

# **Orientação Automática na Baixa Pombalina**

## **Resumo**

O trabalho que aqui se apresenta incide sobre um dos assuntos mais investigados no âmbito da Visão Computacional: o problema da orientação automática. Particularmente, assume-se que neste problema apenas se dispõe de uma única imagem e que o espaço envolvente se enquadra num modelo designado por “mundo de Manhattan”, o que acontece em grande parte dos ambientes interiores e exteriores urbanos, onde predominam arestas aproximadamente orientadas segundo três direcções ortogonais entre si.

Em vez de se seguir uma estratégia baseada na detecção de contornos seguida da aplicação de transformadas de Hough (tarefa difícil e computacionalmente pesada), é utilizado um modelo bayesiano que conduz à estimativa MAP (máximo a posteriori) de uma função de verosimilhança em termos da orientação da câmara, previamente parametrizada à custa de 3 ângulos.

Importa frisar que a implementação prática deste projecto surge como consequência de algumas contribuições originais a nível teórico, de onde gostaria de destacar o desenvolvimento de alguns conceitos específicos a respeito da geometria do problema, e a posterior introdução de pequenas modificações no modelo bayesiano que serviu de base. Assim, pode afirmar-se que a componente prática do projecto aqui apresentado tem raízes fortes nos aspectos teóricos que se procuraram sempre ter em vista, sendo estes a melhor justificação para o caminho seguido. Como exemplo, pode referir-se a investigação adicional do problema da estimação da orientação a partir de uma sequência de vídeo, e que, apesar de não fazer parte dos objectivos iniciais do trabalho, se julgou oportuno abordar, dadas as “portas abertas” pelos fundamentos entretanto introduzidos.

Relativamente ao problema concreto que foi dado a resolver, resta referir que o resultado final alcançado foi a implementação de um algoritmo de estimação da orientação diferente dos que foram até aqui abordados. Este novo algoritmo revelou-se capaz de obter uma taxa de sucesso próxima de 90% com uma velocidade de processamento muito razoável (que se procurou sempre otimizar), o que considero muito satisfatório. Desenvolveu-se ainda uma interface gráfica de demonstração muito simples, com o objectivo de tornar o algoritmo de estimação acessível ao utilizador comum.

---

André Martins

31 de Outubro de 2002

Instituto Superior Técnico, Lisboa

“Foram ao comprido do Rossio, até ao fim. Voltaram, atravessaram-no em diagonal. E pelo lado do Arco do Bandeira, aproximaram-se da Rua do Ouro. Luísa olhava em redor, aflita, procurava uma ideia de ocasião, um acontecimento – e o conselheiro, grave a seu lado, dissertava.”

Eça de Queiroz, *O Primo Basílio*

# Índice de Temas

Lista de Figuras	vii
Lista de Tabelas	x
<b>1 Introdução</b>	<b>1</b>
<b>2 Análise Geométrica do Problema</b>	<b>3</b>
2.1 Mundo de Manhattan . . . . .	3
2.2 Orientação da câmara no espaço tridimensional . . . . .	4
2.2.1 A câmara obscura . . . . .	4
2.2.2 Parametrização da orientação . . . . .	6
2.2.3 Relação entre a orientação e os sistemas de eixos . . . . .	8
2.3 O plano projectivo da imagem . . . . .	8
2.3.1 Pontos de fuga como função da orientação . . . . .	8
2.3.2 Direcção das linhas de Manhattan projectadas no plano . . . . .	10
2.3.3 Propriedades geométricas dos pontos de fuga . . . . .	10
2.4 Orientações equiprojectivas . . . . .	13
2.4.1 Solução geral do problema da orientação dada uma solução particular . . . . .	13
2.4.2 Redução do intervalo onde se garante uma solução particular . . . . .	14
2.5 Análise da câmara em movimento . . . . .	15
2.5.1 Rotações . . . . .	15
2.5.2 Variação da orientação da câmara na hipótese de pequenas rotações . . . . .	16
<b>3 Extracção de Informação de uma Imagem</b>	<b>18</b>
3.1 Introdução . . . . .	18
3.2 O gradiente da intensidade luminosa . . . . .	18
3.2.1 Pré-filtragem gaussiana . . . . .	18
3.2.2 Filtros de Prewitt e de Sobel . . . . .	19
3.3 Quantificação do módulo do gradiente . . . . .	22
3.3.1 Motivação . . . . .	22
3.3.2 Quantificação uniforme . . . . .	23
3.3.3 Quantização não uniforme com saturação . . . . .	24
3.4 Exclusão selectiva de pixels . . . . .	25
3.4.1 Exclusão de pixels com módulo de gradiente reduzido . . . . .	25
3.4.2 Exclusão de pixels redundantes na vizinhança de contornos (non maxima supression) . . . . .	26

3.4.3	Resultados da exclusão selectiva . . . . .	27
3.5	Funcionamento geral da malha de extracção de informação . . . . .	27
<b>4</b>	<b>Estimação da Orientação da Câmara</b>	<b>29</b>
4.1	Introdução . . . . .	29
4.2	Treino e estimação de probabilidades a priori . . . . .	29
4.2.1	As probabilidades a priori . . . . .	29
4.2.2	Estimação das probabilidades $P_{ON}$ e $P_{OFF}$ . . . . .	30
4.2.3	Estimação das probabilidades $P_M$ e modelo para a função de distribuição $P_{ang}$ . . . . .	32
4.3	Modelo bayesiano . . . . .	33
4.4	Estimação do ângulo de compasso para uma câmara no plano horizontal . . . . .	35
4.5	Estimação da orientação no caso geral . . . . .	36
4.5.1	Estimação em bloco dos ângulos de compasso, elevação e torção . . . . .	36
4.5.2	Estimação da orientação em duas partes . . . . .	38
4.6	Estimação da orientação instantânea através de uma sequência de vídeo . . . . .	41
4.7	Classificação de contornos e detecção de objectos parasitas . . . . .	43
4.8	Resultados experimentais . . . . .	45
4.8.1	Condições das experiências . . . . .	45
4.8.2	Estimação da orientação a partir de uma imagem . . . . .	46
4.8.3	Estimação da orientação instantânea a partir de uma sequência de vídeo . . . . .	47
<b>5</b>	<b>Conclusões</b>	<b>49</b>
<b>A</b>	<b>Notas Adicionais Sobre a Geometria do Mundo de Manhattan</b>	<b>51</b>
A.1	Modelo geral da câmara projectiva . . . . .	51
A.2	Relação entre a orientação e os sistemas de eixos . . . . .	53
A.3	Representação de pontos no plano projectivo . . . . .	54
A.4	Localização dos pontos de fuga em função da orientação . . . . .	55
A.5	Cálculo das orientações equiprojectivas . . . . .	56
A.6	Delimitação da região que contém um ponto de fuga ao acaso . . . . .	60
A.7	Variação do ângulo de compasso na hipótese de pequenas rotações . . . . .	64
<b>B</b>	<b>Notas Adicionais Sobre a Malha de Extracção de Informação</b>	<b>67</b>
B.1	Determinação do valor máximo do módulo do gradiente na utilização de máscaras de Prewitt, Sobel e similares . . . . .	67
B.2	Testes comparativos do desempenho de vários medidores de contornos . . . . .	69
<b>C</b>	<b>Estudo de métodos para acelerar a estimação da orientação</b>	<b>72</b>
C.1	Método da procura de máximo local da função de verosimilhança . . . . .	72
C.2	Uso de coordenadas homogéneas para ultrapassar singularidades e reduzir o uso de funções trigonométricas . . . . .	76

<b>D Resultados experimentais para a estimação da orientação</b>	<b>78</b>
D.1 Imagens de interiores . . . . .	78
D.2 Imagens de exteriores . . . . .	82
D.3 Estimações incorrectas . . . . .	88
<b>Referências</b>	<b>91</b>

# Lista de Figuras

2.1	Fotografia de um mundo de Manhattan . . . . .	4
2.2	Sistemas de eixos do mundo de Manhattan e da câmara. . . . .	5
2.3	Projectção de um ponto no plano da imagem . . . . .	6
2.4	Parametrização da orientação da câmara . . . . .	7
2.5	Direcção de uma linha de Manhattan . . . . .	10
2.6	Módulo e argumento do ponto de fuga $\mathbf{F}_z$ . . . . .	11
2.7	Localização dos pontos de fuga . . . . .	12
2.8	Exemplo de orientações equiprojectivas . . . . .	14
2.9	Zona do plano da imagem onde se garante a existência de um ponto de fuga . . . . .	15
2.10	Exemplo de rotação. . . . .	16
3.1	Exemplo de suavização com um filtro gaussiano. . . . .	20
3.2	Resultados estimados para o ângulo do gradiente utilizando um medidor de contornos com filtros de Sobel. . . . .	22
3.3	Ocorrência dos vários níveis de quantificação para um esquema de quantificação uniforme com saturação . . . . .	24
3.4	Ocorrência dos vários níveis de quantificação para um esquema de quantificação logarítmica com saturação . . . . .	25
3.5	Intensidade do módulo do gradiente em cada ponto de uma imagem antes e depois da exclusão selectiva de pontos . . . . .	27
3.6	Esquema ilustrativo da malha de extracção de informação de uma imagem	28
4.1	Histograma das probabilidades $P_{ON}$ e $P_{OFF}$ . . . . .	31
4.2	Modelo utilizado para a função de distribuição da probabilidade $P_{ang}$ . . .	33
4.3	Zona do plano da imagem onde se garante a existência de um ponto de fuga . . . . .	36
4.4	Resultados da aplicação do Algoritmo de estimação da orientação em duas partes. . . . .	41
4.5	Classificação de linhas de Manhattan e detecção de objectos parasitas na imagem de uma sala do IST . . . . .	44
4.6	Classificação de linhas de Manhattan e detecção de objectos parasitas numa imagem do interior do Pavilhão de Civil do IST . . . . .	44
4.7	Exemplo de uma estimação correcta da orientação . . . . .	46
4.8	Exemplo de uma estimação errada da orientação para uma fotografia aérea da Baixa Pombalina. . . . .	47

5.1	Interface gráfica de demonstração da estimação da orientação . . . . .	49
A.1	Aberração esférica . . . . .	52
A.2	Efeito da distorção radial . . . . .	53
A.3	Representação das orientações equiprojectivas no espaço tridimensional .	57
A.4	Localização dos pontos de fuga com $\alpha = -10^\circ$ , $\beta = 50^\circ$ e $\gamma = 35^\circ$ . . . . .	61
A.5	Zonas do plano representadas pelos conjuntos $\mathcal{B}$ , $\mathcal{G}_\varphi$ e $\mathcal{B} \cap \mathcal{G}_\varphi$ . . . . .	62
A.6	Aplicação da hipótese de pequenas rotações . . . . .	65
B.1	Fotografia utilizada nos testes aplicados aos vários medidores de contornos	70
B.2	Histograma dos valores estimados para o ângulo do gradiente numa ima- gem de teste . . . . .	71
C.1	Gráficos de $F_1(\beta, \gamma)$ e $F_2(\alpha)$ obtidos na estimação da orientação em duas etapas utilizando o método do máximo local - caso típico . . . . .	74
C.2	Gráficos de $F_1(\beta, \gamma)$ e $F_2(\alpha)$ obtidos na estimação da orientação em duas etapas utilizando o método do máximo local - caso atípico . . . . .	75
D.1	Estimação da orientação e classificação de contornos para a imagem de uma caixa (Int01) . . . . .	78
D.2	Estimação da orientação e classificação de contornos para a imagem de uma sala do Pavilhão de Civil do IST (Int02) . . . . .	79
D.3	Estimação da orientação e classificação de contornos para uma imagem do interior do Pavilhão de Civil do IST (Int03) . . . . .	79
D.4	Estimação da orientação e classificação de contornos para uma imagem do interior do Pavilhão de Civil (Int04) . . . . .	79
D.5	Estimação da orientação e classificação de contornos para uma imagem do interior do Pavilhão de Civil (Int05) . . . . .	80
D.6	Estimação da orientação e classificação de contornos para uma imagem do interior de um comboio (Int06) . . . . .	80
D.7	Estimação da orientação e classificação de contornos para a imagem de um corredor no Pavilhão Central do IST (Int07) . . . . .	80
D.8	Estimação da orientação e classificação de contornos para a imagem de um corredor no Pavilhão de Civil do IST (Int08) . . . . .	81
D.9	Estimação da orientação e classificação de contornos para a imagem de uma estante com livros (Int10) . . . . .	81
D.10	Estimação da orientação e classificação de contornos para a imagem de um corredor no Pavilhão de Civil do IST (Int11) . . . . .	81
D.11	Estimação da orientação e classificação de contornos para uma imagem da Rua Augusta, na Baixa Pombalina (Ext01) . . . . .	82
D.12	Estimação da orientação e classificação de contornos para uma imagem da Rua Augusta, na Baixa Pombalina (Ext02) . . . . .	82
D.13	Estimação da orientação e classificação de contornos para uma imagem da Rua do Ouro, na Baixa Pombalina (Ext03) . . . . .	83
D.14	Estimação da orientação e classificação de contornos para uma imagem da Praça da Figueira, na Baixa Pombalina (Ext05) . . . . .	83

D.15 Estimação da orientação e classificação de contornos para uma imagem da Praça da Figueira, na Baixa Pombalina (Ext06) . . . . .	84
D.16 Estimação da orientação e classificação de contornos para uma imagem da Praça da Figueira, na Baixa Pombalina (Ext07) . . . . .	84
D.17 Estimação da orientação e classificação de contornos para uma imagem da Praça da Figueira, na Baixa Pombalina (Ext08) . . . . .	84
D.18 Estimação da orientação e classificação de contornos para uma imagem desenhada da Baixa Pombalina (Ext09) . . . . .	85
D.19 Estimação da orientação e classificação de contornos para uma imagem do Rossio, na Baixa Pombalina (Ext10) . . . . .	85
D.20 Estimação da orientação e classificação de contornos para uma imagem do Rossio, na Baixa Pombalina (Ext11) . . . . .	85
D.21 Estimação da orientação e classificação de contornos para uma imagem do Rossio, na Baixa Pombalina (Ext12) . . . . .	86
D.22 Estimação da orientação e classificação de contornos para uma imagem do Rossio, na Baixa Pombalina (Ext13) . . . . .	86
D.23 Estimação da orientação e classificação de contornos para uma imagem da Rua de Santa Justa, na Baixa Pombalina (Ext14) . . . . .	86
D.24 Estimação da orientação e classificação de contornos para uma imagem da Rua de Santa Justa, na Baixa Pombalina (Ext15) . . . . .	87
D.25 Estimação da orientação e classificação de contornos para uma vista aérea da Baixa Pombalina, virada para Norte (Ext16) . . . . .	87
D.26 Estimação da orientação e classificação de contornos para uma imagem aérea da Baixa Pombalina, virada para Sul (Ext19) . . . . .	87
D.27 Exemplos de estimações incorrectas: imagens Int09 e Ext04. . . . .	88
D.28 Exemplos de estimações incorrectas: imagens Ext17 e Ext18. . . . .	88
D.29 Funções de verosimilhança na estimação da orientação para a imagem Int06	89
D.30 Funções de verosimilhança na estimação da orientação para a imagem Int09	90



# Lista de Tabelas

4.1	Conjunto de modelos possíveis para cada pixel da imagem de um mundo de Manhattan e respectiva descrição . . . . .	30
4.2	Estimativas obtidas para as probabilidades $P_M$ . . . . .	32
4.3	Resultados experimentais da aplicação do algoritmo de estimação da orientação em duas partes em imagens de interiores e exteriores . . . . .	48
B.1	Resultados obtidos para vários medidores de contornos aplicados a uma imagem de teste . . . . .	71

## **Agradecimentos**

Gostaria de expressar o meu agradecimento a todos aqueles que contribuíram para a realização deste projecto, começando pelo orientador Prof. Pedro Aguiar e respectivo acompanhante Prof. Mário Figueiredo, e não esquecendo as ajudas preciosas do Prof. Paulo Almeida, da Sarah Almeida e do Ricardo Reis. É também imperativo não esquecer a participação do meu colega Bruno Henriques nos primeiros meses da realização deste trabalho.

# Capítulo 1

## Introdução

Foi em finais dos anos 60 que começou a surgir interesse em implementar sistemas de visão por computador. Nessa altura, tratava-se de um conjunto de problemas cuja investigação pertencia a domínios recentes para a época, como as Redes Neurais ou a Inteligência Artificial. A abordagem consistia tipicamente em partir do estudo de sistemas de visão biológica com o objectivo de simulá-los computacionalmente.

Com o decurso do tempo, verificou-se que o mecanismo de percepção visual é mais complexo do que inicialmente se supunha, o que levou a atacar o problema por outras vertentes, dando menos relevância aos campos da biologia, da psicologia da percepção e das ciências neuronais para incidir-se mais fortemente em áreas como a geometria projectiva, a computação gráfica e a teoria das probabilidades. A autonomia crescente deste tipo de problemas fez com que surgisse toda uma nova disciplina, designada no meio científico por *Visão Computacional*.

O trabalho que aqui se apresenta debruça-se sobre o problema da orientação automática, um dos assuntos mais investigados no domínio da Visão Computacional e sobre o qual todos os avanços ao longo das últimas quatro décadas são poucos quando comparados com tudo o que se encontra ainda por explorar. Em particular, o que aqui se estuda (e se implementa) são métodos alternativos para estimar a orientação a partir de *uma única imagem*, valorizando aspectos como a taxa de sucesso em condições não excepcionais e a rapidez na obtenção da solução.

Evitou seguir-se a estratégia tradicional do processamento da imagem que se baseia na determinação a priori de linhas de contorno seguida da aplicação de transformadas de Hough, uma tarefa difícil e computacionalmente pesada. Em vez disso, utiliza-se um modelo bayesiano que “vê” a imagem ao detalhe de cada pixel, ao qual se atribuem cinco modelos possíveis. Por inferência estatística, obtém-se uma função de verosimilhança em termos da orientação da câmara, sendo esta determinada como a estimativa do máximo a posteriori.

Esta estratégia foi utilizada com sucesso num projecto designado por “Manhattan World” e desenvolvido por Alan Yuille e James Coughlan no Smith-Kettlewell Eye Research Institute, nos Estados Unidos<sup>1</sup>, tendo sido publicados os dois artigos [1] e [2] que serviram de base a este trabalho.

---

<sup>1</sup>Pode encontrar-se mais detalhes sobre este projecto no sítio da Internet [http://www.ski.org/ALYuille\\_lab](http://www.ski.org/ALYuille_lab)

Não se pretendeu reproduzir a teoria e as experiências relacionadas com o projecto “Manhattan World”, mas antes aproveitar a pertinência dos fundamentos aí estabelecidos para introduzir novas contribuições e desenvolver métodos capazes de melhorar os resultados. Teve-se ainda presente o objectivo de generalizar estes fundamentos a domínios mais vastos.

Assim, dos contributos originais que aqui são introduzidos, gostaria de destacar:

- o conceito de *orientações equiprojectivas* e a sua contribuição para o estudo do universo de soluções do problema;
- a delimitação de uma região onde se garante a existência de pelo menos uma solução particular e a localização dos pontos de fuga para esta região;
- a proposta de um método de estimação da orientação *em duas partes* capaz de retirar um grau de liberdade ao problema.
- a introdução da *hipótese de pequenas rotações* da câmara para alargar o estudo ao problema da orientação a partir de uma sequência de vídeo.

O próximo capítulo debruça-se sobre a análise geométrica do problema, procurando clarificar alguns conceitos teóricos e construir um modelo geométrico capaz de servir de alicerce ao desenvolvimento computacional. A maior parte destes conceitos, bem como alguns resultados considerados importantes e as respectivas demonstrações, foram introduzidos pelo autor, constituindo o fundamento sobre o qual residem as abordagens experimentais, pelo que lhes é dado um destaque adicional.

No terceiro capítulo procura descrever-se a forma como a imagem é *captada* pela máquina. Estabelecendo um paralelo com a visão biológica, o que aqui se aborda é a *extracção da informação relevante* que depois é processada no mecanismo de percepção. A extracção da informação é feita, como veremos, através de um processo muito simples, onde se utilizam conceitos conhecidos do processamento de imagem, como a suavização, a medição de contornos e a exclusão selectiva.

O quarto capítulo mostra como se infere estatisticamente a orientação da câmara com base na informação disponível acerca da imagem. Refere-se como uma etapa de treino permite otimizar os parâmetros que são os “pilares” da estatística na estimação da orientação. São ainda estudados métodos para minimizar o tempo de processamento e apresenta-se os principais resultados experimentais.

Finalmente, o quinto capítulo mostra as conclusões finais.

A estes capítulos acrescenta-se um conjunto de anexos contendo demonstrações de resultados teóricos deduzidos pelo autor, resultados detalhados das experiências e o complemento de alguma informação que, por motivos de organização e limitação de espaço, não se julgou oportuno aparecer referida no corpo principal do relatório.

## Capítulo 2

# Análise Geométrica do Problema

### 2.1 Mundo de Manhattan

No contexto da Visão Computacional, é usual designar-se por *mundo* qualquer espaço físico sujeito à observação de um sistema automático de visão, sendo neste caso a *imagem* a representação que é produzida internamente por este mesmo sistema. Não é necessária uma análise muito profunda de um conjunto de imagens para nos apercebermos que estas podem facilmente ser classificadas e agrupadas consoante as propriedades que exibem: por exemplo, imagens de um espaço tipicamente urbano põem em evidência algumas características do mundo que retratam, como a presença de edifícios e avenidas no exterior e corredores e salas em interiores. Estes objectos, pela sua própria forma e pela maneira como se dispõem no espaço sugerem a presença de “linhas-guia” imaginárias que formam entre si uma teia ortogonal.

Estas considerações conduzem-nos à definição seguinte, que embora escapando ao rigor matemático normalmente exigido, atinge o propósito de fazer a ponte entre a realidade e o modelo geométrico que vai ser utilizado.

**Definição 2.1.** *Designa-se por **mundo de Manhattan** um mundo constituído essencialmente por objectos com arestas orientadas aproximadamente segundo um sistema de eixos ortogonais  $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ . Para além disso:*

- *Estes eixos ortogonais designam-se por **eixos de Manhattan**;*
- *As arestas que se encontram “alinhas” com  $\mathbf{x}$ ,  $\mathbf{y}$  ou  $\mathbf{z}$  designam-se por **linhas de Manhattan**;*
- *As arestas que não são rectas ou cujas direcções são substancialmente diferentes das de  $\mathbf{x}$ ,  $\mathbf{y}$  e  $\mathbf{z}$  designam-se por **linhas parasitas**;*
- *Objectos cujas arestas são linhas parasitas designam-se por **objectos parasitas**.*

Esta nomenclatura baseia-se em algumas das designações utilizadas correntemente em [1] e [2], podendo explicar-se a sua origem no carácter especificamente urbano de Manhattan, em Nova Iorque. A imagem da Figura (2.1) é o exemplo de um mundo de Manhattan, podendo facilmente identificar-se nela os eixos, linhas e objectos definidos atrás. Na figura, pode referir-se o candeeiro como exemplo de um objecto parasita.

Ao longo deste estudo debruçar-nos-emos sobre diversas imagens da Baixa Pombalina, por ser a arquitectura desta zona de Lisboa propícia a que se possam encontrar, lado a lado, exemplos mais próximos e mais afastados deste modelo, criando boas condições para a realização de experiências.



Figura 2.1: Fotografia de um mundo de Manhattan: Rua de Santa Justa, na Baixa Pombalina, em Lisboa.

O sistema de eixos  $(\mathbf{x}, \mathbf{y}, \mathbf{z})$  torna possível o uso de coordenadas cartesianas na representação de linhas e pontos do mundo de Manhattan.<sup>1</sup> Este sistema pode ser visto como a “bússola” que referencia qualquer objecto circundante, como por exemplo a própria câmara que capta a imagem. Se também atribuirmos à câmara um sistema de eixos ortogonais  $(\mathbf{a}, \mathbf{b}, \mathbf{c})$  estaremos aptos a definir a “orientação da câmara” como uma relação entre estes dois sistemas de eixos. Na Figura (2.2) representa-se o exemplo em que  $\mathbf{a}$ ,  $\mathbf{b}$  e  $\mathbf{c}$  definem respectivamente a largura, altura e comprimento do paralelepípedo que contém a câmara.

## 2.2 Orientação da câmara no espaço tridimensional

### 2.2.1 A câmara obscura

Ao longo deste estudo será sempre utilizado o modelo da *câmara obscura*<sup>2</sup> para representar o sistema de visão. Este é o modelo mais simples para uma câmara projectiva, sendo aplicável quando esta possui uma boa precisão, uma lente que introduz uma distorção desprezável e os parâmetros de calibração não variam. Quando isso não acontece

<sup>1</sup>Por simplicidade, decidi usar-se nesta fase um modelo euclidiano do espaço tridimensional. Mais adiante, no estudo do plano da imagem, irão ser utilizados alguns conceitos de Geometria Projectiva, por ser este o modelo mais adequado ao permitir ultrapassar algumas singularidades.

<sup>2</sup>Designado por “pinhole camera” na literatura anglo-saxónica. Numa câmara obscura, em vez de uma lente, supõe-se que existe uma abertura pontual para onde convergem todos os raios luminosos que vão dar origem à imagem.

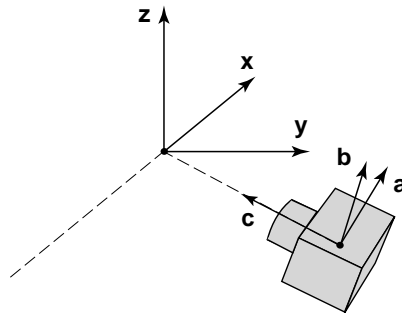


Figura 2.2: Sistema de eixos do mundo de Manhattan e sistema de eixos da câmara.

(como no caso das vulgares *webcams*) deve ter-se em conta a complexidade adicional que deriva do modelo geral da câmara projectiva e que se discute na Secção (A.1) do Apêndice, onde também se aborda superficialmente um método simples para corrigir a *distorção radial*.

Numa câmara obscura, o único parâmetro intrínseco de interesse é a *distância focal*, cujo significado geométrico é introduzido na definição que se segue.

**Definição 2.2.** *Designa-se por:*

- **ponto óptico:** o ponto onde converge o feixe de raios luminosos;
- **plano da imagem:** o plano onde é projectada a imagem bidimensional;
- **eixo principal:** o eixo perpendicular ao plano da imagem que passa pelo ponto óptico;
- **ponto principal:** a intersecção do eixo principal com o plano da imagem;
- **distância focal:** a distância entre o ponto óptico e o plano da imagem.

A Figura (2.3) mostra a projecção equivalente de um ponto no plano da imagem através de uma câmara obscura. Neste exemplo, a câmara mapeia o ponto  $\mathbf{X}$  do mundo no ponto  $\mathbf{X}'$  da imagem, o plano da imagem está a sombreado e o eixo principal é definido pelos pontos  $\mathbf{0}$  e  $\mathbf{P}$ , sendo o primeiro o ponto óptico e o segundo o ponto principal. A distância focal é representada por  $f$ , tendo-se da Definição (2.2) a relação  $f = \overline{\mathbf{0P}} = \|\mathbf{P}\|$ .

No estudo de imagens digitais é conveniente expressar a distância focal em unidades de pixel, procedimento que se mostra no exemplo seguinte.

**Exemplo 2.3.** *A distância focal em unidades de pixel pode ser estimada multiplicando a distância focal em unidades de comprimento pela relação entre a largura da imagem em unidades de pixel (após a digitalização) e a largura da imagem no filme. Por exemplo, imagens obtidas por uma câmara com um filme de 35 milímetros têm dimensões  $36 \times 24$  mm. Se a imagem tiver sido obtida com uma lente que apresenta uma distância focal de 50 mm e tiver sido digitalizada com uma dimensão  $768 \times 512$  pixels, então a distância focal em unidades de pixel será aproximadamente  $f = 50 \times \frac{768}{36} = 1067$  pixels.*

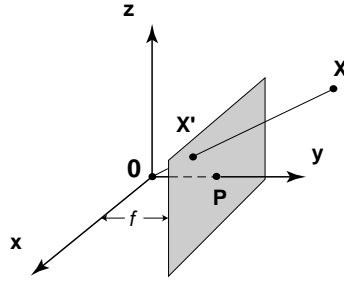


Figura 2.3: Projecção de um ponto no plano da imagem.

Quando se dispõe de um conjunto de imagens obtidas através da mesma câmara, deve ter-se em conta que para a distância focal (em unidades de pixel) ser igual em todas elas é necessário que as imagens digitalizadas apresentem a mesma resolução e que o zoom da câmara seja fixo.

### 2.2.2 Parametrização da orientação

A orientação é a relação entre os sistemas de eixos  $(\mathbf{a}, \mathbf{b}, \mathbf{c})$  da câmara e  $(\mathbf{x}, \mathbf{y}, \mathbf{z})$  do mundo de Manhattan, pelo que pode ser definida como a matriz  $\mathbf{M}$  que verifica  $(\mathbf{a}, \mathbf{b}, \mathbf{c})^\top = \mathbf{M}(\mathbf{x}, \mathbf{y}, \mathbf{z})^\top$ . No entanto, a orientação tem apenas 3 graus de liberdade, pelo que é corrente utilizar 3 parâmetros para representá-la.

A definição seguinte introduz a parametrização da orientação que é utilizada ao longo deste trabalho.

**Definição 2.4.** Representa-se por  $\Psi = (\alpha, \beta, \gamma)$  a **orientação** da câmara parametrizada pelos ângulos  $\alpha, \beta$  e  $\gamma$  definidos no domínio  $\alpha \in ]-\frac{\pi}{2}, \frac{\pi}{2}]$ ,  $\beta \in ]-\frac{\pi}{2}, \frac{\pi}{2}]$  e  $\gamma \in ]-\pi, \pi]$  de acordo com os seguintes princípios:

1.  $\alpha$  é o **ângulo de compasso**<sup>3</sup> definido no plano  $\mathbf{xy}$ ;
2.  $\beta$  é o **ângulo de elevação** entre o eixo principal e o vector obtido após se aplicar o ângulo  $\alpha$ ;
3.  $\gamma$  é o **ângulo de torção**<sup>4</sup> definido no plano da imagem;
4.  $\Psi = \mathbf{0} \Leftrightarrow (\mathbf{x}, \mathbf{y}, \mathbf{z}) = (\mathbf{c}, \mathbf{a}, \mathbf{b})$ ;
5.  $\gamma = 0 \Rightarrow \mathbf{a}^\top \mathbf{z} = 0 \wedge \mathbf{a}^\top \mathbf{y} \geq 0 \wedge \mathbf{b}^\top \mathbf{z} \geq 0$ .

Convém fazer alguns comentários em relação a esta definição:

- O conjunto dos dois ângulos  $\alpha$  e  $\beta$  define o eixo principal  $\mathbf{c}$ ;

<sup>3</sup>Também referido nalguma literatura como “ângulo de azimuth”.

<sup>4</sup>Referido como “twist angle” em [1].



- o ângulo  $\gamma$  define a rotação do plano da imagem (que afecta todos os pontos deste plano excepto o ponto principal, por este estar contido no eixo de rotação). O “rectângulo” da imagem, por sua vez, é descrito pelos vectores  $\mathbf{a}$  e  $\mathbf{b}$ ;
- o ponto (4) assegura que, quando os ângulos de compasso, elevação e torção são simultaneamente nulos ( $\Psi = \mathbf{0}$ ), os sistemas de eixos da câmara e do mundo de Manhattan coincidem, ou seja, nesta situação a câmara está colocada “de frente”;
- o ponto (5) referencia o ângulo  $\gamma$ , impondo que, com este ângulo nulo, a componente de  $\mathbf{a}$  segundo  $\mathbf{z}$  seja nula e a componente segundo  $\mathbf{y}$  seja positiva, o que define respectivamente a direcção e o sentido de  $\mathbf{a}$ .
- qualquer conjunto de valores de  $\alpha$ ,  $\beta$  e  $\gamma$  no domínio referido define univocamente uma orientação desde que  $\beta \neq \frac{\pi}{2}$ . Quando  $\beta = \frac{\pi}{2}$ , os vectores  $(\alpha, \frac{\pi}{2}, \gamma)$  e  $(\alpha + \phi, \frac{\pi}{2}, \gamma + \phi)$  representam a mesma orientação qualquer que seja o valor de  $\phi$ .

A orientação de uma câmara, bem como todos os ângulos e eixos envolvidos na sua definição, estão representados na Figura (2.4).

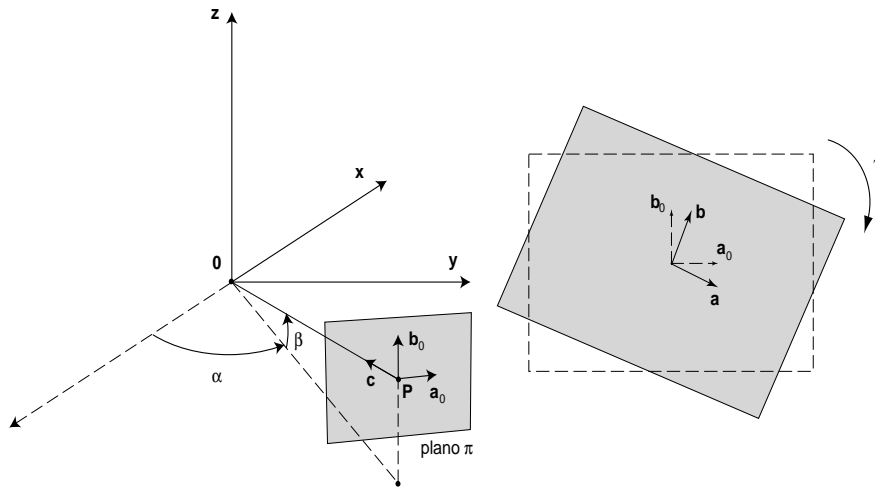


Figura 2.4: Parametrização da orientação da câmara através dos ângulos  $\alpha$ ,  $\beta$  e  $\gamma$ .

Convém notar que esta é apenas uma de muitas formas possíveis para parametrizar a orientação, sendo uma alternativa, por exemplo, à utilização dos ângulos de Euler. É importante ainda ter em conta que *não* existe entre os ângulos  $\alpha$ ,  $\beta$  e  $\gamma$  uma relação trivial de reciprocidade que faça com que uma permutação dos eixos  $(\mathbf{x}, \mathbf{y}, \mathbf{z})$  corresponda a uma idêntica permutação dos ângulos, como o exemplo seguinte comprova. As novas orientações resultantes de permutar os eixos (designadas por *orientações equiprojectivas*) serão estudadas adiante, na Secção (2.4).

**Exemplo 2.5.** *Uma câmara com uma distância focal  $f = 500\sqrt{2} = 707.1$  unidades de pixel está orientada em relação ao mundo de Manhattan de modo que as coordenadas cartesianas do ponto principal são dadas em unidades de pixel por  $\mathbf{P} = (-300, 400, 500)$*

e o ângulo de torção  $\gamma$  é nulo. Tem-se  $\alpha = \arctan \frac{4}{3} \simeq 53.1^\circ$  e  $\beta = 45^\circ$ . Os vectores  $\mathbf{a}$  e  $\mathbf{b}$  são respectivamente  $(0.8, 0.6, 0)$  e  $(0.4243, -0.5657, 0.7071)$ .

Vamos imaginar agora que o eixo  $\mathbf{y}$  passa a ser visto como  $\mathbf{z}$  e que, como consequência, o eixo  $\mathbf{x}$  mantém-se e o novo eixo  $\mathbf{y}$  será simétrico ao antigo eixo  $\mathbf{z}$ . Isto corresponde a uma simples permutação dos eixos. O ponto principal é agora  $\mathbf{P} = (-300, -500, 400)$ . Temos então um novo ângulo de compasso  $\alpha = \arctan \frac{-5}{3} \approx -59.0^\circ$  e um novo ângulo de elevação  $\beta = \arctan \frac{4}{\sqrt{3^2+5^2}} \approx 34.4^\circ$ . Se permutarmos os vectores  $\mathbf{a}$  e  $\mathbf{b}$  da mesma forma que o fizemos para  $\mathbf{x}$  e  $\mathbf{y}$ , resulta que  $a_z \neq 0$ , donde se conclui que **existe agora um ângulo de torção  $\gamma$  que é necessário considerar.**

Assim, uma permutação dos eixos não só conduziu a diferentes valores de  $\alpha$  e  $\beta$  como também introduziu um ângulo  $\gamma$  que dantes supusemos nulo.

### 2.2.3 Relação entre a orientação e os sistemas de eixos

Uma vez escolhida a forma de parametrizar a orientação da câmara, é possível escrever em função desses parâmetros o sistema de eixos  $(\mathbf{a}, \mathbf{b}, \mathbf{c})$  no sistema de coordenadas cartesianas do mundo de Manhattan.

Comecemos por introduzir os vectores  $\mathbf{a}_0 = \mathbf{a} |_{\gamma=0}$  e  $\mathbf{b}_0 = \mathbf{b} |_{\gamma=0}$  que se representaram na Figura (2.4). Para um ângulo de torção  $\gamma$  arbitrário, tem-se:

$$(\mathbf{a}, \mathbf{b})^\top = \mathbf{R}_\gamma(\mathbf{a}_0, \mathbf{b}_0)^\top \quad (2.1)$$

onde  $\mathbf{R}_\gamma$  é a *matriz de torção*, que define uma rotação de  $\gamma$  no plano da imagem e que é dada por:

$$\mathbf{R}_\gamma = \begin{bmatrix} \cos \gamma & -\sin \gamma \\ \sin \gamma & \cos \gamma \end{bmatrix} \quad (2.2)$$

Na Secção (A.2) do Apêndice deduz-se as expressões que relacionam  $\mathbf{a}_0$ ,  $\mathbf{b}_0$  e  $\mathbf{c}$  com a orientação  $\Psi = (\alpha, \beta, \gamma)$ , e que aqui reproduzimos:

$$\begin{cases} \mathbf{a}_0 &= (\sin \alpha, \cos \alpha, 0)^\top \\ \mathbf{b}_0 &= (\cos \alpha \sin \beta, -\sin \alpha \sin \beta, \cos \beta)^\top \\ \mathbf{c} &= (\cos \alpha \cos \beta, -\sin \alpha \cos \beta, -\sin \beta)^\top \end{cases} \quad (2.3)$$

O conjunto das Equações (2.1), (2.2) e (2.3) permite exprimir os eixos  $\mathbf{a}$ ,  $\mathbf{b}$  e  $\mathbf{c}$  nas coordenadas cartesianas do mundo de Manhattan em função da orientação.

## 2.3 O plano projectivo da imagem

### 2.3.1 Pontos de fuga como função da orientação

No problema em estudo, a única informação de que se dispõe acerca do mundo de Manhattan é uma imagem, a partir da qual se quer obter a orientação. Interessa assim estudar quantitativamente o efeito da orientação no plano da imagem, o que será feito recorrendo a alguns conceitos da Geometria Projectiva<sup>5</sup>, pois isso irá permitir que se

<sup>5</sup>Pode encontrar-se inúmeras referências bibliográficas acerca da Geometria Projectiva aplicada a problemas de Engenharia. Em particular, no âmbito da Visão Computacional, encontra-se resumos muito interessantes em [3], [4] e [5].

ultrapassem algumas singularidades. A este respeito, convém fazer referência à Secção (A.3) do Apêndice, que introduz as diferentes notações (cartesiana, homogênea e complexa) que são utilizadas neste capítulo para representar os pontos do plano projectivo  $\mathbb{P}^2$ . A leitura desta secção é assim fundamental para a compreensão do texto que se segue.

Vamos começar pela definição seguinte, que recorre à noção de *ponto ideal* introduzida por este ramo da Geometria:

**Definição 2.6.** *Seja  $\pi$  o plano projectivo da imagem. Dado um eixo  $\mathbf{e}$  em  $\mathbb{R}^3$  que passe pela origem (o ponto óptico  $\mathbf{0}$ ), designa-se por **ponto de fuga** segundo  $\mathbf{e}$  (e representa-se por  $\mathbf{F}_e$ ) a projecção no plano  $\pi$  do ponto ideal de  $\mathbf{e}$ .*

Esta definição garante que para qualquer eixo no espaço tridimensional existe sempre um e um só ponto de fuga, podendo este ser afim ou ideal. Assim não se perde generalidade nos casos em que o eixo é paralelo ao plano da imagem.

Num mundo de Manhattan, onde predominam linhas com as direcções dos eixos  $\mathbf{x}$ ,  $\mathbf{y}$  e  $\mathbf{z}$ , é de esperar a predominância no plano da imagem de linhas convergentes para os três pontos de fuga  $\mathbf{F}_x$ ,  $\mathbf{F}_y$  e  $\mathbf{F}_z$ , pelo que se torna necessário relacionar a orientação da câmara com a localização destes pontos de fuga.

As equações seguintes, deduzidas na Secção (A.4) do Apêndice estabelecem essa relação.

1. Em coordenadas cartesianas:

$$\begin{cases} \mathbf{F}_x = f \left( \frac{a_x}{c_x}, \frac{b_x}{c_x} \right)^\top = f \mathbf{R}_\gamma \left( \frac{\tan \alpha}{\cos \beta}, \tan \beta \right)^\top \\ \mathbf{F}_y = f \left( \frac{a_y}{c_y}, \frac{b_y}{c_y} \right)^\top = f \mathbf{R}_\gamma \left( -\frac{\cot \alpha}{\cos \beta}, \tan \beta \right)^\top \\ \mathbf{F}_z = f \left( \frac{a_z}{c_z}, \frac{b_z}{c_z} \right)^\top = f \mathbf{R}_\gamma (0, -\cot \beta)^\top \end{cases} \quad (2.4)$$

onde  $\mathbf{R}_\gamma$  é dado pela Equação (2.2);

2. Utilizando vectores homogêneos:

$$\begin{cases} \tilde{\mathbf{F}}_x = (fa_x, fb_x, c_x)^\top = \mathbf{H}_{f,\gamma} (\sin \alpha, \cos \alpha \sin \beta, \cos \alpha \cos \beta)^\top \\ \tilde{\mathbf{F}}_y = (fa_y, fb_y, c_y)^\top = \mathbf{H}_{f,\gamma} (\cos \alpha, -\sin \alpha \sin \beta, -\sin \alpha \cos \beta)^\top \\ \tilde{\mathbf{F}}_z = (fa_z, fb_z, c_z)^\top = \mathbf{H}_{f,\gamma} (0, \cos \beta, -\sin \beta)^\top \end{cases} \quad (2.5)$$

com

$$\mathbf{H}_{f,\gamma} = \begin{bmatrix} f \mathbf{R}_\gamma & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix} = \begin{bmatrix} f \cos \gamma & -f \sin \gamma & 0 \\ f \sin \gamma & f \cos \gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

3. Em notação complexa:

$$\begin{cases} \bar{\mathbf{F}}_x = f \left( \frac{\tan \alpha}{\cos \beta} + j \tan \beta \right) \exp(j\gamma) \\ \bar{\mathbf{F}}_y = f \left( -\frac{\cot \alpha}{\cos \beta} + j \tan \beta \right) \exp(j\gamma) \\ \bar{\mathbf{F}}_z = f \cot \beta \exp \left[ j \left( \gamma - \frac{\pi}{2} \right) \right] = -j f \cot \beta \exp(j\gamma) \end{cases} \quad (2.6)$$

### 2.3.2 Direcção das linhas de Manhattan projectadas no plano

Um ponto afim do plano da imagem pode representar-se pelo vector homogéneo  $\mathbf{u} = (u_1, u_2, 1)^\top$ . É importante obter a direcção das linhas que unem  $\mathbf{u}$  a cada um dos pontos de fuga, pois a Definição (2.1) mostra que os objectos de Manhattan, projectados no plano, *têm os seus contornos alinhados com estas linhas*.

Se representarmos por  $\theta_x$ ,  $\theta_y$  e  $\theta_z$  os ângulos que as normais de cada uma destas linhas fazem com a horizontal para cada um dos pontos de fuga  $\mathbf{F}_x$ ,  $\mathbf{F}_y$  e  $\mathbf{F}_z$ , tal como mostrado na Figura (2.5), deduz-se facilmente da Equação (2.5) as expressões:

$$\tan \theta_x = -\frac{fa_x - u_1c_x}{fb_x - u_2c_x}, \quad \tan \theta_y = -\frac{fa_y - u_1c_y}{fb_y - u_2c_y} \quad \text{e} \quad \tan \theta_z = -\frac{fa_z - u_1c_z}{fb_z - u_2c_z} \quad (2.7)$$

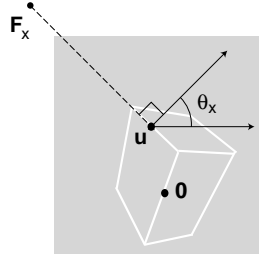


Figura 2.5: Linha de Manhattan originada por  $\mathbf{x}$  no ponto  $\mathbf{u}$  e ângulo formado pela normal a esta linha com a horizontal.

### 2.3.3 Propriedades geométricas dos pontos de fuga

Os pontos de fuga  $\mathbf{F}_x$ ,  $\mathbf{F}_y$  e  $\mathbf{F}_z$  possuem algumas propriedades geométricas de grande interesse. Pode dizer-se que a maior parte das contribuições originais deste trabalho deriva de forma directa ou indirecta de um aprofundamento do estudo destas propriedades, como veremos, aliás, na secção seguinte.

**Propriedade 2.7.** *A localização no plano da imagem do ponto de fuga  $\mathbf{F}_z$  é independente do ângulo de compasso  $\alpha$ .*

*Para além disso:*

- *A distância entre este ponto de fuga e o ponto principal da imagem (ou seja, o módulo de  $\overline{\mathbf{F}}_z$ ) depende apenas do ângulo de elevação  $\beta$ ;*
- *A direcção da linha que une este ponto de fuga ao ponto principal (ou, de outra forma, o argumento de  $\overline{\mathbf{F}}_z$ ) depende apenas do ângulo de torção  $\gamma$ .*

Esta propriedade deriva da parametrização da orientação expressa na Definição (2.4), sendo visível por simples observação da Equação (2.6). A Figura (2.6) ilustra o efeito geométrico descrito por esta propriedade, mostrando como os ângulos  $\beta$  e  $\gamma$  definem por si só a localização do ponto de fuga  $\mathbf{F}_z$ . Pode ainda ver-se que a linha que contém os

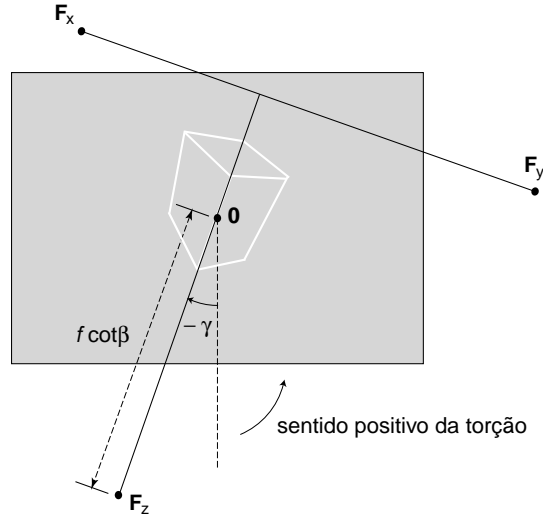


Figura 2.6: O módulo e o argumento do ponto de fuga  $\mathbf{F}_z$  como função dos ângulos de elevação e de torção.

pontos de fuga  $\mathbf{F}_x$  e  $\mathbf{F}_y$  é perpendicular à linha que contém  $\mathbf{F}_z$  e o ponto principal, o que é uma consequência das duas últimas expressões da Equação (2.6).

A propriedade que se apresenta de seguida é também de importância capital, desfazendo a sugestão, trazida pela Propriedade (2.7), do eixo  $\mathbf{z}$  ser “privilegiado” em relação aos outros eixos.

**Propriedade 2.8.** *Os três pontos de fuga  $\mathbf{F}_x$ ,  $\mathbf{F}_y$  e  $\mathbf{F}_z$  são indistinguíveis entre si, isto é, desconhecendo a geometria do espaço tridimensional é impossível identificar univocamente cada um dos pontos de fuga como sendo gerado por  $\mathbf{x}$ ,  $\mathbf{y}$  ou  $\mathbf{z}$ . Isto implica que, ainda que seja conhecida na imagem a posição de todos os pontos de fuga, a impossibilidade de dizer onde está cada um leva a que existam múltiplas soluções  $\{\Psi\}$  para o problema da estimação da orientação.*

As duas propriedades anteriores permitem concluir que *basta localizar um ponto de fuga qualquer* (mesmo nada sabendo acerca dos outros dois) *para determinar de imediato soluções particulares para os ângulos de elevação  $\beta$  e de torção  $\gamma$* . A localização dos restantes pontos de fuga permite, depois, determinar uma solução para o ângulo de compasso  $\alpha$ .

Isto permite que se separe a estimação da orientação em duas partes, como veremos na Secção (4.5).

Vamos agora estudar a localização dos pontos de fuga  $\mathbf{F}_x$ ,  $\mathbf{F}_y$  e  $\mathbf{F}_z$  nos casos particulares em que algum ou alguns dos ângulos  $\alpha$ ,  $\beta$  e  $\gamma$  são nulos.

**Propriedade 2.9.** *Para qualquer orientação  $\Psi$  ocorre um dos seguintes casos possíveis:*

- *Quando os ângulos de elevação e de torção são nulos (tendo-se  $\alpha \neq \frac{k\pi}{2}$ ) o ponto de fuga segundo  $\mathbf{z}$  é ideal, tendo-se  $\tilde{\mathbf{F}}_z = (0, 1, 0)$ . Os pontos de fuga segundo  $\mathbf{x}$*

e segundo  $\mathbf{y}$  estão, por sua vez, situados na linha horizontal que passa pelo ponto principal, como se mostra na Figura (2.7a);

- Quando o ângulo de torção é nulo (tendo-se  $\alpha \neq \frac{k\pi}{2}$  e  $\beta \neq \frac{k\pi}{2}$ ) o ponto de fuga segundo  $\mathbf{z}$  é afim e situa-se na linha vertical que passa pelo ponto principal. Neste caso os pontos de fuga segundo  $\mathbf{x}$  e segundo  $\mathbf{y}$  estão situados numa linha horizontal que não passa pelo ponto principal, como se pode ver na Figura (2.7b);
- No caso geral, qualquer que seja a orientação da câmara  $\Psi = (\alpha, \beta, \gamma)$ , o ponto principal é sempre o **ortocentro** do triângulo definido pelos três pontos de fuga. Um caso geral em que  $\alpha \neq \frac{k\pi}{2}$ ,  $\beta \neq \frac{k\pi}{2}$  e  $\gamma \neq \frac{k\pi}{2}$  encontra-se ilustrado na Figura (2.7c);
- Os casos em que  $\alpha = \frac{k\pi}{2}$  ou  $\beta = \pm \frac{\pi}{2}$  são situações limite de um dos três casos anteriores, passando um dos pontos de fuga a ser ideal.

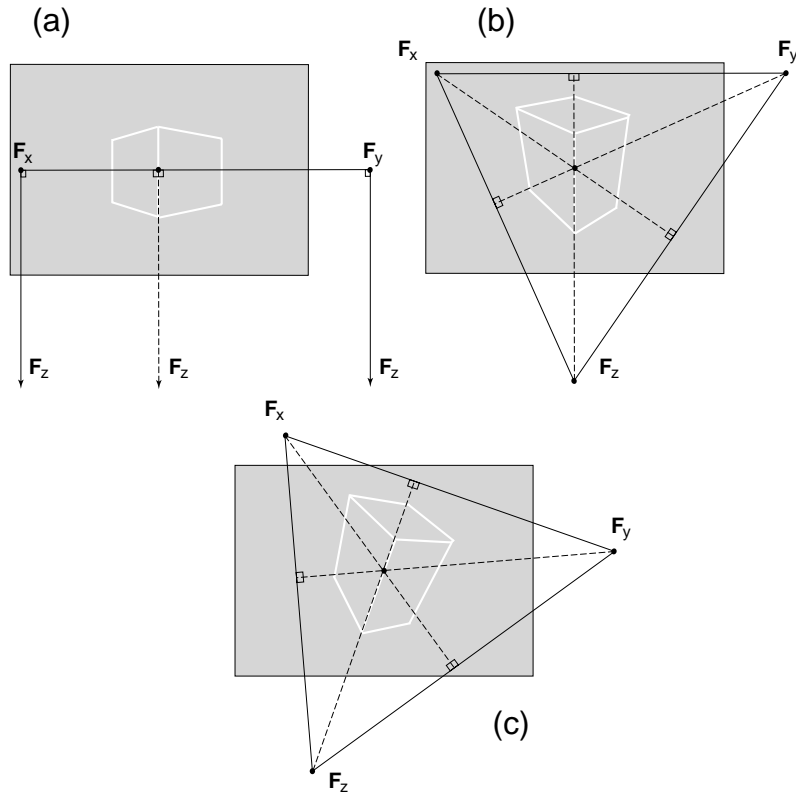


Figura 2.7: Localização dos pontos de fuga para (a) ângulos de elevação e de torção nulos, (b) ângulo de torção nulo e (c) caso geral.

O caso especial da Figura (2.7a) tem especial interesse, pois neste caso a orientação da câmara é dada por  $\Psi = (\alpha, 0, 0)$  e o único grau de liberdade é o ângulo de compasso, o que resulta da câmara se encontrar assente sobre o plano horizontal. É uma hipótese admitida na estimação da orientação da câmara em algumas situações reais, como se mostra em [1].

## 2.4 Orientações equiprojectivas

### 2.4.1 Solução geral do problema da orientação dada uma solução particular

A Propriedade (2.8) deixa patente a ideia de que, mais do que referir-nos a cada ponto de fuga  $\mathbf{F}_x$ ,  $\mathbf{F}_y$  e  $\mathbf{F}_z$  individualmente, interessa introduzir o *conjunto* desses três pontos de fuga. Por outro lado, o Exemplo (2.5) mostrou-nos que diferentes orientações podem conduzir a conjuntos idênticos, o que sugere que se defina uma relação de equivalência entre estas orientações. Somos assim conduzidos à definição de *orientações equiprojectivas*.

**Definição 2.10.** *Seja  $\Psi = (\alpha, \beta, \gamma)$  a orientação de uma câmara em relação ao referencial de Manhattan. O conjunto dos pontos de fuga  $\mathcal{F} = \{\mathbf{F}_x, \mathbf{F}_y, \mathbf{F}_z\}$  pode obter-se de  $\Psi$  a partir de uma função  $f$  tal que  $\mathcal{F} = f(\Psi)$ .*

*Diz-se que duas orientações  $\Psi_1$  e  $\Psi_2$  são equiprojectivas se e só se  $f(\Psi_1) = f(\Psi_2) = \mathcal{F}$ , escrevendo-se nesse caso  $\Psi_1 \equiv \Psi_2$ .*

A relação ‘ $\equiv$ ’ satisfaz as propriedades de transitividade, simetria e reflexão necessárias para ser uma relação de equivalência, pelo que se torna importante caracterizar as *classes de equivalência* daí resultantes. Isto assume um claro interesse devido ao facto de, no problema da estimação da orientação, o processamento da imagem permitir apenas inferir a orientação a partir da localização dos pontos de fuga, *sendo a solução geral o conjunto dos elementos de uma classe de equivalência e não uma orientação específica*.

Interessa assim deduzir expressões que permitam, dada uma orientação, obter todas as que com ela são equiprojectivas.

**Teorema 2.11.** *Qualquer orientação  $\Psi = (\alpha, \beta, \gamma)$  pertence a uma classe de equivalência com  $L \leq 12$  elementos<sup>6</sup>, que são as orientações equiprojectivas com  $\Psi$ . Cada um destes elementos pode ser calculado da seguinte forma:*

1. Começa por obter-se o conjunto  $E_1 = \{\Psi, \Psi', \Psi'', \Psi'''\}$  a partir das expressões:

$$\begin{cases} \Psi' &= (\alpha \pm \frac{\pi}{2}, \beta, \gamma) \\ \Psi'' &= (-\alpha, -\beta, \gamma \pm \pi) \\ \Psi''' &= (-\alpha \pm \frac{\pi}{2}, -\beta, \gamma \pm \pi) \end{cases} \quad (2.8)$$

2. Deste conjunto toma-se um elemento  $\Psi_1 = (\alpha_1, \beta_1, \gamma_1)$  que verifique  $\alpha_1 \in [0, \frac{\pi}{2}]$  e  $\beta_1 \in [0, \frac{\pi}{2}]$ ;<sup>7</sup>

3. Calcula-se  $\Psi_2 = (\alpha_2, \beta_2, \gamma_2)$  e  $\Psi_3 = (\alpha_3, \beta_3, \gamma_3)$  através das expressões:

$$\begin{cases} \alpha_2 &= \arctan\left(\frac{\tan \beta_1}{\sin \alpha_1}\right) \\ \beta_2 &= \arcsin(\cos \alpha_1 \cos \beta_1) \\ \gamma_2 &= \gamma_1 - \arctan\left(\frac{\tan \alpha_1}{\sin \beta_1}\right) \pm \pi \end{cases} \quad e \quad \begin{cases} \alpha_3 &= \operatorname{arccot}\left(\frac{\tan \beta_1}{\cos \alpha_1}\right) \\ \beta_3 &= \arcsin(\sin \alpha_1 \cos \beta_1) \\ \gamma_3 &= \gamma_1 + \arctan\left(\frac{\cot \alpha_1}{\sin \beta_1}\right) \pm \pi \end{cases}$$

<sup>6</sup>Conforme dito anteriormente, consideram-se iguais as orientações  $(\alpha, \frac{\pi}{2}, \gamma)$  e  $(\alpha + \phi, \frac{\pi}{2}, \gamma + \phi)$ ,  $\forall \phi$ . As classes de equivalência destas orientações têm 10 ou menos elementos.

<sup>7</sup>Este elemento existe sempre, como se pode ver na Equação (2.8).

4. Obtém-se através da Equação (2.8) os conjuntos  $E_2 = \{\Psi_2, \Psi'_2, \Psi''_2, \Psi'''_2\}$  e  $E_3 = \{\Psi_3, \Psi'_3, \Psi''_3, \Psi'''_3\}$ . A classe de equivalência é o conjunto  $E = E_1 \cup E_2 \cup E_3$ .

*Demonstração.* A demonstração deste teorema é uma contribuição original e apresenta-se em anexo, na Secção (A.5). ■

**Exemplo 2.12.** Na sequência do Exemplo (2.5), estamos agora em condições de determinar todas as orientações equiprojectivas com  $\Psi = (53.1^\circ, 45^\circ, 0^\circ)$ . Como  $\alpha = 53.1^\circ \in [0, 90^\circ]$  e  $\beta = 45^\circ \in [0, 90^\circ]$  pode fazer-se  $\alpha_1 = \alpha$  e  $\beta_1 = \beta$ , vindo do passo (3) do Teorema (2.11):

$$\begin{cases} \alpha_2 = 51.4^\circ \\ \beta_2 = 25.1^\circ \\ \gamma_2 = 118.0^\circ \end{cases} \quad e \quad \begin{cases} \alpha_3 = 31.0^\circ \\ \beta_3 = 34.4^\circ \\ \gamma_3 = -133.3^\circ \end{cases}$$

e obtendo-se por fim os resultados que se apresentam na Figura (2.8).

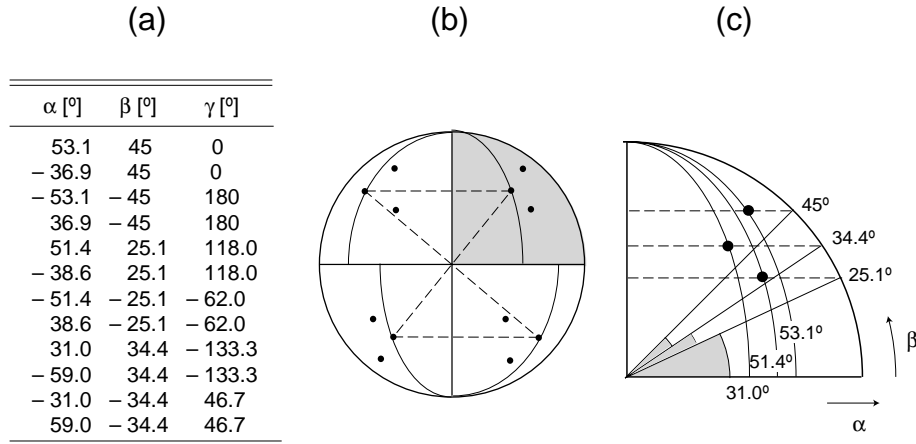


Figura 2.8: Exemplo de orientações equiprojectivas na solução do Exemplo (2.12): (a) Tabela de resultados, (b) e (c) Vista no espaço dos pontos principais das orientações equiprojectivas. Em (b) apresenta-se a sua localização na semi-esfera  $x^2 + y^2 + z^2 = f \wedge x \leq 0$  e em (c) mostra-se a localização de 3 destas orientações no oitavo de esfera definido por  $x^2 + y^2 + z^2 = f \wedge x \leq 0 \wedge y \geq 0 \wedge z \geq 0$ . Uma caracterização geométrica mais completa das orientações equiprojectivas pode ser encontrada na Secção (A.5) do Apêndice.

## 2.4.2 Redução do intervalo onde se garante uma solução particular

Um problema de grande interesse na estimação da orientação consiste em encontrar uma região o mais “pequena” possível que contenha garantidamente uma solução particular, isto é, uma orientação equiprojectiva com a verdadeira orientação da câmara. O



objectivo é limitar a busca a esse intervalo, reduzindo assim o processamento e tornando mais eficiente o algoritmo de estimação.

O domínio em que estão definidos os ângulos  $\alpha$ ,  $\beta$  e  $\gamma$  é, da Definição (2.4),

$$\alpha \in \left]-\frac{\pi}{2}, \frac{\pi}{2}\right] \quad , \quad \beta \in \left]-\frac{\pi}{2}, \frac{\pi}{2}\right] \quad \text{e} \quad \gamma \in ]-\pi, \pi]$$

Após alguma manipulação algébrica é possível obter um intervalo muito mais reduzido onde se garante a existência de uma solução particular, como se mostra no teorema seguinte:

**Teorema 2.13.** *Qualquer orientação da câmara é equiprojectiva com uma orientação  $\Psi = (\alpha, \beta, \gamma)$  com*

$$\alpha \in \left]-\frac{\pi}{4}, \frac{\pi}{4}\right] \quad , \quad \beta \in \left]-\frac{\pi}{4}, \frac{\pi}{4}\right] \quad \text{e} \quad \gamma \in \left]-\frac{\pi}{2} + \arctan \frac{\sqrt{2}}{2}, \frac{\pi}{2} - \arctan \frac{\sqrt{2}}{2}\right]$$

*Demonstração.* A demonstração deste teorema é uma contribuição original e apresenta-se em anexo, na Secção (A.6). ■

O intervalo em graus para o ângulo de torção é aproximadamente  $\gamma \in ]-54.7^\circ, 54.7^\circ]$ . Pode ver-se na Secção (A.6) que este curioso resultado equivale à propriedade seguinte:

**Propriedade 2.14.** *Qualquer que seja a orientação  $\Psi$  da câmara, existe pelo menos um ponto de fuga na zona do plano da imagem representada a sombreado na Figura (2.9).*

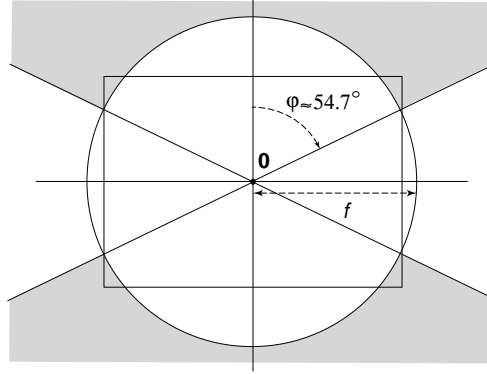


Figura 2.9: Zona do plano onde o Teorema (2.13) garante existir pelo menos um ponto de fuga.

## 2.5 Análise da câmara em movimento

### 2.5.1 Rotações

Um movimento da câmara compõe-se de uma translação e de uma rotação. Como vimos anteriormente, nenhuma translação afecta só por si a orientação  $\Psi$  da câmara,

pois uma translação igual do sistema de eixos de Manhattan coloca os dois sistemas na mesma posição relativa. Por conseguinte, *apenas a componente rotacional do movimento tem relevância no estudo da variação da orientação da câmara  $\Psi(t)$  ao longo do tempo.*

No espaço tridimensional, uma rotação pode ser encarada como uma transformação que tem como domínio todos os pontos do espaço. Se pensarmos em “rodar um ponto em torno de um eixo” estamos a conceber uma rotação que tem como objecto esse ponto particular. Uma rotação fica completamente definida se lhe for associado um eixo de rotação  $\mathbf{n}$  (2 graus de liberdade) e um ângulo  $\varphi$  (1 grau de liberdade), se assumirmos, por exemplo, o sentido dos ponteiros do relógio como o sentido positivo de todas as rotações. Qualquer rotação possui, portanto, 3 graus de liberdade.

O significado geométrico do ângulo  $\varphi$  e do eixo  $\mathbf{n}$  de uma rotação está representado na Figura (2.10).

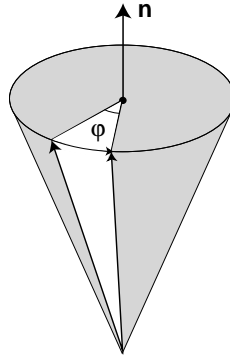


Figura 2.10: Exemplo de uma rotação de um ângulo  $\varphi$  em torno do eixo de rotação  $\mathbf{n}$ .

### 2.5.2 Variação da orientação da câmara na hipótese de pequenas rotações

Consideremos uma câmara de vídeo que se movimenta arbitrariamente ao longo do tempo. Dada a continuidade do movimento, é de esperar que entre tramas consecutivas ocorram pequenas variações da posição da câmara, logo (por terem as translações efeito nulo na orientação) é de esperar que ocorram também pequenas variações da orientação  $\Psi(t)$ , que serão tanto maiores quanto mais rápido for o movimento da câmara ou quanto menor for a frequência de trama. Por conseguinte, estamos em condições de formular uma hipótese que designaremos por *hipótese de pequenas rotações*:

**Hipótese 2.15 (hipótese de pequenas rotações).** *Existe um ângulo  $\varphi_{\max}$  que faz com que, da trama  $t$  para a trama  $t + 1$  de uma sequência de vídeo, o movimento rotacional da câmara possa sempre ser expresso por uma rotação de um ângulo inferior ou igual a  $\varphi_{\max}$  segundo um eixo de rotação arbitrário.*

Na verdade, existem outras hipóteses que melhor descrevem o movimento da câmara entre duas tramas consecutivas, como a consideração de que o eixo de rotação não é

arbitrário mas sim uma pequena variação face ao eixo de rotação anterior (entre  $t - 1$  e  $t$ ) ou, inclusive, considerar um processo de Markov de ordem  $n$  que relacione a probabilidade da orientação no instante  $t$  com as orientações que ocorreram no passado a partir do instante  $t - n$ . No entanto, a hipótese de pequenas rotações permite, devido à sua simplicidade, reduzir a complexidade que advém da dependência da trama actual face às tramas anteriores e, ao mesmo tempo, como veremos, ao considerar com probabilidade nula algumas orientações  $\Psi(t)$  demasiado “afastadas” de  $\Psi(t - 1)$ , permite confinar a estimação da orientação a um conjunto reduzido de possíveis orientações, o que reduz o tempo de processamento do algoritmo e aumenta por conseguinte o seu desempenho.

O teorema seguinte é a principal consequência que advém da hipótese de pequenas rotações:

**Teorema 2.16.** *Sejam  $\Psi(t) = [\alpha(t), \beta(t), \gamma(t)]$  a orientação no instante  $t$  e  $\Psi(t + 1) = [\alpha(t + 1), \beta(t + 1), \gamma(t + 1)]$  a orientação no instante  $t + 1$  para a hipótese de pequenas rotações. São válidas as seguintes propriedades:*

- *O ângulo de torção verifica sempre  $|\gamma(t + 1) - \gamma(t)| \leq \varphi_{\max}$ ;*
- *O ângulo de elevação verifica sempre  $|\beta(t + 1) - \beta(t)| \leq \varphi_{\max}$ ;*
- *Em relação ao ângulo de compasso, se  $|\beta(t + 1) + \beta(t)| \leq \pi - \varphi_{\max}$ , este verifica sempre  $|\alpha(t + 1) - \alpha(t)| \leq \Delta\alpha_{\max}$ , onde:*

$$\cos \Delta\alpha_{\max} = 1 - \frac{\cos \Delta\beta - \cos \varphi_{\max}}{\cos \beta(t) \cos \beta(t + 1)}$$

com  $\Delta\beta = |\beta(t + 1) - \beta(t)|$ .

Se  $|\beta(t + 1) + \beta(t)| > \pi - \varphi_{\max}$ , então qualquer ângulo de compasso é possível no instante  $t + 1$ .

*Demonstração.* A primeira propriedade advém da própria definição de ângulo de torção. O caso mais desfavorável,  $|\gamma(t + 1) - \gamma(t)| = \varphi_{\max}$  ocorre quando de  $t$  para  $t + 1$  a câmara roda um ângulo  $\varphi_{\max}$  num eixo de rotação perpendicular ao plano da imagem.

Já a segunda propriedade é mais difícil de demonstrar. Se o eixo de rotação for tal que o movimento consista numa elevação pura de um ângulo  $\varphi_{\max}$  (isto é, tendo-se  $\alpha(t + 1) = \alpha(t)$  e  $\gamma(t + 1) = \gamma(t)$ ) ter-se-á naturalmente  $|\beta(t + 1) - \beta(t)| = \varphi_{\max}$ . Menos trivial é demonstrar que este é o caso mais desfavorável, embora geometricamente isso seja visível na esfera.

A demonstração da terceira propriedade envolve alguma manipulação algébrica, sendo apresentada na Secção (A.7) do Apêndice. ■

A partir destas propriedades pode limitar-se o intervalo de procura da solução  $\Psi(t + 1)$  quando já se estimou previamente  $\Psi(t)$ , procedendo-se indutivamente para obter estimativas da orientação instantânea.

## Capítulo 3

# Extracção de Informação de uma Imagem

### 3.1 Introdução

Neste capítulo irá tratar-se do processo de *aquisição* da informação relevante contida numa imagem de um mundo de Manhattan. Fazendo uma analogia biológica, este processo é semelhante ao que ocorre quando se observa atentamente uma fotografia procurando aspectos que possam indiciar a posição do fotógrafo no instante da captação.

A informação relevante acerca da “grelha de Manhattan” encontra-se nas arestas dos objectos, pelo que são os *contornos* presentes na imagem que carregam em si a solução para o problema da orientação. Compreende-se assim a necessidade de estabelecer um método capaz de *medir contornos*. O objectivo não é apenas avaliar se um contorno é suficientemente intenso para ser tido em conta, mas também determinar a sua  *direcção*, pois é esta que através da Equação (2.7) permite obter os pontos de fuga e, por conseguinte, a orientação da câmara. Em grande parte dos problemas relacionados com Processamento de Imagem utilizam-se operadores computacionalmente pesados, como as transformadas de Hough, para interpretar linhas de contornos e obter as suas direcções. Neste trabalho, é dado especial ênfase a uma abordagem estatística que conduza a um menor número de cálculos, pelo que se vai evitar o uso deste tipo de operadores<sup>1</sup>, sendo para isso necessário um afastamento da visão biológica para se mergulhar na informação particular que cada pixel da imagem pode fornecer.

### 3.2 O gradiente da intensidade luminosa

#### 3.2.1 Pré-filtragem gaussiana

No problema em estudo utilizam-se imagens monocromáticas, o que apresenta a vantagem, face à utilização de imagens a cores, de exigir um processamento menos

---

<sup>1</sup>Em [1] adoptou-se uma estratégia ainda mais radical, abolindo completamente, para além das transformadas de Hough, a detecção de contornos, no sentido em que esta obriga a tomar decisões prematuras sobre a natureza de cada pixel. No entanto, demonstra-se adiante que a detecção de contornos pode assumir um papel benéfico mantendo a abordagem bayesiana.

pesado na aquisição de informação. Estas imagens são caracterizadas por uma matriz de intensidade luminosa, em que a cada elemento se associa o nível de cinzento<sup>2</sup> do pixel respectivo.

Antes de se iniciar qualquer procedimento relacionado com os contornos da imagem, é usual começar-se por aplicar um filtro que rejeite frequências elevadas. O objectivo desta pré-filtragem consiste em remover o ruído de alta frequência que surge sob a forma de “pixels isolados”, isto é, com valores de intensidade luminosa díspares daqueles que surgem nos pixels vizinhos. Em grande parte dos casos o aparecimento destes pixels isolados prende-se com a existência de fontes de ruído causadas por:

- *Ruído na captação da imagem.* São originados pelos mais diversos factores, ocorrendo durante a captação da imagem na câmara fotográfica analógica ou digital;
- *Ruído na quantificação da intensidade luminosa.* Os erros de quantificação ocorrem quase sempre na proximidade de contornos, e são originados no processo de digitalização quando o nível de intensidade luminosa num pixel é estimado através de uma média ponderada da intensidade em pontos vizinhos. Podem ser atenuados se se utilizar uma resolução mais elevada;
- *Ruído de compressão.* Estes erros ocorrem apenas quando a imagem é comprimida com perda de qualidade, o que acontece por exemplo em ficheiros de imagem com o formato JPEG. Neste caso, formam-se na imagem comprimida pequenas “zonas nebulosas”, podendo estas ser suprimidas ou atenuadas com uma boa pré-filtragem.

Como se pretende rejeitar as frequências altas, isso terá como consequência uma suavização<sup>3</sup> da imagem. Implementa-se quase sempre um filtro gaussiano para cumprir este objectivo. A função gaussiana em duas dimensões é dada por:

$$G_{\sigma}(x, y) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[ -\frac{(x^2 + y^2)}{2\sigma^2} \right]$$

Se discretizarmos esta função a partir de valores inteiros de  $x$  e  $y$  no conjunto  $\{-\frac{N-1}{2}, \dots, -1, 0, 1, \dots, \frac{N-1}{2}\}$  podemos representar uma aproximação da função gaussiana por uma matriz  $N \times N$ , com  $N$  ímpar, que designamos por *máscara gaussiana* e podemos representar por  $\mathbf{G}_{\sigma}(N)$ . Por outro lado, designamos por  $\mathbf{I}$  a matriz de intensidade luminosa, que supomos normalizada no intervalo  $[0, 1]$ . A pré-filtragem gaussiana é realizada através da convolução  $\mathbf{I} * \mathbf{G}_{\sigma}(N)$ .

Na Figura (3.1) está representada uma imagem antes e depois de lhe ser aplicado um filtro gaussiano a partir de uma máscara  $\mathbf{G}_{\sigma}(N)$  com  $\sigma = 1.0$  e  $N = 7$ , sendo evidente o efeito de suavização.

### 3.2.2 Filtros de Prewitt e de Sobel

O objectivo crucial da extracção de informação é caracterizar eficazmente os contornos de uma imagem, dando assim um papel de destaque à teoria da detecção de

<sup>2</sup>Existem tipicamente 256 níveis de intensidade luminosa no processo de quantificação.

<sup>3</sup>Referida na literatura anglo-saxónica como “smoothing”.



Figura 3.1: Fotografia da Rua Augusta, na Baixa Pombalina (a) antes da pré-filtragem (b) após suavização com um filtro gaussiano com  $\sigma = 1.0$  e  $N = 7$ .

contornos. Mais que um detector, pretende-se implementar um *medidor* de contornos, o que pode ser feito recorrendo a técnicas que aproximam o gradiente da intensidade luminosa, estimando para cada ponto<sup>4</sup>:

- O módulo associado ao gradiente nesse ponto, que fornece informação acerca da maior ou menor probabilidade de este pertencer a um contorno e, se pertencer, da *intensidade* desse contorno;
- O ângulo do gradiente, que fornece informação específica acerca da *direcção* do contorno.

Como a detecção de contornos não é o objectivo último deste estudo, apenas são referidos de forma sucinta alguns aspectos que foram tidos em conta e que tiveram utilidade na resolução do problema proposto, podendo obter-se informações mais detalhadas sobre o assunto em [6], [7] e [8].

Alguns métodos, que abordamos na Secção (B.2) do Apêndice, misturam a filtragem gaussiana e a aproximação do gradiente numa só máscara convolutiva, mas para já assume-se que a matriz de intensidade luminosa  $\mathbf{I}$  foi previamente suavizada com um filtro de Gauss, tal como descrito na secção anterior.

Assim, o gradiente consiste numa matriz em que cada elemento é um vector de  $\mathbb{R}^2$  que tem como coordenadas as derivadas parciais nas direcções horizontal e vertical da função de intensidade luminosa. Alternativamente, podemos representá-lo como uma matriz complexa  $\nabla \mathbf{I}$  em que o módulo e o argumento de cada elemento correspondem respectivamente ao módulo e ângulo daquela derivada no pixel correspondente.

As aproximações mais simples para o gradiente  $\nabla \mathbf{I}$  podem obter-se mediante a convolução com as *máscaras de Prewitt*:

$$\mathbf{F}_x = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \quad \text{e} \quad \mathbf{F}_y = -\mathbf{F}_x^\top \quad (3.1)$$

---

<sup>4</sup>Utiliza-se um conjunto discreto de pontos em que a cada elemento corresponde um pixel da imagem. As coordenadas de cada ponto são dadas pelo centro do respectivo pixel.

vindo então:

$$\nabla \mathbf{I} = \mathbf{F}_x * \mathbf{I} + j\mathbf{F}_y * \mathbf{I} = (\mathbf{F}_x + j\mathbf{F}_y) * \mathbf{I}$$

O módulo e ângulo do gradiente em cada ponto são dados respectivamente por:

$$\begin{cases} \text{mod } \nabla \mathbf{I}(x, y) &= \sqrt{[\nabla \mathbf{I}_x(x, y)]^2 + [\nabla \mathbf{I}_y(x, y)]^2} \\ \arg \nabla \mathbf{I}(x, y) &= \arctan \frac{\nabla \mathbf{I}_y(x, y)}{\nabla \mathbf{I}_x(x, y)} \end{cases} \quad (3.2)$$

onde  $\nabla \mathbf{I}_x = \mathbf{F}_x * \mathbf{I}$  e  $\nabla \mathbf{I}_y = \mathbf{F}_y * \mathbf{I}$  são respectivamente as intensidades do gradiente nas direcções horizontal e na vertical.

O método de Sobel é análogo mas utiliza em vez das máscaras  $\mathbf{F}_x$  e  $\mathbf{F}_y$  da Equação (3.1) as *máscaras de Sobel*:

$$\mathbf{F}_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad \text{e} \quad \mathbf{F}_y = -\mathbf{F}_x^\top$$

Empiricamente percebe-se que o filtro de Sobel é uma alternativa ao filtro de Prewitt que visa dar mais ênfase a variações da intensidade luminosa na horizontal e na vertical do que em direcções mais próximas da diagonal.

Os filtros de Prewitt e de Sobel pertencem a uma classe genérica de filtros que utiliza máscaras do tipo:

$$\mathbf{F}_x = \begin{bmatrix} -1 & 0 & 1 \\ -a & 0 & a \\ -1 & 0 & 1 \end{bmatrix} \quad \text{e} \quad \mathbf{F}_y = -\mathbf{F}_x^\top$$

Demonstra-se na Secção (B.1) do Apêndice que o uso deste tipo de filtros na aproximação do gradiente conduz a valores do módulo do gradiente limitados superiormente por:

$$|\nabla \mathbf{I}(x, y)| \leq \sqrt{2(1+a)^2 + 2}$$

Esta relação será particularmente útil quando descrevermos o processo de quantificação do módulo do gradiente.

Verifica-se experimentalmente que é muito vantajosa a inclusão de pré-filtragem gaussiana na implementação de medidores de contornos baseados em filtros de Prewitt e de Sobel, principalmente quando é importante uma estimação precisa do ângulo do gradiente, o que é o caso. A razão para isso prende-se com o facto de, sem esta pré-filtragem, o detector de contornos adquirir um carácter demasiado “local”, pois o comprimento das máscaras convolutivas é de  $3 \times 3$ . Ao suavizar-se a imagem com uma máscara gaussiana de comprimento razoável ( $N \geq 7$ ), atenua-se este carácter local, correlacionando os pixels vizinhos.

Efectuou-se um teste comparativo a vários medidores de contornos, alguns mais sofisticados que os de Prewitt e de Sobel, para avaliar a precisão na estimação de  $\arg \nabla \mathbf{I}$ , utilizando-se como imagem de teste a fotografia de uma aresta recta com uma inclinação de  $79^\circ$  com a vertical, que se representa na Figura (3.2a). Os resultados deste teste, bem como uma descrição sucinta do funcionamento de cada medidor de contornos,

apresentam-se na Secção (B.2) do Apêndice. Desta experiência concluiu-se que o melhor desempenho foi o do medidor de contornos baseado no filtro de Sobel, estando os resultados obtidos para o ângulo do gradiente representados no histograma da Figura (3.2b), onde figuram apenas os valores do ângulo do gradiente em pontos que pertencem a contornos.

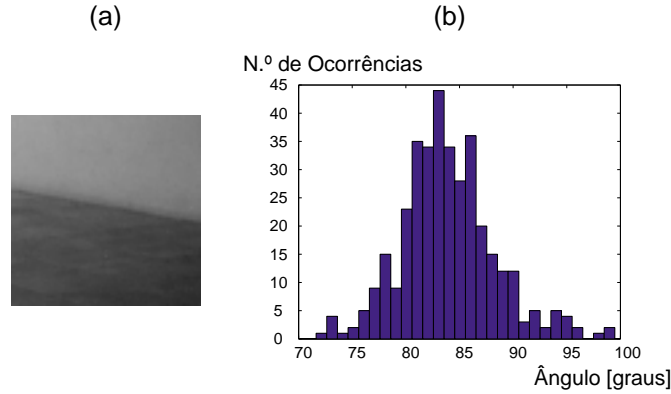


Figura 3.2: (a) Fotografia de um contorno inclinado  $79^\circ$  com a vertical e (b) histograma dos resultados estimados para o ângulo do gradiente em pontos deste contorno. Utilizou-se um medidor de contornos com filtros de Sobel.

### 3.3 Quantificação do módulo do gradiente

#### 3.3.1 Motivação

O módulo do gradiente dá uma ideia da probabilidade de cada ponto pertencer a um contorno e, em caso afirmativo, da intensidade desse contorno. Veremos adiante que é útil dispôr de um conjunto de valores discreto para o módulo do gradiente. Destaca-se duas razões que nos impelem a quantificar esta grandeza:

- Valores discretos tornam menos pesada a computação, aumentando o desempenho final do algoritmo;
- Pode definir-se funções probabilísticas num domínio discreto de forma mais fácil e compreensível que defini-las em intervalos contínuos de comprimento variável. Isto será posto em evidência no próximo capítulo.

O processo de discretização do módulo do gradiente designa-se por *quantificação*. A este processo está associada uma função de quantificação que designamos por  $q(x)$ , e que toma valores no conjunto  $X = \{x_1, x_2, \dots, x_N\}$  onde  $N$  é o número de níveis de quantificação. Tem-se assim, para um ponto  $\mathbf{u}$  da imagem:

$$E_{\mathbf{u}} = q[\text{mod } \nabla \mathbf{I}(\mathbf{u})]$$

onde  $\text{mod } \nabla \mathbf{I}(\mathbf{u})$  é o módulo do gradiente em  $\mathbf{u}$  e  $E_{\mathbf{u}}$  é o seu valor após a quantificação. Tem-se naturalmente  $E_{\mathbf{u}} \in X$ .

Nesta secção irá ser escolhida a função de quantificação  $q(x)$  e justificar essa escolha.



### 3.3.2 Quantificação uniforme

O processo de quantificação mais simples é aquele descrito pela função:

$$q(x) = \left\lceil 1 + N \frac{x}{x_{\max}} \right\rceil$$

e que se designa por *quantificação uniforme*. Nesta função,  $N$  é o número de níveis de quantificação e  $x_{\max}$  é o valor máximo que  $x$  pode tomar<sup>5</sup>.

Claramente se verifica que não deve ser usada esta função de quantificação no caso em estudo, uma vez que numa imagem ocorrem com muito mais frequência valores baixos de  $\text{mod } \nabla \mathbf{I}(\mathbf{u})$  do que valores altos, pelo que os níveis de quantificação mais elevados raramente iriam ser utilizados. Por outro lado, é difícil determinar  $x_{\max}$ , dependendo este parâmetro do medidor de contornos e da utilização ou não de pré-filtragem gaussiana. Uma forma de ultrapassar este problema mantendo a quantificação uniforme consiste em introduzir *saturação*, o que implica redefinir a função  $q(x)$  como:

$$q(x) = \begin{cases} \left\lceil 1 + (N-1) \frac{x}{x_{\text{sat}}} \right\rceil & \text{se } x < x_{\text{sat}} \\ N & \text{se } x \geq x_{\text{sat}} \end{cases}$$

onde  $x_{\text{sat}}$  é um valor de saturação convenientemente escolhido, a partir do qual começa o intervalo correspondente ao nível  $N$ . Define-se ainda o *índice de saturação*  $\epsilon_{\text{sat}}$  por:

$$\epsilon_{\text{sat}} = \frac{x_{\max} - x_{\text{sat}}}{x_{\max}} \quad \Rightarrow \quad x_{\text{sat}} = x_{\max}(1 - \epsilon_{\text{sat}})$$

Aqui,  $x_{\max}$  não é necessariamente o valor máximo que  $x$  pode tomar mas um majorante do conjunto destes valores, devendo o parâmetro  $\epsilon_{\text{sat}}$  ser otimizado experimentalmente em função desse valor. Existem duas abordagens possíveis para escolher o valor de  $x_{\max}$ :

- *Determinar analiticamente um majorante do conjunto de valores possíveis para  $\text{mod } \nabla \mathbf{I}(\mathbf{u})$ .* Na Secção (B.1) do Apêndice demonstra-se que majorantes para o caso de se usarem detectores de contornos de Prewitt ou Sobel são dados pela expressão:

$$x_{\max} = \sqrt{2(1+a)^2 + 2} \quad (3.3)$$

com  $a = 1$  para filtros de Prewitt e  $a = 2$  para filtros de Sobel. Com a utilização de pré-filtragem gaussiana, estes valores de  $x_{\max}$  requerem um nível de saturação  $\epsilon_{\text{sat}}$  elevado.

- *Escolher como majorante o valor do gradiente no ponto da imagem em que ele é máximo*, ou seja,  $x_{\max} = \max_{\mathbf{u}} \text{mod } \nabla \mathbf{I}(\mathbf{u})$ . Este método apresenta a desvantagem de fazer depender a qualidade da quantificação do contraste da imagem. De facto, para imagens de grande contraste, ter-se-á  $x_{\max}$  elevado, pelo que muitos contornos mais fracos (mas que podem ter importância a nível local) corresponderão a níveis de quantização demasiado baixos, para um  $\epsilon_{\text{sat}}$  fixo. Este fenómeno acontece muito

---

<sup>5</sup>O símbolo “[.]” significa “o maior inteiro não superior a”.

em fotografias diurnas onde surgem sombras e reflexos luminosos. Em [8] sugere-se, para ultrapassar este problema, uma divisão da imagem em várias regiões a que corresponderão diferentes valores de  $x_{\max}$ .

No histograma da Figura (3.3) apresenta-se a frequência com que ocorre cada nível de quantificação para a imagem de teste da Figura (3.1), utilizando filtros de Sobel, pré-filtragem gaussiana e quantificação uniforme com saturação  $\epsilon_{\text{sat}} = 0.7$  e  $x_{\max} = \sqrt{20}$ . Pode ver-se que a quantificação uniforme apresenta uma larga ineficiência nos níveis intermédios, que apresentam uma frequência de ocorrência muito reduzida quando comparados com o nível mais baixo.

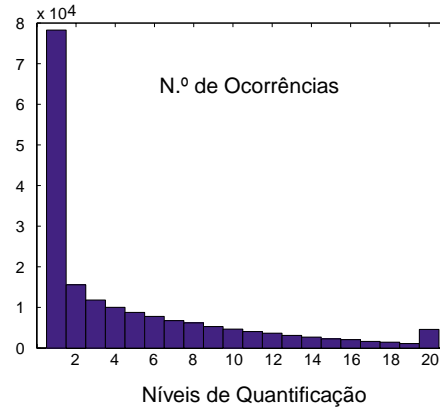


Figura 3.3: Ocorrência dos vários níveis de quantificação para um esquema de quantificação uniforme com saturação.

### 3.3.3 Quantização não uniforme com saturação

A forma mais evidente de melhorar a eficiência do quantificador consiste em utilizar um esquema de quantificação não uniforme, que permita a inclusão de mais níveis para valores baixos de  $x$  e menos níveis para valores mais elevados. A função de quantificação é neste caso dada por:

$$q(x) = \begin{cases} \left[ 1 + (N - 1) f\left(\frac{x}{x_{\text{sat}}}\right) \right] & \text{se } x < x_{\text{sat}} \\ N & \text{se } x \geq x_{\text{sat}} \end{cases}$$

onde  $f(t)$  é uma função definida e com valores em  $[0, 1]$ , apresentando uma curvatura logarítmica, como por exemplo  $f(t) = \log_2(1 + t)$ .

Experimentalmente obtiveram-se bons resultados utilizando funções

$$f_n(t) = f^{(n)}(t) = \underbrace{f \circ f \circ \dots \circ f}_{n \text{ vezes}} = \log_2(1 + f^{(n-1)}(t))$$

com  $n = 4$  e  $n = 5$ .

Pode ver-se na Figura (3.4) o histograma que representa a frequência de utilização de cada nível para a mesma imagem do caso anterior, mas utilizando quantificação não uniforme logarítmica através da função  $f_5(t)$ .

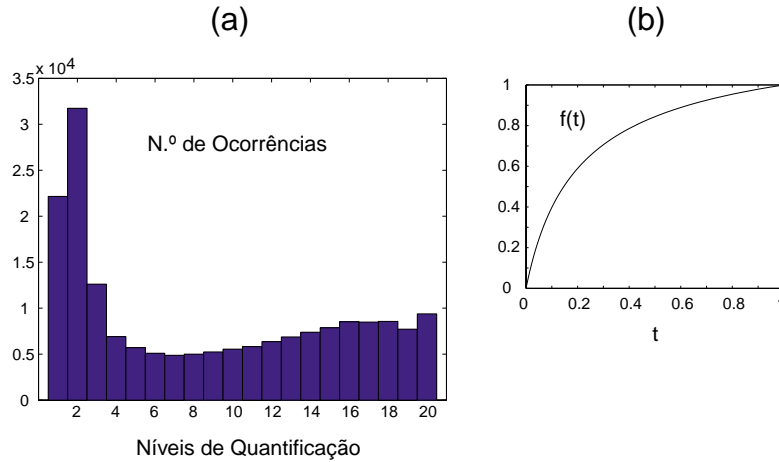


Figura 3.4: Ocorrência dos vários níveis de quantificação para um esquema de quantificação logarítmica com saturação, utilizando a função de quantificação  $f_5(t)$  representada em (b).

## 3.4 Exclusão selectiva de pixels

### 3.4.1 Exclusão de pixels com módulo de gradiente reduzido

Nem todos os pixels de uma imagem contêm informação relevante para o problema da estimação da orientação. Em geral, um ponto  $\mathbf{u}$  que esteja visivelmente fora de um contorno não necessita sequer de ser considerado, pelo que nunca se torna necessário, durante todo o algoritmo, utilizar nos cálculos o valor do gradiente no pixel correspondente. Uma imagem convencional caracteriza-se pela predominância de pixels nestas circunstâncias, para os quais a intensidade do gradiente  $\text{mod } \nabla \mathbf{I}(\mathbf{u})$  é tão baixa que pode à partida assumir-se como estando fora de um contorno. Por conseguinte, é vantajoso considerar um valor limite para esta intensidade, que designaremos por  $x_{\text{lim}}$ , que permita *excluir* qualquer ponto  $\mathbf{u}$  que verifique  $\text{mod } \nabla \mathbf{I}(\mathbf{u}) < x_{\text{lim}}$ .

Com base nestas considerações, pode introduzir-se mais um parâmetro no sistema, que designaremos por *índice limiar de exclusão*  $\epsilon_{\text{lim}}$  e que é dado por:

$$\epsilon_{\text{lim}} = \frac{x_{\text{lim}}}{x_{\text{max}}}$$

Apesar deste procedimento se aproximar da detecção de contornos, deve notar-se que nenhuma decisão é tomada neste sentido em relação aos pixels que não são excluídos, cuja natureza continua em aberto. É também importante ter em conta que a exclusão selectiva de pixels deve ser sempre efectuada *antes da quantificação*, evitando-se assim o processamento supérfluo de quantificar o módulo do gradiente em pontos que depois são suprimidos.

Após algumas experiências, verificou-se que a utilização de  $\epsilon_{\text{lim}} = 0.05$  permite obter bons resultados quando se utilizam filtros de Prewitt ou Sobel com pré-filtragem gaussiana e com  $x_{\text{max}}$  dado pela Equação (3.3), tendo-se eliminado desta forma 60.0% dos pixels da imagem representada na Figura (3.1).

### 3.4.2 Exclusão de pixels redundantes na vizinhança de contornos (non maxima supression)

A aplicação de máscaras convolutivas para suavizar e aproximar o gradiente da imagem faz com que os pontos que residem próximos de um contorno mas que não se encontram no seu *epicentro*<sup>6</sup> acabem também por apresentar módulos de gradiente elevados. A presença destes pontos na matriz  $\nabla \mathbf{I}$  apresenta algumas desvantagens importantes:

- Introduzem redundância na informação a ser processada no algoritmo de estimação da orientação, pelo facto de existirem pixels mais próximos do epicentro do contorno que também contêm informação acerca da direcção desse contorno;
- Quanto mais longe do epicentro do contorno um ponto  $\mathbf{u}$  estiver, maior é o erro cometido, em média, na aproximação da direcção normal do contorno pelo valor de  $\arg \nabla \mathbf{I}(\mathbf{u})$ ; isto faz com que a informação fornecida por estes pixels, além de supérflua, possa ainda ser prejudicial.

A eliminação de pontos redundantes na vizinhança de contornos é feita baseando-se numa técnica desenvolvida por Canny que se designa na literatura anglo-saxónica por *non maxima supression*. O algoritmo que desenvolvemos para o efeito é o que se apresenta de seguida:

**Algoritmo 3.1.** *Para cada pixel  $\mathbf{u}$  no interior da imagem:*

1. *Obtém-se o argumento do gradiente  $\arg \nabla \mathbf{I}(\mathbf{u})$ ;*
2. *Dado esse ângulo, escolhe-se na vizinhança  $3 \times 3$  de  $\mathbf{u}$  dois pixels à esquerda e dois à direita tais que o ângulo obtido no passo anterior esteja (módulo  $\pi$ ) entre esses dois pixels;*
3. *Obtém-se o módulo do gradiente em  $\mathbf{u}$ ,  $\text{mod} \nabla \mathbf{I}(\mathbf{u})$ , e em cada um dos quatro pixels escolhidos no passo anterior;*
4. *Através de uma interpolação do módulo do gradiente nos dois pixels à esquerda e outra nos dois pixels à direita, pesada com a comparação da localização desses pixels com  $\arg \nabla \mathbf{I}(\mathbf{u})$  obtida no passo (1), obtém-se um módulo do gradiente à esquerda de  $\mathbf{u}$ ,  $\text{mod} \nabla \mathbf{I}(\mathbf{u}^-)$  e um módulo do gradiente à direita de  $\mathbf{u}$ ,  $\text{mod} \nabla \mathbf{I}(\mathbf{u}^+)$ ;*
5. *Se  $\text{mod} \nabla \mathbf{I}(\mathbf{u}^-) > \text{mod} \nabla \mathbf{I}(\mathbf{u})$  ou  $\text{mod} \nabla \mathbf{I}(\mathbf{u}^+) > \text{mod} \nabla \mathbf{I}(\mathbf{u})$ , exclui-se  $\mathbf{u}$ . Caso contrário não se exclui, considerando-se que  $\mathbf{u}$  está no epicentro do contorno.*

A aplicação deste algoritmo faz com que os contornos, uma vez eliminados os seus pontos vizinhos, passem a ter uma largura aproximada de 1 pixel, reduzindo assim drasticamente a quantidade de informação que será processada na etapa seguinte.

---

<sup>6</sup>Por “epicentro” entenda-se a zona do contorno onde é mais repentina a variação da intensidade luminosa.

### 3.4.3 Resultados da exclusão selectiva

Uma forma de medir o rendimento da exclusão selectiva numa determinada imagem consiste em calcular o parâmetro  $\eta_{exc}$  dado por:

$$\eta_{exc} = \frac{\text{número de pixels excluídos}}{\text{número de pixels da imagem}} \times 100\% \quad (3.4)$$

Para a imagem da Figura (3.1), onde a medição e quantificação do gradiente foi efectuada como sugerido nas secções anteriores, conseguiu obter-se um rendimento na exclusão  $\eta_{exc} = 85.1\%$ , considerando-se como informação relevante apenas 14.9% dos pixels da imagem. Sem esta supressão dificilmente se conseguiria conceber um algoritmo de estimação da orientação computacionalmente realizável em tempos aceitáveis.

A Figura (3.5) mostra a intensidade do módulo do gradiente em cada ponto antes e depois da exclusão selectiva. Os pontos mais escurecidos da imagem são aqueles cujo módulo é mais elevado, logo, pertencentes a contornos mais fortes. Pontos que foram excluídos representam-se a branco, confundindo-se com zonas da imagem onde não existem contornos.



Figura 3.5: Intensidade do módulo do gradiente em cada ponto de uma imagem (a) antes da exclusão selectiva de pontos (b) após exclusão selectiva de pontos.

## 3.5 Funcionamento geral da malha de extracção de informação

Tendo em conta o estudo realizado nas secções anteriores, pode resumir-se aqui o funcionamento da malha de extracção de informação, cuja implementação servirá os algoritmos de treino e de estimação da orientação, como veremos no capítulo seguinte.

O esquema ilustrativo da malha de extracção de informação está representado na Figura (3.6).

O algoritmo de extracção é o seguinte:

**Algoritmo 3.2.** *Para uma imagem com matriz de intensidade luminosa  $I$ :*

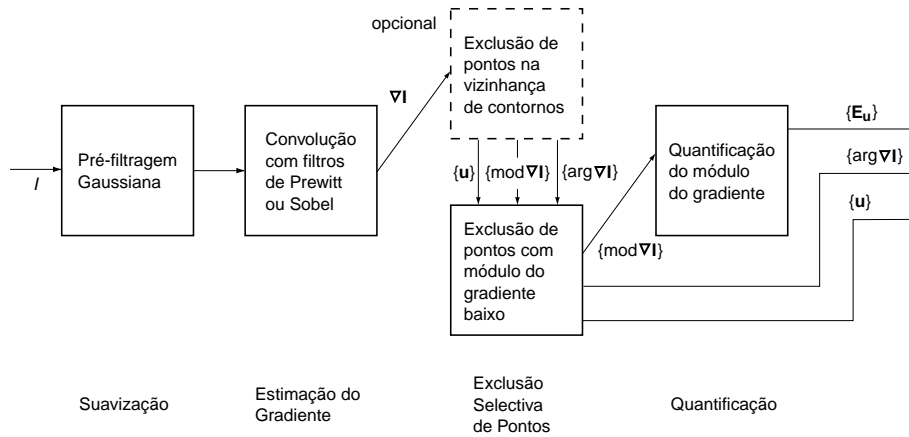


Figura 3.6: Esquema ilustrativo da malha de extração de informação de uma imagem.

1. É feita uma suavização da imagem através de pré-filtragem gaussiana com, por exemplo,  $N = 7$  e  $\sigma = 1.0$ ;
2. Convolui-se  $I$  com as máscaras de Prewitt ou Sobel, obtendo-se o gradiente na forma da matriz complexa  $\nabla I$ ;
3. Opcionalmente, exclui-se a informação relativa aos pixels na vizinhança de contornos, obtendo-se os conjuntos  $\{u\}$ ,  $\{\text{mod } \nabla I(u)\}$  e  $\{\text{arg } \nabla I(u)\}$ . Este passo é executado através do Algoritmo (3.1);
4. Exclui-se a informação relativa aos pixels com módulo de gradiente abaixo de um valor mínimo definido pelo parâmetro  $\epsilon_{\text{lim}}$ , reduzindo-se o número de elementos dos conjuntos  $\{u\}$ ,  $\{\text{mod } \nabla I(u)\}$  e  $\{\text{arg } \nabla I(u)\}$ ;
5. São quantificados os valores do módulo do gradiente em cada ponto  $u$ ,  $\{\text{mod } \nabla I(u)\}$ , preferencialmente através de quantificação logarítmica (com, por exemplo,  $n = 5$ ). Obtem-se assim o conjunto  $\{E_u\}$ , cujos elementos tomam valores em  $X = \{x_1, x_2, \dots, x_N\}$ .

O funcionamento de cada módulo da malha foi explicado nas secções anteriores deste capítulo. Experimentalmente, os resultados obtidos foram muito satisfatórios, como já se referiu em cada uma destas secções, com uma ligeira excepção da estimação da direcção do gradiente, que não pode considerar-se extremamente precisa, como vimos, mas que na globalidade não vai prejudicar o algoritmo final de estimação da orientação, embora dificulte tarefas como a localização na imagem de objectos parasita. Tudo isto vai ser objecto de discussão no próximo capítulo.

## Capítulo 4

# Estimação da Orientação da Câmara

### 4.1 Introdução

No capítulo anterior estudou-se formas de extrair a informação relevante de uma imagem. Vai agora analisar-se a forma como essa informação é processada para, após tratamento estatístico, poder estimar-se a orientação da câmara.

O primeiro passo consiste em definir e estimar o valor óptimo de algumas probabilidades que se supõe constantes, isto é, independentes da imagem que se está a analisar, e que por esse motivo se designam por probabilidades *a priori*. A sua aprendizagem é efectuada numa etapa de *treino*, que é executada apenas uma vez. Conhecidas estas probabilidades, descreve-se um modelo bayesiano que permitirá, com a introdução de algumas aproximações, a implementação de algoritmos de estimação da orientação.

### 4.2 Treino e estimação de probabilidades a priori

#### 4.2.1 As probabilidades a priori

A informação fornecida à entrada da malha de estimação é a que se disponibiliza na saída da malha de extracção da Figura (3.6) e que aqui recapitulamos:

- O conjunto  $\{\mathbf{u}\}$  com as coordenadas de todos os pontos considerados relevantes;
- O valor quantificado do módulo do gradiente para estes pontos, expresso no conjunto  $\{E_{\mathbf{u}}\}$ , em que cada elemento toma valores no conjunto  $X = \{x_1, x_2, \dots, x_N\}$ ;
- O ângulo do gradiente em cada ponto, fornecido pelo conjunto  $\{\arg \nabla \mathbf{I}(\mathbf{u})\}$ .

Por outro lado, tendo em conta aquilo que já vimos nos capítulos anteriores, podemos definir o conjunto  $M = \{m_1, m_2, \dots, m_N\}$  dos modelos possíveis a atribuir a cada pixel da imagem. À semelhança de [1], utiliza-se  $N = 5$  modelos de pixel, estando o significado de cada um deles descrito na Tabela (4.1).

A etapa de treino tem como objectivo estimar os seguintes parâmetros:

Modelos de pixel $m$	
1	Pixel de um contorno consistente com o eixo $x$
2	Pixel de um contorno consistente com o eixo $y$
3	Pixel de um contorno consistente com o eixo $z$
4	Pixel de um contorno parasita
5	Pixel fora de qualquer contorno

Tabela 4.1: Conjunto  $M = \{1, 2, 3, 4, 5\}$  de modelos possíveis para cada pixel da imagem de um mundo de Manhattan e respectiva descrição.

- Para cada  $m_j \in M$ , a probabilidade  $P_M(m_j) = \text{Prob}\{m_{\mathbf{u}} = m_j\}$  de se ter o modelo  $m_j$  para o pixel  $\mathbf{u}$ ;
- Para cada  $x_j \in X$ , as probabilidades:
  - $P_{ON}(x_j) = \text{Prob}\{E_{\mathbf{u}} = x_j | m_{\mathbf{u}} \in \{1, 2, 3, 4\}\}$  de se ter o nível de quantificação  $x_j$  para o módulo do gradiente *dado que* o pixel  $\mathbf{u}$  pertence a um contorno;
  - $P_{OFF}(x_j) = \text{Prob}\{E_{\mathbf{u}} = x_j | m_{\mathbf{u}} = 5\}$  de se ter o mesmo nível de quantificação  $x_j$  *dado que*  $\mathbf{u}$  não pertence a qualquer contorno;
- A função de distribuição contínua  $P_{ang}(t)$ , onde  $t$  é a variável aleatória que representa o erro angular cometido (módulo  $\pi$ ) na aproximação da direcção normal de um contorno (num ponto  $\mathbf{u}$ ) pela estimativa do ângulo do gradiente  $\arg \nabla \mathbf{I}(\mathbf{u})$ .

#### 4.2.2 Estimação das probabilidades $P_{ON}$ e $P_{OFF}$

Uma forma de estimar as probabilidades  $P_{ON}(x_j)$  e  $P_{OFF}(x_j)$  consiste em utilizar um conjunto extenso de imagens cujos contornos foram previamente localizados (manualmente ou utilizando um bom detector de contornos binário) e utilizar esta informação na fase de aprendizagem. O resultado para cada imagem é uma matriz binária  $\{B_{\mathbf{u}}\}$  cujos elementos são 1 em pontos de contorno e 0 nos restantes pontos.

De seguida pode extrair-se informação de cada imagem utilizando o Algoritmo (3.2) *sem a exclusão selectiva de pontos*, isto é, sem efectuar os passos (3) e (4), obtendo-se assim a matriz  $\{E_{\mathbf{u}}\}$ .

As probabilidades  $P_{ON}$  e  $P_{OFF}$  podem então ser aproximadas através das expressões:

$$\begin{cases} P_{ON}(x_j) = \frac{\text{Prob}\{B_{\mathbf{u}} = 1, E_{\mathbf{u}} = x_j\}}{\text{Prob}\{B_{\mathbf{u}} = 1\}} = \frac{n\{\mathbf{u} : B_{\mathbf{u}} = 1 \wedge E_{\mathbf{u}} = x_j\}}{n\{\mathbf{u} : B_{\mathbf{u}} = 1\}} \\ P_{OFF}(x_j) = \frac{\text{Prob}\{B_{\mathbf{u}} = 0, E_{\mathbf{u}} = x_j\}}{\text{Prob}\{B_{\mathbf{u}} = 0\}} = \frac{n\{\mathbf{u} : B_{\mathbf{u}} = 0 \wedge E_{\mathbf{u}} = x_j\}}{n\{\mathbf{u} : B_{\mathbf{u}} = 0\}} \end{cases}$$

onde  $n\{\cdot\}$  designa o número de elementos do conjunto  $\{\cdot\}$ .

Se se repetir este procedimento para várias imagens, podem obter-se estimativas razoáveis para  $P_{ON}$  e  $P_{OFF}$ , tanto melhores quanto mais exacto for o detector de contornos que produz a matriz binária  $\{B_{\mathbf{u}}\}$ . No caso em estudo, utilizou-se para este fim



um detector de Canny com histerese, por possuir um bom desempenho e ser relativamente fácil de implementar.

Os histogramas das Figuras (4.1a-d) mostram os valores obtidos para  $P_{ON}$  e  $P_{OFF}$  no final da etapa de treino em duas experiências onde se analisou 30 imagens de teste: numa, a malha de extracção utiliza máscaras de Prewitt para aproximar o gradiente; na outra utilizou-se para esse fim máscaras de Sobel. Em ambos os casos a imagem foi previamente suavizada com um filtro de Gauss com  $N = 7$  e  $\sigma = 1.0$ , tendo-se aplicado um esquema de quantificação logarítmica com  $n = 5$ .

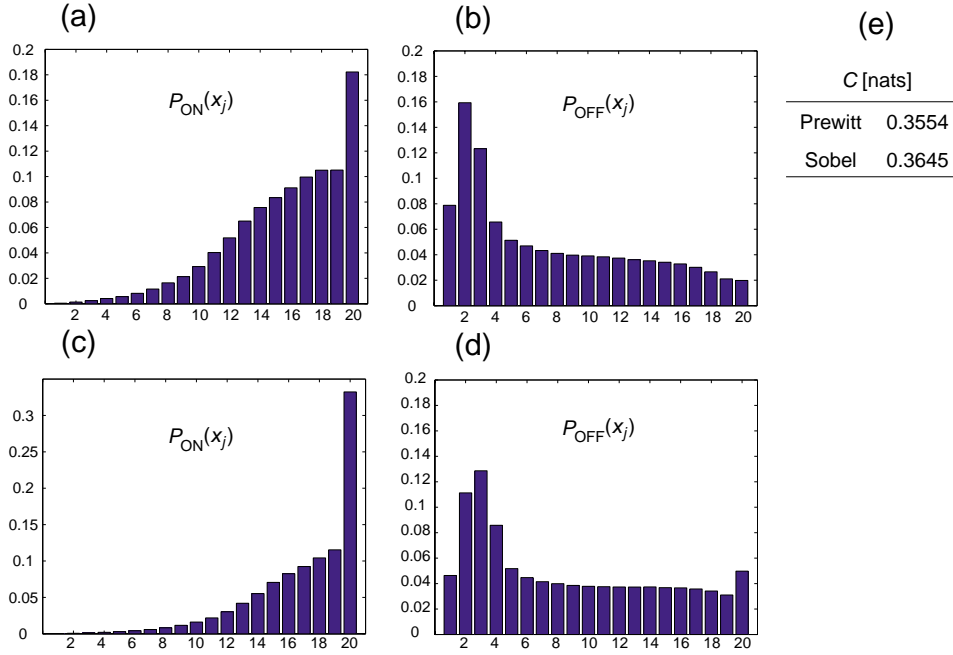


Figura 4.1: Histograma das probabilidades  $P_{ON}$  e  $P_{OFF}$  utilizando (a) e (b) filtros de Prewitt e (c) e (d) filtros de Sobel. Em (e) mostra-se, para cada caso, os valores obtidos para a informação de Chernoff.

Quanto mais diferentes entre si forem as probabilidades  $P_{ON}$  e  $P_{OFF}$ , maior é a certeza sobre a natureza de cada pixel  $\mathbf{u}$  em função do valor de  $E_{\mathbf{u}}$ . Uma boa medida desta diferença, introduzida em [10] é a *informação de Chernoff*, dada por:

$$C(P_{ON}, P_{OFF}) = - \min_{0 \leq \lambda \leq 1} \log \sum_j P_{ON}^\lambda(x_j) P_{OFF}^{1-\lambda}(x_j) \quad [\text{nats}]$$

e que se deverá tentar *maximizar*.

Experimentalmente verifica-se que, utilizando quantificação logarítmica, a informação de Chernoff tende a ser tanto maior quanto maior for o parâmetro de quantificação  $n$ , embora isso não signifique que se devam utilizar valores indiscriminadamente elevados para este parâmetro. Os valores obtidos para a informação de Chernoff nas duas experiências efectuadas estão representados na Figura (4.1e), onde se pode observar o melhor desempenho proporcionado pela utilização de máscaras de Sobel.<sup>1</sup>

<sup>1</sup>Em [1], onde se obteve um resultado inferior para a informação de Chernoff, nomeadamente  $C =$

### 4.2.3 Estimação das probabilidades $P_M$ e modelo para a função de distribuição $P_{ang}$

Para estimar as probabilidades  $P_M(m_j)$ , começa por assumir-se a hipótese:

$$P_M(1) = P_M(2) = P_M(3) = kP_M(4)$$

que resulta da indistinguibilidade dos pontos de fuga posta em evidência na Propriedade (2.8). Esta hipótese conduz-nos a:

$$\sum_{m=1}^4 P_M(m) = (1 + 3k) P_M(4) = 1 - P_M(5) \Leftrightarrow P_M(4) = \frac{1 - P_M(5)}{1 + 3k}$$

Recorrendo de novo à matriz  $\{B_{\mathbf{u}}\}$  temos, por outro lado:

$$P_M(5) = \text{Prob}\{B_{\mathbf{u}} = 0\} = \frac{n\{\mathbf{u} : B_{\mathbf{u}} = 0\}}{n\{\mathbf{u}\}}$$

No caso em estudo tomou-se  $k = \frac{1}{2}$ , o que se revelou uma hipótese razoável<sup>2</sup>. Os resultados obtidos estão representados na Tabela (4.2).

$m_j$	$P_M(m_j)$	
	Prewitt	Sobel
1	0.149	0.138
2	0.149	0.138
3	0.149	0.138
4	0.299	0.276
5	0.253	0.309

Tabela 4.2: Estimativas obtidas para as probabilidades  $P_M$  utilizando filtros de Prewitt e de Sobel na malha de extracção de informação.

Finalmente, a função de distribuição  $P_{ang}(t)$  poderia também ser aprendida na etapa de treino se, para além de um detector de contornos com resposta binária, se usasse um *medidor* capaz de fornecer uma boa estimativa da direcção dos contornos, o que conduziria a uma distribuição aproximadamente normal. No entanto, é possível obter muitas simplificações se utilizarmos o modelo de “caixa” proposto em [1], que define  $P_{ang}(t)$  como<sup>3</sup>:

$$P_{ang}(t) = \begin{cases} \frac{1-\epsilon}{2\tau} & \text{se } t \in [-\tau, \tau] \\ \frac{\epsilon}{\pi-2\tau} & \text{se } t \in ]-\frac{\pi}{2}, -\tau[ \cup ]\tau, \frac{\pi}{2}] \end{cases} \quad (4.1)$$

no domínio  $]-\frac{\pi}{2}, \frac{\pi}{2}]$ .

0.26 nats, afirma-se que outros detectores de contornos mais sofisticados, utilizando imagens a cores, permitem chegar a valores como  $C = 0.51$  nats, mas na prática não surge necessidade de um valor tão elevado.

<sup>2</sup>Este valor de  $k$  equivale a admitir que  $\frac{2}{5} = \frac{1}{1+3k}$  dos contornos da imagem são linhas parasitas.

<sup>3</sup>Na verdade, em [1] a definição da função  $P_{ang}(t)$  é um pouco diferente, pois aí a variável aleatória  $t$  é considerada no domínio  $]-\pi, \pi]$ . No entanto, as duas funções têm o mesmo significado estatístico.

Os valores propostos para os parâmetros  $\epsilon$  e  $\tau$  são (citando [1])  $\epsilon = 0.1$  e  $\tau = 4^\circ$  em fotografias dentro de edifícios e  $\tau = 6^\circ$  em fotografias de exteriores. O gráfico desta função está representado na Figura (4.2), onde é visível o significado destes dois parâmetros.

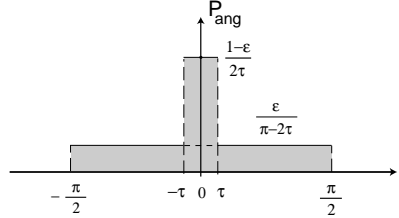


Figura 4.2: Modelo utilizado para a função de distribuição da probabilidade  $P_{ang}$ .

### 4.3 Modelo bayesiano

Com o objectivo de determinar a solução estatisticamente mais provável para a orientação, desenvolveu-se um modelo bayesiano que assenta na teoria exposta em [1] e [2]. Uma forte vantagem desta abordagem face aos métodos usuais consiste em *não forçar decisões prematuras sobre que modelo  $m_1, m_2, \dots, m_5$  corresponde a cada pixel, permitindo que se tenha em conta todas as possíveis interpretações acerca desse pixel*<sup>4</sup>.

Podemos representar por  $\mathbf{E}_{\mathbf{u}}$  a informação extraída do pixel  $\mathbf{u}$ , tendo-se desta forma:

$$\mathbf{E}_{\mathbf{u}} = (E_{\mathbf{u}}, \phi_{\mathbf{u}})$$

onde se fez  $\phi_{\mathbf{u}} = \arg \nabla \mathbf{I}(\mathbf{u}) + k\pi$ , com  $k$  escolhido de modo a ter-se  $\phi_{\mathbf{u}} \in ]-\frac{\pi}{2}, \frac{\pi}{2}]$ .

Uma forma de simplificar os cálculos consiste em admitir que, para qualquer pixel  $\mathbf{u}$ ,  $E_{\mathbf{u}}$  e  $\phi_{\mathbf{u}}$  são estatisticamente independentes, o que nos conduz à factorização seguinte:

$$\text{Prob} \{ \mathbf{E}_{\mathbf{u}} | m_{\mathbf{u}}, \Psi, \mathbf{u} \} = \text{Prob} \{ E_{\mathbf{u}} | m_{\mathbf{u}} \} \cdot \text{Prob} \{ \phi_{\mathbf{u}} | m_{\mathbf{u}}, \Psi, \mathbf{u} \} \quad (4.2)$$

Basicamente, esta hipótese resulta de se admitir que o módulo do gradiente não possui qualquer dependência com a direcção, o que, embora admissível, não é inteiramente verdade, por se utilizarem operadores direccionais no cálculo da matriz  $\nabla \mathbf{I}$ , introduzindo assim dependência entre as duas variáveis aleatórias.

Na Equação (4.2) cada factor relaciona-se com as probabilidades a priori  $P_{ON}$ ,  $P_{OFF}$  e  $P_{ang}$  através das expressões:

$$\text{Prob} \{ E_{\mathbf{u}} | m_{\mathbf{u}} \} = \begin{cases} P_{ON}(E_{\mathbf{u}}) & \text{se } m_{\mathbf{u}} \neq 5 \\ P_{OFF}(E_{\mathbf{u}}) & \text{se } m_{\mathbf{u}} = 5 \end{cases} \quad (4.3)$$

<sup>4</sup>Exceptuam-se os pixels já eliminados no processo de exclusão selectiva, que se consideraram “irrelevantes” e que por isso não são abrangidos por nenhum dos modelos representados na Tabela (4.1).

e

$$\text{Prob}\{\phi_{\mathbf{u}}|m_{\mathbf{u}}, \Psi, \mathbf{u}\} = \begin{cases} P_{ang}[\phi_{\mathbf{u}} - \theta(m_{\mathbf{u}}, \Psi, \mathbf{u}) + k\pi] & \text{se } m_{\mathbf{u}} \in \{1, 2, 3\} \\ U(\phi_{\mathbf{u}}) & \text{se } m_{\mathbf{u}} \in \{4, 5\} \end{cases} \quad (4.4)$$

onde:

- $\theta(m_{\mathbf{u}}, \Psi, \mathbf{u})$  é a direcção normal do contorno obtida analiticamente através da Equação (2.7);
- $k \in \mathbb{Z}$  é escolhido de modo a que o argumento de  $P_{ang}(\cdot)$  esteja contido no intervalo  $]-\frac{\pi}{2}, \frac{\pi}{2}]$ ;
- $U(t) = \frac{1}{\pi}$  é a função de distribuição uniforme no intervalo  $]-\frac{\pi}{2}, \frac{\pi}{2}]$ , o que na expressão mostra a equiprobabilidade de todas as possíveis direcções para o gradiente em pontos que não fazem parte de contornos alinhados com qualquer eixo de Manhattan.

Na abordagem bayesiana que vamos seguir, começa por ser criado um *modelo de mistura* que marginaliza todos os modelos possíveis para o pixel  $\mathbf{u}$ :

$$\text{Prob}\{\mathbf{E}_{\mathbf{u}}|\Psi, \mathbf{u}\} = \sum_{m_{\mathbf{u}}=1}^5 \text{Prob}\{\mathbf{E}_{\mathbf{u}}|m_{\mathbf{u}}, \Psi, \mathbf{u}\} \cdot P_M(m_{\mathbf{u}}) \quad (4.5)$$

O passo seguinte consiste em combinar a informação estatística  $\{\mathbf{E}_{\mathbf{u}}\}$  de todos os pixels, o que se faz assumindo-se que, dada uma orientação  $\Psi$ , a informação  $\{\mathbf{E}_{\mathbf{u}}\}$  é condicionalmente independente de pixel para pixel, o que conduz à expressão:

$$\text{Prob}\{\{\mathbf{E}_{\mathbf{u}}\}|\Psi\} = \prod_{\mathbf{u}} \text{Prob}\{\mathbf{E}_{\mathbf{u}}|\Psi, \mathbf{u}\} \quad (4.6)$$

Esta hipótese de independência condicional negligencia o facto de gradientes em pixels vizinhos estarem fortemente relacionados se se tiver suavizado a imagem, por exemplo, com pré-filtragem gaussiana. No entanto, esta aproximação é muito útil e possibilita a execução computacional de algoritmos baseados neste modelo. Na análise dos resultados deve ter-se em conta o efeito nefasto que ela introduz.

A partir da Equação (4.6), é agora possível, de entre várias orientações da câmara  $\Psi$ , determinar aquela que conduz a uma maior probabilidade  $\text{Prob}\{\{\mathbf{E}_{\mathbf{u}}\}|\Psi\}$ , sendo essa a orientação mais provável, pois de acordo com a fórmula das probabilidades condicionadas:

$$\text{Prob}\{\Psi|\{\mathbf{E}_{\mathbf{u}}\}\} = \frac{\text{Prob}\{\{\mathbf{E}_{\mathbf{u}}\}|\Psi\} \cdot \text{Prob}\{\Psi\}}{\text{Prob}\{\{\mathbf{E}_{\mathbf{u}}\}\}}$$

onde o denominador do segundo membro é constante para todas as orientações  $\Psi$ .

Como tal, a tarefa de estimar a orientação consiste em determinar a estimativa  $\hat{\Psi}$  de *máximo a posteriori* da função  $\text{Prob}\{\{\mathbf{E}_{\mathbf{u}}\}|\Psi\} \cdot \text{Prob}\{\Psi\}$ , designada por *função de verosimilhança*.

Aplicando logaritmos:

$$\begin{aligned}
\hat{\Psi} &= \arg_{\Psi} \max \log [\text{Prob} \{ \{ \mathbf{E}_{\mathbf{u}} \} | \Psi \} \cdot \text{Prob} \{ \Psi \} ] = \\
&= \arg_{\Psi} \max \left( \log \prod_{\mathbf{u}} \text{Prob} \{ \mathbf{E}_{\mathbf{u}} | \Psi, \mathbf{u} \} + \log \text{Prob} \{ \Psi \} \right) = \\
&= \arg_{\Psi} \max \left( \log \text{Prob} \{ \Psi \} + \sum_{\mathbf{u}} \log \text{Prob} \{ \mathbf{E}_{\mathbf{u}} | \Psi, \mathbf{u} \} \right)
\end{aligned}$$

Se não existir nenhuma informação relevante que leve a considerar uma orientação mais provável que outra (como quando se trata de uma única imagem, e não de uma sequência vídeo), esta expressão reduz-se a:

$$\hat{\Psi} = \arg_{\Psi} \max \sum_{\mathbf{u}} \log \text{Prob} \{ \mathbf{E}_{\mathbf{u}} | \Psi, \mathbf{u} \} \quad (4.7)$$

## 4.4 Estimação do ângulo de compasso para uma câmara no plano horizontal

O caso particular em que a câmara está assente sobre o plano horizontal definido por  $\mathbf{x}$  e  $\mathbf{y}$  é assunto de estudo em [1]. Esta referência mostra como se pode implementar um sistema de orientação automática num ambiente urbano através de uma câmara colocada no peito de um indivíduo cego. Assume-se nesta situação que, apesar das oscilações provocadas pelo movimento do indivíduo, a câmara mantém a sua posição aproximadamente horizontal.

Se partirmos desta hipótese, vem imediatamente  $\beta = 0$  e  $\gamma = 0$ , como se pôde ver na Figura (2.7). Assim, apenas é necessário estimar o ângulo de compasso  $\alpha$ , ficando automaticamente definida a orientação  $\hat{\Psi} = (\hat{\alpha}, 0, 0)$ . Na prática, o problema passa a ter apenas 1 grau de liberdade, o que o torna relativamente fácil de resolver.

O algoritmo utilizado para estimar o ângulo de compasso é o que a seguir se reproduz, tendo sido obtido a partir do modelo bayesiano descrito na secção anterior.

**Algoritmo 4.1.** *Dada uma imagem  $\mathbf{I}$ :*

1. *É aplicada a malha de extracção de informação através do Algoritmo (3.2), obtendo-se  $\{\mathbf{u}\}$  e  $\{\mathbf{E}_{\mathbf{u}}\}$ ;*
2. *Para uma colecção de ângulos  $\alpha$  entre  $-45^\circ$  e  $45^\circ$ :*
  - (a) *Obtêm-se os pontos de fuga  $\mathbf{F}_x$  e  $\mathbf{F}_y$  através da Equação (2.5);*
  - (b) *Para cada pixel  $\mathbf{u}$  obtêm-se os valores de  $\theta_x(\mathbf{u})$  e  $\theta_y(\mathbf{u})$  através da Equação (2.7);*
  - (c) *Com  $\Psi = (\alpha, 0, 0)$  obtém-se  $F(\alpha) = \sum_{\mathbf{u}} \log \text{Prob}(\mathbf{E}_{\mathbf{u}} | \Psi, \mathbf{u})$  através das Equações (4.2) a (4.5);*
3. *Obtém-se a estimativa do máximo a posteriori  $\hat{\alpha} = \arg_{\alpha} \max F(\alpha)$ .*

O passo (2) do algoritmo sugere, como primeira abordagem, um *método de força bruta* para a estimativa de máxima verosimilhança, que consiste em abranger na colecção de ângulos  $\alpha$  todos os valores entre  $-45^\circ$  e  $45^\circ$  com incrementos fixos de, por exemplo,  $1^\circ$ , o que permite atingir uma precisão razoável considerando as várias aproximações admitidas pelo modelo bayesiano.

Neste caso particular em que apenas se pretende estimar o ângulo de compasso, a utilização do método de força bruta é viável, embora em termos de eficiência seja recomendável utilizar técnicas mais elaboradas para maximizar a função de verosimilhança. Uma técnica desenvolvida neste estudo e que se revelou particularmente eficiente baseia-se na *procura de um máximo local*, sendo explicada em detalhe na Secção (C.1) do Apêndice.

Independentemente da técnica utilizada, os resultados obtidos com este algoritmo podem considerar-se bastante satisfatórios. A Figura (4.3) representa, a título de exemplo, o ângulo de compasso estimado para a imagem de um mundo de Manhattan, bem como a evolução da função  $F(\alpha)$ . Utilizou-se neste caso o método de força bruta, a fim de poder observar o gráfico da função de verosimilhança em todo o domínio do ângulo de compasso.

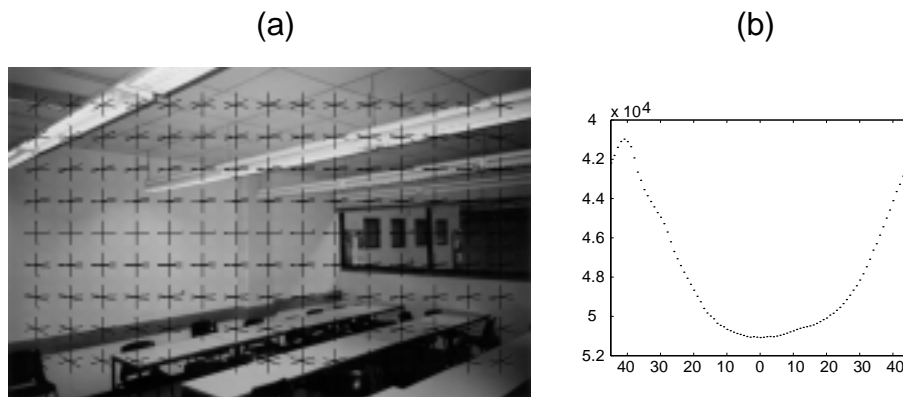


Figura 4.3: Ângulo de compasso estimado para a imagem de uma sala do Pavilhão de Civil, no IST. Em (a) a grelha que se sobrepõe à imagem mostra para vários pontos a direcção das linhas de Manhattan. Em (b) mostra-se a evolução da função  $F(\alpha)$ , onde se pode ver o valor estimado  $\alpha = -41^\circ$  como sendo o argumento que maximiza essa função.

## 4.5 Estimação da orientação no caso geral

### 4.5.1 Estimação em bloco dos ângulos de compasso, elevação e torção

Quando não existe nenhuma restrição para a posição da câmara, todas as orientações são possíveis, tendo que estimar-se, para além do ângulo de compasso  $\alpha$ , os ângulos de

elevação  $\beta$  e de torção  $\gamma$ , todos eles representados na Figura (2.4). Uma imagem assim obtida apresenta contornos como os que se mostram na Figura (2.7c).

A aplicação directa do modelo bayesiano desenvolvido na Secção (4.3) inspira-nos a implementar um algoritmo bastante simples mas, em contrapartida, muito pesado computacionalmente. A ideia é seguir o mesmo raciocínio do Algoritmo (4.1) para estimar, de uma só vez, os ângulos  $\alpha$ ,  $\beta$  e  $\gamma$  que parametrizam a orientação, obrigando a que se maximize uma função de verosimilhança de três variáveis, o que é extremamente custoso.

**Algoritmo 4.2.** *Dada uma imagem  $\mathbf{I}$ :*

1. *É aplicada a malha de extracção de informação através do Algoritmo (3.2), obtendo-se  $\{\mathbf{u}\}$  e  $\{\mathbf{E}_{\mathbf{u}}\}$ ;*
2. *Para uma colecção de orientações  $\Psi = (\alpha, \beta, \gamma)$ , com  $\alpha$  entre  $-45^\circ$  e  $45^\circ$ ,  $\beta$  entre  $-45^\circ$  e  $45^\circ$  e  $\gamma$  entre  $-54.7^\circ$  e  $54.7^\circ$ :*
  - (a) *Obtêm-se os pontos de fuga  $\mathbf{F}_x$ ,  $\mathbf{F}_y$  e  $\mathbf{F}_z$  através da Equação (2.5);*
  - (b) *Para cada pixel  $\mathbf{u}$  obtêm-se os valores de  $\theta_x(\mathbf{u})$ ,  $\theta_y(\mathbf{u})$  e  $\theta_z(\mathbf{u})$  através da Equação (2.7);*
  - (c) *Obtém-se  $F(\Psi) = \sum_{\mathbf{u}} \log \text{Prob}(\mathbf{E}_{\mathbf{u}}|\Psi, \mathbf{u})$  através das Equações (4.2) a (4.5);*
3. *Obtém-se a estimativa do máximo a posteriori  $\hat{\Psi} = \arg_{\Psi} \max F(\Psi)$ ;*
4. *Obtém-se a solução geral do problema calculando as orientações equiprojectivas com  $\hat{\Psi}$  a partir do método descrito no Teorema (2.11).*

Convém notar o seguinte na análise a este algoritmo:

- No passo (2), os valores-limite para os ângulos  $\alpha$ ,  $\beta$  e  $\gamma$  resultam directamente do Teorema (2.13). Se fosse percorrido todo o domínio referido na Definição (2.4), seriam encontrados no passo (3) vários valores de  $\Psi$  para os quais  $F(\Psi)$  é o máximo absoluto: estes valores são as *orientações equiprojectivas* que estudámos na Secção (2.3). Assim, obtém-se uma solução particular e determina-se a partir dela, no passo (4), todo o conjunto de soluções do problema.
- O *método de força bruta* aplicado a este algoritmo revela-se muito ineficiente, sendo a grande limitação o elevadíssimo tempo de processamento que exige. Se a colecção de orientações abranger todos as combinações de  $\alpha$ ,  $\beta$  e  $\gamma$  nos limites considerados com incrementos de  $1^\circ$ , os passos (2a), (2b) e (2c) são executados  $90 \times 90 \times 109.4 \approx 886140$  vezes, o que é claramente excessivo, tendo em conta que numa imagem com uma dimensão de, por exemplo,  $480 \times 640$  unidades de pixel, e considerando um rendimento de exclusão  $\eta_{exc} = 85.1\%^5$ , existem 45773 pixels cuja informação vai ser processada em cada iteração.

---

<sup>5</sup>Para utilizar o exemplo da Secção (3.4).

- Uma alternativa ao método de força bruta consiste em utilizar técnicas de refinamento<sup>6</sup> na procura do máximo a posteriori, como sugerido em [2]. Este procedimento consiste em aplicar inicialmente incrementos elevados aos ângulos  $\alpha, \beta$  e  $\gamma$  e fazê-los diminuir progressivamente, à medida que se prevê estar próximo da solução. Esta técnica, todavia, parece ainda insuficiente para produzir tempos de processamento aceitáveis e garantir ao mesmo tempo que o valor estimado  $\hat{\Psi}$  seja um máximo absoluto e não um pequeno pico local.

#### 4.5.2 Estimação da orientação em duas partes

Com vista a ultrapassar os principais inconvenientes do Algoritmo (4.2), propõe-se como contribuição original um novo método para estimar a orientação. O principal aliciente é a redução dos três graus de liberdade do problema para apenas dois<sup>7</sup>, o que se atinge através de uma separação na estimação dos ângulos  $\alpha, \beta$  e  $\gamma$  que parametrizam a orientação  $\Psi$ . Em vez de proceder a uma estimação em bloco destes ângulos pela máxima verosimilhança, este método subdivide o problema em *duas partes*:

- Na primeira parte, pretende-se estimar pela máxima verosimilhança os ângulos de elevação  $\beta$  e de torção  $\gamma$  (2 graus de liberdade);
- Na segunda parte, pretende-se estimar o ângulo de compasso  $\alpha$  (1 grau de liberdade).

Como vimos na abordagem geométrica, estes ângulos não são independentes entre si, pelo que a separação só é conseguida graças às Propriedades (2.7) e (2.8). Estas propriedades sugerem que comece por ser feita uma busca por um ponto de fuga qualquer e, uma vez localizado, convencionar que se trata de  $\mathbf{F}_z$ , o que determina automaticamente  $\beta$  e  $\gamma$ , concluindo a primeira parte. A segunda parte resume-se a estimar o ângulo  $\alpha$  utilizando as estimativas  $\hat{\beta}$  e  $\hat{\gamma}$  para os dois ângulos anteriores, num problema de natureza semelhante ao da estimação do ângulo de compasso para uma câmara no plano horizontal.

Na primeira parte, o procedimento de “busca” do ponto de fuga  $\mathbf{F}_z$  é feito aplicando o modelo bayesiano desenvolvido na Secção (4.3) acrescido de uma pequena modificação. Como nesta etapa os pontos de fuga  $\mathbf{F}_x$  e  $\mathbf{F}_y$  são do ponto de vista do problema irrelevantes, *contornos relacionados com os eixos  $\mathbf{x}$  e  $\mathbf{y}$  são considerados contornos parasitas*. Isto leva a que se siga um procedimento idêntico para os modelos de pixel  $\{1, 2, 4\} \subset M$ , o que implica que a Equação (4.4) se escreva agora:

$$\text{Prob}(\phi_{\mathbf{u}} | m_{\mathbf{u}}, \Psi, \mathbf{u}) = \begin{cases} P_{ang}(\phi_{\mathbf{u}} - \theta(m_{\mathbf{u}}, \Psi, \mathbf{u}) + k\pi) & \text{se } m_{\mathbf{u}} = 3 \\ U(\phi_{\mathbf{u}}) & \text{se } m_{\mathbf{u}} \neq 3 \end{cases} \quad (4.8)$$

Isto significa admitir mais uma aproximação: ao considerar-se os modelos  $m_{\mathbf{u}} = 1$  e  $m_{\mathbf{u}} = 2$  como equivalentes ao modelo  $m_{\mathbf{u}} = 4$  introduz-se para os dois primeiros

<sup>6</sup>“Coarse-to-fine” na literatura anglo-saxónica.

<sup>7</sup>Ou “dois mais um”, para ser mais exacto. Com rigor, passa-se de uma operação com 3 graus de liberdade para duas operações em que a mais pesada tem 2 graus de liberdade.



modelos uma função de distribuição uniforme para o ângulo  $\phi_{\mathbf{u}}$ , o que não é coerente com a característica gaussiana (dantes aproximada por um modelo de “caixa”) que as direcções destes contornos possuem. Porém, esta aproximação revela-se válida, *sendo no fundo equivalente a situações em que existe uma presença considerável de linhas parasitas que formam a sua própria grelha ortogonal*<sup>8</sup>. Como é de prever, o algoritmo lida bem com esta situação *desde que na imagem exista informação suficiente sobre todos os eixos de Manhattan*.

O algoritmo seguinte descreve o procedimento utilizado na estimação da orientação em duas partes. A primeira parte é constituída pelos passos (2) e (3), enquanto que a segunda está concentrada nos passos (4) e (5).

**Algoritmo 4.3.** *Dada uma imagem  $\mathbf{I}$ :*

1. *É aplicada a malha de extracção de informação através do Algoritmo (3.2), obtendo-se  $\{\mathbf{u}\}$  e  $\{\mathbf{E}_{\mathbf{u}}\}$ ;*
2. *Para uma colecção de ângulos  $\beta$  entre  $-90^\circ$  e  $90^\circ$  e ângulos  $\gamma$  entre  $-90^\circ$  e  $90^\circ$ :*
  - (a) *Obtém-se o ponto de fuga  $\mathbf{F}_z$  através da Equação (2.5);*
  - (b) *Para cada pixel  $\mathbf{u}$  obtém-se o valor de  $\theta_z(\mathbf{u})$  através da Equação (2.7);*
  - (c) *Obtém-se  $F_1(\beta, \gamma) = \sum_{\mathbf{u}} \log \text{Prob}(\mathbf{E}_{\mathbf{u}}|\beta, \gamma, \mathbf{u})$  através das Equações (4.2), (4.3), (4.8)<sup>9</sup> e (4.5);*
3. *Obtém-se as estimativas do máximo a posteriori  $\hat{\beta} = \arg_{\beta} \max F_1(\beta, \gamma)$  e  $\hat{\gamma} = \arg_{\gamma} \max F_1(\beta, \gamma)$ ;*
4. *Para uma colecção de ângulos  $\alpha$  entre  $-45^\circ$  e  $45^\circ$ :*
  - (a) *Obtém-se os pontos de fuga  $\mathbf{F}_x$ ,  $\mathbf{F}_y$  e  $\mathbf{F}_z$  através da Equação (2.5);*
  - (b) *Para cada pixel  $\mathbf{u}$  obtém-se os valores de  $\theta_x(\mathbf{u})$ ,  $\theta_y(\mathbf{u})$  e  $\theta_z(\mathbf{u})$  através da Equação (2.7);*
  - (c) *Obtém-se  $F_2(\alpha) = \sum_{\mathbf{u}} \log \text{Prob}(\mathbf{E}_{\mathbf{u}}|\alpha, \hat{\beta}, \hat{\gamma}, \mathbf{u})$  através das Equações (4.2) a (4.5);*
5. *Obtém-se o máximo a posteriori  $\hat{\alpha} = \arg_{\alpha} \max F_2(\alpha)$ ;*
6. *Obtém-se a solução geral do problema calculando as orientações equiprojectivas com  $\hat{\Psi} = (\hat{\alpha}, \hat{\beta}, \hat{\gamma})$  a partir do método descrito no Teorema (2.11).*

É importante tecer alguns comentários acerca deste algoritmo:

---

<sup>8</sup>Algo parecido com um “submundo de Manhattan parasita” descrito por um sistema de eixos distinto do sistema de eixos de Manhattan. Esta situação é frequente em cenários urbanos e muitas vezes induz em erro o estimador da orientação.

<sup>9</sup>Em vez da Equação (4.4).

- No passo (2), faz-se variar  $\beta$  e  $\gamma$  entre  $-90^\circ$  e  $90^\circ$ , não utilizando, portanto, os mesmos valores-limite do Algoritmo (4.2). A razão para isso prende-se com o facto de no passo (3) as estimativas do máximo a posteriori serem o máximo absoluto da função  $F_1(\beta, \gamma)$ , pelo que se requer que todo o domínio desta função esteja coberto. De facto, ao contrário do Algoritmo (4.2), onde se tinha  $F(\Psi_1) = F(\Psi_2)$  para quaisquer orientações equiprojectivas  $\Psi_1 = (\alpha_1, \beta_1, \gamma_1)$  e  $\Psi_2 = (\alpha_2, \beta_2, \gamma_2)$ , aqui não se tem em geral  $F_1(\beta_1, \gamma_1) = F_1(\beta_2, \gamma_2)$ , pelo que apenas se garante a existência de um *máximo local* no intervalo definido no Teorema (2.13) que pode não ser o máximo absoluto nesse intervalo.
- O *método da força bruta* aplicado aos passos (3) e (5) deste algoritmo permite já a sua realização experimental. Com efeito, se se utilizarem incrementos constantes de  $1^\circ$  fazendo variar  $\alpha$ ,  $\beta$  e  $\gamma$  sobre todo o intervalo, o número de iterações total nestes passos é agora  $180 \times 180 + 90 = 32490$ , cerca de 3.7% das 886140 iterações necessárias no Algoritmo (4.2). Como o tempo de processamento do algoritmo é aproximadamente proporcional a este número de iterações, conclui-se que este método conduz a uma redução muito significativa de 96.3% no tempo necessário para estimar a orientação.  
 Não obstante, este método de força bruta é ainda muito ineficiente. Verifica-se que é possível reduzir em mais 99.5% o tempo de processamento mediante a utilização da técnica descrita na Secção (C.1) do Apêndice, que se baseia na *procura de um máximo local da função de verosimilhança*. Esta técnica conduz a resultados tanto mais rápidos quanto mais próxima do repouso ( $\Psi = \mathbf{0}$ ) estiver a orientação da câmara;
- A única desvantagem face ao Algoritmo (4.2) reside na *propagação do erro*. Se na estimação de  $\beta$  e  $\gamma$ , na primeira parte do algoritmo, se cometerem erros  $\Delta\beta = \hat{\beta} - \beta$  e  $\Delta\gamma = \hat{\gamma} - \gamma$ , o efeito destes erros é amplificado na segunda parte do algoritmo, na estimação de  $\alpha$ . O erro cometido  $\Delta\alpha = \hat{\alpha} - \alpha$  será tanto maior quanto maiores (em módulo) forem  $\Delta\beta$  e  $\Delta\gamma$ . Veremos na Secção (D.3) alguns exemplos de experiências onde a má estimação da orientação teve este factor como uma das causas;
- Uma alternativa (à primeira vista vantajosa) para o passo (4) consiste em não fazer, nesta fase, qualquer processamento relacionado com o eixo  $\mathbf{z}$ , o que se conseguiria suprimindo todos os pixels  $\mathbf{u}$  tais que:

$$\arg_{m_{\mathbf{u}}} \max \text{Prob}(\phi_{\mathbf{u}} | m_{\mathbf{u}}, \hat{\beta}, \hat{\gamma}, \mathbf{u}) = 3$$

Estes pixels são aqueles que se julga fazer parte de contornos originados pelo eixo  $\mathbf{z}$ . Ter-se-ia então, à semelhança da primeira parte do algoritmo, uma equivalência entre os modelos de pixel  $\{3, 4\} \subset M$  que levaria também a reescrever a Equação (4.4): os contornos provocados por  $\mathbf{z}$  seriam, então, considerados parasitas (como o foram, na primeira etapa, os contornos provocados por  $\mathbf{x}$  e  $\mathbf{y}$ ).

A desvantagem deste procedimento é, no entanto, muito maior que o ganho que se obtém com a redução do processamento. Por exemplo, pixels próximos da recta que une os pontos de fuga  $\mathbf{F}_x$  e  $\mathbf{F}_z$  e que pertençam a um contorno, tendem a ser

“vistos” na primeira parte do algoritmo como fazendo parte de uma linha consistente com  $\mathbf{z}$ , pelo que na segunda parte seriam suprimidos sem sequer serem processados. Ao eliminar estes pixels, no entanto, estão também a apagar-se pistas sobre a localização de  $\mathbf{F}_x$ , piorando assim a estimação do ângulo de compasso  $\alpha$ .

Este efeito pode ter graves consequências, porque é frequente em imagens de mundos de Manhattan haver sobreposição de dois contornos originados por eixos diferentes. Outra desvantagem consiste no aumento da propagação do erro nos casos em que ocorrem ligeiros desvios  $\Delta\beta = \hat{\beta} - \beta$  e  $\Delta\gamma = \hat{\gamma} - \gamma$ , pois isso conduz a que se suprimam pixels que não deviam ser suprimidos. Conclui-se assim que esta alternativa não deve ser utilizada, a menos que se encontre uma forma de corrigir estas imperfeições.

Experimentalmente, a qualidade dos resultados obtidos foi muito satisfatória, como se mostra na Figura (4.4). Para o método de força bruta referido obteve-se um tempo de processamento à volta dos 40 minutos<sup>10</sup>, o que é ainda elevado. Todavia, ao utilizar-se o método de procura do máximo local descrito na Secção (C.1) conseguiu reduzir-se o tempo de processamento para 91.4 segundos, obtendo a mesma estimativa para a orientação num algoritmo cerca de 26 vezes mais rápido.



Figura 4.4: Resultados da aplicação do Algoritmo (4.3) ao problema da estimação da orientação: (a) Fotografia da Rua Augusta, na Baixa Pombalina, a que foi sobreposta a grelha de Manhattan e (b) conjunto de soluções possíveis para a orientação.

## 4.6 Estimação da orientação instantânea através de uma sequência de vídeo

Na Secção (2.5) estabeleceram-se alguns conceitos relacionados com a variação da orientação da câmara na *hipótese de pequenas rotações*, abrindo assim caminho para

<sup>10</sup>As condições e resultados destas experiências são um assunto que se aprofunda na Secção (4.8).

a implementação de um método capaz de estimar a função da orientação da câmara ao longo do tempo,  $\Psi(t)$ , utilizando a informação extraída em cada trama de uma sequência de vídeo.

Se atribuírmos a  $\varphi_{\max}$ <sup>11</sup> o valor razoável  $\varphi_{\max} = 5^\circ$ , isso implica que para uma sequência de vídeo digitalizada com 12.5 Hz, o ângulo de rotação por segundo é sempre menor que  $62.5^\circ/s$ . Esta hipótese é viável desde que a câmara não seja submetida a oscilações apreciáveis ou a movimentos demasiado repentinos.

Nas condições da hipótese de pequenas rotações, pode então utilizar-se o seguinte algoritmo para estimar a orientação instantânea:

**Algoritmo 4.4.** *Para cada instante  $t = 0, 1, \dots, T$ , onde  $T$  é o número de tramas:*

1. *Aplica-se à imagem da trama  $t$  a malha de extracção de informação através do Algoritmo (3.2), obtendo-se  $\{\mathbf{u}\} = \{\mathbf{u}(t)\}$  e  $\{\mathbf{E}_u\} = \{\mathbf{E}_u(t)\}$ ;*
2. *Se  $t = 0$  estima-se a orientação inicial  $\hat{\Psi}(0) = (\hat{\alpha}_0, \hat{\beta}_0, \hat{\gamma}_0)$  através do Algoritmo (4.3) e volta-se para o passo (1) fazendo  $t = t + 1$ ; se  $t > 0$ , avança-se para o passo (3);*
3. *Para uma colecção de ângulos  $\beta$  no intervalo  $[\hat{\beta}_{t-1} - \varphi_{\max}, \hat{\beta}_{t-1} + \varphi_{\max}]$  e  $\gamma$  no intervalo  $[\hat{\gamma}_{t-1} - \varphi_{\max}, \hat{\gamma}_{t-1} + \varphi_{\max}]$ , utiliza-se o mesmo procedimento dos passos (2a), (2b) e (2c) do Algoritmo (4.3) para se obter as estimativas do máximo a posteriori  $\hat{\beta}_t = \arg_{\beta} \max F_1(\beta, \gamma)$  e  $\hat{\gamma}_t = \arg_{\gamma} \max F_1(\beta, \gamma)$ ;*
4. *Com base nas estimativas  $\hat{\beta}_{t-1}$  e  $\hat{\beta}_t$  determina-se o valor de  $\Delta\alpha_{\max}$  através das expressões introduzidas no Teorema (2.16);*
5. *Para uma colecção de ângulos  $\alpha$  no intervalo  $[\hat{\alpha}_{t-1} - \Delta\alpha_{\max}, \hat{\alpha}_{t-1} + \Delta\alpha_{\max}]$  utiliza-se o mesmo procedimento dos passos (4a), (4b) e (4c) do Algoritmo (4.3) para se obter a estimativa do máximo a posteriori  $\hat{\alpha}_t = \arg_{\alpha} \max F_2(\alpha)$ ;*
6. *Se  $\hat{\Psi}(t) = (\hat{\alpha}_t, \hat{\beta}_t, \hat{\gamma}_t)$  não estiver na região referida no Teorema (2.13), obtém-se a partir do método descrito no Teorema (2.11) uma orientação equiprojectiva que resida nessa região e assume-se  $\hat{\Psi}(t)$  igual a essa orientação;*
7. *Se  $t < T$  volta-se para o passo (1) fazendo  $t = t + 1$ ; Se  $t = T$  recupera-se a sequência de estimativas  $\{\hat{\Psi}(0), \hat{\Psi}(1), \dots, \hat{\Psi}(T)\}$  e calcula-se para cada instante a solução geral, recorrendo-se novamente ao método descrito no Teorema (2.13). De seguida, procura-se neste conjunto todas as orientações que geram sequências válidas, exigindo que satisfaçam para cada caso as condições da Hipótese (2.15).*

O passo (6) pode ser justificado se atendermos a que o Teorema (2.16) conduz a um intervalo  $[\hat{\alpha}_{t-1} - \Delta\alpha_{\max}, \hat{\alpha}_{t-1} + \Delta\alpha_{\max}]$  tanto mais pequeno quanto mais próximo de  $0^\circ$  forem os valores de  $\hat{\beta}_{t-1}$  e  $\hat{\beta}_t$ , o que sugere que se escolham judiciosamente soluções

---

<sup>11</sup>Este ângulo foi introduzido atrás na Hipótese (2.15).

particulares onde os ângulos  $\{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_{T-1}\}$  sejam o mais próximos possível de  $0^\circ$ . Na abordagem geométrica, vimos que o problema da orientação admite sempre uma solução particular dentro da região referida no Teorema (2.13), pelo que se minimiza o número de cálculos escolhendo para cada instante uma solução particular nessa região.

## 4.7 Classificação de contornos e detecção de objectos parasitas

Vamos agora estudar algumas aplicações da estimação da orientação. Supondo, por exemplo, que para uma dada imagem o Algoritmo (4.3) estima uma orientação  $\hat{\Psi}$  para a câmara, pode regressar-se a cada pixel da imagem e, recorrendo à Equação (4.2), calcular (a menos de uma relação de proporção) a matriz de probabilidades:

$$\text{Prob} \left\{ m_{\mathbf{u}} | \mathbf{E}_{\mathbf{u}}, \hat{\Psi}, \mathbf{u} \right\} \propto \text{Prob} \left\{ \mathbf{E}_{\mathbf{u}} | m_{\mathbf{u}}, \hat{\Psi}, \mathbf{u} \right\} \cdot P_M(m_{\mathbf{u}})$$

para cada  $m_{\mathbf{u}} \in M = \{1, 2, 3, 4, 5\}$ .

Isto permite que se estimem modelos  $\{\hat{m}_{\mathbf{u}}\}$  para cada pixel  $\mathbf{u}$ , recorrendo-se para esse fim à estimativa do máximo a posteriori:

$$\hat{m}_{\mathbf{u}} = \arg_{m_{\mathbf{u}}} \max \left[ \text{Prob} \left\{ \mathbf{E}_{\mathbf{u}} | m_{\mathbf{u}}, \hat{\Psi}, \mathbf{u} \right\} \cdot P_M(m_{\mathbf{u}}) \right] \quad (4.9)$$

Torna-se assim possível *classificar* cada pixel de acordo com um dos modelos mostrados na Tabela (4.1), o que permite realizar tarefas como:

- *Classificação de linhas de Manhattan.* A partir dos pixels  $\mathbf{u}$  com  $\hat{m}_{\mathbf{u}} \in \{1, 2, 3\}$ , classificar contornos como sendo originados pelos eixos  $\mathbf{x}$ ,  $\mathbf{y}$  e  $\mathbf{z}$ ;
- *Detecção de objectos parasitas.* A partir dos pixels  $\mathbf{u}$  com  $\hat{m}_{\mathbf{u}} = 4$  detectar contornos de objectos parasitas na imagem.

Em relação à primeira tarefa, obtém-se experimentalmente um sucesso assinalável mediante a utilização do Algoritmo (4.3), como pode comprovar-se observando os exemplos das Figuras (4.5a) e (4.6a).

Verifica-se, no entanto, que a detecção de objectos parasitas é um pouco mais difícil de realizar. O grande obstáculo consiste no “ruído” que são os pixels  $\mathbf{u}$  pertencentes a contornos originados por  $\mathbf{x}$ ,  $\mathbf{y}$  e  $\mathbf{z}$  mas cuja má estimação da direcção do gradiente conduz a que se obtenha  $\hat{m}_{\mathbf{u}} = 4$ , confundindo-os com contornos de objectos parasitas. Assim, não se consegue distinguir com grande clareza, na imagem, os verdadeiros objectos parasitas, que seriam as cadeiras da última fila na Figura (4.5b) e os quatro indivíduos nos pisos de baixo e de cima na Figura (4.6b). Propõe-se algumas soluções (que não foram exploradas em detalhe por se afastarem um pouco do objectivo deste trabalho):

- *Melhorar a estimação da direcção do gradiente.* Esta solução apresenta a vantagem de servir também para aperfeiçoar os algoritmos de estimação da orientação, pelo que se revela pertinente;

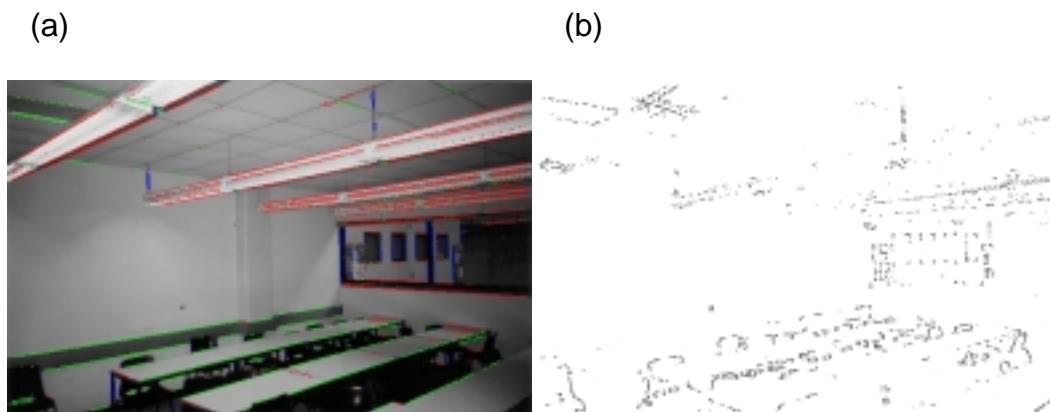


Figura 4.5: Exemplo de aplicações práticas da estimação da orientação: (a) Classificação de linhas de Manhattan e (b) detecção de objectos parasitas na imagem de uma sala do IST. Em (a) representa-se respectivamente a vermelho, verde e azul contornos originados pelos eixos  $x$ ,  $y$  e  $z$ .

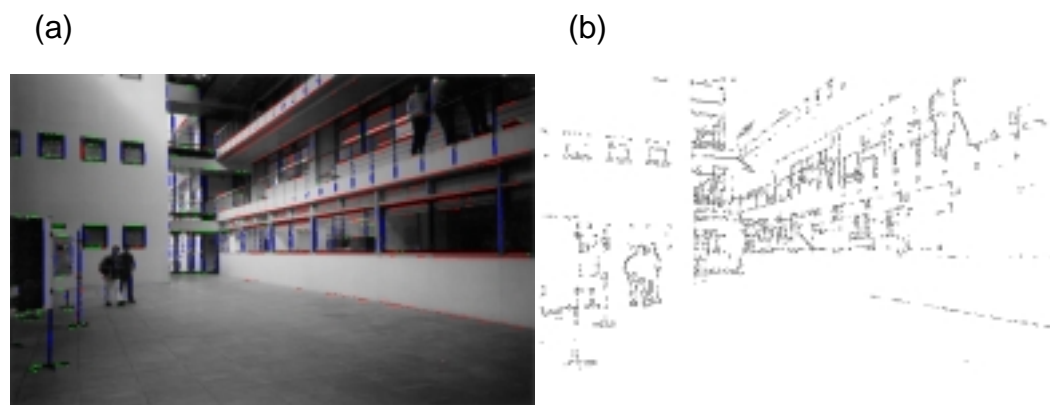


Figura 4.6: Exemplo de aplicações práticas da estimação da orientação: (a) Classificação de linhas de Manhattan e (b) detecção de objectos parasitas numa imagem do interior do Pavilhão de Civil do IST. Em (a) representa-se respectivamente a vermelho, verde e azul contornos originados pelos eixos  $x$ ,  $y$  e  $z$ .

- *Utilizar técnicas de supressão de pixels* (por limiarização ou por isolamento) *depois da estimação da orientação*. Eliminar desta forma os pixels  $\mathbf{u}$  que introduzem ruído por se ter estimado erradamente  $\hat{m}_{\mathbf{u}} = 4$ , tirando partido do facto de estes pixels tenderem a aparecer mais "isolados" na imagem e não apresentarem um pico tão pronunciado na Equação (4.9) para  $\hat{m}_{\mathbf{u}} = 4$ .

## 4.8 Resultados experimentais

### 4.8.1 Condições das experiências

Alguns resultados de experiências foram já mostrados nas secções anteriores, pretendendo ilustrar em cada caso os fundamentos teóricos estabelecidos. Vamos agora incidir numa análise mais crítica e objectiva, debruçando-nos sobre a estimação da orientação e a sua aplicação na classificação de contornos.

Em todas as experiências efectuadas o processador utilizado foi um Pentium IV a 1.5 GHz, com uma memória RAM de 256 Mb. Os algoritmos foram implementados em linguagem MATLAB, procurando otimizar-se a velocidade de execução. Os testes incidiram sobre o Algoritmo (4.3) utilizando os métodos de redução do tempo de processamento que se descrevem nas Secções (C.1) e (C.2) do Apêndice.

A fim de testar a fiabilidade e rapidez proporcionadas por este algoritmo, utilizaram-se 30 imagens de mundos de Manhattan. Não se privilegiou a qualidade deste conjunto de teste, evitando-se a realização de experiências em condições excepcionais. Os casos seguintes são exemplos de imagens com pouca qualidade:

- imagens que apresentam contrastes elevados devido à presença de sombras ou luz excessiva;
- imagens que contêm pouca informação sobre algum ou alguns dos eixos de Manhattan.

A captura foi feita com uma máquina fotográfica convencional (não digital) mantendo a lente completamente retraída para que o zoom permanecesse fixo, tendo-se o cuidado na revelação e digitalização de não deformar nem recortar qualquer das imagens. Nenhum programa foi posteriormente aplicado para eliminar a distorção radial e outros factores prejudiciais.

A partir do processo de estimação dos parâmetros de calibração<sup>12</sup> descrito em [11], estimou-se uma distância focal de 502 unidades de pixel para imagens digitalizadas com  $530 \times 353$  pixels. A distância focal em unidades de comprimento é então:

$$f[\text{mm}] = 502 \times \frac{36}{530} = 34.1 \text{ mm}$$

onde se fez o procedimento inverso do Exemplo (2.3).

---

<sup>12</sup>As imagens utilizadas na calibração foram 10 fotografias de uma grelha quadriculada, variando entre si a orientação da câmara mas colocando sempre o zoom no valor mínimo, com o objectivo de manter constante a distância focal.

### 4.8.2 Estimação da orientação a partir de uma imagem

Conhecidas as probabilidades a priori  $P_M$ ,  $P_{ON}$ ,  $P_{OFF}$ , e  $P_{ang}$  através do processo descrito na Secção (4.2), passou-se para a fase de testes propriamente dita, da qual se apresentam dois exemplos nas Figuras (4.7) e (4.8).

No primeiro caso, a estimação da orientação é satisfatória, o que conduziu a uma classificação correcta das linhas de contorno, como pode ver-se na Figura (4.7b). Objectivamente, pode afirmar-se que a maioria dos testes permitiram atingir resultados igualmente positivos, tendo-se atingido uma taxa de sucesso de 86.7%.

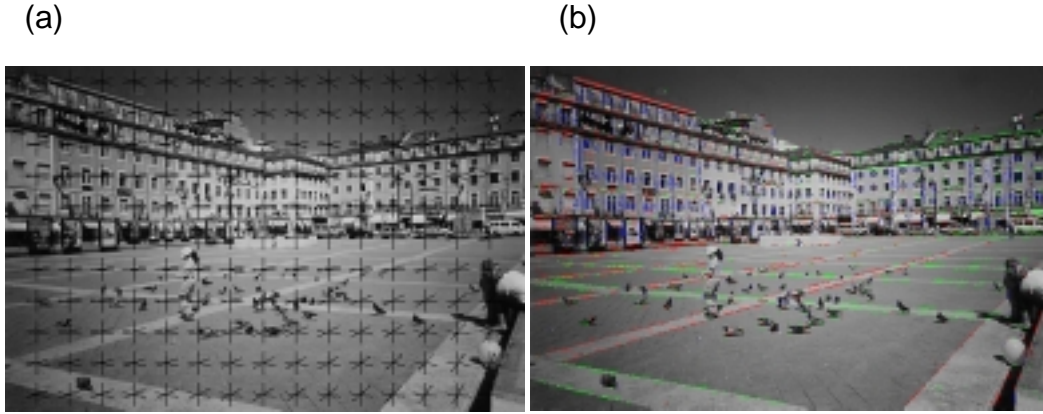


Figura 4.7: Exemplo de uma estimação correcta da orientação: A solução particular encontrada foi  $\alpha = 31^\circ$ ,  $\beta = 1^\circ$  e  $\gamma = 2^\circ$ , tendo-se obtido um tempo de processamento de 49.2 segundos. (a) Grelha de Manhattan e (b) Classificação de linhas de contorno. As cores utilizadas são respectivamente vermelho, verde e azul para os eixos  $\mathbf{x}$ ,  $\mathbf{y}$  e  $\mathbf{z}$ .

O caso da Figura (4.8) é, por seu turno, o exemplo de um resultado insatisfatório: pode ver-se que a estimativa da orientação é claramente errada. A principal causa está relacionada com a existência de *contornos ambíguos*, isto é, pares de contornos originados por eixos diferentes que na primeira parte do algoritmo são erradamente interpretados como sendo ambos originados pelo eixo  $\mathbf{z}$ . Isto faz com que  $\beta$  e  $\gamma$  sejam mal estimados, produzindo como efeito de acumulação um erro ainda maior na estimação do ângulo de compasso. Os valores calculados para as duas funções de verosimilhança apresentam-se nas Figuras (4.8b-c), sendo de notar a evolução algo irregular da função  $F_1(\beta, \gamma)$ , o que indicia pouca certeza no resultado obtido<sup>13</sup>.

Este tipo de situações não é muito frequente, embora mereça preocupação e se deva apontar como um dos problemas a resolver no futuro.

Apresenta-se mais alguns resultados destas experiências nas Secções (D.1), (D.2) e (D.3) do Apêndice. Os tempos de processamento exigidos para cada uma das 30 imagens de teste estão representados na Tabela (4.3). O valor médio obtido para o tempo de processamento foi de 56.7 segundos, o que se considera um resultado excelente, só atingível graças aos processos de optimização descritos nas Secções (C.1) e (C.2). Nalguns

<sup>13</sup>Geralmente, este tipo de situações origina “picos gémeos” nas funções de verosimilhança separados entre si por uma distância inferior a duas vezes o parâmetro  $\tau$  da função de distribuição  $P_{ang}(t)$ .



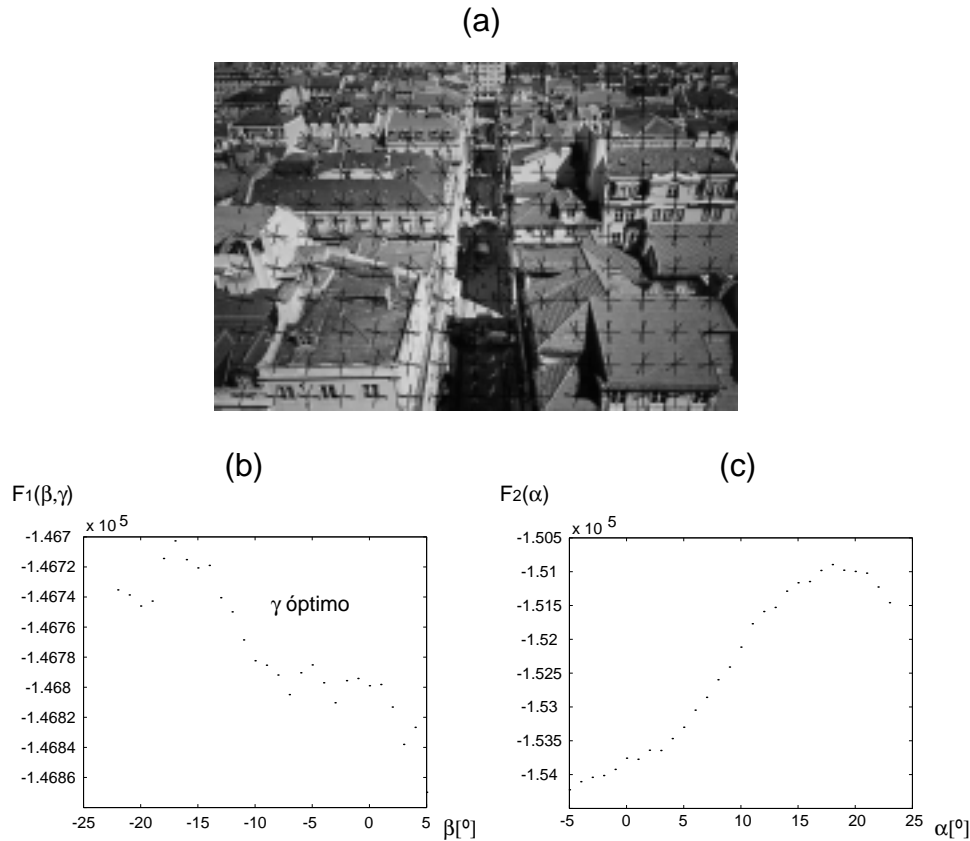


Figura 4.8: Exemplo de uma estimação errada da orientação para uma fotografia aérea da Baixa Pombalina. (a) Grelha de Manhattan, (b) andamento da função  $F_1(\beta, \gamma)$  e (c) o andamento da função  $F_2(\alpha)$ . A solução particular obtida, que não se considera correcta, foi  $\alpha = 18^\circ$ ,  $\beta = -17^\circ$  e  $\gamma = -8^\circ$ .

casos foi mesmo possível descer abaixo dos 30 segundos, com o equipamento e condições referidos atrás.

### 4.8.3 Estimação da orientação instantânea a partir de uma sequência de vídeo

O Algoritmo (4.4) de estimação da função de orientação ao longo do tempo a partir de uma sequência de vídeo apresenta resultados bastante razoáveis, embora se julgue pertinente o seu aperfeiçoamento, dada a dependência que existe entre a qualidade dos resultados e a natureza das tramas da sequência, por vezes conduzindo a estimatórias insatisfatórias. Razões para isso são:

- Contornos que surgem nas várias tramas originados por oscilações ou movimentos relativamente rápidos da câmara, que designamos por *rastos* e que prejudicam a estimação por serem tipicamente linhas rectas e portanto fortemente parasitas;
- Variações súbitas de luminosidade entre tramas consecutivas, fazendo com que

Imagem	$t_p$ [s]	Qualidade (1-5)	Imagem	$t_p$ [s]	Qualidade (1-5)
Int01	3.8	3	Ext05	57.0	5
Int02	22.8	5	Ext06	49.2	5
Int03	21.5	5	Ext07	66.5	5
Int04	34.4	5	Ext08	37.3	5
Int05	20.0	5	Ext09	214.3	3
Int06	35.0	3	Ext10	32.8	5
Int07	16.4	5	Ext11	45.1	5
Int08	13.5	5	Ext12	37.2	5
Int09	28.5	1	Ext13	37.9	5
Int10	35.5	4	Ext14	165.4	5
Int11	18.2	5	Ext15	129.1	5
Ext01	91.5	4	Ext16	67.8	5
Ext02	31.8	5	Ext17	105.6	1
Ext03	40.0	4	Ext18	70.4	1
Ext04	121.5	2	Ext19	50.0	5

Tabela 4.3: Resultados experimentais da aplicação do Algoritmo (4.3) em imagens de interiores e exteriores. Mostra-se para cada uma o tempo de processamento e uma avaliação qualitativa do valor estimado para a orientação: os valores atribuídos variam entre 1 (resultados não aceitáveis) a 5 (resultados excelentes).

algumas sejam sub ou sobre-iluminadas. Ambos os casos podem fazer diminuir o módulo do gradiente e assim tornar “invisíveis” alguns contornos relevantes;

- A digitalização e eventual compressão do ficheiro de vídeo faz com que as tramas percam qualidade. Por exemplo, esquemas de compressão MPEG levam ao surgimento nas imagens de pequenas zonas nebulosas com variações “artificiais” da intensidade luminosa, podendo ser confundidas com contornos. Por outro lado, nestas circunstâncias, os verdadeiros contornos são muitas vezes atenuados;
- A ausência de compressão, que corrige o problema anterior, leva a que seja muito difícil lidar com ficheiros de vídeo pela sua elevada dimensão, tornando o algoritmo pesado e por vezes impossível de ser executado computacionalmente pelo excessivo consumo de memória.

Ainda assim, pode dizer-se que se obteve resultados satisfatórios para a maioria das sequências de vídeo testadas. O tempo de processamento para uma esquema de compressão MPEG-4 com tramas de dimensão  $288 \times 360$  é da ordem dos 10 segundos por trama, o que, para uma frequência de 12.5 tramas por segundo, conduz a tempos de 125 segundos por cada segundo de filme.

## Capítulo 5

## Conclusões

A Figura (5.1) mostra uma imagem da interface gráfica de demonstração que foi desenvolvida através do MATLAB com o objectivo tornar acessível ao utilizador comum a ferramenta implementada para estimar a orientação a partir de uma imagem.



Figura 5.1: Interface gráfica de demonstração da estimação da orientação.

Os resultados obtidos experimentalmente, expostos na Secção (4.8) e complementados com a informação das Secções (D.1), (D.2) e (D.3) do Apêndice, permitem extrair as seguintes conclusões:

- Para uma câmara assente no plano horizontal, o Algoritmo (4.1) permite obter

boas estimativas do ângulo de compasso a partir de uma imagem, bastando que exista, na captação, um esforço intuitivo para manter a câmara nesta posição;

- O Algoritmo (4.3) permite, dada uma imagem, estimar a orientação da câmara quando esta se encontra posicionada livremente. Utiliza-se um método eficiente que divide a estimação da orientação em duas partes distintas, retirando um grau de liberdade ao problema. Os resultados são satisfatórios em cerca de 90% dos casos, mesmo que a imagem disponível não tenha uma qualidade superior a razoável. A utilização de métodos de aceleração do algoritmo teve como resultado, nas experiências efectuadas, um tempo de processamento médio de 56.7 segundos, tendo sido ultrapassada, em vários casos, a fasquia dos 30 segundos, uma vez que a rapidez da solução varia com a própria orientação da câmara.
- A utilização de uma linguagem de programação compilável como C/C++ na implementação do algoritmo poderá aumentar significativamente a velocidade de processamento. Isto parece aliciante, pois faz com que a inclusão deste algoritmo em qualquer programa de processamento de imagem permita um uso fácil e rápido;
- A classificação de contornos é quase sempre efectuada correctamente, dependendo do sucesso obtido na estimação da orientação. O mesmo não se pode dizer da detecção de objectos parasitas, que é difícil mesmo nos casos em que a orientação é correctamente estimada. Isto deve-se a uma insuficiente estimação da direcção do gradiente, que deve ser melhorada em abordagens futuras a este problema;
- O Algoritmo (4.4) de estimação da função de orientação ao longo do tempo (a partir de uma sequência de vídeo) apresenta resultados ainda não completamente satisfatórios, devido aos motivos que se enumeraram na Secção (4.8). No entanto, para a maioria das experiências realizadas obteve-se uma boa estimativa da orientação instantânea, tendo-se como único aspecto a resolver a exigência de se ter cada trama como uma imagem fidedigna de um mundo de Manhattan. Sugere-se que em posteriores abordagens a este problema se aproveite de forma mais eficiente a informação mútua entre tramas como forma de fazer face a tal inconveniente.

Em resumo, os resultados obtidos consideram-se muito satisfatórios. Os principais objectivos deste trabalho foram atingidos e, à medida que alguns assuntos foram aprofundados surgiram novas questões interessantes que também foram exploradas.

Propuseram-se alternativas aos métodos descritos em [1] e [2] que se verificaram experimentalmente ser mais eficientes, o que possibilita a sua inclusão em qualquer programa de processamento de imagem, após proceder, naturalmente, a outros processos de optimização computacional. Os métodos aqui desenvolvidos assentam numa revisão da teoria geométrica e estatística que serve de alicerce ao problema da orientação.

Alguns aspectos, como vimos anteriormente, devem ainda ser melhorados, pelo que se julga recompensatória qualquer abordagem futura a este problema.

## Apêndice A

# Notas Adicionais Sobre a Geometria do Mundo de Manhattan

### A.1 Modelo geral da câmara projectiva

Na Secção (2.2) caracterizou-se o modelo da câmara obscura, onde o único parâmetro intrínseco de interesse consiste na distância focal.

Porém, para a grande parte das câmaras, ou quando os cálculos efectuados exigem maior precisão, é necessário entrar em conta com mais parâmetros, dos quais se destaca a *distorção radial*. Nesta secção irá ser abordado de forma muito ligeira o modelo geral da câmara projectiva.

Uma análise mais completa acerca da calibração de câmaras e de técnicas para corrigir imperfeições na imagem causadas por uma câmara não obscura pode ser encontrada em [11]. Outra perspectiva acerca deste assunto, no contexto da Óptica, surge, por outro lado, em [12].

Comecemos por adicionar ao modelo da câmara obscura os seguintes graus de liberdade:

- *As coordenadas do ponto principal.* Embora no modelo da câmara obscura se assuma que o ponto principal coincide com o centro do rectângulo da imagem, numa câmara real com uma calibração imperfeita pode existir um deslocamento do ponto principal que é necessário considerar. Isso pode também dever-se, numa máquina analógica, a um deslocamento na posição do filme, ou, no processo de digitalização, a um recorte não simétrico da imagem.
- *O coeficiente de aspecto  $\alpha$ .* Após o processo de digitalização pode acontecer que cada pixel da imagem não seja quadrado, isto é, tenha dimensões diferentes na horizontal e na vertical. Nesse caso tem-se  $\alpha \neq 1$  e a imagem sofre uma deformação caso este parâmetro não seja tido em conta.
- *O ângulo entre os eixos ópticos.* Para a maioria das câmaras este ângulo é de  $90^\circ$ , o que faz com que a imagem não apareça deformada. Geralmente este ângulo não sofre mais que ligeiras alterações, mas é sempre um parâmetro a considerar.

Para além destes graus de liberdade, que introduzem *transformações lineares* ao plano da imagem, há ainda a considerar alguns efeitos *não lineares* que são responsáveis pelo aparecimento de deformações:

- *Difusão*. Deve-se a reflexos de superfícies no interior da objectiva e produz perda geral do contraste. Por vezes aumenta devido a reflexos no interior da fotocâmara. Este efeito é atenuado, em algumas máquinas, com a utilização de um pára-sol na objectiva, reduzindo a luz que não desempenha nenhum papel na formação da imagem.
- *Aberração esférica*. Produzir uma objectiva de curvatura esférica tem um preço menor do que outra em que a curvatura varie (asférica). O custo desta operação é a aberração esférica, representada na Figura (A.1), em que os bordos da lente focam as ondas luminosas num ponto que não coincide com o centro da qual não há nitidez. Os raios oblíquos que atravessam a lente projectam-se em partes diferentes no plano da película, de forma confusa, em vez de se sobreporem. Existem várias maneiras de resolver o problema, tratando-se no essencial de uma questão de custos. As objectivas catadióptricas, ou de espelho, com superfícies reflectoras, não introduzem tais aberrações.
- *Distorção radial*. O diafragma impede que os raios de luz oblíquos incidentes passem pelo centro da objectiva e, como as superfícies das lentes não são paralelas aos bordos, a luz que forma a imagem inclina-se. Isso afecta a nitidez e produz uma imagem distorcida, como pode ver-se na Figura (A.2). Se a forma se comprimir, tem-se uma distorção *em barril*; se se alargar, tem-se uma distorção *em almofada*. Geralmente, esta distorção manifesta-se com uma simetria radial, daí designar-se por distorção radial. As objectivas simétricas (que têm elementos complementares na parte anterior e posterior) eliminam a distorção, mas as teleobjectivas e as objectivas retrofoco, sendo assimétricas, são prejudicadas. A imagem do olho de peixe constitui um exemplo extremo da distorção em barril.

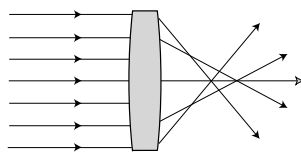


Figura A.1: Aberração esférica.

De todos estes fenómenos, o mais preocupante é a *distorção radial*, que por vezes afasta a câmara real do modelo de câmara obscura atrás referido. Uma forma de lidar com a distorção radial consiste em medi-la, o que possibilita a sua correcção aplicando à imagem algoritmos conhecidos e produzindo como resultado uma imagem regenerada, como se esta tivesse sido obtida através uma câmara sem distorção radial. Descrevem-se alguns destes algoritmos em [11].

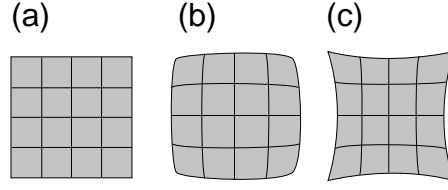


Figura A.2: Efeito da distorção radial: Imagem (a) sem distorção, (b) com distorção em barril e (c) com distorção em almofada.

Uma forma de modelizar a distorção radial consiste em considerá-la como uma função radialmente simétrica

$$F(r) = r(1 + k_1 r^2 + k_2 r^4 + \dots + k_n r^{2n})$$

onde  $r$  é a distância em relação ao ponto central do rectângulo da imagem. A medição da distorção radial é então a estimação dos parâmetros  $k_1, k_2, \dots, k_n$ .

## A.2 Relação entre a orientação e os sistemas de eixos

Nesta secção vai determinar-se as expressões de  $\mathbf{a}$ ,  $\mathbf{b}$  e  $\mathbf{c}$  em função dos ângulos  $\alpha$ ,  $\beta$  e  $\gamma$  que parametrizam a orientação.

Já vimos na Secção (2.2) que:

$$(\mathbf{a}, \mathbf{b})^\top = \mathbf{R}_\gamma(\mathbf{a}_0, \mathbf{b}_0)^\top$$

com:

$$\mathbf{R}_\gamma = \begin{bmatrix} \cos \gamma & -\sin \gamma \\ \sin \gamma & \cos \gamma \end{bmatrix}$$

Resta-nos obter as expressões de  $\mathbf{a}_0$ ,  $\mathbf{b}_0$  e  $\mathbf{c}$ .

Sendo  $\mathbf{P} = (P_x, P_y, P_z)^\top$  as coordenadas cartesianas do ponto principal, deduz-se facilmente que:

$$\begin{cases} P_x &= -f \cos \alpha \cos \beta \\ P_y &= f \sin \alpha \cos \beta \\ P_z &= f \sin \beta \end{cases}$$

Como se tem  $\mathbf{c} = -\frac{\mathbf{P}}{f}$ , vem:

$$\mathbf{c} = (\cos \alpha \cos \beta, -\sin \alpha \cos \beta, -\sin \beta)^\top$$

Por seu turno,  $\mathbf{a}_0$  pode ser escrito recorrendo às propriedades:

$$\begin{cases} \mathbf{c} \perp \mathbf{a}_0 \\ \mathbf{a}_0^\top \mathbf{z} = 0 \\ \mathbf{a}_0^\top \mathbf{y} \geq 0 \\ \mathbf{a}_0^\top \mathbf{a}_0 = 1 \end{cases} \Rightarrow \mathbf{a}_0 = (\sin \alpha, \cos \beta, 0)^\top$$

obtendo-se  $\mathbf{b}_0$  de modo análogo:

$$\begin{cases} \mathbf{a}_0 \perp \mathbf{b}_0 \perp \mathbf{c} \\ \mathbf{b}_0^\top \mathbf{z} > 0 \\ \mathbf{b}_0^\top \mathbf{b}_0 = 1 \end{cases} \Rightarrow \mathbf{b}_0 = (\cos \alpha \sin \beta, -\sin \alpha \sin \beta, \cos \beta)^\top$$

## A.3 Representação de pontos no plano projectivo

### Coordenadas cartesianas e vectores homogéneos

Podem encontrar-se referências interessantes sobre Geometria Projectiva em [3] e um resumo muito elucidativo em [5]. Aqui será feita apenas uma justificação muito rápida do seu uso e uma explicação da notação usada nas várias secções para representar pontos no plano de projecção.

O plano projectivo  $\mathbb{P}^2$  é constituído pelos *pontos afins* (representáveis no plano euclidiano  $\mathbb{E}^2$ ) e pelos *pontos ideais*<sup>1</sup>. A intersecção de duas rectas que no sentido euclidiano designariámos por “paralelas” resulta, como se sabe, num ponto ideal. Um ponto afim  $\mathbf{P}$  representado nas suas coordenadas cartesianas por:

$$\mathbf{P} = (a, b)^\top$$

é representado, no contexto da Geometria Projectiva, pelo vector homogéneo:

$$\tilde{\mathbf{P}} = (a, b, 1)^\top$$

que não é a única representação do ponto em coordenadas homogéneas, pois é uma designação equivalente do mesmo ponto  $\mathbf{P}$  qualquer vector homogéneo  $k(a, b, 1)^\top$ , com  $k \in \mathbb{R} \setminus \{0\}$ . Por outro lado, um ponto ideal  $\mathbf{Q}$  será sempre representado por um vector homogéneo da forma:

$$\tilde{\mathbf{Q}} = (a, b, 0)^\top$$

sendo  $k(a, b, 0)^\top$ , com  $k \in \mathbb{R} \setminus \{0\}$ , uma designação equivalente.

Em contrapartida, as coordenadas cartesianas de um ponto afim  $\tilde{\mathbf{P}} = (a, b, c)^\top$  escrito em notação homogénea são dadas por:

$$\mathbf{P} = \left( \frac{a}{c}, \frac{b}{c} \right)^\top$$

Quando  $c = 0$ , o ponto é ideal, sendo esta expressão elucidativa para se compreender porque razão os pontos ideais se parecem localizar “no infinito”. Toda a recta contém um e um só ponto ideal, partilhado por todas as rectas “paralelas” (no sentido euclidiano) com a primeira. Qualquer ponto ideal  $\tilde{\mathbf{Q}} = (a, b, 0)^\top$  tem 1 grau de liberdade, sendo caracterizado pela direcção da linha que contém a origem e o ponto afim  $\tilde{\mathbf{P}} = (a, b, 1)^\top$ .

---

<sup>1</sup>Também designados por “pontos no infinito”.



### Notação complexa

Por vezes surge a necessidade de representar um ponto  $\mathbf{P}$  do plano projectivo especificando a “distância” em relação à origem e o ângulo que caracteriza a direcção do vector cartesiano a ele associado. Se  $\mathbf{P}$  for afim pode representar-se através de um número complexo:

$$\overline{\mathbf{P}} = re^{j\theta}$$

cuja representação no plano de Argand é feita a partir das suas coordenadas cartesianas.

Para estender esta forma de representação aos pontos ideais podemos usar em vez de  $\mathbb{C}$  o conjunto (que é uma sua extensão):

$$\mathbb{C}_\infty = \{x : x = re^{j\theta}, r \in \mathbb{R} \cup \{\infty\} \wedge \theta \in \mathbb{R}\}$$

Assim, um ponto ideal  $\mathbf{Q}$  pode ser representado em notação complexa por:

$$\overline{\mathbf{Q}} = \infty e^{j\theta}$$

sendo a sua direcção definida (módulo  $\pi$ ) pelo parâmetro  $\theta$ .

Deduz-se portanto que:

$$\infty e^{j\theta} = \infty e^{j(\theta \pm \pi)}$$

## A.4 Localização dos pontos de fuga em função da orientação

Nesta secção deduzem-se as expressões que relacionam as coordenadas dos pontos de fuga com a orientação da câmara.

Se  $\mathbf{X}$  for um ponto qualquer do plano projectivo da imagem, verifica a igualdade:

$$\mathbf{P}^\top \mathbf{X} = f^2$$

onde  $\mathbf{P}$  é o ponto principal e  $f$  é a distância focal.

Como qualquer ponto de fuga pertence ao plano da imagem, tem-se para o ponto de fuga segundo  $\mathbf{x}$ :

$$\begin{cases} \mathbf{P}^\top \mathbf{F}_x = -f \mathbf{c}^\top \mathbf{F}_x = f^2 \\ \mathbf{F}_x^\top \mathbf{y} = \mathbf{F}_x^\top \mathbf{z} = 0 \end{cases} \Leftrightarrow \mathbf{F}_x = -f \frac{1}{\mathbf{c}^\top \mathbf{x}} \mathbf{x} = -f \left( \frac{1}{c_x}, 0, 0 \right)^\top$$

Nas coordenadas do espaço, o vector que identifica este ponto no plano é (o sinal negativo deve-se à inversão da imagem no processo de focagem):

$$\mathbf{U}_x = -(\mathbf{F}_x - \mathbf{P}) = -f \left( -\frac{1}{\mathbf{c}^\top \mathbf{x}} \mathbf{x} + \mathbf{c} \right)$$

Assim, exprimindo  $\mathbf{F}_x = u_x \mathbf{a} + v_x \mathbf{b} = (u_x, v_x)^\top$  nas coordenadas do plano da imagem, vem:

$$\begin{cases} u_x = \mathbf{U}_x \mathbf{a} = -f \left( -\frac{1}{\mathbf{c}^\top \mathbf{x}} \mathbf{x} + \mathbf{c} \right) \mathbf{a} = f \frac{a_x}{c_x} \\ v_x = \mathbf{U}_x \mathbf{b} = -f \left( -\frac{1}{\mathbf{c}^\top \mathbf{x}} \mathbf{x} + \mathbf{c} \right) \mathbf{b} = f \frac{b_x}{c_x} \end{cases} \Leftrightarrow \mathbf{F}_x = f \left( \frac{a_x}{c_x}, \frac{b_x}{c_x} \right)^\top$$

Do mesmo modo,

$$\mathbf{F}_y = f \left( \frac{a_y}{c_y}, \frac{b_y}{c_y} \right)^\top$$

e

$$\mathbf{F}_z = f \left( \frac{a_z}{c_z}, \frac{b_z}{c_z} \right)^\top$$

Pode demonstrar-se que a estas coordenadas dos pontos de fuga correspondem os vectores homogéneos:

$$\begin{cases} \mathbf{F}_x \mapsto (fa_x, fb_x, c_x)^\top \\ \mathbf{F}_y \mapsto (fa_y, fb_y, c_y)^\top \\ \mathbf{F}_z \mapsto (fa_z, fb_z, c_z)^\top \end{cases}$$

cujo uso, que faremos em diante, contempla o caso em que um ou mais dos pontos de fuga não são afins, mantendo assim a generalidade.

Para exprimir os pontos de fuga como função da orientação podemos por fim usar a Equação (2.3) em conjunto com a Equação (2.1), o que nos permite obter em coordenadas cartesianas:

$$\begin{cases} \mathbf{F}_x = f \left( \frac{a_x}{c_x}, \frac{b_x}{c_x} \right)^\top = f \mathbf{R}_\gamma \left( \frac{\tan \alpha}{\cos \beta}, \tan \beta \right)^\top \\ \mathbf{F}_y = f \left( \frac{a_y}{c_y}, \frac{b_y}{c_y} \right)^\top = f \mathbf{R}_\gamma \left( -\frac{\cot \alpha}{\cos \beta}, \tan \beta \right)^\top \\ \mathbf{F}_z = f \left( \frac{a_z}{c_z}, \frac{b_z}{c_z} \right)^\top = f \mathbf{R}_\gamma (0, -\cot \beta)^\top \end{cases}$$

em coordenadas homogéneas:

$$\begin{cases} \tilde{\mathbf{F}}_x = (fa_x, fb_x, c_x)^\top = \mathbf{H}_{f,\gamma} (\sin \alpha, \cos \alpha \sin \beta, \cos \alpha \cos \beta)^\top \\ \tilde{\mathbf{F}}_y = (fa_y, fb_y, c_y)^\top = \mathbf{H}_{f,\gamma} (\cos \alpha, -\sin \alpha \sin \beta, -\sin \alpha \cos \beta)^\top \\ \tilde{\mathbf{F}}_z = (fa_z, fb_z, c_z)^\top = \mathbf{H}_{f,\gamma} (0, \cos \beta, -\sin \beta)^\top \end{cases}$$

com

$$\mathbf{H}_{f,\gamma} = \begin{bmatrix} f \mathbf{R}_\gamma & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix} = \begin{bmatrix} f \cos \gamma & -f \sin \gamma & 0 \\ f \sin \gamma & f \cos \gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

e, por fim, em notação complexa:

$$\begin{cases} \bar{\mathbf{F}}_x = f \left( \frac{\tan \alpha}{\cos \beta} + j \tan \beta \right) \exp(j\gamma) \\ \bar{\mathbf{F}}_y = f \left( -\frac{\cot \alpha}{\cos \beta} + j \tan \beta \right) \exp(j\gamma) \\ \bar{\mathbf{F}}_z = f \cot \beta \exp \left[ j \left( \gamma - \frac{\pi}{2} \right) \right] = -j f \cot \beta \exp(j\gamma) \end{cases}$$

## A.5 Cálculo das orientações equiprojectivas

As orientações equiprojectivas são um conceito introduzido na Secção (2.4). Aqui vai proceder-se à demonstração do Teorema (2.11), que é de grande importância para a estimação da orientação.

Antes, porém, convém clarificar o significado geométrico das orientações equiprojectivas. Dizer que duas orientações distintas conduzem ao mesmo conjunto de pontos de fuga equivale a dizer que possuem alguma relação de simetria no espaço tridimensional.

Se esquecermos por agora o ângulo de torção  $\gamma$  e nos detivermos apenas na posição espacial do ponto principal correspondente a cada uma das orientações que pertencem à mesma classe de equivalência, somos levados a suspeitar que:

- As orientações  $\Psi$ ,  $\Psi'$ ,  $\Psi''$  e  $\Psi'''$  que se representam na Figura (A.3a), cujos pontos principais se encontram repartidos por cada um dos quatro oitavos de esfera<sup>2</sup> possíveis, são equiprojectivas após uma escolha judiciosa do ângulo de torção  $\gamma$ ;
- As orientações  $\Psi_1$ ,  $\Psi_2$  e  $\Psi_3$  representadas na Figura (A.3b), cujos pontos principais se situam no mesmo oitavo de esfera (definido por  $x \leq 0 \wedge y \geq 0 \wedge z \geq 0 \wedge x^2 + y^2 + z^2 = f$ ) são também equiprojectivas a menos de uma rotação  $\gamma$  do plano da imagem.

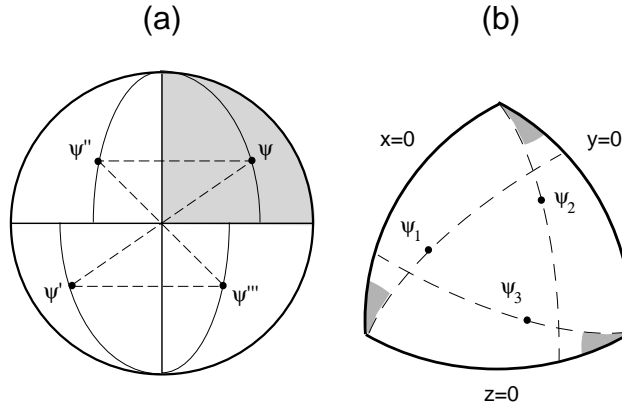


Figura A.3: Representação das orientações equiprojectivas no espaço tridimensional. Em (a) mostra-se como extrapolar um ponto principal de um oitavo da esfera para os restantes três oitavos possíveis; em (b) mostram-se as orientações equiprojectivas cujo ponto principal se encontra no mesmo oitavo de esfera.

Passemos agora para o plano da imagem. O caso ilustrado na Figura (A.3a) obtém-se a partir das duas propriedades que seguem, e que são uma consequência imediata da Equação (2.6):

**Propriedade A.1.** *As orientações  $\Psi = (\alpha, \beta, \gamma)$  e  $\Psi' = (-\alpha, -\beta, \gamma \pm \pi)$  são equiprojectivas. Não ocorre permutação dos pontos de fuga na transformação de  $\Psi$  em  $\Psi'$ .*

**Propriedade A.2.** *As orientações  $\Psi = (\alpha, \beta, \gamma)$  e  $\Psi'' = (\alpha \pm \frac{\pi}{2}, \beta, \gamma)$  são equiprojectivas. Obtêm-se permutando os pontos de fuga  $\mathbf{F}_x$  e  $\mathbf{F}_y$ .*

Pode assim determinar-se uma parte das orientações equiprojectivas. Em particular,  $\Psi''' = (-\alpha \pm \frac{\pi}{2}, -\beta, \gamma \pm \pi)$  também é equiprojectiva com  $\Psi$ ,  $\Psi'$  e  $\Psi''$ . Para obter toda a classe de equivalência basta calcular outras duas permutações distintas dos pontos de fuga.

<sup>2</sup>Da definição de orientação, o ponto principal pode situar-se na semi-esfera  $x^2 + y^2 + z^2 = f$  com  $x < 0$ . Os “quatro oitavos de esfera” referidos são as quatro partes iguais desta semi-esfera.

As Propriedades (A.1) e (A.2) permitem escolher no conjunto  $\{\Psi, \Psi', \Psi'', \Psi'''\}$  uma orientação  $\Psi = (\alpha, \beta, \gamma)$  tal que  $\alpha \in [0, \frac{\pi}{2}]$  e  $\beta \in [0, \frac{\pi}{2}]$ . Restringimo-nos assim ao oitavo de esfera representado na Figura (A.3b). A Equação (2.6) permite-nos então escrever:

$$\left\{ \begin{array}{l} \bar{\mathbf{F}}_1 = \bar{\mathbf{F}}_x(\Psi, f) = \\ \quad = f \left( \frac{\tan^2 \alpha}{\cos^2 \beta} + \tan^2 \beta \right)^{\frac{1}{2}} \exp \left[ j \left( \arctan \left( \frac{\sin \beta}{\tan \alpha} \right) + \gamma \right) \right] \\ \bar{\mathbf{F}}_2 = \bar{\mathbf{F}}_y(\Psi, f) = \\ \quad = f \left( \frac{\cot^2 \alpha}{\cos^2 \beta} + \tan^2 \beta \right)^{\frac{1}{2}} \exp \left[ j \left( \pi - \arctan \left( \frac{\sin \beta}{\cot \alpha} \right) + \gamma \right) \right] \\ \bar{\mathbf{F}}_3 = \bar{\mathbf{F}}_z(\Psi, f) = \\ \quad = f \cot \beta \exp \left[ j \left( -\frac{\pi}{2} + \gamma \right) \right] \end{array} \right. \quad (\text{A.1})$$

em que se fez corresponder a cada ponto  $\mathbf{F}_x$ ,  $\mathbf{F}_y$  e  $\mathbf{F}_z$  os elementos de  $\mathbb{C}_\infty$  (respectivamente)  $\bar{\mathbf{F}}_1$ ,  $\bar{\mathbf{F}}_2$  e  $\bar{\mathbf{F}}_3$ .

Interessa-nos obter as orientações equiprojectivas “permutativas” devidas a considerar cada ponto de fuga  $\mathbf{F}_n$  ( $n = 1, 2, 3$ ) como sendo o ponto de fuga segundo  $\mathbf{x}$ , o que nos conduzirá a uma orientação  $\Psi_n = (\alpha_n, \beta_n, \gamma_n)$ . Temos assim para  $n = 1$  que  $\Psi_1 = (\alpha_1, \beta_1, \gamma_1) = (\alpha, \beta, \gamma)$ , de acordo com a Equação (A.1). As restantes orientações podem obter-se a partir das equações:

$$\left\{ \begin{array}{l} \bar{\mathbf{F}}_1 = \bar{\mathbf{F}}_x(\Psi_1) = \bar{\mathbf{F}}_z(\Psi_2) = \bar{\mathbf{F}}_y(\Psi_3) \\ \bar{\mathbf{F}}_2 = \bar{\mathbf{F}}_y(\Psi_1) = \bar{\mathbf{F}}_x(\Psi_2) = \bar{\mathbf{F}}_z(\Psi_3) \\ \bar{\mathbf{F}}_3 = \bar{\mathbf{F}}_z(\Psi_1) = \bar{\mathbf{F}}_y(\Psi_2) = \bar{\mathbf{F}}_x(\Psi_3) \end{array} \right. \quad (\text{A.2})$$

Mais uma vez, as Propriedades (A.1) e (A.2) permitem garantir que se possa sempre ter  $\alpha_n \in [0, \frac{\pi}{2}]$  e  $\beta_n \in [0, \frac{\pi}{2}]$ . Para cada  $n$  podem obter-se as outras orientações equiprojectivas fazendo  $\Psi'_n = (\alpha_n - \frac{\pi}{2}, \beta_n, \gamma_n)$ ,  $\Psi''_n = (-\alpha_n, -\beta_n, \gamma_n \pm \pi)$  e  $\Psi'''_n = (-\alpha_n - \frac{\pi}{2}, -\beta_n, \gamma_n \pm \pi)$ .

Demonstra-se que não existem outras orientações equiprojectivas senão as dos casos já referidos, pelo que temos um máximo de 4 orientações equiproprojectivas para cada  $n = 1, 2, 3$ , isto é, para cada oitavo de esfera. Assim, para qualquer conjunto  $\mathcal{F}$  de pontos de fuga, o número máximo de orientações equiproprojectivas é  $3 \times 4 = 12$ .

Resta-nos calcular  $\Psi_2$  e  $\Psi_3$ . Para o efeito consideramos, como vimos,  $\alpha_n \in [0, \frac{\pi}{2}]$  e  $\beta_n \in [0, \frac{\pi}{2}]$ ,  $\forall n \in \{1, 2, 3\}$ , o que nos permite transformar algumas implicações em equivalências.

- Cálculo de  $\beta_2$  e  $\gamma_2$ :

$$\begin{aligned}
& \overline{\mathbf{F}}_x(\Psi_1, f) = \overline{\mathbf{F}}_z(\Psi_2, f) \quad \Leftrightarrow \\
& \Leftrightarrow f \left( \frac{\tan^2 \alpha_1}{\cos^2 \beta_1} + \tan^2 \beta_1 \right)^{\frac{1}{2}} \exp \left[ j \left( \arctan \left( \frac{\sin \beta_1}{\tan \alpha_1} \right) + \gamma_1 \right) \right] = \\
& = f \cot \beta_2 \exp \left[ j \left( -\frac{\pi}{2} + \gamma_2 \right) \right] \quad \Rightarrow \\
& \Rightarrow \begin{cases} \cot^2 \beta_2 = \frac{\tan^2 \alpha_1}{\cos^2 \beta_1} + \tan^2 \beta_1 \\ -\frac{\pi}{2} + \gamma_2 = \arctan \left( \frac{\sin \beta_1}{\tan \alpha_1} \right) + \gamma_1 + 2k\pi, k \in \mathbb{Z} \end{cases} \quad \Leftrightarrow \\
& \Leftrightarrow \begin{cases} \beta_2 = \arcsin(\cos \alpha_1 \cos \beta_1) \\ \gamma_2 = \gamma_1 - \arctan \left( \frac{\tan \alpha_1}{\sin \beta_1} \right) \pm \pi \end{cases}
\end{aligned}$$

- Cálculo de  $\alpha_2$ :

$$\begin{aligned}
& \overline{\mathbf{F}}_z(\Psi_1, f) = \overline{\mathbf{F}}_y(\Psi_2, f) \quad \Leftrightarrow \\
& \Leftrightarrow f \cot \beta_1 \exp \left[ j \left( -\frac{\pi}{2} + \gamma_1 \right) \right] = \\
& = f \left( \frac{\cot^2 \alpha_1}{\cos^2 \beta_1} + \tan^2 \beta_1 \right)^{\frac{1}{2}} \exp \left[ j \left( \pi - \arctan \left( \frac{\sin \beta_1}{\cot \alpha_1} \right) + \gamma_1 \right) \right] \quad \Rightarrow \\
& \Rightarrow -\frac{\pi}{2} + \gamma_1 = \pi - \arctan \left( \frac{\sin \beta_2}{\cot \alpha_2} \right) + \gamma_2 + 2k\pi, k \in \mathbb{Z} \quad \Leftrightarrow \\
& \Leftrightarrow \alpha_2 = \arctan \left( \frac{\tan \beta_1}{\sin \alpha_1} \right)
\end{aligned}$$

onde se recorreu ao resultado anterior para  $\beta_2$ .

Idêntico procedimento pode ser seguido para calcular a outra orientação equiprojectiva  $\Psi_3 = (\alpha_3, \beta_3, \gamma_3)$ . Em resumo, obtém-se:

$$\begin{cases} \alpha_2 = \arctan \left( \frac{\tan \beta_1}{\sin \alpha_1} \right) \\ \beta_2 = \arcsin(\cos \alpha_1 \cos \beta_1) \\ \gamma_2 = \gamma_1 - \arctan \left( \frac{\tan \alpha_1}{\sin \beta_1} \right) \pm \pi \end{cases} \quad \text{e} \quad \begin{cases} \alpha_3 = \operatorname{arccot} \left( \frac{\tan \beta_1}{\cos \alpha_1} \right) \\ \beta_3 = \arcsin(\sin \alpha_1 \cos \beta_1) \\ \gamma_3 = \gamma_1 + \arctan \left( \frac{\cot \alpha_1}{\sin \beta_1} \right) \pm \pi \end{cases} \quad (\text{A.3})$$

## A.6 Delimitação da região que contém um ponto de fuga ao acaso

Nesta secção irão deduzir-se mais algumas propriedades relacionadas com a localização dos pontos de fuga no plano, o que culminará na demonstração do Teorema (2.13). Vamos começar por observar duas propriedades importantes que resultam da Equação (2.6):

**Propriedade A.3.** *A diferença de argumentos entre dois pontos de fuga distintos é maior ou igual que  $90^\circ$ . Por outras palavras, não existe mais do que um ponto de fuga em cada quadrante, pelo que os três pontos de fuga estão sempre distribuídos em três quadrantes do plano. Em linguagem matemática,*

$$\forall \mathbf{F}_1, \mathbf{F}_2 \in \mathcal{F}, |\arg \bar{\mathbf{F}}_1 - \arg \bar{\mathbf{F}}_2| \geq \frac{\pi}{2}$$

**Propriedade A.4.** *Para qualquer orientação  $\Psi = (\alpha, \beta, \gamma)$  existem pelo menos dois pontos de fuga no exterior (ou na fronteira) do círculo com raio  $f$  e centro na origem do sistema de coordenadas do plano. Em linguagem matemática,*

$$\forall \Psi, \exists \mathbf{F}_1, \mathbf{F}_2 \in \mathcal{F} : \mathbf{F}_1 \neq \mathbf{F}_2 \wedge \text{mod } \bar{\mathbf{F}}_i \geq \text{mod } \bar{\mathbf{F}}_j \geq f.$$

Como consequência, é impossível existir mais que um ponto no interior desse círculo.

*Demonstração.* A Propriedade (A.3) resulta directamente das expressões já estudadas para os pontos de fuga. Em relação à Propriedade (A.4), existem dois cenários possíveis:

- Se  $\beta \in [-\frac{\pi}{2}, -\frac{\pi}{4}] \cup [\frac{\pi}{4}, \frac{\pi}{2}]$  tem-se  $\text{mod } \bar{\mathbf{F}}_x \geq \text{Im } \bar{\mathbf{F}}_x = f |\tan \beta| \geq f$ . Como  $\text{mod } \bar{\mathbf{F}}_y \geq \text{Im } \bar{\mathbf{F}}_y = \text{Im } \bar{\mathbf{F}}_x$  tem-se também  $\text{mod } \bar{\mathbf{F}}_y \geq f$ .
- Se  $\beta \in [-\frac{\pi}{4}, \frac{\pi}{4}]$  tem-se  $\text{mod } \bar{\mathbf{F}}_z = f |\cot \beta| \geq f$ . Tem-se ainda:

– Se  $\alpha \in [-\frac{\pi}{2}, -\frac{\pi}{4}] \cup [\frac{\pi}{4}, \frac{\pi}{2}]$ :

$$\begin{aligned} \text{mod } \bar{\mathbf{F}}_x &= f \sqrt{\frac{\tan^2 \alpha}{\cos^2 \beta} + \tan^2 \beta} = f \sqrt{\tan^2 \alpha + \frac{\tan^2 \beta}{\cos^2 \alpha}} \geq \\ &\geq f |\tan \alpha| \geq f \end{aligned}$$

– Se  $\alpha \in [-\frac{\pi}{4}, \frac{\pi}{4}]$ :

$$\begin{aligned} \text{mod } \bar{\mathbf{F}}_y &= f \sqrt{\frac{\cot^2 \alpha}{\cos^2 \beta} + \tan^2 \beta} = f \sqrt{\cot^2 \alpha + \frac{\tan^2 \beta}{\cos^2 \alpha}} \geq \\ &\geq f |\cot \alpha| \geq f \end{aligned}$$

pelo que:

$$\text{mod } \bar{\mathbf{F}}_z \geq f \wedge (\text{mod } \bar{\mathbf{F}}_x \geq f \vee \text{mod } \bar{\mathbf{F}}_y \geq f)$$

o que basta para concluir a demonstração. ■

Uma consequência da Propriedade (A.4) é:

**Propriedade A.5.** *Qualquer orientação da câmara é equiprojectiva com uma orientação  $\Psi = (\alpha, \beta, \gamma)$  com  $-\frac{\pi}{4} < \alpha \leq \frac{\pi}{4}$  e  $-\frac{\pi}{4} < \beta \leq \frac{\pi}{4}$ .*

Isto leva-nos a colocar a seguinte questão:

**Problema A.6.** *Qual o menor  $\varphi$  (chamemos-lhe  $\varphi_{\min}$ ) que faz com que qualquer orientação seja equiprojectiva com uma orientação  $\Psi = (\alpha, \beta, \gamma)$  com  $-\frac{\pi}{4} < \alpha \leq \frac{\pi}{4}$ ,  $-\frac{\pi}{4} < \beta \leq \frac{\pi}{4}$  e  $-\varphi_{\min} < \gamma \leq \varphi_{\min}$ ?*

Poderia pensar-se que, por mera analogia com  $\alpha$  e com  $\beta$ , se teria  $\varphi_{\min} = \frac{\pi}{4}$ . No entanto, o exemplo  $\alpha = -10^\circ$ ,  $\beta = 50^\circ$  e  $\gamma = 35^\circ$  comprova que não, como podemos verificar por observação da Figura (A.4). Como calcular então  $\varphi_{\min}$ ?

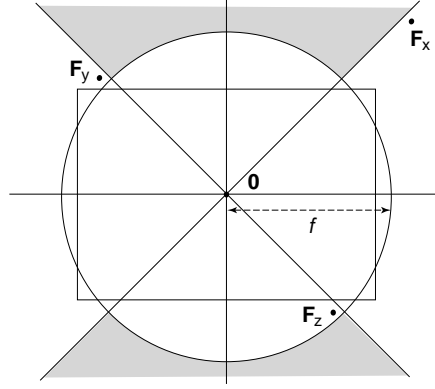


Figura A.4: Localização dos pontos de fuga com  $\alpha = -10^\circ$ ,  $\beta = 50^\circ$  e  $\gamma = 35^\circ$ . A região sombreada é a definida por  $\{\Psi = (\alpha, \beta, \gamma) : -\frac{\pi}{4} < \alpha \leq \frac{\pi}{4} \wedge -\frac{\pi}{4} < \beta \leq \frac{\pi}{4} \wedge -\frac{\pi}{4} < \gamma \leq \frac{\pi}{4}\}$ . Pode ver-se que todos os pontos de fuga se localizam fora desta região.

Para resolver este problema, comecemos por definir os conjuntos seguintes:

$$\mathcal{B} = \{\mathbf{F} \in \mathbb{C}_\infty : \text{mod } \mathbf{F} \geq f\}$$

e

$$\mathcal{G}_\varphi = \left\{ \mathbf{F} \in \mathbb{C}_\infty : \arg \mathbf{F} \in \left[ -\frac{\pi}{2} - \varphi, -\frac{\pi}{2} + \varphi \right] \cup \left[ \frac{\pi}{2} - \varphi, \frac{\pi}{2} + \varphi \right] \right\}$$

As zonas do plano definidas por estes conjuntos estão representadas na Figura (A.5).

Resolver o Problema (A.6) equivale a obter o menor  $\varphi$  que garante a existência de pelo menos um ponto de fuga em  $\mathcal{B} \cap \mathcal{G}_\varphi$ . Para isso vamos recorrer às propriedades obtidas atrás. A Propriedade (A.3) garante que se tenha necessariamente  $\frac{\pi}{4} < \varphi_{\min} \leq \frac{\pi}{2}$ . A Propriedade (A.4), por seu turno, garante que existam pelo menos dois pontos de fuga em  $\mathcal{B}$ . Temos assim dois casos possíveis:

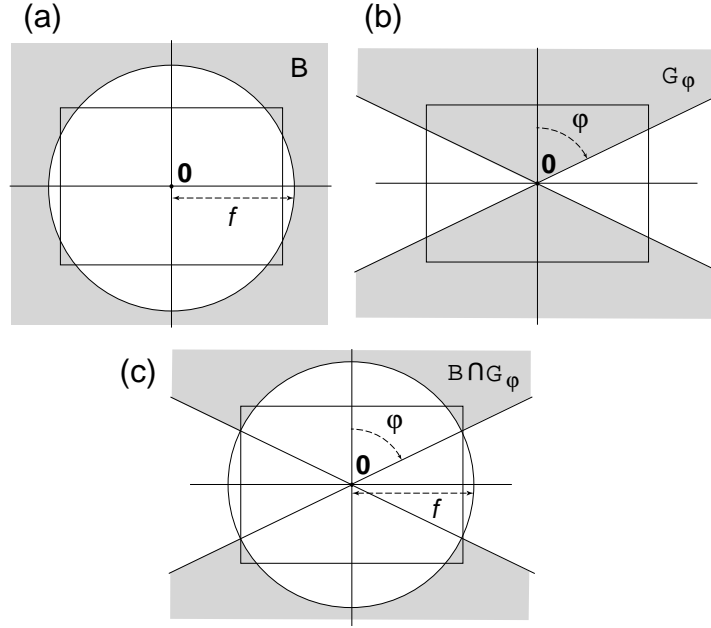


Figura A.5: Zonas do plano representadas pelos conjuntos  $\mathcal{B}$ ,  $\mathcal{G}_\varphi$  e  $\mathcal{B} \cap \mathcal{G}_\varphi$ . Em (a) a Propriedade (A.4) garante a existência de pelo menos 2 pontos de fuga na zona a sombreado; em (b) a Propriedade (A.3) garante existir pelo menos 1 ponto de fuga na região sombreada quando se tem  $\varphi \geq \frac{\pi}{4}$ .

- Se os três pontos de fuga pertencerem a  $\mathcal{B}$ , a Propriedade (A.3) garante que nenhum par desses pontos tem uma diferença de argumentos superior a  $\frac{\pi}{2}$ , pelo que de  $\frac{\pi}{4} < \varphi_{\min} \leq \varphi \leq \frac{\pi}{2}$  pode concluir-se, por simples observação da Figura (A.5b), que existe um ponto de fuga em  $\mathcal{G}_\varphi$ . Como tal, esse ponto de fuga pertence também a  $\mathcal{B} \cap \mathcal{G}_\varphi$ . Esta situação exige apenas que  $\varphi \geq \frac{\pi}{4}$ , pelo que não é restritiva;
- Se existirem dois pontos de fuga em  $\mathcal{B}$  e um em  $\mathcal{B}^c$ , este último (que vamos designar apenas por  $\mathbf{F}$ ) pode pertencer a  $\mathcal{B}^c \cap \mathcal{G}_\varphi$  ou a  $\mathcal{B}^c \cap \mathcal{G}_\varphi^c$ . O segundo caso traz como consequência, devido à Propriedade (A.3), a existência de um outro ponto de fuga em  $\mathcal{G}_\varphi$ . Como este último ponto de fuga está localizado em  $\mathcal{B}$ , pertencerá também a  $\mathcal{B} \cap \mathcal{G}_\varphi$ , pelo que esta situação também não é restritiva.

Por conseguinte, resta averiguar o caso em que  $\mathbf{F} \in \mathcal{B}^c \cap \mathcal{G}_\varphi$ . Em particular, interessa-nos achar o menor valor de  $\varphi$  que torna verdadeira a implicação:

$$\mathbf{F} \in \mathcal{F} \cap \mathcal{B}^c \cap \mathcal{G}_\varphi \Rightarrow \exists \mathbf{F}' \in \mathcal{F} : \mathbf{F}' \neq \mathbf{F} \wedge \mathbf{F}' \in \mathcal{G}_\varphi$$

o que faremos, sem perder generalidade, a partir da suposição de que:

$$\arg \mathbf{F} \in \left[ -\frac{\pi}{2}, -\frac{\pi}{2} + \varphi \right]$$

Assim, se supusermos adicionalmente uma parametrização em que  $\mathbf{F}$  é o ponto de fuga segundo  $\mathbf{z}$ , teremos  $\beta \in [0, \frac{\pi}{4}]$  e  $\gamma \in [0, \varphi]$ . Por outro lado, socorrendo-nos novamente



da Propriedade (A.3), é possível afirmar que, sendo  $\mathcal{F} = \{\mathbf{F}, \mathbf{F}', \mathbf{F}''\}$ , se tem

$$\arg \mathbf{F}' \in \left[0, \frac{\pi}{2} + \varphi\right]$$

e

$$\arg \mathbf{F}'' \in \left[-\pi, -\frac{\pi}{2}\right] \cup \left[\frac{\pi}{2}, \pi\right]$$

o que corresponde a  $\alpha \in \left[0, \frac{\pi}{2}\right]$ .

Tem-se assim, das expressões da Equação (A.3) para as orientações equiprojectivas:

$$\begin{cases} \arg \mathbf{F}' &= -\frac{\pi}{2} + \gamma' = -\frac{\pi}{2} + \gamma - \arctan\left(\frac{\tan \alpha}{\sin \beta}\right) \pm \pi = \\ &= \arctan(\cot \alpha \sin \beta) + \gamma \\ \arg \mathbf{F}'' &= -\frac{\pi}{2} + \gamma'' = -\frac{\pi}{2} + \gamma + \arctan\left(\frac{\cot \alpha}{\sin \beta}\right) \pm \pi = \\ &= \arctan(-\tan \alpha \sin \beta) + \pi + \gamma \end{cases}$$

Para  $\varphi \geq \varphi_{\min}$ , nenhuma orientação  $\Psi = (\alpha, \beta, \gamma)$  poderá então satisfazer simultaneamente:

$$\begin{cases} \arg \mathbf{F}' &\in \left]-\frac{\pi}{2} + \varphi, \frac{\pi}{2} - \varphi\right[ \\ \arg \mathbf{F}'' &\in \left]\frac{\pi}{2} + \varphi, \frac{3\pi}{2} - \varphi\right[ \end{cases}$$

com

$$\alpha \in \left[0, \frac{\pi}{2}\right] \quad , \quad \beta \in \left[0, \frac{\pi}{4}\right] \quad \text{e} \quad \gamma \in [0, \varphi]$$

Logo, deve ser impossível o sistema de inequações:

$$\begin{aligned} &\begin{cases} -\frac{\pi}{2} + \varphi < \arctan(\cot \alpha \sin \beta) + \gamma < \frac{\pi}{2} - \varphi \\ \frac{\pi}{2} + \varphi < \arctan(-\tan \alpha \sin \beta) + \pi + \gamma < \frac{3\pi}{2} - \varphi \end{cases} \Leftrightarrow \\ &\Leftrightarrow \begin{cases} \arctan(\cot \alpha \sin \beta) + \gamma < \frac{\pi}{2} - \varphi \\ \arctan(\tan \alpha \sin \beta) - \pi - \gamma < -\frac{\pi}{2} - \varphi \end{cases} \end{aligned}$$

pois  $\cot \alpha \geq 0$ ,  $\sin \beta \geq 0$  e  $\tan \alpha \geq 0$ .

O valor mais “desfavorável” de  $\beta$  neste sistema é  $\beta = \frac{\pi}{4}$ . Concretizando, resulta a desigualdade simples:

$$f(x) = \arctan x + \arctan \frac{1}{2x} < \pi - 2\varphi$$

onde se fez  $x = \frac{\sqrt{2}}{2} \tan \alpha$ .

O máximo de  $f$  pode ser calculado através do zero da derivada:

$$\begin{aligned} f'(x) &= 0 \Leftrightarrow \\ \Leftrightarrow \frac{1}{1+x^2} + \frac{1}{1+\left(\frac{1}{4x^2}\right)} \cdot \left(-\frac{1}{2x^2}\right) &= 0 \Leftrightarrow \\ \Leftrightarrow x = \pm \frac{\sqrt{2}}{2} \Leftrightarrow \alpha &= \frac{\pi}{4} \end{aligned}$$

Logo,

$$\begin{aligned} \max f(x) &= f\left(\pm \frac{\sqrt{2}}{2}\right) = \arctan \frac{\sqrt{2}}{2} + \arctan \frac{1}{2\frac{\sqrt{2}}{2}} = \\ &= 2 \arctan \frac{\sqrt{2}}{2} \end{aligned}$$

Portanto, como a condição  $\forall x \geq 0, f(x) < \pi - 2\varphi$  deve ser impossível para  $\varphi \geq \varphi_{\min}$ , essa impossibilidade é garantida se e só se:

$$\begin{aligned} \max f(x) &> \pi - 2\varphi \Leftrightarrow \\ \Leftrightarrow 2 \arctan \frac{\sqrt{2}}{2} &> \pi - 2\varphi \Leftrightarrow \\ \Leftrightarrow \varphi &> \frac{\pi}{2} - \arctan \frac{\sqrt{2}}{2} \end{aligned}$$

Conclui-se assim que:

$$\varphi_{\min} = \frac{\pi}{2} - \arctan \frac{\sqrt{2}}{2} \approx 54.7^\circ$$

## A.7 Variação do ângulo de compasso na hipótese de pequenas rotações

Nesta secção deduz-se a expressão que relaciona a variação do ângulo de compasso com o ângulo de elevação, nas condições da hipótese de pequenas rotações.

Para o efeito, comecemos por notar que, de acordo com esta hipótese, o ponto principal da imagem sofre a rotação de um ângulo não superior a  $\varphi_{\max}$  segundo um eixo arbitrário, e portanto no instante  $t+1$  fica situado no “círculo esférico”<sup>3</sup> a sombreado na Figura (A.6). O “raio” deste círculo é um arco de comprimento (em unidades de ângulo) igual a  $\varphi_{\max}$ , sendo o seu centro o ponto principal da imagem no instante  $t$ , que designamos por  $\mathbf{P}(t)$ . O ponto principal em  $t+1$  será designado por  $\mathbf{P}(t+1)$ .

Se no instante  $t+1$  o ângulo de elevação  $\beta(t+1)$  for conhecido e apenas interessar o cálculo do ângulo de compasso (o que sucede, como veremos adiante, a dada altura no algoritmo de estimação da orientação em duas partes) então  $\mathbf{P}(t+1)$  fica confinado ao arco de circunferência representado a negrito na Figura (A.6), e que é paralelo ao plano  $xy$ .

---

<sup>3</sup>Este “círculo esférico” não é mais do que a intersecção de uma superfície de esfera com uma esfera “maciça”, a primeira de raio igual à distância focal e centro em  $\mathbf{0}$  e a outra de raio supostamente pequeno (função de  $\varphi_{\max}$ ) e centro no ponto principal da imagem no instante  $t$ .

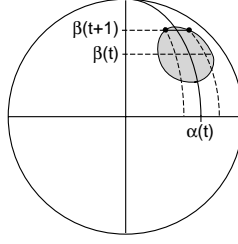


Figura A.6: Aplicação da hipótese de pequenas rotações: a sombreado representa-se a região possível para o ponto principal no instante  $t + 1$ ,  $\mathbf{P}(t + 1)$ .

Assim, os valores-limite para o ângulo de compasso  $\alpha(t + 1)$  serão os valores de  $\alpha$  correspondentes aos dois extremos desse arco. Se  $\mathbf{X}$  é um ponto da esfera de raio  $f$  contido no “círculo esférico”, então obedece à equação:

$$\arccos \frac{1}{f^2} \mathbf{X}^\top \mathbf{P}(t) \leq \varphi_{\max}$$

pelo que vem da Equação (2.3):

$$\begin{cases} \mathbf{P}(t) &= f [-\cos \alpha(t) \cos \beta(t), \sin \alpha(t) \cos \beta(t), \sin \beta(t)]^\top \\ \mathbf{P}(t+1) &= f [-\cos \alpha(t+1) \cos \beta(t+1), \sin \alpha(t+1) \cos \beta(t+1), \sin \beta(t+1)]^\top \end{cases}$$

Substituindo  $\mathbf{X} = \mathbf{P}(t+1)$ :

$$\begin{aligned} \frac{1}{f^2} |\mathbf{P}^\top(t) \mathbf{P}(t+1)| &= |\cos \alpha(t) \cos \beta(t) \cos \alpha(t+1) \cos \beta(t+1) + \\ &\quad + \sin \alpha(t) \cos \beta(t) \sin \alpha(t+1) \cos \beta(t+1) + \\ &\quad + \sin \beta(t) \sin \beta(t+1)| \geq \\ &\geq \cos \varphi_{\max} \end{aligned}$$

Desenvolvendo, e tendo em conta que  $|\beta(t)| \leq \frac{\pi}{2}$  e  $|\beta(t+1)| \leq \frac{\pi}{2}$ , vem:

$$\begin{aligned} &|\cos \beta(t) \cos \beta(t+1) [\cos \alpha(t) \cos \alpha(t+1) + \sin \alpha(t) \sin \alpha(t+1)] + \\ &\quad + \sin \beta(t) \sin \beta(t+1)| \geq \cos \varphi_{\max} \Leftrightarrow \\ \Leftrightarrow &|\cos \beta(t) \cos \beta(t+1) \cos [\alpha(t+1) - \alpha(t)] + \sin \beta(t) \sin \beta(t+1)| \geq \cos \varphi_{\max} \Leftrightarrow \\ \Leftrightarrow &|\cos \Delta \alpha| \leq \frac{-\sin \beta(t) \sin \beta(t+1) \pm \cos \varphi_{\max}}{\cos \beta(t) \cos \beta(t+1)} \leq \\ &\leq 1 - \frac{\cos \Delta \beta - \cos \varphi_{\max}}{\cos \beta(t) \cos \beta(t+1)} \end{aligned}$$

onde se fez:

$$\begin{cases} \Delta \alpha = \alpha(t+1) - \alpha(t) \\ \Delta \beta = \beta(t+1) - \beta(t) \end{cases}$$

Por outro lado, o ângulo de compasso  $\alpha(t+1)$  não fica sujeito a qualquer restrição devida à hipótese de pequenas rotações se se tiver:

$$\left| 1 - \frac{\cos \Delta\beta - \cos \varphi_{\max}}{\cos \beta(t) \cos \beta(t+1)} \right| > 1 \quad \Leftrightarrow \quad |\beta(t)\beta(t+1)| > \pi - \varphi_{\max}$$

Da expressão percebe-se que a capacidade de restringir o valor do ângulo de compasso em  $t+1$  é tanto maior quando mais próximo de zero for o ângulo de elevação. Essa capacidade perde-se rapidamente à medida que o ponto principal se move em direcção a um dos pólos da esfera. Para combater este efeito é útil valermos-nos da teoria estabelecida atrás acerca das orientações equiprojectivas: ao se utilizar instantaneamente um valor de  $\beta$  não necessariamente correspondente à orientação nesse instante, mas a uma pertencente à sua classe de equivalência cujo ponto principal se localize mais próximo do plano xy, garante-se que este nunca se precipite para um dos pólos.

## Apêndice B

# Notas Adicionais Sobre a Malha de Extracção de Informação

### B.1 Determinação do valor máximo do módulo do gradiente na utilização de máscaras de Prewitt, Sobel e similares

Vimos na Secção (3.2) que os filtros de Prewitt e de Sobel são aplicados à matriz  $\mathbf{I}$  de intensidade luminosa através de máscaras do tipo:

$$\mathbf{F}_x = \begin{bmatrix} -1 & 0 & 1 \\ -a & 0 & a \\ -1 & 0 & 1 \end{bmatrix} \quad \text{e} \quad \mathbf{F}_y = -\mathbf{F}_x^\top$$

e que o módulo do gradiente é dado através da expressão:

$$\text{mod } \nabla \mathbf{I}(\mathbf{u}) = \sqrt{[\nabla \mathbf{I}_x(\mathbf{u})]^2 + [\nabla \mathbf{I}_y(\mathbf{u})]^2}$$

com:

$$\begin{cases} \nabla \mathbf{I}_x = \mathbf{F}_x * \mathbf{I} \\ \nabla \mathbf{I}_y = \mathbf{F}_y * \mathbf{I} \end{cases}$$

Pretende agora determinar-se o máximo valor que  $\text{mod } \nabla \mathbf{I}(x, y)$  pode apresentar, supondo à partida que  $\mathbf{I}$  está normalizada, isto é, que:

$$\forall \mathbf{u} : 0 \leq \mathbf{I}(\mathbf{u}) \leq 1$$

Para determinar este máximo, vamos procurar uma submatriz de dimensões  $3 \times 3$  “contida” em  $\mathbf{I}$  cujos elementos multiplicados pelos elementos de  $\mathbf{F}_x$  e  $\mathbf{F}_y$  somam um total que maximiza, no ponto  $\mathbf{u}$  considerado, a expressão  $[\nabla \mathbf{I}_x(\mathbf{u})]^2 + [\nabla \mathbf{I}_y(\mathbf{u})]^2$ .

Seja  $\mathbf{A}$  essa matriz. Então, por observação das máscaras  $\mathbf{F}_x$  e  $\mathbf{F}_y$ , e tendo em conta que  $\mathbf{F}_x(1, 2) = \mathbf{F}_x(2, 2) = \mathbf{F}_x(3, 2) = 0$  e  $\mathbf{F}_y(2, 1) = \mathbf{F}_y(2, 2) = \mathbf{F}_y(2, 3) = 0$ ,  $\mathbf{A}$  deve

ser (a menos de uma transposição) da forma:

$$\mathbf{A} = \begin{bmatrix} x_1 & 1 & x_2 \\ 1 & x & 0 \\ x_3 & 0 & x_4 \end{bmatrix} \quad (\text{B.1})$$

onde  $x$  pode tomar qualquer valor e  $x_1, x_2, x_3$  e  $x_4$  são valores tais que maximizam a expressão:

$$\begin{aligned} E(x_1, x_2, x_3, x_4) &= (a + x_1 + x_2 - x_3 - x_4)^2 + (a + x_1 - x_2 + x_3 - x_4)^2 = \\ &= 2(a + x_1 - x_4)^2 + 2(x_2 - x_3)^2 \end{aligned}$$

Uma simples observação desta expressão é suficiente para verificar que esses valores são (a menos de uma permutação entre  $x_2$  e  $x_3$ ):

$$x_1 = x_2 = 1 \quad \text{e} \quad x_3 = x_4 = 0$$

pelo que da Equação (B.1) vem:

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & x & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

O valor máximo do módulo do gradiente,  $x_{\max} = \max \text{mod } \nabla \mathbf{I}$ , é então dado por:

$$\begin{aligned} x_{\max} &= \left[ \left( \sum_{\mathbf{u}} \mathbf{A}(\mathbf{u}) \mathbf{F}_x(\mathbf{u}) \right)^2 + \left( \sum_{\mathbf{u}} \mathbf{A}(\mathbf{u}) \mathbf{F}_y(\mathbf{u}) \right)^2 \right]^{\frac{1}{2}} = \\ &= \sqrt{E(x_1, x_2, x_3, x_4)} = \\ &= \sqrt{2(1+a)^2 + 2} \end{aligned}$$

ou seja,  $x_{\max} = \sqrt{10}$  quando se utilizam máscaras de Prewitt ( $a = 1$ ) e  $x_{\max} = \sqrt{20}$  quando se utilizam máscaras de Sobel ( $a = 2$ ).

Deve no entanto notar-se que é difícil encontrar numa imagem vulgar, obtida através de uma fotografia, uma submatriz  $\mathbf{A}$  que apresente tais “descontinuidades” entre os seus elementos. Por outro lado, esta improbabilidade transforma-se mesmo em impossibilidade quando a imagem é previamente suavizada através de um filtro de Gauss, situação em que  $x_{\max} = \max \text{mod } \nabla \mathbf{I}$  é *substancialmente inferior* ao valor calculado.

De qualquer forma, pode afirmar-se com segurança que para qualquer imagem  $\mathbf{I}$ , o valor de  $x_{\max}$  calculado é sempre um *majorante* do conjunto  $\{\text{mod } \nabla \mathbf{I}(\mathbf{u})\}$ .

## B.2 Testes comparativos do desempenho de vários medidores de contornos

Para além do medidor de contornos baseado nos filtros de Prewitt e de Sobel, estudado na Secção (3.2), testaram-se outros medidores de contornos um pouco mais complexos. Segue-se um resumo dos resultados obtidos, assim como uma breve descrição de cada medidor experimentado.

- *Filtro Laplaciano da Gaussiana (LoG)*. A ideia que serve de base a este medidor de contornos consiste em identificar pontos de contornos com os zeros da convolução da matriz de intensidade luminosa  $\mathbf{I}$  com a primeira derivada da função gaussiana. Dada a simetria da função gaussiana, só faz sentido estimar a direcção dos contornos se se utilizar em vez da função gaussiana bidimensional duas funções gaussianas direccionais, uma que convolui na horizontal e a outra na vertical. Experimentando com máscaras de várias dimensões e fazendo variar o desvio padrão, pôde concluir-se que os resultados obtidos não são mais satisfatórios do que qualquer um dos filtros de Prewitt e de Sobel, sendo o tempo de processamento superior, pelo que não se justifica o uso deste filtro.
- *Filtro de Canny*. O filtro de Canny é o mais frequentemente utilizado na detecção de contornos. A descrição do seu funcionamento encontra-se explicada em detalhe em [6]. Este filtro não foi utilizado na estimação do gradiente, mas sim, entre outras coisas, no treino para estimar as probabilidades a priori, como se pode ver na Secção (4.2). No entanto, algumas funcionalidades implementadas pelo filtro de Canny têm interesse na malha de extracção de informação, nomeadamente na eliminação de pontos redundantes na vizinhança de contornos<sup>1</sup>.
- *Medidor de contornos SUSAN*<sup>2</sup>. O funcionamento deste medidor de contornos pode ser encontrado em [7], baseando-se no uso de máscaras circulares para determinar o “centro de massa” de pontos próximos de contornos. As experiências efectuadas demonstraram que é muito eficaz na estimação do módulo do gradiente, o que é pouco interessante na óptica do problema em estudo, e tão eficaz na estimação da direcção dos contornos quanto os medidores utilizando filtros de Prewitt e de Sobel. Uma desvantagem importante, que nos levou a abandonar o seu uso, prende-se com o facto de nas experiências efectuadas se terem obtido tempos de processamento mais de dez vezes superiores ao exigido por aqueles dois medidores, reduzindo assim drasticamente o desempenho quando se pretende um algoritmo rápido.
- *Medidor de contornos isométrico*. Este medidor de contornos é o resultado de um estudo que pode ser encontrado em [9]. Em vez de se estimar o gradiente somente através de máscaras na direcção horizontal e vertical, como nos métodos de Prewitt e Sobel, convolui-se a matriz de intensidade luminosa  $\mathbf{I}$  também com máscaras que representam as duas diagonais (direcções a  $-45^\circ$  e a  $45^\circ$  com a vertical), fazendo-se depois uma ponderação apropriada que visa otimizar a estimação da direcção

---

<sup>1</sup>Esta funcionalidade do algoritmo proposto por Canny designa-se na literatura anglo-saxónica por ‘non-maxima supression’.

<sup>2</sup>“Smallest Univalued Segment Assimilating Nucleus”.

do gradiente. No entanto, mais uma vez os resultados obtidos não foram tão satisfatórios quanto o aumento do tempo de processamento exigido, estando este próximo do dobro do tempo de execução dos métodos de Prewitt e Sobel.

É muito importante neste trabalho uma boa precisão na estimação do ângulo do gradiente  $\arg \nabla \mathbf{I}$ , pois todo o processamento estatístico que se segue depende fortemente das direcções estimadas para os contornos.

Assim, foi efectuado um teste de precisão na estimação de  $\arg \nabla \mathbf{I}$  que consistiu na aplicação de vários medidores de contornos à fotografia de uma aresta recta (com uma inclinação de  $79^\circ$  com a vertical), que se representa na Figura (B.1).

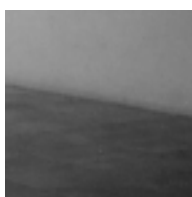


Figura B.1: Fotografia utilizada nos testes aplicados aos vários medidores de contornos. A inclinação do contorno com a vertical é de  $79^\circ$ .

Na Tabela (B.1) apresentam-se os resultados comparativos de alguns métodos experimentados. Deve notar-se que a diferença entre a média dos ângulos estimados e o ângulo real de  $79^\circ$  tem como causa principal a presença de contornos fracos aproximadamente horizontais ( $90^\circ$  com a vertical), em pontos vizinhos de contornos fortes. Esta situação é remediada através da supressão de pontos vizinhos a que fazemos referência na Secção (3.4).

A análise dos dados da Tabela (B.1) leva a que a escolha recaia sobre o uso de filtros de Prewitt ou de Sobel. A opção recaiu sobre os primeiros, estando os resultados obtidos para o ângulo do gradiente representados no histograma da Figura (B.2). Neste histograma figuram apenas os valores do ângulo do gradiente em pontos que pertencem a contornos. O critério utilizado para saber que pontos pertencem a contornos foi um detector de contornos de Canny.



	Prewitt	Sobel	LoG	Isotrópico	SUSAN
Tamanho das máscaras utilizadas	$3 \times 3$	$3 \times 3$	$7 \times 7$	$3 \times 3$	$7 \times 7$
Pré-filtragem gaussiana	Sim	Sim	Não*	Sim	Sim
Tamanho das máscaras no filtro gaussiano	$7 \times 7$	$7 \times 7$	-	$11 \times 11$	$7 \times 7$
Desvio padrão da função gaussiana	1.0	1.0	1.0	1.0/1.5	1.5
Média dos ângulos estimados	84.4°	84.0°	84.1°	84.6°	84.4°
Desvio padrão na estimação dos ângulos	4.5°	4.6°	5.0°	5.3°	6.0°
Relação qualidade/tempo de processamento	boa	boa	razoável	razoável	má

\*As máscaras do filtro LoG permitem estimar o gradiente e aplicar um filtro gaussiano ao mesmo tempo.

Tabela B.1: Resultados obtidos para vários medidores de contornos aplicados à imagem da Figura (B.1).

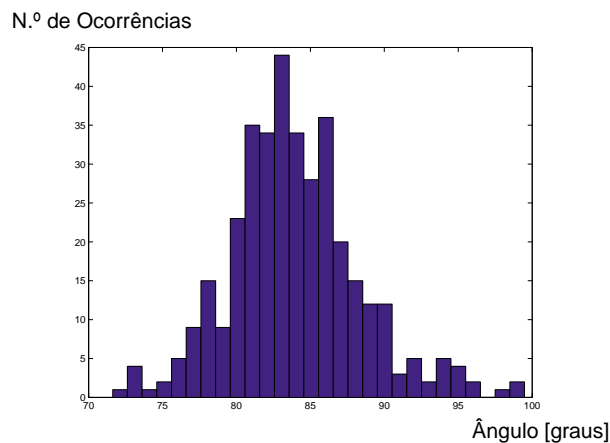


Figura B.2: Histograma dos valores estimados para o ângulo do gradiente para a imagem da Figura (B.1). O método utilizado foi o de Sobel.

## Apêndice C

# Estudo de métodos para acelerar a estimação da orientação

### C.1 Método da procura de máximo local da função de verosimilhança

Os métodos de força bruta para determinar a estimativa do máximo a posteriori, em que se percorre exaustivamente todas as possíveis orientações  $\Psi$  calculando a verosimilhança de cada uma, demonstram-se profundamente ineficientes. No caso do Algoritmo (4.1), por exemplo, os passos (2a), (2b) e (2c), que têm um considerável peso computacional, são executados ciclicamente um grande número de vezes, observando-se que para a maioria dos ciclos o ângulo de compasso  $\alpha$  se encontra afastado da estimativa do máximo a posteriori, como pode ser observado no gráfico da Figura (4.3). Isto sugere que se utilize um método mais eficiente.

Uma possibilidade (não abordada neste estudo) consiste em implementar um algoritmo adaptativo que através de iterações aproxime o valor da solução, o que exige o domínio de algumas técnicas de Análise Numérica. O raciocínio que aqui seguimos, no entanto, é extremamente simples e não faz uso destas técnicas, como pode ver-se no Algoritmo (C.1) que se apresenta de seguida:

**Algoritmo C.1.** *Dada uma função  $f(x)$  definida num domínio  $[x_{\min}, x_{\max}]$  em que a solução do problema  $x_0 = \arg \max f(x)$  se encontra presumivelmente dentro de um intervalo  $[x_{\text{pref min}}, x_{\text{pref max}}]$  e próxima de um valor de palpite  $x_{\text{palp}}$ , com*

$$x_{\min} \leq x_{\text{pref min}} \leq x_{\text{palp}} \leq x_{\text{pref max}} \leq x_{\max}$$

*o valor de  $x_0$  pode ser estimado buscando um máximo local da seguinte forma:*

1. *Fazendo por exemplo  $\varepsilon = 5$  começa por calcular-se  $f(x)$  na vizinhança discreta de  $x_{\text{palp}}$  dada por*

$$V(x_{\text{palp}}) = \{x_{\text{palp}} - \varepsilon\Delta x, \dots, x_{\text{palp}} - \Delta x, x_{\text{palp}}, x_{\text{palp}} + \Delta x, \dots, x_{\text{palp}} + \varepsilon\Delta x\}$$

*que tem  $1 + 2\varepsilon$  elementos;*

2. Encontra-se  $x$  que maximiza  $f(x)$  no intervalo  $V(x_{palp})$ ;
3. Se  $x = x_{palp}$ , a solução do problema é  $x_0 = x_{palp}$ . Caso contrário, faz-se  $x_{palp} = x$  e obtém-se novo conjunto  $V(x_{palp})$ , calculando-se  $f(x)$  nessa vizinhança;
4. Repete-se o passo anterior até se encontrar a solução  $x_0$  ou até  $x_{palp}$  sair fora do intervalo  $[x_{pref\ min}, x_{pref\ max}]$ . Se acontecer a segunda hipótese, coloca-se de novo  $x_{palp}$  com o seu valor inicial e executa-se o passo (5);
5. Agora a busca vai ser efectuada no sentido inverso, calculando-se sequencialmente vizinhanças  $V(x_{palp})$  em torno de valores que vão sendo progressivamente deslocados nessa direcção. Pára-se quando acontecer uma de duas coisas:
  - (a) Se  $x_{palp} = \arg \max f(x)$  no domínio  $V(x_{palp})$ , a solução do problema é  $x_0 = x_{palp}$ ;
  - (b) Se  $x_{palp}$  cair fora do intervalo  $[x_{\min}, x_{\max}]$ , executa-se o passo (6);
6. Inverte-se novamente o sentido da busca. Se este for ascendente, as próximas buscas são feitas ao longo do intervalo  $[x_{pref\ max}, x_{\max}]$ , se for descendente o intervalo é  $[x_{\min}, x_{pref\ min}]$ . Pára-se quando acontecer uma de duas coisas:
  - (a) Se  $x_{palp} = \arg \max f(x)$  no domínio  $V(x_{palp})$ , a solução do problema é  $x_0 = x_{palp}$ ;
  - (b) Se  $x_{palp}$  cair fora do intervalo  $[x_{\min}, x_{\max}]$ , conclui-se que não existe nenhum máximo local de  $f(x)$  que seja superior ao valor da função para cada um dos  $\varepsilon$  argumentos à esquerda e  $\varepsilon$  argumentos à direita de  $x$ . Como tal, segue-se para o passo (7);
7. Utilizando todos os valores de  $f(x)$  já calculados encontra-se o máximo absoluto de  $f(x)$  no domínio  $[x_{\min}, x_{\max}]$ , fazendo-se  $x_0 = \arg \max f(x)$ .

Este algoritmo permite estimar directamente o máximo de qualquer função de uma variável, como as funções  $F_1(\alpha)$  do Algoritmo (4.1) e  $F_2(\alpha)$  do Algoritmo (4.3), que visam estimar o ângulo de compasso. Utilizaram-se nas experiências efectuadas os valores  $\varepsilon = 5$  e  $\Delta x = 1^\circ$ , iniciando-se em ambos os casos o algoritmo com  $x_{palp} = 0^\circ$ .

Quando se pretende maximizar funções de mais que uma variável, como a função  $F_1(\beta, \gamma)$  do Algoritmo (4.3), pode utilizar-se cada uma das variáveis separadamente, deixando variar uma de cada vez enquanto se fixam as outras.

No caso de  $F_1(\beta, \gamma)$ , utilizou-se inicialmente  $\beta_{palp} = 0^\circ$  e, para cada  $\beta$  fixo progredindo de acordo com o Algoritmo (C.1), maximizou-se a função  $f_\beta(\gamma) = F_1(\beta, \gamma)$  de novo através do Algoritmo (C.1) mas agora em função de  $\gamma$ , utilizando-se como  $\gamma_{palp}$  o valor de  $\gamma$  que maximizou  $f_\beta(\gamma)$  para o último  $\beta$  calculado, e  $\gamma_{palp} = 0^\circ$  para o primeiro  $\beta$ .

Os intervalos  $[x_{pref\ min}, x_{pref\ max}]$  utilizados são aqueles que resultam do Teorema (2.13), isto é,  $[-45^\circ, 45^\circ]$  para  $\alpha$  e para  $\beta$  e  $[-54.7^\circ, 54.7^\circ]$  para  $\gamma$ . Por sua vez os intervalos possíveis  $[x_{\min}, x_{\max}]$  são  $[-45^\circ, 45^\circ]$  para  $\alpha$  e  $[-90^\circ, 90^\circ]$  para  $\beta$  e para  $\gamma$ . A

justificação para o uso destes intervalos encontra-se na análise que se fez do Algoritmo (4.3).

Utilizando esta técnica no Algoritmo (4.3), o número total de ciclos nos passos (2) e (4) *varia com a própria orientação* que se pretende estimar, disso dependendo a maior ou menor velocidade de processamento. O caso óptimo ocorre quando  $\alpha = \beta = \gamma = 0^\circ$ , sendo o número total de iterações neste caso dado por  $11 \times 11 + 11 = 132$ , o que corresponde a uma redução de cerca de 99.5% do tempo de processamento quando se faz cada ângulo variar uniformemente no intervalo de valores possíveis. Apesar de não se conseguir em média uma tão grande redução, a utilização desta técnica conduz a tempos de processamento muito aceitáveis, obtendo-se um algoritmo computacionalmente rápido, prático e fácil de utilizar.

As Figuras (C.1) e (C.2) mostram dois exemplos da utilização deste algoritmo em conjunto com o Algoritmo (4.3) para estimar a orientação. No primeiro exemplo, a orientação estimada por este método foi igual ao do método da “força bruta” que se viu na Figura (4.4), mas agora o tempo de processamento foi apenas 91.4 segundos, ou seja, o algoritmo foi cerca de 26 vezes mais rápido. Da observação dos gráficos pode notar-se que, tanto na maximização de  $F_1(\beta, \gamma)$  como na maximização de  $F_2(\alpha)$  o Algoritmo (C.1) terminou no passo (4), que é a situação mais frequente.

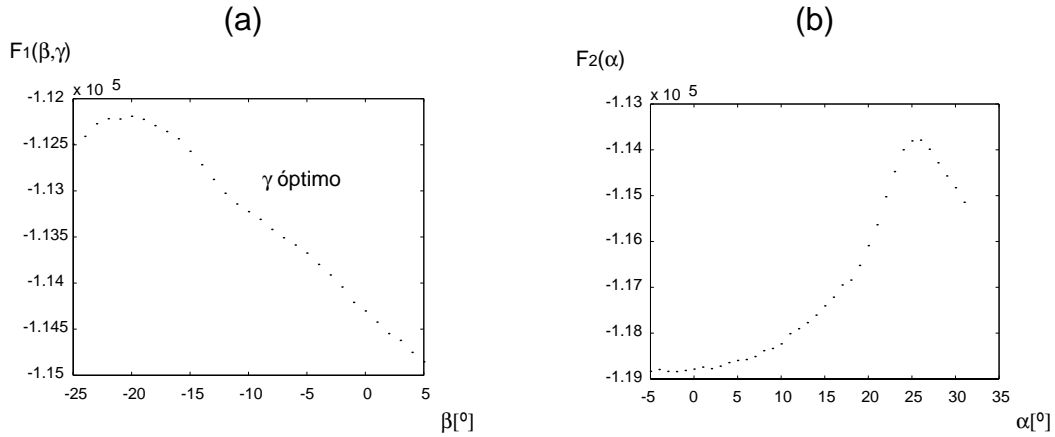


Figura C.1: Gráficos de  $F_1(\beta, \gamma)$  e  $F_2(\alpha)$  obtidos na estimação da orientação em duas etapas utilizando o método do máximo local. A imagem utilizada foi a fotografia da Figura (4.4), obtendo-se idênticos resultados para a orientação. Aqui os máximos locais obtidos foram  $\alpha = 26^\circ$ ,  $\beta = -20^\circ$  e  $\gamma = 1^\circ$ .

Já no exemplo da Figura (C.2) teve de chegar-se até ao passo (5) do mesmo algoritmo para maximizar  $F_1(\beta, \gamma)$ , razão pela qual o tempo de processamento foi desta vez de 165.4 segundos. Uma observação cuidada do gráfico (b) permite compreender o que aconteceu em cada passo.

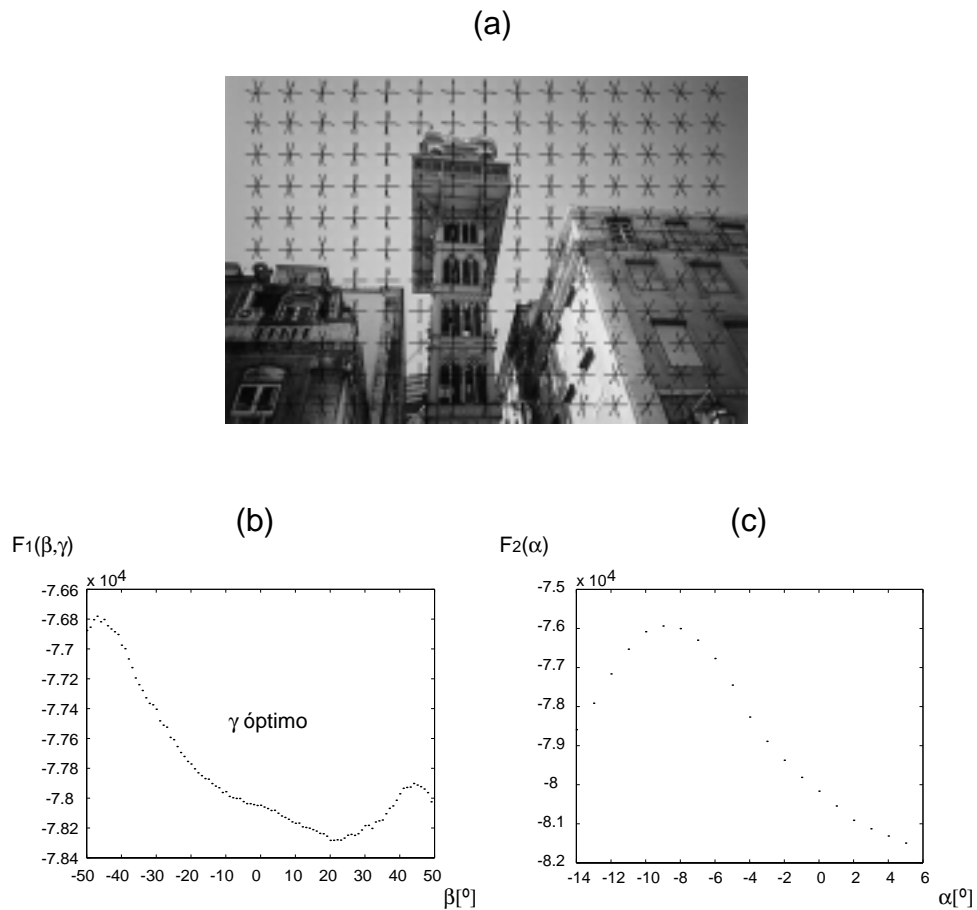


Figura C.2: (a) Fotografia do Elevador de Santa Justa, na Baixa Pombalina, a que foi acrescentada uma grelha que mostra a orientação estimada da câmara. Em (b) mostra-se o andamento da função  $F_1(\beta, \gamma)$  e em (c) o andamento de  $F_2(\alpha)$ . O algoritmo utilizado foi a estimação da orientação em duas etapas utilizando o método do máximo local, tendo sido obtidos os máximos locais  $\alpha = -9^\circ$ ,  $\beta = 44^\circ$  e  $\gamma = -8^\circ$ . No entanto, é mais fácil de perceber na fotografia a orientação equiprojectiva  $\alpha = 9^\circ$ ,  $\beta = -45^\circ$  e  $\gamma = 5^\circ$ , que é também uma solução, como já se suspeitava por observação do gráfico (b).

## C.2 Uso de coordenadas homogéneas para ultrapassar singularidades e reduzir o uso de funções trigonométricas

Os Algoritmos (4.1) e (4.3) fazem um uso alargado de funções trigonométricas, o que computacionalmente pode revelar-se custoso. Por outro lado, algumas singularidades (como quando existe um ponto de fuga ideal) fazem com que seja necessário lidar com excepções na implementação do algoritmo, nomeadamente em cálculos que utilizam as Equações (2.7), (3.2) e (4.4). Verifica-se que em alguns casos se faz uso da função  $\arctan()$  após se ter feito o cálculo da  $\tan()$  de parte do argumento, o que é processamento desnecessário.

Nesta secção vai abordar-se um método simples que evita o uso desnecessário de funções trigonométricas em várias fases do algoritmo. Vimos na Secção (2.3) que os pontos de fuga são dados em coordenadas homogéneas por:

$$\begin{cases} \tilde{\mathbf{F}}_{\mathbf{x}} = (fa_x, fb_x, c_x)^\top = \mathbf{H}_{f,\gamma} (\sin \alpha, \cos \alpha \sin \beta, \cos \alpha \cos \beta)^\top \\ \tilde{\mathbf{F}}_{\mathbf{y}} = (fa_y, fb_y, c_y)^\top = \mathbf{H}_{f,\gamma} (\cos \alpha, -\sin \alpha \sin \beta, -\sin \alpha \cos \beta)^\top \\ \tilde{\mathbf{F}}_{\mathbf{z}} = (fa_z, fb_z, c_z)^\top = \mathbf{H}_{f,\gamma} (0, \cos \beta, -\sin \beta)^\top \end{cases} \quad (\text{C.1})$$

onde

$$\mathbf{H}_{f,\gamma} = \begin{bmatrix} f \cos \gamma & -f \sin \gamma & 0 \\ f \sin \gamma & f \cos \gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Por outro lado, na Secção (3.2) vimos que:

$$\nabla \mathbf{I}(\mathbf{u}) = \nabla \mathbf{I}_x(\mathbf{u}) + j \nabla \mathbf{I}_y(\mathbf{u})$$

Podemos associar a cada elemento da matriz complexa do gradiente um vector homogéneo dado por:

$$\tilde{\mathbf{I}}_{\mathbf{u}} = (\nabla \mathbf{I}_x(\mathbf{u}), \nabla \mathbf{I}_y(\mathbf{u}), \text{mod } \mathbf{I}(\mathbf{u}))^\top$$

que representa a direcção do gradiente.

O ponto  $\mathbf{u}$  é afim e está associado ao vector homogéneo:

$$\tilde{\mathbf{u}} = (u_1, u_2, 1)^\top$$

A direcção da normal do contorno a que pertence  $\mathbf{u}$  obtém-se através do ponto de fuga  $\tilde{\mathbf{F}}(m_{\mathbf{u}}, \Psi) = (F_1, F_2, F_3)^\top$  que está associado ao contorno, e que se pode calcular através da Equação (C.1). Esta direcção pode ser representada também por um vector homogéneo  $\tilde{\mathbf{X}}_{\mathbf{u}}$  tal que:

$$\tilde{\mathbf{X}}_{\mathbf{u}}(m_{\mathbf{u}}, \Psi) = \left( F_1 - u_1 F_3, F_2 - u_2 F_3, \sqrt{(F_1 - u_1 F_3)^2 + (F_2 - u_2 F_3)^2} \right)^\top$$

Debrucemo-nos agora sobre a Equação (4.4). Se fizermos:

$$t_{\mathbf{u}} = \phi_{\mathbf{u}} - \theta(m_{\mathbf{u}}, \Psi, \mathbf{u}) + k\pi$$

temos que:

$$\cos t_{\mathbf{u}} = \mathbf{I}_{\mathbf{u}}^T \mathbf{X}_{\mathbf{u}}(m_{\mathbf{u}}, \Psi) \quad (\text{C.2})$$

onde  $\mathbf{I}_{\mathbf{u}}$  e  $\mathbf{X}_{\mathbf{u}}$  são as coordenadas cartesianas correspondentes a  $\tilde{\mathbf{I}}_{\mathbf{u}}$  e  $\tilde{\mathbf{X}}_{\mathbf{u}}$ .

A função de distribuição  $P_{ang}$ , que definimos na Equação (4.1), é agora equivalente a:

$$P_{ang}(t) = \begin{cases} \frac{1-\epsilon}{2\tau} & \text{se } |\cos t| \geq \cos \tau \\ \frac{\epsilon}{\pi-2\tau} & \text{se } |\cos t| < \cos \tau \end{cases}$$

onde  $\cos t$  é dado pela Equação (C.2).

Com estas considerações, apenas se usam as funções trigonométricas  $\sin()$  e  $\cos()$  que surgem na Equação (C.1), recorrendo-se nos restantes cálculos ao produto interno vectorial expresso na Equação (C.2).

## Apêndice D

# Resultados experimentais para a estimação da orientação

### D.1 Imagens de interiores

As Figuras (D.1) a (D.10) mostram os resultados correctos que se obtiveram para imagens de interiores, onde se verificou uma taxa de sucesso de 91%.

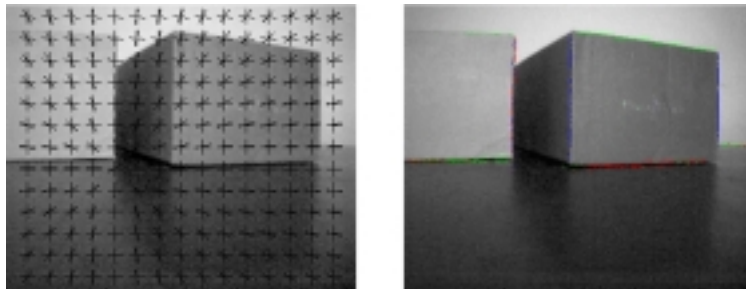


Figura D.1: Estimação da orientação e classificação de contornos para a imagem de uma caixa (Int01). A solução particular encontrada foi  $\alpha = -12^\circ$ ,  $\beta = -4^\circ$  e  $\gamma = -1^\circ$ .





Figura D.2: Estimação da orientação e classificação de contornos para a imagem de uma sala do Pavilhão de Civil do IST (Int02). A solução particular encontrada foi  $\alpha = -41^\circ$ ,  $\beta = 1^\circ$  e  $\gamma = 2^\circ$ .



Figura D.3: Estimação da orientação e classificação de contornos para uma imagem do interior do Pavilhão de Civil do IST (Int03). A solução particular encontrada foi  $\alpha = -33^\circ$ ,  $\beta = -1^\circ$  e  $\gamma = 1^\circ$ .

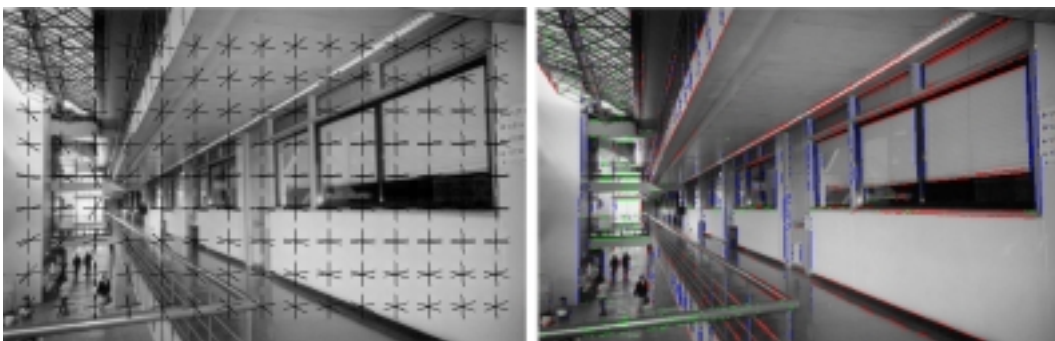


Figura D.4: Estimação da orientação e classificação de contornos para uma imagem do interior do Pavilhão de Civil (Int04). A solução particular encontrada foi  $\alpha = -29^\circ$ ,  $\beta = -3^\circ$  e  $\gamma = 1^\circ$ .

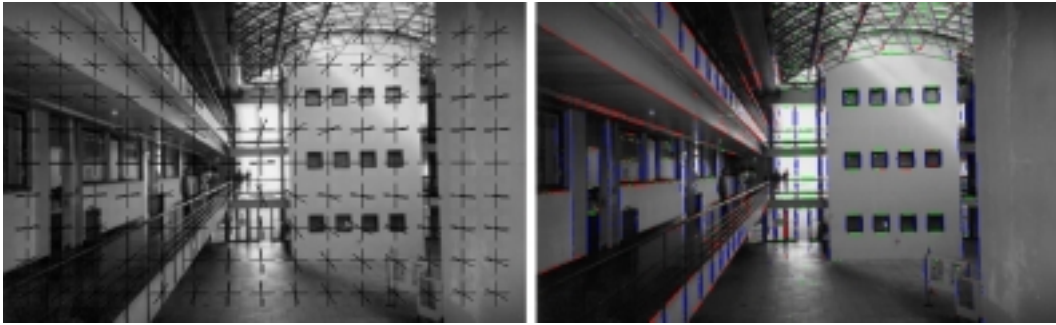


Figura D.5: Estimação da orientação e classificação de contornos para uma imagem do interior do Pavilhão de Civil (Int05). A solução particular encontrada foi  $\alpha = -3^\circ$ ,  $\beta = 0^\circ$  e  $\gamma = 1^\circ$ .



Figura D.6: Estimação da orientação e classificação de contornos para uma imagem do interior de um comboio (Int06). A solução particular encontrada foi  $\alpha = 13^\circ$ ,  $\beta = -5^\circ$  e  $\gamma = 1^\circ$ .

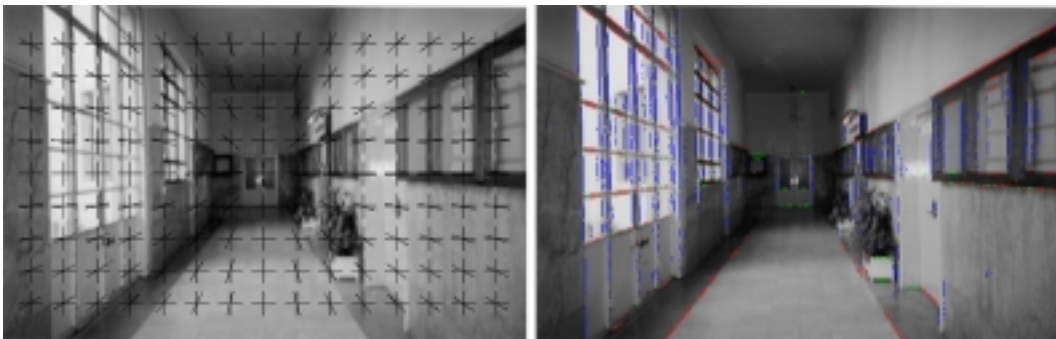


Figura D.7: Estimação da orientação e classificação de contornos para a imagem de um corredor no Pavilhão Central do IST (Int07). A solução particular encontrada foi  $\alpha = 0^\circ$ ,  $\beta = 1^\circ$  e  $\gamma = 1^\circ$ .

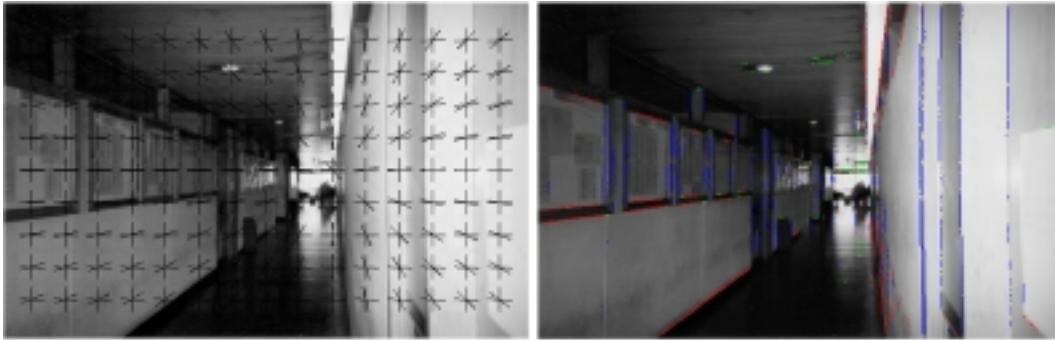


Figura D.8: Estimação da orientação e classificação de contornos para a imagem de um corredor no Pavilhão de Civil do IST (Int08). A solução particular encontrada foi  $\alpha = 14^\circ$ ,  $\beta = -1^\circ$  e  $\gamma = 1^\circ$ .



Figura D.9: Estimação da orientação e classificação de contornos para a imagem de uma estante com livros (Int10). A solução particular encontrada foi  $\alpha = -20^\circ$ ,  $\beta = 0^\circ$  e  $\gamma = 0^\circ$ .



Figura D.10: Estimação da orientação e classificação de contornos para a imagem de um corredor no Pavilhão de Civil do IST (Int11). A solução particular encontrada foi  $\alpha = -27^\circ$ ,  $\beta = 0^\circ$  e  $\gamma = 0^\circ$ .

## D.2 Imagens de exteriores

As Figuras (D.11) a (D.26) mostram os resultados correctos que se obtiveram para imagens de exteriores. Aqui, a taxa de sucesso foi de 84%, um pouco mais baixa, portanto, que a das imagens que vimos anteriormente.



Figura D.11: Estimação da orientação e classificação de contornos para uma imagem da Rua Augusta, na Baixa Pombalina (Ext01). A solução particular encontrada foi  $\alpha = 26^\circ$ ,  $\beta = -20^\circ$  e  $\gamma = 1^\circ$ .



Figura D.12: Estimação da orientação e classificação de contornos para uma imagem da Rua Augusta, na Baixa Pombalina (Ext02). A solução particular encontrada foi  $\alpha = 0^\circ$ ,  $\beta = -1^\circ$  e  $\gamma = 2^\circ$ .





Figura D.13: Estimação da orientação e classificação de contornos para uma imagem da Rua do Ouro, na Baixa Pombalina (Ext03). A solução particular encontrada foi  $\alpha = 23^\circ$ ,  $\beta = 0^\circ$  e  $\gamma = 2^\circ$ .

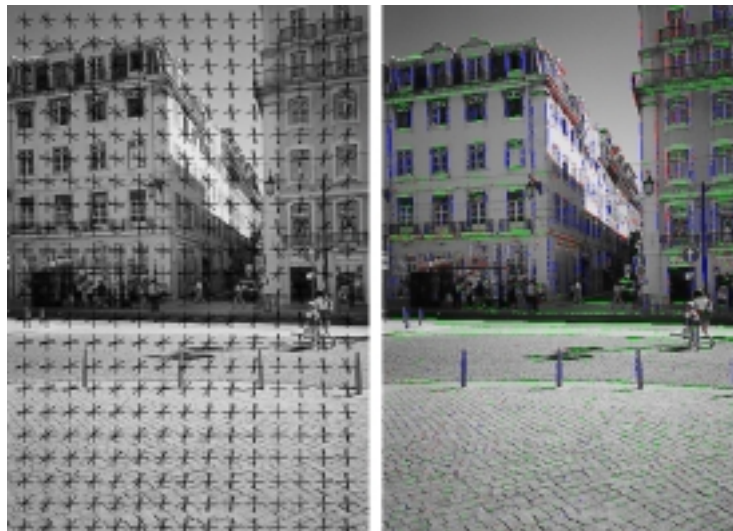


Figura D.14: Estimação da orientação e classificação de contornos para uma imagem da Praça da Figueira, na Baixa Pombalina (Ext05). A solução particular encontrada foi  $\alpha = 11^\circ$ ,  $\beta = 1^\circ$  e  $\gamma = 2^\circ$ .

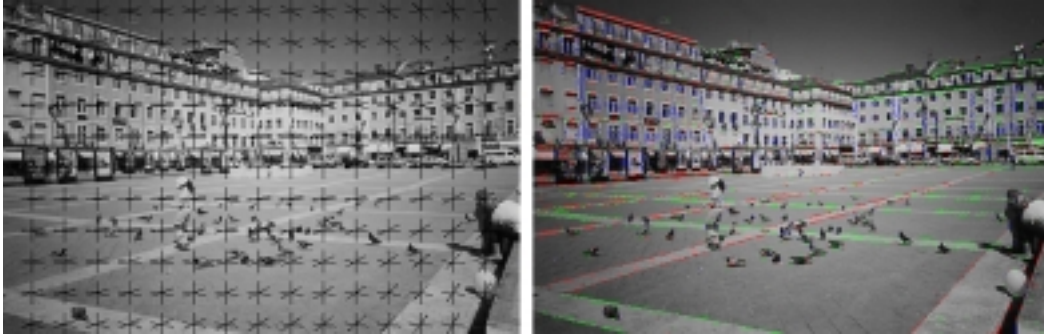


Figura D.15: Estimação da orientação e classificação de contornos para uma imagem da Praça da Figueira, na Baixa Pombalina (Ext06). A solução particular encontrada foi  $\alpha = 31^\circ$ ,  $\beta = 1^\circ$  e  $\gamma = 2^\circ$ .

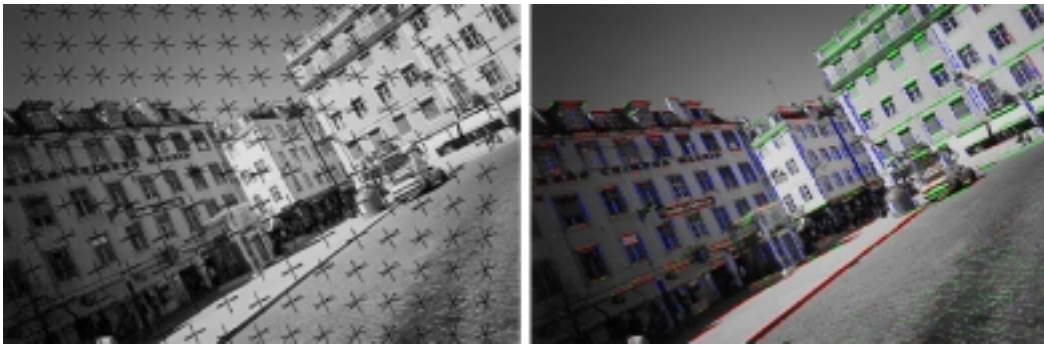


Figura D.16: Estimação da orientação e classificação de contornos para uma imagem da Praça da Figueira, na Baixa Pombalina (Ext07). A solução particular encontrada foi  $\alpha = 29^\circ$ ,  $\beta = -8^\circ$  e  $\gamma = 27^\circ$ .

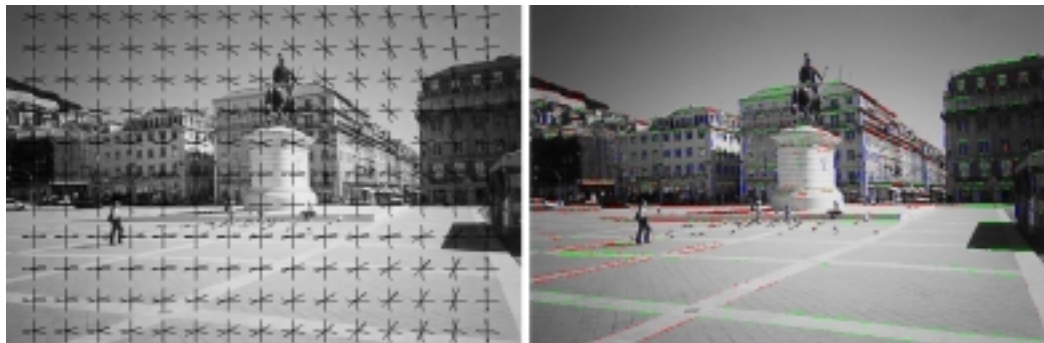


Figura D.17: Estimação da orientação e classificação de contornos para uma imagem da Praça da Figueira, na Baixa Pombalina (Ext08). A solução particular encontrada foi  $\alpha = 25^\circ$ ,  $\beta = -4^\circ$  e  $\gamma = 2^\circ$ .

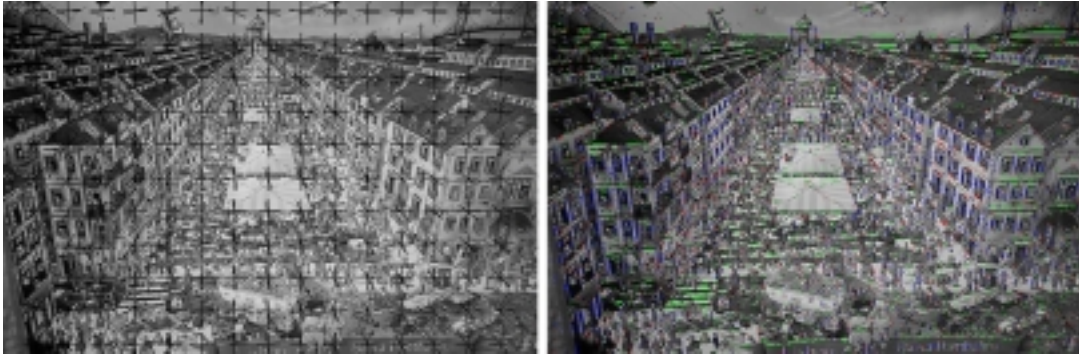


Figura D.18: Estimação da orientação e classificação de contornos para uma imagem desenhada da Baixa Pombalina (Ext09). A solução particular encontrada foi  $\alpha = -7^\circ$ ,  $\beta = 20^\circ$  e  $\gamma = 0^\circ$ .



Figura D.19: Estimação da orientação e classificação de contornos para uma imagem do Rossio, na Baixa Pombalina (Ext10). A solução particular encontrada foi  $\alpha = -2^\circ$ ,  $\beta = 1^\circ$  e  $\gamma = 1^\circ$ .



Figura D.20: Estimação da orientação e classificação de contornos para uma imagem do Rossio, na Baixa Pombalina (Ext11). A solução particular encontrada foi  $\alpha = 18^\circ$ ,  $\beta = -2^\circ$  e  $\gamma = 1^\circ$ .





Figura D.21: Estimação da orientação e classificação de contornos para uma imagem do Rossio, na Baixa Pombalina (Ext12). A solução particular encontrada foi  $\alpha = 5^\circ$ ,  $\beta = -2^\circ$  e  $\gamma = 1^\circ$ .

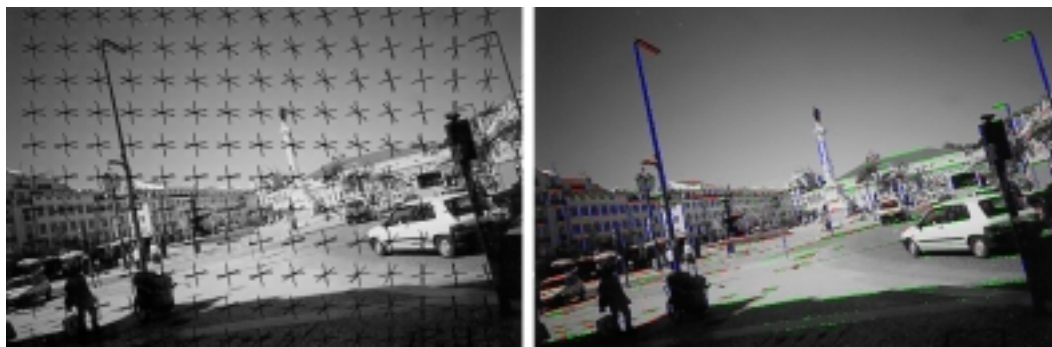


Figura D.22: Estimação da orientação e classificação de contornos para uma imagem do Rossio, na Baixa Pombalina (Ext13). A solução particular encontrada foi  $\alpha = 20^\circ$ ,  $\beta = -5^\circ$  e  $\gamma = 11^\circ$ .

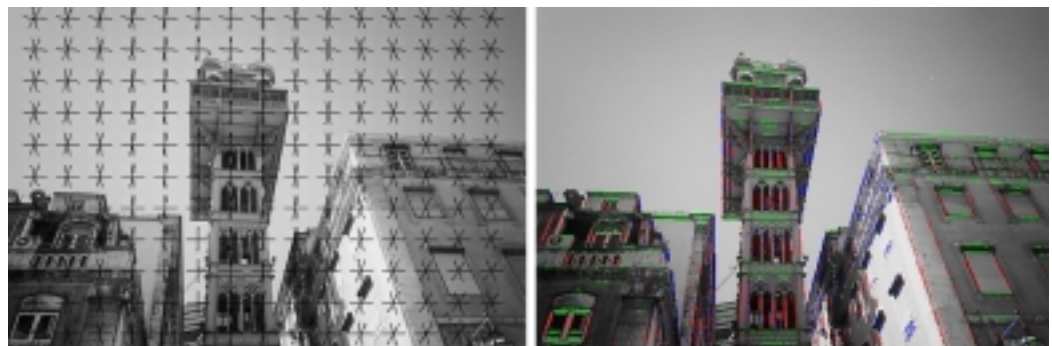


Figura D.23: Estimação da orientação e classificação de contornos para uma imagem da Rua de Santa Justa, na Baixa Pombalina (Ext14). A solução particular encontrada foi  $\alpha = -9^\circ$ ,  $\beta = 44^\circ$  e  $\gamma = -8^\circ$ .



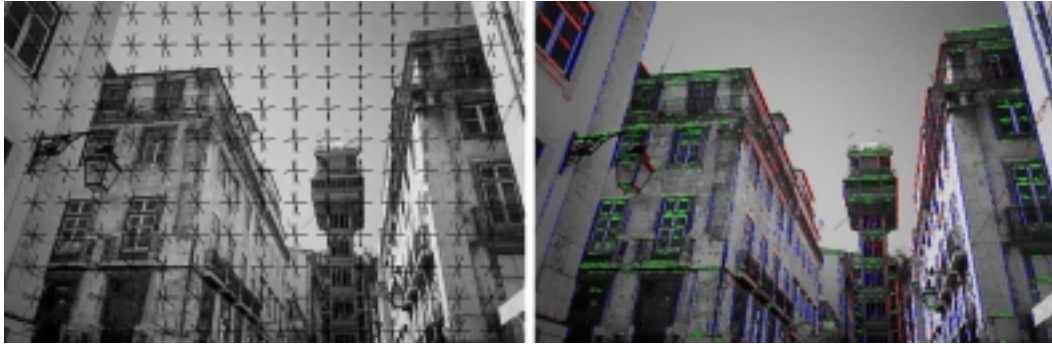


Figura D.24: Estimação da orientação e classificação de contornos para uma imagem da Rua de Santa Justa, na Baixa Pombalina (Ext15). A solução particular encontrada foi  $\alpha = 12^\circ$ ,  $\beta = -31^\circ$  e  $\gamma = -3^\circ$ .

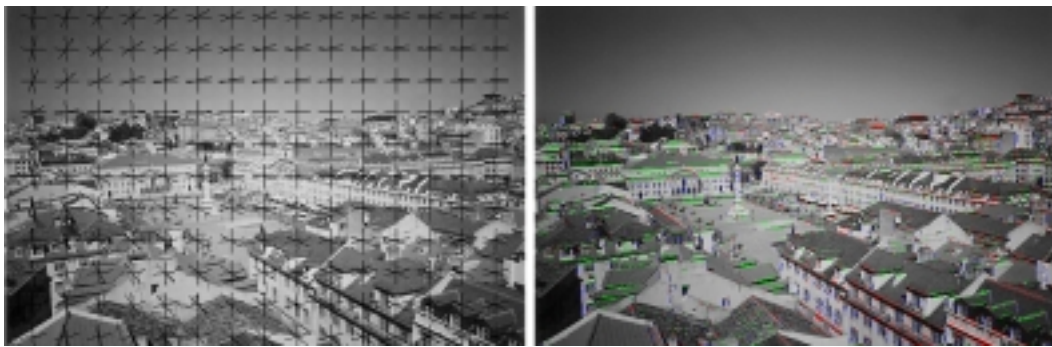


Figura D.25: Estimação da orientação e classificação de contornos para uma vista aérea da Baixa Pombalina, virada para Norte (Ext16). A solução particular encontrada foi  $\alpha = -27^\circ$ ,  $\beta = 7^\circ$  e  $\gamma = 1^\circ$ .



Figura D.26: Estimação da orientação e classificação de contornos para uma imagem aérea da Baixa Pombalina, virada para Sul (Ext19). A solução particular encontrada foi  $\alpha = 25^\circ$ ,  $\beta = 4^\circ$  e  $\gamma = 2^\circ$ .

### D.3 Estimações incorrectas

As Figuras (D.27) e (D.28) mostram os 4 resultados cuja orientação estimada se considera demasiado desviada do valor real. Estes resultados representam 13.3% do total de experiências efectuadas.

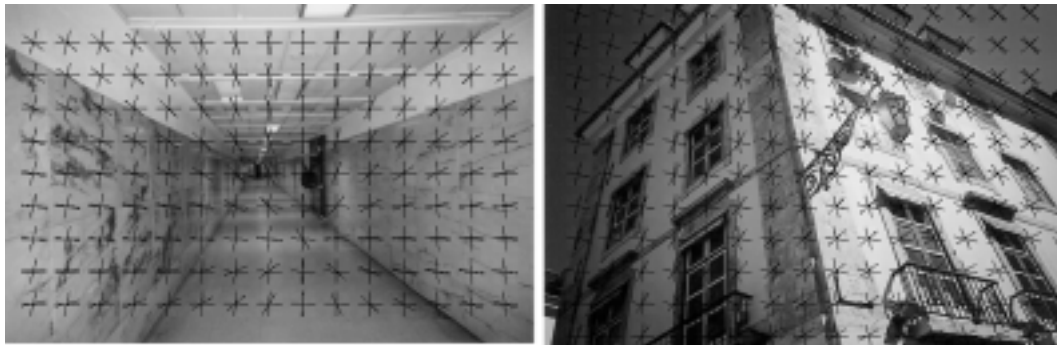


Figura D.27: Estimação da orientação para (a) uma imagem da estação de metro do Saldanha, catalogada como Int09 e (b) a imagem de uma casa da Rua do Carmo, na Baixa Pombalina, catalogada como Ext04. Em ambos os casos a solução particular estimada não se considera satisfatória:  $\alpha = 8^\circ$ ,  $\beta = -13^\circ$  e  $\gamma = -4^\circ$  para a imagem Int09 e  $\alpha = 8^\circ$ ,  $\beta = -44^\circ$  e  $\gamma = -11^\circ$  para a imagem Ext04.

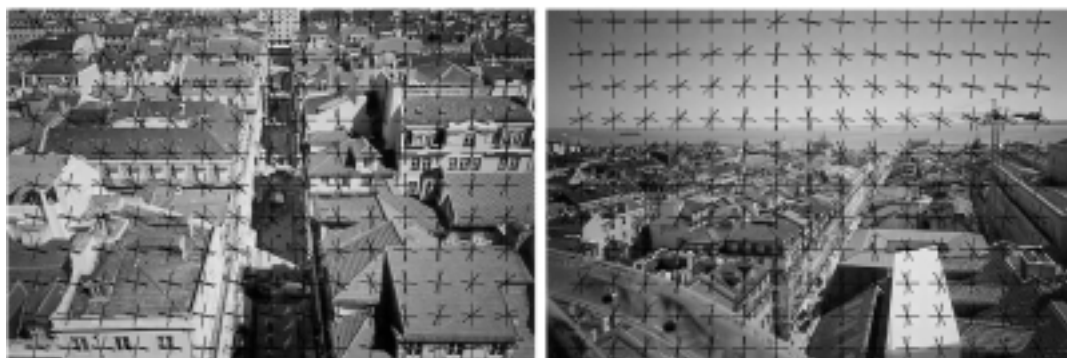


Figura D.28: Estimação da orientação para duas vistas aéreas da Baixa Pombalina, catalogadas como Ext17 e Ext18. Em ambos os casos a solução particular estimada não se considera satisfatória:  $\alpha = 18^\circ$ ,  $\beta = -17^\circ$  e  $\gamma = -8^\circ$  para a imagem Ext17 e  $\alpha = -4^\circ$ ,  $\beta = 19^\circ$  e  $\gamma = -4^\circ$  para a imagem Ext18.

Para além das imagens que aqui se apresentam, algumas outras estimações consideram-se ligeiramente incorrectas, apesar de aceitáveis. A fotografia do comboio que vimos na Figura (D.6) é um desses exemplos, sendo desculpável pelo facto das paredes laterais do comboio não serem verticais, tendo o algoritmo erradamente interpretado uma delas como representativa da direcção do eixo  $\mathbf{z}$ . A Figura (D.29) mostra o gráfico das funções de verosimilhança para este caso específico, sendo evidente, pela presença de

picos gémeos na Figura (D.29a), a "confusão" entre qual das paredes escolher para representar a direcção referida.

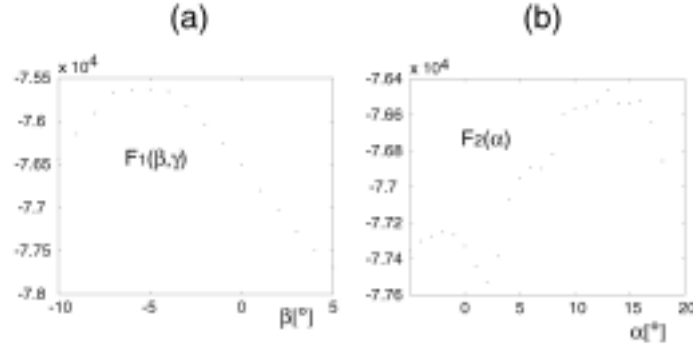


Figura D.29: Funções de verosimilhança na estimação da orientação para a imagem da Figura (D.6).

Mais surpreendente é o resultado negativo na experiência da Figura (D.27a), isto é, com a imagem Int09. Apesar de à primeira vista esta imagem parecer capaz de proporcionar uma boa estimativa, o facto é que não apresenta praticamente nenhuma informação sobre o eixo  $\mathbf{z}$ , uma vez que todos os contornos relacionados com este eixo são fracos (isto é, o gradiente nestes pontos é próximo de zero) e por esse motivo são eliminados ou classificados de acordo com o modelo  $m = 5$ . Assim, neste caso específico, o carácter fortemente “de Manhattan” de nada serve, pois esta informação não passou para a imagem.

A Figura (D.30) mostra a evolução das funções de verosimilhança neste caso, onde não são visíveis picos gémeos na função de verosimilhança de  $\beta$  e  $\gamma$ . O que aqui aconteceu foi que uma das soluções particulares,  $\beta \approx \gamma \approx 0^\circ$ , não originou um máximo da primeira função de verosimilhança, pelo que a segunda solução particular que se deveria obter era  $\beta \approx 90^\circ$  e  $\gamma$  indefinido (resultante de considerar o ponto de fuga  $\mathbf{F}_z$  coincidente com o ponto principal, isto é, o fundo do túnel). Todavia, atingiu-se antes disso um máximo local da função de verosimilhança com  $\beta = 8^\circ$  e  $\gamma = -13^\circ$ , que são estimações erradas. Pode dizer-se que aqui o método de estimar a orientação em duas partes conduziu a resultados errados, o que se deveu à não satisfação da condição necessária para garantir o sucesso deste método, que consiste em *existir informação suficiente sobre cada um dos eixos de Manhattan*.

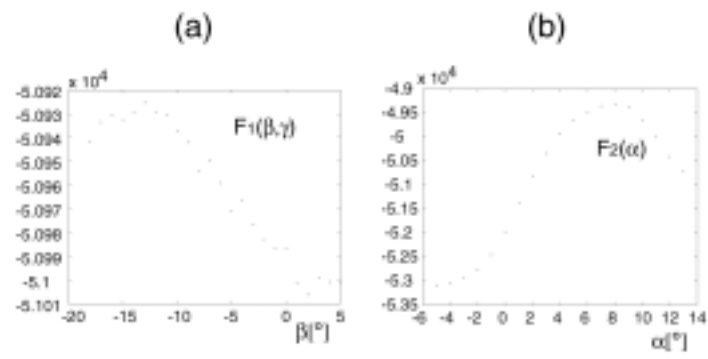


Figura D.30: Funções de verossimilhança na estimação da orientação para a imagem da Figura (D.27).

## Referências

- [1] James M. Coughlan, Alan L. Yuille, “Manhattan World: Compass Direction from a Single Image by Bayesian Inference”, ICCV 1999 - 7th IEEE International Conference On Computer Vision, Kerkyra/Corfu, Greece, September 20-27, 1999
- [2] J.M. Coughlan e A.L. Yuille, “The Manhattan World Assumption: Regularities in scene statistics which enable Bayesian Inference”, NIPS 2000 - Neural Information Processing Systems, Denver, Colorado, November 27-December 2, 2000
- [3] I.Herman, *The Use of Projective Geometry in Computer Graphics*, Lecture Notes in Computer Science, Springer-Verlag, Amesterdam, Netherlands, 1992
- [4] O. Faugeras, *Three-Dimensional Computer Vision: a geometric viewpoint*, MIT Press, Series: Artificial Intelligence, Cambridge/Massachusetts, 1993
- [5] R. Hartley e A. Zissermann, *Multiple View Geometry in Computer Vision*, Cambridge University, UK, 2000
- [6] J.F. Canny, “A computational approach to edge detection”, IEEE Trans. on Pattern Analysis and Machine Intelligence, November 1986.
- [7] S.M. Smith e 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>)
- [8] Artero, A. Olivette e Tommaselli, A.M. Garcia, *Extração de Feições Retas em Imagens Digitais: Adaptação e Avaliação de Diferentes Técnicas*, UNESP Presidente Prudente, Brasil, Fevereiro de 2002 (<http://geodesia.ufsc.br/geodesia-online/Arquivo/2002/02/ta.htm>)
- [9] S. Kunduri, H. Maître, M. Roux, D. Demigny, M. Karabernou, “An Improved Method for Isotropic Edge Orientation Estimation: Application for the Detection of a Roman Cadastral Grid from Multi-Source Images”, École Nationale Supérieure des Télécommunications, TSI, Paris, France.
- [10] T.M.Cover e J.A. Thomas, *Elements of Information Theory*, New York, 1991.
- [11] Jean-Yves Bouguet, *Camera Calibration Toolbox for Matlab* ([http://www.vision.caltech.edu/bouguetj/calib\\_doc](http://www.vision.caltech.edu/bouguetj/calib_doc))
- [12] Freeman, Michael, *The complete guide to photography*, 1991, pp. 32-34.