

Faculdade de Engenharia da Universidade do Porto

Modelos Pontuais de Distribuição em Visão Computacional:
Estudo, Desenvolvimento e Aplicação

Maria João Medeiros de Vasconcelos

Licenciada em Matemática Aplicada à Tecnologia pela Faculdade de Ciências
da Universidade do Porto

Dissertação submetida para satisfação parcial dos requisitos do grau de Mestre em:
Estatística Aplicada e Modelação

Dissertação realizada sob a Orientação de:
Professor Doutor João Manuel R. S. Tavares,
Prof. Auxiliar do Departamento de Engenharia Mecânica e Gestão Industrial
da Faculdade de Engenharia da Universidade do Porto

Outubro de 2005

Resumo

A análise de objectos representados em imagens é uma das áreas da Visão Computacional que mais desenvolvimento tem sofrido nos últimos tempos. A utilização de métodos estatísticos para a modelação de objectos deformáveis representados em imagens tem-se igualmente revelado eficiente e promissora. Assim, modelos flexíveis foram desenvolvidos com o intuito de analisar e simular objectos que apresentam formas variáveis em imagens.

Neste trabalho são estudados os Modelos Pontuais de Distribuição na modelação de objectos deformáveis baseado em análise estatística. Estes modelos permitem obter informação sobre a forma média do objecto em estudo e dos seus desvios admissíveis, sendo construídos através da análise das características de um conjunto de pontos notáveis representativos do objecto a modelar, utilizando para tal a técnica de Análise em Componentes Principais. A combinação da forma geométrica do objecto com os níveis de cinzento que o compõem, permite construir os Modelos de Forma Activa e os Modelos de Aparência Activa que podem ser utilizados no reconhecimento de objectos em novas imagens, e também considerados neste trabalho.

No presente trabalho, estudaram-se detalhadamente os modelos pontuais de distribuição e suas variantes, implementaram-se e automatizaram-se os procedimentos computacionais para a sua construção quando os objectos a considerar são do tipo mão e face. Para os referidos objectos, a construção dos modelos activos tornou-se assim completamente automática, sendo bastante satisfatórios os resultados obtidos na segmentação e reconhecimento deste tipo de objectos em novas imagens.

As técnicas de modelação apresentadas nesta Dissertação têm aplicações em variadas áreas, como por exemplo: em medicina, na localização de ossos e órgãos em imagens médicas; na indústria, na inspecção de produtos; ou em sistemas de segurança, no reconhecimento de faces.

Abstract

One of the recent areas of interest in Computational Vision is image analysis based on flexible models. In this field, the use of statistical methods for object modelling has proved to be suitable to deal with problems in which the objects represented in images have variable shapes.

This work is based on the Pontual Distribution Models for modelling objects through statistical analysis. These models are obtained by capturing the statistics of the coordinates of the landmarks that represent the deformable object in study: after aligning the object shapes, a Principal Component Analysis is made and the mean shape of the object and the main modes of variation are obtained.

The grey levels of the objects can also be modelled and used to build Active Shape Models and Active Appearance Models, also referred in this work, in order to identify the modelled objects in new images.

Statistical models have been very useful for image analysis in different applications of computational vision. For instance, they can be used on several areas like: medicine, for locating bones and organs in medical images; industry, in industrial inspection; and security systems, for face recognition.

The study of Pontual Distribution Models and its variants, the implementation and automatization of models, its construction and application in hand and face images, so that the automatic recognition of these types of objects in new images can be possible, are the main goals of the present work.

Agradecimentos

Ao Prof. João Manuel R. S. Tavares por todo o apoio prestado ao longo desta dissertação, pela sua orientação, incentivo e total disponibilidade apresentada.

A todos os professores, sem exceção, desta edição do Mestrado de Estatística Aplicada e Modelação da Faculdade de Engenharia da Universidade do Porto que permitiram a realização deste trabalho.

Aos meus colegas de mestrado que mais do que ninguém sabem o esforço dispendido ao longo destes anos.

À Camila, ao Pedro, à Liliana e à Rita pela amizade e pelo apoio demonstrados.

Ao Laboratório de Genética e Patologia Molecular pelo incentivo e paciência demonstrados.

À minha família, pais e irmão pelo carinho, incentivo e compreensão...

Índice

Capítulo 1	Introdução	1
1.1	Visão Computacional.....	1
1.2	Modelos Deformáveis e Modelos Pontuais de Distribuição.....	1
1.3	Abordagem Seguida.....	4
1.4	Estrutura da Dissertação	4
1.5	Contribuições Alcançadas.....	6
Capítulo 2	Modelos Pontuais de Distribuição e Variantes	9
2.1	Introdução	9
2.2	Modelo Pontual de Distribuição	10
2.2.1	Etiquetagem	10
2.2.2	Alinhamento.....	11
2.2.3	Variações Admissíveis.....	13
2.3	Níveis de Cinzento.....	15
2.4	Modelo de Forma Activa	17
2.4.1	Estimativa Inicial	17
2.4.2	Fase de Pesquisa	17
2.4.3	Cálculo dos Deslocamentos	18
2.4.4	Encontrar os Parâmetros de Pose e Forma.....	20
2.4.5	Actualização dos Parâmetros de Pose e Forma.....	21
2.5	Multi-resolução.....	22
2.6	Modelo de Aparência Activa	24
2.6.1	Modelo de Aparência.....	24
2.6.2	Escolha dos Pesos da Forma	26
2.6.3	Aproximação a um Novo Exemplo.....	26
2.6.4	Fase de Pesquisa	27
2.6.5	Fase de Aprendizagem.....	27
2.6.6	Algoritmo.....	29
2.7	Resumo	29
Capítulo 3	Colocação e Emparelhamento Automático dos Pontos-chave.....	31
3.1	Introdução	31
3.2	Trabalho Relacionado	32
3.3	Colocação Automática dos Pontos-Chave: Objecto Mão.....	35
3.3.1	Algoritmo de Detecção de Zonas de Pele	35
3.3.2	Pontos-chave do Contorno.....	39
3.3.3	Delimitação da Mão.....	42
3.3.4	Extracção dos Pontos Notáveis.....	43
3.4	Colocação Automática dos Pontos-Chave: Objecto Face.....	43
3.4.1	Algoritmo I: Contornos da face	44
3.4.2	Algoritmo II: Malha Regular	48
3.4.3	Algoritmo III: Malha Adaptativa de Multi-resolução.....	49
3.5	Emparelhamento Automático dos Pontos-Chave	51
3.5.1	Método de Shapiro.....	51
3.5.2	Método de Sclaroff	53

3.6	Resumo	55
Capítulo 4	Implementações Desenvolvidas, Adaptadas e Utilizadas.....	57
4.1	Introdução	57
4.2	Modelo Pontual de Distribuição	58
4.2.1	Etiquetagem do Conjunto de Treino	58
4.3	Modelo de Forma Activa	65
4.4	Modelo de Aparência Activa	68
4.5	Resumo	70
Capítulo 5	Resultados Experimentais.....	73
5.1	Introdução	73
5.2	Imagens Utilizadas.....	73
5.3	Modelação da Mão.....	74
5.3.1	Modelo de Forma Activa	75
5.3.2	Modelo de Aparência Activa	80
5.4	Modelação da Face	85
5.4.1	Modelo de Forma Activa	87
5.4.2	Modelo de Aparência Activa	90
5.5	Resumo	96
Capítulo 6	Conclusões e Perspectivas de Desenvolvimento Futuro.....	97
6.1	Conclusões Finais	97
6.2	Perspectivas de desenvolvimento futuro.....	101
Bibliografia		103
Anexo A		107
Anexo B		109

Lista de Figuras

Figura 2.1 Exemplo de um conjunto de treino.....	10
Figura 2.2 Objecto em questão (a), pontos-chave (b) e o objecto com os pontos-chave (c).....	11
Figura 2.3 Exemplos da variação (a, b, d, e) do primeiro valor próprio em torno da forma média para a mão (c).	14
Figura 2.4 Exemplo do contorno do modelo a aproximar-se da orla de intensidade do objecto.....	18
Figura 2.5 Movimento associado ao ponto-chave segundo a normal.....	18
Figura 2.6 Pirâmide de resolução: imagem original (a), imagens a níveis de resolução 2 (b) e 3 (c).	22
Figura 2.7 Existe sucesso no alinhamento se o deslocamento dX_i a efectuar ao longo do perfil se situar na região central desse perfil.....	23
Figura 2.8 Efeito da variação $(\pm 2sd)$ do primeiro modo de um modelo de aparência construído para uma face.....	26
Figura 3.1 Exemplo de conversão de uma imagem no formato <i>RGB</i> para o espaço <i>YCbCr</i>	36
Figura 3.2 Função de densidade de probabilidade considerada para a construção do modelo representativo da pele da mão usado neste trabalho.	37
Figura 3.3 Imagem original (a), imagem obtida com a representação das probabilidades de existência de pele na imagem original (b), e imagem após binarização da imagem de probabilidades (c).	39
Figura 3.4 Ilustração da conectividade δ do <i>pixel</i> central p_0	39
Figura 3.5 Exemplo de extracção do contorno a partir de duas imagens de um conjunto de treino: imagens originais (a, c), objecto determinado em cada uma das imagens consideradas (b, d) (nas imagens de resultados (b, d) o fundo está representado a cinza, o objecto obtido a azul, e o contorno deste a branco).	40
Figura 3.6 Exemplo de cálculo da <i>k-curvatura</i> no ponto P_i	40
Figura 3.7 Representação da <i>k-curvatura</i> (a) e dos pontos de maior curvatura obtidos numa imagem do conjunto de treino (b).	41

Figura 3.8 Delimitação do contorno numa imagem do conjunto de treino: imagem original (a), etiquetagem das zonas de maior curvatura (b), pontos que delimitam a mão (pontos 1 e 2) (c), contorno final obtido para a mão (d). ..	42
Figura 3.9 Função de densidade de probabilidade considerada para a construção do modelo representativo da pele da face usado.....	44
Figura 3.10 Imagem original (a), imagem com a representação das probabilidades de existência de pele (b), e imagem após binarização por <i>threshold</i> adaptativo (c).	45
Figura 3.11 Exemplo de extracção do contorno da face a partir de duas imagens de um conjunto de treino: imagens originais (a, d), localização obtida do segmento correspondente à face (b, e), objecto determinado em cada uma das imagens consideradas após a determinação do contorno exterior e posterior suavização do mesmo (c, f). (Nas imagens de resultados (c, f), o fundo está representado a cinza, o objecto a azul e o contorno a branco).	45
Figura 3.12 Exemplo de identificação dos olhos e sobrancelhas numa imagem do conjunto de treino através do mapa cromático: imagem original (a), mapa cromático resultante (b), mapa cromático binarizado (c), localização obtida das sobrancelhas (a rosa e a amarelo) e dos olhos (a vermelho e a verde) (d).	46
Figura 3.13 Componente de saturação das imagens da Figura 11 (a, d) e identificação (a vermelho) das zonas da boca respectivas (b, d).	47
Figura 3.14 Dois exemplos de localização do contorno da face (a azul), sobrancelhas (a rosa e a amarelo), olhos (a vermelho e a verde) e boca (os pontos apresentados indicam os “eixos principais” do contorno da boca).....	48
Figura 3.15 Exemplo de colocação da malha rectangular sobre uma imagem do conjunto de treino: contorno da face a azul, centroide da face a vermelho e centroides dos olhos a amarelo (a), malha rectangular 6×10 construída (b) (nodos da malha a vermelho).	49
Figura 3.16 Esquema de construção das malhas adaptativas dos <i>tipos 1</i> e <i>2</i> em (a), exemplo de uma malha adaptativa <i>tipo 1</i> (b) e do <i>tipo 2</i> (c) aplicada sobre uma imagem do conjunto de treino.	50

Figura 3.17 Exemplo de emparelhamento automático dos pontos-chave de duas formas obtido usando o método de Shapiro e considerando emparelhamento global.....	54
Figura 3.18 Exemplo de emparelhamento automático dos pontos-chave de duas formas obtido usando o método de Sclaroff considerando emparelhamento global.....	55
Figura 4.1 Implementação do algoritmo de etiquetagem automática de objectos do tipo mão: menu de entrada de definição dos parâmetros do algoritmo (a), menu para guardar o resultado final obtido (b).....	59
Figura 4.2 Interface da implementação desenvolvida para a visualização dos pontos-chave determinados (a), e exemplo de uma imagem de resultados obtida (b).....	59
Figura 4.3 Exemplo dos contornos extraídos de uma imagem de face do conjunto de treino (a) e dos pontos-chave automaticamente extraídos (b).	60
Figura 4.4 Menu de entrada da implementação desenvolvida para o algoritmo I para etiquetar automaticamente objectos do tipo face (a) e exemplo dos pontos-chave automaticamente extraídos (b) (a branco).	61
Figura 4.5 a) Menu de entrada da implementação desenvolvida para o algoritmo II para etiquetar automaticamente faces (a) e exemplo dos pontos-chave automaticamente extraídos (a branco) (b).....	61
Figura 4.6 Menu de entrada da implementação desenvolvida para o algoritmo III para etiquetar automaticamente faces (a) e exemplo dos pontos-chave automaticamente extraídos (a branco), considerando malha do <i>tipo 1</i> (b), e do <i>tipo 2</i> (c).	62
Figura 4.7 Menus da implementação desenvolvida para a etiquetagem manual: menu inicial (a), definição dos pontos-chave sobre a imagem a considerar (b).	63
Figura 4.8 Implementações adaptadas para emparelhar automaticamente dois conjuntos de pontos-chave: método de Shapiro (a) e método de Sclaroff (b).....	64
Figura 4.9 Definição dos parâmetros de entrada necessários para as implementações adaptadas para o método de Shapiro (a) e para o método de Sclaroff (b). ...	64
Figura 4.10 Exemplo de emparelhamento automático dos pontos-chave de dois objectos obtido usando o método de Shapiro (a) e o método de Sclaroff	

(b). (Nestas imagens os sub-contornos estão ligados entre si por segmentos de recta, assim como os nodos vizinhos e os emparelhamentos obtidos.)	65
Figura 4.11 Menus de entrada do programa computacional usado para a fase de treino dos modelos de forma activa (a, b).....	66
Figura 4.12 Menus de entrada do programa computacional usado para visualizar os modos de variação do modelo activo da forma construído na fase de treino (a, b).....	66
Figura 4.13 Menus de entrada do programa computacional usado para a fase de pesquisa dos modelos de forma activa (a, b).	67
Figura 4.14 Exemplo do efeito de variação de um modo de um modelo da forma construído para a mão (a); imagem de teste com a posição do modelo médio sobreposto, e após a 6 ^a , 17 ^a e 28 ^a iteração do processo de segmentação utilizando um modelo de forma activa previamente construído (b).	67
Figura 4.15 Exemplo de um ficheiro .pts com as coordenadas de um objecto do conjunto de treino (a) e implementação desenvolvida para criar ficheiros nesse formato (b).....	69
Figura 4.16 Exemplo de visualização de um modo de variação da forma (a) e de um modo de variação da aparência (b).	70
Figura 4.17 Fase de pesquisa dos modelos de aparência activa: indicação do modelo a utilizar.	70
Figura 4.18 Fase de pesquisa dos modelos de aparência activa: indicação da imagem a pesquisar.....	71
Figura 5.1 Exemplo de uma imagem do conjunto de treino considerado (a) etiquetada automaticamente usando 79 pontos-chave (b e c).....	74
Figura 5.2 Efeito da variação ($\pm 2sd$) dos primeiros 3 modos do modelo obtido, usando o método <i>A</i> , para a forma da mão.	76
Figura 5.3 Imagens de teste utilizadas para verificar o comportamento dos modelos construídos para a mão numa operação de segmentação.....	78
Figura 5.4 Imagem de teste com a posição do modelo médio obtido para a mão sobreposto, e após a 1 ^a , 9 ^a , 19 ^a , 25 ^a e 29 ^a iteração do processo de segmentação usando o modelo construído.....	78

Figura 5.5 Imagens de teste com o modelo médio obtido para a mão sobreposto (a, c, e, g) e resultados finais do processo de segmentação usando o modelo construído (b, d, f, h).....	78
Figura 5.6 Efeito da variação ($\mp 2sd$) dos três primeiros modos do modelo de textura (à esquerda) e de aparência construído para a mão (à direita).....	81
Figura 5.7 Alguns resultados de iterações do processo de segmentação de uma imagem de teste, usando o modelo de aparência construído para a mão.....	82
Figura 5.8 Imagens de teste após a primeira iteração do modelo de aparência activa (a, c, e, g) e resultados finais do processo de segmentação (b, d, f, h).	82
Figura 5.9 Imagem de teste com a posição do modelo médio obtido para a mão sobreposto, e após a 15ª e 30ª iteração do processo de segmentação usando o modelo de forma activa construído para a mão.....	84
Figura 5.10 Alguns resultados de iterações do processo de segmentação de uma imagem de teste, usando o modelo de aparência construído para a mão.....	84
Figura 5.11 Exemplo de uma imagem do conjunto de treino etiquetada (a) automaticamente usando 44 pontos-chave usando o <i>algoritmo I</i> (b e c).....	85
Figura 5.12 Exemplo de uma imagem do conjunto de treino etiquetada automaticamente usando: 49 pontos-chave pelo <i>algoritmo II</i> (a); 54 pontos-chave pelo <i>algoritmo III-tipo 1</i> (b); e 75 pontos-chave pelo <i>algoritmo III-tipo 2</i> (c).....	85
Figura 5.13 Efeito da variação ($\mp 2sd$) dos primeiros 3 modos do modelo obtido, usando o <i>método A</i> de alinhamento, para a forma da face.....	87
Figura 5.14 Imagens de teste utilizadas para verificar o comportamento dos modelos construídos para a face.	88
Figura 5.15 Imagem de teste com a posição do modelo médio obtido para a face sobreposto, e após a 1ª, 9ª, 19ª, 25ª e 29ª iteração do processo de segmentação usando o modelo construído.....	88
Figura 5.16 Imagens de teste com o modelo médio obtido para a face sobreposto (a, c, e,) e resultados finais do processo de segmentação (b, d, f).	89
Figura 5.17 Efeito da variação ($\mp 2sd$) dos três primeiros modos do modelo de textura e de aparência construído para a face usando o <i>algoritmo I</i> de etiquetagem.....	92

Figura 5.18 Efeito da variação ($\mp 2sd$) dos três primeiros modos do modelo de textura e de aparência construído para a face usando o <i>algoritmo II</i> de etiquetagem.	92
Figura 5.19 Efeito da variação ($\mp 2sd$) dos três primeiros modos do modelo de textura e de aparência construído para a face usando o <i>algoritmo III- tipo 1</i> de etiquetagem.	92
Figura 5.20 Efeito da variação ($\mp 2sd$) dos três primeiros modos do modelo de textura e de aparência construído para a face usando o <i>algoritmo III-tipo 2</i> de etiquetagem.	93
Figura 5.21 Alguns resultados de iterações do processo de segmentação numa imagem de teste, usando o modelo de aparência construído para a face com o <i>algoritmo I</i> de etiquetagem.	94
Figura 5.22 Imagem de teste após a primeira iteração do modelo de aparência activa construído (a, b, c,) e resultados finais do processo de segmentação em novas imagens usando o <i>algoritmo II</i> de etiquetagem (d), <i>algoritmo III-tipo 1</i> de etiquetagem (e) e <i>algoritmo III-tipo 2</i> de etiquetagem(f).	94
Figura A.1 Cálculo dos valores do ângulo e escala necessários para mover o ponto de uma posição para outra.....	109

Lista de Tabelas

Tabela 3.1 Amostras de zonas de pele utilizadas para a construção do modelo representativo da pele da mão usado neste trabalho.....	37
Tabela 3.2 Amostras de zonas de pele utilizadas para a construção do modelo representativo da pele da face usado neste trabalho.	44
Tabela 5.1 Os primeiros 5 valores próprios do modelo obtido para a mão e percentagens de variação retidas por cada um.....	75
Tabela 5.2 Caracterização dos modelos de forma activa gerados para a mão.....	77
Tabela 5.3 Erros obtidos dos modelos de forma construído para a mão nas novas imagens de teste (média e desvio padrão).	79
Tabela 5.4 Caracterização dos modelos de aparência activa considerados.....	81
Tabela 5.5 Erros resultantes da segmentação em imagens de teste, utilizando os modelos de aparência activa construídos.....	83
Tabela 5.6 Os primeiros 10 valores próprios do modelo da face obtido e as percentagens de variação retidas.	86
Tabela 5.7 Erros obtidos dos modelos de forma da face construídos nas imagens de teste consideradas (média e desvio padrão).....	89
Tabela 5.8 Caracterização dos modelos de aparência activa considerados.....	91
Tabela 5.9 Erros resultantes da segmentação do objecto modelado nas imagens de teste, utilizando os modelos de aparência activa construídos para a face. ...	95

Capítulo 1

Introdução

1.1 Visão Computacional

A capacidade de descrever ou interpretar imagens é para o ser humano uma tarefa relativamente simples, devido ao seu poderoso sistema de visão. A importância deste sistema tem levado inúmeros investigadores a tentar reproduzi-lo computacionalmente, através do desenvolvimento de sistemas automáticos de Visão Computacional capazes de executar algumas das suas funções. Assim, a Visão Computacional pode ser entendida como:

“Um conjunto de métodos e técnicas através dos quais sistemas computacionais são capazes de interpretar imagens.”

A segmentação e análise de objectos representados em imagens são uma das áreas da Visão Computacional que maior desenvolvimento tem sofrido nos últimos tempos, sendo várias as abordagens relacionadas com o desenvolvimento de modelos capazes de os caracterizar e, conseqüentemente, reconhecer e simular. Os modelos deformáveis, nos quais se inserem os modelos pontuais de distribuição considerados neste trabalho, são disso um exemplo, permitindo extrair, de forma adequada, as características mais representativas de objectos representados em imagens.

1.2 Modelos Deformáveis e Modelos Pontuais de Distribuição

As principais técnicas de extracção de características de objectos representados em imagens baseadas em modelos deformáveis, resumidamente apresentadas nesta secção, são as seguintes [Blake and Isard, 1998; Tavares, 2000]:

- Contornos activos;
- Templates deformáveis;
- Modelos físicos;
- Modelos pontuais de distribuição.

A utilização de modelos deformáveis na análise e interpretação de imagem surge pela primeira vez em [Kass et al., 1987], sendo apresentados os contornos activos, usualmente designados por “*snakes*”, constituídos por um conjunto de pontos que se adaptam à estrutura a segmentar segundo uma combinação de forças internas e externas. Assim, as forças internas do contorno consideradas traduzem as suas propriedades, nomeadamente a sua flexibilidade e rigidez, e as forças externas permitem atrair o contorno para as zonas importantes da imagem. Neste método, o contorno activo converge quando se atinge um equilíbrio de forças; ou seja, quando o contorno está suficientemente próximo do objecto desejado. Pois, as forças resultantes da imagem, consideradas nesta modelação, são minimizadas quando o contorno activo encontra certas características desejáveis desta, como linhas ou arestas.

Foram sugeridas diversas alterações de forma a melhorar os métodos de contornos activos, como por exemplo a adição de uma força interna de “balão” que garante um enchimento mínimo do contorno na fase inicial do processo. Os vários desenvolvimentos visam, principalmente, permitir a utilização de contornos activos em imagens de maior complexidade, mas também a sua utilização de forma totalmente automática.

Outro tipo de modelos deformáveis são os templates deformáveis que utilizam um conjunto de formas semelhantes (templates) às que se pretendem detectar na imagem, descritas por funções parametrizadas. Por exemplo, [Yuille et al., 1992] constrói um modelo para a detecção de um olho em imagens. Neste caso, o objecto é representado por um círculo que descreve a íris, duas parábolas que caracterizam as pálpebras e ainda a informação das intensidades em torno dessas regiões. Uma desvantagem desta metodologia é o facto da construção dos templates deformáveis ser complexa e fortemente dependente do objecto a considerar.

A modelação física de um objecto permite incorporar um conhecimento prévio do seu comportamento físico no modelo desenvolvido. Em [Pentland and Sclaroff, 1991] descreve-se um método de construção de um modelo físico a partir de um protótipo representado por um conjunto de nodos ligados por molas (elementos elásticos), sendo a massa de cada nodo e a rigidez de cada mola especificadas em duas matrizes. Estas matrizes são usadas para resolver um problema característico do qual se obtêm os modos de vibração do objecto que descrevem as suas variações físicas. Também em [Terzopoulos and Metaxas, 1991], se descrevem os modelos deformáveis superquádricos que mudam de forma consoante as suas propriedades físicas. Estes

modelos combinam as propriedades físicas globais de uma elipse superquádrica com as propriedades locais de uma *spline* e são essencialmente utilizados em operações de segmentação em imagens. No entanto, apesar das técnicas de modelação física serem bastante intuitivas e explorarem as propriedades dos objectos, os modelos resultantes nem sempre produzem representações fiáveis, podendo mesmo originar modelos que representam instâncias inválidas dos objectos modelados.

Por último, os modelos pontuais de distribuição são modelos que extraem as principais características do objecto estudado usando técnicas estatísticas. Estes modelos foram apresentados por [Cootes et al., 1992] e são construídos a partir de um conjunto de imagens do objecto em estudo, sendo as formas a modelar representados por um conjunto de pontos-chave, como por exemplo os pontos notáveis do contorno. Após etiquetagem do conjunto de treino, os pontos notáveis são emparelhados e alinhados, sendo o modelo pontual de distribuição obtido pelo estudo das características do conjunto dos pontos-chave, através de uma análise em componentes principais. Assim, o modelo obtido é formado por um reduzido número de parâmetros linearmente independentes que traduzem a forma média do objecto em estudo e os desvios permitidos para o mesmo.

Os modelos de forma activa e os modelos de aparência activa, apresentados em [Cootes and Taylor, 1992a; Cootes et al., 1998], utilizam os modelos pontuais de distribuição para segmentar e reconhecer objectos representados em imagens. Ambos os modelos utilizam uma combinação do modelo estatístico da forma, ou modelo pontual de distribuição, com os níveis de cinzento que o objecto pode adquirir.

A utilização destes métodos estatísticos, para a segmentação e interpretação de objectos representados em imagens, tem-se revelado interessante e eficiente em variadas áreas, como por exemplo:

- Medicina – localização de ossos e órgãos em imagens médicas 2D/3D;
- Industrial – inspecção de produtos industriais;
- Segurança – identificação e reconhecimento de faces.

Os últimos modelos referidos, os modelos pontuais de distribuição e suas variantes: os modelos de forma activa e os modelos de aparência, constituem o principal objecto de estudo e desenvolvimento deste trabalho.

1.3 Abordagem Seguida

Numa primeira fase deste trabalho, foi realizado um estudo bibliográfico aprofundado sobre os modelos de distribuição pontual. Assim, foram estudadas as metodologias necessárias para a criação destes modelos, identificados exemplos de aplicações e estudadas as recentes variantes dos modelos pontuais de distribuição: os modelos de forma activa e os modelos de aparência activa. Após este estudo, foi seleccionada a plataforma de desenvolvimento e ensaio a utilizar nas implementações inerentes a esta Dissertação.

A segunda fase do trabalho consistiu no desenvolvimento e adaptação de implementações computacionais relacionadas com a metodologia original dos modelos de distribuição pontual, ao qual se seguiram vários ensaios experimentais de forma a validar as implementações utilizadas. Nestes ensaios, as tarefas de determinação e emparelhamento dos pontos-chave foram realizadas manualmente.

A etapa seguinte consistiu na geração automática dos pontos-chave em objectos do tipo mão e face. Assim, estudaram-se algumas metodologias existentes e desenvolveram-se e implementaram-se novas metodologias mais adequadas.

Na quarta fase da Dissertação, estudaram-se técnicas para emparelhar automaticamente os pontos notáveis dos objectos, como as técnicas baseadas em análise modal da forma, ou no método dos elementos finitos, e análise modal com técnicas de optimização [Tavares, 2000; Silva, 2002; Bastos, 2003]. Foram também desenvolvidas e adaptadas implementações destas duas técnicas de emparelhamento.

Na quinta e última fase deste trabalho, foram realizados vários ensaios experimentais de forma a verificar o comportamento das metodologias implementadas para a criação automática dos modelos de distribuição pontual considerados.

1.4 Estrutura da Dissertação

Esta Dissertação está organizada em seis capítulos e dois anexos. Seguidamente, descreve-se de forma resumida os restantes cinco capítulos:

- *Capítulo 2: Modelos Pontuais de Distribuição e Variantes*

Neste capítulo são abordadas as metodologias necessárias para a criação de um modelo de distribuição pontual e para as suas variantes consideradas: o modelo de forma activa e modelo de aparência activa. Assim, são identificados alguns exemplos de aplicações, descritas as várias etapas que permitem construir os modelos pontuais de distribuição, desde a etiquetagem do conjunto de treino ao estudo das variações admissíveis do objecto, e apresentados os fundamentos dos modelos de forma e de aparência activa.

- *Capítulo 3: Colocação e Emparelhamento Automático dos Pontos-chave*

A etapa mais morosa no processo de construção do modelo pontual de distribuição é, sem dúvida, a da colocação e emparelhamento dos pontos-chave. Devido a este facto, considerou-se interessante desenvolver algoritmos para automatizar este processo para objectos do tipo mão e face. Assim, neste capítulo apresentam-se quatro métodos que permitem etiquetar automaticamente objectos do tipo mão e face, sendo um adequado ao primeiro tipo e os restantes ao segundo. Para além de métodos de etiquetagem, são ainda apresentados dois métodos de emparelhamento automático de pontos-chave.

- *Capítulo 4: Implementações Desenvolvidas, Adaptadas e Utilizadas*

Neste capítulo são apresentadas as implementações desenvolvidas e adaptadas para a colocação e emparelhamento automático dos pontos-chave, em objectos do tipo mão e face, e as adaptações efectuadas nos programas computacionais já existentes utilizados neste trabalho para a construção dos modelos de forma activa e de aparência activa.

- *Capítulo 5 Resultados Experimentais*

Nesta Dissertação os modelos construídos são aplicados em objectos de dois tipos: mão e face. Neste capítulo indicam-se as imagens utilizadas e descrevem-se os resultados obtidos para os modelos gerados nas mesmas. Assim, são estudados os modelos de forma e de aparência activa, para os dois tipos de objectos referidos e discutida a influência dos vários parâmetros considerados nos respectivos modelos. Neste capítulo, faz-se ainda a comparação entre os modelos de forma activa e os modelos de aparência activa, através do estudo dos resultados obtidos na segmentação do objecto modelado em novas imagens.

- *Capítulo 6: Conclusões e Perspectivas de Trabalho Futuro*

Por último, neste capítulo são apresentadas algumas conclusões finais sobre o trabalho efectuado ao longo desta Dissertação, bem como indicadas algumas perspectivas de trabalho futuro.

1.5 Contribuições Alcançadas

Como principais contribuições obtidas pelo trabalho realizado ao longo desta Dissertação podem-se referir:

- No domínio do tema da Dissertação, o estudo aprofundado dos modelos pontuais de distribuição e das suas variantes, nomeadamente, dos modelos de forma activa e dos modelos de aparência activa.
- O desenvolvimento de raiz de metodologias computacionais que permitem automatizar o processo de etiquetagem de objectos do tipo mão e face representados em imagens, e, consequentemente, tornar a modelação destes objectos completamente automática.
- O estudo de técnicas capazes de emparelhar, automaticamente, os pontos notáveis de dois objectos e a adaptação das implementações já existentes para realizar tal tarefa: os métodos de Shapiro e de Sclaroff [Tavares, 2000; Silva, 2002; Bastos, 2003]. Assim, caso o procedimento para a colocação automática de pontos-chave não obtenha o emparelhamento implícito dos mesmos, estes algoritmos permitem obter esta correspondência de forma automática.
- A adaptação das implementações já existentes usadas neste trabalho para a construção de modelos de forma e de aparência activa, de forma a incluir os algoritmos de colocação e emparelhamento automático dos pontos-chave desenvolvidos.
- A aplicação dos modelos estudados em objectos do tipo mão e face, permitindo o estudo detalhado da influência dos parâmetros utilizados em cada modelo construído.

- Finalmente, a comparação entre os modelos de forma activa e os modelos de aparência activa, discutindo-se as vantagens e desvantagens de utilizar cada um desses modelos.

Capítulo 2

Modelos Pontuais de Distribuição e Variantes

2.1 Introdução

O Modelo Pontual de Distribuição (*PDM - Point Distribution Model*) permite obter informação da forma média de um objecto assim como os seus desvios admissíveis. Neste capítulo, apresentam-se os fundamentos teóricos necessários para a construção de um modelo pontual de distribuição e para a construção das suas variantes consideradas: o modelo de forma activa (*ASM – Active Shape Model*) e o modelo de aparência activa (*AAM – Active Appearance Model*).

Na secção seguinte, descreve-se o processo de construção de um modelo pontual de distribuição, partindo de um conjunto de imagens do objecto em estudo (imagens de treino), sendo a forma do objecto, presente em cada imagem, representada por um conjunto de pontos-chave (por exemplo, pontos notáveis do contorno). Posteriormente, o modelo é obtido através do estudo das características do conjunto de pontos-chave: depois de se emparelharem os pontos-chave de todas as formas do objecto no conjunto de treino, é realizada uma análise em componentes principais. Deste modo, consegue-se obter um modelo pontual de distribuição, com um reduzido número de parâmetros linearmente independentes, que traduz a forma média do objecto em estudo e os desvios permitidos para a mesma.

Tal como se estuda a forma geométrica do objecto em causa, através dos modelos pontuais de distribuição, é possível estudar os níveis de cinzento que este pode adquirir, obtendo-se assim o modelo de intensidade apresentado na terceira secção deste capítulo. Ao combinar as duas modelações referidas anteriormente, geométrica e níveis de cinzento (intensidade), consegue-se obter para o objecto analisado um modelo mais realístico e robusto. Assim, pode-se construir o modelo de forma activa e o modelo de aparência activa, apresentados na quarta e quinta secção. Estes modelos permitem, entre outras operações, identificar os objectos modelados em novas imagens.

2.2 Modelo Pontual de Distribuição

Como já referido, o modelo pontual de distribuição está na base da construção do modelo de forma activa e do modelo de aparência activa. Com a metodologia usada na construção do modelo pontual de distribuição, pretende-se obter um modelo que represente a forma média de um determinado objecto, assim como os desvios admissíveis para a sua forma, a partir de um conjunto de instâncias do mesmo objecto representadas em imagens que formam um dado conjunto de treino, Figura 2.1.

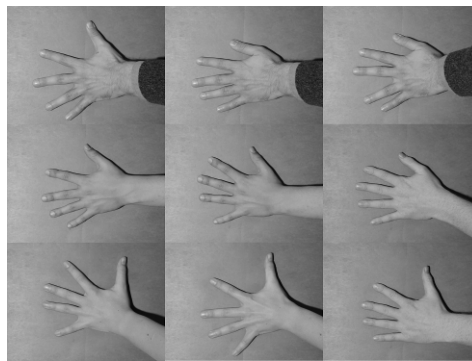


Figura 2.1 Exemplo de um conjunto de treino.

O processo de construção do modelo pontual de distribuição é constituído pelas etapas seguintes:

- 1- Etiquetagem das formas do objecto representadas nas imagens do conjunto de treino;
- 2- Alinhamento das formas do objecto representadas no conjunto de treino, obtendo-se assim a forma média;
- 3- Estudo das variações admissíveis para a forma do objecto modelado.

As secções seguintes explicam cada uma destas etapas.

2.2.1 Etiquetagem

Os objectos (2D/3D) a modelar são representados por um conjunto de pontos notáveis devidamente etiquetados. Estes pontos devem ser colocados de forma consistente em

todas as instâncias do objecto no conjunto de treino; isto é, os pontos-chave devem ser identificados sensivelmente na mesma posição correspondente em cada instância do objecto presente no conjunto treino. O método manual é o mais simples para etiquetar as instâncias dos objectos: no entanto, implica que o utilizador conheça bem o objecto em estudo, de forma a escolher a localização mais apropriada para cada ponto-chave, e seja capaz de colocá-los correctamente nas diferentes imagens que constituem o conjunto de treino. Na verdade, esta é uma operação muito morosa e têm sido estudados vários métodos automáticos ou semi-automáticos para a realização de tal tarefa [Hill and Taylor, 1994; Hill and Taylor, 1996; Hicks et al., 2002], sendo alguns deles apresentados no próximo capítulo.

Geralmente, os pontos-chave são colocados nas zonas que melhor descrevem o objecto em causa, normalmente ao longo do seu contorno ou em outros locais representativos (como por exemplo, pontos de elevada curvatura, vértices, pontos de bifurcação, etc.), Figura 2.2.

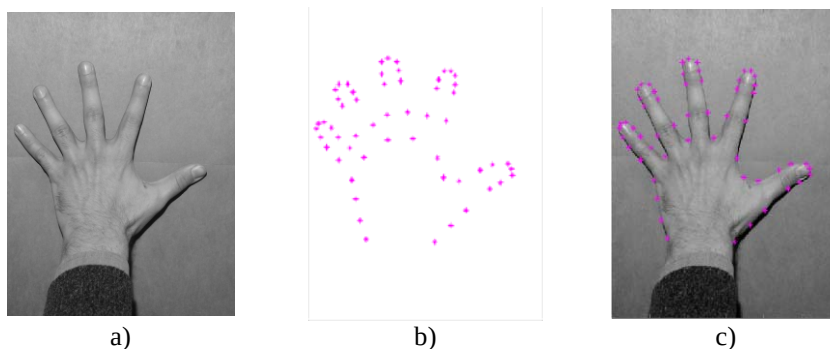


Figura 2.2 Objecto em questão (a), pontos-chave (b) e o objecto com os pontos-chave (c).

2.2.2 Alinhamento

O método de modelação em análise nesta secção consiste em estudar a variação das coordenadas dos pontos-chave no conjunto de treino. Para ser possível esta comparação, entre os pontos equivalentes em diferentes instâncias de um objecto num conjunto de imagens de treino, é necessário que estes estejam devidamente alinhados.

Em [Cootes et al., 1992] é sugerido um método de alinhamento através da rotação, translação e escalamento (transformações rígidas) das instâncias do objecto no conjunto treino. O objectivo é minimizar a soma do quadrado das distâncias entre pontos

equivalentes de instâncias distintas. No referido método, começa-se por considerar as correspondências entre duas instâncias do objecto. Assim, seja:

$$x_i = (x_{i0}, y_{i0}, \dots, x_{ik}, y_{ik}, \dots, x_{in-1}, y_{in-1})^T, \quad (2.1)$$

onde $i = 1 \dots N$, com N a representar o número de instâncias do objecto e n o número de pontos-chave, e seja $M_k(s_k, \theta_k)[x_k]$ a rotação por θ_k e escalamento por s_k . Desta forma, dadas duas instâncias semelhantes, x_i e x_k , pode-se escolher θ_k , s_k e a translação $(t_x, t_y)_k$ que faça corresponder x_i em $M_k(s_k, \theta_k)[x_k]$ de forma a minimizar a soma:

$$E_k = (x_i - M_k(s_k, \theta_k)[x_k])^T W (x_i - M_k(s_k, \theta_k)[x_k]), \quad (2.2)$$

onde:

$$M_k(s_k, \theta_k) \begin{bmatrix} x_{kj} \\ y_{kj} \end{bmatrix} = \begin{bmatrix} (s_k \cos \theta) x_{kj} - (s_k \sin \theta) y_{kj} + t_{kx} \\ (s_k \sin \theta) x_{kj} + (s_k \cos \theta) y_{kj} + t_{ky} \end{bmatrix}, \quad (2.3)$$

e W é a matriz de pesos diagonal para cada ponto (ver anexo A).

Os pesos atribuídos a cada ponto-chave dependem obviamente do tipo de objecto que se pretende modelar, existindo várias opções descritas para o cálculo da matriz de pesos. Neste trabalho, consideram-se dois métodos para a obtenção da matriz de pesos W :

- *Método A*: os pesos são atribuídos conforme a posição de cada ponto-chave seja mais ou menos estável (constante) nas instâncias do objecto no conjunto de treino [Cootes et al., 1992];
- *Método B*: a matriz de pesos é a matriz identidade, considera todos os pontos com o mesmo peso e portanto nenhum ponto-chave é privilegiado.

Segundo o *Método A*, a matriz de pesos é definida do seguinte modo: seja R_{jr} a distância entre os pontos j e r do objecto; $V_{R_{jr}}$ a variância desta distância no conjunto de objectos; então, pode-se escolher um peso, w_j , para o j^{esimo} ponto utilizando:

$$w_j = \left(\sum_{r=0}^{n-1} V_{R_{jr}} \right)^{-1}.$$

Assim, se um ponto tiver tendência para se movimentar muito em relação aos outros pontos-chave do objecto, a soma de variâncias será elevada e será associado um peso reduzido. Se, pelo contrário, um ponto tiver tendência a mover-se pouco em relação aos

outros, a soma de variâncias será reduzida e será atribuído um peso significativo, sendo a correspondência destes pontos, em outras instâncias, prioritária.

Por sua vez, no *Método B*, a matriz de pesos é simplesmente $W = I$. O estudo deste segundo método, em alternativa ao método proposto por [Cootes et al., 1992] permite inferir sobre a influência da utilização de diferentes pesos na construção do modelo pontual de distribuição. A implementação deste método foi desenvolvida de raiz neste trabalho.

O alinhamento de todas as instâncias do objecto, no conjunto de treino, pode ser realizado pelo algoritmo seguinte:

1. *Rodar, redimensionar e transladar cada uma das instâncias do objecto no conjunto de treino com a primeira instância do objecto.*

Repetir os passos seguintes até ser obtida a convergência:

2. *Calcular a média das instâncias alinhadas.*
3. *Rodar, redimensionar e transladar a média para que alinhe com a primeira instância.*
4. *Rodar, redimensionar e transladar, novamente, cada uma das instâncias de forma a coincidir com a média ajustada.*

O terceiro passo da iteração é necessário para renormalizar a média; caso contrário, a média iria variar infinitamente e a convergência não seria atingida.

A condição de convergência pode ser testada através da diferença média entre as transformações necessárias para alinhar cada instância com a média e a transformação identidade. O método de alinhamento apresentado é uma forma da Análise de Procrustes Generalizada [Gower, 1975].

2.2.3 Variações Admissíveis

Os dois passos anteriores permitiram organizar toda a informação necessária para estudar o objecto em questão, o próximo passo será agrupar toda essa informação. Nesta fase estão todas as formas, ou seja, as instâncias do objecto no conjunto de treino, devidamente alinhadas, tendo-se N conjuntos de pontos x_i alinhados. O objectivo desta fase é capturar a variabilidade da forma das instâncias e reduzir a

dimensionalidade dos dados obtidos até então, sem haver perda considerável de informação, para tal, aplica-se uma Análise em Componentes Principais.

A Análise em Componentes Principais é uma forma de identificar padrões em dados e permite expressar esses dados, de forma a evidenciar as semelhanças e diferenças existentes [Duda et al., 2001]. Assim, começa-se por calcular a forma média:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i,$$

e a matriz de covariância S definida por:

$$S = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})^T.$$

Seguidamente, determinam-se os valores próprios, λ_l , e os vectores próprios, p_l , da matriz de covariância pela equação característica:

$$Sp_l = \lambda_l p_l,$$

onde λ_l é o $l^{\text{ésimo}}$ valor próprio de S ($\lambda_l \geq \lambda_{l+1}$). Os vectores próprios p_l , normalizados tal que $p_l^T p_l = 1$, representam os modos de variação da forma do objecto modelado. Os vectores próprios da matriz de covariância S , correspondentes aos valores próprios mais elevados, descrevem grande parte das variações admissíveis e a proporção da variância total explicada por cada vector próprio, é igual ao valor próprio correspondente [Fukunaga and Koontz, 1970]. Assim, as variações mais importantes da forma do objecto modelado podem ser descritas por um reduzido número, t , de vectores próprios.

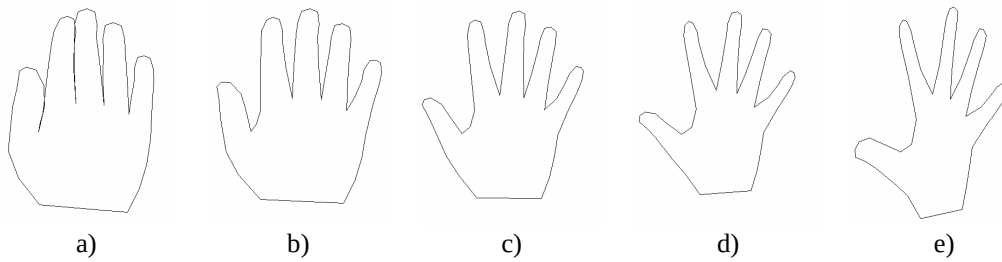


Figura 2.3 Exemplos da variação (a, b, d, e) do primeiro valor próprio em torno da forma média para a mão (c).

A percentagem de variação de cada modo é dada por:

$$\frac{\lambda_l}{\lambda_T} \times 100\%, \quad (2.4)$$

onde $\lambda_T = \sum_{l=1}^{2n} \lambda_l$. Assim, a variância explicada pelos primeiros t modos de variação é determinada por:

$$\frac{\sum_{l=1}^t \lambda_l}{\lambda_T} \times 100\% . \quad (2.5)$$

Deste modo, cada instância do objecto no conjunto de treino pode ser representada pela forma média e pela combinação dos primeiros t vectores próprios obtidos:

$$x = \bar{x} + Pb , \quad (2.6)$$

onde $P = (p_1 p_2 \dots p_t)$ é a matriz dos primeiros t vectores próprios e $b = (b_1 b_2 \dots b_t)$ é o vector de pesos de cada vector próprio.

Os vectores próprios são ortogonais $P^T P = I$ e portanto pode-se obter:

$$b = P^T (x - \bar{x}) . \quad (2.7)$$

As equações anteriores permitem criar novas instâncias do objecto, variando os parâmetros b_i dentro de limites aceitáveis. Admitindo que os parâmetros b_i obedecem a uma distribuição normal, a maior parte das instâncias de objecto situa-se a menos de 3 desvios padrão da média. Como a variância de b_i sobre o conjunto de treino é dada por λ_i , a maioria das instâncias do objecto, cerca de 99%, pode ser gerada estabelecendo-se um limite para os parâmetros b_i tal que:

$$-3\sqrt{\lambda_i} \leq b_i \leq 3\sqrt{\lambda_i} , \quad (2.8)$$

pois a população está concentrada entre três desvios padrões da média [Lima, 2003].

Neste trabalho, o limite imposto pela equação anterior foi alterado para $-2\sqrt{\lambda_i} \leq b_i \leq 2\sqrt{\lambda_i}, i=1 \dots n$ pois, após a análise dos modos de variação em torno da média entre três desvios padrão, quer para o modelo da mão quer para o modelo da face, e tendo em conta as imagens de treino e de teste consideradas, constatou-se que, com esta diminuição, os resultados obtidos na fase de pesquisa são de melhor qualidade.

2.3 Níveis de Cinzento

O modelo de distribuição pontual, usado nos modelos activos, considera, não só informação relativa à forma geométrica do objecto, mas também relativa aos níveis de

cinzento em torno dos pontos-chave. Para tal, considera-se que, como estes pontos correspondem a determinadas zonas do objecto, os níveis de cinzento dos pontos, em diferentes instâncias do mesmo objecto, vão ser, necessariamente, semelhantes. Assim, essa informação é utilizada quando se procura a região da imagem com maior parecença ao modelo construído. A forma de extrair os perfis de intensidade de todos os pontos do objecto é apresentada em seguida.

Para cada ponto-chave, j , da instância do objecto na imagem, i , do conjunto de treino, é extraído o perfil de intensidade, g_{ij} , de dimensão n_p píxeis, centrado nesse ponto.

Neste trabalho, utilizam-se as derivadas normalizadas dos perfis que permitem uniformizar a escala dos níveis de cinzento; no entanto, é também possível utilizar apenas os perfis de intensidade para a fase de pesquisa [Cootes and Taylor, 1993].

O perfil de intensidade do ponto-chave j da imagem i é um vector de dimensão n_p ,

$$g_{ij} = \begin{bmatrix} g_{ij0} & g_{ij1} & \dots & g_{ijn_p-1} \end{bmatrix}^T, \quad (2.9)$$

onde $g_{ijk} = I_i(y_k)$, sendo y_k o k^{esimo} ponto do perfil:

$$y_{ik} = p_{istart} + \frac{k-1}{n_p-1}(p_{iend} - p_{istart}), \quad (2.10)$$

e $I_i(y_k)$ é o nível de cinzento na imagem i neste ponto-chave.

O perfil derivado de dimensão $n_p - 1$ é:

$$g'_{ij} = \begin{bmatrix} g_{ij1} - g_{ij0} & g_{ij2} - g_{ij1} & \dots & g_{ijn_p-1} - g_{ijn_p-2} \end{bmatrix}. \quad (2.11)$$

Por outro lado, o perfil derivado normalizado é dado por:

$$g'_{ij} = \frac{g'_{ij}}{\sum_{k=0}^{n_p-2} g'_{ijk}}. \quad (2.12)$$

A média dos perfis derivados normalizados, $\hat{g}_j$, para cada ponto-chave, j , é obtida por:

$$\hat{g}_j = \frac{1}{N} \sum_{i=1}^N g'_{ij}, \quad (2.13)$$

e a matriz de covariância S_{gj} é dada por:

$$S_{gj} = \frac{1}{N} \sum_{i=1}^N (g'_{ij} - \hat{g}_j)(g'_{ij} - \hat{g}_j)^T. \quad (2.14)$$

Deste modo, obtém-se toda a informação sobre os perfis de intensidade (níveis de cinzento) necessária para a fase de pesquisa a descrever na secção seguinte, particularmente na etapa do cálculo dos movimentos associados aos pontos-chave.

2.4 Modelo de Forma Activa

Depois de construído o modelo pontual de distribuição e obtidos os perfis de intensidade para cada ponto-chave, é então possível localizar instâncias do objecto modelado em novas imagens através da utilização do Modelo de Forma Activa (*ASM – Active Shape Model*) [Cootes and Taylor, 1992a].

2.4.1 Estimativa Inicial

Para que a pesquisa efectuada pelo modelo de forma activa obtenha resultados satisfatórios, o posicionamento inicial do modelo na nova imagem é um factor determinante. Denotando a operação de redimensionamento e rotação da forma representada pelo vector x por s e θ , respectivamente por $M(s, \theta)[x]$, e sendo a translação dada pelo vector $t_i = [t_{xi} \ t_{yi} \ t_{xi} \ \dots \ t_{xi} \ t_{xi}]^T$ de dimensão $2n$, pode-se expressar a estimativa inicial x_i da forma como uma versão redimensionada, rodada e transladada de x_1 :

$$x_i = M(s_i, \theta_i)[x_1] + t_i \quad (2.15)$$

x_1 também pode ser escrito como $x_1 = \bar{x} + dx_1$, com $dx_1 = Pb_1$, ou como $x_1 = \bar{x} + Pb_1$ e então tem-se como estimativa inicial:

$$x_i = M(s_i, \theta_i)[\bar{x} + dx_i] + t_i \quad (2.16)$$

2.4.2 Fase de Pesquisa

A fase de pesquisa propriamente dita começa depois de escolhida a estimativa inicial. Nesta fase procede-se ao ajuste dos pontos do modelo através de um processo iterativo.

Este processo divide-se essencialmente em três etapas: a primeira, consiste no estudo da região em volta de cada ponto-chave para calcular o movimento associado; na segunda, calcula-se os parâmetros de forma e pose do modelo de distribuição pontual a partir desses movimentos associados; e finalmente, na terceira etapa, actualizam-se estes parâmetros tendo em conta os limites impostos pelo modelo previamente construído. Todo este processo é descrito nas secções seguintes.

2.4.3 Cálculo dos Deslocamentos

Esta etapa consiste na determinação dos ajustes necessários para que o modelo se adapte à nova instância do objecto. No caso dos pontos-chave estarem localizados nos contornos do objecto, os ajustes irão fazer com que os pontos se movam para as orlas de maior intensidade, Figura 2.4. Por outro lado, no caso de se ter os perfis de intensidade de cada ponto-chave, a pesquisa implica encontrar a região que melhor se adapta ao modelo dos perfis de intensidade, Figura 2.5. Para cada ponto-chave do modelo da imagem, é extraído o perfil derivado, g' , com determinada dimensão $l(>n_p)$ e este perfil é comparado com o perfil do modelo para que se determine o movimento, ou seja o deslocamento associado a cada ponto-chave.

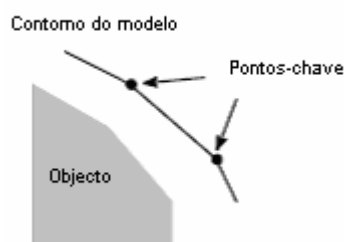


Figura 2.4 Exemplo do contorno do modelo a aproximar-se da orla de intensidade do objecto.

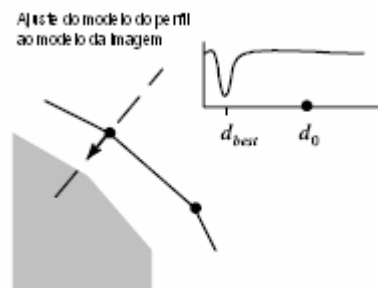


Figura 2.5 Movimento associado ao ponto-chave segundo a normal.

Os perfis de pesquisa são comparados por uma função de semelhança, como por exemplo o desvio quadrático médio, [Hamarneh, 1999], e a distância de *Mahalanobis*, [Cootes et al., 1994], sendo a primeira a utilizada neste trabalho por ser mais adequada [Hamarneh, 1999]. Assim, seja g_{s_j} o perfil de intensidade da imagem e $\hat{g}_j$ a média dos

perfis derivados normalizados para cada ponto-chave, denotando o sub-intervalo de g_{s_j} centrado no $d^{ésimo}$ pixel de g_{s_j} , por $h(d)$, encontra-se o valor de d onde o sub-perfil $h(d)$ é mais semelhante a $\hat{g}_j$. Tal valor pode ser obtido minimizando o desvio quadrático médio:

$$f(d) = (h(d) - \hat{g}_j)^T I (h(d) - \hat{g}_j), \quad (2.17)$$

onde I é a matriz identidade, ou minimizando a distância de *Mahalanobis* para o qual o seu quadrado é definido por:

$$f(d) = (h(d) - \hat{g}_j)^T S_{g_j}^{-1} (h(d) - \hat{g}_j). \quad (2.18)$$

A partir do cálculo anterior, encontra-se o deslocamento dX necessário para ajustar melhor o modelo. No entanto, é importante colocar algumas restrições a esses deslocamentos, para que a deformação do modelo não ultrapasse os limites impostos inicialmente. Assim, suponhamos que d_{best} é a distância entre o ponto do modelo e o ponto que melhor se ajusta. Escolhe-se um deslocamento para o ponto do modelo de dX que é paralelo ao perfil, na direcção do ponto de melhor ajuste, com magnitude:

$$\begin{aligned} &\text{se } |d_{best}| \leq \delta, \text{ então } |dX| = 0 \\ &\text{se } \delta < |d_{best}| < d_{max}, \text{ então } |dX| = 0.5d_{best}. \\ &\text{se } |d_{max}| \leq |d_{best}|, \text{ então } |dX| = 0.5d_{max} \end{aligned} \quad (2.19)$$

Tendo em conta os limites anteriores e os testes experimentais realizados com os limites sugeridos, optou-se por fazer neste trabalho uma ligeira alteração à equação acima indicada e sugerida por [Lima, 2003]. Assim, permite-se que os pontos do modelo se desloquem livremente na situação em que $\delta \leq |d_{best}| \leq d_{max}$; ou seja, se $\delta \leq |d_{best}| \leq d_{max}$ o deslocamento efectivo é de d_{best} e não, como considerado na equação (2.19), de $0.5d_{best}$. Esta alteração permite deslocar o modelo para uma posição mais próxima possível da posição final. Nos restantes intervalos considerados na equação (2.19), mantiveram-se as restrições apresentadas. Quanto aos valores utilizados para δ e d_{max} , após uma fase exaustiva de testes e tendo em conta estudos anteriores, optou-se por utilizar: $\delta = 0.5$ píxeis e $d_{max} = 10$.

2.4.4 Encontrar os Parâmetros de Pose e Forma

Depois de encontrado o movimento associado a cada ponto-chave, é preciso ajustar os parâmetros da pose (rotação, translação e escala) e os parâmetros da forma (os pesos das componentes principais) para mover a estimativa actual x_i o mais próximo possível de $x_i + dx_i$, ou seja:

$$M(s_i(1+ds), \theta_i + d\theta)[x_1] + t_i + dt \rightarrow x_i + dx_i, \quad (2.20)$$

Assim, em primeiro lugar, determina-se o escalamento $(1+ds)$, rotação $d\theta$ e a translação (dt_x, dt_y) necessárias para mover x_i o mais próximo possível de $x_i + dx_i$ (como se descreve no anexo B). Deste modo, resulta os resíduos dos ajustes que apenas podem ser satisfeitos pela deformação da forma x_i . O objectivo é determinar os ajustes dx (a adicionar a x_i), necessários para mover x_i para $x_i + dx_i$, quando combinados com os parâmetros de rotação, translação e escalamento.

Conhecendo $(1+ds)$, $d\theta$ e dt , determinados na etapa anterior, é necessário resolver a equação seguinte em dx :

$$M(s_i(1+ds), \theta_i + d\theta)[x_1 + dx] + t_i + dt = x_i + dx_i, \quad (2.21)$$

usando $x_i = M(s_i, \theta_i)[x_1] + t_i$ obtém-se:

$$M(s_i(1+ds), \theta_i + d\theta)[x_1 + dx] = M(s_i, \theta_i)[x_1] + dx_i - dt, \quad (2.22)$$

como $M^{-1}(s, \theta)[...] = M(s^{-1}, -\theta)[...]$ tem-se:

$$x_1 + dx = M(s_i(1+ds)^{-1}, -(\theta_i + d\theta)) [M(s_i, \theta_i)[x_1] + dx_i - dt], \quad (2.23)$$

ou seja:

$$dx = M(s_i(1+ds)^{-1}, -(\theta_i + d\theta)) [M(s_i, \theta_i)[x_1] + dx_i - dt] - x. \quad (2.24)$$

No geral, o vector resultante dx é de dimensão $2n$, mas como apenas se considera t (menor que $2n$) modos de variação descritores no modelo, pode-se mover a forma em t dimensões descritas pelos primeiros t eixos principais. Então, procura-se o vector que é mais próximo a dx mas na dimensão t . Se se adoptar a aproximação pelos mínimos quadrados, então a solução dx' é a projecção de dx no espaço de dimensão t . Formulando matematicamente, tem-se:

$$dx' = A dx, \quad (2.25)$$

onde A é a matriz de projecção dada por $A = P(P^T P)^{-1} P^T$.

Como as colunas de P são ortonormais e P não é uma matriz quadrada, tem-se que utilizar $P^T P = I$ e assim obter $dx' = PP^T dx$.

Assim, em vez de mover x_1 para $x_1 + dx$ move-se para $x_1 + dx'$. Fazendo dx' como $dx' = Pdb'$ e multiplicando à esquerda por P^T obtém-se:

$$db' = P^T dx'. \quad (2.26)$$

2.4.5 Actualização dos Parâmetros de Pose e Forma

Finalmente, com toda a informação obtida como descrito anteriormente, pode-se actualizar os parâmetros de pose e forma da estimativa inicial. Assim, obtém-se uma nova estimativa $x_i^{(1)}$ onde:

$$x_i^{(1)} = M(s_i(1+ds), \theta_i + d\theta)[x_1 + Pdb'] + t_i + dt. \quad (2.27)$$

Formulado de outro modo, os parâmetros são actualizados da seguinte maneira:

$$\begin{aligned} t_{xi} &\rightarrow t_{xi} + dt_x \\ t_{yi} &\rightarrow t_{yi} + dt_y \\ s_i &\rightarrow s_i(1+ds). \\ \theta_i &\rightarrow \theta_i + d\theta \\ b_l &\rightarrow b_l + db' \end{aligned} \quad (2.28)$$

Podendo os pesos ser adicionados da seguinte forma:

$$\begin{aligned} t_{xi} &\rightarrow t_{xi} + w_t dt_x \\ t_{yi} &\rightarrow t_{yi} + w_t dt_y \\ s_i &\rightarrow s_i(1 + w_s ds), \\ \theta_i &\rightarrow \theta_i + w_\theta d\theta \\ b_l &\rightarrow b_l + W_b db' \end{aligned} \quad (2.29)$$

onde w_t, w_s, w_θ são escalares e W_b é a matriz de pesos diagonal (podendo ser a matriz identidade). Note-se que os limites de b devem ser sempre considerados para que se obtenha sempre uma forma aceitável.

Os factores de deslocamento, w_t , de rotação, w_s , e de escala, w_θ , e os elementos da matriz de pesos diagonal W_b , considerados neste trabalho para a actualização dos parâmetros de pose e de forma foram os seguintes:

$$\begin{aligned}
 w_t &= 1 \\
 w_s &= 1 \\
 w_\theta &= 1 \\
 W_b &= I
 \end{aligned} \tag{2.30}$$

A utilização de valores unitários nestes parâmetros deve-se essencialmente ao facto de não existir, à partida, uma posição ou dimensão mais provável para a estimativa na imagem [Cootes and Taylor, 1993; Hamarneh, 1999; Lima, 2003].

Estas etapas são repetidas, até que não se encontre mudanças significativas entre a nova estimativa e a anterior, ou seja, até que se verifique a convergência no alinhamento entre o modelo e a estimativa, ou até ser atingido um dado número máximo de iterações.

2.5 Multi-resolução

Em 1994 [Cootes et al., 1994] apresenta um melhoramento ao modelo de forma activa descrito nas secções anteriores que consiste em utilizar imagens em vários níveis de resolução (pirâmide de resolução, Figura 2.6). Assim, os perfis de intensidade são estudados para a imagem original e para níveis de resolução inferiores à original, com a finalidade dos modelos de forma activa serem mais rápidos e eficientes.

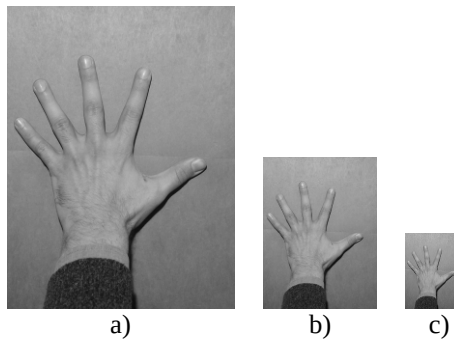


Figura 2.6 Pirâmide de resolução: imagem original (a), imagens a níveis de resolução 2 (b) e 3 (c).

O método usado cria inicialmente uma pirâmide de resolução das imagens, para tal aplica uma máscara Gaussiana de suavização, de dimensão 5×5 à imagem original e reamostra a imagem resultante de modo a obter uma imagem com metade da resolução da anterior, este procedimento é repetido para os restantes níveis da pirâmide. Para uma pirâmide com Q níveis de resolução, as coordenadas do modelo são dadas por:

$$\frac{X}{2^{q-1}}, \quad (2.31)$$

onde $q = 1 \dots Q$, $q = 1$ corresponde à imagem original e o parâmetro de forma b é:

$$\frac{b}{2^{q-1}}. \quad (2.32)$$

A mudança de nível de resolução ocorre quando for atingido um número máximo de iterações ou quando houver sucesso no alinhamento. O sucesso no alinhamento é obtido quando uma fracção γ , do total de pontos-chave do modelo, se situar a uma distância do *pixel* central do perfil de pesquisa não superior a uma fracção β do semi-comprimento desse perfil; ou seja, $|dX_i| \leq \beta \frac{L_p}{2}$ para um perfil de pesquisa de comprimento L_p em torno do ponto X_i , Figura 2.7, [Cootes et al., 1994] utiliza $\gamma = 0.95$ e $\beta = 0.50$.

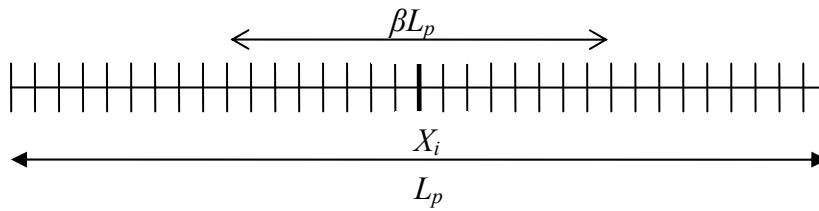


Figura 2.7 Existe sucesso no alinhamento se o deslocamento dX_i a efectuar ao longo do perfil se situar na região central desse perfil.

O número máximo de níveis da pirâmide de resolução considerado na construção dos modelos é determinado neste trabalho pela dimensão da imagem segundo a equação:

$$q_{\max} = \text{int} \left\{ \log_2 \left[\min(l_x, l_y) \right] \right\} - 3, \quad (2.33)$$

onde l_x e l_y representam a dimensão da imagem (em *píxeis*) segundo os eixos x e y [Lima, 2003].

Também neste trabalho, como critério de sucesso no alinhamento para a mudança do nível de resolução, consideraram-se os seguintes valores para os parâmetros γ e β :

$$\begin{aligned} \gamma &= 0.90 \\ \beta &= 0.25 \end{aligned} \quad (2.34)$$

O valor escolhido para β , restringe a posição considerada como aceitável a uma região mais reduzida do que o proposto por [Cootes et al., 1999] ($\beta = 0.50$).

A utilização da multi-resolução permite atingir resultados melhores com menos iterações do que o método original. Outra das vantagens da multi-resolução é a menor probabilidade do modelo ficar retido em máximos ou mínimos locais, pois a pesquisa em imagens de resolução inferior à original, permite procurar um contorno que se encontre em zonas da imagem mais afastadas da localização inicial do modelo.

2.6 Modelo de Aparência Activa

Alguns anos mais tarde, [Cootes et al., 1998] apresenta o Modelo de Aparência Activa (*AAM - Active Appearance Model*) como alternativa ao modelo de forma activa para a segmentação de imagens. O modelo de aparência é formado por uma combinação de um modelo estatístico da forma geométrica com um modelo dos níveis de cinzento do objecto. Resumidamente, o método de pesquisa estuda a relação entre os parâmetros obtidos pelo modelo e os erros residuais existentes entre a nova imagem e o modelo construído durante a fase de treino, com o objectivo de produzir um modelo final mais robusto.

2.6.1 Modelo de Aparência

O primeiro passo na construção do modelo de aparência corresponde a construir o modelo estatístico das variações da forma geométrica, a partir de um conjunto de objectos etiquetados por pontos-chave. Para construir esse modelo estatístico, após etiquetar as instâncias do objecto no conjunto treino, deve-se alinhá-las e representá-las por um vector x onde posteriormente se aplica uma análise em componentes principais. O segundo passo corresponde a construir o modelo estatístico dos níveis de cinzento, para tal deforma-se cada instância do objecto do conjunto treino de forma aos seus pontos-chave coincidirem com a forma média e obtém-se a informação sobre os níveis de cinzento, g_{im} , da imagem do modelo normalizado. Para minimizar o efeito da variação da luminosidade global, normaliza-se cada exemplo aplicando uma escala α e uma translação β

$$g = (g_{im} - \beta) / \alpha. \quad (2.35)$$

Os valores de α e β são escolhidos de forma a ajustar o vector à média normalizada. Seja $\bar{g}$ a média dos dados normalizados, escalados e transladados onde a soma dos elementos é zero e a variância é igual à unidade. Então, os valores de α e β necessários para normalizar g_{im} são dados por:

$$\alpha = g_{im} \cdot \bar{g}, \beta = (g_{im} \cdot 1) / n, \quad (2.36)$$

onde n é o número de elementos dos vectores.

Obter a média dos dados normalizados é então um processo recursivo, visto que a normalização é definida pela média. Uma solução estável pode ser encontrada utilizando um dos exemplos como primeira estimativa da média e alinhando todos os outros a este, depois reestima-se a média e itera-se.

Ao aplicar a análise em componentes principais aos dados normalizados, obtém-se o modelo linear:

$$g = \bar{g} + P_g b_g, \quad (2.37)$$

onde $\bar{g}$ é o vector médio dos níveis de cinzento, P_g é o conjunto ortogonal dos modos de variação e b_g é o conjunto de parâmetros dos níveis de cinzento ou textura.

Assim, a forma geométrica e a aparência de cada objecto modelado pode ser resumida pelos vectores b_s e b_g . Como é bastante provável que existam correlações entre as variações da forma e da textura (ou seja, dos níveis de cinzento), aplica-se uma nova análise em componentes principais com o intuito de reduzir o número de variáveis e, consequentemente, obter um modelo final mais eficiente. Assim, para cada objecto gera-se o vector concatenado:

$$b = \begin{pmatrix} W_s b_s \\ b_g \end{pmatrix} = \begin{pmatrix} W_s P_s^T (x - \bar{x}) \\ P_g^T (g - \bar{g}) \end{pmatrix}, \quad (2.38)$$

onde W_s é a matriz de pesos diagonal para cada parâmetro da forma. Seguidamente, aplica-se uma análise em componentes principais nestes vectores, obtendo-se o modelo:

$$b = Qc, \quad (2.39)$$

onde Q é a matriz que contém os vectores próprios e c é o vector dos parâmetros de aparência que controlam tanto a forma geométrica como os níveis de cinzento do modelo. Como os parâmetros de forma e dos níveis de cinzento têm média zero, c também terá.

A natureza linear do modelo permite expressar a forma e os níveis de cinzento directamente como funções de c :

$$x = \bar{x} + P_s W_s Q_s c, g = \bar{g} + P_g Q_g c, \quad (2.40)$$

onde

$$Q = \begin{pmatrix} Q_s \\ Q_g \end{pmatrix}, \quad (2.41)$$

Na Figura 2.8, apresenta-se um exemplo do primeiro modo de variação de um modelo de aparência construído para uma imagem de face.



Figura 2.8 Efeito da variação ($\pm 2sd$) do primeiro modo de um modelo de aparência construído para uma face.

2.6.2 Escolha dos Pesos da Forma

Os elementos de b_s têm unidades de distância (em *píxeis*), os de b_g têm unidades de intensidade (normalmente entre 0 e 255) logo não podem ser comparados directamente. Para fazer b_s e b_g comparáveis, é necessário estimar o efeito da variação de b_s com g . Para tal, desloca-se sistematicamente cada elemento de b_s do seu valor óptimo em cada imagem treino e representa-se cada imagem dada a forma deslocada. A variação da raiz média quadrática (*root mean square*) em g por unidade no parâmetro b_s dará o peso w_s a ser aplicado na equação (2.38), [Cootes et al., 1998].

2.6.3 Aproximação a um Novo Exemplo

Dada uma nova imagem, etiquetada por um conjunto de pontos-chave, é possível gerar uma aproximação a partir do modelo de aparência previamente construído. Segue-se os passos anteriores para obter b , combinando os parâmetros da forma geométrica e de

aparência que se ajustam ao modelo, e como Q é ortogonal, os parâmetros do modelo de aparência combinado c são dados por:

$$c = Q^T b. \quad (2.42)$$

A reconstrução é assim dada pela aplicação das equações (2.40) que invertem a normalização dos níveis de cinzento, aplicam a posição, a orientação e o escalamento apropriados aos pontos e projectam o vector dos níveis de cinzento na imagem.

2.6.4 Fase de Pesquisa

Em seguida, descreve-se uma solução para ajustar os parâmetros do modelo de uma forma eficiente, partindo do principio que se tem uma nova imagem onde se quer identificar um objecto (segmentar), um modelo de aparência construído para tal objecto como se descreveu anteriormente, e uma boa estimativa inicial.

O objectivo é interpretar este problema como um problema de optimização onde se minimiza a diferença entre a nova imagem e a sintetizada pelo modelo de aparência activa construído. Sendo definido um vector de diferenças:

$$\delta I = I_i - I_m, \quad (2.43)$$

onde I_i é o vector dos níveis de cinzento na nova imagem e I_m é o vector de níveis de cinzento do modelo construído.

Para localizar o melhor ajuste entre o modelo e a nova imagem, minimiza-se a magnitude do vector das diferenças, $\Delta = |\delta I|^2$, pela variação dos parâmetros do modelo c . Este sistema de ajuste resume-se a duas partes: “aprender” a relação entre δI e o erro dos parâmetros do modelo δc e usar esta aprendizagem no algoritmo iterativo para minimizar Δ .

2.6.5 Fase de Aprendizagem

O modelo mais simples a escolher para a relação entre δI e o erro dos parâmetros do modelo é o linear:

$$\delta c = A \delta I. \quad (2.44)$$

Este modelo revelou ser uma boa aproximação para encontrar resultados aceitáveis, [Cootes et al., 1998]. Para encontrar A , aplica-se regressão linear múltipla multi-variada num conjunto de parâmetros δc conhecidos e as correspondentes imagens δI . Pode-se gerar este conjunto de ajustes pela perturbação de parâmetros de modelos de imagens já conhecidas. Estas podem ser imagens do conjunto de treino ou novas imagens criadas com o modelo de aparência construído. No último caso, conhecem-se os parâmetros correctos e as imagens não possuem ruído.

Da mesma forma que existem perturbações nos parâmetros do modelo, também é possível modificar os parâmetros de posição (2D), escala e orientação. Estes quatro parâmetros extra são incluídos na regressão; por uma questão de simplificar a notação, podem ser representados como elementos extra no vector δc . Para manter a linearidade, representa-se a posição e a orientação como (s_x, s_y, t_x, t_y) onde $s_x = s \cos(\theta)$, $s_y = s \sin(\theta)$. Para que se obtenha uma relação “bem comportada”, é importante escolher cuidadosamente a imagem de referência onde a nova imagem é calculada (forma normalizada).

Seguidamente, calcula-se a diferença: Seja c_0 os parâmetros do modelo de aparência conhecidos da imagem. Deslocam-se os parâmetros de uma certa quantidade, δc , para obter os novos parâmetros $c = \delta c + c_0$. Para estes parâmetros, gera-se a forma, x , e os níveis de cinzento normalizados, g_m , usando a equação (2.40) para obter o exemplo normalizado g_s . O erro é então dado por $\delta g = g_s - g_m$.

O algoritmo de treino é simplificado ao aplicar aleatoriamente desvios aos parâmetros do modelo a cada imagem treino e a guardar δc e δg . Em seguida, faz-se a regressão multi-variada para obter a relação:

$$\delta c = A \delta g. \quad (2.45)$$

O melhor intervalo de valores de δc para usar durante o treino é determinado experimentalmente. A relação ideal é aquela que contém o maior conjunto de erros δg possível.

2.6.6 Algoritmo

É possível construir um método iterativo para resolver o problema de optimização, depois de se ter um método que prevê adequadamente a correcção necessária para os parâmetros do modelo. Assim, dada a estimativa inicial dos parâmetros do modelo, c_0 , e o modelo da imagem normalizada, g_s , o procedimento iterativo é o seguinte, [Cootes et al., 1998]:

- 1) *Avaliar o vector de erro $\delta g_0 = g_s - g_m$;*
- 2) *Avaliar o erro $E_0 = |\delta g_0|^2$;*
- 3) *Calcular o deslocamento previsto $\delta c = A\delta g_0$;*
- 4) *Considerar $k = 1$;*
- 5) *Seja $c_1 = c_0 - k\delta c$;*
- 6) *Colocar o modelo na nova posição e calcular o novo vector de erro g_1 ;*
- 7) *Se $|\delta g_1|^2 < E_0$ então aceitar a nova estimativa, c_1 , caso contrário colocar $k = 1.5, k = 0.5, k = 0.25$ e voltar ao passo 5.*

O procedimento anterior é repetido até que não se verifique nenhuma melhoria no erro obtido, $|\delta g|^2$, atingindo deste modo a convergência.

2.7 Resumo

Neste capítulo, abordaram-se os fundamentos teóricos em que se baseiam os modelos pontuais de distribuição, apresentados inicialmente em [Cootes et al., 1992]. Assim, apresentaram-se as diversas etapas necessárias para a construção de um modelo pontual de distribuição, começando pela etiquetagem do conjunto de treino, passando pelo alinhamento do conjunto de treino, e finalizando com o estudo das variações admissíveis do objecto em causa. Descreveu-se também a extracção dos níveis de

cinzento em torno de cada ponto-chave e o cálculo dos perfis de intensidade desses pontos.

Explanaram-se também as diversas etapas da construção do modelo de forma activa, desde o posicionamento inicial do modelo na nova imagem até à conclusão da pesquisa. Nesta fase de pesquisa, descreveu-se o processo iterativo necessário para ajustar os pontos do modelo.

Por fim, apresentou-se a metodologia a usar na construção do modelo de aparência activa. Assim, a construção do modelo de aparência foi descrita, assim como a sua utilização na localização de objectos em novas imagens.

Capítulo 3

Colocação e Emparelhamento Automático dos Pontos-chave

3.1 Introdução

Em todo o processo de construção do modelo de distribuição pontual, a etapa mais morosa é sem dúvida a da colocação e emparelhamento dos pontos-chave nas várias imagens do conjunto de treino. Considerando este facto, optou-se por desenvolver algoritmos que permitissem automatizar este processo para os objectos considerados neste trabalho: a mão e a face.

A primeira fase para obter tais algoritmos, consistiu no estudo bibliográfico de vários processos existentes para tornar a colocação de pontos-chave automática ou semi-automática. Assim, na segunda secção deste capítulo, apresenta-se um resumo dos vários métodos analisados sobre este tema.

Na terceira secção deste capítulo, reúnem-se os fundamentos teóricos necessários para a construção do algoritmo de colocação automática de pontos-chave, desenvolvido para posterior modelação de objectos do tipo mão.

Na secção seguinte, descrevem-se três métodos que permitem automatizar a etapa da colocação de pontos notáveis em faces. Um dos métodos, coloca pontos-chave nas regiões do contorno da face, dos olhos, das sobrancelhas e da boca, e os outros dois consideram os pontos-chave como os nodos de uma malha rectangular colocada adequadamente sobre a face.

Na quinta secção deste capítulo, descrevem-se dois métodos de emparelhamento automático dos pontos-chave, com o intuito de simplificar e automatizar a construção do modelo pontual de distribuição.

Finalmente, na sexta secção, faz-se um resumo dos vários assuntos tratados ao longo deste capítulo.

3.2 Trabalho Relacionado

Para ultrapassar o problema da colocação manual dos pontos-chave e a determinação da correspondência pontual entre instâncias do objecto a modelar, [Hill and Taylor, 1994] apresenta um método para a determinação automática dos pontos (2D) a serem utilizados na geração de modelos pontuais de distribuição, e um método para a determinação da correspondência pontual, sendo apresentados resultados em imagens do coração e da mão. O algoritmo referido é desenvolvido em duas etapas: na primeira, são colocados alguns pontos (posições aproximadas) nos contornos de cada objecto do conjunto de treino, e estabelecidas as correspondências entre pares de instâncias do objecto, utilizando um algoritmo de emparelhamento baseado em programação dinâmica. Assim, o agrupamento de instâncias do objecto em pares permite gerar iterativamente uma forma média, onde qualquer conjunto de pontos colocados nesta forma pode ser projectado para as instâncias do conjunto de treino; na segunda, a localização dos pontos-chave determinados na etapa anterior é afinada: um otimizador não linear ajusta as posições dos pontos, através de um esquema iterativo de maneira a produzir um modelo mais compacto. No entanto, o algoritmo de emparelhamento apresentado, baseado em programação dinâmica, não é suficientemente robusto e preciso para ser de uso geral, para além da sua extensão para o caso de objectos 3D não ser um processo expedito [Hill and Taylor, 1996].

Por outro lado, [Baumberg and Hogg, 1994] descreve um modelo para gerar modelos flexíveis da forma a partir de uma sequência de imagens de pedestres em andamento, sendo os pontos-chave determinados automaticamente. O método considerado, começa por subtrair a imagem de fundo a cada uma das imagens de treino, sendo cada imagem resultante processada de forma a reduzir o ruído e posteriormente binarizada por *threshold* de maneira a obter o contorno do pedestre em análise nesse instante. De seguida, calcula o eixo principal do contorno e identifica um ponto de referência onde este intersecta o mesmo contorno, sendo os pontos-chave determinados a partir deste ponto de referência e igualmente espaçados ao longo do contorno. Embora o processo seja satisfatório para o propósito em causa, a parametrização considerada é arbitrária e apenas são consideradas formas 2D.

Para melhorar o seu algoritmo de colocação automática de pontos-chave anteriormente apresentado, [Hill and Taylor, 1996] propõe um método referenciado pelos autores

como exacto e robusto no estabelecimento de correspondências entre dois contornos fechados. O método em causa constrói uma aproximação poligonal a um dos contornos, gerado a partir do algoritmo de detecção de pontos críticos, e procura emparelhá-lo com o outro contorno através de um esquema de optimização que minimiza uma função de custo. O valor final desta função é usado para aferir sobre a qualidade do ajuste dos contornos envolvidos. Este algoritmo necessita apenas de um parâmetro de controlo associado com a função custo. No trabalho referido, são apresentados resultados para três classes de objectos distintos: mão, ventrículo esquerdo do coração e resistências em circuitos eléctricos.

Tendo por base o estudo anterior, [Brett et al., 1997] apresenta um método para gerar pontos-chave em objectos 3D de forma a construir automaticamente Modelos Pontuais de Distribuição 3D. O algoritmo referido consiste na construção de uma árvore binária de formas combinadas, onde uma vez construída essa árvore é possível identificar um conjunto de pontos notáveis na forma média e propagar estes pontos a todos os “ramos” da árvore. O algoritmo usado na construção desta árvore, necessita de um método de emparelhamento que determine as correspondências entre duas formas, e traduza a qualidade deste ajuste de maneira a determinar quais as formas a emparelhar primeiro. Assim, o método usado consiste na geração de aproximações poliedrais da forma geométrica dos objectos através da utilização da medida de semelhança global Euclidiana entre dois objectos.

Tendo em conta que espécimes biológicos são frequentemente classificados pela sua forma geométrica e que ao longo dos anos se têm realizado inúmeros catálogos com fotografias, desenhos e slides de vários espécimes, [Hicks et al., 2002] apresentam um método para extrair automaticamente os contornos deste tipo de objectos representados em imagens, e utilizam-no para construir Modelos de Forma Activa do grupo taxionómico *diatom*. Assim, consideram as zonas de elevada curvatura dos contornos obtidos a partir das imagens em causa, e utilizam as zonas determinadas no estabelecimento das correspondências entre os pontos-chave do objecto modelado. O método apresentado funciona para o espécime *diatom*, para os quais os pesos e a posição relativa das zonas de curvatura máxima e mínima apresentam variações diminutas. No entanto, é pouco provável que funcione para objectos como a mão, para as quais a variação da forma é considerável.

Posteriormente, em [Angelopoulou and Psarrou, 2004] é apresentado um algoritmo para extrair automaticamente o contorno da mão e construir o respectivo modelo pontual de

distribuição. O processo proposto consiste em três etapas: na primeira, faz a segmentação da imagem e extrai o contorno do objecto em estudo, através de um *threshold* e do filtro de detecção de orlas de intensidade de *Canny* [Deriche 1987; Tavares, 1995]; na segunda etapa, utiliza o *Freeman chain code* [Lim, 1990] para obter a descrição do contorno e assim determinar automaticamente as coordenadas dos *píxeis* e a direcção que o objecto apresenta; na terceira e última etapa, constrói o Modelo de Distribuição Pontual, capaz de descrever a forma geométrica da mão baseado nas coordenadas dos pontos-chave obtidos. O método referido apresenta bons resultados em imagens onde a mão é representada por um contorno fechado e com fundo homogéneo. A detecção automática de pontos notáveis em imagens de faces, como a localização do contorno da face, dos olhos, do nariz ou da boca, também tem sido objecto de estudo de diversos grupos de investigação devido à sua complexidade e utilidade na posterior modelação de faces.

Estudos como os de [Brunelli and Poggio, 1993; Moghaddam and Pentland, 1994], revelam que o reconhecimento de faces baseado em características faciais pode obter bons resultados, considerando que se tornam invariantes a diferentes cortes de cabelo ou maquilhagem, por exemplo. Por seu lado, [Campadelli et al., 2003; Carvalho and Tavares, 2005] descrevem um algoritmo de detecção de zonas de pele que permite localizar, tal como a designação indica, zonas de pele e isolar os objectos que correspondem a faces em imagens. Já em [Gargesha and Panchanathan, 2002; Hsu et al., 2002; Campadelli et al., 2003] são consideradas as características de cor específicas dos olhos e da boca, de maneira a construir mapas cromáticos que indiquem a localização destas entidades nas faces em estudo.

Por outro lado, em [Baker and Matthews, 2004], é considerado o problema da construção automática de Modelos de Aparência Activa como um problema de codificação de imagem. Assim, os autores consideram os pontos-chave como os nodos de uma malha rectangular e constroem o Modelo de Aparência Activa sobre esta malha, apresentando resultados para a modelação de quadrados e de faces.

3.3 Colocação Automática dos Pontos-Chave: Objecto Mão

O algoritmo de colocação automática dos pontos-chave, para posterior modelação da mão, desenvolvido durante este trabalho foi construído tendo por base os métodos propostos em [Angelopoulou and Psarrou, 2004] e [Carvalho and Tavares, 2005].

Assim, a primeira etapa do processo desenvolvido para colocar os pontos-chave, consiste na segmentação do objecto em estudo na imagem a considerar. Tendo em conta que as imagens do conjunto de treino geralmente não possuem um fundo homogéneo, a aplicação usual de um *threshold* seguida de um detector de orlas de intensidade não seria adequada para isolar o objecto em causa. Assim, considerou-se que uma solução interessante seria realizar a detecção de zonas de pele para encontrar o objecto desejado: a mão ou a face.

Na segunda etapa do método desenvolvido, extraem-se os pontos do contorno do objecto e encontram-se as zonas do contorno que possuem elevada curvatura pelo método da *k-curvatura* [Lim, 1990].

Posteriormente, determina-se o contorno final do objecto pretendido; ou seja, como nas etapas anteriores o algoritmo utilizado considera como objecto de interesse as zonas de pele encontradas na imagem, que podem corresponder unicamente à mão ou então a esta mais uma parte do braço, nesta fase extrai-se apenas o contorno correspondente à mão.

A última etapa do método desenvolvido resume-se à extracção dos pontos notáveis do contorno, considerando como pontos de controlo os contidos nas zonas de maior curvatura e repartindo os restantes uniformemente entre cada uma destas zonas. O utilizador da implementação desenvolvida para este método, tem a possibilidade de escolher o número de pontos-chave a reter.

Nas secções seguintes, descreve-se em pormenor cada uma das etapas consideradas no método desenvolvido para colocar os pontos-chave automaticamente em imagens de mãos.

3.3.1 Algoritmo de Detecção de Zonas de Pele

Como já referido, o algoritmo de detecção de zonas de pele foi utilizado com o objectivo de localizar a mão nas imagens a cores a considerar. Para tal, este método utiliza um modelo representativo da pele, construído a partir de um conjunto de imagens

de treino contendo amostras de pele retiradas de diversos pontos da mão de vários indivíduos.

As imagens do conjunto de treino estão representadas originalmente no formato *RGB*, onde as componentes (*R*, *G*, *B*) representam a cor e a luminosidade de cada *pixel* da imagem. Estudos anteriores ([Jones and Rehg, 1999; Zheng et al., 2004]) comprovam que embora a cor da pele abranja uma larga escala de cores no formato *RGB*, a sua variação é muito mais reduzida que a de luminosidade; ou seja, a cor da pele das pessoas é semelhante variando apenas em intensidade. Devido a este facto, o modelo representativo da pele é obtido no espaço cromático: isto é, no formato *YCbCr* (*luminance-chrominance*), onde *Y* representa a intensidade, *Cb* a componente azul e *Cr* a componente vermelha da imagem, com o objectivo de eliminar a influência da luminosidade, Figura 3.1.

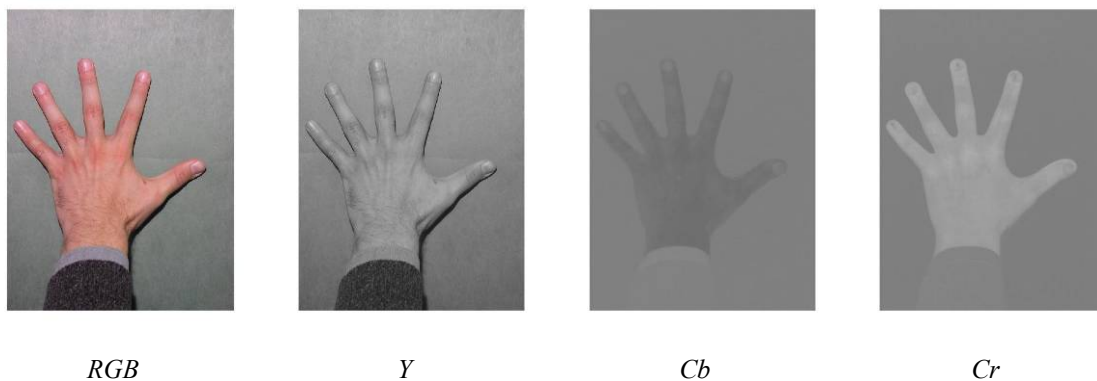


Figura 3.1 Exemplo de conversão de uma imagem no formato *RGB* para o espaço *YCbCr*.

As cores cromáticas são também conhecidas por “cores puras” e são obtidas do formato *RGB* usando a seguinte transformação:

$$\begin{aligned} Cr &= \frac{R}{R+G+B} \\ Cb &= \frac{B}{R+G+B} \end{aligned} \quad (3.1)$$

Esta transformação do formato *RGB* para o espaço cromático é realizada em todas as imagens do objecto a modelar e também nas imagens das amostras de pele, previamente retiradas de diversas zonas da mão para se obter o modelo representativo da pele. Para eliminar algum ruído existente nas componentes *YCbCr*, é aplicado um filtro

passa-baixo em cada uma das imagens do conjunto de treino e em cada uma das imagens de amostras de pele.

Neste trabalho, para desenvolver o modelo representativo da pele, utilizaram-se oito amostras de pele extraídas de várias regiões da mão, de diferentes indivíduos e de tamanho 27×33 *píxeis*, Tabela 3.1.

Tabela 3.1 Amostras de zonas de pele utilizadas para a construção do modelo representativo da pele da mão usado neste trabalho.



Depois de filtrar as imagens das amostras de pele, a função de distribuição da pele pode ser representada por um modelo *Gaussiano* $N(\mu, C)$, onde μ é a média de cada uma das componentes, Cr e Cb , e C a covariância entre estas. Ao obter o modelo representativo da pele, é possível determinar a probabilidade de um dado *pixel* da imagem em estudo pertencer a uma dada região de pele:

$$f(Cr, Cb) = e^{-\frac{1}{2}(x-\mu)^T C^{-1}(x-\mu)}, \quad (3.2)$$

onde $x = (Cr \ Cb)^T$, $\mu = E\{x\}$ e $C = E\{(x-\mu)(x-\mu)^T\}$.

A Figura 3.2 representa graficamente a função de densidade de probabilidade $f(Cr, Cb)$ considerada neste trabalho para o modelo representativo da pele. Os valores médios encontrados para as variáveis Cr e Cb foram respectivamente 152 e 111, valores para os quais a probabilidade de encontrar um *pixel* pertencente a uma região de pele é mais elevada no caso considerado.

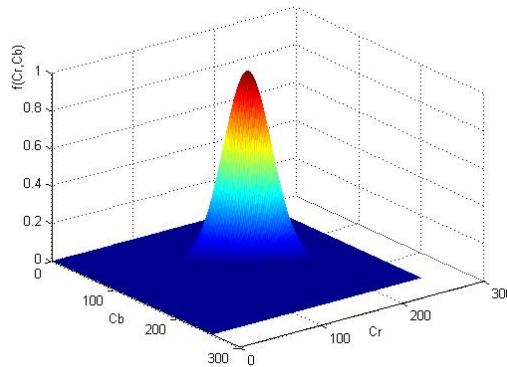


Figura 3.2 Função de densidade de probabilidade considerada para a construção do modelo representativo da pele da mão usado neste trabalho.

Desta forma, o modelo representativo da pele permite transformar uma imagem a cores numa imagem em níveis de cinzento, onde o nível de intensidade de cada *pixel* representa a probabilidade do respectivo *pixel* pertencer a uma dada zona de pele. Visto que as regiões de pele são mais claras que as outras zonas da imagem, as regiões de pele podem ser determinadas na imagem através de um processo de binarização, por exemplo por *threshold*. Assim, seja $g(x, y)$ a imagem em níveis de cinzento, onde x e y correspondem às coordenadas do *pixel* em causa na imagem, e seja T um nível de *threshold* adequado, a imagem binária $b(x, y)$ é obtida segundo a transformação:

$$b(x, y) = \begin{cases} 1 & \text{se } g(x, y) > T \\ 0 & \text{se } g(x, y) \leq T \end{cases} \quad (3.3)$$

Como é necessário segmentar todas as imagens do conjunto de treino, que possuem diferentes tipos de pele, é natural que o valor de *threshold* de cada imagem não seja o mesmo, e daí não ser possível encontrar um valor de *threshold* fixo para todas as imagens. O algoritmo usado aplica então um *threshold* adaptativo que procura o melhor nível de *threshold* para cada imagem.

O processo de *threshold* adaptativo referido baseia-se na observação da variação da região segmentada (que adquire valor 1 à medida em que aumenta o valor do *threshold*). Assim, a região segmentada irá diminuir à medida que o valor de *threshold* se aproxima do máximo, mas irá aumentar bruscamente quando o valor de T for suficientemente reduzido ao ponto de permitir incluir zonas que não são de pele; assim, o valor óptimo de *threshold* será o valor onde a variação do tamanho de região segmentada for mínimo enquanto se diminui o valor de T .

A Figura 3.3, apresenta a aplicação do algoritmo de detecção de zonas de pele numa imagem do conjunto de treino.

O algoritmo descrito permite então obter uma imagem binária com as regiões de pele identificadas, representando o objecto que se pretende analisar. O passo seguinte, para a colocação dos pontos-chave, consiste em obter o contorno da região com interesse identificada na imagem binária.

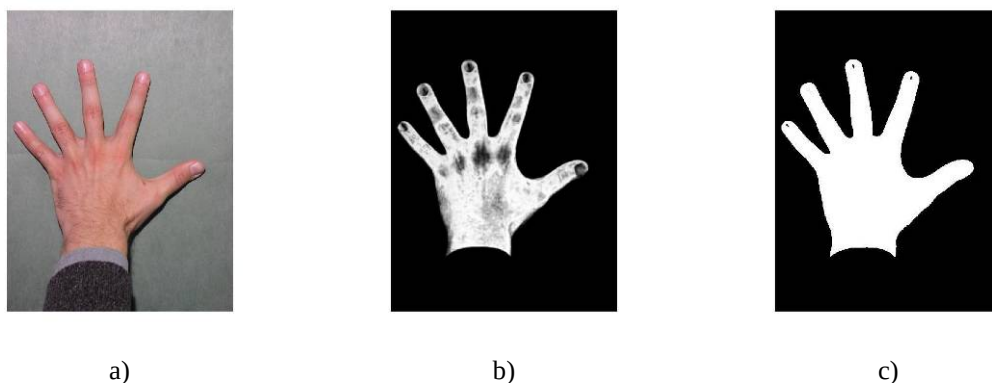


Figura 3.3 Imagem original (a), imagem obtida com a representação das probabilidades de existência de pele na imagem original (b), e imagem após binarização da imagem de probabilidades (c).

3.3.2 Pontos-chave do Contorno

O algoritmo de detecção de zonas de pele obtém uma imagem segmentada que indica a posição do objecto com interesse, a mão, Figura 3.3. Neste trabalho, cada *pixel* da imagem binária possui o valor 1 (branco) no caso de pertencer a uma região de pele, e possui o valor 0 (preto) caso contrário. O objecto em estudo é considerado neste trabalho como formado por um contorno fechado, podendo ou não conter ocos no seu interior. Como neste caso, as imagens do conjunto de treino não apresentam outros objectos, apenas a mão, a região de pele encontrada corresponde exactamente ao objecto a modelar e o número de ocos do objecto encontrado pode ser ignorado. Posteriormente, os pontos do contorno são extraídos considerando uma conectividade 8; ou seja, considera-se que cada *pixel* tem 8 vizinhos como se indica na Figura 3.4.

p1	p2	p3
p4	p0	p5
p6	p7	p8

Figura 3.4 Ilustração da conectividade 8 do *pixel* central p_0 .

Depois de se obter as coordenadas dos *pixels* que constituem o contorno do objecto pretendido, através da etiquetagem da imagem, Figura 3.5, interessa encontrar no contorno obtido os pontos que melhor descrevem o objecto em causa; como por

exemplo, as pontas dos dedos. Para tal, procura-se os pontos do contorno onde a curvatura é máxima.

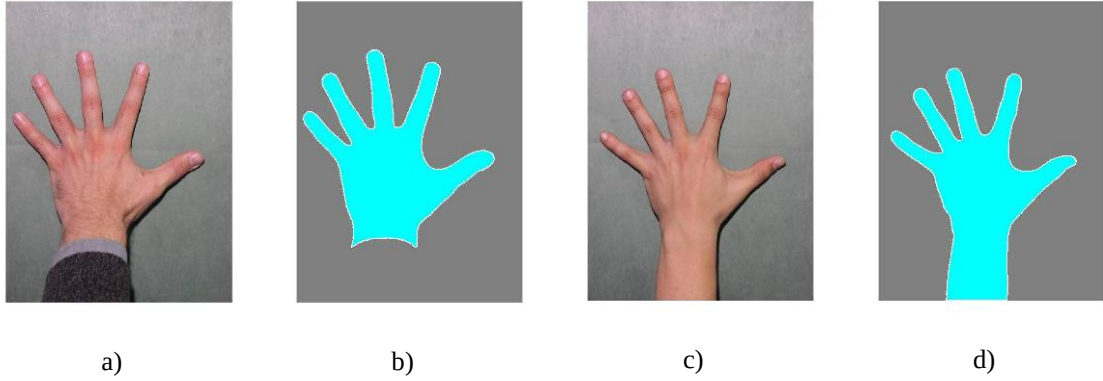


Figura 3.5 Exemplo de extracção do contorno a partir de duas imagens de um conjunto de treino: imagens originais (a, c), objecto determinado em cada uma das imagens consideradas (b, d) (nas imagens de resultados (b, d) o fundo está representado a cinza, o objecto obtido a azul, e o contorno deste a branco).

Como já referido, para calcular as zonas de maior curvatura do contorno utilizou-se a função de *k-curvatura*, utilizada em diversas metodologias como as descritas em [Sheu and Hu, 1999; Wu et al., 2000; Tien et al., 2004]. Assim, para cada ponto do contorno P_i , a sua *k-curvatura* é definida por:

$$c_i(K) = \cos \theta = \frac{\vec{a}_i(K) \cdot \vec{b}_i(K)}{\|\vec{a}_i(K)\| \cdot \|\vec{b}_i(K)\|}, \quad (3.4)$$

onde $\vec{a}_i(K) = (X_i - X_{i+K}, Y_i - Y_{i+K})$ e $\vec{b}_i(K) = (X_i - X_{i-K}, Y_i - Y_{i-K})$, sendo X_i e Y_i as coordenadas do ponto do contorno P_i , θ o ângulo entre $\vec{a}_i(K)$ e $\vec{b}_i(K)$ com K um inteiro positivo, Figura 3.6.

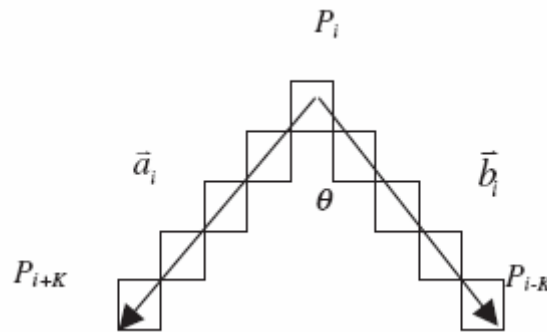


Figura 3.6 Exemplo de cálculo da *k-curvatura* no ponto P_i .

Desta forma, $c_i(K)$ é o co-seno do ângulo entre os vectores $\vec{a}_i(K)$ e $\vec{b}_i(K)$ para um dado *pixel* P_i , variando entre $[-1,1]$ e adquirindo valores positivos quando o ângulo entre estes vectores for inferior a 90° e valores negativos se o ângulo estiver entre 90° e 180° .

Na Figura 3.7, representa-se os valores de *k-curvatura* para cada ponto do contorno de uma imagem do conjunto de treino e as zonas de maior curvatura consideradas. Neste caso, o valor K usado foi *15 píxeis* e considerou-se como zonas de maior curvatura os pontos onde $c_i(K)$ pertence ao intervalo $[-0.75, 0.75]$. Na Figura 3.7a), as zonas relativas aos picos do gráfico correspondem às zonas de curvatura mais elevada, como as pontas dos dedos, as zonas entre os dedos e as próximas do pulso.

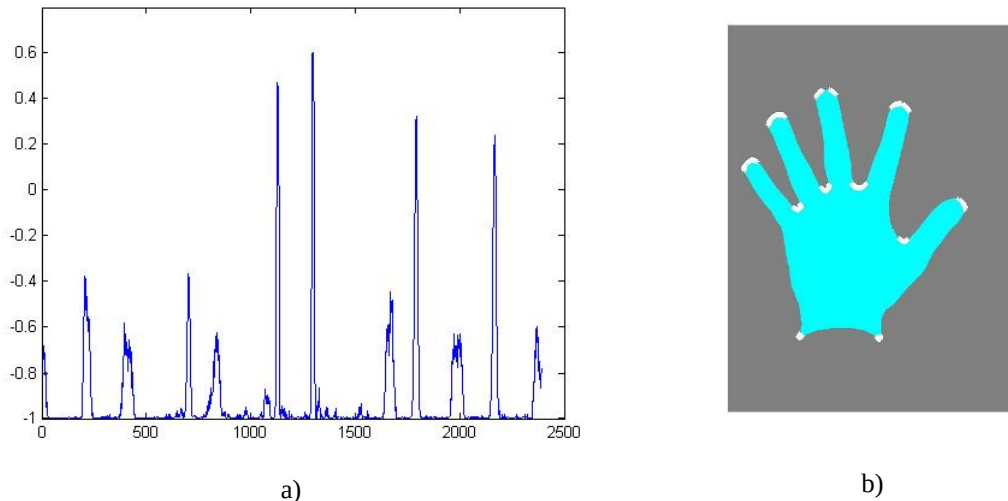


Figura 3.7 Representação da *k-curvatura* (a) e dos pontos de maior curvatura obtidos numa imagem do conjunto de treino (b).

Com a versão actual do algoritmo desenvolvido é possível obter as coordenadas do contorno da mão em todas as imagens do conjunto de treino considerado, assim como extrair as zonas do contorno onde a curvatura é mais elevada. No entanto, no conjunto de treino existem imagens nas quais aparecem, para além da mão, o pulso e o braço (ver, por exemplo, a Figura 3.5a) e c)). Desta forma, o contorno extraído pode não corresponder apenas à zona da mão, o que implica a inclusão de um procedimento suplementar de delimitação do contorno a considerar, descrito na próxima secção.

3.3.3 Delimitação da Mão

Nesta etapa, o objectivo é eliminar os pontos do contorno, obtidos anteriormente, que não pertencem à mão. Visto não haver nenhuma relação no contorno que se possa utilizar para delimitar a zona onde “começa” realmente a mão, a estratégia encontrada para resolver esta situação passa por: encontrar o ponto entre o polegar e o final do pulso, ou entre o dedo mínimo e o final do pulso, que possui maior valor de curvatura; e achar o ponto correspondente no lado oposto.

Assim, o primeiro passo consiste em determinar a que corresponde cada uma das zonas de curvatura de valor mais elevado obtidas. Considerando que a zona lateral direita inferior, segundo as coordenadas, corresponde sempre à zona onde começa o pulso do lado do polegar (*zona 2* - Figura 3.8b)), é então possível etiquetar adequadamente cada uma das restantes zonas.

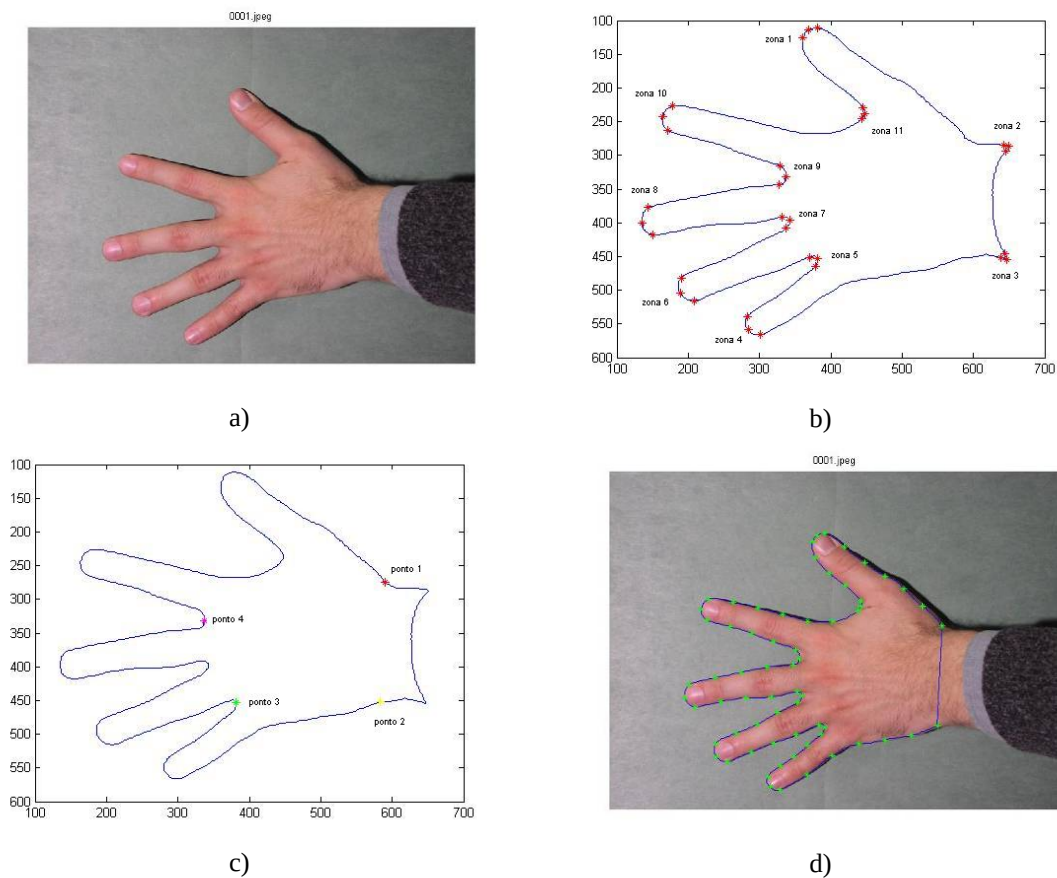


Figura 3.8 Delimitação do contorno numa imagem do conjunto de treino: imagem original (a), etiquetagem das zonas de maior curvatura (b), pontos que delimitam a mão (pontos 1 e 2) (c), contorno final obtido para a mão (d).

Posteriormente, para delimitar a zona da mão, procura-se o ponto do contorno que está entre: a *zona 1* e a *zona 2*, ou a *zona 3* e a *zona 4*, onde o valor da curvatura é mais elevado. O método utilizado para calcular a curvatura é semelhante ao descrito anteriormente: para cada ponto do contorno a curvatura é definida pelo ângulo formado pelas duas rectas, obtidas pelo método dos mínimos quadrados, que melhor aproximam os k pontos antes e depois do ponto em causa.

Depois de determinar o ponto onde o ângulo é mais elevado (*ponto 1* - Figura 3.8c)), resta encontrar o ponto oposto do pulso (*ponto 2* - Figura 3.8c)). Para tal, calcula-se a distância entre o *ponto 1* e o *ponto 3* (Figura 3.8c)) e determina-se o ponto para o qual a distância ao *ponto 4* (Figura 3.8c)) é semelhante à anterior.

3.3.4 Extracção dos Pontos Notáveis

A última etapa, para a colocação automática dos pontos-chave, consiste na escolha do número de pontos a reter para representar o objecto existente em cada imagem, quer nas zonas de valor de curvatura mais elevada, quer entre cada uma dessas zonas.

Com a finalidade de estudar a influência da quantidade de pontos-chave retidos nas zonas de valor de curvatura mais elevada, é possível, ao utilizador da metodologia desenvolvida para o método automático de etiquetagem proposto para objectos do tipo mão, escolher o número de pontos-chave que pretende nestas zonas, assim como entre as mesmas zonas, sendo os pontos distribuídos uniformemente ao longo de cada segmento do contorno, Figura 3.8d).

3.4 Colocação Automática dos Pontos-Chave: Objecto Face

Nesta secção, apresentam-se três algoritmos desenvolvidos ao longo deste trabalho, que permitem automatizar a etapa da colocação de pontos-chave em imagens de faces: o primeiro, considera a detecção de zonas de pele e mapas cromáticos para localizar na imagem o contorno da face, olhos, sobrancelhas e boca; o segundo, considera os pontos-chave como os nodos de uma malha rectangular colocada sobre a zona da face; o terceiro e último, conjuga os dois métodos anteriores, considerando a malha adaptativa à zona da face, mas com maior concentração de nodos na zona dos olhos e da boca. Os

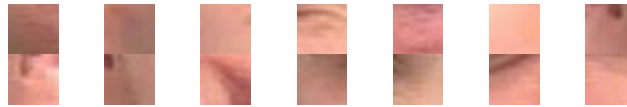
dois últimos algoritmos referidos, serão apenas utilizados neste trabalho na construção dos modelos de aparência activa.

3.4.1 Algoritmo I: Contornos da face

Tal como no caso da mão, o primeiro passo da solução encontrada para a detecção da face em imagens passa por aplicar o algoritmo de detecção de zonas de pele, de forma a obter uma imagem segmentada com as zonas de pele correspondentes.

Para a construção do modelo representativo da pele da face, utilizaram-se neste trabalho catorze amostras de pele extraídas de várias regiões da face, de tamanho 25×25 *píxeis* apresentadas na Tabela 3.2. (Os fundamentos teóricos necessários para a construção deste algoritmo foram já apresentados, na secção 3.3.1).

Tabela 3.2 Amostras de zonas de pele utilizadas para a construção do modelo representativo da pele da face usado neste trabalho.



A função de densidade de probabilidade $f(Cr, Cb)$ considerada para o modelo representativo da pele da face usado está representada na Figura 3.9. Os valores médios encontrados para as variáveis Cr e Cb foram respectivamente 149 e 106, valores para os quais a probabilidade de encontrar um *pixel* pertencente a uma região de pele é mais elevada no caso considerado.

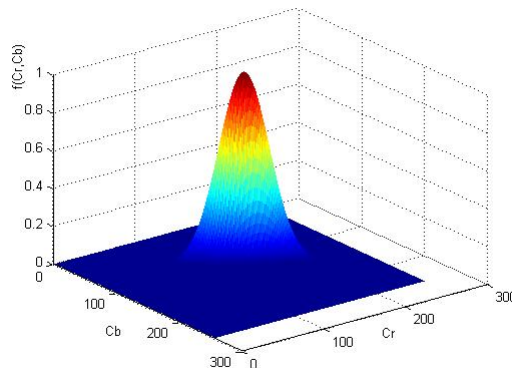


Figura 3.9 Função de densidade de probabilidade considerada para a construção do modelo representativo da pele da face usado.

A Figura 3.10, apresenta a aplicação do algoritmo de detecção de zonas de pele numa imagem do conjunto de treino considerado para a face.

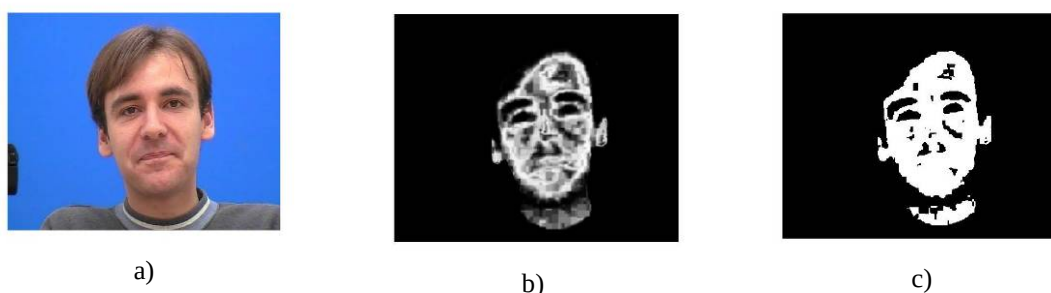


Figura 3.10 Imagem original (a), imagem com a representação das probabilidades de existência de pele (b), e imagem após binarização por *threshold* adaptativo (c).

Depois de se obter a imagem binarizada das regiões de pele, resta determinar qual a região correspondente à face. Assim, a localização do segmento correspondente à face é obtida considerando o número de ocos existentes em cada segmento, pois poderão ser representativos de características faciais como os olhos e as sobrancelhas que não são identificados como regiões de pele [Carvalho and Tavares, 2005] (Figura 3.10c), sendo apenas retido o segmento correspondente à região da face.

Na Figura 3.11, são apresentados os resultados da extracção do contorno da face de duas imagens do conjunto de treino.

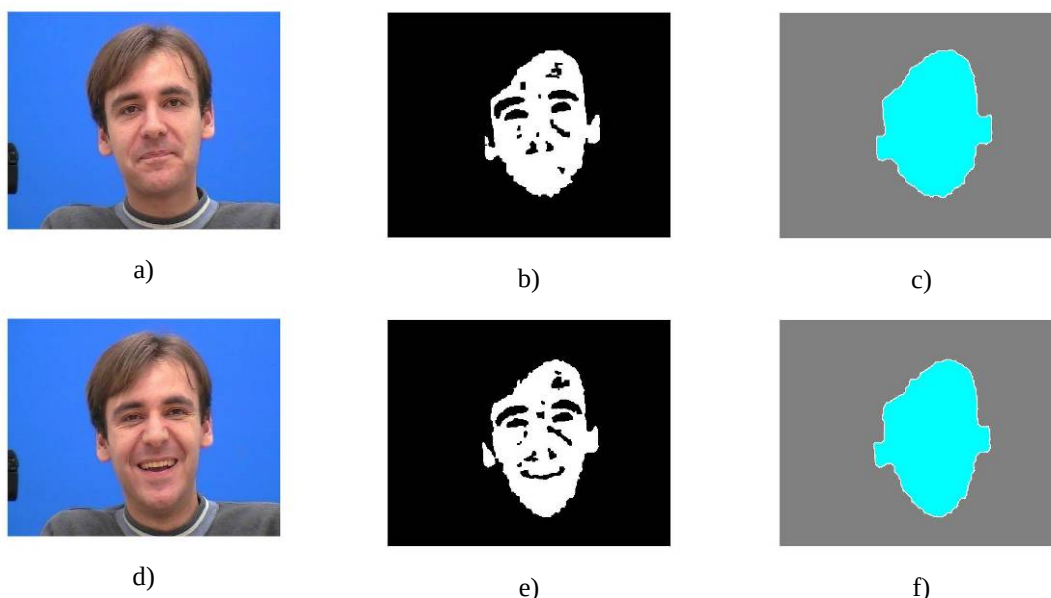


Figura 3.11 Exemplo de extracção do contorno da face a partir de duas imagens de um conjunto de treino: imagens originais (a, d), localização obtida do segmento correspondente à face (b, e), objecto determinado em cada uma das imagens consideradas após a determinação do contorno exterior e posterior suavização do mesmo (c, f). (Nas imagens de resultados (c, f), o fundo está representado a cinza, o objecto a azul e o contorno a branco).

Tendo-se localizado o contorno da face na imagem a considerar, a etapa seguinte do algoritmo prende-se com a localização dos olhos. Para tal, utilizam-se os mapas cromáticos referidos em [Campadelli et al., 2003]. Estes mapas utilizam uma combinação das cores puras da imagem, obtidas como se indica no início deste capítulo aquando da construção do mapa de detecção de zonas de pele (secção 3.3.1.), para a identificação dos olhos.

Assim, os olhos são caracterizados no plano $CbCr$ pelo baixo valor na componente Cr e elevado valor na componente Cb , pelo que o mapa cromático dos olhos considerado é obtido pela transformação:

$$MapaOlhos = \frac{1}{3} \left\{ (Cb^2) + (\hat{Cr})^2 + \left(\frac{Cb}{Cr} \right) \right\}, \quad (3.5)$$

onde Cb^2 , $\hat{Cr}^2$ e Cb/Cr são normalizados ao intervalo $[0, 255]$ e $\hat{Cr}$ é o negativo de Cr , ou seja, $\hat{Cr} = 255 - Cr$.

A transformação cromática descrita permite identificar a zona dos olhos; no entanto, esta transformação nas imagens de treino consideradas neste trabalho permite também localizar a zona das sobrancelhas, como se poderá constatar em seguida. Tal facto não é ignorado neste trabalho; ou seja, a localização das sobrancelhas será também considerada na posterior modelação da face.

Na Figura 3.12, pode-se observar o resultado obtido da aplicação do mapa cromático dos olhos numa imagem de treino, a imagem obtida após a binarização do mapa cromático e a localização dos olhos e das sobrancelhas na face determinadas através da combinação desta última imagem com a imagem que contém o contorno da face (Figura 3.11c).



Figura 3.12 Exemplo de identificação dos olhos e sobrancelhas numa imagem do conjunto de treino através do mapa cromático: imagem original (a), mapa cromático resultante (b), mapa cromático binarizado (c), localização obtida das sobrancelhas (a rosa e a amarelo) e dos olhos (a vermelho e a verde) (d).

Os trabalhos mencionados anteriormente, apresentam também uma transformação cromática capaz de localizar a zona da boca em imagens de faces, partindo do princípio

que a boca é caracterizada no plano $CbCr$ por valores mais elevados na componente Cr e mais reduzidos na componente Cb . Nas imagens de faces utilizadas neste trabalho, esta situação não se verifica como se pode constatar na Figura 3.12a): a cor da boca do indivíduo em causa não é significativamente diferente da cor da sua pele, logo o mapa cromático da boca não conduz a bons resultados de localização. No entanto, se estudarmos as imagens de face sob o formato HSV , onde as componentes (H, S, V) representam a cor, saturação e valor, e se se considerar a componente H da imagem, constata-se que os valores mais elevados desta componente indicam a localização da zona da boca, como se pode observar na Figura 3.13. Através das zonas identificadas a vermelho na Figura 3.13b) e d) retira-se a informação do centro da boca, o centróide, e dos seus extremos. Nos casos nos quais a boca está fechada, os extremos superior e inferior irão coincidir, como se poderá verificar em seguida.

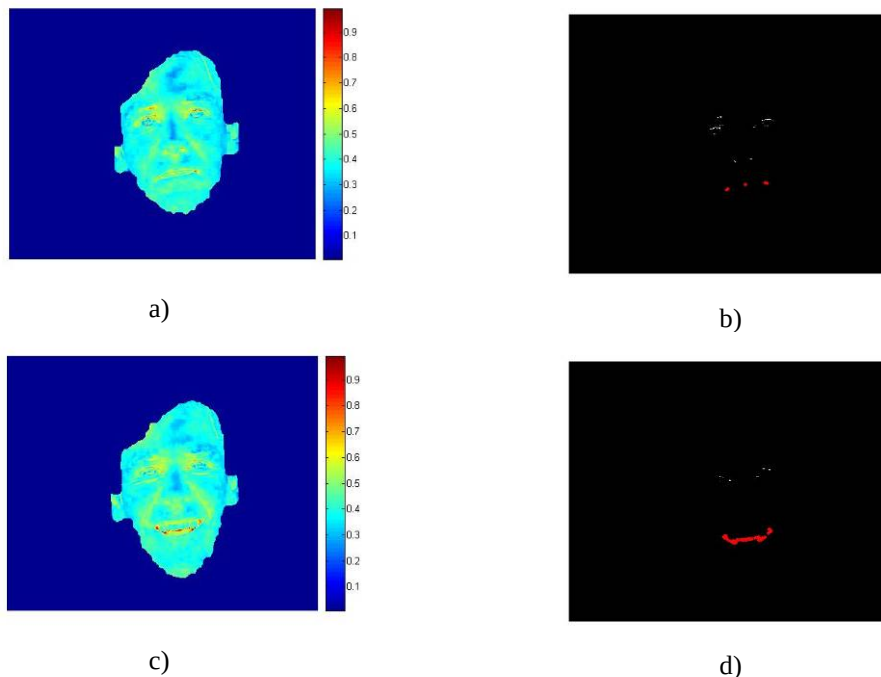


Figura 3.13 Componente de saturação das imagens da Figura 11 (a, d) e identificação (a vermelho) das zonas da boca respectivas (b, d).

Ao conjugar a localização do contorno da face (Figura 3.11), dos olhos e sobrancelhas (Figura 3.12) e da boca (Figura 3.13) obtidos através dos processos descritos anteriormente, é possível retirar os pontos-chave de cada uma das zonas referidas como se pretendia de início. Tendo em conta que a região mais relevante do contorno da face necessária para a sua modelação está situada na zona inferior das orelhas, será esta a

zona utilizada. A referida zona, é posteriormente ajustada por um polinómio de terceiro grau com a finalidade de suavizar o contorno obtido. Assim, os contornos identificados são apresentados na Figura 3.14 para duas imagens do conjunto de treino.

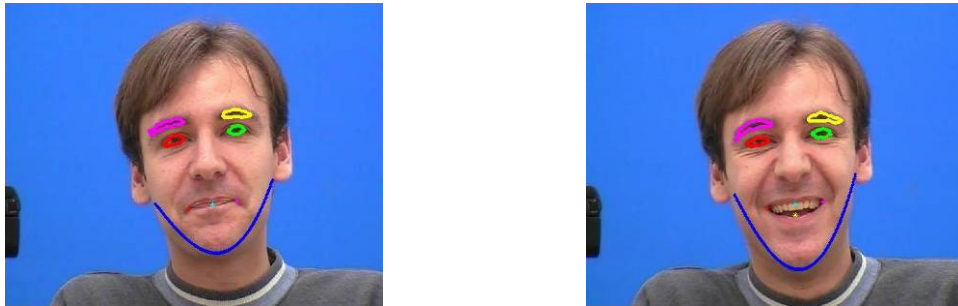


Figura 3.14 Dois exemplos de localização do contorno da face (a azul), sobrancelhas (a rosa e a amarelo), olhos (a vermelho e a verde) e boca (os pontos apresentados indicam os “eixos principais” do contorno da boca).

3.4.2 Algoritmo II: Malha Regular

Em [Baker and Matthews, 2002; Baker and Matthews, 2004] é apresentado um algoritmo capaz de construir automaticamente modelos de aparência activa, considerando o problema da sua construção como um problema de codificação de imagem. O segundo algoritmo desenvolvido neste trabalho, para a extracção automática de pontos-chave em imagens de faces é baseado neste método.

No método de base, os modelos de aparência activa são construídos considerando os pontos-chave como os nodos de uma malha rectangular regular colocada sobre a face a modelar; considerando assim que os pontos-chave não necessitam de ser pontos notáveis do objecto em estudo. Desta forma, o algoritmo usado consiste essencialmente na colocação de uma malha rectangular na imagem.

No algoritmo por nós desenvolvido, a malha utilizada é centrada e rodada adequadamente na zona da face a modelar. Assim, através dos métodos apresentados na secção anterior, localiza-se a região da face e os seus olhos. Posteriormente, as medidas do comprimento e da largura da malha são obtidas a partir do contorno da face determinado, e o seu ângulo de rotação é dado pelo declive da recta formada pelos centróides dos olhos estimados. Desta forma, a malha é colocada adequadamente sobre a face em estudo. Finalmente, para a malha não ficar demasiado justa sobre a face, são

adicionados alguns *píxeis* ao comprimento e à largura da malha. Este número de *píxeis* é definido pelo utilizador da implementação desenvolvida para esta metodologia.

A malha usada, de dimensões $nx \times ny$, é uma malha regular de geometria rectangular que divide o comprimento em nx partes iguais e a largura em ny partes iguais, o que corresponde a $(nx+1) \times (ny+1)$ intersecções da malha, denominados por nodos. Desta forma, na Figura 3.15 apresenta-se uma malha rectangular de dimensões 6×10 , construída para uma imagem do conjunto de treino, que gera automaticamente 77 nodos, ou seja pontos-chave.

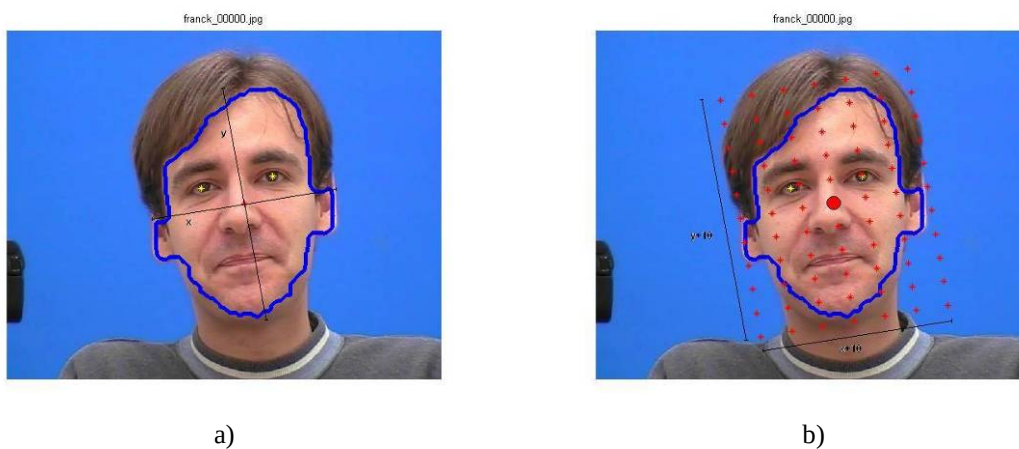


Figura 3.15 Exemplo de colocação da malha rectangular sobre uma imagem do conjunto de treino: contorno da face a azul, centróide da face a vermelho e centróides dos olhos a amarelo (a), malha rectangular 6×10 construída (b) (nodos da malha a vermelho).

Na implementação realizada para este algoritmo, o utilizador tem a possibilidade de escolher os valores nx e ny que pretende para a malha, podendo desta forma controlar o número total de pontos-chave gerados.

3.4.3 Algoritmo III: Malha Adaptativa de Multi-resolução

O terceiro algoritmo, desenvolvido neste trabalho para etiquetar automaticamente faces, foi construído de forma a considerar os pressupostos do primeiro algoritmo apresentado nesta secção, como seja a utilização da localização da face, olhos e boca, e do algoritmo anterior, isto é, a consideração dos pontos-chave como sendo os nodos de uma malha. Este novo algoritmo permite construir uma malha adaptativa de multi-resolução em função da posição da face, dos olhos e da boca.

Na Figura 3.16a) apresenta-se um esquema que traduz o processo de construção da malha adaptativa considerado neste terceiro método. Tal como no algoritmo anterior, a malha é colocada centrada e rodada sobre a face (rectângulo vermelho maior representado na referida figura) de acordo com o declive dado pelos olhos, sendo as suas medidas obtidas a partir do contorno exterior da face. Nesta figura, as zonas delimitadas a preto, indicam a posição das malhas dos olhos e da boca, centradas de acordo com os centróides respectivos. Sendo o comprimento da malha dos olhos corresponde a $2/3$ do comprimento total da face e a largura a $3/10$ da largura total da mesma. No caso da malha da boca, o comprimento corresponde a $1/2$ do comprimento total da face e a largura a $2/10$ da largura total da mesma¹. Para além dos nodos constituintes das malhas dos olhos e boca, este método considera ainda alguns nodos exteriores a estas, indicados a amarelo e a verde no esquema apresentado na mesma figura. O utilizador da implementação desenvolvida para este terceiro método, tem a possibilidade de optar entre a malha *tipo 1* ou a malha *tipo 2*, consoante queira menos ou mais pontos-chave na malha exterior.

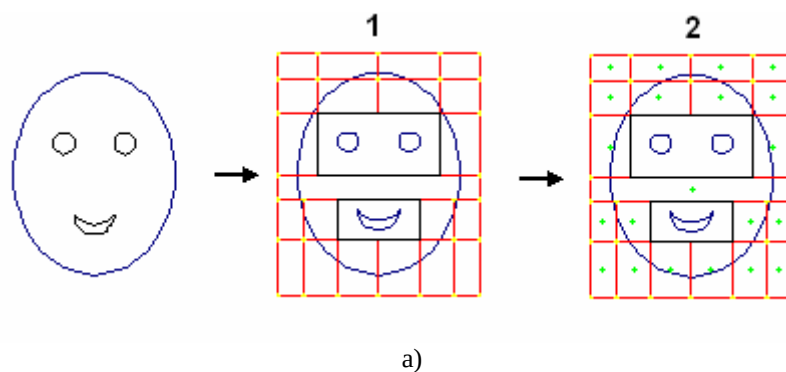


Figura 3.16 Esquema de construção das malhas adaptativas dos *tipos 1 e 2* em (a), exemplo de uma malha adaptativa *tipo 1* (b) e do *tipo 2* (c) aplicada sobre uma imagem do conjunto de treino.

¹ Valores obtidos neste trabalho de forma experimental.

Neste método de etiquetagem automático de faces, o utilizador também tem a possibilidade de escolher os valores nx_o , ny_o , nx_b e ny_b ; ou seja, o número de vezes que pretende dividir a malha dos olhos e da boca respectivamente, além de escolher o tipo de malha adaptativa desejado, *tipo 1* ou *2* como já foi referido. Nas Figura 3.16b) e c), são apresentadas malhas adaptativas obtidas, sendo as malhas dos olhos e da boca de dimensão 5×5 , do *tipo 1* e *2*, respectivamente.

O número total de pontos-chave gerados a partir do algoritmo da malha adaptativa *tipo 1*, é dado pela fórmula $(nx_o+1)*(ny_o+1)+(nx_b+1)*(ny_b+1)+29$. Já o número total de pontos-chave gerados a partir do algoritmo da malha adaptativa *tipo 2*, é dado por $(nx_o+1)*(ny_o+1)+(nx_b+1)*(ny_b+1)+50$.

3.5 Emparelhamento Automático dos Pontos-Chave

Nos casos descritos anteriormente, o emparelhamento dos pontos-chave é obtido automaticamente pela ordem de extracção; no entanto, os algoritmos descritos podem não funcionar para todas as imagens de mãos ou de faces e a colocação manual dos pontos-chave torna-se necessária. Visto a colocação manual dos pontos-chave ser uma tarefa morosa, uma possibilidade de reduzir este tempo é permitir ao utilizador colocar os pontos-chave sem se preocupar com a sua ordem, sendo posteriormente os pontos definidos emparelhados por um processo automático posterior.

Assim, nesta secção serão descritos dois métodos que permitem o emparelhamento automático dos pontos-chave: o método de Shapiro e o método de Sclaroff [Tavares, 2000]. Ambos podem ser utilizados quer nos casos em que a colocação dos pontos-chave é manual sem ordem definida, quer nos casos em que a colocação dos pontos-chave é automática e sem ordem implícita.

3.5.1 Método de Shapiro

Este método, [Shapiro, 1991; Tavares, 2000], recebe como entrada dois conjuntos de pontos: um de dimensão m e outro de dimensão n , respeitantes às formas X_t e X_{t+1} ,

respectivamente. Neste trabalho, a dimensão dos conjuntos será a mesma, no entanto o método funciona também com dimensões distintas.

O primeiro passo deste método de emparelhamento consiste na determinação da matriz de proximidade, H , para cada conjunto de pontos que considera as distâncias mínimas entre os pontos de cada forma. Para a construção da matriz de proximidade H , utiliza-se a medida Gaussiana ponderada:

$$H_{ij} = e^{-\frac{(r_{ij})^2}{2\sigma_x^2}}, \quad (3.6)$$

onde $(r_{ij})^2 = \|X_i - X_j\|^2$ é a distância Euclidiana entre os pontos i e j da forma em questão e σ_x o parâmetro que controla o grau de interacção entre os pontos da mesma. Para valores de σ_x reduzidos, cada ponto tem apenas conhecimento da sua vizinhança local, enquanto que valores de σ_x elevados implica que cada ponto seja influenciado mais globalmente [Tavares, 2000].

O passo seguinte deste método consiste na decomposição da matriz de proximidade em valores e vectores próprios, através da resolução do sistema:

$$H = VDV^T, \quad (3.7)$$

onde a matriz D é a matriz diagonal que contém os valores próprios dispostos por ordem decrescente e V é a matriz ortogonal com os vectores próprios associados a cada valor próprio [Shapiro, 1991].

Por último, determina-se a matriz de correlação, Z , entre os dois conjuntos de vectores característicos, tendo em conta os seguintes aspectos, [Tavares, 2000]:

1. Devido ao número de pontos de cada forma poder ser diferente deve-se considerar, no máximo, $k = \min\{m, n\}$ modos;
2. A fim de considerar que as formas se encontram representados no mesmo sistema de coordenadas globais, para que seja possível a comparação entre ambas, o sinal dos vectores próprios associados à forma X_{t+1} deve ser corrigido face ao sinal dos vectores próprios associados à forma X_t ;
3. A determinação da correspondência entre dois pontos, um de cada forma, é efectuada considerando um dos seguintes critérios [Bastos, 2003]:

- a) Emparelhamento local: o ponto j da forma $t+1$ admite uma correspondência com o ponto i da forma t , se e só se Z_{ij} for mínimo na sua linha e na sua coluna.

Este critério tem a vantagem da sua simplicidade; no entanto, as correspondências, sendo obtidas usando apenas um critério local, não consideram a estrutura global de cada objecto.

- b) Emparelhamento global: problema de programação linear onde o problema de afectação é o seguinte:

$$\begin{aligned} &\text{minimizar } f = \sum_{i=1}^n \sum_{j=1}^n Z_{ij} x_{ij} , \\ &\text{sujeito a } \sum_{i=1}^n x_{ij} = 1, \text{ com } j = 1, \dots, n, \\ &\quad \sum_{j=1}^n x_{ij} = 1, \text{ com } i = 1, \dots, n, \\ &\text{e } x_{ij} \in \{0,1\}, \forall i, j. \end{aligned}$$

Desta forma, são obtidos os emparelhamentos, entre os nodos de cada forma, usando-se um método de optimização global, como por exemplo o método *Simplex*. A vantagem desta solução, é conseguir obter correspondências que melhor satisfaçam todos os nodos das formas em questão, considerando assim a estrutura global de cada objecto.

Na Figura 3.17, apresenta-se um exemplo de resultado de emparelhamento entre duas formas, obtido usando o método de Shapiro resumidamente descrito.

3.5.2 Método de Sclaroff

Este método de emparelhamento permite determinar as correspondências entre pontos de duas formas através de uma modelação física destas, por intermédio de elementos finitos, seguida de uma análise modal para o estabelecimento dos emparelhamentos [Tavares, 2000]. A introdução de características físicas permite, neste método, que cada objecto considerado seja modelado de forma mais realista [Tavares, 2000].

Assim, para a construção do modelo físico de cada forma são atribuídos aos objectos um determinado material virtual caracterizado pelos parâmetros: densidade, ρ , coeficiente de Poisson, ν , e módulo de Young, ξ , que conjuntamente com os dados pontuais que definem as formas, formam a entrada deste método, [Tavares, 2000].

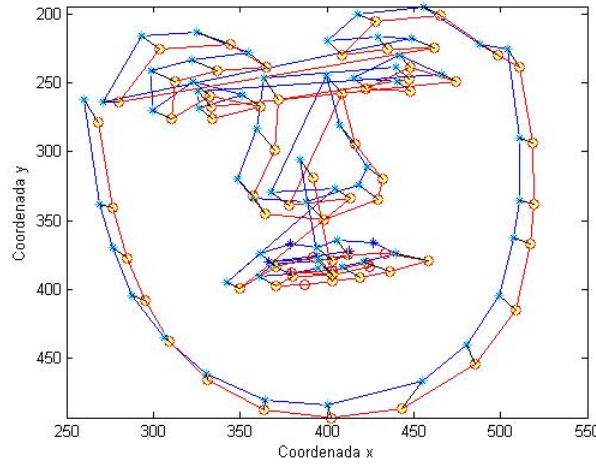


Figura 3.17 Exemplo de emparelhamento automático dos pontos-chave de duas formas obtido usando o método de Shapiro e considerando emparelhamento global.

O método de Sclaroff, [Tavares, 2000], começa por determinar as matrizes de massa, M , e de rigidez, K , dos dois elementos finitos isoparamétricos usados para modelar fisicamente cada uma das formas a emparelhar (um por cada forma). Assim, no caso bidimensional, a matriz de massa é definida por [Sclaroff, 1995; Sclaroff, 1995a; Tavares, 2000]:

$$M = \begin{bmatrix} M_{aa} & 0 \\ 0 & M_{aa} \end{bmatrix}, \quad (3.8)$$

onde $M_{aa} = \rho \pi \sigma^2 H^{-1} \sqrt{H} H^{-1}$ e H representa a matriz de proximidade definida como no método anterior (equação (3.6)).

Por seu lado, a matriz de rigidez é dada por [Sclaroff, 1995; Sclaroff, 1995a; Tavares, 2000]:

$$K = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix}, \quad (3.9)$$

onde os elementos das submatrizes que a compõe, têm a forma:

$$K_{11_{ij}} = \beta\pi \sum_{k,l} a_{ik} a_{jl} \left[\frac{1+\xi}{2} - \frac{\hat{x}_{kl}^2 + \xi \hat{y}_{kl}^2}{4\sigma^2} \right] \sqrt{h_{kl}},$$

$$K_{22_{ij}} = \beta\pi \sum_{k,l} a_{ik} a_{jl} \left[\frac{1+\xi}{2} - \frac{\hat{y}_{kl}^2 + \xi \hat{x}_{kl}^2}{4\sigma^2} \right] \sqrt{h_{kl}},$$

$$K_{12_{ij}} = -\frac{\pi\beta(\alpha+\xi)}{4\sigma^2} \sum_{k,l} a_{ik} a_{jl} \hat{x}_{kl} \hat{y}_{kl} \sqrt{h_{kl}},$$

em que $\hat{x}_{kl} = x_k - x_l$, $\hat{y}_{kl} = y_k - y_l$ e a_{ij} são elementos da matriz H^{-1} .

Após a determinação das matrizes de massa e de rigidez, são calculados os valores e vectores próprios do problema generalizado, $M^{-1}K$, seguindo-se a determinação das correspondências utilizando análise modal, tal como descrito no método anterior [Tavares, 2000].

A título de exemplo, na Figura 3.18 apresenta-se um resultado de emparelhamento de duas formas obtido usando o método de Sclaroff resumidamente descrito nesta secção.

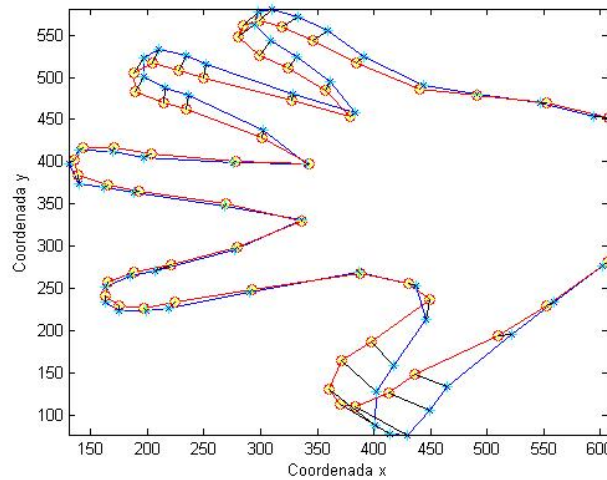


Figura 3.18 Exemplo de emparelhamento automático dos pontos-chave de duas formas obtido usando o método de Sclaroff considerando emparelhamento global.

3.6 Resumo

Neste capítulo, abordaram-se as diversas metodologias existentes para a colocação de pontos-chave, e apresentaram-se novos algoritmos desenvolvidos ao longo deste trabalho capazes de etiquetar automaticamente objectos do tipo mão e face. Assim, um destes algoritmos desenvolvidos destina-se à etiquetagem da mão e os restantes três à

etiquetagem automática de faces. Os referidos métodos mostraram experimentalmente serem eficazes para os objectos em questão. A utilização destes métodos automáticos de etiquetagem, permite que o processo de construção dos modelos de forma activa e de aparência activa seja completamente automático para os tipos de objectos considerados. Apresentaram-se ainda dois métodos distintos de emparelhamento automático de pontos-chave, que facilitam o processo de construção dos modelos pontuais de distribuição caso as correspondências entre pontos-chave não seja dada de forma implícita pela sua ordem.

No próximo capítulo, serão apresentadas as implementações desenvolvidas de raiz ao longo desta Dissertação referentes à colocação automática de pontos-chave, e também as implementações adaptadas de outros autores relativas ao processo de emparelhamento automático e à construção dos modelos de forma e de aparência activa.

Capítulo 4

Implementações Desenvolvidas, Adaptadas e Utilizadas

4.1 Introdução

Neste capítulo, apresentam-se as implementações desenvolvidas e adaptadas em *MATLAB* para a colocação e emparelhamento automático dos pontos-chave, assim como as adaptações efectuadas nos programas, também em *MATLAB*, que foram utilizados para a construção de Modelos de Forma Activa e de Aparência Activa.

Os métodos para a colocação automática e emparelhamento de objectos do tipo mão e face, descritos no capítulo anterior, foram implementados durante este trabalho de raiz em *MATLAB*. Já a metodologia de detecção de zonas de pele considerada, teve por base o algoritmo implementado por [Carvalho and Tavares, 2005], e os métodos de emparelhamento automático dos pontos-chave foram neste trabalho desenvolvidos a partir da implementação previamente realizada em [Silva, 2002] e [Tavares, 2000].

Como já referido, para o estudo do modelo de forma activa foram desenvolvidos algoritmos em *MATLAB*, tendo por base o programa de Modelo de Forma Activa disponibilizado por [Hamarnah, 1999a]. Previamente, foram realizados vários ensaios experimentais para validar o programa base, sendo os resultados obtidos satisfatórios.

Para o estudo do modelo de aparência activa, utilizou-se uma implementação do Modelo de Aparência Activa disponibilizada por [Cootes, 2004]. A referida implementação, foi também validada durante este trabalho através da realização de vários ensaios experimentais.

Os algoritmos por nós desenvolvidos de raiz para etiquetar automaticamente os pontos-chave, foram neste trabalho utilizados para gerar automaticamente modelos de dois tipos de objectos, respectivamente, mão e face. Considerando que o tema principal deste trabalho era centralizado no modelo de distribuição pontual, modelo de forma activa e modelo de aparência activa e no estudo das suas diferenças, considerou-se que a utilização deste tipo de objectos permitiria um estudo adequado sobre cada um destes modelos.

A plataforma computacional de desenvolvimento escolhida, neste trabalho para a implementação da etiquetagem automática e dos modelos de forma activa, foi o *MATLAB 7*, na sua versão *7.0.0.19920 (Release 14)* para sistemas *Microsoft Windows*.

4.2 Modelo Pontual de Distribuição

Tal como foi referido no segundo capítulo desta Dissertação, o processo de construção do modelo de distribuição pontual é constituído por três etapas: na primeira, procede-se à etiquetagem do conjunto de treino; na segunda, processa-se o alinhamento das instâncias do conjunto de treino; e na terceira, determinam-se as variações admissíveis para a forma do objecto considerado. Nesta secção, descreve-se o funcionamento das implementações construídas ou adaptadas para as referidas etapas e ainda, na quarta secção, discute-se a consideração do perfil de intensidade do modelo de distribuição pontual.

4.2.1 Etiquetagem do Conjunto de Treino

A primeira etapa de todo o processo de modelação seguido consiste em etiquetar todos os objectos do conjunto de treino. Assim, as imagens consideradas foram etiquetadas automaticamente, de acordo com os métodos apresentados no capítulo anterior.

4.2.1.1 Etiquetagem Automática da Mão

Tal como já se referiu, foi implementado de raiz um algoritmo que extrai automaticamente os pontos-chave de mãos representadas em imagens. A sua utilização é simples e intuitiva, sendo apenas necessário introduzir alguma informação para a obtenção de uma matriz contendo as coordenadas dos pontos determinados em cada imagem do conjunto de treino e da localização das imagens consideradas. Neste trabalho, as 25 imagens do conjunto de treino considerado estão numeradas para maior facilidade de construção das matrizes finais, sendo a designação dos ficheiros do tipo “mao_1.jpeg”, “mao_2.jpeg”, ..., “mao_25.jpeg”.

Assim, para etiquetar as imagens de treino o utilizador apenas necessita de digitar o comando *hand_contour* na linha de comandos do *MATLAB*, para ter acesso ao menu apresentado na Figura 4.1. Usando o referido menu, o utilizador deverá definir o caminho onde estão localizadas as imagens a etiquetar, o nome das imagens (sem incluir a numeração), o número total de imagens a considerar e o número de pontos-chave que pretende nas zonas de maior curvatura e entre essas zonas. No final da introdução dos parâmetros de utilização do algoritmo o utilizador deverá ainda indicar onde quer guardar o ficheiro de saída com a matriz de coordenadas (Xu) e de localização das imagens (*TrnImgFiles*), Figura 4.1.

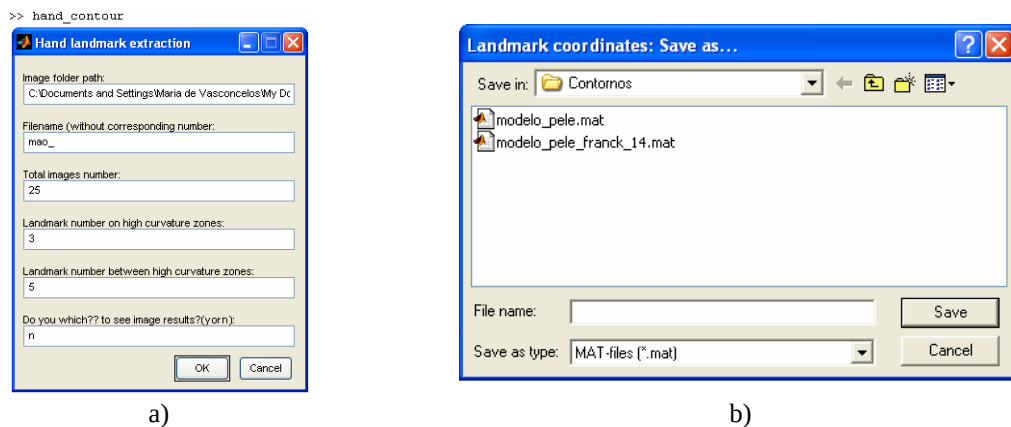


Figura 4.1 Implementação do algoritmo de etiquetagem automática de objectos do tipo mão: menu de entrada de definição dos parâmetros do algoritmo (a), menu para guardar o resultado final obtido (b).

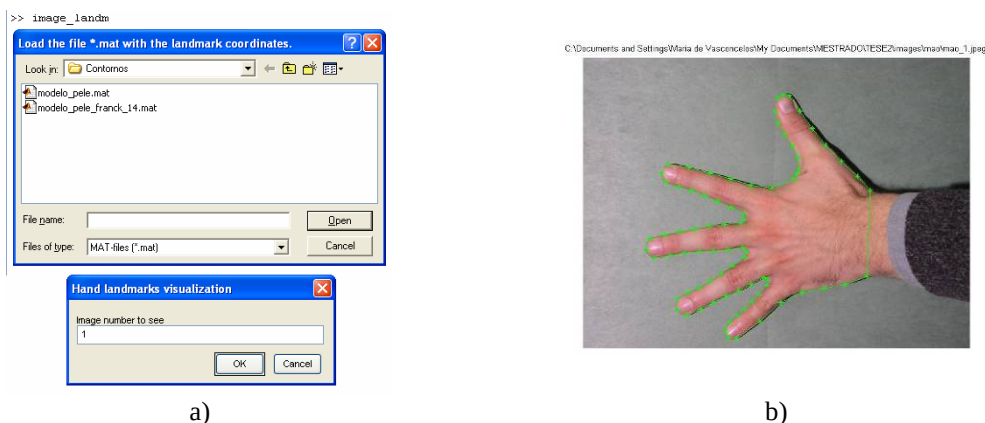


Figura 4.2 Interface da implementação desenvolvida para a visualização dos pontos-chave determinados (a), e exemplo de uma imagem de resultados obtida (b).

Caso o utilizador deseje visualizar as imagens de resultados, como as imagens de probabilidade de zonas de pele, do contorno da mão extraído, das zonas de maior curvatura e da respectiva imagem final etiquetada, deverá inserir *y* no último campo do

menu de parametrização do algoritmo, Figura 4.1a), caso contrário mantém a opção *n*. Se apenas desejar visualizar as imagens etiquetadas, bastará digitar *image_landm* na linha de comandos, indicar o ficheiro onde foram gravados os resultados anteriores e o número da imagem etiquetada que deseja visualizar, Figura 4.2.

4.2.1.2 Etiquetagem Automática de Faces

No capítulo anterior, foram apresentados três algoritmos desenvolvidos de raiz durante este trabalho para etiquetar automaticamente imagens de faces. Em seguida, apresentam-se as implementações desenvolvidas para esse efeito.

Algoritmo I: Contornos da Face

Este algoritmo, encontra automaticamente os contornos das sobrancelhas, dos olhos, da boca e da parte inferior da face representada numa imagem como se verifica na Figura 4.3a). A implementação desenvolvida para este algoritmo extrai por defeito dezoito pontos-chave dos contornos referidos, como apresentado na Figura 4.3b). Para além destes pontos, o utilizador pode ainda escolher o número de pontos que quer colocar entre cada extremidade das sobrancelhas, olhos e face.



Figura 4.3 Exemplo dos contornos extraídos de uma imagem de face do conjunto de treino (a) e dos pontos-chave automaticamente extraídos (b).

Assim, para etiquetar a face usando este algoritmo o utilizador deve digitar na linha de comandos do *MATLAB* o comando *method1* e tem acesso ao menu apresentado na Figura 4.4a. Usando o referido menu, deverá indicar a imagem que pretende etiquetar,

assim como os parâmetros referentes ao número de pontos a considerar entre as extremidades. O vector obtido corresponde às coordenadas dos pontos-chave do contorno determinados, como representado na Figura 4.4.

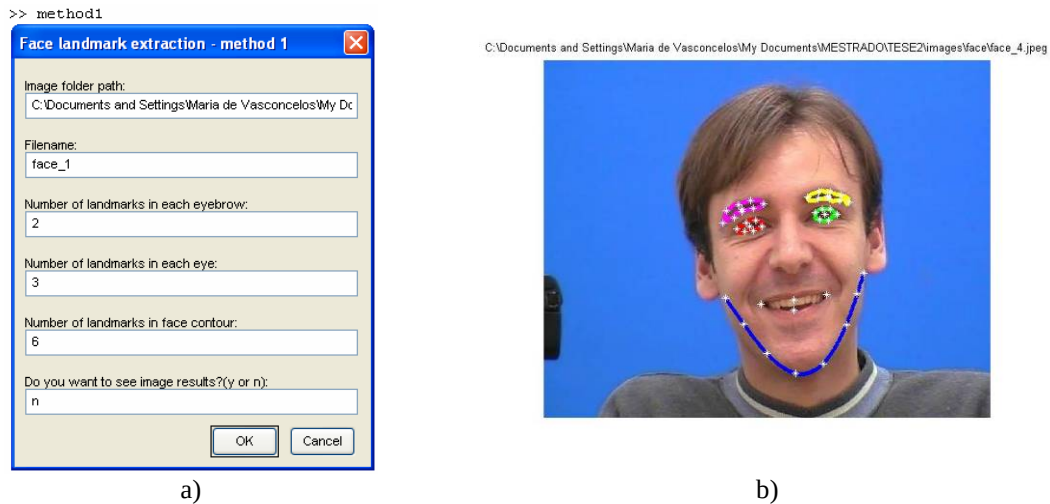


Figura 4.4 Menu de entrada da implementação desenvolvida para o algoritmo I para etiquetar automaticamente objects do tipo face (a) e exemplo dos pontos-chave automaticamente extraídos (b) (a branco).

Algoritmo II: Malha Regular

O segundo algoritmo apresentado no capítulo anterior para etiquetar automaticamente as imagens de faces, detecta a zona da face na imagem e adapta uma malha rectangular à mesma. Na implementação desenvolvida para este algoritmo, o utilizador apenas necessita de indicar a imagem que pretende considerar e escolher o número de divisões da malha pretendido (n_x - comprimento e n_y - largura), Figura 4.5a). Como resultado final, obtém o vector com as coordenadas dos nodos da malha construída, Figura 4.5b).

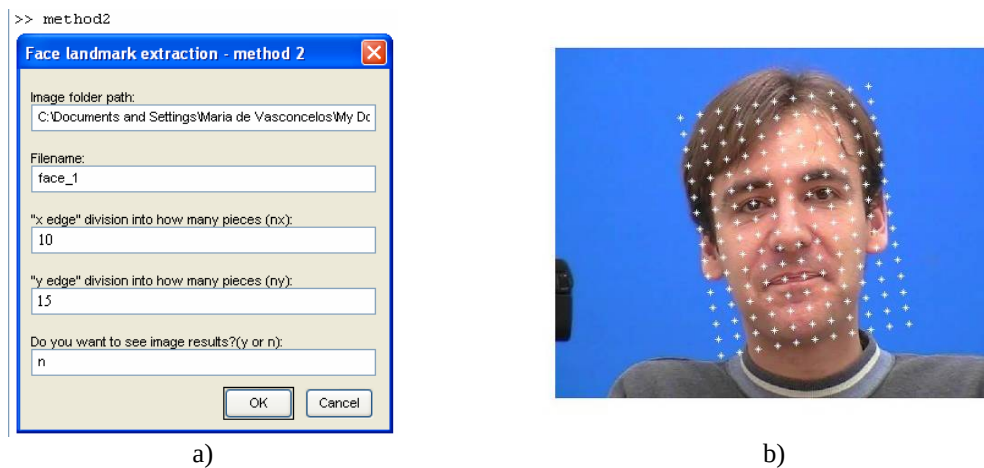


Figura 4.5 a) Menu de entrada da implementação desenvolvida para o algoritmo II para etiquetar automaticamente faces (a) e exemplo dos pontos-chave automaticamente extraídos (a branco) (b).

Algoritmo III: Malha Adaptativa de Multi-resolução

Este algoritmo permite colocar uma malha adaptativa sobre a zona da face numa imagem, tal como referido anteriormente na secção 3.4.3. Para usar a implementação respectiva, o utilizador apenas necessita de indicar a imagem a considerar, o tipo da malha que pretende (1 ou 2) e o número de divisões da malha dos olhos e da boca pretendido, Figura 4.6a), obtendo-se como resultado o vector com as coordenadas dos nodos da malha determinada, Figura 4.6b).

a)



b)



c)

Figura 4.6 Menu de entrada da implementação desenvolvida para o algoritmo III para etiquetar automaticamente faces (a) e exemplo dos pontos-chave automaticamente extraídos (a branco), considerando malha do tipo 1 (b), e do tipo 2 (c).

4.2.1.3 Etiquetagem Manual

No caso das implementações anteriores falharem na colocação automática de pontos-chave para as imagens de mãos ou de faces, ou caso o utilizador pretenda modelar outro tipo de objectos, para os quais as implementações desenvolvidas não sejam adequadas, foi desenvolvida uma implementação que possibilita a sua etiquetagem manual. Para tal o utilizador deverá digitar o comando *label*, na linha de comandos do *MATLAB*, que acciona o respectivo menu, Figura 4.7a), no qual o utilizador deverá indicar o número total de imagens que pretende etiquetar e o número de pontos-chave a definir. Depois de escolher a imagem que pretende, esta torna-se visível, Figura 4.7b), e deverão ser

definidos sobre a mesma os pontos-chave a considerar por intermédio do rato (no caso de se querer apagar o último ponto-chave inserido, deve-se carregar na tecla *delete*).

No final da definição dos pontos-chave, o utilizador tem ainda a possibilidade de gravar as informações armazenadas, tal como acontece nas implementações desenvolvidas para as metodologias automáticas.

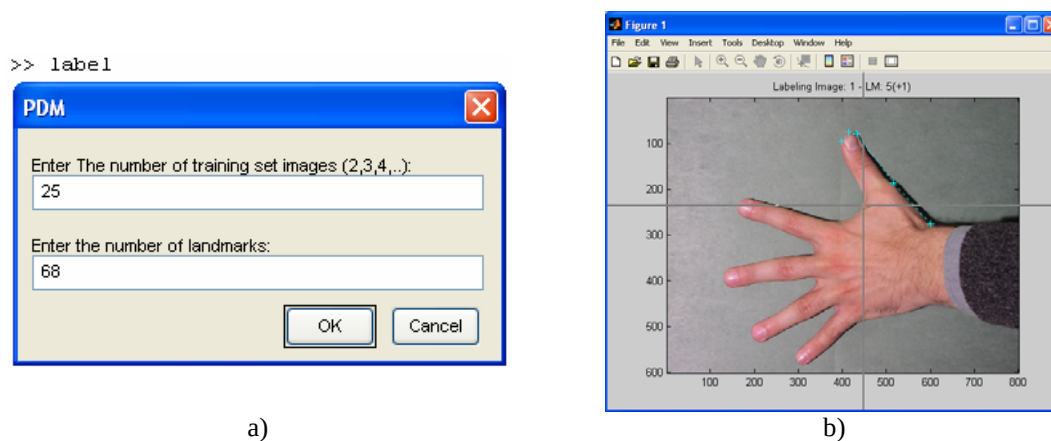


Figura 4.7 Menus da implementação desenvolvida para a etiquetagem manual: menu inicial (a), definição dos pontos-chave sobre a imagem a considerar (b).

4.2.1.4 Emparelhamento Automático dos Pontos-chave

O emparelhamento automático dos pontos-chave é uma mais valia na automatização do processo de etiquetagem, pois garante a correcta ordenação dos pontos-chave, extraídos previamente por qualquer método.

No capítulo anterior, foram resumidamente apresentados dois métodos que permitem o emparelhamento automático dos pontos-chave de dois conjuntos de formas: o método de Shapiro e o método de Sclaroff. A implementação desenvolvida para o efeito, que teve por base as implementações de [Silva, 2002], necessita dos seguintes parâmetros de entrada: para a modelação das formas, os dois conjuntos de dados pontuais a emparelhar (Figura 4.8) e o valor de sigma a utilizar como medida de interacção entre nodos vizinhos das mesmas; e para o estabelecimento dos emparelhamentos, o primeiro modo a considerar, a percentagem dos modos próprios a usar, o valor mínimo a utilizar na definição de casos de simetria e o valor limiar de forma a aceitar cada emparelhamento determinado como aceitável, (Figura 4.9). No caso do método de Sclaroff, é necessário

também introduzir as características do material virtual adoptado para o objecto: a densidade, o coeficiente de Poisson, o módulo de Young e a escala a considerar.

Para a determinação do emparelhamento entre os conjuntos de pontos, o utilizador tem ainda a possibilidade de escolher o critério desejado para a determinação das correspondências: emparelhamento local ou emparelhamento global, usando neste último caso o método *simplex*, tal como referido na secção 3.5 do capítulo anterior.

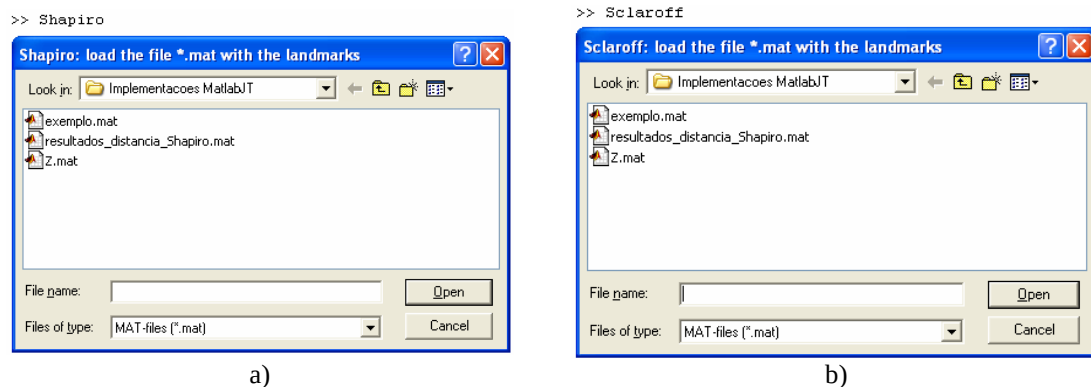


Figura 4.8 Implementações adaptadas para emparelhar automaticamente dois conjuntos de pontos-chave: método de Shapiro (a) e método de Sclaroff (b).

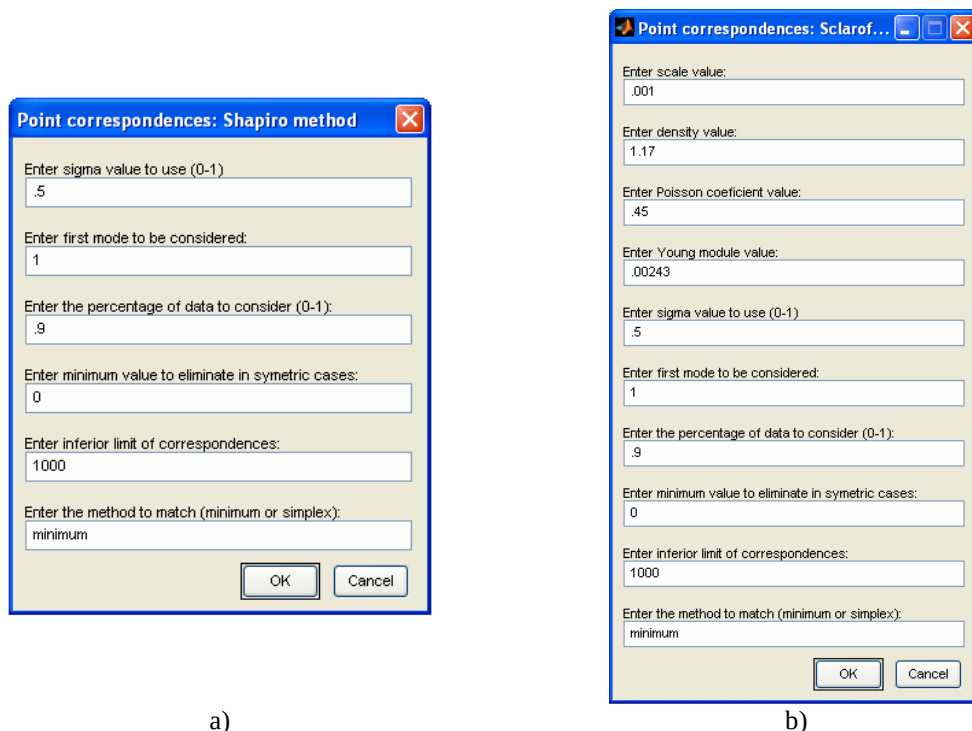


Figura 4.9 Definição dos parâmetros de entrada necessários para as implementações adaptadas para o método de Shapiro (a) e para o método de Sclaroff (b).

A título de exemplo, na Figura 4.10 apresentam-se os resultados obtidos para o emparelhamento dos pontos-chave usando respectivamente o método de Shapiro e de

Sclaroff, e considerando emparelhamento global, para objectos do tipo face e mão. É possível constatar que ambos os algoritmos efectuem o correcto emparelhamento entre os dois conjuntos de pontos considerados.

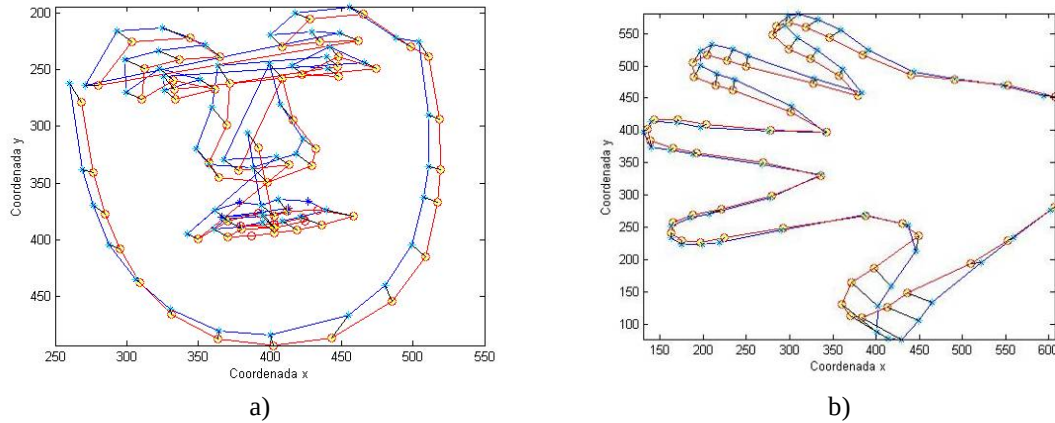


Figura 4.10 Exemplo de emparelhamento automático dos pontos-chave de dois objectos obtido usando o método de Shapiro (a) e o método de Sclaroff (b). (Nestas imagens os sub-contornos estão ligados entre si por segmentos de recta, assim como os nodos vizinhos e os emparelhamentos obtidos.)

4.3 Modelo de Forma Activa

A implementação adoptada neste trabalho para construir os modelos de forma activa é constituída por três programas distintos, para uma melhor funcionalidade: o primeiro, denominado por *training_stage*, permite construir o modelo de forma activa a partir das coordenadas dos pontos-chave do objecto; o segundo, denominado por *variation_modes_stage*, permite ver os modos de variação do modelo construído, e o terceiro programa, denominado por *search_stage*, permite identificar um novo objecto a partir do modelo previamente construído durante a fase de treino.

Para a fase de treino, é necessário indicar o ficheiro que contém as coordenadas dos pontos-chave das instâncias do objecto e a localização das imagens respectivas, sendo o ficheiro construído a partir de um dos métodos referidos na secção 4.3.1, Figura 4.11a). Para além disso, o utilizador tem a possibilidade de escolher: o método de alinhamento dos objectos a usar, *método A* (pelo método dos mínimos quadrados ponderados) ou *método B* (que considera todos os pontos-chave com o mesmo peso, método este desenvolvido de raiz durante este trabalho); a percentagem de variância explicada pelo modelo; e a dimensão do perfil de intensidade a utilizar Figura 4.11b). O modelo final

obtido pode ser guardado num ficheiro, com extensão *mat*, para posterior visualização dos modos de variação ou utilização na fase de pesquisa.

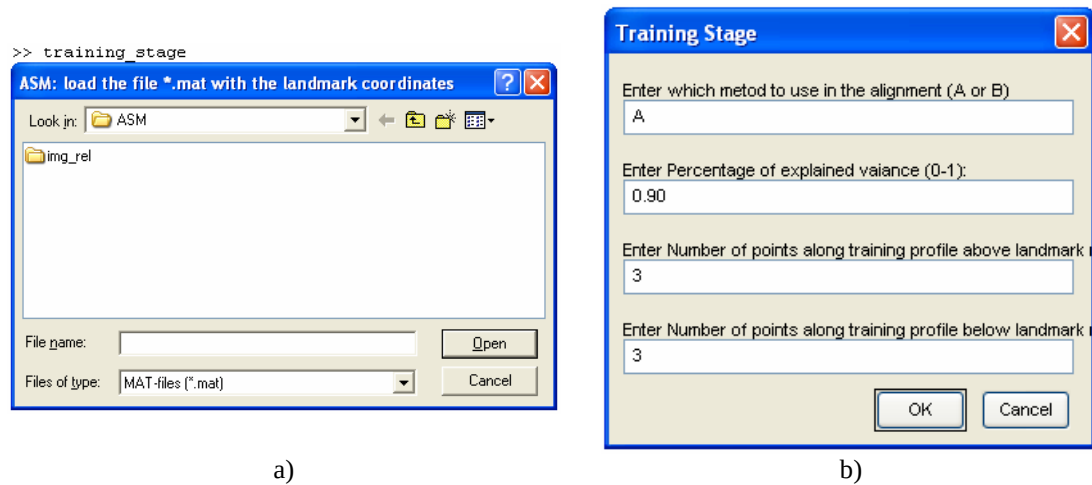


Figura 4.11 Menus de entrada do programa computacional usado para a fase de treino dos modelos de forma activa (a, b).

Na segunda fase de utilização da referida aplicação, é possível visualizar os modos de variação retidos pelo modelo construído na fase de treino; assim, o utilizador tem a possibilidade de observar facilmente o efeito que cada modo de variação produz sobre a forma média obtida. Na Figura 4.12 apresentam-se os menus de entrada para a fase de visualização dos modos de variação. Nesta fase, é exibida uma sequência de imagens da variação de cada modo.

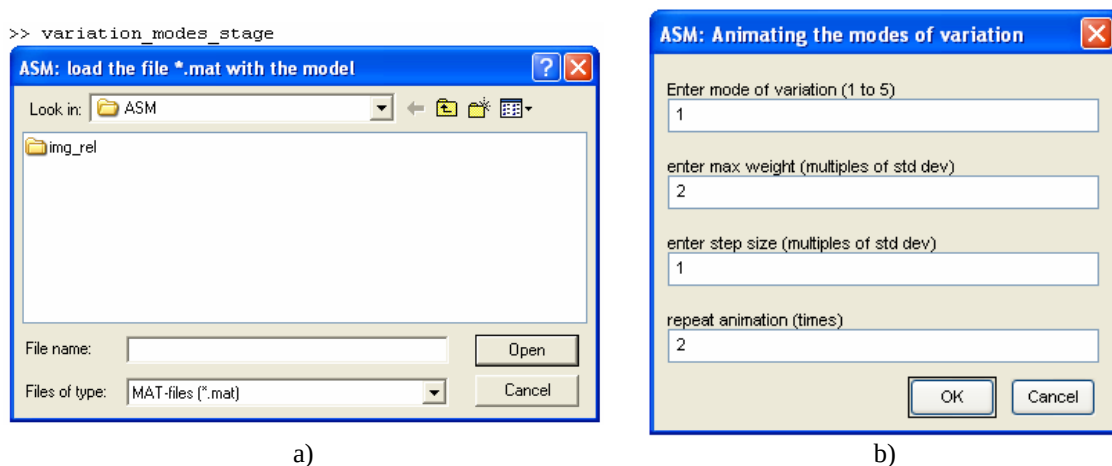


Figura 4.12 Menus de entrada do programa computacional usado para visualizar os modos de variação do modelo activo da forma construído na fase de treino (a, b).

Finalmente, na fase de pesquisa o utilizador indica o modelo a utilizar e a nova imagem na qual pretende identificar (segmentar) o objecto modelado, Figura 4.13a). No programa usado, é também possível alterar a posição do modelo médio no início da

pesquisa por ajuste dos parâmetros de escala, rotação e translação. Nesta fase, o utilizador pode indicar ainda qual o número máximo de iterações que pretende em cada nível de resolução e a dimensão do perfil de pesquisa, Figura 4.13b).

Na Figura 4.14a), apresenta-se um exemplo dos resultados obtidos na segunda fase, a fase de visualização dos modos de variação; e na Figura 4.14b), apresenta-se um exemplo dos resultados de segmentação obtidos para uma nova imagem de teste usando um modelo previamente construído.

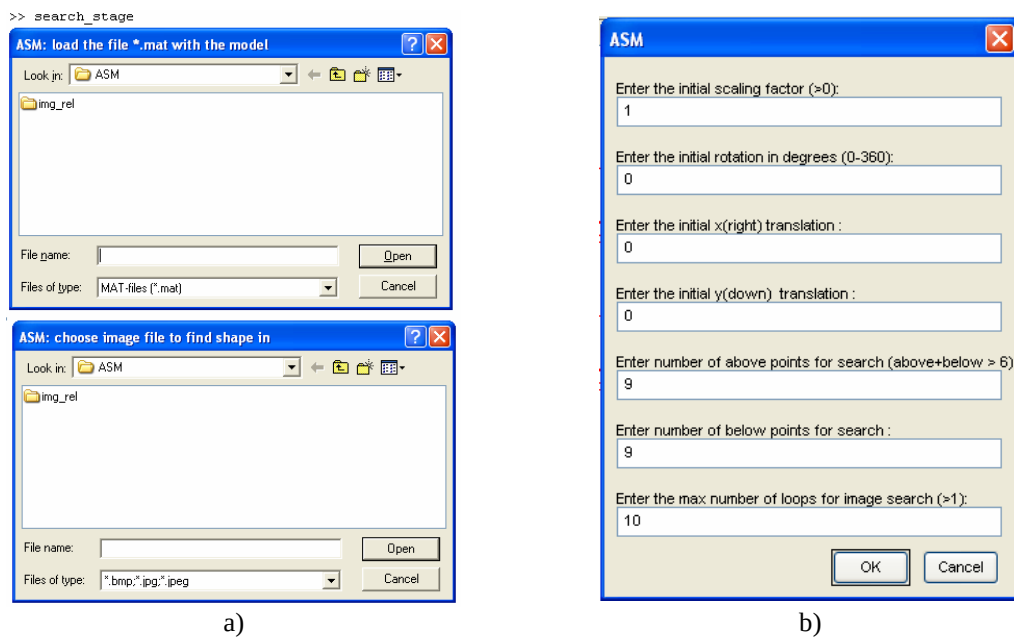


Figura 4.13 Menus de entrada do programa computacional usado para a fase de pesquisa dos modelos de forma activa (a, b).

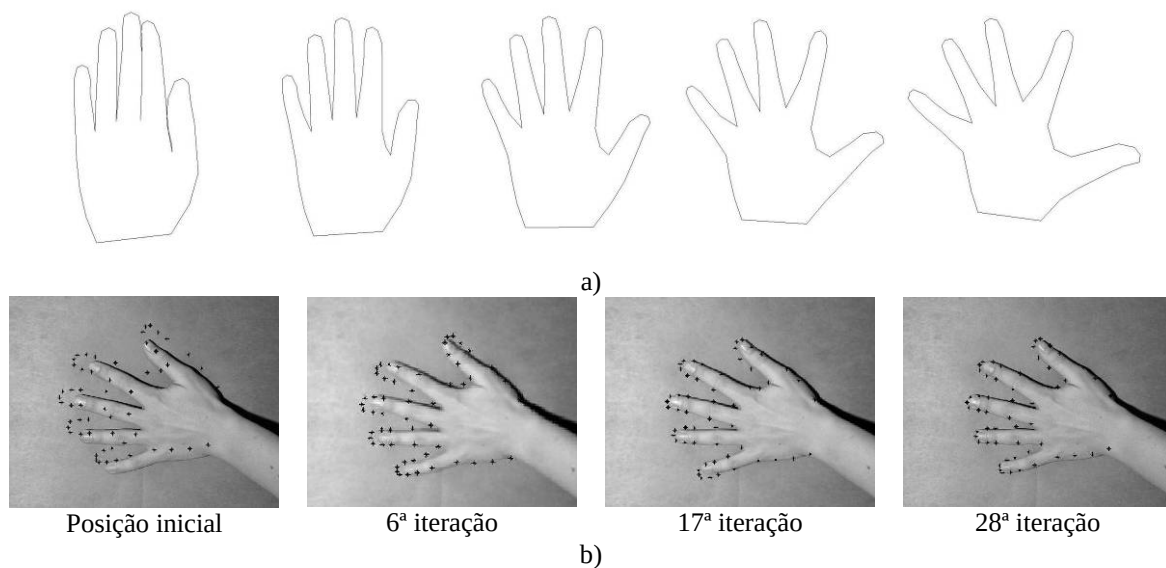


Figura 4.14 Exemplo do efeito de variação de um modo de um modelo da forma construído para a mão (a); imagem de teste com a posição do modelo médio sobreposto, e após a 6ª, 17ª e 28ª iteração do processo de segmentação utilizando um modelo de forma activa previamente construído (b).

4.4 Modelo de Aparência Activa

A implementação usada neste trabalho que permite construir o modelo de aparência activa foi previamente desenvolvida por [Cootes, 2004], e consiste numa série de ficheiros executáveis, sendo o funcionamento global apresentado nesta secção.

A aplicação é constituída por três ficheiros principais: o ficheiro *am_build_apm*, que permite construir os modelos de forma, de textura e de aparência; o ficheiro *am_build_aam*, que constrói o modelo de aparência activa; e o ficheiro *am_markup*, que permite utilizar o modelo activo na segmentação em novas imagens. Para além destes ficheiros, existem ainda os executáveis que permitem visualizar os modelos construídos: *am_view_shape_model* e *am_view_apm*.

Os programas referidos são executados a partir da linha de comandos de uma *shell DOS*, sendo o seu funcionamento bastante simples e intuitivo, como se poderá verificar em seguida. Esta implementação possui ainda documentação referente à utilização das várias ferramentas incluídas, pelo que nesta secção se apresentam apenas o princípio de funcionamento das aplicações base.

De forma a construir o modelo de aparência activa, tal como acontece com o modelo de forma activa, é necessária a informação do conjunto de treino a considerar, as imagens e as coordenadas dos pontos-chave, e definir os parâmetros da percentagem de variação retida por cada modelo construído. Estes ficheiros, necessários para a construção dos modelos de aparência activa, são definidos usando uma aplicação em *MATLAB* desenvolvida de raiz durante este trabalho também a apresentar nesta secção.

Para a execução do procedimento de construção do modelo de aparência activa de um objecto, é necessário que as coordenadas dos pontos-chave das instâncias do objecto a modelar estejam armazenadas em ficheiros com extensão *pts*, segundo a formatação apresentada na Figura 4.15a). Na primeira secção deste capítulo, foram apresentados vários métodos que permitem etiquetar objectos em imagens e guardar as suas coordenadas numa matriz Xu . O programa *criar_pts* permite, tal como a designação indica, criar os ficheiros de extensão *pts* com as coordenadas dos pontos-chave existentes na matriz Xu obtida pelos referidos métodos, Figura 4.15b).

Para além dos ficheiros com as coordenadas dos pontos-chave, são necessários mais dois ficheiros de entrada: um com a informação do nome das imagens do conjunto de treino a considerar e alguns dados sobre a percentagem de variação a reter, com a

extensão *smd*; e outro, com os parâmetros a utilizar durante a fase de pesquisa: como o número de imagens a considerar do conjunto de treino ou os nomes a atribuir para os modelos de forma e aparência a construir, com a extensão *aamprops*. Estes dois ficheiros são extremamente simples de construir, pelo que o utilizador necessita apenas de alterar os nomes dos modelos nos ficheiros fornecidos com o programa principal e definir os parâmetros a considerar.

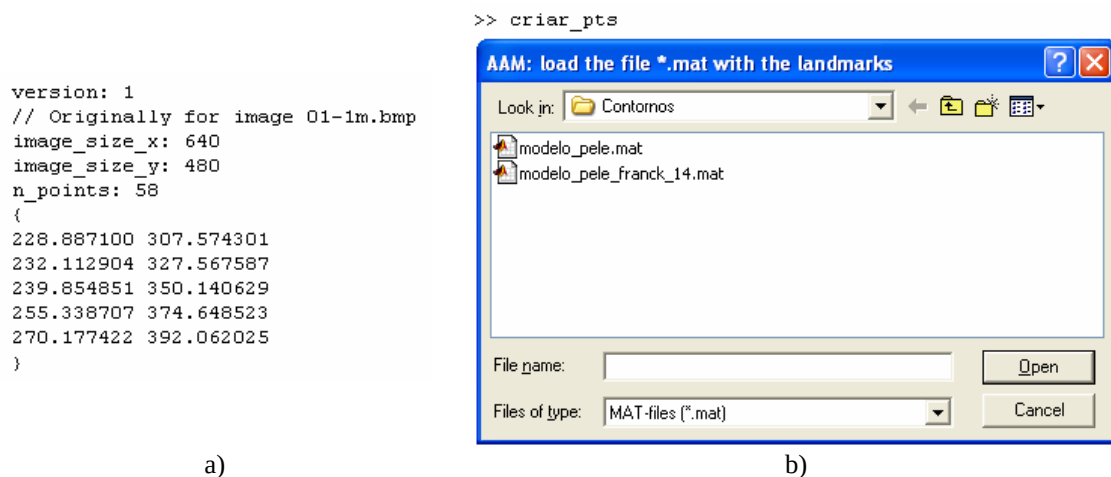
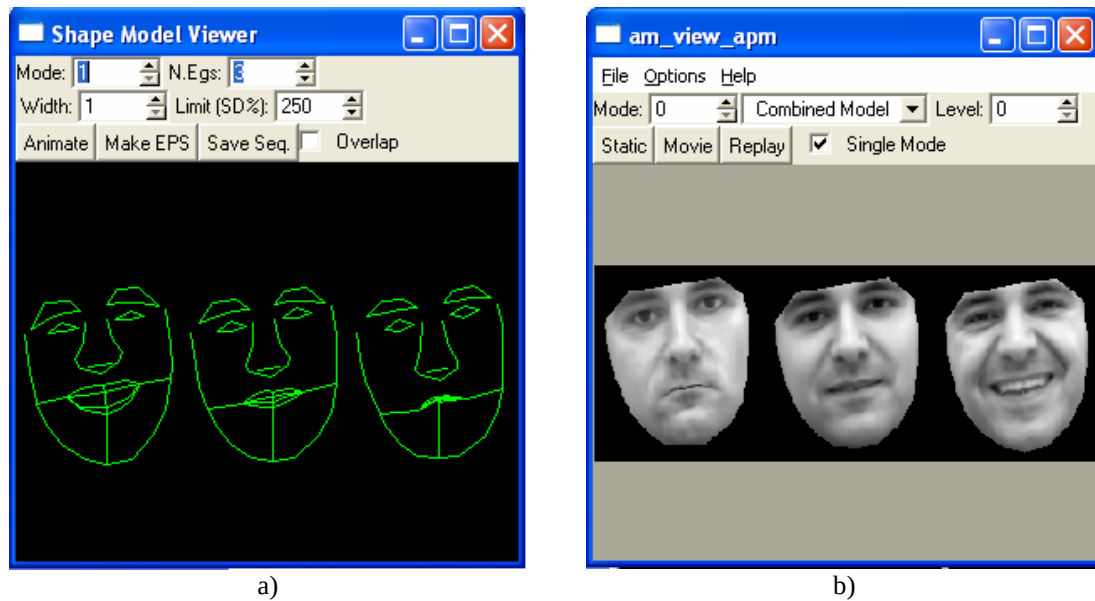


Figura 4.15 Exemplo de um ficheiro *.pts* com as coordenadas de um objecto do conjunto de treino (a) e implementação desenvolvida para criar ficheiros nesse formato (b).

O comando *am_build_apm nomedomodelo* permite construir os modelos de forma, textura e aparência, a partir da linha de comandos de uma *shell DOS*. Por outro lado, o comando *am_build_aam nomedomodelo* permite construir o modelo de aparência activa, a partir dos ficheiros anteriormente descritos. Para o utilizador visualizar os modos de forma, textura ou aparência, apenas necessita de usar o comando *am_view_shape_model nomedomodelo* ou *am_view_apm nomedomodelo*, Figura 4.16. Para pesquisar um objecto numa nova imagem, utilizando o modelo de aparência activa construído para o mesmo, basta executar o ficheiro *am_markup nomedomodelo*, Figura 4.17, indicar a imagem a segmentar na opção *Load Image* do menu *File*, e aceder ao menu de pesquisa, pela opção *Search* do menu *Props*, para definir os parâmetros de procura e dar início à pesquisa, Figura 4.18. Na aplicação dos modelos de aparência activa, o utilizador pode regular diversos parâmetros de pesquisa, assim como a apresentação gráfica da implementação. Além disso, também pode gravar as imagens resultantes ao longo de cada processo de pesquisa.



a) b)

Figura 4.16 Exemplo de visualização de um modo de variação da forma (a) e de um modo de variação da aparência (b).

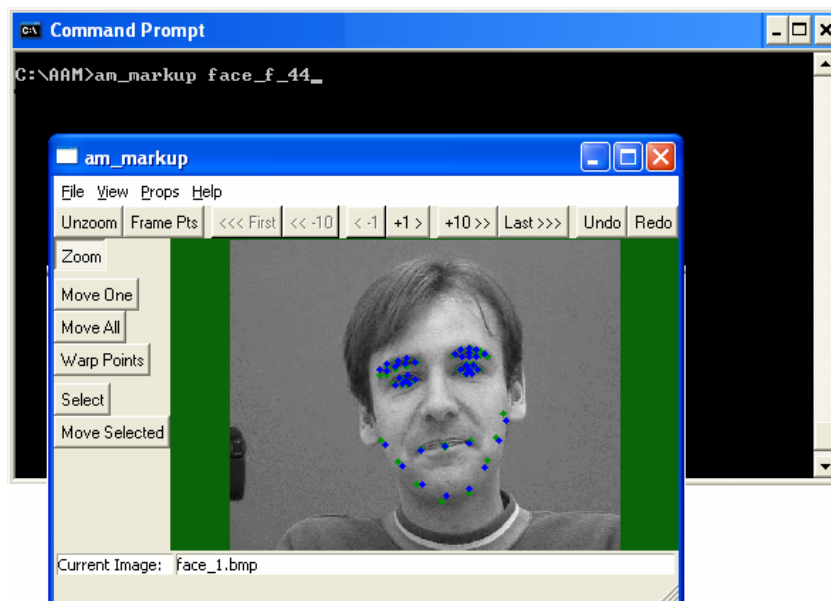


Figura 4.17 Fase de pesquisa dos modelos de aparência activa: indicação do modelo a utilizar.

4.5 Resumo

Neste capítulo, apresentaram-se as implementações desenvolvidas de raiz durante este trabalho, que permitem etiquetar automaticamente objectos do tipo mão e face, e ainda as implementações adaptadas de [Hamarneh, 1999a; Cootes, 2004] para a construção dos modelos de forma activa e de aparência activa.

Os métodos de etiquetagem automática desenvolvidos, mostraram ser eficazes na colocação de pontos-chave nos objectos do tipo considerado, e as implementações construídas para os mesmos são de utilização simples e bastante intuitiva.



Figura 4.18 Fase de pesquisa dos modelos de aparência activa: indicação da imagem a pesquisar.

Apresentam-se ainda as implementações desenvolvidas, tendo por base o trabalho de [Silva, 2002], para o emparelhamento automático dos pontos-chave, podendo estes métodos ser utilizados sempre que as correspondências não sejam dadas implicitamente pela ordem destes pontos. De referir que no caso das implementações desenvolvidas para etiquetar automaticamente objectos do tipo face e mão, esta correspondência é dada de forma implícita pela ordem dos pontos-chave obtidos.

No próximo capítulo, apresentam-se resultados experimentais obtidos pelos modelos estatísticos considerados nesta Dissertação, considerando diversos parâmetros de modelação e verifica-se o seu comportamento através da análise dos resultados de segmentação obtidos.

Capítulo 5

Resultados Experimentais

5.1 Introdução

As aplicações de modelos pontuais de distribuição abrangem variadíssimas áreas como, por exemplo: medicina, na localização de estruturas em imagens médicas; indústria, na inspecção de produtos; ou em sistemas de segurança, no reconhecimento de faces. No presente capítulo, apresentam-se resultados da modelação de dois tipos de objectos representados em imagens: a mão e a face.

Assim, na segunda secção deste capítulo, descrevem-se as imagens utilizadas para a construção dos modelos de forma activa e de aparência activa.

Posteriormente, apresenta-se a modelação da mão utilizando os algoritmos de etiquetagem automática desenvolvidos, descritos nos capítulos anteriores, para a construção dos modelos de forma e aparência activa, variando e analisando o efeito dos diversos parâmetros utilizados. Assim, ao longo desta secção, os modelos considerados são descritos detalhadamente e apresentam-se os resultados obtidos na segmentação da mão em novas imagens, discutindo-se os valores adequados dos parâmetros a utilizar. Faz-se ainda a comparação entre os modelos de forma activa e os de aparência activa, através do estudo dos resultados da segmentação obtidos considerando cinco novas imagens e as mesmas condições iniciais.

Na quarta secção, descrevem-se os modelos de forma e de aparência activa construídos para a modelação da face, utilizando os vários métodos de colocação automática dos pontos-chave apresentados nos capítulos anteriores. Estuda-se também a influência de diversos parâmetros e o comportamento dos modelos construídos, através da análise dos resultados de segmentação obtidos considerando quatro novas imagens.

5.2 Imagens Utilizadas

Neste capítulo foram consideradas 30 imagens digitais de mãos [Stegmann and Gomez, 2002] de quatro indivíduos diferentes, sendo 25 destas imagens utilizadas na fase de

treino e as restantes 5 na fase de teste. As referidas imagens têm dimensão de 800×600 *píxeis*. Já para o caso da face, foram consideradas 26 imagens de faces de um indivíduo [Cootes, 2004], sendo 22 imagens utilizadas na fase de treino e as restantes 4 para testar os modelos construídos. As referidas imagens têm dimensão de 720×576 *píxeis*.

Todas as imagens utilizadas neste trabalho estavam inicialmente no formato *jpeg*. A extracção automática dos pontos-chave foi efectuada a partir das imagens originais a cores, sendo posteriormente convertidas no formato *bmp* e em escala de cinzentos para a construção dos modelos de forma e aparência.

5.3 Modelação da Mão

Tal como se referiu na secção anterior, para a modelação da mão foram utilizadas 25 imagens de treino.

Nesta secção apresenta-se os resultados obtidos com etiquetagem automática de 79 pontos-chave, Figura 5.1, segundo o algoritmo apresentado nos capítulos anteriores, considerando 3 pontos-chave nas zonas de maior curvatura e 5 entre as mesmas zonas. No final desta secção, serão apresentadas algumas observações sobre os resultados obtidos, utilizando um número distinto de pontos-chave para inferir sobre a sua influência no modelo pontual de distribuição obtido.

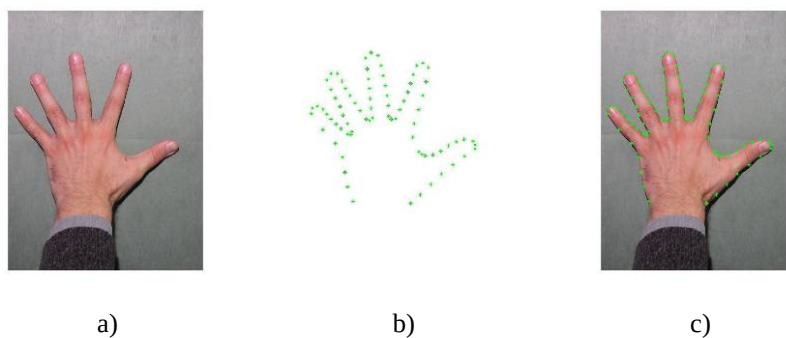


Figura 5.1 Exemplo de uma imagem do conjunto de treino considerado (a) etiquetada automaticamente usando 79 pontos-chave (b e c).

Na Tabela 5.1 são apresentados os valores obtidos para os primeiros 5 valores próprios do modelo pontual de distribuição obtido para a mão em estudo e as percentagens de variação acumuladas respectivas, considerando o alinhamento segundo os *métodos A* e *B* (ver capítulo 2).

A partir dos resultados obtidos, é possível verificar que o modelo obtido para a mão considerando o *método A*, necessita dos primeiros 4 modos de variação para reter 90% da variabilidade total da forma, que os primeiros 6 modos traduzem 96% da variação total da forma, e que os primeiros 11 modos são capazes de traduzir 99% da variabilidade total. Por outro lado, se considerarmos o *método B*, verifica-se que os primeiros 5 modos de variação retêm 92% da variabilidade total da forma, que os primeiros 6 modos traduzem 95% da variação total da forma, e que para traduzir 99% da variabilidade da forma do objecto são necessários os primeiros 11 modos de variação.

Tabela 5.1 Primeiros 5 valores próprios do modelo obtido para a mão e percentagens de variação retidas por cada um.

Método	A		B	
	Valor próprio	Percentagem retida	Valor próprio	Percentagem retida
λ_1	26863	70.73%	21598	65.04%
λ_2	3123.6	78.95%	3379.4	75.22%
λ_3	2866.4	86.50%	2982.3	84.20%
λ_4	1521.7	90.50%	1512.5	88.76%
λ_5	1322.3	93.99%	1382.3	92.92%

Na Figura 5.2 apresenta-se exemplos do efeito da variação dos primeiros 3 modos do modelo da forma construído em torno da forma média obtida, utilizando o *método A*. Na referida figura, pode-se constatar que o primeiro modo traduz essencialmente as variações globais do objecto e que os segundo e terceiro modos se prendem mais com as variações locais, como o deslocamento individual de cada um dos dedos.

5.3.1 Modelo de Forma Activa

Os modelos de forma activa permitem localizar os objectos previamente modelados em novas imagens, como se explica no segundo capítulo desta Dissertação. Em seguida, descrevem-se os vários tipos de modelos gerados e apresentam-se os resultados obtidos na localização dos objectos modelados em novas imagens (operação de segmentação). Com o propósito de analisar as diferenças dos modelos no caso em que a percentagem de variação retida varia, construíram-se dois tipos de modelos: o primeiro que explica

95% da variância total das instâncias da forma, e o segundo que explica 99% da mesma variância.

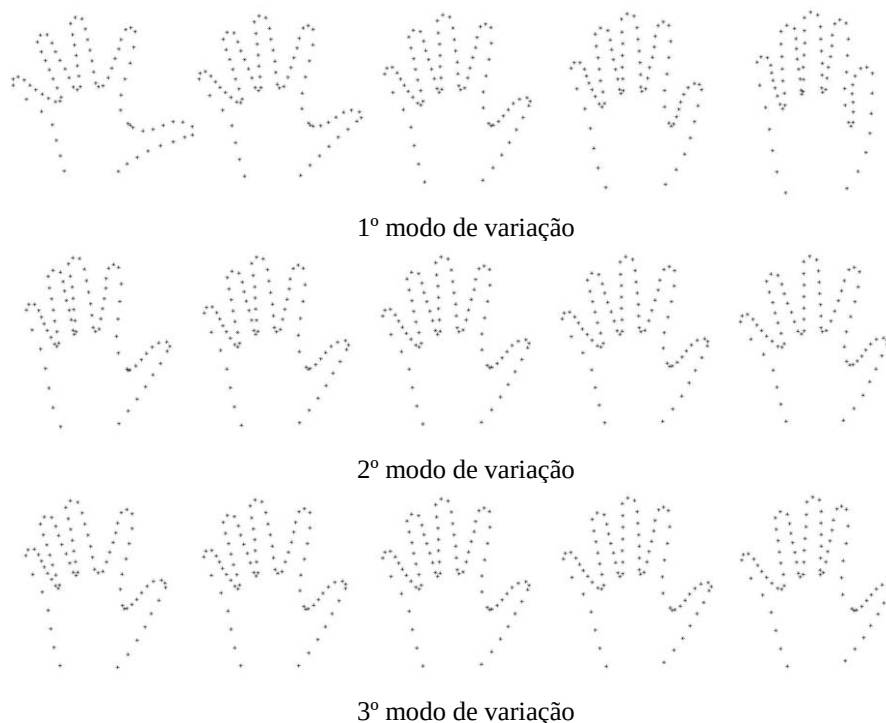


Figura 5.2 Efeito da variação ($\pm 2sd$) dos primeiros 3 modos do modelo obtido, usando o método A, para a forma da mão.

Para construir os modelos de forma activa utilizaram-se também perfis de dimensões distintas, com a finalidade de analisar a sensibilidade dos métodos relativamente à sua variação. Assim, estudaram-se os modelos activos com perfis de dimensão reduzida (7 *píxeis* 3 para cada lado do ponto-chave), média (15 *píxeis*) e elevada (33 *píxeis*).

Deste modo, geraram-se 12 modelos de forma activa cujas características são descritas resumidamente na Tabela 5.2. Para maior facilidade na interpretação de resultados, optou-se por incluir na nomenclatura dos modelos gerados indicações sobre as suas características próprias. Deste modo, a designação de cada modelo indica o objecto modelado, o método de alinhamento utilizado, seguido do valor da percentagem de variação retida e da dimensão do perfil de pesquisa considerado, como se pode observar na tabela referida. Por exemplo, o modelo *Mão_A_95_p3* foi construído para modelar a mão, utilizando o *método A* de alinhamento, retendo 95% da variação total e com dimensão reduzida (3 *píxeis* para cada lado do ponto-chave).

Como já se referiu, as imagens da mão têm dimensões de 800×600 *píxeis*, sendo assim possível construir os modelos considerando no máximo 6 níveis de resolução; no

entanto, verificou-se que para os modelos de forma activa, a utilização de mais do que 3 níveis de resolução não conduzia à convergência do método; dado que com os 4º, 5º e 6º níveis de resolução, os modelos obtidos têm tendência a deslocar-se erradamente pela imagem, não sendo capazes de se fixarem à região esperada. Assim, as pesquisas efectuadas iniciam-se no 3º nível de resolução, no qual os modelos activos localizam a região da mão e nos 2º e 1º níveis ajustam o modelo ao objecto.

Tabela 5.2 Caracterização dos modelos de forma activa gerados para a mão.

Modelo	Método de alinhamento	Percentagem de variação retida	Dimensão do perfil de pesquisa
<i>Mão A 95 p3</i>	A	95%	3
<i>Mão A 95 p7</i>	A	95%	7
<i>Mão A 95 p16</i>	A	95%	16
<i>Mão A 99 p3</i>	A	99%	3
<i>Mão A 99 p7</i>	A	99%	7
<i>Mão A 99 p16</i>	A	99%	16
<i>Mão B 95 p3</i>	B	95%	3
<i>Mão B 95 p7</i>	B	95%	7
<i>Mão B 95 p16</i>	B	95%	16
<i>Mão B 99 p3</i>	B	99%	3
<i>Mão B 99 p7</i>	B	99%	7
<i>Mão B 99 p16</i>	B	99%	16

Neste trabalho, optou-se por caracterizar a adequação da segmentação do objecto modelado em novas imagens, obtida pelos modelos de forma activa construídos, usando o valor mínimo, máximo, médio e desvio padrão das distâncias euclidianas entre os pontos-chave do modelo e da nova instância do objecto.

Para testar os modelos de forma activa gerados para a mão, foram utilizadas 5 imagens de teste, não pertencentes ao conjunto de treino, apresentadas na Figura 5.3.

A Figura 5.4 apresenta o exemplo de um resultado de segmentação, usando um modelo de forma activa construído numa imagem de teste. No referido modelo, os perfis de intensidade para cada ponto-chave são de dimensão 7 *píxeis*, a percentagem de variação retida foi de 95%, o número máximo de iterações por nível de resolução considerado foi de 10 e a segmentação apresentada iniciou-se no nível de resolução 3. Os resultados de segmentação obtidos para as restantes quatro imagens de teste, utilizando o mesmo modelo, são apresentados na Figura 5.5.

Na Tabela 5.3 apresenta-se os valores obtidos da média e do desvio padrão que traduzem a adequação da segmentação de cada modelo construído em cada imagem de teste. Na referida tabela, as células que não apresentam valores referem-se aos casos em

que o modelo considerado não obteve resultados adequados; isto é, os casos em que não foi possível localizar correctamente o objecto na nova imagem.

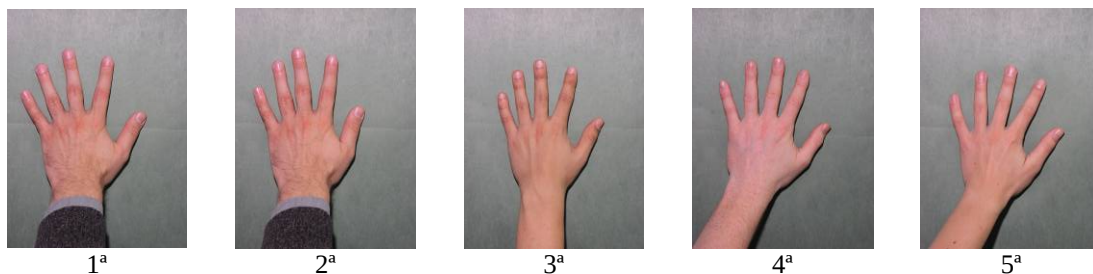


Figura 5.3 Imagens de teste utilizadas para verificar o comportamento dos modelos construídos para a mão numa operação de segmentação.

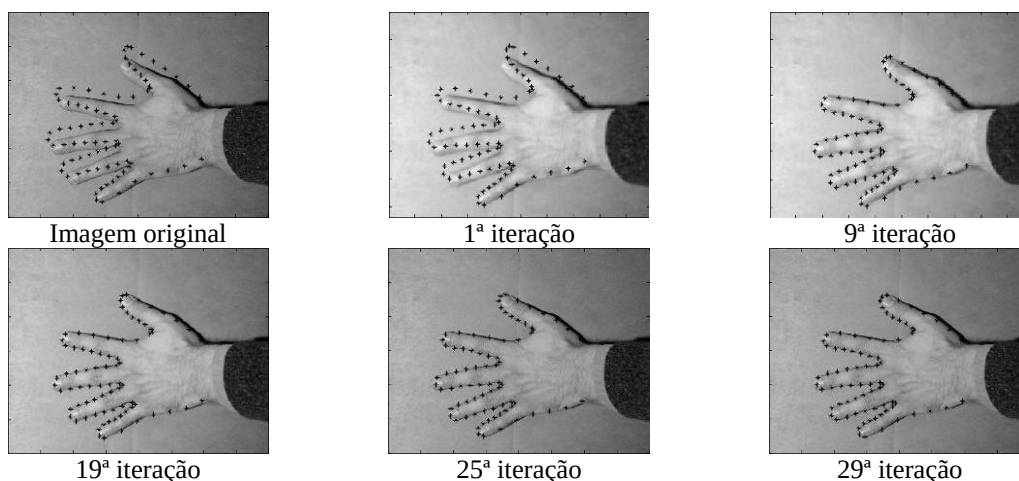


Figura 5.4 Imagem de teste com a posição do modelo médio obtido para a mão sobreposto, e após a 1ª, 9ª, 19ª, 25ª e 29ª iteração do processo de segmentação usando o modelo construído.

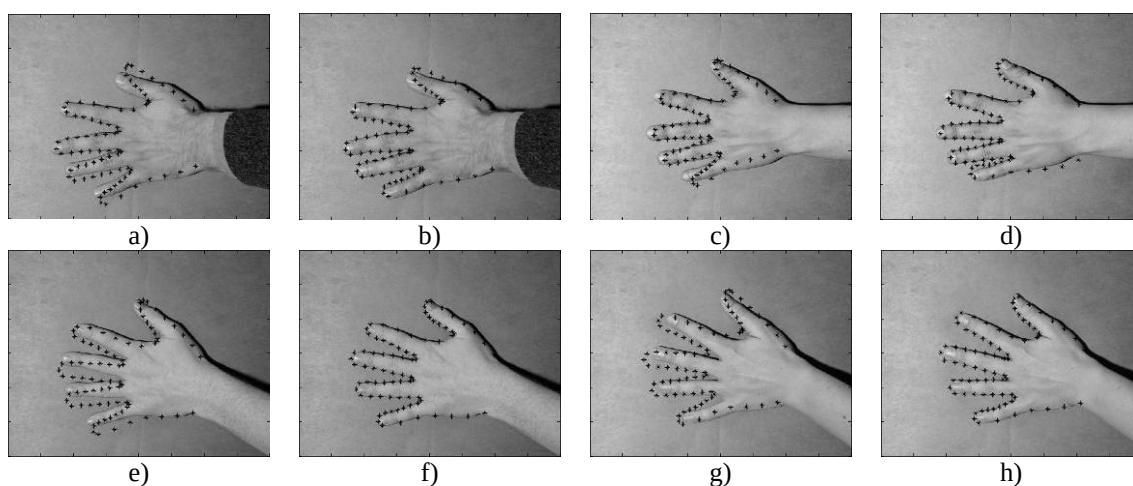


Figura 5.5 Imagens de teste com o modelo médio obtido para a mão sobreposto (a, c, e, g) e resultados finais do processo de segmentação usando o modelo construído (b, d, f, h).

Tabela 5.3 Erros obtidos dos modelos de forma construído para a mão nas novas imagens de teste (média e desvio padrão).

Modelos	Imagem 1	Imagem 2	Imagem 3	Imagem 4	Imagem 5
<i>Mão A 95 p3</i>	5.38 ± 3.18		10.91 ± 4.32	3.64 ± 3.19	5.88 ± 3.77
<i>Mão A 95 p7</i>	6.13 ± 3.67	4.41 ± 3.15		4.25 ± 3.17	6.05 ± 3.25
<i>Mão A 95 p16</i>	5.09 ± 2.96	10.01 ± 5.50			
<i>Mão A 99 p3</i>	7.52 ± 8.06		10.47 ± 3.76	4.10 ± 3.41	4.88 ± 3.23
<i>Mão A 99 p7</i>				3.39 ± 2.69	11.73 ± 5.56
<i>Mão A 99 p16</i>					
<i>Mão B 95 p3</i>	5.42 ± 3.31		12.23 ± 5.02	3.81 ± 3.46	5.14 ± 2.96
<i>Mão B 95 p7</i>	6.79 ± 4.60	4.68 ± 2.63		4.27 ± 3.45	5.49 ± 3.11
<i>Mão B 95 p16</i>		5.28 ± 2.73			
<i>Mão B 99 p3</i>	6.68 ± 7.08	8.64 ± 6.72	8.09 ± 6.08	4.03 ± 3.14	5.35 ± 4.01
<i>Mão B 99 p7</i>		7.09 ± 6.84	8.03 ± 7.46	4.26 ± 3.46	8.32 ± 4.89
<i>Mão B 99 p16</i>	9.26 ± 10.20	8.05 ± 4.74			

Numa primeira análise aos valores da referida tabela, é possível verificar que apenas o modelo *Mão B 99 p3* encontra correctamente o objecto modelado nas 5 imagens de teste consideradas, que 6 dos modelos construídos falham na localização numa imagem de teste, e ainda que os modelos com perfis de pesquisa de dimensão elevada (33 píxeis) geralmente obtêm maus resultados de segmentação do objecto modelado em novas imagens.

Ao utilizar o *método A* de alinhamento, ou seja, quando o alinhamento é obtido tendo em conta as variâncias dos pontos-chave, os modelos obtidos que melhor se ajustam às várias imagens de testes são os que retêm 95% da variância total e utilizam perfis de reduzida ou média dimensão. Por outro lado, ao utilizar o *método B* de alinhamento, ou seja, quando o alinhamento é realizado considerando todos os pontos-chave com o mesmo peso, o mesmo tipo de modelos obtêm melhores resultados na localização do objecto. É de notar que o *modelo B* que retém 99% da variação e utiliza perfis de pequena dimensão, consegue localizar os objectos nas 5 imagens de teste mas obtêm precisões inferiores aos modelos anteriores, daí não ser possível considerá-lo em termos globais como o melhor modelo para a segmentação em estudo.

Também é possível constatar que a utilização quer do *método A* quer do *método B* de alinhamento conduz a bons resultados de segmentação, sendo apenas de referir que em geral o *método A* obtém, nos casos estudados, resultados mais precisos no reconhecimento de objectos modelados em novas imagens.

Observa-se também que nos modelos estudados, considerar os modos que retêm 95% da variação total do objecto conduz a melhores resultados de segmentação do que considerar os modos que retêm 99% da variação, ao contrário do que seria de esperar.

Pela observação do processo de pesquisa dos modelos activos usados, tal situação é justificada pelo facto de ao se utilizar 11 modos de variação (99% da percentagem retida) o modelo activo tem tendência a deformar-se em demasia nas primeiras iterações da pesquisa, conduzindo a piores resultados quando comparando com o comportamento do modelo que considera apenas os 5 primeiros modos de variação (95% da percentagem retida).

A dimensão dos perfis de intensidade usados dos modelos construídos influencia a pesquisa dos objectos modelados em novas imagens. Como já se referiu, verificou-se que a utilização de perfis de dimensão elevada geralmente não conduzia a bons resultados de segmentação dos objectos modelados em novas imagens. Tal resultado deve-se ao facto de, pela dimensão ser elevada, serem considerados muitos pontos que não fazem parte do objecto e portanto não apresentam informação útil para a pesquisa, e que conduzem assim a piores resultados. Por outro lado, a utilização de perfis de reduzida e média dimensão são mais adequados na modelação de objectos como a mão, notando-se apenas uma ligeira subida dos erros ao utilizar os perfis de média dimensão (15 *píxeis*) relativamente aos de dimensão reduzida (7 *píxeis*); tal facto, deve-se à dimensão relativa dos objectos considerados.

5.3.2 Modelo de Aparência Activa

Tal como acontece com os modelos de forma activa, os modelos de aparência activa são capazes de segmentar e reconhecer os objectos modelados em novas imagens.

Neste trabalho, geraram-se modelos de aparência activa a partir de 25 imagens de mãos etiquetadas automaticamente com 79 pontos-chave, tal como foi considerado na secção anterior. Foram construídos dois tipos de modelos, variando a percentagem de variação retida: o primeiro, explica 95% da variância total das instâncias da forma, e o segundo explica 99% da mesma variância. Construíram-se também modelos considerando diferentes números de *píxeis* no modelo de textura (5000, 10000 e 50000 *píxeis*).

Considerando 95% da variação total da mão e 10000 *píxeis* na construção do modelo de textura, extraíram-se 6 modos de variação da forma, 17 modos de textura e 11 modos de aparência. Por outro lado, o modelo de aparência activa construído para reter 99% da variabilidade total da mão e utilizando o mesmo número de *píxeis*, necessita de 11 modos de variação da forma, 23 modos de variação de textura e 19 modos de aparência.

Na Tabela 5.4 apresenta-se as características de cada modelo construído e ainda o número de modos retidos por cada um.

Tabela 5.4 Caracterização dos modelos de aparência activa considerados.

Modelos	Percentagem de variação retida	Nº de píxeis do modelo de textura	Nº de modos de variação da Forma	Nº de modos de variação da Textura	Nº de modos de variação da Aparência
<i>Mão 95 5</i>	95%	5000	6	15	10
<i>Mão 95 10</i>	95%	10000	6	17	11
<i>Mão 95 50</i>	95%	50000	6	19	13
<i>Mão 99 5</i>	99%	5000	11	22	18
<i>Mão 99 10</i>	99%	10000	11	23	19
<i>Mão 99 50</i>	99%	50000	11	24	21

Na Figura 5.6 apresenta-se exemplos do efeito da variação dos primeiros três modos do modelo de textura, ou seja de intensidade dos *píxeis* do objecto, e da aparência construído em torno da forma média. Os modos de variação da forma são os mesmos apresentados no início desta secção. Analisando as referidas figuras, pode-se constatar que os modos de textura consideram a informação dos diferentes indivíduos do conjunto de treino, e que os modos de aparência combinam a forma da mão com a textura.

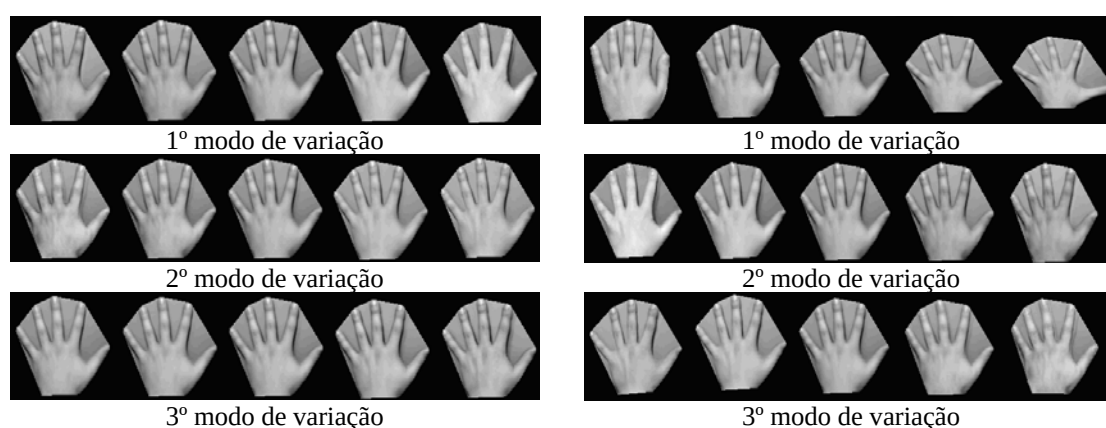


Figura 5.6 Efeito da variação ($\pm 2sd$) dos três primeiros modos do modelo de textura (à esquerda) e de aparência construído para a mão (à direita).

No sentido de estudar a eficácia dos modelos de aparência activa construídos no reconhecimento de objectos, foram utilizadas 5 imagens de teste, Figura 5.3. Na fase de pesquisa, o número máximo de iterações por nível foi limitado a 10 e a segmentação iniciou-se no nível de resolução 5. A Figura 5.7 apresenta algumas iterações do processo computacional do primeiro modelo de aparência activa construído para a

primeira imagem de teste, e a Figura 5.8 apresenta os resultados obtidos para as restantes quatro imagens de teste usando o mesmo modelo.

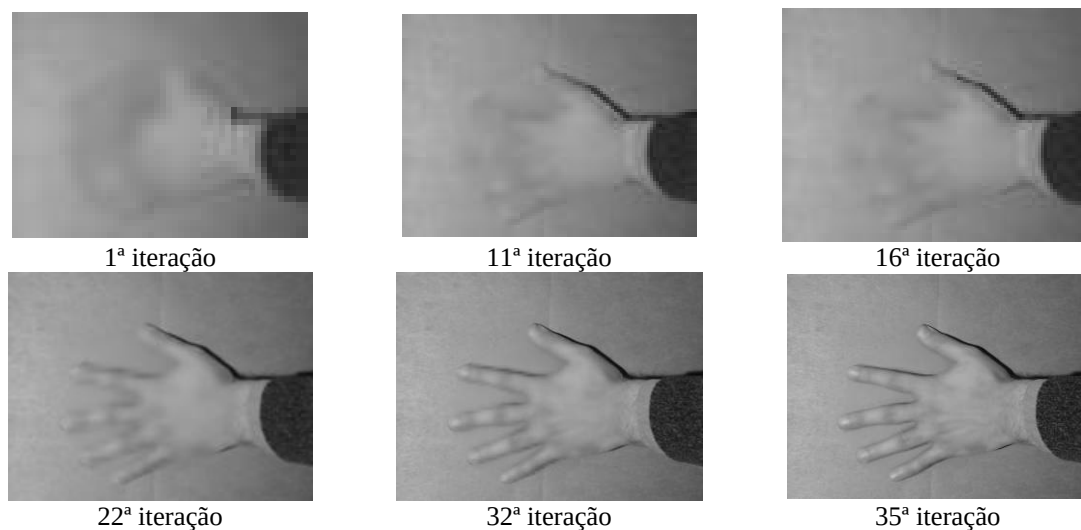


Figura 5.7 Alguns resultados de iterações do processo de segmentação de uma imagem de teste, usando o modelo de aparência construído para a mão.

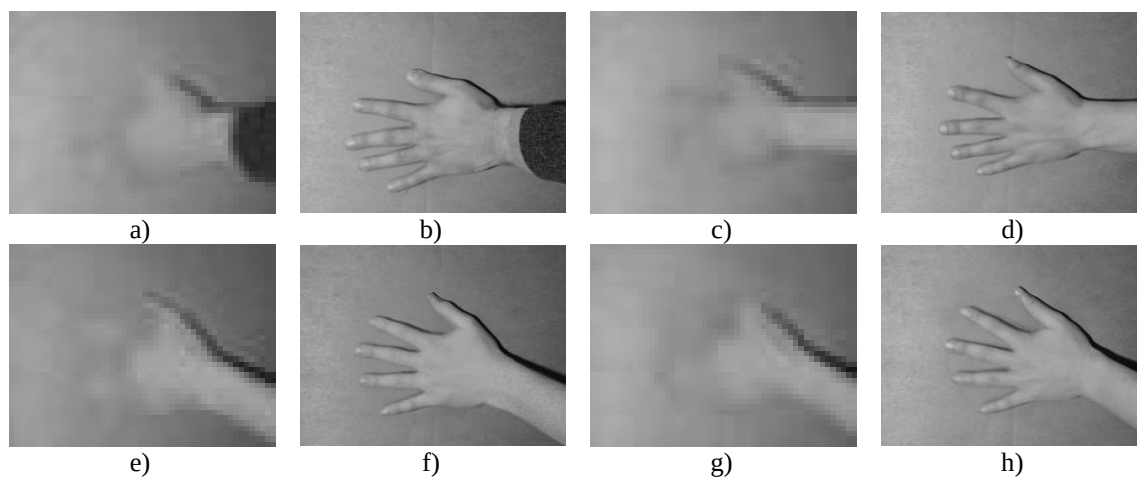


Figura 5.8 Imagens de teste após a primeira iteração do modelo de aparência activa (a, c, e, g) e resultados finais do processo de segmentação (b, d, f, h).

Na Tabela 5.5 apresenta-se os erros resultantes do processo de segmentação do objecto modelado nas várias imagens de teste consideradas, utilizando os modelos descritos anteriormente. Os valores obtidos da média e do desvio padrão dos erros em cada imagem de teste, permitem retirar algumas conclusões sobre os melhores parâmetros a utilizar na modelação da mão considerada. Mais uma vez, nesta tabela, as células que não apresentam valores referem-se aos casos nos quais o modelo não obteve bons resultados; ou seja, onde não foi possível localizar correctamente o objecto na nova imagem, ou a forma do modelo obtido não era adequada para o objecto considerado.

Tabela 5.5 Erros resultantes da segmentação em imagens de teste, utilizando os modelos de aparência activa construídos.

Modelos	Imagem 1	Imagem 2	Imagem 3	Imagem 4	Imagem 5
<i>Mão 95 5</i>	3.00 ± 1.76	4.37 ± 3.54	4.15 ± 3.17		6.64 ± 5.54
<i>Mão 95 10</i>	3.28 ± 1.88	4.21 ± 3.70	3.87 ± 2.91	3.35 ± 2.69	7.63 ± 6.85
<i>Mão 95 50</i>	3.12 ± 1.78	2.91 ± 2.24	3.85 ± 2.73	2.93 ± 1.83	7.23 ± 6.12
<i>Mão 99 5</i>	2.43 ± 1.43	2.41 ± 1.74	2.04 ± 1.23	3.08 ± 3.73	6.13 ± 5.60
<i>Mão 99 10</i>	2.33 ± 1.27	2.18 ± 1.29	2.37 ± 1.75	2.12 ± 2.07	5.77 ± 5.37
<i>Mão 99 50</i>	2.26 ± 1.33	2.15 ± 1.16	2.38 ± 1.82	2.11 ± 1.89	5.11 ± 5.01

Pela análise dos valores da tabela anterior, é possível constatar que apenas o primeiro modelo estudado, *Mão_95_5*, falha na localização do objecto na imagem de teste 4 e que todos os restantes conseguem segmentar, com mais ou menos precisão, os objectos em causa em novas imagens.

Também se pode verificar que os modelos que retêm 99% da variância total da mão obtêm melhores resultados do que os que retêm 95% da variabilidade total do objecto em estudo. Como seria de esperar, quanto maior a percentagem de variação retida pelo modelo, menor o erro obtido na pesquisa em novas imagens. Assim, para os modelos de aparência activa já se obteve o resultado esperado em relação à influência da variação da percentagem retida. No entanto, é de notar que neste tipo de modelos a quantidade de informação utilizada nos modelos obtidos é consideravelmente superior à usada nos modelos de forma activa, daí a diferença dos resultados obtidos entre os dois tipos de modelos; tal permite concluir que os modelos de aparência são mais estáveis e robustos. O número de *píxeis* utilizados na construção dos modelos também influencia, obviamente, a segmentação obtida; no geral, o erro diminui quando este número aumenta, sendo essa relação notória nos modelos que retêm 99% da variância (ver Tabela 5.5). Tal facto também seria de esperar, pois quanto maior o número de *píxeis* usados maior a informação da textura do objecto o que conduz a resultados mais eficazes na pesquisa em novas imagens, esta relação tem que considerar obviamente as dimensões dos objectos em causa.

Neste trabalho, também foram gerados modelos para a mão com número diferente de pontos-chave, e verificou-se que ao utilizar 127 pontos-chave, 5 pontos nas zonas de maior curvatura e 8 pontos entre estas zonas, os modelos de forma activa obtinham melhores resultados na segmentação dos objectos modelados em novas imagens, considerando as mesmas condições iniciais. Assim, constatou-se que erro médio obtido pelos modelos de forma activa utilizando 127 pontos-chave, era de menos 2 *píxeis* quando comparado com os modelos construídos com 79 pontos-chave, e verificou-se o

mesmo tipo de comportamento ao alterar os outros parâmetros: dimensão do perfil, percentagem de variação retida, método de alinhamento. Por outro lado, nos modelos de aparência activa verificou-se que o aumento do número de pontos-chave não se traduzia numa redução dos erros de segmentação obtidos em novas imagens e que os valores obtidos dos erros se mantiveram sensivelmente iguais nas duas situações, confirmando assim mais uma vez que estes modelos são mais estáveis e robustos.

A Figura 5.9 apresenta o exemplo de um resultado de segmentação de um modelo de forma activa numa imagem de teste, utilizando 127 pontos-chave. No modelo utilizado, os perfis de intensidade para cada ponto-chave são de dimensão 7 *píxeis*, a percentagem de variação retida no modelo foi de 95%, o número máximo de iterações por nível de resolução considerado foi de 10, e a segmentação apresentada iniciou-se no nível de resolução 3.

Por seu lado, a Figura 5.10 apresenta algumas iterações do processo computacional de um modelo de aparência activa construído para a primeira imagem de teste. No modelo representado, utilizaram-se 127 pontos-chave, a percentagem de variação retida foi de 95% e foram usados 5000 *píxeis* para construir o modelo de textura associado.

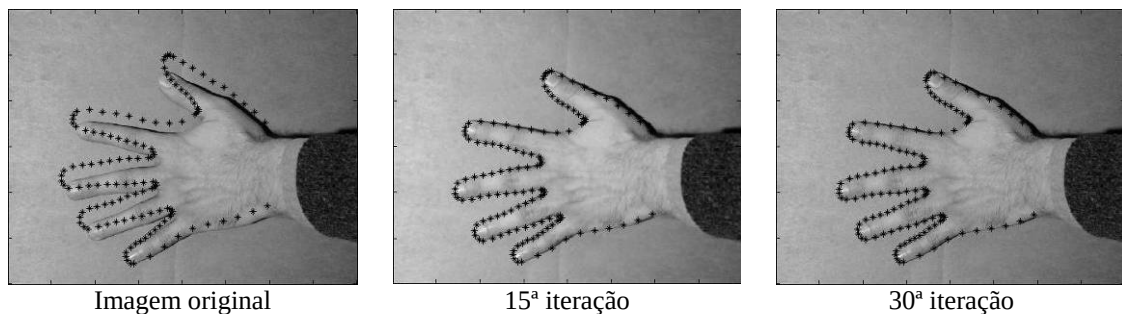


Figura 5.9 Imagem de teste com a posição do modelo médio obtido para a mão sobreposto, e após a 15ª e 30ª iteração do processo de segmentação usando o modelo de forma activa construído para a mão.

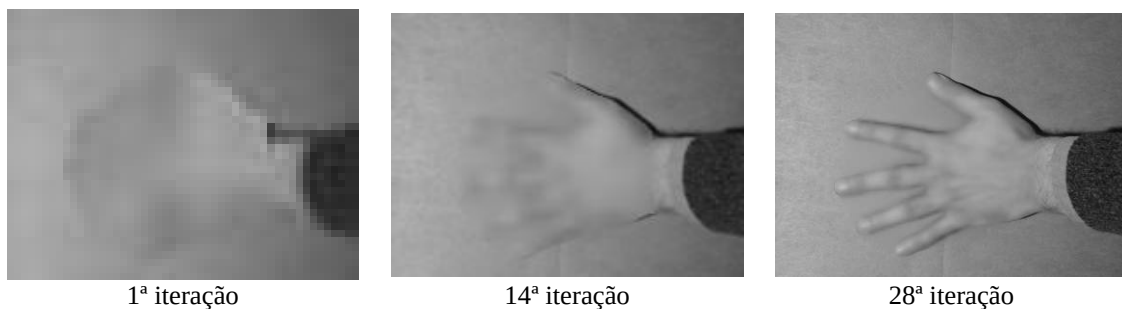


Figura 5.10 Alguns resultados de iterações do processo de segmentação de uma imagem de teste, usando o modelo de aparência construído para a mão.

5.4 Modelação da Face

Como se referiu no início deste capítulo, foram utilizadas 22 imagens de treino para a modelação da face.

Nesta secção, apresentam-se os resultados obtidos pela etiquetagem automática da face segundo os três métodos apresentados nos capítulos anteriores. Para a construção dos métodos de forma activa, usou-se o primeiro método de etiquetagem automática apresentado e para os métodos de aparência activa usaram-se os restantes três métodos de etiquetagem.

Na Figura 5.11 são apresentados resultados utilizando o primeiro método de etiquetagem, onde se consideram 9 pontos-chave em cada sobrancelha, 7 pontos-chave nas regiões dos olhos, 4 pontos-chave na boca e 8 pontos-chave no contorno da face, perfazendo um total de 44 pontos-chave. Já para os modelos de aparência activa são apresentados resultados utilizando uma malha regular de dimensões 7×7 (*algoritmo II*) e uma malha adaptativa de multi-resolução com 54 pontos-chave (*algoritmo III, tipo 1*) e 75 pontos-chave (*algoritmo III, tipo 2*), como se apresenta na Figura 5.12.

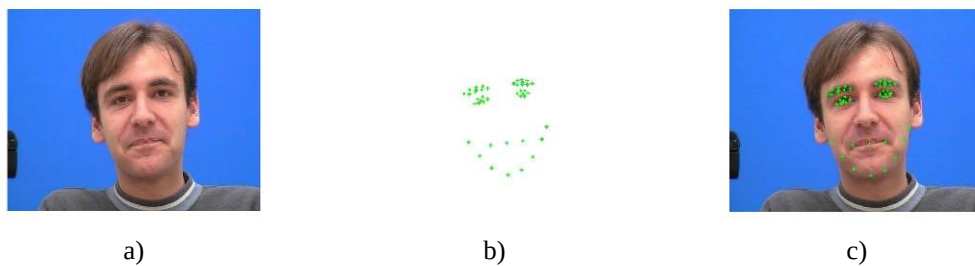


Figura 5.11 Exemplo de uma imagem do conjunto de treino etiquetada (a) automaticamente usando 44 pontos-chave usando o *algoritmo I* (b e c).

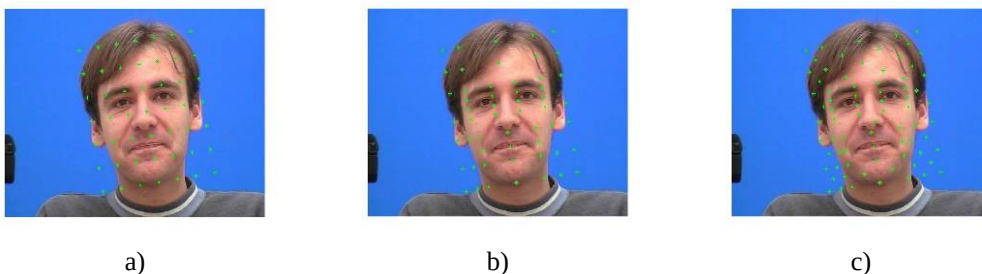


Figura 5.12 Exemplo de uma imagem do conjunto de treino etiquetada automaticamente usando: 49 pontos-chave pelo *algoritmo II* (a); 54 pontos-chave pelo *algoritmo III-tipo 1* (b); e 75 pontos-chave pelo *algoritmo III-tipo 2* (c).

No final desta secção serão também apresentadas algumas observações sobre os resultados obtidos utilizando um número diferente de pontos-chave, nos vários modelos construídos, para inferir sobre sua influência nos modelos pontuais de distribuição obtidos.

Na Tabela 5.6 são apresentados os valores obtidos para os primeiros 10 modos próprios do modelo pontual de distribuição construído para a face em estudo, utilizando o primeiro algoritmo de etiquetagem automática, e as percentagens de variação acumuladas respectivas, considerando o alinhamento segundo os *métodos A e B*. Através da análise da referida tabela, é possível verificar que o modelo obtido para a face, considerando o *método A* de alinhamento, necessita dos primeiros 10 modos de variação para reter 90% da variabilidade total da forma, e que os primeiros 13 modos traduzem 95% da variação total da forma, enquanto que os primeiros 18 modos são capazes de traduzir 99% da variabilidade total. Se considerarmos o *método B* de alinhamento, verifica-se o mesmo comportamento.

Tabela 5.6 Primeiros 10 valores próprios do modelo da face obtido e as percentagens de variação retidas.

Método	A		B	
	Valor próprio	Percentagem retida	Valor próprio	Percentagem retida
λ_1	325.62	23.98	320.42	23.98
λ_2	229.48	40.88	227.56	41.01
λ_3	180.81	54.20	176.73	54.24
λ_4	135.91	64.21	133.13	64.21
λ_5	100.38	71.61	98.00	71.54
λ_6	76.77	77.26	76.03	77.23
λ_7	60.73	81.73	59.17	81.66
λ_8	51.62	85.53	50.66	85.45
λ_9	39.74	88.46	39.56	88.41
λ_{10}	28.94	90.59	28.71	90.56

Na Figura 5.13 apresenta-se exemplos do efeito da variação dos primeiros 3 modos do modelo da forma construído em torno da forma média, utilizando o *método A* de alinhamento. Na mesma figura, pode-se constatar que o primeiro modo traduz essencialmente o movimento de rotação da face, enquanto que o segundo e terceiro modos se prendem mais com as variações locais, como o formato dos olhos e sobrancelhas no segundo modo, e alterações no contorno da face e da boca no terceiro modo.

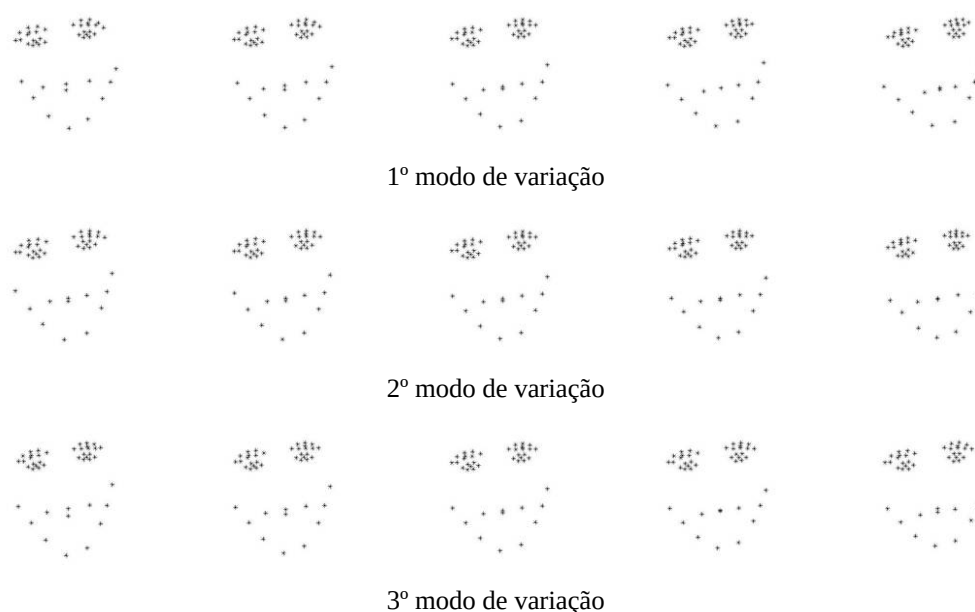


Figura 5.13 Efeito da variação ($\pm 2sd$) dos primeiros 3 modos do modelo obtido, usando o *método A* de alinhamento, para a forma da face.

5.4.1 Modelo de Forma Activa

Em seguida, descrevem-se os vários modelos de forma activa gerados neste trabalho para a face e apresentam-se os resultados obtidos na localização em novas imagens.

Os modelos construídos para a modelação da face são idênticos aos gerados para o caso da mão. Tal como acontece nos modelos da mão, o nome do modelo gerado indica quais os parâmetros considerados na sua construção: *Face_método_%retida_nºpíxeis* usados; ou seja, por exemplo, o modelo *Face_A_95_p3* corresponde ao modelo construído utilizando o *método A* de alinhamento, a percentagem de variação retida no modelo é de 95% e a dimensão do perfil de intensidade é de 7 píxeis. Assim, geraram-se no total 12 modelos de forma activa para a modelação da face, utilizando o primeiro algoritmo de etiquetagem automática, apresentado nos capítulos anteriores, com 44 pontos-chave.

Como já se referiu, as imagens da face têm dimensões de 720×576 píxeis, sendo possível construir modelos considerando no máximo 6 níveis de resolução; no entanto, constatou-se que, para os modelos de forma activa, a utilização de mais do que 3 níveis de resolução não conduzia à convergência do método, tal como aconteceu no caso da mão apresentado anteriormente.

No presente estudo, a adequação da segmentação do objecto em novas imagens, obtida através dos modelos de forma activa construídos, é caracterizada usando o valor

mínimo, máximo, médio e desvio padrão das distâncias euclidianas entre os pontos-chave do modelo construído e do objecto a segmentar na nova imagem.

Para testar os modelos de forma activa construídos para a face, foram usadas 4 imagens de teste, não pertencentes ao conjunto de treino considerado, Figura 5.14.



Figura 5.14 Imagens de teste utilizadas para verificar o comportamento dos modelos construídos para a face.

A Figura 5.15 apresenta um exemplo de segmentação obtido usando um modelo de forma activa numa imagem de teste. No modelo utilizado, os perfis de intensidade para cada ponto-chave são de dimensão 7 *píxeis*, a percentagem de variação retida foi de 95%, o número máximo de iterações por nível de resolução considerado foi de 10 e a segmentação apresentada iniciou-se no nível de resolução 3. Os resultados de segmentação obtidos para as restantes quatro imagens de teste, utilizando o mesmo modelo, são apresentados na Figura 5.16.

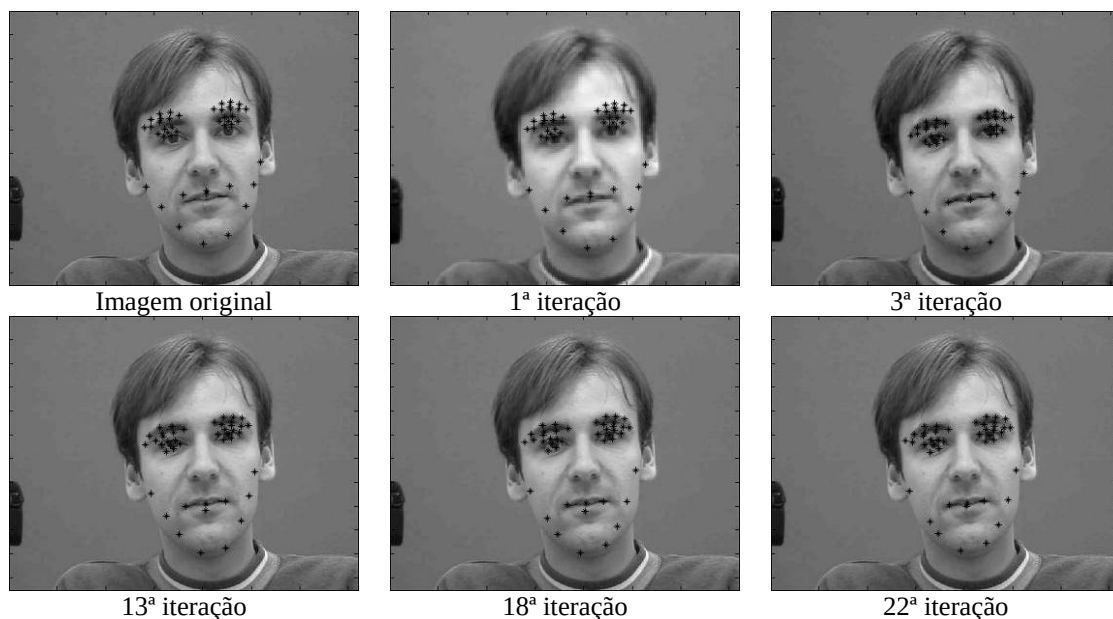


Figura 5.15 Imagem de teste com a posição do modelo médio obtido para a face sobreposto, e após a 1ª, 9ª, 19ª, 25ª e 29ª iteração do processo de segmentação usando o modelo construído.

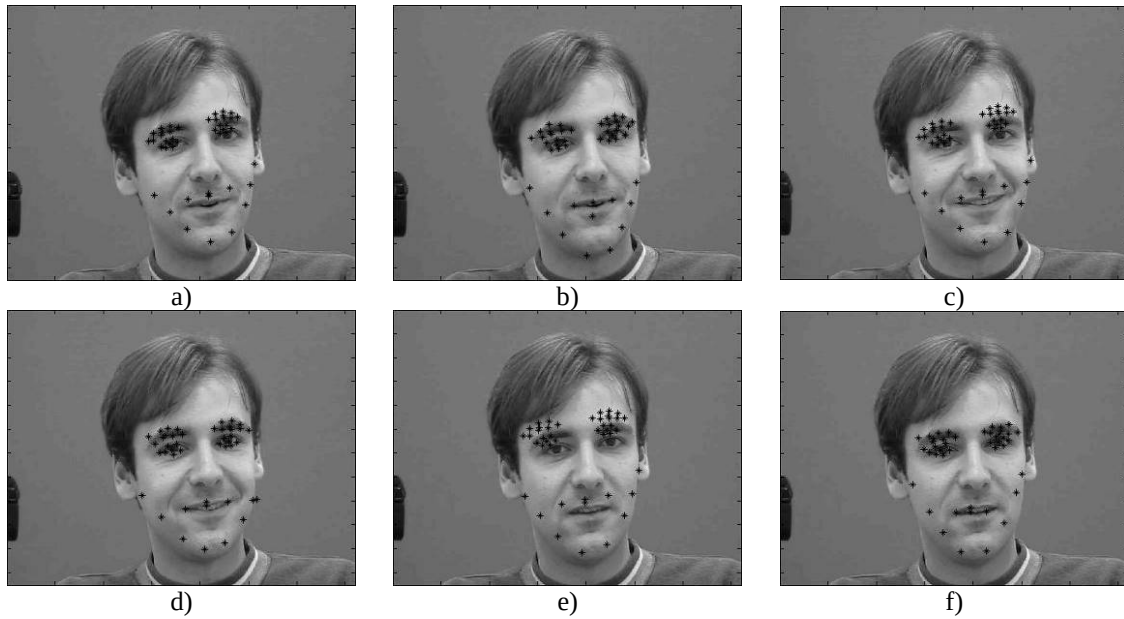


Figura 5.16 Imagens de teste com o modelo médio obtido para a face sobreposto (a, c, e) e resultados finais do processo de segmentação (b, d, f).

Na Tabela 5.7 apresenta-se os valores obtidos da média e do desvio padrão, que traduzem a adequação de cada modelo construído na segmentação do objecto modelado em cada imagem de teste considerada.

Tabela 5.7 Erros obtidos dos modelos de forma da face construídos nas imagens de teste consideradas (média e desvio padrão).

Modelos	Imagem 1	Imagem 2	Imagem 3	Imagem 4
<i>Face A 95 p3</i>	7.76 ± 5.02	7.48 ± 7.71	8.63 ± 8.21	12.10 ± 9.35
<i>Face A 95 p7</i>	6.20 ± 5.26	8.19 ± 7.48	9.05 ± 8.59	7.33 ± 6.16
<i>Face A 95 p16</i>	8.85 ± 6.56	10.23 ± 7.60	10.49 ± 6.87	8.22 ± 6.20
<i>Face A 99 p3</i>	12.45 ± 8.68	10.17 ± 9.30	11.38 ± 10.16	7.40 ± 5.61
<i>Face A 99 p7</i>	7.85 ± 6.65	11.86 ± 8.18	11.27 ± 7.90	9.73 ± 7.32
<i>Face A 99 p16</i>	9.25 ± 6.43	10.49 ± 10.13	9.45 ± 6.71	11.50 ± 8.37
<i>Face B 95 p3</i>	10.75 ± 8.34	8.11 ± 8.47	9.06 ± 8.92	7.53 ± 5.73
<i>Face B 95 p7</i>	6.57 ± 5.51	7.62 ± 6.91	8.85 ± 7.13	10.12 ± 7.88
<i>Face B 95 p16</i>	8.36 ± 6.48	12.58 ± 9.15	10.45 ± 6.87	6.67 ± 6.82
<i>Face B 99 p3</i>	9.59 ± 8.60	10.70 ± 10.22	15.52 ± 11.67	9.31 ± 6.83
<i>Face B 99 p7</i>	10.77 ± 7.54	9.51 ± 6.50	11.53 ± 7.84	7.38 ± 7.16
<i>Face B 99 p16</i>	8.78 ± 6.28	15.00 ± 10.39	9.28 ± 5.71	9.25 ± 5.99

Pela observação dos valores da referida tabela, pode-se verificar que os modelos de forma activa gerados para a face, neste estudo, são sempre capazes de identificar correctamente o objecto modelado em novas imagens. Pode-se constatar ainda, que a utilização de qualquer um dos métodos de alinhamento conduz a bons resultados de segmentação.

Tal como aconteceu no caso da modelação da mão, a utilização dos modelos que retêm 95% da variação total do objecto revela ser mais eficaz do que os modelos que retêm 99% da variação. A justificação para estes resultados é a mesma que para o caso da mão, ou seja, a utilização de mais 5 modos de variação (13 modos retêm 95% da variação e 18 modos retêm 99% da variância total da forma) no processo activo leva a uma maior deformação do modelo nas etapas iniciais da pesquisa, conduzindo assim a piores resultados finais.

Quanto aos resultados obtidos considerando as dimensões dos perfis de intensidade utilizados, verifica-se que os modelos que geram melhores aproximações ao objecto modelado numa nova imagem, são os de perfis de intensidade de dimensão média (15 *píxeis*). A utilização de perfis de intensidade de menor dimensão de (7 *píxeis*) conduzem a resultados menos precisos, devido à dimensão ser demasiado reduzida para localizar correctamente o objecto considerado numa nova imagem. Por outro lado, o uso de perfis de intensidade de dimensão elevada (33 *píxeis*) utilizam informação que não é útil para a pesquisa do objecto modelado em novas imagens, podendo assim conduzir e o modelo a obter resultados menos precisos.

5.4.2 Modelo de Aparência Activa

Nesta secção descreve-se os modelos de aparência activa para a face gerados neste trabalho e os resultados obtidos na segmentação dos objectos modelados em novas imagens.

Neste trabalho construíram-se modelos de aparência activa a partir de 22 imagens de faces etiquetadas automaticamente com 44, 49, 54 e 75 pontos-chave, gerados usando os *algoritmos I, II, III-tipo 1 e III-tipo 2* respectivamente, apresentados nos capítulos anteriores. Assim, geraram-se dois tipos de modelos variando a percentagem de variação retida: o primeiro, explica 95% da variância total das instâncias da forma, e o segundo explica 99% da mesma variância. Construíram-se ainda modelos considerando diferentes números de *píxeis* no modelo de textura (5000, 10000 e 50000 *píxeis*).

Na Tabela 5.8 apresenta-se as características de cada modelo construído, assim como o número de modos retidos por cada um. Verifica-se que o aumento da percentagem de variação da forma retida implica o aumento do número de modos da forma em cada modelo, independentemente do método de etiquetagem utilizado. Observa-se ainda que,

para os modelos construídos com o primeiro algoritmo de etiquetagem (44 pontos-chave) o aumento do número de *píxeis* do modelo de textura provoca pequenas alterações no número de modos de textura retidos. Por último, constata-se que o número de modos de aparência dos modelos construídos com o terceiro algoritmo de etiquetagem utilizando os dois tipos (*tipo 1*: 54 pontos-chave e *tipo 2*: 75 pontos-chave) é idêntico, como se pode observar na tabela referida.

Nas Figura 5.17 a 5.20 apresentam-se exemplos do efeito da variação dos primeiros três modos do modelo de textura e da aparência construídos em torno da forma média, utilizando os vários tipos de etiquetagem. Os modos de variação da forma são os mesmos apresentados no início da secção 5.4. Analisando as referidas figuras, pode-se constatar que os modos de textura consideram a informação das diferentes expressões do indivíduo no conjunto de treino e que os modos de aparência combinam a forma da face com a textura.

Tabela 5.8 Caracterização dos modelos de aparência activa considerados.

Modelos	Percentagem de variação retida	Nº píxeis do modelo de textura	Nº modos de variação da Forma	Nº modos de variação da Textura	Nº modos de variação da Aparência
Face_44_95_5	95%	5000	13	15	12
Face_44_95_10	95%	10000	13	15	12
Face_44_95_50	95%	50000	13	18	13
Face_49_95_5	95%	5000	1	9	4
Face_49_95_10	95%	10000	1	11	5
Face_49_95_50	95%	50000	1	14	6
Face_54_95_5	95%	5000	3	12	6
Face_54_95_10	95%	10000	3	14	8
Face_54_95_50	95%	50000	3	17	9
Face_75_95_5	95%	5000	3	12	6
Face_75_95_10	95%	10000	3	14	7
Face_75_95_50	95%	50000	3	17	9
Face_44_99_5	99%	5000	18	21	19
Face_44_99_10	99%	10000	18	21	19
Face_44_99_50	99%	50000	18	21	19
Face_49_99_5	99%	5000	1	16	11
Face_49_99_10	99%	10000	1	18	13
Face_49_99_50	99%	50000	1	20	15
Face_54_99_5	99%	5000	5	18	14
Face_54_99_10	99%	10000	5	19	15
Face_54_99_50	99%	50000	5	21	18
Face_75_99_5	99%	5000	5	18	14
Face_75_99_10	99%	10000	5	19	15
Face_75_99_50	99%	50000	5	21	18

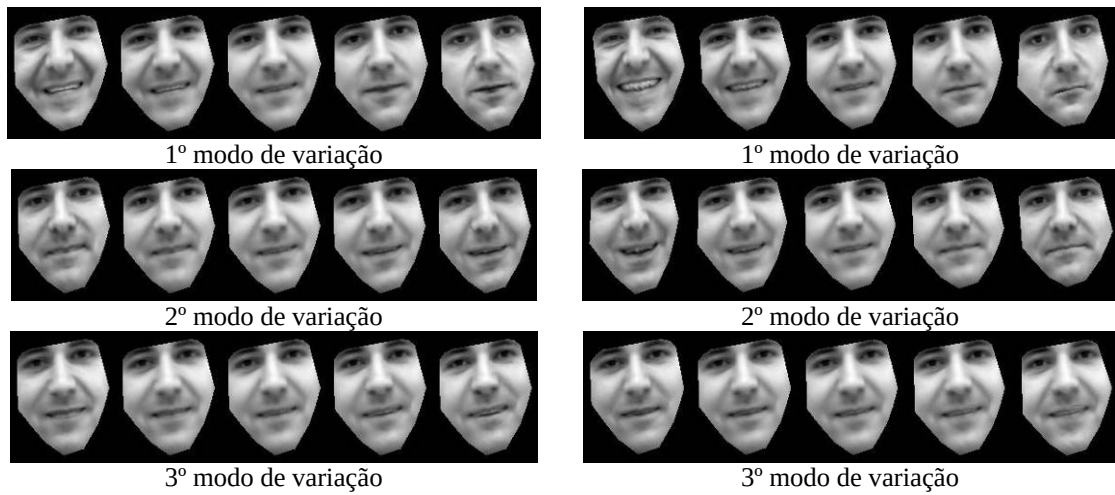


Figura 5.17 Efeito da variação ($\pm 2sd$) dos três primeiros modos do modelo de textura e de aparência construído para a face usando o *algoritmo I* de etiquetagem.

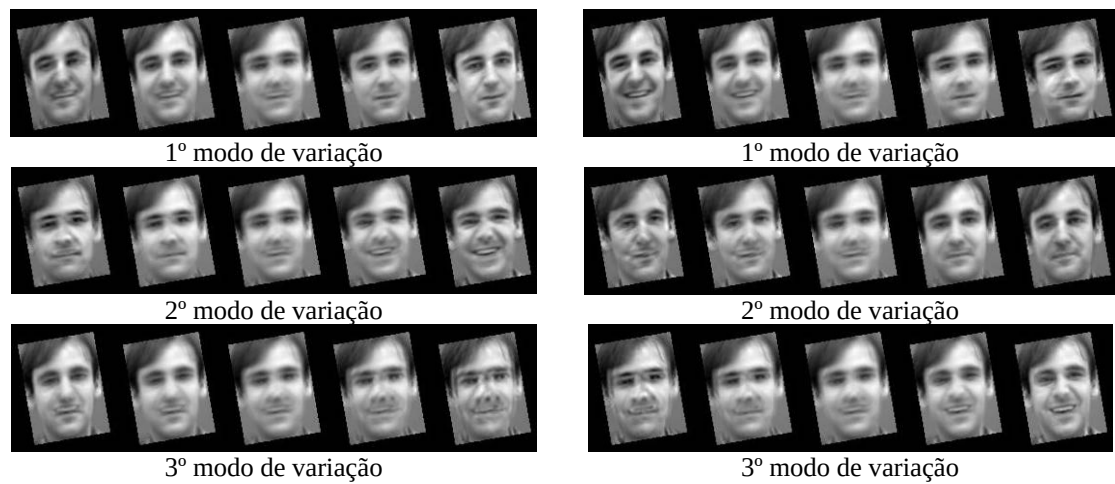


Figura 5.18 Efeito da variação ($\pm 2sd$) dos três primeiros modos do modelo de textura e de aparência construído para a face usando o *algoritmo II* de etiquetagem.

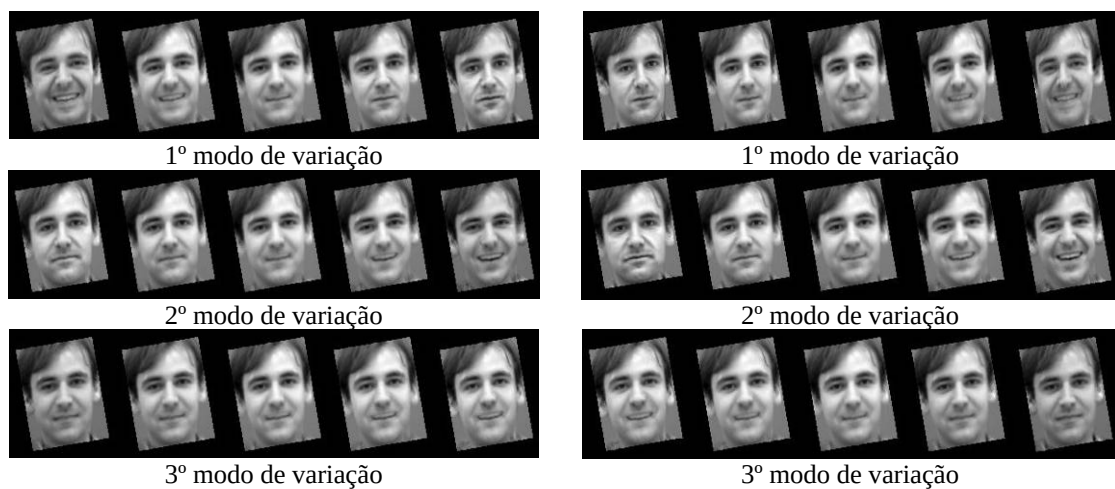


Figura 5.19 Efeito da variação ($\pm 2sd$) dos três primeiros modos do modelo de textura e de aparência construído para a face usando o *algoritmo III- tipo 1* de etiquetagem.

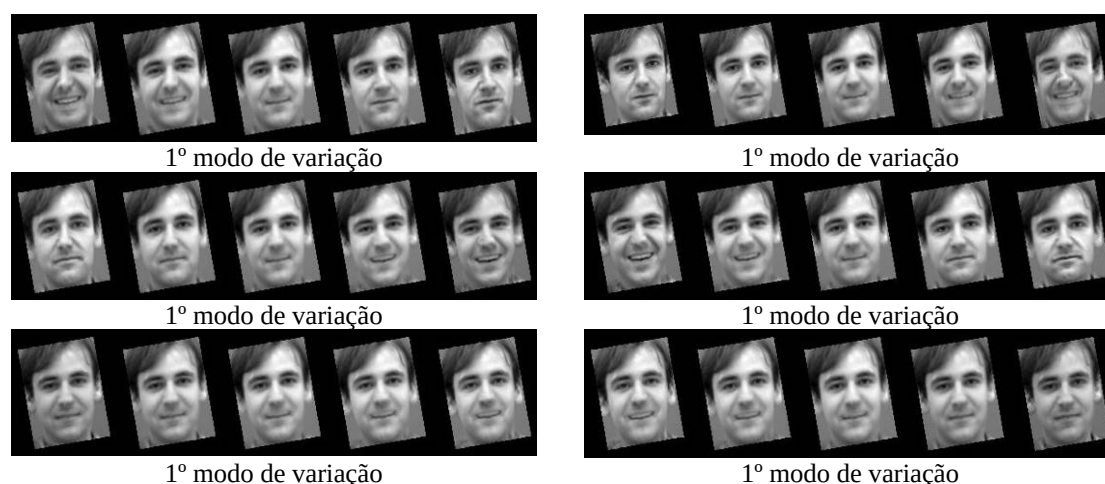


Figura 5.20 Efeito da variação ($\pm 2sd$) dos três primeiros modos do modelo de textura e de aparência construído para a face usando o *algoritmo III-tipo 2* de etiquetagem.

No sentido de estudar a eficácia dos modelos de aparência activa no reconhecimento de objectos do tipo face em novas imagens, foram utilizadas 4 imagens de teste, Figura 5.14. Assim, na fase de pesquisa, o número máximo de iterações por nível foi limitado a 10 e a segmentação iniciou-se no nível de resolução 5. A Figura 5.21 apresenta algumas iterações do processo computacional do segundo modelo de aparência activa construído (com 95% de variação retida e 10000 *píxeis* usados) na primeira imagem de teste considerada usando o *algoritmo I* de etiquetagem automática. Por seu lado, a Figura 5.22 apresenta os resultados obtidos para a mesma imagem de teste usando os restantes algoritmos de etiquetagem automática.

Na Tabela 5.9 apresenta-se os erros resultantes da segmentação do objecto modelado nas imagens de teste consideradas, utilizando os modelos descritos anteriormente. Os valores da média e do desvio padrão dos erros obtidos em cada imagem de teste apresentados na mesma tabela, permitem retirar algumas conclusões sobre os melhores parâmetros a utilizar na modelação da face. Mais uma vez, nesta tabela, as células que não apresentam valores referem-se aos casos nos quais o modelo usado não obteve bons resultados.

Pela análise dos valores da tabela referida, é possível constatar que os modelos considerados para a face conseguem, com mais ou menos precisão, localizar os objectos modelados nas novas imagens de teste usadas.

Tal como acontece no caso de modelação da mão, verifica-se que os modelos que retêm 99% da variância total obtêm melhores resultados do que os que retêm 95% da variabilidade total do objecto em estudo. Mais uma vez, tal resultado justifica-se pelo

facto de quanto maior for a percentagem de variação retida pelo modelo, menor será o erro obtido na pesquisa em novas imagens, pois é utilizada mais informação útil sobre o objecto descrito nas várias imagens que integram o conjunto de treino considerado.

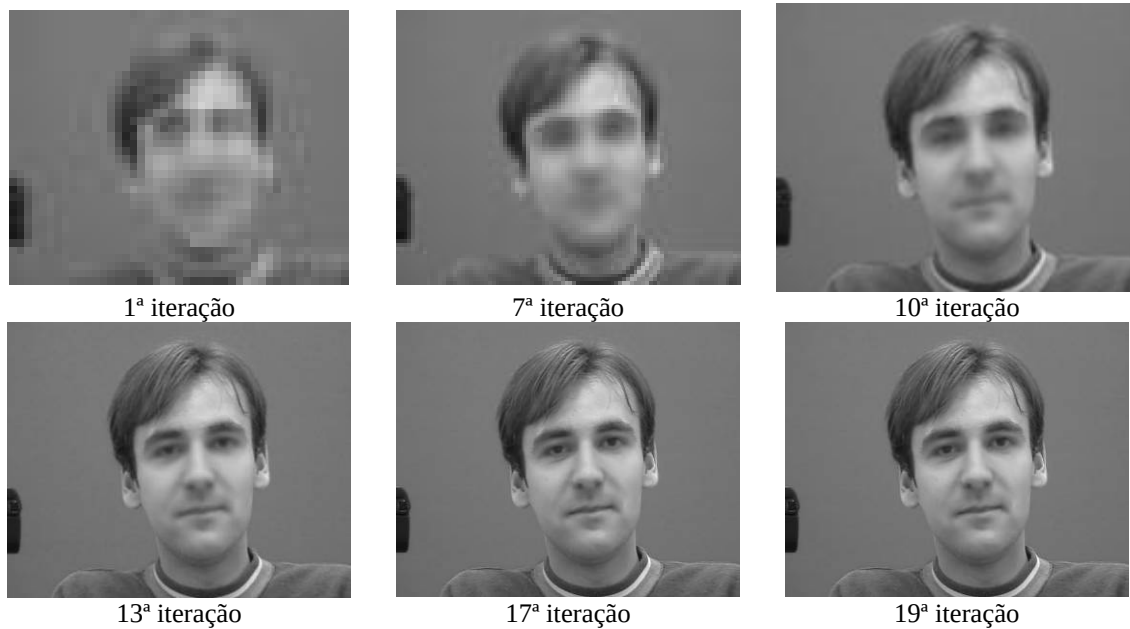


Figura 5.21 Alguns resultados de iterações do processo de segmentação numa imagem de teste, usando o modelo de aparência construído para a face com o *algoritmo I* de etiquetagem.

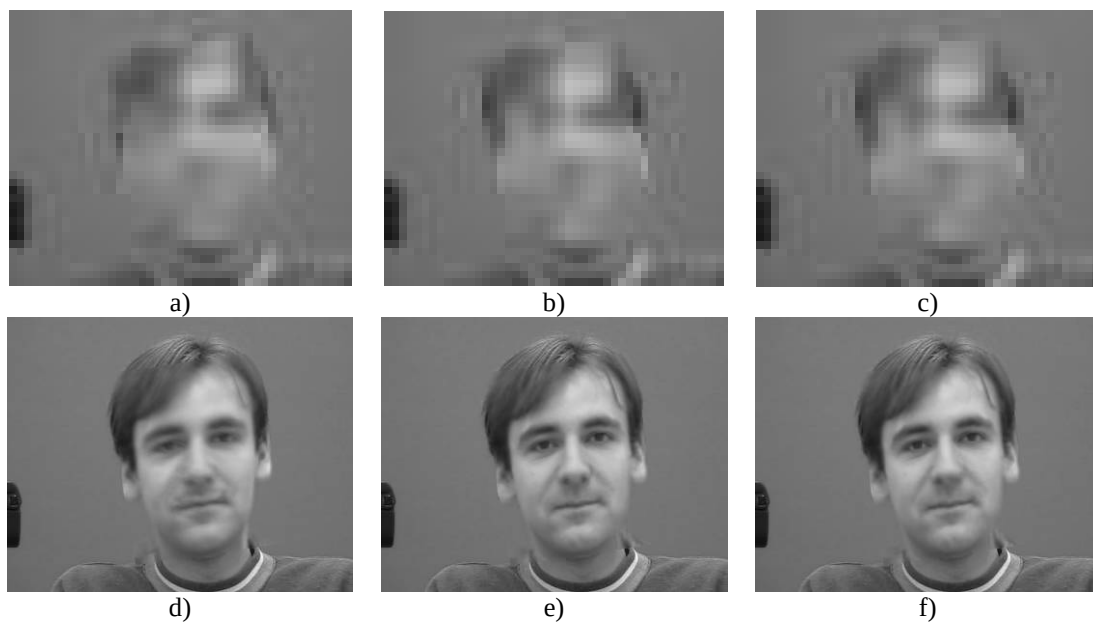


Figura 5.22 Imagem de teste após a primeira iteração do modelo de aparência activa construído (a, b, c) e resultados finais do processo de segmentação em novas imagens usando o *algoritmo II* de etiquetagem (d), *algoritmo III-tipo 1* de etiquetagem (e) e *algoritmo III-tipo 2* de etiquetagem (f).

Tabela 5.9 Erros resultantes da segmentação do objecto modelado nas imagens de teste, utilizando os modelos de aparência activa construídos para a face.

Modelos	Imagem 1	Imagem 2	Imagem 3	Imagem 4
Face_44_95_5	6.14 ± 4.11	4.81 ± 4.50	5.05 ± 3.63	
Face_44_95_10	6.11 ± 4.15	4.74 ± 4.52	5.06 ± 3.70	5.68 ± 4.16
Face_44_95_50	5.17 ± 3.59	4.08 ± 4.46	4.83 ± 3.69	5.37 ± 3.97
Face_49_95_5	3.17 ± 1.09	2.79 ± 0.49	4.92 ± 1.25	
Face_49_95_10	2.93 ± 0.22	1.30 ± 0.10	3.60 ± 1.80	2.54 ± 1.77
Face_49_95_50	2.97 ± 0.54	1.30 ± 0.17	3.51 ± 1.69	2.56 ± 1.76
Face_54_95_5	1.52 ± 0.83	3.63 ± 3.65	2.86 ± 1.52	2.08 ± 1.27
Face_54_95_10	1.58 ± 0.43	3.39 ± 3.31	2.60 ± 1.90	1.70 ± 0.85
Face_54_95_50	1.56 ± 0.64	3.55 ± 3.24	2.70 ± 0.94	1.71 ± 0.91
Face_75_95_5	1.47 ± 0.78	3.59 ± 3.59	2.77 ± 1.61	2.05 ± 1.14
Face_75_95_10	1.65 ± 0.54	3.69 ± 3.57	2.70 ± 1.04	2.08 ± 1.19
Face_75_95_50	1.56 ± 0.60	3.39 ± 3.28	2.53 ± 0.81	1.63 ± 0.85
Face_44_99_5	5.18 ± 3.20	4.68 ± 4.91	5.24 ± 3.57	5.59 ± 4.17
Face_44_99_10	5.23 ± 3.18	4.56 ± 4.71	5.31 ± 3.50	
Face_44_99_50	5.20 ± 3.35	4.31 ± 4.55	4.90 ± 3.48	6.04 ± 5.02
Face_49_99_5	1.25 ± 0.23	1.41 ± 0.06	3.48 ± 1.62	2.46 ± 1.42
Face_49_99_10	1.40 ± 0.22	1.34 ± 0.13	3.54 ± 1.53	2.39 ± 1.56
Face_49_99_50	2.12 ± 0.12	1.26 ± 0.03	3.78 ± 1.49	2.41 ± 1.60
Face_54_99_5	2.17 ± 0.64			
Face_54_99_10	1.88 ± 0.71	3.61 ± 3.53		3.34 ± 1.71
Face_54_99_50	1.67 ± 0.52	3.37 ± 3.54	2.61 ± 1.01	3.05 ± 1.61
Face_75_99_5	2.16 ± 0.60			3.29 ± 1.51
Face_75_99_10	1.80 ± 0.67	3.49 ± 3.51	2.77 ± 1.11	3.29 ± 1.58
Face_75_99_50	1.66 ± 0.53	3.27 ± 3.39	2.60 ± 0.98	3.03 ± 1.58

Quanto à influência do número de *píxeis* utilizados na construção dos modelos de aparência, verifica-se que o erro de segmentação diminui quando este número aumenta. Este resultado era esperado, pois quanto maior o número de *píxeis* usados maior a informação da textura do objecto modelado, conduzindo assim a resultados mais eficazes na pesquisa do mesmo em novas imagens.

Se considerarmos os vários tipos de etiquetagem utilizados e os resultados de segmentação obtidos por cada um dos respectivos modelos, verifica-se que o *algoritmo I* de etiquetagem obtém os piores resultados, e que o *algoritmo III -tipo 2* é o que obtém os melhores resultados, no caso da percentagem de variação retida ser de 95%. Por outro lado, se a percentagem de variação retida for de 99%, o método que obtém resultados mais precisos é o *algoritmo II* de etiquetagem. No entanto, é de notar que tanto o *algoritmo II* como os *algoritmos III* obtêm sempre melhores resultados de segmentação quando comparados com o *algoritmo I*, o que leva a concluir que a colocação dos pontos-chave, no caso da face, não tem de ser necessariamente nos contornos da mesma.

Quanto à utilização de maior número de pontos-chave, os resultados obtidos são idênticos aos da modelação da mão: verifica-se que os erros de segmentação obtidos através dos modelos de forma activa diminuem, e que o mesmo acontece com os modelos de aparência activa, sendo neste caso a diminuição menos notória.

Por último, tendo em conta as modelações da mão e da face, se se considerar os erros resultantes da segmentação, utilizando os modelos de forma activa e os modelos de aparência activa, e tiver em conta que a segmentação em ambos os modelos considera a mesma posição inicial, observa-se que os modelos de aparência activa obtêm melhores resultados, pois a média dos erros e os desvios padrões são sempre menores, como se pode verificar pela comparação das Tabela 5.3 e Tabela 5.5 e das Tabela 5.7 e Tabela 5.9.

5.5 Resumo

Neste capítulo, estudou-se a influência de vários parâmetros na construção dos modelos de forma e de aparência activa, como o número de pontos-chave, o método de alinhamento, a percentagem de variação retida, o comprimento dos perfis de intensidade, e o número de *píxeis* utilizados no modelo de textura. A influência destes parâmetros, foi abordada de acordo com os resultados obtidos na segmentação dos objectos modelados em novas imagens.

Assim, apresentaram-se resultados dos modelos de forma e de aparência activa construídos para objectos do tipo mão e face, e verificou-se o seu comportamento na segmentação destes objectos em novas imagens, utilizando os vários métodos automáticos de etiquetagem desenvolvidos e apresentados nos capítulos anteriores.

No próximo capítulo, são apresentadas as principais conclusões obtidas ao longo desta Dissertação e referidas algumas perspectivas de desenvolvimento futuro.

Capítulo 6

Conclusões e Perspectivas de Desenvolvimento

Futuro

6.1 Conclusões Finais

Nesta Dissertação foi apresentado um estudo detalhado sobre os modelos pontuais de distribuição, as suas variantes principais e as aplicações existentes para este tipo de modelos. Para além disso, procurou-se ainda melhorar a construção dos respectivos modelos, automatizando-se a etapa de colocação e emparelhamento dos pontos-chave para objectos do tipo mão e face.

Assim, o trabalho efectuado ao longo desta Dissertação foi desenvolvido em três fases distintas: na primeira, efectuou-se o estudo das várias etapas necessárias para a construção dos modelos pontuais de distribuição, de forma activa e de aparência activa; na segunda, desenvolveram-se e adaptaram-se métodos computacionais para etiquetar e emparelhar automaticamente objectos representados em imagens, de forma a automatizar totalmente o processo de construção dos modelos considerados; na terceira e última fase, construíram-se vários modelos de forma e de aparência activa para a mão e para a face e analisaram-se detalhadamente os resultados experimentais obtidos.

O modelo pontual de distribuição, dá informação sobre a forma geométrica média de um objecto assim como os seus desvios admissíveis para a mesma, partindo de um conjunto de imagens de treino do mesmo objecto. A construção destes modelos é efectuada em três etapas: na primeira, o objecto em estudo, presente em cada imagem do conjunto de treino, é representado por um conjunto de pontos-chave, como os pontos notáveis do contorno, e os pontos-chave de todas as suas formas em análise no conjunto de treino são emparelhados; na segunda, é realizado o alinhamento dos pontos-chave correspondentes para ser possível a sua posterior comparação; finalmente, na terceira e última etapa é efectuada uma análise em componentes principais, no sentido de capturar a variabilidade da forma do objecto modelado e a reduzir a dimensionalidade dos dados usados no modelo.

Através dos modelos pontuais de distribuição é possível construir modelos capazes de segmentar e reconhecer os objectos modelados em novas imagens, como os modelos de forma activa e os modelos de aparência activa. Estes modelos são construídos através da combinação da forma geométrica dos objectos e dos níveis de cinzento dos mesmos nas várias imagens que integram o conjunto de treino. Os modelos activos gerados podem ser assim aplicados em várias áreas como, por exemplo: em medicina, na localização de ossos e órgãos em imagens médicas; na indústria, na inspecção de produtos industriais; e em segurança, na identificação e reconhecimento de faces.

Como já referido, ao longo deste trabalho, foram ainda desenvolvidos de raiz novos algoritmos para automatizar a etapa de colocação de pontos-chave do objecto nas várias imagens do conjunto de treino. Esta etapa, é a mais morosa de todo o processo de construção dos modelos pontuais de distribuição pois, geralmente, é efectuada de modo manual. Os algoritmos desenvolvidos, são capazes de etiquetar automaticamente objectos do tipo mão e face representados em imagens. A utilização deste tipo de objectos, permitiu neste trabalho um estudo adequado dos modelos estatísticos desenvolvidos.

Foram desenvolvidos quatro algoritmos para a colocação automática de pontos-chave. Um dos algoritmos, aplica-se a objectos do tipo mão, e utiliza um método de detecção de zonas de pele para identificar primeiramente a mão na imagem em questão. Posteriormente, considera o contorno obtido para extrair a região correspondente apenas à mão e considera, como pontos-chave, os pontos contidos nas zonas de maior curvatura e entre cada uma destas zonas. Os restantes algoritmos desenvolvidos, aplicam-se à extracção de pontos notáveis da face, utilizando-se também o método de detecção de zonas de pele para localizar, numa primeira fase, a face na imagem em causa. Assim, um dos métodos coloca pontos-chave nas regiões do contorno da face, dos olhos, das sobrancelhas e da boca; o segundo método, considera os pontos-chave como sendo os nodos de uma malha rectangular colocada adequadamente sobre a face; e o terceiro e último método, conjuga os dois métodos anteriores, considerando assim os pontos como os nodos de uma malha adaptativa e duas sub-malhas, colocadas respectivamente sobre a face, olhos e boca.

Caso o emparelhamento entre os pontos-chave do objecto nas várias imagens do conjunto de treino não seja dado de forma implícita pela ordem destes, tal como acontece com os algoritmos de etiquetagem automática desenvolvidos neste trabalho, torna-se desejável um método de emparelhamento automático. Assim, estudaram-se e

adoptaram-se as implementações computacionais de dois métodos que permitem emparelhar automaticamente os pontos-chave: o método de análise modal das formas a emparelhar, proposto por Shapiro [Tavares, 2000]; e o método baseado na modelação física das mesmas, proposto por Sclaroff [Tavares, 2000]. Para ambos os métodos de emparelhamento, as correspondências podem ser obtidas considerando um critério local de busca ou, pelo contrário, um procedimento global de pesquisa. Esta última possibilidade tem a vantagem, relativamente à primeira, de considerar a estrutura global das formas a emparelhar.

As implementações computacionais previamente existentes usadas neste trabalho para a construção dos modelos de forma activa e dos modelos de aparência activa, foram adaptadas de modo a incluir as implementações desenvolvidas para etiquetar e emparelhar automaticamente os pontos-chave, e certos parâmetros foram alterados de forma a obter-se melhores desempenhos com os objectos considerados. Através destas implementações, construíram-se vários modelos activos, considerando diferentes parâmetros, como o número de pontos-chave, o método de alinhamento, a percentagem de variação retida pelos modelos de forma e de aparência, e a quantidade de informação dos níveis de cinzento a utilizar, de modo a encontrar os parâmetros que melhor se ajustam a objectos considerados (mão e face).

As principais conclusões, retiradas deste trabalho através dos vários resultados experimentais obtidos ao longo do mesmo, são indicadas em seguida.

Modelos de Forma Activa:

- A utilização dos dois métodos de alinhamento apresentados nesta Dissertação, um que tem em conta as variâncias dos pontos-chave e outro que considera os pontos-chave todos com igual peso, conduzem igualmente a bons resultados de segmentação. De uma forma geral, o primeiro método obtém resultados mais precisos no reconhecimento de objectos em novas imagens, pois o peso dos pontos mais significativos do objecto é maior nos modelos gerados.
- Os modelos que obtêm melhores resultados na segmentação de objectos em novas imagens, são os que retêm 95% da variação total da forma; ou seja, ao contrário do previsto, não são os modelos que retêm 99% da variância. Esta situação, ocorre devido ao modelo activo construído com uma percentagem de variância retida de 99% ter tendência para deformar em demasia a forma do

modelo nas primeiras iterações da pesquisa (como utiliza um maior número de modos de variação é possível que inclua mais ruído no modelo).

- A utilização de diferentes dimensões dos perfis de intensidade nos modelos também influencia a pesquisa de objectos em novas imagens. Para a modelação da mão, verificou-se que os modelos que obtêm melhores resultados são os que utilizam perfis de dimensão reduzida (7 *píxeis*) ou média (15 *píxeis*), e para a modelação da face os melhores modelos construídos foram os que usam perfis de dimensão média (15 *píxeis*). Obviamente que estas dimensões estão de acordo com as dimensões dos objectos considerados neste trabalho.

Modelos de Aparência Activa:

- Tal como era esperado, nos casos estudados, os modelos de aparência que obtêm resultados mais eficazes na pesquisa de objectos modelados em novas imagens, são os que retêm 99% da variação do objecto modelado.
- A utilização de um maior número de *píxeis* na construção dos modelos de aparência, traduz-se em melhores resultados de segmentação do objecto modelado em novas imagens. Foram estudados modelos que utilizam 5000, 10000 e 50000 *píxeis* para construir os modelos de textura, e os que obtiveram melhores resultados, tanto na modelação da mão como na modelação da face, foram os que usaram mais *píxeis*. Estes resultados também eram os esperados, pois quanto maior o número de *píxeis* utilizados, até um dado valor máximo função das dimensões do objecto em questão, maior a quantidade de informação da textura do objecto, e portanto menor o erro obtido na localização dos objectos modelados em novas imagens

Conclusões gerais:

- Os modelos de forma activa pesquisam à volta de cada ponto-chave, possuem um maior conhecimento da área circundante do objecto modelado, e obtêm mais rapidamente o objecto em operações de segmentação. Por outro lado, os modelos de aparência activa apenas examinam a região interna do objecto, dificultando assim a operação de segmentação do objecto modelado em novas imagens.
- Os modelos de forma activa apenas utilizam os dados existentes à volta de cada ponto-chave e não consideram a informação dos níveis de cinzento existente ao

longo de todo o objecto, como fazem os modelos de aparência activa, e por isso tendem a ser menos robustos. Por este facto, nos modelos da forma é aconselhável que os pontos-chave sejam colocados em locais característicos dos objectos a modelar, como pontos do contorno ou vértices, assim esta operação torna-se crítica na construção destes modelos.

- Uma outra vantagem dos modelos de aparência activa, é a possibilidade de construir um modelo robusto utilizando um número reduzido de pontos-chave; ao contrário do que acontece com os modelos de forma, que necessitam de um número razoável de pontos-chave para obterem bons resultados. Assim, em casos nos quais a etapa de colocação de pontos-chave não seja um processo automático, a utilização de modelos de aparência é preferível.

6.2 Perspectivas de desenvolvimento futuro

O trabalho desenvolvido ao longo desta Dissertação alcançou resultados bastante satisfatórios; no entanto, pode ser enriquecido, melhorado e continuado segundo diversas vertentes:

- A automatização da construção dos modelos activos é apenas possível para objectos do tipo mão e face. O desenvolvimento de métodos capazes de extrair automaticamente os pontos-chave em objectos de outro tipo, como por exemplo ossos e órgãos em imagens médicas, permitiria uma maior utilização destes modelos na área médica. Outra possibilidade, seria a consideração de objectos 3D.
- A influência do número de imagens de treino usado na construção dos modelos de distribuição pontual, é outra vertente deste trabalho que pode ser considerada. Ainda neste âmbito, deverão ser estudadas alternativas de construção de modelos robustos usando números reduzidos de imagens de treino. Uma possibilidade, poderá passar por incluir a informação obtida a partir de novas imagens sobre o objecto em questão no modelo construído para o mesmo.
- A fase de emparelhamento dos pontos-chave dos objectos poderá ser também tema de desenvolvimentos futuros. Uma possibilidade, seria, por exemplo, a

utilização de técnicas de optimização com restrições de suavização da deformação envolvida.

- A comparação entre estes modelos de base estatística e os modelos construídos segundo princípios físicos, na segmentação de objectos representados em imagens, pode também constituir uma tarefa interessante.
- A possibilidade de englobar algum conhecimento preexistente sobre as propriedades físicas dos objectos modelados nos modelos considerados neste trabalho, poderá melhorar o comportamento futuro destes modelos pois serão mais realistas.
- Algumas das implementações computacionais desenvolvidas em *MATLAB* são bastante morosas. Como, por exemplo, o emparelhamento dos pontos-chave usando a metodologia física de Sclaroff. Uma possibilidade interessante de trabalho futuro, seria portar as implementações desenvolvidas para uma linguagem mais eficiente, como por exemplo a linguagem *C*. Ainda, em termos de implementação, também seria interessante reunir numa única plataforma computacional todas as implementações desenvolvidas, adaptadas ou usadas neste trabalho de forma que, através de uma interface gráfica adequada, os seus utilizadores pudessem usar, testar e comparar as várias metodologias desenvolvidas em diferentes e variados objectos.

Bibliografia

Angelopoulou, A. N. and Psarrou, A. (2004). *Evaluating Statistical Shape Models for Automatic Landmark Generation on a Class of Human Hands*. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Istanbul.

Baker, S. and Matthews, I. (2002). *Automatic Construction of Active Appearance Models as an Image Coding Problem*. IEEE Transactions on Pattern Analysis and Machine Intelligence 26: 1380-1384.

Baker, S. and Matthews, I. (2004). *Automatic Construction of Active Appearance Models as an Image Coding Problem*. IEEE Transactions on Pattern Analysis and Machine Intelligence 26: 1380-1384.

Bastos, M. L. (2003). *Dissertação de Mestrado: Optimização da Determinação das Correspondências entre Objectos Deformáveis no Espaço Modal*. Universidade do Porto.

Baumberg, A. and Hogg, D. (1994). *Learning Flexible Models from Image Sequences*. 3rd European Conference on Computer Vision, Sweden, Stockholm.

Blake, A. and Isard, M. (1998). *Active Contours*, Springer-Verlag.

Brett, A. D., Hill, A. and Taylor, C. J. (1997). *A Method of 3D Surface Correspondence for Automated Landmark Generation*. 8th British Machine Vision Conference, Essex, England.

Brunelli, R. and Poggio, T. (1993). *Face recognition: Features versus templates*. IEEE Transactions on Pattern Analysis and Machine Intelligence 15 (10): 1042-1052.

Campadelli, P., Cusmai, F. and Lanzarotti, R. (2003). *A color based method for face detection*. International Symposium on Telecommunications, Isfahan, Iran.

Carvalho, F. J. S. and Tavares, J. M. R. S. (2005). *Metodologias para identificação de faces em imagens: Introdução e exemplos de resultados*. Congresso de Métodos Numéricos en Ingeniería 2005, Granada, Espanha.

Cootes, T. F. (2004). *Build_aam*,
http://www.wiau.man.ac.uk/~bim/software/am_tools_doc/download_win.html.

Cootes, T. F. (2004). *Talking Face*,
http://www.isbe.man.ac.uk/~bim/data/talking_face/talking_face.html.

Cootes, T. F., Edwards, G. and Taylor, C. J. (1999). *Comparing Active Shape Models with Active Appearance Models*. Department of Medical Biophysics, Manchester University.

- Cootes, T. F., Edwards, G. J. and Taylor, C. J. (1998). *Active Appearance Models*. Proceedings of European Conference on Computer Vision.
- Cootes, T. F. and Taylor, C. J. (1992a). *Active Shape Models - 'Smart Snakes'*. Proceedings of the British Machine Vision Conference, Leeds.
- Cootes, T. F. and Taylor, C. J. (1993). *Active Shape Model Search using Local Grey-Level Models: A Quantitative Evaluation*. British Machine Vision Conference, BMVA Press: 639/648.
- Cootes, T. F., Taylor, C. J., Cooper, D. H. and Graham, J. (1992). *Training Models of Shape from Sets of Examples*. Proceedings of the British Machine Vision Conference, Leeds.
- Cootes, T. F., Taylor, C. J., Cooper, D. H. and Graham, J. (1992). *Training Models of Shape from Sets of Examples*. Proc. British Machine Vision Conference, Leeds.
- Cootes, T. F., Taylor, C. J. and Lanitis, A. (1994). *Active Shape Models: Evaluation of a Multi-Resolution Method for Improving Image Search*. British Machine Vision Conference, BMVA.
- Deriche, R. (1987). *Using Canny's Criteria to derive a Recursively Implemented Optimal Edge Detector*. International Journal of Computer Vision: 167-187.
- Duda, R. O., Hart, P. E. and Stork, D. G. (2001). *Pattern Classification*, John Wiley & Sons, Inc.
- Fukunaga, K. and Koontz, W. L. G. (1970). *Application of the Karhunen-Loeve Expansion to Feature Selection and Ordering*. IEEE Transactions on Computers
- Gargesha, M. and Panchanathan, S. (2002). *A Hybrid Technique for Facial Feature Point Detection*. Fifth IEEE Southwest Symposium on Image Analysis and Interpretation, Santa Fe, New Mexico, USA.
- Gower, J. C. (1975). *Generalized Procrustes Analysis*, Psychometrika.
- Hamarneh, G. (1999). *Active Shape Models, Modeling Shape Variations and Gray Level Information and an Application to Image Search and Classification*.
- Hamarneh, G. (1999a). *ASM (MATLAB)*,
<http://www.cs.sfu.ca/~hamarneh/software/code/asm.zip>.
- Hicks, Y., Marshall, D., Martin, R. R., Rosin, P. L., Bayer, M. M. and Mann, D. G. (2002). *Automatic Landmarking for Building Biological Shape Models*. International Conference of Image Processing, Rochester, USA 2: 801-804.
- Hill, A. and Taylor, C. J. (1994). *Automatic Landmark Generation for Point Distribution Models*. Fifth British Machine Vision Conference, England, York, BMVA Press.

Hill, A. and Taylor, C. J. (1996). *A Method of Non-rigid Correspondence for Automatic Landmark Identification*. 7th British Machine Vision Conference, Edinburgh, Scotland.

Hsu, R., Abdek-Mottaleb, M. and Jain, A. K. (2002). *Face detection in color images*. IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (5): 696-706.

Jones, M. J. and Rehg, J. M. (1999). *Statistical Color Models with application to skin detection*. IEEE Conference on Computer Vision and Pattern Recognition, Ft. Collins, CO, USA.

Kass, M., Witkin, A. and Terzopoulos, D. (1987). *Snakes: Active Contour Models*. International Journal of Computer Vision 1: 321-331.

Lim, J. S. (1990). *Two-Dimensional Signal and Image Processing*, PTR Prentice Hall.

Lima, R. S. (2003). *Dissertação de Mestrado: Modelos de Forma Activos na detecção de contornos pulmonares em radiografias torácicas*. Faculdade de Engenharia da Universidade do Porto. Porto.

Moghaddam, B. and Pentland, A. (1994). *Face recognition using view-based and modular eigenspaces*. Automatic Systems for the Identification and Inspection of Humans 2277(SPIE).

Pentland, A. and Sclaroff, S. E. (1991). *Closed-form solutions for physically based modelling and recognition*. IEEE Transactions on Pattern Analysis and Machine Intelligence 13 7: 715-729.

Sclaroff, S. E. (1995). *PhD Thesis: Modal Matching: A Method for Describing, Comparing, and Manipulating Digital Signals*. Massachusetts Institute of Technology.

Sclaroff, S. E. (1995). *PhD Thesis: Modal Matching: A Method for Describing, Comparing, and Manipulating Digital Signals*.

Sclaroff, S. E. (1995a). *Modal Matching for Correspondence and Recognition*, Massachusetts Institute of Technology, Media Laboratory.

Shapiro, L. (1991). *Towards a Vision-Based Motion Framework*, Robotics Research Group, Department of Engineering Science, Oxford University.

Sheu, H. T. and Hu, W. C. (1999). *Multiprimitive Segmentation of Planar Curves - A Two-Level Breakpoint Classification and Tuning Approach*. IEEE Transactions on Pattern Analysis and Machine Intelligence 21.

Silva, S. (2002). *Relatório: Análise de deformações de objectos visíveis, recorrendo à sua modelização por elementos finitos*. Porto, Instituto de Engenharia Biomédica.

Stegmann, M. B. and Gomez, D. D. (2002). *Hand images*, http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=403.

- Tavares, J. M. R. S. (1995). *Dissertação de Mestrado: Obtenção de Estrutura Tridimensional a Partir de Movimento de Câmara*. Faculdade de Engenharia da Universidade do Porto. Porto.
- Tavares, J. M. R. S. (2000). *Tese de Doutoramento: Análise de Movimento de Corpos Deformáveis usando Visão Computacional*. Faculdade de Engenharia da Universidade do Porto. Porto.
- Terzopoulos, D. and Metaxas, D. (1991). *Dynamic 3D models with local and global deformations: Deformable superquadrics*. IEEE Transactions on Pattern Analysis and Machine Intelligence 13 7: 703-714.
- Tien, F.-C., Yeh, C.-H. and Hsieh, K.-H. (2004). *Automated visual inspection for microdrills in printed circuit board production*. International Journal of Production Research 42, n12: 2477-2495.
- Wu, A., Shah, M. and Lobo, N. V. (2000). *A Virtual 3D Blackboard: 3D Finger Tracking using a Single Camera*.
- Yuille, A. L., Cohen, D. and Hallinan, P. (1992). *Feature extraction from faces using deformable templates*. International Journal of Computer Vision 8: 104-109.
- Zheng, H., Daoudiy, M. and Jedynek, B. (2004). *Blocking Adult Images Based on Statistical Skin Detection*. Electronic Letters on Computer Vision and Image Analysis 4(1-14).

Anexo A

Alinhamento de dois objectos:

Dadas duas instâncias do mesmo objecto, x_1 e x_2 deseja-se encontrar a rotação, θ , a escala s e a translação (t_x, t_y) que transformam $M(s, \theta)[x_2]$ em x_1 de forma a minimizar a soma:

$$E = (x_1 - M(s, \theta)[x_2])^T W (x_1 - M(s, \theta)[x_2]),$$

onde:

$$M(s, \theta) \begin{bmatrix} x_{kj} \\ y_{kj} \end{bmatrix} = \begin{bmatrix} (s \cos \theta) x_{kj} - (s \sin \theta) y_{kj} + t_x \\ (s \sin \theta) x_{kj} + (s \cos \theta) y_{kj} + t_y \end{bmatrix},$$

e W é a matriz de pesos diagonal para cada ponto.

Se se escrever $a_x = s \cos \theta$ e $a_y = s \sin \theta$, então a aproximação de mínimos quadrados leva às seguintes equações lineares:

$$\begin{pmatrix} X_2 & -Y_2 & W & 0 \\ Y_2 & X_2 & 0 & W \\ Z & 0 & X_2 & Y_2 \\ 0 & Z & -Y_2 & X_2 \end{pmatrix} \begin{pmatrix} a_x \\ a_y \\ t_x \\ t_y \end{pmatrix} = \begin{pmatrix} X_1 \\ Y_1 \\ C_1 \\ C_2 \end{pmatrix},$$

sendo:

$$\begin{aligned} X_i &= \sum_{k=0}^{n-1} w_k x_{ik} & Y_i &= \sum_{k=0}^{n-1} w_k y_{ik} \\ Z &= \sum_{k=0}^{n-1} w_k (x_{2k}^2 + y_{2k}^2) & W &= \sum_{k=0}^{n-1} w_k \\ C_1 &= \sum_{k=0}^{n-1} w_k (x_{1k} x_{2k} + y_{1k} y_{2k}) & C_2 &= \sum_{k=0}^{n-1} w_k (y_{1k} y_{2k} - x_{1k} x_{2k}) \end{aligned}.$$

Anexo B

Estimar os parâmetros de ajuste:

Suponha-se que se tem uma forma definida por n pontos num vector x relativo ao centro do modelo (X_c, Y_c) e que se quer encontrar a translação (dX_c, dY_c) , rotação $d\theta$ em torno do centro do modelo, e o factor de escalamento $(1+ds)$ que melhor ajusta o conjunto de pontos, X , ao conjunto de pontos dados por $(X + dX)$. A translação é dada por:

$$dX_c = \frac{1}{n} \sum_{j=0}^{n-1} dX_j \quad dY_c = \frac{1}{n} \sum_{j=0}^{n-1} dY_j.$$

Se retirar-se o efeito da translação, fazendo:

$$dX'_j = dX_j - dX_c, \quad dY'_j = dY_j - dY_c,$$

com:

$$dX'_j = (dX'_j, dY'_j)^T \text{ e } X' = X - X_c,$$

então o problema reduz-se a encontrar a rotação $d\theta$ e o factor de escala $(1+ds)$ que melhor ajusta X' ao conjunto de pontos dados por $(X' + dX')$.

Considere-se o ponto j , o objectivo pretendido é movê-lo para o ponto j' , Figura A.1.

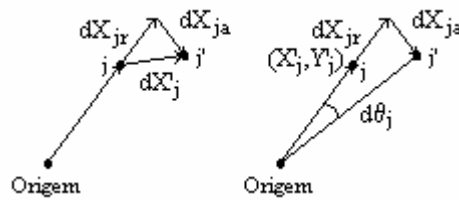


Figura A.1 Cálculo dos valores do ângulo e escala necessários para mover o ponto de uma posição para outra.

É possível mostrar que [Cootes and Taylor, 1992a]:

$$dX_{jr} = \frac{X'_j dX'_j + Y'_j dY'_j}{(X_j'^2 + Y_j'^2)} \begin{pmatrix} X'_j \\ Y'_j \end{pmatrix},$$

$$dX_{ja} = dX'_j - dX_{jr},$$

$$ds_j = \frac{|dX_{jr}|}{\sqrt{X_j'^2 + Y_j'^2}} = \frac{X_j' dX_j' + Y_j' dY_j'}{(X_j'^2 + Y_j'^2)},$$

$$d\theta_j = \frac{|dX_{ja}|}{\sqrt{X_j'^2 + Y_j'^2}}.$$

Então, a rotação e o factor de escalamento são:

$$d\theta = \frac{1}{n} \sum_{j=0}^{n-1} d\theta_j$$

$$ds = \frac{1}{n} \sum_{j=0}^{n-1} ds_j.$$