Research Article

Multithread Face Recognition in Cloud

Dakshina Ranjan Kisku[1] and Srinibas Rana[2]

[1]Department of Computer Science and Engineering, National Institute of Technology Durgapur, Durgapur,
 Bardhaman, West Bengal 713205, India
[2]Department of Computer Science and Engineering, Jalpaiguri Government Engineering College, Jalpaiguri,
 West Bengal 735102, India

Correspondence should be addressed to Dakshina Ranjan Kisku; drkisku@gmail.com

Faces are highly challenging and dynamic objects that are employed as biometrics evidence in identity verification. Recently, biometrics systems have proven to be an essential security tools, in which bulk matching of enrolled people and watch lists is performed every day. To facilitate this process, organizations with large computing facilities need to maintain these facilities. To minimize the burden of maintaining these costly facilities for enrollment and recognition, multinational companies can transfer this responsibility to third-party vendors who can maintain cloud computing infrastructures for recognition. In this paper, we showcase cloud computing-enabled face recognition, which utilizes PCA-characterized face instances and reduces the number of invariant SIFT points that are extracted from each face. To achieve high interclass and low intraclass variances, a set of six PCA-characterized face instances is computed on columns of each face image by varying the number of principal components. Extracted SIFT keypoints are fused using sum and max fusion rules. A novel cohort selection technique is applied to increase the total performance. The proposed protomodel is tested on BioID and FEI face databases, and the efficacy of the system is proven based on the obtained results. We also compare the proposed method with other well-known methods.

1. Introduction

Two-dimensional face recognition [1, 2] is considered an unsolved problem in the achievement of robust performance in the area of human identity. Face analysis with various feature representation techniques has been explored in many studies. Among various feature extraction approaches, appearance-based, feature-based, and model-based techniques are popular. Due to changes in illumination, clutter, head poses, and facial expressions (happy, angry, sad, confused, and surprised), major salient features and occlusion can cause degradation in face recognition performance, even after substantial matching is performed. A limited number of studies address face recognition, which considers noisy features and redundant outliers that are combined with distinctive facial characteristics for matching. These noisy and redundant features are frequently associated with regular facial characteristics during template generation and matching. Face recognition can negatively impact total

performance despite considerable efforts to denoise the effect of redundant characteristics. To overcome this situation, suitable feature descriptors [3] and feature dimensionality reduction techniques [4] can be employed to obtain compact representations. The presence of facial expressions and different lighting conditions can also increase the load in the matching process and complicate face recognition. Face recognition may effectively address these issues.

Due to an increase in subject enrollment and bulk matching, a significant number of computing resources can be housed within an organization's computing facilities. However, these types of facilities have some demerits. To maintain computing resources with existing resources is costly and requires a separate setup. We can overcome these shortcomings by transferring the responsibility of maintaining biometrics resources to a third-party service provider who maintains cloud computing infrastructures, which are housed with their own infrastructures. Integrating cloud computing facilities with a face recognition system can

facilitate the recognition of bulk faces from devices, such as CCTV cameras, webcams, mobile phones, and tablet PCs. This paradigm can be employed to handle a large number of people at different times, whereas cloud-enabled services enable enrollment and the matching process to be remotely conducted.

1.1. Cloud Framework. With the advancement of cloud computing [5, 6], many organizations are rapidly adopting IT-enabled services that are hosted by cloud service providers. Because these services are provided over a network, the cost of hosting services is fixed and predictable. Because cloud computing is very convenient and provides on-demand access to a shared pool of configurable computing resources (servers, networks, storage, applications, and services) over a network, this on-demand service can be availed by organizations who engage in minimal resource efforts and reliable cloud service providers who host cloud infrastructures. Three types of cloud computing models are available, namely, Platform as a Service (PaaS), Software as a Service (SaaS), and Infrastructure as a Service (IaaS). They are collectively known as the SPI model. The SaaS model includes various software and applications that are hosted and run by vendors and service providers and made available to customers over a network. The PaaS model includes delivering operating systems and development tools to customers over a network without the need to download and install them. The IaaS model involves requesting on-demand services of servers, storage, networking equipment, and various support tools over a network.

Cloud-based biometric infrastructures [7, 8] can be developed and hosted at a service provider's location. The on-demand services are available to businesses via network connectivity. Three models (PaaS, SaaS, and IaaS) can be subsequently employed for appropriate physiological or behavioral biometrics applications. Servers and storage can be employed for storing biometric templates, which can be employed for verification or identification. Biometric sensors can be installed on business premises with Internet connectivity via various networks and can be connected to cloud infrastructures to access stored templates for matching and enrollment. In addition, the biometric sensors are employed for enrollment and matching, and the process can run with the help of user interfaces, applications, support tools, networking equipment, storage, servers, and operating systems at the service provider's end, where the biometrics cloud is hosted. Businesses and organizations who want to avail a cloud-based facility for enrollment, authentication, and identification purposes need to have biometrics sensors and Internet connectivity. The SPI model can be employed for preprocessing, feature extraction, template generation, and face matching and decisions, which can be modeled as software models and application programs to be hosted at a service provider's cloud facility.

A few biometrics authentication systems [7–9] have been successfully employed in cloud computing infrastructures. They have facilitated the use of biometrics cloud concepts to minimize the efforts of resource utilization and bulk matching.

1.2. Studies on Baseline Face Recognition. Because we introduce a cloud-based biometric facility to be integrated with a face recognition system, a brief review of baseline face recognition algorithms would be advantageous to develop an efficient cloud-enabled biometric system. Face recognition [1, 2] is a long-standing computer vision problem that has gained the attention of researchers, whereas appearance-based techniques are employed to analyze the face and reduce dimensionality. Projecting a face onto a sufficiently low-dimensional feature space while retaining the distinctive facial characteristics in a feature vector serves a crucial role in recognizing typical faces. The application of appearance-based approaches in face recognition is discussed in [10–15]. Principal component analysis (PCA), linear discriminant analysis (LDA), kernel PCA, Fisher linear discriminant analysis (FLDA), canonical covariate, and the fusion of PCA and LDA are popular approaches in face recognition.

Feature-based techniques [16–18] have been introduced and successfully applied to represent facial characteristics and encode them to invariant descriptors for face analysis and recognition. Many models, such as EBGM [16], SIFT [17–19], and SURF [20], have been employed in face recognition. Local feature-based descriptors can also be employed for object detection, object recognition, and image retrieval problems. These descriptors are robust to lighting conditions, image locations, and projective transformation and are insensitive to noise and image correlations. Local descriptors are localized and detected at local peaks in a scale-space search; after a number of stage filtering approaches, interest points that are stable over transformations are preserved. A focus on two things must be established while local feature descriptors are employed. First, the algorithm must have the ability to create a distinctive feature description for differentiating one interest point from other interest points; second, the algorithm should be invariant to camera position, subject position, and lighting conditions. A situation may arise when a high-dimensional feature space is projected onto a low-dimensional feature space and local descriptors vary from one feature space to another feature space. Thus, the accuracy may change over the same object while interest points are detected in the scale-space over a number of encountered transformations. Due to scaling and low-dimensional projected variables, a higher variance among the observed variables should be retained in the high-dimensional data of a pattern. A reduced number of projected variables retain their characteristics even after they are represented onto a low-dimensional feature space. We can achieve this description, whereas appearance-based techniques are applied to raw images without the need for preprocessing techniques to restore true pixels or remove noise due to image-capturing sensors. A number of representations exist from which we can extract invariant interest points. One appearance-based technique is principal component analysis (PCA) [11, 12].

PCA is a simple dimensionality reduction technique that has many potential applications in computer vision.

However, despite a few shortcomings—it is restricted to orthogonal linear combinations and has implicit assumptions of Gaussian distributions—PCA has been proven to be an acclaimed technique due to its simplicity. In this study, PCA is combined with a feature-based technique (SIFT descriptor) for a number of face column instances to be generated by principal components. Face projections of column vectors range from one to six, in which the interval one can produce high variances in the observed variables of the corresponding face image after projecting the high-dimensional face matrix onto a low-dimensional feature space. Thus, we can obtain a set of low-dimensional feature spaces that correspond to each column of a single face image. Principal components are decided by a sequence of six integer numbers ranging from 1 to 6 to be considered in order and based on which six face instances are generated. Unlike a random sequence, an ordered default sequence is taken from the mathematical definition of eigenvector and the arithmetic distance of any point to its predecessor and successor principal component is always one. A SIFT descriptor, which is a suitable application for these representations, can produce multiple sets of invariant interest points without changing the dimension of each of the keypoint descriptors. This process can change the size of each vector, which consists of keypoint descriptors to be constructed on each projected PCA-characterized face instance. In addition, the SIFT descriptor is robust to partial illumination, projective transform, image locations, rotation, and scaling. The efficacy of the proposed approach has been tested on frontal view face images with mixed facial expressions. However, efficacy is compromised when the head position of a face image is modified.

1.3. Relevant Face Recognition Approaches. In this section, we introduce some related studies and discuss their usefulness in face recognition. For example, the algorithm proposed in [21] discusses a method that employs the local gradient patch around each SIFT point neighborhood and creates PCA-based local descriptors that are compact and invariant. However, the proposed method does not encode a neighborhood gradient patch of each point; instead, it makes a projected feature representation in the low-dimensional feature space with variable numbers of principal components. This method extracts SIFT interest points from the reduced face instances. Another study [22] examines the usefulness of the SIFT descriptor and PCA-WT (WT: wavelet transform) in face recognition. An eigenface is extracted from the PCA-wavelets representation, and SIFT points are subsequently detected and encoded to a feature vector. However, the computational time increases due to the complex and layered representation. A comparative study [23] employs PCA for neighborhood gradient patch representation around each SIFT point and a SURF point for invariant feature detection and encoding. Although PCA reduces the dimensions of the keypoint descriptor and compares the performances of the SIFT and SURF descriptors, it is not employed for face recognition but is applied to the image retrieval problem.

The remainder of the manuscript is organized as follows: Section 2 presents a brief outline of the cloud-based face recognition system. Short descriptions about the SIFT descriptor and PCA are discussed in Section 3. Section 4 exploits the framework and methodology of the proposed method. The fusion of matching proximities and heuristics-based cohort selection are presented in Section 5. Evaluation of the proposed technique and comparison with other face recognition systems are exhibited in Section 6. Section 7 computes time complexity. Conclusions and remarks are made in Section 8.

2. Outline of Cloud-Based Face Recognition

To develop a cloud-based face recognition system, a cloud infrastructure [5, 6] has been setup with the help of remote servers, and a webcam-enabled client terminal and tablet PC are connected to the remote servers via Internet connectivity. Two independent IP addresses are provided to both a client machine and a tablet PC. These two IPs help a cloud engine to identify client machines from the point at which the recognition task is performed. Figure 1 shows the outline of the cloud-enabled face recognition infrastructure, in which we establish three points with three different devices for enrollment and recognition tasks. All other software, application modules (i.e., preprocessing, feature extraction, template generation, matching, fusion, and decision), and face databases are placed on servers, and a storage device is maintained in the cloud environment. During authentication or identification, sample face images are captured via cameras that are installed in both the client machine and tablet PC, and the captured faces are sent to a remote server. In the remote server, application software is invoked to perform necessary tasks. After the matching of probe face images with gallery images that are stored in the database, matching proximity is generated and a decision outcome is sent to the client machine over the network. At the client site, the client machine displays the correct decision on the screen, and the entry of malicious users is restricted. Although the proposed system is a cloud-based face recognition system, our main focus lies on a baseline face recognition system. After giving a brief introduction of cloud-based infrastructures for face recognition and use of publicly available face databases, such as FEI and BioID, we assume that face images are already being captured with the sensor installed in the client machine. Further, they are sent to a remote server for matching and decision.

The proposed approach is divided into the following steps.

(a) As part of the baseline face recognition system, the raw face image is localized and then aligned using the algorithm described in [24]. During the experiment, face images are employed with and without localizing the face part.

(b) A histogram equalization technique [25], which is considered the most elementary image-enhancement technique, is applied in this step to enhance the contrast of the face image.

(c) In this step, PCA [11] is applied to obtain multiple face instances, which are determined from each column
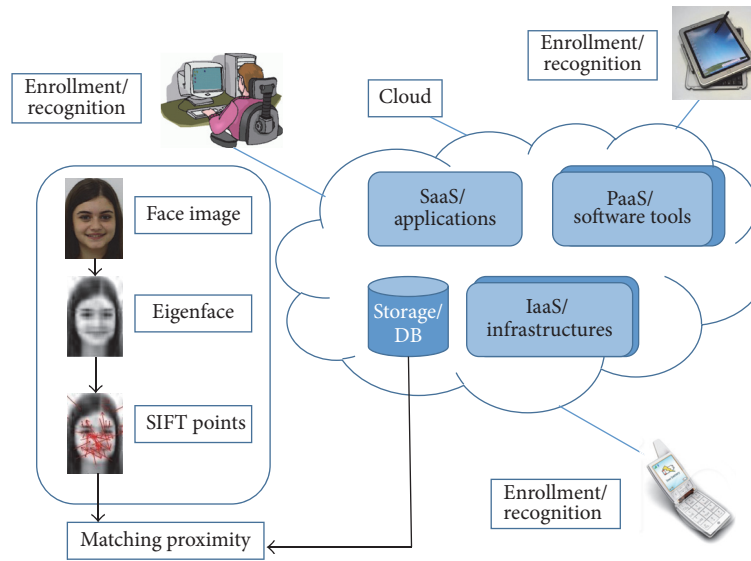
Figure 1: Cloud-enabled face recognition infrastructures.

of the original image (they are not the eigenfaces) by varying principal components from one to six at one distance unit.

(d) From each instance representation, SIFT points [17, 18] are extracted in the scale-space to form an encoded feature vector of keypoint descriptors (K_i) because keypoint descriptors other than spatial location, scale, and orientation are considered feature points.

(e) SIFT interest points that are extracted from the six different face instances (npc: 1, 2, 3, 4, 5, and 6) of a target face form six different feature vectors. They are employed to separately match the feature vectors that are obtained from a probe face. npc refers to the number of principal components.

(f) In this step, matching proximities are determined from the different matching modules and are subsequently fused using "sum" and "max" fusion rules; a decision is made based on the fused matching scores.

(g) To enhance the performance and reduce the computational complexity, we exploit a heuristic-based cohort selection method during matching and apply a T-norm normalization technique to normalize the cohort scores.

3. Brief Review of SIFT Descriptor and PCA

In this section, the scale-invariant feature transform (SIFT) and principal component analysis (PCA) are described. Both the SIFT descriptor and PCA are well-known feature-based and appearance-based techniques that are successfully employed in many face recognition systems.

The SIFT descriptor [17–19] has gained significant attention due to its invariant nature and ability to detect stable interest points around the extrema. It has been proven to be invariant to rotation, scaling, a projective transform, and partial illumination. The SIFT descriptor is robust to image noise and low-level transformations of images. In the proposed approach, the SIFT descriptor can reduce the face matching complexity and computation time while detecting stable interest points on a face image. SIFT points are detected via a four-stage filtering approach, namely, (a) scale-space detection, (b) keypoint localization, (c) orientation assignment, and (d) keypoint descriptor computation. However, keypoint descriptors are employed to generate feature vectors for face matching.

The proposed face matching algorithm aims to produce multiple face (six face instances) representations that are determined from each column of the face image. These face instances exhibit distinctive characteristics that are determined by reducing the dimensions of features that comprise the intensity values. Reduced dimensionality is achieved by applying a simple feature reduction technique, which is known as principal component analysis (PCA). PCA projects a high-dimensional face image onto a low-dimensional feature space, where the face instance of higher variance, such as eigenvectors in the observed variables, is determined. Details on PCA are provided in [11, 12].

4. Framework and Methodology

Main facial feature to frontal face recognition used in the proposed experiment is a set of 128-dimensional vectors of squared patch (region) centred at detected and localized keypoints in multiple scale. This vector describes the local structure around keypoints under computed scale. A keypoint if detected in a uniform region cannot be discriminating because scale or rotational change does not make the point distinguishable from its neighbors. SIFT detected keypoints [17, 18] on frontal face are basically various corner points like corners of lip, corners of eye, nonuniform contour

between nose and cheek, and so forth which exhibit intensity changes in two directions. SIFT detected these keypoints by approximating the Laplacian of Gaussian (LoG) in terms of Difference of Gaussian (DoG). As the Gaussian pyramid builds image at various scales, SIFT extracted keypoints are scale invariant and the computed descriptors remain discriminating from coarse to fine matching. The keypoints could have been detected by Harris corner detection (not scale invariant) or Hessian corner detection method but many of these points detected might not be repeatable under large scale change. Further 128-dimensional feature point descriptor obtained by SIFT feature extraction method is orientation normalized, so rotational invariant. Additionally SIFT feature descriptor is normalized to unit length to reduce the effect of contrast. Maximum value of each dimension of the vector is thresholded to 0.2 and once again normalized to make the vector robust to certain range of irregular illumination.

The proposed face matching method has been developed with the concept of varying number of principal components (npc) (npc: 1, 2, 3, 4, 5, and 6). These variations generate the following improvements, whereas the system computes face instances for matching and recognition.

(a) The projection of a face image onto some face instances facilitates construction of independent face matchers, which can vary their performance, whereas SIFT descriptors are applied for extracting invariant interest points; each instance-based matcher is verified for producing matching proximities.

(b) An individual matcher exhibits its strength to recognize the faces and numbers of SIFT interest points, which are extracted from each face instance, and substantially change from one projected face to another projected face as with the effect of varying the number of principal components.

(c) Its performance as a robust system is rectified when the individual performances are consolidated into a single matcher by fusion of matching scores.

(d) Let ε_i, $i = 1, 2, \ldots, m$, be the m eigenvalues arranged in descending order. Let ε_i be associated with eigenvector e_i (ith principal eigenface in the face space). Then percentage of variance is accounted for by the ith principal component $= (\varepsilon_i / \sum_{i=1}^{m} \varepsilon_i) \times 100$. Generally first few principal components are sufficient to capture more than 95% of variances. But that number of components is dependent on the training set of image space. It varies with the face dataset used. In our experiments we observed that taking as few as only 6 principal components gives a good result and captures the variability which is very close to total variability produced during generation of multiple face instances.

Let training face dataset contain n instances each of uniform size p ($h \times w$ pixels), then face space contains p-dimensional n sample points, and we can derive at most $n - 1$ eigenvectors, but each eigenvector is still p-dimensional ($n \ll p$). Now to compare two face images each containing p number of pixels (i.e., p-dimensional vector) it is required to project each face image onto each of the $n - 1$ eigenvectors (each eigenvector represents one axis of the new $n - 1$ dimensional coordinate system). So from each of the m-dimensional face we derive $n - 1$ scalar values by dot product of mean centred image space face with each of the $n - 1$ face space eigenvectors. Now in the backward direction given $n - 1$ scalar values, we can reconstruct the original face image by the weighted combination of these eigenfaces and adding mean centred data. In this reconstruction process an ith eigenface contributes more than $(i + 1)$th eigenface if they are ordered in decreasing eigenvalues. How accurate the reconstruction is depends on how many principal component (say k, $k = 1, 2, \ldots, n-1$) we take into consideration. Practically it is seen that k need not be equal to $n - 1$ to satisfactorily reconstruct the face. After a specific value of k (say t) contribution from $(t + 1)$th eigenvectors up to $(n - 1)$th vector is so negligible that it may be discarded without losing significant information; indeed there are some methods like Keisar criterion (discard eigenvectors corresponding to eigenvalue less than 1) [11, 12], Scree test, and so forth. Sometimes Keisar method retains too many eigenvectors whereas Scree test retains too few. In essence, the exact value of t is dataset dependent. In Figures 4 and 6 it is clearly shown that when we continue to add one more principal component, the capture of variability is increasing rapidly within first 6 components. But starting from 6th principal component variance capture is almost flat but does not reach the total variability (100% marked line) until last principal component. So despite the small contribution that principal component 7 onwards have, they cannot be redundant.

(e) Distinctive and classified characteristics, which are detected in the reduced low-dimensional face instance, support the integration of local texture information with local shape distortion and illumination changes of neighboring pixels around each keypoint, which comprises a vector of 128 elements of invariant nature.

The proposed methodology is performed with two different perspectives; namely, the first system is implemented without face detection and localization, and the second system is focused on implementing a face matcher with a localized and detected face image.

4.1. Face Matching: Type I. During the initial stage of face recognition, we enhance the contrast of a face image by applying a histogram equalization technique. We apply this contrast enhancement technique to increase the count of SIFT points that are detected in the local scale-space of a face image. The face area is localized and aligned by using the algorithm given in [24]. In the subsequent steps, the face area is projected onto the low-dimensional feature space and an approximated face instance is formed. SIFT keypoints are

FIGURE 2: Six different face instances of a single cropped face image with variations of principal components.

detected and extracted from this approximated face instance, and a feature vector that consists of interest points is created. In this experiment, six different face instances are generated by varying the number of principal components from one to six and they are derived from a single face image.

PCA-characterized face instances are shown in Figure 2; they are arranged according to the order of the considered principal components. The same set of PCA-characterized face instances is extracted from a probe face, and feature vectors that consist of SIFT interest points are formed. Matching is performed with their corresponding face instances in terms of the SIFT points obtained from reference face instances. We apply a k-nearest neighborhood (k-NN) approach [26] to establish correspondence and obtain the number of pairs of matching keypoints. Figure 3 depicts matching pairs of SIFT keypoints on two sets of face instances, which correspond to reference and probe faces. Figure 4 shows the amount of variance captured by all principal components; because the first principal component explains approximately 70% of the variance, we expect that additional components are probably needed. The first four principal components explain the total variability in the face image that is depicted in Figure 2.

4.2. Face Matching: Type II. The second type of face matching strategy utilizes outliers that are available around the meaningful facial area to be recognized. This type of face matching examines the effect of outliers and legitimate features together that are employed for face recognition. Outliers may be located on the forehead above the legitimate and localized face area, around the face area to be considered outside the meaningful area, on both the ears and on the head. However, the effect of outliers is limited because the legitimate interest points are primarily detected in major salient areas, which may be an efficient analysis because face area localization cannot be performed or sometimes outliers are an effective addition to the face matching process.

In the Type I face matching strategy, we employ dimensionality reduction and project an entire face onto a low-dimensional feature space using PCA and construct six different face instances with principal components that vary between one and six. We extract SIFT keypoints from six multiscale face instances and create a set of feature vectors. The face matching task is performed using the k-NN approach, and matching scores are generated as matching proximities from a pair of reference and probe faces. The matching scores are passed through the fusion module and consolidated to form an integrated vector of matching proximities. Figure 5 demonstrates matching between pairs of face instances of an entire face, which corresponds to the reference face and probe

face for a certain number of principal components. Figure 6 shows the amount of variances that were captured by all principal components; because the first principal component explains less than 50% of the variance, we expect that additional components are needed. The first two principal components explain approximately two-thirds of the total variability in the face image depicted in Figure 5.

5. Fusion of Matching Proximities

5.1. Baseline Approach to Fusion. To fuse [26–28] the matching proximities that are computed from all matchers (based on principal components) and form a new vector, we apply two popular fusion rules, namely, "sum" and "max" [26]. Let ms = $[\mathrm{ms}_{ij}]$ be the match scores that are generated by multiple matchers (j), where $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$. Here, n denotes the number of match scores that are generated by each matcher, and m represents the number of matchers that are presented to the face matching process. Consider that the labels ω_0 and ω_1 are two different classes that are referred to as the genuine class and the imposter class, respectively. We can assign ms to either the label ω_0 or the label ω_1 based on the class-conditional probability. The probability of error can be minimized by applying Bayesian decision theory [29] as follows:

$$\text{Assign ms} \longrightarrow \omega_i \text{ if}$$
$$P\left(\omega_i \mid \text{ms}\right) > P\left(\omega_j \mid \text{ms}\right) \tag{1}$$
$$\text{for } i \neq j, \ i, j = 0, 1.$$

The posterior probability $P(\omega_i \mid \text{ms})$ can be derived from the class-conditional density function $p(\text{ms} \mid \omega_i)$ using Bayes formula as follows:

$$P\left(\omega_i \mid \text{ms}\right) = \frac{p\left(\text{ms} \mid \omega_i\right) P\left(\omega_i\right)}{p\left(\text{ms}\right)}. \tag{2}$$

Therefore, $P(\omega_i)$ is the priori probability of the class label ω_i, and $p(\text{ms})$ denotes the probability of encountering ms. Thus, (1) can be rewritten as follows:

$$\text{Assign ms} \longrightarrow \omega_i \text{ if}$$
$$\text{LR} > \tau, \quad \text{where LR} = \frac{p\left(\text{ms} \mid \omega_i\right)}{p\left(\text{ms} \mid \omega_j\right)}, \tag{3}$$
$$i \neq j, \ i, j = 0, 1.$$

The ratio LR is known as the likelihood ratio, and τ is the predefined threshold. The class-conditional density $p(\text{ms} \mid \omega_i)$ can be determined from the training match score vectors using either parametric or nonparametric techniques. However, the class-conditional probability density function can be extended to "sum" and "max" fusion rules. The max fusion can be extended as follows:

$$p\left(\text{ms} \mid \omega_i\right) = \max_{j=1,2\ldots,m; i=0,1} \left(p\left(\text{ms}_j \mid \omega_i\right)\right). \tag{4}$$
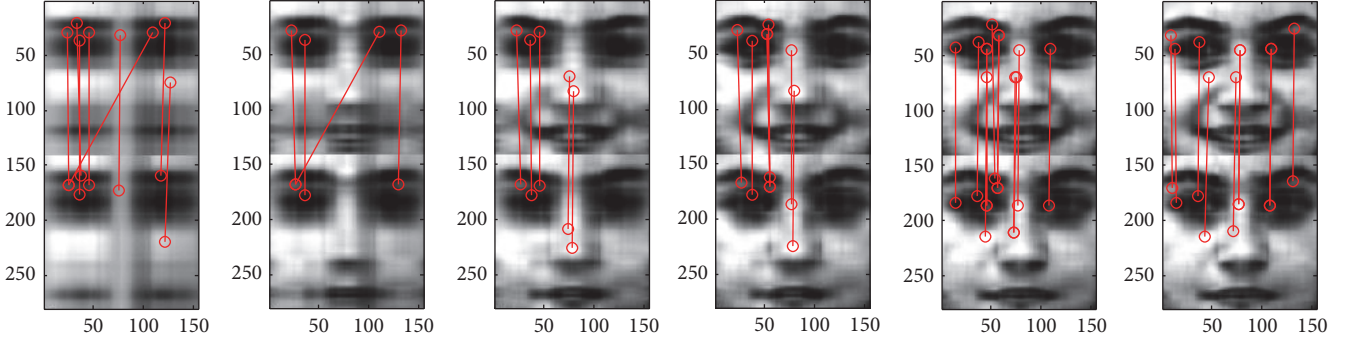
FIGURE 3: Matching pairs of face instances with variations of principal components for a pair of corresponding reference and probe face images.
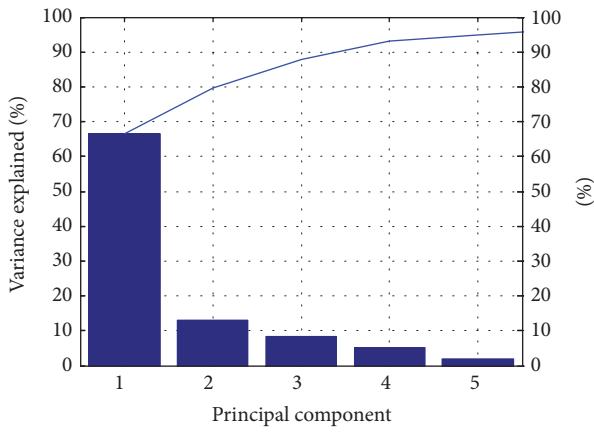

FIGURE 4: Amount of variance accounted by each principal component of the face image, as depicted in Figure 2.

Here, we replace the joint-density function by maximizing the marginal density. The marginal density $p(\mathrm{ms}_j \mid \omega_i)$ for $j = 1, 2, \ldots, m$ and $i = 0, 1$ (i refers to either a genuine sample or an imposter sample) can be estimated from the training vectors of the genuine and imposter scores that correspond to each m matcher. Therefore, we can rewrite (4) as follows:

$$\mathrm{FS}_{\max}^{\omega_{i=0}} = \max_{j=1,2,\ldots,m;i=0,1} \left\{ p\left(\mathrm{ms}_j \mid \omega_{i=0}\right) \right\},$$

$$\mathrm{FS}_{\max}^{\omega_{i=1}} = \max_{j=1,2,\ldots,m;i=0,1} \left\{ p\left(\mathrm{ms}_j \mid \omega_{i=1}\right) \right\}. \tag{5}$$

$\mathrm{FS}_{\max}$ denotes the fused match scores that are obtained by fusing the m matchers in terms of exploiting the maximum scores.

We can easily extend the "max" fusion rule to the "sum" fusion rule by assuming that the posteriori probability does not significantly deviate from the priori probability.

Therefore, we can write an equation for fusing the marginal densities that are known as the "sum" rule as follows:

$$\mathrm{FS}_{\mathrm{sum}}^{\omega_{i=0}} = \sum_{j=1}^{m} p\left(\mathrm{ms}_j \mid \omega_{i=0}\right),$$

$$\mathrm{FS}_{\mathrm{sum}}^{\omega_{i=1}} = \sum_{j=1}^{m} p\left(\mathrm{ms}_j \mid \omega_{i=1}\right). \tag{6}$$

Independently, we apply "max" and "sum" fusion rules to the genuine and imposter scores that correspond to each of six matchers, which are determined from six different face instances for which the principal components vary from one to six. Prior to the fusion of the matching proximities that are produced by multiple matchers, the proximities need to be normalized and the data needs to be mapped to the range of [0-1]. In this case, we use the min-max normalization technique [26] to map the proximities to the specified range, and the T-norm cohort selection technique [30, 31] is applied to improve the performance. To generate matching scores, we apply the k-NN approach [32].

5.2. Cohort Selection Approach to Fusion. Recent studies suggest that cohort selection [30, 31] and cohort-based score normalization [30] can exhibit robust performance and increase the robustness of biometric systems. To understand the usability of cohort selection, a cohort pool is considered. A cohort pool is a set of matching scores obtained from nonmatch templates in the database, whereas a probe sample is matched with the reference samples in the database. The matching process generates matching scores; among this set of scores of the corresponding reference samples, one template is identified as the claimed identity. This claimed identity is the true match to the probe sample, and the matching proximity is significant. In addition to the claimed identity, the remainder of the matched scores are known as cohort scores. We refer to a match score as a true matching proximity, which is determined from the claimed identity. However, the cohort scores and the score that is determined from the claimed identity exhibit similar degradation. To improve the performance of the proposed system, we need to normalize the true matching proximity using the cohort
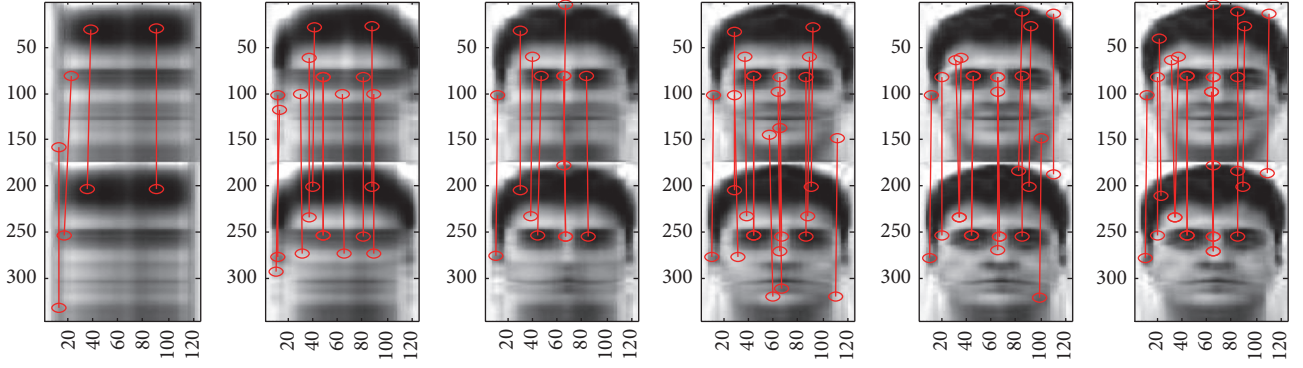
FIGURE 5: Face matching strategy, which shows matching between pairs of face instances with outliers.
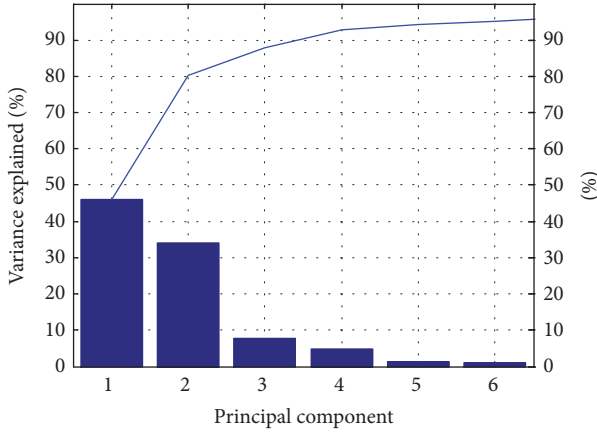

FIGURE 6: Amount of variance accounted by each principal component of the face image depicted in Figure 5.

scores. We can apply simple statistics, such as the mean, standard deviation, and variance, to compute the normalized score of the true reference template using the T-norm cohort normalization technique. We assume that "most similar cohort scores" and "most dissimilar cohort scores" can contribute to computation of the normalized scores, which have more discriminatory information than the normal matching score. As a result, the number of false rejection rates may decrease, and the system can successfully identify a subject from a pool of reference templates.

Two types of probe samples exist: genuine probe samples and imposter probe samples. When a genuine probe face is compared with cohort models, the best-matched cohort model and the few models among the remaining cohort models are expected to be very similar due to the similarity among the corresponding faces. The matching of a genuine probe face with the true cohort model and with the remaining cohort models produced matching scores with the lowest similarity when the true matched template and the remaining of the templates in the database are dissimilar. The comparison of an imposter face with the reference templates in the database can generate matching scores, which are independent of the set of cohort models.

Although cohort-based score normalization is considered extra overhead to the proposed system, it can improve performance. Computational complexity will increase, if the number of comparisons exceeds the number of cohort models. To reduce the overhead of an integrating cohort model, we need to select a subset of cohort models, which contains the majority of discriminating information, and we combine this cohort subset with the true match score to obtain a normalized score. This cohort subset is known as an "ordered cohort subset," which contains the majority of discriminatory information. We can select a cohort subset for each true match template in the database to normalize each true match score when we have a number of probe faces to compare. In this context, we propose a novel cohort subset selection method that utilizes heuristic cohort selection statistics. Because the cohort selection strategy is substantially inspired by heuristic-based T-statistics and a baseline heuristic approach, we refer to this method of hybrid heuristics statistics in which two-stage filtering is performed to generate the majority of discriminating cohort scores.

5.2.1. Methodology: Hybrid Heuristics Cohort. The proposed statistics begin with a cohort score set $\chi^i = \{x_1^i, x_2^i, \ldots, x_n^i\}$, where $i = \{$genuine scores, imposter scores$\}$ and n is the number of cohort scores that consist of the genuine and imposter scores presented in the set χ. Therefore, each score is labeled with $x_j^i \in \{$genuine scores, imposter scores$\}$ and $j \in \{1, 2, \ldots, n\}$ sample scores. From the cohort scores set, we can calculate the mean and standard deviation for the class labels genuine and imposter scores. Let μ^{genuine} and μ^{imposter} be the mean values and let δ^{genuine} and δ^{imposter} be the standard deviations for both class labels. Using T-statistics [33], we can determine a set of correlation scores that correspond to cohort scores:

$$T\left(x_j\right) = \frac{\left|\mu^{\text{genuine}} - \mu^{\text{imposter}}\right|}{\sqrt{\left(\delta_j^{\text{genuine}}\right)/n^{\text{genuine}} + \left(\delta_j^{\text{imposter}}\right)/n^{\text{imposter}}}}. \quad (7)$$

In (7), n^{genuine} and n^{imposter} are the number of cohort scores, which are labeled as genuine and imposter, respectively. We calculate all correlation scores and make a list of all $T(x_j)$

scores. Then, we construct a search space that includes these correlation scores.

Because (7) exhibits a correlation between cohort scores, it can be extended to the baseline heuristics approach in the second stage of the hybrid heuristic-based cohort selection method. The objective of the proposed cohort selection method is to select cohort scores that correspond to the two subsets of highest and lowest correlation scores obtained by (7). These two subsets of cohort scores constitute the cohort subset. We separately collect the correlation scores in a FRINGE data structure or OPEN list; with this initial score, we expand the fringe by adding more correlation scores to the fringe. We also maintain another list, which we refer to as the CLOSED list. After calculating the first $T(x_j)$ score in the fringe, we omit this score from the fringe and expand this score. The next two correlation scores from the search space are removed but retained in the fringe. The correlation score, which was removed from the fringe, is added to the CLOSED list. Because the fringe contains two scores, we arrange them in decreasing order in the fringe by sorting them and removing the maximum score from the fringe. This maximum score is now added to the CLOSED list and maintained in nonincreasing order with other scores in the CLOSED list. We repeat this recursive process in each iteration until the search space is empty. After expanding the search space by moving all correlation scores from the fringe to the CLOSED list, we construct a sorted list. These sorted scores in the CLOSED list are divided into three parts: the first part and last part are merged to create a single list of correlation scores that exhibit the most discriminating features. We establish a cohort subset by determining the most promising cohort scores, which correspond to the correlation scores on the CLOSED list.

To normalize the cohort scores in the cohort subset, we apply the T-norm cohort normalization technique. T-norm describes the property that indicates that the score distribution of each subject class follows a Gaussian distribution. These normalized scores are employed for making decisions and assigning the probe face to one of the two class labels. Prior to making any decisions, we consolidate the normalized scores for six different face models depending on the principal components to be considered, which range between one and six.

6. Experimental Evaluation

The rigorous evaluation of the proposed cloud-enabled face matching technique is conducted with two well-known face databases, namely, FEI [34] and BioID [35]. The face images are presented in the databases with changes in illumination, nonuniform and uniform backgrounds, and facial expressions. For experiments, we have set up a simple protocol of face pair matching and apply two different fusion rules, namely, max fusion rules and sum fusion rules. However, we have implemented the proposed method by considering two perspectives: the Type II perspective indicates that face recognition employs face images, which are provided in the databases without being cropped, and the Type I perspective

Figure 7: Face images from BioID database.

indicates that the face recognition technique uses a manually localized face area after cropping the face images and fixing the size of the face area to 140×140 pixels. The faces of these two face databases are presented with a variety of backgrounds; therefore, a uniform and robust framework should be designed to examine the proposed face matching techniques.

6.1. Databases

6.1.1. BioID Face Database. The face images that are presented in the BioID [35] database are recorded in a degraded environment and are primarily employed for face detection. However, we can also utilize this database for face recognition. Because the faces are captured with a variety of background information and illumination, the evaluation of this database is challenging. Here, we analyze the Type I and Type II face evaluation frameworks. The database consists of 1521 frontal view face images, which were obtained from 23 persons; all faces are gray level images with a resolution of 384×286 pixels. Sample face images from the BioID database are shown in Figure 7. The face images are acquired against a variety of backgrounds, facial expressions, head position changes, and illumination changes.

6.1.2. FEI Face Database. The FEI database [34] is a Brazilian face database of images taken between June 2005 and March 2006. The database consists of 2800 face images of 200 people, who each contributed 14 face images. The faces are captured against a white homogeneous background in an upright frontal position, and all images have scale changes of approximately 10%. Of the 2800 face images, the number of male contributors is equivalent to the number of female contributors; that is, 100 male participants and 100 female participants have contributed an equal number of face images, which total 1400 face images. All images are colorful, and the size of each image is 640×480 pixels. Face images from the FEI database are shown in Figure 8. Faces are acquired against uniform illumination and homogeneous backgrounds with neutral and smiling facial expressions. The database contains faces of varying ages from 18 years to 60 years.

FIGURE 8: Face images from FEI database.

FIGURE 9: Boxplot of principal components versus EER when faces are not cropped and the number of principal components varies between one and six.

6.2. Experimental Protocol. In this experiment, we have developed a uniform framework for examining the proposed face matching technique with the established viability constraints. We assume that all classifiers are mutually random processes. Therefore, to address the biases of each random process, we perform an evaluation with a random distribution of the number of training samples and probe samples. However, distribution is completely dependent on the database that is employed for the evaluation. Because the BioID face database contains 1521 faces of 23 individuals, we equally distribute the face images among the training and test sets. The faces that are contributed by each person are divided into two sets, namely, the training set and the test/probe set. The FEI database contains 2800 face images of 200 people. We devise a protocol for all databases as follows.

Consider that each person has contributed m number of faces and the database size is p (p denotes the total number of face images). We consider that q denotes the total number of subjects/individuals who contributed m number of face images. To extend this protocol, we divide m into two equal groups of face images $m/2$ and retain each group for the training/reference and probe sets. To obtain the genuine and imposter match scores, each face in the training set is compared with $m/2$ number of faces in the probe set, which corresponds to a subject, and each single face is compared with the face images of the remaining subjects. Thus, we obtain genuine match scores of $q \times (m/2)$ dimension and imposter scores of $q \times (q - 1) \times m$ dimension. The k-NN (k-nearest neighbor approach) metric is employed to generate common matching points between a pair of two face images, and we employ a min-max normalization technique to normalize the match scores and map the scores to the range of [0-1]. In this manner, two sets of match scores of unequal dimensions, which correspond to a matcher, are obtained, whereas the face images are compared within intraclass (ω_i) sets and interclass (ω_j) sets, which we refer to as genuine and imposter score sets.

6.3. Experimental Results and Analysis

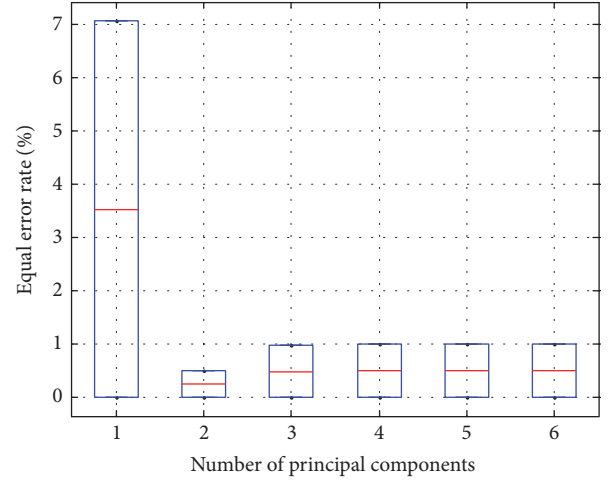*6.3.1. On FEI Database.* The proposed cloud-enabled face recognition system has been evaluated using the FEI face database, which contains neutral and smiling expressions of faces. The neutral faces are utilized as target/training faces, and the smiling faces are used as probe faces. We perform several experiments to analyze the effect of (a) a cloud-enabled environment, (b) face matching without extracting a face area, (c) face matching with extracting a face area, (d) face matching by projecting the face onto a low-dimensional feature space using PCA with varying principal components (one to six) in the conditions mentioned in (a), (b), and (c), and (e) using hybrid heuristic statistics for the cohort subset selection. We depict the experimental results as ROC curves and a boxplot. The ROC curves reveal the performance of the proposed system by GAR versus FAR curves for varying principal components. The boxplot shows how EER varies for different values of principal components.

Figure 9 shows a boxplot when a face with face area has not been extracted, and Figure 12 shows another boxplot when the localized face has been extracted for recognition. In the boxplot in Figure 9, the EER exceeds 7% when the principal component is set to one, and the EER varies between 0% and 1% for the remaining principal components. In the second boxplot, a maximum EER of 10% is attained when the principal component is 1 and the EER varies between 0% and 2.5%. However, an EER of 0.5% is determined for the principal components 2, 3, 4, and 5 after the face area is extracted for recognition. A maximum EER of 1% is attained for the principal components 3, 4, 5, and 6 when face localization was not performed. As shown in Table 1, face recognition performance deteriorates only for the case when the principal component is set to one. For the remaining cases when the principal component varies between two and six, however, low EERs are obtained; a maximum EER is obtained when the principal component is set to four. However, the ROC curves in Figure 10 exhibit higher recognition accuracies for all cases, with the exception of the case when the principal component is set to 1. ROC curves are shown when face images are not localized. Thus, we decouple the information about the
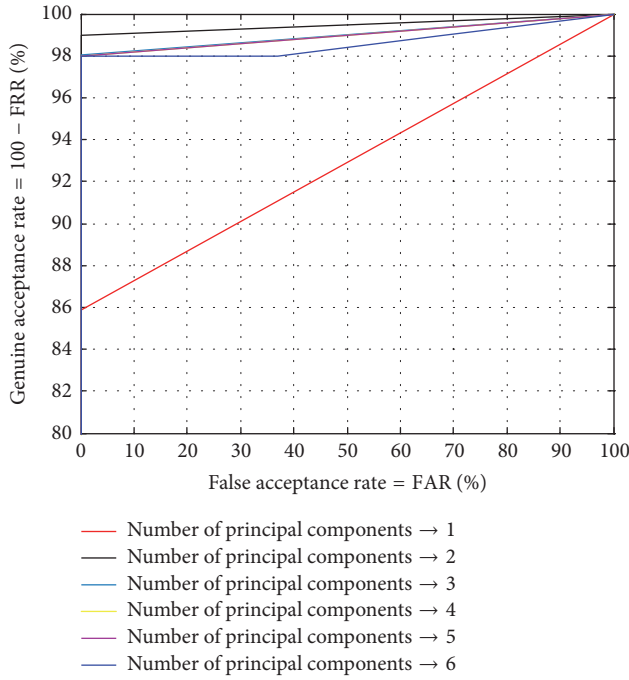
FIGURE 10: Receiver operating characteristics (ROC) curve of GAR versus FAR is shown when faces are not cropped and the number of principal components varies between one and six.

TABLE 1: EER and recognition accuracies of the proposed face matching strategy on FEI database when face localization is not performed and the number of principal components varies between one and six.

Face matching strategies with varying number of principal components	EER (%)	Recognition accuracy (%)
Principal component 1	7.07	92.93
Principal component 2	0.5	99.5
Principal component 3	0.97	99.03
Principal component 4	1.01	98.99
Principal component 5	1	99
Principal component 6	1	99

legitimate face area with explicit information about nonface areas, such as the forehead, hair, ear, and chin areas. These areas may provide crucial information, which is considered additional information in a decoupled feature vector. These feature points contribute to recognizing faces with varying principal components. Figure 11 shows recognition accuracies for different values of principal component determined on FEI database while face localization does not perform. On the other hand, Figure 14 shows recognition accuracies for different values of principal component determined on FEI database while face localization does perform.

Figure 13 shows the ROC curves that are determined from extensive experiments of the proposed algorithm on the FEI face database when a face image is localized and only the face part is extracted. In this case, a maximum EER
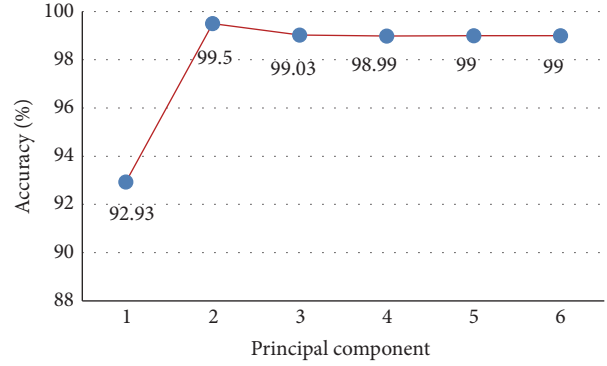

FIGURE 11: Recognition accuracy versus principal component curve when face localization task is not performed.
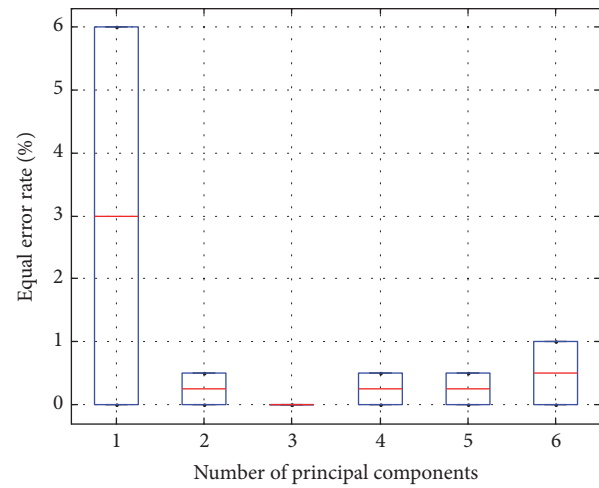

FIGURE 12: Boxplot of principal components versus EER when faces are cropped and the number of principal components varies between one and six.

of 6% is attained when the principal component is set to 1; in other cases, the EERs are as low as the EERs in cases with nonlocalized faces. Table 2 depicts the efficacy of the proposed face matching strategy as recognition accuracies and EERs. With the exception of principal component 1, all remaining cases have shown tremendous improvements over the results listed in Table 1. By integrating feature-based and appearance-based approaches, we have made the algorithm robust not only for facial expressions but also for the areas that correspond to major salient face regions (both eyes, nose, and mouth), which have a significant impact on face recognition performance. The ROC curves in Figure 13 also show that the algorithm is inaccurate when the principal components vary between two and six, even in the case of principal component 3, when an EER and recognition accuracy of 0% and 100%, respectively, are attained.

Based on two major considerations—with and without face localizations—we investigate the proposed algorithm by fusing face instances under a varying number of principal components. To demonstrate the robustness of the system, we apply two fusion rules—the "max" fusion rule and the
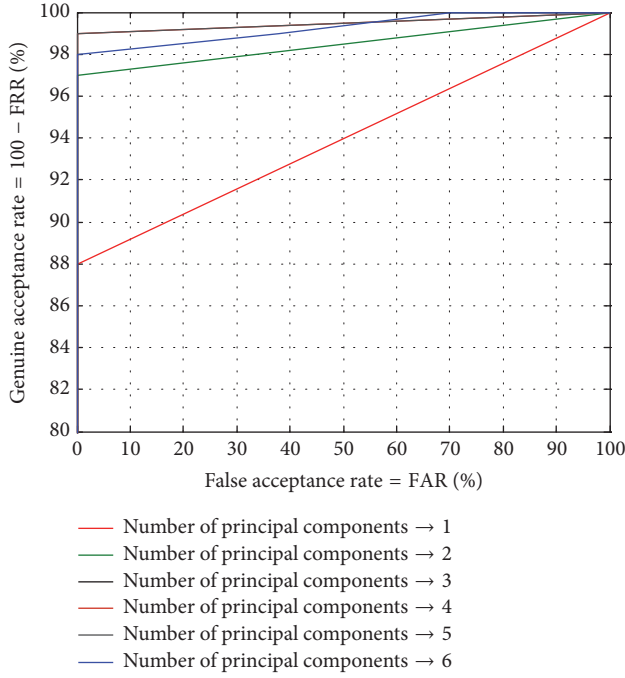
FIGURE 13: Receiver operating characteristics (ROC) curve of GAR versus FAR when faces are cropped and the number of principal components varies between one and six.

TABLE 2: EER and recognition accuracies of the proposed face matching strategy on the FEI database when face localization is performed and the number of principal components varies between one and six.

Face matching strategies with varying number of principal components	EER (%)	Recognition accuracy (%)
Principal component 1	6	94
Principal component 2	0.5	99.5
Principal component 3	0	100
Principal component 4	0.5	99.5
Principal component 5	0.5	99.5
Principal component 6	1	99.0

"sum" fusion rule. However, the effect of localizing the face area and not localizing the face area is investigated, and the conventions with these two fusion rules are integrated. Figure 15 shows the ROC curves that are obtained by fusing the face instances of principal components 1 to 6 without performing face localization, and the matching scores are generated from a single fused classifier, in which all six classifiers are fused in terms of matching scores by applying the "max" and "sum" fusion rules. When we apply the "sum" fusion rule, we obtain 100% recognition accuracy, whereas 98.5% accuracy is obtained when the "max" fusion rule is applied. In this case, the "sum" fusion rule outperforms the "max" fusion rule. When a hybrid heuristic statistics-based cohort selection method is applied to fusion rules for a fusion-based classifier, the "sum" and "max" fusion rules
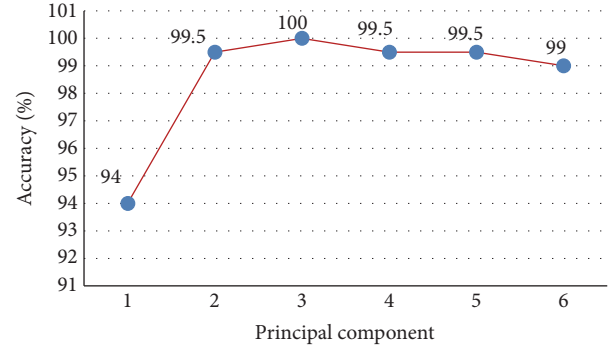

FIGURE 14: Recognition accuracy versus principal component curve when face localization task is performed.

TABLE 3: The table lists EER and recognition accuracies of the proposed face matching strategy on the FEI database when face localization is not performed and principal components are fused using the sum and max fusion rules to form a single set of matching scores. In addition, the hybrid heuristic statistics-based cohort selection technique is applied, which uses T-norm normalization techniques for match score normalization.

Fusion rule/normalization/heuristic cohort selection	EER (%)	Recognition accuracy (%)
Sum fusion rule + min-max normalization	0.0	100
Max fusion rule + min-max normalization	1.5	98.5
Sum fusion rule + T-norm (hybrid heuristic)	0.5	99.5
Max fusion rule + T-norm (hybrid heuristic)	0.5	99.5

achieve 99.5% recognition accuracy. In the general context, the proposed cohort selection method degrades recognition accuracy by 0.5% when it is compared with the "sum" fusion rule-based classifier without applying the cohort selection method. However, hybrid heuristic statistics render the face matching algorithm stable and consistent for both the fusion rules (sum, max) with 99.5% recognition accuracy. In Table 3 the recognition accuracies are shown and in Figure 16 the same has been exhibited for "sum" and "max" fusion rules.

After evaluating the performance of each of the classifiers, in which the face instance is determined by setting the value of the principal component to one value among 1 to 6 and the fusion of face instances without face localization, we evaluate the efficacy of the proposed algorithm by fusing all six face instances in terms of the matching scores obtained from each classifier. In this case, the face image is manually localized and the face area is extracted for the recognition task. Similar to the previous approach, we apply two fusion rules to integrate the matching scores, namely, "sum" and "max" fusion rules. In addition, we exploit the proposed statistical cohort selection technique, which is known as hybrid heuristic statistics. For cohort score normalization, we apply the T-norm normalization technique. This technique
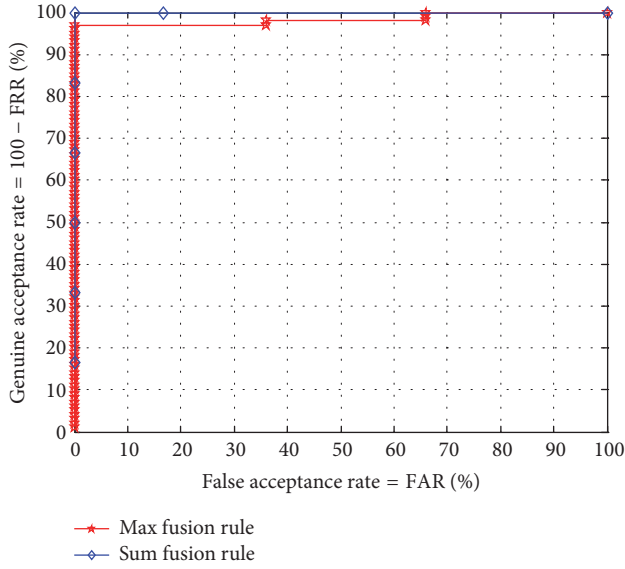
FIGURE 15: Receiver operating characteristics (ROC) curve of GAR versus FAR for two fusion rules, namely, the sum fusion rule and the max fusion rule, when faces are not cropped and the number of principal components varies between 1 and 6.
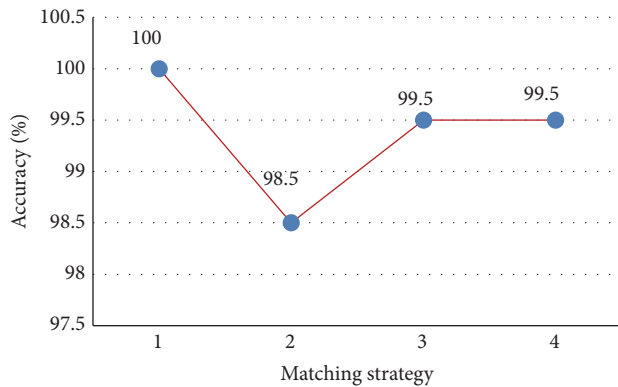

FIGURE 16: Curve of recognition accuracy versus face matching strategy when face instances are fused in terms of matching scores by applying max and sum fusion rules with and without applying hybrid heuristic statistics for the cohort subset selection. The first two cases show the recognition accuracies without applying the cohort selection method, and the last two cases show the recognition accuracies when the cohort selection method is applied. These accuracies are obtained without face localization.

maps the cohort scores into a normalized score set that exhibits the characteristics of each of the scores in the cohort subset and enables the correct match to be rapidly obtained by the system. As shown in Table 4 and Figures 17 and 18, when the "sum" and "max" fusion rules are applied with the min-max normalization technique to the fused match scores, we obtain 100% recognition accuracy. In the next step, we exploit the hybrid heuristic-based cohort selection technique and achieve 100% recognition accuracy in both cases, when the "sum" and "max" fusion rules are applied. The effect of face localization serves a central role in increasing the

TABLE 4: EER and recognition accuracies of the proposed face matching strategy on the FEI database when face localization is performed and the principal components are fused using the sum and max fusion rules to form a single set of matching scores. In addition, the hybrid heuristic statistics-based cohort selection technique, which uses T-norm normalization techniques for match score normalization, is applied.

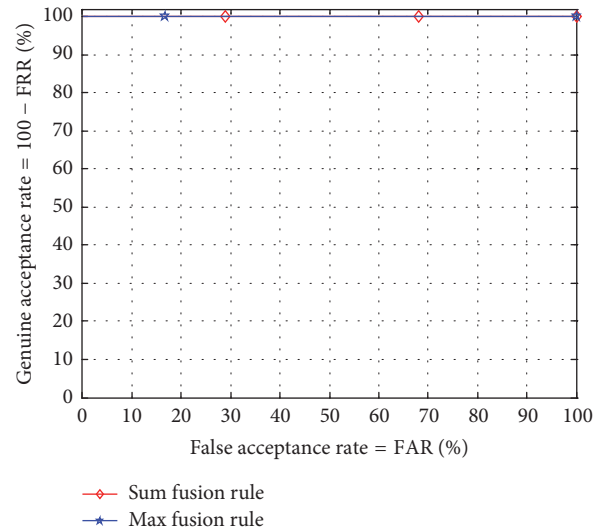| Fusion rule/normalization/heuristic cohort selection	EER (%)	Recognition accuracy (%)
Sum fusion rule + min-max normalization	0.0	100
Max fusion rule + min-max normalization	0.0	100
Sum fusion rule + T-norm (hybrid heuristic)	0.0	100
Max fusion rule + T-norm (hybrid heuristic)	0.0	100

FIGURE 17: Receiver operating characteristics (ROC) curve of GAR versus FAR for two fusion rules, namely, the sum fusion rule and the max fusion rule, when faces are cropped and the number of principal components varies between one and six.

recognition accuracy to cent percent in four cases. Due to face localization, however, the numbers of feature points that are extracted from face images are determined to be dissimilar to the nonlocalized face that is obtained by applying the cohort selection method. These accuracies are obtained after a face is localized.

6.3.2. BioID Database. In this section, we evaluate the performance of the proposed face matching strategy for the BioID face database considering two constraints. Considering the first constraint, the face matching strategy is applied when faces are not localized, and recognition performance is measured in terms of the number of probe faces that are successfully recognized. However, the faces that are provided in the BioID face database are captured in a variety
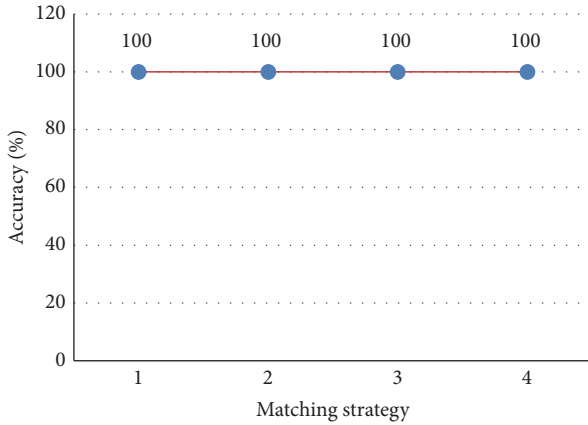
FIGURE 18: Curve of recognition accuracy versus face matching strategy when face instances are fused in terms of matching scores by applying the maximum and sum fusion rules with and without applying hybrid heuristic statistics for the cohort subset selection. The first two cases show the recognition accuracies without applying the cohort selection method, and the last two cases show the recognition accuracies that are obtained by applying the cohort selection method. These accuracies are obtained after a face is localized.
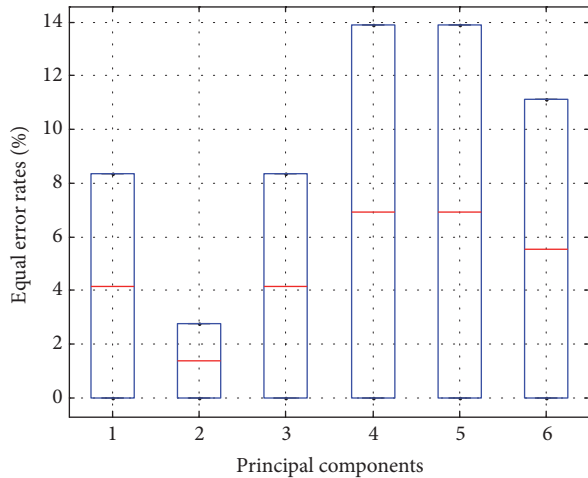

FIGURE 19: Boxplot of principal components versus EER when faces are not cropped and the number of principal components varies between one and six.
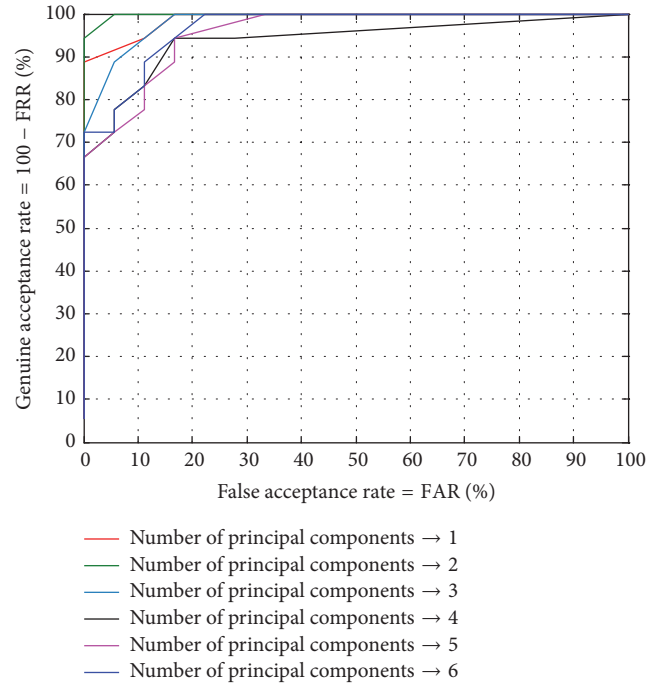

FIGURE 20: Receiver operating characteristics (ROC) curve of GAR versus FAR when faces are not cropped and the number of principal components varies between one and six.

TABLE 5: EER and recognition accuracies of the proposed face matching strategy for the BioID database when face localization is not performed and the number of principal components varies between one and six.

Face matching strategies with varying number of principal components	EER (%)	Recognition accuracy (%)
Principal component 1	8.53	91.47
Principal component 2	2.78	97.22
Principal component 3	8.5	91.5
Principal component 4	13.89	86.11
Principal component 5	13.89	86.11
Principal component 6	11.12	88.88

of environments and illumination conditions. In addition, the faces in the database show various facial expressions. Therefore, evaluating the performance of any face matching is challenging because the positions and locations of frontal view images may be tracked in a variety of environments with changes in illumination. Thus, we need a robust technique that has the capability of capturing and processing all types of distinct features and yields encouraging results in these environments and variable lighting conditions. Face images from the BioID database reflect these characteristics with a variety of background information and illumination changes.

As shown in Figures 19 and 21, the recognition accuracies significantly vary for the principal component range of one to six. For principal component 2, the proposed face matching paradigm yields a recognition accuracy of 97.22%, which is the highest accuracy achieved when the Type II constraint is validated for not localizing the face image. An EER of 2.78%, which is the lowest among all six EERs, is obtained. For principal components 4 and 5, we achieve an EER and recognition accuracy of 13.89% and 86.11%, respectively. Table 5 lists the EERs and recognition accuracies for all six principal components, and Figure 21 shows the same phenomena, which depicts a curve with some points that denote the recognition accuracies that correspond to principal components between one and six. ROC curves determined on BioID database are shown in Figure 20 for unlocalized face.
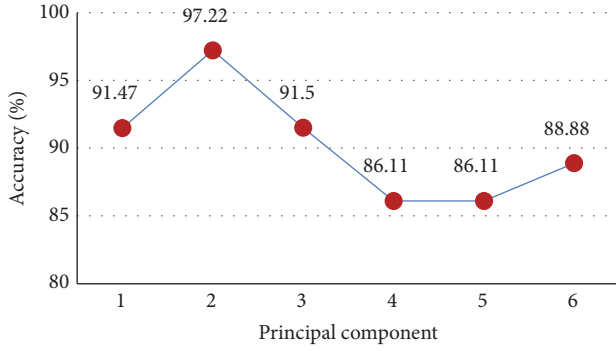
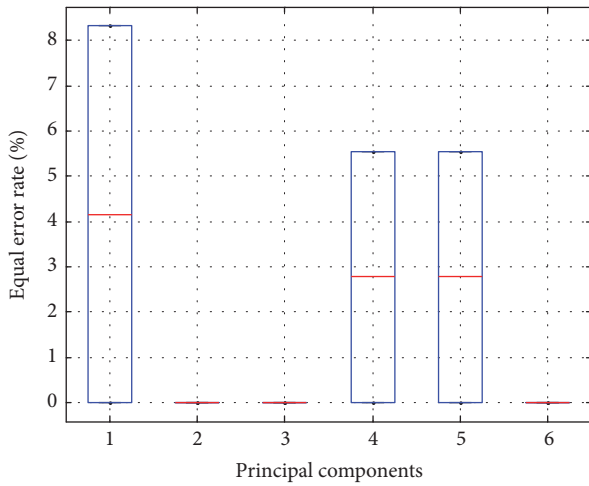FIGURE 21: Recognition accuracy versus principal component when face localization task is not performed.

FIGURE 22: Boxplot of principal components versus EER when faces are localized and the number of principal components varies between one and six.

After considering the Type I constraint, we consider the Type II constraint, in which we obtain the localized face on which the proposed algorithm is applied. Because the face area is localized and localization is primarily performed on degraded face images, we may achieve better results with the Type I constraint.

As shown in Figures 22 and 23, the principal component varies between one and six, whereas the recognition accuracy varies with much better results. However, an EER of 8.5% is obtained for principal component 1, and EERs of 5.59% and 5.98% are obtained for principal components 4 and 5, respectively. For the remainder of the principal components (2, 3, and 6) we achieve a recognition accuracy of 100% in recognizing faces. The ROC curves in Figure 23 show the genuine acceptance rates (GARs) for varying number of principal components and different false acceptance rates (FARs). The principal components (2, 3, and 6) for which the recognition accuracy outperforms other components yield a recognition accuracy of 100%. Figure 24 depicts the recognition accuracies that are obtained for a varying number of principal components. The points on the curve that are
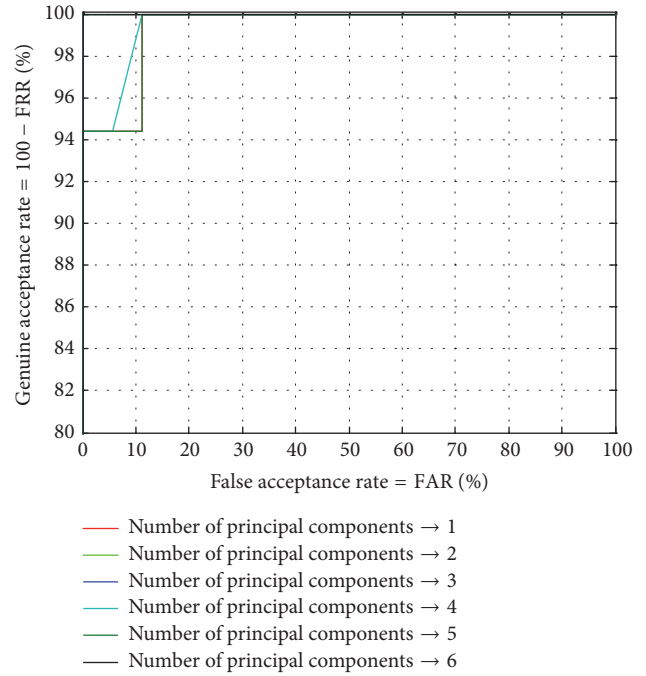

FIGURE 23: Receiver operating characteristics (ROC) curve of GAR versus FAR when faces are cropped and the number of principal components varies between one and six.

TABLE 6: EER and recognition accuracies of the proposed face matching strategy for the BioID database when face localization is performed and the number of principal components varies between one and six.

Face matching strategies with varying number of principal components	EER (%)	Recognition accuracy (%)
Principal component 1	8.5	91.5
Principal component 2	0	100
Principal component 3	0	100
Principal component 4	5.59	94.41
Principal component 5	5.98	94.02
Principal component 6	0	100

marked in red represent the recognition accuracies. Table 6 shows the recognition accuracies when localized face is used.

To validate the Type II constraint for fusions and hybrid heuristics-based cohort selection, we evaluate the performance of the proposed technique by analyzing the effect of a face that is not localized. In this experiment, however, we exploit the same fusion rules, namely, the "sum" and "max" fusion rules, as they are introduced to the FEI database. We also exploit the hybrid heuristics-based cohort selection and the fusion rules. Considering the Type II constraint results, which are obtained on the BioID database, is satisfactory when the fusion rules and cohort selection technique are applied to nonlocalized faces. As shown in Table 7 and from Figure 25, it has been observed that for the first two types of matching strategies, in which the "sum" and
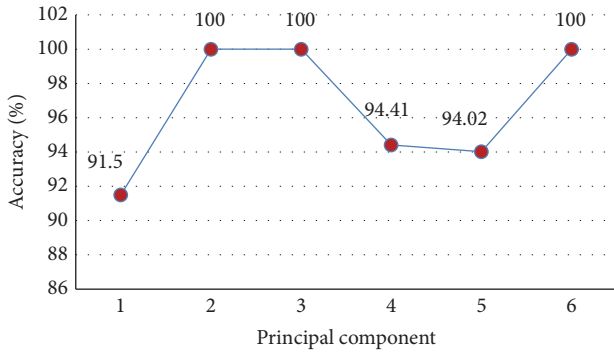
FIGURE 24: Recognition accuracy versus principal component when face localization task is performed.

TABLE 7: EER and recognition accuracies of the proposed face matching strategy for the BioID face database when face localization is performed and the principal components are fused using the sum and the max fusion rules to form a single set of matching scores. In addition, a hybrid heuristic statistics-based cohort selection technique, which employs T-norm normalization techniques for match score normalization, is applied.

Fusion rule/normalization/heuristic cohort selection	EER (%)	Recognition accuracy (%)
Sum fusion rule + min-max normalization	5.55	94.45
Max fusion rule + min-max normalization	0	100
Sum fusion rule + T-norm (hybrid heuristic)	0.5	99.5
Max fusion rule + T-norm (hybrid heuristic)	0	100

"max" fusion rules are applied to fuse the six classifiers, we can achieve a 94.45% recognition accuracy and a 100% recognition accuracy, respectively, whereas EERs of 5.55% and 0%, respectively, are obtained. In this case, the "max" fusion rule outperforms the "sum" fusion rule, which is further illustrated in the ROC curves shown in Figure 26. We achieve 99.5% and 100% recognition accuracies for the next two matching strategies when hybrid heuristics-based cohort selection is applied with the "sum" and "max" fusion rules.

In the last segment of the experiment, we observe the performance to be measured in terms of recognition accuracy and EER when faces are localized. However, the matching paradigms that are listed in Table 7 have also been verified against the Type II constraint. As shown in Table 8 and Figure 27, the first two matching techniques, which employ the "sum" and "max" fusion rules, attained an accuracy of 100%, whereas cohort-based matching techniques show abrupt performance by achieving 99.35% and 100% recognition accuracies. However, a minimal change in accuracy of 0.15% for the combination of the "sum" fusion rule and hybrid heuristics is unreasonable, and the remaining combination of the "max" fusion rule and the hybrid heuristics-based cohort selection method achieved an accuracy of 100%. Therefore, we conclude that the "max" fusion rule outperforms the
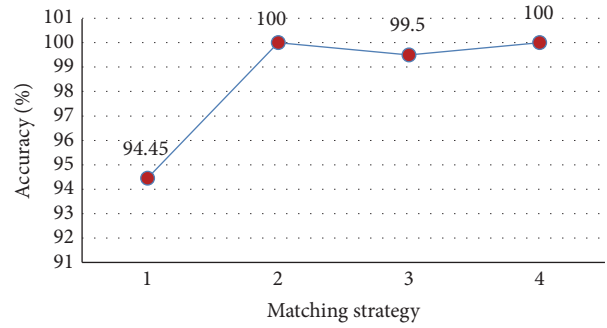

FIGURE 25: Curve of recognition accuracy versus face matching strategy when face instances are fused in terms of matching scores by applying the max and sum fusion rules with and without applying hybrid heuristic statistics to the cohort subset selection. The first two cases show recognition accuracies without applying the cohort selection method, and the last two cases show recognition accuracies that are obtained by applying the cohort selection method. These accuracies are obtained when a face is not localized.
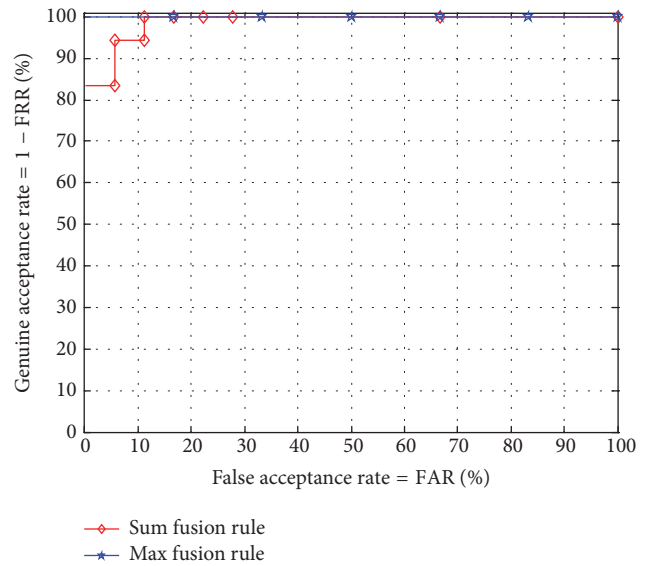

FIGURE 26: Receiver operating characteristics (ROC) curve of GAR versus FAR for two fusion rules, namely, the sum fusion rule and the max fusion rule, when faces are not cropped and the number of principal components varies between one and six.

"sum" rule, which is attributed to a change in the produced cohort subset, for both types of constraints (Type I and Type II). In Figure 28, the recognition accuracies are plotted against the matching strategies, and the accuracy points are marked in blue.

It would be interesting to see how the current ensemble framework would be useful for face recognition in the wild while faces are found in unrestricted conditions. Face recognition in the wild is challenging due to its nature of face acquisition method in unconstrained environment. All the face images are not frontal. Images of the same subject may vary in pose, profile, occlusion, multiple background face, color, and so forth. As our framework is based on only

TABLE 8: EERs and recognition accuracies of the proposed face matching strategy for the BioID database when face localization is performed and the principal components are fused using the sum and the max fusion rules to form a single set of matching scores. In addition, the hybrid heuristic statistics-based cohort selection technique, which employs T-norm normalization techniques for match score normalization, is applied.

Fusion rule/normalization/heuristic cohort selection	EER (%)	Recognition accuracy (%)
Sum fusion rule + min-max normalization	0	100
Max fusion rule + min-max normalization	0	100
Sum fusion rule + T-norm (hybrid heuristic)	0.65	99.35
Max fusion rule + T-norm (hybrid heuristic)	0	100

FIGURE 28: Curve of recognition accuracy versus face matching strategy when face instances are fused in terms of matching scores by applying the maximum and sum fusion rules with and without applying hybrid heuristic statistics for the cohort subset selection. The first two cases show recognition accuracies without applying the cohort selection method, and the last two cases show recognition accuracies that are obtained by applying the cohort selection method. These accuracies are obtained after a face is localized.
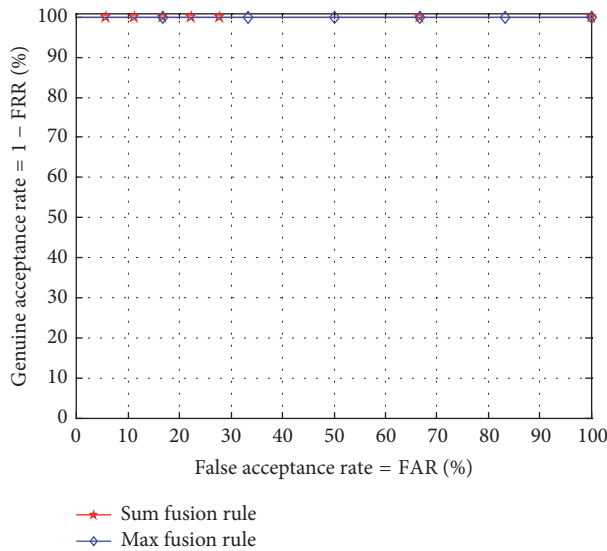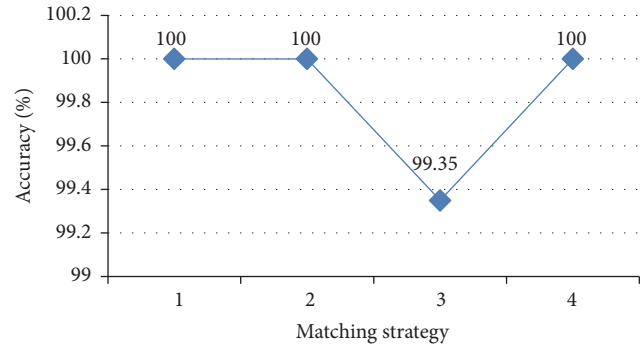

FIGURE 27: Receiver operating characteristics (ROC) curve of GAR versus FAR for two fusion rules, namely, sum fusion rule and max fusion rule, when faces are cropped and the number of principal components varies between one and six.

first six principal components and SIFT features, it requires incorporation of various tools. Tools to detect and crop face region discard background images as far as possible, and detected face regions are to be upsampled or downsampled to make uniform dimensional face image vector to apply PCA. Tools to estimate pose and then apply 2D frontalization to compensate variance lead to less number of principal components to consider.

6.4. Comparison with Other Face Recognition Systems. This section reports a comparative study of the experimental results of the proposed cloud-enabled face recognition protomodel with other well-known face recognition models. These models include some cloud computing-based face

recognition algorithms, which are limited in number, and some traditional face recognition systems, which are not enabled with cloud computing infrastructures. Two different perspectives are considered for which comparisons are performed. The first perspective applies the concept of cloud computing facilities, which are integrated with a face recognition system, whereas the second perspective employs similar face databases to compare with other methods. However, the second perspective does not utilize the concept of cloud computing. To compare the experimental results, we compare the proposed systems with two cloud-based face recognition systems: the first system utilizes eigenface in cloud vision [36], and the second face recognition system utilizes social media with mobile cloud computing facilities [37]. Because cloud-based face recognition models are limited, we presented the results of only these two systems. The system described in [36] employs the ORL face database, which contains 400 face images of 40 individuals, whereas the other system [37] employs a local face database that contains approximately 50 face images. Table 9 shows the performance of the proposed system and the two systems that are exploited in [36, 37] in terms of recognition accuracy. Table 9 also lists the number of training samples that are employed during the matching of face images by different systems. Because the proposed system utilizes two well-known face databases, namely, BioID and FEI databases, and two different face matching paradigms, namely, Type I and Type II, the best recognition accuracies are selected for comparison. The Type I paradigm refers to the face matching strategy, in which face images are localized, and the Type II paradigm refers to the matching strategy in which face images are not localized. Table 9 also shows the results of two face recognition systems [38, 39] that do not employ cloud-enabled infrastructure. The comparative study indicates that the proposed system outperforms other methods, regardless of whether they are cloud-based systems or do not use cloud infrastructures.

Method	Face database	Number of face images	Number of distinct subjects	Recognition accuracy (%)	Error (%)
Cloud vision [36]	ORL	400	40	97.08	2.92
Mobile cloud [37]	Local DB	50	5	85	15
Facial features [38]	BioID	1521	23	92.35	7.65
2D-PCA + PSO-SVM [39]	BioID	1521	23	95.65	4.35
Proposed: Type I	BioID	1521	23	100	0
Proposed: Type I	FEI	2800	20	100	0
Proposed: Type II	BioID	1521	23	97.22	2.78
Proposed: Type II	FEI	2800	20	99.5	0.5
Proposed: Type I + fusion	BioID	1521	23	100	0
Proposed: Type I + fusion	FEI	2800	20	100	0
Proposed: Type II + fusion	BioID	1521	23	100	0
Proposed: Type II + fusion	FEI	2800	20	100	0

7. Time Complexity of Ensemble Network

The time complexity of the proposed ensemble network quantifies the amount of time taken collectively by different modules to run as a set of functions of the length of the input. The time complexity is estimated by counting the number of operations performed by different modules or cascaded algorithms. The ensemble network involves a few cascaded algorithms which together perform face recognition in cloud environment and the algorithms are PCA computation, SIFT features extraction from each face instances, matching, fusion of matching scores using "sum" and "max" fusion rules, and heuristic-based cohort selection. In this section, time complexity of individual modules is computed first and then overall complexity of the ensemble network is computed by summing them together.

(a) Time Complexity of PCA Computation. For PCA algorithm, the computation bottleneck is to derive covariance matrix. Let d be the number of pixels (height × width) determined from each grayscale face image and let the number of face images be N. PCA computation has the following steps.

(i) Finding mean of N sample is $O(Nd)$ (N addition of d dimensional vectors and then summation is divided by N).

(ii) As covariance matrix is symmetric, deriving only upper triangular matrix elements is sufficient. So for each of the $d(d + 1)/2$ elements N multiplications and additions are required leading to $O(Nd^2)$ time complexity. Let the $d \times N$ dimensional covariance matrix be M.

(iii) If Karhunen-Loeve (KL) trick is employed then instead of MM^T ($d \times d$ dimension) compute $M^T M$ ($N \times N$) which requires d multiplications and d additions for each of the N^2 elements, hence $O(N^2 d)$ time complexity (generally $N \ll d$).

(iv) Eigen decomposition of $M^T M$ matrix by SVD method requires $O(n^3)$.

(v) Sorting the eigenvectors (each eigenvector $d \times 1$ dimensional) in descending order of the eigenvalues requires $O(N^2)$. Then taking only first 6 principal component eigenvectors requires constant time.

Projecting the probe image vector on each of the eigenfaces requires a dot product between two vectors resulting in scalar value. Hence d^2 multiplication and d^2 addition for each projection lead to $O(d^2)$. Six such projections require $O(6d^2)$.

(b) Time Complexity of SIFT Keypoints Extraction. Let the dimension of each face image be $M \times N$ pixels and let the face image be represented in column vector of dimension $d(M \times N)$. Gaussian kernel of dimension $w \times w$ is used. Each octave has $s + 3$ scales and total L number of octaves has been used. The important phases of SIFT are as follows.

(i) Extrema detection.

(a) Compute scale.

(1) In each scale, w^2 multiplication and $w^2 - 1$ addition is done by convolution operation for each pixel, so $O(dw^2)$.
(2) $s + 3$ scales are in each octave, so $O(dw^2(s + 3))$.
(3) L is number of octaves, so, $O(Ldw^2(s + 3))$.
(4) Overall $O(dw^2 s)$ is found.

(b) Compute $s + 2$ number of Difference of Gaussians (DoG) for each of the L number of octaves: $O((s+2)xd/2^k)$, $k = 2^{1/s}$, so for L octaves $O(sd)$.
(c) Extrema detection: $O(sd)$.

(ii) Keypoint localization: after eliminating low contrast point and points along edge let αd be number of pixels survived, so $O(\alpha ds)$.

(iii) Orientation assignment: $O(sd)$.

(iv) Keypoint descriptor computation: if $p \times p$ neighborhood of the keypoint is considered, then $O(p^2 d)$ is found.

(c) Time Complexity of Matching. Each keypoint is represented by a feature descriptor of 128 elements. To compare any two such points by Euclidean distance requires 128 subtractions: square of each of the previous 128 subtracted elements, 127 additions to add them up, and one square root at the end. So linear time complexity is $O(n)$, where $n = 128$. Let ith eigenface of probe face image and reference face image have k_1 and k_2 number of survived keypoints. So each keypoint from k_1 will be compared to each of k_2 keypoints by Euclidean distance. So $O(k_1 k_2)$ for a single eigenface pair is $O(6n^2)$ for 6 pairs of eigenfaces (n = number of keypoints). If there are M numbers of reference faces in the gallery then total complexity would be $O(6Mn^2)$.

(d) Time Complexity of Fusion. As the six individual matchers domains are different, therefore, to bring them into uniform domain, min-max normalization technique is used. For each normalized value, computation has two subtraction operations and one division operation. So, it requires constant time $O(1)$. The n pair of probe and gallery images in ith principal component requires $O(n)$. So, for six principal components, it requires $O(6n)$. Finally, the sum fusion requires five summations for each pair of probe and reference face. So, it is a constant time $O(1)$. Subsequently, for n pairs of probe and reference face images, it requires $O(n)$.

(e) Time Complexity of Cohort Selection. Cohort selection requires four operations to be performed: computation of correlation in the search space, insertion of correlation values into the OPEN list, insertion of correlation values at the proper positions into the CLOSED list according to insertion sort, and, finally, division of the CLOSED list of correlation values of size n into three disjoint sets. First two operations take constant time and the third operation takes $O(n^2)$ complexity as it follows the convention of insertion sort. Finally, the last operation takes linear time. Overall time required by cohort selection is $O(n^2)$.

Now, the overall time complexity (T) of ensemble network would be calculated as follows:

$$T = \max \left(\max \left(T_1 \right) + \max \left(T_2 \right) + \max \left(T_3 \right) + \max \left(T_4 \right) \right.$$
$$\left. + \max \left(T_5 \right) \right) = \max \left(O \left(n^3 \right) + O \left(dw^2 s \right) \right. \tag{8}$$
$$\left. + O \left(Mn^2 \right) + O \left(n \right) + O \left(n^2 \right) \right) \approx O \left(n^3 \right),$$

where

T_1 is time complexity of PCA computation,

T_2 is time complexity of SIFT keypoints extractions,

T_3 is time complexity of matching,

T_4 is time complexity of fusion,

T_5 is time complexity of cohort selection.

Therefore, the overall time complexity of ensemble network would be $O(n^3)$ while both enrollment time and verification time through cloud network is assumed to be constant.

8. Conclusion

In this paper, a robust and efficient cloud engine-enabled face recognition system, in which cloud infrastructure has been successfully integrated with a face recognition system, has been proposed. The face recognition system utilizes a baseline methodology, in which face instances are computed by applying a principal component analysis- (PCA-) based texture analysis method to establish six fixed points of principal components, which range from one to six. The SIFT operator is applied to extract a set of invariant points from each face instance, which correspond to the gallery and probe face images. In this methodology, two types of constraints are employed to validate the proposed matching technique: the Type I constraint and the Type II constraint, which denote face matching with face localization and face matching without face localization, respectively. The k-NN method is employed to compute a pair of faces and generate matching points as matching scores. We investigate and analyze various effects on the face recognition system, which directly or indirectly attempt to improve the total performance of the recognition system in various dimensions. To achieve robust performance, we have analyzed the following effects: (a) effect of cloud environment, (b) effect of combining a texture-based method and a feature-based method, (c) effect of using match score level fusion rules, and (d) effect of using a hybrid heuristics-based cohort selection method. After investigating these aspects, we have determined that these crucial and necessary paradigms render the system significantly more efficient than the baseline methodology, whereas remote recognition is achieved either from a remotely placed computer terminal or mobile phones or tablet PCs. In addition, a cloud-based environment reduces the cost required by organizations who would like to implement this integrated system. The experimental results demonstrate high accuracies and low EERs for the types of paradigms that we presented. In addition, the proposed method outperforms other methods.

Competing Interests

The authors declare that they have no competing interests.

References

[1] S. Z. Li and A. K. Jain, Eds., *Handbook of Face Recognition*, Springer, Berlin, Germany, 2nd edition, 2011.

[2] H. Wechsler, *Reliable Face Recognition Methods: System Design, Implementation and Evaluation*, Springer, New York, NY, USA, 2007.

[3] S. Yan, H. Wang, X. Tang, and T. Huang, "Exploring feature descriptors for face recognition," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '07)*, pp. I629–I632, Honolulu, Hawaii, USA, April 2007.

[4] M. Á. Carreira-Perpiñán, "A review of dimension reduction techniques," Tech. Rep. CS-96-09, University of Sheffield, 1997.

[5] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, "Cloud computing and emerging IT platforms: vision, hype, and reality for delivering computing as the 5th utility," *Future Generation Computer Systems*, vol. 25, no. 6, pp. 599–616, 2009.

[6] L. Heilig and S. Vob, "A scientometric analysis of cloud computing literature," *IEEE Transactions on Cloud Computing*, vol. 2, no. 3, pp. 266–278, 2014.

[7] M. Stojmenovic, "Mobile cloud computing for biometric applications," in *Proceedings of the 15th International Conference on Network-Based Information Systems (NBIS '12)*, pp. 654–659, September 2012.

[8] A. S. Bommagani, M. C. Valenti, and A. Ross, "A framework for secure cloud-empowered mobile biometrics," in *Proceedings of the 33rd Annual IEEE Military Communications Conference (MILCOM '14)*, pp. 255–261, Baltimore, Md, USA, October 2014.

[9] K.-S. Wong and M. H. Kim, "Secure biometric-based authentication for cloud computing," in *Proceedings of the 2nd International Conference on Cloud Computing and Services Science (CLOSER '12)*, pp. 86–101, Porto, Portugal, 2012.

[10] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, 71, no. 1, p. 86, 1991.

[11] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 84–91, Seattle, Wash, USA, June 1994.

[12] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Face recognition using LDA-based algorithms," *IEEE Transactions on Neural Networks*, vol. 14, no. 1, pp. 195–200, 2003.

[13] Y. Wen, L. He, and P. Shi, "Face recognition using difference vector plus KPCA," *Digital Signal Processing*, vol. 22, no. 1, pp. 140–146, 2012.

[14] S. Chen, J. Liu, and Z.-H. Zhou, "Making FLDA applicable to face recognition with one sample per person," *Pattern Recognition*, vol. 37, no. 7, pp. 1553–1555, 2004.

[15] D. R. Kisku, H. Mehrotra, P. Gupta, and J. K. Sing, "Robust multi-camera view face recognition," *International Journal of Computers and Applications*, vol. 33, no. 3, pp. 211–219, 2011.

[16] L. Wiskott, J.-M. Fellous, N. Krüger, and C. D. von der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775–779, 1997.

[17] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[18] D. R. Kisku, A. Rattani, E. Grosso, and M. Tistarelli, "Face identification by SIFT-based complete graph topology," in *Proceedings of the IEEE Workshop on Automatic Identification Advanced Technologies (AUTOID '07)*, pp. 63–68, Alghero, Italy, June 2007.

[19] Z. Li, U. Park, and A. K. Jain, "A discriminative model for age invariant face recognition," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 1028–1037, 2011.

[20] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "SURF: speeded up robust features," *Computer Vision and Image Understanding (CVIU)*, vol. 110, no. 3, pp. 346–359, 2008.

[21] K. Yan and R. Sukthankar, "PCA-SIFT: a more distinctive representation for local image descriptors," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04)*, pp. II506–II513, July 2004.

[22] I. A.-A. Abdul-Jabbar, J. Tan, and Z. Hou, "Adaptive PCA-SIFT matching approach for face recognition application," in *Proceedings of the International Multi Conference of Engineers and Computer Scientists*, pp. 1–5, Hong Kong, 2014.

[23] R. E. G. Valenzuela, W. R. Schwartz, and H. Pedrini, "Dimensionality reduction through PCA over SIFT and SURF descriptors," in *Proceedings of the 11th IEEE International Conference on Cybernetic Intelligent Systems (CIS '12)*, pp. 58–63, IEEE, Limerick, Ireland, August 2012.

[24] D. Chen, S. Ren, Y. Wei, X. Cao, and J. Sun, "Joint cascade face detection and alignment," in *Proceedings of the 13th European Conference on Computer Vision*, pp. 109–122, Zurich, Switzerland, September 2014.

[25] R. C. Gonzalez and A. Woods, *Digital Image Processing*, Prentice Hall, 2008.

[26] R. Snelick, U. Uludag, A. Mink, M. Indovina, and A. Jain, "Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 450–455, 2005.

[27] Q. D. Tran, P. Kantartzis, and P. Liatsis, "Improving fusion with optimal weight selection in face recognition," *Integrated Computer-Aided Engineering*, vol. 19, no. 3, pp. 229–237, 2012.

[28] N. Poh, J. Kittler, and F. Alkoot, "A discriminative parametric approach to video-based score-level fusion for biometric authentication," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR '12)*, pp. 2335–2338, Tsukuba, Japan, November 2012.

[29] B. Abidi and M. A. Abidi, *Face Biometrics for Personal Identification: Multi-Sensory Multi-Modal Systems*, Springer, New York, NY, USA, 2007.

[30] A. Merati, N. Poh, and J. Kittler, "User-specific cohort selection and score normalization for biometric systems," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 4, pp. 1270–1277, 2012.

[31] M. Tistarelli, Y. Sun, and N. Poh, "On the use of discriminative cohort score normalization for unconstrained face recognition," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2063–2075, 2014.

[32] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *The American Statistician*, vol. 46, no. 3, pp. 175–185, 1992.

[33] H. Liu, J. Li, and L. Wong, "A comparative study on feature selection and classification methods using gene expression profiles and proteomic patterns," *Genome Informatics*, vol. 13, pp. 51–60, 2002.

[34] C. E. Thomaz and G. A. Giraldi, "A new ranking method for principal components analysis and its application to face image analysis," *Image and Vision Computing*, vol. 28, no. 6, pp. 902–913, 2010.

[35] O. Jesorsky, K. J. Kirchberg, and R. W. Frischholz, "Robust face detection using the hausdorff distance," in *Audio- and Video-Based Biometric Person Authentication: Third International Conference, AVBPA 2001 Halmstad, Sweden, June 6–8, 2001 Proceedings*, J. Bigun and F. Smeraldi, Eds., vol. 2091 of *Lecture Notes in Computer Science*, pp. 90–95, Springer, Berlin, Germany, 2001.

[36] M. K. Suguna, M. R. Shrihari, and M. R. Mahesh, "Implementation of face recognition in cloud vision using eigen faces," *International Journal of Engineering Research and Applications*, vol. 4, no. 7, pp. 151–155, 2014.

[37] P. Indrawan, S. Budiyatno, N. M. Ridho, and R. F. Sari, "Face recognition for social media with mobile cloud computing," *International Journal on Cloud Computing: Services and Architecture*, vol. 3, no. 1, pp. 23–35, 2013.

[38] S. K. Paul, M. S. Uddin, and S. Bouakaz, "Face recognition using facial features," *SOP Transactions on Signal Processing*, In press.

[39] S. Valuvanathorn, S. Nitsuwat, and M. L. Huang, "Multi-feature face recognition based on PSO-SVM," in *Proceedings of the 2012 10th International Conference on ICT and Knowledge Engineering, ICT and Knowledge Engineering*, pp. 140–145, Bangkok, Thailand, November 2012.