Técnicas de processamento de imagens para localização e reconhecimento de faces

Osvaldo Cesar Pinheiro de Almeida

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura:

Técnicas de processamento de imagens para localização e reconhecimento de faces

Osvaldo Cesar Pinheiro de Almeida

Orientador: Prof. Dr. João do Espírito Santo Batista Neto

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação - ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Ciências - Ciências de Computação e Matemática Computacional.

USP - São Carlos Outubro/2006

Dedicatória

Aos meus pais Osvaldo e Cida, meus irmãos Daniel, Carla e Selma, e a minha amável companheira Raquel, por me apoiarem sempre.

Agradecimentos

Ao Prof. João Batista pela orientação, ajuda e incentivo dados durante tantos anos.

Aos meus amigos de república Bruno, Carlos, Carrijo, Diego e Leandro pelos bons momentos que passamos juntos.

A minha prima Juliana que me ofereceu abrigo e companheirismo.

A minha querida Raquel por estar sempre me apoiando, e me auxiliando na revisão do trabalho.

Aos meus irmãos que sempre confiaram em mim, e me ajudaram em todos os momentos.

Aos meus pais pelo grande esforço que fizeram para que eu pudesse me formar e chegar até essa etapa de minha vida, e por sempre acreditarem em mim.

A DEUS por sempre me dar forças, ânimo e coragem nos momentos em que mais precisei.

Resumo

A biometria é a ciência que estuda a mensuração dos seres vivos. Muitos trabalhos exploram as características dos seres humanos tais como, impressão digital, íris e face, a fim de desenvolver sistemas biométricos, utilizados em diversas aplicações (monitoramento de segurança, computação ubíqua, robótica). O reconhecimento de faces é uma das técnicas biométricas mais investigadas, por ser bastante intuitiva e menos invasiva que as demais. Alguns trabalhos envolvendo essa técnica se preocupam apenas em localizar a face de um indivíduo (fazer a contagem de pessoas), enquanto outros tentam identificá-lo a partir de uma imagem. Este trabalho propõe uma abordagem capaz de identificar faces a partir de quadros de vídeo e, posteriormente, reconhecê-las por meio de técnicas de análise de imagens. Pode-se dividir o trabalho em dois módulos principais: (1) – Localização e rastreamento de faces em uma seqüência de imagens (frames), além de separar a região rastreada da imagem; (2) - Reconhecimento de faces, identificando a qual pessoa pertence. Para a primeira etapa foi implementado um sistema de análise de movimento (baseado em subtração de quadros) que possibilitou localizar, rastrear e captar imagens da face de um indivíduo usando uma câmera de vídeo. Para a segunda etapa foram implementados os módulos de redução de informações (técnica Principal Component Analysis – PCA), de extração de características (transformada wavelet de Gabor), e o de classificação e identificação de face (distância Euclidiana e Support Vector Machine - SVM). Utilizando-se duas bases de dados de faces (FERET e uma própria - Própria), foram realizados testes para avaliar o sistema de reconhecimento implementado. Os resultados encontrados foram satisfatórios, atingindo 91,92% e 100,00% de taxa de acertos para as bases FERET e Própria, respectivamente.

Abstract

Biometry is the science of measuring and analyzing biomedical data. Many works in this field have explored the characteristics of human beings, such as digital fingerprints, iris, and face to develop biometric systems, employed in various aplications (security monitoring, ubiquitous computation, robotic). Face identification and recognition are very apealing biometric techniques, as it it intuitive and less invasive than others. Many works in this field are only concerned with locating the face of an individual (for counting purposes), while others try to identify people from faces. The objective of this work is to develop a biometric system that could identify and recognize faces. The work can be divided into two major stages: (1) Locate and track in a sequence of images (frames), as well as separating the tracked region from the image; (2) Recognize a face as belonging to a certain individual. In the former, faces are captured from frames of a video camera by a motion analysis system (based on substraction of frames), capable of finding, tracking and croping faces from images of individuals. The later, consists of elements for data reductions (*Principal Component Analysis* – PCA), feature extraction (Gabor wavelets) and face classification (Euclidean distance and Support Vector Machine – SVM). Two faces databases have been used: FERET and a "home-made" one. Tests have been undertaken so as to assess the system's recognition capabilities. The experiments have shown that the technique exhibited a satisfactory performance, with success rates of 91.97% and 100%for the FERET and the "home-made" databases, respectively.

Sumário

T	Intr	odução	1
	1.1	Motivação	3
	1.2	Objetivo e Justificativa	3
	1.3	Organização	4
2	Bio	metria	5
	2.1	Sistemas Biométricos	6
	2.2	Técnicas de Biometria	8
		2.2.1 Impressão Digital	8
		2.2.2 Íris	9
		2.2.3 Retina	11
		2.2.4 Geometria da mão, dedos e palma da mão	12
		2.2.5 Voz	13
		2.2.6 Assinatura Manuscrita	13
		2.2.7 Dinâmica de Digitação	14
	2.3	Aplicações de Sistemas Biométricos	15
	2.4	Considerações Finais	17

3	Ras	tream	ento e Reconhecimento de Faces	19
	3.1	Carac	terísticas	21
	3.2	Locali	zação e Rastreamento	23
	3.3	Recon	hecimento e Identificação	26
	3.4	Avalia	ção de Algoritmos de Reconhecimento	29
	3.5	Consid	derações Finais	31
4	Téc	nicas j	para Reconhecimento de Face	33
	4.1	Introd	ução a Transformada Wavelet	34
		4.1.1	Transformada de Fourier	35
		4.1.2	Transformada Wavelet	37
		4.1.3	Funções Wavelets Mães	38
		4.1.4	Wavelets Bidimensionais	40
		4.1.5	Exemplos de Wavelets	40
		4.1.6	Aplicações	41
		4.1.7	Considerações Finais Sobre as Wavelets	43
	4.2	Introd	ução a PCA	44
	4.3	Introd	ução ao Aprendizado de Máquina SVM	49
		4.3.1	Aprendizado de Máquina	49
		4.3.2	SVMs	51
		4.3.3	SVMs Lineares com Margens Rígidas	53
		4.3.4	SVMs Lineares com Margens Suaves	54
		4.3.5	SVMs Não Lineares	55
		4.3.6	SVMs Multiclasses	57
		4.3.7	Utilizando SVM - libsvm	57
		4.3.8	Considerações Finais Sobre a SVM	58

5	Me_1	todolog	gia	61
	5.1	Rastre	eamento de Face	62
	5.2	Recon	hecimento de Face	65
		5.2.1	Classificação por Distância Euclidiana	67
		5.2.2	Classificação por SVM	67
		5.2.3	Bases de imagens	68
6	Res	ultado	${f s}$	73
	6.1	Result	ados do Rastreamento de Face	73
	6.2	Result	ados do Reconhecimento de Face	74
		6.2.1	Classificação por Distância Euclidiana	75
		6.2.2	Classificação por SVM	77
	6.3	Discus	ssão dos Resultados	78
7	Cor	ıclusão		81
$\mathbf{R}_{\mathbf{c}}$	eferê	ncias I	Bibliográficas	94

Lista de Figuras

2.1	(a) - Imagem binária de uma impressão digital; (b) - Uma impressão digital após a identificação das minúcias (linhas mais claras)	9
2.2	Imagem ilustrativa do olho humano com a íris em destaque. As informações mais importantes sobre a íris estão relacionadas a sua textura, e não a sua cor	10
2.3	Imagem ilustrativa do fundo de um olho humano, onde podem ser observados os vasos sangüíneos da retina. A quantidade e disposição dos vasos são suas informações mais importantes	11
3.1	Alguns pontos de interesse para o reconhecimento de face. Algumas características que podem ser analisadas são, por exemplo, as distâncias entre os pontos marcados na imagem (distância entre olhos, boca e nariz)	22
4.1	Análise de dois sinais por FT e CWT. A Figura (a) consiste de um sinal de duas freqüências ($sen(10t)$ e $sen(20t)$) sobrepostas, e (b) consiste de um sinal das mesmas freqüências ocupando metade do tempo do sinal cada uma. As Figuras (c) e (d) são os espectros dos sinais de (a) e (b), respectivamente. As Figuras (e) e (f) mostram a magnitude da transformada wavelet (wavelet de Morlet[Morlet, 1981]) dos sinais (a) e (b), respectivamente[Protázio, 2001]	36
4.2	Cobertura no plano de freqüência espacial de um conjunto de wavelets de	40
	Gabor[Lee, 1996]. Usando três escalas e oito orientações	42

4.3	Filtros de respostas ortogonais, após aplicar técnica de eliminação de redundância[Manjunath and Ma, 1996]. Usando quatro escalas e seis orientações	43
4.4	Imagem com um exemplo gráfico de um conjunto de dados em \Re^3 , onde as amostras estão todas em um mesmo plano \Re^2 [Mendonça, 2005]	46
4.5	Exemplo gráfico de um conjunto de treinamento binário e três modelos de classificadores gerados[Smola and Schölkpof, 2002]	52
4.6	Exemplo de um conjunto de treinamento binário dividido por um hiperplano, e suas margens de separação (região entre esse hiperplano e cada hiperplano H_1 e H_2)[Hearst et al., 1998]	54
4.7	Exemplo de um conjunto de treinamento binário com alguns tipos de SVs: livres (em cinza), e limitados (em preto)[Passerini, 2004]	55
4.8	Exemplo de um conjunto de treinamento binário. (a)Dados não lineares; (b)Fronteira não linear de classificação; (c)Fronteira linear de classificação no espaço de características[Müller et al., 2001]	56
5.1	Ilustração do funcionamento do sistema de rastreamento e reconhecimento de face proposto	62
5.2	Imagens de uma face: (a) Imagem original do ambiente que estava sendo filmado; (b) - Imagem com o rosto inteiro do indivíduo; (c) Imagem contendo apenas a região central da face	63
5.3	Ilustração do funcionamento do sistema de localização e rastreamento de face proposto	64
5.4	Ilustração do funcionamento do sistema de reconhecimento de face proposto.	65
5.5	Exemplo de uma imagem do banco de imagens FERET: (a) - Imagem original do banco; (b) - Imagem com o rosto inteiro do indivíduo; (c) Imagem contendo apenas a região central da face	69
6.1	Seqüência de imagens do processo de localização e rastreamento de face implementado	74
6.2	Seqüência de imagens (a) e (c) quando há falha no sistema de localização e rastreamento, e suas respectivas imagens de face capturadas (b) e (d)	74

Lista de Tabelas

2.1	Tabela de comparação entre as técnicas de biometria discutidas [Hong, 1998].	18
4.1	Tabela com as funções Kernels mais comuns e seus parâmetros	57
5.1	Grupos de avaliação e seus atributos quanto a processo de extração de características, processo de classificação e base de imagens	70
6.1	Tabela com as bases de imagens e os resultados obtidos com o processo de verificação usando wavelet de Gabor e distância Euclidiana	76
6.2	Tabela com as bases de imagens e os resultados obtidos com a avaliação dos falsos positivos, para o processo de verificação usando wavelet de Gabor e distância Euclidiana	76
6.3	Tabela com as bases de imagens e os resultados obtidos com o processo de verificação usando PCA com $wavelet$ de Gabor e distância Euclidiana	76
6.4	Tabela com as bases de imagens e os resultados obtidos com a avaliação dos falsos positivos, para o processo de verificação usando PCA com wavelet de Gabor e distância Euclidiana	77
6.5	Tabela com as bases de imagens e os resultados obtidos com o processo de identificação usando wavelet de Gabor e distância Euclidiana	77
6.6	Tabela com as bases de imagens e os resultados obtidos com o processo de identificação usando PCA com $wavelet$ de Gabor e distância Euclidiana	77

6.7	Tabela com as bases de imagens e os resultados obtidos com o processo de	
	identificação usando $wavelet$ de Gabor e SVM	78
6.8	Tabela com as bases de imagens e os resultados obtidos com o processo de identificação usando PCA com wavelet de Gabor e SVM	78
7.1	Tabela com os resultados obtidos usando-se outras abordagens de extração e classificação de dados de imagens da base FERET, e o resultado obtido neste trabalho	84
	neste tradamo	- 64

CAPÍTULO I

Introdução

O processamento de sinais digitais é uma área do conhecimento que estuda os vários fenômenos da natureza capazes de serem representados por um sinal digital, como, por exemplo, uma imagem ou um som. A disseminação do processamento de sinais é ampla em várias áreas da ciência, como biologia, astronomia, química, física, entre outras. Isso é motivado pela constante evolução das técnicas dessa área. Como ocorre, por exemplo, na utilização de técnicas de processamento de imagem[Gonzales and Woods, 1992], um tipo de sinal, nas áreas de medicina e biometria.

A biometria é o segmento da ciência que estuda a mensuração dos seres vivos. Ou seja, por meio de técnicas biométricas é possível realizar medidas em um ser vivo, o que possibilita a análise de várias de suas características. Algumas dessas são intrínsecas a cada ser vivo, ou seja, não há uma medida igual à outra. No caso dos seres humanos, pode-se citar como exemplos, a impressão digital, a face e a voz. O uso da biometria em humanos tem como um dos principais focos a identificação de pessoas, a partir de suas características únicas. Essas podem ser tanto fisiológicas quanto comportamentais, como a impressão digital e voz, respectivamente. Utilizando técnicas de reconhecimento biométrico é possível desenvolver sistemas que realizem a identificação de pessoas para vários fins, como, por exemplo, a identificação de criminosos, por meio da impressão digital.

O uso da biometria em sistemas de reconhecimento humano tem sido estudado

e aplicado há vários anos. Contudo, somente no século XX, com o surgimento dos computadores, foi possível automatizar os sistemas, oferecendo maior precisão na identificação humana. Além disso, foi viabilizada a exploração de outras características, até então impossíveis de serem analisadas sem o auxilio de um sistema computacional, como é o caso da análise de íris. Em conseqüência disso, houve um aumento no interesse em pesquisas relacionadas à biometria humana, contribuindo para o avanço nas técnicas de identificação.

Um dos principais motivos que impulsionam as pesquisas na área de biometria é a segurança. Com a necessidade de proteger o crescente volume de informações, geradas pelos sistemas computacionais, e, locais altamente sigilosos, que possam conter segredos industriais importantes, é que se procura desenvolver sistemas de segurança altamente confiáveis. Isso é possível unindo a esses sistemas, técnicas de biometria que, em geral, são menos suscetíveis a falhas. Além disso, há o interesse em solucionar a deficiência e morosidade em identificar criminosos, com o máximo de precisão possível. Para esse fim, a identificação por meio de sistemas biométricos demonstra-se muito útil.

Para a implementação de um sistema biométrico eficiente e flexível é necessário aliá-lo a um sistema computacional. Por esse motivo, é que o desenvolvimento da biometria está intrinsecamente ligado às áreas de processamento de sinais e de extração de características. Isso porque, a captura das características humanas é realizada por sensores ópticos e/ou eletrônicos que as transformam em sinais, como imagens e sons. A partir desses sinais, muitas vezes são realizadas correções, devido aos erros inerentes ao equipamento de captura. Após essa etapa, é realizada a extração das características importantes desses sinais, as quais irão permitir a identificação de uma pessoa.

O desenvolvimento de um sistema biométrico depende da escolha de dois parâmetros principais: a característica do ser humano, e o poder computacional do sistema. Com isso, é possível dimensionar o nível de detalhes que a característica fornece, além capacidade de análise e extração das informações dos mesmos. Contudo, é preciso considerar a aplicação a qual se destina este sistema e, também, se este será para identificação ou verificação. Com isso, é possível projetar um sistema que atenda aos requisitos desejados, e que ofereça a precisão esperada.

A escolha de uma entre as várias características inerentes ao ser humano deve ser cuidadosa, pois, por existirem inúmeras, a capacidade de identificar um ser humano e o custo para se utilizá-la variam muito. Por isso, é preciso avaliar a relação custo beneficio de cada uma e, atentar para alguns aspectos teóricos importantes (universalidade, singularidade, permanência e mensurabilidade), bem como aspectos práticos (desempenho,

1.1. $MOTIVAÇ\~AO$

aceitabilidade e proteção). Esses pontos ajudam a nortear a escolha da característica.

1.1 Motivação

Muitas características biométricas estão sendo exploradas nos últimos anos. Uma das mais investigadas é a face humana, por ser altamente intuitiva para os seres humanos, além de ser um sistema não intrusivo. Vários métodos foram sugeridos para a localização e reconhecimento de faces[Yang et al., 2002]. Contudo, ainda há muitos problemas a serem solucionados, como a dificuldade de análise de imagens com variações de luminosidade, escala e posicionamento da face. Além disso, muitos trabalhos exploram apenas a localização de uma face, sem determinar a quem pertence, enquanto outros se preocupam apenas com a identificação do indivíduo. Isto torna esta área propícia para novas pesquisas.

Os principais elementos motivadores para se explorar o reconhecimento de face estão nas suas aplicações, como videoconferência, visão cibernética, monitoramento por câmeras, melhoramento de interfaces com os usuários na área de HCI (Human Computer Interaction), entre outras. A identificação de faces torna possível a personalização do ambiente computacional segundo preferências pessoais do indivíduo, tornando o ambiente mais agradável. Outra possibilidade está no uso dessas técnicas aplicadas ao desenvolvimento de visão de robôs[Kuno et al., 2003]. Isso permitiria que um robô pudesse localizar e identificar uma pessoa, possibilitando armazenar informações específicas de cada indivíduo. Há ainda, o grande apelo do sistema em aplicações de monitoramento de segurança, que permite reconhecer as pessoas e verificar se é permitido a elas estarem em determinados locais. Além dessas, existem outras aplicações que podem utilizar essa tecnologia biométrica, como poderá ser observado no Capítulo 3.

1.2 Objetivo e Justificativa

O objetivo deste trabalho é desenvolver um sistema biométrico baseado no reconhecimento de face. Para isso, o trabalho foi dividido em dois módulos: (1) localizar a face de uma pessoa em uma seqüência de imagens (captadas a partir de uma câmera de vídeo), realizar o rastreamento (tracking), que consiste em acompanhar a face (quadro a quadro), e ainda extrair a região da imagem que contém a face; (2) realizar o reconhecimento da face, extraída no processo anterior, identificando a qual pessoa essa face pertence. Assim, o

objetivo consiste em desenvolver e integrar os processos de rastreamento e reconhecimento de face.

Para atingir o objetivo proposto foram utilizadas algumas técnicas de localização e rastreamento de face, e de reconhecimento de face. No módulo de localização e rastreamento utilizou-se a técnica de análise de movimento, onde, comparando-se os quadros das imagens captadas por uma câmera de vídeo, foi possível localizar e rastrear a face de um indivíduo. Já no módulo de reconhecimento foram utilizadas a técnica de PCA (para a redução do número de informações), a transformada wavelet de Gabor (para a extração de características das imagens da face), e os classificadores por distância Euclidiana e por SVM (para identificar a qual indivíduo a face analisada pertence).

Este trabalho permitirá agregar conhecimento da área de biometria ao grupo de pesquisa, além de uma ferramenta biométrica. Isso poderá contribuir com futuros trabalhos, tanto na área de visão computacional, que pode ser explorada pelo grupo de robótica da instituição, quanto no próprio grupo de pesquisa, que pode utilizar essa ferramenta para uma aplicação específica, como reconhecimento de gestos. Ainda, o sistema pode ser explorado pelo grupo de hipermídia, que realiza pesquisas em computação ubíqua. Além disso, o sistema pode ser futuramente combinado com outras técnicas biométricas, melhorando o seu desempenho em aplicações específicas.

1.3 Organização

Para melhor descrever a proposta de trabalho, é necessário abordar com maiores detalhes alguns aspectos sobre biometria, e também sobre reconhecimento de face. Assim, no Capítulo 2, são abordados os principais aspectos da biometria, suas características gerais, os sistemas biométricos, as técnicas mais exploradas e algumas aplicações de sistemas biométricos. No Capítulo 3, serão discutidos os métodos de localização e reconhecimento de face, explorando as principais técnicas utilizadas para esses processos e para realizar o rastreamento, além de discorrer sobre algumas das características faciais mais usadas e sobre os métodos de avaliação dessas técnicas. No Capítulo 4, serão abordados alguns conceitos básicos sobre as técnicas utilizadas para o desenvolvimento desse trabalho. No Capítulo 5 será apresentado a metodologia aplicada para alcançar os objetivos propostos. Já no Capítulo 6 serão apresentados e discutidos os resultados obtidos no projeto.

Capítulo 2

Biometria

A biometria, ciência que estuda a mensuração dos seres vivos, é muito antiga. Informações biométricas como altura e cor dos olhos, são usadas há muito tempo para distinguir pessoas. Essas são maneiras de quantificar características de um determinado indivíduo, para descrevê-lo. A existência de relatos desse tipo são confirmados por artefatos arqueológicos[Hong, 1998]. Desde o final do século XIX, a impressão digital, por exemplo, é utilizada para identificação de criminosos. Todas essas informações, envolvendo os seres humanos e o uso da biometria, mostra a relevância dessa ciência.

A evolução da biometria ganhou força com o surgimento dos computadores. A partir do final do século XX, foi possível explorar a fundo algumas técnicas biométricas. A união com a tecnologia oferecida pelos computadores tornou viável pesquisar aspectos importantes, e que antes eram difíceis de se analisar de maneira adequada, como a análise da íris e da retina. Além disso, foi possível agregar mais confiabilidade e detalhamento nas análises feitas a partir de um computador, viabilizando a criação de sistemas biométricos de qualidade.

Para a desenvolvimento de um sistema biométrico é importante analisar o seu domínio de aplicação. Dessa maneira, é possível levantar qual as necessidades quanto ao poder de processamento do sistema, e também da característica que será utilizada para disponibilizar as informações biométricas. Isso indicará o nível de eficiência do sistema, assim como o custo de seu desenvolvimento. Na seção 2.1, são apresentados

detalhes sobre sistemas biométricos, como o funcionamento e as condições necessárias para serem implementados. Na seção 2.2, são descritas as principais técnicas biométricas, suas características, vantagens e desvantagens, e ainda é apresentada uma comparação entre as técnicas. Pelo reconhecimento de face ser o objeto de pesquisa desse trabalho, será explorado mais detalhadamente que as outras técnicas, no Capítulo 3. Ainda, na seção 2.3, serão discutidas algumas aplicações que utilizam sistemas biométricos.

2.1 Sistemas Biométricos

Um sistema biométrico é essencialmente um sistema de reconhecimento de padrões, de características biológicas, que realiza operações de aquisição dos dados biométricos do indivíduo, por meio de um equipamento, extração das características dos dados adquiridos, e comparação dessas características com os padrões armazenados em sua base de dados[Jain et al., 2003]. Existem duas abordagens distintas para um sistema biométrico:

- Verificação Esse sistema faz a verificação buscando na base de dados a informação sobre o usuário, que o indivíduo diz ser, autenticando-o, caso ele realmente seja esse usuário.
- Identificação Esse sistema faz a identificação, buscando-se em toda a base de dados por algum usuário que tenha se registrado com a mesma informação daquele que está sendo pesquisado, até que este seja encontrado ou não.

O sistema de verificação é tão importante quanto o de identificação, contudo exposto em contextos diferentes. No primeiro, um indivíduo diz ser uma determinada pessoa e o sistema busca em seu banco de dados, verificando se a informação que está sendo introduzida é igual à informação previamente armazenada. Esse é um sistema 1-1 (um para um), pois ele apenas confere a veracidade da informação de entrada com a armazenada. Um exemplo típico desse sistema é o de usuário/senha, onde a informação biométrica do usuário serve como a senha. Já no segundo caso, o sistema procura na sua base de dados por algum usuário que tenha a mesma informação biométrica armazenada daquela que está sendo introduzida, guardando uma certa margem de erro. Esse é um sistema 1-N (um para muitos), pois o dado de um indivíduo é comparado com o dado armazenado de vários usuários. Um exemplo de aplicação desse sistema poderia ser a procura em um banco de dados de uma criança desaparecida, por meio de sua fotografia.

Contudo, existem módulos e bases genéricas para ambos os sistemas. Esses módulos são a base de dados de informações biométricas, e o sistema de registro de usuário, que armazena as informações nessa base para serem, posteriormente, resgatadas e comparadas. Além disso, há a importante definição de qual informação biométrica será utilizada para reconhecer um indivíduo, como a digital, a íris ou a face, e qual técnica será usada para extrair as características dessa informação. Essas escolhas não dependem do contexto do sistema (verificação ou identificação), e sim de outros aspectos, que serão abordados nos próximos parágrafos.

A informação a ser utilizada para reconhecer um indivíduo pode ser, teoricamente, qualquer uma, desde que represente uma característica humana, tanto fisiológica como comportamental. Contudo, é importante seguir alguns critérios no momento da escolha dessa informação, obedecendo aos requisitos necessários, que possibilitará o desenvolvimento de um sistema de reconhecimento de pessoas. Esses requisitos são:

- Universalidade Todos os indivíduos precisam possuir a informação a ser analisada.
- Singularidade Cada indivíduo deve possuir características distintas, únicas, dessa informação.
- Permanência A informação deve ser imutável, ou seja, não pode variar com o tempo.
- Mensurabilidade Significa dizer que a informação deve ser possível de se medir.

Na prática foram reconhecidos alguns outros requisitos que são extremamente importantes para a escolha do tipo de informação a ser utilizada. Esses requisitos práticos são:

- **Desempenho** Refere-se à precisão e velocidade do reconhecimento, os recursos necessários para conseguir um reconhecimento com precisão e velocidade aceitáveis, e aos fatores operacionais e ambientais que afetam a precisão e velocidade do reconhecimento.
- Aceitabilidade Disposição das pessoas em aceitar o sistema biométrico no seu dia-a-dia.

 Proteção – Dificuldade ou facilidade de se enganar o sistema por meio de métodos fraudulentos.

Ao cumprir esses requisitos, a escolha da informação biométrica tende a ser mais confiável. Mas ainda é interessante observar as diferentes classes de informação biométricas, representadas pelas características fisiológicas e comportamentais. A primeira representa características inatas de cada ser humano. Já a segunda representa características adquiridas. Em termos de eficiência e generalidade, as que representam características fisiológicas são mais confiáveis que as comportamentais, já que a variabilidade em uma mesma classe de característica tende a ser menor nas características fisiológicas [Hong, 1998].

2.2 Técnicas de Biometria

Existem atualmente vários tipos de informações biométricas sendo utilizadas em pesquisas. Inicialmente serão mostradas as que representam características fisiológicas: digital, face, íris, retina e geometria da mão, dedos e palma. Posteriormente, as de características comportamentais: voz, assinatura manuscrita e dinâmica de digitação.

2.2.1 Impressão Digital

A impressão digital é uma das características biométricas mais populares e também uma das mais antigas[Hong, 1998]. Por apresentar características únicas em cada indivíduo, e por serem imutáveis durante toda a vida dessa pessoa, é comumente utilizada para a identificação de seres humanos, e é amplamente explorada em desenvolvimento de sistemas biométricos[Venkataramani and Kumar, 2003; Sujan and Mulqueen, 2002].

A impressão digital é constituída de uma série de traços e sulcos na região extrema dos dedos. Esses traços e sulcos formam as características inimitáveis presentes nas digitais, conhecidas como minúcias, que são formadas pelas bifurcações ou terminações entre esses traços e sulcos. A posição relativa das minúcias são atributos imutáveis e inimitáveis de um indivíduo. Pela relação determinada entre as minúcia é possível identificar e distinguir uma pessoa. A Figura 2.1 ilustra uma impressão digital e suas minúcias.

Para o desenvolvimento de um sistema de reconhecimento de impressão digital, é necessário que haja um sensor de aquisição, capaz de capturar a imagem. É importante

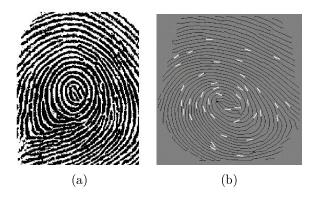


Figura 2.1: (a) - Imagem binária de uma impressão digital; (b) - Uma impressão digital após a identificação das minúcias (linhas mais claras).

destacar que o sensor deve permitir a aquisição de uma imagem de boa qualidade, já que a identificação depende dos detalhes obtidos com a representação dos traços e sulcos. A partir da imagem, é realizado um processo de afinamento dos traços e sulcos, de maneira que sejam transformados em simples linhas. Isso facilita a identificação de suas terminações e bifurcações, possibilitando encontrar as minúcias. É por meio de uma avaliação relativa das posições das minúcias que são realizadas as identificações de indivíduos. Um ponto importante a ser destacado é que a extração das minúcias não considera a disposição global dos traços e sulcos. Além disso, imagens com baixa resolução de detalhes podem dificultar a identificação das minúcias e, conseqüentemente, levar a uma falha de identificação.

A maior desvantagem do uso de sistemas biométricos baseados em impressão digital está na possibilidade desse sistema ser facilmente burlado. Cientistas japoneses, da Universidade de Yokohama, conseguiram burlar um sistema desses utilizando métodos fraudulentos[Matsumoto et al., 2002]. Para isso, foram utilizados moldes de impressões digitais feitas a base de silicone, e também gelatina caseira. O sistema de verificação validou os moldes, como se fossem pessoas. Com essas impressões digitais falsificadas os cientistas conseguiram burlar a segurança do sistema em 80% das tentativas realizadas. Isso demonstra a fragilidade que a impressão digital pode transmitir ao sistema biométrico.

2.2.2 Íris

A íris do olho pode prover uma solução para discriminação biométrica muito melhor do que as oferecidas por impressão digital ou reconhecimento de face[Negin et al., 2000]. A formação da textura da íris no olho humano depende de condições iniciais da mesoderme

embrionária a partir da qual ela se desenvolve. Por essa razão, constitui uma característica única a cada indivíduo, além de ser permanente, ou seja, invariante durante toda a vida de uma pessoa. Além disso, por ser isolada do meio externo, não sofre modificações por meio de cirurgias. Assim, a íris constitui uma característica altamente estável, tornando a técnica de reconhecimento pela íris uma das mais promissoras na área de biometria[Daugman, 1993].

Os sistemas biométricos baseado no reconhecimento da íris humana são considerados os mais seguros, devido as suas propriedades. Essa técnica de reconhecimento humano é considerada simples, porém muito eficiente. Os sistemas desenvolvidos a partir dessa técnica possibilitam uma identificação com um alto grau de precisão, mesmo com um número muito limitado de recursos computacionais[Daugman, 1993]. Por essas razões, são inúmeras as pesquisas que utilizam a íris como característica biométrica[Krichen et al., 2004; Ma et al., 2003]. A Figura 2.2 ilustra a imagem de uma íris do olho humano.

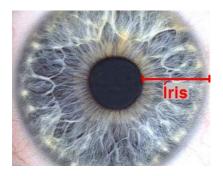


Figura 2.2: Imagem ilustrativa do olho humano com a íris em destaque. As informações mais importantes sobre a íris estão relacionadas a sua textura, e não a sua cor.

Os sistemas que realizam esse tipo de reconhecimento necessitam de um mecanismo de aquisição de imagens. Normalmente, essa aquisição é realizada utilizando-se um scanner, que necessita estar próximo ao olho humano. Além disso, o equipamento de aquisição deve possuir um alto grau de precisão, para obter imagens de boa qualidade e que forneçam as informações sobre textura necessárias para a identificação. Alguns dos principais problemas com o reconhecimento da íris estão no uso do scanner. Isso ocorre pelo fato de muitas pessoas se sentirem desconfortáveis em utilizá-lo, e também pelas restrições técnicas que requerem grande proximidade do olho.

Após a aquisição, é aplicado um método de reconhecimento da íris, de maneira a extrair as principais informações que a caracterizam. Dessa maneira, são gerados identificadores, que são utilizados para distinguir os indivíduos. Para que esses métodos funcionem de maneira adequada, é preciso obter imagem com o menor número de ruí-

dos possível, já que esses podem influenciar na identificação. Muitos métodos podem ser utilizados para realizar a extração, contudo, o que oferece uma melhor generalização e, conseqüentemente, uma melhor identificação é a transformada wavelet de Gabor[Zhang et al., 2003a; Manjunath and Ma, 1996; Chang and Kuo, 1993].

2.2.3 Retina

As características da retina do olho humano, como os vasos sangüíneos, ilustrados na Figura 2.3, constituem umas das características biométricas com o maior grau de estabilidade e unicidade, sendo única em cada pessoa[Hill, 1978]. Isso se deve, possivelmente, ao fato da origem embrionária da retina. As informações da retina não apresentam variabilidade durante a vida de uma pessoa. Contudo, há a possibilidade de haver algum tipo de doença que altere significativamente essas informações, embora, isso só ocorra em situações adversas. Por essa razão, a retina pode possibilitar o desenvolvimento de sistemas biométricos altamente seguros[Marino et al., 2006].

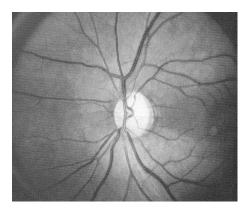


Figura 2.3: Imagem ilustrativa do fundo de um olho humano, onde podem ser observados os vasos sangüíneos da retina. A quantidade e disposição dos vasos são suas informações mais importantes.

A imagem da retina pode ser adquirida por meio da projeção de um laser de baixa intensidade, e a conseqüente captação da imagem por uma câmera. Para o sucesso desse procedimento, é necessário que o olho esteja próximo a câmera, para estabelecer um foco adequado sobre a retina. Em seguida, é aplicado nessa imagem um processo de extração de características, cujo objetivo é o de identificar os vasos da retina e suas bifurcações. A partir dessas informações, pode ser realizada a identificação de uma pessoa. De maneira geral, esse método não sofre influência de alterações do ambiente, como a presença de

óculos de grau, lentes de contato, transplantes, entre outros. Por essa razão, é considerada uma técnica altamente eficiente.

Contudo, esse sistema encontra certas restrições de aceitação, por ser considerado intrusivo. Além disso, há os fatores econômicos, pois o sistema é ainda relativamente caro. Por fim, foi descoberto que o requisito de permanência dessa característica não é totalmente válido, já que, como citado anteriormente, uma pessoa pode adquirir alguma doença que modifique as características da retina, comprometendo o seu reconhecimento. Por outro lado, o fato de uma doença comprometer o reconhecimento (modificação da estrutura dos vasos sanguíneos da retina), pode revelar uma segunda funcionalidade para tais tipos de sistema, que é o auxílio do diagnóstico precoce de algumas doenças relacionadas ao olho humano.

2.2.4 Geometria da mão, dedos e palma da mão

As técnicas de reconhecimento a partir da geometria da mão, dedos e palma da mão são implementadas com base nas informações geométricas dessas características. As aplicações dessas técnicas para sistemas biométricos vêm sendo amplamente exploradas [Kumar et al., 2003; Zhang et al., 2003b]. Para isso, leva-se em consideração o fato de que, em um domínio não muito grande, essas características são únicas para cada indivíduo. Além disso, essas informações costumam ser invariantes ou apresentar pouca variação, após a pessoa ter atingido a idade adulta.

O desenvolvimento de um sistema baseado nessas técnicas utiliza um módulo de aquisição de imagens, que captura a imagem da mão, por exemplo. Em seguida, são analisadas suas informações geométricas, extraindo os principais padrões, que permitem identificar uma pessoa. Esses padrões são geralmente o tamanho dos dedos, a largura e a área. A partir desses dados, é possível desenvolver um sistema biométrico simples, de rápida identificação, reduzido número de rejeições de pessoas cadastradas e baixo custo. Estas são algumas vantagens sobre outras técnicas biométricas, além de possibilitar o armazenamento de dados em um pequeno espaço.

A principal desvantagem dessa técnica está na sua baixa discriminação, ou seja, não possui um bom nível de precisão comparado a outros sistemas. Além disso, há um problema físico, pois esse sistema geralmente ocupa um espaço maior que os demais sistemas de biometria. E por fim, há uma alta suscetibilidade a falhas, sendo simples a adoção de meios fraudulentos para burlar o sistema, caracterizando-o como um sistema de baixa proteção.

2.2.5 Voz

As características vocais do ser humano são totalmente determinadas pela ressonância da região vocal, em função de seu comprimento, do formato da boca e das cavidades nasais[Campbell Jr., 1997]. Devido ao fatores que propiciam a emissão da voz humana, essa característica se torna única para cada indivíduo. Isso possibilita a sua exploração em sistemas biométricos[Wan and Campbell, 2000; Frischholz and Dieckmann, 2000b].

Por meio de um sistema de aquisição de som, como um microfone, é captada uma amostra da voz de um indivíduo. Em seguida, é realizada uma análise dessa amostra, aplicando-se técnicas de processamento de sinais. Normalmente, são analisados os padrões harmônicos do som captado, utilizando-se principalmente análise de Fourier, que analisa a forma de onda do som. Isso permite encontrar o espectro de freqüência que amostra as características da voz. Dessa maneira, diversos sons emitidos por vozes podem ser caracterizados de maneira distinta, possibilitando a identificação de uma pessoa. O procedimento para a identificação de uma pessoa por reconhecimento de voz pode seguir duas abordagens diferentes: (1) dependente do texto; (2) independente do texto. Na primeira, o trecho de fala captado para a identificação deve conter uma frase pré-determinada. Já no segundo caso, o sistema independe da frase pronunciada no momento da aquisição. Essa última abordagem é mais genérica, contudo, mais complexa de ser desenvolvida.

Após um indivíduo ter formado seu padrão de voz em função de fatores físicos e comportamentais, é raro gerar imitações desse padrão, sem que um sistema computacional identifique. Isso torna os sistemas que utilizam esse padrão mais seguros. Contudo, alguns problemas são inerentes a presença de variabilidade da voz, prejudicando o desempenho dos sistemas. Primeiramente, a voz pode ser afetada por fatores ambientais, gerando ruídos e impossibilitando a identificação de um usuário válido, por exemplo. Outros aspectos, e ainda mais críticos, são as variações e alterações da voz de um usuário por razões físicas, como gripes, resfriados e alterações emocionais, como estresse, podem impossibilitar o reconhecimento do usuário. Além disso, o sistema pode ser burlado de maneira simples por meio da gravação da voz de um usuário.

2.2.6 Assinatura Manuscrita

Cada pessoa tem um estilo único de escrever. Assim, a assinatura pode ser considerada como uma espécie de "impressão digital" escrita. Essa característica é única a cada indivíduo, sendo que a sua reprodução eficaz é altamente difícil. Por essa razão, a assinatura

permite o reconhecimento de uma pessoa de maneira única e precisa. Esse conceito de unicidade é utilizado há décadas para identificar pessoas, e validar documentos[Hong, 1998]. Por isso, o uso dessa característica em sistemas biométricos pode ser explorado de maneira ampla, já que as pessoas têm uma percepção intuitiva do seu funcionamento, o que acaba gerando uma grande aceitabilidade do sistema[Fierrez-Aguilar et al., 2005; Munich and Perona, 2003].

A verificação de uma assinatura manuscrita pode ser realizada de maneira estática ou dinâmica. Na primeira, são analisadas as formas geométricas da assinatura do usuário, verificando se há alguma tendência para se escrever em letras arredondadas ou inclinadas, por exemplo. Na segunda, além da análise das formas geométricas, é analisada a maneira com que a assinatura é escrita, avaliando a velocidade, pressão da caneta, identificação dos movimentos da caneta e os pontos em que a caneta é suspensa.

Sistemas que utilizam assinatura dinâmica obtém desempenhos de análise superiores aos que utilizam a estática. Contudo, a implementação da primeira é mais complexa, devido ao número de detalhes observados. A violação do sistema por métodos fraudulentos, como a falsificação de assinaturas, se torna mais difícil em sistemas dinâmicos. Contudo, a falsificação em ambos é considerada difícil. Isso, aliado a intuitividade do sistema, facilita a disseminação de sua usabilidade. Porém, essa técnica apresenta uma desvantagem, que é a grande variação dentro de uma mesma classe, ou seja, uma mesma pessoa acaba assinando de maneiras diferentes. Isso compromete a precisão do sistema.

2.2.7 Dinâmica de Digitação

A técnica de reconhecimento por meio da dinâmica de digitação de um indivíduo, assim como a assinatura manuscrita, analisa o estilo único de uma pessoa digitar. Considerando a variabilidade na maneira das pessoas digitarem, é improvável obter uma imitação perfeita da dinâmica de digitação. Considerando esse fato, o estilo com que uma pessoa digita, que caracteriza sua dinâmica de digitação, pode ser usado como um fator biométrico com um bom grau de segurança e unicidade[Monrose et al., 2002; Robinson et al., 1998].

Um sistema que reconhece a dinâmica de digitação analisa o ritmo da digitação, observando a velocidade, espaço de tempo entre o acionamento de cada tecla, intensidade da pressão, tempo em que se mantém pressionada cada tecla e o tempo de liberação dessas. Ainda, certos dígrafos, ou digitação de duas letras adjacentes, freqüentemente apresentam padrões de tempo únicos, que podem ser usados para auxiliar na caracterização de um

indivíduo. Para que um sistema identifique de fato uma pessoa, é necessário que haja um treinamento prévio. Nessa fase, a pessoa deverá digitar inúmeros textos, para que o sistema associe a ela um modelo de digitação. Esse modelo será utilizado para comparar os usuários, fazendo a sua autenticação. Esse método de reconhecimento é pouco intrusivo e, por essa razão, bem aceito pelas pessoas. Contudo, assim como na técnica de assinatura manuscrita, a variação em uma mesma classe é grande, ou seja, um mesmo usuário digita de várias maneiras diferente. Isso dificulta obter uma boa precisão do sistema.

2.3 Aplicações de Sistemas Biométricos

A utilização de sistemas biométricos não é novidade, pois o ser humano explora muitas dessas técnicas há anos. É possível encontrar relatos do uso de digitais, por exemplo, desde o começo do século XIX, para garantir transações comerciais e, também, seu posterior uso para descobrir suspeitos de crimes[Hong, 1998]. Contudo, foi no século XX que os sistemas biométricos ganharam força, e puderam ser incluídos em sistemas computacionais que, além de facilitar e melhorar antigos sistemas de biometria já conhecidos, criaram espaço para o uso de novas tecnologias, como o reconhecimento da íris. O período entre a descoberta dessas novas técnicas, o aperfeiçoamento das antigas, e o seu uso em sistemas de reconhecimento foi curto.

O principal uso de sistemas de reconhecimento humano está voltado para a área de segurança. Desde a segurança física, em que o objetivo é restringir o acesso a pessoas não autorizadas, até a segurança de informações, onde o objetivo é não permitir que pessoas acessem indevidamente conteúdos que não são pertinentes a ela. Contudo, a utilização de sistemas biométricos é ilimitada, podendo estar presentes em várias áreas de estudos ou do cotidiano dos seres humanos. As mais importantes são:

• Acesso físico — Há muito tempo são utilizados sistemas para impedir que pessoas não autorizadas entrem em um determinado local. Porém, as ferramentas utilizadas para restringir esse acesso eram implementadas utilizando sistemas pouco eficientes, como chave ou tíquete, que podem ser facilmente falsificados. Por essa razão, tornouse necessário criar métodos mais eficazes e seguros de autenticação, como sistemas de biometria, que são mais seguros e precisos que os outros métodos. Alguns sistemas biométricos de identificação podem ser encontrados em uso, como no caso da Disney World nos Estados Unidos, que utiliza a impressão digital para autorizar a entrada em seus parques temáticos, ou mesmo no caso de alguns aeroportos como o

Charlote/Douglas nos Estados Unidos e Flughafen na Alemanha, que utilizam sistemas de reconhecimento de íris para o controle dos passageiros[Liu and Silverman, 2001].

- Acesso virtual Para se realizar o acesso a uma determinada informação importante na *Internet* é necessário, na maior parte das vezes, ser um usuário cadastrado do sistema e utilizar uma senha. Porém, as senhas têm um nível de segurança baixo, e certos sistemas necessitam armazenar as informações de maneira mais segura. Para solucionar esse problema, estão sendo propostos sistemas biométricos para autenticar os usuários, mantendo as informações mais seguras e livre do acesso indevido de pessoas não autorizadas[Everitt and McOwan, 2003].
- Comércio eletrônico O comércio eletrônico ganhou grande espaço nos últimos anos com a popularização da *Internet*, porém, assim como o comércio, aumentou o número de fraudes eletrônicas, por meio de *hacker* utilizando-se de falhas nos sistemas de segurança, e proporcionando enormes prejuízos econômicos para várias empresas de comércio e bancos. Por essa razão, os bancos e empresas de vendas eletrônicas têm demonstrado grande interesse nos sistemas biométricos, que podem garantir as transações com um maior nível de segurança, evitando que *hackers* prejudiquem as transações e, conseqüentemente, espantem seus clientes[Janbandhu and Siyal, 2001].
- Monitoramento A função das aplicações de monitoramento é vigiar locais e áreas com grande concentração de pessoas, a procura de suspeitos e pessoas que possam causar algum tipo de problema ou inconveniência. O sistema faz a análise das pessoas naquele local por meio de vídeos ou fotos, utilizando sistemas de biometria baseadas em técnicas de reconhecimento de face. Esse monitoramento permite o aumento da segurança em áreas públicas. Uma possível aplicação para esse sistema seria em estádios de futebol, onde os torcedores seriam monitorados com o objetivo de encontrar possíveis baderneiros conhecidos.
- Computação Ubíqua O futuro dos sistemas computacionais é ter uma interface simples, não intrusiva, que reconheça o usuário automaticamente por meio da imagem de sua face[Pentland and Choudhury, 2000]. Isso quer dizer que os sistemas tendem a ser simples, transparentes e amigáveis para o usuário. Nesse contexto, os sistemas biométricos possibilitam o reconhecimento do usuário e, conseqüentemente, a aquisição e o acionamento automático de várias de suas preferências. Uma maneira

de desenvolver tal sistema é por meio da utilização da técnica de reconhecimento de faces.

2.4 Considerações Finais

Cada uma das técnicas de biometria possui suas vantagens e desvantagens. Isso quer dizer que não há necessariamente uma técnica que seja melhor que as outras para todos os casos de aplicação. Nesse sentido, cada uma das técnicas vistas são admissíveis, e sua utilização depende do domínio da aplicação. A Tabela 2.1 apresenta uma análise de cada uma das sete técnicas relatadas nessa seção, além da técnica de reconhecimento de face, em relação aos sete requisitos necessários para a escolha de uma técnica biométrica. O valor de cada campo da tabela é o nível com o qual uma técnica atende o requisito: baixo, médio e alto.

Tabela 2.1: Tabela de comparação entre as técnicas de biometria discutidas[Hong, 1998].

Técnica Biométrica	Universalidade	Singularidade	Permanência	Mensurabilidade	Desembenho	Aceitabilidade	Proteção
Impressão Digital	Médio	Alto	Alto	Médio	Alto	Médio	Médio
Face	Alto	Médio	Médio	Alto	Médio	Alto	Baixo
Íris	Alto	Alto	Alto	Médio	Alto	Baixo	Alto
Retina	Alto	Alto	Médio	Baixo	Alto	Baixo	Alto
Geometria da mão	Médio	Baixo	Médio	Alto	Médio	Médio	Baixo
Voz	Médio	Médio	Baixo	Médio	Baixo	Alto	Baixo
Assinatura Manuscrita	Baixo	Médio	Baixo	Alto	Baixo	Alto	Baixo
Dinâmica de Digitação	Baixo	Médio	Baixo	Alto	Baixo	Alto	Baixo

CAPÍTULO

Rastreamento e Reconhecimento de Faces

Uma das características biométricas mais exploradas nesses últimos anos foi a face. A partir da face humana, é possível extrair características como as posições dos olhos, nariz e boca, que permitem localizá-la e, além disso, possibilitam identificar uma pessoa. Com isso, é possível desenvolver sistemas biométricos para uso em sistemas com interfaces amigáveis, que se encaixam nos modelos de HCI (Human Computer Interaction) e que possibilitam uma fácil interação do usuário com o sistema computacional. Também podem ser utilizados em sistemas de monitoramento de segurança e para aplicações em sistemas robóticos, ou no desenvolvimento de sistemas de análise de expressões do rosto, utilizadas em várias áreas da ciência, como a psicologia. A face como uma característica biométrica é largamente explorada por ser intuitiva aos seres humanos, já que todos podem ser diferenciados pelas suas aparências faciais.

Os pesquisadores baseiam-se na intuitividade do reconhecimento da face para explorar suas características e, a partir daí, desenvolver as aplicações. Uma dessas aplicações foi proposta por Sekmen et al. [2002], que procura realizar o controle de movimento de um robô a partir dos movimentos da face humana. Isso pode permitir que deficientes físicos andem em cadeiras de rodas motorizadas, sem necessitar controlá-las com as mãos, mas sim com os movimentos do rosto. Existe também a proposta de Kuno et al. [2003] de desenvolver um robô humanóide que reconheça quando uma pessoa entra no ambiente em que se encontra o robô, e o acompanhe por meio de uma câmera, realizando o rastrea-

mento da face dessa pessoa. O uso de reconhecimento de faces auxilia no desenvolvimento da área de visão cibernética, que busca simular a maneira como os seres humanos utilizam seu sistema de visão.

Além das já citadas aplicações, existem muitas outras que podem fazer uso das informações provenientes da face humana. Lanitis et al. [2004] propõem um algoritmo para determinar a idade de uma pessoa. Com isso, torna-se possível, entre outras coisas, adaptar o ambiente computacional automaticamente, baseado na idade do usuário, tornando o sistema mais ubíquo. Construir um modelo de indexação de imagens por idade, que poderia ser utilizado para consulta de álbuns de fotografia. Desenvolver um programa que simula automaticamente o envelhecimento, que poderia ser usado para estimar a forma da face de uma pessoa desaparecida há alguns anos. E também, auxiliar nas pesquisas da área de psicologia, que procura entender o processo de percepção de idade de um ser humano. No contexto desse último exemplo, Tian et al. [2001] e Gutta et al. [2000] propõem métodos para reconhecimento de expressões faciais que podem identificar o estado emocional de uma pessoa, além de sua etnia. A partir daí, é possível realizar estudos sobre comportamento humano, levantamento étnico e sobre comunicação visual.

Um grande motivador das pesquisas na área de reconhecimento de face é o de monitoramento de segurança. Esses sistemas, aliados a biometria da face, permitem que pessoas sejam reconhecidas em um determinado ambiente. Algumas entidades já fazem uso de sistemas de monitoramento a partir do reconhecimento facial, como algumas companhias comerciais, órgãos governamentais dos Estados Unidos, cassinos e aeroportos[Pentland and Choudhury, 2000]. O objetivo desse monitoramento é preservar a segurança desses ambientes, fazendo, por exemplo, a verificação automática de algum criminoso. Além disso, pode ser utilizado o reconhecimento de face para o monitoramento, em locais públicos, auxiliando órgãos e entidades na procura por pessoas desaparecidas.

O desenvolvimento dessas aplicações, a partir de imagens de faces humanas, exige a utilização de métodos computacionais que permitam localizá-la e, posteriormente, reconhecê-la, ou seja, identificar a face. Para isso, foram propostos inúmeros algoritmos nos últimos anos. As técnicas de localização e reconhecimento de faces em imagens podem ser divididas em quatro categorias[Yang et al., 2002]. Contudo, algumas técnicas podem conter aspectos de mais de uma categoria diferente, devido a combinações com outros algoritmos. Essa combinação é explorada com o objetivo de melhorar o desempenho desses algoritmos, porém, essas quatro categorias são suficientes para classificar as técnicas existentes. São elas:

- Métodos baseados no conhecimento Baseiam-se no conhecimento humano sobre quais características constituem a face. São utilizados principalmente para localização.
- 2. Abordagens de características invariantes Têm como objetivo encontrar e utilizar características estruturais, que não sofrem influência de variações de posição, cor, iluminação, entre outras. São utilizados principalmente para localização.
- 3. Métodos de casamento de padrões (template matching) Utilizam vários modelos de padrões, que descrevem uma face. O processo de reconhecimento é realizado pela comparação entre esses modelos e as imagens. Podem ser utilizados tanto para localização quanto reconhecimento.
- 4. **Métodos baseados na aparência** Diferentemente do casamento de padrões os modelos são ensinados a partir de um treinamento. Dessa maneira, o algoritmo aprende a identificar a face. São utilizados principalmente para reconhecimento.

Esses algoritmos exploram diversas características da face humana, que serão exploradas mais detalhadamente na seção 3.1. A partir desses detalhes, são implementados métodos para localizar a face em uma imagem, podendo ser realizado o rastreamento da face quando seqüências de imagens ou vídeos são utilizadas. Esses métodos serão mais bem discutidos na seção 3.2. Posteriormente, é aplicado um método de extração de características, que possibilita o reconhecimento e identificação da face. Esses métodos serão apresentados na seção 3.3. É importante mencionar a diferença entre localizar e reconhecer uma face. A primeira tem como objetivo simplesmente definir o posicionamento da face na imagem, enquanto o reconhecimento identifica as características inerentes à face. Para avaliar os métodos de reconhecimento podem ser utilizadas diversas abordagens. Na seção 3.4, serão discutidas algumas maneiras para realizar essas avaliações.

3.1 Características

A face humana possui inúmeras características que são utilizadas, de maneira intuitiva, para reconhecer as pessoas, como os olhos, sobrancelhas, boca, nariz e cabelo. Para descrever uma pessoa é necessário analisar o posicionamento, forma e tamanho dessas características, além da distância que separa cada uma delas. A partir dessas informações, é possível caracterizar uma pessoa. Um sistema biométrico faz a extração dessas informações para relacionar um determinado indivíduo à imagem que o identifica. Dessa

maneira, é que são desenvolvidos os sistemas de reconhecimento de face. Na Figura 5.2 é possível observar alguns pontos de interesse de uma face. As características mais exploradas são a boca, o nariz, os olhos e as sobrancelhas[Yang et al., 2002].



Figura 3.1: Alguns pontos de interesse para o reconhecimento de face. Algunas características que podem ser analisadas são, por exemplo, as distâncias entre os pontos marcados na imagem (distância entre olhos, boca e nariz).

Essas informações têm sido utilizadas com sucesso para o reconhecimento de faces frontais. Contudo, em algumas situação as imagens não apresentam todas as informações necessárias, como em imagens de perfil. Esse e outros problemas têm sido fatores de desafio e motivação para as pesquisas nessa área:

- Pose As imagens apresentam faces em diversas poses, com diversos graus de variação, como frontal e de perfil. Isso pode ocluir totalmente ou parcialmente algumas características como o olho e a sobrancelha.
- Outras estruturas A presença ou ausência de outras estruturas no rosto, como barba, bigode, óculos, pode levar a um erro de análise.
- Expressão facial A expressão facial (riso, choro, etc.) altera as características da face.

- Oclusão Algumas imagens apresentam a oclusão parcial ou total de características da face por objetos da cena.
- Orientação da imagem Imagens podem apresentar distinções quanto ao posicionamento da face. Podendo estar rotacionada em diversas direções diferentes.
- Condições da imagem As condições de captura da imagem podem influenciar no reconhecimento. Isso se deve ao fato da possibilidade de haver variações de luminosidade e, também, presença de ruídos devido ao equipamento de aquisição.

Esses fatores influenciam na tomada de decisão sobre o método de reconhecimento de face, além da escolha das informações que serão analisadas. Por propiciarem uma variação muito grande de imagens em ambientes reais, há uma grande quantidade de métodos para reconhecimento de face, contudo, são esses mesmos fatores que possibilitam novas pesquisas nessa área. Como exemplo, podemos citar as pesquisas que procuram combinar o reconhecimento de face padrão, com o de outras características da rosto humano menos utilizadas, como o uso da orelha como uma informação pertinente à identificação de uma pessoa[Chang et al., 2003]. Essa combinação de informações proporciona uma maior flexibilidade aos algoritmos.

3.2 Localização e Rastreamento

A localização e rastreamento de um objeto em uma determinada cena são motivadores de diversas pesquisas, algumas das quais têm como objeto de análise a face humana. O desenvolvimento dessas técnicas é importante para diversas aplicações como o reconhecimento de face, videoconferência, visão cibernética e realidade virtual. Por essa razão, são inúmeras as propostas de algoritmos que tentam viabilizar, de maneira eficiente, a localização e rastreamento de uma face. Contudo, existem inúmeros problemas que dificultam o desenvolvimento de um algoritmo robusto para esse fim. Esses problemas estão relacionados com a grande variabilidade de aspectos como iluminação, objetos que ocluem a face e, também, de ambientes com um fundo muito complexo para análise. Assim, solucionar esses problemas e desenvolver um bom método torna-se um desafio para muitos pesquisadores.

A primeira etapa do processo é a localização da face. Inicialmente, realiza-se uma análise em uma imagem de entrada, e procura-se encontrar um objeto que possa ser uma face. A seguir, é necessário validar o objeto, como sendo uma face ou não. Posteriormente,

caso seja uma face, é iniciada a segunda etapa do processo, que é o rastreamento. Nessa etapa, o algoritmo precisa acompanhar a face durante o período de duração da seqüência de imagens, sem perdê-la.

Muitas abordagens são pesquisadas para solucionar os problemas que envolvem a localização e rastreamento de faces. Algumas dessas, propõem a utilização da cor e/ou textura da face humana. A cor da pele humana pode ser agrupada em uma pequena porção do espaço de cor, e em determinadas condições de iluminação, a cor se torna invariante ao tamanho, escala e orientação do objeto[Wang and Yuan, 2002]. Isso torna essa característica apta a ser utilizada para localizar faces em uma determinada cena. Alguns sistemas de cores são comumente usados para essa prática, como RGB, HSV e YCrCb. Contudo, o sistema HSV é o que permite uma compreensão mais fácil e natural do modelo de cores, e que, por essa razão, é amplamente utilizado para o reconhecimento de faces[Sobottka and Pitas, 1996]. Outro método de localização de faces é o casamento de padrões (template matching)[Yang et al., 2002]. Esses padrões (templates) são construídos a partir de informações observadas da face humana, como a disposição entre os olhos e a boca, e até mesmo o formato da face que se assemelha a uma elipse[Sobottka and Pitas, 1996].

A partir da localização da face é possível realizar o rastreamento em uma seqüência de imagens. Alguns métodos de rastreamento consideram que a face está localizada, e para fins de implementação do algoritmo, realizam a localização inicial da face manualmente[Verma et al., 2003]. Uma possível abordagem para realizar o rastreamento baseia-se na cor da pele. A maioria desses métodos localiza continuamente a face durante a seqüência de imagens, fazendo uma estimativa a partir da localização no quadro anterior. Isso ocasiona a perda da face em movimentos rápidos[Wang and Yuan, 2002]. Wang and Yuan[2002], propõem uma nova abordagem, onde a análise da cor da pele no momento do rastreamento é realizada de maneira global e não local, como citado, o que resolve o problema dos movimentos rápidos. Para isso, são utilizados agentes igualmente distribuídos por toda a área da imagem inicial, que fazem a análise e verificam se há uma face naquela região ou não. Esses agentes se expandem e possibilitam a segmentação da região onde há a face. Com o objetivo de minimizar o tempo computacional do processo, apenas os agentes ativos, aqueles que localizaram uma face, são utilizados nos quadros seguintes.

Outras abordagens fazem uso de treinamento com o objetivo de ensinar ao sistema quais são as características importantes da face. Uma dessas, propõe o uso de redes neurais RBF(Radial Bases Function) para realizar o rastreamento[Yang and Paindavoine, 2003].

Nesse método, o primeiro passo consiste do treinamento da RBF, a partir de algumas imagens exemplo da face humana. Logo após, o sistema está apto a localizar uma face em uma determinada imagem. Esse processo é repetido para todas as seqüências de imagens. Feris et al. [2001], propõem o uso da rede wavelet de Gabor (Gabor Wavelet Networks - GWN) que consiste na combinação das vantagens da RBF e da wavelet de Gabor. A partir da GWN é possível extrair a chamada superwavelet que define matrizes de dilatação e de rotação e um vetor de translação. É por meio da deformação da superwavelet que se ativa o rastreamento. Uma outra abordagem propõe a combinação de análise de movimento e cor, juntamente com uma versão assimétrica do Adaboost, para localizar uma face e, com o auxílio de heurística, realizar o rastreamento [Ruiz-del-Solar et al., 2003]. Usando a técnica Adaboost e alguns exemplos de face e não faces, é realizado o treinamento do classificador. O rastreamento inicia-se com a identificação, na seqüência de imagens, de um objeto se movendo. Para isso é utilizada a técnica de análise de movimento. Posteriormente, é verificado se esse objeto é uma face ou não. A análise de cor serve para otimizar o processo, diminuindo a área de procura.

Existem ainda, pesquisas que se baseiam em métodos probabilísticos para realizar o rastreamento. Uma dessas possíveis abordagens baseia-se na estimativa de densidade posterior usando o método sequencial de Monte Carlo[Li and Chellappa, 2000; Li and Chellappa, 2002]. Utilizando uma classe do método de Monte Carlo, conhecida como sequential importante sampling (SIS), é possível estimar quando, e para onde, um objeto se move. O algoritmo SIS é utilizado para calcular a densidade de uma amostra inicial, que é a região da face, durante toda a seqüência de imagens, a cada quadro. Assim, quando o SIS verifica uma translação do objeto, o rastreamento é iniciado. É importante observar que os autores não se preocupam com a localização inicial do objeto. Outra abordagem propõe o uso de um método probabilístico baseado em wavelet, para localizar a face, e o uso combinado das técnicas de Factored sampling e CONDENSATION (CONditional dENsity propagaTION), como uma espécie de filtro utilizado no rastreamento[Verma et al., 2003]. Inicialmente, é realizado um treinamento probabilístico com faces, em diversas posições, e não faces. Isso possibilita desenvolver o algoritmo de localização que, posteriormente, é aplicado a cada quadro da seqüência de imagens, e, com o auxilio do filtro, que faz uma predição e uma atualização das faces encontradas, constitui-se o processo de rastreamento.

A maior parte das pesquisas envolvendo localização e rastreamento da face humana, utiliza câmeras simples para adquirir as seqüências de imagens. Contudo, há muitos trabalhos que propõem o uso de câmeras estéreo. Uma das vantagens desse método é que a disparidade a torna menos sensível à variação de intensidade de iluminação. Porém,

o cálculo dessa disparidade aumenta sua sensibilidade à textura da imagem[Sung, 2002]. Sung[2002] propõe a integração entre filmes estéreo, watershed e técnicas de gradiente para realizar o rastreamento da face. Primeiramente, calcula-se a disparidade, ou seja, as duas imagens (direita e esquerda) do mesmo quadro são combinadas. Usando um limiar, selecionado do histograma da imagem de disparidade a partir de uma heurística, separa-se a pessoa do fundo da imagem. Aplicando watershed, separa-se a cabeça do restante do corpo humano. Por fim, com o auxilio de técnicas de gradiente, é sobreposta a cada face um contorno na forma de uma elipse. O rastreamento é feito a partir da atualização desse contorno.

3.3 Reconhecimento e Identificação

Classificar um determinado objeto em uma imagem como sendo uma face, consiste no processo de reconhecimento. Já a identificação determina a quem pertence essa face. Para realizar o primeiro processo é necessário extrair as características da face humana, para que se possa diferenciá-la dos demais objetos dentro de um ambiente qualquer. Contudo, essas mesmas informações extraídas, podem ser utilizadas para o processo de identificação de uma pessoa. Por essa razão, ambos os processos serão tratados de maneira análoga.

Existem vários métodos propostos para o reconhecimento de faces humanas. Os que se incluem na classe de casamento de padrões (template matching) requerem a construção de modelos que possuem uma parametrização das características que definem a face. No momento da análise da imagem, é iniciado o processo de casamento de padrões, entre os modelos e as imagens. Nesse momento, é realizado o reconhecimento. Os modelos podem ser construídos de diversas maneiras, como, por exemplo, a partir das definições de bordas ou de blocos, que definem a face e suas características. Para isso, podem ser utilizadas técnicas de reconhecimento de bordas, como Sobel ou operadores Laplaciano, ou ainda obter essas informações por meio da aplicação da PCA(Principal Component Analysis). Esses métodos apresentam um bom resultado prático, contudo, são pouco flexíveis a variações de tamanho e escala. Para solucionar esse problema, alguns métodos exploram a utilização de modelos deformáveis, os quais podem se ajustar de maneira mais adequada às variações citadas. Uma maneira de realizar esse ajuste é associando o uso de snakes[Yang et al., 2002].

A classe de métodos de reconhecimento de faces mais explorada é a de métodos baseados em aparência. Neles, diferentemente da classe de casamento de padrões, o

algoritmo de reconhecimento aprende o modelo. Normalmente, são utilizadas técnicas de análise estatística e aprendizado de máquina, para encontrar as características relevantes da face. Essas características ensinadas podem ser em forma de modelos de distribuição ou funções discriminantes. Um dos métodos mais explorados em pesquisas dessa área é o de eigenfaces [Gao and Leung, 2002]. O seu objetivo é gerar uma pequena base de vetores, que possibilitem a representação das imagens que estão sendo analisadas. Essas bases são geradas a partir do treinamento com uma amostra de imagens de faces, usando a transformada de Karhunen-Loève, também conhecida como PCA ou transformada de Hotelling. A análise da imagem nesse algoritmo possibilita indexar as imagens, permitindo o reconhecimento de faces [Yang et al., 2002].

Outros métodos dessa classe são os baseados em distribuição. Um deles, utiliza um algoritmo modificado do k-means para construir clusters das faces. Em seguida, aplicando medidas de distância Mahalanobis e Euclidiana, computa a distância entre as imagens de treinamento e os clusters. Essas informações são utilizadas para o treinamento do classificador MLP (multilayer perceptron). Outra abordagem faz uso da PCA para gerar os clusters das faces. Posteriormente, a multivariada Gaussiana e uma composição de Gaussianas são usadas no treinamento estatístico das características da face. A densidade de probabilidade pode ser utilizada para fazer o reconhecimento. Outros métodos propiciam o uso de outras técnicas no lugar da PCA, como a FA (factor analysis) e a FLD (Fisher's Linear Discriminant). A FA tem como vantagem em relação a PCA, a definição de um modelo apropriado de densidade para os dados analisados. Enquanto isso, a FLD possui uma melhor projeção que a PCA. Um dos motivos é que a FLD encontra um maior número de direções de projeções discriminantes[Yang et al., 2002].

Vários outros métodos da classe baseada em aparência são relatados em Yang et al. [2002]. Alguns desses, baseados no treinamento usando faces exemplos, utilizam redes neurais, como Kohonen e PDBNN (probabilistic decision-based neural network). Contudo, uma das redes neurais mais explorada é a MLP. Existem também métodos que utilizam o classificador RBF (radial basis function). Há ainda, um novo paradigma de treinamento, conhecido como SVMs (Support Vector Machines). As SVMs têm como princípio básico a redução de erros, chamado de minimização de risco estrutural. A arquitetura SNoW (Sparse Network of Winnows) também é utilizada em métodos de reconhecimento. A SNoW permite explorar o treinamento baseado nas diversas característica da face, dadas como exemplo. Uma abordagem distinta é a do classificador Naive Bayes que, diferentemente de outros métodos, faz uma análise global e local da face. A junção com a análise local permite que algumas características inerentes à face humana sejam mais bem ex-

ploradas, e identificadas. Ainda, existe o modelo HMM (Hidden Markov Model), que considera que o padrão pode ser caracterizado por um processo randômico paramétrico, e os parâmetros desse processo podem ser estimados de uma maneira precisa e bem definida. Para desenvolver um HMM, é necessário decidir a forma do modelo, para estipular o número de estados intermediários. Posteriormente, é realizado o treinamento com exemplos que representam uma seqüência de observações. Isso permite ajustar os parâmetros dos estados, com o objetivo de maximizar a probabilidade de observação dos dados de treinamento. Com essa fase completa, dada uma imagem de face, o algoritmo desenvolvido fará o reconhecimento da face. Para a aplicação específica em faces humanas, uma possível abordagem é utilizar os estados como identificadores de cada característica da face, como olhos, nariz e boca, sendo que essas são observadas de maneira consecutiva.

Essas abordagens de reconhecimento de face mencionadas são alvo de grande parte das pesquisas realizadas na área. Contudo, existem muitos problemas que não são tratados nesses algoritmos inicialmente propostos, sendo que as maiores dificuldades estão relacionadas à variação de intensidade, cor e posicionamento da face. Estudos recentes mostram diversas variações dessas técnicas, com os objetivos de minimizar os problemas e, também, de otimizar as soluções. Gao and Leung[2002], propõem o uso de Line Edge Map (LEM) para o reconhecimento de faces. LEM é uma evolução da Edge Map (EM), que cria um mapa de arestas, para formar modelos de representação da face. Seguindo essa abordagem, a LEM gera modelos a partir de linhas que caracterizam a face. Para calcular a distância entre as LEMs é proposta a LineSegment Hausdorff Distance (LHD). Assim, é possível obter os parâmetros necessários para o reconhecimento. Essa nova abordagem pode ser considerada uma combinação entre casamento de padrões e casamento de características geométricas (geometrical feature matching). Essa técnica se mostrou flexível a variações de iluminação, posicionamento e expressão facial.

Outro trabalho apresenta o reconhecimento de face usando informações globais e locais de texturas, a partir da localização por Bayesian Shape [Yan et al., 2004]. Esse método utiliza wavelet de Gabor para extrair as informações sobre textura da imagem. Tanto as informações locais como globais são extraídas, com o objetivo de propiciar uma análise de alta e baixa freqüência. Em seguida, os parâmetros extraídos são utilizados para localizar a face. Para essa etapa é proposta uma combinação entre o algoritmo Rank-Boost, e um modelo adaptativo de distribuição de probabilidade local. Isso proporciona a parametrização do modelo a ser utilizado para reconhecer a face. Como esse é um modelo estatístico, é necessário que haja um treinamento com exemplos de face, para que o modelo possa ser gerado. Por fim, são utilizados métodos de medidas para mensurar

a similaridade entre o modelo, e o valor obtido na análise da imagem de entrada. Para isso, foram utilizadas três medidas de distância distintas: L1, Euclidiana e co-seno. Esse método se mostrou invariante a luminosidade, escala e posicionamento da face.

Liu e Wechsler [2003], propõem um novo método para reconhecimento de face chamado características independentes de Gabor (independent Gabor features - IGFs). O IGF é constituído de duas partes principais. A primeira realiza a extração das características independentes de Gabor. Na segunda é desenvolvido o método de classificação IGF baseado nas características, usando o modelo PRM (Probabilistic Reasoning Model). O processo completo descrito pelos autores segue, primeiramente, com a extração das informações das imagens aplicando a transformada wavelet de Gabor. Em seguida, essas informações passam por um processo de redução de dimensão. Para isso, é utilizado PCA. Posteriormente, os dados originados da PCA são processados pela ICA (Independent Component Analysis), que tem como objetivo reduzir os dados redundantes, permitindo que as informações representem explicitamente características independentes. Após essas etapas, as informações são processadas pela PRM, que define um classificador linear Bayesiano. Esse PRM é treinado com faces exemplos, para formar o classificador. Para realizar comparações e reconhecer uma face, foram utilizados três medidas de distância: L1, L2 e co-seno. Foi ainda construída uma regra de classificação por média de vizinhança. Com esse processo, os autores desenvolveram um método de reconhecimento invariante a luminosidade, escala e posicionamento da face. Recentemente, foi proposto um ICA melhorado chamado EICA (Enhanced Independent Component Analysis)[Liu, 2004. Essa técnica tem o mesmo objetivo do ICA, contudo, propõe uma maior redução das informações de saída, o que reduziria as informações irrelevantes. O EICA descarta os pequenos valores de autovalores resultantes do processo do ICA. Dessa maneira, o número de dados gerados é menor, como proposto, e ainda, não há perdas substanciais de informações nesse processo. Isso pode otimizar o desempenho de processos que utilizam como entrada os dados gerados pela EICA.

3.4 Avaliação de Algoritmos de Reconhecimento

Existem diversos métodos propostos para localização e reconhecimento de face. Uma grande parte desses métodos passa por uma etapa de treinamento, em que são necessários alguns exemplos de faces, para que o algoritmo desenvolvido possa funcionar corretamente. Além disso, todos esses métodos propostos precisam ser avaliados, para observar todas as suas vantagens e desvantagens, principalmente quanto às variações de iluminação,

escala e posição de face. Juntamente com os métodos propostos seguem metodologias de avaliação de resultados, onde geralmente são criadas bases de imagens próprias, que somente os autores têm acesso. Em conseqüência disso, muitos métodos não podem ser comparados adequadamente, já que os resultados divulgados baseiam-se em testes com bases exclusivas. Por estas razões, é necessário a utilização de bases de imagens comuns, que possam fornecer imagens para treinamento e, também, para a realização de testes, facilitando a comparação entre os resultados obtidos.

Para suprir essa deficiência foram criados, nos últimos anos, algumas bases de imagens. O programa FacE REcognition Technology (FERET) disponibiliza a base de dados FERET. Essa base é constituída de 14.126 imagens de 1.199 pessoas[Phillips et al., 2000], com uma pessoa em cada imagem, e variações de luminosidade, expressão, pose e escala[Liu and Wechsler, 2003]. Por ser uma base com um grande número de imagens, ela é a mais utilizada para realizar comparações entre diversas técnicas de reconhecimento de face. Uma outra base de faces é a da AT&T Cambridge Laboratories, conhecida como base de dados Olivetti, que consiste em 10 diferentes imagens para cada uma das quarenta pessoas. Essas imagens possuem variações de luminosidade, expressão e detalhes faciais, como óculos e bigode. A base de dados de Yale possui 10 imagens frontais de cada uma das 16 pessoas, com variações de expressão facial, luminosidade e faces com e sem óculos. A base de dados UMIST possui 564 imagens de vinte pessoas, com variações de pose, frontal e perfil. Uma outra base é a Purdue AR, que contém 3.276 imagens coloridas frontais com 126 pessoas (70 homens e 56 mulheres). Nessa base, há variações de expressão facial, iluminação e ocorrência de oclusões de face[Yang et al., 2002].

Um aspecto importante a ser considerado, no momento da avaliação de um algoritmo de localização e/ou reconhecimento de face, é a classificação dos erros encontrados durante os testes. É necessário que se avalie o grau de fatalidade de ocorrência de um determinado erro para o sistema desenvolvido. A possibilidade de se obter um falso positivo pode influenciar, de maneira distinta, aplicações diferentes. Por exemplo, um sistema de contagem de pessoas que identifica um objeto qualquer como sendo uma pessoa, pode não trazer tantos problemas quanto um sistema de verificação de identidades, que autentica uma pessoa "A" como sendo uma "B". Essa avaliação é essencial para se realizar um conjunto de testes adequado.

3.5 Considerações Finais

Neste capítulo foi apresentada uma explanação geral sobre os métodos de localização e reconhecimento de faces humanas. Foram discutidas algumas aplicações que utilizam, ou podem utilizar, os algoritmos desenvolvidos nessa área, como sistemas de videoconferência, visão cibernética, segurança, entre outros. A exploração da face humana como uma característica biométrica é facilmente assimilada pelas pessoas, pois é um processo altamente intuitivo, utilizado no dia-a-dia. Além disso, os sistemas de reconhecimento de face são considerados menos intrusivos, pois não necessitam de aparatos intrusivos para adquirir a imagem da face humana. Por essa razão, há uma melhor aceitabilidade de sistemas que utilizam reconhecimento de face, em detrimento de outras técnicas biométricas mencionadas no Capítulo 2.

Na seqüência, foram apresentadas as principais características da face humana empregadas para a localização e reconhecimento, como sobrancelha, olhos e boca. Em seguida, foram abordados os métodos utilizados para realizar a localização e rastreamento, que possibilitam rastrear uma face em uma determinada seqüência de imagens. Uma aplicação para esses métodos é o acompanhamento e monitoramento em tempo real de câmeras de segurança. Unindo esses métodos aos de reconhecimento de face, é possível desenvolver um sistema que identifique pessoas em um determinado ambiente, como em um laboratório, ou até mesmo, possibilitar que um robô se familiarize com as pessoas, identificando-as e armazenando informações sobre cada uma delas, de maneira distinta. É interessante ressaltar a abordagem de alguns trabalhos que têm como objetivo combinar diferentes características da face (que normalmente não são utilizadas em conjunto), para melhorar o desempenho do algoritmo de reconhecimento. Como exemplo, pode-se citar o uso das informações faciais comumente exploradas (olhos, nariz e boca) com informações da orelha[Chang et al., 2003].

Um ponto importante a ser mencionado é a flexibilidade de interação entre os sistemas biométricos. A combinação de técnicas biométricas distintas auxilia no desenvolvimento de sistemas multimodais, que têm como objetivo aumentar o desempenho de reconhecimento do sistema. Nesse contexto, a associação entre a face humana e outras características podem gerar bons resultados, já que o reconhecimento de face se adapta de maneira simples a qualquer ambiente de aplicação. Uma possível associação é entre o reconhecimento de face e de voz[Frischholz and Dieckmann, 2000a], que possibilita ao sistema classificar um indivíduo a partir de um número maior de informações.

Baseado nos resultados e avanços observados na revisão bibliográfica levantada neste capítulo, foram escolhidos alguns métodos para o desenvolvimento deste trabalho. Esses métodos são a transformada wavelet de Gabor, a técnica PCA, e a técnica de aprendizado de máquina SVM. Acredita-se que o uso desses métodos em conjunto, para a área de biometria facial, podem levar a obtenção de bons resultados. O Capítulo 4 apresenta cada uma destas três técnicas.

CAPÍTULO

Técnicas para Reconhecimento de Face

O trabalho descrito nesta dissertação envolveu em um primeiro momento os estudos baseados em biometria, principalmente os assuntos relacionados a localização e ao reconhecimento de faces. Dessa maneira, foi realizado um levantamento bibliográfico sobre a área de pesquisa relacionada. Baseado nas informações coletadas durante a fase de pesquisa bibliográfica verificou-se a necessidade de aprofundar os estudos em algumas técnicas, as quais seriam utilizadas para o desenvolvimento do sistema proposto. Assim, iniciou-se a fase de estudos e entendimento das técnicas de transformada wavelet de Gabor, de PCA (Principal Components Analysis), e de aprendizado de máquina SVM (Support Vector Machine).

A escolha, neste trabalho, de cada uma das técnicas citadas acima se deve ao grande destaque e bons resultados obtidos em aplicações que as utilizam (como pode ser observado no Capítulo 3). A transformada wavelet de Gabor (Seção 4.1) apresenta-se como um ótimo método de extração de características[Yu and Bhanu, 2006; Vukadinovic and Pantic, 2005; Liu and Wechsler, 2003; Arivazhagan and Ganesan, 2003]. A técnica PCA (Seção 4.2) é conceituadamente eficiente para a redução de dados[Lopez-Rubio et al., 2004; Dash et al., 1997]. Já a técnica de aprendizado de máquina SVM (Seção 4.3) vem apresentando ótimos resultados nas aplicações multiclasses[Jiang and Su, 2005; Vatsa et al., 2005; Fierrez-Aguilar et al., 2003; Yuan et al., 2001] (aquelas que apresentam um grande número de classes distintas), como é o caso deste trabalho, devido ao seu grande

poder de generalização e sua robustez. Considerando esse cenário, acredita-se que um sistema de reconhecimento de face, que utiliza essas técnica em conjunto, pode gerar bons resultados.

4.1 Introdução a Transformada Wavelet

A análise de sinais em diversas áreas do conhecimento, como engenharia biomédica, geofísica, processamento de imagens, medicina, robótica, entre outros têm sido de grande importância para a pesquisa e desenvolvimento de trabalhos para essas áreas. Por meio da análise desses sinais é possível conhecer ou descobrir o comportamento de sua fonte, e identificar os parâmetros que a influenciam, com maior ou menor intensidade de atuação. Existem inúmeras técnicas de análise desses sinais, sendo que uma das mais conhecidas e utilizadas é a transformada de Fourier. Contudo, essa técnica possui uma deficiência no que diz respeito à análise de sinais não estacionários, pois não representa a informação temporal do sinal analisado. Dessa maneira, se faz necessário uma outra abordagem para a análise dessa família de sinais, quando tal informação é necessária. Uma proposta bem aceita e que gera bons resultados é a utilização da transformada wavelet.

A transformada wavelet foi sugerida como passo seguinte a uma modificação da transformada de Fourier, a transformada de Fourier por janela, a qual minimizava o problema da análise de sinais não estacionários. Contudo, ainda mantinha a característica de invariabilidade do tamanho da janela durante a análise, melhorando, mas não resolvendo totalmente a perda de informação temporal do sinal. Com a proposta da wavelet obtevese uma forma mais flexível de análise de sinais, onde a informação temporal do sinal analisado é preservada e pode, então, ser utilizada para representar uma característica importante do sinal.

Nessa seção serão explanados os conceitos básicos da transformada wavelet, como a função wavelet mãe. Para isso, será apresentado uma breve descrição sobre a transformada de Fourier, Subseção 4.1.1. Em seguida serão explicados os conceitos da transformada wavelet, Subseção 4.1.2. Também serão tratadas as funções wavelets mães, Subseção 4.1.3, as wavelets bidimensionais, Subseção 4.1.4, alguns exemplos de wavelets, Subseção 4.1.5, e algumas aplicações baseadas em transformada wavelet, Subseção 4.1.6.

4.1.1 Transformada de Fourier

Uma das técnicas mais utilizadas por pesquisadores da área de análise e processamento de sinais é a transformada de Fourier (Fourier Transform - FT), cujo objetivo é transformar um sinal do domínio espacial para o domínio de freqüência, decompondo o sinal original em componentes de (senos e co-senos) de diferentes freqüências[Gonzales and Woods, 1992].

A FT é extremamente eficiente para vários domínios de aplicação. Contudo, existem alguns domínios para o qual essa técnica não se mostra tão eficiente quanto o desejado. Essa transformada possui uma peculiaridade indesejável: no momento da transformação do sinal no domínio espacial para o da freqüência perde-se totalmente a informação sobre a localização temporal do sinal. Isso faz com que a transformada seja eficiente para o tratamento de ondas estacionárias (aquelas que não se alteram no tempo), porém para as não estacionárias ela se mostra ineficiente. Dessa maneira, muitos sinais não estacionários ou transitórios que possuem características importantes (como tendências, mudanças abruptas, inicio e fim de eventos) acabam perdendo tais informações devido à utilização dessa técnica, o que prejudica sua análise.

A equação que define a transformada de Fourier é dada por

$$F(u) = \int f(t)e^{-2i\pi ut} dt, \qquad (4.1)$$

e permite que um sinal seja decomposto em diferentes freqüências, porém sem a informação de localização temporal. Pode-se perceber bem essa perda de informação observando-se o exemplo da Figura 4.1 (a,b), onde dois sinais, o primeiro consistindo de duas freqüências (sen(10t), sen(20t)) sobrepostas durante a duração do sinal, e o outro com as mesmas duas freqüências ocupando cada uma metade do sinal, separadamente. Na Figura 4.1 (c,d) é possível observar o espectro de energia, $|\int (t)|^2$, dos dois sinais respectivamente. Como pode ser observado não há distinção entre os sinais.

Para solucionar essa deficiência, presente na transformada de Fourier no tratamento de sinais não estacionários, se faz necessária a utilização de técnicas capazes de analisar e obter o conteúdo de freqüência localmente no tempo. Para isso, existem duas técnicas que podem ser empregadas para esse fim: transformada de Fourier por janelas (Windowed Fourier Transform - WFT), descrita por Dennis Gabor[Gabor, 1946]; e a transformada wavelet[Daubechies, 1990]. A WFT, conhecida também como transformada de Fourier de curta duração, é uma adaptação da FT, que tem como objetivo resolver

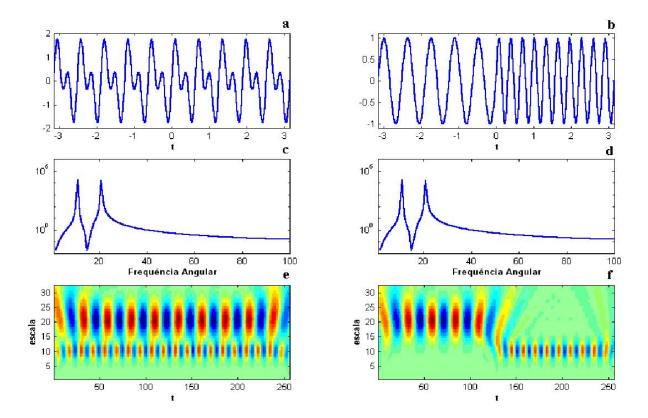


Figura 4.1: Análise de dois sinais por FT e CWT. A Figura (a) consiste de um sinal de duas freqüências (sen(10t) e sen(20t)) sobrepostas, e (b) consiste de um sinal das mesmas freqüências ocupando metade do tempo do sinal cada uma. As Figuras (c) e (d) são os espectros dos sinais de (a) e (b), respectivamente. As Figuras (e) e (f) mostram a magnitude da transformada wavelet (wavelet de Morlet[Morlet, 1981]) dos sinais (a) e (b), respectivamente[Protázio, 2001].

a perda de informação temporal. Para isso, a WFT utiliza-se de uma janela de observação deslocada no domínio do tempo do sinal analisado. Assim, é calculada a TF da porção visível do sinal, por meio da janela, em cada posição que a janela se encontra. Considerando a função j(t) como sendo a janela, pode-se definir a equação WFT de um sinal contínuo f(t) por

$$F(u,b) = \int j(t-b)f(t)e^{-2i\pi ut} dt.$$
 (4.2)

Apesar da WFT apresentar uma solução para a deficiência encontrada pela FT, essa abordagem ainda não é a mais adequada. Isso porque, após a definição do tamanho da janela de observação ela se mantém constante para todas as freqüências. Entretanto, muitos sinais necessitam de uma abordagem mais flexível, onde o tamanho da janela seja

variável, a fim de observar com maior precisão um evento no tempo e na freqüência. Essa necessidade se deve ao fato de ser quase impossível determinar um tamanho de janela ideal, que ofereça uma ótima precisão na localização de eventos nas análises de sinais com freqüências muito distintas. Com isso, se torna natural a exploração de outras técnicas, como a transformada wavelet. Um bom exemplo da capacidade dessa técnica pode ser observada na Figura 4.1 (e,f), que representa a magnitude dos coeficientes obtidos via transformada wavelet contínua (Continuous Wavelet Transform - CWT) para os sinais da Figura 4.1 (a,b), respectivamente. É possível perceber como a técnica é capaz de obter uma distinção entre os sinais analisados, localizando as freqüências de maneira distinta no domínio temporal.

4.1.2 Transformada Wavelet

A idéia fundamental sobre wavelets é a análise de acordo com a escala. Wavelets são funções que satisfazem certos requisitos matemáticos usados na representação de dados ou outras funções. A idéia de se utilizar funções para representar outras funções não é nova, e existe desde que Fourier descobriu que senos e co-senos podem ser utilizados para representar outras funções, porém, na análise de wavelet a escala possui um papel importante no processamento dos dados.

A wavelet pode ser processada utilizando-se diferentes escalas ou resoluções. Assim, quando um sinal é observado em uma grande "janela" de tempo, as características observadas daquele sinal serão mais grosseiras, e quando for utilizada uma pequena "janela" de tempo, as características desse sinal serão mais refinadas. Isso quer dizer que é possível observar tanto características mais aparentes quanto as menos perceptíveis. Como, por exemplo, ao observar uma floresta por meio de uma grande "janela", podese verificar a existência de várias árvores, e, enquanto é analisada essa mesma floresta por meio de uma pequena "janela", torna-se possível observar as folhas dessas diversas árvores.

Para obter a característica de janelas variáveis e, conseqüentemente, analisar os sinais de maneira mais flexível, é realizada a decomposição da função f(t) usando-se uma família de funções de dois parâmetros chamadas de wavelets. Um desses parâmetros é responsável pela translação e o outro pela dilatação da função. Assim, podemos definir a transformada wavelet, considerando um sinal contínuo, por

$$F(a,b) = \int f(t)\psi_{a,b}(t) dt, \qquad (4.3)$$

em que os parâmetros a e b variam em \Re , e as funções $\psi_{a,b}$ são denominadas wavelets e definidas por

 $\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right),\tag{4.4}$

onde o símbolo a é o parâmetro de escala e b o parâmetro de tempo. A variação do valor de a tem como efeito a dilatação (a > 1) ou a contração (a < 1), enquanto a variação de b tem como efeito a análise da função f(t) em torno desse ponto. A transformada wavelet para sinais discretos é definida por

$$F_{n,m}(a,b) = a_0^{-m/2} \int f(t)\psi(a_0^m t - nb_0) dt.$$
 (4.5)

O comportamento da função está baseada em dilatações e translações a partir de uma wavelet mãe $\psi(t)$. A escolha dessa wavelet mãe não é única, no entanto, deve satisfazer as seguintes restrições:

$$\int |\psi(t)|^2 dt = 1, \tag{4.6}$$

e

$$\int \psi(t) dt = 0. \tag{4.7}$$

Isso garante, consecutivamente, que a função $\psi(t)$ tenha energia unitária e que tenha caráter ondulatório, ou seja, comporte-se tal qual uma onda (wavelet).

Para obter os coeficientes de wavelets, para cada escala possível, seria necessário realizar exaustivos cálculos, o que dificultaria a utilização da transformada wavelet contínua. No entanto, a transformada discreta é aplicada sobre um subconjunto de escalas e rotações, sobre o qual são realizados os cálculos. Dentro da transformada discreta, segundo Daubechies[Daubechies, 1992], podem ser observadas duas abordagens distintas:

1) Sistemas redundantes discretos (frames); 2) Ortonormal e outras bases de wavelets. A segunda abordagem considera a estratégia de análise de multiresolução, desenvolvida por Mallat [Mallat, 1989].

4.1.3 Funções Wavelets Mães

Uma das principais preocupações na utilização da transformada wavelet está na escolha da função wavelet mãe $\psi(t)$. Para realizar essa escolha existe uma série de critérios que devem ser considerados[Torrence and Compo, 1998]:

- Ortogonais ou não ortogonais "Na transformada wavelet utilizando-se famílias de wavelets ortogonais [Meyer, 1992], o número de convoluções em cada escala é proporcional à janela da função wavelet escolhida nesta escala. Isto produz um espectro wavelet que contém "blocos" discretos de energia wavelet e é útil no processamento de sinais, pois fornece uma representação mais compacta do mesmo. Infelizmente para analisar séries temporais, um deslocamento não periódico na série produz um espectro wavelet diferente. Reciprocamente, a transformada wavelet obtida utilizando-se famílias de wavelets não ortogonais [Meyer, 1992] é altamente redundante em escalas maiores, onde o espectro wavelet em tempos adjacentes é altamente correlacionado. A transformada wavelet não ortogonal é útil na análise de séries temporais (válido também para séries espaciais), onde atenuações e variações contínuas na amplitude wavelet são esperadas".
- Complexa ou real "Uma função wavelet complexa irá fornecer informação da amplitude e da fase e é mais bem adaptada para capturar comportamentos oscilatórios de séries temporais. Uma função wavelet real fornece apenas informação sobre uma componente e pode ser utilizada apenas para localizar picos e descontinuidades".
- Suporte "A resolução de uma função wavelet é determinada pelo balanço entre seu suporte no espaço real e o seu suporte no espaço de freqüência. Uma função com um suporte mais compacto (mais estreita) terá uma boa resolução no domínio do tempo e uma resolução mais pobre no domínio de freqüência. Enquanto uma função com suporte mais amplo (mais larga) terá uma resolução mais pobre no domínio do tempo e uma boa resolução no domínio de freqüência (características determinadas pelo princípio da incerteza de Heisenberg)".
- Formato "A função wavelet escolhida deve refletir o tipo de características presentes na série temporal. Para séries com picos ou descontinuidades, uma boa escolha seria a wavelet de Haar, enquanto que para séries mais suaves e com variações mais sutis deve-se escolher uma função como a wavelet de Morlet. Se o interesse principal é a obtenção do espectro de energia wavelet, então a escolha da função wavelet não é crítica e qualquer uma delas irá fornecer o mesmo resultado qualitativo".

4.1.4 Wavelets Bidimensionais

As wavelets observadas de maneira unidimensional, até o momento, podem ser facilmente estendidas para a forma bidimensional. Para isso, basta tratar de forma análoga a transformada de wavelet bidimensional, onde os parâmetros $a = (a_1, a_2)$ e $b = (b_1, b_2)$, da função f(t), agora são tratados como vetores. Dessa maneira, para o caso bidimensional tem-se

$$\langle f, \psi_{a,b} \rangle \equiv W f(a,b) = \int \int f(t) \frac{1}{a} \psi \left(\frac{t-b}{a} \right) dt,$$
 (4.8)

e a condição abaixo deve ser satisfeita

$$\int \int \psi(t) dt = 0. \tag{4.9}$$

4.1.5 Exemplos de Wavelets

O poder de análise da transformada wavelet é explorado em diversas áreas de estudo e pesquisa, como astronomia, engenharia, música, matemática, medicina, geologia e processamento de imagens. Essa grande exploração, associada à flexibilidade de escolha da função wavelet mãe, culminou no surgimento de inúmeras funções wavelets. A seguir serão mostradas algumas das funções mais utilizadas:

• Wavelet de Haar - A wavelet de Haar (unidimensional) é a mais simples de todas as wavelets e pertence à família das wavelets ortogonais com suporte compacto, e é definida[Kumar and Foufoula-Georgiou, 1994] como

$$\begin{array}{lll} \psi(t) &=& +1 \ \text{se} \ 0 \leq t < 0.5 \\ \\ \psi(t) &=& -1 \ \text{se} \ 0.5 \leq t < 1 \\ \\ \psi(t) &=& 0 \ \text{caso contrário.} \end{array}$$

• Wavelet Chapéu Mexicano – A wavelet Chapéu Mexicano (unidimensional), muito utilizada para detecção de borda, é a derivada segunda da função gaussiana $f(t) = e^{-1/2t^2}$, pertence à família de wavelets não ortogonais e é definida [Kumar and Foufoula-Georgiou, 1994] como

$$\psi(t) = \frac{2}{\sqrt{3}} \pi^{-0.25} (1 - t^2) e^{-0.5t^2}.$$

41

• Wavelet Morlet – A wavelet Morlet (unidimensional) pertence à família de wavelets não ortogonais e é definida [Kumar and Foufoula-Georgiou, 1994] como

$$\psi(t) = \pi^{-1/4} \left(e^{-i\omega_0 t - e^{-1/2\omega_0^2}} \right) e^{-1/2t^2}.$$

• Wavelet Daubechies – A wavelet de Daubechies (unidimensional) pertence à família de wavelets ortogonais. Para cada inteiro r, a base ortogonal para $L^2(\Re)$ está definida[Daubechies, 1992] como

$$\phi_{r,i,k}(x) = 2^{j/2} \phi_r(2^j x - k), j, k \in \mathbb{Z},$$

onde a função $\phi_r(x)$ em $L^2(\Re)$ tem propriedade que $\phi_r(x-k) \mid k \in \mathbb{Z}$ é uma base seqüencial ortogonal em $L^2(\Re)$. Ainda, j é índice de escala, k é o índice de translação e r é o índice de filtragem.

• Wavelet Gabor – A wavelet de Gabor (bidimensional) é extensivamente utilizada por apresentar bons resultados em sua aplicação. Possui as partes real e imaginária, e forma um completo, mas não ortogonal, conjunto de bases. A função wavelet mãe de Gabor é definida[Manjunath and Ma, 1996] como

$$\psi(x,y) = \left(\frac{1}{2\pi\sigma_x\sigma_y}\right) \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right) + 2\pi jWx\right].$$

A Figura 4.2 mostra a cobertura no plano de freqüência espacial de um conjunto de wavelets de Gabor. Contudo, essa cobertura gera redundâncias, como poder ser observado nas intersecções entre as elipses da Figura 4.2. Para eliminar essas redundâncias Manjunath[Manjunath and Ma, 1996] define uma estratégia que segue os seguintes princípios[Rubner and Tomasi, 2000]: 1) Separação uniforme na orientação; 2) Separação exponencial na escala; 3) Cobertura continua do espaço de freqüência. A cobertura no plano de freqüência espacial do conjunto de wavelets de Gabor após essa definição pode ser observada na Figura 4.3.

4.1.6 Aplicações

Como o conhecimento das transformadas wavelets é amplamente disseminado, principalmente na área acadêmica, existem inúmeros exemplos de aplicações das wavelets, que

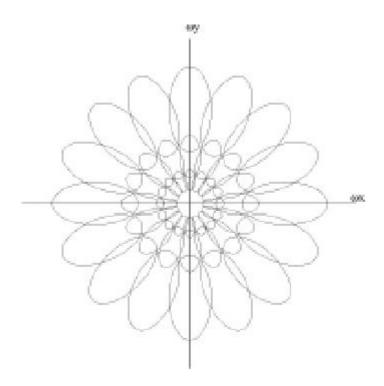


Figura 4.2: Cobertura no plano de freqüência espacial de um conjunto de wavelets de Gabor[Lee, 1996]. Usando três escalas e oito orientações.

se baseiam em diversas funções wavelets mães distintas. Por isso, serão exemplificados apenas alguns trabalhos interessantes que utilizam essa transformada, oferecendo um panorama sobre algumas possíveis aplicações para essa técnica. Segue então uma lista com alguns exemplos:

- A aplicação de wavelet na área de engenharia biomédica, onde se tem, por exemplo, o uso de wavelet de Coiflet 1 para detectar e reconhecer descargas epileptiformes do tipo Espícula-Onda (spikes), possivelmente presentes em sinais de eletroencefalográficos (EEG)[Argoud, 2000].
- Aplicações em sistemas biométricos humanos, onde o principal objetivo é realizar o reconhecimento de um individuo por meio de uma característica biológica intrínseca ao ser humano. Como exemplo, há o uso de wavelet de Gabor[Manjunath and Ma, 1996; Chang and Kuo, 1993] em aplicações que utilizam o reconhecimento por meio da íris do olho humano[Daugman, 1993], garantindo confiabilidade e flexibilidade na operação realizada.
- O uso de wavelet de Gabor e de Daubechies para a indexação e recuperação de

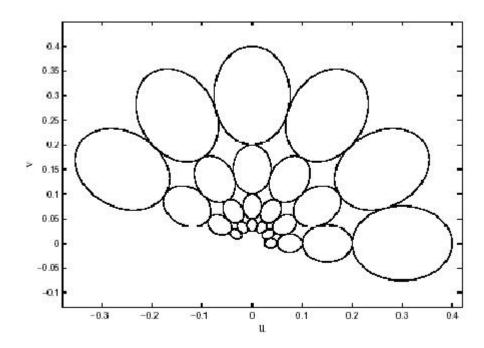


Figura 4.3: Filtros de respostas ortogonais, após aplicar técnica de eliminação de redundância[Manjunath and Ma, 1996]. Usando quatro escalas e seis orientações.

imagens médicas por conteúdo em um banco de imagens [Castañón, 2003]. Nesse processo, a wavelet de Gabor apresentou bons resultados, por se enquadrar melhor com imagens que tenham a textura como principal componente de análise.

• A utilização de wavelet de Daubechies para a análise de eventos geofísicos[Protázio, 2001]. A análise wavelet é aplicada na transferência de escala em perfis geofísicos. O objetivo desse trabalho é a correta avaliação dos parâmetros envolvidos na exploração de reservatórios de petróleo. Oferecendo uma maneira mais segura de obter informações sobre profundidade e volume da jazida, por exemplo.

4.1.7 Considerações Finais Sobre as Wavelets

A necessidade de se utilizar funções que expliquem ou aproximem melhor outras funções, representadas por diversos tipos de sinais, é clara. Isso se deve ao fato de que os sinais existentes na natureza são normalmente de difícil compreensão por meios diretos, inviabilizando a sua investigação nesse sentido. Por conseqüência disso, são inúmeras as pesquisas em diversas áreas de interpretação de sinais, resultando na formalização de técnicas de decomposição de funções, como a transformada de Fourier e transformada wavelet.

A utilização dessas técnicas de decomposição tem sido ampla, principalmente a transformada de Fourier. Contudo, essa técnica não possui a generalização necessária para sinais não estacionários, o que acaba demonstrando uma deficiência dessa técnica perante a necessidade da análise de diversos sinais. Com o objetivo de transpassar essa deficiência é que se formalizou a técnica de transformada wavelet.

A transformada wavelet demonstra-se uma técnica muito poderosa para a análise de sinais. Principalmente quando se trata de sinais não estacionários, oferecendo uma maior flexibilidade na análise do que a transformada de Fourier. Por meio da transformada wavelet é possível gerar observações de um sinal tanto em alto nível (observações mais grosseiras) quanto em baixo nível (observações mais detalhadas). O que proporciona um poder de análise muito grande, tornando-a uma das melhores técnicas para a análise de sinais existentes. Contudo, é importante destacar o fato de que para a utilização dessa técnica se faz necessário a análise e escolha adequada da função wavelet mãe. Escolha tal que pode implicar no sucesso ou fracasso do uso dessa técnica para uma determinada aplicação, já que as diversas funções apresentam características distintas e se adaptam melhor a determinados contextos de aplicação.

Nesse trabalho, o método de extração de características foi implementado baseandose na técnica de transformada wavelet de Gabor. Essa escolha se deve a análise de seu histórico de bons resultados para esse fim[Yu and Bhanu, 2006; Libralao et al., 2005; Vukadinovic and Pantic, 2005; Castañón, 2003]. E principalmente por demonstrar ser uma poderosa técnica na investigação de reconhecimento biométrico, como o reconhecimento de faces[Shen and Bai, 2004; Yang et al., 2002], que é objetivo deste trabalho.

4.2 Introdução a PCA

São diversas as pesquisas em variados domínios de aplicação que encontram um problema em comum, o excesso de dados para serem analisados. Por essa razão é comum encontrar pesquisas nas quais são utilizados algum processo capaz de reduzir o número de informações [Mendonça, 2005]. Contudo, é natural a necessidade da redução desses dados sem que haja perdas significativas de informações, garantindo a qualidade de análises futuras. Considerando essas premissas, encontra-se, nos métodos de análise multivariada, um prático recurso para redução de dados.

A análise multivariada é o ramo da estatística que estuda o comportamento das variáveis dependentes dentro de um determinado conjunto de dados, onde é possível obser-

var n amostras, ou exemplos, de dados que são formados por p variáveis distintas[Johnson, 1982]. Existem diversos métodos de análise multivariada, contudo todas seguem basicamente o mesmo objetivo, que é a redução e a simplificação do problema analisado. Isso significa que esses métodos podem reduzir o número de variáveis a serem computadas, ou mesmo, eliminar variáveis difíceis de serem analisadas ou medidas. Mas, ainda assim, garantindo que as informações mais relevantes não sejam perdidas. Assim, esses métodos permitem aplicar um processo de diminuição da dimensão dos dados de um conjunto, e mesmo havendo perda de informações, normalmente essa perda é reduzida[Oliveira, 1997]. Alguns dos principais objetivos da análise multivariada são:

- Redução de dados Com o objetivo de facilitar a interpretação e análise das informações procura-se redimensionar o espaço dos dados de um problema. Sem que isso cause a perda de informações relevantes;
- **Agrupamento dos dados** Procura-se agrupar dados que possuem alto grau de similaridade. Facilitando o seu entendimento;
- Investigação de dependência entre os dados Comparando-se os dados investigados é possível analisar se as variáveis possuem algum grau de dependência, ou se são independentes entre si;
- **Previsão** Determinando o relacionamento entre as variáveis é possível determinar valores para variáveis de interesse.

A Análise de Componentes Principais (Principal Components Analysis - PCA) é um dos mais antigos e conhecidos métodos de análise multivariada[Jolliffe, 2002; Oliveira, 1997]. Também é conhecido como transformada de Hotelling ou transformada de Karhunen-Loève. São diversos os domínios de aplicação onde essa técnica obtem sucesso[Mendonça, 2005; Yeung and Ruzzo, 2001; De la Torre and Black, 2001; Oliveira, 1997]. O conceito básico da PCA é a de realizar uma projeção dos dados de uma dimensão X para outra Y, onde na dimensão X as variáveis são correlacionadas, e na Y essas novas variáveis não são correlacionadas.

Assim, considerando a aplicação da técnica PCA em um determinado conjunto de dados. As p variáveis correlacionadas de um dado x_i , são projetadas para p variáveis não correlacionadas y_i . Além disso, realiza-se uma ordenação decrescente das variáveis y_i pelas suas variâncias, ou seja, y_1 possui maior variância que y_p . As variáveis y_i são chamadas de componentes principais (PCs), e a i-ésima variável é o i-ésimo componente

principal. Dada a relação de independência entre as variáveis projetadas, é possível escolher apenas algumas q variáveis, onde (q < p), para representar o conjunto de dados com um menor número de informações. Considerando que normalmente há componentes principais irrelevantes, com variância próximo de zero, é possível realizar a diminuição da dimensão dos dados sem que haja perdas significativas. Na Figura 4.4 é ilustrado um exemplo de conjunto de dados de dimensão \Re^3 . Observa-se que todas as amostras estão em um mesmo plano de dimensão \Re^2 , e podem ser fechadas em uma elipse. Ainda é possível observar os eixos formados pela primeira e segunda componentes principais.

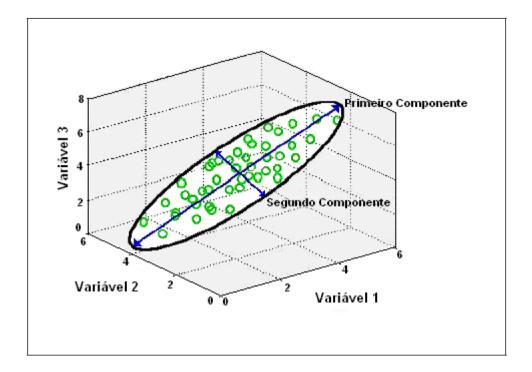


Figura 4.4: Imagem com um exemplo gráfico de um conjunto de dados em \Re^3 , onde as amostras estão todas em um mesmo plano \Re^2 [Mendonça, 2005].

Encontrar uma função que realize a projeção dos dados originais para uma outra dimensão pode ser extremamente complicado. Contudo, encontrar as variâncias das componentes principais resumi-se ao cálculo dos autovalores da matriz de covariância das amostras originais. Assim, dado uma matriz de dados X $(p \times n)$, com p variáveis, a matriz de covariância cov(X) $(p \times p)$ será:

$$cov(X) = \begin{bmatrix} k_{11} & k_{12} & \dots & k_{1p} \\ k_{21} & k_{22} & \dots & k_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ k_{p1} & k_{p2} & \dots & k_{pp} \end{bmatrix}.$$

$$(4.10)$$

Um processo extremamente importante que deve ser realizado antes da criação da matriz de covariância é o de normalização dos dados da matriz X. Com esse procedimento os valores de X passam a formar um conjunto de dados com média zero e variância um. Isso permite que nenhuma variável obtenha maior influência que outra, por problema de dimensão de medida, por exemplo. Dessa maneira, a matriz de covariância cov(X) para os dados normalizados será:

$$cov(X) = \begin{bmatrix} 1 & k_{12} & \dots & k_{1p} \\ k_{21} & 1 & \dots & k_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ k_{p1} & k_{p2} & \dots & 1 \end{bmatrix}.$$
 (4.11)

Com isso, encontra-se uma matriz de covariância que coincide com a matriz de correlação dos dados, onde k_{ij} é a correlação entre X_i e X_j , com $k_{ij} = k_{ji}$. Como é possível observar, a soma dos valores da diagonal principal da matriz de covariância é igual ao número de variáveis dos dados ($\Sigma_{i=1}^{i=p} k_{ij} = p, i = j$). Além disso, por decomposição da PCA tem-se:

$$cov(X)v_i = \lambda_i v_i, \tag{4.12}$$

onde λ_i é o autovalor associado ao autovetor v_i . Os valores do conjunto de λ_i , com i=1,2,...,p, são os autovalores da matriz de covariância. E os v_i , com i=1,2,...,p, são os autovetores. A matriz V de $(p\times p)$, é formada pelos autovetores, onde o i-ésimo vetor v_i constitui a i-ésima coluna de V. Uma característica importante de V é ser ortogonal, ou seja, os autovetores que a constituem são linearmente independentes.

Obtidos os autovalores e autovetores da matriz de covariância, é possível calcular os valores dos dados na nova dimensão. Considerando que as novas coordenadas obtidas são combinações lineares das variáveis originais, temos que:

$$Xv_i = y_i, (4.13)$$

onde y_i é o valor da i-ésima variável na nova coordenada. Assim, considerando X a matriz de $(p \times n)$ com os dados originais, V a matriz de autovetores de $(p \times p)$, e Y a matriz de $(p \times n)$ dos dados na nova coordenada, podemos destacar o processo geral de reconstrução dos novos valores das variáveis como sendo:

$$y_{11} = x_{11}v_{11} + x_{12}v_{21} + \dots + x_{1p}v_{p1}$$

$$y_{21} = x_{21}v_{11} + x_{22}v_{21} + \dots + x_{2p}v_{p1}$$

$$\vdots$$

$$y_{n1} = x_{n1}v_{11} + x_{n2}v_{21} + \dots + x_{np}v_{p1}$$

$$(4.14)$$

Observando os detalhes da teoria da técnica de análise multivariada PCA, podese destacar alguns pontos importantes, observados diretamente nos resultados de sua aplicação[Manly, 1986]:

- As variâncias das componentes principais são iguais aos autovalores da matriz de correlação. Ou seja, sendo λ_i o i-ésimo componente principal, e $y_i = x_1v_{1i} + x_2v_{2i} + \dots + x_pv_{pi}$, então $Var(y_i) = \lambda_i$, onde $Var(y_i)$ é a variação de y_i ;
- Considerando os conceitos de álgebra linear tem-se que, se os autovalores de uma matriz são distintos, então os autovetores correspondentes são linearmente independentes. Isso permite concluir que uma condição necessária para obter V é que os autovalores sejam distintos. Assim, a variância de cada componente principal é única;
- A soma dos autovalores é igual a soma da diagonal principal da matriz de covariância $(\Sigma_{i=1}^{i=p} k_{ij} = \Sigma_{i=1}^{i=p} \lambda_i, i = j)$. Considerando que k_{ij} é a variância de X_i e λ_i é a variância de y_i , então temos que a soma das variâncias das p componentes principais é igual a soma das variâncias das p variáveis originais.

Considerando as características da técnica PCA apresentadas nessa seção, é possível perceber a sua importância para problemas de redução de dados. Com a reestruturação do espaço de características do conjunto de dados iniciais, é possível realizar exclusões de algumas características, sem que isso provoque perdas de informações relevantes. Para isso, basta tratar os autovalores λ_i , ou variâncias, como sendo informações, e quanto maior o seu valor, maior a quantidade de informações que estão sendo representadas.

4.3 Introdução ao Aprendizado de Máquina SVM

A possibilidade de estudar um determinado fenômeno, e a partir de observações conseguir entendê-lo, estimula os pesquisadores de diversas áreas científicas. Uma maneira de interpretar esses fenômenos é encontrando um conjunto de padrões que possam ser distinguidos uns dos outros. Para isso, se faz necessária uma metodologia de classificação, a qual possa identificar corretamente cada conjunto de informações. Contudo, diversas são as técnicas que possibilitam a classificação de um determinado conjunto de dados. Entre elas, podemos destacar as técnicas baseadas no aprendizado de máquina, como Redes Neurais Artificiais (RNAs), Algoritmos Genéticos (AGs), Arvores de Decisão, Raciocínio Baseado em Casos (RBC) e Máquinas de Vetores Suporte (Support Vector Machine - SVMs)[Monard and Baranauskas, 2003].

As SVMs vêm recebendo crescente atenção dos pesquisadores da área de aprendizado de máquinas [Mitchell, 1997]. Seus conceitos são embasados na teoria do aprendizado estatístico [Vapnik, 1995]. Por isso, apresentam algumas características importantes como boa capacidade de generalização. Também são robustas diante de dados com grande dimensão. E ainda têm a convexidade do problema de otimização, o que significa dizer que possui uma função que apresenta apenas um mínimo global [Lorena, 2006]. Essas características das SVMs fazem com que os resultados do uso dessa técnica sejam comparáveis, e muitas vezes, superior aos obtidos com outras técnicas como RNAs [Haykin, 1999; Braga et al., 2000].

Nessa seção são abordados alguns dos conceitos envolvendo as SVMs. Primeiramente são explanados os conceitos gerais de aprendizado de máquina (Subseção 4.3.1). Já na Subseção 4.3.2 há uma breve descrição de conceitos de aprendizado estatístico e SVMs. Na Subseção 4.3.3 são abordadas as SVMs lineares com margens rígidas. Enquanto na Subseção 4.3.4 são abordadas as SVMs lineares com margens suaves. Na Subseção 4.3.6 são apresentadas algumas informações sobre SVMs multiclasses, que foi utilizada no projeto. Na Subseção 4.3.7 há uma breve discussão sobre a biblioteca *libsvm*. Por fim, na Subseção 4.3.8 serão apresentadas algumas considerações finais sobre as SVMs.

4.3.1 Aprendizado de Máquina

O aprendizado de máquina se baseia no princípio de inferência por indução para obter, de modo generalizado, conclusões de um determinado conjunto de informações Lorena,

2006]. Dessa maneira, a partir de alguns dados de exemplo é possível inferir informações sobre o contexto desses dados. Há dois tipos principais de aprendizado indutivo:

- Supervisionado Nesse modo de aprendizado cada classe do conjunto de exemplo é conhecido. Assim, há a idéia da existência de um tutor durante todo o aprendizado.
 O conjunto de informações são passados na forma de pares do tipo: entrada e saída desajada[Haykin, 1999]. No final, espera-se que o modelo encontrado durante o aprendizado seja capaz de gerar corretamente saídas para exemplos ainda não apresentados.
- Não-supervisionado Nesse modo de aprendizado a classe a que pertence o conjunto de exemplos não é conhecido. O agrupamento dos exemplos em determinadas classes é realizada a partir de uma métrica de similaridade. Assim, o próprio sistema toma a decisão de criar novas classes ou agrupar algumas já existentes. Técnicas não-supervisionadas são utilizadas principalmente quando o objetivo é encontrar padrões ou tendências que auxiliem no entendimento dos dados[Lorena, 2006; Souto et al., 2003].

Os conjuntos de dados utilizados em um sistema de aprendizado são normalmente caracterizados por apresentar um vetor de características para cada exemplo, ou dado. Cada característica ou atributo desse vetor expressa um determinado aspecto desse exemplo. Um atributo pode ser definido como nominal ou contínuo. O nominal, ou categórico, pode assumir valores fora de ordem, como cores (amarelo, azul e verde). Já o contínuo assume valores que podem ser definidos em uma seqüência, como os números reais[Monard and Baranauskas, 2003]. O conjunto de dados de um domínio normalmente são agrupados em dois subconjuntos disjuntos: treinamento e teste. O subconjunto de treinamento é utilizado para o aprendizado do conceito. Já o subconjunto de teste é usado para verificar o grau de efetividade do conceito aprendido, ou seja, após o treinamento são apresentados novos exemplos, e verificado se a classe obtida é igual ao do rótulo presente no subconjunto de teste. Dessa maneira é possível estimar os números de acertos e erros da técnica de aprendizado de máquina.

Para a geração de uma boa técnica de aprendizado de máquina é necessário observar alguns aspectos importantes. Um deles é a capacidade de lidar com dados que contenham imperfeições ou ruídos. Uma técnica robusta deve ser capaz de perceber esses ruídos, de modo que a classificação não seja influenciada por essas imperfeições. Além disso, a técnica deve ainda minimizar o impacto de *outliers*, que são exemplos muito

distintos dos demais do conjunto. Esses exemplos extremos podem ser imperfeições dos dados, ou casos muito raros do domínio[Lorena, 2006]. Outro aspecto importante de uma técnica de aprendizado é a capacidade de generalização, que é a capacidade de interpretar corretamente a classe de novos dados. Quando o modelo se especializa nos dados de treinamento, apresentando baixos índices de acerto quando confrontados com novos dados, tem-se a ocorrência de um super-ajustamento (overfitting). Já quando há baixas taxas de acerto mesmo no conjunto de treinamento, tem-se o sub-ajustamento (underfitting). Esses casos podem ocorrer quando os exemplos de treinamento são pouco representativos ou mesmo quando o modelo obtido é muito simples[Monard and Baranauskas, 2003].

4.3.2 SVMs

As Máquinas de Vetores Suporte (SVMs) surgiram pelo emprego direto da teoria do aprendizado estatístico, proposto por Vapnik[Vapnik, 1995], a partir dos estudos iniciados na década de 60[Vapnik and Chervonenkis, 1971; Lorena, 2006]. Muitas aplicações estão sendo desenvolvidas com o uso dessa familia de algoritmos de aprendizado de máquina. O sucesso de sua utilização pode ser observado em diversos domínios de aplicação, como na categorização de textos[Bellili et al., 2003; Dummais, 1998; LeCun et al., 1995], na análise de imagens[Libralao et al., 2005] e reconhecimento de face[Shen and Bai, 2004; Yang et al., 2002]. Os resultados obtidos com essa técnica são comparáveis aos de outras como RNAs[Haykin, 1999; Braga et al., 2000], e em alguns casos superior. Isso se deve as características dessa técnica[Lorena and Carvalho, 2003]:

- Boa capacidade de generalização As SVMs, em geral, conseguem gerar modelos com grande capacidade de generalização. Isso se traduz na eficiência de classificar corretamente os dados que não pertencem ao conjunto de treinamento;
- Robustez em grandes dimensões Normalmente as técnicas de aprendizado de máquina apresentam o problema de *overfitting* na classificação de conjunto de dados com grandes dimensões. Contudo, as SVMs apresentam uma grande robustez diante desse tipo de dados, evitando muitas vezes o problema.
- Convexidade da função objetivo A aplicação das SVMs implica na otimização de uma função quadrática, que possui apenas um mínimo global. Isso representa uma vantagem sobre as RNAs, que possuem mínimos locais na função objetivo.
- Teoria bem definida As SVMs possuem uma teoria matemática bem definida.

A teoria do aprendizado estatístico é a base das SVMs. Assim, é possível explorar as SVMs conhecendo melhor os conceitos do aprendizado estatístico. Para isso, considere um conjunto de testes demonstrados na Figura 4.5. O objetivo do processo de aprendizagem é encontrar um classificador que separe os dados das classes "circulo" e "triângulo". A função que separa, e faz a classificação de cada uma das classes é a linha que está ilustrada na figura entre as classes. Dessa maneira, podemos observar três funções de classificação distintas, onde cada uma apresenta suas peculiaridades.

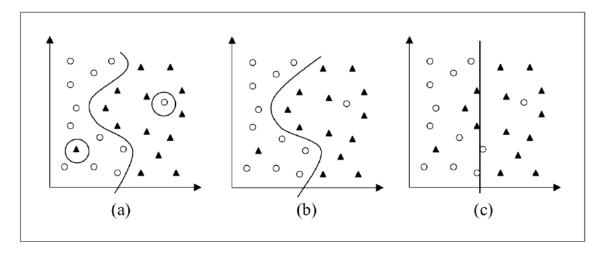


Figura 4.5: Exemplo gráfico de um conjunto de treinamento binário e três modelos de classificadores gerados [Smola and Schölkpof, 2002].

Na Figura 4.5a, pode-se observar uma função que classifica corretamente todos os exemplos do conjunto, inclusive dois que possivelmente são ruídos. Essa é uma função de alta complexidade e com grande sucetibilidade a gerar muitos erros com novos dado, o que seria um super-ajustamento (overfitting) do modelo. Já na Figura 4.5c, o modelo obtido é muito simples, e desconsidera exemplos de classes distintas que estão muito próximas. Nesse caso, tem-se o sub-ajustamento (underfitting) do modelo. Por fim, no caso exposto na Figura 4.5b, há uma função intermediária. É um modelo de complexidade média e que classifica corretamente a maioria dos exemplos, sem se fixar em um ponto individual[Lorena, 2006].

A teoria do aprendizado estatístico oferece meios matemáticos que auxiliam na escolha de um classificador específico para um conjunto de dados. Isso é realizado estimandose o desempenho do classificador com o conjunto de treinamento e a sua complexidade. Dessa maneira, é possível estimar um modelo que obtenha um bom desempenho inclusive para novos dados. Para isso, é utilizado o conceito de minimização do risco estrutural [Vapnik, 1998].

A capacidade de generalização de um modelo de aprendizado pode ser medido por meio do risco (ou erro) esperado de um classificador [Passerini, 2004]. Assim, quanto menor o risco esperado melhor será o classificador. Contudo, não é possível estimar esse erro [Müller et al., 2001]. Por isso, calcula-se o erro empírico que mede o desempenho do classificador no conjunto de dados de treinamento, baseado na taxa de classificações incorretas [Müller et al., 2001]. Espera-se, então, que minimizando esse erro haja uma diminuição do erro sobre os dados de teste. Normalmente, para conjuntos com grande número de dados, essa convergência é válida [Smola and Schölkpof, 2002]. Porém, para conjuntos com número reduzido de dados, e classificadores que memorizam todo o conjunto de treinamento, isso não é garantido. Partindo desse problema surge a idéia de desenvolver limites teóricos para a capacidade de generalização dos modelos de aprendizagem. Com esse objetivo definiu-se o principio indutivo de minimização do risco estrutural [Vapnik, 1998]. Dessa maneira, procura-se minimizar o erro do conjunto de treinamento, juntamente com erro do conjunto de teste. O equilíbrio entre ambos os erros permite obter um modelo com boa generalização e que não apresente super-ajustamento (overfitting).

Toda a apresentação realizada sobre as SVMs até o momento se refere a classificadores binários. Contudo é possível utilizar SVMs para aplicações com multiplas classes. A seguir serão apresentadas quatro tipos diferentes de SVMs: 1) Lineares com margens rígidas, Subseção 4.3.3; 2) Lineares com margens suaves, Subseção 4.3.4; 3) Não lineares, Subseção 4.3.5; 4) Multiclasses, Subseção 4.3.6.

4.3.3 SVMs Lineares com Margens Rígidas

Dado um conjunto de dados de treinamento com duas classes, esse será linearmente separável se for possível separar os dados de cada classe por um hiperplano [Smola and Schölkpof, 2002]. Classificadores que separam os dados por meio de um hiperplano são denominados lineares [Campbell, 2000]. As SVMs são responsáveis por encontrar o melhor hiperplano que divida o conjunto de dados. Para isso, é necessário maximizar as margens de separação. A margem de separação é a menor distância entre o hiperplano que divide os dados, e qualquer ponto do conjunto de treinamento. Na Figura 4.6 a margem de separação é dado por $\frac{d}{2}$, distância entre cada hiperplano H_1 e H_2 , e o hiperplano que divide os dados. O hiperplano ótimo é aquele que gera margens de separação máxima. Como um dado de treinamento nunca pode estar entre as margens de separação, essas SVMs são conhecidas como lineares com margens rígidas. Os dados sobre os hiperplanos H_1 e H_2 são conhecidos como vetores suporte (SVs).

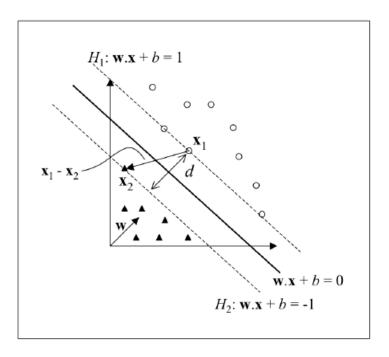


Figura 4.6: Exemplo de um conjunto de treinamento binário dividido por um hiperplano, e suas margens de separação (região entre esse hiperplano e cada hiperplano H_1 e H_2)[Hearst et al., 1998].

4.3.4 SVMs Lineares com Margens Suaves

Na prática, os dados de treinamento em situações reais raramente serão linearmente separáveis. Seja por influência de uma imperfeição, seja pela presença de *outliers*, ou mesmo pela própria natureza do problema, que pode ser não linear. Devido a isso, as SVMs lineares de margens rígidas são estendidas para lidar com conjuntos de treinamento mais gerais. Para isso, permite-se que alguns dados violem as margens de separação imposta. Isso é realizado com a introdução de variáveis de relaxamento [Smola et al., 1999]. Dessa maneira, a SVM permite que alguns dados permaneçam entre as margens de separação, sendo caracterizada como SVM linear com margens suaves. Analisando a Figura 4.7 é possível observar alguns tipos diferentes de vetores suporte (SVs)[Pontil and Verri, 1998]. Há os SVs tradicionais, que estão em tom de cinza na figura, e são chamados de SVs livres. Os dados em preto são SVs limitados, porém corretamente classificação.

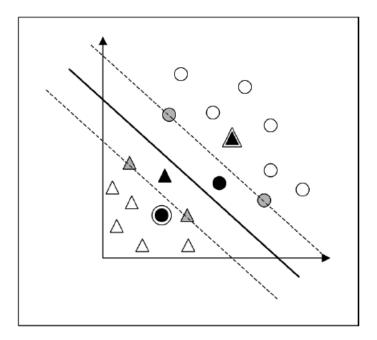


Figura 4.7: Exemplo de um conjunto de treinamento binário com alguns tipos de SVs: livres (em cinza), e limitados (em preto)[Passerini, 2004].

4.3.5 SVMs Não Lineares

As SVMs lineares são eficazes para a classificação de conjuntos de dados lineares, para conjuntos que apresentam distribuição aproximadamente lineares e, ainda, na versão de margens suaves, são tolerantes a presença de algumas imperfeições e até *outliers*[Lorena, 2006]. Contudo, há ainda muitos casos em que os conjuntos de dados não podem ser divididos satisfatoriamente por um hiperplano (conjuntos não lineares). Para este tipo de problema sugere-se o uso de SVMs não lineares. A Figura 4.8(a) representa um exemplo de conjunto de dados em que, para a separação das classes, seria mais adequado o uso de uma fronteira curva, como na Figura 4.8(b).

A solução das SVMs não lineares propõem o mapeamento do espaço original de entrada, para um espaço de maior dimensão, chamado de espaço de características [Hearst et al., 1998]. Escolhendo um mapeamento Φ adequado para a transição dos espaços, é possível aplicar uma SVM linear para o conjunto de treinamento mapeado, realizando, assim, a separação das classes do conjunto. Esse procedimento é motivado pelo teorema de Cover [Haykin, 1999]. Esse teorema afirma que dado um conjunto de dados não linear no espaço de entradas X, este pode ser transformado em um espaço de características \Im , no qual, com grande probabilidade, os dados serão linearmente separáveis. Contudo, há

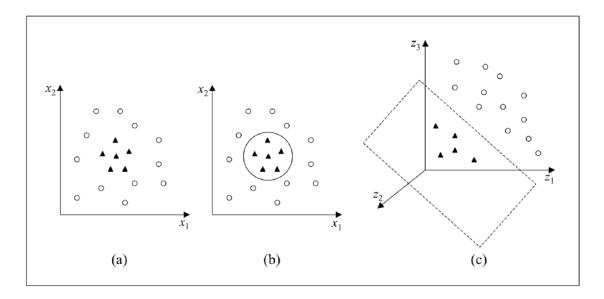


Figura 4.8: Exemplo de um conjunto de treinamento binário. (a)Dados não lineares; (b)Fronteira não linear de classificação; (c)Fronteira linear de classificação no espaço de características[Müller et al., 2001].

duas condições: a primeira é que a transformação seja não linear; e a segunda é que a dimensão do espaço de características seja suficientemente alta. Seguindo esses conceitos, o conjunto de dados da Figura 4.8(a) pode ser mapeado para um espaço de características maior. Assim, pode-se aplicar uma separação usando uma SVM linear, como pode ser observado na Figura 4.8(c).

Como o espaço de características pode ter dimensões muito altas, a computação de Φ pode ser extremamente custosa[Lorena, 2006]. Contudo, a única informação necessária é de como realizar o cálculo de produtos escalares entre os dados do espaço de características. Isso é obtido com o uso de funções conhecidas como kernel K. K é uma função que recebe dois pontos do espaço de entradas e computa o produto escalar desses dados no espaço de características[Herbrich, 2001]. Para garantir a convexidade do problema de otimização, são utilizadas funções kernel que seguem as condições estabelecidas pelo teorema de Mercer[Mercer, 1909; Smola et al., 1999]. Os kernels mais utilizados na prática são os polinomiais, os Gaussianos ou RBF ($Radial\text{-}Basis\ Function$), e os sigmoidais. A Tabela 4.1 apresenta cada um desses kernels, e os parâmetros que devem ser determinados pelo usuário[Chang and Lin, 2006]. O kernel sigmoidal satisfaz as condições de Mercer apenas para alguns valores de δ e κ . Os kernels polinomiais com d=1 também são denominados lineares[Lorena, 2006].

Tipo de Kernel	Função $K(x_i, x_j)$	Parâmetros
Polinomial	$(\delta(x_i \cdot x_j) + \kappa)^d$	δ , κ e d
Gaussiano (RBF)	$\exp(-\sigma \ x_i - x_j\ ^2)$	σ
Sigmoidal	$\tanh(\delta(x_i \cdot x_j) + \kappa)$	$\delta \in \kappa$

Tabela 4.1: Tabela com as funções Kernels mais comuns e seus parâmetros.

4.3.6 SVMs Multiclasses

A princípio, as SVMs são empregadas para problemas de classificação binári, contudo, com algumas modificações é possível extender o seu uso para problemas de multiclasses. Em geral há duas estratégias que podem ser empregadas nesse processo de adaptação: decomposicionais e diretas[Lorena, 2006]. Na primeira o problema multiclasse original é dividido em subproblemas binários, podendo, dessa maneira, determinar os seus classificadores binários. Posteriormente é realizado um processo de reconstrução, onde as saídas dos classificadores encontrados serão combinados, o que permite encontrar o padrão de classificação dos exemplos do problema original[Mayoraz and Moreira, 1996]. Já na estratégia direta o objetivo é o de reformular o algoritmo original da técnica de aprendizado, para uma versão multiclasses. Contudo, esse processo nem sempre é possível ou fácil de se realizar. E quando é possível, pode gerar algoritmos computacionalmente caros[Passerini, 2004; Hsu and Lin, 2002].

Entre as estratégias decomposicionais mais comuns conhecidas estão a um-contratodos [Vapnik, 1995] e todos-contra-todos [Knerr et al., 1990]. Na primeira, sendo κ o número de classes, são gerados κ classificadores. Onde cada um será responsável por separar uma determinada classe das demais. Enquanto isso na estratégia de todos-contratodos, dado que κ é o número de classes, são gerados $\frac{k(k-1)}{2}$ classificadores, em que cada um será responsável pela separação de um par de classes (i,j), em que $i \neq j$ [Lorena, 2006]. O resultado final nessa estratégia é gerado a partir de um esquema de votação por maioria[Krebel, 1999]. Assim, para um novo dado, cada classificador gera um voto para a classe de preferência. A mais votada determina a classe desse dado.

4.3.7 Utilizando SVM - libsym

Para o desenvolvimento desse trabalho foi proposta a aplicação de uma técnica de classificação SVM. Dessa maneira, se fez necessário explorar as possibilidades da utilização de uma biblioteca para realizar essa etapa. Assim, foi realizado um levantamento so-

bre a biblioteca *libsvm*, que consiste de uma biblioteca de aprendizado de máquina SVM para classificação e regressão, e que é simples, fácil de usar e eficiente[Chang and Lin, 2006]. Essa biblioteca possibilita o uso dos classificadores C-SVM e nu-SVM, do uso do one-class-SVM, e das regressões epsilon-SVM e nu-SVM. E ainda disponibiliza uma ferramenta (*grid.py*) de seleção automática do modelo de treinamento para o classificador C-SVM.

Para a utilização da libsvm no projeto, primeiramente foi consultado o seu manual[Hsu et al., 2006]. Posteriormente foi realizado um levantamento das opções de treinamento disponíveis. Entre as opções de escolha havia o tipo de SVM (C-SVC, nu-SVC, one-class SVM, epsilon-SVR e nu-SVR), a função Kernel (linear, polinomial, RBF e sigmoidal), o custo c para alguns tipos de SVM, como o C-SVC, e os parâmetros da função Kernel: d, g (que representa γ) e r (que representa κ). Seguindo as recomendações do manual, o procedimento a ser seguido deveria ser: normalizar os dados do treinamento e do teste; estimar os melhores valores para g e r com a ferramenta grid.py; realizar o treinamento usando os valores de g e r encontrados; e por fim avaliar o resultado do modelo criado, aplicando os dados de teste.

4.3.8 Considerações Finais Sobre a SVM

Nessa seção foram discutidos os principais conceitos básicos envolvendo o aprendizado de máquina, especialmente as Máquinas de Vetores Suporte (SVMs). A atenção dada as SVMs pelos pesquisadores da área de aprendizado de máquina vem crescendo, pois os resultados obtidos em diversos domínios de aplicações são, muitas vezes, superiores ao das RNAs[Haykin, 1999; Braga et al., 2000]. Muitas áreas científicas vêm se beneficiando da eficiência das SVMs, como as áreas de análise de imagens e reconhecimento de faces[Libralao et al., 2005; Shen and Bai, 2004; Yang et al., 2002]. Isso se deve ao fato das SVMs terem uma teoria bem definida, que garante uma boa generalização do problema, além de robustez.

São diversos os tipos de SVMs que podem ser encontradas. Entre elas estão as lineares, com margens rígidas e com margens suaves, e também as não lineares. A princípio essas SVMs foram desenvolvidas para problemas binários. Contudo, pode-se estender o seu domínio para problemas multiclasses aplicando as estratégias de decomposição ou a direta. Devido as restrições computacionais que normalmente são geradas pela estratégia direta, é mais comum o uso da decomposição. Nesse caso, as duas principais técnicas empregadas são a de um-contra-todos, que produz classificadores para cada classe do

conjunto treinamento, e a de todos-contra-todos, que produz um classificador para cada par de classes do conjunto treinamento.

A escolha da SVM como um classificador para esse projeto se deve, a princípio, as grandes contribuições obtidas por essa técnica no meio acadêmico. Os benefícios para a área de reconhecimento de face, tema do trabalho, são reconhecidos[Shen and Bai, 2004; Yang et al., 2002]. Além disso, como o conjunto de classes e dados utilizados no trabalho é grande, é de extrema importância o emprego de classificadores robustos e eficientes. Qualidades essas que a técnica de aprendizado de máquina SVM possui.

Metodologia

Os trabalhos desenvolvidos na área de biometria exploram diversas características distintas dos seres humanos, como a impressão digital, a íris, e a face. Muitos métodos, porém, são considerados invasivos, como é o caso da íris e da retina. Isso dificulta a aceitação desses sistemas, deixando-os restritos ao uso muito específicos, como em sistemas com alto nível de segurança. Contudo, métodos como impressão digital e face são facilmente aceitos, principalmente por serem bastante intuitivos e menos invasivos. Os trabalhos baseados em face humana, muitas vezes, têm como objetivo apenas a localização e caracterização do objeto sob investigação, enquanto outros exploram apenas a identificação de uma pessoa a partir da imagem da face. Poucos são os trabalhos que exploram a localização e a identificação de uma face.

Este trabalho propõe o desenvolvimento de um sistema para reconhecimento de faces (ilustrado na Figura 5.1). Ele é composto de dois módulos:

- 1. A localização e rastreamento de face Implementação de um sistema de rastreamento de faces humanas, a partir de quadros de imagens capturadas por uma câmera digital. O sistema deverá ser capaz de localizar e isolar a face do indivíduo, quando uma pessoa entrar na área de captura, e acompanhar essa face durante todo o período em que a pessoa se encontrar presente (Seção 5.1).
- 2. O reconhecimento de face Implementação de um sistema de reconhecimento

de faces humanas, a partir da imagem da região da face. Será aplicado o algoritmo de reconhecimento que fornecerá as informações necessárias para identificar uma determinada pessoa (Seção 5.2).

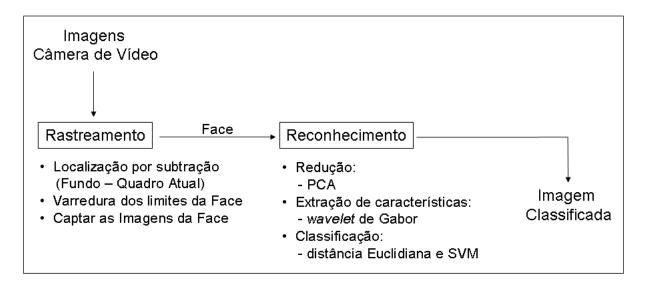


Figura 5.1: Ilustração do funcionamento do sistema de rastreamento e reconhecimento de face proposto.

5.1 Rastreamento de Face

A primeira etapa abordada nesse trabalho foi a implementação de um sistema de localização e rastreamento de face. A partir de uma seqüência de imagens captadas por uma câmera de vídeo, o sistema deverá ser capaz de identificar a face de um indivíduo e em seguida rastreá-la. O processo permite duas formas de extração, gerando dois tipos de imagem de face. Uma imagem com o rosto inteiro do indivíduo, Figura 5.2(a), e outra contendo apenas a região central da face, com olhos, boca e nariz, Figura 5.2(c).

Inicialmente, foram levantadas algumas características específicas da aplicação, que formam o contexto no qual este sistema de localização e rastreamento está inserido. Considera-se, por exemplo, que no ambiente em que está sendo filmado haverá apenas um indivíduo (diminuindo a complexidade da imagem a ser analisada). Além disso, a cabeça dessa pessoa deve estar sempre na parte superior do vídeo, na posição vertical (como na Figura 5.2), ou levemente inclinada. Ainda, espera-se que não haja nenhum outro objeto (acima da linha dos ombros) que possa obstruir a localização da face (como mãos levantadas, chapéus largos, etc). Por fim, leva-se em consideração o formato elíptico da

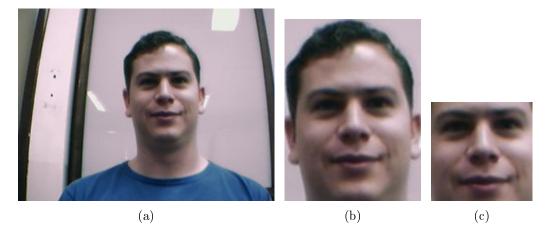


Figura 5.2: Imagens de uma face: (a) Imagem original do ambiente que estava sendo filmado; (b) - Imagem com o rosto inteiro do indivíduo; (c) Imagem contendo apenas a região central da face.

face humana, que apresenta dimensões na vertical e horizontal, segundo uma proporção de $\frac{4}{3}$, aproximadamente (observado empiricamente).

Para alcançar o objetivo proposto foi desenvolvido um sistema flexível, que considerou não somente as dimensões da face, mas também a sua disposição espacial dentro do contexto da aplicação. Nesse sistema é definida uma imagem de fundo (background), que determina o ambiente que está sendo filmado pela câmera, sem a presença de uma face. Em seguida, o sistema verifica quadro a quadro se há alguma alteração na imagem do ambiente. Quando constatado o movimento de algum objeto, identificado pela alteração do ambiente filmado, começa o processo de localização da face. Para isso são considerados alguns aspectos de dimensão da face, assim como sua disposição espacial. Devido ao formato elíptico da face é possível enquadrar a face usando um retângulo (com dimensões proporcionalmente equivalentes as da face). A partir da localização, o processo é repetido quadro a quadro possibilitando realizar o rastreamento da face. Dessa maneira é possível acompanhar o movimento da face durante todo o período em que se localizar perante a câmera. A Figura 5.3 ilustra o processo descrito acima.

Para identificar um objeto na cena em que estão sendo geradas as imagens é realizado o processo de subtração de quadros. Onde o quadro atual é subtraido pelo quadro de fundo (background). Utilizando o resultado obtido, é realizada uma varredura, com o objetivo de procurar algum objeto que não se encontrava no quadro de fundo. Para isso, são varridas todas as linhas do quadro, de cima para baixo, procurando uma região (um conjunto de pixels vizinhos). Caso seja encontrado é considerado a parte superior da face do indivíduo, e em seguida são iniciados os processos de localização dos pontos que

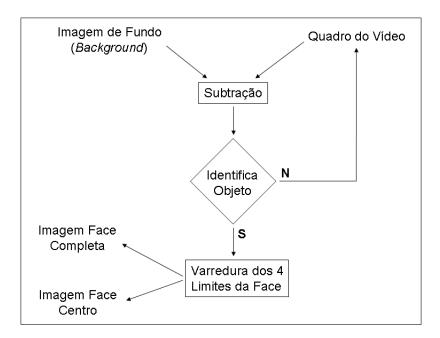


Figura 5.3: Ilustração do funcionamento do sistema de localização e rastreamento de face proposto.

representam o lado direito e esquerdo da face. Para isso são executadas varreduras em todas as colunas da direita para a esquerda, e da esquerda para direita para encontrar, respectivamente, os pontos da direita e da esquerda da face.

Usando os três pontos da face encontrados pelo sistema (superior, direito e esquerdo) é possível estimar um quarto ponto, o inferior. Isso é realizado para permitir enquadrar a região da face em um retângulo, e extrair a imagem relevante para o sistema. Para encontrar esse quarto ponto é considerado uma razão de proporcionalidade, onde a distância entre os pontos na horizontal é aproximadamente um terço maior que a distância entre os pontos da vertical. Essa razão é estimada empiricamente considerando o formato elíptico da face humana.

Com a face localizada é possível extrair as duas imagens que serão utilizadas no sistema de reconhecimento. A primeira, Figura 5.2(a), representa a face com o rosto inteiro do indivíduo. Já a segunda imagem, Figura 5.2(c), representa apenas a parte central da face, preservando seus principais fatores característicos como olhos, nariz e boca. As dimensões das imagens capturadas são variáveis e dependentes do sistema de localização. Contudo, há uma razão de proporcionalidade entre essas dimensões, como mencionado anteriormente.

5.2 Reconhecimento de Face

A segunda etapa desse trabalho aborda a implementação do sistema de reconhecimento de face humana. Esse sistema deve ser capaz de identificar uma determinada pessoa a partir de uma imagem de sua face. Foram implementados dois processos diferentes de reconhecimento. O primeiro consiste da aplicação da transformada wavelet de Gabor diretamente nas imagens. Já no segundo, reduz-se o número de dados usando PCA e posteriormente aplica-se a wavelet de Gabor. Em seguida foi implementado um classificador por SVM, e outro por distância Euclidiana. Para a avaliação foram utilizados os dois tipos de imagens da face (completa e parcial), extraídas no sistema de rastreamento. Empregou-se ainda, as imagens do banco de imagens FERET, usando o mesmo padrão de imagens da face completa e parcial. Assim, foi possível avaliar a eficiência do sistema de reconhecimento, além da contribuição de cada processo avaliado. A Figura 5.4 ilustra o processo descrito acima.

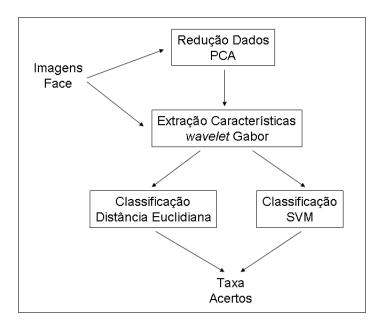


Figura 5.4: Ilustração do funcionamento do sistema de reconhecimento de face proposto.

A primeira ação tomada nessa etapa foi o levantamento de algumas técnicas de extração de características. Em seguida foi implementada a transformada de wavelet de Gabor. Com essa técnica é possível realizar uma análise espacial do conteúdo da imagem da face, e extrair dela uma série de características específicas. Para isso, é necessário informar valores de escala e orientação do processo a ser aplicado na imagem, o que definirá o número de atributos extraídos. Isso permite gerar um vetor de atributos que

caracteriza a imagem original. Esse vetor é de tamanho fixo, dadas as mesmas escala e orientação, independente do tamanho da imagem. Com esse vetor é possível realizar o processo de classificação das imagens.

Inicialmente, a transformada de wavelet de Gabor foi aplicada diretamente sobre uma imagem RGB original, gerando um vetor de atributos (características) de dimensões bastante altas. Quanto maior as dimensões dos vetores, maior o tempo de processamento. Com o intuito de diminuir a quantidade de informações a serem analisadas pela transformada wavelet de Gabor, e consequentemente reduzir o tempo de processamento, implementou-se a técnica de PCA (Seção 4.2).

Esse segundo processo tem início com a aplicação da técnica de PCA sobre a imagem RGB original. O resultado desse processo gera um vetor de informação com um terço do tamanho original. Para isso, considera-se que uma imagem RGB possui três dimensões, ou seja, cada uma das componentes de cor (R, G e B) é considerada uma dimensão distinta. A técnica PCA é utilizada aqui para extrair as informações mais relevantes da imagem, de um modelo com três dimensões para um com apenas uma dimensão.

Para aplicar PCA, foi necessário desenvolver um modelo contendo exemplares e os seus atributos. Assim, os componentes de cor foram considerados os atributos a serem analisados, enquanto cada *pixel* da imagem foi considerado um exemplar do modelo. A partir daí foi possível aplicar PCA, que destaca os atributos por ordem de relevância. Dessa maneira é possível escolher apenas um atributo que represente da melhor maneira possível os três atributos originais. É importante destacar que esse atributo único escolhido é uma composição das três originais, realizada pela PCA.

Tendo escolhido o atributo mais relevante, realiza-se o processo de extração das informações da imagem. Isso ocorre pela reconstrução das informações, baseada na composição do atributo escolhido, dada pelos atributos originais (RGB). Dessa maneira, assim que a imagem original é processada, pela PCA, as informações resultantes são passadas para a transformada wavelet de Gabor. As configurações da técnica de Gabor são as mesmas que as utilizadas quando aplicada diretamente a imagem original. Assim, ao final desse processo, é gerado um vetor de atributos que caracteriza a imagem original.

Após as implementações dos processos de extração de características das imagens, deu-se início a implementação do processo de classificação dos resultados obtidos. Dois processos de classificação foram avaliados. Um utiliza uma medida de distância Euclidiana entre os vetores de atributos. Enquanto o outro processo passa pela classificação através de

aprendizado de máquina baseado em SVM. A partir das classificações é possível analisar e avaliar os resultados dos processos de extração de características, e ainda indicar qual o método mais eficiente para a classificação, no contexto da aplicação.

5.2.1 Classificação por Distância Euclidiana

Para realizar o processo de classificação de distância foi necessário, primeiramente, preparar as imagens do banco de imagens de face. Essa preparação consistiu na divisão das imagens em dois grupos distintos: grupo teste e controle. Assim, uma imagem por individuo foi destinada ao grupo controle, enquanto as demais foram alocadas ao grupo teste. Utilizando as imagens do grupo controle foram gerados os vetores de atributos que, na fase de classificação, serviram para a comparação com os vetores de atributos resultantes das imagens do grupo teste.

O método de classificação de distância utilizada foi baseada na distância Euclidiana (Equação 5.1). Assim, cada vetor de atributos resultantes do grupo teste foi comparado aos vetores de atributos do grupo controle. A partir daí, foi possível classificar a qual indivíduo pertencia determinada imagem, considerando-se como critério de similaridade a menor distância (D) encontrada.

$$D = \sqrt{\sum_{i=1}^{i=n} (x_i - y_i)^2}.$$
 (5.1)

5.2.2 Classificação por SVM

O segundo processo de classificação implementado foi o de aprendizado de máquina utilizando SVM (Seção 4.3). Assim como no processo de classificação de distância, esse processo utilizou dois grupos de imagens: treinamento e teste. O grupo treinamento foi utilizado para treinar a SVM, para que fosse possível generalizar os resultados da classificação. Enquanto o grupo teste foi aplicado para verificar o erro gerado pelo classificador. Nesse sistema de classificação cada indivíduo foi considerado uma classe distinta.

A SVM escolhida para este trabalho foi a SVM multiclasses (Subseção 4.3.6). Isso se deve as características do sistema proposto, onde cada indivíduo representa uma classe distinta, e as imagens de face desses indivíduos são os exemplos de cada classe. Esta SVM foi utilizada a partir da biblioteca *libsvm* (Subseção 4.3.7). Para isso, primeiramente foi definido como padrão o tipo de SVM como C-SVC, e o núcleo como sendo o RBF. Em

seguida, para cada base de imagem, foi necessário encontrar os valores dos parâmetros custo (c) e gama (g), usando a ferramenta grid.py. Após determinar esses valores, foi iniciado o treinamento dos dados, que originou o modelo de classificação. Aplicando-se as imagens de teste sobre o modelo encontrado são obtidas as taxas de acertos e erros do classificador.

5.2.3 Bases de imagens

Para realizar a avaliação dos métodos de extração de características, e dos processos de classificação, foram utilizadas quatro bases distintas de imagens de face. Duas dessas bases foram geradas utilizando a etapa de rastreamento de face. Enquanto as outras duas foram geradas a partir de uma compilação da base FERET de imagens de face. A partir da análise dessas quatro bases de imagens, foi possível verificar qual dessas oferecia uma melhor caracterização da face de um indivíduo.

A primeira base foi formada por imagens próprias que contém a representação do rosto inteiro de um indivíduo, Figura 5.2(a). Dessa maneira, todas as informações relativas ao rosto do indivíduo são consideradas no momento da extração de características. Todas as imagens da base são coloridas (RGB). A base foi formada com quinze imagens frontais da face por indivíduo. Foram utilizados dez indivíduos, resultando em um total de 150 imagens nessa base. Tanto na classificação de distância Euclidiana quanto SVM, essa base foi dividida em 100 imagens para o grupo controle (treinamento) e 50 para o grupo teste. Esta base foi nomeada **Própria - face completa**.

A segunda base foi formada por imagens próprias que contém apenas a parte central do rosto de um indivíduo, Figura 5.2(c). Com isso, as informações coletadas pelo processo de extração de características, preserva os principais fatores relacionados a face humana, como nariz, boca e olhos. As imagens dessa base são todas coloridas (RGB). Assim como a primeira base, essa foi formada por quinze imagens frontais por indivíduo. Foram utilizados nove indivíduos, em um total de 135 imagens. Tanto no processo de classificação por distância Euclidiana quanto SVM foram utilizadas 90 imagens para o grupo controle (treinamento) e 45 para o grupo teste. Esta base foi nomeada **Própria - centro da face**.

A terceira base de imagens foi formada a partir de uma compilação de um subgrupo da base de imagens FERET, como ilustrado na Figura 5.5(b). Para formar esse subgrupo foram escolhidas apenas imagens frontais. Foram selecionados 62 indivíduos que possuíam pelo menos quatro imagens na base. Assim, foi gerado uma base com 248 imagens frontais de 62 indivíduos. Foram utilizadas imagens coloridas (RGB) de dimensões distintas, e que compreendiam todo o rosto do indivíduo. Para ambos os métodos de classificação foram utilizados 186 imagens para o grupo controle (treinamento) e 62 imagens para o grupo teste. Esta base foi nomeada **FERET** - **face completa**.

Já a quarta base, também um subgrupo da base de imagens FERET, foi formada a partir de um processo semelhante ao da terceira base. Contudo, ao invés de serem selecionadas imagens do rosto inteiro, foi selecionado apenas a porção central do rosto, como ilustra a Figura 5.5(c). Nesse caso foram selecionados 50 indivíduos, onde cada um contribuiu com quatro imagens, formando uma base com 200 imagens. Para aplicar os métodos de classificação as imagens foram divididas, formando o grupo de controle (treinamento) com 150 imagens, e o grupo teste com 50 imagens. Esta base foi nomeada **FERET - centro da face**.

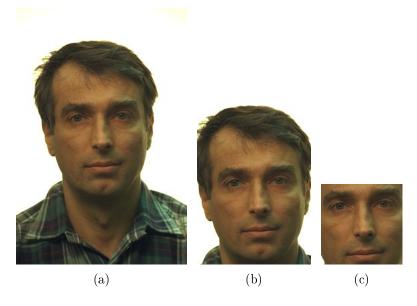


Figura 5.5: Exemplo de uma imagem do banco de imagens FERET: (a) - Imagem original do banco; (b) - Imagem com o rosto inteiro do indivíduo; (c) Imagem contendo apenas a região central da face.

Considerando os processos de extração de características, de classificação, e, por fim, as bases de imagens distintas, foi possível formar dezesseis casos de avaliação. Cada um desses grupos e os atributos utilizados são identificados na Tabela 5.1. Para cada grupo é identificado um processo de extração de características, um processo de classificação e uma base de imagens.

Tabela 5.1: Grupos de avaliação e seus atributos quanto a processo de extração de características, processo de classificação e base de imagens.

	Extra	ação	Classificação	ıção		Bases de	Bases de Imagens	
Grupos de	I				Pró	Própria	FEI	FERET
Avaliação	Gabor	PCA	Distância	SVM	Face	Centro da	Face	Centro da
		Gabor	Euclidiana		Completa	Face	Completa	Face
1	X		X		X			
2	X		X			X		
3	X		X				X	
4	X		X					X
5	X			X	X			
9	X			X		X		
2	X			X			X	
8	X			X				X
6		X	X		X			
10		X	X			X		
11		X	X				X	
12		X	X					X
13		X		X	X			
14		X		X		X		
15		X		X			X	
16		X		X				X

A quantidade de grupos obtidos se deve a três fatores:

- 1. A divisão da etapa de extração de características das imagens em dois processos: um utilizando a técnica de transformada wavelet de Gabor, e outro a técnica de PCA antes da transformada wavelet de Gabor;
- 2. A adoção de dois processos de classificação dos dados obtidos com a extração de características. Um aplicando a distância Euclidiana, e outro usando um classificador gerado a partir da técnica de aprendizado de máquina SVM;
- 3. A utilização das quatro bases de imagens definidas anteriormente: FERET (face completa e centro da face) e Própria (face completa e centro da face).

Para proceder com a etapa de classificação foi necessário realizar a divisão dos dados obtidos, formando um grupo de controle (treinamento) e um grupo de teste. Cada grupo foi formado aleatoriamente, evitando vícios de escolha. Contudo, as classes foram devidamente balanceadas, ou seja, para cada classe existia um mesmo número de exemplos no grupo controle (treinamento), e uma mesma quantidade no grupo de teste. Dessa maneira, foi obtido para cada base de imagens a seguinte quantidade de exemplos:

- 1. **FERET face completa**: 186 para o grupo controle (3 para cada classe); 62 para grupo de teste (1 para cada classe); total de 62 classes;
- 2. **FERET centro da face**: 150 para o grupo controle (3 para cada classe); 50 para grupo de teste (1 para cada classe); total de 50 classes;
- 3. **Própria face completa**: 100 para o grupo controle (10 para cada classe); 50 para grupo de teste (5 para cada classe); total de 10 classes;
- 4. **Própria centro da face**: 90 para o grupo controle (10 para cada classe); 45 para grupo de teste (5 para cada classe); total de 9 classes.



Resultados

Este capítulo apresenta, inicialmente, os resultados obtidos com o desenvolvimento do sistema de localização e rastreamento de face (Seção 6.1). Posteriormente, são apresentados os resultados obtidos com a implementação do sistema de reconhecimento de face (Seção 6.2), segundo os grupos de avaliação definidos no Capítulo 5.

6.1 Resultados do Rastreamento de Face

Nessa seção são mostrados os resultados obtidos com o processo de localização e rastreamento da face. Na Figura 6.1 é possível verificar o resultado da localização e do rastreamento realizados com um indivíduo. Como pode ser observado, dada as peculiaridades do contexto onde o sistema é aplicado, o processo funciona de maneira eficaz (processando uma média de 4,5 quadros por segundo). Contudo, é importante destacar alguns pontos fracos do sistema, o que acarreta em problemas na execução do processo proposto. O primeiro está relacionado a variação brusca de luminosidade, o que provoca falhas no momento da subtração da imagem de fundo com a atual. Outra falha é provocada quando o sistema capta uma dimensão vertical, da imagem da face, excessivamente larga (por exemplo, quando uma mulher está com o cabelo solto), o que gera problemas de enquadramento na imagem da face, devido a taxa de proporcionalidade inserido no processo. Na Figura 6.2, pode-se observar a falha mencionada acima.

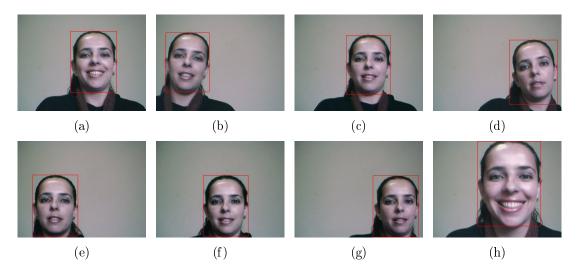


Figura 6.1: Seqüência de imagens do processo de localização e rastreamento de face implementado.

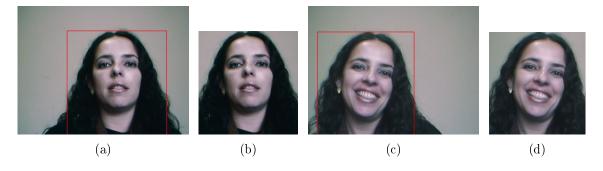


Figura 6.2: Seqüência de imagens (a) e (c) quando há falha no sistema de localização e rastreamento, e suas respectivas imagens de face capturadas (b) e (d).

6.2 Resultados do Reconhecimento de Face

Nessa seção é possível observar os resultados obtidos com o processo de reconhecimento de face, para cada um dos 16 grupos de avaliação descritos no capítulo 5. Para facilitar a observação dos resultados obtidos, esses serão apresentados separadamente por cada processo de classificação utilizado. Assim teremos: classificação por distância Euclidiana (Subseção 6.2.1); e classificação por SVM(Subseção 6.2.2).

Nos resultados obtidos utilizando o classificador de distância Euclidiana, poderão ser observados ainda dois processos de análise biométrico distintos: verificação e identificação. O processo de verificação consiste em validar se um indivíduo é realmente quem ele informa ser. Enquanto o de identificação, o classifica em um dos indivíduos já cadastrados no sistema. No processo de verificação é necessário verificar, além do número de acertos no

reconhecimento de um indivíduo, os falsos positivos. Estes são caracterizados pela falha do sistema que valida um indivíduo a partir de quem ele diz ser, sendo que na verdade ele não é. Devido ao fato da técnica de aprendizado de máquina SVM ser um classificador, são observados apenas as análises do processo de identificação dos indivíduos, sem que haja uma avaliação equivalente para o processo de verificação do classificador de distância Euclidiana.

A técnica de PCA aplicada em algumas avaliações realizadas, teve como objetivo a diminuição da quantidade de informações, que seriam passadas para o processo de extração de características. Originalmente, as imagens eram compostas de três componentes de cor (RGB). Após a aplicação da PCA as imagens passaram a ser representadas por apenas uma componente, que é uma composição das componentes originais. Dessa maneira, reduziu-se o tamanho dos dados em um terço.

Para realizar o processo de extração de características usando a técnica de transformada wavelet de Gabor, foi necessário especificar os parâmetros de escala e orientação. Assim, para todas as avaliações realizadas, foram utilizados os valores de escala igual a 5(E=5), e orientação igual a 4(O=4), resultando na extração de um vetor de 40(E*O*2=40) atributos de cada imagem processada. Isso gerou um conjunto de dados onde cada imagem era representada pelo vetor de atributos encontrado.

6.2.1 Classificação por Distância Euclidiana

São dois os tipos de avaliações realizados com o classificador por distância Euclidiana: verificação e identificação. Inicialmente serão apresentados os resultados obtidos com as avaliações de verificação. Nesse tipo de avaliação são apresentados dois conjuntos de resultados: 1) os acertos e os erros obtidos na avaliação de cada base, onde cada dado deveria ser validado pela sua respectiva classe; 2) os resultados obtidos com a avaliação dos falsos positivos de cada base, onde a validade de verificação para cada dado deveria ser negada para toda classe a que esse dado não pertence.

Na Tabela 6.1 são apresentados os resultados obtidos com o processo de verificação usando wavelet de Gabor, para todas as bases de imagens. Já a Tabela 6.2 contém a avaliação dos falsos positivos, e complementa os resultados desse processo.

Na Tabela 6.3 são apresentados os resultados obtidos com o processo de verificação usando PCA com *wavelet* de Gabor, para todas as bases de imagens. A Tabela 6.4 traz a avaliação dos falsos positivos, e complementa os resultados desse processo.

Tabela 6.1: Tabela com as bases de imagens e os resultados obtidos com o processo de verificação usando wavelet de Gabor e distância Euclidiana.

Base de Imagens	Número de Imagens	Acertos (%)	Erros (%)
FERET - Inteira	62	87,10%	12,90%
FERET - Centro	50	$72,\!00\%$	28,00%
Própria - Inteira	50	$100,\!00\%$	0,00%
Própria - Centro	45	97,78%	2,22%

Tabela 6.2: Tabela com as bases de imagens e os resultados obtidos com a avaliação dos falsos positivos, para o processo de verificação usando wavelet de Gabor e distância Euclidiana.

Base de Imagens	Número de Testes	Falsos Positivos	Falsos Positivos (%)
FERET - Inteira	3782	876	23,16%
FERET - Centro	2450	921	37,59%
Própria - Inteira	450	36	8,00%
Própria - Centro	360	26	7,22%

Após a apresentação dos resultados obtidos com as avaliações dos processos de verificação, são mostrados os resultados obtidos com as avaliações de identificação. Nesse tipo de avaliação cada dado deveria ser rotulado (classificado) corretamente com a sua respectiva classe, considerando o grau de similaridade, que no caso é a menor distância entre o dado e um exemplo do grupo controle.

A Tabela 6.5 apresenta os resultados obtidos com o processo de identificação usando wavelet de Gabor, para todas as bases de imagens.

Na Tabela 6.6 são apresentados os resultados obtidos com o processo de identificação usando PCA com *wavelet* de Gabor, para todas as bases de imagens.

Tabela 6.3: Tabela com as bases de imagens e os resultados obtidos com o processo de verificação usando PCA com wavelet de Gabor e distância Euclidiana.

Base de Imagens	Número de Imagens	Acertos (%)	Erros (%)
FERET - Inteira	62	$85,\!48\%$	14,52%
FERET - Centro	50	78,00%	22,00%
Própria - Inteira	50	$100,\!00\%$	0,00%
Própria - Centro	45	$100,\!00\%$	0,00%

Tabela 6.4: Tabela com as bases de imagens e os resultados obtidos com a avaliação dos falsos positivos, para o processo de verificação usando PCA com *wavelet* de Gabor e distância Euclidiana.

Base de Imagens	Número de Testes	Falsos Positivos	Falsos Positivos (%)
FERET - Inteira	3782	1128	29,83%
FERET - Centro	2450	718	29,31%
Própria - Inteira	450	15	3,33%
Própria - Centro	360	18	$5{,}00\%$

Tabela 6.5: Tabela com as bases de imagens e os resultados obtidos com o processo de identificação usando wavelet de Gabor e distância Euclidiana.

Base de Imagens	Número de Imagens	Acertos (%)	Erros (%)
FERET - Inteira	62	74,19%	$25{,}81\%$
FERET - Centro	50	44,00%	$56,\!00\%$
Própria - Inteira	50	$98,\!00\%$	$2,\!00\%$
Própria - Centro	45	100,00%	0,00%

6.2.2 Classificação por SVM

A classificação dos dados utilizando a técnica de aprendizado de máquina SVM, foi baseada no uso da biblioteca libsvm (Subseção 4.3.7). Foram definidos o C-SVC como tipo de SVM e o RBF como o núcleo. Já os parâmetros custo (c) e gama (g) foram obtidos com o auxílio da ferramenta grid.py. Para isso foi necessário aplicar os dados do grupo de treinamento nessa ferramenta, que ajustou os valores dos parâmetros conforme a eficácia dos treinamentos. Esse processo foi realizado para cada base de imagem utilizada na avaliação. Após encontrar os melhores valores para os parâmetros, foi iniciado o processo de treinamento de cada base. Isso resultou na geração de um modelo de classificação. Aplicando-se os dados do grupo de teste no modelo encontrado, foi possível

Tabela 6.6: Tabela com as bases de imagens e os resultados obtidos com o processo de identificação usando PCA com wavelet de Gabor e distância Euclidiana.

Base de Imagens	Número de Imagens	Acertos (%)	Erros (%)
FERET - Inteira	62	67,74%	$32,\!26\%$
FERET - Centro	50	$38,\!00\%$	$62,\!00\%$
Própria - Inteira	50	$100,\!00\%$	0,00%
Própria - Centro	45	97,78%	$2,\!22\%$

avaliar a qualidade do classificador gerado. Nesse processo de avaliação o dado deveria ser rotulado (classificado) corretamente com a sua respectiva classe.

A Tabela 6.7 mostra os resultados obtidos com o processo de identificação usando wavelet de Gabor, para cada uma das bases de imagens, bem como os respectivos valores dos parâmetros (c) e (g) utilizados para o treinamento.

Tabela 6.7: Tabela com as bases de imagens e os resultados obtidos com o processo de identificação usando wavelet de Gabor e SVM.

Base	Número			Parâmetros
de	de	$\mathbf{Acertos}\ (\%)$	Erros (%)	$(c) \mathbf{e} (g)$
Imagens	Imagens			
FERET - Inteira	62	$91,\!94\%$	8,06%	(c=8) e (g=0,125)
FERET - Centro	50	$68,\!00\%$	$32,\!00\%$	(c = 8192) e (g = 0,0005)
Própria - Inteira	50	$100,\!00\%$	$0,\!00\%$	(c=8) e (g=0,0078)
Própria - Centro	45	$100,\!00\%$	0,00%	(c=32) e (g=0,0078)

Na Tabela 6.8 são apresentados os resultados obtidos com o processo de identificação usando PCA com wavelet de Gabor para cada uma das bases de imagens, bem como os respectivos valores dos parâmetros (c) e (g) utilizados para o treinamento.

Tabela 6.8: Tabela com as bases de imagens e os resultados obtidos com o processo de identificação usando PCA com wavelet de Gabor e SVM.

Base de Imagens	Número de Imagens	Acertos (%)	Erros (%)	Parâmetros (c) e (g)
FERET - Inteira	62	82,26%	7,74%	(c = 32) e (g = 0, 125)
FERET - Centro	50	$60,\!00\%$	40,00%	(c=8) e (g=0,5)
Própria - Inteira	50	$100,\!00\%$	0,00%	(c=8) e (g=0,0078)
Própria - Centro	45	95,56%	4,44%	(c=8) e (g=0,0078)

6.3 Discussão dos Resultados

A Subseção 6.2.1 traz os resultados obtidos com a avaliação de verificação da classificação por distância Euclidiana. Os melhores resultados obtidos nessa avaliação foram: 87,10% de acertos com 23,16% de falsos positivos para a base FERET com a face completa usando wavelet de Gabor; e 100,00% de acertos com 3,33% de falsos positivos para a base

Própria com a face completa usando PCA e wavelet de Gabor. Conclui-se que os melhores resultados são obtidos com bases de imagens contendo a face completa de um indivíduo, e que, ainda, o uso da técnica de PCA no processo de extração de características não é necessário, pois, apesar de apresentar melhores resultados em três dos quatro casos de avaliação, a diferença dos valores não é tão significativa. Comparando as bases FERET e Própria pode-se perceber claramente que a segunda possibilita melhores resultados. Contudo, é importante salientar que devido as características intrínsecas do processo de avaliação, o fato de a base FERET possuir um número maior de classes distintas induz a resultados aparentemente menos satisfatórios, porém mais confiáveis.

Ainda na Subseção 6.2.1 tem-se os resultados obtidos com a avaliação de identificação usando classificação por distância Euclidiana. Nessa avaliação, os melhores resultados obtidos foram: 74,19% de acertos para a base FERET com a face completa usando wavelet de Gabor; e 100,00% de acertos para a base Própria, tanto com a parte central da face usando wavelet de Gabor, quanto com a face completa usando PCA e wavelet de Gabor. Considerando-se esses dados é possível afirmar que as bases de imagens com a face completa permitem obter resultados superiores. O uso da PCA nesse processo de avaliação gerou resultados equivalentes ou inferiores ao do uso da wavelet de Gabor somente. Isso induz a concluir que seu uso não seria adequado. Os melhores resultados obtidos com as bases Próprias em detrimento das bases FERET, nesse caso, se deve ao fato de que o mecanismo de classificação não é robusto o suficiente para se adequar aos exemplos, de uma mesma classe, que possuem grande variação. Como a base FERET possui imagens com grande variação entre seus exemplos, os resultados obtidos tendem a ser inferiores para esse processo de classificação.

Na Subseção 6.2.2 são ilustrados os resultados obtidos com a avaliação de identificação usando classificação por SVM. Os melhores resultados obtidos nesse processo foram: 91,94% de acertos para a base FERET com a face completa usando wavelet de Gabor; e 100,00% de acertos para a base Própria (face completa e parte central) usando wavelet de Gabor, e também para a base Própria com a face completa usando PCA e wavelet de Gabor. Comparando-se os resultados da subseção anterior, pode-se concluir que os melhores resultados são obtidos com bases de imagens com face completa. Além disso, nesse processo de avaliação fica claro que o processo de extração de características sem o uso de PCA permite obter resultados de classificação superiores.

Um ponto importante a ser destacado é a comparação de resultados entre o processo de classificação por distância Euclidiana e SVM. Os melhores resultados na primeira abordagem são: 74,19% para a base FERET e 100,00% para a base Própria. Enquanto

para a segunda são: 91,94% para a base FERET e 100,00% para a base Própria. Isso nos permite concluir que o processo de classificação usando SVM pode gerar resultados superiores aos de distância Euclidiana. Isso se deve a robustez e ao poder de generalização que o processo de classificação por aprendizado de máquina SVM possui, quando se trabalha com um número grande de classes.

Uma observação importante quanto ao uso das bases FERET e Própria deve ser feita. Em todos os casos de avaliação os resultados obtidos são superiores quando utilizada a base Própria. Contudo, como foi observado anteriormente, isso se deve principalmente a grande variabilidade nas imagens da base FERET, em detrimento a base Própria. Porém, é necessário destacar a importância do resultado obtido com a base Própria no processo de classificação usando SVM. Isso porque a robustez do SVM está ligado principalmente a sua capacidade de generalização. Assim, quanto maior o número de exemplos de uma determinada classe, melhor será a capacidade de generalização para um novo exemplo, ao contrário do processo de classificação por distância Euclidiana. Isso leva a acreditar que o resultado obtido usando SVM com a base Própria é altamente relevante. Partindo desse contexto, pode-se considerar que o resultado obtido usando SVM com a base FERET poderia se aproximar do resultado usando a base Própria, se a base FERET (três exemplos de treinamento) dispusesse de um número superior de exemplos, como a base Própria (dez exemplos de treinamento).

Considerando os resultados analisados acima, podemos destacar alguns pontos importantes que foram observados nesse trabalho:

- As imagens contendo a face completa de um indivíduo são mais indicadas para um sistema de reconhecimento, pois é possível obter um número maior de acertos;
- O uso da técnica PCA para a diminuição de informações antes da utilização da wavelet de Gabor é dispensável. Pois, apesar de alguns resultados obtidos serem pouco melhores, esses valores são praticamente equivalentes, e, em outros momentos, os resultados são inferiores:
- O processo de classificação por SVM é melhor que o de distância Euclidiana. A
 SVM apresenta resultados significativamente superiores, principalmente quando são
 utilizadas bases de imagens mais complexas, como a FERET. Isso se deve a sua
 robustez e capacidade de generalização.

CAPÍTULO

Conclusão

Este trabalho teve como objetivo o desenvolvimento de um sistema biométrico capaz de realizar o reconhecimento de uma pessoa a partir de sua face. Para isso, foram propostos o desenvolvimento de dois módulos: (1) um sistema de localização e rastreamento de face, a partir de uma câmera de vídeo, afim de adquirir a imagem da face com a qual seria trabalhada na próxima etapa; (2) um sistema de reconhecimento de face, o qual deverá ser capaz de identificar uma pessoa a partir da imagem de sua face.

Dados os resultados do sistema de localização e rastreamento implementado, é possível perceber que a primeira parte do objetivo deste trabalho foi alcançado. O uso desse sistema permitiu a localização eficaz da face de um indivíduo a partir do uso de uma câmera de vídeo. Além disso, o processo de rastreamento da face obteve êxito em todos os testes realizados. Isso permitiu que as imagens de face, dos indivíduos que participaram dos testes, fossem captadas com rapidez e garantia de qualidade em grande parte dos casos.

Os resultados, observados nos testes para o sistema de localização e rastreamento de faces, foram atingidos devido ao contexto no qual este sistema estava inserido. Isso significa que o sistema utiliza-se do fato de que no ambiente em que está sendo feita a filmagem há apenas uma pessoa. Além disso, a disposição da cabeça do indivíduo sempre estará na parte superior do vídeo. Por fim, leva-se em consideração o fator de proporcionalidade das dimensões da cabeça humana, onde, em geral, temos uma relação

aproximada de $\frac{4}{3}$ entre as dimensões analisadas na vertical e horizontal. Os erros observados no sistema se devem principalmente ao fator de enquadramento da face, quando está sendo realizado o rastreamento. Isso porque, em algumas situações, o sistema identifica de maneira equivocada a dimensão horizontal, como, por exemplo, quando uma pessoa está com cabelos longos e soltos, e o sistema identifica todo o segmento horizontal como sendo a face dessa pessoa. Isso acaba gerando uma falha de enquadramento, já que a relação de proporcionalidade entre as dimensões horizontal e vertical não pode ser mais a pré-determinada.

O desenvolvimento do sistema de reconhecimento de face, o segundo módulo proposto, e no qual foi concentrado o maior esforço, completa o objetivo deste trabalho, e foi alcançado com êxito satisfatório. Isso pode ser comprovado ao analisar-se os resultados, tanto no caso das imagens da base FERET, bem como nas imagens da base Própria, que foi compilada a partir dos resultados do sistema de localização e rastreamento.

Com os resultados obtidos, pode-se inferir alguns aspectos interessantes das técnicas utilizadas neste trabalho. Uma delas é a adoção do uso da técnica de PCA. Aqui, esta técnica foi utilizada para reduzir o número de componentes de cor da imagem original, com o objetivo de reduzir o custo computacional na próxima fase. Os resultados obtidos foram conclusivos quanto a falta de contribuição da PCA nesse caso específico. Isso porque os resultados, quando não eram piores, não apresentaram benefícios significativos, e a redução do custo computacional também não foi significativa. Assim, pode-se afirmar que a PCA não é indicada para a diminuição das componentes de cor, a menos que o sistema em questão tenha em seqüencia um processo altamente custoso computacionalmente, no qual a redução dos dados tenham um impacto maior, caso diferente do observado neste trabalho.

Outro aspecto importante que pôde ser analisado com este trabalho foi o poder de classificação da técnica SVM. Os resultados obtidos pelo classificador baseado na técnica de aprendizado de máquina SVM são bons, principalmente se forem comparados ao classificador baseado em distância Euclidiana. Com as bases de imagens FERET o classificador SVM atingiu um êxito de 91,94%, enquanto que para a base de imagens Própria esse índice chegou a 100,00%. Isso demonstra que a SVM possui grande capacidade de generalização e robustez, o que proporciona, principalmente em sistemas de reconhecimento de face, um ótimo processo de classificação de dados. Acredita-se ainda que os resultados com a base de dados FERET poderiam ser ainda melhores, caso houvesse um número maior de exemplares de treinamento para cada uma das classes de imagens testadas. Isso porque, para classificadores baseados em aprendizado de máquina como

a SVM, é necessário um grande número de exemplares de treinamento (principalmente no caso de classificadores multiclasses com grande número de classes distintas) para que o sistema possa ser realmente capaz de generalizar os dados. Contudo, foi possível utilizar apenas três exemplares para treinamento da base FERET, enquanto que com a base Própria foram utilizados dez exemplares. Como pôde ser observado com está última base de imagens, a SVM conseguiu atingir um bom grau de generalização, o que possibilitou a obtenção de uma ótima taxa de acertos na classificação.

Sobre as bases de imagens utilizadas neste trabalho podem ser feitas algumas observações. A primeira delas é a superioridade nos resultados obtidos pelas imagens de face completa (imagem com o rosto inteiro do indivíduo), em detrimento as imagens de face parcial (imagem contendo apenas a porção central do rosto do indivíduo). Essa vantagem no número de acertos pôde ser observada em ambas as bases de imagens, FERET e Própria, o que indica a real superioridade da face completa. Isso permite concluir que a melhor maneira de analisar biometricamente uma face é utilizando toda a face de uma pessoa, e não somente a porção central, onde estão concentrados a boca, os olhos e o nariz. A segunda observação a ser destacada é a comparação entre as bases FERET e Própria. Os resultados expressam uma significativa superioridade no desempenho da base Própria, contudo, é importante destacar os pontos que permitiram essa disparidade. Um desses é a grande variabilidade encontrada entre as imagens de um mesmo indivíduo na base FERET, o que dificulta a classificação. Além disso, outra questão é a do, já mencionado, número de imagens de treinamento, bem inferior na base FERET, e que ajuda a determinar a qualidade do classificador obtido.

Todas as análises destacadas até o momento estão concentradas apenas na discussão a cerca dos resultados obtidos neste trabalho. Contudo, é de extrema importância realizar uma comparação objetiva entre os resultados obtidos neste trabalho, e os resultados que podem ser encontrados na literatura científica. Pensando nesse propósito é que foi utilizado a base de imagens FERET, pois grande parte dos trabalhos desenvolvidos na área de reconhecimento de face mencionam resultados obtidos a partir do uso de um subconjunto dessa base. Assim, a seguir, serão apresentados (Tabela 7.1) e comparados os resultados obtidos em trabalhos que utilizam outras abordagens para extração e classificação dos dados de imagens de face (base FERET), e o resultado obtido neste trabalho.

O resultado obtido por Liu et al. [2004] usando KFDA (Kernel Fisher Discriminant Analysis) obteve 92,17% de acertos. Nesse trabalho foi destacado os resultados obtidos com o uso do FLDA (Fisher Linear Discriminant Analysis) com 89,69% de acertos,

Tabela 7.1: Tabela com os resultados obtidos usando-se outras abordagens de extração e classificação de dados de imagens da base FERET, e o resultado obtido neste trabalho.

Abordagem Utilizada	Classes	Exemplares	Acertos (%)
wavelet de Gabor e SVM	62	4	91,94%
KFDA	70	6	92,17%
FLDA	70	6	89,69%
KPCA	70	6	83,51%
EICA	369	3/4	95,40%
Eigenfaces	369	3/4	92,40%
Fisherfaces	369	3/4	92,70%

e com o uso do KPCA (Kernel PCA) com 83,51% de acertos. Já Liu[2004] trabalhando com EICA (Enhanced Independent Component Analysis) obteve em seu trabalho um resultado de 95,40% de acertos, e destacou os resultados obtidos com eigenfaces (92,40% de acertos) e fisherfaces (92,70% de acertos). O melhor resultado obtido neste trabalho, usando a base de imagens FERET, usando wavelet de Gabor e o classificador SVM, foi de 91,94% de acertos.

Analisando os resultados obtidos utilizando outras abordagens, é possível observar que o método aplicado neste trabalho obteve resultados satisfatórios. Isto porque, apesar de apresentar uma menor eficiência se comparado ao EICA, com as outras abordagens o resultado obtido acaba sendo similar ou superior. Ainda, vale lembrar que dado um número maior de imagens de treinamento acredita-se que o percentual de acertos aumente.

Referências Bibliográficas

- Argoud, F. I. M. (2000). Contribuição à automação da detecção e análise de eventos epileptiformes em eletroencefalogramas. PhD thesis, Engenharia Biomédica, UFSC.
- Arivazhagan, S. and Ganesan, L. (2003). Texture classification using wavelet transform. In *Pattern Recognition Letters*, volume 24, pages 1513 1521. Elsevier Science Inc., New York, NY, USA.
- Bellili, A., Gilloux, M., and Gallinari, P. (2003). An mlp-svm combination architecture for offline handwritten digit recognition. *International Journal on Document Analysis and Recognition*, 5(4):244 252.
- Braga, A., Carvalho, A. C. P. L. F., and Ludemir, T. B. (2000). Redes Neurais Artificiais: Teoria e Aplicações. Editora LTC.
- Campbell, C. (2000). An introduction to kernel methods. In Howlett, R. J. and Jain, L. C., editors, Radial Basis Function Networks: Design and Applications, pages 155–192. Springer-Verlag.
- Campbell Jr., J. P. (1997). Speaker recognition: A tutorial. *Proceeding of IEEE*, 85(9):1437–1462.
- Castañón, C. A. B. (2003). Recuperação de imagens por conteúdo através de análise multiresolução por waveletstured wavelet transform. Master's thesis, Instituto de Ciências Matemáticas e de Computação ICMC USP.
- Chang, C.-C. and Lin, C.-J. (2006). LIBSVM: a library for support vector machines. Disponível em http://www.csie.ntu.edu.tw/cjlin/libsvm. Último acesso em: 06/2006.

- Chang, K., Bowyer, K., Sarkar, S., and Victor, B. (2003). Comparison and combination of ear and face images in appearance-based biometrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1160–1165.
- Chang, T. and Kuo, C. J. (1993). Texture analysis and classification with tree-structured wavelet transform. *IEEE Transaction on Image Processing*, 2(4):429–441.
- Dash, M., Liu, H., and Yao, J. (1997). Dimensionality reduction of unsupervised data. In *Proceedings of the Ninth IEEE International Conference on Tools with Artificial Intelligence*, 1997., pages 532 539.
- Daubechies, I. (1990). The wavelet transform, time frequency localization, and signal analysis. *IEEE Transactions on Information Theory*, 36(5):961 1005.
- Daubechies, I. (1992). Ten lectures on wavelets. In CBMS-NSF Regional Conference Series in Applied Mathematics, Rutgers University and AT&T Bell Laboratories, volume 61.
- Daugman, J. G. (1993). High confidence visual recognition of persons by a test of statistical independence. *IEEE Transaction PAMI*, 15(11).
- De la Torre, F. and Black, M. J. (2001). Robust principal component analysis for computer vision. In *Eighth International Conference on Computer Vision (ICCV'01)*, volume 1, page 362. IEEE Computer Society.
- Dummais, S. (1998). Using svm for text categorization. *IEEE Intelligence Systems*, pages 21-23.
- Everitt, R. A. J. and McOwan, P. W. (2003). Java-based internet biometric authentication system. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1166 1172.
- Feris, R., Cesar Jr., R., and Kruger, V. (2001). Efficient real-time face tracking in wavelet subspace. In *IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, 2001, pages 113–118.
- Fierrez-Aguilar, J., Ortega-Garcia, J., Garcia-Romero, D., and Gonzalez-Rodriguez, J. (2003). A comparative evaluation of fusion strategies for multimodal biometric verification. In Audio- and Video-Based Biometric Person Authentication, volume 2688 of Lecture Notes in Computer Science, pages 830 837. Springer Berlin/Heidelberg.

- Fierrez-Aguilar, J., Ortega-Garcia, J., and Gonzalez-Rodriguez, J. (2005). Target dependent score normalization techniques and their application to signature verification. *IEEE Transactions on Systems, Man and Cybernetics, Part C*, 35(3):418 425.
- Frischholz, R. and Dieckmann, U. (2000a). Biold: a multimodal biometric identification system. In *IEEE Computer*, volume 33, pages 64–68.
- Frischholz, R. W. and Dieckmann, U. (2000b). Biold: a multimodal biometric identification system. *Computer*, 33(2):64 68.
- Gabor, D. (1946). Theory of communication. Journal of IEE (London), 93:429 457.
- Gao, Y. and Leung, M. (2002). Face recognition using line edge map. *IEEE Transactions* on Pattern Analysis and Machine Intelligence, 24(6):764–779.
- Gonzales, R. C. and Woods, R. E. (1992). *Digital Image Processing*. Addison-Wesley Pub Co.
- Gutta, S., Huang, J., Jonathon, P., and Wechsler, H. (2000). Mixture of experts for classification of gender, ethnic origin, and pose of human faces. *IEEE Transactions on Neural Networks*, 11(4):948–960.
- Haykin, S. (1999). Neural Networks A Compressive Foundation. Prentice-Hall, New Jersey, second edition.
- Hearst, M. A., Schölkopf, B., Dumais, S., Osuma, E., and Platt, J. (1998). Trends and controversies support vector machines. *IEEE Intelligent Systems*, 13(4):18 28.
- Herbrich, R. (2001). Learning Kernel Classifiers: Theory and Algorithms. MIT Press.
- Hill, R. B. (1978). Apparatus and method for identifying individuals through their retinal vasculature patterns. US Patent No. 410937.
- Hong, L. (1998). Automatic Personal Identification Using Fingerprint. PhD thesis, Department of Computer Science, Michigan State University.
- C.-W., C.-C., C.-J.. L. (2006).APrac-Hsu, Chang, and Classification.Disponível GuidetoSupport Vectoremhttp://www.csie.ntu.edu.tw/cjlin/libsvm/papers/guide/guide.pdf. Último acesso em: 06/2006.

- Hsu, C.-W. and Lin, C.-J. (2002). A comparison of methods for multi-class support vector machines. *IEEE Transactions on Neural Networks*, 13(2):415 425.
- Jain, A. K., Ross, A., and Prabhakar, S. (2003). An introduction to biometric recognition.
 IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics.
- Janbandhu, P. K. and Siyal, M. Y. (2001). Novel biometric digital signatures for internet-based applications. In *Information Management & Computer Security*, volume 9, pages 205 212. MCB UP Ltd.
- Jiang, C. and Su, G. (2005). How face pose influence the performance of sym-based face and fingerprint authentication system. In *Advances in Intelligent Computing*, volume 3644 of *Lecture Notes in Computer Science*, pages 310 319. Springer Berlin / Heidelberg.
- Johnson, R. A. anda Wichern, D. (1982). Applied multivariate statistical analysis. Prentice-Hall, New Jersey, second edition.
- Jolliffe, I. T. (2002). Principal Component Analysis. Springer Series in Statistics. Springer.
- Knerr, S., Personnaz, L., and Dreyfus, G. (1990). Single-layer learning revisited: a stepwise procedure for building and training a neural network. pages 41 – 50. Springer-Verlag.
- Krebel, U. (1999). Pairwise classification and support vector machines. In Schölkopf, B., Burges, C. J. C., and Smola, A. J., editors, *Advances in Kernel Methods Support Vector Learning*, pages 185 208. MIT Press.
- Krichen, E., Mellakh, M. A., Garcia-Salicetti, S., and Dorizzi, B. (2004). Iris identification using wavelet packets. In *Proceedings of the 17th International Conference on Pattern Recognition*, 2004. ICPR 2004, volume 4, pages 335 338.
- Kumar, A., Wong, D. C. M., Shen, H. C., and Jain, A. K. (2003). Personal verification using palmprint and hand geometry biometric. In Audio- and Video-Based Biometric Person Authentication, volume 2688 of Lecture Notes in Computer Science, pages 668 678. Springer Berlin/Heidelberg.
- Kumar, P. and Foufoula-Georgiou, E. (1994). Wavelet Analysis in Geophysics: An Introduction, pages 1 43. Academic Press.

- Kuno, Y., Shimada, N., and Shirai, Y. (2003). Look where you're going [robotic wheelchair]. *IEEE Robotics & Automation Magazine*, 10(1):26–34.
- Lanitis, A., Draganova, C., and Christodoulou, C. (2004). Comparing different classifiers for automatic age estimation. *IEEE Transactions on Systems, Man and Cybernetics*, Part B, 34(1):621–628.
- LeCun, Y., Jackel, L., Bottou, L., Brunot, A., Cortes, C., Denker, J., Drucker, H., Guyon, I., Müller, U., Sackinger, E., Simard, P., and Vapnik, V. (1995). Comparison of learning algorithms for handwritten digit recognition. In Fogelman, F. and Gallinari, P., editors, Proceedings of the International Conference on Artificial Neural Networks, pages 53 60.
- Lee, T. S. (1996). Image representation using 2d gabor wavelets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(10):959 971.
- Li, B. and Chellappa, R. (2000). Simultaneous tracking and verification via sequential posterior estimation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2000, volume 2, pages 110–117.
- Li, B. and Chellappa, R. (2002). A generic approach to simultaneous tracking and verification in video. *IEEE Transactions on Image Processing*, 11(5):530–544.
- Libralao, G., Almeida, O. C. P., and Carvalho, A. C. P. L. F. (2005). Classification of ophthalmologic images using an ensemble of classifiers. In *Lecture Notes in Computer Science*, volume 3533, pages 380 389.
- Liu, C. (2004). Enhanced independent component analysis and its application to content based face image retrieval. Transactions on Systems, Man, and Cybernetics Part B: Cybernetics : Accepted for future publication, PP(99):1.
- Liu, C. and Wechsler, H. (2003). Independent component analysis of gabor features for face recognition. *IEEE Transactions on Neural Networks*, 14(4):919–928.
- Liu, Q., Lu, H., and Ma, S. (2004). Improving kernel fisher discriminant analysis for face recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):42 49.
- Liu, S. and Silverman, M. (2001). A practical guide to biometric security technology. In *IEEE IT Professional*, pages 27–32.

- Lopez-Rubio, E., Ortiz-de-Lazcano-Lobato, J. M., Munoz-Perez, J., and Gomez-Ruiz, J. A. (2004). Principal components analysis competitive learning. In *Neural Computation*, volume 16 of 2459 2481. MIT Press Journals.
- Lorena, A. C. (2006). *Investigação de estratégias para a geração de máquinas de vetores suporte multiclasses*. PhD thesis, Instituto de Ciências Matemáticas e de Computação ICMC USP.
- Lorena, A. C. and Carvalho, A. C. P. L. F. (2003). Introdução às máquinas de vetores suporte (Support Vector Machine). Relatório Técnico 192, Instituto de Ciências Matemáticas e de Computação ICMC USP, São Carlos SP.
- Ma, L., Tan, T., Wang, Y., and Zhang, D. (2003). Personal identification based on iris texture analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1519 1533.
- Mallat, S. G. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transaction on Pattern analysis and Machine Intelligence*, 11(7):674 693.
- Manjunath, B. and Ma, W. (1996). Texture features for browsin and retrieval of image data. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 18:837–842.
- Manly, B. F. J. (1986). Multivariate statistical methods. Chapman and Hall, London.
- Marino, C., Penedo, M. G., Penas, M., Carreira, M. J., and Gonzalez, F. (2006). Personal authentication using digital retinal images. *Pattern Analysis & Applications*, 9(1):21 33.
- Matsumoto, T., Matsumoto, H., Yamada, K., and Hoshino, S. (2002). Impact of artificial gummy fingers on fingerprint systems. In *Proceedings of SPIE*, Optical Security and Counterfeit Deterrence Techniques IV, volume 4677, pages 275–289. Int'l Soc. for Optical Engineering.
- Mayoraz, E. and Moreira, M. (1996). On the decomposition of polychotomies into dichotomies. Research Report 96-08, IDIAP, Dalle Molle Institute for Perceptive Artificial Intelligence, Martigny, Valais, Switzerland.
- Mendonça, M. A. (2005). Classificação de gasolinas comerciais através de métodos estatísticos multivariáveis. Master's thesis, Escola Politécnica Poli USP.

- Mercer, J. (1909). Functions of positive and negative type and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society*, 209:415 446.
- Meyer, Y., editor (1992). Wavelets and applications, volume 20 of RMA: Research Notes in Applied Mathematics, Paris. Masson.
- Mitchell, M. (1997). Matchine Learning. McGraw Hill.
- Monard, M. C. and Baranauskas, J. A. (2003). Conceitos de aprendizado de máquina. In Rezende, S. O., editor, *Sistemas Inteligentes*, pages 89 114. Editora Manole.
- Monrose, F., Reiter, M. K., and Wetzel, S. (2002). Password hardening based on keystroke dynamics. In *International Journal of Information Security*, volume 1, pages 69 83. Springer Berlin/Heidelberg.
- Morlet, J. (1981). Sampling theory and wave propagation. In *Proc. 51st Annu. Meet. Soc. Explor. Geophy.*, Los Angeles.
- Müller, K. R., Mika, S., Rätsch, G., Tsuda, K., and Schölkopf, B. (2001). An introduction to kernel-based learning algorithms. *IEEE Transactions on Neural Networks*, 12(2):181 201.
- Munich, M. E. and Perona, P. (2003). Visual identification by signature tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(2):200 217.
- Negin, M., Chmielewski, T. A., Salganicoff, M., Camus, T. A., Seelen, U. M. C., Venetianer, P. L., and Zhang, G. G. (2000). An iris biometric system for public and personal use. In *IEEE Computer*, volume 33, pages 70–75.
- Oliveira, P. R. (1997). Redes neurais artificiais para extração de características. Master's thesis, Instituto de Ciências Matemáticas e de Computação ICMC USP.
- Passerini, A. (2004). Kernel Methods, multiclass classification and applications to computational molecular biology. PhD thesis, Università Degli Studi di Firenze.
- Pentland, A. S. and Choudhury, T. (2000). Face recognition for smart environments. In *IEEE Computer*, volume 33, pages 50–55.
- Phillips, P., Moon, H., Rizvi, S., and Rauss, P. (2000). The feret evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104.

- Pontil, M. and Verri, A. (1998). Support vector machines for 3-d object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(6):637 646.
- Protázio, J. M. B. (2001). Análise wavelet aplicada a sinais geofísicos. Master's thesis, Faculdade de Engenharia Mecânica, Universidade Estadual de Campinas.
- Robinson, J. A., Liang, V. W., Chambers, J. A. M., and MacKenzie, C. L. (1998). Computer user verification using login string keystroke dynamics. *IEEE Transactions on Systems, Man and Cybernetics, Part A*, 28(2):236 241.
- Rubner, Y. and Tomasi, C. (2000). Perceptual Metrics for Image Database Navigation, volume 594 of Series in Engineering and Computer Science. Kluwer Academic Publisher.
- Ruiz-del-Solar, J., Shats, A., and Verschae, R. (2003). Real-time tracking of multiple persons. In 12th International Conference on Image Analysis and Processing, 2003, pages 109–114.
- Sekmen, A., Wilkes, M., and Kawamura, K. (2002). An application of passive human-robot interaction: human tracking based on attention distraction. *IEEE Transactions on Systems, Man and Cybernetics, Part A*, 32(2):248–259.
- Shen, L. and Bai, L. (2004). Gabor feature based face recognition using kernel methods. In Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004, pages 170 176.
- Smola, A. J., Barlett, P., Schölkopf, B., and Schuurmans, D. (1999). Introduction to large margin classifiers. In Smola, A. J., Barlett, P., Schölkopf, B., and Schuurmans, D., editors, Advances in Large Margin Classifiers, pages 1–28. MIT Press.
- Smola, A. J. and Schölkpof, B. (2002). *Learning with Kernels*. The MIT Press, Cambridge, MA.
- Sobottka, K. and Pitas, I. (1996). Face localization and facial feature extraction based on shape and color information. In *International Conference on Image Processing*, volume 3, pages 483–486.
- Souto, M. C. P., Lorena, A. C., Delbem, A. C. B., and Carvalho, A. C. P. L. F. (2003). Técnicas de aprendizado de máquina para problemas de biologia molecular. In *Minicursos de Inteligência Artificial*, *Jornada de Atualização Científica em Inteligência Artificial*, XXIII Congresso da Sociedade Brasileira de Computação, pages 103 152.

- Sujan, V. A. and Mulqueen, M. P. (2002). Fingerprint identification using space invariant transforms. *Pattern Recognition Letters*, 23(5):609 619.
- Sung, E. (2002). Stereo head/face tracking and pose estimation. In 7th International Conference on Control, Automation, Robotics and Vision, 2002. ICARCV 2002, volume 3, pages 1609–1614.
- Tian, Y.-I., Kanade, T., and Cohn, J. (2001). Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2):97–115.
- Torrence, C. and Compo, G. P. (1998). A practical guide to wavelet analysis. Bulletin of the American Meteorololical Society, 79(1):61 78.
- Vapnik, V. N. (1995). The nature of statistical learning theory. Springer-Verlag, New York.
- Vapnik, V. N. (1998). Statistical Learning Theory. John Wiley and Sons.
- Vapnik, V. N. and Chervonenkis, A. Y. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications*, 16(2):283 305.
- Vatsa, M., Singh, R., and Noore, A. (2005). Improving biometric recognition accuracy and robustness using dwt and svm watermarking. *IEICE Electron. Express*, 2(12):362 367.
- Venkataramani, K. and Kumar, B. V. K. V. (2003). Fingerprint verification using correlation filters. In Audio- and Video-Based Biometric Person Authentication, volume 2688 of Lecture Notes in Computer Science, pages 886 894. Springer Berlin/Heidelberg.
- Verma, R., Schmid, C., and Mikolajczyk, K. (2003). Face detection and tracking in a video by propagating detection probabilities. *IEEE Transactions on Pattern Analysis* and Machine Intelligence, 25(10):1215–1228.
- Vukadinovic, D. and Pantic, M. (2005). Fully automatic facial feature point detection using gabor feature based boosted classifiers. In *IEEE International Conference on Systems*, Man and Cybernetics, 2005, volume 2, pages 1692 – 1698.

- Wan, V. and Campbell, W. M. (2000). Support vector machines for speaker verification and identification. In *Proceedings of the IEEE Signal Processing Society Workshop on Neural Networks for Signal Processing X*, 2000, volume 2, pages 775 784.
- Wang, Y. and Yuan, B. (2002). Fast method for face location and tracking by distributed behaviour-based agents. *IEE Proceedings Vision, Image and Signal Processing*, 149(3):173–178.
- Yan, S., He, X., Hu, Y., Zhang, H., Li, M., and Cheng, Q. (2004). Bayesian shape localization for face recognition using global and local textures. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):102–113.
- Yang, F. and Paindavoine, M. (2003). Implementation of an rbf neural network on embedded systems: real-time face tracking and identity verification. *IEEE Transactions on Neural Networks*, 14(5):1162–1175.
- Yang, M.-H., Kriegman, D., and Ahuja, N. (2002). Detecting faces in images: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1):34–58.
- Yeung, K. Y. and Ruzzo, W. L. (2001). Principal component analysis for clustering gene expression data. *Bioinformatics*, 17(9):763 774.
- Yu, J. and Bhanu, B. (2006). Evolutionary feature synthesis for facial expression recognition. In *Pattern Recognition Letters*, volume 27, pages 1289 1298. Elsevier Science Inc., New York, NY, USA.
- Yuan, Q., Doermann, D., and DeMenthon, D. (2001). Hybrid independent component analysis and support vector machinelearning scheme for face detection. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2001. Proceedings. (ICASSP '01)., volume 3, pages 1481 – 1484.
- Zhang, B., Gao, W., Shan, S., and Wang, W. (2003a). Constraint shape model using edge constraint and gabor wavelet based search. In *Audio- and Video-Based Biometric Person Authentication*, volume 2688 of *Lecture Notes in Computer Science*, pages 52 61. Springer Berlin/Heidelberg.
- Zhang, D., Kong, W., You, J., and Wong, M. (2003b). Online palmprint identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1041 1050.