Ranking Methods for Tensor Components Analysis and their Application to Face Images

Tiene A. Filisbino, Gilson A. Giraldi
National Laboratory for Scientific Computing
LNCC
Petropolis, Brazil
{tiene,gilson}@lncc.br

Carlos Eduardo Thomaz
Department of Electrical Engineering
FEI
São Bernardo do Campo, Brazil
cet@fei.edu.br

***Abstract*—Higher order tensors have been applied to model multidimensional image databases for subsequent tensor decomposition and dimensionality reduction. In this paper we address the problem of ranking tensor components in the context of the concurrent subspace analysis (CSA) technique following two distinct approaches: (a) Estimating the covariance structure of the database; (b) Computing discriminant weights through separating hyperplanes, to select the most discriminant CSA tensor components. The former follows a ranking method based on the covariance structure of each subspace in the CSA framework while the latter addresses the problem through the discriminant principal component analysis methodology. Both approaches are applied and compared in a gender classification task performed using the FEI face database. Our experimental results highlight the low dimensional data representation of both approaches, while allowing robust discriminant reconstruction and interpretation of the sample groups and high recognition rates.**

***Keywords*-Dimensionality Reduction; Tensor Subspace Learning; CSA; Face Image Analysis**

I. Introduction

In pattern recognition applications that involve managing of data sets with large number of features we should apply dimensionality reduction for discarding redundancy and reduce the computational cost of further operations. There are numerous works on linear dimensionality reduction including the classical principal component analysis (PCA), linear discriminant analysis (LDA) and multidimensional scaling (MDS) [1], [2]. Linear techniques seek for new variables that obey some optimization criterion and can be expressed as linear combination of the original ones. These methods can be also classified as subspace learning methods in the sense that the output linear space has an optimum subspace for compact data representation.

All of the mentioned methods consider a gray scale $n_1 \times n_2$ image as a high dimensional vector in $\Re^{n_1 \cdot n_2}$ space. Tensor methods, on the other hand, consider gray scale images as second order tensors and colored images as third order tensors applying multilinear methods for subspace learning and analysis. Tensor representation for images was proposed in [3] by using a singular value decomposition method (SVD). Others approaches in this area are the concurrent subspace analysis (CSA) [4], multilinear independent components analysis (MICA) [5], multilinear principal component analysis (MPCA) [6], tensor discriminant analysis (TDA) [7], [8] and tensor rank-one decomposition [9], among others. The applications include face and gait recognition, digital number recognition, signal processing, content analysis, anomaly detection in data mining (see [10] for a survey in tensor techniques and applications).

The problem of ranking tensor components have been considered in the context of the MPCA and TDA [6], [7], [8]. This problem is relevant because the most important components for separating sample groups depend on the kind of patterns we are considering. Therefore, we must identify the most discriminant "directions" for a specific classification task which may be not the ones that most vary in the samples.

In this paper we address this issue in the context of the CSA technique following two approaches: (a) Ranking CSA components using a novel technique for estimating the covariance structure of all the data; (b)Adapt the DPCA technique presented in [11] for data represented by higher order tensors. These approaches are the main contribution of this work. Besides, they are not restricted to the CSA subspace as we shall see later.

In order to complete the task (a) we apply the tensor product (also called outer product) of vectors for general higher order tensor decomposition [9]. We revise this approach and show that CSA principal components and the associated covariance structure arise naturally in this algebraic framework. Next, we address the task (b) by adapting the methodology presented in [11] to tensor spaces. This approach is a ranking method that identifies the most discriminant "directions" for a specific classification task instead of the features that most vary in the samples. So, the original dataset is projected in the CSA subspace. Then, we take the obtained lower dimensional representation and propose to rank the tensor principal components by how well they align with separating hyperplane directions, determined by the corresponding discriminant weights. Such a set of tensor principal components ranked in decreasing order of the discriminant weights is called the tensor discriminant principal components (TDPCA). In this paper we focus on the SVM (Support Vectors Machine) [12] method but any other separating hyperplane could be used.

We investigate the efficiency of both the proposed ranking

methods for tensor components through face image analysis tasks. The experimental results carried out use a gender classification task (female versus male samples) performed using the FEI database [1] The properties of this database are very attractive to test tensor component analysis because the background is controlled while scale and orientation are almost the same in all poses (see first and second paragraphs of section VII). The results show that the proposed methods allow robust reconstruction and interpretation of the database, as well as high recognition rates using less linear features.

The remainder of this work is organized as follows. The section II describes the SVM technique. Next, in sections III-IV, we review the CSA approach and the tensor product framework, respectively. The proposed ranking methods are presented in sections V and VI. In section VII we show the experimental results. In section VIII, we present the conclusions and future works.

II. Support Vector Machines (SVM)

SVM [12] is primarily a two-class classifier that maximizes the width of the margin between classes, that is, the empty area around the separating hyperplane defined by the distance to the nearest training samples [12]. It can be extended to multi-class problems by solving several two-class problems.

Given a training set that consists of N pairs of $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2) \ldots (\mathbf{x}_N, y_N)$, where $\mathbf{x}_i$ denote the n-dimensional training observations and $y_i \in \{-1, 1\}$ the corresponding classification labels, the SVM method [12] seeks to find the hyperplane defined by:

$$f(\mathbf{x}) = (\mathbf{x} \cdot \mathbf{w}) + \mathbf{b} = 0, \tag{1}$$

which separates positive and negative observations with the maximum margin. It can be shown that the solution vector $\mathbf{w}_{svm}$ is defined in terms of a linear combination of the training observations, that is,

$$\mathbf{w}_{svm} = \sum_{i=1}^{N} \alpha_i y_i \mathbf{x}_i, \tag{2}$$

where α_i are non-negative coefficients obtained by solving a quadratic optimization problem with linear inequality constraints [12]. Those training observations x_i with non-zero α_i lie on the boundary of the margin and are called support vectors.

The formulation of the SVM solution does not make any assumption about the data distribution. It focuses on the observations that lie close to the separating margin of the classes, that is, on the observations that most count for classification given a zoom into the details of group differences [12]. Therefore, we expect to select the most discriminant tensor components for a specific classification task when using SVM for computing discriminant weights.

[1] http://fei.edu.br/~cet/facedatabase.html

III. Multilinear Dimensionality Reduction

In the traditional image processing literature, a tensor of order n is just a generalized matrix $\mathbf{X} \in \Re^{m_1 \times m_2 \times \ldots \times m_n}$ [3]. So, it becomes clear that there is an isomorphism between $\Re^{m_1 \times m_2 \times \ldots \times m_n}$ and $\Re^{m_1 \cdot m_2 \cdots m_n}$. Therefore, the addition of tensores and the notions of internal product and norm in $\Re^{m_1 \times m_2 \times \ldots \times m_n}$ are induced, in a natural manner, from the $\Re^{m_1 \cdot m_2 \cdots m_n}$ space [4].

Definition 1. *The internal product between two tensors* $\mathbf{X} \in \Re^{m_1 \times m_2 \times \ldots \times m_n}$ *and* $\mathbf{Y} \in \Re^{m_1 \times m_2 \times \ldots \times m_n}$ *is defined by:*

$$\langle \mathbf{X}, \mathbf{Y} \rangle = \sum_{i_1=1,\ldots,i_n=1}^{m_1,\ldots,m_n} \mathbf{X}_{i_1,..,i_n} \mathbf{Y}_{i_1,..,i_n} \tag{3}$$

Other tensor algebra operations are defined to allow to generate another tensor from one or more input tensors. In areas of physics and mathematics the addition, contraction and multiplication are the most usual operations in tensor algebra [13]. For image processing applications, there are also the following mode-k product and the mode-k flattening operation [4]:

Definition 2. *The mode-k product of tensor* $\mathbf{X} \in \Re^{m_1 \times m_2 \times \ldots \times m_n}$ *with the matrix* $A \in \Re^{m\prime_k \times m_k}$ *is given by:*

$$(\mathbf{X} \times_k A)_{i_1,\ldots,i_{k-1},i,i_{k+1},\ldots,i_n} = \sum_{j=1}^{m_k} \mathbf{X}_{i_1,\cdots,i_{k-1},j,i_{k+1},\cdots i_n} A_{i,j}, \quad i = 1, 2, ..., m\prime_k. \tag{4}$$

Definition 3. *The mode-k flattening of an n-th order tensor* $\mathbf{X} \in \Re^{m_1 \times m_2 \times \ldots \times m_n}$ *into a matrix* $X^k \in \Re^{m_k \times \amalg_{i \neq k} m_i}$, *denoted by* $X^k \Longleftarrow_k \mathbf{X}$, *is defined by expression:*

$$X^k_{i_k,j} = \mathbf{X}_{i_1,\ldots,i_n}, where \quad j = 1 + \sum_{l=1,l\neq k}^{n} (i_l - 1) \amalg_{0=l+1,0\neq k}^{n} m_0. \tag{5}$$

The Figure 1 pictures a mode-1 flattening operation for third-order tensors. Observe that the result of the mode-k flattening operation is a matrix that concatenates side-by-side each $\Re^{m_k}$ array of the input tensor.

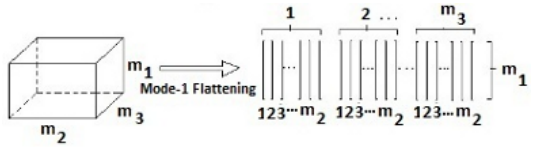

Fig. 1. Mode-1 flattening operation for third-order tensors (Source: Adapted from [4]).

Now, let us consider a database with N element that can be represented through $n-th$ order tensors $\mathbf{X}_i \in \Re^{m_1 \times m_2 \times \ldots \times m_n}$, $i = 1, 2, ..., N$. Let us also consider n projection matrices U^1, U^2,...,U^n, where $U^k \in \Re^{m_k \times m\prime_k}$ and $m_k > m\prime_k, k = 1...n$ (the superscript k in U^k just means the projection matrix index). So, the projection of a tensor $\mathbf{X}_i$ in the space $\Re^{m\prime_1 \times m\prime_2 \times \ldots \times m\prime_n}$ (that means, its low dimensional

representation) is given by a tensor $\mathbf{Y}_i \in \Re^{m\prime_1 \times m\prime_2 \times \ldots \times m\prime_n}$ computed by expression [4]:

$$\mathbf{Y}_i = \mathbf{X}_i \times_1 U^{1^T} \ldots \times_n U^{n^T}, \quad i = 1 \ldots N. \tag{6}$$

Therefore, it can be shown that the reconstruction is a tensor $\mathbf{X}_i^R \in \Re^{m_1 \times m_2 \times \ldots \times m_n}$, defined by [4]:

$$\mathbf{X}_i^R = \mathbf{X}_i \times_1 U^1 U^{1^T} \ldots \times_n U^n U^{n^T} \tag{7}$$

We must define an optimality criterion to seek for suitable matrices U^1, U^2,...,U^n. In [4] this point is addressed by the least square optimization problem:

$$(U^j|_{j=1}^n) = \arg \min_{U^j|_{j=1}^n} \sum_{i=1}^{N} ||\mathbf{X}_i \times_1 U^1 U^{1^T} \ldots \times_n U^n U^{n^T} - \mathbf{X}_i||^2, \tag{8}$$

which solution has the following property [4].

Theorem 1: If $(U^1, \ldots, U^{k-1}, U^{k+1}, \ldots, U^n)$ are known then the ideal matrix U^k is composed by the $m\prime_k$ principal eigenvectors of the covariance matrix $C^k = \sum_{i=1}^{N} X_i^k X_i^{k^T}$, where X_i^k is the matrix generated through the mode-k flattening of the tensor $\mathbf{X}_i^k = \mathbf{X}_i \times_1 U^{1^T} \ldots \times_{k-1} U^{k-1^T} \times_{k+1} U^{k+1^T} \ldots \times_n U^{n^T}$.

Starting from this result it is possible to develop an iterative algorithm, called Concurrent Subspaces Analysis (CSA) in [4], to compute the optimum projection matrices. The CSA input are the image database and lower dimensions $m\prime_k$, $k = 1 \ldots n$. It starts with an initial guess for the projection matrices $U_0^k \in \Re^{m_k \times m\prime_k}$, $k = 1, \ldots n$, by using columnly-orthogonal matrices ($U^{k^T} U^k = I$), which are usually obtained by truncating the number of columns of the identity matrix.

Algorithm 1 CSA Algorithm

1: Projection matrices initialization U_0^k
2: Distribution centroid: $\mathbf{CM} = (1/N) \sum_{i=1}^{N} \mathbf{X}_i$
3: Translation to distribution centroid: $\mathbf{X}_k \Longleftarrow \mathbf{X}_k - \mathbf{CM}$, $k = 1, \ldots, N$
4: **for** $t = 1, \ldots$ to T_{max} **do**
5: **for** $k = 1, \ldots$ to n **do**
6: Mode-k tensor products $\mathbf{X}_i^k$=$\mathbf{X}_i \times_1 U_t^{1^T} \ldots \times_{k-1} U_t^{(k-1)^T} \times_{k+1} U_{t-1}^{(k+1)^T} \ldots \times_n U_{t-1}^{n^T}$
7: Mode-k flattening $\mathbf{X}_i^k$ for X_i^k: $X_i^k \Longleftarrow_k \mathbf{X}_i^k$
8: Covariance matrix: $C^k = \sum_{i=1}^{N} X_i^k X_i^{k^T}$
9: Compute the first $m_k^{'}$ leading eigenvectors of C^k, $C^k U_t^k = U_t^k \Lambda^k$, which constitute the column vectors of U_t^k
10: **end for**
11: **if** $(t > 2$ and $Tr[abs(U_t^{k^T} U_{t-1}^k)]/m_k^{'} > (1 - \epsilon), k = 1, \ldots, n)$ **then**
12: break;
13: **end if**
14: **end for**
15: Output the matrices $U^k = U_t^k, k = 1, \ldots, n$.

Once the initialization ends, the two main loops start with a sequence of operations derived from the Theorem 1. The outer loop depends on the parameter T_{max} that represents the max number of updates for the projection matrices U_t^k, $k = 1, 2, \ldots, n$, which are computed in the inner loop following the philosophy of Theorem 1: taking the matrices $U_t^1, \ldots, U_t^{k-1}, U_{t-1}^{k+1}, \ldots, U_{t-1}^n$ we seek for the optimum U_{t-1}^k by solving the optimization problem given by equation (8) with respect to the matrix U_{t-1}^k.

IV. Tensor Product and Dimensionality Reduction

The concept of tensor can also be introduced following an algebraic approaches, as performed in [9], which is more suitable for CSA tensor component analysis than the one presented in section III. So, let us consider the vector spaces $\Re^{m_1}, \Re^{m_2}, \cdots, \Re^{m_n}$. The tensor product of these spaces, denoted by $\Re^{m_1} \otimes \Re^{m_2} \otimes \cdots \otimes \Re^{m_n}$,

is another vector space whose elements can be represented by the Kronecker product of collumn vectors $v_i \in \Re^{m_i}$, $i = 1, 2, \cdots, n$. Given individual basis $\left\{\mathbf{e}_k^{i_k}, i_k = 1, 2, \cdots, m_k\right\} \subset \Re^{m_k}$ then a natural basis B for the vector space $\Re^{m_1} \otimes \Re^{m_2} \otimes \cdots \otimes \Re^{m_n}$ is generated by:

$$B = \left\{\mathbf{e}_1^{i_1} \otimes \mathbf{e}_2^{i_2} \otimes \cdots \otimes \mathbf{e}_n^{i_n}, \quad \mathbf{e}_k^{i_k} \in \Re^{m_k}\right\}. \tag{9}$$

In this contex, a tensor $\mathbf{X}$ of order n is defined as an element $\mathbf{X} \in \Re^{m_1} \otimes \Re^{m_2} \otimes \cdots \otimes \Re^{m_n}$;that is, an abstract geometric entity that can be expressed as [9]:

$$\mathbf{X} = \sum_{i_1, i_2, \cdots, i_n} \mathbf{X}_{i_1, i_2, \cdots, i_n} \mathbf{e}_1^{i_1} \otimes \mathbf{e}_2^{i_2} \otimes \cdots \otimes \mathbf{e}_n^{i_n}. \tag{10}$$

Now, let us consider the subspaces $\Re^{m\prime_k}$, $k = 1, 2, \cdots, n$, where $m\prime_k \leq m_k$, and a basis for $\Re^{m\prime_1} \otimes \Re^{m\prime_2} \otimes \cdots \otimes \Re^{m\prime_n}$ given by:

$$\widetilde{B} = \left\{\widetilde{\mathbf{e}}_1^{i_1} \otimes \widetilde{\mathbf{e}}_2^{i_2} \otimes \cdots \otimes \widetilde{\mathbf{e}}_n^{i_n}, \quad \widetilde{\mathbf{e}}_k^{i_k} \in \Re^{m\prime_k}\right\}, \tag{11}$$

as well as $U^k \in \Re^{m_k \times m\prime_k}$, the projection matrix in $\Re^{m\prime_k}$ defined by:

$$\mathbf{e}_k^{i_k} = \sum_{j_k=1}^{m\prime_k} U_{i_k j_k}^k \widetilde{\mathbf{e}}_k^{j_k}, \quad k = 1, 2, \cdots, n \quad and \quad i_k = 1, 2, \cdots, m_k. \tag{12}$$

In the tensor product framework, to get the new representation (projection) of the tensor $\mathbf{X}$ in the basis $\widetilde{B}$ it is just a matter of inserting expression (12) in the tensor product representation given by equation (10); that is:

$$\mathbf{Y} = \sum_{i_1, i_2, \cdots, i_n} \mathbf{X}_{i_1, i_2, \cdots, i_n} \left(\sum_{j_1=1}^{m\prime_1} U_{i_1 j_1}^1 \widetilde{\mathbf{e}}_1^{j_1}\right) \otimes \cdots \otimes \left(\sum_{j_n=1}^{m\prime_n} U_{i_n j_n}^n \widetilde{\mathbf{e}}_n^{j_n}\right). \tag{13}$$

A fundamental result for our work is summarized as follows.

Theorem 2: The elements of the tensor $\mathbf{Y} \in \Re^{m\prime_1} \otimes \Re^{m\prime_2} \otimes \cdots \otimes \Re^{m\prime_n}$ defined by expression (13) can be computed through equation (6); that means:

$$\mathbf{Y} = \sum_{j_1, j_2, \cdots, j_n} \mathbf{Y}_{j_1, j_2, \cdots, j_n} \widetilde{\mathbf{e}}_1^{j_1} \otimes \widetilde{\mathbf{e}}_2^{j_2} \cdots \otimes \widetilde{\mathbf{e}}_n^{j_n}, \tag{14}$$

where:

$$\mathbf{Y}_{j_1, j_2, \cdots, j_n} = \left(\mathbf{X} \times_1 U^{1^T} \times_2 U^{2^T} ... \times_n U^{n^T} \right)_{j_1, j_2, \cdots, j_n}. \tag{15}$$

Proof: See [14].

So, the Theorem 2 shows that the expression (6) can be viewed as a dimensionality reduction in each one of the component spaces of the tensor product $\Re^{m_1} \otimes \Re^{m_2} \otimes \cdots \otimes \Re^{m_n}$. This fact will be explored in the next section.

V. Spectral Structure of CSA Subspace

In this paper, the elements of the new basis:

$$\widetilde{B} = \left\{ \widetilde{\mathbf{e}}_1^{i_1} \otimes \widetilde{\mathbf{e}}_2^{i_2} \otimes \cdots \otimes \widetilde{\mathbf{e}}_n^{i_n}, \quad \widetilde{\mathbf{e}}_k^{i_k} \in \Re^{m\prime_k} \right\}, \tag{16}$$

where the projection matrices $U^k \in \Re^{m_k \times m\prime_k}, k = 1...n$, are computed by the CSA Algorithm are called CSA tensor components.

The point is how to sort these components according to the variance explained by each one? The Appendix A of the report [14] shows that we can not perform this task by computing the spectrum of a covariance matrix, in contrast to the PCA or MPCA [1], [6]. However, in the CSA algorithm each subspace

$$\left\{ \widetilde{\mathbf{e}}_k^{j_k}, \quad j_k = 1, 2, \cdots, m\prime_k \right\}, \quad k = 1, 2, \cdots, n,$$

is obtained by taking the first $m\prime_k$ leading eigenvectors of the covariance matrix C^k. So, let

$$\left\{ \lambda_{j_k}^k, \quad j_k = 1, 2, \cdots, m\prime_k \right\}, \quad k = 1, 2, \cdots, n,$$

the associated eigenvalues. The data distribution in each subspace can be represented by the vector:

$$\mathbf{v}_k = \sum_{j_k=1}^{m\prime_k} \lambda_{j_k}^k \widetilde{\mathbf{e}}_k^{j_k}, \quad k = 1, 2, \cdots, n. \tag{17}$$

Therefore, the variance explained by each element of basis $\widetilde{B}$ in expression (16) can be estimated by calculating:

$$\mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \cdots \otimes \mathbf{v}_n = \sum_{j_1, j_2, \cdots, j_n} \lambda_{j_1}^1 \lambda_{j_2}^2 \cdots \lambda_{j_n}^n \widetilde{\mathbf{e}}_1^{j_1} \otimes \widetilde{\mathbf{e}}_2^{j_2} \otimes \cdots \otimes \widetilde{\mathbf{e}}_n^{j_n}.$$

and taking the corresponding coefficient in order to estimate the covariance structure of the CSA subspace. Consequently, we can rank the CSA tensor components by sorting the elements of the set:

$$E = \left\{ \lambda_{j_1, j_2, \cdots, j_n} = \lambda_{j_1}^1 \lambda_{j_2}^2 \cdots \lambda_{j_n}^n, \quad j_k = 1, 2, \cdots, m\prime_k \right\}, \tag{18}$$

to obtain the principal CSA tensor components. The sorted sequence $\{\lambda_{j_1, j_2, \cdots, j_n}\}$ can be re-indexed by just using one index $\{\lambda_i; \quad i = 1, 2, ..., \prod_{k=1}^n m\prime_k\}$ that tells the number of the principal CSA component in the sorted sequence. The expression (17) is the key of this proposal. Once this expression is computed the ranking method proposed can be applied whatever the tensor decomposition used.

VI. Tensor Discriminant Principal Components

In this section we assume that there are only two classes to separate. Then, following the reference [11], we consider to approach the problem of ranking tensor principal components by estimating a linear classifier.

Firstly, we perform dimensionality reduction by computing the low dimensional representation $\mathbf{Y}_i$ of each tensor $\mathbf{X}_i$ through expression (6). The goal of this step is to discard redundancies of the original representation. Then, the linear SVM classifier is estimated (through expressions (1)-(2)) using the projected training examples $\mathbf{Y}_i \in \Re^{m\prime_1 \times m\prime_2 \times ... \times m\prime_n}$ and the corresponding labels (details are given next). The separating hyperplane is defined through a discriminant tensor $\mathbf{W} \in \Re^{m\prime_1 \times m\prime_2 \times ... \times m\prime_n}$ while the discriminant features given by:

$$\begin{aligned} \tilde{y}_1 &= \langle \mathbf{Y}_1, \mathbf{W} \rangle, \\ \tilde{y}_2 &= \langle \mathbf{Y}_2, \mathbf{W} \rangle, \\ &... \\ \tilde{y}_N &= \langle \mathbf{Y}_N, \mathbf{W} \rangle, \end{aligned} \tag{19}$$

are used for classification ($\langle ., . \rangle$ means internal product given in expression 3).

We can investigate the components $\mathbf{W}_{i_1, .., i_n}$ of the discriminant tensor $\mathbf{W}$ to determine the discriminant contribution for each feature. So, following the same idea proposed in [11] for PCA subspaces, these components are weights in expression (19) that determine the discriminant contribution of each feature $\mathbf{Y}_{i; i_1, .., i_n}$. Therefore, if a weight $\mathbf{W}_{i_1, .., i_n}$ is approximately 0 this fact indicates that the corresponding feature $\mathbf{Y}_{i; i_1, .., i_n}$ is not significant to separate the sample groups. In contrast, largest weights (in absolute values) indicate that the corresponding features contribute more to the discriminant score and consequently are important to characterize the differences between the groups [15].

Therefore, we are selecting among the CSA components the "directions" that are efficient for discriminating the sample groups rather than representing all the samples, as performed in section V. The obtained components, ordered in decreasing order of the discriminant weights, are called here the tensor discriminant principal components (TDPCA).

Such idea can be seen from a matrix viewpoint using the Kronecker product. It can be shown that we can vectorize the database tensors $\mathbf{X}_i$ and the reduced representations $\mathbf{Y}_i$ in the column arrays $\mathbf{x}_i^v$ and $\mathbf{y}_i^v$, respectively, such that $\mathbf{y}_i^v = P^T \mathbf{x}_i^v$, where $P = U^n \bigotimes U^{n-1} \bigotimes, ..., \bigotimes U^1$, with the matrices U^j given by the CSA algorithm [14]. Therefore, the expressions in (19) can be re-written as:

$$
\begin{aligned}
\tilde{y}_1 &= y^v_{11}w_1 + y^v_{12}w_2 + ... + y^v_{1m}w_m, \\
\tilde{y}_2 &= y^v_{21}w_1 + y^v_{22}w_2 + ... + y^v_{2m}w_m, \\
&... \\
\tilde{y}_N &= y^v_{N1}w_1 + y^v_{N2}w_2 + ... + y^v_{Nm}w_m,
\end{aligned} \tag{20}
$$

where $m = m\prime_1 \cdot m\prime_2 \ldots m\prime_n$ and $[w_1, w_2, ..., w_m]$ are the weights corresponding to the principal component features calculated by the linear classifier, and $[y^v_{i1}, y^v_{i2}, ..., y^v_{im}]$ are the attributes of each data vector i, where $i = 1, ..., N$, projected on the full rank CSA space [11].

In other words, the zero mean data vectors are projected on the CSA components and reduced to m-dimensional vectors representing the most expressive features. Afterwards, the $N \times m$ data matrix and their corresponding labels are used as input to calculate the SVM separating hyperplanes [11]. Finally, instead of selecting the tensor principal components by estimating their spectral structure (sorting the set E in expression (18) in decreasing order), we select as the first principal TDPCA components the ones with the highest discriminant weights, that is, $[\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_m]$, where $\mathbf{p}_i, \quad i = 1, 2, \ldots m$, are the columns of P corresponding to the largest discriminant weights $|w_1| \geq |w_2| \geq \ldots \geq |w_m|$ given by the separating hyperplane computed through the linear classifier; in this case, the SVM one. We must highlight that SVM can be replaced by any other separating hyperplane. Besides, from expression (19) we conclude that the TDPCA methodology is not restricted to the CSA components.

VII. Experimental results

In our experiments we have used the image face database maintained by the Department of Electrical Engineering of FEI, São Paulo, Brazil. There are 14 images for each of 200 individuals, a total of 2800 images. All images are colorful and taken against a white homogenous background in an upright frontal position with 11 neutral profile rotation of up to about 180 degrees, one facial expression (smile) and two poses with variations in illumination. Scale might vary about 10% and the original size of each image is 640×480 pixels. All faces are mainly represented by students and staff at FEI, between 19 and 40 years old with distinct appearance, hairstyle, and adorns. The number of male and female subjects are exactly the same and equal to 100 [11], [15].

Figure 2 shows an example of 11 neutral profile rotations that have been used in all experiments. For memory requirements, we convert each pose to gray scale before computations. We observe that the images are well scaled and aligned. These features make the FEI database very attractive for testing dimensionality reduction in tensor spaces.

In the first part of the experimental results (section VII-A) we have carried out a gender image analysis (female versus male samples) and visualization. Next, in section VII-B, we have investigated the efficiency of the ranking approaches on recognizing samples.

Fig. 2. The 11 neutral profile rotations of an individual in the FEI database.

A. Understanding and Visualizing the Tensor Components

To avoid the trivial solution for CSA, we run the Algorithm 1 for $U^1_0 \in \Re^{480\times479}$, $U^2_0 \in \Re^{640\times639}$ and $U^3_0 \in \Re^{11\times11}$ which performs a total of $479 \cdot 639 \cdot 11 = 3366891 \approx 3,36 \cdot 10^6$ CSA tensor components, according to expression (16). These components are sorted according to their variances that are computed following expression (18). We have observed that the variances fall in the range $[1.0333 \times 10^4, 2.3 \times 10^{-3}]$. Therefore, we retain all the tensor components performing a full rank CSA subspace with dimension $m = 3366891$.

Now, we determine the discriminant contribution of each CSA component by investigating the weights of the tensor "direction" found by the SVM approach. Table I lists the 60 CSA components with the highest weights in absolute values for discriminating the gender samples (see [14], page 17, for top 100 list). We observe that TDPCA with SVM has selected the $235th$, $184th$, $94th$, and $91th$ principal CSA components between the top 10 most discriminant tensor components. Among the top 60 we get the $2008th$, $945th$ and $829th$ principal CSA components. Since principal components with lower variances describe particular information related to few samples, these results confirm the SVM ability of zooming into the details of group differences.

Gender Experiment: TDPCA components given by SVM									
3	47	91	235	11	184	35	31	4	94
204	132	15	1	24	48	37	66	34	52
64	98	336	469	61	113	23	12	42	54
9	6	157	702	105	190	173	100	608	530
277	543	656	25	143	8	14	862	167	89
945	829	21	782	96	2008	72	214	29	20

TABLE I
TOP 60 (FROM LEFT TO RIGHT AND TOP TO BOTTOM) TENSOR DISCRIMINANT PRINCIPAL COMPONENTS, RANKED BY SVM HYPERPLANE, USING THE FEI DATABASE.

The total variance explained by the 400 CSA most expressive and TDPCA components for the gender experiment is illustrated in Figure 3. This figure shows that as the dimension of the CSA most expressive subspace increases, there is an exponential decrease in the amount of total variance explained by the first CSA tensor principal components. A similar behavior for standard PCA was also reported in [11].

However, the corresponding variances explained by the TDPCA components do not follow the same behavior. Specifically, we observe in Figure 3 some oscillations along the

SVM spectrum. For instance, the amount of the total variance explained by the $101-120$ TDPCA components is a local maximum for the SVM distribution while the amount of the total variance explained by the $181-200$ is a local minimum (page 18 of [14] highlights this fact by considering total variance for 1400 components).

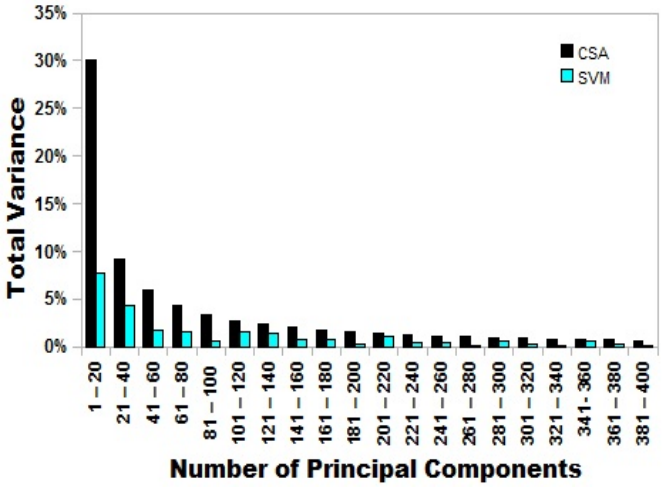

Fig. 3. Amount of total variance explained by the 400 CSA most expressive tensor components (black columns), and 400 SVM most discriminant tensor components (green columns).

These results indicate that despite of the substantial differences between male and female there are some artifacts that vary on several images but they are not related to gender characteristics. So, they should not be considered as discriminant information.

To understand the changes described by the principal tensor directions, we have reconstructed the CSA most expressive features by varying each principal tensor component. Firstly, we shall observe that each CSA tensor component in expression (16) can be written as:

$$\mathbf{\Phi}^{\alpha\beta\gamma} = \sum_{j_1 j_2 j_3}^{\alpha\beta\gamma} \mathbf{\Phi}^{\alpha\beta\gamma}_{j_1 j_2 j_3} \widetilde{\mathbf{e}}_1^{j_1} \otimes \widetilde{\mathbf{e}}_2^{j_2} \otimes \widetilde{\mathbf{e}}_3^{j_3}, \tag{21}$$

where $\mathbf{\Phi}^{\alpha\beta\gamma}_{j_1 j_2 j_3} = 1$ if $(j_1, j_2, j_3) = (\alpha, \beta, \gamma)$ and $\mathbf{\Phi}^{\alpha\beta\gamma}_{j_1 j_2 j_3} = 0$ otherwise. Therefore, expression (7) allows to write the reconstruction of $\mathbf{\Phi}^{\alpha\beta\gamma}$ as:

$$\mathbf{\Phi}_R^{\alpha\beta\gamma} = \sum_{i_1 i_2 i_3} \left(\mathbf{\Phi}^{\alpha\beta\gamma} \times_1 U_1 \times_2 U_2 \times_3 U_3\right)_{i_1 i_2 i_3} \mathbf{e}_1^{i_1} \otimes \mathbf{e}_2^{i_2} \otimes \mathbf{e}_3^{i_3} \tag{22}$$

So, by performing the operation:

$$\mathbf{T}(\lambda) = \mathbf{CM} + \lambda \mathbf{\Phi}_R^{\alpha\beta\gamma}, \quad \lambda \in \Re, \tag{23}$$

with $\mathbf{CM}$ been the centroid computed in line 2 of Algorithm 1, we can visualize each CSA tensor component in the original data space. The Figure 4 shows the frontal pose of tensor $\mathbf{CM}$. Following the work [11], we choose $\lambda \in \left\{\pm 3\sqrt{\lambda_{\alpha\beta\gamma}}, \pm 2\sqrt{\lambda_{\alpha\beta\gamma}}\right\}$, where $\lambda_{\alpha\beta\gamma}$ is the corresponding variance of the tensor component computed in expression (18).

Fig. 4. Average frontal pose.

Figure 5 illustrates the changes on the first two CSA most expressive components against the first two tensor discriminant principal components selected by the SVM hyperplane in the gender experiment. For simplicity, in Figure 5 we are visualizing only the frontal poses of $\mathbf{T}(\lambda)$. It can be seen that the first CSA most expressive directions captures essentially the changes in illumination and some aspects of gender, which are the major variations of all the training samples.

Fig. 5. Visualization of the changes described by the tensor principal components for gender. From top to bottom: the first two CSA most expressive component (first and second rows), the first two TDPCA components selected by SVM (third and fourth rows).

When we compare these results with the ones reconstructed by the TDPCA selected by SVM, illustrated in the third and fourth rows of Figure 5, we observe that it is more noticeable the changes from female (left most) to male (right most) in the third row.

Therefore, even in this situation where the classes are well separated, we observe that the discriminant tensor principal components ranked by SVM are more effective for extracting group differences than the principal CSA ones. This fact should have consequences in the classification rate when considering CSA subspaces selected by both criteria, as we shall see in section VII-B.

For the principal components selected by the SVM separating hyperplane, we must be careful about the above limits for λ because some λ_{j_1,j_2,j_3} can be very small in this case, showing no changes between the samples when we move along

the corresponding principal components. For instance, let us consider the $199th$ TDPCA principal components selected by SVM (the $90th$ principal CSA tensor component). According to the Figure 3, this component counts for less than 5% of the total variance explained by the tensor components. Therefore, if we use only the associated variance, we get no changes, as observed in the first row of Figure 6. Also, the mean variance $\bar{\lambda} = (\sum_{i=1}^{i=m} \lambda_i)/m$, used in [11], is not efficient in this case as we see in the second row (middle) of Figure 6. Other possibility is to observe that the first 90 principal CSA tensor components corresponds to a $51,17\%$ of the total variance. So, we can try to use the partial mean variance $\bar{\lambda}_{90} = (\sum_{i=1}^{i=90} \lambda_i)/90$ in expression (23), which is visualized on the third row of Figure 6 showing few changes in the gender features. Therefore, we decided to use the $\max\{\lambda_i, i = 1, 2, ..., 3366891\}$ whose results are observed in the forth row (bottom) of the Figure 6. In this case the changes in gender features are more noticeable.

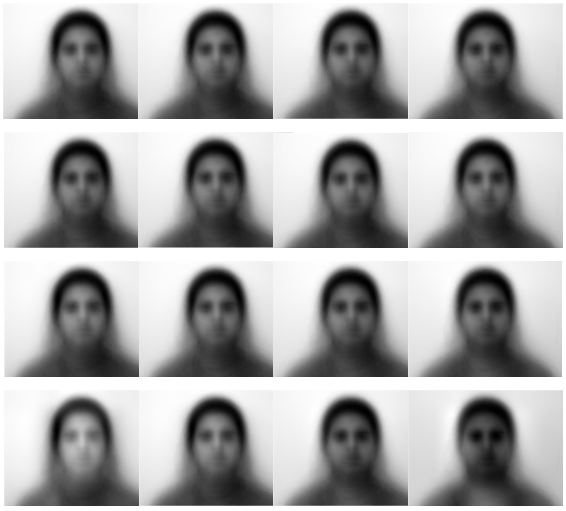

Fig. 6. Visualization of the changes described by the $199th$ TDPCA principal components selected by SVM (the $90th$ principal CSA component). The first row (top) uses λ_{90};the second row uses the global mean $\bar{\lambda}$; the third row applies the partial mean $\bar{\lambda}_{90}$ and the last rows (bottom) uses $\max\{\lambda_i, i = 1, 2, ..., 3366891\}$.

In order to quantify the discriminant power of the principal components, we present in Figure 7 the amount of total discriminant information, in descending order, explained by each one of the first 400 tensor principal components selected by the SVM approach. The proportion of total discriminant information t described by the k^{th} tensor principal component can be calculated as follows:

$$t_k = \frac{|w_k|}{\sum_{j=1}^{m} |w_j|}, \tag{24}$$

where $m = 3366891$, $k = [1, 2, ..., m]$ and $[w_1, w_2, ..., w_m]$ are the weights corresponding to the tensor principal components ranked by the SVM separating hyperplane.

Figure 7 shows that as the dimension of the TDPCA subspace increases there is an exponential decrease in the amount of total discriminant information described by the SVM principal components.

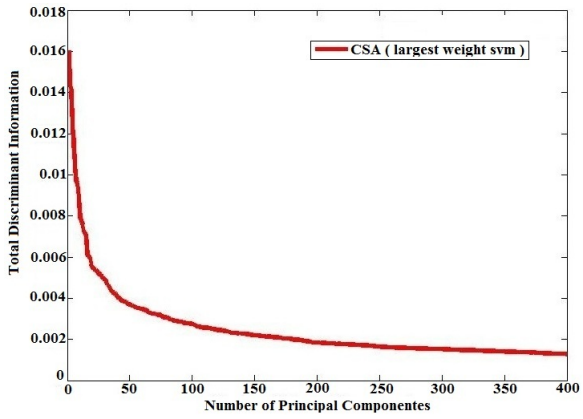

Fig. 7. Amount of total discriminant information calculated in equation (24) and explained by SVM tensor principal components for gender experiment.

B. Recognition Rates of Tensor Components

In this section, we have investigated the effectiveness of the TDPCA components on a recognition task against the CSA components ranked according to the variances computed in expression (18).

The 10-fold cross validation method is used to evaluate the classification performance of the most expressive (CSA largest variances) and most discriminant tensor components selected through SVM weights. In these experiments we have assumed equal prior probabilities and misclassification costs for both groups. We calculate the mean of each class i on the CSA subspace and, given a test observation $\mathbf{X}_t$, we compute the Mahalanobis distance from each class mean $\overline{\mathbf{X}}_i$ to assign that observation to either the male or female groups. That is, we have assigned $\mathbf{X}_t$ to class i that minimizes:

$$d_i(\mathbf{X}_t) = \sum_{j=1}^{k} \frac{1}{\lambda_j} (\mathbf{X}_{t;j} - \overline{\mathbf{X}}_{i;j})^2 \tag{25}$$

where λ_j is the corresponding variance and k is the number of tensor principal components retained. In the recognition experiments, we have considered different number of tensor principal components ($k = 2, 5, 50, 100, 200, 400, 900, 1000, 1200, 1400, 1800, 2000$) to calculate the recognition rates of the corresponding methods of selecting principal components.

Figure 8 shows the average recognition rates of the 10-fold cross validation. For the smaller numbers of components considered ($k = 2$ and $k = 5$) we observe that TDPCA achieves higher or equal recognition rates than the CSA highest variances components. When taking $k \in [50, 1400)$ the TDPCA approach using SVM weights performs better than CSA largest variation subspaces. On the other hand, for $k \in [1400, 2000]$ the principal CSA tensor components gets higher recognition rates.

By using the standard deviation reported on Table II and the average recognition rate pictured on Figure 8 we compute

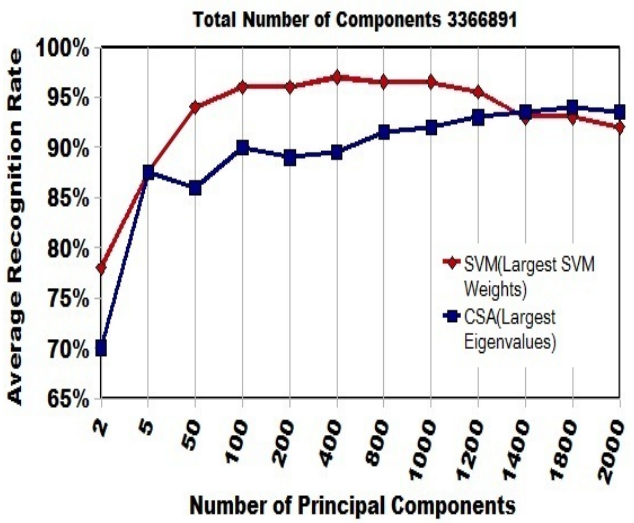

Fig. 8. Gender experiments using the FEI database. Average recognition rate of CSA components selected by the largest variances, and largest SVM weights criteria.

the statistical t-student test [2]. For $50 \leq k \leq 400$ we get $p < 0.0204$ which shows that the observed superiority of TDPCA in this range is statistically significant.

Number of Principal Components (k) × Standard Deviation %												
k	2	5	50	100	200	400	800	1000	1200	1400	1800	2000
CSA	9,5	6,8	8,3	5,5	7,0	8,2	7,4	7,1	6,8	6,7	6,2	6,3
TDPCA	11,7	8,1	4,4	3,7	3,7	3,3	5,0	5,0	6,1	5,1	6,8	8,1

TABLE II
SECOND AND THIRD LINES REPORT STANDARD DEVIATIONS WHEN INCREASING k.

Despite of these nuances the recognition rates for $k > 100$ are higher than 0.875 for both approaches. In particular, for $k = 100$ they are higher than 0.9 which is impressive, since we have retained only 100 components of a full rank subspace with dimension $m = 3366891$. The fact that for smaller numbers of components ($k = 2$) the TDPCA with SVM performs better than the CSA largest variation components can be explained by returning to the Figure 5 and observing that the TDPCA first component (third row) is more efficient to describe the changes from female to male then the CSA one (first row). On the other hand the fact that CSA largest variation components perform quite better than TDPCA for $k > 1400$ may be related with incorporating some noisy information.

The TDPCA approach is limited to two-class problems and its computational cost is higher than the covariance structure estimation because TDPCA includes the separating hyperplane computation. By observing the recognition rates reported above we consider that both approaches are competitive for two-class classification tasks depending on the trade-off between complexity and precision.

VIII. Conclusions and Future Works

In this paper we addressed the problem of sorting tensor components, in the context of the CSA technique. We follow two distinct approaches: (a) Estimating the covariance structure of all the data; (b) Computing discriminant weights, given by separating hyperplanes, to select the most discriminant CSA tensor components (TDPCA technique). In the experimental results, we show a low dimensional data representation of both tensor ranking techniques in a gender experiment performed using the FEI database. Although the TDPCA achieves quite better results, we observe robust discriminant reconstruction and high recognition rates in the classification task of both techniques.

Further research directions are to consider the LDA separating hyperplanes for the TDPCA methodology, test the approaches for colorful image analysis using FEI and FERET [2] databases and compare the methods with MPCA and TDA. Respect to the former, we shall observe that in [11] it is reported that since PCA explains features that most vary in the samples the first principal components do not necessarily represent important discriminant directions to separate sample groups. We will check this fact for MPCA as well.

References

[1] D. Engel, L. Hüttenberger, and B. Hamann, "A Survey of Dimension Reduction Methods for High-dimensional Data Analysis and Visualization," in *Proceedings of IRTG 1131 Workshop 2011*, vol. 27. Germany: Schloss Dagstuhl, 2012, pp. 135–149.

[2] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer, 2001.

[3] D. T. M. Alex O. Vasilescu, "Multilinear analysis of image ensembles: Tensorfaces," vol. 447/460, 2002.

[4] D. Xu, L. Zheng, S. Lin, H.-J. Zhang, and T. S. Huang, "Reconstruction and recognition of tensor-based objects with concurrent subspaces analisys," *IEEE Trans. on Circ. and Syst. for Video Tech.*, vol. 18, no. 1, pp. 36–47, 2008.

[5] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear independent components analysis," in *IEEE COMP. COMP. VIS AND PAT. REC. (CVPR)*, 2005, pp. 547–553.

[6] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "MPCA: Multilinear principal component analysis of tensor objects," *IEEE Trans. on Neural Networks*, vol. 19, no. 1, pp. 18–39, 2008.

[7] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and gabor features for gait recognition," *IEEE Trans. on Patt. Analysis and Machine Intell.*, vol. 29, no. 10, pp. 1700–1715, 2007.

[8] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, and H.-J. Zhang, "Multilinear discriminant analysis for face recognition," *Trans. Img. Proc.*, vol. 16, no. 1, pp. 212–220, jan 2007.

[9] S. Liu and Q. Ruan, "Orthogonal tensor neighborhood preserving embedding for facial expression recognition," *Pattern Recognition*, vol. 44, no. 7, pp. 1497 – 1513, 2011.

[10] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "A survey of multilinear subspace learning for tensor data," *Pattern Recogn.*, vol. 44, no. 7, pp. 1540–1551, July 2011.

[11] C. E. Thomaz and G. A. Giraldi, "A new ranking method for principal components analysis and its application to face image analysis," *Image Vision Comput.*, vol. 28, no. 6, pp. 902–913, June 2010.

[12] V. N. Vapnik, *Statistical Learning Theory*. John Wiley & Sons, INC., 1998.

[13] B. W. Bader, T. G. Kolda, and B. W. Bader, "Matlab tensor classes for fast algorithm prototyping," ACM Trans. Math. Software, Tech. Rep., 2004.

[14] T. A. Filisbino, G. A. Giraldi, and C. E. Thomaz, "Defining and sorting tensor components for face image analysis," Tech. Report LNCC, Tech. Rep., 2013.

[15] G. A. Giraldi, P. S. Rodrigues, E. C. Kitani, J. R. Sato, and C. E. Thomaz, "Statistical learning approaches for discriminant features selection," *J. Braz. Comp. Soc.*, vol. 14, no. 2, pp. 7–22, 2008.

[2] www.itl.nist.gov/iad/humanid/feret/feret_master.html