Face Recognition from Multiple Stylistic Sketches: Scenarios, Datasets, and Evaluation

Chunlei Peng, Nannan Wang$^{(\boxtimes)}$, Xinbo Gao, and Jie Li

State Key Laboratory of Integrated Services Networks,
Xidian University, Xi'an, China
clp.xidian@gmail.com, nnwang@xidian.edu.cn,
{xbgao,leejie}@mail.xidian.edu.cn

Abstract. Matching a face sketch against mug shots, which plays an important role in law enforcement and security, is an interesting and challenging topic in face recognition community. Although great progress has been made in recent years, main focus is the face recognition based on SINGLE sketch in existing studies. In this paper, we present a fundamental study of face recognition from multiple stylistic sketches. Three specific scenarios with corresponding datasets are carefully introduced to mimic real-world situations: (1) recognition from multiple hand-drawn sketches; (2) recognition from hand-drawn sketch and composite sketches; (3) recognition from multiple composite sketches. We further provide the evaluation protocols and several benchmarks on these proposed scenarios. Finally, we discuss the plenty of challenges and possible future directions that worth to be further investigated. All the materials will be publicly available online (Available at http://chunleipeng.com/FRMSketches.html.) for comparisons and further study of this problem.

Keywords: Face recognition · Viewed sketch · Composite sketch · Fusion

1 Introduction

In criminal investigation, face sketch recognition is an essential technique for real-world situations when the photo of the suspect is unavailable or is captured under poor quality. A face sketch is usually generated by the forensic artist [40] or facial composite software [10] based on the information provided by an eyewitness, victim, or poor quality surveillance videos. The sketch is the only clue to identify the suspect. Due to the large domain gap between face sketches and photos, face recognition from a probe sketch remains a challenging and prevalent topic in the community.

The technique of matching a face sketch against photos has been extensively studied in recent years. Existing approaches can be generally classified into four categories based on the types of sketches used: hand-drawn viewed sketch [36,40], hand-drawn semi-forensic sketch [5,25], hand-drawn forensic sketch [15,16,30], and software-generated composite sketch [10]. Early study mainly focuses on

viewed sketch, which is drawn by viewing the photo directly. Because the viewed sketch is relatively reliable for identifying the subject, saturated performance [7, 29–31, 41, 42] has been achieved on the viewed sketch benchmark (CUHK face sketch database, CUFS [40]). However, face sketches are usually unreliable in real-world situations due to the domain gap caused by perceptual bias, descriptive bias, and generating bias [24, 25]. In order to better understand face sketch recognition in real-worlds, semi-forensic sketch is introduced recently, which is drawn based on the recall of the artist after viewing the photo a few minutes ago. Models trained on semi-forensic sketches have shown their possibility to improve face sketch recognition in practice [5, 25]. Forensic sketch is drawn by a forensic artist according to the description of an eyewitness or victim. Forensic sketch recognition is relatively a far from being well studied and solved problem because of the unreliability of forensic sketch. Software-generated composite sketch is widely used in many law enforcement agencies [10]. This is because it is more convenient and efficient to generate a composite sketch by using software than training a skilled forensic artist. However, the composite sketch is composed of isolated facial components. Combining these components to form a face image introduces additional bias than drawing a forensic sketch with pencil in a continuous way. This just makes matching composite sketch with photo being one of the most challenging problems.

Despite the extensive studies on matching a face sketch against photos, they mainly focus on SINGLE sketch based face recognition. However, single sketch can be unreliable in real-world situations[1] according to psychology study [6, 24]. Sometimes this unreliability could be at heightened risk for false identification[2]. Researchers have noticed the unreliability of single sketch in recent years. Nejati *et al.* [24] carefully analyzed the existing biases in matching sketch with photo, and proposed a two-step bias modeling framework. Ouyang *et al.* [26] introduced the fusion of attributes with low-level features to deal with the cross-modal gap. However, these approaches still cannot cope with the defect of relying on single sketch. In forensic investigations multiple sources of information is available about the suspect [4, 9], such as verbal descriptions provided by multiple witnesses or victims, or information obtained from both the verbal description and poor quality surveillance video tracks. These clues can be used to generate multiple face sketches and figure the suspect in different ways with complementary information. Even if there is only one version of description about the suspect, multiple forensic artists can be invited to generate hand-drawn sketches under different styles [8], or multiple software tools can be exploited to generate different stylistic composite sketches [22]. The combination usage of both hand-drawn sketches and composite sketches does exist in real-world cases. The usage of multiple stylistic sketches can be helpful in improving the recognition performance, which has not been rigorously defined and evaluated in the community.

---

[1] http://www.nbcnews.com/news/crime-courts/dying-art-forensic-sketch-artists-face-digital-future-n41421.

[2] http://www.scientificamerican.com/article/do-the-eyes-have-it.

This paper presents a study of face recognition from multiple stylistic sketches. To the best of our knowledge, this is the first study on this essential problem. According to different generation procedures, *i.e.* hand-drawn sketch and software-generated composite sketch, we define three specific scenarios with corresponding datasets and protocols: (1) recognition from multiple hand-drawn sketches drawn by different artists; (2) recognition from a hand-drawn sketch and multiple composite sketches produced by different software tools; (3) recognition from multiple composite sketches from different software tools. We further provide several baseline performances of these three scenarios under pre-defined evaluation protocols. After that we discuss the challenges and possible directions that worth to be investigated in the future, thus making a good start point for research on face recognition from multiple stylistic sketches. All the related materials can be downloaded online (the website will be available after the blind review process) to boost further study on this problem.

In this paper, we make the following three contributions. (1) We present a fundamental study of face recognition from multiple stylistic sketches, and three specific scenarios with corresponding datasets are carefully defined to mimic real-world situations; (2) we provide the evaluation protocols and several benchmarks to address these scenarios, which opens new possibilities of research and further experiments for the benefit of face sketch recognition community; (3) significant challenges and several possible research directions are discussed, which can stimulate future research on this topic.

The remainder of this paper is organized as follows. We first describe related face sketch datasets and representative face sketch recognition methods in Sect. 2. We then define the problem of face recognition from multiple stylistic sketches and three specific scenarios in Sect. 3. Baseline approaches evaluated on the proposed scenarios are introduced in Sect. 4. Evaluations are given in Sect. 5. Conclusion and promising directions are presented in Sect. 6.

2   Related Work

2.1   Face Sketch Recognition

We briefly review representative face sketch recognition methods in this subsection. Existing approaches can be generally classified into four categories based on the types of sketches used: hand-drawn viewed sketch, hand-drawn semi-forensic sketch, hand-drawn forensic sketch, and software-generated composite sketch.

Face sketch recognition started from the seminal work of face sketch synthesis based recognition [36]. Representative approaches include the eigentransformation algorithm [36], the locally linear embedding approach [18], the Markov random field (MRF) based method [40] and a series of improved Markov networks based photo-sketch synthesis methods [29,31,38,44]. However, most of the synthesis-based approaches mainly focused on viewed sketches except for the multiple representations based method [31] evaluated on forensic sketch-photo synthesis and recognition. Later common space projection based approaches were exploited for face sketch recognition. Representative approaches include the

partial least squares (PLS) based method [34] and the multi-view discriminant analysis (MvDA) method [14]. Researchers also attempted to design modality-invariant feature descriptor for viewed sketch recognition, such as the coupled information-theoretic encoding (CITE) based face descriptor [42] and the local radon binary pattern (LRBP) [7].

Bhatt *et al.* [5] firstly introduced the semi-forensic sketches for performance evaluation and bridging the gap between viewed sketches and forensic sketches. They proposed a multiscale circular Weber's local descriptor (MCWLD) and mimetically optimized χ^2 distance for face sketch recognition. Recently, Ouyang *et al.* [25] proposed a memory-aware approach with the corresponding memory gap database. The 1 hour-sketch and 24 hour-sketch in [25] are two types of semi-forensic sketches, which were proven to be helpful for forensic sketch recognition.

The first study to face recognition based on forensic sketch was proposed by [13]. Klare *et al.* proposed a series of frameworks for matching forensic sketches to mug shot photos [15,16]. They further designed the FaceSketchID system [17] for face sketch recognition. Peng *et al.* [30] proposed a graphical representation based method recently. Notice that [15,17,30] evaluated their recognition algorithms on multiple types of sketches respectively including viewed sketches, forensic sketches and composite sketches.

A number of software-generated composite sketch recognition methods were proposed. Han *et al.* [22] firstly proposed a component based approach considering the fact that composite sketches are generated through combining facial components in software tools. Mittal *et al.* proposed a transfer-learning based deep learning approach [23] for composite sketch recognition.

There are several works involved multiple stylistic sketches. Zhang *et al.* [43] studied the fusion of sketches drawn by different artists and compared the performances of humans and a principal component analysis (PCA)-based algorithm. However, the dataset and protocols used in [43] is not available, and they merely used PCA-based algorithm. Gao *et al.* [8] evaluate their sparse representation based face sketch-photo synthesis method on the VIPSL dataset. Multiple stylistic sketches were involved for training, but only single sketch was utilized during the test procedure. In this paper, we assume that multiple stylistic sketches can still be available in test and real-world scenarios. Recently Mittal *et al.* [22] proposed composite sketch recognition using saliency and attribute feedback. Multiple stylistic composite sketches were combined for improving the matching performance. Their experiments showed the latent capacity of face recognition from multiple stylistic sketches. However, they merely involved multiple composite sketches and ignored other styles of sketches like hand-drawn sketches. On the other hand, they merely evaluated three fusion strategies (score-level, rank-level and decision-level) and other strategies, *e.g.* pixel-level and feature-level, were ignored. In this paper, we will conduct extensive evaluations on a variety of baselines with all protocols publicly available for further experiments on face recognition from multiple stylistic sketches.

Table 1. Summary of existing face sketch datasets.

Dataset	Photo	Sketch	Description	Sources of photos
CUFS [40]	606	606	Hand-drawn viewed sketch	CUHK student database, AR [20], and XM2VTS [21]
CUFSF [42]	1,194	1,194	Hand-drawn viewed sketch	FERET [32]
IIIT-D Viewed [5]	238	238	Hand-drawn viewed sketch	FG-NET [1], LFW [12], IIIT-D student & Staff
IIIT-D Semi-Forensic [5]	140	140	Hand-drawn semi-forensic sketch	Same with IIIT-D Viewed database
IIIT-D Forensic [5]	190	190	Hand-drawn forensic sketch	Lois Gibson [9], Karen Taylor [37], and Internet
PRIP-HDC [17]	47	47	Hand-drawn forensic sketch	Internet
PRIP-VSGC [17]	123	123	Software-generated composite sketch	AR [20]
E-PRIP [23]	123	123	Software-generated composite sketch	AR [20]
Memory Gap Database [25]	100	400	Hand-drawn sketch	Internet (mugshots.com)
VIPSL [8]	200	1,000	Hand-drawn viewed sketch	FERET [32], FRAVG2D [33] Indian Face Database [2]

2.2 Face Sketch Datasets

We summarize existing face sketch datasets as shown in Table 1. CUHK face sketch database (CUFS) [40] and CUHK face sketch FERET database (CUFSF) [42] are two publicly available benchmarks provided by the multimedia lab in the Chinese University of Hong Kong (CUHK). These two benchmarks have contributed to the great progress in both face sketch synthesis [39] and face sketch recognition [30] in recent years. Saturated performances have been achieved [27] on these two viewed sketch databases due to the sketches in CUFS and CUFSF are captured under relatively controlled conditions. IIIT Delhi image analysis and biometrics lab published a more challenging IIIT-D Sketch database, which is composed of three types of sketches, namely IIIT-D viewed, IIIT-D semi-forensic, and IIIT-D forensic sketches. The IIIT-D semi-forensic sketches are introduced to bridge the large gap between viewed sketches and forensic sketches. With the great progress on viewed sketches, researchers begin to focus

on more challenging real-world scenarios, such as hand-drawn forensic sketches and software-generated composite sketches. Because forensic sketches usually come from real-world criminal investigations, it is quite difficult to obtain a large scale dataset of forensic sketches. Existing methods often use images from two scanned textbooks [9,37] and other Internet sources. Biometrics research group in Michigan State University published their 47 forensic sketch pairs from the Internet [17]. Considering the fact that law enforcement agencies are now using software tools to generate composite sketches, two software-generated composite sketch datasets PRIP Viewed Software-Generated Composite dataset (PRIP-VSGC) [17] and extended PRIP (E-PRIP) [23] are created. Both the two composite sketch datasets utilized the same 123 photos from the AR dataset [20], and create composites using two kinds of software named Identi-Kit[3] and FACES[4] respectively. Recently a memory gap database[5] was released to investigate the effects of forgetting process and communication process in face sketch recognition. There are 100 subjects in this database, and each subject has four types of sketches drawn after different times of delay. There is another VIPSL dataset [8] which also contain multiple styles of sketches per subject. There are 200 face photos in the VIPSL database, and five sketches are drawn by five different artists for each photo. As shown in Table 1, in several datasets there are multiple stylistic sketches per person, like the memory gap dataset and the VIPSL dataset. There are also several datasets whose photos come from the same source. For example, the AR dataset is the photo source of CUFS, PRIP-VSGC, and E-PRIP. These datasets provide fundamental resources for our study of face recognition from multiple stylistic sketches.

3   Face Recognition from Multiple Stylistic Sketches

3.1   Overview

In this section, we provide an overview of face recognition from multiple stylistic sketches. As motivated in Sect. 1, we aim to present a fundamental study of face recognition from multiple stylistic sketches. Considering there are a variety of styles of sketches, we specify three scenarios as follows:
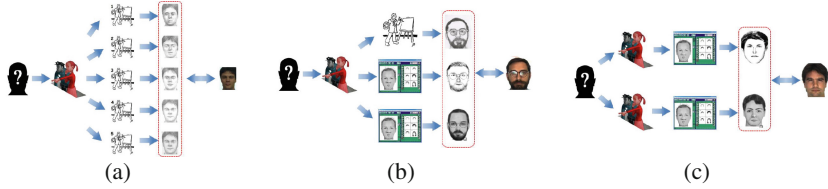
(1) **Scenario-MHS:** recognition from multiple hand-drawn sketches drawn by different artists;
(2) **Scenario-MHCS:** recognition from a hand-drawn sketch and multiple composite sketches produced by different software tools;
(3) **Scenario-MCS:** recognition from multiple composite sketches from different software tools.

   Illustrations of these three scenarios are shown in Fig. 1. We will provide detailed explanations and corresponding dataset settings later in this section. The specific protocols will be introduced in the experimental setup section.

---

[3] http://www.identikit.net/.
[4] http://www.iqbiometrix.com/.
[5] Because this dataset is relatively new, it is still unavailable currently.

Fig. 1. Illustrations of the proposed three scenarios of face recognition from multiple stylistic sketches. (a) Scenario-MHS; (b) Scenario-MHCS; (c) Scenario-MCS. Note that the composite software symbols come from FACES, and some other symbols come from the Internet.
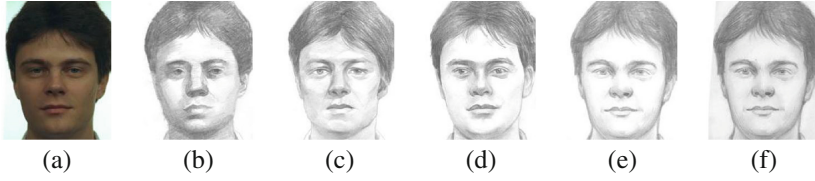
3.2   Scenario-MHS

In this scenario, we consider the situations in which multiple hand-drawn sketches per subject can be available. For example, there are circumstances that multiple witnesses may be interviewed to help produce hand-drawn sketches. Because these witnesses experienced the incident together, their descriptions about the subject will share common information. But each person has his own way of face perception and face description. Therefore, there is complementary information among these different sketches. Another circumstance is when both the witness and surveillance video tracks are available. Multiple sketches can also be drawn based on these different information sources respectively. Even if there is only one clue about the suspect, law enforcement agencies can invite multiple artists to drawn sketches respectively. The artists have different drawing skills and experience, thus multiple stylistic sketches can be obtained.

In order to evaluate this multiple hand-drawn sketches scenario, we adopt the VIPSL dataset[6] [8]. This dataset consists of 200 photos and 1,000 sketches. For each subject, 5 different styles of sketches are drawn by 5 different artists. The photo come from different face databases. There are different skin colors, background colors, and lighting variations in photos of this dataset, and the sketches are drawn with shape exaggeration. Because the sketches are drawn by different artists, there are different styles of sketches in VIPSL, which is appropriate to mimic this scenario. Examples of the sketches utilized in Scenario-MHS are shown in Fig. 2.
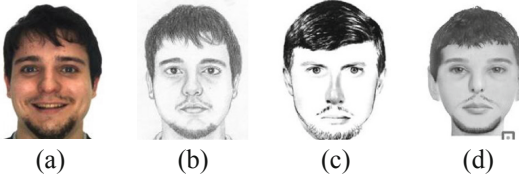
3.3   Scenario-MHCS

This scenario involves both hand-drawn sketches and software-generated composite sketches. More and more law enforcement agencies are using software tools to create composite sketch. Considering the fact that it usually spends years to train a forensic artist, it only takes a few hours to get used to composite-generation software. Therefore, besides the forensic artist, law enforcement agencies can exploit software tools in criminal investigation as well. Furthermore,

---

[6] The VIPSL dataset is available at: http://www.ihitworld.com/index.php/vipsl-database.

Fig. 2. Examples of the photo and the sketches used in Scenario-MHS. (a): Photo of a subject. (b–f): Five hand-drawn sketches of (a) from five artists.

Fig. 3. Examples of the photo and the sketches used in Scenario-MHCS. (a): Photo of a suspect. (b): Hand-drawn sketch of (a). (c)–(d): Two composite sketches of (a).

there exist imperfect communications of the memory of witness [25] when the witness describes the suspect. But the witness can create the composite sketch by himself, thus skip the communication bias.

We utilize part of the CUFS dataset (sketches drawn based on the AR dataset), the PRIP-VSGC dataset, and E-PRIP dataset to simulate this scenario. There are 123 photos from the AR dataset. Because the hand-drawn sketches in CUFS are created strictly based on the AR photos, the sketches and photos in CUFS have exactly the same facial contour, shading, and even hairstyle. This is impossible in real-world conditions. In order to mimic real-world scenario, we randomly replace the photos with another photo of the same identity in the AR dataset. 123 hand-drawn sketches corresponding to the AR identities are used here. The composite sketches in PRIP-VSGC and E-PRIP were created by two different software. Example images used in this scenario are shown in Fig. 3. It can be seen that the hand-drawn sketch contains more texture information, while the contour information in the composite sketch is more distinct. Therefore, there is complementary information in these different stylistic sketches which is favorable to the recognition task.

3.4   Scenario-MCS

According to our conversations with law enforcement agencies, there are many agencies that do not have forensic artists. In contrast, composite-generation software is easier to be obtained. This scenario only involves software-generated composite sketches. In this scenario, multiple witnesses can create multiple composite sketches to figure out the suspect in their memory. Multiple software tools can also be utilized for composite sketch generation. Because the styles of

Table 2. Statistics of the proposed three scenarios.

Scenario name	Photo	Hand-drawn sketch	Software-generated composite sketch
Scenario-MHS	200	1,000	0
Scenario-MHCS	123	123	246
Scenario-MCS	123	0	246

facial components are different in these software tools, the obtained composite sketches vary a lot. This complementary information can be useful for identifying the suspect.

We utilize the PRIP-VSGC dataset and E-PRIP dataset in this scenario. The photos used are the same with Sect. 3.3. For each photo, there are two corresponding composite sketches respectively. Examples are shown in Fig. 3(a), (c), and (d).

Table 2 presents the general statistics of the proposed three scenarios.

4   Baseline Approaches

4.1   Baseline Face Recognition Approaches

To give a benchmark for future comparisons under the settings on aforementioned datasets in above three scenarios, we have conducted experiments using three baseline face recognition algorithms.

(1) **Basic LBP-based face recognition:** We implemented the face recognition algorithm based on Local Binary Pattern (LBP) based texture feature [35]. We modify the original strategy in [35] by replacing LBP with a multi-scale version. There is no training procedure in this method.

(2) **Fisherface:** We then utilize the supervised algorithm Fisherface [3] as baseline approach. We implemented the Fisherface with part of the codes available online[7]. Within class whitening is added to improve the performance.

(3) **VGG-Face:** A convolutional neural networks (CNNs) based face recognition algorithm is further taken as the baseline. We use the pre-trained model VGG-Face [28] available in the MATLAB toolbox MatConvNet. Because the face sketch datasets are relatively small, it is impossible to train a ConvNet from scratch. It is not practical to fine-tune through the network which may lead to overfitting. We therefore utilize the high-level features in the Convnet as deep features. In our experiments, removing the last three fully connected layers can achieve the best performance of 99 % accuracy on CUFS with cosine distance as the similarity metric.

---

[7] http://www.cad.zju.edu.cn/home/dengcai/Data/DimensionReduction.html.

4.2   Baseline Fusion Approaches

We present the fusion techniques used in this paper at five possible levels: pixel level, feature level, score level, rank level, and decision level. The L2 normalization will be used if needed in these fusion approaches.

(1) Pixel level fusion using average summation (**PL-AS**): Pixel level image fusion simply fuses the raw data at pixel level. A simple geometry alignment based on centers of two eyes is pre-processed on the face images. We then average the pixel intensities at the same location of multiple stylistic sketches as the PL-AS result. Experiments show that this kind of fusion significantly improves the performance of hand-drawn sketches, which will be discussed later.

(2) Feature level fusion using feature concatenation (**FL-FC**): The feature descriptors are extracted on these different stylistic sketches. As these features are independent with each other, it is reasonable to simply concatenate them together to form a long vector. This new vector can then be used for recognition.

(3) Score level fusion using equal-weighted sum rule (**SL-SR**): Each sketch provides a matching score with the photo gallery using the face recognition algorithms introduced above. These scores can then be combined through an equal-weighted summation to exploit the complementary information among sketches.

    We further evaluate switching the equal-weighted sum rule with product rule, abbreviate to **SL-PR**, in this paper. Kernel based fusion strategy also shows effective performance. We therefore add two more score level fusion techniques utilizing two-class SVM and one-class SVM, abbreviating as **SL-TSVM** and **SL-OSVM** respectively.

(4) Rank level fusion using highest rank rule (**RL-HR**): In recognition systems, the ranked lists from multiple stylistic sketches can be fused at the rank level by selecting the highest rank among candidates. There is another rank level fusion technique, namely Borda count method [11], in which the sum of the ranks from multiple identification systems is taken as the final rank list. We abbreviate this strategy as **RL-BC** in this paper.

(5) Decision level fusion using majority voting (**DL-MV**): Each recognition system makes its own decision, and a majority vote strategy can then be applied to generate the final decision.

    We adopt these simple fusion techniques to evaluate the proposed three scenarios. Researchers are invited to submit their own algorithms and experimental results on these scenarios later (after the review process).

5   Experiments

5.1   Experimental Setup

All the face images used in this paper are firstly pre-processed by a simple geometry alignment based on the centers of two eyes and cropped to 200×250.

Fig. 4. Examples faces used in this paper: VIPSL photos with enlarged gallery (left) and AR photos with enlarged gallery (right).

The images are divided into patches of size 20×20 with 10 pixels overlapping. In the basic LBP-based face recognition baseline, multi-scale LBP feature is extracted on each patch by concatenating LBP feature descriptors with radius of 1, 3, 5, 7. Therefore, each face image yields a 107,616-D LBP-based feature for recognition. In the Fisherface baseline, 128-D scale invariant feature transform (SIFT) feature [19] is extracted on each image patch, thus leads to a 58,368-D SIFT feature for recognition. In the VGG-Face baseline, after removing the last three fully connected layers, each face image generates a 25,088-D deep feature.

In order to report unbiased performances, we define two views for each scenario. View 1 is used for parameter tuning and view 2 is used for algorithm evaluation. In view 2, protocols of 10 random partitions of dataset are provided and the average performances are reported. We follow the same partition ratio as [15]. For scenario-MHS, 133 subjects are used for training and 67 subjects are left for testing. For scenario-MHCS and scenario-MCS, 82 and 41 subjects are selected for training and testing respectively. The lists of image names are generated randomly, which will be available online.

In order to present results that can better mimic real-world criminal investigation scenarios, we construct an enlarged gallery set of 10,000 subjects. This enlarged gallery is composed of subjects from four sources: FERET [32] (2,437 subjects), XM2VTS [21] (1,180 subjects), MORPH (3,383 subjects), and LFW [12] (3,000 subjects). The face images in the first three datasets are captured under relatively controlled conditions similar to VIPSL and AR. The subjects from LFW are added to increase the diversity of the enlarged gallery. We have shown two groups of example faces in Fig. 4. The left three columns of faces in Fig. 4(a) and (b) are selected from VIPSL photos and AR photos, while the rest are from the enlarged gallery. It can be seen that the quality of these photos are similar, thus the enlarged gallery can affect the performances and help present results closer to real-world scenarios.

Given the experimental settings introduced, we present experimental results under identification scenario[8]. We first present the face recognition performances

---

[8] Due to the limitation of paper length, we only present the identification results in this paper. Verification performances will be released online later.

from single stylistic sketch in Table 3. The five styles of hand-drawn sketches in VIPSL are named style-A, style-B, style-C, style-D, and style-E respectively. For AR photos, the hand-drawn sketches from CUFS are named viewed, while the two styles of composite sketches are called PRIP and EPRIP in this paper. It can be seen that face recognition from the hand-drawn sketches in VIPSL is a relatively easy task, but the rank-1 accuracy can still be improved. Recognition from the composite sketches in AR is very difficult, with very low rank-1 accuracy. We expect this task to be improved by fusion of multiple stylistic sketches.

**Table 3.** Face recognition accuracies from single stylistic sketch on view 2 of each dataset. (accuracy±std, %)

| Dataset	Basic LBP Rank 1	Basic LBP Rank 50	Fisherface Rank 1	Fisherface Rank 50	VGG-Face Rank 1	VGG-Face Rank 50
VIPSL Style-A	49.10 ± 5.47	86.87 ± 3.78	68.36 ± 4.49	96.57 ± 1.87	60.60 ± 3.24	100 ± 0
VIPSL Style-B	45.07 ± 4.44	85.97 ± 2.13	66.27 ± 4.05	97.76 ± 1.06	58.21 ± 2.73	98.36 ± 1.31
VIPSL Style-C	65.67 ± 5.67	97.46 ± 1.42	72.39 ± 2.25	99.55 ± 0.72	59.10 ± 3.16	99.85 ± 0.47
VIPSL Style-D	55.82 ± 4.46	92.69 ± 3.26	71.94 ± 6.20	99.40 ± 0.77	64.18 ± 4.16	100 ± 0
VIPSL Style-E	55.07 ± 6.47	92.54 ± 2.73	75.52 ± 4.83	98.66 ± 1.31	62.24 ± 5.67	100 ± 0
AR Viewed	28.05 ± 6.11	67.56 ± 5.52	58.78 ± 6.13	88.05 ± 4.80	77.07 ± 6.52	96.83 ± 2.31
AR PRIP	3.41 ± 3.49	18.78 ± 5.75	1.46 ± 1.26	10.00 ± 4.22	0.49 ± 1.03	23.90 ± 4.71
AR EPRIP	4.63 ± 2.92	30.10 ± 7.16	1.46 ± 2.36	19.76 ± 7.49	5.12 ± 2.43	46.59 ± 7.75

5.2 Experimental Results

We first evaluate the baseline approaches in scenario-MHS. Figure 5 shows results of the baseline fusion approaches based on three recognition algorithms (Basic LBP-based face recognition, Fisherface, and VGG-Face). Comparing with the performance of recognition from single stylistic sketch in Table 3, fusion of information from multiple stylistic sketches can improve the accuracy. An interesting phenomenon is that pixel level fusion (PL-AS) yields excellent performance in this scenario. This may be because the sketches in scenario-MHS are all hand-drawn viewed sketches. The usage of pixel level fusion technique can be further investigated in the future.

Figure 6 demonstrates the results of baseline approaches in scenario-MHCS. Because this scenario contains both hand-drawn sketches from CUFS and composite sketches from PRIP-VSGC and EPRIP, it is harder than scenario-MHS. From Fig. 6 we can find that score level fusion techniques can achieve better results, while decision level fusion technique (DL-MV) performs poor. This is because there are two styles of composite sketches, whose recognition performances are poor in single sketch scenario as shown in Table 3. They usually takes the majority during voting, thus contributes to the poor performance of DL-MV.

We finally present the baseline performances in scenario-MCS, as shown in Fig. 7. This is the most difficult scenario among the three proposed in this paper.
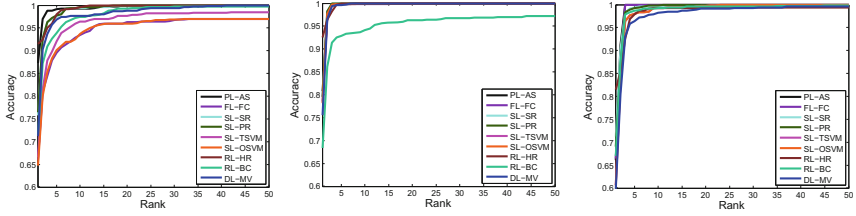
Fig. 5. Recognition performances of using basic LBP (left), Fisherface (middle), and VGG-Face (right) in scenario-MHS.
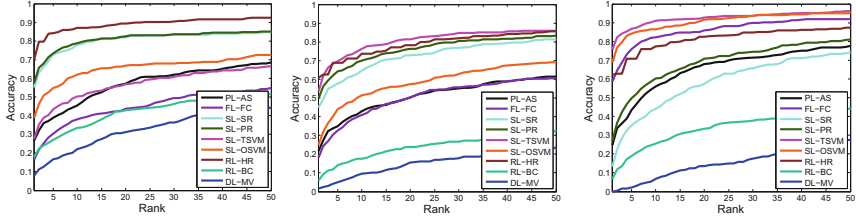
Fig. 6. Recognition performances of using basic LBP (left), Fisherface (middle), and VGG-Face (right) in scenario-MHCS.
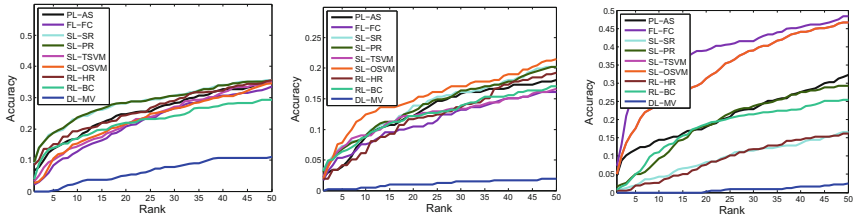
Fig. 7. Recognition performances of using basic LBP (left), Fisherface (middle), and VGG-Face (right) in scenario-MCS.

The Decision level fusion technique (DL-MV) failed to help improve this scenario. Score level fusion techniques achieve better results than other fusion methods, but their performances are still unsatisfying. It remains an unsolved problem of recognition from multiple stylistic sketches in scenario-MCS.

6   Conclusion and Future Directions

This paper presents a fundamental study of face recognition from multiple stylistic sketches. Three scenarios with different degrees of difficulty are proposed to mimic real-world situations. We also describe the corresponding dataset, evaluation protocols and several benchmarks to help illustrate these scenarios. More importantly, preliminary experimental results demonstrate the great challenges of this problem. There are also several interesting findings in the experiments.

For example, pixel level fusion of multiple hand-drawn sketches reveals exciting performance, which is rarely exploited before. However, none of the baseline fusion approaches introduced in this paper can achieve satisfying recognition performance from multiple composite sketches. This remains an open issue for further research.

In the future, we intend to add more baseline approaches into this problem, such as the kernel prototype similarities based method [15] and the graphical representation approach [30]. While we have tried the VGG-face model as the deep learning feature, we merely explored the simplest way. Future research of exploiting deep learning technique to deal with this problem is preferred.

References

1. FG-NET. http://www-prima.inrialpes.fr/FGnet/
2. Indian Face Database. http://vis-www.cs.umass.edu/~vidit/IndianFaceDatabase/
3. Belhumeur, P., Hespanda, J., Kiregeman, D.: Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE TPAMI **19**(7), 711–720 (1997)
4. Best-Rowden, L., Han, H., Otto, C., Klare, B., Jain, A.: Unconstrained face recognition: identifying a person of interest from a media collection. IEEE TIFS **9**(12), 2144–2157 (2014)
5. Bhatt, H., Bharadwaj, S., Singh, R., Vatsa, M.: Memetically optimized MCWLD for matching sketches with digital face images. IEEE TIFS **7**(5), 1522–1535 (2012)
6. Frowd, C.: Eyewitnesses and the use and application of cognitive theory. Introduction to applied psychology (2011)
7. Galoogahi, H., Sim, T.: Face sketch recognition by local radon binary pattern. In: ICIP, pp. 1837–1840 (2012)
8. Gao, X., Wang, N., Tao, D., Li, X.: Face sketch-photo synthesis and retrieval using sparse representation. IEEE TCSVT **22**, 1213–1226 (2012)
9. Gibson, L.: Forensic Art Essentials: A Manual for Law Enforcement Artists. Academic Press, Waltham (2010)
10. Han, H., Klare, B., Bonnen, K., Jain, A.: Matching composite sketches to face photos: a component-based approach. IEEE TIFS **8**(1), 191–204 (2013)
11. Ho, T., Hull, J., Srihari, S.: Decision combination in multiple classifier systems. IEEE TPAMI **16**(1), 66–75 (1994)
12. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical report. 07–49, University of Massachusetts, Amherst (2007)
13. Uhl Jr., R.G., da Victoria Lobo., N.: A framework for recognizing a facial image from a police sketch. In: CVPR, pp. 586–593 (1996)

14. Kan, M., Shan, S., Zhang, H., Lao, S., Chen, X.: Multi-view discriminant analysis. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012. LNCS, vol. 7572, pp. 808–821. Springer, Heidelberg (2012). doi:10.1007/978-3-642-33718-5_58

15. Klare, B., Jain, A.: Heterogeneous face recognition using kernel prototype similarities. IEEE TPAMI **35**(6), 1410–1422 (2013)

16. Klare, B., Li, Z., Jain, A.: Matching forensic sketches to mug shot photos. IEEE TPAMI **33**, 639–646 (2011)

17. Klum, S., Han, H., Klare, B., Jain, A.K.: The FaceSketchID system: matching facial composites to mugshots. IEEE TIFS **9**(12), 2248–2263 (2014)

18. Liu, Q., Tang, X., Jin, H., Lu, H., Ma, S.: A nonlinear approach for face sketch synthesis and recognition. In: CVPR, pp. 1005–1010 (2005)

19. Lowe, D.: Distinctive image features from scale-invariant key-points. IJCV **60**(2), 91–110 (2004)

20. Martinez, A., Benavente, R.: The AR face database. Technical report 24, CVC, Barcelona, Spain, June 1998

21. Messer, K., Matas, J., Kittler, J., Luettin, J., Maitre, G.: XM2VTSDB: the extended M2VTS database. In: AVBPA, Washington, DC, USA, pp. 72–77, April 1999

22. Mittal, P., Jain, A., Goswami, G., Vatsa, M., Singh, R.: Composite sketch recognition using saliency and attribute feedback. Inf. Fusion **33**, 86–99 (2017)

23. Mittal, P., Vatsa, M., Singh, R.: Composite sketch recognition via deep network-A transfer learning approach. In: ICB (2015)

24. Nejati, H., Zhang, L., Sim, T.: Eyewitness face sketch recognition based on two-step bias modeling. In: Wilson, R., Hancock, E., Bors, A., Smith, W. (eds.) CAIP 2013. LNCS, vol. 8048, pp. 26–33. Springer, Heidelberg (2013). doi:10.1007/978-3-642-40246-3_4

25. Ouyang, S., Hospedales, T., Song, Y., Li, X.: ForgetMeNot: memory-aware forensic facial sketch matching. In: CVPR

26. Ouyang, S., Hospedales, T., Song, Y.-Z., Li, X.: Cross-modal face matching: beyond viewed sketches. In: Cremers, D., Reid, I., Saito, H., Yang, M.-H. (eds.) ACCV 2014. LNCS, vol. 9004, pp. 210–225. Springer, Heidelberg (2015). doi:10.1007/978-3-319-16808-1_15

27. Ouyang, S., Hospedales, T., Song, Y.Z., Li, X.: A survey on heterogeneous face recognition: Sketch, infra-red, 3d and low-resolution (2014). http://arxiv.org/abs/1409.5114

28. Parkhi, O.M., Vedaldi, A., Zisserman, A.: Deep face recognition. In: BMVC (2015)

29. Peng, C., Gao, X., Wang, N., Li, J.: Superpixel-based face sketch-photo synthesis. IEEE TCSVT (2015)

30. Peng, C., Gao, X., Wang, N., Li, J.: Graphical representation for heterogeneous face recognition. IEEE TPAMI (2016)

31. Peng, C., Gao, X., Wang, N., Tao, D., Li, X., Li, J.: Multiple representations-based face sketch-photo synthesis. IEEE TNNLS (2015)

32. Phillips, P., Moon, H., Rizvi, S., Rauss, P.: The FERET evaluation methodology for face recognition algorithms. IEEE TPAMI **22**(10), 1090–1104 (2000)

33. Serrano, A., de Diego, I., Conde, C., Cabello, E., Shen, L., Bai, L.: Influence of wavelet frequency and orientation in an SVM based parallel Gabor PCA face verification system. In: IDEAL, pp. 219–228 (2007)

34. Sharma, A., Jacobs, D.: Bypass synthesis: PLS for face recognition with pose, low-resolution and sketch. In: CVPR, pp. 593–600 (2011)

35. Ahonen, T., Hadid, A., Pietikäinen, M.: Face description with local binarypatterns: application to face recognition. IEEE TPAMI **28**(12), 2037–2041 (2006)
36. Tang, X., Wang, X.: Face photo recognition using sketch. In: ICIP, pp. 257–260, September 2002
37. Taylor, K.: Forensic Art and Illustration. CRC Press, Boca Raton (2001)
38. Wang, N., Tao, D., Gao, X., Li, X., Li, J.: Transductive face sketch-photo synthesis. IEEE TNNLS **24**(9), 1364–1376 (2013)
39. Wang, N., Tao, D., Gao, X., Li, X., Li, J.: A comprehensive survey to face hallucination. IJCV **31**(1), 9–30 (2014)
40. Wang, X., Tang, X.: Face photo-sketch synthesis and recognition. IEEE TPAMI **31**(11), 1955–1967 (2009)
41. Zhang, S., Gao, X., Wang, N., Li, J., Zhang, M.: Face sketch synthesis via sparse representation-based greedy search. IEEE TIP **24**(8), 2466–2477 (2015)
42. Zhang, W., Wang, X., Tang, X.: Coupled information-theoretic encoding for face photo-sketch recognition. In: CVPR, pp. 513–520 (2011)
43. Zhang, Y., McCullough, C., Sullins, J., Ross, C.: Hand-drawn face sketch recognition by humans and a pca-based algorithm for forensic applications. IEEE TSMC-Part A **40**(3), 475–485 (2010)
44. Zhou, H., Kuang, Z., Wong, K.: Markov weight fields for face sketch synthesis. In: CVPR, pp. 1091–1097 (2012)