PhD Dissertation

Show, Don't Tell: Non-Verbal Eyewitness Testimony Based on Non-Artistic Face Sketches

Hossein Nejati

Supervisor:

Dr. Terence Sim

October 14, 2013

Abstract

In this text we focus on the problem of eyewitness face sketch recognition, in which we found particular interest due to the presence of a human provider of the information (the eyewitness) and a machine processor of the information (face sketch recognition algorithm). Reviewing the literature of over 30 years of psychological studies, we showed that currently used eyewitness testimony procedures (ETPs) are inaccurate and highly unreliable. We showed that due to these problems, current ETPs not only produce unreliable results (forensic sketches), but also cause distortions to the eyewitness' mental image of the target face. The crucial problems in these vital procedures have drastic consequences, and can sometimes cause death for an innocent human. On the other hand, automatic face sketch recognition methods (FSRs) have been only designed to recognize face sketches which are drawn with a significant similarity to their photo counterparts, and therefore they these automatic methods cannot be applied for recognizing forensic sketches.

Our approach to tackle the eyewitness face sketch recognition problem is to first understand the psychological challenges of the problem, and then based on this understanding, try to avoid sources of unreliability in the ETPs. Based on this strategy we proposed to use non-artistic sketches directly drawn by the eyewitness (Main Sketches), as the medium to retrieve eyewitness' mental image of the target face. Using the directly drawn sketches avoids added distortions of issues such as verbal overshadowing (distortion of a visual memory, due to verbal description of it), piecewise face reconstruction (reconstructing a face, using selecting from different types and shapes of facial components), and implanted ideas. On the other hand, these drawings are also distorted by the eyewitness' mental face perception bias, and face drawing bias, that together we refer to as sketching bias. In our FSR, we therefore proposed to estimate

the sketching bias for each eyewitness and debias the Main Sketch, to reach an estimation of what the eyewitness meant by drawing the Main Sketch. Finally, for matching this estimation to the photo database, we proposed a weighted dynamic point correspondence, which is inspired by psychological suggestions for face perception in humans.

To test our propose method we collected 3 datasets of sketch-photo pairs, including a total of 860 sketches, drawn by 86 human participants. In our tests, we compare our method with the most important previous methods, both on the sketches from our datasets and other publicly available sketch datasets, and we showed the improvements of our method over the others, in terms of accuracy and gallery size. We also provided an important comparison in our tests (not found in previous literature) which is the effect of number of training samples on the accuracy of the algorithm. The importance of this test is rooted in the time consuming procedure of producing sketches by the eyewitness, which eventually results in having only a few sketch samples from each eyewitness to be used for perception bias estimation.

Our reviews in both psychology and computer vision in eyewitness sketch recognition, accompanied with our proposed method and experimental results, suggest a new perspective to develop better eyewitness testimony procedures as well as automatic face sketch recognition methods, which can even shed light on other related computer vision problems. For example, we here present results of applying our proposed concepts on the ear image identification application and showed that with minor problem-related changes, we could surpass previous ear recognition methods.

Finally, in the final chapter of this text we suggest method of combining our approach with traditional eyewitness testimony procedures (to cover cases of poor memory of the target face), possibilities for future works, and final con-

clusions, with the hope that our work can improve computer vision algorithms and more importantly, improving human lives.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Unlike the common belief of the accuracy of eyewitness testimony procedures, studies show that these procedures are not only highly error prone and unreliable, but also the major cause of wrongful convictions. A recent survey by Morgan et al. showed than more than 75% of the convictions overturned through DNA testing since the 1990s were based on eyewitness testimony [Morgan et al., 2007]. The DNA exoneration, as well as a number of other archival analyzes, have led many to the conclusion that false eyewitness identification is the primary cause of wrongful convictions in the United States [Huff et al., 1996, Wells et al., 1998, Scheck et al., 2000, Gross et al., 2005]. Despite these tragic reports, the use of eyewitness testimony for forensic applications is still a common practice, with roots that go back to the beginning of the century [Yarmey, 1997]. The use of these eyewitness testimony procedures (ETPs) is not because of the police ignorance, but because of having no other option than using traditional ETPs in many criminal cases. When an eyewitness has seen the face of a person of interest (also known as the target face), the eyewitness usually attends a police station, in where he/she will be subjected to ETPs. Police artists are trained to draw the target faces based on the verbal

description of the eyewitness, and police officers are trained in effective use of face composite software to reconstruct the face piece by piece (piecewise face reconstruction). These efforts are required because in many situations, the only image of the target face is a mental image in the eyewitness' mind. Eyewitness testimony procedures are therefore *supposedly* designed to use the eyewitness' memory of the target face to find the target identity, either directly (using photographs or lineups) or indirectly (reconstruct the target face through a police artist or a photo-composite software). This reconstructed face (drawn by a police artist, or produced by photo-composite software), known as *forensic sketch* or *eyewitness face sketch*, should then be matched against the police database of faces or distributed in the public.

More than 30 years of psychological studies show that forensic sketches are different from normal exact sketches (artistic sketches drawn from a person or a photo), in terms of accuracy in representing facial features and appearance details (compare forensic sketches in figures 1.1 and 2.8, with exact sketches in figures 2.1 to 2.7). These studies indicate that current eyewitness testimony procedures (ETPs) are highly susceptible to error and should be reformed [Munsterberg, 1927, Morgan et al., 2007, Carlson et al., 2008]. Based on evidence shown in the literature, the flaws in traditional ETPs not only affect the final sketch, but also distort the mental image of the face in the eyewitness's brain, without the eyewitness himself sensing this change [Yarmey, 1997, Morgan et al., 2007]. The extents of these disturbances are so critical that some researchers have suggested entirely avoiding the use of these testimonies in courts [Yarmey, 1997]. As an example of these distortions, the eyewitness can be easily confused by misleading information [Zhu et al., 2010] such as viewing similar faces or subjective questions. Moreover, the piecewise reconstruction of the face in current ETPs causes additional distortions, because it is in-

Figure 1.1: Some examples of unreliable artistic sketches (two left columns from HTTP://depletedcranium.com) and composite sketches (two right columns [Sinha et al., 2006b]).

compatible with the holistic analysis of the human visual system on the faces [Sinha et al., 2006a, Zhang et al., 2010], and as a result, the final reconstructed face significantly deviates from the presumed target face [Sinha et al., 2006a]. Therefore, at the end of these procedures, the sole image from the target face is unrecoverable (as several famous criminal cases also show) [Chabris et al., 2010]. The problems in current ETPs are basically because human memory is fragile, malleable, and susceptible to suggestion [Bernstein and Loftus, 2009], which in turn render results of eyewitness testimonies unreliable.

Regardless of the reliability of the final forensic sketch, police should search for the identity of this reconstructed face. There are several methods proposed for automatic recognition of face sketches (i.e. face sketch recognizers, FSRs). But all of these methods are designed and fine-tuned to recognize exact sketches, drawn by artists, directly from the photographs of faces (similar to portrait sketches). Several proposed FSRs have considered that the amount of

information in face photos is larger than in face sketches, and therefore tried to transform photos to sketch-like images, to prevent information loss. Among the firsts is the work by [Tang and Wang, 2004] in which an eigenface transformation is proposed to project a face photo to the face sketch space, resulting in a sketch-like image. This work reported recognition accuracy of 89%, tested on CUHK face sketch dataset [Wang and Tang, 2009]. This work was followed by [hui Li et al., 2006] in which a sketch-photo pair image is concatenated into a single vector to learn the PCA classifier with correlation to both the sketch and the real face. A non-linear transformation was also presented in [Liu et al., 2005] to replace photo patches with the most similar patch from the sketch gallery (using a PCA-based scoring). The result of this patch replacement classified by non-linear discriminant analysis reported of recognition accuracy of 92% on the CUHK dataset. This method was further improved using multi-scale Markov random field [Wang and Tang, 2009], to synthesize a smooth sketch that marginally improved the accuracy. Xiao et al. [Xiao et al., 2009] proposed a sketch-to-photo transformation in order to transform the problem into a photo-to-photo matching problem. They used an embedded hidden Markov model for patch replacement to synthesize a photo-like image, and then classification using PCA. The experimental results on CUHK dataset reported to have up to 89.1% accuracy in recognition. More recently, FSR methods have been proposed based on Partial Least Squares (PLS) [Sharma and Jacobs, 2011], random forests [Zhang et al., 2011b], support vector regressors [Zhang et al., 2011a], combination of local binary pattern and histogram of Gabors [Galoogahi and Sim, 2012a], and combination of multi-scale LBP and SIFT features [Klare et al., 2011].

Regardless of reported accuracy of the above algorithms, these methods are proposed to address the forensic sketch recognition problem, but all of them

have been tested on *exact sketches* (compare figures 2.1 to 2.7 with figures 1.1 and 2.8), that have significant similarities to their target faces (including exactly similar facial component shape, illumination and shading, skin texture, and even hairstyle). A recent study [Choi et al., 2012] showed an astonishing recognition rate of 85.22% only using hair regions, as well as that the accuracy of an off-the-shelf face photo matcher (merely using shape and edges), even without training, can outperform the currently proposed FSRs [Choi et al., 2012]. In contrast, a real forensic sketch is very likely to be significantly different from its respective target face [Sinha et al., 2006a, Zhang et al., 2010, Klare et al., 2011, Nejati et al., 2011, Choi et al., 2012]. Thus, we argue that although the test results of the previous FSRs show almost perfect performances for exact sketches, these FSRs cannot be used for recognizing forensic sketches (detailed discussion in Chapter 2). We can therefore conclude two main gaps from the literature. First that current eyewitness testimony procedures are unreliable (based on psychological studies); second that current FSRs cannot reliably recognize forensic sketches (based on several tests by [Klare et al., 2011, Choi et al., 2012]).

The motivation for this work is therefore addressing the literally life threatening problems in the eyewitness testimony procedures by (1) designing a new eyewitness testimony procedure for faces that avoids psychological pitfalls; and (2) introducing a robust and practical face sketch recognition based on this new ETPs design. From another perspective, in traditional ETPs the eyewitness contribution is passive (providing verbal description and confirmation), while the artist has the main active contribution that produces the final sketch. In contrast, our approach is to remove the artist from the procedure and transfer the active contribution to the eyewitness, and therefore avoid several important psychological problems of traditional ETPs. In this perspective, when the

tradition ETP is at extreme minimum of eyewitness contribution, our method is at the extreme maximum eyewitness contribution, providing new options for ETPs and FSRs.

We also noted that the process of producing an eyewitness face sketch involves the mental recollections of a human from a target face (eyewitness' mental face image), a method of transferring this mental image (traditionally verbal description), and an artist or machine (traditionally face composite software) which compiles the transferred information into a face sketch. In our proposed methods, we therefore incorporate findings from the human visual system, while particularly focusing on the automatic eyewitness face sketch recognition (FSR) application. We first review currently proposed automatic methods for FSR, their achievements, and their problems that together show the current gaps in addressing eyewitness face sketch recognition problem (Chapter 2). In order to obtain a clear understanding of the face sketch recognition (FSR) problem, we then review the psychological challenges related to the eyewitness testimony procedures (ETPs) to expose the extent of unreliability of these procedures and therefore their results, forensic sketches (Chapter 3). Based on these reviews we then propose a novel eyewitness testimony procedure (ETP), accompanied by a compatible face sketch recognition (FSR), to both avoid psychological pitfalls and implementing a robust and practical automatic face recognition method (Chapter 4). To show the effectiveness of our proposed method, we compare the performances of our methods with the most important previous FSRs on our collected dataset of 860 face sketches as well as on the publicly available CUHK face sketch dataset [Wang and Tang, 2009] with 188 sketches. We analyzed different properties of our method including average accuracy, effect of gallery size, effect of piecewise vs. holistic matching, and the effect of the number of training samples of final performance

(Chapter 5). We also tested the emerged psychologically-inspired framework on another human identification problem, identification between twin siblings based on their ear image, which indicates the capability of application of this framework for a wider range of application with some problem-specific modifications (Chapter 4.4). We finally summarize our works, draw conclusions, and discuss future works in the final chapter, Chapter 6. In the conclusion, we discuss the possibility of incorporating some of the parts of traditional ETPs to our proposed method, to cover fall-back options for our system, particularly for the cases that the eyewitness requires memory triggers to recall the target face structure.

We now continue this chapter with describing our contributions.

1.1 Thesis Contributions

In this text we present our contributions in eyewitness face sketch recognition as follows:

1. A novel, non-verbal eyewitness testimony procedure (ETP), designed based on psychological findings, to deliver non-artistic face sketches, while avoiding many psychological pitfalls of the current procedures. Our ETP provides another option for conducting a more reliable ETP.

2. A accompanying new automatic face sketch recognition method (FSR), designed based psychological findings, to robustly match the non-artistic sketches to the photo database, based on the human's memory properties.

3. The largest face sketch database to date, with non-artistic sketches and new features such as including information about drawers, and sketches from time-delayed face image exposures.

In both parts of our system, ETP and FSR, we combine psychological findings, with image processing techniques to create a unique combination, required to address the eyewitness face sketch recognition problem. This combination of psychology and engineering, not found in previous approaches to this problem, gives our approach the ability to both cope with the special behavior of human's memory of the face, and automation of recognizing the generated face sketch based on mug-shot photos.

Our proposed eyewitness testimony procedure (ETP) is a non-verbal method of retrieving the eyewitness' memory of a face. The basis of our proposed ETP is on non-artistic sketches, drawn directly by the eyewitness. Being the first non-verbal ETP, we prevent adding several types of distortions to the final sketch, by removing the artist from the ETP, avoiding piecewise reconstruction of the face, biased instruction, post-event information, etc. But more importantly, we prevent distorting the mental image of the face in the eyewitness' mind, by avoiding verbal overshadowing distortion of a visual memory, due to verbal description of it), and exposure of the eyewitness to similar faces. While these problematic procedures are regularly practiced in current ETPs, in chapters 3 and 4 we describe the details of how our ETP have a better chance of faithfully retrieving the memory of the face, without distorting this memory in the eyewitness' mind. In terms of eyewitness participation, we also provide another option in which eyewitness has the main contribution to the sketch (by drawing it by him/herself), which is clearly contrasted with current ETPs in which eyewitness has a passive contribution and the police artist has the responsibility of producing the sketch from the verbal description.

In our accompanying eyewitness face sketch recognition (FSR), we particularly focus on the perceptual and sketch drawing biases of each eyewitness, and based on the eyewitness' information from the ETP stage, we try to esti-

mate "what the eyewitness mean" based on "what the eyewitness draws". We first estimate and remove introduced biases to the non-artistic sketch, based on a set of training sample sketch-photo pairs. Then we weight this debiased sketch based on a psychologically-inspired weighting scheme to predict the visually important parts of the sketch. We finally match this weighted sketch to the photo database by imposing a temporal order to the sketch. In chapters 4 and 5 we show that to faithfully recover the target face appearance from the eyewitness' memory recalls, one should account for the processes of face perception and face drawing for each eyewitness.

In order to test our methods, we collected the largest dataset of sketch-photo pairs with unique properties. The interesting properties of our sketch dataset include the use of non-artist sketch drawers, recording of additional information such as race, skin color, and hair color from the perspective of the eyewitness, recording of eyewitness' confidence map, and involving time delay between photo exposure and sketch drawing.

Chapter 2

Automatic Face Sketch Recognition Related Works

Once a face is reconstructed based an eyewitness testimony, it can be matched against the police database of faces. In this stage, automatic face sketch recognition methods (FSRs) are introduced to perform automatic matching between the forensic sketches and the database of mugshots. Several works have been proposed on automatic face sketch recognition (FSR), treating the forensic sketch recognition as yet another face recognition problem, but in a slightly different representation: the sketch sub-space. This is because these methods assume that the forensic sketches are (1) high quality and error prone reconstructions of the target faces, and (2) similar to the target face appearance even in small details. However, as discussed in the Chapters 1 and 3, the first assumption on forensic sketch reliability is falsified by psychological research [Munsterberg, 1927, Morgan et al., 2007, Carlson et al., 2008]. The second assumption is also false as forensic sketches are produces based on eyewitness's verbal description, and in presence of several sources of distortions. Therefore forensic sketches cannot reconstruct details such as hair style or exact shading

of the target face in the mugshot photo [Zhang et al., 2010] (compare figures 2.1 to 2.7 with Figure 2.8). Therefore, while previous FSRs reported accuracy rates as high as 92% [Liu et al., 2005], the applicability of these methods in recognizing forensic sketches is strongly questioned.

We continue this chapter by the review of the current automatic face sketch recognition methods, their assumptions, methods, and problems.

2.1 Automatic Eyewitness Face Sketch Recognition

Regardless of the reliability of a forensic sketch (resulting sketch of an ETP), this sketch, is regarded as a representation of the target face which should be matched against the police face database of criminals. Several different face sketch recognition algorithms (FSRs) are proposed in the literature for recognizing exact face sketches. These exact sketches are drawn by artists while looking at a face photo, and therefore are significantly similar to their face photo counterparts (unlike forensic sketches that are drawn based on verbal description and are highly unreliable). In general, these FSRs can be categorized into the methods that try synthesizing sketch-like images from face photos, and the methods that try performing the opposite, synthesizing photo-like images from face sketches, but definitely not forensic sketch recognition methods.

2.1.1 Matching Exact Sketches

Face sketch recognition methods are ultimately designed to be used for matching forensic sketches, which are drawn based on eyewitness' verbal description (see Section 2), and if they bear similarities to the target face, they surely have

Figure 2.1: An example of photo to sketch eigen transformation proposed in [Tang and Wang, 2004]. From left to right: original photo, eigenface reconstruction of photo, eigen transform reconstruction of sketch, original sketch.

significant distortions from the target face, even in a perfect eyewitness testimony procedure. On the other hand, almost all of FSRs in the literature are designed and tested based on *exact sketches*. Exact sketches are drawn by an artist while looking at the face photo and as shown in figures 2.1 to 2.6. These sketches have significant similarities to their target faces (including exactly similar facial component shape, illumination and shading, skin texture, and even hairstyle), far from the forensic sketches. All of the methods we review in this section have used the exact sketches for their tests (and most likely their designs).

Even when using exact sketches, face sketches and photos are from different modalities and this brings more difficulties for to match a photo and sketch than normal photo to photo matching. One approach to solve the modality difference between sketches and photos is to use a photo-to-sketch transformation, before performing the matching. Among the first to propose an FSR algorithm were Tang et al. [Tang and Wang, 2004] who proposed an eigen transformation to transfer gallery face photos to pseudo-sketch images. This transformation is very similar to eigenface transformation, except that in the reconstruction stage, the projected photo into the eigen space, the weight vec-

Figure 2.2: An example of non-linear photo to sketch transformation proposed in [Liu et al., 2005]. From left to right: photo image, sketch drawn by artist, pseudo-sketch with non-linear method, pseudo-sketch with the eigen transform method [Tang and Wang, 2004].

tor b_p is reconstructed not from the photo training set, but from sketch training set. This transformation decreases the difference between the faces and the sketches, and results in better performance in the next step, matching. In the matching step, these pseudo-sketches were then matched against a gallery of artistic sketch, using a PCA-based algorithm, with a reported recognition accuracy of 89%. An example of the sketch-photo pairs used in this work is presented in figure 2.1.

Liu et al. [Liu et al., 2005] further improved the photo-to-sketch transformation using a non-linear transformation. In this method photos and sketches are first divided into patches and then, each patch in a photo is replaced by the most similar patch from the patches in the sketch gallery. Finding the most similar patch is based on the similarity of the eigenvalues of the photo and sketch patches, based on similar technique introduced in [Tang and Wang, 2004]. The result of the patch replacement (i.e. the pseudo-sketch) is then matched against an artistic sketch gallery, using non-linear discriminant analysis (NLDA) with a reported accuracy of 92%. An example of the sketch-photo pairs used in this work is shown in figure 2.2.

A more recent photo-to-sketch transformation method is proposed by Li et

Figure 2.3: An example of photo-sketch pair used in [hui Li et al., 2006]. The "sketches" used in this work were merely transformed images of the photos, and not real hand-drawn sketches.

al. [hui Li et al., 2006] in which eigenface transformation is similarly employed. In this method, instead of using sketch-only or photo-only vectors, a sketch-photo pair image is concatenated into a single vector to calculate the eigen-vectors, and therefore, the calculated eigen space bears a correlation to both the sketch and the photo spaces. Although this method may have advantages over previous methods like [Tang and Wang, 2004, Liu et al., 2005], this method is only tested on synthetic images, transformed pseudo-sketches from the face photos, and not real sketches drawn by a human, therefore their reported results cannot be compared with other methods. An example of the images pair used in this work is illustrated in figure 2.3.

Wang and Tang introduced another improvement to patch based the photo-to-sketch transformation [Wang and Tang, 2009] in which after the similar patch replacement using eigen-value scoring (similar to [Liu, 2006]), a trained multi-scale Markov random field stitches and warps the patches into a final sketch which results in a smoother final pseudo sketch. This final stitched sketch is then used for sketch-to-sketch matching to find the target face, based on pre-calculated eigen-vectors from the sketch feature space. Authors have reported accuracy of 96% on sketch-photo pairs such as the pairs illustrated in figure 2.4.

While most of the previous works have focused on photo-to-sketch trans-

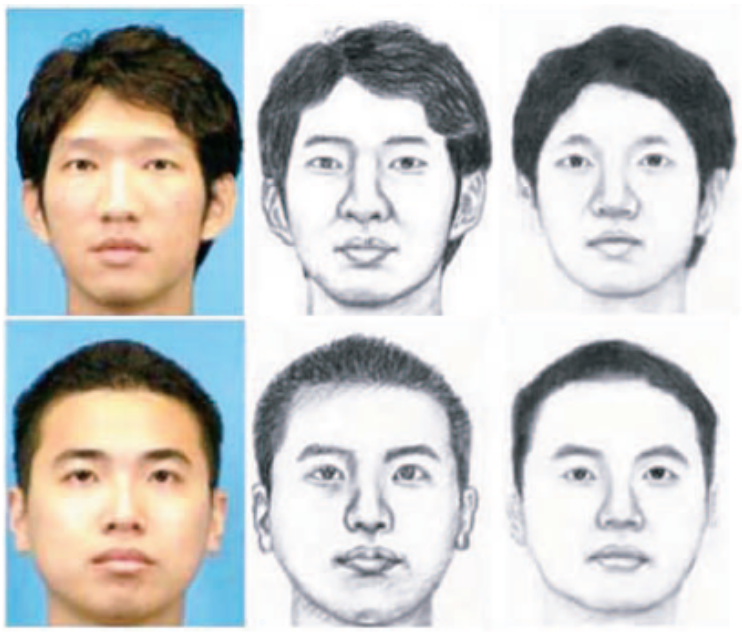

Figure 2.4: Examples of sketch-photo pairs used in [Wang and Tang, 2009]. From left to right: photo, artist sketch, and estimated sketch.

formation, Xiao et al. [Gao et al., 2008b,a, Xiao et al., 2009] proposed an approach which exploits the opposite direction, sketch-to-photo transformation, and tried to change the problem into a photo-to-photo matching problem. In this work, photos and sketches are first divided into patches and then given a sketch a pseudo-photo is generated by replacing the sketch patches with most similar photo patches. In order to find the most similar patches, embedded hidden Markov model (E-HMM) is used to extract the main two-dimensional features in a sketch patch with a moderate computational complexity. The resulting pseudo-photo image is then classified using PCA, with a reporting accuracy of 98%. An example of sketch-photo pairs used in this method is illustrated in figure 2.5.

Authors of [Zhang et al., 2011b] introduced another sketch-to-photo transformation and matching method very similar to [Xiao et al., 2009] by adding support vector regressors to the E-HMM technique, tested on similar sketch-photo pairs with minor improvements in the accuracy.

Figure 2.5: An example of sketch-photo pairs used in [Zhong et al., 2007, Xiao et al., 2009]. From left to right: photo, artist sketch, synthesized photo using method in [Xiao et al., 2009], and synthesized sketch using method in [Zhong et al., 2007]

Other than transforming one of the sketch or photo to the other one's space, some approaches have chosen features with capability of direct comparison between sketch and photos. [Pramanik and Bhattacharjee, 2012] used only a set of geometric face features like eyes, nose, eyebrows, lips, etc. and their length, width and area ratio as the feature vector for matching sketch-photo pairs. Then given a face sketch probe, a KNN classifier was used to find the closet matching face photo, with a reported accuracy rate of 80%. Examples of sketch-photo pairs used in this work is presented in figure 2.6.

Bhatt et al. presented a direct sketch-to-photo matching algorithm [Bhatt et al., 2010] in which discriminating information present in local facial regions are retrieved at different levels of granularity. Both sketches and digital images are decomposed into multi-resolution pyramid to conserve different frequencies of information which forms the discriminating facial patterns. Authors used extended uniform circular local binary pattern descriptors on these patterns to form a unique signature of the face image. In the next step for matching, a genetic optimization algorithm finds the optimum weights corresponding to each facial region. The information obtained from different levels of Laplacian pyramids are combined to improve the identification accuracy. The reported accuracy of this algorithm on artistic sketches such as the ones illustrated in

Figure 2.6: Examples of sketch-photo used in [Pramanik and Bhattacharjee, 2012]

figure 2.7 is 88%.

Galoogahi and Sim presented new face descriptor which is relatively invariant to the sketch/photo modality differences [Galoogahi and Sim, 2012b]. This descriptor called Local Radon Binary Pattern (LRBP) captures face shape characteristics in both the sketch and photo modality, by transforming face image (or sketch image) into Radon space and in this space uses Local Binary Pattern (LBP) to encode local features of the face shape. Then the concatenating histogram of local LBPs (called LRBP) is used for classification. Their experiments on exact sketches of CUHK [Wang and Tang, 2009] and CUFSF [Zhang et al., 2011a] datasets were with 99.51% and 91.12% accuracy, respectively.

Galoogahi and Sim also proposed a more recent FSR based on histogram of averaged oriented gradients (HAOG) to again provide a modality invariant descriptor for face sketch recognition [Galoogahi and Sim, 2012a]. The use of HAOG is motivated by the fact that orientations of stronger gradients, such

Figure 2.7: Examples of sketch-photo pairs used in [Bhatt et al., 2010].

as prominent gradients of facial components, are more modality invariant than weaker gradients, such as fine textures, shadows and wrinkles. Authors showed that using this descriptor on patches with different resolutions, they can reach up to 100% accuracy on CUHK [Wang and Tang, 2009] and AR [Martinez and Benavente, 1998] datasets.

A comparative study is also recently presented by Zhang et al. [Zhang et al., 2010] in which artistic sketch recognition accuracy in humans and PCA-based classification are assessed. In this study, five different artists produce sketches for each face and then humans and algorithms are used to recognize these sketches. The recognition rates of human observers and a PCA-based algorithm showed that the artist styles have significant effect on recognition rates of both humans and PCA classifier. Furthermore, averaging several sketches of a target face results in improvements in recognition rates of both parties; and finally, humans are reported to effectively use tonalities and features such as hair style, and when these features are removed (or unavailable) their performance is significantly affected.

Figure 2.8: Examples of forensic sketch-photo pairs used in [Klare et al., 2011] which is the only work that has tested on forensic sketches, instead of exact sketches. (Two left columns) Two pairs of good quality forensic sketches and the corresponding photographs, and (two right columns) two pairs of poor quality forensic sketches and the corresponding photographs.

2.1.2 Matching Forensic Sketches

Although the above algorithms are proposed to address the forensic sketch recognition problem, all of them have been tested on *exact sketches* (mainly from CUHK [Wang and Tang, 2009] and IIIT-D [Bhatt et al., 2010] databases, examples in figures 2.1,2.2, 2.3, 2.4, 2.5, 2.6, and 2.7) with significant similarities to their target faces (including exactly similar facial component shape, illumination and shading, skin texture, and even hairstyle). A recent study [Choi et al., 2012] showed an astonishing recognition rate of 85.22% only using the hair regions of the sketches and photos. This test reveals that these sketch databases cannot represent real forensic sketches, and therefore their reported accuracies and applicability are questionable. An additional test in Choi et al. [2012] reported that an off-the-shelf face photo matcher that uses merely shape and edges can outperform most of the currently proposed FSRs, even without training [Choi et al., 2012].

In contrast, as figure 2.8 illustrates, a real forensic sketch from current eye-

witness testimonies is very likely to be significantly different, due to problems such as verbal over-shadowing, perception biases, piecewise reconstruction, etc. Note that in exact sketches, the artists tries to produce a sketch as close as possible to a given target face and therefore these sketches contain considerable point-to-point matching geometry, shading, hair style, and small facial details which increases the recognition rates for both human and algorithms. On the other hand, in creating forensic sketches, an eyewitness cannot memorize a target face with detailed information, and moreover, the memorized information cannot be fully delivered to the police by current approaches (i.e. verbal description, or composite face development). This argument is tested in [Klare et al., 2011], in which authors have employed a fusion of SIFT features and multi-scale local binary patterns to recognize exact sketches as well as some forensic sketches (shown in figure 2.8). In the testing phase, in addition to exact sketches, 159 forensic sketches were used with their corresponding photograph of the subject who was later identified by the law enforcement agencies. All of these sketches were drawn by forensic sketch artists working with the eyewitnesses who provided verbal descriptions of the culprit. The interesting result of this work is that this algorithm reported to have 99.47% accuracy in matching exact sketches (the highest accuracy on exact sketches), but when tested on forensic sketches, its accuracy dramatically decreased to 16.33% (rank-1) and remained less than 33% even in rank-50. This work also tested a recent face recognition algorithm[1] (as a representative for state-of-art face recognition algorithms) on matching forensic sketches, with its accuracy reported to be as low as 2.04% and 8.16% in rank-1 and rank-50 respectively. Results of these two tests in [Choi et al., 2012] and [Klare et al., 2011] confirm our argument for dramatic differences between exact sketches and forensic

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

sketches, and that even if an algorithm can accurately recognize exact sketches, it does not necessarily provide reliable results in recognizing forensic sketches.

We can therefore conclude that current FSRs cannot reliably recognize forensic sketches, and there is a need for realistic automatic face sketch recognition.

2.2 Chapter Summary

Based on our literature review on the automatic face sketch recognition methods (FSRs) in this chapter, we showed that FSRs require exact sketches (i.e. having precise sketch to photo similarity) as their input, and cannot be used for recognizing forensic sketches. We showed that exact sketches that are used for performance measurement in FSRs are not proper estimations of real forensic sketches. Therefore, although previous FSRs have reported high accuracy rates on recognizing exact sketches, they are unreliable in recognizing real eyewitness sketches. Moreover, due to modality difference between forensic sketches and face photos, conventional face recognition methods cannot be applied to match forensic sketches. Table 2.1 summarizes our literature review on previously proposed FSRs. As this table also shows how all but one work have focused on recognizing exact sketches, and even the only work on recognizing forensic sketches [Klare et al., 2011] has failed to account for several biases that are added to forensic sketches. In our proposed FSR we not only assume a realistic similarity between the sketch and the target face (unlike in recognizing exact sketches), but also try to model the biases and debias the sketch before matching it to the photo database.

In the next chapter we discuss the psychological problems of currently used eyewitness testimony procedures. We show that regardless of the current

		Exact Sketches	Forensic Sketches
Photo-to-Sketch Transformation	**Pixel-Based**	[Martinez and Benavente, 1998][Tang and Wang, 2003][Tang and Wang, 2004] [hui Li et al., 2006]	-
	Patch-Based	[Liu et al., 2005][Wang and Tang, 2009]	-
Sketch-to-Photo Transformation	**Pixel-Based**	[Gao and Leung, 2002][Gao et al., 2008b]	-
	Patch-Based	[Zhong et al., 2007][Xiao et al., 2009][Gao et al., 2008a]	-
Modality Invariant	**Outline Matching**	[Pramanik and Bhattacharjee, 2012]	-
	Local Binary Pattern	[Galoogahi and Sim, 2012a][Galoogahi and Sim, 2012b]	[Klare et al., 2011][Klare and Jain, 2010]
	Multi-Resolution	[Bhatt et al., 2010]	-
Bias Estimation		-	-

Table 2.1: Summary of previously proposed face sketch recognition methods (input vs. method).

FSR problems, as these FSRs are ultimately supposed to recognize results from eyewitness testimony procedures (ETPs), they are doomed to inherit the unreliability of ETPs.

Chapter 3

Psychological Challenges of Eyewitness Testimony Procedures

Psychological research, going back at least 100 years with Munsterberg's seminal book On the Witness Stand [Munsterberg, 1927], up to more recent works such as [Loftus, 1979, Loftus et al., 1978, Loftus, 2005], [Cutler and Penrod, 1995], [Wells, 1993, Wells et al., 1998, 2006], and [Clark and Godfrey, 2009], have demonstrated the frailties of memory and the influence of suggestion, leaving no doubt that eyewitness testimony procedures (ETPs) are significantly flawed and unreliable, and therefore any automatic face sketch recognition methods (FSRs) that use their results. In this section we discuss the psychological challenges to conduct an unbiased and non-harmful eyewitness testimony that produces reliable results. Here we list the studied psychological challenges with a brief summary of supporting works, with the main purpose of showing that the human memory (unlike common beliefs) neither stores memories like a tape recorder (i.e. with full and *real* details), nor retains and retrieves these memories like a tape (i.e. almost completely unchanged). These are the challenges that each ETP and FSR should address for being reliable.

3.1 General Memory Limitations

General memory limitations refer to nonspecific failures to store or retain information. For example, stress, exposure duration, and retention interval can significantly affect the amount of details stored in the memory and the recognition ability of an eyewitness [Clark and Godfrey, 2009]. It may seem intuitive that eyewitnesses should be less accurate when they have shorter time to observe the perpetrator, make their observations under stressful conditions, or make identifications after long delays, but these predictions have, in some cases, met with counter-intuitive data and controversy [Read, 1995, Clark and Godfrey, 2009] (which we do not presume to resolve here). The important consideration about the general memory limitations is that (1) while they can strongly affect the details remembered by the eyewitness, these limitations are rooted in (both) the innate nature of the human memory and the observation conditions and therefore cannot be controlled or alleviated; and (2) on the other hand regarding these limitations as parts of the eyewitness sketch recognition problem is vital for a proper solution, (which are in many cases ignored as we review the automatic sketch recognition literature).

3.2 Biased Instructions

In many (but not all) jurisdictions, police present eyewitnesses with very standard instructions prior to a show-up or lineup. The instructions often have two key components: (1) that the perpetrator may or may not be in the lineup, and (2) that the eyewitness is not obligated to pick anyone. Such instructions are considered to be unbiased with respect to the perpetrator's presence in the lineup and the responses that eyewitnesses may give. By contrast, the instructions are considered to be biased if they state or imply that the perpetrator is

in the lineup or fail to acknowledge that none of the above is an appropriate response. Such biased instructions presumably increase the eyewitness's willingness to make identification, due either to a lowering of the decision criterion or to a change from a more stringent to a more lenient decision strategy. These shifts should, and do, lead to increases in correct, as well as false identifications [Clark and Davey, 2005]. Moreover, instructions can also implant ideas in the eyewitness' mind that can significantly distort their memory of the face or event [Loftus, 2005, Bernstein and Loftus, 2009]. Although the literature suggested testing several criteria by a psychologist to distinguish between true memories and implanted memories, it should be noted that in most of the cases, the eyewitness cannot distinguish between these two types of memories by himself, and these testing criteria can rarely be applied for face memories.

3.3 Piecewise Reconstruction

It seems that the human brain has a holistic (mainly configural) approach for facial information gathering and encoding, and therefore, the human performance in recognizing faces strongly depends on the configural information of the face, rather than the piecemeal information [Young et al., 1987, Sinha and Poggio, 2002, Jarudi and Sinha, 2003, Jacques and Rossion, 2009, Jones and Bartlett, 2009]. On the contrary, current methods of target face reconstruction, particularly composite sketches, are performed piecewise, and this divergence from configural to piecemeal can result in divergence of the final reconstructed face from the target face [Sinha et al., 2006a, Frowd et al., 2008]. An example of human poor performance in face recognition using a piecemeal approach can be observed in facial composite generation task which includes choosing the best matching facial features from a large collection of images of

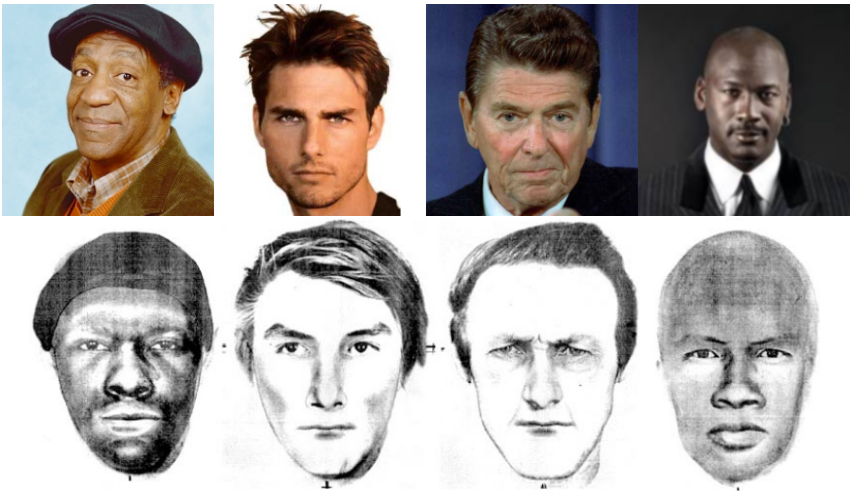

Figure 3.1: Four facial composites generated by an skilled IdentiKit operator. The individuals depicted are all famous celebrities. Degradation of recognition here highlights the problems of using a piecemeal approach in constructing and recognizing faces (from [Sinha et al., 2006b]).

disembodied features and then assemble them to reach a reasonable likeliness to a target face. The mismatch between this piecemeal strategy and the more holistic facial encoding scheme that may actually be used by the brain can lead to problems in the quality of reconstructions as shown in figure 3.1.

3.4 Memory Alteration: Post-event Information

Works by Loftus and her colleagues [Loftus et al., 1978, Loftus, 1979, 2005] have shown that the memory for a given event can be influenced or distorted by exposure to post event information. In eyewitness identification, the most straightforward example of the post event information is the exposure of the eyewitness to the suspect after the staged crime and before the lineup. For example, eyewitnesses may be presented with intervening lineups or mugshots prior to the critical lineup identification task. Several experiments have shown that if the post event information is correct, the eyewitness's accuracy in response to subsequent questions will increase, but if the post event information

is incorrect, the eyewitness's accuracy will decrease [Loftus et al., 1978, Deffenbacher et al., 2006, Clark and Godfrey, 2009].

3.5 Memory Alteration: Viewing Similar Faces

Although the encoded memory of the faces are usually thought to be relatively stable, a recent studies such as Carbon et al. [Carbon et al., 2007] showed that the memory of a face (both configural and local information of the face) can be easily distorted using similar faces. Interestingly, this distortion is even visible for familiar faces which are seen many more times than unfamiliar faces, their encoded memory is different from unfamiliar faces [Megreya and Burton, 2006], and were thought to be more stable [Burton et al., 2005].

We can divide similarity into two categories, namely, similarity of the suspect and the perpetrator, and similarity of the lineup foils to the perpetrator. The first category, suspect and perpetrator similarity is the similarity of the perpetrator as he appeared at the time of the crime to the guilty suspect (i.e. perpetrator) as he appears at the time of the identification; and the similarity of the innocent suspect to the perpetrator. In both cases, results are rather straightforward: If the perpetrator changes his appearance, the correct identification rate decreases and the risk of the wrongful identification of the innocent suspect (a.k.a. innocence risk) increases [Read, 1995, Pozzulo and Marciniak, 2006]. In addition, the more the innocent suspect is similar to the perpetrator, the higher the false identification and the innocence risks are [Clark and Tunnicliff, 2001].

The second similarity category is based on the intuition that if foils can be easily ruled out, the increase in the likelihood of the correct identification of the suspect is the result of the lineup composition, rather than of the eyewitness's

memory. Foil similarity manipulations is instantiated in a variety of ways, but primarily by including foils who mismatch one or several elements of the perpetrator's description (e.g. based on a mugshot match prior to the lineup, or verbal description). In studies where less similar foils were included in the lineup, both correct and false identification rates have increased, with the increase in false identification rates was larger than the increase in correct identification rates [Clark and Godfrey, 2009]. On the other hand, in studies where foils were similarity to the suspect, the risk of incorrect identification increased too. The increase is mainly because when the foils are similar to the suspect, they will be similar to the guilty suspect in both target-present and target-absent lineups (also known as the backfire effect) [Clark and Tunnicliff, 2001].

3.6 Memory Alteration: Verbal Overshadowing

As mentioned, the very first step of all of the current eyewitness testimony procedures is the eyewitness's verbal description about the appearance of the target face. In addition to all memory alteration problems, the verbal description itself degrades the eyewitness' visual memory as well as the "recognizability" of the target face for the eyewitness [Schooler and Engstler-Schooler, 1990, Dodson et al., 1997]. This phenomenon, known as verbal overshadowing, occurs primarily when a principally non-verbal process is disrupted by a task which involves verbalization [Melcher and Schooler, 1996, Dodson et al., 1997], and can reduce the face recognition accuracy down by 50% [Schooler and Engstler-Schooler, 1990, Dodson et al., 1997]. Adding the verbal overshadowing effect to the rest of the memory alteration effects can clearly show the unreliability of current eyewitness testimony procedures.

3.7 Memory Alteration: Mental Norm Biases

In all current eyewitness testimony procedures, in the very first step, the eyewitness provides a verbal description of the appearance of the target face (as we here focus on faces). However, this description is not based on the real appearance of the face, but on the eyewitness' own mental norm, face perception bias, and personal interpretation of the information [Bartlett, 1932, Treadway and McCloskey, 1987]. The theory of reconstructive memory suggests that people store information in the way that is consistent with their norms and previous learning. In other words, humans store a memory by trying to fit the information into a known schema. These schemata can distort a memory so that it fits in the person's existing knowledge and norms, and therefore the memory becomes a personal interpretation, rather than the true information. The effect of personal information can render the eyewitness testimony unreliable [Bartlett, 1932, Treadway and McCloskey, 1987].

For the faces, it is suggested that a face model based on the normal distribution of previously viewed faces (normal face) are used as a basis for memorizing (and then recognizing) faces. Each face is compared with this normal face, and stored based on its differences from the normal distribution [Unnikrishnan, 2009]. When differences of a group of faces are too similar, they are regarded as belonging to the same identity. A well-known example of this norm based perception, motorization and recognition is known as the Other Race Effect, in which a person perceives faces from a different race, as all very similar, and hard to memorize and recognize [Lindsay et al., 1991, O'Toole et al., 1991, Valentine, 1991, Levin, 2000, Furl et al., 2002, Jia et al., 2004, McKone et al., 2007, Ren et al., 2009].

Another less attended alteration due to the mental norms happens when the eyewitness is verbally describing his/her mental image of the target face.

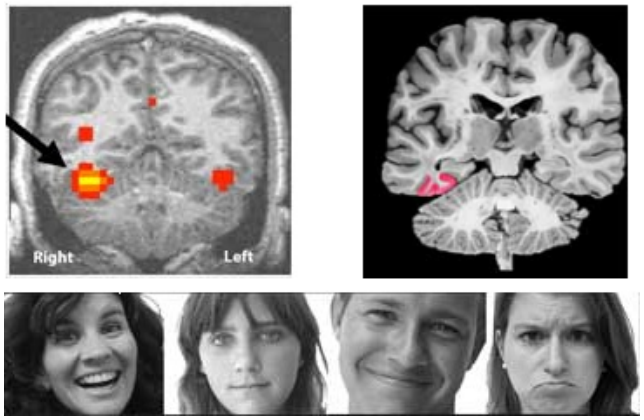

Figure 3.2: The location in the brain that is responsive to faces in typical individuals. This region, called the "Fusiform Face Area" (FFA) is located in a particular location in the temporal lobe called fusiform gyrus and is shown in this functional activation map. Although both sides of the brain are commonly active in response to faces, it is the right side that is usually more active in response to faces (note radiological convention where left and right are reversed in the image). The image on the right of the picture is of the human brain, post mortem, where the fusiform face area is colored in pink. (Image from [Pierce et al., 2001])

It should be noted that this description is based on the eyewitness's norms, but the audiences (police officers or artists) perceive the description based on their own norms and because of this difference in the norms, perceptions would be different.

3.8 Choosing a Psychological Framework

A well-designed face sketch recognition method requires proper understanding of the face perception, remembering, and recognition processes in humans. There are works such as [Sinha et al., 2006a] that present a discreet set of properties of face recognition in humans, but what is required here is a framework that can be used as a basis for our computational models. Therefore, in this section we briefly review the related psychological frameworks.

Human long-term memory holds information about objects, events, and

affective evaluations [Bower et al., 1994] and this encoded representation of objects facilitate their recognition and is a prerequisite for selecting adequate actions [Baddeley, 1990]. The "Fusiform Face Area" (FFA) region, located in fusiform gyrus (Figure 3.2) is a special part of the brain that is believed to handle face perception [Diamond and Carey, 1986, Manjunath et al., 1992, Lades et al., 1993, Wiskott and von der Malsburg, 1996, L. Wiskott and von der Malsburg, 1997, Gauthier et al., 1999, McCandliss et al., 2003, Sinha et al., 2006a, Starrfelt and Gerlach, 2007, Hansen and Atkinson, 2010]. Studies have suggested the use of some sort of face representations in the FFA region that is to some extent illumination [Braje et al., 1998] and view angle [Wallis and Bulthoff, 2001] invariant. But the face representation in the brain should be even more robust as the same face may also appear differently because of changes in expression, hairstyle, age, and speech accompanying movements [Bruce and Langton, 1994, Leder, 2005]. Another interesting aspect of face perception is the speed of face detection and recognition in humans that indicates an abstract and efficient representation of faces in the human visual system [Thorpe et al., 1996]. The representation of the faces is also seem to be different for unfamiliar and familiar faces which makes the recognition of familiar faces significantly more robust than unfamiliar faces [Brooks and Kemp, 2007]. The effect of piecemeal versus holistic images of the faces also have shown that face processing in humans is dominantly holistic, with causes low accuracies and identity hallucinations when face components are concatenated in a piecemeal manner [Young et al., 1985, 1987, Fraser et al., 1990, Sadr, 2002, Sadr et al., 2003]. Several frameworks have been proposed in the literature for face processing in the human visual system to explain the above findings and other properties of face processing. We review these most important suggested frameworks in this section to select the best framework and to use it as the

basis of our computational methods.

There are several frameworks suggested for the underlying mechanisms for the face processing tasks in the human visual system, and here we try to choose the best one to be the basis of our computational algorithms. We search for the most holistic framework in terms of ability to explain more aspects of the human visual system, and the one also backed by better experiments and evidence.

3.8.1 Norm-Based vs. Exemplar-Based Models

Two predominant frameworks suggested for face recognition are "norm-based" and "exemplar-based", both assuming that faces are encoded as vectors in a multi-dimensional "face-space", with the prototypical face located at the center of the face space [Valentine, 1991, Unnikrishnan, 2012]. Each unique measurement which contributes to the construction of a realistic and recognizable face is defined as a separate dimension in the face space. Based on this definition of the space, for example, inter-pupillary distance can be plotted on a two-dimensional graph against nose length, and by adding mouth width a third dimension is created, and so on. Then each face is defined as multi-dimensional point in the face space. Most computer based methods require 250 points or more to construct a good quality line drawing of a recognizable face [Perrett et al., 1994], so their 'face-space' has at least 250 dimensions, each of which has to be precisely measured to correctly place a given face in this multidimensional volume. However, the speed of face recognition [Benson and Perrett, 1991, Bruce and Langton, 1994, Barraclough and Perrett, 2011, Fraser et al., 1990], and the required neural processes [Haxby, 2000, Grill-Spector et al., 2004, Sinha et al., 2006a, Gillam et al., 2009, Jacques and Rossion, 2009], as well as experiments on familiar face recognition in humans [Ellis et al., 1979,

Carbon and Leder, 2005, Carbon et al., 2007, Gillam et al., 2009, Jones and Bartlett, 2009, Schmalzl et al., 2008] argue for a mental representation with lower dimension and more invariance.

The norm-based framework compares individual faces with a prototypical norm, abstracted from all or at least a subset of the faces in the face space; whereas the exemplar-based framework assumes that a norm is not abstracted, but all specific faces are stored and remembered only in relation to their nearest neighbors in the multidimensional face-space [Valentine, 1991, Leopold et al., 2001, 2005, 2006].

Palmer [Palmer, 1975] defined faces in the memory as instances of a perceptual category that abstract information about the prototypical values (or central tendency) of the relevant facial dimensions. Such information would be specific to face processing. Fodor [Fodor, 1983] proposes a similar view of face processing in his suggestion, except that not all seen faces are stored in the memory, but only "favorite candidates", for an eccentric stimulus domain-that is, a domain "whose perceptual analysis requires information that is highly specific to the domain in question." (Fodor [1983], pp. 51-52).

On the other hand, Bruce [Bruce and Langton, 1994] assumed that invariant structural information is abstracted from faces, allowing for their recognition despite many changes. These representations are thought to be based on the component features of the face, as well as on their configuration [Carbon and Leder, 2005, Leder, 2005, Carbon et al., 2007].

From the norm-based point of viewer, a number of theories suggest that face recognition in humans occurs by reference to a prototype [Benson and Perrett, 1991]. For example, some authors [Goldstein and Chance, 1980, Bruce, 1986, Bruce and Young, 1986, Valentine, 1991] have suggested that faces may be encoded by reference to a schematic representation, which emerges as a result of

a person's experience with faces over a lifetime [Schwaninger et al., 2003]. All the works from the norm-based framework (which are more recent) provide experimental results and/or justification against the exemplar-based framework. Therefore, based on our survey in the psychological literature, we reject the exemplar-based frameworks and focus on the norm-based frameworks.

3.8.2 Average Face Model

Among the suggested frameworks in the norm-based category, Average Face Model is one the well-known models. Burton et al. [Burton et al., 2005] presented three experiments in which they compare performances of human observers as well as Principal Component Analysis (PCA) algorithm on recognizing original images of celebrities (i.e. familiar faces), versus an averaged image of each of these identities. Their results show significant improvements in performances of both humans and the PCA algorithm, in recognition of the average face. The average face in their study is the average of all face images from an identity I, when each image is morphed to the average face shape of I (i.e. averaging both shape and texture of the face images of I, illustrated in figure 3.3). These experiments also showed robust recognition over some extents of illumination and pose variations. The Average Face model is able to explain some extent of the rapid face processing, and the effect of face familiarity in improving face recognition performance.

The Average Face Model has also been used on several face related automatic methods addressing face detection (e.g. [Iwata et al., 2002, Chen et al., 2009]), gender detection (e.g. [Guo et al., 2010]), 2D face recognition (e.g. [Chandrasiri et al., 2007]), and 3D face recognition (e.g. [Alyuz et al., 2007]), with reporting improvements over previous methods, particularly PCA-based algorithms. Based on these works, the main advantage of Average Face Model

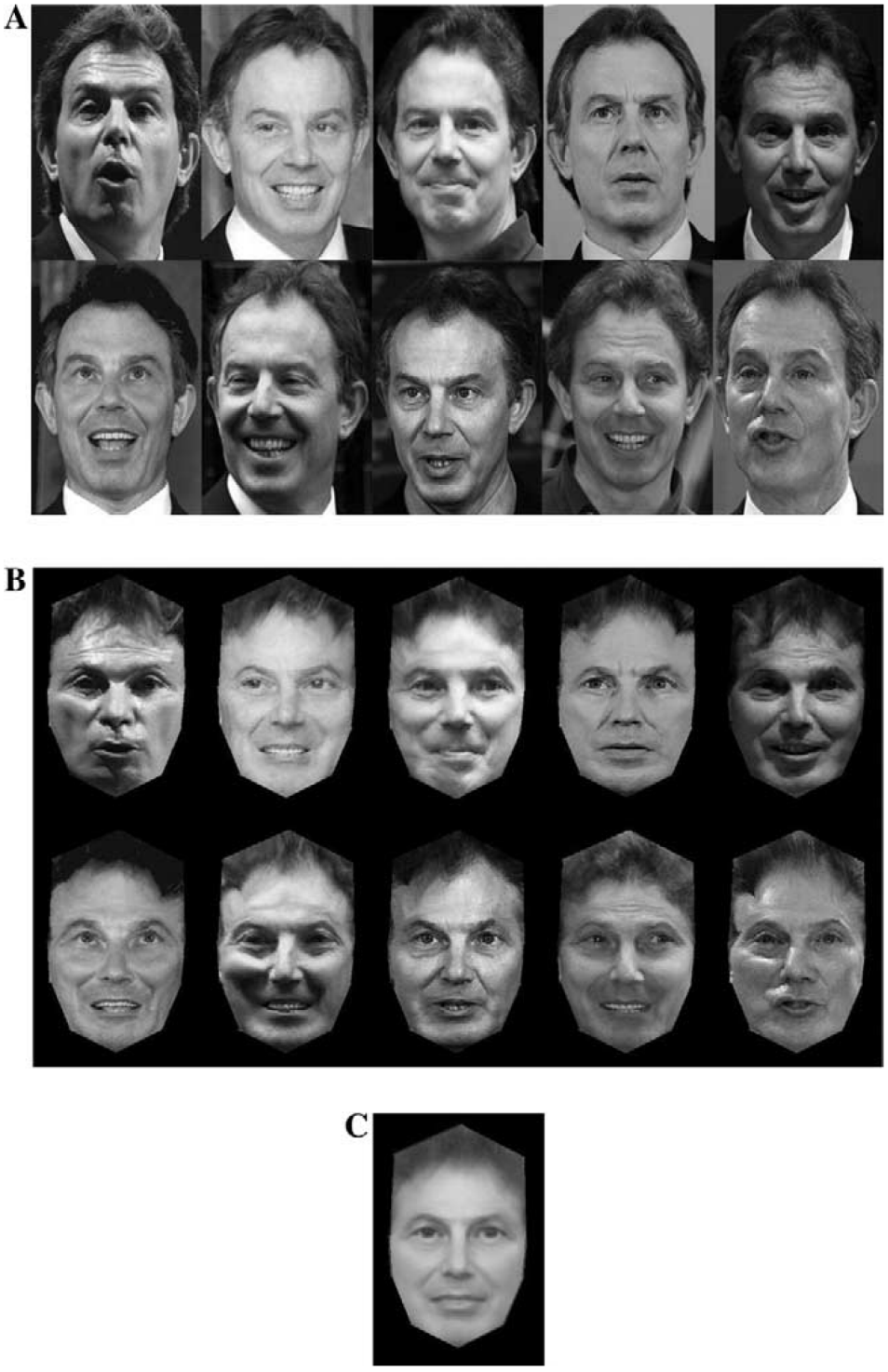

Figure 3.3: The two averaging steps in the Average Face Model over ten images of Tony Blair [Burton et al., 2005]. (A) Shows original images. (B) Shows results of morphing each of these images to a standard shape. (C) Shows the image-average of these shape-standardized images

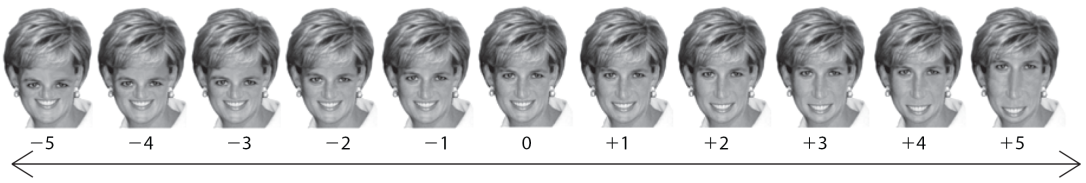

Figure 3.4: The range of configural manipulations in the experiments in Carbon et al. [2007], illustrated by the face of Princess Diana. The scale ranges from -5, up to +5 with zero indicating the original (veridical) version. Subjects selected the faces biased toward the +5 or -5 as the veridical face, after being exposed to +5 or -5 face images respectively. This biased selection indicates a change in the face representation for familiar faces which short exposure to biased stimuli.

for the machine algorithms seems to be the simplicity of the process (simply averaging shape and then texture) and representation of each identity with a single average image.

3.8.3 Exception Report Model

Despite the support of several experimental results, the Average Face Model is recently questioned by several authors (see [Deng et al., 2008, Carbon and Leder, 2005, Carbon et al., 2007, Unnikrishnan, 2009]). One of the important result is presented in [Carbon et al., 2007] which cast doubt upon the belief about central representation of faces. Many psychological frameworks including the Average Face Model assumed that the underlying representation of a face is unlikely to be modified in the course of single incidents [Bruce et al., 1991]. Specially, in recognizing a familiar face, it is a common assumption that incoming perceptual information must be matched against representations of faces stored in memory, which are accumulated over time [Bruce and Young, 1998]. Theories of memory, and particularly Average Face Model, often implicitly claim that these stored representations are both stable and accurate, containing the essential information in a face that allows for its robust recognition (see, e.g., [Bruce and Langton, 1994]). Nevertheless, such a representation

has to be flexible enough to integrate new information that might help one to recognize the most recent appearance of the face. For example, recognizing familiar faces in spite of short-term or long-term changes is probably most efficiently done by integrating these changes into the representation of that person's face. Such an integrating mechanism should also apply to ongoing long-term changes of a face during a human's life span, especially the shifting of the facial configuration from a baby face to a matured appearance. This mechanism should also be capable of integrating short-term changes. For example, the current mental and physical status may alter the facial appearance as much as would hairstyles and make-up. Without a rather flexible mechanism that integrates these types of changes, perceiver's recognition performances would be suboptimal. From an evolutionary point of view, this would be a functional disadvantage due to lacking adaptation. On the other hand, it is also important that (face) representations are sufficiently stable and rigid to allow for reliable recognition. If representations are adapting too rigorously toward recently perceived information, such an over-adaptive mechanism could also be disastrous for a recognition system. Thus, there must be a clever balance between both poles of stability and flexibility, rather than just averaging the faces [Carbon et al., 2007]. Supporting this idea, recent experimental results show that human subjects' judgments on the veridicality of highly familiar faces were strongly influenced by recent visual inputs (i.e. face images) [Leder, 2005, Carbon and Leder, 2005, Carbon et al., 2007]. While the representations for familiar faces are thought to be highly reliable (e.g. in Average Face Model [Bruce et al., 1999, Burton et al., 2005]), experiments in Carbon et al. [2007] showed that after participants were exposed to configurally manipulated versions of familiar faces (see Figure 3.4), they chose a wrong image as the original (veridical) image of the familiar face, and their bias was toward the direction of

manipulated stimuli. This adaptation effect (face identity aftereffect [Leopold et al., 2005]) was demonstrated even 24 hours after exposure to the stimuli. These experiments suggest that the underlying representation for faces could not be based on simple pictorial grounds, an episodic memory effect, or a simple averaging mechanism [Carbon et al., 2007].

In addition to Carbon's experiments, both exemplar-based and norm-based frameworks are uneconomical because they suggest neural processes that have to execute 250 feature-by-feature comparisons of the index face with a standard (either the prototypical norm, or a face that serves as the standard in a certain region of face-space to which the index face seems to belong) [Unnikrishnan, 2012]. To address this shortcoming, a modified norm-based framework, the Exception Report Model (ERM) is recently suggested by Unnikrishnan [Unnikrishnan, 2009, 2012] for face recognition in humans, in which attention is focused exclusively on features that differ significantly from the central representation of all faces, named as the Modal Face. The Modal Face is a normal representation that represents the distribution of values in each of the dimensions in the face space, with the norm defined as all values between the 5th and 95th percentile, regardless of the distribution of the metric (i.e. Gaussian or not). The unusual (exceptional, or deviations from norm) features are then the small number of features which fall into the 10% out of the norm region percentile and the human brain can rapidly and effortlessly recognize the target face by focusing on the exceptional features present in the face. While previous frameworks of face recognition require as many as 200 to 250 features to characterize a face, ERM focuses attention exclusively on features that are significantly different from the average or Modal Face. Thus, only a few ($<$10) unusual features are required to characterize the individuality of a given face. ERM can be quicker and require less mental effort, because the

neural processes employed by this framework are more economical than those employed by previous frameworks of face recognition, which require exhaustive and highly accurate multiple feature-by-feature comparisons between different faces. ERM is probably the underlying mechanism that plays a role in perceiving, remembering and recognizing faces [Unnikrishnan, 2012] as it explains the agility of humans in face recognition tasks (even in crowds), the Other-race effect (for the Other-race effect see [Furl et al., 2002]), and even the mate selection preference in humans [Unnikrishnan, 2012].

In addition to the ability of ERM in explaining face-related behavior in humans, several automatic face recognition methods inspired by the ERM [Hansen and Atkinson, 2010, Nejati and Sim, 2011, Nejati et al., 2011, 2012] have shown the effectiveness of this framework to be used for machines. We therefore choose the ERM over other frameworks to be used as the basis of our understanding of how humans perform face perception, memorization, and recognition.

In addition to employing ERM concepts for eyewitness face sketch recognition, the ability of ERM in explaining the human visual system for face recognition also motivated us to test the applicability of ERM concepts to other visual recognition tasks, such as ear image recognition. Therefore, in addition to our proposal for ETP and FSR, we provide separate sections to propose and test an adapted version of ERM concept for ear image recognition (see Sections 4.4 and 5.4 respectively).

3.9 Chapter Summary

Many challenges in eyewitness sketch recognition lies in the eyewitness testimony procedures. In this chapter we reviewed psychological problems studied

in the literature. Based on our reviews in this section we conclude that human memory is fragile, malleable, and susceptible to suggestion, and current eyewitness testimony procedures, unwittingly, change the memory of the target face in the eyewitness' mind, resulting in production of unreliable reconstructions of the target face. An important effect of the unreliability of these procedures is that any method that uses results of these procedures inherits this unreliability. Despite this, these eyewitness testimony procedures are still being widely used by police departments all over the world (mainly due to the lack of viable alternative methods).

In addition, to acquire a proper understanding of how the human visual system works, we searched through the psychological frameworks suggested for the human perception, and chose the Exception Report Model (ERM) that seemed the best, based on supporting psychological findings. We use this framework as the basis of our automatic models in next chapters.

Based on our literature review in the previous two chapters, we conclude these main points:

1. Based on our reviews on previously proposed face sketch recognition methods (FSRs), we conclude that exact sketches are not a proper estimation of forensic sketches, and performance results on exact sketches cannot provide enough information about performance on forensic sketches.

2. As a result, although previous FSRs have reported high accuracy rates on recognizing exact sketches, as tests show, they are unreliable in recognizing real eyewitness sketches.

3. Based on our previous reviews on psychological challenges in eyewitness testimony procedures, forensic sketches are not also good representations of target faces.

4. Based on our reviews on psychological frameworks proposed for face processing in the human visual system, we selected the Exception Report Model (ERM) as the basis for our computational models.

   (a) ERM suggests that humans use a norm-based face perception process that focuses on the facial features which are deviated from norm, for face perception, memorization, and recognition.

   (b) In our computational methods we only analyze the deviations in face shape, for the sake of simplicity.

Based on the above points, we conclude that more realistic designs are required for automatic face sketch recognition to meet the real case needs. In the next chapter we present our new perspective on the eyewitness testimony procedure and face sketch recognition to address both psychological challenges and automatic matching challenges.

Chapter 4

Reshaping Eyewitness Face Sketch Recognition: The Use of Non-artistic Sketches

Law enforcement agencies seems to have no choice to continue using the flawed ETPs, as there is no proper alternatives to be used when an eyewitness is the only clue to solve a case. On the other hand, the existing challenges facing the eyewitness testimony procedures (ETPs) are so extensive that have made some psychologists argue against any use of ETPs in the courts [Yarmey, 1997]. For example, how can any ETP be used while at the very first step, verbal description of the target face, causes verbal overshadowing and reduce the recognition ability down to 50%? In this chapter we try to answer to this question, as well as the need for an FSR that can robustly recognize the resulting sketch. Our goal is to more faithfully reproduce the mental image in the eyewitness's mind, something which existing methods cannot do.

We discussed the problems of both eyewitness testimony procedures and automatic face sketch recognition methods in the previous chapter and pre-

sented the current gaps in both ETPs and FSRs:

1. Current eyewitness testimony procedures (ETPs) are unreliable. This unreliability also propagates to any process which uses the outcome of these ETPs (i.e. forensic sketches) including face sketch recognition algorithms (FSRs), and law enforcing procedures (Chapter 3).

2. Current automatic face sketch recognition methods (FSRs) are designed to recognize exact sketches, and cannot handle forensic sketches (Chapter 2).

One should note that the problem of face sketch recognition has two main parts: (1) The eyewitness testimony procedure (ETP) in which a sketch is created, (2) the FSR to recognize the resulting sketch from the ETP. Without a reliable ETP like the ones that are currently used in the first part, the second part also provides unreliable results. In this chapter we therefore propose a coupled ETP-FSR to address the problem of face sketch recognition, by considering both of the above gaps together:

1. ETP Part: We propose to employ a modified ETP procedure based on the concepts of ERM model (our preferred psychological model from Section 3.8), using non-artistic sketches as the medium to avoid the disadvantages of current ETPs.

2. FSR Part: We propose an accompanying FSR for recognizing non-artistic sketches by accounting for individual differences among eyewitnesses, and relying on more realistic assumptions of sketch-target face similarity.

4.1 Proposed Eyewitness Testimony Procedure

Now consider the case of an eyewitness who has seen a target face regarding a criminal case. In our ETP, instead of asking the eyewitness to verbally describe the appearance of the target face, we ask him/her to draw a sketch of his memory of the target face, with whatever drawing skills he/she has, that includes the facial component outlines and facial marks (e.g. wrinkles, moles, and scars). Using this "non-artistic" eyewitness sketch that we name the Main Sketch, we avoid verbal overshadowing, memory degradation due to questions, illusive post-event information, viewing other faces, etc., and also piecewise reconstruction:

Verbal Overshadowing: We do not ask the eyewitness for a verbal description.

Biased Questions: We do not ask any question except for drawing the sketch

Implanted Ideas: As we avoid verbal contact and using another person to draw the sketch (police artist), we avoided probable implanted ideas and introducing a third parties mental biases

Post-event Information: We do not expose the eyewitness to mug-shots or other causes post-event information problem

Viewing Similar Faces: We do not expose the eyewitness to any faces during creation of the sketch. When acquiring the drawing profile, we assure the eyewitness that the exposed faces do not have any relation with the target face.

Piecewise Reconstruction: We avoid reconstructing the mental image based on showing separate or accumulating facial components

Figure 4.1: Example of non-artistic sketches and their related target face

Therefore the Main Sketch is more likely to be drawn based on genuine memory of the target face, in contrast with traditional forensic sketches which are drawn based on eyewitness' verbal description.

However, as figure 4.1 illustrates, the result of this ETP is a non-artistic face sketch, which is a crude, noisy, and biased representation of the target face, and cannot be used directly for photo matching. We should therefore process this crude representation of the target face to find the identity-specific features, before we can match it against the database of photos. So although using directly developed non-artistic sketches we avoid many psychology-related problems of the ETP, we have also significantly increased the difference between the sketch and photo modalities, in comparison to exact sketches. To help decrease this modality gap between non-artistic sketches and photos, we also ask the eye-

witness to provide two more types of information. First we ask for additional information about the target face including skin color, iris color, hair color, estimated age, race, and gender (also practiced in real scenarios) which we call *Ancillary Information*, and for each category of the Ancillary Information, we provide a set of predefined values from which the eyewitness can choose: we ask the eyewitness to choose the skin color from Fitzpatrick Scale color pallet (very fair, white, beige, beige with a brown tint, dark brown, black) [Fitzpatrick, 1975], the iris color from Martin–Schultz scale color pallet (gray, blue, green, brown, dark brown, black, red) [Piquet-Thepot, 1968], and the hair color from Fischer–Saller scale color pallet (brown, black, blond, auburn, red, gray/white) [Daniel, 1978]. For estimated age, we group ages in groups of 5 years (e.g. 1-5, 6-10, 11-15...). For the race, we group races into Caucasian, American Indian, Latino, African, Middle Eastern, Indian, and East Asian, and finally for the gender we have male and female.

In addition to the Ancillary Information, we obtain a drawing profile for each eyewitness by asking him/her to sketch a set of known face photos. This drawing profile is a vital key for processing the Main Sketch, and matching this non-artistic sketch to the database of photos as this drawing profile contains the samples of eyewitness' face perception and face drawing biases, which together we call them the *sketching bias*. Based on these samples from the sketching bias we can estimate the sketching bias and remove it from the Main Sketch, to reach a purer representation of the target face, for a more reliable match to the photo database.

The drawing profile is a set of samples from the eyewitness' sketching bias, and based on these samples we try to estimate the sketching bias. Thus the higher the number of these samples are, the more accurate our estimation would be. However, as factors such as anxiety and fatigue disturb the qual-

ity of drawings for the drawing profile, we have to find a minimum required number of sketches for the drawing profile. In addition, as drawing faces is a process related to face perception, it is likely to be affected by factors such as gender and race of both the eyewitness and the target face. It is show that men are more accurate in recognizing female faces Barrett and O'Toole [2009], Bindemann et al. [2009], and the other-race effect O'Toole et al. [1991], Furl et al. [2002], Jia et al. [2004] is a well-known effect in humans' ability in face recognition. Therefore, these parameters should be assessed in search for an ideal number of faces in the drawing profile, as well as the types of faces chosen to be drawn by the eyewitness. In out experiments, presented in Chapter 5 we show that at least 7 to 9 face sketches are required in the drawing profile, for an accurate enough estimation of the sketching bias. A protocol is also required to reach an optimum selection for the faces to be drawn by the eyewitness. However, designing such a protocol requires several psychological experiments which are out of the scope of this thesis.

Thus at the end of our ETP we have three outputs, namely, the Main Sketch (crudely representing the target face), the Ancillary Information (categorical description of the target face), and the drawing profile (describing the eyewitness' sketching bias). Using these three types of information about the target face we try to close the gap between the sketch and photo modalities and match the Main Sketch against our database of face photos.

We continue this chapter by describing our proposed FSR that focuses on the use of these non-artistic sketches to find the target face.

4.2 Proposed Face Sketch Recognition

Finding the correct match of the Main Sketch in the photo database requires a "difference function" that can handle the modality difference between the non-artistic sketches and the face images. We define our difference function based on the Main Sketch and the Ancillary Information. However, the Main Sketch is a crude representation of the target face and even not all parts of this crude representation contains the same amount of information about the target face. There are parts affected by noise, perceptual bias, or ignorant drawing (result of low importance associated with a component) among the parts that represent the actual appearance of the target face. Therefore, we should first (1) estimate the sketching bias, (2) weight each part of the debiased sketch based on the amount of information it bears; and finally, (3) define proper point correspondence to be able to compare a sketch outline and a photo outline. We describe these three requirements for a proper matching between non-artistic sketches and face photos in sections 4.2.1,4.2.2, and 4.2.3.

4.2.1 Sketching Bias Estimation and Removal

In the FSR part, we should first remove the sketching bias from the Main Sketch, to reach a debiased sketch that represents "what the eyewitness meant" by the Main Sketch. We define the sketching bias, as two types of distortions in the non-artistic sketches: First, the distortions from noisy drawing (e.g. shaky hand due to lack of drawing skills) and face completion (i.e. the lines which are only to complete the shape of the face in the sketch, not related to the target face appearance) and second the distortions from mental bias and drawing bias. By mental bias we refer to a function of complex processes such as memory, perceptual bias, and feature representation style that adds biases

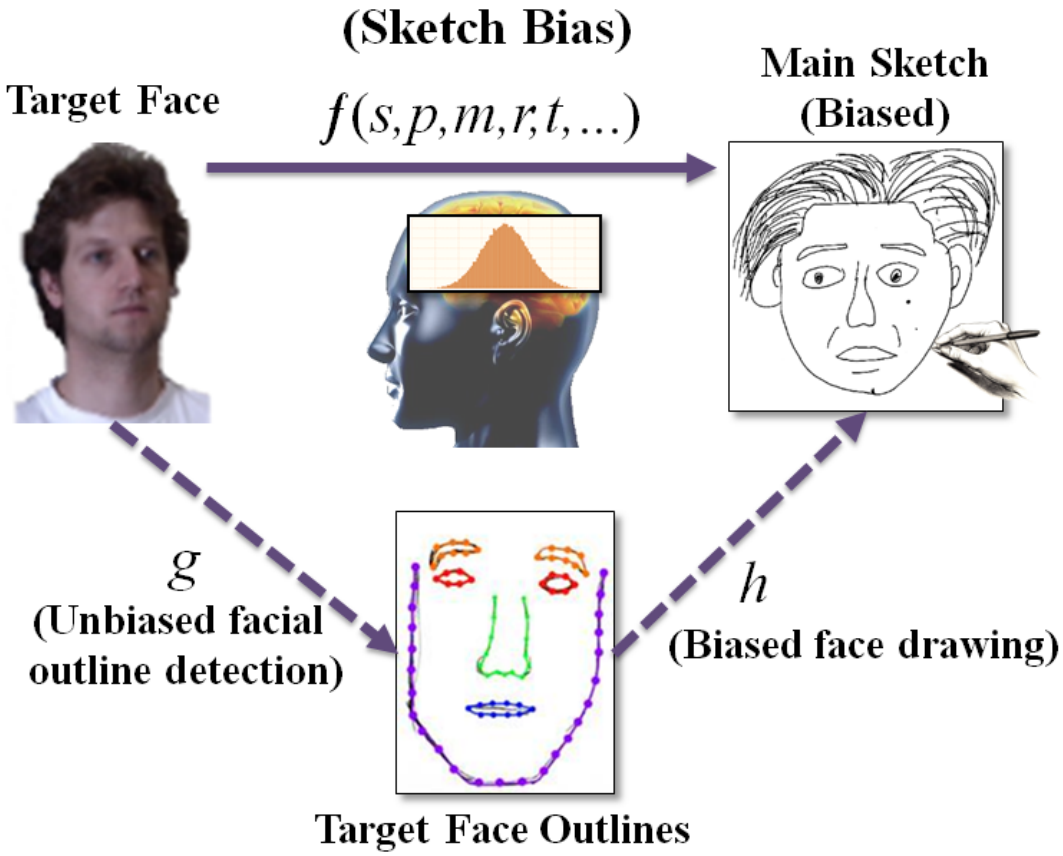

Figure 4.2: Pictorial representation of the process of creating the biased Main Sketch, divided into two steps: first the eyewitness should detect the outlines of the memory of the target face (process g), and then draw the non-artistic sketch based on these outlines (process h). While the mental bias would be added during g and the drawing bias would be added during h, for easier estimation of the sketching bias (combination of mental and drawing biases), we can safely assume a perfect (unbiased) g, and a biased h, where the sketching bias is entirely added during h. Using this assumption we propose our debiasing method in Section 4.2.1.

to the process of creating a non-artistic face sketch from a viewed face. The first type of distortion can be different from one part of the sketch to another, and also can be different from one face to another, based on how the eyewitness remembers the face and the how well s/he can draw (drawing bias). But the second type follows an individual pattern. Based on psychological studies and particularly Exception Report Model [Unnikrishnan, 2009, 2012] the mental face perception, memorization, and recognition biases are formed by a norm-based process in the human visual system, based on the faces seen during the life experiences. As the life experiences of each individual are unique, the mental bias of each individual is also unique. However, as we take an engineering approach to this problem, we try to model the eyewitness' mental bias, and then remove it from the Main Sketch, to reach a relatively unbiased representation of the target face. As mentioned in previous chapter, we are only using face shape as our feature here.

We define the sketching bias as a point-to-point transformation function f, mapping the facial component outlines in a photo, ϕ, to the facial component outlines in it respective sketch, φ. Based on psychological studies we can safely assume that f is (at least) a function of drawing bias, face perception strategy, face memorization, face recognition ability, facial feature visual importance, living environment, gender, and race. But the function f can also be decomposed into two steps, namely, facial component outline detection in the photo (g), and drawing the detected photo outline using a pen to create the sketch (h), illustrated in Figure 4.2:

$$f = h \circ g$$

The two components of the sketching, mental and drawing biases, take place at g (when analyzing the facial features) and h (when drawing the facial features) respectively. But in an engineering point of view, we can easily

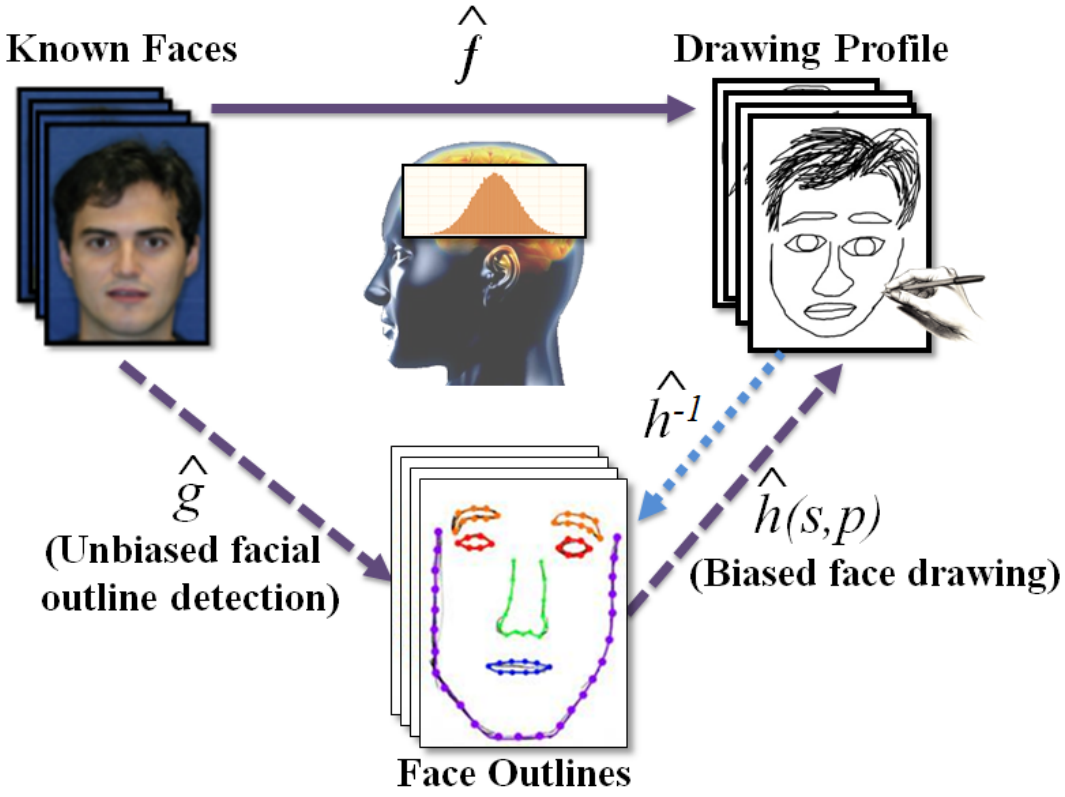

Figure 4.3: Pictorial representation of using the eyewitness drawing profile to debias the Main Sketch. By assuming unbiased g, we estimate g using a facial component detection algorithm ($\hat{g}$), and then learn $\hat{h}^{-1}$, using drawing profile as training samples. We then use $\hat{h}^{-1}$ to debias the Main Sketch. Note that $\hat{h}(s,p)$ is an estimation of original $h(s,p,m,r,t,...)$, when only face perception (p) and drawing bias (s) are considered (for the sake of simplicity).

assume that the entire sketching bias (mental and drawing) take place at h, leaving the g bias free[1]. Using this view, we need to estimate h^{-1} to debias the Main Sketch.

In order to estimate h^{-1}, we use the eyewitness' drawing profile as a set of samples of h, to learn h^{-1}. But we should mention that in this estimation approach, we have assumed that the process of drawing profile $\hat{h}$ (sketches drawn while looking at the photo) is a good estimation of the process of drawing the Main Sketch h (sketches drawn from memory) - Thus we are estimating $\hat{h}^{-1}$:

$$\begin{aligned} \hat{f} &= \hat{h} \circ \hat{g} \\ \hat{h}(\phi) &= s(p(\phi)) + e \end{aligned}$$

where p indicates face perception, s drawing bias, and e is the noise. Although our assumption of $h \approx \hat{h}$ is not true in cases involving memory impairments, this is reasonable assumption for simplification of the problem at this stage.

In order to estimate $\hat{h}^{-1}$, we use the drawing profile as a set of samples of photo outlines ϕ, and sketch outlines$\varphi = \hat{h}(\phi)$. Figure 4.3 illustrates the use of eyewitness' drawing profile, to estimate and remove the sketching bias using $\hat{h}^{-1}$.

It should be noted that based on psychological studies (e.g. [Unnikrishnan, 2012]), the mental bias does not affect all of the facial components in the same way. In addition, the eyewitness' drawing bias for each part of the face may be different. We therefore estimate $\hat{h}^{-1}$ in a piecewise manner. To be able to learn the transformation function $\hat{h}^{-1}$ between sketch-photo pairs, we

[1]See appendix for justifications.

represent each photo, based on the outlines of its facial components (simulating mental function g). Using a simple Active Shape Model (ASM) to provide the data points of the facial component outlines, we fit 16 piecewise cubic Hermite splines [Fritsch and Carlson, 1980] to the outlines of 7 facial components, namely, the eyes, the eyebrows, the nose, the mouth, and the jaw-line (2 splines to each eye, 2 to each eyebrow, 2 to the mouth, 3 to the nose, and 3 to the jaw-line). Monotone interpolation can be accomplished using Cubic Hermite spline with the tangents m_i modified to ensure the monotonicity of the resulting Hermite spline. We select the interpolating tangents for each data point of the ASM, based on the Fritsch–Carlson method [Fritsch and Carlson, 1980]. Let the data points be (x_k, y_k) for $k = 1, ..., n$

1. Compute the slopes of the secant lines between successive points:

$$\Delta_k = \frac{y_{k+1} - y_k}{x_{k+1} - x_k}$$

   for $k = 1, ..., n-1$

2. Initialize the tangents at every data point as the average of the secants,

$$m_k = \frac{\Delta_{k-1} + \Delta_k}{2}$$

   for $k = 2, ..., n-1$. For the endpoints use one-sided differences:

$$m_1 = \Delta_1 \,\&\, m_n = \Delta_{n-1}$$

3. For $k = 1, ..., n-1$, if $\Delta_k = 0$ (if two successive $y_k = y_k + 1$ are equal), then set $m_k = m_{k+1} = 0$ (as the spline connecting these points must be flat to preserve monotonicity) and ignore steps 4 and 5 for those k.

4. Let $\alpha_k = m_k/\Delta_k$ and $\beta_k = m_{k+1}/\Delta_k$. If α or β are computed to be less than zero, then the input data points are not strictly monotone. In that case, piecewise monotone curves can still be generated by choosing $m_k = m_{k+1} = 0$, although global strict monotonicity is not possible.

5. To prevent overshoot and ensure monotonicity, the function

$$\phi(\alpha, \beta) = \alpha - \frac{(2\alpha + \beta - 3)^2}{3(\alpha + \beta - 2)}$$

must have a value greater than (or equal to, if monotonicity need to be strict) zero. One simple way to satisfy this constraint is to restrict the magnitude of vector (α_k, β_k) to a circle of radius 3. That is if $\alpha_k^2 + \beta_k^2 > 9$, then set $m_k = \tau_k \alpha_k \Delta_k$ and $m_{k+1} = \tau_k \beta_k \Delta_k$ where $\tau_k = \frac{3}{\sqrt{\alpha_k^2 + \beta_k^2}}$

Note that only one pass of the algorithm is required. After this pre-processing we have to evaluate the interpolated spline to read cubic Hermite spline, using the data x_k, y_k, and m_k for $k = 1, ..., n$.

To evaluate at x, find the smallest value larger than x, x_{upper}, and the largest value smaller than x, x_{lower}, among x_k such that $x_{lower} \leq x \leq x_{upper}$. Calculate

$$h = x_{upper} - x_{lower} \,\&\, t = \frac{x - x_{lower}}{h}$$

then the interpolant is

$$f_{interpolated}(x) = y_{lower} h_{00}(t) + h m_{lower} h_{10}(t) + h m_{upper} h_{11}(t)$$

where h_{ii} are the basis functions for the cubic Hermite spline.

We then divide each of these 16 splines into four parts (quarter splines), and re-sample each of the parts with 25 equally distributed data points. Finally, we scale and rotate all sketches to the same size and angle based on the position of

eye centers. Note that different splines require different normalizations, based on their sizes, because points do not cover the same amount of area from one spline to another. For example, the eyes are much smaller than the jaw-line, and as they are both sampled with the same number of points, the average data point in the jaw-line spline represents a larger area than the average data point in the eye spline. Therefore, we have to also normalize each spline s_i based on its axis lengths $diam_i$:

$$\bar{s_i} = \frac{s_j}{diam_i})$$

We perform the same point representation and normalization on the sketches. Finally, we can pair the data points from a sketch, to the data points from its related photo outline, as the sampling points to learn $\hat{h}^{-1}$. We use Support Vector Machine Regressors (SVRs) with Radial Basis Function (RBF) kernel, to fit the $\hat{h}^{-1}$function in a piecewise manner. We use a separate SVR for each of the quarter splines (i.e. 4 mapping function for each spline, 64 mapping function for the entire sketch).

4.2.2 Weighting Sketch Outlines

At this step, we have a debiased sketch φ, an estimation of photo ϕ outlines from the Main Sketch outlines, and this debiased sketch should be used to find the target face. However, not all parts of the debiased sketch φ contains the same amount of information about the face photo ϕ. There are parts affected by noise or ignorant drawing (result of low importance associated with a component) and there are parts representing the actual outlines of the face ϕ. Therefore, the estimated target face outlines are also associated with different reliability, which requires a notion of reliability (or importance) weight for each part of the sketch. We define this weight based on the estimated amount of

information that a particular part bears.

In order to estimate the level of information embedded in each part of the sketch, we here use the concepts introduced in Exception Report Model (ERM) [Unnikrishnan, 2009]. The ERM is the psychological framework for face processing in the human visual system which we chose based on our literature survey on psychological frameworks for face processing in humans (Section 3.8). The ERM represents our base understanding of how the human visual system perceives, memorizes, and recognizes faces. ERM suggests that the human brain employs a modified norm based model in which attention is focused exclusively on the deviations from norm (exceptionality) of each features, for rapid and effortless recognition of the target face. The exceptionality of a feature is determined by comparison to the norm representation of that feature in the distribution of different values of that feature. All of the norm representations of all facial features form a Modal Face in the mind of a person, and this Modal Face is the center of face space in the person's mind. However, it is very important to remember that the distributions of the Modal Face, and therefore, the perception of normal and exceptional can be different from one person to another. This is because the Modal Face is formed based on seen faces, and two individuals have been exposed to different faces throughout their life experiences [Unnikrishnan, 2012]. This difference in Modal Face explains the other-race effect, in which individuals from one race have difficulty recognizing people from another race [Furl et al., 2002]. Based on the ERM framework, when a person perceived a face from another (rarely seen) race, an unusually many number of facial features of other race faces are perceived as exceptional, and therefore, the faces from that race are all perceived equally exceptional. Similar phenomena has also been verified in the information theory [Shannon and Weaver, 1962] in which more common verbs carry less infor-

mation, while more information is encoded in less common verbs. Note that in the information theory, the notion of commonality can be defined by the normal distribution and therefore, a less common verb is more exceptional.

Similarly in our application for the eyewitness face sketch recognition, we conclude that two different eyewitnesses may draw the same face differently, not only due to their different drawing bias, but also due to differences in their Modal Faces, that make them perceive different features as normal or exceptional. We should therefore weight each part of the debiased sketch, based on the deviations from the norm in that part, assuming that this deviation represents the level of attention the eyewitness has had on that part and thus an estimation of the level of information the sketch part bears.

We define the level of exceptionality of a feature value as the distance of that feature value to the mean value of that feature, normalized by the standard deviation of that feature. Based on this weighting strategy we weight each point in the sketch or photo:

1. We estimate a normal probability density function (PDF) for each of the 5 points in the fitted splines (described in Section 4.2.1), based on the distribution of their values in the entire Multi-PIE dataset [Gross et al., 2008]. Note that the values are normalized before calculating the PDF.

2. Now we define the weight (exceptionality) of point k of debiased sketch φ_i, as the distance of that point from the mean of its corresponding PDF, normalized by the standard deviation of that PDF:

$$w_k = \sqrt{(\frac{x_k - \mu_{x,k}}{\sigma_{x,k}})^2 + (\frac{y_k - \mu_{y,k}}{\sigma_{y,k}})^2}$$

$$S = W^T \varphi$$

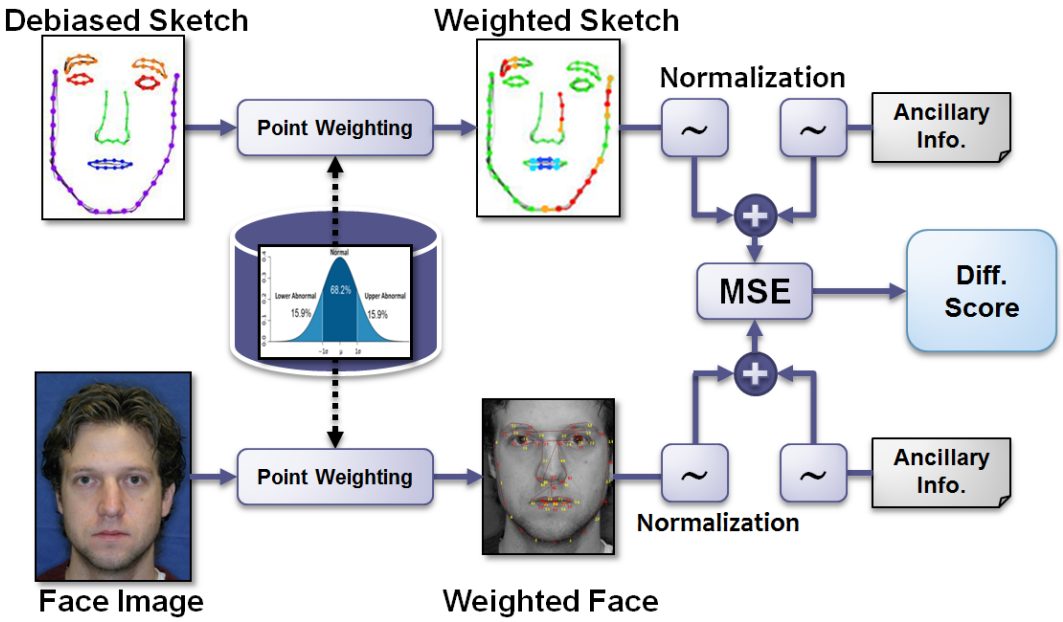

Figure 4.4: Weighting and matching the debiased sketch: different parts of the debiased sketch and photo are weighted based on their deviations from database norm (exceptionality). These points are then normalized and concatenated with their normalized Ancillary Information. Final difference score is calculated based on minimized squared errors to find the closest photo to the sketch.

where w_k, x_k, and y_k are the weight, X and$Y$ coordinates of point k; and $(\mu_{x,k}, \sigma_{x,k})$, $(\mu_{y,k}, \sigma_{y,k})$ are mean and sigma values of the corresponding X coordinate, Y coordinate PDFs; W_i is the vector of all weights associated with all points of sketch φ and finally, S is the weighted sketch. Similar weighting is applied to photo points to reach weighted photo P.

4.2.3 Recognizing the Debiased Sketch

We now have a weighted debiased sketch which is our estimation of the target shape and the Ancillary Information that is the categorical information of the target face. We also use the relative distance between facial components as the third piece of information to find the target face. We calculate the relative distances between the facial components as the distance between their centers

from the center of each of the eyes (as we scale and rotate each sketch based on eye positions).

Because these three pieces of information are from different natures and may differently contribute to the matching process, we should use a proper difference measurement for each feature to be able to combine them into a single difference score. The difference measurement for each of the skin color, iris color, hair color, race, and gender features between a sketch and a photo is assigned heuristically based on common understanding of these features. The difference measurement for each of these features is represented by an $n \times n$ matrix, where n is the number of possible values for that feature. Each cell C_{ij} of this matrix indicates the difference between i^{th} and j^{th} values of that feature, in terms of integers in range $[0, n]$. The difference measurement of facial marks is indicated by minimum squared error (MSE) between facial mark pixels and edge pixels in the same photo region. We define the same photo region as the same-sized rectangle as the bounding box of the facial mark, centered at the same distance from the nearest facial component (see figure 4.5). To avoid noise, we re-size each photo to 128×128 pixels, apply Sobel edge detection, and re-size back the resulting edge image to its normalized size, and then calculate MSE of the facial mark pixels and the edge pixels. Finally, we measure the difference between the sketch and photo points based on Euclidean distance.

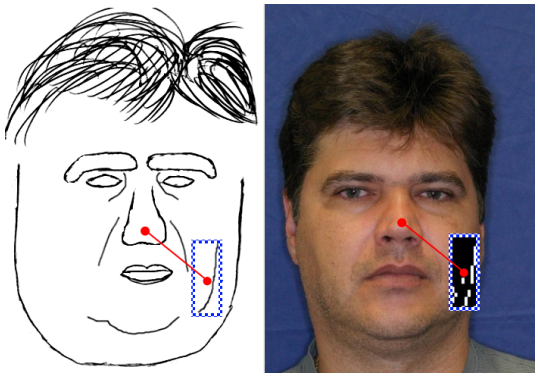

Figure 4.5: Comparing sketch facial marks and image edge pixels, in the same region (i.e. a same size rectangle as the bounding box of the sketch facial mark, centered at the same distance from the nearest facial component).

After calculating all feature differences:

$$\Delta_{point} = (\sum_{i=1}^{16}\sum_{j=1}^{4}\sum_{k=1}^{25} Euc(s_k, p_k)) \quad (4.1)$$

$$\Delta_{mark} = \sum diff_{mark}(S_i, P_j) \quad (4.2)$$

$$\Delta_{skin} = C_{ij,skin} \quad (4.3)$$

$$\vdots$$

$$\Delta_{gender} = C_{ij,gender} \quad (4.4)$$

$$\mathbf{\Delta}(S,P) = \{\Delta_{point}, \Delta_{mark}, \Delta_{skin}, ..., \Delta_{gender}\} \quad (4.5)$$

where Δ_i is difference between sketch and photo regarding feature i, Euc is the Euclidean distance,$diff$ is the feature-related difference function Finally, $\mathbf{\Delta}(S,P)$ is the set of all feature differences between sketch S and photo P. We can now optimally combine all features by calculating the coefficient matrix A that minimizes the sum of squared error E, given a training set of sketch-photo pairs (S_i, P_i):

$$E = \sum_{i=1}^{n} \parallel A(\mathbf{\Delta}(S_i, P_i) \parallel^2$$
$$\Upsilon = \parallel A\mathbf{\Delta}(S,P) \parallel$$

where Υ is the final difference score between Main Sketch φ and database photo ϕ (after sketching bias removal and normalization). Figure 4.4 illustrates the procedure of weighting and matching in our proposed FSR.

Given the difference score Υ, one should be able to measure the confidence of the matches to the photo database. We propose confidence measure C as measurement of the matching confidence, based on the distribution of all sketch-photo difference scores in the database:

$$C = \frac{\sigma^2}{|\mu| + \varepsilon} \tag{4.6}$$

the higher C value indicates the higher confidence of the final matchings. The intuition behind this confidence measurement is that when most of the matching score values are distributed compactly with either a large positive or a large negative mean (small σ, large μ), this indicates that the top-n faces are selected almost randomly. In this case the value of C is small. In contrast, when the difference scores show a distribution with a large variance, with some of the faces having small difference scores (large σ, small μ), this indicates that faces are ordered with a reasonable difference from each other, and therefore the matching is more reliable. In this case the value of C is large.

Using the final score as a confidence measure is not the only possible method. For example, one can use the accuracy of bias estimation in $\hat{h}^{-1}$as another confidence measure. This accuracy can be an indication of how consistent is the eyewitness' drawing pattern, in his/her drawing profile.

4.3 Improving Non-Artistic Sketch Recognition

We have described the details of our two-pronged proposal for a new ETP and an accompanying FSR. In this section we additional methods to improve our

original method. In the Experiments Section (5) we separately present results for the original concept and the improvement techniques to show their effects.

Dynamic Point Correspondence: One of the issues that can be improved in our FSR is regarding the sketch to photo point correspondence. In the previous sections, we used a hard point correspondence, based on the spline sampling index. Obviously in a crude sketch like the Main Sketch, the point correspondence is also not accurate, and therefore assuming a hard, index-based correspondence negatively affects the general accuracy. Therefore, we here propose to impose a temporal order to the sketch and use a weighted dynamic time warping algorithm to find the best point correspondence between the sketch and photo splines. We use a dynamic programming that maximizes the match between sketch points and photo points, constrained by temporal order of the points.

Multi-Distribution Weighting: Imposing temporal order provides a dynamic point correspondence that decreases the errors caused by local displacement of points, but imposing temporal order also brings forth another problem in the weighting step. Using this dynamic point matching, a point from a sketch spline, σ, can be matched to virtually any point ς in the respective photo spline; thus, the definition of "deviation from norm" becomes ambiguous in the weighting step, because the norm should be referenced to a distribution, and the distribution of the sketch point, g_σ, can be different from the distribution of photo point, g_ς. In order to alleviate this ambiguity, we have to account for differences in distributions using an updated weighting scheme, based on distribution comparison.

General-Specific Modeling: One of the most important steps in our FSR is

the estimation of sketching bias. Although given enough samples (sketch-photo pairs in drawing profile) we can robustly estimate the sketching bias, but in reality, to avoid factors such as fatigue and visual distraction, we can only acquire a few training sketch-photo pairs in the drawing profile of each eyewitness. To address the problem of having few samples, we use a general-specific estimation scheme in which we first learn the general sketch-to-photo transformation embedded in the entire sketch database, and then fine tune this transformation to a specific transformation, based on the sketch-photo pairs drawn by a particular eyewitness.

We start with the new weighting scheme, followed by dynamic point matching and then general-specific modeling.

4.3.1 Multi-Distribution Weighting

Based on the ERM concept, we assign higher weights to sketch parts with larger deviations from norm, assuming these parts bear more information from the target face and are memorized better than normal parts (see Section 4.2.2). In our framework, weight of a vector $\sigma \in \varphi$ is defined as the normalized distance of σ from its respective mean (calculated based on the database distribution), which is the inverse of the probability of σ:

$$\omega(\sigma) = 1 - P(\sigma|g)$$

where g is the associated distribution of σ.

When matching two points, σ from sketch φ ($\sigma \in \varphi$), and ς from photo ϕ ($\varsigma \in \phi$), with possibly having different distributions (while we assume normal distributions for all points), we should also account for distribution differences.

Therefore, instead of individually weighting each point (as in Section 4.2.2), we define a matching weight of the two points σ and ς as follows:

$$W(\sigma,\varsigma) \quad = \quad \frac{\omega(\sigma)\omega(\varsigma)}{D^2_{RAD}(g_\sigma, g_\varsigma)\, E^2(\sigma,\varsigma) + \varepsilon} \tag{4.7}$$

where E is the Euclidean distance; g_σ, and g_ς are the corresponding distributions of σ and ς respectively; and D_{RAD} is the Resistor-Average Distance (RAD) between the two distributions. Note that in scoring the match between two points, we account for their individual weights (inverse probability), translational distance, and distributional distance.

The Resistor-Average Distance (RAD) is originally derived from the Kullback-Leibler Divergence (KL) [Cover and Thomas, 1991] which is well-known to measure the distance between probability density functions (PDFs) and defined as:

$$D_{KL}(p \parallel q) \doteq \int p(x) log_2(\frac{p(x)}{q(x)}) dx \tag{4.8}$$

This formula follows an information theory approach to quantify how well a particular PDF $q(x)$ describes samples from another PDF $p(x)$. KL is non-negative and equal to zero iff $p(x) \equiv q(x)$, but KL is asymmetric. The asymmetrical property of KL makes it hard to use this measurement in our frameworks. Therefore, we use an symmetrical extension of KL, known as Resistor-Average Distance (RAD) [Arandjelovic and Cipolla, 2006] as a measure of dissimilarity between two probability densities, defines as:

$$D_{RAD}(p,q) = [D_{KL}(p \parallel q)^{-1} + D_{KL}(p \parallel q)^{-1}]^{-1} \tag{4.9}$$

Similar to KL, RAD is non-negative and equal to zero iff $p(x) \equiv q(x)$, but it is symmetric. It is also notable that when classes C_p and C_q are distributed

according to $p(x)$ and $q(x)$, respectively, $D_{RAD}(p,q)$ reflects the error rate of the Bayes-optimal classifier between C_p and C_q.

4.3.2 Imposing Temporal Order

Based on our newly defined point matching weight, $W(\sigma,\varsigma)$, we proceed to the details of finding point correspondence. We use weighted Dynamic Time Warping (DTW), with the weights of each point σ being $\omega(\sigma)$ to find the best match between the sketch and photo points. Given two outlines from sketch φ and photo ϕ, with points $\sigma_i \in \varphi$ and $\varsigma_j \in \phi$ to be matched, we find the corresponding points in the two outlines with the temporal order constraint, formulated as maximizing of the total matching score, $\Upsilon(\varphi,\phi)$ with constraint Eq. 4.10:

$$\begin{aligned} \Upsilon(\varphi,\phi) &= \{\sigma_{i1}:\varsigma_{j1}, \sigma_{i2}:\varsigma_{j2}, ...\sigma_{ik}:\varsigma_{jk}\} \\ s.t. \quad \forall &\quad \sigma_i,\sigma_s \in \varphi\ ,\ \varsigma_j,\varsigma_t \in \phi \\ if &\quad \sigma_i:\varsigma_j\ ,\ \sigma_s:\varsigma_t \\ then &\quad i<s \Leftrightarrow j<t \end{aligned} \tag{4.10}$$

where i,j,s and t are the point indices and $\sigma_i:\varsigma_j$ represents matching of σ_i to ς_j. The constraint in Eq. 4.10 forces a temporal order for matching points, that given four points which are matched as $\sigma_i:\varsigma_j$ and $\sigma_s:\varsigma_t$ (with $\sigma \in \varphi$ and $\varsigma \in \phi$), if the index of σ_i is smaller than the index of σ_s (σ_i is located before σ_s in the temporal order of spline sampling), then the index of ς_j should be also smaller than ς_t (ς_j should be also located before ς_t in the temporal order of spline sampling). The weighted DTW with the above constraint is presented in Algorithm 4.1, to calculate the match that maximizes the matching score

$\Upsilon(\varphi, \phi)$ between the given sketch and photo outlines, φ and ϕ.

Algorithm 4.1 is to find the best match for point $\sigma_i \in \varphi$ in sketch outline to one of the points ς_j in sketch outline ϕ, given that matches for previous points, $\sigma_1 \cdots \sigma_{i-1}$, are optimal. We use a dynamic programming approach based on Dynamic Time Warping, using Table $\gamma(\varphi, \phi)$ to store the best matches up to the current point. The best match for current point ς_i is selected based on three rules. Rule 1 is the initialization rule for the first element (point) in table γ, giving zero weight to the first point . Rule 2 is the single point matching: if there is only one point to match (i.e. either $|\varphi| = 1$ or $|\phi| = 1$), then the matching weight would be $W(\sigma_i, \varsigma_j)$ (based on Eq. 4.7). And Rule 3 is to use divide and conquer method to recursively divide the sequence of points down to reach a single point to match (based on Rule 2), and then propagate the best matches upwards, until all the sequence is matched. At each step of propagation, if the match σ_i to ς_j is the best option until current point, the new match score, $W(\sigma_i, \varsigma_j)$, is stored in Table γ. Otherwise, the point ς_j is ignored and the algorithm tries to match σ_i to one of the points ς_{j+1} to ς_n. Then, based on the propagation rule, Rule 3, the next row of the γ will be updated for the next point, σ_{i+1}. The Table γ is used to backtrack all possible combinations of points, in order of $O(n \times m)$. The last row in this table, $\gamma_{m,j}(\varphi, \phi)$ contains the accumulated match scores related to all possible matches of the last point σ_m, and therefore $\Upsilon(\phi, \varphi) = \underset{j}{max}(\gamma_{m,j}(\phi, \varphi))$ represents the best point matching of the sketch points φ and the photo points ϕ.

Given two outlines φ, ϕ, our feature vector is the maximum similarity scores for the outlines of 7 facial components (eyes, nose, mouth, eyebrows, and jawline) and the similarity scores of the relative locations of these 7 components, to the center of the eyes. Based on this feature vector, given a new pair of sketch-photo (drawn by the same person), we can label this pair as

Algorithm 4.1 Dynamic programming script to calculate the maximum matching score between feature sequencesφ and ϕ.

$\%\Upsilon(\phi, \varphi)$: Table $\gamma(\varphi, \phi)$ is an $m \times n$ table where $|\varphi| = m$ and $|\phi| = n$.
$\%i$ and j are row index and column index of table γ
$i = 0$
Loop i
    if $i <= 0$ then (% Rule 1, initialization %)
        $\gamma_{i,j}(\varphi, \phi) = 0$
    else if $i == 1$ then (% Rule 2, single point match%)
        $\gamma_{i,1...j}(\varphi, \phi) = maxW(\sigma_i, \varsigma_j)$
    else if $i <= m$ (% Rule 3, divide and conquer%)
        Loop j
        $\gamma_{i,j}(\varphi, \phi) = \max(W(\sigma_i, \varsigma_j), \max(\gamma_{i-1,j}(\varphi, \phi) + W(\sigma_i, \varsigma_j))),\ j' \in [j + 1, n]$
        $\gamma_{i+1,j}(\varphi, \phi) = \gamma_{i+1,j}(\varphi, \phi) + \gamma_{i,j}(\varphi, \phi)$
        Until $j == n$
Until $i == m$
$\Upsilon(\varphi, \phi) = \max_j(\gamma_{m,j}(\varphi, \phi))$

---

matching/non-matching, by minimizing the estimation error E:

$$E = \sum_{\sigma \in \varphi, \varsigma \in \phi} \parallel W(f'(\sigma_i), \varsigma_i) \parallel^2$$

4.3.3 General-Specific Modeling

We proposed to use eyewitness' drawing profile to estimate the sketching bias. The estimation is therefore performed based on a few number of sketch-photo samples in the drawing profile. Here we propose an improvement technique for the cases that a large dataset of drawing profiles (from other eyewitnesses) is available. We propose to improve the transformation estimation by dividing the estimation into two steps of General and Specific modeling: using the entire dataset for learning the General model and using the drawing profiles to fine-tune the General model into each eyewitness' Specific model. Using this two-step modeling, instead of estimating a direct transformation from the sketch space to photo space based on a few samples, we first estimate an

intermediate-transform from the sketch to the intermediate space, a space more similar to the photo space (than the original sketch space) using a large number of samples, and then estimate another transformation from this intermediate space to the photo space using the individual eyewitness samples.

The General model accounts for *general* sketch to photo transformation (general differences between sketch and photo modalities, discusses in several previous works e.g. [Tang and Wang, 2004, Klare et al., 2011, Kiani and Sim, 2012a,b]), assuming all eyewitnesses are the same (i.e. what is assumed by all previous works). In other words, in the General modeling we assume that there is a General transformation used by almost all eyewitnesses in the process of creating a sketch from a face photo. The Specific model then accounts for individual differences in sketching style and mental bias (described in Section 4.2.1) which is ignored by previous works. We can formulate these two transformations as:

$$\bar{\varphi}_i = \tau_i^G(\varphi_i, \{\bar{\chi}_\mathbf{i}, \bar{\mathbf{A}}_\mathbf{i}\}) \tag{4.11}$$

$$\bar{\phi}_i = \tau_i^S(\bar{\varphi}_i, \mathbf{A_i}) \tag{4.12}$$

where $\bar{\varphi}_i$ and $\bar{\phi}_i$ represent the estimation of the photo outlines based on the General model and the combination of General-Specific models, respectively; τ_i^G is the General transformation; τ_i^S is the Specific transformation for the i^{th} eyewitness; $\bar{\chi}_\mathbf{i}$ and $\bar{\mathbf{A}}_\mathbf{i}$ are all training sketch-photo pairs which exclude samples from the i^{th} eyewitness; $\chi_\mathbf{i}$ and $\mathbf{A_i}$ are then the training sketch-photo pairs drawn by the i^{th} eyewitness (i.e. the drawing profile);Ψ is the Main sketch.

We use a separate set of RBF kernel SVRs for learning the General and General-Specific transformations. We first train the first set of SVRs (Gen-

eral SVRs) on a training set sampled from the entire dataset, excluding the eyewitness' drawing profile, and use these trained General SVRs to estimate the intermediate state of the eyewitness' drawing profile. Then we train the second set of SVRs (Specific SVRs) to learn the Specific transformation from the intermediate state drawing profile, to the respective photo outlines. These Specific SVRs are then used to debias the Main Sketch which is then fed to the weighting and matching steps, to calculate the difference score.

4.4 Extended Application: Wonder Ears, Identification of Identical Twins from Ear Images

In this section we extend our framework, introduced in previous chapters for eyewitness face sketch recognition, to another application, identification based on ear images. In this application we used the same weighting scheme (based on deviation from norm), but now on the ear appearance and ear shape, to find the best matches using feature exceptionality levels.

While identical twins identification is a well-known challenge in face recognition, it seems that no work has explored automatic ear recognition for identical twin identification. Ear image recognition has been studied for years, but Iannarelli [1989] appears to be the only work mentioning the twin identification, which was performed manually. We here explore the possibility of twin identification from their ear images using a novel algorithm which focuses on exceptionalities in the ear shapes and appearances. Our algorithm is based on a recently proposed psychological model for face recognition in humans, known as Exception Report Model (ERM), which has been applied in automatic face recognition methods. We test our new approach on 39 pairs of identical twins (78 subjects), with several levels of resolution, occlusion and noise, left ear

vs. right ear training sets, and feature optimization. Our results verify the robustness and optimality of the introduced features for twin identification, and indicate the applicability of ERM to a wider range of tasks in computer vision, than only faces.

Identification based on ear images has been studied for more than two decades. Although other features such as face are more commonly used as a biometric, ear images have several advantages over such features. The ear shape does not change significantly after adulthood, its surface has a relatively uniform color distribution, it is invariant to expression, and ear images are more robust to illumination and head pose changes than features like faces [Burge and Burger, 2000]. The early studies mostly addressed the question of uniqueness of ears, although not always in a forensic context. The most well-known pioneer seems to be Iannarelli [Iannarelli, 1989], in which he performed manual identification over 10,000 ears and found no indistinguishable ears. Iannarelli's studies showed that, given correct point-to-point comparison, ear shape can be considered a biometric identifier as well as more established biometrics like face or voice. Imhofer [Imhofer, 1906] has also found 4 characteristics to uniquely distinguish a set of 500 ears, indicating that the variability between ears is large enough to assume ears as unique identification features Meijerman [2006].

On the basis of the above studies, automatic ear recognition techniques have been introduced, mostly employing methods used in other biometric fields. Eigen-ears [Saleh et al., 2007] could provide high accuracy in recognition in closely controlled conditions, otherwise, having dramatic performance reduction even with slight amounts of rotation. In order to handle rotation in ear images, Abate et al. [Abate et al., 2006] introduced a method based on Generic Fourier Descriptors which is robust to ear rotation and illumination changes.

Yan also presented a complete system [Yan and Bowyer, 2007] including automated segmentation of the ear in a profile view image and 3D shape matching for recognition under constrained conditions with specialized cameras. Ali et al. [Ali et al., 2007] presented another ear recognition method based on manually cropped 2D ear images of profile faces and performed a wavelet transform for feature extraction. Their final labels are then calculated based on Euclidean distance, with a performance similar to [Yan and Bowyer, 2007]. Bustard and Nixon [Bustard and Nixon, 2010] recently proposed an ear registration method that utilized SIFT features followed by a homography transformation, to cope with the occlusion and pose changes. The transformed images are then masked and matched using Euclidean distance. Although this work reported an impressive performance and robustness to occlusion and noise, their semi-automatic ear masking procedure occasionally fails to match correctly to the ear area. Other important approaches for automatic ear recognition are force field transformation [Hurley et al., 2005], local surface patch comparisons using range data [Chen and Bhanu, 2007], Voronoi diagram matching, neural networks, genetic algorithm [Pun and Moon, 2004], geometric feature extraction [Choras, 2004] and ICP, for 3D data [Yan and Bowyer, 2005].

Although several aspects of ear recognition have been explored, there seems to be no work on twins identification using ear images, except Iannarelli [Iannarelli, 1989] who merely performed manual matching of ear images. Therefore, we here present an approach for automatic ear recognition for identification between twin siblings. We propose our approach based on a psychological model, originally suggested for face perception in humans, known as Exception Report Model (ERM) [Unnikrishnan, 2009, 2012], described in Section 4.2.2. The concepts of our proposed ear recognition is similar to our face sketch recognition, but here applied to shape and appearance of the ears, in order to

show the extensibility of our introduced method.

Our proposed system consists of two parts, namely, ear image normalization, and feature weighting and verification. In the first part, ear normalization, we normalize ear images and obtain their shape and appearance information. In order to achieve the ear shape, we use SIFTFlow algorithm [Liu et al., 2008] to calculate a dense correspondence between the gallery ear image and a pre-defined reference ear image. Using this correspondence, we normalize the scale, rotation and illumination of the gallery image. We also use this dense correspondence as the (relative) *ear shape*. In addition to this relative shape information, we use the normalized ear pixel intensities as the *ear appearance*. In the second part, feature weighting and classification, we first weight points in the ear shape and appearance based on their level of exceptionality. Then based on the weighted feature vectors, we train a Support Vector Machine (SVM) classifier to verify whether two given ears belong to the same subject.

We evaluate our system on a dataset of 39 pairs of identical twins (78 subjects), testing ERM possibility and its robustness against five resolution levels, four occlusions levels, and four noise levels, as well as left ear versus right ear training-testing sets. We also test the ERM optimality in automatic ear recognition for both left and right ear images. These results showing high accuracy and robustness, suggest the applicability of ERM to a wider range of automated visual tasks than only faces.

In summary we introduce the following contributions in this chapter:

- We are the first to use ear biometric to identify identical twins (using the largest available twins dataset).
- We preserve and use both ear shape and appearance in our system.
- Motivated by the ERM, we introduce shape and appearance exceptional-

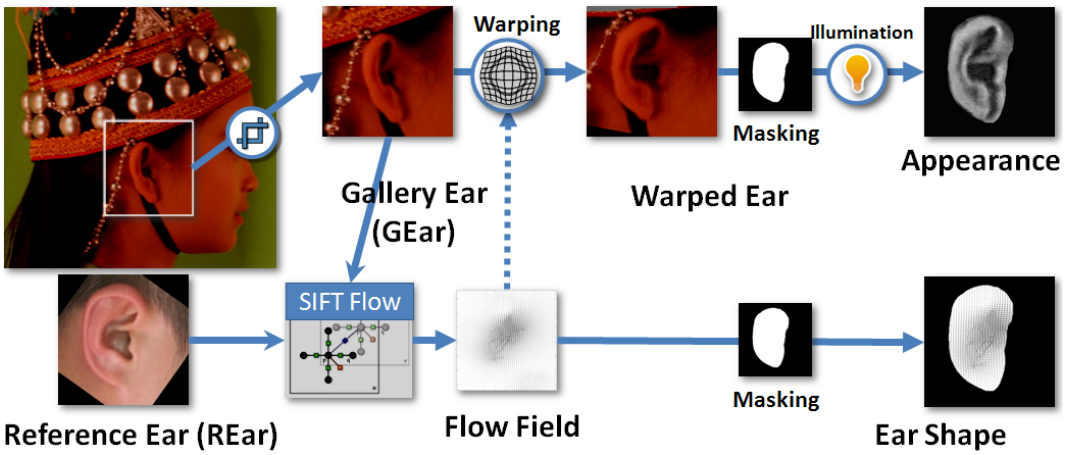

Figure 4.6: The first part of our proposed algorithm, ear normalization: We use SIFTFlow dense matching to acquire the ear flow field (relative ear shape) and then warp the gallery ear image, based on this flow field (ear appearance). Then we mask both shape and appearance and normalize the illumination of the ear appearance.

ity as a new and robust feature for ear recognition, showing the possibility and the optimality of this model.

4.4.1 Ear Recognition Method

In this section we describe the two parts of our twin ear recognition method: (1) ear normalization and (2) weighting and verification (see Fig. 4.6 and 4.7), with a brief comparison with previous methods.

4.4.2 Ear Image Normalization

The first part, ear normalization (see Fig. 4.6), is to crop the ear out of the profile view, and then normalize the rotation, scale, and illumination of this gallery ear image (GEar), based on a reference ear image (REar). It seems that all of the previous works normalized the ear image by selecting two or more fiducial points on the ear area. In previous works, *sparse* point registration has been performed both manually [Iannarelli, 1989, Faez et al., 2008] and auto-

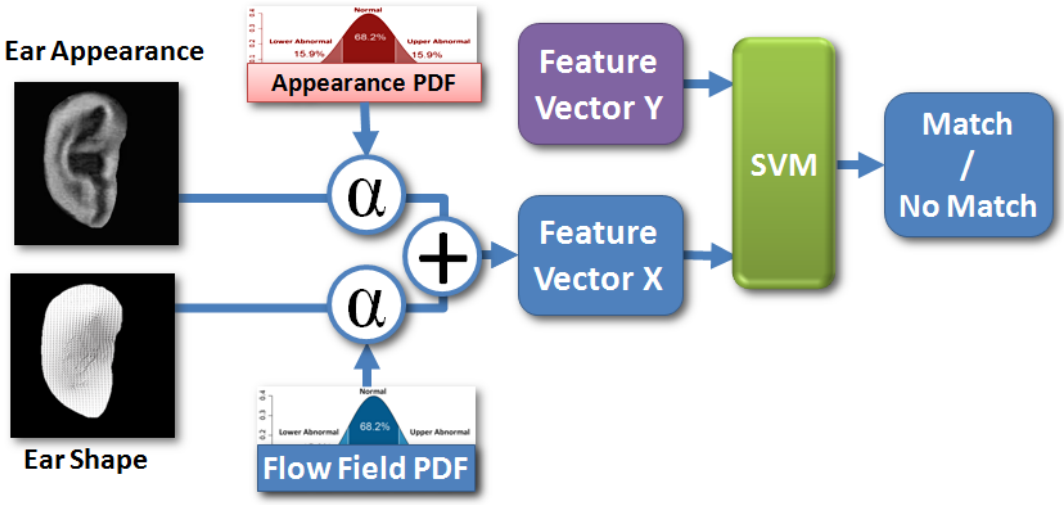

Figure 4.7: The second part of our proposed algorithm, weighting and verification: we weight the shape and appearance points based on their level of exceptionality (α), which is defined by their location in the related PDFs. We then concatenate weighted shape and appearance points into a feature vector and using SVM, we verify whether the identities of two the feature vectors are the same (match) or not (no match).

matically, using methods such as SIFT feature matching [Bustard and Nixon, 2010] and graph matching [Burge and Burger, 2000]. These points are then used to normalize the GEar based on REar coordinates. However, as the transformation function is not stored, when all ears are transformed into a single reference image coordinates, the 3D structure of the ear (i.e. ear shape) would be lost and merely the intensity values (i.e. ear appearance) would remain. In order to avoid losing the ear shape information, we calculate and store the dense correspondence between each GEar and the REar using SIFTFlow. We not only use this dense flow for the scale and rotation normalization, but also treat the flow itself as the relative shape information of each GEar. In addition to representing the shape information, a dense correspondence is particularly required in the ERM as it requires checking all points to measure the level of exceptionality. In the ERM, we weight each point based on its level of exceptionality. The level of exceptionality of a point is not known beforehand, the

only way is to measure the value of each point with respect to its related PDF. Therefore, we need to know the correspondence of each point on the ear to the REar, i.e. dense correspondence (discussed more in Section 4.4.3).

We normalize the GEar images in the following steps:

1. We loosely crop a window of 300×300 pixels out of the profile view (originally 1728×1152 pixels), around the ear-hole which is located by a simple image correlation with very rough precision (Fig. 4.6, crop symbol). The cropping is to reduce the search window of the SIFTFlow algorithm in the next step

2. We apply the SIFTFlow to calculate the dense flow field between the cropped window and the REar (Fig. 4.6, flow field). Although the cropping around the ear-hole is inaccurate, the SIFTFlow can compensate for these errors, thus our final dense point registration is accurate.

3. Based on the calculated flow field, we warp the GEar image to the REar image coordinates, thus normalizing its scale and rotation (Fig. 4.6, warping).

4. As we are only interested in the pixels corresponding to the ear in each image, we mask out the non-ear pixels in both the flow field and the warped ear (Fig. 4.6, masking). The mask is a single pre-defined binary image, manually defined for the REar image. We can use this single mask for all GEar images because the flow field is calculated based on the REar, and the GEar image is then warped to the REar coordinates.

5. The final task is then to normalize the illumination of warped image (Fig. 4.6, illumination). We use Contrast-limited adaptive histogram equalization (CLAHE) [Reza, 2004] for illumination normalization. CLAHE

operates on small regions in the image, enhancing the contrast in each region so that the histogram of the output region approximately matches a specified histogram distribution (here, the REar histogram). The neighboring regions are then combined using bilinear interpolation to eliminate artificially induced boundaries. CLAHE showed better results than global histogram equalization in our experiments.

At the end of this part, we have normalized ear shape (i.e. the masked flow field) and ear appearance (i.e. the masked, illumination normalized warped image), shown in Fig. 4.6, as the final shape and appearance.

4.4.3 Feature Weighting and Verification

The second part, feature weighting and verification, is to apply the Exception Report Model (ERM) concept to our ear recognition method. The ERM suggest that the importance of a feature has a direct relationship with the exceptionality of that feature. As we know, the more a feature is exceptional, the further the location of its value is from the mean value in the related probability density function (PDF). Therefore, the normalized distance of a feature value to the mean value can provide a reasonable measurement for the level of exceptionality of that feature.

Based on the above argument, we weight each point in the ear shape and ear appearance based on its distance to the mean value of the related PDF, in these three steps:

1. We estimate a normal probability density function (PDF) for each shape and appearance point, based on the distribution of their values in our dataset. For a shape point, we estimate a PDF based on the flow vectors of the corresponding points in all flow fields of all GEar images. For an

appearance point, we estimate a PDF based on the intensity values of the corresponding points in all ear appearances.

2. Now we define the level of exceptionality (weight) of point k, as the distance of that point from the mean, normalized by the sigma, in the corresponding PDFs:

$$w_{shape,k} = \sqrt{(\frac{x_k - \mu_{x,k}}{\sigma_{x,k}})^2 + (\frac{y_k - \mu_{y,k}}{\sigma_{y,k}})^2}$$

$$w_{app,k} = \sqrt{(\frac{int_k - \mu_{i,k}}{\sigma_{i,k}})^2}$$

where $w_{shape,k}$ is the shape weight; $w_{app,k}$ is the appearance weight; x_k and y_k are the X and$Y$ coordinates; int_k is intensity value of point k; and $(\mu_{x,k}, \sigma_{x,k})$, $(\mu_{y,k}, \sigma_{y,k})$, and $(\mu_{i,k}, \sigma_{i,k})$ are mean and sigma values of the corresponding X coordinate, Y coordinate, and intensity PDFs.

3. The final feature vector is formed by concatenating weighted shape and appearance points, to represent a GEar image (see Fig. 4.7):

$$\Gamma_i = W_{shape}^T S || W_{app}^T I$$

$$W_{shape} = \begin{bmatrix} w_{shape,1} \\ w_{shape,2} \\ \vdots \\ w_{shape,n} \end{bmatrix}, W_{app} = \begin{bmatrix} w_{app,1} \\ w_{app,2} \\ \vdots \\ w_{app,n} \end{bmatrix}$$

where Γ_i is the concatenated vectors representing GEar image i; S is the ear shape values and I is the ear intensity values.

Given a pair of weighted feature vectors, we now can train an SVM classifier to verify whether the vectors representing the two ears, belong to the same subject.

In the experiments chapter we test the robustness of our ear recognition approach under different resolution, noise, occlusion, training set, and dimensionality reduction.

4.5 Chapter Summary

In this section we presented our eyewitness testimony procedure (ETP) and its accompanied face sketch recognition method (FSR) based on non-artistic sketches, to address the gaps in previously proposed eyewitness testimony procedures and automatic face sketch recognition methods. We showed that based on non-artistic sketches drawn by the eyewitness in our ETP, we can avoid several psychological problems including verbal overshadowing, biased questions, implanted ideas, post-event information, viewing similar faces, and piecewise reconstruction. These problems were among the most important problems of currently used eyewitness testimonies that render the resulting forensic sketches unreliable. We therefore kept the mental image of the target unchanged, and our resulting non-sketch is more reliable, but includes the eyewitness' sketching bias (including drawing bias and face perception bias). In the FSR part we presented our method to debias the non-artistic sketch using eyewitness' drawing profile, and weight this debiased sketch for being matched the photo database. Unlike previous FSRs, we accounted for individual differences between eyewitness, based on a drawing profile of each eyewitness, and our weighting strategy is also based on psychological suggestions for the attention on facial features in humans. Finally we matched the weighted sketch against the photo database based on a weighted combination of facial component outlines, facial marks, skin color, race, etc.

In addition to the original framework, we proposed improvements to our

method to achieve better point correspondence by imposing temporal order, better point weighting by accounting for differences in probability distributions, and better sketching bias estimation by General-Specific modeling that uses the general sketching bias in the entire database.

It is important to note that we used drawing profile, a set of sketches drawn while looking at photos (copy-sketching), to estimate the process of drawing the Main Sketch from the memory (memory-sketching). In the next section, Experiments, we show that how closely copy-sketching and memory-sketching behave, based on our database of sketches.

Other than theoretical discussions that are presented in this chapter, we here briefly discuss possible practical problems in implementing this proposed ETP and FSR, to either substitute or modify current ETP practices.

In our proposed protocol we first ask the eyewitness to draw the target face and provide its Ancillary information, and then asked him/her to provide the drawing profile. In this protocol, drawing the main sketch should be strictly before other inputs, to avoid addition of biases to the eyewitness' memory. The selection of faces to be drawn in the drawing profile is also an important issue. These faces should be selected randomly to avoid distortion due to viewing similar faces, and an ideal set of faces for the drawing profile covers a large spectrum of facial features. Finally, it is important that the eyewitness is properly informed that the drawing profile faces are randomly selected and are not related to the target face. Drawing profile may also cause fatigue which may in turn reduce the fidelity of the sketches. The eyewitness should therefore be granted breaks between drawings that based on common practices can be done in several stages and span even over days (practiced currently by the police departments). There is also possibility of performance anxiety for the eyewitness as this practice is a new approach and the situation of

ETP may also cause anxiety by itself. Performance anxiety can be reduced using warm up stages both in drawing the main sketch and drawing profile. Sketches from warm up stages should be then either discarded or used with a low confidence level. Finally, in real cases the eyewitness may be willing to produce more sketches for the drawing profile (possibly over different days), due to the importance of the task.

In this chapter we also extended the ERM concept to propose the first reliable ear recognition method for automatic twin ear verification. Although verified manually before [Iannarelli, 1989], none of the previous works have addressed this problem. We introduced our ear recognition algorithm, using both shape and appearance of ears, and motivated by Exception Report Model (ERM), a psychological framework for the perception of faces by the brain [Unnikrishnan, 2009, 2012], which has shown good results in face recognition before (see e.g. [M. F. Hansen and Smith, 2010, Nejati and Sim, 2011, Nejati et al., 2011]). We showed that, similar to face recognition, by focusing on deviations from norm (exceptional features) in both the ear shape and appearance, we can accurately identify twins (up to 92%). In our experiments on 39 pairs of twins with different age, gender, and race, we showed the robustness of our algorithm against several variations resolution, noise, and occlusion. However, these experiments also showed the shape and appearance exceptionality features are not the same in the right and left ears. We also performed a dimensionality reduction to show that with only the top 5% features (further than 1.7σ in the PDF curve), we can achieve a fast and accurate recognition.

In conclusion, our results suggest that ears can be considered not only a powerful identification feature among regular subjects, but also among identical twins, in which many other approaches such as face recognition have a poor performance [Phillips et al., 2011]. In addition to addressing the twin iden-

tification from ear images, our work here suggests that the ERM, although originally suggested for face recognition in humans, may be applicable to a wider range object recognition problems, which may also help simulating new frameworks in human visual system studies.

In the next chapter we present our experimental results on non-artistic as well as artistic sketches, and based on these result we present improvements to our algorithms. The experiments on the application of ERM concept on ear recognition problem is also described in the next chapter.

Chapter 5

Experiments

In this chapter we test the performance of our proposed face sketch recognition method (FSR) and compare our method with previously proposed FSRs. For a fair comparison, we need to compare the recognition accuracies of the methods on non-artistic as well as artistic sketches. For the non-artistic sketches, we gathered a total of 860 sketches, and for the artistic sketches we used the CUHK artistic sketch database Wang and Tang [2009]. Additional experiments to show the effectiveness of improvement techniques are also presented in this chapter.

5.1 Data Collection

In order to test our proposed ETP-FSR solution, we have collected three datasets of non-artistic sketches, drawn based on face photos, as the test bed for our algorithm. In all datasets we asked non-artist participants to draw sketches of face photos. The face photos are selected from male individuals, without glasses or facial hair, from Multi-PIE database of faces [Gross et al., 2008]. In collection of each dataset, we instructed the participants to view the target faces on a 14" LCD display, and provide a face sketch that includes

Figure 5.1: Examples of non-artistic sketches and their respective face images from our second dataset.

facial component outlines and main facial marks (wrinkles, moles, etc.), on A4 pages using pen/pencil. The participants were sit behind a desk with the display on it, with their own preference of distance from the display (within a normal range of using desktop computers) and all participants had normal or corrected to normal vision, and could take a break between each sketch drawing to avoid fatigue.

Our first dataset consists of 50 sketches from a set of 25 male target faces from the Multi-PIE face dataset [Gross et al., 2008] (all in frontal pose and normal illumination). We asked 5 non-artist participants to provide in total of 25 sketches of the target faces. Each of the participants provided 5 sketches from 5 faces, while he/she could look at the face images, having unlimited time to deliver (similar to all previous works). We scanned and retouched these 50 sketches to enhance their contrasts and component connectivity. We

used these 50 sketches as the sketch gallery and all 249 faces in the Multi-PIE dataset (all in frontal pose and normal illumination) as the face gallery.

Our second sketch dataset is similar to the first sketch dataset, but consisting of 100 non-artistic sketches of random male target faces (without glasses or facial hair, in normal illumination) from the Multi-PIE dataset, drawn by 10 participants, while looking at the face photos, with unlimited time to deliver (similar to the first dataset). Examples of these sketches are illustrated in figure 5.1.

Our third dataset is significantly larger than the previous two datasets, consisting of 710 non-artistic sketches, drawn by 71 non-artistic participants. In addition, in this dataset we selected 30 male target faces (without glasses or facial hair) from specifically three different races, 10 target faces selected from Caucasian race, 10 from Indian race, and 10 from East Asian race. Then we asked each of the participants to draw the 10 sketches, randomly chosen from *only one of* these three races. We also included exposure time variation in this dataset, by asking participants to deliver the first 5 sketches with no time limit (while they could look at the target faces). We name these first 5 sketches, copy-sketches. For the next 3 sketches, we asked participants to first view the target faces for 10 seconds, then to solve a visual memory puzzle for 1 minute (for the purpose of visual distraction), and after this delay, they could start to sketch. We name these 3 sketches as 10-sec memory-sketches. For the final 2 sketches, we applied the same protocol as for the 10-sec memory sketches, but only allow the participants to view the target faces for 2 seconds and thus naming the 2 later sketches, 2-sec memory-sketches. For the memory-sketches, we also instructed participants to mark the regions that they completely remember, partially remember, and completely unable to remember. Figure 5.2 illustrates an example for copy-sketches, 10-sec memory-sketches, and 2-sec

Figure 5.2: Examples of non-artistic sketches and their respective face images from the same drawer, in our third dataset. From left to right: an example of Immediate Sketch, Long Exposure Sketch, and Short Exposure Sketch from the same participant.

memory-sketches. The motivation behind the limited viewing time was to reveal the facial components which are more important for each participant. Knowing that the face recognition is a significantly fast process in humans [Sinha et al., 2006a], we can assume that as we reduce the viewing time, only the facial components would be stored in the memory that are more important for recognition, and this set of important components can be different from one participant to another.

For our third dataset of sketches, stored additional information about the participant including age, gender, race, and drawing skill level. The participants in this dataset were aged between 19 and 41 (mean 26.98, standard deviation 3.95), both male and female, and from Caucasian, East Asian, Indian, and Middle Eastern races. Some examples of original scan copies of the 10 drawings from a single participant is illustrated in Figure 5.3.

For all of our three datasets, in addition to sketches, we asked each participant to provide Ancillary Information about the target face including skin

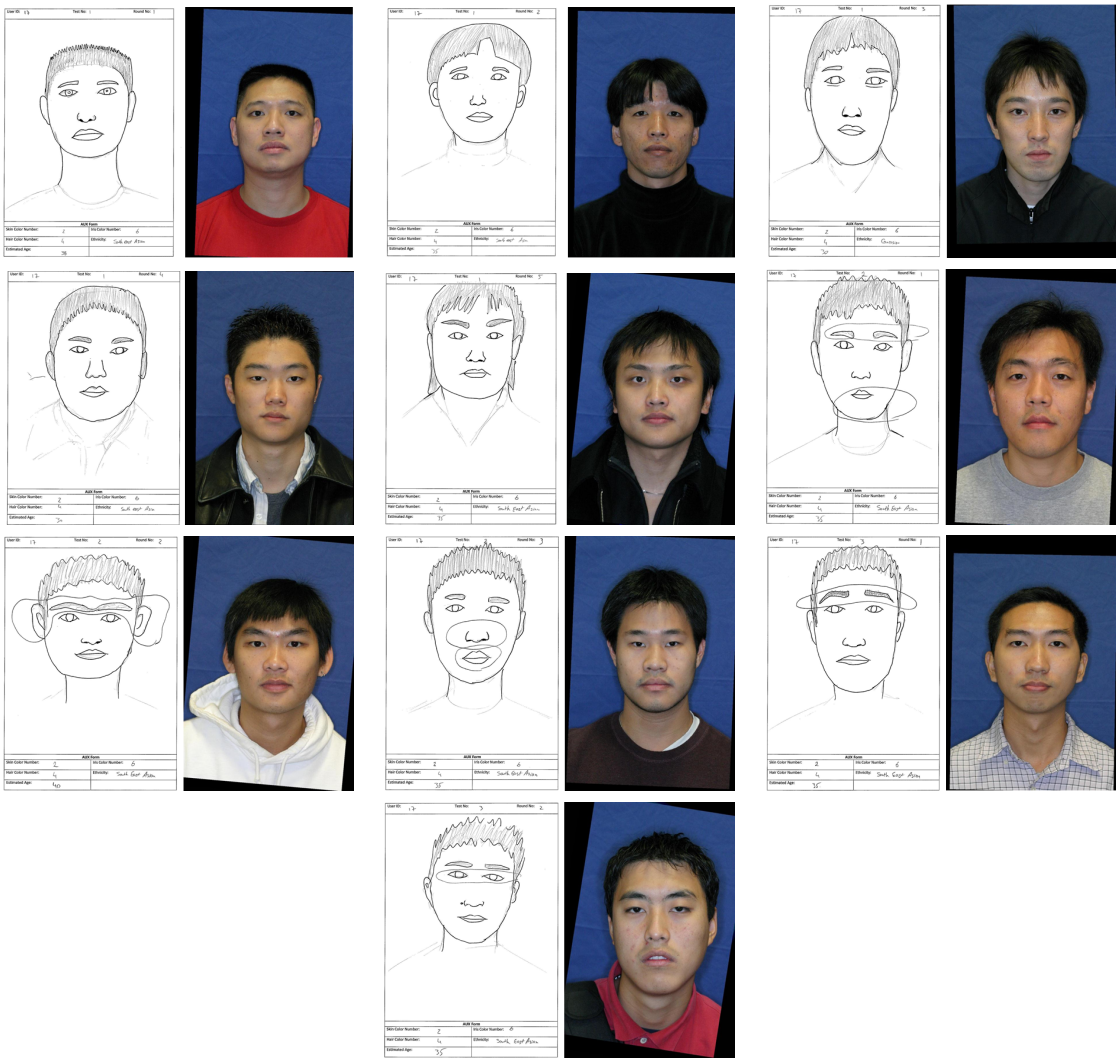

Figure 5.3: Examples of original scanned documents from one non-artistic drawer in our third dataset, drawing East Asian race. From top left to bottom right, stages 1 to 10 (first 5 sketches while looking at the image, 6 to 8 sketches of images viewed for 10 seconds and drawn after 1 minute delay, and 9 and 10 are sketches of images viewed for 2 second after 1 minute delay).

Dataset	No. Sketches	No. Face Images	No. Drawers	Images of Different Races	Limited Exposure	Time Delay	Ancillary Info
#1	50	25	5	No	No	No	**Yes**
#2	100	30	10	**Yes**	No	No	**Yes**
#3	**710**	**30**	**71**	**Yes**	**Yes**	**Yes**	**Yes**

Table 5.1: Summary of our three face sketch datasets.

color, iris color, hair color, estimated age, race, and gender. For each category of the Ancillary Information, we provided a set of predefined classes from which the participant can choose:

- Skin color from Fitzpatrick Scale color pallet (very fair, white, beige, beige with a brown tint, dark brown, black) [Fitzpatrick, 1975],
- Iris color from Martin–Schultz scale color pallet (gray, blue, green, brown, dark brown, black, red) [Piquet-Thepot, 1968]
- Hair color from Fischer–Saller scale color pallet (brown, black, blond, auburn, red, gray/white) [Daniel, 1978].
- Estimated age from ages in 5 year bins (e.g. 1-5, 6-10, 11-15...);
- Race from Caucasian, American Indian, Latino, African, Middle Eastern, Indian, and East Asian
- Gender from male and female.

Table 5.1 illustrates the summarize information about our three datasets. Based on these three datasets, we tested the performance of our algorithms, and compared them to previous works on face sketch recognition. Next in this chapter we discuss details and results of these experiments.

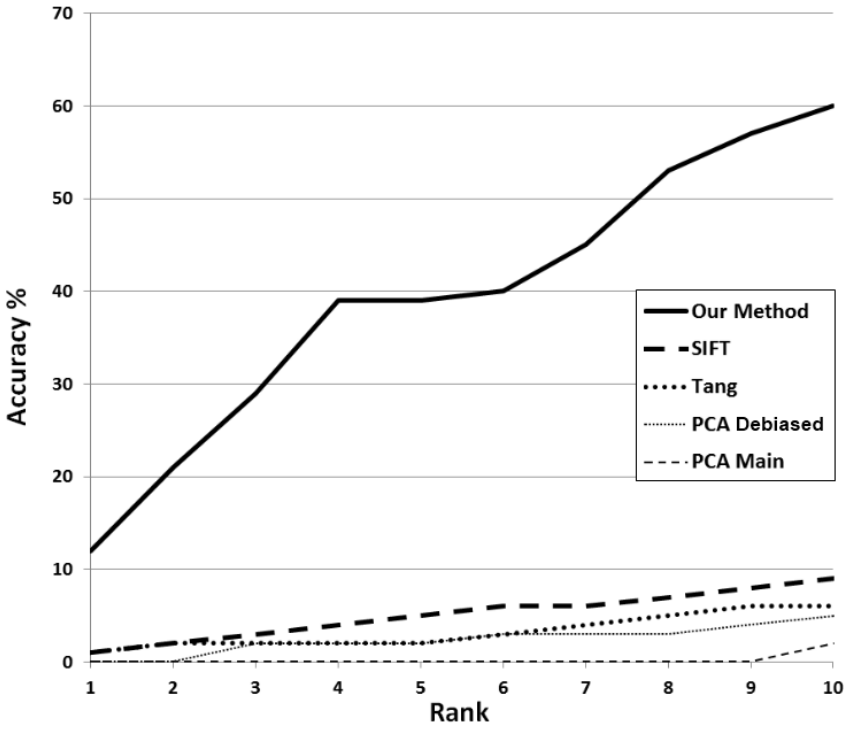

Figure 5.4: CMC curves for PCA on Main Sketches, PCA on debiased sketches, [Tang and Wang, 2004], [Klare et al., 2011], and our FSR, tested on matching non-artistic sketches.

5.2 Experimental Results

In this section we present the experimental results of testing our proposed FSR. We here compare the performance of our FSR on both non-artistic sketches and exact sketches. We compare our results with the performance of PCA, the FSR presented in [Tang and Wang, 2004], and the SIFT-LBP-based FSR presented in [Klare et al., 2011] (reporting the best performance on exact sketches to date) on the combination of our first and second sketch datasets (summing up to 150 non-artistic sketch-photo pairs). We also compare the performance our FSR with these previous methods on matching artistic sketches from CUHK sketch database [Wang and Tang, 2009], used as a test bed in almost all previous works.

Almost all previous methods have used a PCA-based algorithm as their

core (see section 2.1) and therefore, we can safely consider PCA performance as a baseline measurement of accuracy. The FSR introduced in [Tang and Wang, 2004] is reported to have 71% rank-1 (and 96% rank-10) accuracy in recognizing exact sketches and the FSR introduced in [Klare et al., 2011] has reported the highest accuracy on exact sketches from their SIFT-based method to have 97% accuracy for rank-1 exact sketches and 16.33% for rank-1 forensic sketches.

In our experiment for performance comparison on recognizing non-artistic sketches, we used the 249 face photos of the first session of the Multi-PIE face dataset [Gross et al., 2008], as the photo gallery, having a mixture of gender, age, and race. Then each face photo is analyzed by STASM [Milborrow and Nicolls, 2008] to detect facial component outlines and then its Ancillary Information is manually assigned. The non-artistic sketch gallery was the combination of our first and second sketch dataset, consisting of 150 non-artistic sketches.

We performed the sketching bias removal, weighting, and normalization, sketches and photos based on the details in Section 4.2, treating one sketch as the Main Sketch and the rest of the sketch-photo pairs *from the same participant* as the drawing profile. Figure 5.5 shows examples of debiased sketches and their Main Sketch and face photo counterparts. We used the same sketches to be classified using PCA, and FSRs proposed in [Tang and Wang, 2004] and [Klare et al., 2011], to compare their results in Figure 5.4 and table 5.2. In addition, we also tested the PCA on the debiased sketch to test the bias removal step. The CMC curves in this figure show the accuracies up to rank-10 which is similar to top-10 selection of the faces in terms of similarity to the Main Sketch. Based on these results, PCA performed poorly in recognizing both the Main Sketches and debiased sketches, with its rank-50 accuracy re-

Figure 5.5: Examples of debiased sketches and their original Main Sketch counterparts: Red points represent the Main Sketch outlines, green points represent the photo outlines, and the blue points represent the debiased sketch outlines.

	Rank-1	Rank-10	Rank 50
PCA Main	0	0	5
PCA Debiased	0	4	17
Tang et al.	1	6	28
SIFT	1	9	19
Our FSR	**12**	**60**	**93**

Table 5.2: Comparison between accuracy of non-artistic sketch recognition (the first, tenth, and fiftieth ranks), between methods PCA on Main Sketches, PCA on debiased sketches, Tang et al. [Tang and Wang, 2004], SIFT-LBP [Klare et al., 2011], and our proposed FSR.

mained below 10% and 20% respectively. Similarly, Tang et al. and SIFT classifications had poor recognition accuracies, close to PCA performance on debiased sketch. In contrast, the recognition performance of our FSR started with about 12% in rank-1 with a rapid growth as rank increased to about 60% in rank-10. These results therefore indicate the effectiveness of our sketching bias removal, weighting, and matching steps, to transform the crude Main Sketch, into a more reliable and recognizable representation of the target face.

Scalability is another important factor for face sketch recognition methods, as the real database of faces can be in the order of millions. Therefore, in our next experiment we tested the performance of our FSR versus previous method for the effect of photo gallery size. Figure 5.6 demonstrates how the rank-1 recognition accuracies of our FSR, Tang et al., and Klare et al. decreases as the gallery size increases. This figure indicates that although the scalability of our method is better than previous FSRs, its performance sharply drop with the increase in the database size, making it not reliable enough for real cases with large photo databases.

Our next experiment is to compare performances on recognizing exact sketches. In this experiment we used the 188 sketch-photo pairs in the CUHK dataset [Wang and Tang, 2009]. As there was no information about the drawer of these sketches comparable to the drawing profile, we completely omit the

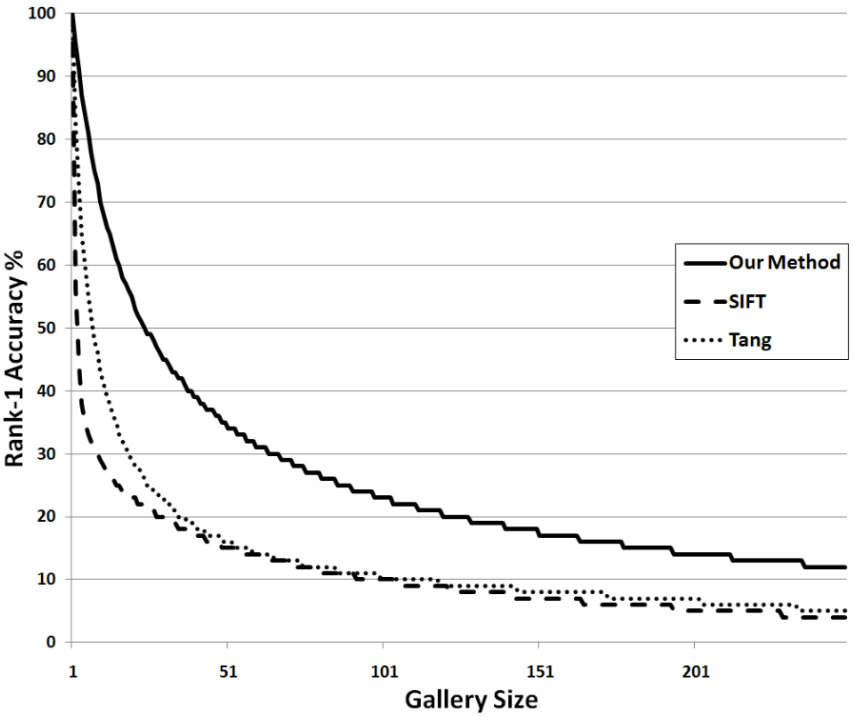

Figure 5.6: Drop in Rank-1 accuracy of our FSR, [Tang and Wang, 2004], and Klare et al. [2011] on the non-artistic sketch recognition, as the photo gallery size increases.

sketching bias removal step for these sketches for our FSR, and we only weight and match the outlines of the sketches and photos. We did not alter the previous methods. Figure 5.7 compares results of our method, PCA, Tang et al.[Tang and Wang, 2004], and SIFT-LBP matching [Klare et al., 2011] on matching exact sketches (see Section 2.1). Results of this second experiment show that, although our method delivers better results than PCA (as the baseline), it is significantly less accurate than the other two FSRs. This is obviously due to the differences in the modalities of artistic and non-artistic sketches, with artistic sketches include a rich combination of accurate facial details, and shading, while non-artistic sketches are merely simple line drawings to represent the essence of the target face appearance (compare figures 2.4 and 5.1). In order to show the sensitivity of previous FSRs to changes small details of exact sketches, we performed our next experiment.

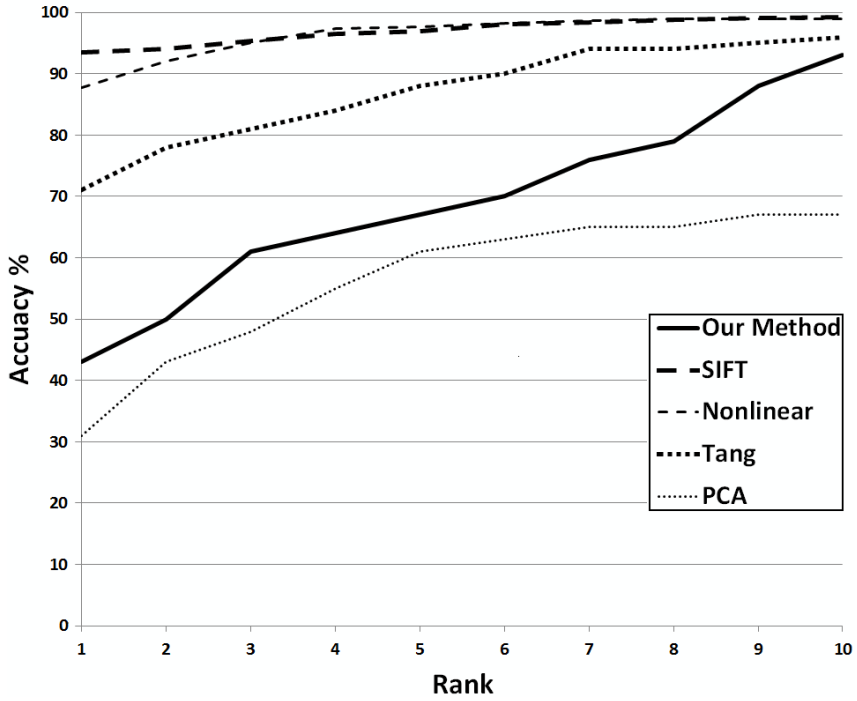

Figure 5.7: CMC curves for PCA, Tang et al. [Tang and Wang, 2004], SIFT (used in [Klare et al., 2011]), and our method for matching artistic sketches from the public dataset of CUHK sketches [Wang and Tang, 2009]

In our next experiment we tested the FSR by Klare et al. [2011] and our method on recognizing reduced exact sketch: tightly cropped exact sketches with their facial shadings removed (See Figure 5.8). In addition to showing the sensitivity of previous FSRs to minute details of exact sketches (such as hair style or shadings), we showed a performance trend for our approach and the FSR by Klare et al. [2011], based on the similarity of sketch to the target face.. Based on the crude sampling illustrated in Figure 5.8, previous face sketch recognition methods are highly sensitive to variation of details such as shading and hair style, and in contrast, our method performs more robustly given stronger variations even in outlines of the face.

In our ETP, we use drawing profile, a set of sketches drawn by the eyewitness while looking at the face photos (copy-sketching), to estimate the process of drawing the Main Sketch, from the memory (memory-sketching). Our final

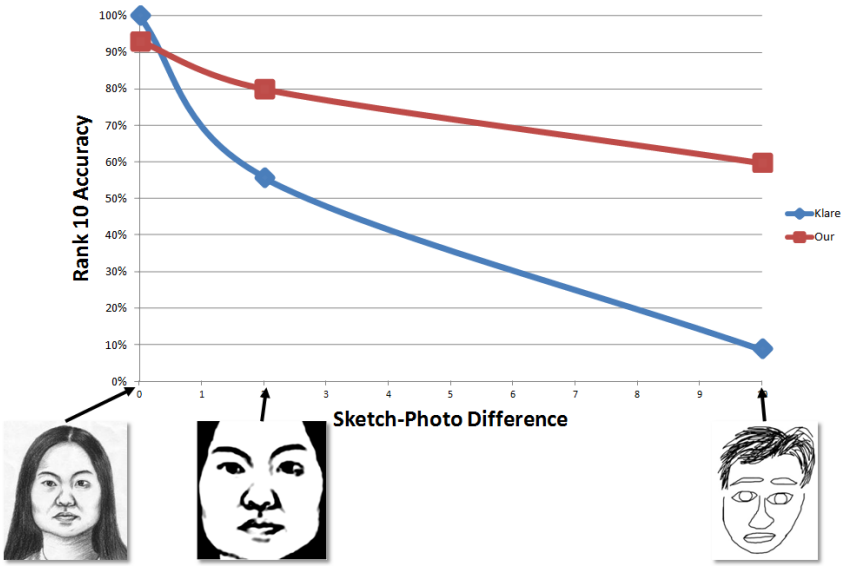

Figure 5.8: Accuracy trend for our approach and the work by Klare et al. Klare et al. [2011] based on the similarity of sketch to the target face

experiment is then to show how closely can copy-sketching $\hat{f}$ estimate memory sketching f. In this experiment we check whether for each face x we have:

$$f(x) \approx \hat{f}(x) \, \forall x$$

or equivalently:

$$x - f(x) \approx x - \hat{f}(x) \, \forall x$$

However, if the face x is seen for the first type of sketching (copy-/memory-sketching), then seeing x again affects results of the next of sketching (memory-/copy-sketching). We therefore check this estimation for facial components, instead of the entire face, with the assumption that facial components from the same gender, age group, and race are almost the same. We therefore check whether for facial components x_i and x'_i we have:

$$x_i - f(x_i) \approx x'_i - \hat{f}(x'_i) \, \forall x_i \approx x'_i$$

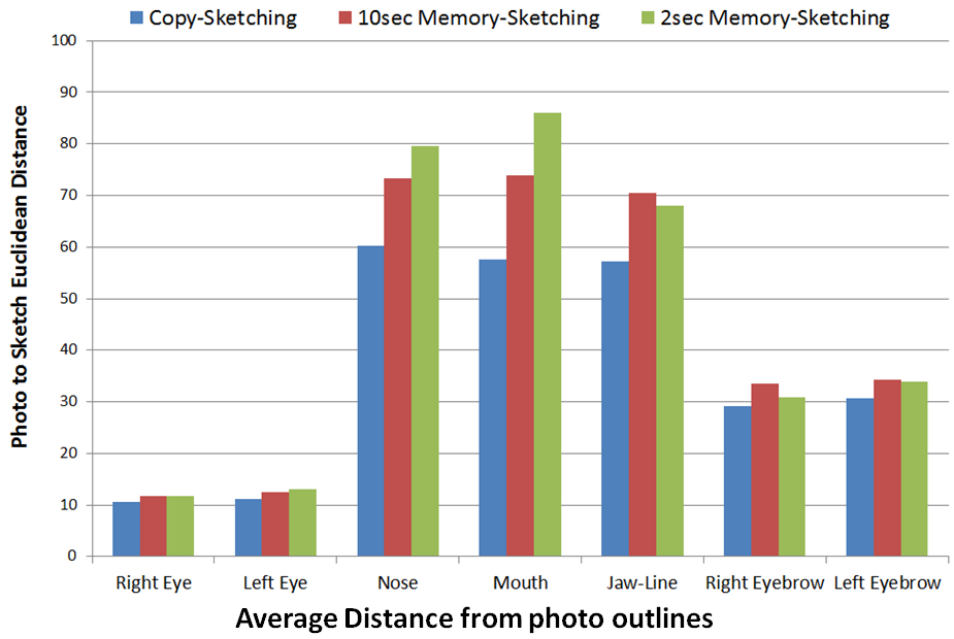

Figure 5.9: Average photo and sketch outline differences based on facial components, in copy-sketching, 10-second memory sketching, and 2-second memory sketching.

In our third dataset of sketches, we have collected two types of memory sketches, 10-sec memory-sketches, and 2-sec memory-sketches. The difference between the photo outlines and the produced sketch outlines is presented in Figure 5.9. It can be seen from this figure that copy-sketching shows a similar pattern to both 10-second and 2-second memory sketches, while, as expected, the reduction of exposure time increases the difference between the photo and sketch outlines.

5.3 Improving Overall Performance

We here used the methods described in Section 4.3 to conduct several experiments, showing the improvements in our FSR. For these experiments we used our third dataset of sketches, that is consisted of 710 sketches, drawn from 249 face images (from Multi-PIE database [Gross et al., 2008]), by 71 subjects,

with different genders (24 female), ages (mean 27.4, standard deviation 3.9), and races (Caucasian, Indian, East Asian, and Middle Eastern) (see Section 5.1).

We tested our face sketch recognition algorithm in verification mode, based on a sample set of all correct sketch-photo pairs (710 pairs) and 10000 incorrect sketch-photo pairs, randomly selected from all possible combinations of sketch-photo pairs (7%-93% correct-incorrect ratio). Based on this pool of sketch-photo pairs, we trained and tested an RBF kernel SVM to label sketch-photo pairs, using cross validation with stratified sampling.

Our first comparison is between the performances based on Specific modeling (i.e. our FSR before improvement, Section 4.2), General modeling, and General-Specific modeling (i.e. improved FSR, Section 4.3.3). We compared the ROC curves for verification performances of these modeling strategies in Figure 5.10. The original FSR (Specific modeling) shows to have a poor performance when the database size increases (from 150 in previous experiments in Section 5, to 710 in this experiment), but using the combination of General and Specific modeling, we can effectively decrease the modality gap and therefore increase the performance.

We also analyzed the effect of number of sketch-photo training pairs per participant on the general-specific model, shown in Figure 5.11, illustrating the ROC curves for sketch verification based on 1 to 9 training sketch-photo pairs (10^{th} sketch is regarded as the Main Sketch). Based on this figure, it seems that the General-Specific model started to perform reasonably well, having 6 or more training pairs per person, but in situations that we have less than 6 training pairs, the performance was not reliable enough to remove the sketching bias and recognize the target face. Nonetheless, acquiring 10 sketches as the drawing profile is not time consuming or labor intensive (in

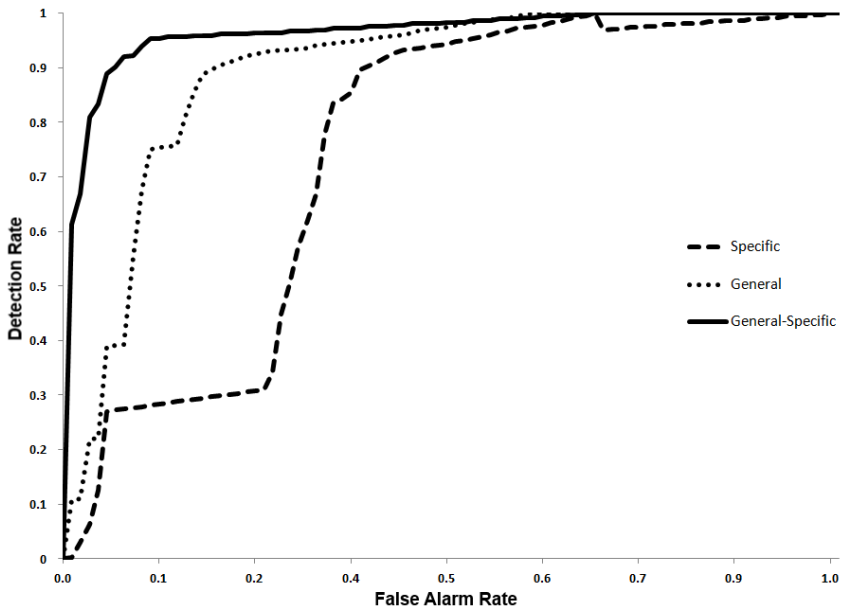

Figure 5.10: ROC curves of original sketches v.s. general, specific, and general specific models, for face sketch verification task.

contrast with traditional ETPs which are stressful and may last up to two days), and therefore, 10 training samples can be set as a baseline for the number of sketch-photo pairs the system requires for a proper recognition.

5.4 Application to Twin Ear Recognition

In this section we evaluate the performance of our algorithm in verification of ear images from 39 pairs of twins (78 subjects). Our ear image dataset is obtained during the Sixth Mojiang International Twins Festival, China, 2010, containing Chinese, Canadian and Russian subjects, each having 2 to 4 real and 20 synthesized images, all in 1728×1152 pixels. Real images are captured from profile view, containing the entire head and shoulder, with some translation and rotation (in-plane and off-plane). Fig. 5.13 displays some of these images (cropped for better illustration). Synthesized images are obtained from real images by adding random amounts of noise, translation, off plane rotation,

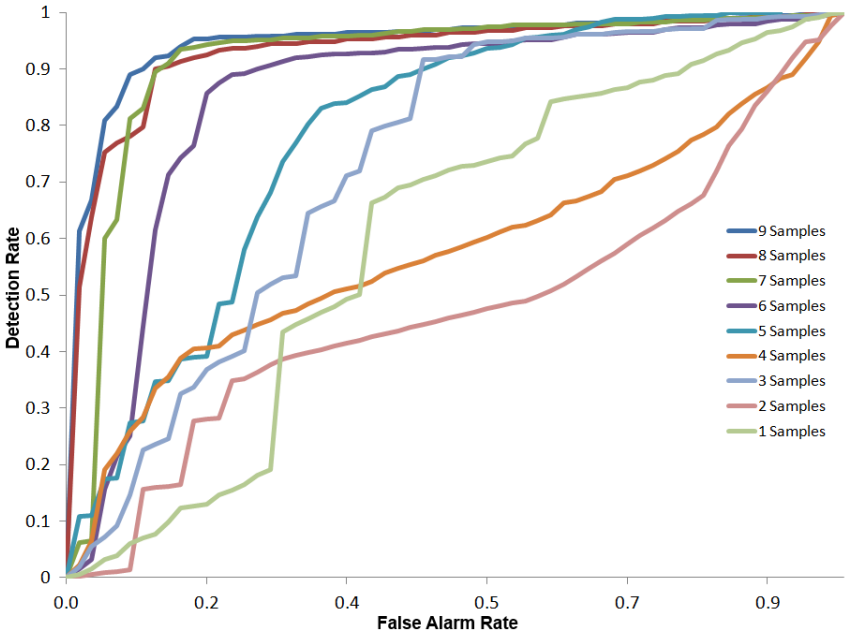

Figure 5.11: Improvement in the performance of General-Specific modeling, with the increase in the number of training sketch-photo pairs per eyewitness.

and realistic motion blur, acquired from [Xu and Jia, 2010]. We used 10 motion kernels from [Xu and Jia, 2010] to synthesize realistic motion blur in ear images, illustrated in figure 5.12 with examples of blurred ear images.

5.4.1 Experimental results

We test our algorithm performance on the Twins dataset with a verification scenario: given a pair of ear images, we verify whether both ears belong to the same subject or not. Under this scenario, we assess the robustness of our algorithm in five experiments:

1. Five different resolutions: We down-sample each GEar image to 300×300, 150×150, 75×75, 37×37, and 18×18 pixels and up-sample it again to 300×300 pixels (Fig. 5.14, top row).

2. Four different noise levels: We add white (Gaussian) noise with $\mu = 0$ and σ =0, 0.1, 0.3, and 0.5 to the GEar images (Fig. 5.14, middle row).

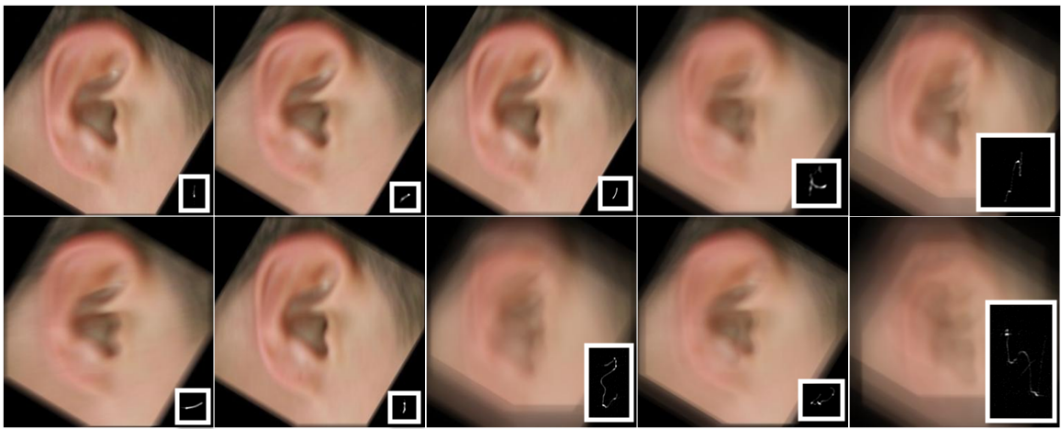

Figure 5.12: Examples of realistic motion blur in synthesized ear images. The motion kernel is from [Xu and Jia, 2010].

Figure 5.13: Some examples of ear images in our dataset (cropped for better illustration).

3. Four different occlusion levels: We simulate 0%, 10%, 30%, and 50%, occlusion levels in right-to-left and top-to-bottom directions (Fig. 5.14, bottom row).

4. Different right ear vs. left ear training-testing sets: We test the performance differences between training and testing on different ear sides. We evaluate two cases of training and testing on the same side (right or left), and two cases of training and testing on different sides.

5. Dimensionality reduction: Motivated by the optimality claim of the Exception Report Model, we test accuracy of our algorithm by applying dimensionality reduction on the level of exceptionality of points. In this test we evaluate how our algorithm performs when we only use points with a minimum level of exceptionality. For example, for the appearance weights we have:

$$W_{app}(dist) = \begin{bmatrix} w_{app,1}(dist) \\ w_{app,2}(dist) \\ \vdots \\ w_{app,n}(dist) \end{bmatrix}$$

$$w_{app,k}(dist) = \begin{cases} w_{app,k}\, if & w_{app,k} > dist \\ 0 & o.w. \end{cases}$$

where $W_{app}(dist)$ indicates the new intensity weights, based on $dist$, the minimum required level of exceptionality. We apply the same weighting strategy for shape weights.

We also compare the performance of our method with a recent work by Bustard and Nixon [Bustard and Nixon, 2010] (B&N). For B&N algorithm,

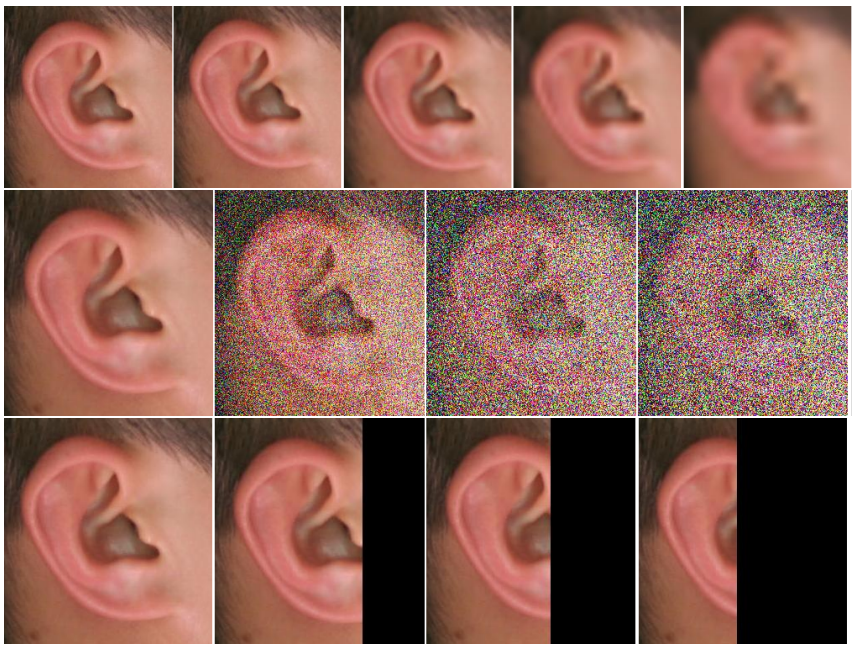

Figure 5.14: An example of resolution (left to right: 300×300, 150×150, 75×75, 37×37, and 18×18 pixels), noise (left to right: standard deviation 0.0, 0.1, 0.3, and 0.5), and occlusion (left to right: 0%, 10%, 30%, and 50%) of ear images.

we corrected the small error in the formula (see Appendix). We also ignore comparisons in which B&N fails to find four corresponding points for calculating the homography transformation.

Accuracy results of ours and B&N algorithms are illustrated in Fig. 5.15 to 5.18, and Table 5.3. Based on our results, our algorithm could perform up to 92% on the Twins dataset, constantly better than B&N. Results also show robustness of our algorithm to resolution, noise, and occlusion.

Regarding resolution variations, results in Fig. 5.15 show that our algorithm is constantly performing better than B&N. However, as resolution decreases, our algorithm's accuracy drops sharper than B&N's. This suggests that the B&N algorithm although low in accuracy, may perform better than our algorithm in very low resolutions (less than 18×18 pixels).

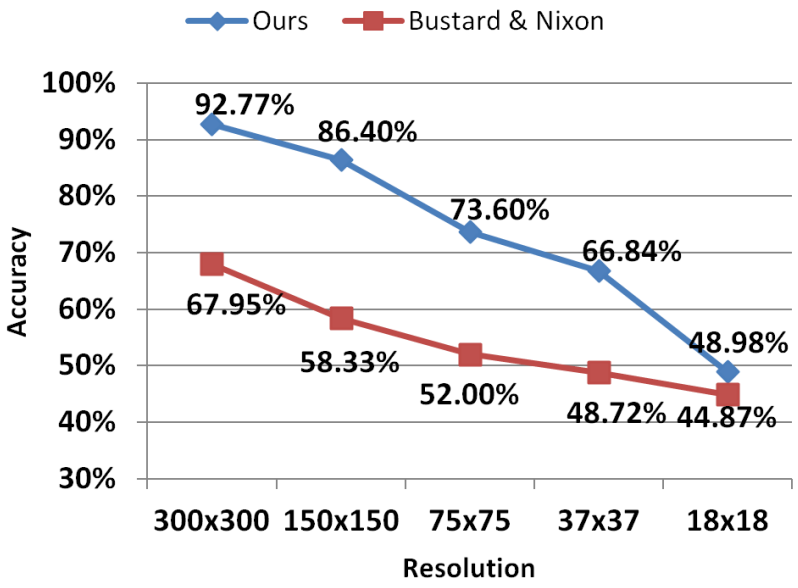

Figure 5.15: Left: Results of verification between sibling across different resolutions.

The noise variation results in Fig. 5.16 indicate that our algorithm is highly robust to noise variations, and even with noise$\sigma = 0.5$, its accuracy is almost the same as the B&N without noise. One reason may be because of the performance of the SIFTFlow dense point registration, which can tolerate noisy data. But it also indicates that the exceptional features are robust against the introduced noise.

Comparing occlusion variation accuracy results in Fig. 5.17 with other tests, it seems that our algorithm is robust towards resolution and noise than the occlusion. This can be because of the loss of strong features (trained in the non-occluded images) in the occlusion variations, while these features, although weakened, are still present in the resolution and noise. In addition, as the accuracy drops more rapidly in the top-to-bottom occlusion curve, it seems that strong features are located more at the top of the ears in our dataset, rather than right of the ears.

Results of training and testing on same or different side ears, presented in Table 5.3, show that the left and right ears in our subjects do not share much

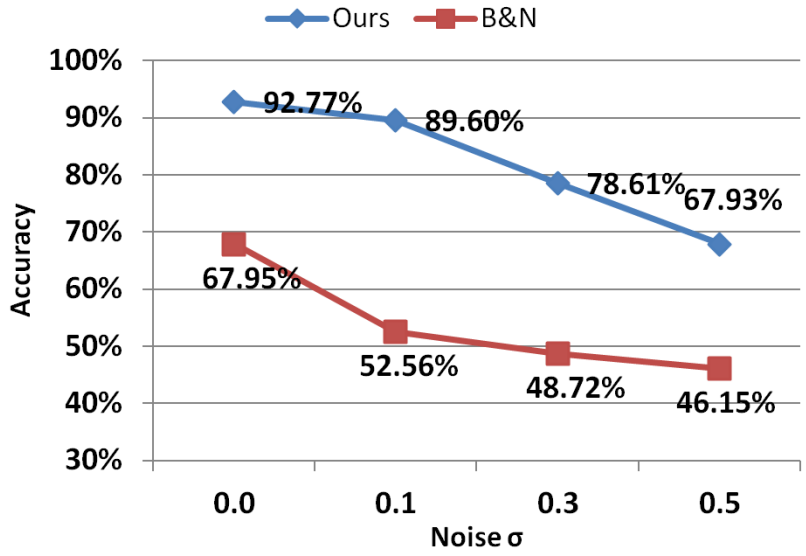

Figure 5.16: Results of verification between sibling with different noise levels

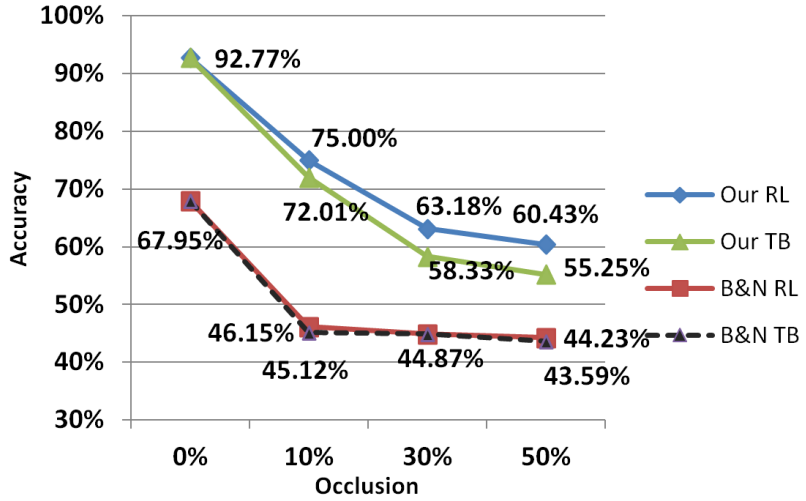

Figure 5.17: Results of verification between sibling with different occlusion levels

Training	Left	Right	Left	Right
Testing	Left	Right	Right	Left
Accuracy %	92.77	92.76	54.78	53.40

Table 5.3: Results of training and testing with left and right ears.

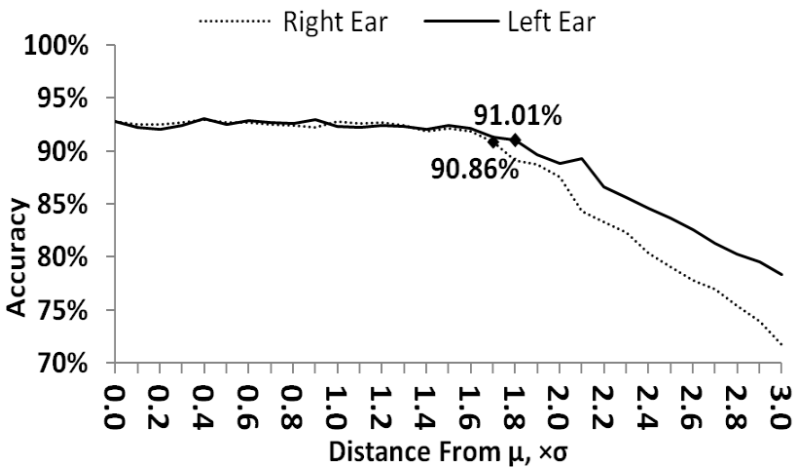

Figure 5.18: Results of dimensionality reduction: Accuracy trends of recognition of the right and left ear, based of the level of exceptionality of the feature (i.e. the normalized distance from the respective mean value). X marks show the largest distance from μ with accuracy higher than 90%.

of their exceptional features. This means that one cannot train only on one side ears and hope to accurately recognize ears from both sides.

Finally, the dimensionality reduction results are presented in Fig. 5.18. These results seem confirm the optimality claim of the Exception Report Model (ERM) that with only about 10% of the features (only exceptional features, with distance more than 1σ from the μ), the brain can accurately and rapidly recognize faces. Similarly, Fig. 5.18 shows that even using only features with distance more than 1.7σ from the μ, we can still achieve more than 90% accuracy in ear recognition. Given a perfect Gaussian distribution of features (assumed in the real world), it means that only using about top 5% of the features, fast and accurate recognition can be performed.

5.5 Chapter Summary

In this section we presented the result of our experiments on our proposed coupled ETP-FSR, and our comparisons with previously proposed methods,

among them the method by Klare et al. with the best reported performance on exact sketches to date. In our experiments we tested our method on both exact and non-artistic sketches from CUHK sketch database and our collected sketch database respectively. We showed that as previous methods are designed to use exact shading cues of artistic sketches, and our method is designed to use the crude outlines in non-artistic sketches, results of our experiments on artistic and non-artistic sketches are almost opposite.

We then showed results of our proposed improvements by imposing temporal order, using generalized point weighting, and using General-Specific modeling of the sketching bias. We showed various results including ranked accumulated accuracy, ROC curves, and number of training samples, and comparing performances with previous methods.

We therefore conclude that our method firstly provides a new perspective on the eyewitness face sketch recognition problem and secondly shows to be reliable in recognizing non-artistic sketches without causing distortions to the eyewitness' mental image of the target face.

Our method however, has a specific assumption to be able to work efficiently. For our algorithm to provide reliable results, the eyewitness should have a certain level of drawing skills. An eyewitness with a poor drawing skills either cannot produce a sketch, or produces a significantly crude sketch that would not bear enough identity-specific information for our algorithm to process. Thus, there should be certain protocols to deal with these cases, by providing drawing aids to the eyewitness, so that a final representation is produced, even with the cost of additional (and perhaps irreversible) biases. In these cases we practically step back towards traditional eyewitness testimony procedures. Our experimental results in Figure 5.8 has another indication here that how our algorithm would behave in cases of higher quality sketches pro-

duced using drawing aids (here with the assumption of no additional biases).

One possible solution for drawing aids can be providing an initial very simple sketch for the eyewitness to start with and manipulate. This initial sketch can be calculated from an average face based on the Ancillary information from the eyewitness, such as gender, age, and race. The eyewitness can then manipulate this sketch by changing the relative location, size, shape, and color of each facial component until a he/she is satisfied with the resulting sketch. However this process can add biases to both mental image of the face and the final results, due to viewing similar faces (as the face becomes more similar to the target face as the manipulation progresses).

As a final note, the core of our proposed methods is trying to understand present biases in an ETP, and find possible ways to remove these biases. This core concept applies in any possible evolutions of our proposed method, either when moving backwards to the traditional ETPs, or moving forwards by introducing new medium of transferring identity-specific information. One should therefore analyze and reduce these biases as much as possible, in any stage of this procedure.

In the next chapter we summarize this entire text, and draw the broader conclusions and consider possibilities for future works.

Chapter 6

Summary and Conclusion

At the final chapter of this text, we summarize our works and achievements, discuss possible fall-back strategies to make our system more robust in cases of poor eyewitness memory, and finally draw conclusions and discuss possible future directions.

Face perception is a complicated process in human's visual system which is learned based on life experiences, and becomes a very personal subconscious process, that cannot be imitated. Humans rapidly and accurately perform face recognition, without intensive calculations, and this swiftness in face perception has helped the human have a vast and diverse social behavior. Thus, many researchers in the automatic human identification field have set the performance of humans as their goal, and developed several biologically inspired methods. In this text we focused on the eyewitness face sketch recognition problem, which includes an interesting mixture of both human's behavioral and perceptual factors (in the eyewitness testimony procedures, ETPs), as well as requirement for machine algorithms (in the face sketch recognition methods, FSRs).

We first reviewed the automatic face sketch recognition methods (FSRs)

in Chapter 2, where we showed that FSRs have unrealistic assumptions for the similarity of forensic sketches to their target faces. We showed that exact sketches that are used for performance measurement in FSRs are not proper estimations of real forensic sketches. Therefore, although previous FSRs have reported high accuracy rates on recognizing exact sketches, they are unreliable in recognizing real eyewitness sketches. Moreover, due to modality difference between forensic sketches and face photos, conventional face recognition methods cannot be applied to match forensic sketches.

We then reviewed the psychologically challenges in currently performed eyewitness testimony procedures (ETPs) in Chapter 3. In this chapter we reviewed the main problems of current ETPs such as general memory limitations, biased instructions, piecewise face reconstruction, post-event information, viewing similar faces, verbal overshadowing, and last but not least, mental norm biases. As we discussed in this chapter, these problems make results of the ETPs (forensic sketches) unreliable. We showed that the studies clearly show that human memory is fragile, malleable, and susceptible to suggestion, and current eyewitness testimony procedures, unwittingly, change the memory of the target face in the eyewitness' mind, resulting in production of unreliable reconstructions of the target face. An important effect of the unreliability of these procedures is that any method that uses results of these procedures inherits this unreliability.

In addition, we also reviewed suggested psychological frameworks for face processes in the human visual system. Based on this review we tried to select the best psychological framework, in terms of explaining the human's visual system behavior. Among all proposed frameworks, we could narrow down our search to the Exception Report Model (ERM), which could explain many of the reported phenomena in human face perception (e.g. the rapid and accurate

processing, other-race effect, mate selection, etc.), and was supported by other works both in psychology and computer vision. Therefore we selected the ERM as our understanding basis for how humans perform face perception, memorization, and recognition.

The importance of the eyewitness face sketch recognition problem, and the presence of these gaps in the ETP and FSR parts shaped our motivation to design a better solution, based on our knowledge of human face perception.

In Chapter 4 we then proposed a new perspective on the eyewitness face sketch recognition problem, by taking into account the individual biases which are involved in perceiving and drawing a face. We proposed a more reliable ETP and an accompanying FSR (coupled ETP-FSR) based on non-artistic sketches directly drawn by the eyewitnesses. This framework is built on more realistic assumptions on the eyewitness' abilities, and testimony results. Our solution to the ETP problems was to ask the eyewitness to draw a non-artistic sketch of the target face (Main Sketch) by him/herself, to avoid any disturbance to the mental image of the target face caused by verbal description, police artist, face composite software, etc. In addition, we not only assess the face sketch, but also focusing on how the sketch is created in the ETP. Creation of a face sketch is a complicated process which involves how a person perceives, remembers, and recognizes faces, which are biases that are evolved based on personal life experiences. Therefore a face sketch not only has information about the face it is representing, but also includes mental biases and individual face perception styles of the drawer. We also asked the eyewitness to provide categorical information about the target face, such as race, skin color, and gender. Finally, in order to estimate the eyewitness' face perception bias and face drawing bias (together as sketching bias) in the Main Sketch, we also acquired a drawing profile of the eyewitness by asking him/her to provide

additional sketches from a set of known face photos. In our FSR solution, we first used the drawing profile as training samples to estimate and remove the sketching bias from the Main Sketch. Then we weighted each part of this debiased sketch, based on deviations from norm, motivated by how humans remember and recognize faces, suggested in ERM. Finally we matched this weighted sketch against the photo database based on a weighted combination of facial component outlines, facial marks, skin color, race, etc.

In our experiments in Chapter 5, we tested our proposed FSR on a datasets of 150 sketches, illustrating performance reliability, comparison with previous FSRs (including Klare et al. [2011] reporting the best performance on exact sketches), effect of the photo gallery size, and on both artistic and non-artistic sketches. Based on these primary results, we proposed three improvements to the sketching bias estimation, weighting, and matching steps, to increase the accuracy and robustness of our FSR. We then tested the performance of our improved method again on a larger dataset of 710 sketches that we have collected from 71 non-artist drawers. We showed that our improved method can reliably recognize the non-artistic sketches using only 6 to 9 sketching samples for each eyewitness. Results of these experiments show that our coupled ETP-FSR can retrieve the mental image of the target face, and recognize it in a database of photo, while avoiding many known psychological problems in current ETPs, and having realistic assumptions of the problem (unlike previous FSRs).

Finally, in Chapter 4.4 we extended our introduced framework for face sketch recognition, to another visual identification application, identification based on ear images. In this application we showed that using the same principles as in our FSR, and with small problem-specific modifications, we can use our framework to identify between twin siblings, based on their ear im-

ages. Similar to Chapter 4, we weighted and matched ear appearance and ear structure based on deviations from norm, and showed that only based on the top 5% deviated features, we can reach 90% accuracy in this application. Our selected psychological framework, ERM, also suggests the use of only a few features that are the most deviated, by the human visual system in face recognition.

Despite our achievements using this new ETP-FSR, there is a human bottleneck in our proposed system. This system starts by the eyewitness providing the crude representation of the target face in terms of the Main Sketch, and the system uses different methods to find the debias this representation into a stage that it is recognizable. However, in cases that the eyewitness cannot create a face out of his memory, the system stops at its very first step. The traditional ETPs use mug shots, or pieces of facial components (as a memory trigger) to help the eyewitness recall the target face, with each of these memory triggers adds a type of bias to the system. However, one can accept a certain amount of bias to achieve at least a biased reconstruction, instead of to absolutely no reconstruction (as our system relies solely on the eyewitness' ability to remember the target face). We therefore here suggest a hybrid approach to use traditional methods, with addition of as little bias as possible.

Looking at our ETP and traditional ETP as the extreme options in terms of in terms of involvement of the eyewitness in producing the final sketch, these two extreme points also can represent the two extremes of added bias to the system, with our ETP having the least amount of bias (only Sketching Bias), and the traditional ETP having the maximum amount of bias (verbal overshadowing, piecewise reconstruction, exposure to similar faces, etc). But having these two extremes, we can also have a hybrid approach that exploits the gray space in between. The more this hybrid approach uses memory triggers,

Figure 6.1: Line drawing samples of isolated facial components, to be used separately to assist the eyewitness in remembering the target face structure.

suggestions, and third party assistance to the eyewitness, the closer it would be to the traditional ETP (minimum eyewitness involvement, maximum bias), and the more the hybrid approach relies on the ability of the eyewitness to remember and draw the sketch, the closer it would be to our proposed ETP (maximum eyewitness involvement, minimum bias). Here we briefly discuss two possibilities to assist the eyewitness, using the schemes from traditional ETPs.

One of the least biased methods to trigger eyewitness' memory can be letting the eyewitness choose isolated facial components that s/he is not sure about, without showing the complete (or composite) face. After deciding on each facial component, the eyewitness can proceed to sketch drawing by him/herself. Using this method the bias is only added to the viewed facial components and not others. The presentation of facial components can be also in a symbolic or line-only manner to reduce the added bias. Figure 6.1 illustrates examples of isolated facial features in line-drawing style, that can be used for this purpose.

Another method of helping the eyewitness in remembering the face structure can be using a generic face (normal face) based on the given Ancillary information (e.g. race, gender, age, skin color, etc.), as a starting point from which the eyewitness can manipulate the image to reach his/her desired face

shape. This method of course would be possible in cases that the Ancillary information is available (i.e. the eyewitness has a vague memory of the target face). The possible bias added to the eyewitness' memory is from the effect of norm face, which is known to create a false familiarity effect Unnikrishnan [2012]. It means the eyewitness may feel that s/he knows (and have seen) this face before, and therefore automatically links this normal face to the target face memory, resulting in memory distortions. The false familiarity effect is stronger when the eyewitness is from the same race as the presented normal face.

The main point in using any method for memory trigger or assisted sketching is to remember that, any of these methods would add a specific type of bias to the system, based on the amount of suggestions that they introduce to the system. Therefore there should be a balance between the amount of information the eyewitness remembers, and the amount of assistant s/he receives from the system.

Based on our reviews, computational model, and experimental results, we conclude that at least in applications where both human and machine are involved (e.g. eyewitness face sketch recognition), and particularly for face processing related applications, there should be a careful accounting for human's mental biases. Currently all of these applications are regarding the same properties for all human entities, and therefore, the differences between each individual are not considered. Thus not only outcome of these applications may become unreliable (due to over simplification of the problem), but also there can be distortions to mental processes of the humans users of these applications. Examples of these distortions are the mental distortions in the current ETPs that cause changes in the eyewitness' mental image of a face. On the other hand, for a reliable design regarding human mental biases, there should

be a reliable framework to guide the design with a big picture of how the human visual system works. In our exploration we selected the ERM as our basis psychological framework, but as our understanding of the human brain evolves, new frameworks can emerge to replace ERM. In addition to the psychological framework, when the application requires information regarding brain stages or memories (e.g. the mental image of a face in the eyewitness' mind), the application design should be in the way to remove distortions as much as possible (e.g. using self-drawn face sketches, instead of verbal description of the faces), and the remaining distortions should be estimated and removed when possible (e.g. sketching bias estimation). We therefore can present this text as a sample of our perspective on designing computational models for analyzing human inputs, particularly inputs directly relating to the brain signals.

6.1 Future Work

Our proposed combination of ETP-FSR can be used in real situations as it requires no trained users, and delivers more reliable results in a shorter time than currently used ETPs. In addition, the same framework introduced here for face sketch recognition, can be applied to other human visual identification tasks such as identification based on ear images [Nejati et al., 2012], and identification based on uncontrolled head and face movements [Zhang et al., 2013].

Nonetheless, our proposed method here can be improved in various ways, as future works on this topic. First of all, it should also be noted that understanding strategies used by the human visual system still has many gaps to be filled, therefore, one can improve our framework by using models better than ERM, or a combination of several psychological frameworks that may together

represent a more complete picture of the human visual system. Similarly, the estimation of mental bias can be extended to a bigger picture, as recent psychological studies revealed the effects of addition (non-visual) information, such as names, in perceiving faces [Hilliar and Kemp, 2008]. Therefore, the bias estimation can be also applied on the categorical information provided by the eyewitness to further assess the eyewitness' perception of race, skin color, or age.

In addition, as discussed in Section 5.5, our method has the assumption of a minimum level of drawing skills for the eyewitness, without which the final results would be unreliable. Therefore, hybrid methods to combine traditional techniques and our method should be developed to handle these cases. The important issue about a hybrid method is the difficulty assessing added biases to the eyewitness' memory and the final sketch, as using drawing aids can add non-measurable biases to the entire protocol.

An important issue in bias estimation that would affect this entire system is the effect of time on the memory of the face. Our third dataset of sketches includes long-exposure sketches (from faces viewed for 10 seconds) and short-exposure sketches (from faces viewed for 2 seconds) with 1 minute delay between the viewing time and the sketching time. Thus an interesting future work would be the analysis of time effect and methods to overcome the added bias.

Time delay has a significant effect on the memory of the eyewitness. As we showed in our experiments (Section 5.2), sketching error increases with reduction of exposure time. In most of the cases, increase of time delay also increases the error. The question of amount of the time delay effect was out of the scope of this thesis, as it is more a psychological question. The analysis of time delay effect on the system may also bring forth situations that the

eyewitness requires some memory triggers to remember the facial structure. Similar to poor drawing skills cases, in these cases we can incorporate fractions of traditional ETPs into our system. The extent of this incorporation and the effects of it (due to addition of biases) on the final sketch as well as on the memory of the face should be closely assessed.

Finally, a complete and practical ETP-FSR system should be able to produce a photo-like reconstruction of the target face to be distributed in public if necessary. This process is possible by adding machine learning components to our proposed system to estimate the facial shading, color, etc.

The aspect of mental bias is also interesting by itself, as it includes identity related information of sketch drawer, shaped throughout the life experiences. Due to differences in individual experiences in life (exposure to different faces), the face perception bias seems to be almost unique for each individual. Therefore, estimating the mental bias (either based on face drawing profile, or brain signals) may lead to new methods of human identification that can not only provides information about the drawer's identity, but also about the race, gender, and even possibly the living environment of the drawer.

Finally, it is important to remind that the main motivation for this work was to create the missing link between psychological findings, automatic face sketch recognition, and real world applications such as eyewitness face sketch recognition, and therefore reduce the chance of wrongful convictions of innocents. We hope that our work stimulate new generations of works based on more realistic assumptions on how humans and machines can operate together, and serves as a step towards methods benefiting human lives.

6.2 List of Related Publications

- H. Nejati and T. Sim, "Face recognition: Different strokes for different folks," in BIC, 2011 (Chapter 3).

- H. Nejati and T. Sim, "A study on recognizing non-artistic face sketches," in Workshop on Applications of Computer Vision (WACV), pp. 240247, 2011 (Chapter 4).

- H. Nejati, T. Sim, and E. Martinez-Marroquin, "Do you see what i see? a more realistic eyewitness sketch recognition," in Biometrics (IJCB), International Joint Conference on, pp. 18, 2011 (Chapter 4).

- H. Nejati, L. Zhang, T. Sim, E. Martinez-Marroquin, and G. Dong, "Wonder ears: Identification of identical twins from ear images," in Pattern Recognition (ICPR), 2012 21st International Conference on, pp. 1201– 1204, 2012 (Chapter 4.4).

- L. Zhang, H. Nejati, L. Foo, K. T. Ma, D. Guo, and T. Sim, "A Talking Profile to Distinguish Identical Twins," in Automatic Face and Gesture Recognition (FG), 2013 (Section 4.3).

- H. Nejati, L. Zhang, and T. Sim, "Eyewitness Face Sketch Recognition Based on Two-Step Bias Modeling," in Computer Analysis of Images and Patterns (CAIP), 2013 (Section 4.3).

Acknowledgment

This research has been supported by NUS FRC Grant R252-000-451-112. The author thanks the Singapore Police Force, and particularly Superintendant Jason Loke for their supports.

Bibliography

A. F. Abate, M. Nappi, D. Riccio, and S. Ricciardi. Ear recognition by means of a rotation invariant descriptor. In *ICPR, Vol 04*, pages 437–440, 2006.

M. Ali, MY Javed, and A. Basit. Ear recognition using wavelets. *Paper, National University of Sciences and Technology, Peshawar Road, Rawalpindi, Pakistan*, 2007.

N. Alyuz, A.A. Salah, and L. Akarun. Alternative average face models for 3d face registration. In *Signal Processing and Communications Applications, 2007. SIU 2007. IEEE 15th*, pages 1 –4, june 2007. doi: 10.1109/SIU.2007.4298697.

O. Arandjelovic and R. Cipolla. An information-theoretic approach to face recognition from face motion manifolds. *Image and Vision Computing*, 24 (6):639 – 647, 2006. ISSN 0262-8856. doi: 10.1016/j.imavis.2005.08.002. Face Processing in Video Sequences.

Alan D. Baddeley. *Human memory : theory and practice*. Allyn and Bacon, Boston [u.a.], 1990.

Nick E. Barraclough and David I. Perrett. From single cells to social perception. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1571):1739–1752, June 2011. doi: 10.1098/rstb.2010.0352. URL `http://dx.doi.org/10.1098/rstb.2010.0352`.

Susan E. Barrett and Alice J. O'Toole. Face adaptation to gender: Does adaptation transfer across age categories? *Visual Cognition*, 17:700 – 715, 2009. ISSN 1350-6285. URL http://www.informaworld.com/10.1080/13506280802332197.

F.C. Bartlett. *Remembering: A Study in Experimental and Social Psychology*. Cambridge University Press, 1932.

P. J. Benson and D. I. Perrett. Perception and recognition of photographic quality facial caricatures: Implications for the recognition of natural images. *Eur. J. Cognitive Psychol.*, 3, no. 1:105–135, 1991.

Daniel M. Bernstein and Elizabeth F. Loftus. How to tell if a particular memory is true or false. *Perspectives on Psychological Science*, 4(4):370–374, 2009. ISSN 1745-6916.

H.S. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa. On matching sketches with digital face images. In *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, pages 1 –7, sept. 2010. doi: 10.1109/BTAS.2010.5634507.

Markus Bindemann, Christoph Scheepers, and A. Mike Burton. Viewpoint and center of gravity affect eye movements to human faces. *J. Vis.*, 9(2):1–16, 2 2009. ISSN 1534-7362. URL http://journalofvision.org/9/2/7/.

G. H. Bower, S. Thompson-Schill, and E. Tulving. Reducing retroactive interference: An interference analysis. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 20:51–66, 1994.

W. L. Braje, D. Kersten, M. J. Tarr, and N. F. Troje. Illumination effects in face recognition. *Psychobiology*, 26:371–380, 1998.

K. R. Brooks and R. I. Kemp. Sensitivity to feature displacement in familiar and unfamiliar faces: Beyond the internal/external feature distinction. *Perception*, 36:1646–1659, 2007.

V. Bruce. Influences of familiarity on the processing of faces. *Perception*, 15: 387–397, 1986.

V. Bruce and S. Langton. The use of pigmentation and shading information in recognizing the sex and identities of faces. *Perception*, 23:803–822, 1994.

V. Bruce and A. W. Young. Understanding face recognition. *Br. J. Psychol.*, 77:305–327, 1986.

V. Bruce and A.W. Young. *In the Eye of the Beholder: The Science of Face Perception*. Oxford: Oxford University Press, 1998.

V. Bruce, T. Doyle, N. Dench, and M. Burton. Remembering facial configurations. *Cognition*, 38:109–144, 1991.

V. Bruce, Z. Henderson, K. Greenwood, P. J. B. Hancock, A. M. Burton, and Miller P. I. Verification of face identities from images captured on video. *J Exp Psychol. Appl.*, 5, no. 4:339–360, 1999.

M. Burge and W. Burger. Ear biometrics in computer vision. In *ICPR*, volume 2, pages 822 –826 vol.2, 2000. doi: 10.1109/ICPR.2000.906202.

A. Mike Burton, Rob Jenkinsa, Peter J.B. Hancock, and David White. Robust representations for face recognition: The power of averages. *Cognitive Psychology*, 51, Issue 3:256–284, 2005.

John D. Bustard and Mark S. Nixon. Toward unconstrained ear recognition from two-dimensional images. *Trans. Sys. Man Cyber. Part A*, 40:486–494, May 2010. ISSN 1083-4427.

C.-C. Carbon and H. Leder. Face adaptation: Changing stable representations of familiar faces within minutes. *Advances in Cognitive Psychology*, 1:1–7, 2005.

Claus-Christian Carbon, Tilo Strobach, Stephen R. H. Langton, Géza Harsanyi, Helmut Leder, and Gyula Kovacs. Adaptation effects of highly familiar faces: Immediate and long lasting. *Memory & Cognition*, 35, no. 8: 1966–1976, 2007.

CA Carlson, SD Gronlund, and SE Clark. Lineup composition, suspect position, and the sequential lineup advantage. *J Exp Psychol Appl*, 14(2):118–28, 2008.

Christopher F. Chabris, Daniel Simons, and Daniel J. Simons. *The Invisible Gorilla: And Other Ways Our Intuitions Deceive Us*. Crown, 2010.

N.P. Chandrasiri, R. Suzuki, N. Watanabe, H. Yoshida, H. Yamada, and H. Harashima. Is average face recognized as the average. In *Signal Processing and Its Applications, 2007. ISSPA 2007. 9th International Symposium on*, pages 1 –4, feb. 2007. doi: 10.1109/ISSPA.2007.4555339.

Hui Chen and Bir Bhanu. Human ear recognition in 3d. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29:718–737, April 2007. ISSN 0162-8828. doi: http://dx.doi.org.libproxy1.nus.edu.sg/10.1109/TPAMI.2007.1005. URL `http://dx.doi.org.libproxy1.nus.edu.sg/10.1109/TPAMI.2007.1005`.

Wei Chen, Tongfeng Sun, Xiaodong Yang, and Li Wang. Face detection based on half face-template. In *Electronic Measurement Instruments, 2009. ICEMI '09. 9th International Conference on*, pages 4–54 –4–58, 16-19 2009. doi: 10.1109/ICEMI.2009.5274642.

Jonghyun Choi, A. Sharma, D.W. Jacobs, and L.S. Davis. Data insufficiency in sketch versus photo face recognition. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*, pages 1–8, 2012. doi: 10.1109/CVPRW.2012.6239208.

Michal Choras. Ear biometrics based on geometrical method of feature extraction. volume 3179 of *Lecture Notes in Computer Science*, pages 51–61. Springer, 2004.

S. E. Clark and S. L. Davey. The target-to-foils shift in simultaneous and sequential lineups. *Law & Human Behavior*, 29:151–172, 2005.

S. E. Clark and J. L. Tunnicliff. Selecting lineup foils in eyewitness identification experiments: Experimental control and realworld simulation. *Law & Human Behavior*, 25:199–216, 2001.

Steven E Clark and Ryan D Godfrey. Eyewitness identification evidence and innocence risk. *Psychonomic bulletin review*, 16(1):22–42, 2009. URL `http://www.ncbi.nlm.nih.gov/pubmed/19145007`.

T. Cover and J. Thomas. *Elements of information theory*. Wiley, New York, 1991.

B. L. Cutler and S. D. Penrod. *Mistaken identification: The eyewitness, psychology, and the law*. New York: Cambridge University Press, 1995.

H. Daniel. Analysis of hair samples of mummies from semna south. *AM. J. PHYS. ANTHROP*, 49:277–282, 1978.

K. A. Deffenbacher, B. H. Bornstein, and S. D. Penrod. Mugshot exposure effects: Retroactive interference, source confusion, and unconscious transference. *Law & Human Behavior*, 30:287–307, 2006.

Weihong Deng, Jun Guo, Jiani Hu, and Honggang Zhang. Comment on "100 *Science*, 321. no. 5891:912, 2008.

R. Diamond and S. Carey. Why faces are and are not special: An effect of expertise. *J Exp Psychol Gen*, 115, no. 2:107–117, 1986.

C. S. Dodson, M. K. Johnson, and J. W. Schooler. The verbal overshadowing effect: Why descriptions impair face recognition. *Memory and Cognition*, 25:129–139, 1997.

H. D. Ellis, J. W. Shepherd, and G. M. Davies. Identification of familiar and unfamiliar faces from internal and external features: some implications for theories of face recognition. *Perception*, 8 (4):431–439, 1979.

K. Faez, S. Motamed, and M. Yaqubi. Personal verification using ear and palm-print biometrics. In *SMC*, pages 3727–3731, 2008.

TB Fitzpatrick. Soleil et peau. *J Med Esthet*, 2:330–334, 1975.

J. A. Fodor. *Modularity of mind.* Cambridge. MA: MIT Press, 1983.

I. H. Fraser, G. L. Craig, and D. M. Parker. Reaction time measures of feature saliency in schematic faces. *Perception*, 19, no. 5:661–673, 1990.

F. N. Fritsch and R. E. Carlson. Monotone piecewise cubic interpolation. *SINUM*, 17:238–246, 1980.

C. Frowd, J. Park, A. McIntyre, V. Bruce, M. Pitchford, S. Fields, M. Kenirons, and P.J.B. Hancock. Effecting an improvement to the fitness function. how to evolve a more identifiable face. In *BLISS '08*, pages 3 –10, 4-6 2008. doi: 10.1109/BLISS.2008.28.

N. Furl, P. J. Phillips, and A. J. O'Toole. Face recognition algorithms and the other-race effect: computational mechanisms for a developmental contact hypothesis. *Cognitive Science*, 26:797–815, 2002.

H.K. Galoogahi and T. Sim. Inter-modality face sketch recognition. In *Multimedia and Expo (ICME), 2012 IEEE International Conference on*, pages 224–229, 2012a. doi: 10.1109/ICME.2012.128.

H.K. Galoogahi and T. Sim. Face sketch recognition by local radon binary pattern: Lrbp. In *Image Processing (ICIP), 2012 19th IEEE International Conference on*, pages 1837–1840, 2012b. doi: 10.1109/ICIP.2012.6467240.

Xinbo Gao, Juanjuan Zhong, Jie Li, and Chunna Tian. Face sketch synthesis algorithm based on e-hmm and selective ensemble. *Circuits and Systems for Video Technology, IEEE Transactions on*, 18(4):487 –496, april 2008a.

Xinbo Gao, Juanjuan Zhong, Dacheng Tao, and Xuelong Li. Local face sketch synthesis learning. *Neurocomputing*, 71(10-12):1921 – 1930, 2008b.

Yongsheng Gao and M.K.H. Leung. Face recognition using line edge map. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(6): 764 –779, jun 2002. ISSN 0162-8828. doi: 10.1109/TPAMI.2002.1008383.

I Gauthier, M J Tarr, A W Anderson, P Skudlarski, and J C Gore. Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nature Neuroscience*, 1999.

Barbara J. Gillam, Barton L. Anderson, and Farhan Rizwi. Failure of facial configural cues to alter metric stereoscopic depth. *J. Vis.*, 9(1):1–5, 1 2009. ISSN 1534-7362. URL `http://journalofvision.org/9/1/3/`.

A. G. Goldstein and J. E. Chance. Memory for faces and schema theory. *Journal of Psychology*, 105:47–59, 1980.

K. Grill-Spector, N. Knouf, and N. Kanwisher. The fusiform face area subserves face perception, not generic within-category identification. *Nat Neurosci*, 7(5):555–562, May 2004. ISSN 1097-6256. doi: http://dx.doi.org/10.1038/nn1224. URL `http://dx.doi.org/10.1038/nn1224`.

R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-pie. In *FG '08*, pages 1–8, Sept. 2008. doi: 10.1109/AFGR.2008.4813399.

S. R. Gross, K. Jacoby, D. J. Matheson, N. Montgomery, and S. Patil. Exonerations in the united states 1989 through 2003. *Journal of Criminal Law & Criminology*, 95:523–560, 2005.

Jing-Ming Guo, Chen-Chi Lin, and Hoang-Son Nguyen. Face gender recognition using improved appearance-based average face difference and support vector machine. In *System Science and Engineering (ICSSE), 2010 International Conference on*, pages 637 –640, july 2010. doi: 10.1109/ICSSE.2010.5551728.

Mark F. Hansen and Gary A. Atkinson. Biologically inspired 3d face recognition from surface normals. *ICEBT*, 2:26–34, 2010.

J. V. Haxby. The distributed human neural system for face perception. *Trends Cogn. Sci.*, 4:223–233, June 2000.

K. F. Hilliar and R. I. Kemp. Barack obama or barry dunham? the appearance of multiracial faces is affected by the names assigned to them. *Perception*, 37:1605–1608, 2008.

C. R. Huff, A. Rattner, and E. Sagarin. *Convicted but innocent*. Thousand Oaks, CA:Sage, 1996.

Yung hui Li, M. Savvides, and V. Bhagavatula. Illumination tolerant face recognition using a novel face from sketch synthesis approach and advanced correlation filters. In *ICASSP 2006*, volume 2, pages II –II, 14-19 2006. doi: 10.1109/ICASSP.2006.1660353.

David J. Hurley, Mark S. Nixon, and John N. Carter. Force field feature extraction for ear biometrics. *Computer Vision and Image Understanding*, 98(3):491–512, 2005.

A. Iannarelli. Ear identification. 1989.

R. Imhofer. Die bedeutung der ohrmuschel fur die feststellung der identitat. *Archiv fur die Kriminologie*, 26:150–163, 1906.

K. Iwata, K. Yamamoto, K. Kato, T. Suzuki, and M. Umemura. Construction of the security system using the front face detection by the average face. In *SICE 2002. Proceedings of the 41st SICE Annual Conference*, volume 1, pages 27 – 32 vol.1, aug. 2002. doi: 10.1109/SICE.2002.1195175.

Corentin Jacques and Bruno Rossion. The initial representation of individual faces in the right occipito-temporal cortex is holistic: Electrophysiological evidence from the composite face illusion. *J. Vis.*, 9(6):1–16, 6 2009. ISSN 1534-7362. URL `http://journalofvision.org/9/6/8/`.

I. Jarudi and P. Sinha. Relative contribution of internal and external features to face recognition. *MIT Libraries*, 2003.

Jingrong Jia, Lijun Yin, and J. Morrissey. On the importance of skin color for "other-race" effect. *ICME*, 1:399–402, 2004.

T.C. Jones and J.C. Bartlett. When false recognition is out of control: the case of facial conjunctions. *Memory & cognition*, 37(2):143–57, 2009. ISSN 0090-502X.

H. Kiani and T. Sim. Inter-modality face sketch receognition. *ICME*, 2012a.

H. Kiani and T. Sim. Face sketch recognition by local radon binary pattern: Lrbp. *ICIP*, 2012b.

B. Klare and A. Jain. Sketch to photo matching: a featurebased approach. *SPIE Conference on Biometric Technology for Human Identification*, 2010.

B. Klare, Zhifeng Li, and A.K. Jain. Matching forensic sketches to mug shot photos. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(3):639 –646, march 2011. ISSN 0162-8828. doi: 10.1109/TPAMI.2010.180.

N. Kruger L. Wiskott, J. M. Fellous and C. von der Malsburg. Face recognition by elastic bunch graph matching. In *Proc. International Conference on Image Processing*, volume 1, pages 129–132 vol.1, 1997. doi: 10.1109/ICIP.1997.647401.

M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Wurtz, and W. Konen. Distortion invariant object recognition in the dynamic link architecture. *IEEE Trans. Comput.*, 42 (3):300–311, 1993. doi: http://dx.doi.org/10.1109/12.210173. URL `http://dx.doi.org/10.1109/12.210173`.

& Carbon C.-C. Leder, H. When context hinders! learn-test compatibility in face recognition. *Quarterly Journal of Experimental Psychology*, 58A: 235–250, 2005.

D. A. Leopold, A. J. O'Toole, T. Vetter, and V. Blanz. Prototype-referenced shape encoding revealed by high-level aftereffects. *Nat Neurosci*, 4(1):89–94, January 2001. ISSN 1097-6256. doi: http://dx.doi.org/10.1038/82947. URL `http://dx.doi.org/10.1038/82947`.

D. A. Leopold, I. V. Bondar, and M. A. Giese. Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature*, 442(7102): 572–575, 2006.

DA Leopold, G Rhodes, KM Müller, and L Jeffery. The dynamics of visual adaptation to faces. *Proceedings of the Royal Society of London B*, 272: 897–904, 2005.

D. Levin. Race as a visual feature: Using visual search and perceptual discrimination tasks to understand face categories and the cross-race recognition deficit. *J Exp Psychol Gen*, 129:559–574, 2000.

D.S. Lindsay, P.C. Jack, and M.A. Christian. Other-race face perception. *Journal of Applied Psychology*, 76(4):587–589, 1991.

C. Liu, J. Yuen, A. Torralba, J. Sivic, and W. Freeman. Sift flow: Dense correspondence across different scenes. In *Proceedings of the 10th European Conference on Computer Vision: Part III*, ECCV 08, pages 28–42, Berlin, Heidelberg, 2008. Springer-Verlag.

Chengjun Liu. Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance. *TPAMI*, 28(5):725–737, 2006. ISSN 0162-8828. doi: http://dx.doi.org/10.1109/TPAMI.2006.90.

Qingshan Liu, Xiaoou Tang, Hongliang Jin, Hanqing Lu, and Songde Ma. A nonlinear approach for face sketch synthesis and recognition. In *CVPR 2005*, volume 1, pages 1005 – 1010 vol. 1, 20-25 2005. doi: 10.1109/CVPR.2005.39.

E. F. Loftus. *Eyewitness testimony.* Cambridge, MA: Harvard University Press., 1979.

E. F. Loftus. Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, 12:361–366, 2005.

E. F. Loftus, D. G. Miller, and H. J. Burns. Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning & Memory*, 4:19–31, 1978.

L. N. Smith M. F. Hansen, G. A. Atkinson and M. L. Smith. 3d face reconstructions from photometric stereo using near infrared and visible light. *ICEBT*, 114:942–951, 2010.

B.S. Manjunath, R. Chellappa, and C. von der Malsburg. A feature based approach to face recognition. *Computer Vision and Pattern Recognition, 1992. Proceedings CVPR '92., 1992 IEEE Computer Society Conference on*, pages 373 –378, Jun 1992. doi: 10.1109/CVPR.1992.223162.

A.M. Martinez and R. Benavente. The ar face database. Technical report, CVC Tech. Report, 24, 1998.

B. D. McCandliss, L. Cohen, and S. Dehaene. The visual word form area: expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences*, 7 (7):293–299, 2003.

E McKone, J L Brewer, S MacPherson, G Rhodes, and W G Hayward. Familiar other-race faces show normal holistic processing and are robust to perceptual stress. *Perception*, 36(2):224–248, 2007.

Ahmed M Megreya and A Mike Burton. Unfamiliar faces are not faces: Evidence from a matching task. *Memory & Cognition*, 34, no. 4:865–876, 2006.

Lynn Meijerman. *Inter- and intra individual variation in earprints.* PhD thesis, University Leiden, 2006.

Joseph M. Melcher and Jonathan W. Schooler. The misremembrance of wines past: Verbal and perceptual expertise differentially mediate verbal overshadowing of taste memory. *Journal of Memory and Language*, 35(2):231–245, 1996.

S. Milborrow and F. Nicolls. Locating facial features with an extended active shape model. *ECCV*, 2008.

C. A. Morgan, M. Hazlett, G. nad Baranoski, A. Doran, S. Southwick, and E.F. Loftus. Accuracy of eyewitness identification is significantly associated with performance on a standardized test of face recognition. *International Journal of Law & Psychiatry*, 30:213–223, 2007.

Hugo Munsterberg. *On the Witness Stand Essays on Psychology and Crime (Foundations of Criminal Justice).* Ams Pr Inc, 1927.

H Nejati and T. Sim. A study on recognizing non-artistic face sketches. *WACV*, pages 240–247, 2011.

H. Nejati, Li Zhang, T. Sim, E. Martinez-Marroquin, and Guo Dong. Wonder ears: Identification of identical twins from ear images. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 1201–1204, 2012.

Hossein Nejati, Terence Sim, and Elisa Martinez-Marroquin. Do you see what i see? a more realistic eyewitness sketch recognition. *Biometrics, International Joint Conference on*, pages 1–8, 2011.

A. J. O'Toole, K. A. Deffenbacher, H. Abdi, and J. C Bartlett. Simulating the "other-race" effect as a problem in perceptual learning. *Connection Science: Neural Comp, Artificial Intelligence & JCN*, 3:163–178, 1991.

Stephen E. Palmer. *Explorations in cognition*, chapter Visual perception and world knowledge: Notes on a model of sensory-cognitive interaction. Freeman, 1975.

D. I. Perrett, K. A. May, and S. Yoshikawa. Facial shape and judgements of female attractiveness. *Nature*, 368(6468):239–242, 1994.

P.J. Phillips, P.J. Flynn, K.W. Bowyer, R.W.V. Bruegge, P.J. Grother, G.W. Quinn, and M. Pruitt. Distinguishing identical twins by face recognition. In *FG 2011*, pages 185 –192, 2011. doi: 10.1109/FG.2011.5771395.

K. Pierce, RA. Muller, J. Ambrose, G. Allen, and E. Courchesne. Face processing occurs outside the fusiform 'face area' in autism: evidence from functional mri. *Brain*, 124:2059–73, 2001.

M.-M. Piquet-Thepot. *Bulletins et Mémoires de la Société d'anthropologie de Paris*, pages 207,208, 1968.

J. D. Pozzulo and S. Marciniak. Comparing identification procedures when the perpetrator has changed appearance. *Psychology, Crime & Law*, 12: 429–438, 2006.

S. Pramanik and D. Bhattacharjee. Geometric feature based face-sketch recognition. In *Pattern Recognition, Informatics and Medical Engineering (PRIME), 2012 International Conference on*, pages 409 –415, march 2012. doi: 10.1109/ICPRIME.2012.6208381.

K.H. Pun and Y.S. Moon. Recent advances in ear biometrics. In *FG*, pages 164 – 169, may 2004. doi: 10.1109/AFGR.2004.1301525.

J. D. Read. The availability heuristic in person identification: The sometimes misleading consequences of enhanced contextual information. *Applied Cognitive Psychology*, 9:91–121, 1995.

D. Ren, W. Chen, C. Hong Liu, and X. Fu. Identity processing in multiple-face tracking. *J. Vis.*, 9(5):1–15, 5 2009.

Ali M. Reza. Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement. *J. VLSI Signal Process. Syst.*, 38:35–44, 2004.

J. Sadr, I. Jarudi, and P. Sinha. The role of eyebrows in face recognition. *Perception*, 32:285–293, 2003.

Mukherjee-S. Thoresz K. & Sinha P. Sadr, J. The fidelity of local ordinal encoding. *In T. Dietterich, S. Becker & Z. Ghahramani (Eds.), Advances in Neural Information Processing Systems 14. MIT Press: Cambridge, MA.*, 2002.

M. I. Saleh, Wyatt C. L., and P. Schaumont. Eigen-face, eigen-ears. using ears for human identification. Master's thesis, Virginia Polytechnic Institute and State University, 2007.

B. Scheck, P. Neufeld, and J. Dwyer. Actual innocence: Five days to execution and other dispatches from the wrongly convicted. *New York: Doubleday*, 2000.

Laura Schmalzl, Romina Palermo, Melissa Green, Ruth Brunsdon, and Max Coltheart. Training of familiar face recognition and visual scan

paths for faces in a child with congenital prosopagnosia. *Cognitive Neuropsychology*, 25:704 – 729, 2008. ISSN 0264-3294. URL http://www.informaworld.com/10.1080/02643290802299350.

JW. Schooler and TY. Engstler-Schooler. Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology*, 22(1):36–71, January 1990.

A. Schwaninger, C.-C. Carbon, and H. Leder. Expert face processing: Specialization and constraints. in g. schwarzer & h. leder (eds.). *The development of face processing*, pages 81–97, 2003.

C. Shannon and W. Weaver. *The mathematical theory of communication*. University of Illinois Press Urbana, 1962.

A. Sharma and D.W. Jacobs. Bypassing synthesis: Pls for face recognition with pose, low-resolution and sketch. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 593–600, 2011. doi: 10.1109/CVPR.2011.5995350.

P. Sinha and T. Poggio. Last but not least: 'united' we stand. *Perception*, 31: 133, 2002.

P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell. Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proc. IEEE*, 94(11):1948–1962, 2006a. ISSN 0018-9219. doi: 10.1109/JPROC.2006.884093.

P. Sinha, B. J. Balas, Y. Ostrovsky, and R. Russell. Face recognition by humans. *Face Recognition: Models and Mechanisms*, 2006b.

R. Starrfelt and C. Gerlach. The visual what for area: Words and pictures in the left fusiform gyrus. *NeuroImage*, 35(1):334 – 342, 2007. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2006.12.003.

Xiaoou Tang and Xiaogang Wang. Face sketch synthesis and recognition. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 687 –694 vol.1, oct. 2003. doi: 10.1109/ICCV.2003.1238414.

Xiaoou Tang and Xiaogang Wang. Face sketch recognition. *T-CSVT*, 14(1): 50–57, jan. 2004. ISSN 1051-8215. doi: 10.1109/TCSVT.2003.818353.

S. Thorpe, D. Fize, and C. Marlot. Speed of processing in the human visual system. *Nature*, 381:520–522, 1996.

M. Treadway and M. McCloskey. Cite unseen: Distortions of the allport and postman rumor study in the eyewitness testimony literature. *Law and Human Behavior*, 11:19–25, 1987. ISSN 0147-7307. 10.1007/BF01044836.

M. Unnikrishnan. Koinophilia revisited: the evolutionary link between mate selection and face recognition. *Current Science*, 102:563–570, 2012.

M.K. Unnikrishnan. How is the individuality of a face recognized? *J. Theor. Biol.*, 261(3):469 – 474, 2009. ISSN 0022-5193. doi: DOI: 10.1016/j.jtbi.2009.08.011.

T. Valentine. A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quarterly Journal of Experimental Psychology*, 43(2):161–204, 1991.

G. Wallis and H. H. Bulthoff. Effects of temporal association on recognition memory. *Proc. Natl Acad Sci*, 98, no. 8:4800–4804,

April 2001. doi: http://dx.doi.org/10.1073/pnas.071028598. URL http://dx.doi.org/10.1073/pnas.071028598.

Xiaogang Wang and Xiaoou Tang. Face photo-sketch synthesis and recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31 (11):1955 –1967, nov. 2009. ISSN 0162-8828. doi: 10.1109/TPAMI.2008.222.

G. L. Wells. What do we know about eyewitness identification? *American Psychologist*, 48:553–571, 1993.

G. L. Wells, M. Small, S. Penrod, R. S. Malpass, S. M. Fulero, and C. A. E. Brimacombe. Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law & Human Behavior*, 22:603–647, 1998.

G. L. Wells, A. Memon, and S. D. Penrod. Eyewitness evidence: Improving its probative value. *Psychological Science in the Public Interest*, 7:45–75, 2006.

L. Wiskott and C. von der Malsburg. Face recognition by dynamic link matching. *Lateral Interactions in the Cortex: Structure and Function*, Chap. 11, 1996.

B. Xiao, X. Gao, D. Tao, and X. Li. A new approach for face recognition by sketches in photos. *Signal Process.*, 89(8):1576–1588, 2009. ISSN 0165-1684. doi: http://dx.doi.org.libproxy1.nus.edu.sg/10.1016/j.sigpro.2009.02.008.

L. Xu and J. Jia. Two-phase kernel estimation for robust motion deblurring. *ECCV*, pages 157–170, 2010.

P. Yan and K.W. Bowyer. Multi-biometrics 2d and 3d ear recognition. *Audio- and Video-based Biometric Person Authentication*, pages 503–512, 2005.

P. Yan and K.W. Bowyer. Biometric recognition using 3d ear shape. *TPAMI*, pages 1297–1308, 2007.

D. A. Yarmey. Probative v. prejudicial value of eyewitness memory research. *Expert Evidence*, 5:89–97, 1997. ISSN 0965-3643. 10.1023/A:1008818205752.

A. W. Young, D. C. Hay, K. H. McWeeny, B. M. Flude, and A. W. Ellis. Matching familiar and unfamiliar faces on internal and external features. *Perception*, 14:737–746, 1985.

A. W. Young, D. Hellawell, and D. C. Hay. Configurational information in face perception. *Perception*, 16:747–759, 1987.

Jiewei Zhang, Nannan Wang, Xinbo Gao, Dacheng Tao, and Xuelong Li. Face sketch-photo synthesis based on support vector regression. In *Image Processing (ICIP), 2011 18th IEEE International Conference on*, pages 1125–1128, 2011a.

L. Zhang, H. Nejati, L. Foo, K. T. Ma, D. Guo, and T. Sim. A talking profile to distinguish identical twins. *Automatic Face and Gesture Recognition (FG)*, 2013.

Wei Zhang, Xiaogang Wang, and Xiaoou Tang. Coupled information-theoretic encoding for face photo-sketch recognition. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 513–520, 2011b. doi: 10.1109/CVPR.2011.5995324.

Yong Zhang, C. McCullough, J.R. Sullins, and C.R. Ross. Hand-drawn face sketch recognition by humans and a pca-based algorithm for forensic applications. *SMC-A*, 40(3):475 –485, may 2010. ISSN 1083-4427. doi: 10.1109/TSMCA.2010.2041654.

J. Zhong, X. Gao, and Tian C. Face sketch synthesis using e-hmm and selective ensemble. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007.*

IEEE International Conference on, volume 1, pages I–485 –I–488, april 2007. doi: 10.1109/ICASSP.2007.366722.

B. Zhu, C. Chen, E. F. Loftus, C. Lin, Q. He, C. Chen, H. Li, G. Xue, Z. Lu, and Q. Dong. Individual differences in false memory from misinformation: Cognitive factors. *Memory*, 18:543–555, 2010.

Appendix

Justification of Bias Estimation Method

We defined the sketching bias as the combination of mental and drawing biases. We now choose an engineering approach to estimate and remove the sketching bias. We should first indicate that while the real process of face perception is not fully understand, the most likely suggestions are from the Exception Report Model (ERM) [Unnikrishnan, 2009, 2012]. The ERM suggests that face perception, memorization, and recognition is performed based on a norm-based model, that focuses on the facial features that are deviated from norm. Based on this model, when the eyewitness views a face, s/he compares this face to his/her individual mental norm, and detects the facial features that are deviated from norm. These deviated features are then used for memorization and future recognition. Therefore, the mental bias (the first component of the sketching bias) is added when the target face is viewed, and is incorporated into the memory of the face. This memory is then used to draw the Main Sketch, and this is where the drawing bias (the second component of the sketching bias) is added. Figure 6.2 illustrate this model in a pictorial representation.

Instead of estimating the mental and drawing bias separately, we can assume that the eyewitness memorized the face perfectly as it is, but when wants to draw it s/he adds both mental and drawing biases at the same time. By this

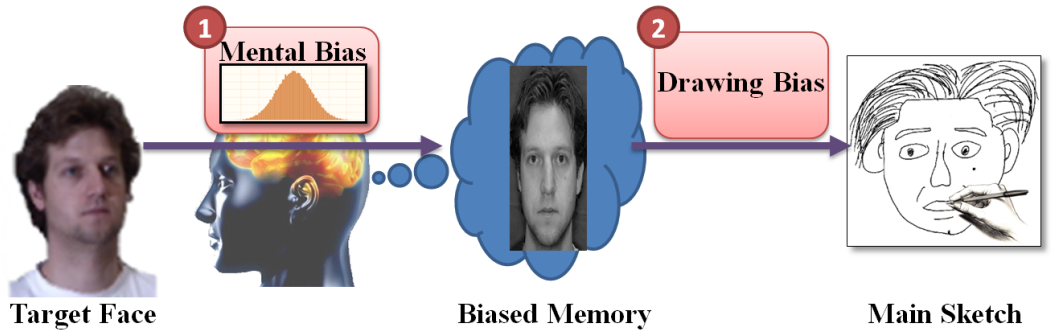

Figure 6.2: The pictorial representation of mental and drawing biases added to the memory and finally the Main Sketch (concluded by the concepts of ERM). The mental bias is added during face perception, base on the mental norm, resulting in a biased memory. Then this memory is used using biased drawing skills to draw the Main Sketch.

assumption, we can divide the process of creating the sketch into two steps of:

1. An unbiased process of face perception and outline detection, g

2. A biases drawing process which adds both mental and drawing biases, h.

We can now easily replicate g using a facial component detection algorithm, and try to estimate h^{-1} as the function to remove sketching bias (including both mental and drawing biases). Figure 6.3 illustrates this approach in a pictorial representation.

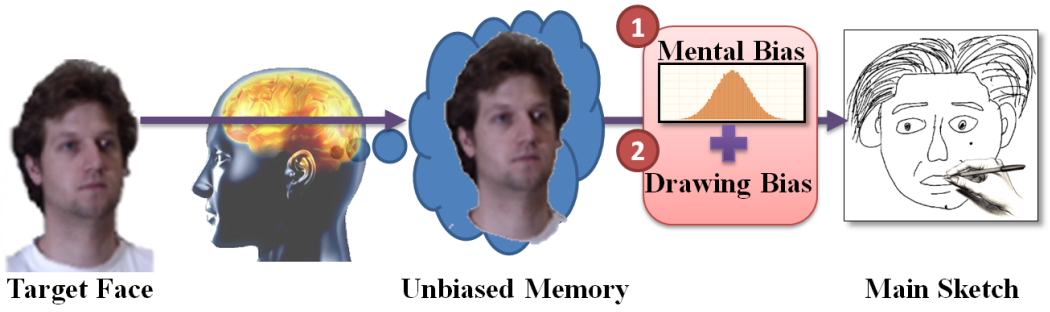

Figure 6.3: The pictorial representation of our perspective change to the process of creating the Main Sketch. We shift all the biases to the final step, while drawing from the memory. Thus we assume an unbiased memory, and addition of both mental bias and drawing bias during drawing from memory. Using this approach, while the final result (the Main Sketch) is not changed, we can easily estimate the sketching bias using the drawing profiles.