Reconhecimento Facial usando Wavelets de Gabor via Álgebra Multilinear

Emanuel D. R. Sena e André L. F. de Almeida

***Resumo*— Neste trabalho é proposto um sistema de reconhecimento facial que explora a natureza multilinear das imagens. A extração de características ocorre em dois estágios. Primeiramente, wavelets de Gabor são utilizadas para a seleção de características. Em seguida, a decomposição tensorial HOSVD (do inglês: *Higher Order Singular Value Decomposition*) é aplicada para separar os fatores multimodais das imagens. Duas abordagens são propostas para a classificação do conjunto de vetores de características, as quais são baseadas na distância Euclidiana e na distância de Mahalanobis, respectivamente. Foi verificado que a abordagem de reconhecimento facial proposta possui taxas de acerto médio superior aos demais métodos simulados. Analisando os resultados para várias configurações de treinamento, a abordagem proposta demonstra estabilidade quando há deficiência na quantidade de imagens de treino.**

***Palavras-Chave*— Álgebra multilinear, Wavelets de Gabor, decomposições tensoriais, imagens multimodais, reconhecimento facial.**

***Abstract*— In this work, we propose a facial recognition system that exploits the multilinear nature of images. The feature extraction occurs in two stages. First, Gabor wavelets are used for feature selection. Then, the higher order singular value decomposition (HOSVD) is applied to separate the multimodal factors of the images. Two approaches based on Euclidean and Mahanalobis distances are proposed, respectively, for the classification of the set of feature vectors. The proposed facial recognition approach exhibits higher average success rates than competing methods. By analyzing the results for various training settings, the proposed approach demonstrates stability when there is deficiency in the amount of training images.**

***Keywords*— Multilinear algebra, Gabor wavelets, tensor decompositions, multimodal images and face recognition.**

I. Introdução

Imagens são uma junção de múltiplos fatores relacionados tais como estrutura da cena, iluminação, posição dos objetos e no caso de imagens faciais a expressão facial, caracterizando assim uma natureza multimodal [1]. Ao longo das últimas décadas, diversos algoritmos para reconhecimento facial foram produzidos, explorando ao máximo a riqueza de informação contida nos múltiplos domínios.

Devido a essa natureza multimodal, os fatores constituintes se misturam, mas a percepção humana é capaz de ser tolerante aos diversos componentes que formam a imagem, tornando-se um sistema robusto a essas variações, o que não é verdade para sistemas computacionais. Em reconhecimento facial os métodos lineares, por exemplo, *Eigenfaces* [2], têm se mostrado ferramentas robustas quando há variação apenas da identidade do indíviduo no conjunto de imagens e são fixados a posição facial, condições de iluminação e expressão.

Emanuel D. R. Sena e André L. F. de Almeida¸ Departamento de Engenharia de Teleinformática, Universidade Federal do Ceará, Fortaleza-CE, Brasil, E-mails: {dario, andre}@gtel.ufc.br. Este trabalho foi parcialmente financiado pela CAPES.

A álgebra multilinear é uma parte da matemática que estende a álgebra linear, em que os espaços vetoriais são generalizados no conceito de espaços tensoriais através do produto tensorial [3]. O interesse em álgebra multilinear tem se expandido para diversas áreas do conhecimento, incluindo o processameto de imagens para o reconhecimento de padrões e visão computacional. Um ramo da álgebra multilinear são as decomposições tensoriais [4], [5], em especial o HOSVD que faz uso da decomposição em valores singulares para seu cálculo [6], oferecendo uma maneira natural para a análise da estrutura multimodal de um conjunto de imagens, conceito primeiramente introduzido por M. Alex O. Vasilescu e Demetri Terzopoulos no método *TensorFaces* [7], [8].

Neste trabalho é proposto um sistema de reconhecimento facial que explora a natureza multilinear das imagens. Para cada imagem facial é realizada uma transformação através das wavelets de Gabor. Em seguida os fatores multimodais são separados pelo uso da álgebra multilinear (HOSVD), formando espaços vetoriais distintos, relacionados com as variações de identidade, posição, iluminação e expressão. Duas abordagens são propostas para a classificação do conjunto de vetores de características, as quais são baseadas na distância Euclidiana e na distância de Mahalanobis, respectivamente. Foi verificado que a abordagem de reconhecimento facial proposta possui taxas de acerto médio superior aos demais métodos simulados. Analisando os resultados para várias configurações de treinamento, a abordagem proposta demonstra estabilidade quando há deficiência na quantidade de imagens de treino.

O conteúdo deste trabalho está distribuído de forma a expor conceitos e definições relacionados, como as wavelets de Gabor na Seção II. Na Seção III, temos uma introdução a álgebra multilinear e as principais ferramentas utilizadas neste artigo. Na Seção IV, faremos uma rápida descrição sobre o método *TensorFaces*. Nas Seção V, o método proposto é descrito. Na Seção VI, os resultados obtidos são apresentados e discutidos.

Notação: Ao longo do artigo para valores escalares, vetores, matrizes e tensores utiliza-se letras minúsculas $(a, b, \ldots)$, negrito com letras minúsculas $(\mathbf{a}, \mathbf{b}, \ldots)$, negrito com letras maiúsculas $(\mathbf{A}, \mathbf{B}, \ldots)$, negrito com letras caligráficas $(\mathcal{A}, \mathcal{B}, \ldots)$, respectivamente. A transposta de uma matriz é denotada por $\mathbf{A}^T$. A norma euclideana é denotada como $\| \cdot \|$. Temos ainda que $\psi_{(\mathrm{f},\theta,\sigma_1,\sigma_2)}(x, y)$ representa uma função de (x, y) com parâmetros $(\mathrm{f}, \theta, \sigma_1, \sigma_2)$. A operação de vetorização de uma matriz é representada pelo operador $vec(\cdot)$. O operador $*$ e $\mathfrak{F}\{\cdot\}$ representam a convolução e transformada de Fourier respectivamente, a multiplicação modo-n é denotada por $\times_n$.

II. WAVELETS

Quando consideramos o espaço $L^2(\mathbb{R})$ de todas as funções mensuráveis de quadrado integrável sobre $\mathbb{R}$, as funções $f(x)$ devem decair para zero, quando $x \to \pm\infty$. Uma base que gera $L^2(\mathbb{R})$ possui funções com essa característica. A idéia principal das wavelets é considerar dilatações, translações ou rotações (wavelets de Gabor) de uma única função $\psi \in L^2(\mathbb{R})$ [9]. Consideramos wavelets $\psi_{\alpha,\beta}(x) = \psi(2^\alpha x - \beta)$, $\alpha, \beta \in \mathbb{Z}$, formando uma base não necessariamente ortogonal.

A. Wavelets de Gabor

As wavelets de Gabor são um conjunto de filtros gerados a partir de uma função Gaussiana modulada por uma exponencial. Definidas conforme [10], [11], [12]:

$$\psi_{(\mathrm{f},\theta,\sigma_1,\sigma_2)}(x,y) = \frac{\mathrm{f}^2}{\pi\sigma_1\sigma_2} e^{-(\alpha^2\mathbf{x}^2+\beta^2\mathbf{y}^2)} e^{j2\pi\mathrm{f}\mathbf{x}} \quad (1)$$

$$\mathbf{x} = x\cos\theta + y\sin\theta \quad (2)$$

$$\mathbf{y} = -x\cos\theta + y\sin\theta \quad (3)$$

Os filtros de Gabor são auto-similares, sendo possível gerar qualquer filtro a partir de $\psi_{(\mathrm{f},\theta,\sigma_1,\sigma_2)}(x,y)$. Quando necessita-se extrair características de uma imagem, utiliza-se um conjunto de filtros com diferentes frequências e rotações:

$$\psi_{u,v} = \psi_{(\mathrm{f}_u,\theta_v,\sigma_1,\sigma_2)} \text{ onde } \mathrm{f}_u = \frac{\mathrm{f}_{max}}{\sqrt{2^u}},\ \theta_v = \frac{v}{8}\pi \quad (4)$$

em que $u = 0, 1, \ldots, U-1$; $v = 0, 1, \ldots, V-1$. Note que f_{max} é a maior frequência que as wavelets de Gabor podem assumir, U é o número de escalas, e V é o número de orientações. Os parâmetros de $\psi_{(\mathrm{f},\theta,\sigma_1,\sigma_2)}(x,y)$ devem ser escolhidos de forma que a extração de características forneça a maior quantidade de informação possível, assim f_{max} deve assumir valores de baixa frequência, devido imagens faciais possuirem sua informação concentrada em baixas frequências. Os valores comumente usados [11] são $\mathrm{f}_{max} = 0.25$, $\sigma_1 = \sigma_2 = \sqrt{2}$ e ainda $\sigma_1 = \frac{\mathrm{f}_u}{\alpha}$, $\sigma_2 = \frac{\mathrm{f}_u}{\beta}$ donde $\alpha = \beta$, mantendo a razão entre a frequência e o formato da Gaussiana constante, ou seja, teremos cópias da mesma função em escalas diferentes quando a orientação for a mesma, a quantidade de escalas é 5, $u = 0, 1, \ldots, 4$ e 8 orientações, $v = 0, 1, \ldots, 7$. A representação de Gabor de uma imagem facial $I(x,y)$ pode ser obtida convoluindo a imagem com a família de wavelets Gabor [11]:

$$\varphi_{u,v}(x,y) = I(x,y) * \psi_{u,v}(x,y) \quad (5)$$

onde $\varphi_{u,v}(x,y)$ denota o resultado da convolução com o filtro de Gabor na orientação u e escala v.

É realizado um *downsampling* por um fator δ em cada $\varphi_{u,v}(x,y)$ obtendo $\varphi^{\delta}_{u,v}(x,y)$, então vetorizado

$$\mathbf{g}_{u,v} = vec(\varphi^{\delta}_{u,v}(x,y)) \quad (6)$$

e normalizado para média zero e variância unitária, gerando um vetor de características $\mathbf{g} = (\mathbf{g}_{0,0}\ \mathbf{g}_{0,1}\ \ldots\ \mathbf{g}_{4,7})^T$ através da concatenação de todos os vetores $\mathbf{g}_{u,v}$.

III. ÁLGEBRA MULTILINEAR

A álgebra multilinear é uma generalização da álgebra linear. De uma maneira simplista, podemos ver que uma função linear é também multilinear em uma variável. A álgebra multilinear é construída sobre o conceito de tensor e espaços tensoriais, análogo aos vetores e espaços vetoriais.

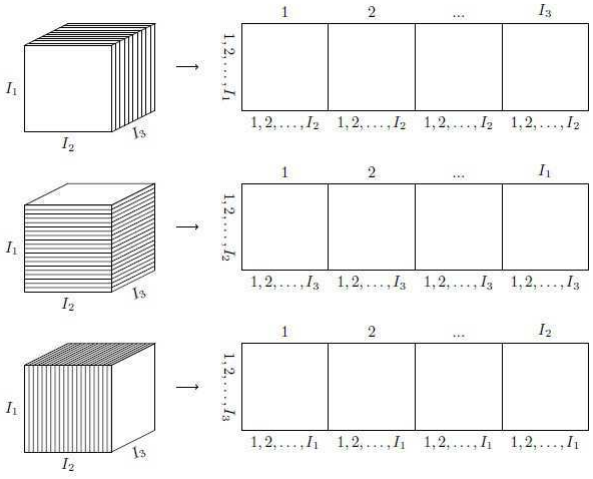

Fig. 1. Tensor matriciado, nos modos 1, 2 e 3.

A. Tensores, fibras, fatias e matriciação de tensores

Um tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ de ordem N pode ser visto como um array multidimensional associado a um produto tensorial de N espaços vetoriais [3] em que seus componentes ou entradas $x_{i_1 i_2 \ldots i_N}$ são acessados através de seus índices.

A ordem do tensor corresponde ao número de dimensões do mesmo, e cada dimensão está associada a um índice. Pela definição acima observamos que escalares, vetores e matrizes também são tensores de ordem zero, um e dois respectivamente. Um subtensor é obtido quando um subconjunto de índices são fixados. Uma fibra de um tensor é um fragmento unidimensional de um tensor obtido a partir da fixação de seus índices exceto por um. Uma fatia (*slice*) de um tensor é uma seção bidimensional de um tensor, obtida fixando seus índices exceto dois.

A operação de matriciação ou desdobramento (do inglês, *unfolding* ou *matricization*) de tensores em matrizes foi proposta por [13], [6], de modo que as fibras de uma determinada dimensão serão as colunas da matriz resultante. A matriciação de um tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ é definida da seguinte maneira [14]: seja os conjuntos ordenados $\gamma = \{\gamma_1, \gamma_2, \ldots, \gamma_L\}$ e $\lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_M\}$ duas partições do conjunto ordenado dos modos $\mathfrak{N} = \{1, 2, \ldots, N\}$ do tensor $\mathcal{A}$. Considere $\mathbf{I}_{\mathfrak{N}} = \{I_1, I_2, \ldots, I_N\}$ o conjunto das dimensões de cada modo de $\mathcal{A}$, definimos a matriciação como:

$$\mathbf{A}_{(\gamma\times\lambda)} \in \mathbb{R}^{J\times K} \text{ em que } J = \prod_{n\in\gamma} I_n \ e\ K = \prod_{n\in\lambda} I_n \quad (7)$$

Quando γ é uma partição de apenas um elemento, isto equivale as fibras do modo n serem as colunas da matriz resultante (Figura 1). A matriciação modo-n de um tensor $\mathcal{D} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ é um caso particular da matriciação:

$$\mathbf{D}_{(n)} \equiv \mathbf{D}_{(\gamma\times\lambda)}, \quad \gamma = \{n\} \text{ e } \lambda = \{1, \ldots, n-1, n+1, \ldots, N\} \quad (8)$$

B. Produto modo-n

O produto modo-n de um tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ por uma matriz $\mathbf{B} \in \mathbb{R}^{J_n \times I_n}$, denotado por $\mathcal{C} = \mathcal{A} \times_n \mathbf{B}$ é definido

como:

$$[\mathcal{A} \times_n \mathbf{B}]_{i_1 i_2 \ldots j_n \ldots i_N} = \sum_{i_n=1}^{I_n} [\mathcal{A}]_{i_1 i_2 \ldots i_n \ldots i_N} [\mathbf{B}]_{j_n i_n} \quad (9)$$

podemos expressar essa definição em forma matricial como:

$$\mathbf{C}_{(n)} = \mathbf{B}\mathbf{A}_{(n)} \quad (10)$$

Seja $\mathbf{C} \in \mathbb{R}^{J_m \times I_m}$, note que a seguinte propriedade é válida:

$$\mathcal{A} \times_n \mathbf{B} \times_m \mathbf{C} = \mathcal{A} \times_m \mathbf{C} \times_n \mathbf{B} \quad (11)$$

C. HOSVD

No contexto das decomposições tensoriais o Higher Order Singular Value Decomposition é uma extensão do SVD (decomposição em valores singulares) que ortogonaliza os N espaços vetoriais relacionados com o tensor $\mathcal{D} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$, no qual podemos expressar o tensor como produtos modo-n das bases desses espaços:

$$\mathcal{D} = \mathcal{Z} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \ \ldots \times_N \mathbf{U}_N \quad (12)$$

O tensor $\mathcal{Z} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ é conhecido como *tensor núcleo*, que governa as interações entres as bases e os vários modos. As matrizes $\mathbf{U}_i$ contém bases para o espaço coluna de $\mathbf{D}_{(i)}$ [6], [16], resultante da operação de matriciação do tensor $\mathcal{D}$ no modo i. O tensor $\mathcal{Z}$ e as matrizes $\mathbf{U}_i$ podem ser calculados de acordo com o Algoritmo 1 [6].

Data: *Tensor* $\mathcal{D}$
Result: *Tensor núcleo* $\mathcal{Z}$ *e matrizes* $\mathbf{U}_i$
$N \longleftarrow$ ordem do tensor;
for $i \leftarrow 1$ **to** N **do**
  $\mathbf{U}_i \leftarrow$ autovetores a esquerda de $\mathbf{D}_{(i)}$
end
$\mathcal{Z} \leftarrow \mathcal{D} \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \ \ldots \times_N \mathbf{U}_N^T$

Algoritmo 1: HOSVD de um tensor

IV. TENSORFACES

A álgebra multilinear oferece uma abordagem natural para a análise de imagens com estrutura multifatores. O método *TensorFaces* consiste em realizar uma modelagem tensorial sobre um conjunto de imagens faciais que são vetorizadas e organizadas como um tensor. Primeiramente devemos construir o tensor de ordem $N = 5$:

$$\mathcal{D} \in \mathbb{R}^{N_{pess} \times N_{pos} \times N_{ilum} \times N_{exp} \times N_{pixels}}$$

Onde $N_{pess}, N_{pos}, N_{ilum}, N_{exp}$ e N_{pixels} representam o número de pessoas, posições faciais, condições de iluminação, expressões faciais e pixels respectivamente, referente a cada fator constituinte na formação da imagem:

$$\underbrace{\text{pessoa}}_{\text{modo 1}} \times \underbrace{\text{posição}}_{\text{modo 2}} \times \underbrace{\text{iluminação}}_{\text{modo 3}} \times \underbrace{\text{expressão}}_{\text{modo 4}} \times \underbrace{\text{pixels}}_{\text{modo 5}}$$

Dessa maneira, o tensor $\mathcal{D}$ pode ser decomposto aplicando-se o HOSVD conforme Algoritmo 1, obtendo-se:

$$\mathcal{D} = \mathcal{Z} \times_1 \underbrace{\mathbf{U}_{pess}}_{\mathbf{U}_1} \times_2 \underbrace{\mathbf{U}_{pos}}_{\mathbf{U}_2} \times_3 \underbrace{\mathbf{U}_{ilum}}_{\mathbf{U}_3} \times_4 \underbrace{\mathbf{U}_{exp}}_{\mathbf{U}_4} \times_5 \underbrace{\mathbf{U}_{pixels}}_{\mathbf{U}_5} \quad (13)$$

onde $\mathcal{Z}$ possui as mesmas dimensões e ordem de $\mathcal{D}$, as matrizes fatores $\mathbf{U}_{pess}, \mathbf{U}_{pos}, \mathbf{U}_{ilum}, \mathbf{U}_{exp}$ e $\mathbf{U}_{pixels}$ possuem dimensões $N_{pess} \times N_{pess}$, $N_{pos} \times N_{pos}$, $N_{ilum} \times N_{ilum}$, $N_{exp} \times N_{exp}$ e $N_{pixels} \times N_{pixels}$ respectivamente.

Matematicamente assumir multilinearidade é assumir linearidade, ou seja, no contexto de reconhecimento facial, o método *Eigenfaces* [2] baseado na análise de componentes principais é parte constituinte do *TensorFaces*, onde cada coluna de $\mathbf{U}_{pixels}$ é um eigenface, pois os mesmos foram calculados executando-se o SVD (decomposição em valores singulares) no modo-5 matriciado $\mathbf{D}_{(pixels)} \in \mathbb{R}^{N_{pixels} \times N_{exp} N_{ilum} N_{pos} N_{pess}}$ do tensor $\mathcal{D}$.

A matriz $\mathbf{U}_{pess}$ é a base do espaço de parâmetros das pessoas, em que cada pessoa na base de dados pode ser representada por um único vetor [8], no qual contém os coeficientes com respeito as bases que serão extraídas do tensor:

$$\mathcal{B} = \mathcal{Z} \times_2 \mathbf{U}_{pos} \times_3 \mathbf{U}_{ilum} \times_4 \mathbf{U}_{exp} \times_5 \mathbf{U}_{pixels} \quad (14)$$

com dimensões idênticas a $\mathcal{Z}$. Note que:

$$\mathcal{D} = \mathcal{B} \times_1 \mathbf{U}_{pess} \Rightarrow \mathbf{D}_{(pess)} = \mathbf{U}_{pess} \mathbf{B}_{(pess)} \quad (15)$$

onde $\mathbf{D}_{(pess)}$ e $\mathbf{B}_{(pess)} \in \mathbb{R}^{N_{pess} \times N_{pixels} N_{exp} N_{ilum} N_{pos}}$. Veja que:

$$\mathbf{d}_1 = \mathbf{D}_{(pixels)}(:, 1) = [\mathbf{D}_{(pess)}(\ 1, 1 : N_{pixels})]^T \quad (16)$$

é a primeira imagem vetorizada do tensor $\mathcal{D}$. De maneira mais geral temos que $\mathbf{d}(i_1, i_2, i_3, i_4)$ é a $(i_1 i_2 i_3 i_4)$-ésima imagem da pessoa i_1, na posição facial i_2, com condição de iluminação i_3 e expressão i_4, onde $i_1 = 1, 2, \ldots, N_{pess}$, $i_2 = 1, 2, \ldots, N_{pos}$, $i_3 = 1, 2, \ldots, N_{ilum}$ e $i_4 = 1, 2, \ldots, N_{exp}$. A partir da Equação (15) obtemos

$$\mathbf{d}(i_1, i_2, i_3, i_4) = \mathbf{B}_{(pess)}^T(i_2, i_3, i_4)\mathbf{u}(i_1) \quad (17)$$

em que $\mathbf{u}(i_1)$ é a i_1-ésima coluna de $\mathbf{U}_{(pess)}^T$, e $\mathbf{B}_{(pess)}(i_2, i_3, i_4)$ é um slice de $\mathcal{B}$, de dimensões $N_{pess} \times N_{pixels}$, obtido pela fixação dos índices i_2, i_3, i_4.

Dado uma imagem de teste $\mathbf{d}_{teste}$ utilizaremos as bases $\mathbf{B}_{(pess)}^T(i_2, i_3, i_4)$ no modo pessoa para projetar $\mathbf{d}_{teste}$ no espaço de parâmetros das pessoas:

$$\hat{\mathbf{u}}_{pess}(i_2, i_3, i_4) = (\mathbf{B}_{(pess)}^T(i_2, i_3, i_4))^{\dagger} \mathbf{d}_{teste} \quad (18)$$

O reconhecimento da imagem $\mathbf{d}_{teste}$ como sendo a pessoa i_1^* consiste em minimizar:

$$i_1^* = \arg \min_{\{i_1, i_2, i_3, i_4\}} \| \hat{\mathbf{u}}_{pess}(i_2, i_3, i_4) - \mathbf{u}(i_1) \|^2 \quad (19)$$

V. MÉTODO PROPOSTO

Nosso método baseia-se em dois fatos, primeiramente na capacidade das wavelets de Gabor extrair e discriminar informações relevantes de uma imagem, onde cada detalhe é capturado devido o conjunto de filtros cobrirem uma ampla variedade de escalas e orientações. Será utilizado o resultado da convolução das wavelets de Gabor com as imagens faciais, operação que será realizada no domínio da frequência fazendo-se uso do Teorema da convolução:

$$\varphi_{u,v}(x, y) = \mathfrak{F}^{-1}\{\mathfrak{F}\{\psi_{u,v}(x, y)\}\mathfrak{F}\{I(x, y)\}\} \quad (20)$$

Após a convolução, cada resposta $\varphi_{u,v}(x,y)$ é decimada por um fator $\delta = 8$, o próximo passo é a vetorização e concatenação das respostas decimadas $\varphi^{\delta}_{u,v}(x,y)$ gerando o vetor $\mathbf{g}$ de características. O segundo fato nos remete ao HOSVD, que ortogonaliza os vários espaços coluna relacionados com cada modo, o que no contexto de reconhecimento facial representa para um conjunto de imagens a separação dos fatores constituintes da sua natureza multimodal. Desta forma, estamos extraindo características com a maior quantidade de informação possível [15] e separando-as por fatores constituintes, tais como: ponto de vista, iluminação e expressão. No *TensorFaces* cada imagem vetorizada $vec(I(x,y))$ é uma coluna de $\mathbf{D}_{(pixels)}$, ao fazermos uso do vetor de características $\mathbf{g}$ estaremos substituindo a imagem vetorizada, gerando um tensor $\mathcal{T} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_5}$ de vetores de características. Note que o vetor $\mathbf{g}$ possui maior capacidade de descrição do que $vec(I(x,y))$ propriamente, proporcionando uma modelagem para $\mathcal{T}$ mais simples, podendo haver fatores com dimensões reduzidas, ou até mesmo com ordem reduzida, caso em que o fator multimodal possui dimenssão 1.

A. Reconhecimento

Assim como no *TensorFaces* e na análise multilinear de uma maneira geral, assume-se uma mistura probabilística do PCA, no caso de imagens faciais esse fato é análogo a modelos baseados em posição facial (*view-based*) quando há diferentes escolhas de bases [17], [18]. O vetor $\mathbf{d}_{teste}$ é substituído pelo vetor de características para teste $\mathbf{g}_{teste}$.

Em essência, a análise multilinear fornece para cada ponto de vista uma gaussiana multidimensional [7], considerando que nossa decisão é baseada em uma função densidade de probabilidade normal [19], podemos derivar dois métodos para classificação.

1) Método 1 – Distância Euclideana Mínima: Classificador que foi utilizado em conjunto com o método *TensorFaces*, quando as características são estatísticamente independentes, e possuem a mesma variância σ, temos que a matriz de covariância é $\sigma^2\mathbf{I}$ para cada classe, ou seja, $\mathbf{C}_i = \sigma^2\mathbf{I}$, esse caso é equivalente a distância euclideana, Equação (19), em que:

$$\hat{\mathbf{u}}_{pess}(i_2,i_3,i_4) = (\mathbf{B}^T_{(pess)}(i_2,i_3,i_4))^{\dagger}\mathbf{g}_{teste} \qquad (21)$$

2) Método 2 – Distância Mahalanobis Mínima, usando matriz de covariância agregada: Considere agora que a matriz de covariância é diferente para todas as classes $\mathbf{C} = \mathbf{C}(i_1)$. É necessário calcularmos as matrizes de covariâncias $\mathbf{C}(i_1)$ no espaço de parâmetros das pessoas. Para assim procedermos projetamos cada vetor de característica $\mathbf{g}(i_1,i_2,i_3,i_4)$ relacionado com cada imagem:

$$\mathbf{g}_{(pess)}(i_1,i_2,i_3,i_4) = (\mathbf{B}^T_{(pess)}(i_2,i_3,i_4))^{\dagger}\mathbf{g}(i_1,i_2,i_3,i_4) \qquad (22)$$

onde $\mathbf{g}_{(pess)}(i_1,i_2,i_3,i_4) \in \mathbb{R}^{N_{pess}\times 1}$, seja $\mathbf{A}(i_1) \in \mathbb{R}^{N_{pess}\times N_{exp}N_{ilum}N_{pos}}$ a matriz de todos os vetores características projetados da pessoa i_1, então $\mathbf{C}(i_1) = \mathbb{E}[\mathbf{A}(i_1)\mathbf{A}^T(i_1)] \in \mathbb{R}^{N_{pess}\times N_{pess}}$. A matriz de covariância agregada $\mathbf{C}$ é dada por:

$$\mathbf{C} = \frac{1}{N}\sum_{i_1=1}^{N_{pess}} N_{i_1}\mathbf{C}(i_1) \qquad (23)$$

tal que N_{i_1} é o número de exemplos de treinamentos da i_1-ésima classe e $N = \sum_{i_1=1}^{N_{pess}} N_{i_1}$ é o número total de exemplos de treinamento. Assim, o reconhecimento de uma imagem de teste $\mathbf{d}_{teste}$ consiste em minimizar:

$$i_1^* = \arg\min_{\{i_1,i_2,i_3,i_4\}}[(\hat{\mathbf{u}}_{pess}(i_2,i_3,i_4) - \mathbf{u}(i_1))^T\mathbf{C}^{-1}(\hat{\mathbf{u}}_{pess}(i_2,i_3,i_4) - \mathbf{u}(i_1))] \qquad (24)$$

VI. Resultados Experimentais

Foi utilizado parte da base de dados Weizmann em http://www.wisdom.weizmann.ac.il/∼vision/FaceBase/, num total de 1080 imagens de 24 indivíduos em 5 posições $(0^o, \pm 17^o, \pm 34^o)$, 3 condições de iluminação e 3 expressões faciais. Os experimentos foram organizados de maneira que fosse possível formar uma bateria de rodadas de treinamento e teste, tornando possível calcularmos estatísticas, como a taxa média de reconhecimento. Na organização do conjunto de treinamento temos possibilidades relacionadas com a variação da dimensão de cada fator multimodal, por exemplo, uma possível configuração para o treinamento seria um conjunto contendo 24 indíviduos em 3 posições, 2 condições de iluminação e 2 expresões. Para os experimentos as faces não foram recortadas conforme vemos na Figura 2 onde temos um indíviduo do banco de dados em 3 posições faciais com a mesma expressão facial e condição de iluminação.

A. Resultados

Considerando todas as possíveis variações de um fator multimodal por vez, foi construído 42 rodadas de treinamento e teste, onde constatamos um aumento na taxa de reconhecimento em casos que há uma pequena quantidade de imagens para treinamento e variações de condições de iluminação. Na Tabela 1 temos a média de reconhecimento considerando todas as rodadas, de maneira que o Método 1 mostrou-se o mais robusto as variações dos fatores multimodais.

Analisando a influência dos fatores multimodais, notou-se que para variações de iluminação e posições o método *TensorFaces* sofre perda na taxa de reconhecimento facial, principalmente quando há apenas uma condição de iluminação

Fig. 2. Imagens de um indíviduo nas posições $-34^o, 0^o, +34^o$ respectivamente, em uma dada condição de iluminação e expressão facial.

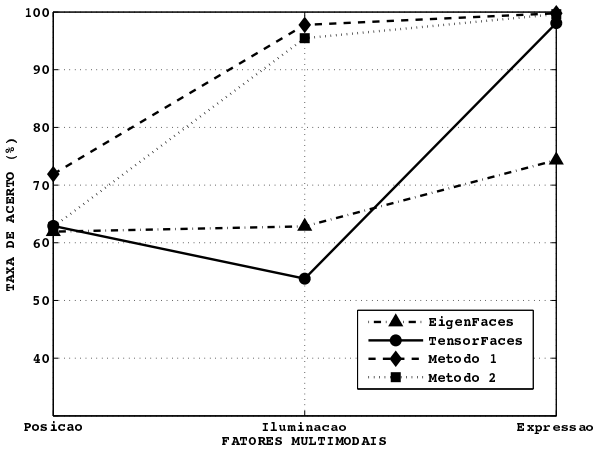

Fig. 3. Taxa de reconhecimento médio para variações de cada fator multimodal.

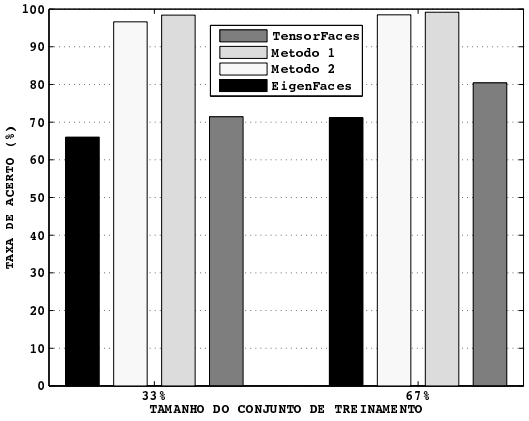

Fig. 4. Taxa de reconhecimento médio para variações do tamanho do conjunto de treinamento considerando variações de expressão e iluminação.

TABELA I

TAXA DE ACERTO CONSIDERANDO TODAS AS RODADAS.

Método	Acerto
Eigenfaces	64%
TensorFaces	67%
Método 1	**80%**
Método 2	73%

ou posição no treinamento, os métodos propostos obtiveram resultados melhores que o *TensorFaces* nessas configurações do conjunto de treinamento. Consequentemente para a variação de cada fator multimodal a taxa de reconhecimento médio dos métodos 1 e 2 apresentaram resultados superiores ao *TensorFaces* e *Eigenfaces* (Figura 3). Quando há mudança no tamanho do conjunto de treinamento devido a variação nas condições de iluminação e expressão, notou-se estabilidade nos métodos propostos, mesmo quando utiliza-se apenas um terço do total de imagens para treino (Figura 4).

Na Figura 5 temos a análise do comportamento dos métodos propostos, levando em consideração a variação na quantidade de posições faciais para treino, conjuntos com $1, 2, 3$ e 4 posições foram criados, representando respectivamente $20\%, 40\%, 60\%$ e 80% do total de imagens. Apesar de todos os algoritmos simulados apresentarem dificuldades com uma

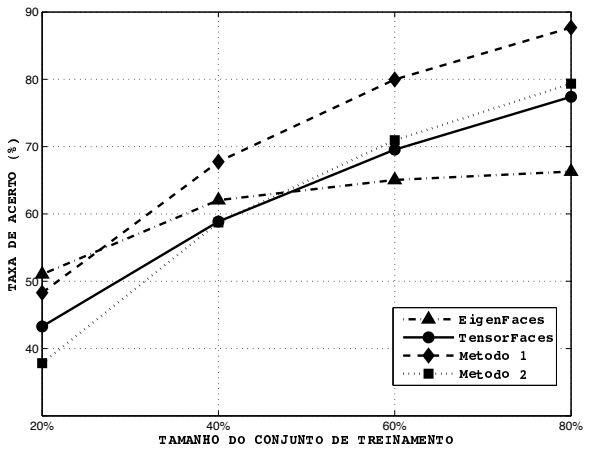

Fig. 5. Taxa de reconhecimento médio para variações do tamanho do conjunto de treinamento considerando variações de posição facial.

pequena quantidade de imagens para treino, o Método 1 mostrou uma curva crescente com a maior taxa de acerto.

REFERÊNCIAS

[1] L. G. Brown, *A survey of image registration techniques. ACM Computing Surveys*, v. 24, n. 4, pp. 325-376, 1992

[2] Turk, M. and Pentland, A., Eigenfaces for Recognition," *Journal of Cognitive Neuroscience*, v. 3, n. 1, pp. 71–86, 1991.

[3] De Lathauwer, L., *Signal processing based on multilinear algebra*, PhD thesis, Addison Katholieke Univ. Leuven, , Belgium, 1997.

[4] Kolda, T., Orthogonal tensor decompositions," *SIAM J. Matrix Anal. Appl.*, v. 23, n. 1, pp. 243–255, 2001.

[5] P. Comon and X. Luciani and A. L. F. de Almeida, "Tensor Decompositions, Alternating Least Squares and Other Tales", " *Journal of Chemometrics*, v. 23, n. 7-8, pp. 393–405, 2009.

[6] De Lathauwer, L., A multilinear singular value decomposition," *SIAM J. Matrix Anal. Appl.*, v. 21, n. 4, pp. 1253-1278, 2000.

[7] Vasilescu, M.A.O. and Terzopoulos, D., Multilinear Analysis of Image Ensembles: TensorFaces," *in Proc. Eur. Conf. Comput. Vis., Copenhagen*, v.1, pp. 447-460, 2002.

[8] M. Alex O. Vasilescu and Demetri Terzopoulos, Multilinear image analysis for facial recognition," *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, v. 2, pp. 511-514, 2002.

[9] Mallat, Stéphane, *A wavelet tour of signal processing*, Academic Press, 2ª ed, 1999.

[10] Tai Sing Lee, Image Representation Using 2D Gabor Wavelets," *IEEE Trans. Pattern Analysis and Machine Intelligence*, v. 18, n. 10, pp. 959–971, 1996.

[11] Shen, LinLin and Bai, Li and Fairhurst, Michael, Gabor wavelets and General Discriminant Analysis for face identification and verification," *Image Vision Comput.*, v. 25, n. 5, pp. 553–563, 2007.

[12] Štruc, Vitomir and Pavešić, Nikola, Gabor-Based Kernel Partial-Least-Squares Discrimination Features for Face Recognition," *Informatica*, v. 20, n. 1, pp. 115–138, 2009.

[13] Kiers, H. A., Towards a standardized notation and terminology in multiway analysis," *J. Chemometrics*, v. 14, n. 3, pp. 105-122, 2000.

[14] Tamara G. Kolda, Multilinear operators for higher-order decompositions, Sandia National Laboratories, Albuquerque, April, 2006.

[15] Okajima, K., Two Dimensional Gabor-type receptive field as derived by mutual information maximization," *Neural Networks*, v. 11, n. 3, pp. 441–447, 1998.

[16] Kolda, Tamara G. and Bader, Brett W., Tensor Decompositions and Applications," *SIAM Rev.*, v. 51, n. 3, pp. 455–500, 2009.

[17] A. Pentland and B. Moghaddam, L., View-based and modular eigenspaces for face recognition," *In Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 84-91, 1994.

[18] M. E. Tipping and C. M. Bishop, Mixtures of probabilistics principal component analysers," *Neural Computation*, v. 11, n. 2, pp. 443-482, 1999.

[19] Richard O. Duda, Peter E. Hart and David G. Stork, Pattern Classification, Wiley Interscience, 2ª ed., 2001.