



UNIVERSIDADE TÉCNICA DE LISBOA  
INSTITUTO SUPERIOR TÉCNICO

## **IMAGENS PANORAMICAS**

Bernardo António F. R. Esteves Pires, nº 47938, AE de Telecomunicações

LICENCIATURA EM ENGENHARIA ELECTROTÉCNICA E DE COMPUTADORES  
Relatório de Trabalho Final de Curso  
05/6/2003/L

Prof. Orientador: Prof. Doutor Pedro M. O. Aguiar

Novembro de 2004



*Rene Magritte*  
*La Condition Humaine*  
*(The Human Nature)*

# Agradecimentos

Gostaria de agradecer em primeiro lugar ao meu orientador, Prof. Pedro Aguiar, pelo seu apoio e disponibilidade e pela sua contribuição constante para a minha aprendizagem.

Para além disto, não posso deixar de expressar a minha gratidão à minha família que me apoiou e ajudou sempre e incondicionalmente, não apenas durante este ano mas em todos os momentos difíceis da minha vida.

Por fim, gostaria ainda de expressar a minha gratidão à Michelle pela sua presença, carinho e apoio constantes.

Muito obrigado a todos. Sem a vossa ajuda não estaria agora a escrever esta página.

# Resumo

O trabalho que aqui se apresenta aborda o problema da construção de uma imagem panorâmica a partir de um conjunto de vistas parciais. Por imagem panorâmica entende-se uma representação da cena global descrita nas vistas parciais. Esta representação é obtida pela junção (alinhamento) de todas as imagens de forma a obter um todo consistente. Trata-se o caso mais geral do problema onde as imagens não se encontram calibradas, isto é, não existe nenhuma informação *a priori* sobre a forma como as imagens se devem juntar para obter o panorama.

O problema é abordado do ponto de vista da teoria de estimação clássica sendo desenvolvida uma estimativa de máxima verosimilhança para o panorama. Esta abordagem permite separar o procedimento em dois passos fundamentais: *Estimação do Movimento*, onde se estima os parâmetros do modelo de movimento que relaciona as imagens, e *Construção do Panorama*, onde se usa as estimativas dos parâmetros obtidas para alinhar as imagens.

Recorrendo à estimação de máxima verosimilhança mostra-se que a *Estimação do Movimento* corresponde à minimização de uma função que mede o erro entre as imagens que se pretende alinhar. Para proceder a esta minimização recorre-se a um algoritmo de multi-resolução que minimiza o erro em cada resolução das imagens com recurso a um algoritmo iterativo.

Para além da abordagem clássica a este problema ser por si só uma contribuição original deste trabalho, a sua aplicação conduz a mais duas contribuições originais. Quando aplicada ao caso do alinhamento de duas imagens permite demonstrar que a estimação óptima do movimento deve recorrer ao uso de uma *janela adaptativa* em vez da tradicional *janela fixa*. Esta contribuição traduz-se numa melhoria substancial na robustez do algoritmo permitindo o alinhamento de imagens com muito pouca sobreposição (área onde, até agora, os métodos *featureless* não conseguiam produzir resultados aceitáveis). Por outro lado quando esta abordagem é aplicada ao alinhamento de múltiplas imagens permite desenvolver um algoritmo global de alinhamento que não só conduz a resultados superiores aos presentes na literatura como permite evitar a propagação de erros sem aumentar drasticamente o peso computacional.

A última contribuição original deste trabalho resulta do estudo dos algoritmos de multi-resolução. Sobre este tema propõe-se um algoritmo de multi-resolução adaptado ao movimento predominante entre as imagens.

A validade de todas as contribuições originais é apoiada por um conjunto de experiências realizadas com recurso ao programa *MatLab*. Para além disso são também apresentados um conjunto de panoramas obtidos com recurso ao conjunto de técnicas descritas.

**Palavras Chave:** Imagem panorâmica, alinhamento de imagens, *featureless mosaicing*.

# Abstract

The work presented here deals with the automatic construction of a panoramic image from a set of partial views. By panoramic image one refers to the global representation of the scene depicted in the partial views. This representation is obtained by registration of all the partial view images. It is assumed that the images are uncalibrated. This means that there is no *a priori* information of how the images should be registered. This is the most generic form of the problem.

The problem is approached using the classical estimation theory and a maximum likelihood estimator for the panoramic image is developed. This approach allows for the problem to be divided in two fundamental steps: *Movement Estimation*, where an estimative for the movement parameters is obtained, and *Panoramic Image Construction*, where the movement parameters estimates are used to register the images.

By using maximum likelihood estimation it is shown that the *Movement Estimation* step is equivalent to the minimization of an function that measures the error between the images. To accomplish this minimization a multi-resolution algorithm is used. In each resolution the error function is minimized by an iterative algorithm.

Besides being, in itself, a new contribution, the classical approach to this problem leads to two other contributions. When applied to the registration of two images it allows to demonstrate that the optimal estimation of the movement should use an *adaptive window* instead of the traditional *fixed window*. This translates in a substantial improvement in the performance of the algorithm allowing for the registration of images with small overlap (area in which the featureless methods had, until now, failed to produce results). On the other hand when applied to the registration of multiple images this approach allows the development of a new global registration algorithm witch is able to avoid error propagation without drastically increasing computational time.

The last original contribution of this work is a result of the study of the multi-resolution algorithm. On this theme we propose a new multi-resolution algorithm adapted to the movement between the images.

All original contributions have been validated by a set of experiences performed by MatLab. Several panoramic images obtained by combining the techniques above are also shown.

**Keywords:** Panoramic image, image registration, featureless mosaicing.

# Conteúdo

<b>Agradecimentos</b>	<b>ii</b>
<b>Resumo</b>	<b>iii</b>
<b>Abstract</b>	<b>iv</b>
<b>índice de Conteúdos</b>	<b>iv</b>
<b>Lista de Figuras</b>	<b>vi</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Motivação e Aplicações . . . . .	1
1.2 Abordagens Correntes e suas Limitações . . . . .	3
1.2.1 Alinhamento de um Par de Imagens . . . . .	3
1.2.2 Alinhamento de Múltiplas Imagens . . . . .	6
1.3 Abordagem Proposta . . . . .	7
1.4 Contribuições Originais . . . . .	8
1.5 Organização do Relatório . . . . .	8
<b>2 Alinhamento de um Par de Imagens: Janela Adaptativa</b>	<b>10</b>
2.1 Alinhamento como Problema de Estimação . . . . .	10
2.1.1 Notação . . . . .	10
2.1.2 Formulação do Problema . . . . .	11
2.1.3 Estimativa de Máxima Verossimilhança . . . . .	12
2.2 Impacto da Janela na Função de Erro . . . . .	14
2.3 Estimação do Movimento entre um Par de Imagens . . . . .	16
2.3.1 Minimização da Função de Erro para Resolução Fixa . . . . .	16
2.3.2 Modelo de Movimento de Translação . . . . .	17
2.3.3 Modelo de Movimento Afim . . . . .	18
2.3.4 Algoritmo de Multi-Resolução . . . . .	20
2.4 Resultados Experimentais . . . . .	20
2.4.1 Comparação entre a Janela Fixa e a Janela Adaptativa . . . . .	20
2.4.2 Alinhamento de Imagens com Pouca Sobreposição . . . . .	22
2.4.3 Alinhamento de Imagens Submarinas . . . . .	23

<b>3</b>	<b>Algoritmos de Multi-Resolução Adaptados ao Modelo de Movimento</b>	<b>25</b>
3.1	Introdução e Motivação . . . . .	25
3.2	Forma Geral dos Algoritmos de Multi-resolução . . . . .	26
3.3	Técnica Tradicional de Construção da Pirâmide . . . . .	26
3.4	Técnica Proposta para a Construção da Pirâmide . . . . .	28
3.5	Resultados Experimentais . . . . .	31
3.5.1	Função de erro . . . . .	31
3.5.2	Alinhamento de duas imagens . . . . .	31
<b>4</b>	<b>Alinhamento de Múltiplas imagens: Estimação Global</b>	<b>34</b>
4.1	Alinhamento de Múltiplas Imagens como Problema de Estimação . . . . .	34
4.1.1	Notação . . . . .	34
4.1.2	Formulação do Problema . . . . .	35
4.1.3	Estimativa de Máxima Verosimilhança . . . . .	36
4.2	Estimação do Movimento entre Múltiplas Imagens . . . . .	40
4.2.1	Minimização da Função de Erro para Resolução Fixa . . . . .	40
4.2.2	Interpretação do Resultado: Algoritmo Global de Estimação . . . . .	43
4.2.3	Algoritmo de Multi-Resolução . . . . .	43
4.3	Resultados Experimentais . . . . .	44
<b>5</b>	<b>Resultados Experimentais</b>	<b>46</b>
<b>6</b>	<b>Conclusão</b>	<b>48</b>
6.1	Sumário . . . . .	48
6.2	Trabalho Futuro . . . . .	48
<b>A</b>	<b>Modelos do Movimento da Câmara</b>	<b>50</b>
A.1	Modelo de Movimento de Translação . . . . .	50
A.2	Modelo de Movimento Afim . . . . .	50
A.3	Modelo de Movimento de Rotação . . . . .	51
<b>B</b>	<b>Dedução da equação (2.16) a partir das equações (2.15) e (2.12)</b>	<b>52</b>
<b>C</b>	<b>Dedução da equação (4.19) a partir das equações (4.18) e (4.15)</b>	<b>54</b>
<b>D</b>	<b>"Registration of Images with Small Overlap"</b>	<b>57</b>
<b>E</b>	<b>"Featureless Global Alignment of Multiple Images"</b>	<b>58</b>

# Lista de Figuras

1.1	Aplicação do problema em estudo: Funcionamento da <i>Wide Field and Planetary Camara 2</i> presente no telescópio Hubble . . . . .	2
1.2	Aplicação do problema em estudo: Construção do panorama do fundo de uma sequência de vídeo a partir de uma série de vistas parciais (para aplicação em compressão de vídeo digital) . . . . .	2
1.3	Par de imagens submarinas onde a detecção de pontos notáveis é particularmente difícil . . . . .	4
2.1	Erro $E(\theta)$ para vários tamanhos da janela $\mathcal{R}$ . . . . .	15
2.2	Comparação do desempenho da janela fixa e da janela adaptativa no alinhamento de imagens com grande sobreposição . . . . .	21
2.3	Comparação do desempenho da janela fixa e da janela adaptativa no alinhamento de imagens com pequena sobreposição . . . . .	21
2.4	Alinhamento de vários pares de imagens não calibradas com pequena sobreposição. (continua na pág. seguinte). . . . .	22
2.5	Alinhamento de vários pares de imagens não calibradas com pequena sobreposição. (continuando a pág. anterior). . . . .	23
2.6	Alinhamento de três pares de imagens submarinas não calibradas . . . . .	24
3.1	Alinhamento de par de imagens não calibradas com pequena sobreposição. Ilustração do funcionamento do método de multi-resolução . . . . .	27
3.2	Par de imagens relacionadas por um movimento de rotação . . . . .	29
3.3	Comparação das funções de erro para a pirâmide tradicional e para a pirâmide proposta . . . . .	32
3.4	Comparação entre o alinhamento de duas imagens usando a pirâmide tradicional e usando a pirâmide proposta . . . . .	33
4.1	Conjunto de sete imagens que se pretende alinhar utilizando o Algoritmo de Alinhamento Global . . . . .	44
4.2	Primeiro Passo do Algoritmo Global: Construção cumulativa de um panorama com sete imagens . . . . .	44
4.3	Resultado Final do Algoritmo Global: Construção correcta do panorama com sete imagens . . . . .	45
5.1	Panorama resultante do alinhamento de quatro imagens . . . . .	46
5.2	Panorama resultante do alinhamento de quatro imagens . . . . .	46



5.3	Panorama resultante do alinhamento de nove imagens . . . . .	47
5.4	Panorama resultante do alinhamento de cinco imagens . . . . .	47

# Capítulo 1

## Introdução

O trabalho que aqui se apresenta aborda o problema da construção automática de uma imagem panorâmica a partir de um conjunto de vistas parciais. Tratamos o caso mais geral em que as vistas parciais correspondem a imagens não calibradas, ou seja, a situação onde não existe informação *a priori* sobre o modo como as imagens se devem alinhar para formar o panorama.

### 1.1 Motivação e Aplicações

O problema da construção automática de panoramas encontra aplicações em áreas tão diversas como a detecção remota [1], compressão e estabilização vídeo digital [2, 3, 4], realidade virtual ou medicina. Em seguida apresentamos, a título ilustrativo, duas aplicações específicas nas áreas da detecção remota e do vídeo digital.

Em detecção remota o problema da construção de panoramas surge devido à necessidade de alinhar imagens provenientes de uma ou várias câmaras<sup>1</sup> no mesmo instante temporal ou em instantes temporais diferentes para obter uma imagem que apresente um campo de visão maior. Como exemplo observe-se a figura Fig. 1.1 onde se mostra a forma como uma das câmaras presentes no telescópio Hubble (neste caso a câmara *Wide Field and Planetary Camera 2*) regista imagens do espaço distante [1]. Cada imagem desta câmara é obtida pelo alinhamento automático por *software* de quatro imagens parcialmente sobrepostas provenientes de sensores diferentes. Note-se, ainda, que uma destas imagens apresenta uma resolução superior às restantes (trata-se da imagem de tamanho menor).<sup>2</sup>

Na área do vídeo digital a construção automática de panoramas permite um grande aumento de taxas de compressão em certas sequências de vídeo. O aumento da taxa de compressão consegue-se alinhando conjuntos de imagens que representam vistas do mesmo panorama. Estes conjuntos (relativamente grandes) de imagens são substituídos por uma única imagem do panorama que representam e pelo conjunto de (poucos) parâmetros da câmara necessários para obter, a cada instante, a imagem de vídeo a partir da imagem do panorama. Como exemplo de aplicação, observe-se a Fig. 1.2 que ilustra a construção de um panorama numa

---

<sup>1</sup>Nesta área habitualmente usa-se o termo sensor uma vez que em muitos casos as imagens a alinhar usam o espectro electromagnético para além do visível.

<sup>2</sup>é importante referir que esta situação corresponde a um caso particular do estudado neste trabalho uma vez que existe informação sobre a forma como as imagens obtidas pelos sensores se relacionam para formar o panorama.

situação real. Note-se que esta situação apresenta uma complexidade adicional uma vez que para além do panorama (previsão meteorológica), existe um "objecto" em movimento no plano da imagem (meteorologista).

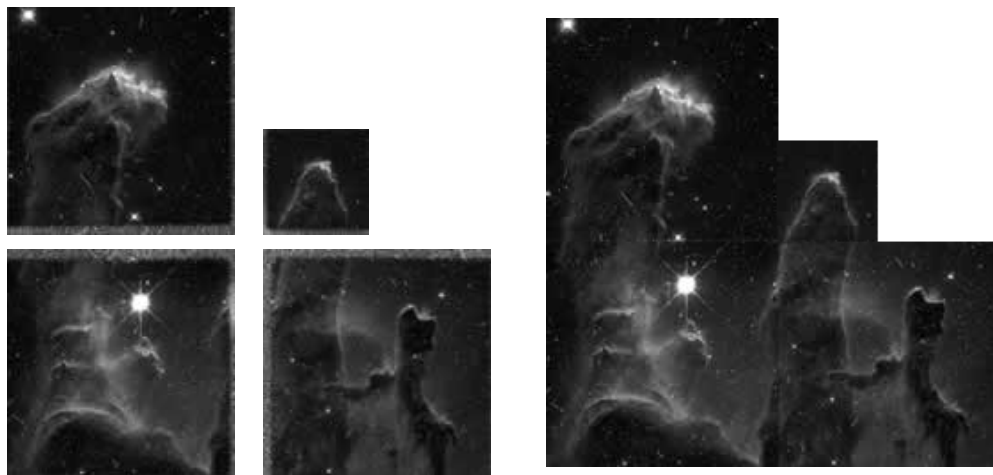


Figura 1.1: Funcionamento da *Wide Field and Planetary Camara 2* presente no telescópio Hubble. À esquerda encontram-se as imagens recolhidas pelos quatro sensores presentes no telescópio; o resultado após alinhamento automático é apresentado na imagem da direita. Imagens retiradas de [1].

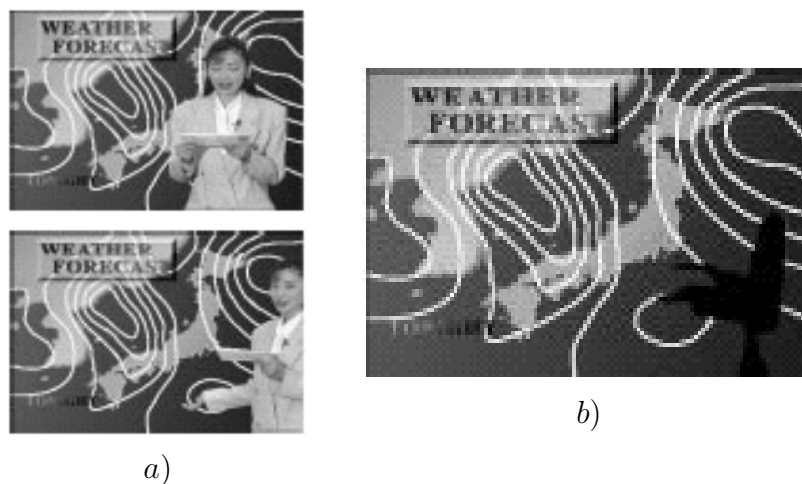


Figura 1.2: Construção do panorama do fundo de uma sequência de vídeo a partir de uma série de vistas parciais (para aplicação em compressão de vídeo digital) a.) Duas imagens da sequência de vídeo. b.) Estimativa do fundo após análise de toda a sequência de vídeo (composta por 100 imagens). Imagens retiradas de [2].

## 1.2 Abordagens Correntes e suas Limitações

Devido à grande complexidade do problema da construção de panoramas a partir de várias imagens, as abordagens correntes consideram em cada instante, unicamente, o alinhamento de um par de imagens (problema que concentra grande parte do trabalho de investigação desenvolvido nesta área).

Para construir um panorama a partir de mais do que duas imagens, existem várias abordagens na literatura que combinam operações de alinhamento de pares de imagens para obter o panorama global.

### 1.2.1 Alinhamento de um Par de Imagens

A generalidade das abordagens ao problema de construção de panoramas assume que o panorama se manteve inalterado enquanto a câmara se move. Este movimento tridimensional (3D) da câmara induz um movimento bidimensional (2D) do padrão de intensidades luminosas no plano da imagem.

Assim, o problema do alinhamento de duas imagens tem sido dividido em duas fases: i) Estimação do Movimento no plano da imagem; e ii) Construção do Panorama.

Nos casos em que faz sentido construir um panorama, i.e., quando a cena é planar ou quando o movimento da câmara se restringe a uma rotação 3D, o movimento 2D no plano da imagem é descrito por uma transformação de coordenadas que depende de um pequeno número de parâmetros<sup>3</sup>. Assim, na primeira fase – Estimação do Movimento – pretende-se estimar os parâmetros da transformação de coordenadas correspondente ao modelo de movimento usado. A segunda fase – Construção do Panorama – é trivial, consistindo apenas de uma média das imagens depois de sujeitas a uma transformação de coordenadas que as coloque num referencial comum.

A dificuldade de resolução do problema encontra-se, então, na fase de estimação dos parâmetros da transformação de coordenadas que relaciona as imagens. É possível agrupar os métodos que resolvem este problema em duas abordagens conforme se baseiem no uso de pontos notáveis, ou o façam directamente a partir dos níveis de intensidade das imagens<sup>4</sup>.

#### Usando pontos notáveis

Um ponto notável (na literatura anglo-saxónica, *feature point*) é uma região da imagem (i.e. um conjunto de *pixels*) que apresenta características singulares que o diferenciam claramente do resto da imagem. As características diferenciadoras podem ser, por exemplo, mudanças bruscas da intensidade luminosa ou configurações de intensidades específicas que não sejam facilmente reproduzíveis.

Na abordagem com recurso a pontos notáveis (na literatura anglo-saxónica *feature-based*) a estimação do movimento é feita em três fases. Em primeiro lugar faz-se a detecção em cada uma das imagens de pontos notáveis segundo um dado critério. Na segunda fase faz-se um emparelhamento dos pontos notáveis presentes nas várias imagens. Por fim, na terceira

---

<sup>3</sup>Um estudo detalhado das situações em que é possível construir um panorama é apresentado no Anexo A.

<sup>4</sup>Para além dos métodos apresentados existem ainda na literatura outras técnicas de estimação do movimento que incluem o uso de Transformadas de Fourier [5, 4] ou de *Integral Projections* [6]

fase estimam-se os parâmetros da transformação de coordenadas a partir da localização dos pontos notáveis.

Assumindo que a detecção e emparelhamento de pontos notáveis foi realizada correctamente, a fase de estimação dos parâmetros da transformação é simples, resumindo-se apenas à resolução de um sistema linear ou de um problema de mínimos quadrados conforme o dobro do número de pontos notáveis disponíveis é igual ou superior ao número de parâmetros que descreve o movimento global do padrão de intensidades das imagens.<sup>5</sup> A simplicidade da fase de estimação dos parâmetros do movimento a partir da correspondência de pontos notáveis faz com que ela tenha sido usada em vários Trabalhos Finais de Curso do I.S.T., como por exemplo [8, 9], e mesmo em trabalhos de projectos de cadeiras como Processamento de Imagem e Visão.

Naturalmente, os problemas desta abordagem prendem-se, precisamente, com as fases de detecção e emparelhamento de pontos notáveis. Na fase de detecção as dificuldades aparecem logo à partida porque não é possível garantir a existência de pontos notáveis claramente identificáveis em todas as situações de interesse. Como exemplo desta dificuldade observe-se a Fig. 1.3 que mostra um par de imagens submarinas onde a detecção de pontos notáveis é particularmente difícil devido à baixa textura e ausência de saliências.

Na fase de emparelhamento, por outro lado, as dificuldades aparecem porque os pontos notáveis encontrados numa imagem podem não ser facilmente identificados noutra devido a efeitos de ruído, alteração de iluminação, oclusão por objectos que se movimentam no plano da imagem ou mesmo porque saíram do campo de visão. Voltando à Fig. 1.3 observe-se que o conjunto de pontos notáveis assinalado na imagem da esquerda não se encontra na imagem da direita porque sai do campo de visão.

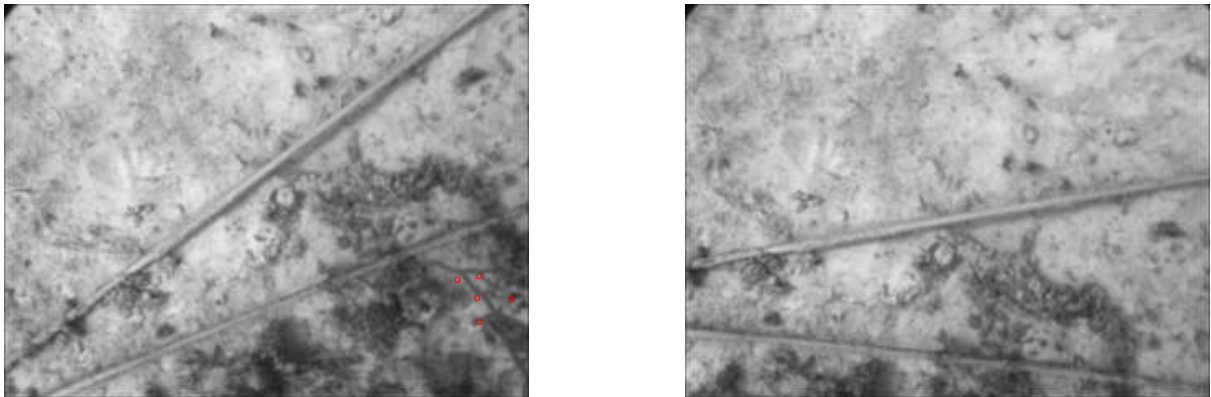


Figura 1.3: Par de imagens submarinas onde a detecção de pontos notáveis é particularmente difícil. Note-se que os pontos notáveis assinalados na imagem da esquerda não têm correspondência com nenhum ponto notável na imagem da direita uma vez que se encontram fora do plano de visão.

---

<sup>5</sup>Note-se que uma correspondência entre dois pontos no plano da imagem restringe a transformação de coordenadas em dois graus de liberdade sendo que o número de graus de liberdade da transformação de coordenadas é igual ao seu número de parâmetros. Note-se, ainda, que esta desigualdade só é válida caso os pontos em questão não sejam *degenerados*. Para mais detalhes consultar [7], págs. 69 a 76

A combinação das dificuldades referidas leva a que seja, por vezes, necessária assistência humana tanto na fase de detecção como na fase de correspondência de forma a garantir um alinhamento correcto entre as imagens. Para tentar obviar estas dificuldades têm sido desenvolvidos, recentemente, métodos para tornar a detecção e emparelhamento de pontos notáveis um processo robusto e automático. No entanto estes métodos usam algoritmos de optimização combinatória que, em geral, resultam computacionalmente pesados. Como exemplo veja-se o método desenvolvido em [10].

### Directamente a partir das imagens

Para tentar obviar as dificuldades acima referidas outros investigadores têm proposto abordagens onde a estimação dos parâmetros do modelo do movimento é feita directamente a partir das imagens, i.e., sem recorrer a pontos notáveis (na literatura anglo-saxónica, *featureless*). Nesta abordagem, a estimativa dos parâmetros do movimento é encontrada por minimização de uma função que mede o erro entre imagens numa janela fixa onde se assume que existe sobreposição [11].

Para tornar a procura do mínimo da função de erro um problema solúvel em tempo útil, esta abordagem recorre, no caso geral, a métodos de minimização baseados em gradiente. Este facto leva a que exista uma possibilidade, presente em qualquer método de minimização por gradiente, de convergência para um mínimo local e não global. Daqui resulta a importância de uma boa inicialização do algoritmo de procura de modo a que o mínimo global se encontre próximo do ponto de inicialização. Ora, nos casos em estudo, não existe nenhuma informação que permita uma inicialização particular uma vez que, tal como já foi referido antes, se trata de imagens não calibradas. Assim, é prática comum usar como inicialização a situação de ausência de movimento o que implica que, para que se obtenham bons resultados, o mínimo global tem que estar próximo desta situação, o que quer dizer que o movimento da câmara entre imagens tem de ser pequeno.

Para lidar com esta limitação, os métodos presentes na literatura recorrem a pirâmides de multi-resolução [12]. Neste processo os parâmetros começam por ser estimados para versões de baixa resolução das imagens (no topo da pirâmide). As estimativas obtidas são refinadas usando imagens de resolução progressivamente superior à medida que o algoritmo desce em direcção à base da pirâmide onde se encontram as imagens originais (de resolução máxima). De um ponto de vista intuitivo<sup>6</sup> a utilização de pirâmides de multi-resolução resolve a dificuldade da estimação de grandes movimentos porque a diminuição da resolução faz com que o movimento entre imagens seja menor e, consequentemente, mais fácil de estimar. Note-se, no entanto, que esta diminuição do movimento é apenas absolutamente válida para parâmetros associados à translação pura entre imagens. Para o parâmetro associado à rotação entre imagens, por exemplo, o uso das pirâmides multi-resolução apresentadas não tem nenhum impacto real uma vez que a amplitude de rotação não é alterada pela diminuição de resolução (uma rotação de  $90^\circ$  é uma rotação de  $90^\circ$  quer em imagens de alta resolução quer em imagens de baixa resolução). Este exemplo simples mostra como a eficiência de pirâmides multi-resolução está fortemente limitada a situações onde o movimento principal entre as

---

<sup>6</sup>A motivação para o uso de pirâmides multi-resolução prende-se ainda com argumentos detalhados no Capítulo 3.

imagens é o de simples translação.

Independentemente da forma como a função de erro é minimizada, os métodos de estimação de movimento directamente a partir das imagens, sofrem de uma grave limitação de desempenho pelo facto de usarem uma janela fixa dado que obrigam a que exista sobreposição entre as imagens a alinhar numa região maior ou igual à janela usada. Uma janela maior conduz a melhores estimativas dos parâmetros de movimento mas obriga a que haja uma grande sobreposição das imagens. Por outro lado, uma janela menor permite que o algoritmo resolva situações onde há menor sobreposição entre as imagens mas conduz a erros de estimação muito superiores. Assim, na prática, o uso da janela fixa limita a aplicação do algoritmo a situações onde existe uma grande sobreposição entre as imagens.

## 1.2.2 Alinhamento de Múltiplas Imagens

A maioria das abordagens correntes parte de uma imagem de base e acrescenta, a cada passo, uma nova imagem ao panorama obtendo o panorama global como o resultado do alinhamento cumulativo de todas as imagens. Apesar de simplificar muito o alinhamento de múltiplas imagens, esta abordagem apresenta um grave problema devido à propagação de erros. Uma vez que a construção é cumulativa, o erro de alinhamento de uma única imagem pode ser propagado para todas as seguintes.

Para tentar obviar este problema têm sido propostos alguns métodos sub-óptimos para a construção do panorama global. Por exemplo, no método desenvolvido por Mann e Pickard em [11], começa-se por construir duas matrizes: a primeira diz-se matriz de suporte e contém as estimativas da sobreposição entre cada par de imagens disponível; enquanto a segunda diz-se matriz de erro e contém as estimativas do erro quadrático médio de alinhamento entre cada par de imagens. A partir destas duas matrizes as imagens são agrupadas num pequeno número de sub-panoramas onde a quantidade de sobreposição entre as imagens é máxima e, simultaneamente, onde o erro entre imagens é mínimo. Por fim, o panorama global é construído pelo alinhamento dos sub-panoramas obtidos. Este novo método reduz o problema da propagação de erros mas envolve um número muito grande de operações de alinhamento de pares de imagens para construir as matrizes de suporte e de erro. Este facto leva a que esta abordagem seja computacionalmente pesada e limita a sua implementação prática a casos onde o número de imagens a alinhar é pequeno.<sup>7</sup>

Existe outro método proposto na literatura, também para tentar obviar o problema da propagação de erros, mas agora apenas aplicável no contexto do alinhamento de imagens utilizando pontos notáveis. Este método baseia-se numa generalização dos métodos de alinhamento de duas imagens com recurso a pontos notáveis propondo a detecção e emparelhamento destes pontos não apenas num par de imagens mas em todo o conjunto de vistas que se pretende alinhar. De facto, se a detecção e o emparelhamento dos pontos notáveis forem feitos de uma forma correcta a estimação do conjunto de parâmetros que relacionam as coordenadas de todas as imagens e a construção do panorama global são triviais. No entanto, tal como já foi referido antes para o emparelhamento de duas imagens usando pontos notá-

---

<sup>7</sup>É importante referir que, embora este método tenha sido desenvolvido por Mann e Pickard no contexto do alinhamento directamente a partir das imagens ele pode ser aplicado sem alterações à situação onde se recorre a pontos notáveis.

veis, também aqui as fases de detecção e emparelhamento são problemáticas. Para manter o método automático os autores propõem, tal como no caso de duas imagens, o uso de métodos de optimização combinatoria. No entanto, mais uma vez, o método resultante é demasiado pesado computacionalmente para ser aplicado em casos onde se pretende alinhar um grande número de imagens [10].

## 1.3 Abordagem Proposta

Neste trabalho optamos por partir de princípios fundamentais formulando o problema no contexto da teoria clássica de estimação. Neste contexto, as imagens são vistas como observações ruidosas de porções distintas do panorama que se pretende estimar. O problema é formulado como a estimação conjunta de todas as variáveis desconhecidas: o panorama e os parâmetros do modelo de movimento que relaciona as imagens.

Mostramos que a estimativa de máxima verosimilhança para o panorama pode ser obtida em forma fechada em função das estimativas dos parâmetros do modelo de movimento. Usando este resultado concluímos que as estimativas de máxima verosimilhança dos parâmetros do modelo de movimento são aquelas que minimizam uma função que mede o erro entre as imagens. Esta função de erro consiste no somatório do quadrado da diferença entre as imagens nos *pixels* onde estas se intersectam.

A função a minimizar obtida desta forma é semelhante à proposta pelos métodos presentes na literatura. No entanto, estes métodos propõem que esta função meça o erro entre imagens numa janela de tamanho fixo. Como se referiu na secção 1.2.1, o uso deste tipo de janelas limita gravemente o domínio de aplicabilidade deste algoritmo uma vez que obriga à existência de uma grande sobreposição entre as imagens a alinhar. Para ultrapassar esta limitação propomos o uso de uma janela adaptativa dada pela maior região onde as imagens se intersectam. Uma vez que não é necessário definir à partida uma janela fixa, o uso da janela adaptativa permite fazer o alinhamento das imagens de uma forma óptima, independentemente da área de intersecção entre elas.

A minimização da função de erro é feita de uma forma iterativa. Em cada iteração construímos uma aproximação de primeira ordem do erro entre imagens em torno da estimativa actual dos parâmetros. A partir desta aproximação minimizamos a função de erro por anulamento das suas derivadas parciais. Este passo corresponde à resolução de um sistema de equações linear. Usamos então o resultado como nova estimativa dos parâmetros repetindo o processo até que se atinja um número limite de iterações ou que a alteração da estimativa seja inferior a um certo limiar. Este procedimento de estimação conduz a alterações constantes dos parâmetros do modelo de movimento. Por outro lado estes parâmetros controlam a posição relativa das imagens pelo que alteram a janela de intersecção entre elas. Assim, a janela onde se calcula a função de erro muda (ou pode mudar) durante o processo de minimização, razão pela qual tem o nome de adaptativa.

Devido à natureza iterativa e baseada em gradiente da solução obtida o desempenho do algoritmo desenvolvido está limitado, à semelhança do que acontece com outros algoritmos de estimação de movimento directamente a partir das imagens, pela possibilidade de convergência para mínimos locais. Para combater esta limitação os métodos propostos na teoria recorrem a



pirâmides multi-resolução. Esta solução apresenta, no entanto, a grave limitação de melhorar o desempenho apenas em situações onde o movimento predominante é o de translação tal como foi descrito na secção 1.2.1. Para obviar este problema propomos uma nova pirâmide adaptada ao tipo de movimento que se admite na imagem.

A formulação que se propõe neste trabalho permite, ainda, lidar com a construção de panoramas a partir de múltiplas imagens. Mostramos que nesta situação a estimativa de máxima verosimilhança para o panorama pode ser obtida de forma fechada em função das estimativas dos parâmetros de movimento. Mostramos, ainda, que a estimativa de máxima verosimilhança dos parâmetros do modelo de movimento que relaciona todas as imagens pode ser obtida por minimização de uma função que mede o erro entre as imagens nas várias regiões onde estas se intersectam. A minimização desta função é feita de uma forma iterativa semelhante à minimização da função de erro no caso de duas imagens. Note-se, no entanto, que a inicialização do algoritmo vai recorrer ao resultado em baixa resolução do método de alinhamento cumulativo. O algoritmo final desenvolvido permite o alinhamento simultâneo e global de um conjunto arbitrário de imagens apresentando uma grande robustez à propagação de erros sem perder drasticamente eficiência computacional.

## 1.4 Contribuições Originais

Identificamos como contribuições originais deste trabalho:

- **Janela adaptativa:** Minimizando a função de erro entre imagens com recurso a uma janela adaptativa tornamos robusta a estimação de movimento directamente a partir das imagens mesmo nos casos onde existe uma pequena sobreposição entre elas;
- **Pirâmide dependente do modelo de movimento:** Usando pirâmides de multi-resolução adaptadas ao modelo de movimento melhoramos a convergência dos métodos iterativos de estimação dos parâmetros de movimento;
- **Estimação global:** Estimando o panorama de uma forma simultânea no caso de múltiplas imagens evitamos a propagação de erros típica das abordagens cumulativas sem aumentar drasticamente o peso computacional;
- **Formulação clássica:** Formulando o problema do alinhamento de imagens no contexto da teoria clássica da estimação, conseguimos provar que a primeira e terceira contribuições correspondem a soluções óptimas para o problema que tentam resolver;

A primeira destas contribuições encontra-se publicada na referência [13] e ainda no Anexo D. A terceira destas contribuições encontra-se submetida para publicação [14] e ainda no Anexo E. Encontra-se ainda em fase de preparação um artigo com o título *Maximum Likelihood Mosaics* com vista a sumeter a *Transactions on Image Processing*.

## 1.5 Organização do Relatório

O presente relatório organiza-se em seis capítulos e quatro anexos.

O primeiro capítulo, Introdução, descreve sucintamente o problema abordado. Neste capítulo mostram-se algumas das motivações para estudar o problema e algumas das suas aplicações em diversas áreas da engenharia. Para além disso apresenta-se de uma forma muito sucinta o estado actual da investigação e resume-se a abordagem proposta. Por fim este capítulo mostra as contribuições que identificamos como originais e indica a estrutura do relatório (precisamente nesta secção).

O segundo capítulo debruça-se sobre o problema do alinhamento de duas imagens. Apresenta-se a formulação deste problema no contexto da teoria clássica da estimação e compara-se a solução obtida com as soluções existentes na literatura mostrando-se que a principal diferença encontrada – o uso de uma janela adaptativa – conduz a uma melhoria de resultados, especialmente nas situações onde existe pouca sobreposição entre as imagens. Ainda neste capítulo apresenta-se o algoritmo de estimação do panorama e mostram-se alguns resultados experimentais que apoiam as conclusões apresentadas.

No terceiro capítulo estuda-se a influência do método de multi-resolução no desempenho global do algoritmo. Propõe-se um método de multi-resolução adaptado ao modelo de movimento utilizado e demonstra-se a sua superioridade face aos métodos na literatura para o modelo de rotação pura da câmara.

No quarto capítulo trata-se o problema do alinhamento de múltiplas imagens. À semelhança do que foi feito no segundo capítulo apresenta-se a formulação deste problema no contexto da teoria clássica da estimação e procede-se à sua resolução. Mostra-se que o algoritmo desenvolvido evita a propagação de erros não sendo, no entanto, demasiado pesado computacionalmente para tratar um número razoável de imagens. Mais uma vez também neste capítulo são apresentadas várias experiências em *MatLab* que suportam os resultados obtidos.

Por fim, e embora os capítulos 2, 3 e 4 apresentem resultados parciais ilustrativos das técnicas propostas, no quinto capítulo apresentamos resultados globais, i.e., mostramos um conjunto de panoramas construídos recorrendo aos métodos descritos. O sexto capítulo conclui o trabalho e apresenta propostas para trabalho futuro.

Para além do corpo principal, fazem ainda parte do trabalho quatro anexos numerados de A a D. O primeiro destes anexos fornece informação adicional sobre em que situações é possível criar panoramas e apresenta os modelos de movimento que são utilizados ao longo de todo o trabalho. Os anexos B e C apresentam deduções de fórmulas presentes nos capítulos 2 e 4 respectivamente. Por fim no anexo D encontra-se o artigo [13] apresentado na *IEEE International Workshop on Multimedia Signal Processing* em Siena, Itália, em Outubro de 2004. Este artigo descreve uma das contribuições originais deste relatório – o uso de uma *janela adaptativa* descrito no capítulo 2.

## Capítulo 2

# Alinhamento de um Par de Imagens: Janela Adaptativa

Neste capítulo concentramo-nos exclusivamente no alinhamento de duas imagens. Na primeira secção usamos a teoria clássica da estimação para encontrar a estimativa de máxima verosimilhança para o panorama e para os parâmetros do modelo de movimento.

Na secção seguinte comparamos o estimador para os parâmetros do modelo de movimento proposto com os estimadores presentes na literatura evidenciando a principal diferença do estimador proposto: o uso de uma *janela adaptativa*.

Na secção 2.3 descrevemos um algoritmo de estimação dos parâmetros do modelo de movimento e, por fim, na última secção expomos um conjunto de experiências em *MatLab* que validam os resultados obtidos.

## 2.1 Alinhamento como Problema de Estimação

### 2.1.1 Notação

Associamos funções às imagens observadas –  $I(\mathbf{x})$  e  $I'(\mathbf{x}')$  – e ao panorama –  $P(\mathbf{x})$ . As funções que representam cada uma das imagens estão definidas em domínios –  $\mathcal{D}$  e  $\mathcal{D}'$  – definidos em sistemas de coordenadas diferentes –  $\mathbf{x}$  e  $\mathbf{x}'$ . A definição seguinte clarifica todos estes conceitos.

**Definição 2.1** *Notação para o alinhamento de um par de imagens:*

$I(\mathbf{x}), I'(\mathbf{x}') :$  *Padrões de intensidades luminosas da primeira e segunda imagens respectivamente;*

$P(\mathbf{x}) :$  *Padrão de intensidade luminosa do panorama;*

$\mathbf{x} = [x \ y]^T, \mathbf{x}' = [x' \ y']^T :$  *Sistemas de coordenadas associados a cada uma das imagens;*

$\mathcal{D}, \mathcal{D}' :$  *Domínios onde os padrões de intensidade luminosa que formam as imagens estão definidas.*

Note-se que, por simplicidade, admitimos a primeira imagem já está alinhada com a imagem panorâmica uma vez que as funções  $I(\mathbf{x})$  e  $P(\mathbf{x})$  partilham o mesmo sistema de coordenadas:  $\mathbf{x} = [x \ y]^T$ . Desta forma pretendemos obter uma representação do panorama no sistema de coordenadas da primeira imagem.

Consideramos que o panorama se mantém constante em ambas as imagens e admitimos que existe apenas movimento tridimensional da câmara que registou as imagens. O movimento tridimensional da câmara induz um conjunto de alterações no padrão bidimensional de intensidades das imagens. Nos casos onde é possível criar panoramas estas alterações podem ser modeladas por uma transformação entre as coordenadas das imagens que depende apenas de um pequeno número de parâmetros [7]. Nesta fase do trabalho, a relação genérica entre as coordenadas das imagens que se apresenta abaixo é suficiente para formular o problema. Este assunto será retomado quando se discutir a implementação prática dos algoritmos de estimação de panoramas (secções 2.3.2 e 2.3.3) e é tratado em pormenor no Anexo A.

**Definição 2.2** *O movimento 3D da câmara induz um movimento 2D no plano da imagem que pode ser modelado por uma **Transformação** entre coordenadas das imagens da forma:*

$$\mathbf{x}' = m(\mathbf{x}, \boldsymbol{\theta}) , \quad (2.1)$$

onde  $\boldsymbol{\theta}$  representa o vector de parâmetros da transformação de coordenadas, também referidos como parâmetros do modelo de movimento. Esta transformação pode ser invertida obtendo-se:

$$\mathbf{x} = m^{-1}(\mathbf{x}', \boldsymbol{\theta}) . \quad (2.2)$$

### 2.1.2 Formulação do Problema

Para formular o problema do alinhamento de imagens segundo a teoria da estimação comecemos por considerar que cada uma das imagens representa um conjunto de amostras do panorama contaminadas por ruído. Assumimos que este ruído é gaussiano e independente de amostra para amostra (de *pixel* para *pixel*) e de imagem para imagem. Assim temos :

$$I(\mathbf{x}) = P(\mathbf{x}) + R(\mathbf{x}) , \quad I'(\mathbf{x}') = P(m^{-1}(\mathbf{x}', \boldsymbol{\theta})) + R'(\mathbf{x}') , \quad (2.3)$$

onde  $R(\mathbf{x})$  e  $R'(\mathbf{x}')$  representam o ruído.

Note-se que, tal como foi introduzido na definição 2.1, a função  $P(\mathbf{x})$  descreve as intensidades luminosas do panorama no referencial da primeira imagem  $\mathbf{x}$ . Como a variável independente na segunda equação de (2.3) é  $\mathbf{x}'$ , recorreremos à equação (2.2) para escrever  $P(\mathbf{x})$  em função de  $\mathbf{x}'$ .

Admitindo que cada uma das amostras de ruído é uma variável aleatória normal com média zero e variância  $\sigma^2$ , a função densidade de probabilidade (f.d.p.) de qualquer amostra de ambas as imagens, genericamente  $\mathbf{x}_i$  ou  $\mathbf{x}'_i$ , é também uma variável aleatória normal agora com média igual ao valor real do panorama nesse ponto e variância igual à do ruído. Isto é:

$$\begin{aligned} r(\mathbf{x}_i) \sim \mathcal{N}(0, \sigma^2) &\Rightarrow I(\mathbf{x}_i) \sim \mathcal{N}(P(\mathbf{x}_i), \sigma^2) , \\ r'(\mathbf{x}'_i) \sim \mathcal{N}(0, \sigma^2) &\Rightarrow I'(\mathbf{x}'_i) \sim \mathcal{N}(P(m^{-1}(\mathbf{x}'_i, \boldsymbol{\theta})), \sigma^2) . \end{aligned} \quad (2.4)$$

### 2.1.3 Estimativa de Máxima Verosimilhança

Escrevendo a função densidade de probabilidade –  $\mathcal{P}(\cdot)$  – explicitamente para cada amostra de cada imagem obtemos:

$$\begin{aligned}\mathcal{P}(I(\mathbf{x}_i)|P(\mathbf{x})) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ \frac{-1}{2\sigma^2} [I(\mathbf{x}_i) - P(\mathbf{x}_i)]^2 \right\} , \\ \mathcal{P}(I'(\mathbf{x}'_i)|P(\mathbf{x}), \boldsymbol{\theta}) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ \frac{-1}{2\sigma^2} [I'(\mathbf{x}'_i) - P(m^{-1}(\mathbf{x}'_i, \boldsymbol{\theta}))]^2 \right\} .\end{aligned}\quad (2.5)$$

Sendo as amostras de ruído independentes entre si, as amostras da imagem também o são. Assim a f.d.p. conjunta de todas as amostras (*pixels*) que formam uma imagem é dada pelo produto das f.d.p. de cada *pixel*. Considerando que existem  $N$  *pixels* em cada imagem temos:

$$\begin{aligned}\mathcal{P}(I(\mathbf{x})|P(\mathbf{x})) &= (2\pi\sigma^2)^{-N/2} \exp \left\{ - (2\sigma^2)^{-1} \sum_{\mathbf{x} \in \mathcal{D}} [I(\mathbf{x}) - P(\mathbf{x})]^2 \right\} , \\ \mathcal{P}(I'(\mathbf{x}')|P(\mathbf{x}'), \boldsymbol{\theta}) &= (2\pi\sigma^2)^{-N/2} \exp \left\{ - (2\sigma^2)^{-1} \sum_{\mathbf{x}' \in \mathcal{D}'} [I'(\mathbf{x}') - P(m^{-1}(\mathbf{x}', \boldsymbol{\theta}))]^2 \right\} .\end{aligned}\quad (2.6)$$

A partir deste resultado, e tendo em conta que o ruído é independente de imagem para imagem, podemos escrever a função de verosimilhança do par de imagens. Esta função é apenas a f.d.p. conjunta das duas imagens condicionada ao panorama,  $P(\mathbf{x})$ , e ao conjunto de parâmetros da transformação de coordenadas,  $\boldsymbol{\theta}$ :

$$\begin{aligned}\mathcal{P}(I(\mathbf{x}), I'(\mathbf{x}')|P(\mathbf{x}), \boldsymbol{\theta}) &= \\ \frac{1}{(2\pi\sigma^2)^N} \exp \left\{ \frac{-1}{2\sigma^2} \left( \sum_{\mathbf{x} \in \mathcal{D}} [I(\mathbf{x}) - P(\mathbf{x})]^2 + \sum_{\mathbf{x}' \in \mathcal{D}'} [I'(\mathbf{x}') - P(m^{-1}(\mathbf{x}', \boldsymbol{\theta}))]^2 \right) \right\} .\end{aligned}\quad (2.7)$$

Para simplificar este resultado escrevemos o segundo somatório em termos de  $\mathbf{x}$  usando a equação (2.1):

$$\begin{aligned}\mathcal{P}(I(\mathbf{x}), I'(\mathbf{x}')|P(\mathbf{x}), \boldsymbol{\theta}) &= \\ \frac{1}{(2\pi\sigma^2)^N} \exp \left\{ \frac{-1}{2\sigma^2} \left( \sum_{\mathbf{x} \in \mathcal{D}} [I(\mathbf{x}) - P(\mathbf{x})]^2 + \sum_{\mathbf{x} \in \mathcal{D}'_{\mathbf{x}}(\boldsymbol{\theta})} [I'(m(\mathbf{x}, \boldsymbol{\theta})) - P(\mathbf{x})]^2 \right) \right\} ,\end{aligned}\quad (2.8)$$

onde  $\mathcal{D}'_{\mathbf{x}}(\boldsymbol{\theta})$  representa o conjunto de valores de  $\mathbf{x}$  onde a segunda imagem foi amostrada. Isto é,  $\mathcal{D}'_{\mathbf{x}}(\boldsymbol{\theta})$  é dado por:

$$\mathcal{D}'_{\mathbf{x}}(\boldsymbol{\theta}) = \{\mathbf{x} : \mathbf{x}' \in \mathcal{D}' \Leftrightarrow m(\mathbf{x}, \boldsymbol{\theta}) \in \mathcal{D}'\} \quad (2.9)$$

Para obter o resultado numa forma mais geral que facilite as deduções que se vão seguir estendemos  $I(\cdot)$  e  $I'(\cdot)$  a todo o  $\mathbb{R}^2$  e definimos as funções:

$$h(\mathbf{x}) = \begin{cases} 1 & \text{se } \mathbf{x} \in \mathcal{D} \\ 0 & \text{caso contrário} \end{cases} , \quad h'(\mathbf{x}, \boldsymbol{\theta}) = \begin{cases} 1 & \text{se } \mathbf{x} \in \mathcal{D}'_{\mathbf{x}}(\boldsymbol{\theta}) \\ 0 & \text{caso contrário} \end{cases} . \quad (2.10)$$

Usando esta notação obtemos o resultado:

**Resultado 2.3** *A Função de Verosimilhança para o problema do alinhamento de duas imagens é dada por:*

$$\mathcal{P}(I(\mathbf{x}), I'(\mathbf{x}')|P(\mathbf{x}), \boldsymbol{\theta}) = \frac{1}{(2\pi\sigma^2)^N} \exp \left\{ \frac{-1}{2\sigma^2} \sum_{\mathbf{x} \in \mathbb{R}^2} \left( [I(\mathbf{x}) - P(\mathbf{x})]^2 h(\mathbf{x}) + [I'(m(\mathbf{x}, \boldsymbol{\theta})) - P(\mathbf{x})]^2 h'(\mathbf{x}, \boldsymbol{\theta}) \right) \right\}. \quad (2.11)$$

Para a discussão que se segue é mais vantajoso usar o simétrico do logaritmo desta função:

$$-\ln(\mathcal{P}(I(\mathbf{x}), I'(\mathbf{x}')|P(\mathbf{x}), \boldsymbol{\theta})) = N \ln(2\pi\sigma^2) + \frac{1}{2\sigma^2} \sum_{\mathbf{x} \in \mathbb{R}^2} \left( [I(\mathbf{x}) - P(\mathbf{x})]^2 h(\mathbf{x}) + [I'(m(\mathbf{x}, \boldsymbol{\theta})) - P(\mathbf{x})]^2 h'(\mathbf{x}, \boldsymbol{\theta}) \right). \quad (2.12)$$

A estimativa de máxima verosimilhança dos parâmetros desconhecidos – o panorama  $P(\mathbf{x})$  e os parâmetros da transformação de coordenadas  $\boldsymbol{\theta}$  – é obtida maximizando a função de verosimilhança (ver [15], págs. 63 a 65) ou, de uma forma equivalente, minimizando o simétrico do logaritmo dessa função:

**Resultado 2.4** *Os estimadores de máxima verosimilhança para o problema do alinhamento de duas imagens são dados por:*

$$\left( \widehat{P(\mathbf{x})}, \widehat{\boldsymbol{\theta}} \right) = \arg \max_{(P(\mathbf{x}), \boldsymbol{\theta})} -\ln(\mathcal{P}(I(\mathbf{x}), I'(\mathbf{x}')|P(\mathbf{x}), \boldsymbol{\theta})), \quad (2.13)$$

onde  $-\ln(\mathcal{P}(I(\mathbf{x}), I'(\mathbf{x}')|P(\mathbf{x}), \boldsymbol{\theta}))$  é definido pela equação (2.12)

Procuramos os estimadores de máxima verosimilhança para o panorama. Derivando o simétrico do logaritmo da função de verosimilhança dado pela equação (2.12) em relação a uma amostra genérica  $(\mathbf{x}_i)$  do panorama e igualando a zero temos:

$$\begin{aligned} \frac{\partial}{\partial P(\mathbf{x}_i)} [-\ln \mathcal{P}(I(\mathbf{x}), I'(\mathbf{x}')|P(\mathbf{x}), \boldsymbol{\theta})] &= 0 \\ \Leftrightarrow \frac{1}{\sigma^2} ([I(\mathbf{x}_i) - P(\mathbf{x}_i)] h(\mathbf{x}_i) + [I'(m(\mathbf{x}_i, \boldsymbol{\theta})) - P(\mathbf{x}_i)] h'(\mathbf{x}_i, \boldsymbol{\theta})) &= 0 \\ \Leftrightarrow \widehat{P(\mathbf{x}_i)} &= \frac{I(\mathbf{x}_i)h(\mathbf{x}_i) + I'(m(\mathbf{x}_i, \boldsymbol{\theta}))h'(\mathbf{x}_i, \boldsymbol{\theta})}{h(\mathbf{x}_i) + h'(\mathbf{x}_i, \boldsymbol{\theta})}. \end{aligned} \quad (2.14)$$

Uma vez que esta equação é válida para qualquer *pixel* do panorama chegamos ao resultado:

**Resultado 2.5** *A estimativa de máxima verosimilhança para o panorama pode ser obtida de uma forma imediata em função da estimativa dos parâmetros da transformação de coordenadas usando a equação:*

$$\widehat{P(\mathbf{x})} = \frac{I(\mathbf{x})h(\mathbf{x}) + I'(m(\mathbf{x}, \boldsymbol{\theta}))h'(\mathbf{x}, \boldsymbol{\theta})}{h(\mathbf{x}) + h'(\mathbf{x}, \boldsymbol{\theta})}. \quad (2.15)$$

Este resultado indica que a estimativa de máxima verosimilhança para o panorama é igual à média das observações na primeira e na segunda imagem na região onde ambas as imagens registaram o panorama (isto é, na região onde  $h(\mathbf{x}_i) = h'(\mathbf{x}_i, \boldsymbol{\theta}) = 1$ ). Na região onde o panorama foi registado por apenas uma das imagens (a primeira se  $h(\mathbf{x}_i) = 1 \wedge h'(\mathbf{x}_i, \boldsymbol{\theta}) = 0$  ou a segunda se  $h(\mathbf{x}_i) = 0 \wedge h'(\mathbf{x}_i, \boldsymbol{\theta}) = 1$ ) a estimativa de máxima verosimilhança é igual à única observação disponível.

Este resultado prova que a estimativa do panorama pode ser obtida de uma forma trivial desde que se conheçam as estimativas dos parâmetros da transformação de coordenadas. Substituindo a equação (2.15) em (2.12) obtemos, após a manipulação descrita no Anexo B:

$$-\ln \left( \mathcal{P} \left( I(\mathbf{x}), I'(\mathbf{x}') | \widehat{P(\mathbf{x})}, \boldsymbol{\theta} \right) \right) = N \ln (2\pi\sigma^2) + \frac{1}{4\sigma^2} \sum_{\mathbf{x} \in \mathbb{R}^2} [I(\mathbf{x}) - I'(m(\mathbf{x}, \boldsymbol{\theta}))]^2 h(\mathbf{x})h'(\mathbf{x}, \boldsymbol{\theta}) . \quad (2.16)$$

Para tornar este resultado mais simples e a sua interpretação mais fácil definimos o conjunto de intersecção entre as duas imagens como:

$$\mathcal{R}_A(\boldsymbol{\theta}) = \mathcal{D} \cap \mathcal{D}'_{\mathbf{x}}(\boldsymbol{\theta}) = \{\mathbf{x} : h(\mathbf{x}) = h'(\mathbf{x}, \boldsymbol{\theta}) = 1\} \quad (2.17)$$

Usando este conjunto podemos escrever:

**Resultado 2.6** *As estimativas de máxima verosimilhança para os parâmetros da transformação de coordenadas são dadas por:*

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} E_A(\boldsymbol{\theta}) , \quad E_A(\boldsymbol{\theta}) = \sum_{\mathbf{x} \in \mathcal{R}_A(\boldsymbol{\theta})} e^2(\mathbf{x}, \boldsymbol{\theta}) , \quad (2.18)$$

onde  $E_A(\boldsymbol{\theta})$  é a função de erro a minimizar sendo  $e(\mathbf{x}, \boldsymbol{\theta})$  o erro entre as imagens dado por:

$$e(\mathbf{x}, \boldsymbol{\theta}) = I(\mathbf{x}) - I'(m(\mathbf{x}, \boldsymbol{\theta})) . \quad (2.19)$$

## 2.2 Impacto da Janela na Função de Erro

De uma forma geral os algoritmos presentes na literatura estimam o panorama de uma forma equivalente à obtida na secção anterior. Embora sem recorrer à estimativa de máxima verosimilhança estes algoritmos acabam por tentar estimar os parâmetros da transformação de coordenadas minimizando uma função de erro semelhante à que foi apresentada na secção anterior. Esta função de erro é dada por:

$$E(\boldsymbol{\theta}) = \sum_{\mathbf{x} \in \mathcal{R}} e^2(\mathbf{x}, \boldsymbol{\theta}), \quad e(\mathbf{x}, \boldsymbol{\theta}) = I(\mathbf{x}) - I'(m(\mathbf{x}, \boldsymbol{\theta})) , \quad (2.20)$$

onde  $\mathcal{R}$  representa uma região (ou janela) fixa de valores de  $\mathbf{x}$ , com forma rectangular e pré-especificada.

Note-se que a grande diferença entre estes algoritmos e o que propomos é o uso da janela fixa  $\mathcal{R}$  ao invés da janela adaptativa  $R_A$ . Neste capítulo estuda-se a influência desta janela no desempenho global do algoritmo.

Quando existe grande área de sobreposição entre imagens, a janela  $\mathcal{R}$  é escolhida como um rectângulo relativamente grande no interior destas. No entanto, quando a área é pequena é difícil fazer, *a priori*, uma escolha apropriada para  $\mathcal{R}$  por duas razões. Em primeiro lugar porque não se sabe onde se encontra a zona de sobreposição pelo que é difícil escolher a localização de  $\mathcal{R}$ . Em segundo lugar porque a imposição *a priori* de uma janela pequena conduz a piores estimativas para os parâmetros  $\theta$  uma vez que o mínimo de  $E(\theta)$  em (2.20) fica menos pronunciado e, simultaneamente, o fenómeno dos mínimos locais se torna mais grave.

Para ilustrar o impacto da janela apresenta-se na Fig. (2.1) a evolução típica do erro  $E$  em (2.20), como função de um único parâmetro  $\theta$ , para vários tamanhos da janela  $\mathcal{R}$ . Naturalmente, quanto maior for  $\mathcal{R}$ , menor é o domínio  $\{\theta\}$  onde podemos calcular  $E(\theta)$ . A existência de vários mínimos e a suavidade do mínimo no valor correcto do parâmetro ( $\theta = 20$ ) nos gráficos de cima (obtidos com janelas pequenas) contrastam com a existência bem marcada de um único mínimo no gráfico do canto inferior direito (obtido com a maior janela). Note-se que a escala vertical dos gráficos é diferente.

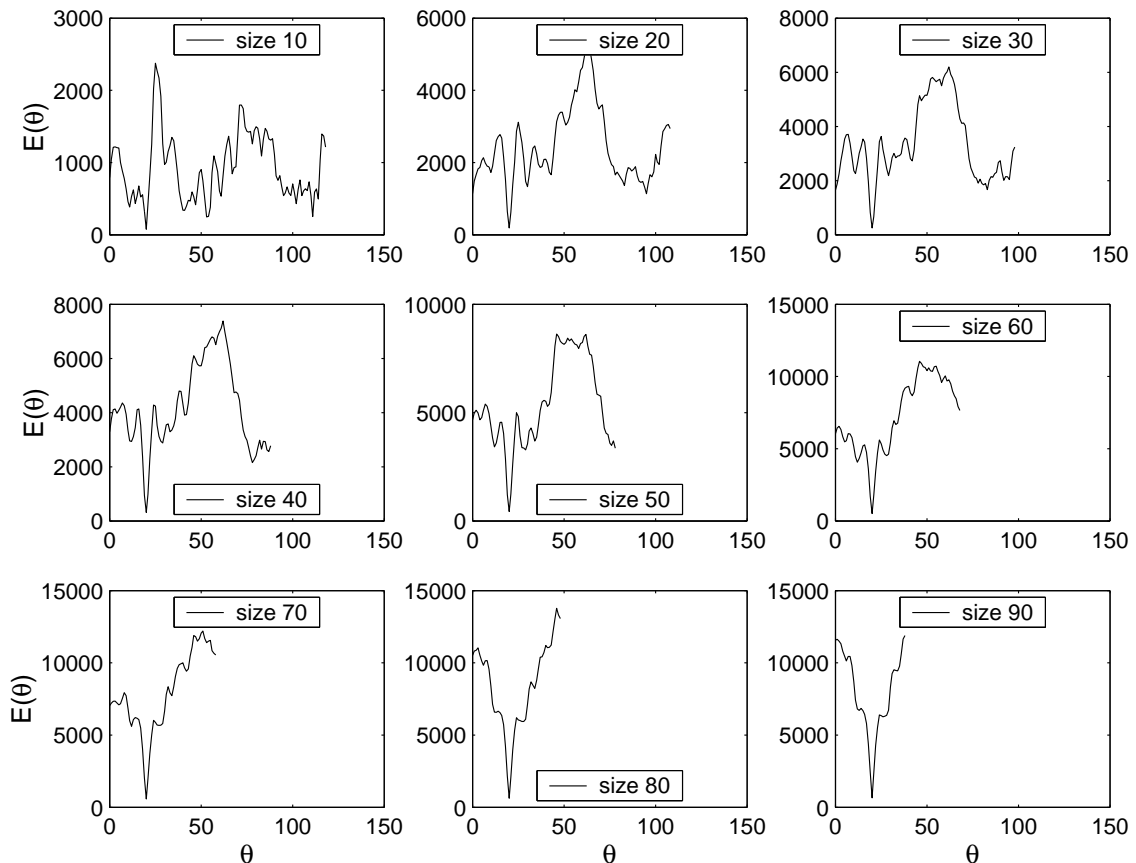


Figura 2.1: Erro  $E(\theta)$  para vários tamanhos da janela  $\mathcal{R}$ , como indicado nos gráficos. Neste caso existe apenas um parâmetro de movimento sendo o seu valor real  $\theta = 20$ . Note-se que quanto maior for  $\mathcal{R}$ , menor é o domínio  $\{\theta\}$  onde podemos calcular  $E(\theta)$



Esta figura mostra também que não é possível avaliar o erro  $e(\boldsymbol{\theta}, \mathbf{x})$  em (2.20) para todos os pares de valores de  $\boldsymbol{\theta}$  e  $\mathbf{x}$ . Este facto leva ao principal problema da janela fixa  $\mathcal{R}$ : o alinhamento das imagens só é possível quando a sobreposição entre estas contém  $\mathcal{R}$ . O uso de algoritmos iterativos – a única forma de minimizar a função de erro em tempo útil para a maioria das aplicações – torna esta limitação ainda mais forte. Isto porque a cada iteração as estimativas de  $\boldsymbol{\theta}$  são alteradas o que faz com que a região onde as imagens se sobrepõem se altere. Ora a limitação da janela fixa obriga a que a sobreposição entre as imagens inclua  $\mathcal{R}$  não apenas na solução final mas também em todas as iterações do algoritmo.

Como ficará claro na secção 2.4.1 o uso da janela fixa dificulta o alinhamento automático das imagens. De facto, especificando *a priori* uma janela fixa  $\mathcal{R}$ , não conseguimos lidar com todas as situações. Ainda que uma janela pequena caiba facilmente na região de sobreposição das imagens, ela pode facilmente aumentar o erro de estimação uma vez que induz o aparecimento de muitos mínimos locais na função de erro tal como é ilustrado nos gráficos de cima da Fig. (2.1). Por outro lado, o uso de uma janela grande limita o alinhamento de imagens a situações onde a sobreposição também é grande.

Para obviar esta limitação propomos o uso de uma janela adaptativa  $\mathcal{R}_A(\boldsymbol{\theta})$  definida em cada instante do algoritmo como a maior região onde é possível calcular o erro  $e(\boldsymbol{\theta}, \mathbf{x})$  como definido em (2.20).

## 2.3 Estimação do Movimento entre um Par de Imagens

### 2.3.1 Minimização da Função de Erro para Resolução Fixa

Para minimizar a função de erro definida no resultado 2.6 vamos utilizar um algoritmo iterativo. Em cada iteração partimos de uma estimativa para o conjunto de parâmetros,  $\boldsymbol{\theta}_0$ , para escrever a nova estimativa dos parâmetros,  $\boldsymbol{\theta}$  como:

$$\boldsymbol{\theta} = \boldsymbol{\theta}_0 + \boldsymbol{\delta} , \quad (2.21)$$

onde a actualização do valor dos parâmetros é estimada por minimização do somatório do quadrado do erro na janela adaptativa,  $\mathcal{R}_A(\boldsymbol{\theta})$ . Isto é:

$$\hat{\boldsymbol{\delta}} = \arg \min_{\boldsymbol{\delta}} \sum_{\mathbf{x} \in \mathcal{R}_A(\boldsymbol{\theta}_0)} e^2(\mathbf{x}, \boldsymbol{\theta}_0 + \boldsymbol{\delta}) , \quad (2.22)$$

onde  $e(\cdot)$  já foi definido no resultado 2.6. Tal como referido em (2.17) a janela adaptativa  $\mathcal{R}_A(\boldsymbol{\theta}_0)$  corresponde á intersecção entre o domínio da primeira imagem,  $I$ , dado por  $\mathcal{D}$  e o domínio de  $I'$  nas coordenadas da primeira imagem dado por  $\mathcal{D}'_{\mathbf{x}}$  definido em (2.9).

Para minimizar a função de erro definida em (2.22) recorremos a um método de Gauss-Newton semelhante ao referido em [11]. Neste método a função de erro  $e(\mathbf{x}, \boldsymbol{\theta})$  é expandida na sua série de Taylor de primeira ordem:

$$e(\mathbf{x}, \boldsymbol{\theta}) \simeq e(\mathbf{x}, \boldsymbol{\theta}_0) + \boldsymbol{\delta}^T \cdot \nabla_{\boldsymbol{\theta}} e(\mathbf{x}, \boldsymbol{\theta}_0) . \quad (2.23)$$

Usando esta equação na função de erro da equação (2.22) e igualando o seu gradiente em  $\boldsymbol{\delta}$  a zero obtemos o seguinte resultado:

**Resultado 2.7** *A estimativa para a actualização dos parâmetros da transformação de coordenadas  $\hat{\boldsymbol{\delta}}$  pode ser obtida por resolução do sistema linear:*

$$\left( \sum_{\mathbf{x} \in \mathcal{R}_A(\boldsymbol{\theta}_0)} \nabla_{\boldsymbol{\theta}} e(\mathbf{x}, \boldsymbol{\theta}_0) \cdot \nabla_{\boldsymbol{\theta}}^T e(\mathbf{x}, \boldsymbol{\theta}_0) \right) \cdot \hat{\boldsymbol{\delta}} + \sum_{\mathbf{x} \in \mathcal{R}_A(\boldsymbol{\theta}_0)} e(\mathbf{x}, \boldsymbol{\theta}_0) \cdot \nabla_{\boldsymbol{\theta}} e(\mathbf{x}, \boldsymbol{\theta}_0) = 0, \quad (2.24)$$

onde  $\nabla_{\boldsymbol{\theta}} e(\mathbf{x}, \boldsymbol{\theta}_0)$  pode ser calculado a partir da sua definição no resultado 2.6 usando a lei de derivação da função composta como:

$$\nabla_{\boldsymbol{\theta}} e(\mathbf{x}, \boldsymbol{\theta}_0) = -\nabla_{\boldsymbol{\theta}} m(\mathbf{x}, \boldsymbol{\theta}_0) \cdot \nabla_{\mathbf{x}} I'(\mathbf{x}) \quad (2.25)$$

### 2.3.2 Modelo de Movimento de Translação

O modelo de movimento de translação pura admite que as coordenadas de ambas as imagens (respectivamente  $\mathbf{x}$  e  $\mathbf{x}'$ ) podem ser relacionadas por uma transformação de coordenadas da forma:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \quad (2.26)$$

Naturalmente esta transformação depende de um vector de parâmetros definido como  $\boldsymbol{\theta} = [b_1 \ b_2]^T$ .

Para estimar o valor destes parâmetros minimizamos a função de erro  $E(\boldsymbol{\theta})$  introduzida no Resultado 2.6. Para minimizar esta função de erro admitimos que existe uma estimativa inicial do conjunto de parâmetros do modelo de movimento  $\boldsymbol{\theta}_0$ . Em cada passo do algoritmo pretendemos refinar esta estimativa recorrendo à expressão:

$$\boldsymbol{\theta} = \boldsymbol{\theta}_0 + \boldsymbol{\delta} \Leftrightarrow \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} b_{1_0} \\ b_{2_0} \end{bmatrix} + \begin{bmatrix} \delta_{b_1} \\ \delta_{b_2} \end{bmatrix} \quad (2.27)$$

onde  $\boldsymbol{\delta}$  corresponde ao vector de actualização dos parâmetros e  $\boldsymbol{\theta}$  corresponde à estimativa inicial a usar no passo seguinte do algoritmo. Na secção 2.3.1 mostrámos que a estimativa de máxima verosimilhança de  $\boldsymbol{\delta}$  é dada, de uma forma genérica, pela equação (2.24) que se reproduz a seguir:

$$\left( \sum_{\mathbf{x} \in \mathcal{R}_A(\boldsymbol{\theta}_0)} \nabla_{\boldsymbol{\theta}} e(\mathbf{x}, \boldsymbol{\theta}_0) \cdot \nabla_{\boldsymbol{\theta}}^T e(\mathbf{x}, \boldsymbol{\theta}_0) \right) \cdot \hat{\boldsymbol{\delta}} + \sum_{\mathbf{x} \in \mathcal{R}_A(\boldsymbol{\theta}_0)} e(\mathbf{x}, \boldsymbol{\theta}_0) \cdot \nabla_{\boldsymbol{\theta}} e(\mathbf{x}, \boldsymbol{\theta}_0) = 0 \quad (2.28)$$

Assim, para encontrar  $\hat{\boldsymbol{\delta}}$  basta aplicar esta equação ao modelo em questão, neste caso o modelo de translação pura. Para tal começamos por calcular o gradiente do modelo,  $m(\mathbf{x}, \boldsymbol{\theta})$ , em relação aos seus parâmetros,  $\boldsymbol{\theta}$ :

$$\nabla_{\boldsymbol{\theta}} m(\mathbf{x}) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (2.29)$$

Seguidamente, e como estamos a resolver o problema para um par de imagens genéricas, representamos respectivamente por  $I'_x(\mathbf{x})$  e  $I'_y(\mathbf{x})$  as derivadas em  $x$  e  $y$  da segunda imagem  $I'(\mathbf{x}')$ . Desta forma podemos escrever o gradiente da segunda imagem em relação às suas coordenadas ( $\mathbf{x}'$ ) como:

$$\nabla_{\mathbf{x}} I'(\mathbf{x}) = \begin{bmatrix} I'_x(\mathbf{x}) \\ I'_y(\mathbf{x}) \end{bmatrix} \quad (2.30)$$

A partir destes dois gradientes podemos usar a equação (2.25) para calcular o gradiente da função de erro  $e(\mathbf{x}, \boldsymbol{\theta})$  em relação a  $\boldsymbol{\theta}$ :

$$\begin{aligned} \nabla_{\boldsymbol{\theta}} e(\mathbf{x}) &= -\nabla_{\boldsymbol{\theta}} m(\mathbf{x}) \cdot \nabla_{\mathbf{x}} I'(\mathbf{x}) \\ &= - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} I'_x(\mathbf{x}) \\ I'_y(\mathbf{x}) \end{bmatrix} \\ &= - \begin{bmatrix} I'_x(\mathbf{x}) \\ I'_y(\mathbf{x}) \end{bmatrix} \end{aligned} \quad (2.31)$$

Aplicando este resultado na equação (2.24) obtemos a estimativa do vector de actualização dos parâmetros:

$$\sum_{\mathbf{x} \in \mathcal{R}_A(\theta_0)} \begin{bmatrix} I_x'^2(\mathbf{x}) & I'_x(\mathbf{x})I'_y(\mathbf{x}) \\ I'_x(\mathbf{x})I'_y(\mathbf{x}) & I_y'^2(\mathbf{x}) \end{bmatrix} \cdot \begin{bmatrix} \hat{\delta}_{b_1} \\ \hat{\delta}_{b_2} \end{bmatrix} = \sum_{\mathbf{x} \in \mathcal{R}_A(\theta_0)} \begin{bmatrix} e(\mathbf{x}, \boldsymbol{\theta}_0) I'_x(\mathbf{x}) \\ e(\mathbf{x}, \boldsymbol{\theta}_0) I'_y(\mathbf{x}) \end{bmatrix} \quad (2.32)$$

### 2.3.3 Modelo de Movimento Afim

O modelo de movimento afim apresenta uma complexidade ligeiramente superior ao modelo de movimento de translação pura. Este modelo admite que as coordenadas de ambas as imagens (respectivamente  $\mathbf{x}$  e  $\mathbf{x}'$ ) podem ser relacionadas por uma transformação de coordenadas da forma:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \quad (2.33)$$

Naturalmente esta transformação depende de um vector de parâmetros que segundo a nossa notação é definido como:

$$\boldsymbol{\theta} = [a_{11}, a_{12}, a_{21}, a_{22}, b_1, b_2]^T \quad (2.34)$$

À semelhança do que foi feito na secção anterior, para estimar o valor destes parâmetros minimizamos a função de erro  $E(\boldsymbol{\theta})$  introduzida no Resultado 2.6. Para minimizar esta função de erro admitimos que existe uma estimativa inicial do conjunto de parâmetros do modelo de movimento  $\boldsymbol{\theta}_0$ . Em cada passo do algoritmo pretendemos refinar esta estimativa recorrendo à expressão:  $\boldsymbol{\theta} = \boldsymbol{\theta}_0 + \boldsymbol{\delta}$  onde  $\boldsymbol{\theta}$  corresponde à estimativa inicial a usar no passo seguinte do algoritmo e  $\boldsymbol{\delta}$  corresponde ao vector de actualização dos parâmetros definido como:

$$\boldsymbol{\delta} = [\delta_{a_{11}}, \delta_{a_{12}}, \delta_{a_{21}}, \delta_{a_{22}}, \delta_{b_1}, \delta_{b_2}]^T \quad (2.35)$$

Na secção 2.3.1 mostrámos que a estimativa de máxima verosimilhança de  $\delta$  é dada, de uma forma genérica, pela equação (2.24) que se reproduz a seguir:

$$\left( \sum_{\mathbf{x} \in \mathcal{R}_A(\theta_0)} \nabla_{\boldsymbol{\theta}} e(\mathbf{x}, \boldsymbol{\theta}_0) \cdot \nabla_{\boldsymbol{\theta}}^T e(\mathbf{x}, \boldsymbol{\theta}_0) \right) \cdot \hat{\delta} + \sum_{\mathbf{x} \in \mathcal{R}_A(\boldsymbol{\theta}_0)} e(\mathbf{x}, \boldsymbol{\theta}_0) \cdot \nabla_{\boldsymbol{\theta}} e(\mathbf{x}, \boldsymbol{\theta}_0) = 0 \quad (2.36)$$

Assim, para encontrar  $\hat{\delta}$  basta aplicar esta equação ao modelo em questão, neste caso o modelo de translação pura. Para tal começamos por calcular o gradiente do modelo,  $m(\mathbf{x}, \boldsymbol{\theta})$ , em relação aos seus parâmetros,  $\boldsymbol{\theta}$ :

$$\nabla_{\boldsymbol{\theta}} m(\mathbf{x}) = \begin{bmatrix} x & y & 0 & 0 & 1 & 0 \\ 0 & 0 & x & y & 0 & 1 \end{bmatrix}^T \quad (2.37)$$

Seguidamente, e como estamos a resolver o problema para um par de imagens genéricas, representamos respectivamente por  $I'_x(\mathbf{x})$  e  $I'_y(\mathbf{x})$  as derivadas em  $x$  e  $y$  da segunda imagem  $I'(\mathbf{x}')$ . Desta forma podemos escrever o gradiente da segunda imagem em relação às suas coordenadas ( $\mathbf{x}'$ ) como:

$$\nabla_{\mathbf{x}} I'(\mathbf{x}) = \begin{bmatrix} I'_x(\mathbf{x}) \\ I'_y(\mathbf{x}) \end{bmatrix} \quad (2.38)$$

A partir destes dois gradientes podemos usar a equação (2.25) para calcular o gradiente da função de erro  $e(\mathbf{x}, \boldsymbol{\theta})$  em relação a  $\boldsymbol{\theta}$ :

$$\begin{aligned} \nabla_{\boldsymbol{\theta}} e(\mathbf{x}) &= -\nabla_{\boldsymbol{\theta}} m(\mathbf{x}) \cdot \nabla_{\mathbf{x}} I'(\mathbf{x}) \\ &= - \begin{bmatrix} x & y & 0 & 0 & 1 & 0 \\ 0 & 0 & x & y & 0 & 1 \end{bmatrix}^T \cdot \begin{bmatrix} I'_x(\mathbf{x}) \\ I'_y(\mathbf{x}) \end{bmatrix} \\ &= - \begin{bmatrix} x I'_x(\mathbf{x}) & y I'_x(\mathbf{x}) & x I'_y(\mathbf{x}) & y I'_y(\mathbf{x}) & I'_x(\mathbf{x}) & I'_y(\mathbf{x}) \end{bmatrix}^T \end{aligned} \quad (2.39)$$

Aplicando este resultado na equação (2.24) obtemos a estimativa do vector de actualização dos parâmetros:

$$\begin{aligned} \sum_{\mathbf{x} \in \mathcal{R}_A(\theta_0)} \begin{bmatrix} x^2 I_x'^2 & xy I_x'^2 & x^2 I_x' I_y' & xy I_x' I_y' & x I_x'^2 & x I_x' I_y' \\ xy I_x'^2 & y^2 I_x'^2 & xy I_x' I_y' & y^2 I_x' I_y' & y I_x'^2 & y I_x' I_y' \\ x^2 I_x' I_y' & xy I_x' I_y' & x^2 I_y'^2 & xy I_y'^2 & x I_x' I_y' & x I_y'^2 \\ xy I_x' I_y' & y^2 I_x' I_y' & xy I_y'^2 & y^2 I_y'^2 & y I_x' I_y' & y I_y'^2 \\ x I_x'^2 & y I_x'^2 & x I_x' I_y' & y I_x' I_y' & I_x'^2 & I_x' I_y' \\ x I_x' I_y' & y I_x' I_y' & x I_y'^2 & y I_y'^2 & I_x' I_y' & I_y'^2 \end{bmatrix} \cdot \begin{bmatrix} \hat{\delta}_{a11} \\ \hat{\delta}_{a12} \\ \hat{\delta}_{a21} \\ \hat{\delta}_{a22} \\ \hat{\delta}_{b1} \\ \hat{\delta}_{b2} \end{bmatrix} = \\ = \sum_{\mathbf{x} \in \mathcal{R}_A(\theta_0)} \begin{bmatrix} e x I_x' \\ e y I_x' \\ e x I_y' \\ e y I_y' \\ e I_x' \\ e I_y' \end{bmatrix} \end{aligned} \quad (2.40)$$

onde se omitiu a dependência de  $I'_x$  e  $I'_y$  em  $\mathbf{x}$  e a dependência de  $e$  em  $\mathbf{x}$  e  $\boldsymbol{\theta}$  para que o resultado ficasse mais compacto.

### 2.3.4 Algoritmo de Multi-Resolução

O método de minimização da função de erro para uma dada resolução descrito na secção anterior baseia-se numa aproximação de Taylor de primeira ordem para a variação da diferença entre as imagens como função dos parâmetros do modelo de movimento. Naturalmente esta aproximação tem uma validade apenas local o que faz com que o algoritmo tal como está apresente uma grande sensibilidade á existência de mínimos locais.

Para tornar a estimação dos parâmetros do modelo de movimento mais robusta aos mínimos locais vamos usar um algoritmo de multi-resolução. Para manter a simplicidade e focar apenas no que foi apresentado usamos a versão tradicional deste algoritmo presente na literatura [12] e descrita em mais pormenor na secção 3.3.

## 2.4 Resultados Experimentais

Nesta secção apresentamos experiências usando três tipos de imagens. Na subsecção 2.4.1 demonstramos experimentalmente a vantagem da janela adaptativa face à tradicional janela fixa. Na secção 2.4.2 mostramos a aplicação do algoritmo desenvolvido ao alinhamento imagens não calibradas com pouca sobreposição, situação em que os algoritmos do tipo *featureless* apresentam pobres resultados. Por fim, na secção 2.4.3 usamos um conjunto de imagens submarinas onde a construção de panoramas é particularmente difícil.

### 2.4.1 Comparação entre a Janela Fixa e a Janela Adaptativa

Para ilustrar como o método da janela adaptativa proposto funciona melhor que o método tradicional da janela fixa, sintetizámos vários pares de imagens "recortando" uma fotografia e adicionando ruído.

Na Fig. 2.2 a sobreposição entre as imagens é grande. A figura de cima na esquerda mostra o alinhamento incorrecto devido à utilização de uma janela muito pequena (de apenas 64 *pixels*). Tanto a janela fixa de 128 *pixels* como a janela adaptativa proposta conduzem a bons resultados como se pode ver nas figuras de cima na direita e de baixo ao centro respectivamente. Note-se, no entanto, que ainda que não seja possível distinguir visualmente a imagem do centro da imagem da direita a estimativa do panorama e dos parâmetros da transformação de coordenadas é mais precisa no caso da janela adaptativa uma vez que a região onde o erro é minimizado é menor.

Na Fig. 2.3 a sobreposição entre as imagens é pequena pelo que é impossível obter um alinhamento correcto usando uma janela de grandes dimensões. Quando se usa uma janela fixa de tamanho 64 *pixels* (figura da esquerda) o algoritmo falha. Por outro lado a imagem da direita mostra que a janela adaptativa proposta consegue alinhar as imagens nesta situação mais difícil.

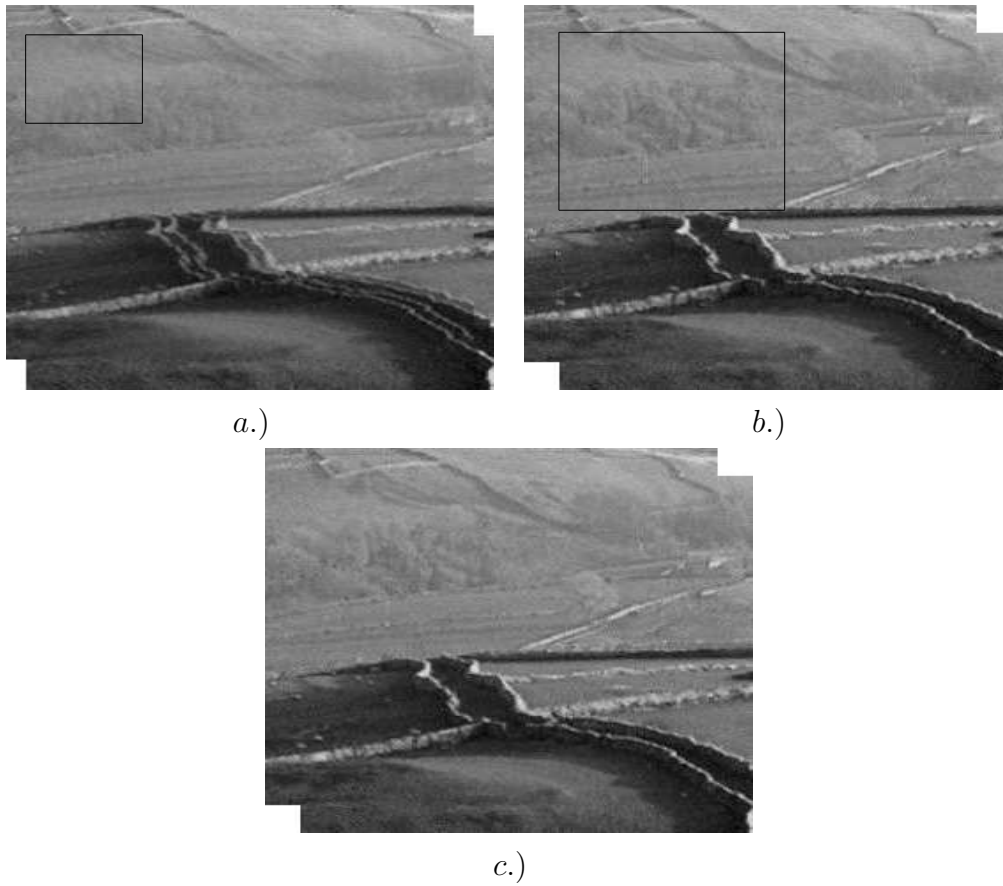


Figura 2.2: Comparação do desempenho da janela fixa e da janela adaptativa no alinhamento de imagens com grande sobreposição: a.) janela fixa de tamanho 64 *pixels* (alinhamento incorrecto); b.) janela fixa de tamanho 128 *pixels*; c.) janela adaptativa.



Figura 2.3: Comparação do desempenho da janela fixa e da janela adaptativa no alinhamento de imagens com pequena sobreposição: à esquerda: usando uma janela fixa de tamanho igual a 64 *pixels* (alinhamento incorrecto); à direita: usando uma janela adaptativa.

### 2.4.2 Alinhamento de Imagens com Pouca Sobreposição

Apresentamos nas figs. 2.4 e 2.5 três exemplos da construção correcta de imagens panorâmicas a partir de imagens reais (não calibradas) com muito pouca sobreposição. Uma vez que a sobreposição entre imagens é muito pequena estas são situações em que o alinhamento é particularmente difícil para os métodos que recorrem à tradicional janela fixa tal como foi demonstrado experimentalmente na secção anterior. Note-se ainda que, por exemplo, a região onde o primeiro par de imagens se intersecta não apresenta praticamente nenhum ponto saliente pelo que o alinhamento destas duas imagens seria igualmente problemático para um método baseado em pontos notáveis. Apesar de todas estas dificuldades a janela adaptativa que propomos não teve qualquer problema em construir todos os panoramas.

Admitimos que as coordenadas de todos os pares de imagens se podiam relacionar por uma transformação afim pelo que foi aplicado o algoritmo descrito na secção 2.3.3. Os panoramas finais mostram que o alinhamento entre as imagens foi correcto.

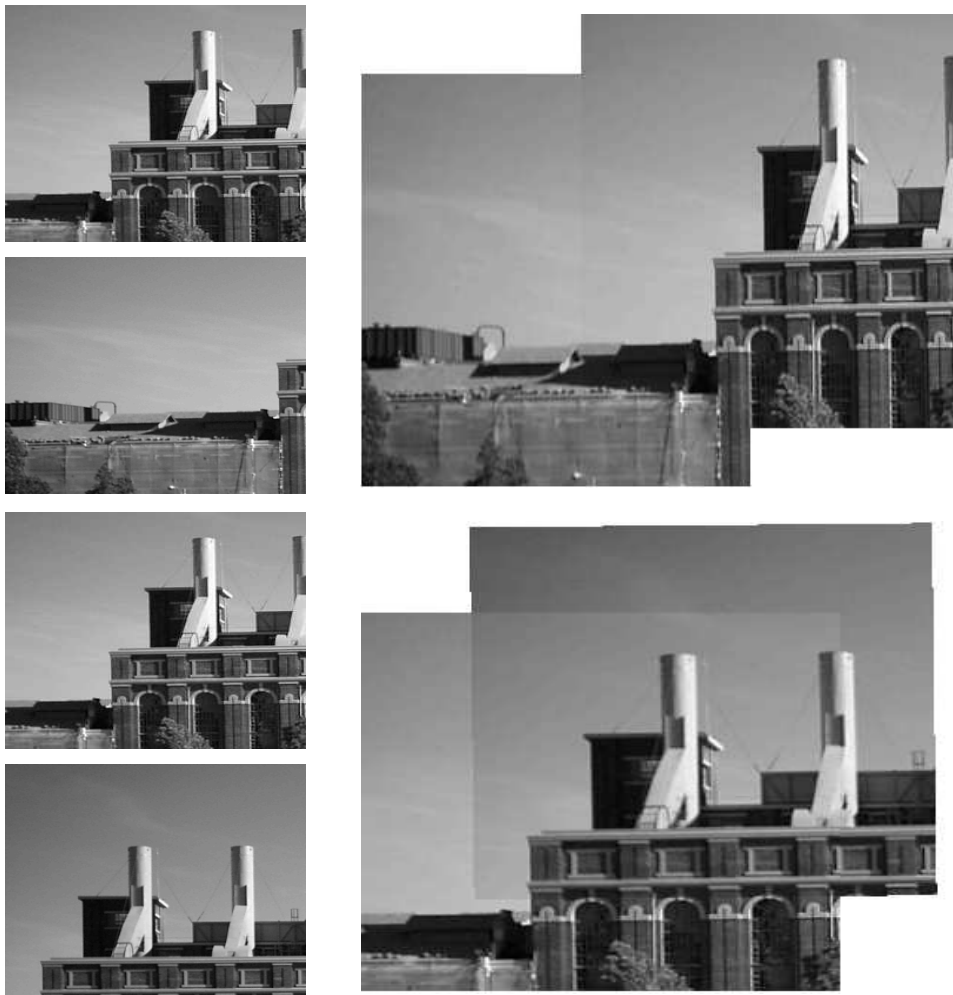


Figura 2.4: Alinhamento de vários pares de imagens não calibradas com pequena sobreposição. (continua na pág. seguinte).



Figura 2.5: Alinhamento de vários pares de imagens não calibradas com pequena sobreposição. (continuando a pág. anterior).

Antes de continuar é importante referir a ligeira descontinuidade dos tons de cinzento quando se passa de uma imagem para a outra. Este efeito deve-se ao facto das imagens apresentarem variações de luminosidade entre si. Este problema pode ser resolvido pela introdução de uma ligeira correcção na forma como o panorama é calculado. No entanto este assunto não faz parte do âmbito deste trabalho pelo que se remete apenas o leitor para a referência [9], págs 19 a 23.

### 2.4.3 Alinhamento de Imagens Submarinas

Como exemplo final apresentamos na fig. 2.6 o resultado do alinhamento de três pares de imagens submarinas. A baixa textura e quase ausência total de pontos notáveis torna estas imagens particularmente difíceis de alinhar. Note-se ainda que apesar da sobreposição entre as imagens não ser muito pequena, não tem forma rectangular. Nesta situação o algoritmo tradicional de janela fixa teria de usar uma janela pequena dentro da região onde existe sobreposição sendo incapaz de utilizar toda a informação disponível ao contrário do que acontece com a janela adaptativa.



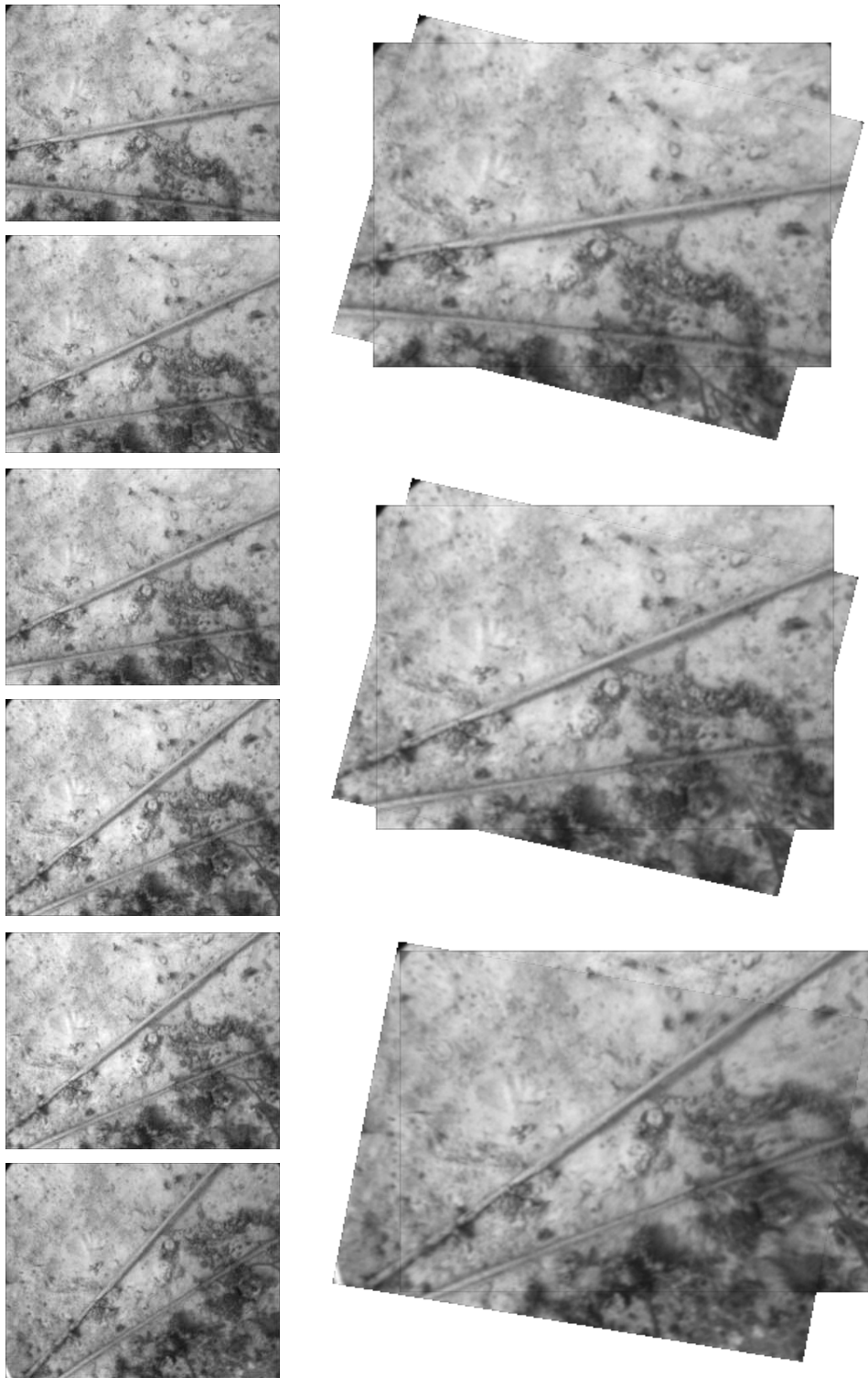


Figura 2.6: Alinhamento de três pares de imagens submarinas não calibradas. À esquerda encontram-se as imagens originais enquanto à direita encontra-se a estimativa do panorama.

## Capítulo 3

# Algoritmos de Multi-Resolução Adaptados ao Modelo de Movimento

Neste capítulo estudamos mais detalhadamente os algoritmos de multi-resolução. Na primeira secção fazemos uma introdução a estes algoritmos e motivamos o seu estudo. Na secção seguinte apresentamos a forma geral do algoritmo de multi-resolução e mostramos que a diferença entre os vários algoritmos consiste na forma como criam a pirâmide de multi-resolução. Na secção 2.3 apresentamos a técnica tradicional de construção da pirâmide e as suas vantagens e limitações. Para tentar ultrapassar as limitações encontradas na pirâmide tradicional propomos um novo método de construção da pirâmide em 2.5. Por fim apresentam-se resultados experimentais que apoiam as conclusões tiradas.

### 3.1 Introdução e Motivação

Tal como foi referido no capítulo de introdução e analisado em mais pormenor no capítulo 2 a dificuldade inerente ao problema de alinhamento de imagens por métodos do tipo *featureless* condensa-se, de uma forma geral, na estimação dos parâmetros do modelo de movimento. A restante grande questão associada ao problema (estimação do panorama a partir dos parâmetros do modelo de movimento) é trivial (tal como é mostrado pelo resultado 2.5).

Neste trabalho a estimação dos parâmetros do movimento é obtida pela maximização da função de verosimilhança. Isto transforma o problema de estimação do movimento num problema de minimização do erro entre imagens como função dos parâmetros do modelo. Este erro é quantificado pela soma do quadrado da diferença entre as imagens na região onde estas se intersectam.

Para proceder à minimização da função de erro (mantendo a resolução das imagens fixa) recorreremos ao método iterativo descrito na secção 2.3.1. Este método assenta numa aproximação de Taylor de primeira ordem para a variação da diferença entre as imagens como função dos parâmetros do modelo de movimento. Naturalmente esta aproximação tem uma validade apenas local o que faz com que o algoritmo deixe de contabilizar o comportamento global da função de erro. Isto conduz a que o algoritmo apresente uma grande sensibilidade à existência de mínimos locais.

Os métodos de multi-resolução que vamos descrever neste capítulo são uma forma de

tornar a convergência do algoritmo para o mínimo global mais robusta, procurando evitar o efeito nefasto da existência de mínimos locais.

## 3.2 Forma Geral dos Algoritmos de Multi-resolução

Antes de passarmos à descrição mais detalhada do algoritmo proposto apresentamos nesta secção uma visão global do funcionamento dos algoritmos de multi-resolução no processo de alinhamento de imagens e construção do panorama.

Para tal resumimos muito brevemente o procedimento de alinhamento no seguinte algoritmo:

1. Construir uma pirâmide de multi-resolução;
2. Para o nível superior da pirâmide (menor resolução), e partindo de uma estimativa inicial arbitrária, recorrer ao processo de minimização para uma resolução fixa (descrito na secção 2.3.1) para estimar os parâmetros do modelo de movimento que relaciona as imagens;
3. Repetir o passo anterior no nível seguinte de resolução da pirâmide, usando como estimativa inicial o resultado da estimação no nível anterior.
4. Tomar como estimativa final dos parâmetros o resultado obtido no último nível da pirâmide (maior resolução, i.e., resolução mais fina ).

Dado que todo este procedimento assenta numa pirâmide de resolução vamos começar por clarificar este conceito. Uma pirâmide de multi-resolução é constituída por uma série de níveis. No nível inferior (base) da pirâmide encontra-se a versão original do conjunto de imagens a alinhar. Para cada um dos níveis acima encontramos versões do conjunto de imagens com resolução progressivamente menor. Neste contexto, menor resolução implica unicamente que cada uma das imagens é representada por um conjunto de amostras menor que a sua representação original. Sendo que este conjunto de amostras não precisa de ser equiespaçado.

## 3.3 Técnica Tradicional de Construção da Pirâmide

A técnica tradicional de construção de pirâmides não é recente [16, 17] sendo considerada fundamental para o sucesso de um algoritmo de estimação do panorama do tipo *featureless* [12]. Nesta técnica colocamos o conjunto de imagens originais na base da pirâmide e construímos os níveis superiores por diminuição da resolução do conjunto de imagens. Isto é, cada quadrado de quatro *pixels* em cada imagem de um dado nível é substituído por um único *pixel* na imagem do nível superior.

Como exemplo de aplicação desta técnica a fig. 3.1 mostra um exemplo da construção correcta de uma imagem panorâmica a partir de imagens reais (não calibradas) com muito pouca sobreposição.

Cada linha na fig. 3.1 corresponde a um nível da pirâmide multi-resolução usada. Em cada linha apresentamos à esquerda as versões de baixa resolução das imagens e à direita o resultado do alinhamento. Na linha de cima apresentamos as versões de mais baixa resolução das imagens. Note-se que o alinhamento entre as imagens ainda é um pouco grosseiro. À medida que descemos na figura a resolução aumenta e a última linha apresenta as imagens originais e o panorama correcto obtido no final do algoritmo.

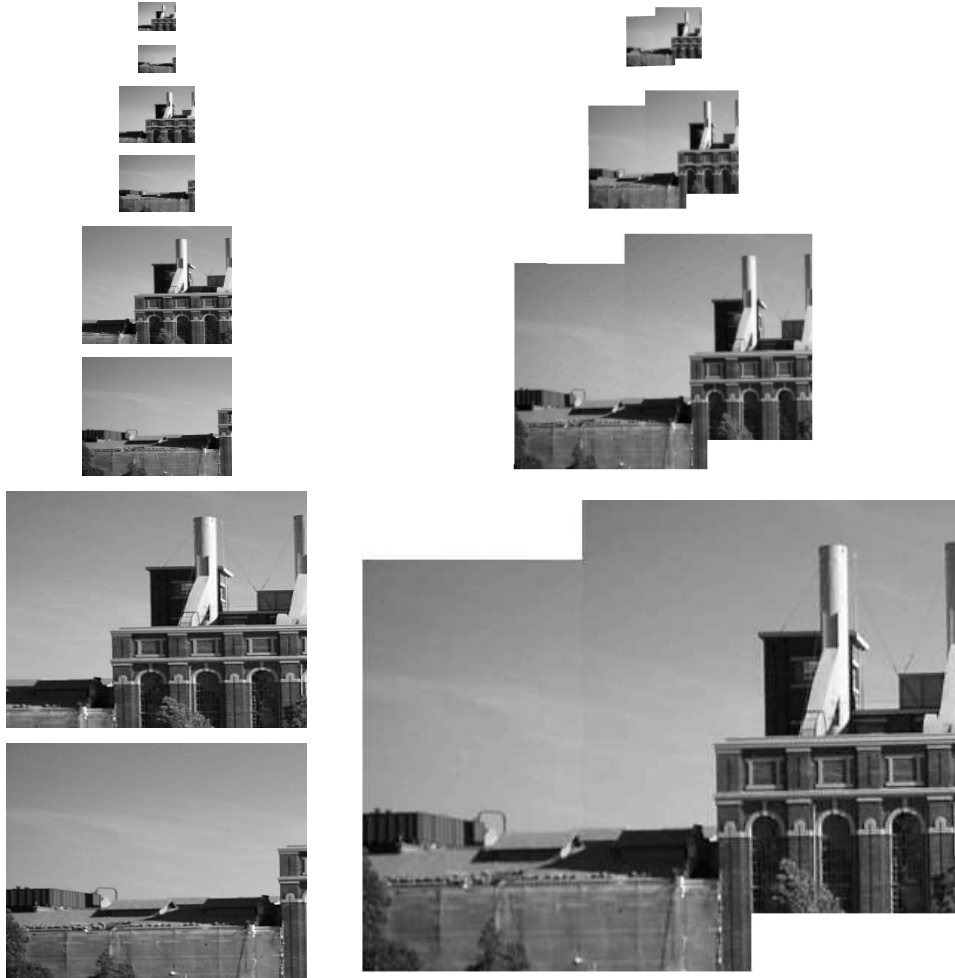


Figura 3.1: Alinhamento de par de imagens não calibradas com pequena sobreposição. As várias linhas ilustram o funcionamento do método de multi-resolução. À esquerda apresentam-se as versões de baixa resolução das imagens e à direita o resultado do alinhamento das imagens da esquerda. A resolução aumenta de cima para baixo sendo que a última linha representa as imagens originais e o panorama final correctamente estimado.

Tal como é mostrado em [12] esta técnica resulta numa pirâmide em que o nível de detalhe nas direcções correspondentes a  $x$  e  $y$  diminui à medida que caminhamos da base para o topo. Esta diminuição do nível de detalhe em  $x$  e  $y$  leva a que a diferença entre as imagens como função do movimento nesses eixos seja uma função mais suave.

Naturalmente, sendo a diferença entre as imagens uma função mais suave nos níveis superiores da pirâmide a aproximação em série de Taylor, usada na estimação de movimento, tem uma validade muito superior o que faz com que as estimativas dos parâmetros do movimento

nesta situação sejam menos sensíveis à existência de mínimos locais. Voltando à fig. 3.1 note-se que o movimento entre as imagens é quase exclusivamente de translação o que faz com que a aplicação da pirâmide tradicional conduza a muito bons resultados.

Mas, se por um lado a diminuição da resolução conduz a uma diferença entre as imagens (como função do movimento em  $x$  e  $y$ ) mais suave, também implica que se perca o detalhe. E, obviamente, esta perda de detalhe leva a que não seja possível obter nos níveis superiores da pirâmide uma estimativa exacta do mínimo da função de erro.

A construção em pirâmide aproveita o melhor de ambas as situações. Quando não temos qualquer informação sobre a localização do mínimo (topo da pirâmide) utilizamos uma função pouco sensível a mínimos locais. Por outro lado quando sabemos aproximadamente a localização do mínimo (base da pirâmide) aproveitamos toda a informação disponível ao utilizar as imagens originais.

Note-se ainda que, para além desta vantagem fundamental, o uso da pirâmide garante ainda uma vantagem acessória que não pode ser negligenciada: o maior número de iterações na procura do mínimo global é realizado nos níveis superiores da pirâmide. Ora, nestes níveis as imagens apresentam uma menor resolução no sentido em que são compostas por menos amostras o que torna o procedimento de minimização dentro de um mesmo nível muito menos pesado do ponto de vista computacional.

Agora que já vimos as principais vantagens do uso do algoritmo de multi-resolução na sua forma tradicional, podemos também encontrar a sua principal limitação: o facto de só aumentar verdadeiramente a robustez em relação aos mínimos locais no movimento segundo os eixos  $x$  ou  $y$ . Isto porque a técnica de diminuição de resolução aumenta a suavidade da função de erro apenas em relação a movimentos em  $x$  e  $y$ .

### 3.4 Técnica Proposta para a Construção da Pirâmide

Para construção da pirâmide propomos uma técnica que tenha em conta o movimento predominante presente na imagem. Para atingir tal objectivo a construção genérica da pirâmide deve ser feita de uma forma semelhante à técnica tradicional: No nível inferior (base) deve ser colocado o conjunto de imagens originais e para obter cada um dos níveis superiores deve ser feita uma diminuição de resolução da imagem que se encontra no nível imediatamente inferior.

No entanto, e ao contrário da técnica tradicional, aqui a diminuição da resolução não é conseguida de uma forma convencional. Antes de diminuir a resolução da imagem fazemos uma mudança nas suas coordenadas de forma que a diminuição afecte de uma forma uniforme não um movimento nas coordenadas da imagem mas um movimento nos parâmetros do modelo de movimento que relaciona as imagens.

O exemplo seguinte clarifica este procedimento.

## Exemplo: Movimento Predominante de Rotação

A fig. 3.2 mostra uma situação onde o movimento entre as imagens corresponde a uma rotação sobre o eixo óptico<sup>1</sup>.

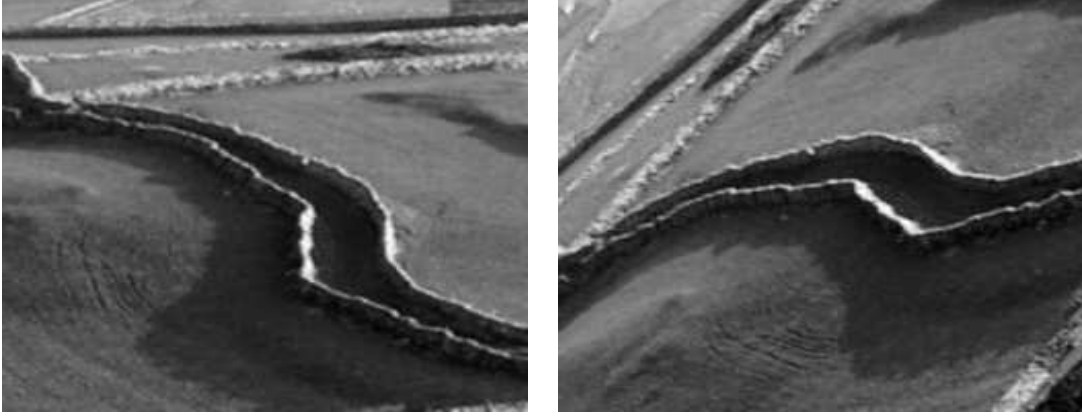


Figura 3.2: Par de imagens relacionadas por um movimento de rotação em torno do eixo óptico

Neste caso o modelo de movimento mais adequado é o da rotação descrito no Anexo A. Este modelo postula que a relação entre as coordenadas das imagens é a indicada em (3.1) onde  $\mathbf{x} = [x \ y]^T$  representa as coordenadas da primeira imagem e  $\mathbf{x}' = [x' \ y']^T$  as coordenadas da segunda.

$$\mathbf{x}' = m(\mathbf{x}, \theta) \Leftrightarrow \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} \quad (3.1)$$

Ora se o movimento a estimar causa uma rotação então não faz sentido recorrer a uma pirâmide tradicional uma vez que neste caso a função de erro fica mais suave para movimentos de  $x$  e de  $y$  mas não para movimentos em termos de  $\theta$ .

Para tornar a função verdadeiramente suave em  $\theta$  propomos uma transformação em  $\mathbf{x}$  e  $\mathbf{x}'$  da forma:

$$\begin{aligned} \mathbf{p} = \mathbf{T}(\mathbf{x}) \Leftrightarrow \begin{bmatrix} r_1 \\ \theta_1 \end{bmatrix} &= \begin{bmatrix} \sqrt{x^2 + y^2} \\ \arctan\left(\frac{y}{x}\right) \end{bmatrix} & \mathbf{x} = \mathbf{T}^{-1}(\mathbf{p}) \Leftrightarrow \begin{bmatrix} x \\ y \end{bmatrix} &= \begin{bmatrix} r_1 \cos(\theta_1) \\ r_1 \sin(\theta_1) \end{bmatrix} \\ \mathbf{p}' = \mathbf{T}(\mathbf{x}') \Leftrightarrow \begin{bmatrix} r_2 \\ \theta_2 \end{bmatrix} &= \begin{bmatrix} \sqrt{x'^2 + y'^2} \\ \arctan\left(\frac{y'}{x'}\right) \end{bmatrix} & \mathbf{x}' = \mathbf{T}^{-1}(\mathbf{p}') \Leftrightarrow \begin{bmatrix} x' \\ y' \end{bmatrix} &= \begin{bmatrix} r_2 \cos(\theta_2) \\ r_2 \sin(\theta_2) \end{bmatrix} \end{aligned} \quad (3.2)$$

Desta forma cada imagem fica representada em termos de coordenadas polares,  $\mathbf{p} = [r_1 \ \theta_1]^T$  e  $\mathbf{p}' = [r_2 \ \theta_2]^T$  para a primeira e segunda imagens respectivamente.

O que é interessante agora é que nestas coordenadas o efeito de uma rotação de  $\theta$  fica equivalente ao efeito de uma translação pura. Podemos observar este facto se substituirmos (3.2)

<sup>1</sup>O eixo óptico corresponde à recta que liga o centro de projecção da câmara ao centro da imagem. Para mais detalhes sobre geometria nos problemas de visão consultar [7] ou [18]

em (3.1):

$$\begin{aligned}
\mathbf{x}' = m(\mathbf{x}, \theta) &\Leftrightarrow \mathbf{T}(\mathbf{x}') = \mathbf{T}(m(\mathbf{x}, \theta)) \\
&\Leftrightarrow \mathbf{p}' = \mathbf{T}(m(\mathbf{T}^{-1}(\mathbf{p}), \theta)) \\
&\Leftrightarrow \begin{bmatrix} r_2 \\ \theta_2 \end{bmatrix} = \mathbf{T} \left( \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \cdot \begin{bmatrix} r_1 \cos(\theta_1) \\ r_1 \sin(\theta_1) \end{bmatrix} \right) \\
&\Leftrightarrow \begin{bmatrix} r_2 \\ \theta_2 \end{bmatrix} = \mathbf{T} \left( r_1 \begin{bmatrix} \cos(\theta) \cos(\theta_1) - \sin(\theta) \sin(\theta_1) \\ \sin(\theta) \cos(\theta_1) + \cos(\theta) \sin(\theta_1) \end{bmatrix} \right) \\
&\Leftrightarrow \begin{bmatrix} r_2 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} r_1 \sqrt{[\cos(\theta) \cos(\theta_1) - \sin(\theta) \sin(\theta_1)]^2 + [\sin(\theta) \cos(\theta_1) + \cos(\theta) \sin(\theta_1)]^2} \\ \arctan \left( \frac{\sin(\theta) \cos(\theta_1) + \cos(\theta) \sin(\theta_1)}{\cos(\theta) \cos(\theta_1) - \sin(\theta) \sin(\theta_1)} \right) \end{bmatrix}
\end{aligned} \tag{3.3}$$

Tendo em conta as igualdades trigonométricas:

$$\begin{aligned}
\cos^2(z) + \sin^2(z) &= 1, & \forall z \\
\cos(y + z) &= \cos(y) \cos(z) - \sin(y) \sin(z), \\
\sin(y + z) &= \sin(y) \cos(z) + \sin(y) \sin(z), & \forall y, z
\end{aligned} \tag{3.4}$$

Verificamos que (3.3) é equivalente a:

$$\begin{aligned}
\mathbf{x}' = m(\mathbf{x}, \theta) &\Leftrightarrow \begin{bmatrix} r_2 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} r_1 \sqrt{1} \\ \arctan \left( \frac{\sin(\theta + \theta_1)}{\cos(\theta + \theta_1)} \right) \end{bmatrix} \\
&\Leftrightarrow \begin{bmatrix} r_2 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} r_1 \\ \theta_1 \end{bmatrix} + \begin{bmatrix} 0 \\ \theta \end{bmatrix} \\
&\Leftrightarrow \mathbf{p}' = \mathbf{p} + \begin{bmatrix} 0 \\ \theta \end{bmatrix}
\end{aligned} \tag{3.5}$$

Fica assim demonstrado que se se aplicar a transformação  $\mathbf{T}$  a ambas as imagens o movimento de rotação entre elas transforma-se num movimento de translação pura. Depois desta transformação baixamos a resolução da forma tradicional (isto é, substituindo cada região quadrada por um só *pixel* contendo a média de todos os *pixels* dentro dessa região) e finalmente invertemos a transformação  $\mathbf{T}$ . O resultado é um par de imagens para o qual a função de erro em termos da coordenada de rotação  $\theta$  é o mais suave possível o que quer dizer que é um par de imagens onde a estimação do movimento de rotação é a mais fácil possível.

## Construção da Pirâmide para um Movimento Genérico

O que foi visto para o caso da rotação pode agora ser generalizado para qualquer tipo de movimento. Para suavizar a função de erro em relação ao conjunto de parâmetros –  $\theta$  – tudo o que temos de fazer é descobrir qual é a transformação invertível  $\mathbf{T}$  que, quando aplicada às coordenadas de ambas as imagens, faz com que o movimento induzido pelos parâmetros  $\theta$  seja equivalente a um movimento de translação pura.

Uma vez conhecendo  $\mathbf{T}$  a pirâmide pode ser construída da mesma forma que uma pirâmide tradicional tendo apenas o cuidado de, antes de baixar a resolução de uma forma uniforme, aplicar a transformação de coordenadas  $\mathbf{T}$  e de depois de baixar a resolução aplicar a transformação inversa  $\mathbf{T}^{-1}$ .

Naturalmente a grande dificuldade deste procedimento encontra-se em encontrar a transformação de coordenadas  $\mathbf{T}$ . No entanto, é também necessário observar que este procedimento não pode ser aplicado para suavizar qualquer número de parâmetros. Uma vez que a imagem só tem duas variáveis independentes este procedimento só pode ser aplicado para suavizar dois parâmetros da transformação de coordenadas independentes. Apesar desta ser uma limitação forte existem modelos de movimento que a verificam. Um exemplo o modelo de *pan* e *tilt* que se aplica quando uma câmara tem o seu centro de projecção fixo e pode apenas rodar para a esquerda e direita (movimento de *pan*) e para cima e para baixo (movimento de *tilt*). Apesar deste modelo ter apenas dois graus de liberdade tem um grande interesse prático uma vez que modela os movimentos de uma câmara de vigilância típica (que não tenha possibilidade de fazer *zoom*).

## 3.5 Resultados Experimentais

### 3.5.1 Função de erro

Para demonstrar a superioridade do desempenho da pirâmide proposta face à pirâmide original vamos construir ambas para um par de imagens com rotação  $\theta = \pi$ . Para cada nível de ambas as pirâmides traçamos o erro entre duas imagens como função do movimento de rotação. O resultado encontra-se na fig. 3.3. Nesta figura observamos que a pirâmide proposta conduz a uma função de erro muito mais suave que a pirâmide tradicional. Este facto indica que, em princípio, a pirâmide proposta vai conduzir a melhores estimativas do que a pirâmide tradicional.

### 3.5.2 Alinhamento de duas imagens

A fig. 3.4 mostra um exemplo claro em que a pirâmide proposta sucede em criar a imagem panorâmica quando a pirâmide tradicional falha. Apresenta-se, para cada nível, as imagens de partida e a estimativa actual do panorama. Esta figura mostra que, pelo menos neste caso, a pirâmide proposta conduz a resultados superiores do que a pirâmide tradicional.



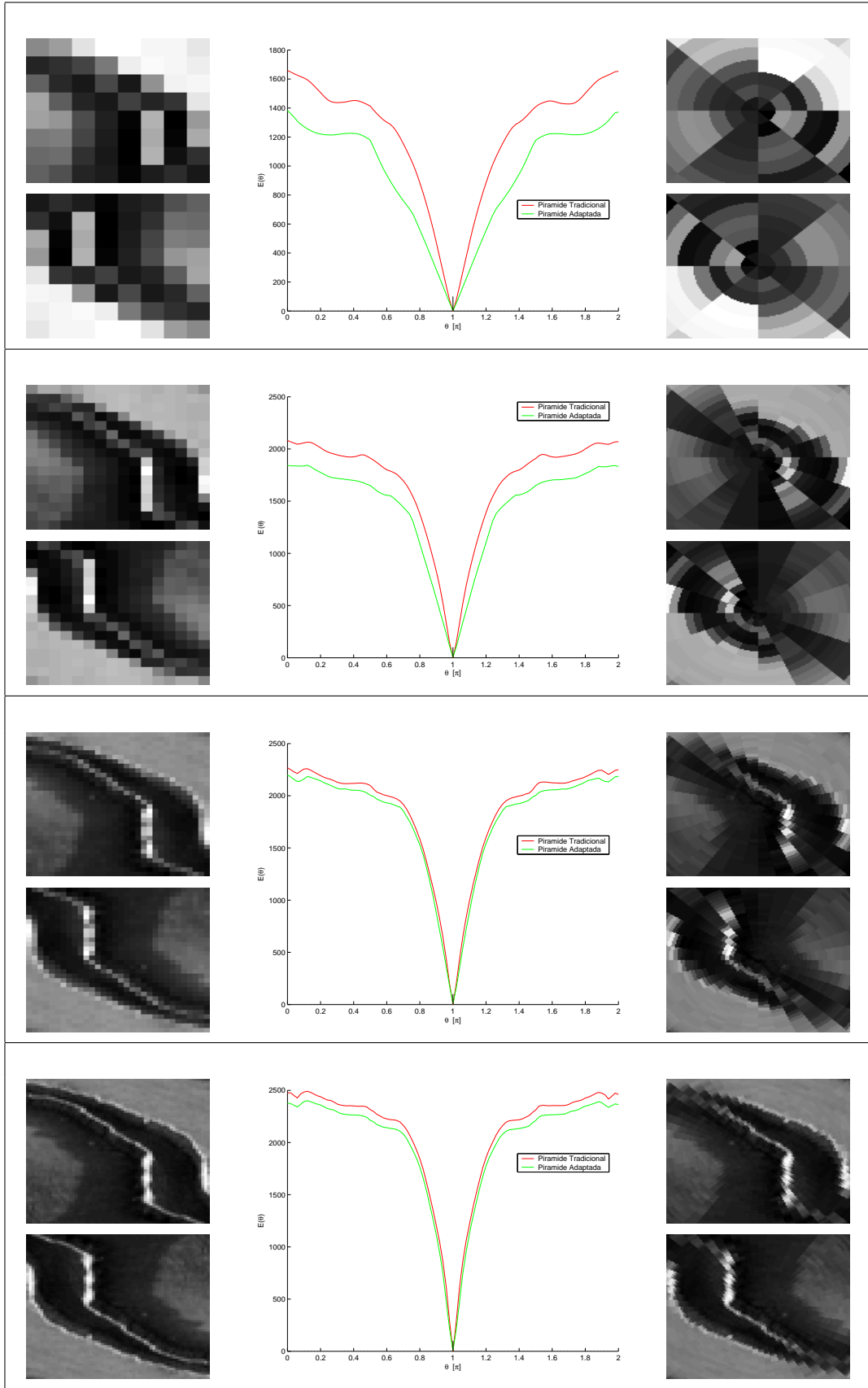


Figura 3.3: Comparação das funções de erro para a pirâmide tradicional e para a pirâmide proposta para movimento de rotação. Esquerda: Pirâmide Tradicional; Centro: Funções de Erro; Direita: Pirâmide Proposta

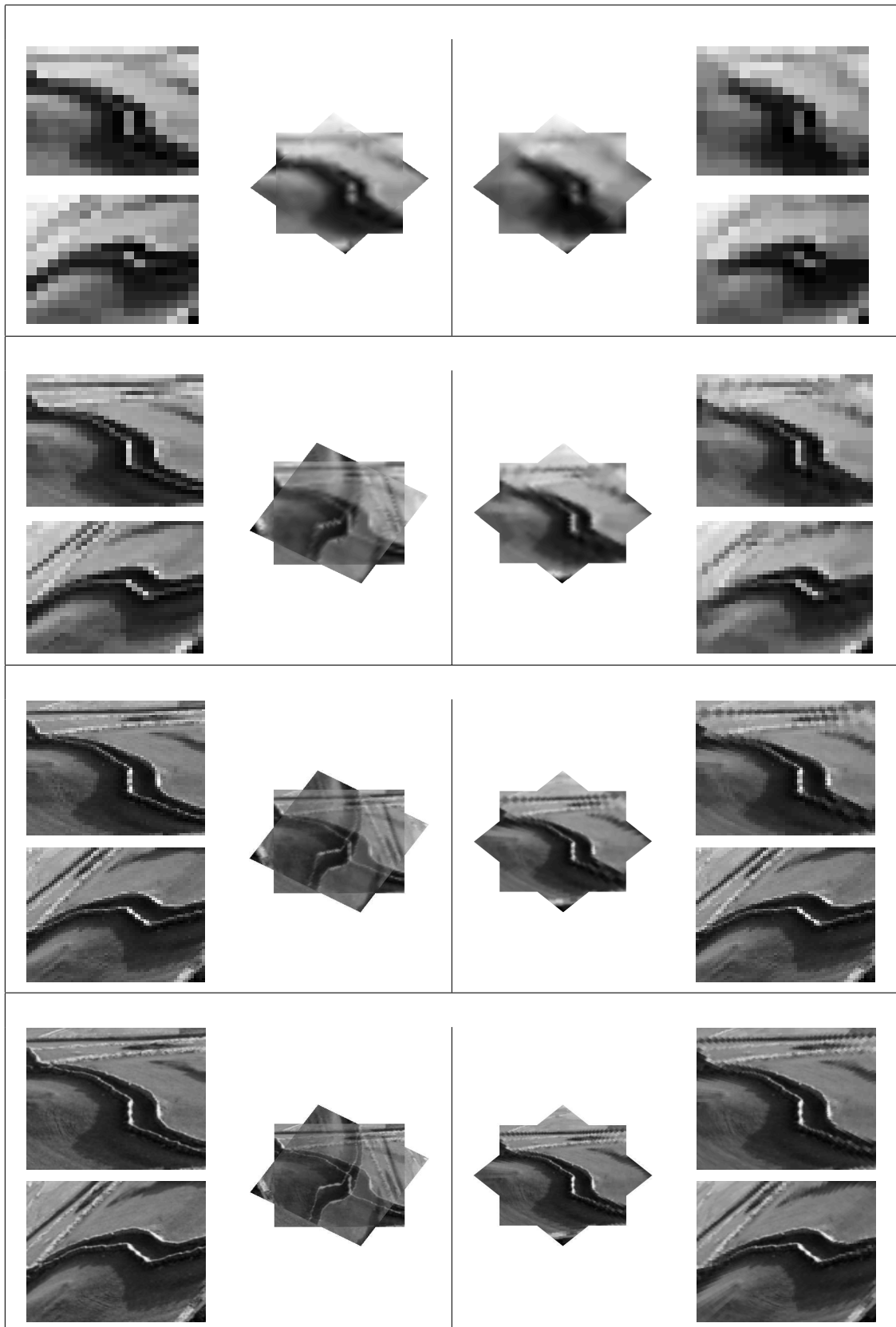


Figura 3.4: Comparação entre o alinhamento de duas imagens usando a pirâmide tradicional e usando a pirâmide proposta. Esquerda: Pirâmide tradicional e alinhamento usando essa pirâmide (alinhamento incorrecto); Direita: Pirâmide proposta e alinhamento usando essa pirâmide (alinhamento correcto)

# Capítulo 4

## Alinhamento de Múltiplas imagens: Estimação Global

No presente capítulo trata-se o problema do alinhamento de múltiplas imagens. À semelhança do que foi feito no segundo capítulo apresenta-se a formulação deste problema no contexto da teoria clássica da estimação e procede-se à sua resolução. Seguidamente expõe-se o algoritmo que permite encontrar a estimativa de máxima verosimilhança para o panorama. Por fim apresentam-se experiências em *MatLab* que provam a validade do algoritmo proposto e mostram que ele é mais resistente à propagação de erros do que os algoritmos tradicionais.

### 4.1 Alinhamento de Múltiplas Imagens como Problema de Estimação

#### 4.1.1 Notação

Começamos por estender a notação usada no capítulo 2. Associamos funções às imagens observadas –  $I^0(\mathbf{x}^0)$  a  $I^n(\mathbf{x}^n)$  – e ao panorama –  $P(\mathbf{x}^0)$ . As funções que representam cada uma das imagens estão definidas em domínios –  $\mathcal{D}^0$  a  $\mathcal{D}^n$  – definidos em sistemas de coordenadas diferentes –  $\mathbf{x}^0$  a  $\mathbf{x}^n$ . A definição seguinte clarifica estes conceitos.

**Definição 4.1** *Notação para o alinhamento de múltiplas imagens:*

$I^0(\mathbf{x}^0)$ : Padrão de intensidades luminosas da imagem de referência;

$I^1(\mathbf{x}^1)$  a  $I^n(\mathbf{x}^n)$ : Padrões de intensidades luminosas das restantes imagens;

$P(\mathbf{x}^0)$  : Padrão de intensidade luminosa do panorama;

$\mathbf{x}^0 = [x^0 \ y^0]^T$ : Sistema de coordenadas associado à imagem de referência (e, consequentemente, ao panorama);

$\mathbf{x}^1 = [x^1 \ y^1]^T$  a  $\mathbf{x}^n = [x^n \ y^n]^T$ : Sistemas de coordenadas associados às restantes imagens;

$\mathcal{D}^0$  a  $\mathcal{D}^n$ : Domínios onde os padrões de intensidade luminosa que formam as imagens estão definidos.

Note-se que, tal como no capítulo 2, definimos logo à partida uma imagem de referência que já está alinhada com a imagem panorâmica dado que as funções  $I^0(\mathbf{x}^0)$  e  $P(\mathbf{x}^0)$  partilham o mesmo sistema de coordenadas, dito de referência:  $\mathbf{x}^0 = [x^0 \ y^0]^T$ . Assim, o objectivo do alinhamento do conjunto de imagens é obter uma representação do panorama no sistema de coordenadas da imagem de referência. é importante frisar que qualquer imagem do conjunto pode ser escolhida para imagem de referência, sendo o panorama resultante o mesmo ainda que dado num sistema de coordenadas diferente (já que a alteração da imagem de referência implica a alteração do sistema de coordenadas de referência).

À semelhança do caso em que existem apenas duas imagens também aqui consideramos que o panorama se mantém constante e admitimos que existe apenas movimento tridimensional da câmara que registou as imagens. O movimento tridimensional da câmara induz um conjunto de alterações no padrão bidimensional de intensidades das imagens. Nos casos onde é possível criar panoramas estas alterações podem ser modeladas por uma transformação do sistema de coordenadas de cada imagem que depende apenas de um pequeno número de parâmetros [7].

**Definição 4.2** *O movimento 3D da câmara induz um movimento 2D no plano da imagem que pode ser modelado por uma **Transformação** entre coordenadas de cada imagem e as coordenadas da imagem de referência. Para o sistema de coordenadas de uma imagem genérica representado por  $\mathbf{x}^i$  esta relação de coordenadas pode ser escrita como:*

$$\mathbf{x}^i = m(\mathbf{x}^0, \boldsymbol{\theta}^i) , \quad (4.1)$$

onde  $\boldsymbol{\theta}^i$  representa o vector de parâmetros da transformação de coordenadas desta imagem. Esta transformação pode ser invertida obtendo-se:

$$\mathbf{x}^0 = m^{-1}(\mathbf{x}^i, \boldsymbol{\theta}^i) . \quad (4.2)$$

Nesta fase do trabalho a relação genérica entre as coordenadas de cada imagem e as coordenadas da imagem de referência apresentada acima é suficiente para formular o problema. Mais tarde este assunto será retomado quando se discutir a implementação prática do algoritmo de estimação de panoramas (ver secções 2.3.2 e 2.3.3), sendo ainda tratado no Anexo A.

### 4.1.2 Formulação do Problema

Para formular o problema do alinhamento de imagens segundo a teoria da estimação comecemos por considerar que cada uma das imagens representa um conjunto de amostras do panorama contaminadas por ruído. Assumimos que este ruído é gaussiano e independente de amostra para amostra (de *pixel* para *pixel*) e de imagem para imagem. Assim temos (onde  $R^0(\mathbf{x}^0)$  a  $R^n(\mathbf{x}^n)$  representam o ruído):

$$\begin{aligned} I^0(\mathbf{x}^0) &= P(\mathbf{x}^0) + R^0(\mathbf{x}^0) , \\ I^1(\mathbf{x}^1) &= P(m^{-1}(\mathbf{x}^1, \boldsymbol{\theta}^1)) + R^1(\mathbf{x}^1) , \\ &\dots \\ I^i(\mathbf{x}^i) &= P(m^{-1}(\mathbf{x}^i, \boldsymbol{\theta}^i)) + R^i(\mathbf{x}^i) , \\ &\dots \\ I^n(\mathbf{x}^n) &= P(m^{-1}(\mathbf{x}^n, \boldsymbol{\theta}^n)) + R^n(\mathbf{x}^n) . \end{aligned} \quad (4.3)$$

Note-se que, tal como foi introduzido na definição 4.1, a função  $P(\mathbf{x}^0)$  descreve as intensidades luminosas do panorama no sistema de coordenadas da imagem de referência,  $\mathbf{x}^0$ . Ora para as imagens que não são de referência – genericamente a imagem  $I^i$  – a variável independente não é  $\mathbf{x}^0$  mas sim  $\mathbf{x}^i$ . Assim, para escrever  $P(\mathbf{x}^0)$  em função de  $\mathbf{x}^i$  na equação (4.3) recorreremos à equação (4.2) que permite obter  $\mathbf{x}^0$  a partir de  $\mathbf{x}^i$ .

Admitindo que cada uma das amostras de ruído é uma variável aleatória normal com média zero e variância  $\sigma^2$ , a qualquer *pixel*, tanto da imagem de referência como das restantes imagens é também uma variável aleatória normal agora com média igual ao valor real do panorama nesse ponto e variância igual à do ruído. Denotando um *pixel* genérico na imagem de referência como  $I^0(\mathbf{x}_j^0)$  e um pixel genérico numa qualquer das restantes imagens como  $I^i(\mathbf{x}_j^i)$  podemos escrever<sup>1</sup>:

$$\begin{aligned} R^0(\mathbf{x}_j^0) &\sim \mathcal{N}(0, \sigma^2) \Rightarrow I^0(\mathbf{x}_j^0) \sim \mathcal{N}(P(\mathbf{x}_j^0), \sigma^2) , \\ R^1(\mathbf{x}_j^1) &\sim \mathcal{N}(0, \sigma^2) \Rightarrow I^1(\mathbf{x}_j^1) \sim \mathcal{N}(P(m^{-1}(\mathbf{x}_j^1, \boldsymbol{\theta}^1)), \sigma^2) , \\ &\dots \\ R^i(\mathbf{x}_j^i) &\sim \mathcal{N}(0, \sigma^2) \Rightarrow I^i(\mathbf{x}_j^i) \sim \mathcal{N}(P(m^{-1}(\mathbf{x}_j^i, \boldsymbol{\theta}^i)), \sigma^2) , \\ &\dots \\ R^n(\mathbf{x}_j^n) &\sim \mathcal{N}(0, \sigma^2) \Rightarrow I^n(\mathbf{x}_j^n) \sim \mathcal{N}(P(m^{-1}(\mathbf{x}_j^n, \boldsymbol{\theta}^n)), \sigma^2) . \end{aligned} \tag{4.4}$$

### 4.1.3 Estimativa de Máxima Verosimilhança

Escrevendo a função densidade de probabilidade –  $\mathcal{P}(\cdot)$  – explicitamente para cada amostra de cada imagem obtemos:

$$\begin{aligned} \mathcal{P}(I^0(\mathbf{x}_j^0)|P(\mathbf{x}^0)) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ \frac{-1}{2\sigma^2} [I^0(\mathbf{x}_j^0) - P(\mathbf{x}_j^0)]^2 \right\} \\ \mathcal{P}(I^1(\mathbf{x}_j^1)|P(\mathbf{x}^0), \boldsymbol{\theta}^1) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ \frac{-1}{2\sigma^2} [I^1(\mathbf{x}_j^1) - P(m^{-1}(\mathbf{x}_j^1, \boldsymbol{\theta}^1))]^2 \right\} \\ &\dots \\ \mathcal{P}(I^i(\mathbf{x}_j^i)|P(\mathbf{x}^0), \boldsymbol{\theta}^i) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ \frac{-1}{2\sigma^2} [I^i(\mathbf{x}_j^i) - P(m^{-1}(\mathbf{x}_j^i, \boldsymbol{\theta}^i))]^2 \right\} \\ &\dots \\ \mathcal{P}(I^n(\mathbf{x}_j^n)|P(\mathbf{x}^0), \boldsymbol{\theta}^n) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ \frac{-1}{2\sigma^2} [I^n(\mathbf{x}_j^n) - P(m^{-1}(\mathbf{x}_j^n, \boldsymbol{\theta}^n))]^2 \right\} \end{aligned} \tag{4.5}$$

Sendo as amostras de ruído em cada *pixel* independentes entre si, os *pixels* da imagem também são independentes entre si. Assim a f.d.p. conjunta de todas os *pixels* que formam uma imagem é dada pelo produto das f.d.p. de cada *pixel*. Considerando que existem  $N$

---

<sup>1</sup>Para clarificar a notação podemos dizer  $I^i(\mathbf{x}_j^i)$  se refere ao *pixel*  $j$  na imagem  $i$

*pixels* em cada imagem temos:

$$\begin{aligned}
\mathcal{P}(I^0(\mathbf{x}^0)|P(\mathbf{x}^0)) &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp \left\{ \frac{-1}{2\sigma^2} \sum_{\mathbf{x}^0 \in \mathcal{D}^0} [I^0(\mathbf{x}^0) - P(\mathbf{x}^0)]^2 \right\} \\
\mathcal{P}(I^1(\mathbf{x}^1)|P(\mathbf{x}^0), \boldsymbol{\theta}^1) &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp \left\{ \frac{-1}{2\sigma^2} \sum_{\mathbf{x}^1 \in \mathcal{D}^1} [I^1(\mathbf{x}^1) - P(m^{-1}(\mathbf{x}^1, \boldsymbol{\theta}^1))]^2 \right\} \\
&\dots \\
\mathcal{P}(I^i(\mathbf{x}^i)|P(\mathbf{x}^0), \boldsymbol{\theta}^i) &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp \left\{ \frac{-1}{2\sigma^2} \sum_{\mathbf{x}^i \in \mathcal{D}^i} [I^i(\mathbf{x}^i) - P(m^{-1}(\mathbf{x}^i, \boldsymbol{\theta}^i))]^2 \right\} \\
&\dots \\
\mathcal{P}(I^n(\mathbf{x}^n)|P(\mathbf{x}^0), \boldsymbol{\theta}^n) &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp \left\{ \frac{-1}{2\sigma^2} \sum_{\mathbf{x}^n \in \mathcal{D}^n} [I^n(\mathbf{x}^n) - P(m^{-1}(\mathbf{x}^n, \boldsymbol{\theta}^n))]^2 \right\}
\end{aligned} \tag{4.6}$$

Para tornar mais compacta a escrita dos resultados seguintes introduzimos agora o vector de valores  $\boldsymbol{\theta}^0$  da seguinte forma:

$$\mathbf{x}^0 = m(\mathbf{x}^0, \boldsymbol{\theta}^0) . \tag{4.7}$$

Note-se que este novo vector,  $\boldsymbol{\theta}^0$ , não corresponde a nenhum novo estimador mas sim ao conjunto de valores dos parâmetros da transformação de coordenadas que fazem com que a aplicação desta transformação ao sistema de coordenadas  $\mathbf{x}^0$  não produza qualquer efeito (isto é, deixe o sistema de coordenadas  $\mathbf{x}^0$  inalterado).

Naturalmente a aplicação da transformação de coordenadas inversa ao sistema de coordenadas  $\mathbf{x}^0$  usando o vector  $\boldsymbol{\theta}^0$  agora introduzido também vai deixar o sistema de coordenadas  $\mathbf{x}^0$  inalterado. Isto é:

$$\mathbf{x}^0 = m^{-1}(\mathbf{x}^0, \boldsymbol{\theta}^0) . \tag{4.8}$$

Esta propriedade permite usar o vector  $\boldsymbol{\theta}^0$  para reescrever a primeira equação de (4.6) numa forma semelhante às restantes equações:

$$\mathcal{P}(I^0(\mathbf{x}^0)|P(\mathbf{x}^0)) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp \left\{ \frac{-1}{2\sigma^2} \sum_{\mathbf{x}^0 \in \mathcal{D}^0} [I^0(\mathbf{x}^0) - P(m^{-1}(\mathbf{x}^0, \boldsymbol{\theta}^0))]^2 \right\} \tag{4.9}$$

Tendo em conta que admitimos que o ruído é independente de imagem para imagem, podemos escrever a função de verosimilhança de todas as imagens. Esta função é apenas a f.d.p. conjunta das imagens condicionada ao panorama,  $P(\mathbf{x}^0)$ , e ao conjunto de parâmetros da transformação de coordenadas,  $\boldsymbol{\theta} = \{\boldsymbol{\theta}^1 \dots \boldsymbol{\theta}^i \dots \boldsymbol{\theta}^n\}$ .

$$\begin{aligned}
\mathcal{P}(\mathcal{I}|P(\mathbf{x}^0), \boldsymbol{\theta}) &= (2\pi\sigma^2)^{-(n+1)N/2} \\
&\exp \left\{ - (2\sigma^2)^{-1} \sum_{i=0}^n \sum_{\mathbf{x}^i \in \mathcal{D}^i} [I^i(\mathbf{x}^i) - P(m^{-1}(\mathbf{x}^i, \boldsymbol{\theta}^i))]^2 \right\}
\end{aligned} \tag{4.10}$$

onde  $\mathcal{I}$  representa o conjunto de todas as imagens nos seus respectivos sistemas de coordenadas:  $\mathcal{I} = \{I^0(\mathbf{x}^0), I^1(\mathbf{x}^1) \dots I^i(\mathbf{x}^i) \dots I^n(\mathbf{x}^n)\}$ .

Para simplificar este resultado escrevemos o segundo somatório em termos de  $\mathbf{x}^0$  usando as equações (4.1) e (4.7):

$$\begin{aligned}
\mathcal{P}(\mathcal{I}|P(\mathbf{x}^0), \boldsymbol{\theta}) &= (2\pi\sigma^2)^{-(n+1)N/2} \\
&\exp \left\{ - (2\sigma^2)^{-1} \sum_{i=0}^n \sum_{\mathbf{x}^i \in \mathcal{D}_0^i(\boldsymbol{\theta}^i)} [I^i(m(\mathbf{x}^0, \boldsymbol{\theta}^i)) - P(\mathbf{x}^0)]^2 \right\} ,
\end{aligned} \tag{4.11}$$

onde  $\mathcal{D}_0^i(\boldsymbol{\theta}^i)$  representa o conjunto de pontos no sistema de coordenadas de referência onde a imagem  $i$  foi amostrada. Isto é,

$$\mathcal{D}_0^i(\boldsymbol{\theta}^i) = \{\mathbf{x}^0 : \mathbf{x}^i \in \mathcal{D}^i \Leftrightarrow m(\mathbf{x}^0, \boldsymbol{\theta}^i) \in \mathcal{D}^i\} \quad (4.12)$$

Estendendo as funções que representam as imagens ( $I^0(\mathbf{x}^0)$  a  $I^n(\mathbf{x}^n)$ ) a todo o  $\mathbb{R}^2$  e definindo as funções:

$$h^i(\mathbf{x}^0, \boldsymbol{\theta}^i) = \begin{cases} 1 & \text{se } \mathbf{x}^0 \in \mathcal{D}_0^i(\boldsymbol{\theta}^i) \\ 0 & \text{caso contrário} \end{cases} \quad (4.13)$$

Podemos escrever o resultado:

**Resultado 4.3** *A Função de Verosimilhança para o problema do alinhamento de múltiplas imagens é dada por:*

$$\mathcal{P}(\mathcal{I}|P(\mathbf{x}^0), \boldsymbol{\theta}) = (2\pi\sigma^2)^{-(n+1)N/2} \exp \left\{ - (2\sigma^2)^{-1} \sum_{\mathbf{x}^0 \in \mathbb{R}^2} \sum_{i=0}^n [I^i(m(\mathbf{x}^0, \boldsymbol{\theta}^i)) - P(\mathbf{x}^0)]^2 h^i(\mathbf{x}^0, \boldsymbol{\theta}^i) \right\} , \quad (4.14)$$

Para a discussão que se segue é mais vantajoso usar o simétrico do logaritmo desta função:

$$\begin{aligned} -\ln(\mathcal{P}(\mathcal{I}|P(\mathbf{x}^0), \boldsymbol{\theta})) &= \frac{(n+1)N}{2} \ln(2\pi\sigma^2) \\ &+ (2\sigma^2)^{-1} \sum_{\mathbf{x}^0 \in \mathbb{R}^2} \sum_{i=0}^n [I^i(m(\mathbf{x}^0, \boldsymbol{\theta}^i)) - P(\mathbf{x}^0)]^2 h^i(\mathbf{x}^0, \boldsymbol{\theta}^i) \end{aligned} \quad (4.15)$$

A estimativa de máxima verosimilhança dos parâmetros desconhecidos – o panorama  $P(\mathbf{x}^0)$  e o conjunto de todos os parâmetros da transformação de coordenadas  $\boldsymbol{\theta}$  – é obtida maximizando a função de verosimilhança (ver [15], págs. 63 a 65) ou, de uma forma equivalente, minimizando o simétrico do logaritmo dessa função:

**Resultado 4.4** *Os estimadores de máxima verosimilhança para o problema do alinhamento de múltiplas imagens são dados por:*

$$\left( \widehat{P(\mathbf{x}^0)}, \widehat{\boldsymbol{\theta}} \right) = \arg \min_{(P(\mathbf{x}^0), \boldsymbol{\theta})} -\ln(\mathcal{P}(\mathcal{I}|P(\mathbf{x}^0), \boldsymbol{\theta})) , \quad (4.16)$$

onde  $-\ln(\mathcal{P}(\mathcal{I}|P(\mathbf{x}), \boldsymbol{\theta}))$  é definido pela equação (4.15)

Começamos por procurar os estimadores de máxima verosimilhança para o panorama. Derivando o simétrico do logaritmo da função de verosimilhança dado pela equação (4.15) em relação a uma amostra genérica ( $\mathbf{x}_j^0$ ) do panorama e igualando a zero temos:

$$\begin{aligned} \frac{\partial}{\partial P(\mathbf{x}_j^0)} [-\ln \mathcal{P}(\mathcal{I}|P(\mathbf{x}^0), \boldsymbol{\theta})] &= 0 \\ \Leftrightarrow \frac{1}{\sigma^2} \sum_{i=0}^n [I^i(m(\mathbf{x}_j^0, \boldsymbol{\theta}^i)) - P(\mathbf{x}_j^0)] h^i(\mathbf{x}_j^0, \boldsymbol{\theta}^i) &= 0 \\ \Leftrightarrow \widehat{P(\mathbf{x}_j^0)} &= \frac{\sum_{i=0}^n I^i(m(\mathbf{x}_j^0, \boldsymbol{\theta}^i)) h^i(\mathbf{x}_j^0, \boldsymbol{\theta}^i)}{\sum_{i=0}^n h^i(\mathbf{x}_j^0, \boldsymbol{\theta}^i)} \end{aligned} \quad (4.17)$$

Esta equação generaliza a equação (2.14) e mostra que a estimativa de máxima verosimilhança para cada um dos *pixels* do panorama é igual à média das observações de todas as imagens que registaram esse pixel no panorama. Uma vez que esta equação é válida para qualquer *pixel* do panorama chegamos ao resultado:

**Resultado 4.5** *A estimativa de máxima verosimilhança para o panorama pode ser obtida de uma forma imediata em função da estimativa dos parâmetros da transformação de coordenadas usando a equação:*

$$\widehat{P(\mathbf{x})} = \frac{\sum_{i=0}^n I^i(m(\mathbf{x}^0, \boldsymbol{\theta}^i)) h^i(\mathbf{x}^0, \boldsymbol{\theta}^i)}{\sum_{i=0}^n h^i(\mathbf{x}^0, \boldsymbol{\theta}^i)} \quad (4.18)$$

Este resultado prova que a estimativa do panorama pode ser obtida de uma forma trivial desde que se conheçam as estimativas dos parâmetros da transformação de coordenadas. Substituindo a equação (4.18) em (4.15) mostramos, após a manipulação descrita no Anexo C, que a função de verosimilhança no caso de múltiplas imagens pode ser escrita como:

$$\begin{aligned} -\ln \left( \mathcal{P} \left( \mathcal{I} | \widehat{P}, \boldsymbol{\theta} \right) \right) &= \frac{(n+1)N}{2} \ln(2\pi\sigma^2) + (4\sigma^2)^{-1} \times \\ &\sum_{\mathbf{x} \in \mathbb{R}^2} H^{-1}(\mathbf{x}^0, \boldsymbol{\theta}) \sum_{i,j=0}^n [I^i(m(\mathbf{x}^0, \boldsymbol{\theta}^i)) - I^j(m(\mathbf{x}^0, \boldsymbol{\theta}^j))]^2 h^i(\mathbf{x}^0, \boldsymbol{\theta}^i) h^j(\mathbf{x}^0, \boldsymbol{\theta}^j) \end{aligned} \quad (4.19)$$

Onde se define a função  $H(\mathbf{x}^0, \boldsymbol{\theta})$  como:

$$H(\mathbf{x}^0, \boldsymbol{\theta}) = \sum_{k=0}^n h^k(\mathbf{x}^0, \boldsymbol{\theta}^k) \quad (4.20)$$

Definimos agora o conjunto de intersecção as imagens genéricas  $i$  e  $j$  como o conjunto de pontos do sistema de referência  $\mathbf{x}^0$  onde tanto a imagem  $i$  como a imagem  $j$  foram amostradas. Naturalmente este conjunto corresponde à intersecção entre os domínios  $D_0^i(\boldsymbol{\theta}^i)$  e  $D_0^j(\boldsymbol{\theta}^j)$ , isto é:

$$\mathcal{R}_{i,j}(\boldsymbol{\theta}^i, \boldsymbol{\theta}^j) = \mathcal{D}_0^i(\boldsymbol{\theta}^i) \cap \mathcal{D}_0^j(\boldsymbol{\theta}^j) = \{ \mathbf{x}^0 : h^i(\mathbf{x}^0, \boldsymbol{\theta}^i) = h^j(\mathbf{x}^0, \boldsymbol{\theta}^j) = 1 \} \quad (4.21)$$

Recorrendo a esta nova notação podemos enunciar o seguinte resultado:

**Resultado 4.6** *As estimativas de máxima verosimilhança para os parâmetros da transformação de coordenadas no caso de múltiplas imagens são dadas por:*

$$\widehat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} E(\boldsymbol{\theta}), \quad E(\boldsymbol{\theta}) = \sum_{i,j=0}^n \sum_{\mathbf{x} \in \mathcal{R}_{i,j}(\boldsymbol{\theta}^i, \boldsymbol{\theta}^j)} \frac{e_{i,j}^2(\mathbf{x}^0, \boldsymbol{\theta}^i, \boldsymbol{\theta}^j)}{H(\mathbf{x}^0, \boldsymbol{\theta})} \quad (4.22)$$

Onde  $E(\boldsymbol{\theta})$  é a função de erro a minimizar sendo  $H(\mathbf{x}^0, \boldsymbol{\theta})$  como definido em (4.20) e  $e_{i,j}(\mathbf{x}^0, \boldsymbol{\theta}^i, \boldsymbol{\theta}^j)$  o erro entre as imagens  $i$  e  $j$  definido como:

$$e_{i,j}(\mathbf{x}^0, \boldsymbol{\theta}^i, \boldsymbol{\theta}^j) = I^i(m(\mathbf{x}^0, \boldsymbol{\theta}^i)) - I^j(m(\mathbf{x}^0, \boldsymbol{\theta}^j)) \quad (4.23)$$



## 4.2 Estimação do Movimento entre Múltiplas Imagens

### 4.2.1 Minimização da Função de Erro para Resolução Fixa

Para minimizar a função de erro definida no resultado 2.6 vamos utilizar um algoritmo iterativo. Dispostos no início de cada iteração de estimativas para os parâmetros do modelo de movimento de cada uma das imagens. Em cada iteração actualizamos apenas os parâmetros correspondentes a uma única das imagens. Por comodidade esta será a imagem  $w$  sendo o conjunto de parâmetros do seu modelo de movimento –  $\theta^w$  – estimado por minimização da função de erro derivada no resultado 4.6.

Antes de proceder à minimização vamos simplificar a função de erro referida de forma a tornar a sua minimização em relação a  $\theta^w$  mais simples. Para tal começamos por escrever esta função de erro de forma a evidenciar a sua dependência nos parâmetros da transformação de coordenadas correspondente à imagem  $w$ :

$$\begin{aligned} E(\theta) &= \sum_{i,j=0}^n \sum_{\mathbf{x} \in \mathcal{R}_{i,j}(\theta^i, \theta^j)} \frac{e_{i,j}^2(\mathbf{x}^0, \theta^i, \theta^j)}{H(\mathbf{x}^0, \theta)} \\ &= \sum_{i=0}^n \sum_{\mathbf{x} \in \mathcal{R}_{i,w}(\theta^i, \theta^w)} \frac{e_{i,w}^2(\mathbf{x}^0, \theta^i, \theta^w)}{H(\mathbf{x}^0, \theta)} + \sum_{j=0}^n \sum_{\mathbf{x} \in \mathcal{R}_{w,j}(\theta^w, \theta^j)} \frac{e_{w,j}^2(\mathbf{x}^0, \theta^w, \theta^j)}{H(\mathbf{x}^0, \theta)} + \\ &\quad + \sum_{\substack{i,j=0 \\ i,j \neq w}}^n \sum_{\mathbf{x} \in \mathcal{R}_{i,j}(\theta^i, \theta^j)} \frac{e_{i,j}^2(\mathbf{x}^0, \theta^i, \theta^j)}{H(\mathbf{x}^0, \theta)} \end{aligned} \quad (4.24)$$

Tendo em conta que para todos os valores do índice mudo  $a$  temos:

$$e_{a,w}^2(\mathbf{x}^0, \theta^a, \theta^w) = e_{w,a}^2(\mathbf{x}^0, \theta^w, \theta^a) \quad (4.25)$$

e

$$\mathcal{R}_{a,w}(\theta^a, \theta^w) = \mathcal{R}_{w,a}(\theta^w, \theta^a) \quad (4.26)$$

podemos condensar os dois primeiros somatórios em 4.24 para obter:

$$E(\theta) = 2 \sum_{i=0}^n \sum_{\mathbf{x} \in \mathcal{R}_{i,w}(\theta^i, \theta^w)} \frac{e_{i,w}^2(\mathbf{x}^0, \theta^i, \theta^w)}{H(\mathbf{x}^0, \theta)} + \sum_{\substack{i,j=0 \\ i,j \neq w}}^n \sum_{\mathbf{x} \in \mathcal{R}_{i,j}(\theta^i, \theta^j)} \frac{e_{i,j}^2(\mathbf{x}^0, \theta^i, \theta^j)}{H(\mathbf{x}^0, \theta)} \quad (4.27)$$

Observamos agora que o segundo somatório não apresenta qualquer dependência em  $\theta^w$  pelo que podemos dizer que a minimização em  $\theta^w$  da função de erro  $E(\theta)$  é completamente equivalente à minimização da função de erro  $E^w(\theta)$  definida como:

$$E^w(\theta) = 2 \sum_{i=0}^n \sum_{\mathbf{x} \in \mathcal{R}_{i,w}(\theta^i, \theta^w)} \frac{e_{i,w}^2(\mathbf{x}^0, \theta^i, \theta^w)}{H(\mathbf{x}^0, \theta)} \quad (4.28)$$

Tal como já foi dito no início desta secção existe em cada iteração uma estimativa inicial para os parâmetros do modelo de movimento correspondentes a cada uma das imagens. Estas estimativas são referidas por  $\theta_0^1$  a  $\theta_0^n$ . A estimativa final para estes parâmetros – referida

por  $\theta^1$  a  $\theta^n$  – será igual à estimativa inicial para os parâmetros correspondentes a todas as imagens excepto a imagem de ordem  $w$  para a qual temos:

$$\theta^w = \theta_0^w + \delta^w \quad (4.29)$$

onde  $\delta^w$  corresponde ao vector de actualização dos parâmetros do modelo de movimento da imagem  $w$ . Naturalmente este vector será estimado por minimização da função de erro  $E(\theta)$  introduzida no resultado 4.6. Esta minimização (em  $E(\theta)$ ) é equivalente à minimização de  $E^w(\theta)$  em (4.28) pelo que obtemos:

$$\widehat{\delta^w} = \arg \min_{\delta^w} \sum_{i=0}^n \sum_{\mathbf{x} \in \mathcal{R}_{i,w}(\theta_0^i, \theta_0^w)} \frac{e_{i,w}^2(\mathbf{x}^0, \theta_0^i, \theta_0^w + \delta^w)}{H(\mathbf{x}^0, \theta_0)} \quad (4.30)$$

Sabendo que  $\theta^w = \theta_0^w + \delta^w$  e admitindo que os valores de  $\delta^w$  são pequenos podemos expandir as funções de erro  $e_{i,w}(\mathbf{x}^0, \theta_0^i, \theta_0^w + \delta^w)$  em série de Taylor de primeira ordem:

$$e_{i,w}(\mathbf{x}^0, \theta_0^i, \theta_0^w + \delta^w) \simeq e_{i,w}(\mathbf{x}^0, \theta_0^i, \theta_0^w) + \delta^{wT} \cdot \nabla_{\theta^w} e_{i,w}(\mathbf{x}^0, \theta_0^i, \theta_0^w) \quad (4.31)$$

onde  $\nabla_{\theta^w} e_{i,w}(\mathbf{x}^0, \theta_0^i, \theta_0^w)$  corresponde ao gradiente do erro em relação aos parâmetros do modelo de movimento da imagem  $w$ . é importante notar desde já que este gradiente não apresenta qualquer dependência nos parâmetros da transformação de coordenadas correspondente a qualquer uma das outras imagens (aqui referidas com recurso à imagem genérica  $i$ ) tal como se pode ver de:

$$\begin{aligned} \nabla_{\theta^w} e_{i,w}(\mathbf{x}^0, \theta_0^i, \theta_0^w) &= \nabla_{\theta^w} [I^i(m(\mathbf{x}^0, \theta_0^i)) - I^w(m(\mathbf{x}^0, \theta_0^w))] \\ &= -\nabla_{\theta^w} I^w(m(\mathbf{x}^0, \theta_0^w)) \\ &= -\nabla_{\theta^w} m(\mathbf{x}, \theta_0^w) \cdot \nabla_{\mathbf{x}} I'(\mathbf{x}) = \nabla_{\theta^w} e_w(\mathbf{x}^0, \theta_0^w) \end{aligned} \quad (4.32)$$

onde a última simplificação se obteve com recurso à regra da cadeia.

Aplicando à equação (4.30) a aproximação apresentada na equação (4.31) e retomando o uso das funções da forma  $h^i(\mathbf{x}^0, \theta^i)$  introduzidas em (4.13) por substituição do uso dos conjuntos da forma  $\mathcal{R}_{i,j}(\theta^i, \theta^j)$  definidos em (4.21) permite escrever:

$$\widehat{\delta^w} = \arg \min_{\delta^w} \sum_{i=0}^n \sum_{\mathbf{x} \in \mathbb{R}^2} \frac{[e_{i,w}(\mathbf{x}^0, \theta_0^i, \theta_0^w) + \delta^{wT} \cdot \nabla_{\theta^w} e_w(\mathbf{x}^0, \theta_0^w)]^2 h^i(\mathbf{x}^0, \theta^i) h^w(\mathbf{x}^0, \theta^w)}{H(\mathbf{x}^0, \theta_0)} \quad (4.33)$$

Aplicando o gradiente em  $\delta^w$  à expressão a minimizar e igualando a zero obtemos:

$$\sum_{i=0}^n \sum_{\mathbf{x} \in \mathbb{R}^2} \frac{2 \nabla_{\theta^w} e_w [e_{i,w} + \delta^{wT} \cdot \nabla_{\theta^w} e_w] h^i h^w}{H} = 0 \quad (4.34)$$

onde se omitiu a dependência de  $e_{i,w}$  em  $\mathbf{x}^0$ ,  $\theta_0^i$  e  $\theta_0^w$ , de  $\nabla_{\theta^w} e_w$  em  $\mathbf{x}^0$  e  $\theta_0^w$ , de  $h^i$  em  $\mathbf{x}^0$  e  $\theta^i$ , de  $h^w$  em  $\mathbf{x}^0$  e  $\theta^w$  e de  $H$  em  $\mathbf{x}^0$  e  $\theta_0$  para obter um resultado mais compacto.

Transpondo ambos os termos da equação e resolvendo o produto obtemos:

$$\sum_{i=0}^n \sum_{\mathbf{x} \in \mathbb{R}^2} \frac{[e_{i,w} + \delta^{wT} \cdot \nabla_{\theta^w} e_w] \nabla_{\theta^w}^T e_w h^i h^w}{H} = 0 \quad (4.35)$$

$$\sum_{i=0}^n \sum_{\mathbf{x} \in \mathbb{R}^2} \frac{e_{i,w} \nabla_{\boldsymbol{\theta}^w}^T e_w h^i h^w}{H} + \sum_{i=0}^n \sum_{\mathbf{x} \in \mathbb{R}^2} \frac{\boldsymbol{\delta}^{wT} \cdot \nabla_{\boldsymbol{\theta}^w} e_w \cdot \nabla_{\boldsymbol{\theta}^w}^T e_w h^i h^w}{H} = \mathbf{0} \quad (4.36)$$

Explorando o facto de que  $\nabla_{\boldsymbol{\theta}^w}^T e_w$  e  $H$  não depende de  $\boldsymbol{\theta}^i$ , obtemos:

$$\sum_{\mathbf{x} \in \mathbb{R}^2} \nabla_{\boldsymbol{\theta}^w}^T e_w h^w \frac{\sum_{i=0}^n e_{i,w} h^i}{H} + \sum_{\mathbf{x} \in \mathbb{R}^2} \boldsymbol{\delta}^{wT} \cdot \nabla_{\boldsymbol{\theta}^w} e_w \cdot \nabla_{\boldsymbol{\theta}^w}^T e_w h^w \frac{\sum_{i=0}^n h^i}{H} = \mathbf{0} \quad (4.37)$$

Usando a definição de  $e_{i,w}$  apresentada em (4.23) e a definição de  $H$  apresentada em (4.20) podemos escrever:

$$\sum_{\mathbf{x} \in \mathbb{R}^2} \nabla_{\boldsymbol{\theta}^w}^T e_w h^w \frac{\sum_{i=0}^n (I^i - I^w) h^i}{H} + \sum_{\mathbf{x} \in \mathbb{R}^2} \boldsymbol{\delta}^{wT} \cdot \nabla_{\boldsymbol{\theta}^w} e_w \cdot \nabla_{\boldsymbol{\theta}^w}^T e_w h^w = \mathbf{0} \quad (4.38)$$

onde foram omitidas as dependências de  $I^i$  em  $m(\mathbf{x}^0, \boldsymbol{\theta}^i)$  e de  $I^w$  em  $m(\mathbf{x}^0, \boldsymbol{\theta}^w)$  para escrever o resultado de uma forma mais compacta.

Desdobrando o parêntesis e tendo em conta que  $I^w$  não apresenta qualquer dependência em  $\boldsymbol{\theta}^i$  obtemos:

$$\sum_{\mathbf{x} \in \mathbb{R}^2} \nabla_{\boldsymbol{\theta}^w}^T e_w h^w \left[ \frac{\sum_{i=0}^n I^i h^i}{H} - I^w \frac{\sum_{i=0}^n h^i}{H} \right] + \sum_{\mathbf{x} \in \mathbb{R}^2} \boldsymbol{\delta}^{wT} \cdot \nabla_{\boldsymbol{\theta}^w} e_w \cdot \nabla_{\boldsymbol{\theta}^w}^T e_w h^w = \mathbf{0} \quad (4.39)$$

Usando novamente a definição de  $H$  apresentada em (4.20), as definições de  $h^w$  e  $\mathcal{D}_0^i(\boldsymbol{\theta}^i)$  apresentadas respectivamente em (4.12) e (4.13) e ainda a definição de  $\widehat{P(\mathbf{x}^0)}$  introduzida no resultado 4.5 obtemos finalmente:

$$\sum_{\mathbf{x} \in \mathcal{D}_0^w} \nabla_{\boldsymbol{\theta}^w}^T e_w [P(\mathbf{x}^0) - I^w] + \boldsymbol{\delta}^{wT} \cdot \left( \sum_{\mathbf{x} \in \mathcal{D}_0^w} \nabla_{\boldsymbol{\theta}^w} e_w \cdot \nabla_{\boldsymbol{\theta}^w}^T e_w \right) = \mathbf{0} \quad (4.40)$$

onde foi omitida ainda a dependência de  $P$  em  $\mathbf{x}^0$  e de  $\mathcal{D}_0^w$  em  $\boldsymbol{\theta}^w$ .

Reordenando os termos e repondo a notação obtemos o resultado:

**Resultado 4.7** *A estimativa para a actualização dos parâmetros da transformação de coordenadas  $\widehat{\boldsymbol{\delta}}$  no caso de múltiplas imagens pode ser obtida por resolução do sistema linear:*

$$\begin{aligned} \sum_{\mathbf{x} \in \mathcal{D}_0^w(\boldsymbol{\theta}_0^w)} \nabla_{\boldsymbol{\theta}^w}^T e_w(\mathbf{x}^0, \boldsymbol{\theta}_0^w) \left[ \widehat{P(\mathbf{x}^0)} - I^w(m(\mathbf{x}^0, \boldsymbol{\theta}^w)) \right] + \\ + \boldsymbol{\delta}^{wT} \cdot \left( \sum_{\mathbf{x} \in \mathcal{D}_0^w} \nabla_{\boldsymbol{\theta}^w} e_w(\mathbf{x}^0, \boldsymbol{\theta}_0^w) \cdot \nabla_{\boldsymbol{\theta}^w}^T e_w(\mathbf{x}^0, \boldsymbol{\theta}_0^w) \right) = \mathbf{0} \end{aligned} \quad (4.41)$$

Onde  $\widehat{P(\mathbf{x}^0)}$  foi calculado no resultado 4.5 e  $\nabla_{\boldsymbol{\theta}^w}^T e_w(\mathbf{x}^0, \boldsymbol{\theta}_0^w)$  pode ser calculado a partir de (4.32)

## 4.2.2 Interpretação do Resultado: Algoritmo Global de Estimação

Apesar de ter sido necessária uma extensa dedução para chegar ao resultado apresentado este revela-se extraordinariamente simples. Ao compararmos este resultado com o resultado equivalente obtido no capítulo 2 (o resultado 2.7) observamos que as equações são equivalentes desde que se faça:

$$e_w(x^0, \theta) = \widehat{P(\mathbf{x}^0)} - I^w(m(\mathbf{x}^0, \theta^w)) . \quad (4.42)$$

Isto quer dizer que o alinhamento de uma só imagem no conjunto de todas as imagens pode ser visto como o alinhamento dessa imagem na estimativa actual do panorama dada por  $\widehat{P(\mathbf{x}^0)}$ .

Este resultado permite refinar em qualquer altura a estimativa do panorama apenas seleccionando uma imagem e alinhando-a novamente com a estimativa actual do panorama. Esta constatação serve de base ao algoritmo de alinhamento global que agora propomos:

**Resultado 4.8 *Algoritmo de Alinhamento Global de um conjunto de múltiplas imagens:***

1. *Fazer um alinhamento sucessivo e cumulativo de todas as imagens (este é um passo de inicialização, na prática corresponde ao algoritmo tradicional);*
2. *Percorrer uma a uma e ciclicamente todas as imagens refinando os parâmetros do seu modelo de movimento através do uso do resultado 4.32; Parar quando o erro entre cada imagem e o panorama passar abaixo de um valor limite ou quando se atingir um número máximo de iterações.*

Observamos assim que a contribuição original deste capítulo consiste em dois pontos fundamentais:

- Uma fórmula (Resultado 4.5) precisa para calcular a estimativa do panorama em qualquer altura;
- Uma forma precisa de refinar a estimativa do panorama por alinhamento de qualquer uma das imagens no panorama global (Resultado 4.7)

## 4.2.3 Algoritmo de Multi-Resolução

À semelhança do que foi dito no Capítulo 2 também aqui o método de minimização da função de erro baseia-se numa aproximação de Taylor de primeira ordem para a variação da diferença entre a imagem e a estimativa actual do panorama como função dos parâmetros do modelo de movimento. Mais uma vez como já foi dito esta aproximação tem uma validade apenas local o que faz com que o algoritmo apresente uma grande sensibilidade à existência de mínimos locais.

Para tornar a estimação dos parâmetros do modelo de movimento mais robusta aos mínimos locais vamos usar um algoritmo de multi-resolução. Para manter a simplicidade e focar apenas no que foi apresentado usamos a versão tradicional deste algoritmo presente na literatura [12] e descrita em mais pormenor na secção 3.3.

## 4.3 Resultados Experimentais

Para mostrar os resultados experimentais vamos acompanhar a construção de um panorama global a partir do conjunto de sete imagens que se apresenta na fig. 4.1. Estas imagens foram criadas "recortando" uma imagem real de forma a garantir que se alinhavam correctamente uma vez que o objectivo desta secção é expor apenas o funcionamento do algoritmo global.



Figura 4.1: Conjunto de sete imagens que se pretende alinhar utilizando o Algoritmo de Alinhamento Global (estas imagens estão ordenadas da esquerda para a direita e de cima para baixo)

Tal como é indicado no Algoritmo Global começamos por proceder ao encaixe cumulativo de todas as imagens. Este procedimento apresenta-se na fig. 4.2

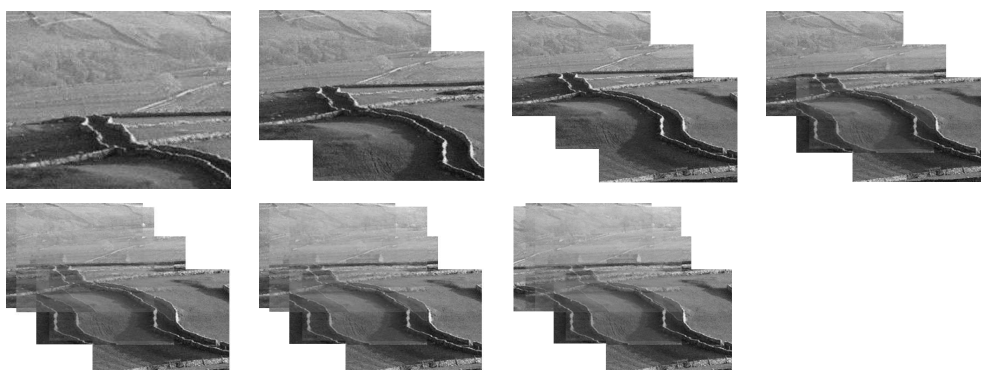


Figura 4.2: Primeiro Passo do Algoritmo Global: Construção cumulativa de um panorama com sete imagens (ordem da construção é da esquerda para a direita e de cima para baixo). Note-se que a quinta imagem é alinhada incorrectamente o que faz com que todas as imagens seguintes também sejam alinhadas erradamente.

A este panorama é aplicado o segundo passo do Algoritmo Global: a posição de cada imagem é ajustada até que se obtém a imagem da fig. 4.3



Figura 4.3: Resultado Final do Algoritmo Global: Construção correcta do panorama com sete imagens.

# Capítulo 5

## Resultados Experimentais

Neste capítulo incluímos alguns panoramas globais produzidos com recurso a uma combinação das técnicas utilizadas.



Figura 5.1: Panorama resultante do alinhamento de quatro imagens

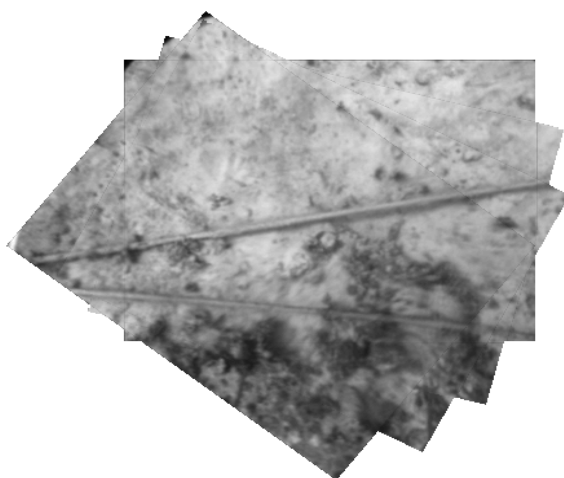


Figura 5.2: Panorama resultante do alinhamento de quatro imagens



Figura 5.3: Panorama resultante do alinhamento de nove imagens



Figura 5.4: Panorama resultante do alinhamento de cinco imagens



# Capítulo 6

## Conclusão

### 6.1 Sumário

Neste trabalho foi estudado em pormenor o problema da construção de imagens panorâmicas.

Foram desenvolvidas quatro contribuições originais sendo que todas elas avançaram no sentido de uma melhor solução do problema.

A primeira contribuição, a Janela Adaptativa, permitiu aumentar grandemente a robustez do algoritmo principalmente nas situações onde o algoritmo era mais fraco: as situações onde existe pouca sobreposição entre as imagens.

A segunda contribuição, Pirâmide dependente do modelo de movimento permitiu obter um maior conhecimento sobre o funcionamento do algoritmo e quais as áreas onde pode ser melhorado. Esta contribuição permitiu ainda desenvolver uma pirâmide óptima para o modelo de movimento de rotação.

A terceira contribuição, Estimação Global, permitiu aumentar grandemente a qualidade dos resultados uma vez que oferece uma forma pouco pesada computacionalmente de garantir a não existência de propagação de erros.

Por fim, a quarta contribuição, Formulação Clássica, permitiu obter um conhecimento mais profundo das bases por detrás do problema possibilitando a primeira e terceira contribuições.

### 6.2 Trabalho Futuro

O estudo intenso do problema em questão permitiu abrir algumas áreas que merecem interesse científico.

A primeira destas áreas prende-se com o estudo mais detalhado dos algoritmos de multi-resolução (tema do terceiro capítulo do trabalho). Nesta área permanecem em aberto os problemas de encontrar pirâmides adaptadas a outro tipo de movimento que não o considerado, nomeadamente o problema de encontrar uma pirâmide adaptada ao movimento de *pan* e *tilt* já discutido. Nesta área existe ainda o problema de seleccionar os parâmetros óptimos a suavizar quando o modelo de movimento tem mais do que dois parâmetros.

Um outro problema que ficou em aberto é a possibilidade de, no segundo passo do Algoritmo Global, escolher de uma forma mais eficiente quais as imagens cujos parâmetros se pretende actualizar em primeiro lugar. Talvez um bom critério seja o erro dessa imagem em

relação ao panorama mas é necessário fazer um estudo mais profundo para poder chegar a uma conclusão.

# Apêndice A

## Modelos do Movimento da Câmara

Note-se que, no mundo real onde os objectos retratados pela câmara têm formas tridimensionais, só faz sentido construir panoramas quando não se altera a posição do centro óptico da câmara. Esta restrição deve-se ao facto de que alterações na posição do centro óptico da câmara num mundo tridimensional fazem com que as imagens mostrem diferentes perspectivas dos objectos. Assim não faz sentido construir uma imagem panorâmica porque esta imagem (à semelhança de qualquer outra imagem bidimensional) só pode representar informação sobre uma única perspectiva.

Esta restrição ao movimento da câmara pode ser relaxada se se considerar que o mundo representado nas imagem é planar (i.e. bidimensional). Nesta situação os objectos pertencentes ao mundo têm apenas uma perspectiva pelo que podem ser retratados por uma imagem panorâmica mesmo que exista movimento tridimensional do centro óptico da câmara.

### A.1 Modelo de Movimento de Translação

O modelo de movimento de translação postula que as coordenadas das imagens se relacionam pela fórmula:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} . \quad (\text{A.1})$$

### A.2 Modelo de Movimento Afim

O modelo de movimento afim postula que as coordenadas das imagens se relacionam pela fórmula:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} . \quad (\text{A.2})$$

### A.3 Modelo de Movimento de Rotação

O modelo de movimento de rotação postula que as coordenadas das imagens se relacionam pela fórmula:

$$\mathbf{x}' = m(\mathbf{x}, \theta) \quad \Leftrightarrow \quad \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} . \quad (\text{A.3})$$

## Apêndice B

### Dedução da equação (2.16) a partir das equações (2.15) e (2.12)

Tal como foi visto na equação (2.12) na página 13 o simétrico do logaritmo da função de verosimilhança é dado por:

$$-\ln (\mathcal{P}(I(\mathbf{x}), I'(\mathbf{x}')|P(\mathbf{x}), \boldsymbol{\theta})) = N \ln (2\pi\sigma^2) + (2\sigma^2)^{-1} \sum_{\mathbf{x} \in \mathbb{R}^2} \left( [I(\mathbf{x}) - P(\mathbf{x})]^2 h(\mathbf{x}) + [I'(m(\mathbf{x}, \boldsymbol{\theta})) - P(\mathbf{x})]^2 h'(\mathbf{x}, \boldsymbol{\theta}) \right) \quad (\text{B.1})$$

Por outro lado na página 13 mostramos que a estimativa de máxima verosimilhança para o panorama pode ser obtida imediatamente a partir da estimativa dos parâmetros da transformação de coordenadas recorrendo à equação (2.15):

$$\widehat{P(\mathbf{x})} = \frac{I(\mathbf{x})h(\mathbf{x}) + I'(m(\mathbf{x}, \boldsymbol{\theta}))h'(\mathbf{x}, \boldsymbol{\theta})}{h(\mathbf{x}) + h'(\mathbf{x}, \boldsymbol{\theta})} \quad (\text{B.2})$$

Para obter a equação (2.16) a partir destas duas equações substituímos a estimativa do panorama,  $\widehat{P(\mathbf{x})}$ , no simétrico do logaritmo da função de verosimilhança:

$$-\ln \left( \mathcal{P} \left( I, I' | \widehat{P}, \boldsymbol{\theta} \right) \right) = N \ln (2\pi\sigma^2) + (2\sigma^2)^{-1} \times \sum_{\mathbf{x} \in \mathbb{R}^2} \left( \left[ I - \frac{I \cdot h + I' \cdot h'}{h + h'} \right]^2 \cdot h + \left[ I' - \frac{I \cdot h + I' \cdot h'}{h + h'} \right]^2 \cdot h' \right) \quad (\text{B.3})$$

onde, para obter um resultado mais compacto, omitimos a dependência de  $I$  e  $h$  em  $\mathbf{x}$ , a dependência de  $I'$  em  $m(\mathbf{x}, \boldsymbol{\theta})$  e a dependência de  $h'$  em  $\mathbf{x}$  e  $\boldsymbol{\theta}$ .

A partir desta equação escrevemos:

$$-\ln \left( \mathcal{P} \left( I, I' | \widehat{P}, \boldsymbol{\theta} \right) \right) = N \ln (2\pi\sigma^2) + (2\sigma^2)^{-1} \times \sum_{\mathbf{x} \in \mathbb{R}^2} \left( \left[ \frac{I \cdot (h+h') - I \cdot h - I' \cdot h'}{h+h'} \right]^2 \cdot h + \left[ \frac{I' \cdot (h+h') - I \cdot h - I' \cdot h'}{h+h'} \right]^2 \cdot h' \right) \quad (\text{B.4})$$

$$\Leftrightarrow -\ln \left( \mathcal{P} \left( I, I' | \widehat{P}, \boldsymbol{\theta} \right) \right) = N \ln (2\pi\sigma^2) + (2\sigma^2)^{-1} \times \sum_{\mathbf{x} \in \mathbb{R}^2} \left( \left[ \frac{I \cdot h + I' \cdot h' - I \cdot h - I' \cdot h'}{h+h'} \right]^2 \cdot h + \left[ \frac{I' \cdot h + I' \cdot h' - I \cdot h - I' \cdot h'}{h+h'} \right]^2 \cdot h' \right) \quad (\text{B.5})$$

$$\Leftrightarrow -\ln \left( \mathcal{P} \left( I, I' | \widehat{P}, \boldsymbol{\theta} \right) \right) = N \ln (2\pi\sigma^2) + (2\sigma^2)^{-1} \times \quad (B.6)$$

$$\times \sum_{\mathbf{x} \in \mathbb{R}^2} \left( \left[ \frac{I \cdot h' - I' \cdot h'}{h + h'} \right]^2 \cdot h + \left[ \frac{I' \cdot h - I \cdot h}{h + h'} \right]^2 \cdot h' \right)$$

$$\Leftrightarrow -\ln \left( \mathcal{P} \left( I, I' | \widehat{P}, \boldsymbol{\theta} \right) \right) = N \ln (2\pi\sigma^2) + (2\sigma^2)^{-1} \times \quad (B.7)$$

$$\times \sum_{\mathbf{x} \in \mathbb{R}^2} \left( [I - I']^2 \cdot \frac{h'^2 \cdot h}{(h + h')^2} + [I - I']^2 \cdot \frac{h^2 \cdot h'}{(h + h')^2} \right)$$

A partir das definições de  $h$  e  $h'$  introduzidas na equação 2.10 da página 2.10 observamos que estas funções tomam apenas os valores 0 e 1 o que as torna invariantes à operação de elevar ao quadrado. Assim (B.8) fica:

$$-\ln \left( \mathcal{P} \left( I, I' | \widehat{P}, \boldsymbol{\theta} \right) \right) = N \ln (2\pi\sigma^2) + (2\sigma^2)^{-1} \sum_{\mathbf{x} \in \mathbb{R}^2} 2 [I - I']^2 \cdot \frac{h \cdot h'}{(h + h')^2} \quad (B.8)$$

Mais uma vez tendo em conta que  $h$  e  $h'$  tomam apenas valores de 0 e 1 observamos que apenas entram para o somatório os termos em que  $h \cdot h' = 1$  o que implica que:

$$h \cdot h' = 1 \quad \Rightarrow \quad h = 1 \wedge h' = 1 \quad \Rightarrow \quad (h + h')^2 = 4 \quad , \quad (B.9)$$

logo, retomando a dependência de  $I$  e  $h$  em  $\mathbf{x}$ , a dependência de  $I'$  em  $m(\mathbf{x}, \boldsymbol{\theta})$  e a dependência de  $h'$  em  $\mathbf{x}$  e  $\boldsymbol{\theta}$ , podemos escrever (B.8) como:

$$-\ln \left( \mathcal{P} \left( I(\mathbf{x}), I'(\mathbf{x}') | \widehat{P}(\mathbf{x}), \boldsymbol{\theta} \right) \right) = N \ln (2\pi\sigma^2) + (4\sigma^2)^{-1} \sum_{\mathbf{x} \in \mathbb{R}^2} [I(\mathbf{x}) - I'(m(\mathbf{x}, \boldsymbol{\theta}))]^2 h(\mathbf{x}) h'(\mathbf{x}, \boldsymbol{\theta}) \quad (B.10)$$

o que corresponde exactamente à equação (2.16).

## Apêndice C

### Dedução da equação (4.19) a partir das equações (4.18) e (4.15)

Tal como foi visto na equação (4.15) na página 38 o simétrico do logaritmo da função de verosimilhança é dado por:

$$\begin{aligned} -\ln(\mathcal{P}(\mathcal{I}|P(\mathbf{x}^0), \boldsymbol{\theta})) &= \frac{(n+1)N}{2} \ln(2\pi\sigma^2) \\ &+ (2\sigma^2)^{-1} \sum_{\mathbf{x}^0 \in \mathbb{R}^2} \sum_{i=0}^n [I^i(m(\mathbf{x}^0, \boldsymbol{\theta}^i)) - P(\mathbf{x}^0)]^2 h^i(\mathbf{x}^0, \boldsymbol{\theta}^i) \end{aligned} \quad (\text{C.1})$$

Por outro lado na página 39 mostramos que a estimativa de máxima verosimilhança para o panorama pode ser obtida imediatamente a partir da estimativa dos parâmetros da transformação de coordenadas recorrendo à equação (4.18):

$$\widehat{P(\mathbf{x}^0)} = \frac{\sum_{i=0}^n I^i(m(\mathbf{x}^0, \boldsymbol{\theta}^i)) h^i(\mathbf{x}^0, \boldsymbol{\theta}^i)}{\sum_{i=0}^n h^i(\mathbf{x}^0, \boldsymbol{\theta}^i)} \quad (\text{C.2})$$

Para obter a equação (4.19) a partir destas duas equações substituímos a estimativa do panorama,  $\widehat{P(\mathbf{x}^0)}$ , no simétrico do logaritmo da função de verosimilhança,  $-\ln(\mathcal{P}(\mathcal{I}|P(\mathbf{x}^0), \boldsymbol{\theta}))$ , obtendo:

$$\begin{aligned} -\ln(\mathcal{P}(\mathcal{I}|\widehat{P}, \boldsymbol{\theta})) &= \frac{(n+1)N}{2} \ln(2\pi\sigma^2) + (2\sigma^2)^{-1} \times \\ &\times \sum_{\mathbf{x} \in \mathbb{R}^2} \sum_{i=0}^n \left[ I^i(m(\mathbf{x}^0, \boldsymbol{\theta}^i)) - \frac{\sum_{j=0}^n I^j(m(\mathbf{x}^0, \boldsymbol{\theta}^j)) h^j(\mathbf{x}^0, \boldsymbol{\theta}^j)}{\sum_{k=0}^n h^k(\mathbf{x}^0, \boldsymbol{\theta}^k)} \right]^2 h^i(\mathbf{x}^0, \boldsymbol{\theta}^i) \end{aligned} \quad (\text{C.3})$$

Para obter um resultado mais compacto omitimos a partir deste ponto a dependência de  $I^i$  em  $m(\mathbf{x}^0, \boldsymbol{\theta}^i)$  e a dependência de  $h^i$  em  $\mathbf{x}^0$  e  $\boldsymbol{\theta}^i$ . Assim esta última equação pode ser re-escrita como:

$$-\ln(\mathcal{P}(\mathcal{I}|\widehat{P}, \boldsymbol{\theta})) = \frac{(n+1)N}{2} \ln(2\pi\sigma^2) + (2\sigma^2)^{-1} \sum_{\mathbf{x} \in \mathbb{R}^2} \sum_{i=0}^n \left[ I^i - \frac{\sum_{j=0}^n I^j h^j}{\sum_{k=0}^n h^k} \right]^2 h^i \quad (\text{C.4})$$

Tendo em conta que a função  $h^i(\mathbf{x}^0, \boldsymbol{\theta}^i)$  toma unicamente o valor de 0 ou de 1 podemos escrever:

$$-\ln(\mathcal{P}(\mathcal{I}|\widehat{P}, \boldsymbol{\theta})) = \frac{(n+1)N}{2} \ln(2\pi\sigma^2) + (2\sigma^2)^{-1} \sum_{\mathbf{x} \in \mathbb{R}^2} \sum_{i=0}^n \left[ I^i h^i - \frac{\left( \sum_{j=0}^n I^j h^j \right) h^i}{\sum_{k=0}^n h^k} \right]^2 \quad (\text{C.5})$$

Resolvendo o quadrado obtemos:

$$-\ln \left( \mathcal{P} \left( \mathcal{I} | \hat{P}, \boldsymbol{\theta} \right) \right) = \frac{(n+1)N}{2} \ln (2\pi\sigma^2) + (2\sigma^2)^{-1} \times \quad (C.6)$$

$$\sum_{\mathbf{x} \in \mathbb{R}^2} \sum_{i=0}^n \left[ (I^i h^i)^2 - 2I^i h^i \frac{\left( \sum_{j=0}^n I^j h^j \right) h^i}{\sum_{k=0}^n h^k} - \left( \frac{\left( \sum_{j=0}^n I^j h^j \right) h^i}{\sum_{k=0}^n h^k} \right)^2 \right]$$

Mais uma vez tendo em conta que a função  $hi(\mathbf{x}^0, \boldsymbol{\theta}^i)$  toma unicamente o valor de 0 ou de 1 podemos escrever:

$$-\ln \left( \mathcal{P} \left( \mathcal{I} | \hat{P}, \boldsymbol{\theta} \right) \right) = \frac{(n+1)N}{2} \ln (2\pi\sigma^2) + (2\sigma^2)^{-1} \times \quad (C.7)$$

$$\sum_{\mathbf{x} \in \mathbb{R}^2} \sum_{i=0}^n \left[ (I^i h^i)^2 - 2I^i h^i \frac{\sum_{j=0}^n I^j h^j}{\sum_{k=0}^n h^k} - h^i \frac{\left( \sum_{j=0}^n I^j h^j \right)^2}{\left( \sum_{k=0}^n h^k \right)^2} \right]$$

Transportando o somatório em  $i$  para dentro dos parêntesis obtemos:

$$-\ln \left( \mathcal{P} \left( \mathcal{I} | \hat{P}, \boldsymbol{\theta} \right) \right) = \frac{(n+1)N}{2} \ln (2\pi\sigma^2) + (2\sigma^2)^{-1} \times \quad (C.8)$$

$$\sum_{\mathbf{x} \in \mathbb{R}^2} \left[ \sum_{i=0}^n (I^i h^i)^2 - 2 \frac{\sum_{i=0}^n I^i h^i \cdot \sum_{j=0}^n I^j h^j}{\sum_{k=0}^n h^k} - \frac{\left( \sum_{i=0}^n h^i \right) \left( \sum_{j=0}^n I^j h^j \right)^2}{\left( \sum_{k=0}^n h^k \right)^2} \right]$$

Simplificando o último termo e resolvendo o quadrado obtemos:

$$-\ln \left( \mathcal{P} \left( \mathcal{I} | \hat{P}, \boldsymbol{\theta} \right) \right) = \frac{(n+1)N}{2} \ln (2\pi\sigma^2) + (2\sigma^2)^{-1} \times \quad (C.9)$$

$$\sum_{\mathbf{x} \in \mathbb{R}^2} \left[ \sum_{i=0}^n (I^i h^i)^2 - 2 \frac{\sum_{i=0}^n I^i h^i \cdot \sum_{j=0}^n I^j h^j}{\sum_{k=0}^n h^k} - \frac{\sum_{i=0}^n I^i h^i \cdot \sum_{j=0}^n I^j h^j}{\sum_{k=0}^n h^k} \right]$$

O que é equivalente a:

$$-\ln \left( \mathcal{P} \left( \mathcal{I} | \hat{P}, \boldsymbol{\theta} \right) \right) = \frac{(n+1)N}{2} \ln (2\pi\sigma^2) + (2\sigma^2)^{-1} \times \quad (C.10)$$

$$\sum_{\mathbf{x} \in \mathbb{R}^2} \left[ \frac{1}{2} \sum_{i=0}^n (I^i h^i)^2 + \frac{1}{2} \sum_{i=0}^n (I^i h^i)^2 - \frac{\sum_{i=0}^n I^i h^i \cdot \sum_{j=0}^n I^j h^j}{\sum_{k=0}^n h^k} \right]$$

Tendo em conta que o termo no segundo sumatório é mudo:

$$-\ln \left( \mathcal{P} \left( \mathcal{I} | \hat{P}, \boldsymbol{\theta} \right) \right) = \frac{(n+1)N}{2} \ln (2\pi\sigma^2) + (2\sigma^2)^{-1} \times \quad (C.11)$$

$$\sum_{\mathbf{x} \in \mathbb{R}^2} \frac{1}{2 \sum_{k=0}^n h^k} \left[ \sum_{k=0}^n h^k \cdot \sum_{i=0}^n (I^i h^i)^2 - 2 \sum_{i=0}^n I^i h^i \cdot \sum_{j=0}^n I^j h^j + \sum_{k=0}^n h^k \cdot \sum_{j=0}^n (I^j h^j)^2 \right]$$



Aproveitando mais uma vez o facto de que a função  $hi(\mathbf{x}^0, \boldsymbol{\theta}^i)$  toma unicamente o valor de 0 ou de 1 e de que todos os somatórios são mudos podemos escrever:

$$-\ln \left( \mathcal{P} \left( \mathcal{I} | \widehat{P}, \boldsymbol{\theta} \right) \right) = \frac{(n+1)N}{2} \ln (2\pi\sigma^2) + (2\sigma^2)^{-1} \times \quad (\text{C.12})$$

$$\sum_{\mathbf{x} \in \mathbb{R}^2} \frac{1}{2 \sum_{k=0}^n h^k} \left[ \sum_{j=0}^n h^j \cdot \sum_{i=0}^n (I^i)^2 h^i - 2 \sum_{i=0}^n I^i h^i \cdot \sum_{j=0}^n I^j h^j + \sum_{i=0}^n h^i \cdot \sum_{j=0}^n (I^j)^2 h^j \right]$$

O que pode ser escrito como:

$$-\ln \left( \mathcal{P} \left( \mathcal{I} | \widehat{P}, \boldsymbol{\theta} \right) \right) = \frac{(n+1)N}{2} \ln (2\pi\sigma^2) + (2\sigma^2)^{-1} \times \quad (\text{C.13})$$

$$\sum_{\mathbf{x} \in \mathbb{R}^2} \frac{1}{2 \sum_{k=0}^n h^k} \left[ \sum_{i,j=0}^n (I^i)^2 h^i h^j - 2 \sum_{i,j=0}^n I^i I^j h^i h^j + \sum_{i,j=0}^n (I^j)^2 h^i h^j \right]$$

Isto é:

$$-\ln \left( \mathcal{P} \left( \mathcal{I} | \widehat{P}, \boldsymbol{\theta} \right) \right) = \frac{(n+1)N}{2} \ln (2\pi\sigma^2) + (2\sigma^2)^{-1} \times \quad (\text{C.14})$$

$$\sum_{\mathbf{x} \in \mathbb{R}^2} \frac{1}{2 \sum_{k=0}^n h^k} \sum_{i,j=0}^n \left[ (I^i)^2 - 2I^i I^j + (I^j)^2 \right] h^i h^j$$

O que pode ser escrito como:

$$-\ln \left( \mathcal{P} \left( \mathcal{I} | \widehat{P}, \boldsymbol{\theta} \right) \right) = \frac{(n+1)N}{2} \ln (2\pi\sigma^2) + (2\sigma^2)^{-1} \times \quad (\text{C.15})$$

$$\sum_{\mathbf{x} \in \mathbb{R}^2} \frac{1}{2 \sum_{k=0}^n h^k} \sum_{i,j=0}^n (I^i - I^j)^2 h^i h^j$$

Retomando a notação completa obtemos a equação:

$$-\ln \left( \mathcal{P} \left( \mathcal{I} | \widehat{P}, \boldsymbol{\theta} \right) \right) = \frac{(n+1)N}{2} \ln (2\pi\sigma^2) + (4\sigma^2)^{-1} \times \sum_{\mathbf{x} \in \mathbb{R}^2} H^{-1}(\mathbf{x}^0, \boldsymbol{\theta}) \sum_{i,j=0}^n [I^i(m(\mathbf{x}^0, \boldsymbol{\theta}^i)) - I^j(m(\mathbf{x}^0, \boldsymbol{\theta}^j))]^2 h^i(\mathbf{x}^0, \boldsymbol{\theta}^i) h^j(\mathbf{x}^0, \boldsymbol{\theta}^j) \quad (\text{C.16})$$

Onde se define a função  $H(\mathbf{x}^0, \boldsymbol{\theta})$  como:

$$H(\mathbf{x}^0, \boldsymbol{\theta}) = \sum_{k=0}^n h^k(\mathbf{x}^0, \boldsymbol{\theta}^k) \quad (\text{C.17})$$

O que corresponde exactamente à equação (4.19) tal como se pretendia demonstrar.

## Apêndice D

### *"Registration of Images with Small Overlap"*

Artigo disponível em "<http://users.isr.ist.utl.pt/~aguiar/2005-ICIP-orig.pdf>".

## Apêndice E

### *"Featureless Global Alignment of Multiple Images"*

Artigo disponível em "<http://users.isr.ist.utl.pt/~aguiar/2004-MMSP-orig.pdf>".

# Bibliografia

- [1] “Hubblesite,” [http://hubblesite.org/sci.d.tech/behind\\_the\\_pictures](http://hubblesite.org/sci.d.tech/behind_the_pictures).
- [2] S.-Y. Chien, S.-Y. Ma, e L.-G. Chen, “Efficient moving object segmentation algorithm using background registration technique,” *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 12, no. 7, pp. 577–586, Julho 2002.
- [3] F. Dufaux e J. Konrad, “Efficient, robust, and fast global motion estimation for video coding,” *IEEE T-IP*, 2000.
- [4] N. Petrovic, N. Jojic, e T. S. Huang, “Hierarchical video clustering,” em *Proceedings of the IEEE International Workshop on Multimedia Signal Processing*, Siena, Italy, Setembro 2004.
- [5] B. Reddy e B. Chatterly, “An FFT-based technique for translation, rotation, and scale-invariant image registration,” *IEEE T-IP*, 1996.
- [6] J. Lee e J. Ra, “Block motion estimation based on selective integral projections,” em *Proc. of IEEE ICIP*, 2002.
- [7] R. Hartley e A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, first edition, reprint edition, 2001.
- [8] L. Nunes e R. Melo, “Mapa 3d a partir de laser num helicóptero,” Trabalho Final de Curso, Instituto Superior Técnico, Lisboa, Portugal, Setembro 2003.
- [9] J. V. Silva, “Imagens panorâmicas,” Trabalho Final de Curso, Instituto Superior Técnico, Lisboa, Portugal, Outubro 2004.
- [10] J. Maciel, *Global Matching: Optimal Solution to Correspondence Problems*, Ph.D. thesis, Instituto Superior Técnico, Lisboa, Portugal, 2001.
- [11] S. Mann e R. Piccard, “Video orbits of the projective group: a simple approach to featureless estimation of parameters,” *IEEE T-IP*, 1997.
- [12] J. R. Bergen, P. Anandan, K. J. Hanna, e R. Hingorani, “Hierarchical model-based motion estimation,” em *European Conf. on Computer Vision*, Santa Margherita Ligure, Itália, 1992.

- [13] B. E. Pires e P. M. Q. Aguiar, “Registration of images with small overlap,” em *Proceedings of the IEEE International Workshop on Multimedia Signal Processing*, Siena, Italy, Setembro 2004.
- [14] B. E. Pires e P. M. Q. Aguiar, “Featureless global alignment of multiple images,” em *Submetido para Proceedings of the IEEE International Conference on Image Processing*, Genova, Italy, Setembro 2005.
- [15] Harry L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Linear Modulation Theory*, John Wiley & Sons, 1968.
- [16] P. Anandan, “A unified perspective on computational technics for the measurement of visual motion,” em *International Conf. on Computer Vision*, London, England, 1987.
- [17] P. J. Burt, R. Hingorani, e R. J. Kolczynski, “Mechanisms for isolating component patterns in the sequential analysis of multiple motion,” em *IEEE Workshop on Visual Motion*, Prindeton, U.S.A., 1991.
- [18] D. A. Forsyth e J. Ponce, *Computer Vision: A Modern Approach*, Prentice Hall, first edition edition, 2003.