# Matching Forensic Sketches to Mug Shot Photos

Brendan F. Klare, *Student Member*, *IEEE*,
Zhifeng Li, *Member*, *IEEE*, and
Anil K. Jain, *Fellow*, *IEEE*

**Abstract**—The problem of matching a forensic sketch to a gallery of mug shot images is addressed in this paper. Previous research in sketch matching only offered solutions to matching highly accurate sketches that were drawn while looking at the subject (viewed sketches). Forensic sketches differ from viewed sketches in that they are drawn by a police sketch artist using the description of the subject provided by an eyewitness. To identify forensic sketches, we present a framework called local feature-based discriminant analysis (LFDA). In LFDA, we individually represent both sketches and photos using SIFT feature descriptors and multiscale local binary patterns (MLBP). Multiple discriminant projections are then used on partitioned vectors of the feature-based representation for minimum distance matching. We apply this method to match a data set of 159 forensic sketches against a mug shot gallery containing 10,159 images. Compared to a leading commercial face recognition system, LFDA offers substantial improvements in matching forensic sketches to the corresponding face images. We were able to further improve the matching performance using race and gender information to reduce the target gallery size. Additional experiments demonstrate that the proposed framework leads to state-of-the-art accuracys when matching viewed sketches.

**Index Terms**—Face recognition, forensic sketch, viewed sketch, local feature discriminant analysis, feature selection, heterogeneous face recognition.

✦

## 1 INTRODUCTION

PROGRESS in biometric technology has provided law enforcement agencies additional tools to help determine the identity of criminals. In addition to DNA and circumstantial evidence, if a latent fingerprint is found at an investigative scene or a surveillance camera captures an image of a suspect's face, then these cues may be used to help determine the culprit's identity using automated biometric identification. However, many crimes occur where none of this information is present, but instead an eyewitness account of the crime is available. In these circumstances, a forensic artist is often used to work with the witness in order to draw a sketch that depicts the facial appearance of the culprit according to the verbal description. Once the sketch image of the transgressor is complete, it is then disseminated to law enforcement officers and media outlets with the hopes of someone knowing the suspect. These sketches are known as forensic sketches and this paper describes a robust method for matching forensic sketches to large mug shot (image) databases maintained by law enforcement agencies.

Two different types of face sketches are discussed in this paper: *viewed sketches* and *forensic sketches* (see Fig. 1). *Viewed sketches* are sketches that are drawn while viewing a photograph of the person

---

- *B.F. Klare and Z. Li are with the Department of Computer Science and Engineering, Michigan State University, 3208 Engineering Building, East Lansing, MI 48823. E-mail: {klarebre, zfli}@msu.edu.*
- *A.K. Jain is with the Department of Computer Science and Engineering, Michigan State University, 3208 Engineering Building, East Lansing, MI 48823, and the Department of Brain and Cognitive Engineering, Korea University, Seoul, Korea. E-mail: jain@cse.msu.edu.*

or the person himself. *Forensic sketches* are drawn by interviewing a witness to gain a description of the suspect. Published research on sketch to photo matching to this point has primarily focused on matching viewed sketches [1], [2], [3], [4], [5], despite the fact that real-world scenarios only involve forensic sketches. Both forensic sketches and viewed sketches pose challenges to face recognition due to the fact that probe sketch images contain different textures compared to the gallery photographs they are being matched against. However, forensic sketches pose additional challenges due to the inability of a witness to exactly remember the appearance of a suspect and her subjective account of the description, which often results in inaccurate and incomplete forensic sketches.

We highlight two key difficulties in matching forensic sketches: 1) matching across image modalities and 2) performing face recognition despite possibly inaccurate depictions of the face. In order to solve the first problem, we use *local feature-based discriminant analysis (LFDA)* to perform minimum distance matching between sketches and photos, which is described in Section 3 and summarized in Fig. 2. The second problem is considered in Section 5, where analysis and improvements are offered for matching forensic sketches against large mug shot galleries.

The contributions of the paper are summarized as follows:

1. We observe a substantial improvement in matching viewed sketches over published algorithms using the proposed local feature-based discriminant analysis.
2. We present the first large-scale published experiment on matching real forensic sketches.
3. Using a mug shot gallery of 10,159 images, we perform race and gender filtering to improve the matching results.
4. All experiments are validated by comparing the proposed method against a leading commercial face recognition engine.

The last point is significant since earlier studies on viewed sketches used PCA (eigenface) matcher as the baseline. It is now well known that the performance of PCA matcher can be easily surpassed by other face matchers.

## 2 RELATED WORK

Most research on sketch matching has dealt with viewed sketches. Much of the early work in matching viewed sketches was performed by Tang et al. [6], [2], [3], [5], [7]. These studies share a common approach in that a synthetic photograph is generated from a sketch (or vice versa) and standard face recognition algorithms are then used to match the synthetic photographs to gallery photographs. The different synthesis methods used include an eigentransformation method (Tang and Wang [6], [2]), local linear embedding (Liu et al. [3]), and belief propagation on a Markov random field (Wang and Tang [5]). Other synthesis methods have been proposed as well [4], [8], [9], [10], [11]. The impact of matching sketches drawn by different artists was studied by Al Nizami et al. [12].

Klare and Jain [1] proposed a method of sketch matching that uses the same feature-based approach that has been successful in other heterogeneous face recognition scenarios (specifically, matching near-infrared face images to visible light). In using SIFT feature descriptors [13], the intrapersonal variations between the sketch and photo modality were diminished while still maintaining sufficient information for interclass discrimination. Such an approach is similar to other methods proposed in the literature [14], [15], [16] of matching near-infrared images (NIR) to visible light images (VIS), where local binary pattern [17] feature descriptors are used to describe both NIR and VIS images.

In this paper, we extend our previous feature-based approach to sketch matching [1]. This is achieved by using local binary patterns (LBP) in addition to the SIFT feature descriptor, which is motivated by LBP's success in a similar heterogeneous matching application by Liao et al. [14]. Additionally, we extend our feature-based matching
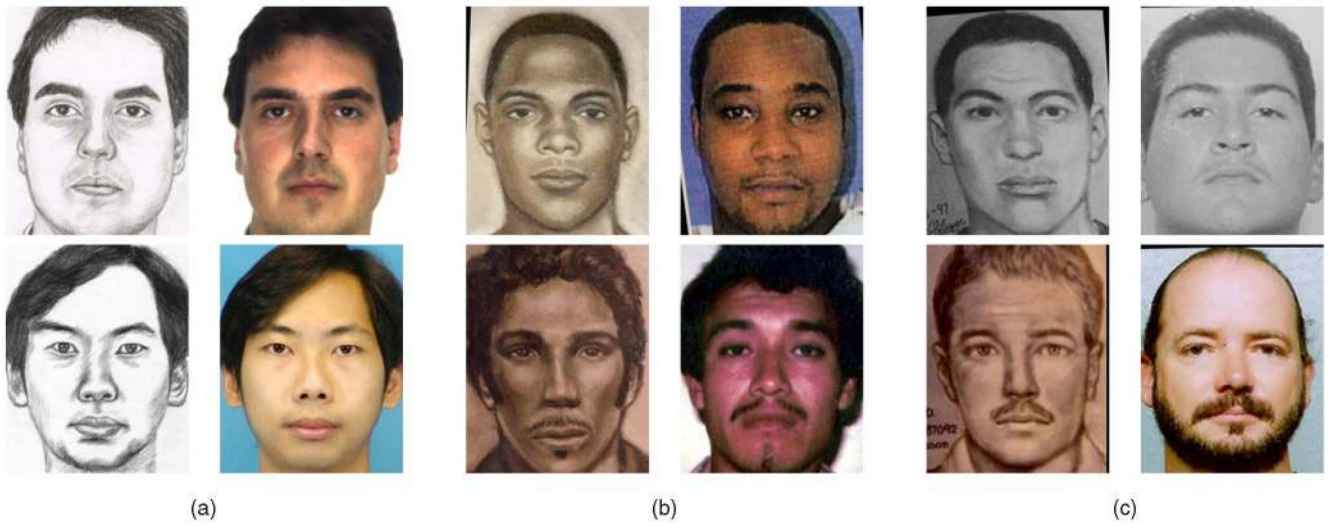
Fig. 1. The difference between viewed sketches and forensic sketches. (a) Viewed sketches and their corresponding photographs, (b) two pairs of good quality forensic sketches and the corresponding photographs, and (c) two pairs of poor quality forensic sketches and the corresponding photographs.

to learn discriminant projections on "slices" of feature patches, which is similar to the method proposed by Lei and Li [15].

## 3    FEATURE-BASED SKETCH MATCHING

Image feature descriptors describe an image or image region using a feature vector that captures the distinct characteristics of the image [18]. Image-based features have been shown to be successful in face recognition, most notably with the use of local binary patterns [19].

### 3.1    Feature-Based Representation

We will now describe how to represent a face with image descriptors. Because most image descriptors are not sufficiently verbose to fully describe a face image, the descriptors are computed
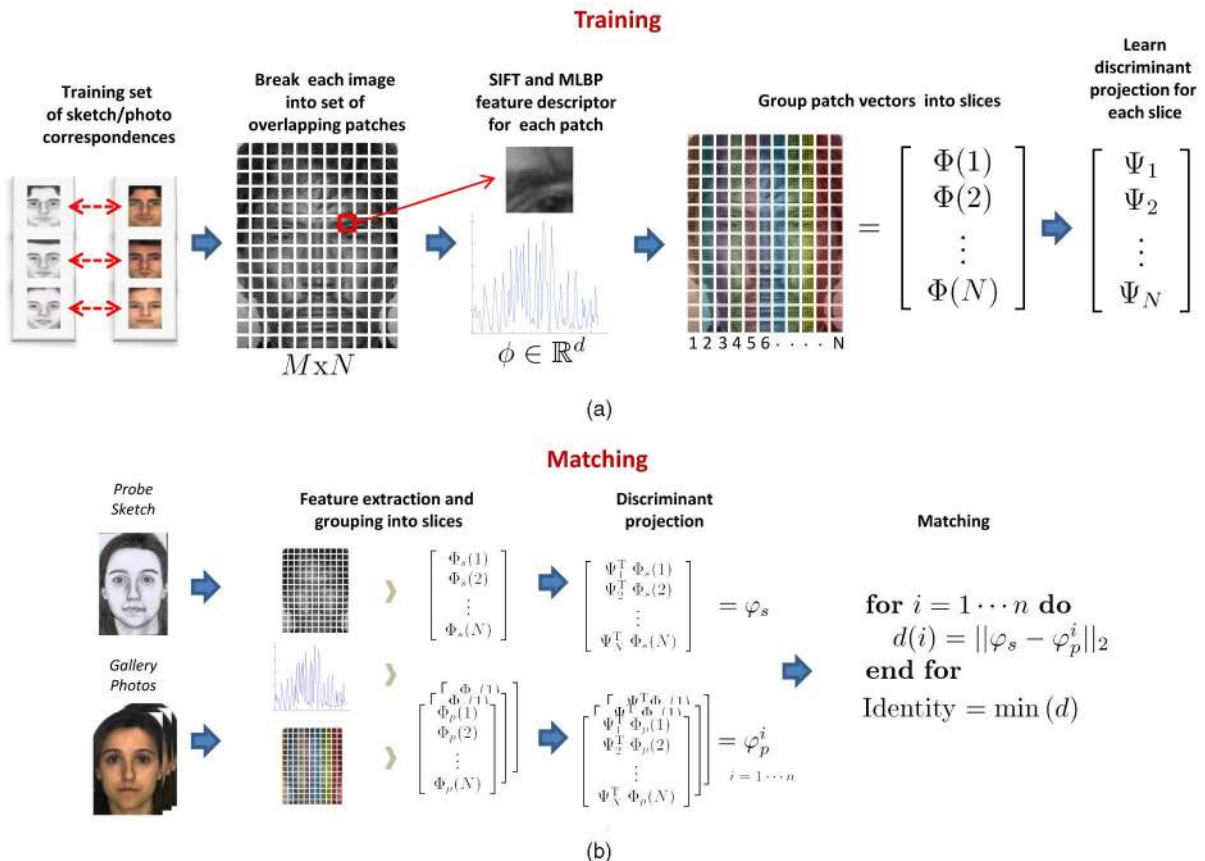


Fig. 2. An overview of the (a) training and (b) recognition using the *LFDA* framework. Each sketch and photo are represented by SIFT and MLBP feature descriptors extracted from overlapping patches. After grouping "slices" of patches together into feature vectors $\Phi(k)(k = 1 \cdots N)$, we learn a discriminant projection $\Psi_k$ for each slice. Recognition is performed after combining each projected vector slice into a single vector $\varphi$ and measuring the normed distance between a probe sketch and a gallery photo.

over a set of uniformly distributed subregions of the face. The feature vectors at sampled regions are then concatenated together to describe the entire face. The feature sampling points are chosen by setting two parameters: a region (or patch) size $s$ and a displacement size $\delta$. The region size $s$ defines the size of the square window over which the image feature is computed. The displacement size $\delta$ states the number of pixels the patch is displaced for each sample; thus, $(s - \delta)$ is the number of overlapping pixels in two adjacent patches. This is analogous to sliding a window of size $s \times s$ across the face image in a raster scan fashion. For an $H \times W$ image, the number of horizontal ($N$) and vertical ($M$) sampling locations is given by $N = (W - s)/\delta + 1$ and $M = (H - s)/\delta + 1$. At each of the $M \cdot N$ patches, we compute the $d$-dimensional image feature vector $\phi$. These image feature vectors are concatenated into one single $(M \cdot N \cdot d)$-dimensional image vector $\Phi$. Whereas $f(I) : I \rightarrow \phi$ denotes the extraction of a single feature descriptor from an image, sampling multiple features using overlapping patches is denoted as $F(I) : I \rightarrow \Phi$. Minimum distance sketch matching can be performed directly using this feature-based representation of subjects $i$ and $j$ by computing the normed vector distance $\|F(I^i) - F(I^j)\|$ [1].

In our sketch matching framework, two feature descriptors are used: SIFT and LBP. The SIFT feature descriptor quantizes both the spatial locations and gradient orientations within an $s \times s$-sized image patch, and computes a histogram in which each bin corresponds to a combination of a particular spatial location and orientation. For each image pixel, the histogram bin corresponding to its quantized orientation and location is incremented by the product of 1) the magnitude of the image gradient at that pixel and 2) the value of a Gaussian function centered on the patch with a standard deviation of $s/2$. Trilinear interpolation is used on the quantized location of the pixel, which addresses image translation noise. The final vector of histogram values is normalized to sum to one. The reader is referred to [13] for a more detailed description of how the SIFT feature descriptor is designed. It is important to reiterate that because we are sampling SIFT feature descriptors from a fixed grid and we do not use SIFT keypoint detection, the SIFT feature descriptor is computed at predetermined locations.

For the local binary pattern feature descriptor [17], we extended the LBP to describe the face at multiple scales by combining the LBP descriptors computed with radii $r \in \{1, 3, 5, 7\}$. We refer to this as the multiscale local binary pattern (MLBP). MLBP is similar to other variants of the LBP, such as MB-LBP [14], but we obtained slightly improved accuracy using MLBP.

The choice of the MLBP and SIFT feature descriptors was based on reported success in heterogeneous face recognition and through a quantitative evaluation of their ability to discriminate between subjects in sketches and photos [20]. Though variants of LBPs have led to substantial success in previous heterogeneous face recognition scenarios, the use of SIFT feature descriptors for this application is quite novel. However, recent work [1] clearly demonstrates the success of SIFT feature descriptors for viewed sketch recognition. SIFT feature descriptors have also been shown to perform comparatively with LBP feature descriptors in a standard face recognition scenario [21]. These feature descriptors are well-suited for sketch recognition because they describe the distribution of the direction of edges in the face; this is the information that both sketches and photos contain. By densely sampling these descriptors, sufficient discriminatory information is retained to more accurately determine a subject's identity over previously used synthesis methods [1].

The feature-based representation requires each sketch and photo image to be normalized by rotating the angle between the two eyes to 0 degree, scaling the images to a 75 interocular pixel distance, and cropping the image size to 200 by 250 pixels. The experimental results reported in Sections 4 and 6 for each of the

two descriptors are based on the sum of score fusion of the match scores generated from computing descriptors with patch sizes of $s = 16$ and $s = 32$. This also holds for the global discriminant described in Section 3.2; we fuse the matching scores computed using two separate patch sizes of 16 and 32. When combining the SIFT and MLBP features, the sum of score fusion is used as well.

## 3.2 Local Feature-Based Discriminant Analysis

With both sketches and photos characterized using SIFT and MLBP image descriptors, we further refine this feature space using discriminant analysis. This is done to reduce the large dimensionality of the feature vector $\Phi$. A straightforward approach would be to apply classical subspace analysis (such as LDA) directly on $\Phi$, and to extract discriminant features for classification. However, there are several problems with this approach. First, the feature dimensionality is too high for direct subspace analysis. In our experiments, each image is divided into either 154 overlapping patches (for $s = 32$) or 720 overlapping patches (for $s = 16$), with each patch producing a 128-dimensional SIFT descriptor or a 236-dimensional MLBP descriptor. The second problem is the possibility of overfitting due to the small sample size (SSS) [22].

In order to handle the combination of a large feature size and small sample size, an ensemble of linear discriminant classifiers called LFDA is proposed. Other discriminant analysis methods have been proposed to handle the SSS problem, such as random sampling LDA [23], regularized LDA [24], and direct LDA [25]. However, we chose the proposed LFDA method because it is designed to work with a feature descriptor representation (as opposed to an image pixel representation), and it resulted in high recognition accuracy.

In the LFDA framework, each image feature vector $\Phi$ is first divided into "slices" of smaller dimensionality, where slices correspond to the concatenation of feature descriptor vectors from each column of image patches. Next, discriminant analysis is performed separately on each slice by performing the following three steps: PCA, within class whitening, and between class discriminant analysis. Finally, PCA is applied to the new feature vector to remove redundant information among the feature slices to extract the final feature vector.

To train the LFDA, we use a training set consisting of pairs of a corresponding sketch and photo of $n$ subjects (which are the $n$ training classes). This results in a total of $2n$ training images with two supports for each subject $i$: the image feature representation of the sketch $\Phi_s^i = F(I_s^i)$ and the photo $\Phi_p^i = F(I_p^i)$. We combine these feature vectors as a column vector in training matrices and refer to them as $X^s = [\Phi_s^1 \, \Phi_s^2 \ldots \Phi_s^n]$ for the sketch, $X^p = [\Phi_p^1 \, \Phi_p^2 \ldots \Phi_p^n]$ for the photo, and $X = [\Phi_s^1 \ldots \Phi_s^n \, \Phi_p^1 \ldots \Phi_p^n]$ for the photo and sketch combined.

The first step in LFDA is to separate the image feature vector into multiple subvectors or slices. Given the $M \times N$ array of patches consisting of SIFT or MLBP descriptors, we create one slice for each of the $N$ patch columns. With a $d$-dimensional feature descriptor, each of the $N$ slices is of dimensionality $(M \cdot d)$. We call this a "slice" because it is similar to slicing an image into $N$ pieces. After separating the feature vectors into slices, the training matrices now become $X_k^s \in \mathbb{R}^{M \cdot d, n}$, $X_k^p \in \mathbb{R}^{M \cdot d, n}$, and $X_k \in \mathbb{R}^{M \cdot d, 2n}$ ($k = 1 \ldots N$), which are all mean-centered.

We next reduce the dimensionality of each training slice matrix $X_k$ using the PCA matrix $W_k \in \mathbb{R}^{M \cdot d, r}$ with $r$ eigenvectors. The purpose is to remove the noisy features which are usually associated with the trailing eigenvectors with the smallest eigenvalues. In our experiments, we use the 100 eigenvectors with the largest eigenvalues (which preserves about 90 percent of the variance). The discriminant extraction proceeds by generating the mean projected class vectors

TABLE 1
Rank-1 Recognition Rates for Matching Viewed Sketches Using the CUHK Public Data Set

| Baseline | | Without LFDA | | LFDA | |
|---|---|---|---|---|---|
| *Method* | *Rank-1 Accuracy (%)* | *Method* | *Rank-1 Accuracy (%)* | *Method* | *Rank-1 Accuracy (%)* |
| FaceVACS [26] | 90.37 | SIFT | 97.00 | SIFT | 99.27 |
| BP Synthesis [5] | 96.30 | MLBP | 96.27 | MLBP | 98.60 |
| SIFT Descriptor-based [1] | 97.87 | SIFT + MLBP | 97.33 | **SIFT + MLBP** | **99.47** |

*The standard deviation across the five random splits for each method in the middle and right columns is less than 1 percent.*

$$Y_k = W_k^{\mathrm{T}} \left( X_k^s + X_k^p \right)/2, \qquad (1)$$

which are used to center the sketch and photo training instances of each class by

$$\begin{aligned} \tilde{X}_k^s &= W_k^{\mathrm{T}} \ X_k^s - Y_k, \\ \tilde{X}_k^p &= W_k^{\mathrm{T}} \ X_k^p - Y_k. \end{aligned} \qquad (2)$$

To reduce the intrapersonal variation between the sketch and the photo, a whitening transform is performed. Whitening the within-class scatter matrix reduces the large feature dimensions that represent the principal intrapersonal variations, which in this case correspond to intrapersonal differences between sketches and photos. To do so, we recombine the training instances into $\tilde{X}_k = [\tilde{X}_k^s \ \tilde{X}_k^p]$. PCA analysis is performed on $\tilde{X}_k$ such that the computed PCA projection matrix $\tilde{V}_k \in \mathbb{R}^{100,100}$ retains all data variance from $\tilde{X}_k$. Let $\Lambda_k \in \mathbb{R}^{100,100}$ be a diagonal matrix whose entries are the eigenvalues of the corresponding PCA eigenvectors $\tilde{V}_k$. The whitening transform matrix is $V_k = (\Lambda_k^{-\frac{1}{2}} V_k^{\mathrm{T}})^{\mathrm{T}}$.

The final step is to compute a projection matrix that maximizes the intraperson scatter by performing PCA on $V^{\mathrm{T}} Y_k$ (which is the whitening transform of the mean class vectors). Using all but one of the eigenvectors in the PCA projection matrix, the resultant projection matrix is denoted as $U_k \in \mathbb{R}^{100,99}$. This results in the final projection matrix for slice $k$:

$$\Psi_k = W_k V_k U_k. \qquad (3)$$

With each local feature-based discriminant trained, we match sketches to photos using the nearest neighbor matching on the concatenated slice vectors. We first separate the feature representation of an image into individual slices

$$\Phi = [\Phi(1)^{\mathrm{T}} \ \Phi(2)^{\mathrm{T}} \ldots \Phi(N)^{\mathrm{T}} ]^{\mathrm{T}}, \qquad (4)$$

where $\Phi(i) \in \mathbb{R}^{M \cdot d}$ is the $i$th slice feature vector. We then project each slice using the LFDA projection matrix $\Psi_k$, yielding the new vector representation $\varphi \in \mathbb{R}^{M \cdot 99}$:

$$\varphi = [ (\Psi_k^{\mathrm{T}} \ \Phi(1))^{\mathrm{T}} \ (\Psi_k^{\mathrm{T}} \ \Phi(2))^{\mathrm{T}} \ldots (\Psi_k^{\mathrm{T}} \ \Phi(N))^{\mathrm{T}} ]^{\mathrm{T}}. \qquad (5)$$

With the LFDA representation of the sketch $\varphi_s$ and photo $\varphi_p$, the normed distance $\|\varphi_s - \varphi_p\|$ is used to select the gallery photo with the minimum distance to the probe sketch.

The proposed LFDA algorithm is a simple yet effective method. From the results in Section 4, we can clearly see that LFDA is able to significantly improve the recognition performance over the basic feature-based sketch matching framework. Similar to other variants of LDA that are designed to handle the small sample size problem [25], [24], [23], LFDA has several advantages over traditional linear discriminant analysis. First, LFDA is more effective in handling large feature vectors. The idea of segregating the feature vectors into slices allows us to work on more manageable sized data with respect to the number of training images. Second, because the subspace dimension is fixed by the number of training subjects, when dealing with the smaller sized slices, the LFDA algorithm is able to extract a larger number of

meaningful features. This is because the dimensionality of each slice subspace is bounded by the same number of subjects as a subspace on the entire feature representation would be.

## 4  VIEWED SKETCH MATCHING RESULTS

In order to compare our proposed LFDA framework to published methods on sketch matching, we evaluated our method using viewed sketches from the CUHK data set[1] [5]. This data set consists of 606 corresponding sketch/photo pairs that were drawn from three face data sets: 1) 123 pairs from the AR face database [27], 2) 295 pairs from the XM2VTS database [28], and 3) 188 pairs from the CUHK student database [6]. Each of these sketch images was drawn by an artist while looking at the corresponding photograph of the subject. Two examples of these viewed sketches are shown in Fig. 1a. For the methods presented in this paper, all results shown are the recognition rates averaged over five separate random splits of 306 training subjects and 300 test subjects.

The results of viewed sketch matching experiment are summarized in Table 1. The first column of the table shows the baseline methods, which includes the top two performing methods in the literature [1], [5] (each used 306 training subjects and 300 test subjects) and Cognitec's FaceVACS commercial face recognition engine [26]. FaceVACS has been shown [1] to perform at the same level as earlier solutions specifically trained for viewed sketch recognition [2]. In the second column, the matching accuracies achieved by directly comparing SIFT and MLBP feature vectors $\Phi$ are listed. The method "SIFT + MLBP" indicates a sum of score fusion [29] of the match scores from SIFT matching and MLBP matching. While both the SIFT and MLBP methods offer similar levels of performance, using LFDA (third column) the accuracy increases to the point where (one average) less than two sketches are incorrectly identified out of the 300 sketches in the probe set.

While LFDA was able to reduce the error in half, the use of LDA actually induced higher error. In the same experiment shown in Table 1, we applied LDA on the entire feature vector $\Phi$ instead of breaking it into slices and performing LDA on each slice vector as is done in LFDA. The accuracy of LDA+SIFT was 95.47 percent, LDA+MLBP was 91.53 percent, and (SIFT+MLBP)+LDA was 97.07 percent. In each case, LDA actually lowered the accuracy from the non-LFDA case. The decrease in accuracy observed when applying the standard LDA is due to the small sample size problem and the curse of dimensionality [22]. Given our large feature representation (for a 32-pixel patch size, the SIFT representation contains 19,712 components and the MLBP representation contains 36,344 components), the subspace projections are too tightly fit to the training data. Because the LFDA method is an ensemble method, it is better suited to avoid this overfitting problem. Other LDA variants have been shown to handle the small sample size problem as well, such as random sampling LDA [23] and regularized LDA [24].

---

1. The CUHK Face Sketch Database is available for download at: http://mmlab.ie.cuhk.edu.hk/facesketch.html.

## 5  MATCHING FORENSIC SKETCHES

The available methods for matching forensic sketches to photos are limited. Uhl and Lobo [30] proposed a now antiquated method of matching sketches drawn by forensic artists using photometric standardization and facial features. Yuen and Man [31] matched lab generated forensic composites to photographs based on point distribution models.

### 5.1  Forensic Sketch Database

In our study, we used a data set consisting of 159 forensic sketches, each with a corresponding photograph of the subject who was later identified by the law enforcement agency. All of these sketches were drawn by forensic sketch artists working with witnesses who provided verbal descriptions after crimes were committed by an unknown culprit. The corresponding photographs (mug shots) are the result of the subject later being identified. The forensic sketch data set used here comes from four different sources:

1. 73 images from the forensic sketch artist Lois Gibson [32],
2. 43 images from the forensic sketch artist Karen Taylor [33],
3. 39 forensic sketches provided by the Michigan State Police Department, and
4. 4 forensic sketches provided by the Pinellas County Sheriff's Office.

In addition to these 159 corresponding forensic sketch and photo pairs, we also made use of a data set of 10,000 mug shot images provided by the Michigan State Police to populate the gallery. Thus, the matching experiments closely replicate real-world scenarios where a law enforcement agency would query a large gallery of mug shot images with a forensic sketch. Examples of the forensic sketches used in our experiments are shown in Figs. 1, 6, and 7.

Certain sketch images in our collection of forensic sketches are of poor quality in terms of not capturing all of the facial features of the suspect. For most of these sketches, it is unlikely that they can be successfully matched automatically to the corresponding photos because they barely resemble the subject. For this reason, we separated our forensic sketches into two categories: *good* quality and *poor* quality. This separation was performed subjectively by looking at the corresponding pairs (sketch and photo) and labeling them as good if the sketch possessed a reasonable resemblance of the subject in the photo and labeling them as poor if the sketch was grossly inaccurate. We believe that this leads to a more accurate portrayal of how accurately forensic sketches can be matched. Fig. 1 shows the difference between *good* quality and *poor* quality sketches.

### 5.2  Large-Scale Forensic Sketch Matching

Matching forensic sketches to large mug shot galleries is different in several respects from traditional face identification scenarios. When presenting face recognition results in normal recognition scenarios, we are generally concerned with exactly identifying the subject in question in a fully automated manner. For example, when preventing multiple passports from being issued to the same person, human interaction should be limited to only ambiguous cases. This is due to the large volume of requests such a system must process. The same is true for matching arrested criminals against existing mug shot databases to confirm their identity. However, when matching forensic sketches, it is not critical for the top retrieval result to be the correct subject as long as it is in the top $R$ retrieved results, say $R = 50$. This is because the culprit being depicted in a forensic sketch typically has committed a heinous crime (e.g., murder, rape, and armed robbery) that will receive a large amount of attention from investigators. Instead of accepting or dismissing only the top retrieved photo, law enforcement officers will consider the top $R$ retrieval results as potential suspects. Generally, many of the

TABLE 2
Demographics of the 159 Forensic Sketch Images
and the 10,159 Mug Shot Gallery Images

|  | Forensic Sketches | Mugshot Gallery |
|---|---|---|
| Caucasian | 58.49% | 46.43% |
| African American | 31.45% | 46.93% |
| Other | 10.06% | 6.64 % |
| Male | 91.19% | 84.33% |
| Female | 8.81% | 15.52% |
| Unknown | 0.00% | 0.03% |

returned subjects can be immediately eliminated as suspects for various reasons, such as if they are currently incarcerated or deceased. The remaining candidates can each then be investigated for their culpability of committing the crime. This scenario is also true of crimes in which a photograph of a suspect is available. Investigators will consider the top $R$ retrieval results instead of only the highest match. Based on the practice followed in forensics, we would like $R$ to be around 50; that is, we are mainly concerned with whether or not the true subject is within the top 50 retrieved images.

In order to improve the accuracy of matching forensic sketches, we utilize ancillary or demographic information provided by the witness, to be used as a soft biometric [34]. For example, suppose the witness reports that the race of the culprit is Caucasian; then we can eliminate all non-Caucasian members of the gallery to not only speed up the matching but also to improve the matching performance. The same is true for gender: If the suspect is reported to be a female, then we disregard any male subject in the gallery. To use this approach, we manually labeled all of the 10,159 mug shot images and all of the forensic sketch/photo pairs in our database with race and gender. For gender, we considered one of three possible categories: male, female, and (in rare cases) unknown. For race, we considered one of three categories: Caucasian, African-American, and "other." The "other" includes individuals who are of Hispanic, Asian, or multiple races. Table 2 lists the percentage of members from each race and gender category in the forensic sketches and the mug shot gallery used in our experiments.

We lack additional ancillary information (e.g., age, height, scars, marks, and tattoos) that could potentially be used to further improve the matching accuracy.

## 6  FORENSIC SKETCH MATCHING RESULTS

Forensic sketch recognition performance using the 159 forensic sketch images (probe set) and 10,159 mug shot images (gallery) will now be presented. In these matching experiments, we use the *local feature-based discriminant analysis* framework presented in Section 3. Our matching uses the sum of score fusion of MLBP and SIFT LFDA, as this was the highest performing method for matching viewed sketches (Table 1).

The performance of matching sketches classified as *good* and *poor* can be found in Fig. 3. There is a substantial difference in the matching performance of *good* sketches and *poor* sketches. Despite the fact that *poor* sketches are extremely difficult to match, the CMC plots in Fig. 3 show that the proposed method performs roughly the same on the *poor* sketches than a state-of-the-art commercial matcher (FaceVACS) does on the *good* sketches.

Fig. 4 and Table 3 show the recognition performance when race and gender information is used to filter the gallery. By utilizing this ancillary information, we can significantly increase the performance of forensic sketch recognition. We noticed a larger performance gain by using race information than the gender information. This is likely due to the more uniform distribution of
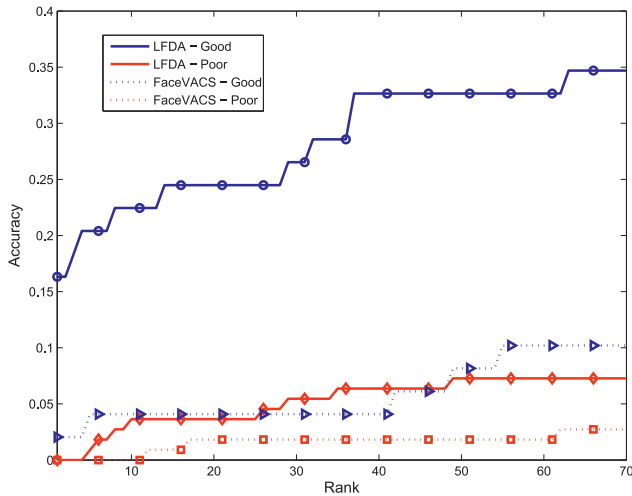
Fig. 3. Performance of matching forensic sketches that were labeled as *good* (49 sketches) and *poor* (110 sketches) against a gallery of 10,159 mug shot images without using race/gender filtering.
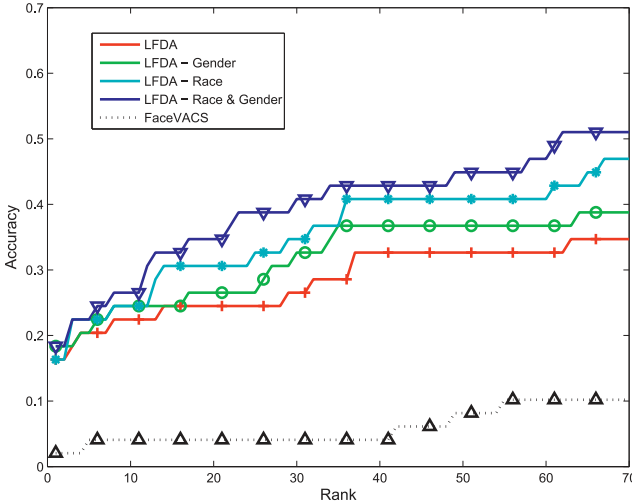


Fig. 4. Performance of matching *good* sketches with and without using ancillary demographic information (race and gender) to filter the results.



Fig. 5. Drop in Rank-50 accuracy using LFDA and FaceVACS on the *good* sketches, without race and gender filtering, as the gallery size increases.

race membership than gender membership in our gallery. The use of other demographic information such as age and height should offer further improvements.

Fig. 5 demonstrates how recognition accuracy decreases as the gallery size increases. This indicates that the proposed LFDA method is scalable with respect to the even larger size galleries that will be encountered in real-world scenarios.

Examples of failed retrievals are shown in Fig. 6. While the top retrieved mug shot is not correct in these two examples, the probe
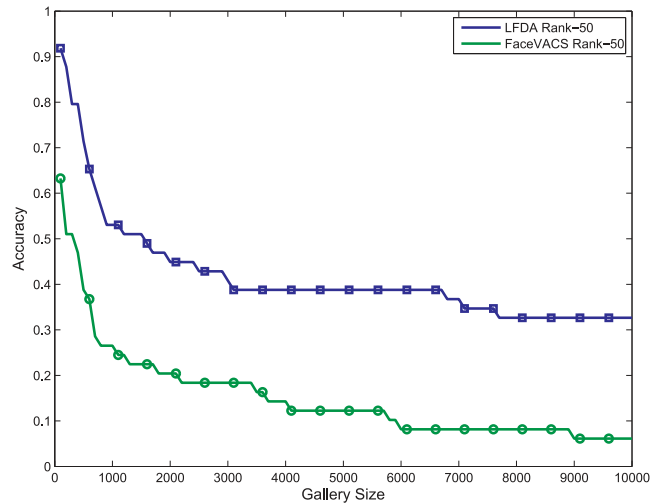
sketch appears to be more similar to the top matched photo than the true photograph. This was nearly always the case: *The top retrieved images appeared highly similar to the probe sketch in the incorrect matchings*.

Fig. 7 shows three of the best matches and three of the worst matches among all of the *good* sketches using the proposed LFDA recognition method. For each image, we have listed the match rank returned by LFDA and FaceVACS.

We performed additional studies to show how the human memory can impact the accuracy of face recognition performance. In these additional experiments, the recognition accuracies of individual facial components (eyes, nose, mouth, and chin) are measured. It is shown that the "external" regions of forensic sketches (chin, hairline, etc.) are more salient than the "internal" regions (eyes, nose, and mouth). While this is contrary to the saliency in face photographs, it is in agreement with cognitive science research on the ability of humans to remember unfamiliar subjects [35]. Due to space limitations, the results are included in a technical report [20].

One limitation of our study is the small number of forensic sketches in our data set, but obtaining a large collection of forensics sketches and the mated photographs from law enforcement agencies is not easy. Not only does this limit the evaluation of our method, but it also affects the performance of our local feature-based discriminant analysis. The LFDA needs a reasonable number of training examples to learn the most discriminative projections. In the case of viewed sketch recognition, we used 306 pairs of sketches and photos for training. For the forensic sketches, even if we performed leave-one-out cross validation, there would still be only a small number of good quality training samples. For this reason, we trained the discriminant on the viewed sketches when

TABLE 3
Recognition Performance of Matching Good Quality Forensic Sketches (49 Sketches) against a Gallery of 10,159 Subjects

| | **Without Race/Gender Filtering** | | |
| | Rank-1 Accuracy (%) | Rank-10 Accuracy (%) | Rank-50 Accuracy (%) |
| --- | --- | --- | --- |
| FaceVACS | 2.04 | 4.08 | 8.16 |
| LFDA | 16.33 | 22.45 | 32.65 |
| | **With Race/Gender Filtering** | | |
| | Rank-1 Accuracy (%) | Rank-10 Accuracy (%) | Rank-50 Accuracy (%) |
| FaceVACS | 2.04 | 8.16 | 26.53 |
| LFDA | 18.37 | 26.53 | 44.90 |

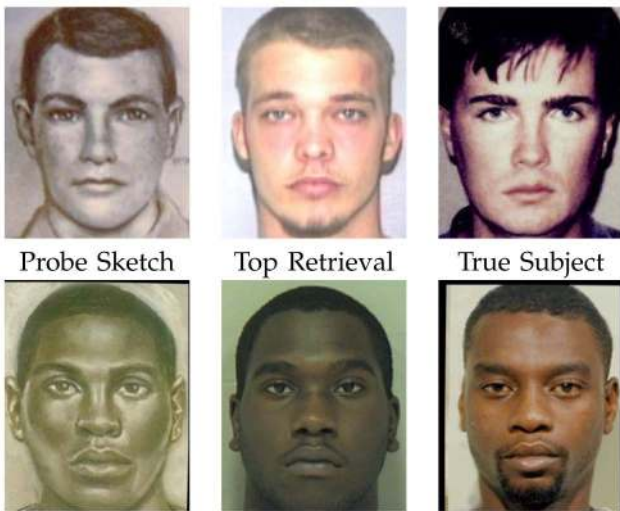*Criminal investigators are generally interested in the top 50 retrieved results.*

Fig. 6. Two examples of failed retrievals on *good* sketches where the first column shows the probe sketches, the second column shows the incorrect Rank-1 matches, and the third column shows the true mated photographs. These are typical cases in which the true subject photo was not retrieved at rank 1, but the impostor subject retrieved at rank 1 visually looks more similar to the sketch than the true subject.

matching forensic sketches. However, we believe that with a larger number of forensic sketches, we could more properly train our discriminant and further improve the matching performance. The bottleneck in finding additional forensic sketches for our experiments is in obtaining the photograph mates. While forensic sketches exist from numerous crimes, even if there is an eventual identification of the subject, the mated sketch and photo are not often stored together in a central database. We are currently working with various law enforcement agencies to increase our data set of forensic sketch pairs.

## 7 CONCLUSIONS

We have presented methods and experiments in matching forensic face sketches to photographs. Matching forensic sketches is a very difficult problem for two main reasons: 1) Forensic sketches are often an incomplete and poor portrayal of the subject's face. 2) We must match across image modalities since the gallery images are photographs and the probe images are sketches.

One of the key contributions of this paper is using SIFT and MLBP feature descriptors to represent both sketches and photos. We improved the accuracy of this representation by applying an ensemble of discriminant classifiers, and termed this framework *local feature discriminant analysis*. The LFDA feature-based representation of sketches and photos was clearly shown to perform better on a public domain-viewed sketch data set than previously published approaches.

Another major contribution of the paper is the large-scale experiment on matching forensic sketches. While previous research efforts have focused on viewed sketches, most real-world problems only involve matching forensic sketches. Using a collection of 159 forensic sketches, we performed matching against a gallery populated with 10,159 mug shot images. Further improvements to the LFDA method were achieved by utilizing ancillary information such as race and gender to filter the 10,159 member gallery. For an unbiased evaluation of our methods, we used a state-of-the-art face recognition system, FaceVACS [26].

Continued efforts on matching forensic sketches are critical for assisting law enforcement agencies in apprehending suspects. A larger data set of forensic sketches and matching photographs needs to be collected to further understand the nature and complexity of the problem.

| Method | Rank |
| --- | --- |
| LFDA | 1 |
| FaceVACS | 320 |

| Method | Rank |
| --- | --- |
| LFDA | 1 |
| FaceVACS | 299 |

| Method | Rank |
| --- | --- |
| LFDA | 1 |
| FaceVACS | 2131 |

| Method | Rank |
| --- | --- |
| LFDA | 775 |
| FaceVACS | 2255 |

| Method | Rank |
| --- | --- |
| LFDA | 1599 |
| FaceVACS | 215 |

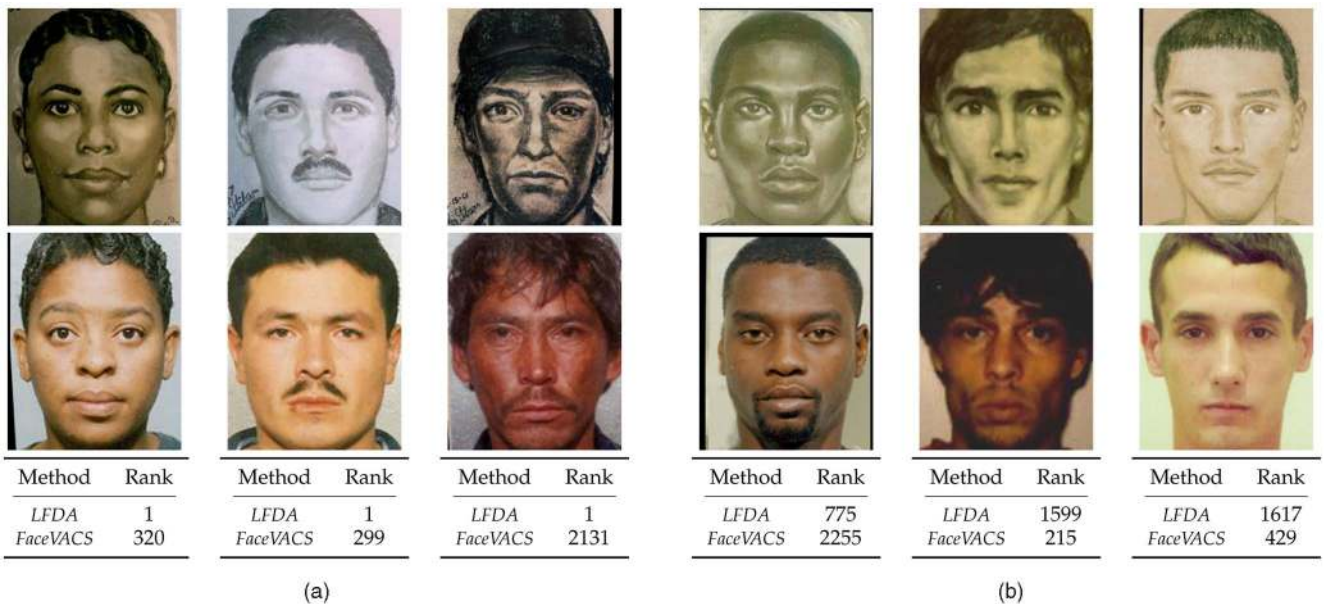| Method | Rank |
| --- | --- |
| LFDA | 1617 |
| FaceVACS | 429 |

(a)        (b)

Fig. 7. Example matches from *good* quality forensic sketches with a background gallery of 10,159 subjects. We show (a) three of the best matches and (b) three of the worst matches. Below each example is the rank of correct match obtained by the proposed LFDA method and FaceVACS.

## REFERENCES

[1] B. Klare and A. Jain, "Sketch to Photo Matching: A Feature-Based Approach," *Proc. SPIE Conf. Biometric Technology for Human Identification VII,* 2010.

[2] X. Tang and X. Wang, "Face Sketch Recognition," *IEEE Trans. Circuits and Systems for Video Technology,* vol. 14, no. 1, pp. 50-57, Jan. 2004.

[3] Q. Liu, X. Tang, H. Jin, H. Lu, and S. Ma, "A Nonlinear Approach for Face Sketch Synthesis and Recognition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 1005-1010, 2005.

[4] J. Zhong, X. Gao, and C. Tian, "Face Sketch Synthesis Using E-HMM and Selective Ensemble," *Proc. IEEE Conf. Acoustics, Speech, and Signal Processing,* 2007.

[5] X. Wang and X. Tang, "Face Photo-Sketch Synthesis and Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 31, no. 11, pp. 1955-1967, Nov. 2009.

[6] X. Tang and X. Wang, "Face Sketch Synthesis and Recognition," *Proc. IEEE Int'l Conf. Computer Vision,* pp. 687-694, 2003.

[7] D. Lin and X. Tang, "Inter-Modality Face Recognition," *Proc. European Conf. Computer Vision,* 2006.

[8] X. Gao, J. Zhong, J. Li, and C. Tian, "Face Sketch Synthesis Algorithm Based on E-HMM and Selective Ensemble," *IEEE Trans. Circuits and Systems for Video Technology,* vol. 18, no. 4, pp. 487-496, Apr. 2008.

[9] B. Xiao, X. Gao, D. Tao, and X. Li, "A New Approach for Face Recognition by Sketches in Photos," *Signal Processing,* vol. 89, no. 8, pp. 1576-1588, 2009.

[10] W. Liu, X. Tang, and J. Liu, "Bayesian Tensor Inference for Sketch-Based Facial Photo Hallucination," *Proc. 20th Int'l Joint Conf. Artificial Intelligence,* 2007.

[11] Y. Li, M. Savvides, and V. Bhagavatula, "Illumination Tolerant Face Recognition Using a Novel Face from Sketch Synthesis Approach and Advanced Correlation Filters," *Proc. IEEE Int'l Conf. Acoustics, Speech and Signal Processing,* 2006.

[12] H. Nizami, J. Adkins-Hill, Y. Zhang, J. Sullins, C. McCullough, S. Canavan, and L. Yin, "A Biometric Database with Rotating Head Videos and Hand-Drawn Face Sketches," *Proc. IEEE Conf. Biometrics: Theory, Applications, and Systems,* 2009.

[13] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int'l J. Computer Vision,* vol. 60, no. 2, pp. 91-110, 2004.

[14] S. Liao, D. Yi, Z. Lei, R. Qin, and S. Li, "Heterogeneous Face Recognition from Local Structures of Normalized Appearance," *Proc. Third Int'l Conf. Biometrics,* 2009.

[15] Z. Lei and S. Li, "Coupled Spectral Regression for Matching Heterogeneous Faces," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 1123-1128, June 2009.

[16] B. Klare and A. Jain, "Heterogeneous Face Recognition: Matching NIR to Visible Light Images," *Proc. Int'l Conf. Pattern Recognition,* 2010.

[17] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 7, pp. 971-987, July 2002.

[18] K. Mikolajczyk and C. Schmid, "A Performance Evaluation of Local Descriptors," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 27, no. 10, pp. 1615-1630, Oct. 2005.

[19] T. Ahonen, A. Hadid, and M. Pietikainen, "Face Description with Local Binary Patterns: Application to Face Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 28, no. 12, pp. 2037-2041, Dec. 2006.

[20] B. Klare, Z. Li, and A. Jain, "On Matching Forensic Sketches to Mug Shot Photos," MSU Technical Report MSU-CSE-10-3, 2010.

[21] E. Meyers and L. Wolf, "Using Biologically Inspired Features for Face Processing," *Int'l J. Computer Vision,* vol. 76, no. 1, pp. 93-104, 2008.

[22] S. Raudys and A. Jain, "Small Sample Size Effects in Statistical Pattern Recognition: Recommendations for Practitioners," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 13, no. 3, pp. 252-264, Mar. 1991.

[23] X. Wang and X. Tang, "Random Sampling for Subspace Face Recognition," *Int'l J. Computer Vision,* vol. 70, no. 1, pp. 91-104, 2006.

[24] J. Lu, K. Plataniotis, and A. Venetsanopoulos, "Regularization Studies of Linear Discriminant Analysis in Small Sample Size Scenarios with Application to Face Recognition," *Pattern Recognition Letters,* vol. 26, no. 2, pp. 181-191, 2005.

[25] H.Y. Jie, H. Yu, and J. Yang, "A Direct LDA Algorithm for High-Dimensional Data—with Application to Face Recognition," *Pattern Recognition,* vol. 34, pp. 2067-2070, 2001.

[26] FaceVACS Software Developer Kit, Cognitec Systems GmbH, http://www.cognitec-systems.de, 2010.

[27] A. Martinez and R. Benavente, "The AR Face Database," CVC Technical Report 24, 1998.

[28] K. Messer, J. Matas, J. Kittler, and K. Jonsson, "XM2VTSDB: The Extended M2VTS Database," *Proc. Int'l Conf. Audio and Video-Based Biometric Person Authentication,* 1999.

[29] A. Ross and A. Jain, "Information Fusion in Biometrics," *Pattern Recognition Letters,* vol. 24, no. 13, pp. 2115-2125, 2003.

[30] R. Uhl and N. Lobo, "A Framework for Recognizing a Facial Image from a Police Sketch," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 1996.

[31] P. Yuen and C. Man, "Human Face Image Searching System Using Sketches," *IEEE Trans. Systems, Man, and Cybernetics,* vol. 37, no. 4, pp. 493-504, July 2007.

[32] L. Gibson, *Forensic Art Essentials.* Elsevier, 2008.

[33] K. Taylor, *Forensic Art and Illustration.* CRC Press, 2001.

[34] A.K. Jain, S.C. Dass, and K. Nandakumar, "Soft Biometric Traits for Personal Recognition Systems," *Proc. Int'l Conf. Biometric Authentication,* 2004.

[35] C. Frowd, V. Bruce, A. McIntyr, and P. Hancock, "The Relative Importance of External and Internal Features of Facial Composites," *British J. Psychology,* vol. 98, no. 1, pp. 61-77, 2007.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.