

Reconhecimento de Imagens de Faces Humanas

3.1 Introdução

Um sistema de reconhecimento de faces é um sistema biométrico que identifica ou verifica seres humanos através de uma característica exclusiva, a face. Sistemas biométricos capturam atributos inerentes a cada indivíduo em particular e que podem ser medidos. Alguns exemplos de características humanas usadas para biometria são mostradas na Tabela 3.1.

Grandeza biométrica	Características observadas	Tipo
Assinatura	O padrão, velocidade, aceleração e a pressão da caneta ao escrever uma assinatura.	Comportamental
Impressão Digital	Padrão dos sulcos cutâneos da superfície da ponta do dedo	Fisiológica
Voz	A maneira como humanos geram sons a partir das regiões vocais, boca, cavidades nasais e lábios.	Comportamental
Íris	A região circular do olho limitada pela pupila e a esclera .	Fisiológica
Retina	O padrão formado pelas veias abaixo da superfície da retina em um olho.	Fisiológica
Mão	Geometria da mão humana.	Fisiológica
Orelha	Geometria da orelha humana.	Fisiológica
Termografia facial	O calor emitido através do rosto.	Fisiológica
Face	Medições de perfil, frontal e forma.	Fisiológica

Tabela 3.1: Exemplos de biometrias.

O reconhecimento de face a partir de imagens estáticas e imagens de vídeo vem há décadas emergindo como uma atividade na área de pesquisa científica com numerosas aplicações comerciais, industriais e legais. Essas aplicações, cada vez mais requerem algoritmos robustos que atuem sobre diferentes condições de iluminação, expressões faciais e orientações.

O problema do reconhecimento de face pode ser classificado como:

- **Identificação** (reconhecimento): Dada uma pessoa com sua face a ser investigada e uma galeria de faces de indivíduos, a tarefa de identificação consiste em encontrar a classe correta para a face investigada (1:n).
- **Verificação** (autenticação da face): Dada uma face a ser investigada, e o conjunto de faces da classe declarada, a tarefa de verificação/autenticação é certificar ou não a informação declarada (1:1).

O trabalho desenvolvido nesta dissertação focaliza-se na abordagem de reconhecimento por identificação. Trabalhos futuros poderão ser ampliados para a abordagem de verificação.

3.2 Medida de desempenho de um sistema de verificação

Durante o processo de autenticação, a tarefa de um classificador é essencialmente identificar duas classes de padrões, i.e., se os mesmos pertencem a uma pessoa ou não. Os padrões que pertencem a uma pessoa genuína são chamados de “população ovelha” ou classe positiva. Os vetores que não pertencem à pessoa, i.e. os impostores, são chamados de “população lobo” ou classe negativa. Quando a saída do classificador for próxima de zero o padrão de entrada é atribuído à classe negativa e quando for próxima de um à classe positiva.

Se cada um dos dois conjuntos de padrões possuem probabilidades de serem classificados por um classificador através de uma distribuição, por exemplo distribuição normal, com o limite da saída entre zero e um, um classificador bem treinado daria o conjunto de resultados mostrados na Figura 3.1:

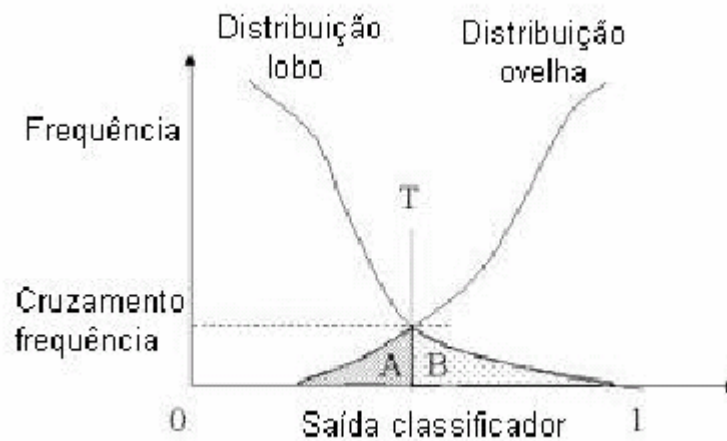


Figura 3.1: Uma típica distribuição das populações de ovelhas e lobos.

Seja T o limiar do classificador, isto implica que, se o classificador responder (por exemplo, a partir de uma saída de neurônio) um valor maior do que T , a pessoa é aceita com a sua identidade declarada, caso contrário ela é rejeitada. A área A na Figura 3.1 (área limitada pela distribuição ovelha, o limiar T e o eixo de saída do classificador), mostra os casos de falsa rejeição e B (área limitada pela distribuição lobo, limiar T e o eixo de saída do classificador) mostra os casos de falsa aceitação. Esta análise é chamada de Análise de Distribuição Ovelha-Lobo e é útil para checar o quanto um classificador tem aprendido ou não.

A área B na Figura 3.1 é proporcional à Taxa de Falsa Aceitação ou Taxa de Alarme Falso (FAR – *False Acceptance Rate*). FAR é também chamado de um Erro Tipo I e é definido pela Equação (3.1).

$$FAR = \frac{\text{Total de Falsa Aceitação}}{\text{Total de Tentativas Falsas}} \quad (3.1)$$

Da mesma maneira, a área A na Figura 3.1 é proporcional à Taxa de Falsa Rejeição (FRR – *False Rejection Rate*). FRR é também chamada de erro Tipo II e é definida pela Equação (3.2).

$$FRR = \frac{\text{Total de Falsa Rejeição}}{\text{Total de Tentativas Verdadeiras}} \quad (3.2)$$

O resultado do deslocamento de T ao longo do eixo de resultados gera diferentes FAR e FRR que, quando “plotados” produzem um gráfico parecido com o da Figura 3.2. Esta análise é chamada de Análise de Limiar e é útil para verificar a Taxa de Erro Idêntico (ERR), dada por $FAR = FRR$.

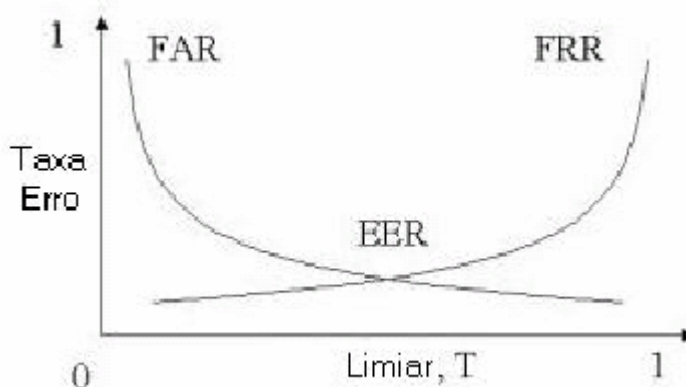


Figura 3.2: FAR e FRR versus Limiar

À medida que T aumenta de zero a um, a FAR decrementa de um a zero e a FRR aumenta de zero a um. Uma FAR alta significa que um impostor teria grandes chances de ser aceito como um usuário verdadeiro, enquanto uma FRR alta significa que um usuário genuíno teria grandes chances de ser rejeitado, quando sua identidade declarada é autêntica. Uma FRR alta causará baixa segurança aos usuários de um sistema de controle de acesso utilizando biometria. Por outro lado, uma FAR alta seria catastrófico pois um impostor poderia facilmente obter um acesso ilegal.

A Frequência de Cruzamento é usualmente exemplo como $(1:X)$, onde X é arredondado para inteiro. A Frequência de Cruzamento e o ERR são frequentemente usados para comparar a qualidade de diferentes classificadores e /ou dados biométricos. ERR pode ser usado para comparar os resultados de dois classificadores ou duas características biométricas. Quanto mais baixo for o ERR melhor, podendo discriminar melhor as duas classes.

“Plotando” FAR versus FRR tem-se a curva de Características de Operação de Receptores ($ROC - Receiver's Operating Characteristics$), que é mostrado na Figura 3.3.

A curva ROC deve este nome ao seu uso original em gerenciamento. Ela também é chamada de detecção do erro da curva de concessão pelo *National Institute of Standards and Technology (NIST)* ou em termos mais gerais como curva de desempenho.

A Figura 3.3 mostra que não é possível minimizar FAR e FRR ao mesmo tempo. Entretanto, um acordo pode ser alcançado, quando $FAR = FRR$, que é o chamado ponto de EER . Em aplicações civis este acerto é alcançável. Um exemplo é o acesso à informação geral, mas em terminal de computador de identidade específica. A linha pontilhada da curva da Figura 3.3 mostra um sistema de melhor qualidade do que o com linha sólida. A curva mais próxima à origem possui um EER menor.

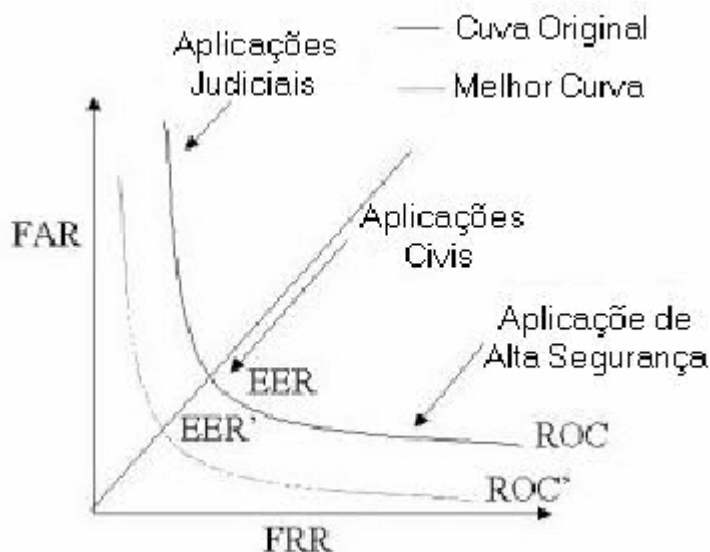


Figura 3.3: FAR versus FRR.

3.3 Detecção de Faces

O trabalho de [YONG R. et al. (1999)] mostra que a interação entre o ser humano e o computador é uma área de intenso nível de pesquisa onde são desenvolvidas interfaces cada vez mais amigáveis. A face humana é um dos meios mais efetivos para se atingir esta meta, pois ela carrega grande número de informações, de modo que computadores poderiam reagir ajustando seu comportamento pelo conhecimento das emoções do usuário, através de suas expressões faciais. Atenção visual é outro exemplo onde computadores podem reagir com base nos interesses de seus usuários. Rumo a este objetivo, o reconhecimento de faces e expressões faciais tem atraído muita atenção recentemente, que vem sendo estudado por mais de vinte anos por psicólogos, neurocientistas e engenheiros. Muitas aplicações interessantes e úteis foram desenvolvidas com estes esforços. A maioria dos métodos existentes consideram que faces humanas devem ser extraídas de imagens estáticas ou de uma seqüência de imagens e focalizadas através de algoritmos de reconhecimento.

Porém, esta é uma tarefa muito desafiante e não mais fácil do que o reconhecimento de faces. Detecção de face é consideravelmente difícil porque envolve a localização da face sem nenhum conhecimento prévio de suas escalas, localização, orientação (ereta, rotacionada), existência ou não de oclusões, diferentes poses (frontal, perfil). Expressões faciais e condições de iluminação também podem alterar por completo as aparências de faces, tornando difícil detectá-las. Além disso, a aparência de faces humanas em uma imagem depende da posição das pessoas e do campo de visão dos dispositivos de aquisição de imagem.

3.4 Reconhecimento de Faces

Os primeiros trabalhos em reconhecimento de faces por computador necessitavam de operadores humanos para a localização de pontos da face cujas posições eram fornecidas como entrada. Dado um conjunto de distâncias de pontos, uma delas poderia ser, por exemplo, a distância da ponta do nariz ao queixo, de uma pessoa desconhecida. A técnica do vizinho mais próximo ou outras regras de classificação eram usadas para identificá-la. Como a extração de características era feita manualmente, o sistema era indiferente a grandes variações na rotação da cabeça, inclinações, qualidade de imagem e contraste. Posteriormente, vieram trabalhos similares, mas sem intervenção humana para a aquisição dos dados de entrada [CHELLAPA, R. (1995)].

3.4.1 Abordagem estatística

Segundo [NEFLAN, A. V. (1996)] as técnicas estatísticas em reconhecimento de faces são utilizadas tanto na extração de características como na tarefa de classificação. Dentre as abordagens estatísticas para reconhecimento de faces, as mais usadas são: i) Métodos de correlação; ii) Métodos de decomposição de valor singular; iii) Métodos baseados em expansão Karhune-Loeve; iv) Métodos baseados em discriminante linear Fisher; v) Métodos baseados em modelos ocultos de Markov.

3.4.1.1 Métodos de correlação

Um procedimento usado para reconhecimento de faces consiste no casamento entre as imagens teste e um conjunto de treinamento de imagens baseado numa medida de correlação. O método de correlação é computacionalmente muito caro e a dependência do reconhecimento sobre a resolução da imagem tem sido investigada. Estudos mostram que o reconhecimento baseado na correlação tem um bom nível de desempenho usando vetores de características pequenos.

3.4.1.2 Métodos baseados em expansão Karhunen-Loève

O método de reconhecimento utilizando auto-faces tem sido implementado com o propósito de comparações, pois ele foi um dos melhores e mais bem sucedidos algoritmos. Este método desenvolvido no *M.I.T.* por M. Turk e A. P. Pentland [TURK, M. A. & PENTLAND, A. P. (1991)], é chamado de Análise de Componentes Principais, onde poucos parâmetros extraídos da face são usados para representação. Estes parâmetros são obtidos pela projeção da face em um sistema de coordenadas dadas por autovetores da matriz de covariância do conjunto de treinamento. Esses autovetores, imagens deles mesmos, são chamados auto-faces e transpõe um vetor de espaço chamado espaço face. Cada face é então codificada por meio de suas coordenadas no espaço face. A comparação de duas faces corresponde a um cálculo da distância Euclidiana entre suas representações do espaço de face.

O trabalho de [LI, S. Z. (1999)] apresenta uma nova abordagem para classificação chamada *Next Linear Combination (NLC)* para reconhecimento de faces baseado em auto-face. Ele considera múltiplos vetores de padrões disponíveis em classes, cada vetor começando de um ponto em um espaço auto-face. Uma combinação linear de vetores pertencentes a uma classe de face é usada para definir uma medida de distância entre um vetor consulta e a classe. A medida é definida como sendo a distância Euclidiana do vetor consulta para a combinação linear mais próxima (portanto *NLC*). Isso contrasta com a classificação por *Next Neighbour (NN)*, onde um vetor consulta é comparado com cada vetor exemplo individualmente. Usando uma combinação linear de vetores exemplo, ao invés de cada um deles individualmente, melhora-se a capacidade de representação dos exemplos pela generalização, através de interpolação e extrapolação. Experimentos mostraram que isto conduziu a melhores resultados do que

métodos de classificação existentes. A Figura 3.4 ilustra o uso da técnica *NLC* para deduzir a posição do ângulo de visão, iluminação ou expressão, relativa a dois exemplos. Nesta figura tem-se:

- Na primeira linha: faces sob mudanças no ângulo de visão. A face-consulta y (à esquerda) está a um ângulo relativamente central das duas faces exemplos x_1 e x_2 , vistas a um ângulo direito e esquerdo respectivamente.
- Na segunda linha: faces sob mudanças de iluminação. A face-consulta y (à esquerda) é iluminada por uma luz a direita e é comparada a duas faces exemplos x_1 e x_2 , uma é iluminada pela esquerda e outra pelo centro, respectivamente.
- Na última linha: faces sob mudanças de expressões.

Com isso ele mostra que a abordagem *NLC* reduz significativamente a taxa de erro com relação à abordagem de classificação *NN* em auto-face para reconhecimento de faces. Este aprimoramento deve-se a capacidade de representação da técnica *NLC* com relação aos padrões exemplo na base de dados: variações na iluminação, ângulo de visão e na expressão entre as imagem de faces. Exemplos são considerados por variações sem seus pesos que determinam a combinação linear.



Figura 3.4: Imagens de faces para uso da técnica *NLC*.

3.4.1.3 Reconhecimento sob condições gerais de visualização

O reconhecimento sob condições gerais de visualização é uma abordagem paramétrica e estende a capacidade do método auto-face para reconhecimento de objetos em imagem 3D sob diferentes condições de iluminação e visualização. Das N imagens de objetos tidas sob P condições de visão e L condições de iluminação, um conjunto de imagem universal é construído de forma a conter todos os dados disponíveis. Dessa maneira um simples espaço paramétrico descreve a identificação do objeto, bem como as condições de visualização e iluminação. A decomposição auto-face deste espaço é usada para extração e classificação de características. Entretanto, para garantir a discriminação entre diferentes classes de objeto o número de auto-vetores usados neste método é maior, quando comparado ao método auto-face clássico [NEFLAN, A. V. (1996)].

3.4.1.4 Reconhecimento usando auto-características

O reconhecimento usando auto-características consiste no uso de características faciais para o reconhecimento de faces. Isto pode ser visto como uma representação modular ou por camadas da face, onde uma descrição grosseira (baixa resolução) de toda a face é definida por detalhes adicionais (alta resolução), salientando as regiões das características. A técnica auto-face foi estendida para detectar características faciais. Para cada característica da face, um espaço de característica é construído pela seleção da maioria de auto características mais significativas (autovetores correspondentes para grandes autovalores da matriz de correlação de características). Na representação auto-característica a equivalente distância do espaço de característica pode ser efetivamente usada para detecção de características faciais [NEFLAN, A. V. (1996)].

3.4.1.5 Método discriminante linear – *Fisherfaces*

No método discriminante linear há a redução da dimensionalidade do espaço de característica pelo uso do Discriminante Linear *Fisher* (*FLD*) [CHOI, M.S. & KIM, W.Y. (2000)]. O *FLD* usa a informação de um grupo de classes e desenvolve um conjunto de vetores de características nos quais variações de diferentes faces são

ênfatisadas, enquanto diferentes exemplos de faces, devido a condições de iluminação, expressões faciais e orientação, são “não ênfatisadas”.

3.4.2.6 Método baseado no modelo oculto de Markov

O modelo oculto de Markov ou *Hidden Markov Model (HMM)* é um conjunto de modelos estatísticos usados para caracterizar propriedades estatísticas de um sinal. *HMM* é feito a partir de dois processos inter-relacionados: (1) uma cadeia de Markov secreta e não observável com finitos números de estados, uma matriz de probabilidade de transição de estado e uma distribuição de probabilidade de estado inicial, (2) um conjunto de funções de densidade probabilidade associado a cada estado.

O modelo *HMM* tem sido usado extensivamente para reconhecimento de voz, onde os dados tem naturalmente uma dimensão ao longo do eixo do tempo. Entretanto, uma cadeia *HMM* equivalente e completamente conectada em duas dimensões dominaria uma grande quantidade de problemas computacionais. Tentativas foram realizadas para usar representação multi-modelos que conduzem a um pseudo *HMM 2D*. Estes modelos são atualmente usados no reconhecimento de caracteres. Foi proposto em [NEFLAN, A. V. (1996)] o uso de *1D HMM* contínuo para o reconhecimento de faces. Supondo-se que cada face está em uma posição ereta e frontal, características ocorrerão em uma ordem previsível, isto é, testa, olhos, nariz, etc. Esta ordenação sugere o uso de um modelo “*Top bottom*” onde somente transições entre estados adjacentes no modo de cima para baixo são permitidos. Os estados do modelo correspondem às características faciais como testa, olho, nariz, boca e queixo. A seqüência de observações é gerada a partir de uma imagem $X \times Y$ usando uma janela de amostra $X \times L$ com $X \times M$ pixels sobrepostos (Figura 3.5) [NEFLAN, A. V. (1996)].

Cada vetor de observação é um bloco de L linhas. Há uma linha M sobreposta entre observações sucessivas. A sobreposição permite que as características sejam capturadas de maneira que a posição vertical seja independente, enquanto um particionamento disjunto da imagem poderia resultar em características nas fronteiras dos blocos.

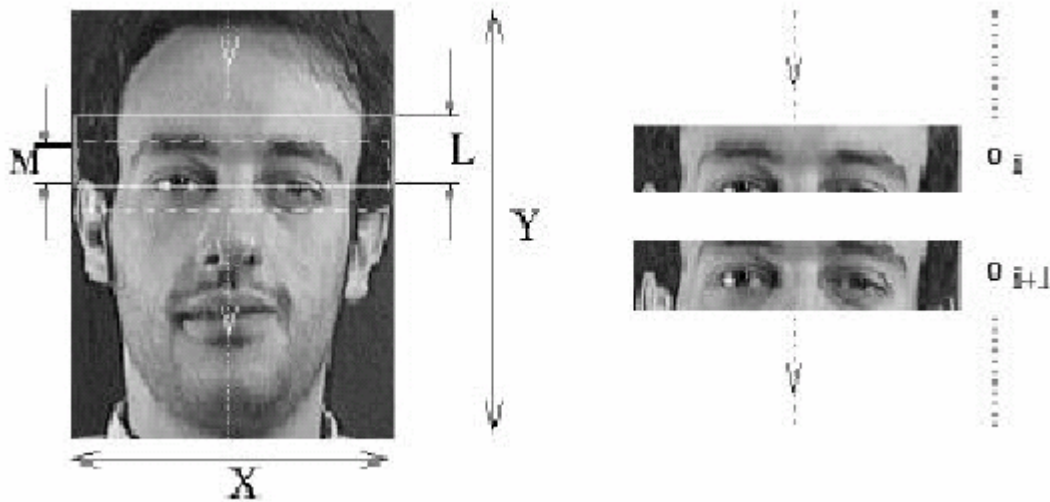


Figura 3.5: Técnica de imagem modelo para reconhecimento *HMM* (NEFLAN, 1996).

3.4.2 RNAs e o reconhecimento de faces

A maioria das aplicações em RNAs (Redes Neurais Artificiais) divide-se em três categorias: classificação (reconhecimento de padrões), onde a rede classifica o padrão de entrada em categorias pré-definidas ou não; predição, onde a rede tenta extrapolar uma série de entradas e controle, onde a rede é usada para guiar interativamente alguns processos externos ou dispositivos. Os dois primeiros são basicamente casos de aproximação, onde se deseja aproximar alguma função tipicamente numérica [GRÖNROOS, M. A. (1998)].

RNAs em reconhecimento de faces vem sendo aplicadas em muitos problemas: classificação por sexo, reconhecimento de faces e classificação de expressões faciais. Uma das primeiras demonstrações foi realizada por [KOHONEN, T. (1995)] através dos chamados mapas associativos. Com um pequeno conjunto de imagens foram realizados experimentos onde a rede deu respostas precisas mesmo quando as imagens de entrada apresentavam muito ruído ou quando partes da imagem estavam faltando [CHELLAPA, R. et al. (1995)].

Atualmente, existem numerosos estudos explorando vários conceitos e problemas no processo de reconhecimento de faces e muitos esforços são empregados na construção de sistemas eficientes com o uso de redes neurais artificiais e outras abordagens. Entretanto, o desempenho da maioria dos esquemas propostos geralmente é sensível à transformação em coordenadas 2D na imagem (por exemplo, escalonamento, translação) [JAIN, A. K. et al. (1996)].

3.5 Considerações Finais

Neste capítulo foram discutidos os principais métodos para reconhecimento de faces humanas, aplicadas à identificação e a verificação. O problema inicial consiste na detecção facial.

Foram apresentadas as abordagens para a avaliação de sistemas de verificação (*FAR* e *FRR*). As aplicações judiciais devem ter um alto *FAR* para então maximizar a chance de se obter um suspeito. Por exemplo, para se identificar suspeitos, um banco de dados criminal pode ser consultado e retornar uma lista de suspeitos que mais se pareçam com o criminoso, ordenados em ordem crescente de similaridade. Por outro lado, para aplicações de alta segurança tal como uma aplicação para *ATM - Automatic Teller Machines* (p.e. caixas automáticos de bancos), é desejável um *FAR* extremamente baixo, pois tais aplicações não podem tolerar a aceitação de um impostor.