Multimedia Tools and Applications (2021) 80:18007–18026
https://doi.org/10.1007/s11042-020-10301-0

Face sketch synthesis: a survey

Hongbo Bi[1] · Ziqi Liu[1] · Lina Yang[1] · Kang Wang[1] · Ning Li[2]

Received: 7 January 2020 / Revised: 6 October 2020 / Accepted: 22 December 2020 /
Published online: 13 February 2021

Abstract
Face sketch synthesis (FSS) has been widely applied to various computer vision tasks, such as criminal detection, information security, digital entertainment, etc. In the past several years, various FSS models with promising performance have been proposed. However, an in-depth understanding of these models in this topic remains lacking. The current survey: i) investigates few models; ii) classifies the models abstractly and monotonously; iii) lacks analysis of existing databases. iv) evaluates models in single evaluation metric. In this paper, we provide a comprehensive survey of the 50 state-of-the-art (SOTA) FSS models. Then we further describe the typical models objectively and analyze the results subjectively. Moreover, we divide these models into two main categories: traditional models and deep learning models. In addition, a novel classification is proposed: coefficient models and regression models. Finally, for the aforementioned problems, we discuss several challenges and highlight some directions of FSS for future research about new database and evaluation strategy.

Keywords Face sketch synthesis (FSS) · Face sketch-photo synthesis · Face hallucination · Traditional models · Deep learning models

1 Introduction

Human face, which plays an important and effective role to recognize human biological characteristic, has attracted extensive attention in the field of recognition. Face recognition [10, 77] has been widely used in a range of real-world applications, such as information security [38], image matching [34], object tracking [23], image conversion [60], etc. In the investigation of criminal case, it is usually difficult to obtain the image information of criminal suspects directly. However, the facial features of suspects described by witnesses are

✉ Ning Li
lnlndy@126.com

[1] School of Electrical and Information Engineering, Northeast Petroleum University, Daqing, China

[2] Chengdu Lead Science & Technology Co. Ltd, Chengdu, China

often readily available. Face sketches, which are drawn by artists based on the recall of the witness, can be used to retrieve the suspect. In addition, because sketches and photos are in different feature spaces, it is difficult to retrieve the face photos directly by using sketches. Further, face sketch synthesis (FSS) algorithm is proposed, which transforms the face photos into corresponding sketches. And the opposite transformation is achieved by exchanging objects. FSS algorithm is not only applicable to law enforcement, but also applicable to digital entertainment [61]. For most people and even artists, transforming the face photos to face sketches quickly and accurately is not an easy task. Nevertheless, it is relatively easy for a computer. For example, in the film industry, with the assistance of the automatic sketch synthesis system, artists can save a great amount of time when drawing cartoon images. Moreover, in mobile internet applications, the sketch filter in video chat allows people to enjoy the conversation. In addition, the system provides a simple tool for people to personalize their identities in digital world, such as Facebook profiles, cartoons, artistic rendering, etc. Considering the styles of sketch, the face photo-sketch synthesis contains two styles: line sketch and stone sketch. The former focuses only on facial lines, and the latter generates a fine-grained sketch. However, neither line nor stone sketch can be recognized by a computer. In 2002, the Chinese University of Hong Kong was committed to the research of face sketch synthesis. Tang et al. [47] employed the feature transformation to construct the face sketch. Specifically, a set of eigenfaces are first generated from training face photos. A weight vector is then obtained by mapping the face photo to the eigenface. Finally, the weighted sum of eigenfaces and calculated weight vectors represents the synthesized sketch. Note that the transformation between photos and sketches is assumed to be a linear mapping. However, it is clearly that the assumption is imprecise. To accord with the process of drawing sketch, Liu et al. [26] proposed a nonlinear model which was composed of local linear combinations. The nonlinear model segments the images into patches and sutures each reconstructed sketch patches. The above models are data-driven models, which have large amounts of computation. Subsequently, Liu et al. [27] proposed a model of learning the mapping relation between photos and sketches. These previous models work in a large photo-sketch pair and are limited to different lighting, poses, etc. Zhang et al. [72] proposed a robust model under few of photo-sketch pairs, which was not limited to frontal face. Recently, deep learning has gained increasing attention and many methods [28, 45, 49, 50, 70] have been developed in various fields. Zhang et al. [19] designed a model based on convolutional neural network (CNN) to learn the end-to-end photo-sketch mapping. Note that the model generates the entire sketch, not the patchwork sketch patches. Isola et al. [41] proposed a Generative Adversarial Network (GANs) to learn the mapping and train the loss function. And the loss function optimizes the relationship by guiding the mapping. In previous work, existing models are classified into two categories: Data-driven and Model-driven. In the Data-driven classification strategy, the models are divided into three types: Subspace learning, Sparse representation and Bayesian inference. Although this classification is detailed, it is hard to understand. In this paper, we compare and evaluate 50 FSS models which are classified into traditional models and deep learning models. In addition, the existing databases are relatively simple which limit the development of FSS. Therefore, we list and analyze the common databases in an attempt to open directions for future work.

Our paper is organized as follows. In Section 2, we summarize and provide details for classification of existing FSS models. In Section 3, we summarize the common databases and evaluation methods for FSS models. We then evaluate several representative FSS models and discuss the challenges. Finally, we conclude this paper and open directions of this field in Section 4.

2 Models and classification

In this section, we provide systematic review of FSS models about the most popular classification and our classification. In addition, we also define models based on the results of synthesized sketches. We list 50 typical models, including 31 traditional models and 19 deep learning models. These models are listed in Table 1.

2.1 The popular classification

In the previous work, the models were divided into Data-driven models and Model-driven models, according to whether the training photo-sketch pairs participated in the online synthesis process. Among them, Data-driven models are divided into three categories according to the rationality: 1) Subspace learning; 2) Sparse representation; 3) Bayesian inference.

i) **Subspace learning**: As far as we know, color image is actually a three-dimensional matrix in mathematics, which is unfavorable for feature extraction. Subspace learning solves the aforementioned problem by mapping images from high dimensional space to low dimensional space. Linear subspace learning includes principal component analysis (PCA) [52] and Linear discriminant analysis algorithm (LDA). Inspired from PCA, Tang et al. [46] approximated the relationship between photos and sketches as a global linear mapping. However, Liu et al. [26] regarded the mapping as locally linear. They obtained the mapping coefficient through different methods, which were then used to carry out linear combination on the training sketches. Finally, the synthetic sketch was obtained.

ii) **Sparse representation**: Sparse representation is a process of decomposing the original signal, which is represented by a dictionary (also known as an over-complete basis) acquired in advance. It represents the input signal as the linear approximation of the dictionary. The natural image itself is a sparse signal, that is, the input signal is linearly represented by an over-complete dictionary. When the coefficients satisfy a certain degree of sparsity, similar signals are obtained. Chang et al. [5] proposed a face sketch synthesis method inspired by the sparse signal representation, which trained over-complete dictionaries in training photo-sketch pairs. The test photos are represented by the photo dictionary to obtain a series of sparse coefficients. Finally, the model gets the corresponding sketches by replacing the photo dictionary with the sketch dictionary.

iii) **Bayesian inference**: Bayesian inference is simply the inference of events s (have yet to occur) based on events t (have already occurred). In face sketch synthesis, bayesian inference is expressed as the maximum posterior probability, i.e.,

$$max_s P(s \mid t) \propto max_s P(t \mid s) P(s) \tag{1}$$

Moreover, there are other classifications [51] which divide the models into image-based models [14, 29] and exemplar-based models [7]. The former models directly generate sketch stones by gray value and edge information of the input images. The latter models focus on learning the drawing style through the photo-sketch pairs. In general, the sketch generated by the model consists of either simple lines with no facial details or shadow effects and details.

Table 1 Current existing models, and sorted by their publication year. T = traditional models, D = deep learning models

No.	Model	Year	Pub	Cat.	
1	LGP [47]	2003	ICCV	–	
2	ET [48]	2004	TCSVT	–	
3	LLE [26]	2005	CVPR	–	
4	MRF [55]	2008	PAMI	–	
5	RMRF [72]	2010	ECCV	–	
6	SFS [17]	2012	TCSVT	–	
7	MDSR [57]	2011	ICIG	–	
8	MWF [87]	2012	CVPR	–	
9	SCDL [58]	2012	CVPR	–	
10	Trans [59]	2013	TNNLS	–	
11	SSD [44]	2014	ECCV	–	
12	SRGS [73]	2015	TIP	–	
13	SST [75]	2016	TIP	Coefficient model	
14	KD-Tree [74]	2016	ECCV	–	
15	MrFSPS [35]	2016	TNNLS	–	
16	2DDCM [51]	2016	TIP	–	T
17	Bayesian [62]	2017	TIP	–	
18	SPSP [76]	2017	TCSVT	–	
19	S-FSPS [36]	2017	TCSVT	–	
20	Unified [63]	2017	SP	–	
21	ArFSPS [25]	2017	NC	–	
22	CMSG [83]	2018	Cybernetics	–	
23	Fast-RSLCR [66]	2018	PR	–	
24	MRNF [84]	2018	IJCAI	–	
25	RL [20]	2018	ICASSP	–	
26	ANI [67]	2018	TCSVT	–	
27	BTI [27]	2007	IJCAI	–	
28	EHMM-SE [15]	2008	NC	–	
29	EHMM [16]	2008	TCSVT	Regression model	
30	SL [69]	2010	NC	–	
31	LR [64]	2017	NC	–	
32	FCN [80]	2015	ICMR	–	
33	BFCN [81]	2017	TIP	–	
34	DGFL [88]	2017	IJCAI	–	
35	CA-GAN [18]	2017	CVPR	–	
36	cGANs [19]	2017	CVPR	–	
37	BP-GAN [68]	2017	PRL	–	
38	PCF [9]	2018	WCACV	–	
39	PS^2-MAN [65]	2018	CS	–	
40	FSSN [21]	2018	PR	–	
41	Col-cGAN [89]	2019	TNNLS	Regression model	D
42	MDAL [82]	2019	TNNLS	–	
43	MS-cGAN [3]	2019	ICIP	–	

Table 1 (continued)

No.	Model	Year	Pub	Cat.
44	TTGAN [71]	2019	NC	–
45	pFCN [30]	2019	NC	–
46	JDRL [22]	2019	TIP	–
47	CLRR [85]	2020	TIP	–
48	IACycleGAN [13]	2020	PR	–
49	NPGM [86]	2020	TNNLS	–
50	msGAN [24]	2020	NC	–

2.2 Proposed classification

The classification of the aforementioned models is overlapped, indistinct and complex. Therefore, according to the mathematical mechanism of the models, a novel category is proposed and divided into the following two categories.

2.2.1 Coefficient models

In simple terms, the coefficient model can calculate linear or nonlinear coefficients between photos and sketches. For the coefficient model, we need first to divide the training sketch-photo pair and the test photo into many image blocks (regular or irregular image blocks). Some coverage between adjacent image blocks should be retained, among which the photo blocks are represented by gray scale or other features. For each test photo block t_i with $i = 1, 2, ..., N$, we choose K nearest neighbors x_k^i with $k = 1, 2, ..., K$. From the training photo block, it is assumed that the test photo block is obtained by linear combination of K nearest neighbors, i.e.,

$$t_i = \sum_{k=1}^{K} a_{ik} x_k^i \tag{2}$$

The training data participates in the synthesis process, and each test photo block is represented by the training photo block linearly or non-linearly. The coefficient model assumes that the composite sketch block s_i and the test photo block t_i have similar manifold distribution. That means that they have the same reconstruction coefficient a_i. Therefore, the key to the problem is how to obtain the reconstruction weight, i.e.,

$$s_i = \sum_{k=1}^{K} a_{ik} y_k^i \tag{3}$$

The test sketches corresponding to the test photos are represented by the calculated coefficient and the training sketch blocks with linear or nonlinear weighted sums. Thus, all the sketch blocks are fused together to form a new sketch.

2.2.2 Regression models

Regression model is a mathematical model to quantitatively describe the statistical relations. In other words, by learning the mapping between different samples (photos and sketches), we quantitatively describe the statistical relationship between them. Different from the training samples required by the coefficient models in the testing process, regression models learn the mapping or regression relationship (linear or nonlinear) from the photo blocks to the sketch blocks in the training process. During the test, for any test photo blocks (location), we then utilize the learned mapping function (corresponding location) for regression. Regression models are divided into linear regression and nonlinear regression. Compared with the coefficient models, regression models have higher synthesis efficiency.

2.3 The classification of results

In addition, the aforementioned classifications focus on the algorithm. In fact, we also can group these models into other groups based on the results. These models can also be divided into two categories: Stitching models and Synthetic models. The former models seam the synthetic sketches of the training patch set. However, the latter models generate the final sketches which are not in the training database. Most of existing models [16, 26, 44, 72] generate the synthetic sketches by stitching the synthetic sketch patches. Some models [57, 87] synthesize a new patch that is not in the training database. To some extent, the former models select the training patches, while the latter models synthesize the sketches. More details can be learnt from the similar tasks [4]. Considering the significant distinction, we analyze the process of models and summarize it as follows:

2.3.1 Stitching models

Stitching models only represent the object sketches by the selected sketch patches from the training sketch patches. Due to the limited training patches, it leads to unreliable results especially for the facial details such as eyes, mouth, etc. The input photo $t = (t_1, t_2, ..., t_i)$ with $i \in 1...N$ is divided into N patches. And $t = (s_1, s_2, ..., s_i)$ with $i \in 1...N$ is the estimated sketch. The joint probability of the input photo and the corresponding sketch is written as

$$p(t_1, ..., t_i, s_1, ..., s_i) = \prod_{j_1, j_2} \Psi(t_{j_1}, t_{j_2}) \prod_{j} \Phi(t_j, s_j) \tag{4}$$

where j is the j^{th} patch. j_1 and j_2 are neighbor patches.

$$\Phi\left(\tilde{t}_j^{\iota}, s_j\right) = exp\left\{- \parallel \tilde{s}_j - s_j \parallel^2 / \sigma_e^2\right\} \tag{5}$$

where $\tilde{t}_j^{\iota}$ is ι^{th} the candidate photo patches corresponding to candidate sketch patches $\tilde{s}_j^{\iota}$.

$$\Psi\left(\tilde{t}_{j_1}^{\iota}, \tilde{t}_{j_2}^{m}\right) = exp\left\{- \parallel d_{j_1 j_2}^{\iota} - d_{j_1 j_2}^{m} \parallel^2 / \sigma_c^2\right\} \tag{6}$$

where $d_{j_1 j_2}^{\iota}$ denotes the intensity or color vector of the $\tilde{t}_{j_2}^{\iota}$ at j_2 in overlapping region between j_1 and j_2. The sketch patches are then computed by maximum a posteriori probability (MAP) $\hat{t}_{jMAP}$ or minimum mean-square error (MMSE) $\hat{t}_{jMMSE}$, i.e.,

$$\hat{t}_{jMAP} = argmax_{[t_j]} max_{[t_i, i \neq j]} P(t_1, ..., t_N \mid s_1, ..., s_N) \tag{7}$$

where M_j^k is the message from node k to its neighbor node j, i.e.,

$$\hat{t}_{jMMSE} = \sum_{t_j} t_j \sum_{[t_i, i \neq j]} P(t_1, ..., t_N \mid s_1, ..., s_N) \tag{8}$$

The MMSE is computed as

$$\hat{t}_{jMMSE} = argmax_{[t_j]}\Phi(t_j, s_j)\prod_{k} M_j^k(t_j) \tag{9}$$

$$M_j^k = max_{[t_k]}\Psi(t_j, t_k)\Phi(t_k, s_k)\prod_{\iota\neq j}\tilde{M}_k^{\iota}(t_k) \tag{10}$$

$$\hat{t}_{jMMSE} = \sum_{t_j}\Psi(t_j, t_k)\Phi(t_k, s_k)\prod_{\iota\neq j}\tilde{M}_k^{\iota}(t_k) \tag{11}$$

2.3.2 Synthetic models

Considering the limitation of the former, the model is proposed to overcome the drawback. Zhou [87] proposed a novel model benefiting from Markov Random Field (MRF). Different from the previous models, each node in MRF represents a candidate patch, and each node corresponds to a series of weights for K candidate patches. The jointly probability of input photo patch t_i and the corresponding weight w_i with $i \in 1, ..., N$ are given by

$$p(t_1, ..., t_N, w_{1,...w_N}) \propto \prod_{i=1}\Phi(t_i, w_i) \prod_{(i,j)\in\Xi} \Psi(w_i, w_j) \tag{12}$$

where

$$\Phi(t_i, w_i) = exp\left\{-\left\| t_i - \sum_{k=1}^{K} w_{i,k}P_{i,k}\right\|^2 / \sigma_D^2\right\} \tag{13}$$

and

$$\Psi(w_i, w_j) = exp\left\{-\left\|\sum_{k=1}^{K} w_{i,k}o_{i,k}^{j} - \sum_{k=1}^{K} w_{j,k}o_{j,k}^{i}\right\|^2 / \sigma_S^2\right\}. \tag{14}$$

$(i, j) \in \Xi$ denotes that the i^{th} and j^{th} are neighbor patches. $o_{i,k}^{j}$ is the overlapping region of k^{th} candidate between i^{th} sketch patch and j^{th} patch. The posterior probability is

$$p(w_1, ..., w_N \mid t_1, ..., t_N) = \frac{1}{Z}p(t_1, ..., t_N, w_1, ..., w_N) \tag{15}$$

where $Z = p(t_1, ..., t_N)$ is the standardized term, and the best weight for candidate patches is calculated through the maximum posterior probability. It can be seen that compared with common MRF, the sketch patch is computed with the combination of candidate patches in each node. The former focuses on the relationship between the input photo and the estimated sketch, while the later emphasizes the weight of candidates.

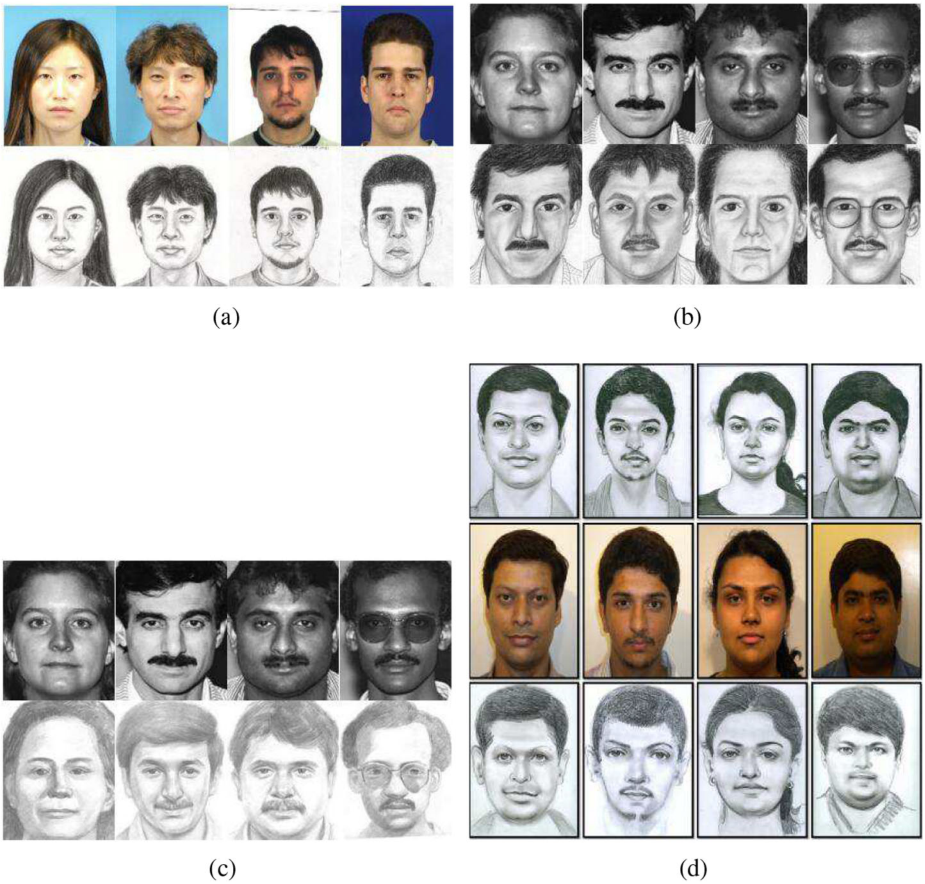
3 Experiment and analysis

3.1 Databases

Face recognition has received extensive attention, and various corresponding databases have been created. These face databases apply to different scenarios: CACD2000 [8] focuses on face age; IMDB-WKI [40] includes face photos with different expressions; CK+ [65] is used to recognize front face with various expressions. However, there are relatively few FSS databases. As far as we know, the common databases are **CUFS** [55], **CUFSF** [78], **VIPSL** [17, 37] and **IIIT-D** [2]. In the following paper, we expound and compare the four databases from three aspects respectively: i) quantity and size; ii) categories and attributes; iii) sketches drawing. Furthermore, the examples of each database are shown in Fig. 1.

CUFS has 606 photo-sketch pairs and includes three sub-databases: CUHK (188 photo-sketch pairs), AR (123 photo-sketch pairs) [31], XM2VTS (295 photo-sketch pairs) [32]. Besides, the size of photos and sketches is 200*250. CUFS considers five attributes: gender, light, pose, skin color and background. Although CUFS considers various attributes in the data set, the pose is controlled to be front. In the real world, this attribute is uncontrolled, especially in digital entertainment applications. Another limitation is the concentration of light on high light and soft light, which lacks diversity. Moreover, the sub-databases cover three backgrounds, including cyan, white and blue backgrounds. Noted that the background is simple and solid, which limits the scope of digital entertainment. In addition, different painting styles produce different sketches. In CUFS, all sketches are drawn by one artist, and the single style may lead to differences in evaluation results.

CUFSF contains 1194 photo-sketch pairs from FERET databases [39], and the size of photos is 200*250. Same as CUFS, CUFSF considers five attributes when constructing the data set. Besides, all face photos are in white and black, which increases the challenge of


(a)

(b)

(c)

(d)

Fig. 1 Samples of four existing databases **a** CUFS **b** CUFSF **c** VIPSL **d** IIIT-D

models due to the low contrast between the face and the background. The data set also adds different poses not limited to the front face. Considering that the light is an important factor affecting performance, the face photos are chosen with multiple lighting variations. In addition, the face area covers the photo and reduces background interference. Each sketch is drawn by the artist in an exaggerated way after looking at the photo. Although this style of paintings changes the facial structure to some extent, we can still distinguish the corresponding face photos. Moreover, the attributes require the models to pay more attention to the sketching process.

VIPSL collects 200 face photos from FERET database, FRAV2D database and Indian face database. Each face photo is drawn by five artists, including 1000 sketches. Besides, the size of photos is 479*724. The database limits the attributes of face to the front, normal light and neutral expression.

IIIT-D comprises three sub-databases: viewed database (238 photo-sketch pairs), semi-forensic database (140 photo-sketch pairs), forensic database (190 photo-sketch pairs). The database consists of photos with various sizes, and the diversity of photos presents a challenge to the models. Moreover, the three sub-databases represent three sketch types: 1) the sketches of viewed databases are drawn for digital images by the professional artist; 2) the artist is asked to sketch based on the memory of the viewer; 3) the sketch is generated from the description of an eyewitness based on his/her recollection of the crime scene.

3.2 Evaluation measure

So far, there are many measurements to evaluate image similarity. In addition, the most popular image quality assessment (IQA) of FSS are FSIM [79], UIQI [53], SSIM [42], VIF [43] and Scoot [11]. We briefly review several popular metrics for FSS evaluation as follows.

FSIM. FSIM utilizes low-level feature based on human visual system for full reference IQA. The index includes two stages: the local quality map computation and mapping operations. The feature similarity measurement between sketch $f_1(x)$ and sketch $f_2(x)$ is separated into two components: PC (phase congruence) and GM (gradient magnitude), i.e.,

$$S_{PC} = \frac{2PC_1(x) \cdot PC_2(x) + T_1}{PC_1^2(x) + PC_2^2(x) + T_1} \tag{16}$$

$$S_G = \frac{2G_1(x) \cdot G_2(x) + T_2}{G_1^2(x) + G_2^2(x) + T_2} \tag{17}$$

where T_1 and T_2 are positive constant. The local quality map is then computed, i.e.,

$$S_L(x) = [S_{PC}(x)]^{\alpha} \cdot [S_G(x)]^{\beta} \tag{18}$$

where α and β are parameters, and they are generally set to 1. Finally, the FSIM is defined as

$$FSIM = \frac{\sum_{x \in \Omega} S_L(x) \cdot PC_M(x)}{\sum_{x \in \Omega} PC_M(x)} \tag{19}$$

where $PC_M = \max(PC_1(x), PC_2(x))$, and Ω denotes the entire image spatial domain.

UIQI. $x = \{x_i | i = 1, 2, ..., N\}$ and $y = \{y_i | i = 1, 2, ..., N\}$ represent standard image and test image respectively. UIQI is defined as:

$$Q = \frac{4\sigma_{xy} \cdot \overline{xy}}{\left(\sigma_x^2 + \sigma_y^2\right)\left[(\overline{x})^2 + (\overline{y})^2\right]} \tag{20}$$

where

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i, \overline{y} = \frac{1}{N}\sum_{i=1}^{N} y_i \tag{21}$$

$$\sigma_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2, \sigma_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \overline{y}) \tag{22}$$

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y}) \tag{23}$$

SSIM. SSIM considers luminance, contrast and structure in three aspects respectively. These feature similarities are defined as follow:

$$\ell(x, Y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{24}$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{25}$$

$$s(x, y) = \frac{2\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{26}$$

where μ_x, μ_y represent mean intensity of image. σ_x, σ_y are defined as variance, and σ_{xy} is covariance. Overall similarity is a function of combination:

$$S(x, y) = f(\ell(x, y), c(x, y), s(x, y)) \tag{27}$$

VIF. VIF approximates image QA problem as the information fidelity problem. The indicator proposes an information measurement for quantitative information from reference image and test image.

$$VIF = \frac{\sum_{j \in subbands} I\left(\overrightarrow{C}^{N,j}; \overrightarrow{F}^{N,j}|\mathrm{s}^{N,j}\right)}{\sum_{j \in subbands} I\left(\overrightarrow{C}^{N,j}; \overrightarrow{E}^{N,j}|\mathrm{s}^{N,j}\right)} \tag{28}$$

where $\mathrm{I}\left(\overrightarrow{C}^{N,j}; \overrightarrow{F}^{N,j}|S^{N,j}\right)$ and $\mathrm{I}\left(\overrightarrow{C}^{N,j}; \overrightarrow{E}^{N,j}|S^{N,j}\right)$ denote the extracted information from j sub-band in reference image and test image. The all of sub-bands are derived from scale-space-orientation wavelet decomposition.

Scoot Measure. Fan [12] proposed the new structure similarity measure to solve previous evaluation problem that is failed to correctly ranked the synthesis sketch. The indicator builds measurement at "image-level" and concatenates two statistics as the Contrast and Energy (CE) feature to represent the image style. Scoot is defined as:

$$E_s = \frac{1}{1+ \parallel \overline{\Psi}\left(X_s'\right) - \overline{\Psi}\left(Y_s'\right) \parallel} \tag{29}$$

where X_s', Y_s' denote the quantized images, and $\overline{\Psi}\left(X_s'\right)$, $\overline{\Psi}\left(Y_s'\right)$ represent the average **CE** feature in four orientations.

3.3 Sketch recognition

In order to further confirm the validity of the models, previous work generally use recognition rate. Sketch recognition includes two strategies: a) all photos are transformed to sketches with synthesis algorithm, and the query sketches are matched with the synthesized sketches. b) the synthesized sketches are transformed to photos, and the synthesized photos then match the real photos in gallery. The existing sketch recognition methods include PCA, Bayesianface (Bayes) [33], Fisherface [1], dual-space LDA [54], null-space LDA [6], and Random Sampling LDA (RS-LDA) [56].

3.4 Experiment and analysis

3.4.1 Evaluation of databases

In this section, we evaluate and analysis the most popular databases CUFS and CUFSF in several aspects. **a)** Alignment. Considering the important role of facial contour in image matching, the sketches which are not aligned with the original face photo will affect the accuracy of recognition. As shown in Fig. 2, the degree of alignment is jagged in CUFS and CUFSF. How to achieve pixel-level alignment is the future direction that we need to consider. **b)** Edge detection. As a future direction, models learning the process of drawing is what we want to achieve. The artist first locates the outline and facial feature, and stone lines and detailed features are then added. The high accuracy of outline contributes to the higher recognition. Therefore, five edge detection algorithms are adopted to detect the outline of face photo. The results indicate that the outlines of face photo are relatively complete. However, the outstanding results are based on the simple background in CUFS and CUFSF. In Fig. 3, we collect a face photo with the complex background from Internet, and the results of edge detection are rough. In reality, the condition of background is out of control. Therefore, the database needs to be further improved.

3.4.2 Evaluation and analysis of models

In this section, we evaluate 9 typical models whose results are collected from the author homepage on the most popular CUHK student sub-database and CUFSF database. The

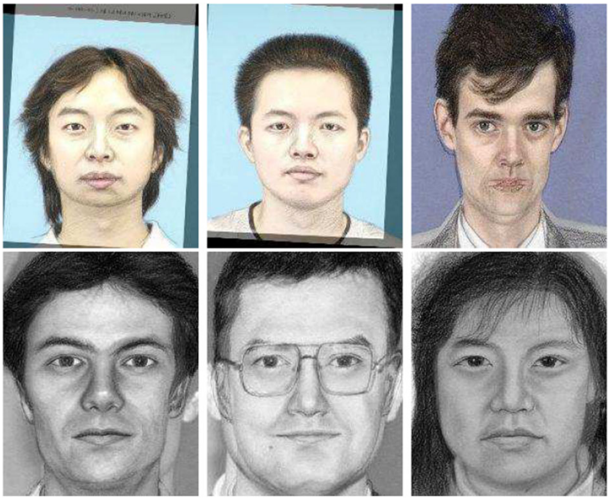


Fig. 2 Synthesized maps (paste the sketches on the original face photos), the first line is CUFS, and the second line is CUFSF

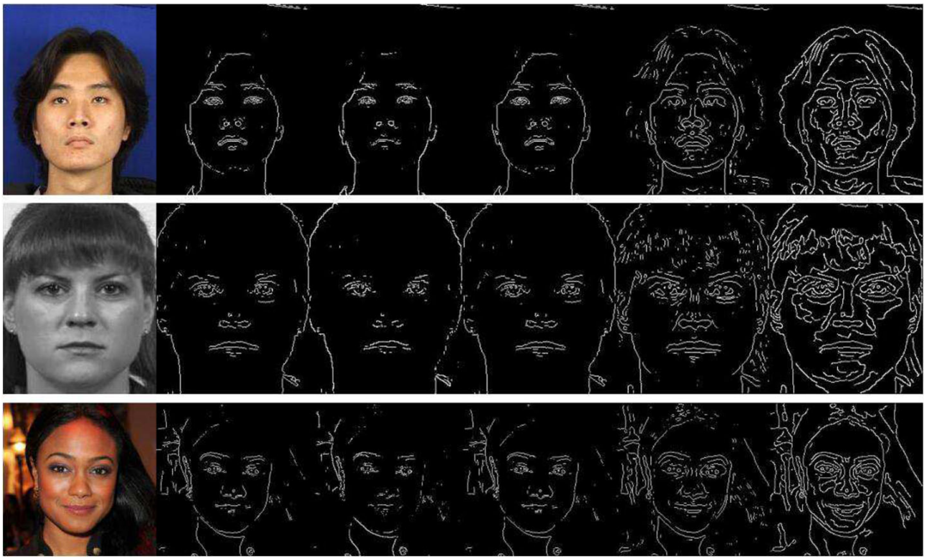


Fig. 3 Edge detection: Sobel; Roberts; Prewitt; Log; Canny. First line is CUFS sample, second line is CUFSF sample, and third line is Internet sample

results and evaluations are listed below. The previous measures, including SSIM, FSIM, UIQI, and VIF, pay attention to both structure information and texture information. However, due to the similarity of two sketches, both of them will be considered. Therefore, the previous measures are not enough to evaluate the sketch. Moreover, existing measures are sensitive to slight image degradation. In fact, sketches drawn by the artists should be robust. Scoot measure is committed to addressing the mentioned issues, and simultaneously evaluates 'block-level' spatial structure and co-occurrence texture statistics. In Fig. 4, we rank the five models based on the results, and the synthesized sketches evaluated by Scoot are consistent with human perception. The evaluation of synthesized sketches can be seen in Table 4. In summary, we analysis the models based on the results of Scoot measure in the following paper.

In Table 2, the top three models are GAN, MRF and LLE, and FCN produces the worst results. As can be seen from Fig. 5, the top three models preserve more detail and texture, and less blur than other models. FCN has a complete outline, but it generates a severe blur. In Table 3, similarly, GAN, MRF, and LLE generate the optimal results, while FCN is relatively poor. As shown in Fig. 6, the blur effect is serious, and especially the glasses are hard to describe. In spite of this, the evaluation index shows that the GAN model obtains excellent results. However, in fact the results have a large deformation compared with original images.

As far as we know, facial features mainly contain eyes, mouth, nose, and facial outline. The better local facial features, the better synthesized sketches. Previous work focused on the similarity of whole sketch and evaluated the models. This strategy ignores the evaluation of local facial features, and it is hard to learn what contributes to the excellent results. Therefore, we remove the mouth, nose, and eyes from the sketch example. In order to ensure the fairness of evaluation, different images are cropped at the same position at the pixel level, as shown in Fig. 7.

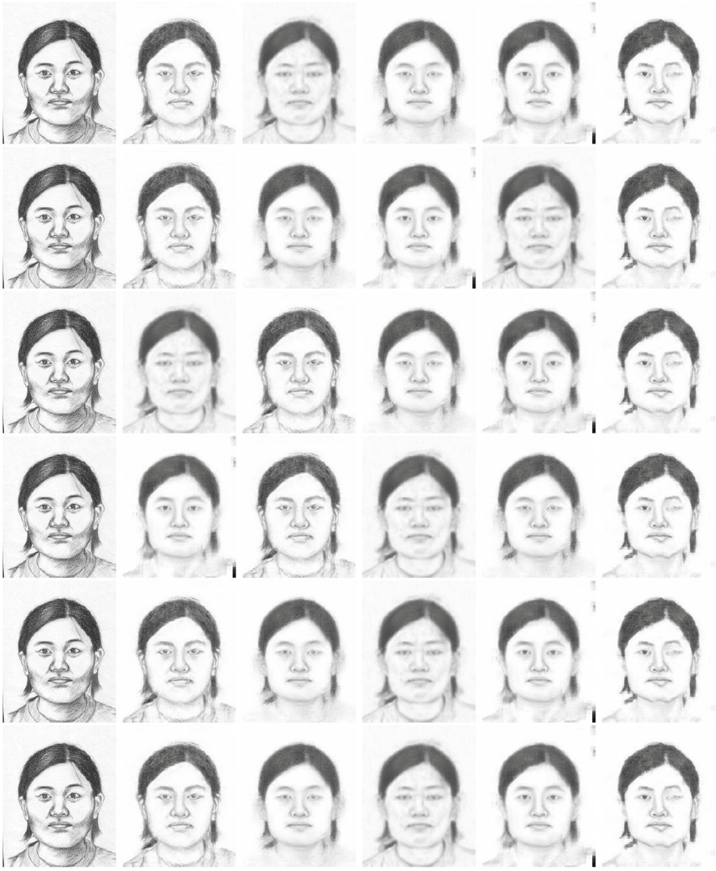


Fig. 4 In the accuracy of popular measures. We rank five models. From the first column to the sixth column are: Sketch, GAN, LLE, FCN, MWF, MRF. From the first row to the sixth row are: VIF, UIQI, SSIM, FSIM, Scoot and human perception. Compared to the human ranking (last line), except for Scoot, other measures do not match the human ranking

Table 2 The results of five evaluation measures on CUHK student database

No.	Model	Scoot	SSIM	FSIM	UIQI	VIF
1	MrFSPS	0.4841	0.6299	0.7194	0.9599	0.0978
2	Fast-RSLCR	0.4704	0.6345	0.7220	0.9624	0.1024
3	FCN	0.4603	0.6090	0.7071	0.9613	0.0896
4	GAN	0.5343	0.5983	0.7314	0.9586	0.0833
5	LLE	0.4921	0.6019	0.7139	0.9600	0.0934
6	MRF	0.5045	0.6023	0.7261	0.9592	0.0797
7	MWF	0.4864	0.6224	0.7347	0.9616	0.0954
8	RSLCR	0.4677	0.6360	0.7226	0.9625	0.1041
9	SSD	0.4735	0.6372	0.7237	0.9595	0.1035

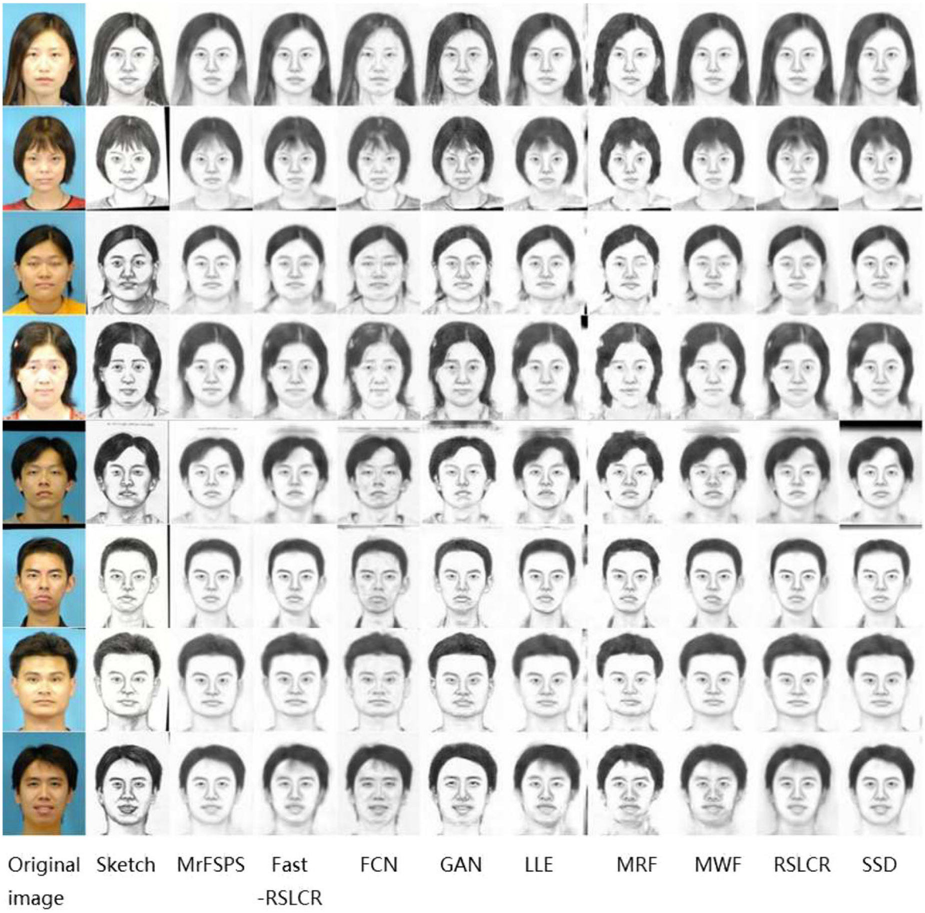


Fig. 5 The sketch examples of CUFS database (Original image, Sketch, MrFSPS, Fast-RSLCR, FCN, GAN, LLE, MRF, MWF, RSLCR, SSD)

Table 3 The results of five evaluation measures on CUFSF database

No.	Model	Scoot	SSIM	FSIM	UIQI	VIF
1	MrFSPS	0.4933	0.4192	0.6813	0.9323	0.0488
2	Fast-RSLCR	0.4608	0.4453	0.9301	0.9398	0.0501
3	FCN	0.4376	0.3619	0.6623	0.9103	0.0295
4	GAN	0.6255	0.3663	0.7060	0.9295	0.0401
5	LLE	0.5065	0.4174	0.7042	0.9369	0.0457
6	MRF	0.6163	0.3722	0.6956	0.9225	0.0356
7	MWF	0.4879	0.4296	0.7028	0.9411	0.0458
8	RSLCR	0.4529	0.4493	0.6649	0.9389	0.0507
9	SSD	0.4135	0.3433	0.6352	0.8709	0.0243

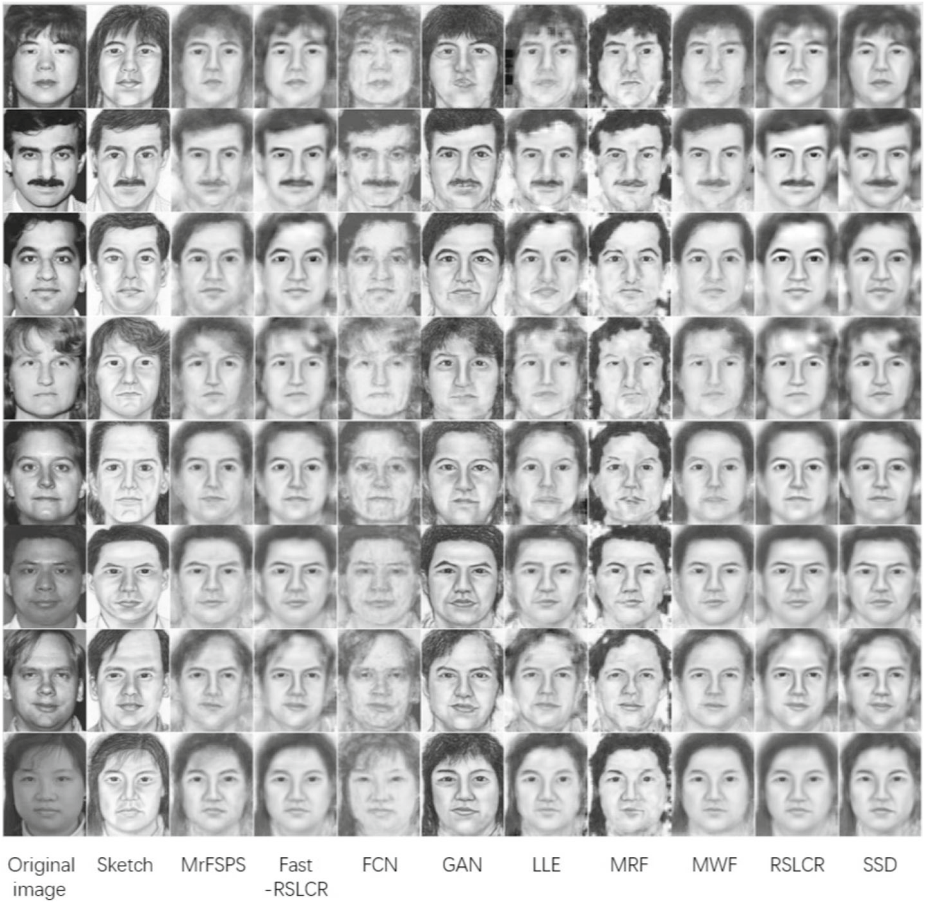


Fig. 6 The sketch examples of CUFSF database (Original image, Sketch, MrFSPS, Fast-RSLCR, FCN, GAN, LLE, MRF, MWF, RSLCR, SSD)

In Table 4, the rank of local facial features and whole sketch is different. The evaluation of testing examples indicates that an excellent model may fail to synthesize a good local feature. This is instructive for us to learn the advantages of models.

4 Conclusion

In this paper, we present the comprehensive survey of FSS models. We first investigate 50 representative FSS models and summarize the existing popular classifications. Compared to the previous work, we then propose a simple and detailed classification: coefficient models and regression models. Moreover, the typical models and popular databases are evaluated and analyzed, which is committed to describe the current situation. Besides, we put forward some personal views based on current existing problems from several new perspectives and future directions.

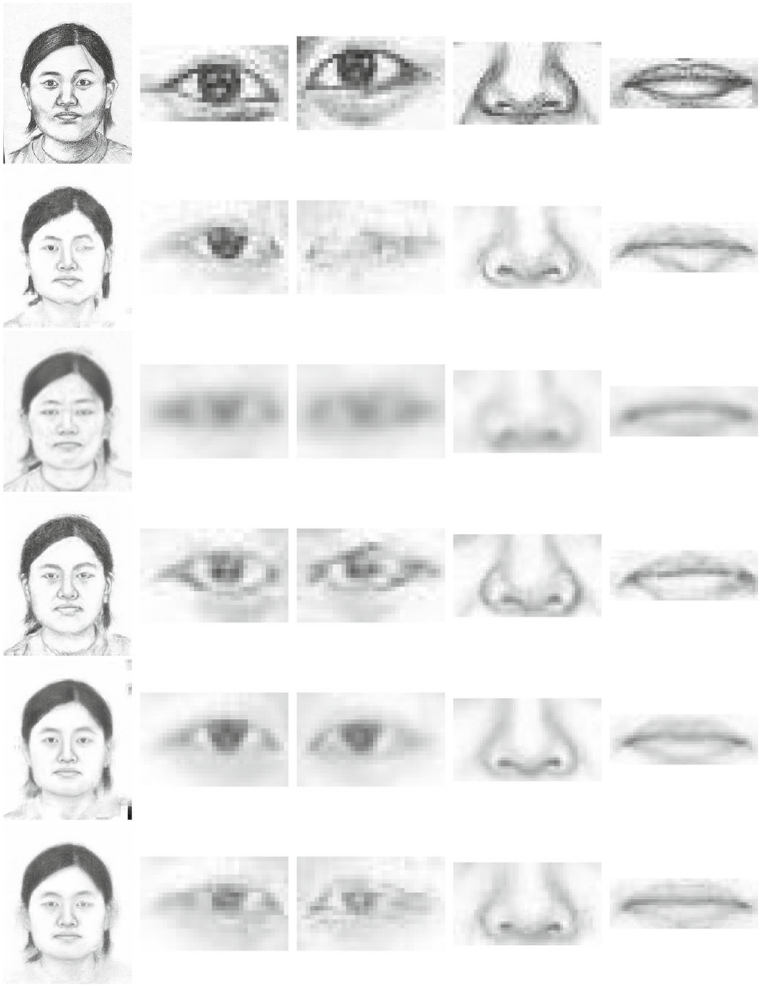


Fig. 7 Local contrast. From the first column to the sixth column are: sketch, left eye, right eye, nose, and mouse. From the first row to the sixth row are: Sketch, MRF, FCN, GAN, MWF, LLE

In summary, we review the research results in recent years and discuss the models based on deep learning. Generally, the GAN model performs better than other models, while the results are still unsatisfactory because of the limited training images. In particular, the GAN

Table 4 The results of facial features with Scoot measure

No.	Model	Sketch	Left eye	Right eye	Nose	Mouth
1	MRF	0.4024	0.1937	0.1713	0.3325	0.2395
2	FCN	0.4172	0.1680	0.1705	0.2631	0.1990
3	GAN	0.4707	0.1912	0.2053	0.3164	0.2414
4	MWF	0.4142	0.1812	0.1753	0.3038	0.2352
5	LLE	0.4251	0.1758	0.1809	0.1809	0.2447

performs worse than some traditional models in local facial features. Furthermore, some models synthesize sketches through the linear combination of training image patches rather than synthesizing new sketches, which also limits the good effect of the model. Therefore, the future work should pay more attention to the process of artists. Not only the number of training images, but also the types of training images should be enriched. Although FSS has made significant progress over the past several decades, there is still much room for improvement. We hope this survey will generate more interest in FSS.

Funding This study was funded by the NEPU Natural Science Foundation under Grant No. 2017$PYZL -$ 05, $JYCX_CX$06_2018 and $JYCX_JG$06_2018.

Declarations

Conflict of Interests Author Hongbo Bi declares that he has no conflict of interest. Author Ziqi Liu declares that she has no conflict of interest. Author Lina Yang declares that she has no conflict of interest. Author Kang Wang declares that he has no conflict of interest. Author Ning Li declares that he has no conflict of interest.

References

1. Belhumeur PN, Hespanha JP, Kriegman DJ (1997) Eigenfaces vs. fisherfaces: recognition using class specific linear projection. Tech. rep. Yale University New Haven United States
2. Bhatt HS, Bharadwaj S, Singh R et al (2012) Memetically optimized MCWLD for matching sketches with digital face images. IEEE Trans Inf Forens Secur (TIFS) 7(5):1522–1535
3. Bi H., Li N., Guan H., Lu D., Yang L. (2019) A Multi-Scale Conditional Generative Adversarial Network for Face Sketch Synthesis. In: 2019 IEEE International Conference on Image Processing (ICIP), pp 3876–3880. https://doi.org/10.1109/ICIP.2019.8803629
4. Cai W, Wei Z (2020) PiiGAN: generative adversarial networks for pluralistic image inpainting. IEEE Access 8:48451–48463
5. Chang L, Zhou M, Han Y et al (2010) Face sketch synthesis via sparse representation. In: International conference on pattern recognition (ICPR), pp 2146–2149
6. Chen LF, Liao HYM, Ko MT et al (2000) A new LDA-based face recognition system which can solve the small sample size problem. Pattern Recognit 33(10):1713–1726
7. Chen H, Xu YQ, Shum HY et al (2001) Example-based facial sketch generation with non-parametric sampling. In: IEEE international conference on computer vision (ICCV), vol 2, pp 433–438
8. Chen B-C, Chen C-S, Hsu WH (2014) Cross-age reference coding for age-invariant face recognition and retrieval. In: European conference on computer vision (ECCV). Springer, pp 768–783
9. Chen C, Tan X, Wong K-YK (2018) Face sketch synthesis with style transfer using pyramid column feature. In: IEEE Winter conference on applications of computer vision. Lake Tahoe, USA, 18
10. Chellappa R, Wilson CL, Sirohey S (1995) Human and machine recognition of faces: a survey. Proc IEEE 83(5):705–741
11. Fan D, Zhang S, Wu Y et al (2018) Face sketch synthesis style similarity: a new structure co-occurrence texture measure. CoRR arXiv:1804.02975
12. Fan D-P, Zhang S, Wu Y-H et al (2019) Scoot: a perceptual metric for facial sketches. In: IEEE international conference on computer vision (ICCV), pp 5612–5622
13. Fang Y, Deng W, Du J et al (2020) Identity-aware CycleGAN for face photo-sketch synthesis and recognition. Pattern Recognit 102
14. Gastal ESL, Oliveira MM (2011) Domain transform for edge-aware image and video processing. ACM Trans Graph 30(4):1–12
15. Gao X, Zhong J, Tao D et al (2008) Local face sketch synthesis learning. Neurocomputing 71(10–12):1921–1930
16. Gao X, Zhong J, Li J et al (2008) Face sketch synthesis algorithm based on E-HMM and selective ensemble. IEEE Trans Circuits Syst Video Technol (TCSTV) 18(4):487–496

17. Gao X, Wang N, Tao D et al (2012) Face sketch–photo synthesis and retrieval using sparse representation. IEEE Trans Circuits Syst Video Technol (TCSTV) 22(8):1213–1226
18. Gao F, Shi S, Yu J et al (2017) Composition-aided sketch-realistic portrait generation. CoRR arXiv:1712.00899
19. Isola P, Zhu J-Y, Zhou T et al (2017) Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp 1125–1134
20. Jiang J, Yu Y, Wang Z et al (2018) Residual Learning for Face Sketch Synthesis. In: 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, pp 1952–1956
21. Jiao L, Zhang S, Li L et al (2018) A modified convolutional neural network for face sketch synthesis. Pattern Recognit 76:125–136
22. Jiang J, Yu Y, Wang Z et al (2019) Graph-regularized locality-constrained joint dictionary and residual learning for face sketch synthesis. IEEE Trans Image Process (TIP) 28:628–641
23. Kim M, Kumar S, Pavlovic V et al (2008) Face tracking and recognition with visual constraints in real-world videos. In: Computer vision and pattern recognition (CVPR). IEEE, pp 1–8
24. Lei Y, Du W, Hu Q (2020) Face sketch-to-photo transformation with multi-scale self-attention GAN. Neurocomputing 396:13–23
25. Li J, Yu X, Peng C et al (2017) Adaptive representation-based face sketch-photo synthesis. Neurocomputing 269:152–159
26. Liu Q, Tang X, Jin H et al (2005) A nonlinear approach for face sketch synthesis and recognition. In: Computer vision and pattern recognition (CVPR), vol 1. IEEE, pp 1005–1010
27. Liu W, Tang X, Liu J (2007) Bayesian tensor inference for sketch-based facial photo hallucination. In: International joint conference on artificial intelligence (IJCAI), pp 2141–2146
28. Liu P, Yu H, Cang S (2019) Adaptive neural network tracking control for underactuated systems with matched and mismatched disturbances. Nonlinear Dyn 98:1447–1464
29. Lu C, Xu L, Jia J (2012) Combining sketch and tone for pencil drawing production. 65–73
30. Lu D, Chen Z, Wu QMJ et al (2019) FCN based preprocessing for exemplar-based face sketch synthesis. Neurocomputing 365:113–124
31. Martínez A, Benavente R (1998) The AR face database. CVC Technical Report 24
32. Messer K (1999) XM2VTSDB : the extended M2VTS database. In: Proceedings of the international conference on audio video based biometric person authentication, vol 964, pp 965–966
33. Moghaddam B, Pentland A (1997) Probabilistic visual learning for object representation. IEEE Trans Pattern Anal Mach Intell 19(7):696–710
34. Park U, Jain AK (2010) Face matching and retrieval using soft biometrics. IEEE Trans Inf Forens Secur (TIFS) 5(3):406–415
35. Peng C, Gao X, Wang N et al (2016) Multiple representations-based face sketch–photo synthesis. IEEE Trans Neural Netw Learn Syst (TNNLS) 27(11):2201–2215
36. Peng C, Gao X, Wang N et al (2017) Superpixel-based face sketch–photo synthesis. IEEE Trans Circuits Syst Video Technol (TCSTV) 27(2):288–299
37. Peng C, Wang N, Gao X et al (2016) Face recognition from multiple stylistic sketches: Scenarios, datasets, and evaluation. In: European conference on computer vision (ECCV). Springer, pp 3–18
38. Petpairote C, Madarasmi S (2014) Improved face recognition with expressions by warping to the best neutral face. In: International conference on emerging trends in computer and image processing (ICETCIP). pp 5–10
39. Phillips PJ, Moon H, Rizvi SA et al (2000) The FERET evaluation methodology for face-recognition algorithms. IEEE Trans Pattern Anal Mach Intell 22(10):1090–1104
40. Rothe R, Timofte R, Van Gool L (2015) Dex: deep expectation of apparent age from a single image. In: Proceedings of the IEEE international conference on computer vision workshops. pp 10–15
41. Roweis ST, Saul LK (2000) Nonlinear dimensionality reduction by locally linear embedding. Science 290(5500):2323–2326
42. Sankarasrinivasan S (2014) Compressed measurement of structural similarity index. NIT Roukela
43. Sheikh HR, Bovik AC (2006) Image information and visual quality. IEEE Trans Image Process (TIP) 15(2):430–444
44. Song Y, Bao L, Yang Q et al (2014) Real-time exemplar-based face sketch synthesis. In: European conference on computer vision (ECCV), vol 8694. Springer, pp 800–813
45. Sun L, Zhao C, Yan Z et al (2019) A novel weakly-supervised approach for rgb-d-based nuclear waste object detection. IEEE Sens J 19:3487–3500
46. Tang X, Wang X (2002) Face photo recognition using sketch. In: 2002 International conference on image processing (2002) Proceedings, 1, I–I. IEEE

47. Tang X, Wang X (2003) Face sketch synthesis and recognition. In: Ninth IEEE international conference on computer vision, 2003. Proceedings, vol 14. IEEE, pp 687–694
48. Tang X, Wang X (2004) Face sketch recognition. IEEE Trans Circuits Syst Video Technol (TCSTV) 14(1):50–57
49. Tang Z, Yu H, Lu C et al (2019) Single-trial classification of different movements on one arm based on ERD/ERS and corticomuscular coherence. IEEE Access 7:128185–128197
50. Tang Z-C, Li C, Wu J-F et al (2019) Classification of EEG-based single-trial motor imagery tasks using a B-CSP method for BCI. Front Inf Technol Electron Eng 20:1087–1098
51. Tu C-T, Chan Y-H, Chen Y-C (2016) Facial sketch synthesis using 2D direct combined model-based face-specific Markov network. IEEE Trans Image Process (TIP) 25(8):3546–3561
52. Turk M, Pentland A (1991) Eigenfaces for recognition. J Cognit Neurosci 3(1):71–86
53. Wang Z, Bovik AC (2002) A universal image quality index. IEEE Signal Process Lett 9(3):81–84
54. Wang X, Tang X (2004) Dual-space linear discriminant analysis for face recognition. In: Computer vision and pattern recognition (CVPR), vol 2, pp II–564–II–569
55. Wang X, Tang X (2008) Face photo-sketch synthesis and recognition. IEEE Trans Pattern Anal Mach Intell 31(11):1955–1967
56. Wang X, Tang X (2004) Random sampling LDA for face recognition. In: Computer vision and pattern recognition (CVPR). IEEE, II–II, p 2
57. Wang N, Gao X, Tao D et al (2011) Face sketch-photo synthesis under multi-dictionary sparse representation framework. In: International conference on image and graphics (ICIG). IEEE, pp 82–87
58. Wang S, Zhang L, Liang Y et al (2012) Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis. In: Computer vision and pattern recognition (CVPR). IEEE, pp 2216–2223
59. Wang N, Tao D, Gao X et al (2013) Transductive face sketch-photo synthesis. IEEE Trans Neural Netw Learn Syst (TNNLS) 24(9):1364–1376
60. Wang N, Tao D, Gao X et al (2014) A comprehensive survey to face hallucination. Int J Comput Vis 106(1):9–30
61. Wang N, Gao X, Li J et al (2016) Evaluation on synthesized face sketches. Neurocomputing 214:991–1000
62. Wang N, Gao X, Sun L et al (2017) Bayesian face sketch synthesis. IEEE Trans Image Process (TIP) 26(3):1264–1274
63. Wang N, Zhang S, Gao X et al (2017) Unified framework for face sketch synthesis. Signal Process 130:1–11
64. Wang N, Zhu M, Li J et al (2017) Data-driven vs. model-driven: fast face sketch synthesis. Neurocomputing 257:214–221
65. Wang L, Sindagi V, Patel VM (2018) High-quality facial photo-sketch synthesis using multi-adversarial networks. In: 13th IEEE international conference on automatic face & gesture recognition, vol 83–90. IEEE Computer Society
66. Wang N, Gao X, Li J (2018) Random sampling for fast face sketch synthesis. Pattern Recognit 76:215–227
67. Wang N, Gao X, Sun L et al (2018) Anchored neighborhood index for face sketch synthesis. IEEE Trans Circuits Syst Video Technol (TCSTV) 28:2154–2163
68. Wang N., Zha W., Li J., Gao X. (2017) Back projection: An effective postprocessing method for GAN-based face sketch synthesis. Pattern Recognition Letters 107:59–65
69. Xiao B, Gao X, Tao D et al (2010) Photo-sketch synthesis and recognition based on subspace learning. Neurocomputing 73(4–6):840–852
70. Yang Z-L, Guo X-Q, Chen Z-M et al (2019) RNN-Stega: linguistic steganography based on recurrent neural networks. IEEE Trans Inf Forens Secur (TIFS) 14:1280–1295
71. Ye L, Zhang B, Yang M et al (2019) Triple-translation GAN with multi-layer sparse representation for face image synthesis. Neurocomputing 358:294–308
72. Zhang W, Wang X, Tang X (2010) Lighting and pose robust face sketch synthesis. In: European conference on computer vision (ECCV), vol 6316. Springer, pp 420–433
73. Zhang S, Gao X, Wang N et al (2015) Face sketch synthesis via sparse representation-based greedy search. IEEE Trans Image Process (TIP) 24(8):2466–2477
74. Zhang Y, Wang N, Zhang S et al (2016) Fast face sketch synthesis via kd-tree search. In: European conference on computer vision (ECCV), vol 9913. Springer, pp 64–77
75. Zhang S, Gao X, Wang N et al (2016) Robust face sketch style synthesis. IEEE Trans Image Process (TIP) 25(1):220–232
76. Zhang S, Gao X, Wang N et al (2017) Face sketch synthesis from a single photo–sketch pair. IEEE Trans Circuits Syst Video Technol (TCSTV) 27(2):275–287

77. Zhao W, Chellappa R, Phillips PJ et al (2003) Face recognition: a literature survey. ACM Comput Surv (CSUR) 35(4):399–458
78. Zhang W, Wang X, Tang X (2011) Coupled information-theoretic encoding for face photo-sketch recognition. In: Computer vision and pattern recognition (CVPR). IEEE, pp 513–520
79. Zhang L, Zhang L, Mou X et al (2011) FSIM: a feature similarity index for image quality assessment. IEEE Trans Image Process (TIP) 20(8):2378–2386
80. Zhang L, Lin L, Wu X et al (2015) End-to-end photo-sketch generation via fully convolutional representation learning. In: International conference on multimedia retrieval (ICMR). ACM, pp 627–634
81. Zhang D, Lin L, Chen T et al (2017) Content-adaptive sketch portrait generation by decompositional representation learning. IEEE Trans Image Process (TIP) 26(1):328–339
82. Zhang S, Ji R, Hu J et al (2019) Face Sketch Synthesis by Multidomain Adversarial Learning. IEEE Trans Neural Netw Learn Syst (TNNLS) 30:1419–1428
83. Zhang M, Li J, Wang N et al (2018) Compositional model-based sketch generator in facial entertainment. IEEE Trans Cybern 48(3):904–915
84. Zhang M, Wang N, Gao X et al (2018) Markov Random Neural Fields for Face Sketch Synthesis. In: International joint conference on artificial intelligence (IJCAI), J Lang, Ed., pp 1142–1148
85. Zhang M, Li Y, Wang N et al (2020) Cascaded face sketch synthesis under various illuminations. IEEE Trans Image Process (TIP) 29:1507–1521
86. Zhang M, Wang N, Li Y et al (2020) Neural probabilistic graphical model for face sketch synthesis. IEEE Trans Neural Netw Learn Syst (TNNLS) 31:2623–2637
87. Zhou H, Kuang Z, Wong K-YK (2012) Markov weight fields for face sketch synthesis. In: Computer vision and pattern recognition (CVPR), vol 1091–1097. IEEE
88. Zhu M, Wang N, Gao X et al (2017) Deep graphical feature learning for face sketch synthesis. In: International joint conference on artificial intelligence (IJCAI), 3574–3580. AAAI Press
89. Zhu M, Li J, Wang N et al (2019) A deep collaborative framework for face photo–sketch synthesis. IEEE Trans Neural Netw Learn Syst (TNNLS) 30(10):3096–3108