

Tensor Discriminant Color Space for Face Recognition

Su-Jing Wang, Jian Yang, *Member, IEEE*, Na Zhang, and Chun-Guang Zhou*

Abstract

Recent research efforts reveal that color may provide useful information for face recognition. For different visual tasks the choice of a color space is generally different. How can a color space be sought for the specific face recognition problem? To address this problem, this paper represents a color image as a 3rd-order tensor and presents the tensor discriminant color space (TDCS) model. The model can keep the underlying spatial structure of color images. With the definition of n-mode between-class scatter matrices and within-class scatter matrices, TDCS constructs an iterative procedure to obtain one color space transformation matrix and two discriminant projection matrices by maximizing the ratio of these two scatter matrices. The experiments are conducted on two color face databases, AR and Georgia Tech face databases, and the results show that both the performance and the efficiency of the proposed method are better than those of the state-of-the-art color image discriminant (CID) model, which involve one color space transformation matrix and one discriminant projection matrix, specifically in a complicated face database with various pose variations.

Index Terms

Face recognition, Color images, Discriminant information, Tensor subspace.

This work was supported by (1) the National Natural Science Foundation of China under Grant No. 60873146, 60973092, 60903097, (2) project of science and technology innovation platform of computing and software science (985 engineering), (3) the Key Laboratory for Symbol Computation and Knowledge Engineering of the National Education Ministry of China. This work was also partially supported by the Program for New Century Excellent Talents in University of China and the National Science Foundation of China under Grants No. 60973098.

S.J Wang, N Zhang and C.G Zhou are with the College of Computer Science and Technology, Jilin University, Changchun 130012, China. (e-mail:wangsj08@mails.jlu.edu.cn; nazhang08@mails.jlu.edu.cn; cgzhou@jlu.edu.cn).

J Yang is now with the School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: csjyang@njit.edu; csjyang@mail.njust.edu.cn).

I. INTRODUCTION

As one of the hottest topics in the field of pattern recognition and artificial intelligence, face recognition has been widely used in public securities, such as crime and terrorist detection, *etc.* There are various subspace transformation methods for recognizing faces. Principal Component Analysis (PCA)[1] is a widely used linear subspace transformation method maximizing the variance of the transformed features in the projective subspace. Linear Discriminant Analysis (LDA)[2] encodes discriminant information by maximizing the between-class covariance, while minimizing the within-class covariance in the projective subspace. Moreover, in order to keep spatial structure information of a gray image, Yang *et al.*[3] proposed an algorithm called two-dimensional principal component analysis (2D-PCA) for face recognition, in which the image covariance (scatter) matrix is directly computed from the image matrix representation. Li and Yuan[4] extended this idea using discriminant information and presented 2D-LDA, which constructs the image between-class covariance matrix and the image within-class covariance matrix. All these methods are methods used to deal with gray face images rather than color face images, because some past researches suggested that color appears to confer no significant face recognition advantages beyond the gray[5].

Recent research efforts[6], [7], [8], [9] revealed that color may provide useful information for face recognition. The experimental results in [6] show that the PCA method using color information can improve the recognition rate compared to the same method using only gray information. In [7], a RGB image of size $I_1 \times I_2$ is transformed into a $I \times 3$ matrix, where $I = I_1 \times I_2$. 2D-PCA is applied on all transformed matrices to recognize the color face images. And the results show that its performance can be improved by about 3% compared to the same method applied on the corresponding gray images. On dealing with the low-resolution face images (20×20 pixels or less), Choi *et al.*[8] demonstrated that facial color cue can significantly improve recognition performance compared with gray-based features. In [9], it is also revealed that the face recognition system on various color spaces (such as RGB, PCS and YIQ) is better than on gray images.

The RGB color space, as a basis for other color spaces which are usually defined by linear or nonlinear transformations of the RGB color space, is dominant in computer vision. The nonlinear transformations color spaces, such as the HSV and L*a*b* color spaces, are generally connected with the human vision system, while those determined by the linear transformations, such as the YUV and YIQ color spaces[10], are usually associated with the color of some hardware.

Given a specific problem like face detection or recognition, the choice of a color space is an important

task[11]. For example, Ikeda[12] chose the HSV color space [12] and Hsu *et al.*[13] applied the YCbCr color space[13] to face detection. For face recognition, Jones[14] and Neagoe[15] used PCA to convert a color facial image from the RGB color space into a color component and two color components respectively. Liu[16] capitalized on a hybrid color space RIQ, which combines the R component image of the RGB color space and the chromatic components I and Q of the YIQ color space, to increase the accuracy of face recognition performance due to the complementary characteristics of its component images. Liu [17] applied PCA, ICA and LDA to obtain the uncorrelated color space (UCS), the independent color space (ICS), and the discriminating color space (DCS) for face recognition, respectively. In [18], they took advantage of ICS to improve the face recognition grand challenge (FRGC)[19] performance. Yang and Liu proposed[20][21] color image discriminant (CID) model borrowing the idea of LDA not only to get the discriminating color space but also to get the optimal spatial transformation matrix.

Recently, a number of researchers have attempted to recognize face using tensor. The methods based on tensor can be divided into 2 categories. In one category[22][23], a high order tensor constructs a multilinear structure and models the multiple factors of facial variation (e.g., different user identities, various user postures and facial expressions, varying lights, etc.) using high-order SVD[24], [25], [26]. In the other category[27], [28], [29], the conventional transformation methods (such as PCA, SVD and LPP[30]) are generalized to tensors. They treat a gray image as a 2nd-order tensor, a color image as a 3rd-order tensor.

In this paper, a RGB color facial image of size $I_1 \times I_2$, defined by a function of two spatial variables and one color variable, is naturally represented by a 3rd-order tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times 3}$. We model the tensor discriminant color space (TDCS) by borrowing the idea of DATER[31]. TDCS seeks two discriminant projection matrices \mathbf{U}_1 , \mathbf{U}_2 corresponding to two spatial variables of color images and one color space transformation matrix \mathbf{U}_3 corresponding to one color variable of the color images. Whereas, CID seeks a discriminant projection matrix \mathbf{P} consisting of one or multiple discriminant projection basis vectors for image discrimination and a color space transformation matrix \mathbf{X} consisting of a set of color component combination coefficients for color image representation. In contrast to CID, the advantages of TDCS are: to keep some underlying spatial structure of color images, to avoid losing a part of the discriminant information and to optimize every column vector of the color space transformation matrix simultaneously.

The rest of this paper is organized as follows: in Section II, we give the related definitions to tensor; in Section III, we will briefly review the color image discriminant (CID) model; in Section IV, we will introduce the tensor discriminant color space model (TDCS) and analyze the relations and distinctions between TDCS and CID theoretically; in Section V, the experiments are conducted on two color face

databases AR and Georgia Tech face database, and the results show that the efficiency and performance of TDCS are better than those of CID, especially in the complicated face database with different poses and clustered background like the Georgia Tech face database; finally in Section VI, conclusions are drawn and several issues for the future works are indicated.

II. TENSOR FUNDAMENTALS

A tensor is a multidimensional array. More formally, an N th-order tensor is an element of the tensor product of N vector spaces, each of which has its own coordinate system. It is the higher-order generalization of scalar (zero-order tensor), vector (1st-order tensor), and matrix (2nd-order tensor). In this paper, lowercase italic letters (a, b, \dots) denote scalars, bold lowercase letters ($\mathbf{a}, \mathbf{b}, \dots$) denote vectors, bold uppercase letters ($\mathbf{A}, \mathbf{B}, \dots$) denote matrices, and calligraphic uppercase letters ($\mathcal{A}, \mathcal{B}, \dots$) denote tensors. The formal definition is given below[26]:

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 $1 \leq i_n \leq I_n, n = 1, 2, \dots, N$.*

If we refer to I_n rank in tensor terminology, we generalize the matrix definition and call column vectors of matrices as 1-mode vectors and row vectors of matrices as 2-mode vectors.

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

We can unfold the tensor \mathcal{A} by taking the n -mode vectors as the column vectors of matrix $\mathbf{A}_{(n)} \in \mathbb{R}^{I_n \times (I_1 \dots I_{n-1} I_{n+1} \dots I_N)}$. These tensor unfoldings provide an easy manipulation in tensor algebra and we can reconstruct the tensor by an inverse process of n -mode unfolding, if necessary.

Definition 3. *The n -mode unfolding matrix of \mathcal{A} , denoted by $(\mathcal{A})_{(n)} \in \mathbb{R}^{I_n \times (I_1 \dots I_{n-1} I_{n+1} \dots I_N)}$, contains the element $a_{i_1 \dots i_N}$ at i_n th row and at $[(i_{n+1}-1)I_{n+2} \dots I_N I_1 \dots I_{n-1} + (i_{n+2}-1)I_{n+3} \dots I_N I_1 \dots I_{n-1} + \dots + (i_N-1)I_1 \dots I_{n-1} + (i_1-1)I_2 \dots I_{n-1} + (i_2-1)I_3 \dots I_{n-1} + \dots + i_{n-1}]$ th column.*

Intuitively, the operation shows slicing the tensor along a given direction depending on the n -mode of unfolding, and then putting the slices side by side in a matrix. The pictorial description is given in Fig. 1 for a 3rd-order tensor.

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

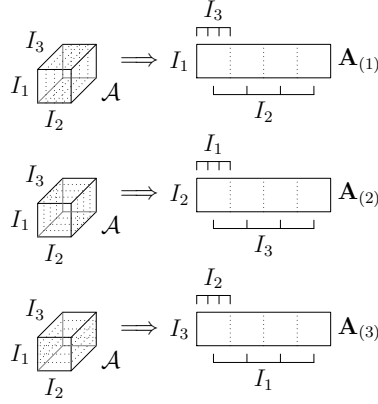


Fig. 1. Unfolding of the $(I_1 \times I_2 \times I_3)$ -tensor \mathcal{A} to the $(I_1 \times I_2 I_3)$ -matrix $\mathbf{A}_{(1)}$, the $(I_2 \times I_3 I_1)$ -matrix $\mathbf{A}_{(2)}$, and the $(I_3 \times I_1 I_2)$ -matrix $\mathbf{A}_{(3)}$ ($I_1 = I_2 = I_3 = 4$).

Definition 4. The n -mode 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:

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

This n -mode product of tensor and matrix can be expressed in terms of unfolding matrices for ease of usage.

$$(\mathcal{A} \times_n \mathbf{U})_{(n)} = \mathbf{U} \cdot \mathbf{A}_{(n)} \quad (2)$$

Given the tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ and the matrices $\mathbf{U} \in \mathbb{R}^{J_n \times I_n}$, $\mathbf{V} \in \mathbb{R}^{J_m \times I_m}$, one has

$$(\mathcal{A} \times_n \mathbf{U}) \times_m \mathbf{V} = (\mathcal{A} \times_m \mathbf{V}) \times_n \mathbf{U} = \mathcal{A} \times_n \mathbf{U} \times_m \mathbf{V} \quad (3)$$

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 $\langle \mathcal{A}, \mathcal{B} \rangle \stackrel{\text{def}}{=} \sum_{i_1} \sum_{i_2} \dots \sum_{i_N} a_{i_1 i_2 \dots i_N} b_{i_1 i_2 \dots i_N}$. 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 $\|\mathcal{A}\|_F \stackrel{\text{def}}{=} \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle}$

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

$$\|\mathcal{A}\|_F = \|(\mathbf{A})_{(n)}\|_F \quad (4)$$

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

$$\mathcal{A} = \mathcal{C} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \dots \times_N \mathbf{U}_N \quad (5)$$

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

$$\mathbf{A}_{(n)} = \mathbf{U}_n \mathbf{C}_{(n)} (\mathbf{U}_N \otimes \dots \otimes \mathbf{U}_{n+1} \otimes \mathbf{U}_{n-1} \otimes \dots \otimes \mathbf{U}_1)^T \quad (6)$$

III. OVERVIEW OF COLOR IMAGES DISCRIMINANT MODEL

The motivation of the CID model[20][21] is to seek a meaningful color space and an effective recognition method in an unified framework. The CID model involves two matrices: the color space transformation matrix \mathbf{X} for the color space and the discriminant projection matrix \mathbf{P} for the image discrimination. CID constructs two iterative procedures to solve the two matrices.

Let \mathbf{A} be a color image of size $I_1 \times I_2$, and let its three color components be \mathbf{R} , \mathbf{G} and \mathbf{B} . Each component is rasterized as I -dimensional vector, where $I = I_1 \times I_2$. Then, a color image \mathbf{A} is expressed as an $I \times 3$ matrix: $\mathbf{A} = [\mathbf{R}, \mathbf{G}, \mathbf{B}] \in \mathbb{R}^{I \times 3}$.

The goal of CID is to find a color space transformation matrix \mathbf{X} to map a color image \mathbf{A} from RGB color space to a new color space, denoted by D -space, in which image representation \mathbf{D} is the best representation of the color image for face recognition.

$$\mathbf{D} = \mathbf{A}\mathbf{X} = [\mathbf{R}, \mathbf{G}, \mathbf{B}]\mathbf{X} \quad (7)$$

In CID, Eq.(7) is rewritten as follows:

$$\mathbf{d}^k = [\mathbf{R}, \mathbf{G}, \mathbf{B}]\mathbf{x}_k, \quad k = 1, 2, 3 \quad (8)$$

where, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3]$ and $\mathbf{D} = [\mathbf{d}^1, \mathbf{d}^2, \mathbf{d}^3]$. \mathbf{d}^k is k th component in the D -space and \mathbf{x}_k consists of optimal coefficients of \mathbf{d}^k .

Let C be the number of individuals, \mathbf{A}_i^c be the i th color face image of c th individual, where $c = 1, 2, \dots, C$, $i = 1, 2, \dots, M_c$, and M_c be the number of color face images of c th individual. The mean image of the c th individual is

$$\overline{\mathbf{A}}^c = \frac{1}{M_c} \sum_{i=1}^{M_c} \mathbf{A}_i^c \quad (9)$$

and the mean image of all individuals is

$$\overline{\mathbf{A}} = \frac{1}{C} \sum_{c=1}^C \overline{\mathbf{A}}^c \quad (10)$$

Given a set of training color face images of individuals, CID builds on the LDA idea to seek the color space transformation matrix \mathbf{X} and the discriminant projection matrix \mathbf{P} .

Similar to LDA, CID defines the between-class scatter matrix $\mathbf{S}_b(\mathbf{x}_k)$ and the within-class scatter matrix $\mathbf{S}_w(\mathbf{x}_k)$ in the D -space as follows:

$$\mathbf{S}_b(\mathbf{x}_k) = \sum_{c=1}^C P_c [(\bar{\mathbf{A}}^c - \bar{\mathbf{A}})\mathbf{x}_k\mathbf{x}_k^T(\bar{\mathbf{A}}^c - \bar{\mathbf{A}})^T] \quad (11)$$

$$\mathbf{S}_w(\mathbf{x}_k) = \sum_{c=1}^C P_c \frac{1}{M_c - 1} \sum_{i=1}^{M_c} [(\mathbf{A}_i^c - \bar{\mathbf{A}}^c)\mathbf{x}_k\mathbf{x}_k^T(\mathbf{A}_i^c - \bar{\mathbf{A}}^c)^T] \quad (12)$$

where $P_c = M_c/M$. The general Fisher criterion in the D -space can be defined as follows:

$$J(\mathbf{P}, \mathbf{x}_k) = \frac{|\mathbf{P}^T \mathbf{S}_b(\mathbf{x}_k) \mathbf{P}|}{|\mathbf{P}^T \mathbf{S}_w(\mathbf{x}_k) \mathbf{P}|} \quad (13)$$

where $|\cdot|$ denotes the determinant operator, and \mathbf{P} is an $I \times d$ transformation matrix. CID defines the general color-space between-class scatter matrix $\mathbf{L}_b(\mathbf{P})$ and the general color-space within-class scatter matrix $\mathbf{L}_w(\mathbf{P})$ as follows:

$$\mathbf{L}_b(\mathbf{P}) = \sum_{c=1}^C P_c [(\bar{\mathbf{A}}^c - \bar{\mathbf{A}})^T \mathbf{P} \mathbf{P}^T (\bar{\mathbf{A}}^c - \bar{\mathbf{A}})] \quad (14)$$

$$\mathbf{L}_w(\mathbf{P}) = \sum_{c=1}^C P_c \frac{1}{M_c - 1} \sum_{i=1}^{M_c} [(\mathbf{A}_i^c - \bar{\mathbf{A}}^c)^T \mathbf{P} \mathbf{P}^T (\mathbf{A}_i^c - \bar{\mathbf{A}}^c)] \quad (15)$$

The following properties are proved in [20]:

$$\text{tr} [\mathbf{P}^T \mathbf{S}_b(\mathbf{x}_k) \mathbf{P}] = \mathbf{x}_k^T \mathbf{L}_b(\mathbf{P}) \mathbf{x}_k \quad (16)$$

$$\text{tr} [\mathbf{P}^T \mathbf{S}_w(\mathbf{x}_k) \mathbf{P}] = \mathbf{x}_k^T \mathbf{L}_w(\mathbf{P}) \mathbf{x}_k = d \quad (17)$$

These coefficient vectors \mathbf{x}_k are required to be $\mathbf{L}_w(\mathbf{P})$ -orthogonal, that is

$$\mathbf{x}_i^T \mathbf{L}_w(\mathbf{P}) \mathbf{x}_j = 0, \quad \forall i \neq j, \quad i, j = 1, 2, 3. \quad (18)$$

First, \mathbf{x}_1 and \mathbf{P} are obtained in Algorithm 1. In Algorithm 1, \mathbf{x}_1 is the generalized eigenvector corresponding to the largest eigenvalue of the matrix pencil $(\mathbf{L}_b(\mathbf{P}), \mathbf{L}_w(\mathbf{P}))$. The remaining two generalized eigenvectors \mathbf{u}_1 and \mathbf{u}_2 are $\mathbf{L}_w(\mathbf{P})$ -orthogonal to \mathbf{x}_1 . And the second combination coefficient vector \mathbf{x}_2 can be computed as follows:

$$\mathbf{x}_2 = (\mathbf{u}_1, \mathbf{u}_2) \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \mathbf{U} \mathbf{y} \quad (19)$$

In order to get \mathbf{x}_2 , there are some small modifications in Step 3. Let $\mathbf{x} = \mathbf{U} \mathbf{y}$, $\mathbf{L}_w(\mathbf{P}) = \mathbf{U}^T \mathbf{L}_w(\mathbf{P}) \mathbf{U}$ and $\mathbf{L}_b(\mathbf{P}) = \mathbf{U}^T \mathbf{L}_b(\mathbf{P}) \mathbf{U}$. Last, \mathbf{z} is the generalized eigenvector corresponding to the smallest eigenvalue of the matrix pencil $(\mathbf{L}_b(\mathbf{P}), \mathbf{L}_w(\mathbf{P}))$. Then, we have $\mathbf{x}_3 = \mathbf{U} \mathbf{z}$.

Algorithm 1 CID Algorithm

Step 1: $i = 0$, and initialize $\mathbf{x} = \mathbf{x}^{[0]}$

Step 2: Compute $\mathbf{S}_b(\mathbf{x})$, $\mathbf{S}_w(\mathbf{x})$ and calculate their generalized eigenvectors $\lambda_1, \lambda_2, \dots, \lambda_d$ corresponding to the d largest eigenvalues. Let $\mathbf{P} = \mathbf{P}^{[i+1]} = [\lambda_1, \lambda_2, \dots, \lambda_d]$.

Step 3: Compute $\mathbf{L}_b(\mathbf{P})$, $\mathbf{L}_w(\mathbf{P})$ and calculate their generalized eigenvector $\mathbf{x}_1^{[i+1]}$ corresponding to the largest eigenvalue.

if $|J(\mathbf{P}^{[i+1]}, \mathbf{x}_1^{[i+1]}) - J(\mathbf{P}^{[i]}, \mathbf{x}^{[i]})| < \epsilon$ **then**

 the iteration terminates and let $\mathbf{P}^* = \mathbf{P}^{[i+1]}$ and $\mathbf{x}^* = \mathbf{x}^{[i+1]}$

else

$\mathbf{x} = \mathbf{x}^{[i+1]}$ and go to Step 2

end if

IV. TENSOR DISCRIMINANT COLOR SPACE MODEL

A. The Algorithm

In this section, we discuss how to use tensor to model the discriminant color space. A color image is naturally represented by a 3rd-order tensor. The 1-mode of tensor is the height of image, the 2-mode of tensor is the width of image and the 3-mode of tensor is the color space of image. For instance, a RGB color image of size $I_1 \times I_2$ is expressed as a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, where $I_3 = 3$. The 3-mode of \mathcal{A} is the color variable in the RGB color space. It has 3 components, which corresponds to \mathbf{R} , \mathbf{G} and \mathbf{B} in RGB space.

Let C be the number of individuals, \mathcal{A}_i^c be the i th color face image of c th individual, let M_c be the number of color face images of c th individual. We use the TDCS algorithm to seek 2 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

$$\mathcal{D}_i^c = \mathcal{A}_i^c \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \times_3 \mathbf{U}_3^T, \quad i = 1, 2, \dots, M_c, \quad c = 1, 2, \dots, C. \quad (20)$$

which ensures that the projected tensors of the same individual's images are distributed as close as possible, while the projected tensors of the images from different individuals are distributed as far away as possible.

The mean image of the c th individual is

$$\bar{\mathcal{A}}^c = \frac{1}{M_c} \sum_{i=1}^{M_c} \mathcal{A}_i^c \quad (21)$$

and the mean image of all individuals is

$$\bar{\mathcal{A}} = \frac{1}{C} \sum_{c=1}^C \bar{\mathcal{A}}^c \quad (22)$$

The between-class scatter of color images is defined as:

$$\Psi_b(\mathcal{A}) = \sum_{c=1}^C \|\bar{\mathcal{A}}^c - \bar{\mathcal{A}}\|_F^2 \quad (23)$$

and within-class scatter of color images is defined as:

$$\Psi_w(\mathcal{A}) = \sum_{c=1}^C \sum_{i=1}^{M_c} \|\mathcal{A}_i^c - \bar{\mathcal{A}}^c\|_F^2 \quad (24)$$

A reasonable idea is to maximize the between-class scatter of projected tensors $\Psi_b(\mathcal{D})$ and to minimize the within-class scatter of projected tensors $\Psi_w(\mathcal{D})$. Then TDCS criterion can be defined as follows:

$$\max J(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3) = \frac{\Psi_b(\mathcal{D})}{\Psi_w(\mathcal{D})} \quad (25)$$

Here, three matrices \mathbf{U}_n need to be simultaneously optimized in maximizing the criterion function J . We can define n -mode between-class scatter matrix $\mathbf{S}_b^{(n)}$ and n -mode within-class scatter matrix $\mathbf{S}_w^{(n)}$ as following,

$$\mathbf{S}_b^{(n)} = \sum_{c=1}^C (\bar{\mathbf{A}}_{(n)}^c - \bar{\mathbf{A}}_{(n)}) \tilde{\mathbf{U}}_n \tilde{\mathbf{U}}_n^T (\bar{\mathbf{A}}_{(n)}^c - \bar{\mathbf{A}}_{(n)})^T \quad (26)$$

and

$$\mathbf{S}_w^{(n)} = \sum_{c=1}^C \sum_{i=1}^{M_c} (\mathbf{A}_{i(n)}^c - \bar{\mathbf{A}}_{(n)}^c) \tilde{\mathbf{U}}_n \tilde{\mathbf{U}}_n^T (\mathbf{A}_{i(n)}^c - \bar{\mathbf{A}}_{(n)}^c)^T \quad (27)$$

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

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

$$\begin{aligned} \Psi_b(\mathcal{D}) &= \sum_{c=1}^C \|\bar{\mathcal{D}}^c - \bar{\mathcal{D}}\|_F^2 \\ &= \sum_{c=1}^C \|(\bar{\mathcal{A}}^c - \bar{\mathcal{A}}) \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \times_3 \mathbf{U}_3^T\|_F^2 \end{aligned} \quad (28)$$

and

$$\begin{aligned} \Psi_w(\mathcal{D}) &= \sum_{c=1}^C \sum_{i=1}^{M_c} \|\mathcal{D}_i^c - \bar{\mathcal{D}}^c\|_F^2 \\ &= \sum_{c=1}^C \sum_{i=1}^{M_c} \|(\mathcal{A}_i^c - \bar{\mathcal{A}}^c) \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \times_3 \mathbf{U}_3^T\|_F^2 \end{aligned} \quad (29)$$

According to the definition of the Frobenius norm for a tensor and that for a matrix, $\|\mathcal{A}\|_F = \|\mathbf{A}_{(n)}\|_F$, and Eq. (6), $\Psi_b(\mathcal{D})$ and $\Psi_w(\mathcal{D})$ can be expressed using the equivalent matrix representation by n -mode unfolding as follows:

$$\begin{aligned}\Psi_b(\mathcal{D}) &= \sum_{c=1}^C \|\bar{\mathbf{D}}_{(n)}^c - \bar{\mathbf{D}}_{(n)}\|_F^2 \\ &= \sum_{c=1}^C \|\mathbf{U}_n^T (\bar{\mathbf{A}}_{(n)}^c - \bar{\mathbf{A}}_{(n)}) \tilde{\mathbf{U}}_n\|_F^2\end{aligned}\quad (30)$$

and

$$\begin{aligned}\Psi_w(\mathcal{D}) &= \sum_{c=1}^C \sum_{i=1}^{M_c} \|\mathbf{D}_{i(n)}^c - \bar{\mathbf{D}}_{(n)}\|_F^2 \\ &= \sum_{c=1}^C \sum_{i=1}^{M_c} \|\mathbf{U}_n^T (\mathbf{A}_{i(n)}^c - \bar{\mathbf{A}}_{(n)}) \tilde{\mathbf{U}}_n\|_F^2\end{aligned}\quad (31)$$

Since $\|\mathbf{A}\|_F = \text{tr}(\mathbf{A}\mathbf{A}^T)$, $\Psi_b(\mathcal{D})$ and $\Psi_w(\mathcal{D})$ can be written in terms of the n -mode between-class scatter matrix and n -mode within-class scatter matrix

$$\begin{aligned}\Psi_b(\mathcal{D}) &= \sum_{c=1}^C \|\mathbf{U}_n^T (\bar{\mathbf{A}}_{(n)}^c - \bar{\mathbf{A}}_{(n)}) \tilde{\mathbf{U}}_n\|_F^2 \\ &= \sum_{c=1}^C \text{tr} \left(\mathbf{U}_n^T (\bar{\mathbf{A}}_{(n)}^c - \bar{\mathbf{A}}_{(n)}) \tilde{\mathbf{U}}_n \tilde{\mathbf{U}}_n^T (\bar{\mathbf{A}}_{(n)}^c - \bar{\mathbf{A}}_{(n)})^T \mathbf{U}_n \right) \\ &= \text{tr} \left\{ \mathbf{U}_n^T \left[\sum_{c=1}^C (\bar{\mathbf{A}}_{(n)}^c - \bar{\mathbf{A}}_{(n)}) \tilde{\mathbf{U}}_n \tilde{\mathbf{U}}_n^T (\bar{\mathbf{A}}_{(n)}^c - \bar{\mathbf{A}}_{(n)})^T \right] \mathbf{U}_n \right\} \\ &= \text{tr} \left(\mathbf{U}_n^T \mathbf{S}_b^{(n)} \mathbf{U}_n \right)\end{aligned}\quad (32)$$

and

$$\begin{aligned}\Psi_w(\mathcal{D}) &= \sum_{c=1}^C \sum_{i=1}^{M_c} \|\mathbf{U}_n^T (\mathbf{A}_{i(n)}^c - \bar{\mathbf{A}}_{(n)}) \tilde{\mathbf{U}}_n\|_F^2 \\ &= \sum_{c=1}^C \sum_{i=1}^{M_c} \text{tr} \left(\mathbf{U}_n^T (\mathbf{A}_{i(n)}^c - \bar{\mathbf{A}}_{(n)}) \tilde{\mathbf{U}}_n \tilde{\mathbf{U}}_n^T (\mathbf{A}_{i(n)}^c - \bar{\mathbf{A}}_{(n)})^T \mathbf{U}_n \right) \\ &= \text{tr} \left\{ \mathbf{U}_n^T \left[\sum_{c=1}^C \sum_{i=1}^{M_c} (\mathbf{A}_{i(n)}^c - \bar{\mathbf{A}}_{(n)}) \tilde{\mathbf{U}}_n \tilde{\mathbf{U}}_n^T (\mathbf{A}_{i(n)}^c - \bar{\mathbf{A}}_{(n)})^T \right] \mathbf{U}_n \right\} \\ &= \text{tr} \left(\mathbf{U}_n^T \mathbf{S}_w^{(n)} \mathbf{U}_n \right)\end{aligned}\quad (33)$$

So, given all the other projection matrices $\mathbf{U}_1, \dots, \mathbf{U}_{n-1}, \mathbf{U}_{n+1}, \dots, \mathbf{U}_N$, the TDCS criterion can be written as follow:

$$\max \frac{\text{tr} \left(\mathbf{U}_n^T \mathbf{S}_b^{(n)} \mathbf{U}_n \right)}{\text{tr} \left(\mathbf{U}_n^T \mathbf{S}_w^{(n)} \mathbf{U}_n \right)}\quad (34)$$

According to Rayleigh quotient, Eq. (34) is maximized if and only if the matrix \mathbf{U}_n consists of the L_n generalized eigenvectors corresponding to the largest L_n generalized eigenvalues of the matrix pencil $(\mathbf{S}_b^{(n)}, \mathbf{S}_w^{(n)})$.

Since the $\mathbf{S}_b^{(n)}$ and $\mathbf{S}_w^{(n)}$ depend on $\mathbf{U}_1, \dots, \mathbf{U}_{n-1}, \mathbf{U}_{n+1}, \dots, \mathbf{U}_N$, we can see that the optimization of \mathbf{U}_n depends on the projections in other modes. An iterative procedure can be constructed to maximize Eq. (25). The pseudocode of the proposed method is summarized in Algorithm 2.

Algorithm 2 TDCS

INPUT: a set of M labeled tensor samples \mathcal{A}_i^c , $i = 1, 2, \dots, M_c$, $c = 1, 2, \dots, C$. the number of reduced dimensions L_n , $n = 1, 2, 3$

OUTPUT: a set of projected tensors \mathcal{D}_i^c and 2 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 $\mathbf{U}_3 \in \mathbb{R}^{I_3 \times L_3}$

Algorithm:

Initialize \mathbf{U}_n with a set of identity matrices

Calculate the mean image of the c th individual $\overline{\mathcal{A}}^c$ and the mean image of all individuals $\overline{\mathcal{A}}$ by Eq. (21) and Eq. (22)

repeat

for $n = 1$ to 3 **do**

 Calculate $\mathbf{S}_b^{(n)}$ and $\mathbf{S}_w^{(n)}$ by Eq. (26) and Eq. (27)

 The protection matrix $\mathbf{U}_n = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{L_n}]$ consists of the L_n generalized eigenvectors corresponding to the largest L_n generalized eigenvalues of the matrix pencil $(\mathbf{S}_b^{(n)}, \mathbf{S}_w^{(n)})$.

end for

 Calculate J_{k+1} by Eq. (25)

until $|J_{k+1} - J_k| < \epsilon$

 Compute a set of projected tensors \mathcal{D}_i^c by Eq. (20)

B. Performance analysis of TDCS

Before analyzing the performance of TDCS, we investigate the relations between CID and TDCS, and see how to get the first column vector \mathbf{x}_1 of the color space transformation matrix \mathbf{X} using TDCS. The $\mathbf{A}_{(3)}$, a 3-mode unfolding matrix of a color image \mathcal{A} , has 3 rows which correspond to three color components \mathbf{R} , \mathbf{G} and \mathbf{B} .

$$\mathbf{A}_{(3)} = [\mathbf{R}, \mathbf{G}, \mathbf{B}]^T = \mathbf{A}^T \quad (35)$$

where \mathbf{A} is the color image matrix in CID. And it can also be considered as a 2nd-tensor in vector space $\mathbb{R}^{I \times 3}$, where $I = I_1 \times I_2$. Then, the problem turns into seeking a discriminant projection matrix $\mathbf{U}_1 \in \mathbb{R}^{I \times d}$ and a color space transformation matrix $\mathbf{U}_2 \in \mathbb{R}^{3 \times 1}$ to transform:

$$\mathbf{D} = \mathbf{A} \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \quad (36)$$

In TDCS, the two initial values are provided for \mathbf{U}_1 and \mathbf{U}_2 . Then $\mathbf{S}_b^{(1)}$ is calculated by using Eq. (26):

$$\begin{aligned} \mathbf{S}_b^{(1)} &= \sum_{c=1}^C \left(\overline{\mathbf{A}}_{(1)}^c - \overline{\mathbf{A}}_{(1)} \right) \tilde{\mathbf{U}}_1 \tilde{\mathbf{U}}_1^T \left(\overline{\mathbf{A}}_{(1)}^c - \overline{\mathbf{A}}_{(1)} \right)^T \\ &= \sum_{c=1}^C \left(\overline{\mathbf{A}}^c - \overline{\mathbf{A}} \right) \mathbf{U}_2 \mathbf{U}_2^T \left(\overline{\mathbf{A}}^c - \overline{\mathbf{A}} \right)^T \end{aligned} \quad (37)$$

Replacing \mathbf{U}_2 with \mathbf{x}_1 in the above equation, we have:

$$\mathbf{S}_b^{(1)} = \sum_{c=1}^C \left(\overline{\mathbf{A}}^c - \overline{\mathbf{A}} \right) \mathbf{x}_1 \mathbf{x}_1^T \left(\overline{\mathbf{A}}^c - \overline{\mathbf{A}} \right)^T \quad (38)$$

By comparing the above equation with Eq. (11), we find that the only difference is the coefficient P_c .

Similarly, the $\mathbf{S}_w^{(1)}$ is calculated by using Eq. (27):

$$\begin{aligned} \mathbf{S}_w^{(1)} &= \sum_{c=1}^C \sum_{i=1}^{M_c} \left(\mathbf{A}_{i(1)}^c - \overline{\mathbf{A}}_{(1)}^c \right) \tilde{\mathbf{U}}_1 \tilde{\mathbf{U}}_1^T \left(\mathbf{A}_{i(1)}^c - \overline{\mathbf{A}}_{(1)}^c \right)^T \\ &= \sum_{c=1}^C \sum_{i=1}^{M_c} \left(\mathbf{A}_i^c - \overline{\mathbf{A}}^c \right) \mathbf{U}_2 \mathbf{U}_2^T \left(\mathbf{A}_i^c - \overline{\mathbf{A}}^c \right)^T \end{aligned} \quad (39)$$

Replacing \mathbf{U}_2 with \mathbf{x}_1 in the above equation, we have:

$$\mathbf{S}_w^{(1)} = \sum_{c=1}^C \sum_{i=1}^{M_c} \left(\mathbf{A}_i^c - \overline{\mathbf{A}}^c \right) \mathbf{x}_1 \mathbf{x}_1^T \left(\mathbf{A}_i^c - \overline{\mathbf{A}}^c \right)^T \quad (40)$$

Comparing the above equation with Eq. (12), we find that the only difference is the coefficient $P_c \frac{1}{M_c - 1}$.

We draw that $\frac{\mathbf{S}_b^{(1)}}{\mathbf{S}_w^{(1)}}$ is equivalent to $\frac{\mathbf{S}_b}{\mathbf{S}_w}$ in CID, so the \mathbf{U}_1 obtained by solving the generalized eigenvalue problem $(\mathbf{S}_b^{(1)}, \mathbf{S}_w^{(1)})$ is equivalent to the discriminant matrix \mathbf{P} obtained by solving the generalized eigenvalue problem $(\mathbf{S}_b, \mathbf{S}_w)$ in CID. Similarly, $\frac{\mathbf{S}_b^{(2)}}{\mathbf{S}_w^{(2)}}$ is equivalent to $\frac{\mathbf{L}_b}{\mathbf{L}_w}$ in CID, so the \mathbf{U}_2 obtained by solving the generalized eigenvalue problem $(\mathbf{S}_b^{(2)}, \mathbf{S}_w^{(2)})$ is equivalent to \mathbf{x}_1 obtained by solving the generalized eigenvalue problem $(\mathbf{L}_b, \mathbf{L}_w)$ in CID. The \mathbf{x}_1 and \mathbf{P} are obtained by repeatedly using the above procedure. From the above analysis, we can see that a color image in DTCS is unfolded on 3-mode in CID. The unfolded operator rasterize each component in color space as I -dimensional vector, where $I = I_1 \times I_2$.

Furthermore, we analyze the underlying reason why TDCS can perform better than CID. One reason is that TDCS extracts all discriminant information and CID only extracts a part of the discriminant

information. In CID, the between-class scatter matrix $\mathbf{S}_b(\mathbf{x}_k)$ and the within-class scatter matrix $\mathbf{S}_w(\mathbf{x}_k)$ are the two matrices with size of $I \times I$. By using linear algebra, we have $\text{rank}(\mathbf{S}_b(\mathbf{x}_k)) = C - 1$ and $\text{rank}(\mathbf{S}_w(\mathbf{x}_k)) = \sum_{c=1}^C (N_c - 1)$. Generally, the number of pattern classes C and the number of all samples are far less than I , so the two matrices $\mathbf{S}_b(\mathbf{x}_k)$ and $\mathbf{S}_w(\mathbf{x}_k)$ are singular. In this case, the generalized eigenvalue problem $(\mathbf{S}_b, \mathbf{S}_w)$ can not directly be solved. For releasing this problem, PCA is applied to reduce dimensionality such that $\mathbf{S}_w(\mathbf{x}_k)$ is nonsingular, followed by LDA for classification. However, one potential problem is that the PCA criterion may not be compatible with the LDA criterion; thus, some information, which may contain the most discriminant information, are thrown away in the PCA step [32], [33]. In contrast to CID, TDCS divides $\mathbf{S}_w(\mathbf{x}_k)$ with $I \times I$ size into two matrices $\mathbf{S}_w^{(1)}$ with $I_1 \times I_1$ size and $\mathbf{S}_w^{(2)}$ with $I_2 \times I_2$ size. When $\sum_{c=1}^C (N_c - 1) \geq \max(I_1, I_2)$, we ensure that $\mathbf{S}_w^{(1)}$ and $\mathbf{S}_w^{(2)}$ are non-singular. Furthermore, if $C \geq \max(I_1, I_2)$, then $\mathbf{S}_b^{(1)}$ and $\mathbf{S}_b^{(2)}$ are full rank. In this case, all discriminant information are extracted.

The other reason is that the three column vectors of the color space transformation matrix are optimized simultaneously. In CID, the discriminant projection matrix \mathbf{P} and the first column vector \mathbf{x}_1 are optimized in the first procedure. In the second procedure, then, the matrix \mathbf{P} and the second column vector \mathbf{x}_2 are optimized. The two procedures are mutually independent except for the result \mathbf{P} in the first procedure as the input for the second procedure. Thus, it is uncertain whether \mathbf{P} is optimal with \mathbf{x}_1 , because it has been optimized and changed with \mathbf{x}_2 on the second procedure. Furthermore, this also raises doubts on whether \mathbf{x}_1 and \mathbf{x}_2 are optimal as well. In TDCS, however, the 3 matrices \mathbf{U}_1 , \mathbf{U}_2 and \mathbf{U}_3 are optimized simultaneously. The color space transformation matrix \mathbf{U}_3 in TDCS is prior to the color space transformation matrix \mathbf{X} in CID.

To sum up, TDCS gets not only more discriminant information but also better color space transformation matrix than CID. Therefore, TDCS has better performance than CID.

C. Efficiency analysis of TDCS

In this section, we analyze the efficiency of TDCS from two aspects, the number of iterations and the time complexity of each iteration. In order to compare the number of iterations, we reveal that the two criterion functions of CID and TDCS are equivalent to each other.

We replace $\mathbf{S}_b(\mathbf{x}_k)$ in Eq. (13) with Eq. (11), the numerator of Eq. (13) can be rewritten as:

$$\begin{aligned}
|\mathbf{P}^T \mathbf{S}_b(\mathbf{x}_k) \mathbf{P}| &= \left| \mathbf{P}^T \sum_{c=1}^C P_c [(\overline{\mathbf{A}}^c - \overline{\mathbf{A}}) \mathbf{x}_k \mathbf{x}_k^T (\overline{\mathbf{A}}^c - \overline{\mathbf{A}})^T] \mathbf{P} \right| \\
&= \left| \sum_{c=1}^C P_c [(\mathbf{P}^T \overline{\mathbf{A}}^c \mathbf{x}_k - \mathbf{P}^T \overline{\mathbf{A}} \mathbf{x}_k)(\mathbf{P}^T \overline{\mathbf{A}}^c \mathbf{x}_k - \mathbf{P}^T \overline{\mathbf{A}} \mathbf{x}_k)^T] \right| \\
&= \left| \sum_{c=1}^C P_c \|\mathbf{P}^T \overline{\mathbf{A}}^c \mathbf{x}_k - \mathbf{P}^T \overline{\mathbf{A}} \mathbf{x}_k\|_F \right|
\end{aligned} \tag{41}$$

Because a matrix is a 2nd-order tensor, the above equation can be rewritten using n -mode product of a tensor by a matrix,

$$|\mathbf{P}^T \mathbf{S}_b(\mathbf{x}_k) \mathbf{P}| = \left| \sum_{c=1}^C P_c \|\overline{\mathbf{A}}^c \times_1 \mathbf{P} \times_2 \mathbf{x}_k - \overline{\mathbf{A}} \times_1 \mathbf{P} \times_2 \mathbf{x}_k\|_F \right| \tag{42}$$

Similarly, the numerator of Eq. (25) be rewritten as following:

$$\begin{aligned}
\Psi_b(\mathcal{D}) &= \sum_{c=1}^C \|\overline{\mathcal{D}}^c - \overline{\mathcal{D}}\|_F^2 \\
&= \sum_{c=1}^C \|\overline{\mathcal{A}}^c \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \times_3 \mathbf{U}_3^T - \overline{\mathcal{A}} \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \times_3 \mathbf{U}_3^T\|_F^2
\end{aligned} \tag{43}$$

Comparing the above two equations, we find that the differences are only a set of coefficients P_c and the number of the order. Similarly, the differences between the denominators of Eq. (13) and Eq. (25) are a set of coefficients $P_c \frac{1}{M_c-1}$ and the number of order. Therefore, the criteria of CID and TDCS are equivalent to each other. On the basis of that, the number of iterations will be compared in Section V-C.

Next, we investigate the time complexity of each iteration of the two algorithms. For each iteration of CID, we need to solve two generalized eigenvector problems $(\mathbf{S}_b(\mathbf{x}), \mathbf{S}_w(\mathbf{x}))$ and $(\mathbf{L}_b(\mathbf{P}), \mathbf{L}_w(\mathbf{P}))$. They require about $14(I_1 \times I_2)^3$ and 14×3^3 floating-point multiplications, respectively[34]. In TDCS, solving three generalized eigenvector problems $(\mathbf{S}_b^{(n)}, \mathbf{S}_w^{(n)})$ ($n = 1, 2, 3$) require about $14 \times I_1^3$, $14 \times I_2^3$ and 14×3^3 floating-point multiplications, respectively. In these two algorithms, the time complexity for calculating each scatter matrix is $O((I_1 \times I_2)^2)$. From the above analysis, we can draw that the time complexity of an iteration in CID is $O((I_1 \times I_2)^3)$ and that in TDCS is $O((I_1 \times I_2)^2)$. So, the time complexity of an iteration in TDCS is one order of magnitude less than that in CID.

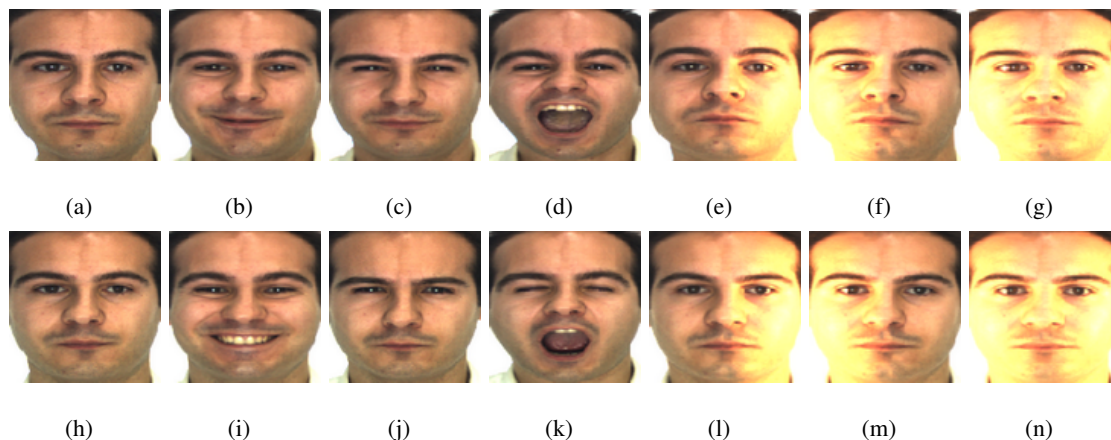


Fig. 2. Sample images of one individual from the AR database.

V. EXPERIMENTS AND RESULTS

A. Database

We conducted the experiments on two well-known color face databases AR[35] and Georgia Tech face databases ¹.

The AR database contains over 4,000 color face images of 126 people (70 men and 56 women). In the experiments, we selected 100 people (50 men and 50 women) of them. Images are frontal view faces with different facial expressions, illumination conditions, and occlusions (sun glasses and scarf). The pictures were taken under strictly controlled conditions. No restrictions on wear (clothes, glasses, etc.), make-up, hair style, etc. were imposed to participants. Each individual participated in two sessions, separated by two weeks (14 days). The same pictures were taken in both sessions. The 14 images of each individual are selected and the occluded face images are excluded in our experiment. All images are cropped into 32×32 pixels. The sample images for one individual of the AR database are shown in Fig. 2, where Fig. 2(a)-Fig. 2(g) are from the first session as the training set, and Fig. 2(h)-Fig. 2(n) are from the second session as the testing set.

Georgia Tech face database contains images of 50 individuals taken in two or three sessions at different times. Each individual in the database is represented by 15 color JPEG images with cluttered background taken at resolution 640×480 pixels. The average size of the faces in these images is 150×150 pixels. The pictures show frontal and/or tilted faces with different facial expressions, lighting conditions and scale.

¹http://www.anejian.com/research/face_reco.htm



Fig. 3. Sample images of one individual from the Georgia Tech database (non-aligned head images).

Each image is manually cropped and resized to 32×32 pixels. The sample images for one individual of the Georgia Tech database are showed in Fig. 3.

B. Experiment setting

For the purpose of evaluating the performance of TDCS, we use *face verification rate* as the criteria. The FERET Verification Testing Protocol [36] recommends using the receiver operating characteristic (ROC) curves to depict the relations between the face verification rate (FVR) and the false accept rate (FAR). In our experiments, the Euclidean distance is used to generate the score matrix. The ROC curves are plotted by the Statistical Learning Toolbox² using the score matrix. For tensor operations, we used the tensor toolbox developed by Bader and Kolda in MATLABTM[37]. All experiments are conducted on

²The `sverifyroc` function in the Statistical Learning Toolbox can only plot the ROC curve illustrating the relations of the false reject rate versus the FAR. We modified it to depict the relations between the FVR and the FAR.

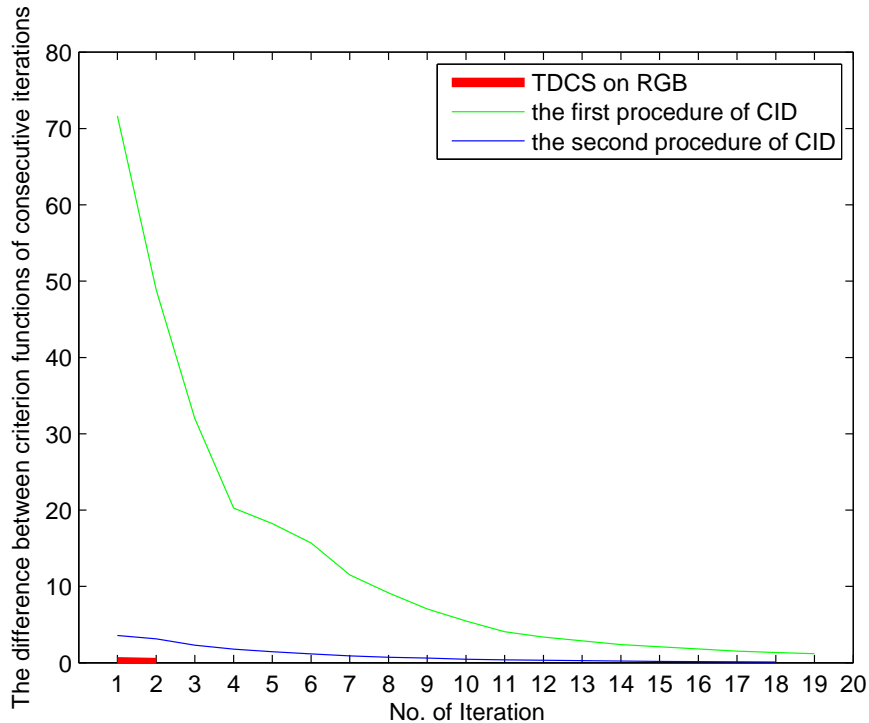


Fig. 4. Illustration of the convergence of the TDCS algorithm and the CID algorithm.

a 2.66 GHz Intel PC with 16 GB memory, with the Microsoft Windows XP 64 bits as OS.

C. Convergence and time complexity

Firstly, the comparisons between the number of iterations in TDCS and that of CID are drawn on all images of the first session in AR database. The convergence threshold ϵ is set as 0.1 for the two algorithms. For CID, the \mathbf{x}_1 is initialized as $[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}]^T$ in the first procedure and the \mathbf{y} is initialized as $[1, 0]^T$ in the second procedure. Fig. 4 illustrates the convergence of the TDCS algorithm and the two procedures of the CID algorithm. That the criterion functions of CID and TDCS are equivalent to each other has been analyzed theoretically in Section IV-C. From Fig. 4, we can see that the difference between the convergences of consecutive procedures. The first procedure of CID still hasn't reached the convergence threshold after 20 iterations. The convergence threshold is reached in the 19th iteration on the second procedure of CID; whereas TDCS only needs 3 iterations to reach the same convergence threshold. The number of all iterations in the two procedures of CID is more than 7 times as many as that of TDCS.

Secondly, we compare TDCS with CID with respect to time efficiency issue on the database introduced above using the various sizes. In order to avoid getting into an infinite iteration, the procedure will be stopped, when the number of iteration reach 20 and the convergence threshold is not reached. Table. I lists the cost times, the number of iterations and the convergence values of the both algorithms on the various sizes. In the first procedure of CID, the convergence threshold is not reached on all sizes. In the second procedure, It is reached on 16×16 , 32×32 and 40×40 , respectively. Even so, TDCS only needs about 2 or 3 iterations to converge. On the size of 16×16 , the cost time of CID is 50 seconds and that of TDCS is only 7 seconds. When the size is 56×56 , the cost time of CID dramatically rises to 19,434 seconds, whereas that of TDCS slowly rises to 77 seconds. The experimental results are consistent with our time complexity analysis in Section IV-C. Therefore, the efficiency of the proposed algorithm is better than that of CID.

TABLE I
COMPARISON OF TDCS AND CID WITH RESPECT TO THE COST TIME AND THE TIMES OF ITERATION

| size | 16×16 | 24×24 | 32×32 | 40×40 | 48×48 | 56×56 |
|---|----------------|----------------|----------------|----------------|----------------|----------------|
| cost time of CID (s) | 50 | 360 | 1814 | 4296 | 10090 | 19434 |
| the times of iteration in CID's the 1st procedure | 20 | 20 | 20 | 20 | 20 | 20 |
| the convergence value in CID's the 1st procedure | 5.07 | 1.83 | 1.17 | 0.61 | 0.19 | 0.25 |
| the times of iteration in CID's the 2nd procedure | 15 | 20 | 19 | 18 | 20 | 20 |
| the convergence value in CID's the 2nd procedure | 0.09 | 0.22 | 0.09 | 0.07 | 0.12 | 0.09 |
| cost time of TDCS (s) | 7 | 10 | 20 | 36 | 66 | 77 |
| the times of iteration in TDCS | 3 | 3 | 3 | 3 | 3 | 2 |
| the convergence vale in TDCS | 0.002 | 0.01 | 0.02 | 0.06 | 0.04 | 0.05 |

D. Experiments and results on the AR database

In this experiment, we train TDCS model and CID model by using 700 color face images from the first session in AR database and test them by all images in the second session. The convergence threshold ϵ is set as 0.1 and \mathbf{x}_1 is initialized as $[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}]^T$. In this case, we get two color space transformation matrices:

$$\mathbf{X} = \begin{bmatrix} -0.1839 & 0.2519 & 0.6209 \\ 0.2059 & 0.5019 & -1.0355 \\ -1.0000 & -0.9629 & 0.4391 \end{bmatrix} \quad (44)$$

and

$$\mathbf{U}_3 = \begin{bmatrix} -0.0840 & 0.2640 & 0.4363 \\ 0.2931 & 0.5008 & -0.8141 \\ -0.9524 & -0.8243 & 0.3832 \end{bmatrix} \quad (45)$$

Using the two matrices, we get 3 color components \mathbf{D}^1 , \mathbf{D}^2 , \mathbf{D}^3 of CID and 3 color components \mathbf{T}^1 , \mathbf{T}^2 , \mathbf{T}^3 of TDCS. These components are illustrated in Fig. 5.

Meanwhile, we conduct LDA and 2D-LDA on corresponding gray images. Because there are 100 individuals in the AR database, only 99 discriminant projection basis vectors are extracted in LDA and CID. For 2D-LDA and TDCS, the two numbers of the spatial reduced dimensions are 10 and 10, respectively. The ROC curves of the four methods are shown in Fig. 6. The results indicate that the TDCS is more effective for improving the AR performance than other three algorithms. The curves also show that the 20 discriminant projection basis vectors in TDCS contain more discriminant information than 100 ones in CID. It is coincided with our performance analysis in Section IV-B. Fig. 6 also shows that CID outperforms 2D-LDA, which illustrates that choosing an optimal color space is more important than keeping spatial structure for recognizing color face images.

Another experiment is conducted as followed. In the experiment, all color images are transformed into one color component using the first column vector of \mathbf{U}_3 and \mathbf{x}_1 to get two color spaces with the first color component. The 2D-LDA is implemented on the first color component of two color spaces and gray images. The results are shown in Fig. 7. It is interesting that the curves for TDCS and CID are superposed practically. Intuitively, the image generated by the first component of CID (see Fig. 5(g)) is very similar to the one generated by the first component of TDCS (see Fig. 5(d)). The results are consistent with our theoretical analysis of the relations between TDCS and CID in section IV-B.

E. Experiments and results on the Georgia Tech face database

Georgia Tech face database is more complex than AR database, because it contains various pose faces with different expressions on cluttered background. In this experiment, we use the first 8 images of each individual as the training set and the remain as the testing set. The TDCS model and CID model are trained and we get two color space transformation matrices:

$$\mathbf{X} = \begin{bmatrix} -1.0000 & 0.4894 & 0.4076 \\ 0.8473 & 0.3595 & -1.0134 \\ -0.2767 & -1.0401 & 0.5332 \end{bmatrix} \quad (46)$$

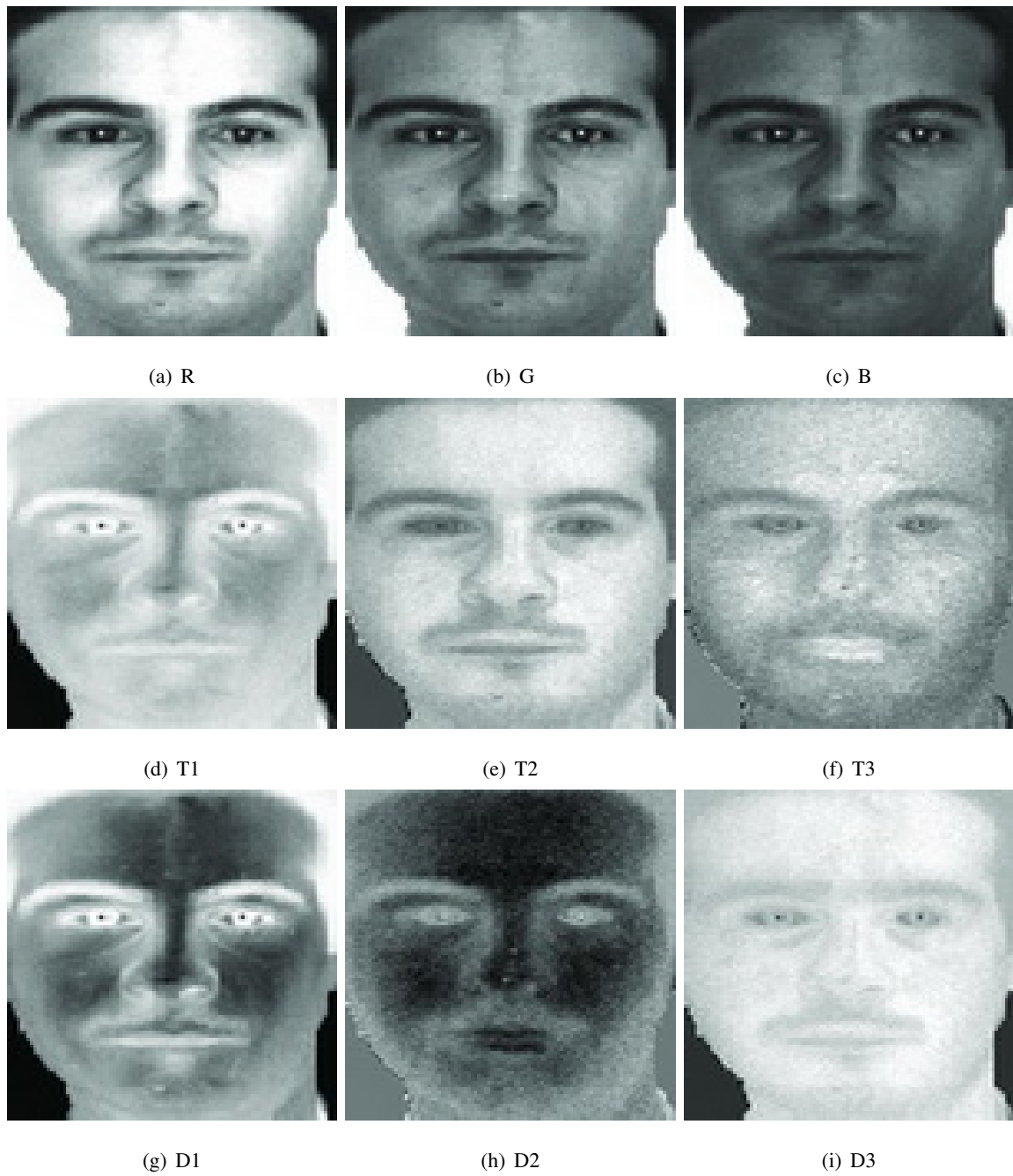


Fig. 5. Illustration of R, G, and B color components and the various components generated by CID and TDCS on the AR face database.

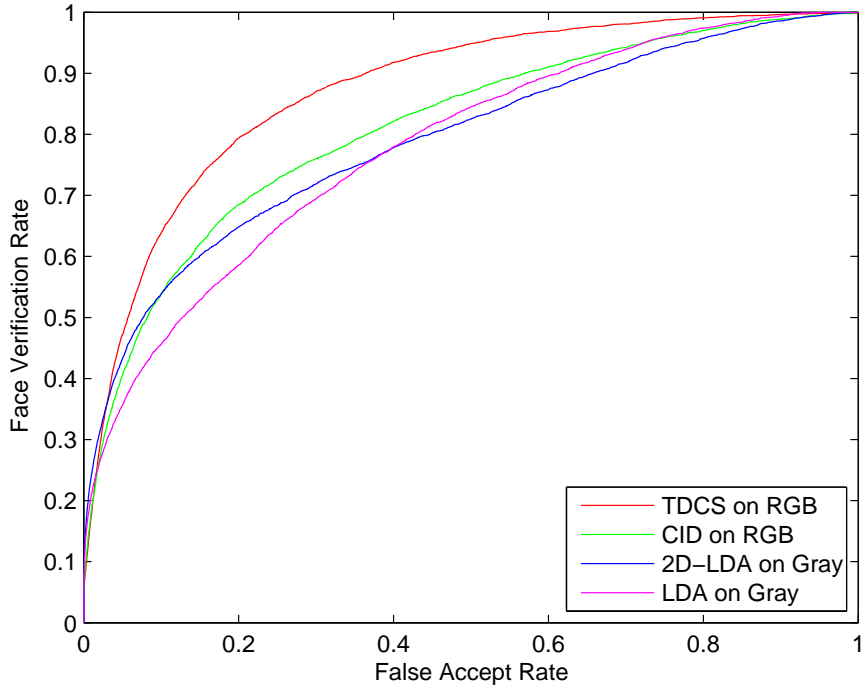


Fig. 6. ROC curves of TDCS, CID, 2D-LDA and LDA on AR face database.

and

$$\mathbf{U}_3 = \begin{bmatrix} 0.1067 & -0.3004 & 0.6192 \\ 0.7589 & 0.7798 & -0.7852 \\ -0.6424 & -0.5492 & 0.0080 \end{bmatrix} \quad (47)$$

The two matrices are not the same as Eq. (44) and Eq. (45) due to the different training sets. Using the two matrices, we get 3 color components \mathbf{D}^1 , \mathbf{D}^2 , \mathbf{D}^3 of CID and 3 color components \mathbf{T}^1 , \mathbf{T}^2 , \mathbf{T}^3 of TDCS. These components are illustrated in Fig. 8.

With 50 individuals in the Georgia Tech database, the 49 discriminant projection basis vectors are extracted in LDA and CID. For 2D-LDA and TDCS, the two numbers of the spatial reduced dimensions are 10 and 10, respectively. The ROC curves of the four methods are solid lines in Fig. 9. The results indicate that TDCS has the best performance compared to the other three algorithms. As shown in Fig. 6, TDCS is better than CID, especially in a complicated face database with different poses as portrayed in Fig. 9.

Similarly, The 2D-LDA is implemented on the first color components of two color spaces and gray images. The results are shown in Fig. 10, in which the three curves are almost consistent. It reveals that

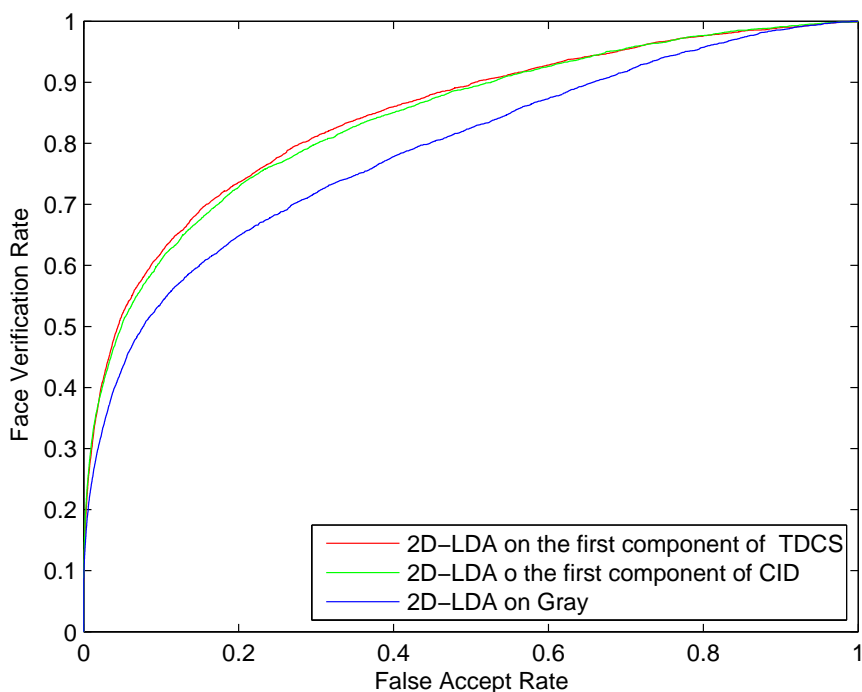


Fig. 7. ROC curves of 2D-LDA on the first component of TDCS, CID and gray images on the AR face database.

one color component is not enough for the discrimination of color images in more complex case.

Lately, all images in the Georgia Tech database are manually aligned (two eyes were aligned at the same position), cropped, and then re-sized to 32×32 pixels. In the cropped images shown in Fig. 11, we retain as much of the facial region as possible, thereby eliminating most of the non-facial regions. The experiments with the same setting are conducted on them. The ROC curves of the four methods are dash-dot lines in Fig. 9. It is interesting that the two ROC curves of TDCS of the head images and the facial images are overlapped and the two ROC curves of 2D-LDA have the same manner because they maintain the spatial structure information of images. Whereas, the corresponding two ROC curves of other two algorithms have distinctions which indicate that TDCS is more robust than CID for the head and face images.

VI. CONCLUSION

The paper presents a novel color space model by learning from the training samples borrowing the idea from DATER. The model is named the tensor discriminant color space (TDCS) which optimizes one color space transformation matrix and two discriminant projection matrices simultaneously. And we

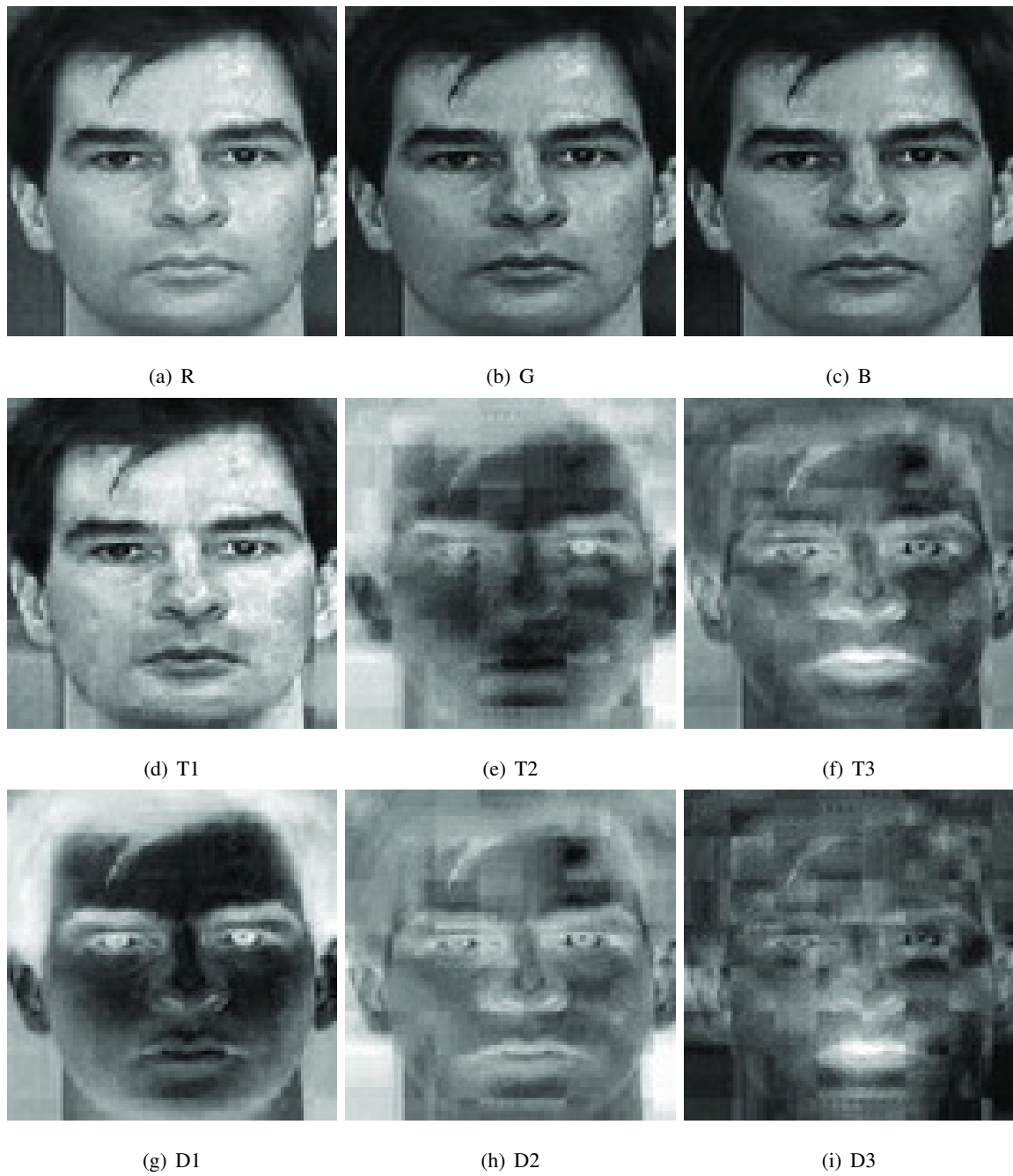


Fig. 8. Illustration of R, G, and B color components and the various components generated by CID and TDCS on the Georgia Tech face database.

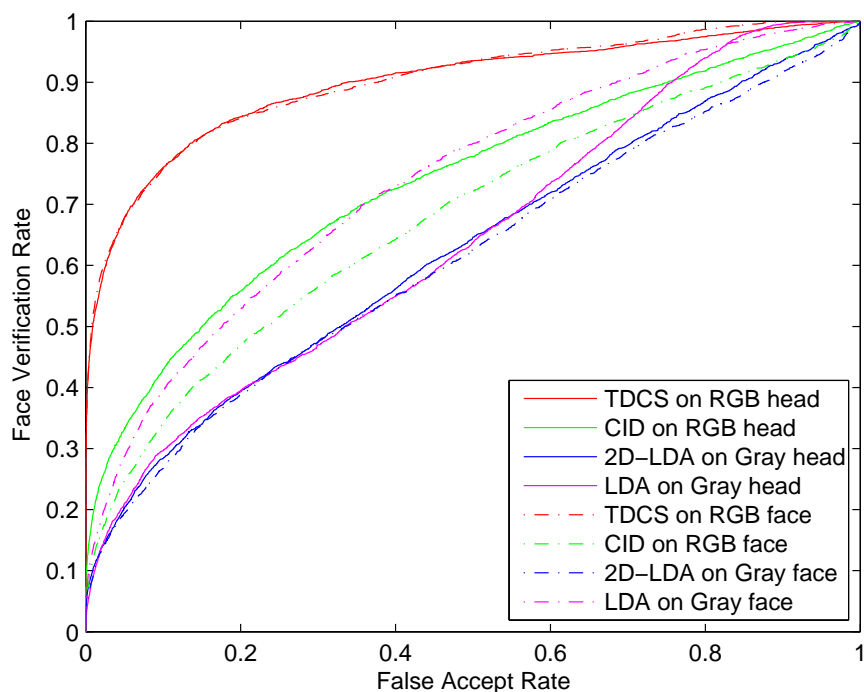


Fig. 9. ROC curves of TDCS, CID, 2D-LDA and LDA on the Georgia Tech face database.

analyze the relations and distinctions between TDCS and CID, theoretically. The experiments on the AR and Georgia Tech color face database have been systematically performed. Their results reveal a number of interesting remarks:

- 1) Theoretically, the time complexity of each iteration in CID is $O((I_1 \times I_2)^3)$ and that in TDCS is $O((I_1 \times I_2)^2)$. From the experiment, moreover, we can observe that the number of all iterations of CID is more than 7 times as many as that of TDCS. Therefore, the efficiency of our proposed algorithm is better than that of CID.
- 2) The TDCS model can achieve the optimal discriminating color space and the optimal spatial transformation matrix simultaneously.
- 3) The TDCS outperforms CID, especially in more complex face database such as the Georgia Tech face database.
- 4) One color component is not enough for the discrimination of color images in more complex cases.
- 5) TDCS is more robust for the face and head images, because it keeps spatial structure information of color images.

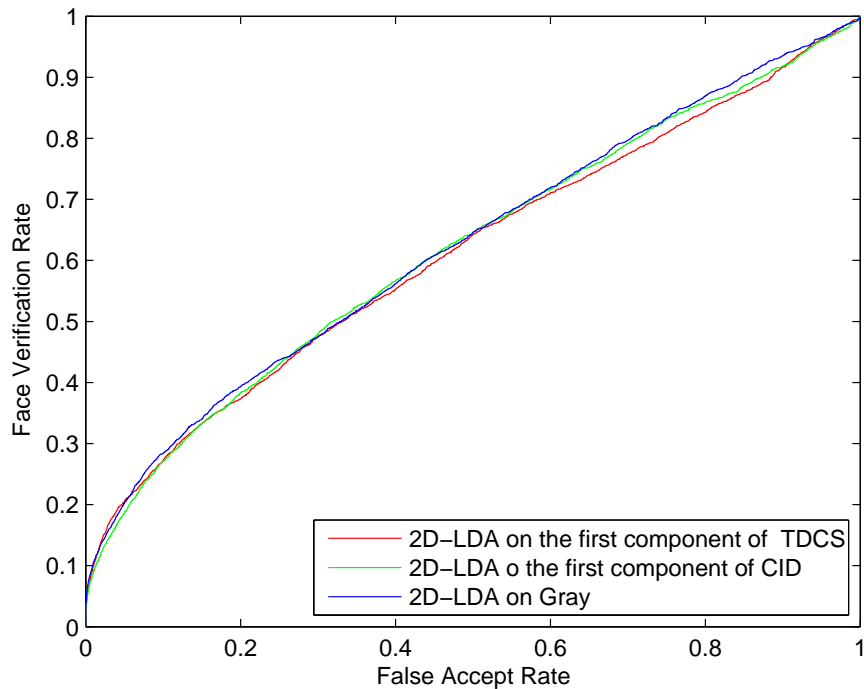


Fig. 10. ROC curves of 2D-LDA on the first component of TDCS, CID and gray images on the Georgia Tech face database.

However, some theoretical justification for the convergence of the algorithm is still missing. Discussions on the convergence of the algorithm in theory will be our future work.

ACKNOWLEDGMENT

The authors would like to thank Xiao-Hua Liu, Xu-Jun Peng, Yu-Hsin Chen and De-Cai Zhang for their valuable comments.

REFERENCES

- [1] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, Jan. 1991.
- [2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, Jul. 1997.
- [3] J. Yang, D. Zhang, A. F. Frangi, and J. Y. Yang, "Two-dimensional PCA: A new approach to appearance-based face representation and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 131–137, 2004.
- [4] M. Li and B. Z. Yuan, "2D-LDA: A statistical linear discriminant analysis for image matrix," *Pattern Recognition Letters*, vol. 26, no. 5, pp. 527–532, 2005.



Fig. 11. Sample images for one individual of the Georgia Tech database (aligned facial images).

- [5] R. Kemp, G. Pike, P. White, and A. Musselman, "Perception and recognition of normal and negative faces: The role of shape from shading and pigmentation cues," *Perception*, vol. 25, no. 1, pp. 37–52, 1996.
- [6] L. Torres, J. Reutter, and L. Lorente, "The importance of the color information in face recognition," in *Proceedings. 1999 International Conference on Image Processing, 1999. ICIP 99.*, vol. 3. IEEE, 1999, pp. 627–631.
- [7] C. Wang, B. Yin, X. Bai, and Y. Sun, "Color face recognition based on 2DPCA," in *19th International Conference on Pattern Recognition, 2008. ICPR 2008.*, 2008, pp. 1–4.
- [8] J. Choi, Y. Ro, and K. Plataniotis, "Color face recognition for degraded face images," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 39, no. 5, pp. 1217–1230, 2009.
- [9] M. Villegas, R. Paredes, A. Juan, and E. Vidal, "Face verification on color images using local features," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08.* IEEE, 2008, pp. 1–6.
- [10] W. H. Buchsbaum, *Color TV Servicing*, 3rd ed. Prentice-Hall, 1975.
- [11] H. Stokman and T. Gevers, "Selection and fusion of color models for image feature detection," *IEEE transactions on pattern analysis and machine intelligence*, pp. 371–381, 2007.
- [12] O. Ikeda, "Segmentation of faces in video footage using HSV color for face detection and image retrieval," in *Proceedings. 2003 International Conference on Image Processing, 2003. ICIP 2003.*, vol. 3. IEEE, 2003.

- [13] R. Hsu, M. Abdel-Mottaleb, and A. Jain, "Face detection in color images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 696–706, 2002.
- [14] C. F. Jones III and A. L. Abbott, "Optimization of color conversion for face recognition," *Eurasip Journal on Applied Signal Processing*, vol. 2004, no. 4, pp. 522–529, 2004.
- [15] V.-E. Neageo, "An optimum 2D color space for pattern recognition," in *Proc. Int. Conf. Image Processing, Computer Vision, Pattern Recognition*, 2006, pp. 526–532.
- [16] Z. Liu and C. Liu, "A hybrid color and frequency features method for face recognition," *IEEE Transactions on Image Processing*, vol. 17, no. 10, pp. 1975–1980, 2008.
- [17] C. Liu, "Learning the uncorrelated, independent, and discriminating color spaces for face recognition," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 2, pp. 213–222, 2008.
- [18] C. Liu and J. Yang, "ICA color space for pattern recognition," *IEEE Transactions on Neural Networks*, vol. 20, no. 2, pp. 248–257, 2009.
- [19] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition CVPR 2005*, vol. 1, 2005, pp. 947–954.
- [20] J. Yang and C. Liu, "Color image discriminant models and algorithms for face recognition," *IEEE Transactions on Neural Networks*, vol. 19, no. 12, pp. 2088–2098, 2008.
- [21] J. Yang, C. Liu, and J. Y. Yang, "What kind of color spaces is suitable for color face recognition?" *Neurocomputing*, vol. 73, no. 10-12, pp. 2140–2146, 2010.
- [22] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear analysis of image ensembles: Tensorfaces," in *ECCV '02: Proceedings of the 7th European Conference on Computer Vision-Part I*. London, UK: Springer-Verlag, 2002, pp. 447–460.
- [23] —, "Multilinear independent components analysis," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR 2005*, vol. 1, Jun. 20–25, 2005, pp. 547–553.
- [24] L. De Lathauwer, B. De Moor, and J. Vandewalle, "On the best rank-1 and rank-(r_1, r_2, \dots, r_n) approximation of higher-order tensors," *Siam Journal On Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1324–1342, 2000.
- [25] —, "A multilinear singular value decomposition," *Siam Journal On Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1253–1278, 2000.
- [26] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *Siam Review*, vol. 51, no. 3, pp. 455–500, 2009.
- [27] J. Ye, "Generalized low rank approximations of matrices," *Machine Learning*, vol. 61, no. 1, pp. 167–191, 2005.
- [28] X. He, D. Cai, and P. Niyogi, "Tensor subspace analysis," in *In Advances in Neural Information Processing Systems 18 (NIPS)*. MIT Press, 2005.
- [29] H. P. Lu, N. P. Konstantinos, and A. N. Venetsanopoulos, "MPCA: Multilinear principal component analysis of tensor objects," *IEEE Transactions on Neural Networks*, vol. 19, no. 1, pp. 18–39, 2008.
- [30] X. F. He and P. Niyogi, "Locality preserving projections," *Advances In Neural Information Processing Systems 16*, vol. 16, pp. 153–160, 2004.
- [31] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, and H. Zhang, "Discriminant analysis with tensor representation," 2005.
- [32] H. Yu and J. Yang, "A direct LDA algorithm for high-dimensional data – with application to face recognition," *Pattern Recognition*, vol. 34, no. 10, pp. 2067–2070, 2001.
- [33] D.-Q. Dia and P. C. Yuen, "Face recognition by regularized discriminant analysis," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 37, no. 4, pp. 1080–1085, 2007.

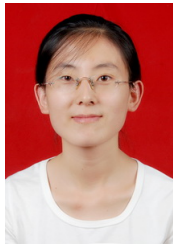
- [34] G. Golub and C. Van Loan, *Matrix computations*. Johns Hopkins Univ Pr, 1996.
- [35] A. Martinez and R. Benavente, "The AR face database," *Univ. Purdue, CVC Tech. Rep.*, vol. 24, 1998.
- [36] H. Moon and P. Phillips, "The FERET verification testing protocol for face recognition algorithms," in *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*. IEEE, 1998, pp. 48–53.
- [37] B. Bader and T.G.Kolda, "Tensor toolbox version 2.3, copyright 2009, sandia national laboratories, <http://csmr.ca.sandia.gov/~tgkolda/TensorToolbox/>."



Su-Jing Wang received the Master's degree from the Software College of Jilin University, Changchun, China, in 2007. From September 2008, he is pursuing to the Ph.D. degree at the College of Computer Science and Technology of Jilin University. He is the author or co-author of more than 10 scientific papers. His current research interests include pattern recognition, computer vision and machine learning.



Jian Yang received the BS degree in mathematics from the Xuzhou Normal University in 1995. He received the MS degree in applied mathematics from the Changsha Railway University in 1998 and the Ph.D. degree from the Nanjing University of Science and Technology (NUST), on the subject of pattern recognition and intelligence systems in 2002. In 2003, he was a postdoctoral researcher at the University of Zaragoza, and in the same year, he was awarded the RyC program Research Fellowship sponsored by the Spanish Ministry of Science and Technology. From 2004 to 2006, he was a postdoctoral fellow at Biometrics Centre of Hong Kong Polytechnic University. From 2006 to 2007, he was a postdoctoral fellow at Department of Computer Science of New Jersey Institute of Technology. Now, he is a professor in the School of Computer Science and Technology of NUST. He is the author of more than 50 scientific papers in pattern recognition and computer vision. His research interests include pattern recognition, computer vision and machine learning. Currently, he is an associate editor of Pattern Recognition Letters and Neurocomputing, respectively.



Na Zhang received the B.E. degree in Computer Science and Technology from Jilin University, Changchun, China, in 2008. From September 2008 till now, she is a graduate student in the Jilin University of College of Computer Science and Technology(CCST). Her research interests include computer simulation and computer vision.



Chun-Guang Zhou PhD, professor, PhD supervisor, Dean of Institute of Computer Science of Jilin University. He is Jilin-province-management Expert, Highly Qualified Expert of Jilin Province, One-hundred Science-Technique elite of Changchun. And he is awarded the Governmental Subsidy from the State Department. He has many pluralities of national and international academic organizations. His research interests include related theories, models and algorithms of artificial neural networks, fuzzy systems and evolutionary computations, and applications of machine taste and smell, image manipulation, commercial intelligence, modern logistic, bioinformatics, and biometric identification based on computational intelligence. he has published over 168 papers in Journals and conferences and he published 1 academic book.