ORIGINAL PAPER

Significance of image representation for face verification

**Anil Kumar Sao · B. Yegnanarayana ·
B. V. K. Vijaya Kumar**

Abstract   In this paper we discuss the significance of representation of images for face verification. We consider three different representations, namely, edge gradient, edge orientation and potential field derived from the edge gradient. These representations are examined in the context of face verification using a specific type of correlation filter, called the minimum average correlation energy (MACE) filter. The different representations are derived using one-dimensional (1-D) processing of image. The 1-D processing provides multiple partial evidences for a given face image, one evidence for each direction of the 1-D processing. Separate MACE filters are used for deriving each partial evidence. We propose a method to combine the partial evidences obtained for each representation using an auto-associative neural network (AANN) model, to arrive at a decision for face verification. Results show that the performance of the system using potential field representation is better than that using the edge gradient representation or the edge orientation representation. Also, the potential field representation derived from the edge gradient is observed to be less sensitive to variation in illumination compared to the gray level representation of images.

A. K. Sao (✉)
Indian Institute of Technology Madras,
Chennai 600036, Tamil Nadu, India
e-mail: anil@cs.iitm.ernet.in

B. Yegnanarayana
International Institute of Information Technology,
Hyderabad 500032, Andhra Pradesh, India
e-mail: yegna@iiit.ac.in

B. V. K. Vijaya Kumar
Carnegie Mellon University, Pittsburgh, PA 15213, USA
e-mail: kumar@ece.cmu.edu

1 Introduction

The objective of an automatic face verification system is to accept or reject the claimed identity of a person using the image of his/her face. This is in contrast to a face identification system, where the system has to determine the probable identity of a given face image [1]. A survey of face recognition techniques can be found in [1,2]. In all the techniques, representation of the face image is crucial. On the basis of representation, face recognition studies can be broadly categorized into two types: Holistic analysis and feature-based analysis. In the holistic analysis, composite features from the entire face image are extracted and used for face recognition. Some of the holistic analysis methods reported in the literature are based on eigen analysis [3], template matching [4,5] and correlation filter [6]. In the feature-based methods, features extracted from the local regions of a face image are used. Some of the features include edge map [7] and Gabor wavelets [8]. Feature-based methods use graph matching [9] technique. It has been argued that both holistic as well as feature-based approaches might be required for face verification [10]. In this work we use holistic analysis.

We explore different representations of a face image for face verification task. For studying the significance of representation, one can used any standard face matching method. In this study the correlation-based minimum average correlation energy filter (MACE) [11] is used for matching two face images. Earlier work with MACE filters employed gray level representation of the face image [6]. The gray level

representation has the drawback that it is sensitive to illumination [7]. This issue was addressed by using face images under different lighting conditions as training examples for designing the MACE filter [12]. Illumination issue can also be addressed by choosing illumination-invariant representation of face images. We consider three representations of a face image: Edge gradient, edge orientation and potential field derived from the edge gradient. The edge gradient and orientation are more robust to changes in illumination, compared to the gray scale representation [7]. But it is difficult to correlate two images using the edge gradient or edge orientation representation. This problem is addressed to some extent by deriving a potential field from the edge gradient.

In this paper the edge-based representations are derived using one-dimensional (1-D) image processing. The 1-D processing provides multiple partial evidences for a given image, one for each direction of processing [13]. Separate MACE filter is derived for each partial evidence. The outputs of all the MACE filters are combined using an auto-associative neural network (AANN) model. The output of the AANN model is used to arrive at a decision for the claimed identity in the verification task. Performance of the proposed representations are compared on Facial Expression Database, collected at Advanced Multimedia Processing (AMP) Laboratory at the Electrical and Computer Engineering Department of Carnegie Mellon University [14]. The database consists of 13 subjects, each having 75 images with varying facial expressions. The size of each face image is scaled down to 30×30 pixels in all the experiments in this paper.

The organization of the paper is as follows: Sect. 2 provides a brief description of the MACE filter. Section 3 deals with the representation of the face image using edge gradient. Along with the edge gradient, edge orientation can also be used to characterize the face image, as explained in Sect. 4. The edge-based representations of image is not suitable for matching two images. Hence Potential field is derived from the edge gradient, as discussed in Sect. 5. Experimental results of face verification using different representations are given in Sect. 6, and Sect. 7 summarizes the work.

2 Minimum average correlation energy (MACE) filter

In this section we give a brief review of the MACE filter, which is used for matching face images in this study. The basic correlation filter is the synthetic discriminant function (SDF) introduced by Hester and Casasent [15]. In the basic SDF filter, the overall approach is to build a filter $h(x, y)$ such that its cross-correlation with an authentic training image $s(x, y)$ has a value 1 at the origin, and a value 0 for the impostor training images. Here, the origin is the center of the correlation output. Such a filter is called an equal correlation peak SDF (ECPSDF) filter. Let $\odot$ denote the two

dimensional (2-D) cross-correlation operator, $c(\tau_x, \tau_y)$ the cross-correlation output, and $f(x, y)$ the input to the filter. Then

$$
\begin{aligned}
c(\tau_x, &\tau_y) \\
&= h(x, y) \odot f(x, y) \\
&= \int \int h^*(x, y) f(x + \tau_x, y + \tau_y) \mathrm{d}x \mathrm{d}y \\
&= \int \int H^*(u, v) F(u, v) \exp[j2\pi(u\tau_x + v\tau_y)] \mathrm{d}u \mathrm{d}v,
\end{aligned}
\tag{1}
$$

where $H(u, v)$ and $F(u, v)$ are the 2-D Fourier transforms of $h(x, y)$ and $f(x, y)$, respectively, and the operator '$*$' denotes the complex conjugate. In the case of ECPSDF filter, the constraint on the peak at the origin of the correlation output is as follows:

$$
\begin{aligned}
c(\tau_x, \tau_y)|_{\tau_x, \tau_y=0} &= h(x, y) \odot s_i(x, y)|_{\tau_x, \tau_y=0} \\
&= \int \int h^*(x, y) s_i(x, y) \mathrm{d}x \mathrm{d}y \\
&= \int \int H^*(u, v) S_i(u, v) \mathrm{d}u \mathrm{d}v \\
&= c_i,
\end{aligned}
\tag{2}
$$

where c_i is 1 for authentic images, and 0 for impostor training images. Here $s_i(x, y)$ represents the ith image in the training set, and $S_i(u, v)$ represents its 2-D Fourier transform. For multiple exposures, the $h(x, y)$ is assumed to be a linear combination of the N training images [15]. That is

$$
h(x, y) = a_1 s_1(x, y) + \cdots + a_N s_N(x, y),
\tag{3}
$$

where the coefficients $a_1, a_2, \ldots, a_N$ are determined to satisfy the condition given in (2). Using (3) we can rewrite (2) as

$$
\sum_{j=1}^{N} a_j R_{ji} = c_i, \quad i = 1, 2, \ldots, N,
\tag{4}
$$

where

$$
R_{ji} = \int \int s_i^*(x, y) s_j(x, y) \mathrm{d}x \mathrm{d}y
\tag{5}
$$

is the inner product of the training images $s_i(x, y)$ and $s_j(x, y)$.

The ECPSDF has the following disadvantages [16]:

(a) The filter does not take into account any noise that might be present in the images.
(b) The filter is designed to control only one point in the correlation output.

The problem of noise is addressed using the Minimum Variance SDF (MVSDF) [17], and the second problem is addressed using the MACE filter. We focus here on the design of

MACE filter. The constraint in the frequency domain given in (2) is reproduced here as

$$\int\int H^*(u,v)S_i(u,v)\mathrm{d}u\mathrm{d}v = c_i. \tag{6}$$

In addition to this constraint, the MACE filter minimizes the average correlation energy (E) of the correlation outputs due to training N images [11], and it is given by

$$E = \frac{1}{N}\sum_{i=1}^{N}\int\int |c_i(\tau_x,\tau_y)|^2 \mathrm{d}\tau_x\mathrm{d}\tau_y$$

$$= \frac{1}{N}\sum_{i=1}^{N}\int\int |S_i(u,v)|^2|H(u,v)|^2\mathrm{d}u\mathrm{d}v. \tag{7}$$

By minimizing E with respect to $\{a_i\}$, the filter design attempts to keep the sidelobes in the correlation output as small as possible. Let $\hat{\mathbf{h}}$ and $\hat{\mathbf{s}}_i$ denote the column vectors of length d formed by sampling $H(u,v)$ and $S_i(u,v)$, respectively. The vectors are arranged from left to right and top to bottom. The quantity d is the number of pixels in the image. Let $\hat{\mathbf{S}}$ be a $d \times N$ matrix whose columns are $\hat{\mathbf{s}}_i$. We can rewrite (6) as

$$\hat{\mathbf{S}}^{+}\hat{\mathbf{h}} = \mathbf{c}, \tag{8}$$

where the superscript '+' represents the complex conjugate transpose of a matrix, and $\mathbf{c} = [c_1, c_2, \ldots, c_N]^T$. If $\mathbf{Y}_i$ is the diagonal matrix containing the elements of $\hat{\mathbf{s}}_i$, then (7) can be written as [11]

$$E = \hat{\mathbf{h}}^{+}\hat{\mathbf{D}}\hat{\mathbf{h}}, \tag{9}$$

where

$$\hat{\mathbf{D}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{Y}_i\mathbf{Y}_i^* $$

is the diagonal matrix containing the average power spectrum of the training images. Minimizing E subject to the condition given (8) yields [11]

$$\hat{\mathbf{h}} = \hat{\mathbf{D}}^{-1}\hat{\mathbf{S}}(\hat{\mathbf{S}}^{+}\hat{\mathbf{D}}^{-1}\hat{\mathbf{S}})^{-1}\mathbf{c}. \tag{10}$$

The filter $\hat{\mathbf{h}}$ given by (10) is used for matching reference with a test image.

3 Image matching using edge gradient representation

One of the problems with using gray scale values of images directly in the correlation filter is that changes in illumination can significantly affect the correlation matching. Changes in illumination are dealt with either by modeling the effect of illumination of faces, or by extracting the features that are less sensitive to illumination. Modeling of illumination requires 3-D model of the face, which in turn requires many samples of

each face. On the other hand, extracting features that are less sensitive to illumination requires appropriate representation of the face image. One of the representations suggested in the literature is the edge map [7,18]. Edge maps are obtained by computing the intensity gradient, and then thresholding the gradient. Selection of the threshold value becomes a major issue in the edge map representation. If the threshold value is low, then spurious edges may show up in the edge map. On the other hand, a high threshold value may remove some important edges.

We propose a continuous edge gradient representation of a face image. This representation has the advantage that it does not require thresholding. The edge gradient computation requires two operations: A smoothing operator to reduce the effect of noise, and a spatial differential operator to compute the gradient [19]. We use a Gaussian function for smoothing, and the derivative of the Gaussian function as differential operator. A problem encountered when smoothing the image before computing the gradient is that the genuine edges may also get smoothed out. The smearing of genuine edges can be reduced by using one-dimensional (1-D) processing of the image [13].

3.1 Extraction of partial edge information

In 1-D processing of a given image, the smoothing operator is applied along one direction, and the derivative operator is applied along the orthogonal direction. By repeating this procedure of smoothing followed by derivative operation along the orthogonal direction also, two edge gradients are obtained, which together represent the intensity gradient of the image along the chosen direction. As the smoothing is done along a direction orthogonal to the direction of the edge extraction, smearing of the edges is reduced.

The 1-D Gaussian function along a line, which makes an angle θ with x-axis, and has a perpendicular distance of ρ from the origin, is given by

$$f_\theta(x) = \frac{1}{\sqrt{2\pi}\sigma_1}\exp\left(\frac{-x^2\sec^2\theta}{2\sigma_1^2}\right), \tag{11}$$

where σ_1 is the standard deviation of the Gaussian function. The derivative operator is given by

$$g_\theta(x) = \frac{-x\sec^2\theta}{\sqrt{2\pi}\sigma_2^3}\exp\left(\frac{-x^2\sec^2\theta}{2\sigma_2^2}\right), \tag{12}$$

where σ_2 is the standard deviation of the Gaussian function whose derivative is computed. For smoothing, 1-D convolution is performed along a straight line making an angle $\theta+90°$ with the x-axis in the image plane. The smoothed

image $i_\theta^s(x, y)$ is given by

$$i_\theta^s\left(x, \frac{\rho + x\sin(\theta+90)}{\cos(\theta+90)}\right)$$
$$= i\left(x, \frac{\rho + x\sin(\theta+90)}{\cos(\theta+90)}\right) \otimes f_{\theta+90}(x), \quad (13)$$

where '$\otimes$' denotes the 1-D convolution operator, and x is the independent variable. Convolution is performed for each value of ρ to obtain the smoothed image. This smoothed image is used to compute the edge gradient i_θ^g, by applying the derivative operator along the orthogonal direction, i.e., along the straight lines making an angle θ with x-axis in the image plane. The edge gradient ($i_\theta^g(x, y)$) is obtained using

$$i_\theta^g\left(x, \frac{\rho + x\sin\theta}{\cos\theta}\right) = i_\theta^s\left(x, \frac{\rho + x\sin\theta}{\cos\theta}\right) \otimes g_\theta(x) \quad (14)$$

for different values of ρ. By performing these two operations along the orthogonal direction we get $i_{\theta+90°}^g$, which along with i_θ^g represents the intensity gradient of the image along the direction θ. Figure 1 shows an example of the gradient maps obtained along different directions for the same face image.

Computation of the edge gradient by the above method is sensitive to background illumination. This can be seen in Fig. 2. The effect of background illumination can be reduced by dividing the edge gradients with the weighted average
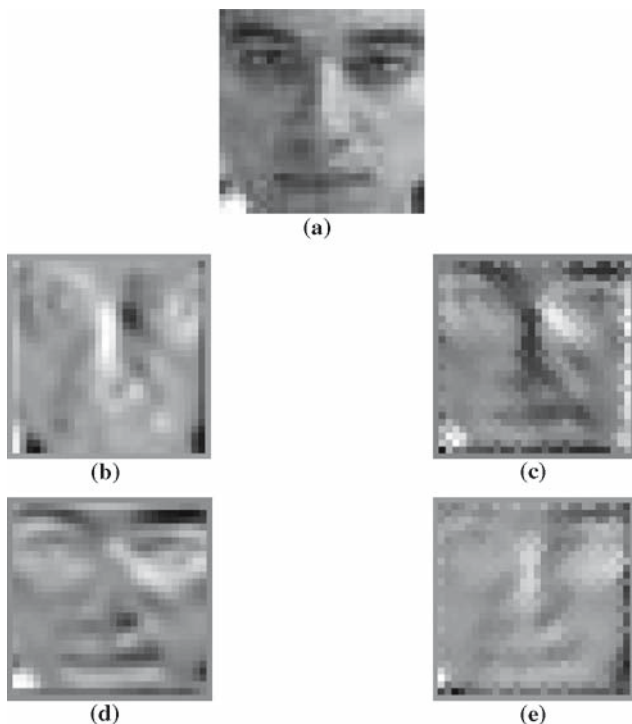

Fig. 1 **a** Gray level image. Edge gradient (i_θ^g) of the face image obtained using $\sigma_1 = 0.9, \sigma_2 = 1$ for **b** $\theta = 0°$, **c** $\theta = 45°$, **d** $\theta = 90°$, and **e** $\theta = 135°$
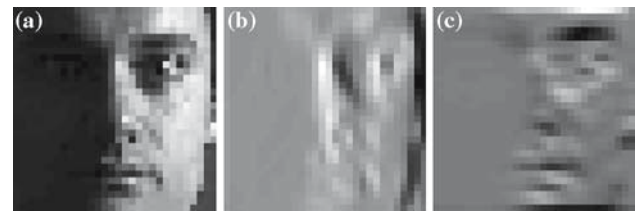
Fig. 2 **a** Gray level image with left shadow. Edge gradient (i_θ^g) of the face image **a**, obtained using $\sigma_1 = 0.9, \sigma_2 = 1$ for **b** $\theta = 0°$, and **c** $\theta = 90°$
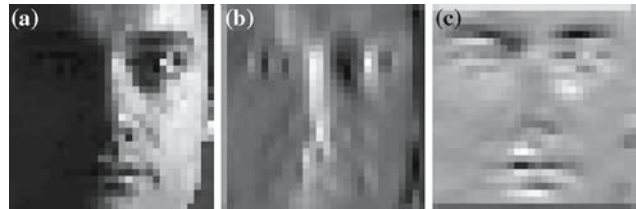

Fig. 3 **a** Gray level image with left shadow. Normalized edge gradient (i_θ^n) of the face image, obtained using $\sigma_1 = 0.9, \sigma_2 = 1$ for **b** $\theta = 0°$, and **c** $\theta = 90°$

value of the smoothed face image. The weigthed average value of the smoothed image is computed using

$$i_\theta^a\left(x, \frac{\rho + x\sin(\theta)}{\cos(\theta)}\right)$$
$$= i_{(\theta+90)}^s\left(x, \frac{\rho + x\sin(\theta)}{\cos(\theta)}\right) \otimes f_\theta(x) \quad (15)$$

for different values of ρ. The normalized edge gradient image is given by

$$i_\theta^n = \frac{i_\theta^g}{i_\theta^a}. \quad (16)$$

The normalized edge gradient images of a face image are shown in Fig. 3. It can be seen that the information in the left half regions is also visible in the normalized edge gradient images (Fig. 3b and c), as compared to the edge gradient image without normalization (Fig. 2b and c). The normalized edge gradient image (i_θ^n) is referred to as the edge gradient in the rest of the paper.

3.2 Face verification using edge gradient and MACE filter

Face verification is performed by cross-correlating the edge gradient (i_θ^n) of a test image with the corresponding MACE filter (h_θ) derived using similar representation in training. The block diagram of the matching process, performed using discrete Fourier Transform (DFT), is shown in Fig. 4. The filter output should have a high peak value at the origin in the correlation plane, when the test face image belongs to the true class. On the other hand, if the test face image belongs to the false class, then the filter output should give a low peak

Fig. 4 Face recognition by
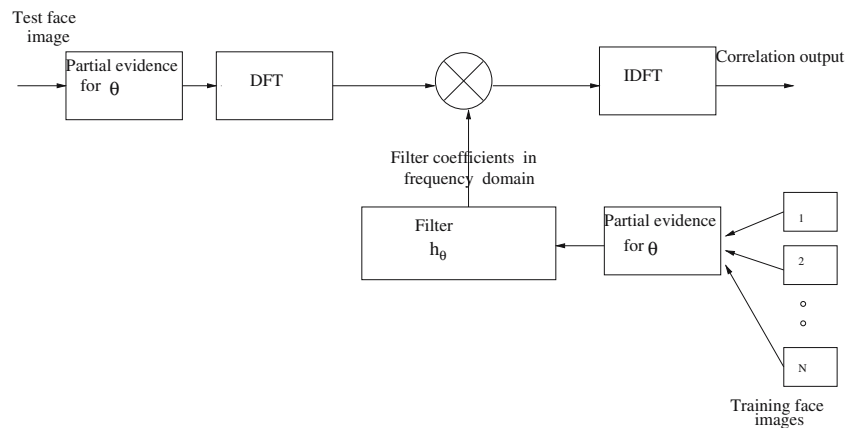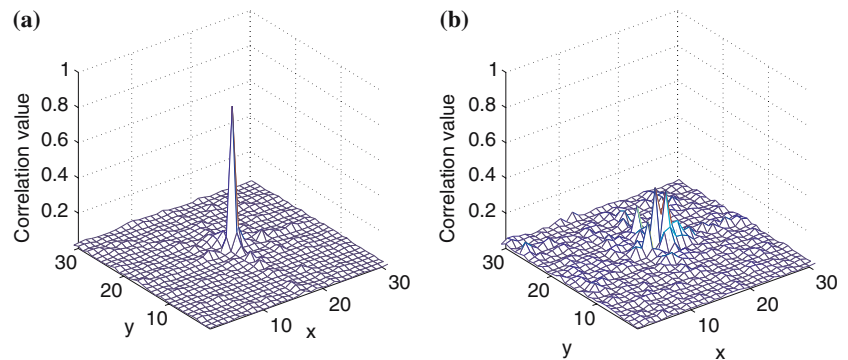correlation filter using edge
gradient representation

Fig. 5 Normalized correlation
output when i_θ^n, obtained for
$\sigma_1 = 0.9$, $\sigma_2 = 1$, $\theta = 0°$, is
used as input to the MACE filter
for **a** true class, and **b** false class

value at the origin. The relative heights of these peaks determine whether the input face is from the true or the false class. Figure 5a shows the normalized correlation output for a true class face image, and Fig. 5b shows the normalized correlation output for a false class face image, both for $\theta = 0°$. The correlation output of a true class image (Fig. 5a) has a high peak value near the origin, whereas for the false class image (Fig. 5b) the correlation output is generally low. The sharpness of the peak in the correlation output is quantified using peak-to-sidelobe ratio (PSR) [6]. The PSR is defined as

$$P = \frac{p - \mu}{\sigma}, \tag{17}$$

where p is the value of the maximum peak in the correlation output, μ is the mean of the correlation output within a window (of size 19×19 pixels) around the peak, and σ is the standard deviation of the values in this window. In actual practice, we leave out a region of size 7×7 pixels in the center of the window while computing μ and σ. The choice of size of these window was arrived at empirically, and other choices can also be made to highlight the peak to sidelobe characteristics.

The spatial spread (σ_2) of the first derivative of the Gaussian function controls the amount of smearing of the edges in the 1-D processing of the image. For low values of σ_2, even a slight deviation in the edge contour reduces the cor-

relation significantly. This leads to high intra-class variance. On the other hand, for high values of σ_2, the edges are smeared, reducing inter-class discrimination. Experimentally, we found that a value of $\sigma_2 \approx 1$ is reasonable for computing the edge gradient.

Another way of controlling the degree of spread of the edge information in the image plane is through the choice of the number of DFT coefficients of the image. Removing some of the high frequency DFT coefficients leads to blurring of the image [20]. The effect of removing the high frequency DFT coefficients is similar to that of increasing σ_2. We have found that removing the last 10 high frequency DFT coefficients along both the frequency axes improves the relative PSR for true and false classes. The size of the 2-D DFT is $(2N - 1) \times (2N - 1)$, where $N = 30$ for an image of 30×30 pixels.

3.3 Combining partial evidences

More information of a face image can be obtained by using the edge gradient (i_θ^n) along different directions (θ), instead of using only one direction. Separate MACE filters are derived for each direction. The PSR value is obtained separately from each filter. A k-dimensional feature vector is formed for a given test face image, where each component corresponds
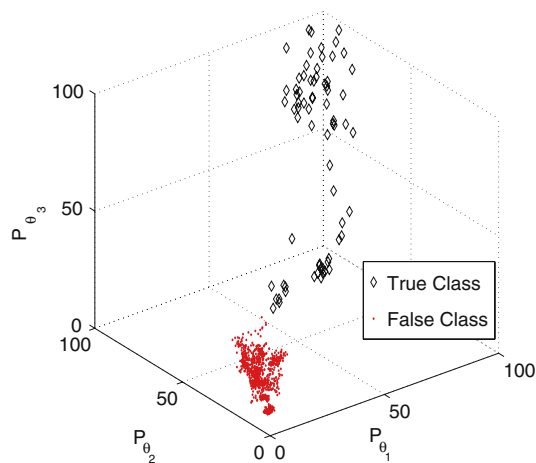
Fig. 6 Scatter plot for a subject using edge gradient representation of face images obtained using $\sigma_1 = 0.9$, $\sigma_2 = 1$, $\theta_1 = 0°$, $\theta_2 = 45°$, $\theta_3 = 90°$. The last 10 high frequency DFT coefficients along both the frequency axes of the 2-D DFT of the edge gradient are eliminated. Three training images are used to derive the MACE filter for each direction. Note that only three values of θ are used to show the scatter plot in 3-D

to the PSR value obtained from the filter designed for one direction, i.e., for one value of θ.

In this study partial evidence is obtained along four different directions ($\theta = 0°$, $45°$, $90°$, $135°$). For each subject in the database, we have a bank of four filters, one for each value of θ. These filters are used to compute the 4-dimensional feature vector (consists of 4 PSR values) for a given test face image. Figure 6 shows the scatter plots showing the feature (PSR) vectors of the true and false class images for a subject. For visualization, the plot shows points using only three ($\theta = 0°$, $45°$, $90°$) of the four component of the vector. For this example, three training images of the subject are used to derive a MACE filter for each direction. The remaining 72 face images of the subject form examples of true class. The PSR vectors are denoted by the diamond ($\diamond$) symbol in the scatter plot. For the false class, 12×75 (i.e., 900) face images are available, and the corresponding PSR vectors are shown by the dot ($\cdot$) symbol in the scatter plot. The scatter plot shows

the distribution of the evidence for the true and false classes. For the templates of a given subject, when a test image of the false class subject (impostor) is applied to bank of filters, the value of the evidence will be low. This is because any image belonging to false class cannot match well with the image of the true class. Note that the points due to an image of the true class generally have higher evidence value. But due to variability in the test image of the true class, the points belonging to the true class will have high variability as can be seen from Fig. 6.

One can generate a large number of points belonging to the false class for a given subject by using the face images of a large number of other subjects. The distribution of cluster of these closely spaced points can be used to represent the false class. If a point due to new test image falls outside the cluster, then the claim of the corresponding test image is accepted. Otherwise the claim is rejected.

The distribution of the scatter points due to false class for each subject can be captured by using any distribution capturing method like Gaussian mixture method (GMM) [21] or auto-associative neural network (AANN) [22] model. The distribution capturing capability of AANN is illustrated in Fig. 7b for the 2-D data shown in Fig. 7a. The structure of the AANN model is 2L 10N 10N 1N 10N 10N 2L, where L denotes a linear unit, and N denotes a nonlinear unit. The number of unit in the input and output layers correspond to the dimension of the data feature vector. The network is trained using the coordinates values of the data as input and output. Figure 7b is obtained using the confidence value derived from the error between the input and output (of the trained network) for each test point in the 2-D plane in the range (-1 to $+1$, -1 to $+1$). The confidence value is computed from the squared error (E_i) for the ith point as

$$C_i = \exp(-E_i/\alpha), \tag{18}$$

where α is a constant. In this example $\alpha = 0.025$. It can be seen that the confidence value is high when the density of points is high in the input data.

The derived AANN model can be used for accepting or rejecting a claim. When the feature vector derived from the

Fig. 7 Distribution capturing ability of AANN model.
a Artificial 2 dimensional data.
b Confidence surface realized by the network structure 2L 10N 10N 1N 10N 10N 2L
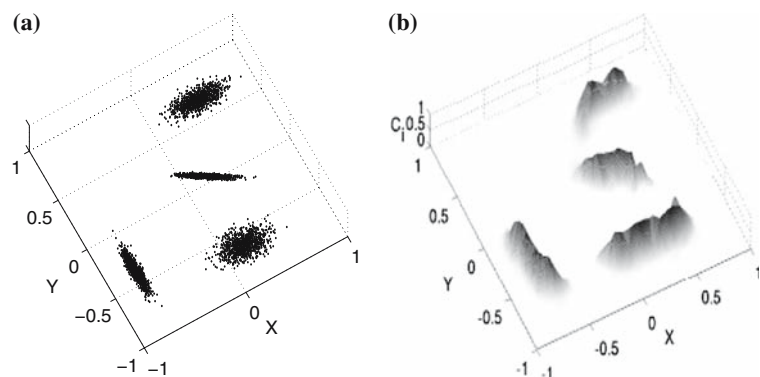
test image belonging to the true class is given to the AANN model, at least some of the partial evidences (component of the feature or PSR vector) could be large. Thus the resulting evidence is not likely to fall into the cluster of points belonging to the false class. Using a suitable threshold for the output of each AANN model, a decision can be made whether to accept the claim of the test input or not. The structure of the AANN model used to capture the distribution of the false class with respect to each subject is $4L\,8N\,2N\,8N\,4L$. Note that the structure depend on the nature of the cluster. In this case since the feature vector has four component (PSR values), the number of units in each of the input and output layers is 4. The number of unit in the intermediate hidden layers depend on the number of clusters and the nature of the clusters. If the clusters are dense and small then the number of the units in the hidden layers also will be small. The number of units in the hidden layers are chosen after experimenting with large amount of data with different structure. The AANN is trained for each subject using backpropagation algorithm for about 10,000 epochs.

We give a brief summary of the complete process discussed above. During training, for each subject, a few images are chosen as training images. The edge gradients along the four directions ($0°, 45°, 90°, 135°$) are computed for the training face images. The edge gradients of all the training images are then used to derive a correlation filter for each direction separately. Several images of false class are presented to bank of correlation filter to obtain a set of 4-D feature (PSR) vectors. These vectors are used to train an AANN model to capture the distribution of the feature vectors belonging to the false class. During testing, the edge gradient along the four directions are computed from the test face image. These edge gradient information is presented to bank of correlation filters to obtain the 4-D feature (PSR) vector. The 4-D feature vector is given as input to the AANN model. The error is computed between output and input of the AANN model. If the error is above the threshold for that model, then the claim is accepted.

Performance of the proposed approach is evaluated using two metrics: false acceptance ratio (FAR) and false rejection ratio (FRR). The trade off between FAR and FRR is a function of the decision threshold. Equal error rate (EER) is the value for which the error rates FAR and FRR are equal. Here, we explain the computation of EER for a single person using three training face images. The correlation filters for four directions ($0°, 45°, 90°,$ and $135°$) are derived using the three training face images. The remaining $72 = (75-3)$ face images form the examples of true class for testing. For the false class, 75×12 (i.e., 900) face images are available. Out of these, 700 face images are used to train the AANN model. The remaining $200 = (900 - 700)$ face images are used for testing. By varying the threshold value of AANN, the receiver operating characteristic (ROC) is obtained as shown in Fig. 8.
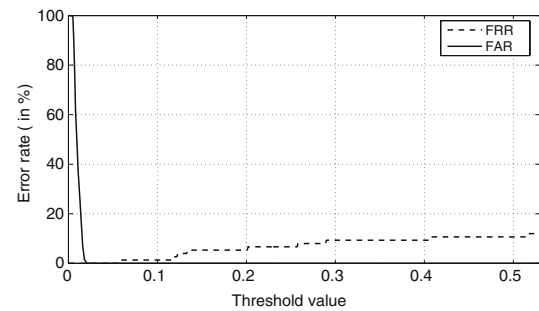

Fig. 8 Variation of false acceptance and false rejection ratio as functions of threshold

The ROC shows that the FAR curve is steep indicating that the corresponding PSR values are clustered around low values. On the other hand, the FRR curve is slowly varying, indicating that the corresponding PSR values are more scattered. The intersection point of FAR and FRR curves gives the EER for the person. Similarly we have computed EER for all subjects, and for cases with different number of training face images. These results are discussed in Sect. 6 (Table 2).

4 Image matching using edge orientation representation

Edge orientation can also be used to represent the face images. Similar to the approach adopted for the edge gradient representation, the edge extraction is carried out along the four directions ($0°, 45°, 90°, 135°$) to obtain the normalized edge gradients $i_0^n, i_{45}^n, i_{90}^n$, and i_{135}^n, respectively. The edge orientation map i_θ^d is given by

$$i_\theta^d = \frac{i_\theta^n}{\sqrt{(i_\theta^n)^2 + (i_{\theta+90}^n)^2}}. \tag{19}$$

Figure 9 shows the edge orientation map obtained for different directions. The matching process is same as that followed for the case of edge gradient representation. The best result was obtained when $\sigma_2 = 1$ and when the last 10 high frequency DFT coefficients along both the axes of the 2-D DFT of the edge orientation were eliminated. The results are discussed in Sect. 6 (Table 3).

5 Image matching using potential field representation

One of the problems with edge gradient and edge orientation representations is that they are very sparse, i.e., most of the values are very small, close to 0. Because of this, even a small deviation in the edge contour for the same face image can significantly reduce the value of the correlation peak. Therefore, the variance of the feature (PSR) vectors obtained by matching intra-class images will be high. This problem

**Table 1** Results for gray level representation: EER (in %) for different number (N) of training face images

Subject	1	2	3	4	5	6	7	8	9	10	11	12	13
EER (N=5)	0	1.3	0	0	0	0	0	0	0	0	0	0	0
EER (N=3)	0	0.9	0	0	1	0	0	0	0	0	0	0	0
EER (N=1)	0	18.3	5.2	0	3.5	0	0	4.1	0	0	0	0	0

Table 2 Results for the edge gradient representation: EER (in %) for different number (N) of training face images for each person and using AANN for combining partial evidences from four directions ($\theta = 0°$, 45°, 90°, 135°)

Subject	1	2	3	4	5	6	7	8	9	10	11	12	13
EER ($N = 5$)	0	2.6	0	0	0	0	0	0	0	0	0	0	0
EER ($N = 3$)	0	12.5	0	0	0	0	0	0	0	0	0	0	0
EER ($N = 1$)	0	23.6	6.6	0	0	0	0	0	0	0	0	0	0

Table 3 Results for the edge orientation representation: EER (in %) for different (N) of training face images and using AANN for combining partial evidences from four directions ($\theta = 0°$, 45°, 90°, 135°)

Subject	1	2	3	4	5	6	7	8	9	10	11	12	13
EER ($N = 5$)	0	6.2	0	0	0	0	0	0	0	0	0	0	0
EER ($N = 3$)	0	12.3	0	15.2	9.2	0	0	4.2	0	0	0	0	0
EER ($N = 1$)	0	23.6	11.2	18.5	10.8	0	0	6.2	0	0	0	10.2	0
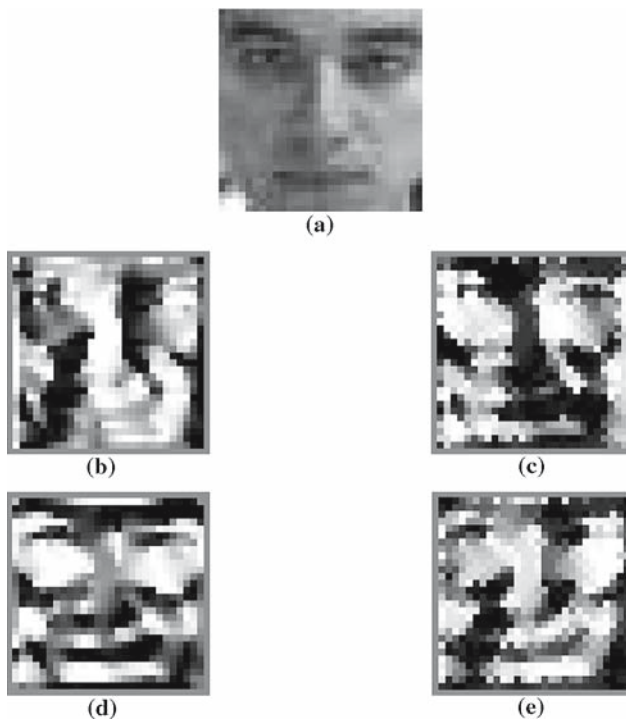
Fig. 9 **a** Gray level face image. Edge orientation (i_θ^d) of the face image obtained using $\sigma_1 = 0.9$, $\sigma_2 = 1$ for **b** $\theta = 0°$, **c** $\theta = 90°$, **d** $\theta = 90°$, **e** $\theta = 135°$

is reduced by spreading the edge information in the edge gradient or edge orientation representations. The edge information can be spread out by either using high values of σ_2,

or by removing some of the high frequency DFT coefficients of the representations. However, spreading the information in this manner leads to smearing of the edges.

Alternatively, the edge information can be spread out using potential field representation. The potential field u_θ derived from the edge gradient i_θ^n, is obtained by minimizing the energy E_θ given by [23]

$$E_\theta = \int \int [\mu((u_{\theta_x})^2 + (u_{\theta_y})^2) + |(i_\theta^n)^2 + (i_{\theta+90}^n)^2| \, |u_\theta - i_\theta^n|^2] \mathrm{d}x \mathrm{d}y, \tag{20}$$

where u_{θ_x} and u_{θ_y} are the partial derivatives of u_θ along the x and y axes, respectively. The parameter μ is the scaling factor used to control the amount of smoothing. This variational formulation follows the standard principle that the result must be smooth when there is no data. In particular, we see that when the gradients i_θ^n and $i_{\theta+90}^n$ are small, the energy is dominated by the partial derivatives of the field, yielding a smooth field. This smoothing term (the first term in the integrand of (20)) is the same term used by Horn and Schunk in their classical formulation of optical flow [24]. On the other hand, when the gradient is large, the second term dominates in the integrand of (20), and it is minimum when $u_\theta = i_\theta^n$. This term is responsible for introducing the gradient information in the potential field. Our approach is similar to the one used in [23] for developing a deformation force of snakes. To minimize the energy E_θ given by (20), the

Table 4 Results for the potential field representation: EER (in %) for different (N) of training face images and using AANN for combining partial evidences from four directions ($\theta = 0°, 45°, 90°, 135°$)

Subject	1	2	3	4	5	6	7	8	9	10	11	12	13
EER ($N = 5$)	0	0	0	0	0	0	0	0	0	0	0	0	0
EER ($N = 3$)	0	0	0	0	0	0	0	0	0	0	0	0	0
EER ($N = 1$)	0	10.2	0	0	0	0	0	0	0	0	0	0	0

following Euler equation is a necessary condition [25]

$$\mu \, \nabla^2 \, u_\theta - (u_\theta - i_\theta^n)|(i_\theta^n)^2 + (i_{\theta+90}^n)^2|^2 = 0, \tag{21}$$

where ∇^2 is the Laplacian operator. We can see that wherever the gradient is zero, (21) will reduce to Laplacian operator. It can be solved by considering u_θ as a function of the variable t, and the solution is given by [23]

$$
\begin{aligned}
u_{\theta\, t}(x, y, t) = {} & \mu \, \nabla^2 \, u_\theta(x, y, t) - [u_\theta(x, y, t) \\
& - i_\theta^n(x, y)][(i_\theta^n(x, y))^2 + (i_{\theta+90}^n(x, y))^2],
\end{aligned}
\tag{22}
$$

where $u_{\theta\, t}$ is the partial derivative of $u(x, y, t)$ with respect to t. The above equation is known as generalized diffusion equation, commonly encountered in heat conduction and reactor physics [26]. Let $\triangle x$ and $\triangle y$ be the inter pixel distance along x and y axes, and $\triangle t$ be the change in t. Then, using the standard approximations for the partial derivatives, (22) can be written as

$$
\begin{aligned}
u_{\theta\, i,j}^{m+1} = {} & (1 - E_{i,j}^2 \, \triangle t) u_{\theta\, i,j}^m + \frac{\mu \, \triangle t}{\triangle x \, \triangle y}(u_{\theta\, i,j+1}^m + u_{\theta\, i+1,j}^m \\
& + u_{\theta\, i-1,j}^m + u_{\theta\, i,j-1}^m - 4u_{\theta\, i,j}^m) + E_{i,j}^2(i_{\theta\, i,j}^n \, \triangle t),
\end{aligned}
\tag{23}
$$

where $E_{i,j} = \sqrt{(i_{\theta\, i,j}^n)^2 + (i_{\theta+90\, i,j}^n)^2}$ and $u_{\theta\, i,j}^m$ is the potential field at the location (i, j) after mth iteration. Equation (23) is stable, provided $E_{i,j}$ and $i_{\theta\, i,j}^n$ are bounded, and the Courant–Friedrich–Lewis restriction $\frac{\mu\triangle t}{\triangle x\triangle y} \leq \frac{1}{4}$ is maintained [25]. The potential fields obtained by (23) for different edge gradients are shown in Fig. 10. The matching process using the potential field is same as that used for the case of the edge gradient representation. The results are discussed in Sect. 6 (Table 4).

6 Experimental results

Of the three representations of the face images for correlation-based face verification, we expect the potential field of the edge gradient to give the smoothest correlation output. By *smooth* we mean the reduction of small peaks around the actual correlation peak. The edge orientation, due to the absence of the magnitude of the edge gradient, is somewhat rough textured. Hence, the correlation filter output obtained
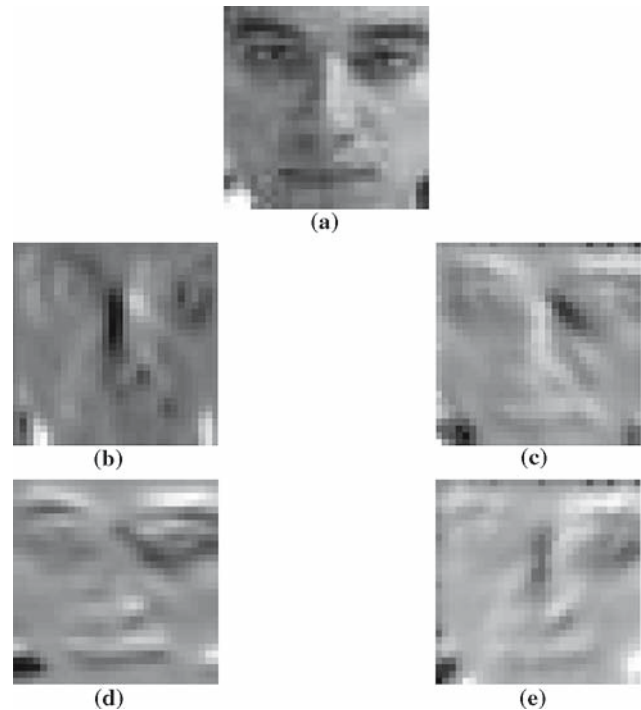

Fig. 10 **a** Gray level image. Potential field (u_θ) developed from the edge gradient of the face image using $\sigma_1 = 0.9, \sigma_2 = 0.6, \mu = 0.005$ for **b** $\theta = 0°$, **c** $\theta = 45°$, **d** $\theta = 90°$, and **e** $\theta = 135°$

for the case of edge orientation is the roughest among the three representations. This is illustrated in Fig. 11, which shows the correlation outputs for the three representations for the case of $\theta = 90°$. Figure 12 shows the scatter plots of the PSR vectors for the three representations using three training images of the subject to derive the MACE filter. For visualization, only 3 of the 4 dimensions ($\theta = 0°, 45°, 90°$) are used in plots. The separation between the two clusters (true and false classes) is better for the case of potential field representation.

The performance (given in terms of EER) for different representations for all the 13 subjects, using different sizes of training set are shown in Tables 2, 3, and 4. Table 1 shows the EER obtained when the MACE filter is derived using gray level representation of the face images [6]. By comparing the results in the tables we observe that the best results are obtained when potential field representation is used. In general, as the number of samples of the training set decreases, the performance also degrades. One can also observe from
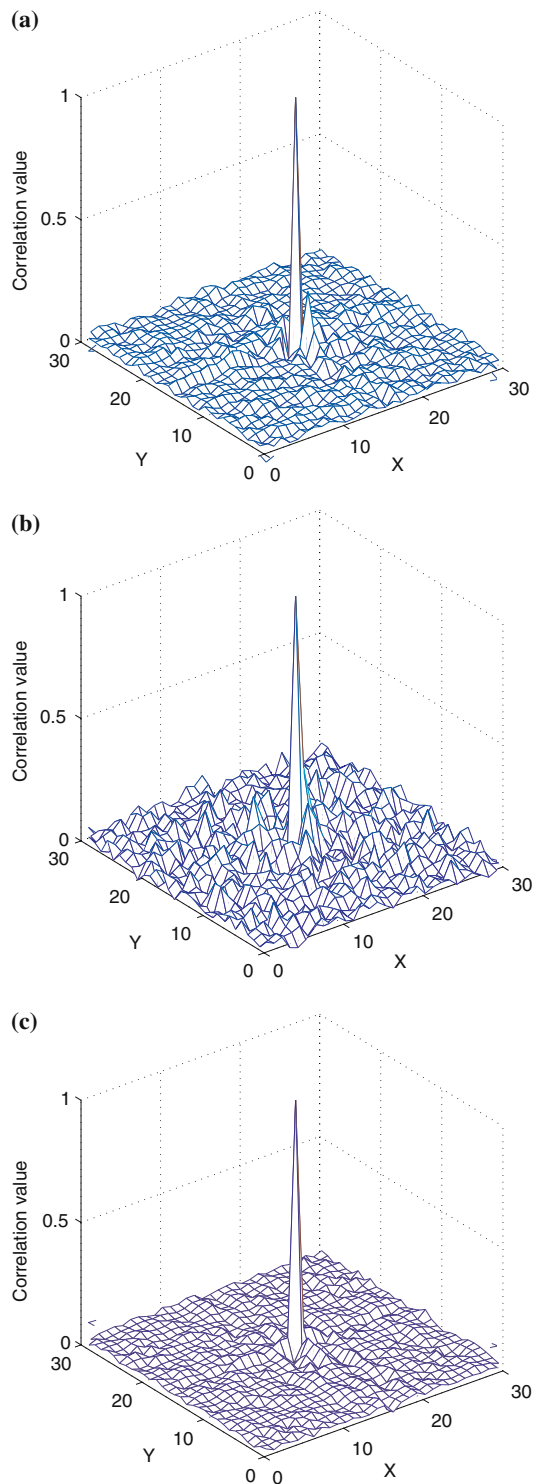
(a)

(b)

(c)

Fig. 11 Correlation outputs using **a** edge gradient representation, **b** edge orientation representation, and **c** potential field representation

the tables that the EER is high for subject 2 using all representations. The reason for this is that subject 2 has more variation in expression than other subjects, and hence it is more difficult to match the test images of the subject with the training images.
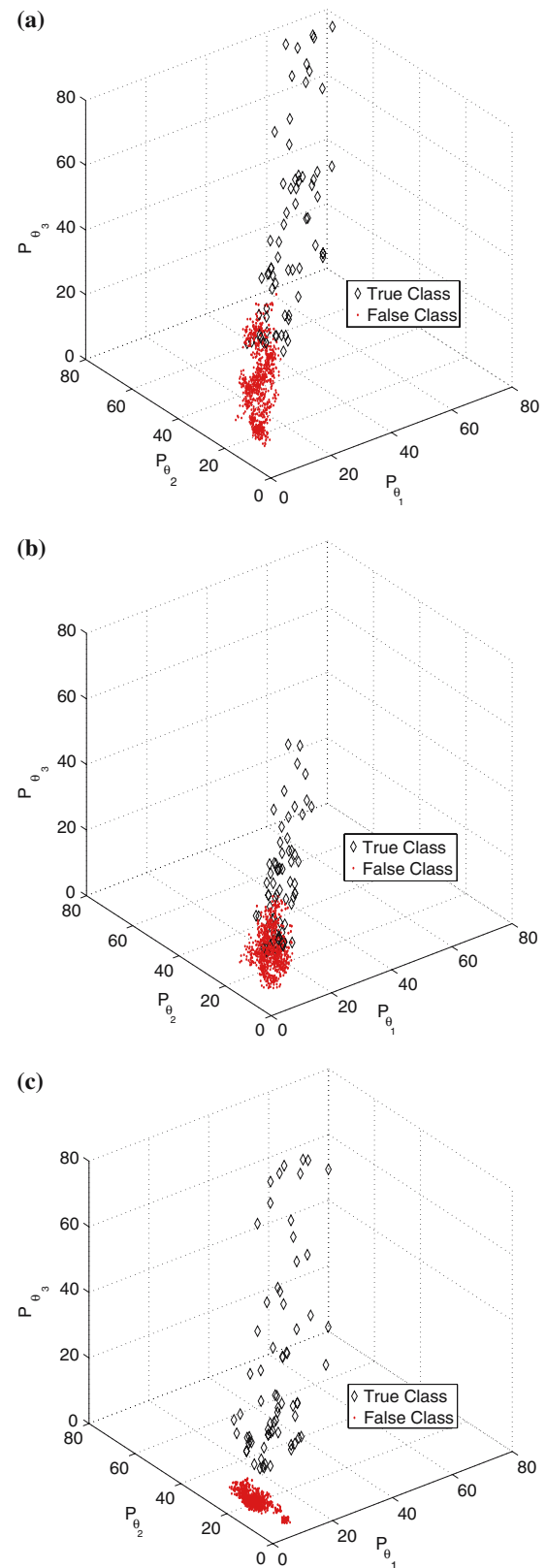
(a)

(b)

(c)

Fig. 12 Scatter plots for a subject using **a** edge gradient representation, **b** edge orientation representation and **c** potential field representation, obtained using $\theta_1 = 0°$, $\theta_2 = 45°$, $\theta_3 = 90°$. Three training face images are used to derive the MACE filter for each direction

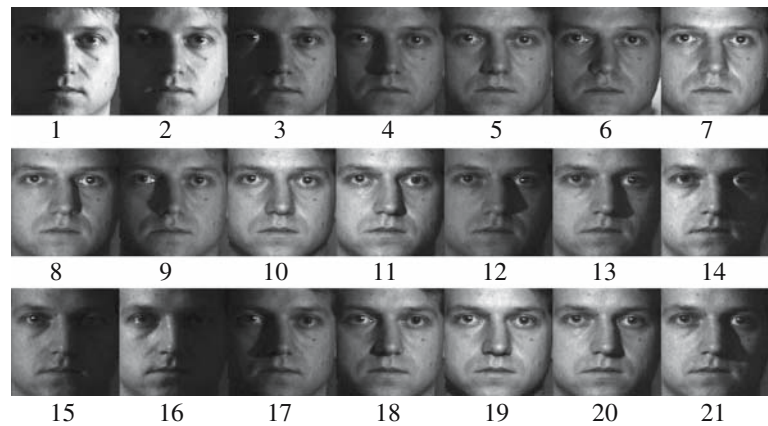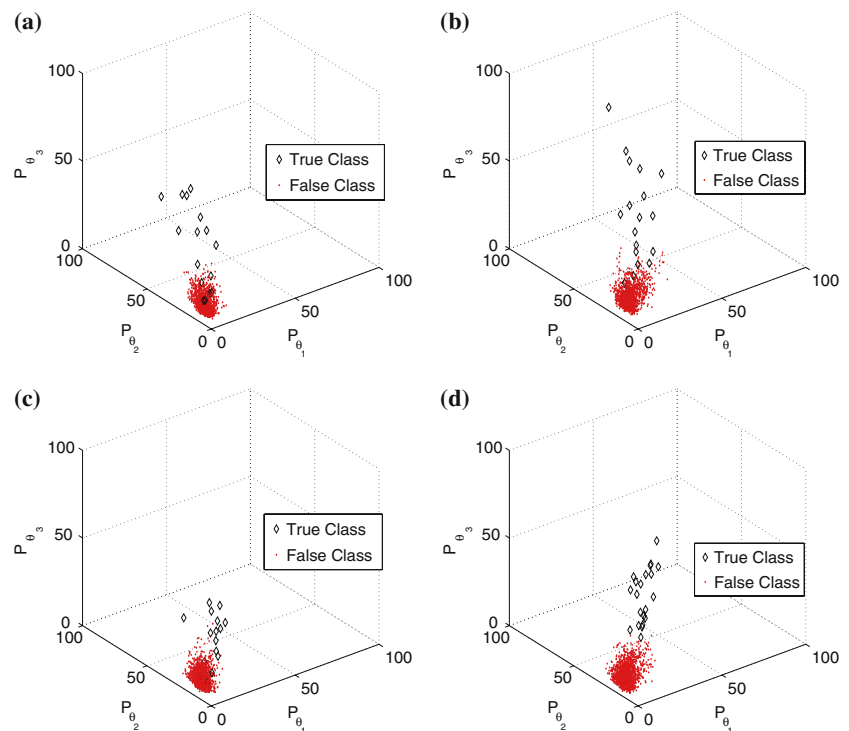**Fig. 13** Sample images of a person from illumination subset of PIE database

Fig. 14 Scatter plots for a subject using potential field representation, obtained from $\theta_1 = 0°$, $\theta_2 = 45°$, $\theta_3 = 90°$, and using training face image with **a** left shadow, **b** frontal view, **c** right shadow, and **d** left shadow, frontal view, and right shadow together


Performance of the potential field representation is also evaluated on a subset of the PIE database (PIE NL dataset) [27], containing 65 subjects, each having 21 face images. Figures 13 show 21 images of a subject. In our experiment, we have derived the MACE filters (along four directions) for each subject using potential field representation for three case of extreme lighting variations, Image 3 (left shadow), Image 7 (frontal lighting), and Image 16 (right shadow). The remaining 18 face images form examples of the true class. There are 21×64 (i.e., 1244) face images are available to be used as samples of false class. Out of these, 900 face images are used to train the AANN model to capture the distribution for each subject. The EER is computed as explained in Sect. 3.3, and an average EER of 1.85% is obtained. On

the other hand, an EER of 0% was reported in [28] when the MACE filters are trained with the same set of training images using gray level representation.

One way to improve the performance of the proposed approach is to use the training face images separately for matching, and then combine the evidence from each for verification. The reason for using training face images separately can be explained from the scatter plots shown in Fig. 14. The figure shows the scatter plots of feature (PSR) vectors using potential field representation for a given subject, using only one face image to design the filter: (Fig. 14a for the left shadow image, Fig. 14b for the frontal view image, and Fig. 14c for the right shadow image). The PSR vectors of the remaining 20 face images of the true class are shown by diamond (◇)

Fig. 15 Scatter plot obtained after combining Fig. 14a–c using (24) for **a** $r = 1$, and **b** $r = 2$
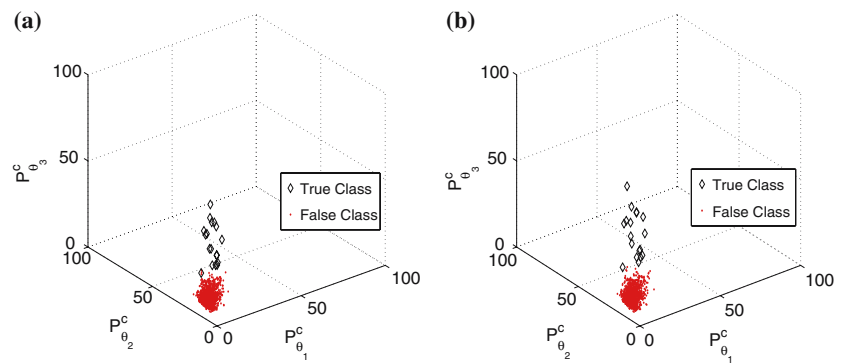
symbol in the scatter plots. If we use all the three training face images (left shadow, frontal view, and right shadow) to design the MACE filter, the resultant scatter plot for the subject is as shown in Fig. 14d. In this case the number of examples of the true class is 18. For the false class, the 1344 feature (PSR) vectors are shown by dot (·) symbol in Figs. 14a–d. The separation between the true and false classes is better in Fig. 14d as compared to Figs. 14a–c. This is due to matching with some kind of interpolated image by the MACE filter in the Fig. 14d.

One can observe from the scatter plots (Figs. 14a–c) that for a single training face image the points due to false class are clustered, whereas the points due to true class are scattered. It is likely that for each true class image, there may be some large evidence value at least from one of the training face images. One way to exploit this behavior is combine the PSR values obtained from each training image.

Let $P_\theta^{t,l}$ is the similarity score (PSR) obtained when the potential field representation along θ direction of the test face image I^t is correlated with the corresponding representation of the training face image I^l.

$$P_\theta^{cl} = \left(\sum_{k=1}^{K} (P_\theta^{kl})^r \right)^{\frac{1}{r}}, \tag{24}$$

where the parameter r decides the weights associated with the scores. For $r \leq 1$, $\min_k [P_\theta^{kl}] \leq P_\theta^{cl} \leq \frac{1}{K} \sum_{k=1}^{K} (P_\theta^{kl})$, and for $r \geq 1$, $\frac{1}{K} \sum_{k=1}^{K} (P_\theta^{kl}) \leq P_\theta^{cl} \leq \max_k [P_\theta^{kl}]$. A low value of r is suitable for the false class, and a high value of r for the true class. One has to choose a suitable value of r such that the separation between true and false classes is enhanced. We have found empirically that $r = 2$ is a good choice. Figure 15 shows the scatter plot obtained using the combined similarity scores for $r = 1$ and $r = 2$. The separation between the true and false class is better in Fig. 15b as compared to Fig. 14d. This is also evident from the EER values obtained using a procedure explained above. An average EER of 0.4 % is obtained using P_θ^{cl} for $r = 2$. In fact an

EER of 0% could be obtained by properly selecting the set of training images.

7 Summary

In this paper, we discussed the significance of representation of face image using correlation-based face verification. Three representations, namely, the edge gradient, the edge orientation and the potential field derived from the edge gradient are considered. The edge gradient and edge orientation representations are expected to be more robust to variation in illumination, compared to the gray scale representation. But these two representations are not well suited for correlation matching due to locality problem. This problem was addressed by using the potential field of the edge gradient.

The three representations are derived using one-dimensional (1-D) processing of the image, which has the advantage of providing multiple evidence for a given image. The evidences are combined using an AANN model to arrive at a decision for face verification. The proposed method of combining the evidences using AANN model has the advantage that it does not require large number of training images of the true class. Experimental results indicate that the potential field representation is superior to the edge gradient and edge orientation representing. Performance of the proposed representation for face verification task was evaluated on the face databases which have expression and illumination variations. The other variation which needs to be addressed is pose variation.

In this paper we have used the cropped face images. But, in practice, it is not easy to obtain cropped face image. This is a major issue that limits the performance of all correlation-based methods for matching face images.

References

1. Chellapa, R., Wilson, C., Sirohey, S.: Human and machine recognition of faces: a survey. Proc. IEEE **83**, 705–740 (1995)
2. Li, S.Z., Jain, A.K.: Handbook of Face Recognition. Springer, New York (2004)

3. Turk, M., Pentland, A.: Eigenfaces for recognition. J. Cogn. Neurosci. **3**, 71–86 (1991)
4. Brunelli, R., Poggio, T.: Face recognition: features versus templates. IEEE Trans. Pattern Anal. Mach. Intell. **15**, 1042–1052 (1993)
5. Baron, R.: Mechanism of human facial recognition. IEEE Trans. Syst. Man, Cybern. **15**, 137–178 (1981)
6. Kumar, B.V.K.V., Savvides, M., Venkataramani, K., Xie, C.: Spatial frequency domain image processing for biometric recognition. In: IEEE Proceedings of International Conference on Image Processing (ICIP), pp. 53–56. Rochester, New York (2002)
7. Gao, Y., Leung, M.K.H.: Face recognition using line edge map. IEEE Trans. Pattern Anal. Machine Intell. **24**, 765–779 (2002)
8. Manjunath, B.S., Chellappa, R., Malsburg, C.V.D.: A features based approach to face recognition. In: Proceedings of IEEE International Conference Computer Vision Pattern recognition. pp. 373–378 (1992)
9. Kotropoulos, C.L., Tefas, A., Pitas, I.: Using support vector machines to enhance the performance of elastic graph matching. IEEE Trans. Pattern Anal. Mach. Intell. **23**, 735–746 (2001)
10. Bruce, V., Hancock, P.J.B., Burton, A.M.: Human face recognition and identification. In: Proc. Face Recognition from Theory to Applications, pp. 373–378 (1998)
11. Mahalanobis, A., Kumar, B.V.K.V., Casasent, D.: Minimum average correlation energy filters. Appl. Opt. **26**, 3633–3640 (1987)
12. Savvides, M.: Kumar, B.V.K.V.: Efficient design of advanced correlation filters for robust distortions-tolerant face identification. In: IEEE International Conference Advanced Video and Signal based Surveillance (2003)
13. Kumar, P.K., Das, S., Yegnanarayana, B.: One-dimensional processing of images. In: Proceedings International Conference Multimedia Processing and Systems, IIT Madras. (2000)
14. http://amp.ece.cmu.edu: (Advanced Multimedia Processing Lab web page at Electrical and Computer Engineering Department at CMU)
15. Hester, C.F., Casasent, D.: Multivariate technique for multiclass pattern recognition. Appl. Opt. **19**, 1758–1761 (1980)
16. Kumar, B.V.K.V.: Tutorial survey of composite filter designs for optical correltors. Appl. Opt. **31**, 773–4801 (1992)
17. Kumar, B.V.K.V.: Minimum variance synthetic discrimination function. J. Opt. Soc. Am. **3**, 1579–1584 (1986)
18. Yilmaz, A., Gokmen, M.: Eigenhill vs. eigenface and eigenedge. Pattern Recogn. **34**, 181–184 (2000)
19. Canny, J.: A computational approach to edge detection. IEEE Trans. Pattern Anal. Mach. Intell. **8**, 679–698 (1986)
20. Oppenheim, A.V., Schafer, R.W.: Discrete-time Signal Processing. Prentice Hall of India, New Delhi (1997)
21. Redner, R.A., Walker, H.F.: Mixture densities, maximum likelihood and the EM algorithm. SIAM-REVIEW **26**, 195–239 (1984)
22. Yegnanarayana, B., Kishore, S.P.: AANN: an alternative to GMM for pattern recognition. Neural Netw. **15**, 459–469 (2002)
23. Xu, C., Prince, J.L.: Gradient vector flow: a new external force for snakes. In: IEEE Proc. Computer Vision Pattern Recognition, pp. 66–71 (1997)
24. Horn, B.K.P., Schunck, B.G.: Determining optical flow. Artif. Intell. **17**, 185–203 (1991)
25. Courant, R., Hilbert, D.: Methods of Mathematical Physics, Vol. 1. Interscience, New York (1953)
26. Charles, A.H., Porsching, T.A.: Numerical Analysis of Partial Differential Equations. Prentice-Hall, Engelwood Cliffs (1990)
27. Sim, T., Baker, S., Bsat, M.: The CMU pose illumination, and expression PIE database of human faces. Tech. Report CMU-RI-TR-01-02, Robotics Institute, Carnegie Mellon University (2001)
28. Savvides, M., Kumar, B.V.K.V.: Illumination normalization using logarithm transforms for face authentication. In: Audio and Video-Based Biometric Person Authentication (2003)