



UNIVERSIDADE DA BEIRA INTERIOR
Covilhã | Portugal

Segmentação de Faces em Imagens no Infravermelho Térmico

Sílvio Brás Filipe

Submetida à Universidade da Beira Interior para a candidatura ao grau de
Mestre em Engenharia Informática no ramo da
Computação e Sistemas Inteligentes

Orientado pelo Prof. Doutor Luís Filipe Barbosa de Almeida Alexandre

Departamento de Informática
Universidade da Beira Interior
Covilhã, Portugal
<http://www.di.ubi.pt>

Agradecimentos

Esta é uma parte importante desta dissertação, uma vez que tenho oportunidade para agradecer às pessoas que de uma forma ou outra foram imprescindíveis para a execução do mesmo.

Em primeiro lugar, gostaria de agradecer ao meu orientador, o Professor Doutor Luís Filipe Barbosa de Almeida Alexandre, por toda a informação que me forneceu, pela forma clara e precisa com que me transmitiu a sua sabedoria e conhecimento científico, pela sua disponibilidade de tempo e acima de tudo pelos seus ensinamentos e conselhos, que penso que foram cruciais para a realização desta dissertação.

Quero também agradecer à minha família, em especial aos meus pais e à minha irmã, pois sempre me apoiaram e ajudaram ao longo desta etapa da minha vida.

Quero agradecer a todos os colaboradores (professores e alunos) do grupo *Soft Computing and Image Analysis Laboratory* (SOCIA Lab), pelo óptimo espírito de equipa e pela grande inter ajuda que existia entre nós dentro do laboratório.

Obrigado!

Resumo

O objectivo desta dissertação é desenvolver um método para a segmentação automática de faces em imagens capturadas no infravermelho térmico, permitindo uma ampla gama de rotações da face e expressões. A motivação por trás desse esforço é de possibilitar um melhor desempenho dos métodos de reconhecimento de faces em imagens no infravermelho térmico. Ao longo desta dissertação são discutidos oito abordagens diferentes e a comparação dos seus desempenhos com outros três métodos publicados anteriormente. As abordagens propostas são baseadas em modelos estatísticos das intensidades dos pixels e a aplicação de contornos activos, contudo outras operações de processamento de imagem são realizadas. Estudamos também o desempenho de três abordagens de fusão sob diferentes regras (votação por maioria, operador lógico *AND* e *OR*). As experiências foram realizadas num total de 893 imagens de teste de 4 bases de dados públicas disponíveis. Os resultados obtidos melhoram os resultados dos métodos existentes até 31.2% para a primeira medida de erro (E_1) e até 39.0% para a segunda medida (E_2), dependendo do método e da base de dados. Quanto ao tempo computacional, as nossas propostas podem melhorar até 75.4% quando comparadas com as outras propostas.

Abstract

The aim of this thesis is to present a method for the automatic segmentation of face images captured in the thermal infrared, allowing for a large range of face rotations and expressions. The motivation behind this effort is to enable better performance of face recognition methods in the thermal infrared images. The thesis discusses 8 different proposed approaches and compares their performance against three other previously published methods. The proposed approaches are based on statistical modelling of pixel intensities and active contour application, although several other image processing operations are also performed. We also study the performance of the fusion of 3 of the approaches under different rules (majority vote, OR and AND fusions). Experiments were performed on a total of 893 test images from 4 public available databases. The obtained results improve on previous existing methods up to 31.2% for the first measure error (E_1) and up to 39.0% for the second measure (E_2), depending on the method and database. Regarding the computational time, our proposals improve up to 75.4% when compared with the other proposals.

Palavras Chave

Segmentação da Face, Imagens Térmicas, Segmentação da Pele, Segmentação de Imagens.

Acrónimos

ADN *Ácido Desoxirribonucleico*

AG *Algoritmos Genéticos*

BERC *Biometrics Engineering Research Center*

CIARP *15th Iberoamerican Congress on Pattern Recognition*

DCT *Discrete Cosine Transform*

DWT *Discrete Wavelet Transform*

EM *Expectation Maximization*

EMD *Empirical Mode Decomposition*

FNR *False Negative Rate*

FFT *Fast Fourier Transform*

FPR *False Positive Rate*

FSU *Florida State University*

ICA *Independent Component Analysis*

IMF *Intrinsic Mode Function*

IPFIB *Identificação dos Pixéis da Face numa Imagem Binária*

IR *Infrared*

IRIS *Imaging, Robotics, and Intelligent Systems laboratory (University of Tennessee)*

LBP *Local Binary Pattern*

LDA *Linear Discriminant Analysis*

LWIR *Long-Wave Infrared*

MACE *Minimum Average Correlation Energy*

MHD *Modified Hausdorff Distance*

MLP *Multilayer Perceptron*

MWIR *Mid-Wave Infrared*

NIR *Near Infrared*

NN *Nearest Neighbour*

OpenCV *Open Computer Vision Library*

OTCBVS *Object Tracking and Classification in and Beyond the Visible Spectrum*

OTSDF *Optimum Trade-off Synthetic Discriminant Function*

PCA *Principal Component Analysis*

R *Red*

RBF *Radial-Basis Function*

RIE *Região de Interesse Elíptica*

RIR *Região de Interesse Rectangular*

SOCIA Lab *Soft Computing and Image Analysis Laboratory*

SVD *Singular Value Decomposition*

SVM *Support Vector Machine*

SWIR *Short-Wave Infrared*

TMP *Thermal Minutia Point*

TPAMI *Transactions on Pattern Analysis and Machine Intelligence*

UND *University of Notre Dame*

Conteúdo

Agradecimentos	iii
Resumo	v
Abstract	vii
Palavras Chave	ix
Acrónimos	xi
Conteúdo	xiii
Lista de Figuras	xvii
Lista de Tabelas	xxi
Lista de Algoritmos	xxiii
1 Introdução	1
1.1 Motivação e Objectivos	5
1.2 Contribuição	6
1.3 Organização	7
2 Estado da Arte	9
2.1 Biometria	9
2.1.1 ADN	9
2.1.2 Orelha	9

2.1.3	Face	10
2.1.4	Face Térmica	10
2.1.5	Geometria do Dedo	10
2.1.6	Impressão Digital	10
2.1.7	Maneira de Andar (<i>Gait</i>)	10
2.1.8	Geometria da Mão	11
2.1.9	Veias da Mão (<i>Palm Vein</i>)	11
2.1.10	Íris	11
2.1.11	Maneira de Escrever usando um Teclado (<i>Keystroke</i>)	11
2.1.12	Impressão da Mão (<i>Palm Print</i>)	12
2.1.13	Retina	12
2.1.14	Assinatura	12
2.1.15	Voz	12
2.2	Reconhecimento Facial em Imagens LWIR	13
2.2.1	Técnicas Holísticas	13
2.2.2	Técnicas Baseadas em Características	16
2.2.3	Factores que Afectam o Reconhecimento de Faces em Imagens no Infravermelho Térmico	18
2.2.4	Fusões	19
2.2.4.1	Fusões Baseadas no Resultado	20
2.2.4.2	Fusões dos Dados	22
2.2.4.3	Reconhecimento Híbrido e Fusões dos Dados	24
2.3	Segmentação Facial em Imagens no Infravermelho Térmico	26
2.4	Conclusão	31
3	Bases de Dados	33
3.1	<i>University of Notre Dame (UND)</i>	33
3.2	<i>Terravic Facial IR</i>	34
3.3	<i>IRIS Thermal/Visible Face</i>	34
3.4	<i>Florida State University (FSU)</i>	35

3.5	Outras	36
3.6	Conclusão	37
4	Métodos de Segmentação Propostos	39
4.1	Operações	39
4.1.1	Região de Interesse Rectangular (RIR)	39
4.1.2	Região de Interesse Elíptica (RIE)	42
4.1.3	Limiarização (<i>Thresholding</i>)	43
4.1.4	Contornos Activos sem Arestas	44
4.1.5	Identificação dos Pixéis da Face numa Imagem Binária (IPFIB)	46
4.2	Métodos de Segmentação	48
4.3	Fusões	48
4.4	Conclusão	50
5	Resultados e Discussão	51
5.1	Segmentação Manual	51
5.2	Medidas de Erro	52
5.3	Resultados Experimentais e Discussão dos Resultados	53
5.4	Conclusão	63
6	Conclusão	65
6.1	Resumo do Trabalho desenvolvido	65
6.2	Trabalho Futuro	66
A	Artigo submetido à CIARP	67
B	Artigo submetido à revista TPAMI	77
	Referências	91

Lista de Figuras

1.1	Representação dos vários espectros existentes, com um destaque para os comprimentos de onda entre o visível e o infravermelho [2].	4
1.2	Exemplo de uma imagem da face capturada em vários espectros [4]. . .	4
1.3	Diagrama de blocos de um sistema de reconhecimento.	7
2.1	Exemplos de <i>eigenfaces</i> extraídos no visível e no infravermelho térmico [84].	14
2.2	A primeira linha contém imagens no visível. A segunda linha contém as correspondentes imagens no espectro NIR. Estas imagens foram retiradas de [59].	15
2.3	Exemplos das componentes independentes extraídos no visível e no infravermelho térmico [76].	16
2.4	Exemplo da extração dos TMP's de uma imagem LWIR.	17
2.5	Exemplo da detecção e substituição dos óculos por um <i>template</i> [51]. . .	20
2.6	Reconhecimento facial multi-espectral [22].	21
2.7	Visão global sobre reconhecimento facial multi-modal [54].	22
2.8	Fusão das características em imagens multi-espectrais da face [81]. . .	24
2.9	Sistema de reconhecimento facial proposto em [49].	25
2.10	O sistema proposto em [16] é composto por três módulos principais: pré-processamento de dados e registo, detecção dos óculos e a fusão das representações face global e local, utilizando as imagens no visível e térmico.	26
2.11	Máscara binária usada em [76].	27

2.12	Exemplo do corte das regiões da face (regiões assinaladas com arestas pretas) e fundo (regiões assinaladas com arestas brancas).	27
2.13	Distribuições das intensidades dos pixéis da face (linha azul) e fundo (linha vermelha) para as quatro bases de dados usadas.	30
2.14	Aplicação do método de segmentação apresentado em [35].	30
3.1	Exemplos de imagens da base de dados UND.	33
3.2	Exemplos de imagens da base de dados Terravic.	34
3.3	Representação do sistema de captura da base de dados IRIS [5].	35
3.4	Exemplo do pré-processamento aplicado às imagens da base de dados IRIS.	36
3.5	Exemplo de imagens da base de dados FSU.	36
3.6	Câmaras e iluminações usadas na captura simultânea das imagens em vários espectros (visível, SWIR, MWIR e LWIR) [89].	37
4.1	Análise da assinatura vertical.	41
4.2	Análise da assinatura horizontal.	42
4.3	Casos possíveis no ajuste da contorno [30].	45
4.4	Exemplo da aplicação da operação IPFIB.	47
4.5	Diagrama de blocos dos métodos de segmentação baseados nas operações apresentadas anteriormente.	49
5.1	Representações gráficas da medida de erro E_1 para os resultados apresentados nas tabelas 5.2, 5.3, 5.4 e 5.5.	53
5.2	Representações gráficas da medida de erro E_2 para os resultados apresentados nas tabelas 5.2, 5.3, 5.4 e 5.5.	54
5.3	Representação gráfica dos tempos de execução apresentados nas tabelas 5.2, 5.3, 5.4 e 5.5.	59
5.4	Resultados da segmentação para os métodos apresentados em [35, 69, 24, 25, 22, 26] e pelos nossos (A , B , ..., H), para as imagens das bases de dados UND (quatro primeiras colunas) e Terravic (quatro últimas colunas).	60

5.5	Resultados da segmentação para os métodos apresentados em [35, 69, 24, 25, 22, 26] e pelos nossos (A, B, . . . , H), para as imagens das bases de dados IRIS (quatro primeiras colunas) e FSU (quatro últimas colunas).	61
5.6	Resultados da segmentação para as fusões apresentadas, para as imagens das bases de dados UND (quatro primeiras colunas) e Terravic (quatro últimas colunas).	62
5.7	Resultados da segmentação para as fusões apresentadas, para as imagens das bases de dados IRIS (quatro primeiras colunas) e FSU (quatro últimas colunas).	62

Lista de Tabelas

5.1	percentagens de pixéis da face fundo em relação ao número total de pixéis das imagens, para as diferentes bases de dados.	52
5.2	Experiment results in the UND database.	55
5.3	Experiment results in the Terravic database.	56
5.4	Experiment results in the IRIS database.	57
5.5	Experiment results in the FSU database.	58

List of Algorithms

1	Estimação dos parâmetros das duas distribuições normais (ω_{s_i} , μ_{s_i} e $\sigma_{s_i}^2$) para os N_s pixéis da região da pele usando o algoritmo EM.	28
2	Algoritmo para a obtenção dos pontos da elipse.	43

Capítulo 1

Introdução

Como parte de uma sociedade da informação, a segurança tornou-se hoje mais do que nunca um activo de grande importância em quase todas as áreas. No processo para alcançar a segurança, a capacidade de identificar com precisão indivíduos com base nas suas características biométricas, tem recebido uma maior preocupação e entusiasmo pela comunidade científica. Os investigadores tentam desenvolver métodos capazes de superar as novas necessidades e os novos desafios. No reconhecimento biométrico, na perspectiva do reconhecimento de padrões, os atributos físicos são capturados, processados e classificados para providenciarem as suas identidades. A utilização de sistemas de reconhecimento, como um substituto para os sistemas tradicionais de segurança, tornou-se uma área de grande interesse tanto pelas instituições públicas como privadas.

A procura e as actividades de pesquisa em reconhecimento de faces humanas em imagens ou em vídeo têm aumentado significativamente nos últimos 30 anos. O reconhecimento de faces tem sido uma área onde muitas pessoas têm focado a sua investigação, pois os meios tradicionais de segurança como cartões de identificação e senhas não são seguros o suficiente ou convenientes [17, 64]. O reconhecimento facial é essencial para a interpretação de expressões faciais, emoções humanas, intenções e comportamentos, o que é um componente chave para os ambientes inteligentes [70]. O reconhecimento facial não é intrusivo e é talvez a forma mais natural de identificação, visto ser a forma como todos os humanos se reconhecem uns aos outros. A face não é a única característica biométrica. Existem outras características fisiológicas que podem ser usadas usadas, tais como impressões digitais, padrões de retina e íris, geometria da mão, e voz. Tais sistemas de identificação biométrica têm de contar com a colaboração

dos participantes. A autenticação usando o reconhecimento facial é intuitiva e não necessita que o sujeito pare as suas actividades. A análise de imagens frontais da face ou de perfil são geralmente eficazes sem a cooperação dos participantes. As áreas onde podem ser aplicadas as tecnologias de reconhecimento facial são amplas, como a correspondência de fotografias do passaporte ou carta de condução, controle de acesso para proteger redes de computadores ou equipamentos, tais como prédios do governo ou tribunais, a autenticação no sector bancário ou operações financeiras, o rastreio automático nos aeroportos de terroristas conhecidos e uso na vídeo vigilância [29]. Essas aplicações vão desde uma correspondência estática de fotografias com formatos controlados a aplicações que analisam, em tempo real, sequências de imagens de um vídeo. Na área de segurança por computador, um sistema de reconhecimento facial pode ser usado continuamente para confirmar a identidade do utilizador do sistema, e para confirmar o nível de autorização que ele possui para a realização de determinadas acções.

Enquanto os seres humanos conseguem de forma rápida e fácil reconhecer faces em situações variáveis ou mesmo após vários anos de separação, o reconhecimento facial automático por parte de uma máquina ainda é uma tarefa desafiadora [36]. A face é inerentemente um objecto 3D, iluminado por uma variedade de fontes de iluminação a partir de diferentes direcções e envolvido por objectos de fundo arbitrário. Portanto, a aparência do rosto varia enormemente quando projectadas numa imagem em 2D. A variação angular da pose pode também causar alterações significativas em 2D. Um sistema de reconhecimento facial robusto tem de ter a capacidade de reconhecer a identidade de uma pessoa apesar destas variações na aparência que a face possa ter. Simultaneamente, o sistema deve ser robusto perante os problemas típicos de aquisição de imagem, como ruído, distorção da câmara de vídeo e a resolução de imagem.

Apesar do sucesso das técnicas de reconhecimento automático de faces em muitas aplicações práticas, o reconhecimento, com base apenas no espectro visível, tendem a deixar de ser consistentes em ambientes não controlados. O desempenho do reconhecimento facial no visível é sensível às variações das condições de iluminação [11]. As variações entre imagens da mesma face, devido a mudanças na iluminação, são maiores que as variações na identidade da face. Outros factores, tais como expressões faciais [92] e variações na pose [19] complicam ainda mais a tarefa de reconhecimento de face. As técnicas de reconhecimento de faces no visível têm dificuldade em identificar os indivíduos usando disfarces e maquilhagens. As aplicações de alta segurança têm

um particular interesse na detecção de disfarces faciais. Os disfarces simples, como um nariz falso ou barba podem modificar substancialmente o aspecto visual de uma pessoa. Obviamente, a identificação visual de gémeos idênticos, ou em que a aparência foi alterada com a cirurgia plástica é quase impossível.

O reconhecimento de faces usando diferentes tipos de imagens, em particular sensores de *Infrared* (IR), tornou-se uma área de interesse crescente [88, 99]. Os espectros electromagnéticos abaixo do espectro visível, como *X-rays* e radiação ultravioleta, são prejudiciais ao corpo humano e não podem ser utilizados para aplicações de reconhecimento de faces. As imagens IR térmicas têm sido sugeridas como uma forte alternativa para a detecção e reconhecimento de faces. Enquanto que as câmaras no visível medem a energia electromagnética na faixa do espectro visível ($0.4 - 0.7\mu\text{m}$), os sensores das câmaras IR térmico medem as radiações térmicas na faixa do espectro infravermelho ($0.7 - 14.0\mu\text{m}$). O espectro infravermelho é composto pelo IR reflectido e o IR térmico. O espectro IR reflectido ($0.7 - 2.4\mu\text{m}$) está associado à radiação solar reflectida e não contém informações sobre as propriedades dos materiais. O *Near Infrared* (NIR) ($0.7-0.9\mu\text{m}$) e o *Short-Wave Infrared* (SWIR) ($0.9-2.4\mu\text{m}$) são espectros de reflexão e as diferenças na aparência entre o visível e o infravermelho reflectido são devido às propriedades dos materiais reflectores. Esta radiação é em grande parte invisível ao olho humano. O IR térmico está associada à radiação térmica emitida pelos objectos. A quantidade de radiação emitida depende tanto da temperatura como da emissividade do material. O espectro infravermelho térmico é dividido em dois grupos principais: o *Mid-Wave Infrared* (MWIR) da faixa espectral $3.0 - 5.0\mu\text{m}$ e o *Long-Wave Infrared* (LWIR) ($8.0-14.0\mu\text{m}$). O rosto humano e o corpo emitem radiações térmicas em ambas as faixas do espectro infravermelho térmico. As câmaras IR térmicas podem medir variações de temperatura na face à distância, e produzem termogramas sob a forma de imagens 2D. O reconhecimento facial no IR térmico é favorecido no LWIR devido a existir uma maior emissão nesta faixa do que no MWIR. A figura 1.1 apresenta os vários espectros existentes, com especial foco para o espectro visível e infravermelho. Na figura 1.2 podemos observar as variações existentes numa imagem da face nos vários espectros.

Apesar de sacrificar as cores no reconhecimento, as técnicas de reconhecimento facial em imagens IR térmicas podem ser usadas para identificar faces quando há pouco ou nenhum controlo sobre as condições de iluminação. Uma vantagem de se usar imagens IR térmicas em vez de imagens no visível resulta do facto de que a luz

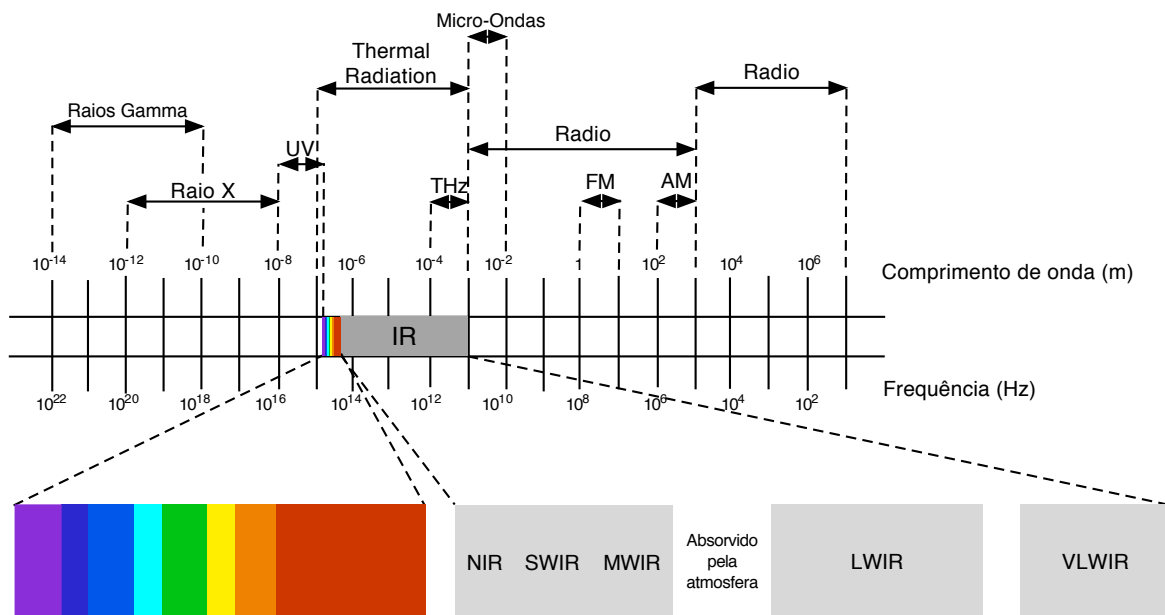


Figura 1.1: Representação dos vários espectros existentes, com um destaque para os comprimentos de onda entre o visível e o infravermelho [2].

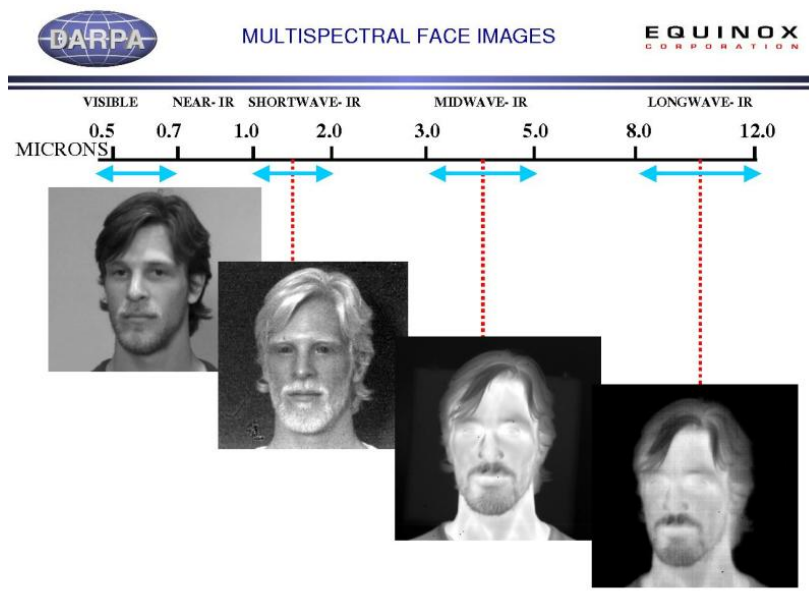


Figura 1.2: Exemplo de uma imagem da face capturada em vários espectros [4].

na faixa do IR térmico é emitida em vez de reflectida [96].

As emissões térmicas da pele são propriedades intrínsecas, independentes da iluminação. Portanto, as imagens da face capturadas usando sensores de infravermelho térmico são invariantes às mudanças da iluminação ambiente. A energia transmitida no

IR pode ser vista em qualquer condição de luminosidade e é menos sujeita a dispersão e absorção pelo fumo ou luz visível. A variabilidade dentro de cada classe também é significativamente menor que a observada em imagens visíveis. No espectro infravermelho foram encontradas vantagens sobre o espectro visível na detecção de faces [38, 39], detecção de faces disfarçadas [68], e reconhecimento facial em condições de fraca iluminação. As imagens IR térmicas têm sido uma ferramenta eficaz no diagnóstico e no estudo da distribuição da temperatura da pele na detecção do cancro de mama [42], reconhecimento de fonte de calor para a inspecção de peças electrónicas [52, 60], detecção de alvos [78] e aplicações militares. A detecção de suspeitos envolvidos em actos ilícitos e actividades potencialmente prejudiciais usando imagens térmicas são propostas em [67]. No entanto os sintomas como o estado de alerta ou ansiedade podem esconder características biométricas, devido à redistribuição do fluxo sanguíneo nos vasos sanguíneos, provocando alterações bruscas na temperatura da pele local.

1.1 Motivação e Objectivos

A informação capturada no visível por uma câmara representa a reflectância de uma face, enquanto a informação medida por um sensor IR térmico é anatómica. Apesar da iluminação e expressão facial se alterar significativamente, a aparência visual da face, as características térmicas desta permanecem quase invariáveis. Nas imagens do infravermelho térmico, o ruído de fundo não é visível e as tarefas de detecção de face, localização e segmentação providenciam uma maior confiança do que em imagens no visível.

Inicialmente, o objectivo deste trabalho era o desenvolvimento de um sistema de reconhecimento facial em imagens LWIR, mas no decorrer deste apenas ficámos pela segmentação. Esta escolha foi feita depois de termos realizado um estudo sobre os métodos que existiam na segmentação e no reconhecimento em imagens LWIR. Neste estudo notámos que existia uma grande lacuna em termos da segmentação, o que não é tão visível no que toca ao reconhecimento (como iremos ver no capítulo 2).

No domínio dos sistemas biométricos, um passo crucial no processo de reconhecimento (ver figura 1.3) é a segmentação. Esta é mais exigente do que uma simples detecção, uma vez que aponta não só para a sua localização, mas também faz a descrição da sua forma. Um sistema robusto de segmentação pode melhorar as taxas

de reconhecimento, independentemente do método de reconhecimento.

A figura 1.3 mostra o esquema geral de um sistema de reconhecimento. Este esquema pode ser utilizado para reconhecimento de faces nos comprimentos de onda visível ou térmico e pode ser usado também para outras modalidades de reconhecimento, tais como as que utilizam imagens de íris [72]. O sistema de reconhecimento é composto por duas partes:

- **Processo de Treino:** Conjunto de treino capturado por uma câmara. É aplicado um algoritmo de segmentação antes de serem extraídas as características que possibilitam o reconhecimento. Estas são armazenadas numa base de dados.
- **Processo de Teste:** Dada uma imagem, esta é segmentada e são extraídas características como no processo de treino. Estas características são comparadas com as armazenadas na base de dados e é produzido o resultado da comparação.

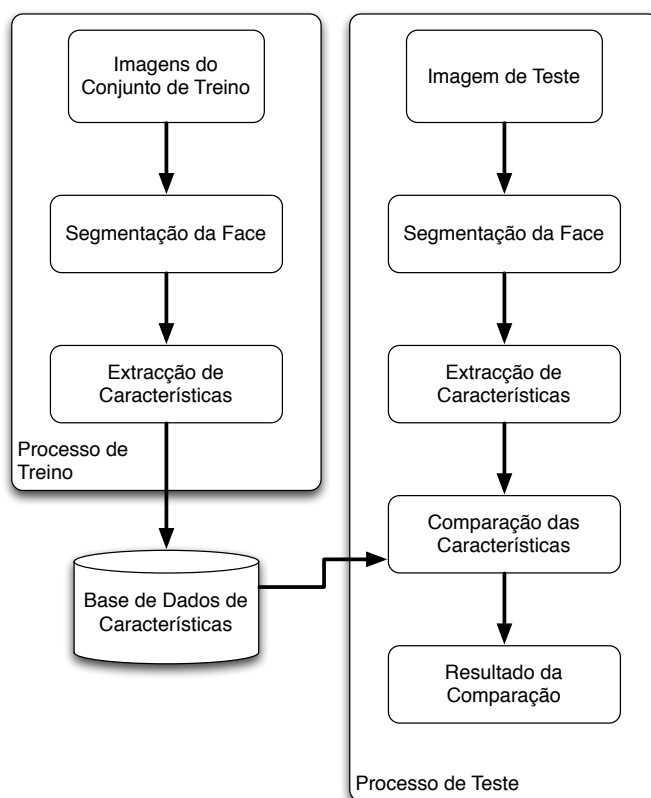


Figura 1.3: Diagrama de blocos de um sistema de reconhecimento.

1.2 Contribuição

Este trabalho vem introduzir uma melhoria significativa na segmentação de faces em imagens IR térmicas. A decisão pela realização deste trabalho teve por base a inexistência de uma sólida bibliografia sobre a segmentação da face em imagens capturadas no IR.

Em contraste com o espectro visível, onde existem numerosos métodos propostos para a realização desta tarefa (baseado na cor [83], geometria [27], etc), em LWIR existe uma grande lacuna de propostas.

Esta dissertação enquadra-se no âmbito do projecto **BioRec - Non-Cooperative Multimodal Biometric Recognition** [3] (PTDC/EIA/69106/2006), financiado pela FCT e FEDER. O projecto **BioRec** tem como principal objectivo a realização de reconhecimento em tempo real através de diversas características biométricas sem a cooperação do sujeito que está a ser avaliado.

1.3 Organização

Este trabalho encontra-se organizado da forma que passamos a descrever. No **Capítulo 1 (Introdução)** foi feita uma introdução ao tema, frisando vários aspectos relacionados. No **Capítulo 2 (Estado da Arte)** são apresentados alguns métodos que existente para o reconhecimento facial usando imagens no infravermelho térmico, visto que este foi o tema inicial desta dissertação. Para colmatar as lacunas existentes nestes métodos, tivemos de mudar o tema que iria ser proposto para esta dissertação, a segmentação de faces em imagens térmicas. Assim sendo, apresentamos também os poucos métodos existentes para a segmentação de faces em imagens capturadas no infravermelho térmico. No **Capítulo 3 (Bases de Dados)** são apresentadas as bases de dados de imagens da face capturadas no infravermelho térmico. No **Capítulo 4 (Métodos de Segmentação Propostos)** apresentamos os métodos desenvolvidos para a segmentação das faces, bem como as operações necessárias para estes métodos. No **Capítulo 5 (Resultados e Discussão)** são apresentados os resultados obtidos usando os vários métodos de segmentação. No último capítulo, apresentamos algumas conclusões/argumentos sobre a realização desta dissertação bem como descrevemos um pouco o que poderá ser acrescentado a este trabalho na tentativa da obtenção de ainda melhores resultados ou validação dos mesmos.

Capítulo 2

Estado da Arte

2.1 Biometria

A biometria refere-se a tecnologias que medem e analisam as características do corpo humano para fins de autenticação [73]. Existe uma grande variedade de características biométricas, sendo as mais comuns usadas para fins biométricos, como apresentamos agora.

2.1.1 ADN

O Ácido Desoxirribonucleico (ADN) é representado através de um código unidimensional e é único para cada pessoa. Embora a maioria do ADN não difira de pessoa para pessoa, 0.1% do genoma inteiro de uma pessoa é único para cada indivíduo, o que representa três milhões de pares bases do ADN. O ADN é altamente propenso a contaminação e não permite o reconhecimento em tempo real.

2.1.2 Orelha

As orelhas têm sido utilizadas para identificação dos sujeitos, especialmente em cenas de crime, onde as impressões digitais estão ausentes. No entanto, a sua singularidade é questionada uma vez que pode ser facilmente coberta, sendo necessário a cooperação do indivíduo que esteja a ser identificado.

2.1.3 Face

A face é a característica biométrica mais comum, visto que os seres humanos a usam para realizar o reconhecimento pessoal. O número de sistemas de reconhecimento implantados comercialmente usando esta característica biométrica é relativamente elevado, mas o desempenho destes sistemas é apenas razoável. Isto deve-se ao facto de que estes sistemas têm dificuldades em lidar com ambientes não controlados.

2.1.4 Face Térmica

Os padrões de calor do corpo humano são considerados únicos para cada pessoa. A face ao ser capturada através de câmaras de infravermelho produz assinaturas faciais únicas, construídas pelo calor que passa através do tecido facial. As assinaturas faciais são denominadas por "termogramas faciais". Este método de reconhecimento não é invasivo, mas pode ser altamente comprometido quando existem por perto superfícies quentes.

2.1.5 Geometria do Dedo

Consiste na avaliação do comprimento do dedo, largura, espessura e curvatura. Apesar de ter uma técnica de aquisição relativamente simples e não invasiva, exige que o sujeito alinhe os dedos contra vários pinos.

2.1.6 Impressão Digital

Usada há vários séculos, através da impressão dos padrões usando tinta, forma como os analfabetos ainda hoje assinam, ou um dispositivo de leitura. O acesso aos detalhes de cristas e sulcos da impressão digital, aliada a alguns *minutiae points*, pode determinar a sua singularidade. Apesar de ser uma tecnologia madura e fácil de usar, também exige a cooperação do indivíduo e é vulnerável ao ruído.

2.1.7 Maneira de Andar (*Gait*)

A forma e dinâmica de caminhar pode ser uma informação potencial para a identificação. A análise das características das séries temporais, comprimentos do passo,

cadência e velocidade ou a silhueta podem permitir a diferenciação do sujeito. No entanto, estas características são vulneráveis às mudanças do piso, velocidade de execução ou condições.

2.1.8 Geometria da Mão

Desde do início dos anos 1980, a geometria da mão é usada para aplicações de baixa segurança. A geometria da mão tem como características a sua forma, o comprimento e a largura dos dedos. Este tipo de sistemas são fáceis de usar e baratos, mas requerem uma superfície de contacto relativamente grande, o que o torna inadequado para fins que possuam um espaço reduzido.

2.1.9 Veias da Mão (*Palm Vein*)

Os padrões das veias do sangue, sendo únicos para cada indivíduo, podem ser capturados com iluminação NIR e constituem um traço biométrico relativamente seguro. A captura das veias da mão é feita no NIR visto que a hemoglobina absorve os raios infravermelhos, levando a que as veias apareçam a preto. Embora este procedimento dificilmente possa ser forjado, ele requer também um nível relativamente elevado de cooperação.

2.1.10 Íris

Os padrões complexos podem conter muitas características distintas que discriminem o sujeito. É relativamente difícil de forjar, visto que até uma íris artificial pode ser detectada, e sua imagem ideal ocorre no NIR através de dispositivos de captura relativamente caros e que exigem cooperação por parte do indivíduo. No entanto, o grupo onde este trabalho foi feito, o SOCIA Lab, tem feito inúmeros progressos no reconhecimento não cooperativo usando imagens da íris [71, 75].

2.1.11 Maneira de Escrever usando um Teclado (*Keystroke*)

Com base na crença de que cada pessoa digita num teclado de uma maneira própria, a análise desses ritmos podem levar à identificação do sujeito. É uma característica

biométrica que necessita uma monitorização contínua para que seja possível analisar os ritmos com que o sujeito tecla, o que reduz a eficácia das falsas tentativas.

2.1.12 Impressão da Mão (*Palm Print*)

Como nas impressões digitais, as palmas da mão contêm padrões únicos. Uma vez que estas são maiores do que os dedos, a distinção entre indivíduos deverá ser mais acentuada, mas também exigem uma superfície maior de aquisição, tornando os dispositivos de captura mais caros.

2.1.13 Retina

A varredura da retina capta os padrões das veias na parte interna do olho. Estes padrões são estáveis, únicos e difíceis de falsificar e, portanto considerados como uma das características biométricas mais seguras. No entanto, a retina é pequena, interna e exige uma cooperação muito grande do indivíduo.

2.1.14 Assinatura

A forma como uma pessoa assina seu nome é uma característica individual a qual permanece durante a vida do indivíduo. A análise clássica deste tratamento é baseado em aparência visual da assinatura, mas a velocidade, a direção e a pressão exercida na escrita também podem ser analisadas. Apesar de possuir uma boa aceitação esta característica pode depender da condição física e psicológica do indivíduo.

2.1.15 Voz

Nesta característica biométrica é utilizada a informação acústica, em vez da visual. O reconhecimento é realizado através da análise das características da voz do indivíduo, e é usado apenas em aplicações de baixa segurança devido à sua variabilidade (devido a constipação, etc.) e baixa performance (devido ao ruído que possa existir em redor).

2.2 Reconhecimento Facial em Imagens no Infravermelho Térmico

O reconhecimento facial é uma área da visão computacional que tem atraído um grande interesse por parte da comunidade científica. A investigação no âmbito do reconhecimento facial térmico é uma área em crescimento. Muitas das técnicas utilizadas no reconhecimento de faces térmicas são inspiradas nas técnicas de reconhecimento no visível (como *Eigenfaces* ou *Principal Component Analysis* (PCA) [93], *Fisherfaces* ou *Linear Discriminant Analysis* (LDA) [40], *Independent Component Analysis* (ICA) [62], *Support Vector Machine* (SVM) [53], ARENA [80], modelo Bayesiano [61], *Boosting algorithms* [63], software comercial de reconhecimento facial (*Facelt*®) [48], etc). Estas técnicas podem ser divididas em abordagens holísticas ou baseadas em características.

Vários resumos do estados de arte sobre as diferentes técnicas de reconhecimento facial no infravermelho podem ser encontrados em [74, 57, 12].

2.2.1 Técnicas Holísticas

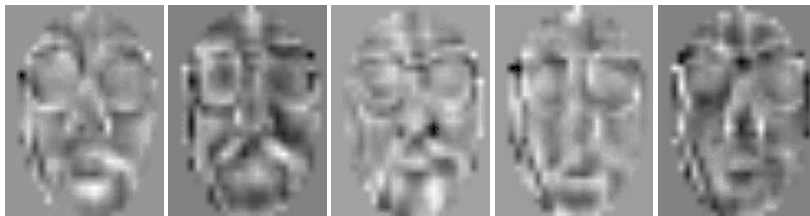
As técnicas mais populares de reconhecimento facial em imagens no infravermelho são técnicas holísticas. A mais utilizada são de longe as *Eigenfaces* ou (PCA) [93]. Essa técnica é usada em muitos casos como uma referência para comparação de desempenho com outras técnicas. Outras duas técnicas populares baseadas em holísticas são a LDA e a ICA.

A PCA permite uma redução na dimensionalidade por uma representação do sub-espaço das faces. A imagem média da face é construída a partir do conjunto de treino e utilizada para obter os vectores próprios que representam uma aproximação ao espaço da face. Partindo do pressuposto de que o conjunto de treino é uma boa representação das possíveis imagens da face, os vectores seleccionados são uma boa aproximação de todas as faces possíveis [84, 85, 88, 76]. Ao seleccionar os vectores próprios correspondentes, os valores próprios mais altos constroem um espaço de pequena dimensão com a projecção de imagens da face. Um dos primeiros trabalhos de reconhecimento facial no infravermelho utiliza a PCA [37]. Os testes foram realizados utilizando imagens de MWIR frontais e com rotação de $\pm 45^\circ$. Desde então, muitos outros autores utilizaram a PCA. *Wilder et al.* [95] também utilizou a PCA, mas este em imagens LWIR.

Socolinsky et al. e *Selinger et al.* [84, 85, 88, 76] realizaram uma comparação entre as imagens no visível e no LWIR utilizando diferentes técnicas, incluindo a PCA (ver figura 2.1). Os testes foram realizados em diversas situações, incluindo faces com e sem óculos, com várias expressões faciais e as várias condições de iluminação. Os autores relataram bons desempenhos da PCA em imagens LWIR.



(a) Exemplos de *eigenfaces* usando imagens no visível.



(b) Exemplos de *eigenfaces* usando imagens no LWIR.

Figura 2.1: Exemplos de *eigenfaces* extraídos no visível e no infravermelho térmico [84].

Friedrich et al. [44] utilizaram as distâncias euclidianas entre as *eigenfaces* das imagens LWIR. Os resultados obtidos mostram que o reconhecimento facial no espectro infravermelho dá melhores resultados, mesmo na presença de alterações na face, como a orientação da cabeça e expressões faciais.

Wu et al. [98, 97] propõem uma técnica para converter imagens em infravermelho LWIR em *blood perfusion data*. As imagens obtidas são destinadas a representar a temperatura interna da pele e são independentes das variações de temperatura ambiente. O reconhecimento facial é feito usando a PCA e redes neuronais *Radial-Basis Function* (RBF). Os autores relatam um aumento nas taxas de reconhecimento utilizando as imagens transformadas em vez da utilização das imagens térmicas originais.

Kang et al. [56] utilizaram a PCA para o reconhecimento facial com imagens SWIR. Neste sistema foi usada uma de iluminação com um comprimento de onda de $940\mu\text{m}$.

Li et al. [59] utilizaram imagens no espectro NIR para reconhecimento de faces. Estes autores usam um sistema de iluminação de $850\mu\text{m}$. Os imagens dos espectros

SWIR e NIR são menos sensíveis às mudanças de iluminação do que as imagens no visível (ver figura 2.2). Esta característica aumenta o desempenho do reconhecimento facial neste tipo de imagens.

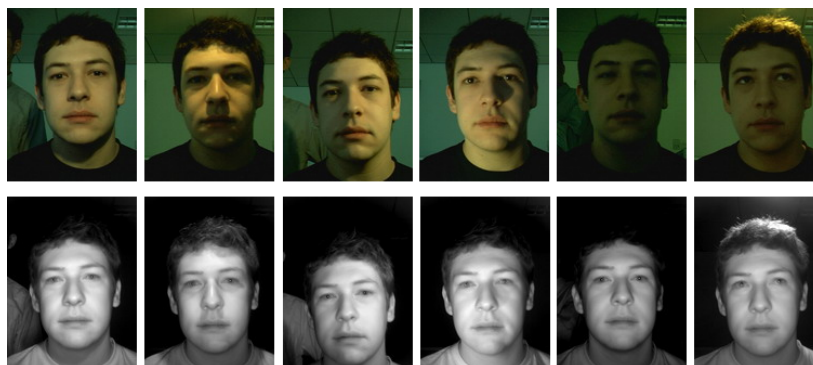


Figura 2.2: A primeira linha contém imagens no visível. A segunda linha contém as correspondentes imagens no espectro NIR. Estas imagens foram retiradas de [59].

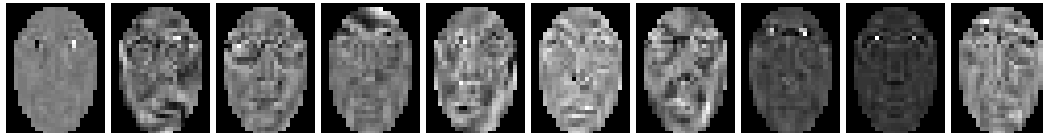
A segunda técnica mais popular de reconhecimento facial no infravermelho é a LDA, também conhecida como *Fisherfaces*. Como a PCA, esta técnica permite uma redução na dimensionalidade através da representação num sub-espço das faces. Em vez da decomposição pelos principais componentes, os dados na LDA são separados linearmente no *Eigenspace* usando uma decomposição singular (*Singular Value Decomposition* (SVD)) [84, 85, 88, 76]. Em [84, 85, 88, 76] estes realizaram comparações entre diferentes técnicas de reconhecimento facial, incluindo a LDA, em imagens no visível e em LWIR. Nos testes realizados, a LDA foi o método que obteve melhor desempenho no reconhecimento de faces no visível e LWIR.

Zou et al. [102, 101] utilizaram a LDA para o reconhecimento facial em imagens NIR. Para a classificação utilizaram quatro técnicas, disponíveis num software de aprendizagem automática denominado por *WEKA* [9], sendo elas redes neurais RBF, o classificador *Adaboost*, SVM, e *Nearest Neighbour* (NN). Os testes mostram que os resultados obtidos com a LDA em imagens NIR superaram os resultados obtidos com imagens no visível, especialmente usando SVM e NN.

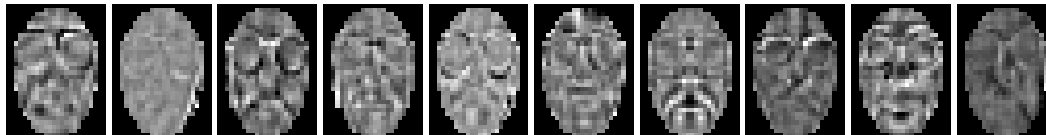
Kang et al. [56] utilizaram a LDA para o reconhecimento facial em imagens SWIR. No NIR, *Li et al.* [59] também utiliza a LDA com o *Local Binary Pattern* (LBP). Neste trabalho, a combinação da LDA com o LBP apresentou melhor desempenho.

Outra técnica holística popular é a ICA [84, 85, 88, 76]. A ICA é uma técnica estatística capaz de encontrar factores ocultos que representam dados aleatórios.

A ICA assume que as variáveis dos dados são uma combinação linear de variáveis desconhecidas e não Gaussianas, mutuamente independentes. *Socolinsky et al.* e *Selinger et al.* [84, 85, 88, 76] utilizam a ICA em imagens LWIR e no visível para o reconhecimento de faces (ver figura 2.3). Eles compararam seu desempenho à PCA e à LDA. A ICA possui um bom desempenho em imagens no visível, seguindo de perto o desempenho melhor técnica da melhor técnica apresentada, sendo ela a LDA.



(a) Exemplos das componentes independentes das imagens no visível.



(b) Exemplos das componentes independentes das imagens no LWIR.

Figura 2.3: Exemplos das componentes independentes extraídos no visível e no infravermelho térmico [76].

Outra técnica holística que pode ser usada para redução da dimensionalidade é a *Discrete Cosine Transform* (DCT). *Zhao et al.* [100] apresentaram um sistema de reconhecimento de faces no NIR usando as frequências mais baixas da DCT. A classificação foi realizada através de um SVM.

2.2.2 Técnicas Baseadas em Características

As abordagens baseadas em características extraem características das imagens da face e utilizam métricas que a partir das características verificam a sua semelhança ou a correspondência no reconhecimento facial. O modelo de correspondência (*template matching*) e cálculo de métricas de distâncias são utilizadas em diversos trabalhos.

Heo et al. [47] fizeram uma comparação entre o reconhecimento facial usando em imagens no visível e no LWIR, sob diferentes condições. O sistema de reconhecimento facial usado foi o *Facelt*® (um software comercial). Este software calcula as distâncias entre várias características extraídas da imagem da face (olhos, nariz, boca, etc.). Os resultados obtidos mostram que imagens térmicas superaram as imagens

no visível, quando não existe a presença de óculos. Mais tarde, os mesmos autores [50, 51] propõem a utilização de filtros de correlação para melhorar o desempenho do reconhecimento facial em imagens LWIR. Os filtros de correlação são conhecidos por serem invariantes à distorção, portanto a combinação de imagens térmicas, que são menos sensíveis às variações de iluminação, podem levar a um aumento do desempenho. No estudo comparativo, que os autores apresentam, a correlação é feita usando dois filtros: filtros *Minimum Average Correlation Energy* (MACE) e filtros *Optimum Trade-off Synthetic Discriminant Function* (OTSDF). Os resultados experimentais mostram que a correlação com os filtros propostos, aplicados a imagens térmicas, superaram a sua utilização em imagens no visível. Neste trabalho, o filtro OTSDF obteve melhor desempenho do que o filtro MACE. Além disso, o filtro OTSDF produziu ainda melhores resultados do que a PCA, *Facelt*® e correlação normalizada.

Pavlidis et al. e *Buddharaju et al.* [69, 24, 25, 22, 26] propuseram uma técnica de reconhecimento facial baseado na extracção de informações fisiológicas da face a partir das imagens térmicas. As informações fisiológicas extraídas representam a rede de vasos sanguíneos sob a pele (ver figura 2.4). Esta rede é única para cada indivíduo. Os autores usam operadores morfológicos para extrair a rede vascular da informação térmica. Os pontos de ramificação do "esqueleto" da rede vascular (figura 2.4(b)) são denominados por *Thermal Minutia Points* (TMPs) (figura 2.4(c)). Os TMPs constituem os dados utilizados no processo de correspondência.

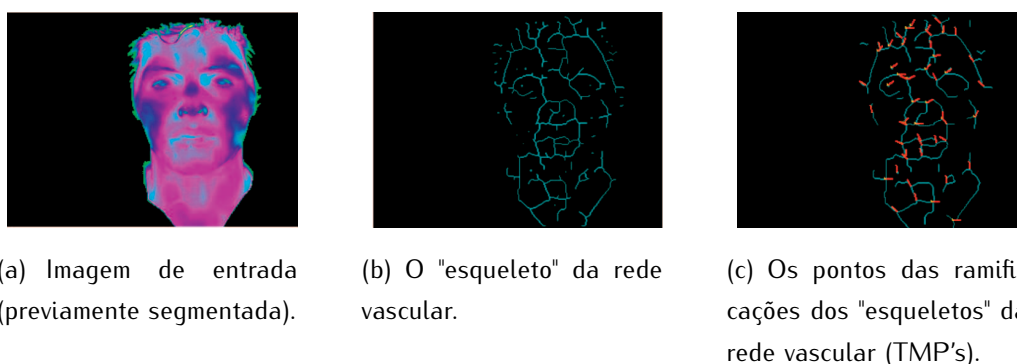


Figura 2.4: Exemplo da extracção dos TMP's de uma imagem LWIR.

Cho et al. [35] propuseram uma modificação ao método proposto em [69, 24, 25, 22, 26]. As modificações que eles propõem passam pelo uso da *Modified Hausdorff Distance* (MHD) em vez da distância Euclidiana e pela uso de uma nova característica. A nova característica é denominada, pelos autores, por *Gravity center point*, sendo

esta o ponto central da face depois de segmentada. Nas suas experiências os autores obtiveram melhores resultados do que abordagens que usam o *Naive Bayes*, RBF, *Multilayer Perceptron* (MLP), *Adaboost* e *Bagging*.

Srivastava et al. e *Buddharaju et al.* [91, 23] propuseram a decomposição de imagens no infravermelho nos seus componentes espectrais usando um conjunto de filtros passa-banda. Para esta decomposição são usados os filtros de *Gabor*. O resultado das imagens filtradas é então utilizado para obter os *Bessel forms* estimados utilizando os momentos de segunda e quarta ordens da imagem filtrada (variância e *kurtosis*). A comparação foi realizada utilizando a métrica L^2 entre os parâmetros dos *Bessels* em imagens diferentes. Os testes foram realizados em imagens LWIR. Os autores mostram que a técnica proposta supera a PCA e a ICA no reconhecimento facial.

Socolinsky et al. [84] usaram uma abordagem baseada em redes neuronais. Estes autores utilizaram um algoritmo denominado por *ARENA* [80], sendo que este primeiro reduz a dimensão dos casos de treino. Isto é feito reduzindo o tamanho da imagem para 15×20 pixels (calculando a média quadrática da sua vizinhança na imagem) e a classificação é realizada usando a NN. Os resultados experimentais mostram que o melhor desempenho foi obtido para imagens LWIR na presença da variação da iluminação, comparando com os resultados das imagens no visível.

Wilder et al. [95] realizaram uma avaliação do desempenho do reconhecimento facial em imagens no visível e LWIR. No seu trabalho, foram usadas transformações na codificação das projecções em escala de cinzentos, *eigenfaces* e filtros de comparação (*matching pursuit filters*). Os autores obtiveram o melhor desempenho no reconhecimento de faces usando as imagens do infravermelho, utilizando as projecções em escala de cinzentos.

2.2.3 Factores que Afectam o Reconhecimento de Faces em Imagens no Infravermelho Térmico

O reconhecimento facial é realizado através da normalização e alinhamento das imagens. Esta normalização é baseada na localização de posições de olhos. *Selinger et al.* [76] estudaram o efeito da localização dos olhos na PCA para o reconhecimento facial. Os olhos são localizados nas imagens LWIR e nas imagens no visível. Para as imagens LWIR os autores utilizaram características *Haar-like* disponíveis no *Open Computer Vision Library* (OpenCV) [8]. Para as imagens no visível eles usaram a

transformada circular de *Hough*, visto que procuram um círculo escuro no centro de um fundo claro. O desempenho da localização dos olhos foi comparada com as regiões localizadas manualmente, e o desempenho de reconhecimento facial foi avaliado em imagens normalizadas usando a localização dos olhos. As experiências da localização dos olhos foram realizada em sujeitos que não usam óculos. Os resultados mostram que o erro na posição dos olhos é mais elevado para as imagens no infravermelho do que nas imagens no visível. Relativamente ao reconhecimento facial, os resultados mostram que a localização errada dos olhos degrada o desempenho do reconhecimento facial. Esta degradação é mais perceptível quando a localização dos olhos é feita nas imagens LWIR. O desempenho do reconhecimento facial, usando a localização dos olhos na normalização das faces, é ainda pior quando utilizam a PCA. O algoritmo desenvolvido pela *Equinox* [4] é referido como sendo o que obtém melhores resultados, mas resultados semelhantes foram obtidos por *Kang et al.* [56] no reconhecimento facial para imagens SWIR usando a PCA e a LDA. As posições dos olhos têm importantes implicações sobre o desempenho de reconhecimento facial.

Siddiqui et al. [79] estudaram o efeito do frio e da presença de óculos de imagens térmicas. O núcleo do sistema de reconhecimento facial apresentado é a PCA. A abordagem proposta detecta e substitui as regiões frias e os óculos pelos valores médios das áreas de outra face. Os autores relataram um aumento na taxa de sucesso com a abordagem proposta.

Heo et al. [50, 51] apresentaram que a presença de óculos degrada o desempenho do reconhecimento facial em imagens térmicas. A detecção dos óculos e substituição destes por um modelo do olho obtido pela média das regiões dos olhos de múltiplas imagens (figura 2.5), levou a um aumento no desempenho do reconhecimento facial na presença de óculos. A mesma abordagem é usada por *Kong et al.* [58], mas com uma estratégia de fusão em multi-escala.

2.2.4 Fusões

O reconhecimento facial usando imagens no infravermelho tem vantagens interessantes sobre o reconhecimento de faces no visível, especialmente quando há variações na iluminação e nas expressões faciais. No entanto, o reconhecimento de faces no visível obtém um melhor desempenho em situações de iluminação controlada e quando o sujeito usa óculos. Nos últimos anos, mais ênfase foi colocada sobre a fusão de

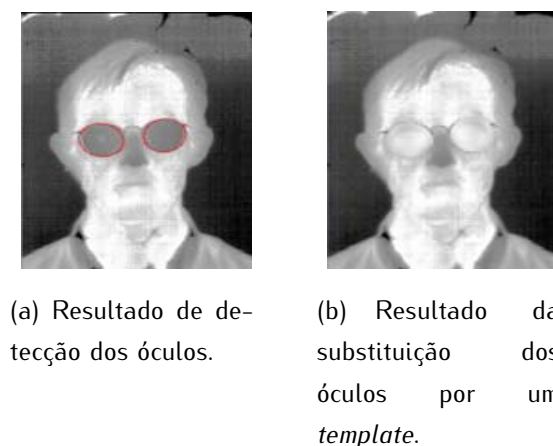


Figura 2.5: Exemplo da detecção e substituição dos óculos por um *template* [51].

imagens no visível e no infravermelho, a fim de atingir taxas de reconhecimento mais altas. Os métodos de fusão podem ser divididos em duas categorias: fusão baseada no resultado e fusão dos dados. A fusão baseada no resultado combina os resultados obtidos nos vários espectros, individualmente. Na fusão dos dados, as imagens fundem-se entre si com uma combinação dos valores dos pixéis.

2.2.4.1 Fusões Baseadas no Resultado

Na fusão baseada no resultado muitos tem sido os métodos propostos para a combinação dos resultados. *Chen et al.* [33, 34] propuseram uma estratégia de fusão baseada no desempenho dos métodos (soma dos desempenhos dos métodos para cada classe) e de uma estratégia baseada na votação (soma dos votos para cada classe). Eles relatam que a fusão das imagens no visível com as do infravermelho supera o reconhecimento facial num espectro individual. Eles também mostram que a estratégia baseada na votação obtém melhores resultados.

Socolinsky et al. [86, 87] utilizaram o somatório das pontuações resultantes do reconhecimento facial em imagens no visível e no LWIR. Os resultados mostram uma melhoria significativa nas taxas de reconhecimento com a estratégia da fusão proposta.

Chen et al. [34] utilizaram uma fusão dos resultados dos reconhecimento facial no visível e no LWIR baseada no *fuzzy integral*. Na sua abordagem é aplicada a PCA em cada um dos espectros. O conjunto das principais componentes da PCA é mapeado para um vector de características e normalizado. A "fuzificação" é aplicada aos vectores

extraídos (do visível e do infravermelho) usando uma modificação do histograma. A fusão usando o *fuzzy integral* permitiu um aumento das taxas de reconhecimento.

Buddharaju et al. [22] propuseram a combinação do método descrito em [69, 24, 25, 26] com o tradicional método de reconhecimento de faces PCA aplicada às imagens no visível (figura 2.6). A abordagem multi-espectral proposta baseia a sua pontuação final na fusão ao nível da decisão, combinando as pontuações individuais dos algoritmos de reconhecimento no visível e infravermelho térmico. A estratégia da fusão proposta fornece resultados melhores do que as duas abordagens individuais sozinhas.

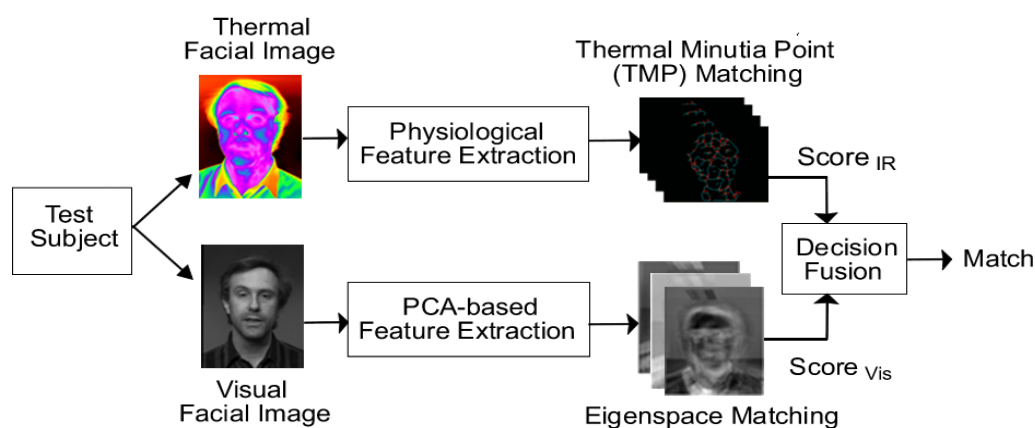


Figura 2.6: Reconhecimento facial multi-espectral [22].

Bowyer et al. [21] estudaram a utilização das imagens no visível, 3D e no infravermelho (LWIR) para o reconhecimento facial. Os autores analisam o desempenho de uma abordagem multi-modal, que combina imagens no visível com imagens 3D, no infravermelho e uma abordagem *multi-sample* (várias imagens da mesma face). Os autores apresentam um desempenho significativamente maior para a combinação de diferentes tipos de imagens (abordagem multi-modal) em relação ao uso das modalidades individuais. O melhor desempenho foi obtido através da combinação de todas as três modalidades (visível, infravermelho e 3D), significativamente maior que a abordagem *multi-sample*. A fusão proposta faz uma ponderação da pontuação da PCA.

Kakadiaris et al. [54, 55] apresentaram uma abordagem multi-modal de reconhecimento facial usando imagens no visível e térmicas (figura 2.7). O modelo proposto utiliza as seguintes características: informação geométrica em 3D, textura 2D no visível e textura 2D no infravermelho. O mapeamento da textura é então utilizado para mapear uma imagem da face num modelo 3D. Os modelos 3D mapeados são projectados num

espaço 2D paramétrico (espaço UV) para produzir o modelo da deformação da face utilizado na fase de reconhecimento. As *wavelets* de *Haar* são então utilizadas para a compressão dos dados. As imagens no infravermelho são utilizadas para extrair as regiões da pele usadas no reconhecimento. A rede vascular também é extraída a partir de imagens térmicas e projectada para o espaço UV paramétrico. O reconhecimento é feito através da comparação das assinaturas obtidas pelas deformações paramétricas da imagem, imagens paramétricas térmicas e pelos mapas das texturas das imagens no visível.

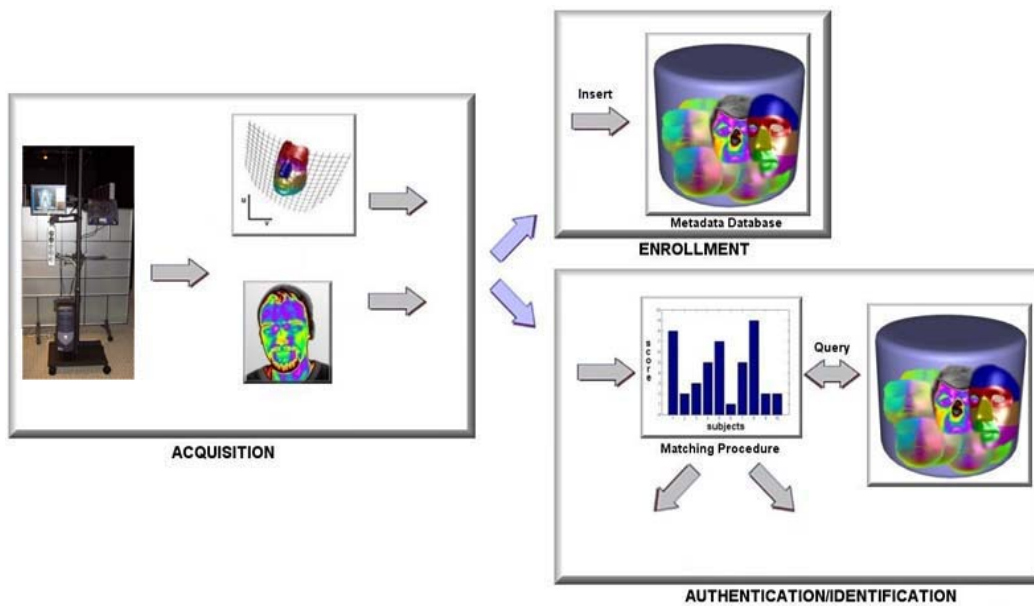


Figura 2.7: Visão global sobre reconhecimento facial multi-modal [54].

2.2.4.2 Fusões dos Dados

A fusão dos dados no reconhecimento de faces no infravermelho tem atraído muito interesse nos últimos anos. A maior parte utiliza a decomposição das *wavelets* para realizar a fusão dos dados. *Gyaourova et al.* e *Singh et al.* [45, 82] usaram um esquema de fusão dos dados das imagens no infravermelho e no visível para o reconhecimento facial. A abordagem proposta visa superar as limitações causadas pela opacidade dos óculos nas imagem no infravermelho. As imagens no visível são menos sensíveis aos óculos do que as imagens no infravermelho, por outro lado, as imagens no infravermelho térmico são menos sensíveis à variação da iluminação. A fusão das duas modalidades pode levar a uma melhoria dos resultados. A estratégia de fusão opera no domínio

das *wavelets*, combinando os coeficientes das *wavelets* de *Haar*. Os coeficientes das *wavelets*, tanto os das imagens no visível como os das imagens no LWIR, são combinados usando Algoritmos Genéticos (AGs). Os AGs escolhem quais os coeficientes que devem ser utilizados e como combiná-los. Os autores apresentam uma melhoria considerável quando são usadas as imagens fundidas e existe a presença de óculos.

As *wavelets* de *Haar* também foram usadas por *Chang et al.* e *Kong et al.* [31, 32, 58]. *Chang et al.* [31, 32] propuseram um esquema de fusão de dados para ajustar a iluminação e melhorar as taxas de reconhecimento facial. Os testes foram realizados na fusão de imagens multi-espectrais. Eles utilizam duas abordagens: a fusão ponderada e *wavelets* de *Haar* para a fusão dos dados. O desempenho no reconhecimento facial foi testado usando *Facelt*® e o melhor desempenho foi obtido com as imagens fundidas.

Kong et al. [58] apresentaram uma estratégia de fusão multi-escala de imagens no visível e térmicas. O método de registo usa as derivadas dos mapas direccionais [20] e a fusão dos dados em multi-escala é realizada usando a *Discrete Wavelet Transform* (DWT). A fusão dos dados consiste numa combinação ponderada dos coeficientes da DWT das imagens no visível e térmicas. A fusão das imagens no infravermelho e no visível é referido como sendo a que possui melhores resultados.

Singh et al. [81] propuseram uma fusão hierárquica das imagens no infravermelho e no visível (figura 2.8). O algoritmo usa as *wavelets* de *Gabor* para realizar a fusão das imagens. No processo de treino, são usados SVMs para seleccionar as características extraídas das imagens, como a amplitude e fase. As características seleccionadas são usadas para produzir as imagens fundidas. Durante o seu trabalho, estes autores analisaram as fusões em diferentes comprimentos de onda. Sendo que a fusão entre imagens da face no visível e no SWIR foram as que produziram melhores resultados.

Bebis et al. [18] propôs uma estratégia de fusão das *eigenfeatures*. As *eigenfeatures* foram extraídos directamente das imagens no visível e no infravermelho. A selecção das *eigenfeatures* mais apropriadas para fusão das imagens no visível e no infravermelho foi realizada recorrendo aos AGs. Os autores apresentam um aumento no desempenho de reconhecimento com a fusão proposta.

Hariharan et al. [46] apresentam uma fusão de imagens LWIR e no visível através de um *Empirical Mode Decomposition* (EMD). O EMD decompõe sinais não-lineares e não-estacionários num *Intrinsic Mode Function* (IMF). As imagens no visível e no infravermelho são transformadas num vector e o EMD é aplicado sobre esses vectores

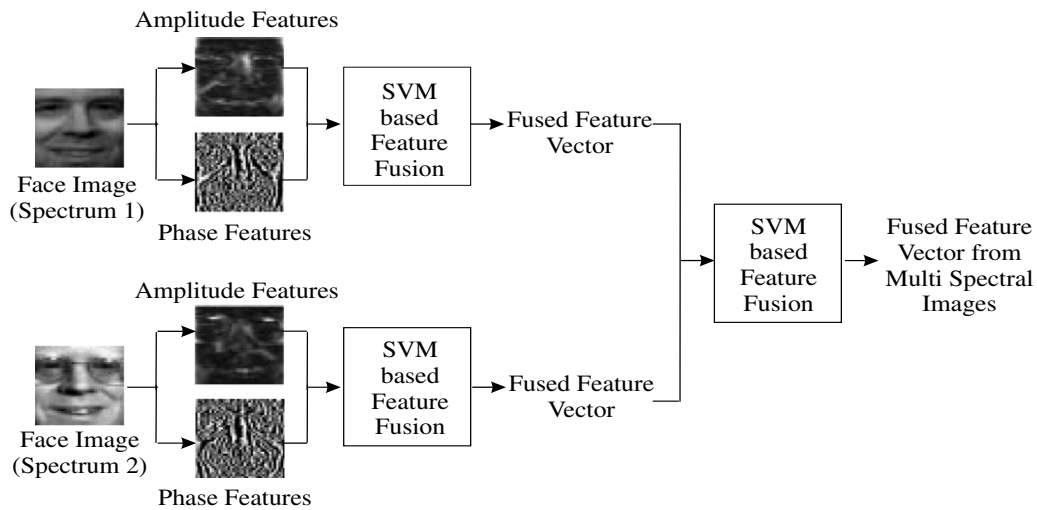


Figura 2.8: Fusão das características em imagens multi-espectrais da face [81].

para obter os IMFs no visível e no infravermelho. Um número de IMFs é seleccionado para a fusão. Os IMFs são pesados, sendo que estes pesos obtidos de forma empírica, para diminuir a informação mútua entre as imagens no visível e no infravermelho. Os IMFs são combinados e o reconhecimento facial é realizado usando a imagem fundida. As experiências foram realizadas em imagens da base de dados Equinox [4]. O reconhecimento facial foi realizado usando o *FaceIt*® e o desempenho foi avaliado em imagens fundidas, obtidas através: da média, da fusão da PCA, fusão com base nas *wavelets* e na fusão EMD. Os autores relatam que o melhor desempenho foi obtido com as imagens da fusão usando o EMD, mesmo na presença de mudança de condições de iluminação e expressões faciais.

2.2.4.3 Reconhecimento Híbrido e Fusões dos Dados

Heo et al. *Abidi et al.* [49, 10] descrevem a fusão das imagens no visível e no infravermelho para o reconhecimento facial. A fusão das imagens é realizada nos dados e nos níveis da decisão (figura 2.9). A fusão a nível dos dados permite a integração das imagens no visível e no infravermelho. A fusão a nível da decisão permite a combinação dos resultados do reconhecimento tanto no visível como no infravermelho. Os autores propõem três técnicas de fusão: fusão das imagens no infravermelho e no visível, fusão da decisão com base na maior votação e fusão na decisão usando a votação média. A proposta da fusão dos dados é realizada através de média ponderada

das imagens no visível e no infravermelho, a fim de obter imagens com uma iluminação invariante. Os pesos são determinados através da distribuição do brilho nas imagens da face. Quando existe uma fraca iluminação o peso das imagens no visível é inferior ao das imagens térmicas. Nas imagens com elevadas temperatura da face, o peso da imagem térmica é menor do que a imagem no visível. A fusão ao nível da decisão é realizada através da combinação das votações individuais no reconhecimento facial no visível e no infravermelho, obtidos através do *FaceIt*®. Os testes foram realizados usando imagens no visível e no LWIR da base de dados Equinox [4]. Os resultados experimentais mostram que as imagens térmicas e que as técnicas de fusão superam o reconhecimento em imagens no visível quando não existem óculos. Quando existem óculos, a combinação entre as técnicas de fusão e um algoritmo de remoção dos óculos aumentam o desempenho.

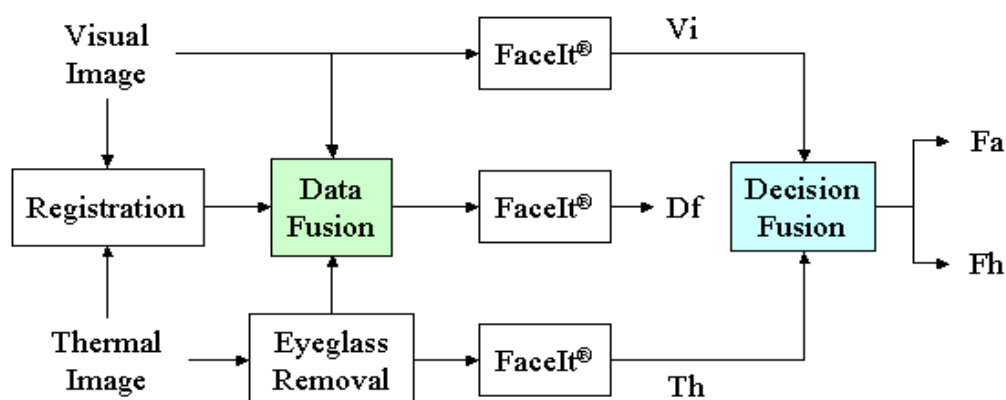


Figura 2.9: Sistema de reconhecimento facial proposto em [49].

Arandjelovic et al. [14, 15, 16] propuseram um esquema de fusão em várias etapas para o reconhecimento facial. A fusão é realizada a nível da decisão e usando as imagens no visível e no LWIR (figura 2.10). A abordagem proposta possui três etapas principais: pré-processamento dos dados, detecção de óculos e fusão global e local com características extraídas das imagens no visível e no térmico. A fusão das características das regiões ao redor dos olhos, da boca e a imagem da face são combinadas usando a PCA. A pontuação de similaridade é obtida através de uma soma ponderada, onde são usados diferentes pesos para as imagens no visível e no térmico. A fusão é realizada por uma combinação ponderada das imagens no visível e no infravermelho. Os pesos são elevados no espectro visível, se as taxas de reconhecimento neste espectro forem altas. O método de fusão proposto com a detecção de óculos superara outras técnicas

exploradas nas sua experiências.

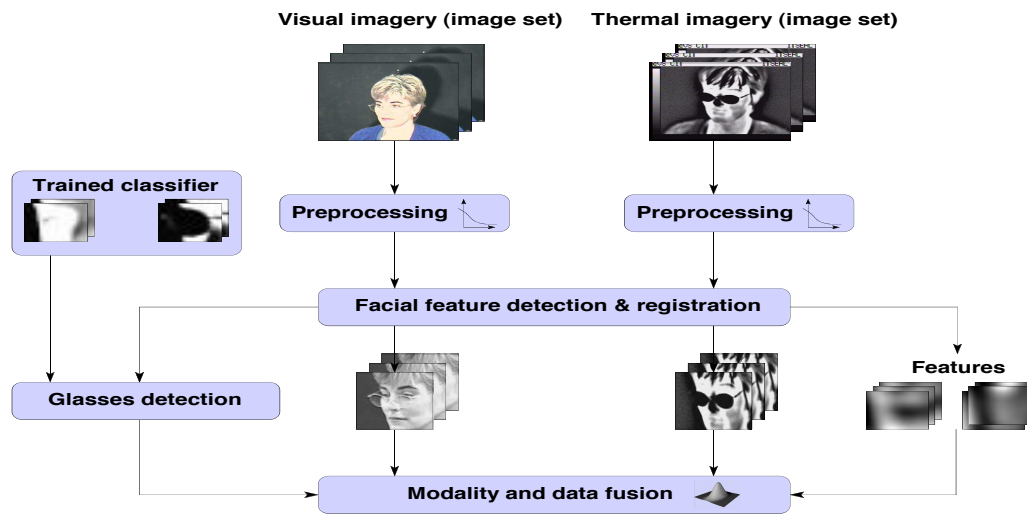


Figura 2.10: O sistema proposto em [16] é composto por três módulos principais: pré-processamento de dados e registo, detecção dos óculos e a fusão das representação face global e local, utilizando as imagens no visível e térmico.

Ali et al. [13] utilizam um conjunto de filtros de *Gabor* para processar as imagens da face. Os filtros obtidos são usados para extrair pontos característicos. Os pontos característicos são seleccionados se representam um máximo do filtro de *Gabor* numa janela predefinida. A estratégia de fusão usa uma média ponderada das imagens no visível e no infravermelho térmico, similar ao apresentado por *Heo et al.* [48, 49]. Os resultados das imagens fundidas são combinados através da soma da média ponderada das imagens no visível e no infravermelho térmico. Os resultados experimentais mostram que a estratégia de fusão proposta supera outras técnicas destes autores.

2.3 Segmentação Facial em Imagens no Infravermelho Térmico

Nesta secção iremos apresentar alguns métodos de segmentação de faces em imagens do infravermelho térmico.

Selinger et al. [76] apresentam um método de segmentação baseado numa máscara binária (figura 2.11) para remover o fundo. A aplicação da máscara é feita usando a localização dos olhos. Para as imagens no infravermelho eles localizaram os olhos

2.3. SEGMENTAÇÃO FACIAL EM IMAGENS NO INFRAVERMELHO TÉRMICO 27

utilizando as características *Haar-like* disponíveis no OpenCV [8]. No visível, eles aplicaram a transformada circular de *Hough* para procurar um círculo escuro no centro de um fundo claro. Na figura 2.3 podem ser observados alguns exemplos da aplicação desta técnica. *Gyaourova et al.* [45] usa este mesmo método de segmentação nas experiências realizadas.

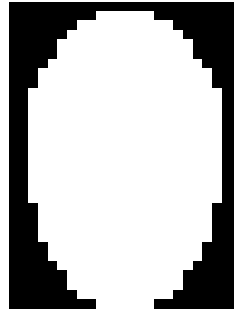


Figura 2.11: Máscara binária usada em [76].

Pavlidis et al. e *Buddharaju et al.* [69, 24, 25, 22, 26] apresentaram um método de segmentação da face usando uma abordagem Bayesiana. Este método é baseado na combinação de duas distribuições normais por classe, sendo estas estimadas usando o algoritmo *Expectation Maximization* (EM). Este algoritmo usa pixels das regiões da pele (s) e de fundo (b) para a fase de treino. As distribuições são obtidas a partir do conjunto de treino, onde fazem parte sub-regiões das imagens que contêm somente pixels de cada um dos tipos (s ou b). A figura 2.12 apresenta a forma como os autores cortam estas sub-regiões.

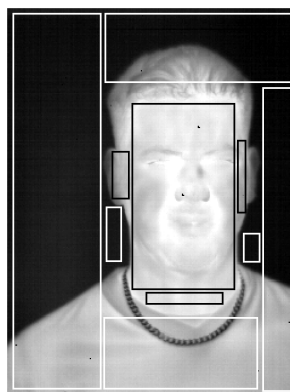


Figura 2.12: Exemplo do corte das regiões da face (regiões assinaladas com arestas pretas) e fundo (regiões assinaladas com arestas brancas).

O algoritmo 1 apresenta como é que são estimados os parâmetros das duas distribuições

normais (ω_{s_i} , μ_{s_i} e $\sigma_{s_i}^2$) para os N_s pixels da região da pele. Para os pixels das regiões do fundo tem de se aplicar o mesmo processo, mas agora para estimar ω_{b_i} , μ_{b_i} e $\sigma_{b_i}^2$ para os N_b pixels das regiões consideradas fundo. No final do processo de treino, a aplicação do algoritmo EM terá estimado 4 médias (μ), 4 variâncias (σ^2) e 4 pesos (ω).

Algoritmo 1 Estimação dos parâmetros das duas distribuições normais (ω_{s_i} , μ_{s_i} e $\sigma_{s_i}^2$) para os N_s pixels da região da pele usando o algoritmo EM.

Input: x pixels das sub-imagens da região da face, N_s número de pixels das sub-imagens da regiões da face e ϵ que é a condição de paragem (10^{-3} , 10^{-4} , ...).

Output: 2 médias (μ_{s_i}), 2 variâncias ($\sigma_{s_i}^2$) e 2 pesos (ω_{s_i}).

$k = 0$

while $|\omega_{s_t}^{(k)} - \omega_{s_t}^{(k+1)}| > \epsilon$ **do**

for $i = 1$ **to** 2 **do**

for $j = 1$ **to** N_s **do**

$$z_{ij}^{(k)} = \frac{\omega_{s_i}^{(k)} (\sigma_{s_i}^{(k)})^{-1} \exp\left\{-\frac{1}{2(\sigma_{s_i}^{(k)})^2} (x_j - \mu_{s_t}^{(k)})\right\}}{\sum_{t=1}^2 \omega_{s_t}^{(k)} (\sigma_{s_t}^{(k)})^{-1} \exp\left\{-\frac{1}{2(\sigma_{s_t}^{(k)})^2} (x_j - \mu_{s_t}^{(k)})\right\}}$$

$$\omega_{s_i}^{(k+1)} = \frac{\sum_{j=1}^{N_s} z_{ij}^{(k)}}{N_s}$$

$$\mu_{s_i}^{(k+1)} = \frac{\sum_{j=1}^{N_s} z_{ij}^{(k)} x_j}{N_s \omega_{s_i}^{(k+1)}}$$

$$(\sigma_{s_i}^{(k+1)})^2 = \frac{\sum_{j=1}^{N_s} z_{ij}^{(k)} (x_j - \mu_{s_i}^{(k+1)})^2}{N_s \omega_{s_i}^{(k+1)}}$$

end for

end for

$k = k + 1$

end while

Na fase de segmentação, para cada pixel existe uma distribuição a priori ($\pi^{(t)}(\theta)$ onde t é a iteração) para verificar se este pixel é pele ($\pi^{(t)}(s)$) ou *background* ($\pi^{(t)}(b) = 1 - \pi^{(t)}(s)$). A probabilidade inicial para a distribuição a priori é igual a $\pi^{(1)}(s) = \frac{1}{2} = \pi^{(1)}(b)$, onde θ é o parâmetro de interesse e pode tomar dois valores (s ou b).

O valor do pixel x_t possui uma distribuição condicional $f(x_t|\theta)$ e se o pixel for pele

2.3. SEGMENTAÇÃO FACIAL EM IMAGENS NO INFRAVERMELHO TÉRMICO 29

obtemos:

$$f(x_t|s) = \sum_{i=1}^2 \omega_{s_i} \mathcal{N}(\mu_{s_i}, \sigma_{s_i}^2), \quad (2.1)$$

onde $\mathcal{N}(\mu_{s_i}, \sigma_{s_i}^2)$ é a distribuição normal com média μ_{s_i} e variância $\sigma_{s_i}^2$.

A distribuição a priori ($\pi^{(t)}(\theta)$) combinada com a probabilidade ($f(x_t|\theta)$) fornece (pelo teorema de *Bayes*) a distribuição a posteriori ($p^{(t)}(\theta|x_t)$), onde, para os pixels da face, temos:

$$p^{(t)}(s|x_t) = \frac{\pi^{(t)}(s)f(x_t|s)}{\pi^{(t)}(s)f(x_t|s) + \pi^{(t)}(b)f(x_t|b)} \quad (2.2)$$

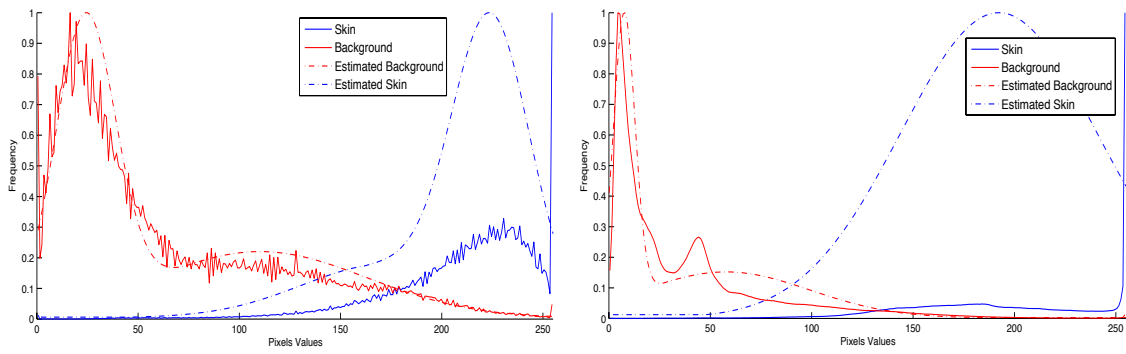
A distribuição a posteriori é também usada para obter a distribuição a priori para a iteração seguinte ($t + 1$):

$$\pi^{(t+1)}(\theta) = \begin{cases} \pi^{(t+1)}(s) = p^{(t)}(s|x_t), & \text{when } \theta = s \\ \pi^{(t+1)}(b) = 1 - \pi^{(t+1)}(s), & \text{when } \theta = b \end{cases} \quad (2.3)$$

Os gráficos da figura 2.13 contêm as distribuições das intensidades dos pixels (linha contínua) das imagens recortadas usadas no conjunto de treino (como apresentamos na figura 2.12), para as quatro bases de dados que foram usadas neste trabalho (serão apresentadas no capítulo 3). As linhas a tracejado são os resultados das combinações *Gaussianas* estimadas pelo EM para cada classe (vermelho para o fundo e azul para a pele).

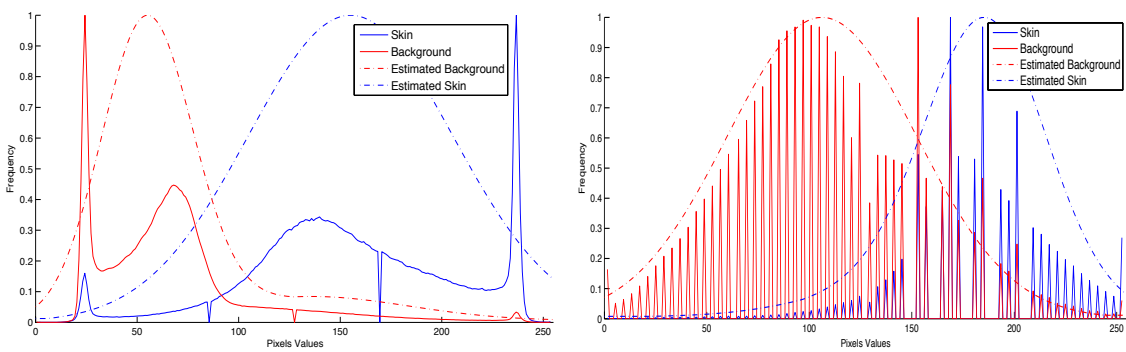
Na figura 2.13(d), as distribuições possuem muitas variações devido a que as imagens da base de dados FSU possuem em média apenas 70 valores diferentes.

Cho et al. [35] apresentam um método para a segmentação da face em imagens LWIR com base no detector de arestas de *Sobel* e operações morfológicas (figura 2.14). O primeiro passo deste método é a remoção de alguns ruído que possa existir na imagem através de um filtro de média (figura 2.14(a)). No resultado do filtro de média, eles detectam as arestas com o método de *Sobel*. No passo seguinte são aplicadas operações morfológicas para inter-ligar os contornos que tenham ficado abertos e remover pequenas áreas (figura 2.14(b)). Por fim eles seleccionam o maior contorno e preenchem essa região (figura 2.14(c)), pois para eles esta região é a que melhor descreve a face.



(a) Distribuição das intensidades dos pixéis para a base de dados UND.

(b) Distribuição das intensidades dos pixéis para a base de dados Terravic.



(c) Distribuição das intensidades dos pixéis para a base de dados IRIS.

(d) Distribuição das intensidades dos pixéis para a base de dados FSU.

Figura 2.13: Distribuições das intensidades dos pixéis da face (linha azul) e fundo (linha vermelha) para as quatro bases de dados usadas.



(a) Imagem da face depois de aplicado o filtro da média.



(b) Arestas detectadas pelo método de *Sobel* depois de serem removidas pequenas áreas.



(c) Máscara final, depois de seleccionarem o maior contorno e preenchem essa região.

Figura 2.14: Aplicação do método de segmentação apresentado em [35].

2.4 Conclusão

Como vimos existem numerosos métodos de reconhecimento no infravermelho térmico, sendo que alguns destes métodos podem ser aplicados tanto a imagens da face no visível como no infravermelho. Outros métodos como os baseados nos TMPs apenas podem ser aplicados às imagens no infravermelho térmico.

Em contraste com o comprimento de onda visível, onde numerosos métodos de segmentação de faces têm sido propostos, tanto baseados na cor [83], na geometria da face [27], etc., no LWIR há uma falta de propostas para melhorar o status actual. No LWIR não existem artigos publicados que apenas falem da segmentação de faces (como acontece no visível). Por norma, nesta área os autores apenas dedicam uma pequena parte do artigo a falar sobre este tema e outros nem o referem. Na nossa opinião, este é um processo muito importante nos sistemas de reconhecimento e que deve de ser tratado com o devido cuidado. Isto deve-se ao facto de que todos os sistemas de reconhecimento possuem uma fase segmentação, isto é, qualquer que seja o padrão que pretendemos reconhecer, este terá de ser segmentado primeiro.

Em suma, depois de analisar as maiores lacunas existentes entre o reconhecimento e segmentação de faces no infravermelho térmico, observamos que esta lacuna estava localizada na segmentação. Com isso, a nossa proposta para o trabalho a realizar nesta dissertação recaiu na segmentação de faces em imagens LWIR.

Capítulo 3

Bases de Dados

3.1 *University of Notre Dame (UND)*

A base de dados *University of Notre Dame (UND)* é apresentada em [41, 34], tendo esta sido capturada pelo *Computer Vision Research Laboratory*. A colecção X1 possui imagens capturadas no visível e no infravermelho térmico, mas para este trabalho apenas iremos usar as imagens do infravermelho térmico. Nesta base de dados existem 2293 imagens LWIR capturadas a 81 pessoas diferentes. As faces presentes nesta base de dados são todas frontais, existe variação das expressões faciais e foi capturada em várias sessões. A captura foi feita usando uma câmara *Merlin Uncooled Long-Wavelength IR* e a resolução desta é de 320×240 píxeis.

Para esta base de dados criámos um conjunto de treino que contém 159 imagens e o conjunto de teste 163. A figura 3.1 apresenta algumas imagens desta base de dados.



Figura 3.1: Exemplos de imagens da base de dados UND.

3.2 *Terravic Facial IR*

A base de dados *Terravic* foi criada pela *Terravic Corporation* e neste momento é o sub-conjunto 4 das bases de dados disponíveis pela *Object Tracking and Classification in and Beyond the Visible Spectrum* (OTCBVS) [6]. Esta base de dados contém 24'508 imagens de 20 pessoas diferentes. As faces capturadas possuem rotações que variam entre os 0° e os 90°, tanto para o lado direito como para o lado esquerdo. Outras características desta base de dados é que possui faces capturadas no interior e no exterior, com e sem óculos e também com e sem chapéu. A câmara usada para a capturar as imagens foi uma *Raytheon L-3 Thermal-Eye 2000AS* e a resolução desta é de 320 × 240 pixéis.

O conjunto de treino criado para esta base de dados contém 235 imagens e o de teste 240. Exemplos de imagens desta base de dados podem ser visualizadas na figura 3.2.



Figura 3.2: Exemplos de imagens da base de dados *Terravic*.

3.3 *IRIS Thermal/Visible Face*

A base de dados *IRIS Thermal/Visible Face* foi criada pelo *Imaging, Robotics, and Intelligent Systems laboratory (University of Tennessee)* (IRIS), sendo disponibilizada por parte da OTCBVS como sendo o seu sub-conjunto número 2 [5]. Esta base de dados conta com 4228 imagens no visível e no infravermelho térmico (apenas iremos usar as imagens do infravermelho térmico e ambas têm uma resolução de 320 × 240 pixéis). A cada uma das 29 pessoas que participaram na construção desta base de dados foram capturadas entre 176 e 250 imagens, nas quais foram variando as rotações da cabeça/poses, as iluminações (cinco tipos de iluminação diferentes) e expressões faciais (três tipos diferentes), usando o sistema de captura apresentado na figura 3.3.

As iluminações foram variando devido às imagens no visível, pois são as únicas que são afectadas pelas diferentes iluminações.

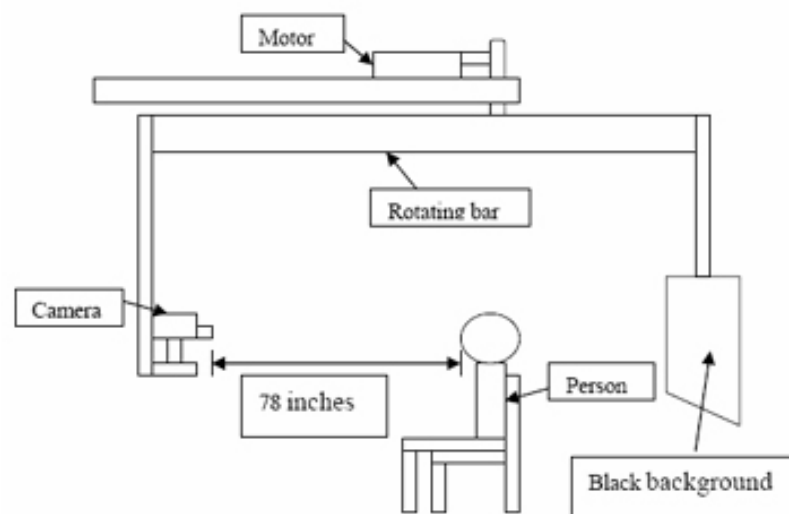


Figura 3.3: Representação do sistema de captura da base de dados IRIS [5].

A câmara usada para capturar as imagens desta base de dados aplica uma moldura às imagens, como se pode ver na figura 3.4(a). Antes de iniciarmos as nossas experiências, tivemos de aplicar um pré-processamento às imagens de forma a remover a moldura, para que esta não interferisse na segmentação. As zonas removidas foram as barras presentes no topo (20 pixéis), esquerda (50 pixéis) e fundo (20 pixéis) da imagem. Com a remoção destas zonas as imagens (figura 3.4(b)) deixaram de ter uma dimensão de 320×240 pixéis para passarem a ter 270×200 pixéis.

O conjunto de treino e de teste para esta base de dados contém cada um 296 imagens.

3.4 Florida State University (FSU)

A base de dados *Florida State University* (FSU) contém 234 imagens de 10 pessoas diferentes, com várias rotações e expressões faciais [90]. Esta base de dados possui duas particularidades. A primeira é que é uma base de dados a cores (figura 3.5) e a outra particularidade é que as imagens não possuem todas a mesma dimensão, o que nos leva a pensar que estas imagens sofreram um pré-processamento por parte dos autores.



(a) Imagem Original com dimensão de 320×240 pixels.



(b) Imagem depois de pré-processada com dimensão de 270×200 pixels.

Figura 3.4: Exemplo do pré-processamento aplicado às imagens da base de dados IRIS.



Figura 3.5: Exemplo de imagens da base de dados FSU.

O conjunto de treino desta base de dados contém 40 imagens e o de teste 194.

3.5 Outras

Ao longo deste trabalho encontramos referências para mais duas bases de dados, sendo estas a base de dados *Equinox* [4] e *Biometrics Engineering Research Center* (BERC).

As imagens da base de dados *Equinox* foram capturadas no visível, SWIR, MWIR e LWIR, onde foram variadas as condições de iluminação (figura 3.6). *D. Socolinsky et al.*

[89] referem que a base de dados contém 175440 imagens da face, onde participaram cerca de 340 pessoas. Neste nosso trabalho não a pudemos usar visto já não estar disponível (informação cedida pelos autores).

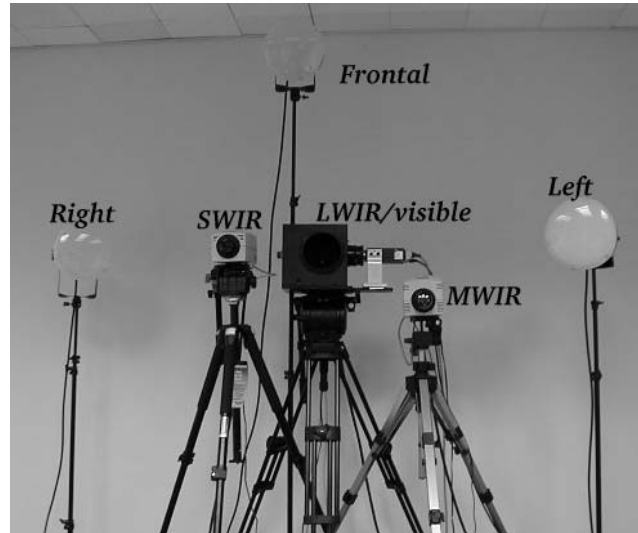


Figura 3.6: Câmaras e iluminações usadas na captura simultânea das imagens em vários espectros (visível, SWIR, MWIR e LWIR) [89].

A outra base de dados que encontramos citada é a BERC. A esta não tivemos acesso devido ao facto de que esta base de dados é privada (informação cedida pelos autores).

3.6 Conclusão

Para este trabalho iremos usar as quatro bases de dados disponibilizadas publicamente. O uso de mais do que uma base de dados irá possibilitar que vejamos os comportamentos dos métodos de segmentação que irão ser apresentados em imagens com diferentes características.

Os nomes das imagens presentes nas divisões realizadas (definição dos conjuntos de treinos e testes) para as várias bases de dados estão disponíveis em http://www.di.ubi.pt/~lfbaa/thermal_face.html.

Capítulo 4

Métodos de Segmentação Propostos

Depois de avaliarmos os métodos propostos por *Cho et al.* [35], e *Pavlidis et al.* e *Buddharaju et al.* [69, 24, 25, 22, 26] verificamos que era possível melhorar os seus resultados, como veremos nas secções seguintes.

4.1 Operações

Nas sub-secções 4.1.1 a 4.1.5 iremos descrever várias operações que irão ser usadas nos métodos que propomos na secção 4.2.

4.1.1 Região de Interesse Rectangular (RIR)

Uma operação interessante seria obtermos uma Região de Interesse Rectangular (RIR) que contivesse a face. Essa operação evitaria os problemas provocados pelas roupas. Este problema aparece devido a que ao existir contacto entre o corpo e as roupas estas também se tornam quentes, visto que o corpo está quente. Isto faz com que a roupa tenha uma temperatura similar à da pele, o que faz com que dificulte a tarefa dos métodos de segmentação baseados nas intensidades dos pixéis.

Para obter a RIR, analisámos as assinaturas verticais e horizontais das imagens. As assinaturas são vectores que contêm a soma das intensidades dos pixéis ao longo

das colunas (equação (4.1)) e linhas das imagens (equação (4.2)), respectivamente:

$$sigV(c) = \sum_{r=1}^{nR} i(c, r) \quad (4.1)$$

$$sigH(r) = \sum_{c=1}^{nC} i(c, r) \quad (4.2)$$

onde c e r são os índices das colunas e linhas da imagem i com dimensão $nC \times nR$.

O primeiro passo para obter a RIR passa pela análise das assinaturas verticais (figura 4.1). A assinatura da figura 4.1(b) representa a assinatura vertical da imagens na figura 4.1(a). Neste sinal, podemos observar que possui várias oscilações de alta frequência que irão aparecer na sua derivada. Para evitar isso, nós propomos que esta seja suavizada usando um filtro Gaussiano 1D (apresentado na figura 4.1(c)). O desvio padrão do filtro Gaussiano é $\sigma = 0.05 \times nC$. Este valor foi obtido depois de estudar a influência de diferentes valores para σ nas imagens do conjunto de treino. A suavização do sinal apresentado na figura 4.1(b) com o filtro apresentado na figura 4.1(c) foi feita recorrendo à convolução unidimensional através da Fast Fourier Transform (FFT). O resultado da convolução é apresentado na figura 4.1(d). Como podemos observar este sinal possui menos oscilações, o que faz com que a derivada deste sinal não possua altas frequências. A derivada do sinal é apresentado na figura 4.1(e). O passo seguinte consiste em determinar os pontos extremos do sinal: nas figuras 4.1(a) e 4.1(e) marcamos o máximo com a linha a vermelho (*colLeft*) e o mínimo a verde (*colRight*). As duas linha indicam a localização da uma grande variação na intensidade da imagem, usando estas como sendo os lados da face.

O passo seguinte, na definição da RIR, é a análise da assinatura horizontal. Para isso consideramos a parte da imagem entre os dois extremos detectados pela análise da assinatura vertical. Esta análise remove os ombros dos indivíduos e supera uma das questões que causava problemas nos métodos propostos em [35, 69, 24, 25, 22, 26]. O processo usado para a análise da assinatura horizontal (apresentado na figura 4.2) é similar ao processo usado na análise da assinatura vertical. A principal diferença é o filtro usado na convolução. Neste caso, o filtro usado é um filtro rectangular (figura 4.2(d)), a dimensão deste filtro é de $0.15 \times nR$. A forma e o tamanho do filtro foram seleccionadas para eliminar variações bruscas que possam aparecer no sinal quando existe a presença de óculos ou porque o sujeito possui o nariz frio. Para obter

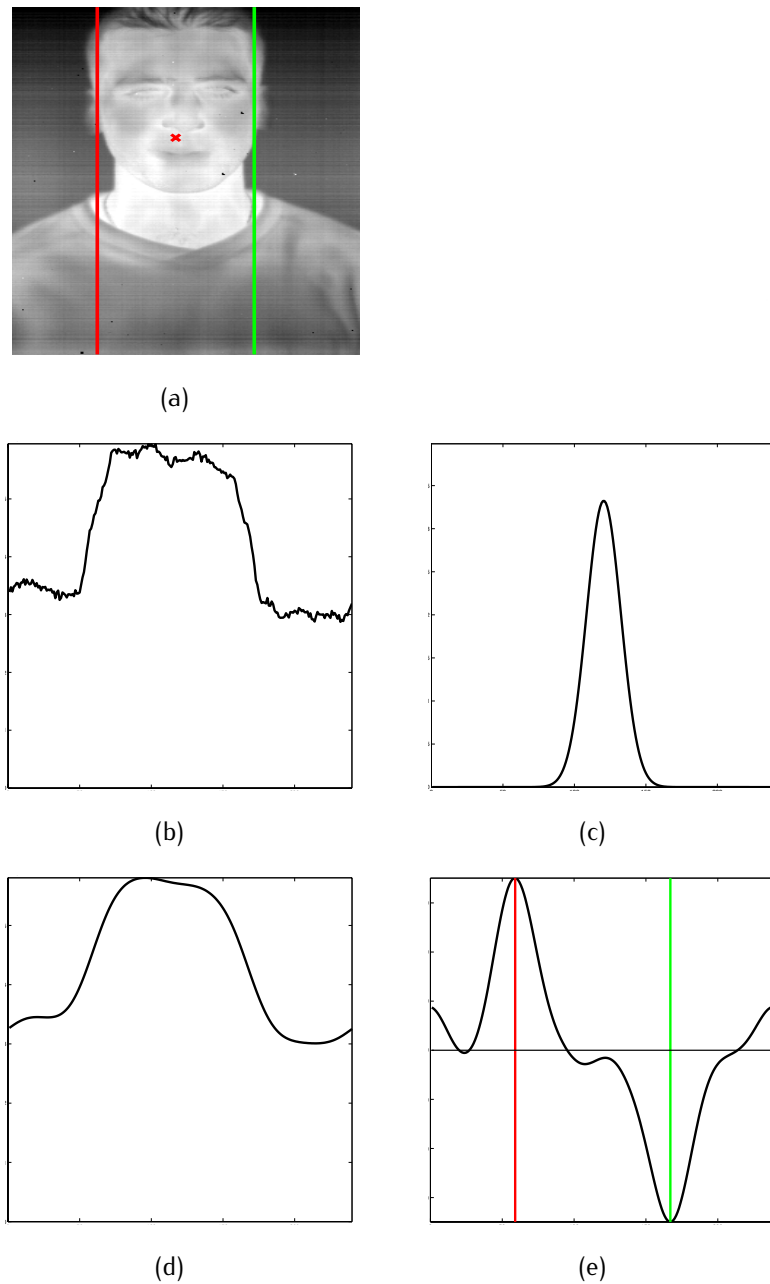


Figura 4.1: Análise da assinatura vertical. A figura 4.1(a) é a imagem original. A figura 4.1(b) é a assinatura vertical da imagem original. A figura 4.1(c) é o filtro Gaussiano usado na convolução, com $\sigma = 0.05 \times nC$. A figura 4.1(d) é o resultado obtido depois de aplicarmos a convolução entre o sinal da figura 4.1(b) e o filtro. A figura 4.1(e) é a derivada do sinal da figura 4.1(d).

os extremos do sinal foi aplicado um processo semelhante ao usado na análise da assinatura vertical. O máximo (*rowUp*) é representado pela linha vermelha e o mínimo

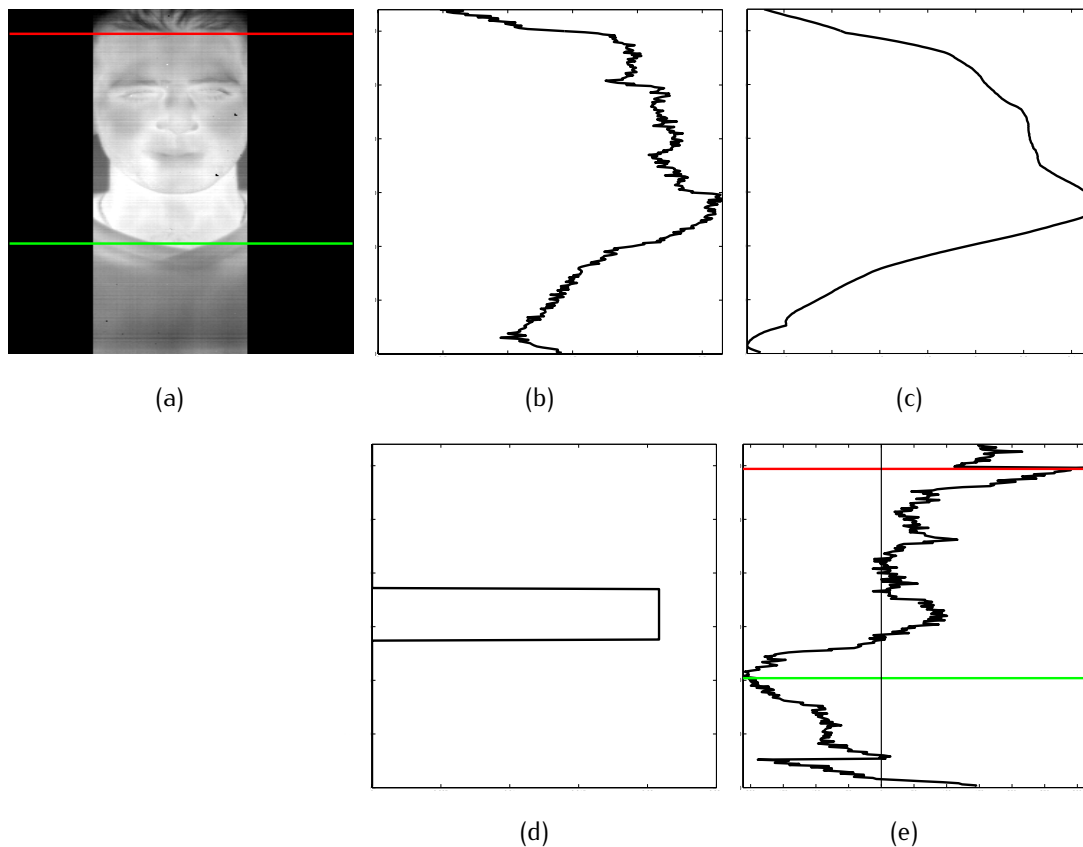


Figura 4.2: Análise da assinatura horizontal. A figura 4.2(a) é a imagem resultante da análise da assinatura vertical. A figura 4.2(b) é a assinatura horizontal. A figura 4.2(c) é o resultado da convolução entre o sinal da figura 4.2(b) e o filtro da figura 4.2(d). A figura 4.2(d) é o filtro usado na convolução. A figura 4.2(e) é a derivada do sinal da figura 4.2(c).

(*rowDown*) pela linha verde, nas figuras 4.2(a) e 4.2(e).

4.1.2 Região de Interesse Elíptica (RIE)

A ideia de definir uma elipse em torno da face provém do facto de que a face tem uma forma que se aproxima a uma elipse. Um exemplo é o trabalho apresentado em [76] onde é definida uma máscara binária em forma de elipse para a segmentação de faces. Nós iremos usar uma máscara elíptica para melhorar a operação RIR em torno da face e para inicializar a máscara usada no método [30] (apresentado na secção 4.1.4).

A elipse é definida no interior da RIR, previamente obtida. Depois de obtermos a RIR iremos calcular o seu ponto central, que irá ser o ponto central da face e da elipse. Para determinar o ponto central ($colCFace, rowCFace$), ponto vermelho na

figura 4.1(a), usamos os pontos extremos obtidos na operação RIR:

$$colCFace = (colLeft + colRight)/2 \quad (4.3)$$

$$rowCFace = (rowUp + rowDown)/2 \quad (4.4)$$

Depois, aplicando o algoritmo 2, obtemos os pontos $\{(X_0, Y_0), \dots, (X_{2\pi}, Y_{2\pi})\}$ da elipse. O algoritmo recebe as coordenadas do centro da face $(colCFace, rowCFace)$ e as coordenadas do canto superior esquerdo da RIR $(colLeft, rowUp)$. Esses pontos são usados para obter a distância entre o centro da face e o lado esquerdo da RIR (denominada por a) e a distância entre o centro da face e o topo da RIR (denominada por b). a e b são usadas para converter as coordenadas polares dos pontos que pertencem à elipse em coordenadas cartesianas $(\{(X_0, Y_0), \dots, (X_{2\pi}, Y_{2\pi})\})$.

Algoritmo 2 Algoritmo para a obtenção dos pontos da elipse.

Entrada: $(colLeft, rowUp)$ e $(colCFace, rowCFace)$

Saída: $\{(X_0, Y_0), \dots, (X_{2\pi}, Y_{2\pi})\}$

$a = |colLeft - colCFace|$

$b = |rowUp - rowCFace|$

for $\theta = 0$ to 2π **do**

$X(\theta) = colCFace + (a * \cos(\theta))$

$Y(\theta) = rowCFace + (b * \sin(\theta))$

end for

4.1.3 Limiarização (*Thresholding*)

Em alguns dos métodos que iremos propor, vamos usar um limiar (*threshold*). O limiar é adaptativo, no sentido que este depende das distribuições de cada base de dados. O objectivo deste limiar é separar o maior número possível de pixels da pele dos de fundo. O valor deste será definido de forma a incluir alguns pixels de fundo para que a maioria dos pixels da pele sejam incluídos.

O primeiro passo na definição do valor de limiar é encontrar o ponto onde as distribuições dos pixels da pele e de fundo se cruzam, ponto onde as linhas contínuas se cruzam nos gráficos da figura 2.13. Por fim, o valor do limiar será metade do valor da intensidade do pixel onde as linhas se cruzam.

4.1.4 Contornos Activos sem Arestas

O método descrito por *Chan et al.* [30, 94] é baseado em técnicas de evolução da curva, como *Mumford-Shah* [65] e pelo modelo *Level Set Formulation* [66]. O contorno activo sem arestas pode detectar objectos cujos limites não são necessariamente definidos pelo gradiente. A minimização da energia pode ser vista como um caso particular do problema da minimização das partições. No *level set formulation*, o problema torna-se num fluxo da curvatura média que envolve o contorno activo. No entanto, o termo de paragem não depende do gradiente da imagem, como nos modelos clássicos de contornos activos.

O método dos contornos activos sem arestas é um método de minimização da energia baseada na segmentação. Supondo que uma imagem ω_0 é composta por duas regiões com intensidades aproximadamente constantes, de valores distintos ω_0^i e ω_0^j , o objecto que queremos segmentar é representado pelos valores de ω_0^i e o contorno é representado por C_0 . Portanto, dentro do objecto temos $\omega_0 \approx \omega_0^i$ e $\omega_0 \approx \omega_0^j$ fora do contorno do objecto. Considerando o termo ajuste:

$$F_1(C) + F_2(C) = \int_{\text{inside}(C)} |\omega_0(x, y) - c_1|^2 dx dy + \int_{\text{outside}(C)} |\omega_0(x, y) - c_2|^2 dx dy, \quad (4.5)$$

onde C é qualquer outra curva variável e c_1, c_2 constantes dependentes de C , sendo estas as intensidades médias de ω_0 no interior e exterior do contorno C , respectivamente. Neste caso, C_0 contorno do objecto é minimizado quando:

$$\inf_C \{F_1(C) + F_2(C)\} \approx 0 \approx F_1(C_0) + F_2(C_0) \quad (4.6)$$

Por exemplo, se a curva C está fora do objecto, então $F_1(C) > 0$ e $F_2(C) \approx 0$. Mas se a curva C está dentro do objecto $F_1(C) \approx 0$ e $F_2(C) > 0$. Se a curva C está dentro e fora do objecto, então $F_1(C) > 0$ e $F_2(C) > 0$. Finalmente, a energia é minimizada se $C = C_0$ ou seja, se a curva C é o contorno do objecto. Na figura 4.3 são apresentadas todas as possibilidades existentes no ajuste do contorno.

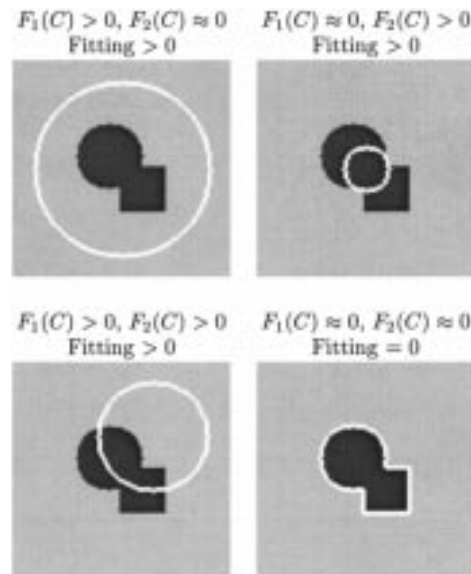


Figura 4.3: Casos possíveis no ajuste da contorno [30].

Com isto, a energia é definida por $F(c_1, c_2, C)$:

$$\begin{aligned}
 F(c_1, c_2, C) = & \mu \cdot \text{Lenght}(C) + \nu \cdot \text{Area}(\text{inside}(C)) \\
 & + \lambda_1 \int_{\text{inside}(C)} |u_0(x, y) - c_1|^2 dx dy \\
 & + \lambda_2 \int_{\text{outside}(C)} |u_0(x, y) - c_2|^2 dx dy,
 \end{aligned} \tag{4.7}$$

onde $\mu, \nu \geq 0, \lambda_1 = \lambda_2 > 0$ são parâmetros fixos. Para a resolução do problema da minimização das partições, os autores propõem que sejam usados os seguintes valores nos parâmetros: $\nu = 0, \lambda_1, \lambda_2 = 1$.

A motivação principal para o uso deste tipo de abordagem é o facto de possuir uma boa capacidade de segmentação de objectos, visto ser um método de segmentação genérico, ao contrário de muitos outros métodos que são desenvolvidos apenas para um determinado tipo de objecto.

Em [43] os autores referem que este método possui uma boa precisão e robustez à custa de um aumento significativo do tempo de processamento. Visto este método ser iterativo, é possível impor restrições de forma a reduzir o seu tempo de processamento. Neste trabalho a restrição imposta foi a limitação no número máximo de iterações. Como já vimos, inicialmente, o critério de paragem deste método era dado pelo $\inf_C \{F_1(C) + F_2(C)\} \approx 0$.

O nosso trabalho, o número de iterações máximas foi de 200, uma vez que este valor obteve bons resultados nas experiências realizadas no conjunto de treino. Nestas experiências existiu uma redução significativa do tempo de processamento sem existirem perdas significativas ao nível da precisão. Como iremos ver mais à frente, a utilização desta abordagem não será mais lenta do que os outros métodos apresentados em [35, 69, 24, 25, 22, 26]. O tempo de processamento deste método dependerá muito do tipo de contorno inicial e da sua localização, sendo que quanto mais afastado da face estiver o contorno inicial mais este irá demorar a convergir.

Este algoritmo foi usado por *Shah et al.* [77] na segmentação dos dentes em imagens de raio-X. Estes autores referem que obtiveram bons resultados na segmentação dos dentes usando este método. Com isto e pela observação das propriedades destas imagens (possuem algumas similaridades com as imagens LWIR), poderemos pressupor que também iremos conseguir obter bons resultados usando este método.

No primeiro teste realizado com este método nas imagens da faces LWIR usámos como contorno inicial um rectângulo centrado na imagem. O tamanho deste rectângulo foi de 90×140 pixéis, onde este foi obtido através de estudo ao tamanho aproximado das faces do conjunto de treino. Mais à frente, apresentaremos outros métodos que usam um tipo diferente de contorno inicial (uma elipse). Como foi referido anteriormente, o contorno inicial poderá afectar o tempo de processamento, por isso é importante escolher um contorno inicial que possua a forma da face.

4.1.5 Identificação dos Pixéis da Face numa Imagem Binária (IPFIB)

Nesta sub-secção, propomos um método para efectuar a selecção dos pixéis da face em imagens binárias (figura 4.4(a)) que resultam da aplicação de outros métodos, tais como [69, 30]. Este método identifica o maior contorno que contenha o ponto central da face, considerando que todos os pixéis no seu interior são pixéis da face com a excepção dos pixéis que pertencem aos óculos (figura 4.4).

O primeiro passo é identificar o ponto central da face, como explicado na operação RIR (cruz vermelha na figura 4.4(a)). Depois disto, aplicamos uma dilatação e uma erosão (uma abertura) usando um elemento estruturante com dimensão de 3×3 e 2×2 , respectivamente. Estas operações morfológicas são usadas para remover pequenas áreas. Com base no resultado das operações morfológicas, extraímos o mapa de arestas usando o detector de arestas de Canny [28]. As arestas obtidas são realçadas usando

uma dilatação com um elemento estruturante com dimensão 3×3 (figura 4.4(b)). A partir destas arestas, seleccionamos a maior que contenha no seu interior o centro da face. Obtendo esse contorno, assumimos que todos os pixéis que estão no seu interior são pixéis da face (figura 4.4(c)).

Para remover os óculos que tenham sido considerados como face no passo anterior, aplicamos a diferença absoluta entre a imagens antes da selecção do maior contorno e imagem resultante do preenchimento do maior contorno (figura 4.4(d)). Com esta diferença, obtemos uma imagem com as regiões que foram alteradas pelo preenchimento. Nesse resultado aplicamos uma abertura usando um elemento estruturante circular com um raio de 10 pixéis (figura 4.4(e)). Usando um elemento estruturante com esta dimensão apenas as maiores regiões, como os óculos, permanecem após a aplicação desta operação morfológica. A imagem resultante é somada à imagem em que o maior contorno foi preenchido e os valores que excedem o maior valor possível (1 ou 255 dependendo da representação usada para o branco na imagem) é ajustado para zero (é considerado como não fazendo parte da face) (figura 4.4(f)).



(a) Exemplo de uma imagem de entrada binária usada nesta operação.



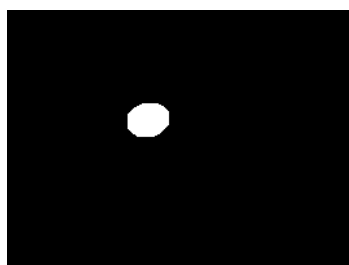
(b) Realce do contorno da imagem binária.



(c) Preenchimento do maior contorno que contém o centro da face.



(d) Diferença absoluta entre a imagem da figura 4.4(a) e 4.4(c).



(e) Resultado da operação morfológica (abertura) com o elemento estruturante circular.



(f) Resultado desta operação, depois de realizada a soma.

Figura 4.4: Exemplo da aplicação da operação IPFIB.

4.2 Métodos de Segmentação

Nesta secção, apresentamos os métodos de segmentação de faces usando imagens no infravermelho térmico que combinam as operações apresentadas anteriormente. O métodos desenvolvidos foram denominados por **A**, **B**, ..., **H**.

O método **A** consiste em considerar os pixéis da face que estejam dentro da região definida pela operação RIR e que sejam considerados como pixéis da face pela aplicação do limiar (*threshold*), operação definida na secção 4.1.3. A figura 4.5(a) apresenta o diagrama de blocos deste método. A aplicação da operação RIR é usada para localizar a face e consideramos todos os pixéis que estejam dentro dessa região como sendo pixéis da face. O limiar define, como pixéis da face, os pixéis com maiores intensidades. A aplicação do operador *AND* nestas duas imagens retorna os pixéis com as maiores intensidades dentro da região definida pela RIR.

O método **B** é similar ao método **A**, com excepção que em vez da operação RIR é usada a operação Região de Interesse Elíptica (RIE), como apresentado na figura 4.5(b).

O método **C** inicia-se com a aplicação da operação RIR e depois é aplicado o método descrito em [69] apenas na região definida pela RIR (figura 4.5(c)). Para o método **D** é aplicada a Identificação dos Pixéis da Face numa Imagem Binária (IPFIB) depois de se ter realizado o método **C** (figura 4.5(d)).

O método **E** é iniciado com corte da região definida pela operação RIR (figura 4.5(e)). Nessa região é aplicada a operação RIE e a elipse obtida irá ser usada para a inicialização do contorno activo do método [30] que irá terminar a segmentação. O método **F** é uma extensão do método **E** (figura 4.5(f)). Esta extensão consiste na aplicação da operação IPFIB depois da aplicação do método **E**.

O método **G** é similar ao método **E**, mas não é realizado o recorte da região definida pela operação RIR (figura 4.5(g)). A diferença entre o método **G** e **H** é a mesma que é verificada entre os método **E** e **F** (figura 4.5(h)).

4.3 Fusões

Também consideramos realizar a fusão de alguns métodos (os que são mais distintos, de modo a os seus erros estejam presentes em partes diferentes das imagens). Os

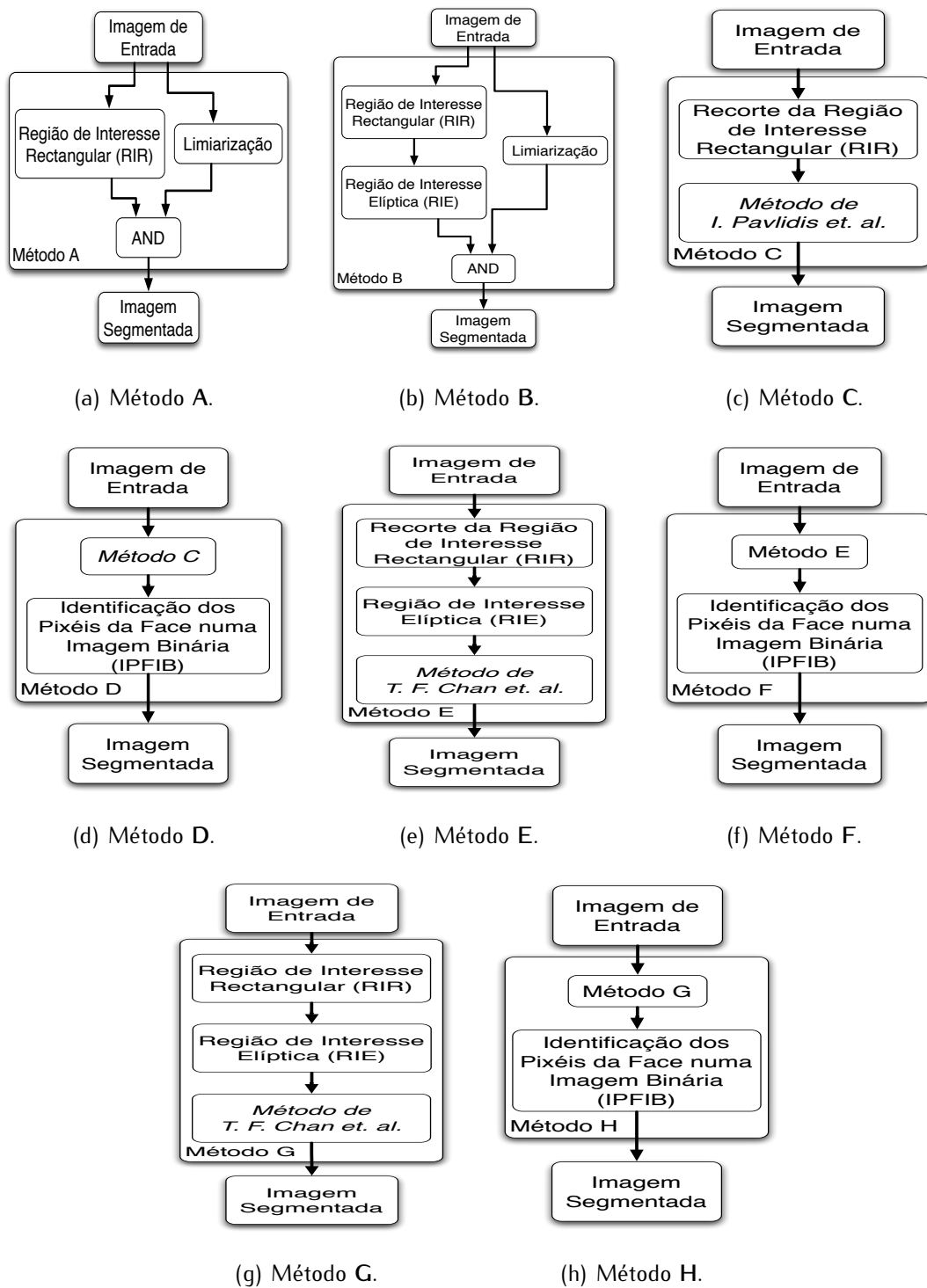


Figura 4.5: Diagrama de blocos dos métodos de segmentação baseados nas operações apresentadas anteriormente.

métodos que foram fundidos são os métodos A, F e H. As fusões usadas são:

1. **Fusão usando a Votação por Maioria:** votação por maioria dos resultados dos métodos A, F and H;
2. **Fusão usando o operador lógico AND:** intersecção dos resultados:

$$FusãoAND(iC, iR) = A(iC, iR) \cap F(iC, iR) \cap H(iC, iR) \quad (4.8)$$

onde iC e iR são os índices das colunas e linhas, e $A(iC, iR)$, $F(iC, iR)$ e $H(iC, iR)$ são os resultados dos métodos A, F and H na coluna iC e linha iR .

3. **Fusão usando o operador lógico OR:** união dos resultados:

$$FusãoOR(iC, iR) = A(iC, iR) \cup F(iC, iR) \cup H(iC, iR). \quad (4.9)$$

4.4 Conclusão

Neste capítulo apresentámos 8 métodos de segmentação de faces em imagens LWIR que na nossa opinião podem vir a aumentar as taxas de reconhecimento usando este tipo de imagens. Além dos 8 métodos ainda apresentámos três abordagens de fusão dos resultados obtidos por três métodos que considerámos possuírem diferenças de modo que a fusão dos seus resultados possa melhorar a segmentação facial em imagens do infravermelho térmico.

Capítulo 5

Resultados e Discussão

5.1 Segmentação Manual

Os conjuntos de teste de todas as bases de dados foram segmentados manualmente, através dum conjunto de máscaras que possuem apenas as regiões de interesse (face). O método [35] não necessita de conjunto de treino e o método [69] usa pixéis das regiões da face e fundo das imagens do conjunto de treino (recortadas manualmente), não sendo assim necessário segmentar as imagens do conjunto de treino.

A tabela 5.1 apresenta a percentagem de pixéis face e fundo presentes no conjunto de teste das bases de dados usadas neste trabalho. Estes valores são obtidos através das imagens segmentadas manualmente. Nesta tabela podemos observar que a única base de dados que possui uma maior percentagem de pixéis da face do que não face é a base de dados FSU. As outras três bases de dados possuem um número superior de pixéis não face, sendo que as percentagens de pixéis da face variam entre 15.40% e os 32.65%.

Os nomes das imagens do conjunto de treino e teste para as diferentes bases de dados usadas neste trabalho estão disponíveis em http://www.di.ubi.pt/~lfbaa/thermal_face.html. Esta listagem irá possibilitar que qualquer pessoa possa replicar os nossos resultados.

	Percentagem de Pixéis (%)	
	Face	Não-Face
UND	15.40	84.60
Terravic	21.11	78.89
IRIS	32.65	67.35
FSU	57.22	42.78

Tabela 5.1: percentagens de pixéis da face fundo em relação ao número total de pixéis das imagens, para as diferentes bases de dados.

5.2 Medidas de Erro

A tarefa de avaliação é bastante simples, sendo que para cada imagem de entrada a sua correspondente imagem segmentada deve ser criada, onde os pixéis pertencentes à face e livres de ruído devem de aparecer a branco, enquanto que os restantes devem de aparecer a preto. O conjunto de teste das bases de dados usada para a medição dos pixéis em concordância entre os mapas binários produzidos pelos algoritmos $O = O_1, \dots, O_n$ e os mapas binários produzidos manualmente $C = C_1, \dots, C_n$.

A taxa de erro (E_1) de um algoritmo é dada pela proporção de pixéis em desacordo, através da disjunção exclusiva (operador lógico ou-exclusivo, XOR), em todas as imagens, dada por:

$$E_1 = \frac{1}{nC \times nR} \sum_{c'} \sum_{r'} O(c', r') \oplus C(c', r') \quad (5.1)$$

onde $O(c', r')$ e $C(c', r')$ são, respectivamente, os pixéis das imagens segmentadas produzidas pelos algoritmos e pelas imagens segmentadas manualmente.

A segunda medida de erro destina-se a compensar a desigualdade entre as probabilidades de pixéis face e não face nas imagens. A taxa de erro E_2 é dada pela média entre a taxa de falsos positivos (*False Positive Rate* (FPR)) e a taxa de falsos negativos (*False Negative Rate* (FNR)):

$$E_2 = 0.5 \times FNR + 0.5 \times FPR \quad (5.2)$$

5.3 Resultados Experimentais e Discussão dos Resultados

Nesta sub-seccção iremos apresentar os resultados da segmentação para os vários métodos de segmentação de faces em imagens LWIR apresentados neste trabalho.

Os gráficos das figuras 5.1 e 5.2 apresentam as taxas de erro para as medidas E_1 (figura 5.1) e E_2 (figura 5.2), que também podem ser observados nas tabelas 5.2, 5.3, 5.4 e 5.5. Ao contrário das tabelas, estes permitem uma visualização rápida da relação entre as nossas diferentes abordagens propostas e as abordagens dos outros autores, para as quatro bases de dados.

É possível que não exista um método único que seja melhor em todas as bases de dados. Isso resulta do facto de que as imagens possuem diferentes características.

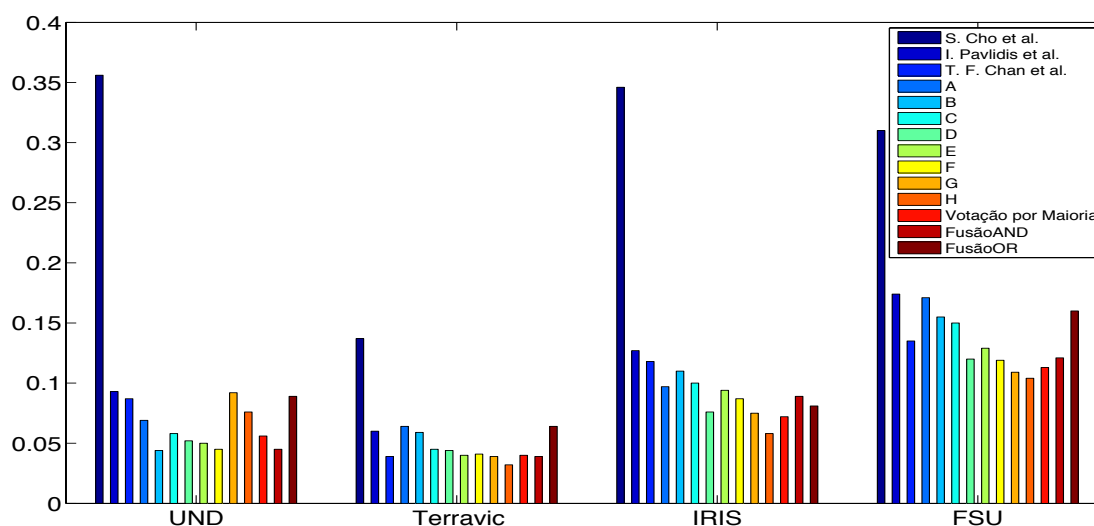


Figura 5.1: Representações gráficas da medida de erro E_1 para os resultados apresentados nas tabelas 5.2, 5.3, 5.4 e 5.5.

A tabela 5.2 contém as taxas de erro de todos os métodos de segmentação apresentados usando as imagens da base de dados UND. Nesta tabela, também são apresentados os tempos médios por imagem. Nesta base de dados, o método que obteve melhores resultados segundo a medida de erro E_1 (método B) não é o mesmo que obteve o melhor resultado segundo a medida de erro E_2 (método A). Isto deve-se a que a medida E_2 favorece os métodos em que as FNR e FPR sejam equilibradas, enquanto que a medida de erro E_1 apenas conta o número de pixels mal classificados. A melhoria existente entre os nosso métodos e os métodos dos outros autores é de

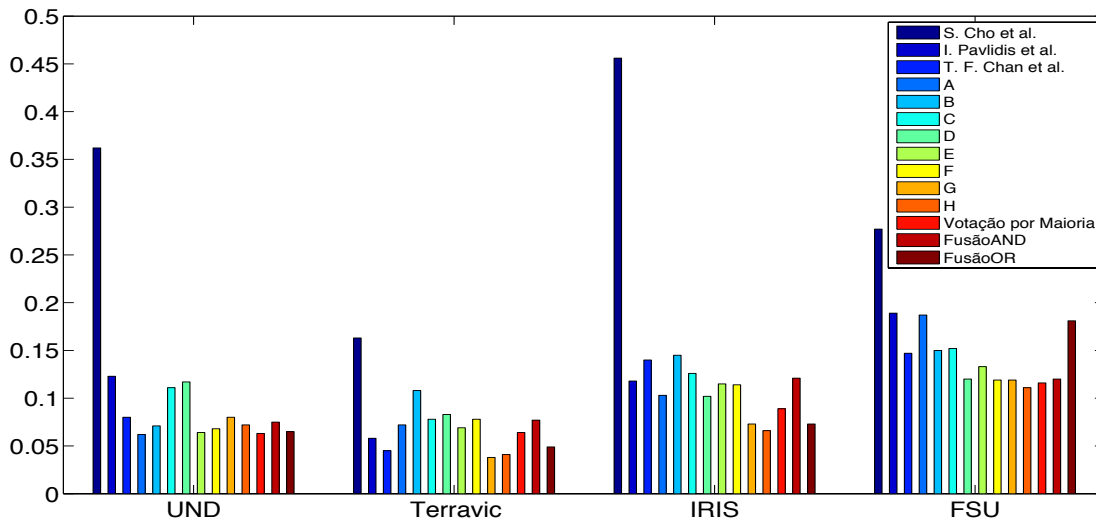


Figura 5.2: Representações gráficas da medida de erro E_2 para os resultados apresentados nas tabelas 5.2, 5.3, 5.4 e 5.5.

31.2% para a medida E_1 e 30.0% para a E_2 .

A tabela 5.3 apresenta as taxas de erro e tempos de processamento na base de dados Terravic. Mais uma vez, o melhor algoritmo segundo a medida de erro E_1 (método H) não é o mesmo que o melhor método segundo a medida E_2 (G). No entanto, a melhoria existente entre os nossos métodos e os métodos dos outros autores é de 10.5% para a medida E_1 e 12.5% para a E_2 .

A tabela 5.4 contém os resultados da segmentação para todos os métodos apresentados usando imagens da base de dados IRIS. Neste caso, os melhores resultados obtidos para as duas medidas de erro foram obtidos pelo mesmo algoritmo (método H). A melhoria entre as abordagens publicadas anteriormente e as nossas foi de 28.8% para a medida E_1 e 39.0% para a E_2 .

Os resultados para a base de dados FSU são apresentados na tabela 5.5. Como acontece na base de dados IRIS, o método H é o que possui melhores resultados em ambas as medidas de erro. A melhoria entre as abordagens publicadas anteriormente e as nossas foi de 20.6% para a medida E_1 e 16.6% para a E_2 .

A partir dos resultados apresentados nas tabelas 5.2, 5.3, 5.4 and 5.5 é visível que de acordo com a medida de erro E_1 , o método H é melhor que os outros em três das quatro bases de dados. Usando a medida de erro E_2 o método H só é melhor em duas das bases de dados.

Os tempos de processamento apresentados foram obtidos em código desenvolvido em

	UND				Time(s)
	FNR	FPR	E_1	E_2	
<i>S. Cho et al.</i> [35]	0.370	0.354	0.356	0.362	1.236
<i>I. Pavlidis et al.</i> [69]	0.166	0.080	0.093	0.123	4.480
<i>T. F. Chan et al.</i> [30]	0.070	0.090	0.087	0.080	9.278
A	0.051	0.073	0.069	0.062	0.012
B	0.110	0.032	0.044	0.071	0.089
C	0.188	0.034	0.058	0.111	1.100
D	0.212	0.023	0.052	0.117	1.294
E	0.085	0.043	0.050	0.064	2.754
F	0.103	0.034	0.045	0.068	2.912
G	0.062	0.097	0.092	0.080	6.808
H	0.067	0.077	0.076	0.072	6.994
Votação por Maioria	0.072	0.054	0.056	0.063	F + H
FusãoAND	0.119	0.031	0.045	0.075	F + H
FusãoOR	0.031	0.099	0.089	0.065	F + H

Tabela 5.2: Experiment results in the UND database.

Matlab (versão R2009b) num *Intel Core 2 Q9300* (2.5GHz), 4Gb RAM (FSB 1066) com sistema operativo *Fedora Core 11*. Observamos que, como esperado, todos os métodos que são baseados no recorte da região definida pela operação RIR (métodos C, D, E and F) são mais rápidos que as propostas originais [69, 24, 25, 22, 26, 30] (figura 5.3).

Observando o método apresentado em [69, 24, 25, 22, 26] e as suas variantes (métodos C e D), a melhoria (em todas as bases de dados), em média, é de 55.7%, enquanto que para a base de dados UND a melhoria foi de 75.4%.

A melhoria no tempo de execução deste método é devida ao uso do recorte, tal que a maioria das roupas são removidas, sendo este um dos problemas identificados no método apresentado em [69, 24, 25, 22, 26].

	Terravic				Time(s)
	FNR	FPR	E_1	E_2	
<i>S. Cho et al.</i> [35]	0.207	0.118	0.137	0.163	1.237
<i>I. Pavlidis et al.</i> [69]	0.055	0.062	0.060	0.058	4.554
<i>T. F. Chan et al.</i> [30]	0.057	0.034	0.039	0.045	10.361
A	0.087	0.058	0.064	0.072	0.012
B	0.194	0.023	0.059	0.108	0.091
C	0.134	0.022	0.045	0.078	1.174
D	0.152	0.015	0.044	0.083	1.330
E	0.119	0.019	0.040	0.069	3.071
F	0.141	0.014	0.041	0.078	3.233
G	0.036	0.040	0.039	0.038	7.115
H	0.055	0.026	0.032	0.041	7.227
Votação por Maioria	0.106	0.022	0.040	0.064	F + H
FusãoAND	0.143	0.012	0.039	0.077	F + H
FusãoOR	0.034	0.064	0.058	0.049	F + H

Tabela 5.3: Experiment results in the Terravic database.

Para o método proposto em [30] e as suas variantes (métodos E e F), a melhoria média foi superior a 62.9% para todas as bases de dados e para a base de dados Terravic esta melhoria chegou aos 70.4%. Para as outras duas variantes (métodos G and H), a melhoria média foi de 28.6%, tendo chegado a 31.3% na base de dados Terravic.

A melhoria presente em todas as variantes do método [30] (métodos E a H) é devido à posição e forma (elipse) do contorno activo inicial. O método [30] usa um contorno inicial centrado na imagem, enquanto que nem todas as imagens estão centradas. Daí a aplicação da operação RIR (para procurar a localização da face e o seu tamanho) antes de usar este método.

	IRIS				Time(s)
	FNR	FPR	E_1	E_2	
<i>S. Cho et al.</i> [35]	0.773	0.139	0.346	0.456	0.959
<i>I. Pavlidis et al.</i> [69]	0.093	0.144	0.127	0.118	3.233
<i>T. F. Chan et al.</i> [30]	0.204	0.077	0.118	0.140	8.717
A	0.122	0.085	0.097	0.103	0.011
B	0.248	0.043	0.110	0.145	0.081
C	0.199	0.052	0.100	0.126	1.277
D	0.175	0.028	0.076	0.102	1.299
E	0.177	0.053	0.094	0.115	3.491
F	0.194	0.034	0.087	0.114	3.631
G	0.068	0.078	0.075	0.073	7.025
H	0.086	0.045	0.058	0.066	7.039
Votação por Maioria	0.139	0.039	0.072	0.089	F + H
FusãoAND	0.214	0.029	0.089	0.121	F + H
FusãoOR	0.048	0.097	0.081	0.073	F + H

Tabela 5.4: Experiment results in the IRIS database.

Em relação aos métodos de fusão, podemos concluir que não apresentaram benefícios, especialmente a nível do tempo de processamento. Estes métodos apresentaram apenas melhorias a nível da redução dos FNRs na fusão usando o operador lógico *OR* e na redução dos FPRs na fusão baseada no operador lógico *AND* (com excepção da base de dados FSU onde o método apresentado em [35] obteve uma FPR mais baixo).

Os tempos de execução das fusões foram apresentados como sendo a soma dos tempos de execução dos métodos **F** e **H**, e não levámos em conta o tempo do método **A**, porque este método é muito rápido e parte das suas operações já são realizadas para obter os outros dois métodos.

Nas figuras 5.4, 5.5, 5.6 e 5.7 são apresentadas algumas das imagens originais (pri-

	FSU				Time(s)
	FNR	FPR	E_1	E_2	
<i>S. Cho et al.</i> [35]	0.506	0.048	0.310	0.277	0.379
<i>I. Pavlidis et al.</i> [69]	0.089	0.289	0.174	0.189	1.296
<i>T. F. Chan et al.</i> [30]	0.058	0.237	0.135	0.147	6.039
A	0.73	0.302	0.171	0.187	0.009
B	0.182	0.118	0.155	0.150	0.078
C	0.141	0.163	0.150	0.152	1.066
D	0.122	0.117	0.120	0.120	1.091
E	0.106	0.160	0.129	0.133	2.759
F	0.119	0.120	0.119	0.119	2.850
G	0.052	0.186	0.109	0.119	3.677
H	0.062	0.159	0.104	0.111	3.746
Votação por Maioria	0.095	0.138	0.113	0.116	F + H
FusãoAND	0.125	0.115	0.121	0.120	F + H
FusãoOR	0.034	0.328	0.160	0.181	F + H

Tabela 5.5: Experiment results in the FSU database.

meira linha), respectivas imagens segmentadas manualmente (segunda linha). Nestas figuras, para cada imagem original (primeira linha de cada figura) o resultado da segmentação é apresentado em duas colunas, sendo a primeira coluna os mapas da segmentação obtidos pelos métodos de segmentação avaliados e a segunda coluna é a aplicação de cada um dos mapas à imagem original.

Nas figuras 5.4 e 5.6 as duas primeiras imagens originais são imagens do conjunto de teste da base de dados UND e as duas últimas colunas da base de dados Terravic. Nas figuras 5.5 e 5.7, a separação é a mesma, mas nestas figuras são apresentadas as imagens das bases de dados IRIS e FSU. Em relação à base de dados FSU, os testes realizados foram realizados usando apenas o canal vermelho das imagens. Esta escolha

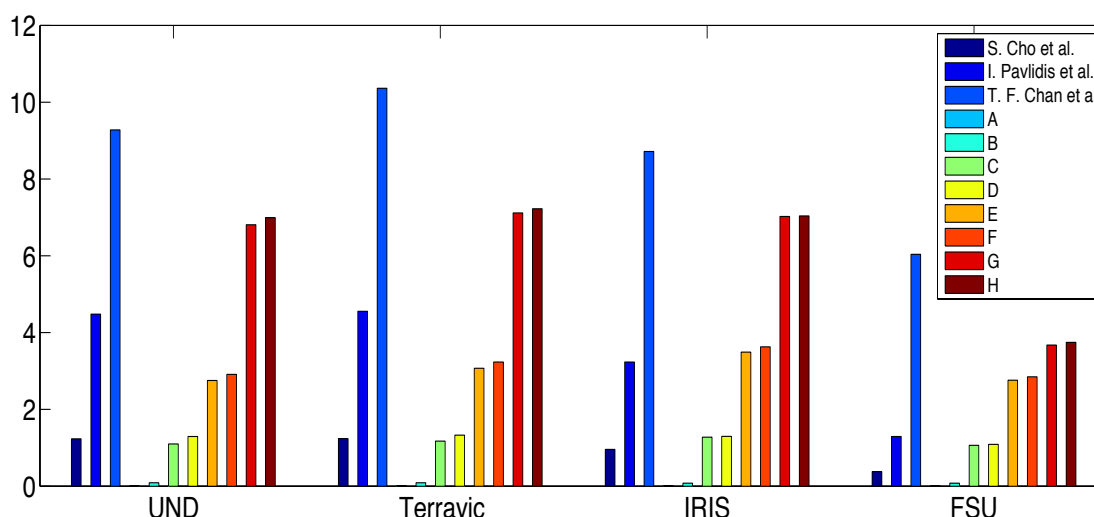


Figura 5.3: Representação gráfica dos tempos de execução apresentados nas tabelas 5.2, 5.3, 5.4 e 5.5.

deve-se ao facto de termos realizados vários testes no início desta dissertação usando os vários canais, como apresentamos no anexo A. Nos resultados apresentados neste anexo podemos ver que o canal que obteve melhores resultados foi o canal *Red* (R).

Nas figuras 5.4 e 5.5, as linhas 3 à 13 apresentam as imagens segmentadas pelos métodos apresentados anteriormente. Nestas figuras, a terceira linha representa a segmentação usando o método apresentado por *Cho et al.* [35]. A quarta linha apresenta as imagens segmentadas pelo método de *Pavlidis et al.* e *Buddharaju et al.* [69, 24, 25, 22, 26]. As restantes linhas (linha 5 até à linha 13) apresentam os resultados da segmentação usando os métodos que nós propomos, isto é, a linha 5 representa os resultados da segmentação usando o método A, linha 6 do método B, ... e a linha 13 o método H.

Nas figuras 5.6 e 5.7 são apresentados os resultados da segmentação usando as fusões apresentadas, sendo que a figura 5.6 contém uma representação das imagens das bases de dados UND (quatro primeiras colunas) e Terravic (quatro últimas colunas), enquanto que a figura 5.7 representa as imagens das bases de dados IRIS (quatro primeiras colunas) e FSU (quatro últimas colunas). Na terceira linha destas figuras, a segmentação é realizada através da votação por maioria entre os resultados dos métodos A, F e H. A quarta linha representa os resultados obtidos através da intersecção dos resultados dos métodos A, F e H (equação 4.8). A última linha apresenta as imagens segmentadas pela união dos resultados dos métodos A, F e H (equação

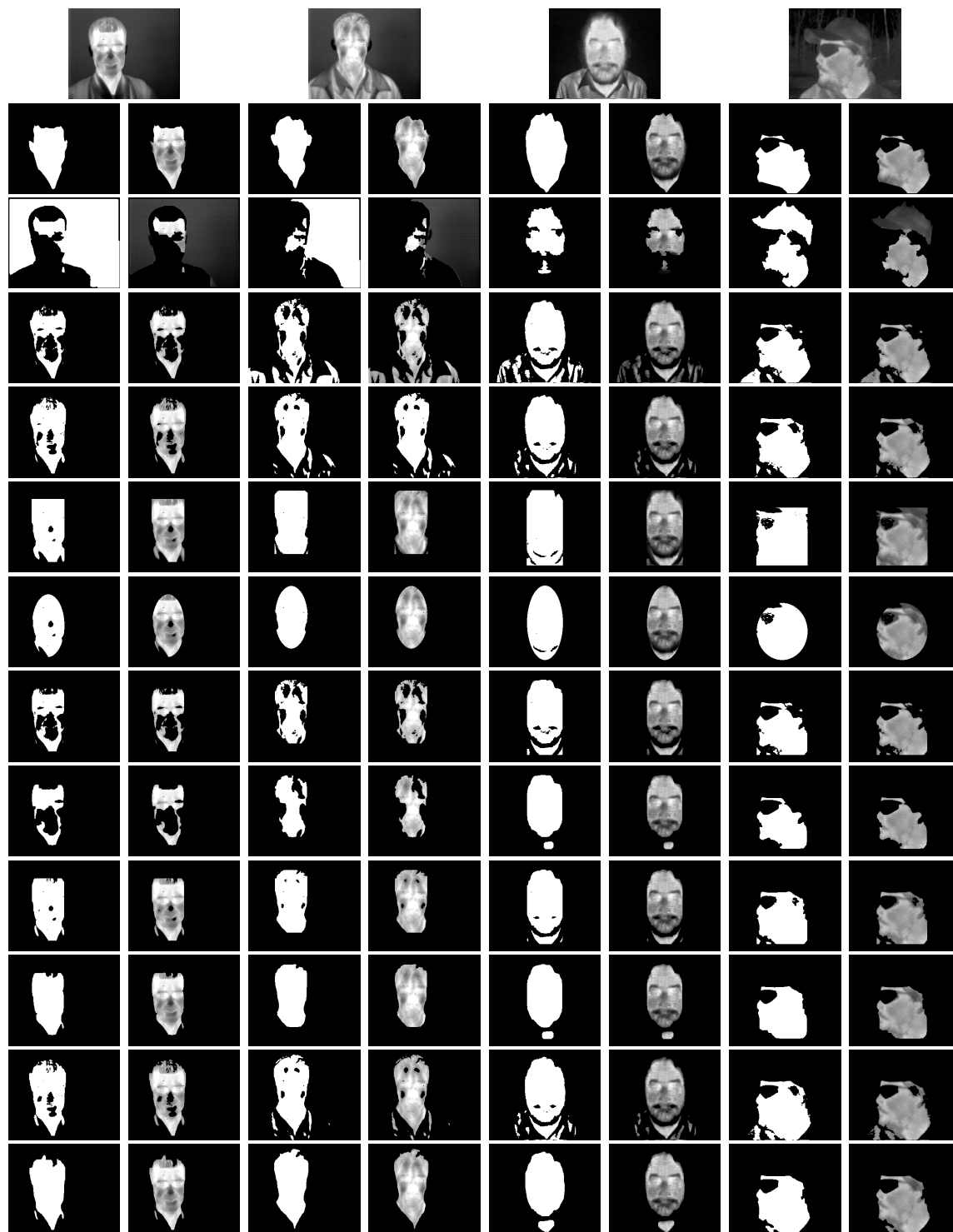


Figura 5.4: Resultados da segmentação para os métodos apresentados em [35, 69, 24, 25, 22, 26] e pelos nossos (A, B, ..., H), para as imagens das bases de dados UND (quatro primeiras colunas) e Terravic (quatro últimas colunas).

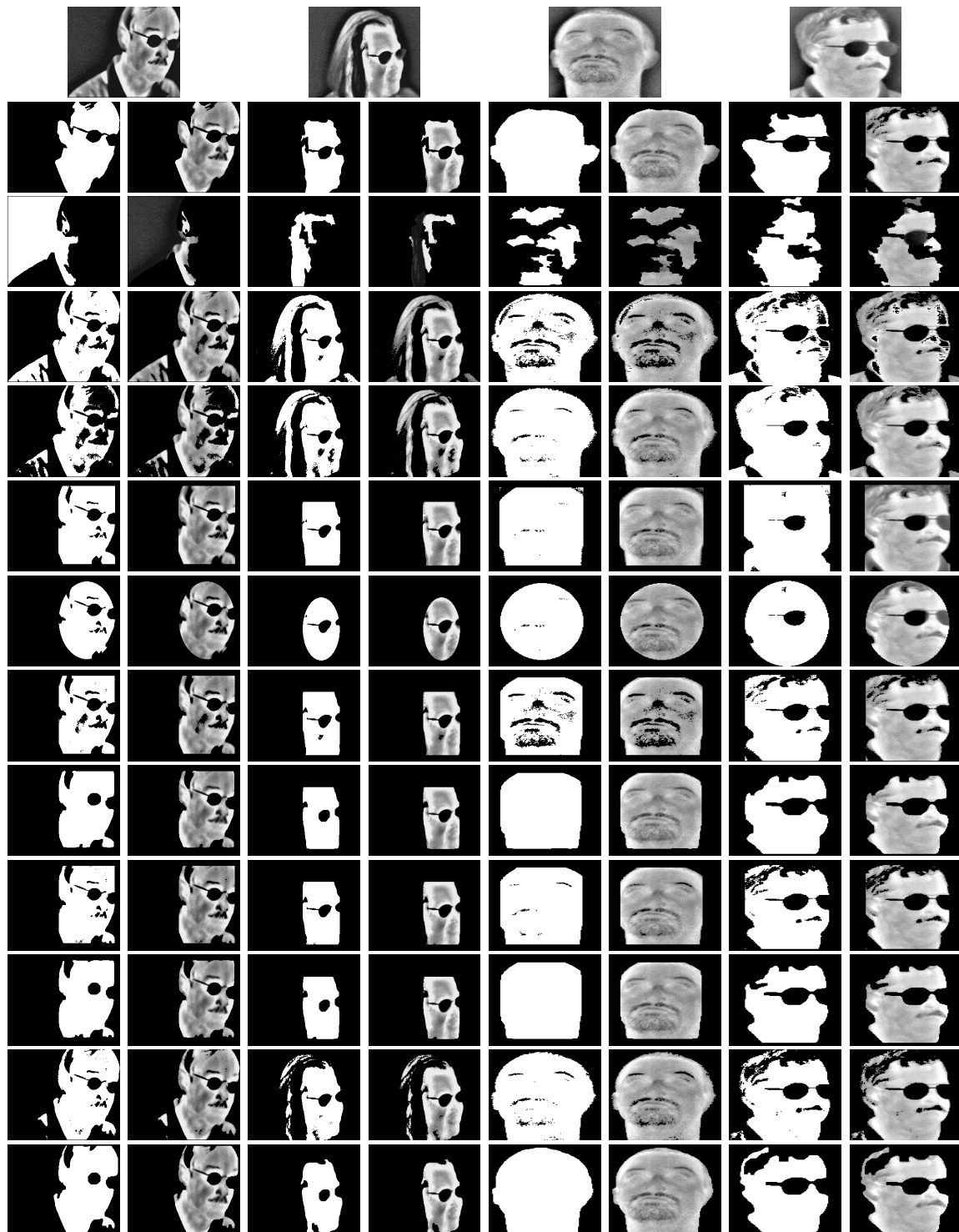


Figura 5.5: Resultados da segmentação para os métodos apresentados em [35, 69, 24, 25, 22, 26] e pelos nossos (A, B, ..., H), para as imagens das bases de dados IRIS (quatro primeiras colunas) e FSU (quatro últimas colunas).

4.9).

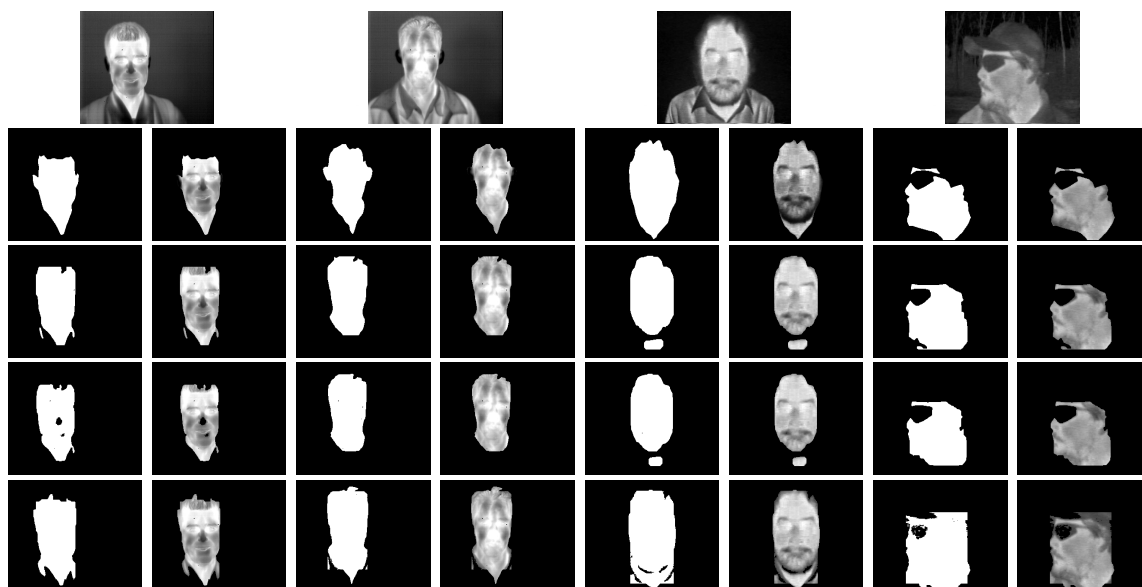


Figura 5.6: Resultados da segmentação para as fusões apresentadas, para as imagens das bases de dados UND (quatro primeiras colunas) e Terravic (quatro últimas colunas).

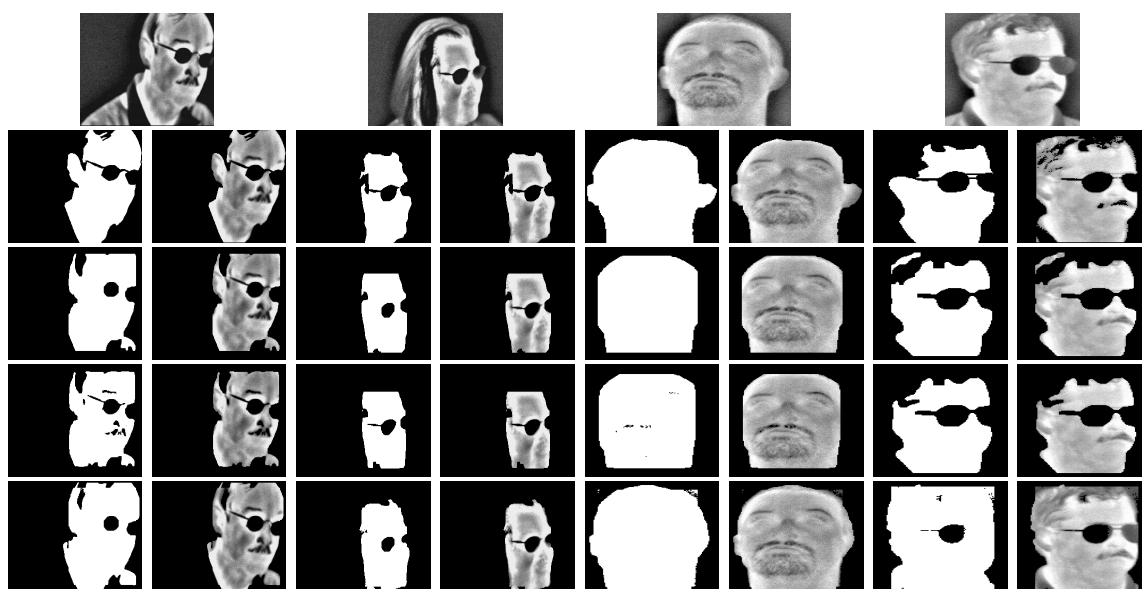


Figura 5.7: Resultados da segmentação para as fusões apresentadas, para as imagens das bases de dados IRIS (quatro primeiras colunas) e FSU (quatro últimas colunas).

5.4 Conclusão

Consideramos que o melhor método para a segmentação de faces em imagens LWIR é o método **H** uma vez que obteve um melhor desempenho em mais bases de dados. Olhando apenas para a medida de erro E_1 , o método **H** apenas não foi o melhor método na base de dados UND. Nesta base de dados o melhor método foi o método **B**, sobretudo devido a que o FPR é mais baixo, visto que apenas considera como face os pixels que estejam no interior da região definida pela elipse.

Se observarmos a medida de erro E_2 , o método **H** é o melhor em duas das quatro bases de dados (IRIS e FSU), sendo o segundo melhor na base de dados Terravic (com uma diferença marginal de 0.003 em relação ao método **G**). Na base de dados UND este método não possui um bom desempenho, como já tinha acontecido na medida de erro E_1 .

A razão do fraco desempenho do método **H** na base de dados UND é que não restringe o processamento a uma pequena área da imagem (região da face), como os métodos **A** a **F** devido ao recorte da região definida pela operação RIR. Essa restrição faz com que estes métodos possuam um FPR baixo, ficando assim beneficiados em relação aos outros métodos. Note-se que as imagens da base de dados UND contêm grandes regiões com roupas, sendo que em várias imagens essas regiões possuem uma área superior à área da região da face.

Com base nestes resultados, fica claro que é possível reduzir o tempo de processamento e, simultaneamente, aumentar a precisão na segmentação da face em imagens LWIR. Note-se que estas imagens apresentam desafios específicos, tais como temperaturas das faces similares às do fundo, como é comum no nariz ou as orelhas e temperatura das roupas semelhante à da face.

Capítulo 6

Conclusão

6.1 Resumo do Trabalho desenvolvido

Este trabalho apresenta um estudo sobre vários métodos de segmentação da face em imagens capturadas no infravermelho térmico. O estudo realizado baseia-se em propostas realizadas por outros autores e nas nossas novas propostas. O objectivo foi desenvolver um método que possuísse um bom desempenho independentemente da pose, rotação ou expressão facial. O desempenho dos métodos apresentados foi medido recorrendo a duas medidas de erro, para destacar tanto o erro global (E_1) como o erro balanceado (E_2) de forma a equilibrar a avaliação, visto que as imagens não possuem o mesmo número de pixels face e não face. As medições foram feitas em quatro bases de dados públicas disponíveis (incluindo 730 imagens de treino e 893 imagens de teste).

Os métodos propostos melhoram os métodos já existentes, com o método **H** a ser a nossa escolha. Ele apresentou os melhores resultados de acordo com a medida de erro E_1 em três das quatro bases de dados e para a medida de erro E_2 apresentou melhores resultados em duas das quatro bases de dados (e numa terceira base de dados foi segundo por uma diferença muito pequena). Além disso, este método é melhor que os métodos existentes [69, 24, 25, 22, 26, 35] em todas as bases de dados e em ambas as medidas de erro, com melhorias até 28.8% segundo a medida de erro E_1 e 38.3% para a E_2 .

Algumas das outras propostas melhoram o tempo de processamento drasticamente (método **A** e **B**) e ainda conseguem bater os resultados dos métodos existentes, ou são apenas marginalmente piores. Na nossa opinião estes métodos podem ser usados

quando existe uma limitação em termos de tempo de processamento, tais como em sistemas em tempo real.

Acreditamos, assim, que o trabalho apresentado é uma importante contribuição para o estado da arte em segmentação de faces em imagens LWIR e que permitirá ajudar a melhorar os métodos de reconhecimento facial em imagens no infravermelho térmico.

Os resultados dos 8 métodos de segmentação desenvolvidos, foram submetidos em artigo para as *Transactions on Pattern Analysis and Machine Intelligence* (TPAMI) [7] (anexo B). Um outro artigo foi submetido para a *15th Iberoamerican Congress on Pattern Recognition* (CIARP) [1] (anexo A), sendo que tanto os métodos como os resultados não foram apresentados nesta dissertação. Este último artigo foi o primeiro estudo que realizámos no início desta dissertação.

6.2 Trabalho Futuro

Futuramente será desenvolvido um novo estudo para que possamos quantificar a relação entre a performance da segmentação e a do reconhecimento. Este estudo passa pela aplicação de métodos de reconhecimento a cada um dos métodos de segmentação e depois avaliaremos se o método que obteve a menor taxa de erro na segmentação também irá obter a melhor taxa de reconhecimento. Um exemplo de um método que é aplicado no reconhecimento em imagens LWIR é o método de extracção das TMPs apresentado por *P. Buddharaju et al.* em [69, 24, 25, 22, 26].

Apêndice A

Improving Face Segmentation in Thermograms using Image Signatures

Neste anexo inclui-se um artigo proposto para o CIARP [1]. De notar que tanto o método apresentado neste artigo como os resultados não são apresentados nesta dissertação.

Improving Face Segmentation in Thermograms using Image Signatures

Silvio Filipe and Luís A. Alexandre

Department of Computer Science
IT - Instituto de Telecomunicações
SOCIA - Soft Computing and Image Analysis Group
University of Beira Interior, 6200-Covilhã, Portugal.
m2213@ubi.pt (S. Filipe) lfbaa@di.ubi.pt (L. A. Alexandre)

Abstract. The aim of this paper is to present a method for the automatic segmentation of face images captured in Long Wavelength Infrared (LWIR), allowing for a large range of face rotations and expressions. The motivation behind this effort is to enable better performance of face recognition methods in the thermal Infrared (IR) images. The proposed method consists on the modelling of background and face pixels by two normal distributions each, followed by a post-processing step of face dilation for closing holes and delimitation based on vertical and horizontal images signatures. Our experiments were performed on images of the University of Notre Dame (UND) and Florida State University (FSU) databases. The obtained results improve on previous existing methods by 2.8% to more than 25% depending on the method and database.

Keywords: Face Segmentation, Human Skin Segmentation, Image segmentation, Infrared Thermal.

1 INTRODUCTION

A large amount of research has been conducted in the field of face recognition, mainly in the visible spectrum. These systems have problems dealing with light variations [6]. Some of the proposed solutions use 3D facial recognition [1] and combine face recognition in both visible and IR spectrum [7].

The growing interest in robust methods (for example, for security applications) has driven the development of facial recognition exclusively in the infrared. Recognition in the LWIR is not affected by light variations.

A crucial step in the process of face recognition is the face segmentation. This is more demanding than simple face detection since it pinpoints not only the face's locations, but also must describe its shape. A robust segmentation system can improve recognition rates regardless of the recognition method.

In contrast with the visible wavelength, where numerous methods have been proposed to accomplish this task (based on color, geometry, etc.), in the LWIR there is a lack of proposals to improve the current status.

In the next sections we present a short description of two available face segmentation methods (section 2) and present our face segmentation method (section 3). In Section 4, we present the datasets used and experimental results, including a small discussion. We end the paper in section 5 with the conclusions.

2 OVERVIEW OF FACE SEGMENTATION IN THERMAL INFRARED IMAGES

Face segmentation, given that it is a preprocessing step for all recognition methods, will lead to their failure if it is not correctly performed. This is not a subject much discussed by the authors of recognition methods in the infrared. Some of the proposed approaches are based only on the creation of an elliptical mask that will be put over the image of the face [5], but these approaches will work only on frontal and centered faces.

Siu-Yeung Cho et al. in [3] present a method for face segmentation in IR images based on the Sobel Edge detector and morphological operations. After the Sobel Edge detector, the largest contour is considered to be the one best describing the face. They apply the morphological operations to the area contained in this outline to connect open contours and remove small areas. Figures 2(e) and 2(f) show the segmented images in figures 2(a) and 2(b) using this method.

I. Pavlidis et al. in [8] describe a method for face segmentation using a Bayesian Approach. This method is based on the combination of two Normal Distributions per class, which are estimated using the Expectation-Maximization (EM) algorithm. This algorithm uses pixels from the skin (s) and background (b) for training. These are obtained from the training set images by selecting subregions that contain only pixels from each of these types. With this, the EM returns 4 means (μ), 4 variances (σ^2) and 4 weights (ω).

In the segmentation stage, for each pixel they have a prior distribution ($\pi^{(t)}(\theta)$ where t is the iteration) to whether that pixel is skin ($\pi^{(t)}(s)$) or background ($\pi^{(t)}(b) = 1 - \pi^{(t)}(s)$). θ is the parameter of interest, which takes two possible values (s and b) with some initial (prior) probability ($\pi^{(1)}(s) = \frac{1}{2} = \pi^{(1)}(b)$).

The input pixel value x_t has a conditional distribution $f(x_t|\theta)$ and if the particular pixel is skin we have:

$$f(x_t|s) = \sum_{i=1}^2 \omega_{s_i} \mathcal{N}(\mu_{s_i}, \sigma_{s_i}^2) \quad (1)$$

where $\mathcal{N}(\mu_{s_i}, \sigma_{s_i}^2)$ is the Normal Distribution with mean μ_{s_i} and variance $\sigma_{s_i}^2$.

The prior distribution ($\pi^{(t)}(\theta)$) combined with the likelihood ($f(x_t|\theta)$) provides (via the Bayes theorem) the posterior distribution ($p^{(t)}(\theta|x_t)$), where, for the skin pixels, according to the Bayes theorem, we have:

$$p^{(t)}(s|x_t) = \frac{\pi^{(t)}(s)f(x_t|s)}{\pi^{(t)}(s)f(x_t|s) + \pi^{(t)}(b)f(x_t|b)} \quad (2)$$

The posterior distribution is also used to obtain the prior distribution for the next iteration:

$$\pi^{(t+1)}(\theta) = \begin{cases} \pi^{(t+1)}(s) = p^{(t)}(s|x_t), & \text{when } \theta = s \\ \pi^{(t+1)}(b) = 1 - \pi^{(t+1)}(s), & \text{when } \theta = b \end{cases} \quad (3)$$

Figures 2(g) and 2(h) show the segmented images in figures 2(a) and 2(b) using this method.

3 PROPOSED METHOD

We evaluated the methods of [3] and [8] and realized that it was possible to improve their results. Our proposal is based on the method of [8]: after analyzing the results of this method (shown in figure 1(b)) we concluded that its main problem is the removal clothing because since the body warms it, clothes have temperatures similar to the skin.

To remove the effect of the clothing, we examined the vertical and horizontal pixel signatures (see figure 1(d) and 1(g)). The vertical and horizontal signatures are the sum of the pixels values along the columns, for the vertical signature, and columns, for the horizontal signature. After this we fill small areas (shown in figure 1(c)) using a dilation with a 4×4 filter. This enables the removal of incorrectly classified pixels using a proportion of the maximum values in the two signatures (see figure 1(e)). This proportion is 20%, i.e., all signatures that have values below 20% of the maximum value are considered as background. This value was obtained searching for the best performance in the training sets of the databases.

After this, we calculate the possible location of the center of the face with new signatures (horizontal in figure 1(f) and vertical in 1(h)) in figure 1(e). The center point is given by the maximum values of the signatures (when more than one maximum value exist in the horizontal or vertical signatures the average of these maximums is used). This possible center location of the face (marked with a cross in figure 1(e)) will be used for the search for the largest contour (see figure 1(j)). Before we look for the largest contour, we apply an erosion followed by dilation with a filter of 3×3 and 2×2 , respectively. This is used to remove some links between areas. For contour extraction we use the Canny edge detector. To enhance the contours extracted by the Canny method used a dilation with a 3×3 filter (see figure 1(i)). Only boundaries that have the center point inside are accepted, producing the end result of figure 1(k).

A possible drawback of this method occurs when the calculated center position of the face is not correct. This may cause the largest contour to be only partially over the actual face.

Figures 2(i) and 2(j) show the segmented images in figures 2(a) and 2(b) using our method.

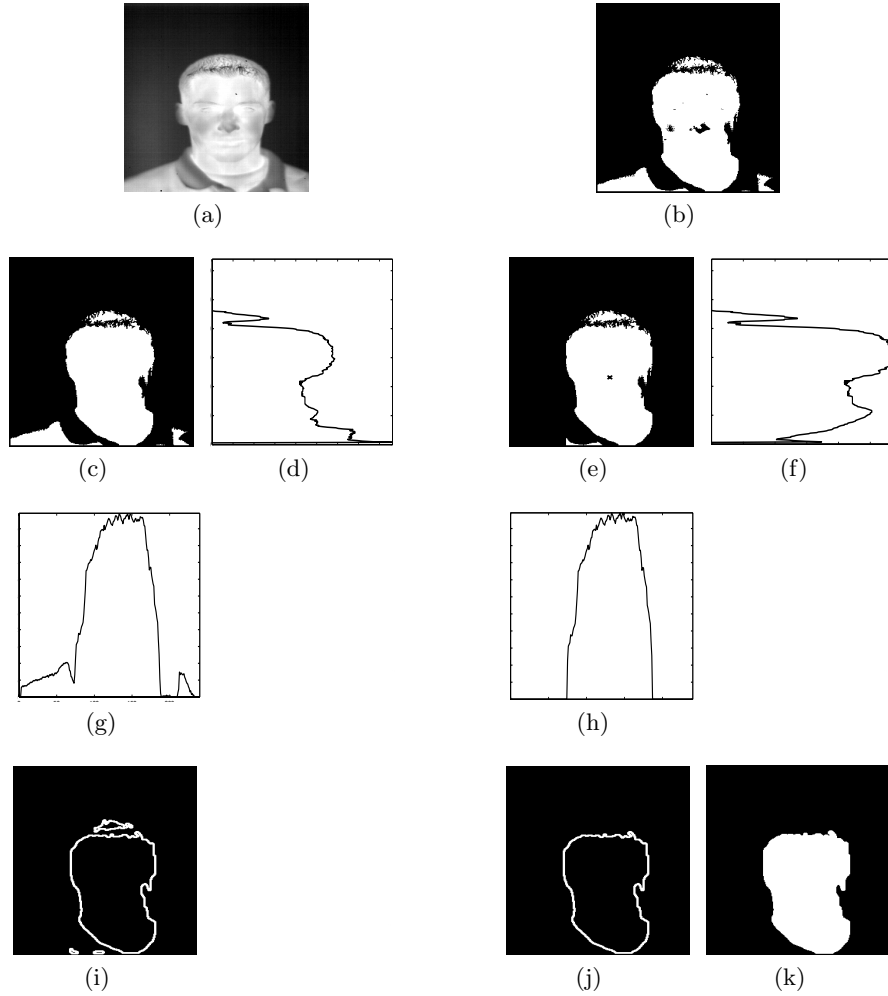


Fig. 1. Thermal face segmentation process. Figure 1(a) is the original image, from the training set of the UND database. Figure 1(b) is the original image (figure 1(a)) segmented by the method [8]. Figure 1(c) is the image (figure 1(b)) after filling small areas. Figure 1(d) is the horizontal signature of figure 1(c). Figure 1(e) is the result of the analysis of horizontal and vertical signatures of figure 1(c). Figure 1(f) is the horizontal signature of figure 1(e). Figure 1(g) is the vertical signature of figure 1(c). Figure 1(h) is the vertical signature of figure 1(e). Figure 1(i) is the result of enhancement (using a dilation with a 3×3 filter) of the contours extracted from the figure 1(e) by the Canny edge detector. Figure 1(j) is the largest contour of the figure 1(i). Figure 1(k) is the result of the face segmentation in the original image (figure 1(a)) using our method, after filling the area inside the contour of the figure 1(j).

4 EXPERIMENTAL RESULTS

4.1 Datasets

The UND database is presented in [4,2]. The "Collection C" of the UND database contains 2293 LWIR frontal face IR images from 81 different subjects. The training set contains 159 images and the test set 163.

The FSU database contains 234 frontal IR images of 10 different subjects, which were obtained at varying angles and facial expressions [9]. The train set contains 40 IR images (four per subject) and the test set 194. The images from this database have color representation. The color channels (Red (R), Green (G) and Blue (B)) and grayscale conversion were processed separately, ie., the algorithm process R, G, B and grayscale independently.

All test set images from both databases were segmented manually to create the ground truth for test sets. Method [3] does not need a training set and method [8] and ours use pixels from manually segmented regions of the training set images avoiding the need for accurate segmentation of the training set.

4.2 Experimental Results and Discussion

The requested task is quite simple: for each input image (see figure 2(a) and 2(b)) a corresponding binary output (shown in figure 2(c) and 2(d)) should be built, where the pixels that belong to the face and are noise-free should appear as white, while the remaining pixels are represented in black. The test set of the databases were used to measure pixel-by-pixel agreement between the binary maps produced by each of the algorithm $O = O_1, \dots, O_n$ (images in figures 2(e), 2(f), 2(g), 2(h), 2(i) and 2(j)) presented earlier and the ground-truth data $C = C_1, \dots, C_n$, manually built *a priori* (shown in figure 2(c) and 2(d)).

The classification error rate (E^1) of the algorithm is given by the proportion of correspondent disagreeing pixels (through the logical exclusive-or operator, see equation 4) over all the image, where $O(c', r')$ and $C(c', r')$ are, respectively, pixels of the output and true class images.

$$E^1 = \frac{1}{c \times r} \sum_{c'} \sum_{r'} O(c', r') \oplus C(c', r') \quad (4)$$

The second error measure aims to compensate the disproportion between the *a priori* probabilities of "face" and "non-face" pixels in the images. The type-I and type-II error rate (E^2) of the images is given by the average between the False Positive Rate (FPR) and False Negative Rate (FNR).

$$E^2 = 0.5 \times FNR + 0.5 \times FPR \quad (5)$$

The results of segmentation for the described methods are presented in table 1. For the UND database, we can observe that error rates obtained with our method improved upon the results of the other two methods. The same was not the case for the FSU database. In this, the FPR increased, but the final

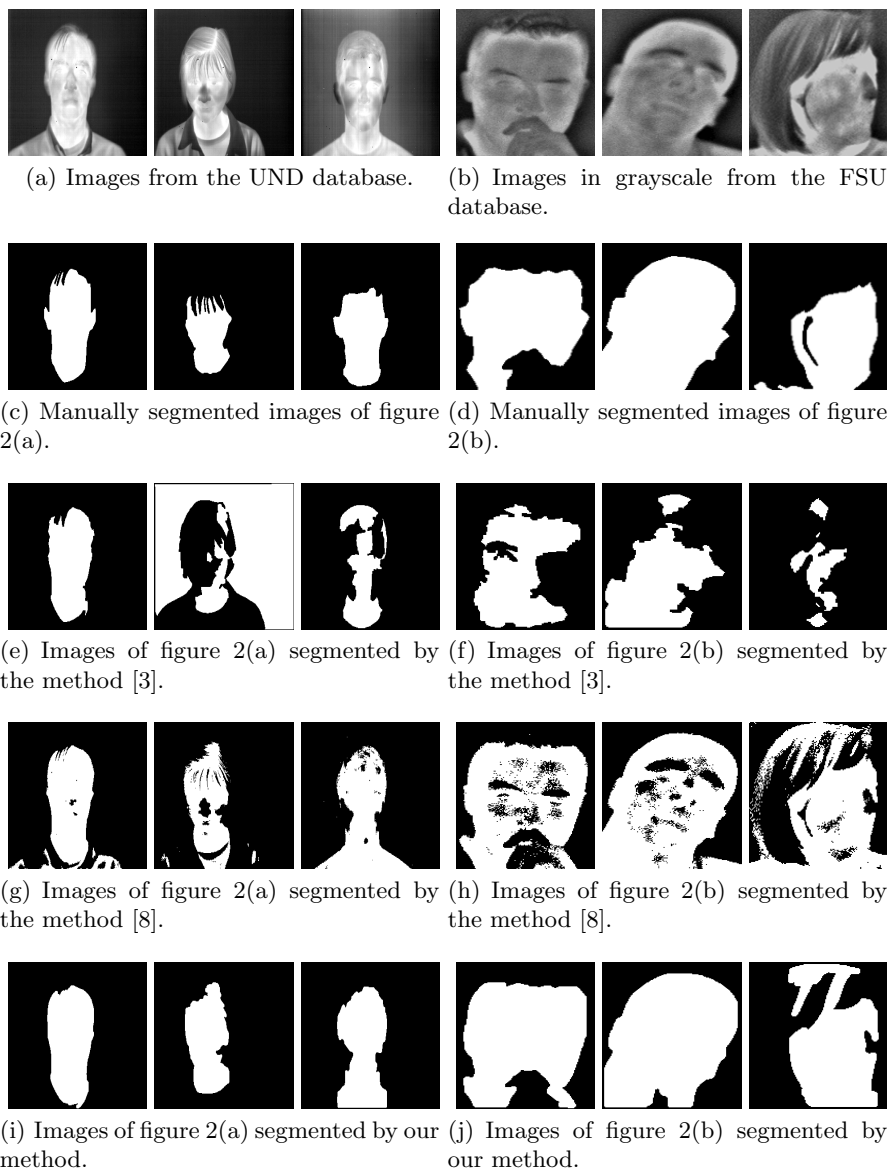


Fig. 2. Input images for the two databases, manually segmented images and the segmented images by the three methods.

two errors (E^1 and E^2) decreased due to the FNR decrease significant. In this database the FPR's increased because when the subject wears glasses and we dilate the images from the method described in [8] we include part of the glasses as face pixels and they do not belong to the face. For the FSU images (shown in

	<i>Siu-Yeung Cho et al.</i> [3]				<i>I. Pavlidis et al.</i> [8]				<i>Ours</i>			
	FNR	FPR	E^1	E^2	FNR	FPR	E^1	E^2	FNR	FPR	E^1	E^2
UND	0.369	0.354	0.356	0.362	0.166	0.080	0.093	0.123	0.145	0.050	0.065	0.097
FSU Gray	0.502	0.048	0.308	0.275	0.200	0.174	0.189	0.187	0.044	0.206	0.114	0.125
FSU R	0.533	0.046	0.325	0.290	0.206	0.169	0.190	0.187	0.058	0.186	0.112	0.122
FSU G	0.502	0.047	0.307	0.275	0.171	0.184	0.171	0.178	0.044	0.206	0.114	0.125
FSU B	0.366	0.085	0.246	0.226	0.170	0.186	0.177	0.178	0.033	0.242	0.123	0.138
FSU <i>Fusion</i> ¹	0.494	0.048	0.303	0.271	0.171	0.184	0.177	0.178	0.045	0.206	0.114	0.125
FSU <i>Fusion</i> ²	0.482	0.049	0.296	0.265	0.171	0.184	0.177	0.178	0.045	0.206	0.114	0.125

Table 1. Segmentation methods results in the test sets from the UND and FSU databases. The FSU *Fusion*¹ is the fusion between the results of segmentation for the R, G and B channels using the majority vote between them. The FSU *Fusion*² is the fusion of the results of grayscale and the R, G and B channels using a weight of 0.3 for the grayscale channel and $\frac{0.7}{3}$ for the other.

figure 2(b)), most of the noise comes from the hair since that the face fills almost entirely the image unlike what happens in the UND database, where there is a large area not covered by the face (see figure 2(a)).

The database FSU was analyzed for each channel independently and two fusions (*Fusion*¹ and *Fusion*²) were made to verify what would be the best approach for the segmentation.

The FSU *Fusion*¹ is the fusion between the results of segmentation for the R, G and B channels using the majority vote between them. The FSU *Fusion*² is the fusion of the results of grayscale and the R, G and B channels using a weight of 0.3 for the grayscale channel and $\frac{0.7}{3}$ for the other.

For the method [3] the best result was obtained with the blue channel (FSU B) for both errors E^1 and E^2 . For the method [8] the smallest error E^1 is obtained with the green (FSU G) and the smallest error E^2 appears in the green (FSU G) and blue (FSU B) channels and for the FSU *Fusion*¹ and FSU *Fusion*².

Our method has the best result in the red channel (FSU R) for both errors E^1 and E^2 . With this, we can say the best result of each method in this database depends on the type of images and that fusions does not always improve the results.

The improvements brought by our approach in relation to method [8] are the removal of clothing and the inclusion of larger number of pixels of the face. Removal of clothing is quite visible in the difference between the images resulting from [8] (figure 2(g)) and our method (figure 2(i)). In these examples it is possible to see that almost all the clothes were removed in the images of the UND database. The inclusion of the pixels of the face is most visible in the FSU database as can be seen in the images of figures 2(h) and 2(j), the first being the result of [8] and ours is the second. With this we minimize the FNRs, causing us to obtain more pixels for face recognition tasks.

5 CONCLUSION

In this paper we proposed a face segmentation method for LWIR images. The method creates two Gaussian distributions for each type of pixel (face and back-

ground) and post-processes the obtained images by closing small holes using morphological operators (decreasing FNRs) and removing the effect of clothes through the analysis of vertical and horizontal image signatures.

The experimental results show that our proposal improves accuracy from 2.8% to over 25% depending on the dataset and the method against which we are comparing.

We are currently searching for new features that may improve segmentation performance.

6 ACKNOWLEDGMENTS

We acknowledge the financial support given by "FCT - Fundação para a Ciência e Tecnologia" and "FEDER" in the scope of the PTDC/EIA/69106/2006 research project "BIOREC: Non-Cooperative Biometric Recognition".

We wish to thank Professor Cho Siu-Yeung David, Assistant Professor in the School of Computer Engineering at Nanyang Technological University (NTU) for the source code of his method[3].

References

1. Bowyer, K., Chang, K., Flynn, P.: A survey of approaches to three-dimensional face recognition. 17th International Conference on Pattern Recognition (ICPR 2004). pp. 358–361
2. Chen, X., Flynn, P., Bowyer, K.: IR and visible light face recognition. *Computer Vision and Image Understanding* 99, 332–358 (2005)
3. Cho, S., Wang, L., Ong, W.: Thermal imprint feature analysis for face recognition. *IEEE International Symposium on Industrial Electronics (ISIE)*, 1875–1880
4. Flynn, P., Bowyer, K., Phillips, P.: Assessment of Time Dependency in Face Recognition: An Initial Study. *Lecture Notes in Computer Science* 2688, 44–51 (2003)
5. Gyaourova, A., Bebis, G., Pavlidis, I.: Fusion of infrared and visible images for face recognition. *Lecture Notes in Computer Science* pp. 456–468
6. Jain, A., Flynn, P., Ross, A.: *Handbook of Biometrics*. Springer-Verlag New York, Inc., Secaucus, NJ, USA (2007)
7. Kong, S., Heo, J., Abidi, B., Paik, J., Abidi, M.: Recent advances in visual and infrared face recognition - a review. *Computer Vision and Image Understanding* (1), 103–135
8. Pavlidis, I., Tsiamyrtzis, P., Manohar, C., Buddhharaju, P.: *Biometrics: Face recognition in thermal infrared*, chap. 29, pp. 1–15. CRC Press (2006)
9. Srivastava, A., Liu, X.: Statistical hypothesis pruning for identifying faces from infrared images. *Image and Vision Computing* pp. 651–661

Apêndice B

Thermal Infrared Face Segmentation: Improving Accuracy and Runtime through Statistical Modelling and Active Contour

Neste anexo inclui-se um artigo proposto para o TPAMI [7].

Thermal Infrared Face Segmentation: Improving Accuracy and Runtime through Statistical Modelling and Active Contour

Silvio Filipe and Luís A. Alexandre

Abstract—The aim of this paper is to present a method for the automatic segmentation of face images captured in Long Wavelength Infrared (LWIR), allowing for a large range of face rotations and expressions. The motivation behind this effort is to enable better performance of face recognition methods in the thermal Infrared (IR) images. The paper discusses 8 different proposed approaches and compares their performance against three other previously published methods. The proposed approaches are based on statistical modelling of pixel intensities and active contour application, although several other image processing operations are also performed. We also study the performance of the fusion of 3 of the approaches under different rules (majority vote, OR and AND fusions). Experiments were performed on a total of 893 test images from 4 public available databases. The obtained results improve on previous existing methods up to 31.2% for the first measure error (E_1) and up to 39.0% for the second measure (E_2), depending on the method and database. Regarding the computational time, our proposals improve up to 75.4% when compared with the other proposals.

Index Terms—Face Segmentation, Thermal infrared, Human skin segmentation, Image segmentation.

1 INTRODUCTION

A LARGE amount of research has been conducted in the field of face recognition, mainly in the visible spectrum. These systems have problems dealing with light variations [1], [2]. Some of the proposed solutions use 3D facial recognition [3] and combine face recognition in both visible and IR spectrum [4], [5].

The growing interest in robust methods (for example, for security applications) has driven the development of facial recognition exclusively in the infrared. Recognition in the LWIR is not affected by light variations.

A crucial step in the process of face recognition (see figure 1) is the face segmentation. This is more demanding than simple face detection since it pinpoints not only the face's locations, but must also describe its shape. A robust segmentation system can improve recognition rates regardless of the recognition method.

In contrast with the visible wavelength, where numerous methods have been proposed to accomplish this task (based on color [6], geometry [7], etc.), in the LWIR there is a lack of proposals to improve the current status.

Figure 1 shows the general scheme for a recognition system. This scheme can be used for face recognition either in the visible or thermal wavelengths and can be used also for other recognition modalities, such as those

that use iris images [8]. The recognition system has two parts:

- *Offline Process:* The training set images are captured by a camera. A segmentation algorithm obtains the face and features are extracted. These are stored in a database.
- *Online Process:* Given an image, it is segmented and features are extracted as in the offline process. These features are compared against the ones stored in the database and a match score is produced.

In the next sections we present a short description of the available face segmentation methods (section 2) and present our face segmentation methods (section 3). In Section 4, we present the datasets used and experimental results, including a discussion. We end the paper in section 5 with the conclusions.

2 OVERVIEW OF FACE SEGMENTATION IN THERMAL INFRARED IMAGES

Face segmentation, given that it is a preprocessing step for all recognition methods, will lead to their failure if it is not correctly performed. One of the proposed approaches is based on the creation of an elliptical mask that will be placed over the face image [9], but this approach will only work on frontal and centered faces.

S. Cho *et al.* in [10] present a method for face segmentation in IR images based on the Sobel Edge detector and morphological operations. After the Sobel Edge detector, the largest contour is considered to be the one best describing the face. They apply the morphological operations to the area contained in this outline to connect

- Department of Computer Science, IT - Instituto de Telecomunicações, SOCIA - Soft Computing and Image Analysis Group, University of Beira Interior, 6200-Covilhã, Portugal.
E-mail: m2213@ubi.pt (S. Filipe) and lfbaa@di.ubi.pt (L. A. Alexandre).
- Additional material available in: http://www.di.ubi.pt/~lfbaa/thermal_face.html

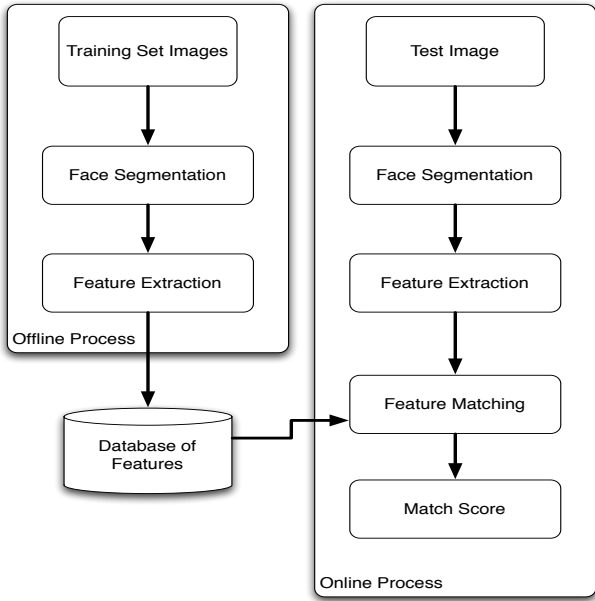


Fig. 1. Block diagram of face recognition system.

open contours and remove small areas. Row 3 of figure 10 shows the segmented images using this method.

I. Pavlidis *et al.* in [11] describe a method for face segmentation using a Bayesian Approach. This method is based on the combination of two Normal Distributions per class, which are estimated using the Expectation-Maximization (EM) algorithm. This algorithm uses pixels from the skin (s) and background (b) for training. These are obtained from the training set images by selecting sub-regions that contain only pixels from each of these types (see figure 2). With this, the EM returns 4 means (μ), 4 variances (σ^2) and 4 weights (ω).



Fig. 2. Sample crop of face regions (boxes with black borders) and background (boxes with white borders).

In the segmentation stage, for each pixel there is a prior distribution ($\pi^{(t)}(\theta)$) where t is the iteration) to whether that pixel is skin ($\pi^{(t)}(s)$) or background

($\pi^{(t)}(b) = 1 - \pi^{(t)}(s)$). θ is the parameter of interest, which takes two possible values (s and b) with initial prior probability equal to $\pi^{(1)}(s) = \frac{1}{2} = \pi^{(1)}(b)$.

The input pixel value x_t has a conditional distribution $f(x_t|\theta)$ and if the particular pixel is skin we have:

$$f(x_t|s) = \sum_{i=1}^2 \omega_{s_i} \mathcal{N}(\mu_{s_i}, \sigma_{s_i}^2), \quad (1)$$

where $\mathcal{N}(\mu_{s_i}, \sigma_{s_i}^2)$ is the Normal Distribution with mean μ_{s_i} and variance $\sigma_{s_i}^2$.

The prior distribution ($\pi^{(t)}(\theta)$) combined with the likelihood ($f(x_t|\theta)$) provides (via the Bayes theorem) the posterior distribution ($p^{(t)}(\theta|x_t)$), where, for the skin pixels, we have:

$$p^{(t)}(s|x_t) = \frac{\pi^{(t)}(s)f(x_t|s)}{\pi^{(t)}(s)f(x_t|s) + \pi^{(t)}(b)f(x_t|b)} \quad (2)$$

The posterior distribution is also used to obtain the prior distribution for the next iteration:

$$\pi^{(t+1)}(\theta) = \begin{cases} \pi^{(t+1)}(s) = p^{(t)}(s|x_t), & \text{when } \theta = s \\ \pi^{(t+1)}(b) = 1 - \pi^{(t+1)}(s), & \text{when } \theta = b \end{cases} \quad (3)$$

Figure 3 contains the pixel intensity distributions (solid line) from the cropped images used for training (as the ones presented in figure 2), for the four databases used in this paper. The dashed lines are the mixtures of Gaussian distributions estimated with EM for each pixel type.

In figure 3(d), the distributions are choppy given that the images in FSU database only have around 70 different values.

Row 4 of figure 10 shows images segmented using this method.

3 PROPOSED METHODS

We evaluated the methods of [10] and [11] and realized that it was possible to improve their results, as we will see in the following sections.

In sub-sections 3.1 to 3.5 we describe several operations that will be used on the methods proposed in 3.6.

3.1 Rectangular Region of Interest (RROI)

An interesting operator would give the Rectangular Region of Interest (RROI) that contains the face. This would avoid the problems caused by the clothes since, as the body warms it, clothes have temperatures similar to the skin and may difficult the pixel intensity-based segmentation approaches.

To obtain the RROI we will analyse the vertical and horizontal image signatures. These are 1D vectors that contain the sum of the intensity of the pixels along the columns (equation (4)) and rows (equation (5)), respectively:

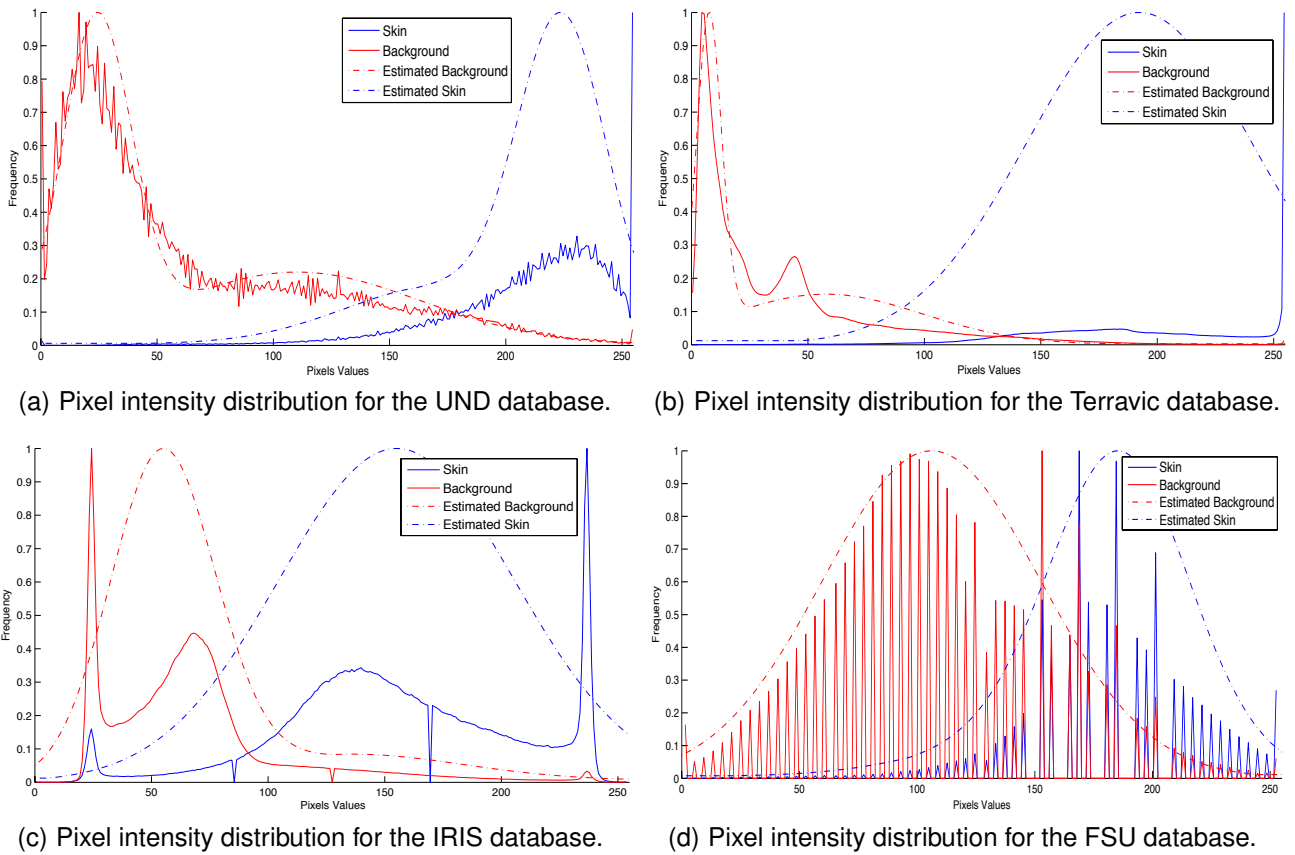


Fig. 3. Face pixel intensity distribution (blue line) and background (red line) for the four databases used.

$$\text{sig}V(c) = \sum_{r=1}^{nR} i(c, r) \quad (4)$$

$$\text{sig}H(r) = \sum_{c=1}^{nC} i(c, r) \quad (5)$$

where c and r are the indexes of column and row for image i of dimension $nC \times nR$.

The first step to obtain the RROI is now described (see figure 4). We start by analysing the vertical signature. The signal in figure 4(b) represents the vertical signature of figure 4(a). This signal has several high frequency oscillations that will appear in its derivative. This can be avoided by smoothing it with a 1D Gaussian filter (we used the one in figure 4(c)). The standard deviation of the Gaussian filter is $\sigma = 0.05 \times nC$. This value was obtained by studying the influence of different values of σ in the training set images. The result of the convolution is in figure 4(d). The derivative of this signal is in figure 4(e). The following step consists in determining the extrema of this signal: in figures 4(a) and 4(e) we marked the maximum in red (*colLeft*) and the minimum in green (*colRight*). The two lines indicate the location of a large variation in image intensity that we identify with the sides of the face.

The next step in defining the RROI is the analysis of the horizontal signature. For this we only consider the part of the image between the two extrema detected in the vertical signature analysis. This removes the shoulders of the subjects and overcomes one of the issues that was causing problems in the previous approaches. The process used in the analysis of the horizontal signature (see figure 5)) is similar to the one used to analyse the vertical signature. The main difference is the filter used: in this case its width is $0.15 \times nR$. The shape and size of the filter were selected to remove sudden variations that would appear in the signal when the subjects are wearing glasses or have a cold nose. Next we used a process similar to the one for the vertical signature in order to obtain the extrema of the smoothed signal. The maximum (*rowUp*) is represented by a red line in figures 5(a) and 5(e) and the minimum (*rowDown*) by a green line.

3.2 Elliptical Region of Interest (EROI)

The idea of defining an ellipse to enclose the face is appealing since a face has approximately the shape of an ellipse. An example is the work in [9] where the segmentation approach uses such an ellipse. We will also use an ellipse to improve the RROI around the face and to initialize the mask used in the method [12] (discussed

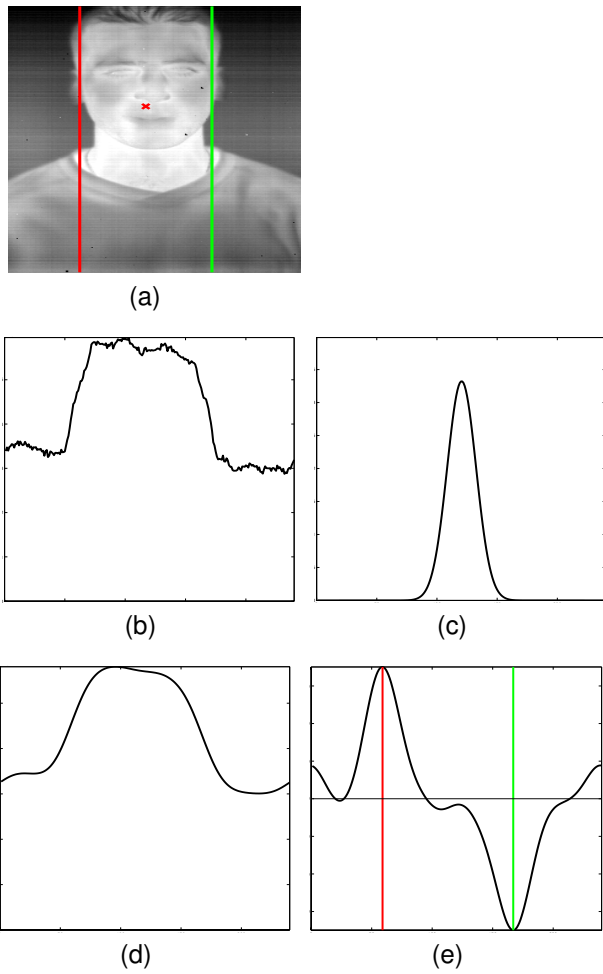


Fig. 4. Vertical signature analysis. Figure 4(a) is the original image. Figure 4(b) is its vertical signature. Figure 4(c) is the Gaussian filter used in the convolution. Figure 4(d) is result of applying this filter to 4(b). Figure 4(e) is the derivative of figure 4(d).

in section 3.4).

The ellipse will be defined inside the previously obtained RROI. We start by finding the center of the face which we will use as the center of the ellipse. To determine this center point $(colCFace, rowCFace)$, the red cross on the image in figure 4(a), we use the extrema obtained while searching for the RROI:

$$colCFace = (colLeft + colRight)/2 \quad (6)$$

$$rowCFace = (rowUp + rowDown)/2 \quad (7)$$

Then, using algorithm 1, we can obtain the $\{(X_0, Y_0), \dots, (X_{2\pi}, Y_{2\pi})\}$ points of the ellipse. The algorithm receives the coordinates of the face center, $(colCFace, rowCFace)$ and the coordinates of the left upper corner of the RROI, $(colLeft, rowUp)$. These points are used for obtaining the distance from the center of the face to the left side of the RROI (which is denoted by a) and the distance from the center

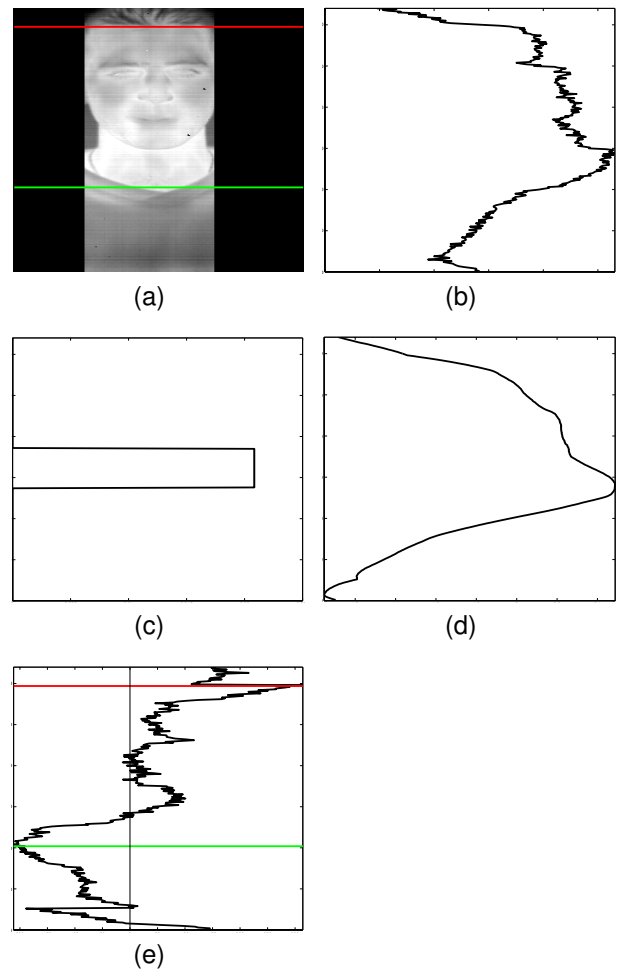


Fig. 5. Horizontal signature analysis. Figure 5(a) is the original image. Figure 5(b) is its horizontal signature. Figure 5(c) is the filter used in the convolution. Figure 5(d) is result of applying this filter to 5(b). Figure 5(e) is the derivative of figure 5(d).

to the top of the RROI (denoted by b). a and b are used to convert the polar coordinates of the points that belong to the ellipse to cartesian coordinates as $\{(X_0, Y_0), \dots, (X_{2\pi}, Y_{2\pi})\}$.

Algorithm 1 Algorithm for the Elliptical Region of Interest (EROI).

Input: $(colLeft, rowUp)$ and $(colCFace, rowCFace)$

Output: $\{(X_0, Y_0), \dots, (X_{2\pi}, Y_{2\pi})\}$

$a = |colLeft - colCFace|$

$b = |rowUp - rowCFace|$

for $\theta = 0$ to 2π **do**

$X(\theta) = colCFace + (a * \cos(\theta))$

$Y(\theta) = rowCFace + (b * \sin(\theta))$

end for

3.3 Thresholding

We will use a threshold step in some of the proposed methods. The threshold is adaptive in the sense that it depends on the training set distributions for each database. The goal of this threshold is to separate most of the skin pixels from the background, and so it will be chosen to include some of the background pixels but to guarantee that most of the skin pixels will be included.

First identify the point at which the the distributions (solid lines in figure 3) for skin and background pixels intersect. The threshold value is chosen as half the pixel value identified.

3.4 Active Contours without Edges

Based on the Mumford-Shah minimal partition functional [13], Chan and Vese [12], [14] proposed a level set model for active contours to detect objects whose boundaries are not necessarily defined by the gradient, as with the classical active contour.

The main motivation for the use of this type of algorithms is their excellent ability to segment objects present in images. We choose for this step the use of the active contours without edges [12]. In [15] the authors refer that this method achieves greater accuracy and robustness at the cost of a major reduction in speed. For this work we imposed a restriction: the maximum number of iterations is now 200, since this is a value that yielded good results on experiments performed with the training sets this will reduce the computational cost without visible accuracy loss, according to some training set experiments performed. As will be seen below, the use of this approach will not be slower than the other methods [10] and [11]. The processing time will depend of the type of initial contour and of its position: the further away the initial contour is from the face, the longer it will take to converge.

An example of using this algorithm can be seen in [16] where it is used to segment teeth and where we can see that the X-ray images have some similarities with the images of the LWIR.

The result of segmenting the images in the first row of figure 10 with this method are in row 5 of the same figure. The initial boundary was a centered rectangle of 90×140 pixels. This size was obtained by approximating the face size of the images in the training set. We will see below other approaches that use this method but with a different type of initial boundary (an ellipse). As we mentioned previously, the initial contour will affect the processing time, hence the importance of choosing an initial contour with a shape similar to a face.

3.5 Face Pixel Identification from Binary Image (FPIBI)

We propose a method for selecting the face pixels in binary images (see figure 6(a)) that result from the application of the methods [11], [12]. This method identifies the largest contour that contains the face center,

considers all pixels inside this contour as face pixels with the exception of the pixels that belong to glasses (see figure 6).

We start by identifying the center of the face as explained in the RROI operation (red cross in the figure 6(a)). Then we apply a dilation followed by an erosion (an opening) using structural elements of sizes 3×3 and 2×2 , respectively. These morphological operations are used to remove small areas. Then, an edge map is obtained using the Canny edge detector [17]. The obtained edges are enhanced through a dilation with a structuring element of size 3×3 (see figure 6(b)). From these edges we select the largest that contains the face center. We now assume that all the pixels inside this largest contour are face pixels (see figure 6(c)).

To remove glasses that may have been considered as being face in the previous step, we make the absolute difference between the image before the selection of the largest contour and the image that results from filling this largest contour (see figure 6(d)). With this difference we will obtain the image regions that were altered by the filling. We apply an opening with a circular structuring element of 10 pixel radius (see figure 6(e)). Only the largest regions, such as the glasses, remain after the application of this morphological operator. This resulting image is summed with the the one that results from filling the largest contour and the values that exceed the largest possible value (1 or 255 depending on the representation used for white in the images) are clipped to zero (not considered as face pixel) (see figure 6(f)).

3.6 Segmentation Methods

In this section we present some methods that combine several of the previously described operations. We shall call them **A**, **B**, ..., **H**. The result of the application of these methods to the images in the first row of figure 10 is presented in the same figure in rows 6 to 13, respectively.

Method **A** consists on considering face pixels those that are both in the region defined by RROI and those that result from applying the threshold defined in section 3.3 (see figure 7(a)). The application of RROI is used to localise the face and consider all the pixels inside the region as face pixels. The threshold will define which are the hottest pixels in the region. The application of an **AND** operator to these two images gives us only the hottest of the pixels inside the RROI.

Method **B** is similar to method **A** except that instead of using RROI, in this case we use operation EROI (see figure 7(b)).

Method **C** starts with the application of method RROI and then applies the method in [11] to the obtained RROI (see figure 7(c)). Method **D** applies FPIBI after the application of method **C** (see figure 7(d)). Method **E** starts by extracting the ellipse to the cropped RROI (see figure 7(e)). The obtained ellipse is used as the first iteration of method [12] which is then used to



(a) Example of a binary input image for this operation. (b) Enhanced Edges of the binary image.



(c) Fill the largest contour containing the center of the face (red cross in figure 6(a) and 6(b)). (d) Absolute difference between the figures 6(a) and 6(b).



(e) Result of the morphological operator (opening) with a circular structuring sum element of 10 pixel radius. (f) Final result of the operation, after making the with a circular structuring sum.

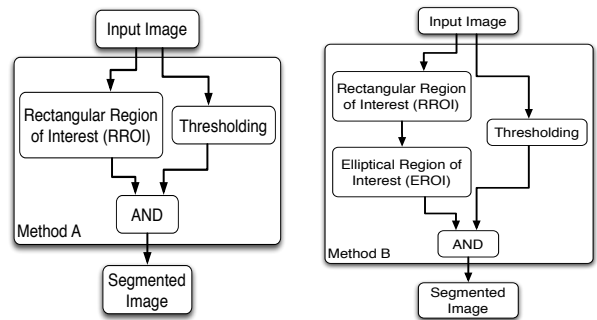
Fig. 6. Step by step example of the application of the FPIBI operation (see text for detailed description).

finish the segmentation. Method F extends method E (see figure 7(f)): it consists on the application of FPIBI after method E is applied. Method G is similar to method E but without cropping the RROI (see figure 7(g)). The difference between methods G and H is the same as between E and F (see figure 7(h)).

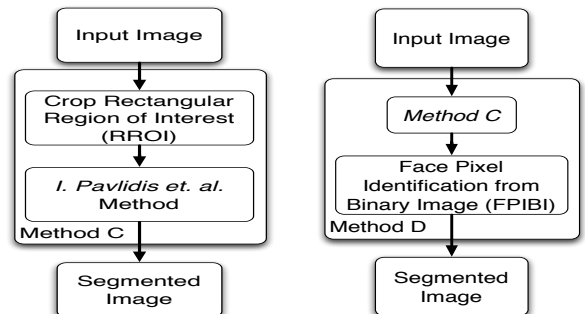
The main problem we had identified in method [11] (the clothes have a temperature similar to the skin) is overcome by these methods, specially D, F and H. This is due to the use of FPIBI to select the largest region of the image that contains a particular point (the face center).

3.7 Fusion

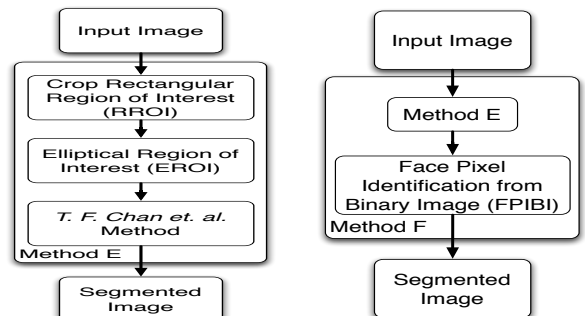
We also considered fusing the results of some of the methods (the ones that are more diverse, such that their errors are made in different parts of the images). The fused methods are A, F and H. The fusion approaches



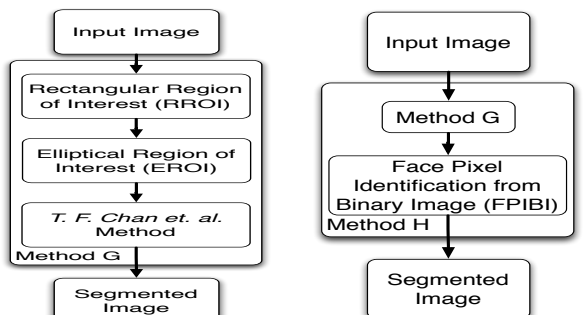
(a) Block diagram of the method A. (b) Block diagram of the method B.



(c) Block diagram of the method C. (d) Block diagram of the method D.



(e) Block diagram of the method E. (f) Block diagram of the method F.



(g) Block diagram of the method G. (h) Block diagram of the method H.

Fig. 7. Block diagrams of the segmentation methods.

used are:

- 1) **Majority Vote Fusion:** majority vote from the results of methods **A**, **F** and **H**;
- 2) **AND Fusion:** intersection between the results:

$$ANDFusion(iC, iR) = A(iC, iR) \cap F(iC, iR) \cap H(iC, iR)$$

where iC and iR are the indexes of the columns and rows and $A(iC, iR)$, $F(iC, iR)$ and $H(iC, iR)$ are the results of the methods **A**, **F** and **H**, respectively.

- 3) **OR Fusion:** union of the results:

$$ORFusion(iC, iR) = A(iC, iR) \cup F(iC, iR) \cup H(iC, iR).$$

The result of the application of these fusion approaches to the images in the first row of figure 11 is presented in the same figure in rows 3 to 5, respectively.

4 EXPERIMENTAL RESULTS

4.1 Datasets

The University of Notre Dame (UND) database is presented in [18], [19]. The "Collection X1" of the UND database contains 2293 LWIR frontal face IR images from 81 different subjects. The training set contains 159 images and the test set 163. Two images from this database are in the first row, columns 1 and 2 of figure 10.

The "Dataset 04: Terravic Facial IR Database" is a subset of the Object Tracking and Classification in and Beyond the Visible Spectrum (OTCBVS) database [20]. This database contains 24508 images of 20 different persons. It has different poses (rotations front, left, right), images captured indoor and outdoor and images of people with glasses and hats. The training set has 235 images and the test set has 240. Two images from this database are in the first row, columns 3 and 4 of figure 10.

The "Dataset 02: IRIS Thermal/Visible Face Database" is a subset of the OTCBVS database [21]. The 4228 images were acquired in the Imaging, Robotics, and Intelligent Systems laboratory (University of Tennessee) (IRIS) with 11 images per rotation (images for each expression and illumination) yielding between 176 and 250 images per person. This database was acquired with different illuminations for visible wavelength work. These differences do not affect the LWIR so we ignore the different versions due to illuminations changes. The training and test sets have both 296 images. Two images from this database are in the first row, columns 5 and 6 of figure 10.

The Florida State University (FSU) database contains 234 frontal IR images of 10 different subjects, which were obtained at varying angles and facial expressions [22]. The training set contains 40 IR images (four per subject)

and the test set 194. Two images from this database are in the first row, columns 7 and 8 of figure 10.

The test set images from all databases were segmented manually to create the test set ground truth (samples shown in row 2 of figure 10). Method [10] does not need a training set and method [11] and ours use pixels from manually segmented regions of the training set images avoiding the need for accurate segmentation of the training set.

Table 1 shows the percentage of face and background pixels present in the databases used in this paper. These values are obtained based on the manually segmented images. We can see that the FSU database is the only that has more face than background pixels (the other databases have a face to background pixel ratio between 15.40% and 32.65%).

	Pixel percentage (%)	
	Face	Non-Face
UND	15.40	84.60
Terravic	21.11	78.89
IRIS	32.65	67.35
FSU	57.22	42.78

TABLE 1

Face and background pixel ratios to the total number of image pixels, for the different databases.

A list with the names of the images used in the train and test sets is available at http://www.di.ubi.pt/~lfbaa/thermal_face.html.

4.2 Experimental Results and Discussion

The requested task is quite simple: for each input image (as the ones shown in the first row of figure 10) a corresponding binary output (as those shown in the second row of the same figure) should be built, where the pixels that belong to the face and are noise-free should appear as white, while the remaining pixels are represented in black. The test set of the databases were used to measure pixel-by-pixel agreement between the binary maps produced by each of the algorithms $O = O_1, \dots, O_n$ (these maps are shown in figure 10, rows 3 to 13) and the ground-truth data $C = C_1, \dots, C_n$, manually built *a priori* (see examples in row 2 of figure 10).

The classification error rate (E_1) of the algorithm is given by the proportion of correspondent disagreeing pixels (through the logical exclusive-or operator, see equation 8) over all the image, where $O(c', r')$ and $C(c', r')$ are, respectively, pixels of the output and true class images:

$$E_1 = \frac{1}{nC \times nR} \sum_{c'} \sum_{r'} O(c', r') \oplus C(c', r') \quad (8)$$

The second error measure aims to compensate the disproportion between the *a priori* probabilities of "face"

and “non-face” pixels in the images. The type-I and type-II error rate (E_2) of the images is given by the average between the False Positive Rate (FPR) and False Negative Rate (FNR):

$$E_2 = 0.5 \times FNR + 0.5 \times FPR \quad (9)$$

The graphs in figures 8 and 9 show the values of the error measures E_1 and E_2 , which can also be observed in tables 2, 3, 4 and 5. It is possible to see that there is no single method that is the best in all databases. This results from the fact that images from different databases possess different features. These graphs also allow for a fast visualization of the relation between the proposed segmentation approaches and the proposals of other authors, across the four databases.

Table 2 contains the error rates for all the segmentation methods on the UND database. It also contains the average execution time per image. In this case, the method with the best result (method **B**) for error rate E_1 is not the same with best result for measure E_2 (method **A**). This is due to the fact that E_2 favors methods that produce balanced FNR and FPR, whereas E_1 only counts the number of wrong classified pixels. The improvement from previous published approaches was of 31.2% for E_1 and 30.0% for E_2 .

	UND				Time(s)
	FNR	FPR	E_1	E_2	
<i>S. Cho et al.</i>	0.370	0.354	0.356	0.362	1.236
<i>I. Pavlidis et al.</i>	0.166	0.080	0.093	0.123	4.480
<i>T. F. Chan et al.</i>	0.070	0.090	0.087	0.080	9.278
A	0.051	0.073	0.069	0.062	0.012
B	0.110	0.032	0.044	0.071	0.089
C	0.188	0.034	0.058	0.111	1.100
D	0.212	0.023	0.052	0.117	1.294
E	0.085	0.043	0.050	0.064	2.754
F	0.103	0.034	0.045	0.068	2.912
G	0.062	0.097	0.092	0.080	6.808
H	0.067	0.077	0.076	0.072	6.994
Mean Fusion	0.072	0.054	0.056	0.063	F + H
AND Fusion	0.119	0.031	0.045	0.075	F + H
OR Fusion	0.031	0.099	0.089	0.065	F + H

TABLE 2
Experiment results in the UND database.

Table 3 shows the error rates and execution times for the Terravic database. Again, the best algorithm for E_1 (method **H**) is not the same as the best for E_2 (method **G**). The improvement from previous published approaches was of 10.5% for E_1 and 12.5% for E_2 .

Table 4 contains the results for all the segmentation methods on database IRIS. In this case the best results regarding both error measures are obtained by the same method (method **H**). The improvement from previous published approaches was of 28.8% for E_1 and 39.0% for E_2 .

	Terravic				Time(s)
	FNR	FPR	E_1	E_2	
<i>S. Cho et al.</i>	0.207	0.118	0.137	0.163	1.237
<i>I. Pavlidis et al.</i>	0.055	0.062	0.060	0.058	4.554
<i>T. F. Chan et al.</i>	0.057	0.034	0.039	0.045	10.361
A	0.087	0.058	0.064	0.072	0.012
B	0.194	0.023	0.059	0.108	0.091
C	0.134	0.022	0.045	0.078	1.174
D	0.152	0.015	0.044	0.083	1.330
E	0.119	0.019	0.040	0.069	3.071
F	0.141	0.014	0.041	0.078	3.233
G	0.036	0.040	0.039	0.038	7.115
H	0.055	0.026	0.032	0.041	7.227
Mean Fusion	0.106	0.022	0.040	0.064	F + H
AND Fusion	0.143	0.012	0.039	0.077	F + H
OR Fusion	0.034	0.064	0.058	0.049	F + H

TABLE 3
Experiment results in the Terravic database.

	IRIS				Time(s)
	FNR	FPR	E_1	E_2	
<i>S. Cho et al.</i>	0.773	0.139	0.346	0.456	0.959
<i>I. Pavlidis et al.</i>	0.093	0.144	0.127	0.118	3.233
<i>T. F. Chan et al.</i>	0.204	0.077	0.118	0.140	8.717
A	0.122	0.085	0.097	0.103	0.011
B	0.248	0.043	0.110	0.145	0.081
C	0.199	0.052	0.100	0.126	1.277
D	0.175	0.028	0.076	0.102	1.299
E	0.177	0.053	0.094	0.115	3.491
F	0.194	0.034	0.087	0.114	3.631
G	0.068	0.078	0.075	0.073	7.025
H	0.086	0.045	0.058	0.066	7.039
Mean Fusion	0.139	0.039	0.072	0.089	F + H
AND Fusion	0.214	0.029	0.089	0.121	F + H
OR Fusion	0.048	0.097	0.081	0.073	F + H

TABLE 4
Experiment results in the IRIS database.

The results for the FSU database are in table 5. As was the case for the IRIS database, the same method is the best regarding both error measures (method **H**). The improvement from previous published approaches was of 20.6% for E_1 and 16.6% for E_2 .

From the results in tables 2, 3, 4 and 5 it is visible that, according to the E_1 error measure, method **H** was better than the others in three out of four databases. Using the E_2 error measure it only beats the others in two databases.

The execution times presented were obtained using *Matlab R2009b* on an *Intel Core 2 Q9300* (2.5GHz), 4Gb RAM (FSB 1066) and *Fedora Core 11* operative system. We observe that, as expected, all the methods that used the crop from the RROI (**C**, **D**, **E** and **F**) are faster when compared with the original proposals [11] and [12].

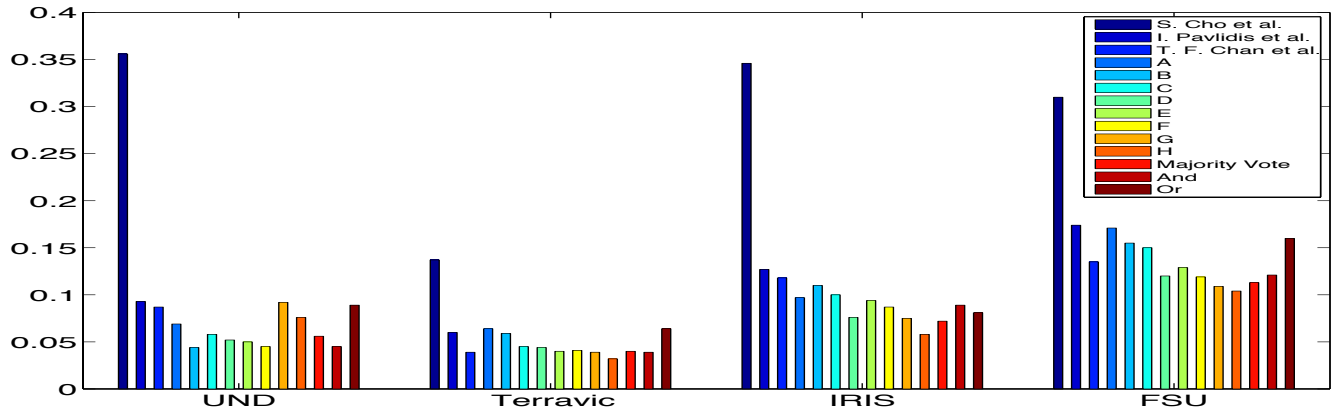


Fig. 8. Error values E_1 from tables 2, 3, 4 and 5.

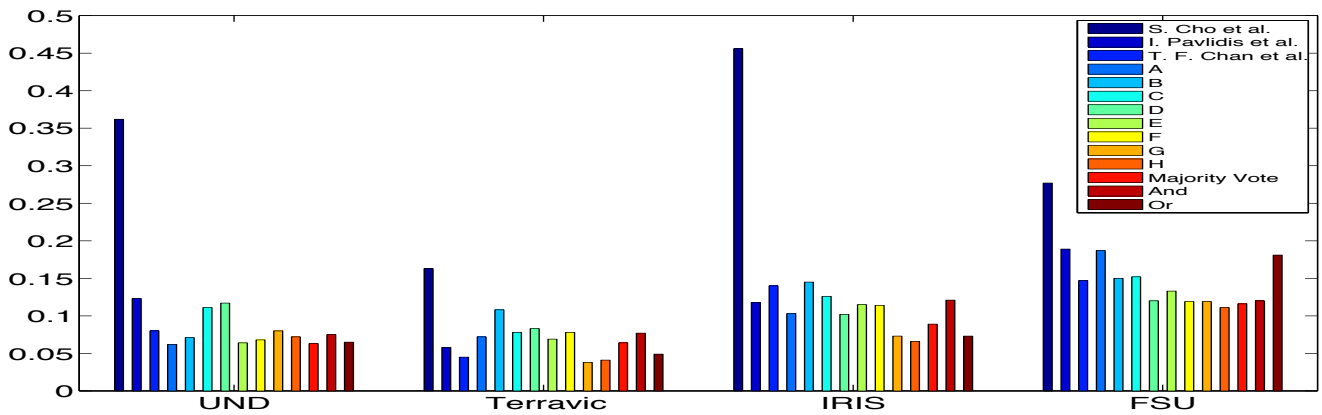


Fig. 9. Error values E_2 from tables 2, 3, 4 and 5.

	FSU				Time(s)
	FNR	FPR	E_1	E_2	
<i>S. Cho et al.</i>	0.506	0.048	0.310	0.277	0.379
<i>I. Pavlidis et al.</i>	0.089	0.289	0.174	0.189	1.296
<i>T. F. Chan et al.</i>	0.058	0.237	0.135	0.147	6.039
A	0.73	0.302	0.171	0.187	0.009
B	0.182	0.118	0.155	0.150	0.078
C	0.141	0.163	0.150	0.152	1.066
D	0.122	0.117	0.120	0.120	1.091
E	0.106	0.160	0.129	0.133	2.759
F	0.119	0.120	0.119	0.119	2.850
G	0.052	0.186	0.109	0.119	3.677
H	0.062	0.159	0.104	0.111	3.746
Mean Fusion	0.095	0.138	0.113	0.116	F + H
AND Fusion	0.125	0.115	0.121	0.120	F + H
OR Fusion	0.034	0.328	0.160	0.181	F + H

TABLE 5
Experiment results in the FSU database.

Regarding the method in [11] and its variants (C and D), the average (across all databases) improvement is

55.7%, with as much as 75.4% on the UND database.

The improvement in execution time of this method is due to the use of the crop such that most of the clothes are removed from subsequent processing (note that this was one of the problems identified in the method of [11]).

For the method proposed in [12] and its variants (E and F) the improvements were even higher at 62.9% for the average across all databases and for the Terravic database it reached 70.4%. For the other two variants (G and H), the average (across all databases) improvement is 28.6% and with as much as 31.3% on the Terravic database.

The improvement in variants E to H with regard to the method in [12] is due to the position and shape (ellipse) of the initial contour. The method in [12] uses an initial contour that is centered in the image, whereas some of the faces are not centered. Hence we applied RROI (searching for the face position and size) before using the mentioned method (see the variants E to H).

Regarding the fusion approaches, we conclude that they were not beneficial, specially if we consider the execution time. The improvements are basically on the reduction of FNR when using **OR Fusion** and reduction

of the FPR when using **AND Fusion** (except on the FSU database where the method from [10] had the smallest FPR).

We consider that the best method for segmenting LWIR images is method **H** since it has the best overall performance across databases. Regarding the E_1 error, method **H** was the best method in all but the UND database. In this database the best method was **B** mainly because it had a small FPR since it requires the face pixels to belong to an elliptical region. The execution times of the fusions were presented as the execution times of methods **F** plus **H** and did not take into account the time of the method **A** because this method is both very fast and part of its processing is already done to obtain the other two methods.

Regarding the E_2 error, method **H** is the best for two of the databases (IRIS and FSU), the second best for the Terravic database (marginal 0.003 difference from the best which was method **G**) and it did not show a good performance in the UND database, as it had previously happened with error measure E_1 .

The reason for the poor performance of method **H** in the UND database is that it does not restrict the processing to a small image region as the methods **A** to **F** do, since these use the RROI operation. This restriction makes the FPR small and benefits these methods when compared to the others. Note that the UND database images contain large regions with clothes: sometimes the area of the face is even smaller than that of the clothes.

From these results it is clear that it is possible to reduce processing time and simultaneously increase face segmentation accuracy in LWIR images. Note that these images present particular challenges such as face temperatures similar to background ones, as is common with the nose and ears.

5 CONCLUSION

This paper presents a study of several face segmentation methods for images captured in the LWIR. Previous proposals are evaluated and new ones are presented. The goal was to develop a method that would perform well regardless of face pose, rotation and expression. We made the evaluations using four public available databases (including 730 training and 893 test images) and used two error measures to highlight both the overall error (E_1) and the balanced error (E_2) such that the evaluation was not distorted by having a different number of test images from the different classes.

The proposed methods improve on the previous existing ones, with method **H** being the best overall. It presented the best results according to E_1 in three of the four databases and two of the four according to E_2 (and in a third database was second best by a very small amount). It improves over the existing methods in all four databases and in both error measures, with improvements of up to 28.8% according to E_1 and 38.3% according to E_2 .

Some of the other proposals improve in terms of processing time dramatically (methods **A** and **B**) and still beat the previously existing methods or are only marginally worst, and can be used if time constraints are critical such as in real-time systems.

We thus believe that the paper presents an important contribution to the state-of-the-art in LWIR face segmentation and that it can help to improve face recognition methods in the LWIR.

ACKNOWLEDGMENTS

We acknowledge the financial support given by "FCT - Fundação para a Ciência e Tecnologia" and "FEDER" in the scope of the PTDC/EIA/69106/2006 research project "BIOREC: Non-Cooperative Biometric Recognition".

We wish to thank Professor Cho Siu-Yeung David from the School of Computer Engineering at Nanyang Technological University (NTU) for the source code of his method [10].

REFERENCES

- [1] A. Ross, K. Nandakumar, and A. Jain, *Handbook of Multibiometrics (International Series on Biometrics)*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.
- [2] A. Jain, P. Flynn, and A. Ross, *Handbook of Biometrics*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2007.
- [3] K. Bowyer, K. Chang, and P. Flynn, "A survey of approaches to three-dimensional face recognition," *17th International Conference on Pattern Recognition (ICPR 2004)*, vol. 1, pp. 358–361, 2004.
- [4] S. Kong, J. Heo, B. Abidi, J. Paik, and M. Abidi, "Recent advances in visual and infrared face recognition - a review," *Computer Vision and Image Understanding*, vol. 97, no. 1, pp. 103–135, 2005.
- [5] M. Akhloufi, A. Bendada, and J. Batsale, "State of the art in infrared face recognition," *Quantitative InfraRed Thermography Journal*, vol. 5, no. 1, pp. 3–26, 2008.
- [6] K. Sobottka and I. Pitas, "Segmentation and tracking of faces in color images," in *Second International Conference on Automatic Face and Gesture Recognition*. 1996, pp. 236–241, IEEE Comput. Soc. Press.
- [7] C. Butakoff and A. Frangi, "Multi-view face segmentation using fusion of statistical shape and appearance models," *Computer Vision and Image Understanding*, vol. In Press,, pp. –, 2009.
- [8] H. Proença and L. Alexandre, "Iris recognition: Analysis of the error rates regarding the accuracy of the segmentation stage," *Image and Vision Computing*, vol. 28, pp. 202–206, 2010.
- [9] A. Gyaourova, G. Bebis, and I. Pavlidis, "Fusion of infrared and visible images for face recognition," *Lecture Notes in Computer Science*, pp. 456–468, 2004.
- [10] S. Cho, L. Wang, and W. Ong, "Thermal imprint feature analysis for face recognition," *IEEE International Symposium on Industrial Electronics*, , no. ISIE, pp. 1875–1880, 2009.
- [11] I. Pavlidis, P. Tsiamyrtzis, C. Manohar, and P. Buddharaju, *Biometrics: Face recognition in thermal infrared*, chapter 29, pp. 1–15, CRC Press, 2006.
- [12] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 10, no. 2, pp. 266–277, 2001.
- [13] D. Mumford and J. Shah, "Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems," *Comm. Pure Applied Math.*, vol. 42, pp. 577–685, 1989.
- [14] L. A. Vese and T. F. Chan, "A multiphase level set framework for image segmentation using the Mumford and Shah model," *International Journal of Computer Vision*, vol. 50, no. 3, pp. 271–293, 2002.
- [15] D. Freedman, R. J. Radke, D. M. Lovelock, and G. T. Y. Chen, "Model-based segmentation of medical imagery by matching distributions," *IEEE Transactions on Medical Imaging*, vol. 24, no. 3, pp. 281–292, 2005.



Fig. 10. Sample test set results for images from the UND database (columns 1 and 2), Terravic database (columns 3 and 4), IRIS database (columns 5 and 6) and FSU database (last two columns). Original images are in the first row and manual masks in the second. Approaches by [10], [11] and [12] are in rows 3 to 5 respectively. The remaining rows have the results of our proposals **A** to **H**, respectively.



Fig. 11. Sample test set results for images from the UND database (columns 1 and 2), Terravic database (columns 3 and 4), IRIS database (columns 5 and 6) and FSU database (last two columns). Original images are in the first row and manual masks in the second. The remaining rows have the results of the fusions **Majority Vote**, **AND** and **OR** between the results of methods **A**, **F** and **H**, respectively.

- [16] S. Shah, A. Abaza, A. Ross, and H. Ammar, "Automatic Tooth Segmentation Using Active Contour Without Edges," in *2006 Biometrics Symposium: Special Session on Research at the Biometric Consortium Conference*, Baltimore, MD, 2006, pp. 1–6, IEEE.
- [17] J. Canny, "A computational approach to edge detection," *IEEE Trans. PAMI*, vol. 8, no. 6, pp. 628–633, September 1986.
- [18] P. Flynn, K. Bowyer, and P. Phillips, "Assessment of Time Dependency in Face Recognition: An Initial Study," *Lecture Notes in Computer Science*, vol. 2688, pp. 44–51, 2003.
- [19] X. Chen, P. Flynn, and K. Bowyer, "IR and visible light face recognition," *Computer Vision and Image Understanding*, vol. 99, pp. 332–358, 2005.
- [20] "IEEE OTCBVS WS Series Bench; Roland Mieziako, Terravic Research Infrared Database.
- [21] "IEEE OTCBVS WS Series Bench; DOE University Research Program in Robotics under grant DOE-DE-FG02-86NE37968; DOD/TACOM/NAC/ARC Program under grant R01-1344-18; FAA/NSSA grant R01-1344-48/49; Office of Naval Research under grant #N000143010022.3.
- [22] A. Srivastava and X. Liu, "Statistical hypothesis pruning for identifying faces from infrared images," *Image and Vision Computing*, vol. 21, pp. 651–661, 2003.

Referências

- [1] *15th Iberoamerican Congress on Pattern Recognition*. <http://www.ciarp.org>.
- [2] *Absolute Astronomy*. http://www.absoluteastronomy.com/topics/Infrared_vision.
- [3] *BioRec - Non-Cooperative Multimodal Biometric Recognition*. <http://www.di.ubi.pt/~lfbaa/biorec.html>.
- [4] *Equinox Corporation*. <http://www.equinoxsensors.com/products/HID.html>.
- [5] *IEEE OTCBVS WS Series Bench; DOE University Research Program in Robotics under grant DOE-DE-FG02-86NE37968; DOD/TACOM/NAC/ARC Program under grant R01-1344-18; FAA/NSSA grant R01-1344-48/49; Office of Naval Research under grant #N000143010022.3*. <http://www.cse.ohio-state.edu/otcbvs-bench/>.
- [6] *IEEE OTCBVS WS Series Bench; Roland Mieziako, Terravic Research Infrared Database*. <http://www.cse.ohio-state.edu/otcbvs-bench/>.
- [7] *IEEE Transaction on Pattern Analysis and Machine Intelligence*. <http://www.computer.org/portal/web/tpami/>.
- [8] *Open Computer Vision Library (OpenCV)*. <http://sourceforge.net/projects/opencvlibrary/>.
- [9] *Weka Machine Learning Project*. <http://www.cs.waikato.ac.nz/~ml/index.html>.

- [10] B. Abidi, S. Huq, and M. Abidi. Fusion of visual, thermal, and range as a solution to illumination and pose restrictions in face recognition. In *International Carnahan Conference on Security Technology*, pages 325–330, 2004.
- [11] Y. Adini, Y. Moses, and S. Ullman. Face recognition: the problem of compensating for changes in illumination direction. *Transactions on Pattern Analysis and Machine Intelligence*, 19(7):721–732, 1997.
- [12] M. Akhloufi, A. Bendada, and J. Batsale. State of the art in infrared face recognition. *Quantitative InfraRed Thermography Journal*, 5(1):3–26, 2008.
- [13] U. Ali, J. Taj, T. Hussain, and M. Hussain. Real-Time Efficient Parallel Thermal and Visual Face Recognition Fusion. In *International Conference on Electro/information Technology*, pages 569–574, East Lansing, MI, 2006.
- [14] O. Arandjelović, R. Hammoud, and R. Cipolla. Multi-Sensory Face Biometric Fusion (for Personal Identification). In *Computer Vision and Pattern Recognition Workshop*, page 128, 2006.
- [15] O. Arandjelović, R. Hammoud, and R. Cipolla. On Person Authentication by Fusing Visual and Thermal Face Biometrics. In *International Conference on Video and Signal Based Surveillance*, page 50, Sydney, Australia, 2006.
- [16] O. Arandjelović, R. Hammoud, and R. Cipolla. *Towards Person Authentication by Fusing Visual and Thermal Face Biometrics*, chapter 6, pages 75–90. Signals and Communication Technology. Springer Berlin Heidelberg, 2007.
- [17] R. Baron. Mechanisms of human facial recognition. *International Journal of Man-Machine Studies*, 15(2):137–178, 1981.
- [18] G. Bebis, A. Gyaourova, S. Singh, and I. Pavlidis. Face recognition by fusing thermal infrared and visible imagery. *Image and Vision Computing*, 24(7):727–742, 2006.
- [19] J. Ben-arie and D. Nandy. A volumetric/iconic frequency domain representation for objects with application for pose invariant face recognition. *Transactions on Pattern Analysis and Machine Intelligence*, 20(5):449–457, 1998.

- [20] F. Boughorbel, A. Koschan, and M. Abidi. Registering Multi-sensor Datasets from a Robotic Inspection Platform. In *SPIE Defense and Security Symposium*, pages 36–41, Orlando, Florida, 2004.
- [21] K. Bowyer, K. Chang, P. Flynn, and X. Chen. Face Recognition Using 2-D, 3-D, and Infrared: Is Multimodal Better Than Multisample? In *Proceedings of the IEEE*, volume 94, pages 2000–2012, 2006.
- [22] P. Buddharaju and I. Pavlidis. *Multispectral Face Recognition: Fusion of Visual Imagery with Physiological Information*, chapter 7, pages 91–108. Springer Berlin Heidelberg, 2007.
- [23] P. Buddharaju, I. Pavlidis, and I. Kakadiaris. Face Recognition in the Thermal Infrared Spectrum. In *Computer Vision and Pattern Recognition Workshop*, page 133, 2004.
- [24] P. Buddharaju, I. Pavlidis, and P. Tsiamyrtzis. Physiology-based face recognition. In *Advanced Video and Signal Based Surveillance*, pages 354–359, 2005.
- [25] P. Buddharaju, I. Pavlidis, and P. Tsiamyrtzis. Pose-Invariant Physiological Face Recognition in the Thermal Infrared Spectrum. In *Computer Vision and Pattern Recognition Workshop (CVPRW'06)*, page 53, 2006.
- [26] P. Buddharaju, I. Pavlidis, P. Tsiamyrtzis, and M. Bazakos. Physiology-based face recognition in the thermal infrared spectrum. *Transactions on Pattern Analysis and Machine Intelligence*, 29(4):613–626, 2007.
- [27] C. Butakoff and A. F. Frangi. Multi-view face segmentation using fusion of statistical shape and appearance models. *Computer Vision and Image Understanding*, 114(3):311–321, 2009.
- [28] J. Canny. A computational approach to edge detection. *IEEE Trans. PAMI*, 8(6):628–633, September 1986.
- [29] J. Center. Practical application of Facial recognition: Automated Facial Recognition Access Control system. In H. Wechsler, J. P. Phillips, V. Bruce, F. F. Soulie, T. S. Huang, editor, *Face recognition: From theory to applications*, pages 402–408, Springer-Verlag, Berlin, Germany, 1998.

- [30] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, January 2001.
- [31] H. Chang, H. Harishwaran, M. Yi, A. Koschan, B. Abidi, and M. Abidi. An indoor and outdoor, multimodal, multispectral and multi-illuminant database for face recognition. In *Computer Vision and Pattern Recognition Workshop*, page 54, 2006.
- [32] H. Chang, M. Yi, H. Harishwaran, B. Abidi, A. Koschan, and M. Abidi. Multispectral fusion for indoor and outdoor face authentication. In *Biometrics Symposium: Special Session on Research at the Biometric Consortium Conference*, pages 1–6, Baltimore, MD, 2006.
- [33] X. Chen, P. Flynn, and K. Bowyer. PCA-based face recognition in infrared imagery: baseline and comparative studies. In *International Workshop on Analysis and Modeling of Faces and Gestures*, pages 127–134, 2003.
- [34] X. Chen, P. Flynn, and K. Bowyer. IR and visible light face recognition. *Computer Vision and Image Understanding*, 99(3):332–358, 2005.
- [35] S. Cho, L. Wang, and W. Ong. Thermal imprint feature analysis for face recognition. In *International Symposium on Industrial Electronics*, pages 1875–1880, Seoul, 2009.
- [36] I. Craw, N. Costen, T. Kato, and S. Akamatsu. How should we represent faces for automatic recognition? *Transactions on Pattern Analysis and Machine Intelligence*, 21(8):725–736, 1999.
- [37] R. Cutler. Face recognition using infrared images and eigenfaces. Technical report, University of Maryland, 1996.
- [38] J. Dowdall, I. Pavlidis, and G. Bebis. A face detection method based on multi-band feature extraction in the near-IR spectrum. In *Workshop on Computer Vision Beyond the Visible Spectrum*, Kauai, Hawaii, 2001.
- [39] J. Dowdall, I. Pavlidis, and G. Bebis. Face detection in the near-IR spectrum. *Image and Vision Computing*, 21(7):565–578, 2003.

- [40] K. Etemad and R. Chellappa. Discriminant analysis for recognition of human face images. In *First International Conference on Audio- and Video-Based Biometric Person Authentication*, pages 125–142. Springer-Verlag London, UK, 1997.
- [41] P. Flynn, K. Bowyer, and P. Phillips. Assessment of Time Dependency in Face Recognition: An Initial Study. *Lecture Notes in Computer Science*, 2688:44–51, 2003.
- [42] K. Foster. Thermographic detection of breast cancer. *Engineering in Medicine and Biology Magazine*, 17(6):10–14, 1998.
- [43] D. Freedman, R. J. Radke, D. M. Lovelock, and G. T. Y. Chen. Model-based segmentation of medical imagery by matching distributions. *IEEE Transactions on Medical Imaging*, 24(3):281–292, 2005.
- [44] G. Friedrich and Y. Yeshurun. Seeing people in the dark: face recognition in infrared images. In *Second International Workshop on Biologically Motivated Computer Vision*, pages 348–359. Springer-Verlag London, UK, 2002.
- [45] A. Gyaourova, G. Bebis, and I. Pavlidis. Fusion of infrared and visible images for face recognition. In *European Computer Vision Conference (ECCV)*, pages 456–468, 2004.
- [46] H. Hariharan, A. Koschan, B. Abidi, A. Gribok, and M. Abidi. Fusion of Visible and Infrared Images using Empirical Mode Decomposition to Improve Face Recognition. In *International Conference on Image Processing*, pages 2049–2052, Atlanta, GA, 2006.
- [47] J. Heo, B. Abidi, S. Kong, and M. Abidi. Performance Comparison of Visual and Thermal Signatures for Face Recognition. In *The Biometric Consortium Conference*, Crystal City, VA, 2003.
- [48] J. Heo, B. Abidi, J. Paik, and M. Abidi. Face recognition: evaluation report for Facelt identification and surveillance. In *Sixth International Conference on Quality Control by Artificial Vision*, volume 5132, pages 551–558, 2003.
- [49] J. Heo, S. Kong, B. Abidi, and M. Abidi. Fusion of Visual and Thermal Signatures with Eyeglass Removal for Robust Face Recognition. In *Computer Vision and Pattern Recognition Workshop*, page 122, 2004.

- [50] J. Heo, M. Sawvides, and B. Vijayakumar. *Advanced Correlation Filters for Face Recognition Using Low-Resolution Visual and Thermal Imagery*, pages 1089 – 1097. Springer-Verlag Berlin Heidelberg, 2005.
- [51] J. Heo, M. Sawvides, and B. V. K. Vijayakumar. Performance Evaluation of Face Recognition using Visual and Thermal Imagery with Advanced Correlation Filters. In *Computer Vision and Pattern Recognition Workshop*, page 9, San Diego, CA, USA, 2005.
- [52] W. Jiuqing and L. Xingshan. PCB infrared thermal imaging diagnosis using support vector classifier. In *4th World Congress on Intelligent Control and Automation*, pages 2718–2722, 2002.
- [53] K. Jonsson, J. Matas, J. Kittler, and Y. P. Li. Learning support vectors for face verification and recognition. In *Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 208–213, Grenoble, 2000.
- [54] I. Kakadiaris, G. Passalis, T. Theoharis, G. Toderici, I. Konstantinidis, and N. Murtuza. 8D - THERMO CAM: Combination of Geometry with Physiological Information for Face Recognition. In *Computer Vision and Pattern Recognition*, pages 1183–1183, 2005.
- [55] I. Kakadiaris, G. Passalis, T. Theoharis, G. Toderici, I. Konstantinidis, and N. Murtuza. Multimodal Face Recognition: Combination of Geometry with Physiological Information. In *Computer Vision and Pattern Recognition*, pages 1022–1029, 2005.
- [56] J. Kang, A. Borkar, A. Yeung, N. Nong, M. Smith, and M. Hayes. Short wavelength infrared face recognition for personalization. In *International Conference on Image Processing*, pages 2757–2760, Atlanta, GA, 2006.
- [57] S. Kong, J. Heo, B. Abidi, J. Paik, and M. Abidi. Recent advances in visual and infrared face recognition - a review. *Computer Vision and Image Understanding*, 97(1):103–135, 2005.
- [58] S. Kong, J. Heo, F. Boughorbel, Y. Zheng, B. Abidi, A. Koschan, M. Yi, and M. Abidi. Multiscale Fusion of Visible and Thermal IR Images for Illumination-Invariant Face Recognition. *International Journal of Computer Vision*, 71(2):215–233, 2006.

- [59] S. Li, L. Zhang, S. Liao, X. Zhu, R. Chu, M. Ao, and R. He. A Near-infrared Image Based Face Recognition System. In *7th International Conference on Automatic Face and Gesture Recognition*, pages 455–460, Southampton, 2006.
- [60] K. Lin, W. Chen, S. Chen, and F. Wang. Diagnostic technique for classifying the quality of circuit boards using infrared thermal image. In *9th IFSA World Congress and 20th NAFIPS International Conference*, pages 464–469, Vancouver, BC, 2001.
- [61] C. Liu and H. Wechsler. A unified Bayesian framework for face recognition. In *International Conference on Image Processing*, pages 151–155, Chicago, IL, 1998.
- [62] C. Liu and H. Wechsler. Comparative assessment of independent component analysis (ICA) for face recognition. In *Second International Conference on Audio- and Video-based Biometric Person Authentication*, Washington, DC, 1999.
- [63] J. Lu, K. Plataniotis, A. Venetsanopoulos, and S. Li. Ensemble-based discriminant learning with boosting for face recognition. *Transactions on Neural Networks*, 17(1):166–178, 2006.
- [64] B. Miller. Vital signs of identity [biometrics]. *IEEE Spectrum*, 31(2):22–30, 1994.
- [65] D. Mumford and J. Shah. Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems. *Comm. Pure Applied Math.*, 42:577–685, 1989.
- [66] S. Osher and J. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics*, 79(1):12–49, 1988.
- [67] I. Pavlidis, J. Levine, and P. Baukol. Thermal image analysis for anxiety detection. In *International Conference on Image Processing*, pages 315–318, Thessaloniki, 2001.
- [68] I. Pavlidis and P. Symosek. The imaging issue in an automatic face/disguise detection system. In *Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications*, pages 15–24, Hilton Head, SC, 2000.

- [69] I. Pavlidis, P. Tsiamyrtzis, C. Manohar, and P. Buddharaju. *Biometrics: Face recognition in thermal infrared*, chapter 29, pages 1–15. CRC Press, 3 edition, 2006.
- [70] A. Pentland and T. Choudhury. Face recognition for smart environments. *Computer*, 33(2):50–55, 2000.
- [71] H. Proença. On the feasibility of the visible wavelength, at-a-distance and on-the-move iris recognition. In *Workshop on Computational Intelligence in Biometrics: Theory, Algorithms, and Applications*, pages 9–15, Nashville, TN, March 2009.
- [72] H. Proença and L. A. Alexandre. Iris recognition: Analysis of the error rates regarding the accuracy of the segmentation stage. In *International Conference on Computational Intelligence and Security*, pages 1771–1774, Guangzhou, 2006.
- [73] H. Proença. *Towards non-cooperative biometric iris recognition*. Ph.d., University of Beira Interior, 2006.
- [74] F. Prokoski. History, current status, and future of infrared identification. In *Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications*, pages 5–14, Hilton Head, SC, 2000.
- [75] G. Santos and H. Proença. On the Role of Interpolation in the Normalization of Non-ideal Visible Wavelength Iris Images. In *International Conference on Computational Intelligence and Security*, pages 315–319, Beijing, 2009.
- [76] A. Selinger and D. Socolinsky. Appearance-based facial recognition using visible and thermal imagery: a comparative study. Technical Report 02-01, Equinox Corporation, 2002.
- [77] S. Shah, A. Abaza, A. Ross, and H. Ammar. Automatic Tooth Segmentation Using Active Contour Without Edges. In *2006 Biometrics Symposium: Special Session on Research at the Biometric Consortium Conference*, pages 1–6, Baltimore, MD, 2006. IEEE.
- [78] M. Shirvaikar and M. Trivedi. A neural network filter to detect small targets in high clutter backgrounds. *Transactions on Neural Networks*, 6(1):252–257, 1995.

- [79] R. Siddiqui, M. Sher, and K. Rashid. Face Identification based on Biological Trait using Infrared Images after Cold Effect Enhancement and Sunglasses Filtering. In *12th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, pages 161–164, University of West Bohemia, Campus Bory, Plzen-Bory, Czech Republic, 2004.
- [80] T. Sim, R. Sukthankar, M. Mullin, and S. Baluja. Memory-based face recognition for visitor identification. In *Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 214–220, Grenoble, 2000.
- [81] R. Singh, M. Vatsa, and A. Noore. Hierarchical fusion of multi-spectral face images for improved recognition performance. *Information Fusion*, 9(2):200–210, 2008.
- [82] S. Singh, A. Gyaourova, G. Bebis, and I. Pavlidis. Infrared and visible image fusion for face recognition. In *SPIE Defense and Security Symposium (Biometric Technology for Human Identification)*, volume 5404, pages 585–596, 2004.
- [83] K. Sobottka and I. Pitas. Segmentation and tracking of faces in color images. In *Second International Conference on Automatic Face and Gesture Recognition*, pages 236–241, Killington, VT, 1996.
- [84] D. Socolinsky, C. Eveland, L. Wolff, and J. Neuheisel. Illumination invariant face recognition using thermal infrared imagery. In *Computer Vision and Pattern Recognition*, pages 527–534, 2001.
- [85] D. Socolinsky and A. Selinger. A comparative analysis of face recognition performance with visible and thermal infrared imagery. In *16th International Conference on Pattern Recognition*, pages 217–222, 2002.
- [86] D. Socolinsky and A. Selinger. Thermal face recognition in an operational scenario. In *Computer Vision and Pattern Recognition*, pages 1012–1019, 2004.
- [87] D. Socolinsky and A. Selinger. Thermal face recognition over time. In *17th International Conference on Pattern Recognition*, pages 187–190, 2004.
- [88] D. Socolinsky, A. Selinger, and J. Neuheisel. Face recognition with visible and thermal infrared imagery. *Computer Vision and Image Understanding*, 91(1-2):72–114, 2003.

- [89] D. Socolinsky, L. Wol, and C. Eveland. *Face Recognition in the Thermal Infrared*. Advances in Pattern Recognition. Springer London, 2005.
- [90] A. Srivastava and X. Liu. Statistical hypothesis pruning for identifying faces from infrared images. *Image and Vision Computing*, 21:651–661, 2003.
- [91] A. Srivastava, X. Liu, B. Thomasson, and C. Heshner. Spectral Probability Models for IR Images with Applications to IR Face Recognition. In *Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications*, Kauai, Hawaii, 2001.
- [92] Y. Tian, T. Kanade, and J. Cohn. Recognizing action units for facial expression analysis. *Transactions on Pattern Analysis and Machine Intelligence*, 23(2):97–115, 2001.
- [93] M. Turk and A. Pentland. Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition*, volume 591, pages 586–591, Maui, HI, 1991.
- [94] L. A. Vese and T. F. Chan. A multiphase level set framework for image segmentation using the Mumford and Shah model. *International Journal of Computer Vision*, 50(3):271–293, 2002.
- [95] J. Wilder, P. J. Phillips, and S. Wiener. Comparison of visible and infra-red imagery for face recognition. In *Second International Conference on Automatic Face and Gesture Recognition*, pages 182–187, Killington, VT, 1996.
- [96] L. Wolff, D. Socolinsky, and C. Eveland. Quantitative measurement of illumination invariance for face recognition using thermal infrared imagery. In *Workshop on Computer Vision Beyond the Visible Spectrum*, pages 140–151, Kauai, Hawaii, 2003.
- [97] S. Wu, Z. Fang, Z. Xie, and W. Liang. *Blood Perfusion Models for Infrared Face Recognition*, chapter 13, pages 183–206. I-Tech, Vienna, Austria, 2008.
- [98] S. Wu, W. Song, L. Jiang, S. Xie, F. Pan, W. Yau, and S. Ranganath. *Infrared face recognition by using blood perfusion data*, volume 3546, pages 320–328. Springer Berlin / Heidelberg, 2005.

- [99] Y. Yoshitomi, T. Miyaura, S. Tomita, and S. Kimura. Face identification using thermal image processing. In *6th IEEE International Workshop on Robot and Human Communication*, pages 374–379, Sendai, 1997.
- [100] S. Zhao and R. Grigat. An Automatic Face Recognition System in the Near Infrared Spectrum. *Machine Learning and Data Mining in Pattern Recognition*, 3587:437–444, 2005.
- [101] X. Zou, J. Kittler, and K. Messer. Ambient Illumination Variation Removal by Active Near-IR Imaging. *Advances in Biometrics*, 3832:19–25, 2005.
- [102] X. Zou, J. Kittler, and K. Messer. Face Recognition Using Active Near-IR Illumination. In *British Machine Vision Conference*, pages 153–164, Oxford Brookes University, Oxford, U.K., 2005.