FACESKETCHID: A SYSTEM FOR FACIAL SKETCH TO MUGSHOT MATCHING

by

Scott Jeffrey Klum

A THESIS

Submitted
to Michigan State University
in partial fulfillment of the requirements
for the degree of

Computer Science—Master of Science

2014

ABSTRACT

FACESKETCHID: A SYSTEM FOR FACIAL SKETCH TO MUGSHOT MATCHING

by

Scott Jeffrey Klum

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 legacy method of identifying suspects is slow, tedious, and may not even lead to the timely apprehension of the suspect. Hence, there is a need for a method that can automatically and efficiently match facial composites to large police mugshot databases. As a result 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, to our knowledge, no facial composite to mugshot matching system has been documented that is readily deployable. The contributions of this thesis 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 complementary algorithms (holistic and component-based), and (iii) a study of the effects of training data on algorithm performance. Experimental results are presented 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.

For my family, whose support is the reason these pages are filled.

ACKNOWLEDGEMENTS

I would first like to acknowledge and thank my advisor, Dr. Anil Jain. I've felt an overwhelming sense of personal growth while writing this thesis, nearly all of which I owe to Dr. Jain's exceptional guidance. I am forever indebted to him for putting up with me during the difficult times of my graduate degree, and helping me thoroughly enjoy the good times. I can confidently say that Dr. Jain has positively influenced not only my intellectual development, but my life as a whole, as much as anyone that I have known.

My thesis committee, which includes Dr. Xiaoming Liu and Dr. Arun Ross in addition to Dr. Jain, have been tremendously helpful as I've worked towards my Master's. The greatest impact both Dr. Liu and Dr. Ross have had on me is simply through example; both are model computer scientists and researchers. If I can hold myself to the standards that they have set for themselves, I will undoubtedly be successful.

Next, I would like to acknowledge my friend and personal mentor, Dr. Brendan Klare. Brendan's steadfast positivity and enthusiasm have been an indescribable boon to what I've achieved while working towards my Master's. Engineers like Brendan, who are not only incredibly knowledgeable and creative but also well-spoken and personable, are one of the biggest reasons I love the discipline I've chosen.

I'd also like to thank my family for their unwavering support. I am lucky to have a sister, Adrienne, who has been a constant source of encouragement (and plenty of laughs) throughout my time in college. My mother and father, Cindy and Jeff, are the greatest parents a young man could ask for. My father has shown how selfless he is during the course of my Master's, fulfilling any request without the slightest hesitation. My mother, as she always has been, was a source of calm during my Master's work; knowing I have her unconditional love has been indescribably reassuring. I have a wonderful extended family who has also been a source of joy these past two years. My Grandma Ellie and Grandpa Bill merit special recognition as, without their support, it would have been much more difficult to attend college.

I have worked with a special group of people who have contributed in some way to my Master's, and undoubtedly, I'll miss acknowledging someone in the following list: Josh Klontz, Charles Otto, Serhat Bucak, Soweon Yoon, Sunpreet Arora, Hu Han, Lacey Best-Rowden, Radha Chitta, Scott McCallum and Jake Roberto. Josh and Charles deserve a special mention, as both have been incredible friends and colleagues and have helped me become a better computer scientist. Sunpreet has also been a sounding board for me, not to mention a good friend. My roommates, Akos Nagy and Nick Skaff, have become lifelong friends and I am thankful to have met them during my Master's.

Finally, I want to thank my best friend Kristin Houck. The English language lacks the descriptive power to detail what Kristin has meant and means to me; trying to describe it would only cheapen things. I will just say:

I love you Kristin. Thank you for everything.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

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 eyewitness' or victim's descriptions are the only form of evidence available [13]. 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.

1.1 Background

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 [14]. Examples of high profile cases in which a hand-drawn composite was used are shown in Figs. 1.1 (a,b,c).

(ii) *Software-generated composites*: Facial composites created using software kits which allow an operator to select various facial components (Figs. 1.1 (d,e,f)). Software-generated composites have become a popular and more affordable al-

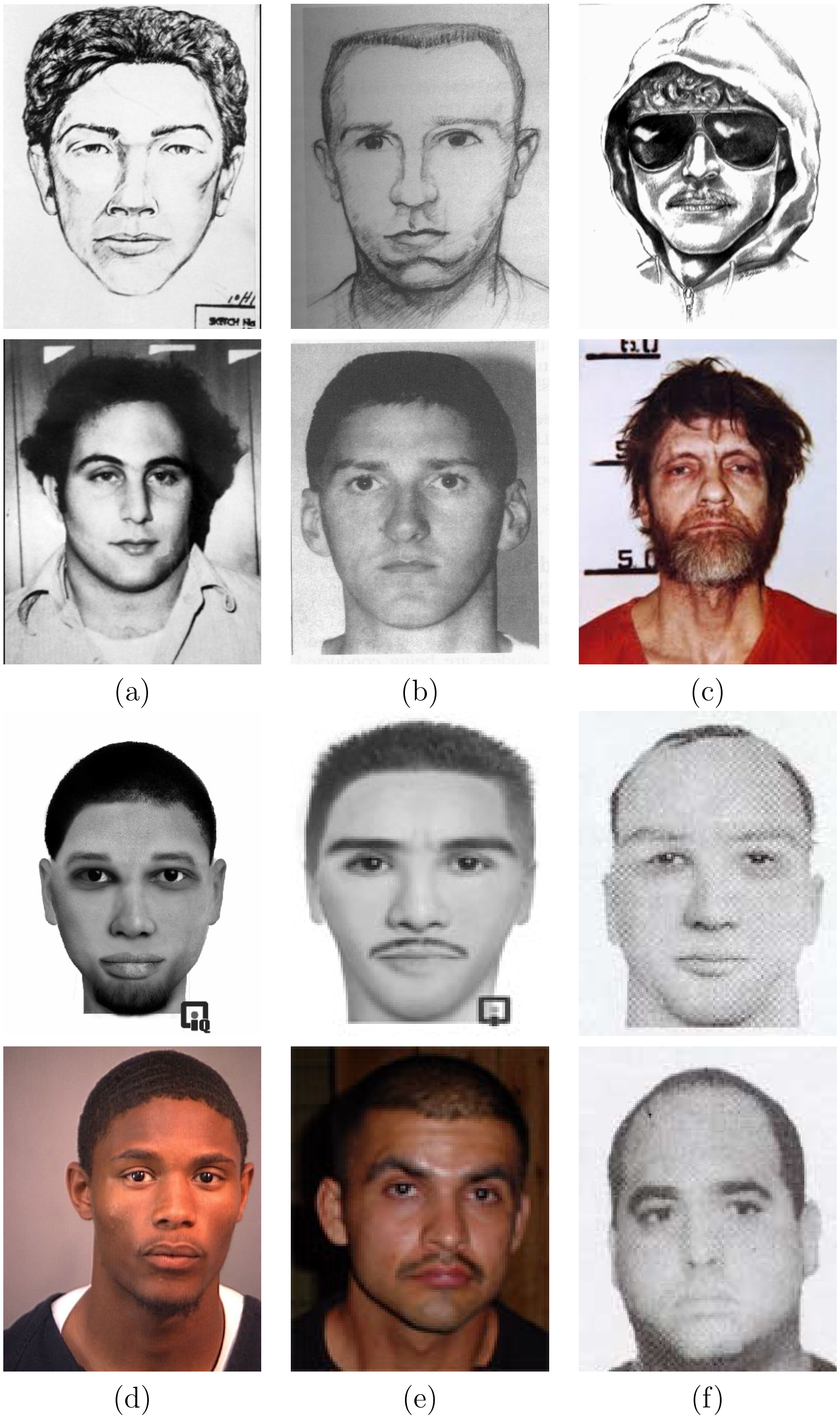


Figure 1.1: Examples of facial composites that were used in cases in which the suspect was successfully apprehended. Examples of hand-drawn composites and their mugshot mates are shown for David Berkowitz (Son of Sam) (a) [1], Timothy McVeigh (the Oklahoma City bomber) (b) [2], and Ted Kaczynski (the Unabomber) (c) [3]. Software-generated composites (d,e,f) that were created using the software FACES [4] are shown with mated mugshots.

ternative to hand-drawn composites. According to [14], 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 [14].

(iii) *Surveillance composites*: Facial composites drawn by forensic artists based on poor quality surveillance images. Surveillance composites are used in scenarios when commercial-off-the-shelf (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 packages rely 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. We note that in [15], we referred to hand-drawn composites as "forensic sketches" and software-generated composites as "composite sketches". The naming conventions have been corrected in this thesis to reflect the prevailing law enforcement terminology.

While several methods that match viewed[1] and hand-drawn composites to mugshots have been reported in the literature [16, 17, 18, 19, 20, 21, 22, 23], only a few methods have been published for automatic matching of software-generated composites to mugshots [19, 12]. In all the previous studies reported on software-generated composites, with the exception of [15], composites were created while the operator was viewing the high quality mugshot.

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

This type of viewed composite does not accurately reflect the creation of composites used 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 when viewing images of the suspect, these images are of poor quality compared with mugshots. To the best of our knowledge, no studies have reported performance when matching surveillance composites to mugshots. An extended review of facial composite to mugshot matching literature is presented in Chapter 2. Chapter 3 describes the construction process for each type of facial composite used in this thesis.

To the best of our knowledge, no matcher 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, this thesis presents 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 Chapter 4. The FaceSketchID System uses two complementary algorithms when matching facial composites to mugshots: (i) a holistic algorithm and (ii) a component-based algorithm, both of which are described in Chapter 4. It is important to point out that the matching performance of the FaceSketchID System critically depends on the accuracy of the composite (in terms of its resemblance to the suspect's face) 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.

Chapter 5 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 Chapter 6. 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 the Face Vendor Recognition Test (FRVT)[2]. We also investigate the influence of the type of data used to train the algorithms leveraged by the FaceSketchID System. To facilitate comparisons with previously published results, we also detail the FaceSketchID System's performance when matching viewed software-generated composites to photographs.

1.2 Contributions

The primary contributions of this thesis are:

(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 complementary algorithms (holistic and component-based).

(iii) A study of the effects of training data on algorithm performance.

[2] http://www.nist.gov/itl/iad/ig/frvt-home.cfm

CHAPTER 2

RELATED WORK

Automated face matching between two facial photographs is a well studied problem in computer vision and biometrics [24]. However, matching facial composites to photographs is a more challenging problem with only a limited amount of published work, some of which include: [16, 17, 18, 19, 20, 22, 21, 23, 12, 15]. 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 [14].

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 [19], but it required user input in the form of relevance feedback in the matching or recognition phase. Further, the authors in [19] used a small gallery in their experiments (300 facial photographs). The method proposed by Han et al. [12], 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

Table 2.1: Prior work on facial composite to photograph matching.

	Publication	Approach	Limitations
Viewed Hand-Drawn	Tang and Wang [16]	Photograph-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. [17]	Photograph-to-composite conversion using locally linear embedding	
	Gao et al. [25]	Photograph-to-composite conversion using embedded hidden Markov model	
	Wang and Tang [20]	Photograph-to-composite conversion using multiscale Markov random field model	
	Lin and Tang [18]	Common discriminant feature extraction	
	Zhang et al. [26]	Principle Component Analysis (PCA) based algorithm	
Hand-Drawn	Uhl and Lobo [27]	Photometric standardization	Software-generated composites, which are widely used in law enforcement, were not considered.
	Klare and Jain [22]	SIFT and MBLP feature descriptors with local-feature based discriminant analysis	
	Bhatt et al. [21]	Multi-scale circular Weber's local descriptor	
Software-Generated	Yuen and Man [19]	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. [12]	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.

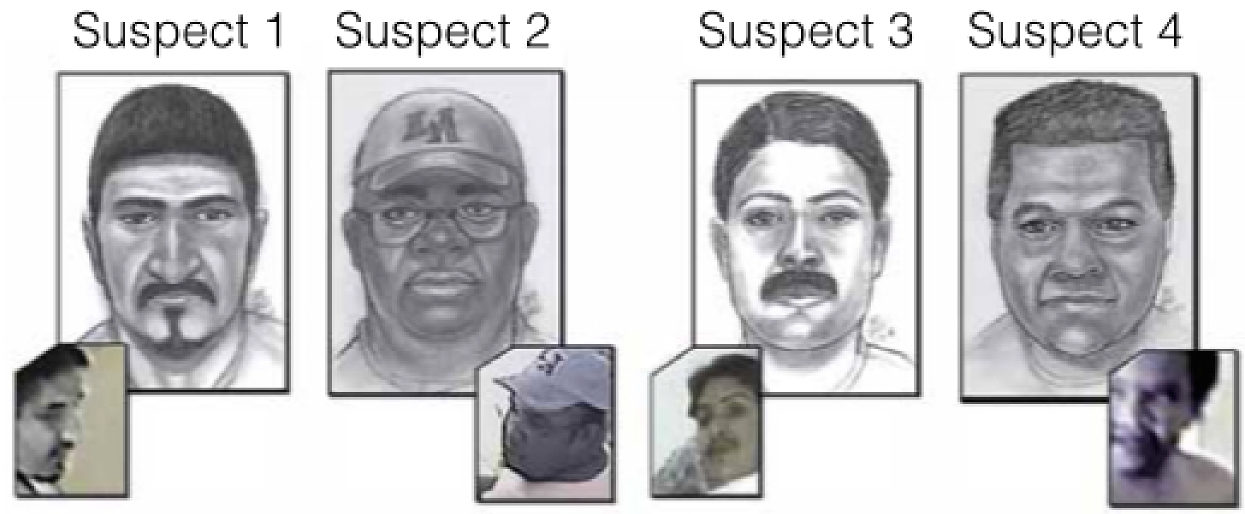


Figure 2.1: Examples of surveillance images that are of sufficiently poor quality such that COTS matchers are expected to fail to find the true mate in a mugshot database. Surveillance composites can be drawn based on these images, which can be used to identify a suspect more accurately. Surveillance composites shown are drawn by Sandra Enslow [5].

the match scores of two different algorithms (holistic and component-based) are fused. We also detail the use of surveillance composites, which have not previously been reported in the literature (Fig. 2.1). Additional examples of surveillance composites can be found in Figs. 3.2 and 6.13. 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 2.1.

Chapter 3 describes the process of creating the types of facial composites used in this thesis. We also detail the databases used to evaluate the FaceSketchID System.

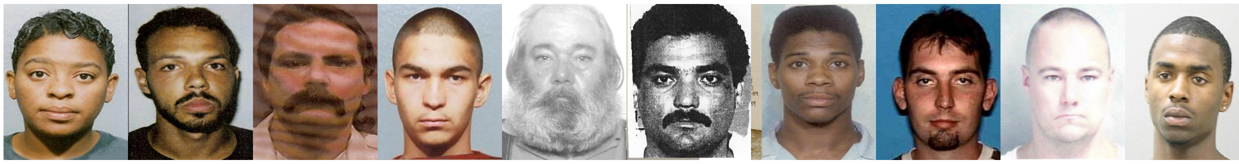
CHAPTER 3

CONSTRUCTING A FACIAL COMPOSITE

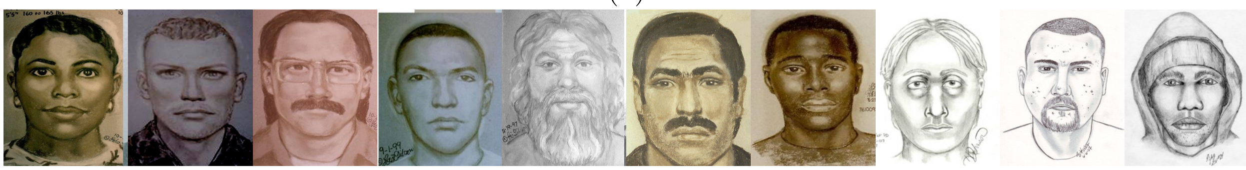
As previously mentioned, law enforcement agencies rely on three modalities of facial composites: (i) hand-drawn composites, (ii) software-generated composites and (iii) surveillance composites. Hand-drawn composites (Section 3.1) are drawn based on a verbal description. Typically, hand-drawn composites are drawn by a forensic artist with special training. Similarly, software-generated composites (Section 3.2) are drawn based on a verbal description, but are created using menu-driven software. In most composite software packages, 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 are expected to fail on the original face images, are described in Section 3.3. Fig. 3.1 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.

3.1 Hand-Drawn Composites

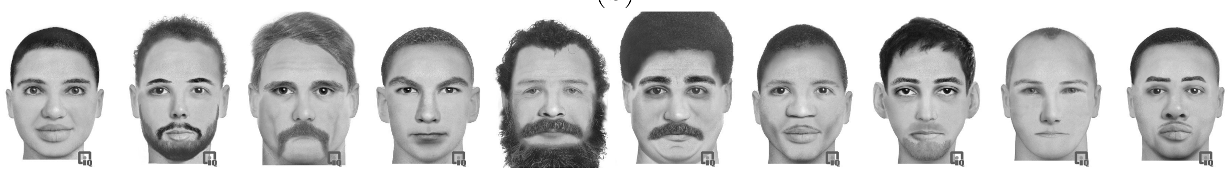
All hand-drawn composites used in our study were created by forensic artists for real-world criminal investigations. To create a hand-drawn composite, an artist draws a face based on descriptions provided by either one or multiple eyewitnesses. For this type of composite, the time between observation 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 [6], 43 were drawn by Karen Taylor [2], 56 were provided by



(a)



(b)



(c)

Figure 3.1: Examples of mugshots (a) and mated hand-drawn composites (b) and software-generated composites (c) created using FACES [4]. The hand-drawn composites shown were drawn by either Lois Gibson [6] or forensic artists at the Michigan State Police (MSP).

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.

3.2 Software-Generated Composites

A number of software systems are available to create composites: E-FIT [28], EvoFit [29], FACES [4], Identi-Kit [30], Mac-a-Mug [31], and Photo-Fit [31]. Of these, Identi-Kit and FACES are most widely used by law enforcement agencies in the United States [14]. Both Identi-Kit and FACES allow users to choose from a set of candidate components or features (e.g. 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 [12]. 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 [32]. 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 (the author of this thesis) 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 [33] to enhance the volunteer's memory of the suspect's facial features in the mugshot. To reduce the problem of operator contamination [32], 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 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.

3.3 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 older-generation 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. However, in most cases the surveillance imagery is of extremely poor quality and the COTS face matchers fail to find the corresponding individual within a mugshot gallery. To make use of these poor quality face images, 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

(a) (b) (c)

Figure 3.2: 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.

upon the matching performance of COTS matchers using a set of (surveillance composite, mugshot) pairs provided to us by the PCSO (Fig. 3.2). We note that while the surveillance images used in this thesis are of relatively high quality, we introduce the use of surveillance composites as a proof-of-concept to show that they can be used successfully. A more realistic use case of surveillance composites would involve the composites depicted in Fig. 2.1, but we do not have the mated mugshots for these composites.

3.4 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)[3]. The CUHK Face Sketch database [20] contains 188 pairs from the CUHK student database, 123 pairs from the AR database [7], 295 pairs from the XM2VTS database [34], and 1,194 pairs [35] from the FERET database [36]. We will refer to this set of 1,800 (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

[3] `http://mmlab.ie.cuhk.edu.hk/facesketch.html`

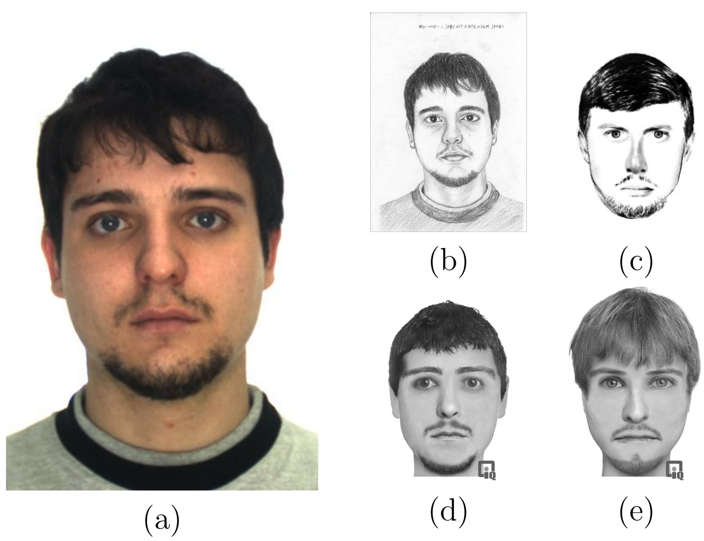


Figure 3.3: For a given photograph (a) from the AR database [7], the viewed facial composites used in our study consist of a hand-drawn composite (b), a composite created using Identi-Kit (c), and two composites created using FACES (d, e) by two different operators.

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 [12]. 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. 3.3.

Chapter 4 describes the FaceSketchID System in detail, including system specifications as well as the algorithms that are used in composite recognition.

CHAPTER 4

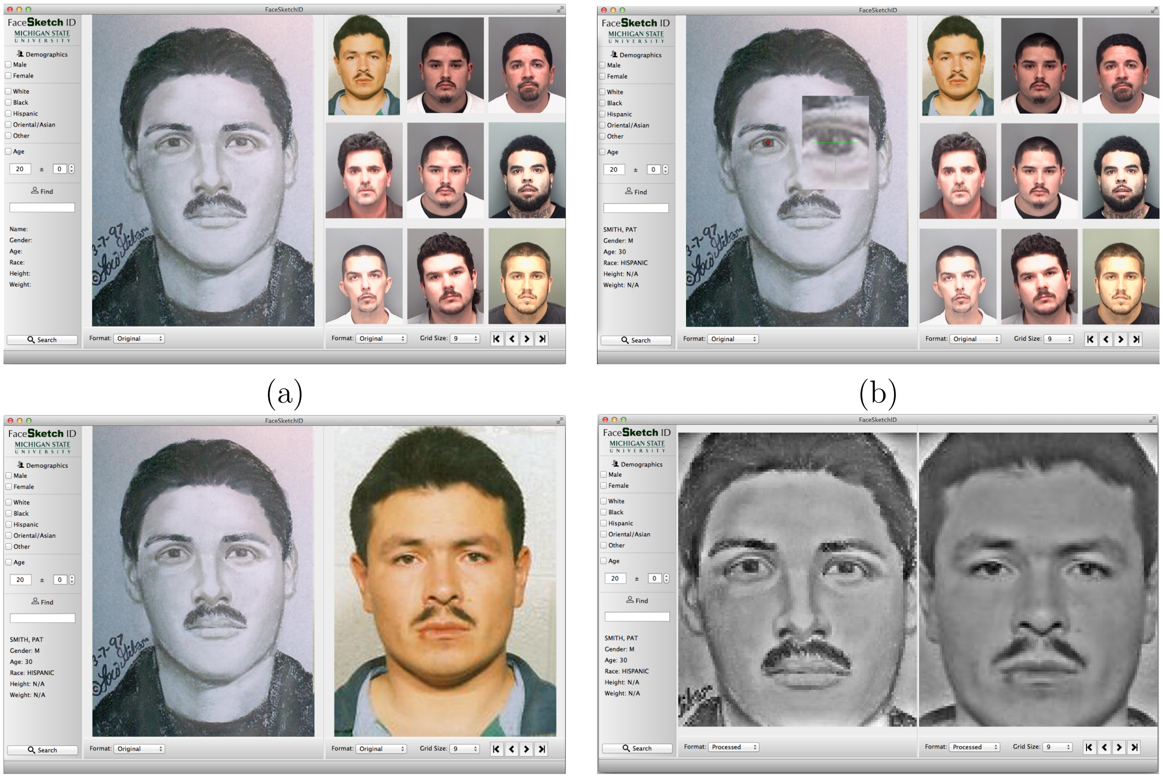
THE FACESKETCHID SYSTEM

The FaceSketchID System was developed to address the lack of a fully automatic and effective means to match facial composites to mugshots. System specifications for the FaceSketchID System are described in Section 4.1, while the matching algorithms used by the FaceSketchID System are described in Section 4.2.

4.1 System Specifications

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, 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 [37] as a matrix library, Eigen [38] for statistical learning, and Qt [39] for the GUI. Some modules of the FaceSketchID System are available in OpenBR [40]. 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 are approximately 5.73 KB in size. Specifically, templates contain feature vectors for the two matching algorithms in addition to demographic information (age, race, and gender) and detected eye locations. We note that the combination of a reasonably small template size in conjunction with rapid recognition capabilities should facilitate a large usability improvement over the traditional method of using composite data



(a) (b) (c) (d)

Figure 4.1: The FaceSketchID System graphical user interface (a). Options exist for manually modifying query eye locations (b), one-to-one comparison of query and target images (c), and viewing images after algorithmic normalization and preprocessing (d). The composite shown was drawn by Lois Gibson [6].

(e.g. dissemination to media outlets). A video demonstrating the matching process can be found at `http://biometrics.cse.msu.edu/images/ImgProjects/help_match.mp4`.

4.2 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 [23], is designed for use in heterogenous 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. [12], was originally designed for software-generated composite to mugshot matching. Modifications made to the original algorithms in [23] and [12] 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 [41]. Each algorithm is described in detail in the following sections. We note that both algorithms have been extensively tuned to perform well in (composite, mugshot) recognition.

4.2.1 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 [23]. 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 [16] and [25]. Thus, the algorithm for 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 [42] is then applied to the composite to compensate for the differences related to the change in modality between composite and mugshot. Subsequently, SIFT [8] features are extracted from a dense grid across the face. In parallel to the CSDN preprocessing and SIFT feature extraction pipeline, the normalized composite is processed with the Tan & Triggs pipeline [43] (consisting of a Gamma filter, a difference of Gaussian filter, and contrast equalization) and multi-scale local binary pattern (MLBP) [9] features are extracted. 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. Although [23] uses additional filters to improve generalization across multiple modalities, our empirical study shows that using only the CSDN filter prior to extracting SIFT features and the Tan & Triggs pipeline prior

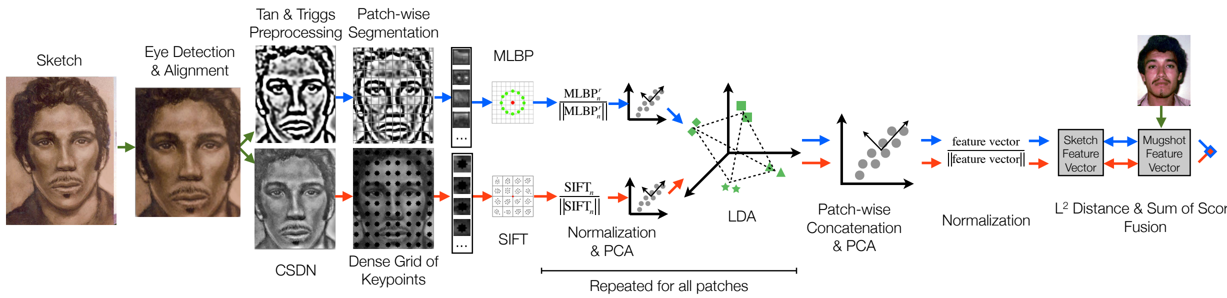


Figure 4.2: The holistic algorithm pipeline used by the FaceSketchID System. Following normalization and preprocessing, SIFT [8] and MLBP [9] features are extracted. Optimal subspaces are learned for each patch and the projected features are concatenated. After a PCA step to reduce template size, the final feature vector is normalized using the L^2 norm. The similarity scores based on the SIFT and MLBP feature vectors when comparing composite and mugshot are fused using the sum rule after z-score normalization.

to extracting MLBP features, results in the best matching performance for composites.

The Tan & Triggs preprocessing pipeline, in particular, resulted in significantly better qualitative performance. That is, after incorporating this pipeline into both the holistic and component-based algorithms, impostor matches returned at low ranks were observed to be more similar in appearance to the query composite (which could be argued to be a measure robustness for this recognition task). We note that although in [43] the preprocessing pipeline is used to handle differences in lighting, our qualitative analysis indicates that it helps handle differences in image quality (resolution, compression artifacts, etc.) as well.

For both SIFT and MLBP features, optimal subspaces are learned for each patch using linear discriminant analysis (LDA) after applying Principal Component Analysis (PCA) to reduce redundancy in the extracted features. Klare and Jain used a random-sample LDA (RS-LDA) technique introduced in [44] to handle the small sample size problem. Since we have relatively more training data, RS-LDA has been replaced by LDA in the FaceSketchID System to improve algorithm speed.

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 for both feature representations. 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. 4.2.

4.2.2 Component-Based Algorithm

As mentioned earlier, the component-based method used by the FaceSketchID System was proposed in [12] to match software-generated composites (created using FACES and Identi-Kit) to photographs. Similar to the holistic algorithm, the process for representing an image is not dependent on its modality. Thus, the process for representing a composite using the component-based method will be described.

In the component-based algorithm, facial components are automatically localized by detecting landmarks with an active shape model (ASM) via the STASM library [10]. As in [41], the ASM is initialized using eye locations provided by a COTS eye detector that is bundled with the FaceSketchID System. We note that 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 (via misalignment).

After facial components are extracted and normalized to a specific width and height and processed using the Tan & Triggs preprocessing pipeline, MLBP [9] 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 MLBP 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.

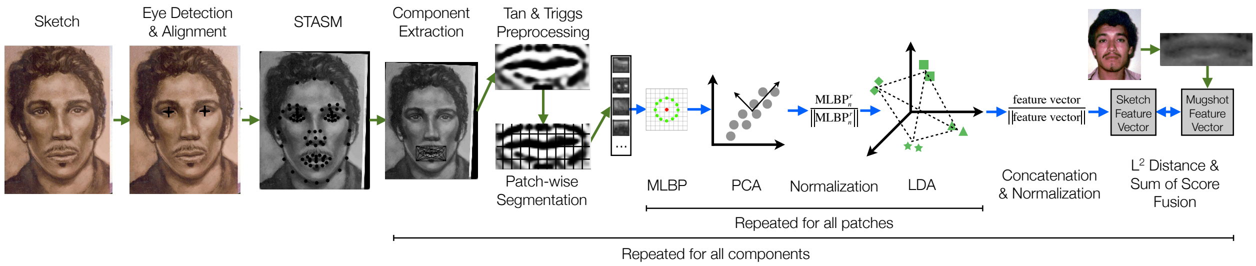


Figure 4.3: The component-based algorithm pipeline used by the FaceSketchID System. Following normalization, STASM [10] is used to detect 77 facial landmarks. Each of the 3 components used (nose, mouth, and eyes) are extracted and normalized to a specific width and height. MLBP features are extracted 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 after z-score normalization.

In [12], cosine similarities between corresponding patches of facial components are computed and the overall component similarity is the average of the patch-wise similarities. Concatenating the patch-wise feature vectors prior to computing an overall component similarity reduces template size and increases comparison speed within the framework of the FaceSketchID System. Further, the two comparison techniques result in comparable performance. As with [12], the most accurate components to be used during score fusion are determined empirically. Scores are normalized prior to fusion using z-score normalization and equal weights are applied to all components. For both the hand-drawn and software-generated composites, the component-based algorithm uses the mouth, nose, and eye components. A diagram of the component-based algorithm pipeline is shown in Fig. 4.3.

A simple sum-fusion rule is used when fusing the match scores from both algorithms after z-score normalization. When matching both hand-drawn and software-generated composites to mugshots, weights of 0.6 and 0.4 are assigned to the match scores of the holistic and component-based representations, respectively (i.e. the holistic SIFT and MLBP match scores have weights of 0.2, while each of the component-based match scores have weights of approximately 0.13).

Chapter 5 describes the experimental protocols used to evaluate the FaceSketchID System in detail.

CHAPTER 5

EXPERIMENTAL PROTOCOL

The gallery set for all experiments (with the exception of those listed in Section 5.4) consists of mated mugshots for the listed composites plus a set of 100,000 mugshots from the PCSO database. Experimental results are reported with and without filtering the mugshot gallery using demographic information (in the form of age range, race, and gender). Ground-truth demographic information was provided for the 100,000 mugshots used to extend the gallery from the PCSO. We estimate demographic information for the mugshots for which we do not have ground-truth information (the mated composites are defined to have an age range $\pm$ 5 years relative to the estimated age of the mugshot). Three commercial face matchers are used as baselines, which will be referred to as COTS-1, COTS-2, and COTS-3. Note that we are unable to train any of the COTS systems. Table 5.1 summarizes the databases used in our study. Fig. 5.1 shows example mugshots used to extend the gallery in matching experiments. Table 5.2 shows the distribution of demographics in the extended gallery.

Table 5.1: A summary of databases used in this study.

Database	Details	No. of Pairs
PRIP-HDC	**H**and-**d**rawn **c**omposites (HDC) with mugshot mates	265
PRIP-SGC	**S**oftware-**g**enerated **c**omposites (SGC) with mugshot mates	75
PRIP-VHDC	**V**iewed **h**and-**d**rawn **c**omposites (VHDC) with mugshot mates	93
PRIP-VSGC	**V**iewed **s**oftware-**g**enerated **c**omposites (VSGC) with photograph mates	123
CUHK-VHDC	**V**iewed **h**and-**d**rawn **c**omposites (VHDC) with photograph mates	1800

Figure 5.1: Example images from the 100,000 PCSO mugshots used to extend the experimental gallery.

Table 5.2: Demographic distribution of the 100,000 mugshots used to extend the gallery.

Males						
Ethnicity	Age Range					
	< 20	20-30	30-40	40-50	50-60	> 60
Asian	1	182	155	89	46	17
Black	101	7447	6511	4171	2953	917
Hispanic	10	2480	2945	1599	619	184
White	46	11441	11424	10013	9133	3285
Other	1	83	78	91	40	30
Females						
Ethnicity	Age Range					
	< 20	20-30	30-40	40-50	50-60	> 60
Asian	0	60	53	37	26	16
Black	2	2011	1660	1111	634	151
Hispanic	0	312	308	135	105	16
White	0	5002	4930	3944	2783	667
Other	0	19	28	9	6	3

5.1 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 of 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. For these pairs, the training set that results in the best recognition performance is used. Note that cross-validation is not necessary in these experiments since we are training and testing on different data.

Finally, we report the performance of a composite related to the Boston Marathon bombing suspect (see [45]) 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 trial[4]. 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.

5.2 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.

We include a comparison between the 75 composites from the PRIP-SGC database and the corresponding 75 composites from the PRIP-HDC database [15]. In this experiment, only the performance after fusing the match scores of both algorithms is reported.

5.3 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 optimal training sets as determined by the experiments in Section 5.1. 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. We report the retrieval rank achieved using the most accurate composite and the most accurate frame from the surveillance video for each subject. We also include a surveillance composite depicting Tamerlan Tsarnaev (the older brother alleged to be involved in the Boston Marathon bomb-

[4] `http://www.businessinsider.com/dzhokhar-tsarnaev-court-room-sketch-2013-7`

ing) [11] and report the FaceSketchID System's performance when matching the composite to photographs from [45]. Example surveillance data used in this study can be found in Figs. 3.2 (a,b,c).

5.4 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 [12]. A gallery of 10,000 mugshots from the PCSO is used to extend the gallery in these experiments to be consistent with that in [12]. While we have age range, gender and race information to filter the gallery set, only gender-based demographic filtering was reported in [12]. Thus, to provide a fair comparison with [12], all retrieval rates for the (viewed software-generated composite, photograph) experiments do not involve any demographic filtering.

Chapter 6 discusses the results of the experiments listed above.

CHAPTER 6

EXPERIMENTAL RESULTS

The following sections describe the recognition performance of the FaceSketchID System. Section 6.1 and Section 6.2 describe the performance of hand-drawn and software-generated composites, respectively. In Section 6.3, we discuss the performance of the FaceSketchID System on surveillance composites. Finally, we report the FaceSketchID System's performance on viewed software-generated composites in Section 6.4 to allow for comparisons against previously published results.

6.1 Hand-Drawn Composites

Fig. 6.1 shows rank retrieval results of the experiments involving hand-drawn composites. For the holistic algorithm (Fig. 6.1 (a)), the choice of data used to train the algorithm has little effect on the recognition performance, with the exception of training on only the pairs from the CUHK-VHDC database. Comparing Rank-100 and Rank-200 retrieval rates, the best training set (PRIP-HDC+CUHK-VHDC+PRIP-VHDC) differs from the worst training set (CUHK-VHDC) by approximately 2.4% and 2.6%, respectively. It is interesting to note that while the viewed hand-drawn composite combinations of training data exhibit similar performance, training the holistic algorithm on the CUHK-VHDC database alone is markedly worse than any other composite database. One explanation for this is that training on only (viewed hand-drawn composite, mugshot) pairs likely causes the subspaces to be biased towards (viewed hand-drawn composite, photograph) matching (in which texture and structure are extremely accurate). In the ideal case, in which the (hand-drawn composite, mugshot) pairs are extremely similar, using the CUHK-VHDC database to train the holistic algorithm may be advantageous. A more common case in composite recognition is that certain aspects of the composite are inaccurate or contain a significant amount of

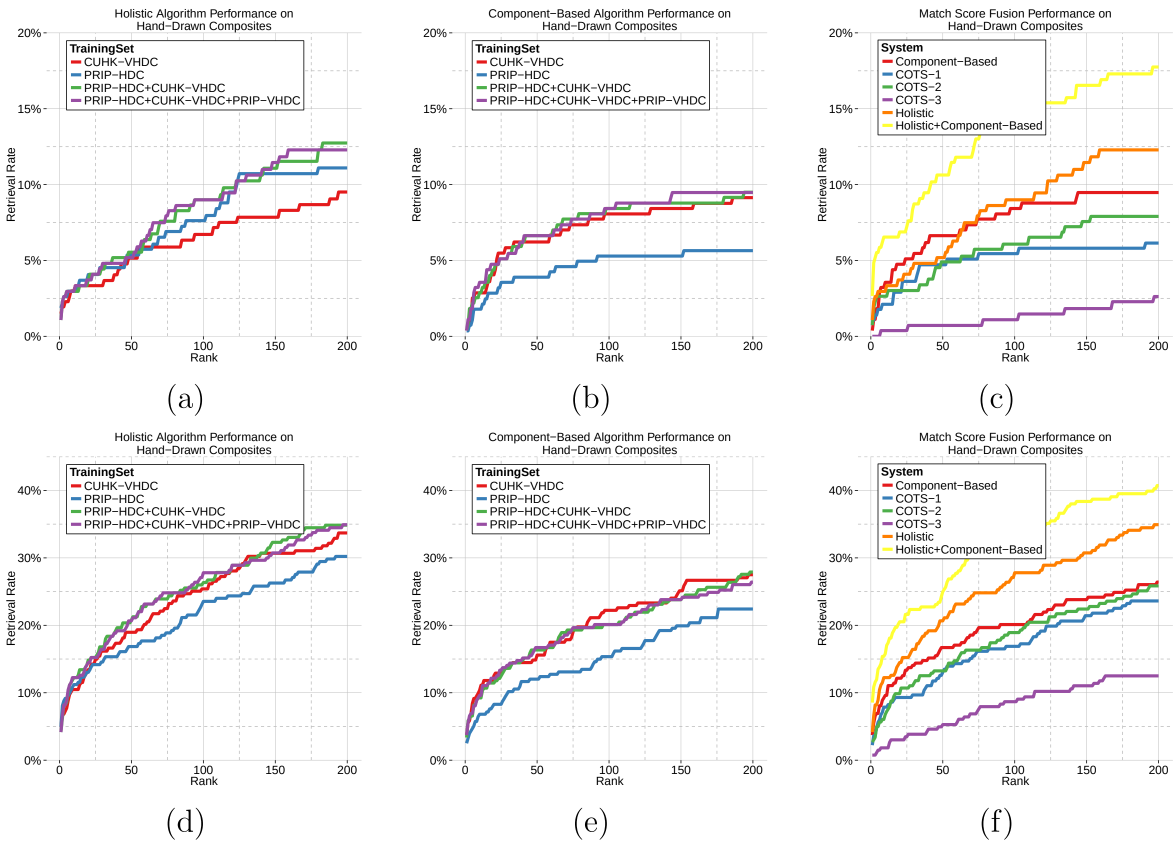


Figure 6.1: Results for the holistic (a,d) and component-based (b,e) algorithms when matching hand-drawn composites to mugshots. The best performance is achieved through a fusion of the holistic and component-based algorithm match scores (c,f). Three COTS matchers are included in (c,f) as a baseline.

noise (introduced by the artist "filling in the blanks" of the witness' or victim's description). Thus, it is understandable that a training set consisting of both (hand-drawn composite, mugshot) and (viewed hand-drawn composite, mugshot) pairs results in the best recognition performance for the holistic algorithm.

By contrast, the component-based algorithm (Fig. 6.1 (b)) has the worst recognition performance when trained on the PRIP-HDC database alone. The reason for this is that, at the component level (e.g. only considering the mouth), there can be a significant amount of noise if the components are misaligned. In cases where the component alignment is difficult, subspaces learned are likely influenced by noise more than the structure and texture of the

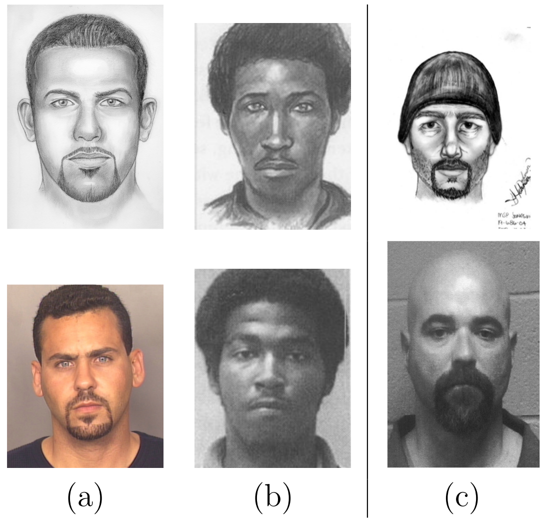


Figure 6.2: Examples of successful Rank-1 matches of (hand-drawn composite, mugshot) pairs (a,b). A failure case (c) shows a relatively accurate composite which was returned at a rank higher than 200.

component itself. Since facial landmark detection is more stable in the case of the (viewed hand-drawn composite, mugshot) pairs, using this data to train likely reduces noise present in the learned subspaces as a consequence of having more data with accurate alignment. This explanation justifies the fact that training using 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 in combination with the hand-drawn composites (which have a less accurate, but more forensically relevant, appearance), the component-based method is able to perform reasonably 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. 6.1 (c)). The Rank-200 performance after match score fusion is approximately 5.7% better than that of the holistic algorithm and approximately 8.3% better than that of the component-based algorithm. We include COTS retrieval rates in Fig. 6.1 (c) as a baseline against our best achieved matching performance. Examples of successful and unsuccess-

Method	Rank
Holistic	30
Component-Based	4961
Fused	8

Method	Rank
Holistic	556
Component-Based	206
Fused	34

Figure 6.3: Examples of retrieval rank improvement when matching hand-drawn composites to mugshots after match score fusion of the holistic and component-based algorithms.

ful (hand-drawn composite, mugshot) match cases can be found in Fig. 6.2. Examples of (hand-drawn composite, mugshot) match cases in which match score fusion improved the performance can be found in Fig. 6.3. 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. 6.1 (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 40%.

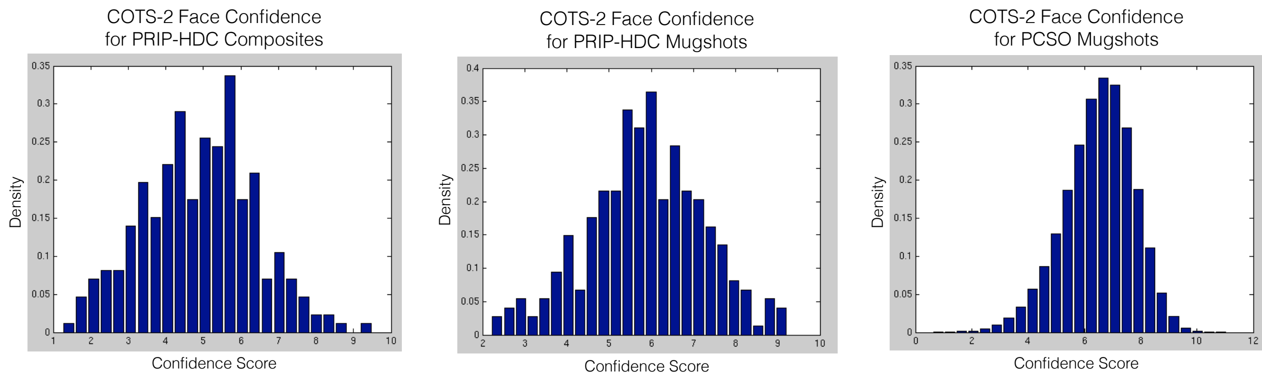
The data in the PRIP-HDC database varies significantly in terms of artist skill and image quality (resolution, compression artifacts, etc.) as it consists of images collected from multiple sources. Some composites have little information other than the outlines of major facial features (Fig. 6.4 (a)), whereas others have less anatomically correct facial proportions resulting in a “cartoon-like” appearance (Fig. 6.4 (b)). It is difficult to overcome the challenges intrinsic to these poor quality composites and, as expected, the performance of the FaceSketchID System suffers in these cases. We also note that the face confidence of the composites in the PRIP-HDC database (as determined by COTS-2) has been observed to be more consistent with that of the PRIP-HDC mugshots (Figs. 6.5 (a, b)) as compared

(a) (b)

Figure 6.4: Examples of poor quality composites with little information other than outlines of facial components (a) or unrealistic appearance (b). Both pairs shown are retrieved at higher than Rank-5000 after demographic filtering.

to the mugshots from the PCSO (used to extend the experimental gallery) (Fig. 6.5 (c)). This limitation of our experimental databases may imply that using different data that has consistent facial quality may result in different performance for all of the evaluated face matchers. However, we expect that the relative performance of the FaceSketchID System and the COTS face matchers will remain consistent with the results reported above.

Figs. 6.6 (a), (b), and (c) show example retrievals from the sequestered set after demographic filtering (using the PRIP-HDC+CUHK-VHDC+PRIP-VHDC database to train both algorithms). 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



(a) (b) (c)

Figure 6.5: COTS-2 face confidence scores for the PRIP-HDC database composites (a) ($\mu = 4.89$, $\sigma = 1.48$), the PRIP-HDC database mugshots ($\mu = 5.89$, $\sigma = 1.37$), and the PCSO database mugshots (c) ($\mu = 6.51$, $\sigma = 1.37$). The difference in quality is relatively smaller between the composites and the PRIP-HDC mugshots versus that of the PCSO mugshots. Therefore, the reported performance for all face matchers may be boosted due to the fact that it is easier to match images with similar facial quality.

(a) (b) (c)

Figure 6.6: 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.

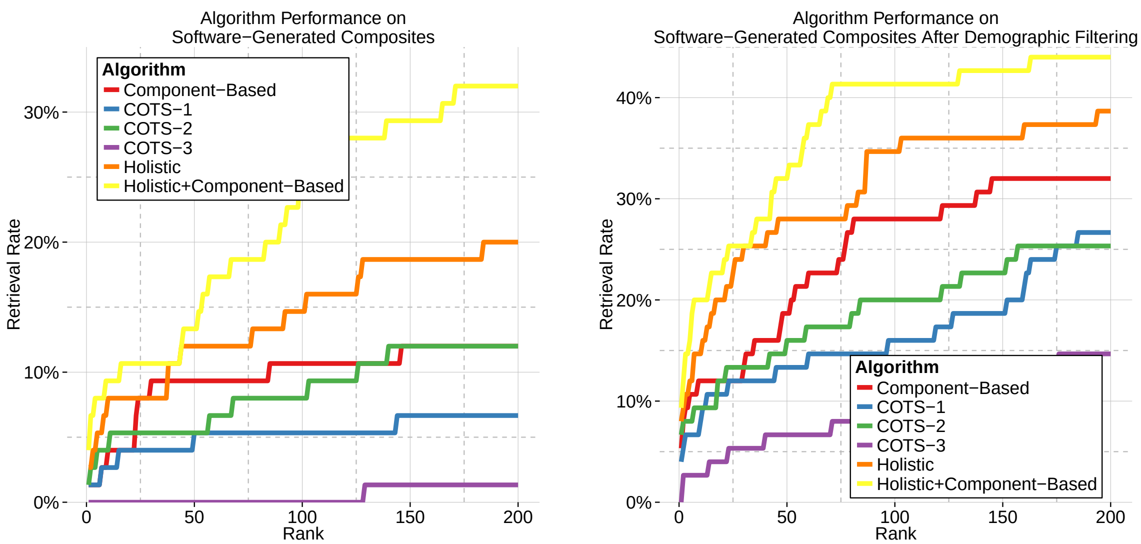
Figure 6.7: A composite and photographs depicting Dzhokhar Tsarnaev. Retrieval ranks for the FaceSketchID System are listed below the corresponding photograph. All ranks are after demographic filtering (15-25 year old, white, male).

after demographic filtering. Note that matching 13 out of 32 pairs by Rank-200 equates to a retrieval rate of 40.63%, which is roughly equal to the best results reported in the cross-validation experiments above.

Fig. 6.7 shows the composite and the photographs depicting Djokar Tsarnaev along with the rank at which they were retrieved (after demographic filtering). As noted in [45], 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).

6.2 Software-Generated Composites

Fig. 6.8 (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, achieving a 8% retrieval rate by Rank-25, 14.6% by Rank-100, and 20% by Rank-200. In contrast with the holistic algorithm, the component-based algorithm performs worse, achieving Rank-25, -100, and -200 retrievals rates of 8%,



(a) (b)

Figure 6.8: 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.

10.6%, and 12%, respectively. This may be due to the fact that it is difficult to precisely recognize (i.e. match at low ranks) an individual based purely on a set of components from the software-generated composite, as it is hard to achieve the same level of specificity as is possible when drawing facial composites by hand. 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 to those in the mugshot. Filtering with demographic information improves the Rank-200 performance of the holistic and composite-based algorithms when matching (software-generated composite, mugshot) pairs (Fig. 6.8 (b)) to 38.6% and 32%, respectively.

As mentioned in Section 6.1, both algorithms benefit from the use of training data that mirrors the data used during testing. Because we do not have enough (software-generated composite, mugshot) pairs, the only training data used were from the CUHK-VHDC databases. Thus, while we achieve respectable Rank-200 retrieval rates of approximately 20% and 12% for the holistic and component-based algorithms, respectively, we

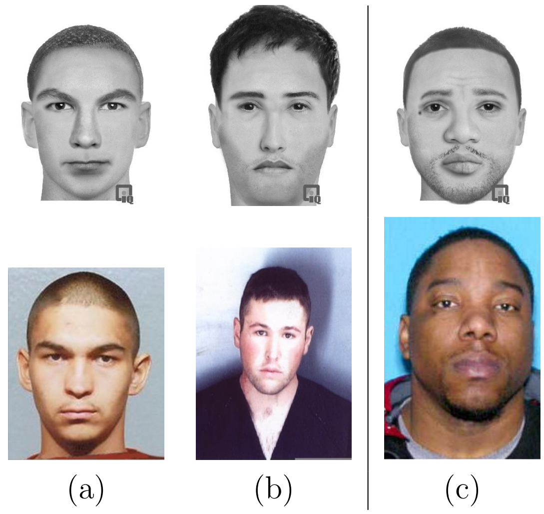


Figure 6.9: Examples of successful Rank-1 matches of (software-generated composite, mugshot) pairs (a,b). A failure case (c) shows a relatively accurate composite which was returned at a rank higher than 200.

believe that the performance of both algorithms could be improved if more operational software-generated composite data were available.

As with the hand-drawn composites, score fusion (using the sum rule) 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 to approximately 32% before demographic filtering and 44% after demographic filtering. Examples of successful and unsuccessful (software-generated composite, mugshot) match cases can be found in Fig. 6.9. Examples of (software-generated composite, mugshot) match cases in which match score fusion improved the performance can be found in Fig. 6.10. True accept rates (TAR) at false accept rates (FAR) of 0.1% and 1.0% are reported in Table 6.1 for both the PRIP-HDC and the PRIP-SGC databases to facilitate comparisons with other 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

Method	Rank
Holistic	5
Component-Based	467
Fused	1

Method	Rank
Holistic	492
Component-Based	146
Fused	6

Figure 6.10: Examples of retrieval rank improvement when matching software-generated composites to mugshots after match score fusion of the holistic and component-based algorithms.

Table 6.1: 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.

Matcher	Testing Database	TAR @ FAR = 0.1%	TAR @ FAR = 1.0%
FaceSketchID	PRIP-HDC	17.9% ± 3.9%	54.6% ± 3.1%
	PRIP-SGC	27.1%	65.4%
COTS-1	PRIP-HDC	12.2% ± 4.9%	38.1% ± 4.7%
	PRIP-SGC	12.9%	47.3%
COTS-2	PRIP-HDC	11.5% ± 2.8%	49.1% ± 25.5%
	PRIP-SGC	17%	100%
COTS-3	PRIP-HDC	6.4% ± 1.8%	24.9% ± 5%
	PRIP-SGC	6.9%	23.2%

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. 6.11.

Fig. 6.12 shows the performance of the FaceSketchID System on 75 hand-drawn compos-

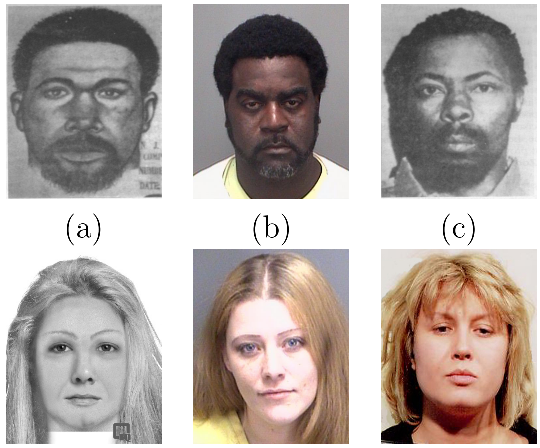

(a) (b) (c)
(d) (e) (f)

Figure 6.11: 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 appearence to the composite than the genuine mugshot match (c,f).

ites and 75 software-generated composites depicting the same individuals both before and after demographic filtering [15]. For these 75 subjects, our experimental results indicate that the software-generated composites are more accurate than the hand-drawn composites (32% vs. 12% before filtering at Rank-200, respectively). In an attempt to provide an unbiased comparison, the algorithms are trained on the (viewed hand-drawn composite, photograph) pairs from the CUHK-VHDC database. Because of this, it is surprising that the software-generated composites perform significantly better than the hand-drawn composites.

The most probable explanation for this is that the textures in the software-generated composites, as compared with the hand-drawn composites, more closely match the mugshots. Specifically, FACES uses components from visible-light images to create a composite (giving the composite a photo-realistic appearance). Thus, while the hand-drawn composite may be as accurate as its software-generated counterpart, the nature of the problem (recognition by comparison against a mugshot gallery) may favor the software-generated composites. However, we need to collect more software-generated composite data (using other software kits) to make a firm claim regarding the superiority of a given composite modality.

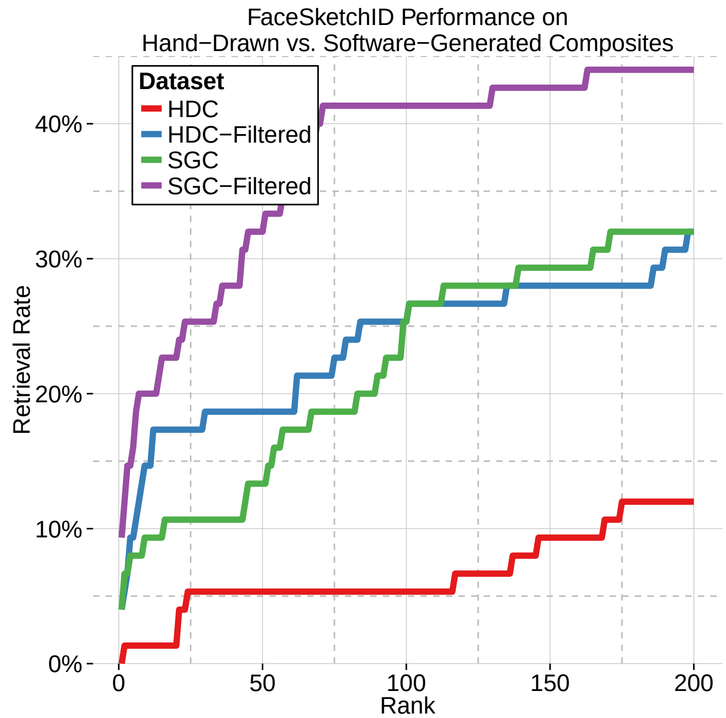


Figure 6.12: Results for the FaceSketchID System when matching 75 hand-drawn composites and 75 software-generated composites to mugshots before and after demographic filtering. COTS matchers are included as a baseline.

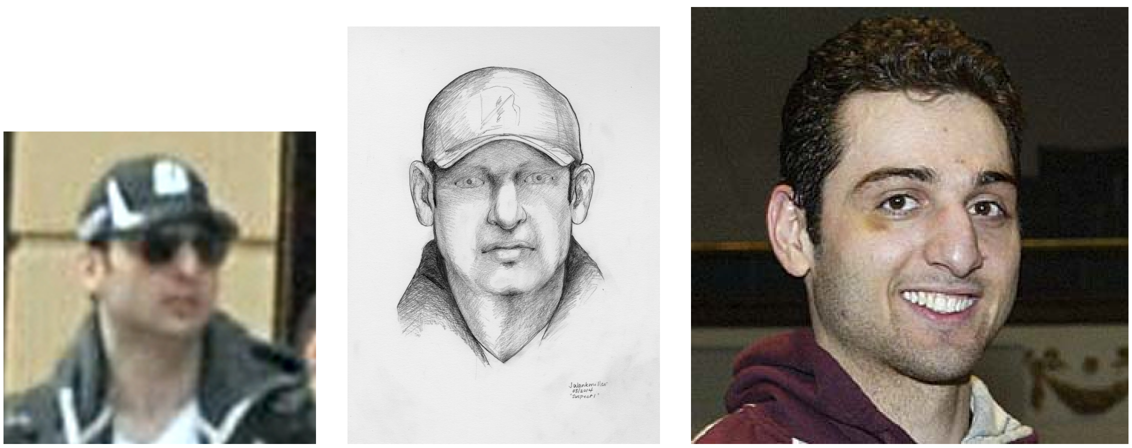
6.3 Surveillance Composites

Table 6.2 shows the retrieval ranks of (surveillance composite, mugshot) pairs for the FaceSketchID System and (surveillance image, mugshot) pairs for the COTS matchers after demographic filtering. When using surveillance composites, the FaceSketchID System is able to achieve results comparable to COTS systems when using surveillance frames.

We note that it is difficult to make meaningful inferences from the small number of (surveillance composite, mugshot) pairs used in this study. Further, the surveillance images used in our study are of reasonably high quality such that, in general, the use of surveillance composites is unnecessary (as the COTS matchers do well on the surveillance frames). However, cases with highly off-pose, occluded, or blurred faces present challenges to state-of-the-art unconstrained face recognition systems (such as those in Fig. 2.1). We hope to

Table 6.2: 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 failed to enroll the query surveillance mugshot. All retrieval ranks reflect a gallery of 100,000 subjects after demographic filtering.

Subject	FaceSketchID	COTS-1	COTS-2	COTS-3
1	114	60	**20**	95,962
2	**1**	**1**	1	5,131
3	32	**2**	FTE	367
4	22	**1**	**1**	733



(a) (b) (c)

Figure 6.13: A surveillance frame (a) of Tamerlan Tsarnaev was used to create a high-quality surveillance composite (drawn by Jane Wankmiller [11]) (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.

acquire additional surveillance data to evaluate the performance of the FaceSketchID System in more challenging surveillance cases.

Fig. 6.13 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 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,

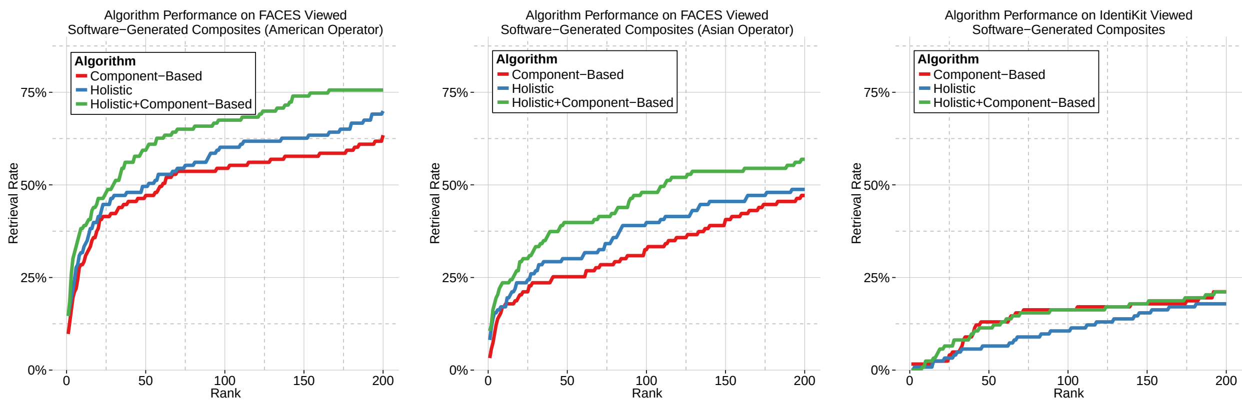


Figure 6.14: Results for the holistic, component-based, and fused algorithms on viewed software-generated composites created using FACES and IdentiKit. For the composites created using FACES, two operators (an American and an Asian) created composites [12].

photograph) pair reported in [45].

6.4 Viewed Software-Generated Composites

Han et al. [12] 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 comparable Rank-1, Rank-100, and Rank-200 retrieval rates of approximately 14.6%, 67.4%, 75.6% on the same dataset after fusing the match scores of the holistic and component-based algorithms (Fig. 6.14).

Han et al. also report the matching performance when FACES composites are created by an Asian user. Similar to the American-created composites, the FaceSketchID System is able to achieve higher retrieval rates, with an improvement at Rank-200 of approximately 4.9% over that in [12]. 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 [12]. For the composites created with Identi-Kit, match score fusion via the sum rule does not improve the performance of the FaceSketchID System because of the poor performance of the holistic algorithm (although another fusion methodology may be helpful). This can be explained by the fact that the Identi-Kit composites have little information other than outlines of facial components. For these pairs, the component-based algorithm detailed in this thesis achieves comparable results with the highest Rank-200 retrieval rate reported in [12].

CHAPTER 7

SUMMARY AND FUTURE WORK

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. This thesis improves upon the usability and recognition performance of these composites through: (i) An exploration of composite use cases, (ii) development of the FaceSketchID System, which provides law enforcement agencies a tool with which to more effectively make use of facial composite data, and (iii) an investigation of the effect of training the two algorithms on different (composite, photograph) databases to achieve the best matching performance.

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 also show the performance of surveillance composites, which are used for poor quality surveillance images where COTS systems are expected to 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 in all matching experiments.

To further improve the system's performance and robustness would require acquiring additional (composite, photograph) pairs. A special emphasis should be given to finding surveillance composites, specifically those with severe off-pose or occluded faces. We are also currently developing 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.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Turner Entertainment Networks, "Son of Sam," 2014. [Online]. Available: http://www.trutv.com/library/crime/photogallery/serial-killer-sketches.html?curPhoto=4

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

[3] Turner Entertainment Networks, "Ted Kaczynski: The Unabomber," 2014. [Online]. Available: http://www.crimelibrary.com/terrorists_spies/terrorists/kaczynski/5.html

[4] FACES 4.0, IQ Biometrics, http://www.iqbiometrix.com.

[5] I. Lovett, "Los angeles officials identify video assault suspects," *New York Times*, January 2011.

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

[7] A. M. Martinez and R. Benavente, "The AR face database," *CVC Technical Report*, vol. 24, June 1998.

[8] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, January 2004.

[9] 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.

[10] S. Milborrow and F. Nicolls, "Locating facial features with an extended active shape model," in *Proc. ECCV*, 2008.

[11] L. Best-Rowden, H. Han, C. Otto, B. Klare, and A. Jain, "Unconstrained face recognition: Identifying a person of interest from a media collection," *MSU Tech. Report*, 2014.

[12] 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.

[13] A. Jain, B. Klare, and U. Park, "Face matching and retrieval in forensics applications," *IEEE Multimedia*, vol. 19, no. 1, pp. 20–28, 2012.

[14] D. McQuiston-Surrett, 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.

[15] S. Klum, B. Klare, H. Han, and A. Jain, "Sketch based face recognition: Forensic vs. composite sketches," in *Proc. IEEE ICB*, 2013.

[16] X. Tang and X. Wang, "Face sketch recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 50 – 57, January 2004.

[17] 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.

[18] D. Lin and X. Tang, "Inter-modality face recognition," in *Proc. ECCV*, 2006, pp. 13–26.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] S. Li and A. Jain (eds.), *Handbook of Face Recognition*, 2nd ed. Springer, 2011.

[25] 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.

[26] 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.

[27] 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.

[28] E-FIT, VisionMetric Ltd., http://www.vision-metric.com.

[29] 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.

[30] Identi-Kit, Identi-Kit Solutions, http://www.identikit.net/.

[31] G. Wells and L. Hasel, "Facial composite production by eyewitnesses," *Current Directions in Psychological Science*, vol. 16, no. 1, pp. 6–10, February 2007.

[32] C. Frowd, D. Carson, H. Ness, D. McQuiston-Surrett, 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.

[33] 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.

[34] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, "Xm2vtsdb: the extended m2vts database," in *Proc. of International Conference on Audio- and Video-Based Person Authentication*, vol. 24, June 1999, pp. 72–77.

[35] W. Zhang, X. Wang, and X. Tang, "Coupled information-theoretic encoding for face photo-sketch recognition," in *Proc. IEEE CVPR*, 2011.

[36] P. J. Phillips, H. Moon, 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.

[37] Itseez, "The OpenCV Library," www.opencv.org, 2000.

[38] G. Guennebaud, B. Jacob *et al.*, "Eigen v3," http://eigen.tuxfamily.org, 2010.

[39] Digia, "Qt5," http://qt-project.org/, 2013.

[40] J. Klontz, B. Klare, S. Klum, A. Jain, and M. Burge, "Open source biometric recognition," in *Proc. IEEE BTAS*, 2013.

[41] 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.

[42] E. Meyers and L. Wolf, "Using biologically inspired features for face processing," *Int. J. Comput. Vision*, vol. 76, no. 1, pp. 93–104, 2008.

[43] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Trans. Image Proc.*, vol. 19, no. 6, pp. 1635 – 1650, June 2010.

[44] X. Wang and X. Tang, "Random sampling for subspace face recognition," *Int. J. Comput. Vision*, vol. 70, no. 1, pp. 91–104, 2006.

[45] J. Klontz and A. Jain, "A case study on unconstrained facial recognition using the boston marathon bombings suspects," *MSU Tech. Report*, 2013.