



Pontifícia Universidade Católica do Rio Grande do Sul  
PUCRS

## Detecção de Faces e Características Faciais

Eduardo Costa Lopes

Pós-Graduação em Ciência da Computação

Orientador: Prof. Dr. José Carlos Bins Filho.

RELATÓRIO TÉCNICO No. 44

**Contato:**

[eduardo.lopes@vetorial.net](mailto:eduardo.lopes@vetorial.net)

<http://www.inf.pucrs.br/~elopes>

[bins@inf.pucrs.br](mailto:bins@inf.pucrs.br)

<http://www.inf.pucrs.br/~bins>

Eduardo Costa Lopes é aluno do curso de mestrado do Programa de Pós-Graduação em Ciência da Computação da Faculdade de Informática (PPGCC/FACIN), PUCRS. Engenheiro de Computação pela Fundação Universidade Federal do Rio Grande (FURG). Desenvolveu trabalhos na área de redes neurais, algoritmo genéticos e otimização de arquiteturas de redes neurais. Atualmente sua pesquisa se concentra nas áreas de Visão Computacional e Processamento de Imagens, para o desenvolvimento de um sistema automático de reconhecimento de faces humanas. Sua bolsa de pesquisa é patrocinada pela CAPES, e dedica-se exclusivamente à pesquisa.

Copyright©Faculdade de Informática - PUCRS  
Published by PPGCC/FACIN, PUCRS  
Av. Ipiranga, 6681  
90619-900 Porto Alegre-RS-Brasil

# Sumário

Lista de Figuras . . . . .	4
Lista de Siglas e Abreviaturas . . . . .	5
<b>1 Introdução</b>	<b>6</b>
<b>2 Métodos Estatísticos</b>	<b>9</b>
2.1 Introdução . . . . .	9
2.2 <i>Eigenfaces</i> . . . . .	9
2.2.1 Calculando as <i>Eigenfaces</i> . . . . .	10
2.2.2 Classificação de faces através de <i>eigenfaces</i> . . . . .	11
2.3 Análise por Discriminante Linear . . . . .	12
2.3.1 Discriminante Linear de Fisher . . . . .	12
2.4 Modelos de Markov . . . . .	14
<b>3 Redes Neurais Artificiais</b>	<b>17</b>
3.1 Introdução . . . . .	17
3.2 Redes Neurais para o Reconhecimento de Faces . . . . .	17
<b>4 Conclusão</b>	<b>22</b>

# Lista de Figuras

1.1	Principais componentes de um Sistema de Reconhecimento de Faces . . . .	7
2.1	Imagem original . . . . .	10
2.2	Imagem da Figura 2.1 projetada no espaço de <i>eigenfaces</i> . . . . .	10
2.3	HMM para reconhecimento de faces [14] . . . . .	15
2.4	Imagem de face e os blocos de extração[1] . . . . .	15
2.5	Sistema de reconhecimento de faces utilizando HMM. O modelo que apresentar a maior probabilidade fornece a identidade da face [14] . . . . .	16
2.6	Alguns resultados de reconhecimento de faces utilizando HMM[1] . . . . .	16
3.1	Sistema de reconhecimento de face combinando redes neurais e <i>eigenfaces</i> .	18
3.2	Uma típica rede de convolução [4] . . . . .	19
3.3	Diagrama de blocos do método RNA-HMM . . . . .	20
3.4	Diagrama de blocos do método HMM-RNA . . . . .	20

# Lista de Siglas e Abreviaturas

RAF	-	Reconhecimento Automático de Faces
KLT	-	Karhunen-Loève Transform
HMM	-	Hidden Markov Models
LDA	-	Linear Discriminant Analysis
RNAs	-	Redes Neurais Artificiais
PCA	-	Principal Component Analysis
FLD	-	Fisher Linear Discriminant
RBF	-	Radial Basis Function
ERBF	-	Ensemble of RBF
FBG	-	Face Bunch Graph

# Capítulo 1

## Introdução

Apesar de nos últimos anos a humanidade estar experimentando uma crescente evolução tecnológica, desafios como tornar a sociedade segura ainda não foram resolvidos. No caso da segurança, com o crescimento da população humana e o aumento proporcional dos problemas sociais, ferramentas de segurança tem se tornado uma necessidade vital em qualquer tipo de organização.

Dentro deste cenário de crescente necessidade de segurança e organização, métodos de identificação e autenticação automática são uma chave tecnológica para problemas em várias áreas como: controle de acesso a prédios do governo ou laboratórios de alta periculosidade; controle de acesso em computadores; atividades do dia a dia como transações bancárias; investigação criminal, entre outras.

Normalmente o acesso de pessoas a locais seguros se baseia no uso de elementos que somente elas conhecem, como senhas, ou algo que possuam, como um cartão magnético. O inconveniente desta técnica é que senhas podem ser esquecidas, ou descobertas por um invasor e cartões podem ser extraviados. Para superar a desvantagem do uso de senhas, a biometria passou a ser usada como um recurso de autenticação, pois se baseia em características físicas do sujeito a ser reconhecido. Dentre os métodos biométricos, os mais comuns são o reconhecimento de retina e de impressões digitais. No entanto, segundo Fromherz [20], mesmo estes métodos continuam sendo pouco difundidos.

Porém, dentro da biometria, o reconhecimento de faces tem se tornado uma área mais ativa nos últimos anos [5], por ser a forma mais natural e direta de reconhecimento de uma pessoa. Inspirados na habilidade humana de reconhecer faces como objetos especiais, pesquisadores de visão computacional e fisiologistas tem estudado métodos para o reconhecimento automático e eficiente de faces humanas.

O problema do Reconhecimento Automático de Faces (RAF), consiste na resolução de problemas menores, pois envolve a detecção de faces de uma imagem real, a extração de características da face, verificação e identificação de uma pessoa, Figura 1.1. Embora, sendo constituído de vários componentes um sistema de RAF tem como ponto mais importante a extração de características de uma face, onde se tenta representar da forma mais eficiente uma determinada face, para que posteriormente esta informação possa ser utilizada em possíveis propostas de identificação.



Figura 1.1: Principais componentes de um Sistema de Reconhecimento de Faces

Na extração de características o principal desafio se manifesta em como representar os dados de entrada em um pequeno espaço de características, cujos pontos correspondem a diferentes poses de um mesmo indivíduo, sem haver confusão entre os demais pontos que representam outros indivíduos. Além disso, deve-se levar em conta as diferenças de expressão facial, orientação da face, condições de luz, imagem de fundo, entre outros, que tornam a tarefa de extração de características uma atividade complexa.

Segundo Kamran e Chellapa [5], os métodos de reconhecimento de faces se subdividem em três categorias, de acordo com a técnica de extração de características e a técnica de classificação utilizadas:

- Métodos Estatísticos ou métodos baseados em aparência
- Métodos baseados nas características da face
- Métodos não-estatísticos

Os métodos estatísticos surgiram da necessidade de representar uma face de maneira otimizada e compacta. Os algoritmos pertencentes a esta categoria, geralmente, se baseiam na redução dimensional da imagem. Alguns autores como Turk e Pentland [22] tratam a imagem como um vetor de alta dimensão onde cada pixel é mapeado para um componente do vetor. Eles usam a transformada KLT (*Karhunen-Loève Transform*) para achar os componentes principais da distribuição dos vetores de imagens ou os autovetores <sup>1</sup> da matriz de covariância do conjunto de imagens de faces. Uma vez obtido o conjunto de autovetores, cada imagem do banco de dados pode ser representada como uma combinação linear deles, mas com um número de dimensões bem menor do que a imagem original. O reconhecimento será possível através da comparação de um imagem projetada no espaço de eigenfaces e com as projeções das imagens de faces conhecidas.

O LDA (*Linear Discriminant Analysis*, ou *Fisherfaces*), é outro método estatístico para redução dimensional da imagem, mas que leva em consideração informações qualitativas relativas aos membros de cada classe de imagens de face. Nesta técnica uma base de vetores característicos é formada, na qual, as variações entre diferentes faces é realçada, enquanto

---

<sup>1</sup> *eigenvectors*

que, várias instâncias da mesma face, com diferentes condições de iluminação e expressão facial são "desrealçadas". Tal técnica também é utilizada no reconhecimento genérico de objetos em sistemas de visão de robôs móveis. Ela é útil quando se tem um conjunto de faces que devem ser associadas a um *label*.

Por último, HMMs (*Hidden Markov Models*) são utilizados para o reconhecimento de faces devido as suas propriedades de caracterização estatística de um sinal. Nefian [13] utiliza um HMM, no qual, cada estado do modelo tem associado um elemento da face na seguinte ordem: testa, olhos, boca, nariz, queixo. Cada face a ser reconhecida é modelada por um HMM particular e na classificação de uma face, aquele que apresenta maior probabilidade, define a identidade da pessoa a ser reconhecida.

Em contraste, os métodos não estatísticos, são aqueles, que trabalham diretamente na imagem de intensidade. As Redes Neurais Artificiais (RNAs) são as ferramentas mais comuns destes métodos, sendo utilizadas para o reconhecimento de faces devido as características inerentes ao modelo, que os permite realizar tarefas de classificação de forma inteligente. Jamil [7] descreve um experimento para reconhecimento de faces que utiliza *eigenfaces* e redes neurais. As *Eigenfaces* são aplicadas para extrair as informações relevantes de uma face, com objetivo de reduzir o número de entradas e evitar o uso da imagem inteira. As redes neurais, então, são usadas para classificar as faces, através do aprendizado correto dos seus respectivos coeficientes resultantes do algoritmo de *eigenfaces*. Além disso, existem várias maneiras de se utilizar as RNAs como classificadores, bem como existem várias arquiteturas e algoritmos de treinamento que podem ser utilizados para se construir um reconhecedor de faces.

Finalmente, os métodos baseados na geometria da face <sup>2</sup> tentam localizar os componentes de maior relevância na face ou pontos característicos de uma face. Os pontos característicos, uma vez determinados, são utilizados para calcular as distâncias e os tamanhos relativos entre os componentes de maior relevância da face, que uma vez normalizados, constituirão um vetor característico que será usado na classificação. Os sistemas que utilizam este método também se baseiam em algoritmos de janela deslizante para processamento local juntamente com algum outro algoritmo de procura que localiza os pontos característicos da face. Estes métodos são mais tolerantes a diferenças de escalas, tamanho e localização da face na imagem. Porém são mais sensíveis a erros provocados quando existe algum objeto sobrepondo a face, como por exemplo, uma imagem de alguém com óculos. O mesmo problema se manifesta no caso de expressões faciais e pequenas variações de pose. Geralmente as técnicas baseadas em geometria da face tentam construir um molde de uma determinada face (*template*). O reconhecimento se dará através da comparação de um rosto com um conjunto de *templates*, que de acordo com uma função de energia será determinada qual *template* modela a face sendo reconhecida e a que indivíduo pertence.

Neste trabalho, nos capítulos que se seguem serão abordadas as principais técnicas de reconhecimento de faces, bem como, serão dados exemplos do uso das mesmas em problemas reais.

---

<sup>2</sup>*features-based*



# Capítulo 2

## Métodos Estatísticos

### 2.1 Introdução

Neste capítulo serão apresentadas algumas das técnicas mais comuns de reconhecimento de faces, utilizadas em vários sistemas de reconhecimento devido as suas propriedades de compressão e representação eficiente dos dados que caracterizam uma determinada imagem.

O desenvolvimento de um modelo computacional para o reconhecimento de faces é uma tarefa complexa, pois a face em si é complexa e multidimensional. Muitos trabalhos em reconhecimento de faces partem do pressuposto que medidas contidas na própria face são relevantes e suficientes. Isto sugere um método de codificação e decodificação de informação (imagens de faces) que realce as características locais e globais. Tais características não se referem obrigatoriamente a noção de características da face como: nariz, boca, olhos, etc.

Em uma linguagem mais formal, queremos extrair informações relevantes de uma imagem e codificá-la da maneira mais otimizada possível e comparar cada face codificada com um banco de dados de modelos codificados de maneira semelhante, permitindo que se realize a identificação de um determinado rosto.

### 2.2 *Eigenfaces*

O método *eigenfaces* proposto por Turk e Pentland [22] é baseado na transformada de Karhunen-Loève KLT, ou PCA (*Principal Component Analysis*), e é motivada pelo trabalho anterior de Sirovich e Kirby [12] devido a sua eficiência na representação de figuras.

Segundo Pentland e Turk [22] as imagens de faces não se encontram distribuídas de forma aleatória em um espaço de alta dimensionalidade, portanto elas podem ser descritas de alguma forma em um espaço de dimensão menor. De acordo com esta idéia, faz-se o uso da transformada KLT (*Karhunen-Loève Transform*) ou também conhecido como PCA (*Principal Component Analysis*), com o objetivo de achar os vetores que melhor descrevem a distribuição de imagens dentro do espaço de imagens inteiro. Tais vetores são denominados *eigenfaces* devido a semelhança que possuem com imagens de faces, Figuras 2.1 2.2.



Figura 2.1: Imagem original



Figura 2.2: Imagem da Figura 2.1 projetada no espaço de *eigenfaces*

### 2.2.1 Calculando as *Eigenfaces*

Considerando  $I(x, y)$  uma imagem <sup>1</sup> de duas dimensões  $N \times N$ , ou um vetor de dimensão  $N^2$ . A partir de um conjunto de imagens de testes  $I = I_1, I_2, \dots, I_M$ , encontra-se a face média  $fm$  do conjunto definida por:

$$fm_j = \frac{1}{M} \sum_{i=1}^M I_i^j \quad (2.1)$$

onde  $I_i^j$  refere-se ao pixel  $j$  da imagem  $i$ .

Cada face difere da média por:

$$\phi_i = I_i - fm \quad (2.2)$$

O conjunto  $I$  é submetido ao PCA que acha um conjunto de  $M$  vetores ortonormais  $u_k$  e seus respectivos *eigenvalues*  $\lambda_k$  que melhor descrevem a distribuição do conjunto  $I$ . Os vetores  $u_k$  e os escalares  $\lambda_k$  são os respectivamente os autovetores e os autovalores da matriz de covariância definida por:

$$C = \frac{1}{M} \sum_{n=1}^M \phi_n \phi_n^T = AA^T \quad (2.3)$$

onde a matriz  $A = \phi_1, \phi_2, \dots, \phi_M$ . A dimensão de  $A$  é  $N^2 \times N^2$ , o que torna o cálculo dos  $N^2$  *eigenvalues* e *eigenvectors* uma tarefa intratável para imagens de tamanho típico

---

<sup>1</sup>quando nos referimos a imagem estamos considerando que a mesma se refere a imagem de uma face

como 128x128 pixels, que neste caso nos fornece uma matriz de covariância de 128<sup>2</sup>x128<sup>2</sup>. Precisamos de outro método de se calcular os *eigenvectors* que não seja tão dispendioso.

Felizmente existe uma maneira de se obter os *eigenvectors* através da resolução de uma matriz de dimensão  $M \times M$  e realizar-se combinações lineares dos vetores resultantes [22]. Tal análise, chamada de *snapshot method*, permite uma grande redução dos cálculos da ordem de  $N^2$  (número de pixels da imagem) para a ordem  $M$  (número de imagens no conjunto de treinamento). Como o conjunto de treinamento  $M$  é, geralmente muito pequeno, ou seja,  $M \ll N^2$ , teremos um número de cálculos bastante reduzidos. Os *eigenvalues* associados a cada *eigenvector* nos permite organizar cada *eigenvector* de acordo com a sua utilidade na caracterização das variações entre as imagens.

## 2.2.2 Classificação de faces através de *eigenfaces*

Uma vez que tenhamos determinado as *eigenfaces*, o processo de identificação, se torna agora, uma tarefa de reconhecimento de padrões. As *eigenfaces* definem um subespaço de dimensão  $M'$  dentro do espaço de imagens original de dimensão  $N^2$ . Este novo espaço é formado dos  $M'$  *eigenvectors* com os maiores *eigenvalues* associados. O número de *eigenfaces* utilizadas para formar a nova base de vetores é determinada heurísticamente de acordo com o valor dos *eigenvalues*, geralmente toma-se os  $k$  vetores com os maiores *eigenvalues* associados, tidos como os mais relevantes, que sejam suficientes para representar as imagens de forma eficiente.

Quando uma imagem de teste  $I_t$  é submetida ao reconhecimento, esta é projetada no espaço de *eigenfaces* pela operação:

$$w_k = u_k^T(I_t - fm) \quad (2.4)$$

O pesos obtidos pela equação 2.4 acima formam o vetor  $W_t = [w_1, w_2, \dots, w_M]$  que descrevem a imagem de teste no espaço de *eigenfaces*. Ele será usado para achar qual das faces conhecidas que mais se aproxima da imagem de entrada. Esse procedimento consiste, normalmente no cálculo da distância Euclidiana  $e_k$  entre a imagem de entrada e as faces conhecidas definido por:

$$e_k = \|W_t - W_k\|, \quad (2.5)$$

onde  $W_k$  é o vetor de pesos de uma determinada face conhecida.

A imagem  $I_t$  tem sua identidade determinada quando a distância entre ela e uma face conhecida for menor que um limiar (*threshold*)  $l$ . A identidade será fornecida pela imagem que apresenta uma distância  $e_k$  mínima da imagem  $I_t$ . Se isto não ocorrer em nenhum momento do processo de identificação, então a face é classificada como desconhecida [16] [22].

De acordo com o que foi dito no capítulo anterior, um dos problemas no reconhecimento de faces é como identificar a presença de faces em uma determinada imagem. Segundo Pentland [22], pode-se usar a técnica de *eigenfaces* para detectar a presença de uma ou mais faces dentro de uma imagem, partindo da idéia de que uma face não muda radicalmente

quando projetada no espaço de *eigenfaces*, como podemos ver nas figuras 2.1 e 2.2. Sistemas que utilizam esta técnica para o reconhecimento de faces são razoavelmente robustos a variações de iluminação e orientação da face, mas apresentam a desvantagem de degradarem muito rápido quando existe variação de escala nas imagens de faces [16] [14].

## 2.3 Análise por Discriminante Linear

Em [5] é proposto um novo método de redução de dimensionalidade do espaço de características através do uso do FLD ou Discriminante Linear de Fisher (*Fisher's Linear Discriminant-FLD*). Esta técnica leva em conta informações qualitativas pertinentes as classes de faces [14] realçando as variações entre faces de classes diferentes e "desrealçando" as variações provocadas pelas condições de iluminação, expressão facial e orientação de faces de uma mesma classe[10].

### 2.3.1 Discriminante Linear de Fisher

Para utilizar esta técnica necessitamos de um conjunto de imagens relativamente grande. Este conjunto deve conter várias imagens de uma mesma pessoa com diferentes expressões faciais que corresponderá a uma classe. Portanto, diferentes sujeitos pertencerão a diferentes classes. Essa divisão de imagens em classes permite que seja aplicado algum método de separação que permita a formação de *clusters* no espaço de imagens de faces. O cálculo das matrizes de separação interna e entre as classes é definida por:

$$S_b = \sum_{i=1}^c P_i (\mu_i) \quad (2.6)$$

$$S_w = \sum_{i=1}^c P_i \sum_{j=1}^{N_i} (x_j^i - \mu_i)(x_j^i - \mu_i)^T \quad (2.7)$$

onde  $x_j^i$  é o  $j$ th vetor pertencente a classe  $i$ ,  $\mu_i$  a média da classe  $i$  definido por:

$$\mu_i = \frac{1}{N_i} \sum_{j=1}^{N_i} x_j^i \quad (2.8)$$

e  $\mu$  é a média global de todos os vetores:

$$\mu = \frac{1}{\sum_{i=1}^c N_i} \sum_{i=1}^c \sum_{j=1}^{N_i} x_j^i, \quad (2.9)$$

onde  $S_w$  é a matriz de espalhamento interna da classe, ou seja a média de afastamento entre os vetores da classe  $i$ , enquanto  $S_b$  representa o afastamento da média  $\mu_i$  da classe  $i$  em relação a média global  $\mu$ . Se  $S_w$  for uma matriz singular, o *LDA* achará outra matriz

$V_{otimo} \in R^{n \times k}$  que forneça um valor mínimo para o determinante da matriz  $S_w$  e um valor máximo para a matriz  $S_b$ , isto é:

$$V_{otimo} = \underset{V}{\operatorname{argmax}} \frac{|V^T S_b V|}{|V^T S_w V|} = [\nu_1, \nu_2, \dots, \nu_k] \quad (2.10)$$

onde os termos  $\nu_i$  com  $i = 1, 2, \dots, k$  é o conjunto de autovalores (*eigenvectors*) de  $S_b$  e  $S_w$  correspondendo ao conjunto de autovalores (*eigenvalues*)  $\{\lambda_i | i = 1, 2, \dots, k\}$  :

$$S_b \nu_i = \lambda_i S_w \nu_i \quad (2.11)$$

Infelizmente a matriz  $V_{otimo}$  não pode ser obtida diretamente da equação 2.10 se ela for singular [14], o que na maioria das vezes é verdade. Isto ocorre devido ao fato de que a dimensão de  $S_w$  ser menor que  $N - c$ , onde  $N$  é o número de imagens de treinamento e  $c$  é o número de classes e o número de pixels das imagens ser bem maior que o número de imagens. O problema da matriz  $S_w$  ser singular pode ser contornado por um método chamado *Fisherface*, no qual o conjunto de imagens é projetado em um espaço de dimensões reduzidas ( $N - c$ ), através do PCA resultando em uma matriz  $S_w$  não-singular. Logo após, a equação 2.10 é utilizada para se reduzir a dimensão para  $c - 1$ .  $V_{otimo}$  pode ser definida também por:

$$V_{otimo} = V_{fld} V_{pca} \quad (2.12)$$

$$V_{pca} = \underset{V}{\operatorname{argmax}} |V^T C V| \quad (2.13)$$

$$V_{fld} = \underset{V}{\operatorname{argmax}} \frac{|V^T V_{pca}^T S_b V_{pca} V|}{|V^T V_{pca}^T S_w V_{pca} V|} \quad (2.14)$$

onde  $C$  é a matriz de covariância das imagens que compõe o conjunto de treinamento da equação 2.3. As colunas de  $V_{otimo}$  são vetores ortogonais também chamados de *Fisherfaces*. Uma vez que se tenha encontrado os *Fisherfaces* o procedimento de reconhecimento pode ser feito de maneira semelhante ao método adotado nas *eigenfaces*, ou seja, o reconhecimento pode ser efetuado através do cálculo da distância Euclidiana no espaço característico definido pelos *Fisherfaces* [14].

Alguns autores [10], afirmam que o LDA é capaz de superar o PCA quando se tem um conjunto de imagens muito grande. Os métodos apesar de semelhantes apresentam uma diferença fundamental, uma vez que o método de *eigenfaces* é **global**, ou seja, é aplicado sobre todos os dados indistintamente, enquanto que o LDA é **local** sendo aplicado sobre um subconjunto de dados do conjunto total de dados. Essa diferença pode ser notada uma vez que, o algoritmo de *eigenfaces* é aplicado sobre uma matriz de covariância dos dados e o LDA sobre uma matriz de separabilidade entre os dados.

## 2.4 Modelos de Markov

Historicamente os Modelos Escondidos de Markov (*Hidden Markov Models - HMM*) tem sido uma ferramenta útil no reconhecimento de fala [15]. Tais modelos possuem como principal característica a capacidade de caracterização de propriedades estatísticas de um sinal. HMM são compostos pelos seguintes elementos:

- $S = S_1, S_2, \dots, S_N$  o conjunto de estados;
- $N$ , número de estados do modelo;
- $M$ , número de símbolos diferentes;
- $V = v_1, v_2, \dots, v_M$  o conjunto de diferentes símbolos de observação;
- $A$ , matriz de probabilidade de transição e  $a_{ij}$  um elemento da coluna  $i$  e linha  $j$  da matriz  $A$ , onde

$$a_{ij} = P[q_t = S_j | q_{t-1} = S_i] \quad 1 \leq i, j \leq N, \quad (2.15)$$

considerando que  $q_t$  é um estado do modelo no instante  $t$  ( $1 \leq t \leq T$ ),  $q_t \in S$  e  $T$  o tamanho da seqüência de observação,  $0 \leq a_{i,j} \leq 1$  e  $\sum_{j=1}^N a_{ij} = 1, 1 \leq i \leq N$

- $B = b_{j(k)}$ , matriz de probabilidades dos símbolos de observação, onde

$$b_{j(k)} = P[O_t = v_k | q_t = S_j] \quad 1 \leq j \leq N \leq k \leq M \quad (2.16)$$

e  $O_t$  é o símbolo observado no instante  $t$ .

- $\Pi$ , distribuição inicial de estados onde  $\Pi = \pi_i$  e

$$\pi_i = P[q_1 = S_i], 1 \leq i \leq N$$

Podemos também definí-los pela tripla  $\lambda = (A, B, \Pi)$ .

Nos últimos anos alguns pesquisadores vem estudando a utilidade dos HMM no reconhecimento de faces. Nefian [14][1] utiliza HMM tanto para detecção como para o reconhecimento de faces frontais. Em sua técnica cada estado do HMM está associado a uma região da face (cabelo, testa, olhos, nariz e boca, sempre nesta ordem), figura 2.3. Para cada pessoa a ser reconhecida existe um HMM que a representa, e este deve passar por uma fase de treinamento, onde um conjunto de imagens consistindo de diferentes imagens de face de uma mesma pessoa, é utilizado para estimar os parâmetros  $A, B, \Pi$  do HMM.

Cada imagem do conjunto de treinamento possui largura  $W$  e altura  $H$  e, é dividida em blocos que se sobrepõe de altura  $L$  e largura  $W$ , Figura 2.4. Todos os blocos extraídos das imagens são arranjados em forma de um vetor-coluna consistindo de todos os pixels que pertencem ao bloco e, logo após constrói-se uma matriz de covariância que, através do uso da transformada KLT obtem-se uma nova base (*KLT basis*). Em seguida cada vetor é projetado na nova base, formando-se assim um conjunto de *vetores de observação* com

um número de dimensões reduzidas. Portanto, a partir do conjunto de imagens de um indivíduo extrai-se um conjunto de vetores de observação e cada um deste é associado a um estado do HMM e usado para obter as estimativas iniciais da matriz  $B$ . O restante do treinamento é feito através do uso dos algoritmo de segmentção uniforme e do algoritmo de Viterbi. O treinamento é um processo iterativo e pára quando a probabilidade de segmentação Viterbi é menor do que um limiar. Os parâmetros finais do HMM são obtidos através do procedimento recursivo Bam-Welch.

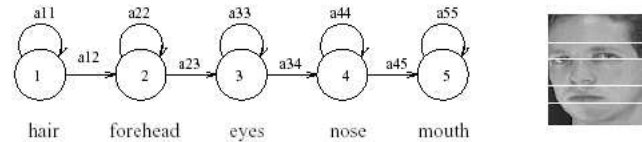


Figura 2.3: HMM para reconhecimento de faces [14]

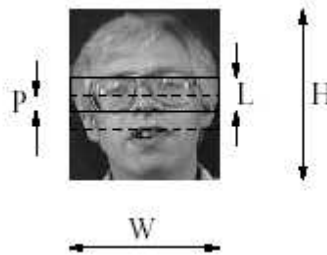


Figura 2.4: Imagem de face e os blocos de extração[1]

O procedimento de reconhecimento, ilustrado na Figura 2.5 ocorre da seguinte maneira: após a extração dos vetores de observação da imagem de teste, da mesma maneira como é feita no treinamento, a probabilidade da seqüência de observação, formada pelos vetores de observação, é calculada em cada HMM através de um reconhecedor Viterbi. O modelo que fornecer a maior probabilidade denotada por:

$$P(O^t|\lambda_k) = \max_n P(O^t|\lambda_n) \quad (2.17)$$

fornecerá a identidade da face.

De acordo com Nefian [13], os HMMs permitem a construção de sistemas de reconhecimento de faces robustos para casos onde as imagens apresentam várias orientações, expressões faciais e diferenças de escala. Além disso a complexidade computacional pode ser reduzida com o uso de algum método de redução de dimensões, como por exemplo, *eigenfaces*. Na figura 2.6 é apresentado algumas das imagens de faces utilizadas como teste por Nefian [1]. As imagens marcadas indicam que estas foram classificadas incorretamente.

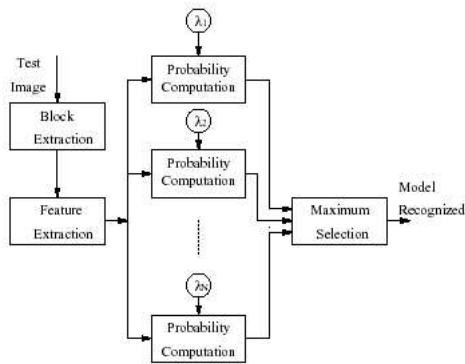


Figura 2.5: Sistema de reconhecimento de faces utilizando HMM. O modelo que apresentar a maior probabilidade fornece a identidade da face [14]



Figura 2.6: Alguns resultados de reconhecimento de faces utilizando HMM[1]

Com o que foi dito aqui podemos concluir que os métodos discutidos, além de serem úteis no reconhecimento também podem ser usados para compressão de imagens. A principal característica desses métodos, exceto aqueles que fazem uso de HMM, é que eles se aplicam a imagem inteira, sem a necessidade de encontrar determinadas regiões. Em contraste, veremos no próximo capítulo técnicas que exigem a extração de determinadas características da face, exigindo uma etapa de pré-processamento para a extração das mesmas, e com elas implementar uma estratégia de classificação baseadas nas características da face.



# Capítulo 3

## Redes Neurais Artificiais

### 3.1 Introdução

Vários problemas em ciência da computação tem solução a partir de inspirações biológicas. A própria pesquisa em visão computacional se baseia na possibilidade de imitar um sentido que se acredita ser de vital importância para a construção de máquinas inteligentes [17]. As redes neurais surgiram, então, na tentativa de imitar o cérebro humano [3], e tem apresentado ótimos resultados em áreas como previsão de clima, classificação e visão computacional [9]. Na próxima seção mostraremos algumas técnicas baseadas em redes neurais para o reconhecimento de faces.

### 3.2 Redes Neurais para o Reconhecimento de Faces

Embora as redes neurais tenham sido muito usadas no reconhecimento de faces a maioria das propostas segue uma abordagem básica: é feita uma extração de atributos sobre os dados, seja através de extração de características faciais, seja através de transformação no espaço de dados (PCA, LDA), e as redes neurais são usadas para fazer a classificação das faces sobre estes atributos [16, 19, 23, 21, 2]. Nesta seção serão apresentadas as formas mais comuns desta abordagem, bem como será descrito uma estratégia que une RNA e HMM para criar métodos híbridos de classificação.

Em [11] é apresentado um sistema de reconhecimento de faces, ilustrado na figura 3.1. A rede neural do tipo *feed-forward* treinada com o algoritmo de retropropagação é utilizada no processo de classificação (reconhecimento) da face. Para reduzir a dimensionalidade dos dados de entrada da rede, projeta-se as imagens no espaço de *eigenfaces* (capítulo 2) e utiliza-se a nova representação para realizar o treinamento da rede. Os autores afirmam que esta técnica é superior a técnica de Turk e Pentland [22], que utilizava somente *eigenfaces*, atingindo taxas de reconhecimento de 94% na classificação correta de faces fora do conjunto de treinamento. Além do reconhecimento de faces [11] afirmam que as redes neurais são capazes de resolver problemas como reconhecimento de raças e classificação de expressões faciais. O inconveniente do uso do algoritmo de treinamento retropropagação é a demora

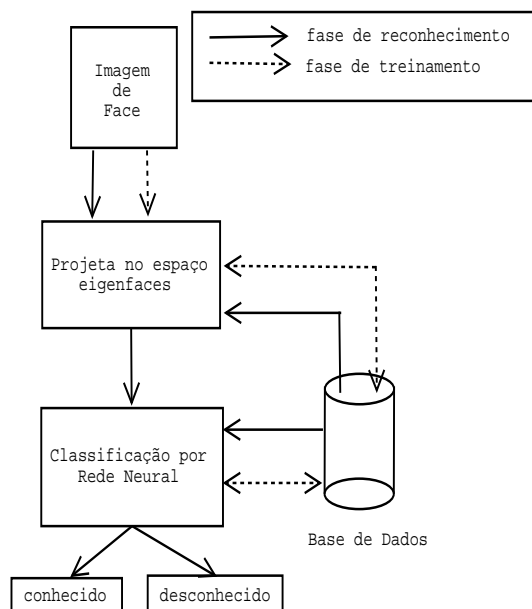


Figura 3.1: Sistema de reconhecimento de face combinando redes neurais e *eigenfaces*

na convergência em um ponto ótimo onde a rede apresenta um erro baixo e uma grande capacidade de generalização, além disso não se tem a garantia de que ele atingirá tal ponto.

Um dos problemas de utilizar redes neurais para o reconhecimento de faces é que muitas vezes elas não generalizam bem para pontos fora do espaço de treinamento, ou seja, não existe um conjunto de faces suficiente que permita construir um conjunto de treinamento que forneça informação suficiente para a rede modelar e reconhecer qualquer tipo de face. Além deste problema é desejável que ela seja capaz de tratar imagens com distorções causadas por translação ou rotação da face. Para o último problema as redes neurais de convolução possuem alguns mecanismos que permitem um certo grau de tolerância as deformações. Este modelo baseia-se em três idéias: campos receptivos, compartilhamento de pesos e subamostragem espacial. Um exemplo deste tipo de rede é mostrado na figura 3.2. A rede é formada de várias camadas de vários planos. Na primeira camada é apresentada uma imagem aproximadamente centralizada e, em cada plano existem unidades que recebem entradas vindas de um pequeno conjunto de unidades dos planos anteriores. Cada plano funciona como um extrator de uma determinada característica e a rede é treinada com um variante do algoritmo de retropropagação adaptado para esta arquitetura [4]. Em [6] é apresentado um sistema baseado em redes de convolução para reconhecimento de faces em multiescala e robusto quanto as variações de expressão facial.

Da mesma forma que se utiliza apenas uma arquitetura de rede neural para implementar um sistema de reconhecimento, também podemos utilizar várias redes através da combinação de suas saídas e obter um resultado superior do que se fosse utilizado apenas uma rede. Gutta [18] apresenta um sistema de reconhecimento de faces para tarefas de vigilância que utiliza um grupo de redes neurais do tipo RBF (*Radial Basis Function*) com-

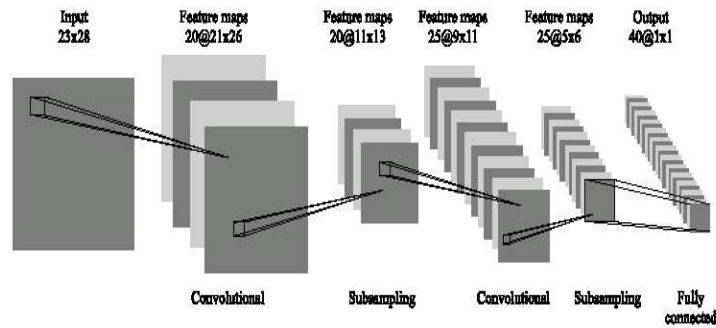


Figura 3.2: Uma típica rede de convolução [4]

pondo um conjunto chamado ERBF (*Ensemble of RBF*). O ERBF de Gutta é composta de nove redes RBF subdividas em grupos de três. As primeiras três redes são treinadas com imagens de faces sem nenhum ruído, outras três redes são treinadas com imagens que possuem borrões e as últimas três são treinadas com imagens que possuem distorções geométricas. O treinamento das redes se dá através da utilização de exemplos positivos e negativos. A decisão final do ERBF (aceita ou rejeita) leva em conta a saída de cada rede se desenrolando da seguinte maneira: se a norma da média das saídas das 9 redes for maior que um valor  $\theta$  então a imagem de entrada é aceita (reconhecida) senão rejeita. Se o valor da norma cair dentro de um intervalo  $(\theta - \delta_\theta, \theta)$  então o resultado é indeciso e não se tira nenhuma conclusão a respeito da imagem de entrada e o ERBF dará como saída o *label* REJECT/UNDECIDED. Segundo Gutta a taxa de acerto do ERBF pode chegar até 99.89% quando  $\delta_\theta = 0.14$  e  $\theta = 0.67$ . Também o autor destaca a utilidade da rede RBF em classificação de imagens devido a sua habilidade para clusterização antes de classificá-la. Esses modelos podem ser usados como um aproximador universal de funções e no caso do reconhecimento de imagens elas são úteis para classificação em termos de raça, sexo e idade antes do reconhecimento final. Maiores detalhes sobre RBF podem ser encontrados em Gutta [18] e Shang-Hung [21].

Até agora vimos como usar uma RNA ou um conjunto de RNAs para o reconhecimento de faces, mas Kang e Young [8] apresentam duas técnicas de reconhecimento híbridas que une HMM e RNAs. Em uma delas as saídas de uma rede neural são enviadas para um HMM (método RNA-HMM). A RNA é utilizada como um etiquetador (*labeler*) para as entradas do HMM. As saídas dos neurônios da camadas de saída estão conectadas as entradas do HMM subsequente e utiliza-se o método *TOP-X* para enviar as informações da RNA para o HMM. O esquema *TOP-X* consiste em estabelecer um *score* para cada saída e escolher as mais altas, ou seja, se  $X$  vale 2, então somente as duas saídas mais altas da RNA serão utilizadas como a seqüência de observação do HMM e, sobre a qual será calculada uma probabilidade via algoritmo Viterbi. Na Figura 3.3 temos um diagrama de blocos ilustrando o método RNA-HMM. Em testes realizados por Kang e Young [8] os melhores resultados foram obtidos quando usou-se a esquema TOP-3.

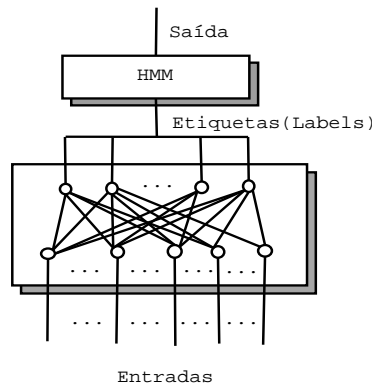


Figura 3.3: Diagrama de blocos do método RNA-HMM

Outra implementação possível, é o método HMM-RNA. Nesta técnica cada indivíduo é representado por um HMM, cujas saídas estão ligadas a entrada de uma RNA. Se duas pessoas apresentam características semelhantes, os respectivos HMMs que as representam terão saídas semelhantes, e diante deste problema o próximo estágio faz uso de uma rede neural para aumentar o poder de discriminação do sistema como um todo, Figura 3.4.

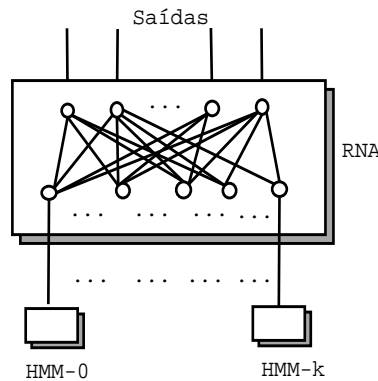


Figura 3.4: Diagrama de blocos do método HMM-RNA

Utilizou-se 600 imagens, sendo 300 para o treinamento e 300 para teste e algumas imagens delas apresentam rostos usando óculos. A taxa de reconhecimento é medida de acordo com os reconhecimentos corretos e segundo Kang e Young [8] o modelo RNA-HMM apresentou uma taxa de 95.7% de reconhecimento quando o esquema TOP-3 é utilizado, a taxa de aprendizado  $\alpha$  e o termo *momentum*  $\gamma$  possuem valor 0.3 e número de estados do HMM é cinco. Em contraste o modelo HMM-RNA alcançou 100% de reconhecimento com os mesmos parâmetros ( $\alpha = 0.3, \gamma = 0.3, \dots$ ) do modelo RNA-HMM.

Os exemplos aqui fornecidos são apenas uma pequena parte do que realmente existe em reconhecimento de faces utilizando redes neurais. Existem vários métodos utilizando as mais diversas arquiteturas de redes ou a combinação delas, abordá-las aqui vai além do

escopo deste trabalho. Com o que foi apresentado pode-se concluir que esta se mostra uma ferramenta útil para a construção de sistemas de reconhecimento de faces robustos para aplicações em problemas reais.

# Capítulo 4

## Conclusão

De acordo com o que foi discutido neste trabalho podemos notar que não existe um algoritmo genérico que seja capaz de resolver o problema de reconhecer faces tal qual como é feito pelos seres humanos. As técnicas aqui apresentadas, são propostas de reconhecimento de faces em determinadas situações, pois podemos notar que alguns métodos são insensíveis a determinados fenômenos como variação de iluminação e expressão. O tempo de resposta e a complexidade dessas técnicas também são importantes, principalmente quando se deseja implementar algum sistema de reconhecimento de faces para problemas do mundo real, como o caso de vigilância de prédios. Além disso, alguns métodos funcionam bem quando o conjunto de dados é relativamente pequeno, enquanto outros degradam nesta situação.

Portanto podemos concluir que as técnicas de reconhecimento de faces são úteis em determinados casos, e a escolha por uma delas será influenciada pela natureza das imagens que estarão disponíveis para o reconhecimento, bem como a quantidade das mesmas e o tempo de resposta dos algoritmos de reconhecimento, e ainda, vale lembrar que mais de uma técnica pode ser utilizada de forma que compensem as deficiências de cada uma.

Com base no que foi estudado e apresentado neste trabalho pretende-se construir um sistema de reconhecimento de faces para controlar automaticamente o acesso e pessoas em prédios. Este sistema deverá tratar um número pequeno de indivíduos(em torno de 100) e usar apenas imagens frontais. Maiores detalhes serão apresentados no Trabalho Individual II.

# Referências Bibliográficas

- [1] M.H. Hayes III A. Nefian. Face detection and recognition using hidden markov models. In *International Conference on Image Processing*, pages 141–145, October 1998.
- [2] William A. Barret. A survey of face recognition algorithms and testing results. In *Conference Record of the Thirty-First Asilomar Conference on Signals, Systems and Computers*, pages 301–305, October 5-7 1998.
- [3] João de Fernandes Teixeira. *Mentes e Máquinas Uma Introdução à Ciência Cognitiva*. Ed. Artes Médicas, 1998.
- [4] Steve Lawrence C. Lee Giles Ah Chung Tsoi Andrew D.Jack. Face recognition a convolutional neural network approach. In *IEEE Transaction on Neural Networks and Pattern Recognition*, volume 8, pages 98–113, 1997.
- [5] Kamran Etermad and Rama Challapa. Face recognition using discriminant eigenvectors. In *IEEE International Conference on Acoustic, Speech and Signal Processing*.
- [6] Beat Fasel. Multiscale facial expression recognition using convolutional neural networks. In *Proceedings of the third Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP'2002), Ahmedabad, India, 2002*.
- [7] Nazish Jamil Samreen Iqbal Naveela Iqbal. Face recognition using neural networks.
- [8] Young Kug Ham Kang Sik Yoon and Rae-Hong Park. Hybrid approaches to frontal view face recognition using the neural network. In *IEEE Proceedings*, pages 1359–1362, 1997.
- [9] Ben J. Kröse and P.Patrick van der Smagt. *An Introduction To Neural Networks*. 5 edition, January 1993.
- [10] Aleix M. Martínéz and Avinash C. Kak. Pca versus lda. In *IEEE Transactions Analysis and Machine Intelligence*, volume 23, pages 228–233, 2001.
- [11] A. Sehad A. Hadid H.Hocini M.Djeddi and S.Ameur. Face recognition using neural networks and eigenfaces. In *ISCA 15th International conference on Computers and their Applications*,, USA, march 29-31 2000.

- [12] M.Kirby and J. Sirovich. Application of the karhunen-loève procedure for the characterization of human faces. In *IEEE Transactions on Pattern and Machine Intelligence*, volume 12, pages 103–108, January 1990.
- [13] A. Nefian and M. Hayes. Hidden markov models for face recognition. In *IEEE Conference on Acoustic Speech and Signal Processing*, pages 2721–2724, 98.
- [14] Ara V. Nefian. *Statistical Approachs To Face Recognition*. PhD thesis, Georgia Institute of Technology, School of Electrical Engineering, December 1996.
- [15] Lawrence R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. In *Proceedings of the IEEE*, volume 77, February 1989.
- [16] Charles L. Wilson Rama chellapa and Saad Sirohey. Human and machine recognition of faces: A survey. In *Proceedings of IEEE*, volume 83, pages 705–740, May 1995.
- [17] Linda G. Shapiro and George C. Stockman.
- [18] S. Gutta J. Huang B. Takacs and H. Wechler. Face recognition using ensembles of networks. In *International Conference on Pattern Recognition ICPR*, Vienna Austria, 1996.
- [19] Ashit Talukder. Adaptive activation function neural net for face recognition. In *IEEE International Conference on Neural Networks*, pages 549–552, 2001.
- [20] Peter Stucki Thomaz Fromherz and Martin Bichsel. A survey of face recognition. Technical Report 97.01, University of Zurich, 1997.
- [21] Meng Joo Er Shiqian Wu Juwei Lu Hock Toh. Face recognition with radial basis function rbf neural networks. In *IEEE Transactions on Neural Networks*, volume 13, pages 697–709, May 2002.
- [22] Matthew A. Turk and Alex P. Pentland. Face recognition using eigenfaces. In *Proceedings of International Conference on Pattern Recognition*, pages 586–591, 1991.
- [23] Marcos Foundéz Zanuy Virginia Espinosa Duró. Face identification by means of a neural net classifier. In *IEEE Proceedings, International Carnahan Conference on Security Technology*, Marid, Spain, October 5-7 1999.