Fusion Tensor Subspace Transformation Framework

Su-Jing Wang[1,2,3,*], Chun-Guang Zhou[2,3], Xiaolan Fu[1]

1 State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China
2 College of Computer Science and Technology, Jilin University, Changchun, Jilin, China.
3 Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun, 130012, P.R. China.
∗ **E-mail: Corresponding wangsujing@psych.cn.cn**

Abstract

Tensor subspace transformation, a commonly used subspace transformation technique, has gained more and more popularity over the past few years because many objects in the real world can be naturally represented as multidimensional arrays, i.e. tensors. For example, a RGB facial image can be represented as a three-dimensional array (or 3rd-order tensor). The first two dimensionalities (or modes) represent the facial spatial information and the third dimensionality (or mode) represents the color space information. Each mode of the tensor may express a different semantic meaning. Thus different transformation strategies should be applied to different modes of the tensor according to their semantic meanings to obtain the best performance. To the best of our knowledge, there are no existing tensor subspace transformation algorithm which implements different transformation strategies on different modes of a tensor accordingly. In this paper, we propose a fusion tensor subspace transformation framework, a novel idea where different transformation strategies are implemented on separate modes of a tensor. Under the framework, we propose the Fusion Tensor Color Space (FTCS) model for face recognition.

Introduction

Subspace transformation (or subspace analysis [1]), a main type of feature extraction, has gained huge popularity over the past few years. Principal Component Analysis (PCA) [2] seeks the optimal projection directions according to maximal variances. Linear Discriminant Analysis (LDA) [3] uses discriminant information to search for the directions which are most effective for discrimination by maximizing the ratio between the between-class and within-class scatters. Both PCA and LDA aim to preserve global structures of the samples. Locality Preserving Projections (LPP) [4] aims to preserve the local structure of the original space in the projective subspace. Discriminant Locality Preserving Projections (DLPP) [5] encodes discriminant information into LPP to further improve the discriminant performance of LPP for face recognition. These algorithms need to vectorize the objects (samples).

In the real world, however, many objects are naturally represented by multidimensional arrays, i.e., tensors, such as a color facial image used in face recognition (see Fig. 1). If these objects are vectorized, their natural structure information will be lost [6]. As such, a great deal of interests are aroused in the field of tensor [7] [8] [9] [10] [11]. Among the subspace transformation techniques, tensor subspace transformation has also become a highly discussed topic. Multilinear Principal Component Analysis (MPCA) [12], a tensor version of PCA, applies PCA transformation on each mode (or dimensionality) of tensors. Similarly, Discriminant Analysis with Tensor Representation (DATER) [13], General Tensor Discriminant Analysis (GTDA) [14], Tensor Subspace Analysis (TSA) [15], and Discriminant Tensor Subspace Analysis (DTSA) [16] apply LDA, Maximum Scatter Difference (MSD) [17], LPP, and DLPP to transform each mode of tensors, respectively. These tensor subspace transformation methods use a certain vector subspace transformation method to transform every modes of tensors.

However, each mode of tensors may express a different semantic meaning. For example, a color facial image can be treated as a 3rd-order tensor, where mode-1 and mode-2 represent the facial spatial infor-
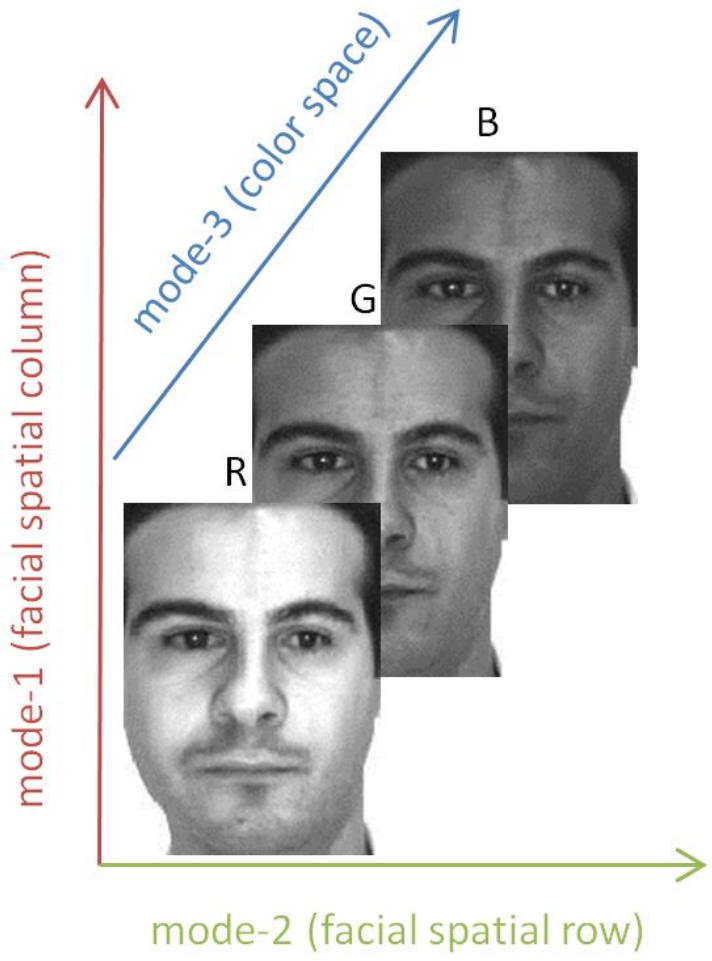
Figure 1. the tensorial represents of a color facial image.

mation and mode-3 representing the color space information (see Fig. 1). The facial spatial information and color space information are two different types of information, which should be handled by two different transformations to obtain better performance. In other words, for color facial images, we should implement a transformation strategy on the first two modes and another transformation strategy on the third mode. As such, each type of information should implement a transformation strategy best suited for the semantic meaning.

To the best of our knowledge, there are no existing tensor subspace transformation algorithm, which implements different transformation strategies on different modes of tensors according to their semantic meanings. To address this problem, we propose the fusion tensor subspace transformation framework, which shows an novel idea that different transformation strategies can be implemented on different modes of tensors. Under the framework, we propose the Fusion Tensor Color Space (FTCS) model for face recognition.

Materials and Methods

Tensor Fundamentals and Denotations

A tensor is a multidimensional array. It is the higher-order generalization of scaler (zero-order tensor), vector (1st-order tensor), and matrix (2nd-order tensor). In this paper, lowercase italic letters (a, b, ...) denote scalars, bold lowercase letters ($\mathbf{a}$, $\mathbf{b}$, ...) denote vectors, bold uppercase letters ($\mathbf{A}$, $\mathbf{B}$, ...) denote matrices, and calligraphic uppercase letters ($\mathcal{A}$, $\mathcal{B}$, ...) denote tensors. The formal definition is given below [18]:

Definition 1 *The* order *of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ is N. An element of $\mathcal{A}$ is denoted by $\mathcal{A}_{i_1 i_2 \dots i_N}$ or $a_{i_1 i_2 \dots i_N}$, where $1 \leq i_n \leq I_n$, $n = 1, 2, \dots, N$.*

Definition 2 *The* mode-n vectors *of $\mathcal{A}$ are the I_n-dimensional vectors obtained from $\mathcal{A}$ by fixing every index but index i_n.*

Definition 3 *The* mode-n unfolding matrix *of $\mathcal{A}$, denoted by $(\mathcal{A})_{(n)} \in \mathbb{R}^{I_n \times (I_1 \dots I_{n-1} I_{n+1} \dots \times I_N)}$, contains the element $a_{i_1 \dots i_N}$ at i_nth row and at jth column, where*

$$j = 1 + \sum_{k=1, k \neq n}^{N} (i_k - 1) J_k, \qquad with \qquad J_k = \prod_{m=1, m \neq n}^{k-1} I_m. \tag{1}$$

We can generalize the product of two matrices to the product of a tensor and a matrix.

Definition 4 *The* mode-n product *of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ by a matrix $\mathbf{U} \in \mathbb{R}^{J_n \times I_n}$, denoted by $\mathcal{A} \times_n \mathbf{U}$, is an $(I_1 \times I_2 \times \dots \times I_{n-1} \times J_n \times I_{n+1} \times \dots \times I_N)$-tensor of which the entries are given by:*

$$(\mathcal{A} \times_n \mathbf{U})_{i_1 i_2 \dots i_{n-1} j_n i_{n+1} \dots i_N} \stackrel{def}{=} \sum_{i_n} a_{i_1 i_2 \dots i_{n-1} i_n i_{n+1} \dots i_N} u_{j_n i_n}. \tag{2}$$

Definition 5 *The* scalar product *of two tensors $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$, denoted by $\langle \mathcal{A}, \mathcal{B} \rangle$, is defined in a straightforward way as $\langle \mathcal{A}, \mathcal{B} \rangle \stackrel{def}{=} \sum_{i_1} \sum_{i_2} \dots \sum_{i_N} a_{i_1 i_2 \dots i_N} b_{i_1 i_2 \dots i_N}$. The* Frobenius norm *of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ is then defined as $\|\mathcal{A}\|_F \stackrel{def}{=} \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle}$*

From the definition of the mode-n unfolding matrix, we have

$$\|\mathcal{A}\|_F = \|(\mathbf{A})_{(n)}\|_F \tag{3}$$

By using tensor decomposition, any tensor $\mathcal{A}$ can be expressed as the product

$$\mathcal{A} = \mathcal{C} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \ldots \times_N \mathbf{U}_N \tag{4}$$

where $\mathbf{U}_n$, $n = 1, 2, \ldots, N$, is an orthonormal matrix and contains the ordered principal components for the nth mode. $\mathcal{C}$ is called the *core tensor*. Unfolding the above equation, we have

$$\mathbf{A}_{(n)} = \mathbf{U}_n \mathbf{C}_{(n)} (\mathbf{U}_N \otimes \ldots \otimes \mathbf{U}_{n+1} \otimes \mathbf{U}_{n-1} \otimes \ldots \otimes \mathbf{U}_1)^T \tag{5}$$

where operator $\otimes$ is the Kronecker product of the matrices.

The Connection among PCA, 2D-PCA and MPCA

Before introducing the fusion tensor subspace transformation framework, we firstly investigate the connection among PCA, 2D-PCA [19] and MPCA. From the previous section, we know that a tensor is the higher-order generalization of scaler, vector and matrix. Similarly, MPCA is the higher-order generalization of PCA and 2D-PCA.

Suppose there are M N-order tensors $\mathcal{X}_i \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$, $i = 1, 2, \ldots, M$. MPCA seeks N projection matrices $\mathbf{U}_1 \in \mathbb{R}^{I_1 \times L_1}$, $\mathbf{U}_2 \in \mathbb{R}^{I_2 \times L_2}, \ldots, \mathbf{U}_N \in \mathbb{R}^{I_N \times L_N}$ in order to transform $\mathcal{X}_i$ as

$$\begin{aligned} \mathcal{Y}_i &= \mathcal{X}_i \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \times_3 \ldots \times_N \mathbf{U}_N^T, \\ i &= 1, 2, \ldots, M. \end{aligned} \tag{6}$$

such that $\mathcal{Y}_i, (i = 1, 2, \ldots, M)$ captures most of the variations observed in the original tensor objects $\mathcal{X}_i$.

Here, N projection matrices $\mathbf{U}_n$ $(n = 1, 2, \ldots, N)$ need to be determined. Generally, we fix $N-1$ projection matrices $\mathbf{U}_1, \ldots, \mathbf{U}_{n-1}, \mathbf{U}_{n+1}, \ldots, \mathbf{U}_N$ to solve for $\mathbf{U}_n$. In MPCA, $\mathbf{U}_n$ consists of the L_n eigenvectors corresponding to the largest L_n eigenvalues of the matrix

$$\begin{aligned} \mathbf{\Phi}^{(n)} &= \sum_{m=1}^{M} (\mathbf{X}_{m(n)} - \bar{\mathbf{X}}_{(n)}) \cdot \mathbf{U}_{\mathbf{\Phi}^{(n)}} \\ &\cdot \mathbf{U}_{\mathbf{\Phi}^{(n)}}^T \cdot (\mathbf{X}_{m(n)} - \bar{\mathbf{X}}_{(n)})^T \end{aligned} \tag{7}$$

where $\bar{\mathbf{X}}_{(n)}$ denotes the mode-n unfolding matrix of the mean values of $\mathcal{X}$, $\mathbf{X}_{m(n)}$ denotes the mode-n unfolding matrix of the mean values of $\mathcal{X}_m$ and

$$\mathbf{U}_{\mathbf{\Phi}^{(n)}} = (\mathbf{U}_{n+1} \otimes \mathbf{U}_{n+2} \otimes \ldots \otimes \mathbf{U}_N \otimes \mathbf{U}_1 \otimes \ldots \otimes \mathbf{U}_{n-1}). \tag{8}$$

In Eq. (7), $\mathbf{U}_{\mathbf{\Phi}^{(n)}}$ is to use the fixed $N-1$ projection matrices $\mathbf{U}_1, \ldots, \mathbf{U}_{n-1}, \mathbf{U}_{n+1}, \ldots, \mathbf{U}_N$ to transform the corresponding $N-1$ modes.

When $\mathcal{X}_i$ $(i = 1, 2, \ldots, M)$ is a 2nd-order tensor, Eq. (6) is simplified to

$$\mathbf{Y}_i = \mathbf{X}_i \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T, \qquad i = 1, 2, \ldots, M. \tag{9}$$

If we only transform mode-2, $\mathbf{U}_1$ is an identity matrix of size I_1. Then, $\mathbf{U}_{\mathbf{\Phi}^{(2)}}$ is also an identity matrix and $\mathbf{X}_{m(2)} = \mathbf{X}^T$. In this case, Eq. (7) is simplified to

$$\begin{aligned} \mathbf{\Phi}^{(2)} &= \sum_{m=1}^{M} (\mathbf{X}_m^T - \bar{\mathbf{X}}^T) \cdot (\mathbf{X}_m^T - \bar{\mathbf{X}}^T)^T \\ &= \sum_{m=1}^{M} (\mathbf{X}_m - \bar{\mathbf{X}})^T \cdot (\mathbf{X}_m - \bar{\mathbf{X}}) \end{aligned} \tag{10}$$

Eq. (10) is exactly the image covariance (scatter) matrix $\mathbf{G}_t$ in 2D-PCA [19]. So, 2D-PCA is a special case of MPCA. When objects are represented by matrices and only rows of matrices need to be transformed, MPCA degenerates into 2D-PCA.

When $\mathcal{X}_i, (i = 1, 2, \ldots, M)$ are 1st-order tensors, Eq. (6) is simplified to

$$\mathbf{y}_i = \mathbf{x}_i \times_1 \mathbf{U}_1^T, \qquad i = 1, 2, \ldots, M. \tag{11}$$

In this case, Eq (7) is simplified to

$$\boldsymbol{\Phi}^{(1)} = \sum_{m=1}^{M} (\mathbf{x}_m - \bar{\mathbf{x}}) \cdot (\mathbf{x}_m - \bar{\mathbf{x}})^T \tag{12}$$

The above equation is exactly the scatter matrix in PCA. So, PCA is a special case of MPCA. When objects are represented by vectors, MPCA degenerates into PCA.

Following the above analysis, 2D-PCA applies PCA transformation on rows of matrices, and MPCA applies PCA transformation on all modes of tensors.

Similarly, through the above analysis, one can notice that DATER also applies LDA transformation on all modes of tensors. Likewise, GTDA, TSA and DTSA also applies MSD, LPP and DLPP transformation on all modes of tensors respectively. There are several other tensor subspace transformation methods that also applies a single type of transformation on all modes of tensors, however due to page limit we chose to only mention a portion of these algorithms.

Fusion tensor subspace transformation framework

Tensor subspace transformation method firstly initializes N projection matrices $\mathbf{U}_1, \ldots, \mathbf{U}_{n-1}, \mathbf{U}_n, \mathbf{U}_{n+1}, \ldots, \mathbf{U}_N$ as identity matrices or random matrices, then fixes $N-1$ projection matrices $\mathbf{U}_1, \ldots, \mathbf{U}_{n-1}, \mathbf{U}_{n+1}, \ldots, \mathbf{U}_N$. Following, the matrices are used to transform $\mathcal{X}_i$, and the transformed results are unfolded on mode-n. Finally, $\mathbf{U}_n$ is obtained by implementing a certain transformation on the mode-n unfolding matrices. We can see that the solution of $\mathbf{U}_n$ depends on the other projection matrices. N projection matrices are solved by constructing an iterative procedure.

Existing tensor subspace transformation methods only implement one transformation strategy on all modes. In the real world, each mode of tensors may represent a different type of information. We should implement different transformation strategies on different modes according to their semantic meaning. So we propose a Fusion Tensor Subspace Transformation (FTSA) framework, which is described in Algorithm 1. In Algorithm 1, the statement denoted (*) is the core statement in the framework. For a certain represented object, different transformations are used on different modes according to their semantic meaning.

In the algorithm, we use the maximal iterative times T_{max} to deal with the problem that the algorithms may be not convergent. Actually, the convergence of many tensor subspace analysis algorithms cannot be generally proved, the classification results based on these algorithms show to be stable after rounds of iterations as illustrated in these papers (e.g. DATER, 2D LDA [20]). The convergence of FTSA depends on the specific transformations.

Under the framework, we developed the Fusion Tensor Color Space (FTCS) model for face recognition.

Fusion tensor color space model

Recently, researches showed that color information may help to improve the face recognition accuracy. While, the R, G, and B component images in the RGB color space are correlated. Decorrelation among the components of these images helps reduce redundancy and is an important strategy to improve the accuracy of subsequent recognition method [21]. Liu [22] proposed the Uncorrelated Color Space (UCS),

---

Algorithm 1 Fusion tensor subspace transformation framework

INPUT: M N-order tensors $\mathcal{X}_i \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$, $i = 1, 2, \ldots, M$.
OUTPUT: $\mathbf{U}_n$, $n = 1, 2, \ldots, N$.
Initialize $\mathbf{U}_n$ with a set of identity matrices;
for $t = 1$ to T_{max} **do**
  for $n = 1$ to N **do**
    for $i = 1$ to M **do**
      $\mathcal{Y}_i^{(n)} = \mathcal{X}_i \times_1 \mathbf{U}_1^T \ldots \times_{n-1} \mathbf{U}_{n-1}^T \times_{n+1} \mathbf{U}_{n+1}^T \ldots \times_N \mathbf{U}_N^T$;
      $\mathbf{Y}_{i(n)} \leftarrow$ the mode-n unfolding matrix of $\mathcal{Y}_i^{(n)}$;
    end for
    n-th transformation on M matrices $\mathbf{Y}_{i(n)}$ to obtain $\mathbf{U}_n$; (*)
  end for
  if $t > 2$ and $\|\mathbf{U}_n - \mathbf{U}_n^{pre}\|^2 < \varepsilon_n, n = 1, 2, \ldots, N$, where $\mathbf{U}_n^{pre}$ is $\mathbf{U}_n$ in the previous iteration. **then**
    break;
  end if
end for

---

the Independent Color Space (ICS), and the Discriminating Color Space (DCS). Specifically, the UCS applies PCA to decorrelate the R, G, and B component images. The ICS and DCS further enhance the discriminating power for the subsequent recognition method by means of Independent Component Analysis (ICA [23]) and LDA, respectively. The experimental results showed that ICS obtains the best color space because its components are not only uncorrelated but also independent.

Many papers have reported that the discriminant analysis methods on facial images can enhance subsequent recognition method [3] [5]. Color Image Discriminant model (CID) [24], borrowing the idea of LDA, aims to seek an optimal color space and an effective recognition method of color images in a unified framework. Tensor Discriminant Color Space (TDCS) [25] model, borrowed the idea of DATER [13], seeks two discriminant projection matrices $\mathbf{U}_1$, $\mathbf{U}_2$ corresponding to the facial spatial information and one color space transformation matrix $\mathbf{U}_3$ corresponding to the color space. Actually, TDCS uses LDA transformation on both facial spatial information and color space information. They [26] also used elastic net to propose Sparse Tensor Discriminant Color Space (STDCS).

For color space information, however, ICA transformation is better than LDA transformation [22]. Motivated by the insights, we explore a Fusion Tensor Color Space (FTCS) model which applies discriminant analysis on the facial spatial information and applies ICA on the color space information.

A color facial image is naturally represented by a 3rd-order tensor, where mode-1 and mode-2 of a tensor are facial spatial information and mode-3 of tensor is the color space information. For instance, a RGB image with size $I_1 \times I_2$ is represented as a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, where $I_3 = 3$. The mode-3 of $\mathcal{A}$ is the color variable in the RGB color space which has 3 components corresponding to $\mathbf{R}$, $\mathbf{G}$ and $\mathbf{B}$ in RGB space. FTCS uses LDA on the first two modes and ICA on the third mode.

Assuming C is the number of individuals, $\mathcal{X}_i^c$ is the ith color facial image of the cth individual, and M_c is the number of color facial images of the cth individual, where $M = M_1 + M_2 + \ldots + M_C$. the FTCS algorithm seeks two discriminant projection matrices $\mathbf{U}_1 \in \mathbb{R}^{I_1 \times L_1}$, $\mathbf{U}_2 \in \mathbb{R}^{I_2 \times L_2}$ and a color space transformation matrix $\mathbf{U}_3 \in \mathbb{R}^{I_3 \times L_3}$ (usually $L_1 < I_1$, $L_2 < I_2$ and $L_3 \leq I_3$) for transformation

$$\mathcal{Y}_i^c = \mathcal{X}_i^c \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \times_3 \mathbf{U}_3^T,$$
$$i = 1, 2, \ldots, M_c, \quad c = 1, 2, \ldots, C. \tag{13}$$

where $\mathbf{U}_1$ and $\mathbf{U}_2$ are obtained by using discriminant analysis and $\mathbf{U}_3$ is obtained by using ICA.

The mean image of the c-th individual and the mean image of all individuals are defined by:

$$\overline{\mathcal{X}}^c = \frac{1}{M_c}\sum_{i=1}^{M_c}\mathcal{X}_i^c \qquad \text{and} \qquad \overline{\mathcal{X}} = \frac{1}{C}\sum_{c=1}^{C}\overline{\mathcal{X}}^c \tag{14}$$

The between-class scatter and within-class of color images are defined as:

$$\Psi_b(\mathcal{X}) = \sum_{c=1}^{C}\|\overline{\mathcal{X}}^c - \overline{\mathcal{X}}\|_F^2 \tag{15}$$

and

$$\Psi_w(\mathcal{X}) = \sum_{c=1}^{C}\sum_{i=1}^{M_c}\|\mathcal{X}_i^c - \overline{\mathcal{X}}^c\|_F^2. \tag{16}$$

We can define mode-n between-class scatter matrix $\mathbf{S}_b^{(n)}$ and mode-n within-class scatter matrix $\mathbf{S}_w^{(n)}$ as:

$$\mathbf{S}_b^{(n)} = \sum_{c=1}^{C}\left(\overline{\mathbf{X}}_{(n)}^c - \overline{\mathbf{X}}_{(n)}\right)\tilde{\mathbf{U}}_n\tilde{\mathbf{U}}_n^T\left(\overline{\mathbf{X}}_{(n)}^c - \overline{\mathbf{X}}_{(n)}\right)^T \tag{17}$$

and

$$\mathbf{S}_w^{(n)} = \sum_{c=1}^{C}\sum_{i=1}^{M_c}\left(\mathbf{X}_{i(n)}^c - \overline{\mathbf{X}}_{(n)}^c\right)\tilde{\mathbf{U}}_n\tilde{\mathbf{U}}_n^T\left(\mathbf{X}_{i(n)}^c - \overline{\mathbf{X}}_{(n)}^c\right)^T, \tag{18}$$

where $\tilde{\mathbf{U}}_n = \mathbf{U}_N \otimes \ldots \otimes \mathbf{U}_{n+1} \otimes \mathbf{U}_{n-1} \otimes \ldots \otimes \mathbf{U}_1$, $n = 1, 2, 3$.

Then, the between-class scatter of the projected tensors $\Psi_b(\mathcal{Y})$ and the within-class scatter of the projected tensors $\Psi_w(\mathcal{Y})$ can be rewritten as follows:

$$\Psi_b(\mathcal{Y}) = tr\left(\mathbf{U}_n^T\mathbf{S}_b^{(n)}\mathbf{U}_n\right) \tag{19}$$

and

$$\Psi_w(\mathcal{Y}) = tr\left(\mathbf{U}_n^T\mathbf{S}_w^{(n)}\mathbf{U}_n\right). \tag{20}$$

So, given $\mathbf{U}_2$ and $\mathbf{U}_3$ (or $\mathbf{U}_1,\mathbf{U}_3$), $\mathbf{U}_1$ (or $\mathbf{U}_2$) can be obtained by the following discriminant analysis:

$$\max \frac{tr\left(\mathbf{U}_n^T\mathbf{S}_b^{(n)}\mathbf{U}_n\right)}{tr\left(\mathbf{U}_n^T\mathbf{S}_w^{(n)}\mathbf{U}_n\right)} \qquad n = 1, 2 \tag{21}$$

According to Rayleigh quotient, Eq. (21) is maximized if and only if the matrix $\mathbf{U}_n$ consists of L_n generalized eigenvectors, which corresponds to the largest L_n generalized eigenvalues of the matrix pencil $(\mathbf{S}_b^{(n)}, \mathbf{S}_w^{(n)})$, which satisfies:

$$\mathbf{S}_b^{(n)}\mathbf{u} = \lambda\mathbf{S}_w^{(n)}\mathbf{u} \qquad n = 1, 2 \tag{22}$$

Since $\mathbf{S}_b^{(n)}$ and $\mathbf{S}_w^{(n)}$ are dependent on $\mathbf{U}_1, \ldots, \mathbf{U}_{n-1}, \mathbf{U}_{n+1}, \ldots, \mathbf{U}_N$, we can see that the optimization of $\mathbf{U}_n$ depends on the projections of other modes.

In order to obtain $\mathbf{U}_3$, we use ICA [1]. to decorrelate the RGB color space. we use $\mathbf{U}_1$ and $\mathbf{U}_2$, which are obtained through the above discriminant analysis, to transform:

$$\begin{aligned}\mathcal{Y}_i^{c(3)} &= \mathcal{X}_i^c \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T, \\ i &= 1, 2, \ldots, M_c, \quad c = 1, 2, \ldots, C.\end{aligned} \tag{23}$$

---

[1] For ICA operations, we used Hyvarinen's fixed-point algorithm http://www.cis.hut.fi/projects/ica/fastica/
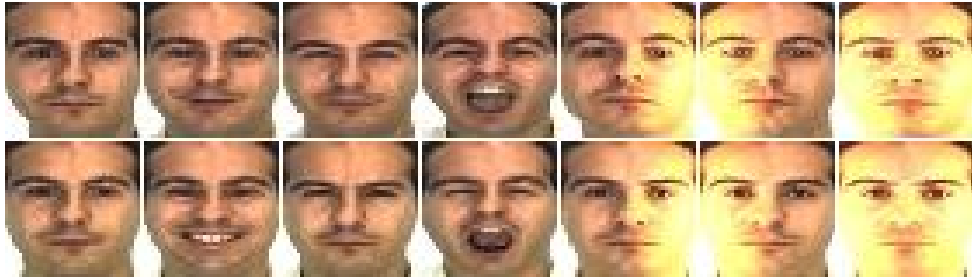
Figure 2. Sample images of one individual from the AR database.

where $\mathcal{Y}_i^{c(3)} \in \mathbb{R}^{L_1 \times L_2 \times I_3}$. M 3rd-order tensor $\mathcal{Y}_i^{c(3)}$ are concatenated to a 4th-order tensor $\mathcal{F} \in \mathbb{R}^{L_1 \times L_2 \times I_3 \times M}$. The mode-3 unfolding matrix $\mathbf{F}_{(3)}$ is a $3 \times K$ matrix, where $K = L_1 \times L_2 \times M$ and the three rows of $\mathbf{F}_{(3)}$ corresponding to the three components in RGB space, respectively.

The color space transformation matrix $\mathbf{U}_3$ may be derived using ICA on $\mathbf{F}_{(3)}$. The ICA of $\mathbf{F}_{(3)}$ factorizes the covariance matrix Σ_F into the following form:

$$\Sigma_F = \mathbf{U}_3^{-1} \nabla \mathbf{U}_3^{-T} \tag{24}$$

where $\nabla \in \mathbb{R}^{3 \times 3}$ is diagonal real positive and $\mathbf{U}_3$ transforms RGB color space to a new color space whose three components are independent or the most independent three component possible. The $\mathbf{U}_3$ in Eq. (24) may be derived using Comon's ICA algorithm by calculating mutual information and high-order statistics. As a result, an iterative procedure can be constructed to obtain $\mathbf{U}_1$, $\mathbf{U}_2$ and $\mathbf{U}_3$.

Results

Experiments and results on the AR database

The AR database contains over 4,000 color facial images of 126 people. Each individual participated in two photo sessions. In both sessions, the pictures were taken under identical requirements and conditions. In our experiments, we selected 100 people, where 14 images of each individual are selected and occluded face images are excluded. These facial images have been cropped [27] and can be downloaded from the AR face database official web[2]. All images are cropped and resized to 32×32 pixels. The sample images for one individual of the AR database are shown in Fig. 2, where the images on the top row are from the first session as the training set, and the images on the bottom row are from the second session as the testing set.

In this experiment, we trained FTCS, TDCS[3] and CID. The convergence threshold ϵ was set as 0.1 and $\mathbf{x}_1$ was initialized as $[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}]^T$. In this case, we got three color space transformation matrices:

$$\mathbf{U}^{CID} = \begin{bmatrix} 0.4126 & -0.2107 & -0.5558 \\ -0.0261 & -0.4683 & 1.0536 \\ 1.0000 & 0.9739 & -0.5524 \end{bmatrix}, \tag{25}$$

$$\mathbf{U}_3^{TDCS} = \begin{bmatrix} 0.1267 & -0.2084 & 0.3358 \\ -0.2128 & -0.4168 & -0.7897 \\ 0.9689 & 0.8848 & 0.5134 \end{bmatrix} \tag{26}$$

---

[2] http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html

[3] Although STDCS [26] is better than TDCS, we still did not compare FTCS to STDCS. Because the motivation of the paper is to implement different transformations on different modes of a tensor.
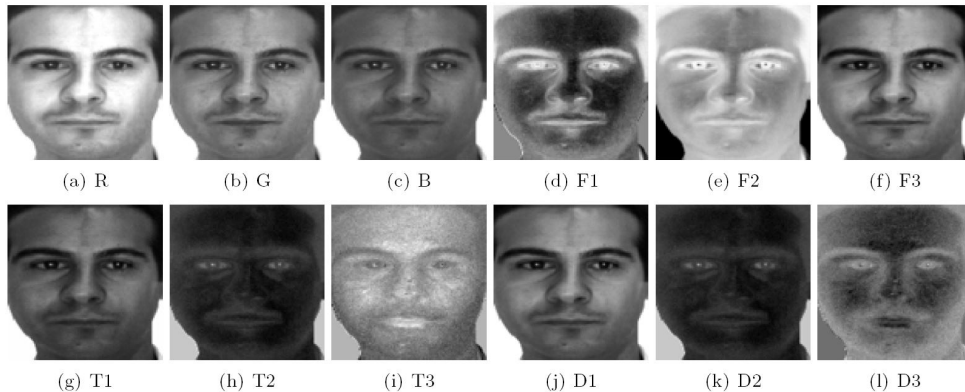
(a) R (b) G (c) B (d) F1 (e) F2 (f) F3

(g) T1 (h) T2 (i) T3 (j) D1 (k) D2 (l) D3

Figure 3. Illustration of R, G, and B color components and the various components generated by CID, TDCS and FTCS on the AR face database.

and

$$\mathbf{U}_3^{FTCS} = \begin{bmatrix} -0.2220 & -0.1853 & 0.6189 \\ -0.6473 & 0.0507 & 0.2239 \\ 0.7292 & -0.9814 & 0.7529 \end{bmatrix} \tag{27}$$

Using these three matrices, we obtained three color components $\mathbf{D}^1$, $\mathbf{D}^2$, $\mathbf{D}^3$ of CID; three color components $\mathbf{T}^1$, $\mathbf{T}^2$, $\mathbf{T}^3$ of TDCS and three color components $\mathbf{F}^1$, $\mathbf{F}^2$, $\mathbf{F}^3$ of FTCS (see Fig. 3).

Meanwhile, we carried out LDA and 2D-LDA on corresponding gray images. In LDA and CID, only 99 discriminant projection basis vectors were extracted. For 2D-LDA, TDCS and FTCS, the spatial dimensions of the two modes are both reduced to 10. The score matrices were generated by Manhattan distance and Euclidean distance, respectively. The ROC curves of the five methods are shown in Fig. 4. The results indicate that the performance of FTCS with Manhattan distance obtains the best performance. However, the space between two curves of FTCS is narrower than the space between two curves of TDCS. This shows that FTCS is more robust to the type of distance used and results of both Manhattan and Euclidean distance produces closer results than those of TDCS.

In the five algorithms, LDA and CID are used on vectorized face images. Overall, their performances are poorer than the other three algorithms based on tensorized face images. This shows that the facial spatial structure information is important to the face recognition. Specially, when the false accept rate is less than 0.03, the performance of 2D-LDA outperforms that of TDCS with the color information. In the case, the color information fails to work for face recognition. This is due to the fact that the color space information is transformed by LDA, which is not an optimal transformation for color space information in comparison to ICA. Whereas, FTCS uses ICA on the color space information. As a results, FTCS obtains the best performance.

Table 1. Verification rate (in percent) comparison of the five methods, respectively, when the FAR is 0.1

	FTCS	TDCS	CID	2D-LDA	LDA
Manhattan	71.97	48.10	42.99	54.82	30.29
Euclidean	70.73	62.63	52.70	53.14	33.50

Table 1 list the verification rates of the five methods with 0.1 FAR. In both cases of Manhattan distance and Euclidean distance, FTCS gets the best verification rate among the five methods. For Man-
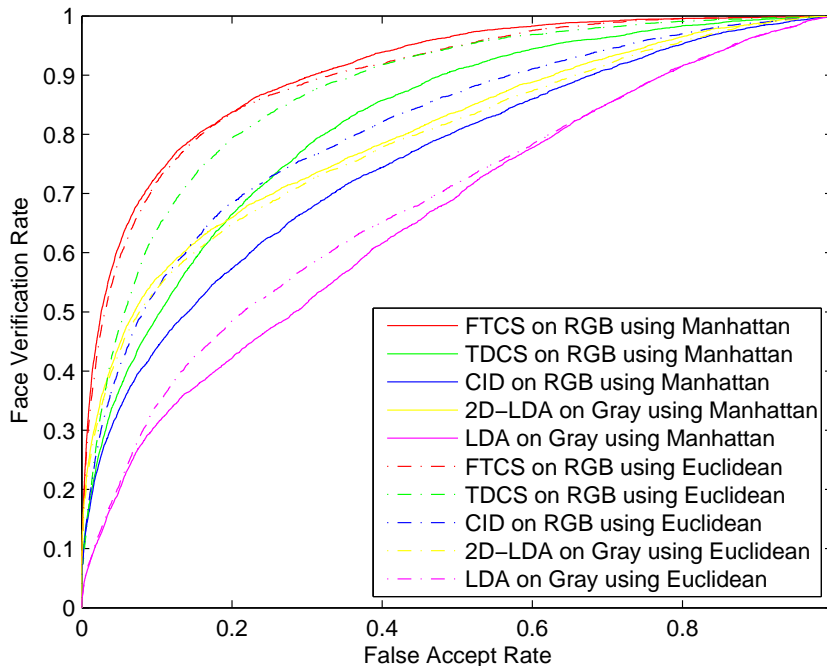
Figure 4. ROC curves of FTCS, TDCS, CID, 2D-LDA and LDA on the AR face database.

hattan distance, 2D-LDA without the color information is better than TDCS with the color information. Whereas, FTCS outperforms 2D-LDA. This also shows that ICA is better than LDA to transform the color space information.

Experiments and results on the LFW face database

LFW database[4] is designed for studying the problem of unconstrained face recognition. It contains more than 13,000 images of faces collected from the web. In our experiments, we chose 1,251 images from 86 people each with 11-20 images. Each image was manually cropped and resized to 32×32 pixels. Sample images of one individual from the LFW database are shown in Fig. 5. From Fig. 5, we can see that the skin color of the same person are different due to various cameras. In this experiment, we randomly selected $\lfloor p/2 \rfloor$ images of each person (the person has p images) as the training set and the remaining images as the testing set. The CID, TDCS and FTCS models were trained and three color space transformation matrices were obtained as follow:

$$\mathbf{U}^{CID} = \begin{bmatrix} 1.0000 & -1.0089 & -0.3600 \\ 0.3166 & 0.2805 & 0.9981 \\ -0.2236 & 0.9078 & -0.6621 \end{bmatrix}, \tag{28}$$

$$\mathbf{U}_3^{TDCS} = \begin{bmatrix} -0.9103 & 0.6923 & 0.1885 \\ 0.4085 & -0.7043 & -0.7688 \\ -0.0673 & -0.1571 & 0.6111 \end{bmatrix} \tag{29}$$

---

[4]http://vis-www.cs.umass.edu/lfw/

Figure 5. Sample images of one individual from the LFW database.

and

$$\mathbf{U}_3^{FTCS} = \begin{bmatrix} -0.3602 & 0.8170 & -0.9108 \\ 0.9032 & -0.5766 & 0.1319 \\ 0.2336 & -0.0017 & -0.3912 \end{bmatrix} \qquad (30)$$

These components are illustrated in Fig. 6.

These three matrices are not the same as Eq. (25), Eq. (26) and Eq. (27) due to the different training sets. Using these three matrices, we got three color components $\mathbf{D}^1$, $\mathbf{D}^2$, $\mathbf{D}^3$ of CID; three color components $\mathbf{T}^1$, $\mathbf{T}^2$, $\mathbf{T}^3$ of TDCS and three color components $\mathbf{F}^1$, $\mathbf{F}^2$, $\mathbf{F}^3$ of FTCS.

In the same way, FTCS, TDCS, CID, 2D-LDA and LDA are conducted on the LFW face database. The results indicate that FTCS has the best performance compared to the other four algorithms from Fig. 7. We can also see that the curves of 2D-LDA and LDA for gray images are almost consistent. For the unconstrained face recognition, to use only the spatial structure information can not improve the performance. Whereas, the color space information further improves the performance. One is also drawn from Fig. 7 that the color information plays more important role than the spatial structure information for the unconstrained face recognition.

Discussion

Recently, tensor subspace transformation is a highly mentioned topic, because many real objects can be represented by tensors. For different objects, the semantic meaning of tensorial modes are different. Even when the objects are the same, each mode of tensors may express a different semantic meaning. To the best of our knowledge, there aren't any existing tensor subspace transformation algorithms which implements different transformation strategies on different mode of tensors according to their semantic meaning. In this paper, we propose the fusion tensor subspace analysis framework, which shows an novel idea that different transformation strategies can applied on different modes of tensor. Under the framework, we propose FTCS for face recognition. The experimental results show the performances of the proposed algorithm is better than existing tensor subspace transformation algorithms. FTCS is only
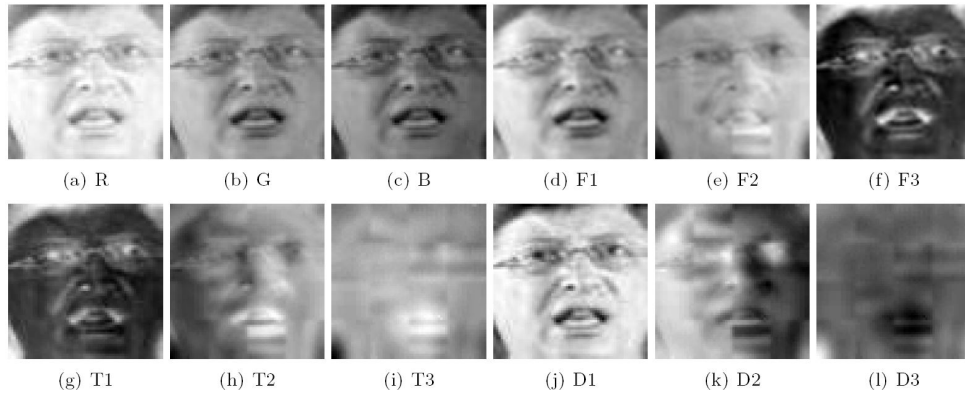
Figure 6. Illustration of R, G, and B color components and the various components generated by CID, TDCS and FTCS on the LFW face database.
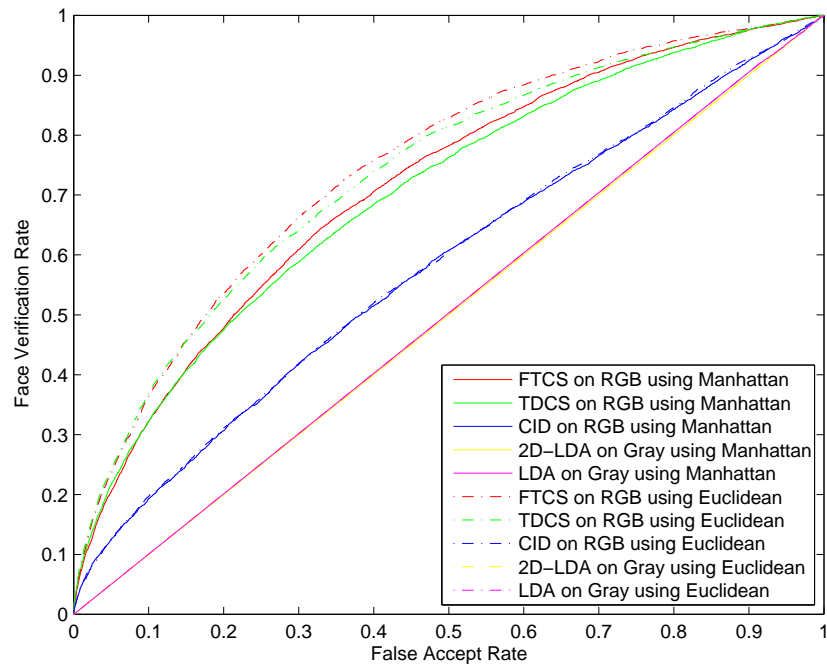

Figure 7. ROC curves of FTCS, TDCS, CID, 2D-LDA and LDA on the LFW face database.

an example of fusion tensor subspace transformation framework. Under the framework, many algorithms can be developed for action recognition, micro-expression recognition, EEG recognition and so on.

References

1. Wang X, Tang X (2004) A unified framework for subspace face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 26: 1222-1228.

2. Turk M, Pentland A (1991) Eigenfaces for recognition. Journal of Cognitive Neuroscience 3: 71–86.

3. Belhumeur PN, Hespanha JP, Kriegman DJ (1997) Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence 19: 711–720.

4. He XF, Niyogi P (2004) Locality preserving projections. Advances In Neural Information Processing Systems 16 16: 153–160.

5. Yu WW, Teng XL, Liu CQ (2006) Face recognition using discriminant locality preserving projections. Image and Vision Computing 24: 239–248.

6. Lu H, Plataniotis K, Venetsanopoulos A (2011) A survey of multilinear subspace learning for tensor data. Pattern Recognition 44: 1540-1551.

7. Tao D, Li X, Wu X, Hu W, Maybank SJ (2007) Supervised tensor learning. Knowledge and Information Systems 13: 1–42.

8. Zhang L, Zhang L, Tao D, Huang X (2011) A multifeature tensor for remote-sensing target recognition. Geoscience and Remote Sensing Letters, IEEE 8: 374–378.

9. Tao D, Li X, Wu X, Maybank S (2008) Tensor rank one discriminant analysis - convergent method for discriminative multilinear subspace selection. Neurocomputing 71: 1866–1882.

10. Tao D, Song M, Li X, Shen J, Sun J, et al. (2008) Bayesian tensor approach for 3-d face modeling. Circuits and Systems for Video Technology, IEEE Transactions on 18: 1397–1410.

11. Sun J, Tao D, Papadimitriou S, Yu PS, Faloutsos C (2008) Incremental tensor analysis: Theory and applications. ACM Transactions on Knowledge Discovery from Data (TKDD) 2: 11.

12. Lu HP, Konstantinos NP, Venetsanopoulos AN (2008) MPCA: Multilinear principal component analysis of tensor objects. IEEE Transactions on Neural Networks 19: 18-39.

13. Yan S, Xu D, Yang Q, Zhang L, Tang X, et al. (2005) discriminant analysis with tensor representation. CVPR 2005 .

14. Tao D, Li X, Wu X, Maybank S (2007) General tensor discriminant analysis and gabor features for gait recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence : 1700–1715.

15. He X, Cai D, Niyogi P (2005) Tensor subspace analysis. In: In Advances in Neural Information Processing Systems 18 (NIPS). MIT Press.

16. Wang SJ, Zhou CG, Zhang N, Peng XJ, Chen YH, et al. (2011) Face recognition using second-order discriminant tensor subspace analysis. Neurocomputing 74: 2142–2156.

17. Song F, Zhang D, Mei D, Guo Z (2007) A multiple maximum scatter difference discriminant criterion for facial feature extraction. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 37: 1599–1606.

18. Kolda TG, Bader BW (2009) Tensor decompositions and applications. Siam Review 51: 455–500.

19. Yang J, Zhang D, Frangi AF, Yang JY (2004) Two-dimensional PCA: A new approach to appearance-based face representation and recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 26: 131–137.

20. Ye J, Janardan R, Li Q (2004) Two-dimensional linear discriminant analysis. In: Neural Information Processing Systems. pp. 1569-1576.

21. Fukunaga K (1990) Introduction to statistical pattern recognition. Academic Pr.

22. Liu C (2008) Learning the uncorrelated, independent, and discriminating color spaces for face recognition. IEEE Transactions on Information Forensics and Security 3: 213-222.

23. Comon P (1994) Independent component analysis, a new concept? Signal processing 36: 287–314.

24. Yang J, Liu C (2008) Color image discriminant models and algorithms for face recognition. IEEE Transactions on Neural Networks 19: 2088-2098.

25. Wang SJ, Yang J, Zhang N, Zhou CG (2011) Tensor discriminant color space for face recognition. IEEE Transactions on Image Processing 20: 2490–2501.

26. Wang SJ, Yang J, Sun MF, Peng XJ, Sun MM, et al. (2012) Sparse tensor discriminant color space for face verification. IEEE Transactions on Neural Networks and Learning Systems 23: 876 -888.

27. Martinez AM, Kak AC (2001) PCA versus LDA. IEEE Transactions on Pattern Analysis and Machine Intelligence 23: 228–233.