

# Robust Face Recognition Using Color Information

Zhiming Liu and Chengjun Liu

Department of Computer Science, New Jersey Institute of Technology  
Newark, New Jersey 07102, USA  
z19@njit.edu

**Abstract.** This paper presents a robust face recognition method using color information with the following three-fold contributions. First, a novel hybrid color space, the  $RC_rQ$  color space, is constructed out of three different color spaces: the  $RGB$ ,  $YC_bC_r$ , and  $YIQ$  color spaces. The  $RC_rQ$  hybrid color space, whose component images possess complementary characteristics, enhances the discriminating power for face recognition. Second, three effective image encoding methods are proposed for the component images in the  $RC_rQ$  hybrid color space: (i) a patch-based Gabor image representation for the  $R$  component image, (ii) a multi-resolution LBP feature fusion scheme for the  $C_r$  component image, and (iii) a component-based DCT multiple face encoding for the  $Q$  component image. Finally, at the decision level, the similarity matrices generated using the three component images in the  $RC_rQ$  hybrid color space are fused using a weighted sum rule. The most challenging Face Recognition Grand Challenge (FRGC) version 2 Experiment 4 shows that the proposed method, which achieves the face verification rate of 92.43% at the false accept rate of 0.1%, performs better than the state-of-the-art face recognition methods.

## 1 Introduction

Robust face recognition methods require that algorithms be able to extract and exploit multiple facial features efficiently. Many current face recognition methods, however, start with one single feature extraction process, such as extracting the principal components by the Eigenfaces method [1] and extracting the linear discriminating components by the Fisherfaces method [2], and then utilize one classifier, such as the nearest neighbor classifier, for face recognition. In order to take advantage of the rich facial features resided in human face images, a better face recognition method should apply multiple feature extraction processes, and fuse the multiple face features for improving recognition performance. Fusion of multiple face features can take place by applying either various image features or various subspace transformed features.

The fusion of local and global features for improving face recognition performance has been studied recently [3], [4]. Generally speaking, local features such as those extracted by the Gabor [5] and the LBP [6] methods are different from

global features such as those derived by the Eigenfaces and the Fisherfaces methods, as different features deal with different scaled facial details. If the features are complementary to each other, the correlation between the outputs of the local and the global classifiers would be reduced. As a result, fusing the classification outputs at the decision level could lead to improved overall performance [7]. Currently, most methods extract the complementary features from the gray image only, which leads to fulfilling very limited complementary characteristics between the local and global features. A better method should seek a natural and much more powerful complementary face representation by capitalizing on the multiple images in a color space.

This paper presents a robust face recognition method integrating the local and global features derived from a novel hybrid color space. The motivation of our method is to explore the complementary facial information in order to boost face recognition performance by fusing their classification outputs. The novelty of our method is three-fold. First, a novel hybrid color space, the  $RC_rQ$  color space, is constructed out of three different color spaces: the  $RGB$ ,  $YC_bC_r$ , and  $YIQ$  color spaces. The  $RC_rQ$  hybrid color space, whose component images possess complementary characteristics, enhances the discriminating power for face recognition. Second, three effective image encoding methods are proposed for the component images in the  $RC_rQ$  hybrid color space: (i) a patch-based Gabor image representation for the  $R$  component image, (ii) a multi-resolution LBP feature fusion scheme for the  $C_r$  component image, and (iii) a component-based DCT multiple face encoding for the  $Q$  component image. For each method, the Enhanced Fisher Model (EFM) [8] is applied to extract features for classification. Finally, at the decision level, the similarity matrices generated using the three component images in the  $RC_rQ$  hybrid color space are fused using a weighted sum rule. The most challenging Face Recognition Grand Challenge (FRGC) version 2 Experiment 4 shows that the proposed method, which achieves the face verification rate of 92.43% at the false accept rate of 0.1%, performs better than the state-of-the-art face recognition methods.

## 2 The Proposed Method

### 2.1 A Hybrid Color Space: $RC_rQ$

Color provides powerful information for object detection, indexing and retrieval, as "humans can discern thousands of color shades and intensities, compared to about only two dozen shades of gray" [9]. Color information is helpful for improving the performance of face recognition due to the complementary characteristics among the color component images. This paper addresses face recognition in a novel hybrid color space instead of the conventional color spaces. As the  $R$  component image in the  $RGB$  color space is more effective than other component images for face recognition [10], we define a new hybrid color space  $RC_rQ$ , where  $C_r$  and  $Q$  are from the  $YC_bC_r$  color space and the  $YIQ$  color space, respectively. Fig. 1 shows the component images in the  $RC_rQ$  color space. Note that the  $R$  component image in Fig. 1 has the fine face region, which is suitable for the



**Fig. 1.** The  $RCrQ$  hybrid color space. Note that the  $R$  component image has the fine face region while the  $C_r$  and  $Q$  component images contain partial face contour information.

extraction of Gabor features, while the  $C_r$  and  $Q$  component images contain partial face contour information.

## 2.2 The Patch-Based Gabor Image Representation for the $R$ Image

The Gabor Image Representation (GIR) of an image captures salient visual properties such as spatial location, orientation selectivity, and spatial frequency characteristics [11]. Specifically, the GIR is the convolution of the image with a family of Gabor kernels that may be formulated as follows [11]:

$$\psi_{\mu,\nu}(z) = \frac{\|k_{\mu,\nu}\|^2}{\sigma^2} e^{-\frac{\|k_{\mu,\nu}\|^2 \|z\|^2}{2\sigma^2}} \left[ e^{ik_{\mu,\nu}z} - e^{-\frac{\sigma^2}{2}} \right] \quad (1)$$

where  $\mu$  and  $\nu$  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:

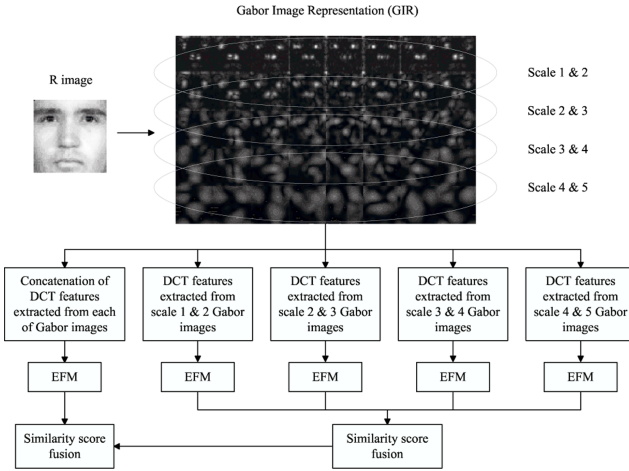
$$k_{\mu,\nu} = k_\nu e^{i\phi_\mu} \quad (2)$$

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  $R(x, y)$  represent the  $R$  component image, the convolution of  $R(x, y)$  and a Gabor kernel  $\psi_{\mu,\nu}$  may be formulated as follows:

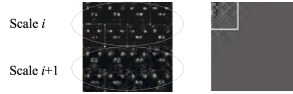
$$O_{\mu,\nu}(z) = R(z) * \psi_{\mu,\nu}(z) \quad (3)$$

where  $z = (x, y)$ ,  $*$  denotes the convolution operator, and  $O_{\mu,\nu}(z)$  is the convolution result corresponding to the Gabor kernel at orientation  $\mu$  and scale  $\nu$ . Commonly used Gabor kernels contain five different scales,  $\nu \in \{0, \dots, 4\}$ , and eight orientations,  $\mu \in \{0, \dots, 7\}$ . The set  $\mathcal{S} = \{O_{\mu,\nu}(z) : \mu \in \{0, \dots, 7\}, \nu \in \{0, \dots, 4\}\}$ , thus, forms the GIR of the image  $R$ .

The advantage of GIR stems from the integration of different spatial frequencies, spatial localities, and orientation selectivities. The GIR thus contains rich information for face recognition, which can be applied to extract features both locally and holistically. Fig. 2 shows the outline of face recognition using the GIR. For local GIR feature extraction, we separate the GIR into an ensemble of patches. The GIR is disintegrated into 4 patches along the horizontal direction, with the adjacent scale images forming one group. As the GIR patch images reside in a very high dimensional space (16 times the original image size), DCT



**Fig. 2.** The outline of face recognition using the GIR.

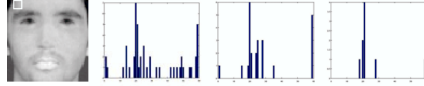


**Fig. 3.** A reshaped Gabor patch image and its DCT domain image, where  $i \in \{1, \dots, 4\}$ . A frequency set selection scheme, which selects the low frequency set located in the upper-left corner, is used to reduce the dimensionality and decorrelate the redundancy of the GIR patch images.

is applied for dimensionality reduction for improving computational efficiency. To facilitate the DCT feature extraction, each GIR patch image is reshaped to a square, as shown in Fig. 3. After transforming the GIR patch image to the DCT domain, a frequency set selection scheme using a square mask is applied to select the low frequency feature set located in the upper-left corner. Then, the Enhanced Fisher Model (EFM) is used to classify these DCT feature sets. The four GIR patches generate four similarity matrices, which are fused by means of the sum rule. For holistic GIR feature extraction, the GIR is considered as a whole for classification. In particular, each of the Gabor convolved images is processed for dimensionality reduction using DCT. The DCT features derived from the 40 Gabor images are then concatenated to form an augmented vector, which is classified by the EFM. Previous research shows that Gabor kernels with different scales help improve performance differently, which implies that different number of DCT features should be used to capture the discriminant information corresponding to the Gabor kernels with different scales. The similarity matrix generated in this approach is subsequently fused with one generated from local approach by means of the sum rule, as shown in Fig. 2.

### 2.3 The Multi-resolution LBP Feature Fusion for the $C_r$ Image

The success of Local Binary Patterns (LBP) [6] in face recognition is due to its robustness in terms of gray-level monotonic transformation. In practice, the face



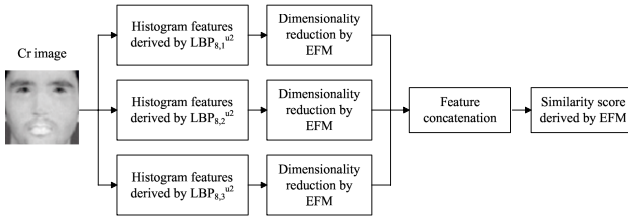
**Fig. 4.** A  $C_r$  image and three local histograms corresponding to the three scale operators:  $LBP_{8,1}^{u2}$ ,  $LBP_{8,2}^{u2}$ , and  $LBP_{8,3}^{u2}$ , from a subwindow of  $9 \times 9$  pixels

consists of uneven skin surface, which usually leads to nonmonotonic gray-level transformation. In this case, the performance of LBP degrades significantly, while Gabor kernel filters display excellent capabilities of resisting such variations. Compared with the  $R$  and  $Y$  images, the  $C_r$  image lacks the detailed information of skin surface, hence leads to less nonmonotonic gray-level transformation. Our research reveals that LBP outperforms Gabor filter in extracting discriminating features from the  $C_r$  image for face recognition.

In a  $3 \times 3$  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:

$$LBP = \sum_{p=0}^7 2^p s(i_p - i_c) \tag{4}$$

where  $s(i_p - i_c)$  equals 1, if  $i_p - i_c \geq 0$ ; and 0, otherwise. Two extensions of the basic LBP were further developed [12]. The first extension allows LBP to deal with any size of neighborhoods by using circular neighborhoods and bilinearly interpolating the pixel values. The second extension defines the so called *uniform patterns*. When the binary string is considered circular, we can call LBP uniform if there are at most two bitwise transitions from 0 to 1 or vice versa. After extensions, LBP can be expressed as:  $LBP_{P,R}^{u2}$ , where  $P, R$  means  $P$  sampling points on a circle of radius  $R$ .



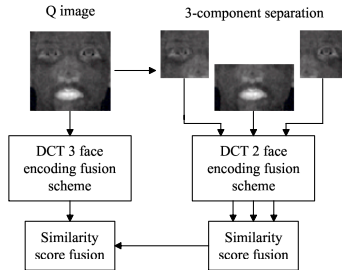
**Fig. 5.** Multiple resolution LBP feature fusion scheme

Note that the Gabor image representation encompasses the features corresponding to five scales for improving face recognition performance. Inspired by this idea, we combine the multiple-resolution information from the LBP operators. First, three LBP operators,  $LBP_{8,1}^{u2}$ ,  $LBP_{8,2}^{u2}$  and  $LBP_{8,3}^{u2}$ , are used to extract the multi-resolution histogram features from the  $C_r$  image. Fig. 4 shows three local histograms derived from three scale LBP operators in a subwindow

of an image. It is easy to see 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  $LBP_{8,1}^{u2}$ ,  $LBP_{8,2}^{u2}$  and  $LBP_{8,3}^{u2}$ . However, this operation will result in the problem of high dimensionality. In this paper, we propose an LBP multiple-resolution feature fusion scheme, as shown in Fig. 5. For each global LBP histogram, the 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 process. In particular, we first normalize and then concatenate the three reduced features and derive 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  $\mu_i$  and  $\delta_i$  are the mean and standard deviation of feature  $\mathbf{X}_{h_i}$ . 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.

#### 2.4 The Component-Based DCT Multiple Face Encoding for the $Q$ Image

In the  $YIQ$  color space,  $Y$ ,  $I$ , and  $Q$  represent luminance, hue, and saturation, respectively. In terms of face image quality, the  $Q$  component image seems contain too much noise to be applied for face recognition by local feature extraction methods, such as GIR or LBP. The rough outlines of facial components in the  $Q$  image, such as nose, eyes and mouth, however, still contain valuable discriminating information. Based on this observation, we propose a holistic method to utilize the  $Q$  component image — the fusion of component-based DCT multiple face encoding.



**Fig. 6.** Component-based DCT multiple face encoding fusion scheme

Component-based methods [13], [14] have been shown effective for improving face recognition performance, as the statistical variations caused by illumination and pose in each component image may be smaller than those in the whole face image [13]. We consider a simple separation of three facial components shown in Fig. 6. As eyes and the vicinities have the most important discriminant information, they split up into the left eye component and the right eye component.

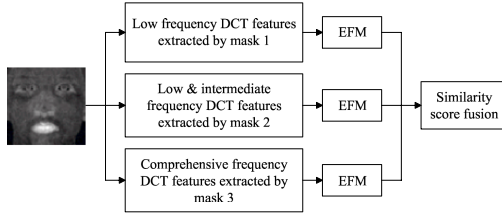


Fig. 7. DCT multiple face encoding fusion scheme

While the bottom half of face has weak discriminant capability, it is kept as an entity. Note that there are overlapping regions among the adjacent components, and each of three components is processed by the DCT 2 (using two DCT masks) face encoding fusion scheme to generate a similarity matrix. These three similarity matrices are further fused using the sum rule to generate a new similarity matrix.

The whole  $Q$  image is processed by the DCT 3 (using three DCT masks) face encoding fusion scheme to generate a similarity matrix. This similarity matrix is further fused with the similarity matrix derived using the facial components to derive the final similarity matrix.

The proposition of the DCT multiple face encoding fusion scheme is based on the observation that the reconstructed images of different DCT feature sets display different facial details. That is, these DCT feature sets are supposed to be complementary to each other. When fusing their classification outputs, the final classification result should be improved. Fig. 7 shows the outline of the three DCT face encoding fusion scheme for the whole  $Q$  image. Three masks, which are defined in the DCT domain shown in Fig. 3, are used to select three DCT feature sets. The decision of the mask size will be discussed in the experiment section.

### 3 Experiments

This section assesses the proposed method on a face recognition task using the FRGC version 2 Experiment 4 database [15]. The training set consists of 12,776 images that are either controlled or uncontrolled. The target set includes 16,028 controlled images and the query set contains 8,014 uncontrolled images. The sizes of images used in our experiments are  $128 \times 128$   $R$  images for Gabor feature extraction, and  $64 \times 64$   $C_r$  and  $Q$  images.

#### 3.1 Effectiveness of the $RC_rQ$ Hybrid Color Space

To evaluate the effectiveness of the new hybrid color space  $RC_rQ$ , we first conduct experiments on the  $Y$ ,  $C_r$ ,  $Q$  and  $R$  component images by applying the EFM with the 1,000 features and the cosine similarity measure. In particular, the  $R$  and  $Y$  images have the same face region as the  $C_r$  and  $Q$  images shown in Fig. 1. The size of all images is  $64 \times 64$ . At the decision level, prior to the similarity fusion, all similarity matrices are normalized using the z-score normalization.

**Table 1.** Face verification rates at 0.1% false accept rate of color component images

FVR (ROC III) at 0.1% FAR				
$Y$	$C_r$	$Q$	$R$	$RC_rQ$
56.41%	57.17%	53.97%	62.12%	79.98%

**Table 2.** Face verification rates at 0.1% false accept rate using the  $R$  images. Fusion I means the fusion of the four patches, and Fusion II means fusing the results of the whole GIR image and the Fusion I.

FVR (ROC III) at 0.1% FAR						
Scale 1&2	Scale 2&3	Scale 3&4	Scale 4&5	Fusion I	Whole GIR	Fusion II
68.27%	72.56%	73.99%	69.61%	82.81%	80.81%	84.12%

The results derived from the ROC III curves are listed in Table 1, which clearly indicates that the  $R$  component image carries much more discriminant information than the grayscale image ( $Y$ ). Furthermore, by fusion at the decision level, the  $RC_rQ$  hybrid color space boosts the face recognition performance significantly. Although the  $C_r$  and  $Q$  images have lower than 60% verification rates, the overall verification performance is improved significantly when they are combined with the  $R$  component image. Therefore, the  $RC_rQ$  hybrid color space constitutes an excellent platform, from which one can focus on improving performance of each component image and expect to achieve good final performance by fusing their results.

### 3.2 Effectiveness of the Proposed Method

We first conduct experiments on  $R$  images using the proposed patch-based Gabor image representation. To alleviate the effect of illumination variations, an illumination normalization procedure [4] is applied to the  $R$  images. The illumination normalization is not applied to the  $C_r$  and  $Q$  images, because the irregular intensity values in  $C_r$  and  $Q$  images usually lead to unstable illumination normalization results that compromise face recognition performance.

When the GIR is considered as a whole, we reduce the dimensionality using the DCT domain mask defined in Fig. 3. As different Gabor kernel scales contribute to the recognition performance differently, the size of masks is empirically chosen as follows to comply with such a characteristic:  $8 \times 8$ ,  $14 \times 14$ ,  $17 \times 17$ ,  $19 \times 19$ , and  $20 \times 20$ , corresponding to scales 1 to 5, respectively. The resulting feature vector with size 10,480 is then processed by EFM ( $m = 1,100$ ), producing the FVR (ROCI) of 80.81% at FAR of 0.1%. For each of the GIR patch images, the  $64 \times 64$  DCT features are chosen via masking. EFM, with  $m = 1,350$ , is used to classify these features. The face verification results using the  $R$  component image are listed in Table 2.

We then conduct experiments on  $C_r$  image using LBP features. To extract the LBP features, we divide a face image of  $64 \times 64$  into 144 ( $12 \times 12$ ) overlapping windows of  $9 \times 9$  pixels (3 pixel overlapping). The EFM ( $m = 1,300$ ) is used



**Table 3.** Face verification rates at 0.1% false accept rate of  $C_r$  image

FVR (ROC III) at 0.1% FAR			
$LBP_{8,1}^{u2}$	$LBP_{8,2}^{u2}$	$LBP_{8,3}^{u2}$	Fusion
61.04%	62.63%	53.84%	69.38%

**Table 4.** Face verification rates at 0.1% false accept rate of  $Q$  image. LE and RE mean the left eye component and the right eye component.

Image	Holistic			LE		RE		Bottom half		
Mask	$M_{15 \times 15}$	$M_{26 \times 26}$	$M_{64 \times 64}$	$M_{15 \times 15}$	$M_{39 \times 39}$	$M_{15 \times 15}$	$M_{39 \times 39}$	$M_{15 \times 15}$	$M_{39 \times 64}$	
Feature #	223	600	770	223	510	223	510	223	650	
FVR (ROC III)	41.48%	53.12%	54.46%	28.97%	32.79%	34.25%	36.41%	26.55%	32.97%	
	56.17%			33.59%		38.34%		33.40%		
						54.81%				
	58.83%									

**Table 5.** Comparison of the proposed method with the state-of-the-art

FVR (ROC III) at 0.1% FAR			
Method in [4]	Method in [3]	Method in [16]	Our method
83.6%	85.8%	87.5%	92.4%

to derive the discriminating features from each of three scale LBP histograms. After concatenation, another EFM ( $m = 390$ ) is used to process the augmented feature vector. The face verification results using the  $C_r$  component image are listed in Table 3.

We finally conduct experiments on  $Q$  images using the proposed component-based DCT multiple face encoding. In particular, for the  $Q$  image of size  $64 \times 64$ , the sizes of the left eye, the right eye region and the bottom half of face components are  $39 \times 39$ ,  $39 \times 39$ , and  $39 \times 64$ , respectively. For the DCT multiple face encoding, the selection of the DCT mask sizes affects the performance after fusion. As the number of subjects of training data is 222, the rank of the between-class scatter matrix is at most 221. In order to derive the 221 EFM features, the input feature vector should reside in a space whose dimensionality is larger than 221. We therefore choose  $M_{15 \times 15}$  as the smallest mask. The selection of the largest mask is determined by the size of images. For the holistic image, the size of the middle mask is empirically chosen to contain the low and intermediate frequencies. This size is  $26 \times 26$  in our experiments. The experiment results are given in Table 4, where the numbers of features used by EFM are also included.

After generating three similarity matrices corresponding to the three component images in the  $RC_rQ$  color space, we fuse them by means of weighted sum. In our experiments, we empirically set the weights to 1.0, 0.6, and 0.8, respectively, based on the different roles of the  $R$ ,  $C_r$ , and  $Q$  component images for face recognition. The final face verification results are shown in Table 5. Specifically,

our proposed method, which achieves the face verification rate of 92.43% at the false accept rate of 0.1%, performs better than the state-of-the-art face recognition methods, such as [4] with the face verification rate of 83.6%, [3] with the face verification rate of 85.8%, and [16] with the face verification rate of 87.5% at the same false accept rate using the same face image database.

## References

1. Turk, M., Pentland, A.: Eigenfaces for recognition. *Journal of Cognitive Neuroscience* 13, 71–86 (1991)
2. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Analysis and Machine Intelligence* 19, 711–720 (1997)
3. Su, Y., Shan, S., Chen, X., Gao, W.: Hierarchical ensemble of global and local classifiers for face recognition. In: *Proc. IEEE International Conference on Computer Vision (ICCV 2007)* (2007)
4. Tan, X., Triggs, B.: Fusing gabor and lbp feature sets for kernel-based face recognition. In: Zhou, S.K., Zhao, W., Tang, X., Gong, S. (eds.) *AMFG 2007. LNCS*, vol. 4778, pp. 235–249. Springer, Heidelberg (2007)
5. Liu, C., Wechsler, H.: Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition. *IEEE Trans. on Image Processing* 11, 467–476 (2002)
6. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: Application to face recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence* 28 (2006)
7. Kittler, J., Hatef, M., Robert, P., Matas, J.: On combining classifiers. *IEEE Trans. Pattern Analysis and Machine Intelligence* 20, 226–239 (1998)
8. Liu, C., Wechsler, H.: Robust coding schemes for indexing and retrieval from large face databases. *IEEE Trans. on Image Processing* 9, 132–137 (2000)
9. Gonzalez, R., Woods, R.: *Digital Image Processing*, 2nd edn. Prentice-Hall, Englewood Cliffs (2002)
10. Shih, P., Liu, C.: Comparative assessment of content-based face image retrieval in different color spaces. *International Journal of Pattern Recognition and Artificial Intelligence* 19, 873–893 (2005)
11. Daugman, J.: Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional cortical filters. *J. Optical Soc. Am.* 2, 1160–1169 (1985)
12. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Analysis and Machine Intelligence* 24, 971–987 (2002)
13. Kim, T., Kim, H., Hwang, W., Kittler, J.: Component-based lda face description for image retrieval and mpeg-7 standardisation. *Image and Vision Computing* 23, 631–642 (2005)
14. Heisele, B., Serre, T., Poggio, T.: A component-based framework for face recognition and identification. *International Journal of Computer Vision* 74, 167–181 (2007)
15. Phillips, P., Flynn, P., Scruggs, T., Bowyer, K., Chang, J., Hoffman, K., Marques, J., Min, J., Worek, W.: Overview of the face recognition grand challenge. In: *Proc. IEEE Conference on Computer Vision and Pattern Recognition* (2005)
16. Kumar, V., Savvides, M., Xie, C.: Correlation pattern recognition for face recognition. *Proceedings of the IEEE* 94, 1963–1976 (2006)