Técnicas de Seleção de Atributos e de Classificação para Reconhecimento de Faces

Projeto de Pesquisa - Exame de Qualificação

aluno: Teófilo Emídio de Campos

Bolsista Fapesp (processo no. 99/01488-8)

orientador: Roberto Marcondes Cesar Junior

{teo, cesar}@ime.usp.br

Departamento de Ciência da Computação - IME - USP

Rua do Matão, 1010, 05508-900 São Paulo - SP - Brasil

18 de setembro de 2000

Resumo

Reconhecimento de faces é uma área de pesquisa muito promissora que ainda possui muitos problemas não resolvidos completamente, sendo um exemplo o problema de reconhecimento de pessoas em cenas naturais usando seqüências de vídeo. Dentre os principais obstáculos a serem atravessados, está o de se obter uma representação compacta de faces que possibilite distinguir os indivíduos rapidamente. Este trabalho aborda parte desse problema utilizando, principalmente, seleção automática de características. Alguns resultados já obtidos mostram que essa abordagem é promissora. Para possibilitar a obtenção desses resultados e a elaboração de um plano de trabalho, foi realizado um estudo das técnicas de reconhecimento estatístico de padrões, principalmente em se tratando de redução de dimensionalidade dos dados, além de uma revisão de métodos de reconhecimento de faces. Pretende-se aprofundar um pouco mais o estudo de técnicas de seleção de características e implementar o método inovador de reconhecimento a partir de seqüências de vídeo proposto no plano de trabalho, o que deverá levar à conclusão do projeto de mestrado do aluno.

Sumário

1 Introdução

Métodos de identificação de pessoas sempre foram muito importantes para toda a sociedade. No mundo moderno, as pessoas sempre precisam carregar consigo algum documento para identificá-las, como se a individualidade só existisse para pessoas que portam tais documentos. Partindo do fato de que não existem pessoas completamente idênticas, a necessidade da utilização de tais documentos se extingue quando se dispõe de métodos capazes de diferenciar cada indivíduo sem confundi-lo com seus semelhantes. Provavelmente esse é o principal objetivo da pesquisa em Biométrica. Um sistema biométrico é um sistema de reconhecimento de padrões que estabelece a autenticidade de uma característica fisiológica ou comportamental possuída por um usuário [1].

Dentre as técnicas de reconhecimento biométrico de pessoas que são utilizadas atualmente, as mais precisas são as baseadas em imagem de íris [1]. A confiabilidade de sistemas de reconhecimento de íris é tão grande que já existem bancos adotando-os para identificar seus usuários. A desvantagem dessa abordagem está em seu caráter invasivo, pois é necessário que o usuário permaneça parado em uma posição definida e com os olhos abertos enquanto o *scanner* de íris ou uma câmera captura a imagem. Nesse ponto está a vantagem de um sistema de reconhecimento baseado em imagens de faces. A pesquisa em reconhecimento de faces vem se desenvolvendo no sentido da criação de sistemas que identifiquem pessoas mesmo que essas não percebam que estão sendo observadas, de forma que, no futuro, possivelmente uma criança desaparecida poderá ser localizada através de imagens de câmeras espalhadas em pontos estratégicos de uma cidade, como estações de metrô, cruzamentos de avenidas ou até mesmo satélites.

Além dessa aplicação, várias outras motivantes para a pesquisa nessa área foram analisadas por Chellappa et al. [2], como:

- identificação pessoal para banco, passaporte, fichas criminais;
- sistemas de segurança e de controle de acesso;
- monitoramento de multidões em estações, *shopping centers* etc.;
- criação de retrato falado;

- busca em fichas criminais;
- envelhecimento computadorizado para auxiliar na busca por desaparecidos;
- interfaces perceptuais homem-máquina com reconhecimento de expressões faciais.

Devido a sua importância prática e aos interesses dos cientistas cognitivos, a pesquisa em reconhecimento de faces é tão antiga quanto a própria visão computacional [3]. Em [2] há uma análise de trinta anos de pesquisa em reconhecimento de faces humano e por máquina que cita 221 trabalhos. Outra evidência do crescimento dessa área de pesquisa é a existência de conferências específicas de reconhecimento de face e gestos [4, 5, 6, 7], bem como a existência de revistas com seções temáticas nessa área [8].

O reconhecimento óptico automático de faces é uma área da visão computacional que consiste no estudo de métodos para extração de informações a partir de imagens digitais. Em reconhecimento de face, o objetivo é identificar pessoas que aparecem em imagens. Devido à idade da área de pesquisa e ao número de pesquisadores trabalhando nesse assunto, as técnicas empregadas para efetuar a classificação das informações extraídas das imagens de faces são as mais variadas do reconhecimento de padrões. As seções 3.1 e 6 possuem descrições a respeito de métodos de reconhecimento de padrões e classificação.

Para possibilitar um reconhecimento preciso, primeiramente é necessário segmentar as faces para que somente essas sejam tratadas, não considerando os objetos que estiverem atrás do sujeito a ser reconhecido, os quais podem influenciar na tomada de decisão do classificador. Para isso é utilizado um método de detecção de faces.

No caso de seqüências de vídeo, a segmentação deve ser feita em todas as imagens da seqüência. O problema é que geralmente imagens de cenas dinâmicas apresentam menor qualidade devido a borramentos proporcionados pelo próprio movimento dos objetos, além disso, essas ocupam muito espaço na memória do computador. Para evitar o problema de complexidade em espaço, em geral são adotadas imagens com menor resolução em seqüências de vídeo, o que compromete mais a qualidade das imagens [2]. Para possibilitar o reconhecimento em tempo real, é necessário que os processos sejam muito rápidos. Por isso, em geral é adotado um método de segmentação apenas no primeiro quadro em que a pessoa aparece na seqüência, sendo subseqüentelemte aplicado um procedimento de perseguição (ou rastreamento - *tracking*), que é mais rápido e considera informações obtidas no quadro anterior para evitar a realização de buscas por toda a imagem. Como exemplos de métodos rápidos de detecção e perseguição de faces em seqüências de imagens, pode-se citar: [9, 10, 11].

Além disso, o processo de reconhecimento também deve ser rápido, o que causa a necessidade da utilização de bons algoritmos de redução da dimensionalidade dos dados. Detalhes sobre redução de dimensionalidade e métodos de classificação eficientes computacionalmente podem ser encontrados neste texto (seções 4 e 5).

Na próxima seção estão tratados os objetivos e o escopo deste trabalho de mestrado. A seção 3 contém um panorama da área em que esse trabalho se insere, destacando métodos de reconhecimento de padrões e algumas abordagens propostas para reconhecimento de faces. Nas seções 4, 5 e 6 há uma revisão um pouco mais detalhada de métodos de extração de características, redução de dimensionalidade e de classificação. Os resultados preliminares desse projeto obtidos até o presente momento estão descritos na seção 7. Por final, na seção 8 há uma descrição do plano de trabalho proposto para ser implementado até março de 2001, que é a data prevista para a finalização do desenvolvimento do mestrado e início da preparação para a defesa da dissertação.

2 Objetivos

Este trabalho se restringe ao estudo de técnicas de redução de dimensionalidade utilizando principalmente seleção de características visando a criação de um método de reconhecimento (classificação) de faces utilizando seqüências de vídeo.

Técnicas de detecção e rastreamento de faces em seqüências de vídeo não fazem parte do escopo deste projeto. Serão utilizadas seqüências temporais de imagens com a face já segmentada e normalizada com relação à escala e orientação. A obtenção dessas imagens de forma automática é responsabilidade do aluno Rogério Schmidt Feris (mestrado - DCC/IME/USP).

Pretende-se realizar testes utilizando seqüências de imagens capturadas em ambientes sem iluminação controlada e com pessoas se movimentando naturalmente. Pequenas bases de vídeo podem ser criadas a partir de equipamento disponível no Laboratório de Processamento de Imagens (IME/USP), mas o maior problema desse tipo de teste é a falta de disponibilidade de bases de seqüências de imagens com muitas pessoas diferentes. Por isso poderão ser feitos testes utilizando bases de imagens estáticas que possuem muitas pessoas diferentes. Também pretende-se estudar formas efetivas de avaliação de desempenho do método de reconhecimento a ser implementado, para possibilitar ajustes visando aperfeiçoá-lo.

Já foram dados alguns passos no sentido de investigação de técnicas de extração e seleção de características visando reduzir a dimensionalidade dos dados e melhorar o desempenho de classificadores (vide seção 7). Para alcançar os objetivos finais foi feito um plano de trabalho detalhado que se encontra na seção 8.

3 Revisão Bibliográfica

Visão computacional é uma área de pesquisa em que se utilizam elementos de processamento de imagens e de reconhecimento de padrões para extrair e interpretar informações contidas em imagens. O estudo realizado neste trabalho focaliza algumas das técnicas de visão computacional que podem ser aplicadas ao reconhecimento de faces.

Inicialmente serão descritas algumas técnicas de reconhecimento de padrões (seção 3.1) que podem ser aplicadas ao reconhecimento de faces. Posteriormente (seção 3.2) será feita uma revisão de alguns dos métodos de reconhecimento de faces.

3.1 Reconhecimento de Padrões

Reconhecimento de padrões é a área de pesquisa que tem por objetivo a classificação de objetos (padrões) em um número de categorias ou classes [12]. No caso de reconhecimento de faces, as imagens de façes são os objetos e as classes são seus nomes ou identificações. Há também o problema de categorização de faces, que trata da classificação das pessoas em categorias discriminando, por exemplo, gênero, faixa etária e etnia. Nesse caso, as classes são a categoria em que a pessoa pertence. Como este trabalho é centrado no problema de reconhecimento (identificação), não serão detalhadas técnicas de categorização.

Nos últimos 50 anos de pesquisa neste área, o problema geral de reconhecimento de padrões complexos com orientação, localização e escala arbitrárias permanece sem solução. A aplicação ao reconhecimento de faces requer técnicas robustas e eficientes. Além de reconhecimento de faces, as técnicas de reconhecimento de padrões podem ser aplicadas em vários outros domínios, dentre os quais, Jain et. al. [13] destaca:

- bio-informática, análise de seqüências de proteínas ou DNA;
- mineração de dados (*data mining*), procurando por padrões significativos em espaços multi-dimensionais;
- classificação de documentos de Internet;

- análise de imagens de documentos para reconhecimento de caracteres (*Optical Character Rocognition - OCR*);
- inspeção visual para automação industrial;
- busca e classificação em base de dados multimídia;
- reconhecimento biométrico, incluindo faces, íris ou impressões digitais;
- sensoriamento remoto por imagens multiespectrais;
- reconhecimento de fala.

Um ponto em comum entre essas aplicações é que usualmente as características disponíveis nos padrões de entrada, que tipicamente são milhares, não são diretamente utilizadas. Normalmente são utilizadas características extraídas dos padrões de entrada otimizadas usando-se procedimentos guiados pelos dados, como PCA.

Dado um padrão, seu reconhecimento/classificação consiste em uma das seguintes tarefas: **classificação supervisionada**, em que o padrão de entrada é identificado como um membro de uma classe pré-definida pelos padrões de treinamento, que são rotulados com suas classes; e **classificação não supervisionada**, em que o padrão é associado a uma classe que é aprendida com base na similaridade entre os padrões de treinamento. Neste trabalho será enfocado o primeiro tipo de classificação (supervisionada), o qual é mais usual em reconhecimento de faces.

Uma característica importante de reconhecimento de faces, assim como várias outras aplicações atuais, é que nenhuma abordagem individual é ótima, e que métodos e abordagens múltiplas devem ser utilizados, combinando várias modalidades de sensores e métodos de classificação [13]. O projeto de sistemas de reconhecimento de padrões essencialmente envolve três aspectos: aquisição de dados e pré-processamento, representação dos dados e tomada de decisões. Assim, geralmente o desafio encontra-se na escolha de técnicas para efetuar esses três aspectos.

Em geral, acredita-se que um problema de reconhecimento de padrões bem definido e restrito (com pequenas variações intra-classe e grandes variações inter-classes) permitirá uma representação compacta dos padrões e uma estratégia de decisão simples. Mas nem sempre os padrões a serem reconhecidos possuem essas características. Nesse fato reside a importância de algoritmos de extração e seleção de características.

Segundo Jain et. al. [13], as quatro abordagens mais bem conhecidas de reconhecimento de padrões são: casamento (*template matching*), abordagem estatística, sintática e redes neurais. A seguir há uma subseção que descreve resumidamente a abordagem estatística e, posteriormente, uma subseção que contém um resumo das outros métodos. É importante lembrar que não serão descritas abordagens que não foram aplicadas a reconhecimento de faces.

Nas seções seguintes, a abordagem estatística será descrita mais detalhadamente, principalmente no que diz respeito à extração e seleção de características. Quanto aos outros métodos, haverão novas refêrencias a eles, principalmente na seção 3.2, mas não serão descritas com maiores detalhes, visto que não pertencem ao escopo central deste trabalho.

3.1.1 Abordagem estatística

Basicamente, um sistema de reconhecimento estatístico de padrões pode ser composto pelas seguintes partes [14, 13] (vide figura 1): um sistema de aquisição de dados (por exemplo: sensores ou câmeras); um sistema de pré-processamento, para eliminar ruídos ou distorções; um extrator de características, que cria um vetor de características com dados extraídos dos objetos adquiridos, reduzindo os dados a atributos, propriedades ou características; um seletor de características, que analisa o conjunto de características e elimina as mais redundantes; e um classificador, que analisa um padrão obtido e toma uma certa decisão.

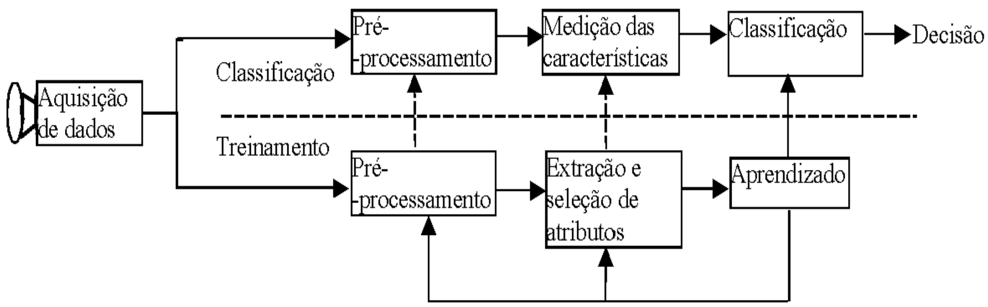

Figura 1: Um sistema genérico de reconhecimento de padrões (baseada em [14] e [13]).

O classificador toma decisões baseando-se no aprendizado obtido a partir de um conjunto de treinamento, o qual contém exemplos de padrões de todas as classes existentes no sistema. O reconhecedor de padrões é avaliado através de um conjunto de testes, preferencialmente composto por padrões de todas as classes, mas que não estão no conjunto de treinamento. Além do classificador, o pré-processamento, o extrator e o seletor de características podem ser dependentes dos dados de treinamento.

Essa estrutura não é particular dos métodos de reconhecimento estatístico, sendo também comum em outras abordagens de reconhecimento de padrões. Mas todos os sistemas de reconhecimento possuem um extrator e/ou um seletor de atributos. Na abordagem de *template matching* e em alguns casos de redes neurais, são utilizados dados no mesmo formato em que esses são adquiridos para classificá-los. No caso de sistemas estatísticos, quando o problema abordado for muito complexo, torna-se essencial o uso de extração e seleção de característica. Exemplos de problemas complexos são aqueles em que há muitas classes ou quando a dimensão dos padrões no formato em que são adqüiridos for muito alta.

Na abordagem estatística, cada padrão é representado em termos de d características (*features*) ou atributos, sendo visto como um ponto em um espaço d-dimensional.

O objetivo é que tal espaço seja gerado de forma que suas características permitam que vetores de padrões pertencentes a categorias diferentes fiquem dispostos de maneira que o processo de classificação seja simplificado. Por exemplo, pode ser desejável que os vetores ocupem regiões compactas e disjuntas em um espaço de características de d dimensões. A efetividade do espaço de características é determinada pelo quanto os padrões de classes diferentes podem ser separados.

Assume-se que cada padrão é uma observação obtida aleatoriamente segundo uma certa probabilidade condicional a uma determinada classe. Dado um conjunto de padrões de treinamento contendo elementos de todas as classes, o objetivo é estabelecer fronteiras de decisão no espaço de características que separem os padrões de classes diferentes. Na teoria de decisão estatística, a fronteira de decisão é determinada para cada classe, que pode ser especificada (aprendizado supervisionado) ou aprendida (aprendizado não-supervisionado). No caso de sistemas de reconhecimento de faces, normalmente é realizado aprendizado supervisionado [2], isto é, as imagens de treinamento possuem um rótulo que identifica de quem é a fotografia. Por esse motivo não serão não serão descritos métodos não-supervisionados de aprendizado, detalhes a respeito desse assunto podem ser encontrados em [15].

Para efetuar a classificação, também pode-se usar uma abordagem baseada em análise de funções discriminantes: primeiramente uma forma paramétrica da fronteira de decisão é especificada (por exemplo, linear ou quadrática), posteriormente a melhor fronteira de decisão da forma especificada é encontrada com base na classificação dos padrões de treinamento. Tais fronteiras podem ser construídas usando, por exemplo, o critério de erro médio quadrático [13, 12].

Existem várias técnicas estatísticas de reconhecimento de padrões, cada qual com seu nicho de aplicações. Nas seções 4, 5 e 6 há mais detalhes sobre alguns métodos estatísticos de extração e seleção de atributos, e classificação.

3.1.2 Outras Abordagens para o Reconhecimento de Padrões

Casamento (*template matching*)

Trata-se da abordagem mais antiga e mais simples. De maneira geral, dispõe-se de um modelo (*template*) que é um protótipo do padrão a ser reconhecido. No caso de imagens, o modelo é uma forma 2D. Para classificar, é realizado o casamento do modelo com o sinal (por exemplo, imagem) que deve conter o padrão a ser reconhecido, levando em consideração variações com translação, rotação e mudanças na escala. A medida de similaridade, normalmente relacionada à idéia de correlação, deve ser otimizada com base no conjunto de treinamento.

A maior desvantagem de *template matching* está em seu custo computacional. No caso do uso de *templates* rígidos, ocorrem falhas ao reconhecer padrões distorcidos ou com grande variação intra-classe. Os modelos de *templates* deformáveis podem ser usados quando as deformações não podem ser modeladas diretamente.

Em [12], há um capítulo com detalhes a respeito dessa abordagem.

Abordagem sintática

Trata-se de uma abordagem em que os padrões são tratados sob uma perspectiva hierárquica, ou seja, um padrão é composto por sub-padrões mais simples, os quais também são compostos por outros mais simples. Os sub-padrões elementares são chamados de primitivas, e um dado padrão complexo é representado em termos de inter-relações entre essas primitivas. Uma analogia formal pode ser traçada com a sintaxe de uma linguagem em que pode-se ver os padrões como sentenças pertencentes a uma linguagem, as primitivas como seu alfabeto, e as sentenças são geradas de acordo com a respectiva gramática. A gramática para cada classe de padrões deve ser inferida a partir dos exemplos de treinamento [13].

Uma vantagem de reconhecimento estrutural de padrões está no fato de que, além de fazer a classificação, essa abordagem também provém uma descrição de como os padrões são gerados a partir de primitivas que podem ser repedidas ou obedecerem determinadas ordens no sinal. Esse paradigma pode ser aplicado em padrões com estrutura definida, como imagens com textura e análise de formas de contornos.

Um exemplo de modelo de classificação sintática dependente do contexto em que observa-se uma ascensão recente é o de *Hiddem Markov Models* (HMM). Trata-se de um autômato finito estocástico [12] que gera uma cadeia de observação, isto é, a seqüência de vetores observados. Uma HMM consiste de um número de estados, e a cadeia observada é produzida como um resultado de sucessivas transições de um estado para outro, cada transição pode ocorrer com uma certa probabilidade, a qual é determinada na fase de treinamento. Na prática, os estados correspondem a certas características físicas, como, por exemplo, sons distintos. Em reconhecimento de fala, o número de estados depende do número de fenômenos sonoros esperados em cada palavra. Na realidade, tipicamente três ou quatro estados são utilizados para cada fenômeno, porém, o número exato de estados é freqüentemente resultado de experimentos e não pode ser determinado precisamente a priori. Além de reconhecimento de fala, HMM podem ser aplicados a reconhecimento de caracteres, e de eventos temporais de vídeo, como gestos [16] e expressões faciais.

As desvantagens da abordagem sintática estão na necessidade de, inicialmente, segmentar os padrões de ruídos para detectar primitivas, e na inferência de gramática dos dados de treinamento, pois essa abordagem pode permitir uma explosão combinatória das possibilidades a serem investigadas, o que demanda grandes conjuntos de treinamento e um alto esforço computacional [13].

Redes neurais

Uma rede neural artificial é um modelo de grafo orientado disposto em uma rede em que os nós representam neurônios artificiais e as arestas orientadas são as conexões entre as entradas e saídas dos neurônios. Redes neurais podem ser vistos como máquinas massivamente paralelas com muitos processadores simples e muitas interconexões [13].

As redes neurais podem ser hábeis a aprender relações entre entradas e saídas complexas e não lineares. Dentre outras características, inclui-se a possibilidade de usar procedimentos seqüenciais de treinamento, e o poder de adaptação aos dados.

Há várias famílias de redes neurais. Dentre elas, uma das mais comuns em reconhecimento de padrões é a *feed-forward network*, que inclui *perceptron* multicamadas e redes funções de bases radiais (*radial-basis function* - *RBF*). Essas redes são organizadas em camadas possuindo conexões unidirecionais entre uma camada e outra. Outra classe popular de redes é a dos mapas auto organizáveis (*self-organizing maps* - *SOM*), ou redes de Kohonen, as quais são usadas principalmente para agrupamento de dados (*clustering*) e em mapeamento de características. O processo de aprendizado envolve atualização da arquitetura da rede e dos pesos das conexões, de forma que a rede pode efetuar tarefas específicas de classificação e análise de agrupamentos (*clustering*).

A popularidade de redes neurais cresceu devido ao fato que, aparentemente, elas têm uma baixa dependência a um específico domínio, de forma que o mesmo tipo de rede pode ser aplicado em vários problemas, o que não ocorre com as abordagens baseadas em modelos ou em regras. Outro fator importante na popularidade é o fato de existirem muitos algoritmos eficientes de aprendizado com redes neurais.

As redes neurais disponibilizam um novo conjunto de algoritmos não lineares para extração de características (usando camadas escondidas) e classificação (p. ex. *mult-layer perceptrons*). Além disso, algoritmos de extração de características podem ser mapeados em arquiteturas de redes neurais para implementação eficiente em hardware.

Apesar de possuírem aparentemente princípios diferentes, a maioria dos modelos de redes neurais populares são implicitamente equivalentes ou similares a métodos clássicos de reconhecimento estatístico de padrões. Entretando, redes neurais oferecem várias vantagens, como abordagens unificadas para extração e classificação, e procedimentos flexíveis para encontrar boas soluções não lineares.

Lógica Nebulosa

A lógica nebulosa foi criada com inspiração no comportamento humano, que se baseia na interpretação do mundo sem precisão e na descrição desse por atributos lingüísticos.

A partir de conjuntos não-nebulosos (*crisp*), a criação de seu respectivos conjuntos nebulosos (*fuzzyficação*) é feita através da determinação do grau de pertinência de cada um dos elementos aos conjuntos. As funções de pertinência são baseadas na relação entre cada elemento e o(s) suporte(s) dos conjuntos. Essa relação pode ser, por exemplo, inversamente proporcional à distância entre os elementos e os suportes. Um suporte pode ser determinado, por exemplo, pelo centróide do conjunto de treinamento.

A classificação com lógica nebulosa normalmente é feita através de regras nebulosas e métodos de *defuzzificação*. As regras nebulosas normalmente são regras de inferência que utilizam delimitadores dos valores de pertinência de cada padrão aos conjuntos. O processo de *defuzzificação* consiste da utilização dos resultados das regras para a criação de conjuntos *crisp* que contenha informações a respeito da classificação dos elementos.

Maiores detalhes a respeito de conjuntos nebulosos, lógica fuzzy e aplicações podem ser encontrados em [17]. No presente trabalho, o uso de conjuntos nebulosos se aplica em algoritmos de seleção de características. Detalhes sobre esse assunto estão na seção 5.2.

3.1.3 Problemas de generalização

Nesta seção, discute-se os problemas de generalização de classificadores. Tais problemas são muito relevantes no projeto de sistemas de reconhecimento estatístico de padrões, mas que também podem ser comuns a sistemas não estatísticos, como redes neurais.

Não importando qual o classificador utilizado, ele deve ser treinado usando os exemplos de treinamento. Como resultado, o desempenho do classificador depende tanto do número de exemplos de treinamento como dos valores específicos dos exemplos, ou seja, da qualidade desses exemplos. Ao mesmo tempo, o objetivo do projeto de um sistema de reconhecimento é classificar futuros exemplos de teste que, preferencialmente, não sejam os mesmos que os de treinamento.

Porém, a otimização de um classificador para maximizar seu desempenho no conjunto de treinamento nem sempre produz um bom resultado para o conjunto de testes. A habilidade de **generalização** de classificadores se refere a seu desempenho ao classificar padrões de teste que não foram utilizados durante o treinamento.

Deficiências na capacidade de generalização de um classificador podem ser atribuídas aos seguintes fatores: sobre-ajuste (*overfitting*), sobre-treinamento (*overtraining*) e "maldição da dimensionalidade" (*curse of dimensionality*). A seguir encontram-se alguns detalhes a respeito desses fatores.

Sobre-ajuste (*ovefitting*)

Basicamente, há dois problemas que ocorrem se um classificador gerar superfícies de decisão mais complexas do que o requisitado pelo problema que esse aborda. O primeiro, e mais óbvio, é que o custo computacional dos processos de classificação é proporcional à complexidade do sistema classificador utilizado. Assim, se o classificador for mais complexo que o necessário, haverá desperdício de esforço computacional. O outro problema é que um classificador muito complexo possui menos poder de generalização.

A complexidade da superfície de decisão de um classificador é proporcional ao número de parâmetros livres que ele possui, ou seja, o número de parâmetros utilizados para a determinação das superfícies de decisão.

Intuitivamente pode parecer lógico que, para que um classificador seja poderoso, esse deve construir fronteiras de decisão detalhadas. Mas quando o número de parâmetros livres é grande, o classificador tende a se adaptar a detalhes específicos da base de treinamento, o que pode causar uma redução da taxa de acerto. Esse fenômeno é conhecido como *overfitting* (sobre-ajuste). Para evitá-lo, é desejável que o classificador seja o mais simples possível, pois assim será dada mais importância às maiores regularidades nos dados e as menores serão ignoradas, pois essas podem ser resultantes de ruídos. Também é possível evitar esse problema aumentando o número de padrões de treinamento do classificador, utilizando padrões que representem o máximo de variações possíveis dos dados. Mas obviamente isso não reduzirá o custo computacional do classificador.

Maiores detalhes sobre *overfiting* podem ser encontrados em [12].

Sobre-treinamento (*Overtraining*)

O problema de sobre-treinamento (*overtraining*) pode ser observado principalmente em sistemas de redes neurais, mas, de certa forma, também ocorre em sistemas estatísticos. Uma característica em comum com sobre-ajuste é o fato de ambos serem causados pela adaptação do classificador às peculiaridades específicas do conjunto de treinamento. Mas no caso de sobre-treinamento, o problema ocorre quando o classificador é treinado com um conjunto muito grande de exemplos de padrões com pequena variação intra-classe (no caso de classificadores estatísticos) ou com muitas iterações de treinamento (no caso de classificadores neurais). A conseqüência desse fato é que a capacidade de generalização do classificador é reduzida, proporcionando muitas falhas quando esse é utilizado para classificar padrões não pertencentes ao conjunto de treinamento. No caso de redes neurais, esse problema ocorre quando são realizadas muitas iterações de treinamento com pares de entrada e saída. A figura 2 ilustra esse fato.

Ainda no caso de redes neurais, uma maneira de evitar *overtraining* determinar qual é o número ideal do iterações de treinamento. Para isso, pode-se criar um gráfico como o da figura 2 e identificar o ponto de mínimo da curva de erro do conjunto de testes. Mas isso só é possível quando o tamanho do conjunto de

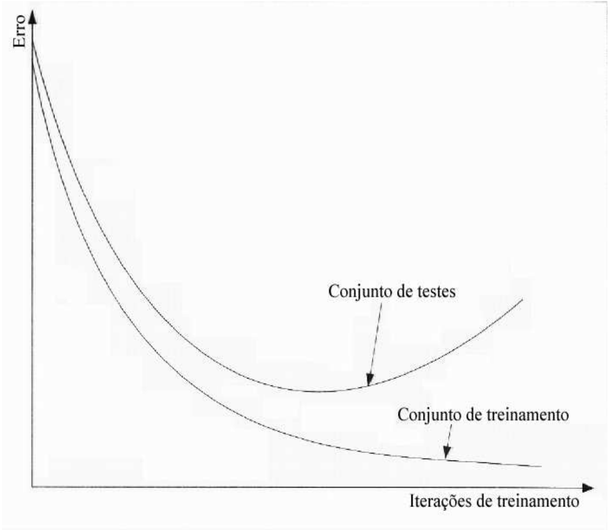

Figura 2: Curvas ilustrando o efeito de sobre-treinamento (baseada em [12]).

treinamento for grande o suficiente. No caso de sistemas estatísticos, uma solução é utilizar, para treinamento, dados que realmente representem todas as possíveis variações entre padrões de mesma classe.

Maldição da dimensionalidade (*Curse of dimensionality*) e o fenômeno do pico (*peaking phenomena*)

Trata-se do seguinte fenômeno: o número de elementos de treinamento requeridos para que um classificador tenha um bom desempenho é uma função exponencial da dimensão do espaço de características. Um caso ilustrativo em que esse fenômeno ocorre é o da técnica de particionamento do espaço de características para classificação baseada em árvores de decisão. Nessa técnica, cada reta suporte dos vetores da base do espaço de características é segmementada em intervalos regulares. A intercecção entre esses intervalos forma células no espaço. O reconhecimento de padrões é feito através da associação de uma classe a cada célula. Esse é um exemplo de sistema de classificação em que é bastante intuitivo verificar que, para que não sobrem células com classificação indefinida, é necessário que o número de elementos de treinamento seja uma função exponencial da dimensão do espaço de características. Esse fenômeno é bem conhecido pela comunidade de reconhecimento de padrões. Maiores detalhes podem ser encontrados em [13].

Nos casos em que o número de elementos de treinamento é arbitrariamente grande ou a distribuição estatística das classes é completamente conhecida, a probabilidade de erro de classificação de uma regra de decisão não aumenta com o aumento do número de características consideradas. Porém, nos problemas práticos, observa-se que a adição de características pode prejudicar o desempenho de um classificador se o número de exemplos de treinamento não for grande o suficiente em relação ao número de características. Esse fenômeno, chamado fenômeno do pico *peaking phenomena* é uma conseqüência da maldição da dimensionalidade e também foi amplamente estudado (por exemplo: [18, 19]). Todos os classificadores amplamente utilizados, inclusive redes neurais multi-camadas com retro-alimentação (*feed-forward*) sofrem da maldição da dimensionalidade.

3.1.4 Redução de *dimensionalidade*

O termo *dimensionalidade* é atribuído ao número de características de uma representação de padrões, ou seja, a dimensão do espaço de características. As duas principais razões para que a dimensionalidade seja a menor possível são: custo de medição e precisão do classificador. Quando o espaço de características contém somente as características mais salientes, o classificador será mais rápido e ocupará menos memória. Além disso, conforme discutido acima, quando o conjunto de exemplos de treinamento não é muito grande, um espaço de características pequeno pode evitar a maldição da dimensionalidade e propiciar pequenas taxas de erro ao classificador (vide seção 3.1.3) .

Além da necessidade de utilizar a menor dimensionalidade possível, há outro fator analizado pelo teorema do "patinho feito" de Watanabe [20], que diz que é possível fazer dois padrões arbitrários ficarem similares se esses forem codificados com um número suficientemente grande de características similares. Isso enfatiza a necessidade de uma escolha cuidadosa de características.

Para efetuar redução de dimensionalidade, basicamente há dois métodos: extração de características e seleção de características. Basicamente, os algoritmos de extração de características criam novas características a partir de transformações ou combinações do conjunto de características original. Já os algoritmos de seleção de características selecionam, segundo determinado critério, o melhor subconjunto do conjunto de características original.

Freqüentemente extração de características precede seleção de características, de forma que, inicialmente é feita a extração de características a partir dos dados de entrada, a seguir um algoritmo de seleção de características elimina os atributos mais irrelevantes segundo um determinado critério, reduzindo a dimensionalidade.

A escolha entre seleção e extração de características depende do domínio de aplicação e do conjunto específico de dados de treinamento disponíveis. Em geral a seleção de características reduz o custo de medição de dados, e as características selecionadas mantém sua interpretação física original, mantendo as propriedades que possuíam quando foram criadas. Já as características transformadas geradas por extração de características podem prover uma habilidade de discriminação melhor que o melhor subconjunto das características originais, mas as novas características (combinações lineares ou não lineares das características originais) podem não possuir um significado físico.

É importante salientar que, se a redução de dimensionalidade for excessiva, o classificador pode perder o poder de discriminação. Por isso é importante analisar a variação do comportamento do classificador com a dimensionalidade, de forma que seja possível estimar a dimensionalidade ideal para determinado classificador e conjunto de dados.

Maiores detalhes sobre extração e seleção de atributos estão nas seções 4 e 5.

3.2 Reconhecimento de Faces

Conforme dito na seção 1, devido à idade da pesquisa em reconhecimento de faces e devido à importância dessas pesquisas, esse problema foi amplamente abordado por vários cientistas, utilizando técnicas muito distintas. Nesta seção será feita uma breve revisão geral de métodos de reconhecimento de faces frontais.

Através da revisão de Chellappa et al. [2], publicada em 1995 e de outros trabalhos propostos após essa publicação [3], [4], [5], [6], [7] e [21], pode-se agrupar os métodos de reconhecimento de face nas seguintes categorias: por atributos (locais), holísticas, baseadas na transformada de Gabor, tridimensionais e de seqüências de vídeo. A seguir essas abordagens serão comentadas:

3.2.1 Por atributos

A abordagem de reconhecimento de faces por atributos é bastante intuitiva. Alguns dos métodos dessa categoria baseiam-se na construção de um vetor de características a partir de medidas de distância e ângulos entre pontos característicos da face, como pontos dos cantos dos lábios, centro da boca, nariz, narinas, pupilas, pontos extremos dos olhos, sobrancelhas, orelhas, pontos do contorno do queixo etc. e combinações dessas distâncias [22].

Esses métodos também são aplicados a imagens de faces de perfil, em que os pontos característicos são, por exemplo, a ponta do nariz, o ponto entre os lábios, a sobrancelha, testa, queixo, pescoço, etc. A vantagem do reconhecimento a partir de perfis está no uso de informações que não ficam disponíveis em imagens frontais bidimensionais, como o tamanho do nariz e do papo, além do fato de não ser difícil extrair esses pontos quando o fundo é uniforme.

O reconhecimento é feito através do uso de um sistema de reconhecimento de padrões, seja estatístico, seja por redes neurais, usando como espaço de características esses vetores. Com a ausência de detectores automáticos de pontos característicos da face, muitos autores utilizaram operadores humanos.

Alguns dos métodos automáticos de extração desses pontos se baseiam em contornos. Também podem ser consideradas como abordagens locais os sistemas de reconhecimento que utilizam informações como projeções horizontais dos mapas de bordas binários (verticais e horizontais) de imagens de face ou de partes da face, como nariz e boca [23]. É muito comum utilizar o reconhecimento por atributos para imagens de pessoas de perfil. O uso de energia de curvatura do contorno do perfil como forma de extração o vetor de características também é uma abordagem baseada em atributos.

A vantagem de abordagens locais é a invariância a translação, e a desvantagem está nos problemas ocasionados com alterações devido a expressões faciais e a rotações.

3.2.2 Métodos holísticos

Os métodos holísticos consideram todos os *pixels* da imagem ou de regiões características da face. Nessa abordagem a dimensionalidade dos dados é igual ao número de pixels das imagens consideradas. Para resolver esse problema podem ser utilizados métodos estatísticos de redução de dimensionalidade, como PCA [24], discriminantes lineares [19] ou redes neurais [25, 26]. O método de reconhecimento com PCA é o mais popular, tendo sido utilizado várias vezes em associação com pré-processamentos de normalização de imagens para melhorar o desempenho. A classificação pode ser feita de diversas maneiras, geralmente através de redes neurais ou sistemas estatísticos. Métodos baseados em pirâmides para reconhecimento de faces também podem ser classificados como holísticos.

Em [23], é feito um estudo comparando o desempenho de métodos por atributos (locais) com um método holístico. O método holístico testado foi o de *template matching*, testando o desempenho de imagens de olhos, narizes, bocas e também de toda a face. Os métodos holísticos proporcionaram resultados melhores que os locais. A vantagem de abordagens globais está no fato de que pequenas variações locais não prejudicam muito o reconhecimento. A principal desvantagem está nos problemas de variação de iluminação e, em alguns casos, no custo computacional.

3.2.3 Técnicas baseadas em Gabor

Neste item serão comentadas abordagens baseadas em extratores de características que utilizam transformada de Gabor em regiões da imagem, e não na imagem toda, o que diferencia essa abordagem das abordagens holísticas. Esses métodos são um pouco mais recentes e muito promissores. A mais conhecida é uma abordagem local que utiliza *jets*. Um *jet* é um vetor em que cada posição é determinada através de uma transformada de Gabor bidimensional com a janela Gaussiana (modulada por uma exponencial complexa) em um determinado local da imagem, como, por exemplo, o centro do nariz, as pupilas, os cantos da boca etc. Cada campo de um vetor *jet* é determinado pelo cálculo da transformada com uma janela de escala (variância da Gaussiana), orientação e/ou freqüência diferente. Cada *jet* é calculado em uma posição diferente de uma imagem (modelo) de treinamento. Na abordagem de *Elastic Graph Matching* [27, 28, 29], os *jets* são tratados como nós e as ligações entre os nós são arestas, formando um grafo.

Para efetuar a classificação, antes é necessário realizar um processo de combinação com imagens, que se refere ao posicionamento dos *jets* em determinados pontos da imagem. Para isso, um procedimento tenta localizar a posição para cada nó do grafo que, ao mesmo tempo, maximiza a similaridade das características e minimiza o custo da topografia. Esse processo é feito para as imagens de treinamento e de testes, de forma que os pontos importantes da face são localizados e os *jets* são calculadas nessas posições. As imagens de teste são classificadas de acordo com a similaridade com as imagens de treinamento, com base na combinação dos grafos e os *jets* obtidos. Essa abordagem pode apresentar problemas com variações na iluminação e com a imagem do fundo.

Uma outra abordagem que também utiliza a transformada de Gabor é a de *Gabor Wavelet Networks* [30], que consiste de uma técnica que representa um modelo discreto da face como a combinação linear de funções da wavelet Gabor 2D.

3.2.4 Métodos tridimensionais

Há vários métodos de reconhecimento baseados em informações tridimensionais, como os que utilizam *range images* e os baseados em visão estéreo. A vantagem desses métodos está na possibilidade de obtenção de informações relevantes das superfícies da cena real que não podem ser obtidas através de imagens bidimencionais. A desvantagem está na necessidade de sistemas de aquisição mais caros, como um *scanner* tridimensional (para *range images*) e de uma câmera extra (para visão estéreo), além do fato do custo computacional para tratar tais informações é maior. A pesquisa em métodos tridimensionais para reconhecimento de faces não avançou muito nos últimos anos pois o uso de sistemas bidimensionais baseados em treinamento com múltiplos pontos de vista (ou várias orientações da face) [31], bem como sistemas de rastreamento e determinação da orientação tridimensional a partir de imagens bidimensionais [32], possibilitam a obtenção de bons resultados dispensando sistemas mais caros.

3.2.5 Vídeo

Até 1995, a pesquisa em sistemas de reconhecimento baseada em seqüências de vídeo estava começando a se desenvolver. Alguns testes com seqüências de vídeo realizados apenas para mostrar vários resultados de classificação, como em [33], em que, para cada quadro, foi feito um teste de classificação individual para comparar o método baseado em *eigenfaces* com o de combinação de grafos em seqüências com variação de expressão facial. Além disso, alguns pesquisadores empregavam seqüências temporais para integrar as informações estáticas da câmera com sons de sílabas. Esses sistemas possuíam dois módulos: o de vídeo, o qual capturava apenas um quadro para efetuar o reconhecimento empregando algum método comum; e o de audio, que normalmente utiliza informações da freqüência da voz do indivíduo para efetuar o reconhecimento. Os resultados de classificação eram combinados obtendo uma taxa de reconhecimento melhor que a dos sistemas isolados. Como exemplos de métodos dessa abordagem, tem-se os que estão descritos em [34, 35].

Com o surgimento de métodos eficientes de detecção e rastreamento de pessoas, juntamente com o uso de treinamento com faces em diversas orientações, existem sistemas melhores de reconhecimento de faces em seqüências de vídeo com procedimentos de escolha de "bons" quadros das seqüências [36].

Porém, até o presente momento não há um sistema conhecido de reconhecimento ou categorização em seqüências de vídeo bem conhecido que realmente aproveite o movimento para extrair informações extras, como movimentos que possam caracterizar pessoas (*gait*). Em [37], há um estudo psicofísico que mostra que informações a respeito do jeito das pessoas andarem também podem ser utilizadas para indentificá-las quando a imagem da face não tiver qualidade suficiente ou estiver ocluída. Por outro lado, a pesquisa em reconhecimento de gestos (de cabeça [16] e de mão), bem como a de reconhecimento de movimentos do corpo, como danças e interpretações dramáticas está se desenvolvendo muito rapidamente e com resultados bastante promissores, mas não foi aplicada para identificação.

3.2.6 Cronologia e estado-da-arte

Segundo Pentland [3], o primeiro sistema famoso de reconhecimento automático de faces provavelmente é o de Kohonen, proposto em 1989. Kohonen demonstrou que uma simples rede neural pode desempenhar reconhecimento de faces usando imagens de faces registradas (normalizadas e alinhadas). Foi empregada uma rede que computa a descrição das faces através da aproximação dos auto-vetores da matriz de auto-correlação das imagens de face. Como sabemos, esses auto-vetores mais tarde ficaram conhecidos como *eigenfaces*. O sistema de Kohonen não foi um sucesso prático, pois ele depende de alinhamento e normalização das faces.

Nos anos seguintes, muitos pesquisadores tentaram esquemas de reconhecimento de faces baseados em características locais, como bordas, distâncias entre pontos característicos e outras abordagens com o emprego de redes neurais. Enquanto muito sucesso foi obtido em bases de imagens pequenas com faces alinhadas, nenhum trabalho obteve sucesso em problemas mais realísticos de grandes bases de dados e com localização, orientação e escala da face são desconhecidas [3].

O método de reconhecimento de faces utilizando a transformada de Karhunen Loève foi proposto em [38] e está descrito com maiores detalhes na seção 4.2. Em [24], foi demonstrado que o erro residual da codificação usando *eigenfaces* pode ser usada tanto para detectar faces em imagens naturais como para a determinação precisa da localização, escala e orientação de faces na imagem. Também foi mostrado que esse método pode ser associado ao método de reconhecimento usando *eigenfaces* para obter reconhecimento de faces confiável em imagens com pouca restrição.

A partir de 1993, surgiram vários outros sistemas de reconhecimento robustos a imagens não normalizadas. Segundo Pentland [3], os três melhores são:

- **Aprendizado Visual Probabilístico** [39, 40]: utiliza a transformada de *eingeface* seguida por um discriminante linear.
- **Análise de Discriminantes Lineares** [41]: também se baseia na transformada de *eigenface*, mas utiliza um discriminante quadrático.
- ***Elastic Graph Matching*** [27]: conforme descrito anteriormente, trata-se da computação dos "Gabor jets" seguida pela realização de uma comparação baseada em um modelo flexível (*flexible template*) entre os descritores da imagem

usando um algoritmo de casamento de grafos (*graph-matching*).

Para avaliar os algoritmos de reconhecimento de faces, há o programa FERET (Face Recognition Technology) [42], o qual é conhecido por ser o teste mais abrangente proposto até agora. A base de dados do FERET possui faces com variações em posição, escala e iluminação de modo consistente com as fotografias 3×4 ou as de carteira de habilitação para motoristas americanos. Há imagens de pessoas obtidas de fotos tiradas em datas diferentes (a diferença chega a um ano).

O maior teste do FERET possui imagens de 1196 pessoas diferentes. Nesse teste, os algoritmos citados acima possuem desempenho muito similar. Em imagens frontais do mesmo dia, o desempenho dos algoritmos foi de mais de 95% de acerto. Para imagens obtidas com câmeras e iluminações diferentes, o desempenho foi entre 80 e 90%. Para imagens tomadas um ano depois, a taxa de reconhecimento típica foi de 50%. A diferença entre os algoritmos foi menor que 0.5%.

Para testes com 200 pessoas, os três algoritmos praticamente não erraram. Entretanto, nesse experimento, mesmo um simples método de combinação por correlação pode, algumas vezes, propiciar o mesmo resultado, com a diferença de se tratar de um método lento. Por isso, Pentland [3] sugere que, para que um novo algoritmo seja considerado potencialmente competitivo, esse deve ser testado com bases de dados com, no mínimo, 200 indivíduos, devendo resultar em uma taxa de reconhecimento maior que 95%.

Porém, como sabemos, esses resultados são válidos para imagens estáticas, e ainda não há um bom método definitivo de testes de algoritmos destinados a reconhecimento de pessoas a partir de seqüências de vídeo.

4 Extração de características

Dado um espaço de características de dimensão M, os métodos de extração de características determinam um sub-espaço apropriado de *dimensionalidade* m ($m < M$). Há métodos lineares e não lineares de extração de características, sendo que, de modo geral, processos lineares de extração de características podem ser definidos de acordo com a equação 1:

$$Y = H \cdot X, \tag{1}$$

onde X é a matriz $M \times n$ de padrões em que cada coluna contém um padrão de treinamento, H é a matriz $m \times M$ de transformação linear e Y é a matriz $m \times n$ dos padrões derivados, sendo que n é o número de padrões de treinamento. A seção 4.2 descreve como a matriz X é criada quando os padrões são imagens de faces.

Já os extratores de características não lineares não apresentam uma fórmula genérica. Dentre os extratores lineares estão: transformada de Fourier, análise de componentes principais (PCA), análise de discriminantes lineares (por exemplo, o discriminante de Fisher) e outras projeções lineares em geral. Em relação aos extratores não lineares, pode-se citar as redes neurais e outros métodos de projeção não lineares, como os heurísticos.

4.1 Transformada de Fourier

A transformada de Fourier é uma ferramenta muito importante em processamento de imagens. Através da transformada de Fourier é possível decompor um sinal em seus componentes de freqüência (senos e cosenos). Assim, um coeficiente de Fourier reflete a importância de determinada freqüência para o sinal.

Dado um sinal contínuo e unidimensional $u(t)$, sua transformada de Fourier é definida por:

$$U(f) = F(u(t)) = \int_{-\infty}^{\infty} u(t) e^{-i2\pi ft} dt, \tag{2}$$

onde f denota freqüência e t denota tempo. A inversa da transformada de Fourier é definida por:

$$u(t) = F^{-1}(U(f)) = \int_{-\infty}^{\infty} U(f) e^{i2\pi ft} dt, \tag{3}$$

A trasformada discreta de Fourier é definida como:

$$U(\frac{s}{NT}) = \sum_{n=0}^{N-1} u(nT)e^{-i2\frac{\pi}{N}ns}, s = 0, 1, \cdots, N-1 \tag{4}$$

Analogamente, obtém-se a transformada discreta inversa de Fourier:

$$u(NT) = \sum_{n=0}^{N-1} U(\frac{s}{nT})e^{i2\frac{\pi}{N}ns}, s = 0, 1, \cdots, N-1 \tag{5}$$

onde T é o intervalo de amostragem de $u(t)$.

Dentre as principais aplicações da transformada de Fourier, encontram-se análise, filtragem, reconstrução e compressão de imagens, bem como reconhecimento de padrões e de objetos. Detalhes a respeito da construção dessa transformada, bem como suas propriedades, a versão bidimensional e suas aplicações podem ser encontrados em [43, 44, 45].

4.2 PCA

Segundo Jain et al. [13], o melhor extrator de características linear conhecido é o de análise de componentes principais (PCA). Essa transformada é amplamente utilizada pela comunidade de reconhecimento de padrões e de reconhecimento de faces.

Conforme dito anteriormente, a transformada PCA, também conhecida por transformada de Hotelling e por expansão de Karhunen-Loève, é amplamente utilizada para efetuar reconhecimento de faces [38, 24, 2, 26].

Visando tratar imagens como padrões em um espaço linear, sendo h o número de linhas de uma imagem e w o número de colunas, pode-se dizer que uma imagem é um padrão de $h \cdot w$ características, ou um vetor no espaço $(h \times w)$-dimensional, o qual chamaremos de "espaço de imagens".

Dada uma imagem representada como uma matriz $h \times w$, pode-se construir sua representação como um vetor através de uma leitura linha a linha (*raster*) da imagem, colocando o valor de cada *pixel* da imagem em um vetor coluna, o qual é inserido na matriz X. Ou seja, dada uma matriz I^i de h linhas e w colunas representando a imagem i, $X_{i,j+w(k-1)} = I^i_{j,k}$, para $j = 1, 2, 3, \cdots, w$, $k = 1, 2, 3, \cdots, h$ e $i = 1, 2, 3, \cdots, n$, em que n é o tamanho do conjunto de treinamento e $h \cdot w = M$ (vide equação 1).

Em reconhecimento de padrões, é sempre desejável dispor de uma representação compacta e com um bom poder de discriminação de classes de padrões. Para isso, é desejável que não haja redundância entre as diferentes características dos padrões, ou seja, que não haja covariância entre as bases do espaço de características. Mas, obviamente, pode-se notar que o espaço de imagens é altamente redundante quando usado para descrever faces, pois cada *pixel* é altamente correlacionado com outros *pixels*.

Para verificar se há covariância entre as características (ou variáveis), utiliza-se a matriz de covariância dos padrões [46, 14]. Dada uma matriz de padrões X conforme definida na equação 1, a matriz Σ de covariância de X pode ser obtida por:

$$\Sigma = (X - \mu) \cdot (X - \mu)^t, \tag{6}$$

em que μ é a matriz $M \times n$ (mesma dimensão que X), em que todas suas colunas contém o valor esperado de X, o qual pode ser obtido pela média aritmética das colunas de X. Ou seja:

$$\mu_{l,i} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_{l,i}, \tag{7}$$

para $l = 1, 2, 3, \cdots, M$.

É importante notar que $\Sigma_{l,l}$ é a variância da característica l dos padrões ($X_{l,i}$, para todo $i = 1, \cdots, n$). Ou seja, os elementos da diagonal da matriz de covariância se referem à variância das características. Já os elementos fora da diagonal $\Sigma_{l,o}$ representam a covariância entre a característica l e o. Se duas características, l e o, são estatisticamente independentes, a covariância é nula ($\Sigma_{l,o} = 0$).

Conforme dito anteriormente, é desejável que os padrões sejam representados em um espaço em que não haja covariância entre características diferentes. Um espaço vetorial com essa propriedade possui uma base cuja matriz de covariância de seus vetores é diagonal. Partindo-se de um conjunto de exemplos de padrões de treinamento, para obter uma base com tal propriedade basta utilizar uma transformada que diagonalize a matriz de covariância da base atual do espaço. Com a diagonalização da matriz de covariância, a variância das variáveis (características) será maximizada e a covariância entre uma variável e outra será nula.

Devido ao processo de criação da matriz de covariância, pode-se mostrar que ela é diagonalizável [12, 14, 47]. A diagonalização é realizada através dos auto-vetores de Σ [47], ou seja, a matriz H $(m \times M)$ da equação 1 deve ser obtida através dos auto-vetores da matriz Σ dispostos cada um em uma linha. Maiores detalhes a respeito de obtenção de auto-valores e auto-vetores podem ser encontrados em [47].

É importante lembrar que os auto-valores refletem a importância dos auto-vetores. No caso de PCA, os auto-valores da matriz de covariância são iguais à variância das características transformadas [12]. Assim, se um auto-vetor possui auto-valor grande, significa que este fica em uma direção em que há uma grande variância dos padrões. A importância desse fato está na classificação, pois, em geral, é mais fácil distinguir padrões usando uma base em que seus vetores não são correlacionados e que apontam para a direção da maior variância dos dados.

Através das figuras 3, 4 e 5 pode-se visualizar o efeito da transformada PCA para o caso bidimensional. Pode-se notar que é realizada uma rotação da base do espaço vetorial de forma que o primeiro vetor da nova base fique na direção em que há maior variância dos dados e o segundo fique perpendicular ao primeiro, na direção da segunda maior variação.

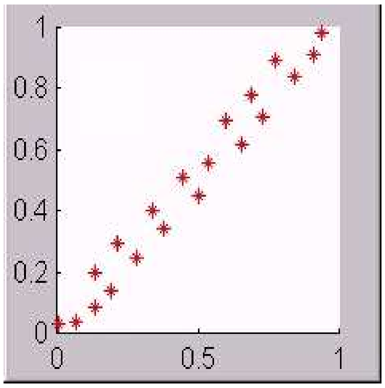

Figura 3: Dados artificiais bidimensionais.

O número de auto-vetores obtido é, no máximo, igual ao número de *pixels* da imagem (ou variáveis dos padrões de entrada). Porém, conforme dito acima, se a matriz H for construída de forma que forem escolhidos somente os auto-vetores contendo os maiores auto-valores, a variância total dos padrões de entrada não sofre grandes mudanças. Assim, é possível realizar redução de dimensionalidade utilizando-se, na construção de H, somente os m maiores auto-vetores, de forma que

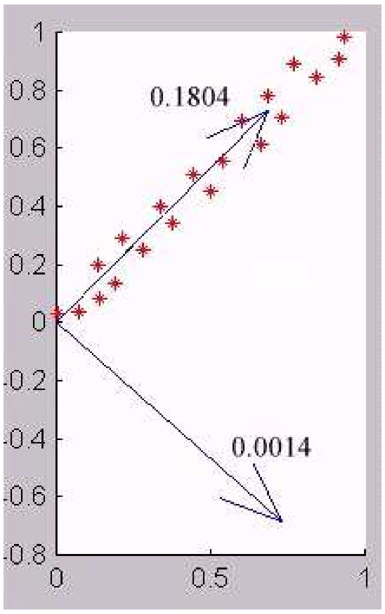

Figura 4: Dados de teste com os auto-vetores da matriz de covariância e seus respectivos auto-valores.

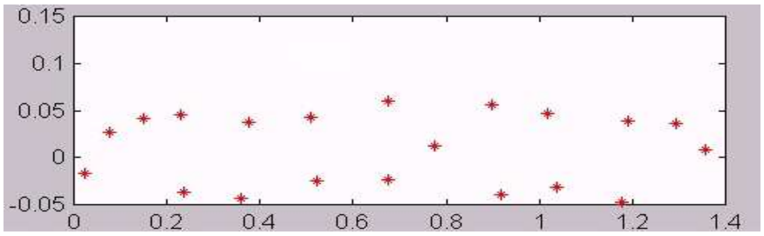

Figura 5: Dados no espaço criado.

a dimensionalidade de Y será $m \times n$, o que significa uma redução de dimensionalidade de $M - m$ dimensões. Após a redução da dimensionalidade com PCA, em geral, para reconhecimento de faces, a classificação é efetuada utilizando-se o classificador do vizinho mais próximo (vide seção 6).

A transformada PCA é relativamente simples, mas o processo de treinamento é complexo, visto que, dentre outras operações, é necessário efetuar $M \times M \times n$ multiplicações para criar a matriz de covariância Σ [18]. Porém a aplicação é muito rápida e, em geral, produz bons resultados para reconhecimento de faces. Mas segundo [14], PCA é uma técnica de extração de características não supervisionada propícia para dados com distribuição Gaussiana, mas não se tem certeza de que as faces possuam tal distribuição. Através das figuras 6, 7 e 8 pode-se observar um caso simples bidimensional em que, caso seja feita uma redução de dimensionalidade, será muito mais difícil distinguir os dois padrões, pois o segundo auto-vetor é melhor para distinguir as duas classes do que o primeiro.

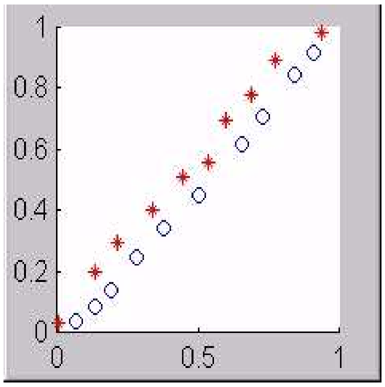

Figura 6: Dados artificiais de teste: duas classes em um espaço bidimensional.

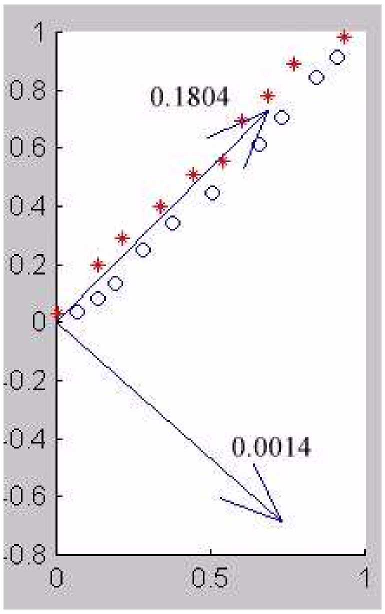

Figura 7: Dados de teste de duas classes com os auto-vetores da matriz de covariância e seus respectivos auto-valores.

Conforme será descrito posteriormente, uma forma de eliminar esse problema consiste na aplicação de um algoritmo de seleção de características (vide seção 5).

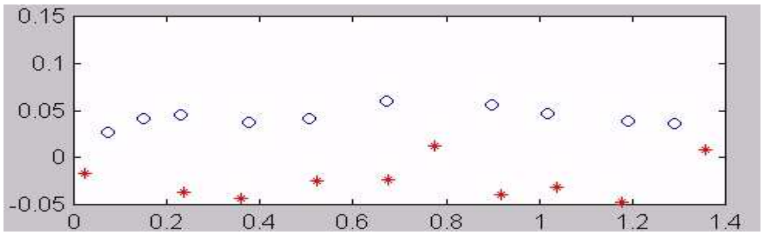

Figura 8: Dados no espaço criado: note que o primeiro auto-vetor não possui poder de discriminação.

4.3 Discriminantes Lineares

Como pode-se observar na seção 4.2, a transformada de PCA é um método linear não supervisionado de extração de características que assume que os dados possuam uma distribuição Gaussiana [13]. Essas características possibilitam a ocorrência de problemas como aquele ilustrado nas figuras 6, 7 e 8. Para evitar tais problemas, pode-se aplicar algoritmos de seleção de características ou utilizar extratores de características que utilizam informações da distribuição das classes no espaço original.

Através de análise de discriminantes esses problemas podem ser evitados, pois trata-se de um método que utiliza informações das categorias associadas a cada padrão para extrair linearmente as características mais discriminantes. Em análise de discriminantes, a separação inter-classes é enfatizada através da substituição da matriz de covariância total do PCA por uma medida de separabilidade como o critério Fisher.

O critério Fisher se trata da determinação dos auto-vetores de $S_w^{-1} \cdot S_b$, sendo que S_w é a matriz de espalhamento intra-classes e S_b é a matriz de espalhamento inter-classes. Ou seja, esse método extrai coeficientes que minimizam o espalhamento de padrões de mesma classe entre classes e maximizam o espalhamento de padrões de classes diferentes.

Em geral, essa abordagem possibilita a obtenção de resultados melhores que o PCA para redução de dimensionalidade. Por exemplo, no caso ilustrado na figura 3, o discriminante linear de Fisher iria determinar, como primeiro vetor da base, exatamente aquele determinado pelo segundo auto-vetor no caso de PCA, ou seja, o vetor cujo auto-valor é 0.0014 na figura 7.

O principal problema dessa abordagem está em sua incapacidade em obter bons resultados se aplidada a classes com distribuição côncava e com interceção com outras classes, como no caso de dados com distribuição similar aos da figura 9 (em todas as dimensiões). Nesse caso, a transformada vai tentar minimizar a variação intra-classe e maximizar a variação inter-classes, o que pode resultar em uma representação dos dados pior do que a original para classificadores como os k-vizinhos mais próximos.

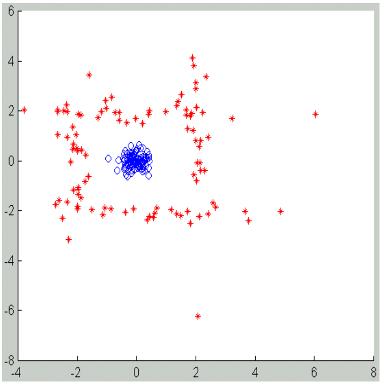

Figura 9: Exemplo de distribuição que pode falhar com um discriminante linear.

Maiores detalhes a respeito de discriminantes lineares podem ser obtidos em [12].

5 Seleção de Características

Métodos automáticos de seleção de características são importantes em muitas situações em que tem-se disponível um conjunto grande de características e é desejável selecionar um subconjunto adequado. Além de ser uma forma de redução de dimensionalidade, uma aplicação importante é a fusão de dados procedentes de múltiplas modalidades de sensores ou de múltiplos modelos de dados. A importância de redução de dimensionalidade está explicita na seção 3.1.4.

Seleção automática de características é uma técnica de otimização que, dado um conjunto de D características, tenta selecionar um subconjunto de tamanho d que maximiza uma função critério.

Formalmente, dado um conjunto Y de D características, o algoritmo de seleção de características deve encontrar um subconjunto $X \subseteq Y$ tal que $|X| = d$, e

$$J(X) = \max_{Z \subseteq Y, |Z| = d} J(Z), \tag{8}$$

em que $J(.)$ é a função critério, a qual pode, por exemplo, ser definida por, $1 - p_e$, sendo p_e é a probabilidade de erro.

É desejavel que a função critério seja maior quanto menor for a redundância entre as características, e quanto maior a facilidade de discriminar padrões de classes diferentes.

Dessa forma, o algoritmo de seleção de características poderá reduzir a dimensionalidade sem que um classificador perca o poder de distinguir classes no espaço de características. Uma conseqüência da aplicação de um bom algoritmo de seleção de atributos é a redução do número necessário de amostras de treinamento para se obter bons resultados com um classificador, ou seja, a redução de *curse of dimensionality*.

Além da escolha da função critério, também é importante determinar a dimensionalidade apropriada do espaço de características reduzido. Uma forma simples de resolver esse problema é efetuar seleção de características para vários valores de d. Em [13], os autores defendem que há uma relação entre o tamanho do conjunto de treinamento e a dimensionalidade ideal para um sistema de classificação. Sendo n o tamanho do conjunto de treinamento, é seguro que não ocorre o *curse of* *dimensionality* se forem usadas menos que $n/10$ características.

Apesar da importância de seleção de atributos, não há regras ou procedimentos definitivos para essa tarefa em cada aplicação particular [43], principalmente quando o número de características disponíveis é grande. Por esse motivo, um grande conjunto de algoritmos de seleção de atributos tem sido proposto. Em [48], Jain et al. propôs uma taxonomia sobre este tópico. A seguir serão descritos separadamente alguns algoritmos de seleção de características e algumas funções critério.

5.1 Algoritmos de seleção

Há vários métodos diferentes de seleção de características. De acordo com Jain et al. [48], tais abordagens podem ser agrupados nas seguintes categorias de acordo com a taxonomia exibida na figura 10.

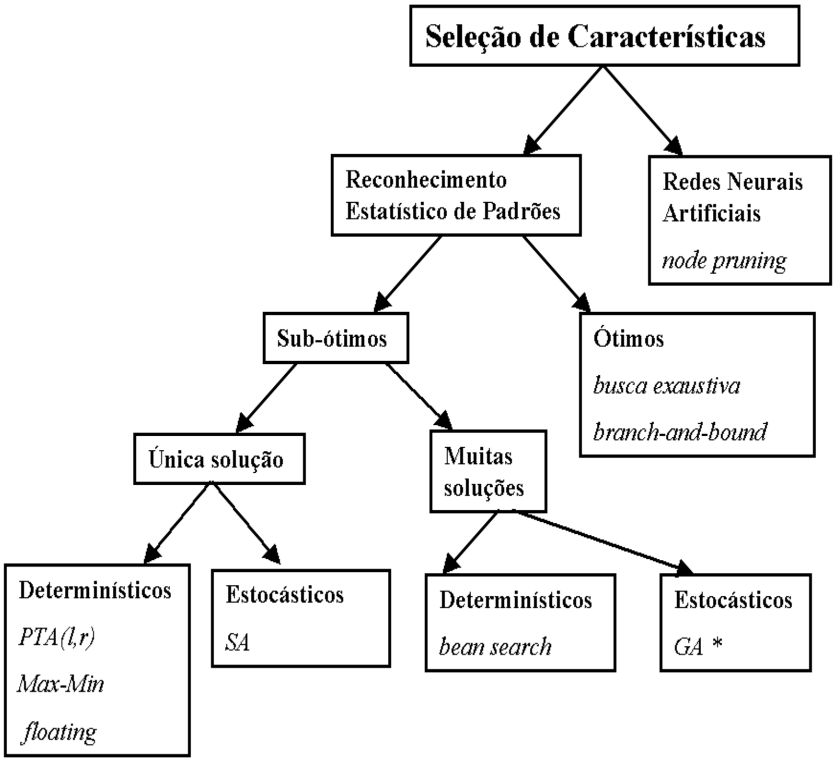

Figura 10: Taxonomia dos métodos de seleção de características. Adaptada da figura 1 contina em [48].

5.1.1 Redes Neurais

Um método de seleção de características bem conhecido que utiliza uma rede neural é chamado **Node Pruning** [49], ou "corte de nós". Basicamente o algoritmo funciona através de uma rede neural multi-camadas com retro-alimentação, utilizando um algoritmo de aprendizado baseado em retro-propagação (*backpropagation*). É definida uma medida de "saliência de nós" e utilizado um algoritmo que elimina os nós menos salientes. Dessa forma, a complexidade da rede pode ser reduzida após seu treinamento. A eliminação dos nós de entrada significa a eliminação de características do conjunto de características. A saliência de um nó é definida pela soma do aumento no erro sobre todos os padrões de treinamento, como um resultado da remoção daquele nó.

Inicialmente a rede neural é treinada, sendo posteriormente realizada a eliminação de nós seguida de um re-treinamento da rede, repetindo-se o processo até que seja alcançada a dimensão desejada. A vantagem do método node-pruning é que ele simultaneamente determina o melhor subconjunto de características e o classificador ótimo.

5.1.2 Métodos ótimos

Em termos de resultado obtido, o único método realmente ótimo é o da **busca exaustiva**, que avalia todos os $\binom{D}{d}$ subconjuntos possíveis de tamanho d. Essa abordagem é muito cara computacionalmente mesmo para conjuntos não muito grandes, pois sua complexidade é exponencial.

Algumas funções critério possuem uma propriedade chamada **monotonicidade**. Uma função é monotônica se $J(A \bigcup B) \geq J(A)$, para todo $A, B \subseteq Y$, ou seja, o valor da função critério é sempre maior para conjuntos de características maiores. Para este caso, há o algoritmo de busca em árvores chamado **branch-and-bound** [50]. Esse algoritmo pode retornar a solução ideal sem verificar todas as possibilidades, mas sabemos que, devido a *curse of dimensionality* (vide seção 3.1), para situações em que o conjunto de treinamento não é grande o suficiente, a função critério não é monotônica. Por esse fato, o algoritmo *branch-and-bound* não pode ser aplicado em qualquer situação.

Outra desvantagem desse método é que, no pior caso, todas as configurações são consultadas, o que faz com que o algoritmo tenha complexidade exponencial no pior caso, o que o torna impraticável para grandes conjuntos de características.

5.1.3 Métodos estocásticos com múltiplas soluções

Essa classe de métodos engloba o uso de algoritmos genéticos para seleção de características, que foi introduzido em [51]. Nessa abordagem, o conjunto de características é representado como uma cadeia binária de caracteres de tamanho D em que 0 ou 1 na posição i indica a ausência ou presença da característica i. Essa cadeia é chamada "cromossomo".

Inicialmente, uma população aleatória de cromossomos é criada. Cada cromossomo é avaliado para determinar sua aptidão (*fitness*), a qual determina se o cromossomo irá "sobreviver" à próxima geração ou "morrer". A partir de mutações ou cruzamentos dos cromossomos atuais são criados novos cromossomos.

Após várias iterações, a aptidão geral da população será melhorada e sempre haverão várias soluções, porém, como os resultados são obtidos a partir de processos estocásticos (portanto não-determinísticos), normalmente são obtidos sub-conjuntos diferentes quando o algoritmo é aplicado ao mesmo conjunto em outro momento.

Em [52] essa técnica foi aplicada para efetuar a classificação de formas biológicas.

5.1.4 Métodos determinísticos de múltiplas soluções

Alguns métodos tratam o espaço de subconjuntos de características como um grafo, chamado "reticulado de seleção de características", em que cada nó representa um subconjunto e uma aresta representa a relação de pertinência. Para selecionar os melhores conjuntos, aplica-se um algoritmo padrão de busca em grafos. Em [51], os métodos "**best-first search**" e uma versão restrita chamada "**beam search**" foram testados para seleção de características.

5.1.5 Métodos Determinísticos com Solução Única

Há vários métodos de seleção de características determinísticos de solução única. A seguir serão descritos alguns desses métodos que são baseados em técnicas de busca.

Melhor Característica Individual [48, 12]

Esse método consiste na avaliação de todas as características tomadas individualmente e seleção das d melhoreres. Trata-se de um método bastante intuitivo e computacionalmente simples, mas não é garantido que o melhor subconjunto seja determinado, pois algumas características podem ser boas tomadas individualmente, mas podem formar um conjunto ruim quando associadas entre si.

Busca seqüencial para frente (SFS) [48, 12]

Trata-se de um método *botton-up* em que, dado um conjunto de características já selecionadas (inicialmente nulo), a cada iteração seleciona a característica que, unida ao conjunto determinado pela iteração anterior, produz a o melhor resultado da função critério. Essa característica é adicionada ao conjunto de características anterior e uma nova iteração é realizada. São realizadas d iterações. A desvantagem desse método é que, uma vez que uma característica tenha sido selecionada, ela não pode ser descartada do subconjunto ótimo. Isso proporciona o chamado efeito "nesting" que pode ocorrer quando o subconjunto ótimo não contém elementos do conjunto já selecionado. A vantagem é o custo computacional quando se deseja obter conjuntos pequenos em relação ao total de caracteríscias.

Busca seqüencial para trás (SBS) [48, 12]

Esse algoritmo é uma versão *top-down* do algoritmo anterior. A diferença entre SBS e SFS é que o SBS é iniciado com o conjunto de características completo (contendo todas as D características) e vai eliminando as menos importantes, ou seja, as que menos alteram a função critério quando são eliminadas. Assim como o método de busca seqüencial para frente, a desvantagem desse método é que, uma vez eliminada uma característica, ela não retornará ao subconjunto ótimo novamente. Como conseqüência também pode ocorrer o efeito "nesting" caso o melhor subconjunto contenha alguma das características que foram eliminadas. A vantagem desse método é o custo computacional, quando se deseja obter conjuntos grandes em relação ao total de características.

Algoritmos de busca seqüencial generalizada (GSFS e GSBS) [53, 12]

Tratam-se de algoritmos de seleção que inserem (no caso do GSFS) ou removem (no caso do GSBS) tuplas (subconjuntos) de características ao invés de fazerem com apenas uma característica por vez. O que possibilita o funcionamento dos algoritmos generalizados é a utilização de funções que determinam a significância de tuplas.

Mais l - menos r (PTA) [53, 12]

Esse método, cujo nome original é "Plus l - Take Away r" (PTA), foi criado visando evitar o efeito "nesting". Basicamente, primeiro o algoritmo aumenta o conjunto de características em l elementos usando o método de seleção para frente, e depois elimina r características usando a busca seqüencial para trás. Os valores de l e r devem ser determinados pelo usuário. Na versão *botton-up*, l deve ser maior que r. Já na versão *top-down*, $l < r$.

Métodos de busca seqüencial flutuante para frente e para trás (SFFS e SFBS) [54, 12]

Tratam-se de generalizações do método "mais l - menos r", em que os valores de l e r são determinados e atualizados dinamicamente. Esses métodos provém soluções muito próximas da solução ótima com um pequeno custo computacional. O método de busca para frente é a versão *botton-up*, e o de busca para trás, *top-down*. Segundo Jain et al. [48, 13] esses são os métodos que melhor combinam tempo de execução com qualidade dos resultados.

Métodos adaptativos de busca seqüencial flutuante para frente e para trás [53] (ASFFS e ASFBS)

Esses métodos são generalizações dos métodos de busca seqüencial flutuante, ou seja, podem ser inseridos e eliminados não somente características, mas também conjuntos (tuplas) de características por vez. Esses conjuntos têm seu tamanho e seu conteúdo determinados dinamicamente. A vantagem desses métodos é que, em termos de resultado da função critério, os resultados obtidos são mais próximos dos resultados ótimos. A desvantagem é que, quando os conjuntos de dados que são inseridos e eliminados por vez é grande, o algoritmo se torna muito mais lento e complexo que os métodos de busca seqüencial flutuante. Esse resultado está descrito na seção 7.2.1 e em [9].

5.2 Funções critério

Conforme mencionado anteriormente, uma das partes mais importantes na redução da dimencionalidade é a escolha de uma função critério. Em seleção de características, o objetivo das funções critério é minimizar o erro de classificadores. Conforme mencionado anteriormente, um exemplo de função critério: $J(X) = 1 - pe$, sendo pe a probabilidade de erro de um classificador no espaço de características. Essa probabilidade de erro pode ser determinada através da taxa de acerto de um classificador ou da distância entre as classes de padrões de treinamento no espaço de características.

A seguir serão descritos sucintamente algumas funções critério conhecidas.

5.2.1 Desempenho de um Classificador

Um critério amplamente utilizado é o de erro de classificação de um subconjunto de características. Basicamente, quando não se dispõe de informações a respeito da distribuição dos dados, cada conjunto de características é utilizado para treinar e testar um classificador. A taxa de reconhecimento é utilizada como critério de decisão, de forma que, quanto maior a taxa de reconhecimento melhor é o conjunto de características.

Segundo Kohn [55], deve-se tomar o cuidado de não empregar o conjunto de treinamento (utilizado para projetar o classificador) para estimar a probabilidade de erro, pois o classificador estará ajustado especificamente para o conjunto de treinamento e a estimativa da probabilidade de erro será muito otimista.

Basicamente essa abordagem possui dois problemas. O primeiro é que o erro de classificação por si só não pode ser confiavelmente estimado quando a razão entre o tamanho do conjunto de exemplos e o do conjunto de características é pequena. O segundo e principal problema dessa abordagem é que a escolha de um classificador é um problema por si só, e o subconjunto selecionado ao final claramente depende do classificador [13].

5.2.2 Distâncias entre Classes

Visando otimizar o conjunto de características para minimizar a probabilidade de erro de qualquer classificador, deve-se maximizar a distância entre padrões de classes diferentes no espaço de características.

Quando se dispõe de um conjunto de amostras treinamento de cada classe, pode-se supor que tal conjunto possui uma boa representação das mesmas e estimar a distância entre as classes.

Nesse caso, há várias formas de se medir a distância entre conjuntos de classes diferentes no espaço de características. Dentre elas pode-se citar [12, 55]:

- **distância entre os centróides das classes**: em que se deve primeiramente determinar os centróides das classes, posteriormente medir a distância entre eles e finalmente somar os resultados;
- **distância entre vizinhos mais próximos, mais distantes e média**: consideram-se, respectivamente, o mínimo, o máximo ou a média das distâncias entre pares de elementos de classes diferentes;
- **distâncias baseadas nas matrizes de espalhamento**: utilizam medidas de separabilidade baseadas em análise de discriminante;
- **distância de Mahalanobis**: trata-se de uma medida que considera a distribuição estatística das classes, bem como a distância espacial entre elas;
- **distância de Bhattacharyya e divergência**: trata-se de distâncias baseadas nas funções “densidade de probabilidade das classes”, a distância espacial não é considerada.
- **distâncias nebulosas**: são medidas de distância que utilizam informações obtidas a partir da *fuzzyficação* dos conjuntos, como os suportes dos conjuntos

e os coeficientes de pertinência dos padrões. Em [56], há uma revisão bastante completa de distâncias nebulosas aplicadas a processamento de imagens. Em [57] foi utilizada uma distância nebulosa como função critério de um algoritmo de seleção de características. Os resultados obtidos com essa abordagem estão descritos na seção 7.2.2.

Maiores detalhes e informações sobre outras medidas de distância podem ser encontrados em [55, 12, 14, 56].

6 Classificação de Padrões

Como os métodos de classificação não são os principais componentes deste trabalho, nesta seção serão citadas apenas informações gerais sobre alguns dos métodos mais conhecidos de classificação e de super-classificação (subseção 6.1). Uma atenção maior será dada aos classificadores de mínima distância ao protótipo e de k vizinhos mais próximos nas seções 6.2.1 e 6.2.2, uma vez que esses foram utilizados nos testes até então realizados.

6.1 Visão Geral

Conforme dito na seção 3.1, classificador é o elemento que possui como entrada um padrão e, como saída, o rótulo que identifica a que classe tal padrão pertence. Ou seja, são os elementos que, de fato, fazem o reconhecimento de padrões. Com exceção de alguns classificadores sintáticos e baseados em lógica fuzzy, eles devem ser treinados utilizando um conjunto de amostras. Dentre os métodos de reconhecimento de padrões melhor conhecidos, pode-se destacar:

- **Estatísticos**: mínima distância ao protótipo, k-vizinhos mais próximos (vide seção 6.2.1 e 6.2.2), classificador de Bayes e outros.
- **Não-estatísticos**: redes neurais, correlação, classificadores sintáticos, árvore de decisão e baseados em lógica fuzzy (vide seção 3.1).
- **Super-classificadores**: também conhecidos como "métodos de combinação", são utilizados quanto se dispõe de vários resultados de classificação e deseja-se combinar os resultados para decidir a qual classe os dados pertencem. Tais esquemas podem ser aplicados quando são utilizados sistemas de multi-sensores e vários classificadores diferentes para classificar um determinado conjunto de dados, ou quando vários padrões separados formam um conjunto que pode pertencer à mesma classe. Um exemplo desse caso é o de seqüências de vídeo. Dentre os super-classificadores mais conhecidos estão: soma, média e mediana dos resultados e o esquema de votação [13].

.

Maiores detalhes a respeito de classificadores podem ser encontrados em [13, 12, 14].

6.2 Classificadores utilizados

Nesta seção serão traçados comentários a respeito dos classificadores que foram utilizados nos testes realizados até o presente momento. São eles: distância ao protótipo e k-vizinhos mais próximos. Uma característica comum a ambos é a facilidade de implementação, mas há grandes diferenças entre eles, tanto no algoritmo quanto em suas vantagens e desvantagens.

6.2.1 Distância ao Protótipo

O classificador de distância ao protótipo é bastante simples em termos de esforço computacional, tanto na fase de treinamento quanto na de teste. Essa característica se deve à simplicidade de seu algoritmo. A fase de treinamento consiste na determinação dos protótipos, um para cada classe. Um exemplo de protótipo é a média (baricentro) das classes. Na fase de teste, cada padrão é classificado de acordo com o protótipo mais próximo.

Assim, para cada padrão de teste, é necessário realizar apenas $C - 1$ comparações, sendo C o número de classes existentes. A desvantagem dessa abordagem é a qualidade dos resultados, pois os protótipos não contém informações suficientes sobre a forma da distribuição das classes.

6.2.2 Regra dos K-vizinhos mais próximos

Trata-se de um método de classificação que não possui processamento na fase de treinamento. Para cada padrão de teste, é medida a distância entre ele e todos os padrões de treinamento. Verifica-se a quais classes pertencem os K padrões mais próximos e a classificação é feita associando-se o padrão de teste à classe que for predominante.

Tomando-se $k = 1$, obtém-se o classificador de vizinho mais próximo, que é muito comum em aplicações de reconhecimento de faces após a extração de características usando PCA.

A principal vantagem desse método é que ele cria uma superfície de decisão que se adapta à forma de distribuição dos dados de treinamento de forma detalhada, possibilitando a obtenção de boas taxas de acerto quando o conjunto de treinamento é grande ou representa todas as variações possíveis dos dados, principalmente no caso de $k = 1$. Valores maiores de k são promovem bons resultados quando a distribuição dos conjuntos possui poucas sobreposições, e quando, por exemplo, um padrão de treinamento da uma classe A cercado por vários da classe B pode ser tratado como uma excessão, e os padrões de teste que ficarem região podem ser classificados como elementos da classe B.

A principal desvantagem está na complexidade do teste, visto que, para cada padrão de teste, além de ser necessário realizar P medições de distância, sendo P o número de padrões de treinamento, é necessário ordenar os resultados.

7 Primeiros resultados

Conforme dito anteriormente, dentre os objetivos práticos finais desse trabalho está o de possibilitar a criação um método de reconhecimento de pessoas em tempo real a partir de seqüencias de vídeo com poucas restrições em relação a iluminação e movimentos das pessoas. Dentre as características importantes de um sistema desse tipo encontram-se:

- necessidade de um sistema de extração de características e classificação rápido e barato, para que não ocorra perda das *deadlines*, que, nesse caso podem ser determinadas pela taxa de aquisição de quadros por segundo (para classificação de cada quadro) e pelo tempo exigido para que seja dada uma resposta ao usuário [58];
- possibilidade de haver muitas imagens por pessoa tanto para treinamento quanto nos testes [2]. Porém, dependendo do classificador utilizado, é desejável utilizar poucos exemplos de treinamento para permitir a obtenção de resultados em tempo real. Já a disponibilidade de muitas imagens por pessoa na fase de testes permite o uso de métodos de super-classificação para obter melhores resultados.
- falta de controle com relação a iluminação, escala e orientação da face. [36]

Visando obter um sistema rápido e que não tenha um grande custo em relação a memória, é desejável que a dimensionalidade dos dados não seja grande. Além disso, há outras vantagens de efetuar redução de dimensionalidade que foram comentadas na seção 3.1.

Basicamente, foram estudadas duas abordagens para redução da complexidade dos dados de um reconhecedor de faces. A primeira, e mais óbvia, é a de reduzir a dimensionalidade dos dados observados através da simples utilização de uma janela na imagem. A segunda abordagem testada consiste na aplicação de um algoritmo de seleção de características, selecionando somente os atributos com maior poder de discriminação das classes. Tais testes estão descritos a seguir.

7.1 Uso de regiões menores da imagem

Os métodos baseados em Análise dos Componentes Principais (PCA) estão entre os que possibilitam a obtenção dos melhores resultados em termos de reconhecimento de faces frontais. Apesar da qualidade dos resultados obtidos, conforme dito anteriormente, essa técnica tem a desvantagem de ser um tanto cara computacionalmente, pois todos os pixels da imagem são utilizados para se obter sua representação em função da covariância entre essa imagem e todas as outras imagens da base de dados.

Alguns pesquisadores utilizaram *eigenfaces* e *eigenfeatures* para efetuar o reconhecimento. O termo *eigenfeature* foi criado por Baback em [31] e refere-se à utilização de PCA em regiões restritas da imagem, obtendo os componentes principais de pontos característicos da face, como boca (*eigenmouth*), nariz (*eigennose*) e olhos (*eigeneyes*). Segundo Moghddam e Pentland [31], estudos de movimentos dos olhos indicam que essas regiões particulares das faces representam marcas importantes para reconhecimento, especialmente em uma tarefa de tentativa de discriminação para identificação de pessoas.

Em [23] foram obtidos resultados interessantes, em que surpreendentemente os resultados alcançado com a utilização de um quadro (*template*) abrangendo somente a região dos olhos foram melhores que os resultados com um quadro que cobria toda a face.

No trabalho de Baback [31], os resultados obtidos com *eigenfeatures*, que incluíram olhos, nariz e boca, foram melhores que o de *eigenfaces*.

Além desses fatores, em [31] discute-se de uma vantagem potencial do uso de regiões características das faces. Trata-se da eliminação da possibilidade de ocorrência de erros provocados pelo uso ou não de barba, bigode, chapéu, variações no comprimento do cabelo, presença de feridas na face, etc. Esses elementos podem prejudicar o desempenho quando utiliza-se a imagem de toda a face, mas não quando forem utilizadas somente as regiões características.

Neste trabalho de mestrado, foram realizados testes visando verificar os resultados de Brunelli em um sistema de reconhecimento baseado em PCA, mas comparando apenas o desempenho do classificador com imagens de faces versus com imagens contendo os olhos. Esse trabalho foi publicado em [18]. A base de imagens que foi utilizada para esses testes é composta por imagens de dezesseis homens adultos, seis imagens por pessoa. Várias imagens continham pessoas usando óculos, bigode ou barba e com diferentes comprimentos de cabelo. Além disso, as imagens possuem grandes variações na iluminação, escala e na orientação da face. Porém as imagens consideradas não possuem problemas de auto-oclusão dos olhos. Exemplos de imagens de uma pessoa podem ser vistos na figura 11.

Figura 11: Exemplo de imagens de um indivíduo da base utilizada.

Dois testes foram realizados: no primeiro utilizando três imagens por pessoa para treinar o sistema e, no segundo, cinco. Em ambos os testes foi utilizada apenas uma imagem de teste por pessoa diferente das de treinamento.

As imagens usadas para construir as *eigenfaces* foram criadas a partir de recortes da base original para que os cabelos e o fundo da imagem não influenciassem no reconhecimento, pois esses podem apresentar muitas variações. Tais recortes englobavam a região entre a testa e o queixo dos indivíduos. Já para a construção dos *eigeneyes*, foram utilizados recortes que englobam somente a região os dois olhos. O tamanho desses recortes foi determinado de acordo com uma proporção baseada na distância entre os olhos. A figura 12 mostra um exemplo desses recortes.

A determinação da posição dos olhos foi feita por um operador humano, pois segmentação não faz parte do escopo deste trabalho. Nos próximos experimentos, pretende-se utilizar os resultados de um sistema automático de detecção de pontos característicos das faces, conforme descrito na seção 8.

Como as imagens apresentavam grandes variações na orientação da cabeça e na escala, antes de efetuar os recortes das faces e dos olhos, foi realizada a rotação das

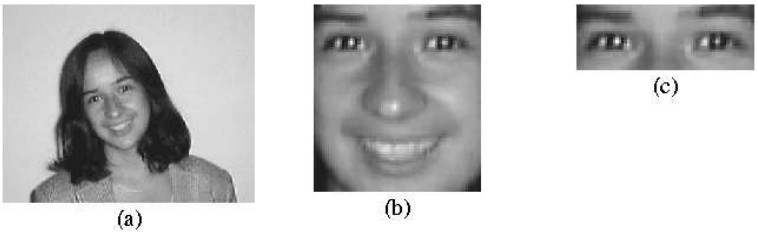

Figura 12: Processo de obtenção das imagens de face e de olhos: (a) imagem original, de 512×342 pixels ; (b) recorte de face; (c) recorte de olhos.

imagens fazendo com que os olhos ficassem na mesma linha horizontal. Após realizar os recortes, para viabilizar o uso de PCA, foi necessário redimensionar as imagens para que todas ficassem com a mesma resolução. Foi efetuado o redimensionamento utilizando o método de "vizinho mais próximo" para determinar o valor dos *pixels* da imagem de saída.

A resolução escolhida foi de 64×64 pixels, pois essa engloba faces mesmo nas imagens em que a pessoa está mais afastada da câmera. Além disso, essa é uma resolução que equilibra custo computacional com qualidade das imagens, já que é desejável utilizar as menores imagens possíveis mas sem perder muitos detalhes.

O pré-processamento descrito acima foi realizado em todas as imagens da base, criando assim, um conjunto de imagens de olhos e outro de faces. Posteriormente as imagens de treinamento da base de faces são utilizadas para treinar uma transformada PCA, obtendo-se, dessa forma, os *eigenfaces*. O mesmo é feito para as imagens de olhos possibilitando a obtenção dos *eigeneyes*. Alguns *eigeneyes* e *eigenfaces* obtidos a partir de uma base treinada com 5 imagens por pessoa são mostrados na figura 13. A seção 4.2 contém maiores detalhes a respeito da transformada PCA.

Os resultados obtidos são bastante satisfatórios como meio de comparação entre faces e olhos para reconhecimento de pessoas. Em todos os testes realizados com imagens que pertenciam ao conjunto de treinamento, a taxa de acerto foi de 100%. Os resultados dos testes realizados com imagens que não pertenciam ao conjunto de treinamento estão ilustrados nas tabelas 1 (com treinamento usando 3 imagens por pessoa) e 2 (com treinamento usando 5 imagens por pessoa). O classificador utilizado foi o de vizinho mais próximo.

Através desses resultados, pode-se notar que, em geral, o reconhecimento com

# auto-vetores	Olhos %	Faces %
3	25,00	31,25
4	25,00	37,50
5	50,00	37,50
10	56,25	43,75
13	62,50	43,75
15	62,50	43,75
24	62,50	43,75
48	62,50	43,75

Tabela 1: Desempenho do classificador para reconhecimento de olhos e de faces quando treinado com 3 elementos por pessoa.

# auto-vetores	Olhos %	Faces %
3	40,00	46,67
15	73,33	66,67

Tabela 2: Desempenho do classificador para reconhecimento de olhos e de faces quando treinado com 5 elementos por pessoa.

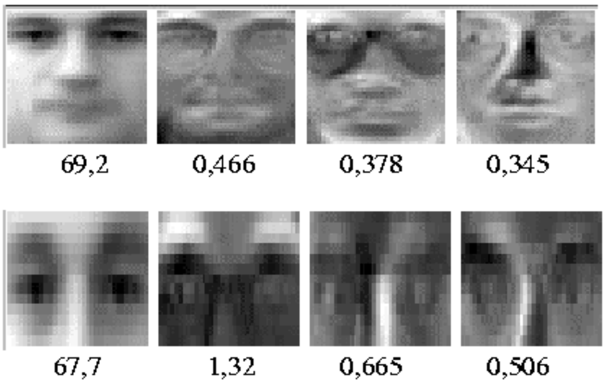

Figura 13: Os quatro primeiros auto-vetores mostrados como imagens e seus respectivos auto-valores, obtidos através da base de faces (acima) e da base de olhos (abaixo)

olhos é melhor que com faces. Esse fato é aparentemente inesperado, já que as imagens de faces contém mais informações que as de olhos. Mas, devido à *curse of dimensionality*, sabe-se que o aumento na dimensionalidade dos dados deve ser compensado por um aumento exponencial do número de exemplos de treinamento para que a taxa de acerto permaneça estável. Isso justifica o fato de que as taxas de reconhecimento aumentam significativamente para ambos os sistemas de classificação quando aumenta-se o tamanho do conjunto de treinamento. Além disso, pode-se notar que, quando treinado com 3 imagens por pessoa, o desempenho do sistema não melhora quando são utilizados mais que 13 auto-vetores, pois 13 é o número suficiente para discriminar os padrões e o restante não adiciona informações relevantes para a classificação. Maiores detalhes sobre *curse of dimensionality* estão na seção 3.1.3.

Outra razão para a queda do desempenho do sistema que utiliza imagens de toda a face está no fato de que a inclusão das regiões do nariz e da boca podem reduzir o desempenho do classificador, pois a boca não é um objeto rígido como os olhos, e a forma do nariz sofre grandes alterações com variações na orientação da cabeça. Por esses fatores, as imagens de faces são mais distorcidas, o que causa uma maior dificuldade em obter boas taxas de reconhecimento usando tais imagens com um conjunto de treinamento pequeno, já que o nariz e a boca são características que podem ser muito correlacionadas e ruidosas.

Assim, para possibilitar a obtenção de boas taxas de acerto utilizando imagens de toda a face, dever-se-ia aumentar o tamanho do conjunto de treinamento [18].

7.2 Testes com Seleção de características

Foram estudados e testados alguns métodos de seleção de características sob dois aspectos: o algoritmo de seleção e a função critério. Tais trabalhos serão descritos, respectivamente, nas seções 7.2.1 e 7.2.2.

7.2.1 Teste de algoritmos

Em [9], nós fizemos um estudo comparando o desempenho de quatro formas de seleção de características, dos quais dois baseados em busca automática, para discriminar classes de padrões obtidos a partir de descritores de Fourier. O problema abordado em [9] trata de discriminação de imagens contendo faces versus imagens não contendo faces de uma forma rápida. O método de discriminação, que nós propusemos em [59] constitui-se de:

- obtenção do mapa de bordas horizontais binário da imagem através do Laplaciano de uma janela Gaussiana unidimensional vertical [44];
- formação de um sinal unidimensional a partir de uma "varredura" vertical do mapa de bordas obtido;
- obtenção de 30 descritores de Fourier [44, 45] desse sinal unidimensional.

Para efetuar a seleção automática de características (no caso, descritore de Fourier), os únicos métodos de busca testados foram os algoritmos SFS e o ASFS (Métodos de Busca Seqüencial Flutuante e suas versões Adaptativas). Isso se justifica pelo fato desses serem os melhores algoritmos de busca para seleção de característica (vide seção 5). Os resultados obtidos foram avaliados para determinar o melhor seletor de características e a melhor dimensionalidade. A arquitetura desse processo está ilustrada na figura 14.

Esse processo foi realizado em 219 imagens de faces e em 219 imagens de outros objetos (não-faces). Desse total, 2/3 foi utilizado para treinar um classificador de mínima distância ao protótipo, e o restante para testá-lo.

As quatro técnicas de seleção de características testadas foram:

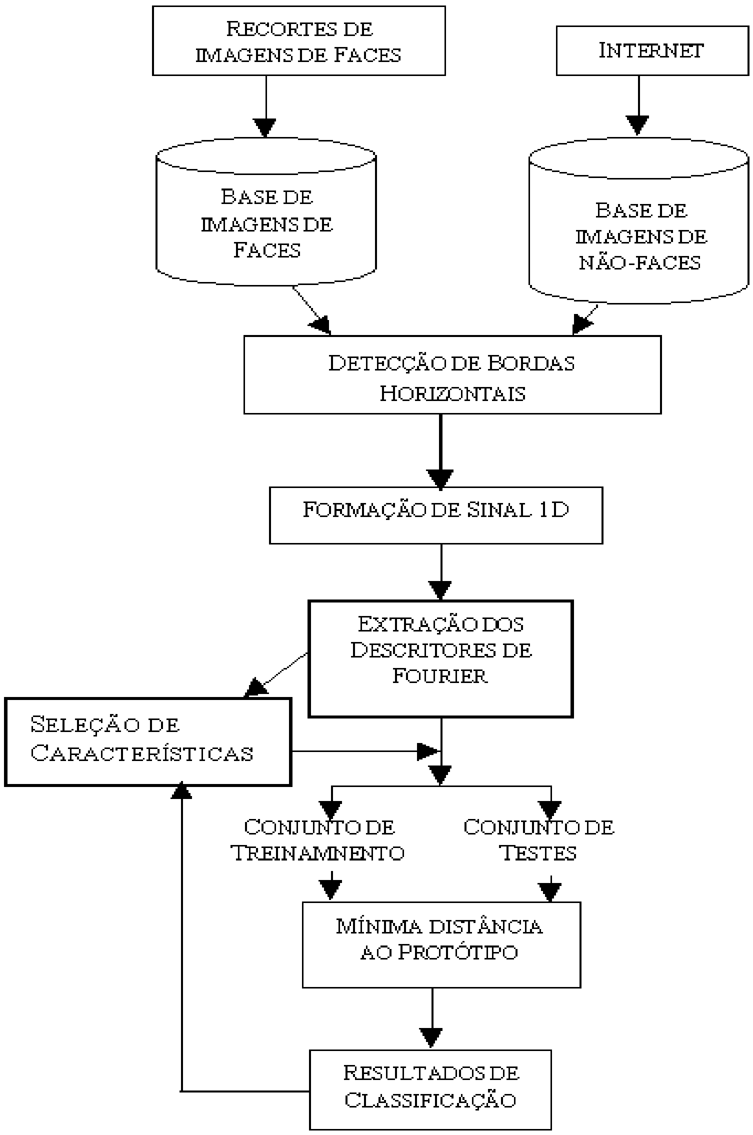

Figura 14: Esquema do sistema de discriminação faces $\times$ não-faces.

- utilização dos d primeiros coeficientes (abordagem mais comum em se tratando de coeficientes de Fourier);
- utilização dos d maiores coeficientes (comum em Wavelets);
- métodos de seleção SFS (vide seção 5);
- métodos de seleção ASFS (vide seção 5).

Para os métodos de seleção de características SFS e ASFS, foi usado, como função critério, o desempenho de um classificador de mínima distância ao protótipo. É importante notar que o número de padrões de treinamento é de 146 faces e 146 não-faces, enquanto que o número de classes é apenas 2 (face e não-face). Com isso, o uso do desempenho do classificador como função critério para seleção de características é justificado (vide seção 5). Através da figura 15 pode-se observar os resultados obtidos, que comprovam a superioridade dos métodos de seleção de característica, principalmente o ASFS. É interessante notar também que os melhores resultados foram obtidos utilizando dimensionalidade menor que 20, o que confirma que um aumento no número de características não garante melhora no desempenho do classificador.

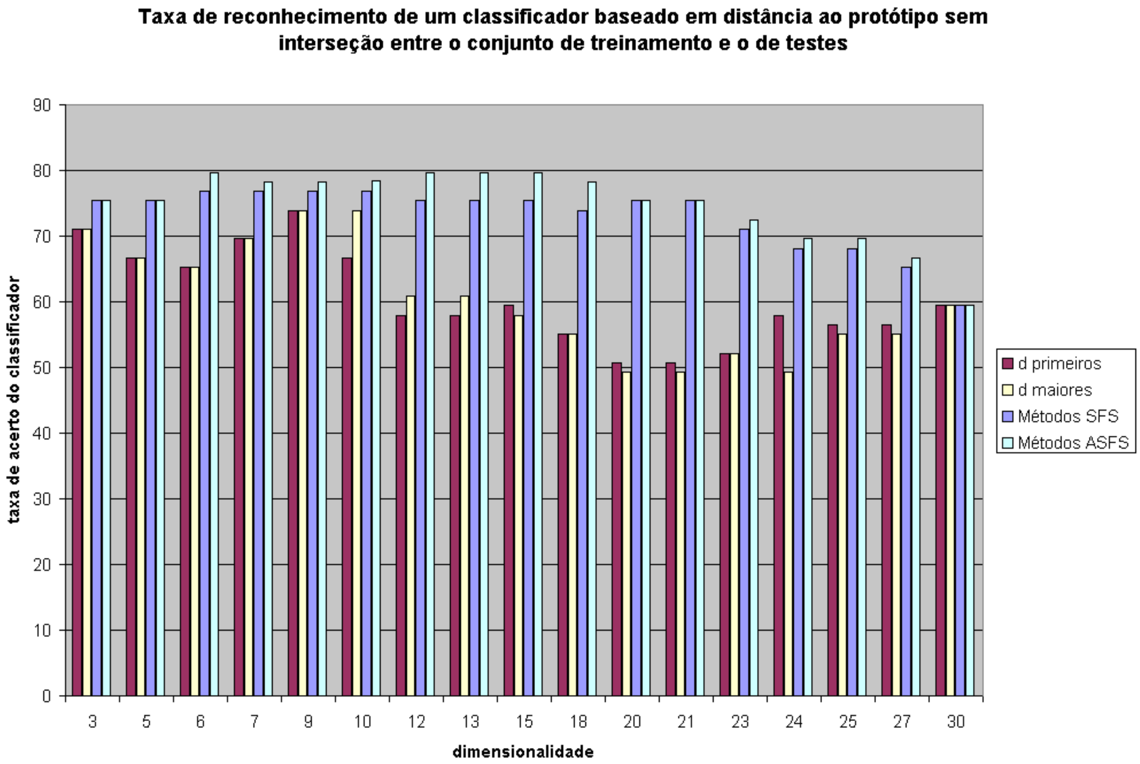

Figura 15: Resultados obtidos (em % de taxa de acerto do classificador) pelos conjuntos de características selecionados.

No pior caso, a maior diferença entre o resultados de SFS e ASFS foi de 4,35%. Mas, em termos de tempo de execução, no pior caso o algoritmo SFS levou 2 segundos, enquanto o ASFS levou 4 horas e 22 minutos para determinar o conjunto de características. Esse é um fator muito relevante na escolha do algoritmo de seleção.

É importante lembrar que o total de características disponíveis é 30 e, por isso, todos os métodos "convergem" para o mesmo resultado quando a dimensionalidade vai para 30. Detalhes sobre esse trabalho podem ser vistos em [57].

7.2.2 Função Critério

As medidas de distância entre agrupamentos ou classes mais utilizadas como função critério para seleção de característica são mais adequadas a conjuntos convexos, tendendo a privilegiar conjuntos linearmente separáveis (por exemplo, a distância de Mahalanobis). O problema é que com esses critérios de distância não é possível detectar bons agrupamentos côncavos ou com médias próximas, como do da figura 9, em que a distribuição dos padrões de uma classe encontra-se no interior da distribuição dos padrões de outra classe, mas as distribuições das classes não se interceptam. Nesse caso, dificilmente um critério de distância tradicional identificaria o potencial desses agrupamentos, apesar do fato dos dois agrupamentos estarem bem definidos, possibilitando a obtenção de boas taxas de reconhecimento com um classificador de k-vizinhos mais próximos.

Visando evitar esse problema, foram realizados testes com seleção de características utilizando uma função critério que, juntamente com o algoritmo de seleção, maximiza a distância entre padrões que pertencem a classes diferentes e minimiza a distância entre elementos que pertencem à mesma classe. Isso é feito independentemente da forma da distribuição dos padrões no espaço de características.

Um critério de distância que possui essas propriedades é a distância fuzzy baseada em tolerância, proposta em [60]. Nesse critério, a distância é determinada através de uma vizinhança em torno de cada padrão de treinamento.

Para utilizar esse critério de distância deve-se, inicialmente, transformar os conjuntos de treinamento em conjuntos nebulosos (*fuzzy*). Foram realizados testes com essa função critério e publicados em [57], numa colaboração com a pesquisadora francesa Isabelle Block. A função de *fuzzyficação* utilizada foi a seguinte:

$$\mu_A(x) = \frac{1}{1 - d(x, p_A)}, \tag{9}$$

em que x é um padrão, $\mu_A(x)$ é a função de pertinência desse padrão ao conjunto A, p_A representa o suporte da classe A e $d()$ é a distância Euclidiana. O suporte pode ser definido, por exemplo, como o baricentro do conjunto A, de forma que a função de pertinência é inversamente proporcional à distância do padrão ao suporte de cada classe.

Para definir a distância fuzzy baseada em tolerância, inicialmente é definida uma distância local:

$$d_x^\tau(\mu_A, \mu_B) = \inf_{y,z \in B(x,\tau)} |\mu_A(y) - \mu_B(z)|, \tag{10}$$

em que $B(x, \tau)$ denota uma bola (espacial) fechada, de raio τ centrada em x. Assim, pode-se definir a distância fuzzy baseada em tolerância por [60]:

$$d_p^\tau(\mu_A, \mu_B) = [\int_S [d_x^\tau(\mu_A, \nu_B)]^p dx]^{1/p}, \tag{11}$$

onde S representa todo o espaço de características.

Para avaliar o desempenho desse critério de distância com o intuito de efetuar seleção de características, foram feitos testes com os métodos de busca SFS [54] em dados artificiais. O critério de distância fuzzy baseado em tolerância foi comparado com o desempenho de um classificador de distância ao protótipo. Para avaliar o desempenho dos conjuntos de dados foram utilizados dois classificadores: k-vizinhos mais próximos e o de mínima distância ao protótipo.

Esse algoritmo foi testado com um conjunto de dados artificiais de seis dimensões, duas classes, com 100 exemplos por classe. As figuras 16, 17 e 18 mostram a distribuição para cada par de dimensões. Esse conjunto de dados foi criado para permitir que seja feita seleção de características reduzindo a dimensionalidade para 2, o que permite visualização dos dados.

Antes de realizar a seleção de características, os dados foram normalizados de forma a ficarem com média 0 e variância unitária. Isso é importante para evitar problemas com os classificadores, pois esses são baseados em distância Euclidiana [19].

Para avaliar os resultados, os dados com as distribuições descritas nas figuras 16, 17 e 18, com 100 amostras para cada classe, foram gerados 100 vezes. A distância foi utilizada com $\tau = 0.5$ e o classificador de k-vizinhos com $k = 3$.

Conforme esperado, o seletor de características baseado em distância nebulosa selecionou as características 5 e 6 em todos os 100 testes. Já o algoritmo de seleção com o desempenho do classificador frequentemente selecionou as caracterís 1 e 2, mas várias outras combinações de características também foram selecionadas.

Quanto ao desempenho obtido com o classificador de k-vizinhos mais próximos, os resultados estão na tabela 3. Quando o classificador k-vizinhos foi testado som

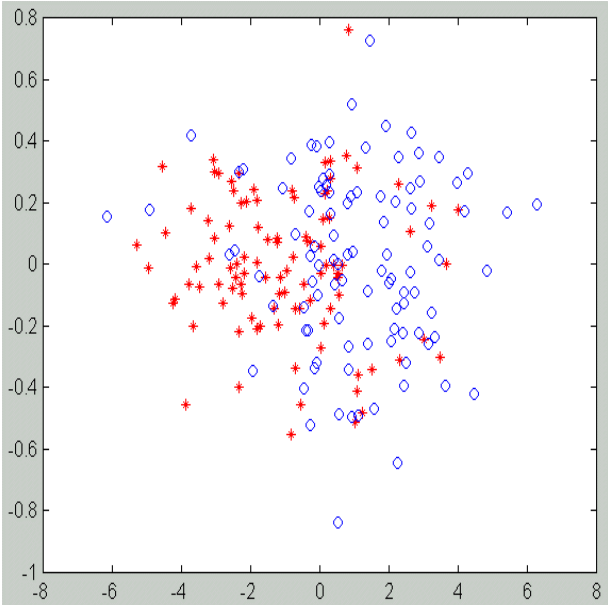

Figura 16: Dados artificiais utilizados em [57]. (*) = classe A, (o) = classe B. Nas características 1 e 2, os padrões possuem distribuições Gaussianas com médias diferentes.

Classificador ——— Critério	Conjunto de treinamento igual ao conjunto de testes	Sem interseção entre o conjunto de treinamento e o de testes
Classificador	95,5611%	89,4706%
Distância nebulosa	99,9950%	95,0735%

Tabela 3: Desempenho médio (após 100 testes) do classificador k-vizinhos para os sub-conjuntos de características selecionados.

interseção entre o conjunto de treinamento e o de testes, foram utilizados 67 padrões para treinar o classificador e 33 para testá-lo.

Os resultados obtidos mostram que a distância nebulosa baseada em tolerância permite a obtenção de bons resultados para conjuntos côncavos ou com conjuntos apresentando sobreposição entre classes diferentes. Maiores detalhes sobre esse trabalho podem ser obtidos em [57].

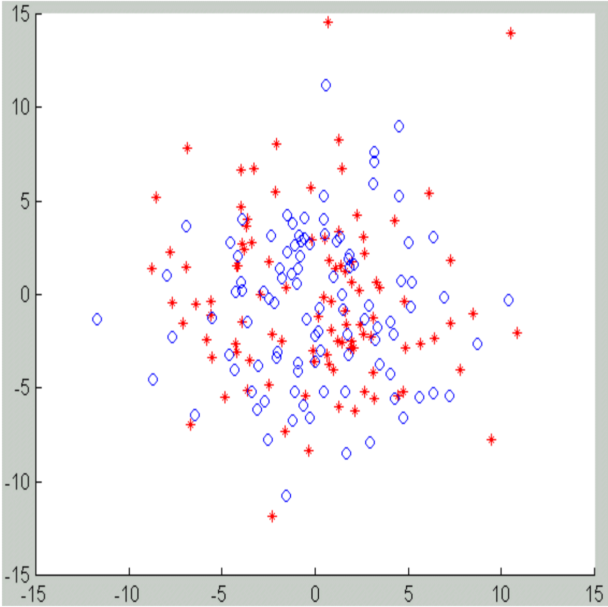

Figura 17: Continuação da figura 16. Distribuições ruidosas nas características 3 e 4.

8 Plano de Trabalho

Esta seção contém o plano de trabalho proposto para ser implementado até março de 2001, quando as implementações e testes deverão ser finalizados e deverá ser iniciada a fase final da redação da dissertação de mestrado e a preparação para a defesa.

8.1 Metas

Os estudos realizados até então caminham para possibilitar a implementação de um projeto de sistema de reconhecimento de pessoas a partir de seqüências de vídeo. O trabalho está sendo desenvolvido nas linhas da seção 7, a qual mostra que já foram feitos vários testes e obtidos resultados preliminares. Pretende-se seguir na direção de obter resultados mais conclusivos e estatisticamente mais relevantes.

Também faz parte do plano de trabalho aprofundar os estudos visando entender melhor a formalização dos métodos, fazer outras simulações, etc. Além disso, pretende-se realizar testes com seqüências de vídeo. Os objetivos principais a serem salientados são:

- realizar mais testes com uso de regiões menores da imagem;

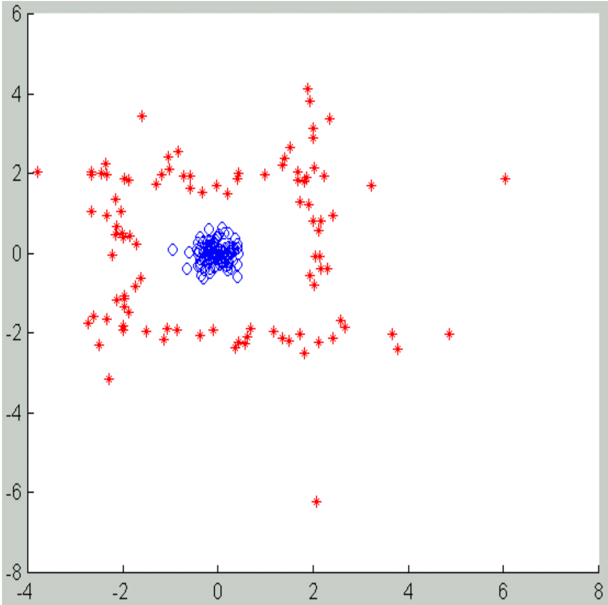

Figura 18: Continuação da figura 17. Nas características 5 e 6, a classe B possui distribuição Gaussiana "dentro" da classe A, a qual é gerada como uma mistura de 4 Gaussianas formando um anel.

- aprofundar o estudo de seleção de características;
- aplicar as técnicas desenvolvidas ao problema de reconhecimento a partir de seqüências de vídeo;
- estudar o estado-da-arte de reconhecimento de faces [3] e realizar testes com algum método promissor de reconhecimento de faces que possa ser associado a seleção de características.

Mais especificamente, pretende-se tentar realizar avanços nos seguintes tópicos:

- Seleção de características usando lógica nebulosa. O critério de distância nebulosa baseada em tolerância poderá ser melhorado através a utilização da versão estendida dessa medida de distância, a qual foi proposta em [61]. Trata-se de um algoritmo que calcula dinamicamente o tamanho bola $B(x, \tau)$ da equação 10. Outros avanços que poderão ser realizados nessa parte do projeto é o de implementar e testar outras formas de *fuzzyficação* dos conjuntos, diferentes da equação 9.
- Pretende-se estudar a possibilidade de usar, para classificar, informações relevantes que podem ser extraídas do movimento. Para tal, pretende-se iniciar

os estudos com base em [62, 36].

8.2 Projeto para Reconhecimento a partir de Seqüências de Vídeo

Esta seção contém a proposta de um algoritmo que relaciona todos os tópicos que estão sendo estudados e desenvolvidos para a conclusão do projeto de mestrado. O fluxograma dessa proposta pode ser visto na figura 19 e foi documentado em [63].

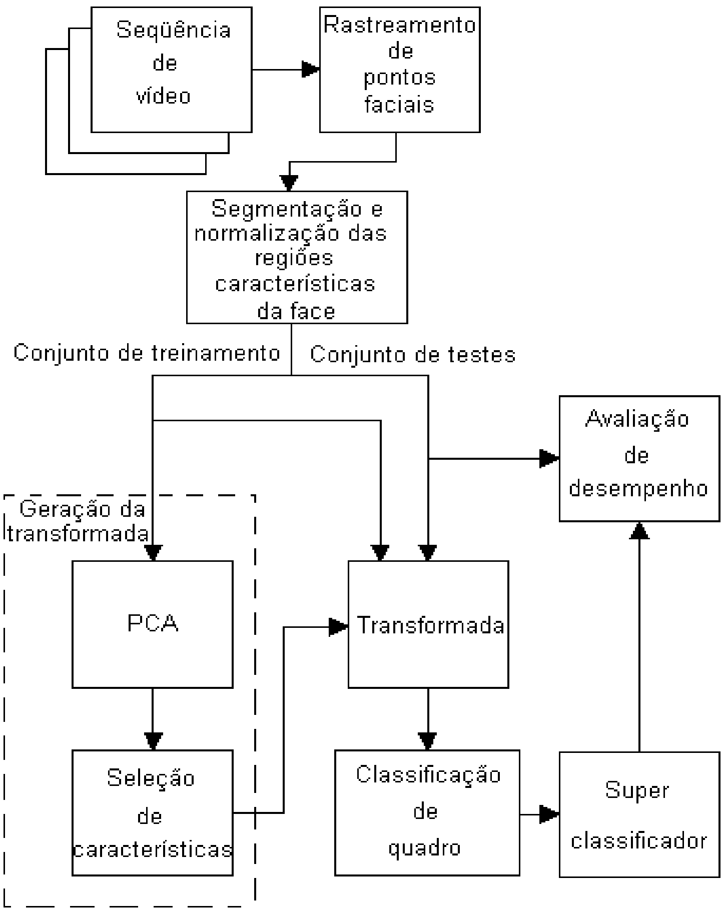

Figura 19: Esquema do projeto de reconhecimento a partir de seqüências de vídeo.

Basicamente, esse projeto foi criado a partir da união das duas idéias para redução da complexidade discutidas na seção 7, ou seja, redução da dimensionalidade através da redução da imagem de entrada e de seleção de características.

Assim, o sistema de reconhecimento utilizará quatro recortes da imagem de entrada: para os dois olhos, o nariz e a boca. As tarefas de detecção e perseguição de pontos característicos da face, bem como a de normalização das imagens de olhos, nariz e boca não fazem parte do escopo deste projeto. Mas é importante ressalvar que essas tarefas serão realizadas através de um método baseado em Gabor Wavelet Networks [10]. Esse método detecta e persegue os pontos característicos e determina os parâmetros da transformação afim que leva uma imagem frontal e a uma determinada escala à posição em que os pontos se encontram. Através desses parâmetros, pode-se realizar a inversa da transformação afim e obter imagens normalizadas. Esse processo de normalização é importante para reduzir as variações dos padrões introduzidas pelos movimentos da face, o que melhora o desempenho do sistema de reconhecimento.

Assim, as imagens utilizadas tanto para treinar quanto para testar o sistema de reconhecimento serão constituídas por imagens das regiões características normalizadas com relação à transformação afim.

Para efetuar o treinamento, será utilizada uma seqüência de vídeo por pessoa. O reconhecimento será feito utilizando análise de componentes principais (PCA), com uma base para cada característica. Alguns detalhes sobre PCA encontram-se na seção 4.2. Assim, será criada uma base para olhos esquerdos, outra para olhos direitos, uma para os narizes e outra para as bocas, obtendo-se as *eigenfeatures* (*eigenlefteyes, eigenrighteyes, eigennoses* e *eigenmouth*). Detalhes sobre os termos recém-citados encontram-se na seção 7.1.

Após a obtenção de todas as *eigenfeatures*, essas são concatenadas de forma a criar um espaço de características que consiste na soma dos espaços vetoriais formados por todas as *eigenfeatures* (detalhes sobre soma de espaços vetoriais podem ser encontrados em [47]). Para reduzir a dimensionalidade desse espaço, será aplicado o algoritmo de se seleção de características descrito na seção 7.2.2. A figura 20 esquematiza o método descrito.

A utilização de um algoritmo de seleção de características é motivada por poder-se dizer que essa abordagem constitui-se de um tipo de fusão de multisensores, se considerarmos que cada região característica da face é obtida através de um meio de aquisição diferente. Nesse ponto surge a necessidade de reduzir a dimensionalidade de maneira a valorizar os sensores com maior poder de discriminação. Além

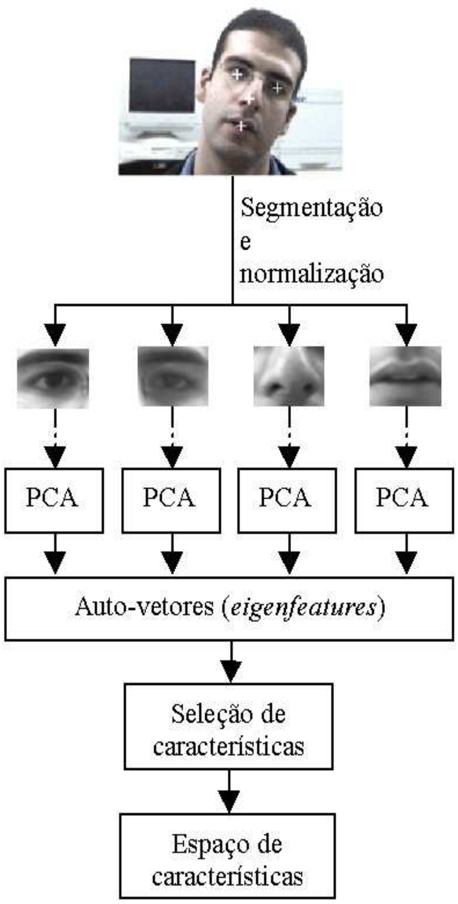

Figura 20: Geração do espaço de características.

disso, como pode-se concluir da seção 7.2.1, a aplicação de algoritmos de seleção de características pode proporcionar melhora na taxa de acerto de classificadores.

Além disso, conforme dito anteriormente, a transformada PCA faz uma rotação no espaço de características de forma que o primeiro vetor da base fique na direção em que há mais variação entre os padrões, o segundo vetor na direção em que ocorre a segunda maior variação perpendicular ao primeiro, e assim por diante. Ou seja, a variação específica entre elementos de classe diferente não é otimizada.

Em [13], os autores mostram os resultados de uma abordagem de reconhecimento parecida com a abordagem proposta aqui. Trata-se da aplicação de seleção de características usando a técnica de busca flutuante (métodos SFS) sobre as características obtidas a partir da transformada PCA sobre imagens de dígitos.

Outro fator motivador para a aplicação de seleção de características sobre PCA está em um dos resultados obtidos em [19], em que o desempenho de um sistema de reconhecimento de pessoas baseado em PCA foi melhorado com a eliminação dos três primeiros auto-vetores. Os autores de [19] justificam que há algumas evidências de que esses auto-vetores são influenciados pelas mudanças de iluminação, e não por variações inter-classes. Provavelmente, esse fato se deve principalmente ao fato de que em [19] foram realizados testes com imagens apresentando grandes variações de iluminação, e os primeiros auto-vetores apontam para o sentido em que há maior variação dos dados. Esse resultado provem evidencias de que é possível obter resultados melhores aplicando um método de seleção de características sobre as *eigenfeatures* ao invés de utilizar simplesmente os primeiros auto-vetores.

Em [31] os autores declararam que não estava definida uma estratégia de realizar fusão ótima das informações obtidas das diferentes regiões da face. Tanto em [31], como em [23], foi utilizado um classificador para cada região da face e os resultados foram combinados utilizando um super-classificador. No caso de [23], a classificação das regiões foi feita usando *template matching* e a combinação foi realizada através da soma dos resultados (graus de similaridade dos templates de cada pessoa). Já em [31], a classificação das regiões foi feita por vizinho mais próximo no *eigenspace* e a combinação, através do esquema de votação. A estrutura proposta aqui (seleção de *eigenfeatures*) é uma forma de fundir os dados para a utilização de um único classificador para todas as regiões das imagens. Um super-classificador será utilizado somente para combinar os resultados de classificação de cada quadro da seqüência de vídeo.

Para efetuar o reconhecimento de pessoas em seqüências de vídeo, primeiro os quadros deverão ser representados no espaço de características criado a partir de *eigenfeature selection*. Inicialmente o espaço de características deverá ser povoado pelos elementos de treinamento obtidos a partir de seqüências de vídeo em que as pessoas variem a pose e a expressão facial. Dessa forma, cada classe poderá ter muitos elementos de treinamento. Posteriormente, para cada pessoa, uma outra seqüência de vídeo será utilizada para testar o sistema. Cada quadro das seqüências de teste será classificado individualmente através de um classificador de mínima distância ao protótipo ou o k-vizinhos mais próximos (descritos na seção 6). Conforme mensionado anteriormente, um super-classificador (ou método de combinação) será utilizado para decidir o resultado da classificação a partir dos resultados obtidos pelos quadros individuais da seqüência. Pretende-se utilizar o esquema de votação para efetuar essa tarefa, mas poderão ser utilizados outros. Alguns super classificadores foram descritos e avaliados em [13].

Esse trabalho incluirá avaliação de desempenho do classificador através de uma estratégia parecida com a "deixe um fora" (*leaving-one-out*) [14]. Dado que se dispões de n amostras por classe, a estratégia *leaving-one-out* consiste de treinar o classificador com $n - 1$ padrões por classe e testá-lo com uma amostra, e alternar o elemento de teste, de forma que todos os elementos sejam testados a partir do treinamento dos restantes. Nossa proposta é de dispor de algumas seqüências de vídeo para cada pessoa e realizar testes alternando seqüências para testar e para treinar o classificador.

Através dessa avaliação poderá ser determinado qual dos classificadores de quadros será utilizado (K-vizinhos ou distância ao protótipo), bem como o valor K, caso o classificador de k-vizinhos mais próximo seja escolhido. Também poderão ser testadas várias estratégias de super classificação.

Uma outra aplicação possível para esse trabalho se refere ao reconhecimento de expressões faciais. Para isso basta realizar o mesmo processo, com a diferença de que as classes consideradas nas imagens de treinamento serão classes de expressões faciais, e não identificação de pessoas.

Referências

[1] Sharath Pankanti, Rund M. Bolle, and Anil Jain. Biometrics: The future of identification. *Computer*, pages 46–49, February 2000.

[2] Rama Chellappa, Charles L. Wilson, and Saad Sirohey. Human and machine recognition of faces: A survey. *Proceedings of the IEEE*, 83(5):703–740, May 1995.

[3] Alex Pentland. Looking at people: Sensing for ubiquitous and wearable computing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):107–119, January 2000.

[4] Martin Bichsel, editor. *First International Conference on Face and Gesture Recognition*. MultiMedia Laboratory Department of Computer Science university of Zurich, Zurich, Switzerland, June 1995. Proceedings.

[5] I. Essa, editor. *Second IEEE International Conference on Face and Gesture Recognition*. IEEE Computer Society Press, Killington, USA, October 1996. Proceedings.

[6] Masahiko Yachida, editor. *Third IEEE International Conference on Face and Gesture Recognition*. IEEE Computer Society Press, Nara, Japan, April 1998. Proceedings.

[7] James L. Crowley, editor. *Forth IEEE International Conference on Face and Gesture Recognition*. IEEE Computer Society Press, Grenoble, France, March 2000. Proceedings.

[8] Rangachar Kasturi, editor. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. IEEE Computer Society Press, July 1997. Theme Section of the Journal - Face and Gesture Recognition.

[9] Teófilo Emídio de Campos, Rogério Schmidt Feris, and Roberto Marcondes Cesar Junior. Improved face $\times$ non-face discrimination using fourier descriptors through feature selection. In *Proceedings of 13th SIBGRAPI*. IEEE Computer Society Press, October 2000.

[10] Rogério Schmidt Feris and Roberto Marcondes Cesar Junior. Detection and tracking of facial landmarks using gabor wavelet networks. In Sameer Singh,

Nabeel Murshed, and Walter Kropatsch, editors, *Lecture Notes in Computer Science*, Rio de Janeiro, Brasil, March 2001. The International Association for Pattern Recognition, Springer-Verlag Press.

[11] Rogério Schmidt Feris, Teófilo Emídio de Campos, and Roberto Marcondes Cesar Junior. Detection and tracking of facial features in video sequences. In *Lecture Notes in Artificial Intelligence*, volume 1973, pages 129–137, Acapulco, Mexico, April 2000. The Mexican Society for Artificial Intelligence and Acapulco Institute of Technology, Springer-Verlag Press.

[12] Sergios Theodoridis. *Pattern Recognition.* Academic Press, USA, 1st edition, 1999.

[13] Anil K. Jain, Robert P. W. Duin, and Jianchang Mao. Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):4–37, January 2000.

[14] Richard O. Duda and Peter E. Hart. *Pattern Classification and Scene Analysis.* Wiley-Interscience, USA, 1st edition, 1973.

[15] Anil K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: a review. *ACM Computing Surveys*, 31(3):264–323, January 1999.

[16] Carlos Hitoshi Morimoto, Yaser Yacoob, and Larry Davis. Recognition of head gestures using hidden markov models. In *ICPR*, Vienna, Austria, August 1996.

[17] Didier Dubois, Henri Prade, and Ronald R. Yager, editors. *Fuzzy Information Engineering.* Wiley Computer Publishing, USA, 1997.

[18] Teófilo Emídio Campos, Rogério Schmidt Feris, and Roberto Marcondes Cesar Jr. Eigenfaces versus eigeneyes: First steps towards performance assessment of representations for face recognition. In *Lecture Notes in Artificial Intelligence*, volume 1973, pages 197–206, Acapulco, Mexico, April 2000. The Mexican Society for Artificial Intelligence and Acapulco Institute of Technology, Springer-Verlag Press.

[19] Peter N. Belhumeur, João P. Hespanha, and David J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Tran-*

sactions on Pattern Analysis and Machine Intelligence, 19(7):711–720, July 1997.

[20] S. Watanabe. *Pattern Recognition: Human and Mecanical.* New York: Wiley, 1985.

[21] Matthew Turk, editor. *1998 Workshop on Perceptual User Interfaces.* Microsoft - http://research.microsoft.com/PUIWorkshop, San Francisco, USA, November 1998. Proceedings.

[22] Ingemar J. Cox, Joumana Ghosn, and Peter N. Yianilos. Feature-based recognition using mixture-distance. Technical report, NEC Research Institute, October 1995.

[23] R. Brunelli and T. Poggio. Face recognition: Features versus templates. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 15(10):1042–1052, October 1993.

[24] Mathew A. Turk and Alex P. Pentland. Face recognition using eigenfaces. In *Proc. of the IEEE con Computer Society Conferece.* IEEE Computer Society Press, 1991.

[25] Steve Lawrence, C. Lee Giles, Ah Chung Tsoi, and Andrew D. Back. Face recognition: A hybrid neural network approach. Technical report, Institute for Advanced Computer Studies, University of Maryland, 1996.

[26] Sami Romdhani. Face recognition using principal component analysis. Master's thesis, Department of Electronics and Electrical Engineering, University of Glasgow, UK, March 1996.

[27] Leurenz Wiskott, Jean-Marc Fellous, Norbert Krüger, and Christoph von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):775–779, July 1997.

[28] Leurenz Wiskott, Jean-Marc Fellous, Norbert Krüger, and Christoph von der Malsburg. Face recognition and gender determination. In *International Conference on Automatic Face and Gesture Recognition*, pages 92–97, University of Zurich, 1995. MultiMedia Laboratory.

[29] Martin Lades, Jan C. Vorbrüggen, Joachim Buhmann, Jörg Lange, Christoph von der Malsburg, Rolf P. Würtz, and Wolfgang Konen. Distortion invariant object recognition in the dynamic link architecture. *Transactions on Computers*, 42(3):300–311, 1993.

[30] Voker Krüger and Gerald Sommer. Gabor wavelet networks for object representation. In *22. DAGM Symposium*, Kiel, Germany, September 2000. accepted for publication.

[31] Baback Moghaddam and Alex Pentland. Face recognition using view-based and modular eigenspaces. In *Automatic Systems for the identification and Inspection of Humans*, volume 2277. SPIE, 1994.

[32] Marco La Cascia and Stam Sclaroff. Fast, reliable head tracking under varying illumination. In *IEEE Conference on Computer Vision and Pattern Recognition*, June 1999.

[33] Yaser Yacoob, Heung-Man Lam, and Larry S. Davis. Recognizing faces showing expressions. In *International Conference on Automatic Face and Gesture Recognition*, Zurich, Swistzerland, 1995.

[34] R. Brunelli, D. Falavigna, T. Poggio, and L. Stringa. Automatic person recognition by acoustic and geometric features. *MVA*, 8:317–325, 1995.

[35] R. Brunelli and D. Falavigna. Person identification using multiple cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17:955–966, 1995.

[36] Stephen McKenna, Shaogang Gong, and Yogesh Raja. Face recognition in dynamic scenes. In *British Machine Vision Conference (BMVC)*. Essex, 1997.

[37] A. Mike Burton, Stepehn Wilson, and Michelle Cowan. Face recognition in poor quality video: Evidence from security surveillance. *Psychological Science*, 1999.

[38] M. Kirby and L Sirovich. Application of the karhunen-loève procedure for the characterization of human faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):103–108, January 1990.

[39] Baback Moghaddam and Alex P. Pentland. Probabilistic visual learning for object representation. *IEEE Transactions on Patterns Analysis and Machine Intelligence*, 19(7):696–710, July 1997.

[40] Baback Moghaddam, Wasiuddin Wahid, and Alex Pentland. Beyond eigenfaces: Probabilistic matching for face recognition. In *Third IEEE International Conference on Automatic Face and Gesgure Recognition*, Nara, Japan, April 1998. IEEE Computer Society Press.

[41] Wen Yi Zhao, Rama Chellappa, and P. J. Phillips. Subspace linear discriminant analysis for face recognition. *IEEE Transactions on Image Processing*, 1999. under revision.

[42] P. Phillips, H. Wechsler, J. Huang, and P Rauss. The feret database and evaluation procedure for face recognition algorithms. *Image and Vision Computing*, 16(5):295–306, 1998.

[43] K. R. Castleman. *Digital Image Processing.* Englewood Cliffs, NJ, 1996.

[44] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing.* Addison-Wesley Publishing Compuany, 1992.

[45] Roberto Marcondes Cesar Junior. *Análise Multi-Escala de Formas Bidimensionais.* PhD thesis, IFSC - USP, São Carlos, 1997.

[46] John B. Kennedy and Adam M. Neville. *Basic Statistical Methods for Engineers and Scientists.* Harper and Row, Publishers, third edition, 1986.

[47] Carlos A. Callioli, Hygino H. Rodriguese, and Roberto C. F. Costa. *Álgebra Linear e Aplicações.* Editora Atual, SP, sexta edition, 1998.

[48] Anil K. Jain and Douglas Zongker. Feature-selection: Evaluation, application, and small sample performance. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(2):152–157, February 1997.

[49] J. Mao, K. Mohiuddin, and A. K. jain. Parsimonious network design and feature selection through node pruning. In *Proc. 12th ICRP*, pages 622–624, Jerusalem, 1994.

[50] P. M. Narendra and K. Fukunaga. A branch and bound algorithm for feature subset selection. *IEEE Trans. Computers*, 26(9):917–922, September 1977.

[51] W. Siedleki and J. Sklansky. A note on genetic algorithms for large-scale feature selection. *Pattern Recognition Letters*, 10:335–347, November 1989.

[52] O.M. Bruno, R.M. Cesar-Jr., L.A. Consularo, and L. da F. Costa. Automatic feature selection for biological shape classification in synergos. In *Proc. Brazilian Conference on Computer Graphics, Image Processing and Vision, (SIBGRAPI-98, Rio de Janeiro - RJ, Out 1998)*, pages 363–370. IEEE Computer Society Press, 1998.

[53] P. Somol, P. Pudil, J. Novovicová, and P. Paclík. Adaptive floating search methods in feature selection. *Pattern Recognition Letters*, 20:1157–1163, 1999.

[54] P. Pudil, J. Novovicová, and J. Kittler. Floating search methods in feature selection. *Pattern Recognition Letters*, 15:1119–1125, November 1994.

[55] André Fábio Kohn. Reconhecimento de padrões, uma abordagem estatística. Apostila, Escola Politécnica, USP, 1998.

[56] Isabelle Block. On fuzzy distances and their use in image processing under imprecision. *Pattern Recognition*, 11(32):1873–1895, 1999.

[57] Teófilo Emídio de Campos, Isabelle Bloch, and Roberto Marcondes Cesar Junior. Feature selection based on fuzzy distances between clusters: First results on simulated data. In Sameer Singh, Nabeel Murshed, and Walter Kropatsch, editors, *Lecture Notes in Computer Science*, Rio de Janeiro, Brasil, March 2001. The International Association for Pattern Recognition, Springer-Verlag Press.

[58] Jean-Marie Farines, Joni da S. Fraga, and Rômulo S. de Oliveira. *Sistemas de Tempo Real*. IME/USP, São Paulo-SP, Brasil, July 2000. XII Escola Nacional de Computação.

[59] Teófilo Emídio de Campos, Rogério Schmidt Feris, and Roberto Marcondes Cesar Junior. Discriminação de faces $\times$ não-faces usando descritores de fourier. In *52a. Reunião Anual da SBPC*, UnB, Brasilia - DF, Brasil, July 2000. Sociedade Brasileira para o Progresso da Ciência.

[60] R. Lowen and W. Peeters. Distances between fuzzy sets representing gray level images. *Fuzzy Sets and Systems*, 99(2):143–153, 1998.

[61] R. Lowen and W. Peeters. On various classes of semi-pseudometrics used in pattern recognition. In *Seventh IFSA World Congress*, volume I, pages 232–237, Prague, June 1997.

[62] Shaogang Gong, Stephen McKenna, and Alexandra Psarrou. *Dynamic Vision: From Images to Face Recognition.* Imperial College Press, UK, May 2000.

[63] Teófilo Emídio de Campos, Rogério Schmidt Feris, and Roberto Marcondes Cesar Junior. A framework for face recognition from video sequences using gwn and eigenfeature selection. In Roberto Marcondes Cesar Junior, editor, *Proceedings of I-WAICV*, Atibaia - Brasil, November 2000. IME - USP. submitted.