Pattern Recognition 109 (2021) 107579

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

Iterative local re-ranking with attribute guided synthesis for face sketch recognition

Decheng Liu[a], Xinbo Gao[a,d,*], Nannan Wang[b,*], Chunlei Peng[c], Jie Li[a]

[a] State Key Laboratory of Integrated Services Networks, School of Electronic Engineering, Xidian University, Xi'an 710071, Shaanxi, P. R. China
[b] State Key Laboratory of Integrated Services Networks, School of Telecommunications Engineering, Xidian University, Xi'an 710071, Shaanxi, P. R. China
[c] State Key Laboratory of Integrated Services Networks, School of Cyber Engineering, Xidian University, Xi'an 710071, Shaanxi, P. R. China
[d] Chongqing Key Laboratory of Image Cognition, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

ARTICLE INFO

Article history:
Received 2 August 2019
Revised 25 March 2020
Accepted 4 August 2020
Available online 5 August 2020

Keywords:
Face sketch recognition
Facial attribute
Re-ranking

ABSTRACT

Because of the large texture and spatial structure discrepancies between face sketches and photos, face sketch recognition becomes a challenging problem in face recognition community. For example, in law enforcement and security, the specific face sketch generation process could introduce some inevitable biases which results in poor face sketch recognition performance. In order to mimic the modality gap introduced by the biases during face sketch creation process, the novel iterative local re-ranking with attribute guided synthesis method is proposed for face sketch recognition, which does not require any extra manually annotation or human interaction. The clues of face attributes are utilized to generate images with varying local characteristic from probe sketches, which could help eliminate the unavoidable biases. Considering the special property of face sketches, the iterative local re-ranking algorithm is designed to encode the contextual information integrated with local invariant discriminative information for matching sketches with photos. Experimental results on multiple face sketch databases demonstrate that the proposed method achieves superior performances compared with state-of-the-art methods.

1. Introduction

Face recognition is a challenging and important application in computer vision. Recently great progress has been achieved, however there still exist many challenging scenarios in the real world. Especially in the law enforcement agency, there exist many scenes where the mug shot of the suspect is not available or only poor-quality images are captured in video surveillance. Due to the lack of suspects' photographs, law enforcement agencies have started to generate face sketches according to the description provided by eyewitness or blurry surveillance videos. With technological improvement, more forensic artists have begun to utilize the generation software to produce composite sketches as the placement of hand-drawn sketches. However, it is due to the complex and special generation process of face sketches, there always exist shape exaggerations and distortions in face sketches. More importantly, in law enforcement face sketches are utilized to determine the identity of criminals where only the description of eyewitnesses is available. Forensic psychology related works [1] prove that face sketch recognition is affected by forgotten memory of eyewitness and imperfect communication. In summary, the sketch-photo modality difference, the inaccurate memory of eyewitnesses and the biased communication of memory all bring about unavoidable biases in the generated face sketches. Thus, face sketch recognition remains a difficult and challenging task in the real-world scenario.

Existing face sketch recognition methods mostly focus on three aspects: 1) extracting modality invariant features [2–4], which contain the identity discriminative information; 2) projecting different modality face images into a latent common space [5–8], where face sketches and gallery photos could be matched directly; 3) transforming images in one modality to another modality [9–12], which would make these images in homogeneous scenarios, and then the traditional homogeneous face recognition methods could be directly utilized. However, face sketches in real-world scenes differ from gallery photos because of the avoidable perceptual bias, descriptive bias and generating bias [13]. These inevitable biases caused in the generation procedures could enlarge the gap of different modalities. Only transforming sketches and photos in homogeneous scenarios is not enough to eliminate these biases when matching face sketches. To mimic the modality gap mentioned be-

* Corresponding authors.
E-mail addresses: dcliu.xidian@gmail.com (D. Liu), gaoxb@cqupt.edu.cn (X. Gao), nnwang@xidian.edu.cn (N. Wang), clpeng@xidian.edu.cn (C. Peng), leejie@mail.xidian.edu.cn (J. Li).

https://doi.org/10.1016/j.patcog.2020.107579
0031-3203/

fore, a new framework called iterative local re-ranking method is proposed.

In real-world scenario, given a face sketch of the suspect, the police want to search the according photo belongs to the same person in a large mug shot gallery. When considering face sketch recognition as a retrieval process, re-ranking is an import and critical procedure to effectively improve performance. Generally, the initial ranking list is calculated with the pairwise similarities between the query sketch and the gallery photos, and then the re-ranking procedure is used to refine the initial ranking list by taking account of the neighborhood relations among all images [14]. Yet in face sketch recognition community, limited researches have been devoted to re-ranking. It is because that there generally only exist sketch-photo pairs, with lack of forensic artists and the difficulty of sketches generation. It inspires us to generate more neighbors of face sketches as the base of re-ranking method.

This paper proposes a novel iterative local re-ranking with attribute guided synthesis method for face sketch recognition, which could effectively eliminate inevitable biases in the generation process, and simultaneously not require any human interaction or manually annotated data. The proposed method consists three steps. Firstly, the attribute guided face sketch synthesis method generates the synthesized face photos with clues of attributes. These synthesized photos with varying local characteristic could help eliminate the unavoidable biases introduced by generation process. Then, cross modality invariant features of both synthesized photos and gallery photos are extracted to obtain initial distance matrix. Finally, an iterative local re-ranking algorithm is proposed to integrate the face local discriminative information to further boost the recognition performance. The main reasons of superior performance are that synthesized face photos with varying local characteristic could eliminate the unavoidable biases in the sketches generation process, and the local iterative re-ranking method would integrate the local discriminative information of face components. The framework of the proposed algorithm is illustrated in Fig. 1.

The main contributions of this paper are summarized as follows:

1. The attribute guided face sketch synthesis method is utilized to generate images with varying local characteristic, which could eliminate the face perception biases in the face sketch generation process.
2. To the best of our knowledge, it is the first exploration to introduce re-ranking technique for face sketch recognition. The proposed iterative local re-ranking method is designed to integrate the inherent local invariant information of face sketches. More importantly, the proposed post-process procedure does not require any human interaction and could be applied in an unsupervised manner.
3. The proposed re-ranking method is proved to be effective on multiple face sketch databases, and achieves superior performance compared with the state-of-the-art methods.

The rest of this paper is organized as follows. Section 2 gives a review of face sketch recognition methods and related works. Section 3 presents a novel iterative local re-ranking with attribute guided synthesis method for face sketch recognition. Section 4 shows the experimental results and parameters analysis, and the conclusion is drawn in Section 5.

2. Related work

In this section, the representative face sketch recognition methods are reviewed in mentioned three categories: feature descriptor-based methods, common space-based methods and synthesis-based methods.

Feature descriptor-based methods: [15] firstly utilized a difference of Gaussian filter for matching heterogeneous images. [2] explored the multiple hand-crafted features and proposed a local feature based discriminant analysis. Prototype random subspace (P-RS) [16] was proposed to utilize nonlinear kernel similarities for matching. [3] presented a transfer learning based representation method for face sketch recognition. [17] presented a common encoding model to capture common discriminant information. [4] proposed an unsupervised feature learning method which learns features from raw pixels. [18] utilized the character of face sketch and fused complementary discriminant information for face sketch recognition. However, this kind of methods would be utilized with high computational complexity; **Common space-based methods:** [5] firstly proposed a common discriminant feature extraction (CDFE) approach. [19] presented a coupled spectral regression based method for matching. A multi-view discriminant analysis (MvDA) method [6] was proposed to exploit both inter-view and intra-view correlations of heterogeneous face images. [7] proposed a margin based cross-modality metric learning to address the gap of different modalities. [20] explored the deep attribute guided representation for face sketch recognition, which effectively integrates face attribute discriminant information. Yet the projection procedure may losses some discriminative

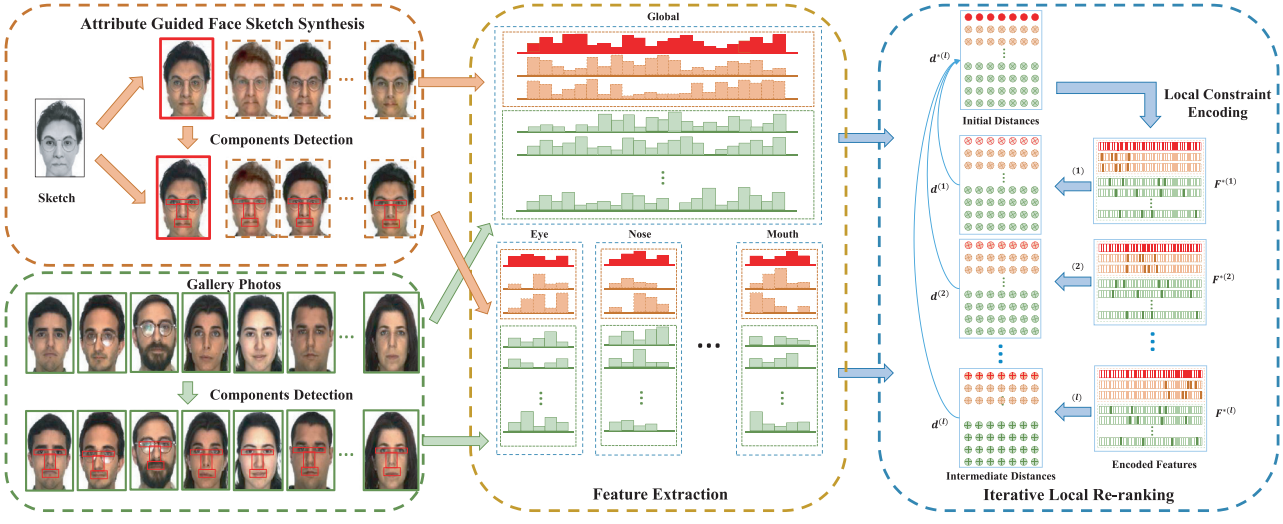


Fig. 1. Framework of the proposed iterative local re-ranking with attribute guided synthesis for face sketch recognition.

information. **Synthesis-based methods:** [9] presented an Eigen-transform algorithm for sketch-photo synthesis. [10] employed local linear embedding (LLE) to synthesize face sketches. Considering the relationship between face image patches and its neighboring patches, [11] exploited the Markov random field (MRF) for synthesis. [21] proposed a Markov weight field model to select candidates to construct the MRF Instead of directly matching synthesized sketches. [22] employed embedded hidden Markov model to represent the non-linear relationship. [23] presented a transductive face sketch-photo synthesis method. [24] proposed a multiple representations based approach to enhance the generalizability in forensic science. Instead of directly matching synthesized sketches, [12] proposed a novel graphical representation where Markov network are employed to represent image patches separately. [25] incorporated the synthesized images with asymmetric joint learning for face sketch recognition. Recently, generative adversarial networks (GAN) attracts growing attentions in image generation task. A conditional GAN [26] is proposed to perform image to image task[27]. exploited an unpaired image to image translation framework[28]. utilized multi-adversarial networks to generate high resolution face images[29]. exploited a composition-aided generative adversarial network for sketch-photo synthesis. These methods yield fine sketch-realistic textures generated by GAN model, but mostly introduce noise among the generated results. Apparently the quality of synthesized sketches could influence the matching performance and the image synthesis process is a complex problem itself. Most synthesis methods focus on the modality difference between sketches and photos, ignoring the unavoidable shape exaggerations and distortions introduced from the generation process. Besides, many researches utilized the inherent properties of face sketches to improve sketch matching performance. The memory-aware framework [13] is proposed for forensic sketch recognition, which takes the domain gap between forensic sketches and photos into consideration. Motivated by the sketch generation process, [18,30] proposed the component based representation algorithm for matching face sketches.

The re-ranking methods have been applied widely in object retrieval application. And these methods [31] mostly utilize the neighborhood relationship to refine the initial ranking list to boost performance. In particular, [32] proposed the sparse contextual activation algorithm (SCA) which encodes the local distribution of images and measures the dissimilarity in Jaccard distance. [33] introduced a k-reciprocal feature (k-RF) by encoding the k-reciprocal nearest neighbors. [34] introduced the regularized ensemble diffusion and learned the weights of similarities automatically for retrieval, which could effectively suppress the negative impacts of noisy similarities[35]. proposed the supervised smooth manifold as a generic tool, which can easily boost performances of person re-identification algorithms[36]. summarized fusion methods and proposed the unified ensemble diffusion algorithm, which has a faster diffusion step and is more robust to noise. However, due to the pairwise samples in sketch database lacking of contextual neighbors, limited effort has been devoted to re-ranking in face sketch recognition community. *Considering these inevitable biases may result in shape exaggerations and distortions, the synthesized face photos with varying local characteristic would help eliminate these biases, which is the basic motivation of our proposed method.*

3. Proposed algorithm

3.1. Problem definition

In the real-world scenario, face sketches are always generated from forensic artists according to the description of eyewitnesses, and the probe sketch is utilized to match mugshot photos in a large gallery. Consider the probe sketch s and the gallery photos set P with N images $P = \{p_i | i = 1, 2, \ldots, N\}$. Let $f(s)$ and $f(p_i)$ denotes the extracted cross modality invariant features of the probe sketch s and the gallery photos p_i respectively. More details about feature extraction are described in Section 4.1.3. Following the initial distance $d(s, p_i)$ between the probe sketch s and the gallery photo p_i can be measured by Euclidean distance. The smaller distance value is, the smaller dissimilarity of these images is.

Here for each probe sketch s, the initial ranking list $L(s, P) = \{p_1, p_2, \ldots, p_N\}$ could be obtained by the $d(s, p_i)$, where $d(s, p_i) < d(s, p_{i+1})$. Obviously our goal is to refine the initial ranking list $L(s, P)$ to make the gallery photo with the same identity rank higher in the list, which results in improving the face sketch recognition performance.

3.2. Attribute guided face sketch synthesis

Different from other modalities face images (*e.g.* near infrared images and thermal infrared images), face sketches are always utilized in law enforcement when no visual images are available. Forensic artists would generate face sketches by hands or with specific generation softwares just according to the description of eyewitnesses. The forgotten memory of eyewitnesses and inaccurate description of memory could introduce inevitable biases in the face sketches. It could be found that most forensic artists always exaggerate the local characteristic of sketches to make them easier to recognize. It inspires us to generate images with varying local characteristic to eliminate these unavoidable biases mentioned.

In this section, the details of attribute guided face sketch synthesis method are proposed. It consists two stages. In the first stage, the unpaired sketch-photo synthesis architecture is adopted from [27]to mimic the large sketch-photo modality gap. It is because that there always lacks of enough paired sketch-photo images in real-world scenario. Given the probe sketch s, the sketch-photo generator mapping $G^0_{p^0}(s) : s \rightarrow p^0$, and the photo-sketch generator mapping $G^0_s(p^0) : p^0 \rightarrow s$. $D^0_{p^0}$ would be trained to distinguish between "real" photo p^0 and the generated "fake" photo $G^0_{p^0}(s)$, and D^0_s distinguish between "real" sketch s and the generated "fake" sketch $G^0_s(p^0)$. For convenience, the adversarial loss is defined as,

$$L_{GAN}(G, D) = \mathrm{E}_{y \sim p_{data(y)}}[logD(y)] + \mathrm{E}_{x \sim p_{data(x)}}[log(1 - D(G(x)))]. \quad (1)$$

The full object function is:

$$L(G^0_{p^0}, G^0_s, D^0_{p^0}, D^0_s) = L_{GAN}(G^0_{p^0}, D^0_{p^0}) + L_{GAN}(G^0_s, D^0_s) + \lambda_c L_{cyc}(G^0_{p^0}, G^0_s), \quad (2)$$

where

$$L_{cyc}(G^0_{p^0}, G^0_s) = \mathrm{E}_{p^0 \sim p_{data(p^0)}}[\left\| p^0 - G^0_{p^0}(G^0_s(p^0)) \right\|_1] + \mathrm{E}_{s \sim s_{data(s)}}[\left\| s - G^0_s(G^0_{p^0}(s)) \right\|_1].$$

Here λ_c controls the importance of objectives and the whole model is trained as:

$$G^{0*}_{p^0}, G^{0*}_s = \arg \min_{G^0_{p^0}, G^0_s} \max_{D^0_{p^0}, D^0_s} L(G^0_{p^0}, G^0_s, D^0_{p^0}, D^0_s). \quad (3)$$

In the second stage, in order to eliminate the biases introduced by the forgotten memory and inaccurate communication, the attribute guided sketch synthesis framework is proposed from [37]. This architecture model could effectively learn the mapping among different attribute domains with a single generator. However, only single attribute of synthesized photos is revised to avoid change of face sketch's identity. Here the generator G^a is trained to generate the synthesized photo p^a with different attribute labels a, $G^a(p^0, a)$: $p^0 \rightarrow p^a$. CelebA database [38] is utilized to train the

generator G^a. Here five local discriminative attributes are chosen, where $a \in$ {Black Hair, Blond Hair, Brown Hair, Big Nose, Chubby}. $D^a : p^0 \rightarrow \{D_{src}(p^0), D_{att}(p^0, a)\}$, $D_{src}(p^0)$ produces probability distributions over sources, and $D_{att}(p^0, a)$ predicts the presence of the desired attribute. Finally, the objective function for discriminator D^a is defined as

$$L_{D^a} = -L_{GAN}(G^a, D_{src}) + \lambda_{att}\mathrm{E}_{p^0,a'}[-logD_{att}(a'|p^0)], \tag{4}$$

and objective function for generator G^a is defined as

$$\begin{aligned} L_{G^a} &= L_{GAN}(G^a, D_{src}) \\ &\quad +\lambda_{att}\mathrm{E}_{p^0,a}[-logD_{att}(a|G^a(p^0, a))] \\ &\quad +\lambda_{con}\mathrm{E}_{p^0,a,a'}[\left\| p^0 - G^a(G^a(p^0, a), a')\right\|_1], \end{aligned} \tag{5}$$

where a' means original attribute label of sketch s ($\lambda_{att} = 1, \lambda_{con} =$ 10 by default). λ_{att} and λ_{con} control the importance of attribute classification and reconstruction losses respectively.

Finally, when the probe sketch s is acquired, the trained generator $G^0_{p^0}(s)$ and $G^a(p^0, a)$ could synthesize photos p^a with varying local characteristic for matching to boost performance.

3.3. Iterative local re-ranking

3.3.1. Local constraint nearest neighbors

Motivated by [32], the neighborhood set of the probe sketch s is defined as $N(s, k)$, which contains the top-k candidates in $L(s, P)$:

$$N(s, k) = \{p_1, p_2, \ldots, p_k\}. \tag{6}$$

Considering the property of face sketch generation process, the proposed method aims to integrate the local invariant discriminative information from face components to address the modality gap between face sketches and photos. Following [18], the holistic face image is divided into five components: hair, brows, eyes, nose and mouth. And then the pre-trained network [39] provides a common representation for face sketches and face photos which could diminish intraperson variations. The face components neighborhood set of the probe sketches is denoted as $N^c(s, k)$, $c \in$ {*Hair, Brows, Eye, Nose, Mouth*}. For ranking the gallery photo p with similar components features higher, the local constrain nearest neighbors are defined as $C^c(s, k)$,

$$\begin{aligned} C^c_h(s, k) &= \{p_i | (p_i \in N(s, k) \wedge (p_i \in N^c(s, k)))\} \\ &= N(s, k) \cap N^c(s, k). \end{aligned} \tag{7}$$

Compared with the nearest neighborhood set $N(s, k)$, the intersection local constrain nearest neighbors $C^c_h(s, k)$ are more likely to be positive samples. However, when the gallery photo p_i ranks high in the $C^c_h(s, k)$, p_i should rank high in both ranking list $N(s, k)$ and $N^c(s, k)$. Thus, some positive photos may be excluded from the intersection local constrain nearest neighbors $C^c_h(s, k)$. Naturally, the union local constrain nearest neighbors $C^c_l(s, k)$ is defined with looser constrain,

$$\begin{aligned} C^c_l(s, k) &= \{p_i | (p_i \in N(s, k) \vee (p_i \in N^c(s, k)))\} \\ &= N(s, k) \cup N^c(s, k). \end{aligned} \tag{8}$$

Following some negative gallery photos may be included in the union local constrain nearest neighbors $C^c_l(s, k)$, which would disturb the matching performance. In the proposed algorithm, these two neighbors would be integrated simultaneously, and more contrasting positive samples are introduced in the proposed nearest neighbor set compared with the original nearest neighbor set.

3.3.2. Local constraint encoding

Intuitively, the assumption is shown that when the probe sketch and photo are similar, the neighbors of them are also similar. In other words, the more similar the sketch and photo are, the more overlap neighbors they contain. Here the Jaccard distance is adopted to represent the dissimilarity of probe sketches and gallery photos:

$$d_J(s, p) = 1 - \frac{|C^c(s, k) \cap C^c(p, k)|}{|C^c(s, k) \cup C^c(p, k)|}. \tag{9}$$

Where $|\cdot|$ means the number of neighbors in the set. $C^c(s, k)$ means the local constraints nearest neighbor set mentioned before. Although Jaccard distance could be utilized to represent the dissimilarity between sketches and photos with the relationship of their neighbor sets, there exists some shortcomings [33]. Inspired from [32], the local constraint nearest neighbor set of the probe sketch s is encoded into a N dim feature $F_s = [F_{s,p_1}, F_{s,p_2}, \ldots, F_{s,p_N}]$. Additionally, in order to make full advantage of contextual information, the neighbors are assigned with different weights when they occupy different positions. Here the similar neighbors are set with larger weights. However, only taking the contextual information of probe sketches into account is not enough to acquire discriminative information, for the large gap between sketches and photos. Hence, the inherent properties of face sketches are derived to encode more discriminative features. Motivated by [30], the local constraint encoding is designed to integrate more local invariant discriminative information, which is consistent with the process of face sketch generation. For convenience, the Gaussian kernel is applied to measure pairwise dissimilarity, which is proved to be effective in visual retrieval task [32,33].

$$F_{s,p_i} = \begin{cases} \exp(-d(s, p_i)), & p_i \in C^c(s, k) \\ 0, & \text{otherwise}. \end{cases} \tag{10}$$

Here $d(s, p_i)$ represents the dissimilarity of probe sketch s and gallery photo p_i. $C^c(s, k) \in \{C^c_h(s, k), C^c_l(s, k)\}$. $C^c(s, k)$ means the intersection local constrain nearest neighbors $C^c_h(s, k)$ or the union local constrain nearest neighbors $C^c_l(s, k)$ in Eq. (7), (8). k is denoted by k_1 to be distinguished from the latter k in Eq. (12). With the proposed local constraint encoding, more divergent neighbors according to local invariant information could be utilized to make the encoded features more discriminative. In this way, the encoded features could benefit from the diversity of synthesized photos with varying local characteristic, which means more divergent attributes guided discriminative information could help eliminate the unavoidable biases of face sketches. For convenience, the interaction and union set are calculated as

$$\begin{aligned} |C^c(s, k) \cap C^c(p_i, k)| &= \left\| \min(F^*_s, F^*_{p_i}) \right\|_1 \\ |C^c(s, k) \cup C^c(p_i, k)| &= \left\| \max(F^*_s, F^*_{p_i}) \right\|_1. \end{aligned} \tag{11}$$

Besides, there exists a rule that the encoded features of the probe sketches with the same identity should be similar. In real-world scenarios, there indeed exist some inevitable biases in face sketch generation process, which results in shape exaggerations and distortions of them. Thus the attribute guided face sketch synthesis method could help increase the interperson diversity and diminish intraperson variations. Emulating the idea [32] that images form the same category may share similar encoder features, it is assumed that synthesis faces with varying local characteristic would maintain the same identity, and naturally the encoded features of them should be similar. Furthermore, the designed new feature is encoded to integrate more divergent attributes information:

$$F^*_{s,p_i} = \frac{1}{|C^c(s,k)|} \sum_{i \in C^c(s,k)} F_i, \tag{12}$$

where F_i means the local constraint encoded feature according to Eq. (10). F^*_s means the new encoded feature of probe sketch s, which integrate more divergent information. k is denoted by k_2 to be distinguished from the k in Eq. (10). Finally the new dissimilar-

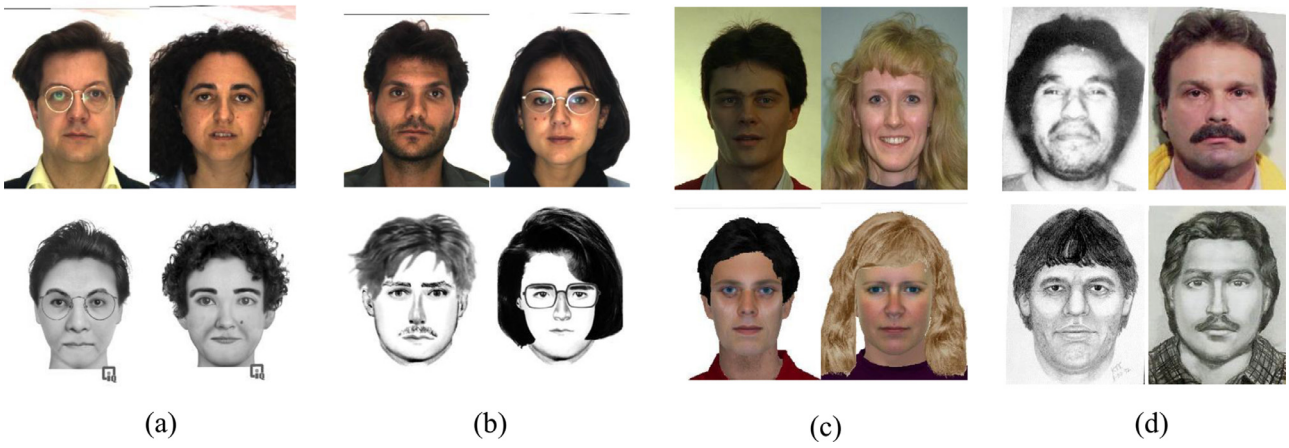


Fig. 2. The illustration of face sketch databases. (a): E-PRIP Sketch Database. (b): PRIP-VSGC Sketch Database. (c): UoM-SGFS Sketch Database. (d): Forensic Sketch Database.

ity metric could be designed as

$$d(s,p_i) = 1 - \frac{1}{2}(w \cdot \frac{\sum_{j=1}^{N} \min(F^{*h}_{s,p_j}, F^{*h}_{p_i,p_j})}{\sum_{j=1}^{N} \max(F^{*h}_{s,p_j}, F^{*h}_{p_i,p_j})} + (1-w) \cdot (\frac{\sum_{j=1}^{N} \min(F^{*l}_{s,p_j}, F^{*l}_{p_i,p_j})}{\sum_{j=1}^{N} \max(F^{*l}_{s,p_j}, F^{*l}_{p_i,p_j})})), \quad (13)$$

here $w \in [0, 1]$ balances the importance of two different types of local constraint encoding features.

3.3.3. Iterative local aggregation

Actually, when exploring the face sketch generation process, it could be found that different local regions provide different complement discriminative information. It is because that forensic artists always are concerned with some obvious local information to make generated face sketches more discriminative. Derived from this, the contextual information is further encoded in an iterative manner to improve performance. Here the intermediate distance is calculated in the same way as Eq. (13),

$$d^{(l)}(s,p_i) = 1 - \frac{1}{2}(w \cdot \frac{\sum_{j=1}^{N} \min(F^{*h^{(l)}}_{s,p_j}, F^{*h^{(l)}}_{p_i,p_j})}{\sum_{j=1}^{N} \max(F^{*h^{(l)}}_{s,p_j}, F^{*h^{(l)}}_{p_i,p_j})} + (1-w) \cdot (\frac{\sum_{j=1}^{N} \min(F^{*l^{(l)}}_{s,p_j}, F^{*l^{(l)}}_{p_i,p_j})}{\sum_{j=1}^{N} \max(F^{*l^{(l)}}_{s,p_j}, F^{*l^{(l)}}_{p_i,p_j})})). \quad (14)$$

Inspired from [14] in which the iterative scheme is firstly introduced to fuse information, our goal is to iteratively refine the initial distance $d^{*(l)}(s, p_i)$ by integrating extra complementary components discriminative information. The iterative local aggregation function is formulated as follows:

$$d^{*^{(l)}}(s,p_i) := (1-\lambda) d^{*^{(l)}}(s,p_i) + \lambda d^{(l)}(s,p_i), \quad (15)$$

where $\lambda \in [0, 1]$ denotes the iterative local aggregation factor. It is noted that the intermedia distance is calculated by the extra local region pairwise distance. Naturally, five-iteration feature encoding is conducted to fuse complementary information for recognition, due to local discriminative information extracted from five local regions (Hair, Brows, Eye, Nose and Mouth).

4. Experiments

In this section, the recognition performance of the proposed method is evlauated on multiple face sketch databases. These face sketch databases are similar to real scenarios: the Extend PRIP database (E-PRIP) [3,40], PRIP Viewed Software-Generated Composite database (PRIP-VSGC) [41], Extended UoM-SGFS database [42] and Forensic Sketch Databases. Example face images are shown in Fig. 2. Meanwhile, there always exists limited number of face sketches because of the complex generation procedure. The face attributes information of these face sketches could be directly acquired without extra labor in the generation process.

The details about face sketch databases and evaluation metric are shown firstly. Then the proposed method is compared on multi face sketch databases to confirm that our method could achieve superior performance. Finally, the effect of different parameters is investigated on the recognition performance, and the ablation study of the proposed method is further exploited. For most experiment databases, the database is randomly split into the training set and the testing set. The accuracies shown in this section are statistical results over 10 random partitions, expected when noted.

4.1. Databases and setting

Four face sketch databases are shown in this section. Example face images are shown in Fig. 2. For all experiment databases, the whole dataset is randomly split into the training set and the testing set. And recognition accuracies shown in the following sections are statistical results over several random partitions.

4.1.1. Databases

Extended PRIP Database (E-PRIP) contains 123 subjects, with photos from the AR database [43] and composite sketches [44] are generated by FACE software.

PRIP Viewed Software-Generated Composite Database (PRIP-VSGC) also contains 123 subjects, with photos from the AR database [43] and composite sketches are generated by Identi-Kit software [45].

Extended UoM-SGFS database [42] contains 600 subjects, with photos from the Color FERET database and 1200 facial sketches in the extended UoM-SGFS database. Two sets are present in this database: Set A containing sketches created using EFIT-V, and Set B containing sketches in Set A which are lightly altered using an image editing program to make the sketches more realistic.

Forensic Database contains 168 subjects, with mug shot photos and corresponding sketches from real world. The forensic sketches are drawn by sketch artists with the description of eyewitnesses or victims. The eyewitnesses's forgotten memory, inaccurate description of memory and even the artists's perceptual experience when drawing details of sketches could lead to differences between forensic sketches and photos.

To make results much closer to the real scenarios, the proposed method is evaluated on the enlarged gallery [12]. The enlarged

gallery contains 10,000 face photo images of 5329 subjects which mimic the real-world face retrieval scenarios.

4.1.2. Evaluation metric

In real-world scenario, it is not necessary to find the rank-1 photo as the suspect mug shot for the large modality difference between sketches and photos. Instead, the rank-n photos would also be found as the candidates of suspect. Thus, the cumulative match characteristic would be utilized as the evaluation metric in following face sketch recognition experiments.

4.1.3. Feature representations

Because of the successfully application of deep convolutional network methods in traditional face recognition, the proposed method aims to extract cross modality invariant features with convolutional network method. Due to small number of face sketch database, experiment [46] shows the fine-tuning in networks would result in overfitting and inferior results. Hence, the 29 layers LightCNN network [47] is pre-trained with the CASIA-WebFace and MS-Celeb-1M database sequentially, and then extract both face sketches and photos features at the second pool layer. Besides, the NLDA algorithm [48] is applied to make these cross modality features more discriminative. The mentioned method is utilized to extract features of synthesis photos and gallery photos and then utilize the Euclidean metric to calculate dissimilarities, which is regarded as the baseline of our face sketch recognition framework.

4.2. Experiments on face sketch databases

4.2.1. Experiments on E-PRIP sketch database

The same protocol [3] is followed and the dataset is split into two parts randomly with 10 partitions: the 48 sketch-photo pairs for training and the rest 75 sketch-photo pairs for testing. As for the parameters in our algorithm, k_1 is set to 26, k_2 is set to 3, w is set to 0.2 and λ is set to 0.2.

Comparison with re-ranking methods Compared with two popular re-ranking methods, sparse contextual activation (SCA) and k-reciprocal encoding (k-RF), the proposed method outperforms them in rank-10 accuracy, and gains a increase of 3.74% compared with baseline method. It is because the unavoidable biases are taken into consideration and extract local discriminative information.

Comparison with state-of-the-art methods The proposed approach is proposed with the state-of-art methods in Table 1 with protocol in [3]. It is noted that the Fisherface [49] is widely used for VIS homogeneous face recognition, however it yields poor performance in face sketch recognition. It is because the great modality gap between face sketch and conventional VIS images make it more challenging. Thus the face sketch recognition method is designed for the specialised face recogntion scenarios. [42] applied transfer learning to make features learned specifically for this task. However, these methods ignore the local inherent discriminative information of face sketch. It can be seen that the proposed method outperforms existing methods and reached rank-10 accuracy of 83.47% on the E-PRIP database.

Table 1
Rank-10 recognition accuracies of the state-of-the-art approaches and our method on the E-PRIP and PRIP-VSGC databases.

Algorithms	Accuracy(E-PRIP)	Accuracy(PRIP-VSGC)
Fisherface [49]	35.30%	21.87%
MCWLD [50]	24.00%	15.40%
SSD-based [40]	53.30%	45.30%
Transfer Learning [3]	60.20%	52.00%
CNNs [46]	65.60%	51.50%
DEEPS [42]	80.80%	54.90%
DCCNN [51]	68.60%	67.40%
SP-Net [52]	80.00%	-
SGR-DA [53]	-	70.00%
DLFace [54]	82.80%	**76.40%**
Baseline	79.73%	64.53%
SCA [32]	78.13%	66.40%
k-RF [33]	80.93%	68.27%
Proposed	**83.47%**	68.67%

4.2.2. Experiments on PRIP-VSGC sketch database

With the same protocol [3], the 48 sketch-photo pairs are selected randomly as the training set and the rest pairs are the testing set. k_1 is set to 36, k_2 is setto 3, w is set to 0.2 and λ is set to 0.2 in this database. The performance of our method with comparison is reported in Table 1. The proposed method outperforms existing methods and reached rank-10 accuracy of 68.67% on the PRIP-VSGC database.

4.2.3. Experiments on UoM-SGFS sketch database

The same protocol is followed with [42] to evaluate perform on this database. The 450 subjects are selected at random for training and the remaining 150 subjects are assigned as the test set. Besides, 1521 photos are randomly selected from the enlarged gallery mentioned [12] to simulate the mugshot galleries in law enforcement agencies. k_1 is set to 3, k_2 is set to 2, w is set to 0.2 and λ is set to 0.2 in this dataset.

The proposed iterative local re-ranking with attribute guided synthesis method is compared with state-of-the-art methods on the UoM-SGFS database Set A and Set B as shown in Table 2, 3 respectively. The proposed method achieves 84.53% at rank 1, which increase 52.93% compared with state-of-the-art method [42] on the UoM-SGFS Set A database. Due to the lightly alteration from Set A sketches to Set B sketches, the proposed method could achieve 89.73% at rank 1 on the UoM-SGFS Set B database.

Table 2
Recognition accuracies of the state-of-the-art methods and the proposed iterative local re-ranking method on the UoM-SGFS SetA Database.

Algorithms	Rank-1 Accuracy	Rank-10 Accuracy
PCA [55]	2.80%	8.40%
CBR [30]	5.73%	18.80%
VGGFace [56]	9.33%	31.07%
P-RS [16]	22.13%	49.33%
DEEPS [42]	31.60%	66.13%
DLFace [54]	64.80%	92.13%
Baseline	80.27%	97.07%
SCA [32]	82.67%	94.27%
k-RF [33]	83.73%	97.20%
Proposed	**84.53%**	**98.13%**

Table 3
Recognition accuracies of the state-of-the-art methods and the proposed iterative local re-ranking method on the UoM-SGFS SetB Database.

Algorithms	Rank-1 Accuracy	Rank-10 Accuracy
PCA [55]	5.33%	9.87%
CBR [30]	7.60%	25.47%
VGGFace [56]	16.13%	48.00%
P-RS [16]	40.80%	70.80%
DEEPS [42]	52.17%	82.67%
DLFace [54]	72.53%	94.80%
Baseline	84.00%	98.00%
SCA [32]	86.67%	94.40%
k-RF [33]	88.80%	98.40%
Proposed	**89.73%**	**98.40%**

Table 4
Rank-10 recognition accuracies of the state-of-the-art methods and the proposed iterative local re-ranking method on the Forensic Sketch Database.

Algorithms	Accuracy	Algorithms	Accuracy
PCA [55]	6.07%	Fisherface [49]	10.71%
P-RS [16]	11.21%	G-HFR [12]	18.21%
SGR-DA [53]	39.00%	DLFace [54]	40.73%
Baseline	24.64%	SCA [32]	31.25%
k-RF [33]	34.29%	**Proposed**	**41.43%**

4.2.4. Experiments on forensic sketch database

Experiments on matching real-world forensic sketches with mug shot photos are conducted finally. For experiemnts on the forensic sketch database, the CUHK AR database is chosen as the training dataset. Following the same protocol in [12], 112 subjects are selected randomly as the training set, and the remaining 56 subjects are uses as the testing set. To mimic the large modality gap between forensic sketches and mug shot photos, the similar strategy is followed with [24]. Here the CUHK AR database is taken as the training pairs. Firstly, photos in mugshot dataset are transformed to synthesize photos by performing the aforementioned face sketch synthesis method. Then the forensic sketches are also transformed into synthesized photos with the same synthesis method. When transforming both forensic sketches and mugshot photos with the same image style, the proposed iterative local re-ranking method is utilized for matching sketches.

Benefiting from the iterative local re-ranking with attribute guided synthesis strategy, the proposed method could achieves 41.43% which gains an impressive increase of 23.22% in rank-10 accuracy compared with the state-of-the-art method [12] as shown in Table 4.

4.3. Parameters analysis

In this subsection, the impact of several parameters is analyzed in the proposed algorithm on the face sketch recognition performance. The proposed iterative local re-ranking with attribute guided synthesis method is evaluated on E-PRIP Sketch Database under different parameter settings in Fig. 4 and Fig. 5.

Fig. 3 shows example probe sketches and corresponding synthesized photos with proposed attribute guided face sketch synthesis method. In the second row, the attribute Big Nose guided synthesized photo seems more similar with mugshot photo compared with the original face sketch. It is the unavoidable biases in the generation process that introduce shape distortions. Experiments show the proposed synthesis algorithm could help eliminate these biases.

Influence of parameter k_1 The effect of parameter k_1 is investigated in the E-PRIP sketch database. k_2 is fixed at 3, w is fixed at 0.2 and λ is fixed at 0.2. As k_1 grows, the rank-10 accuracy first rises with fluctuations and after arriving at the optimal point it initiates a slow descent with fluctuations. The optimal point is around $k_1 = 26$. When k_1 is too large, superfluous similarities from neighbours could be taken into consideration, which would make performance poor.

Influence of parameter k_2 The effect of parameter k_2 is also investigated in the E-PRIP sketch database. k_1 is fixed at 26, w is fixed at 0.2 and λ is fixed at 0.2. As k_2 increases, the rank-10 accuracy increase in a range and then decline steadily. It could be found that when k_2 reaches approximately 3 leading better performance. It is because when too much value of k_2 is assigned, more images of different identities would be included in the neighbor enhancement, resulting in a decline in performance.

Influence of parameter w The impact of the parameter w is shown in the left subfigure of Fig. 5. k_1 is fixed at 26, k_2 is fixed at 3 and λ is fixed at 0.2. As illustrated before, the w balances the importance of two different types of local constraint encoding features. With increase of w, the rank-10 accuracy first increase with fluctuations and after reaching the peak at $w = 0.2$, it starts a decline. Experiment shows the combination of them is better, which proves these two features contain complementary discriminative information.

Influence of parameter λ The impact of the parameter λ is shown in the right subfigure of Fig. 5. k_1 is fixed at 26, k_2 is fixed at 3 and w is fixed at 0.2. The optimal point is around $\lambda = 0.2$. As illustrated before, the λ denotes the important aggregating factor for iterative local aggregation. The rank-10 accuracy first rises and then drops steadily. Experiment shows the proposed iterative local aggregation method indeed adds more discriminative contextual information from different face components to improve performance.

4.4. Further evaluations

Ablation Study In the ablation study, experiments are conducted to further evaluate the performance of each component of proposed algorithm on E-PRIP Sketch Database and PRIP-VSGC Sketch Database. For convenience of comparison, the same parameters setting is utilized as shown in Section 4.3. As shown in

Fig. 3. Synthesized results of face sketches on the E-PRIP Sketch Database. (a) Face sketches. (b) Mugshot photos. (c) Synthesized photos in the first stage. (d)-(h) separately shows the synthesized photos guided with attributes Black Hair, Blond Hair, Brown Hair, Big Nose and Chubby in the second stage.

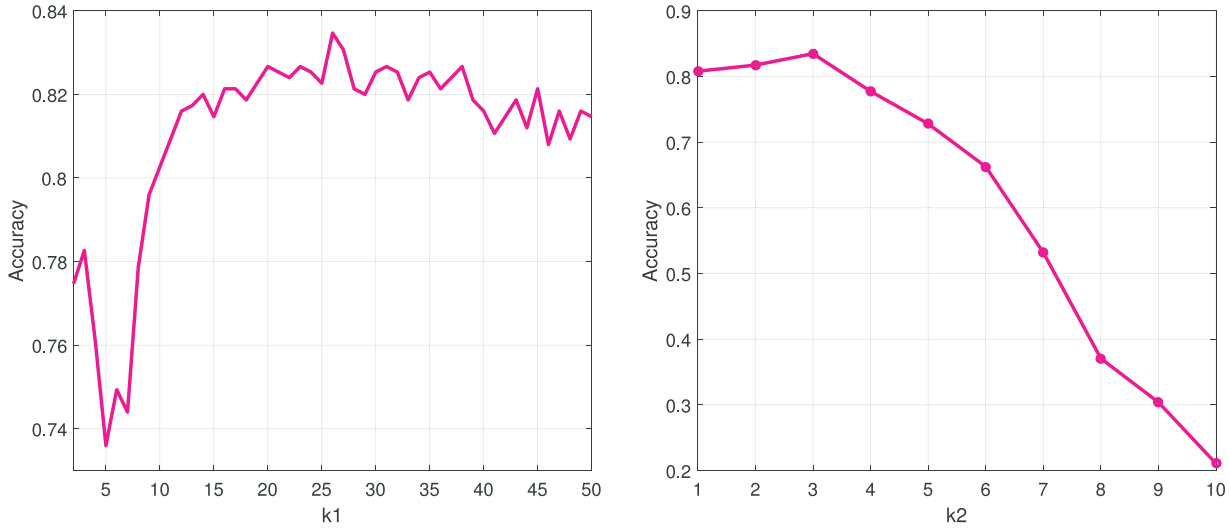


Fig. 4. Left subfigure shows the accuracies of different numbers of k_1 at rank-10; right subfigure shows the accuracies of different numbers of k_2 at rank-10. All the experiments are conducted on the E-PRIP database.

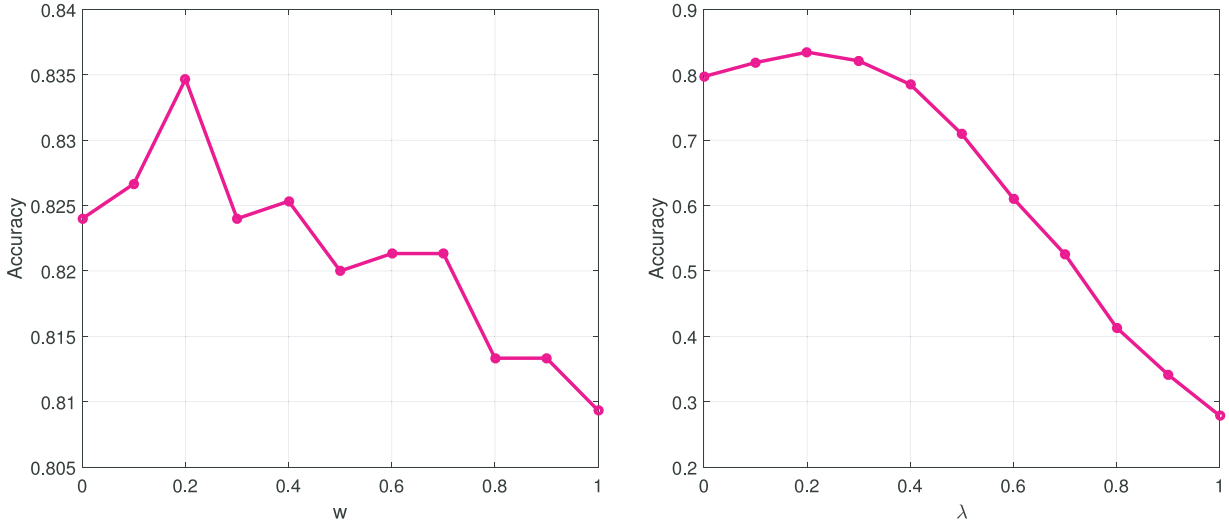


Fig. 5. Left subfigure shows the accuracies of different numbers of w at rank-10; right subfigure shows the accuracies of different numbers of λ at rank-10. All the experiments are conducted on the E-PRIP database.

Table 5
Ablation study on the E-PRIP and PRIP-VSGC Sketch Databases.

Components	Rank-10(E-PRIP)	Rank-10 (PRIP-VSGC)
Without Attribute Guided Face Sketch Synthesis	78.13%	66.00%
Without Iterative Local Re-ranking	79.73%	64.53%
Proposed full algorithm	**83.47%**	**68.67%**

Table 5, removing the attribute guided face sketch synthesis will degrade the recognition accuracy. It is because that these generated images with varying local characteristic could help eliminate unavoidable biases, which are introduced in face sketch generation process. Additionally, it is clear that utilizing iterative local re-ranking could effectively integrate the inherent local information to improve face recognition performance. It is noted that the proposed local iterated re-ranking is inspired from the iterative scheme introduced in [14], but differs from that. The differences are clarified on three aspects: 1. [14] aims to explore the diverse information embedded in features and fuses divided features for re-ranking. Different from that, our goal is to utilize more extra local modality-invariant discriminative information to dress the modality gap, which is inspired from researches [18,30] in face sketch community; 2. The intermedia distance in [14] includes the neighborhood information given by the initial distance. However, the proposed intermedia distance in Eq. (14) is calculated by the extra local region discriminative information but not the initial distance; 3. In [14] authors manually conduct two-iteration encoding in experiments. In the proposed method, five-iteration feature encoding is conducted naturally to fuse complementary information for recognition, due to local discriminative information extracted from five local regions (Hair, Brows, Eye, Nose and Mouth). Considering the large shape exaggerations and distortions in face sketches, the proposed iteration local re-ranking aims to integrate inherent local invariant information to mimic the modality gap.

Example Results Example probe sketches matching results are shown in Fig. 6. Experimental results show the proposed method could make mug shot photos ranks higher in the ranking list, which would be missed in the ranking list of the baseline method.

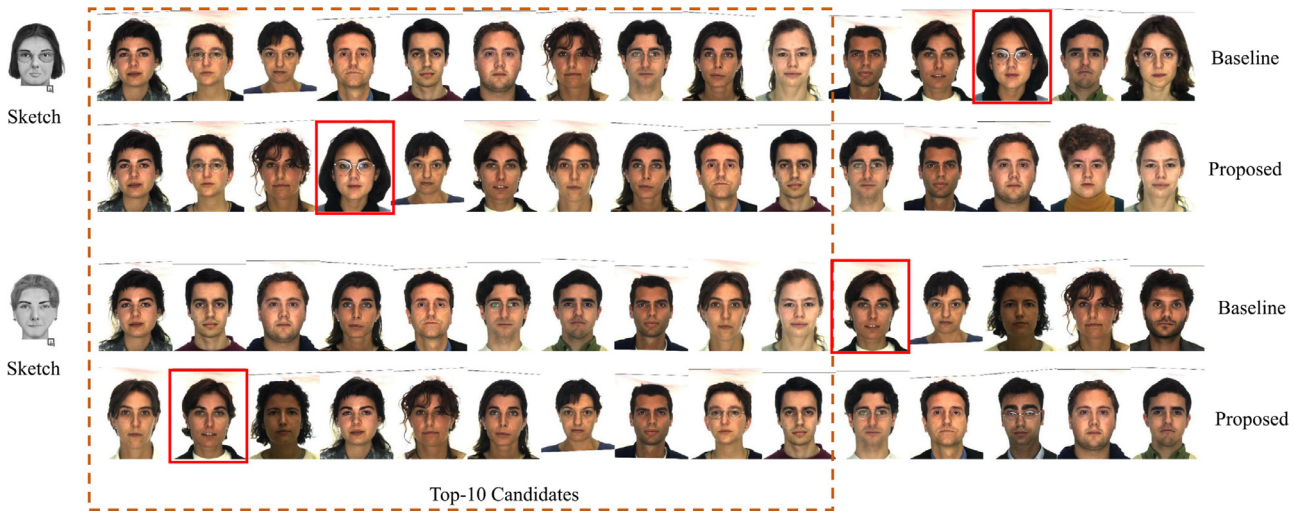


Fig. 6. Example results of two probe sketches on the E-PRIP sketch database. For each probe sketch, the first and the second rows correspond to the ranking results produced by Baseline and the proposed method respectively. Photo surrounded by the red box denotes the same identity as the sketch. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Example results show that the proposed iterative local re-ranking with attribute guided synthesis algorithm could effectively boost the recognition performance.

5. Conclusion

A novel iterative local re-ranking with attribute guided synthesis algorithm is proposed for face sketch recognition in this paper. Considering the unavoidable biases introduced in the generation process, the clues of face attributes are utilized to synthesis photos with varying local characteristic, which could help mimic the large difference between sketches and photos. Besides, an iterative local re-ranking algorithm is proposed to extract contextual features integrated with local discriminative information. Experiments on E-PRIP database, PRIP-VSGC database, UoM-SGFS database and Forensic Sketch database illustrate the effectiveness of the proposed method. The key benefit of the proposed method is that the inevitable biases elimination and the local invariant discriminate information are crucial for face sketch recognition. In the future, further researches would focus on: (1) evaluating the recognition performance on more heterogeneous face recognition scenarios, (2) discovering the cross modality attribute information to eliminate unavoidable biases of face sketches, (3) integrating the local invariant discriminative information more effectively.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

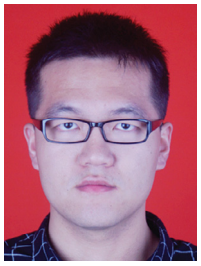
Ackonwledgements

This work was supported in part by the National Key Research and Development Program of China under Grant 2018AAA0103202 and Grant 2016QY01W0200, in part by the National Natural Science Foundation of China under Grant 61922066, Grant 61876142, Grant 61806152, Grant 61671339, Grant 61772402, Grant U1605252, and Grant 61432014, in part by the National High-Level Talents Special Support Program of China under Grant CS31117200001, in part by the China Post-Doctoral Science Foundation under Grant 2018M631124 and Grant 2019T120880, in part by the Fundamental Research Funds for the Central Universities under Grant JB190117 and Grant JB191502, in part by the China 111 Project under Grant B16037, in part by Natural Science Basic Research Plan in Shaanxi Province of China under Grant 2019JM-289, in part by the Key Research and Development Program of Shaanxi under Grant 2020ZDLGY08-08, in part by the Innovation Fund of Xidian University, and in part by the Xidian University-Intellifusion Joint Innovation Laboratory of Artificial Intelligence.

References

[1] C.D. Frowd, W.B. Erickson, J.M. Lampinen, F.C. Skelton, A.H. McIntyre, P.J. Hancock, A decade of evolving composites: regression-and meta-analysis, Journal of Forensic Practice 17 (4) (2015) 319–334.
[2] B. Klare, Z. Li, A. Jain, Matching forensic sketches to mug shot photos, IEEE Trans. Pattern Anal. Mach. Intell. 33 (3) (Mar. 2011) 639–646.
[3] P. Mittal, M. Vatsa, R. Singh, Composite sketch recognition via deep network, in: Proc. Int. Conf. Biom., 2015, pp. 1091–1097.
[4] J. Lu, V.E. Liong, J. Zhou, Simultaneous local binary feature learning and encoding for homogeneous and heterogeneous face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 40 (8) (2018) 1979–1993.
[5] D. Lin, X. Tang, Inter-modality face recognition, in: Proc. Eur. Conf. Comput. Vis., 2006, pp. 13–26.
[6] M. Kan, S. Shan, H. Zhang, S. Lao, X. Chen, Multi-view discriminant analysis, IEEE Trans. Pattern Anal. Mach. Intell. 38 (1) (2016) 188–194.
[7] J. Huo, Y. Gao, Y. Shi, W. Yang, H. Yin, Heterogeneous face recognition by margin-based cross-modality metric learning, IEEE Trans. Cybern. 48 (6) (2017) 1814–1826.
[8] J. Huo, Y. Gao, Y. Shi, H. Yin, Cross-modal metric learning for auc optimization, IEEE Trans. Neural Netw. Learn. Syst. 29 (10) (2018) 4844–4856.
[9] X. Tang, X. Wang, Face sketch synthesis and recognition, in: Proc. IEEE Int. Conf. Comput. Vis., 2003, pp. 687–694.
[10] Q. Liu, X. Tang, H. Jin, H. Lu, S. Ma, A nonlinear approach for face sketch synthesis and recognition, in: Proc. IEEE Conf. Comput. Vis. Pattern Recogn., 2005, pp. 1005–1010.
[11] X. Wang, X. Tang, Face photo-sketch synthesis and recognition, IEEE Trans. Pattern Anal. Mach. Intell. 31 (11) (2009) 1955–1967.
[12] C. Peng, X. Gao, N. Wang, J. Li, Graphical representation for heterogeneous face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 39 (2) (2017) 301–312.
[13] S. Ouyang, T.M. Hospedales, Y.-Z. Song, X. Li, Forgetmenot: memory-aware forensic facial sketch matching, in: Proc. IEEE Conf. Comput. Vis. Pattern Recogn., 2016, pp. 5571–5579.
[14] R. Yu, Z. Zhou, S. Bai, X. Bai, Divide and fuse: A re-ranking approach for person re-identification, in: British Machine Vision Conference, 2017.
[15] S. Liao, D. Yi, Z. Lei, R. Qin, S.Z. Li, Heterogeneous face recognition from local structures of normalized appearance, in: Proc. Int. Conf. Biom., 2009, pp. 209–218.
[16] B.F. Klare, A.K. Jain, Heterogeneous face recognition using kernel prototype similarities, IEEE Trans. Pattern Anal. Mach. Intell. 35 (6) (2013) 1410–1422.
[17] D. Gong, Z. Li, W. Huang, X. Li, D. Tao, Heterogeneous face recognition: a common encoding feature discriminant approach, IEEE Trans. Image Process. 26 (5) (2017) 2079–2089.
[18] D. Liu, J. Li, N. Wang, C. Peng, X. Gao, Composite components-based face sketch recognition, Neurocomputing 302 (2018) 46–54.

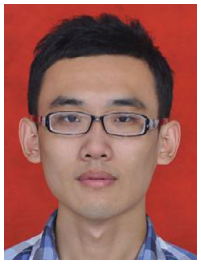
[19] Z. Lei, S. Li, Coupled spectral regression for matching heterogeneous faces, in: Proc. IEEE Conf. Comput. Vis. Pattern Recogn., 2009, pp. 1123–1128.
[20] D. Liu, N. Wang, C. Peng, J. Li, X. Gao, Deep attribute guided representaiton for heterogeneous face recognition, in: Proc. Int. Joint. Conf. Artificial Intell., 2018, pp. 835–841.
[21] H. Zhou, Z. Kuang, K.-Y.K. Wong, Markov weight fields for face sketch synthesis, in: Proc. IEEE Conf. Comput. Vis. Pattern Recogn., 2012, pp. 1091–1097.
[22] J. Zhong, X. Gao, C. Tian, Face sketch synthesis using e-hmm and selective ensemble, in: Proc. IEEE Int. Conf. Acoustics, Speech Signal Process, 2007, pp. 485–488.
[23] N. Wang, D. Tao, X. Gao, X. Li, J. Li, Transductive face sketch-photo synthesis, IEEE Trans. Neural Netw. Learn. Syst. 24 (9) (Sep. 2013) 1364–1376.
[24] C. Peng, X. Gao, N. Wang, D. Tao, X. Li, J. Li, Multiple representations-based face sketch–photo synthesis, IEEE Trans. Neural Netw. Learn. Syst. 27 (11) (2016) 2201–2215.
[25] B. Cao, N. Wang, X. Gao, J. Li, Asymmetric joint learning for heterogeneous face recognition, in: Proc. AAAI, 2018, pp. 1–8.
[26] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: Proc. IEEE Conf. Comput. Vis. Pattern Recogn., 2017, pp. 5967–5976.
[27] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proc. IEEE Int. Conf. Comput. Vis., 2017, pp. 2242–2251.
[28] L. Wang, V. Sindagi, V. Patel, High-quality facial photo-sketch synthesis using multi-adversarial networks, in: Proc. IEEE Int. Conf. Face Gesture Recogn., 2018, pp. 83–90.
[29] J. Yu, S. Shi, F. Gao, D. Tao, Q. Huang, Composition-aided face photo-sketch synthesis, arXiv preprint arXiv:1712.00899 (2017).
[30] H. Han, B. Klare, K. Bonnen, A. Jain, Matching composite sketches to face photos: a component-based approach, IEEE Trans. Inf. Forens. Security 88 (1) (Nov. 2013) 191–204.
[31] T. Mei, Y. Rui, S. Li, Q. Tian, Multimedia search reranking: a literature survey, ACM Comput Surv 46 (3) (2014) 38.
[32] S. Bai, X. Bai, Sparse contextual activation for efficient visual re-ranking, IEEE Trans. Image Process. 25 (3) (2016) 1056–1069.
[33] Z. Zhong, L. Zheng, D. Cao, S. Li, Re-ranking person re-identification with k-reciprocal encoding, in: Proc. IEEE Conf. Comput. Vis. Pattern Recogn., 2017, pp. 3652–3661.
[34] S. Bai, Z. Zhou, J. Wang, X. Bai, L. Jan Latecki, Q. Tian, Ensemble diffusion for retrieval, in: Proc. IEEE Int. Conf. Comput. Vis., 2017, pp. 774–783.
[35] S. Bai, X. Bai, Q. Tian, Scalable person re-identification on supervised smoothed manifold, in: Proc. IEEE Conf. Comput. Vis. Pattern Recogn., 2017, pp. 2530–2539.
[36] S. Bai, P. Tang, P.H. Torr, L.J. Latecki, Re-ranking via metric fusion for object retrieval and person re-identification, in: Proc. IEEE Conf. Comput. Vis. Pattern Recogn., 2019, pp. 740–749.
[37] Y. Choi, M.-J. Choi, M. Kim, J.-W. Ha, S. Kim, J. Choo, Stargan: Unified generative adversarial networks for multi-domain image-to-image translation, in: Proc. IEEE Conf. Comput. Vis. Pattern Recogn., 2018, pp. 8789–8797.
[38] Z. Liu, P. Luo, X. Wang, X. Tang, Deep learning face attributes in the wild, in: Proc. IEEE Int. Conf. Comput. Vis., 2015, pp. 3730–3738.
[39] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556 (2014).
[40] P. Mittal, A. Jain, G. Goswami, R. Singh, M. Vatsa, Recognizing composite sketches with digital face images via ssd dictionary, in: Proc. IEEE Int. Joint Con. Biometrics, 2014, pp. 1–6.
[41] S.J. Klum, H. Han, B.F. Klare, A.K. Jain, The facesketchid system: matching facial composites to mugshots, IEEE Trans. Inform Forensics. Securi. 9 (12) (2014) 2248–2263.
[42] C. Galea, R.A. Farrugia, Matching software-generated sketches to face photographs with a very deep cnn, morphed faces, and transfer learning, IEEE Trans. Inf. Forens. Security 13 (6) (2018) 1421–1431.
[43] A.M. Martinez, The ar face database, CVC Technical Report24 (1998).
[44] FACES. [Online]. Available: http://www.iqbiometrix.com.
[45] Identi-Kit. [Online]. Available: http://www.identikit.net.
[46] S. Saxena, J. Verbeek, Heterogeneous face recognition with cnns, in: Proc. Eur. Conf. Comput. Vis., 2016, pp. 483–491.
[47] X. Wu, R. He, Z. Sun, T. Tan, A light CNN for deep face representation with noisy labels, IEEE Trans. Inform. Forens. Security 13 (11) (2018) 2884–2896.
[48] L.-F. Chen, H.-Y.M. Liao, M.-T. Ko, J.-C. Lin, G.-J. Yu, A new LDA-based face recognition system which can solve the small sample size problem, Pattern Recogn. 33 (10) (2000) 1713–1726.
[49] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces *vs.* fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711–720.
[50] H.S. Bhatt, S. Bharadwaj, R. Singh, M. Vatsa, Memetically optimized mcwld for matching sketches with digital face images, IEEE Trans. Inf. Forens. Security 7 (5) (2012) 1522–1535.
[51] H. Kazemi, S. Soleymani, A. Dabouei, M. Iranmanesh, N.M. Nasrabadi, Attribute-centered loss for soft-biometrics guided face sketch-photo recognition, in: Proc. IEEE Conf. Comput. Vis. Pattern Recogn., 2018, pp. 499–507.
[52] H. Cheraghi, H.J. Lee, Sp-net: a novel framework to identify composite sketch, IEEE Access 7 (2019) 131749–131757.
[53] C. Peng, X. Gao, N. Wang, J. Li, Sparse graphical representation based discriminant analysis for heterogeneous face recognition, Signal Processing 156 (2019) 46–61.
[54] C. Peng, N. Wang, J. Li, X. Gao, Dlface: deep local descriptor for cross-modality face recognition, Pattern Recognit 90 (2019) 161–171.
[55] I. Jolliffe, Principal Component Analysis, in: International encyclopedia of statistical science, Springer, 2011, pp. 1094–1096.
[56] O.M. Parkhi, A. Vedaldi, A. Zisserman, et al., Deep face recognition, in: Proc. British Conf. Machine Vis., 2015, pp. 1–12.



Decheng Liu received the B.Eng. degree in electronic and information engineering from Xidian University, Xi'an, China, in 2016, where he is currently pursuing the Ph.D. degree in intelligent information processing with the School of Electronic Engineering. His current research interests include computer vision and machine learning, especially for heterogeneous image analysis and its application.

Xinbo Gao received the B.Eng., M.Sc. and Ph.D. degrees in electronic engineering, signal and information processing from Xidian University, Xi'an, China, in 1994, 1997, and 1999, respectively. From 1997 to 1998, he was a research fellow at the Department of Computer Science, Shizuoka University, Shizuoka, Japan. From 2000 to 2001, he was a post-doctoral research fellow at the Department of Information Engineering, the Chinese University of Hong Kong, Hong Kong. Since 2001, he has been at the School of Electronic Engineering, Xidian University. He is currently a Cheung Kong Professor of Ministry of Education of P. R. China, a Professor of Pattern Recognition and Intelligent System of Xidian University and a Professor of Computer Science and Technology of Chongqing University of Posts and Telecommunications. His current research interests include Image processing, computer vision, multimedia analysis, machine learning and pattern recognition. He has published six books and around 300 technical articles in refereed journals and proceedings. Prof. Gao is on the Editorial Boards of several journals, including Signal Processing (Elsevier) and Neurocomputing (Elsevier). He served as the General Chair/Co-Chair, Program Committee Chair/Co-Chair, or PC Member for around 30 major international conferences. He is a Fellow of the Institute of Engineering and Technology and a Fellow of the Chinese Institute of Electronics.

Nannan Wang received the B.Sc. degree in information and computation science from the Xi'an University of Posts and Telecommunications in 2009 and the Ph.D. degree in information and telecommunications engineering from Xidian University in 2015. From September 2011 to September 2013, he was a Visiting Ph.D. Student with the University of Technology, Sydney, NSW, Australia. He is currently a Professor with the State Key Laboratory of Integrated Services Networks, Xidian University.He has published over 100 articles in refereed journals and proceedings, including IEEE T-PAMI, IJCV, NeurIPS, ECCV etc. His current research interests include computer vision, pattern recognition, and machine.



Chunlei Peng received the B.Sc degree in electronic and information engineering from Xidian University, Xi'an, China, in 2012. He received his Ph.D. degree in information and telecommunications engineering in 2017. Now, he works with the School of Cyber Engineering at Xidian University. From September 2016 to September 2017, he has been a visiting Ph.D. student with the Duke University, NC, USA. His current research interests include computer vision, pattern recognition, and machine learning.

Jie Li received the B.Sc. degree in electronic engineering, the M.Sc. degree in signal and information processing, and the Ph.D. degree in circuit and systems, from Xidian University, Xi'an, China, in 1995, 1998, and 2004, respectively. She is currently a Professor in the School of Electronic Engineering, Xidian University, China. Her research interests include image processing and machine learning. In these areas, she has published around 50 technical articles in refereed journals and proceedings including IEEE T-NNLS, T-IP, T-CSVT, Information Sciences etc.