begin{bmatrix} x_d & y_d & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} x_e \\ y_e \\ 1 \end{bmatrix} = 0$$

ou, de outra forma,

$$f_{11}x_ex_d + f_{12}x_dy_e + f_{13}x_d + f_{21}x_ey_d + f_{22}y_ey_d + f_{23}y_d + f_{31}x_e + f_{32}y_e + f_{33} = 0$$

Reformulando a equação anterior como uma equação linear homogénea, com nove parâmetros, o seu aspecto será

$$\mathbf{g}_i \mathbf{f} = 0 \quad (2.28)$$

onde

$$\mathbf{g}_i = \begin{bmatrix} x_ex_d & x_dy_e & x_d & x_ey_d & y_ey_d & y_d & x_e & y_e & 1 \end{bmatrix}$$

e

$$\mathbf{f} = \begin{bmatrix} f_{11} & f_{12} & f_{13} & f_{21} & f_{22} & f_{23} & f_{31} & f_{32} & f_{33} \end{bmatrix}^T$$

O próximo passo no processo de estimação pode ser descrito da forma que se segue. Se tivermos n pontos correspondentes (logo i a variar de 1 a n), usando a equação (2.28), poderemos construir um sistema de equações do tipo $\mathbf{A}\mathbf{f} = \mathbf{0}$, com

$$\mathbf{A} = \begin{bmatrix} \mathbf{g}_1 & \mathbf{g}_2 & \cdots & \mathbf{g}_n \end{bmatrix}^T \quad (2.29)$$

A estimação da matriz fundamental, \mathbf{F} , fica concluída quando se obtém a solução deste sistema linear homogéneo.

O sistema linear criado pode ter zero, uma, ou mais do que uma soluções, dependendo da característica da sua matriz \mathbf{A} , como se constata de imediato.

Pela equação (2.28), este sistema de equações homogéneo tem nove incógnitas, pois o vector \mathbf{f} tem exactamente esse número de elementos. No entanto, devido ao facto de \mathbf{F} ter so-

mente sete graus de liberdade, apenas sete desses elementos são linearmente independentes. Assim, a matriz \mathbf{A} tem que ter uma característica limitada entre sete e nove.

Tendo delimitado as características que \mathbf{A} pode ter, convém esclarecer quais as situações em que cada uma acontece, começando pela situação óptima. Assim, quando a matriz \mathbf{A} tiver característica oito, do sistema $\mathbf{A}\mathbf{f} = \mathbf{0}$ resultam apenas oito elementos linearmente independentes, dos nove que compõem \mathbf{f} . No entanto, como qualquer matriz fundamental é definida a menos de um factor multiplicativo, a solução obtida define univocamente a matriz fundamental. Para a obter recorre-se à definição de espaço nulo direito de uma matriz, pois o vector \mathbf{f} é o espaço nulo direito de \mathbf{A} . Esta solução é única porque, através da expressão (2.25), tem-se $n = 9$, $C(\mathbf{A}) = 8$ e, por consequência, apenas um espaço nulo. Logo, qualquer situação que saia fora deste caso é degenerativa, pois irão ser encontradas mais ou menos soluções que a obtida neste caso. Assim, pode dizer-se, que um conjunto de pontos correspondentes cria uma matriz \mathbf{A} , do tipo (2.29), geometricamente degenerada em relação à matriz fundamental, \mathbf{F} , se a dimensão do seu espaço nulo direito for diferente de um.

A situação em que a característica de \mathbf{A} é igual a sete gera dois espaços nulos, obtendo mais do que uma solução. Na estimação de \mathbf{F} , esta situação aparece quando o sistema $\mathbf{A}\mathbf{f} = \mathbf{0}$ tem apenas sete equações linearmente independentes. Deste caso resulta uma ou três soluções para o vector \mathbf{f} [Hartley 00]. Posteriormente apresentar-se-á o método numérico para as calcular.

Se, por fim, os dados de entrada para a estimação de \mathbf{f} não forem exactos, devido ao ruído nas coordenadas dos pontos correspondentes, então não existe solução. Esta é a situação em que a característica de \mathbf{A} é nove e, portanto, não existem espaços nulos associados à matriz \mathbf{A} . Nesta situação terá que se usar o método dos mínimos quadrados para obter uma solução aproximada. Assim, como o sistema de equações $\mathbf{A}\mathbf{f} = \mathbf{0}$ não tem solução, garante-se sempre uma.

No apêndice B apresenta-se a informação necessária à utilização de uma função para o MATLAB®(mfest.m), na qual se faz a estimação da matriz fundamental para qualquer das situações referidas.

De seguida abordar-se-ão as seguintes questões importantes para a estimação da matriz fundamental:

- restrição da característica

- normalização
- eliminação de falsas correspondências

Serão, também, apresentados dois dos mais importantes métodos de estimação da matriz fundamental: os algoritmos de sete e de oito pontos.

Restrição da característica

Sendo \mathbf{A} uma matriz $m \times n$, a restrição de característica aplica-se quando se quer obter a matriz que mais perto esteja de \mathbf{A} e que tenha uma menor característica, ou quando se quer saber a distância entre essas duas matrizes. Segundo Datta [Datta 94], a distância entre uma matriz \mathbf{A} e a matriz de menor característica (mais próxima dela) é dada pelo menor valor singular de \mathbf{A} , diferente de zero. Em particular, se \mathbf{A} for uma matriz quadrada, não singular, aquele valor é sempre a distância à matriz mais próxima.

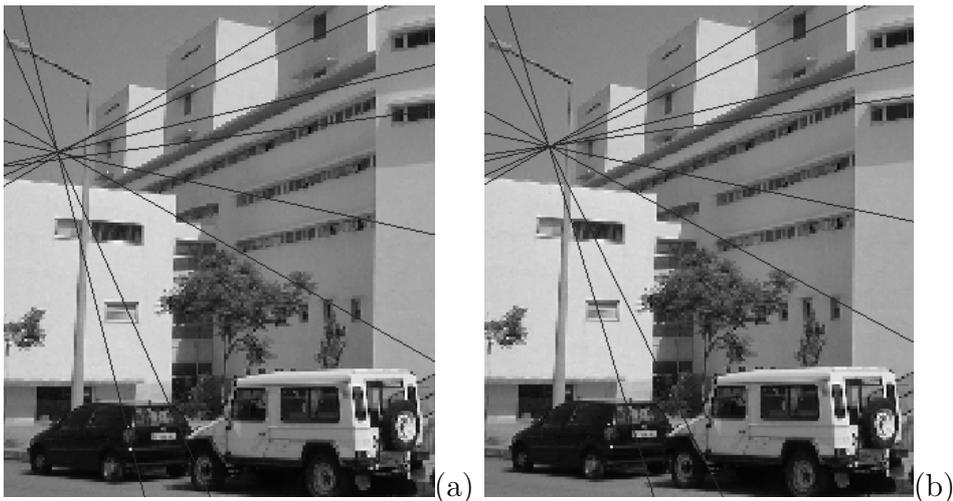


Figura 2.6: (a) Efeito de uma matriz fundamental não singular no cálculo e respectiva representação das rectas epipolares. (b) Efeito de uma matriz fundamental singular no cálculo e respectiva representação das rectas epipolares. Esta matriz é obtida forçando a matriz da figura (a) a ter característica dois.

Uma das mais importantes propriedades da matriz fundamental, \mathbf{F} , é a de ser singular, especificamente, de ter característica dois. Sendo assim, muitas das aplicações que usam a matriz fundamental baseiam-se nessa propriedade. Por exemplo, se \mathbf{F} não for singular, então as rectas epipolares não se intersectam num único ponto, como se pode verificar na

figura 2.6. Isto acontece porque não existe a possibilidade de encontrar os espaços nulos de \mathbf{F} , e, portanto, os epipolos.

Assim, observando a solução do sistema linear $\mathbf{A}\mathbf{f} = \mathbf{0}$, que resulta das equações (2.28) e (2.29), verifica-se que, excepto para pontos correspondentes muito precisos, a matriz fundamental, \mathbf{F} , depois de estimada, não tem determinante nulo, ou, por outras palavras, não tem característica dois. Esta é uma das principais falhas dos métodos lineares de estimação da matriz \mathbf{F} , e é uma das razões para a sensibilidade da matriz fundamental ao ruído. Portanto, para que a estimação da matriz fundamental seja considerada completa, \mathbf{F} deve ter característica dois. Se assim não acontecer, deve ser executado algum processo de imposição dessa característica.

O processo de imposição da restrição de característica pode ser feito de duas maneiras, explicitamente dentro do método de estimação [Agapito 97] ou posteriormente à estimação. Não sendo objectivo desta dissertação um estudo aprofundado deste processo, apresentam-se, de seguida, apenas os dois métodos mais significativos para o fazer: o da imposição da característica *a posteriori* e o algoritmo de minimização algébrica.

O método da imposição da característica *a posteriori* consiste na substituição da matriz \mathbf{F} estimada, por uma matriz \mathbf{F}' que minimiza a norma de Frobenius, $\|\mathbf{F}-\mathbf{F}'\|$, sujeita a $\det(\mathbf{F}') = 0$ [Datta 94]. Isto é conseguido, impondo ao menor dos valores singulares de \mathbf{F} o valor zero. Se a decomposição em valores singulares de \mathbf{F} é dada por

$$\mathbf{F} = \mathbf{U} \begin{bmatrix} \tau_1 & 0 & 0 \\ 0 & \tau_2 & 0 \\ 0 & 0 & \tau_3 \end{bmatrix} \mathbf{V}^T$$

em que $\tau_1 \geq \tau_2 \geq \tau_3$ são os valores singulares de \mathbf{F} , impõem-se que τ_3 seja igual a zero. A nova matriz fundamental que respeita a restrição da característica dois virá igual a

$$\mathbf{F}' = \mathbf{U} \begin{bmatrix} \tau_1 & 0 & 0 \\ 0 & \tau_2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{V}^T$$

Este método foi usado pela primeira vez por Tsai e Huang [Tsai 84] na estimação da matriz essencial, a qual, à semelhança da matriz fundamental, teria que ter característica dois (ver equação (2.17)). Segundo Hartley [Hartley 95], a imposição da característica *a posteriori* torna-se mais tolerante a falhas quando se aplica aos dados de entrada uma

operação numérica denominada normalização. Veremos mais à frente em que consiste esta operação.

Este método peca por não considerar a diferença de importância existente entre os elementos de \mathbf{F} . Na verdade alguns estão muito mais ligados aos pontos correspondentes de entrada do que outros.

O procedimento mais correcto seria calcular a matriz de covariância a partir dos elementos de \mathbf{F} em termos dos dados de entrada e, de seguida, obter a matriz singular que mais perto está de \mathbf{F} , através da minimização da distância de *Mahalanobis*. Como esta minimização não pode ser feita linearmente para qualquer matriz de covariância [Hartley 00], este método deixa de ser atractivo.

A alternativa possível, com resultados muito semelhantes ao processo de minimização da distância de *Mahalanobis*, é a de estimar a matriz singular, \mathbf{F} , directamente. Para tal, o procedimento mais conhecido é o algoritmo de minimização algébrica.

Este algoritmo começa com a substituição de $(\mathbf{C}_d \mathbf{R} \mathbf{C}_e^{-1})^{-T}$ por \mathbf{H} , na equação (2.24), da qual resulta

$$\mathbf{F} = \mathbf{H}[\mathbf{e}_e]_{\wedge} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} [\mathbf{e}_e]_{\wedge} = \begin{bmatrix} h_{12} - h_{13}e_y & h_{13}e_x - h_{11} & h_{11}e_y - h_{12}e_x \\ h_{22} - h_{23}e_y & h_{23}e_x - h_{21} & h_{21}e_y - h_{22}e_x \\ h_{32} - h_{33}e_y & h_{33}e_x - h_{31} & h_{31}e_y - h_{32}e_x \end{bmatrix} \quad (2.30)$$

A matriz \mathbf{H} é não singular e $[\mathbf{e}_e]_{\wedge}$ é a matriz antisimétrica e singular, que usa as coordenadas homogéneas do epipolo, dadas por $\mathbf{e}_e = [e_x \ e_y \ 1]^T$. Considerando que os elementos de um vector $\mathbf{f} = [f_{11} \ f_{12} \ f_{13} \ f_{21} \ f_{22} \ f_{23} \ f_{31} \ f_{32} \ f_{33}]^T$ são os mesmos que os da matriz \mathbf{F} , e que os elementos da matriz \mathbf{H} são os mesmos que os de um vector $\mathbf{h} = [h_{11} \ h_{12} \ h_{13} \ h_{21} \ h_{22} \ h_{23} \ h_{31} \ h_{32} \ h_{33}]^T$, então a equação (2.30) pode ser reformulada em

$$\mathbf{f} = \begin{bmatrix} -[\mathbf{e}_e]_{\wedge} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -[\mathbf{e}_e]_{\wedge} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -[\mathbf{e}_e]_{\wedge} \end{bmatrix} \mathbf{h} = \mathbf{G}\mathbf{h} \quad (2.31)$$

pois

$$\mathbf{G}\mathbf{h} = \begin{bmatrix} -[\mathbf{e}_e]_{\wedge} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -[\mathbf{e}_e]_{\wedge} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -[\mathbf{e}_e]_{\wedge} \end{bmatrix} \mathbf{h} =$$

$$= \begin{bmatrix} 0 & 1 & -e_y & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & e_x & 0 & 0 & 0 & 0 & 0 & 0 \\ e_y & -e_x & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -e_y & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & e_x & 0 & 0 & 0 \\ 0 & 0 & 0 & e_y & -e_x & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -e_y \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & e_x \\ 0 & 0 & 0 & 0 & 0 & 0 & e_y & -e_x & 0 \end{bmatrix} \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \\ h_{33} \end{bmatrix} = \begin{bmatrix} h_{12} - h_{13}e_y \\ h_{13}e_x - h_{11} \\ h_{11}e_y - h_{12}e_x \\ h_{22} - h_{23}e_y \\ h_{23}e_x - h_{21} \\ h_{21}e_y - h_{22}e_x \\ h_{32} - h_{33}e_y \\ h_{33}e_x - h_{31} \\ h_{31}e_y - h_{32}e_x \end{bmatrix}$$

A estimação da matriz fundamental, \mathbf{F} , parte do sistema de equações $\mathbf{A}\mathbf{f} = \mathbf{0}$, como foi visto atrás. Sabendo que a matriz \mathbf{F} é definida a menos de um factor multiplicativo, assume-se, sem perda de generalidade, que a norma do vector \mathbf{f} é um [Hartley 00]. Como tal, a solução deste problema passa por

$$\min_{\mathbf{f}} \|\mathbf{A}\mathbf{f}\| \quad s.a \quad \|\mathbf{f}\| = 1 \quad (2.32)$$

Substituindo a equação (2.31) em (2.32), fica-se com

$$\min_{\mathbf{G}\mathbf{h}} \|\mathbf{A}\mathbf{G}\mathbf{h}\| \quad s.a \quad \|\mathbf{G}\mathbf{h}\| = 1 \quad (2.33)$$

Da solução para a minimização (2.33) resultam os elementos da matriz \mathbf{H} , e, por consequência, os elementos da matriz \mathbf{F} pretendida. No entanto, este procedimento só é possível se a matriz \mathbf{A} tiver característica oito ou nove.

Se o epipolo não for conhecido *a priori* terá que se recorrer à iteratividade, devido ao facto da equação $\det(\mathbf{F}) = 0$ ser cúbica, em vez de linear. Assim, começa-se por estimar uma matriz fundamental, \mathbf{F}_i , representada pelo vector \mathbf{f}_i , da qual se retira o epipolo $\mathbf{e}_i = \mathcal{N}_d(\mathbf{F}_i)$ (neste caso $i = 0$). A partir deste epipolo encontra-se um novo vector \mathbf{f}_i que minimiza a equação (2.33), e com ele calcula-se o erro algébrico $\epsilon_i = \mathbf{A}\mathbf{f}_i$. Posteriormente, utiliza-se o algoritmo de Levenberg-Marquardt para variar \mathbf{e}_i iterativamente de forma a minimizar $\|\epsilon_i\|$. O vector para onde todos os vectores \mathbf{f}_i convergem é o vector que representa a matriz fundamental desejada.

O processo acima descrito pode, também, com base na equação (2.23), partir de

$$\mathbf{F} = [\mathbf{e}_d]_{\wedge} \mathbf{H} = \begin{bmatrix} h_{31}e_y - h_{21} & h_{32}e_y - h_{22} & h_{33}e_y - h_{23} \\ h_{11} - h_{31}e_x & h_{12} - h_{32}e_x & h_{13} - h_{33}e_x \\ h_{21}e_x - h_{11}e_y & h_{22}e_x - h_{12}e_y & h_{23}e_x - h_{13}e_y \end{bmatrix}$$

onde $e_d = (e_x, e_y)$. Para este caso, sendo \mathbf{I}_3 a matriz identidade 3×3 , a equação (2.31) passa a ser dada por

$$\mathbf{f} = \begin{bmatrix} \mathbf{0} & -\mathbf{I}_3 & e_y \mathbf{I}_3 \\ \mathbf{I}_3 & \mathbf{0} & -e_x \mathbf{I}_3 \\ -e_y \mathbf{I}_3 & e_x \mathbf{I}_3 & \mathbf{0} \end{bmatrix} \mathbf{h} = \mathbf{Gh} \quad (2.34)$$

pois

$$\mathbf{Gh} = \begin{bmatrix} \mathbf{0} & -\mathbf{I}_3 & e_y \mathbf{I}_3 \\ \mathbf{I}_3 & \mathbf{0} & -e_x \mathbf{I}_3 \\ -e_y \mathbf{I}_3 & e_x \mathbf{I}_3 & \mathbf{0} \end{bmatrix} \mathbf{h} =$$

$$= \begin{bmatrix} 0 & 0 & 0 & -1 & 0 & 0 & e_y & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & e_y & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & e_y \\ 1 & 0 & 0 & 0 & 0 & 0 & -e_x & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -e_x & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -e_x \\ -e_y & 0 & 0 & e_x & 0 & 0 & 0 & 0 & 0 \\ 0 & -e_y & 0 & 0 & e_x & 0 & 0 & 0 & 0 \\ 0 & 0 & -e_y & 0 & 0 & e_x & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \\ h_{33} \end{bmatrix} = \begin{bmatrix} h_{31}e_y - h_{21} \\ h_{32}e_y - h_{22} \\ h_{33}e_y - h_{23} \\ h_{11} - h_{31}e_x \\ h_{12} - h_{32}e_x \\ h_{13} - h_{33}e_x \\ h_{21}e_x - h_{11}e_y \\ h_{22}e_x - h_{12}e_y \\ h_{23}e_x - h_{13}e_y \end{bmatrix}$$

Apresenta-se no apêndice B a informação necessária à utilização das funções `imprank.m` e `mfminasi.m` no MATLAB©, as quais fazem, respectivamente, a imposição da restrição de característica *a posteriori* e a estimação de \mathbf{F} através do método da minimização algébrica. A função `mfminasi.m` produz, além da matriz fundamental, \mathbf{F} , uma matriz \mathbf{H} .

Normalização

Devido à falta de estabilidade nos sistemas lineares criados através das equações (2.28) e (2.29), Hartley em [Hartley 95] argumenta que, normalizando o conjunto de pontos correspondentes a introduzir nesse sistema, antes da estimação de \mathbf{f} , se consegue melhorar o desempenho do algoritmo de estimação. De acordo com Hartley, a análise do problema conduz à aplicação, separadamente para cada conjunto de pontos da respectiva imagem, das seguintes acções:

- translacionar as coordenadas do ponto correspondente, de forma a que a sua origem fique no centro de massa do conjunto de pontos. Normalmente o centro das coor-

denadas da imagem está no canto superior esquerdo, tornando obscuros os valores significativos das coordenadas;

- escalonar os eixos da imagem de modo a que a distância média à origem seja $\sqrt{2}$. Este escalonamento é conseguido calculando, para todos os pontos correspondentes, a média euclidiana da distância à origem. Depois, obtém-se o factor de escala com a divisão de $\sqrt{2}$ pela média encontrada. Este processo termina quando todas as coordenadas, do conjunto dos pontos correspondentes, forem multiplicados pelo factor de escala.

Esta transformação das coordenadas *pixel*, das imagens, tem a vantagem de melhor balancear os elementos de \mathbf{f} . Por isso, esta acção deverá ser executada antes da formação das equações (2.28), do sistema. Matematicamente, a operação de normalização transforma as coordenadas homogêneas da imagem, \mathbf{p}_e e \mathbf{p}_d , de acordo com

$$\mathbf{p}_{en} = \mathbf{M}_e \mathbf{p}_e \quad \mathbf{p}_{dn} = \mathbf{M}_d \mathbf{p}_d$$

onde \mathbf{M}_e e \mathbf{M}_d são as matrizes que integram todas as transformações de normalização descritas acima.

Depois da matriz fundamental, \mathbf{F} , estar estimada, onde se inclui a imposição da característica dois, deve ser-lhe aplicada uma desnormalização, para que qualquer ponto não normalizado da imagem possa interagir com ela. A operação de desnormalização da matriz normalizada, \mathbf{F}_n , é dada por

$$\mathbf{F} = \mathbf{M}_d^T \mathbf{F}_n \mathbf{M}_e$$

A operação de desnormalização não é propriamente inversa à da normalização pois, apesar da matriz poder voltar a interagir com os pares de pontos correspondentes não normalizados das imagens, foi eliminado da matriz desnormalizada algum ruído que permaneceria se não fosse feito o processo de normalização/desnormalização.

Uma vez mais, remete-se para o apêndice B a apresentação do modo de utilização, no MATLAB©, da função de normalização de um conjunto de pontos de uma imagem, `htnorm2d.m`.

Convém mencionar que, devido aos seus resultados na redução do ruído, o processo de normalização/desnormalização aqui descrito deve ser usado em qualquer método onde intervenham dados bidimensionais [Hartley 98].

Eliminação de falsas correspondências

Como foi visto, a estimação da matriz fundamental faz-se tendo por base os pontos correspondentes nas imagens. Todos os métodos apresentados consideram que a posição desses pontos nas imagens é quase exacta (poderão ter apenas um pequeno ruído). No entanto, esta hipótese nem sempre é satisfeita. Neste sentido, não é estranho que, por vezes, na estimação da matriz fundamental se encontre um processamento de eliminação de pontos que não fazem sentido no seu conjunto global. Este processo é conhecido como eliminação de falsas correspondências.

Um dos métodos que se pode usar para executar este processamento é o RANSAC, o qual será explicado com pormenor na secção 2.7. A utilização do RANSAC na estimação da matriz fundamental, \mathbf{F} , integra-se num processo iterativo que começa por escolher sete pares de pontos correspondentes aleatórios, com os quais calcula a matriz \mathbf{F} . O facto de serem apenas sete traz a vantagem da matriz fundamental ser estimada já com característica dois, sem ter que a impôr. Outra razão para esta opção, em vez do uso de oito pares de pontos correspondentes, é que o número de iterações que deverão ser feitas, para assegurar uma alta probabilidade de eliminação de verdadeiros pontos correspondentes que não interessam à estimação, é exponencial em relação ao tamanho do conjunto de dados de entrada ($7^n < 8^n$) [Hartley 00]. A única desvantagem do uso de sete pares de pontos correspondentes na estimação de \mathbf{F} é a de poderem resultar três matrizes em vez de uma, tornando o esforço de cálculo três vezes maior.

Posteriormente, com a matriz \mathbf{F} , verifica-se quão perto de satisfazerem a geometria epipolar, estão todos os pontos correspondentes, do conjunto de dados inicial. Como essa verificação se baseia na distância, poderão existir duas escolhas claras para o cálculo dessa distância. A primeira é conhecida como erro de reprojecção [Zhang 96], enquanto que a segunda é referida como aproximação de Sampson ao erro de reprojecção [Hartley 00]. No entanto, devido à complexidade de ambas as maneiras, optou-se por usar a qualidade da correspondência da matriz fundamental (ver apêndice A), pois obtém resultados satisfatórios [Zhang 96].

Depois de se fazer esta verificação dá-se início a nova iteração. O resultado final é o que proporciona uma matriz que tem mais pares de pontos correspondentes a satisfazerem a geometria epipolar, dentro de um nível de erro especificado.

A informação necessária ao uso, no MATLAB®, da função que estima \mathbf{F} eliminando as

falsas correspondências existentes no conjunto de dados de entrada, `mfransac.m`, é apresentada no apêndice B.

Algoritmo dos sete pontos

Este algoritmo é aplicado apenas quando a característica da matriz \mathbf{A} , dada pela equação (2.29), é sete. Na verdade, este método só encontra soluções para o sistema $\mathbf{A}\mathbf{f} = \mathbf{0}$ porque faz uso da propriedade de singularidade da matriz fundamental, \mathbf{F} ($\det(\mathbf{F}) = 0$).

A solução do sistema de equações $\mathbf{A}\mathbf{f} = \mathbf{0}$, neste caso, está no espaço bidimensional de

$$\alpha\mathbf{F}_1 + (1 - \alpha)\mathbf{F}_2 \quad (2.35)$$

onde α é um escalar. Por seu lado, as matrizes \mathbf{F}_1 e \mathbf{F}_2 são dadas pelo espaço nulo direito da matriz \mathbf{A} , que nesta situação tem dimensão dois ($\mathcal{N}_d(\mathbf{A}) = \{\mathbf{f}_1, \mathbf{f}_2\}$).

Usando agora a restrição de singularidade da matriz \mathbf{F} encontram-se as soluções da equação polinomial cúbica

$$\det(\alpha\mathbf{F}_1 + (1 - \alpha)\mathbf{F}_2) = 0 \quad (2.36)$$

que na realidade são os possíveis valores que α pode ter de forma a satisfazer a equação (2.36). Essas soluções serão todas reais ou, então, serão duas complexas e uma real. As soluções complexas nunca se consideram [Hartley 00].

Com as soluções reais obtêm-se as matrizes fundamentais através da equação (2.35). A melhor solução, caso hajam três soluções reais para α , terá posteriormente que ser obtida comparando todas as matrizes encontradas. Para mais pormenores veja-se [Zhang 96] e [Hartley 00].

No apêndice B, mostra-se a informação necessária à utilização, no MATLAB®, da função `mfmin.m`, a qual estima \mathbf{F} quando a característica de \mathbf{A} é apenas sete. Aí, a comparação entre as matrizes fundamentais encontradas é feita com base na distância dos pontos usados na estimação às respectivas rectas epipolares. Esta comparação designa-se por qualidade da correspondência da matriz fundamental (veja-se o apêndice A).

Algoritmo dos oito pontos

Como já foi referido, a solução óptima na estimação da matriz fundamental ocorre quando a matriz \mathbf{A} , do sistema $\mathbf{A}\mathbf{f} = \mathbf{0}$, tem característica oito. Assim, surgiu a ideia de forçar essa característica, impondo a um dos elementos de \mathbf{f} o valor 1. À partida, parece que com esta

mudança se está a encontrar uma solução particular, mas no caso da matriz fundamental, \mathbf{F} , em que a solução final é definida a menos de um factor de escala, obtém-se a melhor solução possível. A restrição que é imposta, tem a propriedade de eliminar a solução trivial do sistema homogéneo inicial, que normalmente não é a pretendida [Zhang 96].

Este tipo de resolução, aplicado somente quando se pretendem soluções definidas a menos de um factor multiplicativo, foi descrito e usado pela primeira vez por Longuet-Higgins [Longuet-Higgins 81] na estimação da geometria epipolar. Ficou conhecido, nesse campo de estimação, por algoritmo de oito pontos. Apesar de bastarem apenas sete pontos para determinar a matriz fundamental, o algoritmo de oito pontos é o processo mais comum na estimação de \mathbf{F} . Nele utilizam-se, sempre, pelo menos oito pontos.

A escolha da variável a restringir não é arbitrária, como é discutido em [Luong 93]. Para melhor se perceber este facto, seja $\mathbf{f} = [f_1 \ f_2 \ \cdots \ f_n]^T$, a solução óptima final de um qualquer sistema linear de equações. Suponha-se que $f_i \ll 1$ (ou $f_i \gg 1$), e que $f_k \simeq 1$, com $1 \leq i, k \leq n$. Se no processo de estimação do vector \mathbf{f} restringíssemos $f_k = 1$, obteríamos um valor mais próximo da solução óptima final, do que se tivéssemos feito $f_i = 1$. Como é obvio, se forçarmos uma das variáveis a ser um, e na realidade ela é muito mais pequena (ou muito maior), o resultado da estimação será um desastre. Como não sabemos, *a priori*, qual das variáveis está mais próxima da unidade, a solução para este problema consiste em testar todas as combinações e guardar a melhor estimação.

Torna-se conveniente nesta fase abordar a técnica de passagem de um sistema linear do tipo $\mathbf{A}\mathbf{f} = \mathbf{0}$ para o seu equivalente do tipo $\mathbf{A}'\mathbf{f}' = \mathbf{b}$. Recuperando a equação (2.28), atribui-se à variável f_{33} o valor de um, para uma melhor explanação. Assim, reformulando a equação (2.28) através da aplicação deste conhecimento obtém-se

$$\mathbf{g}'_i \mathbf{f}' = -1$$

onde

$$\mathbf{g}'_i = \begin{bmatrix} x_e x_d & x_d y_e & x_d & x_e y_d & y_e y_d & y_d & x_e & y_e \end{bmatrix}$$

e

$$\mathbf{f}' = \begin{bmatrix} f_{11} & f_{12} & f_{13} & f_{21} & f_{22} & f_{23} & f_{31} & f_{32} \end{bmatrix}^T$$

Tendo um conjunto de n pontos correspondentes, formamos, à semelhança do que foi

feito na equação (2.29), um sistema linear de n equações e oito incógnitas $\mathbf{A}'\mathbf{f}' = \mathbf{b}$, onde

$$\mathbf{A}' = \begin{bmatrix} \mathbf{g}'_1 \\ \mathbf{g}'_2 \\ \dots \\ \mathbf{g}'_n \end{bmatrix}$$

e o vector \mathbf{b} é composto por n valores -1 .

Tal como se viu para f_{33} , pode-se atribuir a qualquer outro elemento do vector \mathbf{f} , o valor 1. Nesse caso, para f_{ij} , com $1 \leq i, j \leq 3$, elimina-se da equação (2.28) os elementos g_{ij} e f_{ij} , e cria-se um vector \mathbf{b} com n valores iguais a $-g_{ij}$. Esta situação faz com que o vector \mathbf{f} passe a ter norma 1, em vez da inicial norma 0. O facto da norma do vector \mathbf{f} ser um é, na verdade, o algoritmo dos oito pontos.

Se, neste método, forem usados exactamente oito pontos correspondentes, o resultado teria muito provavelmente grandes erros, devido à inexactidão das posições tridimensionais dos pontos, escolhidos nas imagens [Hartley 95]. O uso de mais pontos leva-nos, assim, apesar de se ficar com um sistema sobre-determinado, a uma melhor solução. Nesta situação, o processo matemático de estimação pode ser obtido através de um método numérico que aplique os mínimos quadrados, não esquecendo de sujeitar a solução à condição da sua norma ser um. A matriz fundamental resultante, só em casos raros origina uma matriz que valida a restrição de característica dois.

No apêndice B, encontra-se a informação necessária à utilização, no MATLAB®, da função que estima \mathbf{F} através do método dos oito pontos, `mf8pta.m`.

Outras considerações

Foram apresentadas, nesta secção, várias maneiras de estimação da matriz fundamental. Torna-se, no entanto, importante mencionar algumas notas sucintas sobre qual o método mais desejável para certas pretensões.

Assim, se for pretendido um método rápido e fácil de implementar, deve-se usar o algoritmo dos oito pontos normalizado. Este método origina quase sempre resultados adequados, sendo ideal como primeira fase de outros métodos, como por exemplo o método da minimização algébrica. Se, por outro lado, for desejável mais precisão na estimação da matriz fundamental, deve-se usar o método da minimização algébrica, com ou sem a iteração sobre a posição do epipolo.

Os métodos de estimação que se apresentou são, segundo Hartley [Hartley 00], suficientes para a obtenção de excelentes estimações. No entanto, a estimação da matriz fundamental pode ser feita, alternativamente, através de outras formas [Zhang 96]. Por opção, essas maneiras não fazem parte deste estudo, uma vez que não são imprescindíveis para se atingir o trabalho proposto.

2.4.3 Obtenção da distância focal

Segundo [Nayar 98] é possível recuperar a distância focal de uma câmara a partir da matriz fundamental, \mathbf{F} . Este dado é um dos mais importantes a conhecer quando se executam, por exemplo, processos de reconstrução tridimensional. A sua determinação é feita da maneira que se segue.

Sendo λ_1 e λ_2 os respectivos valores próprios positivos e negativos de $\mathbf{F}_s = \mathbf{F} + \mathbf{F}^T$ e sendo \mathbf{v}_1 e \mathbf{v}_2 os respectivos vectores próprios associados a λ_1 e λ_2 , define-se

$$\mathbf{m} = \sqrt{\lambda_1}\mathbf{v}_1 + \sqrt{-\lambda_2}\mathbf{v}_2$$

ou

$$\mathbf{m} = \sqrt{\lambda_1}\mathbf{v}_1 - \sqrt{-\lambda_2}\mathbf{v}_2$$

Esta ambiguidade pode ser resolvida da seguinte forma, aquela que das expressões anteriores for equivalente a $\mathbf{l} = e_e \wedge e_d$ é ignorada. A recuperação da distância focal f é feita através da condição

$$\frac{\mathbf{l}^T \mathbf{w} \mathbf{m}}{\sqrt{(\mathbf{l}^T \mathbf{w} \mathbf{l})(\mathbf{m}^T \mathbf{w} \mathbf{m})}} = 0$$

onde

$$\mathbf{w} = \begin{bmatrix} f^2 & 0 & 0 \\ 0 & f^2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

2.5 Transformações projectivas

O importante problema da correspondência, além de poder ser tratado através do uso da geometria epipolar, pode também ser resolvido com recurso às transformações projectivas.

Por definição, chama-se transformação projectiva (ou colineação) de pontos do espaço vectorial \mathcal{P}^n em pontos do espaço vectorial \mathcal{P}^m , a toda a aplicação linear, em coordenadas homogéneas, que se exprime, na sua forma matricial, por

$$\begin{aligned} \mathcal{P}^n &\longrightarrow \mathcal{P}^m \\ \mathbf{P}_n &\longmapsto \mathbf{P}_m = \mathbf{H}\mathbf{P}_n \end{aligned} \tag{2.37}$$

A matriz \mathbf{H} , que define esta transformação, tem $(m+1) \times (n+1)$ elementos.

Convém recordar que os vectores \mathbf{P}_n e \mathbf{P}_m estão definidos a menos de um factor multiplicativo, pois são usadas as suas coordenadas homogéneas. Este facto implica que a matriz \mathbf{H} seja, também ela, definida a menos de um factor de escala. Por sua vez, esta situação leva a que qualquer transformação projectiva de \mathcal{P}^n em \mathcal{P}^m tenha exactamente $(m+1) \times (n+1) - 1$ graus de liberdade.

Dentro das muitas transformações projectivas existentes, interessará a esta dissertação a transformação projectiva entre dois espaços vectoriais iguais ($m = n$). Neste caso, todas as transformações projectivas são designadas por homografias e definidas através de uma matriz não singular. Para um dado n , o conjunto de todas as homografias possíveis forma um grupo que se designa por grupo projectivo de grau n . Particularizando ainda mais, irá ser abordado apenas o caso do grupo projectivo de grau 2 ($m = n = 2$). A razão para se fazer apenas esta abordagem é óbvia, pois pretende-se relacionar imagens, que são espaços vectoriais bidimensionais, de forma a serem encontrados pontos correspondentes entre elas.

2.5.1 Cálculo de transformações projectivas

O cálculo de qualquer transformação projectiva, \mathbf{H} , a menos de um factor de escala, não é mais do que a determinação dos seus $(m+1) \times (n+1) - 1$ parâmetros independentes. Para isso, como se pode constatar através da equação (2.37), é necessário, à partida, o conhecimento de alguns pontos correspondentes nos dois espaços vectoriais.

No processo de cálculo de transformações projectivas, por cada correspondência entre um ponto de \mathcal{P}^n e a sua imagem de \mathcal{P}^m , formam-se $(m+1)$ equações lineares. No entanto, durante a estimação dos parâmetros independentes da transformação projectiva, o factor multiplicativo, integrado nas $(m+1) \times (n+1)$ incógnitas, é eliminado, ficando-se apenas com m equações linearmente independentes. Esta situação leva a que sejam necessários $(n+1 + \frac{n}{m})$ pares de pontos correspondentes, para a completa definição da transformação \mathbf{H} .

No caso das homografias, é necessário conhecer pelo menos $(n + 2)$ pares de pontos correspondentes para determinar todos os seus parâmetros. Além disso, existindo o conhecimento de mais do que $(n + 2)$ pares de pontos correspondentes, é necessário que haja uma base projectiva de exactamente $(n + 2)$ pontos em cada espaço vectorial. Caso isto não aconteça não se pode estimar a homografia.

De acordo com o objectivo do presente trabalho, centra-se, seguidamente, a atenção no problema da estimação de homografias de grau 2, definidas a menos de um factor multiplicativo, devido à necessidade de relacionar imagens entre si, bem como uma imagem e um plano 3D. Sejam então $\mathbf{p}_e = k_e [x_e \ y_e \ 1]^T$ e $\mathbf{p}_d = k_d [x_d \ y_d \ 1]^T$, os pontos correspondentes em cada espaço vectorial, isto é, na respectiva imagem esquerda e direita. Expandindo a equação (2.37) para este caso ($m = n = 2$) obtém-se

$$k\mathbf{p}_d = \mathbf{H}\mathbf{p}_e \Leftrightarrow k \begin{bmatrix} x_d \\ y_d \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x_e \\ y_e \\ 1 \end{bmatrix} \quad (2.38)$$

ou, de outra forma,

$$\begin{cases} kx_d = h_{11}x_e + h_{12}y_e + h_{13} \\ ky_d = h_{21}x_e + h_{22}y_e + h_{23} \\ k = h_{31}x_e + h_{32}y_e + h_{33} \end{cases}$$

com $k = \frac{k_d}{k_e}$, um factor de escala diferente de zero, que agrega as coordenadas homogéneas de ambos os pontos.

Fazendo a eliminação de k nas duas primeiras equações do sistema anterior e reformulando-as de modo que as incógnitas sejam os coeficientes da homografia \mathbf{H} , obtém-se, na forma matricial, a equação seguinte

$$\begin{bmatrix} x_e & y_e & 1 & 0 & 0 & 0 & -x_e x_d & -x_d y_e - x_d \\ 0 & 0 & 0 & x_e & y_e & 1 & -x_e y_d & -y_e y_d - y_d \end{bmatrix} \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \\ h_{33} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Leftrightarrow \mathbf{A}_i \mathbf{h} = \mathbf{0} \quad (2.39)$$

Confirmando o que foi dito atrás, para cada par de pontos correspondentes são geradas duas equações. Assim, comprova-se a necessidade de pelo menos quatro pares de pontos correspondentes para estimar a homografia, \mathbf{H} . Esta estimação fica concluída quando se encontrar a solução do sistema de equações $\mathbf{A}\mathbf{h} = \mathbf{0}$, sendo a matriz \mathbf{A} constituída pelas sub-matrizes \mathbf{A}_i , da equação (2.39), resultantes do uso dos diferentes pares de pontos correspondentes. Esta maneira de obter a matriz \mathbf{H} é conhecida como estimação não homogénea.

A forma homogénea de estimação da homografia \mathbf{H} parte de

$$\mathbf{p}_d \wedge \mathbf{H}\mathbf{p}_e = 0 \Leftrightarrow [\mathbf{p}_d]_{\wedge} \mathbf{H}\mathbf{p}_e = 0 \quad (2.40)$$

Esta equação define o conhecimento de que o produto externo entre dois vectores iguais é o vector nulo (notar que $\mathbf{p}_d = \mathbf{H}\mathbf{p}_e$). Como as equações (2.39) e (2.40) são equiparáveis, o modo de obtenção da solução é feito da mesma maneira. Devido ao facto das homografias serem definidas a menos de um factor multiplicativo, pode-se omitir uma das equações geradas por (2.40) [Hartley 00].

Como se pode verificar pela equação (2.39), a matriz homográfica de grau 2 tem nove elementos e é definida a menos de um factor de escala. Logo, a matriz \mathbf{A} , deste sistema, deve ter característica oito (solução óptima) ou característica nove.

Quando a matriz \mathbf{A} tiver característica oito, o sistema $\mathbf{A}\mathbf{h} = \mathbf{0}$ tem uma única solução para \mathbf{h} , já definida a menos de um factor multiplicativo. Esta solução, que pode ser obtida por vários métodos numéricos lineares, é dada pelo espaço nulo direito da matriz \mathbf{A} .

Se os dados de entrada para a estimação de \mathbf{h} não forem exactos, devido ao ruído nas coordenadas dos pontos correspondentes, então a característica de \mathbf{A} é nove. Nesta situação terá que se usar o método dos mínimos quadrados para obter a solução. Neste caso, a justificação reside no facto do sistema de equações $\mathbf{A}\mathbf{h} = \mathbf{0}$, neste caso, poder não ter solução, e, assim, garante-se sempre uma. Deixa-se para a secção 2.7, desta dissertação, a explicação do método dos mínimos quadrados.

À semelhança do que acontece com a estimação da matriz fundamental, se forem usados apenas quatro pares de pontos correspondentes, muito provavelmente o resultado da estimação terá grandes erros, devido à inexactidão das posições tridimensionais dos pontos correspondentes escolhidos. É aconselhável, portanto, o uso de mais pares de pontos para se obter uma solução óptima, apesar de se poder cair na situação da matriz \mathbf{A} ter característica nove.

No apêndice B apresenta-se a informação necessária à utilização, no MATLAB®, da

função `mhest.m`, a qual faz a estimação da matriz homográfica para qualquer das situações referidas.

De seguida abordar-se-ão as seguintes questões importantes para a estimação da matriz homográfica:

- normalização
- eliminação de falsas correspondências
- algoritmos de estimação

Normalização

Segundo Hartley [Hartley 98] [Hartley 00], o cálculo da transformação \mathbf{H} deve integrar a operação de normalização aos seus dados de entrada. As razões para a sua aplicação são as mesmas que as apresentadas na secção da estimação da matriz fundamental.

Devido à necessidade de, neste trabalho, estimar homografias entre uma imagem e um plano tridimensional e face à falta de estabilidade nos sistemas lineares, além da normalização do conjunto de dados bidimensionais (como foi visto na secção 2.4.2), deve-se normalizar, também, o conjunto de dados tridimensionais para melhorar o desempenho do algoritmo de estimação, conforme argumenta Hartley. Assim, de acordo com este autor, a análise do problema conduz à aplicação, separadamente para cada conjunto de pontos tridimensionais, das seguintes acções:

- translacionar as coordenadas do ponto correspondente, de forma a que a sua origem fique no centro de massa do conjunto de pontos;
- escalonar os eixos do sistema tridimensional do mundo, de modo a que a distância média à origem seja $\sqrt{3}$. Este escalonamento é conseguido calculando, para todos os pontos 3D, a média euclidiana da distância à origem. Depois, obtém-se o factor de escala com a divisão de $\sqrt{3}$ pela média encontrada. Este processo termina quando todas as coordenadas, do conjunto dos pontos 3D, forem multiplicados pelo factor de escala.

Depois de feita a estimação da homografia, deve ser-lhe aplicada uma desnormalização, tal como acontece com a transformação dos dados bidimensionais, para que qualquer ponto

não normalizado do mundo 3D possa interagir com ela. Assim, para a matriz \mathbf{H} , esta operação difere da aplicada para a matriz fundamental pois é feita através de

$$\mathbf{H} = \mathbf{M}_d^{-1} \mathbf{H}_n \mathbf{M}_e$$

sendo \mathbf{H}_n a homografia obtida com um conjunto de dados normalizados.

Convém mencionar que, devido aos seus resultados na redução do ruído, o processo de normalização/desnormalização aqui descrito deve ser usado em qualquer método onde intervenham dados tridimensionais [Hartley 98].

Uma vez mais, remete-se para o apêndice B a apresentação do modo de utilização, no MATLAB®, da função de normalização de um conjunto de pontos de uma imagem, `htnorm3d.m`.

Eliminação de falsas correspondências

De forma a eliminar as falsas correspondências do conjunto de dados inicial, pode ser aplicado ao cálculo da homografia, à semelhança do que foi explicado para a matriz fundamental, o método do RANSAC (este método será explicado na secção 2.7). A utilização do RANSAC na estimação de homografias de grau 2 integra-se num processo iterativo que começa por escolher, aleatoriamente, quatro pares de pontos correspondentes do conjunto de dados inicial. Com esses pontos calcula-se a homografia, que será a base para a verificação de quão perto estão os pontos do conjunto de dados inicial dos seus correspondentes calculados. Esta verificação, à semelhança da que é feita para a matriz fundamental, usa a qualidade da correspondência da matriz homográfica (ver apêndice A) pois obtém resultados satisfatórios [Zhang 96]. Depois de se fazer esta verificação dá-se início a nova iteração. O resultado final é o que proporciona uma matriz que tem mais pares de pontos a coincidirem com os seus correspondentes calculados, dentro de um nível de erro especificado.

No apêndice B é apresentada a informação necessária ao uso, no MATLAB®, da função `mhransac.m`, a qual estima \mathbf{H} eliminando as falsas correspondências existentes no conjunto de dados de entrada.

Algoritmos de estimação

Salvaguardadas as devidas diferenças, os métodos para a obtenção da solução do sistema $\mathbf{A}\mathbf{h} = \mathbf{0}$, da equação (2.39), são os mesmos que os utilizados na estimação da matriz

fundamental.

Como exemplo desta afirmação veja-se o algoritmo dos oito pontos. A aplicação deste método na estimação da homografia, \mathbf{H} , é linear, mas inspira algum cuidado. Segundo Hartley [Hartley 00], se a origem do sistema de coordenadas de um dos espaços vectoriais tiver um correspondente no infinito, obtido através de uma matriz \mathbf{H} , o elemento h_{33} dessa matriz deve ser zero. Ao ser-lhe atribuído o valor 1, falha-se completamente a estimação de \mathbf{H} . Remete-se, mais uma vez, para o apêndice B, a informação necessária ao modo de funcionamento, no MATLAB©, da função que estima \mathbf{H} através do algoritmo dos oito pontos, mh8pta.m.

2.5.2 Homografias infinitas

Para se introduzir o conceito de homografia infinita ou homografia de um plano no infinito, seja $P = (X, Y, Z)$, um ponto 3D, cujas projecções nas imagens esquerda e direita, são p_e e p_d , respectivamente. As coordenadas homogêneas, quando este ponto está no infinito, são dadas pelo vector $\mathbf{P} = \begin{bmatrix} X & Y & Z & 0 \end{bmatrix}^T$. Considere-se, também, que as duas câmaras, representadas na figura 2.5, têm matrizes de parâmetros intrínsecos \mathbf{C}_e e \mathbf{C}_d e que as suas matrizes de orientação e vectores de posição são \mathbf{R}_e , \mathbf{R}_d , \mathbf{t}_e e \mathbf{t}_d , respectivamente.

Para a imagem esquerda, com base na equação (2.6), tem-se

$$k\mathbf{p}_e = \mathbf{C}_e \begin{bmatrix} \mathbf{R}_e & \mathbf{t}_e \end{bmatrix} \mathbf{P} \Leftrightarrow P = (\mathbf{C}_e \mathbf{R}_e)^{-1} k\mathbf{p}_e \quad (2.41)$$

Substituindo P , numa equação equivalente a (2.41) para a imagem direita, obtém-se

$$k\mathbf{p}_d = \mathbf{C}_d \mathbf{R}_d (\mathbf{C}_e \mathbf{R}_e)^{-1} k\mathbf{p}_e \quad (2.42)$$

Eliminando o factor de escala da equação (2.42) conclui-se que a homografia infinita é dada por

$$\mathbf{H}_\infty = \mathbf{C}_d \mathbf{R}_d \mathbf{R}_e^{-1} \mathbf{C}_e^{-1} = \mathbf{C}_d \mathbf{R} \mathbf{C}_e^{-1} \quad (2.43)$$

onde \mathbf{R} é a matriz de rotação entre as câmaras, representada na figura 2.5.

Esta homografia é a mesma que se obtém no caso da câmara efectuar uma rotação pura em torno do seu centro de projecção. Quando a diferença entre as posições da câmara está só na rotação, todos os planos têm a mesma homografia, tirando os que contêm o centro óptico, para os quais ela não está definida. No caso de uma câmara realizar uma translação

pura, mantendo os seus parâmetros intrínsecos constantes, esta homografia passa a ser dada pela matriz identidade.

As razões para a apresentação do conceito das homografias infinitas são duas. A primeira é que a homografia infinita tem um papel importante nas reconstruções e calibrações afins. A segunda, prende-se com o facto do uso desta matriz reduzir o espaço de procura de pontos correspondentes, sobre as rectas epipolares, quando se usa a matriz fundamental [Hartley 00].

2.5.3 Relações entre a matriz fundamental e as homografias

Para se obter a relação entre a matriz fundamental \mathbf{F} e a homografia \mathbf{H} , começa-se por considerar p_d , um ponto genérico da imagem direita, e_d , o epipolo dessa mesma imagem e q , um ponto qualquer pertencente à recta epipolar que passa por p_d . Esta recta epipolar é dada por

$$\mathbf{p}_d^T [\mathbf{e}_d]_{\wedge}^T \mathbf{q} = 0 \Leftrightarrow (\mathbf{e}_d \wedge \mathbf{p}_d)^T \mathbf{q} = 0 \quad (2.44)$$

Como q pode ser um ponto qualquer da recta epipolar, então seja

$$\mathbf{q} = \mathbf{p}_d = \mathbf{H}\mathbf{p}_e \quad (2.45)$$

Assim, da substituição da equação (2.45) em (2.44) obtém-se

$$\mathbf{p}_d^T [\mathbf{e}_d]_{\wedge}^T \mathbf{H}\mathbf{p}_e = 0 \Leftrightarrow \mathbf{p}_d^T \mathbf{F}\mathbf{p}_e = 0$$

com

$$\mathbf{F} = [\mathbf{e}_d]_{\wedge}^T \mathbf{H} = -[\mathbf{e}_d]_{\wedge} \mathbf{H} \quad (2.46)$$

Uma outra maneira de se obter esta mesma relação passa pela definição da homografia \mathbf{H} e das rectas epipolares da imagem direita que passam pelo epipolo e_d , dadas por $r_e = \mathbf{F}\mathbf{p}_e$. Ambos os pontos p_e e p_d são a projecção, nas imagens, de um ponto genérico 3D, sendo portanto correspondentes. Assim, p_d deve pertencer à recta epipolar de p_e . Utilizando estes conhecimentos obtém-se

$$r_e = \overline{e_d p_d} = \mathbf{F}\mathbf{p}_e \Leftrightarrow [\mathbf{e}_d]_{\wedge} \mathbf{p}_d = \mathbf{F}\mathbf{p}_e \Leftrightarrow [\mathbf{e}_d]_{\wedge} \mathbf{H}\mathbf{p}_e = \mathbf{F}\mathbf{p}_e$$

de onde se chega a

$$\mathbf{F} = [\mathbf{e}_d]_{\wedge} \mathbf{H} \quad (2.47)$$

À primeira vista, parece que os resultados apresentados nas equações (2.46) e (2.47) não são coerentes, mas isso não é verdade. Como a matriz fundamental é definida a menos de um factor multiplicativo, então a relação geral entre \mathbf{F} e \mathbf{H} é

$$\mathbf{F} = k[\mathbf{e}_d]_{\wedge} \mathbf{H} \quad (2.48)$$

com k um qualquer factor multiplicativo.

Todo este raciocínio pode ser feito com o uso do epipolo da imagem esquerda, e_e , onde esta relação é dada por

$$\mathbf{F} = k\mathbf{H}^{-T}[\mathbf{e}_e]_{\wedge} \quad (2.49)$$

com k um qualquer factor de escala.

Conhecendo agora estas relações, pode-se dizer, com ajuda das equações (2.23) e (2.24), que a homografia \mathbf{H} é dada por

$$\mathbf{H} = k\mathbf{C}_d\mathbf{R}\mathbf{C}_e^{-1}$$

onde k é um qualquer factor de multiplicativo. No entanto, esta não é a sua única expressão. Da maneira como foram deduzidas as equações (2.48) e (2.49), constata-se que a matriz \mathbf{H} pode ser dada por qualquer homografia existente entre as imagens [Hartley 00]. A razão para tal, encontra-se no facto de bastarem quaisquer quatro projecções de pontos 3D coplanares para a definirem.

A equação

$$\mathbf{F}^T \mathbf{H} + \mathbf{H}^T \mathbf{F} = \mathbf{0}$$

estabelece uma outra relação entre as homografias e a matriz fundamental. Além disso, pode servir para verificar a coerência de uma homografia estimada com a matriz fundamental e vice-versa. A prova desta equação baseia-se nas equações (2.48), (2.49) e na propriedade da matriz antisimétrica, na qual $[\mathbf{v}]_{\wedge}^T = -[\mathbf{v}]_{\wedge}$, para qualquer vector \mathbf{v} .

2.6 Superfícies Reflectoras

Recorde-se que, para determinar a informação 3D a partir de imagens, explorando a disparidade entre os pontos projectados, são necessários vários pontos de vista (dois no mínimo). Além disso, se a cena possui elementos dinâmicos, as imagens devem ser captadas em simultâneo. No entanto, a maioria dos sistemas de aquisição de imagens vídeo de hoje,

não permite trabalhar com entradas simultâneas e sincronizadas. Além destes factos, e tendo em mente que a introdução de processos no meio industrial fica mais facilitada se os seus custos forem reduzidos, tornou-se imperativo, neste trabalho, passar pela introdução de superfícies reflectoras em processos de determinação de informação 3D.

De forma a não cair na generalidade, este trabalho concentra-se somente nos espelhos planares. Existem três grandes categorias de espelhos planares. O vulgar espelho caseiro, conhecido pelo nome técnico de espelho de metal, devido a ter na parte detrás da superfície de vidro prateada, uma malha metálica. É esta malha, protegida por uma camada de tinta, que faz com que os raios sejam reflectidos. Além deste tipo, existem ainda os espelhos dieléctricos, os líquidos e os separadores de raios. Este trabalho irá usar somente os espelhos de metal.

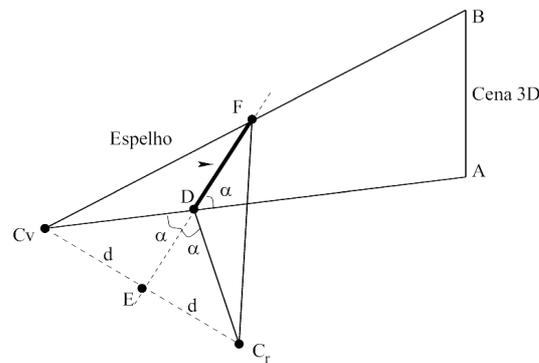


Figura 2.7: *Modo de funcionamento dos espelhos. O $\triangle ABC_v$ é o maior corte lateral do volume de cobertura da câmara. C_v é o centro óptico da câmara virtual equivalente ao centro óptico C , da câmara real, que resulta da introdução do espelho. C_v fica à mesma distância, também medida na perpendicular, a que C_r está do espelho, do seu lado oposto.*

Antes de passar aos sistemas de aquisição de imagens que recorrem a espelhos, passemos em revista os princípios que os regem. O potencial dos espelhos planares assenta na sua capacidade em redireccionar os feixes de luz sem distorcer os ângulos a que eles chegam aos espelhos. Isto é, quando um feixe de luz encontra um espelho, a um ângulo α , é reflectido com um ângulo $(180 - \alpha)$. Este efeito de redireccionamento, chamado de dobra do feixe de luz, pode ser melhor compreendido com o auxílio da figura 2.7.

Relativamente a esta figura, há que salientar alguns aspectos. O primeiro está na redução do espaço que vai desde a câmara até à cena 3D com a introdução do espelho. Esta redução

ocorre sempre que o número de dobras do feixe de luz aumenta. No entanto, é necessária muita atenção a este número pois ele condiciona a intensidade dos feixes. A operação de redução de espaço também pode ser feita se alargarmos o tamanho do espelho e o movermos mais para perto da cena 3D a captar.

Um outro aspecto a salientar da figura 2.7, tem por base o triângulo aí representado ($\triangle ABC_v$). Este triângulo é sempre o maior corte lateral do volume de cobertura da câmara, dado estar-se sempre a lidar tridimensionalmente com um cone ou um prisma rectangular de luz. A dobra feita a este triângulo leva a câmara a mudar de posição, para conseguir ver a mesma cena do mundo 3D. Nesta situação, à representação da câmara que fica atrás do espelho e que tem as mesmas características ópticas da câmara real (simbolizada por C_r), chama-se câmara virtual. O centro óptico da câmara virtual fica à mesma distância, também medida na perpendicular, a que o centro óptico real está do espelho, do seu lado oposto. A explicação geométrica deste facto é dada de seguida.

Por definição $\angle ADF = \angle C_r DE = \alpha$. Como o $\triangle ABC_v$ é o mesmo que o $\triangle ABC_r$, com a diferença da sua ponta estar dobrada, então $\overline{C_r D} = \overline{C_v D}$. Por outro lado, sendo $\angle C_v DE = \angle ADF = \alpha$, pois são simetricamente opostos, e $\overline{EF} \perp \overline{C_v C_r}$, então

$$\overline{C_r E} = \overline{C_v E} = d$$

Para mais pormenores veja-se [Baker 98].

Quando se fala de espelhos, tem-se, obrigatoriamente, que interpretar a imagem no espelho, ou por outras palavras, compreender a simetria entre a imagem no espelho e a cena 3D vista. De forma a uma melhor explicação, considere-se a figura 2.8. Nesta figura mostra-se que a direcção normal à superfície do espelho da cena 3D é invertida quando vista no espelho. A imagem no espelho depende da localização do espelho relativamente à cena do mundo 3D. Esta inversão, conhecida por inversão frente-para-trás, leva a que algo que esteja mais perto do espelho, e portanto mais longe do ponto de vista real, fique na imagem no espelho, olhada daquele ponto de vista, numa posição mais perto desse mesmo ponto de vista.

2.6.1 Zona de cobertura dos sistemas de aquisição

Revistos os princípios gerais que regem os espelhos, passa-se a tratar das questões ligadas aos sistemas de aquisição de imagens, compostos por um conjunto de espelhos e uma



Figura 2.8: *Simetria criada pelo espelho. Um espelho inverte a direcção normal à superfície do espelho da cena 3D. Dependendo do posicionamento do espelho, a inversão pode ser feita da direita para a esquerda (parte esquerda da figura) ou do topo para o fundo (parte direita da figura). Quando ambas as inversões são feitas ao mesmo tempo obtém-se o mostrado na parte central da figura.*

única câmara. Uma vez mais, esta opção foi tomada para que a dissertação não caísse na generalidade e não se perdesse nas imensas possibilidades. O objectivo destes sistemas é o de adquirir várias imagens de uma mesma cena do mundo 3D. Para que isso aconteça, tem que se conseguir uma geometria com os espelhos, tal que seja simulada, a partir da única câmara existente, a presença de várias câmaras virtuais. Para que as imagens sejam da mesma zona da cena, cada câmara virtual tem que possuir uma zona em comum, à qual se chamará de zona de cobertura. Para uma melhor percepção deste conceito, assumindo sempre o uso de câmaras de CCD rectangular, veja-se a figura 2.9. Ao interior da figura geométrica, produzida pelos feixes de luz limítrofes de entrada no CCD, chama-se volume de cobertura e a intersecção dessa figura geométrica com os diferentes planos intervenientes designa-se por área de cobertura nesses planos. Esta questão merece ser estudada pois como há simulação de câmaras através do uso dos espelhos, torna-se necessário saber qual é a zona comum a essas câmaras para que a cena 3D pretendida seja lá enquadrada.

Quando se tem apenas uma câmara CCD, o volume de cobertura é dado pelo interior de um prisma rectangular (ver figura 2.9). Na simulação de duas câmaras, por acção dos espelhos, o volume de cobertura é obtido através da intersecção de dois prismas rectangulares, onde o resultado, conforme o posicionamento dos espelhos, é uma figura geométrica mais ou menos complexa.

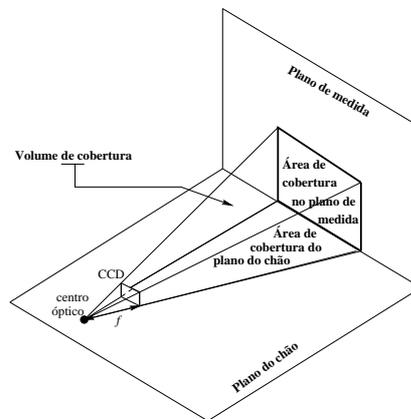


Figura 2.9: Zona de cobertura. Zona da cena 3D que a câmara consegue captar. Ao interior da figura geométrica, produzida pelos feixes de luz limítrofes de entrada no CCD, chama-se volume de cobertura. As intersecções dessa figura geométrica com os planos de medida e do chão designam-se por área de cobertura desses respectivos planos.

A área de cobertura, segundo um dado plano, é obtida pela intersecção da referida figura geométrica com esse plano. No entanto, de todos os planos possíveis, os que verdadeiramente interessam são os que estão representados na figura 2.9, designadamente o plano onde assentam todos os espelhos (plano do chão) e o plano paralelo aos espelhos e perpendicular ao plano do chão (plano de medida). A área no plano do chão informa-nos sobre os limites onde a cena 3D, a captar, se deve situar. Por seu lado, a área de cobertura no plano de medida relata até que altura, em relação ao plano do chão, a câmara consegue captar. Como as áreas de cobertura nos diferentes planos dependem dos espelhos, faremos só o estudo para planos rectangulares e para a área de cobertura no plano do chão. A razão deve-se ao facto dos cálculos serem semelhantes para ambas as áreas (podendo diferir apenas com o tamanho dos espelhos).

Entre as possibilidades de se organizar este estudo, optou-se por agrupar as geometrias dos sistemas deste tipo, em termos de resultados de imagem. Serão apresentados somente exemplos de resultados estereoscópicos, pois o modo de aproximação para outras situações é exactamente o mesmo. Assim, temos como primeiro grupo aquele cujo resultado são duas imagens da mesma cena a uma notável diferença de distância de captação, enquanto que no segundo grupo aparecem as imagens onde essa diferença não é significativa. A distância de captação prende-se com a distância que os raios de luz percorrem até chegar à câmara.

Obviamente, os que andam mais, produzem imagens mais distantes.

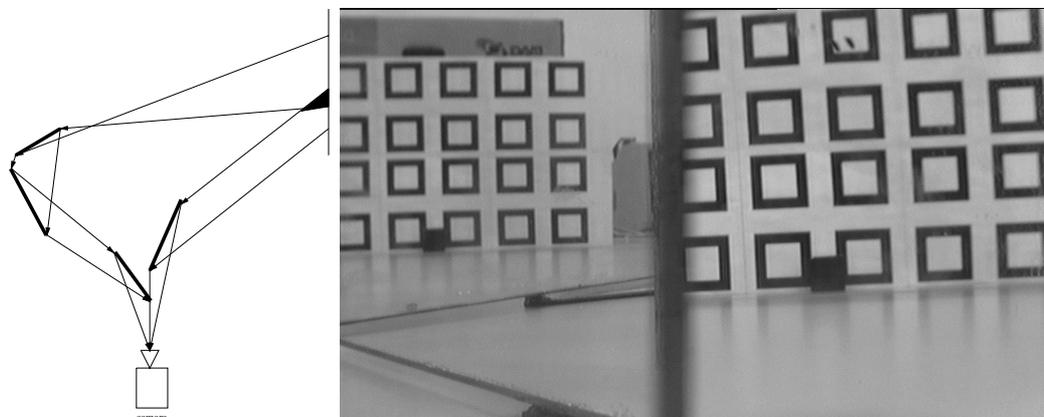


Figura 2.10: *Vista perpendicular, em relação ao plano do chão, de uma configuração exemplificativa, a qual resulta de notáveis diferenças de distância de captação e imagem resultante desta configuração. A zona pintada, no esquema, é a área de cobertura segundo esse plano.*

Tendo sempre como linha de pensamento o processo da reconstrução, e sabendo que se irá recorrer à correspondência entre as imagens, o primeiro grupo não tem interesse para processo, como se pode constatar pelo exemplo mostrado na figura 2.10.

O segundo grupo ainda pode ser subdividido, segundo o posicionamento dos espelhos, entre aqueles que têm uma geometria simétrica a um eixo vertical ou horizontal e aqueles onde não existe essa simetria. O primeiro subgrupo será mencionado como grupo simétrico, enquanto que o segundo será designado por grupo não simétrico. Nestes dois grupos existem, obviamente, muitas configurações específicas. No entanto, serão apresentados apenas dois exemplos que, além de serem suficientes em termos de explicação de como se executa a definição de uma dada configuração, serão parte integrante do trabalho desenvolvido.

O tamanho do CCD da câmara está relacionado com os ângulos máximos de entrada dos feixes na câmara. Assim, para os cálculos dos dois exemplos seguintes, serão considerados os ângulos θ_e e θ_d , medidos em relação ao feixe central de entrada de luz na câmara e aos limites de entrada esquerdo e direito, respectivamente, de feixes de luz impostos pela forma do CCD. Para estes dois casos, a medida A será considerada como a área de cobertura no plano do chão.

O grupo simétrico é exemplificado com a figura 2.11. Esta configuração é semelhante à divulgada pela Inria [Mathieu 95], diferindo no facto de $a \neq c$. A razão para esta pequena

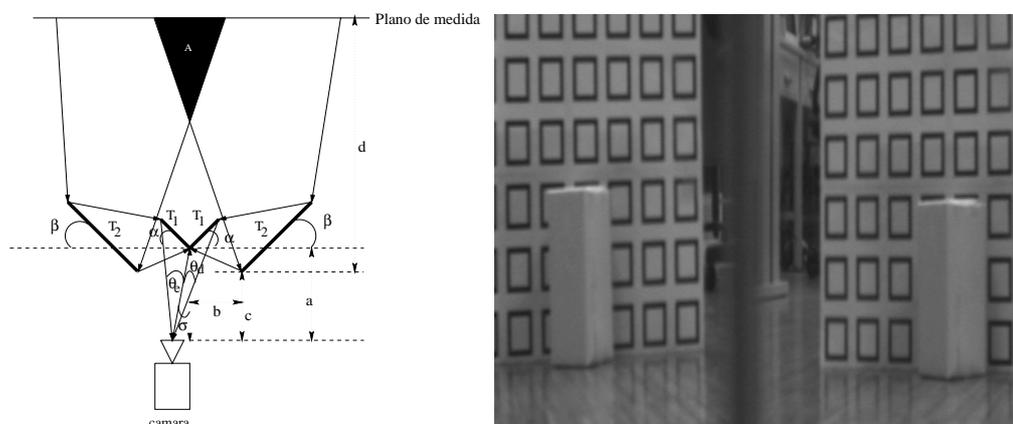


Figura 2.11: *Geometria simétrica. Vista perpendicular ao plano do chão da configuração simétrica, escolhida para os espelhos e para a câmara, e imagem resultante do uso desta geometria. θ_e e θ_d transmitem a projecção do tamanho do CCD no plano do chão. T_1 e T_2 são as medidas das larguras dos espelhos pequenos e grandes, respectivamente. A zona pintada, A , é a área de cobertura no referido plano.*

mudança deve-se à tentativa de generalização desta configuração, através da possibilidade da diminuição do espaço ocupado pela câmara e pelo conjunto de espelhos.

O primeiro destaque na notação apresentada na figura 2.11 vai para o ângulo σ , que é o ângulo que o feixe central de entrada de luz na câmara faz com o eixo de simetria (cuja direcção é segundo o eixo das ordenadas do plano do chão). Torna-se evidente que o valor de σ não pode ser tal que a câmara apareça na imagem captada. T_1 e T_2 são as medidas das larguras dos espelhos pequenos e grandes, respectivamente. α e β são os ângulos que os espelhos T_1 e T_2 fazem, respectivamente, com o eixo das abcissas do plano do chão. Por fim, a é a distância medida sobre o eixo de simetria entre a câmara e a junção dos espelhos T_1 , enquanto que c é a mesma medida em relação aos espelhos T_2 . O valor de b é a distância, medida sobre a perpendicular ao eixo de simetria, entre os espelhos T_1 e T_2 . A distância d é a que fica entre o extremo, mais próximo da câmara, do espelho T_2 e do plano de medida. Torna-se obvio que, nesta configuração, os ângulos α e β têm que ser ajustados de maneira a se conseguir uma área de cobertura com a mesma forma da que foi apresentada na figura 2.11. Assim, o valor do ângulo α tem que ser tal que os feixes de luz não se percam entre os espelhos, enquanto que o ângulo β tem que ser sempre superior ao ângulo α .

Da configuração apresentada na figura 2.11, surgem, devido ao modo de funcionamento

dos espelhos, algumas relações que ajudam no dimensionamento de um sistema que use essa configuração. Quando a câmara estiver posicionada do lado esquerdo do eixo de simetria, teremos

$$T_1 = \frac{a \sin(\theta_e)}{\cos(\sigma) \cos(\alpha + \theta_e - \sigma)} = \frac{a \sin(\theta_d)}{\cos(\sigma) \cos(\alpha + \theta_d + \sigma)}$$

$$T_2 = \frac{(T_1 \sin \alpha + a - c) \tan(2\alpha + \theta_e - \sigma) + T_1 \cos \alpha - b}{\sin \beta \tan(2\alpha + \theta_e - \sigma) + \cos \beta} = \frac{(T_1 \sin \alpha + a - c) \tan(2\alpha + \theta_d + \sigma) + T_1 \cos \alpha - b}{\sin \beta \tan(2\alpha + \theta_d + \sigma) + \cos \beta}$$

Por outro lado, se a câmara estiver posicionada do lado direito do eixo de simetria, teremos

$$T_1 = \frac{a \sin(\theta_e)}{\cos(\sigma) \cos(\alpha + \theta_e + \sigma)} = \frac{a \sin(\theta_d)}{\cos(\sigma) \cos(\alpha + \theta_d - \sigma)}$$

$$T_2 = \frac{(T_1 \sin \alpha + a - c) \tan(2\alpha + \theta_e + \sigma) + T_1 \cos \alpha - b}{\sin \beta \tan(2\alpha + \theta_e + \sigma) + \cos \beta} = \frac{(T_1 \sin \alpha + a - c) \tan(2\alpha + \theta_d - \sigma) + T_1 \cos \alpha - b}{\sin \beta \tan(2\alpha + \theta_d - \sigma) + \cos \beta}$$

Quanto à área de cobertura, A , desta geometria, ela é dada por

$$A = \frac{d \tan[2\beta - (2\alpha \pm \sigma)] - b^2}{2 \tan[2\beta - (2\alpha \pm \sigma)]} + \frac{d \tan[2\beta - (2\alpha \mp \sigma)] - b^2}{2 \tan[2\beta - (2\alpha \mp \sigma)]}$$

onde o sinal ligado a σ depende do lado em que a câmara está em relação ao eixo de simetria. O sinal que aparece por cima acontece na situação em que a câmara está no lado esquerdo do eixo de simetria, enquanto que o sinal que aparece em baixo refere-se à situação da câmara posicionada no lado direito desse eixo.

Um sistema de aquisição de imagens baseado nesta configuração poderá tornar-se muito útil num processo ligado à visão computacional devido à sua simetria, tal como são os casos dos trabalhos [Goshtasby 93] e [Mathieu 95].

Como exemplo do grupo não simétrico tem-se o apresentado na figura 2.12. A notação usada sofre algumas alterações em relação à anterior. T_e e T_d são as medidas dos dois espelhos usados nesta configuração, enquanto que os ângulos α e β são os ângulos que os espelhos T_e e T_d fazem com o eixo paralelo ao plano de medida. Por seu lado, as medidas a e b são, respectivamente, os catetos do triângulo retângulo, cuja hipotenusa é o feixe de entrada de luz que bate no extremo do espelho T_e mais próximo da câmara. Por fim, d é a distância entre o extremo do espelho T_e , que mais perto fica da câmara, e o plano de medida.

Neste exemplo surgem equações que se podem utilizar na construção de um sistema baseado na configuração apresentada na figura 2.12. Através de um cálculo trigonométrico

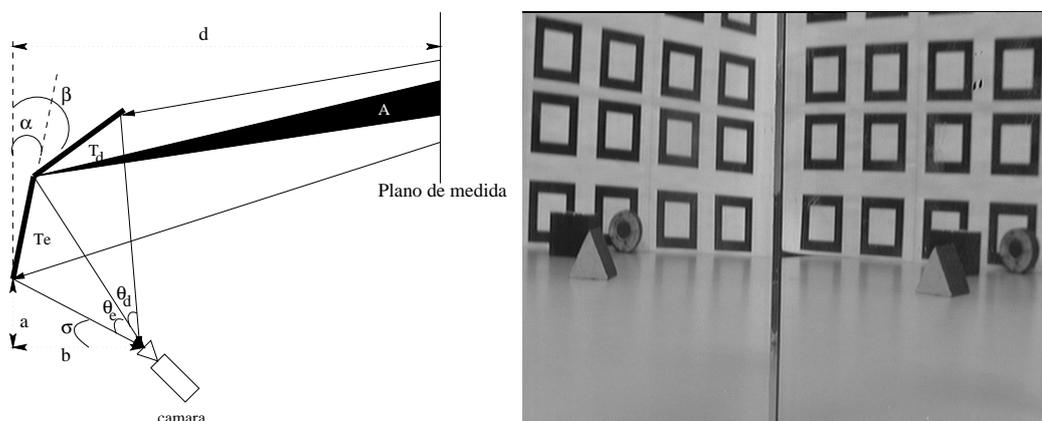


Figura 2.12: *Geometria não simétrica. Vista perpendicular ao plano do chão da configuração não simétrica, escolhida para os espelhos e para a câmara, e imagem resultante do uso desta geometria. θ_e e θ_d transmitem o tamanho da projecção do CCD no plano do chão. T_e e T_d são as medidas das larguras dos espelhos. A zona pintada, A , é a área de cobertura no plano do chão.*

simples, as equações são

$$T_e = \frac{\sqrt{a^2 + b^2} \sin \theta_e}{\cos(\sigma - \alpha + \theta_e)} \quad T_d = \frac{\sqrt{(a + T_e \cos \alpha)^2 + (b - T_e \sin \alpha)^2} \sin \theta_d}{\cos(\sigma - \beta + \theta_e + \theta_d)}$$

onde

$$\sigma = \arctan\left(\frac{a}{b}\right)$$

A área de cobertura, A , para esta geometria é dada por

$$A = \frac{(d - T_e \sin \alpha)^2}{2} |\tan(\sigma + \theta_e - 2\alpha) - \tan(\sigma + \theta_e + 2\beta)|$$

Como se pode verificar, nas figuras 2.10, 2.11 e 2.12 aparece uma falha. Essa falha deve-se à quebra existente entre os espelhos, podendo ser diminuída, embora nunca eliminada, com o aumento da distância da câmara aos espelhos ou mudando as características dessa câmara.

2.6.2 A geometria epipolar e os espelhos

Para se perceber melhor a relação existente entre os espelhos e a geometria epipolar, considere-se a figura 2.13. Esta figura mostra uma configuração genérica de dois espelhos

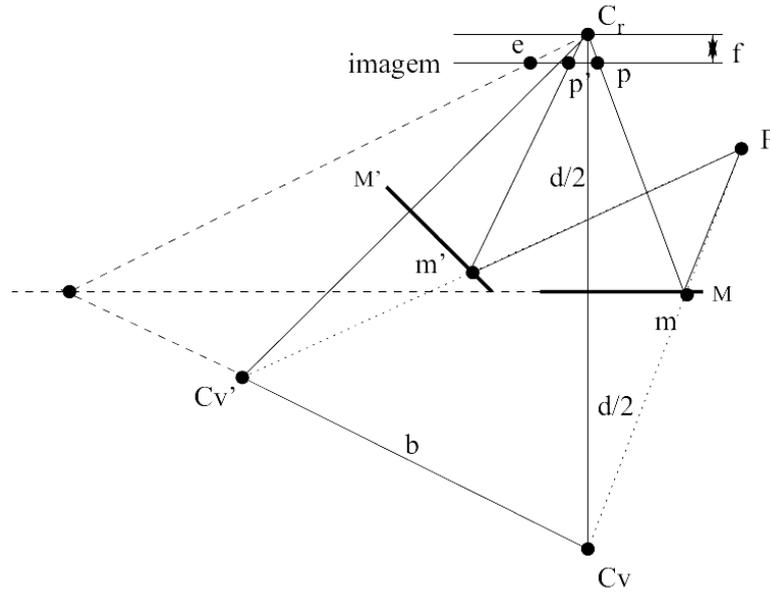


Figura 2.13: *Obtenção de informação epipolar através dos espelhos. M e M' são dois espelhos planares. Os pontos C_v e $C_{v'}$ são os centros ópticos virtuais, enquanto que C_r é o centro óptico real.*

planares M e M' e uma câmara definida pelo seu centro óptico C_r , a uma distância $\frac{d}{2}$ do espelho M . É possível observar-se o plano imagem a uma distância f de C_r . Os pontos C_v e $C_{v'}$ são os centros ópticos virtuais, provocados respectivamente pelos espelhos M e M' . A distância entre estes pontos e os respectivos espelhos é igual à distância entre o ponto C_r e esses espelhos. Considera-se que a origem do sistema de coordenadas está situada no ponto C_v .

Sendo P um ponto da cena 3D, as suas projecções no plano imagem são os pontos p e p' . Os raios correspondentes a estes pontos intersectam os espelhos nos pontos 3D m e m' , respectivamente.

Para o ponto genérico $p' = (x', y')$, na imagem, a expressão da sua recta epipolar correspondente é dada pela projecção no plano imagem, da recta que resulta da intersecção do plano que passa por C_v , $C_{v'}$ e m' com o espelho M . Sabendo que $C_{v'} = (v'_x, v'_y, v'_z)$ e que $\overline{m'C_{v'}} = (d_x, d_y, d_z)$, a equação do plano epipolar é expressa por

$$AX + BY + CZ = 0 \quad (2.50)$$

com

$$A = d_z v'_y - d_y v'_z \quad B = d_x v'_z - d_z v'_x \quad C = d_y v'_x - d_x v'_y$$

A equação do espelho planar M , é

$$Z = \frac{d}{2} \quad (2.51)$$

Substituindo Z da equação (2.51) em (2.50), fica-se com a intersecção do plano epipolar com o espelho, dada por

$$2AX + 2BY + Cd = 0 \quad (2.52)$$

A projecção em perspectiva de um qualquer ponto $m = (X_m, Y_m, Z_m)$ desta linha no plano imagem, de forma a obter $p = (x, y)$, é conseguida através de

$$x = \frac{fX_m}{d + \frac{AX_m + BY_m}{C}} \quad y = \frac{fY_m}{d + \frac{AX_m + BY_m}{C}} \quad (2.53)$$

A equação da recta epipolar resulta da substituição da equação (2.53) em (2.52). Estas rectas correspondem às projecções das linhas 3D, resultantes da intersecção da família de planos que passa pela baseline virtual (linha que une os dois centros ópticos virtuais), b , com o espelho M .

É facilmente visto que estas linhas convergem no ponto de intersecção da recta b com o espelho M . Da projecção deste ponto no plano imagem resulta o epipolo $e = (e_x, e_y)$. A dedução da expressão do epipolo começa por assumir que o vector direcção da recta b é $\vec{b} = (\lambda, \mu, \nu)$. Assim, a recta b fica definida por

$$\frac{X}{\lambda} = \frac{Y}{\mu} = \frac{Z}{\nu}$$

Projectando na imagem a intersecção da recta b com o espelho M , os resultados imediatos são as coordenadas do epipolo

$$e_x = \frac{\lambda f}{\nu} \quad e_y = \frac{\mu f}{\nu}$$

Todo o raciocínio aqui desenvolvido pode ser feito de forma a encontrar as expressões para as rectas epipolares que correspondem às projecções das linhas 3D, resultantes da intersecção da família de planos que passa pela baseline virtual, b , com o espelho M' e para o seu respectivo epipolo. Para mais pormenores veja-se [Sameer 98].

2.7 Métodos numéricos

Os métodos que aqui se pretende apresentar estão relacionados com a obtenção de soluções para quaisquer sistemas lineares, do tipo $\mathbf{Ax} = \mathbf{b}$, usando apenas critérios lineares. A vantagem destes critérios é a da não iteratividade dos métodos que a utilizam, enquanto que a sua desvantagem é a da criação de sensibilidade ao ruído, independentemente da dimensão do conjunto de dados de entrada. Esta desvantagem deve-se, principalmente, ao não condicionamento dos dados de entrada e de saída. Existem, no entanto, vários métodos numéricos que tentam de alguma forma controlar essa sensibilidade, como veremos mais à frente com a apresentação de um desses métodos.

O problema de resolução de sistemas lineares, de forma genérica $\mathbf{Ax} = \mathbf{b}$, aparece num vasto leque de aplicações. Como acontece em todos os sistemas de equações, nos sistemas lineares o número de soluções pode variar entre nenhuma, uma, e mais do que uma soluções. Assim, quando as linhas de \mathbf{A} são linearmente independentes, os sistemas lineares têm, normalmente, uma única solução, enquanto que, quando as linhas são linearmente dependentes, têm, normalmente, infinitas soluções. Porém, em ambas as situações estes sistemas poderão não ter solução. Pode-se, então, definir dois tipos de sistemas lineares quanto ao seu número de soluções, os sistemas não singulares e os sistemas singulares.

Os sistemas não singulares são aqueles que têm uma única solução, e onde a matriz \mathbf{A} é quadrada e invertível. Estes sistemas podem ser solucionados por métodos directos, que aplicam um número finito de passos até obter a solução, ou por métodos iterativos, que encontram a solução por aproximações. Como exemplos dos primeiros temos a eliminação de Gauss, a regra de Cramer e a factorização QR [Datta 94][Recipes 92]. Por seu lado, os métodos de Jacobi, de Gauss-Seidel e do gradiente conjugado [Datta 94] [Recipes 92] são exemplos de métodos iterativos. Uma vez que os sistemas não singulares raramente aparecem na determinação de informação 3D, não lhe dedicaremos mais tempo.

Por sua vez, os sistemas singulares são aqueles que têm mais do que uma ou nenhuma soluções. Existem várias soluções quando o número de equações é maior do que o número de incógnitas. Esta situação geralmente não interessa, e pode ser ultrapassada acrescentando equações de zeros até se obter tantas equações quanto o número de incógnitas [Hartley 00]. Por outro lado, quando existirem mais equações do que incógnitas, só se obtêm soluções se o vector \mathbf{b} pertencer ao espaço vectorial das colunas da matriz \mathbf{A} . Também se enquadram neste grupo os sistemas cuja matriz \mathbf{A} é quadrada, mas não invertível.

Os sistemas singulares estão, muitas vezes, agregados a processos que necessitam sempre de uma solução. Nestes casos, quando a solução não existe, continua a fazer sentido procurar uma para a atribuir ao processo. Um exemplo comum desta situação surge quando o processo introduz erros nos dados que servem à criação do sistema de equações. A forma de resolução linear de toda a problemática que envolve os sistemas singulares é feita através da aplicação do método dos mínimos quadrados, que se expõe de seguida.

2.7.1 Método dos mínimos quadrados

Nos sistemas singulares, o melhor procedimento para garantir sempre uma solução, é aproximar o mais possível o vector \mathbf{Ax} do vector \mathbf{b} . Por outras palavras, procura-se encontrar a solução, dada pelo vector \mathbf{x} , tal que a norma do vector $\mathbf{Ax} - \mathbf{b}$ seja a menor possível. Quando se está a trabalhar no espaço euclidiano, esta solução é referida como a solução dos mínimos quadrados, aplicando-se a norma euclidiana. Assim, o problema da obtenção da solução, \mathbf{x} , pelo método dos mínimos quadrados resume-se a

$$\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|^2 \quad (2.54)$$

Por definição, o quadrado da norma euclidiana de um vector é a soma dos quadrados dos componentes do vector [Datta 94].

Antes de avançar, é importante fazer algumas referências a um caso particular dos sistemas lineares, os sistemas homogêneos, cuja forma geral é $\mathbf{Ax} = \mathbf{0}$. Estes sistemas lineares têm sempre uma solução trivial, $\mathbf{x} = \mathbf{0}$, embora também possam ter outras soluções [Datta 94]. Para as encontrar, normalmente são impostas restrições de forma a eliminar a solução trivial. Um exemplo desta metodologia é o algoritmo de oito pontos aplicado à estimação das matrizes fundamental e homográfica de grau 2. É na descoberta das outras soluções que estes sistemas (agregados com a respectiva imposição das restrições) também recorrem à solução dos mínimos quadrados. Neste caso, a equação (2.54) passa a ser dada por

$$\min_{\mathbf{x}} \|\mathbf{Ax}\|^2 \quad s. a \ (\dots) \quad (2.55)$$

Esta situação aparece muito frequentemente nos problemas de reconstrução.

Tendo em conta os problemas com que esta dissertação tem que lidar, limita-se o estudo do método dos mínimos quadrados a situações em que o número de linhas de \mathbf{A} é maior ou igual ao número de colunas. Para estes casos existe sempre uma solução para qualquer

sistema de equações [Datta 94]. Essa solução é única se e só se, a característica de \mathbf{A} for igual ao seu número de colunas. Se o número de colunas for maior que a característica de \mathbf{A} , então existem múltiplas soluções. De seguida, serão discutidos alguns dos métodos computacionais para resolver estes casos.

A decomposição em valores singulares

Ao problema de se encontrar as matrizes ortogonais \mathbf{U} e \mathbf{V} ($\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I}$, com \mathbf{I} a matriz identidade) e a matriz diagonal $\mathbf{S} = \text{diag}(\tau_1, \tau_2, \dots, \tau_r, 0, \dots, 0)$ de uma matriz \mathbf{A} , $m \times n$, tal que

$$\mathbf{A} = \mathbf{USV}^T \quad (2.56)$$

chama-se decomposição em valores singulares da matriz \mathbf{A} [Datta 94][Recipes 92]. Esta decomposição é mais conhecida pelas iniciais SVD, que advêm do seu nome em inglês, *single value decomposition*. Os elementos τ_i , da matriz \mathbf{S} chamam-se valores singulares, enquanto que as colunas de \mathbf{U} e \mathbf{V} são designadas por vectores singulares. Quando m for igual a n , $\mathbf{UU}^T = \mathbf{VV}^T = \mathbf{I}$. Quando se usa o SVD, assume-se, sem perda de generalidade, que na matriz \mathbf{A} , $m \geq n$, pois quando $m < n$, a matriz a considerar deverá ser \mathbf{A}^T .

Muitas áreas da engenharia, de que é exemplo a da visão por computador, estão a dar cada vez mais importância a este método de decomposição, cujas aplicações típicas são as de calcular uma característica, uma base ortonormada, projecções, distâncias entre matrizes, etc, na presença de ruído nos dados de entrada. Os valores e vectores singulares são as ferramentas numéricas mais fiáveis para este tipo de cálculos. Segundo Data [Datta 94], o SVD é a aproximação numérica mais efectiva para resolver o problema nos mínimos quadrados, especialmente os casos de deficiências na característica.

A partir de $\mathbf{Ax} = \mathbf{b}$, usando o SVD, o problema dos mínimos quadrados é resolvido da seguinte maneira. Substituindo a equação (2.56) em (2.54), obtém-se

$$\min_{\mathbf{x}} \|\mathbf{USV}^T\mathbf{x} - \mathbf{b}\|^2 = \min_{\mathbf{x}} \|\mathbf{U}(\mathbf{SV}^T\mathbf{x} - \mathbf{U}^T\mathbf{b})\|^2$$

Como é o vector \mathbf{x} que está a ser minimizado, a matriz \mathbf{U} pode desaparecer da equação anterior, dando origem a

$$\min_{\mathbf{y}} \|\mathbf{Sy} - \mathbf{b}'\|^2$$

onde $\mathbf{y} = \mathbf{V}^T\mathbf{x}$ e $\mathbf{b}' = \mathbf{U}^T\mathbf{b}$. Com esta simplificação, o problema passa a ser de solução trivial, devido ao facto da matriz \mathbf{S} ser diagonal. Assim, o vector $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \dots & y_n \end{bmatrix}^T$

é dado por

$$y_i = \begin{cases} \frac{b'_i}{\tau_i} & \text{se } \tau_i \neq 0 \\ 0 & \text{se } \tau_i = 0 \end{cases}$$

A solução para o sistema, obtida por

$$\mathbf{x} = \mathbf{V}\mathbf{y}$$

é única quando $m = n$, ou, é a solução óptima que minimiza o vector \mathbf{x} , quando $m > n$.

Esta dedução, quando aplicada a um sistema homogéneo $\mathbf{Ax} = \mathbf{0}$ resulta na solução trivial $\mathbf{x} = \mathbf{0}$. Esta solução pode ser eliminada optando por restringir um dos coeficientes de \mathbf{x} ao valor um. No entanto, isto faz com que esse coeficiente da solução \mathbf{x} prevaleça sobre os outros, introduzindo grande sensibilidade ao ruído.

Assim, para os sistemas homogéneos, se \mathbf{x} for solução de $\mathbf{Ax} = \mathbf{0}$, então $k\mathbf{x}$ (com k um qualquer escalar) também o é. Logo, a imposição da restrição $\|\mathbf{x}\|^2 = 1$ é razoável, e serve para eliminar a solução trivial $\mathbf{x} = \mathbf{0}$. Esta nova restrição permite que todos os coeficientes de \mathbf{x} tenham o mesmo peso, evitando grandes erros quando comparados com os valores reais de \mathbf{x} . Sendo assim, o método dos mínimos quadrados para esta situação é resolvido da seguinte maneira. Substituindo a equação (2.56) em (2.55) obtém-se

$$\min_{\mathbf{x}} \|\mathbf{USV}^T \mathbf{x}\|^2 \quad \text{s. a} \quad \|\mathbf{x}\|^2 = 1 \Leftrightarrow \min_{\mathbf{x}} \|\mathbf{SV}^T \mathbf{x}\|^2 \quad \text{s. a} \quad \|\mathbf{x}\|^2 = 1 \quad (2.57)$$

Definindo $\mathbf{y} = \mathbf{V}^T \mathbf{x}$, e sabendo que $\|\mathbf{V}\mathbf{x}\| = \|\mathbf{x}\|$, devido à matriz \mathbf{V} ser ortogonal, pode-se apresentar a expressão (2.57) através de

$$\min_{\mathbf{x}} \|\mathbf{SV}^T \mathbf{x}\|^2 \quad \text{s. a} \quad \|\mathbf{V}^T \mathbf{x}\|^2 = 1 \Leftrightarrow \min_{\mathbf{y}} \|\mathbf{S}\mathbf{y}\|^2 \quad \text{s. a} \quad \|\mathbf{y}\|^2 = 1 \quad (2.58)$$

Como \mathbf{S} é uma matriz diagonal, o vector \mathbf{y} que minimiza (2.58) é tal que todos os seus elementos são zero, com excepção de apenas um deles que é um. A posição deste elemento no vector está associada à posição do menor valor próprio de \mathbf{S} . Assim, a solução do sistema $\mathbf{Ax} = \mathbf{0}$ é dada pela coluna da matriz \mathbf{V} que está associada ao menor valor próprio de \mathbf{S} .

A solução aplicada aos sistemas homogéneos também pode ser usada nos sistemas não homogéneos $\mathbf{Ax} = \mathbf{b}$. Para tal, começa-se por criar uma matriz $\mathbf{A}' = \begin{bmatrix} \mathbf{A} & \mathbf{b} \end{bmatrix}$ e usá-la num sistema $\mathbf{A}'\mathbf{x}' = \mathbf{0}$ (onde $\mathbf{x}' = \begin{bmatrix} \mathbf{x} & 1 \end{bmatrix}$). Depois de resolver este novo sistema através do método descrito atrás, a solução pretendida, \mathbf{x} , é dada por $\mathbf{x} = \frac{\mathbf{x}'_{(1:n-1)}}{\mathbf{x}'_{(n)}}$, onde n é o tamanho do vector \mathbf{x}' . Esta solução, comparada com a anterior, elimina o efeito de um dos

coeficientes da solução \mathbf{x} prevalecer sobre os outros, reduzindo então a sua sensibilidade ao ruído.

O apêndice B contém a informação necessária à utilização, no MATLAB[®], da função `lsqsvd.m`, a qual encontra a solução para qualquer sistema de equações através do método dos mínimos quadrados, feito com base no SVD.

Análise dos valores próprios

O método da análise dos valores próprios surge como alternativa ao método anterior, para a resolução de sistemas homogêneos. Inicia-se a sua explicação transformando a equação (2.57) com os multiplicadores de Lagrange em

$$\min_{\mathbf{x}} \mathcal{F}(\mathbf{x}, \lambda)$$

onde

$$\mathcal{F}(\mathbf{x}, \lambda) = \|\mathbf{A}\mathbf{x}\|^2 + \lambda(1 - \|\mathbf{x}\|^2)$$

com λ o multiplicador de Lagrange. Através da equação $\frac{d\mathcal{F}}{d\lambda} = 0$ pode obter-se

$$\mathbf{A}^T \mathbf{A} \mathbf{x} = \lambda \mathbf{x}$$

Com este resultado, conclui-se que a solução \mathbf{x} , do sistema homogêneo $\mathbf{A}\mathbf{x} = \mathbf{0}$, terá que ser um vector próprio unitário da matriz $\mathbf{A}^T \mathbf{A}$, em que λ é o seu correspondente valor próprio [Datta 94] [Recipes 92]. Se a matriz $\mathbf{A}^T \mathbf{A}$ for simétrica e semi-definidamente positiva, todos os seus valores próprios são reais positivos ou nulos [Zhang 96], e a solução \mathbf{x} é o vector próprio unitário da matriz $\mathbf{A}^T \mathbf{A}$, associado ao mais pequeno valor próprio dessa matriz.

Todo este cálculo da solução \mathbf{x} pode ser obtido directamente pelo método de Householder [Datta 94] [Recipes 92], seguido da recuperação do valor e do vector próprio pretendidos, através do método QR [Datta 94] [Recipes 92], ou usando o SVD [Datta 94] [Recipes 92].

A análise dos valores próprios também pode ser usada nos sistemas não homogêneos $\mathbf{A}\mathbf{x} = \mathbf{b}$ seguindo o mesmo procedimento que é feito no SVD para sistemas homogêneos. Isto faz com que os resultados obtidos com o método da análise dos valores próprios sejam equivalentes aos obtidos pelo SVD para sistemas homogêneos, pois existe uma relação entre ambos os métodos. Essa relação está no facto dos valores próprios da matriz $\mathbf{A}^T \mathbf{A}$ serem iguais ao quadrado dos valores singulares de \mathbf{A} ($eig(\mathbf{A}^T \mathbf{A}) = svd^2(\mathbf{A})$).

No apêndice B apresenta-se a informação necessária ao modo de funcionamento, no MATLAB©, da função que encontra a solução para qualquer sistema de equações através do método dos mínimos quadrados, feito com base na análise dos valores próprios, `eigsol.m`.

A pseudo-inversa

Sendo \mathbf{A} uma matriz $m \times n$, a matriz $\mathbf{A}^+ = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$, quando $m \geq n$, ou $\mathbf{A}^+ = \mathbf{A}^T (\mathbf{A} \mathbf{A}^T)^{-1}$, quando $m < n$, é chamada de pseudo-inversa. A matriz pseudo-inversa é também designada por generalização inversa Moore-Penrose [Datta 94].

A matriz pseudo-inversa pode também ser calculada através do SVD [Datta 94], da seguinte forma. Sabendo que $\mathbf{A} = \mathbf{U} \mathbf{S} \mathbf{V}^T$, com $\mathbf{S} = \text{diag}(\tau_1, \tau_2, \dots, \tau_r, 0, \dots, 0)$, a pseudo-inversa é $\mathbf{A}^+ = \mathbf{V} \mathbf{S}^+ \mathbf{U}^T$, onde $\mathbf{S}^+ = \text{diag}(\frac{1}{\tau_1}, \frac{1}{\tau_2}, \dots, \frac{1}{\tau_r}, 0, \dots, 0)$ [Datta 94].

A única solução dos mínimos quadrados para um problema sobre-determinado do tipo $\mathbf{A} \mathbf{x} = \mathbf{b}$, de característica igual ao número de incógnitas, é

$$\mathbf{x} = \mathbf{A}^+ \mathbf{b}$$

Quando nesta solução se usa a pseudo-inversa, calculada sem recurso ao SVD, diz-se que a solução foi obtida pelo método das equações normais.

Para que este método possa ser usado nos sistemas de equações homogêneas, terá que ser condicionada uma das variáveis da solução \mathbf{x} , de forma a poder ser criado o vector \mathbf{b} , à semelhança da solução protagonizada pelo SVD, para sistemas não homogêneos. Este vector terá como elementos os valores da coluna da matriz \mathbf{A} que estiver associada à variável (similar a um algoritmo de oito pontos, generalizado para qualquer matriz). Introduce-se, tal como a solução SVD para sistemas não homogêneos, sensibilidade a ruído, devido ao efeito de um coeficientes da solução \mathbf{x} prevalecer sobre os outros.

Remete-se para o apêndice B a informação necessária ao uso das funções `psinvsol.m` e `psinvsvd.m`, no MATLAB©, as quais encontram a solução para qualquer sistema de equações através do método dos mínimos quadrados, feito com base na pseudo-inversa, respectivamente, via equações normais ou via SVD.

2.7.2 Consenso aleatório de amostras - RANSAC

Com o intuito de controlar a sensibilidade introduzida pelo não condicionamento dos dados de entrada (característica dos métodos expostos anteriormente), pode recorrer-se ao

método do consenso aleatório de amostras, também conhecido pelas iniciais RANSAC. Esta designação tem origem na terminologia inglesa *Random Sample Consensus*. O RANSAC, apresentado pela primeira vez por Fischler e Bolles [Fischler 81] para uma aplicação de análise de imagens, é um método robusto de adaptação de dados a um modelo específico e tem encontrado um número crescente de aplicações na área da visão por computador. Através da interpretação dos dados, o RANSAC é capaz de suavizar a acção de uma grande percentagem de erros na estimação do modelo. No modelo a estimar utilizam-se também outros métodos numéricos, como os três descritos atrás.

O procedimento do RANSAC é oposto ao das técnicas convencionais de eliminação de erros. Em vez de usar o maior número de dados possível para obter uma solução inicial, e depois eliminar os dados inválidos, o RANSAC usa o conjunto de dados mais pequeno possível, para a definição completa do modelo, e, quando pode, alarga esse conjunto com dados consistentes com o modelo.

Este método pode ser formalmente descrito como se segue. Dado um modelo que disponha de um conjunto de dados maior ou igual a n , e necessite de um mínimo de n dados para que todos os seus parâmetros sejam conhecidos, escolhe-se, aleatoriamente, um subconjunto de n dados e calcula-se o modelo que lhes corresponde. Com esse modelo, determina-se, entre todos os dados, quantos são consistentes com ele, dentro de uma certa tolerância pré-definida. Ao conjunto de dados consistentes que se obtém, chama-se conjunto de consenso.

Se o cardinal do conjunto de consenso for maior ou igual a um limite mínimo escolhido, estima-se o modelo final com esses dados, pois todos os outros dados são considerados como erros. Se, pelo contrário, o cardinal for menor, reinicia-se o processo. Se depois de um pré-determinado número de vezes o grupo de consenso não for obtido, termina-se o processo e volta-se a aplicar o método aumentando o número de dados. A alternativa é ficar com o conjunto de consenso mais próximo do limite mínimo imposto, consoante o objectivo final pretendido.

Como foi possível verificar, existem três parâmetros que dependem do utilizador, a tolerância usada para eliminar dados, o limite mínimo de dados para o conjunto de consenso, e o número de iterações a fazer caso não se encontre o consenso. Dependendo da aplicação, alguns destes parâmetros poderão ser ignorados. Fischler e Bolles, em [Fischler 81], apresentam um estudo sobre o comportamento destes parâmetros.

Existem alguns melhoramentos que se podem introduzir no RANSAC. Um deles é o de

o tornar determinístico na maneira de escolher os dados. O outro, uma vez estabelecido o grupo de consenso final, é o de adicionar novos dados consistentes com o modelo estimado e calcular o novo modelo com o conjunto de dados alargado.

Um algoritmo de amostragem de dados aleatório mais recente, o MINPRAN (minimização da probabilidade de aleatoriedade), foi descrito por Stewart [Stewart 95]. Este método não necessita, ao invés do RANSAC, do conhecimento, *a priori*, da tolerância para eliminar os dados.

2.7.3 Comportamento dos métodos numéricos apresentados

Nesta subsecção comparam-se os métodos numéricos lineares apresentados anteriormente, nomeadamente o método da decomposição em valores singulares (SVD), o método da análise dos valores próprios (EIG) e o método da pseudo-inversa (PSI).

Qualquer aplicação de um método numérico implica, obviamente, um conjunto de dados de entrada, um modelo matemático que os processa segundo a forma desejada, e um conjunto de dados de saída resultantes do processamento. Qualquer erro nos dados de entrada ou no método numérico afecta os resultados de saída.

Sendo assim, inicia-se esta comparação com um simples caso de intersecção de várias rectas. Este caso foi escolhido não só pela sua simplicidade, mas também por ser uma das operações necessárias a um dos processos de reconstrução defendidos nesta dissertação.

	<i>P0</i>	<i>P1</i>
recta 1	(72, 182)	(644, 008)
recta 2	(72, 182)	(682, 443)
recta 3	(72, 182)	(271, 548)

Tabela 2.2: *Conjunto de dados de entrada base, para a comparação entre processos numéricos.*

Dentro deste caso, considere-se como primeira situação o conjunto de dados de entrada contidos na tabela 2.2. O resultado numérico da intersecção das rectas da tabela 2.2, aplicando os métodos PSI, SVD e EIG, é sempre o mesmo, o ponto $e \equiv (72, 182)$, tal como se comprova pela sua representação na figura 2.14(a).

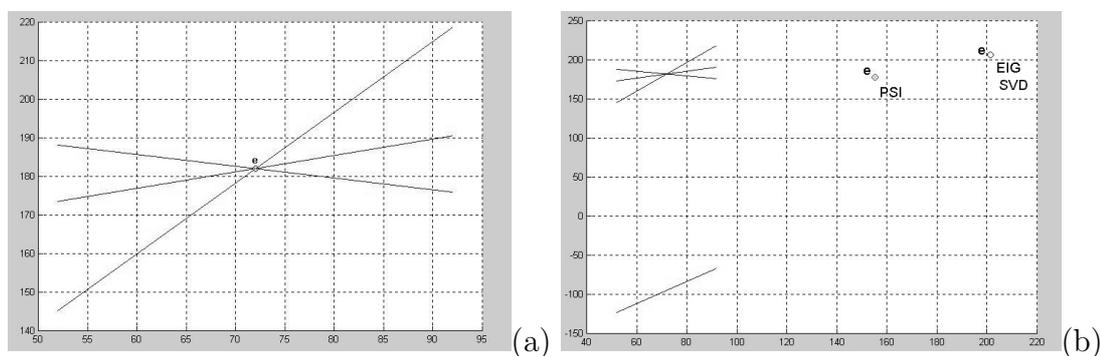


Figura 2.14: Resultado gráfico da intersecção das rectas, usando os métodos PSI, SVD e EIG. (a) Para a tabela 2.2; (b) Para a tabela 2.3.

Com a introdução, no conjunto de dados de entrada base da tabela 2.2, de uma nova recta, que não passe pelo ponto e , representado na figura 2.14(a), apresenta-se a segunda situação. Os dados de entrada para esta segunda situação são os da tabela 2.3.

	$P0$	$P1$
recta 1	(072, 182)	(644, 008)
recta 2	(072, 182)	(682, 443)
recta 3	(072, 182)	(271, 548)
recta 4	(460, 453)	(159, 028)

Tabela 2.3: Novo conjunto de dados de entrada para a comparação entre processos numéricos, onde aos dados base, da tabela 2.2, foi acrescentada uma recta errada.

Com esta introdução cria-se um erro grosseiro no conjunto de dados de entrada, o qual fica espelhado no resultado do cálculo do ponto de intersecção feito por cada um dos métodos, como se pode confirmar na tabela 2.4 e no gráfico da figura 2.14(b).

Se nesta segunda situação aplicarmos a cada um dos três métodos o consenso aleatório de amostras (RANSAC), os resultados voltam a ser iguais aos obtidos na figura 2.14(a). Graficamente, desta aplicação resulta o apresentado na figura 2.15.

Com a apresentação deste caso constata-se que, com dados de entrada erróneos, nenhum dos métodos usados consegue obter uma solução certa ou, pelo menos, muito aproximada. Para a obter, torna-se necessário aplicar-se-lhes um método que filtre os dados de entrada de forma a eliminar os erros. Tal como foi mostrado, um desses métodos poderá ser o

<i>Método numérico</i>	<i>Resultado da intersecção</i>
PSI	(151.15, 177.87)
SVD	(201.19, 206.60)
EIG	(201.19, 206.60)

Tabela 2.4: *Resultado numérico da intersecção das rectas da tabela 2.4, para cada um dos métodos PSI, SVD e EIG.*

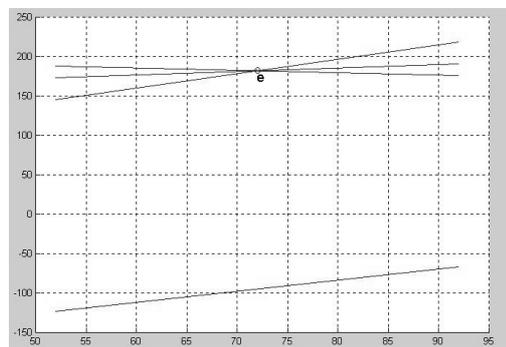


Figura 2.15: *Resultado gráfico da intersecção das rectas da tabela 2.4, para os métodos PSI, SVD e EIG, aplicando-lhes o RANSAC.*

RANSAC, o qual será o usado ao longo desta dissertação, quando houver essa necessidade de filtragem.

O caso de intersecção apresentado usa dados sintéticos. Na prática, o conjunto de entrada para uma intersecção de rectas pode conter, ao mesmo tempo, algumas rectas totalmente desenquadradas, outras rectas com pequenos erros e ainda outras completamente consistentes com o modelo (rectas concorrentes no mesmo ponto). A figura 2.16 ilustra graficamente esta situação. Como se nota, tendo em mente a solução certa (da intersecção de várias rectas só pode surgir um ponto), as rectas correctas e as que contêm pequenos erros transmitem a ideia de onde deverá estar posicionado correcto do ponto de intersecção. No entanto, as rectas grosseiramente erradas fazem com que a solução final se afaste dessa posição.

Numa situação destas, o resultado final após a aplicação do RANSAC poderá não passar por qualquer das rectas correctas, como indica a figura 2.17(a). Este facto deve-se à tolerância usada pelo RANSAC que, quanto maior for, mais dados errados deixará de fil-

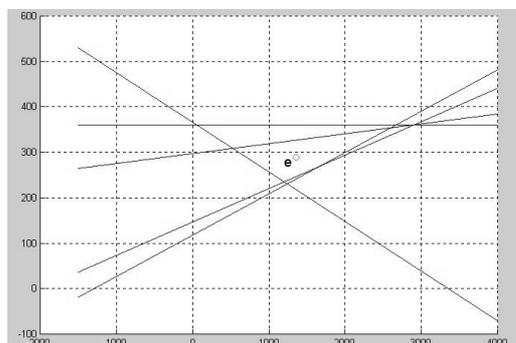


Figura 2.16: Resultado gráfico da intersecção de várias rectas, algumas totalmente correctas e outras com pequenos e grandes erros.

trar. No entanto, uma demasiada limitação desta tolerância poderá não trazer benefícios mas sim mais problemas, como se pode ver na figura 2.17(b), onde o ponto de intersecção obtido não corresponde à solução óptima.

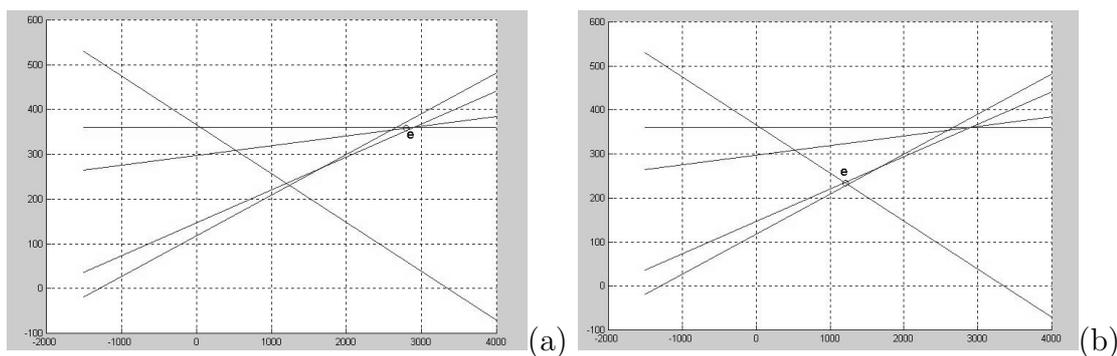


Figura 2.17: Resultado gráfico da intersecção de várias rectas, algumas totalmente correctas e outras com pequenos e grandes erros, aquando da aplicação do RANSAC. (a) Com uma tolerância que permite eliminar as rectas grosseiramente erradas; (b) Com uma tolerância muito pequena.

Para todas as situações em que houve eliminação de falsas rectas, através da aplicação do RANSAC, o processo foi feito com base na distância do ponto de intersecção calculado, a cada recta.

Continuando a descrição do comportamento prático dos três métodos numéricos (PSI, SVD e EIG), passemos à observação da estimação da matriz fundamental para um conjunto de pontos correspondentemente conhecidos, pois este é um dos cálculos necessários ao trabalho

aqui desenvolvido. Assim, serão apresentados dois casos: um primeiro que usa todos os pontos correspondentes possíveis encontrados nas imagens (73 pontos), e um segundo caso que utiliza apenas os pontos necessários à obtenção de uma solução (8 pontos). Para cada caso, e por cada método, apresentam-se quatro estimações da matriz fundamental. A primeira estimativa utiliza a forma geral do método sem nenhum condicionamento e é designada por G. A segunda, designada por N+G, condiciona apenas os dados de entrada. Por seu lado, a terceira condiciona somente os dados de saída e é designada por G+C2. Na última estimativa, N+G+C2, são feitos ambos os condicionamentos.

Com os resultados da estimativa são calculados alguns valores de grande interesse a esta análise, os quais podem ser vistos na tabela 2.5. Com essa tabela, pretende-se mostrar se os três métodos lineares referidos respeitam todas as restrições do problema a resolver (matriz fundamental ter que ser singular), através da segunda coluna (valor do determinante), da quarta coluna (posição do epipolo na imagem esquerda) e da quinta coluna (posição do epipolo na imagem direita). Os valores para as duas últimas colunas são obtidos através do cálculo do espaço nulo da respectiva matriz fundamental e da sua transposta.

Para uma mais fácil e melhor comparação dos resultados destes métodos, existiu a necessidade de se encontrar um factor mesurável que os pudesse relacionar (é claro que este factor deverá ser aplicado ao caso específico que se está a tratar). Para a estimativa da matriz fundamental, são vários os autores que usam a norma de frobenius [Datta 94] aplicada à diferença entre a matriz estimada e uma matriz de referência ($\|\mathbf{F} - \mathbf{F}_{ref}\|$) [Zhang 94]. Apesar deste factor relacionar as matrizes através do ângulo entre os seus vectores unitários, e necessitar muito pouco tempo de processamento, ele não é muito apropriado, podendo levar a resultados enganadores [Zhang 94]. Outro método utilizado, proposto por Stéphane Laveau [Zhang 94], usa como factor de comparação a média das distâncias entre pontos aleatórios de cada uma das rectas epipolares, provenientes das duas matrizes fundamentais que se querem comparar. Apesar deste factor de comparação ser muito bom, tem como grande desvantagem a necessidade de muito tempo de processamento. O factor de comparação escolhido é apresentado na terceira coluna da tabela 2.5. Esta é uma situação muito mais leve em termos de tempo de processamento, e foi designada como factor de qualidade da matriz fundamental, Q_f . Este factor é obtido através do cálculo da raiz da média quadrática da distância entre os pontos correspondentes conhecidos, e as respectivas rectas epipolares. O seu cálculo está descrito no apêndice A. No entanto, em termos gerais, não há diferença entre os resultados de uma comparação com este factor de qualidade ou

com o factor usado por Stéphane Laveau.

A primeira referência a fazer à tabela 2.5 é que as situações G e N+G, de qualquer dos métodos, não originam uma solução verdadeira para a matriz fundamental. De facto, apesar dos valores do determinante serem muito pequenos, podendo ser considerados zero, a matriz fundamental não é singular, pois com o seu espaço nulo não se obtêm os epipolos. Isto leva a grandes erros, quando estas soluções são usadas posteriormente.

É um facto que, muitas das vezes, a aplicação geral, de um destes três métodos lineares, falha por não poder considerar algumas restrições do problema a resolver. Essa falha só acontece quando os sistemas têm mais do que uma solução. Daí chamar-se a atenção para a importância da verificação de todas as restrições do problema. Quando essas restrições não são totalmente respeitadas, há que fazer um pós-processamento. No caso que se está a apresentar, como à matriz fundamental obtida lhe faltava verificar apenas a restrição da característica 2, com as situações G e N+G, ela teve que ser imposta depois. Assim, a análise da tabela 2.5 terá que se concentrar apenas nas situações G+C2 e N+G+C2.

A segunda referência à tabela 2.5 prende-se com os piores resultados obtidos para o caso 2. Torna-se claro que estes métodos lineares são muito sensíveis a pequenos erros (ruído) existentes nos dados de entrada. Quando se tem poucos dados de entrada, a presença do ruído é mais notada já que, com muito mais dados ele é dissolvido.

A melhor situação que é apresentada na tabela 2.5 é a N+G+C2. A razão para isto deve-se a um pré-processamento antes da aplicação do método. Este pré processamento, conhecido como normalização dos dados, serve apenas para diluir o ruído pelos dados de entrada existentes. Obviamente, tirando importância ao ruído, os resultados melhoram, como se constata.

Não podendo dizer que um dos métodos se sobrepõe aos outros, podemos, no entanto, afirmar qual o melhor, para problemas que têm várias soluções (até porque estes são os problemas reais com que lidamos nos processos de reconstrução).

Dos dados da tabela 2.5 verifica-se que os métodos da decomposição dos valores singulares e da análise dos valores próprios atingem sempre melhores soluções, usando para isso o factor de qualidade da solução, Q_f . Este facto deve-se, unicamente, a que, no método da pseudo-inversa, um dos coeficientes da solução prevalece sobre os outros, enquanto que nos métodos anteriores esse peso é igual para todos os coeficientes. No entanto, é de notar que os resultados do SVD foram obtidos através da aplicação dos mínimos quadrados para sistemas homogéneos.

Em conclusão, convém estar atento ao método numérico a usar no problema de forma a não obter resultados menos bons. A experiência adquirida foi a de que, muitas vezes, esses resultados menos bons (como são parecidos com os melhores resultados) são a causa de falta de robustez da solução para o problema.

Métodos	$\det(\mathbf{F})$	Q_f	e_e	e_d
Caso 1- 73 pontos				
<u>Pseudo-inversa</u>				
G	$-8.01e - 10$	$1.90e - 01$	—	—
N+G	$-7.16e - 11$	$2.80e - 03$	—	—
G+C2	$-8.01e - 24$	$1.91e - 01$	(-155.88, 488.33)	(-171.22, 492.94)
N+G+C2	$+2.52e - 26$	$2.87e - 03$	(-15380, 19333)	(-169970, 212120)
<u>SVD</u>				
G	$+8.02e - 10$	$1.90e - 01$	—	—
N+G	$+7.85e - 12$	$2.78e - 03$	—	—
G+C2	$-9.02e - 23$	$1.91e - 01$	(-157.36, 489.94)	(-172.74, 494.6)
N+G+C2	$+1.40e - 26$	$2.82e - 03$	(-3489.2, 4049.8)	(-4396.6, 5050.2)
<u>Valores próprios</u>				
G	$+8.02e - 10$	$1.90e - 01$	—	—
N+G	$+7.85e - 12$	$2.78e - 03$	—	—
G+C2	$-9.02e - 23$	$1.91e - 01$	(-157.36, 489.94)	(-172.74, 494.6)
N+G+C2	$+1.40e - 26$	$2.82e - 03$	(-3489.2, 4049.8)	(-4396.6, 5050.2)
Caso 2- 8 pontos				
<u>Pseudo-inversa</u>				
G	$-2.48e - 07$	$3.43e - 10$	—	—
N+G	$+3.65e - 04$	$2.00e - 13$	—	—
G+C2	$-4.58e - 22$	$3.14e + 00$	(68.903, 115.32)	(49.251, 101.54)
N+G+C2	$-1.12e - 19$	$3.09e - 02$	(67.378, 109.44)	(50.94, 106.45)
<u>SVD</u>				
G	$-2.43e - 07$	$3.53e - 07$	—	—
N+G	$-1.01e - 08$	$3.81e - 14$	—	—
G+C2	$-8.31e - 23$	$3.14e + 00$	(68.904, 115.32)	(49.251, 101.54)
N+G+C2	$-1.11e - 23$	$3.09e - 02$	(67.378, 109.44)	(50.94, 106.45)
<u>Valores próprios</u>				
G	$-2.43e - 07$	$3.53e - 07$	—	—
N+G	$-1.01e - 08$	$3.81e - 14$	—	—
G+C2	$-8.31e - 23$	$3.14e + 00$	(68.904, 115.32)	(49.251, 101.54)
N+G+C2	$-1.11e - 23$	$3.09e - 02$	(67.378, 109.44)	(50.94, 106.45)

Tabela 2.5: Comportamento dos métodos de estimação da matriz fundamental. G refere-se às estimações que utilizam a forma geral método sem condicionamentos. Nas linhas $N+G$ e $G+C2$ acrescentou-se, respectivamente, um condicionamento dos dados de entrada e um dos dados de saída. Nas linhas $N+G+C2$ aplicaram-se ambos os condicionamentos.

Capítulo 3

Recuperação de pontos 3D

Neste capítulo é apresentado o trabalho desenvolvido com vista à criação de um sistema de determinação de informação 3D, baseado em espelhos e em imagens não calibradas.

3.1 Introdução

Conforme já foi referido, o grande objectivo deste trabalho é o de desenvolver um método linear robusto, baseado em superfícies reflectoras e na geometria epipolar, destinado à determinação de informação 3D, euclidiana, a partir de imagens não calibradas. Uma das tarefas que um processo deste tipo pode realizar é a recuperação de pontos 3D. Este capítulo descreve, passo a passo, como essa tarefa foi desenvolvida.

Começou-se por desenvolver uma primeira aproximação, centrada na matriz fundamental, cujo diagrama modular é o apresentado na figura 1.1, com a designação de (a). Esta aproximação, descrita na primeira secção deste capítulo, seguiu a linha dos processos clássicos de reconstrução, embora tenha tratado de maneira diferente a questão da calibração. Apesar da validade do trabalho realizado nesta fase, concluiu-se ser um método instável devido à obtenção das características da câmara (calibração) ser feita através da matriz fundamental.

Consequentemente, passou-se ao desenvolvimento da aproximação (b), conforme a figura 1.1, ou seja, à reconstrução centrada nos espelhos, tratada na terceira secção deste capítulo, e com a qual foi atingida a meta proposta. Este método teve como pilares de suporte o posicionamento dos espelhos, as transformações projectivas e o cálculo da matriz

fundamental (simplificado significativamente face à aproximação (a)). Estas bases permitiram reduzir o processo clássico de recuperação de pontos 3D euclidianos, a partir de imagens, a uma intersecção de rectas 3D com parâmetros conhecidos e perfeitamente definidos.

3.2 Reconstrução centrada na matriz fundamental

Esta secção descreve a primeira aproximação à criação de um sistema experimental compacto, rápido e de baixo custo, para obtenção das coordenadas tridimensionais de pontos em objectos estáticos e opacos. Para tal utilizou-se imagens estéreo (explorando o potencial dos espelhos), transformações projectivas, geometria epipolar (matriz fundamental) e triangulação.

Com o fim de baixar os custos dos sistemas de aquisição de imagens, pode-se usar o potencial dos espelhos, tal como foi visto no ponto 2.6, para reduzir o número de câmaras. Assim, esta aproximação simula uma geometria estéreo binocular, usando a configuração estudada e apresentada na figura 2.11, para a obtenção das imagens.

Na correspondência entre os pontos nas imagens é explorada a relação entre duas câmaras, colocadas em posições diferentes, pois, como se confirma pela figura 2.11, a imagem resultante da aplicação da configuração utilizada é a junção de duas imagens da mesma cena 3D. Desta forma, é reduzida a complexidade da geometria epipolar e conseguida uma computação mais robusta da matriz fundamental. Quando se pretende fazer apenas corresponder as projecções de pontos pertencentes a um plano tridimensional, utilizam-se as transformações projectivas, especificamente as homografias, de forma a otimizar o processo. No entanto, a estimação da matriz fundamental torna-se sempre necessária devido à sua utilização na fase de reconstrução.

A solução para a obtenção dos pontos 3D é baseada no processo clássico de reconstrução para duas câmaras. No entanto, esta reconstrução, tal como foi mostrado na secção 2.3, necessita do conhecimento dos parâmetros intrínsecos e extrínsecos das câmaras, ou, por outras palavras, de um processo de calibração das mesmas. Para isso, foi considerado o problema da auto-calibração, no contexto em que as duas câmaras estão posicionadas lateralmente com os eixos ópticos coplanares a intersectarem-se. Cada câmara tem, no entanto, a possibilidade de variação do ângulo de vergência. Com este método directo, os parâmetros necessários à calibração são retirados da matriz fundamental uma vez que se

conhece a relação entre os sistemas de coordenadas das câmaras virtuais e do mundo.

3.2.1 O sistema de aquisição

Para recuperar uma cena tridimensional a partir de imagens, explorando a disparidade entre os pontos projectados, o sistema criado (baseado na estereoscopia) é formado por quatro espelhos e uma única câmara, como se pode observar na figura 3.1. A sua geometria

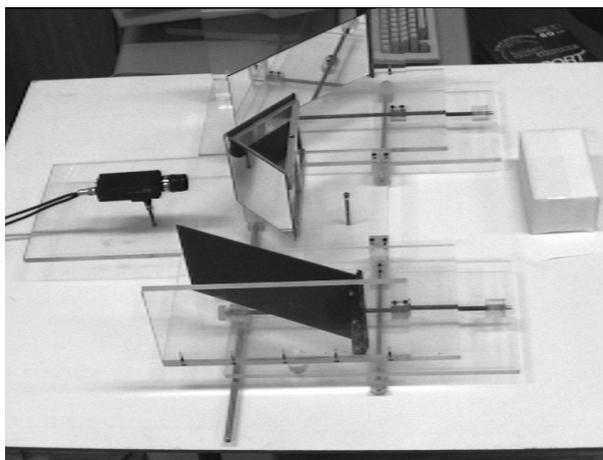


Figura 3.1: *Imagem do sistema experimental para ser utilizada a aproximação (a).*

permite a aquisição de duas imagens de uma só vez, vistas de posições diferentes, por câmaras virtuais com as mesmas características da câmara real, sendo a imagem resultante a junção das duas imagens da mesma cena.

Recorda-se que os espelhos possibilitam a multiplicação de sensores caros, sem alteração de resultados, além de permitirem a criação de sistemas compactos e fáceis de manejar, por se poder controlar o seu tamanho e volume. Logo, esta aproximação tem como grandes vantagens o baixo custo e a grande versatilidade.

Na configuração geométrica em que se baseia esta aproximação, o ângulo σ é dado por

$$\tan(\sigma) = \frac{|a - c|}{b}$$

Através de testes realizados para a obtenção de imagens estéreo com uma boa área comum, $\sigma \cong 0$. Então, sem perda significativa de precisão, optou-se por restringir $\sigma = 0$. Esta situação conduz a $a = c$, passando a configuração usada para a que se apresenta na figura 3.2.

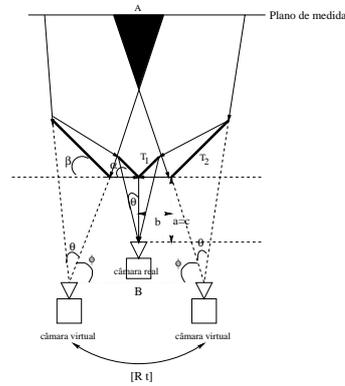


Figura 3.2: Corte lateral da configuração seguida na aproximação (a).

A principal característica desta configuração é a simetria dos espelhos em relação à câmara. Esta simetria serve essencialmente para facilitar o cálculo da medida B da figura 3.2. Esta medida, normalmente chamada de *baseline*, é a distância entre os centros ópticos das câmaras (neste caso virtuais). Assim, através da igualdade de triângulos e tendo por base a figura 2.11, a forma geral da *baseline*, B , é dada por

$$B = 2 \left[b + \cos(\phi) \frac{a+b}{\cos(\sigma)} \right]$$

ou por

$$B = 2 \left[b + (a + b)\cos(\phi) \right] \quad (3.1)$$

quando $\sigma = 0$, com os valores a , b e ϕ , representados na figura 3.2. Os parâmetros a e b são dois dos dados fixos do sistema e os seus valores obtêm-se por medição. Na tabela 3.1 são apresentados todos esses dados métricos fixos, usados para a aquisição das imagens estéreo.

Dados Físicos		Dados controlados
$T_1 = 15.0cm$	$T_2 = 29.5cm$	$a = 6.85cm$
$b = 16cm$	$\sigma = 0^\circ$	$c = 6.85cm$
$\alpha = 45^\circ$	$\beta = 55^\circ$	$\theta = 31.28^\circ$

Tabela 3.1: Dados métricos fixos usados pelo sistema de aquisição de imagens criado.

Como se constata na equação (3.1), para o conhecimento completo dos parâmetros do

sistema de aquisição de imagens, falta ainda saber o valor de ϕ . Este dado será obtido de seguida, através do uso da geometria epipolar.

3.2.2 Correspondência

Duas vistas de uma só cena ou objecto, ambas obtidas a partir de uma câmara móvel ou de duas câmaras estéreo, são relacionadas pela geometria epipolar, podendo ser descritas pela matriz fundamental, \mathbf{F} . Nesta aproximação, a solução para a correspondência utiliza a restrição resultante da configuração geométrica dos sistemas de coordenadas das câmaras. Para poder usar esta restrição, basta conhecer a relação entre as câmaras. De facto, o posicionamento dos espelhos facilita a obtenção desta relação, como se irá verificar para o ângulo ϕ . Veremos também a forma como esta geometria simplifica as equações na estimação da matriz fundamental, implicando somente o conhecimento de seis pontos correspondentes nas imagens estéreo. Apesar do caso ser específico à configuração apresentada, faremos esta dedução de forma generalizada, para duas câmaras com vergências diferentes.

Tendo por base a configuração do sistema, apresentada na figura 3.2, consideram-se as duas câmaras virtuais para a dedução da matriz fundamental. Devido ao facto de os sistemas de coordenadas das câmaras virtuais e do mundo não possuírem, neste caso, movimento relativo, e devido ao posicionamento dos espelhos, é assumido, sem perda de generalidade, que $Y_e = Y_d = -Y$ e que X_eZ_e , X_dZ_d e XZ estão no mesmo plano. Por outras palavras, considera-se que o sistema não tem ciclotorsão. Esta geometria é apresentada de uma forma mais clara na figura 3.3, onde as câmaras virtuais, esquerda e direita, têm as vergências ϕ_e e ϕ_d , respectivamente, e onde os seus eixos ópticos contêm P_f , chamado ponto de fixação dos eixos ópticos das câmaras. Nas figuras 3.2 e 3.3, verifica-se, também, que as duas câmaras estão relacionadas através de uma translação \mathbf{t} ao longo da *baseline*, B , no plano XZ , e também através de uma rotação \mathbf{R} em torno do eixo Y . Tendo por base a equação (2.4), o vector \mathbf{t} e a matriz \mathbf{R} são dados por

$$\mathbf{t} = \begin{bmatrix} B \sin(\phi_e) \\ 0 \\ B \cos(\phi_e) \end{bmatrix} \quad \mathbf{R} = \begin{bmatrix} -\cos(\phi_e + \phi_d) & 0 & \sin(\phi_e + \phi_d) \\ 0 & 1 & 0 \\ -\sin(\phi_e + \phi_d) & 0 & -\cos(\phi_e + \phi_d) \end{bmatrix}$$

Logo, a matriz fundamental, \mathbf{F} , é definida da mesma forma que a apresentada na tabela

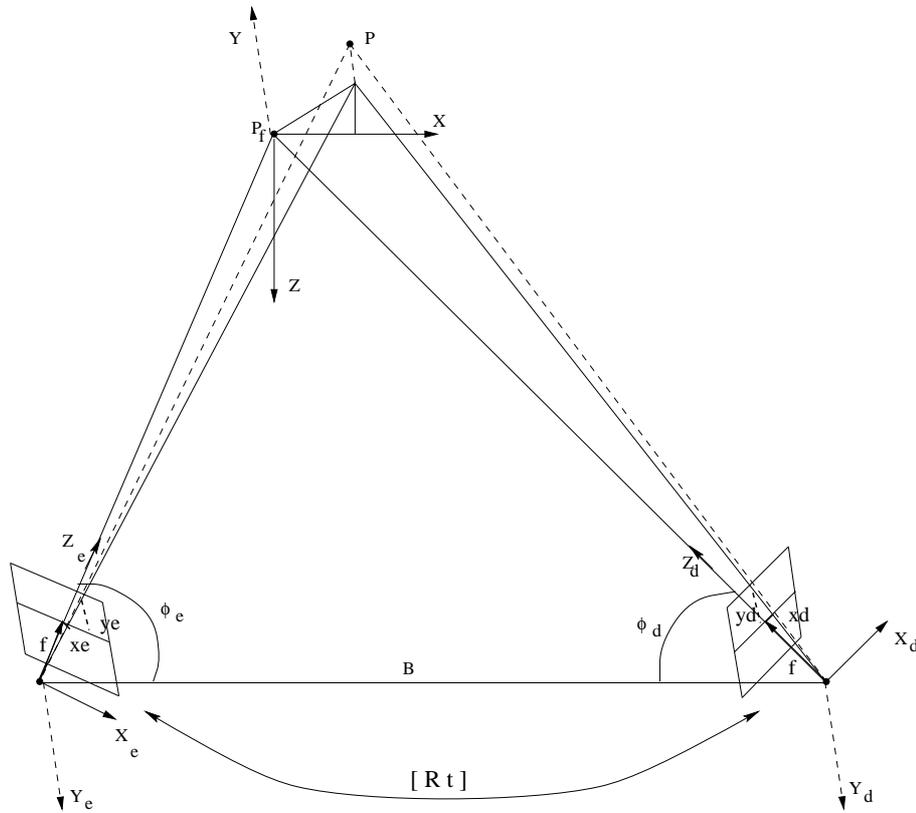


Figura 3.3: Representação do sistema de coordenadas de duas câmaras virtuais e do mundo. O eixo óptico de cada câmara passa em P_f . Nesta configuração $Y_e = Y_d = -Y$ e $X_e Z_e$, $X_d Z_d$ e XZ estão no mesmo plano, ou seja, o sistema não tem ciclotorsão.

2.1, para movimentos planares no plano XZ , por

$$\mathbf{F} = \begin{bmatrix} 0 & f_{12} & f_{13} \\ f_{21} & 0 & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}$$

O problema considerado em sequência é o da estimação de \mathbf{F} , a partir de um conjunto suficiente de pontos correspondentes $(\bar{p}_{e_i}, \bar{p}_{d_i}) : i = 1, \dots, n$, onde $n \geq 6$, similarmente ao que foi mostrado na subsecção 2.4.2. A única diferença reside na equação (2.28) para cada par de pontos nas imagens, que agora fica reduzida a

$$\mathbf{g}_i = \begin{bmatrix} x_r y_l & x_r & x_l y_r & y_r & x_l & y_l & 1 \end{bmatrix}$$

e

$$\mathbf{f} = \begin{bmatrix} f_{12} & f_{13} & f_{21} & f_{23} & f_{31} & f_{32} & f_{33} \end{bmatrix}^T$$

Esta redução do número de pontos correspondentes em cada imagem, torna-se vantajosa quando existem poucos pontos característicos (potenciais correspondentes) nas imagens. Por outro lado, se tivermos muitos pontos correspondentes, esta redução possibilita um aumento de 33.(3)%, na probabilidade da matriz \mathbf{F} obtida ser a mais correcta em relação à estimação de \mathbf{F} pelo algoritmo dos 8 pontos.

Finalmente, para terminar a explanação sobre o que foi feito no processo de correspondência, há que falar sobre a procura de outros pontos correspondentes nas imagens. Para isso, basta usar um dos muitos programas existentes para fazer uma pesquisa automática desses pontos nas rectas epipolares. Caso se queira otimizar este processo, isto quando se pretender apenas reconstruir superfícies planares, pode-se recorrer à estimação de homografias (uma por cada superfície), as quais relacionam as projecções de pontos dessas superfícies, em cada imagem. Como se verá de seguida, nesta aproximação a matriz fundamental tem sempre que ser determinada.

3.2.3 Auto-calibração

Todas as matrizes fundamentais guardam em si a informação sobre a geometria em que as imagens foram captadas. Veremos que a matriz fundamental, para este caso, não foge à normalidade. Sabendo que a geometria usada é a da figura 3.2, constata-se que ambas as câmaras virtuais têm a mesma vergência ϕ ($|\phi_e| = |\phi_d|$). Tendo este facto em conta, e utilizando as equações (2.6) e (2.19), a matriz \mathbf{F} vem dada pela seguinte forma

$$\mathbf{F} = \begin{bmatrix} 0 & f_{12} & f_{13} \\ f_{21} & 0 & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} = \begin{bmatrix} 0 & \frac{B \cos(\phi)}{f^2 k_x k_y} & -c_y f_{12} \\ \frac{B \cos(3\phi)}{f^2 k_x k_y} & 0 & \frac{-B[c_x \cos(3\phi) + f k_x \sin(3\phi)]}{f^2 k_x k_y} \\ -c_y f_{21} & \frac{-B[c_x \cos(\phi) + f k_x \sin(\phi)]}{f^2 k_x k_y} & -c_y (f_{23} + f_{32}) \end{bmatrix} \quad (3.2)$$

Fazendo primeiro a estimação de \mathbf{F} pode-se, posteriormente, e mediante o conhecimento da equação (3.2), conseguir dois dados importantíssimos na obtenção dos pontos 3D. Esses dados são o ângulo de vergência das câmaras, ϕ , e o quociente de aparência das imagens estéreo, $k_a = \frac{k_y}{k_x}$, necessários ao processo de reconstrução propriamente dito, que são dados

por

$$|\phi| = \arcsin\left(\frac{\sqrt{f_{12}-f_{21}}}{2\sqrt{f_{12}}}\right) \quad k_a = \frac{2(c_x f_{12} + f_{32})^2 \sqrt{3f_{12} + f_{21}}}{B(f_{12} - f_{21})\sqrt{f_{12}}}$$

3.2.4 Reconstrução tridimensional dos pontos

A reconstrução clássica dos pontos 3D é baseada nas projecções de duas imagens. Uma vez estabelecida a correspondência unívoca entre essas projecções, o próximo passo a dar é o da recuperação das coordenadas tridimensionais desses pontos.

Considerando a geometria da figura 3.3, onde é conhecido o ângulo de vergência das câmaras, $|\phi_e| = |\phi_d| = \phi$, a *baseline*, B , e com base na equação (2.6), define-se para cada câmara virtual, em coordenadas homogêneas,

$$\mathbf{P}_e = \begin{bmatrix} \mathbf{R}_e & \mathbf{t} \end{bmatrix} \mathbf{P} \quad \mathbf{P}_d = \begin{bmatrix} \mathbf{R}_d & \mathbf{t} \end{bmatrix} \mathbf{P} \quad (3.3)$$

com as matrizes de rotação e o vector de translação dados por

$$\mathbf{R}_e = \begin{bmatrix} \sin(\phi) & 0 & \cos(\phi) \\ 0 & -1 & 0 \\ \cos(\phi) & 0 & -\sin(\phi) \end{bmatrix} \quad \mathbf{R}_d = \begin{bmatrix} \sin(\phi) & 0 & -\cos(\phi) \\ 0 & -1 & 0 \\ -\cos(\phi) & 0 & -\sin(\phi) \end{bmatrix} \quad \mathbf{t} = \begin{bmatrix} 0 \\ 0 \\ \frac{B}{2\cos(\phi)} \end{bmatrix}$$

pois é considerado, sem perda de generalidade, que a origem do sistema é o ponto P_f .

Considere-se o ponto genérico tridimensional, $P = (X, Y, Z)$, da figura 3.3, e respectivas projecções nas imagens esquerda e direita $p_e = (x_e, y_e)$ e $p_d = (x_d, y_d)$. Tendo como base as equações (2.1), (2.2) e (2.6), e sabendo que na equação (3.3), $\mathbf{P}_e(2) = \mathbf{P}_d(2)$ (notar que $\mathbf{A}(i)$ é a linha i da matriz \mathbf{A}), então

$$X = \frac{C_1 C_2 \tan(\alpha)}{1 - C_1 \tan(\alpha)} + C_2$$

e

$$Z = \frac{C_2 \tan(\alpha)}{1 - C_1 \tan(\alpha)}$$

com

$$C_1 = \frac{y_e - y_d}{y_d + y_e} \tan \phi \quad C_2 = \frac{B(y_d - y_e)}{2(y_d + y_e) \cos^2(\phi)} \quad \tan(\alpha) = \frac{Z}{X} = \frac{x_d y_e - x_e y_d}{x_d y_e + x_e y_d} \tan(\phi)$$

Por seu lado, a coordenada Y é obtida usando a semelhança entre os triângulos, representados na figura 3.3, com a expressão

$$\frac{f_y P(2)[P_e(3) - P_d(3)]}{y_d - y_e} = \frac{f_x [P_d(1)P_e(3) - P_e(1)P_d(3)]}{x_d - x_e}$$

onde f_x e f_y são dados por $f_i = fk_i$, $i = x, y$, com f a distância focal da câmara real (que é igual à das câmaras virtuais) e k_i o factor de conversão, no eixo i da imagem, do sistema métrico do mundo para o sistema métrico da imagem (*pixels*). Como resultado desta aplicação obtém-se

$$Y = \frac{1}{k_a} \frac{y_d - y_e}{x_d - x_e} \left\{ \sin(\phi) \left[C_2 + \frac{[C_1 + \tan(\alpha)]C_2 \tan(\alpha)}{1 - C_1 \tan(\alpha)} \right] - \frac{B \tan(\alpha)}{2 \cos(\phi)} \right\}$$

em que $k_a = \frac{f_y}{f_x} = \frac{k_y}{k_x}$ é o quociente de aparência das imagens estéreo, obtido através da matriz fundamental, na secção anterior. Em [Weinshall 87] pode ver-se a dedução da $\tan(\alpha)$ em termos das coordenadas das imagens.

A diferença entre a reconstrução projectiva e euclidiana, está, exactamente, na questão do conhecimento de k_a . Neste caso, a reconstrução obtida é euclidiana. No entanto, se não for conhecido valor de k_a , nada saberemos sobre a métrica entre os pontos reconstruídos, pelo que a reconstrução é projectiva.

3.2.5 Resultados experimentais

A imagem da figura 3.4 foi adquirida pelo protótipo apresentado na figura 3.1. Dessa

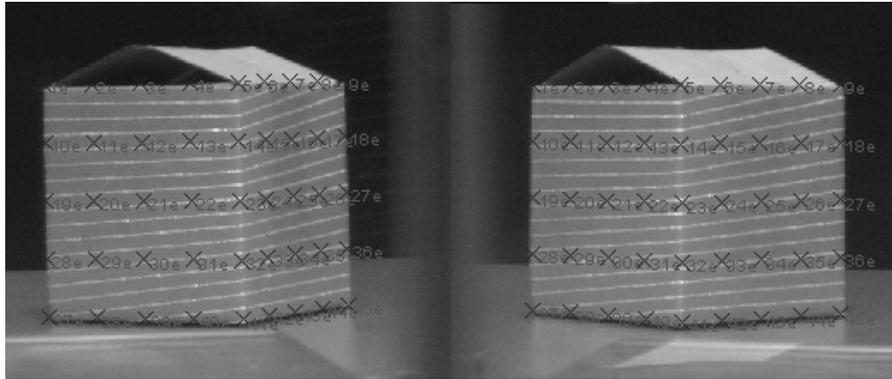


Figura 3.4: Imagem resultante da aquisição feita pelo protótipo criado, com a representação dos pontos correspondentes usados para a reconstrução.

imagem escolheram-se 45 pontos correspondentes em cada uma das sub-imagens. Este processo é válido pois estas sub-imagens, apesar de juntas, são independentes.

Usando, então, esses pontos correspondentes, representados na figura 3.4, o resultado

da estimação da matriz fundamental é

$$\mathbf{F} = \begin{bmatrix} 0 & 5.6739e-6 & -3.2698e-4 \\ 1.2927e-5 & 0 & -9.2957e-2 \\ -3.54e-3 & 8.916e-2 & 9.9166e-1 \end{bmatrix}$$

A qualidade da matriz \mathbf{F} obtida é dada por $Q_f = 0.0191$ *pixels*, usando na estimação o método linear da análise dos valores próprios (o cálculo da qualidade da matriz fundamental encontra-se explicitado no anexo A).

Este resultado foi obtido sem nenhuma otimização do método linear utilizado, o que demonstra algumas potencialidades para este processo de reconstrução tridimensional. De notar, também, que os sinais dos elementos da matriz resultante estão conforme os da equação (3.2). No entanto, um sinal de instabilidade surge com a tentativa de confirmar o valor de c_y a partir de \mathbf{F} . A média dos valores obtidos é de $c_y = 198$, enquanto que o verdadeiro valor é de $c_y = 288$.

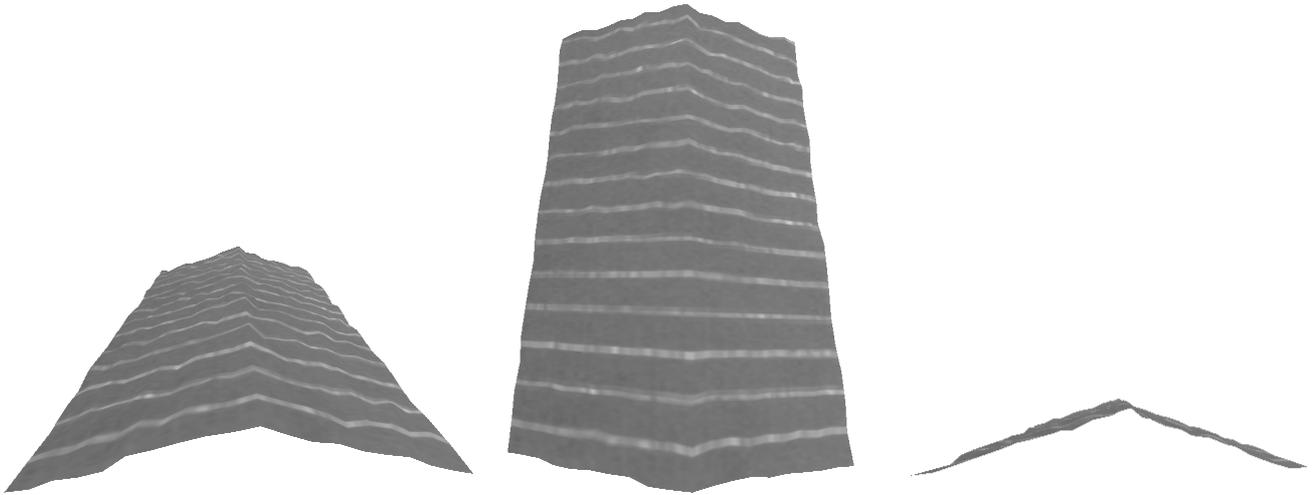


Figura 3.5: Resultado do método de reconstrução aplicado ao objecto representado na figura 3.4. Este resultado é obtido a partir da informação contida nas imagens e respectiva geometria das câmaras virtuais.

Os resultados da reconstrução tridimensional dos pontos representados na figura 3.4 aparecem na figura 3.5. Para uma maior clareza dos resultados, mostra-se a ligação entre os pontos reconstruídos. Além dos dados apresentados na tabela 3.1, $\phi = 34.43^\circ$, $B = 69.7cm$

e $k_a = -75.83$. A reconstrução das faces da caixa, reforça a validade do processo estudado, apesar deste apresentar algumas falhas.

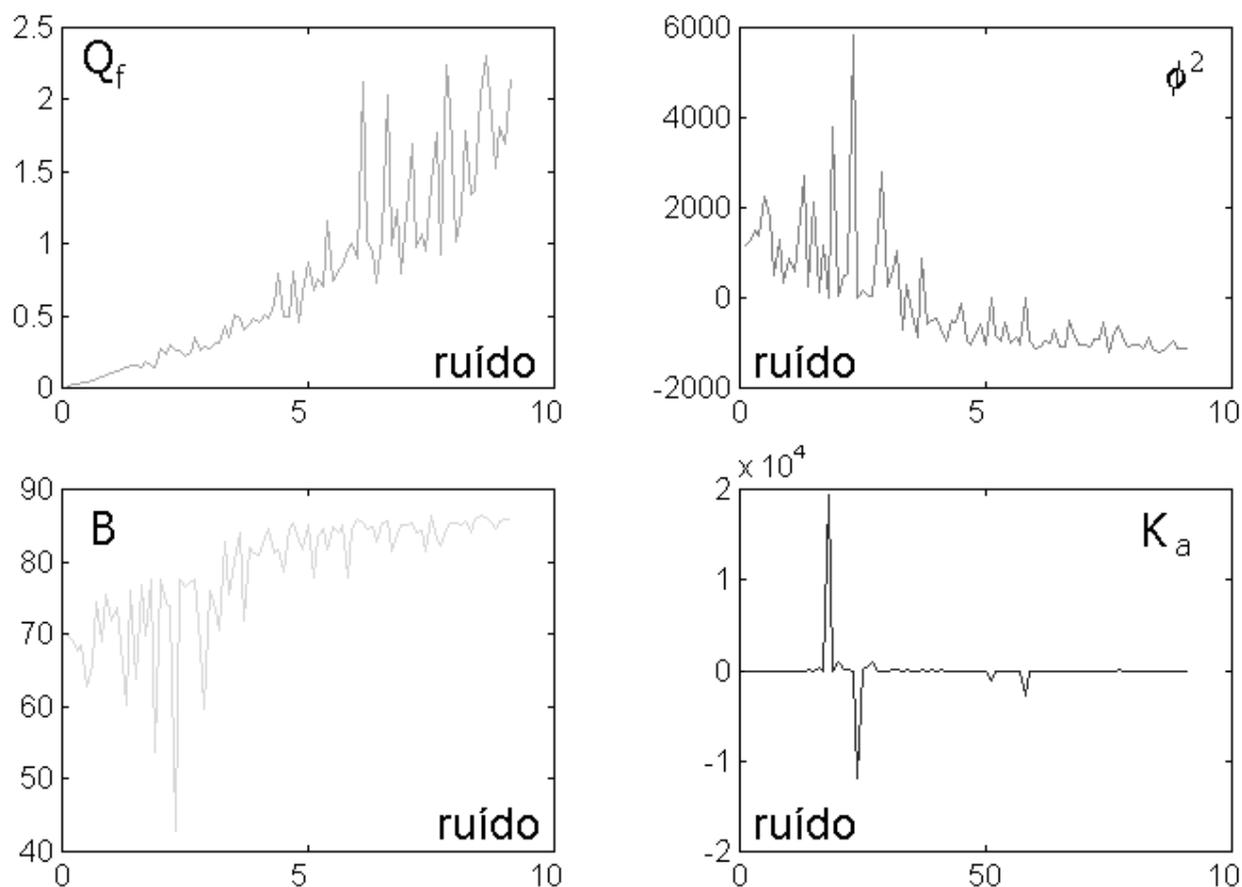


Figura 3.6: Dados sobre a estabilidade da aproximação (a).

Numa tentativa de concluir sobre a estabilidade deste processo, começou-se por introduzir ruído com uma distribuição normal gaussiana, de média zero e variância um, nos pontos correspondentes usados no cálculo de \mathbf{F} . Dos quadros da figura 3.6 retiram-se as conclusões que se apresentam de seguida.

Com o incremento do ruído, a qualidade da matriz fundamental, Q_f , deteriora-se, como já se esperava. Quanto ao valor de B , devido à sua enorme dependência de ϕ , só estabiliza quando ϕ também o faz. Por seu lado, o ângulo ϕ é o que sofre mais com a instabilidade da matriz fundamental. A razão é que, quando é calculada a diferença entre os elementos f_{12} e f_{21} , de \mathbf{F} , por vezes o resultado é negativo, levando a que o valor de ϕ seja um número

complexo e o processo falhe. Quanto a k_a , os seus valores não são tão instáveis pois, apesar de também dependerem dos elementos f_{12} e f_{21} , de \mathbf{F} , não têm operações de diferença no seu seio.

Daqui se conclui que, para que este processo seja mais robusto, o método de estimação da matriz fundamental tem que ter em conta os sinais e os valores dos seus elementos. Este facto leva-nos até à ideia dos métodos não lineares para a estimação da matriz fundamental.

Os dados exemplificativos que referimos, resumem a conclusão de todos os testes feitos ao processo. Apesar de válido, e até promissor em certos aspectos, ele sofre de instabilidade. Essa falta de robustez advém de todo o processo se apoiar na matriz fundamental, e, como se disse anteriormente, ela ser bastante sensível a ruídos quando estimada por métodos lineares.

Como esta dissertação não pretende investigar outros métodos de estimação da matriz fundamental que não os lineares, e como o incremento de robustez deste processo necessita de métodos não lineares para melhorar a estimação da matriz fundamental [Zhang 96] [Torr 96], então avançou-se para outra aproximação.

3.3 Reconstrução centrada nos espelhos

Nesta aproximação, designada na figura 1.1 por (b), mantiveram-se as grandes linhas em relação ao método anterior. Isto é, continuaram a ser usadas as imagens estéreo, as transformações projectivas (homografias), a geometria epipolar (matriz fundamental) e a triangulação.

A primeira diferença entre as duas aproximações aparece na maneira de multiplicar a única câmara usada, já que, em vez de usar quatro espelhos, como acontecia na reconstrução centrada na matriz fundamental, usa apenas dois, como se pode comprovar pela figura 3.7.

A correspondência continua a ser atingida por intermédio da matriz fundamental ou das homografias. Porém, a partir de agora, avança-se para uma gestão implícita da singularidade da matriz fundamental, \mathbf{F} , através de uma formulação simples, ao invés da imposição dessa singularidade *a posteriori* (aplicada na aproximação anterior). Esta forma de estimação de \mathbf{F} , como será mostrado, conduz ao seu bom condicionamento e estabilidade, restaurando as virtudes dos métodos lineares afastados por causa da sua instabilidade. A estimação da matriz \mathbf{F} baseia-se, então, na relação existente entre a matriz fundamental e



Figura 3.7: *Uma perspectiva do sistema utilizado.*

uma homografia. Com o uso dos espelhos, consegue-se obter a matriz fundamental apenas com um par de pontos em cada imagem.

A recuperação dos pontos tridimensionais, para esta aproximação, apoia-se na triangulação e no posicionamento dos espelhos. Classicamente, a triangulação passa pela obtenção de todos os parâmetros de quatro planos, cuja intersecção dá o ponto 3D desejado (cada dois planos formam a recta que passa pelo centro óptico de uma câmara e pelo ponto tridimensional a recuperar). O posicionamento dos espelhos, adoptado nesta aproximação, simplifica esta tarefa pois facilita o cálculo dos centros ópticos das câmaras, bem como o conhecimento do posicionamento de certos pontos tridimensionais. Toda esta informação permite a obtenção das rectas projectantes que passam pelo ponto 3D, que se pretende recuperar, e pelos centros ópticos das câmaras. Por consequência, recupera-se o próprio ponto 3D.

3.3.1 O sistema de aquisição

Como se constata pela figura 3.7, usamos dois espelhos atrás da cena que se pretende captar e uma única câmara. Os espelhos ficam sempre assentes no plano XY , podendo o ângulo de abertura entre eles ser controlado. Esta geometria, estudada na secção 2.6.1, é semelhante à apresentada na figura 2.12. No entanto, a imagem resultante em vez de só conter as projecções das reflexões dos espelhos, também contém a projecção da cena 3D. Esta situação, criada devido à aproximação dos espelhos à cena, servirá posteriormente

para a simplificação do problema da correspondência. Assim, com esta configuração para os espelhos, capta-se uma única imagem que corresponde à junção de três imagens distintas, tal como se pode conferir na figura 3.8.

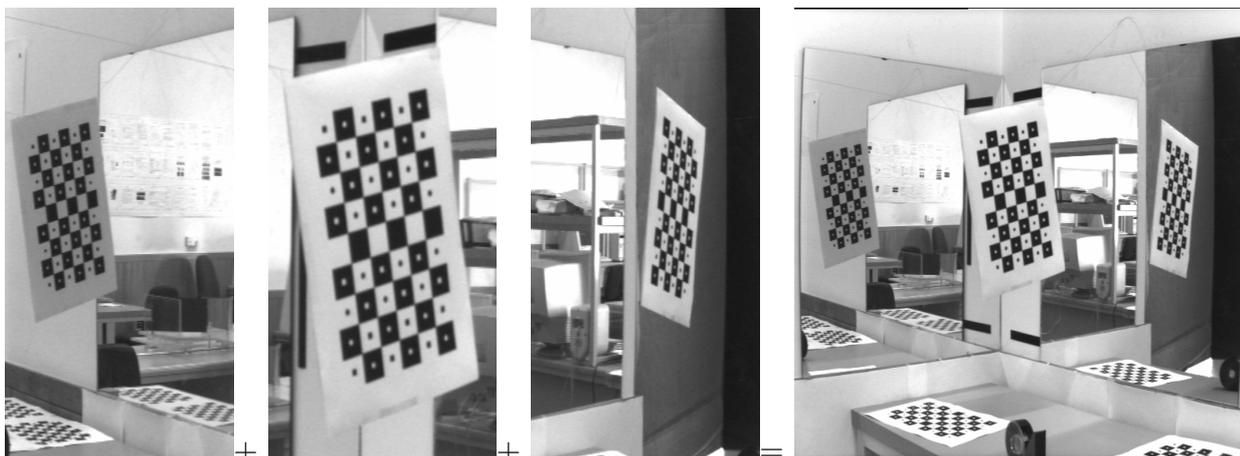


Figura 3.8: *Divisão da imagem captada pelo sistema de aquisição.*

A situação criada simula a geometria triocular, pois temos uma câmara real e duas câmaras virtuais (criadas pelos dois espelhos) a observarem a mesma cena de três pontos de vista diferentes. No entanto, a geometria apresentada tem grandes vantagens em relação à geometria triocular que usa três câmaras reais. É que, apesar dessas três câmaras reais poderem ser idênticas, os seus parâmetros intrínsecos raramente são iguais. Já no que respeita à geometria escolhida, os parâmetros intrínsecos das câmaras virtuais são exactamente os mesmos, pois os espelhos reproduzem integralmente a única câmara real existente. Outra vantagem refere-se ao histograma de intensidades das três imagens obtidas por cada câmara. Na situação criada, temos a certeza de que todos os histogramas são semelhantes, o que dificilmente acontece no caso da geometria multi-ocular. Apesar da importância destes benefícios, a maior vantagem está no conteúdo da própria imagem resultante pois, como se verá na próxima secção, facilita muito a obtenção dos epipolos. Em qualquer outra geometria multi-ocular, uma das tarefas mais difíceis é a da determinação exacta da posição dos epipolos.

Apesar de todo este potencial, o problema da ocultação não é eliminado. Este problema consiste na existência de zonas ocultas, por sobreposição ou mesmo por desaparecimento de certas partes da cena nas diferentes imagens, tiradas de diferentes pontos de vista. Nesta

aproximação o problema da ocultação pode ser minorado através do controlo do ângulo entre os espelhos, de forma a dar a todas as câmaras (real ou virtuais) um idêntico campo de visão.

3.3.2 Correspondência

Como este processo de reconstrução tem por base o método da triangulação, à semelhança do que aconteceu na aproximação anterior, torna-se imprescindível conhecer as projecções do mesmo ponto 3D. Para isso, recorre-se à estimação de uma colineação, \mathbf{H} , quando se pretende recuperar a estrutura de planos simples, ou à matriz fundamental, \mathbf{F} , quando se pretende reconstruir alguma estrutura mais complexa.

A grande diferença da reconstrução centrada nos espelhos em relação à aproximação estudada na secção anterior tem a ver com a obtenção da matriz fundamental. Em vez de a estimar explicitamente, na reconstrução que agora se descreve, começa-se por procurar a posição de um epípolo numa das imagens e depois, estima-se uma homografia de um plano entre as duas imagens. A partir desses conhecimentos, e usando as equações (2.48) ou (2.49), obtém-se a matriz fundamental. Ao processo de obtenção dos epípolos chamamos método dos epípolos por construção.

Se a estimação da homografia passar pela criação de uma base projectiva, para cada imagem, o número de parâmetros que a compõem será reduzido, como se mostrará posteriormente. Com esta mudança astuciosa do ponto de referência nas imagens e com o uso do potencial dos espelhos, aplicado à obtenção dos epípolos, mostra-se que é possível reduzir para dois o número de pontos correspondentes, em cada imagem, necessários à estimação da matriz fundamental.

Como é óbvio, se apenas se pretender reconstruir planos simples, a mudança do ponto de referência nas imagens não traz qualquer vantagem, pois continua a ser necessário pelo menos quatro pontos correspondentes para estimar a homografia. Sendo assim, só deverá ser feita a simplificação do cálculo da homografia se for pretendido recuperar estruturas complexas, não planares.

O procedimento para a estimação da matriz fundamental, pode ainda ser optimizado, quando se conhecem pelo menos oito pares de pontos correspondentes. Nestas circunstâncias, não é necessário fazer-se o cálculo da homografia. Aplicando o algoritmo da minimização algébrica, descrito no capítulo anterior, e com a utilização desses pontos juntamente

com o epipolo conhecido, obtêm-se, simultaneamente, a matriz fundamental e uma homografia.

A estimação da matriz fundamental, aqui proposta, apresenta a vantagem de uma gestão implícita da sua característica. Através de uma formulação simples, e apesar da redução do número de pontos a usar, atinge-se um sistema bem condicionado. A estabilidade dos resultados restaura parcialmente as virtudes do método linear, afastado devido aos resultados obtidos pela aproximação anterior.

De seguida será apresentado o aspecto geométrico da obtenção dos epipolos e, posteriormente, será explanada também a metodologia algébrica para a estimação da matriz fundamental.

Aspecto geométrico do método dos epipolos por construção

A explicação do aspecto geométrico do método dos epipolos por construção é, obviamente, igual tanto para a obtenção do epipolo esquerdo como para o direito. Assim, faz-se apenas a explanação referente à determinação do epipolo direito.

Seja π um plano do espaço tridimensional, definido por três pontos não colineares P_1 , P_2 e P_3 . Além disso, considerem-se os pontos P_4 e P , pertencentes ao mesmo espaço vectorial, mas não coplanares com os três pontos precedentes (ver figura 3.9). Esse conjunto de pontos é observado por duas câmaras, cujos centros de projecção são respectivamente $C_{v'}$ e C_v , originando uma imagem esquerda e uma imagem direita.

Continuando a observar a figura 3.9, a recta $\overline{C_{v'}P_4}$ corta o plano π no ponto Q_4 , enquanto que a recta $\overline{C_{v'}P}$ corta o plano π no ponto Q . Os pontos P_4 e Q_4 projectam-se na imagem esquerda no mesmo ponto p_{4e} , enquanto que na imagem direita projectam-se em dois pontos diferentes, p_{4d} e q_{4d} , respectivamente. Isto é conhecido como efeito de paralaxe, e é verdadeiro quando o movimento que liga as duas câmaras comporta uma componente não nula da translação. De uma forma similar, os dois pontos P e Q são projectados na imagem esquerda num único ponto p_e e na imagem direita em dois pontos diferentes, p_d e q_d .

Na imagem direita, as duas rectas $\overline{p_{4d}q_{4d}}$ e $\overline{p_dq_d}$, conhecidas como rectas paralaxe, são as projecções das duas rectas tridimensionais $\overline{C_{v'}P_4}$ e $\overline{C_{v'}P}$, respectivamente. Por definição, $\overline{p_{4d}q_{4d}}$ e $\overline{p_dq_d}$ são as rectas epipolares correspondentes aos dois pontos p_{4e} e p_e da imagem esquerda. Assim, como se pode ver pela figura 3.9, a intersecção destas duas rectas dá o

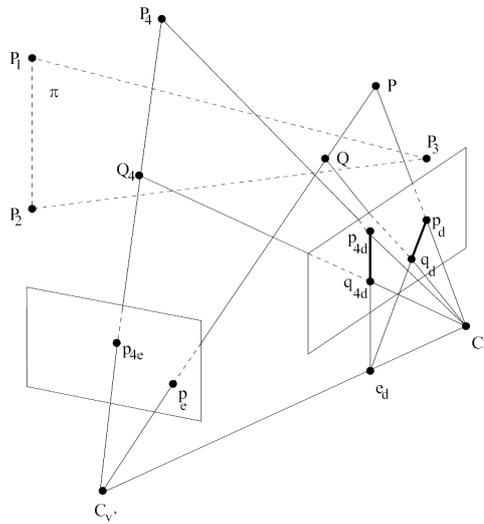


Figura 3.9: Construção geométrica do epipolo da imagem direita.

epipolo pretendido.

Contudo, fica por resolver um problema que é o do desconhecimento das localizações de q_{4d} e q_d na imagem direita. A colocação da geometria epipolar tal como descrita até aqui não é uma novidade em si, no entanto o modo de obtenção dos pontos q_{4d} e q_d é original. A determinação desses pontos é facilmente conseguida com o apoio dos espelhos, como se verá de seguida.

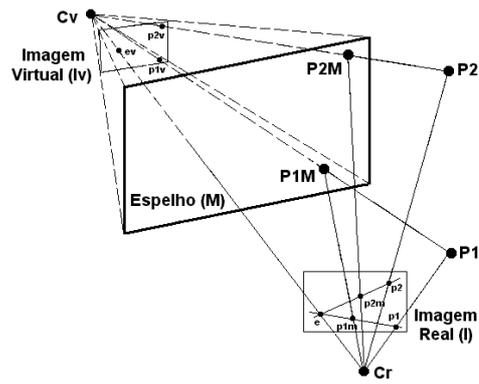


Figura 3.10: Visão tridimensional da utilização do espelho na obtenção dos pontos q_{4d} e q_d .

Considere-se, então, a figura 3.10, na qual se representa, graficamente, a configuração

usada para a aquisição das imagens (ver figura 3.7). Na figura 3.10 verifica-se que os pontos 3D, P_1 e P_2 , são reflectidos nos espelhos M , gerando os pontos P_{1M} e P_{2M} , respectivamente. Por sua vez, estes pontos são captados pela imagem I , originando as projecções p_1, p_2, p_{1m} e p_{2m} . O centro óptico da imagem I é o ponto C_r , enquanto que o centro óptico da imagem virtual (criado pelo espelho M) é o ponto C_v .

Por definição, os pontos P_1, C_r e C_v formam o plano epipolar π_p , associado a P_1 . Logo, como P_{1M} pertence ao raio projectivo de P_1 , também pertence ao plano π_p . Assim, a recta formada por p_1 e por p_{1m} é uma recta epipolar, ou, por outras palavras, uma recta de paralaxe. Como se pode observar, o mesmo se passa para os pontos P_2, P_{2M}, p_2 e p_{2m} . A intersecção destas duas rectas origina o epipolo e .

Nesta explicação geométrica, verifica-se que aos pontos P_{1M} e P_{2M} correspondem os pontos Q e Q_4 da figura 3.9, enquanto que a P_1 e P_2 correspondem os pontos P e P_4 .

Solução algébrica

Como se sabe, o cálculo da matriz fundamental é feito tendo por base uma das equações (2.48) ou (2.49). Como tal, é fácil perceber a necessidade do conhecimento de um epipolo e de uma homografia que relacione as projecções de um plano da cena 3D, nas duas imagens. Assim, para uma melhor explanação, divide-se o método que aqui se propõe para atingir a correspondência, em três fases: a da obtenção do epipolo por construção, a da estimação da homografia e a da estimação da matriz fundamental.

De forma a simplificar a explicação, denomina-se, relativamente à figura 3.8, as imagens da reflexão dos espelhos esquerdo e direito como imagens esquerda e direita, respectivamente. A imagem da cena 3D que não é a reflexão de nenhum espelho, será referenciada como imagem central. Os pontos de cada uma destas partes da imagem serão distinguidos pela primeira letra da parte a que pertencem, isto é, se forem da parte esquerda, central ou direita terão como índice, respectivamente, a letra e, c ou d.

A fase da obtenção do epipolo, começa pela observação da imagem não calibrada da figura 3.8, que se considera conter apenas dois epipolos. A razão para esta situação advém do facto de que, ao se considerarem as verdadeiras projecções da cena 3D (e não as projecções das reflexões) como parte integrante no cálculo dos epipolos, está-se a definir que a imagem adquirida é composta por apenas duas imagens distintas (esquerda e direita). Assim, para se obter cada epipolo terão que se usar os pontos da parte central da imagem e os pontos

da parte vista pela câmara virtual correspondente.

Analiticamente, o cálculo de cada epipolo é o mesmo, pelo que mostra-se apenas como se obtém o epipolo da imagem da parte direita. Seja P um ponto 3D, genérico, e p_c e p_d as suas respectivas projecções na parte central e direita da única imagem captada. A projecção p_c é obtida pela câmara real. Por seu lado, a projecção p_d é obtida pela câmara virtual direita, ou seja, é a projecção da reflexão de P no espelho direito, visto pela câmara real.

A recta formada pelos pontos p_c e p_d é dada por

$$(p_d \wedge p_c)^T q = 0 \quad (3.4)$$

com q um ponto genérico da recta. Tendo um conjunto de n pontos correspondentes, com $n \geq 2$, nestas duas imagens, pode-se formar um sistema linear $\mathbf{Ax} = \mathbf{b}$, com n equações e duas incógnitas. Da solução deste sistema resultam as coordenadas do epipolo desejado. Para se obter esta solução matemática utiliza-se um dos métodos lineares apresentados na secção 2.7. Outra maneira de concluir que a única solução do sistema criado só pode ser o epipolo direito advém da comparação das equações (2.44) e (3.4).

<i>Caso</i>	<i>Número de pontos usados</i>	<i>Epipolo esquerdo</i>	<i>Epipolo direito</i>
1	2	(-6974.3,-871.3)	(1910.2,-801.0)
2	4	(-2225.0,-098.0)	(1801.6,-707.3)
3	8	(-2959.7,-224.6)	(1629.0,-616.2)
4	8	(-0878.6,-591.2)	(0186.2,-525.3)

Tabela 3.2: *Resultados do cálculo dos epipolos. Os primeiros três casos usam o método dos epipolos por construção. O quarto caso estima a matriz fundamental, obtendo, de seguida, o seu espaço nulo.*

Na tabela 3.2 podem ser vistos os epipolos obtidos para quatro casos diferentes. Para os primeiros três casos desta tabela, os resultados foram obtidos detectando manualmente alguns pontos correspondentes em cada parte da imagem captada. Da união dos pontos da parte esquerda e central da imagem captada, surgiram as rectas que, ao serem intersectadas, deram origem ao epipolo esquerdo. Da mesma forma, o epipolo direito surge da intersecção das rectas obtidas com a união dos pontos da parte direita e central da imagem captada. Na última situação recorre-se ao método clássico, ou seja, estima-se primeiro a matriz fundamental e depois calcula-se o seu espaço nulo, obtendo-se, finalmente, o epipolo.

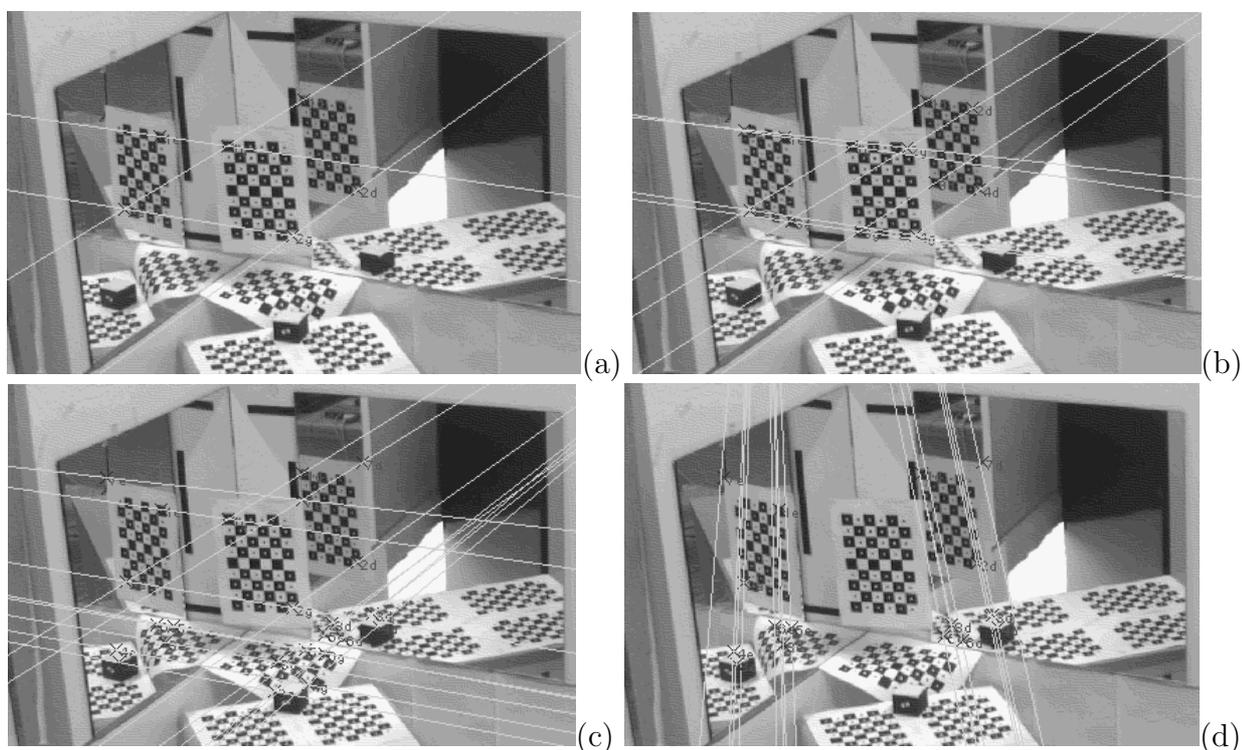


Figura 3.11: *Rectas formadas para a obtenção dos epipolos, através da sua intersecção. (a) Caso onde se usam apenas dois pontos; (b) Caso onde se usam quatro pontos; (c) Caso onde se usam oito pontos; (d) Rectas epipolares obtidas através da matriz fundamental, estimada pelo algoritmo dos oito pontos.*

A figura 3.11 mostra as rectas epipolares que deram origem aos resultados apresentados na tabela 3.2. A figura 3.11(a) usa apenas um par de pontos, em cada uma das partes da imagem captada, e corresponde ao primeiro caso apresentado na tabela 3.2. A figura 3.11(b), com quatro pontos em cada uma das partes da imagem captada, corresponde, por sua vez, ao segundo caso apresentado na tabela 3.2, e assim sucessivamente. As duas últimas situações são tratadas com oito pontos, também correspondentes, em cada uma das partes da imagem captada.

A imagem adquirida, mostrada na figura 3.11, tem um pormenor que convém ser esclarecido. No momento da aquisição, os espelhos encontram-se ambos assentes no plano XY e estão perpendiculares entre si. Isto implica que, nas figuras 3.11(a), 3.11(b) e 3.11(c), alguns pontos da parte direita e central da imagem captada só podem ser correspondentes se for considerado que a folha não tem espessura, isto é que a folha representa um plano

transparente.

Numa primeira comparação entre o método dos epipolos por construção e o método clássico, torna-se óbvio que o primeiro necessita apenas de dois pontos correspondentes em cada imagem (caso 1 da tabela 3.2), enquanto que o outro método precisa de oito (caso 4 da mesma tabela). O número de pontos correspondentes é muito importante porque afecta a sua sensibilidade ao ruído. Não esquecer que, quando o número de pontos for maior que o necessário, o sistema torna-se sobre-dimensionado, e qualquer pequeno engano nesses pontos afecta a solução final, pois esta é obtida através de um processo de minimização linear.

Como o nosso método é um processo gráfico puro, a sensibilidade ao ruído torna-se mais suave que no método clássico, mas não desaparece. Analisando os três primeiros casos da tabela 3.2, nota-se que, com o incremento do número de pontos usados o ruído é reduzido, tornando os resultados mais estáveis. Comparando os dois últimos casos, verifica-se uma grande diferença entre os valores das coordenadas dos epipolos. Este facto acontece devido à estimação de \mathbf{F} ter sido feita com o mínimo de pontos correspondentes possível, o que eleva grandemente a sua sensibilidade ao ruído. É, no entanto, muito importante focar que o sinal das coordenadas dos epipolos é o mesmo para todos os casos.

Apesar deste exemplo e das suas conclusões, ele não confirma, inequivocamente, a validade do método, pois não se consegue ver a projecção do centro óptico do espelho. Assim, o melhor exemplo que se pode dar, no sentido de confirmar definitivamente a validade deste método é que põe a câmara real numa posição tal, que o espelho fica com a sua reflexão. Neste caso, como se pode confirmar pela figura 3.12, o epipolo tem que estar na parte da imagem que contém a projecção da reflexão da câmara no espelho. Nessa figura fez-se apenas a aquisição da cena 3D e a sua reflexão no espelho esquerdo, ignorando o espelho direito. Na obtenção do epipolo foram escolhidos manualmente alguns pontos correspondentes, sem qualquer critério definido. Cada recta é formada por uma projecção da parte central e uma da parte esquerda da imagem captada. As coordenadas do epipolo foram determinadas aplicando o RANSAC à intersecção destas rectas.

No que diz respeito à fase da estimação da homografia, lembre-se a necessidade do conhecimento de, pelo menos, quatro pares de pontos correspondentes.

Recuperando os três pontos P_1 , P_2 e P_3 , que definem o plano π , da figura 3.9, considere-se que p_{1e} , p_{2e} e p_{3e} são as suas projecções na parte esquerda da imagem e p_{1d} , p_{2d} e p_{3d} na parte direita da imagem. Sabendo que três pontos não colineares do espaço tridimensional

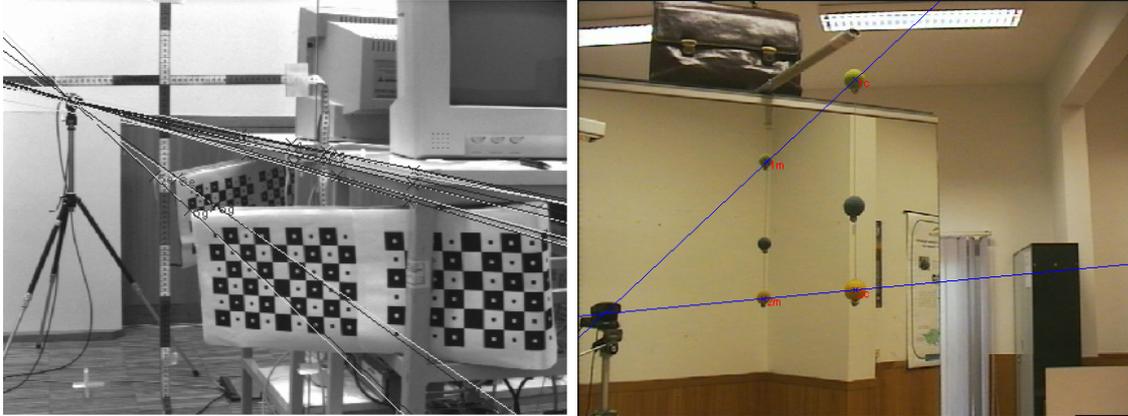


Figura 3.12: Exemplos de obtenção do epipolo pelo método dos epipolos por construção.

definem sempre um plano, é evidente que estas projecções verificam a relação (2.38).

As equações obtidas através de (2.38) podem ser simplificadas, efectuando uma mudança de base para cada um dos dois conjuntos de projecções (parte esquerda e direita da imagem). Essa simplificação é feita passando das coordenadas *pixels* usuais, às coordenadas projectivas definidas numa base canónica do espaço projectivo de dimensão 2. Para tal, efectua-se uma transformação projectiva do plano que afecta as coordenadas

$$p_a = \begin{bmatrix} \tau & \tau & \tau \end{bmatrix}^T \quad p_b = \begin{bmatrix} \alpha & 0 & 0 \end{bmatrix}^T \quad p_c = \begin{bmatrix} 0 & \beta & 0 \end{bmatrix}^T \quad p_d = \begin{bmatrix} 0 & 0 & \gamma \end{bmatrix}^T$$

aos pontos da parte esquerda da imagem, p_{0e} , p_{1e} , p_{2e} e p_{3e} , respectivamente, com p_{0e} um ponto qualquer nessa parte. Para a parte direita da imagem, procede-se da mesma maneira. O ponto p_{0d} , da imagem direita, não tem que ser, contudo, o correspondente de p_{0e} . É preciso notar que os quatro pontos, em cada uma das duas imagens, não devem ser, três a três, colineares.

Como as homografias são invertíveis, podemos então efectuar todos os cálculos com as coordenadas projectivas novas e depois voltar às coordenadas *pixels* iniciais. Assim, com a escolha das novas coordenadas nas duas partes da imagem, e utilizando a relação (2.38) para os três pares de pontos correspondentes, (p_{1e}, p_{1d}) , (p_{2e}, p_{2d}) e (p_{3e}, p_{3d}) , a matriz \mathbf{H} fica reduzida à forma

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

Esta matriz tem três parâmetros, mas como todas as colineações são definidas a menos de um factor de escala, somente dois deles são independentes. Para além disso, estes três parâmetros não podem ser nulos, sob pena de \mathbf{H} não ser invertível, mas sim uma homografia singular que transformaria um plano numa recta. Por consequência, podemos fixar um dos três parâmetros com o valor 1, por exemplo o parâmetro γ . Desta forma a matriz fica reduzida a

$$\mathbf{H} = \begin{bmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3.5)$$

Antes de avançar, convém ver o comportamento das homografias estimadas a partir da equação (2.38). Para a estimação das homografias usa-se o método numérico da análise dos valores próprios.

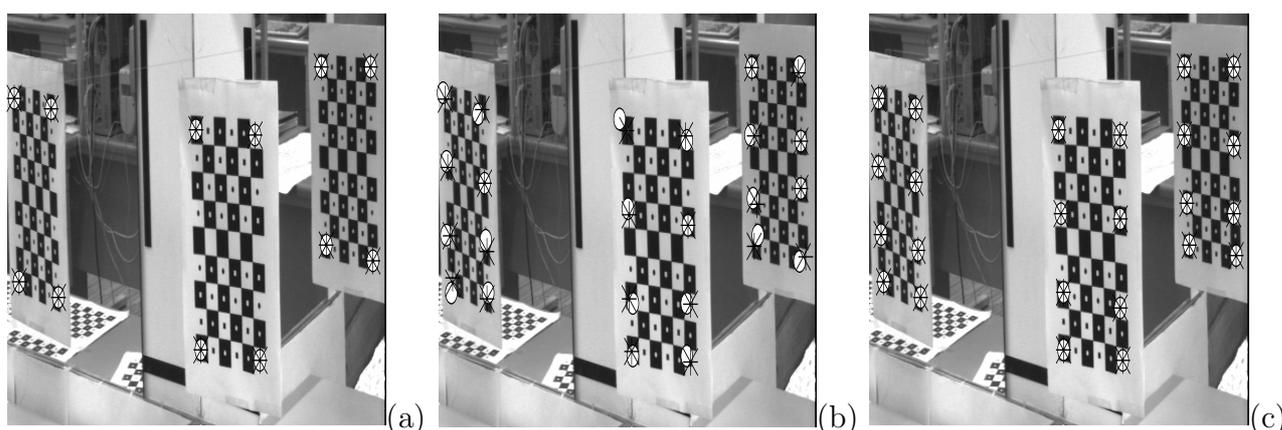


Figura 3.13: *Representação dos pontos usados e obtidos na estimação das homografias: (a) Caso dos quatro pontos normalizados; (b) Caso dos oito pontos não normalizados; (c) Caso dos oito pontos normalizados.*

A imagem mostrada na figura 3.13 representa uma folha sem espessura, ou seja, uma superfície planar perfeita. Sobre essa imagem, considere-se três situações de cálculo da homografia. Numa primeira, usa-se o mínimo de pontos correspondentes necessários à estimação, entre as três partes da imagem captada (figura 3.13(a)). Numa segunda situação, utilizam-se oito trios de pontos correspondentes não normalizados (figura 3.13(b)). Finalmente, numa terceira situação, os oito trios de pontos são normalizados antes de se proceder à estimação da matriz homográfica (figura 3.13(c)).

Uma vez que a imagem da figura 3.13 é a junção de três sub-imagens, existem três homografias que fazem a correspondência entre o plano tridimensional projectado em cada uma dessas sub-imagens. Sendo assim, os pontos marcados com \times são os usados nas estimações, enquanto que os pontos marcados com $+$ e com \circ são os que resultam da aplicação dessas homografias (cada ponto, de cada parte da imagem capturada, pode ser obtido recorrendo a duas homografias diferentes). Para conhecê-las todas, basta estimar apenas duas delas. A razão desta afirmação é simples e prova-se da seguinte maneira. Como todas as homografias são invertíveis, a partir da equação (2.38) tem-se

$$\mathbf{p}_e = \mathbf{H}_{ec}\mathbf{p}_c \Leftrightarrow \mathbf{p}_c = \mathbf{H}_{ec}^{-1}\mathbf{p}_e \quad \mathbf{p}_d = \mathbf{H}_{dc}\mathbf{p}_c \quad (3.6)$$

substituindo a primeira equação de (3.6) na segunda, obtém-se o pretendido, ou seja, a terceira homografia, dada por

$$\mathbf{p}_d = \mathbf{H}_{de}\mathbf{p}_e = \mathbf{H}_{dc}\mathbf{H}_{ec}^{-1}\mathbf{p}_e \quad (3.7)$$

<i>No. de pontos</i>	<i>Pontos normalizados</i>	\mathbf{H}_{ec}	\mathbf{H}_{dc}	\mathbf{H}_{ed}	\mathbf{H}_{eda}
<i>Distância média entre todos os pontos e as suas reprojeções</i>					
4	Sim	1.93 pixels	1.27 pixels	1.29 pixels	1.29 pixels
8	Não	4.89 pixels	1.85 pixels	1.35 pixels	3.32 pixels
8	Sim	2.01 pixels	1.43 pixels	1.23 pixels	1.19 pixels
<i>Desvio padrão da distância entre todos os pontos e as suas reprojeções</i>					
4	Sim	2.43 pixels	1.63 pixels	1.42 pixels	1.42 pixels
8	Não	2.68 pixels	0.76 pixels	0.75 pixels	1.71 pixels
8	Sim	0.65 pixels	0.69 pixels	0.57 pixels	0.49 pixels

Tabela 3.3: Resultados da estimação das homografias. A matriz \mathbf{H}_{ed} foi estimada com a primeira equação inserida em 3.7, enquanto que \mathbf{H}_{eda} foi obtida através da segunda equação inserida em 3.7.

De forma a complementar as conclusões sobre a estimação das homografias, considere-se também a tabela 3.3, a qual apresenta alguns dos resultados da estimação de todas as homografias envolvidas na figura 3.13. Analisando os seus resultados, nota-se que com a

normalização dos dados de entrada, os resultados tornam-se melhores (o segundo caso da tabela 3.3 tem uma média e um desvio padrão da soma das distâncias entre todos os pontos e as suas projecções, piores que o terceiro caso). Outro facto que resulta da comparação da primeira e última situações da tabela 3.3, é que o uso de mais pontos correspondentes correctos (sem ruído) melhora o desempenho da correspondência (o desvio padrão do último caso é o melhor). Assim, conclui-se que o problema dos métodos numéricos, isto é, a sua sensibilidade aos ruídos existentes nos dados de entrada, permanece.

Existe ainda algo de interessante a considerar na tabela 3.3, mais especificamente nas suas duas últimas colunas. A coluna da matriz \mathbf{H}_{ed} contém os dados obtidos a partir da equação (2.38), enquanto que a coluna da matriz \mathbf{H}_{eda} contém os dados obtidos a partir da equação (3.7). Como se pode verificar, os resultados de \mathbf{H}_{ed} só são melhores que os de \mathbf{H}_{eda} para a situação dos pontos não normalizados, sendo semelhantes quando é feita a operação de normalização. A razão para este facto prende-se com a propagação dos erros. Quando uma matriz é estimada, ela contém quase sempre alguns erros (vistos como sensibilidade da matriz). Ao multiplicarmos duas dessas matrizes, o resultado virá afectado com o erro de ambas. No entanto, depois dos dados de entrada serem normalizados, o erro que eles carregam quase desaparece, levando a que a sua propagação quase não tenha efeitos. Conclui-se, então, que a operação de normalização é muito importante para o cálculo das homografias.

Finalmente, passemos à fase da estimação da matriz fundamental. Suponha-se que o epipolo na imagem direita, em coordenadas homogéneas, é dado por $e_d(e_x, e_y, e_c)$. Sendo \mathbf{H} a matriz, 3×3 , que descreve a homografia escrita sob a forma (3.5), a matriz fundamental procurada, \mathbf{F} , dada pela equação (2.48), pode ser descrita por

$$\mathbf{F} = \begin{bmatrix} 0 & e_c & -e_y \\ -e_c & 0 & e_x \\ e_y & -e_x & 0 \end{bmatrix} \begin{bmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & e_c\beta & -e_y \\ -e_c\alpha & 0 & e_x \\ e_y\alpha & -e_x\beta & 0 \end{bmatrix} \quad (3.8)$$

Substituindo a equação (3.8) na equação (2.10), obtém-se

$$\mathbf{p}_d^T \mathbf{F} \mathbf{p}_e = 0 \Leftrightarrow \begin{bmatrix} a_d & b_d & c_d \end{bmatrix} \begin{bmatrix} 0 & e_c\beta & -e_y \\ -e_c\alpha & 0 & e_x \\ e_y\alpha & -e_x\beta & 0 \end{bmatrix} \begin{bmatrix} a_e \\ b_e \\ c_e \end{bmatrix} = 0$$

ou, de uma outra forma,

$$a_e(c_d e_y - b_d e_c)\alpha + b_e(a_d e_c - c_d e_x)\beta + c_e(e_x b_d - e_y a_d) = 0 \quad (3.9)$$

É de notar que os pontos das imagens, \mathbf{p}_e e \mathbf{p}_d , e o epipolo, \mathbf{e} , têm que passar pela mesma mudança de base por que passaram aquando da estimação da homografia. Daí a necessidade de se considerar as três componentes das suas coordenadas, pois a sua coordenada homogénea pode ser transformada em zero (o que não quer dizer que o ponto esteja no infinito). Se, em vez do epipolo direito, fosse conhecido o esquerdo, todo o processo seria feito da mesma forma, com a excepção de se usar a equação (2.49) em vez da (2.48).

Mesmo que a cena seja coplanar (todos os pontos pertencem ao plano π), a equação (3.9) continuará válida. Neste caso particular, qualquer um dos pontos da imagem pode ser o epipolo. Contudo, a homografia \mathbf{H} continua a verificar a restrição imposta por (2.38). Isto garante uma coerência ao nível da singularidade da homografia, mesmo com esta degeneração.

Na equação (3.9) só os parâmetros α e β são desconhecidos. Como tal, esta equação pode ser transformada em

$$\begin{bmatrix} a_e(c_d e_y - b_d e_c) & b_e(a_d e_c - c_d e_x) \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} c_e(e_y a_d - e_x b_d) \end{bmatrix} \quad (3.10)$$

Se tivermos n pontos correspondentes, com $n \geq 2$, usando a equação (3.10) poderemos construir um conjunto de equações lineares, ficando então com um sistema de equações do tipo $\mathbf{Ax} = \mathbf{b}$. A estimação da matriz fundamental, \mathbf{F} , e, automaticamente, da homografia, \mathbf{H} , fica concluída, com a obtenção da solução desse sistema através de um qualquer método numérico.

Por outro lado, se tivermos n pontos correspondentes, com $n \geq 8$, poderemos obter resultados similares. Nessas circunstâncias, basta aplicar ao epipolo descoberto e ao conjunto dos n pontos, o algoritmo da minimização algébrica sem iteração, para obter a matriz fundamental e uma homografia, entre as várias possíveis. O facto da matriz \mathbf{G} , da equação (2.31), não ter característica igual ao número de colunas impede que esta homografia seja única.

A matriz \mathbf{F} , assim definida, verifica a equação (2.10). Para além disso, ela é, por construção, uma matriz de característica 2 (o determinante de \mathbf{F} é sempre nulo), sendo, por consequência, formalmente a matriz fundamental que descreve a geometria epipolar entre a parte esquerda e a parte direita da imagem.

Voltando à imagem da figura 3.11, a tabela 3.4 apresenta alguns dos resultados da estimação das matrizes fundamentais que lhe dizem respeito. Relembra-se que a parte central

No. pontos	Matrizes Fundamentais			Qualidade de F	α e β	Epípolos (Espaço nulo de F)	
2	$F_{ec} =$	0	-1.0	-871.3	$Q_{fec} = 2.0985$	$\alpha_{ec} = 1.0$ $\beta_{ec} = 1.0$	$\mathbf{e}_e = (-6974.3, -871.3)$
		1.0	0	6974.3			
	$F_{dc} =$	871.3	-6974.3	0	$Q_{fdc} = 1.3938$	$\alpha_{dc} = 1.0$ $\beta_{dc} = 1.0$	$\mathbf{e}_d = (2107.2, -942.3)$
		0	-1.0	-942.3			
4	$F_{ec} =$	0	-1.1	-486.5	$Q_{fec} = 1.4482$	$\alpha_{ec} = 1.0584$ $\beta_{ec} = 1.0106$	$\mathbf{e}_e = (-4228.4, -459.7)$
		1.0	0	4273.0			
	$F_{dc} =$	459.7	-4228.4	0	$Q_{fdc} = 1.1366$	$\alpha_{dc} = 1.0172$ $\beta_{dc} = 0.9793$	$\mathbf{e}_d = (1730.3, -723.7)$
		0	-1.0	-708.8			
8	$F_{ec} =$	0	-1.0	-231.0	$Q_{fec} = 0.7201$	$\alpha_{ec} = 1.0281$ $\beta_{ec} = 1.005$	$\mathbf{e}_e = (-2959.7, -224.6)$
		1.0	0	2974.6			
	$F_{dc} =$	224.6	-2959.7	0	$Q_{fdc} = 1.1567$	$\alpha_{dc} = 1.0072$ $\beta_{dc} = 0.9909$	$\mathbf{e}_d = (1617.3, -621.8)$
		0	-1.0	-616.2			
		1.0	0	-1629.7			
		620.6	1614.2	0			

Tabela 3.4: Resultados da estimação das matrizes fundamentais. A qualidade de F e as coordenadas dos epípolos estão em pixels. Os epípolos são obtidos através do cálculo do espaço nulo de F .

da imagem, no cálculo dos epípolos, é considerada como pertencente a ambas as partes esquerda e direita, e portanto, isso implica a existência de apenas duas matrizes fundamentais. Através do factor de qualidade da matriz F , confirma-se a propriedade subjacente aos métodos numéricos, segundo a qual, com o incremento de pontos correspondentes correctos conseguem-se melhores resultados. Comparando a tabela 3.2 com a tabela 3.4, verifica-se que as coordenadas dos epípolos são semelhantes (a situação ideal seria a igualdade, o que aconteceria se fosse aplicado o algoritmo da minimização algébrica). Apesar deste processo de estimação ser mais robusto do que o da aproximação centrada na matriz fundamental, continua sensível ao ruído. A solução para reduzir este problema passa pela filtragem do ruído, do conjunto de dados de entrada, com o RANSAC, aplicado à estimação de α e β .

A figura 3.14 permite visualizar a aplicação das matrizes fundamentais obtidas, isto é, mostra as rectas epipolares. Comparando-a com a figura 3.11, esclarece-se o facto evidente de que a geometria epipolar foi estimada com base nos epípolos.

3.3.3 Reconstrução tridimensional dos pontos

A fase de reconstrução começa com a definição do sistema de coordenadas do mundo. Como os espelhos são planares, assume-se, sem perda de generalidade, que o plano XZ contém o espelho esquerdo e que o eixo Z é perpendicular ao plano onde assentam ambos

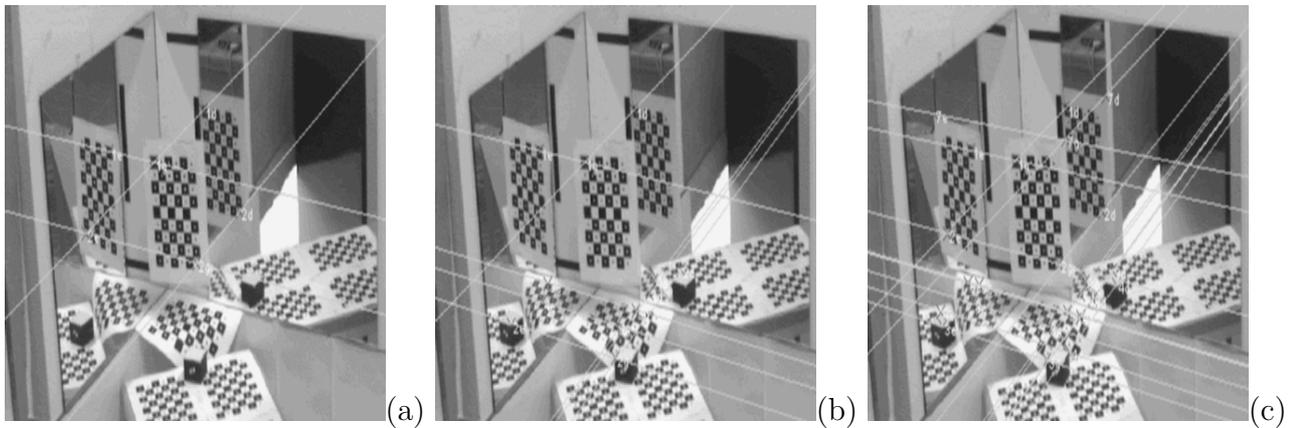


Figura 3.14: *Rectas epipolares resultantes da estimação das matrizes fundamentais: (a) Estimação com dois pontos; (b) Estimação com quatro pontos; (c) Estimação com oito pontos.*

os espelhos, tal como mostra a figura 3.15. Conforme for a melhor situação para o processo

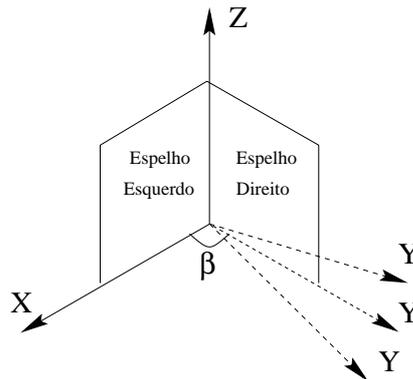


Figura 3.15: *Sistema de coordenadas do sistema de aquisição e reconstrução.*

de reconstrução, o eixo X , pode ser qualquer recta 3D paralela à recta que passa pela base do espelho esquerdo. Por fim, o eixo Y posiciona-se conforme a obrigatoriedade da perpendicularidade existente para com os eixos X e Z . Isto faz com que o eixo Y possa estar para lá do espelho direito, sobre o espelho direito ou aquém do espelho direito, porque, como se vê na figura 3.15, a abertura entre os dois espelhos é sempre fixa (para cada processo de reconstrução), e igual a β .

Conforme o que foi dito, a imagem adquirida pode ser considerada como a junção das duas imagens captadas pelas câmaras virtuais, posicionadas atrás dos espelhos. Assim, o

processo de reconstrução baseia-se na intersecção geométrica dos raios de luz que passam pelos centros ópticos dessas câmaras virtuais e pelo ponto 3D genérico, P , a recuperar.

Denominando o raio de luz genérico que passa pelo centro óptico C_v , da câmara virtual esquerda, como recta (tridimensional) esquerda, e o raio de luz genérico que passa pelo centro óptico $C_{v'}$, da câmara virtual direita, como recta (tridimensional) direita, a equação dessas rectas, ambas restringidas a passar pelo ponto a recuperar, pode ser dada por

$$P = C_i + \lambda \frac{(P_i - C_i)}{\|P_i - C_i\|} \quad (3.11)$$

com $i \in \{v, v'\}$ e $\lambda \neq 0$. Como se constata da equação (3.11), além dos centros ópticos de cada câmara virtual, precisa-se também de conhecer os pontos tridimensionais P_e e P_d , pertencentes às respectivas rectas, esquerda e direita, para se atingir a recuperação de P . Da equação anterior também se consegue perceber que, por cada nova reconstrução são, necessários novos pontos P_e e P_d . Os pontos P_e e P_d não são quaisquer pontos das rectas, tendo, obviamente, uma relação que facilita o processo de reconstrução. Veremos posteriormente qual é essa relação e em que é que ela simplifica a recuperação de qualquer ponto P .

Para facilitar a explanação, a fase de reconstrução dos pontos tridimensionais é dividida em três partes, a da obtenção dos pontos P_e e P_d (cálculo dos pontos do plano espelho), a da obtenção dos centros ópticos virtuais C_v e $C_{v'}$ (cálculo dos centros ópticos) e, por fim, a da obtenção do ponto genérico P (recuperação de pontos genéricos).

Cálculo dos pontos do plano espelho

Os conhecimentos até este momento resumem-se ao das projecções que estão na imagem não calibrada e ao de que ambos os espelhos planares podem ser considerados como planos tridimensionais. Assim, para a obtenção genérica dos pontos tridimensionais P_e e P_d , irá ser usada uma matriz de transformação, \mathbf{T} , que relaciona as projecções da imagem com pontos euclidianos de um plano do sistema de coordenadas do mundo.

À primeira vista, a relação (2.7) fornece a possibilidade de se obter a matriz \mathbf{T} . No entanto, a partir da forma como está apresentada, não se conseguem determinar os pontos 3D desejados pois no sistema de equações resultante existem três incógnitas para duas equações. De forma a ultrapassar este problema surgiu a ideia de aproveitar o facto do espelho esquerdo estar no plano XZ e, a partir daí, obter os pontos genéricos P_e e P_d .

Neste contexto, como se tem dois espelhos, terão de ser criadas duas matrizes de transformação, que relacionem matematicamente os pontos tridimensionais dos espelhos com as suas correspondentes projecções na imagem.

O ponto P_e corresponde à intersecção do raio projectivo que chega à câmara virtual esquerda, e que passa pelo ponto a recuperar, com o plano espelho esquerdo. Como tal, para a obtenção deste ponto genérico considera-se $P_e = (X_e, 0, Z_e)$, pois o espelho esquerdo, ao estar contido no plano XZ (ver figura 3.15), faz com que todas as coordenadas dos pontos que lhe pertencem tenham ordenada igual a zero. Considere-se, também, que o ponto P_e é projectado na imagem resultante do sistema de aquisição no ponto $p_e = (x_e, y_e)$. A substituição deste conhecimento na relação (2.7) origina

$$k \begin{bmatrix} x_e \\ y_e \\ 1 \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & t_{13} & t_{14} \\ t_{21} & t_{22} & t_{23} & t_{24} \\ t_{31} & t_{32} & t_{33} & t_{34} \end{bmatrix} \begin{bmatrix} X_e \\ 0 \\ Z_e \\ 1 \end{bmatrix} \Leftrightarrow k\mathbf{p}_e = \begin{bmatrix} t_{11} & t_{13} & t_{14} \\ t_{21} & t_{23} & t_{24} \\ t_{31} & t_{33} & t_{34} \end{bmatrix} \begin{bmatrix} X_e \\ Z_e \\ 1 \end{bmatrix}$$

ou, de forma mais reduzida,

$$k\mathbf{p}_e = \mathbf{T}_{xz} \begin{bmatrix} X_e \\ Z_e \\ 1 \end{bmatrix}$$

onde k é um factor multiplicativo qualquer, diferente de zero.

Com esta simplificação conseguiu-se fazer com que a matriz de transformação se tornasse 3×3 e, automaticamente, invertível. Desta forma, todos os pontos P_e determinam-se apenas com as projecções das reflexões na imagem, através de

$$k \begin{bmatrix} X_e \\ Z_e \\ 1 \end{bmatrix} = \mathbf{T}_{xz}^{-1} \mathbf{p}_e \quad (3.12)$$

Por seu lado, o ponto P_d corresponde à intersecção do raio projectivo que chega à câmara virtual direita, e que passa pelo ponto a recuperar, com o plano espelho direito. Devido ao facto do espelho direito poder estar sobre o plano YZ ou noutras posições (ver figura 3.15), a obtenção do ponto genérico P_d pode sofrer uma pequena alteração face à determinação de P_e . Sendo D a coordenada horizontal no plano que contém o espelho direito, através de

uma geometria simples obtém-se, para $0^\circ \leq \beta < 180^\circ$ (intervalo de abertura permitida),

$$\begin{cases} X_d = D \cos \beta \\ Y_d = D \sin \beta \end{cases} \Leftrightarrow \begin{cases} X_d = \frac{Y_d}{\tan \beta} \\ D = \frac{Y_d}{\sin \beta} \end{cases} \quad (3.13)$$

substituindo (3.13) na relação (2.7) e considerando que $p_d = (x_d, y_d)$ corresponde à projecção do ponto P_d na imagem, fica-se com

$$k \begin{bmatrix} x_d \\ y_d \\ 1 \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & t_{13} & t_{14} \\ t_{21} & t_{22} & t_{23} & t_{24} \\ t_{31} & t_{32} & t_{33} & t_{34} \end{bmatrix} \begin{bmatrix} \frac{Y_d}{\tan \beta} \\ Y_d \\ Z_d \\ 1 \end{bmatrix} \Leftrightarrow k \mathbf{p}_d = \begin{bmatrix} \frac{t_{11}}{\tan \beta} + t_{12} & t_{13} & t_{14} \\ \frac{t_{21}}{\tan \beta} + t_{22} & t_{23} & t_{24} \\ \frac{t_{31}}{\tan \beta} + t_{32} & t_{33} & t_{34} \end{bmatrix} \begin{bmatrix} Y_d \\ Z_d \\ 1 \end{bmatrix}$$

ou, de forma mais reduzida,

$$k \mathbf{p}_d = \mathbf{T}_{yz} \begin{bmatrix} Y_d \\ Z_d \\ 1 \end{bmatrix}$$

onde k é um factor multiplicativo qualquer, diferente de zero.

Tal como para P_e , todos os pontos P_d determinam-se através de

$$k \begin{bmatrix} Y_d \\ Z_d \\ 1 \end{bmatrix} = \mathbf{T}_{yz}^{-1} \mathbf{p}_d \quad (3.14)$$

com apenas com as projecções das reflexões do espelho direito na imagem.

Como as homografias \mathbf{T}_{xz} e \mathbf{T}_{yz} são matrizes 3×3 , definidas a menos de um factor de escala, a sua estimação necessita, no mínimo, de um conjunto de quatro pontos tridimensionais do respectivo plano, bem como das suas correspondentes projecções 2D, cujos pontos genéricos definiremos por (x_m, y_m) . A respectiva estimação destas matrizes passa pela resolução dos sistemas $\mathbf{A} \mathbf{t} \mathbf{x} \mathbf{z} = \mathbf{0}$ e $\mathbf{A} \mathbf{t} \mathbf{y} \mathbf{z} = \mathbf{0}$, aplicando um qualquer método numérico, com

$$\mathbf{t} \mathbf{x} \mathbf{z} = \left[t_{11} \quad t_{13} \quad t_{14} \quad t_{21} \quad t_{23} \quad t_{24} \quad t_{31} \quad t_{33} \quad t_{34} \right]^T$$

$$\mathbf{t} \mathbf{y} \mathbf{z} = \left[\frac{t_{11}}{\tan \beta} + t_{12} \quad t_{13} \quad t_{14} \quad \frac{t_{21}}{\tan \beta} + t_{22} \quad t_{23} \quad t_{24} \quad \frac{t_{31}}{\tan \beta} + t_{32} \quad t_{33} \quad t_{34} \right]^T$$

Para cada um dos sistema de equações acima mencionado, a matriz \mathbf{A} é composta pelas respectivas sub-matrizes $\mathbf{A} \mathbf{x} \mathbf{z}$ e $\mathbf{A} \mathbf{y} \mathbf{z}$. Estas são criadas a partir de cada par de pontos

correspondentes, e são dadas por

$$\mathbf{Axz} = \begin{bmatrix} X_M & Z_M & 1 & 0 & 0 & 0 & -X_M x_m & -Z_M x_m - x_m \\ 0 & 0 & 0 & X_m & Z_M & 1 & -X_M y_m & -Z_M y_m - y_m \end{bmatrix}$$

e

$$\mathbf{Ayz} = \begin{bmatrix} Y_M & Z_M & 1 & 0 & 0 & 0 & -Y_M x_m & -Z_M x_m - x_m \\ 0 & 0 & 0 & Y_m & Z_M & 1 & -Y_M y_m & -Z_M y_m - y_m \end{bmatrix}$$

Utilizando marcas naturais (como por exemplo, os cantos dos espelhos) ou marcas artificiais colocadas nos espelhos, conseguimos saber o posicionamento das projecções dessas marcas na imagem (ou seja, os pontos p_e e p_d). Por outro lado, se essas marcas forem medidas em relação à origem do sistema de coordenadas do mundo (que se situa nos espelhos), obtemos também os pontos P_e e P_d , e juntamos a tão desejada métrica euclidiana no processo de reconstrução. As marcas deverão ser tais que tornem o processo natural, isto é, não sejam necessárias tarefas adicionais à reconstrução. Por isso, como se conhecem as medidas dos espelhos, as marcas relacionadas com esse facto são as preferenciais.

O número de pares de pontos correspondentes, necessários à estimação, pode ser optimizado, ou seja, diminuído, se forem usados os mesmos pontos, tanto na equação (3.12) como na equação (3.14). Se os espelhos esquerdo e direito estiverem juntos, lado a lado, todos os pontos pertencentes à linha de junção dos espelhos têm essa capacidade. Apesar deste facto, não se podem escolher todos os pontos 3D necessários dessa linha pois isso degeneraria as soluções obtidas para as transformações \mathbf{T}_{xz} e \mathbf{T}_{yz} . No máximo só se poderão escolher três pontos pertencentes à linha de junção dos espelhos, o que implica que para as duas transformações bastará conhecer apenas cinco pares de pontos diferentes.

Cálculo dos centros ópticos

Através das projecções de dois pontos 3D na imagem e das matrizes de transformação \mathbf{T}_{xz} e \mathbf{T}_{yz} , obtidas no ponto anterior, obtêm-se as posições tridimensionais correspondentes a essas projecções, nos planos dos espelhos esquerdo e direito. Se, a estes pontos dos espelhos, lhes juntarmos as posições tridimensionais reais das projecções usadas, podemos formar rectas tridimensionais. Como cada espelho simboliza uma câmara (virtual), teremos duas rectas 3D por cada câmara virtual, pois os dois pontos conhecidos do mundo real originam dois pontos em cada espelho. Cada recta tridimensional obtida é um raio projectante que

chega à câmara virtual. Como se sabe, a intersecção de dois raios projectantes que chegam à mesma câmara tem como resultado o centro óptico dessa mesma câmara.

Embora correcto, este raciocínio teórico tem na sua aplicação prática alguns problemas relevantes. Sendo assim, passemos à descrição do procedimento prático utilizado no cálculo destes centros ópticos. Como a maneira de os obter é a mesma para qualquer das duas câmaras virtuais, a descrição irá ser feita de forma genérica, sem especificar se o cálculo é para a câmara virtual esquerda ou direita.

A execução começa com a marcação e a medição de dois pontos da cena 3D, P_{1a} e P_{2a} , não pertencentes aos espelhos. Após a aquisição da imagem não calibrada da cena, fica-se a conhecer as projecções correspondentes a esses pontos. Multiplicando essas projecções pela inversa da matriz de transformação (\mathbf{T}_{xz} ou \mathbf{T}_{yz} , conforme o caso), obtêm-se as reflexões no espelho dos pontos marcados na cena, respectivamente, P_{1b} e P_{2b} . Com estes pontos tridimensionais definem-se as rectas

$$r_1 \equiv P_{1a} - P + \lambda \frac{P_{1b} - P_{1a}}{\|P_{1b} - P_{1a}\|} = 0 \quad r_2 \equiv P_{2a} - P + \lambda \frac{P_{2b} - P_{2a}}{\|P_{2b} - P_{2a}\|} = 0$$

A intersecção destas rectas dará o centro óptico virtual. É aqui que podem surgir problemas.

Quando se fala em intersecção de duas rectas tridimensionais, primeiro há que verificar qual a relação entre elas. Assim, podem existir as seguintes possibilidades: as rectas não são paralelas e não se intersectam, as rectas intersectam-se, as rectas são paralelas, mas não coincidem e, finalmente, as rectas são coincidentes.

Teoricamente, só a segunda possibilidade é que deveríamos ter neste processo de reconstrução. No entanto, basta haver algum erro de medição ou na escolha das projecções na imagem, para que outras possibilidades aconteçam. Na prática, a probabilidade de acontecer alguma das duas últimas possibilidades é quase nula porque, além da escolha das projecções ser feita manualmente, só uma medição muito incorrecta daria esses resultados. Entre as quatro hipóteses referidas, a primeira situação pode acontecer frequentemente, bastando para isso pequenos erros na escolha das projecções (por exemplo, em vez de escolher o ponto certo na imagem de 768×576 *pixels*, escolhe-se o ponto que está seis ou sete *pixels* mais afastado). Por tudo isto, o procedimento tem de fazer uma verificação da relação entre as rectas encontradas, e, a partir daí, se for possível, ultrapassar os erros que possam existir.

Sendo P_{1a} e P_{1b} os pontos pertencentes a uma das rectas (a qual denominaremos por recta 1), e P_{2a} e P_{2b} os pontos pertencentes à outra recta (que designaremos por recta 2),

o procedimento segue com a formação dos vectores paralelos a essas rectas, dados por

$$\vec{r}_1 = P_{1a}\vec{P}_{1b} \qquad \vec{r}_2 = P_{2a}\vec{P}_{2b}$$

e, também, com a formação do vector \vec{u} que une um ponto da recta 1 a um ponto da recta 2. Com os dados que possuímos, esse vector pode ser obtido através de qualquer uma das formas seguintes

$$\vec{u} = P_{1a}\vec{P}_{2a} \qquad \vec{u} = P_{1a}\vec{P}_{2b} \qquad \vec{u} = P_{1b}\vec{P}_{2a} \qquad \vec{u} = P_{1b}\vec{P}_{2b}$$

Para o procedimento, não faz qualquer diferença a escolha da forma de obtenção do vector \vec{u} .

O vector perpendicular a ambas as rectas 1 e 2 e, também, paralelo à recta que passa pelo ponto onde as rectas 1 e 2 se encontram mais próximas pode ser formado através de

$$\vec{v} = \vec{r}_1 \wedge \vec{r}_2$$

onde \wedge é o operador de produto vectorial.

Se o vector \vec{v} for nulo, então, as rectas 1 e 2 são paralelas. Esta situação representa as terceira e quarta possibilidades das relações entre rectas acima referidas, que, como se disse, são devidas a erros grosseiros. Se durante um processo de reconstrução este caso ($\vec{v} = 0$) acontecer, o procedimento pára imediatamente, de forma a poder ser reformulado algum valor das medições.

Por outro lado, se $\vec{v} \neq 0$, o procedimento continua, com o cálculo da distância mínima entre as rectas 1 e 2, através de

$$d_m = \frac{|\vec{u} \cdot \vec{v}|}{|\vec{v}|}$$

Se d_m for zero, então as rectas intersectam-se. Se d_m for diferente de zero, então as rectas não se intersectam. A forma de ultrapassar esta última situação, já que ela resulta, obviamente, de um erro, consiste em encontrar o ponto médio entre os pontos das rectas 1 e 2 mais próximos um do outro. Esta solução também se adequa ao caso das rectas se intersectarem porque, sendo $d_m = 0$, o ponto médio entre um ponto e ele próprio é exactamente esse mesmo ponto.

Nesta fase do cálculo ($\vec{v} \neq 0$) define-se o plano π_1 , o qual contém a recta 1, através de

$$\pi_1 \equiv \det \begin{bmatrix} \vec{r}_1 \\ \vec{v} \\ P\vec{P}_{1a} \end{bmatrix} = 0$$

com P um ponto genérico do plano π_1 . Em vez do ponto P_{1a} poderia ser usado o ponto P_{1b} sem que isso alterasse o procedimento.

Da mesma forma, define-se π_2 , o plano que contém a recta 2, através de

$$\pi_2 \equiv \det \begin{bmatrix} \vec{r}_2 \\ \vec{v} \\ P\vec{P}_{2a} \end{bmatrix} = 0$$

com P um ponto genérico do plano π_2 . Também aqui, poderia ter sido usado o ponto P_{2b} sem que isso modificasse o procedimento.

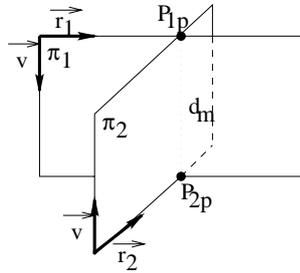


Figura 3.16: Aspecto gráfico do método de intersecção entre duas rectas não paralelas e não coincidentes.

Da intersecção da recta 1 com o plano π_2 surge o ponto P_{1p} , enquanto que o ponto P_{2p} resulta da intersecção da recta 2 com o plano π_1 . Estes pontos estão, entre si, à distância d_m , tal como se confirma pela figura 3.16, e pertencem, respectivamente, às rectas 1 e 2.

Para se obter a solução matemática destas intersecções, terá que se ultrapassar o problema do sobre-dimensionamento do sistema de equações resultante, pois este sistema contém quatro equações (a equação do plano correspondente e três equações relativas à recta 3D) para três incógnitas (as coordenadas X , Y e Z do ponto de intersecção). Para resolver este problema deverá ser usado um método linear (por exemplo o método dos valores próprios ou o SVD) que encontre a solução óptima do sistema.

Por fim, o centro óptico correcto, ou o melhor dentro da possível situação de erro, é dado por

$$C_o = \frac{P_{2i} - P_{1i}}{2}$$

Uma vez obtidos os centros ópticos virtuais, torna-se fácil obter o centro óptico da câmara real. Este dado é fisicamente bastante difícil de ser confirmado, pois qualquer

centro óptico encontra-se posicionado no interior da câmara. No entanto, o valor obtido para o centro óptico da câmara real, trará mais uma confirmação sobre a correcção do método, pois indica a zona onde a câmara real deverá estar posicionada. Poder-se-ia tirar mais partido do centro óptico real, neste processo de reconstrução, se fosse conhecido um plano tridimensional, paralelo ou coincidente ao plano imagem, com o qual se pudessem relacionar as projecções da imagem, relativas apenas à cena 3D.

A forma de obtenção do centro óptico, na prática, é descrita de seguida. Conforme foi dito atrás, um epipolo é a projecção do centro óptico da câmara na imagem, pelo que através das equações (3.12) e (3.14) podemos obter as posições tridimensionais nos espelhos (P_{Me} e P_{Md}) dos epipolos obtidos anteriormente por construção (ver subsecção anterior). Com estes pontos e com os centros ópticos das câmaras virtuais podem-se formar duas rectas, cuja intersecção é o centro óptico da câmara real, C_r . A determinação desta nova informação terá que passar pelo método que foi descrito para solucionar o problema ligado à intersecção de rectas tridimensionais. Os centros ópticos das câmaras virtuais são os pontos 3D que, como vimos, é necessário conhecer.

Outra forma de cálculo do centro óptico da câmara real assenta no facto dos centros ópticos real e virtual esquerdo pertencerem à mesma recta 3D, a qual contém o ponto P_{Me} . Esta restrição associada ao posicionamento do sistema de eixos, mostrado na figura 3.15, e ao facto da equação do plano espelho ser $Y = 0$, impõe que as coordenadas X e Z de C_r terão que ser as mesmas que P_{Me} . Segundo Sameer [Sameer 98], os centros ópticos virtuais estão à mesma distância do espelho que o centro óptico real. Logo, a coordenada Y de C_r é a inversa da coordenada Y do centro óptico virtual esquerdo.

Obtidos os centros ópticos o sistema fica calibrado. Além disso, mantém-se calibrado enquanto não se alterarem as posições dos espelhos e da câmara real. Assim, com uma calibração apenas, pode-se fazer o número de reconstruções que se desejar. Para isso, basta alterar a cena 3D, adquirir nova imagem e fazer o processo descrito no próximo ponto para todas as projecções das reflexões que se vêm em ambos os espelhos.

Recuperação de pontos genéricos

Com a intersecção das rectas que passam pelos pontos genéricos P_e e P_d (correspondentes ao ponto a recuperar) e pelos centros ópticos C_v e $C_{v'}$, pode-se recuperar qualquer ponto tridimensional que não esteja oculto a pelo menos uma das câmaras virtuais.

Sendo as coordenadas do ponto a recuperar $P(X, Y, Z)$, as coordenadas dos centros ópticos das câmaras virtuais $C_v = (C_{vx}, C_{vy}, C_{vz})$ e $C_{v'} = (C_{v'x}, C_{v'y}, C_{v'z})$, e as coordenadas dos pontos dos planos dos espelhos, correspondentes a $P_e(X_e, Y_e, Z_e)$, $P_d(X_d, Y_d, Z_d)$ e P , a equação (3.11) pode ser transformada em

$$\begin{bmatrix} (C_{vy} - Y_e) & (X_e - C_{vx}) & 0 \\ (C_{vz} - Z_e) & 0 & (X_e - C_{vx}) \\ (C_{v'y} - Y_d) & (X_d - C_{v'x}) & 0 \\ (C_{v'z} - Z_d) & 0 & (X_d - C_{v'x}) \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} C_{vx}Y_e - X_eC_{vy} \\ C_{vx}Z_e - X_eC_{vz} \\ C_{v'x}Y_d - X_dC_{v'y} \\ C_{v'x}Z_d - X_dC_{v'z} \end{bmatrix} \quad (3.15)$$

Aplicando qualquer método numérico à equação (3.15), recupera-se a informação perdida do ponto P . Este modo de se obter os pontos reconstruídos é uma alternativa ao método que foi usado no cálculo dos centros ópticos. Obviamente que a equação (3.15) poderia também ter sido usada na obtenção dos centros ópticos.

Outras informações importantes

É possível determinar para este sistema de aquisição mais algumas informações importantes, como as rotações e translações entre os diferentes sistemas de coordenadas (mundo, câmara real e câmaras virtuais), que podem ser de grande utilidade para o processo de reconstrução tridimensional. De notar que estas informações estão todas referenciadas ao sistema de coordenadas do mundo.

Assim, começa-se por considerar as transformações rígidas, do tipo (2.3), entre o centro óptico real e os centros ópticos virtuais,

$$C_v = \mathbf{R}_v C_r + \mathbf{t}_v = \begin{bmatrix} \mathbf{R}_v & \mathbf{t}_v \\ 0 & 1 \end{bmatrix} \begin{bmatrix} C_r \\ 1 \end{bmatrix} = \mathbf{D}_v \begin{bmatrix} C_r \\ 1 \end{bmatrix} \quad (3.16)$$

$$C_{v'} = \mathbf{R}_{v'} C_r + \mathbf{t}_{v'} = \begin{bmatrix} \mathbf{R}_{v'} & \mathbf{t}_{v'} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} C_r \\ 1 \end{bmatrix} = \mathbf{D}_{v'} \begin{bmatrix} C_r \\ 1 \end{bmatrix} \quad (3.17)$$

As matrizes \mathbf{D}_v e $\mathbf{D}_{v'}$ são conhecidas como matrizes de reflexão.

Tendo por base as equações (3.16) e (3.17), chega-se à conclusão que a orientação relativa entre os centros ópticos virtuais é dada por

$$C_{v'} = \mathbf{R}_{v'} \mathbf{R}_v^{-1} C_v + \mathbf{t}_{v'} - \mathbf{R}_{v'} \mathbf{R}_v^{-1} \mathbf{t}_v = \begin{bmatrix} \mathbf{R}_{v'} \mathbf{R}_v^{-1} & \mathbf{t}_{v'} - \mathbf{R}_{v'} \mathbf{R}_v^{-1} \mathbf{t}_v \\ 0 & 1 \end{bmatrix} \begin{bmatrix} C_v \\ 1 \end{bmatrix}$$

ou, de uma forma mais compacta, por

$$C_{v'} = \mathbf{D}_{v'} \mathbf{D}_v^{-1} \begin{bmatrix} C_v \\ 1 \end{bmatrix} = \mathbf{D} \begin{bmatrix} C_v \\ 1 \end{bmatrix} \quad (3.18)$$

Representando cada um dos espelhos pelas suas respectivas normais e relembando que a distância entre os centros ópticos virtuais e os espelhos é igual à distância entre o centro óptico real e os respectivos espelhos [Sameer 98], então, segundo [Nayar 98], a matriz de reflexão para cada um dos espelhos é dada por

$$\mathbf{D}_v = \begin{bmatrix} \mathbf{I} - 2\mathbf{n}_e \mathbf{n}_e^T & 2d_e \mathbf{n}_e \\ 0 & 1 \end{bmatrix} \quad \mathbf{D}_{v'} = \begin{bmatrix} \mathbf{I} - 2\mathbf{n}_d \mathbf{n}_d^T & 2d_d \mathbf{n}_d \\ 0 & 1 \end{bmatrix} \quad (3.19)$$

onde a matriz \mathbf{I} é a matriz identidade, as distâncias d_e e d_d são, respectivamente, as distâncias dos espelhos esquerdo e direito ao centro óptico real, \mathbf{n}_e o vector normal ao espelho esquerdo e \mathbf{n}_d o vector normal ao espelho direito.

De acordo com o que foi apresentado anteriormente, o vector normal ao espelho esquerdo é dado por $\mathbf{n}_e = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T$, pois este espelho tem como equação geral $Y = 0$, e o vector normal ao espelho direito é dado por $\mathbf{n}_d = \begin{bmatrix} 1 & -\frac{1}{\tan\beta} & 0 \end{bmatrix}^T$, devido a este espelho ter como equação geral $-X + \frac{1}{\tan\beta}Y = 0$.

Por outro lado, substituindo as expressões de (3.19) na equação (3.18) obtém-se a orientação relativa entre os planos espelho esquerdo e espelho direito, dada de uma forma compacta por

$$\mathbf{D} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix}$$

com a rotação, \mathbf{R} , dada por

$$\mathbf{R} = \mathbf{I} + 4(\mathbf{n}_e \cdot \mathbf{n}_d) \mathbf{n}_e \mathbf{n}_d^T - 2\mathbf{n}_e \mathbf{n}_e^T - 2\mathbf{n}_d \mathbf{n}_d^T \quad (3.20)$$

e a translação, \mathbf{t} , dada por

$$\mathbf{t} = 2d_e \mathbf{n}_e - [2d_e(\mathbf{n}_e \cdot \mathbf{n}_d) + 2d_d] \mathbf{n}_d$$

Como os centros ópticos real e virtuais são conhecidos, a expressão matemática das distâncias d_e e d_d é obtida através de

$$d_e = \frac{\overline{C_r C_v}}{2} = \frac{\sqrt{(C_x - C_{vx})^2 + (C_y - C_{vy})^2 + (C_z - C_{vz})^2}}{2}$$

e

$$d_d = \frac{\overline{C_r C_{v'}}}{2} = \frac{\sqrt{(C_x - C_{v'x})^2 + (C_y - C_{v'y})^2 + (C_z - C_{v'z})^2}}{2}$$

considerando que o centro óptico real é $C_r = (C_x, C_y, C_z)$, que o centro óptico virtual esquerdo é $C_v = (C_{vx}, C_{vy}, C_{vz})$ e que o centro óptico virtual direito é $C_{v'} = (C_{v'x}, C_{v'y}, C_{v'z})$.

Uma outra informação que se pode retirar da configuração geométrica deste sistema de aquisição, também importante em processos de reconstrução tridimensional, é o conhecimento de um plano de referência. Uma das possibilidades normalmente utilizada é o plano que contém os três centros ópticos. Este plano é conhecido como plano de movimento [Nayar 98]. A equação deste plano é obtida por

$$\det \begin{bmatrix} C_{vx} - C_x & C_{vy} - C_y & C_{vz} - C_z \\ C_{v'x} - C_x & C_{v'y} - C_y & C_{v'z} - C_z \\ X - C_x & Y - C_y & Z - C_z \end{bmatrix} = 0 \Leftrightarrow AX + BY + CZ + D = 0$$

A partir desta equação pode-se retirar a sua normal, que é $n = [A \ B \ C]^T$. Com esta nova informação e com base na equação (3.20), obtém-se a rotação existente entre este plano e os planos espelho esquerdo e espelho direito.

Além disto, podemos verificar se as posições dos epipolos na imagem estão correctas. Pelo que foi dito atrás, \mathbf{e}_e e \mathbf{e}_d são formados pela intersecção da *baseline* $\overline{C_v C_{v'}}$ com as respectivas imagens virtuais. Para o epipolo \mathbf{e}_e , considere-se a recta 3D que contém os pontos C_r e C_v . A sua equação é

$$\frac{X - C_x}{C_{vx} - C_x} = \frac{Y - C_y}{C_{vy} - C_y} = \frac{Z - C_z}{C_{vz} - C_z}$$

Da intersecção desta recta com o plano espelho esquerdo, cuja equação é $Y = 0$, resulta o ponto 3D

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} \frac{-C_y(C_{vx} - C_x)}{C_{vy} - C_y} + C_x \\ 0 \\ \frac{-C_y(C_{vz} - C_z)}{C_{vy} - C_y} + C_z \end{bmatrix}$$

Com este ponto, usando a equação (3.12) obtém-se as coordenadas do epipolo \mathbf{e}_e na imagem. Por seu lado, para a obtenção do epipolo \mathbf{e}_d , considere-se a recta 3D que contém os pontos C_r e $C_{v'}$. A sua equação é

$$\frac{X - C_x}{C_{v'x} - C_x} = \frac{Y - C_y}{C_{v'y} - C_y} = \frac{Z - C_z}{C_{v'z} - C_z}$$

Da intersecção desta recta com o plano espelho direito, cuja equação é $-X + \frac{1}{\tan\beta}Y = 0$, resulta o ponto 3D

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} \frac{C_x(C_{v'y} - C_y) - C_y(C_{v'x} - C_x)}{(C_{v'y} - C_y) - \tan\beta(C_{v'x} - C_x)} \\ \tan\beta X \\ \frac{(X - C_x)(C_{v'z} - C_z)}{(C_{v'x} - C_x)} + C_z \end{bmatrix}$$

Com este ponto, usando a equação (3.14) obtêm-se as coordenadas do epipolo \mathbf{e}_d na imagem. Este modo de cálculo destes epipolos simplifica o método criado em [Sameer 98], pois evita o conhecimento prévio dos parâmetros intrínsecos e extrínsecos da câmara.

3.3.4 Resultados experimentais

Experiências com imagens reais validam o nosso método simples de reconstrução, das quais se apresentam alguns resultados significativos. Antes de avançar, convém separá-los em três grupos, classificados quanto ao ângulo de abertura, β , entre os espelhos (ver figura 3.15). Com esta classificação pretende-se dar uma visão mais clara sobre as potencialidades e problemas (especificamente o da oclusão) do processo de reconstrução apresentado. Assim, os resultados do primeiro grupo são para todos os casos em que $\beta = 90^\circ$, enquanto que o segundo grupo é relativo aos casos em que $\beta < 90^\circ$. Por fim, o terceiro grupo é para os casos em que $\beta > 90^\circ$.

$$\beta = 90^\circ$$

Este grupo constitui uma situação especial. Aqui o espelho esquerdo pertence ao plano XZ e o espelho direito pertence ao plano YZ . Este último facto simplifica a matriz de transformação \mathbf{T}_{yz} pois $\frac{t_{11}}{\tan\beta} = \frac{t_{21}}{\tan\beta} = \frac{t_{31}}{\tan\beta} = 0$.

Como primeira experiência deste grupo, tentou-se apenas a reconstrução pontos pertencentes ao mesmo plano tridimensional. A figura 3.17(a) representa uma imagem não calibrada de um plano 3D. Como seria impossível ver fisicamente o plano 3D naquela posição, em ambos os espelhos, utilizou-se uma folha, na qual foi colocado um padrão. Desta forma simulou-se a transparência da folha. Os pontos representados na figura 3.17(a) através da marca \times são os que foram reconstruídos. O resultado dessa reconstrução é mostrado na figura 3.17(b), onde foi colocado o padrão do plano sobre os pontos obtidos.

Para se ficar com uma melhor visão sobre a precisão deste processo de reconstrução, na tabela 3.5 apresentam-se alguns resultados numéricos referentes à figura 3.17, como as

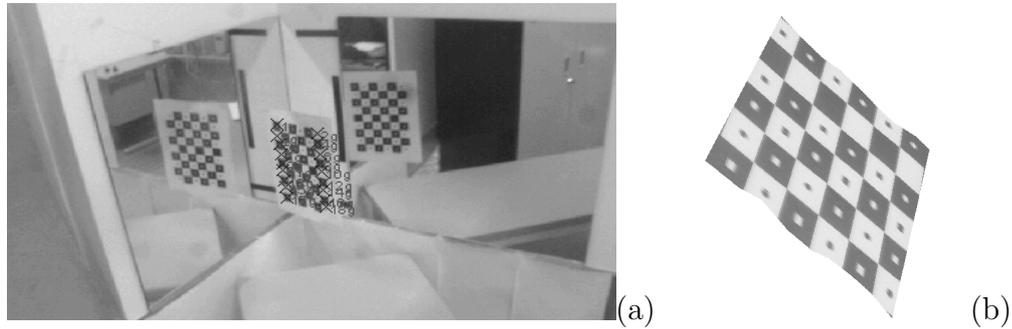


Figura 3.17: (a) Imagem não calibrada de um plano 3D, de onde se recuperaram os pontos assinalados com um \times ; (b) Resultados da reconstrução planar.

Posição na imagem	X	Y	Z	Distância entre os pontos reais e recuperados	Distância horizontal	Distância vertical
1	194.12	087.78	350.58	15.54	144.90	26.50
2	193.87	232.18	362.71	16.85	144.90	19.68
3	194.43	092.58	324.51	11.25	142.69	20.83
4	195.46	234.04	343.18	20.00	142.69	32.66
5	194.59	093.61	303.70	09.76	136.10	29.11
6	197.66	229.48	310.91	12.32	136.10	24.77
7	194.66	093.28	274.59	10.63	138.39	19.99
8	199.21	231.11	286.24	11.76	138.39	21.71
9	194.46	089.72	254.92	13.63	144.08	26.86
10	200.46	233.35	264.67	14.41	144.08	30.25
11	194.41	090.14	228.05	13.08	144.98	30.88
12	202.00	234.78	234.49	10.63	144.98	24.76
13	194.70	095.38	197.62	09.71	137.41	21.97
14	202.91	232.00	209.91	08.97	137.41	32.38
15	194.31	091.80	175.94	11.61	142.21	24.80
16	204.19	233.66	177.59	06.33	142.21	24.80
17	194.08	091.59	141.78	15.45	145.62	-
18	205.17	236.36	152.96	08.61	145.62	-

Tabela 3.5: Informação sobre os pontos recuperados do plano 3D. Os valores da sexta e sétima coluna referem-se à distância entre o ponto recuperado e os seus vizinhos de baixo e lateral, respectivamente.

coordenadas dos pontos reconstruídos (relativas à origem do sistema de coordenadas do mundo), a distância entre os pontos reais e os reconstruídos e a distância entre pontos vizinhos (na vertical e na horizontal). Estes dados foram obtidos usando os cantos dos espelhos, como pontos 3D conhecidos (pois sabia-se que o tamanho de cada espelho era de $600\text{mm} \times 600\text{mm}$), na obtenção das matrizes \mathbf{T}_{xz} e \mathbf{T}_{yz} . A fase da correspondência, entre as projecções de cada parte da imagem adquirida, foi ultrapassada recorrendo apenas às homografias, porque o que estava em jogo era apenas um plano 3D.

Da quinta coluna da tabela 3.5 resulta uma informação importante, a do erro médio entre as coordenadas dos pontos reais e reconstruídos. Neste caso, a coordenada X tem

um erro médio de $\Delta X = 12.31mm$, enquanto que a coordenada Y tem um erro médio de $\Delta Y = 4.09mm$. Por fim, o erro médio na coordenada Z é de $\Delta Z = 3.94mm$.

Como se constata na figura 3.17(a), o padrão estampado na folha é feito de rectângulos de tamanhos iguais. Os pontos escolhidos situam-se no centro de cada rectângulo, estando, por isso, todos à distância de $138mm$ em relação aos vizinhos horizontais e de $25mm$ em relação aos vizinhos verticais.

A sexta e sétima colunas da tabela 3.5 mostram que os pontos recuperados mantêm mais ou menos as suas posições, em relação aos seus vizinhos. Destas colunas resulta que a distância média entre vizinhos horizontais é de $141.82mm$, enquanto que entre os vizinhos verticais é de $25.75mm$. A razão para a diferença entre a distância exacta, na cena 3D, e a distância média entre os pontos recuperados, está no ruído introduzido aquando da escolha das projecções na imagem.

Com o conhecimento de $(192, 103, 353)$ e $(207, 228, 152)$, dois pontos 3D da cena, a estimação dos centros ópticos, em milímetros, dá os seguintes resultados

$$C_v = (1210.6, -739.0, 718.7)$$

$$C_{v'} = (-1490.1, 1094.9, 857.8)$$

$$C_r = (1082.4, 852.4, 831.4)$$

onde, recorde-se, C_v é o centro óptico da câmara virtual esquerda, $C_{v'}$ é o centro óptico da câmara virtual direita e C_r é o centro óptico da câmara real. A posição do ponto C_r , apesar de fisicamente ser muito difícil conhecê-la, coincide com a zona onde a câmara estava.

Aumentando a complexidade, passou-se para a tentativa de uma recuperação de pontos de várias superfícies planares, ambos captados na mesma imagem. A figura 3.18(a) representa essa imagem não calibrada, enquanto que o resultado da reconstrução dessa imagem é mostrado na figura 3.18(b). Sobre os pontos recuperados foram colocados os padrões que os planos tinham na cena 3D. Tal como o exemplo anterior, as transformações \mathbf{T}_{xz} e \mathbf{T}_{yz} foram estimadas através dos pontos situados nos cantos dos espelhos e respectivas projecções na imagem. No entanto, na correspondência, em vez das homografias, recorreu-se à matriz fundamental. A razão para esta mudança ficou a dever-se ao facto de os pontos a recuperar pertencerem a diferentes superfícies da cena 3D.

Repare-se, através da observação das figuras 3.17(a) e 3.18(a), que, para este grupo, os pontos das superfícies que se podem recuperar são apenas os que pertencem às superfícies perpendiculares aos espelhos. No caso limite, através do truque dos padrões colados nas

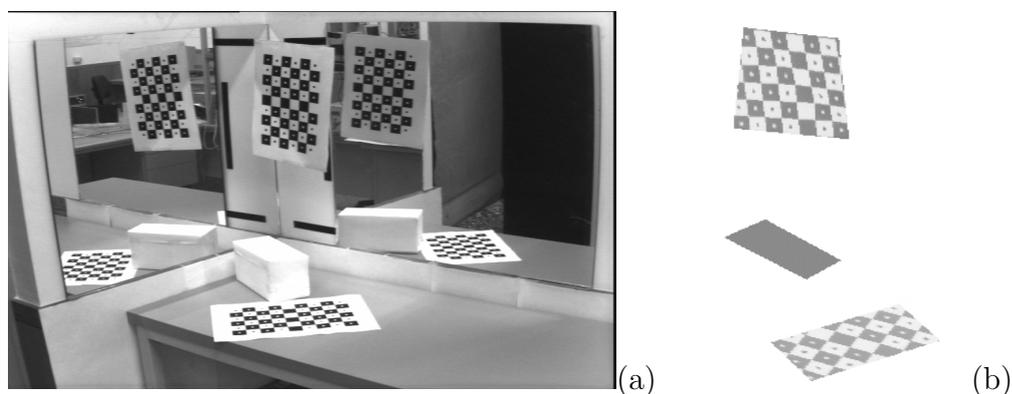


Figura 3.18: (a) Imagem não calibrada, da qual se recuperaram pontos de diferentes superfícies planares; (b) Resultados da reconstrução, onde sobre os pontos recuperados se pôs o padrão original das superfícies da cena 3D.

superfícies, podem-se reconstruir apenas os pontos das superfícies planares perpendiculares ao plano onde assentam os espelhos. A razão para este facto deve-se às câmaras virtuais terem uma *baseline* grande. A única forma de evitar este problema de oclusão é a de abrir os espelhos, ou seja, aumentar o ângulo β .

Como se pôde verificar pela tabela 3.5, em comparação com a aproximação (a), esta aproximação revela uma muito menor sensibilidade ao ruído, pois os cálculos não se apoiam fundamentalmente numa matriz.

Apesar dos bons resultados de reconstrução, a aplicação deste método é extremamente morosa, pois é feita ponto a ponto. Além disso, não consegue recuperar mais do que posições. Isto implica que, para uma melhor apresentação da cena 3D recuperada, tenha que ser feito um processamento de atribuição de cores (ou texturas) aos pontos recuperados. Juntando isto ao facto do método ser manual e ter a limitação de só poder recuperar as projecções das reflexões nos espelhos torna-se necessário o estudo de um outro método de reconstrução que ultrapasse estes problemas.

$$\beta < 90^\circ$$

Neste grupo apenas se podem recuperar pontos das superfícies perpendiculares aos espelhos. A razão é a mesma apontada para o grupo anterior. Além disso, sabendo que, quando o ângulo entre os espelhos é menor do que 90° , começam a ser reflectidas as próprias reflexões do espelho, o espaço de imagem com zonas correspondentes diminui. Para este grupo

não são apresentados nenhuns exemplos porque não acrescentariam mais nada ao que já foi mostrado, especificamente com os dois exemplos anteriores, relativos ao caso em que $\beta = 90^\circ$.

$\beta > 90^\circ$

Por fim, havendo um ângulo entre os espelhos maior do que 90° , começa-se a poder recuperar pontos de outras superfícies, além daquelas que são perpendiculares aos espelhos. Obviamente que essa possibilidade vai aumentando com o ângulo β . A razão para este facto está na diminuição da *baseline* entre os centros ópticos virtuais.

A triangulação necessária à recuperação da informação sobre o posicionamento tridimensional do ponto, visto através das suas projecções na imagem, apoia-se também nos planos XZ e YZ . No entanto, existe uma diferença obvia em relação ao grupo onde $\beta = 90^\circ$. Um dos pontos para se executar a triangulação, é automaticamente retirado do espelho esquerdo, pois este espelho está assente no plano XZ . O outro ponto necessita de algo mais, antes de ser usado na triangulação. Numa primeira fase, obtêm-se as coordenadas da projecção desse ponto no plano YZ (coordenadas Y e Z no espelho direito). De seguida, usando a equação (3.13), obtêm-se as coordenadas tridimensionais finais do ponto no espelho direito. Em suma, do espelho esquerdo tem-se $P_e = [X_e \ 0 \ Z_e]^T$ e do espelho direito tem-se $P_d = \left[\frac{Y_d}{\tan \beta} \ Y_d \ Z_d \right]^T$.

Exceptuando esta diferença quanto ao modo de obtenção dos pontos P_e e P_d , todo o restante processo deste grupo é exactamente igual ao processo do grupo em que $\beta = 90^\circ$, pelo que aqui se prescinde da sua exposição.

Capítulo 4

Reconstrução de superfícies

A reconstrução tridimensional de superfícies a partir de imagens é um processo com grande potencial de aplicação em diferentes domínios de actividade produtiva, como, por exemplo, em sistemas de localização e seguimento, jogos de vídeo, realidade virtual, produção cinematográfica ou até em publicidade na internet.

Sendo este assunto tão importante não poderemos deixar de avançar, nesta dissertação, para a recuperação dessa informação, no sentido da obtenção da estrutura tridimensional da cena. Assim, este capítulo começará por fazer uma introdução ao tema da reconstrução de superfícies. De seguida, apresenta a técnica de reconstrução que será usada, denominada por escavação do espaço tridimensional (*space carving*). Esta técnica baseia-se em silhuetas. Depois faz-se a sua aplicação, em troca com a triangulação, à última aproximação vista no capítulo anterior. Por fim, abandonando o modo de aquisição de imagens baseado nos espelhos, apresenta-se um sistema prático e completo, no que diz respeito à reconstrução tridimensional de modelos de cenas do mundo 3D a partir de uma sequência de silhuetas bidimensionais. Estas silhuetas pertencem a imagens previamente calibradas.

4.1 Introdução

É cada vez maior a utilização de computadores na produção industrial de objectos físicos a partir de modelos digitais. Tradicionalmente, em gráficos computadorizados, os modelos 3D são construídos utilizando-se software de design polígono a polígono. Essa abordagem consome imenso tempo e a qualidade da informação que se obtém depende, em muito, das

capacidades do utilizador. O processo inverso e correspondente à inferência da descrição digital de objectos físicos a partir das suas imagens, tem recebido menos atenção. Este processo é normalmente referido como uma actividade de engenharia reversiva ou mais especificamente, perscrutação tridimensional (*3D scanning*). Existem várias características tridimensionais num objecto que podem ter interesse em serem recuperadas, onde se incluem a forma, as cores e as propriedades do material do objecto. Este capítulo tratará só o problema da recuperação de formas tridimensionais, também designado por reconstrução de superfícies. O propósito da reconstrução de superfícies é a determinação da informação geométrica sobre a forma da superfície, normalmente um conjunto de distâncias ou orientações, medidas entre a superfície e um referencial global. É usual chamar a este conjunto de informações tridimensionais mapas de profundidade ou mapas 3D.

Existem diferentes técnicas de adquirir informação sobre uma superfície tridimensional. Uma das técnicas consiste em adquirir os dados tridimensionais através do uso de sondas de tacto, montadas em sistemas mecânicos de medição de coordenadas. Esta forma de aquisição de dados é muito precisa, mas morosa e dispendiosa, estando limitada aos materiais que podem resistir ao contacto mecânico. Esta técnica é ainda muito utilizada na indústria automóvel e na indústria aeronáutica. Menos fiáveis e mais baratas as sondas digitais manuais determinam posicionamentos através de campos magnéticos ou ultra sons. Este tipo de digitalização requer, no entanto, uma significativa intervenção do homem [Hoppe 94].

Aproveitando as potencialidades criadas e testadas na área da visão por computador, começou-se, recentemente, a substituir as sondas mecânicas por sondas laser de profundidade (laser range scanner), uma vez que estas, para além da rapidez, permitem a obtenção de um conjunto de dados densos e precisos, a elevadas larguras de banda. Estas sondas iluminam o objecto com o feixe laser e medem a distância ao referencial usando triangulação, interferência ou o tempo de voo, produzindo os mapas de profundidade com as distâncias do sensor ao objecto que está a ser sondado. Se o sensor laser e o objecto estiverem fixos então poderá haver oclusão e apenas os pontos visíveis podem ser completamente recuperados. Para se reconstruírem objectos de uma forma global e completa, deverá haver movimento entre o sensor e o objecto ou existirem vários mapas de profundidade. Neste último caso o conjunto dos mapas podem conter a estrutura do objecto mas a fusão dos vários mapas que daria a reconstrução global da superfície, não é trivial [Almeida 99] [Ayache 96].

Mais recentemente, devido às técnicas anteriores serem muito dispendiosas, requererem cuidadosa calibração, não funcionarem muito bem com superfícies polidas ou que reflectam

pouca luz, ou funcionarem com objectos de tamanho limitado, começaram a ser populares os métodos em que a aquisição é feita através de luz estruturada (lasers) e/ou de câmaras. Estes métodos baseiam-se nas relações geométricas estabelecidas entre os dispositivos de aquisição, iluminação e as superfícies a recuperar. No entanto, a maior parte dos métodos de reconstrução feitos nesta área necessitam de conhecimentos adicionais, como a estrutura dos dados recolhidos, as características internas dos sensores e as informações sobre a orientação da iluminação, da superfície ou dos sensores em relação a um referencial [Martins 99] [Ferreira 00]. Ao permitir-se que os modelos 3D possam ser reconstruídos automaticamente a partir de uma sequência de imagens, a estrutura formada por estas técnicas fornece uma solução cujo custo é reduzido e a eficiência é aumentada. Adicionalmente, os sistemas baseados nestas técnicas também podem lidar com objectos de vários tamanhos e com reflexões variadas. O método descrito neste capítulo insere-se no domínio destas técnicas.

Outro ponto importante na análise das várias técnicas de reconstrução de superfícies é o da representação tridimensional que é gerada pela técnica. Essa representação terá que ser compatibilizada com a descrição tridimensional dos objectos em computador e poderá ser realizada de diferentes maneiras. Uma delas, muito popular, baseia-se numa representação de linhas e curvas na superfícies do objecto (frames). Uma alternativa é utilizar uma representação volumétrica do objecto, através de agregações de primitivas elementares da representação volumétrica, vulgarmente conhecida como *voxel* ou *pixels* 3D [Kutulakos 99] [Wong 01]. Cada *voxel* representa um pequeno volume do espaço tridimensional, podendo estar ocupado (ou não) e reter características correspondentes às propriedades físicas da cena nesse volume elementar. Uma segunda alternativa é a representação com pequenas áreas de superfícies conhecidas que, agregadas, permitem a obtenção da reconstrução tridimensional. As superfícies que mais se utilizam são pequenos pedaços de planos ou agregações de superfícies representadas por curvas - ver [Ferreira 00] para uma solução com curvas NURBS. O processo de modelação apresentado neste artigo segue a segunda forma de abordagem através de uma representação volumétrica baseada em *voxels*.

4.2 Aplicações potenciais

Com o desenvolvimento de sistemas rápidos e baratos para a reconstrução tridimensional houve um crescimento dos domínios de aplicação destas técnicas. Este crescimento, visível

no texto seguinte, impôs a necessidade do estudo deste problema, nesta dissertação.

Uma dessas áreas é a antropometria. Os dados antropométricos são muito usados no design de produtos tais como: calçado, vestuário, equipamento de segurança, móveis e outros objectos com que o ser humano interage. Este tipo de dados é, por vezes, de importância capital. Como exemplo disso, poder-se-á referir a necessidade de precisão nos equipamentos de protecção ou a problemática da ergonomia de alguns equipamentos. A maioria dos dados antropométricos de uso generalizado deriva de informações adquiridas, há cerca de 50 anos atrás, por medição manual de uma amostra da população. As melhorias tecnológicas nos anos mais recentes tornaram esses dados obsoletos. Esforços recentes para a solucionar este problema saíram frustrados pela ausência de um método adequado à aquisição rápida de dados tridimensionais densos e precisos do corpo humano. O advento das recentes tecnologias de recuperação tridimensional com imagens veio colmatar esta falha [Márquez 00].

Ao nível industrial, o facto de se poder manusear modelos digitais dos objectos físicos traz vantagens visíveis de redução custos. Por exemplo, uma simulação computacional com os modelos pode dar a conhecer problemas que podem ser resolvidos antes da sua realização. Esta tecnologia também pode ser usada no controlo de produção, especificamente em inspecções e metrologia dimensional, aumentando a velocidade e qualidade da produção. Além disto, com a tecnologia emergente SFF (solid free-form fabrication), as reconstruções tridimensionais resultantes podem ser modificadas ou incorporadas na criação de novos produtos, obtendo-se, rapidamente, os protótipos de objectos tridimensionais [Weiss 97]. Também é iniciada a possibilidade do envio/recepção, para qualquer parte do mundo, de qualquer objecto de uma forma extremamente rápida (fax tridimensional) [Levoy 00]. Com os protótipos pode-se ainda estudar o comportamento de novos produtos no mercado.

O marketing é uma outra das áreas que beneficiou desta tecnologia. Com os resultados das reconstruções podem ser criadas bases de dados tridimensionais de produtos que podem ser publicitados pela Internet ou visualizados em computadores, explorando as potencialidades que a tridimensionalidade contém. Uma dessas potencialidades é a do uso virtual do produto, por exemplo mobiliário, onde o utilizador pode, em sua casa, fazer medições ou compor o produto que mais lhe interessa com os vários componentes existentes.

Outro campo de utilização desta tecnologia é a ciência forense. Alguns dos exemplos mais importantes do seu uso são os da identificação de corpos, especialmente na reconstrução facial e na medição das cenas do crime [Improofs www].

Outras áreas que este tipo de tecnologia revolucionou foram os meios audiovisuais, onde

a realidade virtual, até aí apenas possível através da criatividade de alguns animadores, entrou numa era de realismo notável. Exemplos da utilização de reconstrução tridimensional de superfícies existem no campo da produção cinematográfica e televisiva [3D Matic [www](#)].

No campo cultural e educativo esta tecnologia permitiu recriar vários ambientes de outras épocas ou artefactos de civilizações perdidas, como o caso de Sagalassos, na Turquia [3D Murale [www](#)]. Com a ajuda da reconstrução tridimensional pode-se preservar os objectos que a arqueologia resgatou, usando os modelos digitais em vez dos reais, e proporcionar a qualquer pessoa uma melhor aprendizagem sobre os hábitos dos antepassados.

Um dos campos de maior potencialidade de disseminação desta tecnologia é o da medicina, onde numerosos exemplos de possibilidades da sua utilização podem ser encontrados. Esses exemplos vão desde a exame, o diagnóstico, o planeamento e a simulação cirúrgica e o projecto e fabrico de próteses, entre outros. Aliás, a integração de dados obtidos por sistemas de radiologia e similares com dados tridimensionais adquiridos pela recuperação tridimensional são exemplo dos avanços mais recentes deste campo [3D Matic [www](#)] [Visible [www](#)]. Outras aplicações da reconstrução tridimensional enquadram-se na pesquisa da cura para certas doenças. Um exemplo disso, é o desenvolvimento de um sistema de análise de imagens tridimensionais para microscópicos electrónicos, aplicado ao estudo da doença de Parkinson e que se utiliza uma modelização volumétrica [Márquez].

4.3 Reconstrução por escavação do espaço 3D

Apesar de existirem diversas técnicas que solucionam o problema da reconstrução tridimensional de superfícies a partir de imagens, estas soluções são numericamente estáveis para problemas bem definidos e restritos [Marr 76] [Cipolla 92]. Recentemente, começaram a ser desenvolvidas alternativas a estas técnicas, com algoritmos de modelação volumétrica baseados no princípio da escavação do espaço tridimensional. Foi demonstrado que podem ser recuperadas estruturas muito complexas com algoritmos baseados neste princípio [Kutulakos 99] [Wong 01].

4.3.1 Introdução

A teoria da escavação do espaço é uma técnica que consegue uma solução para o problema da reconstrução tridimensional de superfícies de uma forma geral, a partir de im-

agens [Kutulakos 99]. Esta técnica define, no início do processo de escavação, um espaço volumétrico onde cabe a superfície a reconstruir. Em cada iteração, este espaço volumétrico é escavado até a superfície resultante ser consistente com as imagens captadas, isto é, até que todos os pontos dentro do espaço volumétrico sejam consistentes com essas imagens. Um ponto tridimensional, pertencente ao espaço volumétrico, é consistente com um ponto numa imagem, se a sua côr resultar da radiação do ponto tridimensional. Isto é válido para modelos de reflexão que são calculados localmente, sendo gerado um volume tridimensional consistente com a radiação captada na imagem obtida daquele ponto de vista.

O objectivo do critério de verificação de consistência é decidir se existe um valor de radiação que possa ser atribuído a um ponto no espaço, de forma a que ele seja considerado consistente com as imagens de entrada. Em cada iteração é calculado o critério de consistência para um *voxel* e, se ele for consistente, é-lhe atribuído o valor de radiação. Caso contrário não fará parte da superfície que se está a reconstruir. Este processo é repetido até não ser possível eliminar mais nenhum *voxel*. O que resta do espaço volumétrico deverá ser a forma da superfície que se desejava recuperar. Kutulakos e Seitz provaram que se esta superfície for encontrada, então é consistente com todas as imagens da superfície.

Nesta dissertação é escolhida esta abordagem de intersecção do volume devido à sua capacidade de descrever objectos de topologias mais complexas (por exemplo, objectos com buracos). Baseado em [Szeliski 93] e [Wong 01], é apresentado um algoritmo que cria uma *octree*, utilizando silhuetas a partir de múltiplas imagens, usando uma variante ao critério de verificação de consistência. Deste modo, em alternativa à função que modeliza a radiação da cena, utilizam-se as silhuetas e as projecções dos *voxels* nas imagens. Com este critério, obtém-se a taxa de ocupação da superfície em cada *voxel*. As silhuetas (ou contornos) são uma característica dominante nas imagens e podem ser extraídas com relativa facilidade. Fornecem informações importantes no que diz respeito à forma dos objecto e são a única informação disponível no que toca a superfícies suaves quase sem textura.

Além da estrutura recuperada o processo aqui descrito obtém, também, a textura da cena 3D. Esta informação advém da côr que os *voxels* têm nas imagens.

4.3.2 Representação de uma *Octree*

Uma *octree* [Jackins 80] é uma estrutura de dados em árvore, na qual cada nodo tem, pelo menos, oito ramificações. É geralmente utilizada em computação gráfica para a repre-

sentação volumétrica de objectos, onde cada nodo na árvore representa um *voxel* (elemento do volume) no espaço tridimensional.

A representação dos objectos é feita, então, da seguinte maneira. O nodo da raiz da *octree* consiste num só *voxel* de grandes dimensões que define o volume envolvente do objecto. A *octree* é construída subdividindo-se recursivamente cada *voxel* da árvore em oito sub-*voxels*, que são representados por oito nodos filho. A cada nodo da árvore é atribuído uma de três côres (preto, cinzento ou branco), de acordo com a sua taxa de ocupação. Um nodo preto representa um *voxel* que está totalmente preenchido, um nodo cinzento representa um *voxel* que está parcialmente preenchido e um nodo branco representa um *voxel* que está completamente vazio. É de notar que tanto os nodos pretos como os nodos brancos não têm qualquer ramificação. Os nodos cinzentos têm ramificações, podendo esses seus filhos terem qualquer uma das três côres. Eles representam um *voxel* que se encontra na superfície do objecto. A figura 4.1 mostra um volume simples representado por uma *octree*.

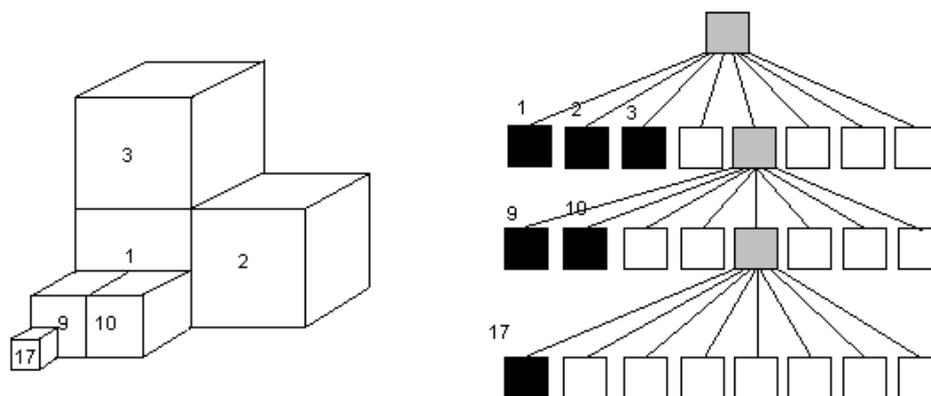


Figura 4.1: Representação de um volume simples através de uma *octree*.

Poderão ser encontrados mais detalhes sobre as representações, construções e manipulações de *octree* em [Jackins 80] [Chen 88].

Na implementação apresentada neste capítulo, os *voxels* são cubos e cada subdivisão produz oito sub-cubos idênticos. Cada nodo na *octree* armazena a côr (taxa de ocupação), o comprimento e as coordenadas do centro do cubo que representa. Também contém indicadores das suas ramificações, se existirem. A fim de se conseguir um acesso mais rápido

aos cubos de um determinado nível, todos os cubos no mesmo nível são armazenados numa lista dinâmica.

4.3.3 Algoritmo de escavação baseado em silhuetas

O algoritmo estabelece inicialmente o nodo da raiz da *octree* é como um cubo cinzento que envolve completamente o objecto. A partir daqui, a fim de aperfeiçoar o resultado da recuperação da superfície, vão-se formando novos níveis, subdividindo cada cubo cinzento do nível anterior em oito sub-cubos. É de notar que os cubos brancos e pretos não precisam de ser subdivididos, uma vez que todos os seus possíveis nodos filhos teriam a mesma classificação que os nodos originais.

Antes de se iniciar o teste de consistência dos cubos com as imagens, assume-se que todos esses cubos do novo nível estarão dentro do objecto, atribuindo-lhes a cor preta. O teste de consistência é feito projectando os cubos em cada imagem, e com os pontos resultantes verificando se estes se intersectam com a silhueta na imagem. Com este teste, determina-se a taxa de ocupação do respectivo cubo. A projecção em cada imagem da sequência é feita utilizando uma matriz de projecção que está associada às imagens. Neste processo, as cores dos cubos são actualizadas.

Se a projecção dos pontos se situar completamente fora da silhueta na imagem corrente, o cubo deve encontrar-se completamente fora do objecto. É-lhe atribuída, então, a cor branca e não é necessário mais nenhuma verificação nas outras imagens. Se a projecção se situar parcialmente dentro da silhueta na imagem corrente, o cubo deve encontrar-se perto do limite do objecto. A sua cor é, então, actualizada para cinzenta, a fim de indicar que a sua taxa de ocupação é ambígua e que é necessário ser mais aperfeiçoado. Finalmente, se a projecção se situar completamente dentro da silhueta na imagem corrente, a sua taxa de ocupação não pode ser determinada de imediato e, por isso, apenas mantém a sua cor actual. Se o cubo continuar a ser preto após a verificação de todas as projecções em todas as silhuetas da sequência, o cubo deve então encontrar-se completamente dentro do objecto. Note-se que os cubos só podem ser removidos ou "esculpidos" do objecto que está a ser reconstruído.

O processo atrás descrito é repetido até não existir nenhum cubo cinzento no nível corrente ou, na prática, até se atingir o nível de resolução pré-estabelecido. O algoritmo para a reconstrução, utilizando imagens calibradas da mesma cena tridimensional, captadas

de múltiplas posições, é dado pelo seguinte pseudo-código:

```
Estabelecer o tamanho do menor cubo que envolva o objecto;
Estabelecer o numero de niveis da octree;
Colocar o nodo raiz da octree a cinzento;
Enquanto o nivel maximo nao for atingido faz,
  Se nao houver nenhum cubo cinzento no nivel corrente entao,
    Termina o processo;
  Para cada nodo cinzento do nivel corrente faz,
    Subdivide-o em 8 sub-nodos;
    Para cada sub-nodo faz,
      Coloca a sua cor a preto;
      Para cada imagem da sequencia faz,
        Projecta o cubo na imagem;
        Se a projeccao estiver completamente fora da silhueta entao,
          Coloca a cor do nodo correspondente a branco;
          Ignora o resto das imagens;
        Senao,
          Se a projeccao estiver parcialmente dentro da silhueta entao,
            Muda a cor do nodo para cinzenta;
          Senao,
            Mantem a cor do nodo;
```

Este algoritmo é muito eficaz quando os cubos brancos são identificados na sua etapa o mais inicial possível, evitando, assim, desnecessárias projecções, testes de intersecção e subdivisões do cubo. Uma vez que somente os cubos cinzentos do nível corrente são considerados e refinados durante cada iteração do processo, é preciso ter cuidado em não classificar o cubo como preto (i.e. completamente dentro) ou branco (i.e. completamente fora) a menos que se tenha a certeza. Pelo contrário, se um cubo branco ou preto for erradamente classificado como cinzento, apenas indicará que a taxa de ocupação desse cubo é ambígua e que o cubo será reconsiderado e refinado no próximo nível.

Neste algoritmo, quanto maior for a divisão do volume inicial, ou seja, o número de níveis da *octree* (NN), melhor será a resolução da reconstrução. No entanto, para uma maior resolução o processo será computacionalmente mais demorado, pois o número total de cubos

possíveis de testar é de 8^{NN} . Depois do processo de remoção de *voxels* estar completo, os que restaram correspondem à superfície reconstruída que assegura uma consistência geométrica e fotométrica das imagens da superfície. A amostragem da superfície é feita recorrendo à biblioteca gráfica e à informação contida na *octree* (mostram-se todos os *voxels* pretos). Para desenhar um cubo, o programa de visualização lê o tamanho do respectivo *voxel* e o seu centro. As coordenadas de todos os *voxels* da *octree* são relativas ao mesmo referencial.

4.3.4 Extração das silhuetas e teste de intersecção

Para a extração das silhuetas nas imagens é utilizado o filtro de *canny*. De forma a otimizar o teste de consistência, as silhuetas são representadas por imagens binárias. Além disso, num pré-processamento das imagens, a todos os *pixels* pertencentes aos objectos, delimitados pela silhueta, é-lhes atribuída a cor branca, enquanto que ao resto da imagem é atribuída a cor preta. Este processo pode ser representado pela figura 4.2.

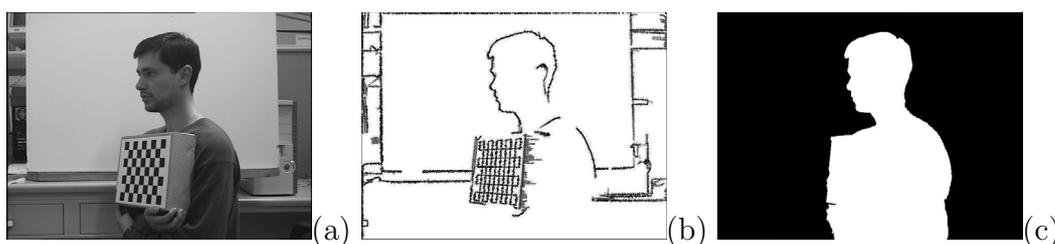


Figura 4.2: (a) Imagem original. (b) Silhueta extraída com a ajuda do filtro de *canny*. (c) Imagem no teste de consistência.

Para classificar um cubo dentro da *octree*, os oito vértices do cubo são projectados na imagem binária (ver figura 4.2(c)). A projecção de um cubo numa imagem é geralmente um hexágono. De modo a facilitar o algoritmo, escolheu-se o menor rectângulo que engloba todas as oito projecções dos vértices do cubo. Como este rectângulo é sempre maior ou igual que a projecção actual do cubo, tem que se realçar algumas considerações.

Se este rectângulo estiver completamente dentro da silhueta (todos os *pixels* são brancos), então o cubo também estará, para essa imagem. Similarmente, se este rectângulo estiver completamente fora da silhueta (todos os *pixels* são pretos), então isso significa que o cubo também estará. O problema poderia estar na situação em que o rectângulo estivesse parcialmente dentro da silhueta. No entanto, não se verifica este problema, pois como o cubo

é considerado ambíguo, vai ser repartido em oito sub-cubos, que são testados novamente, até se chegar a uma das situações atrás referidas.

4.3.5 Recuperação das texturas

A obtenção das texturas é feita na mesma fase em que se remove ou se mantém os *voxels*. A ideia presente à recuperação desta informação é a de que a textura de uma superfície não é mais do que o conjunto de côres que se vêem na imagem dessa superfície. Assim, da projecção dos *voxels* resulta, como foi visto, um rectângulo em cada imagem. À medida que se aumenta o nível da *octree*, os *voxels* tornam-se mais pequenos, bem como as suas projecções. No limite, a projecção de um *voxel* na imagem restringe-se a um único ponto. Conhecendo esse ponto na imagem tem-se a respectiva côr do *voxel*.

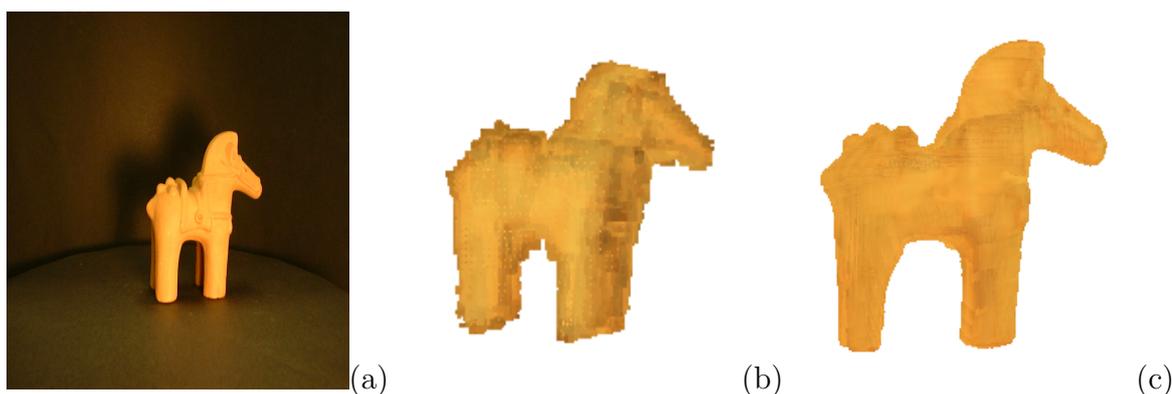


Figura 4.3: *Resultado da recuperação da textura. (a) Uma das imagens usadas; (b) Uso da média da côr das projecções do centro do voxel nas imagens; (c) Uso da média da côr das projecções dos vértices do voxel nas imagens.*

Na prática, como se usam várias imagens das superfícies tridimensionais e como estas imagens são capturadas de posições diferentes, a côr em cada imagem, referente a um mesmo ponto tridimensional, pode ser diferente. A solução passa, então, por usar a média das côres resultantes de cada projecção da superfície do *voxel* nas imagens. Nesse processo o *voxel* pode ter multiplas côres. Devido ao grande tempo de processamento que esse processo leva, pode ser usado apenas a média das côres resultantes da projecção do centro do *voxel* nas imagens, ou, como solução intermédia, a média das côres resultantes da projecção dos oito vértices do *voxel* nas imagens. Esta última solução gera resultados muito semelhantes à côr

original da superfície 3D. O resultado da utilização destas duas soluções podem ser vistos na figura 4.3.

Na obtenção da textura devem ser usadas todas as imagens, utilizadas na escavação, de forma a suavizar ruídos. Estes ruídos podem resultar de erros no cálculo das silhuetas. Basta que para isso a silhueta seja maior do que a imagem do objecto a recuperar.

Outro problema que existe na recuperação da textura é o da atribuição de côres para zonas que não aparecem na imagem (por exemplo, a parte de trás de uma superfície, quando se tem apenas as imagens da parte da frente). Como as únicas informações que se têm são as imagens, a côr atribuída às zonas ocultas vai ser a do *voxel* que está mais perto da câmara, segundo as rectas projectantes. Isto faz com que, neste caso, as superfícies resultantes possam ser pouco claras, pois as texturas que aparecem nas imagens são repetidas. Mais à frente será mostrado um exemplo deste problema.

4.3.6 Resultados experimentais

De seguida, apresentar-se-ão os resultados experimentais da aplicação do algoritmo explicado anteriormente. Primeiramente será vista essa aplicação ao sistema desenvolvido no capítulo 3, referido como aproximação (c1) na figura 1.1. Posteriormente, abandonando o modo de aquisição de imagens baseado nos espelhos, mostrar-se-ão os resultados da aplicação deste algoritmo a imagens calibradas, adquiridas de vários pontos de vista (aproximação (c2) na figura 1.1).

Reconstrução baseada nos espelhos

Tal como foi visto, qualquer aplicação do algoritmo de escavação do espaço tridimensional necessita de usar imagens calibradas, isto é, tem que conhecer todas as matrizes de projecção associadas a essas imagens. Assim, como se pretende aplicar este algoritmo ao sistema de reconstrução mostrado na secção 3.3 (aproximação (c1) na figura 1.1), há que fazer alguns ajustes e procurar as informações que faltam.

Apesar das matrizes \mathbf{T}_{xz} e \mathbf{T}_{yz} , dadas respectivamente pelas equações (3.12) e (3.14), terem algo em comum com as matrizes de projecção, elas apenas permitem a projecção entre o plano imagem e um plano previamente definido. Logo, para aplicar o algoritmo de escavação do espaço tridimensional a este sistema tem que se alargar essas matrizes, de forma a conhecer-se completamente a relação entre o mundo tridimensional e as imagens.

Especificamente, temos que estimar as matrizes de calibração das câmaras (as quais são as mesmas para todas as câmaras devido às propriedades dos espelhos), juntamente com as transformações (rotações e translações) entre os quatro sistemas de coordenadas, mundo 3D, câmara real, câmara virtual esquerda e câmara virtual direita. Este processo é designado por calibração das câmaras.

O problema da calibração das câmaras, necessário à extracção de informação 3D a partir de imagens, tem sido amplamente estudado. O seu objectivo é o cálculo da transformação do espaço vectorial da câmara em qualquer outro espaço vectorial, normalmente o do mundo 3D. Zhang, em [Zhang 02], descreve genericamente as técnicas de calibração de acordo com a dimensão dos objectos de calibração. Segundo este autor, a calibração 3D é feita com base na observação do objecto de calibração cuja geometria no mundo 3D é conhecida com muito boa precisão [Faugeras 93]. A calibração 2D é executada observando o mesmo padrão plano com diferentes orientações [Zhang 00]. Por seu lado, a calibração 1D baseia-se na observação de um objecto, composto por um conjunto de pontos colineares (devendo um deles estar fixo), de várias posições diferentes (mínimo seis) [Zhang 02]. A calibração 0D é feita a partir das relações geométricas entre a cena estática e uma câmara móvel [Hartley 00].

Existem, no entanto, outras técnicas que são difíceis de encaixar na classificação descrita acima. Kim, em [Kim 98], classificou os métodos de calibração em cinco categorias: as técnicas que envolvem uma optimização não linear em toda a escala, as técnicas que envolvem o cálculo de equações lineares na obtenção da matriz de transformação em perspectiva, as técnicas baseadas no método dos dois planos, as técnicas que recorrem ao método dos dois estados e as técnicas da auto-calibração adaptativa.

Os dois métodos aqui propostos têm características que os fazem pertencer à calibração 2D e 1D, de acordo com a classificação de Zhang, porque observam planos e necessitam conhecer um conjunto de pontos. No entanto, tem algumas vantagens quando comparados com estes dois tipos de calibração. Os planos observados pela câmara, os espelhos, são vistos de uma única posição, em vez de várias, e os pontos que se precisam conhecer são apenas dois, e não seis, além de não precisarem de ser colineares nem fixos. De acordo com a classificação de Kim, as técnicas que se vão apresentar de seguida envolvem o cálculo de equações lineares na obtenção da matriz de transformação em perspectiva.

O primeiro método aqui proposto é extremamente simples, pois recupera a matriz de projecção, \mathbf{T} , associada à câmara real fazendo uso das matrizes \mathbf{T}_{xz} e \mathbf{T}_{yz} , referidas na subsecção 3.3.3, e definidas respectivamente pelas equações (3.12) e (3.14). Assim, consid-

eremos as matrizes

$$\mathbf{T} = \begin{bmatrix} \mathbf{t}_{11} & \mathbf{t}_{12} & \mathbf{t}_{13} & \mathbf{t}_{14} \\ \mathbf{t}_{21} & \mathbf{t}_{22} & \mathbf{t}_{23} & \mathbf{t}_{24} \\ \mathbf{t}_{31} & \mathbf{t}_{32} & \mathbf{t}_{33} & \mathbf{t}_{34} \end{bmatrix} \quad \mathbf{T}_{xz} = \begin{bmatrix} \mathbf{txz}_{11} & \mathbf{txz}_{12} & \mathbf{txz}_{13} \\ \mathbf{txz}_{21} & \mathbf{txz}_{22} & \mathbf{txz}_{23} \\ \mathbf{txz}_{31} & \mathbf{txz}_{32} & \mathbf{txz}_{33} \end{bmatrix} = k_{xz} \begin{bmatrix} \mathbf{t}_{11} & \mathbf{t}_{13} & \mathbf{t}_{14} \\ \mathbf{t}_{21} & \mathbf{t}_{23} & \mathbf{t}_{24} \\ \mathbf{t}_{31} & \mathbf{t}_{33} & \mathbf{t}_{34} \end{bmatrix}$$

$$\mathbf{T}_{yz} = \begin{bmatrix} \mathbf{tyz}_{11} & \mathbf{tyz}_{12} & \mathbf{tyz}_{13} \\ \mathbf{tyz}_{21} & \mathbf{tyz}_{22} & \mathbf{tyz}_{23} \\ \mathbf{tyz}_{31} & \mathbf{tyz}_{32} & \mathbf{tyz}_{33} \end{bmatrix} = k_{yz} \begin{bmatrix} \frac{\mathbf{t}_{11}}{\tan \beta} + \mathbf{t}_{12} & \mathbf{t}_{13} & \mathbf{t}_{14} \\ \frac{\mathbf{t}_{12}}{\tan \beta} + \mathbf{t}_{22} & \mathbf{t}_{23} & \mathbf{t}_{24} \\ \frac{\mathbf{t}_{13}}{\tan \beta} + \mathbf{t}_{32} & \mathbf{t}_{33} & \mathbf{t}_{34} \end{bmatrix}$$

Olhando atentamente para a composição destas matrizes, pode-se constatar que se obtém explicitamente da matriz \mathbf{T}_{xz} a primeira, terceira e quarta colunas da matriz de projecção \mathbf{T} , definidas a menos de um factor multiplicativo. Por seu lado, da matriz \mathbf{T}_{yz} pode-se, também, obter explicitamente a terceira e quarta colunas da matriz de projecção \mathbf{T} , definidas, da mesma forma, a menos de um factor multiplicativo.

Da comparação entre as duas últimas colunas das matrizes \mathbf{T}_{xz} e \mathbf{T}_{yz} encontra-se a relação entre elas, dada pelo factor de escala $\frac{k_{yz}}{k_{xz}}$, pois ambas as colunas de cada matriz devem ser iguais entre si. Com base neste factor, obtém-se a informação que faltava conhecer, a segunda coluna da matriz \mathbf{T} . Assim, como a matriz \mathbf{T} também é definida a menos de um factor de escala, a sua composição pode ser dada por

$$\mathbf{T} = \begin{bmatrix} \mathbf{t}_{11} & \mathbf{t}_{12} & \mathbf{t}_{13} & \mathbf{t}_{14} \\ \mathbf{t}_{21} & \mathbf{t}_{22} & \mathbf{t}_{23} & \mathbf{t}_{24} \\ \mathbf{t}_{31} & \mathbf{t}_{32} & \mathbf{t}_{33} & \mathbf{t}_{34} \end{bmatrix} = \begin{bmatrix} \mathbf{txz}_{11} & k_{yz}\mathbf{tyz}_{11} - \frac{\mathbf{txz}_{11}}{\tan \beta} & \mathbf{txz}_{12} & \mathbf{txz}_{13} \\ \mathbf{txz}_{21} & k_{yz}\mathbf{tyz}_{21} - \frac{\mathbf{txz}_{21}}{\tan \beta} & \mathbf{txz}_{22} & \mathbf{txz}_{23} \\ \mathbf{txz}_{31} & k_{yz}\mathbf{tyz}_{31} - \frac{\mathbf{txz}_{31}}{\tan \beta} & \mathbf{txz}_{32} & \mathbf{txz}_{33} \end{bmatrix}$$

pois é assumido, sem perda de generalidade, que $k_{xz} = 1$. Como se reparou, a abertura entre os espelhos, β , tem sempre que ser conhecida. Relembra-se que apenas precisamos do valor de β porque ambos os espelhos assentam no plano XY .

Como se constata pelo atrás referido, este método calcula apenas a matriz de projecção que relaciona o mundo 3D com a câmara real. Isto deve-se, simplesmente, à relação entre β , o que cada câmara virtual consegue ver da outra (reflexão da reflexão), e o espaço útil de reconstrução. Quanto mais as câmaras virtuais se conseguirem ver, menor é β e menor é o espaço útil à reconstrução (problema da oclusão - ver subsecção 3.3.4 para $\beta < 90^\circ$). No entanto, para se calcularem as matrizes de projecção virtuais deveria ser possível ver pelas câmaras virtuais, ao menos uma reflexão de um plano da cena 3D criada pelo outro espelho.



Figura 4.4: *Imagens usadas na reconstrução baseada no método das homografias \mathbf{T}_{xz} e \mathbf{T}_{yz} .*

Relembra-se que cada câmara virtual está associada a um espelho. Como estas restrições não são complementares, só se pode obter a matriz de projecção associada à câmara real.

A aplicação deste método começa com a aquisição das imagens apresentadas na figura 4.4. Com base nestas imagens e sabendo que a abertura entre os espelhos é de 141 graus ($\beta = 141^\circ$), obtiveram-se as respectivas matrizes \mathbf{T}_{xz} e \mathbf{T}_{yz} . Assim, para a figura 4.4(a) essas matrizes são dadas por

$$\mathbf{T}_{xz} = \begin{bmatrix} -1.7747e-3 & -1.0244e-4 & 7.3745e-1 \\ 6.2271e-6 & -1.7652e-3 & 6.7539e-1 \\ 1.4557e-7 & -4.0144e-7 & 1.8845e-3 \end{bmatrix} \quad \mathbf{T}_{yz} = \begin{bmatrix} 2.1319e-3 & -4.7793e-5 & 7.3745e-1 \\ 5.3545e-4 & -1.7546e-3 & 6.7539e-1 \\ -6.4952e-7 & -2.6508e-7 & 1.8845e-3 \end{bmatrix}$$

Da utilização da figura 4.4(b) resultaram as matrizes

$$\mathbf{T}_{xz} = \begin{bmatrix} -1.9011e-3 & -5.6901e-5 & 8.5630e-1 \\ -2.5657e-5 & -1.5625e-3 & 5.1647e-1 \\ -1.0461e-6 & -2.0540e-7 & 2.2768e-3 \end{bmatrix} \quad \mathbf{T}_{yz} = \begin{bmatrix} 2.0807e-3 & -1.4572e-4 & 8.5630e-1 \\ 5.4612e-5 & -1.5668e-3 & 5.1647e-1 \\ -1.2353e-6 & -4.3981e-7 & 2.2768e-3 \end{bmatrix}$$

Por fim, para a figura 4.4(c) as matrizes estimadas são

$$\mathbf{T}_{xz} = \begin{bmatrix} 1.4579e-3 & 1.5307e-5 & -7.3836e-1 \\ 1.5495e-4 & 1.2139e-3 & -6.7440e-1 \\ 8.9992e-7 & 7.6457e-8 & -1.7898e-3 \end{bmatrix} \quad \mathbf{T}_{yz} = \begin{bmatrix} 1.6621e-3 & -1.1903e-4 & 7.3836e-1 \\ -1.4479e-4 & -1.2569e-3 & 6.7440e-1 \\ -8.2786e-7 & -3.2619e-7 & 1.7898e-3 \end{bmatrix}$$

Como se constata, os valores dos elementos das segunda e terceira colunas das matrizes \mathbf{T}_{xz} e \mathbf{T}_{yz} apresentam uma ligeira diferença entre si. A justificação é a sensibilidade introduzida pelo método numérico que os permitiu obter, pois foi utilizado no processo de calibração um escasso número de pontos 2D e 3D e existia ruído na correspondência entre a cena 3D e a imagem.

A partir das seis matrizes anteriores calcularam-se, então, as matrizes de projecção, associadas à imagem i , com $i = a, b, c$, da figura 4.4, \mathbf{T}_i , dadas por

$$\mathbf{T}_a = \begin{bmatrix} -1.7747e-3 & -1.0088e-4 & -1.0244e-4 & 7.3745e-1 \\ 6.2271e-6 & 5.4329e-4 & -1.7652e-3 & 6.7539e-1 \\ 1.4557e-7 & -4.6638e-7 & -4.0144e-7 & 1.8845e-3 \end{bmatrix}$$

$$\mathbf{T}_b = \begin{bmatrix} -1.9011e-3 & -3.1105e-4 & -5.6901e-5 & 8.5630e-1 \\ -2.5657e-5 & 2.2333e-5 & -1.5625e-3 & 5.1647e-1 \\ -1.0461e-6 & -2.5513e-6 & -2.0540e-7 & 2.2768e-3 \end{bmatrix}$$

$$\mathbf{T}_c = \begin{bmatrix} 1.4579e-3 & 1.7221e-4 & 1.5307e-5 & -7.3836e-1 \\ 1.5495e-4 & 3.3973e-4 & 1.2139e-3 & -6.7440e-1 \\ 8.9992e-7 & 1.9600e-6 & 7.6457e-8 & -1.7898e-3 \end{bmatrix}$$

Com base nas matrizes anteriores e seguindo a ideia apresentada em [Silva 94], obtiveram-se os parâmetros intrínsecos da câmara, dados por

$$\mathbf{C}_a = \begin{bmatrix} 4.7128e-3 & 0 & -3.8515e-9 \\ 0 & 1.8424e-3 & 2.7019e-10 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{C}_b = \begin{bmatrix} 4.8602e-3 & 0 & 2.3616e-9 \\ 0 & 1.5652e-3 & 3.5480e-10 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{C}_c = \begin{bmatrix} 1.4681e-3 & 0 & 1.6507e-9 \\ 0 & 1.2700e-3 & 8.9815e-10 \\ 0 & 0 & 1 \end{bmatrix}$$

Por ter sido usada a mesma câmara, a matriz de calibração resultante das diferentes matrizes \mathbf{T} deveria ser a mesma. A diferença advém da presença de ruído nas matrizes que as originaram, conforme referido acima. O desvio existente não é, contudo, muito significativo.

Da mesma fonte obtiveram-se os seguintes parâmetros extrínsecos, associados a cada imagem da figura 4.4,

$$\mathbf{R}_a = \begin{bmatrix} -9.9673e-1 & -5.6658e-2 & -5.7536e-2 \\ 3.3716e-3 & 2.9416e-1 & -9.5575e-1 \\ 1.4557e-7 & -4.6638e-7 & -4.0144e-7 \end{bmatrix} \quad \mathbf{t}_a = \begin{bmatrix} 4.1418e+2 \\ 3.6569e+2 \\ 1.8845e-3 \end{bmatrix}$$

$$\mathbf{R}_b = \begin{bmatrix} -9.8645e-1 & -1.6140e-1 & -2.9525e-2 \\ -1.6416e-2 & 1.4290e-2 & -9.9976e-1 \\ -1.0461e-6 & -2.5513e-6 & -2.0540e-7 \end{bmatrix} \quad \mathbf{t}_b = \begin{bmatrix} 4.4431e+2 \\ 3.3045e+2 \\ 2.2768e-3 \end{bmatrix}$$

$$\mathbf{R}_c = \begin{bmatrix} 9.9304e-1 & -1.1730e-1 & 1.0426e-2 \\ 1.2200e-1 & -2.6750e-1 & 9.5580e-1 \\ 8.9992e-7 & -1.9600e-6 & 7.6457e-8 \end{bmatrix} \quad \mathbf{t}_c = \begin{bmatrix} -5.0292e+2 \\ -5.3101e+2 \\ -1.7898e-3 \end{bmatrix}$$

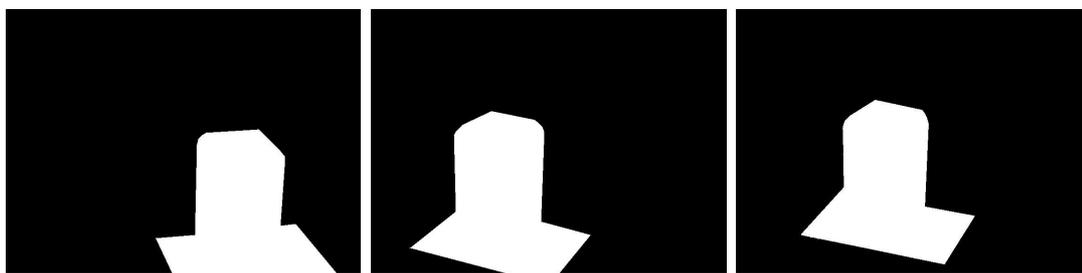


Figura 4.5: *Resultado da binarização das imagens da figura 4.4.*

Tendo as câmaras calibradas, pôde-se avançar para a reconstrução através do algoritmo da escavação do espaço 3D. Assim, após o processo de binarização das imagens da figura 4.4 obtiveram-se as imagens da figura 4.5. O resultado final desta reconstrução para dez níveis é apresentado na figura 4.6. Como se pode verificar,



Figura 4.6: *Diferentes vistas do resultado da reconstrução da cena 3D, usando as homografias \mathbf{T}_{xz} e \mathbf{T}_{yz} .*

a reconstrução é excelente tendo em conta o uso de apenas seis pares de pontos correspondentes. Com a apresentação deste exemplo, a intenção foi realçar a potencialidade do método recorrendo apenas à informação mínima necessária.

A grande vantagem deste método, como se pôde verificar, é a facilidade com que se obtêm as matrizes de projecção em perspectiva, \mathbf{T} . A sua desvantagem é a sensibilidade ao ruído, introduzida aquando da estimação das homografias \mathbf{T}_{xz} e \mathbf{T}_{yz} . Como foi explicado na secção 2.5, um pequeno erro no conjunto de dados de entrada pode resultar em grandes erros nas homografias. Assim, a fim de melhorar os

resultados das reconstruções deve-se usar o maior número de pontos correspondentes de entrada possíveis e aplicar o RANSAC para evitar falsas correspondências.

Através da comparação das figuras 3.17/3.18(b) e 4.6 verificam-se notórias melhorias na reconstrução. Enquanto que na aproximação (b), usada nas figuras 3.17 e 3.18(b), só era possível reconstruir, morosa e manualmente, superfícies planas ou quase planas, nesta aproximação, (c1), consegue-se recuperar, rápida e automaticamente, qualquer tipo de superfície.

Contudo, além do ruído, o escasso número de imagens usadas impede uma melhor definição da cena (problema da oclusão). Este problema fica a dever-se ao facto da câmara não conseguir ver a parte de trás da cena 3D e o seu raio de acção limitar-se às posições onde consegue ver, simultaneamente, a cena e as suas reflexões nos dois espelhos. Para ultrapassar esta questão avançou-se para a aproximação (c2).

O centro óptico da câmara é, geralmente, uma informação de difícil determinação para qualquer técnica de calibração. No caso do segundo método de calibração, aqui proposto, aproveita-se a facilidade da obtenção da posição do centro óptico da câmara, proporcionada pelo uso dos espelhos. Além disso, chama-se a atenção para a simplicidade na obtenção dos parâmetros intrínsecos e na definição do plano imagem, apenas com base em relações geométricas. A desvantagem deste método é a sua sensibilidade ao ruído devido à linearidade das equações usadas. Para ultrapassar este problema usa-se a normalização de Hartley para os pontos de entrada, 2D e 3D [Hartley 98].

Passemos, então, à explanação deste segundo método. Convém mencionar que, à semelhança do método anterior, este método reata o ponto onde se ficou na subsecção 3.3.3, retirando-lhe a parte relativa à reconstrução dos pontos. Outra ressalva diz respeito ao facto deste método usar apenas o espelho esquerdo na estimação da matriz de calibração das câmaras. Esta opção justifica-se pela simplificação de cálculos que a equação do espelho esquerdo, $Y = 0$, permite.

Assim, tal como foi definido, o sistema de coordenadas do mundo 3D está afecto aos espelhos. Devido a este facto, o centro óptico real, C_r , não é mais do que a translação \mathbf{t} entre o sistema de coordenadas do mundo 3D e o sistema de coordenadas da câmara real. Assumindo o seu conhecimento, $C_r = \mathbf{t} = \begin{bmatrix} t_x & t_y & t_z \end{bmatrix}^T$, juntamente

com o conhecimento da matriz que relaciona o espelho esquerdo com a imagem, \mathbf{T}_{XZ} , e partindo das equações (2.4), (2.6) e (3.12) obtemos

$$k\mathbf{T}_{XY} = \mathbf{C} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_3 & \mathbf{t} \end{bmatrix} \quad (4.1)$$

ou, de outro modo,

$$\begin{aligned} k \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{bmatrix} &= \begin{bmatrix} fk_x & \gamma & c_x \\ h_{21} & fk_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{13} & t_x \\ r_{21} & r_{23} & t_y \\ r_{31} & r_{33} & t_z \end{bmatrix} = \\ &= \begin{bmatrix} r_{11}fk_x + r_{21}\gamma + r_{31}c_x & r_{13}fk_x + r_{23}\gamma + r_{33}c_x & t_xfk_x + t_y\gamma + t_zc_x \\ r_{21}fk_y + r_{31}c_y & r_{23}fk_y + r_{33}c_y & t_yfk_y + t_zc_y \\ r_{31} & r_{33} & t_z \end{bmatrix} \end{aligned}$$

pois a coordenada Y é sempre igual a zero.

Da igualdade entre os elementos (3, 3) de cada matriz da equação (4.1), obtemos

$$k = \frac{t_z}{t_{33}} \quad (4.2)$$

Usando a equação (4.2) na igualdade entre os elementos (3, 1) de cada matriz da equação (4.1), obtemos

$$r_{31} = \frac{t_z t_{31}}{t_{33}} \quad (4.3)$$

Recorrendo ao mesmo processo do qual resultou a equação prévia, para os elementos (3, 2), obtemos

$$r_{33} = \frac{t_z t_{32}}{t_{33}} \quad (4.4)$$

Devido ao produto interno entre os mesmos dois vectores de \mathbf{R} ser unitário, tal como se viu na equação (2.5), e porque conhecemos r_{31} and r_{33} , calculamos

$$r_{32} = \pm \sqrt{1 - r_{31}^2 + r_{33}^2}$$

Não nos preocupamos com o sinal de r_{32} porque este parâmetro será usado na criação de uma recta 3D, onde o seu factor de escala absorverá esse sinal, tal como veremos mais à frente.

Como se sabe, segundo a equação (2.3), a matriz de rotação e o vector de translação definem a transformação de um ponto genérico P , do referencial do mundo 3D, para o ponto P_c , no referencial da câmara real. Tendo por base esse conhecimento, juntamente com as propriedades da rotação, pode-se obter a relação inversa através de

$$P = \mathbf{R}^{-1}(P_c - \mathbf{t}) = \mathbf{R}^T(P_c - \mathbf{t})$$

Isto significa que o vector $\vec{\mathbf{r}}_3 = \begin{bmatrix} r_{31} & r_{32} & r_{33} \end{bmatrix}$ é o vector director do eixo óptico, o qual é dado por

$$\mathbf{r}_{eo} \equiv C_r + \lambda \vec{\mathbf{r}}_3 = 0$$

Da intersecção desta recta 3D, \mathbf{r}_{eo} , com o plano do espelho ($Y = 0$), obtemos o ponto P_{oM} . Da sua aplicação à equação (3.12) resultam as coordenadas do ponto principal da imagem, isto é

$$p_{om} = (c_x, c_y) = \mathbf{T}_{XZ} \mathbf{P}_{oM}$$

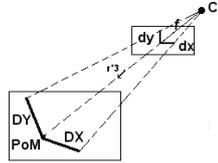


Figura 4.7: Cálculo das relações entre os sistemas métricos utilizados (metro e pixels).

As relações horizontal (k_x) e vertical (k_y) entre os sistemas métricos utilizados (pixels na imagem e metros no mundo 3D) obtêm-se, tal como demonstrado na figura 4.7, respectivamente, por

$$k_x = \frac{dx}{DX} \quad k_y = \frac{dy}{DY}$$

onde

$$dx = \overline{p_1 - p_0} \quad dy = \overline{p_2 - p_0} \quad DX = \overline{\mathbf{T}_{XY}^{-1} \mathbf{p}_1 - \mathbf{T}_{XY}^{-1} \mathbf{p}_0} \quad DY = \overline{\mathbf{T}_{XY}^{-1} \mathbf{p}_2 - \mathbf{T}_{XY}^{-1} \mathbf{p}_0}$$

Note-se que os pontos p_0 e p_1 devem formar uma recta horizontal, centrada na imagem, enquanto que os pontos p_0 e p_2 devem formar uma recta vertical, também centrada na imagem.

Com a informação obtida até agora, podemos estimar a distância focal da câmara real, na métrica do mundo 3D, baseando-nos na semelhança entre triângulos, através de

$$f = \frac{\mathbf{t}_z \overline{p_{om} e_e}}{P_{oM} - P_{eM}}$$

onde p_{om} e P_{oM} são os pontos atrás referidos, $e_e = (e_x, e_y)$ é o epipolo na imagem esquerda, estimado da secção 3.3.2, e P_{eM} é o ponto de intersecção da recta tridimensional $\overline{C_r C_v}$ com o plano espelho esquerdo. Para que o valor de f seja o correcto, todos os valores usados no seu cálculo devem estar na mesma métrica. Assim, as coordenadas x e y de todos os pontos da imagem devem ser divididas, respectivamente, por k_x e k_y . Devido ao facto de se ter recorrido à semelhança entre triângulos, devemos sempre utilizar $\overline{C_r C_v}$ pois é a única recta perpendicular ao espelho. Além disso, devemos utilizar a recta $\overline{C_r P_{oM}}$ porque ela é a única recta que é perpendicular ao plano imagem.

Utilizando os parâmetros f , k_x , k_y , c_x e c_y obtidos geometricamente, a equação (4.3) e a equação (4.2) na igualdade entre os elementos (2,1) de cada matriz da equação (4.1), obtemos

$$r_{21} = \frac{kt_{21} - c_y r_{31}}{fk_y}$$

Da mesma maneira para a igualdade dos elementos (2,2) de cada matriz da equação (4.1), recorrendo agora à equação (4.4), obtemos

$$r_{23} = \frac{kt_{22} - c_y r_{33}}{fk_y}$$

Seguindo o mesmo processo para a igualdade entre os elementos (1,3) de cada matriz da equação (4.1), chegamos a

$$\gamma = \frac{kt_{13} - c_x t_z - fk_x t_x}{t_y}$$

Finalmente, usando todos os parâmetros estimados até agora, calculamos

$$r_{11} = \frac{kt_{11} - c_x r_{31} - \gamma r_{21}}{fk_x} \quad r_{13} = \frac{kt_{1,2} - c_x r_{33} - \gamma r_{23}}{fk_x}$$

Conhecendo os vectores de rotação \mathbf{r}_1 e \mathbf{r}_3 e utilizando as propriedades das matrizes de rotação, obtemos

$$\mathbf{r}_2 = \mathbf{r}_1 \wedge \mathbf{r}_3 = \det \left(\begin{bmatrix} r_{11} & r_{21} & r_{31} \\ r_{13} & r_{23} & r_{33} \\ \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \end{bmatrix} \right) = \begin{bmatrix} r_{21}r_{33} - r_{23}r_{31} \\ r_{13}r_{31} - r_{11}r_{33} \\ r_{11}r_{23} - r_{13}r_{21} \end{bmatrix} \quad (4.5)$$

Resumindo, nesta fase conhecem-se já as matrizes de calibração de todas as câmaras ($\mathbf{C} = \mathbf{C}_{virtual} = \mathbf{C}_{virtual'}$), bem como a orientação relativa entre a câmara real e o mundo 3D (\mathbf{R} e \mathbf{t}).

Se utilizarmos a imagem virtual esquerda e os pontos do espelho esquerdo poderemos obter, usando o processo descrito na subsecção 3.3.3, outra matriz com as mesmas características de \mathbf{T}_{XZ} . Então, como conhecemos a posição do centro óptico virtual esquerdo, seguindo o mesmo processo acima descrito, obtemos os parâmetros extrínsecos inerentes à câmara virtual esquerda (\mathbf{R}_v e \mathbf{t}_v).

A orientação relativa entre as câmaras real e virtual esquerda é dada por

$$P_c = \mathbf{R}\mathbf{R}_v^{-1}(P_v - \mathbf{t}_v) + \mathbf{t}$$

onde P_c é um ponto genérico no referencial da câmara real e P_v é um ponto genérico no referencial da câmara virtual esquerda.

O processo de cálculo da orientação relativa entre a câmara virtual direita e o mundo 3D ($\mathbf{R}_{v'}$ e $\mathbf{t}_{v'}$) é semelhante ao desenvolvido anteriormente para as outras câmaras. No essencial as diferenças residem no uso da matriz \mathbf{T}_{YZ} , em vez da \mathbf{T}_{XZ} , e na utilização da equação do espelho direito, $X - \frac{1}{\tan \beta}Y = 0$.

Com os conhecimentos adquiridos até agora, é possível obter a equação do plano imagem, na métrica do mundo 3D. Assim, sejam os pontos da imagem $p_0 = [x_0 \ y_0]$, $p_1 = [x_1 \ y_1]$ e $p_2 = [x_2 \ y_2]$. Estes pontos, medidos em *pixels*, devem estar em posições diferentes e não serem todos colineares entre si. As suas coordenadas no sistema métrico do mundo 3D são as seguintes

$$P_0 = \left[\begin{array}{cc} \frac{x_0 - c_x}{k_x} & \frac{y_0 - c_y}{k_y} \\ \frac{t_z \sqrt{\left(\frac{x_0 - c_x}{k_x}\right)^2 + \left(\frac{y_0 - c_y}{k_y}\right)^2}}{T_{XY}^{-1} \mathbf{p}_0 - \mathbf{P}_{eM}} \end{array} \right]$$

$$P_1 = \begin{bmatrix} \frac{x_1 - c_x}{k_x} & \frac{y_1 - c_y}{k_y} & \frac{t_z \sqrt{\left(\frac{x_1 - c_x}{k_x}\right)^2 + \left(\frac{y_1 - c_y}{k_y}\right)^2}}{T_{XY}^{-1} \mathbf{p}_1 - \mathbf{P}_{eM}} \end{bmatrix}$$

$$P_2 = \begin{bmatrix} \frac{x_2 - c_x}{k_x} & \frac{y_2 - c_y}{k_y} & \frac{t_z \sqrt{\left(\frac{x_2 - c_x}{k_x}\right)^2 + \left(\frac{y_2 - c_y}{k_y}\right)^2}}{T_{XY}^{-1} \mathbf{p}_2 - \mathbf{P}_{eM}} \end{bmatrix}$$

A equação do plano imagem é, então, dada por

$$\det \left(\begin{bmatrix} P_1 - P_0 \\ P_2 - P_0 \\ P - P_0 \end{bmatrix} \right) = 0 \Leftrightarrow AX + BY + CZ + D = 0 \quad (4.6)$$

onde P é um ponto 3D genérico e $\vec{\mathbf{n}}_i = [A \ B \ C]$ é o vector normal a esse plano.

A intersecção do plano imagem com o plano espelho esquerdo é uma recta 3D, cuja equação, usando (4.6), é

$$\mathbf{r}_i \equiv AX + CZ + D = 0$$

A rotação, \mathbf{R} , entre o mundo e a câmara real é feita em torno desta recta, cujo ângulo de rotação é o ângulo entre o plano espelho esquerdo e o plano imagem. Como o vector normal ao plano espelho é $\vec{\mathbf{n}}_e = [0 \ 1 \ 0]$, então esse ângulo é dado por

$$\varrho = \arccos\left(\frac{\mathbf{n}_i \cdot \mathbf{n}_e}{\|\mathbf{n}_e\| \|\mathbf{n}_i\|}\right)$$

O vector director da recta \mathbf{r}_i , o eixo de rotação, é dado por

$$\mathbf{vd} = \frac{\mathbf{n}_e \wedge \mathbf{n}_i}{\|\mathbf{n}_e \wedge \mathbf{n}_i\|} = [vd_x \ vd_y \ vd_z]^T$$

Com este ângulo e este eixo de rotação podemos confirmar os valores dos elementos da rotação, \mathbf{R} , através do uso dos quaterniões. Assim, sendo o quaternião dado por

$$\mathbf{q} = \begin{bmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{bmatrix} = \begin{bmatrix} \cos\left(\frac{\varrho}{2}\right) \\ vd_x \sin\left(\frac{\varrho}{2}\right) \\ vd_y \sin\left(\frac{\varrho}{2}\right) \\ vd_z \sin\left(\frac{\varrho}{2}\right) \end{bmatrix}$$

com $|\mathbf{q}|^2 = 1$, a rotação que lhe está associada é

$$\mathbf{R} = \begin{bmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1q_2 - q_0q_3) & 2(q_1q_3 + q_0q_2) \\ 2(q_1q_2 + q_0q_3) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(q_2q_3 - q_0q_1) \\ 2(q_1q_3 - q_0q_2) & 2(q_2q_3 + q_0q_1) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{bmatrix}$$

Reconstrução sem recurso aos espelhos

No sistema de aquisição baseado nos espelhos, visto anteriormente, o campo de visão das câmaras virtuais é menor que o da câmara real, pois está confinado à dimensão dos espelhos. Além disso, a imagem relativa a cada espelho capta, para além de parte da cena, a reflexão da imagem captada pelo outro espelho, que não é usada na reconstrução. Assim, de cada imagem captada por estas câmaras virtuais só é aproveitada uma pequena parcela da mesma. Isto faz com que hajam pontos ocultos, não sendo possível a reconstrução completa da cena. Em face desta situação, avança-se para uma captação de imagens sem recurso aos espelhos.

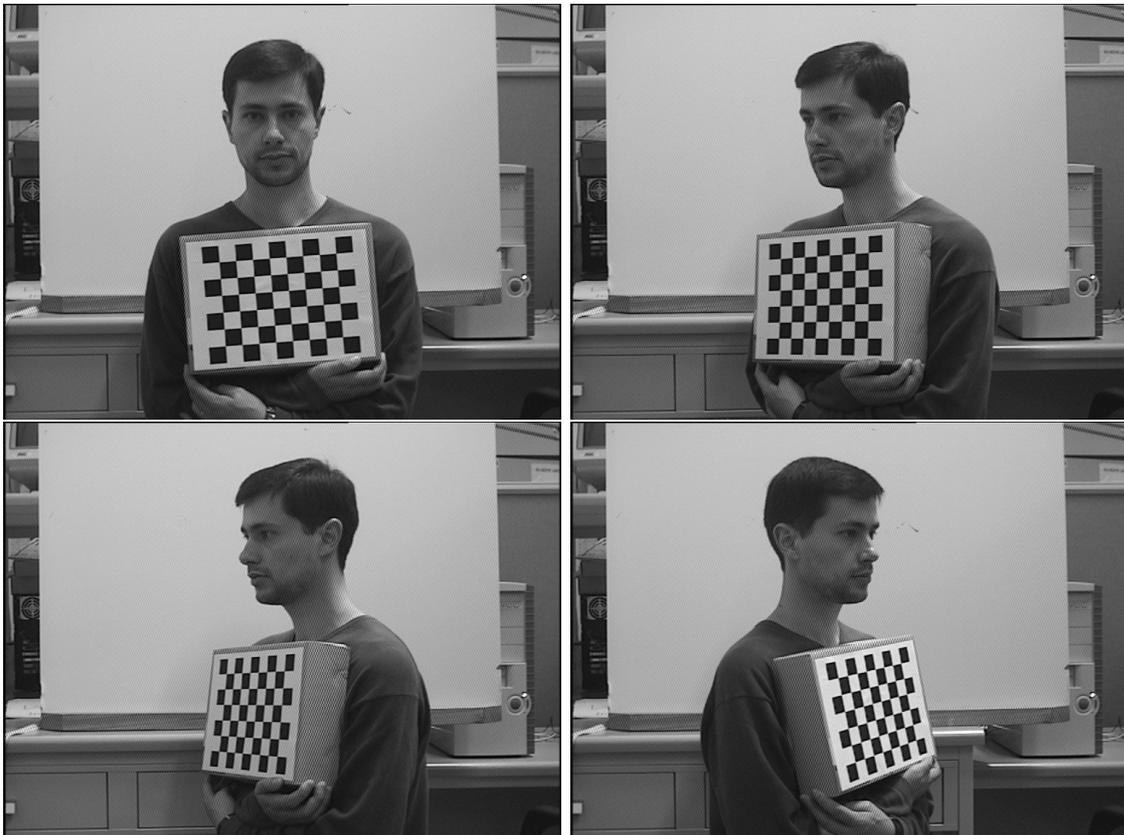


Figura 4.8: *Imagens usadas no processo de reconstrução de superfícies.*

As imagens representadas na figura 4.8, foram captadas por uma câmara que se ía movimentando em relação à cena. As posições das câmaras podem ser vistas



Figura 4.9: Posicionamento das câmaras que adquiriram as imagens da figura 4.8.

na figura 4.9 e confirmadas através das seguintes matrizes de rotação e vectores de translação,

$$\begin{aligned}
 \mathbf{R}_1 &= \begin{bmatrix} -0.075461 & -0.990485 & -0.115086 \\ -0.996900 & 0.077518 & -0.013492 \\ 0.022285 & 0.113711 & -0.993264 \end{bmatrix} & \mathbf{t}_1 &= \begin{bmatrix} 34.103115 \\ 7.813111 \\ 158.628082 \end{bmatrix} \\
 \mathbf{R}_2 &= \begin{bmatrix} 0.023894 & -0.672645 & -0.739580 \\ -0.997506 & 0.033101 & -0.062332 \\ 0.066408 & 0.739225 & -0.670176 \end{bmatrix} & \mathbf{t}_2 &= \begin{bmatrix} 26.615583 \\ 7.830751 \\ 155.473236 \end{bmatrix} \\
 \mathbf{R}_3 &= \begin{bmatrix} 0.094305 & -0.398762 & -0.912193 \\ -0.994858 & -0.003755 & -0.101210 \\ 0.036934 & 0.917047 & -0.397066 \end{bmatrix} & \mathbf{t}_3 &= \begin{bmatrix} 23.056980 \\ 7.939978 \\ 156.151489 \end{bmatrix} \\
 \mathbf{R}_4 &= \begin{bmatrix} -0.201053 & -0.676850 & 0.708133 \\ -0.979270 & 0.157063 & -0.127909 \\ -0.024646 & -0.719170 & -0.694397 \end{bmatrix} & \mathbf{t}_4 &= \begin{bmatrix} 37.531162 \\ 5.871248 \\ 169.935303 \end{bmatrix}
 \end{aligned}$$

Para cada aquisição, foi feita uma calibração da câmara, usando as rotinas de calibração do OpenCV, obtendo-se, além das matrizes anteriores, a seguinte matriz de calibração

$$\mathbf{C} = \begin{bmatrix} 1539.884521 & 0 & 167.125214 \\ 0 & 1433.421997 & 394.105225 \\ 0 & 0 & 1 \end{bmatrix}$$

Para cada imagem, usando todas estas matrizes, foi calculada uma matriz de transformação do tipo (2.7). Convém mencionar que o OpenCV é apenas uma biblioteca

de funções. Isto significa que para se obter a calibração das imagens teve que ser feita uma aplicação.

As imagens binárias usadas para a reconstrução, referentes à figura 4.8, são apresentadas na figura 4.10.



Figura 4.10: *Imagens binárias usadas no processo de reconstrução de superfícies.*

Os resultados da reconstrução, apoiada numa *octree* de nove níveis, podem ser vistos nas figuras 4.11 e 4.12. A primeira figura apresenta os resultados obtidos pelos diferentes níveis da *octree*, mostrando, apenas, os limites dos *voxels* para demonstrar que o número de *voxels* aumenta consideravelmente à medida que se sobe de nível. A figura 4.12 permite visualizar os resultados da reconstrução, para nove níveis da *octree*, segundo diferentes pontos de vista. Os resultados da figura 4.12 apresentam alguns problemas devido à não utilização de imagens da parte de trás da cena 3D. Nestas circunstâncias, a atribuição da textura para os *voxels* referentes a essa zona da cena, é feita com base nos *voxels* da frente. Consequentemente, para algumas das vistas a reconstrução torna-se pouco clara.

O site <http://svr-www.eng.cam.ac.uk/~kykw2/research.html>, permitiu obter as imagens calibradas, representadas na figura 4.13, pelo método descrito em [Wong 01]. A partir das suas matrizes de transformação verificou-se que essas imagens foram captadas por uma única câmara que se ia movimentando em relação à cena ou por várias câmaras estáticas em posições diferentes. As posições das câmaras podem ser confirmadas através da figura 4.14.

As figuras 4.15 e figura 4.16 permitem visualizar os resultados da reconstrução, para oito níveis da *octree* e para dez níveis da *octree* (segundo diferentes pontos de vista), respectivamente.

No exemplo de reconstrução anterior a câmara foi rodando em torno do objecto a recuperar. De forma a mostrar que a este algoritmo se pode aplicar quaisquer imagens do objecto verifique-se o próximo exemplo. As imagens da figura 4.18 foram captadas por uma câmara nas posições apresentadas na figura 4.17.

As figuras 4.19 e figura 4.20 permitem visualizar os resultados da reconstrução, para oito níveis da *octree* e para nove níveis da *octree* (segundo diferentes pontos de vista), respectivamente.

Estes resultados mostram a boa qualidade de reconstrução de superfícies que se obtém com a aplicação do método da escavação do espaço 3D. Pela facilidade de implementação e fracas exigências (uso de apenas uma câmara e um processo de calibração aceitável), este método revela grandes potencialidades de aplicação a qualquer necessidade de reconstrução.

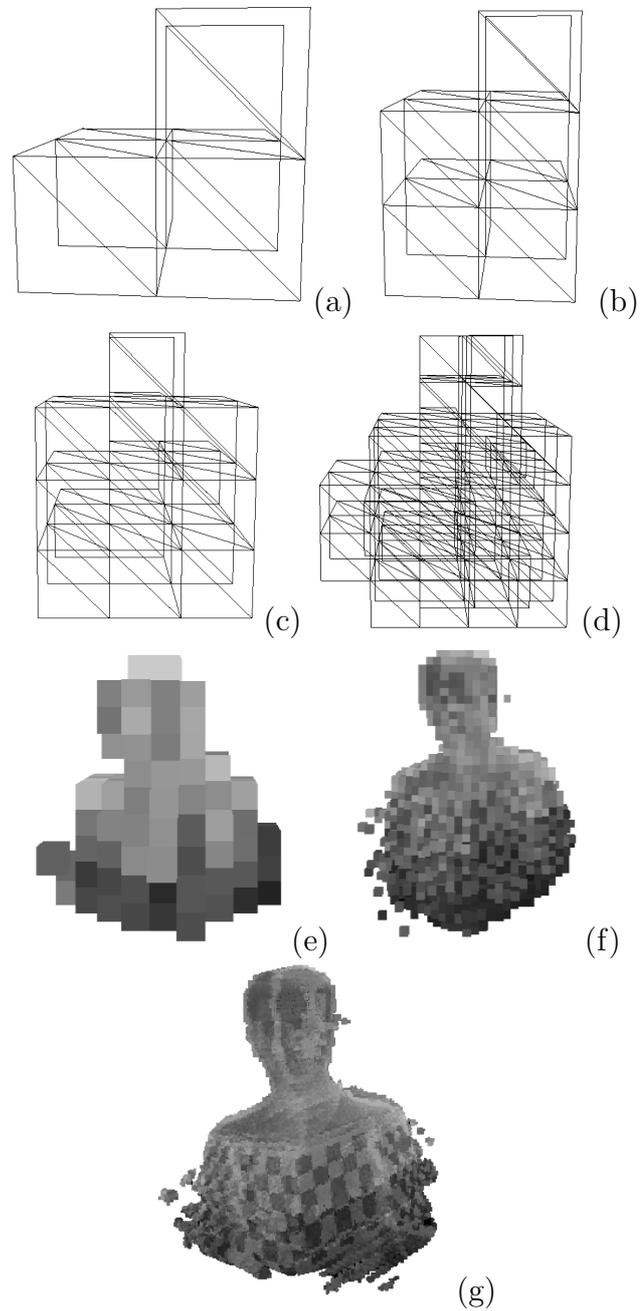


Figura 4.11: Resultados da escavação por níveis (visualização dos limites dos voxels): (a) Nível 2; (b) Nível 3; (c) Nível 4; (d) Nível 5; (e) Nível 6; (f) Nível 7; (g) Nível 8.

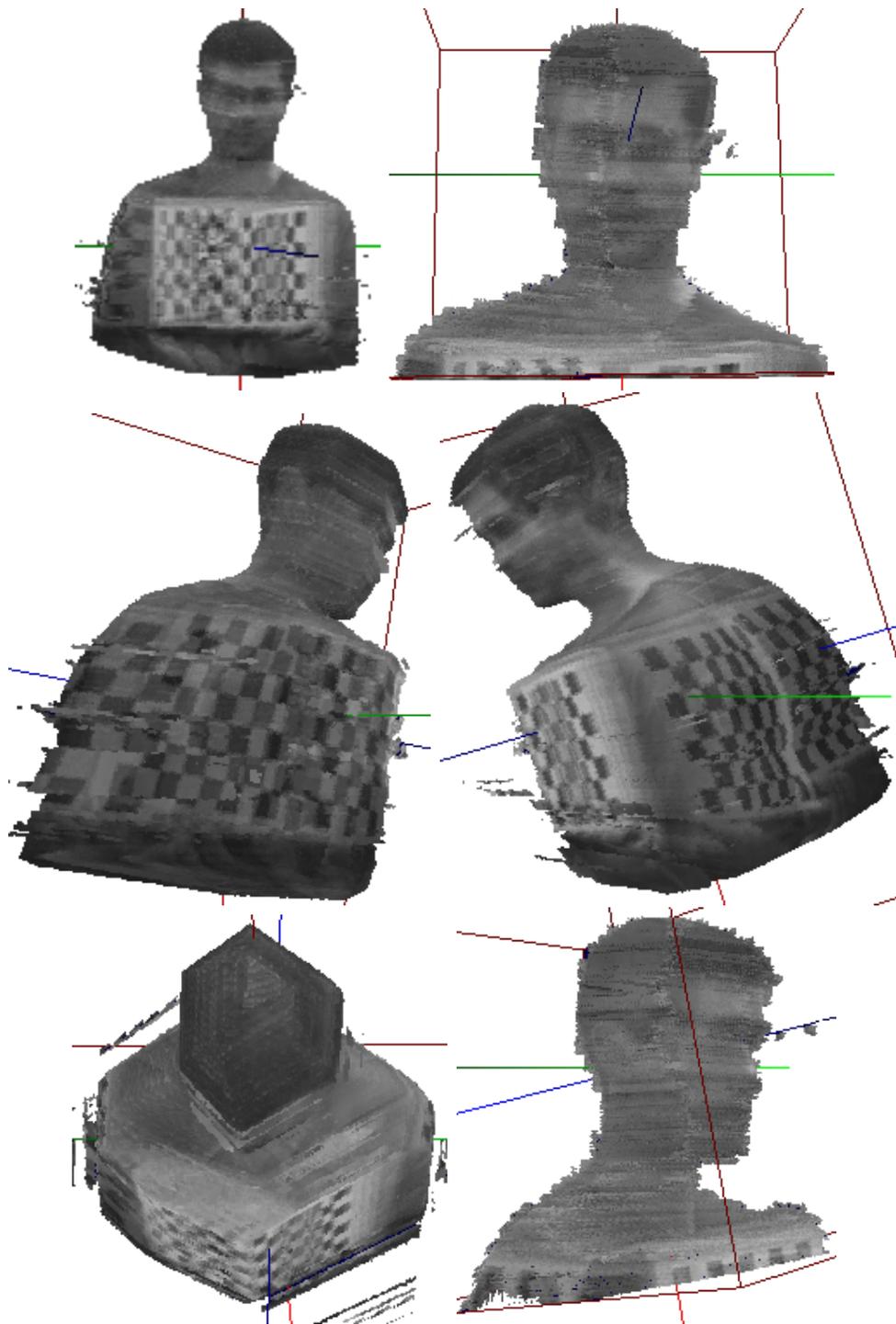


Figura 4.12: *Diferentes vistas do resultados da escavação para nove níveis.*

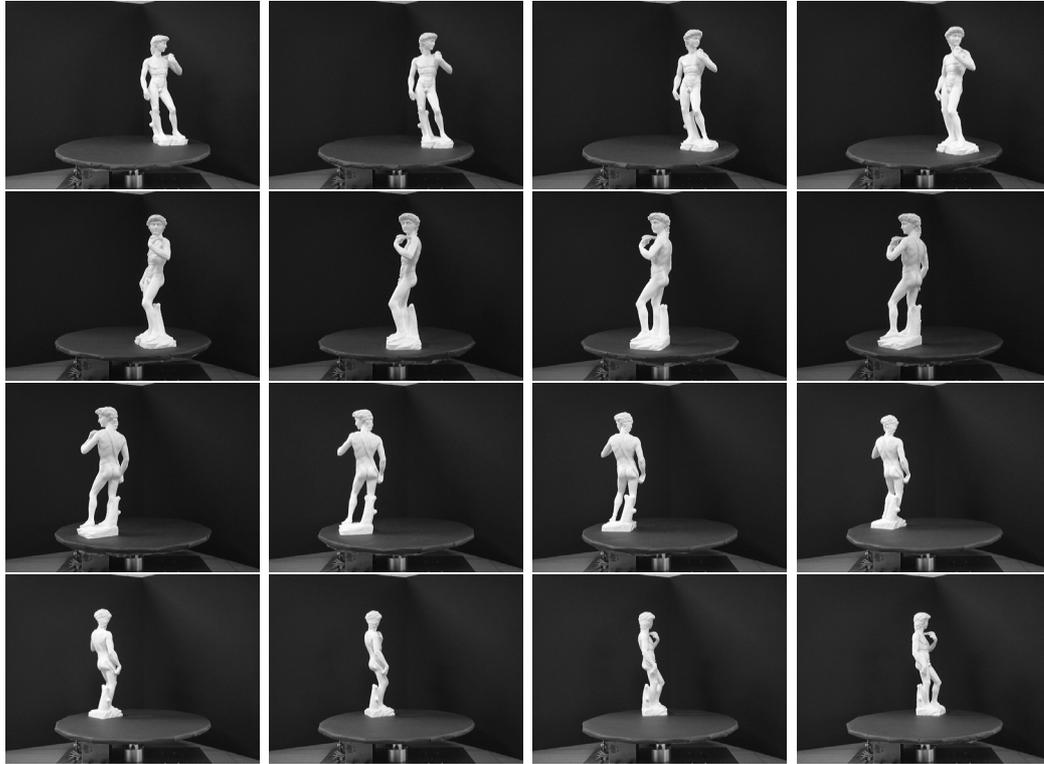


Figura 4.13: *Imagens usadas no processo de reconstrução de superfícies.*

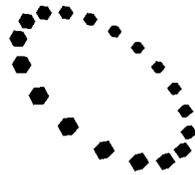


Figura 4.14: *Posicionamento das câmaras que adquiriram as imagens da figura 4.13.*

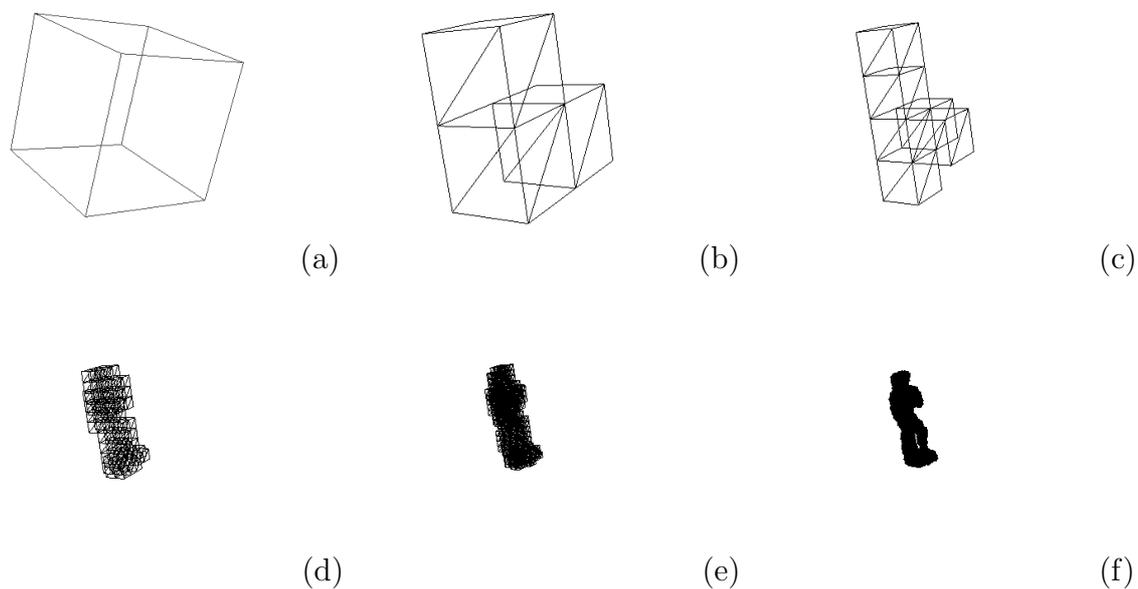


Figura 4.15: Resultados da escavação por níveis: (a) Nível 1; (b) Nível 2; (c) Nível 3; (d) Nível 5; (e) Nível 6; (f) Nível 8.

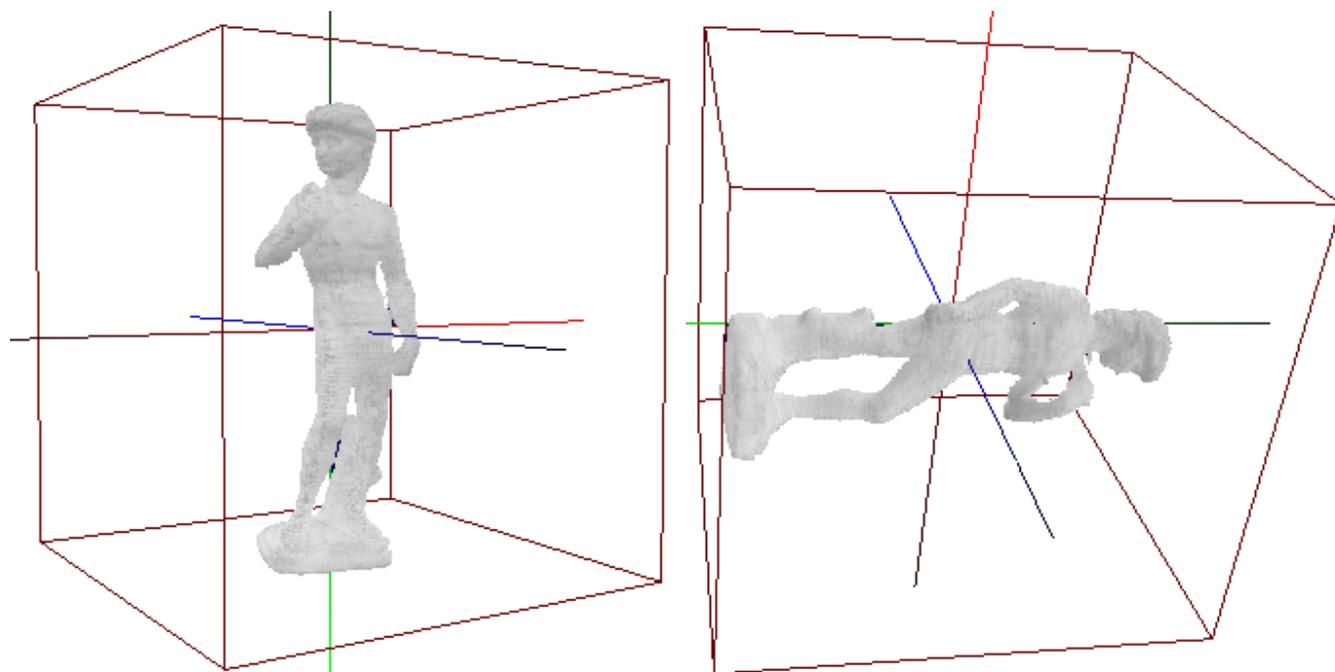


Figura 4.16: Diferentes vistas do resultados da escavação para dez níveis.

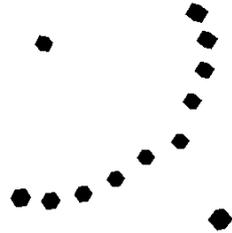


Figura 4.17: *Posicionamento das câmeras que adquiriram as imagens da figura 4.18.*

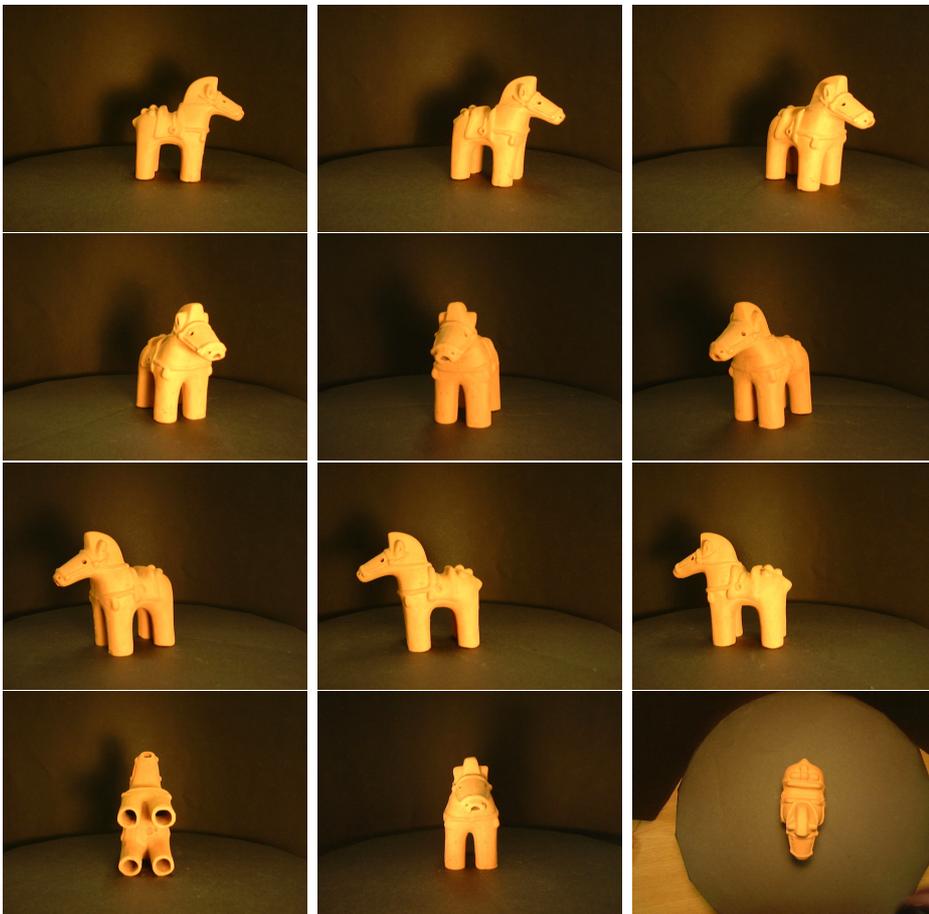


Figura 4.18: *Imagens usadas no processo de reconstrução de superfícies.*

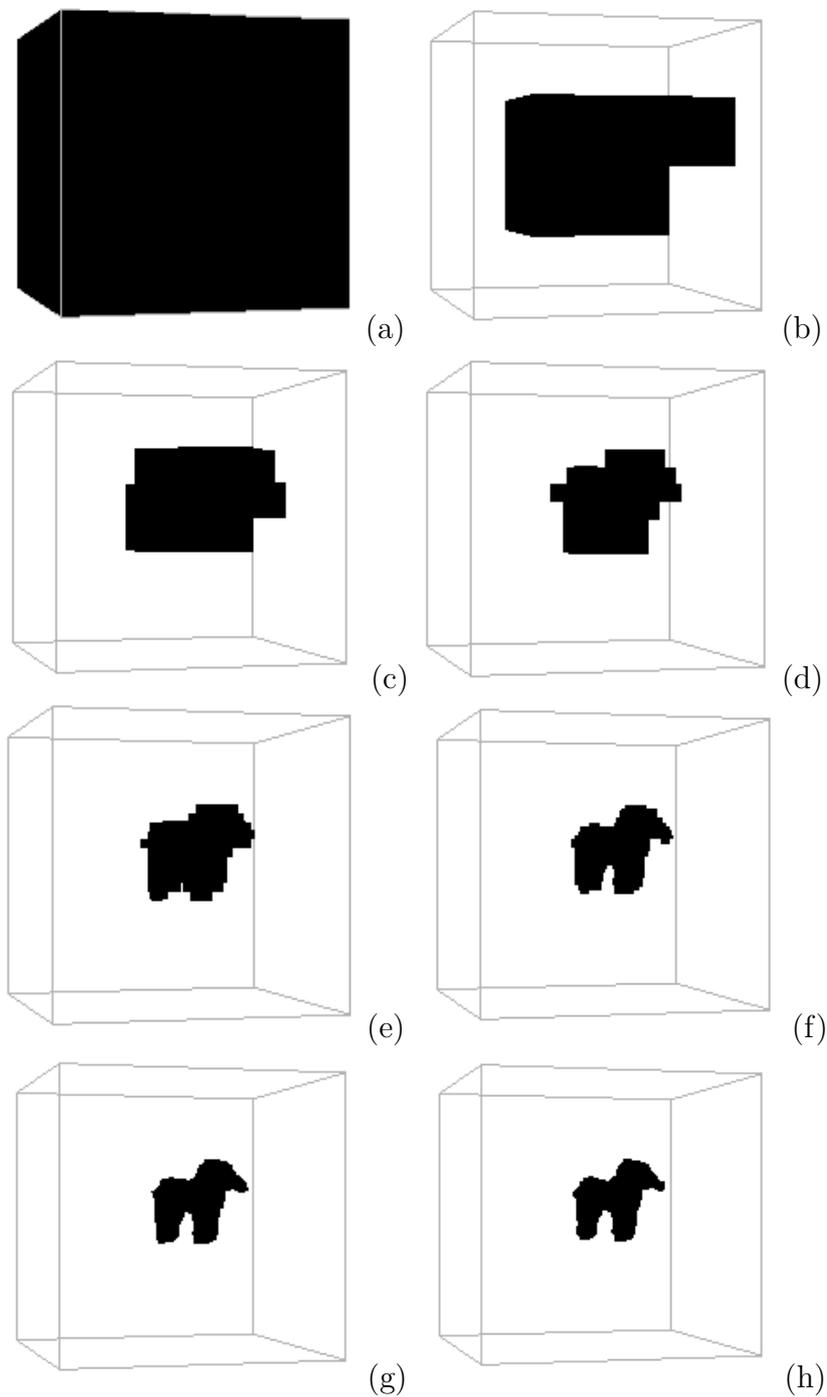


Figura 4.19: Resultados da escavação por níveis: (a) Nível 1; (b) Nível 2; (c) Nível 3; (d) Nível 4; (e) Nível 5; (f) Nível 6; (g) Nível 7; (h) Nível 8.

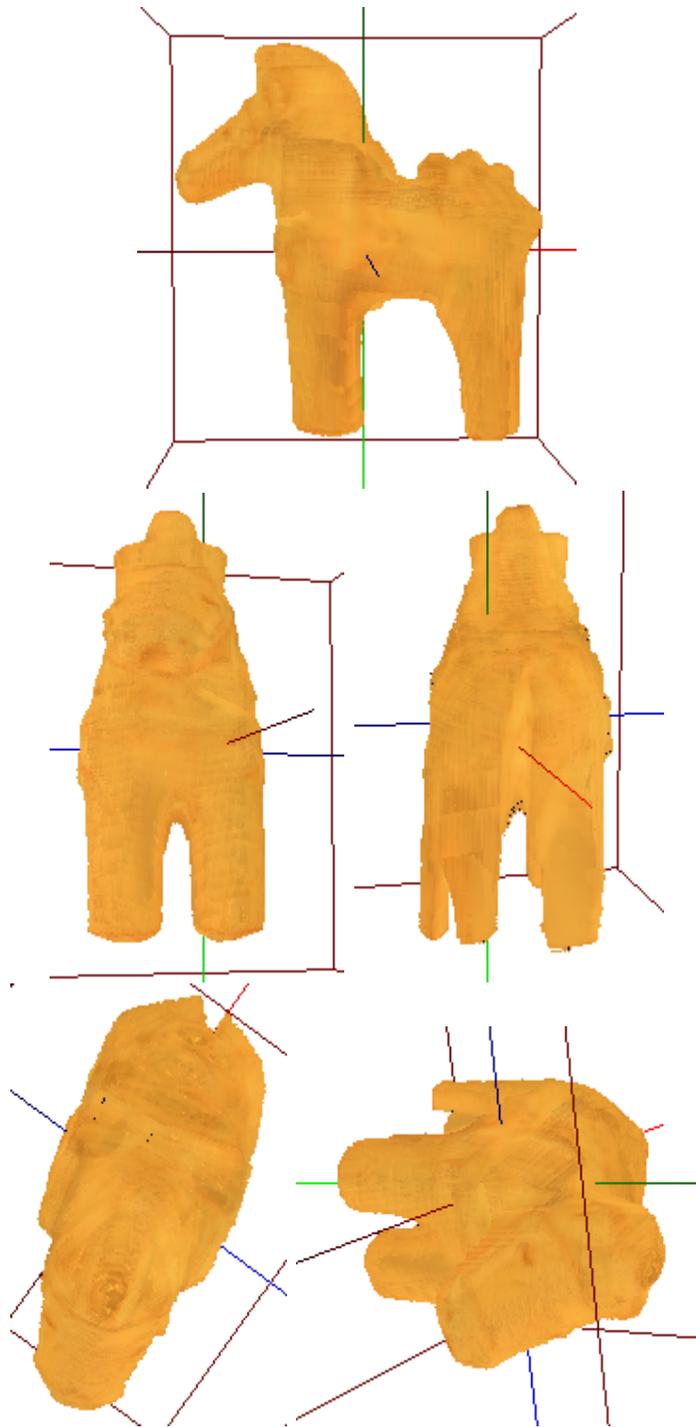


Figura 4.20: *Diferentes vistas do resultados da escavação para nove níveis.*

Capítulo 5

Conclusões e trabalho futuro

Neste capítulo apresentam-se as conclusões do trabalho desenvolvido e são sugeridas linhas de aplicação futuras.

5.1 Conclusões

O potencial contido nas imagens levou o homem a utilizá-las cada vez mais ao longo dos anos. Há, neste momento, um grande dinamismo nas áreas ligadas à imagem, sendo a visão por computador uma delas. Este trabalho inseriu-se exactamente nesta área e pretendeu contribuir, de alguma forma, para o aumento dos conhecimentos nela envolvidos.

Nesta dissertação pretendeu-se implementar processos robustos de recuperação de informação tridimensional a partir de imagens. Por se tratar de um assunto já investigado, procurou-se renovar esse estudo, através da introdução da ideia, menos explorada na área da visão, da utilização de superfícies reflectoras.

Começou-se por estudar formas simples de aquisição de imagens baseadas em espelhos. A partir daí, e de forma a tornar o sistema de aquisição o menos dispendioso possível, definiu-se que se iriam utilizar apenas uma câmara e um conjunto de espelhos planares.

O passo seguinte consistiu na criação e desenvolvimento de um primeiro sistema, que permitiu um envolvimento prático com os problemas da reconstrução, nomeada-

mente as correspondências entre imagens, as calibrações, as triangulações e os métodos numéricos. Embora válido, o trabalho produzido até aqui era um pouco instável. Essa instabilidade devia-se ao facto de a reconstrução se basear exclusivamente na matriz fundamental (dos espelhos apenas se tinha retirado a capacidade da duplicação de imagens).

Assim, passou-se à construção de um novo sistema que retirasse mais do potencial dos espelhos e que não se apoiasse tanto na matriz fundamental. A ideia base surgiu do trabalho de Sameer [Sameer 98], que mostra a possibilidade do conhecimento da geometria epipolar entre duas imagens de uma cena 3D vistas por espelhos. Com este suporte, procurou-se inovar no sentido de facilitar a obtenção da geometria epipolar. Verificou-se que, efectuando uma determinada mudança de base para cada conjunto de projecções correspondentes, bastava conhecer apenas as projecções nas imagens de dois pontos tridimensionais, bem como as suas reflexões num dos espelhos (quatro projecções reais na imagem), em vez das oito projecções reais necessárias no método de Sameer [Sameer 98], para se conhecer a geometria epipolar. A partir daqui, criou-se um sistema para a determinação da estrutura tridimensional dos objectos imóveis da cena. Neste sistema, foram utilizados espelhos na ajuda à triangulação, através de relação existente entre os pontos dos planos espelho e imagem.

Devido às limitações deste sistema, alterou-se a forma de recuperar a informação 3D. Em vez de ser usada a triangulação, foi aplicada a escavação do espaço tridimensional, baseada nas silhuetas. Com esta técnica concluiu-se que os resultados, embora prometedores, necessitavam de mais imagens do que as que o sistema de aquisição usado poderia facultar.

Para verificar todas as potencialidades da técnica da escavação do espaço tridimensional, a captação de imagens calibradas passou a ser feita por um conjunto de câmaras, e não por recurso aos espelhos. Os resultados obtidos demonstraram uma excelente capacidade de reconstrução de cenas complexas, por exemplo, de objectos com buracos.

Em síntese:

- a reconstrução baseada apenas na matriz fundamental e na triangulação requer

uma enorme precisão nas projecções correspondentes;

- a reconstrução baseada nos espelhos e na triangulação é muito mais tolerante a falhas introduzidas pelo conjunto de projecções correspondentes, do que a reconstrução baseada na matriz fundamental;
- a reconstrução baseada nos espelhos e na triangulação ainda pode ser melhorada;
- a reconstrução baseada nos espelhos e na escavação do espaço tridimensional só gera bons resultados quando usa muitas imagens;
- a técnica da escavação do espaço tridimensional, baseada nas silhuetas, fornece excelentes resultados para qualquer cena 3D, desde que use imagens bem calibradas, que abranjam toda a cena.

5.2 Trabalho futuro

Para melhor descrever as propostas de trabalho futuro, faz-se a separação entre as que se aplicam ao trabalho descrito no capítulo 3 e as que são inerentes ao trabalho do capítulo 4.

No que diz respeito aos sistemas apresentados no capítulo 3, conforme foi mostrado na secção 3.3, a imagem captada pelo sistema pode ser dividida em três partes. Cada parte está associada a um centro óptico (centro óptico real, Co_r , centro óptico virtual direito, Co_d , e centro óptico virtual esquerdo, Co_e). Este trabalho apenas reconstruiu pontos nas imagens captadas pelos centros ópticos virtuais (Co_e/Co_d). Assim, uma primeira melhoria a fazer, como trabalho futuro, poderia ser tentar a recuperação de todos os pontos possíveis nos três casos que se conhecem: Co_e/Co_d , Co_e/Co_r e Co_r/Co_d . Esta alteração iria possibilitar a recuperação de pontos 3D, ocultos a alguma das três partes da imagem.

Com a adição de mais espelhos ao sistema de aquisição poder-se-ia tentar um campo de visão o mais alargado possível. A vantagem que poderia advir desse procedimento seria, no seguimento da ideia atrás referida, a obtenção de um maior número de pontos ocultos.

Na sequência disto, uma tarefa interessante a desenvolver seria a junção de todos os pontos reconstruídos (depois de eliminados os repetidos), mostrando-os graficamente, sem conhecimento prévio sobre a sua organização. Isto faria com que, através de uma boa interpolação, fosse possível recuperar também a forma do objecto.

Outra ideia interessante a explorar, numa tentativa de recuperar algo mais do que pontos, seria a da questão das curvas de perfil. Este estudo poderia facilitar a recuperação de superfícies através da utilização do sistema de aquisição apresentado.

Uma outra questão a abordar poderia ser a da utilização do trifocal, já que existem três centros ópticos. Sendo introduzidos mais espelhos, simplificar-se-ia o procedimento com as relações trifocais, eliminando uma sobrecarga de processamento que existiria na continuação do relacionamento dos espelhos dois a dois, através da geometria epipolar.

No que diz respeito ao trabalho do capítulo 4, uma das propostas para melhorar a visualização seria a criação de uma rede de triângulos (*meshes*) a partir dos *voxels*. Isto tornaria o processamento gráfico muito mais leve.

Outra proposta seria a implementação de um sistema multi-ocular de reconstrução, ligado a um sistema de aquisição de imagens automaticamente calibradas em tempo real, que fosse capaz de obter, sem qualquer intervenção humana, a estrutura e textura de qualquer objecto que se lhe aplicasse. Uma das tarefas teria que ser a automatização do processo de binarização das imagens.

Uma ideia interessante a aplicar a todo este trabalho seria a utilização de métodos numéricos não lineares em vez dos utilizados.

Por fim, deixa-se no ar a possibilidade da introdução de emissores laser. Neste caso, crê-se que o processo de reconstrução beneficiaria, principalmente em termos de restrições a usar (ou seja, baixaria a sua complexidade), na medida em que estes emissores podem ser vistos como câmaras que, ao invés de captarem luz, emitem-na.

Apêndice A

A qualidade da correspondência

Não existindo qualquer regra prática para a definição da qualidade da correspondência, nesta dissertação optou-se por baseá-la nas distâncias. Mais especificamente, sendo a correspondência dada através de uma matriz fundamental ou homográfica, essas distâncias são medidas entre os pontos usados para a estimação da matriz e os resultados da aplicação dessa matriz aos ditos pontos.

A.1 Matriz fundamental

Para testar a qualidade de uma matriz fundamental, \mathbf{F} , definida através de um conjunto de pontos correspondentes, (p_{ei}, p_{di}) , $i = 1 \dots n$, nas imagens I_e e I_d , foi utilizada a fórmula seguinte

$$Q_f = \sum_{i=1}^n \frac{d(\mathbf{p}_{di}, \mathbf{F}\mathbf{p}_{ei}) + d(\mathbf{p}_{ei}, \mathbf{F}^T \mathbf{p}_{di})}{2n} \quad (\text{A.1})$$

com Q_f a designar-se qualidade da matriz fundamental, \mathbf{F} .

Em (A.1), d é a distância euclidiana de um ponto à sua recta epipolar correspondente. Note-se que os pontos usados são dados pelas suas coordenadas *pixel* na imagem, fazendo com que a distância d seja medida também em *pixels*. Por outras palavras, Q_f é a distância média da distâncias de todos os pontos usados às suas correspondentes epipolares.

Apresenta-se no apêndice B a informação necessária à utilização, no MATLAB©, da função `mfq.m`, a qual calcula este factor de qualidade.

A.2 Matriz homográfica

Por seu lado, a qualidade de uma matriz homográfica, \mathbf{H} , definida através de um conjunto de pontos correspondentes, (p_{ei}, p_{di}) , $i = 1 \dots n$, nas imagens I_e e I_d , foi testada com a utilização da fórmula seguinte

$$Q_h = \sum_{i=1}^n \frac{d(\mathbf{p}_{di}, \mathbf{H}\mathbf{p}_{ei}) + d(\mathbf{p}_{ei}, \mathbf{H}^{-1}\mathbf{p}_{di})}{2n} \quad (\text{A.2})$$

com Q_h a qualidade da homografia, \mathbf{H} .

Em (A.2), d é a distância euclidiana de um ponto ao resultado da aplicação da homografia ao seu correspondente. Os pontos usados são dados pelas suas coordenadas *pixel* na imagem, pelo que a distância d é também medida em *pixels*. Q_h é, assim, a distância média das distâncias de todos os pontos usados aos pontos reprojectados (obtidos pela aplicação da homografia aos pontos usados).

Remete-se para o apêndice B a informação necessária ao modo de funcionamento, no MATLAB©, da função que calcula este factor de qualidade, `mhq.m`.

Apêndice B

Rotinas para MATLAB®

As rotinas de condicionamentos de dados, de qualidade da correspondência, de resolução de sistemas lineares (mínimos quadrados), de estimação da matriz fundamental, de estimação da matriz homográfica de grau 2 e de eliminação de falsas correspondências usadas na dissertação estão disponíveis em www.isr.uc.pt/~nmartins/work.html.

B.1 Condicionamentos de dados

```
% HTNORM2D Normalize an input matrix of 2D points, as Hartley's proposal. The
% matrix structure is [x1 coordinate, y1 coordinate; ... ;xn coordinate, yn
% coordinate].
%
% Usage : [Ptn,Mnorm] = Htnorm2d(Pt) or Ptn=Htnorm2d(Pt)
% Input : One vector of 2D points, Pt
% Output : The normalized vector, Ptn
%          : The matrix that achieve the normalization, Mnorm

% HTNORM3D Normalize an input matrix of 3D points, as Hartley's proposal. The
% matrix structure is [x1 coordinate, y1 coordinate, z1 coordinate; ... ;xn
% coordinate, yn coordinate, zn coordinate].
%
% Usage : [Ptn,Mnorm] = Htnorm3d(Pt) or Ptn=Htnorm3d(Pt)
% Input : One vector of 3D points, Pt
% Output : The normalized vector, Ptn
```

```

%           : The matrix that achieve the normalization, Mnorm

% IMPRANK Impose the rank n to a square matrix A with rank n+1, minimizing the
% Frobenius norm, if only the determinant of matrix A is not zero.
%
% Usage : Ar = Imprank(A)
%   Input  : The matrix with rank n+1, A
%   Output : The matrix with rank n, Ar

% MFMINASI Calculate the fundamental matrix and the homography between two vectors
%           of 2D points, using algebraic minimization without interaction.
%
% Usage : [F,H] = Mfminasi(Pte,Ptd,e)
%   input  : Two vectors of 2D points, with at least 8 points, Pe and Pd
%           : The known epipole, e(ex ey)
%   output : The fundamental matrix, F
%           : The homography between the vectors of points, H

```

B.2 Qualidade da correspondência

```

% MFQ Calculate the quality factor of a fundamental matrix, which is the mean
% distance of the sum of all distances from the points, in both sets, to its
% correspondent epipolar lines.
%
% Usage : Qf = Mfq(Pt1,Pt2,F)
%   Input  : Two matrices of 2D points, Pt1 and Pt2
%           : The fundamental matrix, F
%   Output : The quality factor, Qf

% MHQ Calculate the quality factors of a homographic matrix (p2=Hp1). One is
% the mean distance of the sum of all distances from the known points and
% the obtained ones, in both sets. The other is the standard deviation of
% the calculated distance.
%
% Usage : [Qh1,Qh2] = Mhq(Pt1,Pt2,H)
%   Input  : Two matrices of 2D points, Pt1 and Pt2

```

```
%           : The homographic matrix, H
%   Output : The quality factors, Qh1 (mean) and Qh2 (standard deviation)
```

B.3 Resolução de sistemas lineares (mínimos quadra dos)

```
% LSQSVD Computes the least-squares solution x of any system Ax=b (including b=0),
%           using SVD method. A is a (m x n) matrix, with m>=n. By itself x=Lsqsvd(A)
%           computes the least-squares solution x of system Ax=0.
```

```
%
```

```
% Usage : x = Lsqsvd(A,b) or x = Lsqsvd(A)
%   input : Coefficient matrix of the system, A
%           : Result vector of the system, b
%   output : Solution vector, x
```

```
% EIGSOL Computes the least-squares solution x of any system Ax=b (including b=0),
%           using the analysis of the eigen values. A is a (m x n) matrix, with m>=n.
%           By itself x=Eigsol(A) computes the least-squares solution of system Ax=0.
```

```
%
```

```
% Usage : x = Eigsol(A,b) or x = Eigsol(A)
%   input : Coefficient matrix of the system, A
%           : Result vector of the system, b
%   output : Solution vector, x
```

```
% PSINVSVD Computes the least-squares solution x of any system Ax=b (with vector b
%           not null), using pseudo-inverse solution, based on the svd method. A is
%           a (m x n) matrix, with m>=n.
```

```
%
```

```
% Usage : x = Psinvsvd(A,b)
%   input : Coefficient matrix of the system, A
%           : Result vector of the system, b
%   output : Solution vector, x
```

```
% PSINVSOL Computes the least-squares solution x of any system Ax=b (with vector b
%           not null), using pseudo-inverse solution, based on the normal equation
%           method. A is a (m x n) matrix, with m>=n.
```

```

%
% Usage : x = Psinvsol(A,b)
%   input  : Coefficient matrix of the system, A
%           : Result vector of the system, b
%   output : Solution vector, x

```

B.4 Estimação da matriz fundamental

```

% MFMIN Calculate the fundamental matrix between two vectors of 2D points,
%   for its minimum case - only 7 linearly independent equations.

```

```

%
% Usage : F = Mfmin(Pe,Pd)
%   input  : Two vectors of 2D points, with at least 7 points, Pe and Pd
%   output : The fundamental matrix, F
%
% see HTNORM2D and MFQ

```

```

% MF8PTA Estimate the fundamental matrix between two vectors of 2D points,
%   using 8-point algorithm.

```

```

%
% Usage : F = Mf8pta(Pe,Pd,type)
%   Input  : Two vectors of 2D points, with at least 8 points, Pte and Ptd
%           : Method to execute the least square if necessary, type
%           type=0          -> svd method (default)
%           type=1          -> eigen values analysis
%           type=2          -> pseudo-inverse by svd method
%           type=other number -> pseudo-inverse by normal equations
%   Output : The fundamental matrix, F
%
% see HTNORM2D, LSQSVD, EIGSOL, PSINVSVD, PSINVSOL, IMPRANK and MFQ

```

```

% MFEST Estimate the fundamental matrix between two vectors of 2D points.

```

```

%
% Usage : F = Mfest(Pe,Pd,type)
%   Input  : Two vectors of 2D points, with at least 7 points, Pe and Pd
%           : Method to execute the least square if necessary, type

```

```

%           type=0           -> svd method (default)
%           type=1           -> eigen values analysis
%           type=2           -> pseudo-inverse by svd method
%           type=other number -> pseudo-inverse by normal equations
%   Output : The fundamental matrix, F
%
% see HTNORM2D, MFMIN, IMPRANK and MF8PTA

```

B.5 Estimação da matriz homográfica de grau 2

```

% MH8PTA Estimate the homographic matrix between two vectors of 2D points,
%   using 8-point algorithm.
%
% Usage : H = Mh8pta(Pe,Pd,type)
%   Input : Two vectors of 2D points, with at least 4 points, Pte and Ptd
%           : Method to execute the least square if necessary, type
%           type=0           -> svd method (default)
%           type=1           -> eigen values analysis
%           type=2           -> pseudo-inverse by svd method
%           type=other number -> pseudo-inverse by normal equations
%   Output : The homographic matrix, H
%
% see HTNORM2D, LSQSVD, EIGSOL, PSINVSVD, PSINVSOL and MHQ

% MHEST Estimate the homographic matrix between two vectors of 2D points.
% Usage : H = Mhest(Pe,Pd,type)
%
%   Input : Two vectors of 2D points, with at least 4 points, Pe and Pd
%           : Method to execute the least square if necessary, type
%           type=0           -> svd method (default)
%           type=1           -> eigen values analysis
%           type=2           -> pseudo-inverse by svd method
%           type=other number -> pseudo-inverse by normal equations
%   Output : The homographic matrix, H
%
% see HTNORM2D and MH8PTA

```

B.6 Eliminação de falsas correspondências

```
% MFRANSAC Calculate the fundamental matrix between Pte and Ptd, using
% RANSAC method to eliminate the false matches.
%
% Usage : [F,Pea,Pda] = Mfransac(Pte,Ptd,iter,th,type);
% Input  : Two vectors of 2D points, with at least 7 points, Pte and Ptd
%          The number of iterations of the process, iter (by default 100)
%          The threshold, th (by default 0.1)
% Output : The fundamental matrix, F
%          Two vectors of 2D points without the false matches, Pea and Pda
%
% see NUMALEAF, MFMIN and MFEST

% MHRANSAC Calculate the homographic matrix between Pte and Ptd, using
% RANSAC method to eliminate the false matches.
%
% Usage : [H,Pea,Pda] = Mhransac(Pte,Ptd,iter,th,type);
% Input  : Two vectors of 2D points, with at least 4 points, Pte and Ptd
%          The number of iterations of the process, iter (by default 100)
%          The threshold, th (by default 0.1)
% Output : The homographic matrix, H
%          Two vectors of 2D points without the false matches, Pea and Pda
%
% see NUMALEAH and MHEST
```

Apêndice C

Propriedades das matrizes antisimétricas

Sejam os vectores tridimensionais $\mathbf{u} = \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix}^T$ e $\mathbf{v} = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix}^T$. O produto vectorial entre estes dois vectores é dado por

$$\mathbf{u} \wedge \mathbf{v} = \begin{bmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = [\mathbf{u}]_{\wedge} \mathbf{v} = \begin{bmatrix} u_2 v_3 - u_3 v_2 \\ u_3 v_1 - u_1 v_3 \\ u_1 v_2 - u_2 v_1 \end{bmatrix} \quad (\text{C.1})$$

Sendo $\mathbf{M} = \begin{bmatrix} \mathbf{m}_1 & \mathbf{m}_2 & \mathbf{m}_3 \end{bmatrix}^T$, uma matriz invertível, 3×3 , com base na equação (C.1), temos

$$(\mathbf{M}\mathbf{u}) \wedge (\mathbf{M}\mathbf{v}) = \begin{bmatrix} (\mathbf{m}_2^T \mathbf{u})(\mathbf{m}_3^T \mathbf{v}) - (\mathbf{m}_3^T \mathbf{u})(\mathbf{m}_2^T \mathbf{v}) \\ (\mathbf{m}_3^T \mathbf{u})(\mathbf{m}_1^T \mathbf{v}) - (\mathbf{m}_1^T \mathbf{u})(\mathbf{m}_3^T \mathbf{v}) \\ (\mathbf{m}_1^T \mathbf{u})(\mathbf{m}_2^T \mathbf{v}) - (\mathbf{m}_2^T \mathbf{u})(\mathbf{m}_1^T \mathbf{v}) \end{bmatrix} \quad (\text{C.2})$$

Por definição, a adjunta da matriz, \mathbf{M} e a matriz dos cofactores de \mathbf{M} é

$$\text{adj}(\mathbf{M}) = [\text{cof}(\mathbf{M})]^T = \det(\mathbf{M})\mathbf{M}^{-1}$$

Logo, através de simples cálculos numéricos pode-se concluir que

$$\text{cof}(\mathbf{M}) = [\text{adj}(\mathbf{M})]^T = \det(\mathbf{M})\mathbf{M}^{-T} = \begin{bmatrix} (\mathbf{m}_2 \wedge \mathbf{m}_3)^T \\ (\mathbf{m}_1 \wedge \mathbf{m}_3)^T \\ (\mathbf{m}_1 \wedge \mathbf{m}_2)^T \end{bmatrix} \quad (\text{C.3})$$

Com base na equação (C.1) e (C.3), sabe-se que

$$\text{cof}(\mathbf{M})(\mathbf{u} \wedge \mathbf{v}) = \begin{bmatrix} (\mathbf{m}_2 \wedge \mathbf{m}_3)^T \\ (\mathbf{m}_1 \wedge \mathbf{m}_3)^T \\ (\mathbf{m}_1 \wedge \mathbf{m}_2)^T \end{bmatrix} \begin{bmatrix} u_2 v_3 - u_3 v_2 \\ u_3 v_1 - u_1 v_3 \\ u_1 v_2 - u_2 v_1 \end{bmatrix}$$

ou de outra forma, fazendo as multiplicações,

$$\begin{bmatrix} (\mathbf{m}_2 \wedge \mathbf{m}_3)^T \\ (\mathbf{m}_1 \wedge \mathbf{m}_3)^T \\ (\mathbf{m}_1 \wedge \mathbf{m}_2)^T \end{bmatrix} (\mathbf{u} \wedge \mathbf{v}) = \begin{bmatrix} (\mathbf{m}_2^T \mathbf{u})(\mathbf{m}_3^T \mathbf{v}) - (\mathbf{m}_3^T \mathbf{u})(\mathbf{m}_2^T \mathbf{v}) \\ (\mathbf{m}_3^T \mathbf{u})(\mathbf{m}_1^T \mathbf{v}) - (\mathbf{m}_1^T \mathbf{u})(\mathbf{m}_3^T \mathbf{v}) \\ (\mathbf{m}_1^T \mathbf{u})(\mathbf{m}_2^T \mathbf{v}) - (\mathbf{m}_2^T \mathbf{u})(\mathbf{m}_1^T \mathbf{v}) \end{bmatrix} \quad (\text{C.4})$$

Juntando as equações (C.2), (C.3) e (C.4) conclui-se que

$$(\mathbf{M}\mathbf{u}) \wedge (\mathbf{M}\mathbf{v}) = \det(\mathbf{M})\mathbf{M}^{-T}(\mathbf{u} \wedge \mathbf{v}) \quad (\text{C.5})$$

Aplicando o conhecimento contido na equação (C.1) em (C.5), esta última pode ser escrita como

$$[\mathbf{M}\mathbf{u}]_{\wedge} \mathbf{M}\mathbf{v} = \det(\mathbf{M})\mathbf{M}^{-T}[\mathbf{u}]_{\wedge} \mathbf{v} \quad (\text{C.6})$$

Eliminando o vector \mathbf{v} de ambos os membros da equação (C.6), e assumindo que $\mathbf{M}\mathbf{u} = \mathbf{t}$ (ou $\mathbf{u} = \mathbf{M}^{-1}\mathbf{t}$), obtém-se

$$[\mathbf{t}]_{\wedge} \mathbf{M} = \det(\mathbf{M})\mathbf{M}^{-T}[\mathbf{M}^{-1}\mathbf{t}]_{\wedge}$$

Bibliografia

- [3D Matic www] "3D-MATIC", www.faraday.gla.ac.uk, "Faraday Home Frameset", EPSRC Sponsored Faraday Partnership Administered by the Department of Computing Science, University of Glasgow, last revision - 31/08/00, University of Glasgow, 2000.
- [3D Murale www] "3D-MURALE project", www.prip.tuwien.ac.at/Research/Murale/.
- [Agapito 97] L. de Agapito, D. Q. Huynh and M. J. Brooks, "Self-calibrating a stereo head: An error analysis in the neighbourhood of degenerate configurations", *Proceedings of Fifth International Conference on Computer Vision (ICCV'97)*, 1997.
- [Ahuja 89] N. Ahuja and J. Veenstra, "Generating octrees from object silhouettes in orthographic views", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 11(2), pp. 137-149, February, 1989.
- [Almeida 99] Luís Almeida, Jorge Dias, "Dense Depth Maps using Stereo Vision Head", *Proceedings of SIRS99 the 7th International Symposium on Intelligent Robotic Systems*, Coimbra, July 20-23, 1999.
- [Ayache 96] J. Felmar and N. Ayache, "Rigid, Affine and Locally Affine Registration of Free-Form Surfaces", *IJCV*, Vol. 18(2), pp. 99-119, 1996.
- [Baker 98] Simon Baker and Shree K. Nayar, "A theory of catadioptric image formation", *Proceedings of the Sixth International Conference on Computer Vision (ICCV'98)*, Bombay, India, January, 1998.
- [Batista 99] J. Batista, H. Araújo and A. de Almeida, "Iterative Multi-Step Explicit Camera Calibration", *IEEE Trans. on Robotics and Automation*, vol.15, no.5, October, 1998.
- [Boyer 97] E. Boyer and M. O. Berger, "3D surface reconstruction using occluding contours", *Int. Journal of Computer Vision*, Vol. 22(3), pp. 219-233, March 1997.
- [Chen 88] H. H. Chen and T. S. Huang, "A survey of construction and manipulation of octrees", *Computer Vision, Graphics and Image Processing*, Vol. 43(3), pp. 409-431, September, 1988.

- [Chien 86] C. H. Chien and J. K. Aggarwal, "Volume/surface octrees for the representation of three-dimensional objects", *Computer Vision, Graphics and Image Processing*, Vol. 36(1), pp. 100-113, October, 1986.
- [Cipolla 92] R. Cipolla and A. Blake, "Surface shape from the deformation of apparent contours", *Int. Journal of Computer Vision*, Vol. 9(2), pp. 83-112, 1992.
- [Devernay 95] F. Devernay and O. Faugeras, Rapport technique, n^o2725, Institut National de Recherche en informatique et en automatique, Novembre, 1995.
- [Dias 91] J. Dias, A. de Almeida, H. Araújo and J. Batista, "Improving camera calibration by using multiple frames in hand-eye robotic systems", IROS 91, Osaka, Japan, November, 1991.
- [Dias 94] Jorge Dias, "Reconstrução tridimensional utilizando visão dinâmica", dissertação de doutoramento, Departamento de Engenharia Electrotécnica da Faculdade de Ciências e Tecnologia da Universidade de Coimbra, 1994.
- [Datta 94] Biswa Nath Datta, "Numerical Linear Algebra and applications", Brooks/Cole Publishing Company, ISBN 0-534-17466-3, 1994.
- [Fischler 81] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography", *Communications of the ACM*, Vol. 24, pp. 381-395, June, 1981.
- [Faugeras 87] O. Faugeras and G. Toscani, "Camera calibration for 3D computer vision", In *Proceedings of international Workshop on machine vision and machine intelligence*, Tokyo, Japan, 1987.
- [Faugeras 92] O. Faugeras, "What can be seen in three dimensions with an uncalibrated stereo rig?", In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision*, pp. 563-578, Santa Margherita Ligure, Italy, Springer-Verlag, May, 1992.
- [Faugeras 93] Olivier Faugeras, "Three-Dimensional Computer Vision: a Geometric Viewpoint", MIT Press, 1993.
- [Ferreira 00] J. Ferreira and J. Dias, "A 3D Scanner - Three-Dimensional Reconstruction From Multiple Images", *Proceedings CONTROL 2000 - Portuguese Conf. on Automatic Control - Student Forum*, pp. 690-695, University of Minho, Portugal, 2000.
- [Francisco 94] A. Francisco, "Active structure acquisition by continuous fixation movements", Ph. D. dissertation, Dept. of Numerical Analysis and Computing Science, KTH (Royal Institute of Technology), June 1994. ISRN KTH/NA/P-94/17-SE.
- [Fu 87] K. S. Fu, R. C. Gonzalez and C. S. G. Lee, "Robotics: Control, sensing, vision and intelligence", McGraw-Hill, ISBN 0-07-100421-1, 1987.

- [Giblin 87] P. J. Giblin and R. S. Weiss, "Reconstructions of surfaces from profiles", *In Proc. 1st Int. Conference on Computer Vision*, pp. 136-144, London, UK, June 1987.
- [Goshtasby 93] Ardeshir Goshtasby and William A. Gruver, "Design of a single-lens stereo camera system", *Pattern recognition*, Vol. 26, n^o6, pp. 923-937, 1993.
- [Hartley 92] R. I. Hartley, R. Gupta and T. Chang, "Stereo from uncalibrated cameras", *In Proceedings of the Conference on Computer Vision and Pattern Recognition*, pp. 761-764, Urbana-Champaign, Illinois, U.S.A., 1992.
- [Hartley 95] R. I. Hartley, "In defence of the 8-point algorithm", *Proceedings of Fifth International Conference on Computer Vision (ICCV'95)*, pp. 1064-1070, Cambridge, Massachusetts, June, 1995.
- [Hartley 95a] R. I. Hartley and P. Sturm, "Triangulation", *Lecture Notes in Computer Science*, Vol. 970, pp. 190-197, 1995.
- [Hartley 98] R. I. Hartley, "Minimizing Algebraic Error", *In IEEE Int. Conf. Computer Vision*, Bombay, January, 1998.
- [Hartley 00] R. I. Hartley and A. Zisserman, "Multiple view geometry in computer vision", ISBN 0521623049, University press, Cambridge, UK, 2000.
- [Hecht 74] E. Hecht and A. Zajac, "Optics", Addison-Wesley, 1974.
- [Hong 85] T. H. Hong and M. O. Shneier, "Describing a robot's workspace using a sequence of views from a moving camera", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 7(6), pp. 721-726, November, 1985.
- [Hoppe 94] H. Hoppe, T. DeRose, T. Duchamp, M. Halstead, H. Jin, J. McDonald, J. Schweitzer, W. Stuetzle. "Piecewise Smooth Surface Reconstruction", *Computer Graphics (SIGGRAPH 1994 Proceedings)*, pp. 295-302, 1994.
- [Improofs www] "IMPROOFS project", www.esat.kuleuven.ac.be/~konijn/improofs.html.
- [Jackins 80] C. L. Jackins and S. L. Tanimoto, "Octrees and their use in representing three-dimensional objects", *Computer Graphics Image Processing*, Vol. 14(3), pp. 249-270, November, 1980.
- [Kim 98] W. S. Kim, H.S. Cho, "A novel omnidirectional image sensing system for assembling parts with arbitrary cross-section shapes", *IEEE/ASME Trans. Mechatronics* 3, 1998.
- [Koenderink 89] J. J. Koenderink and A. J. Van Doorn, "Affine structure from motion", Technical report, Utrecht University, Utrecht, The Netherlands, October 1989.

- [Kutulakos 99] K. N. Kutulakos and S. M. Seitz, "A theory of shape by space carving", *In Proc. 7th Int. Conf. on Computer Vision*, pp. 307-314, Kerkyra, Greece, September, 1999.
- [Levoy 00] M. Levoy, K. Pulli, B. Curless, S. Rusinkiewicz, D. Koller, L. Pereira, M. Ginzton, S. Anderson, J. Davis, J. Ginsberg, J. Shade and D. Fulk, "The Digital Michelangelo Project: 3D scanning of large statues", *SIGGRAPH 2000*, New Orleans, LA, pp. 24-28 July, 2000.
- [Li 96] F. Li, M. Brady and C. Wiles, "Fast computation of the fundamental matrix for an active stereo vision system", *Proceedings of European Conference on Computer Vision*, Lecture notes in computer science 1065, Vol. 1, pp. 157-166, Cambridge, UK, B. Buxton and R. Cipolla editors, 1996.
- [Longuet-Higgins 81] H. Longuet-Higgins, "A computer algorithm for reconstructing a scene from two projections", *Nature*, 293:133-135, 1981.
- [Luong 93] Quang-Tuan Luong, Rachid Deriche, Olivier Faugeras, Theo Papadopoulos, "On determining the fundamental matrix: Analysis of different methods and experimental results", Technical report *n°1894*, INRIA, April, 1993.
- [Márquez] J. Márquez, P. Anglade, E. Hirsch e Y. Agid, "Développement d'un Système d'Analyse d'Images 3D pour la Microscopie Electronique: Application à l'étude des Noyaux des Neurones Dopaminergiques de la Substance Nigra dans la Maladie de Parkinson", Collaboration with the INSERM U289 laboratory, Hôpital de la Salpêtrière, Paris.
- [Márquez 00] J. Márquez, T. Bousquet, I. Bloch, F. Schmitt e C. Grangeat, "Laser-Scan Acquisition of Head Models for Dosimetry of Hand-held Mobile Phones", www.tsi.enst.fr/marquez/ABS-BEMS2000/text-BEMS2000.html.
- [Marr 76] D. Marr and T. Poggio, "Cooperative computation of stereo disparity", *Science*, 194, pp. 283-287, 1976.
- [Marr 79] D. Marr and T. Poggio, "A computational theory of human stereo vision", *Proceedings Royal Society of London*, Vol. B 204, 1979.
- [Martin 83] W. N. Martin and J. K. Aggarwal, "Volumetric descriptions of objects from multiple views", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 5(2), pp. 150-158, March, 1983.
- [Martins 98] N. Martins and J. Dias, "A Visual Inspection System Based on Fundamental Matrix and Structured Light", *WAC98 - World Automation Congress - Special on session Sixth International Symposium on Manufacturing and Applications*, Anchorage, Alaska, USA, May 10-14, 1998.

- [Martins 99] N. Martins and J. Dias, "Visual Inspection based on mirror images", *Proceedings of SIRS99 the 7th International Symposium on Intelligent Robotic Systems*, Coimbra, Portugal, July 20-23, 1999.
- [Martins 00] N. Martins and J. Dias, "Visual Inspection and 3D Reconstruction Based on Image-To-Mirror Planes", *IEEE International Conference on Industrial Technology - IEEE ICIT'2000*, Goa, India, January 19-22, 2000.
- [Martins 01] J. Ferreira, N. Martins, L. Agnelo and J. Dias, "Imagiologia Tridimensional Digital para Construção de Protótipos Industriais", *Robótica*, No. 45, pp. 28-34, 4o. Trimestre 2001.
- [Mathieu 95] Hervé Mathieu et Frédéric Devernay, "Système de miroirs pour la stéréoscopie", Rapport technique, n°172, Institut National de Recherche en informatique et en automatique, Juin, 1995.
- [Mitsumoto 92] H. Mitsumoto, S. Tamura, K. Okazaki, N. Kajimi and Y. Fukui, "3D reconstruction using mirror images based on a plane symmetry recovery method", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, n°9, pp. 941-945, 1992.
- [Mohr 90] R. Mohr and E. Arbogast, "It can be done without camera calibration", Technical report RR 805-I-IMAG 106 LIFIA, Lifa-Imag, February 1990.
- [Mohr 91] R. Mohr and L. Morin, "Relative positioning from geometric invariants", In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pp. 139-144, Maui, Hawaii, U.S.A., June 1991.
- [Mundy 92] J. L. Mundy and A. Zisserman, editors, "Geometric invariance in computer vision", MIT Press, 1992.
- [Nayar 88] S. K. Nayar, "Sphereo: Recovering depth using a single camera and two specular spheres", In *Proceedings of SPIE: Optics, illumination and image sensing for machine vision II*, November, 1988.
- [Nayar 98] Joshua Gluckman and Shree K. Nayar, "Planar catadioptric stereo: geometry and calibration", New York, USA, 1998.
- [Noborio 88] H. Noborio, S. Fukuda and S. Arimoto, "Construction of the octree approximating three-dimensional objects by using multiple views", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 10(6), pp. 769-782, November, 1988.
- [Poelman 94] C. J. Poelman and T. Kanade, "A paraperspective factorisation method for shape and motion recovery", In J. O. Eklundh, editor, *Proceedings of the 3rd European Conference on Computer Vision*, pp. 97-108, Stockholm, Sweden, May 1994.

- [Pollard 85] S. B. Pollard, J. E. W. Mayhew and J. P. Frisby, "PMF: A stereo correspondence algorithm using a disparity gradient limit", *Perception*, 14, pp. 449-470, 1985.
- [Potmesil 87] M. Potmesil, "Generating octree models of 3D objects from their silhouettes in a sequence of images", *Computer Vision, Graphics and Image Processing*, Vol. 40(1), pp. 1-29, October, 1987.
- [Recipes 92] W. H. Press, B. P. Flannery, S. A. Teukolsky and W. T. Vetterling, "Numerical recipes in c: The art of scientific computing", Cambridge University Press, 1992.
- [Rees 70] D. W. Rees, "Panoramic television viewing system", United States patent *n*^o3505465, April, 1970.
- [Rothwell 95] C. Rothwell, G. Csurka, O. Faugeras, "A comparison of projective reconstruction methods for pairs of views", Rapport technique, *n*^o2538, Institut National de Recherche en informatique et en automatique, Avril, 1995.
- [Sameer 98] Sameer A. Nene and Shree K. Nayar, "Stereo with mirrors", *Proceedings of the Sixth International Conference on Computer Vision (ICCV'98)*, Bombay, India, January, 1998.
- [Silva 94] Jorge Alves da Silva, "Aquisição e processamento de informação tridimensional", dissertação de doutoramento, Departamento de Engenharia Electrotécnica e de Computadores da Faculdade de Engenharia da Universidade do Porto, 1994.
- [Sparr 91] G. Sparr, "Projective invariants for affine shapes of point configurations", In *Proceedings of the DARPA-ESPRIT Workshop on Applications of Invariants in Computer Vision*, pp. 151-170, Reykjavik, Iceland, March 1991.
- [Srivastava 90] S. K. Srivastava and N. Ahuja, "Octree generation from object silhouettes in perspective views", *Computer Vision, Graphics and Image Processing*, Vol. 49(1), pp. 68-84, January, 1990.
- [Stewart 95] C. V. Stewart, "MINPRAN, a new robust estimator for computer vision", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 17, *n*^o10, pp. 925-938, 1995.
- [Szeliski 93] R. Szeliski, "Rapid octree construction from image sequences", *Computer Vision, Graphics and Image Processing*, Vol. 58(1), pp. 23-32, July, 1993.
- [Szeliski 98] R. Szeliski and R. Weiss, "Robust shape recovery from occluding contours using a linear smoother", *Int. Journal of Computer Vision*, Vol. 28(1), pp. 27-44, June, 1998.
- [Tomasi 91] C. Tomasi and T. Kanade, "Factoring image sequences into shape and motion", In *Proceedings of IEEE Workshop on Visual Motion, Princeton, New Jersey*, pp. 21-28, Los Alamitos, California, U.S.A., 1991.

- [Torr 96] P. H. S. Torr and D. W. Murray, "The development and comparison of robust methods for estimating the fundamental matrix", In *Int. J. Computer Vision*, Kluwer Academic Publishers, Boston, July, 1996.
- [Tsai 84] R. Y. Tsai and T. S. Huang, "The perspective view of three points", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 6:13-27, 1984.
- [Vaillant 92] R. Vaillant and O. D. Faugeras, "Using extremal boundaries for 3D object modeling", *IEEE Trans. On Pattern Analysis and Machine Intelligence*, Vol. 14(2), pp. 157-173, February 1992.
- [Visible www] "The Visible Human Project", www.nlm.nih.gov/research/visible.
- [Weinshall 87] D. Weinshall, "Qualitative depth and shape from stereo, in agreement with psychophysical evidence", Massachusetts Institute of Technology Artificial Intelligence Laboratory and Center for Biological Information Processing Whitaker College, December, 1987.
- [Weinshall 93] D. Weinshall, "Model-based invariants for 3D vision", *International Journal of Computer Vision*, Vol. 10(1), pp. 27-42, 1993.
- [Weiss 97] L. Weiss, "SFF Processes", JTEC/WTEC Panel Report on Rapid Prototyping in Europe and Japan, March 1997.
- [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, 1986.
- [Wong 99] K.-Y. K. Wong, P. R. S. Mendonça, and R. Cipolla, "Reconstruction and estimation from apparent contours under circular motion", In *T. Pridmore and D. Elliman, ed., Proc. British Machine Vision Conference*, Vol. 1, pp. 83-92, Nottingham, UK, September, 1999.
- [Wong 01] K.-Y. K. Wong and R. Cipolla, "Structure and motion from silhouettes", In *proc. 8th Int. Conf. on computer vision*, Vol. 2, pp. 217-222, Vancouver, B C, Canada, July, 2001.
- [Wrobel 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.
- [Zhang 94] Z. Zhang, Rachid Deriche, Olivier Faugeras and Quang-Tuan Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry", Technical report *n°2273*, INRIA, May, 1994.
- [Zhang 96] Z. Zhang, "Determining the epipolar geometry and its uncertainty: A review", Technical report *n°2927*, INRIA, July, 1996.
- [Zhang 97] Z. Zhang and G. XU, "A general expression of the fundamental matrix for both perspective and affine cameras", *Proceedings of Fifteenth International Joint Conference on Artificial Intelligence*, Nagoya, Japan, August 23-29, 1997.

- [Zhang 00] Z. Zhang, "A flexible new technique for camera calibration", IEEE Transactions on Pattern Analysis and Machine Intelligence, No. 22, Vol. 11,; pp. 1330-1334, 2000.
- [Zhang 02] Z. Zhang, "Camera Calibration with One-Dimensional Objects", ECCV, Springer-Verlag, Berlin, pp. 161-174, 2002.