On Matching Forensic Sketches to Mugshot Photos

Under review TPAMI

Brendan Klare, Zhifeng Li, and Anil K. Jain

Abstract

The problem of matching a forensic sketch to a gallery of mugshot images is addressed in this paper. Previous research in sketch matching offered solutions to matching highly accurate sketches that were drawn while looking at the subject. We refer to these as viewed sketches. Forensic sketches differ from viewed sketches in that they are drawn by a police sketch artist using the description provided by an eye-witness who is typically unfamiliar with the subject. To solve the problem of matching forensic sketches, we propose a robust framework, called local feature-based discriminant analysis (LFDA). In LFDA, we first represent both sketches and photos using gradient location and orientation histograms and multi-scale local binary patterns. A discriminant projection method is then used on the feature-based representation for minimum distance matching. We apply the LFDA method to match a dataset of forensic sketches against a mugshot gallery containing 10,159 images. Compared to a leading commercial face recognition system, LFDA offers substantial improvements to matching forensic sketches to the corresponding images. We are able to further improve the matching performance using race and gender information to reduce the target gallery size. Experiments on matching viewed sketches also demonstrate that LFDA matching method exceeds all other previous methods of sketch matching.

Index Terms

Face Recognition, forensic sketch, viewed sketch, feature-based, discriminant, feature selection

I. Introduction

The maturation of biometric technology has provided criminal investigators 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 eye-witness account of the crime is available. In these circumstances a forensic artist is often used to work with the witness or the victim 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 who the suspect is. These sketches are known as forensic sketches and this paper describes a robust method for matching forensic sketches to mughshot (image) databases maintained by law enforcement agencies.

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

Brendan Klare, Zhifeng Li, and Anil K. Jain are with the Department of Computer Science and Engineering, Michigan State University, 3115 Engineering Building, East Lansing, MI 48824-1226, U.S.A.

Anil K. Jain is also with the Department of Brain and Cognitive Engineering, Korea University, Seoul, Korea

E-mail: klarebre,zfli,jain@cse.msu.edu

(a) (b) (c)

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. (c) Two pairs of poor quality forensic sketches and the corresponding photograph.

TABLE I
EXISTING METHODS IN LITERATURE FOR MATCHING VIEWED SKETCHES TO PHOTOGRAPHS.

Author and Year	Method	Database	# Probe / # Gallery	Rank-1 Performance (%)
Tang and Wang 2003 [7]	Eigen-transform	CUHK	300 / 300	81.3
Tang and Wang 2004 [2]	Eigen-transform	CUHK	300 / 300	90.0
Liu et al. 2005 [3]	Nonlinear Synthesis	CUHK	300 / 300	87.7
Zhong et al. 2007 [4]	E-HMM Synthesis	CUHK	300 / 300	95.24
Wang and Tang 2009 [5]	MRF-based synthesis	CUHK	300 / 300	96.3
Klare and Jain 2010 [1]	SIFT feature-based	CUHK	300 / 300	97.87
Lin and Tang 2009 [8]	Common Discriminant	FERET	700 / 350	92.87
Xiao et al. 2008 [9]	E-HMM Synthesis	CUHK	56 / 56	100.0

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, despite the fact that real world scenarios only involve forensic sketches [1][2][3][4][5]. Both forensic sketches and viewed sketches complicate 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 a witness's inability to exactly remember the appearance of a suspect and her subjective account of the description, which often results in inaccurate 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 II and summarized in Figure 2. The second problem is considered in Section III, where analysis and improvements are offered for matching forensic sketches against large mugshot galleries.

The contributions of the paper are summarized as follows: (i) We observe a substantial improvement in matching viewed sketches over published algorithms using the proposed local feature-based discriminant analysis; (ii) We present the first large-scale published experiment on matching real forensic sketches; (iii) Using a mugshot gallery of 10,159 images, we perform race and gender filtering to improve the matching results; (iv) All experiments are validated by comparing proposed methods against a leading commercial face recognition engine.

II. MATCHING VIEWED SKETCHES

This section focuses on matching viewed sketches. Matching viewed sketches is a hypothetical situation because, to our knowledge, there is no practical scenario where a viewed sketch would be used for identification instead of the photograph it was drawn from. That said, methods for robustly matching viewed sketches are of critical importance to matching forensic sketches. This is because, in general, an algorithm that can successfully match viewed sketches will only fail in matching forensic sketches when the forensic sketch is incomplete or of poor quality. By developing a highly robust viewed sketch matching algorithm, we can focus strictly on the problem of sketch quality when progressing to forensic sketches.

Matching viewed sketches to photographs is an aspect of heterogeneous face recognition [6]. These problems involve matching two face images acquired in different modalities. Some example modalities matched in heterogeneous face biometrics are: spectral bands (e.g. near infared and visible), image sensors (e.g. CCD and CMOS), resolutions (e.g. low and high), and (in this case) sketches and photos. In Section II-A we discuss the published papers on matching viewed sketches, which are summarized in Table I. Next, we build on the effective feature-based sketch matching approach proposed by Klare and Jain [1] by finding the optimal set of image feature descriptors (gradient location and orientation histograms, and multi-scale local binary patterns) for representing sketches and photos (Sections II-B and II-C). After representing sketches and photos with the selected feature descriptors, in Section II-D we present a local feature-based discriminant analysis method (LFDA) that further reduces the dimensionality of the sketch and photo representations and projects them in a direction that is better suited for classification. An overview of this entire process is shown in Figure 2. In Section II-E we experimentally demonstrate the effectiveness of these methods for matching viewed sketches.

A. Prior Work

Most of the work in matching viewed sketches was performed by Tang et al. [7][2][3][5][8]. Tang and Wang first approached the problem using an eigentransformation method [2] to either project a sketch image into a photo subspace, or to project a photo image into a sketch subspace. Once projected into the same image subspace, they were matched using a PCA-based matcher. In [7], Tang and Wang performed a similar eigen-based transformation; however, they applied the transformation to the shape and texture of the face separately. In [3], a synthetic sketch generation method was described by Liu et al. that would convert a photo image into a geometry preserving synthetic sketch (or vice-versa) with Local Linear Embedding (LLE). An improvement to this method was offered by Wang and Tang [5], where the relationship between sketch and photo image patches was modeled with a Markov random field. Belief propagation was used to minimize the energy between the selected

patches and their corresponding sketch or photo mates as well as their selected neighboring patches. In both [3] and [5], the synthetic sketches generated were matched to a gallery of photographs using a variety of standard face recognition algorithms.

In [8], Lin and Tang presented a general method for heterogeneous face biometerics that was applied between the sketch and photo modalities (as well as the NIR and visible spectra). To match sketches and photos they learned a linear transformation method that projected images from the two domains into a common feature space. Zhong et al. [4] used an embedded hidden Markov model (E-HMM) to synthesize a sketch from a photo. Gao et al. presented a similar synthesis method using embedded hidden Markov models in [10] and [11]. In [9], Xiao et al. improved the E-HMM synthesis method by breaking the images into patches and focusing on the conversion of sketch to photo (instead of photo to sketch). Wei et al. [12] presented a synthesis (or hallucinating faces) method that builds on the non-linear method in [3]. Al Nizami et al. [13] demonstrated the intrapersonal variation that occurs in sketch images drawn by different artists.

A feature-based method for matching sketches was presented by Klare and Jain [1], which serves as the motivation for the sketch matching method presented in this paper. The feature-based sketch matching approach uniformly samples both sketch and photo images using SIFT feature descriptors [14] at different scales. To match a sketch with its corresponding photo, Klare and Jain presented two separate feature-based methods: direct matching and common representation matching. The direct matching method measured the distance between the SIFT descriptors in the sketch and photo domains. It was shown in [1] that the SIFT descriptors are largely invariant to the sketch and photo modalities, which allowed for successful matching by directly comparing these descriptors. The common representation method uses a training set of sketch-photo correspondences to first measure the distance of probe sketches and gallery photos to the training set using the SIFT feature representation. The distance between a sketch and photo is then measured by how similar they are to the subjects in the training set. Using the common representation method removes the need to directly compare image descriptors across modalities based on the assumption that if the computed image features from sketches of two different people are similar (or dissimilar), then the same features computed on the corresponding photos of those same two persons will also be similar (or dissimilar).

Most previously discussed methods for matching viewed sketches and photos are based on either performing a global conversion from one domain to the other, or by using a patch-based conversion that replaces each sketch image patch by a photo patch in a training set. These methods have the advantage of being able to generate a synthetic photograph of a sketch (or vice-versa), which can then be matched using existing face recognition algorithms. However, the performance of these methods is limited because they are dependent on the amount of training data. When removing the constraints provided by a closed data set, these methods are subject to errors in estimating the synthetic images due to the large dimensionality and range of pixel values in the synthetic image.

B. 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 [15]. In this section we develop a framework for representing sketches and photos using an image descriptor, and define three key properties these descriptors should possess for sketch matching. In Section II-C, we will then use these properties to guide the selection of actual image features that offer the most robust sketch matching.

Let us define the following notations. $f(I) : I \rightarrow \phi$ is an image descriptor that maps image (or image region) $I \in \mathbb{R}^{W,H}$ to the vector $\phi \in \mathbb{R}^d$. Denoting I as I_s or I_p indicates the image is a sketch or photo, and denoting I as I^i or I^j denotes the image is from person i or j. For example, I_s^j is a sketch image of person j and I_p^i is a photo image of person i. Two images of the same person i in the same domain are denoted as I^{i_1} and I^{i_2}. Using these notations, we will now define three properties that an image feature descriptor must satisfy in order to be useful for matching sketches and photos:

Property 1: Provide a low dimensional representation of a high dimensional WxH image.

$$d \ll WH \tag{1}$$

A compact representation of a face image is important to allow for successful recognition. This property formed the basis of Turk and Pentland's eigenface approach [16].

Property 2: Maintain sufficient image information for face recognition.

$$||\phi^i - \phi^j|| \gg ||\phi^{i_1} - \phi^{i_2}|| \tag{2}$$

Though rather intuitive, this property is important to state nonetheless. If an image descriptor is unable to retain the discriminative information in a face image then it cannot be used to identify an individual.

Property 3: Be largely invariant to changes between the sketch and photo domain.

$$||f(I_p^i) - f(I_s^i)|| \approx ||f(I_p^{i_1}) - f(I_p^{i_2})|| \tag{3}$$

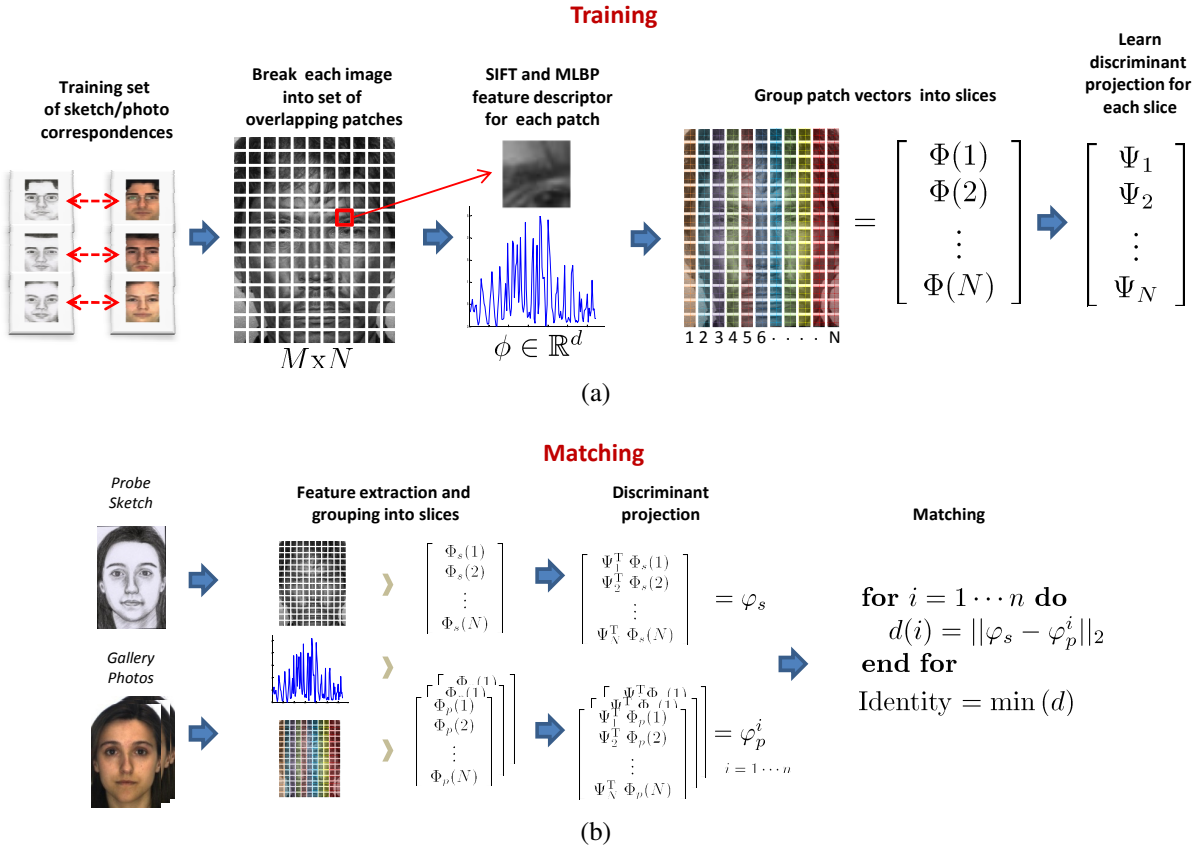


Fig. 2. An overview of the training (a) and matching (b) using the *LFDA* framework. Each sketch and photo are represented by SIFT and MLBP descriptors extracted from overlapping patches. After grouping "slices" of patches together into feature vectors $\Phi(k)$ $(k = 1 \cdots N)$, we learn a discriminant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

In order to directly compare sketches and photos we need an image descriptor that has roughly the same output for two images that differ only in modality. Without satisfaction of this property, changes in the feature vector could be due to either a different identity or from a different modality (sketch or photo), making recognition infeasible.

If an image descriptor can satisfy all the above three properties then: 1) we can predict the identity of individuals in this feature space representation, and 2) the descriptor will be mostly the same regardless of which image domain it is being extracted from (sketch or photo). Finding descriptors that best satisfy these conditions will help solve the problem of sketch matching.

We will now describe how to represent a face with an image descriptor. Most image descriptors are not large enough to fully describe a face image (i.e. they fail Property 2). Because of this, we compute feature descriptors 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, and a displacement size. The region size s defines the size of the square window the image feature is computed from. The displacement size δ states the number of pixels the patch is displaced for each sample. This is analagous to sliding a window of size sxs across the face image in a raster scan fashion. After moving it δ pixels we compute the image descriptor and continue the scan. This is similar to dividing the face image into overlapping patches of size s that overlap by δ pixels. If an image has a height of H pixels and width of W pixels, then the number of horizontal (N) and vertical (M) sampling locations are given by

$$\begin{aligned} N &= (W - s)/\delta + 1 \\ M &= (H - s)/\delta + 1 \end{aligned} \tag{4}$$

At each of the $M \cdot N$ patches, we compute the d dimensional image feature vector ϕ. These image feature vectors are concatenated into one single $M \cdot N \cdot d$ dimensional image vector Φ. 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 simply measuring the normed vector distance $||F(I^i) - F(I^j)||$ [1].

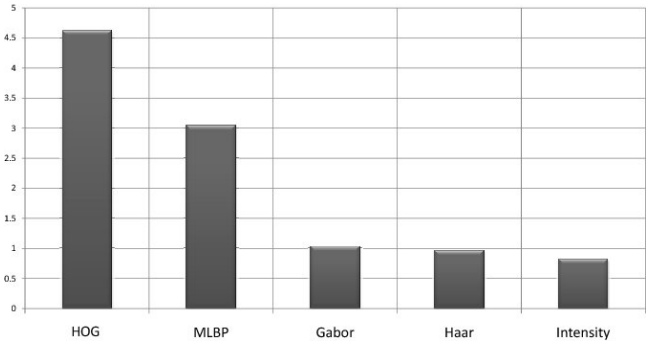


Fig. 3. The distance $\hat{d}$ (Eq. 5) between the same person comparisons and different person comparisons across sketch and photo images for each candidate feature descriptor. A larger distance indicates the feature descriptor has large invariance between image modalities and thus is more suitable for directly comparing sketches and photos.

C. Feature Descriptor Selection

We now experimentally determine the feature descriptors that best satisfy the three stated properties. The following set of candidate image descriptors are considered: SIFT, MLBP, Haar, Gabor, and intensity. The SIFT feature descriptor was originally proposed as part of the larger SIFT framework [14]; we are only using the descriptor, not keypoint detection. The MLBP feature descriptor is the multi-scale local binary pattern. This is simply the concatenation of LBP feature descriptors [17] of radii $r \in \{1, 3, 5, 7, 9\}$. The Gabor descriptor is the output of convoluting an image with 40 Gabor jets (8 orientations and 5 scales). The intensity descriptor is simply a vector of the image intensities at each pixel in the region. Haar features are low cost texture estimates that compare the difference between the summed intensity values in different subregions [18].

Given these sets of candidate features, we would like to find the descriptors that best satisfy the three stated properties. We first test the most critical property: Property 3, invariance between the sketch and photo domains. The satisfiability of this property is empirically determined by measuring distances between feature representations of sketches and photos. Two separate distances are measured in each of the candidate feature spaces: the distance between photos and sketches of the same person, and the distance between the photos and sketches of different people. We randomly selected 100 sketch/photo pairs from the CUHK database (described in Section II-E), which are denoted I_s^i for the sketches, and I_p^i for the photos ($i = 1 \ldots 100$). For each candidate image feature f, we generated the distance matrix $D_f \in \mathbb{R}^{100,100}$, where $D_f(i,j) = ||F(I_s^i) - F(I_p^j)||_2$. We denote the mean of the diagonal entries in D_f as μ_f^{same}, which is the average distance of the same image comparisons across image domains. We denote the mean of all non-diagonal entries in D_f as μ_f^{diff}, which is the average distance of different image comparisons across image domains. Thus, we prefer image feature f_1 over image feature f_2 if

$$\hat{d}_1 = \frac{(\mu_{f_1}^{\text{same}} - \mu_{f_1}^{\text{diff}})^2}{\sigma_{f_1}^2} > \frac{(\mu_{f_2}^{\text{same}} - \mu_{f_2}^{\text{diff}})^2}{\sigma_{f_2}^2} = \hat{d}_2 \tag{5}$$

which is the Mahalanobis distance of the true match distances from the impostor match distances (where σ_f^2 is the variance of entries in D_f). The measured distance thus quantifies the satisfiability of the third property.

Figure 3 lists the distances for each candidate image feature. It is observed that the two most domain invariant image descriptors are the SIFT and MLBP. Both of these feature descriptors share two attributes. The first is that they both describe regions based on a pixel's relative values. In the case of SIFT, this is the form of the vertical and horizontal gradients, and for local binary patterns this is accomplished by thresholding neighborhood pixels by the center value. In describing a region based on its relative values, the change between domains is minimized yet the structure and shape of the face are still captured. The second similarity between these two feature descriptors is that they are distribution-based descriptors. The SIFT and MLBP vectors are normalized to sum to one so that they form a probability measure. This normalization step removes the effect of the different absolute vector magnitudes between different domains.

Satisfaction of Property 1 is simple to determine based on the nature of the descriptors. To determine whether or not Property 2 is satisfied, we refer to the literature. Gabor images have been applied in various approaches for face recognition, most notably in the elastic graph bunch matching algorithm [19]. More recently, a framework for face recognition using local binary patterns was proposed by Ahonen et al. [20]. The use of SIFT descriptors for face recognition in [1] has already been discussed. Finally, the use of image intensity is the common choice to represent a face (typically followed by a subspace projection).

Table II summarizes the candidate descriptors and the properties they satisfy. The only two descriptors that satisfy all the three properties are the SIFT descriptors, which we subsequently will use to represent both sketches and photos. Results reported in Section II-E for each of the two descriptors are a sum score fusion of the match scores generated from computing descriptors with a patch size of $s = 16$ and $s = 32$. This also holds for our global discriminant described in Section II-D: we fuse the matching scores computed using the two separate patch sizes of 16 and 32.

TABLE II
FEATURE DESCRIPTORS AND THE PROPERTIES THEY SATISFY.

Image Descriptor	Property 1	Property 2	Property 3
SIFT	✓	✓	✓
MLBP	✓	✓	✓
Gabor Image	x	✓	x
Haar Feature	✓	x	x
Image Intesity	x	✓	x

TABLE III
RANK-1 RECOGNITION RATES FOR MATCHING VIEWED SKETCHES USING THE CUHK PUBLIC DATASET. THE STANDARD DEVIATION ACROSS THE FIVE RANDOM SPLITS FOR EACH METHOD IN THE MIDDLE AND RIGHT COLUMNS ARE LESS THAN 1%.

Baseline		**Without LFDA**		**LFDA**	
Method	*Rank-1 Accuracy (%)*	*Method*	*Rank-1 Accuracy (%)*	*Method*	*Rank-1 Accuracy (%)*
FaceVACS [21]	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**

D. Local Feature-based Discriminant Analysis

With both sketches and photos characterized using either the aforementioned SIFT or MLBP image feature representation, we further refine this feature space using discriminant analysis. This is motivated by the fact that though the image feature representation is of a reduced dimensionality from the original pixel representation, the total data size is still too large due to the use of overlapping patches in the feature extraction process. In order to handle the large dimensionality efficiently while retaining as much of discriminatory information, a framework called *local feature-based discriminant analysis (LFDA)* is proposed. In this framework each image feature vector is first divided into slices of smaller dimensionality. Next, discriminant analysis is performed separately on each slice to extract the discriminant features. The projected features from all the slices are then combined together for recognition.

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 $X^s = [\Phi_s^1\ \Phi_s^2 \dots \Phi_s^n]$ for the sketch, $X^p = [\Phi_p^1\ \Phi_p^2 \dots \Phi_p^n]$ for the photo, and $X = [\Phi_s^1 \dots \Phi_s^n\ \Phi_p^1 \dots \Phi_p^n]$ for the photo and sketch combined.

The first step in LFDA is to separate the image feature vector into multiple sub-vectors or slices. Given the MxN array of patches (Eq. 4) consisting of SIFT or MLBP feature vectors, we create one slice per 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 $X_k \in \mathbb{R}^{M \cdot d, 2n}$ $(k = 1 \dots N)$, which are all mean centered.

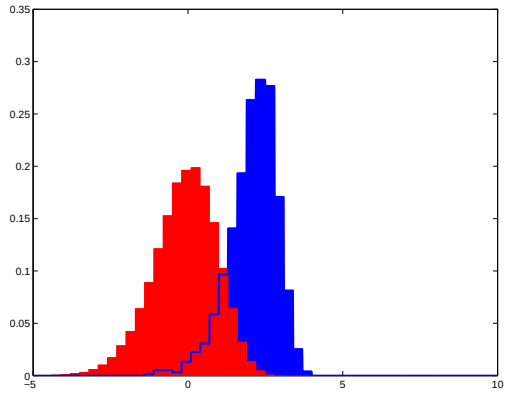
We next reduce the dimensionality of each training slice matrix X^k using PCA. The PCA projection matrix $W_k \in \mathbb{R}^{M \cdot d, 100}$ uses the leading 100 eigenvectors of the training covariance matrix, which generally preserves about 90% of the total variance. The discriminant extraction proceeds by generating the mean projected class vectors

$$Y_k = W_k^{\mathrm{T}}(X_k^s + X_k^p)/2 \tag{6}$$

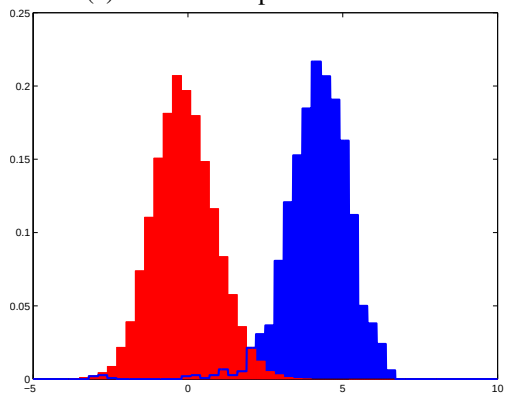
which are used to center the sketch and photo training instances of each class by

$$\begin{aligned} \tilde{X}_k^s &= W_k^{\mathrm{T}}\, X_k^s - Y_k \\ \tilde{X}_k^p &= W_k^{\mathrm{T}}\, X_k^p - Y_k \end{aligned} \tag{7}$$

A whitening transform is performed to reduce the inter-person variation between the sketch and photo. This is achieved by recombining 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 of the 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 $V_k = \left(\Lambda_k^{-\frac{1}{2}} V_k^{\mathrm{T}}\right)^{\mathrm{T}}$. The final step is to compute a projection matrix that maximizes the between person 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



(a) LFDA with pixel intensities



(b) LFDA with MLBP & SIFT features

Fig. 4. (Normalized) Match scores using the local feature-based discriminant with pixel intensities (a) and image features (b). While the discriminant is able to improve the recognition between two different modalities, these plots show that the feature representation of the image modalities (sketch and photo) must already have high levels of similarity for accurate matching. Careful selection of image features (Section II-C) combined with the discriminant helps to separate the distributions of true (blue) and imposter (red) match scores.

$$\Psi_k = W_k V_k U_k \tag{8}$$

With each local feature-based discriminant trained, we can proceed to match sketches to photos as follows. We first separate the feature representation of an image into individual slices

$$\Phi = [\Phi(1)^{\mathrm{T}}\ \Phi(2)^{\mathrm{T}} \ldots \Phi(N)^{\mathrm{T}}\ ]^{\mathrm{T}} \tag{9}$$

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 Ψ_k yielding the new vector representation $\varphi \in \mathbb{R}^{M \cdot 99}$

$$\varphi = \left[(\Psi_k^{\mathrm{T}}\, \Phi(1))^{\mathrm{T}}\ (\Psi_k^{\mathrm{T}}\, \Phi(2))^{\mathrm{T}} \ldots (\Psi_k^{\mathrm{T}}\, \Phi(N))^{\mathrm{T}}\ \right]^{\mathrm{T}} \tag{10}$$

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

The LFDA framework has several advantages. As with any discriminant analysis method, it is able to generate discriminative projections for comparing sketch to photo. Because of the large data size, the approach of segregating the feature vectors into slices allows us to work on more manageable sized data with respect to the number of training images. Without this step, the length of the feature vector would make the discriminant analysis prohibitive in terms of computational requirements. Finally, we will see in Section II-E that LFDA offers significant improvements to feature-based sketch matching.

E. Viewed Sketch Matching Results

To demonstrate the effectiveness of matching viewed sketches using LFDA, we tested our method using the CUHK dataset[1] from [5]. This dataset consists of 606 corresponding sketch/photo pairs that were drawn from three existing datasets: (1) 123

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

(a) (b) (c)

Fig. 5. An example of the internal (b) and external (c) features of the face (a). Humans use the internal features more for recognizing people they are familiar with, and the external features more for recognizing people they are unfamiliar with [27]. Witnesses of a crime are generally unfamiliar with the culprit, therefore the external facial features should be more salient in matching forensic sketches.

pairs from the AR face database [22], (2) 295 pairs from the XM2VTS database [23], and 188 pairs from the CUHK student database [7]. Each of these sketch images were drawn by an artist while looking at the corresponding photograph of the subject. Two examples of these viewed sketches are shown in Figure 1(a). Each sketch and photo image was normalized by rotating the angle between the two eyes to 0°, scaling the images to a 75 interocular pixel distance, and cropping the image size to 200 by 250 pixels. In this experiment, we used 306 pairs for training the LFDA and 300 pairs for testing the accuracy of the system, which allows for a direct comparison against existing methods that used the same data split (see Table I). 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 III. The first column of the table shows the baseline methods, which includes the two top performing methods in literature [1][5] and Cognitec's FaceVACS commercial face recognition engine [21]. In the second column the matching accuracies achieved by directly comparing SIFT and MLBP feature vectors Φ are listed. The method 'SIFT + MLBP' indicates a sum score fusion [24] of the match scores from SIFT matching and MLBP matching. Both the SIFT and MLBP methods offer similar levels of performance, which is of the same order as the top performing previous methods [1][5]. Using the LFDA framework (third column), the recognition performances increase to the point where only two sketches are incorrectly identified out of the 300 sketches in the probe set.

While the LFDA method offers a significant performance improvement, it is necessary to first represent both the sketch and photo in an image feature space that is largely invariant to the change in modality. If the image pixel intensities are used with LFDA instead of the SIFT or MLBP representation, then the average Rank-1 accuracy falls to 83.67%. This is demonstrated in Figure 4, which plots the distribution of the match scores using LFDA both with and without image features. Without image features (i.e. using pixel intensities as the feature vector), there is a significant overlap between true and impostor match scores after discriminant projection. This is because the two supports used in training each class (one sketch and one photo) differ significantly in the pixel intensity representation, which increases the distance from each training item to its mean class location.

III. Matching Forensic Sketches

In this section, we address the problem of matching forensic sketches. We briefly present human cognition studies on witness memory and forensic sketches in Section III-A. We then investigate the saliency of individual face regions (eye, nose, mouth, etc.) in Section III-B. In Section III-C, we improve the forensic sketch matching for a large-scale mugshot database by using ancillary information (race and gender) to filter the gallery. Experimental results from matching a dataset of forensic sketches that were drawn by forensics artists based on an actual verbal description provided by a witness/victim are provided in Section III-D. In these experiments we use a gallery of 10,159 mugshots and a probe set of forensic 159 sketches.

The available methods for matching forensic sketches to photos is limited. Uhl and Lobo [25] proposed a method of matching forensic sketches drawn by forensic artists using photometric standardization and facial features. However, their approach is now antiquated with respect to modern methods, and experimental validation used only 7 probe sketches. Yuen and Man [26] presented a system for matching forensic composites to photographs based on point distribution models, but the sketches were generated using a composite generation software and the subjects were images from the FERET database and another existing face database.

In our study on forensic sketch recognition we used a dataset consisting of 159 forensic sketches, each with a corresponding photograph of the subject that 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 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 forensic sketch artist Lois Gibson that are all published in [28]. (2) 43 images from forensic sketch artist Karen Taylor that are all published in [29]. (3) 39 forensic sketches provided to us by the Michigan State Police Department. (4) 4 forensic sketches provided to us by the Pinellas County Sherrif's Office. In addition to these 159 corresponding forensic sketch and photo pairs, we also made use of a dataset of 10,159 mugshot images provided by the

Fig. 6. Masks used for region based forensic sketch matching. Shown above are the mean photo patches of each patch used for a particular region. The mosaic effect is due to the fact that face patches are extracted in an overlapping manner.

Michigan State Police to populate the gallery. Thus, the matching scenarios in our experiments exactly replicate real world scenarios where a law enforcement agency would query a large gallery of mugshot images with a forensic sketch. Examples of the forensic sketches used in our experiments are shown in Figures 1, 11, and 12.

Certain sketch images in our collection of forensic sketches are of poor quality in terms of not capturing all the facial features of the suspect. For most of these sketches, it is unlikely that they can be matched 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 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 this leads to a more accurate portrayal of how accurately forensic sketches can be matched. Figure 1 shows the difference between *good* quality and *poor* quality sketches.

A. Human Memory and Forensic Sketches

A distinct difference between a viewed sketch and a forensic sketch is that the forensic sketch may have many inaccuracies due to the witness's inability to correctly remember the suspect's face. A body of psychological research exists that focuses on a person's ability to successfully recall the appearance of an individual they are unfamiliar with and whom they viewed only momentarily [27], [30], [31]. A consistent finding of these studies is that the facial features used to recognize someone depends on the level of familiarity. In this respect, facial features are separated into internal and external features (see Figure 5).

When we are familiar with the person we are attempting to recognize (e.g. a co-worker, family member, or celebrity), we predominantly make use of the internal facial features for identification [27], [30]. These features include the nose, eyes, eye brows, and mouth. Most research in automatic face recognition has observed that these internal features are also the most discriminative areas [32]. When we are attempting to recognize someone who is unfamiliar to us, the external features of the face are predominantly used to establish identity [27], [30]. External features consist of the outer region of the face.

Frowd et al. [31] studied whether humans are best able to match forensic sketches using the internal or external features of the face. In their experiments, test subjects were shown the photograph of a celebrity they were unfamiliar with and given approximately one minute to remember the appearance. Two days later the subjects worked with a forensic sketch artist to draw a sketch of the person they viewed earlier in the photograph. Using these composites, a separate set of subjects that had familiarity with the same celebrities were asked to identify two different versions of the sketches: (i) sketches with only the interior regions of the face shown, and (ii) sketches with only the exterior regions of the face shown. The experiments concluded that higher identification rates were achieved using the exterior regions of the face [31].

Frowd et al.'s results are based on a controlled experiment, so we must tread lightly in using them for automated face recognition. One of the most important properties for a biometric trait is permanence [33], which the external regions of the face do not satisfy well. By growing or removing a beard, changing hairstyles, or donning headgear, a person can drastically change the appearance of their external face regions. Therefore, assigning a higher prior probability to the decisions made from external forensic sketch regions over internal regions may not be a wise choice.

B. Forensic Sketch Region Saliency

Due to the observation in the human cognition studies that different regions of the face may be more salient than others, we measure the performance of automatic sketch matching using only certain regions of the face. For our feature-based framework, it is quite easy to implement this by only selecting the patches in the face that correspond to a given region. We considered six separate face regions for localized identification: (1) internal, (2) external, (3) eyes, (4) nose, (5) mouth, and (6) chin. Figure 6 shows the patches used for each of these face regions (with patch size $s = 32$) and the average intensity for each patch (the

TABLE IV
DEMOGRAPHICS OF THE 159 FORENSIC SKETCH IMAGES AND THE 10,159 MUGSHOT 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%

TABLE V
RECOGNITION PERFORMANCES MATCHING GOOD QUALITY FORENSIC SKETCHES (49 SKETCHES) AGAINST A GALLERY OF 10,159 SUBJECTS. CRIMINAL INVESTIGATORS ARE GENERALLY INTERESTED IN THE TOP 50 RETRIEVED RESULTS.

	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

mean patch). Thus, when matching using one of the masks, we performed distance matching using only the patches shown in each mask.

In Section III-D we will show the results of forensic sketch matching using only these face regions.

C. Large-Scale Forensic Sketch Matching

Matching forensic sketches to large mugshot 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 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 mugshot 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 N retrieval results, say $N = 50$. This is because the culprit being depicted in a forensic sketch typically has committed a heinous crime (e.g murder, rape, armed robbery) that will receive a large amount of attention from investigators. Instead of accepting or dismissing only the top retrieval result, law enforcement officers will consider the top N retrieval results as potential suspects. Generally, many of the returned subjects can be immediately eliminated as suspects for various reasons, such as 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 N retrieval results instead of only the highest match. In our experiments we would like N 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 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 sped up the matching but also to improve matching performance. The same is true for gender: if the suspect is reported to be a female then we disregard any male subjects in the gallery. To use this approach, we manually labeled all of the 10,159 mugshot images and all the forensic sketch/photo pairs 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 IV lists the percentage of members from each race and gender category in the forensic sketches and the mugshot gallery used in our experiments.

We lack additional ancillary information that could potentially be used to further improve the matching accuracy. In addition to race and gender, information such as age and height can typically be used to filter the gallery. This is common information that is recorded in both mugshot and driver license records.

D. Forensic Sketch Matching Results

Forensic sketch recognition performance using the 159 forensic sketch images (probe set) and 10,159 mugshot images (gallery) will now be discussed. In these matching experiments we use the *local feature-based discriminant analysis (LFDA)*

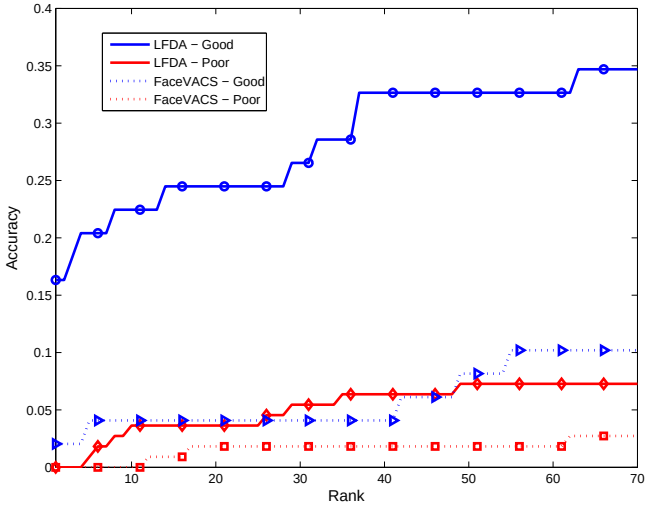


Fig. 7. Performance of matching forensic sketches that were labeled as *good* (49 sketches) and *poor* (110 sketches) against a gallery of 10,159 mugshot images without using race/gender filtering.

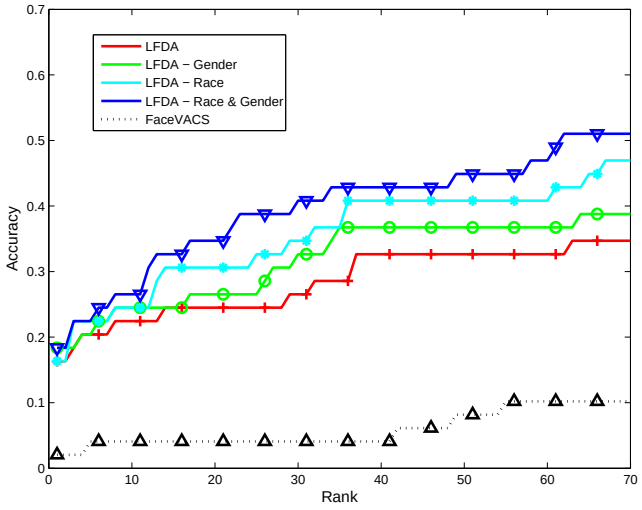


Fig. 8. Performance of matching *good* sketches with and without using ancillary demographic information (race and gender) to filter the results.

framework presented in Section II. Our matching uses sum-score fusion of MLBP and SIFT LFDA, as this was the highest performing method for matching viewed sketches (Table III). As mentioned previously, we segerated the 159 sketches into 49 *good* sketches and 110 *poor* sketches based on how accurately the sketch depicts the photo. Figure 1 shows examples of these two different categories of sketches.

The performance of matching sketches classified as *good* and *poor* can be found in Figure 7. There is a substantial difference in how well a forensic sketch can be matched if it has a reasonably accurate depiction of the subject (*good* sketches) and an inaccurate depiction of the face (*poor* sketches). Despite the fact that *poor* sketches are extremely difficult to match, the CMC plots in Figure 7 shows 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.

Figure 8 and Table V show the recognition performance when race and gender information is used to filter the gallery. By utilizing this ancillary information, we can significantly increased the performance of forensic sketch recognition. We noticed a larger performance gain from the race information than the gender information. This is likely due to the more uniform distribution of race membership than gender membership in our gallery. The use of other information such as age and height should offer further improvements.

Figure 9 demonstrates how recognition accuracy decreases as the gallery size increases. It can be seen that the rate of decrease begins to taper as the gallery grows. This indicates that the proposed LFDA method is scalable with respect to even larger sized galleries that will be encountered in real world scenarios.

Discriminatory information contained in individual face regions (eyes, nose, mouth, etc.) is shown in Figure 10. Again, this is achieved by first applying the masks in Figure 6 to the face features patches. These results mostly agree with cognitive science research (Section III-A) that indicated the external regions of the face were the most informative in matching forensic

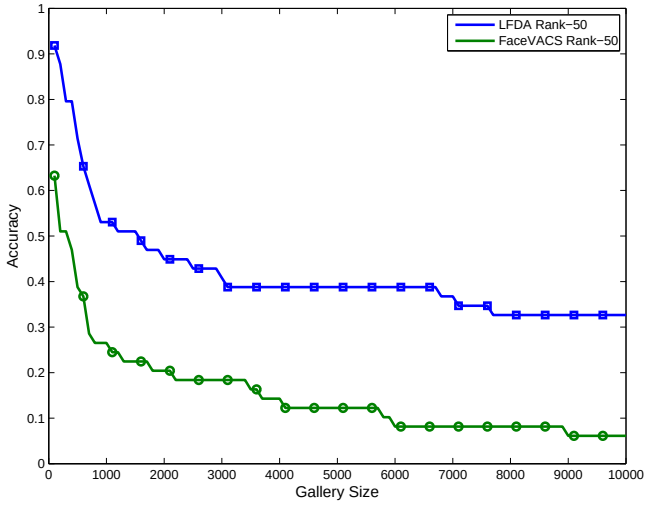


Fig. 9. Drop in Rank-50 accuracy using LFDA and FaceVACS on the *good* sketches without race and gender filtering as the gallery size increases.

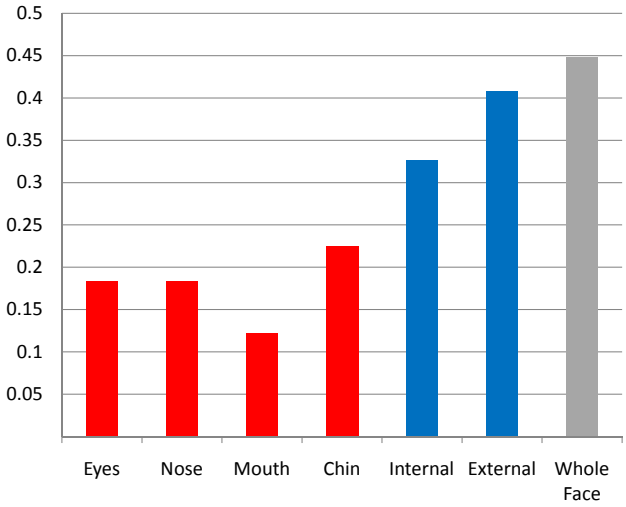


Fig. 10. Matching performance on the *good* sketches using race/gender filtering with SIFT and MLBP feature-based matching on only specific face regions.

sketches. Between the eyes, nose, mouth, and chin, we found the chin to be the most informative region of the face. In fact, only using the chin region for region recognition we were able to achieve Rank-50 accuracy of 22.45% with a gallery size of 10,159 images, which is interesting given the fact that the chin is not generally regarded as an overly valuable feature in face recognition.

Examples of failed retrievals are shown in Figure 11. While the top retrieved mugshots are wrong in these examples, the probe sketch appears to be more similar to these photos than the true photograph. This was nearly always the case: the top retrieved images appeared highly similar to the probe sketch in the incorrect matching.

Figure 12 shows three of the best matches and three of the worst matches amongst all the *good* sketches using the proposed LFDA recognition method. For each image, we have listed the match rank achieved by LFDA and FaceVACS, as well as from each of the individual facial regions discussed in Section III-B. By looking at the face region match ranks we can see that for different sketches different face regions are more salient.

One limitation of our study is the small number of forensic sketches, but obtaining a large collection of forensics sketches 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 had 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 matching forensic sketches. However, we believe that with a larger number of forensic sketches we could more properly train our discriminant and improve the matching performance. The problem in finding forensic sketches for our experiments is having the photograph mate. While forensic sketches exist from numerous crimes, even if there is an eventual identification of the subject, the mated

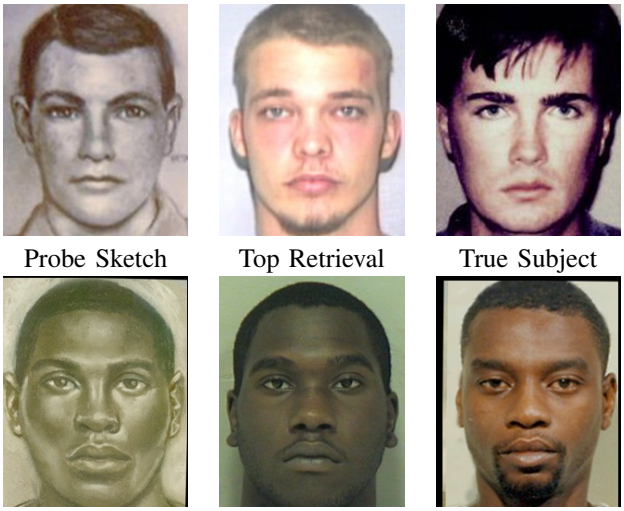


Fig. 11. Two examples of failed retrievals on *good* sketches, where the first column shows the probe sketches, the second column shows the incorrect Rank-1 matches, and the third column shows the true mated photographs. These are typical cases in which the true subject photo was not retrieved at rank 1, but the impostor subject retrieved at rank 1 looks more similar to the sketch than the true subject.

Method	Rank
LFDA	1
FaceVACS	320
Eyes	210
Nose	152
Mouth	1
Chin	33
Internal	3
External	3

Method	Rank
LFDA	1
FaceVACS	299
Eyes	198
Nose	159
Mouth	31
Chin	304
Internal	2
External	31

Method	Rank
LFDA	1
FaceVACS	2131
Eyes	5
Nose	2
Mouth	823
Chin	24
Internal	1
External	6

Method	Rank
LFDA	775
FaceVACS	2255
Eyes	166
Nose	298
Mouth	1776
Chin	3508
Internal	101
External	2231

Method	Rank
LFDA	1599
FaceVACS	215
Eyes	3237
Nose	2974
Mouth	1018
Chin	3742
Internal	3012
External	2402

Method	Rank
LFDA	1617
FaceVACS	429
Eyes	1992
Nose	3634
Mouth	3725
Chin	52
Internal	2246
External	692

Fig. 12. Example matches from the *good* quality forensic sketches with a background gallery of 10,159 subjects. We show three of the best matches (left) and three of the worst matches (right). Below each example are the rank scores obtained by using the proposed LFDA method, FaceVACS, and feature-based matching with different regions of the face using the masks shown in Figure 6.

sketch and photo are not often stored together in a central database. We are currently working with various law enforcement agencies to increase our dataset of forensic sketch pairs.

IV. Conclusions

In this paper we presented methods and experiments in matching forensic 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 a sketch recognition method, called *local feature-discriminant analysis (LFDA)*. Using SIFT and MLBP features in conjunction with a feature reduction method using discriminant analysis, we were able to perform robust sketch recognition. In matching viewed sketches LFDA performed better than 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 mugshot images. Further improvements to the LFDA method were achieved by utilizing ancillary information such as race and gender to filter of the 10,159 member gallery. We conducted an investigation of the saliency of different facial regions, which supported previous findings that the exterior regions of forensic sketches are more salient than the internal regions (a sharp contrast from traditional face recognition).

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

ACKNOWLEDGMENT

The authors would like to thank Professor Xiaoou Tang for sharing the CUHK sketch database. We would also like to thank Lois Gibson, Karen Taylor, Sheila Meese (and the rest of the Michigan State Police forensic artists), and Scott McCallum (PCSO) for providing us with forensic sketches and the mated photographs. We would like to further thank Lois Gibson for granting us permission to publish her sketches. Anil Jain's research was partially supported by the World Class University program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology (R31-2008-000-10008-0).

REFERENCES

[1] B. Klare and A. Jain, "Sketch to photo matching: A feature-based approach," in *Proc. SPIE Conference on 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, 2004.
[3] Q. Liu, X. Tang, H. Jin, H. Lu, and S. Ma, "A nonlinear approach for face sketch synthesis and recognition," in *Proc. of IEEE Conference on Computer Vision & Pattern Recognition*, 2005, pp. 1005–1010.
[4] J. Zhong, X. Gao, and C. Tian, "Face sketch synthesis using e-hmm and selective ensemble," in *Proc. of IEEE Conference on Acoustics, Speech and Signal Processing*, 2007.
[5] X. Wang and X. Tang, "Face photo-sketch synthesis and recognition," *IEEE Trans. Pattern Analysis & Machine Intelligence*, vol. 31, no. 11, pp. 1955–1967, Nov. 2009.
[6] S. L. (ed.), *Encyclopedia of Biometrics*. Springer, 2009.
[7] X. Tang and X. Wang, "Face sketch synthesis and recognition," in *Proc. of IEEE International Conference on Computer Vision*, 2003, pp. 687–694.
[8] D. Lin and X. Tang, "Inter-modality face recognition," in *Proc. of European Conference on Computer Vision*, 2006.
[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] X. Gao, J. Zhong, D. Tao, and X. Li, "Local face sketch synthesis learning," *Neurocomputing*, vol. 71, no. 10-12, pp. 1921–1930, 2008.
[11] X. Gao, J. Zhong, J. Li, and C. Tian, "Face sketch synthesis algorithm based on e-hmm and selective ensemble," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 4, pp. 487–496, April 2008.
[12] W. Liu, X. Tang, and J. Liu, "Bayesian tensor inference for sketch-based facial photo hallucination," in *Proc. of 20th International Joint Conference on Artificial Intelligence*, 2007.
[13] 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," in *Proc. of IEEE Conference on Biometrics: Theory, Applications and Systems*, 2009.
[14] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
[15] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Analysis & Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.
[16] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
[17] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Analysis & Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.
[18] R. Lienhart and J. Maydt, "An extended set of haar-like features for rapid object detection," in *Proc. of International Conference on Image Processing*, 2002.
[19] L. Wiskott, J.-M. Fellous, N. Kruger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," *Proc. of International Conference on Image Processing*, vol. 1, p. 129, 1997.
[20] T. Ahonen, A. Hadid, and M. Pietikinen1, "Face recognition with local binary patterns," in *Proc. of European Conference on Computer Vision*, 2004.
[21] FaceVACS Software Developer Kit, Cognitec Systems GmbH, http://www.cognitec-systems.de.
[22] A. Martinez and R. Benavente, "The AR face database," in *CVC Technical Report 24*, 1998.
[23] K. Messer, J. Matas, J. Kittler, and K. Jonsson, "XM2VTSDB: The extended M2VTS database," in *Proc. of Audio and Video-based Biometric Person Authentication*, 1999.
[24] A. Ross and A. Jain, "Information fusion in biometrics," *Pattern Recognition Letters*, vol. 24, no. 13, pp. 2115–2125, 2003.
[25] R. Uhl and N. Lobo, "A framework for recognizing a facial image from a police sketch," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, 1996.
[26] 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.
[27] A. W. Young, D. Hay, K. H. McWeeny, B. M. Flude, and A. W. Ellis, "Matching familiar and unfamiliar faces on internal and external features," *Perception*, vol. 14, pp. 737–746, 1985.
[28] L. Gibson, *Forensic Art Essentials*. Elsevier, 2008.
[29] K. Taylor, *Forensic Art and Illustration*. CRC Press, 2001.
[30] V. Bruce, Z. Henderson, K. Greenwood, P. Hancock, A. Burton, and P. Miller, "Verification of face identities from images captured on video," *Journal of Experimental Psychology: Applied*, vol. 5, no. 4, pp. 339–360, 1999.
[31] C. Frowd, V. Bruce, A. McIntyr, and P. Hancock, "The relative importance of external and internal features of facial composites," *British Journal of Psychology*, vol. 98, no. 1, pp. 61–77, 2007.
[32] P.-H. Lee, G.-S. Hsu, and Y.-P. Hung, "Face verification and identification using facial trait code," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, June 2009, pp. 1613–1620.

[33] A. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4–20, Jan. 2004.
[34] A. K. Jain, S. C. Dass, K. Nandakumar, and K. N, "Soft biometric traits for personal recognition systems," in *Proc. of International Conference on Biometric Authentication*, 2004.