

ÁLGEBRA MULTILINEAR APLICADA AO RECONHECIMENTO FACIAL

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

Universidade Federal do Ceará

{dario, andre}@gtel.ufc.br

Resumo – Nesse artigo de revisão o problema de reconhecimento facial é investigado do ponto de vista da álgebra multilinear, mais especificamente por meio de decomposições tensoriais fazendo uso das wavelets de Gabor. A extração de características ocorre em dois estágios: primeiramente as wavelets de Gabor são aplicadas de maneira holística na seleção de características; em segundo as imagens faciais são modeladas como um tensor de ordem superior de acordo com o fatores multimodais presentes. Com isso aplicamos a decomposição tensorial Higher Order Singular Value Decomposition (HOSVD) para separar os fatores que influenciam na formação das imagens. O método de reconhecimento facial proposto possui uma alta taxa de acerto e estabilidade quando há variação nos diversos fatores multimodais, tais como, posição facial, condição de iluminação e expressão facial. Propomos ainda uma maneira sistemática para realização da validação cruzada em modelos tensoriais para estimação da taxa de erro em sistemas de reconhecimento facial que exploram a natureza multilinear do conjunto de imagens. Através do particionamento aleatório dos dados organizado como um tensor, a validação cruzada modo- n proporciona a criação de folds extraindo subtensores no modo desejado, caracterizando um método estratificado e susceptível a repetições da validação cruzada com diferentes particionamentos.

Palavras-chave – Reconhecimento Facial, Wavelets de Gabor, Álgebra Multilinear, Decomposições Tensoriais, Validação Cruzada.

Abstract – In this review, the face recognition problem is investigated from the standpoint of multilinear algebra, more specifically the tensor decomposition, and by making use of Gabor wavelets. The feature extraction occurs in two stages: first the Gabor wavelets are applied holistically in feature selection; Secondly facial images are modeled as a higher-order tensor according to the multimodal factors present. Then, the HOSVD is applied to separate the multimodal factors of the images. The proposed facial recognition approach exhibits higher average success rate and stability when there is variation in the various multimodal factors such as facial position, lighting condition and facial expression. We also propose a systematic way to perform cross-validation on tensor models to estimate the error rate in face recognition systems that explore the nature of the multimodal ensemble. Through the random partitioning of data organized as a tensor, the mode- n cross-validation provides folds as subtensors extracted of the desired mode, featuring a stratified method and susceptible to repetition of cross-validation with different partitioning.

Keywords – Face Recognition, Gabor Wavelets, Multilinear Algebra, Tensor Decomposition, Cross-Validation.

1 INTRODUÇÃO

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 [1]. O interesse em álgebra multilinear tem se expandido para diversas áreas do conhecimento, incluindo o processamento de imagens para o reconhecimento de padrões e visão computacional. Um ramo da álgebra multilinear são as decomposições tensoriais [2, 3], em especial a HOSVD que faz uso da Decomposição em Valores Singulares (SVD) para seu cálculo [4], 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 [5, 6]. Imagens são sinais na qual sua formação depende da 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 [7]. Ao longo das últimas décadas, diversos algoritmos para reconhecimento facial foram produzidos, explorando ao máximo a riqueza de informação contida nos múltiplos domínios [8, 9]. Devido a essa natureza multimodal, os fatores constituintes se misturam. Fato imperceptível aos seres humanos devido a nossa percepção ser tolerante aos diversos componentes que formam a imagem, tornando-se um sistema robusto a essas variações. O mesmo não é verdade para sistemas computacionais de reconhecimento facial. Em reconhecimento facial os métodos lineares [10, 11], por exemplo, EigenFaces [12], mostraram-se ferramentas robustas quando há variação apenas da identidade do indivíduo no conjunto de imagens e são fixados a posição facial, condições de iluminação e expressão.

2 CONCEITOS IMPORTANTES DE ÁLGEBRA LINEAR

Trataremos a álgebra linear do ponto de vista matricial e introduziremos as operações matriciais mais importantes para o tratamento tensorial. Nesse artigo \mathbb{F}^n denotará o espaço n -dimensional sobre o corpo $\mathbb{F} = \mathbb{R}$ ou \mathbb{C} .

Definição 1 (Produto de Hadamard) Seja $A, B \in \mathbb{F}^{m \times n}$, definimos o produto de Hadamard como

$$A * B = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & \dots & a_{1n}b_{1n} \\ a_{21}b_{21} & a_{22}b_{22} & \dots & a_{2n}b_{2n} \\ \dots & \dots & \ddots & \dots \\ a_{m1}b_{m1} & a_{m2}b_{m2} & \dots & a_{mn}b_{mn} \end{bmatrix}, \quad (1)$$

que em forma escalar fica $[A * B]_{ij} = a_{ij}b_{ij}$. Segue ainda que $A * B \in \mathbb{F}^{m \times n}$.

Definição 2 (Produto de Kronecker) O produto de Kronecker de duas matrizes $A \in \mathbb{F}^{m \times n}$ e $B \in \mathbb{F}^{p \times q}$ denotado por $A \otimes B \in \mathbb{F}^{mp \times nq}$ é definido como:

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \dots & a_{1n}B \\ a_{21}B & a_{22}B & \dots & a_{2n}B \\ \dots & \dots & \ddots & \dots \\ a_{m1}B & a_{m2}B & \dots & a_{mn}B \end{bmatrix}, \quad (2)$$

o termo $a_{ij}B$ produz um bloco na matriz $A \otimes B$ bem definido. Na forma escalar temos $[A \otimes B]_{(i-1)p+1:ip, (j-1)q+1:jq} = a_{ij}B$.

Definição 3 (Produto Externo) Definimos o produto externo de dois vetores $u \in \mathbb{F}^m$ e $v \in \mathbb{F}^n$, como:

$$u \circ v = uv^\top = \begin{bmatrix} u_1v_1 & u_1v_2 & \dots & u_1v_n \\ u_2v_1 & u_2v_2 & \dots & u_2v_n \\ \dots & \dots & \ddots & \dots \\ u_mv_1 & u_mv_2 & \dots & u_mv_n \end{bmatrix} \text{ em que } u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{bmatrix} \text{ e } v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}. \quad (3)$$

A representação em forma escalar fica $[u \circ v]_{ij} = u_iv_j$ para $i = 1, 2, \dots, m$ e $j = 1, 2, \dots, n$.

Definição 4 (Matriz rank-1) O rank de uma matriz $A \in \mathbb{F}^{m \times n}$ denotado por $r(A)$ é o número máximo de colunas (ou linhas) Linearmente Independente (LI). Uma matriz A é dita rank-1, i.e., $r(A) = 1$ se a mesma for composta pelo produto externo de dois vetores $u = [u_1, u_2, \dots, u_m]^\top$ e $v = [v_1, v_2, \dots, v_n]^\top$:

$$A = u \circ v = \begin{bmatrix} u_1v_1 & u_1v_2 & \dots & u_1v_n \\ u_2v_1 & u_2v_2 & \dots & u_2v_n \\ \dots & \dots & \ddots & \dots \\ u_mv_1 & u_mv_2 & \dots & u_mv_n \end{bmatrix} \quad (4)$$

Definição 5 (Soma Direta) Se U e V são espaços vetoriais (sobre o mesmo corpo), sua soma direta é o espaço vetorial W , denotado por $U \oplus V$, no qual seus elementos são os pares ordenados $[x, y]^\top$ com $x \in U$ e $y \in V$, com as operações lineares abaixo bem definidas

$$\alpha_1 \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \alpha_2 \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} \alpha_1x_1 + \alpha_2x_2 \\ \alpha_1y_1 + \alpha_2y_2 \end{bmatrix} \quad (5)$$

em que $x_1, x_2 \in U$, $y_1, y_2 \in V$ e $[x_1, y_1]^\top, [x_2, y_2]^\top \in W = U \oplus V$.

O espaço dos vetores $[x, 0]^\top$ é um subespaço de W isomorfo a U , similarmente $y \in V$ pode ser identificado como os vetores $[0, y]^\top \in W$. Então, para todos os propósitos podemos considerar U e V como subespaços de W , o que nos leva ao seguinte:

Teorema 1 Seja U e V espaços vetoriais. A dimensão da soma direta de U e V é a soma de suas dimensões.

Prova 1 A prova pode ser encontrada na pág. 30 de [13].

Teorema 2 (Singular Value Decomposition) Seja $A \in \mathbb{F}^{m \times n}$ com rank $r(A) = R$. Então existem matrizes $U_1 \in \mathbb{F}^{m \times r}$, $U_2 \in \mathbb{F}^{n \times r}$ e $S \in \mathbb{F}^{r \times r}$ tal que $U_1U_1^\top = U_2U_2^\top = I_r$, ou seja, U_1 e U_2 são matrizes ortogonais e S é uma matriz diagonal de valores singulares $S = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_R^2)$. Assim, A pode ser escrita como,

$$A = U_1SU_2^\top \quad (6)$$

Prova 2 A prova pode ser encontrada na pág. 19 de [14].

Em matemática aplicada, a aproximação de baixo rank, do inglês, Low Rank Approximation (LRA), é um problema de minimização em que a função custo mede o quanto podemos aproximar uma matriz A por outra \hat{A} de rank menor.

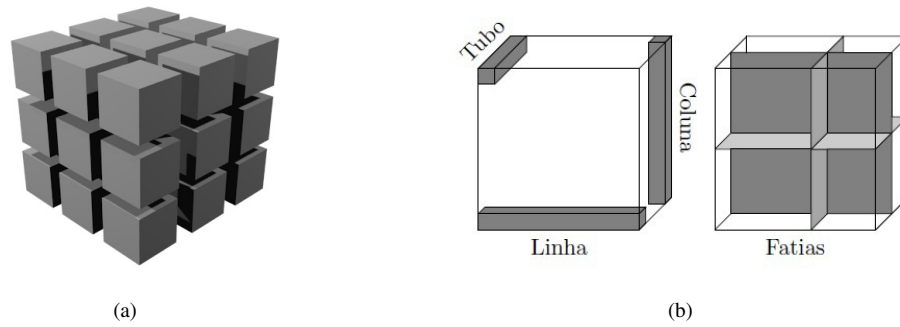


Figura 1: (a) Tensor de ordem 3, (b) Exemplos de fibras (linhas, colunas e tubos) e fatias de um tensor de ordem 3.

Teorema 3 (Teorema de Eckart-Young) *Seja $\mathbf{A} \in \mathbb{R}^{m \times n}$ com $\text{rank } r(\mathbf{A}) = R \leq p = \min\{m, n\}$, e seja*

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^\top = \sum_{i=1}^R \sigma_i^2 \mathbf{u}_i \mathbf{v}_i^\top \quad (7)$$

sua decomposição em valores singulares. Defina,

$$\mathbf{A}_k = \mathbf{U}_k \mathbf{S}_k \mathbf{V}_k^\top = \sum_{i=1}^k \sigma_i^2 \mathbf{u}_i \mathbf{v}_i^\top \quad (8)$$

então $\hat{\mathbf{A}} = \mathbf{A}_k$ é a aproximação ótima de \mathbf{A} com função custo

$$\min_{r(\hat{\mathbf{A}})=k} \|\mathbf{A} - \hat{\mathbf{A}}\|_F = \|\mathbf{A} - \mathbf{A}_k\|_F \quad (9)$$

Em que $\|\mathbf{A}\|_F = \sqrt{\sum_{i,j} a_{ij}^2}$ é a norma de Frobenius.

Prova 3 *A prova pode ser encontrada em [15].*

3 ÁLGEBRA MULTILINEAR

Nesta seção será realizada uma revisão sobre álgebra multilinear e decomposições tensoriais. Serão apresentados conceitos fundamentais, formalismo matemático, notações e operações envolvendo tensores. Na maioria das definições focamos no uso de coordenadas devido a natureza numérica dos problemas práticos em reconhecimento de padrões. Primeiramente abordaremos o conceito de produto tensorial de maneira informal:

Definição 6 (Produto Tensorial de Vetores) *Seja \mathbf{u} um vetor com m coordenadas e \mathbf{v} um vetor com n coordenadas, a saber $\{u_i\}$ e $\{v_j\}$ respectivamente. O produto tensorial de \mathbf{u} e \mathbf{v} denotado por $\mathbf{u} \bullet \mathbf{v}$, é um vetor com mn coordenadas dadas por $\{u_i v_j\}$.*

Definição 7 (Produto Tensorial de Espaços Vetoriais) *Suponha que um produto tensorial $\mathbf{u} \bullet \mathbf{v}$ esteja bem definido para $\mathbf{u} \in U \subset \mathbb{F}^m$ e $\mathbf{v} \in V \subset \mathbb{F}^n$. O produto tensorial de $U \bullet V$ é definido como o espaço vetorial de todas as combinações lineares dos vetores da forma $\mathbf{x} \bullet \mathbf{y}$ onde $\mathbf{x} \in U$ e $\mathbf{y} \in V$.*

Observe que o produto externo (3) e o produto de Kronecker (2) são produtos tensoriais. Como exemplo, seja o espaço vetorial das matrizes $\mathbf{A} \in \mathbb{F}^{m \times n}$ e $\mathbf{B} \in \mathbb{F}^{p \times q}$, respectivamente. Então o conjunto de todas as combinações lineares das matrizes do tipo $\mathbf{A} \otimes \mathbf{B} \in \mathbb{F}^{m \times p \times n \times q}$ é um espaço vetorial. Dito isso, podemos observar que dados dois produtos tensoriais entre U e V , a saber, $U \bullet_1 V$ e $U \bullet_2 V$, existe um único isomorfismo L [16–18]. Desse modo, nos restringiremos aos produtos externo (3) e o de Kronecker (2). A álgebra multilinear também é fundamentada no conceito de transformação multilinear, que definimos de acordo com [19].

Definição 8 (Transformação Multilinear) *Uma transformação*

$$\psi : U_1 \times U_2 \times \dots \times U_N \mapsto U$$

é multilinear se for linear em cada variável, i.e., para todo $\mathbf{u}_i, \mathbf{v} \in U_i$, $i = 1, \dots, N$ e para todo $\alpha \in \mathbb{F}$ temos

$$\psi(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_i + \mathbf{v}, \dots, \mathbf{u}_N) = \psi(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_i, \dots, \mathbf{u}_N) + \psi(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{v}, \dots, \mathbf{u}_N) \quad (10a)$$

$$\psi(\mathbf{u}_1, \mathbf{u}_2, \dots, \alpha \mathbf{u}_i, \dots, \mathbf{u}_N) = \alpha \psi(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_i, \dots, \mathbf{u}_N) \quad (10b)$$

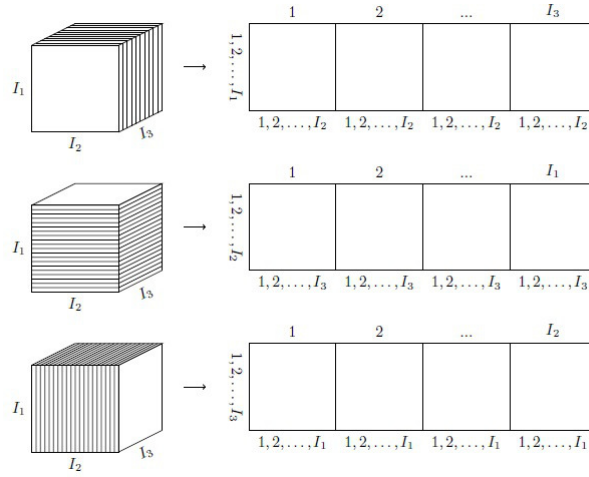


Figura 2: Matriciação de um tensor de ordem 3 nos modo-1, modo-2 e modo-3, respectivamente.

Iremos definir tensores de ordem superior da mesma forma em que foram propostos em [1, 20–22].

Definição 9 (Tensor de ordem N) Seja U_1, U_2, \dots, U_N espaços vetoriais de dimensões I_1, I_2, \dots, I_N respectivamente. Considere N vetores $\mathbf{u}_1 \in U_1, \mathbf{u}_2 \in U_2, \dots, \mathbf{u}_N \in U_N$. Denotamos por $\mathbf{u}_1 \circ \mathbf{u}_2 \circ \dots \circ \mathbf{u}_N$ a transformação multilinear sobre $U_1 \times U_2 \times \dots \times U_N$, definida por

$$(\mathbf{u}_1 \circ \mathbf{u}_2 \circ \dots \circ \mathbf{u}_N)(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) = \langle \mathbf{u}_1, \mathbf{x}_1 \rangle_1 \langle \mathbf{u}_2, \mathbf{x}_2 \rangle_2 \dots \langle \mathbf{u}_N, \mathbf{x}_N \rangle_N \quad (11)$$

onde $\langle \mathbf{u}_i, \mathbf{x}_i \rangle_i$ é o produto interno de \mathbf{u}_i com um vetor arbitrário $\mathbf{x}_i \in U_i$.

O espaço gerado por todos os elementos $\mathbf{u}_1 \circ \mathbf{u}_2 \circ \dots \circ \mathbf{u}_N$ é chamado de espaço tensorial sobre $U_1 \times U_2 \times \dots \times U_N$. Um elemento desse espaço tensorial é chamado de tensor de ordem N . Assim, um tensor $\mathcal{A} \in \mathbb{F}^{I_1 \times I_2 \times \dots \times I_N}$ pode ser visto como um *array multidimensional* associado a um produto tensorial de N espaços vetoriais [1] em que seus componentes ou entradas $a_{i_1 i_2 \dots i_N}$ 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. A Figura 1(a) representa um tensor de terceira ordem. Cada elemento do tensor é acessado por três índices i, j e k . Incorporado ao conceito de tensor temos a seguinte definição:

Definição 10 Seja \mathcal{A} de ordem N , um subtensor $\mathcal{B} = \mathcal{A}_{i_n = \alpha}$ de ordem $N - 1$ é obtido quando fixamos o índice n em α . De uma maneira mais geral 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 (ver Figura 1(b)). A operação de matriciação ou desdobramento (do inglês, *unfolding* ou *matricization*) de tensores em matrizes foi proposta por [4, 23], de modo que as fibras de uma determinada dimensão serão as colunas da matriz resultante. A matriciação de um tensor $\mathcal{A} \in \mathbb{F}^{I_1 \times I_2 \times \dots \times I_N}$ é definida como em [24]: Seja $\mathbb{N} = \{1, 2, \dots, N\}$ o conjunto de modos presentes no tensor \mathcal{A} e $\mathbb{I}_N = \{I_1, I_2, \dots, I_N\}$ o conjunto dos tamanhos de cada modo. Considere os dois conjuntos ordenados $\mathbb{R} = \{r_1, r_2, \dots, r_L\}$ e $\mathbb{C} = \{c_1, c_2, \dots, c_M\}$ uma partição de \mathbb{N} , a matriciação (*unfolding*) de um tensor pode ser definida como um remodelamento (*reshape*) de \mathcal{A} tal que

$$\mathcal{A}_{(\mathbb{R} \times \mathbb{C})} \in \mathbb{F}^{J \times K} \text{ em que } J = \prod_{n \in \mathbb{R}} I_n \text{ e } K = \prod_{n \in \mathbb{C}} I_n \quad (12)$$

Dito isso, definimos a matriciação modo- n como:

Definição 11 (matriciação modo- n) Seja $\mathbb{R} = \{r_1\}$ uma parte de \mathbb{N} com um único elemento $r_1 = n$ e $\mathbb{C} = \{c_1, c_2, \dots, c_{n-1}\}$, então as fibras do modo- n são as colunas de $\mathcal{A}_{(\mathbb{R} \times \mathbb{C})}$ (veja Figura 2). A forma matriciada $\mathcal{A}_{(n)}$ é definida como

$$\mathcal{A}_{(n)} \equiv \mathcal{A}_{(\mathbb{R} \times \mathbb{C})}, \mathbb{R} = \{n\} \text{ e } \mathbb{C} = \{1, \dots, n-1, n+1, \dots, N\} \quad (13)$$

Definição 12 (produto modo- n) O produto modo- n de um tensor $\mathcal{A} \in \mathbb{F}^{I_1 \times I_2 \times \dots \times I_N}$ com uma matriz $\mathbf{B} \in \mathbb{C}^{J_n \times I_n}$, denotado por $\mathcal{C} = \mathcal{A} \times_n \mathbf{B}$ é definido como:

$$[\mathcal{A} \times_n \mathbf{B}]_{i_1 i_2 \dots j_n \dots i_N} = \sum_{i_n=1}^{I_n} [\mathcal{A}]_{i_1 i_2 \dots i_n \dots i_N} [\mathbf{B}]_{j_n i_n} \quad (14)$$

que pode ser expresso em forma matricial como

$$\mathbf{C}_{(n)} = \mathbf{B}\mathbf{A}_{(n)} \quad (15)$$

Seja $\mathbf{C} \in \mathbb{R}^{J_m \times I_m}$, note que as seguintes propriedades seguem diretamente da definição

$$\mathbf{A} \times_n \mathbf{B} \times_m \mathbf{C} = \mathbf{A} \times_m \mathbf{C} \times_n \mathbf{B} \quad (16)$$

3.1 Decomposições Tensoriais

O conceito de decomposição tensorial surgiu em 1927 através dos trabalhos de Hitchcock [25, 26] e Cattell em 1944 [27]. Posteriormente, esses conceitos tiveram maior desenvolvimento no contexto da Psicometria com o modelo desenvolvido por Carroll & Chang [28] denominado *Canonical Decomposition (CANDECOMP)* e com Harshman [29], que de forma independente propôs um modelo semelhante, denominado *Parallel Factor Analysis (PARAFAC)*. Mais recentemente L. De Lathauwer propôs um modelo de decomposição tensorial que estende o conceito de SVD para tensores de ordem superior [4].

Revisitando a SVD

Formulemos a SVD na forma de produto modo- n , a fim de facilitar o entendimento da HOSVD, aqui tomaremos $\mathbb{F} = \mathbb{C}$, pois é o caso mais geral.

Teorema 4 (SVD) *Qualquer matriz $\mathbf{A} \in \mathbb{C}^{I_1 \times I_2}$ pode ser escrita como o produto*

$$\mathbf{A} = \mathbf{U}_1 \mathbf{S} \mathbf{V}_2^H = \mathbf{S} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \quad (17)$$

em que

1. $\mathbf{U}_1 = (\mathbf{u}_1^{(1)}, \mathbf{u}_2^{(1)}, \dots, \mathbf{u}_{I_1}^{(1)})$ é uma matriz unitária.
2. $\mathbf{U}_2 = (\mathbf{u}_1^{(2)}, \mathbf{u}_2^{(2)}, \dots, \mathbf{u}_{I_2}^{(2)}) (= \mathbf{V}_2^H)$ é uma matriz unitária.
3. $\mathbf{S} \in \mathbb{C}^{I_1 \times I_2}$ é uma matriz com as seguintes propriedades:

1. pseudo-diagonalidade

$$\mathbf{S} = \text{diag} \{ \sigma_1, \sigma_2, \dots, \sigma_{\min\{I_1, I_2\}} \}$$

2. ordenação

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{\min\{I_1, I_2\}} \geq 0$$

Os valores σ_i são os autovalores de \mathbf{A} e os vetores $\mathbf{u}_i^{(1)}$ e $\mathbf{u}_i^{(2)}$ são respectivamente o i -ésimo autovetor a esquerda e o i -ésimo autovetor a direita.

HOSVD

No contexto das decomposições tensoriais a HOSVD é uma extensão natural da SVD que ortogonaliza N espaços vetoriais V_1, V_2, \dots, V_N relacionados ao espaço tensorial que $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$ pertence. Com a HOSVD somos capazes de expressar o tensor \mathcal{A} como N produtos modo- n de cada base dos espaços V_i . Mais precisamente temos o seguinte teorema:

Teorema 5 (HOSVD) *Qualquer tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$ pode ser escrito como o produto*

$$\mathcal{A} = \mathcal{Z} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \dots \times_N \mathbf{U}_N \quad (18)$$

em que

1. $\mathbf{U}_n = (\mathbf{u}_1^{(n)}, \mathbf{u}_2^{(n)}, \dots, \mathbf{u}_{I_n}^{(n)})$ é uma matriz unitária.
2. $\mathcal{S} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$ em que os subtensores $\mathcal{S}_{i_n=\alpha}$ obtidos fixando o índice n em α possui as seguintes propriedades:

1. ortogonalidade: dois subtensores $\mathcal{S}_{i_n=\alpha}$ e $\mathcal{S}_{i_n=\beta}$ são ortogonais para todo n, α e β tal que $\alpha \neq \beta$:

$$\langle \mathcal{S}_{i_n=\alpha}, \mathcal{S}_{i_n=\beta} \rangle = 0 \text{ quando } \alpha \neq \beta$$

2. ordenação

$$\|\mathcal{S}_{i_n=1}\| \geq \|\mathcal{S}_{i_n=2}\| \geq \dots \geq \|\mathcal{S}_{i_n=I_n}\|$$

para todos os possíveis valores de n .

A norma de Frobenius $\sigma_i^{(n)} = \|\mathcal{S}_{i_n=i}\|$ são os autovalores singulares do modo- n de \mathcal{A} . Os vetores $\mathbf{u}_i^{(n)}$ são autovetores modo- n .

A representação matricial da HOSVD pode ser obtida da Equação (18) como:

$$\mathcal{A}_n = \mathbf{U}_n \mathcal{S}_{(n)} (\mathbf{U}_{n+1} \otimes \mathbf{U}_{n+2} \otimes \dots \otimes \mathbf{U}_N \otimes \mathbf{U}_1 \otimes \mathbf{U}_2 \otimes \dots \otimes \mathbf{U}_{n-1})^H \quad (19)$$

Na decomposição (18) o tensor $\mathcal{Z} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$ é conhecido como *core tensor* ou *tensor núcleo*, os coeficientes de \mathcal{Z} governam as iterações entre as bases \mathbf{U}_i . A matriz \mathbf{U}_i é a base do espaço coluna de $\mathbf{A}_{(i)}$ [4, 30]. O tensor núcleo \mathcal{Z} e as matrizes $\{\mathbf{U}_i\}_{i=1}^N$ podem ser calculadas de acordo com o Algoritmo 1. Se considerarmos uma partição $\mathbf{A} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_P\}$ e $\mathbf{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_Q\}$ do conjunto \mathbb{N} então podemos definir a HOSVD *reduzida* de \mathcal{A} restrita aos modos de \mathbf{A} , ver Algoritmo 2:

$$\mathcal{A} = \mathcal{T} \times_{\mathbf{a}_1} \mathbf{U}_{\mathbf{a}_1} \times_{\mathbf{a}_2} \mathbf{U}_{\mathbf{a}_2} \dots \times_{\mathbf{a}_P} \mathbf{U}_{\mathbf{a}_P} \quad (20)$$

em que

$$\mathcal{T} = \mathcal{Z} \times_{\mathbf{b}_1} \mathbf{U}_{\mathbf{b}_1} \times_{\mathbf{b}_2} \mathbf{U}_{\mathbf{b}_2} \dots \times_{\mathbf{b}_Q} \mathbf{U}_{\mathbf{b}_Q} \quad (21)$$

Algoritmo 1 Higher Order Singular Value Decomposition

```

1: procedure HOSVD( $\mathcal{A}$ )
2:    $N \leftarrow$  ordem de  $\mathcal{D}$ 
3:   for  $i = 1, 2, \dots, N$  do
4:      $\mathbf{U}_i \leftarrow$  autovetores a esquerda de  $\mathbf{A}_{(i)}$ 
5:   end for
6:    $\mathcal{Z} \leftarrow \mathcal{A} \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \dots \times_N \mathbf{U}_N^T$ 
7:   return  $\mathcal{Z}, \{\mathbf{U}_i\}_{i=1}^N$  ▷ Tensor núcleo e matrizes  $\mathbf{U}_i$ .
8: end procedure

```

Algoritmo 2 HOSVD reduzida

```

1: procedure REDUCEDHOSVD( $\mathcal{A}, \mathbf{A}$ )
2:   for  $i \in \mathbf{A}$  do
3:      $\mathbf{U}_i \leftarrow$  autovetores a esquerda de  $\mathbf{A}_{(i)}$ 
4:   end for
5:    $\mathcal{T} \leftarrow \mathcal{A} \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \dots \times_P \mathbf{U}_P^T$ 
6:   return  $\mathcal{T}, \{\mathbf{U}_i\}_{i=1}^P$ 
7: end procedure

```

4 Modelos Tensoriais em Reconhecimento Facial

A álgebra multilinear oferece uma abordagem natural para a análise de imagens com estrutura multifator. Os métodos tensoriais consistem em realizar uma modelagem tensorial sobre um conjunto de imagens faciais que são vetorizadas e organizadas como um tensor. Nessa abordagem os pressupostos acerca do conjunto de imagens, a saber, o espaço das imagens e o espaço das faces são perfeitamente válidos. De fato, é levado em consideração os fatores que influenciam a formação das imagens faciais: posição, iluminação e expressão. O conjunto de fatores multimodais formam um conjunto de espaços vetoriais. Dados multimodais que são modelados como tensores de ordem superior devem obedecer um ordenamento, uma lei de formação para os tensores. Para fins de ilustração, considere um tensor de dados de ordem 4, a saber $\mathcal{A} \in \mathbb{R}^{3 \times 3 \times 3 \times 3}$. Se considerarmos o modo-4 como sendo o espaço de parâmetros, ou seja, as colunas de $\mathbf{A}_{(4)}$ são os padrões, teríamos de organizar os dados conforme ilustrado na Figura (3). Essa organização está diretamente relacionada com a maneira em que matriciamos um tensor e vice-versa. Esses espaços estão intrinsecamente relacionados através do produto tensorial entre os mesmos, onde cada elemento do espaço tensorial é uma instância do conjunto de dados que pode ser decomposto através da HOSVD.

4.1 TensorFaces

No método tensorial TensorFaces são considerados 5 espaços [5, 6]:

- Espaço dos indivíduos;
- Espaço relacionado as posições faciais;
- Espaço das expressões faciais;
- O espaço das condições de iluminação;
- Espaço das faces ou pixels.

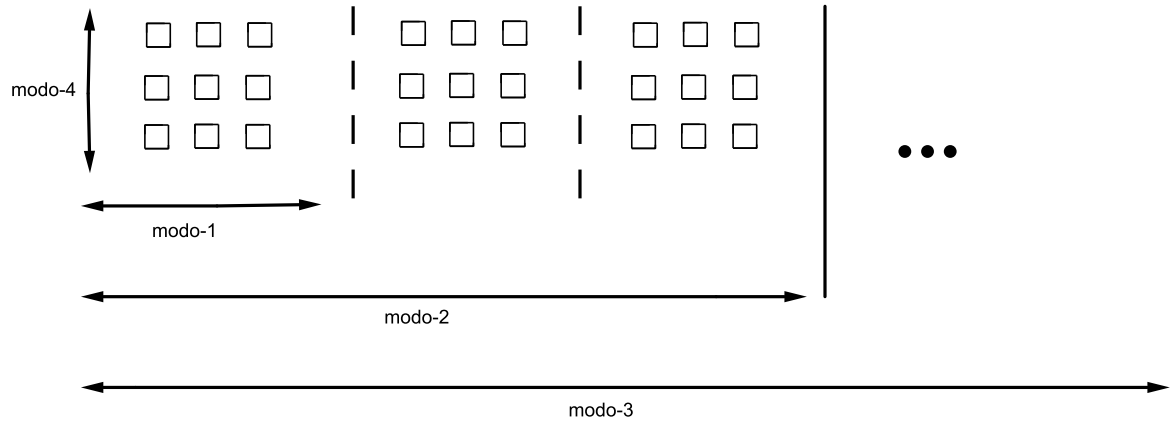


Figura 3: Modelo tensorial de ordem 4, onde cada coluna representa um padrão.

Dessa forma, através da ordenação dos dados devemos construir o tensor de ordem $N = 5$:

$$\mathcal{D} \in \mathbb{R}^{N_{pe} \times N_v \times N_i \times N_e \times N_p}$$

em que N_{pe} , N_v , N_i , N_e e N_p 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:

$$\mathcal{D} = \mathcal{Z} \times_1 \underbrace{U_{pe}}_{U_1} \times_2 \underbrace{U_v}_{U_2} \times_3 \underbrace{U_i}_{U_3} \times_4 \underbrace{U_e}_{U_4} \times_5 \underbrace{U_p}_{U_5} \quad (22)$$

onde \mathcal{Z} possui as mesmas dimensões e ordem de \mathcal{D} , as matrizes fatores U_{pe} , U_v , U_i , U_e e U_p possuem dimensões $N_{pe} \times N_{pe}$, $N_v \times N_v$, $N_i \times N_i$, $N_e \times N_e$ e $N_p \times N_p$, respectivamente. Matematicamente, assumir multilinearidade é assumir linearidade, ou seja, no contexto de reconhecimento facial, o método EigenFaces [12] baseado na análise de componentes principais é parte constituinte do TensorFaces, onde cada coluna de U_p é uma autoface, pois os mesmos foram calculados executando-se o SVD no modo-5 matriciado $\mathcal{D}_{(5)} = \mathcal{D}_{(p)} \in \mathbb{R}^{N_p \times N_e N_i N_v N_{pe}}$ do tensor \mathcal{D} . É importante notar que o modo-5 é o modo de medidas dos pixels, o qual não é relevante para a geração das imagens. Sendo assim, em vez de calcularmos diretamente o tensor núcleo \mathcal{Z} , podemos aplicar a HOSVD reduzida (veja Alg. 2) em relação a esse modo:

$$\mathcal{D} = \mathcal{T} \times_1 U_{pe} \times_2 U_v \times_3 U_i \times_4 U_e \quad (23)$$

em que $\mathcal{T} = \mathcal{Z} \times_5 U_p$ é denominado o *tensor núcleo aumentado* [31]. No modelo multilinear (23), estamos recuperando o tensor de dados a partir de sua decomposição HOSVD. Podemos ser mais específicos quanto a informação a ser recuperada, por exemplo, todas as faces de um único indivíduo, ou todas as faces de um único indivíduo, ou todas as faces de uma única pessoa em uma posição facial específica. De maneira mais geral, uma amostra específica de um indivíduo i_1 na posição i_2 em uma dada condição de iluminação i_3 na expressão facial i_4 pode ser recuperada pela equação

$$\mathcal{D}(i_1, i_2, i_3, i_4) = \mathcal{T} \times_1 \mathbf{u}_{pe, i_1}^\top \times_2 \mathbf{u}_{v, i_2}^\top \times_3 \mathbf{u}_{i, i_3}^\top \times_4 \mathbf{u}_{e, i_4}^\top, \quad (24)$$

em que $\mathbf{u}_{n, j}^\top$ é a linha j da matriz fator U_n . Observe que $\mathcal{D}' = \mathcal{D}(i_1, i_2, i_3, i_4) \in \mathbb{R}^{1 \times 1 \times 1 \times 1 \times N_p}$ e que a amostra desejada $d = \mathcal{D}'_{(5)}$. Se quisermos recuperar todas as imagens de um indivíduo i_1 , temos

$$\mathcal{D}(i_1, :, :, :) = \mathcal{T} \times_1 \mathbf{u}_{pe, i_1}^\top \times_2 U_v \times_3 U_i \times_4 U_e.$$

Existem outras informações desejadas que podem ser recuperadas.

- imagens na expressão i_4 :

$$\mathcal{D}(:, :, :, i_4) = \mathcal{T} \times_1 U_{pe} \times_2 U_v \times_3 U_i \times_4 \mathbf{u}_{e, i_4}^\top;$$

- imagens na condição de iluminação i_3 :

$$\mathcal{D}(:, :, i_3, :) = \mathcal{T} \times_1 \mathbf{U}_{pe} \times_2 \mathbf{U}_v \times_3 \mathbf{u}_{i_3}^\top \times_4 \mathbf{U}_e;$$

- imagens na posição facial i_2 :

$$\mathcal{D}(:, i_2, :, :) = \mathcal{T} \times_1 \mathbf{U}_{pe} \times_2 \mathbf{u}_{v, i_2}^\top \times_3 \mathbf{U}_i \times_4 \mathbf{U}_e;$$

Na equação (24), fomos capazes de recuperar uma amostra específica de um indivíduo i_1 . O vetor linha U_{pe, i_1} sempre será usado quando for necessário recuperar qualquer imagem facial do indivíduo i_1 . Assim a matriz \mathbf{U}_{pe} é a base do espaço de parâmetros das pessoas, em que cada pessoa na base de dados pode ser representada por um único vetor [5], no qual contém os coeficientes com respeito as bases que serão extraídas do tensor:

$$\mathcal{B} = \mathcal{T} \times_2 \mathbf{U}_v \times_3 \mathbf{U}_i \times_4 \mathbf{U}_e \quad (25)$$

com as mesmas dimensões de \mathcal{Z} . Note que:

$$\mathcal{D} = \mathcal{B} \times_1 \mathbf{U}_{pe} \Rightarrow \mathbf{D}_{(pe)} = \mathbf{U}_{pe} \mathbf{B}_{(pe)} \quad (26)$$

em que $\mathbf{D}_{(pe)}$ e $\mathbf{B}_{(pe)} \in \mathbb{R}^{N_{pe} \times N_p N_e N_i N_v}$. Ainda mais:

$$\mathbf{d}_1 = \mathbf{D}_{(p)}(:, 1) = \mathbf{D}_{(pe)}^\top(1, 1 : N_p) \quad (27)$$

é a primeira imagem 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 $i_1 = 1, 2, \dots, N_{pe}$, $i_2 = 1, 2, \dots, N_v$, $i_3 = 1, 2, \dots, N_i$ e $i_4 = 1, 2, \dots, N_e$. A partir da Equação (26) obtemos

$$\mathbf{d}(i_1, i_2, i_3, i_4) = \mathbf{B}_{(pe)}^\top(i_2, i_3, i_4) \mathbf{u}(i_1) \quad (28)$$

em que $\mathbf{u}(i_1)$ é a i_1 -ésima coluna de \mathbf{U}_{pe}^\top , e $\mathbf{B}_{(pe)}^\top(i_2, i_3, i_4)$ é um slice de \mathcal{B} , de dimensões $N_{pe} \times N_p$, obtido pela fixação dos índices i_2, i_3, i_4 . Dado uma imagem de teste \mathbf{d}_t utilizaremos as bases $\mathbf{B}_{(pe)}^\top(i_2, i_3, i_4)$ no modo pessoa para projetar \mathbf{d}_t no espaço de parâmetros das pessoas:

$$\hat{\mathbf{u}}_{pe}(i_2, i_3, i_4) = (\mathbf{B}_{(pe)}^\top(i_2, i_3, i_4))^\dagger \mathbf{d}_t \quad (29)$$

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

$$i_1^* = \arg \min_{\{i_1, i_2, i_3, i_4\}} \|\hat{\mathbf{u}}_{pe}(i_2, i_3, i_4) - \mathbf{u}(i_1)\|^2 \quad (30)$$

4.2 Wavelets de Gabor via Álgebra Multilinear

Faremos uso das wavelets de Gabor em uma forma holística, onde toda a face é usada na extração de características. Iremos extrair um vetor de características aumentado a resultante da concatenação de cada convolução da imagem com a família de wavelets de Gabor, consistindo de toda a informação extraída de diferentes frequências, orientações e localizações, essa representação pode produzir características discriminantes para reconhecimento. As wavelets de Gabor são autosimilares sendo possível gerar qualquer filtro $\psi_{u,v}(x, y)$ a partir de uma wavelet mãe $\psi(x, y)$. Definimos as wavelets de Gabor segundo [32–35]:

$$\psi(x, y) = \frac{f_u^2}{\pi \xi \kappa} e^{-\left(\frac{f_u^2}{\xi^2} \lambda_1^2 + \frac{f_u^2}{\kappa^2} \lambda_2^2\right)} e^{j2\pi f_u \lambda_1} \quad (31a)$$

$$\lambda_1 = x \cos \theta + y \sin \theta. \quad (31b)$$

$$\lambda_2 = -x \sin \theta + y \cos \theta. \quad (31c)$$

Para a extração de características de uma dada imagem é usado um banco de filtros, a saber, a família de wavelets de Gabor com diferentes frequências e rotações:

$$\psi_{u,v}(x, y) = \psi_{(f_u, \theta_v)}(x, y) \text{ tal que } f_u = \frac{f_{\max}}{\sqrt{2^u}}, \theta_v = \frac{v}{8} \pi \quad (32)$$

em que $u = 0, 1, \dots, 4$; $v = 0, 1, \dots, 7$. Ou seja, teremos um conjunto de 40 wavelets divididas em 5 diferentes frequências (escalas) com 8 rotações para cada escala.

Representação de uma imagem nas Wavelets de Gabor

A representação das wavelets de Gabor de uma imagem é a convolução da mesma com a família descrita em (31). Seja $I(x, y)$ uma dada imagem, a convolução da imagem I com a wavelet $\psi_{u,v}$ é definida como

$$\varphi_{u,v}(x, y) = I(x, y) * \psi_{u,v}(x, y), \quad (33)$$

em que $*$ é o operador de convolução e $\varphi_{u,v}(x, y)$ é o resultado da convolução da imagem $I(x, y)$ com a wavelet de Gabor $\psi_{u,v}(x, y)$ em uma escala u e uma dada rotação v . Assim o conjunto $G = \{\varphi_{u,v}(x, y) : u = 0, 1, \dots, 4; v = 0, 1, \dots, 7\}$ formam a representação das wavelets de Gabor da imagem $I(x, y)$ [36]. Aplicando o teorema da convolução [37], podemos obter via transformada de Fourier o resultado desejado:

$$\mathfrak{F}\{\varphi_{u,v}(x, y)\} = \mathfrak{F}\{I(x, y)\}\mathfrak{F}\{\psi_{u,v}(x, y)\}, \quad (34)$$

implicando em

$$\varphi_{u,v}(x, y) = \mathfrak{F}^{-1}\{\mathfrak{F}\{I(x, y)\}\mathfrak{F}\{\psi_{u,v}(x, y)\}\} \quad (35)$$

em que \mathfrak{F} e \mathfrak{F}^{-1} são a transformada de Fourier e sua inversa, respectivamente.

Note que f_{\max} é a maior frequência que as wavelets de Gabor podem assumir. Os parâmetros de $\psi_{(f_u, \theta_v)}(x, y)$ 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 possuírem sua informação concentrada em baixas frequências. Os valores comumente usados [33] são $f_{\max} = 0.25$, $\xi = \kappa = \sqrt{2}$, mantendo a razão entre a frequência e o formato da Gaussiana constante. Com o intuito de agregar todos os resultados das convoluções de uma única imagem, são realizadas a vetorização e concatenação das respostas, gerando um vetor de características aumentado. Esse vetor torna-se rapidamente muito grande. Por exemplo, se tomarmos uma imagem 64×64 , seu vetor aumentado teria dimensão $64.64.40 = 163840$, o que é computacionalmente intratável. Assim, antes de realizarmos a vetorização de $\varphi_{u,v}(x, y)$ é aplicado um fator de *downsampling* para diminuirmos a dimensão gerando $\varphi_{u,v}^e(x, y)$ como mostrado na Equação (36). Em seguida, então normalizamos para média zero com variância unitária [36].

$$\mathbf{g}_{u,v} = \text{vec}(\varphi_{u,v}^e(x, y)) \quad (36)$$

Assim, geramos o conjunto

$$\mathbf{G}_{\text{vec}} = \{\text{vec}[\varphi_{u,v}^e(x, y)] : u = 0, 1, \dots, 4; v = 0, 1, \dots, 7\} \quad (37)$$

de vetores intermediários para formar o vetor de características aumentado

$$\mathbf{g} = (\mathbf{g}_{0,0}^\top \mathbf{g}_{0,1}^\top \dots \mathbf{g}_{4,7}^\top)^\top. \quad (38)$$

Os modelos tensoriais aqui tratados levam em consideração aspectos relativos à formação da imagem, ou seja, fatores multimodais. Um dado vetor de características aumentado \mathbf{g} proveniente do processo de extração de características descrito acima, ainda tem sua origem em uma dada imagem, à qual foi afetada, em sua formação, por um conjunto de fatores. De uma forma mais geral podemos ver o vetor \mathbf{g} como $\mathbf{g}(i_1, i_2, i_3, i_4)$, ou seja, o resultado da aplicação da família das wavelets de Gabor em uma imagem facial do indivíduo i_1 , na posição facial i_2 , em uma condição de iluminação i_3 com uma certa expressão facial i_4 . Dessa maneira, podemos construir um modelo tensorial fundamentado no fato da natureza multimodal do conjunto de imagens faciais não ter sido afetada pelas transformações a elas aplicadas, logo o modelo consiste em

$$\mathcal{G} \in \mathbb{R}^{N_{pe} \times N_v \times N_i \times N_e \times N_f}$$

formado pela permutação ou organização de todas as $\mathbf{g}(i_1, i_2, i_3, i_4)$, em que $i_1 = 1, 2, \dots, N_{pe}$, $i_2 = 1, 2, \dots, N_v$, $i_3 = 1, 2, \dots, N_i$ e $i_4 = 1, 2, \dots, N_e$. Observe que, $N_f = \frac{40(N_p)^2}{\sigma^2}$ é a dimensão do espaço de características das wavelets. Para separar os espaços vetoriais relacionados a cada fator multimodal imerso no tensor \mathcal{G} realizaremos a HOSVD deste tensor:

$$\mathcal{G} = \mathcal{Z} \times_1 \mathbf{U}_{pe} \times_2 \mathbf{U}_v \times_3 \mathbf{U}_i \times_4 \mathbf{U}_e \times_5 \mathbf{U}_f \quad (39)$$

Como \mathbf{U}_f é a base do espaço de parâmetros e estamos particularmente interessados na decomposição nos espaços de fatores, então podemos modificar a HOSVD na Equação (39) para sua forma reduzida:

$$\mathcal{G} = \mathcal{V} \times_1 \mathbf{U}_{pe} \times_2 \mathbf{U}_v \times_3 \mathbf{U}_i \times_4 \mathbf{U}_e \quad (40)$$

em que $\mathcal{V} = \mathcal{Z} \times_5 \mathbf{U}_f$. Através desse método híbrido de extração de características, estamos obtendo através das wavelets de Gabor a maior quantidade de informação possível [38], separando seus fatores multilineares constituintes através da HOSVD, tais como posição facial, iluminação e expressão facial. Como no TensorFaces, assumimos uma mistura probabilística do PCA, que é análogo a métodos de separação de subespaços baseados em posição facial [39, 40]. Logo, para cada posição no modelo tensorial existe uma Gaussiana multidimensional [5].

Algoritmo 3 Classificação com o TenWav

```

1: procedure TENWAV(Imgs,  $\mathbf{g}_{test}$ ) ▷ Conjunto de imagens.
2:   for  $I \in \text{Imgs}$  do
3:     Filtrar a imagen  $I$  com as wavelets Gabor
4:     downsample todos  $\mathbf{g}_{u,v}$  do fator  $\delta$ 
5:      $\mathbf{g}_I \leftarrow (\mathbf{g}_{0,0} \mathbf{g}_{0,1} \cdots \mathbf{g}_{4,7})^\top$ 
6:   end for
7:   Construa  $\mathbf{G}_{(5)}$  com todos  $\mathbf{g}_I$ 
8:   Remodelar  $\mathbf{G}_{(5)}$  para um tensor  $\mathcal{G}$ 
9:    $\mathbf{A} \leftarrow \{5\}$ 
10:   $\mathcal{V}, \mathbf{U}_{pe}, \mathbf{U}_p, \mathbf{U}_i, \mathbf{U}_e \leftarrow \text{reducedHOSVD}(\mathcal{G}, \mathbf{A})$ 
11:   $\mathcal{W} \leftarrow \mathcal{V} \times_2 \mathbf{U}_p \times_3 \mathbf{U}_i \times_4 \mathbf{U}_e$ 
12:   $\hat{\mathbf{g}}_{pe}(i_2, i_3, i_4) \leftarrow [\mathbf{W}_{(pe)}^\top(i_2, i_3, i_4)]^\dagger \mathbf{g}_{test}$ 
13:   $i_1^* \leftarrow \arg \min_{\{i_1, i_2, i_3, i_4\}} \|\hat{\mathbf{g}}_{pe}(i_2, i_3, i_4) - \mathbf{g}(i_1)\|^2$ 
14:  return  $i_1^*$ 
15: end procedure

```



Figura 4: Imagens de um indivíduo em uma dada condição de iluminação, expressão facial e nas posições $-34^\circ, -17^\circ, 0^\circ, +17^\circ, +34^\circ$, respectivamente.

Múltiplas Projeções Lineares no Espaço das Wavelets de Gabor

Cada indivíduo no conjunto de imagens pode ser representado por um vetor de coeficientes na matriz de autovetores do espaço de pessoas \mathbf{U}_{pe} . Este vetor de coeficientes é relativo ao espaço de gerado pelas projeções das wavelets de Gabor em cada fator multimodal (posição, iluminação e expressão). Da mesma forma que podemos pensar em uma assinatura do indivíduo no modelo TensorFaces, para o modelo tensorial com wavelets de Gabor podemos pensar em uma assinatura multiescala da pessoa no modelo tensorial. Ou seja,

$$\mathcal{W} = \mathcal{V} \times_2 \mathbf{U}_p \times_3 \mathbf{U}_i \times_4 \mathbf{U}_e \quad (41)$$

implica em

$$\hat{\mathbf{g}}_{pe}(i_2, i_3, i_4) = [\mathbf{W}_{(pe)}^\top(i_2, i_3, i_4)]^\dagger \mathbf{g}_{test} \quad (42)$$

em que a operação de classificação é realizada através da distância euclidiana para cada base relativa aos fatores multimodais:

$$i_1^* = \arg \min_{i_1, i_2, i_3, i_4} \|\hat{\mathbf{g}}_{pe}(i_2, i_3, i_4) - \mathbf{g}(i_1)\|^2 \quad (43)$$

Uma vez que a dimensão do vetor de característica tenha sido reduzida e a habilidade de discriminação tenha sido aumentada por algum método de análise de subespaços, podemos usar simplesmente o classificador *naïve bayes* com a distância

4.3 Simulações e Resultados

Em nossos experimentos utilizamos dois bancos de imagens faciais:

- Weizmann Face Image Database [41]. Foi utilizado parte do banco consistindo de 24 indivíduos em 5 posições faciais, 3 condições de iluminação e 3 expressões faciais, totalizando 1080 imagens, na Figura 4 temos exemplo de um indivíduo que compõem esse banco de faces.
- Extended Yale Face Image Database [42]. Foi utilizado parte do banco consistindo de 27 indivíduos em 9 posições faciais, 15 condições de iluminação, totalizando 3645 imagens.

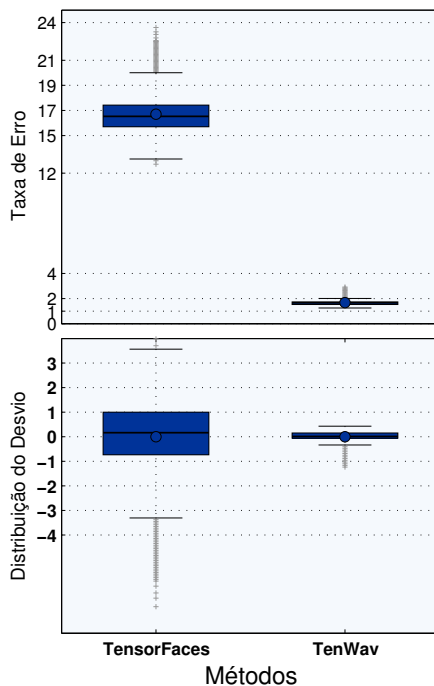


Figura 5: Taxa de erro e estimativa da distribuição do desvio, resultantes de 8500 repetições do 10-fold VC no modo híbrido *pos-ilum* de dimensão 15.

4.4 Experimentos no Weizmann Face Image Database

Para os experimentos nesse banco de faces os fatores multimodais considerados são: posição facial, condição de iluminação e expressão facial. Ainda são componentes do modelo tensorial o número de indivíduos e a dimensão de cada imagem facial vetorizada para o modelo TensorFaces e a dimensão de cada resposta (resultante da filtragem das faces com o banco de filtros das Wavelets de Gabor) para o TenWav. Dessa forma, para a construção dos modelos tensoriais TensorFaces e TenWav devemos organizar nosso dataset como um tensor $\mathcal{D} \in \mathbb{R}^{24 \times 5 \times 3 \times 3 \times N_p}$ de ordem 5 onde cada modo tem correspondência biunívoca com os fatores multimodais que compõem o banco de imagens Weizmann. O espaço de medidas ou espaço dos pixels tem dimensão N_p , que pode assumir valores variáveis conforme a necessidade, e. g., uma operação de redimensionamento da imagem original, o fator de *downsampling* ρ no modelo TenWav, dentre outras operações de pré-processamento desejado.

Para o tensor \mathcal{D} mencionado acima, temos que o fator multimodal de maior dimensão é a posição facial com tamanho 5. Experimentos anteriores com TensorFaces foram realizados com essa configuração [5, 6, 31], onde foi considerado 4 posições para treino e uma para teste. O problema nessa abordagem é evidente, uma vez que a taxa de acerto e consequentemente a taxa de erro apresentada não resulta de um experimento fidedigno. Ainda se fosse utilizada uma abordagem sistemática como 5-fold VC nesse modo, devido à dimensão ser igual ao número de *folds* na validação cruzada teríamos apenas uma possibilidade para a configuração das partições, sendo possível apenas uma repetição do 5-fold VC. Uma abordagem mais apropriada a fim de contornar esses problemas consiste em realizar uma fusão modo- n para dois fatores multimodais, produzindo um tensor com fator multimodal híbrido de dimensão elevada.

Fusão nos modos posição facial e iluminação

No próximo experimento, realizamos a fusão dos modos posição facial e iluminação, criando um tensor $\mathcal{D} \in \mathbb{R}^{24 \times 15 \times 3 \times N_p}$ de ordem 4. Ao fazermos isso, estamos diminuindo a complexidade do modelo, no sentido do número de parâmetros a ser calculado, ou seja, uma SVD a menos para ser calculada. Como pode ser visto na representação da Figura 6. O novo modo *pos-ilum* de dimensão 15 foi escolhido para a realização do 10-fold VC modo- n , tal que, $\lfloor \frac{15}{10} \rfloor = 1$ e $\gamma = 15 - \lfloor \frac{15}{10} \rfloor 10 = 5$. Assim, temos cinco *folds* $\mathcal{D}_i \in \mathbb{R}^{24 \times 2 \times 3 \times N_p}$ e cinco $\mathcal{F}_i \in \mathbb{R}^{24 \times 1 \times 3 \times N_p}$. O número total de partições é $\rho = \frac{15!}{2!2!2!2!2!} \approx 4,08 \times 10^{10}$, tornando o cálculo de todos $\varepsilon_\rho^{(i)}$ ($i = 1, \dots, \rho$) dispendioso. O objetivo principal nesse primeiro experimento é analisar o desempenho do 10-fold VC modo- n a longo prazo. Escolhemos 8500 amostras aleatórias do total de ρ partições, ou seja, repetimos o experimento 8500 vezes. Como podemos ver na Figura 5, a distribuição de $\varepsilon_r^{(i)} - \varepsilon_r$ não sofre de uma alta variância, i.e., uma estimativa única do 10-fold VC estará próxima de ε_r , o que confirma a viabilidade da VC modo- n . Ambos os métodos TensorFaces e TenWav apresentam baixa variância, principalmente o TenWav. A figura apresenta ainda a taxa de erro estimada, onde fica evidente que o TenWav é mais adequado que o TensorFaces quando os fatores multimodais posição facial e condição de iluminação são variados no conjunto de faces. Os resultados apresentados relativos à estimação do erro e distribuição do

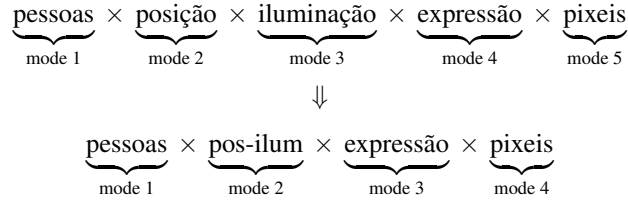


Figura 6: Representação pictográfica da estrutura interna do modelo tensorial com fusão nos modos posição e iluminação.

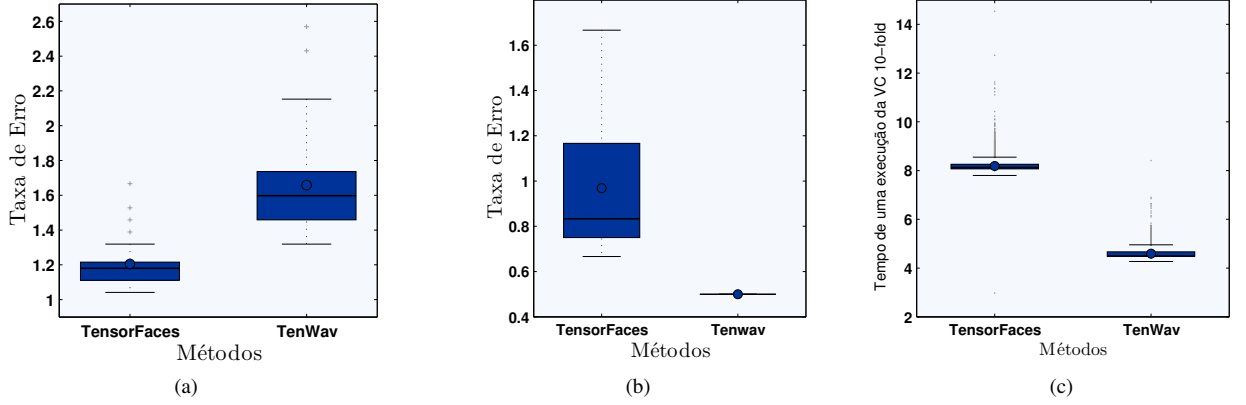


Figura 7: Resultados da taxa de erro para a base Weizmann Face Image: (a) 100 repetições do 10-fold VC no modo *pos-exp* de dimensão 15, (b) 1000 repetições do 5-fold VC no modo *ilum-exp* de dimensão 9, (c) tempo médio de cada validação cruzada em 8500 repetições do 10-fold VC no modo *pos-ilum* de dimensão 15.

desvio de $\varepsilon_p^{(i)}$ em relação a ε_p , mostra-nos que se tomarmos amostras de tamanho menor, tais como 10, 50 ou 100, teremos ainda boas estimativas da taxa de erro. Outros dois experimentos foram realizados a fim de se verificar o comportamento dos métodos TenWav e TensorFaces. Para o experimento *Fusão nos modos posição facial e expressão* foram realizadas 100 repetições do 10-fold VC modo- n . No segundo experimento *Fusão nos modos iluminação e expressão* foram realizadas 1000 repetições do 5-fold VC modo- n .

Fusão nos modos posição facial e expressão

Neste experimento observamos como a variação conjunta dos fatores posição facial e expressão afetam a taxa de erro dos modelos considerados. Para isso é necessária fusão dos modos correspondentes a esses fatores. Primeiramente devemos fazer uma mudança na estrutura interna do tensor original $\mathcal{D} \in \mathbb{R}^{N_{pe} \times N_v \times N_i \times N_e \times N_p}$ para um novo tensor $\mathcal{D}^* \in \mathbb{R}^{N_{pe} \times N_v \times N_e \times N_i \times N_p}$ de forma que os modos posição facial e expressão tornem-se vizinhos, possibilitando sua fusão, conforme é representado na Figura 8. Após a fusão, o tensor resultante $\mathcal{D} \in \mathbb{R}^{24 \times 15 \times 3 \times N_p}$ com o novo modo *pos-exp* de dimensão 15 foi escolhido para a

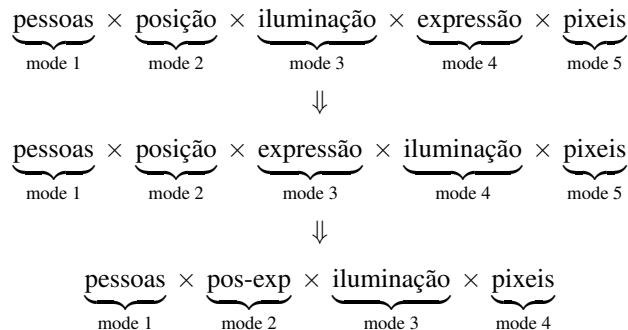


Figura 8: Representação pictográfica da mudança da estrutura interna do modelo tensorial com fusão nos modos posição e expressão.

realização do 10-fold VC modo- n . Realizando a mesma análise quanto aos *folds* que podem ser construídos, vemos que temos cinco *folds* $\mathcal{D}_i \in \mathbb{R}^{24 \times 2 \times 3 \times N_p}$ e cinco $\mathcal{F}_i \in \mathbb{R}^{24 \times 1 \times 3 \times N_p}$ com o mesmo número de partições ρ do experimento anterior. Na Figura 7(a), apresentamos um sumário dos resultados desse experimento, onde vemos taxas de erros próximas (diferença de 0.5%), com o TensorFaces apresentando um melhor desempenho.

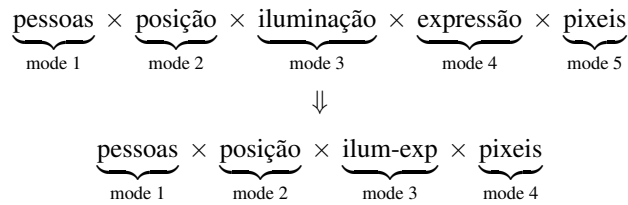


Figura 9: Representação pictográfica da estrutura interna do modelo tensorial com fusão nos modos iluminação e expressão.

Experimento	Número de repetições	TensorFaces	TenWav
<i>pos-ilum</i>	8500	$\approx 7.08 \times 10^4$	$\approx 3.97 \times 10^4$
<i>ilum-exp</i>	1000	$\approx 3.41 \times 10^3$	$\approx 1.89 \times 10^3$
<i>pos-exp</i>	100	$\approx 1.29 \times 10^3$	$\approx 7.76 \times 10^2$

Tabela 1: Tempo total em segundos da HOSVD nos modelos TenWav e TensorFaces

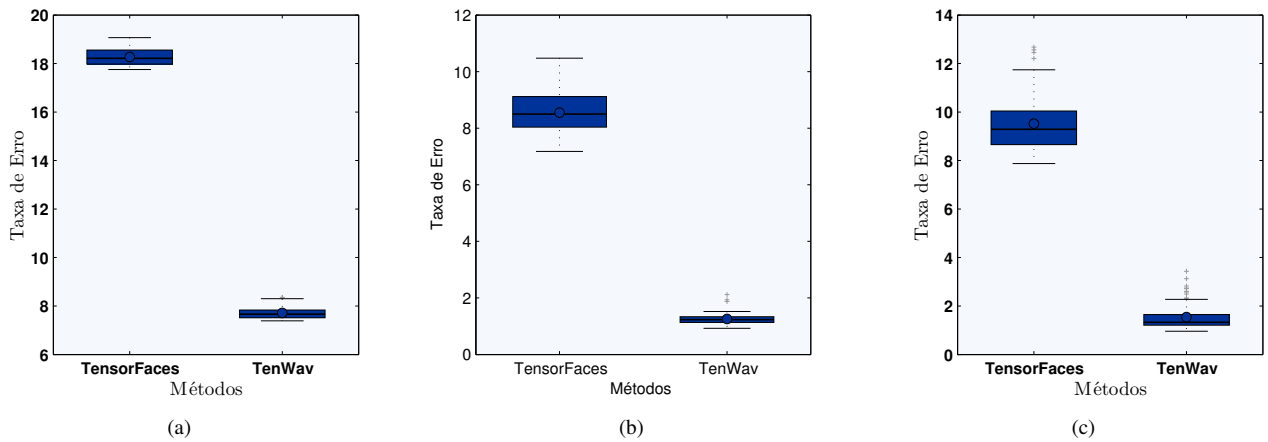


Figura 10: Resultados da taxa de erro para a base Extended Yale Face Image: (a) 100 repetições do 5-fold VC no modo posição facial de dimensão 9, (b) 100 repetições do 10-fold VC no modo iluminação de dimensão 15, (c) 100 repetições do 5-fold VC no modo iluminação de dimensão 15.

Fusão nos modos iluminação e expressão

Nesse experimento analisamos os efeitos causados por mudanças nos fatores iluminação e expressão. Segue que as dimensões desses modos são ambas 3, logo o fator híbrido *ilum-exp* possui dimensão 9 (ver Fig. 9), impossibilitando o uso do 10-fold VC modo- n . Como alternativa viável consideramos o 5-fold VC modo- n com 1000 repetições. Na Figura 9 podemos ver a representação da estrutura interna do modelo tensorial com fusão nos modos iluminação e expressão. Desse modo é possível criar quatro tensores $D_i \in \mathbb{R}^{24 \times 5 \times 2 \times N_p}$ e um $\mathcal{F}_i \in \mathbb{R}^{24 \times 5 \times 1 \times N_p}$. Os resultados podem ser visualizados na Figura 7(b), a qual nos indica que dois modelos possuem taxas de erros próximas.

Análise do tempo de execução

É também de interesse estimarmos o desempenho dos modelos tensoriais quanto ao tempo de execução, em específico o tempo da HOSVD em cada modelo. Primeiramente, analisamos o tempo médio de uma execução do 10-fold VC modo- n do experimento *pos-ilum*. Na Figura 7(c), pode-se concluir que o tempo médio de execução do TenWav é menor que o TensorFaces. Isso ocorre deve-se ao fato do modo pixel possuir dimensão N_p reduzida, essa redução é devido ao fator de *downsampling* ϱ . De fato, para uma imagem $I \in \mathbb{R}^{n \times n}$, temos que $\text{vec}(I) \in \mathbb{R}^{n^2}$. No modelo TensorFaces temos $N_p = n^2$, enquanto que para o TenWav temos $N_p^* = \frac{uv}{\varrho^2} n^2$. Segue que $\frac{uv}{\varrho^2} \leq 1$ na maioria dos casos de aplicações práticas. No nosso caso $u = 5, v = 8$ e $\varrho = 8$, de forma que $N_p^* = 0.625N_p$. Na Tabela 1 vemos o tempo total para cada um dos três experimentos realizados sobre a base Weizmann.

4.5 Experimentos no Extended Yale Face Image Database

A base de dados *Extended Yale Face Database B* contém 16128 imagens de 28 indivíduos em 9 posições faciais e 64 condições de iluminação, para os experimentos foram utilizados 27 indivíduos em 9 posições faciais e 15 condições de iluminação. Foram realizados dois experimentos com 100 repetições do k -fold VC:

- O primeiro experimento consiste em realizar 5-fold VC modo- n sobre o modo posição facial, fazendo-se uso de quatro tensores $\mathcal{D}_i \in \mathbb{R}^{27 \times 2 \times 15 \times N_p}$ e um tensor $\mathcal{F}_i \in \mathbb{R}^{27 \times 1 \times 15 \times N_p}$.
- O segundo experimento consiste em realizar 5-fold VC modo- n sobre o modo iluminação, fazendo-se uso de cinco tensores $\mathcal{D}_i \in \mathbb{R}^{27 \times 9 \times 3 \times N_p}$.
- O terceiro experimento consiste em realizar 10-fold VC modo- n sobre o modo iluminação, fazendo-se uso de cinco tensores $\mathcal{D}_i \in \mathbb{R}^{27 \times 9 \times 2 \times N_p}$ e cinco tensores $\mathcal{F}_i \in \mathbb{R}^{27 \times 9 \times 1 \times N_p}$.

Os resultados do primeiro experimento podem ser vistos na Figura 10(a), onde fica evidente que o TensorFaces teve uma aumento na taxa de erro devido à variação das posições faciais, efeito também visto no TenWav de forma mais amena. Nas Figuras 10(b) e 10(c) apresentamos os resultados dos experimentos 2 e 3. Observe que a iluminação é o fator multimodal que menos afeta o desempenho dos modelos tensoriais. A diferença entre esses dois experimentos é a quantidade de *folds* em cada validação cruzada, possibilitando uma análise da melhor divisão do *dataset* na VC. Observa-se que, embora a mesmo a VC com 5 *folds* seja "mais estratificada" (tensores do mesmo tamanho) é preferível utilizar o 10-fold VC, pois a quantidade de *outliers* é reduzida.

REFERÊNCIAS

- [1] L. De Lathauwer. "Signal processing based on multilinear algebra". Ph.D. thesis, Katholieke Univ. Leuven, Leuven, Belgium, 1997.
- [2] T. Kolda. "Orthogonal tensor decompositions". *SIAM J. Matrix Anal. Appl.*, vol. 23, no. 1, pp. 243–255, July 2001.
- [3] P. Comon, X. Luciani and A. L. F. de Almeida. "Tensor Decompositions, Alternating Least Squares and Other Tales". *Journal of Chemometrics*, vol. 23, no. 7-8, pp. 393–405, 2009.
- [4] L. De Lathauwer. "A multilinear singular value decomposition". *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1253–1278, 2000.
- [5] M. Vasilescu and D. Terzopoulos. "Multilinear image analysis for facial recognition". volume 2, pp. 511–514 vol.2, 2002.
- [6] M. Vasilescu and D. Terzopoulos. "Multilinear subspace analysis of image ensembles". volume 2, pp. II–93–9 vol.2, June 2003.
- [7] L. G. Brown. "A survey of image registration techniques". *ACM Computing Surveys*, vol. 24, no. 4, 1992.
- [8] Y. Zhu and J. Xue. "Face recognition based on random subspace method and tensor subspace analysis". *Neural Computing and Applications*, vol. 28, no. 2, pp. 233–244, Feb 2017.
- [9] H. Lu, K. N. Plataniotis and A. N. Venetsanopoulos. *Multilinear Subspace Learning: Fundamentals and foundations*. Taylor & Francis/CRC, 2014.
- [10] M. Hassaballah and S. Aly. "Face recognition: challenges, achievements and future directions". *IET Computer Vision*, vol. 9, no. 4, pp. 614–626, 2015.
- [11] K. Solanki and P. Pittalia. "Review of face recognition techniques". *International Journal of Computer Applications*, vol. 133, no. 12, pp. 20–24, 2016.
- [12] M. Turk and A. Pentland. "Eigenfaces for Recognition". *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
- [13] P. Halmos. *Finite-Dimensional Vector Spaces*. Undergraduate Texts in Mathematics. Springer New York, 1974.
- [14] J. Magnus and H. Neudecker. *Matrix differential calculus with applications in statistics and econometrics*. Wiley series in probability and statistics. John Wiley, 1999.
- [15] C. Eckart and G. Young. "The approximation of one matrix by another of lower rank". *Psychometrika*, vol. 1, no. 3, pp. 211–218, 1936.
- [16] A. Knapp. *Basic Algebra: Along with a Companion Volume Advanced Algebra*. Number v. 1 in Cornerstones (Birkhäuser Verlag). Birkhäuser Boston, 2006.
- [17] R. Merris. *Multilinear Algebra*. Algebra, logic, and applications. Taylor & Francis, 1997.
- [18] M. Artin. *Algebra*. Pearson Education, 2014.
- [19] L. Hogben. *Handbook of Linear Algebra*. Discrete Mathematics and Its Applications. Taylor & Francis, 2006.
- [20] S. Lang. *Algebra*. Graduate Texts in Mathematics. Springer New York, 2002.

- [21] D. Leibovici and R. Sabatier. “A Singular Value Decomposition of a k-Ways Array for a Principal Components Analysis of Multi-Way Data, the PTA-k”. *Linear Algebra and its Applications*, vol. 269, pp. 307–329, 1998.
- [22] W. Greub. *Multilinear algebra*. Universitext - Springer-Verlag. Springer-Verlag, 1978.
- [23] H. A. Kiers. “Towards a standardized notation and terminology in multiway analysis”. *J. Chemometrics*, vol. 14, no. 3, pp. 105–122, 2000.
- [24] T. G. Kolda. “Multilinear operators for higher-order decompositions”. Technical report, 2006.
- [25] F. L. Hitchcock. “The expression of a tensor or a polyadic as a sum of products”. *J. Math. Phys. Camb.*, , no. 6, pp. 164–189, 1927.
- [26] F. L. Hitchcock. “Multiple invariants and generalized rank of a p-way matrix or tensor”. *J. Math. Phys. Camb.*, , no. 7, pp. 39–70, 1927.
- [27] R. B. Cattell. “Parallel proportional profiles and other principles for determining the choice of factors by rotation”. *Psychometrika*, , no. 9, pp. 267–283, 1944.
- [28] J. D. Carroll and C. J. J. “Analysis of individual differences in multidimensional scaling via an N -way generalization of “Eckart-Young” decomposition”. *Psychometrika*, , no. 35, pp. 283–319, 1970.
- [29] R. A. Harshman. “Foundations of the PARAFAC procedure: Model and conditions for an “explanatory” multi-mode factor analysis”. *UCLA Working Papers in Phonetics*, vol. 16, pp. 1–84, Dec. 1970.
- [30] T. G. Kolda and B. W. Bader. “Tensor Decompositions and Applications”. *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.
- [31] M. A. O. Vasilescu. “A Multilinear (Tensor) Algebraic Framework for Computer Graphics, Computer Vision, and Machine Learning”. Ph.D. thesis, University of Toronto, , Belgium, 2009.
- [32] T. S. Lee. “Image Representation Using 2D Gabor Wavelets”. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 10, pp. 959–971, 1996.
- [33] L. Shen, L. Bai and M. Fairhurst. “Gabor wavelets and General Discriminant Analysis for face identification and verification”. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pp. 553–563, 2007.
- [34] V. Štruc and N. Pavešić. “Gabor-based kernel-partial-least-squares discrimination features for face recognition”. *Informatica (Vilnius)*, vol. 20, no. 1, pp. 115–138, 2009.
- [35] V. Štruc and N. Pavešić. “The Complete Gabor-Fisher Classifier for Robust Face Recognition”. *EURASIP Advances in Signal Processing*, vol. 2010, pp. 26, 2010.
- [36] C. Liu and H. Wechsler. “Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition”. *IEEE Transactions on Image Processing*, vol. 11, no. 4, pp. 467–476, April 2002.
- [37] A. Oppenheim and R. Schaffer. *Discrete-Time Signal Processing*. Pearson Education, Limited, 2010.
- [38] K. Okajima. “Two-dimensional Gabor-type receptive field as derived by mutual information maximization”. *Neural Networks*, vol. 11, no. 3, pp. 441 – 447, 1998.
- [39] A. Pentland and L. B. Moghaddam. “View-based and modular eigenspaces for face recognition”. *In Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 84–91, 1994.
- [40] M. E. Tipping and C. M. Bishop. “Mixtures of probabilistic principal component analysers”. *Neural Computation*, vol. 11, no. 2, pp. 443–482, 1999.
- [41] “The Weizmann Facebase”.
- [42] A. Georghiades, P. Belhumeur and D. Kriegman. “From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose”. *IEEE Trans. Pattern Anal. Mach. Intelligence*, vol. 23, no. 6, pp. 643–660, 2001.