Orientação Automática na Baixa Pombalina: Classes de Equiprojectividade e Estimação Sequencial a Partir de Vídeo

André Martins

11 de Setembro de 2005

Resumo

O problema de inferir a orientação de uma câmara que adquire uma sequência de vídeo é de grande interesse em Visão por Computador, surgindo aplicações nas áreas de robótica móvel, calibração e reconstrução 3D.

As abordagens tradicionais requerem um passo intermédio de detecção de padrões (cantos ou linhas de contorno) em cada imagem e a sua posterior correspondência entre as várias imagens; este passo é computacionalmente pesado e requer o ajuste cuidadoso de muitos parâmetros. Em vez disso, este trabalho sugere uma abordagem probabilística de estimação sequencial fazendo uso de um modelo adequado a cenas urbanas, dito *mundo de Manhattan*, segundo o qual a maioria dos contornos se alinha em três direcções ortogonais.

As principais contribuições são: (i) a definição de classes de equivalência de orientações *equiprojectivas*; (ii) a introdução de um modelo de *pequenas rotações* adequado ao movimento da câmara; e (iii) a separação de cada estimação em duas partes reduzindo a complexidade de $\mathcal{O}(N^3)$ para $\mathcal{O}(N^2)$. A redução do peso computacional viabiliza a sua aplicação em tempo real. O desempenho é avaliado utilizando sequências de vídeo da Baixa Pombalina.

Palavras-chave: Visão por Computador, orientação da câmara, estimação sequencial, cenas urbanas, calibração da câmara.

Classificação ACM: **I.4** (IMAGE PROCESSING AND COMPUTER VISION), **I.5** (PATTERN RECOGNITION), **I.3** (COMPUTER GRAPHICS), **I.6** (SIMULATION AND MODELING).

Conteúdo

Capítulo 1

Introdução

1.1 Motivação

Aplicações em áreas como Vídeo Digital, Realidade Virtual, Robótica Móvel e Navegação Automática requerem métodos eficientes para estimar a "pose" (*i.e.*, a posição e a orientação) de uma câmara de vídeo, ao longo do tempo, a partir da sequência de imagens que esta captura. De modo análogo, em Processamento de Imagem, Reconhecimento de Padrões e Pesquisa de Imagens surge muitas vezes a necessidade de, a partir de uma única imagem, caracterizar a estrutura tridimensional (3D) de objectos, o que muitas vezes requer, de alguma forma, estimar a pose da câmara relativamente a esses objectos. Em resumo, pode afirmar-se que o desenvolvimento de métodos eficientes para a estimação da posição e/ou orientação de uma câmara, quer a partir de uma imagem, quer a partir de uma sequência de imagens, é actualmente uma das maiores preocupações em Visão por Computador.

As abordagens tradicionais baseiam-se na detecção de "padrões" em imagens; estes padrões consistem geralmente em cantos de objectos ou em linhas de contorno. Em aplicações envolvendo múltiplas imagens, uma vez detectados estes padrões em cada imagem, o passo seguinte consiste em correspondê-los entre diferentes imagens, por exemplo através de *seguimento* (pode encontrar-se alguns exemplos em [1, 2, 3]). Em aplicações que utilizam uma única imagem, os métodos mais comuns envolvem o *agrupamento* de padrões (ver por exemplo [4, 5, 6]). Todavia, é consensual que tanto a correspondência como o agrupamento automático de padrões são problemas difíceis e para os quais os resultados até agora atingidos se revelam pouco satisfatórios; existe um sério compromisso entre robustez e comportabilidade computacional que muitas vezes inviabiliza o seu uso em aplicações práticas. Para além disso, o facto de se basear toda a inferência num conjunto de padrões geralmente pequeno (com relação à totalidade da imagem) faz com que informação útil possa ser prematuramente desprezada.

No caso de múltiplas vistas, têm sido propostos métodos que estimam a estrutura 3D directamente a partir dos valores da intensidade das imagens, *i.e.*, sem envolver a detecção e correspondência de padrões – exemplos disso encontram-se em [7, 8]. Porém, estas abordagens conduzem quase sempre a algoritmos complexos, de convergência lenta, e demasiado dependentes da hipótese de que, de vista para vista, o padrão de brilho em pontos correspondentes permanece aproximadamente constante, tornando-os por isso muito sensíveis a ruído.

No caso de uma única imagem, J. Coughlan e A. Yuille propuseram recentemente uma abordagem diferente em [9, 10] que evita a detecção e agrupamento de padrões. A ideia consiste em utilizar conhecimento prévio sobre a estrutura do "mundo". De facto, em grande parte das cenas urbanas, muitos contornos estão alinhados com uma de três

direcções ortogonais, definindo um sistema de eixos. Sob este modelo que designaram por *mundo de Manhattan*, J. Coughlan e A. Yuille utilizaram técnicas de inferência bayesiana para estimar a componente rotacional da pose 3D (*i.e.*, a orientação) da câmara com relação àquele sistema de eixos. O modelo de mundo de Manhattan foi aplicado depois disso (paralelamente ao trabalho que aqui se apresenta) em [11] para auto-calibração da câmara e estendido em [12] para ambientes urbanos mais genéricos.

O trabalho aqui apresentado inspira-se no modelo de mundo de Manhattan para propor um novo método de estimação da orientação 3D a partir de sequências de imagens de cenas urbanas. As contribuições originais são:

- enquanto que em [9, 10] o modelo de mundo de Manhattan é utilizado para estimar a orientação a partir de uma única imagem, este método estende o seu uso para *sequências* de imagens;
- é introduzido um novo modelo de *pequenas rotações* que expressa o facto de a câmara de vídeo movimentar-se suave e continuamente no espaço 3D;
- define-se o conjunto das orientações 3D em termos de classes de equivalência de *orientações equiprojectivas*, onde são consideradas equivalentes as orientações que conduzem ao mesmo *conjunto* de pontos de fuga e, por isso, são indistinguíveis do ponto de vista da estimação. Mostra-se como cada classe de equivalência tem 24 elementos e se relaciona geometricamente com o grupo octaédrico das simetrias próprias do cubo. Reduz-se o espaço de procura da solução para uma região mais pequena que contém o conjunto quociente;
- para cada imagem, decompõe-se em dois passos a estimação dos três ângulos que parametrizam a orientação da câmara: um passo de complexidade $\mathcal{O}(N^2)$ onde são estimados dois destes ângulos (elevação e torção), e outro passo de complexidade $\mathcal{O}(N)$ onde é estimado o terceiro ângulo (azimute). Esta decomposição reduz consideravelmente o peso computacional de cada estimação, pois reduz a complexidade de $\mathcal{O}(N^3)$ para $\mathcal{O}(N^2 + N) = \mathcal{O}(N^2)$;
- mostra-se teoricamente como o modelo de mundo de Manhattan é um caso particular de uma classe de mundos estruturados de dimensão arbitrária e como, uma vez formalizada essa estrutura, todos os resultados fundamentais podem ser extrapolados.

1.2 Conteúdo

Este trabalho está organizado da seguinte forma: no Capítulo II introduz-se a geometria inerente ao modelo de mundo de Manhattan, definindo-se o conceito de orientações equiprojectivas e o modelo de pequenas rotações; mostra-se também como os mesmos resultados teóricos podem ser extrapolados para uma classe de mundos estruturados de dimensão arbitrária. O Capítulo III debruça-se sobre a abordagem probabilística e consequente implementação do algoritmo de estimação sequencial da orientação da câmara, tendo em conta os conceitos definidos no Capítulo II; discute-se ainda os resultados experimentais da aplicação deste algoritmo a sequências de vídeo reais. Finalmente, o Capítulo IV expõe as conclusões e sugere desenvolvimentos futuros.

Capítulo 2

Geometria do mundo de Manhattan

2.1 Mundo de Manhattan

Em Visão por Computador, designa-se por *mundo* o espaço 3D observado por um sistema de visão, geralmente constituído por uma ou mais câmaras de vídeo ligadas a uma unidade de processamento. O sistema de navegação aqui proposto é constituído por uma única câmara de vídeo digital ligada a um processador, capturando uma sequência de imagens $\{\mathtt{I}_1, \mathtt{I}_2, \ldots, \mathtt{I}_n\}$. O objectivo do sistema consiste em obter uma sequência de estimativas $\{\widehat{\mathtt{O}}_1, \widehat{\mathtt{O}}_2, \ldots, \widehat{\mathtt{O}}_n\}$ da orientação da câmara para cada um destes instantes. Considera-se parte do objectivo obter desempenhos capazes de viabilizar o uso do sistema em tempo real, pelo que se dará ênfase à rapidez do processamento.

Em [9, 10], J. Coughlan e A. Yuille sugeriram uma nova abordagem para estimar a orientação da câmara a partir de uma única imagem. O seu trabalho, desenvolvido para o Smith-Kettlewell Eye Research Institute, em São Francisco, tinha como objectivo implementar um sistema de ajuda para cegos e amblíopes capaz de orientá-los nas ruas, avenidas ou no interior de edifícios de uma cidade; o sistema seria constituído por uma câmara acoplada no peito do utilizador, cujas imagens seriam processadas de modo a emitir, por exemplo, um aviso sonoro se a trajectória tendesse a desviá-lo para a estrada ou contra um obstáculo como uma parede.

J. Coughlan e A. Yuille decidiram tirar partido da natureza "urbana" do mundo para onde o sistema de navegação foi idealizado e, como alternativa aos métodos tradicionais, propuseram seguir uma abordagem estatística baseada numa modelização prévia do mundo capaz de captar essa natureza "urbana". Assim surgiu o conceito de "mundo de Manhattan", que se procurará aqui definir com alguma informalidade.

Definição 2.1 Mundo de Manhattan *(do inglês* Manhattan world*) é um modelo do mundo segundo o qual este é constituído essencialmente por objectos cujas arestas são linhas rectas que, no seu conjunto, estão alinhadas segundo três direcções ortogonais. O referencial ortonormado* $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ *definido por estas direcções designa-se por* **referencial de Manhattan***, sendo* $\mathbf{x}$*,* $\mathbf{y}$ *e* $\mathbf{z}$ *designados por* **eixos de Manhattan***.*

As Figuras 2.1a-c mostram uma representação esquemática de um mundo de Manhattan e dois exemplos reais de mundos passíveis de serem modelizados como sendo "de Manhattan": uma fotografia de uma cena interior e outra de uma exterior. Grande parte das cenas "urbanas" podem ser classificadas como mundos de Manhattan; as direcções ortogonais são devidas à presença de salas, corredores, ruas, avenidas, edifícios, etc. A

própria designação é inspirada em Manhattan (Nova Iorque), que se caracteriza pelo desenho ortogonal das suas ruas e avenidas e pela paisagem "paralelepipédica" conferida pelos arranha-céus.

A Baixa Pombalina, em Lisboa, é também um exemplo real de mundo de Manhattan. Por esse motivo, grande parte das experiências realizadas resultaram de sequências de vídeo obtidas nesse local.

Recentemente, outros trabalhos adoptaram modelos inspirados no mundo de Manhattan para cenários ligeiramente diferentes. Um exemplo é o *mundo de Atlanta* [12], que pode ser visto como uma generalização do mundo de Manhattan adaptado a cenas onde predominam edifícios paralelepipédicos que partilham uma direcção comum – a vertical –, mas que fazem um ângulo de azimute entre si.

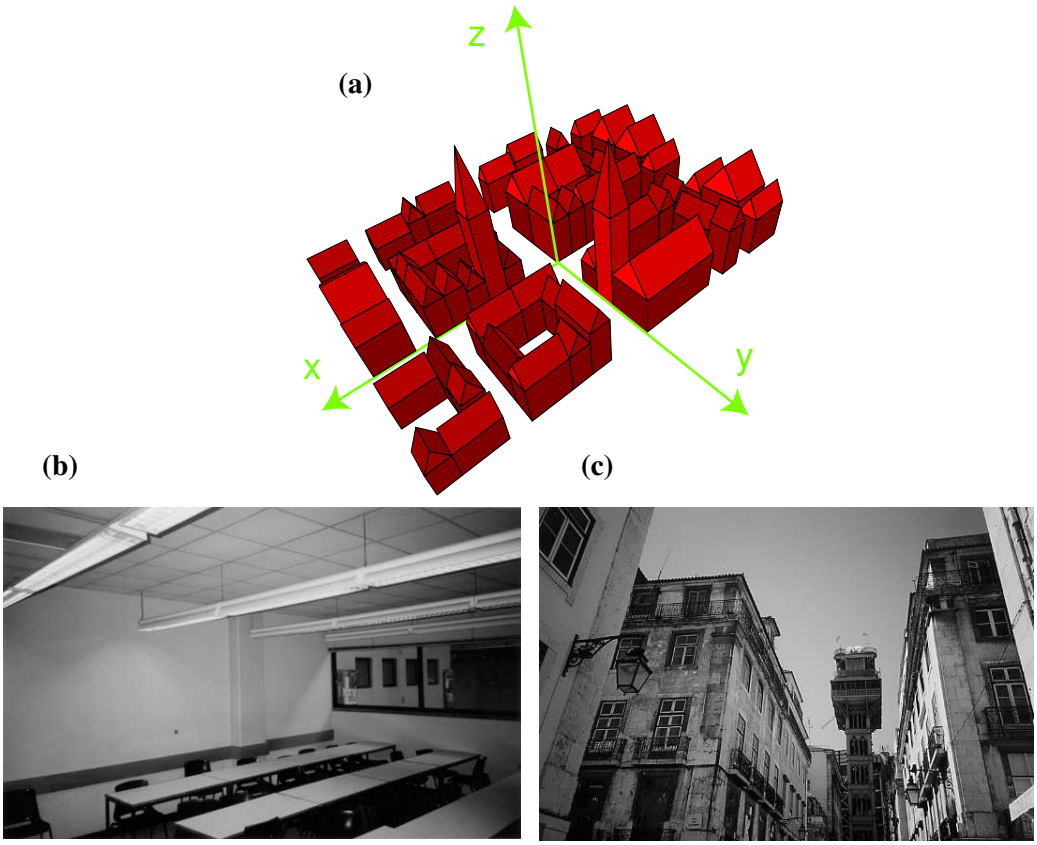

Figura 2.1: Mundos de Manhattan: (a) Representação esquemática; (b) Fotografia de uma cena interior: sala de aula; (c) Fotografia de uma cena exterior: Rua de Santa Justa, na Baixa Pombalina de Lisboa. Note-se como em (b) e (c) são visíveis as três direcções ortogonais dos eixos de Manhattan.

2.2 Orientação da câmara

O modelo utilizado para a câmara é o da *câmara escura*[1]. Para uma câmara calibrada, a adopção deste modelo não implica perda de generalidade, pois o conhecimento dos parâmetros intrínsecos da câmara permite rectificar as imagens de modo a simular uma

[1]Designado na literatura anglo-saxónica por *pinhole camera*.

câmara escura – em [13] e [14], por exemplo, são dados alguns exemplos de métodos de calibração.

De acordo com este modelo, todos os raios ópticos convergem num ponto – o **centro óptico** – e são projectados numa superfície planar – o **plano da imagem**. A distância entre o centro óptico e o plano da imagem é designada por **distância focal** e o ponto do plano da imagem mais próximo do centro óptico designa-se por **ponto principal**. A Figura 2.2 representa esquematicamente a câmara escura.

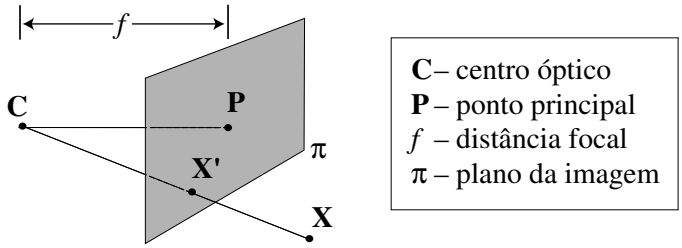

Figura 2.2: Representação esquemática da câmara escura. Um ponto $\mathbf{X}$ no espaço tridimensional é projectado no ponto $\mathbf{X}'$ do plano da imagem.

Sejam respectivamente $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ e $(\mathbf{n}, \mathbf{h}, \mathbf{v})$ os referenciais do mundo de Manhattan e da câmara. Estes relacionam-se através da equação

$$(\mathbf{n}, \mathbf{h}, \mathbf{v}) = (\mathbf{x}, \mathbf{y}, \mathbf{z}) \cdot \mathtt{O} \ , \tag{2.1}$$

onde $\mathtt{O} \in \mathrm{SO}(3)$ é uma matriz de rotação que representa a **orientação** da câmara. Adoptando o sistema de coordenadas definido pelo referencial de Manhattan, vem:

$$\mathtt{O} = [\mathbf{n}, \mathbf{h}, \mathbf{v}]. \tag{2.2}$$

A orientação tem três graus de liberdade e pode ser parametrizada por três ângulos exprimindo três rotações sucessivas (ver Fig. 2.3):

- α, o ângulo de *compasso* ou azimute, correspondendo a uma rotação em torno do eixo $\mathbf{z}$;
- β, o ângulo de *elevação* sobre o plano $\mathbf{xy}$;
- γ, o ângulo de *torção* em torno da linha de vista.

No texto que se segue, a orientação é frequentemente representada em função destes parâmetros, $\mathtt{O} \equiv \mathtt{O}(\alpha, \beta, \gamma)$.

De acordo com esta parametrização, os ângulos α, β e γ relacionam-se com $\mathbf{n}$, $\mathbf{h}$ e $\mathbf{v}$ da seguinte forma:

$$\mathbf{n} = [\cos\alpha\cos\beta, \sin\alpha\cos\beta, \sin\beta]^T \tag{2.3}$$

e

$$[\mathbf{h}, \mathbf{v}] = [\mathbf{h}_0, \mathbf{v}_0] \begin{bmatrix} \cos\gamma & -\sin\gamma \\ \sin\gamma & \cos\gamma \end{bmatrix}, \tag{2.4}$$

onde, representando por "$\times$" o produto externo vectorial,

$$\mathbf{h}_0 = \frac{\mathbf{z} \times \mathbf{n}}{||\mathbf{z} \times \mathbf{n}||} = [-\sin\alpha, \cos\alpha, 0]^T \quad \text{e} \tag{2.5}$$

$$\mathbf{v}_0 = \mathbf{n} \times \mathbf{h}_0 \ . \tag{2.6}$$

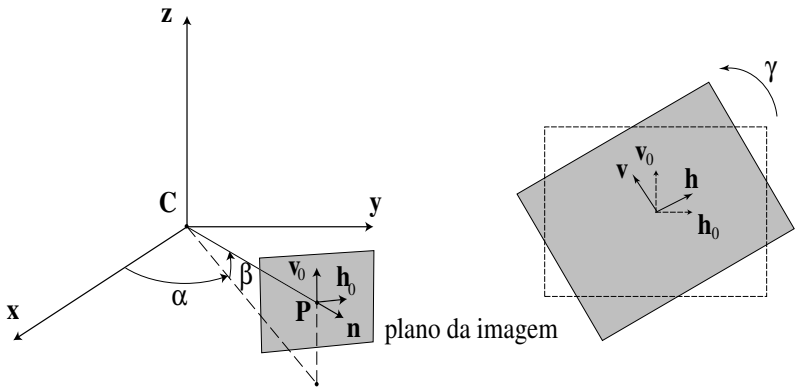

Figura 2.3: Parametrização da orientação da câmara. À esquerda: o ângulo de compasso α e o ângulo de elevação β. À direita: o ângulo de torção γ representado no plano da imagem.

Utilizando coordenadas homogéneas[2], e escolhendo o centro óptico $\mathbf{C}$ como origem do sistema de coordenadas, *i.e.*, $\mathbf{C} = [0,0,0,1]^T$, vem para o ponto principal $\mathbf{P}$ e para o plano da imagem π:

$$\mathbf{P} = \begin{bmatrix} f\mathbf{n} \\ 1 \end{bmatrix} \quad \text{e} \quad \pi = \begin{bmatrix} \mathbf{n} \\ -f \end{bmatrix}, \tag{2.7}$$

onde f representa a distância focal. Sendo $(\mathbf{P}; \mathbf{n}, \mathbf{h}, \mathbf{v})$ o referencial afim da câmara, as coordenadas homogéneas de um ponto neste referencial obtêm-se das suas coordenadas homogéneas no referencial $(\mathbf{C}; \mathbf{x}, \mathbf{y}, \mathbf{z})$ multiplicando à esquerda por uma matriz de transformação 4×4 dada por

$$\mathtt{T} = \begin{bmatrix} \mathbf{n}^T & -f \\ \mathbf{h}^T & 0 \\ \mathbf{v}^T & 0 \\ \mathbf{0}^T & 1 \end{bmatrix} = \begin{bmatrix} n_x & n_y & n_z & -f \\ h_x & h_y & h_z & 0 \\ v_x & v_y & v_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{2.8}$$

2.3 Pontos de fuga

No espaço projectivo $\mathbb{P}^3$, a intersecção de quaisquer duas rectas com a mesma direcção $\mathbf{d}$ é um ponto ideal. A projecção desse ponto ideal no plano da imagem designa-se por **ponto de fuga** segundo $\mathbf{d}$. Naturalmente, escolhida uma direcção $\mathbf{d}$, o respectivo ponto de fuga pode obter-se calculando a intersecção do plano da imagem com a recta de direcção $\mathbf{d}$ que passa no centro óptico.

Posto isto, designando por $\mathbf{W_x}$, $\mathbf{W_y}$ e $\mathbf{W_z}$ os vectores com as coordenadas homogéneas dos pontos de fuga segundo cada um dos eixos de Manhattan $\mathbf{x}$, $\mathbf{y}$ e $\mathbf{z}$, estes serão necessariamente da forma $[w_{x1}, 0, 0, w_{x4}]^T$, $[0, w_{y2}, 0, w_{y4}]^T$ e $[0, 0, w_{z3}, w_{z4}]^T$, respectivamente, e satisfarão $\pi^T\mathbf{W_x} = \pi^T\mathbf{W_y} = \pi^T\mathbf{W_z} = 0$, o que, recorrendo a (2.7), conduz a:

$$\mathbf{W_x} = [f, 0, 0, n_x]^T \quad , \quad \mathbf{W_y} = [0, f, 0, n_y]^T \quad \text{e} \quad \mathbf{W_z} = [0, 0, f, n_z]^T. \tag{2.9}$$

Mudando para o referencial da câmara, o que é feito utilizando a matriz $\mathtt{T}$ de (2.8), correspondem aos mesmos pontos as coordenadas homogéneas $\mathtt{T}\mathbf{W_x} = [0, fh_x, fv_x, n_x]^T$, $\mathtt{T}\mathbf{W_y} = [0, fh_y, fv_y, n_y]^T$ e $\mathtt{T}\mathbf{W_z} = [0, fh_z, fv_z, n_z]^T$, respectivamente.

[2]Ver [15] para uma breve explicação desta e de outras noções de Geometria Projectiva.

Identificando agora o plano da imagem π com o plano projectivo $\mathbb{P}^2$, os mesmos pontos de fuga, vistos como pontos 2D, descrevem-se no referencial afim $(\mathbf{P}; \mathbf{h}, \mathbf{v})$ através dos vectores de coordenadas homogéneas $\mathbf{v_x}$, $\mathbf{v_y}$ e $\mathbf{v_z}$ (ver Figura 2.4a) dados por

$$\mathbf{v_x} = [fh_x, fv_x, n_x]^T = \mathtt{R}_\gamma \cdot [-f\sin\alpha, -f\cos\alpha\sin\beta, \cos\alpha\cos\beta]^T \ , \quad (2.10)$$

$$\mathbf{v_y} = [fh_y, fv_y, n_y]^T = \mathtt{R}_\gamma \cdot [f\cos\alpha, -f\sin\alpha\sin\beta, \sin\alpha\cos\beta]^T \ , \quad (2.11)$$

$$\mathbf{v_z} = [fh_z, fv_z, n_z]^T = \mathtt{R}_\gamma \cdot [0, f\cos\beta, \sin\beta]^T . \quad (2.12)$$

com

$$\mathtt{R}_\gamma = \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} . \quad (2.13)$$

onde se recorreu a (2.3)-(2.6). Confrontando as expressões (2.10)-(2.12) com (2.2), pode ver-se que os pontos de fuga $\mathbf{v_x}$, $\mathbf{v_y}$ e $\mathbf{v_z}$ se relacionam com as linhas da matriz de orientação de acordo com

$$[\mathbf{v_x}, \mathbf{v_y}, \mathbf{v_z}] = \begin{bmatrix} 0 & f & 0 \\ 0 & 0 & f \\ 1 & 0 & 0 \end{bmatrix} \mathtt{O}^T \ . \quad (2.14)$$

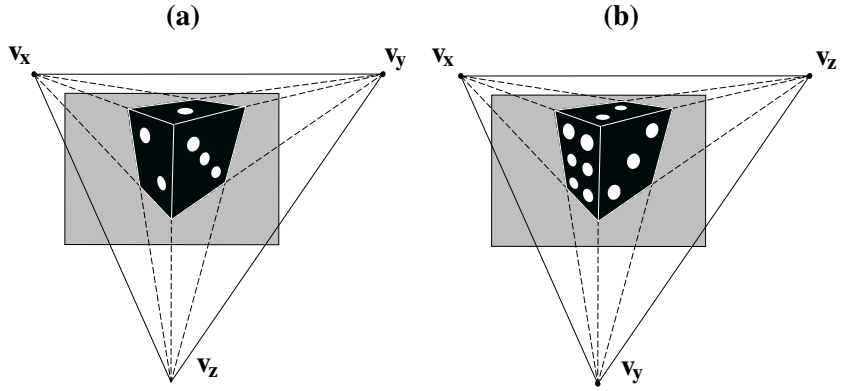

Figura 2.4: Pontos de fuga dos eixos de Manhattan sobre duas imagens de um dado. A orientação da câmara é diferente nas imagens (a) e (b); porém, os pontos de fuga, no seu conjunto, têm a mesma localização. Note-se como, de (a) para (b), os pontos de fuga $\mathbf{v_y}$ e $\mathbf{v_z}$ são permutados.

2.4 Orientações equiprojectivas

A expressão (2.14) mostra que a orientação da câmara $\mathtt{O}$ pode ser estimada mediante a localização na imagem dos pontos de fuga dos eixos de Manhattan, $\mathbf{v_x}$, $\mathbf{v_y}$ e $\mathbf{v_z}$. Porém, partindo apenas de uma imagem de um mundo de Manhattan, sem informação adicional sobre a geometria do mundo, tais pontos de fuga são indistinguíveis, *i.e.*, localizá-los equivale a conhecer o conjunto $\{\mathbf{v_x}, \mathbf{v_y}, \mathbf{v_z}\}$, não se sabendo, no entanto, qual o ponto de fuga que corresponde a cada eixo de Manhattan – a possibilidade de *permutar* os pontos de fuga conduz a múltiplas soluções para a orientação. Esta situação é ilustrada nas Figuras 2.4a-b, onde se mostra como duas orientações da câmara distintas originam o mesmo conjunto de pontos de fuga. Esta ambiguidade motiva o conceito de *equiprojectividade*.

Definição 2.2 (Orientações equiprojectivas) *Seja* $\mathcal{V}(\mathtt{O}) = \{\mathbf{v_x}, \mathbf{v_y}, \mathbf{v_z}\}$ *o conjunto dos pontos de fuga determinado por uma orientação* $\mathtt{O}$. *Duas orientações* $\mathtt{O}$ *e* $\mathtt{O}'$ *denominam-se* **equiprojectivas** sse *possuem idênticos conjuntos de pontos de fuga, i.e.,* sse $\mathcal{V}(\mathtt{O}) = \mathcal{V}(\mathtt{O}')$.

A equiprojectividade, como se acaba de definir, satisfaz as propriedades de reflexividade, simetria e transitividade; portanto, é uma relação de equivalência. O resultado seguinte mostra como pode obter-se a classe de equivalência $\mathcal{E}(\mathtt{O})$ de uma dada orientação, *i.e.*, o conjunto de todas as orientações equiprojectivas com $\mathtt{O}$.

Proposição 2.3 *Designe-se por* $\mathrm{SP}^+(n)$ *o grupo das* **matrizes de permutação com sinal** $n \times n$ **cujo determinante é positivo** *(i.e., o conjunto das matrizes* $n \times n$ *com entradas em* $\{-1, 0, 1\}$*, com exactamente um elemento não nulo por linha e por coluna e com determinante positivo, necessariamente igual a* 1*).*

Duas orientações $\mathtt{O}$ *e* $\mathtt{O}'$ *são equiprojectivas* sse *existir* $\mathtt{M} \in \mathrm{SP}^+(3)$ *tal que* $\mathtt{O}' = \mathtt{MO}$. *Cada classe de equivalência de orientações equiprojectivas tem exactamente* 24 *elementos, o número de elementos de* $\mathrm{SP}^+(3)$.

Demonstração: Seja $\mathtt{V}(\mathtt{O}) = [\mathbf{v_x}, \mathbf{v_y}, \mathbf{v_z}]$ a matriz definida pelos pontos de fuga associados a $\mathtt{O}$. $\mathtt{O}$ e $\mathtt{O}'$ são equiprojectivas *sse* possuem os mesmos pontos de fuga a menos de um factor de escala e de uma permutação, *i.e.*, *sse* existem uma matriz de permutação $\mathtt{P}$ e uma matriz diagonal $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3)$ tais que $\mathtt{V}(\mathtt{O}) = \mathtt{V}(\mathtt{O}')\Lambda\mathtt{P}$. De (2.14) tem-se

$$\mathtt{V}(\mathtt{O}) = \mathtt{K}\mathtt{O}^T = \begin{bmatrix} 0 & f & 0 \\ 0 & 0 & f \\ 1 & 0 & 0 \end{bmatrix} \mathtt{O}^T \ , \tag{2.15}$$

pelo que multiplicando ambos os membros à esquerda por $\mathtt{K}^{-1}$ ($\mathtt{K}$ é obviamente invertível) conclui-se que $\mathtt{O}$ e $\mathtt{O}'$ são equiprojectivas *sse* $\mathtt{O}^T = \mathtt{O}'^T\Lambda\mathtt{P}$, *i.e.*, *sse* $\mathtt{O}' = \Lambda\mathtt{P}\mathtt{O} = \mathtt{MO}$, com $\mathtt{M} = \Lambda\mathtt{P}$, ou $m_{ij} = \lambda_i p_{ij}$. Dado que $\mathtt{O}, \mathtt{O}' \in \mathrm{SO}(3)$, vem $|\lambda_1| = |\lambda_2| = |\lambda_3| = 1$, *i.e.*, $m_{ij} = \pm p_{ij}$, e $\det \mathtt{M} = 1$; isto equivale a ter $\mathtt{M} \in \mathrm{SP}^+(3)$. Ora, o grupo $\mathrm{SP}^+(n)$ é isomorfo ao grupo octaédrico das simetrias "próprias" do n-cubo e tem ordem $\frac{1}{2}\, n!\, 2^n$, o total de possíveis combinações de um número par de operações de permutação e troca de sinal com n elementos. Para $n = 3$, vem $\frac{1}{2}\, 3!\, 2^3 = 24$, sendo este o número de elementos de cada classe de equivalência definida pela relação de equiprojectividade. ■

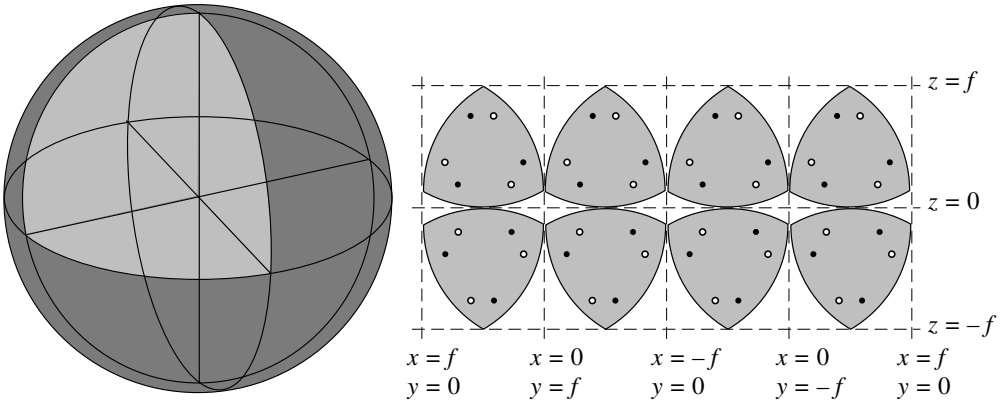

Figura 2.5: Localização 3D dos pontos principais de orientações equiprojectivas, nos octantes de uma esfera de raio f centrada em $\mathbf{C}$. Representam-se duas classes de equivalência: os pontos brancos e os pontos pretos; os segundos são o "reflexo" dos primeiros.

A Figura 2.5 ilustra duas classes de equivalência formadas por orientações equiprojectivas, em que uma é o "reflexo" da outra.

O conceito de equiprojectividade revela-se útil em qualquer problema de estimação de orientação ou de localização de pontos de fuga, visto que permite reduzir os espaços de procura. Assim, em vez de se procurar sobre todo o domínio SO(3), pode recorrer-se à relação de equivalência definida pela equiprojectividade e utilizar um espaço de procura para a orientação mais pequeno que contenha o conjunto quociente $\mathrm{SO}(3)/\mathrm{SP}^{+}(3)$. A proposição seguinte formaliza este reciocínio:

Proposição 2.4 *Toda a orientação* $\mathtt{O} \in \mathrm{SO}(3)$ *tem uma equiprojectiva* $\mathtt{O}' \in \mathcal{R}$*, onde* $\mathcal{R}$ *é uma região em* SO(3) *definida por:*

$$\mathcal{R} = \left\{ \mathtt{O}(\alpha,\beta,\gamma) \in \mathrm{SO}(3) : \alpha \in \left]-\frac{\pi}{4},\frac{\pi}{4}\right], \;\; \beta \in \left]-\frac{\pi}{4},\frac{\pi}{4}\right], \;\; e \;\; \gamma \in \left]-\varphi,\varphi\right] \right\}, \qquad (2.16)$$

com $\varphi = \arctan\sqrt{2} \approx 54.7°$*. Uma afirmação equivalente é: qualquer que seja a orientação da câmara, existe pelo menos um ponto de fuga na região representada na Fig. 2.6.*

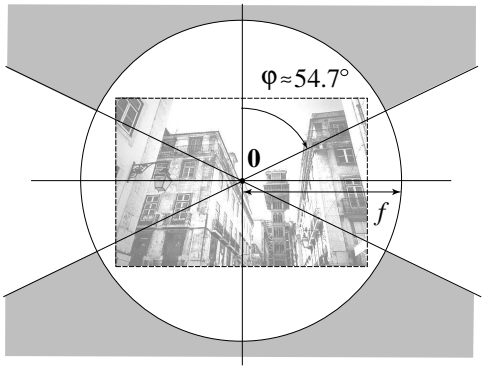

Figura 2.6: Representação do plano da imagem. É garantida a existência de pelo menos um ponto de fuga na região a sombreado.

Demonstração: Por simplicidade, vamos utilizar coordenadas cartesianas para os pontos de fuga, *i.e.*, $\widetilde{\mathbf{v}}_i = [v_{i1}/v_{i3}, v_{i2}/v_{i3}]^T = [fh_i/n_i, fv_i/n_i]^T$ para $i \in \{x,y,z\}$, sendo $\mathbf{v}_i = [v_{i1}, v_{i2}, v_{i3}]^T$ as coordenadas homogéneas. Como se poderá ver recorrendo a limites, a eventual existência de pontos "no infinito", *i.e.*, para os quais $v_{i3} = 0$, não implica perda de generalidade. De (2.10)-(2.12) tem-se (para $i,j \in \{x,y,z\}$)

$$\widetilde{\mathbf{v}}_i^T \widetilde{\mathbf{v}}_j = \begin{cases} f^2\left(\frac{1}{n_i^2} - 1\right) & \text{se } i = j \\ -f^2 & \text{se } i \neq j, \end{cases} \qquad (2.17)$$

o que permite obter tanto a distância euclidiana $d_i = (\widetilde{\mathbf{v}}_i^T \widetilde{\mathbf{v}}_i)^{1/2}$ entre os pontos $\widetilde{\mathbf{v}}_i$ e $\widetilde{\mathbf{p}} = [0,0]^T$ como o ângulo $\theta_{ij} = \arccos \frac{\widetilde{\mathbf{v}}_i^T \widetilde{\mathbf{v}}_j}{d_i d_j}$ formado pelas linhas $[\widetilde{\mathbf{p}}\widetilde{\mathbf{v}}_i]$ e $[\widetilde{\mathbf{p}}\,\widetilde{\mathbf{v}}_j]$, com $i \neq j$.

Considere-se agora o disco $\mathcal{D}$ com raio f e centro em $\widetilde{\mathbf{p}}$, *i.e.*, $\mathcal{D} = \{(u,v) \in \mathbb{R}^2 : u^2 + v^2 \leq f^2\}$. Temos que $\widetilde{\mathbf{v}}_i \in \mathcal{D}$ *sse* $d_i \leq f$, o que, por (2.17), é equivalente a $n_i^2 \geq 1/2$. Visto que $n_x^2 + n_y^2 + n_z^2 = 1$, a condição $n_i^2 \geq 1/2$ implica $n_j^2 \leq 1/2$ para qualquer $j \neq i$, o que significa que não pode existir mais que um ponto de fuga no interior do círculo $\mathcal{D}$. Para além disso, os três pontos de fuga estão todos na fronteira ou no exterior de $\mathcal{D}$ *sse* $n_i^2 \leq 1/2$, para $i \in \{x,y,z\}$.

Para completar a demonstração, é ainda necessário o seguinte resultado intermédio:

Lema 2.5 *Quaisquer dois pontos de fuga* $\widetilde{\mathbf{v}}_i$ *e* $\widetilde{\mathbf{v}}_j$*, com* $i \neq j$*, verificam* $\cos\theta_{ij} \leq 0$*. Para além disso, se* $\widetilde{\mathbf{v}}_k \in \mathcal{D}$*, com* $k \neq i$ *e* $k \neq j$*, então* $\cos\theta_{ij} \geq -\frac{1}{3}$*.*

Demonstração (do lema): A primeira afirmação é consequência imediata de (2.17). Para demonstrar a segunda afirmação, obtém-se $\min\cos\theta_{ij} = -\frac{f^2}{\min d_i d_j}$ em função de n_i e n_j, no domínio definido por $n_i^2 + n_j^2 \leq 1/2$. O mínimo ocorre em $|n_i| = |n_j| = \frac{1}{2}$ com valor $-1/3$. ■

Uma vez que $\frac{1}{2}\arccos(-\frac{1}{3}) = \arctan\sqrt{2} \approx 54.7°$, a existência de um ponto de fuga na região a sombreado da Figura 2.6 é uma simples consequência do Lema 2.5. Para se obter (2.16), considere-se uma orientação $\mathtt{O}$ e seja $\widetilde{\mathbf{v}}_i$ um ponto de fuga localizado nessa região a sombreado. A Proposição 2.3 garante então a existência de uma orientação equiprojectiva $\mathtt{O}'$ com pontos de fuga $\{\widetilde{\mathbf{v}}'_{\mathbf{x}}, \widetilde{\mathbf{v}}'_{\mathbf{y}}, \widetilde{\mathbf{v}}'_{\mathbf{z}}\}$ satisfazendo: (i) $\widetilde{\mathbf{v}}'_{\mathbf{z}} = \widetilde{\mathbf{v}}_i$, e (ii) $d'_x \leq d'_y$. De (2.10) – (2.13) temos, devido a (i), que $\beta' \in]-\pi/4, \pi/4]$ e $\gamma' \in \left]-\arctan\sqrt{2}, \arctan\sqrt{2}\right]$, e devido a (ii), que $\alpha' \in]-\pi/2, \pi/2]$, o que conclui a demonstração. ■

2.5 Modelo de pequenas rotações

Vamos agora assumir que a câmara se move e adquire uma sequência de imagens $\{\mathtt{I}_1, \ldots, \mathtt{I}_N\}$. Seja $\mathtt{O}_k = \mathtt{O}(\alpha_k, \beta_k, \gamma_k)$ a orientação na k-ésima imagem. A sequência de orientações $\{\mathtt{O}_1, \ldots, \mathtt{O}_N\}$ depende apenas da componente rotacional do movimento, *i.e.*, é independente da translação da câmara. Numa sequência de vídeo típica, a orientação da câmara varia suave e continuamente. Esta propriedade é formalizada introduzindo o modelo de **pequenas rotações**, que a seguir se descreve.

Definição 2.6 (modelo de pequenas rotações) *Seja* $\mathtt{R}_k(\rho_k, \mathbf{e}_k) \in \mathrm{SO}(3)$ *a componente rotacional do movimento da câmara entre a* $(k-1)$*-ésima e a* k*-ésima imagem, onde* ρ_k *e* $\mathbf{e}_k$ *representam o ângulo e o eixo de rotação, respectivamente. Independentemente de* $\mathbf{e}_k$*, diz-se que a câmara é consistente com o modelo das pequenas rotações* $\mu(\xi)$ sse *existe um "pequeno" ângulo fixo* ξ *tal que* $|\rho_k| \leq \xi$ *para qualquer* k*.*

Nas experiências realizadas (ver Secção 3.5), usou-se o modelo de pequenas rotações $\mu(5°)$, o que implica que para uma taxa de amostragem de 12.5 Hz o ângulo de rotação é sempre menor que 62.5° em cada segundo; trata-se de uma hipótese intuitivamente razoável.

A proposição seguinte expressa como as variações dos ângulos de compasso, elevação e torção entre imagens consecutivas podem ser limitadas devido ao modelo de pequenas rotações.

Proposição 2.7 *Se o movimento da câmara é consistente com o modelo de pequenas rotações* $\mu(\xi)$*, então, em qualquer instante* k*, aplicam-se as seguintes restrições:*

- *A variação do ângulo de elevação,* $\Delta\beta = \beta_k - \beta_{k-1}$*, satisfaz*

$$|\Delta\beta| \leq \xi. \tag{2.18}$$

- *A variação do ângulo de compasso,* $\Delta\alpha = \alpha_k - \alpha_{k-1}$*, satisfaz*

$$|\Delta\alpha| \leq \mathtt{a}_\xi(\beta_k, \beta_{k-1}) \equiv \begin{cases} \arccos\left(1 - \frac{\cos|\Delta\beta| - \cos\xi}{\cos\beta_{k-1}\cos\beta_k}\right) & \text{se } |\beta_{k-1} + \beta_k| \leq \pi - \xi \\ \frac{\pi}{2} & \text{caso contrário.} \end{cases} \tag{2.19}$$

Se $\mathtt{O}_{k-1}$ *se encontra na região* $\mathcal{R}$ *expressa em* (2.16)*, então, independentemente de* β_k *e* β_{k-1}*:*

$$|\Delta\alpha| \leq \max_{|\beta_{k-1}|\leq\frac{\pi}{4}} \mathtt{a}_\xi(\beta_k, \beta_{k-1}) = \arccos(2\cos\xi - 1). \tag{2.20}$$

- *A variação do ângulo de torção,* $\Delta\gamma = \gamma_k - \gamma_{k-1}$*, satisfaz*

$$|\Delta\gamma| \leq \mathtt{g}_\xi(\beta_{k-1}), \tag{2.21}$$

onde $\mathtt{g}_\xi$ *é uma função par crescente em* $[0, \frac{\pi}{2}]$ *com* $\mathtt{g}_\xi(0) = \xi$ *e* $\mathtt{g}_\xi(\frac{\pi}{2}) = \pi$*. Se* $\mathtt{O}_{k-1} \in \mathcal{R}$*, então* $|\beta_{k-1}| \leq \frac{\pi}{4}$ *e*

$$|\Delta\gamma| \leq \mathtt{g}_\xi\left(\frac{\pi}{4}\right), \tag{2.22}$$

A Fig. 2.7 mostra o gráfico da função $\mathtt{g}_\xi$ *no subdomínio* $[0, \frac{\pi}{4}]$*, para* $\xi = 5°$*; este valor de* ξ *conduz a* $|\Delta\gamma| \leq 7.08°$*.*

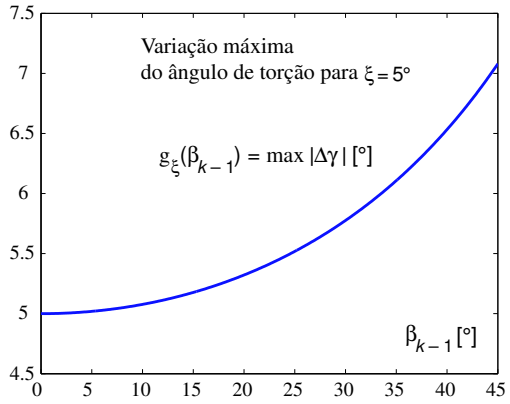

Figura 2.7: Variação máxima do ângulo de torção como função do ângulo de elevação inicial, utilizando um modelo de pequenas rotações $\mu(5°)$.

Demonstração: Considere-se $\mathtt{R}_k(\rho_k, \mathbf{e}_k)$ como a composição de duas rotações: $\mathtt{R}_{k_1}(\rho_{k_1}, \mathbf{e}_{k_1})$, que transforma $\mathbf{n}_{k-1}$ em $\mathbf{n}_k$ tendo como eixo de rotação $\mathbf{e}_{k_1} = \frac{(\mathbf{n}_{k-1} \times \mathbf{n}_k)}{||\mathbf{n}_{k-1} \times \mathbf{n}_k||}$, seguida de $\mathtt{R}_{k_2}(\rho_{k_2}, \mathbf{e}_{k_2})$ que "torce" a câmara em torno do eixo principal, *i.e.*, com $\mathbf{e}_{k_1} = \mathbf{n}_k$. Compondo estas duas rotações, e tendo em conta que $\mathbf{e}_{k_1} \perp \mathbf{e}_{k_2}$, obtém-se $\cos\frac{\rho_k}{2} = \cos\frac{\rho_{k_1}}{2}\cos\frac{\rho_{k_2}}{2}$. Portanto, a condição de pequenas rotações $|\rho_k| \leq \xi$ implica tanto $\cos\frac{\rho_{k_1}}{2} \geq \cos\frac{\xi}{2}$ como $\cos\frac{\rho_{k_2}}{2} \geq \cos\frac{\xi}{2}$, *i.e.*, $|\rho_{k_1}| \leq \xi$ e $|\rho_{k_2}| \leq \xi$. Uma vez que $\cos\rho_{k_1} = \mathbf{n}_k^T\mathbf{n}_{k-1}$, de (2.3) obtém-se

$$\cos\xi \leq \cos\rho_{k_1} = \cos\beta_k\cos\beta_{k-1}\cos\Delta\alpha + \sin\beta_k\sin\beta_{k-1} \leq \cos\Delta\beta, \tag{2.23}$$

o que basta para provar (2.18). De (2.23), obtém-se ainda

$$\cos\Delta\alpha \geq (\cos\xi - \sin\beta_k\sin\beta_{k-1})/(\cos\beta_k\cos\beta_{k-1}) \ ,$$

o que simplificando conduz a (2.19). Se $\mathtt{O}_{k-1} \in \mathcal{R}$, então de (2.16) vem $|\beta_k + \beta_{k+1}| \leq \pi/4 + \pi/4 + \xi \leq \pi - \xi$. O valor máximo de $\Delta\alpha$ ocorre para $\beta_k = \beta_{k-1} = \frac{\pi}{4}$, conduzindo assim a (2.20).

Quanto a $\Delta\gamma$, não é possível chegar a uma expressão simples para $\mathtt{g}_\xi(\beta_{k-1})$; no entanto, visto que ρ_k é função de β_{k-1}, β_k, $\Delta\alpha$ e $\Delta\gamma$, pode estudar-se $\mathtt{g}_\xi$ fazendo $\alpha_{k-1} = \gamma_{k-1} = 0$. Por simetria esférica tem-se que $\mathtt{g}_\xi$ é uma função par; por outro lado, uma simples inspecção geométrica mostra que $\mathtt{g}_\xi(\beta_{k-1})$ aumenta com $|\beta_{k-1}|$. Escrevendo $\mathtt{R}_k$ como a composição de três rotações – uma para o compasso, outra para a elevação e a última para a torção –, e recorrendo à fórmula para o produto de quaterniões, chega-se a

$$|\Delta\gamma| = 2\arccos\frac{AB - C\sqrt{B^2+C^2-A^2}}{B^2+C^2}, \tag{2.24}$$

onde $A = \cos\frac{\rho_k}{2}$, $B = \cos\frac{\Delta\alpha}{2}\cos\frac{\Delta\beta}{2}$ e $C = \sin\frac{\Delta\alpha}{2}\left(\cos\frac{\Delta\beta}{2}\sin\beta_k - \cos\beta_k\sin\Delta\beta\right)$. A maximização numérica de (2.24) com respeito a $\Delta\alpha$ e β_k (para $\rho_k = \xi$) aproxima $\mathtt{g}_\xi$. ■

Se a orientação $\mathtt{O}_{k-1}$ residir na região $\mathcal{R}$ definida em (2.16), o espaço de procura para $\mathtt{O}_k$ é significativamente reduzido graças às restrições impostas pela Proposição 2.7. Em particular, com $\xi = 5°$, tem-se nestas condições

$$|\Delta\alpha| \leq 7.08° \ , \ |\Delta\beta| \leq 5° \text{ e } |\Delta\gamma| \leq 7.08°. \tag{2.25}$$

Mesmo se $\mathtt{O}_{k-1} \notin \mathcal{R}$, a Proposição 2.4 garante a existência de uma orientação equiprojectiva $\mathtt{O}'_{k-1} \in \mathcal{E}(\mathtt{O}_{k-1})$ tal que $\mathtt{O}'_{k-1} \in \mathcal{R}$. Isto mostra como se pode utilizar conjuntamente o modelo de pequenas rotações e as orientações equiprojectivas para reduzir significativamente o espaço de procura.

2.6 Generalização para outros mundos estruturados

Muitos dos resultados obtidos nas secções anteriores, em que sempre se teve em mente o modelo de mundo de Manhattan no espaço tridimensional, são na verdade generalizáveis a outros modelos de mundos, não necessariamente tridimensionais, podendo ser empregues para estimar a orientação da câmara em cenários de diferente estrutura ou para aplicar a problemas de dimensão superior, não necessariamente provenientes da área de Visão por Computador.

De facto, o mundo de Manhattan pode ser visto como um caso particular de mundo estruturado, tridimensional, onde a maior parte da informação relevante está "alinhada" segundo três direcções ortogonais, $\mathbf{x}$, $\mathbf{y}$ e $\mathbf{z}$. Ora, este raciocínio pode generalizar-se a uma classe de mundos estruturados, n-dimensionais, onde estão em jogo m direcções não necessariamente ortogonais, $\mathbf{x}_1, \cdots, \mathbf{x}_m \in \mathbb{R}^n$, com $m \geq n$. Sendo $(\mathbf{a}_1, \cdots, \mathbf{a}_n)$ o referencial do mundo e $(\mathbf{b}_1, \cdots, \mathbf{b}_n)$ um outro referencial, designado por referencial da **câmara**, onde $\{\mathbf{a}_i\}_{i=0}^{i=n}$ e $\{\mathbf{b}_i\}_{i=0}^{i=n}$ são duas bases ortonormadas de $\mathbb{R}^n$, os dois referenciais relacionam-se por

$$(\mathbf{b}_1, \cdots, \mathbf{b}_n) = (\mathbf{a}_1, \cdots, \mathbf{a}_n)\mathtt{O}, \tag{2.26}$$

onde $\mathtt{O}$ é uma matriz de transformação de coordenadas a que se pode chamar **orientação** e que pertence ao grupo das matrizes ortogonais, $\mathrm{O}(n)$; caso a natureza do problema exija que esta transformação preserve a orientação dos eixos, *i.e.*, seja uma *rotação*, tem-se $\mathtt{O} \in \mathrm{SO}(n)$, onde $\mathrm{SO}(n)$ é o sub-grupo de $\mathrm{O}(n)$ formado pelas matrizes ortogonais com determinante positivo. Em ambos os casos, utilizando o sistema de coordenadas definido pelo referencial canónico do mundo, *i.e.*, $(\mathbf{a}_1, \cdots, \mathbf{a}_n) = ([0, 0, \cdots, 1]^T, \cdots, [1, 0, \cdots, 0]^T)$, vem $\mathtt{O} = [\mathbf{b}_1, \cdots, \mathbf{b}_n]$.

No plano da imagem (um sub-espaço de dimensão $n-1$), pode associar-se a cada um dos pontos de fuga segundo as direcções $\mathbf{x}_1, \cdots, \mathbf{x}_m$ respectivamente os vectores de

coordenadas homogéneas $\mathbf{v}_1, \cdots, \mathbf{v}_m \in \mathbb{R}^n$; introduzindo as matrizes $n \times m$ definidas como $\mathtt{V} = [\mathbf{v}_1, \cdots, \mathbf{v}_m]$ e $\mathtt{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_m]$, tem-se

$$\mathtt{V} = \mathtt{K}\mathtt{O}^T\mathtt{X}, \tag{2.27}$$

onde $\mathtt{K}$ é uma matriz invertível do tipo da de (2.15). Aliás, para um mundo de Manhattan tridimensional tem-se $n = 3$ e $\mathtt{X}$ igual à matriz identidade 3×3, obtendo-se precisamente (2.15).

O objectivo é estimar a orientação supondo que se dispõe de um mecanismo capaz de "sondar" o mundo e obter um conjunto de estimativas para a localização dos pontos de fuga, $\mathcal{V} = \{\widehat{\mathbf{v}}_1, \cdots, \widehat{\mathbf{v}}_m\}$. Define-se a relação de equiprojectividade em função das direcções $\mathbf{x}_1, \cdots, \mathbf{x}_m$ da seguinte forma: duas orientações $\mathtt{O}$ e $\mathtt{O}'$ são **equiprojectivas** *sse* os pontos de fuga segundo estas direcções tiverem, no seu conjunto, a mesma localização, ou seja, usando (2.27), *sse* existirem uma matriz de permutação $\mathtt{P}$ e uma matriz diagonal Λ tais que

$$\mathtt{K}\mathtt{O}^T\mathtt{X} = \mathtt{K}\mathtt{O}'^T\mathtt{X}\Lambda\mathtt{P}. \tag{2.28}$$

Como $\mathtt{O}, \mathtt{O}' \in \mathrm{SO}(n)$, deve ter-se $\mathtt{O}' = \mathtt{M}\mathtt{O}$ com certo $\mathtt{M} \in \mathrm{SO}(n)$; multiplicando ambos os termos de (2.28) à esquerda por $\mathtt{O}'\mathtt{K}^{-1}$ e à direita por $\mathtt{X}^T(\mathtt{X}\mathtt{X}^T)^{-1}\mathtt{O}$, obtém-se por fim:

$$\mathtt{M} = \mathtt{X}\Lambda\mathtt{P}\mathtt{X}^T(\mathtt{X}\mathtt{X}^T)^{-1} \quad \wedge \quad \mathtt{M} \in \mathrm{SO}(n), \tag{2.29}$$

o que permite caracterizar a classe de equivalência das orientações equiprojectivas para qualquer mundo nesta classe de mundos estruturados. Note-se que alargando o domínio da orientação para $\mathrm{O}(n)$ (*i.e.*, retirando a exigência de que $\mathtt{O}$ seja uma rotação) obtém-se exactamente os mesmos resultados substituindo $\mathrm{SO}(n)$ por $\mathrm{O}(n)$.

Exemplo 2.8 *Tome-se como exemplo um mundo tridimensional ($n = 3$) constituído por edifícios que são prismas de base hexagonal,* i.e.*, com direcções definidas pelos vectores normalizados de $\mathbb{R}^3$: $\mathbf{x}_1 = [0, 0, 1]^T$, $\mathbf{x}_2 = [0, 1, 0]^T$, $\mathbf{x}_3 = [\sqrt{3}/2, 1/2, 0]^T$ e $\mathbf{x}_4 = [-\sqrt{3}/2, 1/2, 0]^T$. Fazendo $\mathtt{X} = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4]$ e tendo em conta que a multiplicação por uma matriz de* $\mathrm{SO}(3)$ *preserva a norma, o que implica $\Lambda = \mathrm{diag}(\pm 1, \pm 1, \pm 1)$ obtém-se, utilizando* (2.29)*, o seguinte conjunto de soluções para $\mathtt{M}$:*

$$\mathtt{M} \in \left\{ \begin{bmatrix} \pm 1 & 0 & 0 \\ 0 & \pm 1 & 0 \\ 0 & 0 & \pm 1 \end{bmatrix}, \begin{bmatrix} k_1\frac{1}{2} & k_3\frac{\sqrt{3}}{2} & 0 \\ k_2\frac{\sqrt{3}}{2} & k_4\frac{1}{2} & 0 \\ 0 & 0 & \pm 1 \end{bmatrix} \right\} \cap \mathrm{SO}(3), \tag{2.30}$$

com $|k_1| = |k_2| = |k_3| = |k_4| = 1$ e $k_1k_2k_3k_4 = -1$. Cada classe de equivalência formada pelas orientações equiprojectivas tem portanto, para este modelo de mundo, $\frac{1}{2}(2^3 + 2^4) = 12$ elementos, o que coincide com a ordem do grupo de simetrias do prisma hexagonal.

Exemplo 2.9 (mundo de Manhattan n-dimensional) *Os mundos de Manhattan n-dimensionais são um caso particular desta classe de mundos estruturados, aqueles para os quais se tem como "directrizes" n vectores ortonormados de $\mathbb{R}^n$,* i.e.*, para os quais as direcções $\mathbf{x}_1, \cdots, \mathbf{x}_n$ satisfazem, para quaisquer i, j, $\mathbf{x}_i^T\mathbf{x}_j = \delta_{ij}$, onde δ_{ij} representa o delta de Kronecker e portanto a matriz $\mathtt{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_n]$ é uma matriz de permutação, fazendo com que* (2.29) *se reduza a $\mathtt{M} = \Lambda\mathtt{P} \ \wedge \ \mathtt{M} \in \mathrm{SO}(n)$, o que equivale a $\mathtt{M} \in \mathrm{SP}^+(n)$.*

Assim, as classes de equivalência de orientações equiprojectivas, num mundo de Manhattan n-dimensional, são isomorfas ao grupo $\mathrm{SP}^+(n)$ *das* matrizes de permutação

com sinal cujo determinante é positivo, *tendo por isso* $\frac{1}{2}2^n n!$ *elementos (se se alargar o domínio da orientação para* $\mathrm{O}(n)$*, tem-se* $\mathtt{M} \in \mathrm{SP}(n)$*, onde* $\mathrm{SP}(n)$ *é o grupo das* matrizes de permutação com sinal, *que contém* $\mathrm{SP}^+(n)$ *como subgrupo, e tem* $2^n n!$ *elementos).*

Em $\mathbb{R}^n$ *a orientação tem* $(n-1)+(n-2)+...+1 = n(n-1)/2$ *graus de liberdade. Assumindo que o mecanismo consegue obter cada uma das estimativas dos pontos de fuga* $\{\widehat{\mathbf{v}}_1, ..., \widehat{\mathbf{v}}_n\}$ *independentemente, é possível obter estimativas de cada linha da matriz de orientação ao mesmo tempo que se estima cada um dos pontos de fuga, através de um procedimento recursivo com* $n-1$ *passos: no primeiro passo, estima-se a localização do ponto de fuga* $\widehat{\mathbf{v}}_1$ *no espaço n-dimensional – isto permite obter o sub-espaço de dimensão* $n-1$ *que contém todos os outros pontos de fuga; no segundo passo localiza-se* $\widehat{\mathbf{v}}_2$ *nesse sub-espaço obtendo-se um sub-espaço de dimensão* $n-2$ *que contém os restantes; e assim por diante até ao último passo onde se estima* $\widehat{\mathbf{v}}_{n-1}$ *e* $\widehat{\mathbf{v}}_n$ *no espaço unidimensional. Em cada um destes sub-espaços de dimensão p é possível obter regiões* $\mathcal{R}_p$ *semelhantes à da Proposição 2.4 onde se garante a existência de um ponto de fuga* $\widehat{\mathbf{v}}_{n-p+1}$*.*

Capítulo 3

Estimação sequencial da orientação

3.1 Critério de estimação

Para estimar a sequência de orientações da câmara, $\{\mathtt{O}_1, \ldots, \mathtt{O}_N\}$, a partir da sequência de imagens observadas, $\{\mathtt{I}_1, \ldots, \mathtt{I}_N\}$, propõe-se uma abordagem probabilística de estimação sequencial, fazendo uso dos modelos de mundo de Manhattan (ver Secção 2.1) e de pequenas rotações (ver Secção 2.5).

Como vimos atrás, o modelo de mundo de Manhattan assume que os contornos presentes na imagem $\mathtt{I}_k$ estão, na sua maioria, "alinhados" com os eixos de Manhattan $\mathbf{x}$, $\mathbf{y}$ e $\mathbf{z}$. Ora, é sabido que o *gradiente* de intensidade é uma medida da magnitude e direcção destes contornos, facto que é utilizado em numerosas técnicas de processamento de imagem. Por conseguinte, pode considerar-se que a estatística do gradiente de intensidade de cada imagem, $\nabla\mathtt{I}_k$, transporta informação sobre a correspondente orientação da câmara através de uma função de verosimilhança $P(\nabla\mathtt{I}_k|\mathtt{O}_k)$ (como apontado em [9, 10]). Neste trabalho, esta consideração é tida em conta e adaptada a um modelo de estimação sequencial, utilizando-se um critério de *maximum a posteriori* (MAP) onde a estimativa da orientação para cada instante k é dada por

$$\widehat{\mathtt{O}}_k = \arg\max_{\mathtt{O}_k} \left\{ P(\mathtt{O}_k|\nabla\mathtt{I}_k, \widehat{\mathtt{O}}_{k-1}, ..., \widehat{\mathtt{O}}_1) \right\} = \tag{3.1}$$

$$= \arg\max_{\mathtt{O}_k} \left\{ P(\nabla\mathtt{I}_k|\mathtt{O}_k) \frac{P(\mathtt{O}_k|\widehat{\mathtt{O}}_{k-1}, ..., \widehat{\mathtt{O}}_1)}{P(\nabla\mathtt{I}_k)} \right\} = \tag{3.2}$$

$$= \arg\max_{\mathtt{O}_k} \left\{ P(\nabla\mathtt{I}_k|\mathtt{O}_k) P(\mathtt{O}_k|\widehat{\mathtt{O}}_{k-1}, ..., \widehat{\mathtt{O}}_1) \right\}, \tag{3.3}$$

onde de (3.1) para (3.2) se assume que cada $\nabla\mathtt{I}_k$ depende apenas da orientação $\mathtt{O}_k$ nesse mesmo instante, e não de toda a "história" anterior. Uma estimação sequencial completamente bayesiana requeriria métodos Monte Carlo computacionalmente muito pesados (veja-se [16], [17]). Como alternativa, considera-se este processo como um *modelo de Markov oculto* (HMM[1]) que percorre estados $\{\mathtt{O}_k\}$ emitindo símbolos $\{\mathtt{I}_k\}$, e em que cada estado só depende do estado anterior, *i.e.*, $P(\mathtt{O}_k|\widehat{\mathtt{O}}_{k-1}, ..., \widehat{\mathtt{O}}_1) = P(\mathtt{O}_k|\widehat{\mathtt{O}}_{k-1})$. De acordo com esta formalização, e aplicando logaritmos a (3.3), vem então:

$$\widehat{\mathtt{O}}_k = \arg\max_{\mathtt{O}_k} \left\{ \log P(\nabla\mathtt{I}_k|\mathtt{O}_k) + \log P(\mathtt{O}_k|\widehat{\mathtt{O}}_{k-1}) \right\} \tag{3.4}$$

[1] Do inglês *hidden Markov model*.

Analisando (3.4), pode encarar-se a probabilidade *a priori* $P(\mathtt{O}_k|\widehat{\mathtt{O}}_{k-1})$ como uma forma de penalizar grandes variações entre estimações consecutivas da orientação. Experimentalmente verifica-se que este critério simplificado conduz a bons resultados e, explorando as potencialidades da "equiprojectividade" e do modelo de pequenas rotações introduzidos atrás, pode ser implementado com vista a funcionar em tempo real. Para a primeira imagem da sequência, $k = 1$, utiliza-se

$$\widehat{\mathtt{O}}_1 = \arg\max_{\mathtt{O}_1} \{\log P(\nabla \mathtt{I}_1|\mathtt{O}_1)\}\,, \tag{3.5}$$

que se obtém de (3.4) suprimindo o segundo termo, por se considerar, na ausência de estimativas anteriores, que todas as orientações são equiprováveis.

3.2 Função de verosimilhança

Nesta secção, para simplificar a notação, omite-se o índice temporal k e deriva-se a função de verosimilhança $P(\nabla \mathtt{I}|\mathtt{O})$ para uma imagem genérica. O método utilizado para obter a função de verosimilhança é inspirado em [9, 10]; segue-se uma breve descrição.

3.2.1 Gradiente de intensidade

Em todas as imagens considera-se apenas a informação monocromática, *i.e.*, a quantidade total de luz presente em cada *pixel*. Existem muitos métodos para obter estimativas do gradiente de intensidade (ver por exemplo [18, 19, 20, 21, 22] ou uma análise comparativa de vários destes métodos em [23]). O que aqui se utiliza, pela sua simplicidade e rapidez computacional, é o método da *convolução com máscaras de Sobel*. Dada uma imagem $\mathtt{I}$, elimina-se o ruído de alta frequência convoluindo-a com um filtro gaussiano – *i.e.*, uma máscara construída com a função gaussiana, $G(x, y) = 1/(2\pi\sigma^2)\exp[-(x^2 + y^2)/(2\sigma^2)]$ – obtendo-se uma imagem *suavizada* $\mathtt{I}_s$; de seguida, o gradiente $\nabla \mathtt{I} = (\nabla \mathtt{I}_x, \nabla \mathtt{I}_y)$ é estimado através das convoluções $\nabla \mathtt{I}_x = \mathtt{I}_s * \mathtt{S}_x$ e $\nabla \mathtt{I}_y = \mathtt{I}_s * \mathtt{S}_y$, onde

$$\mathtt{S}_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \text{ e } \mathtt{S}_y = -\mathtt{S}_x^T$$

são as máscaras de Sobel.

Por razões de desempenho computacional, são rejeitados todos os *pixels* cuja informação se considera inútil; tais são:

- Os *pixels* próximos de contornos mas para os quais existem vizinhos cujo gradiente é, em módulo, superior. Este critério de rejeição é também utilizado no algoritmo de detecção de contornos de Canny (ver [20]) com o intuito de obter contornos com a espessura média de 1 pixel. Designa-se na literatura anglo-saxónica por *non-maxima supression*.

- Os *pixels* cujo gradiente, em módulo, está abaixo de um determinado limiar, o que garante que não pertencem a contornos; este critério designa-se por *limiarização*.

Nas experiências efectuadas, a aplicação destes critérios de rejeição permitiu eliminar cerca de 80% dos *pixels* da imagem.

3.2.2 Classes de *pixel*

Cada *pixel* da imagem, $\mathbf{u} = [u, v, 1]^T$, que não tenha sido rejeitado na etapa anterior, pode ser associado a uma *classe* $m_{\mathbf{u}} \in \{1, 2, 3, 4, 5\}$. As classes são definidas da seguinte forma: as classes 1, 2 e 3 referem-se a *pixels* de contornos alinhados com um dos eixos de Manhattan, **x**, **y** e **z** respectivamente. A classe 4 abrange os *pixels* que pertencem a contornos não alinhados com qualquer daqueles eixos. Finalmente, a classe 5 inclui os *pixels* que não pertencem a contornos. Estas classes distribuem-se com probabilidades *a priori* $\{P_m(m_{\mathbf{u}})\}$, obtidas *off-line* através de um detector de contornos binário e fixando heuristicamente $P_m(1) = P_m(2) = P_m(3) = 0,5 \times P_m(4)$. A Figura 3.1 e a Tabela 3.1 ilustram cada uma das classes de *pixel*.

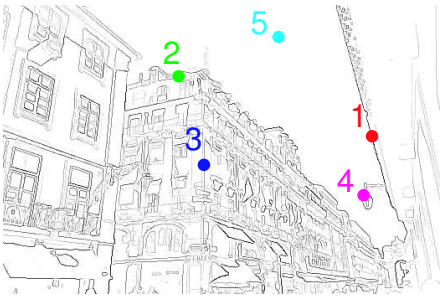

Figura 3.1: Classes de *pixel* para uma fotografia da Rua Augusta, na Baixa Pombalina, em Lisboa.

$m_{\mathbf{u}}$	Descrição	$P_m(m_{\mathbf{u}})$
1	*Pixel* de um contorno alinhado com **x**	0,138
2	*Pixel* de um contorno alinhado com **y**	0,138
3	*Pixel* de um contorno alinhado com **z**	0,138
4	*Pixel* de um contorno não alinhado com **x**, **y** ou **z**	0,276
5	*Pixel* não pertencente a qualquer contorno	0,309

Tabela 3.1: Classes de *pixel*.

3.2.3 Funções de probabilidade da magnitude do gradiente

Seja $\mathbf{E_u} = (E_{\mathbf{u}}, \phi_{\mathbf{u}})$ a representação polar do gradiente da intensidade da imagem $\nabla \mathtt{I}$ no *pixel* $\mathbf{u}$, onde $E_{\mathbf{u}} = \mathtt{Q}[(\nabla \mathtt{I}_x^2(\mathbf{u}) + \nabla \mathtt{I}_y^2(\mathbf{u}))^{1/2}] \in \{1, ..., N\}$ é a magnitude do gradiente quantificada por uma função $\mathtt{Q}$ com N níveis de quantificação, e $\phi_{\mathbf{u}} = \arctan(\nabla \mathtt{I}_y(\mathbf{u})/\nabla \mathtt{I}_x(\mathbf{u}))$ é a direcção do gradiente. A magnitude e a direcção do gradiente são condicionalmente independentes, dada a classe do pixel. Naturalmente, a magnitude do gradiente é também condicionalmente independente da orientação da câmara e da localização do pixel. Logo,

$$P(\mathbf{E_u}|m_{\mathbf{u}}, \mathtt{O}, \mathbf{u}) = P(E_{\mathbf{u}}|m_{\mathbf{u}})\, P(\phi_{\mathbf{u}}|m_{\mathbf{u}}, \mathtt{O}, \mathbf{u}), \tag{3.6}$$

onde

$$P(E_{\mathbf{u}}|m_{\mathbf{u}}) = \begin{cases} P_{\text{on}}(E_{\mathbf{u}}), & \text{se } m_{\mathbf{u}} \neq 5 \\ P_{\text{off}}(E_{\mathbf{u}}), & \text{se } m_{\mathbf{u}} = 5, \end{cases} \tag{3.7}$$

e $P_{\text{on}}(E_{\mathbf{u}})$ e $P_{\text{off}}(E_{\mathbf{u}})$ são as funções de probabilidade da magnitude do gradiente quantificado, *condicionadas* ao evento de o pixel **u** pertencer (*on*) ou não (*off*) a um contorno,

respectivamente. Estas probabilidades são, também elas, obtidas através de um processo de treino que decorre *off-line*, em que se utiliza um detector de contornos binário. Os seus valores (para uma quantificação logarítmica e $N = 20$) encontram-se representados na Figura 3.2.

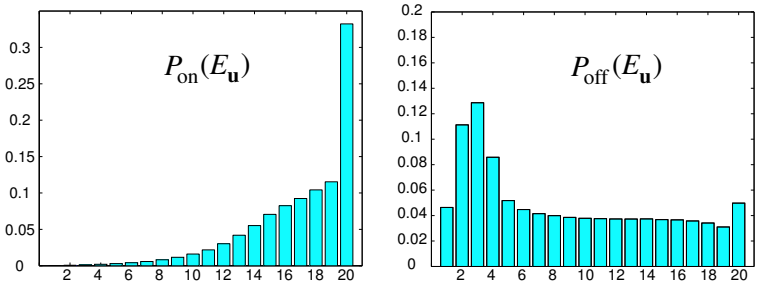

Figura 3.2: Histogramas das funções P_{on} e P_{off}. No caso de P_{on}, como seria de esperar, a probabilidade de se ter valores elevados para a magnitude do gradiente em *pixels* de contornos é muito maior do que a probabilidade de se ter valores reduzidos.

3.2.4 Funções de probabilidade da direcção do gradiente

Sejam $\theta_{\mathbf{x}}(\mathsf{O}, \mathbf{u})$, $\theta_{\mathbf{y}}(\mathsf{O}, \mathbf{u})$ e $\theta_{\mathbf{z}}(\mathsf{O}, \mathbf{u})$ as direcções do gradiente idealmente observado no pixel $\mathbf{u}$ se $m_{\mathbf{u}} = 1$, 2 e 3, respectivamente. Estas direcções obtêm-se directamente dos pontos de fuga, dados em função de O por (2.14). A função de probabilidade da direcção do gradiente é

$$P(\phi_{\mathbf{u}}|m_{\mathbf{u}}, \mathsf{O}, \mathbf{u}) = \begin{cases} P_{ang}\left(\phi_{\mathbf{u}} - \theta_{\mathbf{x}}(\mathsf{O}, \mathbf{u})\right) & \text{se } m_{\mathbf{u}} = 1 \\ P_{ang}\left(\phi_{\mathbf{u}} - \theta_{\mathbf{y}}(\mathsf{O}, \mathbf{u})\right) & \text{se } m_{\mathbf{u}} = 2 \\ P_{ang}\left(\phi_{\mathbf{u}} - \theta_{\mathbf{z}}(\mathsf{O}, \mathbf{u})\right) & \text{se } m_{\mathbf{u}} = 3 \\ \mathsf{U}\left(\phi_{\mathbf{u}}\right) & \text{se } m_{\mathbf{u}} \in \{4, 5\}, \end{cases} \tag{3.8}$$

onde $\mathsf{U}(\cdot)$ é a função densidade de probabilidade uniforme em $]-\frac{\pi}{2}, \frac{\pi}{2}]$ e P_{ang} é modelizada como uma função "caixa", *i.e.*,

$$P_{ang}(t) = \begin{cases} (1-\epsilon)/(2\tau) & \text{se } t \in [-\tau, \tau] \\ \epsilon/(\pi - 2\tau) & \text{se } t \in]-\pi/2, -\tau[\cup]\tau, \pi/2]. \end{cases}$$

Nas experiências efectuadas, estes parâmetros foram afinados para $\epsilon = 0.1$ e $\tau = 4^\circ$ (ver Figura 3.3).

3.2.5 Função de verosimilhança

Finalmente, a verosimilhança conjunta é obtida de (3.6) marginalizando (*i.e.*, somando) sobre todos os possíveis modelos em cada pixel e assumindo independência entre *pixels* diferentes:

$$P(\nabla\mathtt{I}|\mathsf{O}) = P(\{\mathbf{E}_{\mathbf{u}}\}|\mathsf{O}) = \prod_{\mathbf{u}} \sum_{m_{\mathbf{u}}=1}^{5} P(E_{\mathbf{u}}|m_{\mathbf{u}})\, P(\phi_{\mathbf{u}}|m_{\mathbf{u}}, \mathsf{O}, \mathbf{u})\, P(m_{\mathbf{u}})\,. \tag{3.9}$$

Aplicando logaritmos,

$$\log P(\nabla\mathtt{I}|\mathsf{O}) = \sum_{\mathbf{u}} \log \left\{ \sum_{m_{\mathbf{u}}=1}^{5} P(E_{\mathbf{u}}|m_{\mathbf{u}})\, P(\phi_{\mathbf{u}}|m_{\mathbf{u}}, \mathsf{O}, \mathbf{u})\, P(m_{\mathbf{u}}) \right\}. \tag{3.10}$$

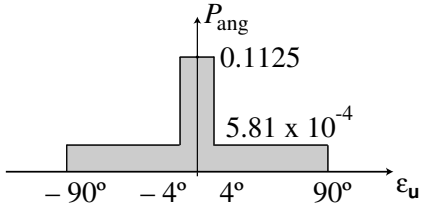

Figura 3.3: Gráfico da função P_{ang}, a densidade de probabilidade de erro na estimação da direcção do contorno a partir da direcção do gradiente. Naturalmente a densidade é maior para erros pequenos. Por simplicidade e vantagem computacional, é modelizada como uma "caixa", onde τ controla a precisão da estimativa e ϵ é a probabilidade de se cometer um erro superior a τ.

3.3 Procedimento de localização das estimativas

Como se descreveu atrás, a maximização expressa em (3.4) permite estimar a orientação da câmara para cada imagem da sequência de vídeo. O segundo termo desta expressão, $\log P(\mathtt{O}_k|\widehat{\mathtt{O}}_{k-1})$, pode ser descrito através de uma função definida *a priori*; o primeiro termo é a função de verosimilhança descrita em (3.10). Esta maximização é um problema de optimização tridimensional em ordem a α, β e γ, que pode ser resolvida por um algoritmo de procura exaustiva nos intervalos de variação destes ângulos, com complexidade $\mathcal{O}(N^3)$, onde N descreve a frequência de amostragem daqueles intervalos.

Neste trabalho, propõe-se uma solução aproximada que separa o problema em dois passos mais simples: uma optimização bidimensional em ordem a β e γ, seguida por uma procura unidimensional em ordem a α. A complexidade do algoritmo de procura exaustiva correspondente a estas duas optimizações sucessivas é $\mathcal{O}(N^2 + N) = \mathcal{O}(N^2)$. Esta aproximação advém do facto de o ponto de fuga $\mathbf{v_z}$ ser independente do ângulo de compasso α, como é claro de (2.12).

3.3.1 1.° passo: Estimação de β e γ

Dada a k-ésima imagem $\mathtt{I}_k$, com $k > 1$, procede-se, como primeiro passo, à estimação de β_k e γ_k, de acordo com

$$\left(\widehat{\beta}_k, \widehat{\gamma}_k\right) = \arg\max_{\beta,\gamma} \left\{ \log P(\nabla \mathtt{I}_k|\beta,\gamma) + \log P(\beta,\gamma|\widehat{\beta}_{k-1}, \widehat{\gamma}_{k-1}) \right\}, \tag{3.11}$$

onde a verosimilhança $\log P(\nabla \mathtt{I}_k|\beta,\gamma)$ é uma versão de (3.10) que apenas modeliza a informação de direcção dos contornos que são consistentes com o eixo $\mathbf{z}$. Mais especificamente, em vez de (3.8), utiliza-se aqui

$$P(\phi_{\mathbf{u}}|m_{\mathbf{u}},\beta,\gamma,\mathbf{u}) = \begin{cases} P_{ang}\left(\phi_{\mathbf{u}} - \theta_{\mathbf{z}}(\beta,\gamma,\mathbf{u})\right) & \text{se } m_{\mathbf{u}} = 3 \\ \mathsf{U}\left(\phi_{\mathbf{u}}\right) & \text{se } m_{\mathbf{u}} \in \{1,2,4,5\}. \end{cases} \tag{3.12}$$

Note-se que a utilização de uma distribuição uniforme é apenas uma forma de ignorar a informação de direcção do gradiente vinda de todos os *pixels* excepto daqueles que estão associados com o eixo $\mathbf{z}$ (*i.e.*, com classe $m_{\mathbf{u}} = 3$) durante a estimação de β_k e γ_k; tal não significa que aquelas direcções sejam de facto uniformemente distribuídas.

O segundo termo de (3.11) faz intervir $P(\beta,\gamma|\widehat{\beta}_{k-1}, \widehat{\gamma}_{k-1})$, uma função densidade de probabilidade (f.d.p.) em duas variáveis que se modeliza como sendo gaussiana e

"truncada" na região de interesse, *i.e.*:

$$P(\beta,\gamma|\widehat{\beta}_{k-1},\widehat{\gamma}_{k-1}) = \begin{cases} \lambda \mathsf{G}(\beta,\gamma), & \text{se } (\beta,\gamma) \in I_\beta \times I_\gamma \\ 0 & \text{c.c.} \end{cases}, \tag{3.13}$$

onde:

- $\mathsf{G}(\beta,\gamma)$ é a f.d.p. gaussiana em duas variáveis com valores médios $(\mu_\beta,\mu_\gamma) = (\widehat{\beta}_{k-1},\widehat{\gamma}_{k-1})$ e variâncias $(\sigma_\beta^2,\sigma_\gamma^2)$;
- $I_\beta =]\widehat{\beta}_{k-1}-\xi, \widehat{\beta}_{k-1}+\xi]$ é o intervalo de variação do ângulo de elevação nas condições do modelo de pequenas rotações $\mu(\xi)$, expresso em (2.18);
- $I_\gamma =]\widehat{\gamma}_{k-1}-\mathsf{g}_\xi(\widehat{\beta}_{k-1}), \widehat{\gamma}_{k-1}+\mathsf{g}_\xi(\widehat{\beta}_{k-1})]$ é o intervalo de variação do ângulo de torção nas mesmas condições, expresso em (2.21);
- λ é determinado de modo a ter-se $\int_{I_\beta \times I_\gamma} P(\beta,\gamma|\widehat{\beta}_{k-1},\widehat{\gamma}_{k-1}) d\beta d\gamma = 1$.

A função (3.13) é definida *a priori* e formaliza o modelo de pequenas rotações, bem como a suavidade na variação de β_k e γ_k. As variâncias σ_β^2 e σ_γ^2 afinam o compromisso entre a suavidade da sequência estimada de ângulos e a precisão dessas estimativas.

Na primeira imagem da sequência, $k = 1$, define-se *a priori* $P(\beta,\gamma)$ como uniforme (*i.e.*, constante) no domínio dado por (2.16), $I_\beta \times I_\gamma =]-45°, 45°] \times]-54.7°, 54.7°]$, pelo que em vez de (3.11) usa-se:

$$\left(\widehat{\beta}_1, \widehat{\gamma}_1\right) = \arg\max_{\beta,\gamma} \log P(\nabla \mathtt{I}_1|\beta,\gamma). \tag{3.14}$$

3.3.2 2.° passo: Estimação de α

Dadas as estimativas $\widehat{\beta}_k$ e $\widehat{\gamma}_k$, pode estimar-se o ângulo de compasso α_k utilizando

$$\widehat{\alpha}_k = \arg\max_{\alpha} \left\{ \log P(\nabla \mathtt{I}_k|\alpha,\widehat{\beta}_k,\widehat{\gamma}_k) + \log P(\alpha|\widehat{\alpha}_{k-1},\widehat{\beta}_{k-1},\widehat{\beta}_k) \right\}, \tag{3.15}$$

onde a verosimilhança $\log P(\nabla \mathtt{I}_k|\alpha,\widehat{\beta}_k,\widehat{\gamma}_k)$ é dada directamente por (3.10) e a função *a priori* $P(\alpha|\widehat{\alpha}_{k-1},\widehat{\beta}_{k-1},\widehat{\beta}_k)$ é também uma f.d.p. gaussiana "truncada":

$$P(\alpha|\widehat{\alpha}_{k-1},\widehat{\beta}_{k-1},\widehat{\beta}_k) = \begin{cases} \lambda \mathsf{G}(\alpha), & \text{se } \alpha \in I_\alpha \\ 0 & \text{c.c.} \end{cases}, \tag{3.16}$$

onde:

- $\mathsf{G}(\alpha)$ é a f.d.p. gaussiana com valor médio $\mu_\alpha = \widehat{\alpha}_{k-1}$ e variância σ_α^2;
- $I_\alpha =]\widehat{\alpha}_{k-1}-\mathsf{a}_\xi(\widehat{\beta}_k,\widehat{\beta}_{k-1}), \widehat{\alpha}_{k-1}+\mathsf{a}_\xi(\widehat{\beta}_k,\widehat{\beta}_{k-1})]$ é o intervalo de variação do ângulo de compasso nas condições do modelo de pequenas rotações $\mu(\xi)$, expresso em (2.19);
- λ é determinado de modo a ter-se $\int_{I_\alpha} P(\alpha|\widehat{\alpha}_{k-1},\widehat{\beta}_{k-1},\widehat{\beta}_k) d\alpha = 1$.

Também aqui a variância σ_α^2 afina o compromisso entre a suavidade da sequência estimada de ângulos e a precisão dessas estimativas.

Para a primeira imagem, $k = 1$, define-se $P(\alpha)$ como uniforme em todo o domínio dado por (2.16), $I_\alpha =]-45°, 45°]$, usando-se portanto em vez de (3.15)

$$\widehat{\alpha}_1 = \arg\max_{\alpha} \log P(\nabla \mathtt{I}_1|\alpha,\widehat{\beta}_1,\widehat{\gamma}_1). \tag{3.17}$$

3.3.3 Localização das estimativas

Com base nas considerações estabelecidas acima, o algoritmo de estimação da orientação pode ser formalizado como se segue.

Algoritmo 3.1 *Objectivo: Dada uma sequência de imagens* $\{\mathtt{I}_1, \ldots, \mathtt{I}_N\}$ *de um mundo de Manhattan, estimar a correspondente sequência de orientações da câmara,* $\{\mathtt{O}_1, \ldots, \mathtt{O}_N\}$.

- *Para cada imagem* $\mathtt{I}_k$ *da sequência de vídeo:*
  1. *Calcula-se o gradiente* $\nabla \mathtt{I}_k$*;*
  2. *Através de* non maxima supression *e limiarização, selecciona-se os* pixels $\{\mathbf{u}\}$ *que contêm informação relevante, obtendo-se* $\{\mathbf{E_u}\} = \{(E_\mathbf{u}, \phi_\mathbf{u})\}$*;*
  3. *Estima-se* $\widehat{\beta}_k$ *e* $\widehat{\gamma}_k$ *pelo método de procura exaustiva. Se* $k = 1$*, usa-se* (3.14) *com o espaço de procura* $I_\beta \times I_\gamma =]{-}45°, 45°] \times]{-}54.7°, 54.7°]$*, graças a* (2.16)*; se* $k > 1$*, usa-se* (3.11) *e o espaço de procura tem em conta as estimativas anteriores, sendo para um modelo de pequenas rotações* $\mu(5°)$*:* $I_\beta \times I_\gamma =]\widehat{\beta}_{k-1} - 5°, \widehat{\beta}_{k-1} + 5°] \times]\widehat{\gamma}_{k-1} - 7.08°, \widehat{\gamma}_{k-1} + 7.08°]$*, conforme* (2.25)*.*
  4. *Estima-se* $\widehat{\alpha}_k$ *pelo método de procura exaustiva. Se* $k = 1$*, usa-se* (3.17) *com o espaço de procura* $I_\alpha =]{-}45°, 45°]$*; se* $k > 1$*, usa-se* (3.15) *e o espaço de procura tem em conta as estimativas anteriores, sendo para o mesmo modelo* $\mu(5°)$*:* $I_\alpha =]\widehat{\alpha}_{k-1} - 7.08°, \widehat{\alpha}_{k-1} + 7.08°]$*.*
  5. *Obtém-se a classe de equivalência* $\mathcal{E}(\widehat{\mathtt{O}}_k)$ *das orientações equiprojectivas com* $\widehat{\mathtt{O}}_k = \mathtt{O}(\widehat{\alpha}_k, \widehat{\beta}_k, \widehat{\gamma}_k))$*, recorrendo à Proposição 2.3. Escolhe-se em* $\mathcal{E}(\widehat{\mathtt{O}}_k)$ *uma orientação* $\mathtt{O}'_k \in \mathcal{R}$*, cuja existência é garantida por* (2.16)*. Faz-se* $\widehat{\mathtt{O}}_k := \mathtt{O}'_k$*. Como foi dito no último parágrafo da Secção 2.5, isto permite manter pequenos espaços de procura na próxima iteração dos passos 3 e 4.*
- *Como passo final, parte-se da sequência obtida,* $\{\widehat{\mathtt{O}}_1, \ldots, \widehat{\mathtt{O}}_N\}$ *a que se associam classes de equivalência* $\{\mathcal{E}(\widehat{\mathtt{O}}_1), \ldots, \mathcal{E}(\widehat{\mathtt{O}}_N)\}$ *e selecciona-se em cada* $\mathcal{E}(\widehat{\mathtt{O}}_k)$ *uma orientação* $\mathtt{O}'_k$ *tal que a sequência* $\{\mathtt{O}'_1, \ldots, \mathtt{O}'_N\}$ *satisfaça, em cada instante* k*, o modelo das pequenas rotações. Faz-se então* $\{\widehat{\mathtt{O}}_1, \ldots, \widehat{\mathtt{O}}_N\} := \{\mathtt{O}'_1, \ldots, \mathtt{O}'_N\}$*.*

3.4 Hipótese nula e classificação de contornos

Tal como referido em [9, 10], a estimação da orientação a partir de uma imagem permite *a posteriori*:

(i) Verificar a validade do modelo de mundo de Manhattan para a imagem dada;

(ii) Classificar cada *pixel* da imagem de acordo com as classes $\{1, \ldots, 5\}$ definidas na Tabela 3.1.

Em (i), o objectivo é evitar estimativas erradas testando estatisticamente a hipótese do mundo de Manhattan; este teste permite detectar quando uma imagem se afasta deste modelo e, nessa eventualidade, abster-se de estimar a orientação.

Neste trabalho, em que a estimação é sequencial, esta verificação tem ainda a mais-valia de permitir detectar perdas de sincronismo no modelo de pequenas rotações, *i.e.*, situações em que devido à estimação errada de várias orientações consecutivas, a dependência do modelo das estimativas anteriores coloca-o "à deriva". Uma vez detectadas, estas situações podem ser corrigidas forçando a estimação seguinte a decorrer

independentemente da estimativa anterior, *i.e.*, como se fosse a primeira imagem de uma nova sequência.

A verificação da validade do modelo de mundo de Manhattan é feita construindo um modelo de hipótese nula, o que é feito modificando (3.8) para $P(\phi_{\mathbf{u}}|m_{\mathbf{u}}, \mathsf{O}, \mathbf{u}) = \mathsf{U}(\phi_{\mathbf{u}})$ e calculando

$$P_{\text{null}}(\{\mathbf{E_u}\}) = \prod_{\mathbf{u}} \left[\left(\sum_{m_{\mathbf{u}}=1}^{4} P_m(m_{\mathbf{u}}) \right) \cdot P_{\text{on}}(E_{\mathbf{u}}) + P_m(5) \cdot P_{\text{off}}(E_{\mathbf{u}}) \right] \mathsf{U}(\phi_{\mathbf{u}}), \quad (3.18)$$

o que significa deixar de distinguir diferentes tipos de contorno e deixar de assumir que a estatística da imagem reflecte qualquer tipo de estrutura tridimensional. A verosimilhança do modelo de hipótese nula é comparada com a do modelo de mundo de Manhattan, que é aproximada por

$$P_{\text{manh}}(\{\mathbf{E_u}\}) \approx P(\{\mathbf{E_u}\}|\widehat{\mathsf{O}})P(\widehat{\mathsf{O}}). \quad (3.19)$$

onde $\widehat{\mathsf{O}}$ é a orientação estimada. Se $\log P_{\text{manh}}(\{\mathbf{E_u}\}) - \log P_{\text{null}}(\{\mathbf{E_u}\}) < \delta$ para um certo $\delta \geq 0$ escolhido como margem de confiança, decide-se que o modelo de mundo de Manhattan não é suficientemente verosímil e descarta-se $\widehat{\mathsf{O}}$.

Em (ii), o objectivo é classificar *a posteriori* os *pixels* da imagem com os valores $1, \ldots, 5$, o que é feito tomando (3.6) e fazendo:

$$\widehat{m}_{\mathbf{u}} = \arg\max_{m_{\mathbf{u}}} P(E_{\mathbf{u}}|m_{\mathbf{u}})\, P(\phi_{\mathbf{u}}|m_{\mathbf{u}}, \mathsf{O}, \mathbf{u}). \quad (3.20)$$

A classificação dos *pixels* pode ser levada a cabo com o objectivo de captar a geometria do mundo de Manhattan: um procedimento possível é agrupar os *pixels* classificados como 1, 2 e 3 em linhas com a direcção dos eixos de Manhattan e computar as intersecções destas linhas, permitindo a sua reconstrução 3D por troços a menos de escalamentos. Apesar de este procedimento poder ser efectuado utilizando apenas uma imagem, o uso de uma sequência de vídeo permite refinar a reconstrução das linhas através, por exemplo, de restrições de proximidade baseadas no modelo de pequenas rotações. Pode ainda estabelecer-se correspondências com estas linhas entre as várias imagens, a fim de estimar a translação da câmara e combinar com a orientação já estimada para obter a pose completa (posição e orientação) a menos de um escalamento. Por outro lado, os *pixels* que forem classificados com $\widehat{m}_{\mathbf{u}} = 4$ pertencem a contornos não consistentes com os eixos de Manhattan; estes podem ser utilizados para detectar objectos não alinhados com esta estrutura (*e.g.*, pessoas num ambiente urbano), o que tem especial interesse em aplicações envolvendo reconhecimento de padrões.

3.5 Experiências e Resultados

A fim de avaliar a qualidade do método proposto, aplicou-se o Algoritmo 3.1 a um conjunto de sequências de vídeo adquiridas na Baixa Pombalina com uma câmara vulgar, em movimento livre e sem nenhum cuidado especial para evitar vibrações e outros efeitos indesejáveis. As sequências de vídeo utilizadas apresentam uma resolução baixa (cada imagem tem 288×360 *pixels*) e uma grande compressão (formato MPEG-4). Apesar da pouca qualidade das sequências, algumas contendo imagens sub- ou sobre-expostas devido a fortes contrastes luz/sombra, e à distorção radial que afasta a câmara do modelo da "câmara escura", concluiu-se que o Algoritmo 3.1 é capaz de estimar correctamente a orientação da câmara, de que é exemplo a Figura 3.4.

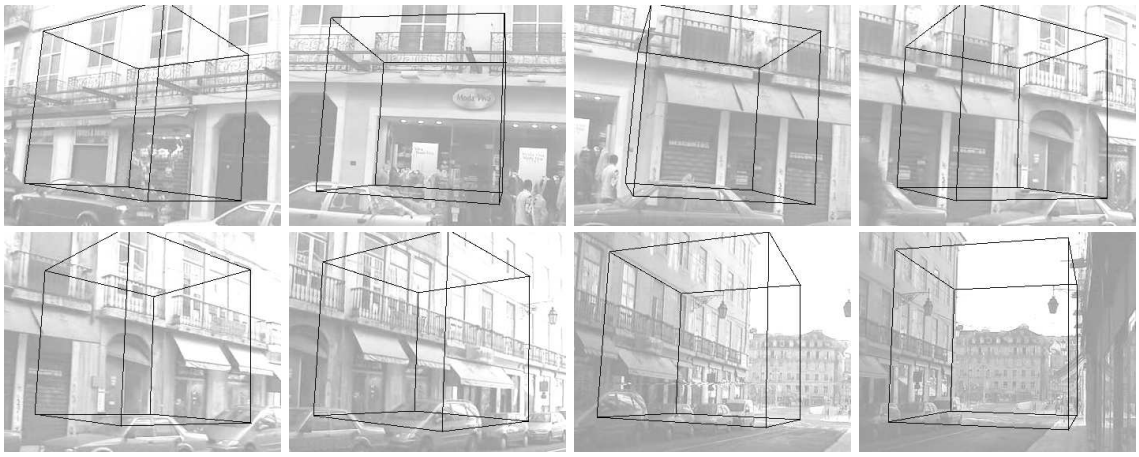

Figura 3.4: Estimativas da orientação da câmara para uma sequência na Rua da Prata, em Lisboa. Os cubos sobrepostos nas imagens representam a orientação estimada para os eixos de Manhattan. Representa-se a estimativa na primeira imagem da sequência ($k = 1$) e para imagens seguintes, em intervalos regulares.

As imagens das Figuras 3.5 e 3.6 provêm de duas outras sequências. Note-se que em ambos os casos a orientação é correctamente estimada, mesmo na presença de muitos contornos não alinhados com os eixos de Manhattan (por exemplo, as pessoas na Figura 3.6). Os gráficos em cada uma destas figuras representam as estimativas dos ângulos que parametrizam a orientação (*i.e.*, os ângulos de compasso, elevação e torção) para as duas sequências. Observando o gráfico, pode ver-se que as estimativas para a sequência da Figura 3.6 são um pouco mais ruidosas que as da Figura 3.5, o que se deve à qualidade inferior das imagens. A suavidade destas estimativas é controlada pelas variâncias postas em jogo nas f.d.p. (3.13) e (3.16), e que são escolhidas *a priori*; nas experiências realizadas, essas variâncias são iguais para os três ângulos e para ambas as sequências. Naturalmente, existe um compromisso entre esta suavidade e a capacidade de captar com precisão rotações mais bruscas da câmara.

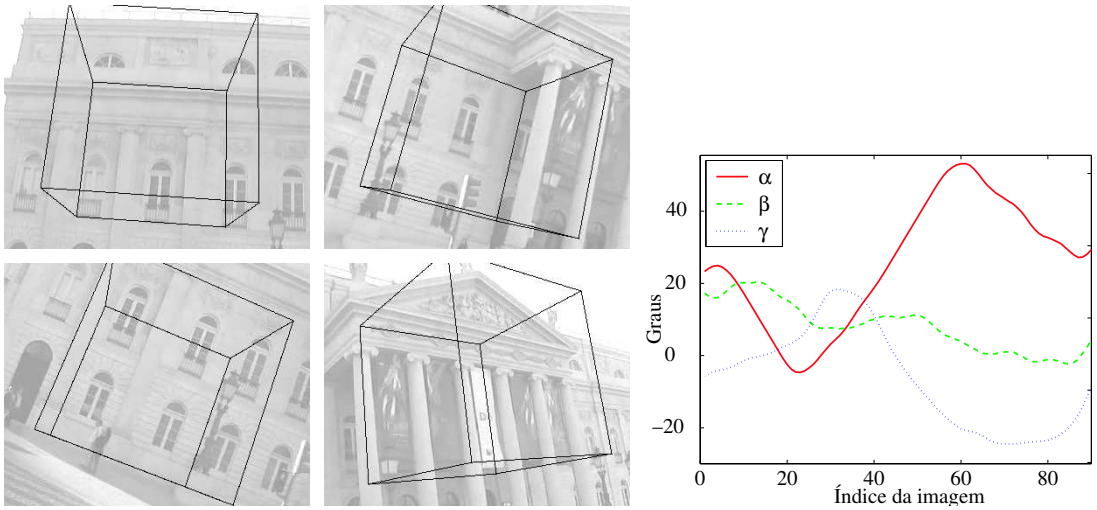

Figura 3.5: Esquerda: imagens 20, 30, 40 e 50 de outra sequência de vídeo, no Teatro D. Maria II, em Lisboa. Direita: gráfico temporal com as estimativas dos ângulos de compasso, elevação e torção que parametrizam a orientação da câmara.

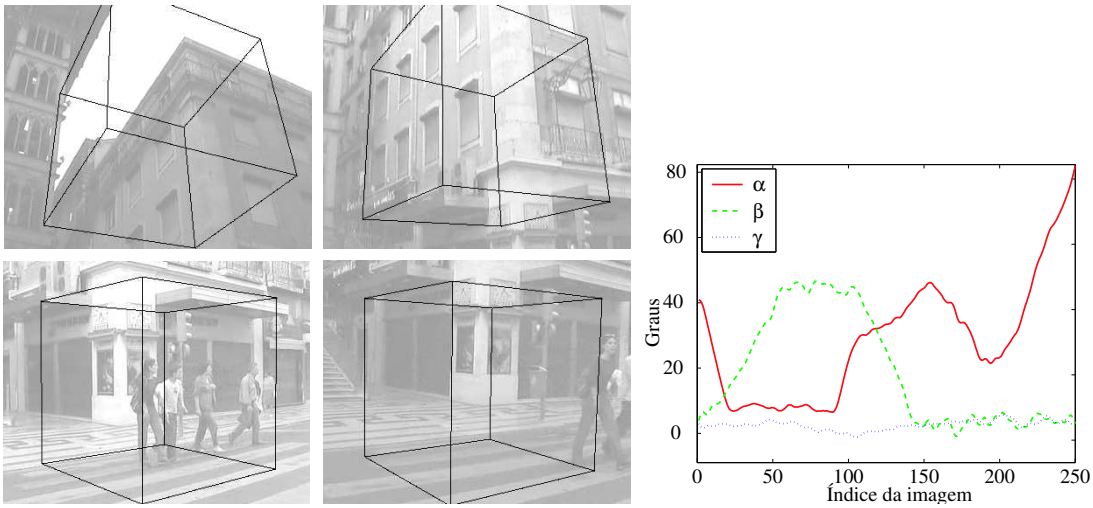

Figura 3.6: Esquerda: imagens 110, 130, 150 e 170 de uma terceira sequência de vídeo, no Elevador de Santa Justa, em Lisboa. Direita: gráfico temporal com as estimativas dos ângulos de compasso, elevação e torção que parametrizam a orientação da câmara.

A Figura 3.7 ilustra ainda dois exemplos de classificação de contornos, segundo o método descrito na Secção 3.4; um para uma imagem interior e outro para uma imagem exterior. Pode ver-se que a classificação é mais ruidosa no caso da imagem exterior, o que é representativo das restantes experiências efectuadas.

O tempo de processamento típico para cada imagem das sequências é menor que 1 segundo, tendo os testes sido efectuados num PC com processador Pentium IV a 3.0 GHz e a implementação do algoritmo no programa MATLAB, sem utilizar código pré-compilado. Neste momento, o algoritmo está a ser implementado em C/C++ com o objectivo de ser utilizado em tempo real.

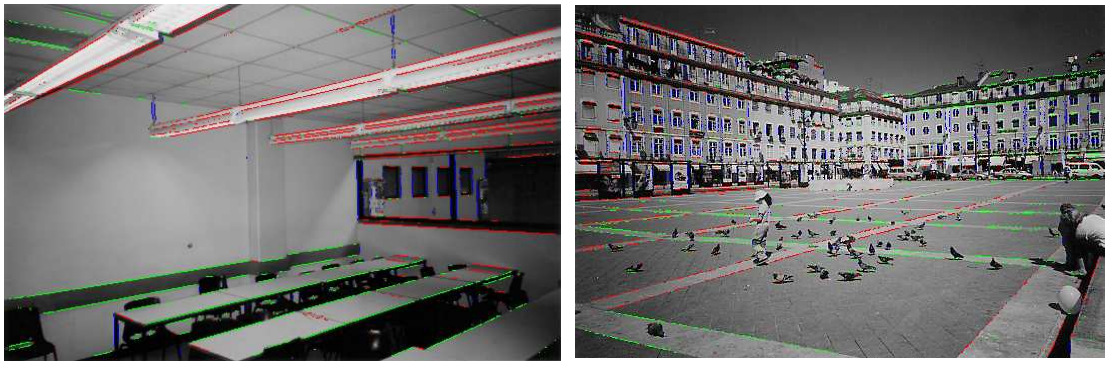

Figura 3.7: Exemplos de classificação de contornos para uma imagem interior e outra exterior. Representa-se respectivamente a vermelho, verde e azul *pixels* de contornos alinhados com os eixos de Manhattan $\mathbf{x}$, $\mathbf{y}$ e $\mathbf{z}$. Esquerda: Sala de aula. Direita: Praça da Figueira, Baixa Pombalina, Lisboa.

Capítulo 4

Conclusão

Neste trabalho, propõe-se uma abordagem probabilística para estimar a orientação da câmara a partir de sequências de vídeo em cenários urbanos. A abordagem utilizada tira partido da existência de "regularidades" nos contornos da imagem, o que é expresso através do modelo de mundo de Manhattan. O método proposto evita os passos intermédios tradicionalmente empregues, como a detecção e correspondência automática de padrões (cantos ou linhas de contorno), que por serem pouco robustos e computacionalmente pesados limitam o seu uso em aplicações práticas.

Os resultados experimentais demonstram que este método é capaz de lidar correctamente com sequências de vídeo de baixa qualidade e baixa resolução, ainda que o ambiente em causa se afaste razoavelmente do modelo de mundo de Manhattan, devido à presença de muitos contornos "espúrios". O algoritmo que aqui se introduz reduz a complexidade do problema e possibilita obter tempos de processamento que viabilizam o seu uso num sistema que funcione em tempo real.

As considerações teóricas introduzidas neste trabalho mostram como se pode tirar partido das classes de equivalência de orientações equiprojectivas para reduzir o domínio de procura da solução. Generaliza-se este conceito para uma classe genérica de mundos estruturados, n-dimensionais, de que o modelo de mundo de Manhattan tridimensional é um caso particular. Mostra-se como o mesmo conceito pode ser útil ao desenvolvimento de algoritmos de estimação em que existe conhecimento *a priori* da estrutura dos dados, independentemente da dimensão do espaço e do âmbito do problema, não se exigindo que este esteja relacionado com Visão por Computador.

Desenvolvimentos futuros incluem:

- optimização do algoritmo com vista à implementação prática de um sistema protótipo de estimação da orientação 3D em tempo real (trabalho corrente);
- desenvolvimento de métodos para estimar adicionalmente a posição 3D da câmara (dada a sua orientação) no sentido de obter uma calibração completa e permitir reconstruir a trajectória e efectuar uma reconstrução 3D do mundo;
- investigação do problema de detectar automaticamente a "estrutura" do mundo, nos casos em que não existe conhecimento prévio desta estrutura;
- avaliação da possibilidade de estender o método a problemas provindos de outras áreas, nomeadamente em Aprendizagem, Regressão e Classificação Automática, com possíveis aplicações em Inteligência Artificial, Bioinformática e Processamento de Língua Natural.

Bibliografia

[1] O. Faugeras. *Three-Dimensional Computer Vision*. MIT Press, Cambridge, MA, 1993.

[2] R. Hartley and A. Zisserman. *Multiple View Geometry*. Cambridge University Press, 2000.

[3] J. Mundy and A. Zisserman, editors. *Geometric Invariants in Computer Vision*. MIT Press, Cambridge, MA, 1992.

[4] E. Lutton, H. Maître, and J. Lopez-Krahe. Contribution to the determination of vanishing points using Hough transform. *IEEE Trans. on PAMI*, 16(4):430–438, 1994.

[5] S. Utcke. Grouping based on projective geometry constraints and uncertainty. In *Proc. IEEE Int. Conf. on Computer Vision*, Bombay, India, 1998.

[6] J. Kosecka and W. Zhang. Video compass. In European Conf. on Comp. Vision, Springer Verlag, LNCS 2350, 2002.

[7] B. Horn and E. Weldon Jr. Direct methods for recovering motion. *Int. Jour. Computer Vision*, 2(1):51–76, 1988.

[8] G. P. Stein and A. Shashua. Model-based brightness constraints: On direct estimation of structure and motion. *IEEE Trans. on PAMI*, 22(9):992–1015, 2000.

[9] J. Coughlan and A. Yuille. Manhattan world: Compass direction from a single image by Bayesian inference. In *Proc. IEEE Int. Conf. on Computer Vision*, Corfu, Greece, 1999.

[10] J. Coughlan and A. Yuille. The Manhattan world assumption: Regularities in scene statistics which enable Bayesian inference. In *Proc. Neural Information Processing Systems*, Denver, CO, USA, 2000.

[11] J. Deutscher, M. Isard, and J. MacCormick. Automatic camera calibration from a single Manhattan image. In *Proc. European Conf. on Computer Vision*, Springer Verlag, LNCS 2350, 2002.

[12] G. Schindler and F. Dellaert. Atlanta world: An expectation-maximization framework for simultaneous low-level edge grouping and camera calibration in complex man-made environments. In *Proc. IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition*, Washington, DC, 2004.

[13] Jean-Yves Bouguet. Camera Calibration Toolbox for Matlab. (http://www.vision.caltech.edu/bouguetj/calib_doc)

[14] Zhengyou Zhang. Flexible Camera Calibration by Viewing a Plane from Unknown Orientations. In *Proc. IEEE Int. Conf. on Computer Vision*, Kerkyra, Greece, 1999.

[15] I. Herman. *The Use of Projective Geometry in Computer Graphics.* Lecture Notes in Computer Science, Springer-Verlag, Amesterdam, Netherlands, 1992.

[16] N. Gordon A. Doucet, N. Freitas. *Sequential Monte Carlo Methods in Practice.* Springer-Verlag, New York, 2001.

[17] M. Isard and A. Blake. CONDENSATION – conditional density propagation for visual tracking. *Int. Jour. of Computer Vision*, 29(3):5–28, 1998.

[18] E.C. Hildreth Implementation of a theory of edge detection. M.I.T. Artificial Intell. Lab., Cambridge, MA, Tech. Rep. AITR579, 1980.

[19] R.M. Haralick Digital step edges from zero crossing of second directional derivatives. In *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1984.

[20] J.F. Canny. A computational approach to edge detection. In *IEEE Trans. on Pattern Analysis and Machine Intelligence*, November 1986.

[21] R.C. Gonzalez, R.E. Woods Digital image processing. AddisonWesley Publishing Company, Inc, 1993.

[22] S.M. Smith and J.M. Brady. SUSAN - A New Approach to Low Level Image Processing. Department of Engineering Science, Oxford University, Oxford, UK, 1995 (http://www.fmrib.ox.ac.uk/~steve/susan/susan/susan.html)

[23] M. Heath, S. Sarkar, T. Sanocki, K.W. Bowyer A Robust Visual Method for Assessing the Relative Performance of Edge-Detection Algorithms. In *IEEE Trans. on Pattern Analysis and Machine Intelligence*, December 1997.