Do You See What I See? A More Realistic Eyewitness Sketch Recognition

Hossein Nejati and Terence Sim
National University of Singapore
School of Computing, National University of Singapore, Singapore 117417
{nejati,dcssimt}@nus.edu.sg
Elisa Martinez-Marroquin
La Salle School of Engineering - Ramon Llull University
St Joan de la Salle, 42. 08022 - Barcelona. Spain
elisa@salle.url.edu

Abstract

Face sketches have been used in eyewitness testimonies for about a century. However, 30 years of research shows that current eyewitness testimony methods are highly unreliable. Nonetheless, current face sketch recognition algorithms assume that eyewitness sketches are reliable and highly similar to their respective target faces. As proven by psychological findings and a recent work on face sketch recognition, these assumptions are unrealistic and therefore, current algorithms cannot handle real world cases of eyewitness sketch recognition.

In this paper, we address the eyewitness sketch recognition problem with a two-pronged approach. We propose a more reliable eyewitness testimony method, and an accompanying face sketch recognition method that accounts for realistic assumptions on sketch-photo similarities and individual eyewitness differences. In our eyewitness testimony method we first ask the eyewitness to directly draw a sketch of the target face, and provide some ancillary information about the target face. Then we build a drawing profile of the eyewitness by asking him/her to draw a set of face photos. This drawing profile implicitly contains the eyewitness’ mental bias. In our face sketch recognition method we first correct the sketch for the eyewitness’ bias using the drawing profile. Then we recognize the resulting sketch based on an optimized combination of the detected features and ancillary information. Experimental results show that our method is 12 times better than the leading competing method at Rank-1 accuracy, and 6 times better at Rank-10. Our method also maintains its superiority as gallery size increases.

I. Introduction

In 1984, based on the testimony of five eyewitnesses, Kirk Bloodsworth was convicted of the rape and murder of a nine-year-old girl and sentenced to the gas chamber. After Bloodsworth served nine years in prison, DNA testing proved him to be innocent [1]. Such devastating mistakes by eyewitnesses are not rare, and more than 75% of the convictions overturned through DNA testing since the 1990s were based on eyewitness testimony [2].

The eyewitness testimony for forensic applications has had a long history, with roots that go back to the beginning of the nineteenth century [3]. It has been an important tool for law enforcement agencies in difficult situations where no other clue is available. However, more than 30 years of psychological studies show that current eyewitness testimony methods are highly susceptible to error and should be reformed [4], [2], [5] (see figure 1 for some examples). For example, in the very first step of current eyewitness testimony methods (ETMs), the eyewitness provides a verbal description of the target face, not based on the real appearance of the face, but on own mental norm and bias [6]. Moreover, this verbal description itself degrades the eyewitness’ visual memory as well as the recognizability of the target face, in a phenomenon known as verbal overshadowing [7]. The eyewitness’ memory can also be altered due to misleading information arising from viewing similar faces or answering subjective questions raised by the police [8], [9]. Finally, piecewise reconstruction of the target face, specially in composite sketches, degrades the eyewitness’ mental image of the target face and the recognition result [10]. These problems are basically because human memory is fragile, malleable, and susceptible to suggestion [8], which in turn render
current ETMs unreliable. Despite this, current methods are still widely used by police departments all over the world. This is mainly due to the lack of viable alternative methods.

The resulting sketch from an eyewitness testimony, which we term the forensic sketch (which is drawn by a police artist, or produced by a photo-composite software), should be matched against the police database of faces. There exist several automatic face sketch recognition algorithms (FSRs) proposed in the literature to perform this matching, with some reporting recognition accuracies as high as 92% [12]. The main problem with all but one of these FSRs is that they were tested on exact artistic sketches, instead of forensic sketches. These exact artistic sketches bear great similarity with their target faces, including exact point-to-point geometry, skin texture, lighting, and even hairstyle (see figure 2, and the next section). In contrast, forensic sketches usually differ greatly from the target face, due to verbal overshadowing, memory degradation (due to questions, viewing other faces, etc), and piecewise reconstruction. However, as the result of this ETM is a non-artistic face sketch, we should process this crude representation of the target face to find the identity-specific features. In addition to the Main Sketch, we propose to ask the eyewitness to provide Ancillary Information about the target face, such as its gender, skin tone, ethnicity, etc (which is also practiced in real scenarios). Finally, we acquire an individual drawing profile for the eyewitness by asking him/her to sketch a set of face photos (in our experiments we asked eyewitnesses to draw at least 4 of these face photos).

In the FSR part, we first remove the noise and eyewitness’ mental bias from the Main Sketch. To do this, we estimate the true shape of the target face, from the Main Sketch, based on the eyewitness’ drawing profile—by learning individual drawing and feature representation style. We then linearly combine all features by optimally weighting different parts of the estimated shape and each of the Ancillary Information. Finally, we use this linear combination to match the modified Main Sketch against a gallery of face photos.

We test our algorithm which we call Non-Artistic Face Sketch Recognition (NAFSR), on a dataset of 100 sketches, hand drawn by 10 eyewitnesses, from 249 face photos of the Multi-PIE face dataset [14]. We compared the results of NAFSR with a PCA classification algorithm, an FSR introduced in [15], and a SIFT-based classification (used in [16]). In this comparison, NAFSR shows a significant improvement over the other methods. These results show that with a more reliable ETM, and by accounting for eyewitness bias in the FSR, we can recognize eyewitness sketches more accurately.

II. Related Works

The resulting sketch of an eyewitness testimony, the forensic sketch, is 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 this representation of the target face. Among the first to propose

![Figure 1. Some examples of unreliable artistic sketches (two left columns from http://depletedcranium.com) and composite sketches (two right columns [11]).](image)
an FSR algorithm were Tang et al. [15]. They proposed an eigenface transformation to transfer gallery face photos to sketch-like images. This transformation decreases the difference between the faces and the sketches, and results in better performance in PCA classification. The sketch-like images were then matched against an artistic sketch gallery using a PCA-based algorithm, with a reported recognition accuracy of 89%.

Another photo-to-sketch transformation method is proposed by Li et al. [17] in which eigenface transformation is similarly employed. In this method, a sketch-photo pair image is concatenated into a single vector to learn the PCA classifier towards an eigenspace with correlation to both the sketch and the real face. However, the test results of this work are limited to transformed version of the original image, and no real sketches.

Liu et al. [12] further improved the photo-to-sketch transformation using a non-linear transformation in which the photo is divided into patches and replacing each patch by the most similar patch from the sketch gallery. The patch similarity is calculated using a PCA-based classifier. The result of this transformation is then matched against an artistic sketch gallery, using Non-linear discriminant analysis with a reported accuracy of 92%.

Another improvement to patch based the photo-to-sketch transformation is introduced by Wang and Tang [18] in which a trained multi-scale Markov random field stitches and warps the patches into a final sketch. This final sketch is then used for sketch-to-sketch matching in order to find the target face, by a PCA-based algorithm.

While most of previous works have focused on photo-to-sketch transforming, Xiao et al. [19], [20] proposed an approach which exploits a sketch-to-photo transformation. In this work, the authors reconstructed a photo-like image from the provided artistic sketches, in order to transform the problem into a photo-to-photo matching problem. In this work, embedded hidden Markov model (EHMM) is used to reconstruct each sketch, with the most similar patches from the photo database. The resulting photo-like image is then classified using PCA.

Although the above algorithms are proposed to address the forensic sketch recognition problem, all of them have been tested on exact artistic sketches. As shown in figure 2, these exact artistic sketches have significant similarities to their target faces (including exactly similar facial component shape, illumination and shading, skin texture, and even hairstyle). In contrast, as figure 1 illustrates, a real forensic sketch from current eyewitness testimonies is very likely to be significantly different, due to verbal overshadowing, perception biases, piecewise reconstruction, etc (see [7], [4], [3], [2], [5]). Therefore, although the test results of the above methods show that they can accurately recognize exact artistic sketches, they provide no clue about their performance on recognizing forensic sketches.

The only work which has stepped further by testing forensic sketches is [16]. In this work, Klare et al. employed a fusion of SIFT features and multiscale local binary patterns to recognize some forensic sketches as well as exact artistic sketches. The interesting result of this work is that this algorithm reported to have 99.47% accuracy in matching exact artistic 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. These results support our argument that even if an algorithm can accurately recognize exact artistic sketches, it does not necessarily provide reliable results in recognizing forensic sketches. This work also tested a state-of-the-art face recognition software [21] (as a representative for conventional 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.

Based on the discussions in this section, we conclude three main points. First, although previous FSRs have reported high accuracy rates, they seem unreliable in recognizing real eyewitness sketches. Second, exact artistic sketches are not a proper estimation of real forensic sketches. And third, conventional face recognition methods cannot be applied to match forensic sketches.

III. Our Method

The problem of face sketch recognition has two main parts: (1) The eyewitness testimony method (ETM) in which a sketch is created, (2) the recognizing the resulting sketch from the ETM. Without a reliable ETM, the second part also provides unreliable results. Therefore, as all of the previous works have ignored the unreliability of current ETMs and merely focused on the final sketch, they could not accurately match forensic sketches.

Here we present a Non-Artistic Face Sketch Recognition (NAFSR) approach to address the face sketch recognition problem by accounting for both of its parts. We de-
scribe a psychologically plausible and reliable eyewitness testimony method, accompanied by a sketch recognition method, to faithfully retrieve and recognize eyewitness’ memory of the target face (figure 3).

A. Proposed Eyewitness Testimony Method

The basic task of an eyewitness testimony is to retrieve the eyewitness’ mental image of a face, and encode it into a usable format for other processes such as FSRs. Information retrieval is a critical step in eyewitness testimony in which memory disturbances should be avoided and individual differences should be accounted for. In the ETM part, we gather various types of information regarding each target face. First, we ask the eyewitness to draw a sketch of the target face by himself, called the Main Sketch, that includes outlines of facial features and facial marks (e.g. wrinkles, moles, or scars). Examples of these Main Sketches are presented in figure 4. By asking the eyewitness to draw the Main Sketch himself, we avoid introducing changes to the eyewitness memory. The Main Sketch is therefore 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 (see section I). On the other hand, the Main Sketch is a crude and non-artistic representation of the target face, which requires processing before being used (describe in section III-B).

In addition to the Main Sketch, we ask the eyewitness to provide Ancillary Information about the target face including skin color, iris color, hair color, estimated age, ethnicity, and gender. For each class of the Ancillary Information, we provide a set of predefined classes 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) [22], the iris color from Martin–Schultz scale color pallet (gray, blue, green, brown, dark brown, black, red) [23], and the hair color from Fischer–Saller scale color pallet (brown, black, blond, auburn, red, gray/white) [24]. For estimated age, we group ages in groups of 5 years (e.g. 1-5, 6-10, 11-15...). For the ethnicity, 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.

Finally, we obtain a drawing profile for each eyewitness by asking him/her to sketch a set of face photos. Here we
asked the eyewitnesses to draw at least 4 faces from a set of 30 face photos. The three outputs of the ETM is illustrated in figure 3.

B. Proposed Face Sketch Recognition

We define the sketch recognition process as finding photo(s) with the smallest the difference score to a given sketch, in the photo dataset (see figure 3, FSR section). The difference score is defined based on two of the three inputs from the ETM: the Main Sketch, the Ancillary Information. However, as shown in figure 4, the Main Sketch is a non-artistic representation of the target face, and requires processing before being used in the difference score. Therefore we use the third input from the ETM, the drawing profile, to estimate the true shape of the target face from the Main Sketch.

1) Target Face Shape Estimation: The Exception Report Model (ERM) of face recognition, recently proposed by Unnikrishnan [13], suggests that the human brain employs a modified norm based model in which attention is focused exclusively on the unusual features for rapid and effortless recognition of the target face. ERM is probably the underlying mechanism that plays a role in perceiving, remembering and recognizing faces [13]. While previous models 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 model are more economical than those employed by previous models of face recognition, which require exhaustive and highly accurate multiple feature-by-feature comparisons between different faces.

Motivated by ERM, we can separately account for normal and unusual facial features presented in the Main Sketches. The unusual nature of a facial feature is determined on the basis of the normal distribution of that feature. However, the normal distribution in the eyewitness’ mind (MND) can be different from a systematically calculated normal distribution (SND) based on the photo dataset. Moreover, two different eyewitnesses may have different MNDs based on their personal life experiences. When two individuals have different MNDs, they may differently perceive and remember a particular face as they may identify a different set of unusual features in that face. Similarly, when two eyewitnesses draw a face, they may present different features as normal or unusual. In addition, personal drawing styles should also be considered [25]. Thus, to be able to estimate “what the eyewitness means” from “what the eyewitness draws”, we should calculate an individually tailored mapping function, from each eyewitness’ MND, to our system’s SND.

In order to learn the mapping function from MND to SND, we learn the user’s individual drawing style based on his/her drawing profile. For each sketch-photo pair in the drawing profile, we first scale and rotate all sketches to the same size and angle based on the position of eye centers. We then (using a simple Active Shape Model) fit 16 Piecewise Cubic Hermite Interpolating Polynomial splines [26] to the outlines of 7 facial components, the eyes, the eyebrows, the nose, the mouth, and the jaw-line (2 to each eye, 2 to each eyebrow, 2 to the mouth, 3 to the nose, and 3 to the jaw-line). Each of these 16 splines are then divided into four parts (quarter splines), and each part is sampled in 25 equally distributed data points. Then we use these data points to train Support Vector Machine regressors for each quarter spline (i.e. 4 mapping function for each spline, 64 mapping function for the entire sketch). Finally, we estimate the true shape of the target face, by mapping each point in the Main Sketch by its respective quarter spline mapping function. The training SVRs and target face shape estimation is indicated in processes 1 and 2 in figure 3. After removing individual biases from the Main Sketch, we proceed to sketch recognition procedure.

2) Recognizing the Main Sketch: In order to recognize the Main Sketch (matching process), we define a difference score between a given sketch and a photo, based on the estimated target shape and the Ancillary Information. However, note that each of these features has a different nature, range, and significance to the matching process. Therefore, we should use a proper difference measurement for each feature (figure 3, processes 3, 4, and 5), normalize these measurements (figure 3, process 6), and finally combine them into a single score (figure 3, process 7).

The difference measurement of sketch and photo points is indicated by sum of squared error (SSE). We calculate three different SSEs, one for normal points, one for unusual points.
After calculating all feature differences, we normalize each feature difference by dividing each one by its respective standard deviation over the entire dataset (figure 3, process 6). Note that the spline differences require yet another normalization. Because of the differences in spline sizes, points do not cover the same amount of area from one spline to another (the eyes are much smaller than the jaw-line). Therefore, we have to normalize the effect of each spline based on its axis lengths too. Equations 1 to 3 illustrate the first regularization step:

\[
\Delta_{\text{unus}} = \left( \sum_{i=1}^{16} \sum_{j=1}^{100} \frac{\text{diff}(S_{\text{point}}, P_{\text{point}})}{\text{diam}_i} \right)
\]

\[
\Delta_{\text{mark}} = \sum \text{diff}(S_{\text{mark}}, P_{\text{edge}})
\]

\[
\Delta_{\text{set}}(S, P) = \left\{ \frac{\Delta_{\text{norm}}}{\sigma_{\text{norm}}}, \frac{\Delta_{\text{unus}}}{\sigma_{\text{unus}}}, \frac{\Delta_{\text{mark}}}{\sigma_{\text{mark}}} \right\}
\]

where \(\Delta_{\text{unus}}\) is unusual points feature difference; \(\text{diff}\) is the respective difference measurement function; \(\text{diam}_i\) is the diameter length of the \(i^{th}\) spline; \(\Delta_{\text{mark}}\) is the facial mark feature difference; \(\sigma_{\text{norm}}, \sigma_{\text{unus}},\) and \(\sigma_{\text{mark}}\) are standard deviations for normal points, unusual points, and facial marks respectively. Finally, \(\Delta_{\text{set}}\) is the normalized set of feature differences between sketch \(S\) and photo \(P\) after the normalization.

We finally optimally combine all features by calculating the coefficient matrix \(A\) that minimizes the sum squared error \(E\), given a set of sketch features \(S = \{s_1, \cdots, s_k\}\) and a set of photo features \(P = \{p_1, \cdots, p_k\}\):

\[
E = \sum_{i=1}^{k} \| A(\text{diff}(s_i, p_i)) \|^2
\]

\[
\text{final_score} = \| A\Delta_{\text{set}}(S, P) \|^2
\]

Where \(\text{final_score}\) is the difference measurement between sketch \(S\) and photo \(P\) (figure 3, process 7).

IV. Experimental Results

In this section we present the experimental results of applying NAFSR, on 100 non-artistic sketches. We also compare our classification results with PCA, the method presented by Tang et al. [15], and SIFT classification. Almost all previous methods have used a PCA-based algorithm; therefore, we can consider the PCA accuracies as a generalized measurement of their recognition accuracy. The FSR introduced in [15] is reported to have 71% rank-1 (and 96% rank-10) accuracy in recognizing exact artistic sketches. The SIFT classification is reported to have accuracy rate of 97% (rank-1) for exact artistic sketches and 16.33% (rank-1) for forensic sketches [16].

The photo dataset is the 249 faces of the first session of the Multi-PIE face dataset [14], having a mixture of gender, age, and ethnicity. Each photo is analyzed by STASM [27] to detect facial component outlines. For each photo, all Ancillary Information except age are assigned manually. The age feature is assigned to the average estimated age of 16 individuals (none are eyewitnesses), in terms of 5-year bins.

The sketch dataset consists of 100 non-artistic sketches of random target faces from the photo dataset, drawn by 10 eyewitnesses, while looking at the face photos, with
unlimited time to deliver (similar to all previous works). We use a simple active shape model to detect facial component outlines, and normalize all photos and sketches based on the eye positions. We estimate the target face shape of each sketch, by treating the rest of the sketches from the same eyewitness as the drawing profile.

Figure 6 and table I compare the results of NAFSR, PCA, the FSR proposed in [15], and SIFT classification. The input to these methods is the Main Sketch. In addition we also test the PCA on the estimated target shapes. The CMC curves show the accuracies up to rank-50 (50 guesses); however, in real cases the possible number of guesses can be around 10. Note that the traditional recognition accuracy is the rank-1 value in the CMC curve.

Based on the results, PCA performs poorly in recognizing both the Main Sketches and estimated target shapes, with its rank-50 accuracy remains below 10% and 20% respectively. Similarly, Tang et al. and SIFT classifications have poor recognition accuracies. In contrast, the recognition accuracy of NAFSR starts with 12% in rank-1 with a rapid growth as rank increases (60% in rank-10).

Figure 7 demonstrates how the rank-1 recognition accuracies of NAFSR, Tang et al., and SIFT decrease as the gallery size increases. This figure indicates the scalability of our method to be used in datasets of very large size.

It should be noted that the direct comparison between these methods is not completely possible as the input and assumptions are different. All previous methods have used artistic sketch as their input and assumed high sketch-photo similarity. In contrast, we use non-artistic sketch as input, directly drawn by the eyewitnesses and we have a realistic assumption for differences between these sketches and the target faces. The non-artistic sketches have a very larger modality difference to face photos, than the exact artistic sketches. Therefore it can be expected beforehand that previous methods perform poorly. However, considering the reliability of these non-artistic sketches (see section III-A), our results show that if processed carefully, they can be used to find the target face. And the key to a proper processing of these sketches, lies in assessing the procedure of creating them (ETM) - which is totally ignored by previous works.

V. Summary and Conclusion

More than 30 years of psychological studies show that forensic sketches are highly unreliable due to problems such as verbal overshadowing in the very first steps of eyewitness testimony methods (ETMs) [7], and piecewise reconstruction in the next steps [10]. However, no practical alternative ETM is successfully proposed and police departments continue to use the unreliable traditional methods. We showed that the previous works for recognizing forensic sketches are unreliable in handling real cases because of their unrealistic assumptions of forensic sketch reliability and sketch/photo similarity (section II).

Here we proposed both a more reliable ETM and an accompanying face sketch recognition method (FSR) which account for more realistic situations. We are the first to provide a solution for face sketch recognition by not only assessing the face sketch, but also focusing on how the sketch is created in the ETM. And we are the first to account for individual eyewitness differences in mental biases and information delivering style.

In our ETM we asked the eyewitness to sketch the
target face by him/herself, and therefore avoided introducing memory disturbances such as verbal overshadowing, piecewise reconstruction, or addition of external memories. Then we asked the eyewitness to provide some additional information about the target face, such as skin tone and ethnicity. We finally acquired a drawing profile of the eyewitness based on his/her sketches of a set of faces.

In our FSR, we estimated the shape of the target faces from the eyewitness sketch, based on the eyewitness’s drawing profile (section III-A). Then we optimally combine all features such as facial component shape unusuality and relative distances into a single difference score, using global least squares optimization (section III-B). We finally compared our method with PCA, and two other proposed methods in [15] and [16] which reported high accuracy (up to 97%) in recognizing exact artistic face sketches. Results of recognizing 100 non-artistic sketches from 10 eyewitnesses show that our method has a significant improvement over other methods, with a fast growth in accuracy as the rank increases (section IV).

Furthermore, this combination of ETM/FSR can be easily used in real situations as it requires no trained users, and delivers more reliable results in a relatively short time.

Nonetheless, it is clear that improvements such as better ways of target face shape estimation or addition of new sources of information can further improve the final recognition accuracy. For example, further assessment on perception of race, skin color, or age by individuals may produce more accurate estimation of mental biases and therefore better matching rates. In addition, a more realistic situation can be tested by using sketches drawn merely based on eyewitness’ memory, when there is a time delay between viewing the face and sketching.

The main motivation for this work was to create the missing link between psychological findings, automatic face sketch recognition, and real world applications, and therefore reduce the chance of wrongful convictions of innocents, such as the Kirk Bloodsworth. We hope that this work serves as a first step for better methods benefiting both computer vision and forensic sciences.

References