ABSTRACT

FACE RECOGNITION USING MULTIPLE FEATURES IN DIFFERENT COLOR SPACES

by
Zhiming Liu

Face recognition as a particular problem of pattern recognition has been attracting substantial attention from researchers in computer vision, pattern recognition, and machine learning. The recent Face Recognition Grand Challenge (FRGC) program reveals that uncontrolled illumination conditions pose grand challenges to face recognition performance. Most of the existing face recognition methods use gray-scale face images, which have been shown insufficient to tackle these challenges. To overcome this challenging problem in face recognition, this dissertation applies multiple features derived from the color images instead of the intensity images only.

First, this dissertation presents two face recognition methods, which operate in different color spaces, using frequency features by means of Discrete Fourier Transform (DFT) and spatial features by means of Local Binary Patterns (LBP), respectively. The DFT frequency domain consists of the real part, the imaginary part, the magnitude, and the phase components, which provide the different interpretations of the input face images. The advantage of LBP in face recognition is attributed to its robustness in terms of intensity-level monotonic transformation, as well as its operation in the various scale image spaces. By fusing the frequency components or the multi-resolution LBP histograms, the complementary feature sets can be generated to enhance the capability of facial texture description. This dissertation thus uses the fused DFT and LBP features in two hybrid color spaces, the RIQ and the VIQ color spaces, respectively, for improving face recognition performance.

Second, a method that extracts multiple features in the CID color space is presented for face recognition. As different color component images in the CID color space display different characteristics, three different image encoding methods, namely, the patch-based

Gabor image representation, the multi-resolution LBP feature fusion, and the DCT-based multiple face encodings, are presented to effectively extract features from the component images for enhancing pattern recognition performance. To further improve classification performance, the similarity scores due to the three color component images are fused for the final decision making.

Finally, a novel image representation is also discussed in this dissertation. Unlike a traditional intensity image that is directly derived from a linear combination of the R, G, and B color components, the novel image representation adapted to class separability is generated through a PCA plus FLD learning framework from the hybrid color space instead of the RGB color space. Based upon the novel image representation, a multiple feature fusion method is proposed to address the problem of face recognition under the severe illumination conditions.

The aforementioned methods have been evaluated using two large-scale databases, namely, the Face Recognition Grand Challenge (FRGC) version 2 database and the FERET face database. Experimental results have shown that the proposed methods improve face recognition performance upon the traditional methods using the intensity images by large margins and outperform some state-of-the-art methods.

FACE RECOGNITION USING MULTIPLE FEATURES IN DIFFERENT COLOR SPACES

by
Zhiming Liu

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Computer Science

Department of Computer Science

January 2011

APPROVAL PAGE

FACE RECOGNITION USING MULTIPLE FEATURES IN DIFFERENT COLOR SPACES

Zhiming Liu

Dr. Chengjun Liu, Dissertation Advisor — Date
Associate Professor of Computer Science, New Jersey Institute of Technology

---

Dr. James Geller, Committee Member — Date
Professor of Computer Science, New Jersey Institute of Technology

---

Dr. Andrew Sohn, Committee Member — Date
Associate Professor of Computer Science, New Jersey Institute of Technology

---

Dr. Usman W. Roshan, Committee Member — Date
Associate Professor of Computer Science, New Jersey Institute of Technology

---

Dr. Xiaoguang Lu, Committee Member — Date
Research Scientist, Siemens Corporate Research, Inc., Princeton, New Jersey

BIOGRAPHICAL SKETCH

Author: Zhiming Liu

Degree: Doctor of Philosophy

Date: January 2011

Undergraduate and Graduate Education:

- Doctor of Philosophy in Computer Science,
  New Jersey Institute of Technology, Newark, New Jersey, 2011
- Master of Science in Computer Engineering,
  University of Nevada, Reno, Nevada, 2006
- Master of Science in Electrical Engineering,
  Sichuan University, Chengdu, Sichuan, China, 2001
- Bachelor of Science in Electrical Engineering,
  Sichuan University, Chengdu, Sichuan, China, 1997

Major: Computer Science

Publications:

Z. Liu, J. Yang, and C. Liu, "Extracting Multiple Features in the CID Color Space for Face Recognition," *IEEE Transactions on Image Processing*, vol. 19, no. 9, pp. 2502-2509, 2010.

Z. Liu and C. Liu, "Fusion of Color, Local Spatial and Global Frequency Information for Face Recognition," *Pattern Recognition*, vol. 43, no. 8, pp. 2882-2890, 2010.

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.

Z. Liu and C. Liu, "Fusion of the Complementary Discrete Cosine Features in the YIQ Color Space for Face Recognition," *Computer Vision and Image Understanding*, vol. 111, no. 3, pp. 249-262, 2008.

Z. Liu, C. Liu, and Q. Tao, "Learning-based Image Representation and Method for Face Recognition," *IEEE Third International Conference on Biometrics: Theory, Applications and Systems (BTAS'09)*, Sept 28 - 30, 2009, Arlington, Virginia, USA.

Z. Liu and C. Liu, "Robust Face Recognition Using Color Information," *The 3rd IAPR/IEEE Conference on Biometrics (ICB'09)*, June 2 - 5, 2009, Italy.

Z. Liu and Q. Tao, "Face Recognition Using New Image Representations," *2009 International Joint Conference on Neural Networks (IJCNN'09)*, June 14 - 19, 2009, Atlanta, USA.

Z. Liu and C. Liu, "Fusing Frequency, Spatial and Color Features for Face Recognition," *IEEE Second International Conference on Biometrics: Theory, Applications and Systems (BTAS'08)*, Sept 29 - Oct 1, 2008, Arlington, Virginia, USA.

To My Beloved Parents

ACKNOWLEDGMENT

I owe deep gratitude to those who have made everything possible for me to complete this thesis. First and foremost, I would like to thank my advisor, Prof. Chengjun Liu for his support and encouragement, and for introducing me to the problem of color face recognition. Along with Prof. James Geller, Prof. Andrew Sohn, Prof. Usman W. Roshan at NJIT, and Dr. Xiaoguang Lu at Siemens Corporation, which were a great source of help and encouragement as my committee members. Special thanks also go to former lab members, Prof. Jian Yang at Nanjing University of Science and Technology in China, and Dr. Hui Kong at Ohio State University. I really have learned a lot from discussing research ideas with them.

Last, but not least, I would like to thank my friends at NJIT: Shuo Chen, Shengyan Gao, Jingjing Zhang, and Xinfa Hu for their generous help. Additionally, I want to thank Jichao Sun and Venkata Gopal Edupuganti for their dedication towards our collective tournament badminton.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

Chapter **Page**

TABLE OF CONTENTS
(Continued)

Chapter **Page**

LIST OF TABLES

LIST OF TABLES
(Continued)

Table **Page**

LIST OF FIGURES

LIST OF FIGURES
(Continued)

Figure **Page**

LIST OF FIGURES
(Continued)

Figure **Page**

CHAPTER 1

INTRODUCTION

Face recognition, a typical problem in computer vision, pattern recognition, and machine learning, has been attracting more and more research attention recently, due to the complexity of the problem itself and the enormous applications in the commercial and government sectors (Bowyer et al. 2006; Zhao et al. 2003; Jain et al. 2004). Although numerous methods in face recognition have been proposed in the last two decades, there are still many open problems that remain to be unsolved. The recent Face Recognition Grand Challenge (FRGC) program reveals that uncontrolled conditions, such as illumination, expression, blurring, and so on, pose grand challenges to face recognition performance (Phillips et al. 2005). Some of the typical image variations on face images are shown in Figure 1.1. Therefore, some new techniques are in great demand to achieve a breakthrough in solving these obstacles. In this chapter, some popular face recognition approaches will be briefly introduced first, and then the outline of the work in this dissertation will be presented.

1.1 Face Recognition Using Appearance-based Methods

Pattern recognition relies heavily on the particular choice of features utilized by the classifiers. Therefore, feature selection and feature extraction are crucial to many pattern classification problems, e.g., face recognition. One typical block diagram of face recognition system is shown in Figure 1.2. The common objective of feature selection and extraction is to map the original measurements into more effective features, which show significant differences from one class to another, so that the classifiers can be designed more easily with better performance (Fukunaga 1990). One of the dominant and most successful approaches in face recognition is appearance-based feature extraction. Figure 1.3 shows some major methods in this category, which consists of two aspects: statistics-level and

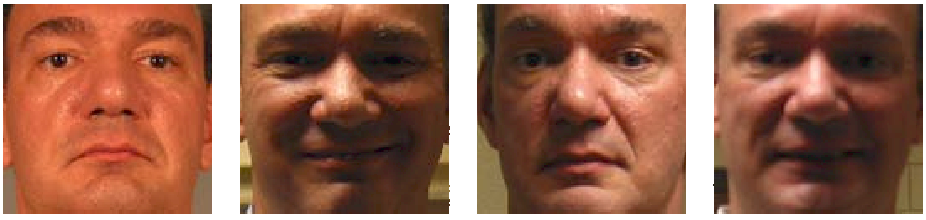

Figure 1.1 Examples of face image variations. Large image variations in illumination, expression, and blurring pose challenges to face recognition.

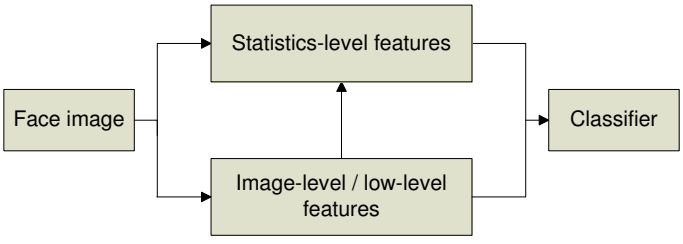

Figure 1.2 Block diagram of face recognition system.

image-level features. Another important category is model-based feature extraction, which utilizes the shape, texture, and 3D depth information for face recognition. More details about these methods can be found in a survey (Lu 2003). In this section, some state-of-the-art face recognition methods belonging to the first category are briefly presented, because the work in this dissertation mainly makes use of features derived from holistic and local face appearance.

1.1.1 Statistics-level Feature Extraction

The face images reside usually in a high-dimensional image space. There is a great demand to find the meaningful and compact patterns in such a space for developing robust face recognition methods so as to meet two requirements: enhanced discrimination ability and computational efficiency. Therefore, most appearance-based face recognition algo-

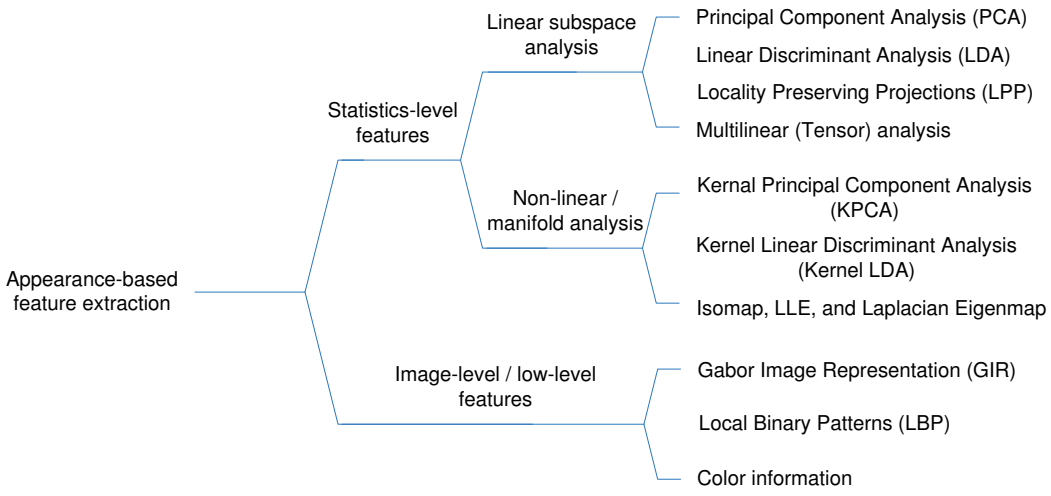

Figure 1.3 Hierarchical structure of the appearance-based feature extraction in face recognition.

rithms usually start with the dimensionality reduction by using some popular linear subspace methods. In the following sections, several major statistical methods are introduced.

Principal Component Analysis (PCA)

As an optimal linear transformation in the sense of minimum Mean Square Error (MSE), Principal Component Analysis (PCA) (Turk & Pentland 1991; Kirby & Sirovich 1990) has been a leading technique for dimensionality reduction of input data. Given a set of d-dimensional column image vectors $\{X_{ij}\}$, where $X_{ij} \in \mathbb{R}^d$ is the j-th image of class i. Let the training set consist of c persons and l_i sample images for person i. Thus, the number of training samples is $m = \sum_{i=1}^{c} l_i$. For face recognition, each person is a class with prior probability of λ_i. The within-class scatter matrix is defined as:

$$S_w = \sum_{i=1}^{c} \frac{\lambda_i}{l_i} \sum_{j=1}^{l_i} (X_{ij} - \overline{X_i})(X_{ij} - \overline{X_i})^T, \tag{1.1}$$

where $\overline{X_i} = \frac{1}{l_i}\sum_{j=1}^{l_i} X_{ij}$ is the mean of person i. The between-class scatter matrix S_b and the total (mixture) scatter matrix S_t are defined respectively as:

$$S_b = \sum_{i=1}^{c} \lambda_i (\overline{X_i} - \overline{X})(\overline{X_i} - \overline{X})^T, \tag{1.2}$$

$$S_t = \sum_{i=1}^{c} \frac{\lambda_i}{l_i} \sum_{j=1}^{l_i} (X_{ij} - \overline{X})(X_{ij} - \overline{X})^T, \tag{1.3}$$

where $\overline{X} = \frac{1}{m}\sum_{i=1}^{c}\sum_{j=1}^{l_i} X_{ij}$ is the grand mean.

PCA seeks a principal subspace of lower dimensionality to maximize the data reconstruction capability of the features. As a result, the features in this subspace can represent the original data accurately. The objective function of PCA can be defined as:

$$W^* = \underset{\|W\|=1}{\operatorname{argmax}} |W^T S_t W|. \tag{1.4}$$

Maximizing the above equation can be solved via eigenvalue-eigenvector analysis. That is, the matrix W^* can be constructed by obtaining the k principal eigenvectors corresponding to the k largest eigenvalues of S_t.

Linear Discriminant Analysis (LDA)

The best representation of data may not perform well from the classification point of view, as the total scatter matrix consists of both the within- and between-class variations. To obtain the discrimination of features for differentiating face images of one person from the others, one needs to manipulate the within- and between-class variations separately. To that end, face recognition using Linear Discriminant Analysis (LDA) (Swets & Weng 1996; Belhumeur et al. 1997; Etemad et al. 1997) has been an area of increasing interest. LDA is also known as Fisher Linear Discriminant (FLD). In this dissertation, the terms LDA and FLD are used interchangeably. The objective function of LDA can be defined as:

$$W^* = \underset{W}{\operatorname{argmax}} \frac{|W^T S_b W|}{|W^T S_w W|}. \tag{1.5}$$

Equation (1.5) is called the Fisher criterion. To maximize the ratio value of this criterion, LDA seeks an optimal subspace W^* that separates the different classes as far as possible and compresses the same classes as compactly as possible. To derive W^*, LDA solves the generalized eigenvectors of $S_bW = \lambda S_wW$, and chooses the k principal eigenvectors corresponding to the k largest eigenvalues.

Locality Preserving Projections (LPP)

While PCA and LDA aim to preserve the global Euclidean structure, the local manifold structure is more important in many real-world applications, especially when nearest-neighbor based classifiers are used for classification (He et al. 2005). Given the high-dimensional data lies on a low dimensional manifold embedded in the ambient space, a novel linear learning algorithm, called Locality Preserving Projections (LPP) (He et al. 2005), has been proposed to find an optimal linear transformation to preserve the local geometric structure of the face image space. The objective function of LPP is as follows (He et al. 2005):

$$\min \sum_{ij} \|y_i - y_j\|^2 S_{ij}, \tag{1.6}$$

where y_i is the one-dimensional representation of X_i by the linear transformation $y_i = W^T X_i$. The matrix S is a similarity matrix. A possible way to define S is given as (He et al. 2005):

$$S_{ij} = \begin{cases} \exp(-\|X_i - X_j\|^2/t) & \|X_i - X_j\|^2 < \varepsilon \\ 0 & \text{otherwise,} \end{cases} \tag{1.7}$$

where $\varepsilon > 0$ is very small and defines the radius of the local neighborhood. As the neighboring points X_i and X_j are mapped far apart, i.e., $\|y_i - y_j\|^2$ is large, the objective function incurs a heavy penalty. Therefore, minimizing (1.6) makes an effort to ensure that, if X_i and X_j are close, then y_i and y_j are close as well (He et al. 2005).

Defining D as a diagonal matrix whose entries are column (or row since S is symmetric) sums of S, i.e., $D_{ii} = \sum_j S_{ji}$ (He et al. 2005), one can derive Laplacian matrix as $L = D - S$. Following some simple algebraic steps, the objective function (1.6) is rewritten as:

$$W^* = \underset{W}{\operatorname{argmin}} W^T X L X^T W \quad \text{s.t. } W^T X D X W = 1. \tag{1.8}$$

The optimal transformation vector W can be derived by the minimum eigenvalue solution to the generalized eigenvector problem: $XLX^TW = \lambda XDX^TW$. By using LPP, a new face recognition method called Laplacianfaces has been proposed to map the face images into a face subspace for analysis and has demonstrated more discriminative power than Eigenfaces and Fisherfaces for face recognition (He et al. 2005).

Tension-based Image Representation

Currently, most subspace learning methods handle image data in the form of 1-D vectors, by concatenating the columns or rows of an image into a single vector. This process most likely results in the *curse of dimensionality* dilemma and the *small sample size problem*, because the dimensionality of features usually is much larger than the sample size. However, image essentially is matrix, i.e., second-order tensors, which has motivated researchers to propose some novel subspace methods, such as 2-D PCA (Yang et al. 2004) and 2-D LDA (Ye et al. 2004), to extract features directly on images. In recent years, advances have focused on encoding face images as second- or higher-order tensors and extending the traditional PCA, LDA, and LPP into their tensorized versions (Yan et al. 2007a; Yan et al. 2007b; Tao et al. 2007; Xu et al. 2008; Fu & Huang 2008a). These emerging methods have showed the superiority over some traditional subspace learning methods in both recognition accuracy and computational efficiency.

Kernel PCA

Linear analysis methods have met with difficulties in extracting the effective features for face recognition, when the face images suffer from the large variations due to illumination conditions, viewing directions or poses, facial expression, aging, and disguises such as facial hair, glasses, or cosmetics. Encountering these challenges, face recognition schemes are required to possess enhanced discrimination abilities. Since the late 1990s, enormous efforts have been put into developing nonlinear analysis methods (e.g., kernel-based methods). By taking advance of a nonlinear kernel mapping, one is able to improve the discriminating power of the feature representation remarkably.

To obtain higher order correlations beyond variances and covariances between input variables, linear PCA was extended to a nonlinear form, Kernel PCA (KPCA) (Schölkopf et al. 1998). The basic idea of KPCA is mapping the original input space to a high-dimensional feature space, where standard PCA is actually conducted. By applying kernel functions that implement a canonical dot product in the low-dimensional input space instead of the high-dimensional feature space, KPCA implicitly achieves the nonlinear mapping between the input space and the feature space, such that the expensive computation in directly nonlinear mapping is avoided subtly.

Given a set of samples $x_1, x_2, \ldots, x_m \in \mathbb{R}^d$ in the input space, the kernel induces a nonlinear mapping between the input space and the feature space denoted by $\phi : \mathbb{R}^d \to H$. Then the kernel mapped data is first centered and then is used to form a data matrix in the feature space: $D = [\phi(x_1), \phi(x_2), \ldots, \phi(x_m)] \in H$. Let $K \in \mathbb{R}^{m \times m}$ be a kernel matrix defined by the dot product in the feature space. The element K_{ij} in K is given as:

$$K_{ij} = k(x_i, x_j) = (\phi(x_i) \cdot \phi(x_j)), \tag{1.9}$$

where $k(\cdot)$ is a kernel function and is usually defined as three types: polynomial kernels, Gaussian kernels, and sigmoid kernels (Schölkopf & Smola 2002). Rather than directly eigen-decomposing the covariance matrix, $C = \frac{1}{m} \sum_{i=1}^{m} \phi(x_i) \phi(x_i)^t$, whose dimensionality is

very high or even infinite in the feature space, Schölkopf et al. (1998) derived the equivalent eigenvalue problem:

$$m\lambda\alpha = K\alpha, \tag{1.10}$$

where $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_m]^t$ is a column vector. Then an eigenvector V of KPCA is computed as follows:

$$V = D\alpha = \sum_{i=1}^{m} \alpha_i \phi(x_i). \tag{1.11}$$

Given a training/test sample x, the corresponding feature in the feature space is $\phi(x)$. The k-th KPCA feature of x is computed as the projection of $\phi(x)$ onto the eigenvector V^k:

$$\beta(x)_k = V^k \phi(x) = \sum_{i=1}^{m} \alpha_i^k K(x_i, x). \tag{1.12}$$

Kernel LDA

Similar to linear PCA, KPCA still captures only the overall variance of input patterns, which is not significant for discrimination purpose. To account for the nonlinear interactions among patterns, linear LDA needs to be extended to a nonlinear form, kernel LDA, to obtain the enhanced discrimination power. The idea of kernel LDA was first proposed by Mika et al. (1999), whose method deals with two-class pattern classification problems. Subsequently, Baudat and Anouar (2000) proposed another type of kernel LDA, called Generalized Discriminant Analysis (GDA), to address multiclass pattern classification problems. Since then, research has witnessed the development of a bunch of kernel LDA algorithms (Roth & Steinhage 2000; Cooke 2002; Yang et al. 2002; Lu et al. 2003; Yang et al. 2005). Thus, there are several kinds of kernel LDA formulations. One among them is called KPCA + LDA (Yang et al. 2005), whose straightforward insight into the nature of kernel LDA makes it easier to understand and implement kernel LDA, particularly

for new investigators and programmers (Yang et al. 2005).

As revealed in Yang's paper, the essence of kernel LDA consists of two steps. That is, KPCA is first used to reduce (or increase) the dimensionality of the input space to n, which is the rank of the covariance matrix C in the feature space, i.e., the rank of the centralized Gram matrix K defined in Equation (1.9). Generally, $n = m - 1$, where m is the number of training samples. Next, LDA is implemented for further feature extraction in the KPCA-transformed space $\mathbb{R}^n$ (Yang et al. 2005). Let y_{ij} be the KPCA feature of the j-th training sample in class i, calculated by Equation (1.12). The within- and between-class scatter matrices in $\mathbb{R}^n$ are defined as:

$$S_w = \sum_{i=1}^{c} \frac{\lambda_i}{l_i} \sum_{j=1}^{l_i} (y_{ij} - \overline{y_i})(y_{ij} - \overline{y_i})^T, \tag{1.13}$$

$$S_b = \sum_{i=1}^{c} \lambda_i (\overline{y_i} - \overline{y})(\overline{y_i} - \overline{y})^T, \tag{1.14}$$

where l_i is the number of training samples in class i with prior probability of λ_i, $\overline{y_i}$ is the mean of the training samples in class i, $\overline{y}$ is the grand mean. The ordinary LDA procedures then derive an optimal transformation matrix U. Finally, given a training/test image, its kernel LDA features are defined as:

$$Z = U^t V^t \phi(x) = U^t \Omega^t \Theta, \tag{1.15}$$

where Ω and Θ have similar definitions as α and $K(x_i, x)$ in KPCA. Please refer to the paper (Yang et al. 2005) for more details.

Manifold Learning

For conducting nonlinear dimensionality reduction on the high-dimensional data that can be considered as a set of geometrically correlated points lying on or nearly on a smooth low-dimensional manifold, some popular manifold learning algorithms include Lo-

cally Linear Embedding (LLE) (Roweis & Saul 2000), Isomap (Tenenbaum et al. 2000), Laplacian Eigenmaps (LE) (Belkin & Niyogi 2003) have recently been developed. LLE maps the input data to a lower dimensional space in a manner that preserves the relationship between the neighboring points; Isomap finds the low-dimensional representations for a data set by approximately preserving the geodesic distances of the data pairs; LE preserves the similarities of the neighboring points, and its linearized form is LPP. Although these algorithms are derived on the basis of different motivations, they all can be unified within the Graph Embedding (GE) framework and its linear/kernel/tensor extensions (Yan et al. 2007). These nonlinear methods do yield impressive results on some benchmark artificial data sets. However, the generated maps are defined only on the training data points, and how to evaluate the maps on novel test data points remains unclear (He et al. 2005). Therefore, these nonlinear manifold learning algorithms might not be suitable for some computer vision tasks, such as face recognition (He et al. 2005).

Recent studies also reveal that locality features and intrinsic geometric structures in the input space may take on additional discriminating power for classification, assuming that Locally Embedded Analysis (LEA) (Fu & Huang, 2005), and Locally Preserving Projections (LPP) (He et al., 2005). Both nonlinear kernel mapping and locally preserved graph embedding help improve the discriminating power of feature representation. However, they generally have the drawback of high-computational cost in classification and overfitting (Fu et al., 2008; Kim et al., 2005).

1.1.2 Image-level Feature Extraction

Image variations cause the changes of data distribution in high-dimensional image space. If raw images are used directly for face recognition, such changes pose burden to the process of statistical feature extraction via either linear or nonlinear way. To facilitate the statistical feature extraction for classification, some image feature extraction techniques have been developed to process raw images for extracting image features, which are usually invariant

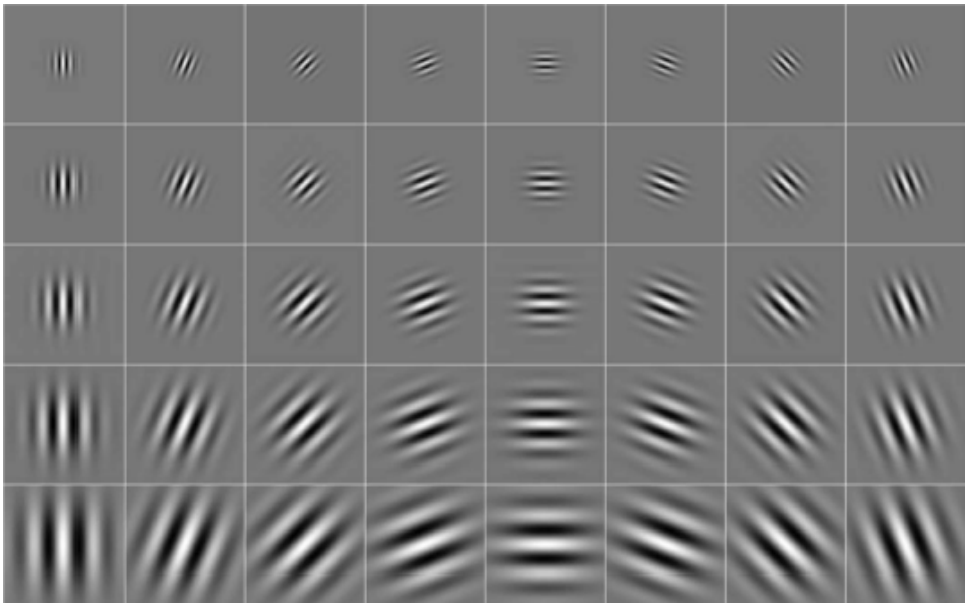

Figure 1.4 Gabor wavelets: the real part of the Gabor kernels at five scales and eight orientations.

to image variations. Some popular methods in image feature extraction include Gabor Image Representation (GIR) (Daugman 1985; Donato et al. 1999; Liu & Wechsler 2002), and Local Binary Patterns (LBP) (Ojala et al. 1996; Ojala et al. 2002).

Gabor Image Representation (GIR)

GIR of an image captures salient visual properties such as spatial location, orientation selectivity, and spatial frequency characteristics (Daugman 1985), displaying robust characteristics in dealing with image variabilities. Specifically, GIR is the convolution of the image with a family of Gabor kernels that may be formulated as follows (Daugman 1985):

$$\psi_{\mu,\nu}(z) = \frac{\|k_{\mu,\nu}\|^2}{\sigma^2} e^{-\frac{\|k_{\mu,\nu}\|^2 \|z\|^2}{2\sigma^2}} \left[e^{ik_{\mu,\nu} z} - e^{-\frac{\sigma^2}{2}} \right] \tag{1.16}$$

where μ and ν define the orientation and scale of the Gabor kernels, $z = (x, y)$, $\|\cdot\|$ denotes the norm operator, and the wave vector $k_{\mu,\nu}$ is defined as follows:

Figure 1.5 Gabor image representation (GIR): the magnitude representation.

$$k_{\mu,\nu} = k_\nu e^{i\phi_\mu} \tag{1.17}$$

where $k_\nu = k_{max}/f^\nu$ and $\phi_\mu = \pi\mu/8$. k_{max} is the maximum frequency, and f is the spacing factor between kernels in the frequency domain. Let $I(x,y)$ represent a face image, the convolution of $I(x,y)$ and a Gabor kernel $\psi_{\mu,\nu}$ may be formulated as follows:

$$O_{\mu,\nu}(z) = I(z) * \psi_{\mu,\nu}(z) \tag{1.18}$$

where $z = (x,y)$, $*$ denotes the convolution operator, and $O_{\mu,\nu}(z)$ is the convolution result corresponding to the Gabor kernel at orientation μ and scale ν. Commonly used Gabor kernels contain five different scales, $\nu \in \{0,...,4\}$, and eight orientations, $\mu \in \{0,...,7\}$. The set $\mathscr{S} = \{O_{\mu,\nu}(z) : \mu \in \{0,...,7\}, \nu \in \{0,...,4\}\}$, thus, forms GIR of the image I. Figure 1.4 shows the the real part of Gabor kernels at different scales and orientations, and Figure 1.5 shows the magnitude of GIR derived from the convolution of a face image with these Gabor kernels. Usually, the concatenated magnitude images are considered as Gabor features for face recognition (Liu et al. 2002).

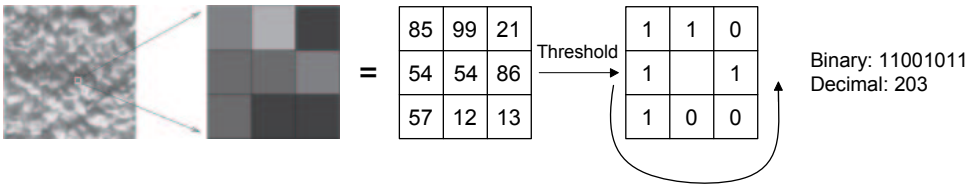

Figure 1.6 An example of the basic LBP operation (Ahonen et al. 2006).

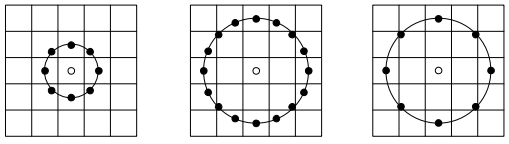

Figure 1.7 An example of LBP circular neighborhoods: (8,1), (16,2), and (8,2). When the sampling point is not at the center of a pixel, the pixel value is bilinearly interpolated.

Local Binary Patterns (LBP)

LBP, which was originally introduced for texture analysis (Ojala et al. 1996), has been successfully extended to describe face images and demonstrated effective for face recognition, due to the finding that face images can be viewed as a composition of micro-patterns that are well described by the LBP operators. In a 3×3 neighborhood of an image, the basic LBP operator assigns a binary label 0 or 1 to each surrounding pixel by thresholding at the gray value of the central pixel and replacing its value with a decimal number converted from the 8-bit binary number. Formally, the LBP operator is defined as follows:

$$LBP = \sum_{p=0}^{7} 2^p s(i_p - i_c) \tag{1.19}$$

where $s(i_p - i_c)$ equals 1, if $i_p - i_c \geq 0$; and 0, otherwise. Figure 1.6 illustrates an example of the basic LBP operator.

Two extensions of the basic LBP were further developed by Ojala et al. (2002). The first extension allows LBP to deal with any size of neighborhoods by using circular neighborhoods and bilinearly interpolating the pixel values. The notation (P, R) thus represents P points on a circle of radius of R. Figure 1.7 shows an example of circular neighborhoods.

The second extension defines the so called *uniform patterns*. When the binary string is traversed circularly, LBP can be called uniform if there are at most two bitwise transitions from 0 to 1 or vice versa. For example, the patterns 00000000 (0 transitions), 00011100 (2 transitions) and 11000011 (2 transitions) are uniform whereas the patterns 11010011 (4 transitions) and 00101001 (6 transitions) are not. In the computation of LBP histograms, every uniform pattern has its own bin and all nonuniform patterns are assigned to a separate bin. Ahonen et al. (2006) have found that 90.6% of the patterns in the (8,1) neighborhood and 85.2% of the patterns in the (8,2) neighborhood are uniform when processing FERET facial images. After extensions, LBP can be expressed as: $LBP^{u2}_{P,R}$.

1.1.3 Color Information for Face Recognition

Most existing face recognition methods work with the gray-scale images, which have been revealed to be suffering from severe image variations by the recent Face Recognition Grand Challenge (FRGC) program (Phillips et al. 2005). Color information has been widely applied in face detection, but not in face recognition. Recent studies show that different color spaces transformed from the RGB color space display different discriminating power for pattern recognition (Hsu et al. 2002; Geusebroek et al. 2001; Torres et al. 1999; Ross & Govindarajan 2005; Shih et al. 2005). As different color component images in a color space exhibit the various representations of faces, they could provide the complementary information to each other. Such a characteristic implies that when some facial features produce the incorrect classification output due to image variations on the luminance-related images, such as Y, R, and V of HSV, they could work well on the chromatic images, such as I, Q, Cr, and so on. As a result, the sets of face images misclassified by the different color component images would not necessarily overlap. Therefore, data fusion at the image level or the decision level can be used to combine several color component images to improve the performance of face recognition, when compared to the methods using gray-scale images alone.

There are a few approaches in the literature to address the feature extraction in color space. Xie and Kumar (2005) proposed the quaternion correlation filter technique for color face recognition that processes all of the color channels jointly, where the quaternion representation encodes the three color components (such as in the RGB and NTSC color spaces) in the imaginary parts of the quaternion number. Jones and Abbott (2006) further proposed to use the quaternion representation to extend the Gabor filters from the complex domain to the hypercomplex domain for color face recognition. To overcome the large illumination variations on color in face images, Kim and Choi (2007) proposed the construction of a tensor of color image ensemble. They formed a 4-way tensor whose coordinates are associated with rows and columns of face images, color, and samples, and then employed the higher-order SVD of the tensor to extract such features. The other approaches of extracting features in color face images include the application of Non-negative Matrix Factorization (NMF) (Rajapakse et al. 2004) and LBP (Chan et al. 2007).

This section first details ten commonly used color spaces in computer vision and pattern recognition, then introduces an emerging technique of color space normalization in face recognition. The ten color spaces include: the RGB color space, the *rgb* color space, the $I_1I_2I_3$ (decorrelated RGB) color space, the human perceptual color spaces HSV and HSI, the video transmission efficiency color spaces YIQ, YUV and YCbCr, the CIE-XYZ color space, and the CIE perceptually uniform color space CIE-$L^*a^*b^*$.

The rgb color space

The RGB colors are sensitive to surface orientation, illumination direction, and illumination intensity. To alleviate such sensitivities, one can derive the normalized colors by projecting the R, G, B values onto the R = G = B = max{R,G,B} plane. The projection spans a normalized *rgb* chromaticity triangle. The transformation is defined as:

$$
\begin{aligned}
r &= R/(R+G+B) \\
g &= G/(R+G+B) \\
b &= B/(R+G+B).
\end{aligned}
\tag{1.20}
$$

The $I_1I_2I_3$ color space

An effective color space to stabilize RGB images is $I_1I_2I_3$, proposed by Ohta et al. (1980), which uses Karhunen-Loève transformation (KLT) to decorrelate the RGB components. The conversion from RGB to $I_1I_2I_3$ based on Ohta's experimental model is given by the simple linear transformation:

$$
\begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 0 & -1/2 \\ -1/2 & 1 & -1/2 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \tag{1.21}
$$

Human perceptual color spaces

The HSI and HSV color spaces are motivated by the human vision system in a sense that one can describe color by means of hue, saturation, and brightness. Hue and saturation define chrominance, while intensity or value specifies luminance. The HSI color space is defined as follows:

$$
\begin{aligned}
I &= (R+G+B)/3 \\
S &= 1-(1/I)min(R,G,B) \\
H &= \begin{cases} \theta & \text{if } B \leq G \\ 360-\theta & \text{otherwise,} \end{cases}
\end{aligned}
\tag{1.22}
$$

where

$$\theta = \cos^{-1}\left\{\frac{\frac{1}{2}[(R-G)+(R-B)]}{[(R-G)^2+(R-G)(G-B)]^{\frac{1}{2}}}\right\}. \tag{1.23}$$

The HSV color space is defined as:

$$\text{Let}\begin{cases} \text{MAX} &= \max(R,G,B) \\ \text{MIN} &= \min(R,G,B) \\ \delta &= \text{MAX}-\text{MIN} \end{cases}$$

$$H = \begin{cases} 60(\frac{G-B}{\delta}) & \text{if MAX} = \text{R} \\ 60(\frac{B-R}{\delta}+2) & \text{if MAX} = \text{G} \\ 60(\frac{R-G}{\delta}+4) & \text{if MAX} = \text{B} \\ \text{not defined} & \text{if MAX} = 0 \end{cases} \tag{1.24}$$

$$S = \begin{cases} \frac{\delta}{\text{MAX}} & \text{if MAX} \neq 0 \\ 0 & \text{if MAX} = 0 \end{cases}$$

$$V = \text{MAX}.$$

In order to confine H within the range of [0,360],

$$H = H + 360 \quad \text{if } H < 0. \tag{1.25}$$

Note that the R, G, B values in both Equations (1.22) and (1.24) are scaled to $[0,1]$.

Video transmission efficiency color spaces

The YUV and YIQ color spaces are widely used in video for transmission efficiency. The YUV color space is applied by Phase Alternation by Line (PAL) and the System Electronique Couleur Avec Memoire (SECAM), and the YIQ color space is adopted by the National Television System Committee (NTSC) video standard in reference to RGB NTSC. The YUV color space is defined as:

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.2990 & 0.5870 & 0.1140 \\ -0.1471 & -0.2888 & 0.4359 \\ 0.6148 & -0.5148 & -0.1000 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \tag{1.26}$$

The I and Q components are derived from their counterparts, U and V, by a clockwise rotation (33^o), and then the YIQ color space is defined as follows:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.2990 & 0.5870 & 0.1140 \\ 0.5957 & -0.2745 & -0.3213 \\ 0.2115 & -0.5226 & 0.3111 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \tag{1.27}$$

The YCbCr color space is a scaled and offset version of the YUV color space. To derive YCbCr, the RGB components are distributed into luminance (Y), chrominance blue (Cb), and chrominance red (Cr). The Y component has 220 levels ranging from 16 to 235, while the Cb, Cr components have 225 levels ranging from 16 to 240:

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.4810 & 128.5530 & 24.9660 \\ -37.7745 & -74.1592 & 111.9337 \\ 111.9581 & -93.7509 & -18.2072 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \tag{1.28}$$

where the R, G, B values are scaled to $[0,1]$.

CIE uniform color spaces

In the CIE (Commission Internationale de 1'Éclairage) systems, the starting point for all color specification is CIE XYZ. XYZ is known as tristimulus values, which lead to the other CIE perceptually uniform color spaces, such as the $L^*a^*b^*$ color space. A linear

transformation from RGB space to XYZ space is defined as:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.607 & 0.174 & 0.201 \\ 0.229 & 0.587 & 0.114 \\ 0.000 & 0.066 & 1.116 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \tag{1.29}$$

The L*a*b* color space is one of the most popular color spaces and is modeled based on the human vision system. The L* component represents brightness from 0 (black) to 100 (white). The a* component corresponds to the measurement of redness (positive values) or greenness (negative values), and the b* component corresponds to the measurement of yellowness (positive values) and blueness (negative values). Based on the XYZ tristimulus, the L*a*b* color space is defined as:

$$\begin{aligned} L^* &= 116 f(\tfrac{Y}{Y_o}) - 16 \\ a^* &= 500 \left[f(\tfrac{X}{X_o}) - f(\tfrac{Y}{Y_o}) \right] \\ b^* &= 200 \left[f(\tfrac{Y}{Y_o}) - f(\tfrac{Z}{Z_o}) \right] \end{aligned} \tag{1.30}$$

where

$$f(x) = \begin{cases} x^{1/3} & \text{if } x > 0.008856 \\ 7.787x + \frac{16}{116} & \text{otherwise.} \end{cases} \tag{1.31}$$

In addition to the above existing color spaces, a few researchers have devoted to generating some new color spaces for face recognition (Jones III & Abbott 2004; Liu 2008; Yang et al. 2008; Liu & Yang 2009). The main idea of these methods can be described as: first, the color-space total, between-class, and within-class scatter matrices are constructed using pixel values containing different color components, then some linear learning algorithms, such as PCA, LDA, and Independent Component Analysis (ICA), are applied to these scatter matrices to derive the transformation vectors, which generate one or several new color components from original color components. The motivation behind generating new color spaces is to make the new color components more uncorrelated, more discrimina-

tive, and more independent than original color components for improving the performance of face recognition.

Color Space Normalization for Face Recognition

The most important issue of color-based face recognition is to choose the appropriate color spaces, which are able to provide more discriminative power than the others. As a straightforward way, the exhaustive enumeration strategy has been applied in the reference (Shih et al. 2005) to seek the best color spaces. Their research reveals that different color spaces transformed from the RGB color display different discriminating abilities for face recognition. Specifically, the YIQ color space provides better face recognition performance in comparison with other color spaces. One nature question arises "what kind of color spaces is suitable for color face recognition?", namely, what characteristics are related to the discriminative color spaces for face recognition? To answer this question, Yang et al. (2010b) have proposed the concept of Double-Zeros-Sum (DZS) in color spaces.

By discovering what characteristics make the $I_1I_2I_3$, YUV, and YIQ color spaces more powerful than the RGB and XYZ color spaces for face classification, Yang et al. (2010a) found an interesting observation in color transformation matrices: the sums of the elements in the second and third rows of the transformation matrix are both zero or approximate zero. This characteristic is so called Double-Zeros-Sum (DZS) (Yang et al. 2010b). However, the RGB and XYZ color spaces do not hold such a property. For example, the transformation matrix of the RGB color space is given as follows:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \tag{1.32}$$

Note the transformation matrix is an identity matrix, indicating the sums of the elements

in second and third rows are not zero. A similar situation also appears in other weak discriminating color spaces, such as XYZ.

Inspired by this finding, Yang et al. (2010a) further proposed the concept of color space normalization (CSN) and developed two CSN techniques. These CSN techniques can work well in all color spaces that are derived by a linear transformation of the RGB color space, so that the normalized color spaces possess the same property as the strong discriminating color spaces. The two normalization techniques, according to whether the computation procedures happen within a color component or across three color components, are named as *within-color-component* normalization (CSN-I) and *across-color-component* normalization (CSN-II), respectively. Next, the CSN-I technique will be briefly presented, and readers can refer to the paper (Yang et al. 2010a) for the details of the CSN-II.

Let C_1, C_2, and C_3 be the three color components derived by the following linear transformation of the RGB color space:

$$\begin{bmatrix} C_1 \\ C_2 \\ C_3 \end{bmatrix} = \mathbf{A} \begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} \mathbf{A_1} \\ \mathbf{A_2} \\ \mathbf{A_3} \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \tag{1.33}$$

The mean values of the second and the third rows in $\mathbf{A}$ are $m_2 = (a_{21} + a_{22} + a_{23})/3$ and $m_3 = (a_{31} + a_{32} + a_{33})/3$, respectively. Subtracting m_2 and m_3 from the second and third rows results in a normalized transformation matrix $\tilde{\mathbf{A}}$, which generates a normalized color space $\tilde{C}_1\tilde{C}_2\tilde{C}_3$ (Yang et al. 2010a):

$$\begin{bmatrix} \tilde{C}_1 \\ \tilde{C}_2 \\ \tilde{C}_3 \end{bmatrix} = \tilde{\mathbf{A}} \begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} \mathbf{A_1} \\ \tilde{\mathbf{A}}_2 \\ \tilde{\mathbf{A}}_3 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$
$$= \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21}-m_2 & a_{22}-m_2 & a_{23}-m_2 \\ a_{31}-m_3 & a_{32}-m_3 & a_{33}-m_3 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \quad (1.34)$$

For example, the normalized RGB and XYZ color spaces via the CSN-I are defined as:

$$\begin{bmatrix} \tilde{R} \\ \tilde{G} \\ \tilde{B} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -1/3 & 2/3 & -1/3 \\ -1/3 & -1/3 & 2/3 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix},$$
$$\begin{bmatrix} \tilde{X} \\ \tilde{Y} \\ \tilde{Z} \end{bmatrix} = \begin{bmatrix} 0.607 & 0.174 & 0.201 \\ -0.0343 & 0.2537 & -0.2193 \\ -0.3940 & -0.3280 & 0.7220 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \quad (1.35)$$

Some Preliminary Evaluations of Different Color Spaces for Face Recognition

To evaluate the feasibility of color information in face recognition, a set of experiments has been executed on the FRGC version 2 Experiment 4, the most challenging FRGC experiment, which contains 12,776 face images of 222 subjects in the training set, 16,028 face images of 466 subjects in target set, and 8,014 face images of 466 subjects in the query set. Image normalization is first used to align the centers of the eyes to specified positions and fixed interocular distance. To be specific, the centers of the eyes are provided by the FRGC

Table 1.1 Face Verification Rate (FVR) ROC III of the FRGC Database Using Several Different Color Spaces

Color space	FVR ROC III				
	color 1	color 2	color 3	Sum rule	Product rule
RGB	62.12%	53.78%	39.93%	59.14%	56.54%
XYZ	59.16%	56.41%	40.94%	49.53%	34.38%
HSV	10.87%	53.99%	61.81%	66.95%	53.92%
$L^*a^*b^*$	59.30%	47.84%	35.22%	68.77%	40.43%
YUV	56.41%	52.36%	57.20%	72.25%	71.63%
$I_1I_2I_3$	55.09%	51.56%	52.67%	77.97%	74.36%
YIQ	56.41%	54.20%	53.97%	78.58%	67.19%
$\tilde{R}\tilde{G}\tilde{B}$	62.12%	52.67%	51.39%	78.06%	75.24%
$\tilde{X}\tilde{Y}\tilde{Z}$	59.16%	58.20%	51.24%	71.10%	71.69%
RIQ	62.12%	54.20%	53.97%	79.03%	76.97%

database and the predefined positions in the 64 $\times$ 64 images are (17, 18) and (47, 18). A subimage procedure then crops the face image to the size of 64 $\times$ 64 to extract the facial region. Some example images are shown in Figure 1.1. Finally, an LDA-based algorithm and the cosine similarity measure are implemented to recognize faces.

The classification results of individual color components are first derived, and then the sum rule and the product rule are applied, respectively, to fuse these classification outputs for deriving the performance of color spaces for face recognition. Experimental results using the different color spaces are derived from the Receiver Operating Characteristic (ROC) curves. Table 1.1 summarizes the Face Verification Rate (FVR) from the ROC III curves. As can be seen, when the FVR (56.41%) of the luminance Y is considered as benchmark, most of color spaces improve the face recognition performance significantly. In particular, some strong discriminating color spaces, such as YIQ, $I_1I_2I_3$, YUV, can achieve good performance in face recognition, while some weak discriminating color spaces, such as RGB and XYZ, achieve low performance. On the other hand, experimental results show that the color normalization technique indeed causes the weak color spaces RGB and XYZ to enhance the discrimination for face recognition. In addition, two nonlinear color spaces

HSV and $L^*a^*b^*$ do not demonstrate any advantage over some strong discriminating color spaces. It should be noted that the hybrid color spaces, such as RIQ, demonstrate the advantage in improving recognition performance over other color spaces.

1.2 Topics Overview

Most of the existing color-based face recognition methods pay attention to the selection and generation of color spaces, largely ignoring the research on the low-level feature extraction in different color spaces. For example, the paper (Shih et al. 2005) assesses comparatively face recognition performance in different color spaces using a standard PCA algorithm. The research reveals that some color configurations, such as YV in the YUV color space and YI in the YIQ color spaces, can help improve face recognition performance. However, nothing about the low-level feature extraction has been investigated in this paper. To fill in such a gap, this dissertation focuses on face recognition by addressing facial feature extractions in different color component images. Specifically, some image feature extraction methods have been investigated and developed to extract the effective features from color images for face recognition. The feasibility and effectiveness of the proposed methods have been evaluated by two large-scale face databases, namely, the Face Recognition Grand Challenge (FRGC) version 2 database and the FERET database. The FRGC version 2.0 Experiment 4, the most challenging FRGC experiment, contains 12,776 training images, 16,028 controlled target images, and 8,014 uncontrolled query images. To assess the generalization of the proposed methods, the experiments using the FERET database are also conducted. The two most challenging probe sets, Dup I and Dup II, which are used for analyzing the effect of aging on the recognition performance, are evaluated in the experiments. The Biometric Experimentation Environment (BEE) system of FRGC and the rank-1 accuracy rate of FERET reveal that the proposed methods improve performance of face recognition significantly, which not only take advantages over the traditional methods using gray-scale images but also outperform some state-of-the-art face recognition meth-

ods.

Chapter 2 presents a novel multiple feature fusion method for face recognition by fusing the frequency, spatial, and color features for improving the face recognition grand challenge performance. In particular, the hybrid color space RIQ is constructed, according to the discriminating properties among the individual component images. For each component image, the frequency features are extracted from the magnitude, the real and imaginary parts in the frequency domain of an image. Then, a variant of Regularized Linear Discriminant Analysis (RLDA) extracts discriminating features from the frequency data for similarity computation using a cosine similarity measure. Finally, the three similarity matrices generated using the three components in the RIQ color space are fused by means of the sum rule — the decision level fusion — to derive the final similarity matrix for face recognition. The effectiveness of the proposed method is demonstrated using two large-scale face databases, namely, the FRGC version 2 and the FERET databases. In particular, experiments on FRGC and BEE show that for the most challenging FRGC version 2 Experiment 4, the proposed method achieves the face verification rate (ROC III) of 82.49% at the false accept rate of 0.1%. For the FERET Dup I probe set, the proposed method achieves the rank-1 accuracy of 81.30%.

Chapter 3 describes a novel face recognition method, embodying the "Color + LBP + LDA" strategy. First, another hybrid color space VIQ is constructed by replacing the luminance Y in YIQ by the V component in HSV. As there is more mutual independence information in VIQ than in the RGB, the hybrid VIQ color space is more feasible as a complementary representation of face images for classification than some color spaces, such as RGB. On each component image of VIQ, a multipscale LBP fusion is proposed to encompass the different LBP histogram features. By utilizing such a fusion scheme, both the microstructures and macrostructures of face images are fused to extract the discriminating features, which can contain much more discrimination information than the one a single LBP operator can provide. Regarding the extraction of discriminating features, a variant

of RLDA is used to extract the complete discriminating information, which resides in both the principal and null spaces of the within-class scatter matrix. Experiments on the FRGC version 2 and FERET databases show that the proposed method using the "Color + LBP + LDA" strategy can significantly improve face recognition performance under difficult conditions.

Chapter 4 is concerned with the application of different feature extraction methods to different color component images. Now that the different color component images display the various discriminating properties and face representations, it is more feasible to apply a variety of feature extraction methods than the single-feature extraction methods to some color component images for face recognition. This chapter first introduces the derivation of a new color space, the Color Image Discriminant (CID) color space (Yang et al. 2008). Then an emphasis is laid on using Gabor Image Representation (GIR), Local Binary Patterns (LBP), and the frequency features in the different color component images of the CID color space, respectively, for improving the performance of face recognition. In particular, experiments on the FRGC data set show that the proposed method can achieve the FVR of 91.6% at 0.1% FAR.

Chapter 5 discusses the generation of some new image representations for face recognition. Most existing face recognition methods use the gray-scale image, which is the simple linear combination of the primary colors R, G, and B. However, the disadvantage of such an usage is that R, G, and B have strong correlation to each other. Since the decorrelation property is essential for pattern recognition, the image representations derived from three primary color components are not ideally suited for face recognition. This chapter first derives some new image representations from the other color spaces through a learning algorithm, and then proposes a novel method for addressing the problem of face recognition under the difficult illumination conditions. Experiments on two large-scale face databases, FRGC and FERET, show the effectiveness of the proposed method.

The conclusions and future work are presented in Chapter 6, where the major con-

tribution of this dissertation are summarized and future research directions are discussed.

CHAPTER 2

FUSING FREQUENCY AND COLOR FEATURES FOR FACE RECOGNITION

The preceding chapter has experimentally demonstrated the feasibility of color information for improving the performance of face recognition. In particular, the YIQ color space displays better discrimination than some other color spaces. Under the precondition that the complementary property is guaranteed, if one of color component images is replaced as other one with stronger discriminating ability, the overall classification performance provided by integrating several component images could be enhanced. Recent research has demonstrated that the R color component image possesses stronger discriminating power than the luminance Y for face recognition (Shih et al. 2005). A hybrid color space RIQ instead of the YIQ color space, thus, is considered in this chapter to provide color information for face recognition.

As an important tool in image processing, Discrete Fourier Transform (DFT) has been widely employed in designing image filters. After processing the input data, DFT generates four components in the frequency domain: real part, imaginary part, magnitude, and phase angel. Since these frequency components have the different interpretations for input data, the potential complementary characteristics could be provided by them. Thus, it is feasible to combine these DFT frequency components for face recognition (Hwang et al. 2006).

This chapter investigates the application of frequency and color information to improve the performance of face recognition. The method is hereby carried out in a hybrid color space, RIQ, which is revealed to have more discriminating capabilities than the YIQ color space. To represent face image efficiently, the multiple complementary frequency features, comprising the real part, the imaginary part, and the magnitude of the Fourier transform, are first extracted from each component image of RIQ. Then, the concatenated real and imaginary parts, and the magnitude are processed, respectively, to derive the dis-

criminating features using a variant of Regularized Linear Discriminant Analysis (RLDA). The resulting discriminating features are then concatenated to form an augmented vector, which is further processed by RLDA. Besides, the spatial information contained in the multiple scaled face images is investigated to improve recognition performance. Finally, at the decision level, the similarity matrices corresponding to the three color component images are fused by the sum rule to generate the final similarity matrix for face recognition. To improve the generalization performance of the FLD method, RLDA analyzes the eigenvalue spectrum of the within-class scatter matrix and replaces its small-valued trailing eigenvalues with a constant. The new eigenvalue spectrum thus replaces the original spectrum of the within-class scatter matrix for the simultaneous diagonalization of the within- and the between-class scatter matrices to extract the discriminating features.

To evaluate the effectiveness of the proposed method, experiments are conducted on two large-scale face databases, the FERET and FRGC databases. Experimental results show that (i) the hybrid color space improves face recognition performance, and (ii) the complementary frequency and spatial features further improve face recognition performance.

2.1 The Selection of Hybrid Color Space: RIQ

The complementary characteristics of color spaces can be applied to improve face recognition performance (Torres et al. 1999; Shih et al. 2005). As the multiple imaging in a color space encodes the complementary and different representations for the same face image, the data fusion occurred in the image level or the decision level could lead to improved overall performance. The recent research reveals that the YIQ color space is among the best ones for face recognition, and the R component image has better discriminating capability than the gray image Y (Liu et al. 2008; Shih et al. 2005; Liu 2006). Therefore, it may be a fact that the hybrid color space RIQ contains more discriminating power than the YIQ color space for face recognition.

Next, the procedures are presented to calculate the amount of discriminating capability contained in a color space. Assume that $\mathbf{A}$ be a color face image with a size of $m \times n$, consisting of three component images $\mathbf{C}_1$, $\mathbf{C}_2$, and $\mathbf{C}_3$, where $\mathbf{C}_1, \mathbf{C}_2, \mathbf{C}_3 \in \mathbb{R}^N$ are the column vectors and $N = m \times n$. The color image $\mathbf{A}$ then can be expressed as an $N \times 3$ matrix: $\mathbf{A} = [\mathbf{C}_1, \mathbf{C}_2, \mathbf{C}_3] \in \mathbb{R}^{N \times 3}$. Let L be the number of subjects in training data set, and $\mathbf{A}_{ij}$ be the j-th color image in subject i, where $i = 1, 2, ..., L$, and $j = 1, 2, ..., M_i$, and M_i denotes the number of training samples in subject i. Then, the mean of the training samples in subject i is defined as

$$\bar{\mathbf{A}}_i = \frac{1}{M_i} \sum_{j=1}^{M_i} \mathbf{A}_{ij} = [\bar{\mathbf{C}}_{1i}, \bar{\mathbf{C}}_{2i}, \bar{\mathbf{C}}_{3i}]. \tag{2.1}$$

The grand mean of training samples is defined as

$$\bar{\mathbf{A}} = \frac{1}{M} \sum_{i=1}^{L} \sum_{j=1}^{M_i} \mathbf{A}_{ij} = [\bar{\mathbf{C}}_1, \bar{\mathbf{C}}_2, \bar{\mathbf{C}}_3], \tag{2.2}$$

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

Based on Equations (2.1) and (2.2), the color space between-class and within-class scatter matrices S_b and S_w are defined as follows:

$$S_b = \sum_{i=1}^{L} P_i [(\bar{\mathbf{A}}_i - \bar{\mathbf{A}})^T (\bar{\mathbf{A}}_i - \bar{\mathbf{A}})] \tag{2.3}$$

$$S_w = \sum_{i=1}^{L} P_i \frac{1}{M_i - 1} \sum_{i=1}^{M_i} [(\mathbf{A}_{ij} - \bar{\mathbf{A}}_i)^T (\mathbf{A}_{ij} - \bar{\mathbf{A}}_i)] \tag{2.4}$$

where P_i is the priori probability of subject i, and $S_b, S_w \in \mathbb{R}^{3 \times 3}$. The amount of discriminating capability of a color space, consisting of $\mathbf{C}_1$, $\mathbf{C}_2$, and $\mathbf{C}_3$, can be calculated by using the Fisher criterion:

$$J_4 = \frac{tr(S_b)}{tr(S_w)}. \tag{2.5}$$

Given that the color image $\mathbf{A}$ consists of three component images R, I, Q, or Y, I, Q, their J_4 values are derived, respectively, by using Equation (2.5). Note that the intensity

Table 2.1 The Values of the Criterion J_4 among R, I, Q, and Y, I, Q Component Images

Color space	Criterion J_4
RIQ	0.4464
YIQ	0.4319

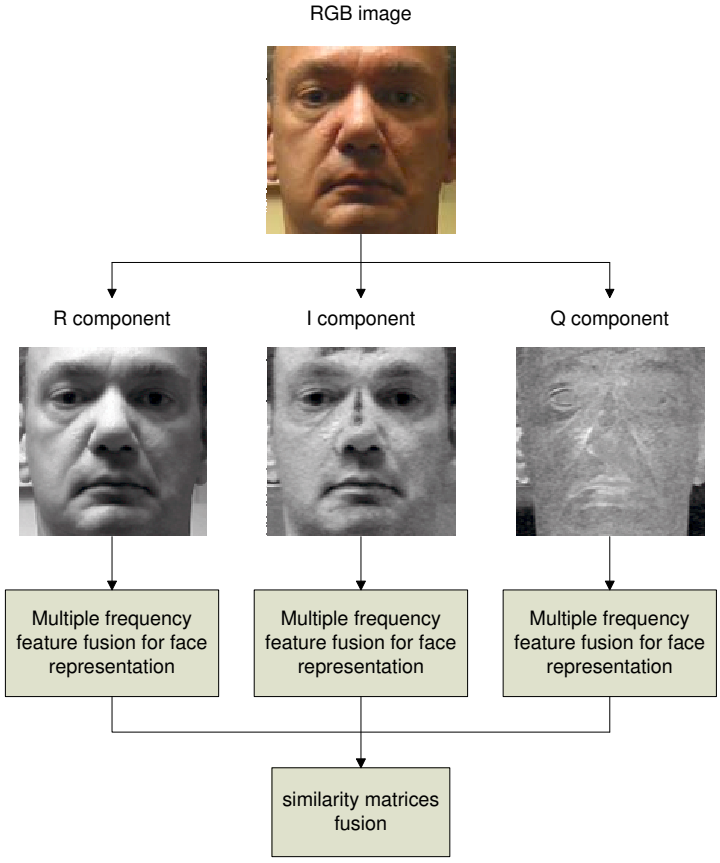

Figure 2.1 System architecture of the proposed method.

values of three component images are normalized to zero mean and unit variance before computation. The results are given in Table 2.1, which shows that the hybrid color space RIQ contains more discriminating information than the conventional color space YIQ for face recognition.

Therefore, the hybrid color space RIQ is selected to provide the color information in the method to recognize face images. Figure 2.1 shows the system architecture of the

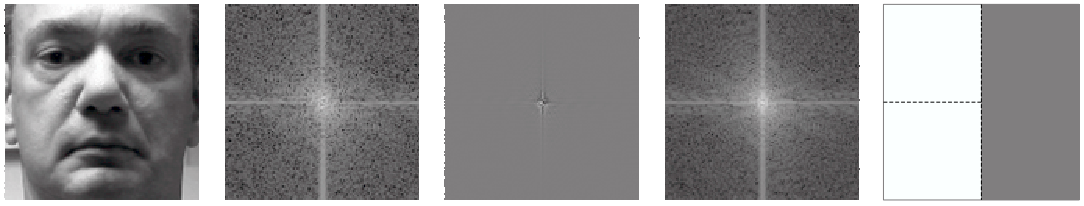

Figure 2.2 The 2D discrete Fourier transform of a face image: the real part (log plot), the imaginary part, and the magnitude (log plot). The frequency features used in the method are extracted from the right two quadrants, as indicated by the gray area.

proposed method. For each component image in the RIQ hybrid color space, the multiple frequency feature fusion for face representation extracts the multiple frequency features derived from the Fourier domain of face image to calculate the similarity scores. At the decision level, the similarity scores derived from the color component images in the RIQ hybrid color space are fused to improve face recognition performance.

2.2 Multiple Frequency Feature Fusion for Face Representation

Fourier transform is able to convert an image from the spatial domain to the frequency domain, where the image is decomposed into the combination of various frequencies. Applying this technique, one can extract the salient image properties in the frequency domain that are often not available in the spatial domain. For an image $f(x,y)$ with a spatial resolution of $N = m \times n$ ($x = 0,1,\cdots,m-1$ and $y = 0,1,\cdots,n-1$), let $F(u,v)$ be its 2D Discrete Fourier Transform (DFT) (Gonzalez & Woods 2002):

$$F(u,v) = \sum_{x=0}^{m-1}\sum_{y=0}^{n-1} f(x,y)e^{-j2\pi(ux/m+vy/n)} \tag{2.6}$$

where $u = 0,1,\cdots,m-1$ and $v = 0,1,\cdots,n-1$. Generally speaking, the DFT $F(u,v)$ is complex, comprising the real and imaginary parts: $R(u,v)$ and $I(u,v)$. The Fourier spec-

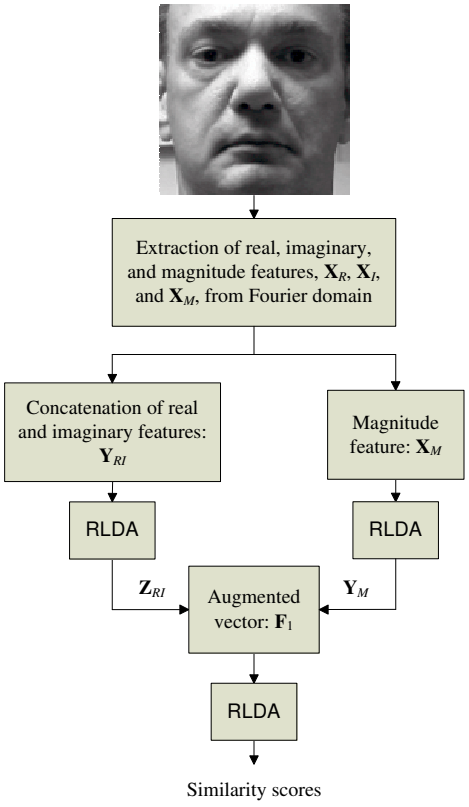

Figure 2.3 The multiple frequency features fusion scheme for the R component image.

trum or magnitude is defined as follows (Gonzalez & Woods 2002):

$$|F(u,v)| = [R^2(u,v) + I^2(u,v)]^{1/2} \tag{2.7}$$

Figure 2.2 shows the 2D discrete Fourier transform of a face image: the real part, the imaginary part, and the magnitude (log plot). As the Fourier domain is symmetric with respect to the origin, the frequency features contained in the right two quadrants are chosen to reduce the feature size in the Fourier domain.

Figure 2.3 shows the outline of the multiple frequency feature fusion scheme for the R component image. The real, imaginary, and magnitude parts are first extracted from the Fourier domain. Let $\mathbf{X}_R$, $\mathbf{X}_I$, and $\mathbf{X}_M$ be the real part, imaginary part, and magnitude part of frequency features from the right two quadrants, respectively. These features are

then converted into column vectors, $\mathbf{X}_R, \mathbf{X}_I, \mathbf{X}_M \in \mathbb{R}^{N/2}$. The concatenated form of real and imaginary parts can be expressed as $\mathbf{Y}_{RI} = (\mathbf{X}_R; \mathbf{X}_I) \in \mathbb{R}^N$. Subsequently, $\mathbf{Y}_{RI}$ will be processed by RLDA for discriminating feature extraction, and the resulting feature vector is $\mathbf{Z}_{RI} \in \mathbb{R}^d$, where d is less than L, the number of subjects. As the magnitude feature $\mathbf{X}_M$ is related to the real and imaginary parts, it is fused with the counterpart at the feature level after the RLDA feature extraction. Let $\mathbf{Y}_M \in \mathbb{R}^d$ be the feature vector after the RLDA process of $\mathbf{X}_M$, then an augmented vector is defined as:

$$\mathbf{F}_1 = (\frac{\mathbf{Z}_{RI} - \mu_{RI}}{\delta_{RI}}; \frac{\mathbf{Y}_M - \mu_M}{\delta_M}) \tag{2.8}$$

where μ_{RI}, μ_M, and δ_{RI}, δ_M are the mean values and the standard deviations of $\mathbf{Z}_{RI}$ and $\mathbf{Y}_M$, respectively. Finally, the augmented vector $\mathbf{F}_1$ will be processed by RLDA to derive the similarity scores, through a cosine similarity measure.

The multiple frequency feature fusion scheme applies to the I and Q component images as well to generate the similarity scores, respectively. At the decision level, the similarity scores derived from the these color component images are fused for improving face recognition performance.

2.3 A Variant of Regularized Linear Discriminant Analysis

As face images usually reside in a very high dimensional space, learning to recognize human faces demands great capabilities in finding the compact representations of the meaningful features with low dimensionality from the high dimensional face space. Fisher Linear Discriminant analysis (LDA or FLD) is an efficient way to extract such discriminative features as it defines the separation between classes to be the ratio of the between-class scatter S_B to the within-class scatter S_W and attempts to maximize this ratio $J_{LDA}(W_{opt}) = \max(|W^T S_B W|/|W^T S_W W|)$. Thus some LDA-based face recognition methods have been extensively studied and achieved promising results (Liu & Wechsler 2000; Huang et al.

2002; Jiang et al. 2008).

In the paper (Liu & Wechsler 2000), analysis in eigen-spectrum shows that the limited number of training samples results in the poor estimates of the small-valued trailing eigenvalues of S_W. As these eigenvalues tend to capture noise, their inverses in the whitening step may cause LDA to fit misleading variations, as a result, overfitting occurs. Therefore, Principal Component Analysis (PCA) is used to reduce the dimensionality of input data and the principal space of S_W is chosen for face recognition (Liu & Wechsler 2000). By contrast, the null space approach (Huang et al. 2002) assumes that the most discriminative information resides in the null space of S_W. In this method, the Fisher criterion can be maximized by first removing the null space of the total scatter matrix S_T and then deriving the null space of S_W for the transformed data. One common disadvantage of the methods mentioned above is that they all ignore some discriminative information either in the null or in the principal space.

In the paper (Jiang et al. 2008), a regularization-based subspace approach is proposed for face recognition. Specifically, the eigenspace of S_W is decomposed into three subspaces: the face, noise, and null subspaces. Eigenfeatures are regularized differently in these subspaces based on an eigenspectrum model to alleviate problems of instability, overfitting, or poor generalization (Jiang et al. 2008). Due to their merit in exploiting comprehensive discriminative information, a variant of the RLDA methods is adopted in this chapter. The major steps are as follows.

- Extracting the principal space of S_T. Apply PCA on S_T and select the first n principal eigenvectors to construct the principal space P_n, where $n < r$ and r is the rank of S_T. Then $S_W^{'} = P_n^T S_W P_n$ and $S_B^{'} = P_n^T S_B P_n$.

- Subspace decomposition of $S_W^{'}$. Apply PCA on $S_W^{'}$ to derive the eigenspace $\Phi = \{\phi_i\}_{i=1}^{l}$ and the eignvalues that are sorted in descending order $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_l$. Φ can

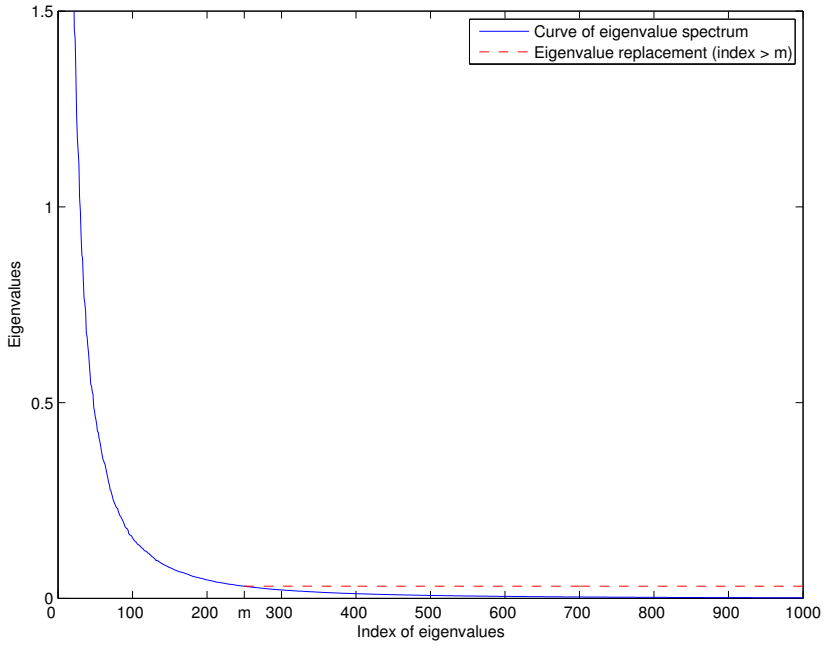

Figure 2.4 The eigenvalue spectrum of $S_W^{'}$. While the large eigenvalues are unchanged, the small eigenvalues with indices larger than m are replaced by a constant, $\rho = \lambda_{m+1}$, where λ_{m+1} is the $(m+1)th$ eigenvalue in the eigenvalue spectrum of $S_W^{'}$.

be expressed as the form of a matrix:

$$\Phi = [\phi_1, \phi_2, \ldots, \phi_l]_{n \times l} \tag{2.9}$$

where $\phi_i \in R^n$ is an eigenvector of $S_W^{'}$. The Φ of $S_W^{'}$ can be decomposed into two subspaces: a principal subspace $\{\phi_i\}_{i=1}^{m}$ and its complement, a trailing subspace $\{\phi_i\}_{i=m+1}^{l}$. The m can be estimated by using a median-based operation (Jiang et al. 2008).

- Eigenspectrum regularization. In the trailing subspace $\{\phi_i\}_{i=m+1}^{l}$, the largest eigenvalue λ_{m+1} is used to replace the remaining eigenvalues λ_i $(m+2 \leq i \leq l)$. Now, for Φ, one can form a new eigenvalue matrix Γ_m, whose diagonal elements are the union of the eigenvalues in $\{\phi_i\}_{i=1}^{k}$ and the replaced ones in $\{\phi_i\}_{i=k+1}^{m}$. An example of the regularized eigenspectrum is given in Figure 2.4.

- Discriminative feature extraction. After whitening $S_W^{'}$, the new between-class scatter matrix can be given as:

$$K_B = \Gamma_m^{-1/2} \Phi^t S_B^{'} \Phi \Gamma_m^{-1/2}. \tag{2.10}$$

After deriving the eigenspace Θ of K_B, the transformation matrix to extract the discriminative features is:

$$T = P_n \Phi \Gamma_m^{-1/2} \Theta. \tag{2.11}$$

2.4 Experiments

2.4.1 Effectiveness of the Hybrid Color Space

To evaluate the effectiveness of the hybrid color space RIQ, experiments are first implemented on the FRGC version 2 Experiment 4 and the FERET Dup I probe set using the original R, Y, I, and Q component images, and then the performance of the RIQ and YIQ color spaces are assessed by fusing the similarity matrices of their corresponding color components. For the color FERET database, 1,660 images comprising of 830 persons (two images fa/fb per person) are randomly selected to construct the training set. The gallery set fa contains 967 images. The probe set Dup I, which is used for analyzing the effect of aging on the recognition performance, is evaluated in experiments. The Dup I set contains 722 images.

Specifically, the RLDA method first processes each individual component image to derive discriminating features. These features then apply the cosine similarity measure to generate a similarity matrix. Based on the z-score (Jain et al. 2005) normalized similarity matrix, the BEE system generates three ROC curves (ROC I, ROC II, and ROC III) to derive the face verification rate for the FRGC database, while for the FERET database the algorithm generates the rank-1 accuracy. For the YIQ and RIQ color spaces, the three

Table 2.2 FRGC Version 2 Experiment 4 (ROC III) Face Verification Rates at 0.1% False Accept Rate Using Different Color Component Images and Color Spaces

Color component/space	FVR (ROC III) at 0.1% FAR	Number of m
R	63.61%	800
Y	58.64%	850
I	56.77%	600
Q	56.85%	450
YIQ	80.46%	-
RIQ	80.71%	-

Table 2.3 FERET Dup I Rank-1 Face Recognition Rates of Different Color Component Images and Color Spaces

Color component/space	Rank-1 accuracy	Number of m
R	62.18%	600
Y	54.29%	600
I	62.04%	600
Q	70.36%	600
YIQ	78.39%	-
RIQ	81.02%	-

z-score normalized similarity matrices corresponding to their component images are first fused to form a new similarity matrix using the sum rule (Kittler et al. 1998). The new similarity matrix is further normalized using the z-score normalization method and then analyzed to generate the final classification results. Tables 2.2 and 2.3 list the experimental results derived from the face verification rate of the curve ROC III at 0.1% false accept rate and the rank-1 accuracy, respectively, for the FRGC and FERET databases. The numbers of m used in RLDA are also given in two tables. The results can conclude that (i) the R component image possesses more discriminating capabilities than the Y component image, (ii) fusion of individual color component images boosts the performance significantly, and (iii) the RIQ color space achieves better face recognition performance than the YIQ color space does.

2.4.2 Multiple Frequency Feature Fusion for Face Recognition

In the method, the frequency features (the real and imaginary parts, the magnitude) are extracted from the right two quadrants in the Fourier domain, because of the symmetry property with respect to the origin. These frequency features are processed by a two-level fusion strategy, as shown in Figure 2.3, i.e., the image fusion (concatenation at the image level) for the real and imaginary parts, and the feature fusion (formation of augmented vector) for the transformed features via RLDA upon the concatenated real and imaginary parts and the magnitude. The resulting vector is $\mathbf{F}_1$ in Equation (2.8). Of course, there are three other fusion strategies to fuse the real and imaginary parts and the magnitude, i.e., image fusion, feature fusion and decision fusion.

For image fusion, the real and imaginary parts and the magnitude are first centered, respectively, and then are fused by the concatenation operation. Let μ_R, μ_I, μ_M, and δ_R, δ_I, δ_M be the mean values and standard deviations of $\mathbf{X}_R$, $\mathbf{X}_I$, and $\mathbf{X}_M$, respectively. The concatenated vector can be obtained by:

$$\mathbf{F}_2 = (\frac{\mathbf{X}_R - \mu_R}{\delta_R}; \frac{\mathbf{X}_I - \mu_I}{\delta_I}; \frac{\mathbf{X}_M - \mu_M}{\delta_M}) \tag{2.12}$$

For feature fusion, the real and imaginary parts and the magnitude are first processed by RLDA to derive the corresponding features, $\mathbf{Y}_R$, $\mathbf{Y}_I$, and $\mathbf{Y}_M$, respectively, and then these resulting features are centered and concatenated to form:

$$\mathbf{F}_3 = (\frac{\mathbf{Y}_R - \mu_R}{\delta_R}; \frac{\mathbf{Y}_I - \mu_I}{\delta_I}; \frac{\mathbf{Y}_M - \mu_M}{\delta_M}) \tag{2.13}$$

where μ_R, μ_I, μ_M, and δ_R, δ_I, δ_M are the mean values and standard deviations of $\mathbf{Y}_R$, $\mathbf{Y}_I$, and $\mathbf{Y}_M$, respectively.

For decision fusion, since the matching scores derived from the various features and classifiers are heterogenous, the score normalization is a crucial step to transform these scores into a common domain prior to combining them (Jain et al. 2005). One of the most commonly used score normalization techniques is the z-score that transforms the

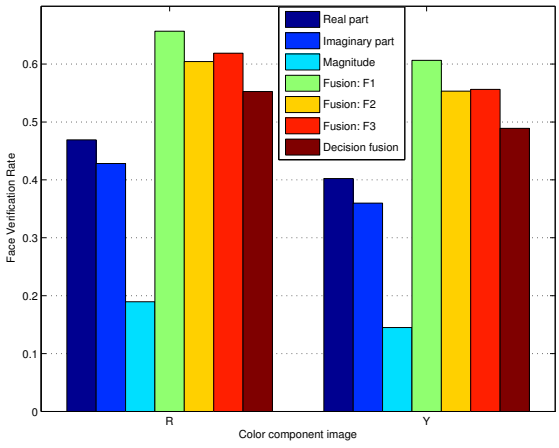

Figure 2.5 The face verification rates (ROC III) at 0.1% false accept rate of the R and Y component images using different fusion strategies to fuse the real part, the imaginary part, and the magnitude.

matching scores by calculating the arithmetic mean and standard deviation of the given data. Suppose that there exist m target vectors $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_m$. Given a query vector $\mathbf{Y}$, a similarity score vector $\mathbf{S} = [S_1, S_2, \cdots, S_m]^T = [\delta_{cos}(\mathbf{X}_1, \mathbf{Y}), \delta_{cos}(\mathbf{X}_2, \mathbf{Y}), \cdots, \delta_{cos}(\mathbf{X}_m, \mathbf{Y}),]^T$ can be obtained by calculating the cosine similarity measure between each pair of $\mathbf{X}_i$ and $\mathbf{Y}$: $\delta_{cos}(\mathbf{X}_i, \mathbf{Y}) = -\frac{\mathbf{X}_i^T \mathbf{Y}}{\|\mathbf{X}_i\| \|\mathbf{Y}\|}$. The normalized scores are given by

$$S_i' = \frac{S_i - \mu}{\delta}, \tag{2.14}$$

where μ is the arithmetic mean of $S_1, S_2, \cdots, S_m$ and δ is the corresponding standard deviation. Specifically, for a given face image, the features of the real and imaginary parts and the magnitude, $\mathbf{X}_R$, $\mathbf{X}_I$, and $\mathbf{X}_M$, are first processed by RLDA to derive the discriminating features, respectively, and then these features are used to calculate the cosine similarity scores to form three normalized similarity score vectors, $\mathbf{S}_R$, $\mathbf{S}_I$, and $\mathbf{S}_M$, by the z-score normalization, respectively. The next step is to combine these vectors by means of the sum rule, which has been demonstrated to be more effective than all the other fusion techniques on fusing the similarity scores (Kittler et al. 1998; Jain et al. 2005). The final similarity score vector is then $\mathbf{S}_R + \mathbf{S}_I + \mathbf{S}_M$.

To compare the effectiveness of these fusion strategies discussed above, Figure 2.5

Figure 2.6 Example FRGC images used in experiments that are already cropped to the size of 64 $\times$ 64 with two different scales, the scale 1 in the top row and the scale 2 in the bottom row. In particular, the images from left to right are the R, Y, I, and Q component images, respectively.

shows the experimental results of the R and Y component images using these fusion methods to fuse the real and imaginary parts and the magnitude on the FRGC database. It clearly shows that the vector $\mathbf{F}_1$, i.e., combination of two fusion strategies, achieves better performance than the others. As the magnitude is related to the real and imaginary parts, the fusion at the image level achieves somewhat lower performance than the fusion at the feature space via RLDA. Furthermore, the magnitude produces the low performance. As a result, the overall performance after fusion at the decision is the lowest. Here, the analysis results on the FRGC database are shown only, and the similar results are achieved on the FERET database.

2.4.3 Multiple Spatial Feature Fusion for Face Recognition

The multiple spatial features are contained in the multiple scaled face images. Now that the different scaled images encode different discriminating information (Hwang et al. 2006) that can produce the complementary classification outputs, the fusion of these classification outputs could further improve the overall face recognition performance. Besides the images used in the previous sections, which contain the partial face contour information, the images containing the fine face region are cropped from the original FRGC images. Figure 2.6

Table 2.4 FRGC Version 2 Experiment 4 (ROC III) Face Verification Rates at 0.1% False Accept Rate of the R, Y, I, and Q Component Images, Applying the Multiple Frequency and Spatial Feature Fusion Scheme. The Numbers of m Used in RLDA are Included in Parentheses

Method	FVR (ROC III) at 0.1% FAR							
	R		Y		I		Q	
	Scale 1	Scale 2	Scale 1	Scale 2	Scale 1	Scale 2	Scale 1	Scale 2
$\mathbf{X}_M$	18.95% (600)	21.50% (550)	14.50% (600)	15.85% (600)	16.96% (400)	15.45% (350)	9.02% (200)	10.34% (150)
$\mathbf{Y}_{RI}$	63.49% (850)	61.84% (650)	58.61% (850)	55.70% (700)	56.75% (650)	52.89% (500)	56.91% (450)	54.00% (400)
$\mathbf{F}_1$	65.68% (430)	65.41% (430)	60.65% (430)	60.04% (430)	58.50% (430)	56.00% (430)	57.22% (300)	54.98% (250)
Fusion	69.21%		64.44%		60.89%		59.64%	

shows these two scaled face images that are referred to as the scale 1 image and the scale 2 image, respectively. They both have the same resolution of 64 $\times$ 64. Note that the concept of the multiple scaled images differs from that used in the paper (Singh et al. 2008), where multiple-resolution images containing the same face region are applied for face recognition. The method in this chapter lays an emphasis on containing the different regions with an intent to generate the complementary information.

Experiments on the R, Y, I, and Q component images are conducted by using the multiple frequency feature fusion scheme across two spatial scales. The experimental results and the numbers of m in RLDA are listed in Table 2.4 and Table 2.5 for the FRGC and FERET databases, respectively. Generally speaking, the scale 1 images contain the more discriminating information than the scale 2 images for face recognition, which means that the face contour is an important factor that contributes to improving performance. The results in Table 2.4 and Table 2.5 also reveal that fusing the classification outputs of two scaled images can lead to improved performance. Finally, to evaluate the effectiveness of the proposed method utilizing the different color information, experiments are carried out by fusing the classification outputs from the individual color component images at the

Table 2.5 FERET Dup I Rank-1 Accuracy of the R, Y, I, and Q Component Images, Using the Multiple Frequency and Spatial Feature Fusion Scheme. The Numbers of m Used in RLDA are Included in Parentheses

Method	Rank-1 accuracy							
	R		Y		I		Q	
	Scale 1	Scale 2	Scale 1	Scale 2	Scale 1	Scale 2	Scale 1	Scale 2
$\mathbf{X}_M$	41.27% (170)	38.91% (175)	37.25% (195)	35.87% (155)	44.45% (163)	40.44% (171)	40.30% (167)	38.36% (141)
$\mathbf{Y}_{RI}$	61.77% (583)	57.61% (376)	57.34% (645)	53.04% (316)	63.43% (583)	59.14% (372)	71.88% (380)	67.72% (385)
$\mathbf{F}_1$	65.51% (375)	57.89% (299)	59.00% (373)	52.90% (290)	65.92% (371)	59.55% (375)	70.63% (388)	67.31% (386)
Fusion	66.20%		60.39%		65.10%		75.21%	

Table 2.6 Experimental Results of the RIQ and YIQ Color Spaces Using the Proposed Method

Color space	FRGC ver. 2 Exp. 4			FERET Dup I
	FVR (ROC I)	FVR (ROC II)	FVR (ROC III)	Rank-1
RIQ	82.99%	82.92%	82.49%	81.30%
YIQ	82.04%	82.03%	81.95%	79.50%

decision level. The final results are shown in Table 2.6. As can be seen, applying the multiple color, frequency and spatial feature fusion is indeed able to improve face recognition performance significantly.

2.5 Conclusion

This chapter presents a novel multiple feature fusion method that fuses the frequency, spatial and color features in a hybrid color space to improve the face recognition grand challenge performance. In particular, the frequency features are extracted from the magnitude, the real and imaginary parts in the Fourier domain of an image; the spatial features are derived from two different scales of a face image; and the color features are from a hybrid color space, namely, the RIQ color space. To extract the discriminative features efficiently,

a variant of RLDA is presented in this chapter. Experiments show that the proposed method achieves the face verification rate (ROC III) of 82.49% at the false accept rate of 0.1% on the FRGC version 2 Experiment 4 and the rank-1 accuracy of 81.30% on the FERET Dup I probe set.

CHAPTER 3

FUSING LBP AND COLOR FEATURES FOR FACE RECOGNITION

This chapter presents a novel LBP-based method, which fuses the complementary features derived from the multiscale LBP histograms in the hybrid VIQ color space, for face recognition. The complementary characteristics of the features thus come from both the multiple imaging (three component images) in the VIQ color space and the multiple spatial scales (different LBP operators) in the LBP features.

Face recognition has been an attractive research topic in pattern recognition, computer vision, and machine learning in the last two decades. However, improving the recognition performance under difficult conditions (such as illumination changes, pose variations, and aging) is still an open issue. Substantial efforts have been put into developing sophisticated nonlinear learning methods, e.g., kernel-based methods and manifold learning, to enhance the generalization capability of algorithms, at the expense of computational efficiency. This chapter shows that the significant improvement of performance under difficult conditions can be achieved easily, by utilizing some simple but effective features without resort to the complex learning methods.

The method is straightforward and can be described briefly as the “Color + LBP (Ahonen et al. 2006) + LDA" strategy. For extracting effective image features, a hybrid color space VIQ is selected, which is able to provide an intrinsically complementary representation for each face image based on dependence analysis. On each component image of VIQ, a multiscale LBP fusion scheme is proposed to integrate the LBP histogram features at three scales for increasing the description capability of LBP operators for face skin texture. For extracting statistical features, a variant of Regularized Linear Discriminant Analysis (RLDA), which operates in both the principal and null spaces of the within-class scatter matrix, is adopted to extract the complete discriminative information for classification.

3.1 Independence Analysis for Selecting Color Spaces for Face Recognition

Face recognition using different imaging methodologies has become an area of increasing interest. One representative example is to fuse visible and thermal infrared (IR) images, by capturing complementary information of reflectance and radiation from the face (Bebis et al. 2006). However, the high cost of IR sensors may pose a potential limitation to implement the IR-based methods in practical face recognition. Recently, advances have focused on using color information, which offers a promising alternative to IR imaging to generate multiple representations for the face, to improve the performance of face recognition (Yang et al. 2008; Yang et al. 2010a). This section discusses the selection of color information in face recognition based on dependence analysis.

Analyzing the dependence among different feature components is an important and effective criterion for the process of feature selection. One of important aspects in designing color-based face recognition methods is to select appropriate color spaces, whose high independence among color components makes the features extracted from different color components as mutually complementary and uncorrelated as possible. As a result, data fusion either in feature level or in decision level across different color components usually yields the improved recognition performance when compared to the individual color components. In information theory, the mutual information is a quantity that measures the mutual dependence of two random variables X and Y. Formally, the definition for the discrete case can be given as:

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in Y} p(x,y) \log\left(\frac{p(x,y)}{p_1(x)p_2(y)}\right), \tag{3.1}$$

where $p(x,y)$ is a joint density and $p_1(x)$ and $p_2(y)$ are the marginal densities.

Table 1.1 has shown that both R and V (in HSV) color components can achieve better accuracy than the luminance Y in face recognition. Thus, another hybrid color space VIQ is considered for face recognition in this chapter. The dependence information between different color components is quantitatively analyzed, based on the training set of

Table 3.1 Mutual Information in Different Color Spaces

Color space	Mutual information I
RGB	0.4037
VIQ	0.0855

Figure 3.1 Color component images for two subjects in the FERET (top row) and FRGC (bottom row) databases, respectively. From left to right, color component images: R, G, B, V, I, and Q.

the Face Recognition Grand Challenge (FRGC) database, a large-scale face database with complex image variations. Specifically, all the training images in one color component are concatenated into a long feature vector. Then the joint and marginal densities $p(x,y)$, $p_1(x)$, and $p_2(y)$ are estimated from any pair of feature vectors. Finally, the mutual information between two feature vectors is computed using Equation (3.1). Table 3.1 shows the average values of mutual information of three color components in different color spaces. The closer the value I comes to zero, the more mutual independence information exists among features. Thus, the hybrid VIQ color space is better than the RGB color space to generate a multiple representation of face images for classification. Some examples of color component images are provided in Figure 3.1. Generally, a competent color space for face recognition consists of one luminance component (such as Y, R, and V) and two chrominance components (Yang et al. 2010a).

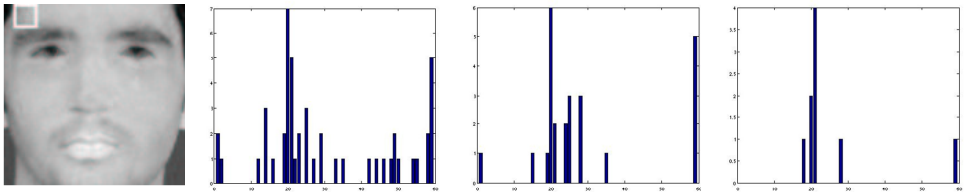

Figure 3.2 Three LBP histograms corresponding to the three LBP operators: $LBP^{u2}_{8,1}$, $LBP^{u2}_{8,2}$, and $LBP^{u2}_{8,3}$, in a subwindow of 9×9 pixels.

3.2 Fusion of Multiscale LBP Features

Using local features for face recognition has been attracting substantial attention because they are more robust against variations in pose and illumination in patch images than holistic features. As a representative approach in local feature category, Local Binary Patterns (LBP) has shown great capabilities in describing face images for classification due to its invariance to monotonic gray-level changes and computational efficiency (Ahonen et al. 2006).

In a 3×3 neighborhood of an image, the basic LBP operator generates a binary string by thresholding each surrounding pixel i_p with the gray value of the central pixel i_c. Formally, the LBP operator is defined as follows:

$$LBP = \sum_{p=0}^{7} 2^p s(i_p - i_c), \tag{3.2}$$

where $s(i_p - i_c)$ equals 1, if $i_p - i_c \geq 0$; and 0, otherwise. After the introduction of *uniform patterns* (Ojala et al. 2002), LBP can be expressed as: $LBP^{u2}_{P,R}$, where P, R means P sampling points on a circle of radius R. The histograms obtained from a LBP operator in local image regions are usually concatenated into a global histogram for face recognition. Figure 3.2 shows three local histograms derived from three scale LBP operators in a subwindow of an image. Most of LBP-based methods measure the similarity between two face images using histogram matching, which has been proved to be inadequate to utilizing the discriminative information in histograms for classification when compared to statistical methods, such as

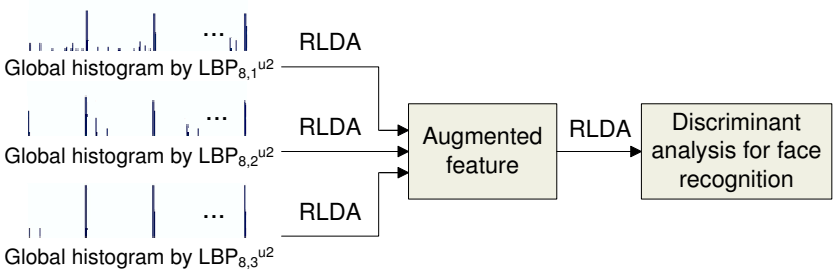

Figure 3.3 Multi-resolution LBP feature fusion scheme.

LDA (Shan et al. 2006).

In this chapter, a multiscale LBP fusion scheme is proposed based on statistical method. Note that the Gabor image representation encompasses the features corresponding to five scales for improving face recognition performance. Inspired by this idea, the multiple-resolution information from the LBP operators is combined. First, three LBP operators, $LBP^{u2}_{8,1}$, $LBP^{u2}_{8,2}$, and $LBP^{u2}_{8,3}$, are used to extract the multi-resolution histogram features from the image. It is easy to see from Figure 3.2 that these histograms are complementary to one another. Smaller scale operators extract more detailed information (microstructure) and maintain the similar profile (macrostructure) as larger operators do. Second, three global histogram features are fused to form an augmented feature. One straightforward way is to concatenate the three global histograms, corresponding to $LBP^{u2}_{8,1}$, $LBP^{u2}_{8,2}$, and $LBP^{u2}_{8,3}$. However, this operation will result in the problem of high dimensionality. In the proposed method, a LBP multiple-resolution feature fusion scheme, as shown in Figure 3.3, is proposed. For each global LBP histogram, a variant of RLDA is used to extract features and reduce dimensionality. Let X_{h1}, X_{h2}, and X_{h3} be the reduced features after the RLDA process. In particular, the three reduced features first are normalize and then are concatenated into an augmented feature vector, $Y = (\frac{X_{h1}-\mu_1}{\delta_1}; \frac{X_{h2}-\mu_2}{\delta_2}; \frac{X_{h3}-\mu_3}{\delta_3};)$, where μ_i and δ_i are the mean and standard deviation of feature X_{hi}. By applying this fusion scheme, both the microstructures and the macrostructures of face image are utilized to extract the

discriminating features, which contain much more face information than what a single LBP operator can provide.

3.3 Illumination Normalization Procedures

One of the most challenging variations on face image is the illumination variation. The dramatic changes of face appearance caused by illumination variation pose the complexity to the feature extraction of face image, hence leading to the degradation of face recognition performance. Therefore, the preprocessing to alleviate the effect of illumination variations is an essential procedure for a robust and reliable face recognition algorithm. This section presents an efficient illumination normalization method, which comprises the adjustment of DCT coefficients in the logarithm domain (Chen et al. 2006), the Difference of Gaussian (DoG) filtering, and the contrast equalization (Tan & Triggs 2007a).

An image $f(x,y)$ may be characterized by the production of illumination component $i(x,y)$ and reflectance component $r(x,y)$ (Gonzalez & Woods 2002):

$$f(x,y) = i(x,y)r(x,y). \tag{3.3}$$

Usually, Equation (3.3) is operated by taking the logarithm transform on the left and right sides, so that the frequency components of illumination and reflectance can be processed separately (Gonzalez & Woods 2002). Thus the subsequent procedures will occur in the logarithm domain.

The illumination and reflectance components have their own characteristics. The illumination component of an image generally makes a feature of slow spatial variations, while the reflectance component inclines to change quickly, particularly at the junctions of dissimilar objects (Gonzalez & Woods 2002). This differentiation determines that one can recover the reflectance of faces and remove the effect of illumination by the frequency analysis. Rather than using the Fourier transform in Homomorphic filter, the method resorts

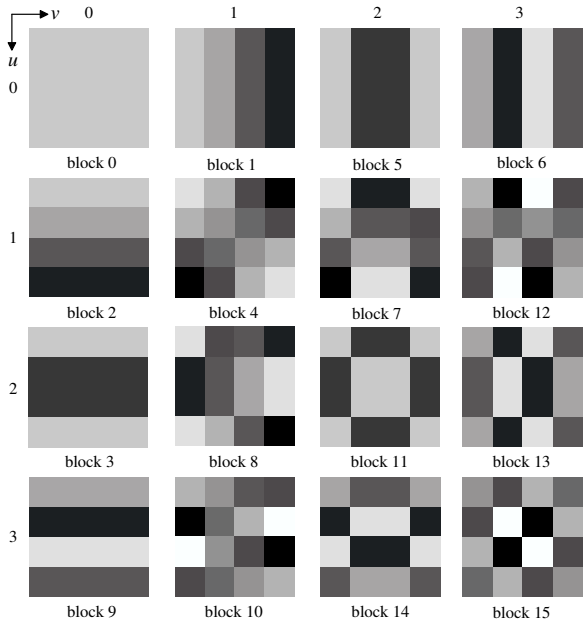

Figure 3.4 Discrete-cosine basis functions for $N = 4$.

to the Discrete Cosine Transform (DCT), which can demonstrate a straightforward explanation that the intensity variations of an image are encoded in the discrete-cosine basis functions. Besides, the recent research (Chen et al. 2006) has revealed that DCT has some advantages over the Fourier transform to remove the illumination effects on face images.

Figure 3.4 shows an example of discrete-cosine basis functions for the case $N = 4$ (for the images of size 4×4) (Gonzalez & Woods 2002). It can be seen that (i) block 0 represents the overall illumination of an image, i.e., the average intensity, (ii) block 1 and block 2 represent the intensity variations along the horizontal and vertical orientations, respectively, (iii) while block 4 represents the intensity variations in both orientations. Based on these facts, most effects of illumination variations upon face images can be alleviated by adjusting the coefficient values obtained using these three basis functions, i.e., block 1, block 2, and block 4, as shown in Figure 3.4.

Let $C(0,1)$, $C(1,0)$, and $C(1,1)$ be the coefficients corresponding to three basis functions after the DCT transform of a face image. Since these coefficients represent the illumination variations, each of them is set to a value of zero, i.e., $C(0,1) = 0$, $C(1,0) = 0$,

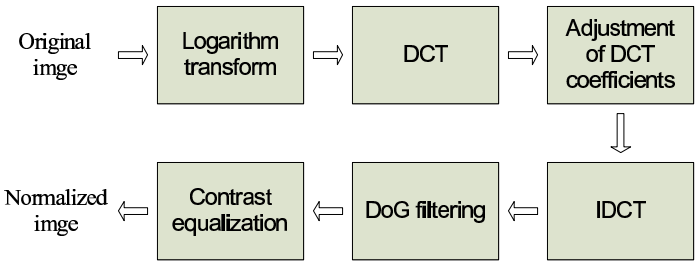

Figure 3.5 The diagram of illumination normalization procedures.

Figure 3.6 Examples of the illumination normalized gray-scale images. The first column: original gray images; the second column: IDCT reconstructed images in the logarithm domain; the last column: the normalized images after DoG filtering and contrast equalization.

and $C(1,1) = 0$, and then the inverse DCT transform is conducted to form the reconstructed image in the logarithm domain. Generally, the adjustment of DCT coefficients still can not remove some shadow effects, caused by the 3D structure of face surface on some small areas, which make the spatial details unclear. Thus, the Difference of Gaussian (DoG) filtering is applied to preserve the facial details, particularly in the shading regions. Finally, a contrast equalization procedure used in (Tan & Triggs 2007a) transforms the intensity values of images to the specified range. Figure 3.5 summarizes the illumination normalization procedures presented above.

Figure 3.6 shows some examples of the illumination normalized face images. Note

that the illumination normalization method is applied to Y and V but not to I and Q component images, because the chromatic component images I and Q do not hold the characteristic of illumination-reflectance model.

3.4 Experiments

This section assesses the proposed method using two large-scale face databases, the FERET and FRGC databases (Phillips et al. 2000; Phillips et al. 2005). For the color FERET, 1,000 images comprising of 500 persons (two images fa/fb per person) are randomly selected to construct the training set. The image number of the gallery set is 967. The two most challenging probe sets, Dup I and Dup II, which are used for analyzing the effect of aging on the recognition performance, are evaluated in experiments. Their image numbers are 722 and 228, respectively. In the FRGC version 2 database Experiment 4, the training set contains 12,776 images that are either controlled or uncontrolled. The target set has 16,028 controlled images and the query set has 8,014 uncontrolled images. The size of face images used in experiments is 64×64.

3.4.1 Evaluation of Color Spaces for Face Recognition

This set of experiments evaluates the effectiveness of the different color spaces on face recognition. An LDA-based method with an Enhanced Fisher Model (EFM) (Liu & Wechsler 2000) and the cosine similarity measure are used. Table 3.2 shows the performances of the color components R, G, B, V, Y, I, and Q in face recognition. As can be seen, the commonly used luminance Y is not ideally suited for face recognition, as its performance is lower than those of the color components R and V in both the FERET and FRGC databases. To evaluate the effect of different color spaces on face recognition performance, the z-score normalization technique (Jain et al. 2005) is first used to normalize the similarity scores generated by the different color images, and then the normalized similarity scores are fused

Table 3.2 Experimental Results Using Different Color Components

Color component & space	FERET Rank-1 accuracy		FRGC FVR (ROC III)
	Dup I	Dup II	Exp. 4
R	52.8%	29.8%	62.1%
G	46.4%	22.0%	53.8%
B	44.0%	20.6%	40.0%
V	52.8%	29.8%	61.8%
Y	49.2%	28.0%	56.4%
I	57.6%	39.9%	54.2%
Q	63.4%	49.6%	54.0%
RGB	49.8%	24.1%	59.1%
YIQ	70.0%	47.4%	78.6%
VIQ	72.0%	55.3%	79.0%

via the sum rule for classification. The results in Table 3.2 indicate that the face recognition performance can be improved significantly using the data fusion in color spaces. Specifically, the hybrid color space VIQ achieves the best performance among the color spaces used in this chapter.

Table 3.2 leads to the following conclusions. The individual color components have a high variability in face recognition performance with varying illumination. For example, when luminance Y is considered as benchmark, the I and Q components achieve higher performance in the FERET database while lower performance in the FRGC database, owing to the different illumination conditions. It is this characteristic that causes a controversy about the feasibility of color information in face recognition. However, when data fusion is applied in color spaces, the performance improvement can be expected in most cases, due to the mutual independence between color components. Therefore, the color-based methods should use the color spaces instead of the individual color components for face recognition.

Table 3.3 Experimental Results Using the LBP Features

Database		Color	$LBP^{u2}_{8,1}$	$LBP^{u2}_{8,2}$	$LBP^{u2}_{8,3}$	Fusion
FERET (Rank-1)	Dup I	V	68.4%	72.0%	71.3%	74.4%
		I	70.1%	71.7%	69.7%	79.9%
		Q	61.5%	65.9%	66.1%	70.9%
	Dup II	V	40.8%	54.4%	58.8%	60.1%
		I	56.1%	61.0%	58.8%	71.1%
		Q	55.7%	62.7%	67.5%	71.9%
FRGC (FVR ROCIII)	Exp. 4	V	69.4%	70.3%	63.0%	73.5%
		I	54.8%	61.2%	55.1%	65.8%
		Q	51.9%	57.2%	50.1%	63.8%

Table 3.4 Experimental Results Using the Proposed Method

FERET (Rank-1)		FRGC (FVR ROCIII)
Dup I	Dup II	Exp. 4
87.1%	85.5%	85.6%

3.4.2 Experiments with the Proposed Method

In the proposed method, each component image of VIQ is used to generate the LBP features. Specifically, a face image of size 64×64 is divided into 144 (12×12) overlapping windows of 9×9 pixels (3 pixels overlapping). Then the multiplescale LBP fusion scheme works on each component image for face recognition. Finally, three similarity scores derived by three component images are fused via the sum rule for the final classification. Experimental results based on the LBP features and the proposed method are given in Table 3.3 and Table 3.4, respectively. As can be seen, the LBP features improve the recognition performance not only on V but also on I and Q components. The proposed multiplescale LBP fusion scheme is effective for further improving the recognition performance in comparison with a single LBP operator, and the proposed method can boost the performance significantly.

Table 3.5 Experimental Results Using the LBP Features on the Normalized V Image

Method	FERET Rank-1 accuracy		FRGC FVR (ROC III)
	Dup I	Dup II	Exp. 4
$LBP^{u2}_{8,1}$	83.7%	83.7%	70.4%
$LBP^{u2}_{8,2}$	82.1%	84.6%	71.0%
$LBP^{u2}_{8,3}$	74.8%	74.6%	61.6%
Fusion	86.0%	86.4%	74.5%

Table 3.6 Comparison of the Proposed Method with the Others

Method	FERET Rank-1 accuracy		FRGC FVR (ROC III)
	Dup I	Dup II	Exp. 4
Yao et al. 2008	87.0%	85.0%	N/A
Tan et al. 2007(b)	90.0%	85.0%	83.6%
Su et al. 2007	N/A	N/A	85.8%
Shan et al. 2006	92.0%	88.9%	N/A
In this chapter	91.0%	91.2%	86.1%

3.4.3 Experiments with the Illumination Normalization on the V Images

The illumination normalization method is applied to the luminance-like component image V for eliminating the effect of illumination variations, and the associated experimental results using the LBP features are provided in Table 3.5. It is demonstrated that the illumination normalization is helpful for improving the discrimination capabilities of the LBP features, especially on the FERET database. Finally, the proposed method is compared with some state-of-the-art methods in Table 3.6. As can be seen, the proposed method can achieve the competitive performance on two large-scale face databases, the FERET and FRGC databases.

It is of interest to investigate the feasibility of classifying the test set in one face database using the training set in another database. As the FRGC training set contain more complex conditions in image variabilities, it is chosen to train the classifiers in the proposed

Table 3.7 The Rank-1 Accuracy of the FERET Database Using the FRGC Training Set

Dataset	Color	LBP	FERET rank-1 accuracy		
Dup I	V	$LBP^{u2}_{8,1}$	82.68%	86.28%	92.80%
		$LBP^{u2}_{8,2}$	83.10%		
		$LBP^{u2}_{8,3}$	67.72%		
	I	$LBP^{u2}_{8,1}$	82.54%	83.10%	
		$LBP^{u2}_{8,2}$	79.91%		
		$LBP^{u2}_{8,3}$	73.40%		
	Q	$LBP^{u2}_{8,1}$	68.00%	75.34%	
		$LBP^{u2}_{8,2}$	72.29%		
		$LBP^{u2}_{8,3}$	64.54%		
Dup II	V	$LBP^{u2}_{8,1}$	73.24%	82.89%	90.35%
		$LBP^{u2}_{8,2}$	78.94%		
		$LBP^{u2}_{8,3}$	62.71%		
	I	$LBP^{u2}_{8,1}$	68.42%	70.17%	
		$LBP^{u2}_{8,2}$	65.35%		
		$LBP^{u2}_{8,3}$	59.21%		
	Q	$LBP^{u2}_{8,1}$	60.96%	68.42%	
		$LBP^{u2}_{8,2}$	61.40%		
		$LBP^{u2}_{8,3}$	55.70%		

method for testing the FERET probe sets. In particular, the parameters used in this set of experiments are the same as those in the foregoing experiments. Table 3.7 shows the the rank-1 accuracy on the FERET database, indicating the performance is still pretty high even though the different training set is used. Note that the V component image is processed via the illumination normalization. Meanwhile, the experiments imply that the proposed method could provide a potential solution to the *one sample per person* problem (Tan et al. 2006), where only one available training sample per person poses grand challenges to the face recognition performance.

3.5 Conclusion

A novel method for face recognition has been proposed in this chapter. The large-scale experiments on the FERET and FRGC databases have shown the strategy “Color + LBP + LDA" is effective to improve the recognition performance significantly under the difficult illumination and aging conditions. In particular, the FRGC as training set has been applied to test the FERET probe sets. The excellent experimental results show that the proposed method could provide a promising alternative for the *one sample per person* problem, an often encountered challenge in practical face recognition.

CHAPTER 4

EXTRACTING MULTIPLE FEATURES FOR FACE RECOGNITION

Face recognition has been intensely studied for more than two decades. Although numerous methods have been developed, there are still many difficulties in addressing complex image variabilities caused by illumination, pose, facial expression, aging, etc. As one strives to prevent these challenges from compromising the performance in some traditional methods that usually use a single feature, an effective face recognition method may rely on multiple features. Thus, efficiently exploiting and extracting multiple facial features are crucial for face recognition. If the multiple features are complementary to each other, the diversity among misclassifications of the corresponding classifiers could be enhanced (Kittler et al. 1998). As a result, the fusion at the feature level or the decision level can improve the performance of face recognition. The complementarity between feature sets can be achieved by extracting the holistic and local features (Su et al. 2007) as well as utilizing the multiple component images in a color space (Liu et al. 2008; Yang et al. 2008).

Color information has been widely used for face detection (Hsu et al. 2002). Recently, some efforts have been put into utilizing color information for improving the accuracy in face recognition. For example, the quaternion domain has been applied to the RGB color space for face recognition (Xie et al. 2005; Jones et al. 2006). Research in (Shih et al. 2005) reveals that different color spaces transformed from the RGB color space display different discriminating power for face recognition. Specifically, the YUV and YIQ color spaces provide better face recognition performance in comparison with other color spaces (Shih et al. 2005). Besides, the color transformation can be learned by the statistical methods, such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Independent Component Analysis (ICA), in order to derive some effective color image representations for face recognition purpose (Liu 2008; Yang et al. 2008).

The motivation behind this work is to exploit the multiple complementary facial

features in a novel discriminant color space, so as to improve face recognition performance through fusing the classification outputs of multiple features at the decision level. First, a novel CID color space is constructed by an iterative discriminant analysis upon the Color Image Discriminant (CID) model (Yang et al. 2008), aimed at generating the new component images D^1, D^2, and D^3 that are optimal with respect to a discriminant criterion. As three new component images display the different image characteristics and discriminating capacities, three effective methods are proposed to extract the features on the component images, respectively. They are (1) the patch-based Gabor image representation for the D^1 component image, (2) the multi-resolution LBP feature fusion for the D^2 component image, and (3) the DCT-based multiple face encodings for the D^3 component image. Second, the component images of CID are orthogonal to each other and the three types of resulting feature sets are complementary. Such properties determine that the sets of face images misclassified by the different feature sets upon the corresponding component images would not necessarily overlap. This implies that at the decision level the derived similarity scores can be fused to enhance discriminating power for face recognition.

4.1 Color Image Discriminant (CID) Model

The CID model (Yang et al. 2008) is developed on the basis of the RGB color space, since it is a fundamental color space, from which a number of color spaces can be generated. Let $\mathbf{A}$ be a color image with a size of $m \times n$ and its three primary color components be $\mathbf{R}$, $\mathbf{G}$, and $\mathbf{B}$. Without loss of generality, let $\mathbf{R}$, $\mathbf{G}$, and $\mathbf{B}$ be column vectors: $\mathbf{R}, \mathbf{G}, \mathbf{B} \in \mathbb{R}^N$, where $N = m \times n$. The color image $\mathbf{A}$ is then expressed as an $N \times 3$ matrix: $\mathbf{A} = [\mathbf{R}, \mathbf{G}, \mathbf{B}] \in \mathbb{R}^{N \times 3}$.

The CID model seeks a set of optimal coefficients x_1, x_2, and x_3 to linearly combine the color components, $\mathbf{R}$, $\mathbf{G}$, and $\mathbf{B}$, such that the generated component is an optimal representation of the color image $\mathbf{A}$ with respect to a discriminant criterion. Specifically, Let c be the number of pattern classes, $\mathbf{A}_{ij}$ be the j-th color image in class i, where $i = 1, 2, ..., c$, and $j = 1, 2, ..., M_i$, and M_i denotes the number of training samples in class i. Let $\bar{\mathbf{A}}_i$ be

the mean image of training samples in class i and $\bar{\mathbf{A}}$ be the mean image across all training samples. By combining the three color components of the color image $\mathbf{A}_{ij} = [\mathbf{R}_{ij}, \mathbf{G}_{ij}, \mathbf{B}_{ij}]$, a new image is given as

$$\mathbf{D}_{ij} = x_1\mathbf{R}_{ij} + x_2\mathbf{G}_{ij} + x_3\mathbf{B}_{ij} = [\mathbf{R}_{ij}, \mathbf{G}_{ij}, \mathbf{B}_{ij}]\mathbf{X}, \tag{4.1}$$

where $\mathbf{X} = [x_1, x_2, x_3]^T$ is a vector of color component combination coefficients.

Let $\bar{\mathbf{D}}_i$ be the mean vector of combinative component images in class i and $\bar{\mathbf{D}}$ be the grand mean vector. It is evident that $\bar{\mathbf{D}}_i = \bar{\mathbf{A}}_i\mathbf{X}$ and $\bar{\mathbf{D}} = \bar{\mathbf{A}}\mathbf{X}$. The between-class scatter matrix $\mathbf{S}_b(\mathbf{X})$ and the within-class scatter matrix $\mathbf{S}_w(\mathbf{X})$ in the D-space are defined as follows:

$$\mathbf{S}_b(\mathbf{X}) = \sum_{i=1}^{c} P_i[(\bar{\mathbf{A}}_i - \bar{\mathbf{A}})\mathbf{X}\mathbf{X}^T(\bar{\mathbf{A}}_i - \bar{\mathbf{A}})^T], \tag{4.2}$$

$$\mathbf{S}_w(\mathbf{X}) = \sum_{i=1}^{c} P_i \frac{1}{M_i - 1} \sum_{j=1}^{M_i} [(\mathbf{A}_{ij} - \bar{\mathbf{A}}_i)\mathbf{X}\mathbf{X}^T(\mathbf{A}_{ij} - \bar{\mathbf{A}}_i)^T], \tag{4.3}$$

where P_i is the *a priori* probability for class i and commonly evaluated as $P_i = M_i/M$, $M = \Sigma_{i=1}^{c} M_i$.

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

$$J(\mathbf{P}, \mathbf{X}) = \frac{|\mathbf{P}^T\mathbf{S}_b(\mathbf{X})\mathbf{P}|}{|\mathbf{P}^T\mathbf{S}_w(\mathbf{X})\mathbf{P}|}, \tag{4.4}$$

where $|\cdot|$ denotes the determinant operator, and $\mathbf{P}$ is an $N \times d$ transformation matrix that is formed by a set of projection basis vectors $\phi_1, \phi_2, \ldots, \phi_d$.

Maximizing the general Fisher criterion is equivalent to solving the problem of twin generalized eigenvalues as follows (Yang et al. 2008):

$$\begin{cases} \mathbf{S}_b(\mathbf{X})\phi = \lambda\mathbf{S}_w(\mathbf{X})\phi \\ \mathbf{L}_b(\mathbf{P})\mathbf{X} = \mu\mathbf{L}_w(\mathbf{P})\mathbf{X}, \end{cases} \tag{4.5}$$

where $\mathbf{L}_b(\mathbf{P})$ and $\mathbf{L}_w(\mathbf{P})$ are, respectively, the color-space between-class and within-class

scatter matrices defined as follows:

$$\mathbf{L}_b(\mathbf{P}) = \sum_{i=1}^{c} P_i[(\bar{\mathbf{A}}_i - \bar{\mathbf{A}})^T \mathbf{P}\mathbf{P}^T(\bar{\mathbf{A}}_i - \bar{\mathbf{A}})], \tag{4.6}$$

$$\mathbf{L}_w(\mathbf{P}) = \sum_{i=1}^{c} P_i \frac{1}{M_i - 1} \sum_{j=1}^{M_i} [(\mathbf{A}_{ij} - \bar{\mathbf{A}}_i)^T \mathbf{P}\mathbf{P}^T(\mathbf{A}_{ij} - \bar{\mathbf{A}}_i)]. \tag{4.7}$$

Then the problem in Equation (4.5) can be solved iteratively by the following CID algorithm:

The CID Algorithm:

---

1. Set $k = 0$, and provide an initial value for $\mathbf{X} : \mathbf{X} = \mathbf{X}^{[k]}$.
2. Construct $\mathbf{S}_b(\mathbf{X})$ and $\mathbf{S}_w(\mathbf{X})$, and calculate their generalized eigenvectors $\phi_1, \phi_2, \ldots, \phi_d$ corresponding to the d largest eigenvalues. Let $\mathbf{P} = \mathbf{P}^{[k+1]} = [\phi_1, \phi_2, \ldots, \phi_d]$.
3. Construct $\mathbf{L}_b(\mathbf{P})$ and $\mathbf{L}_w(\mathbf{P})$, and calculate their generalized eigenvector $\mathbf{X}^{[k+1]}$ corresponding to the largest eigenvalue.
4. If $|J(\mathbf{P}^{[k+1]}, \mathbf{X}^{[k+1]}) - J(\mathbf{P}^{[k]}, \mathbf{X}^{[k]})| < \varepsilon$, the iteration terminates and let $\mathbf{P}^* = \mathbf{P}^{[k+1]}$ and $\mathbf{X}^* = \mathbf{X}^{[k+1]}$. Otherwise, let $\mathbf{X} = \mathbf{X}^{[k+1]}$ and go to step 1.

Using the CID algorithm, an optimal color component combination coefficient vector $\mathbf{X}_1 = \mathbf{X}^* = [x_{11}, x_{21}, x_{31}]^T$ can be obtained. Actually, the other two color component combination coefficient vectors, $\mathbf{X}_2$ and $\mathbf{X}_3$, can be obtained similarly under the $\mathbf{L}_w(\mathbf{P})$-orthogonal constraint. The CID color space is then defined by the following transformation:

$$[\mathbf{D}^1, \mathbf{D}^2, \mathbf{D}^3] = [\mathbf{R}, \mathbf{G}, \mathbf{B}][\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3], \tag{4.8}$$

where $\mathbf{D}^1$, $\mathbf{D}^2$, and $\mathbf{D}^3$ are the three color components of image $\mathbf{A} = [\mathbf{R}, \mathbf{G}, \mathbf{B}]$ in the CID color space.

Figure 4.1 The color component images R, G, and B (top row), and the new color component images D^1, D^2, and D^3 (bottom row) for one subject.

4.2 Extracting Multiple Features in the CID Color Space for Face Recognition

The color component combination coefficients, $\mathbf{X}_1$, $\mathbf{X}_2$, and $\mathbf{X}_3$, can be learned from the standard training set of the Face Recognition Grand Challenge (FRGC) version 2 database Experiment 4 (Phillips et al. 2005), which contains 12,776 face images of 222 subjects. The three resulting coefficient vectors are $\mathbf{X}_1 = [1.0000, 0.0138, -0.2079]^T$, $\mathbf{X}_2 = [-0.8620, 0.9622, -0.1003]^T$, and $\mathbf{X}_3 = [0.0353, -0.9519, 0.9165]^T$, respectively. Based on these coefficients, the three discriminating color components D^1, D^2, and D^3 can be obtained for each color image. Figure 4.1 shows an example of the R, G, and B images and the three new color component images for one subject. Since the new images possess different discriminating properties and different face representations, it is feasible to apply different feature extraction methods to the D^1, D^2, and D^3 images, respectively. On the other hand, multiple feature sets, which potentially offer complementary information about the face images to be classified, could be utilized to improve the classification performance by the data fusion at the feature level or the decision level (Kittler et al. 1998). Therefore, this section proposes three effective image encoding methods to extract the multiple feature sets on the new color component images.

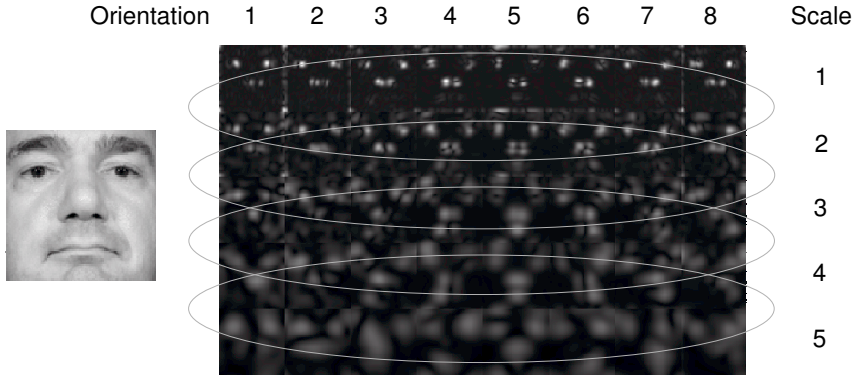

Figure 4.2 GIR derived from the convolution of one D^1 face image with the Gabor kernels with five scales and eight orientations.

4.2.1 The Patch-based Gabor Image Representation for the D^1 Image

Gabor Image Representation (GIR) of an image is the convolution of image with a family of Gabor kernels that commonly contain five scales and eight orientations for face recognition (Daugman 1985):

$$\psi_{\mu,\nu}(z) = \frac{\|k_{\mu,\nu}\|^2}{\sigma^2} e^{-\frac{\|k_{\mu,\nu}\|^2 \|z\|^2}{2\sigma^2}} \left[e^{ik_{\mu,\nu} z} - e^{-\frac{\sigma^2}{2}} \right], \tag{4.9}$$

where $\mu \in \{0,...,7\}$ and $\nu \in \{0,...,4\}$ define the orientation and scale of Gabor kernels, $z = (x,y)$, and $\|\cdot\|$ denotes the norm operator. The wave vector $k_{\mu,\nu}$ is defined as: $k_{\mu,\nu} = k_\nu e^{i\phi_\mu}$. Figure 4.2 shows GIR of one D^1 face image. Note that the D^1 component image in Figure 4.2 has the fine face region that is helpful for the extraction of the Gabor features without the face contour information.

The edge of GIR arises from the combination of different spatial frequencies, spatial localities, and orientation selectivities. GIR is thus able to classify faces using rich information, which can be applied to both local and holistic feature extractions. Figure 4.3 shows the outline of face recognition based on GIR.

For the extraction of local GIR feature, GIR is separated into an ensemble of the patches along the horizontal direction, with the images in two adjacent scales and all the

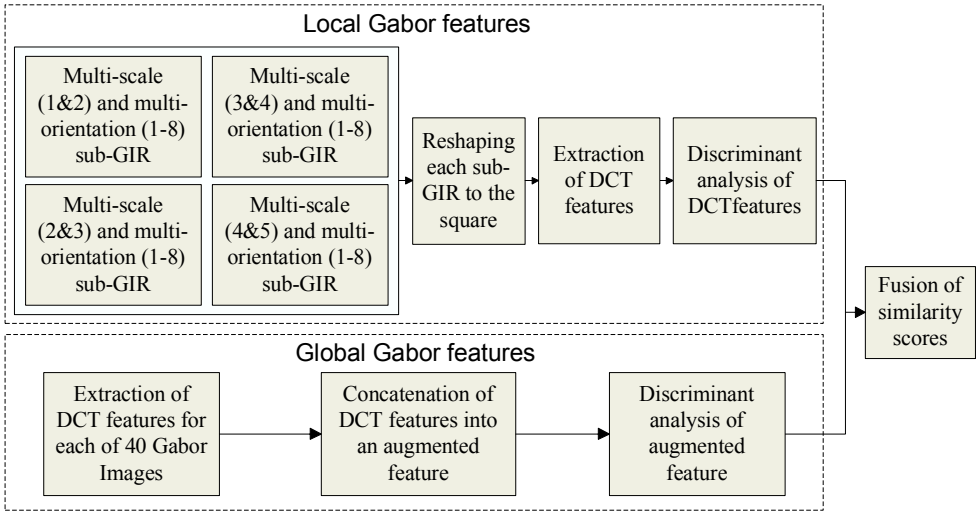

Figure 4.3 The extraction of local and global Gabor features from GIR for face recognition.

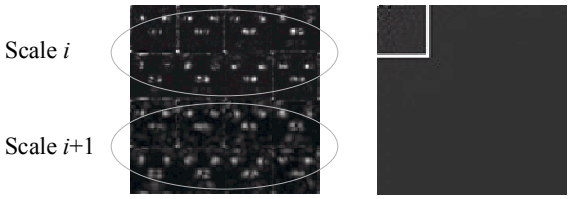

Figure 4.4 A reshaped sub-GIR patch image with $i \in \{1,2,3,4\}$, and a subset of frequencies in the DCT domain to encode the sub-GIR patch images.

eight orientations forming one sub-GIR as illustrated in Figure 4.2. The rationale of grouping two adjacent scales is due to the assumption that the redundancy in GIR is caused mainly by the similarities between all of the adjacent Gabor filters. Discrete Cosine Transform (DCT) can thus be used to reduce dimensionality and redundancy for improving computational efficiency and recognition performance. To facilitate the DCT feature extraction, each sub-GIR patch image is first reshaped to a square as shown in Figure 4.4. After transforming a reshaped sub-GIR patch image to the DCT domain, a frequency set selection scheme via a square mask is used to select a set of the low frequency features located in the upper-left corner. Then this feature set undergoes the linear discriminant analysis by an Enhanced Fisher Model (EFM) (Liu & Wechsler 2000) for face recognition.

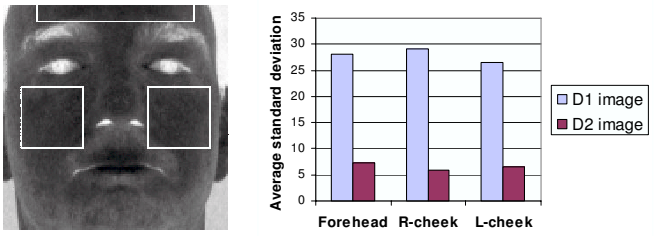

Figure 4.5 The forehead, right cheek and left cheek regions on one D^2 image and their average standard deviations of intensity values of all the D^1 and D^2 training images.

For the extraction of holistic GIR feature, GIR is considered as a whole for classification. In particular, each of the Gabor filtered images is processed by the dimensionality reduction using DCT. The DCT features derived from the 40 Gabor images are then concatenated to form an augmented feature vector, which is used for face recognition by EFM. The similarity score generated in this approach is subsequently fused with the ones from the four local Gabor feature sets through the sum rule for final classification.

4.2.2 The Multi-resolution LBP Feature Fusion for the D^2 Image

The success of Local Binary Patterns (LBP) in face recognition is attributed to its robustness in terms of gray-level monotonic transformation (Ahonen et al. 2006). However, faces consist of uneven skin surface, which usually leads to nonmonotonic gray-level transformation as illumination changes. In this case, the performance of LBP degrades significantly, while Gabor kernel filters display excellent capabilities of resisting the severe image variations, such as those in the FRGC database. Compared with the D^1 component image, the D^2 component image has the characteristic of keeping away from the nonmonotonic gray-level transformation on face skin surface. To validate this observation, three subregions as shown in Figure 4.5 are chosen for all the D^1 and D^2 component images in the FRGC training set. Then the average standard deviations of intensity values in these subregions are computed, respectively. The results in Figure 4.5 show that the D^2 component image has a lower average of standard deviations than the D^1 component image, hence leading to less nonmonotonic gray-level transformation. Such a property suggests that the D^2 component

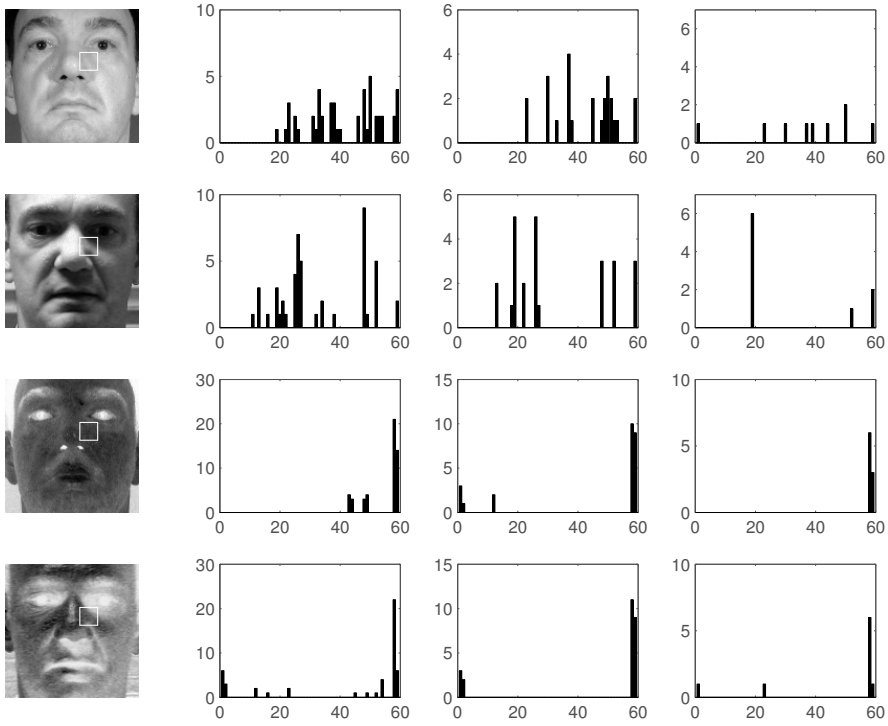

Figure 4.6 The comparison of the LBP histograms in a window on face images. The top two rows are the target and query D^1 component images of one subject. The bottom two rows are their associated D^2 component images.

image is desirable for the extraction of the LBP features. Furthermore, the robustness of the D^2 component image to illumination variations is demonstrated through the comparison of the LBP histograms. Figure 4.6 illustrates the comparison results, which show clearly that there is a perfect matching between the target and query LBP histograms of the same subject for the D^2 component image, even though the large illumination variations occur.

Noting that GIR integrates the features corresponding to five different scales for improving the performance, the multi-resolution information obtained from the different LBP operators can be combined for face recognition. First, the three LBP operators, $LBP^{u2}_{8,1}$, $LBP^{u2}_{8,2}$, and $LBP^{u2}_{8,3}$, are used to extract the multi-resolution histogram features from the D^2 component image, respectively. Second, three global histogram features are fused to form an augmented feature. A straightforward way is to concatenate them. However, this process leads to the problem of high dimensionality. In this chapter, a multi-resolution LBP

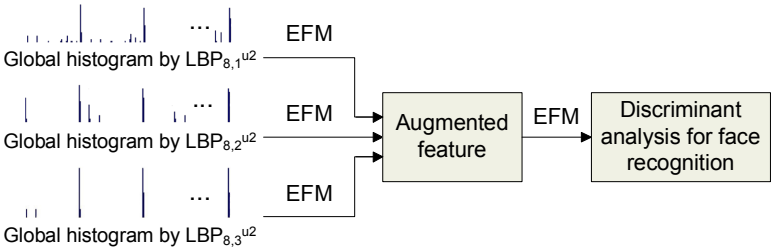

Figure 4.7 Multi-resolution LBP feature fusion scheme.

feature fusion scheme is adopted as shown in Figure 4.7. For each global LBP histogram, an EFM is used to extract features and reduce dimensionality. Let $\mathbf{X}_{h_1}$, $\mathbf{X}_{h_2}$, and $\mathbf{X}_{h_3}$ be the reduced features after the EFM analysis. Specifically, they are normalized and then concatenated into an augmented feature vector, $\mathbf{Y} = (\frac{\mathbf{X}_{h_1}-\mu_1}{\delta_1}; \frac{\mathbf{X}_{h_2}-\mu_2}{\delta_2}; \frac{\mathbf{X}_{h_3}-\mu_3}{\delta_3})$, where μ_i and δ_i are the mean and standard deviation of the feature $\mathbf{X}_{h_i}$. As such, both microstructures and macrostructures of face images are utilized to extract the discriminating features that contain more discriminative power than the one a single LBP operator can provide.

4.2.3 The DCT-based Multiple Face Encodings for the D^3 Image

The preceding sections propose classifying the D^1 and D^2 face images by applying the Gabor and LBP methods, which concentrate primarily on the extraction of local image features. The main purpose of this chapter is to extract the multiple complementary feature sets for face representation, such that the diversity of misclassifications can be enhanced. Based on this, DCT is applied to extract the global features from the D^3 component image for face recognition. DCT transforms images from the spatial domain to the frequency domain, where an image is decomposed into a combination of various and uncorrelated frequency components. DCT is thus able to extract the features in the frequency domain to encode different facial details that are not directly accessible in the spatial domain. Due to the specific properties, DCT has been successfully applied to face recognition recently (Hafed et al. 2001; Chen et al. 2006).

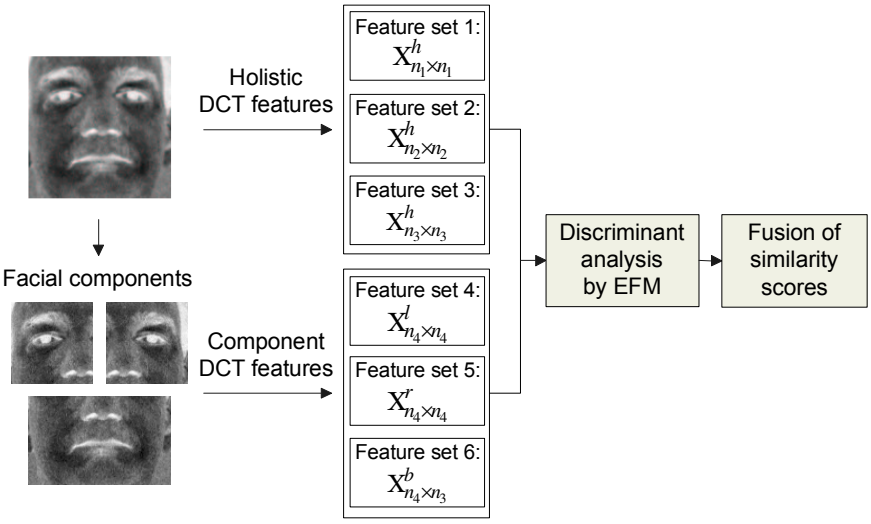

Figure 4.8 DCT-based multiple face encoding scheme for the D^3 image.

In the following, a method, called the DCT-based multiple face encodings, is presented to extract multiple DCT feature sets to classify the D^3 face image. The architecture of method is outlined in Figure 4.8. Similar to the idea of extracting the Gabor features, the DCT-based method is applied to extract the feature sets in terms of holistic and component views. For each category, three DCT feature sets, which describe the different facial information, represent a multiple face encoding. In the DCT domain, one DCT feature set, $\mathbf{X}_{l\times k}$, is defined as: $\mathbf{X}_{l\times k} = \cup F(u,v)$, $\forall u,v$, $u = 0,1,\ldots l-1$, $v = 0,1,\ldots k-1$, where $F(u,v)$ is the DCT coefficient of an image $f(x,y)$ at the location (u,v) in the DCT domain.

The different DCT coefficients correspond to the different spatial information. The upper-left subset in the DCT domain encodes most of the energy in a face image, i.e., the low frequency information associated with the profile information in space. When this subset extends toward right and down (same as the variable mask in Figure 4.4), the more high frequency information is included, the more facial details are displayed in space. Based on this observation, the three holistic DCT feature sets, $\mathbf{X}^h_{n_1\times n_1}$, $\mathbf{X}^h_{n_2\times n_2}$, and $\mathbf{X}^h_{n_3\times n_3}$, with the sizes, $n_1 \times n_1$, $n_2 \times n_2$, and $n_3 \times n_3$, respectively, are used to encode the different levels of facial details. Since the three DCT feature sets contain the different discriminating

information for face images, their classification outputs could be complementary to each other. Thus, after the EFM discriminant analysis, the generated similarity scores can be fused to improve the performance.

Component-based face recognition methods (Kim et al. 2005) have been shown effective for combating image variations caused by illumination and pose in component or patch images, because the variations in these small regions are less than those in the whole face image. The proposed method only considers a simple separation of the three facial components as shown in Figure 4.8, i.e., the upper-left component, the upper-right component, and the bottom-half component. Note that there are overlapping regions that preserve the adjacency relationships among the neighboring components (Kim et al. 2005). The three facial components yield the three component DCT feature sets $\mathbf{X}^{l}_{n_4 \times n_4}$, $\mathbf{X}^{r}_{n_4 \times n_4}$, and $\mathbf{X}^{b}_{n_4 \times n_3}$, with the sizes $n_4 \times n_4$, $n_4 \times n_4$, and $n_4 \times n_3$, respectively. Then these DCT feature sets can generate three similarity scores by the EFM discriminant analysis. Finally, the six similarity scores, resulting from both holistic and component categories, are fused by the sum rule at the decision level to derive a new similarity score for classification.

4.3 Experiments

This section assesses the proposed method using the FRGC version 2 database Experiment 4, the most challenging FRGC experiment (Phillips et al. 2005). Specifically, the training set contains 12,776 images that are either controlled or uncontrolled. The target set has 16,028 controlled images and the query set has 8,014 uncontrolled images. The image sizes in experiments are 128×128 for the D^1 image to extract the Gabor features, and 64×64 for the D^2 and D^3 images to extract the LBP and DCT features, respectively. To evaluate the generalization of the proposed method, experiments also are conducted on the FERET database (Phillips et al. 2000) at the end of this section.

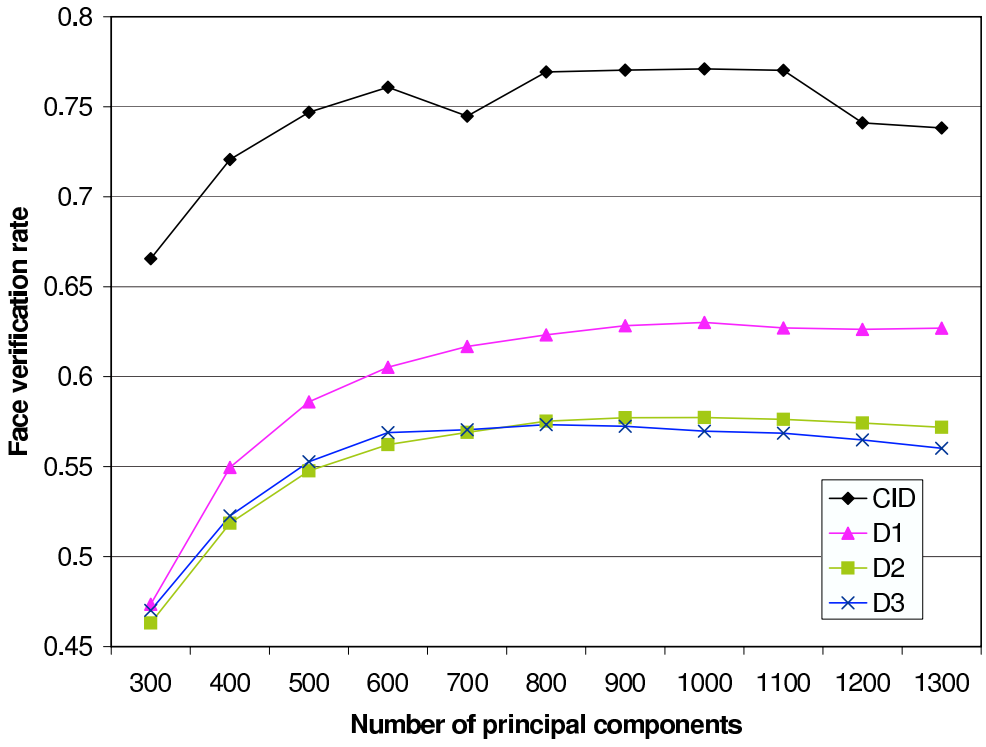

Figure 4.9 The performance of the CID color space vs. the number of principal components.

4.3.1 Effectiveness of the CID Color Space for Face Recognition

Experiments are first conducted on the individual component images, D^1, D^2, D^3, R, G, and B, through applying the EFM method and the cosine similarity measure on the FRGC database. In particular, the face region of all the images is the same as one in Figure 4.1 and the size is 64×64. For comparison purpose, the color component images r, g, b, Y, U, and V (in YUV) are also included in experiments. Figure 4.9 shows that the performances of the D^1, D^2, and D^3 images vary, while changing the number of principal components in EFM. Specifically, the 1,000 principal components achieve the best recognition performance for most color component images. To evaluate the effectiveness of color spaces for face recognition, the classification outputs of different color component images are fused at the decision level by the sum rule. The Face Verification Rates (FVR) derived from the

Table 4.1 FVR (ROC III) at 0.1% FAR Using Different Color Images and Spaces

FVR (ROC III) at 0.1% FAR		
D^1 63.01%	R 62.12%	Y 56.41%
D^2 57.73%	G 53.78%	U 52.36%
D^3 56.97%	B 39.93%	V 57.20%
CID 77.10%	RGB 59.14%	YUV 72.25%
D^{1*} 64.18%	R^* 64.32%	r 34.00%
D^{2*} 45.10%	G^* 61.13%	g 44.36%
D^{3*} 42.41%	B^* 48.39%	b 30.28%
CID^* 62.74%	RGB^* 66.43%	rgb 49.42%

ROC III curves at the False Accept Rate (FAR) of 0.1% are listed in Table 4.1.

The results in Table 4.1 show that the new color space, CID, achieves better performance than the others. This is due to that the new color component images D^1, D^2, and D^3 are derived from an optimization procedure upon the CID model, which not only represents the effective color combination coefficients but also contains the discriminant projection basis vectors for image classification (Yang et al. 2008). Moreover, the new color components are $\mathbf{L}_w(\mathbf{P})$-orthogonal to each other. The fusion at the decision level can thus boost the performance significantly. It should be noted that the performance of the commonly used Y image is lower than those of the new color component images. This implies that Y is not ideally suited for face recognition.

Finally, an illumination normalization procedure described in Chapter 3 is adopted to alleviate the effect of illumination variations on the component images R, G, and B. This procedure comprises three steps: (1) the adjustment of DCT coefficients in logarithm domain (Chen et al. 2006), (2) the Difference of Gaussian (DoG) filtering to preserve the facial details, and (3) the contrast equalization to transform the intensity values to a specified range. Then the illumination normalized R, G, and B component images are used to generate the CID color space. The corresponding experimental results denoted with asterisks in Table 4.1 show that the illumination normalization can not help the CID model

Table 4.2 FVR (ROC III) at 0.1% FAR Using Different Gabor Patch Images

FVR (ROC III) at 0.1% FAR	
Scale 1: 35.06%	Orientation 1: 16.36%
Scale 2: 51.73%	Orientation 2: 30.34%
Scale 3: 59.31%	Orientation 3: 37.92%
Scale 4: 56.35%	Orientation 4: 46.69%
Scale 5: 37.22%	Orientation 5: 32.51%
-	Orientation 6: 44.78%
-	Orientation 7: 37.74%
-	Orientation 8: 26.69%
Decision fusion: 73.44%	Decision fusion: 72.23%

improve face recognition performance.

4.3.2 Experiments Using the Patch-based GIR for the D^1 Image

To generate GIR, the parameters of the Gabor filters are chosen as follows: the spacing factor between filters in the frequency domain $\sqrt{2}$, the standard deviation 2π, and the maximum frequency $\pi/2$. The experiments first aim to investigate the roles of scale and orientation of the Gabor filters in face recognition. With that goal, GIR is divided into five scale patches (containing eight orientations) along the horizontal direction and eight orientation patches (containing five scales) along the vertical direction, respectively. Table 4.2 provides the FVR on the FRGC database using the DCT features obtained from these Gabor patches. It is clear that the scale patches possess more discriminative power than the orientation patches. Thus, GIR should be processed along the horizontal direction instead of the vertical direction to extract the local GIR features.

The patch-based GIR method then extracts the DCT features from each of sub-GIR patches using a DCT mask with size of 64×64. Prior to this operation, each sub-GIR patch is reshaped into a square array. To assess the feasibility of this idea, experiments are conducted on the sub-GIR patches before and after the reshaping operation, respectively. Table 4.3 shows the comparative performances, which indicate that the reshaping opera-

Table 4.3 FVR (ROC III) at 0.1% FAR Using the Local and Global GIR Features

Method	FVR (ROC III) at 0.1% FAR		
	Original image, original sub-GIR	Original image, reshaped sub-GIR	Normalized image, reshaped sub-GIR
Scales 1&2	48.91%	53.87%	67.57%
Scales 2&3	61.67%	65.10%	72.84%
Scales 3&4	67.16%	68.37%	75.10%
Scales 4&5	62.82%	64.16%	70.61%
Fusion I	75.60%	76.94%	83.29%
Whole GIR	-	74.12%	81.28%
Fusion II	-	78.16%	84.50%
Fusion III	-	76.15%	82.39%
Scales 1&3	-	61.96%	72.15%
Scales 2&4	-	67.47%	75.13%
Scales 3&5	-	65.93%	73.17%
Scales 1&5	-	54.41%	67.25%
Fusion I	-	75.09%	82.36%

tion indeed helps extract more discriminating DCT features in the GIR method. As far as the whole GIR is concerned, its dimensionality is reduced by using a DCT domain mask defined in Figure 4.4 as well. For computation efficiency, each of Gabor convolved images is reduced to a vector of size 256 (16×16) in the DCT domain. The resulting augmented vector of size 10,240 is then processed by the EFM method.

The Fusion I in Table 4.3 represents the decision fusion of classification outputs of the four sub-GIR patches, and the Fusion II is the decision fusion of outputs of the Fusion I and the whole GIR, both using the sum rule. For comparison purpose, the fusion of local and global GIR features is considered at the feature level. When concatenating these feature vectors directly, the length of resulting feature vector is 26,624, which is too high to be processed during the PCA computation by computers. Thus, EFM is used to reduce the dimensionality of individual local and global GIR feature sets and then an augmented feature vector is formed for classification. This idea is similar to that detailed in Figure 4.7 The Fusion III in Table 4.3 generates the classification results using this fusion strategy, indicating that the performance of data fusion at the feature level is not as good as that at

Table 4.4 FVR (ROC III) at 0.1% FAR Using LBP for the D^1 and D^2 Images

Method	FVR (ROC III) at 0.1% FAR			
	D^1		D^2	
	χ^2	EFM	χ^2	EFM
$LBP^{u2}_{8,1}$	7.34%	65.18%	18.69%	59.50%
$LBP^{u2}_{8,2}$	11.15%	65.99%	21.81%	61.75%
$LBP^{u2}_{8,3}$	12.01%	56.42%	21.69%	54.38%
Fusion	11.79%	70.82%	25.14%	70.02%

the decision level in the approach.

To extract the local GIR features, the two adjacent GIR scales are grouped into a sub-GIR patch. This strategy may not be optimum, as there are so many other combinations of GIR scales into sub-GIR patches. Nevertheless, it is reasonable under the assumption that the redundancy in GIR is caused mainly by the similarities between the adjacent Gabor filters. Due to the lack of space, only one of other combinations, i.e., scales 1&3, scales 2&4, scales 3&5, and scales 1&5, is chosen to classify faces for comparison purpose. The corresponding FVRs in Table 4.3 are lower than those of the combination using the adjacent scales, implying that more redundancy among sub-GIR patches can be reduced when applying DCT to the sub-GIR patches consisting of adjacent scales, at least in this comparison instance. The experimental results derived by the illumination normalized D^1 image are listed in Table 4.3 as well.

4.3.3 Experiments Using the LBP Features for the D^2 Image

To extract the LBP features, a face image of size 64×64 is divided into 144 (12×12) overlapping windows of 9×9 pixels (3 pixels overlapping). The length of the LBP feature vectors is thus 8,496. This section is first concerned with the face description capability of the LBP operators on the D^1 and D^2 images using the non-statistical method, e.g., histogram matching. Table 4.4 records the experimental results on the FRGC database using the Chi square distance measure upon the target and query LBP histograms. As can be

seen, the D^2 image is more suitable than the D^1 image for the LBP descriptors with respect to histogram matching for face analysis. However, due to the large illumination variations, the performance achieved by histogram matching is not good enough for contribution to the overall recognition performance. On the other hand, statistical pattern recognition is able to learn the characteristics of input data distribution for each category and find a proper discriminant function from the training samples. The supervised classification method (e.g., LDA) thus can be applied to extract the discriminating features from the LBP histogram vectors for face recognition.

For each of the three global LBP histograms, an EFM analysis is first used to derive the discriminating features with lower dimensionality. After feature concatenation, EFM is applied again to analyze the augmented feature vector to choose the most discriminating features for classification. Table 4.4 lists the corresponding FVRs, which indicate clearly that the fusion of multi-resolution LBP features via EFM helps improve the performance significantly. Specifically, the D^1 and D^2 images can achieve the comparable performances using the proposed method.

4.3.4 Experiments Using the DCT Features for the D^3 Image

For the extraction of holistic DCT features, the three feature sets are used to encode the different levels of facial details. The largest set, which includes comprehensive frequencies, has the same size 64×64 ($n_3 = 64$) as the original images to encode the entire facial details. The smallest set, which includes low frequencies to encode the facial profile information, is chosen according to the following rule. Because the subject number of the FRGC training set is 222, the rank of between-class scatter matrix is at most 221. To derive the 221 EFM features, the input feature vector should reside in a space whose dimensionality is larger than 221. Thus, 15×15 ($n_1 = 15$) is chosen as the size of the smallest set. The remaining issue is to determine the size of middle set, namely, low frequencies plus intermediate frequencies. It is hard in practice to explicitly choose the exact boundary between interme-

diate and high frequencies in the DCT domain. To solve this problem, *root-mean-square error* (e_{rms}) is used.

To calculate the e_{rms}, the reconstructed images using the DCT feature sets are needed. For example, the reconstructed image, $\hat{f}_{15\times15}(x,y)$, of the DCT feature set $\mathbf{X}^h_{15\times15}$ is obtained by conducting the inverse DCT transform upon the DCT domain, where the upper-left subset of size 15×15 is kept unchanged while the values in the remaining area are set to zeros. Similarly, for the DCT feature set $\mathbf{X}^h_{64\times64}$, another reconstructed image is generated, which is actually equivalent to the original image $f(x,y)$. Next, let $\hat{f}_{n_2\times n_2}(x,y)$ be the reconstructed image of the DCT feature set $\mathbf{X}^h_{n_2\times n_2}$. Now, the three e_{rms} values, $e_{rms}(15)$, $e_{rms}(n_2)$, and $e_{rms}(64)$ are computed as follows:

$$e_{rms}(15) = \left[\frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\left[\hat{f}_{15\times15}(x,y) - f(x,y)\right]^2\right]^{1/2}, \tag{4.10}$$

$$e_{rms}(n_2) = \left[\frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\left[\hat{f}_{n_2\times n_2}(x,y) - f(x,y)\right]^2\right]^{1/2}, \tag{4.11}$$

where $M \times N$ ($M = N = 64$) indicates image size, and $e_{rms}(64)$ = 0 according to definition. The size of the DCT feature set $\mathbf{X}^h_{n_2\times n_2}$ can then be chosen by solving the following optimization problem:

$$n_2^* = \underset{n_2}{\operatorname{argmin}} \left| e_{rms}(n_2) - (e_{rms}(15) + e_{rms}(64))/2 \right|. \tag{4.12}$$

Based on Equation (4.12), n_2 = 28 can be estimated from the FRGC training set. 28×28 is thus chosen as the size of middle feature set.

As for the extraction of component DCT feature sets, the sizes of feature sets are just set to the same as ones of the facial component images. In particular, the upper-left and upper-right component images have the same size 39×39, and the bottom-half component

Table 4.5 FVR (ROC III) at 0.1% FAR Using the DCT Features for the D^3 Image

Method	FVR (ROC III) at 0.1% FAR
Holistic 15 × 15 DCT features	41.83%
Holistic 28 × 28 DCT features	57.48%
Holistic 64 × 64 DCT features	58.57%
Upper-left component DCT features	35.78%
Upper-right component DCT features	31.43%
Bottom-half component DCT features	24.89%
Decision fusion	61.31%

image has size 39×64. Thus, $n_4 = 39$ is used for the sets $\mathbf{X}^l_{n_4 \times n_4}$, $\mathbf{X}^r_{n_4 \times n_4}$ and $n_4 = 39$, $n_3 = 64$ for the set $\mathbf{X}^b_{n_4 \times n_3}$. Table 4.5 lists the FVRs derived from the different DCT feature sets and their fusion result at the decision level on the FRGC database.

4.3.5 Effectiveness of the Proposed Method

After having generated the three similarity matrices, which correspond to the FVR of 84.5% for the illumination normalized D^1 image (see Table 4.3), the FVR of 70.02% for the D^2 image, and the FVR of 61.31% for the D^3 image, respectively, they can be fused by means of the sum rule. Specifically, the method achieves the FVR (ROC III) of 90.4% at the FAR of 0.1% on the FRGC version 2 Experiment 4. Currently, the significant portion of computational time is spent in the extraction of the GIR features, taking about 1.9 seconds to process one image using Matlab on 2.0 GHz P4 PC. The faster speed can be expected through C++ programming.

The main concern in this chapter is to exploit the multiple complementary feature sets on face images, so that the fusion of classification outputs at the decision level can improve the face recognition performance as much as possible. To that end, this chapter proposes extracting the GIR, LBP, and DCT feature sets appropriate for the component images in the CID color space. To validate this idea, all the feature extraction methods have been applied to the three component images for face recognition. The experimental results

Table 4.6 FVR (ROC III) at 0.1% FAR Using Different Features in the CID Color Space

Method	FVR (ROC III) at 0.1% FAR		
	D^1	D^2	D^3
GIR	78.16%	22.18%	27.02%
LBP	70.82%	70.02%	64.45%
DCT	65.68%	60.19%	61.31%

Table 4.7 Some Experimental Results Using the Different Combinations of Features

GIR	LBP			DCT			FVR
D^1	D^1	D^2	D^3	D^1	D^2	D^3	(ROC III)
√		√				√	90.4%
√		√	√				88.9%
√	√	√			√		89.5%
√	√		√		√		90.0%
√	√	√	√	√	√	√	91.6%

Table 4.8 Comparison of the Proposed Method with the Others on the FRGC Database

Method	FVR (ROC III) at 0.1% FAR
Method in (Su et al. 2007)	85.8%
Method in (Kumar et al. 2006)	87.5%
The proposed method	91.6%

on the FRGC database are given in Table 4.6. As can be seen, all the methods are effective for improving the recognition accuracy on the three component images, except GIR that fails in working on the D^2 and D^3 images. Furthermore, the different combinations of feature sets can lead to the various classification results, some of which are provided in Table 4.7. In particular, when fusing all the classification outputs of different methods on the three component images, the proposed method can achieve the FVR of 91.64% at 0.1% FAR, better than the performances achieved by some state-of-the-art methods as shown in Table 4.8. Note that the GIR features used in the experiments as indicated in Table 4.7 are derived from the illumination normalized D^1 image.

Table 4.9 The Rank-1 Recognition Rate on the FERET Database

Method	Rank-1 recognition rate	
	Dup I	Dup II
Method in (Zou et al. 2007)	85.0%	79.5%
Method in (Yao et al. 2008)	86.0%	83.0%
The proposed method (1)	88.4%	87.3%
The proposed method (2)	88.8%	86.4%
The proposed method (3)	87.1%	86.8%

To assess the generalization of the proposed method, experiments are implemented on the color FERET database. Three sets of experiments are designed to investigate the recognition performance based on the different training sets. They are (1) FRGC for both the CID model and feature extraction, (2) FRGC for the CID model and FERET for feature extraction, and (3) FERET for both the CID model and feature extraction. Specifically, 1,000 images comprising of 500 persons (two images fa/fb per person) are randomly selected to construct the FERET training set. This set can derive the other set of the CID transformation vectors: $\mathbf{X}_1 = [0.7097, 0.4294, -1.0]^T$, $\mathbf{X}_2 = [0.4674, -0.9702, 0.5028]^T$, and $\mathbf{X}_3 = [-0.1105, 1.4258, -1.3154]^T$. The two most challenging probe sets, Dup I and Dup II, which are used for analyzing the effect of aging on the recognition performance, are evaluated in experiments. For the color FERET database, the image numbers of the gallery, Dup I, and Dup II probe sets are 967, 722, and 228, respectively. Table 4.9 gives the rank-1 recognition rates, showing that the proposed method can also achieve a good performance on the color FERET database using the different training sets. The performances achieved by some other methods using gray images are given in Table 4.9 as well. It should be noted that one can not establish a direct comparison between the proposed method and the others, because of the differences between the color and gray FERET databases.

4.4 Conclusion

This chapter proposes a novel face recognition method that extracts multiple features in the CID color space, where three new color component images, D^1, D^2, and D^3, are derived using an iterative algorithm. Three different image encoding methods are also presented to effectively extract features from the component images in the new CID color space for enhancing pattern recognition performance. The similarity scores from the three color component images are fused for the final decision making. Experiments using two large-scale face databases, namely, the Face Recognition Grand Challenge (FRGC) version 2 database and the FERET database, show the effectiveness of the proposed method.

CHAPTER 5

LEARNING IMAGE REPRESENTATION FOR FACE RECOGNITION

Feature selection and feature extraction are crucial to many pattern classification problems, e.g., face recognition. The common objective of feature selection and extraction is to map the original measurements into more effective features, which show significant differences from one class to another, so that the classifiers can be designed more easily with better performance. Most existing methods in face recognition scenario mainly focus on extracting features either from image space via techniques such as Gabor kernels (Daugman 1985) and Local Binary Patterns (LBP) (Ojala et al. 2002), or from transformed spaces via statistical methods such as Principal Component Analysis (PCA) (Turk et al. 1991) and Linear Discriminant Analysis (LDA) (Belhumeur et al. 1997). Image representation, as the beginning stage of face recognition, actually can be considered as another important feature. Unfortunately, the role of image representation is often ignored in recent studies. Numerous face recognition methods usually start with image representation that is the directly linear combination of three primary colors, R, G, and B, i.e., the luminance image Y = 0.299R + 0.58G + 0.114B or the intensity image I = (R + G + B)/3. But more theoretical evidences and experimental results are in great demand to support that such image representations or color transformations are optimal for image classification, especially for face recognition. Hence, the novel image representation adapted to class separability is highly desired in order to simplify the classifier design and improve the classification performance for face recognition.

The contribution to face recognition in this chapter is to generate the novel image representations and present the corresponding face recognition method. Generally speaking, the hybrid color configurations have less correlation between their components than the RGB color space. Such a property determines that the hybrid color configurations usually achieve better performance than the RGB color space in face recognition, as decorrelation

is important for pattern recognition. Thus some hybrid configurations of color components are selected as source images to generate image representations. Furthermore, the proposed method obtains the optimal color transformation via a statistical learning algorithm, PCA plus FLD. PCA seeks a principal subspace of lower dimensionality to maximize the data reconstruction capability of the features. As a result, features in this subspace can represent the original data accurately. One can thus apply PCA to derive an optimal color transformation to generate a new image representation that best represents the data information from several original color components. But the best representation of data may not perform well from the classification point of view because the total scatter matrix is made up of both the within- and between-class variations. To obtain the discrimination of image representation for face recognition, one needs to handle the within- and between-class variations separately. The proposed method further applies FLD to generate the discrimination-driven image representations to fit face recognition, by following an idea in the Color Image Discriminant (CID) model (Yang et al. 2008). Furthermore, to overcome the effect of illumination variations on face image, a face image is partitioned into several small patches. Depending on the discriminative power embedded in different facial regions, the proposed method can then apply the learning algorithm to obtain the different color transformations via either PCA alone or PCA plus FLD for these patch images. Therefore, a novel image representation based on a patch scheme can be derived for face recognition. Finally, a multiple feature method, extracting both holistic and local features, is presented for face recognition using the proposed image representation. The large-scale experiments on the Face Recognition Grand Challenge (FRGC) version 2 Experiment 4 and the FERET Dup I and Dup II probe sets have been implemented to show the effectiveness of the proposed method.

Table 5.1 Correlation Coefficients between Different Color Components

	R	G	B	V	Y	I	Q	Cb	Cr
R	1.0	0.93	0.87	0.99	0.97	0.78	0.03	-0.77	0.72
G	-	1.0	0.94	0.93	0.99	0.55	-0.22	-0.67	0.46
B	-	-	1.0	0.87	0.93	0.42	-0.06	-0.46	0.37
V	-	-	-	1.0	0.97	0.79	0.04	-0.77	0.73
Y	-	-	-	-	1.0	0.64	-0.11	-0.71	0.56
I	-	-	-	-	-	1.0	0.28	-0.84	0.97
Q	-	-	-	-	-	-	1.0	0.22	0.46
Cb	-	-	-	-	-	-	-	1.0	-0.72
Cr	-	-	-	-	-	-	-	-	1.0

5.1 Hybrid Configurations of Color Components

RGB color space is commonly used in some face recognition methods (Xie et al. 2005; Jones et al. 2006; Kim et al. 2007; Yang et al. 2008) for color image. However, correlation among components R, G, and B is so strong that the resultant improvement in performance is limited when three components are combined for pattern recognition, such as face recognition. Therefore, other color spaces, especially the hybrid color configurations, have been adopted to perform face recognition in recent studies (Shih et al. 2005; Liu et al. 2008; Sadeghi et al. 2007).

To construct the hybrid configurations of color components, some candidates of color components need to be chosen first. Research in (Shih et al. 2005) has revealed that V in HSV is more effective than the others for face recognition. This is reasonable because the V color component prefers the maximal value among R, G, and B components according to the definition V = *max*(R,G,B). Such a characteristic makes the V color component look like brighter than the luminance Y when the large illumination change occurs on face image. Thus V is chosen on behalf of R, G, and B. Among other commonly used color spaces for face image analysis are YIQ and YCbCr. By calculating the correlation coefficients resided between individual color components V, Y, I, Q, Cb, and Cr, some rational hybrid configurations of color components can be chosen. Table 5.1 records the

corresponding correlation coefficients calculated from the training database of FRGC. By preferring the low coefficient magnitude, some color configurations such as VCrQ, YIQ, and YCrQ are selected as the source color images to generate the new image representations. It is worthwhile of note that the Q color component has a higher chance to be chosen in the method due to its decorrelationship with the others.

5.2 Learning Image Representation

In the $C^1C^2C^3$ color configuration, an image of resolution $m \times n$ consists of three color components $\mathbf{C}^1$, $\mathbf{C}^2$, and $\mathbf{C}^3$. Without loss of generality, let $\mathbf{C}^1$, $\mathbf{C}^2$, and $\mathbf{C}^3$ be column vectors: $\mathbf{C}^1, \mathbf{C}^2, \mathbf{C}^3 \in R^N$, where $N = m \times n$. Each vector $\mathbf{C}$ is normalized to have zero mean and unit variance. A data matrix $\mathbf{X} \in R^{3 \times Nl}$ can be formed using all the training images:

$$\mathbf{X} = \begin{bmatrix} \mathbf{C}_1^1, & \mathbf{C}_1^2, & \mathbf{C}_1^3 \\ \mathbf{C}_2^1, & \mathbf{C}_2^2, & \mathbf{C}_2^3 \\ \vdots, & \vdots, & \vdots \\ \mathbf{C}_l^1 & \mathbf{C}_l^2, & \mathbf{C}_l^3 \end{bmatrix}^t \tag{5.1}$$

where l is the number of training images. In $\mathbf{X}$, each column is an observation and each row is a variable. The covariance matrix Σ_X may be formulated as $\Sigma_X = \frac{1}{Nl-1}\tilde{\mathbf{X}}\tilde{\mathbf{X}}^t \in R^{3\times 3}$, where $\tilde{\mathbf{X}}$ is the centered data matrix. PCA is used to factorize Σ_X into the following form: $\Sigma_X = \Phi\Lambda\Phi^t$, where $\Phi = [\Phi_1, \Phi_2, \Phi_3] \in R^{3\times 3}$ is an orthonormal eigenvector matrix and $\Lambda = diag\{\lambda_1, \lambda_2, \lambda_3\} \in \mathbb{R}^{3\times 3}$ is a diagonal eigenvalue matrix with diagonal elements in decreasing order ($\lambda_1 \geq \lambda_2 \geq \lambda_3$).

By projecting three color component images $\mathbf{C}^1, \mathbf{C}^2, \mathbf{C}^3$ of an image onto Φ_1, then a new image representation $\mathbf{U} \in R^N$ can be obtained, which is associated with the largest eigenvalue. So one has:

$$\mathbf{U} = [\mathbf{C}^1, \mathbf{C}^2, \mathbf{C}^3]\Phi_1 \tag{5.2}$$

According to properties of PCA, $\mathbf{U}$ is optimal for data representation but not for data classification. Recently, the CID model (Yang et al. 2008) has been proposed to derive a more reasonable color transformation for face recognition. Thus, the novel image representation can be further generated based on $\mathbf{U}$ via FLD implemented in CID. Let $\bar{\mathbf{U}}_i$ be the mean vector of class ω_i and $\bar{\mathbf{U}}$ be the grand mean vector. Then the between- and within-class scatter matrices $\mathbf{S}_b$ and $\mathbf{S}_w$ are defined as follows:

$$\mathbf{S}_b = \sum_{i=1}^{k} P(\omega_i)(\bar{\mathbf{U}}_i - \bar{\mathbf{U}})(\bar{\mathbf{U}}_i - \bar{\mathbf{U}})^t \tag{5.3}$$

$$\mathbf{S}_w = \sum_{i=1}^{k} P(\omega_i)\mathscr{E}\left\{(\mathbf{U} - \bar{\mathbf{U}}_i)(\mathbf{U} - \bar{\mathbf{U}}_i)^t | \omega_i\right\} \tag{5.4}$$

where $P(\omega_i)$ is the prior probability of class ω_i, and k is the number of classes, and $\mathbf{S}_b, \mathbf{S}_w \in R^{N\times N}$. The general Fisher criterion in the U image space can be defined as follows:

$$J(\mathbf{P}) = \frac{|\mathbf{P}^t \mathbf{S}_b \mathbf{P}|}{|\mathbf{P}^t \mathbf{S}_w \mathbf{P}|} \tag{5.5}$$

Maximizing this criterion can be solved by deriving the optimal transform matrix $\mathbf{P} = [\psi_1, \psi_2, \ldots, \psi_d] \in R^{N\times d}$, where $\psi_1, \psi_2, \ldots, \psi_d$ are chosen from the generalized eigenvectors of $\mathbf{S}_b\Psi = \lambda\mathbf{S}_w\Psi$ corresponding to the d largest eigenvalues.

Let $\mathbf{C} = [\mathbf{C}^1, \mathbf{C}^2, \mathbf{C}^3]$ be the original color configuration with the $N(0,1)$ normalization. By projecting $\mathbf{C}$ onto the matrix $\mathbf{P}$, one can define the general color-space between-class scatter matrix $\mathbf{L}_b$ and within-class scatter matrix $\mathbf{L}_w$ as follows (Yang et al. 2008):

$$\mathbf{L}_b = \sum_{i=1}^{k} P(\omega_i)(\bar{\mathbf{C}}_i - \bar{\mathbf{C}})^t \mathbf{P}\mathbf{P}^t (\bar{\mathbf{C}}_i - \bar{\mathbf{C}}) \tag{5.6}$$

$$\mathbf{L}_w = \sum_{i=1}^{k} P(\omega_i)\mathscr{E}\left\{(\mathbf{C} - \bar{\mathbf{C}}_i)^t \mathbf{P}\mathbf{P}^t (\mathbf{C} - \bar{\mathbf{C}}_i) | \omega_i\right\} \tag{5.7}$$

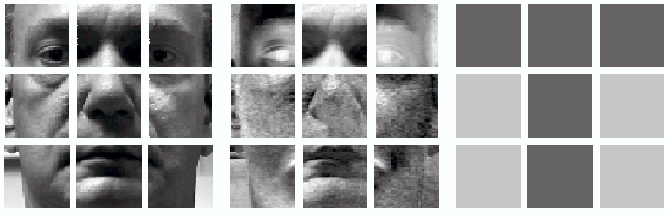

Figure 5.1 The patch-based novel image representation. The left image: the partition of an original image. The middle image: a novel representation derived from the color configuration YCrQ based on a patch scheme. The right image: the U-based representation (gray region) and the D-based representation (dark region).

where $\bar{\mathbf{C}}_i$ is the mean of class ω_i and $\bar{\mathbf{C}}$ is the grand mean, and $\mathbf{L}_b, \mathbf{L}_w \in R^{3\times 3}$. By obtaining the generalized eigenvectors ξ_1, ξ_2, ξ_3 of $\mathbf{L}_b \Xi = \lambda \mathbf{L}_w \Xi$, the ξ_1 corresponding to the largest eigenvalue is selected as the optimal color transformation, which can generate the discriminating image representation suitable for face recognition. Finally, a novel image representation $\mathbf{D} \in R^N$ can be derived by projecting three color component images $\mathbf{C}^1, \mathbf{C}^2, \mathbf{C}^3$ of an image onto ξ_1:

$$\mathbf{D} = [\mathbf{C}^1, \mathbf{C}^2, \mathbf{C}^3]\xi_1 \tag{5.8}$$

5.3 Patch-based Novel Image Representation for Face Recognition

Since the illumination variations appear on face image unevenly, the patch-based methods have become appealing to address this problem and have achieved the improved performance in face recognition (Kim et al. 2005; Heisele et al. 2007). To alleviate the effect of illumination variations and other factors, the face image is partitioned into several small patches, which will learn their own color transform vectors, respectively, via either PCA or PCA plus FLD.

Figure 5.1 shows the partition of a 64×64 face image into nine patches. Because the experiment Section will show that the best image representation for face recognition in FRGC can be generated by the YCrQ color configuration, YCrQ is chosen to derive the optimal color transformation for each patch. In practice, some patches on face cheek

Table 5.2 The Values of Fisher Criterion Derived by PCA and PCA plus FLD Learning Algorithm for Nine Patches

	Image representation	
	U-based	D-based
Patch 1	0.7180	0.7347
Patch 2	0.6218	0.6540
Patch 3	0.7086	0.7136
Patch 4	0.3938	0.3819
Patch 5	0.4410	0.4450
Patch 6	0.3493	0.3375
Patch 7	0.3553	0.3273
Patch 8	0.3174	0.3837
Patch 9	0.3577	0.3192

area contain too insufficient discriminating information to distinguish face classes well. In these patches, the D-based representation defined by Equation (5.8) actually achieves lower performance than the U-based representation defined by Equation (5.2) for face recognition. To guarantee the obtaining of a better representation for each face patch, the Fisher criterion in Equation (5.5) is rewritten as follows:

$$\alpha = \arg\max_{\alpha,\mathbf{P}} \frac{|\mathbf{P}^t \mathbf{S}_b(\alpha)\mathbf{P}|}{|\mathbf{P}^t \mathbf{S}_w(\alpha)\mathbf{P}|} \tag{5.9}$$

where $\alpha \in \{\Phi_1, \xi_1\}$ represents one of color transform vectors used in Equations (5.2) and (5.8) to generate the U-based or D-based image representation. During the training procedure, the criterion in Equation (5.9) is used to choose the better one from Φ_1 and ξ_1 to generate the optimal image representation for each face patch.

Table 5.2 records the values of Fisher criterion generated by the training data set for nine different patches labeled from top to bottom and from left to right in face image. The U-based image representation, derived by PCA, and the D-based image representation, derived by PCA plus FLD, generate the different values of Fisher criterion for each patch, as shown in Table 5.2. The higher the criterion value, the more discriminating information one patch has. From the classification point of view, some face patches (such as facial

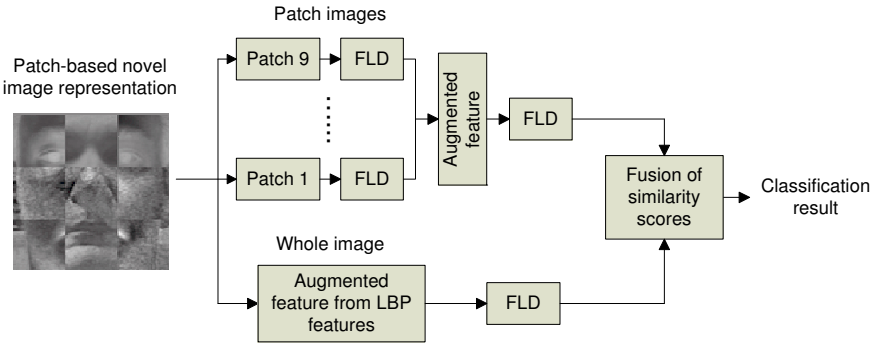

Figure 5.2 Architecture of face recognition using the patch-based novel image representation.

component regions) containing the sufficient facial information should be characterized by the D-based image representation, while the others (such as facial cheek regions) containing the poor facial information should be characterized by the U-based image representation. Therefore, by choosing one of two kinds of image representations in favor of a larger Fisher criterion value for each patch, one can derive a more reasonable face image representation, that is, a mixture of the U-based and D-based image representations based on a patch scheme. Following this idea, a novel representation for face image is illustrated by the right image in Figure 5.1, where the U-based and D-based image representations are symbolized by gray region and dark region, respectively. The middle image in Figure 5.1 is the actual representation for the original image using the proposed method.

Furthermore, a method is presented for face recognition using multiple features on the basis of the patch-based novel image representation. The multiple features, which usually consist of holistic and local features, have been demonstrated effective in improving the performance of face recognition, due to the complementary characteristic among features. The architecture of the multiple features for face recognition in the proposed method is illustrated in Figure 5.2. The holistic and local features are extracted respectively from face image, and their classification outputs are fused at the decision level. On the one hand, the multiple-scale LBP (Liu et al. 2009) is applied to extract a holistic augmented feature. Note that although the LBP operators work on sub-images, the LBP features can

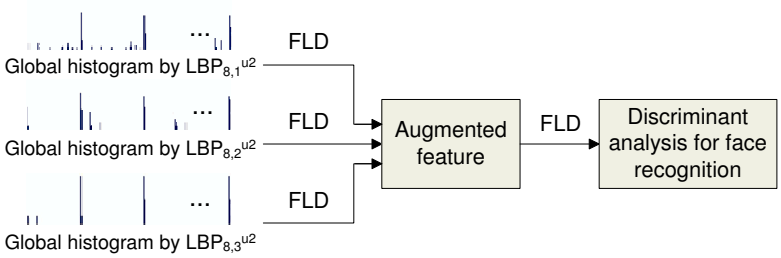

Figure 5.3 Multi-resolution LBP feature fusion scheme.

be regarded as holistic one in this chapter because the global histograms are derived from the whole image. Three LBP operators, $LBP^{u2}_{8,1}$, $LBP^{u2}_{8,2}$ and $LBP^{u2}_{8,3}$ (Ojala et al. 2002), are first used to extract the multi-resolution global histogram features from the whole face image, respectively. Then FLD is applied to extract the discriminating feature and reduce dimensionality for each histogram feature. Finally, three reduced features are concatenated into an augmented feature, which will undergo the discriminant analysis for face recognition. As such, both the microstructure and the macrostructure described by LBP operators in face image are utilized to extract the discriminating features, which contain much more discriminative power than the one a single LBP operator can provide. The foregoing method is illustrated in Figure 5.3. On the other hand, similar to the derivation of the holistic augmented feature, a local augmented feature obtained from the patch images can be extracted by concatenating the FLD features as shown in Figure 5.2. At the decision level, two kinds of classification outputs are first processed by the z-score normalization (Jain et al. 2005) individually and then are fused by the sum rule (Kittler et al. 1998) for the final classification.

5.4 Experiments

This section assesses the performance of the proposed method using the Face Recognition Grand Challenge (FRGC) version 2 Experiment 4, the most challenging FRGC experiment (Phillips et al. 2005), and the another de facto standard face database, the FERET

database. For the FRGC database, the training set contains 12,776 images that are either controlled or uncontrolled. The target set has 16,028 controlled images and the query set has 8,014 uncontrolled images. The FRGC baseline algorithm, which in essence is a PCA algorithm, reveals that the uncontrolled factors pose grand challenges to the face recognition performance. The Biometric Experimentation Environment (BEE) system generates three Receiver Operating Characteristic (ROC) curves (ROC I, ROC II, and ROC III) corresponding to the images collected within semesters, within a year, and between semesters, respectively (Phillips et al. 2005). The FERET database (Phillips et al. 2000) consists of more than 13,000 facial images corresponding to more than 1,500 subjects. Because images were acquired during different photo sessions, the illumination conditions and the size of the face may change. The diversity of the FERET database features across ethnicity, gender, and age. The images were acquired without any restrictions imposed on facial expression and with at least two frontal images shot at different times during the same photo session.

5.4.1 Results of the FRGC Database

Face recognition using different image representations

According to the correlation criterion presented in Section 5.2, some ideal configurations of color components VCrQ, VCbQ, VIQ, YCrQ, YCbQ, and YIQ are selected, while other configurations such as RGB, VCbCr, VCbI, VCrI, YCbCr, YCbI, and YCrI are also included in experiments for comparison purpose. Then the FRGC training images, whose size has been rescaled to 32×32, are used to derive the color transformation vectors Φ_1 and ξ_1 by using either PCA or PCA plus FLD learning algorithm.

To evaluate the effectiveness of the novel image representations for face recognition, a set of experiments on FRGC version 2 Experiment 4 is carried out using FLD and the cosine similarity measure. For the experiments of face recognition, the size of images is 64×64. The experimental results, which includes the Face Verification Rate (FVR)

Table 5.3 FVR (ROC III) at 0.1% FAR Using Different Image Representations

Color configuration	FVR at 0.1% FAR		Color configuration	FVR at 0.1% FAR	
	U-based	D-based		U-based	D-based
YCrQ	66.59%	67.35%	VCrQ	65.76%	66.74%
VIQ	64.80%	65.23%	YIQ	64.95%	64.92%
VCrI	61.31%	63.02%	YCrI	61.65%	62.81%
VCbQ	60.22%	62.09%	YCbQ	61.17%	61.84%
Method	YCrQ	RGB	Luminance Y	56.41%	
in (Yang et al. 2008)	65.29%	63.01%	Intensity I	55.14%	
			V in HSV	61.78%	

at 0.1% False Accept Rate (FAR) corresponding to ROC III curve, are provided in Table 5.3. Here, Table 5.3 only lists the color configurations that can generate the novel image representations with FVR over 61.78%, produced by V in the HSV color space. For the comparison purpose, experimental results derived from the luminance Y and the intensity I are also included in Table 5.3. It is evident that the novel image representations produced by the proposed method outperform other image representations for face recognition. Specifically, the performance derived from the YCrQ color configuration is beyond those of conventional gray-scale images (luminance Y and intensity I) by more than 10%.

The proposed learning method is based partially on the CID model (Yang et al. 2008). However, it has its own advantages over CID when the input is hybrid color configuration instead of RGB. Table 5.3 also provides the experiment results using CID upon the YCrQ and RGB color configurations for comparison. Note that experiments use CID to generate one component image from three source color components for face recognition. As can bee seen, the proposed learning method achieves a better performance than CID when working on the YCrQ color configuration. On the other hand, CID works in an iterative manner, which spends more computation time during training than the proposed method. The above merits of the proposed learning method are attributed to the PCA procedure. As PCA derives an optimal image with respect to data representation from the source images, the U-based image representation is able to generate a good recognition

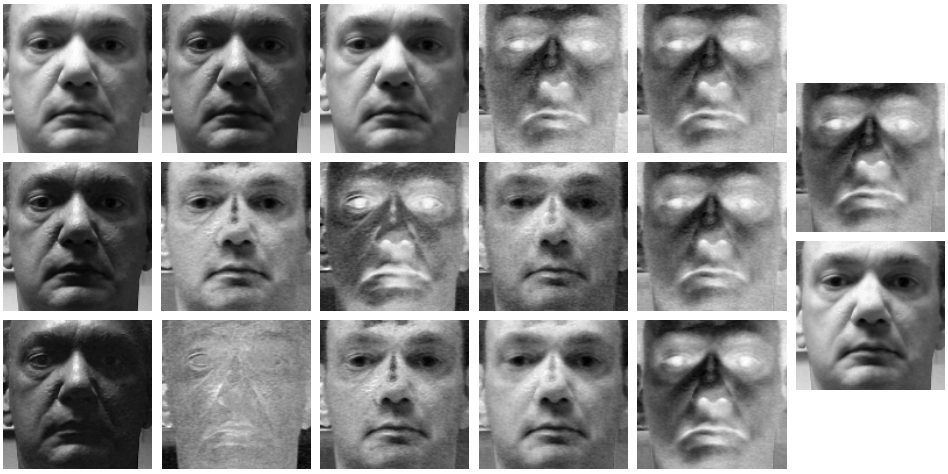

Figure 5.4 The image representations from the first column to the sixth column: R, G, and B; Y, I, and Q; V, Cb, and Cr; U_{YCrQ}, U_{VCrQ}, and U_{VIQ}; D_{YCrQ}, D_{VCrQ}, and D_{VIQ}; the images derived from YCrQ and RGB using CID model (Yang et al. 2008).

performance close to optimal one. This procedure facilitates the subsequent FLD to derive an optimal color transform vector and the iteration in CID can be skipped.

To visualize the novel image representations, Figure 5.4 displays the images generated from the color configurations YCrQ, VCrQ, VIQ, and some existing color images R, G, B, Y, I, Q, V, Cb, Cr, as well as the images derived by the CID algorithm. For the YCrQ color configuration that achieves the best performance, the color transform coefficients produced by the proposed method are [-0.5457;-0.7322;-0.4075] for U_{YCrQ} and [-0.6568;-0.6924;-0.2988] for D_{YCrQ}, respectively. Also, the color transform coefficients derived by CID are [-0.6803;-1.0000;-0.1086] and [1.0000;0.0138;-0.2079] for the YCrQ and RGB color configurations, respectively. In particular, due to correlation in the RGB color space and that the R component has a leading performance among all color components for the FRGC database, the optimal image generated by R, G, and B components using CID is close to the R component in visualization. Obviously, the novel image representations using the hybrid color configurations inherit the image characteristics from other color components, which help improve the performance of face recognition.

Table 5.4 FVR (ROC III) at 0.1% FAR of Nine Patch Images

	Image representation	
	U-based	D-based
Patch 1	9.66%	21.21%
Patch 2	13.77%	15.33%
Patch 3	17.97%	23.77%
Patch 4	14.46%	5.84%
Patch 5	5.87%	7.69%
Patch 6	20.11%	7.88%
Patch 7	21.19%	4.07%
Patch 8	4.16%	5.59%
Patch 9	11.01%	4.17%

Table 5.5 Experimental Results Using the Patch-Based Representation

Method	FVR (ROC III) at 0.1% FAR			
	Y	U-based	D-based	Proposed one
Decision fusion	46.15%	60.10%	60.85%	63.11%
Image fusion	56.41%	68.29%	69.69%	70.55%
FLD fusion	60.82%	69.47%	70.86%	72.35%

Face recognition using patch-based novel image representation

The next set of experiments aims to evaluating the effectiveness of the proposed patch-based image representation method for face recognition. Experiments first use the patch images alone. Table 5.4 provides the FVR at 0.1% FVR for each patch image generated by both the U-based and D-based image representations. The performances show that the two image representations prefer some particular patches to others in face region, respectively, which is consistent with the conclusion reached in Table 5.2. Next, experiments assess the performance of the patch-based representation for face recognition by combining nine patch images via three ways: fusion at the decision level, fusion at the image level (i.e., the input image in Figure 5.2), and fusion at the FLD feature level (i.e., the augmented feature in Figure 5.2). Table 5.5 lists the corresponding experiment results, which are obtained from the luminance Y, the U-based image, the D-based image, and the novel one

Table 5.6 Experimental Results Using the Multi-Resolution LBP Features

Method	FVR (ROC III) at 0.1% FAR
$LBP^{u2}_{8,1}$	65.47%
$LBP^{u2}_{8,2}$	66.12%
$LBP^{u2}_{8,3}$	56.16%
FLD fusion	71.86%

Table 5.7 Experimental Results Using the Multiple Feature Fusion Strategy

	FVR (ROC III) at 0.1% FAR	
Method	Luminance Y	Proposed one
Holistic feature	71.16%	71.86%
Local feature	60.82%	72.35%
Decision fusion	76.04%	80.71%

consisting of two different image representations. The analysis of performances in Table 5.5 shows that the patch-based methods augmented by new image representation improve the performance of face recognition significantly. In particular, when considering the FVR of 56.41% of the luminance Y as the baseline, one can conclude that the improvement caused by the patch-based scheme itself is at most 5% (from 56.41% to 60.82%), while the improvement due to the new image representation is over 10% (from 56.41% to 70.55% or from 60.82% to 72.35%).

The experiments carried out so far using the FRGC data set assess only the novel image representations of the proposed face recognition framework, namely, the U_{YCrQ}, the D_{YCrQ}, and their combination. Now experiments are implemented with the multiple feature fusion using the patch-based novel image representation. The experimental results derived from the LBP features are provided in Table 5.6, which indicates that the fusion of multi-resolution LBP features via FLD indeed helps improve the face recognition performance due to the complementary characteristic among the LBP features. Finally, the face recognition performance is derived by fusing the classification outputs of the holistic and local features at the decision level. The FVRs are given in Table 5.7, where the experimental

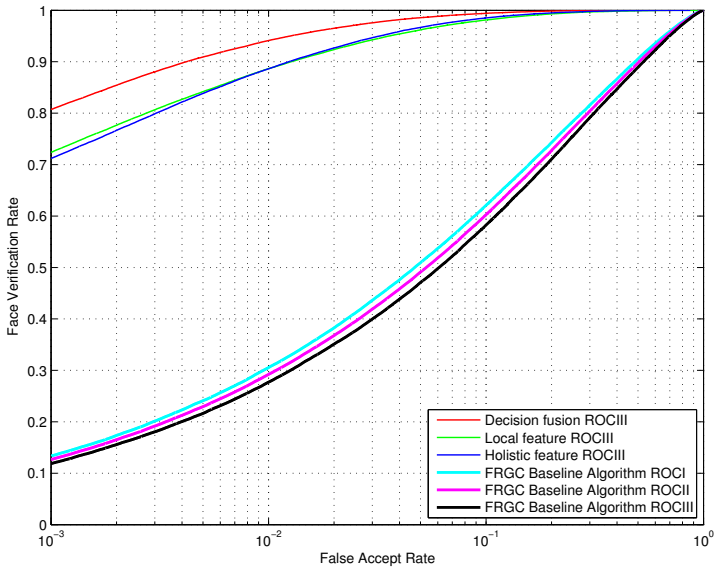

Figure 5.5 The ROC curves that are obtained using the proposed method from the classification outputs of the holistic and local features, as well as their fusion at the decision level via the sum rule. The curves derived from the FRGC baseline algorithm using the gray-scale images are also included for comparison.

Table 5.8 Experimental Results Fusing the Proposed New Image and the Y Image

Method	FVR (ROC III) at 0.1% FAR
Holistic feature on the Y	71.16%
The proposed method	80.71%
Decision fusion	84.14%

results using the luminance Y are also included for comparison. The corresponding ROC curves are shown in Figure 5.5. It has been experimentally found that fusing the classification outputs of the new image representations and the Y image can further improve the face recognition performance (Liu & Tao 2009). The results of the proposed method and the LBP features on the Y image can thus be fused through the sum rule to derive better performance, as shown in Table 5.8. In particular, the proposed method achieves the FVR (ROC III) of 84.14% at 0.1% FAR, better than performances of some recent methods such as 83.6% in (Tan & Triggs 2007b) and 74.33% in (Hwang et al. 2006) using the same database.

Table 5.9 Experimental Results Using the New Image Representations on the FERET Database

Training set	Image	Coefficient	Rank-1 accuracy	
			Dup I	Dup II
FRGC	U_{YCrQ}	[-0.5457;-0.7322;-0.4075]	78.94%	75.43%
	D_{YCrQ}	[-0.6568;-0.6924;-0.2988]	80.74%	77.19%
	D_{RGB}	[1.0000; 0.0138;-0.2079]	55.54%	38.15%
	Y	-	50.13%	28.07%
RERET	U_{YCrQ}	[-0.2677;-0.7279;-0.6313]	73.68%	73.24%
	D_{YCrQ}	[-0.2661;-0.8564;-0.4425]	73.40%	73.68%
	D_{RGB}	[0.7097; 0.4294;-1.0000]	60.38%	56.57%
	Y	-	49.20%	28.00%

5.4.2 Results of the FERET Database

Experiments are carried out on the Dup I and Dup II sets, the two most challenging probe sets from the FERET database. For the color FERET database, the image numbers of the gallery, Dup I, and Dup II probe sets are 967, 722, and 228, respectively. The first set of experiments assesses face recognition performance using the learning-based U and D images from the YCrQ hybrid color space. Besides the parameters derived from the FRGC training set, another set of parameters is generated using 1000 images from the FERET fa and fb sets. The rank-1 accuracies based on different parameters and training sets are shown in Table 5.9 using FLD and the cosine similarity measure. As can be seen, the new image representations U_{YCrQ} and D_{YCrQ} are capable of improving the performance significantly when comparing to the traditional Y image. Also, the image generated from the RGB color space using the CID model (Yang et al. 2008) is included for comparison. In particular, experimental results based on the FRGC training set demonstrate better classification performance than the FERET training set, because the large FRGC training set can avoid the *small sample size problem* that is often encountered during using the FERET training set. Hereafter, all experiments will be conducted based on the FRGC training set.

The second set of experiments evaluates the overall face recognition framework,

Table 5.10 Experimental Results Using the Proposed Method on the FERET Database, and Comparison with Two State-of-the-Art Methods

Method	Rank-1 accuracy	
	Dup I	Dup II
Holistic feature on the Y image	78.11%	63.15%
Local feature on the new image	80.05%	75.43%
Holistic feature on the new image	84.34%	79.38%
Decision fusion	91.41%	89.04%
In paper (Yao et al. 2008)	87.00%	85.00%
In paper (Tan et al. 2007(b))	90.00%	85.00%

which capitalizes on the multiple feature fusion on the patch-based novel image representation. Table 5.10 shows the experimental results based on the rank-1 face recognition rate. In particular, when fusing the classification outputs of the proposed method and the Y image, the overall performances of the Dup I and the Dip II are 91.41% and 89.04%, respectively, outperforming two state-of-the-art methods, as shown in Table 5.10.

5.5 Conclusion

This chapter introduces a novel method for face recognition. Unlike the conventional gray-scale image that is directly converted from the RGB color space, the discrimination-driven image representations can be generated by applying the learning algorithm to derive the optimal color transform vectors from some other decorrelated color configurations. By partitioning face image into several patches, the learning algorithm is applied to obtain the different image representations for different patches, thus generating a patch-based novel image representation. Finally, an effective method working on the proposed image representation is presented for face recognition. The large-scale experiments on the most challenging FRGC version 2 Experiment 4 and the FERET Dup I and Dup II probe sets have been carried out to demonstrate the effectiveness of the proposed method.

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

Face recognition is an important task in computer vision because of its immense application potential. In this dissertation, the author has taken the preliminary steps to develop a methodology to extract multiple features sets in different color spaces for face recognition, with an attempt to overcome the grand challenge issues, such as the image variabilities in terms of illumination, facial expression, aging, etc., which are often encountered in traditional methods based on intensity images. Experiments on two large-scale face databases, namely, the face recognition grand challenge (FRGC) version 2 database and the FERET database, have shown that the effectiveness of the proposed methods on addressing the challenging problems in face recognition. The contributions of this dissertation are listed below:

1. The author has developed an DFT-based method in the RIQ color space for face recognition. This method leverages the multiple DFT frequency components in the different color images, thus defining the complementary representations for face images to enhance the discriminating capability for face recognition.
2. The author has proposed a multi-resolution LBP feature fusion in the VIQ color space for face recognition. This method fuses the LBP features in three scales by Fisher Linear Discriminant analysis, thus extracting more discriminating information provided by both microstructure and macrostructure of face images than the one a single LBP operator can provide. Applying such a LBP feature in the VIQ color space, namely the "Color + LBP + LDA" strategy, further boosts the face recognition performance.
3. The author has proposed a novel method to extract multiple features in the CID color space. As different color components in the CID color space display different char-

acteristics, three different image encoding methods using the GIR, LBP, and DCT features, respectively, are presented for face recognition. To further improve classification performance, the similarity scores due to the three color component images are fused for the final decision making.

4. The author has proposed a novel image representation for face recognition. In particular, the decorrelated color components Y, Cr, and Q instead of the R, G, and B color components are used. A PCA plus FLD learning framework is presented to derive the optimal coefficients to combine the Y, Cr, and Q color components to generate a discriminating image representation for face recognition.

In the past few years, research on face recognition is moving toward recognizing faces on unconstrained real-world photos. This is largely due to an increasing need for robust face recognition for consumers to tag digital photos and facilitate their organization and online sharing. By unconstrained, the author means faces that show a large range of the variations seen in everyday life. This includes *natural* variabilities in pose, lighting, focus, resolution, facial expression, age, gender, race, make-up, and so on (Huang et al. 2007). Although traditional face databases, such as FRGC and FERET, contain a large image variability, they can be categorized into the moderately controlled database. The web-based image databases, such as Labeled Faces in the Wild (LFW) (Huang et al. 2007) that displays much more natural variations in face images, belong to the category of the uncontrolled environment. The future work will be mainly focused on the following two aspects:

Applying the color information for face recognition in real world

The proposed methods in this dissertation have shown that color is feasible and effective in improving face recognition performance on two large-scale databases, FRGC and FERET, which mainly are related to access control scenarios. Equipped with some advanced local

image descriptors, such as the learning-based (LE) descriptor (Cao et al. 2010) and the T2-S2-9 descriptor (Winder & Brown 2007), the color information could improve face matching performance achieved by some state-of-the-art methods (Hua & Akbarzadeh 2009; Cao et al. 2010) that use intensity images alone.

Applying the novel image representation for other face image analyses, such as face detection and gender classification in real world

Image-based face analysis, such as face detection and gender classification, has been an active research topic in computer vision and pattern recognition. Since some multiclass problems can be converted into two-class problems, the novel image representation learned for face recognition can be adapted to face detection and gender classification, by using two-class learning methods. To that end, the future work will apply some discriminant analysis methods, such as Asymmetric Principal and Discriminant Analysis (APCDA) (Jiang 2009) and Subclass Discriminant Analysis (SDA) (Zhu & Martinez 2006), to generate other novel image representations.

REFERENCES

Ahonen, T., Hadid, A., & Pietikäinen, M. (2006). Face description with local binary patterns: Application to face recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037-2041.

Baudat, G. & Anouar, F. (2000). Generalized discriminant analysis using a kernel approach. *Neural Computation*, vol. 12, no. 10, pp. 2385-2404.

Bebis, B., Gyaourova, A., Singh, S. & Pavlidis, I. (2006). Face recognition by fusing thermal infrared and visible imagery. *Image and Vision Computing*, vol. 24, pp. 727-742.

Belhumeur, P.N., Hespanha, J.P., & Kriegman, D.J. (1997). Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720.

Belkin, M. & Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, vol. 15, no. 6, pp. 1373-1396.

Beveridge, J.R., Bolme, D., Draper, B.A., & Teixeira, M. (2005). The CSU face identification evaluation system: Its purpose, features, and structure. *Machine Vision and Applications*, vol. 16, no. 2, pp. 128-138.

Bowyer, K.W., Chang, K., & Flynn, P.J. (2006). A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition. *Computer Vision and Image Understanding*, vol. 101, no. 1, pp. 1-15.

Cao, Z., Yin, Q., Tang, X., & Sun, J. (2010). Face recognition with learning-based descriptor. in *Proc. IEEE International Conference on Computer Vision and Pattern Recognition (CVPR'10)*.

Chan, C., Kittler, J., & Messer, K. (2007). Multi-scale local binary pattern histograms for face recognition. in *Proc. IAPR/IEEE 2nd International Conference on Biometrics (ICB'07)*.

Chen, W., Er, M.J., & Wu, S. (2006). Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain. *IEEE Trans. Systems, Man, and Cybernetics – Part B: Cybernetics*, vol. 36, no. 2, pp. 458-466.

Comon, P. (1994). Independent component analysis, a new concept? *Signal Processing*, vol. 36, pp. 287-314.

Cooke, T. (2002). Two variations on Fisher's linear discriminant for pattern recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 2, pp. 268-273.

Daugman, J.G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional cortical filters. *J. Optical Soc. Am.*, vol. 2, no. 7, pp. 1160-1169.

Donato, G., Bartlett, M.S., Hager, J.C., Ekman, P., & Sejnowski, J. (1999). Classifying facial actions. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 974-989.

Ekenel, H.K., & Stiefelhagen, R. (2006). Analysis of local appearance-based face recognition: Effects of feature selection and feature normalization. in *Proc. the 2006 IEEE International Conference on Computer Vision and Pattern Recognition Workshop*, June 17-22.

Etemad, K., & Chellappa, R. (1997). Discriminant analysis for recognition of human face images. *J. Opt. Soc. Am. A*, vol. 14, pp. 1724-1733.

Finlayson, G.D., Hordley, S.D., & Hubel, P.M. (2001). Color by correlation: A simple, unifying framework for color constancy. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1209-1221.

Fisher, R.A. (1936). The use of multiple measures in taxonomic problems. *Ann. Eugenics*, vol. 7, pp. 179-188.

Fu, Y., & Huang, T.S. (2005). *Locally linear embedded eigenspace analysis.* Tech. Rep. IFP-TR, ECE, University of Illinois Urbana-Champaign.

Fu, Y., & Huang, T.S. (2008a). Image classification using correlation tensor analysis. *IEEE Trans. Image Processing*, vol. 17, no. 2, pp. 226-234.

Fu, Y., Li, Z., Yuan, J., Wu, Y., & Huang, T.S. (2008b). Locality versus globality: query-driven localized linear models for facial image computing. *IEEE Trans. Circuits and Systems for Video Technology*, vol. 18, no. 12, pp. 1741-1752.

Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition* (2^{nd} ed.). Academic Press.

Garcia, C., & Tziritas, G. (1999). Face detection using quantized skin color regions merging and wavelet packet analysis. *IEEE Trans. Multimedia*, vol. 1, no. 3, pp. 264-277.

Geusebroek, J.M., van den Boomgaard, R., Smeulders, A.W.M., & Geerts, H. (2001). Color invariance. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 12, pp. 1338-1350.

Gonzalez, R.C., & Woods, R.E. (2002). *Digital Image Processing* (2^{nd} ed.). Prentice Hall.

Hafed, Z.M., & Levine, M.D. (2001). Face recognition using the discrete cosine transform. *International Journal of Computer Vision*, vol. 43, no. 3, pp. 167-188.

Hakyin, S. (1994). *Neural Networks - A Comprehensive Foundation*. Macmillan College Publishing Company.

He, X., Yan, S., Hu, Y., Niyogi, P., & Zhang, H.J. (2005). Face recognition using laplacianfaces. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 328-340.

Hjelmas, E., & Low, B.K. (2001). Face detection: A survey. *Computer Vision and Image Understanding*, vol. 83, pp. 236-274.

Hsu, R.-L., Abdel-Mottaleb, M., & Jain, A. K. (2002). Face detection in color images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 696-706.

Hua, G., & Akbarzadeh, A. (2009). A robust elastic and partial matching metric for face recognition. in *Proc. IEEE International Conference on Computer Vision (ICCV'09)*.

Huang, R., Liu, Q., Lu, H., & Ma, S. (2002). Solving the small sample size problem of LDA. in *Proc. IAPR International Conference on Pattern Recognition (ICPR'02)*.

Huang, G.B., Ramesh, M., Berg, T., & Learned-Miller, E. (2007). *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments.* Technical Report 07-49, University of Massachusetts, Amherst.

Hwang, W., Park, G., Lee, J., & Kee, S.C. (2006). Multiple face model of hybrid fourier feature for large face image set. in *Proc. IEEE International Conference on Computer Vision and Pattern Recognition (CVPR'06)*.

Jain, A.K., Pankanti, S., Prabhakar, S., Hong, L., & Ross, A. (2004). Biometrics: A grand challenge. in *Proc. IAPR International Conference on Pattern Recognition (ICPR'04)*, pp. 935-942.

Jain, A.K., Nandakumar, K., & Ross, A. (2005). Score normalization in multimodel biometric systems. *Pattern Recognition*, vol. 38, pp. 2270-2285.

Jiang, X., Mandal, B., & Kot, A. (2008). Eigenfeature regularization and extraction in face recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 30, no. 3, pp. 383-394.

Jiang, X. (2009). Asymmtric principal component and discriminant analyses for pattern recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 5, pp. 931-937.

Jones III, C.F. & Abbott, A.L. (2004). Optimization of color conversion for face recognition. *Eurasip Journal on Applied Signal Processing*, no. 4, pp. 522-529.

Jones III, C.F., & Abbott, A.L. (2006). Color face recognition by hypercomplex Gabor analysis. in *Proc. IEEE 7th International Conference on Automatic Face and Gesture Recognition (FGR'06)*.

Kim, Y., & Choi, S. (2007). Color face tensor factorization and slicing for illumination-robust recognition. in *Proc. IAPR/IEEE 2nd International Conference on Biometrics (ICB'07)*.

Kirby, M., & Sirovich, L. (1990). Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 103-108.

Kittler, J., Hatef, M., Robert, P.W., & Matas, J. (1998). On combining classifiers. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226-239.

Kumar, V., Savvides, M., & Xie, C. (2006). Correlation pattern recognition for face recognition. *Proceedings of the IEEE*, vol. 94, no. 11, pp. 1963-1976.

Liu, C., & Wechsler, H. (2000). Robust coding schemes for indexing and retrieval from large face databases. *IEEE Trans. on Image Processing*, vol. 9, no. 1, pp. 132-137.

Liu, C. & Wechsler, H. (2001). A shape and texture based enhanced Fisher classifier for face recognition. *IEEE Trans. on Image Processing*, vol. 10, no. 4, pp. 598-608.

Liu, C. & Wechsler, H. (2002). Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition. *IEEE Trans. on Image Processing*, vol. 11, no. 4, pp. 467-476.

Liu, C. (2006). Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 725-737.

Liu, C. (2007). The Bayes decision rule induced similarity measures. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1086-1090.

Liu, C. (2008). Learning the uncorrelated, independent, and discriminating color spaces for face recognition. *IEEE Trans. on Information Forensics and Security*, vol. 3, no. 2, pp. 213-222.

Liu, Z., & Liu, C. (2008). Fusion of the complementary discrete cosine features in the YIQ color space for face recognition. *Computer Vision and Image Understanding*, vol. 111, no. 3, pp. 249-262.

Liu, Z., Liu, C., & Tao, Q. (2009a). Learning-based image representation and method for face recognition. in *Proc. IEEE 3rd International Conference on Biometrics: Theory, Applications and Systems (BTAS'09)*.

Liu, Z., & Tao, Q. (2009b). Face Recognition Using New Image Representations. in *Proc. 2009 International Joint Conference on Neural Networks (IJCNN'09)*.

Lu, J., Plataniotis, K.N., & Venetsanopoulos, A.N. (2003). Face recognition using kernel direct discriminant analysis algorithms. *IEEE Trans. on Neural Networks*, vol. 14, no. 1, pp. 117-126.

Lu, X. (2003). *Image analysis for face recognition - A brief survey.* Retrieved from https://www.msu.edu/ lvxiaogu/publications/publications.htm.

Mika, S., Rätsch, G., Weston, J., Schölkopf, B., & Müller, K.-R. (1999). Fisher discriminant analysis with kernels. in *Proc. IEEE International Workshop Neural Networks for Signal Processing IX*, pp. 41-48.

Moghaddam, B. (2002). Principal manifolds and probabilistic subspaces for visual recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 6, pp. 780-788.

Moon, H., & Phillips, P.J. (2001). Computational and performance aspects of PCA-based face-recognition algorithms. *Perception*, vol. 30, pp. 303-321.

Ohta, Y., Kanade, T., & Sakai, T. (1980). Color information for region segmentation. *Comput. Graph. Image Process.*, vol. 13, pp. 222-241.

Ojala, T., Pietikäinen, M., & Harwood, D. (1996). A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, vol. 29, no. 1, pp. 51-59.

Ojala, T., Pietikäinen, M., & Mäenpää, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971-987.

O'Toole, A.J., Abdi, H., Jiang, F., & Phillips, P.J. (2007a). "Fusing face recognition algorithms and humans," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 37, no. 5, pp. 1149-1155.

O'Toole, A.J., Phillips, P.J., Jiang, F., Ayyad, J., Penard, N., & Abdi, H. (2007b). Face recognition algorithms surpass humans matching faces across changes in illumination. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 9, pp. 1642-1646.

Phillips, P.J., Moon, H., Rizvi, S., & Rauss, P., (2000). The FERET evaluation methodology for face recognition algorithms. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090-1104.

Phillips, P.J., Flynn, P.J., Scruggs, T., Bowyer, K.W., Chang, J., Hoffman, K., Marques, J., Min, J., & Worek, W. (2005). Overview of the face recognition grand challenge. in *Proc. the 2005 IEEE International Conference on Computer Vision and Pattern Recognition Workshop*.

Pratt, W.K. (1971). Spatial transform coding of color images. *IEEE Trans. Communication Technology*, vol. 19, no. 6, pp. 980-995.

Rajapakse, M., Tan, J., & Rajapakse, J. (2004). Color channel encoding with NMF for face recognition. in *Proc. IEEE International Conference on Image Processing (ICIP'04)*.

Ross, A., & Govindarajan, R. (2005). Feature level fusion using hand and face biometrics. in *Proc. of SPIE Conference on Biometric Technology for Human Identification II*, pp. 196-204.

Roth, V., & Steinhage, V. (2000). Nonlinear discriminant analysis using kernel functions. in *Advances in Neural Information Processing Systems*, S.A. Solla, T.K. Leen, and K.-R. Miller, eds., vol. 12, pp. 568-574, MIT Press.

Roweis, S., & Saul, L. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, vol. 290, no. 5500, pp. 2323-2326.

Scholkopf, B., Smola, A., & Miller, K. (1998). Nonlinear component analysis as a kernels eigenvalue probelm. *Neural Computation*, vol. 10, pp. 1299-1319.

Schölkopf, B., & Smola, A. (2002). *Learning with Kernels: Support Vector Machine, Regularization, Optimization and Beyond*, Cambridge, MA:MIT Press.

Shan, S., Zhang, W., Su, Y., Chen, X., & Gao, W. (2006). Ensemble of piecewise FDA based on spatial histograms of local (Gabor) binary patterns for face recognition. in *Proc. IAPR International Conference on Pattern Recognition (ICPR'06)*.

Shih, P., & Liu, C. (2005). Comparative assessment of content-based face image retrieval in different color spaces. *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 19, no. 7, pp. 873-893.

Singh, R., Vatsa, M., & Noore, A. (2008). Integrated multilevel image fusion and match score fusion of visible and infrared face images for robust face recognition. *Pattern Recognition*, vol. 41, no. 3, pp. 880-893.

Sobottka, K., & Pitas, I. (1996). Segmentation and tracking of faces in color images. in *Proc. IEEE 2nd International Conference on Automatic Face and Gesture Recognition (FGR'96)*.

Su, Y., Shan, S., Chen, X., & Gao, W. (2007). Hierarchical ensemble of global and local classifiers for face recognition. in *Proc. IEEE International Conference on Computer Vision (CVPR'07)*.

Swets, D.L., & Weng, J. (1996). Using discriminant eigenfeatures for image retrieval. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 831-836.

Tan, T., & Ikeuchi, K. (2001). Separating reflection components of textured surfaces using a single image. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 2. pp. 178-193.

Tan, X., Chen, S., Zhou, Z.-H., & Zhang, F. (2006). Face recognition from a single image per person: A survey. *Pattern Recognition*, vol. 39, no. 9, pp. 1725-1745.

Tan, X., & Triggs, B. (2007a). Enhanced local texture feature sets for face recognition under difficult lighting conditions. in *Proc. IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG'07)*.

Tan, X., & Triggs, B. (2007b). Fusing Gabor and LBP features sets for kernel-based face recognition. in *Proc. IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG'07)*.

Tao, D., Li, X., Wu, X., & Maybank, S.J. (2007). General tensor discriminant analysis and Gabor features for gait recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 10. pp. 1700-1715.

Tenenbaum, J., de Silva, V., & Langford, J.C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, vol. 290, no. 5500, pp. 2319-2323.

Terrillon, J.C., Shirazi, M.N., Fukamachi, H., & Akamatsu, S. (2000). Comparative performance of different skin chrominance models and chrominance space for the automatic detection of human faces in color images. in *Proc. IEEE 4th International Conference on Face and Gesture Recognition (FGR'00)*.

Torres, L., Reutter, J.Y., & Lorente, L. (1999). The importance of color information in face recognition. in *Proc. IEEE International Conference on Image Processing (ICIP'99)*.

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, vol. 13, no. 1, pp. 71-86.

Wang, X., & Tang, X. (2004). Dual-space linear discriminant analysis for face recognition. in *Proc. IEEE International Conference on Computer Vision and Pattern Recognition (CVPR'04)*.

Winder, S., & Brown, M. (2007). Learning local image descriptors. in *Proc. IEEE International Conference on Computer Vision and Pattern Recognition (CVPR'07)*.

Xie, C., & Kumar, V. (2005). Quaternion correlation filters for color face recognition. in *Proc. of SPIE Security, Steganography, and watermarking of Multimedia Contents*, vol. 5681.

Xie, C., & Kumar, V. (2007). Comparison of Kernel Class-dependence Feature Analysis (KCFA) with Kernel Discriminant Analysis (KDA) for face recognition. in *Proc. IEEE 1st International Conference on Biometrics: Theory, Applications and Systems (BTAS'07)*.

Xu, D., Yan, S., Zhang, L., Lin, S., Zhang, H., & Huang, T.S. (2008). Reconstruction and Recognition of Tensor-based Objects with Concurrent Subspaces Analysis. *IEEE Trans. Circuits and Systems for Video Technology*, vol. 18, no. 1, pp. 36-47.

Yan, S., Xu, D., Zhang, B., Zhang, H., Yang, Q., & Lin, S. (2007a). Graph embedding and extensions: A general framework for dimensionality reduction. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 1. pp. 40-51.

Yan, S., Xu, D., Yang, Q., Zhang, L., Tang, X., & Zhang, H.J. (2007b). Multilinear discriminant analysis for face recognition. *IEEE Trans. Image Processing*, vol. 16, no. 1. pp. 212-220.

Yang, M.H., Ahuja, N., & Kriegman, D. (2000). Face recognition using kernel eigenfaces. in *Proc. IEEE International Conference on Image Processing (ICIP'00)*.

Yang, M.H. (2002). Kernel eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods. in *Proc. IEEE 5th International Conference on Automatic Face and Gesture Recognition (FGR'02)*.

Yang, J., & Yang, J.-y. (2003). Why can LDA be performed in PCA transformed space? *Pattern Recognition*, vol. 36, pp. 563-566.

Yang, J., Frangi, A.F., Yang, J.-y., Zhang, D., & Jin, Z. (2005). KPCA plus LDA: A complete kernel Fisher discriminant framework for feature extraction and recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 2. pp. 230-244.

Yang, J., & Liu, C. (2008). Color image discriminant models and algorithms for face recognition. *IEEE Trans. Neural Networks*, vol. 19, no. 12. pp. 2088-2098.

Yang, J., Liu, C., & Zhang, L. (2010a). Color space normalization: Enhancing the discriminating power of color spaces for face recognition. *Pattern Recognition*, vol. 43, pp. 1454-1466.

Yang, J., Liu, C., & Yang, J.-y. (2010b). What kind of color spaces is suitable for color face recognition? *Neurocomputing*, vol. 73, pp. 2140-2146.

Zhao, W., Chellapa, R., Philips, J., & Rosenfeld, A. (2003). Face recognition: A literature survey. *ACM Computing Surveys*, vol. 35, no. 4, pp. 399-458.

Zhu, M., & Martinez, A.M. (2008). Subclass discriminant analysis. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 8, pp. 1274-1286.