

The FaceSketchID System: Matching Facial Composites to Mugshots

Scott J. Klum, *Student Member, IEEE*, Hu Han, *Member, IEEE*, Brendan F. Klare, *Member, IEEE*,
and Anil K. Jain, *Fellow, IEEE*

Abstract—Facial composites are widely used by law enforcement agencies to assist in the identification and apprehension of suspects involved in criminal activities. These composites, generated from witness descriptions, are posted in public places and in the media with the hope that some viewers will provide tips about the identity of the suspect. This method of identifying suspects is slow, tedious, and may not lead to the timely apprehension of a suspect. Hence, there is a need for a method that can automatically and efficiently match facial composites to large police mugshot databases. Because of this requirement, facial composite recognition is an important topic for biometrics researchers. While substantial progress has been made in non-forensic facial composite (or viewed composite) recognition over the past decade, very little work has been done using operational composites relevant to law enforcement agencies. Furthermore, no facial composite to mugshot matching systems have been documented that are readily deployable as standalone software. Thus, the contributions of this paper include: (i) an exploration of composite recognition use cases involving multiple forms of facial composites, (ii) the FaceSketchID System, a scalable and operationally deployable software system that achieves state-of-the-art matching accuracy on facial composites using two algorithms (holistic and component-based), and (iii) a study of the effects of training data on algorithm performance. We present experimental results using a large mugshot gallery that is representative of a law enforcement agency's mugshot database. All results are compared against three state-of-the-art commercial-off-the-shelf (COTS) face recognition systems.

Index Terms—Facial composite recognition, hand-drawn composite, software-generated composite, surveillance composite, mugshot, holistic face recognition, component-based face recognition

I. INTRODUCTION

FACIAL composites are commonly used in law enforcement to assist in identifying suspects involved in a crime when no facial image of the suspect is available at the crime scene (e.g., from a surveillance camera or a mobile phone). After a composite of a suspect's face is created, authorities disseminate the composite to law enforcement and media outlets with the hope that someone will recognize the individual and provide pertinent information leading to an arrest. Facial composites are particularly valuable when

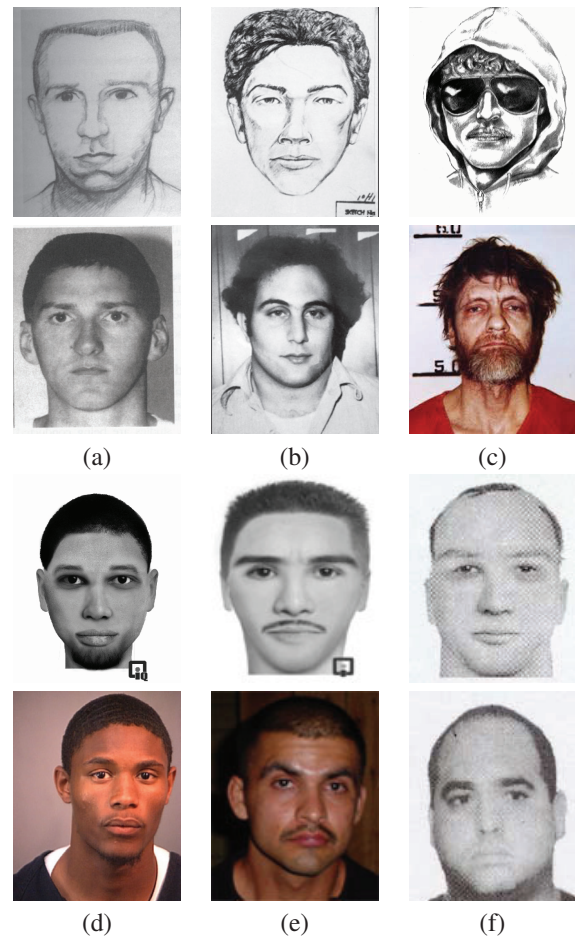


Fig. 1. Examples of facial composites that were used in cases in which the suspect was successfully apprehended. Examples of hand-drawn composites and their mugshot mates are shown for Timothy McVeigh (the Oklahoma City bomber) (a) [2], David Berkowitz (Son of Sam) (b)¹, and Ted Kaczynski (the Unabomber) (c)². Software-generated composites (d,e,f)³ that were created using the software FACES [3] are shown with mated mugshots.

eyewitness' or victim's descriptions are the only form of evidence available [1]. Unfortunately, this process is inefficient and does not leverage all available resources, in particular, the extensive mugshot databases maintained by law enforcement agencies. Successful techniques for automatically matching facial composites to mugshots will improve the effectiveness of facial composites and allow for faster apprehension of suspects.

Copyright © 2013 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

S.J. Klum, H. Han, and A.K. Jain are with the Department of Computer Science and Engineering, Michigan State University, Room 3115, 428 S. Shaw Lane, Engineering Building, East Lansing, MI 48824-1226. E-mail: {klumscot, hhan, jain}@msu.edu.

B.F. Klare is with Noblis, 3150 Fairview Park Drive, Falls Church, VA 22042. E-mail: brendan.klare@noblis.org.



Fig. 2. Examples of surveillance images that are of sufficiently poor quality such that COTS matchers are expected to fail to find the true mate in a mugshot database. Surveillance composites can be drawn based on these images, which can be used to identify a suspect more accurately. Surveillance composites shown are drawn by Sandra Enslow [4].

Facial composites used in law enforcement can be divided into three categories:

- (i) *Hand-drawn composites*: Facial composites drawn by forensic artists based on the description provided by a witness. Hand-drawn composites have been used in criminal investigations dating as far back as the 19th century [5]. Examples of high profile cases in which a hand-drawn composite was used are shown in Figs. 1 (a,b,c).
- (ii) *Software-generated composites*: Facial composites created using software kits which allow an operator to select various facial components (Figs. 1 (d,e,f)). Software-generated composites have become a popular and more affordable alternative to hand-drawn composites. According to [5], 80% of law enforcement agencies report using some form of software to create facial composites of suspects. We note that, based on conversations with law enforcement agencies, the actual adoption and use of composite-generation software may be lower than reported in [5].
- (iii) *Surveillance composites*: Facial composites drawn by forensic artists based on poor quality surveillance images (Fig. 2). Surveillance composites are used in scenarios when COTS systems are expected to fail on query (probe) face images (due to poor lighting, off-pose faces, occlusion, etc.).

Whereas forensic artists typically require a few years of training to become proficient in drawing composites, only a few hours of training are typically required before a police officer can start using composite-generation software. Irrespective of the quality and capability of the software, most composite software relies on choosing a set of facial components (e.g., eyes, nose, mouth) based on the information contained in the witness' description. It is important to emphasize that irrespective of the method used to generate the composite, the

quality of the resulting composite (namely, its resemblance to the suspect's real face) mainly depends on the accuracy of the description provided by the witness and the skill of the artist/operator. Section III describes the construction process for each type of facial composite used in this study. We note that in [6], we referred to hand-drawn composites as "forensic sketches" and software-generated composites as "composite sketches". The naming conventions have been corrected in this study to reflect the prevailing law enforcement terminology.

While several methods that match viewed⁴ and hand-drawn composites to mugshots have been reported in the literature [7], [8], [11], [16], [10], [15], [14], [18], only a few methods have been published for automatic matching of software-generated composites to mugshots [16], [17]. In all the previous studies reported on software-generated composites, with the exception of [6], composites were created while the operator was viewing the high quality mugshot. This type of viewed composite does not accurately reflect the creation of composites for use in criminal investigations because the mugshot of the suspect is unknown or unavailable. Indeed, there would be no need to create the composite if we knew the suspect and had his mugshot. While surveillance composites are also created while viewing images of the suspect, these images are of poor quality compared with mugshots. We know of no studies that report performance when matching surveillance composites to mugshots. An extended review of facial composite to mugshot matching literature is presented in Section II.

To our knowledge, no system is available that is designed for facial composite to mugshot matching and is deployed at law enforcement agencies, though investigators often attempt to use COTS matchers in this manner with limited success. To address this need, we have developed the FaceSketchID System as a standalone software system that can match facial composites to their mugshot mates with state-of-the-art accuracy. System specifications can be found in Section IV. The FaceSketchID System uses two complementary al-

¹<http://www.trutv.com/library/crime/photogallery/serial-killer-sketches.html?curPhoto=4>

²http://en.wikipedia.org/wiki/Ted_Kaczynski

³http://www.facesid.com/mediacenter_frontline_stories.html

⁴Many studies on facial composite to photograph matching have relied on viewed composites in which the composite is drawn by hand while viewing the photograph.

TABLE I
PRIOR WORK ON FACIAL COMPOSITE TO PHOTOGRAPH MATCHING

	Publication	Approach	Limitations
Viewed Hand-Drawn Composites	Tang and Wang [7]	Photo-to-composite conversion using eigentransform	Viewed composites are not of any value in law enforcement and forensics applications. Methods that convert composite to photograph or vice versa are often solving a more difficult problem than the facial composite to photograph matching task.
	Liu et al. [8]	Photo-to-composite conversion using locally linear embedding	
	Gao et al. [9]	Photo-to-composite conversion using embedded hidden Markov model	
	Wang and Tang [10]	Photo-to-composite conversion using multiscale Markov random field model	
	Lin and Tang [11] Zhang et al. [12]	Common discriminant feature extraction Principle Component Analysis (PCA) based algorithm	
Hand-Drawn Composites	Uhl and Lobo [13]	Photometric standardization	Software-generated composites, which are widely used in law enforcement, were not considered.
	Klare and Jain [14]	SIFT and MBLP feature descriptors with local-feature based discriminant analysis	
	Bhatt et al. [15]	Multi-scale circular Weber's local descriptor	
Software- Generated Composites	Yuen and Man [16]	Point distribution model and geometrical relationship	Composites were created while viewing the photograph of the subject (viewed software-generated composites). Hand-drawn composites were not considered.
	Han et al. [17]	Component based representation using MLBP descriptors	
			Contributions
Proposed Method		Facial composite to mugshot matching algorithms are deployed in the FaceSketchID System. We fuse the match scores of two different (holistic and component-based) algorithms to boost the matching performance.	Hand-drawn composites, software-generated composites, and surveillance composites are considered. We investigate the effects of training the algorithms on different types of (composite, photograph) data.

gorithms when matching facial composites to mugshots: (i) a holistic algorithm and (ii) a component-based algorithm, that are described in Sections V-A and V-B, respectively. It is important to point out that the matching performance of the FaceSketchID System critically depends on the quality of composite as well as the difference between the time the mugshot in the database was captured and the time the composite was created. Nevertheless, while the accuracy of composite to mugshot matching is significantly lower than mugshot to mugshot matching, composite to mugshot matching systems are needed to maximize the opportunity of apprehending suspects in heinous and egregious crimes where the evidence in the form of a suspect's photograph is lacking.

Section VI describes the experiments used to evaluate the FaceSketchID System. Experimental results when matching hand-drawn, software-generated, and surveillance composites to their mugshot mates are reported in Section VII. Three COTS face matchers are used to establish baseline recognition accuracy when matching facial composites to mugshots. All COTS face matchers used in our experiments have been studied in FRVT⁵. We also investigate the influence of the type of data used to train algorithms used in the FaceSketchID System. To facilitate comparison with previously published results, we also detail the FaceSketchID's performance when matching viewed software-generated composites to photographs.

II. RELATED WORK

Automated face matching between two facial photographs is a well studied problem in computer vision and biometrics [19]. However, matching facial composites to photographs is

a more challenging problem with only a limited amount of published work, some of which include: [7], [8], [11], [16], [10], [14], [15], [18], [17], [6]. Of these, most studies have used composites drawn while viewing the mugshot or photograph (viewed hand-drawn composites). Further, the studies that considered operational hand-drawn composites did not address the use of software-generated composites which are reported to be widely used by law enforcement agencies.

To our knowledge, only two previous studies focused on automatic face recognition systems using software-generated composites. The first used a combination of local and global features to represent composites [16], but it required user input in the form of relevance feedback in the matching or recognition phase. Further, the authors in [16] used a small gallery in their experiments (300 facial photographs). The method proposed by Han et al. [17], used a component based approach to match facial composites to mugshots. While Han et al. used a larger gallery with 10,000 mugshots and created a matching method that is fully automatic, the software-generated composites used were created while viewing the mugshot photograph (viewed software-generated composites) and therefore do not reflect operational scenarios.

Our work uses hand-drawn composites from criminal investigations and software-generated composites created using descriptions from volunteers given two days after viewing a mugshot, mimicking a witness of an actual crime scene. Furthermore, we compare the recognition accuracy of hand-drawn composites when algorithms are trained using different training data sets. We show the improved performance of matching facial composites to mugshots when the match scores of two different algorithms (holistic and component-based) are fused. We also detail the use of surveillance

⁵<http://www.nist.gov/itl/iad/ig/frvt-home.cfm>

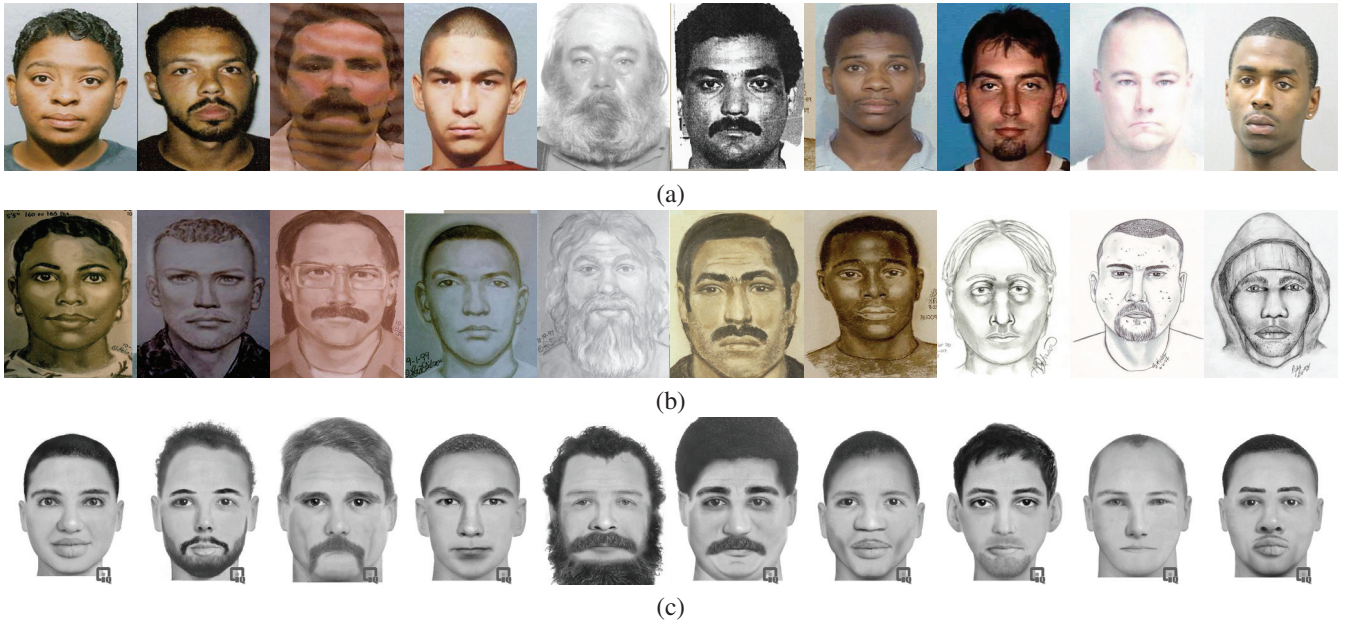


Fig. 3. Examples of mugshots (a) and mated hand-drawn composites (b) and software-generated composites (c) created using FACES [3]. The hand-drawn composites shown were drawn by either Lois Gibson [20] or forensic artists at the Michigan State Police (MSP).

composites, which have not previously been reported in the literature. All experimental results are based on comparisons against mugshot mates for the specified facial composites and an extended gallery of 100,000 mugshots. The size of our gallery is representative of a law enforcement agency's mugshot database. A summary of related work can be found in Table I.

This publication extends upon the work in [6] in the following ways: (i) the background database is increased to 100,000 versus 10,000 in [6], (ii) integrated and generalized the approaches in [6], [17], [14] and [21] into a unified framework, (iii) improved algorithms such that the recognition performance is comparable to [6] after increasing the background gallery size by an order of magnitude, (iv) evaluated the FaceSketchID System on the largest reported collection of operational (composite, mugshot) pairs reported in the literature, (v) evaluated the effect of the type of training data on the performance of our recognition system, and (vi) demonstrated that composites made from poor quality surveillance imagery can be used to identify the suspect that could not be identified based on the original image.

III. CONSTRUCTING A FACIAL COMPOSITE

As previously mentioned, law enforcement agencies rely on three modalities of facial composites: (i) hand-drawn composites, (ii) software-generated composites and (iii) surveillance composites. Hand-drawn composites (Section III-A) are drawn based on a verbal description. Typically, hand-drawn composites are drawn by a forensic artist with special training. Similarly, software-generated composites (Section III-B) are drawn based on a verbal description, but are created using menu-driven software. In most composite software, the operator selects from a set of facial components to synthesize a face.

Facial composites can also be created using poor quality or off-pose facial images. These surveillance composites, which are used when COTS face matchers fail, are described in Section III-C. Figure 3 shows example hand-drawn and software-generated composites along with mated mugshots that are used in our experiments. For the remainder of this paper we will use the (query modality, target modality) ordered pair convention to denote matching scenarios.

A. Hand-Drawn Composites

All hand-drawn composites used in our study were created by forensic artists for real-world criminal investigations. To create a hand-drawn composite, an artist draws a face based on descriptions provided by either one or multiple eyewitnesses. For this type of composite, the time between observation and recall by a witness varies depending on the circumstances. A total of 265 hand-drawn composites along with their mated mugshots are used in our experiments, which we will refer to as the Pattern Recognition and Image Processing (PRIP) Hand-Drawn Composite (PRIP-HDC) database. Of the 265 total hand-drawn composites, 73 were drawn by Lois Gibson [20], 43 were drawn by Karen Taylor [2], 56 were provided by the Pinellas County Sheriff's Office (PCSO), 46 were drawn by forensic artists employed by the Michigan State Police (MSP), and 47 were downloaded from the Internet.

It is worth noting that some of the hand-drawn composites and their mugshot mates, namely those acquired from [2], were scanned out of a textbook. These pairs, while captured at a high resolution, consist of small printer dots. To compensate for this artifact, a Gaussian smoothing ($\sigma=1.25$) was applied to each pair from this set. In general, this significantly degraded the matching performance of composites that were smoothed, presumably because a source of bias (the printer dots) was eliminated. Specifically, 23 of the 43 total composites were not

TABLE II
MITIGATING THE EFFECT OF POOR QUALITY COMPOSITES

	Before Smoothing	After Smoothing
Hand-drawn composites from [2]		
Recognized by Rank-200 (%)	72.09	18.06
PRIP-HFC Rank-200 Performance (%)	26.03	21.50

Applying a Gaussian smoothing ($\sigma=1.25$) significantly degrades the matching accuracy for the 43 (hand-drawn composite, mugshot) pairs from [2] using the holistic algorithm. However, the overall matching performance only degrades slightly, suggesting the FaceSketchID System is more robust when trained on the artifact-corrected (hand-drawn composite, mugshot) pairs.

longer matched by Rank-200 when using the holistic algorithm (described in Sec. V-A). One would expect the overall matching performance to degrade by about 8.8% (corresponding to 23 out of the 265 hand-drawn composites in the PRIP-HDC database). However, the retrieval ranks of other hand-drawn composites improved such that the overall Rank-200 matching performance only dropped by about 5% (Table II). Other studies which have used sketches from [2] have not reported eliminating this bias.

B. Software-Generated Composites

A number of software systems are available to create composites: E-FIT [22], EvoFit [23], FACES [3], Identi-Kit [24], Mac-a-Mug [25], and Photo-Fit [25]. Of these, Identi-Kit and FACES are most widely used by law enforcement agencies in the United States [5]. Both Identi-Kit and FACES allow users to choose from a set of candidate components or features (i.e. eyes, mouth, nose). FACES provides a larger number of features and options, and it has been observed to be more accurate in capturing facial characteristics than Identi-Kit [17]. For these two reasons, we used FACES to create composites for our matching experiments.

To create the software-generated composites, we used a procedure designed to mimic real-world composite synthesis detailed in [26]. Volunteers (adults ranging from 20-40 years of age) were asked to view a mugshot of a suspect for one minute. Two days later they were asked to describe the mugshot to the FACES software operator (one of the authors of this paper) who had not seen the mugshot. Volunteers also provided demographic information about the suspect to the best of their ability (gender, race/ethnicity, age range). During the description process, the FACES operator used a cognitive interview technique [27] to enhance the volunteer's memory of the suspect's facial features in the mugshot. To reduce the problem of operator contamination [26], in which previously created composites influence the creation of the current composite, a random face was generated initially which was then modified based on the volunteer's description.

We note that there are certain limitations in creating software-generated composites. For example, it is difficult to achieve certain types of shading and skin texture in the composite. The options for localizing a component on the face are limited, and therefore achieving the desired alignment of components is also challenging. In total, 75 software-generated

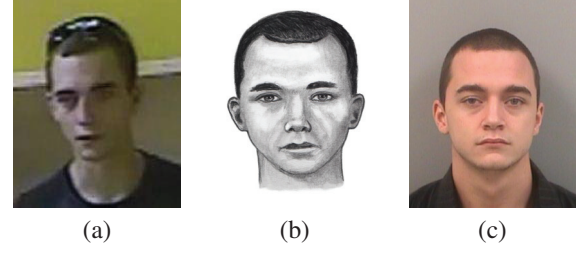


Fig. 4. An example surveillance image (a) and the corresponding surveillance composite (b) and mugshot (c) used in this study. All surveillance data was provided by the PCSO.

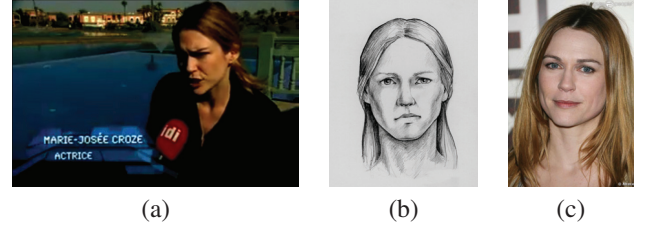


Fig. 5. An example video frame (a) and the corresponding surveillance composite (b) and photograph (c) from [28]. The subject shown is Marie-Josée Croze.

composites were synthesized, each taking 30 minutes to create, on average. This database will be referred to as the PRIP Software-Generated Composite (PRIP-SGC) database.

C. Surveillance Composites

Given the ubiquity of surveillance technology, law enforcement agencies attempt to make use of all of the facial image data at their disposal regardless of its quality. At the lower end of the quality spectrum are images captured by cell phones, retail surveillance cameras, and ATM cameras which often are blurred, have significant shadows or occlusion, or contain an off-pose face. In some cases, these facial images are of sufficient quality to be used in a COTS face matcher. For the majority of these images, however, COTS face matchers fail to find the corresponding individual within a mugshot gallery. To make use of this poor quality data, law enforcement agencies often employ a forensic artist to create a high quality facial composite from the surveillance face image. We investigate the possibility of using this form of facial composite to improve upon the matching performance of COTS matchers using a set of (surveillance composite, mugshot) pairs provided to us by the PCSO (Fig. 4) and a set of pairs from [28] (Fig. 5).

D. Viewed Composites

While viewed composites are not applicable in forensic scenarios, we have found them to be useful during algorithm training. Eighteen-hundred (viewed hand-drawn composite, photograph) pairs used in our study are available from the Chinese University of Hong Kong (CUHK)⁵. The CUHK Face Sketch database [10] contains 188 pairs from the CUHK student database, 123 pairs from the AR database [29], 295 pairs from the XM2VTS database [33], and 1,194 pairs [34] from the FERET database [35]. We will refer to this set of

TABLE III
DIFFERENCES IN SYSTEM CHARACTERISTICS WITH RELATED WORK

Publication	Gallery Size	Enrollment Speed	Comparison Speed	Template Size
Tang and Wang [7]	100	Not reported	Not reported	Not reported
Liu et al. [8]	300	Not reported	Not reported	Not reported
Gao et al. [9]	Not reported	63% faster than [8]	Not reported	Not reported
Wang and Tang [10]	300	0.0055 per sec. ^a	Not reported	Not reported
Lin and Tang [11]	350	Not reported	Not reported	Not reported
Zhang et al. [12]	100	Not reported	Not reported	Not reported
Uhl and Lobo [13]	16	Not reported	Not reported	Not reported
Klare and Jain [14]	10,159	Not reported	Not reported	Not reported
Bhatt et al. [15]	7,063	10.41 per sec. ^b	Not reported	Not reported
Yuen and Man [16]	300	Not reported	Not reported	Not reported
Han et al. [17]	10,159	Not reported	Not reported	Not reported
Proposed	100,000+	1.07 per sec. per thread ^c	22,000 per sec. per thread	5.73 KB

^aUnspecified 3 GHz machine

^b2 GHz Intel Core Duo (time reported for descriptor computation only)

^c2.9 GHz Intel Core i7 (time includes landmark detection)

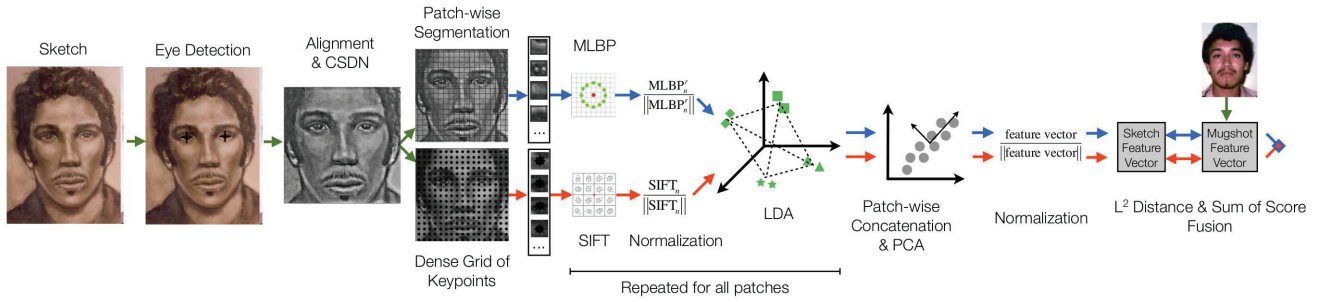


Fig. 7. The holistic algorithm pipeline used by the FaceSketchID System. Following normalization and application of a center-surround divisive normalization (CSDN) [30] filter, SIFT [31] and MLBP [32] features are extracted. Optimal subspaces are learned for each patch and the projected features are concatenated. After a PCA step to reduce template size, the final feature vector is normalized using the L^2 norm. The similarity scores from the SIFT and MLBP feature vectors when comparing composite and mugshot are fused using the sum rule.

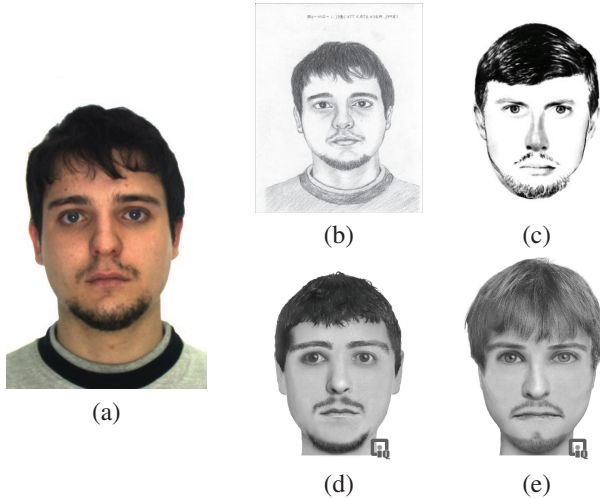


Fig. 6. For a given photograph (a) from the AR database [29], the viewed facial composites used in our study consist of a hand-drawn composite (b), a composite created using Identi-Kit (c), and two composites created using FACES (d, e) by two different operators.

1800 (viewed hand-drawn composite, photograph) pairs as the CUHK-VHDC database.

Additionally, a set of viewed hand-drawn composites were drawn by forensic artists at the MSP for 93 of the 265 mugshots in the PRIP-HDC database. We investigate the effects of training our matching algorithms on these composites, which we will refer to as the PRIP Viewed Hand-Drawn Composite (PRIP-VHDC) database. Eighty-five of the PRIP-VHDC subjects have a single facial composite, and the remaining 8 have two composites each.

To demonstrate the strength of the FaceSketchID System compared to previously reported results, we include a set of viewed software-generated composites in our matching experiments from [17]. This data will be referred to as the PRIP Viewed Software-Generated Composite (PRIP-VSGC) database. For each of the 123 photographs from the AR database used in the PRIP-VSGC database, three composites were created. Two composites were created using FACES and the third was created using Identi-Kit. Examples of viewed composites used in our experiments can be found in Fig. 6.

⁵<http://mmlab.ie.cuhk.edu.hk/facesketch.html>

IV. THE FACEskETCHID SYSTEM

The FaceSketchID System was developed to address the lack of a fully automatic means to match facial composites to mugshots. The matching algorithms used by the FaceSketchID System are described in Sections V-A and V-B. The FaceSketchID System supports a drag-and-drop enrollment interface with options for manually modifying detected eye locations, viewing both probe and target images after algorithm processing, viewing similarity heatmaps of mugshot matches, and searching for known individuals by name within the mugshot matches. The FaceSketchID System also supports filtering the mugshot gallery via demographic information in the form of age range, race, and gender. To simplify deployment, gallery images can be enrolled to and accessed from remote locations (e.g. an off-site server).

The FaceSketchID System is compatible with Windows, OSX, and Ubuntu Linux environments. Source code for the FaceSketchID System is written in C++. The FaceSketchID System uses OpenCV [36] as a matrix library, Eigen [37] for statistical learning, and Qt [38] for the GUI. Some modules of the FaceSketchID System are available in OpenBR [39]. On a 2.9 GHz Intel Core i7 laptop with 8 GB of RAM, enrollment (including eye detection) and matching speeds are 1.07 templates per second per thread and about 22,000 comparisons per second per thread, respectively. Templates (including demographic information) are approximately 5.73KB in size. Table III lists the major systematic and experimental differences between this study and related work. A video demonstrating the matching process can be found at http://biometrics.cse.msu.edu/images/ImgProjects/help_match.mp4.

V. FACIAL COMPOSITE TO MUGSHOT MATCHING ALGORITHMS

The FaceSketchID System leverages two complementary algorithms when matching facial composites to their mugshot mates. The first, developed by Klare and Jain [18], is designed for use in heterogeneous face recognition. That is, it is effective not only when matching (facial composite, mugshot) pairs, but also in near-infrared, thermal, and cross-distance matching scenarios. The second, developed by Han et al. [17], was originally designed for software-generated composite to mugshot matching. Modifications made to the original algorithms in [18] and [17] to improve matching accuracy, template size, and algorithm speeds are also discussed. In the case of the component-based algorithm, modifications are heavily influenced by the algorithm in [21]. Each algorithm is described in detail in the following sections.

A. Holistic Algorithm

The holistic algorithm used by the FaceSketchID System has been found to be an effective technique for matching a facial composite probe against a gallery of mugshots [18]. One strength of the holistic algorithm is that it represents both facial composites and mugshots with local descriptor-based features, eliminating the need to synthesize a pseudo-composite from the mugshot as is done in [7] and [9]. Thus, the algorithm for

TABLE IV
HOLISTIC ALGORITHM DIFFERENCES COMPARED TO [18]

Modification	Effect
Improved eye localization	Improved accuracy
Removed Gaussian and difference of Gaussian (DoG) filter pipelines	Smaller templates, faster algorithm, improved accuracy
Changed RS-LDA to LDA	Smaller templates, faster algorithm
C++ implementation (versus Matlab)	Smaller templates, faster algorithm
Parameter tuning (MLBP radii, keypoint density, PCA variance retained, etc.)	Smaller templates, faster algorithm, improved accuracy

representing a mugshot is analogous to a facial composite and will be omitted from the following description.

After detecting eye locations, the facial composite is normalized to a fixed height and width and transformed such that right and left eyes are at the same position for every composite. The center-surround divisive normalization (CSDN) filter [30] is then applied to the composite to compensate for the differences related to the change in modality between composite and mugshot. Although [18] uses additional filters to improve generalization across multiple modalities, our empirical study shows that using only the CSDN filter results in the best matching performance for composites, and the resulting reduction in complexity increases the algorithm speed.

Subsequently, SIFT [31] and multi-scale local binary pattern (MLBP) [32] features are extracted in parallel from a dense grid and uniform patches across the face, respectively. Note that the dense grid used as keypoints for SIFT descriptor extraction and the patches used to compute MLBP features correspond to the same locations within the composite. For both SIFT and MLBP features, optimal subspaces are learned for each patch using linear discriminant analysis (LDA). Klare and Jain used a random-sample LDA (RS-LDA) technique introduced in [41] to handle the small sample size problem. Since we have a relatively more training data, RS-LDA has been replaced by LDA in the FaceSketchID System to improve algorithm speed and template size.

After learning an optimal subspace for each patch and projecting the patch-wise features into their respective subspaces, the projected features are concatenated to form a single feature vector. PCA is applied to the feature vector to reduce template size, and the resulting feature vector is normalized using the L^2 norm. To measure the similarity between feature vectors, the holistic algorithm uses the L^2 similarity measure. After z-score normalization, scores from the SIFT and MLBP representations are fused via a sum-of-score fusion rule with equal weight applied to both representations. A diagram of the holistic algorithm pipeline is shown in Fig. 7. Table IV summarizes key differences between the holistic algorithm in the FaceSketchID System vs. the algorithm presented in [18].

B. Component-Based Algorithm

As mentioned earlier, the component-based method used by the FaceSketchID System was proposed in [17] to match software-generated composites (created using FACES and

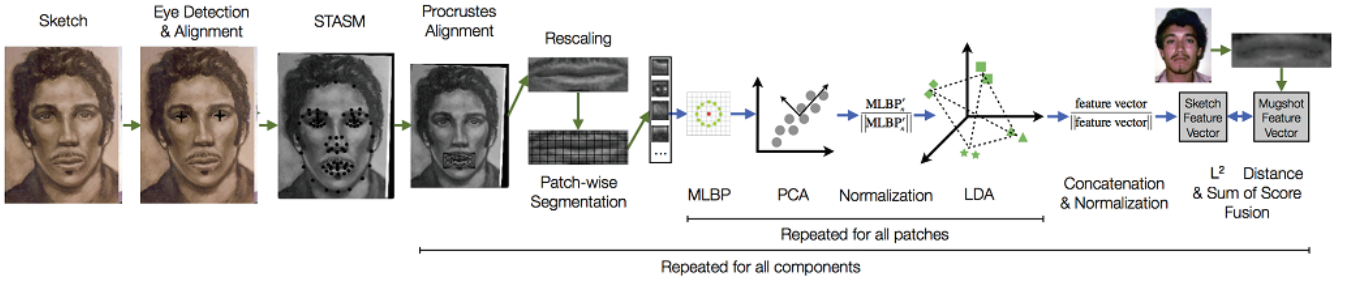


Fig. 8. The component-based algorithm pipeline used by the FaceSketchID System. Following normalization, STASM [40] is used to detect 77 facial landmarks. For each of the 6 components used, procrustes alignment is used to ensure maximum alignment of landmarks. MLBP features are extracted from patches inside the facial component. A PCA step is used to reduce the noise in the MLBP descriptor for a given patch. Optimal subspaces are learned for each patch and the projected features are concatenated and normalized. In the interest of brevity, we omit the final PCA step used to reduce the template size in this diagram. L^2 similarity scores are generated for each of the selected components. The overall (facial composite, mugshot) similarity score is the sum of the individual component scores.

Identi-Kit) to photographs. In the component-based algorithm, facial components are automatically localized by detecting landmarks with an active shape model (ASM) via the STASM library [40]. As in [21], the ASM is initialized using eye locations provided by a COTS eye detector that is bundled with the FaceSketchID System. To evaluate the landmarking performance of STASM on mugshots versus composites, we have performed an experiment in which 20 (composite, mugshot) pairs are randomly selected and marked with 5 ground-truth landmarks (both eye centers, nose tip, and mouth corners). After running STASM on each composite and mugshot, we calculate the average landmark location error (in pixels) after normalizing based on the interpupillary distance (IPD). For the 20 mugshots and 20 composites, the average errors are 0.0421 ($\sigma=0.0362$) and 0.0417 ($\sigma=0.0386$), respectively. From these results, we contend that although the STASM models were developed for facial landmark detection in photographs, they generalize well to facial composites.

While alignment is critical for the holistic algorithm, it is arguably more important when extracting landmarks used by the component-based representation because the relatively small size of the facial components limits the descriptive tolerance to noise. Bonnen et al. [21] showed that Procrustes alignment of landmarks improved the matching accuracy, especially when matching off-pose facial images. In the facial composite to mugshot matching scenario, the problem of off-pose facial images is not severe. Nevertheless, using Procrustes alignment and a Delaunay triangulation-based texture mapping framework improved the performance of the component-based algorithm in the FaceSketchID System.

After component alignment and texture mapping, MLBP [32] descriptors are used to capture the texture and structure of patches in each facial component. A PCA step is used to reduce the noise present in the patch-wise MBLP representation for the given facial component. Similar to holistic algorithm, the component-based method uses LDA to learn the optimal subspace and improve recognition accuracy. A final PCA step is used to reduce template size.

In [17], cosine similarities between corresponding patches of facial components are computed and the overall compo-

TABLE V
COMPONENT-BASED ALGORITHM DIFFERENCES COMPARED TO [17]

Modification	Effect
Improved eye localization	Improved accuracy
Seeding ASM with eye locations	Improved accuracy
Procrustes alignment and texture mapping	Improved accuracy
Concatenate component and patches	Smaller templates, faster algorithm, improved accuracy
Changed RS-LDA to LDA	Smaller templates, faster algorithm
C++ implementation (versus Matlab)	Smaller templates, faster algorithm
Parameter tuning (patch size, component bounding boxes, PCA variance retained, etc.)	Smaller templates, faster algorithm, improved accuracy

nent similarity is the average of the patch-wise similarities. Experimental results indicate that concatenating the patch-wise feature vectors prior to computing an overall component similarity improves matching accuracy. More importantly, this technique reduces template size and increases comparison speed by orders of magnitude within the framework of the FaceSketchID System. As with [17], the most accurate components to be used during score fusion are determined empirically for each composite modality. Scores are normalized prior to fusion using z-score normalization and equal weights are applied to all components. The mouth, nose, and hair components are used when matching (hand-drawn composite, mugshot) pairs whereas the mouth, nose, hair, and brow components are used when matching (software-generated composite, mugshot) pairs. A diagram of the component-based algorithm pipeline is shown in Fig. 8.

A simple sum-fusion rule is used when fusing the match scores from both algorithms after z-score normalization. We note that we are using the L^2 distance metric for both algorithms for simplicity and consistency reasons. In the case of (hand-drawn composite, mugshot) recognition, equal weights (empirically determined) are assigned to match scores. However, when matching software-generated composites to

TABLE VI
DATABASE SUMMARY

Database	Details	No. of Pairs
PRIP-HDC	Hand-drawn composites (HDC) with mugshot mates	265
PRIP-SGC	Software-generated composites (SGC) with mugshot mates	75
PRIP-VHDC	Viewed hand-drawn composites (VHDC) with mugshot mates	93
PRIP-VSGC	Viewed software-generated composites (VSGC) with photograph mates	123
CUHK-VHDC	Viewed hand-drawn composites (VHDC) with photograph mates	1800



Fig. 9. Example mugshots from the 100,000 PCSO mugshots used to extend the gallery in our experiments.

mugshots, weights of 0.6 and 0.4 are assigned to the match scores of the holistic and component-based representations, respectively. Table V summarizes key difference between the component-based algorithm in the FaceSketchID System versus the algorithm presented in [17].

VI. EXPERIMENTAL PROTOCOL

As mentioned earlier, the gallery set for all experiments (with the exception those listed in Section VI-D) consists of mated mugshots for the listed composites plus a set of 100,000 mugshots from the PCSO. Experimental results are reported with and without filtering the mugshot gallery using demographic information (age range, race, and gender). Three commercial face matchers (referred to as COTS-1, COTS-2, and COTS-3) are used as baselines. Note that we are unable to train any of the COTS systems. Table VI summarizes the databases used in our study. Figure 9 shows example mugshots used to extend the gallery in matching experiments.

A. Hand-Drawn Composites

Experimental results for matching (hand-drawn composite, mugshot) pairs are reported based on a 5-fold cross-validation scheme. Training and testing sets are disjoint; that is, no subject that was used to train an algorithm was used when testing its performance. The 265 subjects in the PRIP-HDC dataset are assigned to a cross-validation fold via an MD5 hashing function based on the subject's identifier. Thus, the number of subjects in the testing subset of a given fold varies but is, on average, 53. The different training sets used to train

the two algorithms in the hand-drawn composite to mugshot matching experiments are as follows:

1) *PRIP-HDC*: Facial composite to mugshot matching algorithms are trained on the set of approximately 212 (hand-drawn composite, mugshot) pairs available per cross-validation fold in the PRIP-HDC database. Each subject in the PRIP-HDC database is only associated with two images: a hand-drawn composite and a mugshot.

2) *CUHK-VHDC*: Algorithms are trained on the set of 1,800 (viewed hand-drawn composite, photograph) pairs from the CUHK-VHDC database. Similar to the PRIP-HDC database, the CUHK-VHDC database contains only two images per subject. Note that although we no longer need to use cross-validation because we are training on one dataset and testing on another, we maintain the testing splits used in (1) to allow for comparison between experiments. This procedure is used in all the following experiments in which cross-validation is not necessary.

3) *PRIP-HDC + CUHK-VHDC*: Algorithms are trained on the set of approximately 212 (hand-drawn composite, mugshot) pairs available in the PRIP-HDC database as in protocol (1). For the training subset of each fold of the PRIP-HDC database, the entire CUHK-VHDC database is added.

4) *PRIP-HDC + CUHK-VHDC + PRIP-VHDC*: Algorithms are trained with the same protocol as in protocol (3). When the testing set does not contain subjects from the PRIP-VHDC dataset, those pairs are added to the training set.

We also report performance on a set of 32 (hand-drawn composite, mugshot) pairs which have been sequestered from the training experiments reported above. Note that cross-validation is not necessary in this experiment since we are training and testing on different data.

Finally, we report the performance of a composite of one of the Boston Marathon bombing suspects (see [42]) using the same experimental setup as the sequestered pairs. The composite depicts the younger brother, Dzhokhar Tsarnaev, drawn by an artist in the court during a preliminary trial⁴. While it can be considered as a viewed hand-drawn composite, we include the performance as an additional example of the capabilities of the FaceSketchID System.

B. Software-Generated Composites

Due to the relatively small size of the PRIP-SGC database, we could only train the two matching algorithms using the CUHK-VHDC database. Thus, there is no need for cross-validation when testing on software-generated composites.

C. Surveillance Composites

To evaluate the performance of the FaceSketchID System on (surveillance composite, mugshot) pairs, we train the holistic and component-based algorithms using the CUHK-VHDC database. We compare retrieval ranks when using (surveillance composite, mugshot) pairs in the FaceSketchID System versus retrieval ranks when using (surveillance image, mugshot) pairs in COTS matchers. For a given subject in this dataset, there are, on average, 6 composites created by different artists. Examples of this surveillance data can be found in Figs. 4

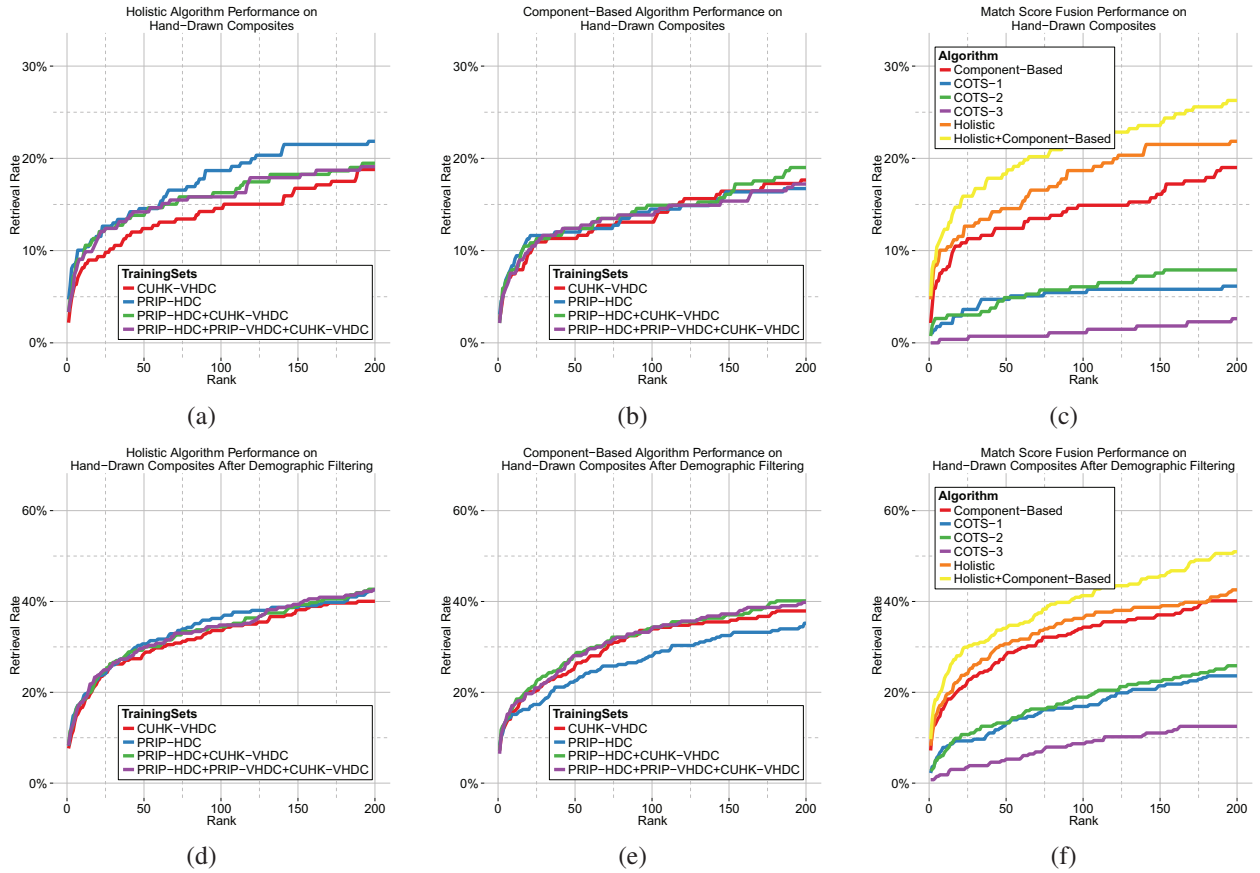


Fig. 10. Results for the holistic (a,d) and component-based (b,e) algorithms when matching hand-drawn composites to mugshots. The best performance is achieved through a fusion of the holistic and component-based algorithm match scores (c,f). Three COTS matchers are included in (c,f) as a baseline.

(a,b,c). We report the retrieval rank achieved using the most accurate composite and the most accurate frame from the surveillance video for each subject. We include a surveillance composite depicting Tamerlan Tsarnaev (the older brother alleged to be involved in the Boston Marathon bombing) [28] and report the FaceSketchID System’s performance when matching the composite to photographs from [42].

We also evaluate the performance of the FaceSketchID System on a set of 21 surveillance composites created while viewing videos from the YouTube Faces (YTF) [43] database for [28]. In [28], the mated photographs used for these composites are from the Labeled Faces in the Wild (LFW) [44] database. To best replicate a mugshot matching scenario, we have instead used mated photographs found on the Internet that have a more “mugshot-like” appearance. For these composites, the performance of the FaceSketchID System when matching surveillance composites to photographs is compared against performance of COTS-1 and COTS-2⁶ when matching video frames to photographs. For the COTS systems, an overall similarity measure between a video and a photograph is computed by fusing the scores from individual frames via the sum rule (as reported in [28]). Examples of these composites and their mated photographs can be found in Fig. 5.

⁶We are unable to perform additional experiments using COTS-3 since our license has expired and we cannot renew the license.

D. Viewed Software-Generated Composites

We include results when matching (viewed software-generated composite, photograph) pairs from the PRIP-VSGC database to compare against previously published results. As noted earlier, the PRIP-VSGC database contains 123 pairs from the AR database, with three composites available per pair. Composites generated using FACES by both the American and Asian operators as well as the composites generated using Identi-Kit are used to evaluate the performance of both algorithms. Both the holistic and component-based algorithms are trained on the CUHK-VHDC database for this experiment. We note that a subset of the CUHK-VHDC database contains pairs from the AR database which are not included since they would bias the matching performance if used during algorithm training. We do not include the performance of COTS systems in this experiment as they can be found in [17]. A gallery of 10,000 mugshots from the PCSO is used to extend the gallery in these experiments to be consistent with that in [17]. While we have age range, gender and race information to filter the gallery set, only gender-based demographic filtering was reported in [17]. Thus, to provide a fair comparison with [17], all retrieval rates for the (viewed software-generated composite, photograph) experiments do not involve any demographic filtering.

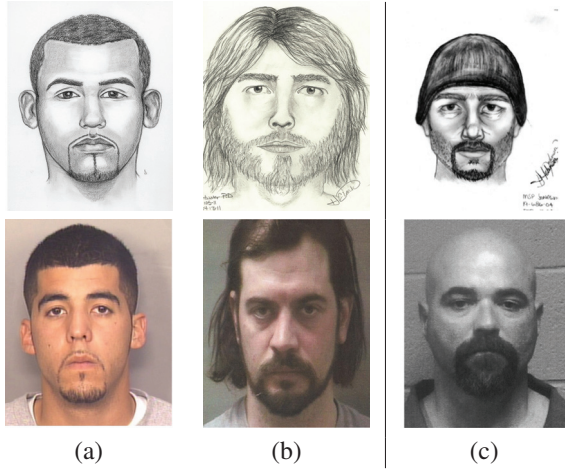


Fig. 11. Examples of successful Rank-1 matches of (hand-drawn composite, mugshot) pairs (a,b). A failure case (c) shows a relatively accurate composite which was returned at a rank higher than 200.

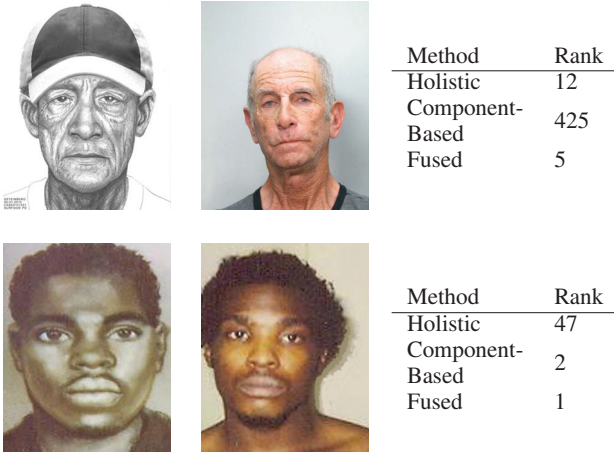


Fig. 12. Examples of retrieval rank improvement when matching hand-drawn composites to mugshots after match score fusion of the holistic and component-based algorithms.

VII. EXPERIMENTAL RESULTS

A. Hand-Drawn Composites

Figure 10 shows rank retrieval results of the experiments listed in Section VI-A. For the holistic algorithm (Fig. 10 (a)), the choice of data used to train the algorithm has a noticeable effect on retrieval rate. Comparing Rank-100 and Rank-200 retrieval rates, the best training set (PRIP-HDC) differs from the worst training set (CUHK-VHDC) by approximately 4% and 3%, respectively. It is interesting to note that while the viewed hand-drawn composite combinations of training data exhibit similar performance, especially at lower ranks, training the holistic algorithm on the PRIP-HDC database alone is markedly better than including any other composite database. One explanation for this is that training on (hand-drawn composite, mugshot) pairs alone allows the subspaces to be tightly tuned to the (hand-drawn composite, mugshot) matching task. The introduction of viewed hand-drawn composites to the training set likely causes the subspaces to be biased

towards (viewed hand-drawn composite, photograph) matching (in which texture and structure are extremely accurate). Furthermore, the number of (viewed hand-drawn composite, photograph) pairs relative to (hand-drawn composite, mugshot) pairs when the training datasets are combined also biases the subspace towards the viewed hand-drawn composite matching scenario, resulting in performance degradation.

By contrast, the component-based algorithm (Fig. 10 (b)) exhibits more consistency between the best and worst training sets (19.2% versus 16.9% at Rank-200, respectively). However, retrieval results indicate that, for this algorithm, training on the PRIP-HDC database alone is the worst choice of the training sets. While the holistic algorithm likely benefits from a subspace that is highly tuned to the (hand-drawn composite, mugshot) matching scenario, the performance of the component-based algorithm is likely degraded by that same specificity to the problem. The reason for this is that, at the component level (e.g. only considering the mouth), there can be a significant amount of noise due to the misalignment of components. Thus, subspaces learned in this scenario are likely influenced by noise more than the structure and texture of the component itself. This explanation justifies the fact that training using a combination of viewed hand-drawn composite and hand-drawn composite data results in the best performance of the component-based algorithm when matching (hand-drawn composite, mugshot) pairs. That is, with the stability provided to the learned subspace by the viewed hand-drawn composites (when we consider comparing facial components between subjects), the component-based algorithm is able to match (hand-drawn composite, mugshot) pairs well.

We note that in [6], the difference in performance between the holistic and component-based algorithms was explained to be due to the different training methodologies. In this study, both algorithms use the same training methodology, additional explanations are needed to account for this difference. Primarily, the component-based method does not consider any spatial relationship between individual components, which is encoded in the holistic representation. Furthermore, it may simply be more difficult to match composites at low ranks based on the use of a small number of components used in the component-based algorithm. However, when these components are accurately described by a witness and accurately depicted by a forensic artist, the component-based method does well.

The best performance that the FaceSketchID System is able to achieve on (hand-drawn composite, mugshot) pairs results from fusing the holistic and component-based match scores (Fig. 10 (c)). Different weighting methods were investigated, but assigning equal weights to both similarity scores resulted in the best performance. The Rank-200 performance after match score fusion is approximately 4% better than that of the holistic algorithm and approximately 6% better than that of the component-based algorithm. We include COTS retrieval rates in Fig. 10 (c) as a baseline against our best achieved matching performance. Examples of successful and unsuccessful (hand-drawn composite, mugshot) match cases can be found in Fig. 11. Examples of (hand-drawn composite, mugshot) match cases in which match score fusion improved the performance can be found in Fig. 12.

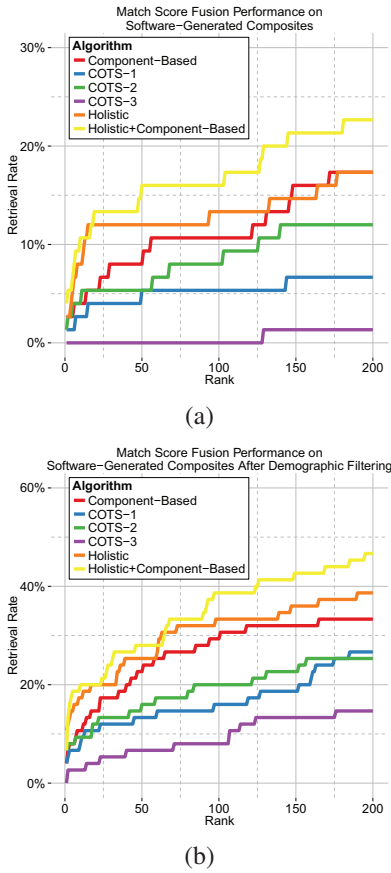


Fig. 13. Results for the holistic and component-based algorithms when matching software-generated composites to mugshots before (a) and after (b) demographic filtering. COTS matchers are included as a baseline.

Filtering the gallery based on demographic information (age, gender, and race) significantly improves the retrieval rates for both holistic and component-based algorithms as well as the three COTS matchers (Figs. 10 (d,e,f)). Again, the highest retrieval rate performance is achieved by the match score fusion of the holistic and component-based algorithms, with a Rank-200 retrieval rate of approximately 51%. Klum et al. explain in [6] the phenomenon that the performance of the component-based algorithm is boosted more by demographic filtering than that of the holistic algorithm. This is due primarily to the fact that before filtering, most components provide relatively poor recognition accuracy individually, so there is little benefit derived from their fusion. After filtering, the recognition accuracy of each component is significantly improved leading to a significantly higher fused component-based accuracy.

Figs. 16 (a), (b), and (c) show example retrievals from the sequestered set. For these 32 (hand-drawn composite, mugshot) pairs, the FaceSketchID System is able to match two subjects at Rank-1, nine by Rank-100 and 13 by Rank-200 after demographic filtering. Note that matching 13 out of 32 pairs by Rank-200 equates to a retrieval rate of 40.63%, which is comparable to the best results reported in the cross-validation experiments above.

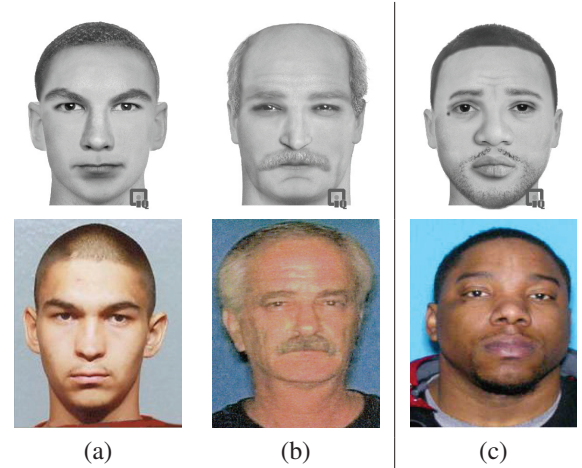


Fig. 14. Examples of successful Rank-1 matches of (software-generated composite, mugshot) pairs (a,b). A failure case (c) shows a relatively accurate composite which was returned at a rank higher than 200.

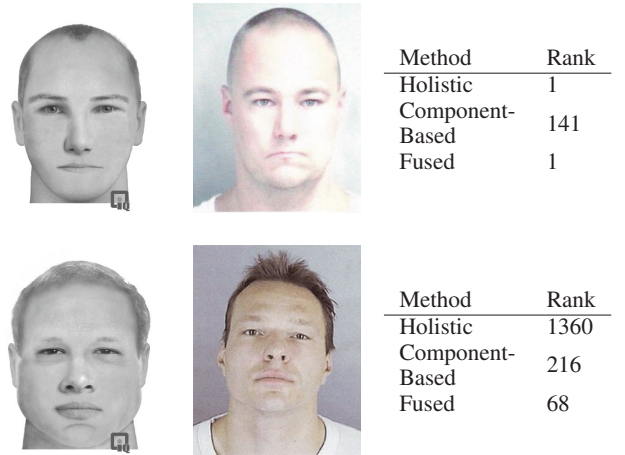


Fig. 15. Examples of retrieval rank improvement when matching software-generated composites to mugshots after match score fusion of the holistic and component-based algorithms.

Fig. 17 shows the composite and the photographs depicting Djokar Tsarnaev, one of the two Boston Bombing suspects, along with the rank at which they were retrieved (after demographic filtering). As noted in [42], these photographs would likely not be in a law enforcement agency's database. Nevertheless, the FaceSketchID System is able to retrieve multiple photographs at low ranks (correctly recognizing five of six photographs by Rank-50).

B. Software-Generated Composites

Figure 13 (a) shows the performance of the holistic and component-based algorithms in addition to the COTS matchers when matching (software-generated composite, mugshot) pairs. The holistic algorithm performs well at low ranks, achieving a 12% retrieval rate by Rank-25. As mentioned in Section VII-A, the holistic method benefits heavily from training data that mirrors the data used during testing. Because we do not have enough (software-generated composite, mugshot)

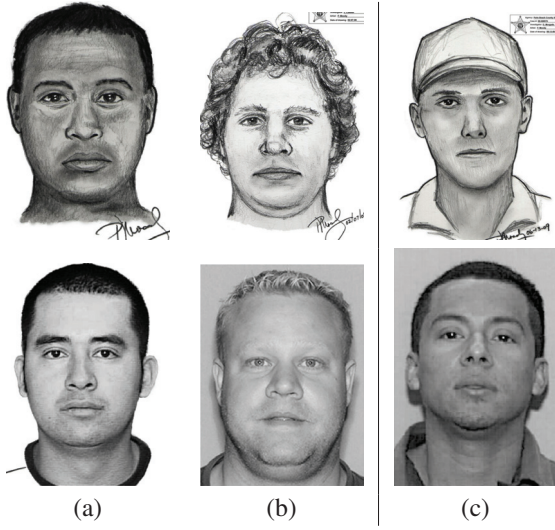


Fig. 16. Examples of successful matches of (hand-drawn composite, mugshot) pairs (a,b) from the sequestered dataset. A failure case (c) shows a relatively accurate composite which was returned at a rank higher than 200.



Fig. 17. A composite and photographs depicting Dzhokhar Tsarnaev. Retrieval ranks for the FaceSketchID System are listed below the corresponding photograph. All ranks are after demographic filtering (15-25 year old, white, male).

pairs, the only training data used were from the CUHK-VHDC databases. Thus, while we achieve a respectable retrieval rate of approximately 17% at Rank-200, we believe that the holistic algorithm's performance could be improved if operational software-generated composite data were available.

In contrast with the holistic algorithm, the component-based algorithm performs relatively poorly at lower ranks but matches the holistic algorithm's performance at higher ranks. This may be due to the fact that it is difficult to precisely recognize (i.e. match at Rank-1) an individual based purely on a set of components from the facial composite. Two scenarios contribute to this difficulty: (i) a particular component matches very highly with an impostor and/or, (ii) overall, the components in the composite are not sufficiently similar those in the mugshot. However, by Rank-50 we have likely overcome these inherent challenges with the component-based algorithm, and thus are able to achieve performance similar to that of the holistic algorithm.

As with the hand-drawn composites, score fusion of the two algorithms results in the best retrieval rates for the FaceSketchID System on (software-generated composite, mugshot) pairs. Fusion improves the Rank-200 performance of both

TABLE VII
TRUE ACCEPT RATES AT FALSE ACCEPT RATES FOR EVALUATED SYSTEMS

Matcher	Testing Database	TAR @ FAR =	
		0.1%	1.0%
FaceSketchID	PRIP-HDC	0.255 ± 0.020	0.565 ± 0.028
	PRIP-SGC	0.251	0.613
COTS-1	PRIP-HDC	0.122 ± 0.049	0.381 ± 0.047
	PRIP-SGC	0.129	0.473
COTS-2	PRIP-HDC	0.115 ± 0.028	0.491 ± 0.255
	PRIP-SGC	0.170	1.000
COTS-3	PRIP-HDC	0.064 ± 0.018	0.249 ± 0.050
	PRIP-SGC	0.069	0.232

True accept rates (TAR) at false accept rates (FAR) of 0.1% and 1% for the FaceSketchID System and three different COTS matchers after demographic filtering. Scores listed for the FaceSketchID System reflect using the training sets which provide the highest retrieval rates for both algorithms and fusing the match scores. Standard deviations of the 5-fold cross-validation are reported when testing on the PRIP-HDC database.

the holistic algorithm and the component-based algorithm by approximately 6%. Filtering with demographic information improves the performance of the holistic and composite-based algorithms (after match score fusion) when matching (software-generated composite, mugshot) pairs (Fig. 13 (b)) to 48% at Rank-200. Examples of successful and unsuccessful (software-generated composite, mugshot) match cases can be found in Fig. 14. Examples of (software-generated composite, mugshot) match cases in which score fusion improved the performance can be found in Fig. 15. True accept rates (TAR) at false accept rates (FAR) of 0.1% and 1.0% are reported in Table VII for both the PRIP-HDC and the PRIP-SGC databases to facilitate comparisons with published results.

It is worth noting that in many cases where facial composites were not successfully matched to their correct mugshot mate, the Rank-1 retrieval is more similar in appearance to the composite than the true mate. We view this not as a failure of the FaceSketchID System, but a result of the inherent difficulties in synthesizing an accurate composite which can be due to (i) inaccurate or rather vague description of the suspect provided by the witness, (ii) the age difference between the time the suspect's mugshot was captured and when he was seen by a witness/victim, and (iii) inexperience of the forensic artists or the limitations of the composite software. Examples of such (hand-drawn composite, mugshot) and (software-generated composite, mugshot) pairs are shown in Fig. 18.

C. Surveillance Composites

Table VIII shows the retrieval ranks of (surveillance composite, mugshot) pairs for the FaceSketchID System and (surveillance image, mugshot) pairs for the COTS matchers for the 4 subjects from the PCSO. Performance of the COTS matchers seems to vary significantly for these subjects. While there is variability in the performance of composites between artists, the most accurate composites achieve the lowest rank for 3 out of the 4 subjects using the FaceSketchID System.

Using the composites from [28], the FaceSketchID System

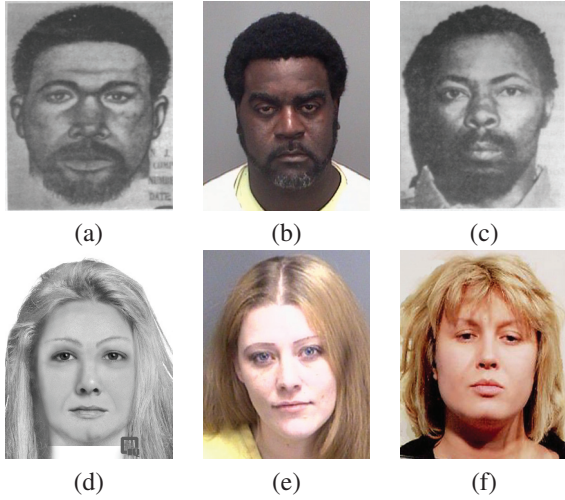


Fig. 18. Examples of a hand-drawn composite (a) and a software-generated composite (d) for which the Rank-1 match is an impostor (b,e) that is more similar in appearance to the composite than the genuine mugshot match (c,f).

TABLE VIII
RETRIEVAL RANKS FOR SURVEILLANCE IMAGERY FROM THE PCSO

Subject	FaceSketchID	COTS-1	COTS-2	COTS-3
1	124	3,007	300	95,962
2	1	1	3	5,131
3	89	72	FTE	367
4	253	1,122	497	733

Retrieval ranks for (surveillance composite, mugshot) obtained from the FaceSketchID System and (surveillance image, mugshot) obtained from the three COTS matchers. Ranks marked as FTE indicate the COTS matcher returned “failure to match”. All retrieval ranks reflect a gallery of 100,000 and do not involve any demographic filtering.

is able to achieve the best performance on 9 of the 21 subjects (Table IX). By comparison, COTS-1 and COTS-2 are able to achieve the best performance on 7 and 5 subjects, respectively. The FaceSketchID System also outperforms the COTS matchers in terms of rank retrieval performance, correctly matching 8 composites by Rank-200 (versus COTS-1 and COTS-2 retrieving 6 and 5 by Rank-200, respectively). However, there is no significant advantage in using surveillance composites to perform recognition with the FaceSketchID System versus using surveillance video with the COTS matchers as the YTF videos are, in general, of adequate quality. Indeed, for all of the YTF videos, the Viola-Jones face detector [45] is able to successfully find a face in at least 48 consecutive frames. Thus, in forensic investigations in which the investigator has access to surveillance data of the same quality as the YTF videos, our experimental results indicate it may be valuable to use both the surveillance composite and the surveillance video. However, cases with highly off-pose, occluded, or blurred faces (like those in Fig. 2) present challenges to state-of-the-art unconstrained face recognition systems.

Fig. 19 shows a surveillance composite of Tamerlan Tsarnaev which was created using a poor quality surveillance frame. As with Djokar Tsarnaev, the photographs shown of Tamerlan would not be in the mugshot database. However, the FaceSketchID System is able to match Tamerlan’s composite

TABLE IX
RETRIEVAL RANKS FOR SURVEILLANCE IMAGERY FROM [28]

Subject	FaceSketchID	COTS-1	COTS-2
Adrien Brody	1480	4743	FTM
Ahmet Necdet Sezer	850	1209	FTD
Akbar Hashemi Rafsanjani	30	7	FTM
Akhmed Zakayev	6599	3031	3107
Alastair Campbell	5790	2081	FTD
Aleksander Kwasniewski	893	964	5
Alex Zanardi	1873	4301	FTD
Alexandre Despatie	4022	500	FTD
Alvaro Uribe	2496	3519	FTD
Amanda Marsh	3538	FTD	FTM
Angelo Reyes	55	83	3
Carlos Iturgaitz	50	6918	21
Chuck Amato	58	1374	FTD
Conchita Martinez	3018	181	1131
Erskine Bowles	1911	1676	FTD
Filip De Winter	1570	4868	FTD
Jacques Chirac	424	69	32
Luis Ernesto Derbez Bautista	4	135	FTD
Marie-Josée Croze	169	2578	497
Na Na Keum	52	15	FTD
Thomas Ferguson	46	1235	43

Retrieval ranks for the FaceSketchID System when matching (surveillance composite, photograph) pairs and COTS systems after demographic filtering. The FaceSketchID System is able to perform the best on 8 of the total 21 subjects whereas COTS-1 and COTS-2 perform the best on 7 and 5 subjects, respectively. Ranks marked as FTD indicate that facial landmark detection failed on either the query or target images. Ranks marked as FTM indicate the COTS matcher returned “failure to match”.

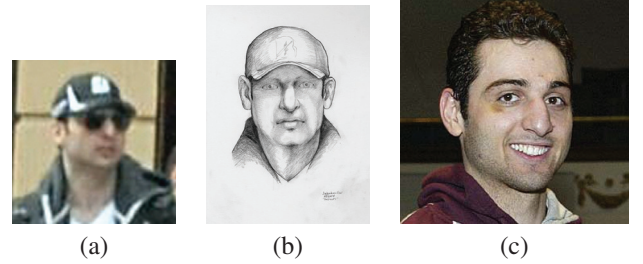


Fig. 19. A surveillance frame (a) of Tamerlan Tsarnaev was used to create a high-quality surveillance composite (drawn by Jane Wankmiller [28]) (b) to improve performance when matching against a photograph (c). After demographic filtering (20-30 year old, white, male), the FaceSketchID System is able to retrieve the photograph by Rank-20.

to his photograph at Rank-2113 and Rank-20 before and after demographic filtering, respectively. This result is comparable to the performance achieved using the most accurate COTS system when matching the best (surveillance frame, photograph) pair reported in [42].

D. Viewed Software-Generated Composites

Han et al. [17] report Rank-1, Rank-100, and Rank-200 retrieval rates of 10.6%, 65%, and 73.2% when matching viewed software-generated composites (created by an American operator) to photographs, respectively. The FaceSketchID System is able to achieve Rank-1, Rank-100, and Rank-200 retrieval rates of approximately 24%, 79%, 87% on the same dataset after fusing the match scores of the holistic and component-based algorithms (Fig. 20). Han et al. also report the matching performance when FACES composites are created by an Asian user. Similar to the American-created

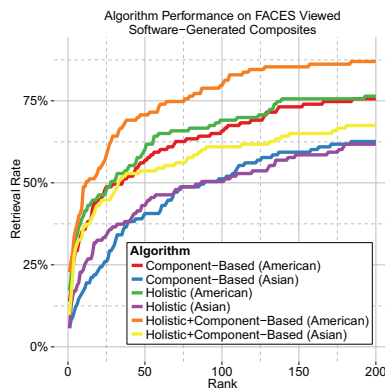


Fig. 20. Results for the holistic, and component-based algorithms on viewed software-generated composites created using FACES.

composites, the FaceSketchID System is able to achieve consistently higher retrieval rates, with an improvement at Rank-200 of approximately 14%. When matching software-generated composites created using Identi-Kit to mugshots using the FaceSketchID System, we observe a similar performance degradation compared with composites created using FACES as was reported in [17]. For the composites created with Identi-Kit, match score fusion does not improve the performance of the FaceSketchID System because of the poor performance of the holistic algorithm. This can be explained by the fact that the Identi-Kit composites have little information other than outlines of facial components. However, the component-based algorithm still improves upon the highest Rank-200 retrieval rate reported in [17] by 16%.

VIII. SUMMARY

Facial composites drawn by forensic artists (hand-drawn composites) or created using software (software-generated composites) are routinely used by law enforcement agencies to assist in identification and apprehension of suspects involved in criminal activities, especially when no photograph of the suspect at the crime scene is available. We have developed the FaceSketchID System to provide law enforcement agencies a tool with which to more effectively make use of this facial composite data. The FaceSketchID System combines the strengths of two different representation and matching algorithms (holistic and component-based) to achieve state-of-the-art accuracies for both (hand-drawn composite, mugshot) and (software-generated composite, mugshot) pairs. We investigate the effect of training our two algorithms on different databases to achieve the best matching performance. We also show the performance of surveillance composites, which are used for poor quality surveillance images where COTS systems fail. This scenario is likely to become an important application of facial composites to mugshot matching given the growing number of surveillance cameras around the globe. Three state-of-the-art commercial matchers were used as baselines for our experiments. Filtering of the large gallery (100,000 mugshots) based on demographic information showed a significant improvement in retrieval accuracy. Future work will involve

acquiring more surveillance composites, with an emphasis on those with severe off-pose or occluded faces. We will also develop matching algorithms that can directly use the verbal description (from a witness) in addition to the facial composite to improve the matching performance of the FaceSketchID System.

ACKNOWLEDGMENT

The authors would like to thank the team at the Pinellas County Sheriff's Office for their support of this research. We would also like to thank Lois Gibson, Karen Taylor, the forensic artists from the MSP, and the forensic artists that collaborate with the PCSO for providing facial sketch and mated mugshot pairs. This research was supported by the National Institute of Justice (NIJ) grant 2011-IJ-CX-K057.

REFERENCES

- [1] A. Jain, B. Klare, and U. Park, "Face matching and retrieval in forensics applications," *IEEE Multimedia*, vol. 19, no. 1, pp. 20–28, 2012.
- [2] K. Taylor, *Forensic Art and Illustration*. CRC Press, 2001.
- [3] FACES 4.0, IQ Biometrics, <http://www.iqbiometrix.com>.
- [4] I. Lovett, "Los angeles officials identify video assault suspects," *New York Times*, January 2011.
- [5] D. McQuiston, L. Topp, and R. Malpass, "Use of facial composite systems in US law enforcement agencies," *Psychology, Crime and Law*, vol. 12, no. 5, pp. 505–517, 2006.
- [6] S. Klum, B. Klare, H. Han, and A. Jain, "Sketch based face recognition: Forensic vs. composite sketches," in *Proc. IEEE ICB*, 2013.
- [7] X. Tang and X. Wang, "Face sketch recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 50–57, January 2004.
- [8] Q. Liu, X. Tang, H. Jin, H. Lu, and S. Ma, "A nonlinear approach for face sketch synthesis and recognition," in *Proc. IEEE CVPR*, June 2005, pp. 1005–1010.
- [9] X. Gao, J. Zhong, and C. Tian, "Face sketch synthesis algorithm based on e-hmm and selective ensemble," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 4, pp. 487–496, April 2008.
- [10] X. Wang and X. Tang, "Face photo-sketch synthesis and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 11, pp. 1955–1967, November 2009.
- [11] D. Lin and X. Tang, "Inter-modality face recognition," in *Proc. ECCV*, 2006, pp. 13–26.
- [12] Y. Zhang, C. McCullough, J. Sullins, and C. Ross, "Hand-drawn face sketch recognition by humans and a pca-based algorithm for forensic applications," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 40, no. 3, pp. 475–485, May 2010.
- [13] R. Uhl and N. Lobo, "A framework for recognizing a facial image from a police sketch," in *Proc. IEEE CVPR*, June 1996, pp. 586–593.
- [14] B. Klare, Z. Li, and A. Jain, "Matching forensic sketches to mug shot photos," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 639–646, March 2011.
- [15] H. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa, "Memetically optimized MCWLD for matching sketches with digital face images," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 5, pp. 1522–1535, October 2012.
- [16] P. Yuen and C. Man, "Human face image searching system using sketches," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 37, no. 4, pp. 493–504, July 2007.
- [17] H. Han, B. Klare, K. Bonnen, and A. Jain, "Matching composite sketches to face photos: A component based approach," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 1, pp. 191–204, January 2013.
- [18] B. Klare and A. Jain, "Heterogeneous face recognition using kernel prototype similarities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1410–1422, June 2013.
- [19] S. Li and A. Jain (eds.), *Handbook of Face Recognition*, 2nd ed. Springer, 2011.
- [20] L. Gibson, *Forensic Art Essentials*. Elsevier, 2008.
- [21] K. Bonnen, B. Klare, and A. K. Jain, "Component-based representation in automated face recognition," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 1, pp. 239–253, January 2012.
- [22] E-FIT, VisionMetric Ltd., <http://www.vision-metric.com>.

- [23] C. Frowd, P. Hancock, and D. Carson, "EvoFIT: A holistic, evolutionary facial imaging technique for creating composites," *ACM Trans. Appl. Percept.*, vol. 1, no. 1, pp. 19–39, July 2004.
- [24] Identi-Kit, Identi-Kit Solutions, <http://www.identikit.net/>.
- [25] G. Wells and L. Hasel, "Facial composite production by eyewitnesses," *Current Directions in Psychological Science*, vol. 16, no. 1, pp. 6–10, February 2007.
- [26] C. Frowd, D. Carson, H. Ness, D. McQuiston, J. Richardson, H. Baldwin, and P. Hancock, "Contemporary composite techniques: The impact of a forensically-relevant target delay," *Legal and Criminological Psychology*, vol. 10, no. 1, pp. 63–81, February 2005.
- [27] R. Geiselman, R. Fisher, D. MacKinnon, and H. Holland, "Eyewitness memory enhancement with the cognitive interview," *American Journal of Psychology*, vol. 99, pp. 385–401, 1986.
- [28] L. Best-Rowden, H. Han, C. Otto, B. Klare, and A. Jain, "Unconstrained face recognition: Identifying a person of interest from a media collection," *IEEE Trans. Inf. Forensics Security*, 2014, To appear.
- [29] A. M. Martinez and R. Benavente, "The AR face database," *CVC Technical Report*, vol. 24, June 1998.
- [30] E. Meyers and L. Wolf, "Using biologically inspired features for face processing," *Int. J. Comput. Vision*, vol. 76, no. 1, pp. 93–104, 2008.
- [31] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, January 2004.
- [32] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution grayscale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, August 2002.
- [33] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, "XM2VTSDB: The Extended of M2VTS Database," in *Proc. of International Conference on Audio- and Video-Based Person Authentication*, vol. 24, June 1999, pp. 72–77.
- [34] W. Zhang, X. Wang, and X. Tang, "Coupled information-theoretic encoding for face photo-sketch recognition," in *Proc. IEEE CVPR*, 2011.
- [35] P. J. Phillips, H. M. P. J. Rauss, and S. Rizvi, "The feret evaluation methodology for face recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, 2000.
- [36] Itseez, "The OpenCV Library," www.opencv.org, 2000.
- [37] G. Guennebaud, B. Jacob *et al.*, "Eigen v3," <http://eigen.tuxfamily.org>, 2010.
- [38] Digia, "Qt5," <http://qt-project.org/>, 2013.
- [39] J. Klontz, B. Klare, S. Klum, A. Jain, and M. Burge, "Open source biometric recognition," in *Proc. IEEE BTAS*, 2013.
- [40] S. Milborrow and F. Nicolls, "Locating facial features with an extended active shape model," in *Proc. ECCV*, 2008.
- [41] X. Wang and X. Tang, "Random sampling for subspace face recognition," *Int. J. Comput. Vision*, vol. 70, no. 1, pp. 91–104, 2006.
- [42] J. Klontz and A. Jain, "A case study on unconstrained facial recognition using the boston marathon bombings suspects," *MSU Tech. Report*, 2013.
- [43] L. Wolf, T. Hassner, and I. Maoz, "Face recognition in unconstrained videos with matched background similarity," in *Proc. IEEE CVPR*, 2011, p. 529534.
- [44] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," *Univ. of Massachusetts, Amherst Technical Report*, 2007.
- [45] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vision*, pp. 137–154, 2004.



Scott Klum is a graduate student in the Department of Computer Science and Engineering at Michigan State University. In 2012, he received his B.S. degree in Computer Science from the Department of Computer Science and Engineering at Michigan State University. His research interests include biometrics, computer vision, and software engineering. He is a student member of the IEEE.



Hu Han is a research associate in the Department of Computer Science and Engineering at Michigan State University, East Lansing. He received the B.S. degree in Computer Science from the Shandong University, Jinan, China, in 2005 and the Ph.D. degree from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2011. His research interests include computer vision, pattern recognition, and image processing, with applications to biometrics, forensics, law enforcement, and security systems. He is a member of the IEEE.



Brendan Klare received the B.S. and M.S. degrees in computer science from the University of South Florida in 2007 and 2008, respectively, and the Ph.D. degree in computer science from Michigan State University in 2012. He is a lead scientist at Noblis. From 2001 to 2005, he served as an airborne ranger infantryman in the 75th Ranger Regiment. His research interests include pattern recognition, image processing, and computer vision. He has authored several papers on the topic of face recognition, and he received the Honeywell Best Student Paper

Award at the 2010 IEEE Conference on Biometrics: Theory, Applications, and Systems (BTAS). He is a member of the IEEE.



Anil K. Jain is a University Distinguished Professor in the Department of Computer Science and Engineering at Michigan State University, East Lansing. His research interests include pattern recognition and biometric authentication. He served as the editor-in-chief of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (1991-1994). He is the coauthor of a number of books, including Handbook of Fingerprint Recognition (2009), Handbook of Biometrics (2007), Handbook of Multibiometrics (2006), Handbook of Face Recognition (2011), BIOMETRICS: Personal Identification in Networked Society (1999), and Algorithms for Clustering Data (1988). He served as a member of the Defense Science Board and The National Academies committees on Whither Biometrics and Improvised Explosive Devices. He received the 1996 IEEE TRANSACTIONS ON NEURAL NETWORKS Outstanding Paper Award and the Pattern Recognition Society best paper awards in 1987, 1991, and 2005. He has received Fulbright, Guggenheim, Alexander von Humboldt, IEEE Computer Society Technical Achievement, IEEE Wallace McDowell, ICDM Research Contributions, and IAPR King-Sun Fu awards. He is a fellow of the AAAS, ACM, IAPR, SPIE, and IEEE.