We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS BOOK CITATION INDEX INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

14

Discriminating Color Faces for Recognition

Jian Yang[1], Chengjun Liu[2] and Jingyu Yang[1]
[1]*School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, P. R. China*
[2]*Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102, USA*

1. Introduction

Color provides useful and important information for object detection, tracking and recognition, image (or video) segmentation, indexing and retrieval, etc. [1-15]. Color constancy algorithms [13, 14] and color histogram techniques [5, 10-12], for example, provide efficient tools for indexing in a large image database or for object recognition under varying lighting conditions. Different color spaces (or color models) possess different characteristics and have been applied for different visual tasks. For instance, the *HSV* color space and the YC_bC_r color space were demonstrated effective for face detection [2, 3], and the modified $L^*u^*v^*$ color space was chosen for image segmentation [7]. Recently, a selection and fusion scheme of multiple color models was investigated and applied for feature detection in images [15].

Although color has been demonstrated helpful for face detection and tracking, some past research suggests that color appears to confer no significant face recognition advantage beyond the luminance information [16]. Recent research efforts, however, reveal that color may provide useful information for face recognition. The experimental results in [17] show that the principle component analysis (PCA) method [35] using color information can improve the recognition rate compared to the same method using only luminance information. The results in [18] further reveal that color cues do play a role in face recognition and their contribution becomes evident when shape cues are degraded. Other research findings also demonstrate the effectiveness of color for face recognition [19-22, 38].

If color does help face recognition, then a question arises: how should we represent color images for the recognition purpose? One common practice is to convert color images in the RGB color space into a grayscale image by averaging the three color component images before applying a face recognition algorithm for recognition. However, there are neither theoretical nor experimental justifications for supporting that such a grayscale image is a good representation of the color image for the recognition purpose. Other research effort is to choose an existing color space or a color component configuration for achieving good recognition performance with respect to a specific recognition method. For instance, Rajapakse et al. [19] used the RGB color space and nonnegative matrix factorization (NMF) method for face recognition. Torres et al. [17] suggested using the YUV color space or the configuration of S and V components from the HSV color space together with PCA for

Source: Recent Advances in Face Recognition, Book edited by: Kresimir Delac, Mislav Grgic and Marian Stewart Bartlett,
ISBN 978-953-7619-34-3, pp. 236, December 2008, I-Tech, Vienna, Austria

feature extraction. Shih and Liu [21] showed that the color configuration YQC_r, where Y and Q color components are from the YIQ color space and C_r is from the YC_bC_r color space, was effective for face recognition using the enhanced Fisher linear discriminant (FLD) model [23]. In summary, current research efforts apply a separate strategy by first choosing a color image representation scheme and then evaluating its effectiveness using a recognition method. This separate strategy cannot theoretically guarantee that the chosen color image representation scheme is best for the subsequent recognition method and therefore cannot guarantee that the resulting face recognition system is optimal in performance.

The motivation of this chapter is to seek a meaningful representation and an effective recognition method of color images in a unified framework. We integrate color image representation and recognition into one discriminant analysis model: color image discriminant (CID) model. In contrast to the classical FLD method [24], which involves only one set of variables (one or multiple discriminant projection basis vectors), the proposed CID model involves two sets of variables: a set of color component combination coefficients for color image representation and one or multiple discriminant projection basis vectors for image discrimination. The two sets of variables can be determined optimally and simultaneously by the developed, iterative CID algorithm. The CID algorithm is further extended to generate three color components (like the three color components of the RGB color images) for further improving face recognition performance.

We use the Face Recognition Grand Challenge (FRGC) database and the Biometric Experimentation Environment (BEE) system to assess the proposed CID models and algorithms. FRGC is the most comprehensive face recognition efforts organized so far by the US government, and it consists of a large amount of face data and a standard evaluation system, known as the Biometric Experimentation Environment (BEE) system [25, 26]. The BEE baseline algorithm reveals that the FRGC version 2 Experiment 4 is the most challenging experiment, because it assesses face verification performance of controlled face images versus uncontrolled face images. We therefore choose FRGC version 2 Experiment 4 to evaluate our algorithms, and the experimental results demonstrate the effectiveness of the proposed models and algorithms.

2. CID model and algorithm

In this section, we first present our motivation to build the color image discriminant model and then give the mathematical description of the model and finally design an iterative algorithm for achieving its optimal solution.

2.1 Motivation

We develop our general discriminant model based on the RGB color space since it is a fundamental and commonly-used color space. Let **A** be a color image with a resolution of $m \times n$, and let its three color components be **R**, **G**, and **B**. Without loss of generality, we assume that **R**, **G**, and **B** are column vectors: $\mathbf{R}, \mathbf{G}, \mathbf{B} \in R^N$, where $N = m \times n$. The color image **A** is then expressed as an $N \times 3$ matrix: $\mathbf{A} = [\mathbf{R}, \mathbf{G}, \mathbf{B}] \in R^{N \times 3}$.

How should we represent the color image **A** for the recognition purpose? Common practice is to linearly combine its three color components into one grayscale image:

$$\mathbf{E} = \tfrac{1}{3}\mathbf{R} + \tfrac{1}{3}\mathbf{G} + \tfrac{1}{3}\mathbf{B} \tag{1}$$

The grayscale image **E** is then used to represent **A** for recognition. However, theoretical explanation is lacking in supporting that such a grayscale image is a good representation of image **A** for image recognition.

The motivation of this chapter is to seek a more effective representation of the color image **A** for image recognition. Our goal is to find a set of optimal coefficients to combine the **R**, **G**, and **B** color components within a discriminant analysis framework. Specifically, let **D** be the combined image given below:

$$\mathbf{D} = x_1\mathbf{R} + x_2\mathbf{G} + x_3\mathbf{B}, \tag{2}$$

where x_1, x_2 and x_3 are the color component combination coefficients. Now, our task is to find a set of optimal coefficients so that **D** is the best representation of the color image **A** for image recognition.

Given a set of training color images with class labels, we can generate a combined image **D** for each image $\mathbf{A} = [\mathbf{R}, \mathbf{G}, \mathbf{B}]$. Let us discuss the problem in the *D*-space, i.e., the pattern vector space formed by all the combined images defined by Eq. (2). In order to achieve the best recognition performance, we borrow the idea of Fisher linear discriminant analysis (FLD) [24] to build a color image discriminant (CID) model. Note that the CID model is quite different from the classical FLD model since it involves an additional set of variables: the color component combination coefficients x_1, x_2 and x_3. In the following, we will show the details of a CID model and its associated CID algorithm for deriving the optimal solution of the model.

2.2 CID model

Let c be the number of pattern classes, A_{ij} be the j-th color image in class i, where $i = 1, 2, \cdots, c$, $j = 1, 2, \cdots, M_i$, and M_i denotes the number of training samples in class i.

The mean image of the training samples in class i is

$$\bar{\mathrm{A}}_i = \frac{1}{M_i}\sum_{j=1}^{M_i}\mathrm{A}_{ij} = [\bar{\mathrm{R}}_i, \bar{\mathrm{G}}_i, \bar{\mathrm{B}}_i]. \tag{3}$$

The mean image of all training samples is

$$\bar{\mathrm{A}} = \frac{1}{M}\sum_{i=1}^{c}\sum_{j=1}^{M_i}\mathrm{A}_{ij} = [\bar{\mathrm{R}}, \bar{\mathrm{G}}, \bar{\mathrm{B}}], \tag{4}$$

where M is the total number of training samples, i.e., $M = \sum_{i=1}^{c} M_i$.

The combined image of three color components of the color image $\mathrm{A}_{ij} = [\mathrm{R}_{ij}, \mathrm{G}_{ij}, \mathrm{B}_{ij}]$ is given by

$$\mathrm{D}_{ij} = x_1\mathrm{R}_{ij} + x_2\mathrm{G}_{ij} + x_3\mathrm{B}_{ij} = [\mathrm{R}_{ij}, \mathrm{G}_{ij}, \mathrm{B}_{ij}]\mathrm{X} \tag{5}$$

Let $\bar{\mathrm{D}}_i$ be the mean vector of the combined images in class i and $\bar{\mathrm{D}}$ the grand mean vector:

$$\bar{\mathrm{D}}_i = \bar{\mathrm{A}}_i \mathrm{X} \tag{6}$$

$$\bar{\mathrm{D}} = \bar{\mathrm{A}} \mathrm{X} \tag{7}$$

The between-class scatter matrix $\mathrm{S}_b(\mathrm{X})$ and the within-class scatter matrix $\mathrm{S}_w(\mathrm{X})$ in the *D*-space are defined as follows:

$$\begin{aligned} \mathrm{S}_b(\mathrm{X}) &= \mathrm{s} \\ &= \sum_{i=1}^{c} P_i [(\bar{\mathrm{A}}_i - \bar{\mathrm{A}})\mathrm{X}][(\bar{\mathrm{A}}_i - \bar{\mathrm{A}})\mathrm{X}]^T \\ &= \sum_{i=1}^{c} P_i [(\bar{\mathrm{A}}_i - \bar{\mathrm{A}})\mathrm{X}\mathrm{X}^T (\bar{\mathrm{A}}_i - \bar{\mathrm{A}})^T] \end{aligned} \tag{8}$$

$$\begin{aligned} \mathrm{S}_w(\mathrm{X}) &= \sum_{i=1}^{c} P_i \left(\frac{1}{M_i - 1} \sum_{j=1}^{M_i} (\mathrm{D}_{ij} - \bar{\mathrm{D}}_i)(\mathrm{D}_{ij} - \bar{\mathrm{D}}_i)^T \right) \\ &= \sum_{i=1}^{c} P_i \frac{1}{M_i - 1} \sum_{j=1}^{M_i} [(\mathrm{A}_{ij} - \bar{\mathrm{A}}_i)\mathrm{X}][(\mathrm{A}_{ij} - \bar{\mathrm{A}}_i)\mathrm{X}]^T \\ &= \sum_{i=1}^{c} P_i \frac{1}{M_i - 1} \sum_{j=1}^{M_i} [(\mathrm{A}_{ij} - \bar{\mathrm{A}}_i)\mathrm{X}\mathrm{X}^T (\mathrm{A}_{ij} - \bar{\mathrm{A}}_i)^T] \end{aligned} \tag{9}$$

where P_i is the prior probability for Class *i* and commonly evaluated as $P_i = M_i / M$. Since the combination coefficient vector **X** is an unknown variable, the elements in $\mathrm{S}_b(\mathrm{X})$ and $\mathrm{S}_w(\mathrm{X})$ can be viewed as linear functionals of **X**.

The general Fisher criterion in the *D*-space can be defined as follows:

$$J(\boldsymbol{\varphi}, \mathbf{X}) = \frac{\boldsymbol{\varphi}^T \mathbf{S}_b(\mathbf{X}) \boldsymbol{\varphi}}{\boldsymbol{\varphi}^T \mathbf{S}_w(\mathbf{X}) \boldsymbol{\varphi}}, \tag{10}$$

where $\boldsymbol{\varphi}$ is a discriminant projection basis vector and **X** a color component combination coefficient vector.

Maximizing this criterion is equivalent to solving the following optimization model:

$$\begin{cases} \max\limits_{\boldsymbol{\varphi}, \mathbf{X}} \mathrm{tr}\{\boldsymbol{\varphi}^T \mathbf{S}_b(\mathbf{X}) \boldsymbol{\varphi}\} \\ \text{subject to } \boldsymbol{\varphi}^T \mathbf{S}_w(\mathbf{X}) \boldsymbol{\varphi} = 1 \end{cases}, \tag{11}$$

where $\mathrm{tr}(\cdot)$ is the trace operator. We will design an iterative algorithm to simultaneously determine the optimal discriminant projection basis vector $\boldsymbol{\varphi}^*$ and the optimal combination coefficient vector **X*** in the following subsection.

2.3 CID algorithm

First of all, let us define the color-space between-class scatter matrix $\mathbf{L}_b(\varphi)$ and the color-space within-class scatter matrix $\mathbf{L}_w(\varphi)$ as follows

$$\mathbf{L}_b(\varphi) = \sum_{i=1}^{c} P_i[(\overline{\mathbf{A}}_i - \overline{\mathbf{A}})^T \varphi\varphi^T (\overline{\mathbf{A}}_i - \overline{\mathbf{A}}), \tag{12}$$

$$\mathbf{L}_w(\varphi) = \sum_{i=1}^{c} P_i \frac{1}{M_i - 1} \sum_{j=1}^{M_i} [(\mathbf{A}_{ij} - \overline{\mathbf{A}}_i)^T \varphi\varphi^T (\mathbf{A}_{ij} - \overline{\mathbf{A}}_i)]. \tag{13}$$

$\mathbf{L}_b(\varphi)$ and $\mathbf{L}_w(\varphi)$ are therefore 3×3 non-negative definite matrices. Actually, $\mathbf{L}_b(\varphi)$ and $\mathbf{L}_w(\varphi)$ can be viewed as dual matrices of $\mathbf{S}_b(\mathbf{X})$ and $\mathbf{S}_w(\mathbf{X})$.
Based on the definition of $\mathbf{L}_b(\varphi)$ and $\mathbf{L}_w(\varphi)$, we give the following proposition:

Proposition 1: $\varphi^T \mathbf{S}_b(\mathbf{X})\varphi = \mathbf{X}^T \mathbf{L}_b(\varphi)\mathbf{X}$, and $\varphi^T \mathbf{S}_w(\mathbf{X})\varphi = \mathbf{X}^T \mathbf{L}_w(\varphi)\mathbf{X}$.

Proof:

$$\varphi^T \mathbf{S}_b(\mathbf{X})\varphi = \sum_{i=1}^{c} P_i[\varphi^T (\overline{\mathbf{A}}_i - \overline{\mathbf{A}})\mathbf{X}][\mathbf{X}^T (\overline{\mathbf{A}}_i - \overline{\mathbf{A}})^T \varphi]$$

$$= \sum_{i=1}^{c} P_i[\mathbf{X}^T (\overline{\mathbf{A}}_i - \overline{\mathbf{A}})^T \varphi][\varphi^T (\overline{\mathbf{A}}_i - \overline{\mathbf{A}})\mathbf{X}]$$

$$= \mathbf{X}^T \{\sum_{i=1}^{c} P_i[(\overline{\mathbf{A}}_i - \overline{\mathbf{A}})^T \varphi\varphi^T (\overline{\mathbf{A}}_i - \overline{\mathbf{A}})\}\mathbf{X}$$

$$= \mathbf{X}^T \mathbf{L}_b(\varphi)\mathbf{X}.$$

Similarly, we can derive that $\varphi^T \mathbf{S}_w(\mathbf{X})\varphi = \mathbf{X}^T \mathbf{L}_w(\varphi)\mathbf{X}$. □

The model in Eq. (11) is a constrained optimization problem, which can be solved using the Lagrange multiplier method. Let the Lagrange functional be as follows:

$$L(\varphi, \mathbf{X}, \lambda) = \varphi^T \mathbf{S}_b(\mathbf{X})\varphi - \lambda(\varphi^T \mathbf{S}_w(\mathbf{X})\varphi - 1), \tag{14}$$

where λ is the Lagrange multiplier. From Proposition 1, we have

$$L(\varphi, \mathbf{X}, \lambda) = \mathbf{X}^T \mathbf{L}_b(\varphi)\mathbf{X} - \lambda(\mathbf{X}^T \mathbf{L}_w(\varphi)\mathbf{X} - 1), \tag{15}$$

First, take the derivative of $L(\varphi, \mathbf{X}, \lambda)$ in Eq. (14) with respect to φ:

$$\frac{\partial L(\varphi, \mathbf{X}, \lambda)}{\partial \varphi} = 2\mathbf{S}_b(\mathbf{X})\varphi - 2\lambda \mathbf{S}_w(\mathbf{X})\varphi \tag{16}$$

Equate the derivative to zero, $\frac{\partial L(\varphi, \mathbf{X}, \lambda)}{\partial \varphi} = 0$, then we have the following equation:

$$\mathbf{S}_b(\mathbf{X})\boldsymbol{\varphi} = \lambda \mathbf{S}_w(\mathbf{X})\boldsymbol{\varphi} \tag{17}$$

Second, take the derivative of $L(\boldsymbol{\varphi}, \mathbf{X}, \lambda)$ in Eq. (15) with respect to **X**:

$$\frac{\partial L(\boldsymbol{\varphi}, \mathbf{X}, \lambda)}{\partial \mathbf{X}} = 2\mathbf{S}_b(\boldsymbol{\varphi})\mathbf{X} - 2\lambda \mathbf{S}_w(\boldsymbol{\varphi})\mathbf{X} \tag{18}$$

Equate the derivative to zero, $\frac{\partial L(\boldsymbol{\varphi}, \mathbf{X}, \lambda)}{\partial \mathbf{X}} = 0$, then we have the following equation:

$$\mathbf{L}_b(\boldsymbol{\varphi})\mathbf{X} = \lambda \mathbf{L}_w(\boldsymbol{\varphi})\mathbf{X} \tag{19}$$

And finally, take the derivative of $L(\boldsymbol{\varphi}, \mathbf{X}, \lambda)$ in Eq. (14) with respect to λ and equate it to zero, and we have the following equation:

$$\boldsymbol{\varphi}^T \mathbf{S}_w(\mathbf{X})\boldsymbol{\varphi} = 1, \tag{20}$$

which is equivalent to

$$\mathbf{X}^T \mathbf{L}_w(\boldsymbol{\varphi})\mathbf{X} = 1. \tag{21}$$

Therefore, finding the optimal solutions $\boldsymbol{\varphi}^*$ and **X*** of the optimization problem in Eq. (11) is equivalent to solving the following two sets of equations:

$$\textit{Equation Set I}: \quad \begin{cases} \mathbf{S}_b(\mathbf{X})\boldsymbol{\varphi} = \lambda \mathbf{S}_w(\mathbf{X})\boldsymbol{\varphi} \\ \boldsymbol{\varphi}^T \mathbf{S}_w(\mathbf{X})\boldsymbol{\varphi} = 1 \end{cases} \tag{22}$$

$$\textit{Equation Set II}: \quad \begin{cases} \mathbf{L}_b(\boldsymbol{\varphi})\mathbf{X} = \lambda \mathbf{L}_w(\boldsymbol{\varphi})\mathbf{X} \\ \mathbf{X}^T \mathbf{L}_w(\boldsymbol{\varphi})\mathbf{X} = 1 \end{cases} \tag{23}$$

Theorem 1 [27] Suppose that **A** and **B** are two $n \times n$ nonnegative definite matrices and **B** is nonsingular. There exist *n* eigenvectors $\xi_1, \cdots, \xi_n$ corresponding to eigenvalues $\lambda_1, \cdots, \lambda_n$ of the generalized eigen-equation $\mathrm{A}\xi = \lambda \mathrm{B}\xi$, such that

$$\boldsymbol{\xi}_i^T \mathbf{A} \boldsymbol{\xi}_j = \begin{cases} \lambda_i & i = j \\ 0 & i \neq j \end{cases} \quad i, j = 1, \cdots, n \tag{24}$$

and

$$\boldsymbol{\xi}_i^T \mathbf{B} \boldsymbol{\xi}_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \quad i, j = 1, \cdots, n. \tag{25}$$

From Theorem 1, we know the solution of *Equation Set I*, i.e., the extremum point $\boldsymbol{\varphi}^*$ of $J_F(\boldsymbol{\varphi}, \mathbf{X})$, can be chosen as the eigenvector of the generalized equation $\mathbf{S}_b(\mathbf{X})\boldsymbol{\varphi} = \lambda \mathbf{S}_w(\mathbf{X})\boldsymbol{\varphi}$ corresponding to the largest eigenvalue, and the solution of *Equation*

Set II, i.e., the extremum point $\mathbf{X}^*$ of $J_F(\boldsymbol{\varphi}, \mathbf{X})$, can be chosen as the eigenvector of the generalized equation $\mathbf{L}_b(\boldsymbol{\varphi})\mathbf{X} = \lambda \mathbf{L}_w(\boldsymbol{\varphi})\mathbf{X}$ corresponding to the largest eigenvalue. Based on this conclusion, we can design an iterative algorithm to calculate the extremum points $\boldsymbol{\varphi}^*$ and $\mathbf{X}^*$.

Let $\mathbf{X} = \mathbf{X}^{[k]}$ be the initial value of the combination coefficient vector in the *k*-th iteration. In the first step, we construct $\mathbf{S}_b(\mathbf{X})$ and $\mathbf{S}_w(\mathbf{X})$ based on $\mathbf{X} = \mathbf{X}^{[k]}$ and calculate their generalized eigenvector $\boldsymbol{\varphi} = \boldsymbol{\varphi}^{[k+1]}$ corresponding to the largest eigenvalue. In the second step, we construct $\mathbf{L}_b(\boldsymbol{\varphi})$ and $\mathbf{L}_w(\boldsymbol{\varphi})$ based on $\boldsymbol{\varphi} = \boldsymbol{\varphi}^{[k+1]}$ and calculate their generalized eigenvector $\mathbf{X}^{[k+1]}$ corresponding to the largest eigenvalue. $\mathbf{X} = \mathbf{X}^{[k+1]}$ is used as initial value in the next iteration.

The CID algorithm performs the preceding two steps successively until it converges. Convergence may be determined by observing when the value of the criterion function $J(\boldsymbol{\varphi}, \mathbf{X})$ stops changing. Specifically, after k+1 times of iterations, if $| J(\boldsymbol{\varphi}^{[k+1]}, \mathbf{X}^{[k+1]}) - J(\boldsymbol{\varphi}^{[k]}, \mathbf{X}^{[k]}) | < \varepsilon$, we think the algorithm converges. Then, we choose $\boldsymbol{\varphi}^* = \boldsymbol{\varphi}^{[k+1]}$ and $\mathbf{X}^* = \mathbf{X}^{[k+1]}$. The CID algorithm is illustrated in Figure 1.

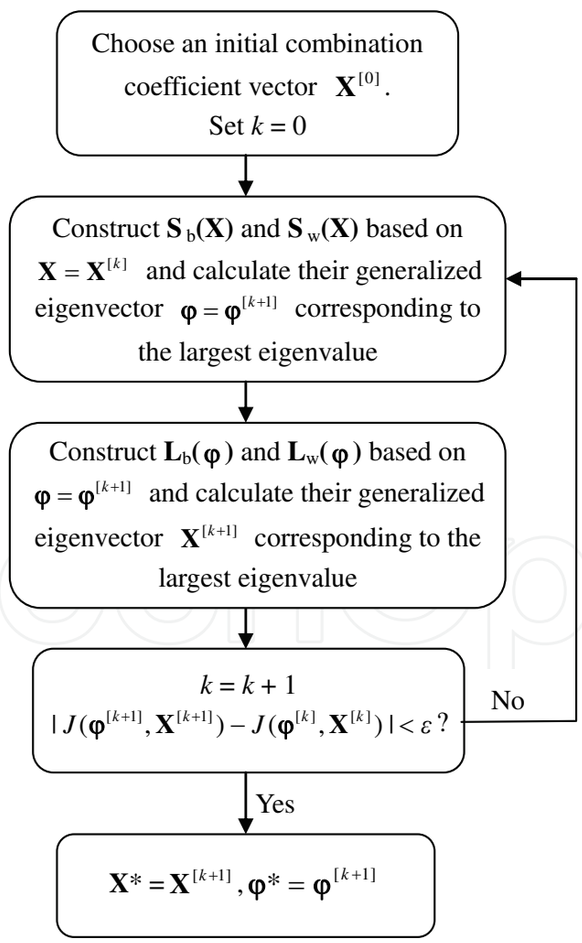


Fig. 1. An overview of the CID Algorithm

2.4 Extended CID algorithm for multiple discriminating color components

Using the CID algorithm, we obtain an optimal color component combination coefficient vector $\mathbf{X}^*=[x_{11}, x_{21}, x_{31}]^T$, which determines one discriminating color component $\mathrm{D}^1 = x_{11}\mathrm{R} + x_{21}\mathrm{G} + x_{31}\mathrm{B}$. In general, one discriminating color component is not enough for the discrimination of color images. Actually, analogous to the three color components in the RGB color space, we can derive three discriminating color components for image recognition. Let us denote the three discriminating color components of the color image $\mathrm{A} = [\mathrm{R}, \mathrm{G}, \mathrm{B}]$ as follows:

$$\mathbf{D}^i = x_{1i}\mathbf{R} + x_{2i}\mathbf{G} + x_{3i}\mathbf{B} = [\mathbf{R}, \mathbf{G}, \mathbf{B}]\,\mathbf{X}_i,\ i = 1, 2, 3 \tag{26}$$

where $\mathbf{X}_i$ $(i = 1, 2, 3)$ are the corresponding combination coefficient vectors. These coefficient vectors are required to be $\mathbf{L}_w(\varphi)$-orthogonal[1], that is

$$\mathbf{X}_i^T \mathbf{L}_w(\varphi)\mathbf{X}_j = 0, \quad \forall\, i \neq j, \quad i, j = 1, 2, 3. \tag{27}$$

Let the first combination coefficient vector be $\mathbf{X}_1 = \mathbf{X}^*$ and $\varphi = \varphi^*$, which have been determined in the foregoing subsection. Since the second combination coefficient vector is assumed to be $\mathbf{L}_w(\varphi)$-orthogonal to the first one, we can choose it from the $\mathbf{L}_w(\varphi)$-orthogonal complementary space of $\mathbf{X}_1$. We know $\mathbf{X}_1$ is chosen as the generalized eigenvector $\mathbf{u}_1$ of $\mathbf{L}_b(\varphi)$ and $\mathbf{L}_w(\varphi)$ corresponding to the largest eigenvalue after the CID algorithm converges. Let us derive $\mathbf{L}_b(\varphi)$ and $\mathbf{L}_w(\varphi)$'s remaining two generalized eigenvectors $\mathbf{u}_2$ and $\mathbf{u}_3$, which are $\mathbf{L}_w(\varphi)$-orthogonal to the first one. We choose the second combination coefficient vector $\mathbf{X}_2 = \mathbf{u}_2$ and the third combination coefficient vector $\mathbf{X}_3 = \mathbf{u}_3$.

After calculating three color component combination coefficient vectors $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$, we can obtain the three discriminating color components of the color image **A** using Eq. (26). In order to further improve the performance of the three discriminating color components, we generally center $\mathbf{X}_2$ and $\mathbf{X}_3$ in advance so that each of them has zero mean.

3. Experiments

This section assesses the performance of the proposed models and algorithms using a large scale color image database: the Face Recognition Grand Challenge (FRGC) version 2 database [25, 26]. This database contains 12,776 training images, 16,028 controlled target images, and 8,014 uncontrolled query images for the FRGC Experiment 4. The controlled images have good image quality, while the uncontrolled images display poor image quality, such as large illumination variations, low resolution of the face region, and possible blurring. It is these uncontrolled factors that pose the grand challenge to face recognition performance. The Biometric Experimentation Environment (BEE) system [25] provides a computational experimental environment to support a challenge problem in face

[1] This conjugate orthogonality requirement is to eliminate the correlations between combination coefficient vectors. The justification for this is given in Ref. [36].

recognition, and it allows the description and distribution of experiments in a common format. The BEE system uses the PCA method that has been optimized for large scale problems as a baseline algorithm, and it applies the whitened cosine similarity measure. The BEE baseline algorithm shows that FRGC Experiment 4, which is designed for indoor controlled single still image versus uncontrolled single still image, is the most challenging FRGC experiment. We therefore choose the FRGC Experiment 4 to evaluate our method. In our experiments, the face region of each image is first cropped from the original high-resolution still images and resized to a spatial resolution of 32×32. Figure 2 shows some example FRGC images used in our experiments.

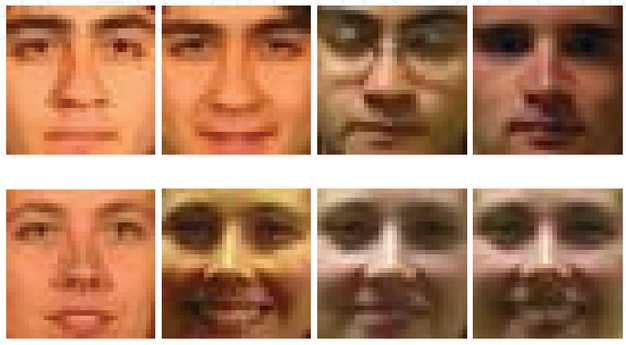


Fig. 2. Example FRGC images that have been cropped to 32x32.

According to the FRGC protocol, the face recognition performance is reported using the Receiver Operating Characteristic (ROC) curves, which plot the Face Verification Rate (FVR) versus the False Accept Rate (FAR). The ROC curves are automatically generated by the BEE system when a similarity matrix is input to the system. In particular, the BEE system generates three ROC curves, ROC I, ROC II, and ROC III, corresponding to images collected within semesters, within a year, and between semesters, respectively. The similarity matrix stores the similarity score of every query image versus target image pair. As a result, the size of the similarity matrix is $T \times Q$, where T is the number of target images (16,028 for FRGC version 2 Experiment 4) and Q is the number of query images (8,014 for FRGC version 2 Experiment 4).

3.1 Face recognition based on one color component image

Following the FRGC protocol, we use the standard training set of the FRGC version 2 Experiment 4 for training. The initial value of the CID algorithm is set as $\mathbf{X}^{[0]} = [1/3, 1/3, 1/3]$, and the convergence threshold of the algorithm is set to be $\varepsilon = 0.1$. After training, the CID algorithm generates one optimal color component combination coefficient vector $\mathbf{X}_1 = [x_{11}, x_{21}, x_{31}]^T$ and a set of 220 optimal discriminant basis vectors since there are 222 pattern classes. The combination coefficient vector $\mathbf{X}_1$ determines one discriminating color component $\mathbf{D}^1 = x_{11}\mathbf{R} + x_{21}\mathbf{G} + x_{31}\mathbf{B}$ for color image representation and the set of discriminant basis vectors determines the projection matrix for feature extraction. In comparison, we also implement the FLD algorithm on grayscale images and choose 220 discriminant features.

For each method mentioned, the cosine measure [33] is used to generate the similarity matrix. After score normalization using Z-score [34], the similarity matrix is analyzed by the

BEE system. The three ROC curves generated by BEE are shown in Figure 3 and the resulting face verification rates at the false accept rate of 0.1% are listed in Table 1. The performance of the BEE baseline algorithm is also shown in Figure 3 and Table 1 for comparison. Figure 3 and Table 1 show that the proposed CID algorithm achieves better performance than the classical FLD method using grayscale images. In particular, the CID algorithm achieves a verification rate of 61.01% for ROC III, which is a nearly 10% increase compared with the FLD method using the grayscale images.

Method	ROC I	ROC II	ROC III
BEE Baseline	13.36	12.67	11.86
FLD on grayscale images	52.96	52.34	51.57
CID	**60.49**	**60.75**	**61.01**

Table 1. Verification rate (%) comparison when the false accept rate is 0.1%

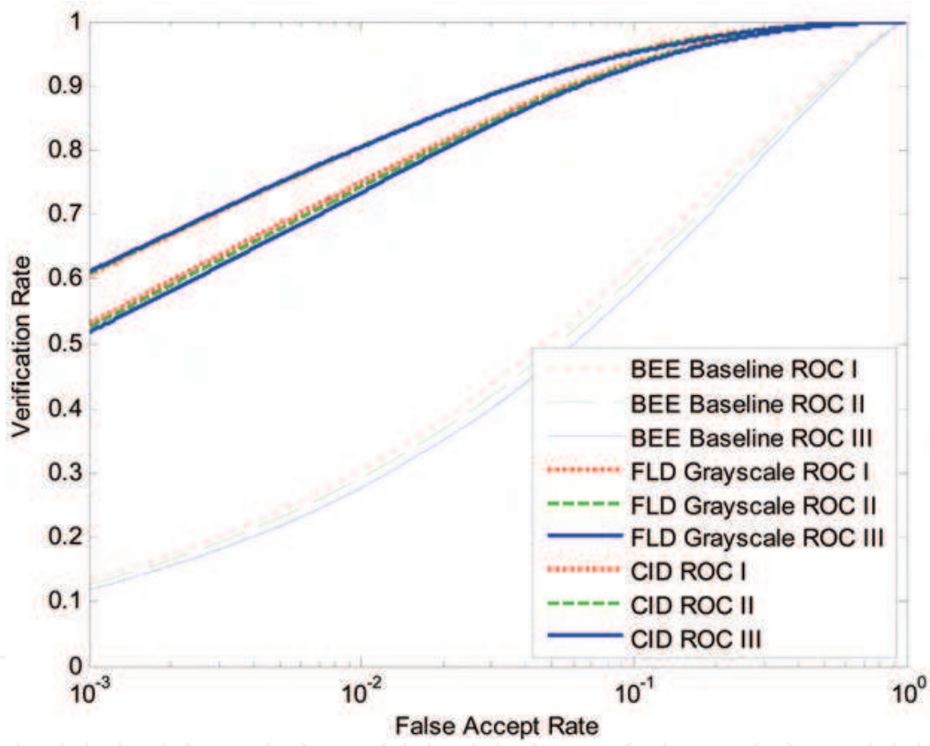


Fig. 3. ROC curves corresponding to the BEE Baseline algorithm, FLD on grayscale images and the CID Algorithm

It should be pointed out that the convergence of the CID algorithm does not depend on the choice of the initial value of $\mathbf{X}^{[0]}$. We randomly generate four set of initial values (four three-dimensional vectors). The convergence of the CID algorithm corresponding to these four set of initial values and the originally chosen initial values $\mathbf{X}^{[0]} = [1/3, 1/3, 1/3]$ is illustrated in Figure 4. Figure 4 shows that the convergence of the CID algorithm is independent of the choice of initial value of $\mathbf{X}^{[0]}$. The algorithm consistently converges to a very similar value of the criterion function $J(\varphi, \mathbf{X})$, and its convergence speed is fast: it always converges within 10 iterations if we choose $\varepsilon = 0.1$.

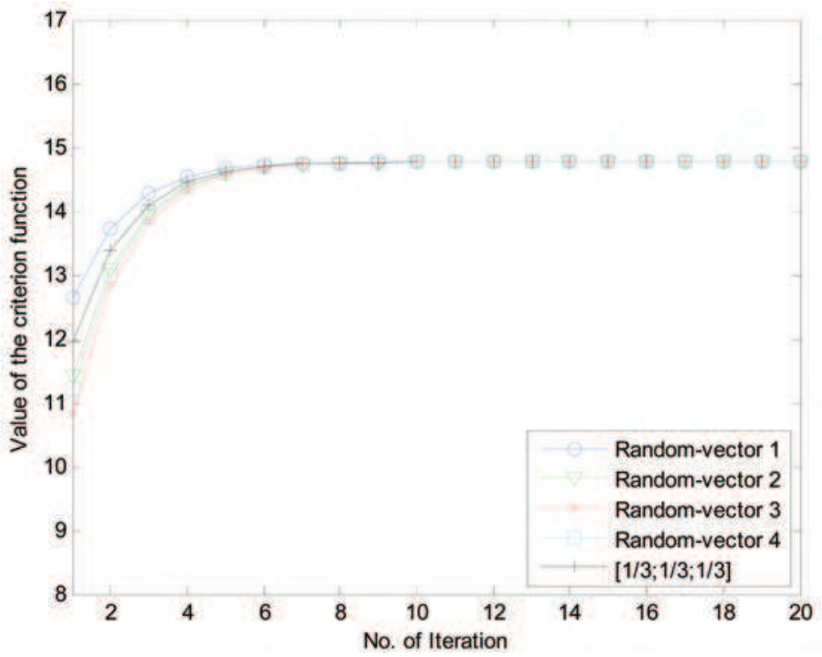


Fig. 4. Illustration of the convergence of the CID algorithm

3.2 Face recognition based on three color component images

In this experiment, we train the extended CID algorithm using the standard training set of the FRGC version 2 Experiment 4 to generate three color component combination coefficient vectors $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$, and based on these coefficient vectors we obtain three discriminating color components $\mathbf{D}^1$, $\mathbf{D}^2$ and $\mathbf{D}^3$ for each color image. The three discriminating color component images corresponding to one orginal image are shown in Figure 5.

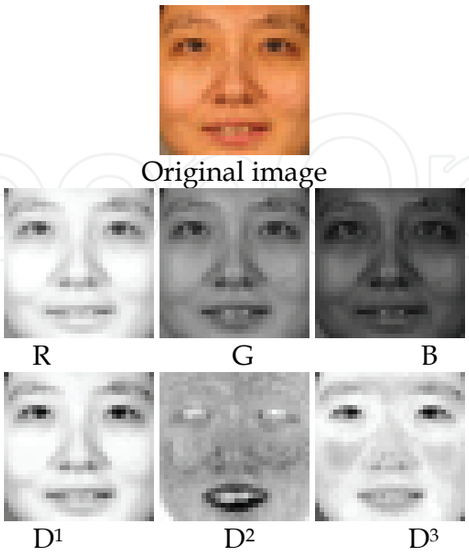
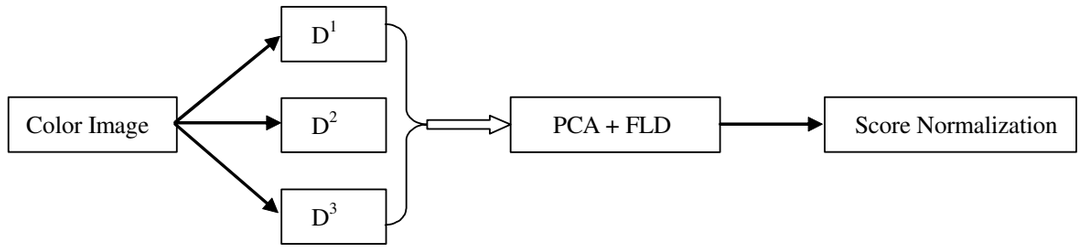
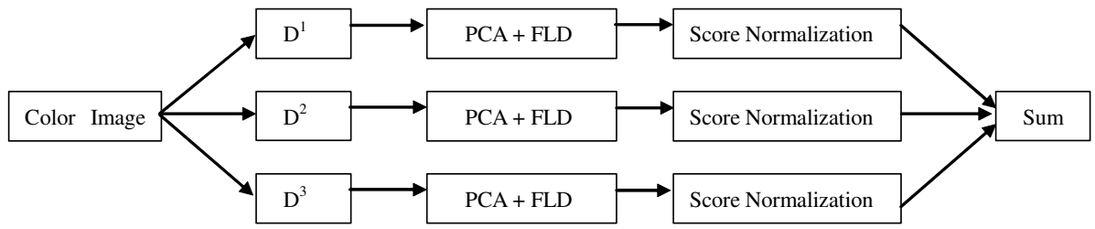


Fig. 5. Illustration of R, G, B color component images and the three color component images generated by the proposed method

We employ two fusion strategies, i.e., decision-level fusion and image-level fusion, to combine the information within the three discriminating color component images for recognition purpose. The decision-level fusion strategy first extracts discriminant features from each of the three color component images, then calculates the similarity scores and normalizes them using Z-score, and finally fuses the normalized similarity scores using a sum rule. The image-level fusion strategy first concatenates the three color components $\mathbf{D}^1$, $\mathbf{D}^2$ and $\mathbf{D}^3$ into one pattern vector and then performs PCA+FLD [23] on the concatenated pattern vector. To avoid the negative effect of magnitude dominance of one component image over the others, we apply a basic image normalization method by removing the mean and normalizing the standard deviation of each component image before the concatenation. To avoid overfitting, we choose 900 principal components (PCs) in the PCA step of the decision-level fusion strategy and 1000 PCs in the PCA step of the image-level fusion strategy. The frameworks of the two fusion strategies are shown in Figure 6.



(a) Outline of the image-level fusion



(b) Outline of the decision-level fusion

Fig. 6. Illustration of the decision-level and image-level fusion strategy frameworks

For comparison, we apply the same two fusion strategies to the R, G and B color component images and obtain the corresponding similarity scores. The final similarity matrix is input to the BEE system and three ROC curves are generated. Figure 7 shows the three ROC curves corresponding to each of three methods: the BEE Baseline algorithm, FLD using RGB images, and the extended CID algorithm using the decision-level fusion strategy. Figure 8 shows the ROC curves of the three methods using the image-level fusion strategy. Table 2 lists the face verification rates at the false accept rate of 0.1%. These results indicate that the fusion of the three discriminant color components generated by the extended CID algorithm is more effective for improving the FRGC performance than the fusion of the original R, G and B color components, no matter what fusion strategy is used.

In addition, by comparing the results of the two fusion strategies shown in Table 2, one can see that the three color components generated by the extended CID algorithm demonstrates quite stable face recognition performance while the R, G and B color components do not. For the three color components generated by the extended CID algorithm, the performance

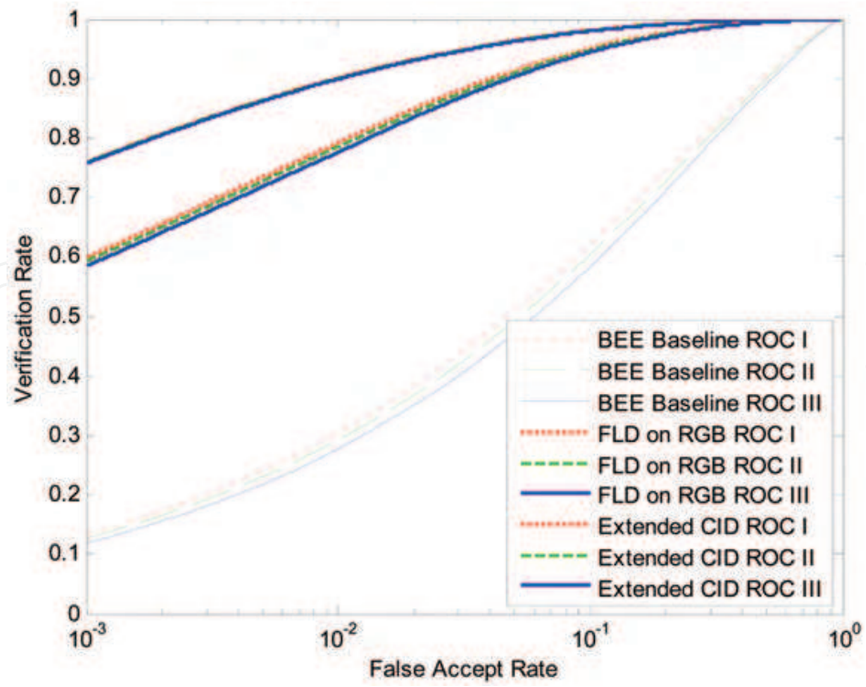


Fig. 7. ROC curves corresponding to the BEE baseline algorithm, FLD using the RGB images, and the extended CID algorithm (for three color components) using the decision-level fusion strategy

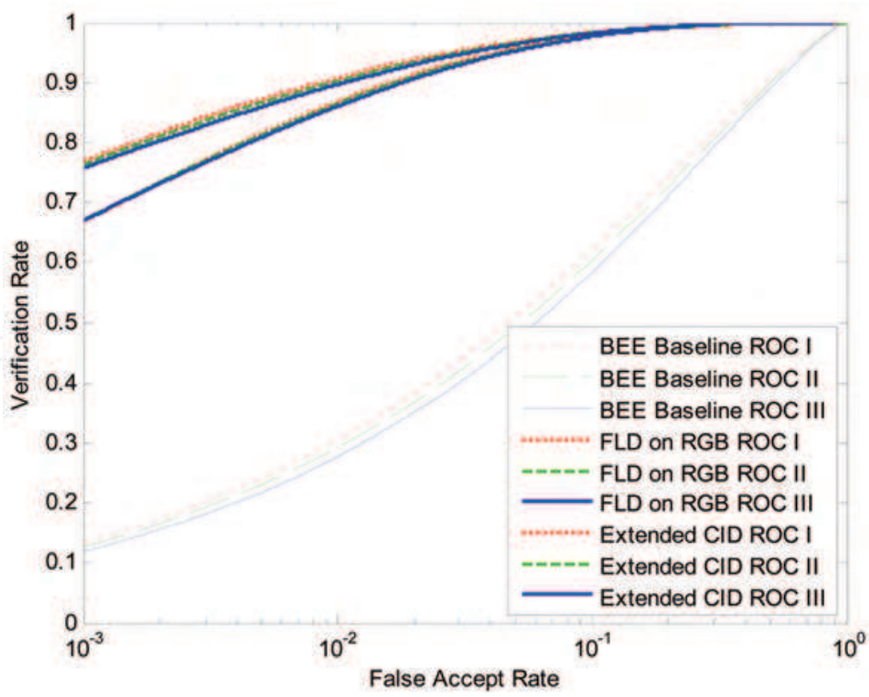


Fig. 8. ROC curves corresponding to the BEE baseline algorithm, FLD using the RGB images, and the extended CID algorithm (for three color components) using the image-level fusion strategy

difference between the two fusion strategies is at most 1.01%, whereas for the R, G and B color components, the performance difference between the two fusion strategies is as large as 8.55%. The RGB color space performs much worse when the decision-level fusion strategy is used.

Fusion strategy	Method	ROC I	ROC II	ROC III
Decision-level fusion	FLD on RGB images	59.75	59.14	58.34
	Extended CID	**75.73**	**75.74**	**75.66**
Image-level fusion	FLD on RGB images	66.68	66.85	66.89
	Extended CID	**76.72**	**76.25**	**75.64**

Table 2. Verification rate (%) comparison when the false accept rate is 0.1% using all of the three color components images

4. Conclusions

This chapter seeks to find a meaningful representation and an effective recognition method of color images in a unified framework. We integrate color image representation and recognition tasks into one discriminant model: color image discriminant (CID) model. The model therefore involve two sets of variables: a set of color component combination coefficients for color image representation and a set of projection basis vectors for color image discrimination. An iterative CID algorithm is developed to find the optimal solution of the proposed model. The CID algorithm is further extended to generate three color components (like the three color components of RGB color images) for further improving the recognition performance. Three experiments using the Face Recognition Grand Challenge (FRGC) database and the Biometric Experimentation Environment (BEE) system demonstrate the performance advantages of the proposed method over the Fisher linear discriminant analysis method on grayscale and RGB color images.

5. Acknowledgments

This work was partially supported by the National Science Foundation of China under Grants No. 60503026, No. 60632050, and the 863 Hi-Tech Program of China under Grant No. 2006AA01Z119. Dr. Chengjun Liu was partially supported by Award No. 2006-IJ-CX-K033 awarded by the National Institute of Justice, Office of Justice Programs, US Department of Justice.

6. References

[1] J. Luo, D. Crandall, "Color object detection using spatial-color joint probability functions" *IEEE Transactions on Image Processing*, June 2006, 15(6), pp. 1443 – 1453

[2] R. L. Hsu, M. Abdel-Mottaleb, and A.K. Jain, "Face detection in color images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, no. 5, pp. 696–706, 2002.

[3] O. Ikeda, "Segmentation of faces in video footage using HSV color for face detection and image retrieval", *International Conference on Image Processing* (ICIP 2003), 2003.

[4] Y. Wu; T.S. Huang, "Nonstationary color tracking for vision-based human-computer interaction", *IEEE Transactions on Neural Networks*, July 2002, 13(4), pp. 948 – 960.

[5] T. Gevers, H. Stokman, "Robust histogram construction from color invariants for object recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Jan. 2004, 26(1), pp. 113 – 118

[6] A. Diplaros, T. Gevers, and I. Patras, "Combining color and shape information for illumination-viewpoint invariant object recognition", *IEEE Transactions on Image Processing*, Jan. 2006, 15(1), pp. 1 – 11

[7] G. Dong, M. Xie, "Color clustering and learning for image segmentation based on neural networks", *IEEE Transactions on Neural Networks*, July 2005, 16(4), pp. 925 – 936

[8] H. Y. Lee; H. K. Lee; Y. H. Ha, "Spatial color descriptor for image retrieval and video segmentation", *IEEE Transactions on Multimedia*, Sept. 2003, 5(3), pp. 358 – 367

[9] A. W. M. Smeulders, M. Worring, S Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Anal. Machine Intell.*, 2000, 22(12), pp. 1349–1380.

[10] M. J. Swain and D.H. Ballard, "Color indexing," *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11–32, 1991.

[11] B. V. Funt, G.D. Finlayson, "Color constant color indexing", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, May 1995, 17(5), Page(s):522 - 529

[12] D. A. Adjeroh, M. C. Lee, "On ratio-based color indexing", *IEEE Transactions on Image Processing*, Jan. 2001, 10(1), Page(s): 36 – 48.

[13] G. Healey and D. A. Slater, "Global color constancy: Recognition of objects by use of illumination invariant properties of color distributions," *Journal of the Optical Society of America A*, vol. 11, no. 11, pp. 3003–3010, 1994.

[14] G. D. Finlayson, S. D. Hordley, and P.M. Hubel, "Color by correlation: A simple, unifying framework for color constancy," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, no. 11, pp. 1209–1221, 2001.

[15] H. Stokman; T. Gevers, "Selection and Fusion of Color Models for Image Feature Detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, March 2007, 29(3), Page(s): 371 – 381

[16] 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*, 25, 37-52. 1996.

[17] L. Torres, J.Y. Reutter, L. Lorente, "The importance of the color information in face recognition", *International Conference on Image Processing* (ICIP 99), Oct. 1999, Volume 3, Page(s): 627 - 631

[18] A. Yip and P. Sinha, "Role of color in face recognition," *MIT tech report* (ai.mit.com) AIM-2001-035 CBCL-212, 2001.

[19] M. Rajapakse, J. Tan, J. Rajapakse, "Color channel encoding with NMF for face recognition", *International Conference on Image Processing* (ICIP '04), Oct. 2004, Volume 3, Page(s):2007- 2010.

[20] C. Xie, B. V. K. Kumar, "Quaternion correlation filters for color face recognition", *Proceedings of the SPIE*, Volume 5681, pp. 486-494 (2005).

[21] P. Shih and C. Liu, "Improving the Face Recognition Grand Challenge Baseline Performance Using Color Configurations Across Color Spaces", *IEEE International Conference on Image Processing*, *ICIP 2006*, 2006, October 8-11, Atlanta, GA.

[22] C. Jones III, A. L. Abbott, "Color face recognition by hypercomplex gabor analysis", *7th International Conference on Automatic Face and Gesture Recognition* (FGR 2006), April, 2006.

[23] C. Liu and H. Wechsler, "Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition", *IEEE Trans. Image Processing*, vol. 11, no. 4, pp. 467-476, 2002.

[24] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, second edition, 1990.

[25] 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," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2005.

[26] P. J. Phillips., P. J. Flynn, T.Scruggs, K.W. Bowyer, W. Worek, "Preliminary Face Recognition Grand Challenge Results", *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition* (FGR'06).

[27] P. Lancaster and M. Tismenetsky, The Theory of Matrices (Second Edition), Academic Press, INC. Orlando, Florida, 1985.

[28] D. L. Swets and J. Weng. "Using discriminant eigenfeatures for image retrieval", *IEEE Trans. Pattern Anal. Machine Intell.*, 1996,18(8), pp. 831-836.

[29] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriengman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection", *IEEE Trans. Pattern Anal. Machine Intell.* 1997, 19 (7), pp. 711-720.

[30] W. Zhao, A. Krishnaswamy, R. Chellappa, D. Swets, and J. Weng, "Discriminant analysis of principal components for face recognition", in *Face Recognition: From Theory to Applications*, Eds. H. Wechsler, P.J. Phillips, V. Bruce, F.F. Soulie and T.S. Huang, Springer-Verlag, pp. 73-85, 1998.

[31] J. Yang, J.Y. Yang, "Why can LDA be performed in PCA transformed space?" *Pattern Recognition*, 2003, 36(2), pp. 563-566.

[32] J. Yang, A. F. Frangi, J.-Y. Yang, D. Zhang, Z. Jin, "*KPCA Plus LDA: A Complete Kernel Fisher Discriminant Framework for Feature Extraction and Recognition", IEEE Trans. Pattern Anal. Machine Intell.,* 2005, 27(2), pp. 230-244.

[33] C. Liu, "Capitalize on Dimensionality Increasing Techniques for Improving Face Recognition Grand Challenge Performance", *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 28, no. 5, pp. 725-737, 2006.

[34] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodel biometric systems", *Pattern Recognition*, 38 (2005), 2270-2285.

[35] M. Turk and A. Pentland, "Eigenfaces for recognition", J. Cognitive Neuroscience, 1991, 3(1), pp. 71-86.

[36] Z. Jin, J.Y. Yang, Z.S. Hu, Z. Lou, "Face Recognition based on uncorrelated discriminant transformation", *Pattern Recognition*, 2001,33(7), 1405-1416.

[37] J. Ye, R. Janardan, and Q. Li. "Two-Dimensional Linear Discriminant Analysis", *Neural Information Processing Systems (NIPS 2004)*.

[38] J. Yang, C. Liu, "A General Discriminant Model for Color Face Recognition", *Eleventh IEEE International Conference on Computer Vision (ICCV 2007*), Rio de Janeiro, Brazil, October 14-20, 2007.

Recent Advances in Face Recognition

Edited by Kresimir Delac, Mislav Grgic and Marian Stewart Bartlett

ISBN 978-953-7619-34-3
Hard cover, 236 pages
Publisher InTech
Published online 01, June, 2008
Published in print edition June, 2008

The main idea and the driver of further research in the area of face recognition are security applications and human-computer interaction. Face recognition represents an intuitive and non-intrusive method of recognizing people and this is why it became one of three identification methods used in e-passports and a biometric of choice for many other security applications. This goal of this book is to provide the reader with the most up to date research performed in automatic face recognition. The chapters presented use innovative approaches to deal with a wide variety of unsolved issues.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Jian Yang, Chengjun Liu and Jingyu Yang (2008). Discriminating Color Faces For Recognition, Recent Advances in Face Recognition, Kresimir Delac, Mislav Grgic and Marian Stewart Bartlett (Ed.), ISBN: 978-953-7619-34-3, InTech, Available from: http://www.intechopen.com/books/recent_advances_in_face_recognition/discriminating_color_faces_for_recognition

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen