Robust Face Recognition by Fusion

K. Srinivasa Rao Lilesh S. Ghadi Y. Anoop Kumar A. N. Rajagopalan *
Image Processing and Computer Vision Laboratory
Department of Electrical Engineering
IIT Madras, Chennai - 600 036
email:{ raju@ee.iitm.ernet.in}

Abstract

We propose a novel probabilistic framework that combines information acquired from different facial features for robust face recognition. The features used are the entire face, the edginess image of the face, and the eyes. In the training stage, individual feature spaces are constructed using Principle Component Analysis (PCA) and Fisher's Linear Discriminant (FLD). By using the distance in feature space (DIFS) values of the training images, the distributions of the DIFS values in each feature space are computed. For a given image, the distributions of the DIFS values yield confidence weights for the three facial features extracted from the image. The final score is computed using a probabilistic fusion criterion and the match with the highest score is used to establish the identity of the person. A new pre-processing scheme for illumination compensation is advocated.

1. Introduction

Automatic face recognition is a challenging problem in computer vision. Computers that recognise faces can be applied to a wide variety of problems, including criminal identification, security systems, and human-computer interactions. In recent years, many different techniques have been applied to this task, and there is considerable literature on face recognition.

Here, we propose a novel probabilistic scheme for fusing different facial features for robust face recognition. Facial features of the face when considered along with the entire face image provide cues for better discrimination. Our main contribution is in formulating a mathematical framework to integrate the information coming from multiple facial features. The three facial features that we consider are the entire face, the edginess image of the face [1, 2], and the eyes. The edginess image is robust to variations in illumination while the eyes are robust to facial expressions and occlusions. We use Principle Component Analysis (PCA) [3] in conjunction with Fisher's Linear Discriminant (FLD) [4, 5] to encode the facial features in a lower-dimensional feature space. Three individual spaces are constructed corresponding to the three facial features. The Distance In Feature Space (DIFS) values are calculated for the training images in each of the feature spaces and these values are used to compute the distributions of the DIFS values. The distributions of the DIFS values play an important role in characterizing the differences between imposters and true persons. Given a test image, the three facial features are first extracted and their DIFS values are computed in each feature space. Each feature provides an opinion on the claim in terms of a confidence value which is measured by integrating the DIFS distributions of each feature space w.r.t the DIFS value computed in that feature space. The confidence values of all the three features are fused for final recognition. The identity established by our fusion technique is more reliable compared to the case when features are used individually. As a preprocessing step, we also propose a new Block Histogram Equalization (BHE) technique that is quite effective in compensating for local changes in illumination. When tested on the standard FERET database with variations in facial expressions and ambient illumination, the proposed fusion method yields significant improvement in final recognition accuracy over the accuracy achieved with face only. The work proposed here is an extension of our work in [2].

*Funding from MHRD, India is gratefully acknowledged.

2. Feature Selection and Pre-Processing

In this section, we first explain the motivation for the choice of the specific facial features that we have considered in this paper. This is followed by a novel illumination compensation method which serves as a preprocessor.

2.1. Facial Features

Face recognition approaches that consider only the entire face as a feature do not take into account just what other aspects of the face stimuli are important for recognition [7]. Utilizing complementary information should improve performance. For our recognition algorithm, we have considered, in addition to the entire face image, two other features; namely, the edginess image of the face, and the eyes. The motivation for incorporating local features into a recognition system stems from the fact that it is possible for humans to recognise a face from only parts of it.

The edginess image is a global facial feature that is reasonably robust to illumination. It is a measure of the change in intensity from one pixel to the next. To extract a good edginess image map, we employ 1-D processing [1] along orthogonal directions as follows. To detect the horizontal component of edginess, a 1-D Gaussian filter is first used to smooth the image horizontally. This helps in reducing the effect of noise. The Gaussian smoothing filter is given by

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-x^2}{2\sigma^2}} \tag{1}$$

where σ is the standard deviation of the filter. A discrete approximation of this filter appears in Table 1(a). A differential operator which is a first-order derivative of the 1-D Gaussian function is used next in the orthogonal direction (i.e., vertically) to find the horizontal component of edginess. The differential operator is given by

$$c(y) = \frac{-y}{\sqrt{2\pi}\sigma^3} e^{\frac{-y^2}{2\sigma^2}} \tag{2}$$

and its discrete approximation is given in Table 1(b).

.0001	.0440	.054	.242	.3989	.242	.054	.0440	.0001

(a)

.0005	.0133	.108	.242	0.0	-.242	-.108	-.0133	-.0005

(b)

Table 1. Filter coefficients for a typical (a) 1-D Gaussian filter, and (b) Differential operator.

(a) (b) (c)

Figure 1. (a) A gray-scale face image, (b) it's edginess image, and (c) the cropped eyes.

The vertical component of edginess is computed in a similar manner by carrying out the above steps in the orthogonal direction. The final edginess image is obtained by taking the absolute sum of the horizontal and the vertical components. Figure 1 shows a gray-scale face image, the corresponding edginess image, and the extracted eyes. Note that the edginess image is also a gray-valued image.

2.2. Intensity Normalization

A face recognition system must recognize a face from it's novel image despite variations in illumination. Unfortunately, till todate, no revolutionary solution exists for the intensity normalization problem. However, approaches have been proposed to alleviate the effect of illumination variations [8].

We propose a simple Block Histogram Equalization (BHE) technique for illumination compensation. We assume that a reference image taken under well-controlled lighting conditions is available. Let X and Y be the input and the reference images, respectively, of size $N \times N$ pixels. The goal is to bring the illumination level of the input image X to that of the reference image Y by applying BHE. Consider a block image B_I from the input image X with pixel locations ranging from 1 to M and also a block image B_R from the reference image Y at the corresponding pixel locations (Figure 2). We would like to apply histogram modification to the input image block B_I to make the pixel intensity distribution of B_I equivalent to the pixel intensity distribution of B_R.

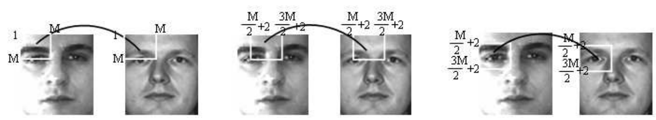

Figure 2. Block histogram equalization. In each image pair, the left one is the input image while the right one is the reference image.

Consider the input block image (i.e., B_I) with pixel value $x \geq 0$ to be a random variable with probability density function $p_x(x)$ and cumulative probability distribution $F_x(x)$ given by $F_x(x) = \int_0^x p_x(u)du$. Let the reference block image (i.e., B_R) with pixel intensity $y \geq 0$ be a random variable with probability density function $p_y(y)$ and cumulative probability distribution $F_y(y)$ given by $F_y(y) = \int_0^y p_y(u)du$. The final output block image B_O with pixel intensity value $z \geq 0$ should have the density function $p_y(y)$ and cumulative distribution $F_y(y)$ and is given by

$$z = F_y^{-1}[F_x(x)]. \tag{3}$$

The histogram modified block image intensity values are scaled with a windowing filter H is given by

$$B_O(n,m) = B_O(n,m)H(n,m) \quad 1 \leq n,m \leq M \tag{4}$$

$$H(n,m) = \begin{cases} \frac{4nm}{M^2} & 1 \leq n,m \leq \frac{M}{2} \\ \frac{4m(M-n+1)}{M^2} & \frac{M}{2} < n \leq M, 1 \leq m \leq \frac{M}{2} \\ \frac{4n(M-m+1)}{M^2} & 1 \leq n \leq \frac{M}{2}, \frac{M}{2} < m \leq M \\ \frac{4(M-n+1)(M-m+1)}{M^2} & \frac{M}{2} < n,m \leq M \end{cases} \tag{5}$$

By simultaneously shifting the blocks in both the horizontal and the vertical directions in steps of $\frac{M}{2}+1$ pixel locations (as shown in Figure 2), and adding pixel intensity values in overlapping regions, we arrive at the final image Z. The intensity changes are smoothed out across adjacent blocks. The blocks are overlapped to avoid edges and patches from appearing in the illumination compensated image. The window H is defined such that the sum of the weights in the overlapping region is 1. In Figure 3, we give a few examples of images taken under different illumination directions and the corresponding intensity normalized images using our method. The reference image was kept the same for all the images. The proposed BHE intensity normalization technique is simple to implement and effective.

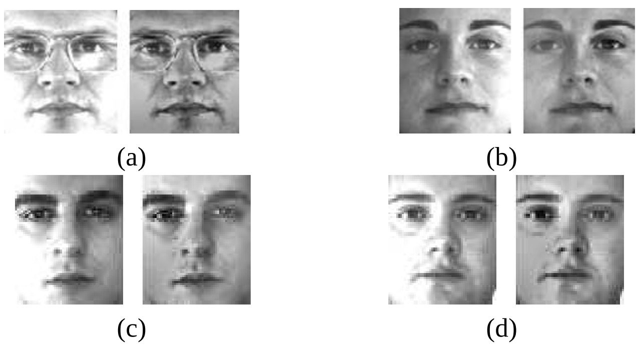

Figure 3. Images before and after intensity normalization with BHE.

3. Eigen and Fisher Analysis

In this section, we discuss Eigen and Fisher theory in brief. The analysis here is common to all the facial features considered in this paper. Let $f_{i,m}$ denote the N^2 element training vector representing the m^{th} image (size of $N \times N$) of the i^{th} person. Let the average image of the entire data set be ψ and the covariance matrix be C [3]. Using PCA, the weight vector corresponding to the m^{th} training image $\phi_{i,m}$ of the i^{th} person can be derived as $w_{i,m} = E_{pca}^T \phi_{i,m}$ where E_{pca} contains the significant eigenvectors of C given by $E_{pca} = [e_1, e_2, ..., e_{K'}]$. The average weight vector of the i^{th} person is given by $w_i = \frac{1}{M}\sum_{m=1}^{M} E_{pca}^T \phi_{i,m}$

Fisher's Linear Discriminant (FLD) [5] is a class-specific method to shape the scatter in order to make it more reliable for classification. Mathematically, FLD selects the projection matrix E_{fld} in such a way that the ratio of the determinant of the between-class scatter matrix S_b to the within-class scatter matrix S_w of the projected samples is maximized [5]. We give as input to the FLD the reduced dimension weight vector derived from PCA. The projection matrix E_{fld} is chosen as

$$E_{fld} = \underset{E'}{\arg\max} \frac{|E'^T S_b E'|}{|E'^T S_w E'|} \tag{6}$$

The final projection matrix E_{opt} is given by $E_{opt}^T = E_{fld}^T E_{pca}^T$. The final mean weight vector corresponding to the i^{th} person is then derived as $w_i' = \frac{1}{M}\sum_{m=1}^{M} E_{opt}^T \phi_{i,m}$

In the recognition stage, a new test image feature γ is transformed into its respective feature space and the weight vector is derived as $w = E_{opt}^T(\gamma - \psi)$. The simplest method for determining which class provides the best description of the test image is to find the class j that minimizes the Distance In Feature Space (DIFS) value computed as

$$\varepsilon_i = \| (w - w_i') \| \qquad 1 \leq i \leq I \tag{7}$$

The test image is classified as belonging to class j if the minimum value corresponds to ε_j and is below some chosen threshold θ.

4. Recognition by Fusion

In this section, we propose a probabilistic fusion strategy to integrate information coming from multiple facial features. To assign confidence weights, we propose to compute the distributions of the DIFS values in each feature space. The distributions are calculated empirically from the training data as explained below.

4.1. DIFS Distributions

As will be shown, the DIFS distributions are very useful in characterising how an imposter will differ from the true person.

In order to simplify the analysis, we consider a single feature (say) face. After constructing the face space using PCA and FLD (as discussed in Section 3), the training data itself is used to compute the DIFS distributions. The training images are transformed to the face space and their DIFS values are computed w.r.t. all the identities in the database. Let ϵ^α (a random variable with density function $f^\alpha(\epsilon^\alpha)$)

represent the DIFS values between an individual's images and his own weight vector. Thus, ϵ^{α} describes the distribution of the DIFS values for a true person. Without loss of generality, we assume that for a given individual, ϵ_1, ϵ_2,...., ϵ_I are the DIFS values arranged in an increasing order and that these are statistically independent. Let ϵ_i (a random variable with density function $f_i(\epsilon_i)$) denote the DIFS value at rank i. For a given individual, the relative difference in the DIFS value at rank i, defined as the difference between the DIFS value at rank i to his/her own DIFS value is

$$\Delta_i = \epsilon_i - \epsilon^{\alpha} \qquad 1 \leq i \leq I \tag{8}$$

Note that I is the total number of individuals/identities in the database. The DIFS distribution $f_i(\Delta_i)$ at rank i is

$$f_i(\Delta_i) = \int_{-\infty}^{\infty} z_i(\epsilon_i, \Delta_i) \;\; d\epsilon_i \tag{9}$$

where

$$z_i(\epsilon_i, \Delta_i) = \frac{I!}{(I-i)!(i-1)!} F(\epsilon_i)^{i-1} [1 - F(\epsilon^{\alpha})]^{I-i} f(\epsilon_i) f(\epsilon^{\alpha})$$

Here, $z_i(\epsilon_i, \Delta_i)$ describes the joint distribution of the i^{th} rank DIFS value (i.e., ϵ_i) and its distance from the true person's DIFS value (i.e., ϵ^{α}). The term $f(\epsilon_i)$ is the probability density function of the absolute DIFS value at rank i and $F(\epsilon_i)$ is the corresponding distribution function. For more details on the distribution of distances, see [9]. Since closed form expressions for $f(\epsilon_i)$, $f(\epsilon^{\alpha})$, $F(\epsilon_i)$, and $F(\epsilon^{\alpha})$ are not available for the problem on hand, we derive $f_i(\Delta_i)$ and $f(\varepsilon^{\alpha})$ empirically by directly using the DIFS values computed from the training dataset. The mean value of the DIFS distributions $f_i(\Delta_i)$ increases with rank i which implies that an imposter at higher ranks is quite different from the actual person in terms of the DIFS values.

4.2. Recognition

The confidence weights of the identities are computed according to the DIFS values and the positional ranking in the feature space in which they appear. The three features (face, edginess, and eyes) are assumed to be independent. We make this approximation for mathematical convenience. The weight to be assigned to rank i depends on the positional ranking as well as the proximity of the DIFS value at rank i with the top rank DIFS value.

When a new test image γ arrives, it's DIFS values $\epsilon_1^{'}, \epsilon_2^{'}, ..., \epsilon_I^{'}$ are arranged in an increasing order (in the respective feature spaces). Let $\epsilon^{\alpha^{'}}$ denote the DIFS value for the top rank in the face space. The relative DIFS values $\Delta_i^{'}$ are computed using equation (8). The confidence weight assigned to the identity at rank *i* is computed as

$$P_{face}(i) = \Big[P_i(\Delta_i^{'}) \;\; P_{order}(i)\Big]_{face} \qquad 1 \leq i \leq I \tag{10}$$

where $P_i(\Delta_i^{'})$ describes how close the i^{th} rank DIFS value is to the top rank and

$$P_i(\Delta_i^{'}) = \int_{\Delta_i^{'}}^{\infty} f_i(\Delta_i) \;\; d\Delta i \tag{11}$$

Note that for a feature the top rank identity need not always be the correct identity. For example, let us assume that the 3rd rank identity is the actual person. Then, the DIFS values for rank 1 and 2 cannot be much less than the DIFS value at rank 3 (i.e., $\Delta_2^{'}$ and $\Delta_3^{'}$ will be small). Depending on the DIFS value at rank 3 and the distribution of $f_3(\Delta_3)$, the weight assigned to the 3rd rank could even be higher than the weight assigned to the 2nd rank identity. This allows us to accommodate higher-ranked identities also.

The term $P_{order}(i)$ assigns an appropriate weight to an identity depending on the positional ranking as well as the top rank DIFS value. If the top rank DIFS value is very small, then he/she is most likely the actual identity. Hence, $P_{order}(i)$ should fall very sharply as rank i increases. If the top rank DIFS value is large, then the top person may not be the actual identity, and hence the confidence weight should fall gradually to accommodate even individuals at higher ranks. Thus, depending on the top rank DIFS value, we give relative weightage to the person at rank i. We define $P_{order}(i)$ using a Gamma distribution with λ = 1 and β=1 as

$$P_{order}(i) = \frac{\beta}{\gamma(\lambda)} \left(\frac{(i-1)}{\theta}\right)^{(\lambda\beta - 1)} e^{(-\frac{(i-1)}{\theta})}, \;\; \theta = \frac{1}{p}, \tag{12}$$
$$1 \leq i \leq I \quad \text{and} \quad p = \int_{\varepsilon_1^{'}}^{\infty} f^{\alpha}(\varepsilon^{\alpha}) \; d\varepsilon^{\alpha}$$

In the above equation, $\varepsilon_1^{'}$ is the DIFS value at rank 1. For p=1, the Gamma curve falls very sharply compared to other values of p. When the top rank identity is not a genuine one, the combined effect of $P_i(\Delta_i^{'})$ and $P_{order}(i)$ is to accommodate identities at even higher ranks. In an exactly similar manner, we compute $P_{edge}(i)$ and $P_{eye}(i)$ for the other two features. If $I_1, I_2, ..., I_I$ are the identity indicators of the individuals in the database, then the final confidence weight of an identity I_i is obtained by multiplying the confidence weights contributed from each feature space of that identity

$$P(I_i) = P_{face}(I_i) P_{edge}(I_i) P_{eye}(I_i) \tag{13}$$

Here, $P_{face}(I_i), P_{edge}(I_i)$, and $P_{eye}(I_i)$ are confidence values acquired from each feature for the identity I_i. The identity ID(γ) for a given individual γ is determined by the following criterion:

$$ID(\gamma) = \begin{cases} I_k & \text{if } P(I_k) > \tau \\ \text{Imposter} & \text{otherwise} \end{cases}$$
$$\text{where} \quad P(I_k) = max\{P(I_1), P(I_2), ..., P(I_I)\} \tag{14}$$

The threshold τ is chosen such that an untrained person should not be recognised at all.

5. Experimental results

In this section, we demonstrate the performance of the proposed method on the standard FERET database. A commonly used performance measure for face recognition is the Cumulative Match Scores (CMS), i.e., the recognition accuracy in the top n ranks. The required facial features were cropped with reference to the eye locations which were provided with the dataset. The eye locations were used to account for rotation and scaling, when necessary. All images were intensity normalized using the BHE technique described in Section 2.2.

Probe Category	Gallery size	Probe set size
FaFb	1196	1195
duplicate I	1196	194
FaFc	1196	722
duplicate II	1196	234

Table 2. Gallery and probe information for FERET.

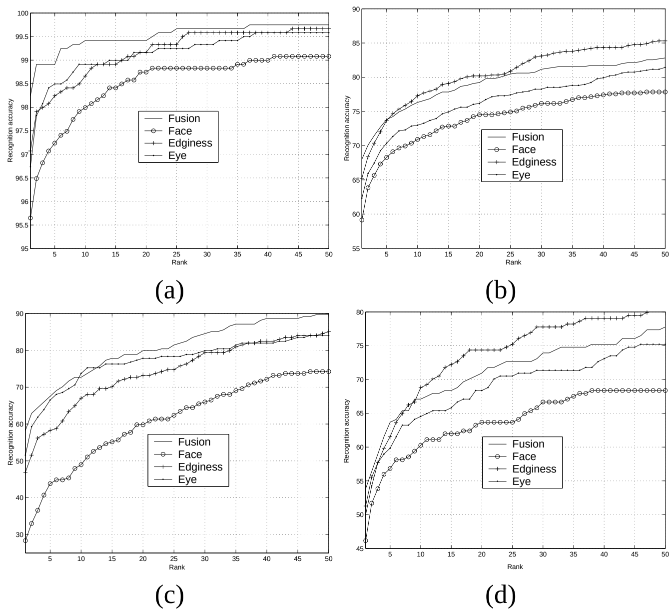

Figure 4. CMS plots for (a) the FaFb probe set, (b) duplicate I, (c) the FaFc probe set, and (d) duplicate II.

The FERET database contains 14,126 images comprising of 1,199 individuals. Since the images are acquired during different photo sessions, this dataset contains significant variations in pose, illumination and facial expressions. We have compared our system performance with the FERET evaluation results [6]. The FERET evaluation in [6] provides a comprehensive picture of the state-of-the-art in face recognition. In Table 2, details of all the four probe categories are given. The FA images (regular frontal faces of persons) were used as the gallery set, where as four categories of probe sets were used to compare against the gallery set. The first probe category was the FaFb probe set. This indicates an alternative frontal image, taken seconds after the corresponding FA images. The second probe category contained all duplicate frontal images in the FERET database and is referred to as the duplicate I probe set. The third category of probe set is the FaFc set which contains images taken on the same day but with different camera and illumination. The fourth category of probe set is called the duplicate II set. These images are duplicates of FA images but taken at least one year between the acquisition of the gallery images (FA) and the probe images.

When tested on all the probe categories, the CMS plots for the top n ranks are shown in Figure 4. In Table 3, we have compared the performance of the proposed method with the partially automatic face recognition algorithms that appear in [6]. From Table 3, we observe that the performance of the proposed method is comparable to the best reported results. For the FaFb, duplicate I and duplicate II probe sets, our method has better accuracy compared to others. On the FaFc probe set, we come second.

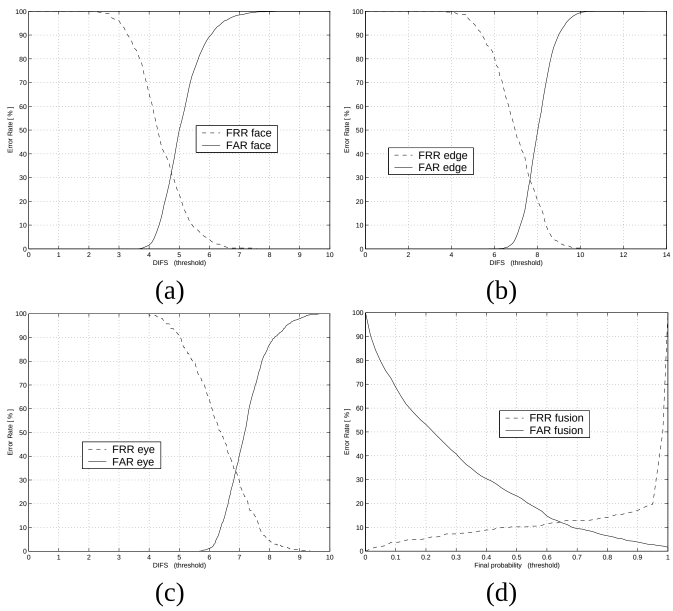

Figure 5. FAR and FRR plots for the FERET database using (a) face only, (b) edginess only, (c) eyes only, and (d) fusion.

The performance of our system was next tested on untrained people to check how well it rejects unknown persons. This is done using the False Acceptance Rate (FAR) and the False Rejection Rate (FRR) curves. FAR is the probability that an untrained person is falsely accepted as a known identity while FRR is the probability that a known

Probe set	Recognition accuracy at rank 1 [%]					
	Fusion	UMD 97	USC	MIT 96	Baseline cor	Baseline EF
FaFb	98.3	96.5	95	94.8	82.5	79.5
duplicate I	68	46	58	57	35	42
FaFc	59	59	82	32	7	18
duplicate II	54	21	46	34	16	22

Table 3. Recognition accuracy for different algorithms with the FERET database.

person is falsely rejected as an unknown person. The relation between the two rates is controlled by the acceptance threshold of the system. If the threshold is set to a very high value, there will be no false acceptances (i.e., FAR = 0), but it will be impossible to accept even a true (known) person who is in the training data (i.e., FRR = 100%). Setting too low a threshold will cause the situation to reverse. The value of FRR and FAR at the point where the plots cross is called the Equal Error Rate (ERR). System performance can be specified in terms of the ERR value. For a good recognition system, the ERR value should be as small as possible.

For the FERET database, when our method was tested against trained and untrained individuals, the FAR and FRR plots for face, edginess, eyes, and fusion are shown in Figure 5(a), (b), (c), and (d), respectively. The system was trained with 482 individuals out of the 1,199 individuals in the database. A total of 1,446 images were used for training, 3 images per subject. To compute the FAR plots, 1,440 face images were used as probe images from the remaining 717 untrained individuals. Initially, we fixed the acceptance threshold for each feature space and also for the fusion method. For face, edginess, and eyes, the threshold value is in terms of the top rank DIFS value, whereas for the fusion method it is in terms of the final confidence value of the top rank. We compute the error rate by testing on all the probe images. The value of the error rate indicates the ratio between the number of face images that are accepted as known identities over the total number of images. Error rates were computed for all the threshold values by varying the threshold value from the lowest possible to the highest possible value. While computing FRR, 304 untrained facial images were used as probe images from 482 trained individuals (i.e., known identities) and we computed the error rates for all the threshold values. From Figure 5, we observe that the ERR values for face, edginess, and eye are quite high (31%, 30%, 34%, respectively). In contrast, the ERR value for the fusion method is much lower and is only 12%. Thus, the fusion method is very useful in rejecting imposters.

6. Conclusions

We have described a system that uses different facial features for robust recognition. The proposed probabilistic fusion scheme combines information coming from the face, the edginess image of the face, and the eyes. A new algorithm for illumination compensation is also given. The method has been validated by testing it on the FERET database. It has been shown that fusion improves overall recognition accuracy. The improvement is particularly significant under facial occlusions, variations in facial expressions, and illumination changes.

References

[1] B.S. Venkatesh, S. Palanivel, and B. Yegnanarayana, “Face detection and recognition in an image sequence using eigenedginess,” *Indian Conference on Computer Vision, Graphics and Image Processing,* (Ahmedabad, India), pp. 97-101, (2002).

[2] K. Srinivasa Rao and A.N. Rajagopalan, “Probabilistic fusion of facial features,” *Workshop on Computer Vision Graphics and Image Processing,* (Gwalior, India), pp. 45-48, (2004).

[3] M. Turk and A. Pentland, “Eigenfaces for recognition,” *Journal of Cognitive Neuroscience,* vol. 3, pp. 76-96, 1991.

[4] P.N. Belhumer, J.P. Hespanha, and D.J. Kriegman, “Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 19, pp. 711-720, 1997.

[5] R. Fisher, “The use of multiple measures in taxonomic problems,” *Ann. Eugenic,* vol. 7, pp. 179-188, 1936.

[6] P.J. Phillips, H. Moon, S.A. Rizvi, and P.J. Rauss, “The FERET evaluation methodology for face-recognition algorithms,” *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 22, pp. 1090-1104, 2000.

[7] P.S. Penev and J.J. Atick, “Local feature analysis: A general statistical theory for object representation,” *Network: Computation in Neural Systems,* vol. 7, pp. 477-500, 1996.

[8] A. Shashua and T. Riklin-Raviv, “The Quotient Image: Class-based re-rendering and recognition with varying illuminations,” *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 23, pp. 129-139, 2001.

[9] E.J. Gumbel, “*Statistics of extremes*”, Columbia University Press. (1958).