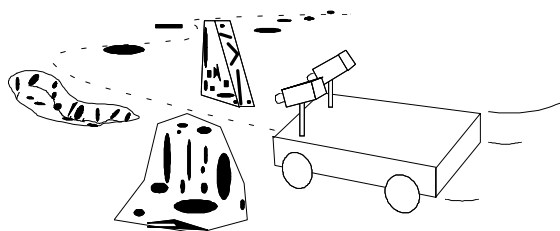


# UNIVERSIDADE TÉCNICA DE LISBOA INSTITUTO SUPERIOR TÉCNICO



## Visão para Robótica Móvel: Detecção de Obstáculos sobre Pavimento Plano

José António da Cruz Pinto Gaspar  
(Licenciado)

Dissertação para a obtenção do grau de Mestre em  
Engenharia Electrotécnica e de Computadores

Setembro de 1994

Tese realizada sob a orientação de  
**João José dos Santos Sentieiro**

Professor Associado do  
Departamento de Engenharia Electrotécnica e de Computadores  
INSTITUTO SUPERIOR TÉCNICO

## **AGRADECIMENTOS**

A realização de um curso de mestrado significa, antes de mais, um conjugar de esforços e de boa vontade, de um vasto conjunto de pessoas, a quem quero justamente agradecer.

Começo naturalmente por agradecer ao Professor João Sentieiro e ao Eng.º José Alberto dos Santos Victor, o infinito apoio e orientação que sempre me prestaram.

Agradeço aos meus colegas do grupo de visão, Franc van Trigt, Fernando Pires, José Varandas, Paulo Soberano, Alexandre, César, Marco e Vítor, as sugestões, as discussões técnicas e filosóficas, enfim, todas as trocas de opiniões e experiências.

Agradeço também aos meus colegas de mestrado, João Hespanha, Constantino Lagoa, Daniel Fryxell, Paulo Pereira e Luís Bernardo, o sentido de entreajuda que demonstraram durante este curso. Agradeço em particular ao Francisco Pedrosa, o software de detecção de cantos em imagens, utilizado nesta tese.

Ao meu amigo de há muitos anos, Francisco Moreira, agradeço o excelente companheirismo de que sempre deu provas.

Aos meus pais, que me proporcionaram as condições para aqui chegar, e à minha irmã, que me mostra quão simples podem ser as situações mais complicadas, devo o meu mais profundo sentimento de gratidão. Dedico-lhes, por isso, esta dissertação.



## RESUMO

Nesta tese descreve-se um sistema de detecção visual de obstáculos para robots móveis, desenvolvido segundo o paradigma de visão por objectivos. Uma única câmara e um conjunto de espelhos planos são usados para aquisição de *imagens stereo* da área à frente do robot móvel. A hipótese de deslocação do robot sobre um pavimento plano é usada explicitamente na caracterização do problema perceptual. Um obstáculo é definido como qualquer ponto não pertencente ao plano do pavimento.

Durante a fase de inicialização, um conjunto de pontos correspondentes é usado para a estimação robusta de um conjunto de 8 parâmetros que descrevem de uma forma completa o mapa de disparidades associado ao plano do pavimento. O mapa de disparidades, *Mapa de Referência*, contém a disparidade para cada ponto 3D, desde que pertencente ao plano do pavimento. Durante a operação normal, a detecção de obstáculos consiste na identificação de pontos para os quais o mapa de referência é violado. Foi realizada uma implementação em tempo real e foram obtidos bons resultados.

Descreve-se também um algoritmo alternativo de detecção de obstáculos sobre pavimento plano, baseado no cálculo do fluxo óptico de uma sequência de imagens adquirida por uma única câmara em movimento. O algoritmo é analisado no plano formal, discutindo-se a dualidade e a complementaridade face ao algoritmo stereo implementado.

**Palavras chave :** Robótica Móvel, Visão robótica, Reconstrução, Visão Qualitativa e por Objectivos, Detecção de Obstáculos sobre Pavimento Plano.

## **ABSTRACT**

This thesis presents a system for vision-based obstacle detection for mobile robots, developed within the framework of purposive vision. A single camera and a set of flat mirrors are used to acquire two distinct images of the area in front of a mobile robot. The assumption that the robot is moving on the ground plane is explicitly used to constrain the perception problem. An obstacle is defined as any point laying outside of the ground plane.

During an initialisation phase, a set of correspondent points are used to robustly estimate the set of 8 parameters that fully describe the disparity map of the ground plane. The disparity map, Reference Map, establishes the disparity for every image pixel, provided that the corresponding 3D point is on the ground plane. During operation, to detect the presence of obstacles, the system has to check for the points where the Reference Map is violated. A real time implementation was developed yielding good results.

An alternative algorithm for ground plane obstacle detection, based on the calculation of the flow field of an image sequence acquired by a moving camera, is described. This approach, dual to the stereo system proposed in the thesis, is formally discussed and its main features analysed.

**Keywords** : Mobile Robotics, Robot Vision, Reconstruction, Qualitative and Purposive Vision, Ground Plane Obstacle Detection (GPOD).

# ÍNDICE

	Pág.
AGRADECIMENTOS.....	i
RESUMO .....	iii
ABSTRACT .....	iv
ÍNDICE .....	v
CAPÍTULO 1 Introdução. Robótica Móvel e Visão. ....	1
1.1. Motivação e apresentação do problema.....	1
1.2. Visão reconstrutiva. Visão qualitativa e por objectivos. Visão activa. ....	5
1.3. Organização da tese .....	7
CAPÍTULO 2 Visão.....	9
2.1. Modelo da câmara.....	9
2.2. Stereo binocular .....	13
2.2.1. Cálculo de profundidade .....	15
2.2.2. Emparelhamento ("Matching") .....	15
2.3. Sequências de imagem. Fluxo óptico. ....	20
2.4. Resumo.....	22
CAPÍTULO 3 Detecção de Obstáculos baseada em Stereo Binocular .....	25
3.1. Introdução .....	25
3.2. Inicialização .....	30
3.2.1. Estimção de parâmetros do mapa de referência.....	33
3.2.2. Obtenção de pontos correspondentes .....	35
3.2.3. Métodos alternativos para a obtenção do mapa de referência.....	36
3.3. Detecção.....	37
3.3.1. Filtragem morfológica .....	38
3.3.2. Diagrama de blocos da fase de detecção.....	39
3.4. Evoluções possíveis .....	39
3.5. Resumo.....	40

CAPÍTULO 4 Aparato Experimental.....	43
4.1. O sistema de aquisição.....	44
4.2. O sistema de dois espelhos .....	45
4.2.1. Duas câmaras virtuais.....	45
4.2.2. Espaço visível.....	49
4.2.3. Resolução. Altura mínima detectada .....	50
4.3. A placa de aquisição .....	51
4.4. Resumo .....	53
CAPÍTULO 5 Imagens Stereo vs Sequências de Imagem.....	55
5.1. Detecção de Obstáculos sobre Pavimento Plano baseada em Fluxo Óptico... 55	
5.2. Complementaridade.....	63
5.3. Resumo .....	64
CAPÍTULO 6 Resultados Experimentais .....	67
6.1. Tempos de cálculo.....	75
6.2. Resumo .....	75
CAPÍTULO 7 Conclusões .....	77
7.1. Trabalho Futuro .....	79
Bibliografia.....	81
Apêndice A: Detecção de Pontos de Canto.....	87
Apêndice B: Correção da Distorção .....	89
Apêndice C: Ajuste de um Parabolóide a uma Superfície de SSDs .....	93
Cálculo do mínimo do parabolóide (mínimo da SSD).....	94
Apêndice D: Configuração do sistema de detecção de obstáculos.....	95



# **CAPÍTULO 1**

## **Introdução. Robótica Móvel e Visão.**

### **1.1. Motivação e apresentação do problema**

A robótica é uma disciplina científica multidisciplinar em torno da qual outras disciplinas têm tido desenvolvimentos específicos expressivos. O crescimento e diversidade das suas aplicações, bem como o interesse que cada vez mais grupos de investigação revelam, fazem antever uma cada vez maior importância desta área em domínios como os dos sistemas produtivos, dos serviços e do lazer.

As aplicações da robótica vão, hoje, desde as associadas à substituição do homem em tarefas repetitivas, não interessantes, susceptíveis de erro, degradantes ou sub-utilizadores das suas capacidades e sensibilidade, até às associadas à assistência a pessoas inválidas ou idosas, e vigilância ou operação em ambientes hostis (ex.: ambientes radioactivos, zonas profundas dos mares, espaço, etc).

Dispositivos robóticos têm sido desenvolvidos com várias configurações e objectivos. Os mais vulgares são sem dúvida os manipuladores robóticos, "braços robot", que começaram por ganhar popularidade nas linhas de montagem de automóveis. Estes manipuladores de base fixa são, no entanto, limitados no seu volume de trabalho. O alargamento do volume de trabalho tem vindo a ser realizado com "robots móveis". Os "robots móveis" começam a ser usados por exemplo em ambientes industriais para transporte de "peças" entre as estações de fabrico e de armazenagem.

Um robot móvel consiste numa plataforma móvel, sobre a qual é integrada, de forma inteligente, percepção e acção sobre o mundo que rodeia o robot. A "acção" guia e controla o robot durante o seu movimento pelo mundo. A "percepção" recolhe os dados

dos sensores, interpreta-os de modo a melhorar a compreensão do mundo e propõe "acções".

A utilização de imagens para percepção, por robots móveis, do mundo envolvente tem sido um desafio importante para as equipas de investigação que desenvolvem actividade na área da Visão Robótica.

A área da visão robótica encontra na observação do mundo animal uma das suas fontes inspiradoras mais importantes. De facto, o próprio ser humano usa o sistema visual, com muita eficiência, em muitas das suas actividades. As tentativas já realizadas, de simulação e implementação artificial de capacidades visuais, têm revelado estar-se perante um problema muito complexo e de resolução global difícil. Esta dificuldade tem levantado a questão de qual o grau em que as capacidades visuais humanas e dos restantes animais podem ser imitadas e utilizadas em robots móveis. A resposta a esta questão tem vindo a ser procurada, por alguns investigadores, em estudos de psico-fisiologia da visão.

A introdução de visão num robot móvel traduz-se num significativo aumento das suas capacidades sensoriais e portanto num correspondente aumento de versatilidade e segurança na operação do robot. Esperam-se por exemplo, da visão robótica, soluções ou simplificações para problemas de detecção e localização de objectos ou obstáculos, úteis no desempenho de tarefas de manipulação ou navegação.

Vários protótipos de robots móveis com navegação, "guiamento" ou planeamento suportados em informação visual têm sido desenvolvidos.

Um destes exemplos [Dickmanns92], tem por base um veículo automóvel de mercadorias / comercial equipado com todos os necessários actuadores e "interfaces" de modo a poder ser conduzido por computador. Duas câmaras de vídeo, suportadas por uma plataforma electromecânica de dois graus de liberdade ("pan", "tilt") montada próxima do vidro pára-brisas dianteiro, permitem a visualização das bermas da estrada (fig. 1.1) e de eventuais obstáculos (outros veículos na estrada). Este conjunto experimental foi construído para investigação e teste de sistemas de visão. Os resultados do seguimento visual da estrada e detecção / desvio de obstáculos, foram notáveis, tendo o veículo atingido de forma segura velocidades de 50 km/h.

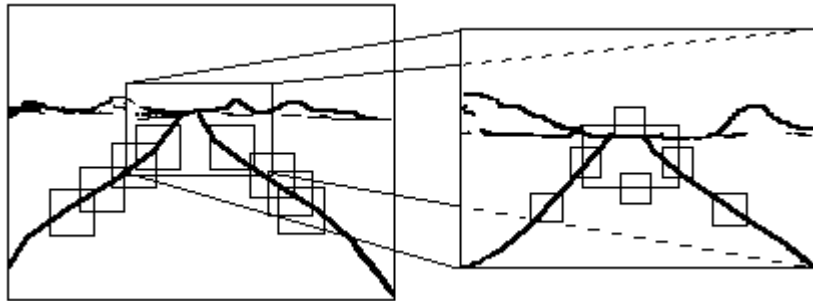


Fig. 1.1 Seguimento das bermas da estrada. Imagem esquerda obtida com uma lente de grande angular; imagem direita obtida com uma lente de pequena angular (zoom) para uma visualização pormenorizada. [Dickmanns92].

Num outro exemplo, baseado numa plataforma móvel e num sistema de três câmaras (montadas sobre a plataforma), o sistema detecta segmentos de recta 3D e usa as suas projecções (fig. 1.2) sobre o plano do pavimento para planeamento de trajectórias sem colisões (evita obstáculos). [Buffa92].

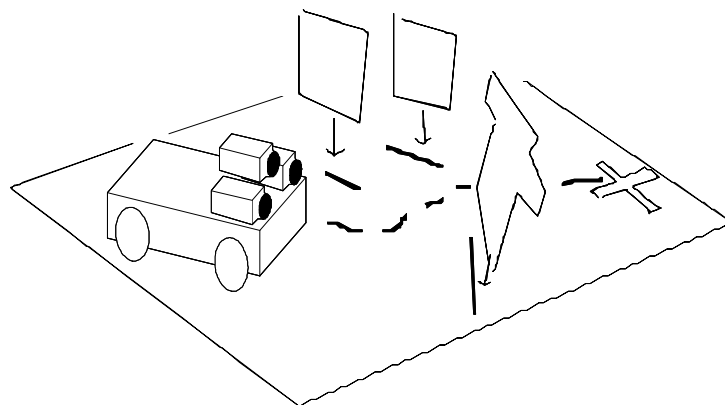


Fig. 1.2 Informação de projecção dos obstáculos sobre o chão. [Buffa90], [Buffa92].

Num terceiro exemplo de navegação visual [Blake92], são detectados e seguidos contornos (linhas) na imagem. O robot (suporte da câmara) realiza movimentos exploratórios, provocando deslocações dos contornos na imagem (ver fig.1.3.a), cuja análise permite a selecção de linhas de perfil dos obstáculos e a obtenção de informação

qualitativa sobre a sua posição (obstáculo do lado esquerdo ou do lado direito da linha de perfil). Esta informação permite ao robot mover-se para um dado ponto do seu mundo, contornando os obstáculos (fig.1.3.b).

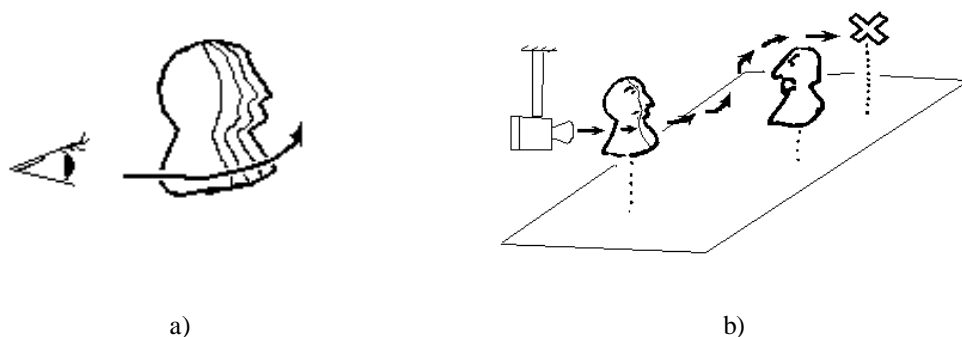


Fig. 1.3 A linha limite (perfil) de um volume, varia de acordo com a direcção de observação (imagem 'a'). Esta variação pode ser explorada para o desenho de algoritmos de contorno de obstáculos (imagem 'b'). [Blake92].

Nesta tese trata-se o problema da detecção visual, por um robot móvel, de obstáculos sobre o pavimento plano onde aquele se desloca. A hipótese de movimento sobre pavimento plano ([Ferrari90], [Mallot91], [Zielke90]) é válida para um número elevado de aplicações e permite uma definição simples de obstáculo, como um conjunto de pontos situados fora do plano do pavimento.

A detecção de obstáculos é uma tarefa fundamental quando se pretende que um robot móvel se desloque sem colisões num ambiente onde existam à partida obstáculos fixos ou móveis ou onde obstáculos móveis possam surgir durante o curso de uma missão. Em particular o sistema descrito nesta tese destina-se a sistemas reactivos de desvio de obstáculos ("obstacle avoidance"). A ideia de comportamento robótico reactivo, mais uma vez, surge por analogia com o comportamento humano: quando um peão distraído de repente se apercebe que está na eminência de chocar com outro peão, esquece momentaneamente o seu lugar de destino e simplesmente escolhe virar um pouco à direita ou à esquerda, ou simplesmente parar.

## 1.2. Visão reconstrutiva. Visão qualitativa e por objectivos. Visão activa.

A visão robótica, é desde há trinta anos uma disciplina perfeitamente identificada. Durante este período, a maior parte da investigação e desenvolvimento tem estado concentrada na inversão do processo de formação de imagem para reconstrução tridimensional do mundo (**visão reconstrutiva**) ou na identificação de objectos, suas localizações e orientações [Blake92, Aloimonos91].

Desde o trabalho pioneiro de David Marr (nos anos setenta), sistematizando e propondo uma teoria para a organização da percepção visual, a visão reconstrutiva tem sofrido grandes avanços e conhecido importantes desenvolvimentos.

Como exemplos de módulos visuais de reconstrução podem ser referidos [Aloimonos90]: recuperação de forma a partir de sombreamento, de textura, de contornos ou a partir de stereo; recuperação de profundidade a partir de stereo; recuperação de estrutura a partir do movimento; recuperação da direcção da luz a partir da intensidade do brilho; recuperação de descontinuidades físicas a partir de descontinuidades no brilho; recuperação de movimento a partir de derivadas da intensidade do brilho; etc.

Todos estes exemplos são problemas inversos. Uma grande parte dos problemas inversos são mal postos, no sentido de que não têm solução, não têm solução única ou não dependem de forma continua dos dados. Para estes casos, são usuais as técnicas de regularização, baseadas em funcionais de custo onde se incluem hipóteses adicionais, normalmente condições de suavidade [Tikhonov77, Horn86, Terzopoulos84, Poggio85].

Uma característica comum dos sistemas visuais de reconstrução, é a necessidade de calibração métrica, i.e., cálculo de parâmetros da(s) câmara(s) (distância focal, centro óptico, etc) e das medidas dos suportes (especialmente no caso stereo é necessária a estimação da relação entre câmaras). Normalmente o problema da calibração é complexo porque os sistemas envolvidos são não lineares, motivando processos morosos e pouco robustos.

Para obviar este facto, têm vindo a ser desenvolvidos sistemas visuais projectados segundo uma abordagem qualitativa. No paradigma da **visão qualitativa**, o termo

qualitativo é geralmente usado com o significado de *não utilização de calibração métrica* ("nonmetric", [Lim91]); em alguns casos é simplesmente dada ênfase à diminuição da confiança sobre a precisão métrica característica da visão reconstrutiva.

Apesar de não ser usada calibração métrica continua a ser possível extrair informação útil das imagens. Existem técnicas de reconhecimento com invariantes [Zisserman92], técnicas de detecção, técnicas de seguimento, etc.

Os algoritmos enquadrados no paradigma da visão qualitativa podem ser mais robustos porque não se baseiam em discriminações métricas finas, dependentes do processo de calibração, em restrições na geometria ou na estrutura da imagem, etc.

Por outro lado existe uma motivação biológica para realizar investigação seguindo o paradigma da visão qualitativa. O afastamento ou diminuição da "métrica" está de acordo com a nossa sensibilidade: se é normalmente fácil distinguir entre dois objectos qual o mais próximo, já não é tão fácil indicar por uso exclusivo de visão, as distâncias a que se encontram os dois objectos.

As técnicas de reconstrução, normalmente pouco robustas e muitas vezes pesadas computacionalmente, têm motivado a procura pela comunidade científica, de perspectivas mais pragmáticas para a realização de sistemas de visão. Um paradigma recente, **Visão por Objectivos** ("Purposive Vision") [Aloimonos90], caracteriza-se pelo desenho de módulos visuais que desempenham tarefas simples e específicas, de resultados imediatamente úteis para a realização de acções. A integração destes módulos permite atingir um comportamento geral inteligente. A especificação ou identificação das tarefas simples, recorre normalmente a analogias biológicas.

O ênfase deste paradigma é sobre a identificação da tarefa, e não sobre a máxima extracção de informação da imagem para pós-processamento.

Continuando uma evolução pragmática, um paradigma novo tem-se vindo a impor: *visão activa* ([Bajcsy85, Aloimonos90, Ballard91, Blake92]).

A **visão activa** tornou-se importante, em parte, devido a dificuldades computacionais das técnicas de visão tradicionais e também devido a novas oportunidades

tecnológicas (especialmente do domínio da robótica, por exemplo com o aparecimento das primeiras cabeças robot).

Visão activa não significa desenvolvimento de sistemas visuais que contêm a sua própria fonte de radiação como o nome poderia indicar por oposição a sistemas visuais passivos.

Os sistemas de Visão activa supõem normalmente mecanismos que permitem um controlo activo de parâmetros da câmara tais como posição, orientação, focagem, zoom, abertura e vergência (num sistema de duas câmaras) em resposta a solicitações das tarefas ou dos estímulos externos.

As abordagens de visão qualitativa e a visão por objectivos têm vindo a ser incluídas em visão activa porque:

- não é fácil uma calibração métrica contínua, por exemplo, de uma "cabeça stereo robot" (o "movimento de olhos" é uma característica fundamental da Visão activa).
- fornecem de facto informação útil para esquemas reactivos ou de controlo.
- identificam e fornecem um conjunto de "tasks" (tarefas) às cabeças robóticas, permitindo-lhes exibir um comportamento inteligente ([Aloimonos90]).

O sistema de detecção visual de obstáculos proposto nesta tese, enquadra-se num paradigma de visão qualitativa e por objectivos. É qualitativo porque não usa calibração métrica (fornecendo por consequência resultados de carácter qualitativo). É "objectivo" porque se destina a um fim específico (detecção de obstáculos para um sistema de "obstacle avoidance" reactivo para plataformas móveis), tendo sido especialmente desenvolvido nesse sentido.

### **1.3. Organização da tese**

Ao longo da tese é descrito um sistema de detecção visual de obstáculos. Este sistema é baseado em pares de imagens stereo para inicialização e operação, e gera como resultado imagens com indicação de pontos obstáculo.

Como principais tópicos da tese, podem ser referidos:

- Filosofia da abordagem diferente da tradicional reconstrução: o sistema de detecção de obstáculos proposto segue uma abordagem "qualitativa e por objectivos".

## *Capítulo 1: Introdução*

- Formalização e utilização da descrição global de um mapa de disparidades (correspondências) de referência. Utilização de "plane to plane projective transformation" (há muito conhecida na geometria mas só recentemente introduzida na visão) para descrição global do mapa de referência. Estimação de parâmetros.

- Estudo de um sistema de dois espelhos para a realização de um sistema stereo de duas câmaras virtuais.

O texto da tese está organizado em sete capítulos, com o seguinte conteúdo:

Capítulo 2: introdução de conceitos e definições úteis para os restantes capítulos (modelo da câmara, ponto correspondente, disparidade, emparelhamento, linha epipolar, fluxo óptico, foco de expansão, etc)

Capítulo 3: descrição do algoritmo principal do sistema: Algoritmo de Detecção de Obstáculos sobre Pavimento Plano. Apresenta a forma do mapa de disparidades (correspondências) de referência e a estimação dos seus parâmetros. Descreve a fase de detecção de obstáculos.

Capítulo 4: descrição do aparato experimental, desde a estrutura de dois espelhos e uma câmara para realizar um sistema stereo de duas câmaras virtuais, até à placa de aquisição. São em particular discutidos os tópicos de espaço visível e de resolução do sistema (altura mínima detectada).

Capítulo 5: resultados experimentais.

Capítulo 6: descrição formal de um algoritmo de detecção de obstáculos baseado em sequências de imagens e no movimento da câmara, dual do algoritmo stereo implementado. É discutida a dualidade e a complementaridade dos dois algoritmos.

Finalmente, no capítulo 7, são tecidos comentários finais sobre o sistema descrito e implementado, e são propostos alguns objectivos para trabalho futuro.



# CAPÍTULO 2

## Visão

Pretende-se neste capítulo introduzir alguns conceitos e definições úteis para o resto do texto desta tese.

### 2.1. Modelo da câmara

Do ponto de vista do desenvolvimento de sistemas de visão, o problema de aquisição de imagem está simplificado dada a variedade de câmaras de vídeo e placas de aquisição de imagem existentes no mercado (fig. 2.1). Estas placas digitalizam sinais de vídeo produzidos pelas câmaras (tipicamente baseadas em sensores CCD<sup>1</sup>) e memorizam as imagens obtidas.

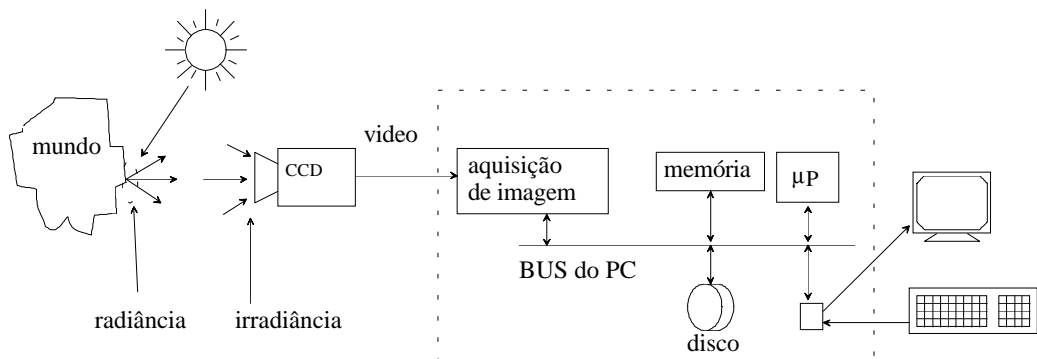


Fig. 2.1 Formação de imagem (nível de brilho) e aquisição de imagem.

Para o desenvolvimento de algoritmos de visão, é importante perceber o processo de formação de imagem, ou seja a ligação entre o mundo real e a imagem  $I(x,y)$  capturada

---

<sup>1</sup>CCD: "Charge Coupled Device"

pela câmara. Duas questões importantes estão associadas a este processo ([Horn86]). A primeira questão é geométrica e preocupa-se com a obtenção da posição na imagem de um dado ponto tridimensional. A segunda questão está relacionada com a determinação na imagem das cores ou níveis de brilho dos objectos do mundo: é um problema de radiometria.

Para a questão geométrica, a resposta mais usual é um modelo de **câmara pontual** ("pin-hole camera model"), que representa uma projecção perspectiva (ilustrado na fig. 2.2). Este modelo supõe que a câmara é constituída por uma caixa negra com uma pequena abertura (um ponto) numa face e um sensor de imagem na face interna oposta.

A projecção de um ponto  $P$  no plano de imagem é determinada pela intersecção do raio passando por  $P$  e pelo centro de projecção  $O$ , com o plano de imagem.

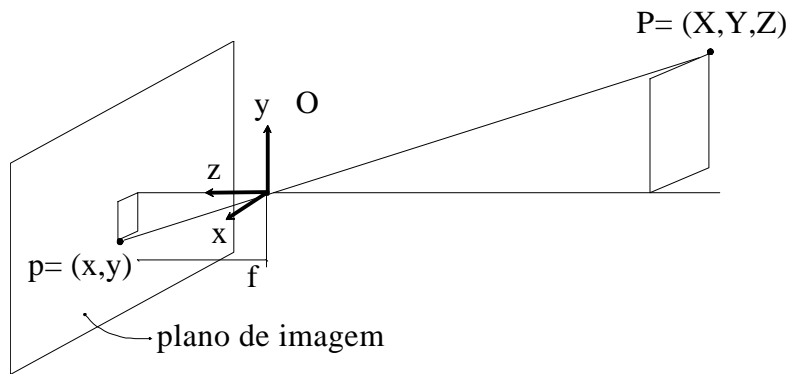


Fig. 2.2 Modelo de câmara pontual.

Este modelo, não é fisicamente realizável porque uma pequena abertura significa uma pequena passagem de energia luminosa (insuficiente para ser detectável). É contudo uma boa aproximação para o caso real de uma câmara equipada com uma lente, que permite simultaneamente a projecção de um ponto 3D como um ponto 2D e além disso concentra a energia de um cone de direcções, suficiente para a detecção.

Nesta tese, será usado um modelo semelhante. Em particular o plano de imagem é colocado à frente do sensor e o sistema de coordenadas é reorientado (fig. 2.3). Este modelo afasta-se da realidade física, mas adapta-se melhor aos actuais dispositivos de aquisição e "display" de imagem.

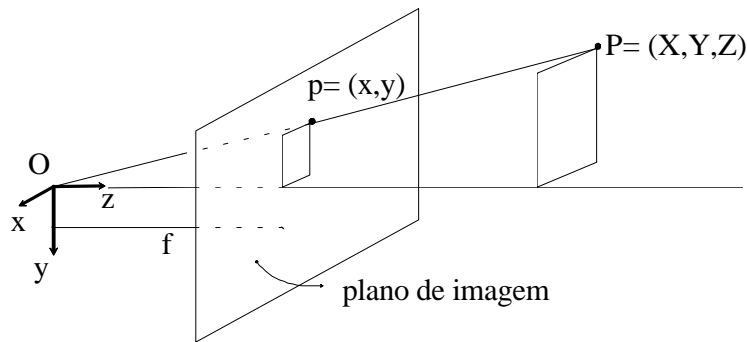


Fig. 2.3 Modelo de câmara pontual (plano de imagem à frente do centro de projecção).

O modelo de câmara pontual é caracterizado por uma equação que relaciona coordenadas tridimensionais de um ponto com as coordenadas bidimensionais da sua projecção:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \frac{f}{Z} \begin{bmatrix} X \\ Y \end{bmatrix} \quad (1)$$

onde  $(X, Y, Z)$  denota um ponto do espaço tridimensional,  $(x, y)$  é a sua projecção sobre o plano de imagem e  $f$  representa a distância entre o plano de imagem e o centro de projecção.

Muitas vezes  $f$  é referida como *distância focal*, aludindo à concretização do modelo pontual por uma câmara equipada com uma lente. A distância entre plano de imagem e centro de projecção é semelhante à distância focal quando a distância entre objecto observado e lente é muito maior que a distância focal.

A determinação do nível de brilho de um dado pixel<sup>2</sup> da imagem, motivado por um ou mais pontos do cenário, depende da iluminação e das características do material observado.

---

<sup>2</sup>pixel: "picture element"

Definindo "irradiância" como a potência luminosa recebida por unidade de área e "radiância" como a potência luminosa emitida por unidade de área num cone de direcções com ângulo sólido unitário, o brilho de um ponto no plano de imagem corresponde à irradiância recebida nesse ponto e depende da radiância do ponto do mundo (fig. 2.1) e naturalmente da abertura e focagem da lente.

A radiância de uma superfície do mundo depende da iluminação (irradiância recebida pela superfície) e do tipo de superfície. Por exemplo uma superfície Lambertiana radia, de igual forma em todas as direcções, toda a energia recebida (a energia recebida é função da direcção da superfície comparada com a direcção de iluminação).

A informação que se pode extrair de uma imagem, está normalmente corrompida com ruído e rodeada de bastante redundância. No sentido de reduzir o ruído e (ou) a redundância, são usuais as operações de segmentação.

Procura-se normalmente com a segmentação encontrar regiões de brilho, cor ou textura uniformes, ou encontrar outras características relevantes. Recorrendo a estudos sobre a visão humana, observa-se que a capacidade de reconhecimento se mantém bastante elevada mesmo quando as imagens são reduzidas aos seus contornos. Os contornos são portanto características cuja extracção é útil para sistemas visuais.

Formalmente, um contorno é uma linha formada por uma sequência de pontos de contorno, sendo um **ponto de contorno** ("edge") definido como ponto da imagem onde existe variação rápida do nível de brilho. Definido o conceito, vários têm sido os detectores propostos (laplaciano, laplaciano da gaussiana (Marr-Hildreth), detector de Canny [Canny83], etc), normalmente baseados em operadores diferenciais e com maior ou menor filtragem de ruído.

Outra característica importante de uma imagem são os seus "cantos". Um **ponto de canto** é um ponto da imagem onde a linha (curva) de contorno apresenta variação brusca na direcção. Para referências em detectores de cantos ver por exemplo [Eryurtlu92].

Nesta tese foi usado um detector de cantos, o detector de Moravec ([Nasrabadi92]). Este detector selecciona pontos da imagem para os quais todas as direcções apresentam uma variação elevada do nível de brilho (definição alternativa de ponto de canto).

## 2.2. Stereo binocular

Um sistema stereo binocular é um sistema baseado em pares de imagens adquiridas em simultâneo (veja-se fig.2.4) e é o mais frequente nos sistemas de visão biológicos e robóticos, sendo muitas vezes referido simplesmente como stereo. O sistema de visão do homem é um exemplo de visão stereo binocular. Um exemplo de stereo não binocular pode ser visto em [Buffa92], onde conjuntos de três imagens ("trinocular stereo") são usados para retirar ambiguidade a algumas correspondências e aumentar a qualidade de outras.

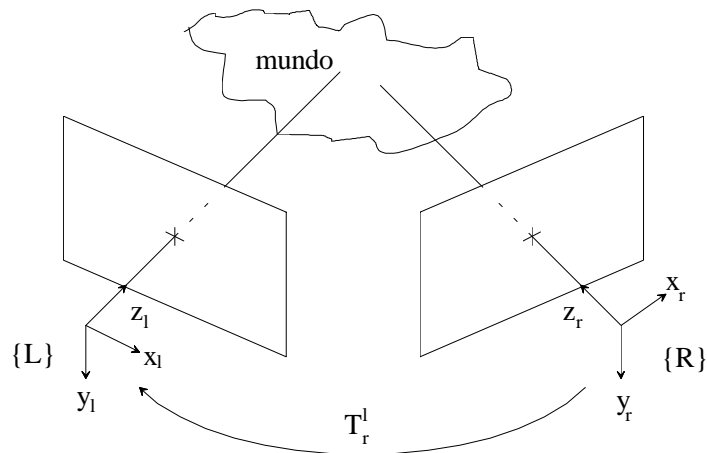


Fig. 2.4 Sistema stereo binocular: 2 câmaras.

Nos sistemas artificiais mais vulgares as câmaras têm uma posição rígida uma em relação à outra. Esta relação de transformação de coordenadas é representável por uma matriz homogénea  $T$  ( $4 \times 4$ ) ou pelo conjunto de uma matriz de rotação  $R$  ( $3 \times 3$ ) e de um vector de translação  $P$  ( $3 \times 1$ ), constantes.

$$[X_l \ Y_l \ Z_l \ 1]^T = T_r^l [X_r \ Y_r \ Z_r \ 1]^T \quad (2)$$

$$T_r^l = \begin{bmatrix} r_{11} & r_{12} & r_{13} & p_x \\ r_{21} & r_{22} & r_{23} & p_y \\ r_{31} & r_{32} & r_{33} & p_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3)$$

Nos sistemas artificiais mais recentes ("cabeças robóticas de visão stereo" [Santos-Victor94+]) as câmaras podem-se mover pelo que  $T$  pode não ser constante (este caso não será desenvolvido durante esta tese).

Entre duas imagens obtidas por um sistema stereo convergente (ambas as câmaras direccionadas para uma mesma zona do cenário) existem naturalmente semelhanças. O conceito de **pontos correspondentes** (ou **homólogos**) e o conceito de **disparidade** formalizam esta ligação de informação entre imagens.

Dois pontos  $(x_r, y_r)$  e  $(x_l, y_l)$  dizem-se correspondentes se forem projecções do mesmo ponto tridimensional  $(X_r, Y_r, Z_r)$  (ou  $(X_l, Y_l, Z_l)$ ) usando o sistema de coordenadas da câmara esquerda) pertencente a uma dada superfície do mundo.

Disparidade num ponto  $(x_r, y_r)$  é um vector definido pela diferença de coordenadas para o ponto correspondente  $(x_l, y_l)$ :

$$disp_r(x_r, y_r) = \begin{bmatrix} x_l \\ y_l \end{bmatrix} - \begin{bmatrix} x_r \\ y_r \end{bmatrix} = \begin{bmatrix} x_l - x_r \\ y_l - y_r \end{bmatrix} \quad (4)$$

A partir do conceito de disparidade pode ser definido o "mapa de disparidades" como um conjunto de vectores de disparidade (um por cada  $(x_r, y_r)$  da imagem).

Disparidade também pode ser definida tomando por referência a imagem esquerda:

$$disp_l(x_l, y_l) = \begin{bmatrix} x_r \\ y_r \end{bmatrix} - \begin{bmatrix} x_l \\ y_l \end{bmatrix} = \begin{bmatrix} x_r - x_l \\ y_r - y_l \end{bmatrix}$$

Nesta tese será sempre usada a primeira forma, i.e., tomada por referência a imagem direita.

É possível obter um mapa de disparidades referenciado sobre a imagem esquerda a partir de um mapa de disparidades referenciado sobre a imagem direita, excepto nas regiões de **oclusão** (regiões do mundo visíveis numa imagem e encobertas na outra) para as quais não existirão normalmente estimativas de disparidades. A relação entre mapas (excluindo oclusões) é simples:  $disp_l(x_l, y_l) = -disp_r(x_r, y_r)$  (sendo  $(x_l, y_l)$  e  $(x_r, y_r)$  pontos homólogos).

### 2.2.1. Cálculo de profundidade

Uma só imagem não permite geralmente fazer afirmações absolutas<sup>3</sup> sobre dimensões ou **profundidades** (distâncias entre plano de imagem e pontos do mundo) das diversas zonas do cenário observado. Vêr-se-á à frente que o cálculo de profundidades é possível a partir de imagens stereo. Esta é uma razão importante para a utilização de sistemas stereo.

Considerando ambas as câmaras de um sistema stereo binocular, pontuais e admitindo relação rígida entre câmaras obtém-se o seguinte sistema ([Horn86]):

$$\begin{cases} \frac{x_l}{f} Z_l = \left( r_{11} \frac{x_r}{f} + r_{12} \frac{y_r}{f} + r_{13} \right) Z_r + p_x \\ \frac{y_l}{f} Z_l = \left( r_{21} \frac{x_r}{f} + r_{22} \frac{y_r}{f} + r_{23} \right) Z_r + p_y \\ Z_l = \left( r_{31} \frac{x_r}{f} + r_{32} \frac{y_r}{f} + r_{33} \right) Z_r + p_z \end{cases} \quad (5)$$

que pode ser resolvido em ordem a  $Z_r$  (profundidade no ponto  $(x_r, y_r)$ ), combinando a primeira e terceira equações (ou segunda e terceira equações).

O cálculo de  $Z_r$  é possível desde que sejam conhecidos a relação entre câmaras  $T$ , a distância focal  $f$  e o ponto correspondente de  $(x_r, y_r)$ :  $(x_l, y_l)$ .

### 2.2.2. Emparelhamento ("Matching")

É portanto importante para o cálculo de profundidade, encontrar o ponto da imagem esquerda  $(x_l, y_l)$  correspondente de  $(x_r, y_r)$  da imagem direita. O processo de procura de pontos correspondentes é designado de **emparelhamento** ("**matching**").

Um critério usual de comparação de níveis de brilho é a *soma das diferenças ao quadrado dos níveis de brilho* ("Sum of the Squared Differences": **SSD**):

---

<sup>3</sup>A recuperação de valores relativos de profundidade (formas de superfícies) é possível a partir de imagens únicas, usando por exemplo técnicas de "shape from shading" e conhecendo as propriedades de reflectância das superfícies observadas.

$$I_{SSD_{test}}(x_r, y_r) = \sum_{i=-W}^W \sum_{j=-W}^W [ I_l(x_l + i, y_l + j) - I_r(x_r + i, y_r + j) ]^2 \quad (6)$$

Este critério é uma medida simplificada de correlação, tem um valor mínimo quando os níveis de brilho dos pontos comparados (e respectivos vizinhos) são idênticos e é útil para o processo de emparelhamento<sup>4</sup>.

Como atrás foi definido, dois pontos são correspondentes se se referem a um mesmo ponto tridimensional do mundo. Espera-se a partir desta definição que os níveis de brilho dos dois pontos (e respectivos vizinhos) sejam idênticos. Esta observação sugere para o processo de emparelhamento, a procura do ponto  $(x_l, y_l)$  na imagem esquerda que minimiza o valor de  $I_{SSD}(x_r, y_r)$  (fig. 2.5).

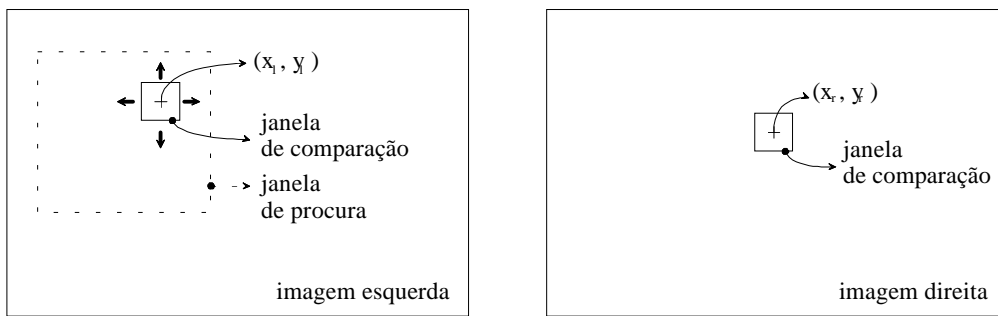


Fig. 2.5 Processo de emparelhamento. A janela de comparação move-se dentro da janela de procura até encontrar o  $(x_l, y_l)$  que melhor corresponde a  $(x_r, y_r)$  (SSD mínimo: brilhos semelhantes).

A partir da equação (5) pode ser deduzida uma restrição de procura: a **restrição epipolar**. Sendo conhecidas a relação entre câmaras e as respectivas distâncias focais, a restrição epipolar indica para cada ponto  $(x_r, y_r)$  uma linha de possíveis pontos correspondentes  $(x_l, y_l)$ : **linha epipolar** (ver exemplo na fig. 2.6).

<sup>4</sup>O facto de serem considerados pontos vizinhos na comparação (SSD) é importante para atenuar o efeito do ruído sempre presente nas imagens.



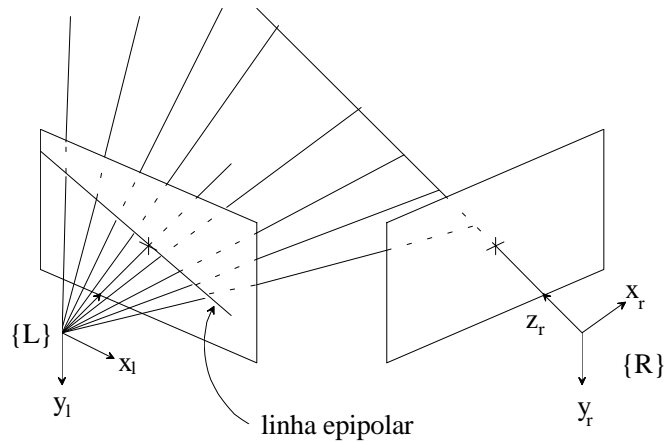


Fig. 2.6 Exemplo de linha epipolar.

Geometricamente a linha epipolar definida para o ponto  $(x_r, y_r)$  corresponde à intersecção do plano epipolar (definido pelos centros ópticos das duas câmaras e pelo ponto  $(x_r, y_r, f)$ ), com o plano de imagem da câmara esquerda.

O conhecimento do sistema stereo (relação entre câmaras e distâncias focais), permite então a obtenção das linhas epipolares, reduzindo desta forma o espaço de procura de pontos correspondentes e otimizando portanto o processo de emparelhamento (fig.2.7).

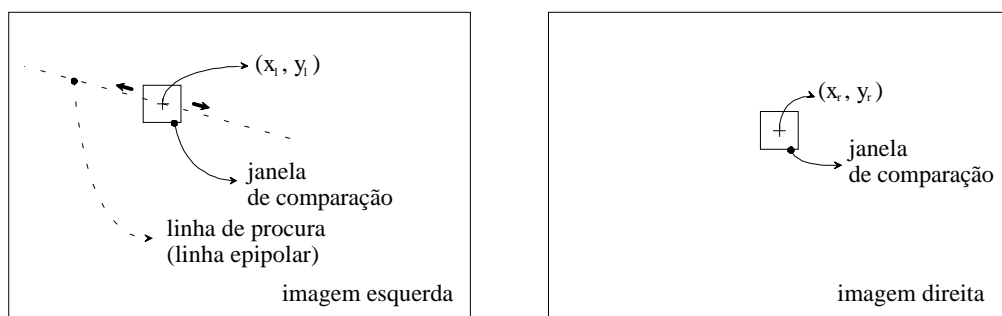


Fig. 2.7 Processo de emparelhamento. A janela de comparação move-se sobre a linha de procura até encontrar o  $(x_l, y_l)$  que melhor corresponde a  $(x_r, y_r)$  (SSD mínimo: brilhos semelhantes). O espaço de procura é significativamente menor quando comparado com o caso da fig.2.5: o processo de emparelhamento é mais eficiente.

### Dificuldades no processo de emparelhamento

O processo de emparelhamento de pontos é difícil ou impossível em regiões de imagem com informação insuficiente, tais como regiões com nível de brilho constante ou sobre contornos com direcções das linhas epipolares (fig.2.8). A razão desta dificuldade assenta na existência de vários mínimos locais (ou mínimos absolutos consecutivos) de SSD.

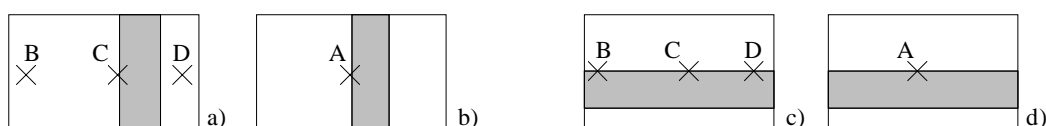


Fig. 2.8 Dificuldade de emparelhamento de pontos sobre contornos com direcção da restrição epipolar. Imagens a,b constituem um par stereo e c,d constituem outro par stereo. Por hipótese as linhas epipolares são horizontais. No caso a,b é fácil decidir que o ponto correspondente de A é o C. No caso c,d não existe informação suficiente para decidir qual o ponto correspondente de A (pode ser B, C, D ou qualquer outro na mesma linha horizontal).

Também as oclusões dificultam o processo de emparelhamento, podendo não existir (não ser visível) para um ponto na imagem de referência o ponto correspondente na outra imagem.

O ruído, estando sempre presente em qualquer imagem real, implica erros de localização (emparelhamento) quando as janelas de comparação são pequenas. Por outro lado janelas grandes de comparação podem conter variações de disparidades motivadas pelo efeito da perspectiva ([Okutomi92]). Também este facto pode induzir erro de emparelhamento (desvio em relação ao emparelhamento correcto).

Técnicas multi-nível são uma solução tradicional para o problema da escolha da dimensão das janelas de comparação: o emparelhamento é realizado primeiro em baixa resolução atenuando os erros do ruído e depois refinado em alta resolução, evitando os efeitos da perspectiva. Técnicas recentes apontam para soluções de compromisso: janelas de dimensão adaptativa, baseadas na informação local sobre nível de ruído ([Okutomi92]).

Um **problema mal posto** no sentido de Hadamard é um problema que não tem solução, a solução não é única, ou a solução não depende de forma continua dos dados.

Vários mínimos ou mínimos locais (de SSD), erros de localização devido a ruído em janelas pequenas, possibilidade de não existência de correspondência (oclusões), erros de localização devidos a diferenças nos padrões de disparidade entre janelas grandes em imagens diferentes, são dificuldades intrínsecas que levam à classificação do emparelhamento como um problema mal posto.

O facto de o emparelhamento ser um problema mal posto leva à utilização de técnicas de regularização, normalmente baseadas na hipótese de suavidade dos mapas de disparidade (muitas vezes complementada com detecção local de discontinuidades). Contudo, estas técnicas exigem usualmente um esforço computacional excessivo.

O emparelhamento de características tais como contornos -segmentos ou splines-, cantos, regiões, grafos, etc, resolve alguns dos problemas apontados do emparelhamento de pontos, mas introduz a dificuldade extra de detecção das características nas imagens. A detecção de características da imagem implica normalmente um processo de segmentação, que também é usualmente um problema mal posto.

O emparelhamento de segmentos de recta é útil em cenários artificiais (interiores). É um processo mais eficiente que a procura de correspondência para todos os pontos da imagem, mas exige uma atenção especial aos segmentos interrompidos.

Os pontos de canto nas imagens, são bons candidatos para o processo de emparelhamento porque implicam uma convexidade local da superfície de testes de SSD ([Anandan89]): a SSD apresenta um mínimo único (dentro de uma janela local).

A escolha de cantos não evita contudo a possibilidade de existência de vários mínimos (p.e. com repetição de padrões) nem de mínimos locais (padrões semelhantes).

## Reconstrução de mapa de profundidades

Aplicando a técnica de *cálculo de profundidade para um ponto* ( $x_p, y_p$ ) a todos os pontos ( $x_r, y_r$ ) da imagem, obtém-se (reconstrói-se) um **mapa de profundidades**.

A reconstrução de um mapa de profundidades<sup>5</sup> depende portanto do emparelhamento, problema mal posto (tal como foi visto na descrição de cálculo de profundidade), pelo que também é um problema mal posto.

---

<sup>5</sup>Muitas vezes referida simplesmente como "reconstrução tridimensional".

O carácter mal posto do problema tem como consequência a inexistência de informação de profundidade para certas regiões da imagem. Este facto leva normalmente à utilização de interpolações por meio de técnicas de regularização ([Terzopoulos84], [Gaspar92]) ou técnicas de triangulação (ex. triangulação de Delaunay).

### 2.3. Sequências de imagem. Fluxo óptico.

O **fluxo óptico** define-se como a velocidade de deslocação dos padrões de brilho na imagem, tendo por objectivo aproximar o **campo de movimento**, i.e., a projecção sobre o plano de imagem de velocidades 3D de pontos das superfícies do mundo.

Existem várias técnicas para o cálculo de fluxo óptico [Barron92]. As mais conhecidas são as técnicas diferenciais (1ª e 2ª ordem) e o seguimento de regiões ("Region-based matching").

A técnica diferencial de 1ª ordem [Horn86], a única que aqui será descrita, tem por base a hipótese de conservação da intensidade (nível de cinzento) do pixel, após eventual movimento  $(\Delta x, \Delta y)$  originado num pequeno intervalo de tempo  $\Delta t$  (para a existência de fluxo óptico não nulo é necessário movimento no cenário ou da câmara):

$$I(x + \Delta x, y + \Delta y, t + \Delta t) = I(x, y, t) \quad (7)$$

Expandindo em série de Taylor o membro do lado esquerdo da equação anterior, e desprezando termos de ordem superior à 1ª (os índices associados a  $I(x, y, t)$  significam derivadas parciais):

$$I(x + \Delta x, y + \Delta y, t + \Delta t) = I(x, y, t) + \Delta x \cdot I_x(x, y, t) + \Delta y \cdot I_y(x, y, t) + \Delta t \cdot I_t(x, y, t) \quad (8)$$

Combinando as equações (7) e (8) obtém-se finalmente a equação da **restrição fundamental do fluxo óptico**:

$$\boxed{I_x u + I_y v + I_t = 0} \quad (9)$$

onde foram omitidas as variáveis das funções, e  $(u, v)$  é o vector de fluxo óptico:

$$u = \lim_{\Delta t \rightarrow 0} \frac{\Delta x}{\Delta t}, \quad v = \lim_{\Delta t \rightarrow 0} \frac{\Delta y}{\Delta t}$$

Sendo uma só equação, e duas as incógnitas, não é possível a determinação completa do vector de fluxo óptico, pelo que normalmente se rearranja a equação (9) pondo em evidência o produto interno entre gradiente da imagem e fluxo óptico:

$$\begin{bmatrix} I_x & I_y \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = -I_t \quad (10)$$

e atribuindo-se então a (u,v) o valor da projecção do fluxo óptico sobre a normal ao contorno:

$$\begin{bmatrix} u \\ v \end{bmatrix} = -\frac{I_t}{\|(I_x, I_y)\|^2} \begin{bmatrix} I_x \\ I_y \end{bmatrix} \quad (11)$$

Esta dificuldade na determinação 2D do vector de fluxo óptico (nada se sabe na direcção do contorno) costuma ser referida como o **problema da abertura** ("aperture problem").

A determinação do fluxo óptico é portanto um problema mal posto (admite infinitas soluções). Uma técnica usual para estes casos é a regularização. É desenhado um funcional de custo onde se incluem os dados fornecidos por (9) e uma hipótese de suavidade do fluxo óptico ao longo da imagem. O resultado final surge após discretização e arranjo do funcional de custo, sob a forma de um algoritmo iterativo (relaxação).

Quando uma câmara se aproxima por exemplo de um poster, observa-se que os padrões de brilho do poster se expandem, a partir de um ponto: um foco ([Ballard82], [Sandini90]).

Geometricamente um foco é determinado pela intersecção com o plano de imagem da linha recta imaginária, que passa pelo centro óptico da câmara e tem direcção coincidente com translação. Em particular temos um **foco de expansão (FOE)**<sup>6</sup> se nos aproximarmos do ponto de fixação e os vectores de velocidade radiam a partir do foco; temos um **foco de contracção (FOC)** se a câmara se afasta e os vectores de velocidade colapsam / caem sobre o foco.

Formalmente, sendo o movimento da câmara uma translação pura com velocidade  $W=(W_x, W_y, W_z)$ , o FOE (FOC) pode ser calculado por:

$$FOE = (FOE_x, FOE_y) = (W_x, W_y) \cdot \frac{f}{W_z} \quad (12)$$

---

<sup>6</sup>FOE: "Focus of Expantion". FOC: "Focus of Contraction".

Notar que não existe determinação quando  $Wz=0$  (foco no infinito): movimento paralelo ao plano de imagem.

A partir do FOE radiam todos os vectores de fluxo óptico (admitindo cenário estático e movimento de translação da câmara não nulo segundo perpendicular ao plano da imagem). Adicionando esta restrição à equação da restrição fundamental do fluxo óptico é possível a determinação completa (2D) dos vectores de fluxo óptico, excepto nas regiões de brilho constante ou com variações só nas direcções perpendiculares às radiais do FOE <sup>7</sup>. Este problema é dual do problema de emparelhamento de pontos sobre contornos com direcção da restrição epipolar (fig.2.8).

Entre fluxo óptico e disparidade existe uma dualidade: dado o movimento do veículo que transporta a câmara, o fluxo óptico pode ser transformado em disparidade (usando o integral do fluxo óptico).

Com base na dualidade entre fluxo óptico e disparidade podem ser replicados alguns algoritmos: o algoritmo de detecção de obstáculos apresentado no capítulo 3 baseado em disparidades / correspondências é apresentado no capítulo 6 baseado em fluxo óptico.

Convém no entanto realçar a complementaridade da informação temporal *versus* espacial. Há informação que pode ser obtida por um sistema stereo (por exemplo distância a um objecto móvel) e não pode ser obtida por um sistema baseado em sequências de imagem (e vice versa: a detecção de movimento só é possível analisando imagens separadas temporalmente).

## 2.4. Resumo

Foram neste capítulo introduzidos alguns conceitos e definições, que serão úteis para os restantes capítulos desta tese.

Foram em particular apresentados os conceitos de ponto correspondente e disparidade inseridos no problema de recuperação de profundidade (stereo), e de fluxo óptico (sequências de imagem).

---

<sup>7</sup>O problema da abertura é parcialmente resolvido.

Entre disparidade e fluxo óptico existe uma dualidade que pode ser útil para o desenho de novos algoritmos.

Os problemas de cálculo de disparidade e de fluxo óptico são problemas mal postos pelo que normalmente exigem técnicas adicionais de filtragem (regularização).

A estimativa de disparidade e de fluxo óptico é fiável próximo de cantos na imagem. Em regiões cujos contornos tenham direcções aproximadas das linhas epipolares (sendo o objectivo estimar disparidade) ou direcções aproximadamente radiais em relação ao FOE/C (sendo o objectivo estimar fluxo óptico) e em regiões sem variações de brilho, a estimativa é pouco fiável ou impossível.





## CAPÍTULO 3

### Detecção de Obstáculos baseada em Stereo Binocular

#### 3.1. Introdução

Considerando o objectivo genérico de navegação autónoma para robótica móvel, um dos primeiros níveis a construir é naturalmente o que evita obstáculos eventualmente inesperados.

Para evitar um obstáculo é necessário em primeiro lugar detectá-lo e com este objectivo vários sensores têm vindo a ser usados, tais como sonares, sensores infra-vermelhos, lasers, etc. A visão, embora possuindo um grande potencial, tem sido excluída destes sistemas devido, em grande parte, ao grande peso computacional requerido pelas abordagens mais clássicas de visão.

No entanto, nos sistemas biológicos a visão é frequentemente usada para resolver, de forma eficiente, um grande número de problemas complexos como sejam navegação, manipulação, auto-localização, etc. Assim, é legítimo esperar que a visão possa também desempenhar um papel chave num vasto conjunto de problemas no domínio da robótica.

Procurava-se tradicionalmente, com informação visual, de alguma forma reconstruir tridimensionalmente um cenário, para depois calcular uma trajectória segura. Este era de facto um objectivo aliciante mas computacionalmente muito dispendioso para além de ser intrinsecamente um problema mal posto (ver capítulo 2). É neste contexto que na área da visão têm vindo a surgir correntes de pensamento<sup>1</sup> mais pragmáticas, onde os objectivos são especializados e as soluções surgem mais facilmente (normalmente com uma carga

---

<sup>1</sup>na nomenclatura inglesa são típicas as designações "active vision" e "purposive vision" para estas correntes de pensamento (ver cap. 1).

computacional significativamente menor) inspiradas com frequência nos sistemas de visão biológicos, psicologia, psico-física, etc. [Gibson51].

Nesta tese é proposto um método de detecção de obstáculos, para robots móveis, usando como hipótese fundamental o facto de o robot móvel se deslocar sobre um pavimento plano, e definindo um obstáculo como todo e qualquer ponto 3D que esteja fora deste plano. Neste enquadramento e para esta tarefa específica, é possível (dentro da filosofia "purposive vision") desenhar algoritmos de percepção visual rápidos e robustos, contrastando com as abordagens clássicas (reconstrução). Para a solução deste problema, têm sido essencialmente exploradas duas abordagens, uma baseada em "Fluxo Óptico" (como em [Mallot91] e [Guarnotta]) e a outra baseada em sistemas stereo (como em [Zielke90], [Ferrari90] e [Ferrari92]).

A abordagem baseada em "Fluxo Óptico" depende, para efeitos de inicialização e de operação, do movimento do robot. Pelo contrário a abordagem stereo pode ser testada independentemente do movimento do robot (pode inclusivamente ser testada na situação estática) e não invalida futuras extensões com técnicas de Fluxo Óptico. Estas observações levaram à escolha da abordagem stereo para o Algoritmo de Detecção de Obstáculos sobre Pavimento Plano (ADOPP) que neste capítulo se descreve.

Uma só imagem não permite geralmente fazer afirmações absolutas sobre dimensões ou profundidades das diversas zonas do cenário observado. Duas imagens adquiridas simultaneamente (par stereo) já permitem recuperações de profundidades para pontos tridimensionais simultaneamente visíveis em ambas as imagens, i.e., pontos cujas projecções, uma na imagem direita  $(x_r, y_r)$  e outra na imagem esquerda  $(x_l, y_l)$ , estão dentro dos limites das respectivas imagens (ver capítulo 2).

O estabelecimento da correspondência entre  $(x_r, y_r)$  e  $(x_l, y_l)$  ("matching") é necessário para o cálculo da profundidade. O desenho tradicional de sistemas stereo de recuperação visual de profundidades começa então pela pesquisa de correspondências: "*Dado um pixel  $(x_r, y_r)$ , qual o correspondente pixel  $(x_l, y_l)$  no sentido de que ambos são projecções de um ponto desconhecido  $(X, Y, Z)$ ?*". A primeira resposta é naturalmente a de pesquisa de toda a imagem esquerda à procura do ponto  $(x_l, y_l)$  cujo brilho seja o mais semelhante ao do ponto  $(x_r, y_r)$ . Na realidade esta pesquisa completa é desnecessária e não

é realizada, porque o conhecimento dos parâmetros das câmaras e da transformação de coordenadas entre elas, restringe as possibilidades de correspondências a linhas denominadas por "linhas epipolares" (ver capítulo 2).

Esta é a estratégia tradicional de reconstrução 3D: para todos os  $(x_r, y_r)$  procuram-se na imagem esquerda os correspondentes  $(x_l, y_l)$ , respeitando a restrição epipolar.

Supondo agora que as câmaras só vêem pontos do plano do pavimento conhecido (existe uma relação conhecida entre  $X, Y$  e  $Z$ ), acontece que para um dado pixel  $(x_r, y_r)$  existe um e um só ponto  $(X, Y, Z)$  e existe um e um só pixel  $(x_l, y_l)$ , i.e., o ponto correspondente é encontrado analiticamente sem qualquer pesquisa (fig.3.1).

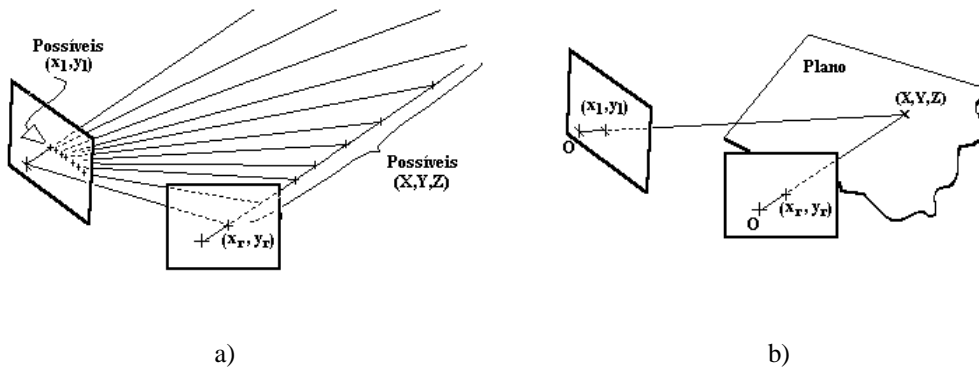


Fig. 3.1 Simplificação introduzida pela hipótese de observação de um "pavimento plano" conhecido.  
 a) Para um  $(x_r, y_r)$ , são possíveis infinitos  $(X, Y, Z)$  (o cenário é livre), pelo que são possíveis vários  $(x_l, y_l)$ .  
 b) Para um  $(x_r, y_r)$ , o conhecimento do plano do pavimento e da relação entre câmaras leva à existência de um e um só  $(X, Y, Z)$  e de um e um só  $(x_l, y_l)$ .

A junção da informação da restrição epipolar com a hipótese de que a superfície 3D visualizada é um plano, define completamente as correspondências. O conjunto de todas as correspondências constitui um *mapa de correspondências*. Este mapa é guardado para futura referência e simplificando a linguagem, diz-se que guarda a correspondência entre ambas as imagens.

Com o movimento do robot suporte das câmaras, estas verão não só pavimento como também obstáculos. *Obstáculos* são definidos como pontos localizados fora do plano do pavimento (incluindo buracos no pavimento)<sup>2</sup>. Estes obstáculos induzirão correspondências diferentes das calculadas analiticamente para o plano do pavimento,

<sup>2</sup>Para serem detectados é necessário que exista textura nas superfícies visualizadas.

peço que o teste de comparação de brilhos nos pontos  $(x_r, y_r)$  e  $(x_l, y_l)$  deve resultar falso. Na fig.3.2 pode ser observado um exemplo simples de alteração do mapa de disparidades devida à presença de uma caixa (obstáculo) sobre o pavimento.

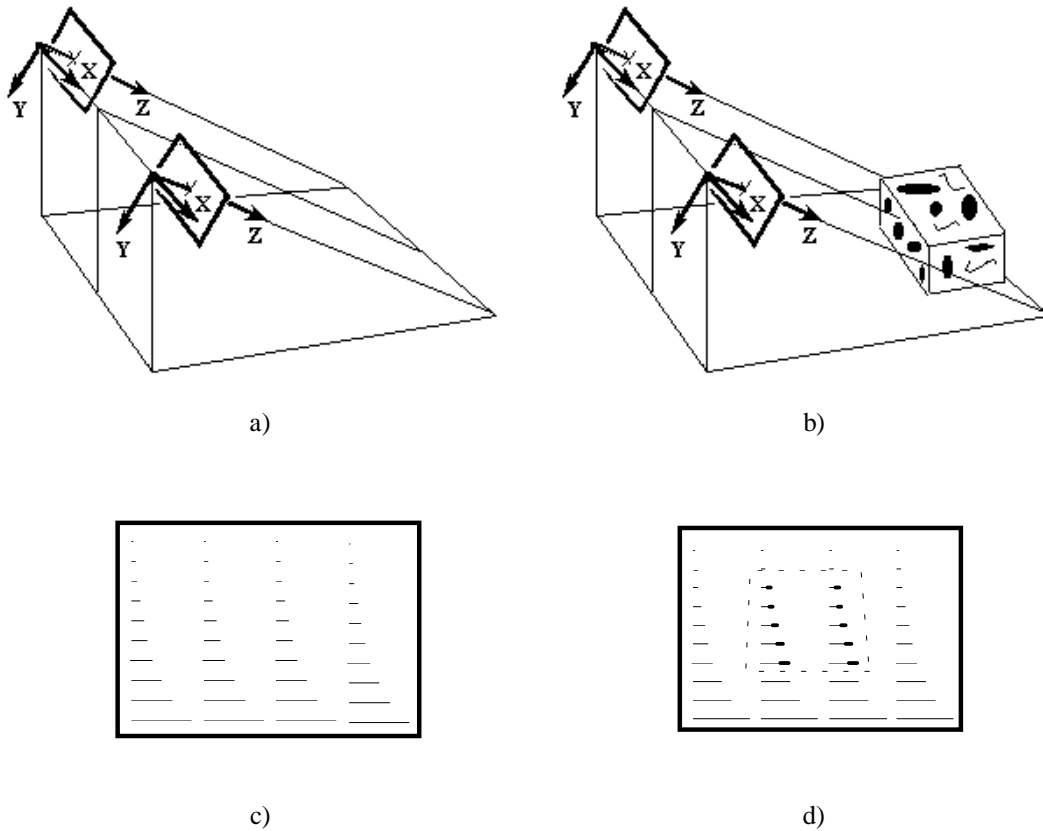


Fig. 3.2 Exemplo simples de alterações do mapa de disparidades, devido à presença de uma "caixa" sobre o pavimento; 'c' é o mapa de disparidades correspondente ao pavimento e resulta da situação da figura 'a'; 'd' resulta da situação da figura 'b' e mostra o mapa de disparidades correspondente ao pavimento com algumas disparidades aumentadas devido à presença do caixote (zona a tracejado apresenta disparidades diferentes em relação à figura 'c').

Esta é a ideia base do método proposto: conhecido o plano onde o robot suporte das câmaras se movimentam, os parâmetros da geometria stereo e os parâmetros das câmaras, é construído um mapa de correspondências de referência<sup>3</sup>, cujo teste falha nos pontos onde existem obstáculos (obstáculos são violações no mapa de referência).

---

<sup>3</sup>ver-se-á mais à frente que não são necessários explicitamente estes parâmetros; bastará para o cálculo do mapa de referência um conjunto de pontos correspondentes, observado directamente das imagens de inicialização.

Comparando o algoritmo de reconstrução 3D e o ADOPP, verifica-se neste último que a hipótese de movimento sobre um "pavimento plano" conhecido, evita as tradicionais pesquisas de correspondências, com a criação de um "mapa de correspondências de referência" a ser testado durante a operação<sup>4</sup>. O facto de tão somente serem testados valores de referência e não serem efectuadas pesquisas (e portanto não serem encontradas correspondências correctas para os pontos pertencentes aos obstáculos), limita a informação final a indicações de existências de obstáculos deixando para níveis superiores afirmações sobre respectivas alturas (ou profundidades). É ainda este facto que faz do ADOPP um algoritmo rápido comparado com a reconstrução 3D.

Na continuação deste capítulo é detalhado o ADOPP, com realce para a existência de duas fases (ver fluxograma da fig.3.3): Inicialização e Detecção.

Da breve introdução ao ADOPP, ficou patente a necessidade de cálculo de um mapa de referência que antecipe e guarde as correspondências entre as imagens direita e esquerda para os diversos pontos do plano do pavimento. Este mapa é calculado na fase de *inicialização*. Terminada a inicialização, começa o ciclo de testes que constitui a fase de *detecção*. Nesta fase, o tempo de computação deve ser minimizado, permitindo ao veículo um movimento rápido e seguro (seguro no sentido de que reconhece com a antecedência adequada eventuais obstáculos).

---

<sup>4</sup>um mapa de referência mantém-se válido enquanto não existirem variações da geometria stereo, dos parâmetros das câmaras ou ainda reposicionamentos da(s) câmara(s) relativamente ao plano do pavimento.

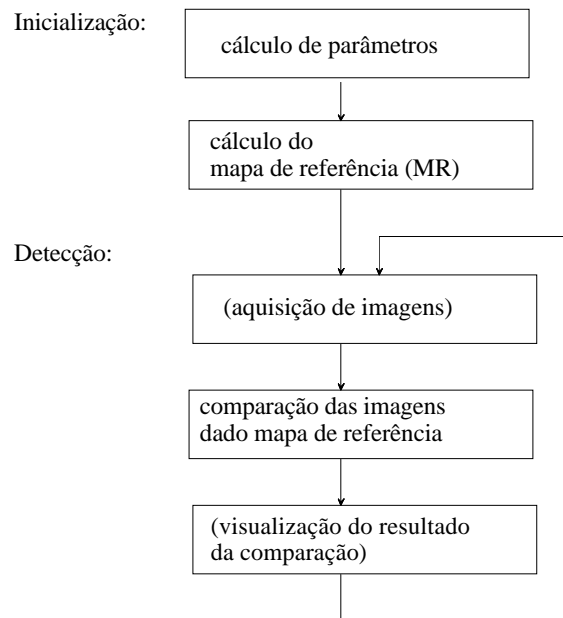


Fig. 3.3 Fluxograma do Algoritmo de Detecção de Obstáculos sobre Pavimento Plano (ADOPP).

### 3.2. Inicialização

Como atrás foi dito, antes da fase de detecção é necessária uma inicialização da qual resulta essencialmente uma tabela bidimensional: o mapa de referência, que descreve a disparidade (ou correspondência) para qualquer pixel relativo a um ponto do pavimento. Este mapa concentra informação da geometria do sistema (relação entre câmaras), do sistema de projecção e do plano.

A forma analítica do mapa de referência pode ser obtida a partir de um conjunto de pressupostos que têm vindo a ser introduzidos ao longo deste capítulo e que agora se completam e resumem:

1) o processo de formação de imagem, do ponto de vista geométrico, pode ser representado pelo modelo de câmara pontual (do inglês "pin-hole camera"); i.e., são realizadas projecções perspectiva, dadas por:

$$\begin{bmatrix} x_l \\ y_l \end{bmatrix} = \frac{f_l}{Z_l} \begin{bmatrix} X_l \\ Y_l \end{bmatrix} \quad \begin{bmatrix} x_r \\ y_r \end{bmatrix} = \frac{f_r}{Z_r} \begin{bmatrix} X_r \\ Y_r \end{bmatrix} \quad (1)$$

onde  $(X_l, Y_l, Z_l)$  e  $(X_r, Y_r, Z_r)$  são representações do mesmo ponto tridimensional no sistema de coordenadas da câmara esquerda e da câmara direita respectivamente;  $f_l$  e  $f_r$  são distâncias focais;  $(x_l, y_l)$  e  $(x_r, y_r)$  são as projecções (2D), respectivamente na câmara direita e na câmara esquerda, do referido ponto tridimensional.

2) as câmaras têm uma posição rígida uma em relação à outra<sup>5</sup>. Esta relação de transformação de coordenadas é representável por uma matriz homogénea  $T$  (4x4) ou pelo conjunto de uma matriz de rotação  $R$  (3x3) e de um vector de translação  $P$  (3x1).

$$\begin{bmatrix} X_l & Y_l & Z_l & 1 \end{bmatrix}^T = T_r^l \begin{bmatrix} X_r & Y_r & Z_r & 1 \end{bmatrix}^T \quad (2)$$

$$T_r^l = \begin{bmatrix} R_r^l & P_r^l \\ \vec{0} & 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & p_x \\ r_{21} & r_{22} & r_{23} & p_y \\ r_{31} & r_{32} & r_{33} & p_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3)$$

3) ambas as câmaras observam um plano, descrito por:

$$Z_r = Z_0 + A.X_r + B.Y_r \quad (4)$$

expresso nas coordenadas da câmara direita e representado pelos parâmetros  $A, B, Z_0$ .

Da manipulação algébrica destes pressupostos (1,2,3), surge então a expressão para o mapa de referência (mapa de correspondências)  $\tilde{\mathcal{F}}$ , que para cada pixel da imagem direita, determina as coordenadas do pixel correspondente na imagem esquerda no pressuposto de que o ponto visualizado pertença ao plano do pavimento:

---

<sup>5</sup>neste momento supõe-se, sem perda de generalidade, que o sistema dispõe de duas câmaras, i.e., não importa se são duas câmaras simuladas por uma só com um par de espelhos acoplado.

$$\begin{aligned}
 \begin{bmatrix} x_l \\ y_l \end{bmatrix} &= \tilde{\mathcal{F}}(x_r, y_r) \\
 &= \begin{bmatrix} \mathcal{F}_x(x_r, y_r) \\ \mathcal{F}_y(x_r, y_r) \end{bmatrix} \\
 &= \frac{\begin{bmatrix} b_1 x_r + b_2 y_r + b_3 \\ c_1 x_r + c_2 y_r + c_3 \end{bmatrix}}{a_1 x_r + a_2 y_r + 1}
 \end{aligned} \tag{5}$$

onde  $a_1, a_2, b_1, b_2, b_3, c_1, c_2, c_3$  têm as seguintes expressões<sup>6</sup>:

$$\begin{aligned}
 [a_1 \quad a_2] &= \frac{1}{(r_{33}Z_0 + p_z)f_r} [r_{31}Z_0 - p_z A \quad r_{32}Z_0 - p_z B] \\
 [b_1 \quad b_2 \quad b_3] &= \frac{f_l}{(r_{33}Z_0 + p_z)f_r} [r_{11}Z_0 - p_x A \quad r_{12}Z_0 - p_x B \quad f_r(r_{13}Z_0 + p_x)] \\
 [c_1 \quad c_2 \quad c_3] &= \frac{f_l}{(r_{33}Z_0 + p_z)f_r} [r_{21}Z_0 - p_y A \quad r_{22}Z_0 - p_y B \quad f_r(r_{23}Z_0 + p_y)]
 \end{aligned}$$

A expressão (5) é conhecida e generalizada na teoria dos invariantes [Zisserman92] como uma *transformação projectiva* de um *plano projectivo* sobre outro, sendo normalmente representada por uma matriz 3x3 num sistema de coordenadas homogéneo:

$$\begin{bmatrix} \lambda x_l \\ \lambda y_l \\ \lambda \end{bmatrix} = \begin{bmatrix} b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \\ a_1 & a_2 & 1 \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ 1 \end{bmatrix}$$

A expressão (5) é ainda referenciada com uma forma semelhante em [Mallot92] e em [Zielke90]. Mallot e Zielke realizam "inversão da distorção introduzida pela perspectiva de observação do pavimento" para ambas as imagens do par stereo. No trabalho descrito nesta tese, só a imagem esquerda é transformada: pela expressão (5) é mapeada sobre a imagem direita.

---

<sup>6</sup>estas fórmulas pretendem mostrar que os pressupostos estão incluídos em (5); no seguimento do texto verificar-se-á que é realizada uma estimação directa de  $a_1, a_2, b_1, b_2, b_3, c_1, c_2, c_3$  sem conhecimento dos parâmetros de câmaras, geometria stereo e pavimento plano.



A expressão (5) indica para cada pixel  $(x_r, y_r)$  o pixel correspondente  $(x_l, y_l)$  supondo que ambos são projecções de um ponto pertencente ao plano do pavimento. Para cada ponto  $(x_r, y_r)$  pode ser alternativamente indicada uma disparidade. O conjunto destas disparidades constitui o mapa de disparidades de referência  $\bar{\mathcal{A}}(x_r, y_r)$  :

$$\begin{aligned}\bar{\mathcal{A}}(x_r, y_r) &= \begin{bmatrix} x_l \\ y_l \end{bmatrix} - \begin{bmatrix} x_r \\ y_r \end{bmatrix} \\ &= \bar{\mathcal{F}}(x_r, y_r) - \begin{bmatrix} x_r \\ y_r \end{bmatrix} \\ &= \frac{\begin{bmatrix} -a_1 x_r^2 + (b_1 - 1)x_r - a_2 x_r y_r + b_2 y_r + b_3 \\ c_1 x_r - a_1 x_r y_r + (c_2 - 1)y_r - a_2 y_r^2 + c_3 \end{bmatrix}}{a_1 x_r + a_2 y_r + 1}\end{aligned}\quad (6)$$

$\bar{\mathcal{F}}(x_r, y_r)$  e  $\bar{\mathcal{A}}(x_r, y_r)$  (expressões (5) e (6)) são conceptualmente idênticos, têm o mesmo conteúdo informativo, sendo por isso muitas vezes usada a expressão mapa de referência para indicar indistintamente um dos dois. Na prática, na programação do algoritmo é preferível usar  $\bar{\mathcal{F}}$ , enquanto na visualização é preferível usar  $\bar{\mathcal{A}}$ .

### 3.2.1. Estimação de parâmetros do mapa de referência

Observando a expressão do mapa de referência (5), verifica-se que ela depende de um conjunto de oito parâmetros  $(a_1, a_2, b_1, b_2, b_3, c_1, c_2, c_3)$ . Conclui-se então que o mapa de referência pode ser obtido facilmente quando forem conhecidos esses parâmetros pelo que a primeira etapa consiste na sua estimação.

Fazendo uma ligeira manipulação algébrica, separando e reescrevendo a equação vectorial (5) obtém-se:

$$\begin{cases} x_l x_r a_1 + x_l y_r a_2 - x_r b_1 - y_r b_2 - b_3 = -x_l \\ y_l x_r a_1 + y_l y_r a_2 - x_r c_1 - y_r c_2 - c_3 = -y_l \end{cases}\quad (7)$$

Para encontrar os oito parâmetros são necessárias pelo menos oito equações linearmente independentes, o que se consegue com quatro cópias das duas equações anteriores, naturalmente com um mínimo de quatro conjuntos  $(x_r, y_r, x_l, y_l)$  distintos não colineares 3 a 3 ([Zisserman92]). Se estiverem disponíveis mais conjuntos de pontos correspondentes, é construído um sistema sobre-determinado.

Definindo então o vector  $u$  que agrupa o conjunto de parâmetros a estimar:

$$u = [a_1 \ a_2 \ b_1 \ b_2 \ b_3 \ c_1 \ c_2 \ c_3]^T \quad (8)$$

e sendo P o conjunto de dados (N pontos correspondentes) necessários para a estimação dos parâmetros:

$$P = \{(x_{l_1}, y_{l_1}, x_{r_1}, y_{r_1}), \dots, (x_{l_N}, y_{l_N}, x_{r_N}, y_{r_N})\} \quad (9)$$

as duas equações (7) são repetidas N vezes, pelo que se obtêm um sistema de 2N equações a oito incógnitas. Este sistema pode ser facilmente representado em notação matricial:

$$Mu = v \quad (10)$$

mais concretamente:

$$\begin{bmatrix} x_{l_1}x_{r_1} & x_{l_1}y_{r_1} & -x_{r_1} & -y_{r_1} & -1 & 0 & 0 & 0 \\ y_{l_1}x_{r_1} & y_{l_1}y_{r_1} & 0 & 0 & 0 & -x_{r_1} & -y_{r_1} & -1 \\ x_{l_2}x_{r_2} & x_{l_2}y_{r_2} & -x_{r_2} & -y_{r_2} & -1 & 0 & 0 & 0 \\ y_{l_2}x_{r_2} & y_{l_2}y_{r_2} & 0 & 0 & 0 & -x_{r_2} & -y_{r_2} & -1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{l_i}x_{r_i} & x_{l_i}y_{r_i} & -x_{r_i} & -y_{r_i} & -1 & 0 & 0 & 0 \\ y_{l_i}x_{r_i} & y_{l_i}y_{r_i} & 0 & 0 & 0 & -x_{r_i} & -y_{r_i} & -1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{l_N}x_{r_N} & x_{l_N}y_{r_N} & -x_{r_N} & -y_{r_N} & -1 & 0 & 0 & 0 \\ y_{l_N}x_{r_N} & y_{l_N}y_{r_N} & 0 & 0 & 0 & -x_{r_N} & -y_{r_N} & -1 \end{bmatrix} \cdot u = \begin{bmatrix} -x_{l_1} \\ -y_{l_1} \\ -x_{l_2} \\ -y_{l_2} \\ \vdots \\ -x_{l_i} \\ -y_{l_i} \\ \vdots \\ -x_{l_N} \\ -y_{l_N} \end{bmatrix} \quad (11)$$

Este sistema tem solução desde que existam pelo menos oito equações linearmente independentes. A solução usualmente adoptada para sistemas sobre-determinados é a pseudo-inversa (correspondendo a uma estimativa de mínimos quadrados):

$$u = [M^T M]^{-1} M^T v \quad (12)$$

A partir do momento que o vector de parâmetros  $\mathbf{u}$  está calculado, o mapa de referência é obtido simplesmente aplicando a equação (5) para todos os pontos (x,y) dentro dos limites de uma imagem<sup>7</sup>.

---

<sup>7</sup>um ponto (x,y) na imagem direita pode não ser observável na imagem esquerda (i.e. pode estar fora dos limites da imagem esquerda)

### 3.2.2. Obtenção de pontos correspondentes

O cálculo dos parâmetros  $\mathbf{u} = [a_1 \ a_2 \ b_1 \ b_2 \ b_3 \ c_1 \ c_2 \ c_3]^T$  (pela expressão (12)) necessita de um mínimo de quatro pares de pontos correspondentes  $\{x_r, y_r, x_l, y_l\}$ .

Uma estimativa inicial dos parâmetros  $\mathbf{u}$  é obtida à custa de quatro pares de pontos correspondentes, seleccionados manualmente. Para tal, são colocados sobre o pavimento quatro padrões que facilitam a tarefa de emparelhamento.

O algoritmo prossegue com a aplicação de um detector de cantos nas imagens e de um processo de emparelhamento destes pontos, guiado pela estimativa inicial dos parâmetros (ver fluxograma da fig.3.4).

O emparelhamento é refinado, de forma a ser encontrado para cada ponto (canto) da imagem direita o ponto da imagem esquerda que melhor lhe corresponde. A procura local da melhor correspondência é realizada com a obtenção do mínimo de SSD<sup>8</sup> numa janela de procura ou com uma evolução iterativa à volta da correspondência original. Pode ainda ser usado um algoritmo que ajusta localmente um parabolóide de valores de SSD e escolhe o mínimo (esta é uma técnica típica dos sistemas stereo que calculam profundidades e fornece -com alguma limitação- correspondências com resolução sub-pixel).

A obtenção automática de pontos correspondentes é repetida uma ou duas vezes. Quando a parametrização ainda não está suficientemente bem feita, falham algumas correspondências. Espera-se que a iteração actual melhore a parametrização conseguindo na próxima iteração juntar alguns dados recusados actualmente. Pretende-se assim melhorar a parametrização do plano.

---

<sup>8</sup>SSD: "Sum of the Squared Differences". Medida de semelhança entre janelas de imagens, definida no capítulo 2. Um mínimo de SSD significa uma semelhança máxima.

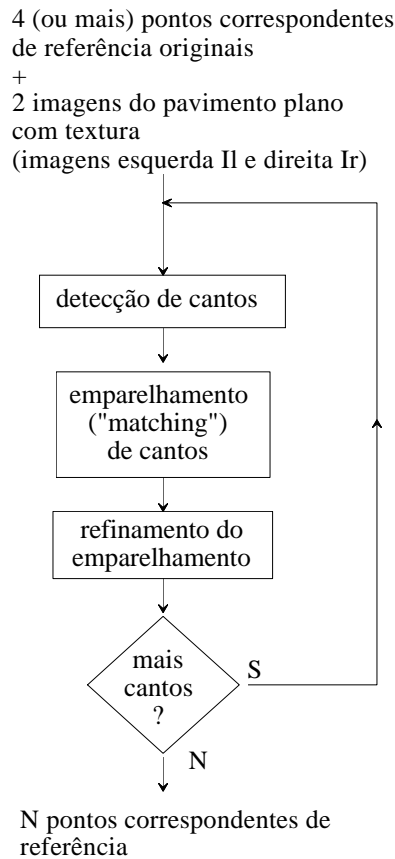


Fig. 3.4 Fluxograma: melhoria de dados (pontos correspondentes) para cálculo de parâmetros do mapa de referência. "Refinamento do emparelhamento" altera localmente  $(x_l, y_l)$  de modo a melhorar a semelhança (SSD) entre  $I_l(x_l, y_l)$  e  $I_r(x_r, y_r)$  numa janela de comparação.

### 3.2.3. Métodos alternativos para a obtenção do mapa de referência

Alternativamente à construção do mapa de referência por via de uma pré-determinação de oito parâmetros, pode ser usado um "stereo matcher" por exemplo de contornos ou de cantos complementado por um algoritmo de interpolação de dados esparsos<sup>9</sup>.

Para interpolação de dados esparsos podem ser usadas técnicas de regularização aplicadas à reconstrução de superfícies ([Terzopoulos84], [Geman84], [Gaspar92]) ou técnicas de triangulação (por exemplo triangulação de Delaunay).

---

<sup>9</sup>esta é uma perspectiva tradicional, com um peso computacional significativo, no entanto esta é uma etapa de inicialização, portanto o tempo não é crítico.

### 3.3. Detecção

Dada a tabela de correspondências entre as imagens direita e esquerda para os diversos pontos do plano do pavimento, a "detecção" consiste na sua simples verificação (detecção de eventuais violações) para cada novo par de imagens.

Para cada ponto da imagem direita  $(x_r, y_r)$ , a tabela indica o ponto correspondente na imagem esquerda  $(x_l, y_l)$ , supondo que ambos se referem a um ponto  $(X, Y, Z)$  do pavimento. Se os pixels  $(x_r, y_r)$  e  $(x_l, y_l)$  de facto se referem ao mesmo ponto  $(X, Y, Z)$  do pavimento, então os seus brilhos devem ser semelhantes. Caso os brilhos não sejam semelhantes os pontos não são correspondentes: o ponto  $(x_r, y_r)$  não é uma projecção de um ponto do pavimento, i.e., pertence a um obstáculo.

Portanto a verificação consiste numa simples comparação de brilhos, onde as diferenças são interpretadas como pontos de obstáculo.

A verificação de semelhança entre dois pontos correspondentes, costuma ser alargada à vizinhança, para atenuar o inevitável efeito do ruído. O cálculo de diferença (ao quadrado) acumulado ao longo de uma vizinhança é uma medida de comparação usual, e normalmente referida como SSD ("Sum of the Squared Differences"). A SSD toma o valor nulo quando os dois pontos a serem comparados têm o mesmo brilho (assim como os seus vizinhos mais próximos). É uma medida simplificada de correlação:

$$I_{SSD_{test}}(x, y) = SSD(I_l(\vec{\mathcal{A}}(x, y)), I_r(x, y)) \\ = \sum_{i=-W}^W \sum_{j=-W}^W [I_l(f_x(x, y) + i, f_y(x, y) + j) - I_r(x + i, y + j)]^2 \quad (13)$$

Esta forma de SSD compara uma janela centrada em  $(x, y)$  com uma janela centrada em  $\vec{\mathcal{A}}(x, y)$ . Foi este o critério originalmente usado neste trabalho.

No entanto, para o método proposto, é conhecida a disparidade esperada para cada ponto, pelo que é fácil calcular a transformação de toda a janela (uma janela quadrada na imagem direita aparece distorcida na imagem esquerda). Propõe-se assim um SSD ligeiramente diferente para levar em conta a realidade da distorção da janela (este efeito só será significativo para janelas de comparação grandes):

$$\begin{aligned}
 I_{SSD_{test}}(x, y) &= SSD(I_l(\tilde{\mathcal{F}}(x, y)), I_r(x, y)) \\
 &= \sum_{i=-W}^W \sum_{j=-W}^W \left[ I_l(f_x(x+i, y+j), f_y(x+i, y+j)) - I_r(x+i, y+j) \right]^2
 \end{aligned} \tag{14}$$

Resumindo, para todos os pontos  $(x_r, y_r)$  da imagem direita são encontrados e comparados os pontos  $(x_l, y_l) = \tilde{\mathcal{F}}(x_r, y_r)$  da imagem esquerda, por via da SSD. Se a SSD tiver um valor nulo, o ponto  $(x_r, y_r)$  pertence ao pavimento ou a uma zona sem textura do obstáculo. Se a SSD tiver um valor significativo (a determinar experimentalmente), então o ponto  $(x_r, y_r)$  pertence ao obstáculo.

### 3.3.1. Filtragem morfológica

A filtragem morfológica surge da constatação experimental de dois factos: o primeiro nota que grandes variações de brilho no pavimento (exemplo variação branco-preto dos referenciais colocados sobre o pavimento) motivam, após teste de comparação de brilhos, pontos isolados "fortes" (valor de SSD  $\gg 0$ ); o segundo facto nota que as regiões onde ocorrem violações serão tanto mais largas quanto mais altos forem os obstáculos.

Os pontos isolados "fortes", surgem devido à "discretização" do mapa de referência (aproximação por valores inteiros das concretizações da expressão analítica do mapa de referência). Estes pontos são indicadores erróneos de obstáculos pelo que devem ser evitados.

Ponderando os dois factos, constrói-se a ideia de que pontos isolados devem ser penalizados. A forma escolhida, é heurística e baseia-se na contagem de pontos "fortes" em janelas locais.

Filtro morfológico:

$$\begin{aligned}
 f_{thres1}(a) &= \begin{cases} 0, & a < thres1 \\ 1, & a \geq thres1 \end{cases} \\
 f_{thres2}(b) &= \begin{cases} 0, & b < thres2 \\ 255, & b \geq thres2 \end{cases} \\
 I(x, y) &= f_{thres2} \left[ \sum_{c=x-W_2}^{x+W_2} \sum_{l=y-W_2}^{y+W_2} f_{thres1}(I_{SSD_{test}}(c, l)) \right]
 \end{aligned} \tag{15}$$

Nas expressões, *thres1* ("threshold 1" - limiar 1) está relacionado com os níveis esperados para o *SSD<sub>test</sub>*, e estabelece o limiar de separação entre zona semelhante e zona diferente.

O *thres2* ("threshold 2" - limiar 2) controla a largura dos "traços de violação" e está naturalmente relacionado com a largura da janela de teste  $W_2$ . A heurística usada foi  $thres2=2*W_2-1$ . Se os traços de violação forem estreitos (ou forem pontos isolados) então é provável que digam respeito a erros de discretização. Um traço de largura 1 numa janela  $W_2 \times W_2$  tem  $W_2$  pontos, e um traço de largura 2 na mesma janela tem  $2*W_2$  pontos (a heurística está colocada entre estes dois valores).

### 3.3.2. Diagrama de blocos da fase de detecção

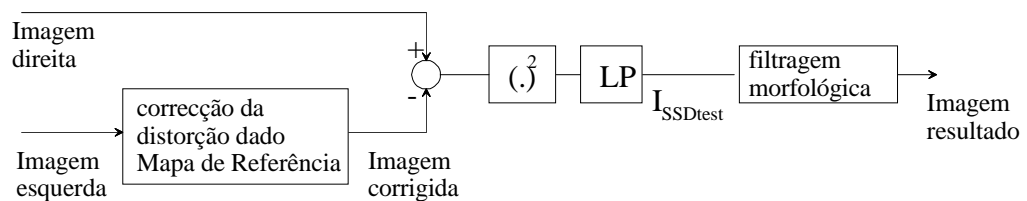


Fig. 3.6 Diagrama de blocos da fase de detecção do ADOPP

Nota específica de implementação: a SSD foi simplificada pela utilização de módulos no lugar de quadrados.

### 3.4. Evoluções possíveis

A investigação tendo por objectivo base a detecção de obstáculos sobre pavimento plano continua em progresso. Como foi referido na introdução, duas abordagens têm vindo a ser exploradas: a abordagem stereo e a abordagem baseada em fluxo óptico (a descrever no capítulo 6).

Na "abordagem stereo", por exemplo, começam a ser realizados estudos no sentido de relaxar a hipótese de "relação constante entre câmaras", porque começa a ser frequente nos laboratórios de visão, a existência de "cabeças stereo". Uma cabeça stereo é constituída por um par de câmaras e por um conjunto de motores associados, organizados

no sentido de simularem os graus de liberdade dos "olhos humanos". Neste caso a relação entre câmaras deixa de ser constante. Uma extensão ao ADOPP, consiste na salvaguarda de vários mapas de referência eventualmente interpoláveis ([Cornell92]). As técnicas de regularização podem ser úteis para compensar pequenas variações na geometria.

A informação do ADOPP concentra-se nas regiões com textura ou de contornos. Uma pré-deteção de contornos para diminuição do número de testes, pode diminuir a quantidade de processamento. A dúvida neste caso consiste na comparação de velocidade de um detector de contornos comparado com "um verificador de SSDs".

Pode também ser útil uma fusão da abordagem "stereo" com a abordagem baseada em fluxo óptico, especialmente porque existem padrões de brilho motivados por obstáculos, somente identificáveis por uma das abordagens.

### **3.5. Resumo**

Apresentou-se neste capítulo um "Algoritmo de Detecção de Obstáculos sobre Pavimento Plano - ADOPP" (na nomenclatura inglesa: "Ground Plane Obstacle Detection - GPOD"). Este algoritmo é constituído por duas fases essenciais: fase de inicialização e fase de detecção.

Da inicialização resulta o mapa de referência que antecipa e guarda as correspondências entre as imagens direita e esquerda para os diversos pontos do plano do pavimento. Este mapa é calculado de forma paramétrica.

A fase de "detecção" é constituída por um ciclo eterno de testes do mapa de referência, sendo cada teste motivado por um novo par de imagens. O resultado dos testes são imagens com indicações de obstáculos (fronteiras).

Em relação às aproximações tradicionais de reconstrução tridimensional, o ADOPP não necessita de calibração métrica (a inicialização é realizada exclusivamente com imagens) e substitui procuras locais por testes locais (reconstrução supõe uma etapa de "matching" de dados entre imagens, o que se traduz num problema de procura -ver cap.2-). A substituição das procuras locais por testes permite que o algoritmo seja rápido.

A não utilização de calibração métrica impede afirmações tais como "obstáculo com h metros" ou "obstáculo a x metros": o resultado final é obtido em "pixels obstáculo na imagem direita". Esta informação é contudo suficiente para a implementação de um



comportamento reactivo de "obstacle avoidance", baseado num mapa de acções, que associa a zonas da imagem certas atitudes (exemplo: "obstáculo na metade esquerda da imagem, então virar à direita...").



## CAPÍTULO 4

### Aparato Experimental.

Pretende-se neste capítulo apresentar o aparato experimental necessário à realização e teste do Algoritmo de Detecção de Obstáculos sobre Pavimento Plano (ADOPP) apresentado no capítulo 3.

A detecção de obstáculos pressupõe alguma informação de profundidade. Para a obtenção de informação de profundidade usando visão, é conhecida a necessidade de um conjunto de imagens. Este conjunto pode ser formado por duas imagens adquiridas simultaneamente (stereo binocular) ou por um conjunto de imagens adquiridas sequencialmente<sup>1</sup>.

Em ambos os casos existe para além da aquisição de imagem o problema adicional da "aquisição simultânea" ou da "aquisição em sequência". Na solução deste problema, podem ser usadas várias técnicas, que recorrem com maior ou menor grau à placa de aquisição de imagens.

A escolha dos parâmetros da câmara e da placa de aquisição, influencia directamente as imagens obtidas. Por exemplo, uma câmara equipada com uma lente de pequena angular vê obstáculos mais pequenos a uma maior distância, mas numa faixa mais reduzida, em comparação com uma lente de grande angular (a distância focal da lente está directamente relacionada com o ângulo de visão da câmara e com as distâncias observadas). Este é um tópico importante porque dele depende a quantidade de informação que o algoritmo de detecção de obstáculos pode fornecer.

---

<sup>1</sup>a expressão "alguma informação de profundidade" salienta a ideia central desta tese de que para a detecção de obstáculos é suficiente informação relativa ou simplificada, de profundidade: são por exemplo suficientes um *mapa de disparidades de referência* e *pares de imagens stereo* para testar este mapa de referência (não é necessário o tradicional mapa de profundidades necessariamente precedido por uma calibração cuidadosa).

Assim neste capítulo apresentam-se as estratégias consideradas para a aquisição simultânea, referem-se as capacidades da placa de aquisição e por fim faz-se uma caracterização do sistema escolhido.

## **4.1. O sistema de aquisição**

O algoritmo de detecção de obstáculos, descrito no capítulo 3, está concebido para operar sobre pares de imagens stereo (no capítulo 6 é apresentado um algoritmo dual baseado em sequências de imagens).

Quando se opta por usar imagens stereo, surge naturalmente a necessidade de projectar e dimensionar o sistema de aquisição constituído pela(s) câmara(s), respectivo suporte e placa de aquisição.

Existem várias formas de adquirir imagens stereo. A solução mais comum usa duas câmaras fixas, sobre um suporte horizontal, separadas por uma distância 'b' ("baseline"), possivelmente com vergência não nula (i.e., tendo um ponto de fixação).

Este aparato dá muitas vezes origem a problemas na estimação ou teste de correspondências devido à diferença de ganhos entre câmaras, de aberturas (íris), de distâncias focais, etc.

Nesta tese usa-se uma solução alternativa, que tem por base um conjunto de espelhos planos e uma única câmara, ultrapassando portanto o problema das diferenças entre os parâmetros das câmaras. São em particular descritas duas estruturas de espelhos. Uma das estruturas contém quatro espelhos (fig. 4.1.a) e a outra contém simplesmente dois (fig. 4.1.b).

---

<sup>2</sup>Charge Coupled Device

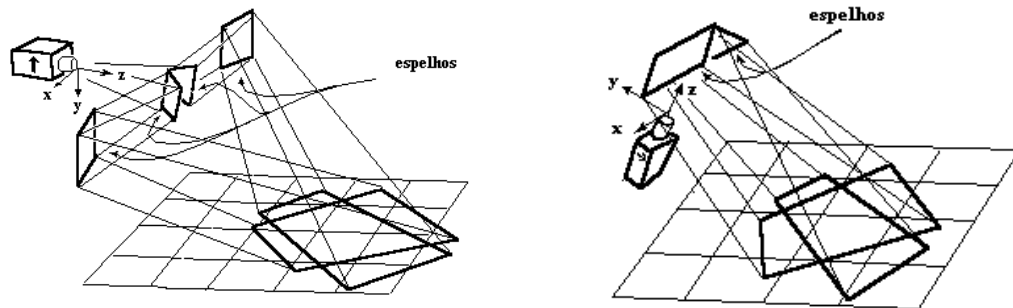


Fig. 4.1: Sistema de aquisição de imagens stereo baseado numa única câmara e num conjunto de quatro e ou de dois espelhos.

Apesar da utilização da estrutura baseada em espelhos se traduzir numa redução do campo de visão (a menos que sejam utilizados espelhos convexos), permite uma única aquisição de imagem (por par stereo), correspondendo a metade direita da imagem a uma câmara virtual e a metade esquerda da imagem à outra câmara virtual (na secção seguinte é definido o conceito de câmara virtual).

O sistema de dois espelhos implica inversão horizontal (ou vertical) nas imagens adquiridas permitindo no entanto uma estrutura mais compacta quando comparada com o sistema de quatro espelhos. Por esta razão, foi escolhido neste trabalho o sistema de dois espelhos ilustrado na figura 4.1.b.

## 4.2. O sistema de dois espelhos

### 4.2.1. Duas câmaras virtuais

A imagem reflectida por um espelho plano  $E$  e capturada por uma câmara  $C$ , pode ser obtida por uma câmara  $C_v$ , designada de virtual, de características semelhantes (distância focal, resolução, ganho, etc) colocada numa posição e orientação de acordo com as posições e orientações de  $C$  e de  $E$ .

Mais especificamente, sendo  $O_c$  o centro de projecção e  $(\hat{x}, \hat{y}, \hat{z})$  o sistema de coordenadas associado à câmara  $C$ ,  $n$  a normal ao espelho  $E$ , e  $D_{cm}$  a distância entre  $C$  e  $E$  medida segundo o eixo  $\hat{z}$ , então o centro de projecção  $O_{cv}$  e a orientação  $(\hat{x}_v, \hat{y}_v, \hat{z}_v)$  da câmara virtual  $C_v$  têm as seguintes expressões:

$$\begin{aligned}
 O_{C_v} &= O_c + D_{cm}\hat{z} + D_{cm}r(n, \hat{z}) \\
 (\hat{x}_v, \hat{y}_v, \hat{z}_v) &= (r(n, \hat{x}), r(n, \hat{y}), r(n, \hat{z})) \\
 r(n, v) &= v_1 - 2(n_1|v_1).n_1, \quad v_1 = v/|v|, \quad n_1 = n/|n|
 \end{aligned} \tag{1}$$

onde  $r(n, v)$  significa a reflexão do vector  $v$  pelo espelho plano caracterizado pela normal  $n$ .

Os sistemas de coordenadas de  $C$  e de  $C_v$  têm uma regra diferente de ordenação dos seus versores, o que se verifica na prática por uma reflexão horizontal (ou vertical) entre imagens adquiridas por uma câmara no lugar da câmara virtual ou pelo conjunto de uma câmara e um espelho.

Uma estrutura de espelhos correctamente desenhada e associada a uma câmara, permite a criação de várias câmaras virtuais. Em particular o sistema de dois espelhos e uma câmara escolhido (fig.4.1.b), gera duas câmaras virtuais (fig.4.2.a), existindo portanto um par stereo de câmaras que lhe é equivalente.

A modelização do sistema de dois espelhos é importante para uma melhor compreensão da geometria (útil por exemplo para simulações de espaço visível -ver secção seguinte) e permite a ligação a trabalho de investigação já realizado (por exemplo estudo da resolução de sistemas stereo: altura mínima detectada).

Na fig.4.2 apresenta-se um modelo para o sistema de dois espelhos e identificam-se os principais parâmetros.

A estrutura de dois espelhos tem, por construção, cinco graus de liberdade: ângulo entre espelhos  $\alpha$ , ângulo dos espelhos com o pavimento  $\gamma$ , ângulo da câmara com o pavimento  $\beta$ , distância entre câmara e espelhos  $D_{cm}$  e altura da (lente da) câmara em relação ao pavimento  $H$ . A lente da câmara constitui um grau de liberdade adicional, implicando a sua escolha a determinação do campo de visão da câmara: ângulo horizontal  $\theta$ , ângulo vertical  $\phi$ . (ganho controlado na placa de aquisição, e resolução das imagens são parâmetros extra também disponíveis).

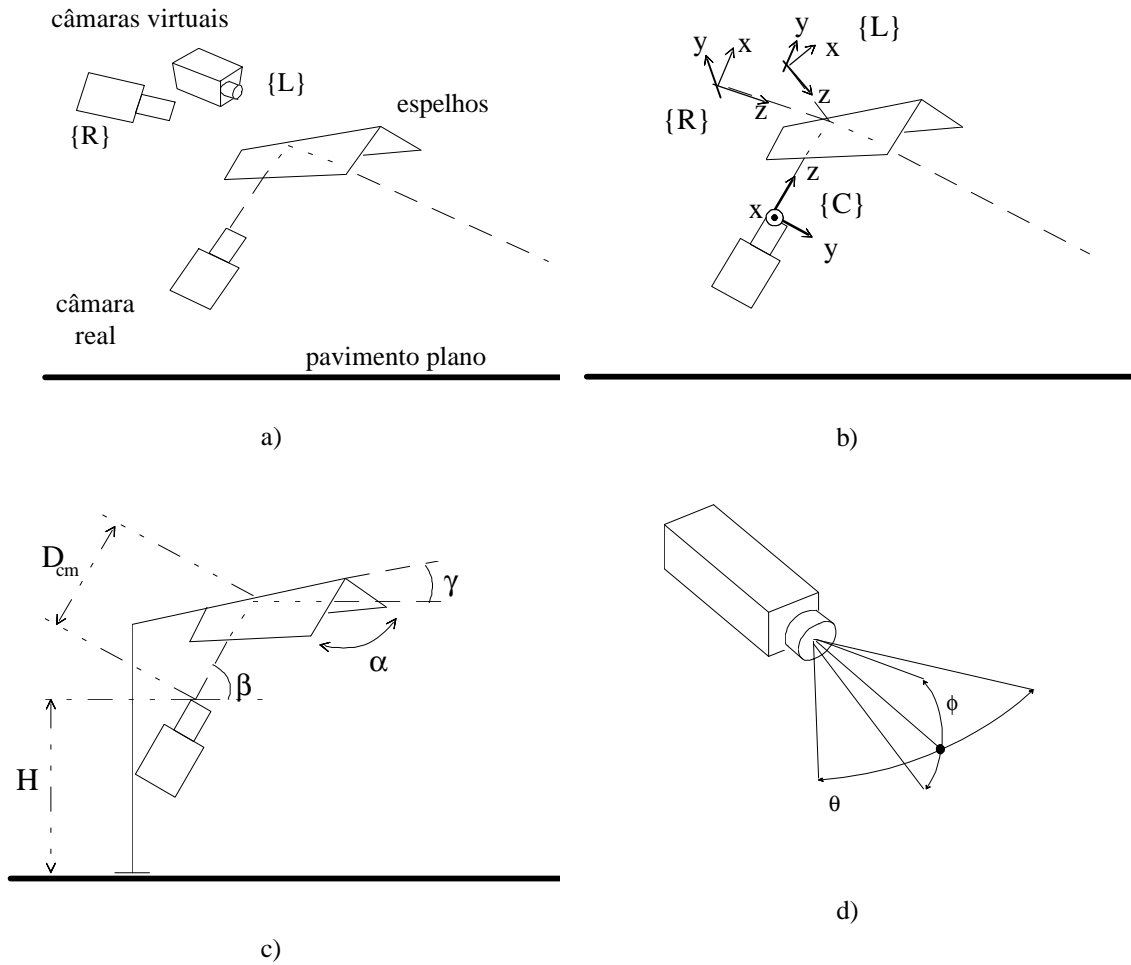


Fig. 4.2: Sistema de dois espelhos e uma câmara para aquisição de imagens stereo. a) Duas câmaras virtuais. b) Sistemas de coordenadas. c) Principais parâmetros. d) Campo de visão da câmara (lente).

A orientação dos dois espelhos relativamente à câmara real (dada pelos ângulos  $\alpha$ ,  $\beta$  e  $\gamma$ ) pode ser expressa de forma conveniente por dois vectores normais aos espelhos ( $n_r$  e  $n_l$ ):

$$\begin{aligned} n_r &= (n_{rx}, n_{ry}, n_{rz}) = (-\cos(\alpha/2), \text{sen}(\alpha/2).\cos(\gamma), -\text{sen}(\alpha/2).\text{sen}(\gamma)) \\ n_l &= (n_{lx}, n_{ly}, n_{lz}) = (\cos(\alpha/2), \text{sen}(\alpha/2).\cos(\gamma), -\text{sen}(\alpha/2).\text{sen}(\gamma)) \end{aligned} \quad (2)$$

A partir destas normais e de cálculos simples de reflexões de vectores e de intersecções de rectas com planos, é possível simular visualmente o campo visual do sistema de espelhos.

A cada uma das câmaras virtuais pode ser associado um sistema de coordenadas (fig.4.2.b). Entre os sistemas de coordenadas das câmaras virtuais e o sistema de

coordenadas da câmara real podem ser definidas transformações de coordenadas (usando eqs.(2)):

$$\mathbf{T}_R^C = \begin{bmatrix} R_R^C & P_R^C \\ \bar{0} & 1 \end{bmatrix} = \begin{bmatrix} r(n_r, \hat{x}) & r(n_r, \hat{y}) & r(n_r, \hat{z}) & D_{cm}(\hat{z} + r(n_r, \hat{z})) \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{T}_L^C = \begin{bmatrix} R_L^C & P_L^C \\ \bar{0} & 1 \end{bmatrix} = \begin{bmatrix} r(n_l, \hat{x}) & r(n_l, \hat{y}) & r(n_l, \hat{z}) & D_{cm}(\hat{z} + r(n_l, \hat{z})) \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

onde:

$\mathbf{T}_R^C$  = transformação (homogénea) de coordenadas  
da câmara virtual direita para o sistema de coordenadas  
associado à câmara real

$\mathbf{T}_L^C$  = transformação (homogénea) de coordenadas  
da câmara virtual esquerda para o sistema de coordenadas  
associado à câmara real

$\mathbf{T}_R^L$  = transformação (homogénea) de coordenadas  
entre as duas câmaras virtuais

Dadas as transformações entre câmaras virtuais e câmara real, pode agora ser obtida a transformação entre câmaras virtuais:

$$\begin{aligned} \mathbf{T}_R^L &= \begin{bmatrix} R_R^L & P_R^L \\ \bar{0} & 1 \end{bmatrix} \\ &= (\mathbf{T}_L^C)^{-1} \mathbf{T}_R^C \\ &= \begin{bmatrix} (R_L^C)^T & -(R_L^C)^T P_L^C \\ \hline \bar{0} & 1 \end{bmatrix} \mathbf{T}_R^C \\ &= \begin{bmatrix} (R_L^C)^T R_R^C & (R_L^C)^T (P_R^C - P_L^C) \\ \hline \bar{0} & 1 \end{bmatrix} \\ &= \text{função}(n_l, n_r, D_{cm}, \hat{x}, \hat{y}, \hat{z}) \end{aligned} \quad (3)$$

A calibração do sistema baseado em espelhos, ou seja a estimação da relação entre câmaras virtuais  $\mathbf{T}_R^C$ , e deste sistema em relação pavimento (e ainda a estimação de parâmetros das câmaras), seria importante para efeitos de aplicação / teste de algoritmos



de reconstrução. No entanto, a aplicação desejada segue uma filosofia qualitativa, não necessitando desta operação, tradicionalmente delicada.

#### 4.2.2. Espaço visível

A modelização e identificação dos principais parâmetros do sistema de dois espelhos, serve de base para a simulação do trajecto óptico das imagens, que por sua vez permite:

1) calcular a dimensão dos espelhos para lentes com distâncias focais maiores ou iguais a 4.8mm (o campo de visão de uma câmara varia inversamente com a distância focal da lente usada, pelo que escolhida a distância focal mínima, está definido o campo de visão máximo e por consequência as dimensões máximas dos espelhos).

2) avaliar o espaço visível pelo sistema, i.e., região observada simultaneamente em ambas as imagens do par stereo.

Na fig.4.3 mostram-se resultados desta simulação. Dois dos casos (fig.4.3.a e fig.4.3.c) mostram situações extremas nas quais o espaço visível é nulo porque a abertura (ângulo) dos espelhos é excessiva ou insuficiente. O caso ilustrado na figura 4.3.b é o que maximiza o espaço visível e portanto o que é desejado.

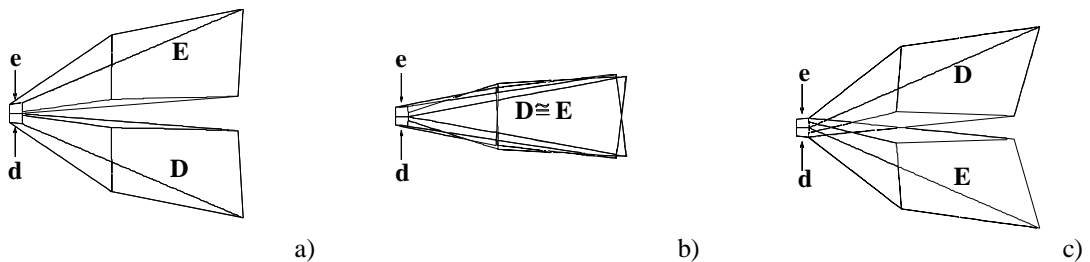


Fig. 4.3: Espaço visível =  $D \cap E$ , em função da abertura dos espelhos direito 'd' e esquerdo 'e' (resultados de simulação): a) ângulo, entre espelhos, excessivo b) ângulo, entre espelhos, que maximiza espaço visível c) ângulo, entre espelhos, insuficiente.

O espaço visível depende directamente da lente usada e corresponde sensivelmente a metade do campo de visão possível para esta lente. É no entanto possível alargar o espaço visível usando espelhos convexos em lugar de espelhos planos. Esta hipótese não foi desenvolvida nesta tese.

### 4.2.3. Resolução. Altura mínima detectada.

*Resolução* de um sistema de detecção de obstáculos sobre pavimento plano, define-se pela altura mínima (relativamente ao plano do pavimento) que um obstáculo deve ter, de modo a ser detectável.

A resolução de um sistema stereo genérico ([Bobet94]) depende essencialmente da distância entre câmaras ("baseline") e da distância focal da câmara que captura a imagem de referência. Mais concretamente, a resolução melhora (altura mínima diminui) com o aumento destes parâmetros.

Estas conclusões são importantes para o dimensionamento do sistema de aquisição (baseado em dois espelhos), proposto neste capítulo.

Como foi referido atrás, o sistema de dois espelhos e uma câmara, pode ser interpretado como um sistema stereo de duas câmaras virtuais.

Sendo a distância focal das câmaras virtuais a mesma da câmara real, um aumento da distância focal da (lente da) câmara real melhora a resolução do detector de obstáculos.

Uma interpretação qualitativa do efeito da distância focal sobre a resolução pode ser feita atendendo à relação inversa entre distância focal e campo de visão da câmara. A um aumento da distância focal corresponde uma diminuição da largura máxima de cenário observado. Mantendo-se constante o número de pixels por imagem, no caso em que o cenário observado é menor, existe uma melhoria da descrição do cenário. Esta melhoria de descrição permite valores de disparidades mais precisos, pelo que a discriminação de profundidade é melhor, o que beneficia a capacidade de detecção de obstáculos. Ou seja, o aumento da distância focal melhora a resolução.

Desta interpretação qualitativa conclui-se ainda que o número de pixels por imagem adquirida (resolução da imagem) é um factor muito relevante. Aumentando a resolução das imagens adquiridas, melhora-se a resolução do detector de obstáculos.

O comprimento da "baseline" no sistema de espelhos, corresponde ao módulo do vector que liga os sistemas de coordenadas das câmaras virtuais. Desenvolvendo a equação (3):

$$\text{comprimento da "baseline"} = |P_R^L| = |P_R^C - P_L^C| = 2 \cdot D_{cm} \cdot \text{sen}(\alpha) \cdot \cos(\gamma)$$

onde  $\alpha$  é o ângulo entre os espelhos e  $\gamma$  é o ângulo do conjunto de espelhos com o eixo  $y$  ("tilt"). Conclui-se desta expressão para o comprimento da "baseline", que a resolução depende ainda de  $D_{cm}$ ,  $\alpha$  e  $\gamma$ . Na prática  $\gamma$  e  $\alpha$  são determinados pela escolha e maximização do espaço visível. Nomeadamente  $\gamma$  é escolhido de modo a colocar a região de detecção de obstáculos próxima da plataforma móvel<sup>3</sup> e  $\alpha$  é escolhido de modo a maximizar o espaço visível, dependendo portanto da distância focal. Resta a distância entre a câmara e os espelhos  $D_{cm}$ , este sim um parâmetro livre para o ajuste da resolução: a resolução melhora com o aumento de  $D_{cm}$ .

Resumindo, a resolução do sistema de dois espelhos, proposto, melhora com o aumento da distância focal, com a resolução das imagens adquiridas e com a distância entre câmara e espelhos.

A optimização da resolução continua a ser um importante tópico de investigação porque determina a versatilidade do sistema de detecção visual de obstáculos (quanto mais baixos puderem ser os obstáculos detectáveis, mais plataformas podem usar este sistema). Este estudo é ainda importante para a optimização da estrutura, sendo eventualmente possível projectar uma pequena estrutura para os espelhos, compacta e robusta, a colocar à frente da lente da câmara e suportada por esta.

### 4.3. A placa de aquisição

O problema de aquisição de imagem, i.e., da captura (tipicamente precedida por digitalização) de uma imagem contida num sinal de vídeo, há muito que foi identificado. A sua solução através de placas de aquisição é útil para todos os que desejam tratar digitalmente uma imagem. Aplicações geográficas, militares, de visualização médica ou de simples criação de documentos contendo imagens, significam um mercado extenso motivador do aparecimento de várias placas de aquisição. Estas distinguem-se fundamentalmente pela resolução em número de cores ou níveis de cinzento por pixel e pelo número de pixels por imagem.

---

<sup>3</sup>O algoritmo de detecção de obstáculos destina-se a um sistema reactivo onde se espera que os obstáculos estejam de facto próximo da plataforma.

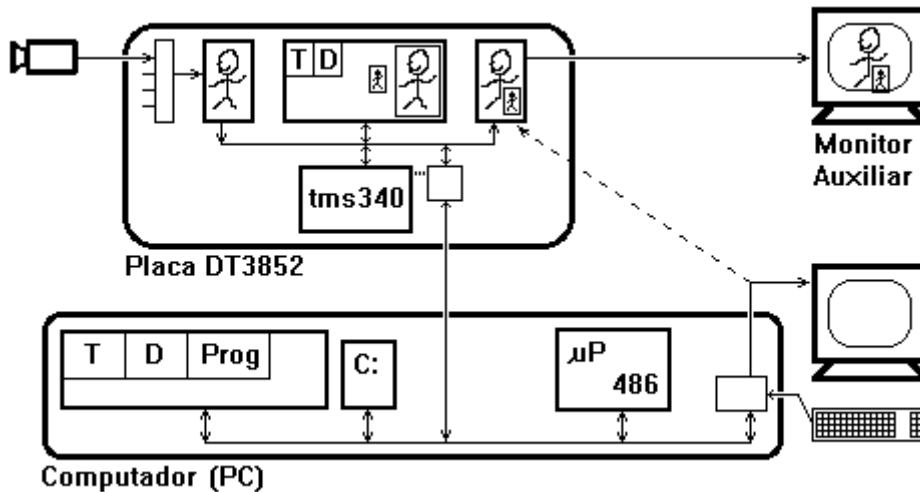


Fig. 4.4: Principais blocos da placa de aquisição de imagem DT3852.

A placa de aquisição de imagem usada neste trabalho é uma placa monocromática versátil da Data Translation®: a DT3852 (fig.4.4). Esta placa contém um conjunto típico de características tais como 'buffers' para imagem de entrada e de saída, e um multiplexer de entradas de vídeo. Contém ainda um conjunto específico de características de onde sobressai um microprocessador TMS340, uma memória interna (máx. 8Mb, útil para aquisição de seqüências de imagens e/ou instalação para execução de software) para além da possibilidade de programação das frequências de sincronismo dos sinais de entrada<sup>4</sup>, da frequência de digitalização e da resolução (é uma placa com multi-resolução na entrada e na saída).

A possibilidade de multi-resolução permite o ajuste da dimensão das imagens (número de pixels por imagem) à capacidade de processamento disponível. Imagens grandes implicam grandes tempos de transferência e de processamento. Imagens pequenas podem não ter informação suficiente. É importante a possibilidade de escolha de um compromisso.

Velocidades de aquisição:

---

<sup>4</sup>a capacidade de programação das frequências de sincronismo de vídeo, permite a aquisição de vários tipos de sinais de vídeo, por ex. RS-170/NTSC e CCIR/PAL.

1)	1 canal	24.64	imgs / seg
2)	2 canais não sincronizados	1.49	pares imgs / seg
3)	2 canais sincronizados	12.3	pares imgs / seg (aprox. 24 imgs/seg)
4)	2 canais sincronizados, aquisição + transferência para a memória do PC (resolução aprox. 200x200)	4.9	pares imgs adquiridas e transferidas / seg

Tabela com velocidades de aquisição da DT3852. Os números indicados foram obtidos contando o número de imagens adquiridas em 100 seg e incluem software interno de 'tracing'.

Da observação da tabela de velocidades de aquisição, conclui-se que a placa é limitada na aquisição de sinais não sincronizados (existe uma previsão interna dos impulsos de sincronismo que é útil para sinais com sincronismos frágeis / deficientes mas penaliza em excesso a aquisição de sinal de duas fontes distintas).

Uma outra observação importante refere-se ao tempo de transferência das imagens da memória da placa para a memória do PC:

- tempo de aquisição + transferência de um par de imagens é de cerca de 204 ms
- tempo de aquisição de um par de imagens é de cerca de 81ms
- tempo de transferência de um par de imagens (resolução aprox. 200x200) é de cerca de 123ms (aproximadamente 150% do tempo de aquisição).

Ou seja, é bastante significativo o tempo de transferência da memória da placa para a memória do PC (a transferência de imagens da placa de aquisição para a memória do PC é um estrangulamento típico que se prevê ser brevemente ultrapassado pelos novos standards de bus para PC).

#### 4.4. Resumo

Neste capítulo foi descrito o aparato experimental de aquisição de imagens stereo, baseado numa estrutura de dois espelhos e numa única câmara. Esta solução não só evita eventuais diferenças de parâmetros, usuais nas configurações de duas câmaras, como também pode conduzir a soluções compactas e robustas.

Foi dada atenção particular neste capítulo à relação entre espaço visível e resolução com parâmetros físicos do sistema. A compreensão desta relação é fundamental para o sucesso do sistema, em particular na etapa da construção / montagem.

Sobre o espaço visível máximo, conclui-se que corresponde sensivelmente a metade do campo de visão possível para a lente usada na câmara. Lentes de grandes angulares (baixas distâncias focais) permitem grandes espaços visíveis (mas implicam piores resoluções).

Sobre a resolução (altura mínima em relação ao pavimento que um obstáculo deve ter de modo a ser detectável), conclui-se que depende essencialmente da distância focal da lente usada, da resolução das imagens adquiridas e da distância entre câmara e espelhos. A altura mínima diminui com o aumento destes parâmetros.

Para além do espaço visível e da resolução, são também importantes a dimensão do sistema e a frequência de operação. O aumento da resolução das imagens permite detecções de obstáculos mais baixos, mas o tempo de processamento e de transferência de dados aumenta. O aumento do espaço visível (por exemplo com lentes de menor distância focal) implica um aumento da dimensão dos espelhos (e da estrutura em geral). São naturalmente necessárias soluções de compromisso.

Na fig. 4.5 mostra-se uma imagem do aparato experimental (conjunto câmara e dois espelhos) usado nesta tese.

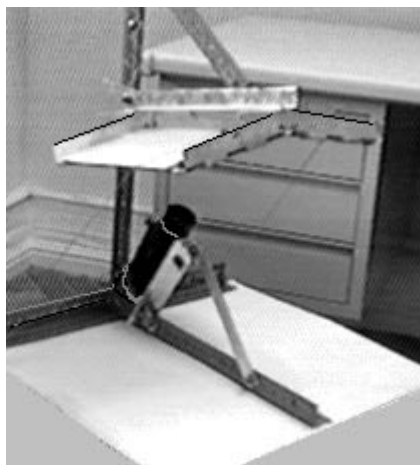


Fig. 4.5: Uma imagem do sistema experimental (conjunto câmara e dois espelhos).

## **CAPÍTULO 5**

### **Imagens Stereo vs Sequências de Imagem**

Ao longo desta tese tem por várias vezes sido referida a hipótese de desenho de um algoritmo de Detecção de Obstáculos sobre Pavimento Plano (ADOPP) baseado em sequências de imagens monoculares, mais explicitamente nos campos de Fluxo Óptico.

Existem vários métodos de detecção de obstáculos baseados em Fluxo Óptico. Como exemplo salientam-se as técnicas de inversão de perspectiva [Mallot91], o estudo da divergência do Fluxo Óptico [Nelson89] e a construção de um mapa de referência de Fluxo Óptico [Guarnotta].

A detecção de obstáculos baseada em sequências de imagens é interessante porque permite usar uma estrutura mecânica mais simples e menor quantidade de hardware (uma só placa de aquisição).

Neste capítulo será então apresentada uma alternativa dual ao algoritmo stereo descrito ao longo da tese. O algoritmo alternativo baseia-se no cálculo de Fluxo Óptico de uma sequência de imagens adquirida por uma única câmara em movimento ([Santos-Victor94]). Este algoritmo é analisado no plano teórico / formal discutindo-se a dualidade e a complementaridade desta alternativa face ao algoritmo implementado.

#### **5.1. Detecção de Obstáculos sobre Pavimento Plano baseada em Fluxo Óptico**

O Algoritmo de Detecção de Obstáculos sobre Pavimento Plano (ADOPP) baseado em imagens stereo, apresentado no capítulo 3, foi desenvolvido a partir de três hipóteses: câmaras pontuais, posição relativa entre câmaras constante e observação pelas câmaras de um pavimento plano.

A hipótese de pavimento plano é importante e central porque abre um caminho diferente da tradicional reconstrução 3D por pesquisas locais de disparidades

(normalmente um problema mal posto). Nomeadamente as disparidades passam a ser completamente previsíveis para o plano do pavimento, e os obstáculos podem ser simplesmente definidos como objectos texturados colocados acima ou abaixo do plano do pavimento.

A previsibilidade das disparidades para os pontos do plano do pavimento permite a criação de um mapa de referência, com uma forma analítica baseada em oito parâmetros, resultante da manipulação algébrica das hipóteses<sup>1</sup>.

A estimação dos oito parâmetros a partir de um conjunto de pontos correspondentes colhido de um par de imagens stereo, permite o cálculo do mapa de referência, concluindo-se desta forma a fase de inicialização. A fase de operação (detecção de obstáculos) resume-se à repetição consecutiva do teste do mapa de referência à procura de eventuais violações indicadoras de obstáculos.

O desenvolvimento do ADOPP baseado em fluxo óptico é semelhante ao desenvolvimento do ADOPP baseado em stereo. De igual modo, são postuladas três hipóteses (idênticas ao caso stereo) a partir das quais resulta o conceito e a forma analítica do mapa de referência<sup>2</sup>. Também este mapa de referência é baseado em oito parâmetros, estimáveis a partir de dados<sup>3</sup> colhidos em imagens.

Resumindo então as três hipóteses:

1) câmara é representável pelo modelo pontual

$$\begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{Z} \begin{bmatrix} f_x X \\ f_y Y \end{bmatrix} \quad (1)$$

onde  $f_x, f_y$  concentram a informação da distância focal da lente com as resoluções vertical e horizontal do CCD da câmara.

---

<sup>1</sup>Mapa de referência: mapa que antecipa e guarda as correspondências entre as imagens direita e esquerda para os diversos pontos do plano do pavimento.

<sup>2</sup>Neste caso o mapa de referência conterà vectores de fluxo óptico enquanto no caso stereo continha correspondências ou disparidades.

<sup>3</sup>Os dados serão naturalmente vectores de Fluxo Óptico (F.O.)



2) o cenário é estático e a câmara tem movimento de translação e rotação constante caracterizado por  $T$  e  $\omega$ ; supõe-se ainda que campo de movimento e fluxo óptico são sempre semelhantes<sup>4</sup>. Estes dados combinados com pressuposto 1 geram fórmulas conhecidas para o campo de movimento ( $u(x,y)$ ,  $v(x,y)$ ):

$$\begin{aligned}\vec{T} &= [T_x \quad T_y \quad T_z]^T \\ \vec{\omega} &= [\omega_x \quad \omega_y \quad \omega_z]^T \\ v(x,y) &= f_y \left[ \frac{\frac{y}{f_y} T_z - T_y}{Z(x,y)} + \omega_x \left( 1 + \frac{y^2}{f_y^2} \right) - \omega_y \frac{xy}{f_x f_y} - \omega_z \frac{y}{f_y} \right] \\ u(x,y) &= f_x \left[ \frac{\frac{x}{f_x} T_z - T_x}{Z(x,y)} + \omega_x \frac{xy}{f_x f_y} - \omega_y \left( 1 + \frac{x^2}{f_x^2} \right) + \omega_z \frac{y}{f_y} \right]\end{aligned}\quad (2)$$

3) a câmara normalmente observa um plano representado pelos parâmetros  $A, B, Z_0$ :

$$Z(X, Y) = Z_0 + AX + BY \quad (3)$$

Manipulando a expressão anterior, incluindo o modelo da câmara pontual, obtém-se uma nova expressão para o plano do pavimento mas expressa agora em coordenadas 2D do plano de imagem:

$$Z(x, y) = \frac{Z_0}{1 + A \frac{x}{f_x} + B \frac{y}{f_y}} \quad (4)$$

Juntando fórmulas do campo de movimento (2), obtém-se finalmente a forma do mapa de referência (mapa com vectores de fluxo óptico relativos a pontos do pavimento):

$$\begin{cases} v(x, y) = v_0 + v_x x + v_y y + v_{xy} xy + v_{yy} y^2 \\ u(x, y) = u_0 + u_x x + u_y y + u_{xy} xy + u_{xx} x^2 \\ u_{xy} = v_{yy} \quad , \quad u_{xx} = v_{xy} \end{cases}$$

(5)

---

<sup>4</sup>Na realidade Campo de Movimento e Fluxo Óptico são grandezas distintas: campo de movimento é constituído por projecções 2D sobre o plano de imagem de velocidades de pontos 3D, enquanto o fluxo óptico se refere a movimentos 2D de zonas de brilho na imagem. Podem ser realizadas experiências conceptuais que mostram esta diferença ([Horn 86]) mas na prática espera-se que o fluxo óptico seja uma razoável aproximação do campo de movimento.

Mostra esta expressão que os vectores de fluxo óptico para o plano do pavimento são previsíveis bastando para tal conhecer um conjunto de 8 parâmetros.

Definindo  $\theta$  como o vector de parâmetros:

$$\theta = [v_0 \ v_x \ v_y \ v_{xy} \ v_{yy} \ u_0 \ u_x \ u_y]^T \quad (6)$$

e sendo  $D$  o conjunto de dados (valores de fluxo óptico em dados  $x,y$ ) necessários para a estimação dos parâmetros:

$$D = \{(x_1, y_1, u_1, v_1), \dots, (x_N, y_N, u_N, v_N)\} \quad (7)$$

com estes  $N$  dados, as duas equações (4) são repetidas  $N$  vezes, pelo que se obtêm um sistema de  $2N$  equações a oito incógnitas. Este sistema pode ser facilmente representado em notação matricial:

$$M\theta = d \quad (8)$$

mais concretamente:

$$\begin{bmatrix} 1 & x_1 & y_1 & x_1 y_1 & y_1^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_1^2 & x_1 y_1 & 1 & x_1 & y_1 \\ 1 & x_2 & y_2 & x_2 y_2 & y_2^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_2^2 & x_2 y_2 & 1 & x_2 & y_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_i & y_i & x_i y_i & y_i^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_i^2 & x_i y_i & 1 & x_i & y_i \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_N & y_N & x_N y_N & y_N^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_N^2 & x_N y_N & 1 & x_N & y_N \end{bmatrix} \cdot \theta = \begin{bmatrix} v_1 \\ u_1 \\ v_2 \\ u_2 \\ \vdots \\ v_i \\ u_i \\ \vdots \\ v_N \\ u_N \end{bmatrix} \quad (9)$$

Este sistema tem solução desde que existam pelo menos oito equações linearmente independentes<sup>5</sup>. A solução usualmente adoptada para sistemas sobre-determinados é a pseudo-inversa (correspondendo a uma estimativa de mínimos quadrados):

$$\theta = [M^T M]^{-1} M^T d \quad (10)$$

---

<sup>5</sup>São necessários pelo menos 4 vectores de fluxo óptico. Na realidade serão usados bastantes mais de forma a melhorar a estimativa dos parâmetros.

Calculado o conjunto de parâmetros  $\theta$ , pode finalmente ser construído o mapa de referência (fica concluída a inicialização) e pode começar a operação normal.

Na operação normal é realizada continuamente a detecção de obstáculos. Obstáculos são simplesmente definidos como volumes com textura acima (ou abaixo) do plano de referência (pavimento). Estes volumes motivam fluxos ópticos diferentes, ponto a ponto, das referências guardadas no mapa de referência (motivam fluxos maiores se estiverem mais próximos que o plano de referência<sup>6</sup>, menores se estiverem mais distantes). Portanto a operação normal resume-se à comparação do campo de fluxo óptico actual com o campo de fluxo óptico guardado no mapa de referência.

Em imagens stereo existem regiões próximas dos cantos das imagens, que não são visíveis em ambas as imagens. Este facto é notado nos exemplos do ADOPP stereo. Estas regiões podem ser analisadas pelo ADOPP baseado em fluxo óptico. Esta nova informação pode ser útil em sistemas de visão activa por exemplo para reposicionar o ponto de fixação do stereo.

### **A razão das hipóteses**

- É necessário que a plataforma (suporte da câmara) tenha um movimento caracterizado por velocidades de translação e de rotação ( $T$  e  $\omega$ ) constantes, caso contrário não seria possível a geração de um Mapa de referência constante.

Por exemplo, um aumento da velocidade de translação da câmara  $T$ , em relação à etapa de inicialização, aumenta as leituras de fluxo óptico. Mantendo o mapa de referência, acontece que a textura do pavimento atinge níveis de fluxo óptico maiores que a referência pelo que o próprio pavimento passaria a ser classificado como obstáculo. Portanto, sempre que existe uma variação de  $T$ , é necessária uma adaptação do mapa de referência (neste caso, variação de  $T$ , a actualização consiste num simples escalamento).

---

<sup>6</sup>Mais especificamente, serão gerados fluxos maiores, para pontos do obstáculo, que estejam mais próximos da câmara, que os respectivos pontos ocluídos contidos no plano de referência.

Uma análise semelhante poderia ser feita para uma alteração de  $\omega$  (só que neste caso a actualização do mapa de referência já não é tão simples).

- Para a operação correcta do ADOPP baseado em fluxo óptico, é importante a verificação da hipótese de cenário estático porque sendo certo que alguns movimentos no cenário podem facilitar a detecção de obstáculos, outros há que a podem impedir.

No exemplo da figura seguinte, estão representados dois objectos móveis (A e B) que se deslocam com velocidades  $2v$  e  $v$ , nas alturas  $2h$  e  $h$  respectivamente, em relação ao pavimento. O objecto móvel 'A' transporta uma câmara e observa 'B' e o 'pavimento' com velocidades relativas coincidentes ( $V_{\text{pavimento}} = -f_y \cdot 2v / (2h) = -f_y \cdot v / h = V_B$ , considerando que o movimento se efectua segundo o eixo  $y$  da câmara). Esta coincidência de velocidades entre o objecto 'B' e o pavimento impossibilita a detecção do objecto móvel 'B' pelo ADOPP baseado em fluxo óptico.

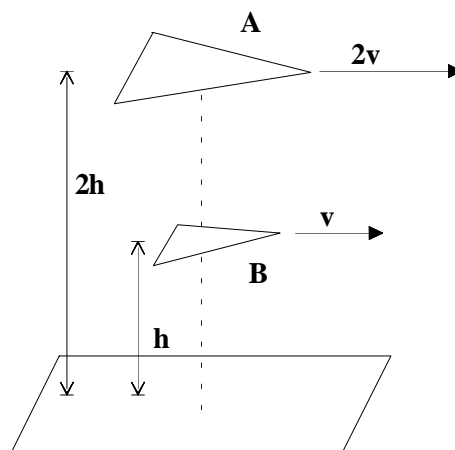


Fig. 5.1 Para o objecto móvel A, o objecto móvel B é indiscernível do pavimento, usando o ADOPP baseado em fluxo óptico.

### **Algumas das observações são semelhantes às realizadas para o ADOPP stereo:**

- A informação (detecção de pontos pertencentes a obstáculo(s)) resume-se às zonas com textura ou contornos (zonas onde existem variações de brilho).

- Há linhas não observáveis (exemplo: no caso do stereo imagine-se uma grande parede, que olhamos de frente, e da qual só vemos o rodapé; não conseguimos saber a que distância está a referida parede, se ignoramos a informação de "shape-from-blurr").

No caso stereo, a detecção falha para os contornos pertencentes ao obstáculo, cuja direcção coincide ponto a ponto com a direcção da disparidade de referência. No caso do fluxo óptico, a detecção falha para os contornos pertencentes ao obstáculo cuja direcção coincide ponto a ponto com a direcção dos fluxos ópticos de referência.

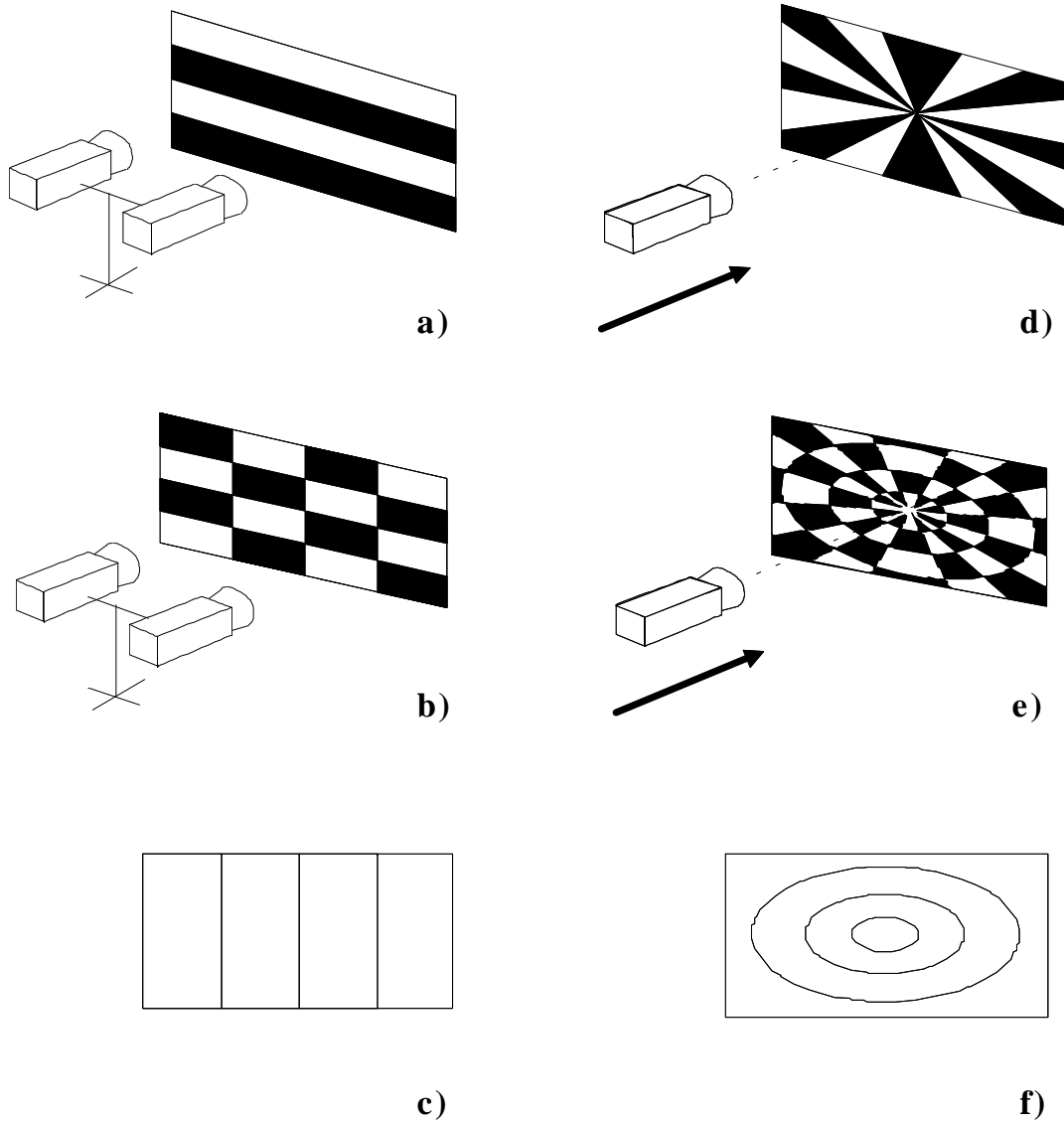


Fig. 5.2 Direcções não detectáveis: com a geometria stereo da imagem 'a', com o movimento da câmara na imagem 'd' e dados os cenários indicados, especialmente escolhidos, em ambos os casos não existe detecção do obstáculo (a parede com as texturas indicadas nas imagens). Nas imagens 'b' e 'e', a textura da parede já não é tão desfavorável e os resultados previsíveis de detecção mostram-se nas imagens 'c' e 'f'.

O exemplo escolhido para a demonstração desta observação (fig. 5.2) aproveita casos para os quais são fáceis de prever as disparidades e os fluxos ópticos. No caso 5.2.a (câmaras alinhadas) as restrições epipolares são horizontais bem como as disparidades<sup>7</sup>. Também os contornos observados são horizontais, pelo que existe uma possibilidade de "escorregamento horizontal do teste de disparidade" (em particular a disparidade de referência encontra sempre um padrão semelhante<sup>8</sup>). No caso de fluxo óptico (5.2.d), o FOE (capítulo 2) coincide com o centro da figura e os fluxos ópticos (distribuídos radialmente do centro para o exterior) coincidem em direcção com os contornos da imagem observada, pelo que acontecerá um problema semelhante de "escorregamento" ao descrito no caso do stereo.

Este problema é solucionado em stereo, usando uma terceira câmara ou alterando momentaneamente a geometria. De forma semelhante em sequências temporais de imagens, pode ser realizada uma mudança momentânea do movimento de translação da câmara.

- O resultado dos ADOPP depende da altura dos obstáculos: no stereo a largura dos traços de detecção aumenta com a altura dos obstáculos; no fluxo óptico aumenta a diferença entre a referência e valor de teste quando os obstáculos são mais altos.

### **Relaxação das hipóteses:**

Tem vindo a ser realizado um trabalho de investigação bastante intenso no sentido de relaxar algumas das hipóteses admitidas.

- Cenário não estático: uma pré-deteção das zonas em movimento é possível (basta parar durante um momento o movimento da câmara e observar os movimentos na imagem). Dadas as zonas em movimento podem ser usados algoritmos de seguimento.

---

<sup>7</sup>Esta coincidência de direcção entre restrição epipolar e disparidade está relacionada com a escolha da relação entre câmaras (os seus sistemas de coordenadas diferem simplesmente por uma translação).

<sup>8</sup>Escolhido um ponto  $x,y$  na imagem direita e procurando o seu correspondente na imagem esquerda, existem muitos candidatos possíveis; um destes candidatos de acordo com o ADOPP será testado e o teste não gerará suspeita de obstáculo.

Separadas as zonas em movimento, pode continuar a aplicação de ADOPP nas restantes regiões.

- $T$  e  $\omega$  variáveis: é um problema equivalente ao caso da geometria variável do stereo. A proposta de inversão da perspectiva (criação de uma câmara virtual, de modo a obter um plano de imagem paralelo ao pavimento) contorna este problema desenhando um algoritmo de detecção cujo mapa de referência é constante para todos os  $x,y$  da imagem, passando os obstáculos a ser detectados por ex. por algoritmos de detecção de contornos sobre resultados de comparação entre referências e campos de fluxo óptico ([Mallot91], [Zielke90]).

## **5.2. Complementaridade**

Explorada a dualidade da informação stereo e das sequências de imagem no desenho de algoritmos ADOPP, não pode deixar de ser referida a complementaridade da informação base das medidas de disparidade e fluxo óptico.

Disparidades lidas em imagens stereo permitem a obtenção de profundidades (posições 3D de pontos) dados os parâmetros intrínsecos e a relação entre posições das câmaras. Por sua vez o fluxo óptico contém informação única de movimento. A junção de informação stereo com informação temporal, permitirá, a determinação de posição, de velocidade ou de aceleração em valores 3D. No fluxograma da figura 5.3 é dado um exemplo de utilização do ADOPP stereo, complementado com informação de fluxo óptico. Neste caso é possível não só detectar obstáculos, como também distinguir entre obstáculos fixos e móveis.

A utilização simultânea, de algoritmos de detecção de obstáculos sobre pavimento plano, baseados em imagens stereo e em fluxo óptico, pode ser útil, por exemplo, para a análise de regiões da imagem não visíveis em stereo (regiões do pavimento visíveis por uma só câmara, devido às dimensões finitas das imagens), para a detecção de obstáculos em movimento (fig.5.1) e para a redução das direcções de textura não detectáveis (fig.5.2).

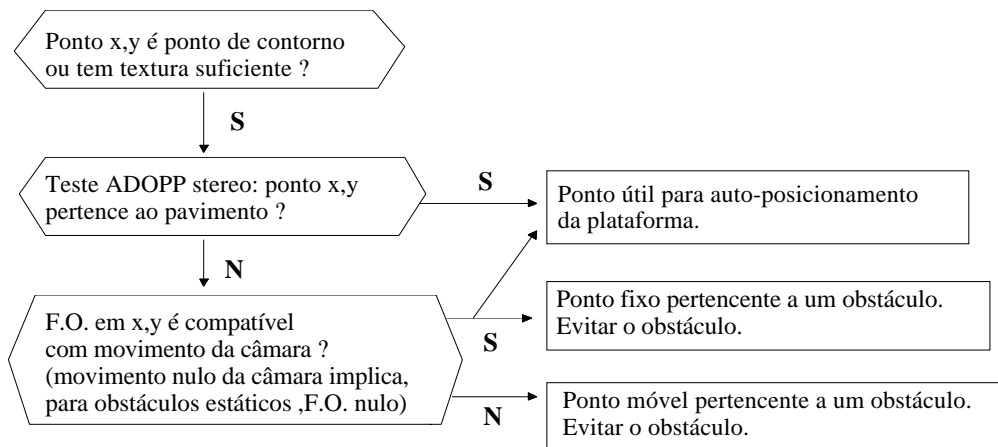


Fig. 5.3 Exemplo de um algoritmo que tira proveito da complementaridade da informação stereo vs seqüência temporal. Supõe-se que este algoritmo está a ser usado por uma plataforma móvel equipada com duas câmaras.

### 5.3. Resumo

Foi neste capítulo apresentado um Algoritmo de Detecção de Obstáculos sobre Pavimento Plano (ADOPP) baseado em seqüências de imagem, dual e alternativo do algoritmo baseado em imagens stereo que tem vindo a ser explorado ao longo desta tese.

Detecção de obstáculos baseada em seqüências de imagem, é uma abordagem atractiva porque permite a mais simples estrutura mecânica e a menor quantidade de hardware (uma só placa de aquisição). No entanto, comparando com o algoritmo baseado em imagens stereo, existem mais cuidados a tomar na realização de testes, nomeadamente no controlo da trajetória do veículo (o 'caso stereo' pode ser testado sem movimento) e na medida do fluxo óptico.

Para além da dualidade é particularmente interessante relembrar a complementaridade da informação temporal vs stereo. Existe informação que só pode ser obtida analisando imagens no tempo, como por ex. a distinção entre obstáculos (objectos) estáticos ou móveis. Por outro lado a recuperação de grandezas 3D como a posição, a velocidade ou a aceleração de pontos ou corpos no espaço, sem necessidade de hipóteses sobre o cenário, só é possível juntando informação de imagens stereo. Trabalhos



interessantes têm vindo a ser realizados sob esta perspectiva de integração e atendendo à franca evolução na análise individual dos dois tipos de dados (imagens stereo e sequências de imagem), prevêem-se e esperam-se trabalhos integrados cada vez mais importantes.



## CAPÍTULO 6

### Resultados Experimentais

Neste capítulo são documentados vários exemplos e apresentados resultados obtidos com o sistema de detecção visual de obstáculos que tem vindo a ser apresentado ao longo desta tese.

Por cada par de imagens stereo obtido pelo sistema de espelhos (descrito no capítulo 4), é gerada uma imagem resultado, seguindo o algoritmo de detecção de obstáculos proposto no capítulo 3. Cada pixel da imagem resultado tem um de três níveis de brilho possíveis: "preto", "branco" ou "cinzento". Estes brilhos traduzem *detecção*, *não detecção* ou *indeterminação de existência de obstáculo* para cada pixel da imagem. Nomeadamente o "preto" indica pixel pertencente ao obstáculo, o "branco" indica pixel não pertencente a obstáculo ou pertencente a uma zona de obstáculo com textura insuficiente e finalmente o "cinzento" indica pixel sobre o qual nada se pode afirmar porque se situa numa zona da imagem direita sem correspondente visível na imagem esquerda (considerando um dado plano de referência).



Violações do mapa de referência: pontos do obstáculo.



Zona da imagem direita não observável na imagem esquerda (zona desprezada pelo ADOPP implementado).



Não existe obstáculo ou a informação de brilho é insuficiente (não existem variações significativas ou textura) para a sua detecção.

Legenda dos níveis de brilho possíveis nas imagens resultantes do "algoritmo de detecção de obstáculos sobre Pavimento plano".

No primeiro exemplo ambas as imagens mostram exclusivamente pontos pertencentes ao plano do pavimento. Tendo sido este o plano escolhido para referência, naturalmente nenhum obstáculo foi sinalizado (fig. 6.1).

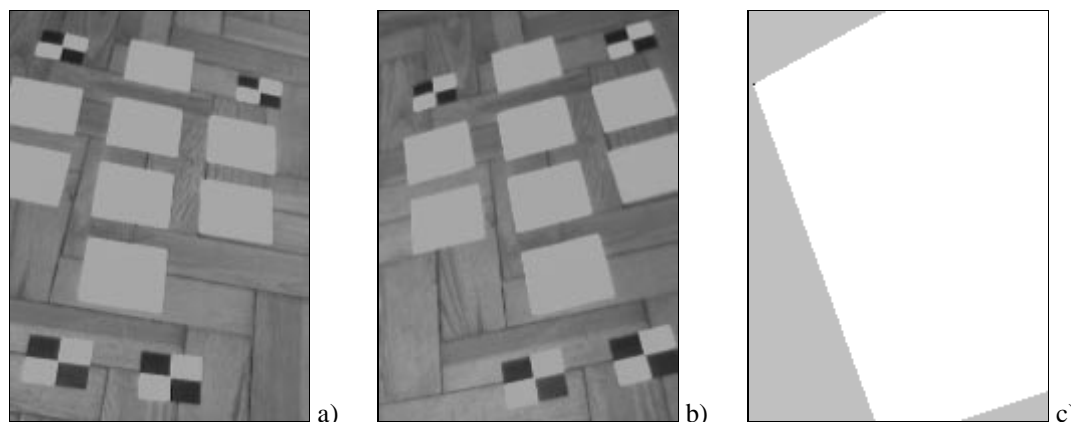


Fig. 6.1 Imagens 'a','b' constituem o par stereo do plano do pavimento e a imagem 'c' é o resultado da detecção de obstáculos.

Este é um teste importante na avaliação da qualidade da parametrização. É realizado várias vezes durante a fase de inicialização, quando os parâmetros ainda estão a ser ajustados. Observou-se em particular, (na etapa de estimação de parâmetros), ser importante um número significativo de pontos correspondentes relativos ao plano do pavimento<sup>1</sup> e serem úteis técnicas de refinamento destes pontos.

Observa-se ainda a região "cinzenta" típica para todos os testes realizados. Esta região corresponde a pontos da imagem direita, por hipótese projecções do plano do pavimento, para os quais não existem correspondentes dentro dos limites da imagem esquerda.

No segundo exemplo foram colocados alguns obstáculos sobre o pavimento. A detecção de obstáculos revelou-se essencialmente sobre os contornos da imagem, correspondentes a arestas / fronteiras dos objectos ou a variações fortes de brilho dos mesmos. As imagens adquiridas e os resultados encontram-se na figura 6.2.

---

<sup>1</sup>Nos testes realizados foram obtidos, após emparelhamento de pontos de canto das imagens esquerda e direita, entre 25 e 35 pontos correspondentes.

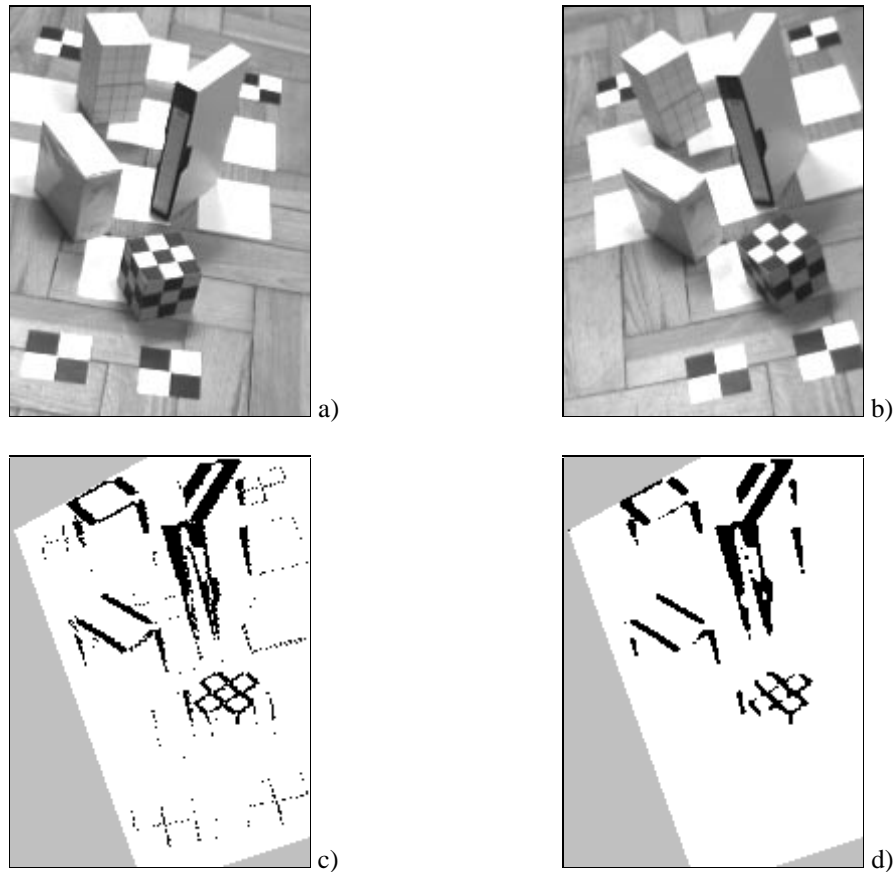


Fig. 6.2 Vários obstáculos e efeito da filtragem morfológica: 'a' e 'b' constituem o par stereo original; as imagens 'c' e 'd' mostram o resultado da detecção de obstáculos sem e com filtragem morfológica.

Variações fortes de brilho contidas no plano do pavimento (por exemplo as variações "preto - branco" dos pequenos referenciais ainda colocados nas suas posições originais) motivaram pontos isolados "fortes". Estes pontos, devidos essencialmente à "digitalização" do mapa de referência, foram eliminados com a filtragem morfológica descrita no capítulo 3. Excepto os exemplos onde é indicado o contrário, todos os resultados mostrados neste capítulo contêm uma filtragem morfológica.

O exemplo seguinte mostra que existem infinitas formas de obstáculos detectáveis por este algoritmo. A detecção de formas humanas, tal como a mão, são importantes porque realçam o factor de protecção humana na interacção com sistemas robóticos.

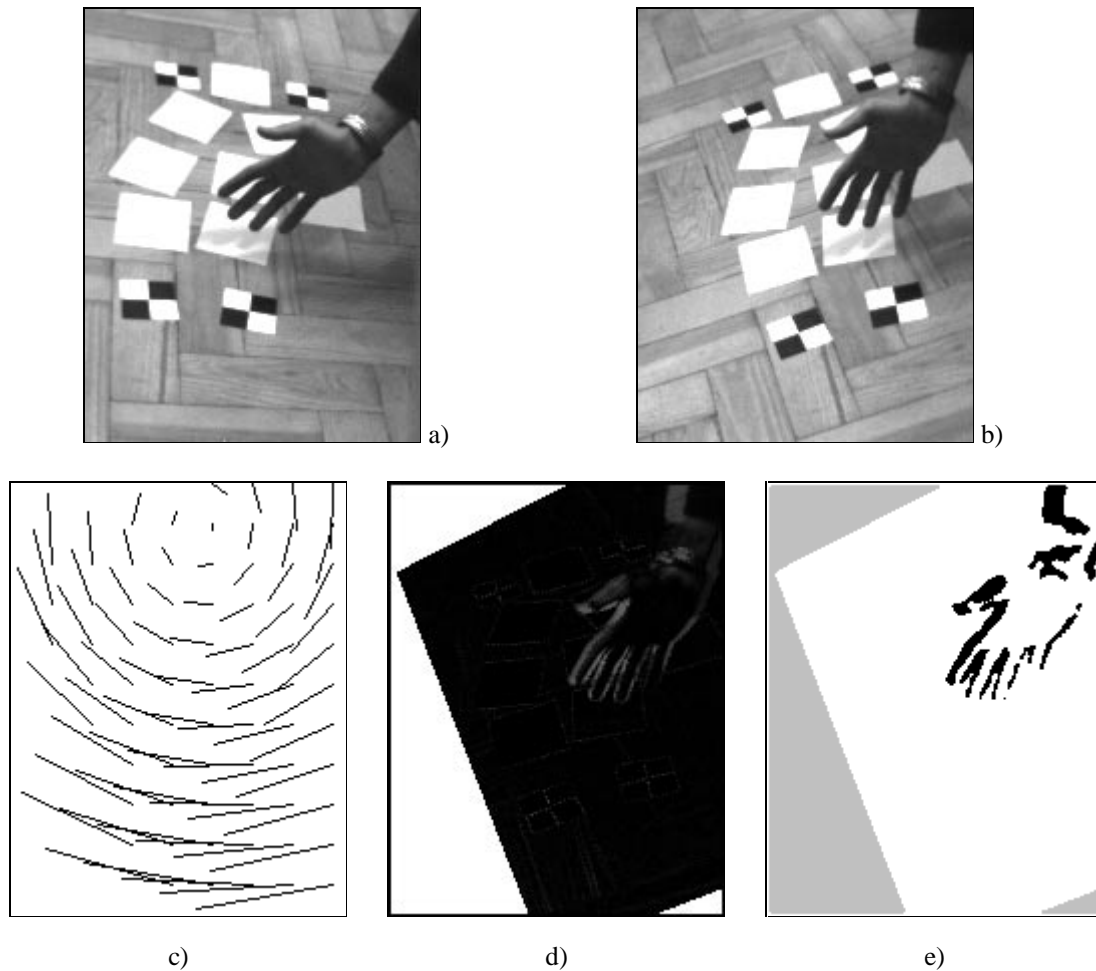


Fig. 6.3 Detecção do obstáculo "mão"; Mapa de referência. Imagens 'a','b' constituem o par stereo, a imagem 'c' é uma visualização do mapa de referência como mapa de disparidades, a imagem 'd' é o resultado simples de SMD e a imagem 'e' é o resultado final diferindo de 'd' porque contém filtragem morfológica e os níveis de brilho escolhidos para representação foram invertidos

O mapa de referência aqui apresentado e visualizado como um mapa de disparidades (figura 6.3.c) é um exemplo típico. Para todos os exemplos existe, tal como foi referido na descrição do Algoritmo de Detecção de Obstáculos sobre Pavimento Plano (ADOPP: capítulo 3), um mapa de referência, i.e. uma tabela de correspondências entre as imagens direita e esquerda para os diversos pontos do plano escolhido para referência. Os mapas de referência são semelhantes entre os vários exemplos e serão mesmo iguais entre os exemplos obtidos para a mesma configuração da estrutura física e do 'hardware'.

A imagem 'd' é um resultado intermédio tipo. Este resultado consiste numa filtragem de média de módulos de diferenças, entre imagem direita e imagem esquerda (mapeada

sobre imagem direita)<sup>2</sup>. A imagem 'e' é obtida a partir da 'd' após filtragem morfológica e transformação dos níveis de brilho para análise de acordo com a legenda apresentada no início deste capítulo.

O quarto exemplo ilustra a sensibilidade do método utilizado, relativamente à altura mínima dos objectos detectados. A figura 6.4 mostra as imagens adquiridas sendo o obstáculo constituído por uma mão.

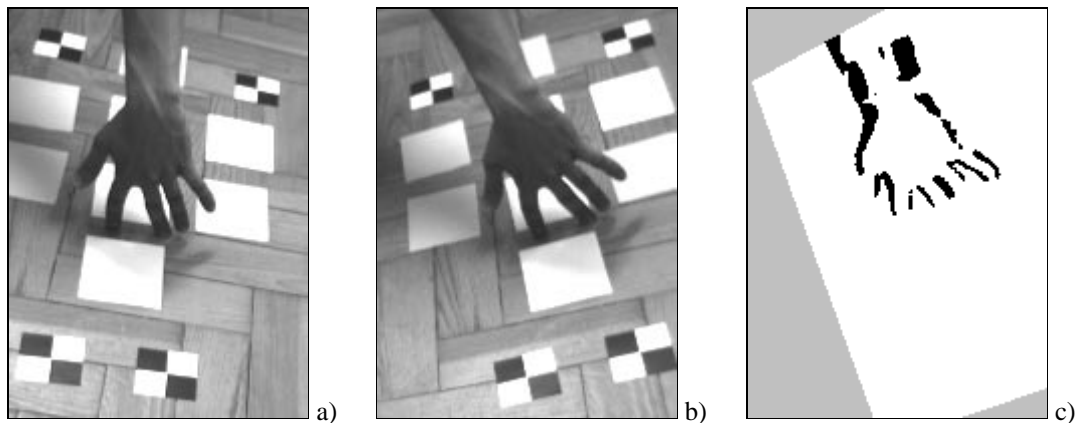


Fig. 6.4 Altura mínima detectada: obstáculo constituído por uma mão (os dedos tocam no pavimento)..

É de notar que a altura mínima detectável corresponde a cerca de 5 cm (metade inferior dos dedos da mão).

Seguidamente ilustra-se a detecção de pontos abaixo do plano do pavimento, ou seja buracos a contornar pelo robot móvel.

Foram realizadas duas parametrizações para este exemplo. A primeira usa para referência o plano do pavimento. A segunda usa para referência o plano que contém os referenciais e que está, aproximadamente, 10cm acima do plano do pavimento. Mostra este exemplo que o algoritmo também detecta obstáculos abaixo do plano de referência, porque marca fronteiras dos "objectos desenhados no pavimento". As imagens são apresentadas na figura 6.5.

---

<sup>2</sup>Na nomenclatura inglesa: "Sum of the Modules of the Differences" ("SMD"). Nota-se o uso de módulos no lugar de quadrados, por razões de eficiência computacional (i.e., foram usadas SMD em lugar das tradicionais SSD, "Sum of the Squared Differences")

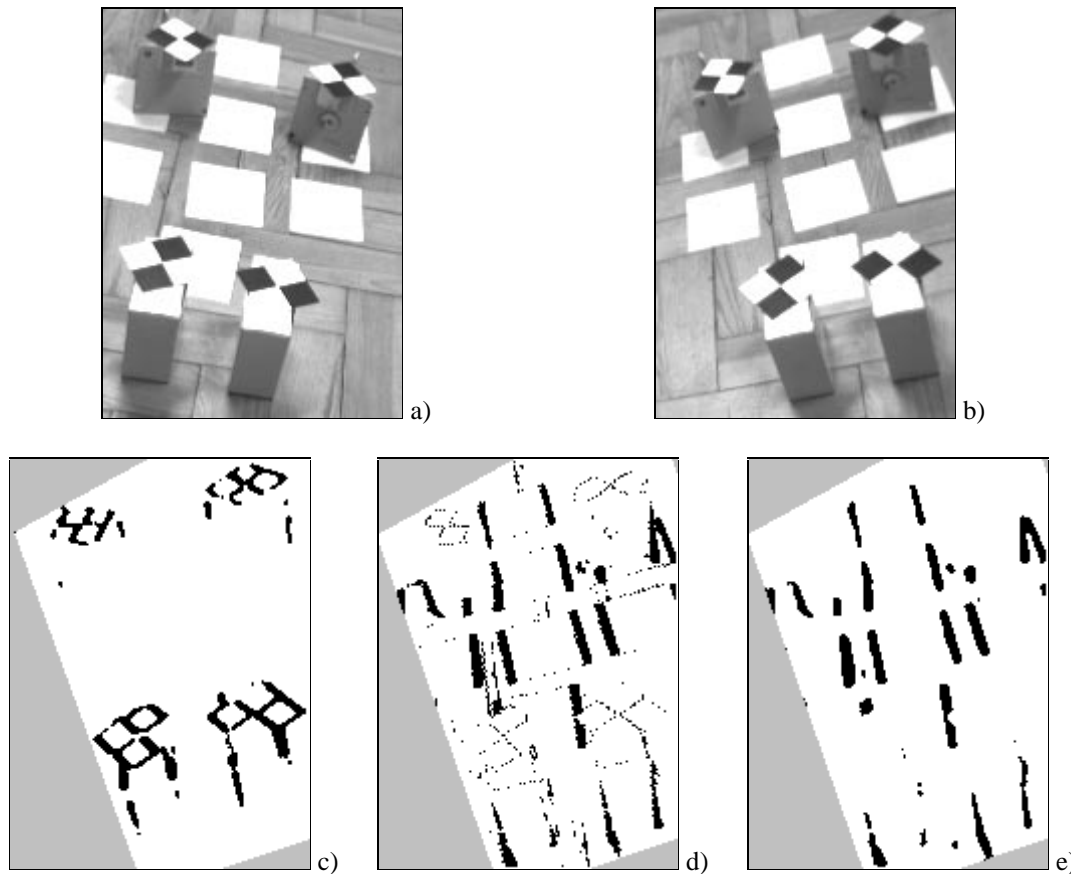


Fig. 6.5 Dois planos: detecção não só de obstáculos acima do pavimento mas também de buracos.

A imagem 'c', mostra detecção tendo sido parametrizado o plano do pavimento. As imagens 'd','e' mostram detecção tendo sido parametrizado o plano que contém os referenciais (a imagem 'd' difere da 'e' porque lhe foi anulada a filtragem morfológica para mais fácil análise de resultados por comparação com imagem direita).

No teste que se segue, ilustra-se que existem direcções para os contornos dos obstáculos que inviabilizam a sua detecção como obstáculos. Para tal testou-se um obstáculo constituído por uma folha de papel orientada em várias direcções (fig.6.6).



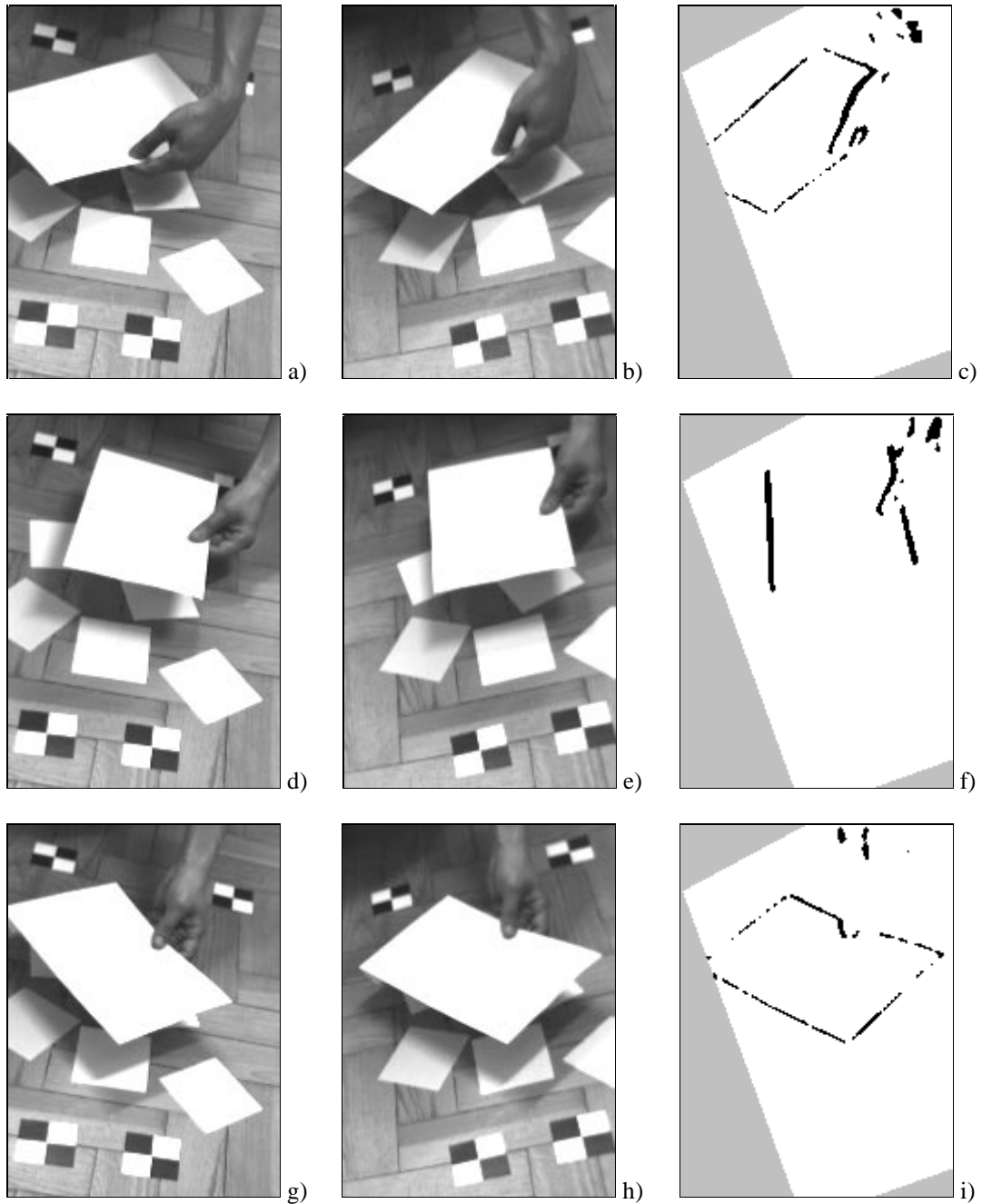


Fig. 6.6 Testes com a folha de papel: certas direcções de textura em obstáculos não são detectáveis. .

Observa-se que nem todos os contornos dos obstáculos são detectáveis. Ponto a ponto existe uma direcção (a direcção da restrição epipolar) que não pode ser detectada porque existe um escorregamento local nas comparações.

O exemplo final, ilustra que quanto maior for a distância do objecto ao plano do pavimento maior é a região de violação detectada (ver fig.6.7).

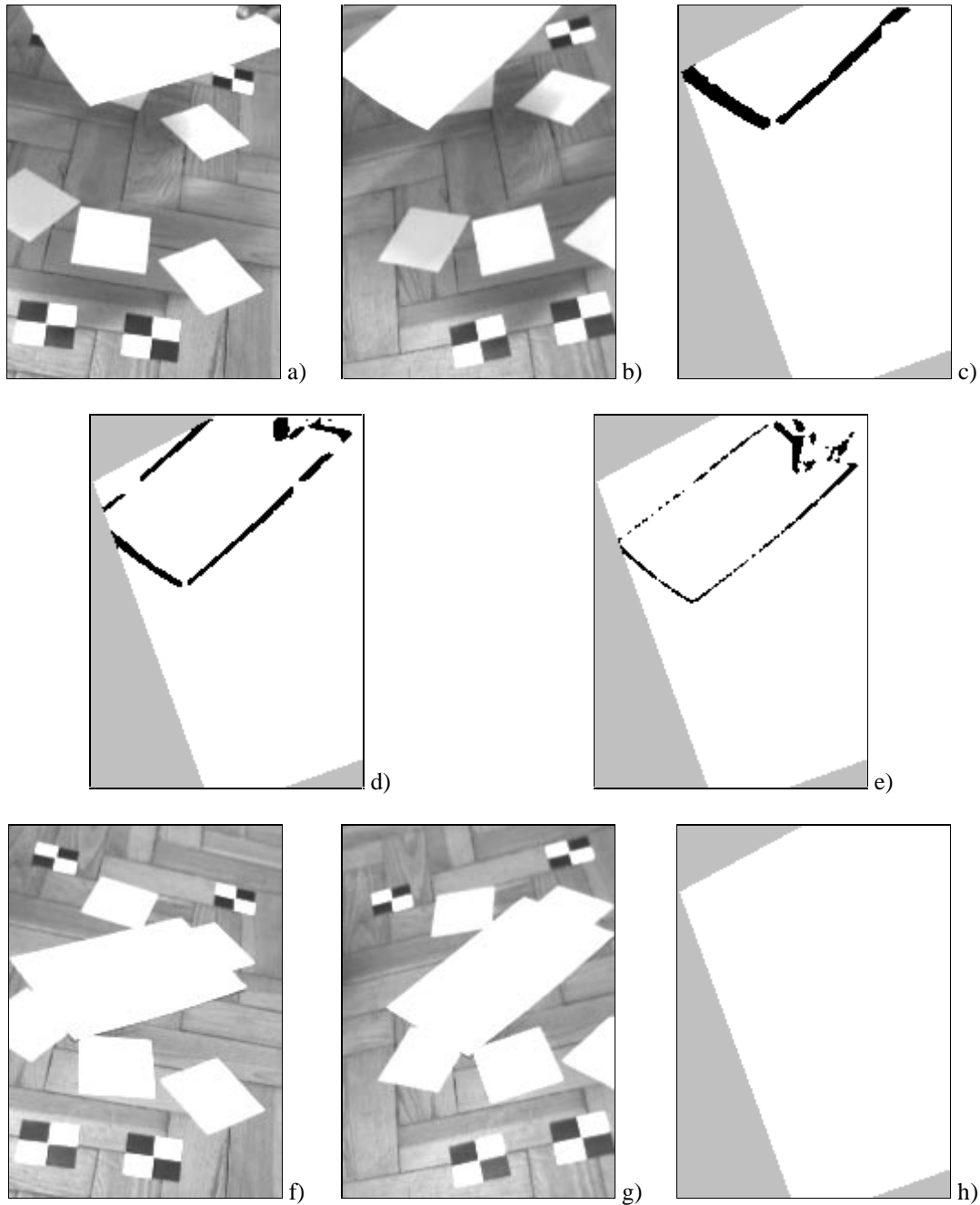


Fig. 6.7 Testes com a folha de papel: variação da "largura de detecção" com a altura do obstáculo. A largura das zonas de detecção aumenta com o aumento da altura dos obstáculos (mais precisamente, com a diminuição da distância em relação à câmara).

## 6.1. Tempos de cálculo

Nas experiências realizadas com a folha de papel, as imagens têm uma resolução de 149x204 (colunas x linhas) pixels e o tempo de cálculo de cada teste foi de aproximadamente 1.07seg, ou seja, foi atingida uma frequência de operação de aproximadamente 0.9 Hz (esta medida foi obtida ao fim de 428 testes que demoraram 455 seg). Este tempo contém aquisição de imagem, detecção de obstáculos e 'display' de resultados.

Nas primeiras experiências de detecção de uma "mão", as imagens têm dimensão 167x216 e a frequência de operação foi de aproximadamente 0.6 Hz.

Ambos os resultados foram obtidos com um computador PC-80486/50MHz.

A configuração (medidas) do sistema de espelhos, usada na realização dos exemplos descritos neste capítulo, está descrita em apêndice.

## 6.2. Resumo

Mostrou-se neste capítulo um conjunto de resultados obtidos com o sistema de detecção de obstáculos sobre pavimento plano, que tem vindo a ser descrito ao longo desta tese.

Durante a fase de inicialização (etapa de estimação de parâmetros), observou-se ser importante definir um número significativo de pontos correspondentes relativos ao plano do pavimento (foram usados entre 25 e 35; o número mínimo como se viu no capítulo 3 é de quatro pontos correspondentes) e serem úteis técnicas de refinamento destes pontos.

A partir dos resultados ilustrados, verifica-se que a detecção de obstáculos se concentra sobretudo sobre os seus contornos (tipicamente arestas ou zonas da superfície com variações fortes de brilho).

Nota-se ainda que a altura do obstáculo é relevante para a largura dos "traços de detecção": quanto maior a altura, maior a largura de detecção.

A simplicidade da definição de obstáculo, obtida com a hipótese de movimento sobre um pavimento plano, permite que o algoritmo de detecção seja rápido e eficiente mantendo contudo uma vasta gama de aplicações possíveis. Particularmente gratificante é

a possibilidade de detecção de obstáculos com uma infinidade de formas. Esta característica foi tornada real porque o tradicional objectivo de reconhecimento foi moderado e redefinido como simples detecção de zona(s) de obstáculo(s)<sup>3</sup>.

---

<sup>3</sup>Não existe a tentativa de segmentação de obstáculos, nem se tenta impedir uma sub-divisão dos mesmos. As regiões detectadas como pertencentes a obstáculos, são simples conjuntos de pixels, correspondentes a um ou mais obstáculos.

# CAPÍTULO 7

## Conclusões

Nesta tese foi apresentado um sistema de detecção visual de obstáculos, para robots móveis. O sistema pressupõe que o robot se move sobre um pavimento plano, sendo então possível definir um **obstáculo** como todo e qualquer ponto fora do plano do pavimento<sup>1</sup>.

A filosofia subjacente ao trabalho proposto na tese enquadra-se na visão por objectivos. Abdicando da generalidade duma abordagem reconstrutiva, foi possível definir um método de percepção visual com o propósito específico de detectar os obstáculos de frente a um veículo móvel que se desloca sobre um pavimento plano. Como contrapartida, ganha-se ainda velocidade de cálculo e robustez.

O algoritmo de detecção de obstáculos sobre pavimento plano, baseia-se em pares de imagens stereo. A hipótese de pavimento plano, permite ainda a criação de um mapa de referência que descreve as disparidades (ou correspondências) para pixels que sejam projecções de pontos do pavimento. Este mapa é calculado na fase da inicialização, de forma paramétrica, a partir de um conjunto de pontos correspondentes relativos ao plano do pavimento.

Na fase de detecção, o algoritmo limita-se a verificar para todos os pixels da imagem direita, se a disparidade (correspondência) esperada, salvaguardada no mapa de referência, é observada (o pixel em análise é a projecção de um ponto do pavimento) ou se é violada (o pixel em análise é a projecção de um ponto do obstáculo).

O facto de ser testado, por cada pixel da imagem, um valor de disparidade (correspondência) guardado no mapa de referência, não sendo necessário o moroso

---

<sup>1</sup>É necessária a existência de textura nas regiões dos obstáculos.

processo de procura típico em sistemas de visão reconstrutiva (emparelhamento), permite que este seja um algoritmo rápido.

A aquisição de imagens stereo foi realizada tomando por base um sistema de dois espelhos e uma câmara, evitando-se desta forma eventuais diferenças de parâmetros, usuais nas configurações de duas câmaras.

Apresentou-se também um algoritmo de detecção de obstáculos alternativo, baseado no cálculo do fluxo óptico de uma sequência de imagens adquirida por uma câmara em movimento. De igual modo pressupôs-se que o veículo se move sobre um pavimento plano e foi definido obstáculo como qualquer ponto fora do plano do pavimento. De novo se obteve uma expressão paramétrica para o mapa de referência (contendo valores de fluxo óptico relativos a pontos do pavimento), a calcular na fase de inicialização. Dado o mapa de referência, limita-se também a fase de detecção a verificar para todos os pixels da imagem se os valores de fluxo óptico de referência se verificam (o pixel em análise é a projecção de um ponto do pavimento) ou são violados (o pixel em análise é a projecção de um ponto do obstáculo).

Este algoritmo foi analisado no plano formal, discutindo-se a dualidade e a complementaridade da informação obtida face ao algoritmo stereo implementado.

Os resultados obtidos durante o teste do sistema de detecção de obstáculos baseado em imagens stereo, mostraram ser possível a estimação paramétrica do mapa de disparidades (correspondências) de referência, garantindo-se desta forma uma consistência global para este mapa. Verificou-se em particular ser importante, para a estimação de parâmetros, o número significativo de pontos correspondentes relativos ao plano do pavimento obtido pelo emparelhamento de pontos de canto, e serem úteis as técnicas usadas para o refinamento destes pontos.

Os resultados obtidos demonstram que o algoritmo de detecção de obstáculos proposto é rápido e eficiente. Os obstáculos são detectados essencialmente sobre as suas fronteiras (perfis) e sobre as regiões com textura.

## 7.1. Trabalho Futuro

Ao longo da tese foram sendo indicadas várias possíveis direcções de desenvolvimento e evolução para o sistema de detecção visual de obstáculos proposto. Pretende-se nesta secção resumir e propor como trabalho futuro algumas destas direcções.

A continuação do estudo do sistema de espelhos e uma câmara, para a aquisição de imagens stereo, em particular nos aspectos de espaço visível e resolução, pode eventualmente tornar possível a construção de uma pequena estrutura para os espelhos, compacta e robusta, a colocar à frente da lente da câmara e suportada por esta.

Os resultados apresentados no capítulo 6, foram obtidos com o sistema numa posição estática. Prevê-se naturalmente a utilização do sistema de detecção de obstáculos, num robot móvel, para a implementação de um sistema reactivo de desvio de obstáculos.

O aparecimento nos laboratórios de visão de "cabeças stereo" motiva a extensão do estudo do sistema de detecção de obstáculos para o caso de geometrias stereo variáveis. Os parâmetros do mapa de referência, estimados na fase de inicialização, variam no tempo. A estimação de parâmetros para um conjunto finito de geometrias e consecutiva interpolação para geometrias intermédias é uma evolução possível. Algum trabalho começa a ser realizado nesta área ([Cornell92]), mas é ainda um campo em investigação.

A alternativa de detecção de obstáculos baseada em fluxo óptico calculado sobre sequências de imagem obtidas por uma câmara em movimento, necessita de mais cuidados (cenário estático e estimação precisa de fluxo óptico) e deve ser desenvolvida também para movimento variável.

As características complementares dos métodos stereo e dos métodos baseados em fluxo óptico fazem pressupor que a utilização simultânea de ambos os processos (como aliás sucede com o ser humano) permita lidar com cenas sucessivamente mais complexas, incluindo múltiplos objectos estáticos ou em movimento, oclusões, etc.





## Bibliografia

- [Aloimonos88] "Active Vision",  
Y. Aloimonos, I. Weiss, A. Bandyopadhyay  
International Journal of Computer Vision, 1(4):333-356, Janeiro 1988.
- [Aloimonos90] "Purposive and Qualitative Active Vision",  
Y. Aloimonos  
proc. 10th International conference on Pattern Recognition, Atlantic City, NY,  
USA, 346-360, Junho 1990.
- [Aloimonos91] "Computer Vision",  
Y. Aloimonos, A. Rosenfeld  
Science vol. 253, articles, 1249-1254, Setembro 1991.
- [Anandan89] "A computational framework and an algorithm for the measurement of visual motion",  
P. Anandan  
International Journal of Computer Vision, 2, 283-310, 1989.
- [Bajcsy85] "Active Perception vs. Passive Perception",  
R. Bajcsy,  
Proceedings of IEEE workshop on Computer Vision, Bellair, MI, pp. 55-59,  
Outubro 1985.
- [Ballard82] "Computer Vision",  
D. Ballard, C. Brown  
Prentice-Hall, 1982.
- [Ballard91] "Animate Vision",  
D. Ballard  
Artificial Intelligence, 48:57-86, 1991.
- [Barron92] "Performance of Optical Flow Techniques",  
J. Barron, D. Fleet, S. Beauchemin,  
Robotics and Perception Laboratory,  
Department of Computing and Information Science, Queen's University,  
Kingston, Ontario, Canada K7L 3N6, Julho 1992.

## *Bibliografia*

- [Blake92] "Visual Exploration of Free-space",  
A. Blake, A. Zisserman, R. Cipolla  
in "Active Vision", pp.175-188, A. Blake, A. Yuille, Eds,  
MIT Press, 1992.
- [Bobet94] "Obstacle Detection Analysis",  
P. Bobet, C. Schmid, J-M. Bedrune, J. Crowley  
CVPR 94, Washington, 19-23 Junho 1994
- [Buffa90] "Obstacle avoidance and trajectory planning for an indoors mobile robot using  
stereo vision and Delaunay triangulation"  
M. Buffa, O. Faugeras, Z. Zhang  
Proceedings of Roundtable Discussion on Vision-Based Vehicle Guidance'90,  
pp12.1-12.8, Science University of Tokyo, IEEE, Julho 90
- [Buffa92] "A complete navigation system for a mobile robot, using a real-time  
stereovision and Delaunay triangulation",  
M. Buffa, O. Faugeras, Z. Zhang  
MVA'92 (IAPR Workshop on Machine Vision Applications), pp.191-194,  
Dezembro 1992
- [Canny83] "Finding edges and lines in images",  
MIT, Artif. Intell. Lab., Cambridge, MA, TR-720 1983.
- [Cornell92] "GPOD under variable camera geometry using a predictive stereo matcher",  
S. Cornell, J. Porril, J. Mayhew  
BMVC 1992.
- [Dickmanns92] "Expectation Based Dynamic Scene Understanding",  
E.D. Dickmanns  
in "Active Vision", pp303-335, A. Blake, A. Yuille, Eds,  
MIT Press, 1992.
- [Eryurtlu92] "A comparative study of grey level corner detectors",  
F. Eryurtlu and J. Kittler,  
Signal processing VI: Theories and applications,  
J. Vandewalle, R.Boite, M. Moonen, A. Oosterlinck (eds.), Elsevier Science, 1992.
- [Ferrari90] "A stereo vision system for real time obstacle avoidance in unknown  
environment",  
F. Ferrari, E. Grosso, G. Sandini, M. Magrassi  
IEEE international workshop on Intelligent Robots and Systems, IROS 1990.
- [Ferrari92] "VOILA",  
F. Ferrari, G. Garibotto, S. Masciangelo, G. Sandini  
ESPRIT Project P2502 (ECCV 1992 Esprit Open Day).

- [Gaspar92] "Reconstrução de superfícies usando regularização",  
J. A. Gaspar, José A. Santos Victor  
IST/ISR Grupo de Visão Robótica, relatório interno, Outubro 1992
- [Gaspar93] "Biblioteca para manipulação de imagem: img (v6.0)",  
J. A. Gaspar,  
IST/ISR Grupo de Visão Robótica  
Documento contido no pacote de distribuição da biblioteca, Junho 1993.
- [Gaspar94] "Ground Plane Obstacle Detection with a Stereo Vision System",  
J. A. Gaspar, J. S.-Victor, J. Sentieiro,  
Proceedings do International workshop on Intelligent Robotic Systems IRS94,  
Julho 1994.
- [Gaspar94] "Visual Obstacle Detection",  
J. A. Gaspar, J. S.-Victor, J. Sentieiro,  
1º Encontro Português de Controlo Automático: Controlo'94, Instituto Superior  
Técnico, Lisboa - Portugal, Setembro 1994.
- [Geman84] "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of  
images",  
S. Geman, D. Geman  
IEEE T-PAMI-6, pp721-741, Novembro 1984.
- [Gibson51] "The perception of the visual world",  
J. Gibson,  
Houghton Mifflin, Boston, 1951.
- [Guarnotta] "Studio de un problema di 'obstacle avoidance' per un robot mobile",  
F. Guarnotta, D. Rizzieri, F. Tarocchi, M. Tistarelli, G. Sandini  
DIST Genova, relatório interno.
- [Haralick92] "Computer and Robot Vision",  
R. Haralick, L. Shapiro,  
Addison-Wesley, 1992.
- [Horn86] "Robot Vision",  
B.K.P. Horn  
MIT Press, 1986.
- [Lim91] "AAAI-90 Workshop on Qualitative Vision"  
W. Lim, P. Kahn, D. Weinshall, A. Blake  
report of the workshop presentations and discussions, 1991.
- [Mallot91] "Inverse perspective mapping simplifies optical flow computation and obstacle  
detection"  
H.A. Mallot, H.H. Bülthoff, J.J. Little, S. Bohrer

## *Bibliografia*

Biological Cybernetics 64, pp.177-185,  
Springer-Verlag 1991.

[Marr82] "Vision: A Computational Investigation into the Human Representation and Processing of Visual Information"  
D. Marr  
W.H. Freeman & Co., San Francisco, 1982.

[Nasrabadi92] "Hopfield Network for Stereo Vision Correspondence",  
N.M. Nasrabadi, C.Y. Choo,  
IEEE Transactions on Neural Networks, Janeiro 1992.

[Nelson89] "Obstacle Avoidance Using Flow Divergence",  
R. Nelson, J. Aloimonos,  
IEEE T-PAMI, vol.11, No.10, Outubro 1989.

[Okutomi92] "A Locally Adaptive Window for Signal Matching",  
M. Okutomi, T. Kanade  
International Journal of Computer Vision, 7:2, 143-162, Jan. 1992.

[Poggio85] "Computational vision and regularization theory",  
T. Poggio, V. Torre, C. Koch,  
1985.

[Sandini90] "Active tracking strategy for monocular depth inference over multiple frames",  
G. Sandini, M. Tistarelli,  
IEEE T-PAMI, vol.12, No.1, Janeiro1990.

[Santos-Victor94+] "MEDUSA - A Stereo Head for Active Vision",  
J. S.-Victor, F. Trigt, J. Sentieiro,  
Proceedings do International workshop on Intelligent Robotic Systems IRS94,  
Julho 1994.

[Santos-Victor94] "Uncalibrated Obstacle Detection using Normal Flow",  
J.A. Santos-Victor, G. Sandini,  
DEEC/IST/ISR Lisboa, LIRA/DIST Genova,  
submetido a "Machine Vision and Applications", Junho 1994.

[Tancock91] "Broadcast Television Fundamentals",  
M. Tancock  
Pentech Press, Londres, 1991.

[Terzopoulos84] "Multiresolution computation of visible surface representations",  
D. Terzopoulos  
Ph.D. dissertation, Dep. Elec. Eng. Comput. Sci.,  
MIT Cambridge MA, Janeiro 1984.

[Tikhonov77] "Solutions of ill posed problems",  
A. Tikhonov, V. Arsenin,  
Winston & Sons, Washington DC, 1977.

[Zielke90] "Adapting computer vision systems to the visual environment: topographic  
mapping"  
T. Zielke, K. Storjohann, H. A. Mallot, W. Seelen  
ECCV 1990.

[Zisserman92] "Notes on geometric invariance in vision"  
A. Zisserman,  
BMVC92 tutorial.

## *Bibliografia*

## Apêndice A: Detecção de Pontos de Canto

Tal como foi definido no capítulo 2, um ponto de canto é um ponto onde a "curva de contorno"  $\theta(t)$ , sofre uma variação brusca de direcção, i.e.,  $|\partial\theta(t)/\partial t| \rightarrow \infty$  ([Haralick92]).

Alternativamente, canto também pode ser definido como um ponto da imagem para o qual todas as direcções apresentam um módulo de derivada elevado.

O detector de cantos de Moravec, usado neste trabalho ([Nasrabadi92]), tira partido desta segunda definição, e realiza estimativas de módulo de derivada em quatro direcções (vertical, horizontal, diagonal a 45° e diagonal a 135°).

Se o *mínimo* destas estimativas num dado ponto (x,y) é elevado, então todas as direcções exibem um módulo de derivada elevado, pelo que (x,y) é um potencial ponto de canto. O ponto (x,y) será classificado definitivamente como ponto de canto se o seu mínimo das estimativas de módulo de derivada, for *máximo* quando comparado com os mínimos dos pontos vizinhos numa dada janela (subdivisão da imagem, por exemplo com 20x20 pontos<sup>1</sup>). Impede-se desta forma a classificação de vários pontos vizinhos como pontos de canto e obtém-se um conjunto de pontos de interesse distribuído por toda a imagem.

Um problema conhecido deste detector de cantos, é o facto de ter uma resposta anisotrópica (ou seja dependente da direcção), uma vez que é baseado num conjunto discreto de direcções. Este problema não é significativo para a aplicação desejada nesta tese, emparelhamento de pontos de canto, porque um canto falso numa imagem não deve ter correspondente, pelo que na fase de emparelhamento é desprezado.

---

<sup>1</sup>Foi usada uma sobreposição de 1 ponto de largura entre janelas, para evitar a detecção de cantos em regiões com um nível de brilho constante.

### Aplicação ao ADOPP stereo:

Os pontos de canto nas imagens, são bons candidatos para o processo de emparelhamento porque implicam uma convexidade local da superfície de testes de SSD ([Anandan89]): a SSD apresenta um mínimo único (dentro de uma janela local).

Foram, por esta razão, usados na inicialização do Algoritmo de Detecção de Obstáculos sobre Pavimento Plano (ADOPP) para a obtenção de pontos correspondentes (dados necessários ao cálculo de parâmetros do mapa de referência).

Um exemplo dos resultados obtidos é ilustrado na figura A.1.

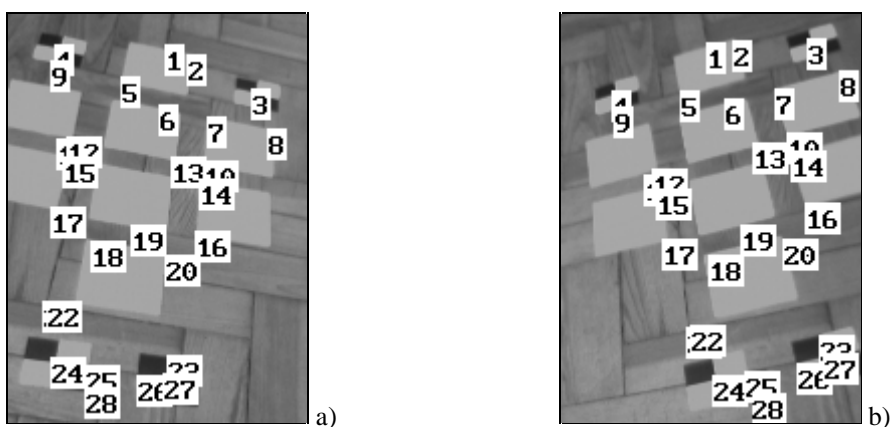


Fig. A.1 Cantos detectados e emparelhados para uma janela de detecção de cantos 20x20.



## Apêndice B: Correção da Distorção

Neste apêndice é analisado o resultado da correção da distorção (mapeamento da imagem esquerda sobre a imagem direita) contida no ADOPP stereo descrito no capítulo 3.

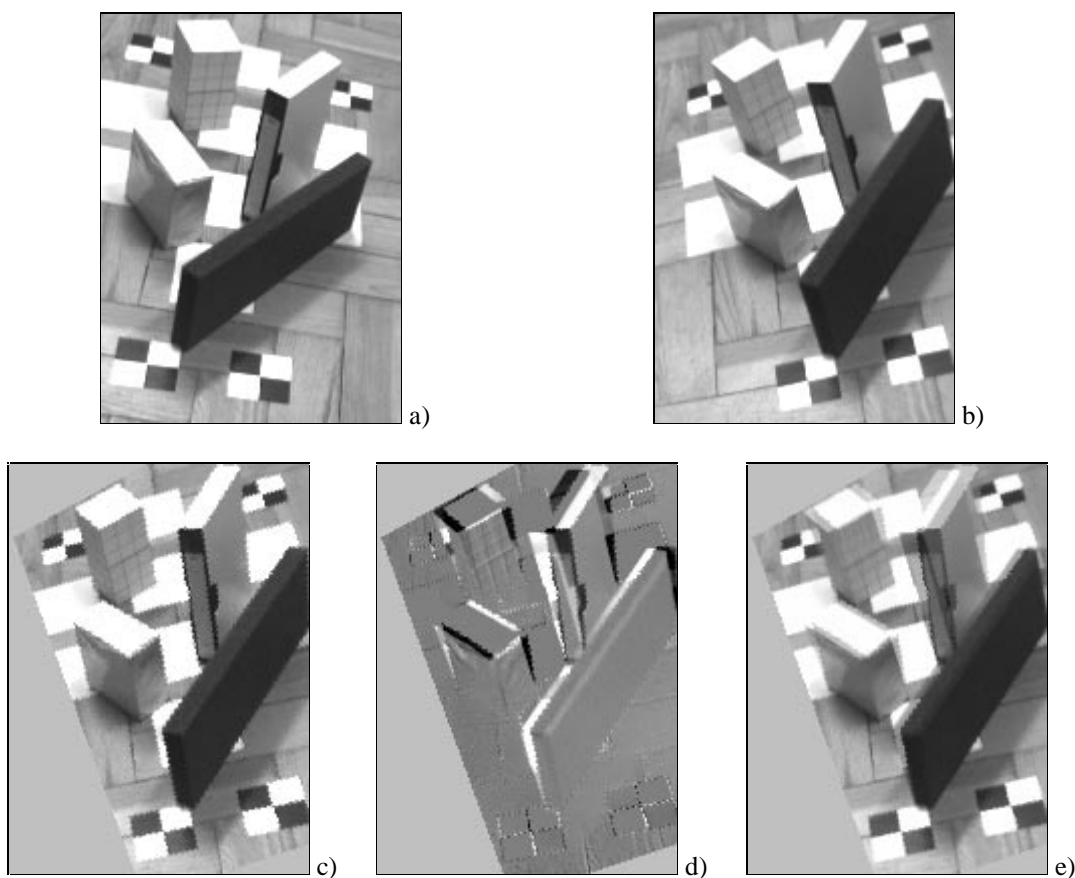


Fig. B.1 Correção da distorção: a,b são as imagens originais (esquerda e direita), 'c' é a imagem esquerda 'a' mapeada/corrigida sobre a direita 'b', 'd' é a diferença entre a imagem 'c' e a verdadeira imagem direita 'b' (mais um offset de 128), 'e' é a soma das imagens 'b' e 'c'

Notar que a correção da distorção consegue mapear a zona de pavimento da imagem esquerda sobre a imagem direita de uma forma praticamente coincidente. Por outro lado, quanto maior a altura do objecto (mais longe do pavimento) mais errada se revela a correção (esta é uma característica essencial para o ADOPP).

A correção da distorção é mais perfeita se for realizada uma interpolação de imagem (i.e., o mapa de referência pode passar a conter referências não inteiras, por ex. "float").

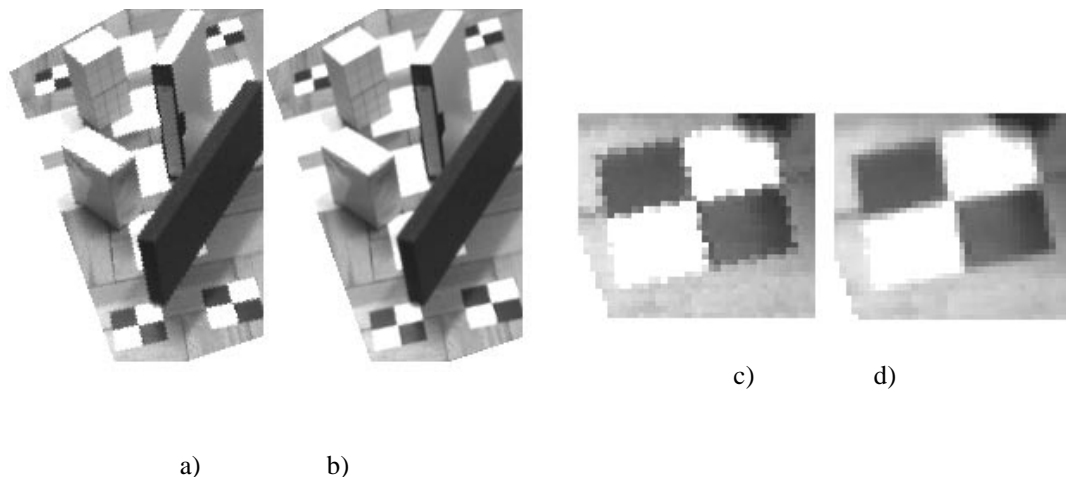


Fig. B.2 Comparação entre correção de distorção sem interpolação (imagem 'a') e com interpolação (imagem 'b'): 'c' e 'd' são pormenores de 'a' e 'b' ampliados.

A interpolação realizada, é uma interpolação bidimensional, com polinómios de Lagrange. O exemplo foi obtido com polinómios de grau 1, o que equivale visualmente à ligação por segmentos de recta, dos quatro pontos mais próximos do ponto a interpolar, como se ilustra na figura B.3.

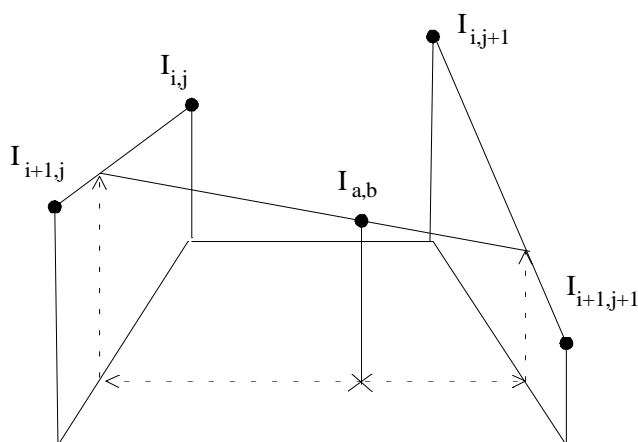


Fig. B.3 Interpolação bidimensional com polinómios de grau 1: "qual o nível de brilho no ponto (a,b) dados o níveis de brilho em (i,j), (i+1,j), (i,j+1) e em (i+1,j+1)"

Esta interpolação é usada na fase de inicialização, para ajuste de pontos correspondentes (implica salvaguarda de coordenadas de correspondentes na imagem esquerda com resolução superior: 'float').



## Apêndice C: Ajuste de um Parabolóide a uma Superfície de SSDs

A equação de um parabolóide:

$$z = ax^2 + bx + cy^2 + dy + exy + f \quad (\text{C.1})$$

pode ser escrita de forma linear, separando o vector de parâmetros:

$$z = M\theta \quad (\text{C.2})$$

$$\theta = [a \quad b \quad c \quad d \quad e \quad f] \quad (\text{C.3})$$

$$z = [x^2 \quad x \quad y^2 \quad y \quad xy \quad 1]\theta^T \quad (\text{C.4})$$

Estando disponíveis N dados (x,y,z) obtém-se um sistema sobre-determinado, solucionável por uma técnica de pseudo-inversa:

$$M = \begin{bmatrix} x_1^2 & x_1 & y_1^2 & y_1 & x_1 y_1 & 1 \\ x_2^2 & x_2 & y_2^2 & y_2 & x_2 y_2 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_N^2 & x_N & y_N^2 & y_N & x_N y_N & 1 \end{bmatrix} \quad (\text{C.5})$$

$$\theta = [M^T M]^{-1} M^T Z \quad (\text{C.6})$$

Uma análise mais atenta do problema mostra que é possível estimar  $\theta$  sem necessidade de criar a matriz M:

$$M^T M = \begin{bmatrix} m_{40} & m_{30} & m_{22} & m_{21} & m_{31} & m_{20} \\ m_{30} & m_{20} & m_{12} & m_{11} & m_{21} & m_{10} \\ m_{22} & m_{12} & m_{04} & m_{03} & m_{13} & m_{02} \\ m_{21} & m_{11} & m_{03} & m_{02} & m_{12} & m_{01} \\ m_{31} & m_{21} & m_{13} & m_{12} & m_{22} & m_{11} \\ m_{20} & m_{10} & m_{02} & m_{01} & m_{11} & N \end{bmatrix}, \quad m_{ab} = \sum_{i=1}^N x_i^a y_i^b \quad (\text{C.7})$$

$M^T M$  é naturalmente simétrica e tem somente 15 elementos  $m_{ab}$  distintos ( $a+b \leq 4$ ), onde:

$$M^T Z = [d_{20} \quad d_{10} \quad d_{02} \quad d_{01} \quad d_{11} \quad d_{00}]^T, \quad d_{ab} = \sum_{i=1}^N x_i^a y_i^b z_i \quad (C.8)$$

são mais 6 acumuladores. O ajuste da parábola para N dados (x,y,z) pode ser realizado por simples actualização dos 15+6 acumuladores por cada novo dado. O conjunto final de parâmetros obtêm-se calculando (C.6) ou seja, invertendo a matriz (C.7) e multiplicando o resultado pelo vector (C.8).

### Cálculo do mínimo do parabolóide (mínimo da SSD)

A partir de (C.1) pode ser calculado analiticamente um ponto de extremo por procura do ponto onde as derivadas parciais se anulam:

$$\begin{cases} \frac{\partial z}{\partial x} = 0 \\ \frac{\partial z}{\partial y} = 0 \end{cases} \Rightarrow \begin{cases} 2ax_{\min Z} + b + ey_{\min Z} = 0 \\ 2cy_{\min Z} + d + ex_{\min Z} = 0 \end{cases} \Rightarrow \begin{cases} x_{\min Z} = \frac{-2bc + de}{4ac - e^2} \\ y_{\min Z} = \frac{-2ad + be}{4ac - e^2} \end{cases} \quad (C.9)$$

Dado que z(x,y) ajusta localmente uma superfície de SSDs, a definição de SSD permite a afirmação de que o extremo é um mínimo (uma SSD tem um mínimo quando existe uma boa correspondência).

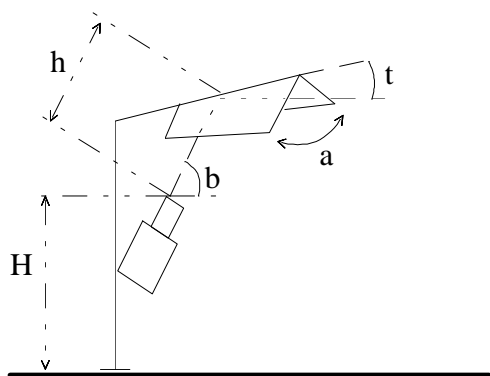
Este algoritmo está incluído na fase de inicialização do ADOPP, como uma das alternativas de refinamento do emparelhamento de pontos de referência.

## Apêndice D: Configuração do sistema de detecção de obstáculos

Neste apêndice descrevem-se as dimensões do sistema de espelhos, os parâmetros da câmara e da placa de aquisição, e algumas opções do software, usados para a obtenção dos resultados do capítulo 6.

O ajuste destes parâmetros reflecte-se no espaço visível, na altura mínima detectada, na estabilidade do sistema e naturalmente no tempo de cálculo.

### Sistema de espelhos:



$a$ = abertura dos espelhos=  $120^\circ$

$t$ = inclinação dos espelhos em relação à horizontal=  $11^\circ$

$b$ = ângulo do eixo da câmara em relação à horizontal=  $67^\circ$

$h$ = distância da lente aos espelhos (na direcção do eixo óptico)= 10 cm

$H$ = distância da lente ao pavimento= 61.5 cm

### Lente da câmara:

16 mm ou 8 mm ou 4.8 mm: 8 mm (compromisso entre campo de visão e resolução).

### **Software da placa de aquisição de imagem (DT3852):**

Resolução das imagens: as imagens foram sub-amostradas, tendo sido usadas somente as colunas e as linhas ímpares (multi-resolução é uma capacidade específica da placa de aquisição); compromisso entre velocidade e resolução.

Dimensão das imagens (máximo 256x256): foram usadas duas dimensões para as imagens 149x204 ou 167x216.

### **Software ADOPP:**

#### Inicialização:

Criação automática de novos pontos (correspondentes) de referência (por detecção e 'matching' de cantos) [S/N]: Sim

Métodos de refinamento do emparelhamento de pontos de referência:

- a) escolha do mínimo da SSD numa janela de procura.
- b) escolha do mínimo de SSD por evolução iterativa à volta do ponto de comparação original.
- c) escolha do mínimo de SSD por ajuste de uma parábola (aos valores de SSD) e cálculo analítico do mínimo.
- d), e) e f) iguais aos métodos a), b) e c) respectivamente, mas com uma correcção prévia da distorção (a ser anulada após a obtenção do mínimo da SSD).

Para o refinamento dos pontos correspondentes de referência dos exemplos mostrados no capítulo 6, foi sempre usada correcção prévia da distorção (métodos "d", "e" e "f").

#### Operação normal / Fase de detecção:

Critério de comparação de níveis de brilho (SSD / SMD): SMD

Limiares do filtro morfológico (limiar de violação significativa, dimensão da janela de contagem e limiar de ponto isolado):  $\text{threshold}_1=20$ ,  $W_2 \times W_2=3 \times 3$ ,  $\text{threshold}_2=5$  ou 6.