

# A Probabilistic Fusion Methodology for Face Recognition

**K. Srinivasa Rao**

*Image Processing and Computer Vision Laboratory, Department of Electrical Engineering, Indian Institute of Technology Madras, Chennai-600 036, India*  
*Email: srinu\_iitm2003@yahoo.co.in*

**A. N. Rajagopalan**

*Image Processing and Computer Vision Laboratory, Department of Electrical Engineering, Indian Institute of Technology Madras, Chennai-600 036, India*  
*Email: raju@ee.iitm.ernet.in*

*Received 11 May 2004; Revised 30 March 2005; Recommended for Publication by Satya Dharanipragada*

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 principal 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 a person. A new preprocessing scheme for illumination compensation is also advocated. The proposed fusion approach is more reliable than a recognition system which uses only one feature, trained individually. The method is validated on different face datasets, including the FERET database.

**Keywords and phrases:** face recognition, block histogram modification, edginess image, probabilistic fusion, distance in feature space.

## 1. INTRODUCTION

Automatic face recognition is a challenging problem in computer vision. Computers that recognize faces can be applied to a wide variety of problems such as information mining from face databases, security systems, and human-computer interactions [1]. Existing face recognition methodologies may be broadly classified into two categories, holistic and analytic [2]. Here, we refer to some well-known works in both categories. The survey is by no means exhaustive. Holistic approaches consider global properties of the face. The eigenface-based face recognition system proposed by Turk and Pentland [3] uses principal component analysis (PCA) to compute linear projections of face images to arrive at a compact representation. The method has good recognition rate but is sensitive to variations in facial expressions and ambient illumination. For higher discriminability, Fisher's linear discriminant (FLD) analysis in conjunction with PCA has been proposed in the literature [4, 5]. Elastic graph matching for face recognition is yet another popular approach [6]. It uses a novel dynamic-link architecture

based on multiscale morphological dilation and erosion for authentication of frontal images. The idea is to weight the graph nodes according to their discriminating power. Duc et al. [7] propose automatic weighting of nodes by employing local discriminants. In analytic approaches, a set of geometrical features acquired from a face are used for recognition. Brunelli and Poggio [8] have proposed template matching for face recognition. Wu and Huang [1] have developed a recognition analysis based on feature vectors extracted from profiles. A face profile is first reduced to outline curves, and then a number of fiducial marks are extracted and used for recognition. Recognition systems that combine holistic and analytic approaches are known to be more reliable. In recent years, multimodal identification techniques have been receiving a lot of attention. The idea is to integrate different biometric cues such as face, fingerprint, speech, and more recently gait [9, 10, 11, 12] for improving overall recognition performance.

In this paper, we propose a novel probabilistic scheme for fusing different facial features for robust face recognition. Facial features when considered along with the entire

TABLE 1: Filter coefficients for a typical 1D Gaussian filter and differential operator.

1D Gaussian filter	Differential operator
0.0001	0.0005
0.044	0.0133
0.054	0.1080
0.242	0.2420
0.3989	0.0
0.242	-0.2420
0.054	-0.1080
0.044	-0.0133
0.0001	-0.0005

face image provide cues for better discrimination [8, 13]. 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 (i.e., the gray-level image of the face), the edginess image of the face [14, 15], and the eyes. The edginess image is robust to variations in illumination while the eyes are robust to facial expressions and occlusions. We use principal component analysis (PCA) [3, 16] 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 all the images in the training set and in each of the feature spaces. These values are used to compute the distributions of the DIFS values which play an important role in characterizing the differences between imposters and true persons. Given a new 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 with respect to 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 propose a new block histogram equalization (BHE) technique that is quite effective in compensating for local changes in illumination.

The performance of our method has been evaluated on three different face datasets, namely, the AR face database [17], the IPCV Lab database (collected in our laboratory), and the FERET database [18]. The performance is evaluated in terms of a cumulative match score (CMS) which describes how well the system is able to recognize a person in the top  $n$  ranks. When compared with the standard FERET evaluation test results [18], our algorithm is found to perform quite well on all the probe sets in the FERET evaluation. It compares favorably with the best face recognition algorithms that currently exist in the literature [18]. The ability to reject unknown people (i.e., people not in the training dataset) is also important in face recognition. We evaluate the performance

of our algorithm on trained as well as untrained individuals in terms of the false acceptance (FA) and the false rejection (FR) rates. We show that our system is very good at recognizing trained individuals and at rejecting untrained people.

In Section 2, we discuss facial features and a preprocessing step for intensity normalization. Section 3 describes generalized Eigen and Fisher analysis for each feature. A probabilistic fusion methodology is proposed in Section 4. Experimental results are given in Section 5, while the conclusions are summarized in Section 6.

## 2. FEATURE SELECTION AND PREPROCESSING

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

In recent years, there has been a lot of interest in multimodal biometric person verification systems, since the identity established by such systems is generally more reliable than a single-mode system. In a multimodal system, even if any one cue fails, other cues could provide the required information for establishing identity. 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 [19]. 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 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 1D processing [14] along orthogonal directions as follows. To detect the horizontal component of edginess, the face image is convolved with a 1D Gaussian filter to smooth the image horizontally. This helps in reducing the effect of noise. The Gaussian smoothing filter is given by

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

where  $\sigma$  is the standard deviation of the filter. A discrete approximation of this filter appears in Table 1, left column. The smoothed image is then convolved in the orthogonal direction (i.e., vertically) with a differential operator which is a first-order derivative of the 1D Gaussian function to find the horizontal component of edginess. The differential operator is given by

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

and its discrete approximation is given in Table 1, right column.

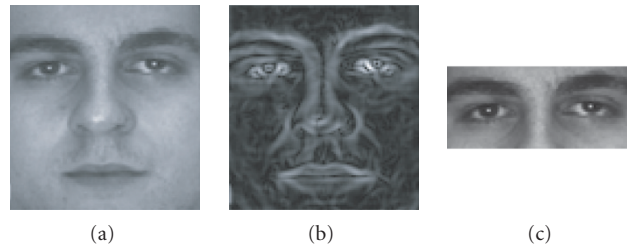


FIGURE 1: (a) A gray-scale face image, (b) its edginess image, and (c) the cropped eyes.

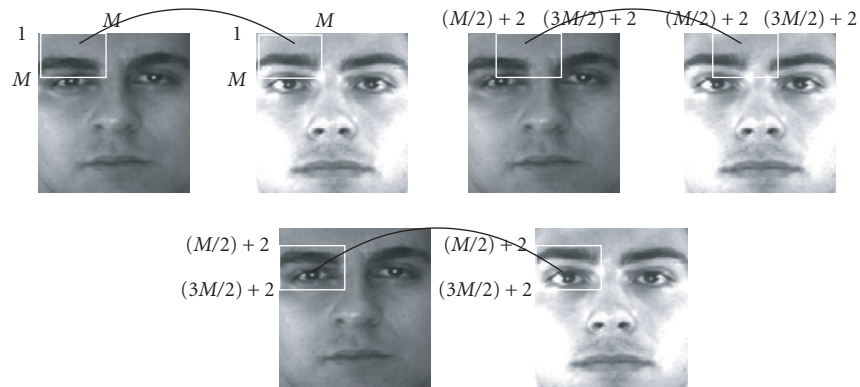


FIGURE 2: Block histogram equalization. In each image pair, the left one is the input image while the right one is the reference image.

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.

The motivation for incorporating local features into a recognition system stems from the fact that it is possible for humans to recognize a face from only parts of it. For example, if a face is occluded in its lower part, a human being may still be able to recognize it despite the occlusion. The eyes, in particular, are very important local features that help in the recognition process [8]. We conducted experiments in our lab and found that, on most occasions, people could recognize faces just from the eyes. Eyes are quite robust to facial expressions and occlusions. Each of the three facial features discussed in this section could be used either individually or jointly for recognition. We assume that eye positions are given so that these facial features can be extracted.

## 2.2. Intensity normalization

A face recognition system must recognize a face from its novel image despite variations in illumination. Unfortunately, till today, no revolutionary solution exists for the intensity normalization problem. However, approaches have been proposed to alleviate the effect of illumination variations. Adini et al. [20] have proposed different face representations such as the edge map, the image intensity derivative,

and convolution with a 2D Gabor filter for intensity normalization. The idea is to represent a face by illumination invariant features but this is difficult to actually achieve in practice. The quotient image proposed by Shashua and Riklin-Raviv [21] works under the assumption of a Lambertian model. A main drawback of this technique is that it fails under shadow effects [21]. The illumination cones technique proposed by Georgiades et al. [22] considers the effect of shadow and outperforms most existing methods. But it is computationally intensive and requires at least seven input images per person.

In this paper, 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  $B_R$ .

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

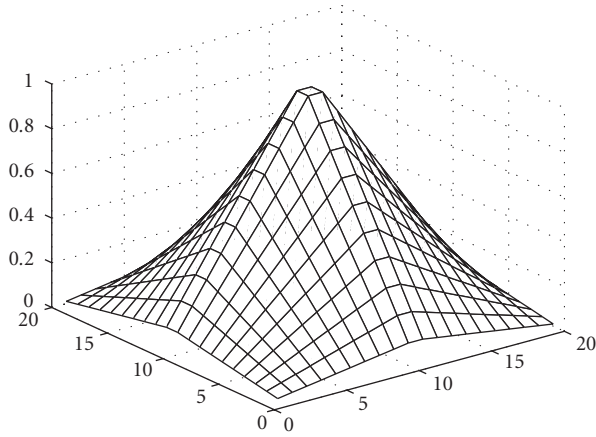


FIGURE 3: The windowing filter.

(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

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

The histogram-modified block image intensity values are scaled with a windowing filter  $H$  which is as defined below and is pictorially shown in Figure 3:

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

where

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

By simultaneously shifting the blocks in both the horizontal and the vertical directions in steps of  $M/2 + 1$  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 4, 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. We also checked the performance of our BHE method for errors in detected eye locations. This was done by artificially introducing errors in the positions of the eye locations as shown in Figure 4c. The standard deviation of the error was chosen to be 2 pixels. The intensity-normalized images in Figure 4d reveal that the BHE method is able to handle small errors in the eye positions. The proposed BHE intensity normalization technique is both simple to implement and effective.

### 3. FEATURE REPRESENTATION

Each of the facial features discussed in Section 2 can be viewed as a vector in a high-dimensional hyperspace in the pixel domain. However, such a representation is too detailed and not an optimal one. Since facial features are similar in overall configuration across individuals and are not randomly distributed in the pixel domain, they can be described by a relatively low-dimensional representation. For this purpose, we use PCA in conjunction with FLD. The idea of using eigenfeatures is motivated by a technique developed by Sirovich and Kirby [23] for efficiently representing pictures of faces. FLD provides improved discrimination over PCA. We now briefly describe feature representation using PCA and FLD. The analysis is common to all the facial features considered in this paper.

Let a training feature image  $f(x, y)$  be a two-dimensional  $N \times N$  array of intensity values. An image may also be considered as a vector of dimension  $N^2$ . Let  $f_{i,m}$  denote the  $N^2$  element training vector representing the  $m$ th image of the  $i$ th person. If in the training dataset, there are  $I$  number of people, each having  $M$  number of images, we have a total of  $K = I \cdot M$  images. The average image of the entire data set is given by

$$\psi = \frac{1}{K} \sum_{i=1}^I \sum_{m=1}^M f_{i,m}. \quad (6)$$

The covariance matrix  $C$  of the feature dataset can be written as

$$C = \frac{1}{K} \sum_{i=1}^I \sum_{m=1}^M \phi_{i,m} \phi_{i,m}^T, \quad (7)$$

where  $\phi_{i,m}$  is the mean subtracted image of  $f_{i,m}$  and is given by  $\phi_{i,m} = f_{i,m} - \psi$ . The weight vector corresponding to the  $m$ th training image  $\phi_{i,m}$  of the  $i$ th person is derived as  $w_{i,m} = E_{\text{pca}}^T \phi_{i,m}$  where  $E_{\text{pca}}$  consists of only the significant eigenvectors  $K'$  of  $C$  and is given by  $E_{\text{pca}} = [e_1, e_2, \dots, e_{K'}]$ , where  $K' \ll N^2$ . This is achieved by ignoring eigenvectors with small eigenvalues. We choose  $K'$  based on the cumulative sum of the eigenvalues such that

$$\sum_{k=1}^{K'} \frac{\lambda_k}{T_K} > R, \quad (8)$$

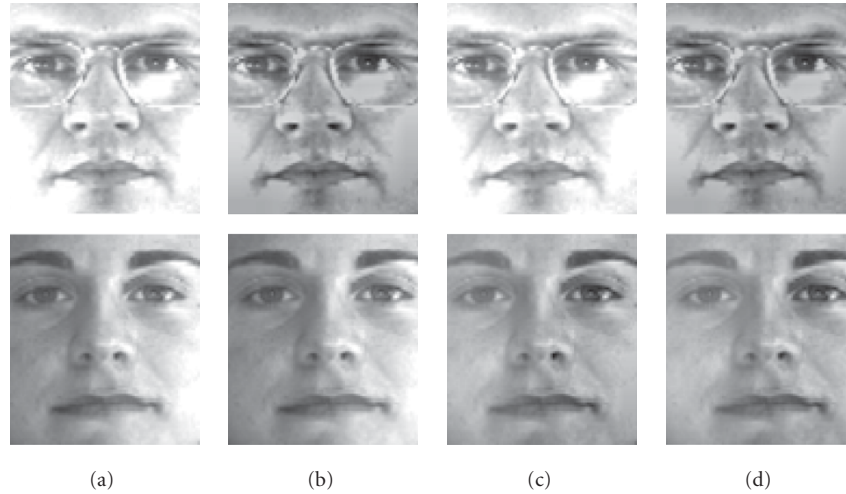


FIGURE 4: Images before and after intensity normalization with BHE. (a) Input images with exact eye locations, (b) corresponding output images after applying BHE, (c) input images with deviations from the exact eye locations, and (d) the corresponding output images.

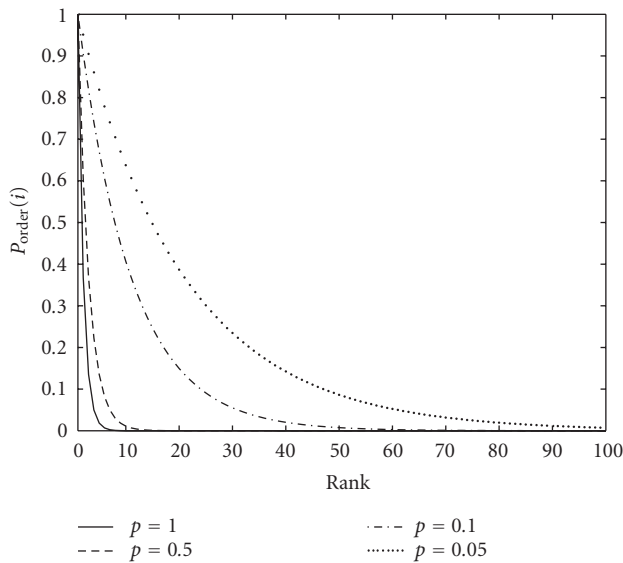


FIGURE 5: Confidence weights based on the Gamma distribution for different values of  $p$ .

where  $T_K = \sum_{l=1}^K \lambda_l$  and  $\lambda_l$  denotes the  $l$ th eigenvalue. A typical value for  $R$  is 0.99. The eigenvalues are arranged in the order of decreasing magnitude. For each facial feature, we derive the transform matrix  $E_{\text{pca}}$  corresponding to that feature from the training data.

The eigenfaces method derives only the most expressive features (MEF). The PCA inspired features do not necessarily provide for good discrimination. Fisher's linear discriminant (FLD) is an example of a class-specific method to shape the scatter for reliable classification. Mathematically, FLD selects the projection matrix  $E_{\text{fld}}$  in such a way that the ratio of the determinant of the between-class scatter matrix to the within-class scatter matrix of the projected samples is

maximized. The FLD is capable of distinguishing within and between-class scatters and yields better discriminating features for classification [24]. Since the original image space is high dimensional, most methods [4, 5] first perform dimensionality reduction using PCA and then apply FLD. Otherwise, one is confronted with the difficulty of dealing with a singular within-class scatter matrix.

For an  $I$ -class problem, the between-class scatter matrix is defined as

$$S_b = \sum_{i=1}^I (w_i - \bar{w})(w_i - \bar{w})^T, \quad (9)$$

where  $w_i$  is the average weight vector of the  $i$ th class and  $\bar{w}$  is the average weight vector of all the classes in the PCA space. The within-class scatter matrix is defined as

$$S_w = \sum_{i=1}^I \left( \frac{1}{M} \sum_{m=1}^M (w_{i,m} - w_i)(w_{i,m} - w_i)^T \right). \quad (10)$$

The projection matrix  $E_{\text{fld}}$  is chosen as

$$E_{\text{fld}} = \mathbf{E}' \frac{\arg \max |E'^T S_b E'|}{|E'^T S_w E'|} \quad (11)$$

and can be determined by solving the generalized eigenvalue problem [4]  $S_b e'_i = \lambda_i S_w e'_i$ . The projection matrix  $E_{\text{opt}}$  which is a combination of the eigen- and Fisher projections is given by  $E_{\text{opt}}^T = E_{\text{fld}}^T E_{\text{pca}}^T$ . The final feature vector corresponding to the  $i$ th person is

$$w'_i = \frac{1}{M} \sum_{m=1}^M E_{\text{opt}}^T \phi_{i,m}. \quad (12)$$

This is how the feature space is constructed for any one feature.

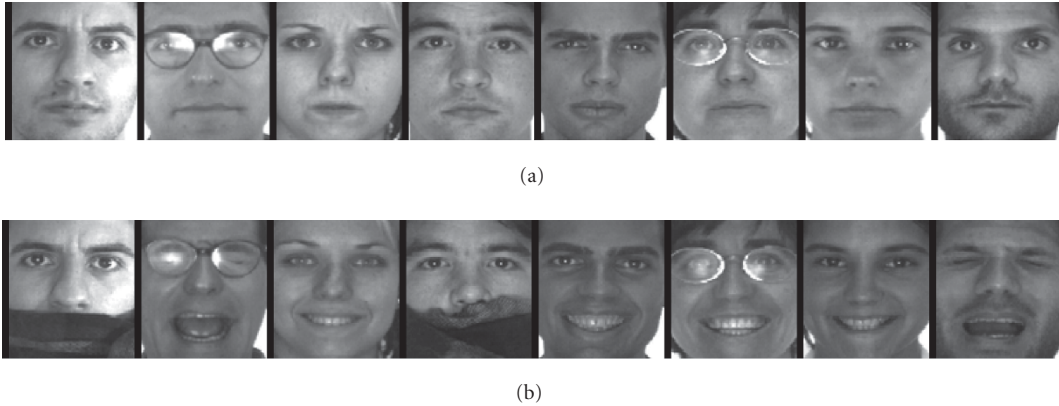


FIGURE 6: The AR face database: (a) training set and (b) probe set.

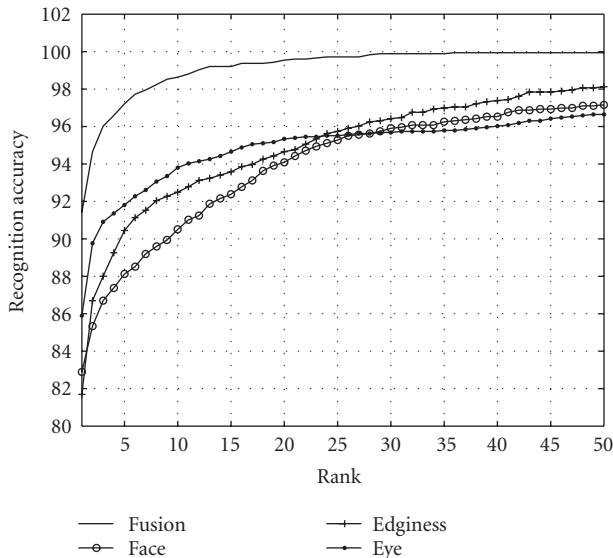


FIGURE 7: CMS curves for the AR database.

In the recognition stage, a new test image feature  $\gamma$  is transformed into its respective feature space and its feature vector is derived as  $w = E_{opt}^T(\gamma - \psi)$ . The simplest (but not the best) way for determining the class to which the test image belongs is to find the distance in feature space (DIFS) value defined as

$$\varepsilon_i = \|(w - w'_i)\|, \quad 1 \leq i \leq I. \quad (13)$$

The test image can be classified as belonging to class  $j$  if the minimum value corresponds to  $\varepsilon_j$  and is below some chosen threshold  $\theta$ .

It must be mentioned here that during the enrollment phase, the face must first be detected for facial feature extraction. Automatic and accurate face detection in cluttered backgrounds remains a challenging problem [25]. Following other works [18], we assume that the face has been located

and eye locations are provided. The three facial features are cropped with reference to these eye locations.

#### 4. RECOGNITION BY FUSION

In this section, we propose a probabilistic fusion strategy to integrate information coming from multiple facial features. Broadly speaking, the term fusion encompasses any area which deals with utilizing a combination of different sources of information, either to generate one representational format, or to reach a decision. A classifier system generally integrates multiple cues at one of the following three levels.

- (i) *Sensor-level fusion.* Raw data from the sensors is combined without any confidence associated with the labels. In this case, a simple majority rule may be employed to reach a reliable decision.
- (ii) *Feature-level fusion.* A set of possible labels ranked in each feature space with decreasing confidence value. But the confidence values themselves are not used for final fusion.
- (iii) *Decision fusion.* At the output of each feature space, labels are assigned with associated confidence values. These values are fused for decision making.

In our recognition algorithm, we use the decision fusion approach. 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 characterizing how an imposter will differ from the true person. We use the relative difference in the DIFS values instead of the absolute values, since the relative DIFS values are invariant to the mean shift of the absolute DIFS values.

In order to simplify the analysis, we consider a single feature, say face. After constructing the face space using PCA

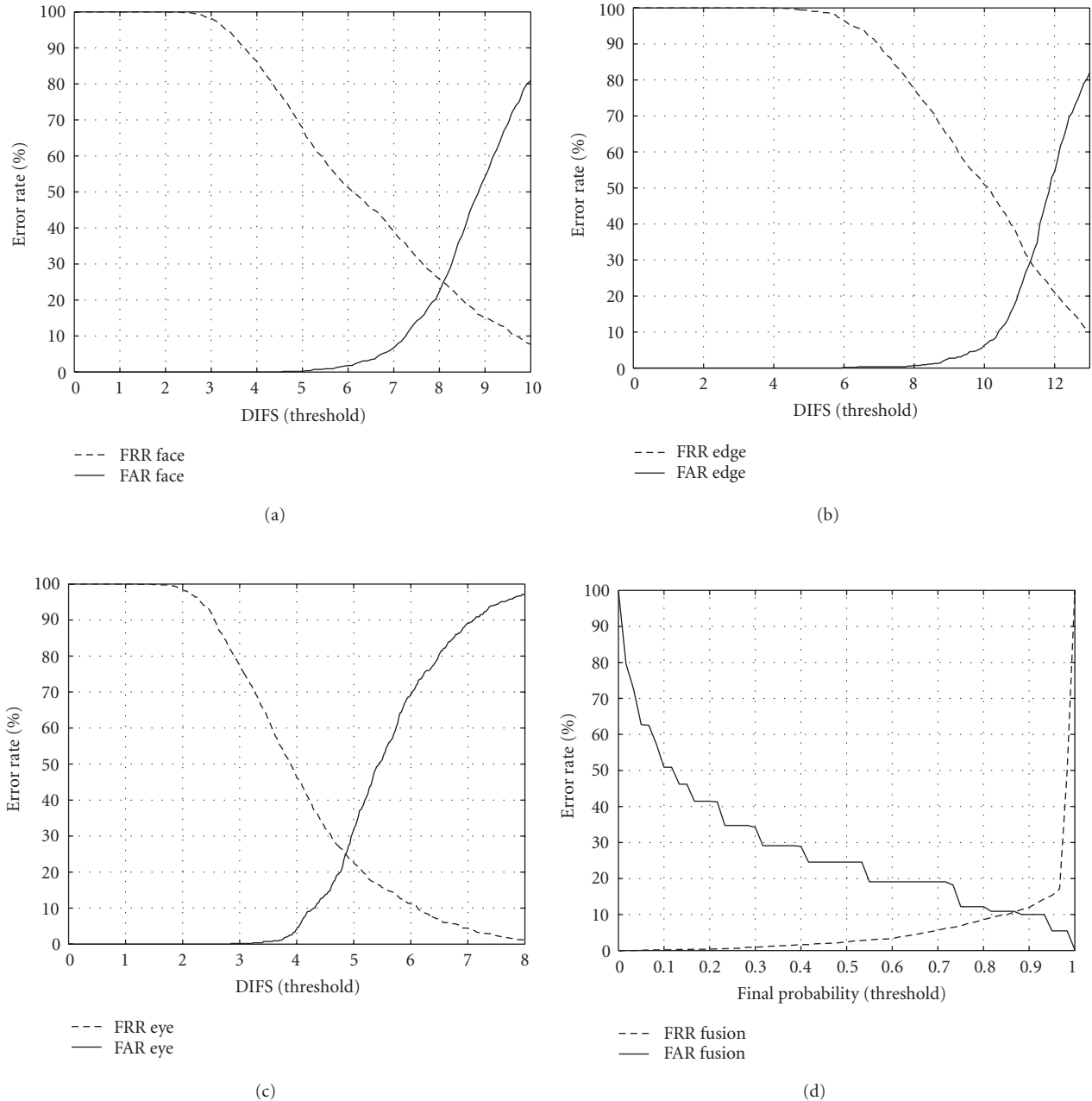


FIGURE 8: FAR and FRR plots for the AR database using (a) face only, (b) edginess only, (c) eyes only, and (d) fusion.

and FLD (as discussed in Section 3), the training data itself is used to compute the DIFS distributions as follows. Let  $\epsilon^\alpha$  (a random variable with density function  $f^\alpha(\epsilon^\alpha)$ ) represent the DIFS value of all individuals' training images with their own classes. That is,  $f^\alpha(\epsilon^\alpha)$  describes the distribution of the DIFS values for the genuine case. When we project a training image onto the PCA-FLD face space, we can derive DIFS values for that image with respect to all the individuals in the database. We arrange the DIFS values in an increasing order and repeat this exercise for all the images in the training set. Let  $\epsilon_i$  (a random variable with density function  $f_i(\epsilon_i)$ ) denote the DIFS value at rank  $i$ . Without loss of generality, it

can be assumed that  $\epsilon_i$  is statistically independent of  $\epsilon_j$  for  $i \neq j$ . Note that  $\epsilon^\alpha$  and  $\epsilon_i$  are based on all the images in the training set and are not individual specific. The relative difference in the DIFS value at rank  $i$  is defined as

$$\Delta_i = \epsilon_i - \epsilon^\alpha, \quad 1 \leq i \leq I, \quad (14)$$

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

$$f_i(\Delta_i) = \int_{-\infty}^{\infty} z_i(\epsilon_i, \Delta_i) d\epsilon_i, \quad (15)$$

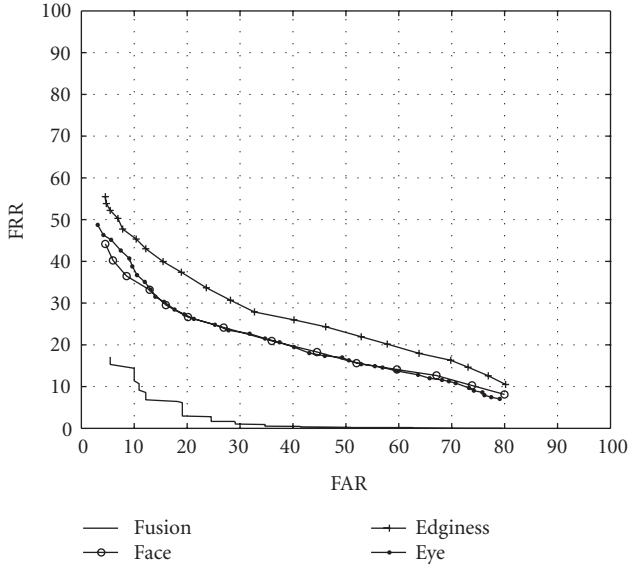


FIGURE 9: ROC plots for the AR database.

where

$$z_i(\epsilon_i, \Delta_i) = \frac{I!}{(I-i)!(i-1)!} F(\epsilon_i)^{i-1} [1-F(\epsilon^\alpha)]^{I-i} f(\epsilon_i) f(\epsilon^\alpha). \quad (16)$$

Here,  $z_i(\epsilon_i, \Delta_i)$  describes the joint distribution of the  $i$ th-rank DIFS value (i.e.,  $\epsilon_i$ ) and its distance from the true person's DIFS value (i.e.,  $\epsilon^\alpha$ ). 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 [26]. Since closed-form expressions are not available for the problem on hand, we derive  $f_i(\Delta_i)$  and  $f^\alpha(\epsilon^\alpha)$  empirically by using the DIFS values for  $\epsilon_i$  and  $\epsilon^\alpha$  from the images in the training set. Since there are  $I$  individuals each having  $M$  training images, we have  $I \cdot M$  DIFS values or sample realizations for the random variable  $\epsilon^\alpha$  and  $I \cdot M$  DIFS values for  $\epsilon_i$  corresponding to each rank  $i$ ,  $i = 1, 2, \dots, I$ . The DIFS distribution at rank 1 (i.e.,  $f_1(\Delta_1)$ ) will be almost equal to a Delta function because most of the time the top rank identity will be the actual identity (i.e.,  $\epsilon_1 \approx \epsilon^\alpha$ ). The mean value of the DIFS distributions  $f_i(\Delta_i)$  will increase with rank  $i$ .

#### 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 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  $\gamma$  arrives, its DIFS values  $\epsilon'_1, \epsilon'_2, \dots, \epsilon'_I$  are arranged in an increasing order. Let  $\epsilon^{\alpha'}$  denote the DIFS value for the top rank in the face space. The relative DIFS values  $\Delta'_i$  are computed using (14) as  $\Delta'_i = \epsilon'_i - \epsilon^{\alpha'}$ . The confidence weight assigned to the hypothesis that the image

$\gamma$  belongs to that of the identity at rank  $i$  is computed as

$$P_{\text{face}}(i) = [P_i(\Delta'_i) \cdot P_{\text{order}}(i)]_{\text{face}}, \quad 1 \leq i \leq I, \quad (17)$$

where  $P_i(\Delta'_i)$  describes how close the  $i$ th-rank DIFS value is to the top-rank and

$$P_i(\Delta'_i) = \int_{\Delta'_i}^{\infty} f_i(\Delta_i) d\Delta_i. \quad (18)$$

The term  $P_{\text{order}}(i)$  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 for the given image  $\gamma$  is very small, then it will most likely correspond to the actual identity. Hence,  $P_{\text{order}}(i)$  should fall very sharply as rank  $i$  increases. On the other hand, 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 lower ranks (i.e., higher values of  $i$ ). Depending on the top rank DIFS value  $\epsilon^{\alpha'}$ , we give relative weightage to the person at rank  $i$  using the Gamma distribution. The Gamma distribution is given by

$$P_{\text{order}}(i) = \frac{\beta}{\gamma(\lambda)} \left( \frac{i-1}{\theta} \right)^{(\lambda\beta-1)} e^{-(i-1)/\theta}, \quad \theta = \frac{1}{p}, \quad 1 \leq i \leq I. \quad (19)$$

For our problem,

$$p = \int_{\epsilon^{\alpha'}}^{\infty} f^\alpha(\epsilon^\alpha) d\epsilon^\alpha, \quad (20)$$

where  $\epsilon^{\alpha'}$  denotes the DIFS value for the top rank. We define  $P_{\text{order}}(i)$  with  $\lambda = 1$  and  $\beta = 1$ . In Figure 5, we show confidence plots for different values of  $p$ . For  $p = 1$ , the 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  $P_{\text{order}}(i)$  is to accommodate identities at even lower ranks.

In an exactly similar manner, we compute  $P_{\text{edge}}(i)$  and  $P_{\text{eye}}(i)$  in their respective PCA-FLD feature spaces. For mathematical convenience, the three features (face, edginess, and eyes) are assumed to be independent. If  $I_1, I_2, \dots, I_I$  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, that is,

$$P(I_i) = P_{\text{face}}(I_i) \cdot P_{\text{edge}}(I_i) \cdot P_{\text{eye}}(I_i). \quad (21)$$

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

$$\text{ID}(T) = \begin{cases} I_k & \text{if } P(I_k) > \tau, \\ \text{Imposter} & \text{otherwise,} \end{cases} \quad (22)$$



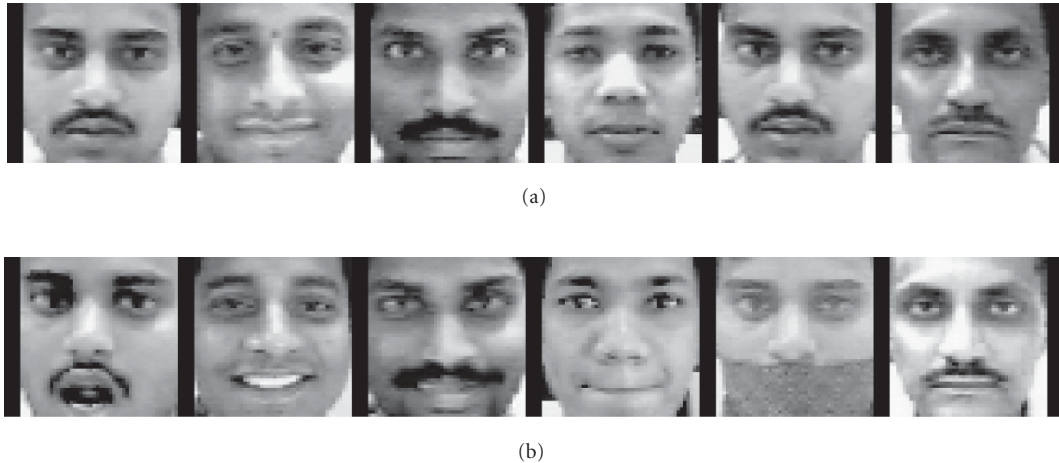


FIGURE 10: IPCV database: (a) training set and (b) probe set.

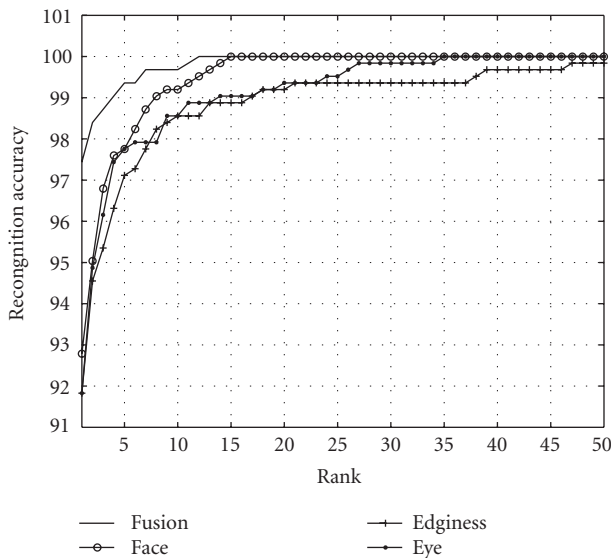


FIGURE 11: CMS curves for the IPCV database.

where

$$P(I_k) = \max \{P(I_1), P(I_2), \dots, P(I_l)\}. \quad (23)$$

The threshold  $\tau$  is chosen such that an untrained person should not be recognized at all. An individual not trained with our system should be declared as an imposter and this works as follows. For an imposter, it is unlikely that in all the features the person will come in the top rank with respect to the same identity. For example, if in the unlikely event that for any one feature, identity  $I_k$  comes in the top rank, he/she is very likely to go down in ranking in the other two features. The final confidence value  $P(I_k)$  (which is a product of the weights acquired in the three features) will be less than the threshold value  $\tau$ .

## 5. EXPERIMENTAL RESULTS

In this section, we demonstrate the performance of the proposed method on different types of face datasets. A commonly used performance measure for face recognition is the cumulative match scores (CMS), that is, the recognition accuracy in the top  $n$  ranks. We also demonstrate system rejection performance in terms of the false acceptance rate (FAR) and the false rejection rate (FRR) by testing against trained as well as untrained individuals. We validate our method on three different datasets. The first is the AR face database which is available in the public domain [17]. The second is a set of face images collected in our lab. The third is the standard FERET database [18] which has been widely used for evaluating face recognition algorithms. For all the three datasets, the required facial features were cropped with reference to the eye locations. 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.

### 5.1. The AR face database

This face database is due to Martinez and Benavente [17] at the Computer Vision Center (CVC) of the University of Alabama at Birmingham. It contains rich variations in expressions and many facially occluded images. Images are also photographed in different photo sessions so as to contain different lighting conditions and scale changes.

#### 5.1.1. Training

During training, we selected 630 images comprising 126 individuals with five images per subject. Figure 6a shows some typical training images (all images are frontal). For face and edginess, the images were cropped to a size of  $100 \times 90$  pixels with reference to the eye locations. For eyes, the image size was  $30 \times 90$  pixels. The top 150 eigenvectors were used for each feature.

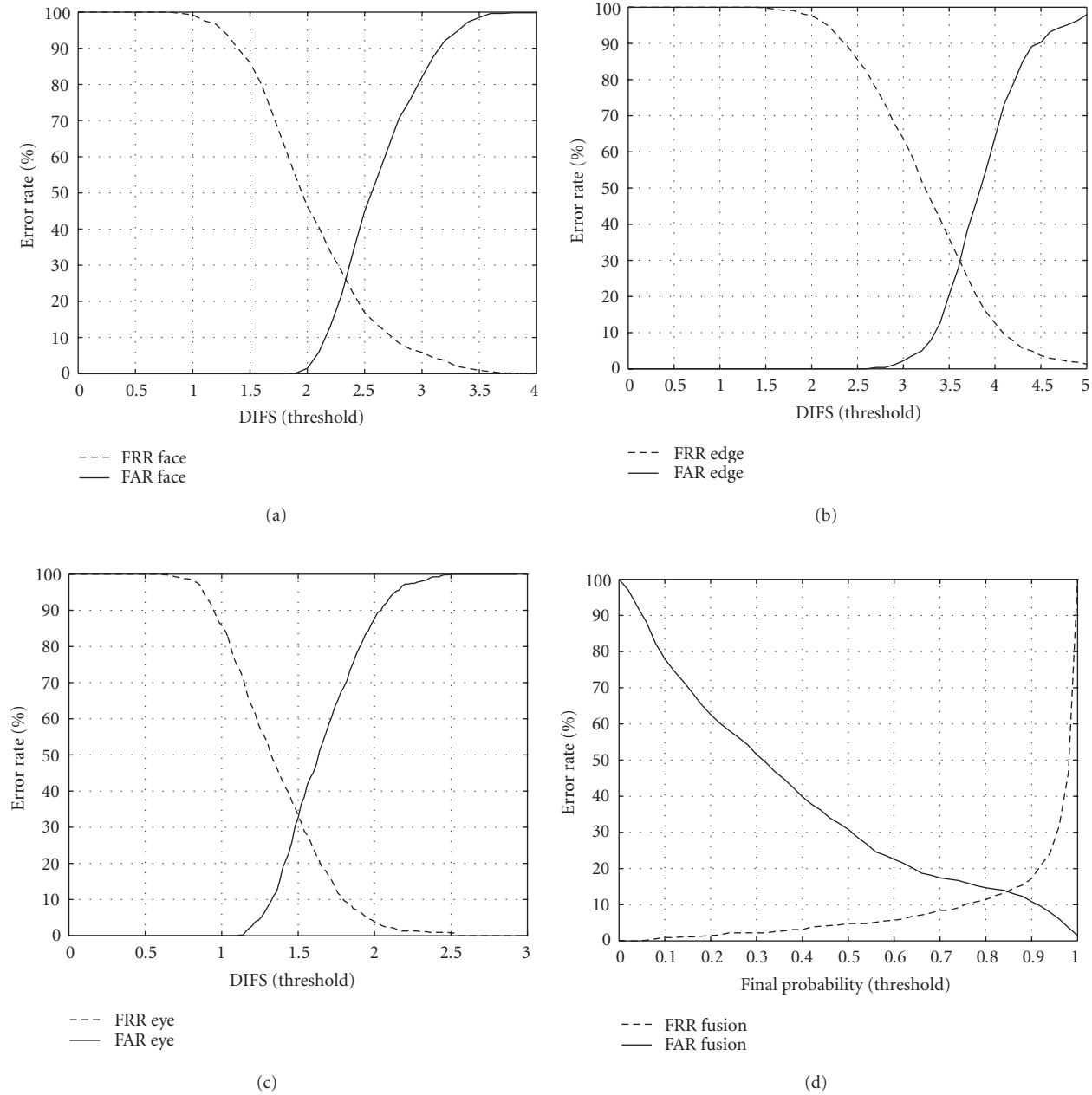


FIGURE 12: FAR and FRR plots for the IPCV database using (a) face only, (b) edginess only, (c) eyes only, and (d) fusion.

After constructing the individual feature spaces, the required DIFS distributions ( $f^\alpha(\epsilon^\alpha)$  and  $f_i(\Delta_i)$  where  $1 \leq i \leq I$ ) are computed off-line using the training dataset as follows. An image  $f_{i,m}$  labeled with identity  $I_i$  is processed to extract the three facial features. The features are transformed to their respective spaces and the DIFS values ( $\epsilon_i$ ,  $1 \leq i \leq I$ ) are arranged in an increasing order in each feature space. The DIFS value of the identity  $I_i$  contributes sample values to the random variable  $\epsilon^\alpha$ . The difference between  $\epsilon^\alpha$  and the DIFS value at rank  $i$  (i.e.,  $\epsilon_i$ ) contributes sample values to the random variable  $\Delta_i$ . By using all the labeled training images, we compute 630 sample values for all the random variables.

These values are then used to compute the DIFS distributions at each rank.

### 5.1.2. Testing

The recognition performance was tested using a probe set containing 1759 images of all the trained individuals. This subset contains different expressions and facial occlusions as shown in Figure 6b. The images in the test set are different from the training set.

In general, for a face recognition algorithm, one can ask how good the algorithm is at identifying a probe image. The question is not always “is the top match correct?” but “is

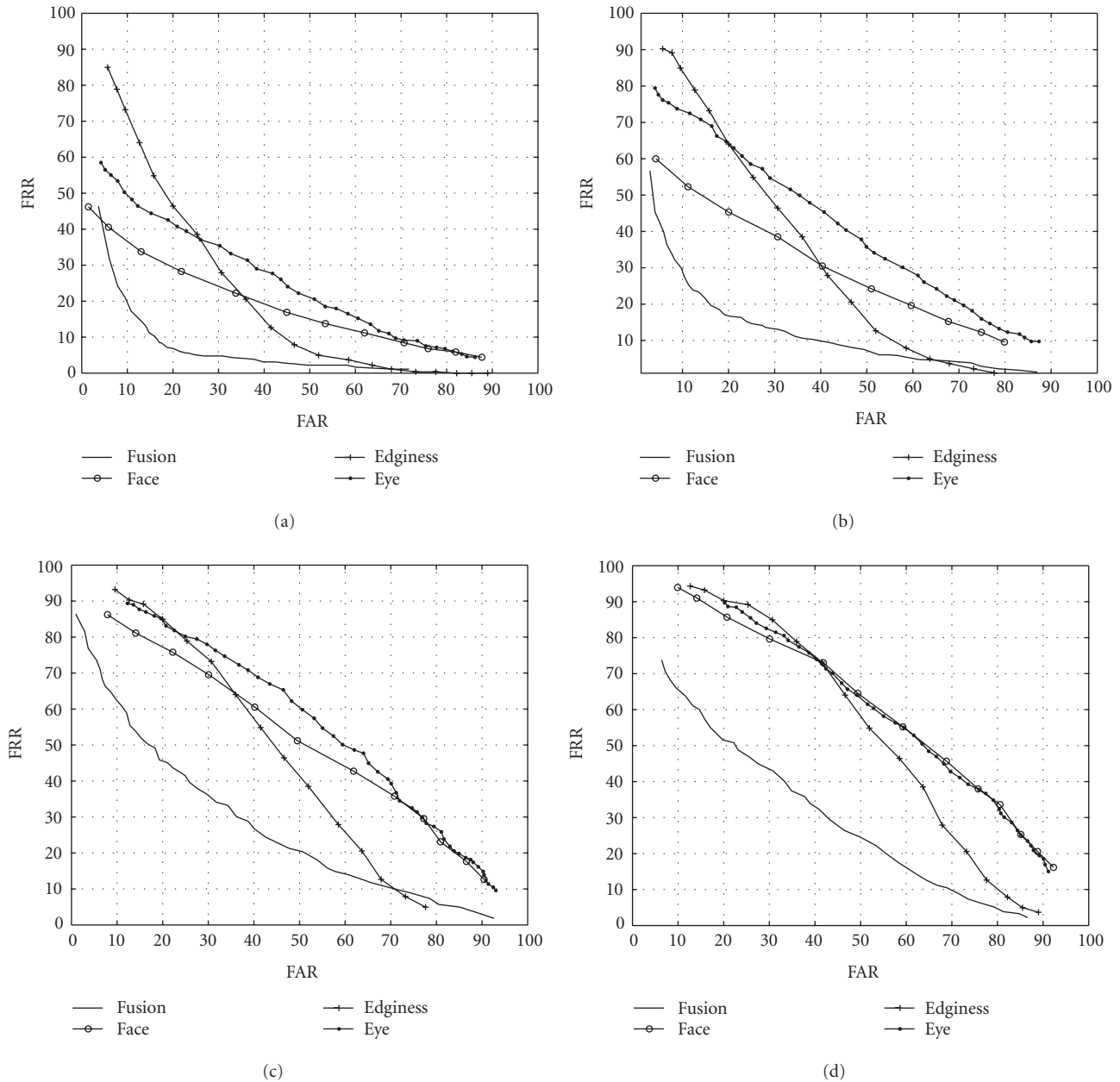


FIGURE 13: These ROC curves correspond to (a) no error, and (b), (c), (d) error with standard deviation 1, 2, and 3 pixels, respectively, in the position of the eye coordinates.

the correct answer in the top  $n$  matches?" The system performance can be reported by using cumulative match scores (CMS) plots in which the rank is plotted along the horizontal axis, and the vertical axis is the percentage of correct matches. The cumulative recognition rate was found by first ranking each probe image in relation to the gallery set. The rank indicates how well a probe face matches an intended gallery face. For a probe face with rank 1, this indicates that the probe face has the smallest Euclidean distance to an intended gallery face. Therefore, one can generally conclude that a probe face with the  $n$ th smallest Euclidean distance to the intended gallery face has a rank  $n$ . Once the rank is found

for each probe face, a cumulative rank score is calculated for each rank. This is found by summing the instances of a probe face which are within or less than a rank  $n$  and then dividing by the total number of probe faces.

The CMS plots for the top  $n$  ranks are shown in Figure 7 for the AR face database. The graphs show the performance for different features, namely, face, edginess, and eyes individually, as well as for the fusion case. For  $n = 1$ , the graphs indicate that the recognition accuracy of face, edginess, and eyes, when used individually, is 83.0%, 79.5%, and 86.0%, respectively. The recognition rate after fusion increases to 92.0% for the top rank. Thus, we observe that there

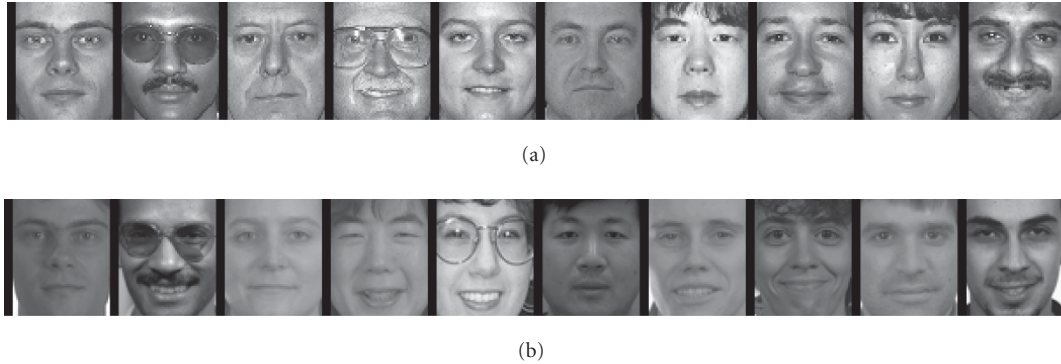


FIGURE 14: The FERET database: (a) training images, and (b) probe images.

is an improvement of almost 6% with fusion as compared to the best accuracy achieved with a single feature. For  $n = 10$ , the accuracy using fusion is almost 99%.

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 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.,  $\text{FAR} = 0$ ), but it will be impossible to accept even a true (known) person who is in the training data (i.e.,  $\text{FRR} = 100\%$ ). 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). For a good recognition system, the ERR value should be as small as possible.

For the AR face 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 Figures 8a, 8b, 8c, and 8d, respectively. While computing FAR plots, the system was trained with 90 persons out of 126 members in the AR face database. A total of 380 probe images of the remaining untrained people were used as probe images. Initially, we fix 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 the FRR values, the system is again trained with the complete dataset (i.e., with 126 persons). The probe set of 1209 images comprising all 126 persons are used as known identities and we computed the error rates for all the threshold values. From Figure 8, we observe that the ERR values for face, edginess, and eyes are

quite high (25%, 29.5%, and 25%, respectively). In contrast, the ERR value for the fusion method is much lower and is only 10%. Thus, the fusion method is very useful in rejecting imposters. The ERR is just one operating point and not necessarily the one that could be used in practice. For real applications, one would choose  $\tau$  in (22) such that the FAR is very low even if this is at the expense of a slightly higher FRR. The receiver operating characteristics (ROC) for our system has been shown in Figures 9 and the value of  $\tau$  can be decided from it. Based on the ROC curves, we observe that fusion clearly outperforms the other three features.

## 5.2. The IPCV database

We tested the proposed method on yet another dataset collected in our own lab. We call it the IPCV dataset. This database has variations in position, scale, facial expression, occlusion, and illumination. All images were intensity normalized using the BHE technique to compensate for illumination variations. The database has 2000 facial images of 105 individuals. Of these, 1365 images were selected for training with 13 images per person, while the remaining 635 facial images were used as probe. Some example training and probe set images are shown in Figures 10a and 10b, respectively. Face images were cropped to a size of  $50 \times 40$  pixels for face as well as edginess images. The eye images were of size  $25 \times 40$  pixels. The top 300 eigenvectors are used for face and edginess features, while the top 200 eigenvectors are used for the eye feature. The DIFS distributions are computed from the training data. The CMS curves for the top  $n$  ranks are shown in Figure 11. For the top rank, the graph indicates that the recognition accuracy for face, edginess, and eyes, taken individually, is 92.8%, 91.8%, and 92%, respectively. The recognition rate after fusion increases to 97.5% for the top rank.

For computing the FAR plots, the system was trained with 90 persons out of the 105 members in the dataset. A total of 500 probe images of the remaining untrained people were used. Figures 12a, 12b, 12c, and 12d show the FAR and FRR plots for face, edginess, eyes, and fusion, respectively. From the figure, we observe that the ERR values for face, edginess, and eye are quite high (26%, 30%, and 33%, respectively). In contrast, the ERR value for the fusion method

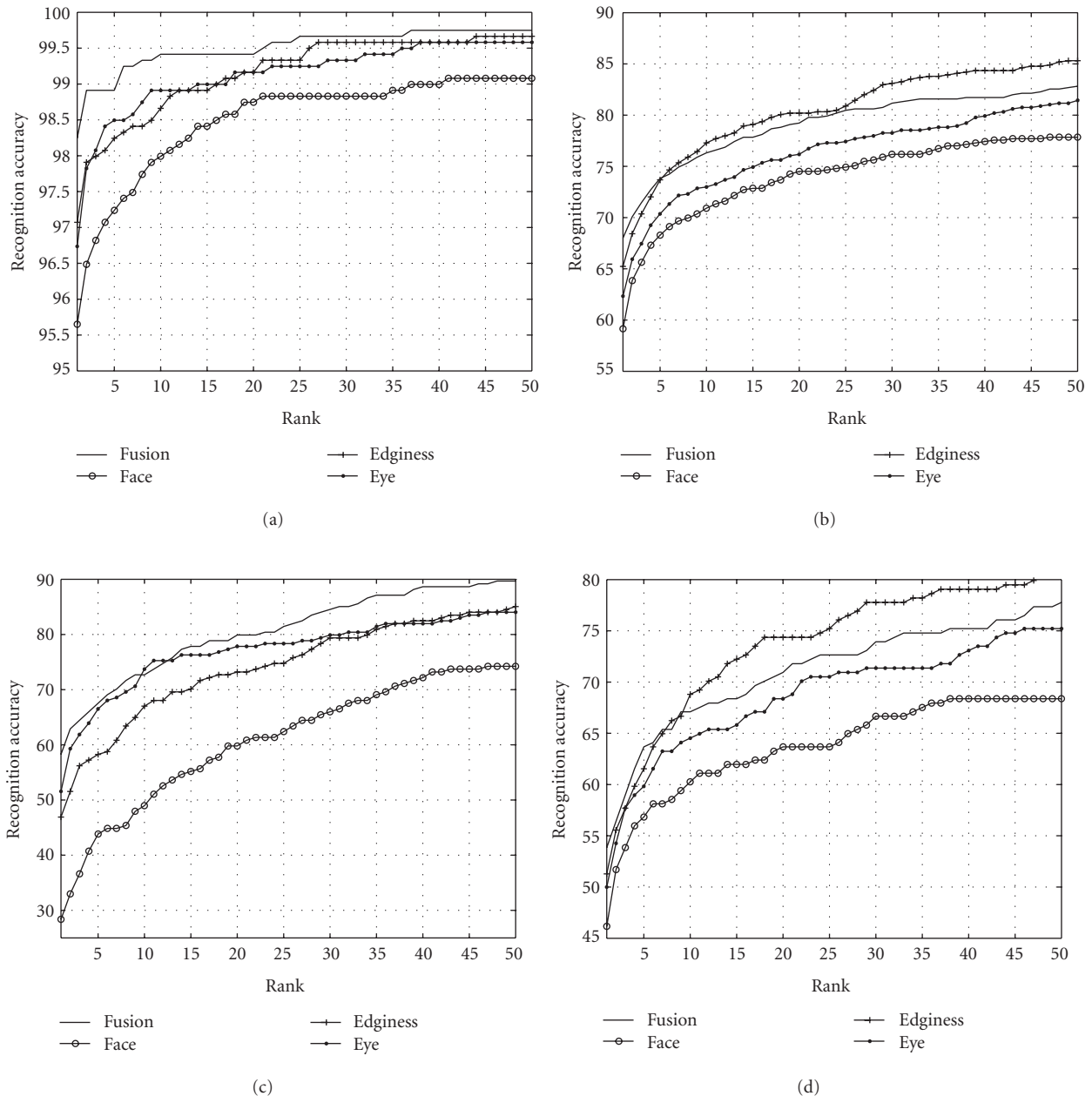


FIGURE 15: CMS plots for (a) the FB probe set, (b) duplicate I, (c) the FC probe set, and (d) duplicate II.

is only 12%. In Figure 13a, the ROC curves for the IPCV database have been shown for fusion, and the three facial features when used individually.

For face recognition methods to work, the position of the eyes should be known very accurately [18]. To study the effect of errors in the eye position on recognition accuracy, we assumed exact eye positions for the training images but introduced Gaussian distributed errors in the position of the eye coordinates for the probe images. The standard deviation of the error was chosen to be 1, 2, and 3 pixels and the ROC curves for each of these cases are shown in Figures 13b, 13c, and 13d.

### 5.3. The FERET database

The proposed algorithm was finally tested on the standard FERET database [18]. 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 [18]. The FERET evaluation in [18] provides a comprehensive picture of the state of the art in face recognition. Figure 14 depicts some example training and probe images from the FERET database. In Table 2, details of all the four probe categories are given. The

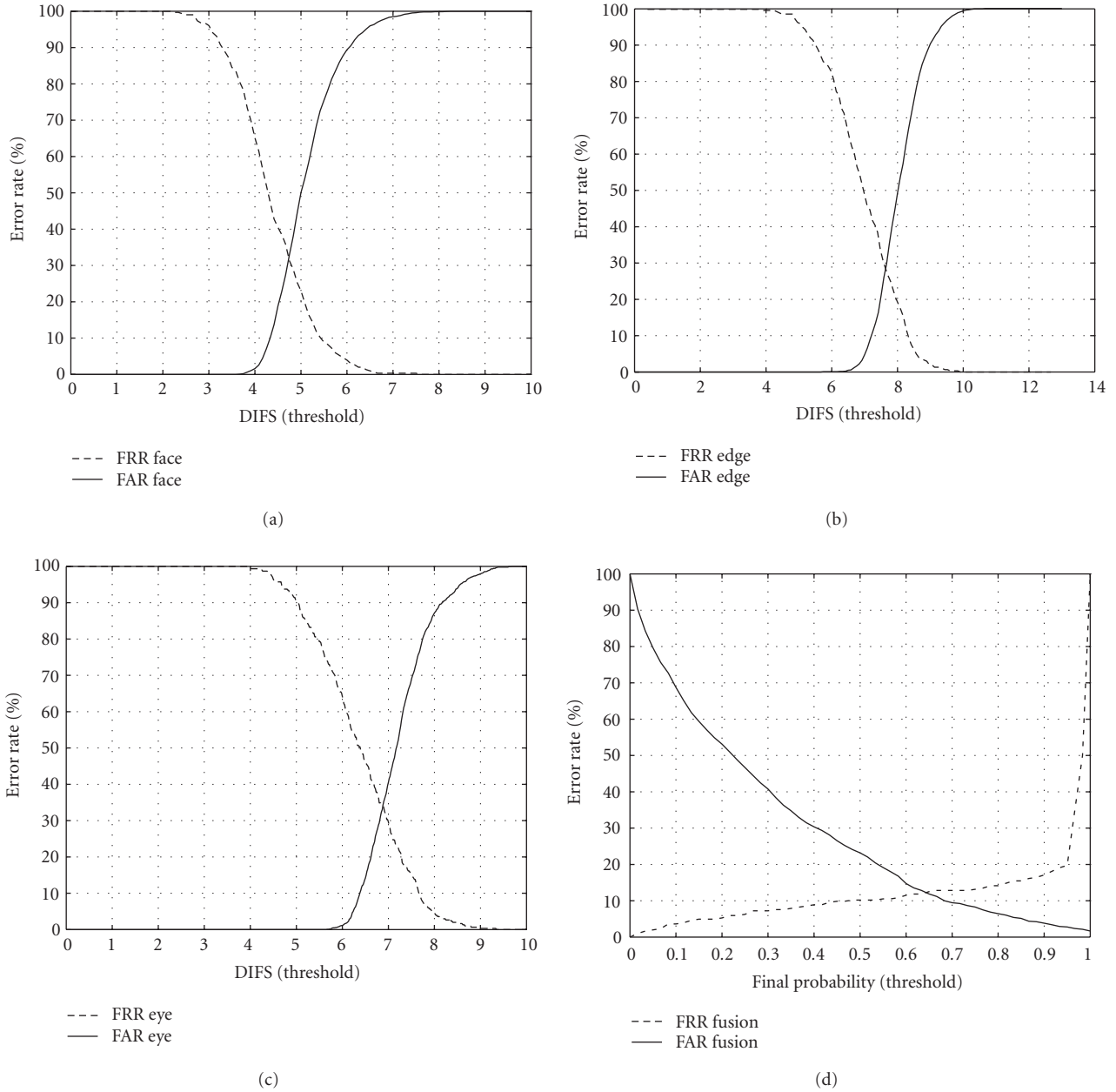


FIGURE 16: FAR and FRR plots for the FERET database using (a) face only, (b) edginess only, (c) eyes only, and (d) fusion.

FA images (regular frontal faces of persons) were used as the gallery set, whereas four categories of probe sets were used to compare against the gallery set. The first probe category was the FB 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 FC 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 probe.

TABLE 2: Gallery and probe information for FERET.

Probe category	Gallery size	Probe set size
FB	1196	1195
Duplicate I	1196	722
FC	1196	194
Duplicate II	1196	234

When tested on all the probe categories, the CMS plots for the top  $n$  ranks are shown in Figure 15. In Table 3, we have compared the performance of the proposed method with the

TABLE 3: Recognition accuracy for different algorithms with the FERET database.

Serial number	Probe set	Recognition accuracy at rank 1					
		Fusion	UMD 97	USC	MIT 96	Baseline corr.	Baseline EF
1	FB	98.3	96.5	95	94.8	82.5	79.5
2	Duplicate I	68	46	58	57	35	42
3	FC	59	59	82	32	7	18
4	Duplicate II	54	21	46	34	16	22

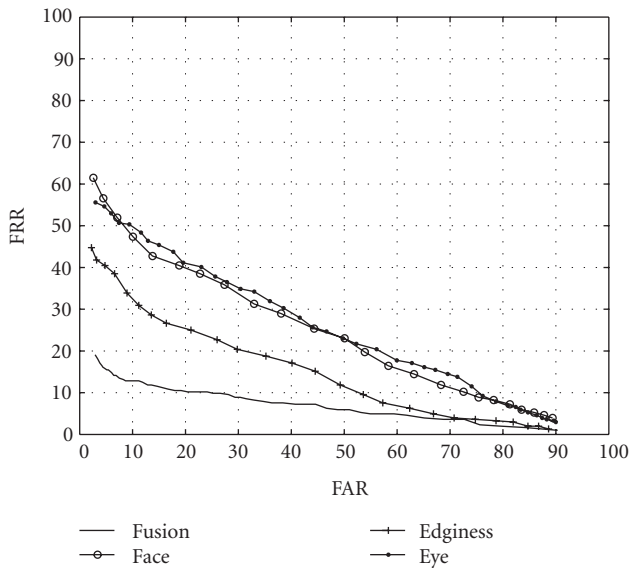


FIGURE 17: ROC curves for the FERET database.

partially automatic face recognition algorithms that appear in [18]. (The recognition accuracy for these algorithms has been deciphered from the plots in [18]. Hence, there may be small variations in the actual values.) From Table 3, we observe that the performance of the proposed method is comparable to the best reported results. For the FB, duplicate I and duplicate II probe sets, our method has better accuracy compared to others. On the FC probe set, we come second.

The FAR and FRR plots were also computed for the FERET database. Figures 16a, 16b, 16c, and 16d show the FAR and FRR plots for face, edginess, eyes, and fusion, respectively. The system was trained with 482 people out of the 1 199 people 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 from the remaining 717 untrained individuals. While computing FRR, 304 facial images were used from 482 trained individuals. From Figure 16, we observe that the ERR values for face, edginess, and eye are quite high (31%, 30%, and 34%, resp.). The ERR value after fusion reduces to 12%. Figure 17 shows the ROC plots corresponding to fusion and the individual facial features for the FERET database.

## 6. CONCLUSIONS

We have described a system that uses different facial features for robust recognition. A probabilistic fusion scheme has been proposed that 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 three different face datasets, including the FERET database. It has been shown that fusion of different facial features improves overall recognition accuracy. The improvement is particularly significant under facial occlusions, variations in facial expressions, and illumination changes. Fusion is also very useful in rejecting imposters.

## ACKNOWLEDGMENTS

The authors are thankful to the reviewers for their useful comments which helped in improving the presentation of the paper considerably. Funding from MHRD, India, is gratefully acknowledged.

## REFERENCES

- [1] C. J. Wu and J. S. Huang, "Human face profile recognition by computer," *Pattern Recognition*, vol. 23, no. 3-4, pp. 255-259, 1990.
- [2] K.-M. Lam and H. Yan, "An analytic-to-holistic approach for face recognition based on a single frontal view," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, no. 7, pp. 673-686, 1998.
- [3] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71-86, 1991.
- [4] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 7, pp. 711-720, 1997.
- [5] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol. 7, no. 1, pp. 179-188, 1936.
- [6] C. Kotropoulos, A. Tefas, and I. Pitas, "Frontal face authentication using morphological elastic graph matching," *IEEE Trans. Image Processing*, vol. 9, no. 4, pp. 555-560, 2000.
- [7] B. Duc, S. Fischer, and J. Bigun, "Face authentication with Gabor information on deformable graphs," *IEEE Trans. Image Processing*, vol. 8, no. 4, pp. 504-516, 1999.
- [8] R. Brunelli and T. Poggio, "Face recognition: features versus templates," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, no. 10, pp. 1042-1052, 1993.
- [9] L. Hong and A. Jain, "Integrating faces and fingerprints for personal identification," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, no. 12, pp. 1295-1307, 1998.

- [10] S. Ben-Yacoub, Y. Abdeljaoued, and E. Mayoraz, "Fusion of face and speech data for person identity verification," *IEEE Trans. Neural Networks*, vol. 10, no. 5, pp. 1065–1074, 1999.
- [11] G. Shakhnarovich and T. Darrell, "On probabilistic combination of face and gait cues for identification," in *Proc. 5th IEEE International Conference on Automatic Face and Gesture Recognition (FGR '02)*, pp. 169–174, Washington, DC, USA, May 2002.
- [12] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, no. 3, pp. 226–239, 1998.
- [13] R. Brunelli and D. Falavigna, "Person identification using multiple cues," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, no. 10, pp. 955–966, 1995.
- [14] B. S. Venkatesh, S. Palanivel, and B. Yegnanarayana, "Face detection and recognition in an image sequence using eigenedginess," in *Proc. Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP '02)*, pp. 97–101, Ahmedabad, India, December 2002.
- [15] K. S. Rao and A. N. Rajagopalan, "A probabilistic fusion technique for face recognition," in *Proc. 12th European Signal Processing Conference (EUSIPCO '04)*, Vienna, Austria, September 2004.
- [16] P. Watanapongse and H. H. Szu, "Application of the principal wavelet component in pattern classification," in *Wavelet Applications V*, vol. 3391 of *Proceedings of SPIE*, pp. 194–205, Orlando, Fla, USA, April 1998.
- [17] A. M. Martinez and R. Benavente, "The AR face database," Tech. Rep. #24, Computer Vision Center (CVC), Bellaterra, Barcelona, Spain, 1998.
- [18] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 10, pp. 1090–1104, 2000.
- [19] P. S. Penev and J. J. Atick, "Local feature analysis: a general statistical theory for object representation," *Network: Computation in Neural Systems*, vol. 7, no. 3, pp. 477–500, 1996.
- [20] Y. Adini, Y. Moses, and S. Ullman, "Face recognition: the problem of compensating for changes in illumination direction," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 7, pp. 721–732, 1997.
- [21] A. Shashua and T. Riklin-Raviv, "The quotient image: class-based re-rendering and recognition with varying illuminations," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, no. 2, pp. 129–139, 2001.
- [22] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, no. 6, pp. 643–660, 2001.
- [23] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," *Journal of the Optical Society of America A*, vol. 4, no. 3, pp. 519–524, 1987.
- [24] C. Liu and H. Wechsler, "A shape- and texture-based enhanced Fisher classifier for face recognition," *IEEE Trans. Image Processing*, vol. 10, no. 4, pp. 598–608, 2001.
- [25] A. N. Rajagopalan, K. S. Kumar, J. Karlekar, et al., "Locating human faces in a cluttered scene," *Graphical Models and Image Processing*, vol. 62, no. 5, pp. 323–342, 2000.
- [26] E. J. Gumbel, *Statistics of Extremes*, Columbia University Press, New York, NY, USA, 1958.

**K. Srinivasa Rao** received the B. Tech. degree in electronics and communication engineering from Nagarjuna University, India, in 2002 and the M.S. degree from the Indian Institute of Technology (IIT), Madras, in 2005. In his M.S. research work, he developed a probabilistic fusion technique for face recognition under the supervision of Dr. A. N. Rajagopalan. Presently he is working in Benchmark Vision Systems Pvt. Ltd., Chennai, India. His research interests include automatic face recognition, object recognition, and image restoration.



**A. N. Rajagopalan** received his B. Tech. and M. Tech. degree from Nagpur University. He obtained the Ph.D. degree in electrical engineering from the Indian Institute of Technology, Bombay, in 1998. During the summer of 1998, he was a Visiting Scientist at the Image Communication Laboratory, University of Erlangen, Erlangen, Germany. Between October 1998 and September 2000, he worked as an Assistant Research Scientist at the Center for Automation Research, University of Maryland, College Park. Since October 2000, he has been serving as an Assistant Professor in the Department of Electrical Engineering at IIT Madras. He was a Visiting Scientist in summer 2002 and a Visiting Assistant Professor in summer 2005 at the Center for Automation Research, University of Maryland. His research interests include depth recovery from defocused images, image restoration, particle filters, image super resolution, face detection and recognition, and higher-order statistical learning. He is a coauthor of the book *Depth From Defocus: A Real Aperture Imaging Approach* published by Springer, New York, in 1999.

