

# From Scores to Face Templates: A Model-Based Approach

Pranab Mohanty, *Student Member, IEEE*, Sudeep Sarkar, *Senior Member, IEEE*, and Rangachar Kasturi, *Fellow, IEEE*

**Abstract**—Regeneration of templates from match scores has security and privacy implications related to any biometric authentication system. We propose a novel paradigm to reconstruct face templates from match scores using a linear approach. It proceeds by first modeling the behavior of the given face recognition algorithm by an affine transformation. The goal of the modeling is to approximate the distances computed by a face recognition algorithm between two faces by distances between points, representing these faces, in an affine space. Given this space, templates from an independent image set (*break-in*) are matched only once with the enrolled template of the targeted subject and match scores are recorded. These scores are then used to embed the targeted subject in the approximating affine (nonorthogonal) space. Given the coordinates of the targeted subject in the affine space, the original template of the targeted subject is reconstructed using the inverse of the affine transformation. We demonstrate our ideas using three fundamentally different face recognition algorithms: Principal Component Analysis (PCA) with Mahalanobis cosine distance measure, Bayesian intra-extrapersonal classifier (BIC), and a feature-based commercial algorithm. To demonstrate the independence of the break-in set with the gallery set, we select face templates from two different databases: the Face Recognition Grand Challenge (FRGC) database and the Facial Recognition Technology (FERET) database. With an operational point set at 1 percent False Acceptance Rate (FAR) and 99 percent True Acceptance Rate (TAR) for 1,196 enrollments (FERET gallery), we show that at most 600 attempts (score computations) are required to achieve a 73 percent chance of breaking in as a randomly chosen target subject for the commercial face recognition system. With a similar operational setup, we achieve a 72 percent and 100 percent chance of breaking in for the Bayesian and PCA-based face recognition systems, respectively. With three different levels of score quantization, we achieve 69 percent, 68 percent, and 49 percent probability of break-in, indicating the robustness of our proposed scheme to score quantization. We also show that the proposed reconstruction scheme has 47 percent more probability of breaking in as a randomly chosen target subject for the commercial system as compared to a hill climbing approach with the same number of attempts. Given that the proposed template reconstruction method uses distinct face templates to reconstruct faces, this work exposes a more severe form of vulnerability than a hill climbing kind of attack where incrementally different versions of the same face are used. Also, the ability of the proposed approach to reconstruct the actual face templates of the users increases privacy concerns in biometric systems.

**Index Terms**—Face template reconstruction, probability of break-in, multidimensional scaling, security and privacy issues in biometric systems, hill climbing attack.

## 1 INTRODUCTION

RECENTLY, biometric technologies have become an integral part of many secure access systems. Biometric-based authentication systems are being deployed in both low-risk secure systems such as laptops and cell phones to relatively high-risk secure systems such as military bases and airports. The increasing demands of biometric technologies can be well justified with its advantages over password or smart-card-based technologies, such as user convenience, high security, and less fraud. However, like many other authentication technologies, biometric-based systems also possess vulnerable points of security breaches in biometric-based authentication systems [1]. The cost of replacing a biometric token or template is higher when compared to that of a password or a smart card, with severe security and privacy implications. The templates can be reused over digital networks or can be

used to reproduce synthetic biometric templates such as fake fingers or model faces [2], [3]. In case of face templates, there is an additional risk that the identity of a person using a biometric access system in a highly secure facility can be revealed. Several authors have successfully pointed out various sources of security breaches in biometric-based authentication systems [4], [5]. Lately, some countermeasures have also been proposed to nullify such threats [6], [7], [8], and the standardized biometric application programming interface (BioAPI) has been continuously updated with countermeasure guidelines such as to encrypt templates, avoid storage and transmission of original templates, and perform quantization of match scores [9].

In general, most biometric authentication systems have four major modules [10], a biometric template acquisition sensor, a matching module to compare a new template to an enrolled template, a decision module using predefined thresholds for particular operational points, and a database for enrolled templates (gallery). In many applications, it is not possible to integrate all these modules to one unit. In such scenarios, the information from one unit to the other is passed through digital channels and/or stored in digital media for offline processing. As reported by many authors [3], [8], each of these modules possesses different levels of security threats,

• The authors are with the Computer Science and Engineering Department, University of South Florida, 4202 E. Fowler Ave., ENB 118, Tampa, FL 33620-5399. E-mail: {pkmohant, Sarkar, chair}@cse.usf.edu.

Manuscript received 11 Apr. 2006; revised 17 Oct. 2006; accepted 20 Feb. 2007; published online 8 Mar. 2007.

Recommended for acceptance by S. Baker.

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0282-0406. Digital Object Identifier no. 10.1109/TPAMI.2007.1129.

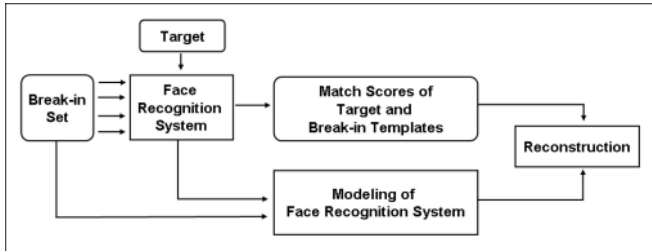


Fig. 1. Schematic of processes in the proposed model-based approach. The proposed approach is a model-based one-shot method employing multiple face templates—the break-in set.

and different countermeasures are necessary to nullify such threats. For instance, liveness detection at the sensor unit will detect any attempts to hack the system with synthetic templates. Similarly, a secure database or a secure digital channel will prevent any unauthorized access of templates over a network. In this paper, we made a successful attempt to explore one such point of vulnerability between a matching module and a decision module. In applications, where the matching module and decision module are not integrated together, we need to store the “match scores” in a digital media or transmit the match score through a digital channel to a decision module [11]. This scenario can arise in distributed network cluster biometric systems with a central decision unit. Such networks can arise in wide area monitoring contexts. In this paper, we consider the question: “Can unauthorized access of these match scores result in security and privacy breaches?”

We propose a model-based template reconstruction scheme from match scores using distinct face images. The schematic diagram of the proposed method is shown in Fig. 1. We consider a set of face images different from the gallery and probe sets and name it as the “break-in” set. Then, we build an affine transformation model of the face recognition algorithm that is assumed to be known. It can be noted that the face recognition system is treated as a complete black box, and we do not perform any reverse engineering on the recognition system. The assumption of the knowledge of the face recognition algorithm is a weak one. (It might even be possible to identify the recognition algorithm given score matrices of known algorithms. However, we do not explore that angle here.) The modeling of the recognition system is an offline procedure and needs to be constructed only once for a given recognition algorithm. Once we have built such a model, we present the templates from our break-in set to the system to be broken and observe the match scores to an assumed identity. Therefore, in real-time scenarios, our proposed method only requires access to a set of match scores equal to the number of images in the break-in set. These match scores are then used to embed the unknown template of a targeted subject in the modeled affine space. Finally, we use the inverse of the affine transformation to reconstruct the unknown template of a targeted subject in the original image space. We validate our proposed template reconstruction scheme on three different types of face recognition systems using two standard public databases, FERET [12] and FRGC [13]. Two template-based algorithms, namely, Principal Component Analysis with cosine distance measure [14] (widely accepted as a baseline algorithm) and Moghaddam and Pentland’s algorithm, popularly known as the Bayesian intrapersonal/extrapersonal classifier with Maximum Likelihood (ML) estimation [15],

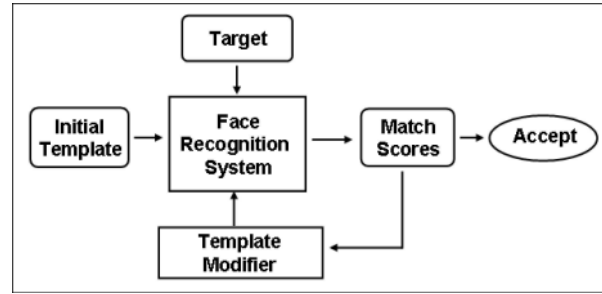


Fig. 2. Schematic of processes in a hill climbing attack. The hill climbing attack is an iterative process that starts from a face template and then iteratively updates the template until an accept decision is generated by the system.

and a feature-based commercial face recognition system are used for this experiment.

## 2 PREVIOUS APPROACHES: HILL CLIMBING-BASED ATTACKS

The dominant approach for a match score-based attack on a biometric system is based on hill climbing. Here, we outline the relevant work in this regard. In Section 5.4, we present a more systematic evaluation of our proposed method against one version of the hill climbing attack and demonstrate the efficiency of our method both qualitatively in terms of reconstructed templates and quantitatively in terms of the probability of breaking into a recognition system.

Soutar [16] was the first to propose an iterative template adaptation scheme, popularly known as the hill climbing attack, to break into a biometric system based on match scores. The proposed scheme attacks the account of a subject, referred to as the targeted subject, by starting from an arbitrary face template and iteratively refining it. At every iteration, if the modified template results in a better score than the previous match score, then the modified template is retained, or else, it is discarded. The process is iterated until the template is accepted as the targeted subject. The basic block diagram of the hill climbing attack is shown in Fig. 2. Note that with this method one might break into a system using a final template that does not look like any face as long as it “fools” the system. In other words, it is not a face reconstruction method but rather a break-in strategy. Though Soutar did not report any quantitative results of biometric template reconstruction, good performance of similar approaches has been reported by several others [17], [18].

One countermeasure for the first generation of hill climbing approaches is to quantize the scores. With appropriate quantization, it will not be possible to get an incremental feedback as is needed by these approaches. Therefore, Adler [17] proposed a modified hill climbing attack for a face recognition system with quantized match scores using an additional independent set of eigenfaces. The recognition systems that output quantized match scores do not alter the match scores with small changes in input images, which can prevent the first generation of hill climbing attacks. After initializing the process with an arbitrary face template, at every iteration, the previously updated template is multiplied with randomly selected eigenfaces with different weights. This is expected to generate templates farther away from the previous template. The face template that results in a

better match score is retained as the updated image for the next iteration. The process terminates when there is no further improvement in match scores. Experimental results on a commercial face recognition algorithm show that after nearly 4,000 attempts, a high match score is achieved with 99 percent confidence. Later, Adler [19] extended this idea to work with encrypted face templates.

Security breaches are possible not only in face biometrics but in other biometrics too. Uludag and Jain [18] extended the hill climbing attack idea to break into minutiae-based fingerprint recognition algorithms. Initially, random minutiae templates are created and matched against the targeted user by a fingerprint matching system. The best matched template is then used to generate another set of minutiae templates by randomly adding and deleting existing minutiae. The iteration process is continued till the system accepts the template. The authors reported that all 160 enrolled accounts could be broken with less than 1,000 attempts for each account. Lopresti and Raim [20] proposed an attack on an online handwriting recognition system by randomly generating feature vectors through a generative model of human handwriting. A set of different text samples from the enrolled users was fed to the generative model. With few text samples from the enrolled users, the model reproduced different text templates through random partition and concatenation of the input text until a template was accepted as a successful match. Preliminary results show that this attack succeeded 49 percent of the time.

Although hill climbing-based attacks can quite successfully break a particular targeted account, effective countermeasures for such type of attacks can also be generated. One property of hill climbing-based attacks is that they require a large number of attempts before success. Hence, one possible countermeasure for such attacks is to restrict the number of consecutive unsuccessful attempts. However, this still leaves the system vulnerable to a spyware-based attack that interlaces its false attempts with the attempts by genuine users (successful attempts) and collects information to iterate over a period of time. However, in most hill climbing-based attacks, the templates at the  $i$ th attempt (iteration) are generated from the  $(i - 1)$ th attempts (iterations) and are similar to each other. Hence, if we monitor all unsuccessful attempts for a particular targeted account within a fixed time interval, we will discover a pattern of similar faces with decreasing dissimilarity scores (see Fig. 3). Therefore, a continuous observation of unsuccessful match scores will help to detect hill climbing-based spyware attacks. In this paper, we expose a more severe form of vulnerability where such countermeasures will be hard to design since we use scores from distinct face images with no obvious patterns in the scores.

In Fig. 3, we present a schematic visualization of the search process to illustrate the differences between a hill climbing attack and our proposed linear scheme. Our algorithm requires the distances or scores between the face of the targeted subject and a set of faces from the break-in set that is distributed throughout the space. Whereas a hill climbing-based attack computes scores for faces along a trajectory of incremental scores from an arbitrary template to that of targeted subject, there are no obvious patterns in the scores needed by our approach; hence, the proposed scheme is not incrementally iterative. As discussed earlier, the statistically decreasing dissimilarity scores generated by a

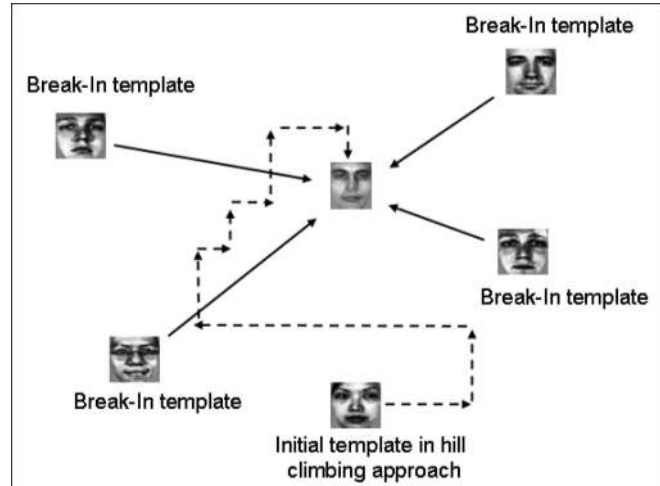


Fig. 3. Visualization of the search process of a hill climbing attack and the proposed model-based approach. The dotted line represents one of possible path in a hill climbing attack starting from a chosen template. At each point on the path, the distance to the template is needed. The solid arrow represents a one-time comparison of templates of the targeted subject with members of the break-in set templates that is needed by the proposed approach.

hill climbing-based approach can indeed be used to detect such attacks, but a similar strategy cannot be applied to our proposed method. The hill climbing approach is considered as a break-in strategy to a recognition system, whereas the proposed method is a template reconstruction scheme for any face recognition system. In our case, the break-in performance shows the accuracy and confidence in reconstructed templates. As a result, the proposed algorithm has vulnerability implications related to both security and privacy issues of the users. Also, the numbers of attempts in our break-in scheme are predefined by the number of images in the break-in set, which allows such attacks to be more feasible in real-time applications.

### 3 MODELING OF THE RECOGNITION ALGORITHM

The heart of our proposed template reconstruction approach is the modeling of a face recognition algorithm using an affine transform and then inverting this transformation to reconstruct templates. The inputs to any face recognition system are two face templates  $x_i$  and  $x_j$ , and the output is a dissimilarity or similarity score  $d_{ij}$ . Typically, a recognition algorithm transforms the given image into a point in a low-dimensional space, followed by a distance measure on this model space. This low-dimensional space is usually constructed by statistical methods such as PCA [14], linear discriminant analysis (LDA) [21], or independent component analysis (ICA) [22] or constructed out of low-level features detected in the images, such as in the elastic bunch graph matching approach [23]. We adopt a black box-based modeling approach. Given the success of template-based recognition approaches such as the Bayesian, PCA, LDA, and ICA that rely on linear transformations of the original image space, we seek to model the given face recognition algorithm, even feature-based ones, by an affine transformation, followed by euclidean distance computation in this model space. However, unlike previous template-based approaches, we allow the transformation to be nonorthogonal. We seek an

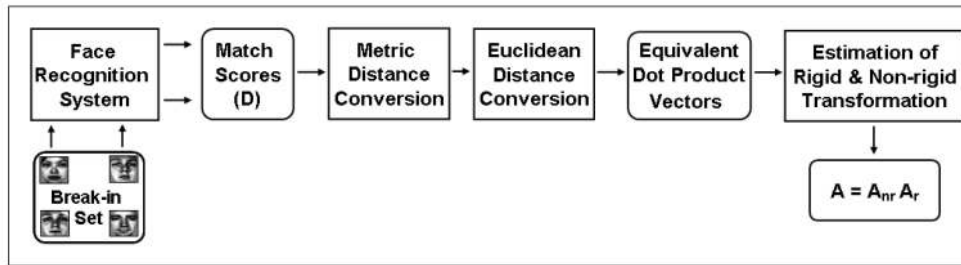


Fig. 4. Block diagram of the modeling strategy. Starting with a set of face templates (break-in set), we estimate a rigid and a nonrigid transformation that model the behavior of the recognition algorithm in terms of distance measure on these templates.

approximating affine transformation  $\mathbf{A}$  that is a composition of an orthogonal (or rigid) and a nonrigid (shear and stretch) transformation  $\mathbf{A} = \mathbf{A}_{nr}\mathbf{A}_r$ . Similar questions have also been considered by Liu et al. [24] but just for orthogonal transformations, that is,  $\mathbf{A}_{nr} = \mathbf{I}$ , the identity matrix. They cast it as an optimization problem over the space of possible transformations. We offer a more direct method to find the general affine transformation, not necessarily orthogonal. We have found that the nonrigid part of the transformation does help in enhancing performance [25]. The approximating affine transformation preserves the distances among the templates generated by the face recognition system. In Fig. 4, we outline the steps involved in designing the affine model space. Given this affine space, we can embed any template in this space based on its distance  $d$  from a known set of templates—the set of break-in templates. Once we have the embedded affine coordinates for template  $\mathbf{y}_z$ , we can reconstruct the face by inverting the affine transformation.

The objective of the modeling is to determine an affine transform  $\mathbf{A}$  such that when the given images  $\mathbf{x}_i$ 's, are transformed to the affine space, the euclidean distance between the transformed coordinates of two face images are similar to the distances computed between them by the face recognition algorithm. In this notation,  $\mathbf{x}_i$ 's are  $N$ -dimensional row-scanned representations of the images, and the affine transformation  $\mathbf{A}$  has dimensions  $M \times N$ , with  $M < N$ . We find this model in two steps: First, we express the given distances (or their monotonically increasing transformation) between known images as a dot product distance between vectors. Then, we construct the affine transformation between these vectors and the images. Pekalska et al. [26] discuss a similar mechanism to derive models for standard classifiers such as the nearest neighborhood rule, linear discriminant analysis, and linear programming problems from the dissimilarity scores between objects. The distances are embedded to the euclidean/pseudoeuclidean space depending on the presence/absence of the euclidean property of the original distance matrix, and then, unknown objects are projected to the embedded space and classified accordingly with a euclidean distance measure.

### 3.1 Dot Product Distances

Let  $d_{ij}$  be the distance between two images  $\mathbf{x}_i$  and  $\mathbf{x}_j$ , ( $\mathbf{x}_i^T \in \mathbb{R}^N$ ) as computed by the given face recognition algorithm. Here, we assume that the face recognition algorithm outputs the dissimilarity scores of two templates. However, if a recognition algorithm computes similarities instead of distances, we can always convert the similarity scores  $s_{ij}$  into distances using a variety of transformations such as  $(1 - s_{ij})$ ,  $-\log(s_{ij})$ ,  $\frac{1}{s_{ij}} - 1$ , and so forth. Then, these distances can be

arranged as a  $K \times K$  matrix  $\mathbf{D} = [d_{ij}^2]$ , where  $K$  is the number of images in the breaking set. In this paper, we will denote matrices by bold capital letters  $\mathbf{A}$  and column vectors by bold small letters  $\mathbf{a}$ . We will denote the identity matrix by  $\mathbf{I}$ , a vector of ones by  $\mathbf{1}$ , a vector of zeros by  $\mathbf{0}$ , and the transpose of  $\mathbf{A}$  by  $\mathbf{A}^T$ . For each image, we would like to find vectors  $\mathbf{y}_i$  such that  $\mathbf{y}_i^T \mathbf{y}_j = f(d_{ij})$ , where  $f(\cdot)$  is a monotonically increasing function, and  $\mathbf{y}_i^T \in \mathbb{R}^M$ . For biometric systems, if the original match score between two templates is not modified based on other templates on the gallery, then a monotonically increasing transformation of the distances does not affect the model of the system. The choice of this monotonically increasing function depends on the face recognition algorithm under consideration. However, a common approach is to express this monotonically increasing transformation as a composition of two transformations, the first one transforms the given distances into euclidean distances, and the second one “centers” the euclidean distances into dot product distances. For many recognition algorithms, the underlying distance measure may not be euclidean, and in some case, the observed dissimilarity matrix may not exhibit metric properties as well. In such cases, we need to transform the distance matrix  $\mathbf{D}$  to the equivalent euclidean distance matrix  $\mathbf{D}_E$ . Although the process of converting a noneuclidean distance matrix to an equivalent euclidean distance matrix is not feasible in all cases, an approximation to noneuclidean distance matrix  $\mathbf{D}$  can be used for such embedding [26]. In the rest of this section, we discuss the mathematical derivation of configuration points  $\mathbf{y}_i$  from the euclidean distance matrix  $\mathbf{D}_E$ . Note that if the original observed distance matrix  $\mathbf{D}$  is euclidean, then  $\mathbf{D}_E = \mathbf{D}$ . On the other hand, if  $\mathbf{D}$  is a noneuclidean distance matrix, then  $\mathbf{D}_E$  represents a monotonically modified equivalent euclidean distance matrix of the original distance matrix  $\mathbf{D}$ . For better readability, the derivation of  $\mathbf{D}_E$  from the noneuclidean distance matrix  $\mathbf{D}$  is discussed in Section 3.3. For now, we will assume that we have the score matrix  $\mathbf{D}_E$ .

Although many different schemes [27], [28] can be used to arrive at a set of configuration points that preserve the pairwise distances given by an input distance matrix, for this experiment, we followed a simple scheme commonly known as classical scaling or metric multidimensional scaling (MDS) [29], [30]. Given the euclidean distance matrix  $\mathbf{D}_E$ , here, the objective is to find  $K$  vectors  $\{\mathbf{y}_1, \dots, \mathbf{y}_K\}$  such that

$$\mathbf{D}_E(i, j) = (\mathbf{y}_i - \mathbf{y}_j)^T (\mathbf{y}_i - \mathbf{y}_j). \quad (1)$$

Note that the above configuration points  $\mathbf{y}_i$ 's are not unique. Any translation or rotation of vectors  $\mathbf{y}_i$ 's can also be a solution to (1). To reduce such degrees of freedom of

the solution set, we constrain the solution set of vectors to be centered at the origin and the sum of the vectors to zero, that is,  $\sum_i \mathbf{y}_i = \mathbf{0}$ .

Equation (1) can be compactly represented in matrix form as

$$\mathbf{D}_E = \mathbf{c} \cdot \mathbf{1}^T + \mathbf{1} \cdot \mathbf{c}^T - 2\mathbf{Y}^T\mathbf{Y}, \quad (2)$$

where  $\mathbf{Y}$  is a matrix constructed using the vectors  $\mathbf{y}_i$  as the columns  $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_K]$ , and  $\mathbf{c}$  is a column vector of the magnitudes of the vectors  $\mathbf{y}_i$ 's. Thus,

$$\mathbf{c} = [\mathbf{y}_1^T \mathbf{y}_1, \dots, \mathbf{y}_K^T \mathbf{y}_K]^T. \quad (3)$$

To simplify (2), if we pre and postmultiply each side of the equation by centering matrix  $\mathbf{H} = (\mathbf{I} - \frac{1}{K}\mathbf{1}\mathbf{1}^T)$ , we have

$$\mathbf{H}\mathbf{D}_E\mathbf{H} = \mathbf{H}\mathbf{c} \cdot \mathbf{1}^T\mathbf{H} + \mathbf{H}\mathbf{1} \cdot \mathbf{c}^T\mathbf{H} - 2\mathbf{H}\mathbf{Y}^T\mathbf{Y}\mathbf{H} = -2\mathbf{Y}^T\mathbf{Y}, \quad (4)$$

where we have used the constraint that we are looking for the centered solution set, that is,  $\sum_i \mathbf{y}_i = \mathbf{0}$ ; thus,  $\mathbf{H}\mathbf{c} = \mathbf{0}$  and  $\mathbf{H}\mathbf{Y}^T = \mathbf{Y}^T$ . Using  $\mathbf{B}$  to represent  $-\frac{1}{2}\mathbf{H}\mathbf{D}_E\mathbf{H}$ , the search for the coordinates can be cast as

$$\mathbf{B} = -\frac{1}{2}\mathbf{H}\mathbf{D}_E\mathbf{H} = \mathbf{Y}^T\mathbf{Y}. \quad (5)$$

Since  $\mathbf{D}_E$  is a euclidean matrix, the matrix  $\mathbf{B}$  is also a distance matrix, representing dot product distances between the vectors  $\mathbf{y}_i$ 's and is a symmetric positive semidefinite matrix [29], [30].

### 3.2 Distance to Vectors

The next task is to find a set of vectors such that  $\mathbf{Y}^T\mathbf{Y} = \mathbf{B}$ , where  $\mathbf{B}$  is the dot product distances derived from the monotonically increasing transformation of the distances computed by the face recognition algorithm being modeled. One such solution strategy is to use the eigenvalue decomposition (EVD) of  $\mathbf{B}$ . Since  $\mathbf{B}$  is a symmetric positive semidefinite matrix, let us assume that the rank of  $\mathbf{B}$  is  $M \leq N$ , so  $\mathbf{B}$  has  $M$  nonnegative eigenvalues and  $N - M$  zero eigenvalues. Hence,

$$\mathbf{B} = \mathbf{V}_{\text{EVD}}\mathbf{\Delta}_{\text{EVD}}\mathbf{V}_{\text{EVD}}^T, \quad (6)$$

where  $\mathbf{\Delta}_{\text{EVD}}$  is  $N \times N$  diagonal matrices where the first  $M$  diagonal entries represents the nonzero eigenvalues of matrix  $\mathbf{B}$  sorted in ascending order.  $\mathbf{V}_{\text{EVD}}$  represents the corresponding eigenvectors of  $\mathbf{B}$ . The solution is then given by

$$\mathbf{Y} = (\mathbf{V}_{\text{EVD}}^M \mathbf{\Delta}_{\text{EVD}}^{\frac{1}{2}})^T, \quad (7)$$

where  $\mathbf{\Delta}_{\text{EVD}}^M$  is  $M \times M$  diagonal matrices consisting of  $M$  nonzero eigenvalues of  $\mathbf{B}$ , and  $\mathbf{V}_{\text{EVD}}^M$  represents the corresponding eigenvectors of  $\mathbf{B}$ .

### 3.3 Noneuclidean Distance to Euclidean Distance

In this section, we will discuss the derivation of euclidean distance matrix ( $\mathbf{D}_E$ ) from a noneuclidean distance matrix ( $\mathbf{D}$ ). The procedure discussed in this section was first proposed by Gower and Legendre [31] and later followed by many authors [29], [30], [32]. We are presenting a similar discussion adapted from [26]. In order to understand the details of the euclidean matrix, let us recall the metric property and a related theorem on the euclidean distance matrix.

**Definition 1 (Metric property).** A distance measure  $\mathbf{d}$  is called a metric if it satisfies the following properties:

1.  $d(x, y) = 0$  iff  $x = y$  (reflexive).
2.  $d(x, y) \geq 0 \quad \forall x \neq y$  (positivity).
3.  $d(x, y) = d(y, x)$  (symmetry).
4.  $d(x, y) \leq d(x, z) + d(z, y)$  (triangle inequality).

**Theorem 3.1.** A distance matrix  $\mathbf{D}$  is euclidean iff  $\mathbf{B} = -\frac{1}{2}\mathbf{H}\mathbf{D}_E\mathbf{H}$  is a positive semidefinite metric [29], [31].

Let the observed distance matrix  $\mathbf{D}$  be noneuclidean. Then, either the distance measure in  $\mathbf{D}$  is not a metric and/or  $\mathbf{D}$  is not a positive semidefinite matrix. In this section, we will try to reinforce the missing properties of a euclidean distance matrix without affecting the overall recognition performance from  $\mathbf{D}$ . First, let us examine each of the metric properties more closely on  $\mathbf{D}$ . In most of the applications such as biometric template matching, the reflexive and positivity properties are straightforward. As we know, two different templates with little variation always produce a nonzero dissimilarity. Hence, we can assume that the reflexive and positivity properties always hold unless small scores are forcefully suppressed to zero. Even if the scores are rounded off to the nearest number or small scores are suppressed to zero, as long as we do not have a sparse distance matrix with few positive entries, we still can find an embedding in the model space that can approximate the distance matrix  $\mathbf{D}$ . Now, if the distance matrix  $\mathbf{D}$  violates the symmetric property, then this property can be reinstated by replacing  $\mathbf{D}$  with  $\frac{1}{2}(\mathbf{D} + \mathbf{D}^T)$ . Although this simple solution will change the performance of the algorithm, this correction can be viewed as a first cut fix for our modeling transformation to the algorithms that violate the symmetric property of match scores. Finally, if  $\mathbf{D}$  violates the triangle inequality, then we can also enforce the triangle inequality by adding a constant factor  $\varsigma$  to nondiagonal entries of  $\mathbf{D}$ , where  $\varsigma \geq \max_{i,j,k} |d_{ij} + d_{jk} - d_{ik}|$  (see [29, p. 21]). The value of  $\varsigma$  is learned using break-in set templates only. Using an offline copy of the face recognition system, we compute pairwise distance between every template of the break-in set; as a result, we have a full distance matrix  $\mathbf{D}$  with diagonal elements representing self-distance that is set to zero. While computing the distance from the targeted subject to each template in the break-in set, we use the learned value of  $\varsigma$ . Note that  $\varsigma$  is added to the nondiagonal entries of  $\mathbf{D}$  irrespective of genuine and impostor scores; therefore, the overall performance of the face recognition system, as represented by the distance matrix, is not affected by the addition of  $\varsigma$  to the computed distances.

With the above discussion, at this point, we can assume that  $\mathbf{D}$  is a metric distance, and if  $\mathbf{D}$  is noneuclidean, then it must be violating the positive semidefinite property of the distance matrix. In that case,  $\mathbf{B}$  has negative eigenvalues, and we can not use (7) to derive configuration points  $\mathbf{y}_i$ 's. As discussed in [26, Section 3.2]), there are two alternatives to derive configuration points  $\mathbf{y}_i$ . One possible solution is to consider only  $M$  positive eigenvalues of  $\mathbf{B}$ . This approach is useful when the magnitude of the largest negative eigenvalue is much smaller than the largest positive eigenvalues of  $\mathbf{B}$ , and the total energy contributed by positive eigenvalues must be greater than that of negative eigenvalues. This indicates that the original distance matrix  $\mathbf{D}$  is closed to a euclidean distance matrix, and the negative eigenvalues may occur due to error in observing the distance. However, for biometric applications where noneuclidean distance measures are used successfully, this method may not be suitable. Therefore, in this experiment, we follow the alternate approach where the

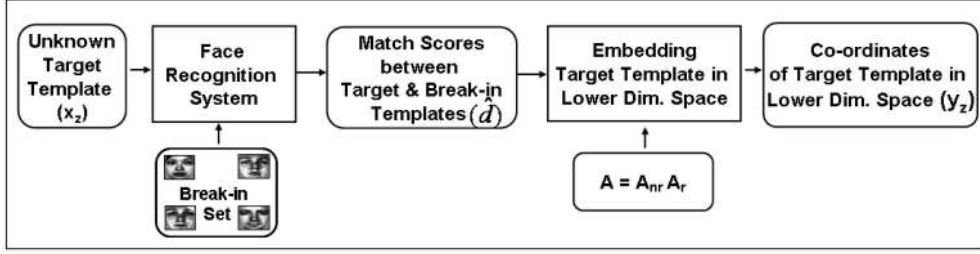


Fig. 5. Block diagram of the embedding scheme. After estimating the affine transformation  $\mathbf{A}$ , we use the distance between a template of the targeted subject and selected break-in set templates to compute the coordinate of the targeted subject in lower dimensional space.

original distance matrix is modified according to Theorem 3.2 to create a new euclidean distance matrix  $\mathbf{D}_E$ .

**Theorem 3.2.** *If  $\mathbf{D}$  is a metric distance, then there exists a constant  $h$  such that the matrix with elements  $(d_{ij}^2 + h)^{\frac{1}{2}}$ ,  $i \neq j$  is euclidean, where  $h \geq -2\lambda_n$  is the smallest (negative) eigenvalue of  $\mathbf{HDH}$ , where  $\mathbf{H} = (\mathbf{I} - \frac{1}{K}\mathbf{1}\mathbf{1}^T)$  [29], [31], [32].*

### 3.4 Affine Transformation

So far, we have seen how to find a set of coordinates  $\mathbf{Y}$  such that the euclidean distance between these coordinates is related to the distances computed by the recognition algorithm by an additive constant. We now find an affine transformation  $\mathbf{A}$  that will relate these coordinates  $\mathbf{Y}$  to the images  $\mathbf{X}$  such that

$$\mathbf{Y} = \mathbf{A}(\mathbf{X} - \mu), \quad (8)$$

where  $\mu$  is the mean of the images in the break-in set, that is, average face. We do not restrict this transformation to be orthonormal or rigid. We consider  $\mathbf{A}$  to be composed of two subtransformations: nonrigid transformation  $\mathbf{A}_{nr}$  and rigid transformation  $\mathbf{A}_r$ , that is,  $\mathbf{A} = \mathbf{A}_{nr}\mathbf{A}_r$ . The rigid part  $\mathbf{A}_r$  can be arrived at by any analysis that computes an orthonormal subspace from the given set of training images. In this experiment, we use PCA for the rigid transformation. Let the PCA coordinates corresponding to the nonzero eigenvalues, that is, non-null subspace, be denoted by  $\mathbf{X}_r = \mathbf{A}_r(\mathbf{X} - \mu)$ . The nonrigid transformation  $\mathbf{A}_{nr}$  relates these rigid coordinates  $\mathbf{X}_r$  to the distance-based coordinates  $\mathbf{Y}$

$$\mathbf{Y} = \mathbf{A}_{nr}\mathbf{X}_r. \quad (9)$$

Substituting (7) in (9), we have

$$\mathbf{A}_{nr}\mathbf{X}_r = \left( \mathbf{V}_{\text{EVD}}^M \mathbf{\Delta}_{\text{EVD}}^M \right)^T. \quad (10)$$

Multiplying both sides of (10) by  $\mathbf{X}_r^T$  and using the result that  $\mathbf{X}_r \mathbf{X}_r^T = \mathbf{\Lambda}_{\text{PCA}}$ , where  $\mathbf{\Lambda}_{\text{PCA}}$  is the diagonal matrix with the nonzero eigenvalues computed by PCA, we have

$$\mathbf{A}_{nr} = \left( \mathbf{V}_{\text{EVD}}^M \mathbf{\Delta}_{\text{EVD}}^M \right)^T \mathbf{X}_r^T \mathbf{\Lambda}_{\text{PCA}}^{-1}. \quad (11)$$

This nonrigid transformation, allowing for shear and stress, and the rigid transformation, computed by PCA, together model the face recognition algorithm. Note that the rigid transformation is not dependent on the face recognition algorithm; it is only the nonrigid part that is determined by the distances computed by the recognition algorithm. An alternative interpretation could be that the nonrigid transformation captures the difference between a PCA-based

recognition strategy—the baseline—and the given face recognition algorithm.

## 4 EMBEDDING AND RECONSTRUCTION

For the break-in scenario, we will not have access to the template of targeted subject; however, we will be able to retrieve the distances of the targeted subject to any given image. Therefore, we need a mechanism to be able to compute the coordinates of the targeted subject from the given distances, that is, embed the image of the targeted subject in the modeling affine space. Given the embedded coordinates, we will use the inverse transformation to reconstruct the face template of the targeted subject. In this section, we explain the embedding solution, outlined Fig. 5. Let  $\mathbf{y}_z$  be the unknown template coordinate vector of the targeted subject in the affine space. Let  $\mathbf{d} = [\hat{d}_1, \hat{d}_2, \dots, \hat{d}_K]^T$  be the vector of distances of  $\mathbf{y}_z$  from the  $K$  images  $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_K]$  in the break-in set as computed by the face recognition algorithm, along with the euclidean correction factor that was found during the estimation of the recognition algorithm (see Section 3.1). Based on the nature of the construction of the affine space, these distances would be equal to the euclidean distance between the vectors  $\mathbf{y}_z$  and  $\mathbf{y}_i$ .

Mathematically,

$$\hat{d}_i^2 = \|\mathbf{y}_i - \mathbf{y}_z\|^2 = \|\mathbf{y}_i\|^2 + \|\mathbf{y}_z\|^2 - 2\mathbf{y}_i^T \mathbf{y}_z, \quad \forall i = 1, \dots, K. \quad (12)$$

Subtracting  $\hat{d}_i^2$  from  $\hat{d}_{i+1}^2$  and simplifying, we have

$$\begin{aligned} \mathbf{E}\mathbf{y}_z &= \mathbf{F} \\ \mathbf{y}_z &= \mathbf{E}^\dagger \mathbf{F}, \end{aligned} \quad (13)$$

where

$$\mathbf{E}^T = \left[ (\mathbf{y}_2 - \mathbf{y}_1)^T, (\mathbf{y}_3 - \mathbf{y}_2)^T, \dots, (\mathbf{y}_K - \mathbf{y}_{K-1})^T \right], \quad (14)$$

$$\mathbf{F}^T = [f_i], \quad f_i = \frac{1}{2} \left[ \left( \hat{d}_i^2 - \|\mathbf{y}_i\|^2 \right) - \left( \hat{d}_{i+1}^2 - \|\mathbf{y}_{i+1}\|^2 \right) \right], \quad (15)$$

and  $\mathbf{E}^\dagger$  represents the pseudoinverse of  $\mathbf{E}$ . Here, we assume that  $\mathbf{E}$  does not map all points to the null space of  $\mathbf{F}$ ; hence, the pseudoinverse of  $\mathbf{E}$  exists. Since,  $\mathbf{E}$  consists of all projected points  $\mathbf{y}_i$ 's in the model space, a very low rank of  $\mathbf{E}$ , say, two or three, indicates that either the face recognition algorithm computes the distance between two templates in such low-dimensional space, or the templates in the break-in set are similar to each other and, hence, lying in a subspace of dimension two or three. The later scenario can be avoided by selecting distinct templates in the break-in set; however, if the

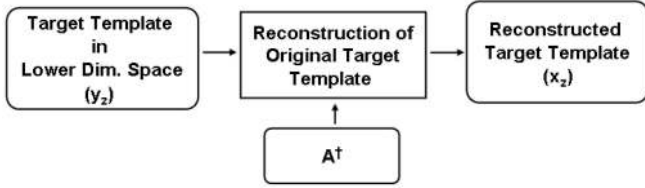


Fig. 6. Block diagram of the reconstruction scheme. Once the coordinates of the targeted subject in the lower dimensional space is known, we simply use the pseudoinversion of the affine transformation to reconstruct the original template of the targeted subject.

recognition algorithm projects the templates to two or three-dimensional spaces, then the performance of the system will have a low False Acceptance Rate (FAR), and any arbitrary template has a high probability of breaking into the system.

Now, given distances  $\hat{\mathbf{d}} = [\hat{d}_1, \hat{d}_2, \dots, \hat{d}_K]$  of any unknown template  $\mathbf{y}_z$  from  $K$  images in the break-in set, we can use (13) to compute the coordinates of  $\mathbf{y}_z$  in the approximating affine space.

Once we obtain the coordinate of any unknown template in the affine space, we invert the transformation to reconstruct the template (see Fig. 6). Mathematically, if  $\mathbf{y}_z$  is the embedding coordinate of unknown template  $\mathbf{x}_z$ , then

$$\begin{aligned} \mathbf{A}_{nr} \mathbf{A}_r \mathbf{x}_z &= \mathbf{y}_z, \\ \mathbf{x}_z &= \mathbf{A}_r^T \mathbf{A}_{nr}^\dagger \mathbf{y}_z. \end{aligned} \quad (16)$$

In summary, the individual steps involved in reconstructing a template of the targeted subject follows:

#### Inputs

1. knowledge of the face recognition algorithm,
2. a set of  $K$  face images (break-in set), and
3. a set of match scores between the templates from break-in set to the assumed identity's template.

#### Modeling

1. Compute distance matrix  $\mathbf{D}$  between these  $K$  templates using the underlying face recognition algorithm.
2. If  $\mathbf{D}$  is not euclidean, then compute the equivalent euclidean distance matrix  $\mathbf{D}_E$ .
3. Calculate  $\mathbf{X}_{EVD}$  from  $\mathbf{D}_E$ :
  - a. Construct the matrix  $\mathbf{B}$  by double centering with  $\mathbf{H}$ . This step “centers” the given distances and converts them into equivalent dot product distances  $\mathbf{B} = -\frac{1}{2} \mathbf{H} \mathbf{D} \mathbf{H}$ .
  - b. Compute the EVD of  $\mathbf{B}$  as
$$\mathbf{B} = \mathbf{V}_{EVD} \mathbf{\Delta}_{EVD} \mathbf{V}_{EVD}^T.$$
  - c. Compute coordinates  $\mathbf{Y}$  as  $\mathbf{Y} = (\mathbf{V}_{EVD}^M \mathbf{\Delta}_{EVD}^{M \frac{1}{2}})^T$ .
4. Build the affine transformation  $\mathbf{A} = \mathbf{A}_{nr} \mathbf{A}_r$ :
  - a. The rigid part  $\mathbf{A}_r$  of the affine transformation can be arrived at by PCA. Let the PCA coordinates be denoted by  $\mathbf{X}_r = \mathbf{A}_r \mathbf{X}$ . The nonrigid part  $\mathbf{A}_{nr}$  of the transformation is given by

$$\mathbf{A}_{nr} = \left( \mathbf{V}_{EVD}^M \mathbf{\Delta}_{EVD}^{M \frac{1}{2}} \right)^T \mathbf{X}_r^T \mathbf{\Lambda}_{PCA}^{-1},$$

where  $\mathbf{\Lambda}_{PCA}$  is the diagonal matrix with the PCA eigenvalues.

#### Embedding and Reconstruction

5. Find the MDS coordinate  $\mathbf{y}_z$  of the targeted subject  $\mathbf{x}_z$ :
  - a. Compare the templates in the break-in set with the template of the target subject to create distances vector  $\hat{\mathbf{d}}$ .
  - b. The coordinate of the targeted subject  $\mathbf{y}_z$  in MDS space is constructed as  $\mathbf{y}_z = \mathbf{E}^\dagger \cdot \mathbf{F}$ .
6. Reconstruct the unknown template  $\mathbf{x}_z$  using (16).

## 5 EXPERIMENTAL SETUP AND RESULTS

We demonstrate our reconstruction scheme using three fundamentally different face recognition algorithms: PCA with the Mahalanobis cosine distance measure, the Bayesian intra-extrapersonal classifier (BIC), and a feature-based commercial algorithm. In order to emphasize the true independence of the break-in set and gallery set, we use two distinct public databases. The FERET [12] database is used for the gallery images, and the FRGC database [13] is used to construct different break-in sets. In this section, we first provide an overview of the two databases and face recognition algorithms used in our experiments. Then, we present the reconstructed templates and corresponding break-in performance for each of the face recognition algorithms. Later, we compare our approach with hill climbing-based attacks and show the efficiency of our proposed method over a hill climbing-based approach [17] both in terms of quality of reconstructed templates and break-in performance. Finally, we demonstrate the robustness of our proposed algorithm to score quantization.

### 5.1 Experimental Setup

#### 5.1.1 Database

The face images used in this experiment are selected from the FERET [12] and FRGC face databases [13]. To ensure the distinctiveness of the break-in set with the gallery set, we choose our break-in set from a subset of the FRGC training set and reconstructed all the images from the FERET gallery set containing 1,196 images from 1,196 subjects. The FERET face database is a widely used public database, and the gallery set is predefined (*feret\_gallery.srt* in [33]) in that database. We use the Colorado State University (CSU) Face Identification Evaluation System to normalize the original face images [33]. The normalized face images have the same eye locations, the same size ( $150 \times 130$ ), and similar intensity distribution. Few preprocessed face images are shown in Fig. 7. For break-in sets, we selected a subset of the FRGC training set with 600 controlled images from the first 150 subjects (in the increasing order of their numeric ID) with four images per subject. In order to validate the effectiveness of the proposed template reconstruction scheme and break-in strategy, it is necessary that the selected face recognition algorithms have high recognition rates at low FAR. Since most of the face recognition algorithms perform poorly on a data set with one or more variations in face images [13], we restrict our experiments to controlled frontal face images only. Similarly, current template-based algorithms require the images to be scaled to the same size with the same eye location, so a preprocessing step is inevitable for such algorithms.

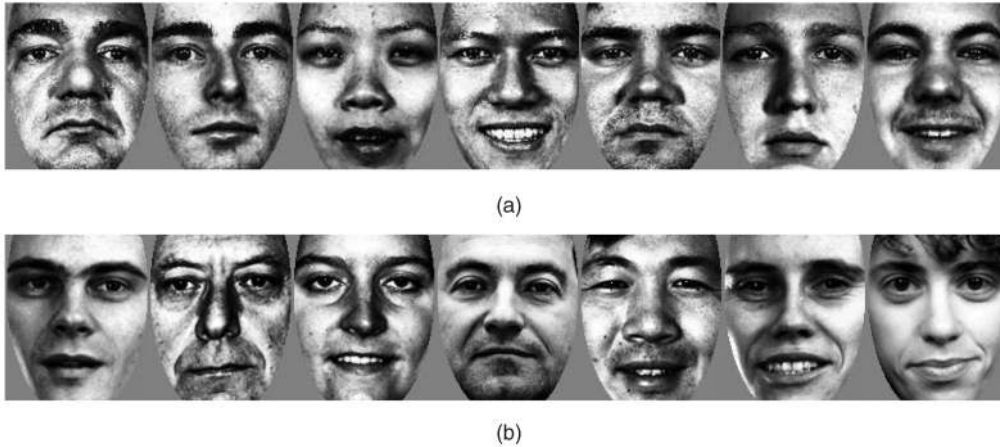


Fig. 7. Sample images from (a) the break-in set and (b) the gallery set. The break-in set and gallery set are independent of each other and have no subjects in common.

However, if a face recognition system has high performance without such restriction on the variation of face images or size of the face images, then the proposed scheme can be extended naturally to such systems.

### 5.1.2 Face Recognition Algorithms

We evaluate the proposed reconstruction scheme on two template-based algorithms and one feature-based face recognition system: 1) PCA approach with Mahalanobis cosine angle as the distance measure, which, by default, is considered as the baseline algorithm for the face recognition system [14], 2) Moghaddam and Pentland's algorithm, popularly known as the Bayesian intrapersonal/extrapolational classifier [15], and 3) a commercial face recognition system. The commercial system is based on a Local Feature Analysis of face images and widely regarded as being among the best available at present. Both the baseline and Bayesian algorithms were trained using the break-in set from the FRGC training set, but the commercial algorithm did not require any training process and was used as a black box in all of our experiments. Since all the face images are normalized with fixed eye coordinates ( $\theta$ ) and fixed-size ( $150 \times 130$ ) face images, we did not utilize the face and eye detector module embedded in the commercial face recognition system. Using the *fafb* probe set of the FERET distribution, we observe that the baseline, the Bayesian, and the commercial algorithms have 97 percent, 95 percent, and 99 percent True Acceptance Rate (TAR) at 1 percent FAR, respectively, on the *fafb* experiment in the FERET database.

### 5.1.3 Distance Measure

The three algorithms used in this experiment have completely different approaches of comparing two faces and generate similarity and/or dissimilarity scores with different distance measures. The baseline algorithm uses a Mahalanobis cosine angle and has dissimilarity scores between  $-1$  and  $1$ . Similarly, the Bayesian maximum likelihood classifier reports the similarity between two faces in terms of probability of difference image to the intrapersonal/extrapolational space. For this experiment, we use the CSU implementation of the Bayesian algorithm [33], where a negative logarithm transformation is applied to the probabilistic similarity score to convert them into a distance

measure [34]. However, in order to have an upper bound for the dissimilarity scores, we row normalize the distances to the interval  $[0, 1]$ . The similarity measure used in the feature-based commercial algorithm is not known, but the similarity scores are within a finite range of  $[S_{min}, S_{max}]$ . We convert similarity scores to distances by simply subtracting each match score  $S_{ij}$  from the maximum possible match score ( $S_{max} - S_{min}$ ). In our experiments, we use raw match scores from the commercial system without any score normalization. We observe that all the three distance measures used by respective algorithms exhibit the symmetric property but violate the triangle inequality property. Hence, we reinforce the triangle inequality property in the respective distance matrices, as discussed in Section 3.3. The values of  $\zeta$  learned from the break-in set are 1.297, 2.094, and 19.970 for the baseline, the Bayesian, and the commercial algorithms, respectively.

## 5.2 Affine Modeling

Our first objective is to model the behavior of each face recognition algorithm in terms of an affine transformation. In other words, the distance between two templates computed by these algorithms should be close to the euclidean distance between the two templates in the respective affine spaces. Here, we present some of the intermediate results showing the accuracy of our modeling scheme and the behavior of the constructed affine spaces.

In Fig. 8, we plot the eigenvalues of the transformed distance matrices  $\mathbf{B}$  defined in (5). The eigenvalues of the individual algorithms reflect the nature of the affine space for each individual algorithm. The plots for the eigenvalues of the three distance matrices from the three algorithms appear different due to different scales of eigenvalues for each algorithm. In Fig. 8a, we can observe that the eigenvalues drop from 9.4 to zero at 360 index of the eigenvector, which is about 60 percent of the total number of images. Thus, it can be inferred that the baseline algorithm uses top eigenvectors that contribute 60 percent of the total energy. Moreover, Fig. 8 also provides estimation for the dimension of each affine space. For example, we can expect that for the baseline algorithm, any break-in set with more than 360 images will result in approximately the same probability of break-in. In other words, 360 images (attempts) are sufficient to achieve an optimal break-in performance for the baseline algorithm.



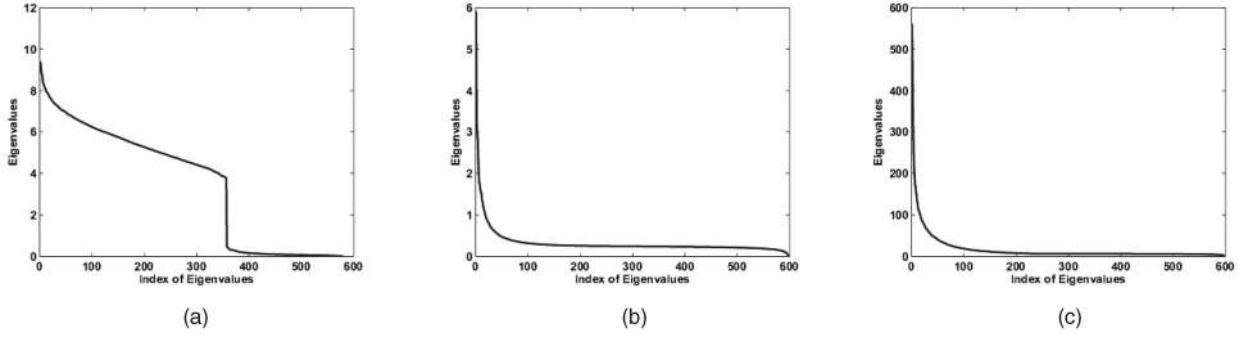


Fig. 8. The eigenvalues of the B matrix for the three face recognition algorithms: (a) baseline algorithm, (b) Bayesian algorithm, and (c) commercial algorithm. This distribution of eigenvalues provides an estimation of the dimension of the corresponding MDS space for each algorithm.

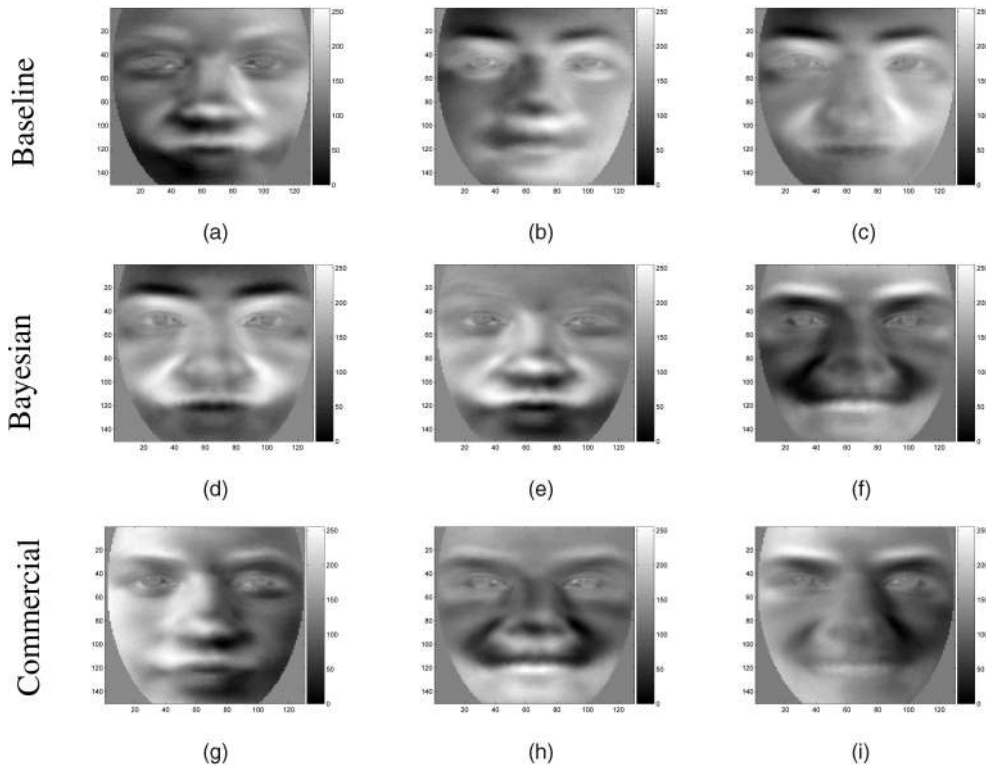


Fig. 9. Top three dimensions of the affine approximation to the three face recognition algorithms: (a) baseline algorithm, (b) Bayesian algorithm, and (c) commercial algorithm. The darker shades represent the variation captured along that particular dimension.

However, in the case of the Bayesian algorithm, it appears that eigenvalues do not drop to zero even with 600 images in the break-in set. Hence, we can expect the loss of sharp features in the reconstructed images for the Bayesian algorithm with 600 or less images in the break-in set. Similarly, for the commercial algorithm, we can expect a near optimal performance with 600 images in the break-in set. Fig. 9 represents the top three dimensions of the affine approximation of the three face recognition algorithms. These dimensions indicate the amount of variations (darker shades) captured by the affine transformation along the corresponding dimensions. Although individual dimensions of the affine transformation for each algorithm differs, a collective observation at the top three dimensions reveals that each algorithm tries to capture a similar variation in face images in the first three dimensions of the respective affine transformation. For example, the second dimension of affine transformation of the commercial algorithm (Fig. 9h) is similar to the third dimension of affine

transformation of the Bayesian algorithm (Fig. 9f). Similarly, the first dimension of affine transformation for the baseline algorithm (Fig. 9a) captures similar variations as that of the second dimension of affine transformation for the Bayesian algorithm (Fig. 9e). To quantify the modeling error, we compute the euclidean distance between the projected images in the affine space and compare it with the actual distance matrices computed by the respective algorithms after the correction of the additive constant factor. The normalized error  $\varepsilon$  is then computed as follows:

$$\varepsilon = \frac{\tilde{d}_{ij} - d_{ij}}{d_{ij}},$$

where  $\tilde{d}_{ij}$  represents the euclidean distance between projected images  $i$  and  $j$  in the affine space, and  $d_{ij}$  represents the actual distance computed by the recognition algorithm. We observe that the mean of the normalized



Fig. 10. Reconstructed face templates using a break-in set with 600 images: The first row represents the original templates. The second, third, and fourth rows represent the reconstructed templates for the baseline, Bayesian, and commercial algorithms, respectively.

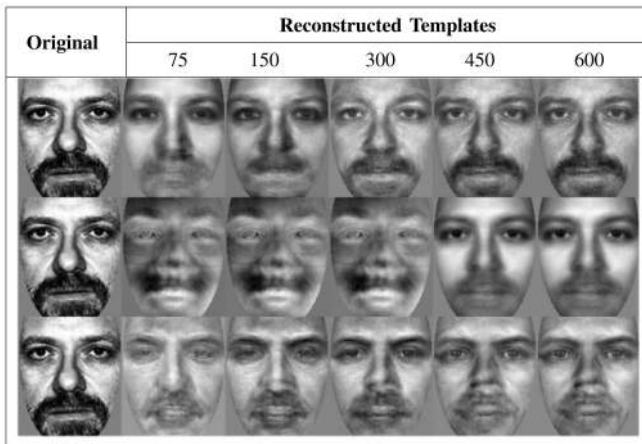


Fig. 11. Variation in reconstructed templates for the three algorithms using five different break-in sets with 75, 150, 300, 450, and 600 images. The first column represents the original template of the targeted subject. The first, second, and third rows represent the reconstructed templates for the baseline, Bayesian, and commercial algorithms, respectively.

errors  $\varepsilon$  are 0.002, 0.0457, and 0.1881 with standard deviations of 0.1563, 0.0915, and 0.2554 for the baseline, Bayesian, and commercial algorithms, respectively.

### 5.3 Reconstruction and Break-In

To study the effect of the number of images in the break-in set on the quality of reconstructed templates and break-in performance, we created five different break-in sets from the FRGC training set. Two break-in sets contain 75 and 150 images with one image per subject, and the other three break-in sets contain 300, 450, and 600 images with multiple images per subject. Sample images from the break-in set and gallery set are shown in Fig. 7. We reconstructed all the 1,196 images in the FERET gallery set using each of the five break-in sets. In Fig. 10, we present some of the reconstructed images using a break-in set with 600 images. In Fig. 11, we show the reconstruction templates of a particular targeted subject with all the five break-in sets. As expected, the reconstruction of the targeted subject's template improves with the number of images in the break-in sets. The noise in the reconstructed images is due to the fact that the break-in set and

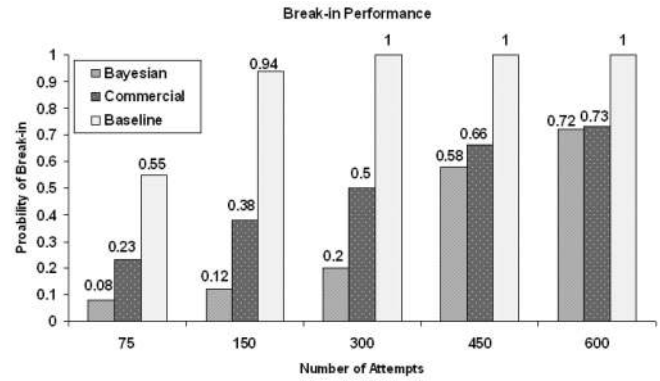


Fig. 12. Probability of break-in using five different break-in sets for the three algorithms at 1 percent FAR on the FERET gallery set.

gallery set are from two distinct databases collected in a totally different environment. In the case of the Bayesian algorithm, the reconstructed images appear much smoother than the original image. As mentioned earlier, the Bayesian algorithm requires more than 600 numbers of images in the break-in set for better reconstruction. To quantify the performance of breaking into a system with reconstructed images, we compute the *probability of break-in*, which is defined as the *probability of breaking a randomly chosen targeted subject*:



















$$\text{Prob. of break-in} = \frac{\text{No. of successfully verified targeted subject using reconstructed images}}{\text{Total no. of enrolled subjects}} \quad (17)$$

The reconstructed templates of the targeted subject are matched against the original templates of the targeted subject, and a dissimilarity matrix is generated for each of the three algorithms. From each of these dissimilarity matrices, we compute the TAR at 1 percent FAR from the respective Receiver Operating Characteristic (ROC) curves. The probability of breaking into any particular face recognition system is computed as the ratio between the number of accounts successfully broken at 1 percent FAR and the total number of accounts. Note that this performance measure for break-in is different from that used in other related works [18], [20], where the number of attempts is considered as a measure of the breaking into a security system. In our case, the number of attempts is fixed and the same as the number of images in the break-in set. In Fig. 12, we demonstrate the trade-off between probabilities of break-in for the system and the number of attempts, which is the same as the number of images in the break-in set. For the baseline algorithm, only 300 attempts is sufficient to achieve a 100 percent success rate to break into the system. For the feature-based commercial algorithm, 600 numbers of attempts are needed to break into the system with a probability of 0.73. For the template-based Bayesian algorithm, 600 attempts are required to break into the system with a probability of 0.72. Note that at 1 percent FAR, the probability of break-in with any random template after 600 attempts is 0.01.

### 5.4 Comparison with the Hill Climbing Approach

In this section, we compare our proposed template reconstruction scheme with a hill climbing-based approach [17], [18] on the commercial face recognition system. The eigenfaces required to modify the previous template in a hill climbing approach are created using 600 images from the break-in set. At each iteration, a randomly selected eigenface

TABLE 1  
Comparison of Reconstructed Template Using Our Approach against the Hill-Climbing Approach

Target Subject	Hill Climbing Approach			Our Approach	
	Initial Guess	300 Attempt	600 Attempt	300 Attempt	600 Attempt
 Easy	 Reject	 Reject	 Accept	 Accept	 Accept
 Moderate	 Reject	 Reject	 Reject	 Reject	 Accept
 Difficult	 Reject	 Reject	 Reject	 Reject	 Reject

is added or subtracted from the previous template. Due to the computational demand of the hill climbing process, we restrict our version of the hill climbing method to the first 100 subjects of the FERET gallery set, and a maximum of 600 attempts are allowed per subject. The commercial algorithm is set to operate at 1 percent FAR with 99 percent TAR, and we let the system decide the acceptance or rejection of a probe template based on this operational setup. We count the number of targeted subjects that are successfully broken by the hill climbing method and compare that with the number of successfully accepted reconstructed templates using our break-in set with 600 images. It should be noted that once we reconstruct a targeted subject’s face template, we treat the reconstructed template as an original face template and match it with the gallery set. This comparison shows the efficiency of our approach against the hill climbing approach after 600 iterations. In Table 1, we present few reconstructed templates from the hill climbing approach at 300 and 600 iterations and the corresponding reconstructed templates with our approach using the same number of comparisons. In the first column in Table 1, we show three different targeted subjects enrolled with templates marked as easy, moderate, and hard accounts to break in. The first row in Table 1 represents a targeted subject (*easy*) whose account is broken by both the hill climbing approach and our approach. However, it should be noted that the hill climbing approach requires 600 attempts to break into this easy account, whereas the same result can be achieved with only 300 iterations using our proposed scheme. Similarly, in the second row in Table 1, we present a targeted subject (*moderate*) whose account cannot be broken by the hill climbing approach after 600 attempts, but the proposed scheme successfully broke that account with 600 attempts. Finally, in the third row, we present a targeted subject (*hard*) whose account cannot be hacked by either scheme.

In Fig. 13, we compare the overall break-in performance of both schemes using the first 100 subjects from the FERET gallery set. Note that the probability of breaking into the system with any random template is equal to the FAR of the system, which is 0.01 in all of our experiments. We can observe that the proposed scheme has a 47 percent higher chance of breaking into a random account compared to the hill climbing attack with 600 attempts. It is worth to mention here that in [17], Adler shows that this particular hill climbing-based approach requires approximately 3,000 to 4,000 iterations to successfully break an account, which is much higher compared to the 600 iterations we used here. This count does not include the comparisons needed during the modeling procedure, which is done offline.

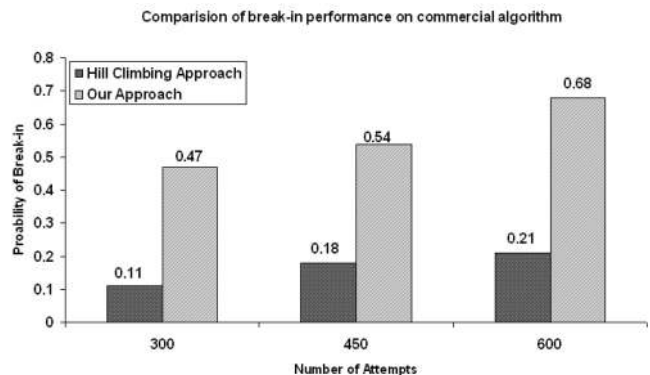


Fig. 13. Comparison of the probability of break-in at 1 percent FAR for the commercial algorithm with the first 100 subjects on the FERET gallery set. The commercial algorithm is set to operate at a predefined threshold such that TAR = 99 percent at 1 percent FAR.

## 5.5 Score Quantization

One countermeasure to the first-generation hill climbing attack is to quantize match scores. The systems with quantized match scores do not alter the output match scores with small changes in input images, which can prevent general hill climbing attacks. In such cases, if two similar probe templates, when matched with a template of the targeted subject, have the original match scores, say, 42.56 and 43.4, in the range  $[0, 100]$  and if the system quantizes the output match scores to the nearest integer (round-off), then both the scores will be quantized to 43. For such type of quantized scores, a hill climbing-based approach will fail to observe the improvement in the modified template and will fail to regenerate a template for the next iteration. However, such quantization of match scores has minimal effect on the proposed break-in scheme. Though, in [19], Adler proposed a modified hill climbing approach for systems with quantized match scores, our version of the hill climbing approach failed with quantized match scores and, therefore, we did not compare the break-in performance of the hill climbing approach with our approach on quantized match scores. In our proposed scheme, we compare different face templates to the targeted subject and do not need to observe any improvement in the match scores; hence, the proposed scheme is robust to the system with quantized match scores. In this experiment, we compute the probability of break-in using quantized match scores for the commercial face recognition system. We define a quantization index  $Q_s$  that controls the level of quantization:

$$S_{quant} = \left\lfloor \frac{S_{orig} - S_{min}}{\Delta S} \right\rfloor \cdot \Delta S + S_{min}, \quad (18)$$

$$Q_s = \frac{\Delta S}{(S_{max} - S_{min})},$$

where  $S_{orig}$ ,  $S_{quant}$ ,  $S_{max}$ , and  $S_{min}$  represent the original, the quantized, the minimum, and the maximum match scores of the recognition system, respectively. In (19), the parameter  $\Delta S$  controls the level of the quantization of the original scores and is defined as the length of the quantized intervals, that is, the difference between two successive quantized scores. To be consistent with the variable range of match scores for different algorithms, we define quantization index  $Q_s$  by normalizing  $\Delta S$  over a possible range of match scores of a recognition system. If the quantization index is set to 0.1, then the original scores are quantized at 10 different points, and if  $Q_s$  equals to 0.01, then the original scores are quantized at 100 different points. For this experiment, we use four

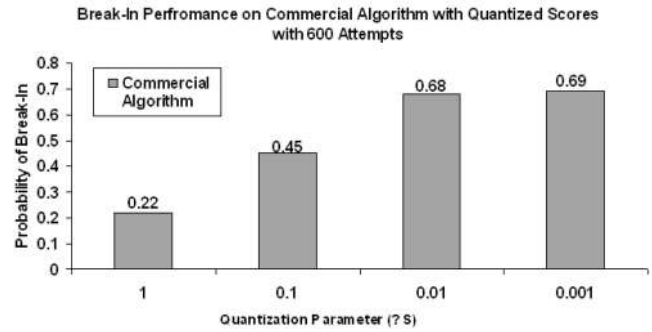







Fig. 14. Probability of break-in at 1 percent FAR for the commercial algorithm with quantized scores. The value of quantization parameter  $\Delta S$  in (19) is set to 0.001, 0.01, 0.1, and 1 to quantize the original match score at four different levels.

different levels of quantization by setting the value of  $Q_s$  to 0.0001, 0.001, 0.01, and 0.1. Fig. 14 shows the probability of break-in at 1 percent FAR for the commercial algorithm with the proposed quantization of match scores. We observe that the probabilities of break-in do not change significantly when the original scores quantized with  $Q_s$  equals to 0.0001 or 0.001, and the probability of break-in drops from 0.68 to 0.45 when  $Q_s$  equals to 0.01. However, we can observe that for  $Q_s$  equal to 0.1, the probability of break-in dropped from 0.45 to 0.22. In Table 2, we demonstrate the effect of quantization on a reconstructed template along with the acceptance/rejection decision from the system using that particular reconstructed template. As we can observe, with an increasing value of  $Q_s$ , the quality of the reconstructed template starts to degrade and is eventually rejected by the system. If the system outputs a very high level of quantized scores, for example, with  $Q_s = 0.1$ , then the original match scores are highly distorted, and the affine modeling of the underlying algorithm is erroneous, and as a result, the overall break-in performance is affected. However, it should be observed that such quantization of match scores has a trade-off with the operational flexibility of a system. For example, if the recognition system, with the range of original scores in the interval  $[0, 100]$ , quantizes the original scores at 10 different points with  $Q_s$  equal to 0.1 (that is, output scores as a multiplier of 10), then the system is restricted to operate only at these 10 distinct operational points (thresholds) and lose the flexibility to operate at any intermediate threshold or FARs.

TABLE 2  
Effect of Quantization of Match Scores on Reconstructed Templates

Target Subject	Quantization Index $Q_s$			
	0.0001	0.001	0.01	0.1
				
	Accept	Accept	Reject	Reject

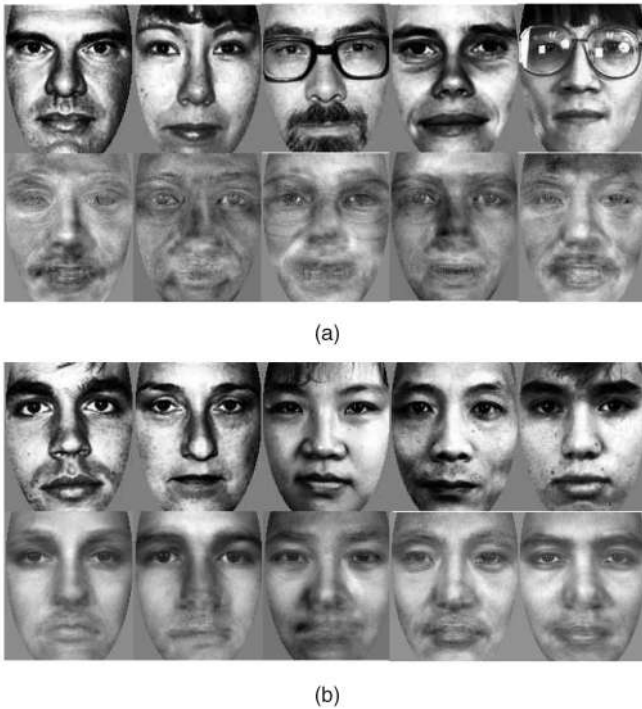


Fig. 15. Easy/difficult face images for the commercial algorithm. (a) Sample of difficult target subjects that cannot be broken with 600 attempts. (b) Sample of easy target subjects that can be broken with only 75 attempts.

## 6 CONCLUSION

In this paper, we discussed a novel scheme to reconstruct face images from match scores and exposes a potential source for security breach in the face recognition systems. We used an affine transformation to approximate the behavior of the face recognition system using an independent set of face templates termed as the break-in set. Selected subsets of templates in the break-in set were then matched only once with the enrolled templates of the targeted subject. Given the distances of the targeted subject's template, we embedded that template in the learned affine space and inverted the modeling affine transformation to arrive at the original template. We used three fundamentally different face recognition algorithms and observed that with the proposed scheme, only 300 attempts were required to achieve a 100 percent probability of breaking into the baseline face recognition algorithm, and 600 attempts were required for the Bayesian algorithm to achieve a 72 percent success. For the commercial algorithm, we achieved a 73 percent success rate to break into the system with 600 attempts. This observation leads us to investigate further on the easiness/hardness property of a particular target subject. In Fig. 15, we presented few target subjects that were hard to break with 600 numbers of attempts, and we showed few target subjects that were easy to break with only 75 attempts. It will be interesting to investigate further on the hardness/easiness of a particular target subject and the abilities to quantize any face template in terms of a hard/easy template to be used to sneak into a system.

A cursory look at match scores from a biometric system may not appear to be a weak link in terms of security and privacy issues; however, with our proposed scheme, we revealed that even match scores carry sufficient information for the reverse engineering of the original templates and should be protected in the same way as the original

templates. The major advantage of the proposed scheme over the earlier proposed hill climbing attack is that it is not based on a local search, and the number of attempts is less. Using two different data sets for gallery and break-in templates, we demonstrated that our proposed modeling scheme is also generalized across databases. Our scheme uses distinct templates in each attempt when compared to a targeted subject. Therefore, such an attack is difficult to detect automatically and cannot be neutralized by a simple quantization of match scores. Thus, future face recognition systems should emphasize issues related to the privacy of the face template and system robustness to such types of attacks. Recently, *cancelable biometrics* have been proposed to encrypt both gallery and probe templates in an attempt to restrict the unauthorized access of biometric templates [35]. Similarly, in [36], Boulton proposed *revocable biometrics* with a robust distance measure where face templates are encrypted, along with the redistribution of match scores, within a predefined range for each subject in the gallery set. These developments have the potential to be used as countermeasures to our proposed template reconstruction scheme; however, this remains to be demonstrated.

## ACKNOWLEDGMENTS

This research was funded in part by Florida I4 Hightech Corridor Funds, Unisys Corp., and the Intelligence Technology Innovation Center (ITIC) within the US Central Intelligence Agency (CIA).

## REFERENCES

- [1] S. Prabhakar, S. Pankanti, and A. Jain, "Biometric Recognition: Security and Privacy Concerns," *IEEE Security and Privacy Magazine*, vol. 1, no. 2, pp. 33-42, Apr. 2003.
- [2] T. Matsumoto, H. Matsumoto, K. Yamada, and S. Hoshino, "Impact of Artificial Gummy Fingers on Fingerprint Systems," *Proc. SPIE, Optical Security and Counterfeit Deterrence Techniques IV*, vol. 4667, pp. 275-289, 2002.
- [3] A.K. Jain and S. Li, *Handbook of Face Recognition*. Springer, 2005.
- [4] G. Hachez, F. Koeune, and J. Quisquater, "Biometrics, Access Control, Smart Cards: A Not So Simple Combination," *Proc. Fourth Smart Card Research and Advanced Applications Conf.*, vol. 4, pp. 273-288, Sept. 2000.
- [5] S. Schimke, C. Vielhauer, and T. Kalker, "Security Analysis for Biometric Data in ID Documents," *Proc. SPIE-IS&T Symp. Electronic Imaging*, vol. 5681, pp. 474-485, 2005.
- [6] C. Soutar, D. Roberge, S.A. Stojanov, R. Gilroy, and B.V.K. Vijaya, "Biometric Encryption Using Image Processing," *Proc. SPIE, Optical Security and Counterfeit Deterrence Techniques II*, vol. 3314, pp. 178-188, 1998.
- [7] U. Uludag, S. Pankanti, S. Prabhakar, and A.K. Jain, "Biometric Cryptosystems: Issues and Challenges," *Proc. IEEE, special issue on enabling security technologies for digital rights management*, vol. 92, no. 6, pp. 948-960, June 2004.
- [8] N. Ratha, J. Connell, and R. Bolle, "An Analysis of Minutiae Matching Strength," *Proc. Int'l Conf. Audio and Video-Based Biometric Person Authentication*, 2001.
- [9] BioAPI, *BioAPI 2.0—International Version*, BioAPI Consortium, <http://www.bioapi.org/internationalversion.html>, 2005.
- [10] R. Bolle, J.H. Connell, S. Pankanti, N.K. Ratha, and A.W. Senior, *Guide to Biometrics*. Springer, 2004.
- [11] S. Nanavati, M. Thieme, and R. Nanavati, *Biometrics: Identity Verification in a Networked World*. John Wiley & Sons, 2005.
- [12] P. Phillips, H. Wechsler, J. Huang, and P.J. Rauss, "The FERET Database and Evaluation Procedure for Face Recognition Algorithms," *Image and Vision Computing*, vol. 16, pp. 295-306, 1998.
- [13] P.J. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the Face Recognition Grand Challenge," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 947-954, June 2005.

- [14] M.A. Turk and P. Pentland, "Face Recognition Using Eigenfaces," *Proc. IEEE Conf. Conf. Computer Vision and Pattern Recognition*, pp. 586-591, June 1991.
- [15] B. Moghaddam and A. Pentland, "Beyond Eigenfaces: Probabilistic Matching for Face Recognition," *Proc. Int'l Conf. Automatic Face and Gesture Recognition*, pp. 30-35, 1998.
- [16] C. Soutar, "Biometric System Security," *Secure*, vol. 5, pp. 46-49, 2002.
- [17] A. Adler, "Images Can Be Regenerated from Quantized Biometric Match Score Data," *Proc. Canadian Conf. Electrical and Computer Eng.*, pp. 469-472, May 2004.
- [18] U. Uludag and A. Jain, "Attacks on Biometric Systems: A Case Study in Fingerprints," *Proc. SPIE-EI 2004, Security, Steganography and Watermarking of Multimedia Contents*, pp. 622-633, Jan. 2004.
- [19] A. Adler, "Vulnerabilities in Biometric Encryption System," *Proc. Int'l Conf. Audio and Video-Based Biometric Person Authentication*, pp. 1100-1109, July 2005.
- [20] D. Lopresti and J. Raim, "The Effectiveness of Generative Attacks on an Online Handwriting Biometric," *Proc. Int'l Conf. Audio and Video-Based Biometric Person Authentication*, pp. 1090-1099, July 2005.
- [21] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, July 1997.
- [22] P. Comon, "Independent Component Analysis, a New Concept?" *Signal Processing*, vol. 36, no. 3, Apr. 1994.
- [23] L. Wiskott, J. Fellous, N. Kruger, and C. Malsburg, "Face Recognition by Elastic Bunch Graph Matching," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 1160-1169, July 1997.
- [24] X. Liu, A. Srivastava, and K. Gallivan, "Optimal Linear Representations of Images for Object Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26, no. 5, pp. 662-666, May 2004.
- [25] P. Mohanty, S. Sarkar, and R. Kasturi, "Designing Affine Transformations Based Face Recognition Algorithms," *Proc. IEEE Workshop Face Recognition Grand Challenge*, June 2005.
- [26] E. Pekalska, P. Paclik, and R. Duin, "A Generalized Kernel Approach to Dissimilarity-Based Classification," *J. Machine Learning Research*, vol. 2, no. 2, pp. 175-211, 2002.
- [27] J. Tenenbaum, V. Silva, and J. Langford, "A Global Geometric Framework for Nonlinear Dimensionality Reduction," *Science Magazine*, vol. 290, no. 5500, 2000.
- [28] S. Roweis and L. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," *Science Magazine*, vol. 290, no. 5500, 2000.
- [29] T. Cox and M. Cox, *Multidimensional Scaling*, second ed. Chapman and Hall, 1994.
- [30] I. Borg and P. Groenen, "Modern Multidimensional Scaling," *Springer Series in Statistics*, Springer, 1997.
- [31] J. Gower and P. Legendree, "Metric and Euclidean Properties of Dissimilarity Coefficients," *J. Classification*, vol. 3, pp. 5-48, 1986.
- [32] V. Roth, J. Laub, M. Kawanabe, and J.M. Buhmann, "Optimal Cluster Preserving Embedding of Nonmetric Proximity Data," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1540-1551, Dec. 2003.
- [33] R. Beveridge, D. Bolme, M. Teixeira, and B. Draper, "The CSU Face Identification Evaluation System," *Machine Vision and Applications*, vol. 16, no. 2, pp. 128-138, 2005.
- [34] D. Bolme, "The Bayesian Intrapersonal/Extrapersonal Classifiers," master's thesis, Colorado State Univ., Computer Science Dept., July 2003.
- [35] N. Ratha, J. Connell, and R. Bolle, "Enhancing Security and Privacy in Biometrics-Based Authentication Systems," *IBM Systems J.*, vol. 40, no. 3, 2001.
- [36] T. Boulton, "Robust Distance Measures for Face-Recognition Supporting Revocable Biometric Tokens," *Proc. Int'l Conf. Automatic Face and Gesture Recognition*, pp. 560-566, 2006.



**Pranab Mohanty** received the MS degree in mathematics from the Utkal University, Orissa, India, in 1997 and the MS degree in computer science from the Indian Statistical Institute, Calcutta, India, in 2000. He is currently a PhD candidate at the University of South Florida. His research interests include biometrics, image and video processing, computer vision, and pattern recognition. He is a student member of the IEEE.



**Sudeep Sarkar** received the BTech degree in electrical engineering from the Indian Institute of Technology, Kanpur, India, in 1988 and the MS and PhD degrees in electrical engineering, on a University Presidential Fellowship, from Ohio State University, Columbus, in 1990 and 1993, respectively. Since 1993, he has been with the Computer Science and Engineering Department at the University of South Florida, Tampa, where he is currently a professor. His research inter-

ests include perceptual organization in single images and multiple image sequences, biometrics, gait recognition, color-texture analysis, and performance evaluation of vision systems. He has coauthored one book and coedited another book on perceptual organization. He is the recipient of the US National Science Foundation Faculty Early Career Development (CAREER) award in 1994, the University of South Florida (USF) Teaching Incentive Program Award for undergraduate teaching excellence in 1997, the Outstanding Undergraduate Teaching Award in 1998, and the Theodore and Venette Askounes-Ashford Distinguished Scholar Award in 2004. He served on the editorial boards for the *IEEE Transactions on Pattern Analysis and Machine Intelligence* (1999-2003) and *Pattern Analysis & Applications Journal* (2000-2001). He currently serves on the editorial board of the *Pattern Recognition* and the *IEEE Transactions on Systems, Man, and Cybernetics, Part-B*. He is a senior member of the IEEE and the IEEE Computer Society.



**Rangachar Kasturi** received the BE (electrical) degree from Bangalore University, India, in 1968 and the MSEE and PhD degrees from Texas Tech University in 1980 and 1982, respectively. He was a professor of computer science and engineering and electrical engineering at the Pennsylvania State University during 1982-2003 and was a Fulbright Scholar during 1999. He has been elected to serve as the 2008 President of the IEEE Computer Society. He was the

President of the International Association for Pattern Recognition (IAPR) during 2002-2004. He has served as the editor in chief of the *IEEE Transactions on Pattern Analysis and Machine Intelligence* and the *Machine Vision and Applications* journals. He has received the Penn State Engineering Society Premier Research Award and has been inducted into the Texas Tech Electrical Engineering Academy. His research interests are in computer vision and pattern recognition. He is an author of the textbook *Machine Vision* and has published numerous papers and research reference books. He has directed many research projects in document image analysis, video sequence analysis, and biometrics. In particular, he is directing a project that evaluates research progress in detection and tracking of faces, people, text, and vehicles in video sequences. He is a fellow of the IEEE and a fellow of the International Association for Pattern Recognition (IAPR).

► For more information on this or any other computing topic, please visit our Digital Library at [www.computer.org/publications/dlib](http://www.computer.org/publications/dlib).