ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 2, 107-111 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Deep Feature Representation for Face Sketch Recognition

Weiguo Wan[1], Hyo Jong Lee [1, 2, *]

[1]*Division of Computer Science and Engineering, Chonbuk National University, 54896, Korea*

[2]*Center for Advanced Image and Information Technology, Chonbuk National University, 54896, Korea*

ARTICLEINFO

Article history:
Received: 07 January, 2019
Accepted: 28 February, 2019
Online: 15 March, 2019

Keywords:
Face sketch recognition
Face sketch synthesis
VGG-Face network

ABSTRACT

Face sketch recognition aims at matching face sketch images to face photo images. The main challenge lies in modality discrepancy between face photo and sketch images. In this work, we propose a new facial sketch-to-photo recognition approach by adopting VGG-Face deep learning network, with which face images can be represented by compact and highly discriminative feature vectors. Different from existing VGG-Face based methods which directly match face sketches to photos, we firstly transform the gallery photos to sketches for decreasing the modality difference. Experimental results on multiple face photo-sketch datasets indicate the superiority of our method.

1. Introduction

Face sketch to photo recognition is a key branch of face recognition to address the lack of face photos, which has wide application in forensics [1]. If crime happened, and only incomplete information is obtained about the suspect because of the bad quality of monitoring videos, face sketches which drawn by the artists with the description of witnesses are used to identity the possible suspect. The police can shrink the range of suspects by searching the law enforcement face databases or surveillance cameras with the drawn face sketch images [2].

Face sketch to photo recognition approaches can be mainly classified into common space projection approaches and local feature descriptor approaches. Common space projection approaches attempt to transform facial images from various modalities to common subspace where the modality difference is reduced. After that, facial photos and sketches could be directly identified in the common space. Lin et al. [3] translated heterogeneous features to common feature space with common discriminant feature extraction (CDFE). In [4], Lei et al. proposed a subspace learning framework named coupled spectral regression (CSR) for heterogeneous face recognition. Then, they improved the CSR algorithm through learning a mapping from all modalities to all samples [5]. In [6], Sharma et al. employed partial least squares (PLS) algorithm to project heterogeneous facial images into common subspace. Mignon et al. [7] proposed to learn a discriminative latent space by using cross modal metric learning (CMML) method. In [8], Kan et al. suggested a multiple view analysis algorithm, with which the dependencies from intra-view and inter-view can be utilized to obtain a discriminant subspace for heterogeneous face recognition. However, the space projection based methods may lose the underlying information of the source images and lead to degradation of the recognition effect.

Local feature representation approaches aim at extracting robust and modality-invariant face feature representations from the heterogeneous face images. In [9], Klare et al. put forward a face sketch recognition approach with a local feature discriminant analysis (LFDA) framework which taking advantage of multiscale local binary pattern (MLBP) [10] and scale invariant feature transform (SIFT) [11]. In [12], Zhang et al. suggested a coupled information theoretic encoding based local face descriptor for face sketch to photo recognition. And later, the coupled information theoretic encoding algorithm was modified to the random forests by using various sampling methods. Galoogahi et al. [13] put forward a facial feature descriptor named local radon binary pattern (LRBP) for face sketch recognition, which firstly projects the face images into the radon space and then encodes them with the local binary patterns (LBP). In addition, a local face descriptor based on histogram of averaged oriented gradients (HAOG) is suggested by Galoogahi et al. to decrease the modality difference [14]. In [15], Lei et al. put forward a heterogeneous face

*Corresponding Author: Hyo Jong Lee, hlee@chonbuk.ac.kr

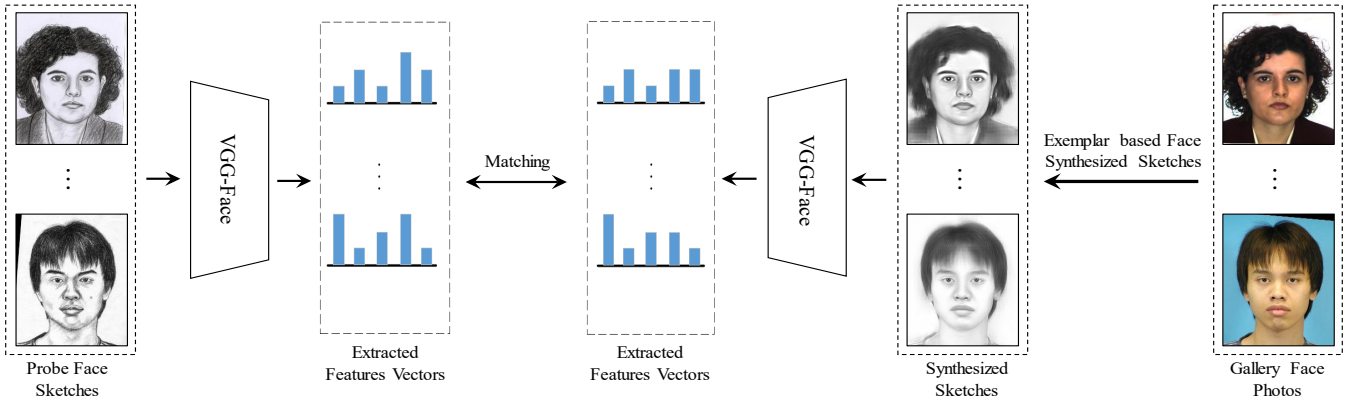

Figure 1. The illustration of the proposed face sketch recognition approach.

recognition method based on discriminant image filter learning. Alex et al. [16] suggested a local difference of Gaussian binary pattern algorithm for heterogeneous face recognition. However, when extracting the face feature representation, the local feature descriptor based approaches lose the sight of the holistic spatial structure of face, which is important for face sketch recognition.

Moreover, because of the huge differences between the face photo and sketch images, it is difficult for traditional homogeneous face recognition approaches to achieve good performance by directly matching the sketches to photos. In this paper, we transform the photo images into sketch images by adopting an exemplar-based sketch generation approach to decrease the modality difference between face photo and sketch images. In [17], Parkhi et al. presented a deep learning network, namely VGG-Face, which is able to map a face images to a compact space where distances can be used for measuring the face similarity. Peng et al. [18] mentioned that it can be adopted for face sketch recognition. Motivated by its efficient performance, we adopt it to extract the discrimination face feature representation in this paper. With the VGG-Face network, the feature vectors which represent the corresponding face images can be obtained. The illustration of the proposed approach is showed in Figure 1.

2. Related Works

2.1. Exemplar-based Sketch Generation

The exemplar-based face sketch generation approach needs a group of training dataset which contains some face photo-sketch pairs. First, each test photo **T** is divided into several image patches with overlapping between adjacent patches. Then, for each test photo patch **t**, several number of closest photo patches are chose from the training photo images which are cut image patches with the same approach. At the same time, we can get *K* corresponding sketch candidate patches with the obtained photo patches. Finally, the sketch patches are generated by combining the *K* candidate sketch patches with the corresponding weighting coefficients. The sketch generation operation can be represented as follow:

$$\min_{w} \| \mathbf{t} - \mathbf{X} \cdot \mathbf{w} \|_2^2, \ \ s.t. \ \mathbf{1}^T \mathbf{w} = 1 \tag{1}$$

where **X is** the *K* selected photo patches from training data. **w** is the weighting vector to combine candidate sketch patches.

Then we can obtain the target sketch patch ***s*** by:

$$\mathbf{s} = \mathbf{Y} \cdot \mathbf{w} \tag{2}$$

where **Y** represents the corresponding sketch patches of the test photo patch **t**.

2.2. VGG-Face Networks

The VGG-Face networks [17] are built on the basis of VGG-Very-Deep-16 convolutional neural network (CNN) architecture. The VGG-Face networks consist of a series of convolutional, pool, and fully-connected layers. The first eight blocks are convolutional layers and the remaining three blocks are fully-connected layers. For each convolutional layers, a ReLU activation layer is followed. The output dimensions of the first two fully-connected layers are 4096 and the output dimension of final fully-connected layer is 2622. The filters of convolutional layers are with size of 3×3 while the pool layers perform subsampling with a factor of two. Figure 2 displays the framework of the VGG-Face model.

In the proposed method, a pretrained VGG-Face model with 37 basic layers is utilized to extract deep face features. The pretrained model is trained on a big face dataset which has 982,803 images from 2622 subjects. The pretrained VGG-Face model only can recognize the people in the training dataset, however, we are able to extract face features from its bottleneck layers for any face images by forwarding the face images through the whole networks. The extracted features are highly discriminative, compact, and interoperable encodings of the input face images. When the face features are obtained from the bottleneck layers of the VGG-Face model, the sketch to photo matching can be performed based on these obtained face features. In this paper, the MatConvNet toolbox, which provides the pretrained implementation of the VGG-Face model, is utilized to extract deep face features.

3. The Proposed Method

In this paper, we presented a new facial sketch-to-photo recognition approach with the VGG-Face network and face sketch generation. The proposed method mainly consists of face image preprocessing, face sketch synthesis, and face sketch recognition three steps, which will be described later.

3.1. Face Image Preprocessing

Face photo and sketch images vary in properties, e.g. varying resolution, pose, and deformation. We firstly normalized all the

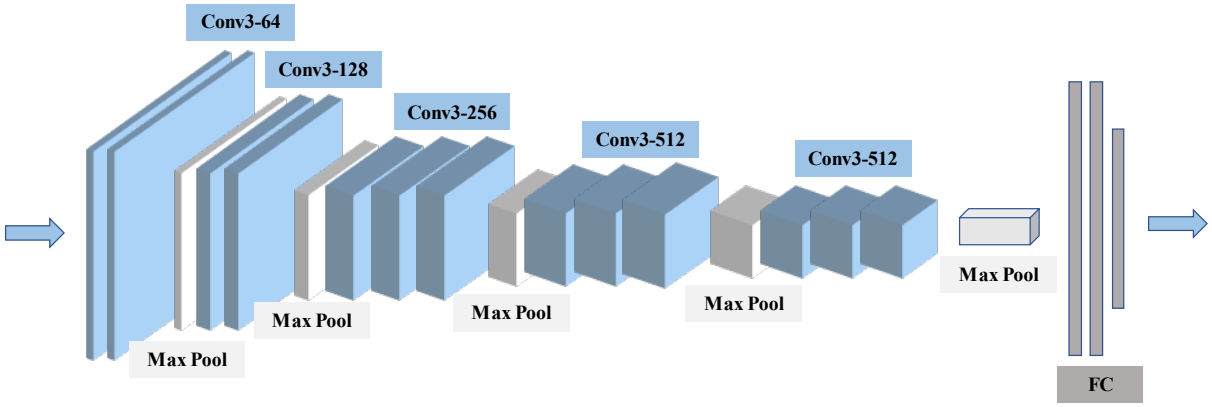

Figure 2. The VGG-Face CNN architecture.

sketch and photo images by fixing the coordinates of eyes centers. The common-used Dlib library was used to detect eyes position, and then translation, rotation, and scaling operations are conducted to align face images based on the detected eyes coordinates. Figure 3 shows some samples of original and preprocessed face photo and sketch images.

3.2. Face Photo to Sketch Synthesis

Assuming there are M geometrically aligned training photo-sketch pairs. We first divide these photo images and sketch images in training data into patches. Then, we reshape every patch to a vector. For each patch position, the searching area is extended c pixel around the patch. Thus, we can obtain $(2c+1)^2$ patches in the searching area for each patch position. We will obtain $(2c+1)^2M$ face photo and sketch patches from the training data, respectively. We employ the random sampling algorithm for selecting K face photo patches $\mathbf{U}^{(i,j)} \in \mathbf{R}^{2p^2 \times K}$ and sketch patches $\mathbf{V}^{(i,j)} \in \mathbf{R}^{2p^2 \times K}$ from training data. (i, j) is the patch position at the m-th row and the n-th column. Each image patch can be reshaped to a d-dimensional column vector.

For each test photo patch $\mathbf{t}_1^{(i,j)}$, we can compute the reconstruction weight by using equation (3):

$$\min_{\mathbf{w}^{(i,j)}} \| \mathbf{t}_1^{(i,j)} - \mathbf{U}^{(i,j)}\mathbf{w}^{(i,j)} \|_2^2 + \lambda \| \mathbf{d}^{(i,j)} \square \mathbf{w}^{(i,j)} \|, s.t. \mathbf{1}^T \mathbf{w}^{(i,j)} = 1 \quad (3)$$

where $\mathbf{d}^{(i,j)} \in \mathsf{R}^{K\times 1}$ represents the L2 distance of the test photo patch $\mathbf{t}_1^{(i,j)}$ and the selected photo patches $\mathbf{U}^{(i,j)}$ from training data, $\mathbf{w}^{(i,j)} \in \mathsf{R}^{K\times 1}$ represents the weighting coefficients of the test photo patch $\mathbf{t}_1^{(i,j)}$.

Equation (3) has the closed-form solution:

$$\begin{aligned} \mathbf{w}'^{(i,j)} &= (C^{i,j} + \lambda \mathrm{diag}(\mathbf{d}^{(i,j)})) \setminus \mathbf{1} \\ \mathbf{w}^{(i,j)} &= \mathbf{w}'^{(i,j)} / \mathbf{1}^T \mathbf{w}'^{(i,j)} \end{aligned} \quad (4)$$

where $\mathbf{1}$ represent a column vector, its elements all are 1. $C^{i,j} = (\mathbf{U}^{(i,j)} - \mathbf{1}\mathbf{t}_1^{(i,j)^T})(\mathbf{U}^{(i,j)} - \mathbf{1}\mathbf{t}_1^{(i,j)^T})^T$ is the covariance matrix, $\mathrm{diag}(\mathbf{d}^{(i,j)})$ is a diagonal matrix which is extended from $\mathbf{d}^{(i,j)}$.

By linearly combining K selected sketch patches of the training data and the weighting coefficients $\mathbf{w}^{(i,j)}$, the result sketch block

Figure 3. Facial photo-sketch examples in CUFS dataset and CUFSF dataset.

$\mathbf{s}^{(i,j)}$ could be generated:

$$\mathbf{s}^{(i,j)} = \mathbf{V}^{(i,j)}\mathbf{w}^{(i,j)} \quad (5)$$

We can generate the final target sketch with overlapping area averaged when get all the target sketch patches,. Fig. 3 shows some generated examples with the exemplar-based face sketch generation. From Figure 3, we can observe that the modality difference between the sketch image and the transformed sketch image smaller than the photo image.

3.3. Face Sketch Recognition with VGG-Face Network

After the photo images are generated to sketch images, the VGG-Face network was adopted to get the facial feature vectors of synthesized sketch images. We then call the synthesized sketch images gallery, and the test face sketch images probe.

We extract the face features by utilizing the VGG-face networks which is provided by the MatConvNet toolbox. The VGG-face architecture is composed of several convolution layers with 3×3 filters, several pooling layers with a factor of 2, and 3 fully-connected layers. In VGG deep networks, the early layers are

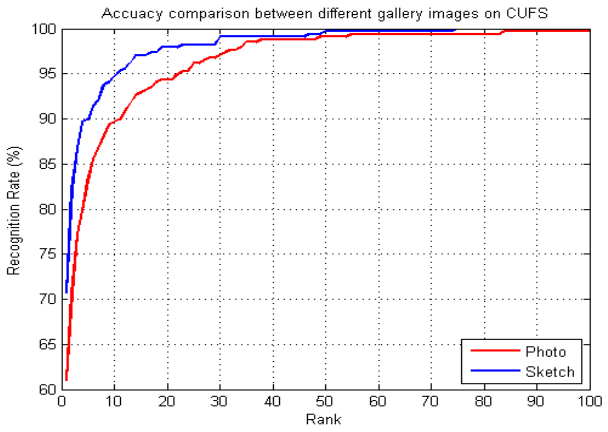

Figure 4. Performance comparison by using different gallery images.

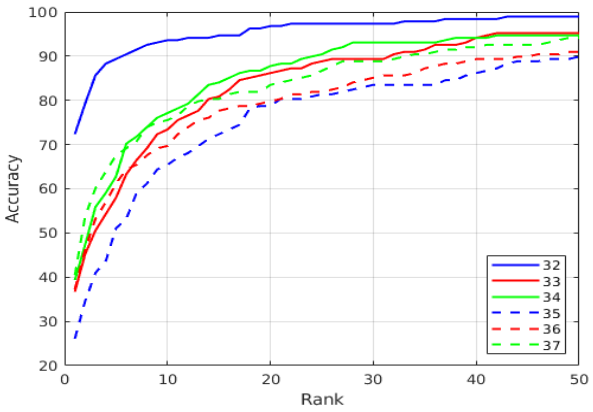

Figure 5. Performance comparison by different VGG-Face layers.

the low-level representation for input images such as edge and corner, the intermediate layers represent mid-level features like parts, and the high-level representations exist in the last layers. In order to find the optimum high-discrimination face features for face sketch-to-photo recognition, we selected activation maps from 32nd layer to 37th layer as the deep feature representation. The performance comparison results, which will be analyzed in the experiment section, indicated that the 32nd layer obtains the highest recognition accuracy. In this case, each face image was represented by a 25,088-dimension deep feature vector.

By using the well-trained VGG-Face network model, we can map all the sketch images in gallery into 25,088-dimension feature vectors. And a same length feature vector can be obtained by using the pre-trained network for each input probe image. We use the squared Euclidean distance of the face feature vectors to calculate the facial similarities. Face images from the different subjects have large distances while facial images of the same person have small distances. Hence, the face sketch recognition can be regarded as a nearest neighbor classification problem. The illustration of the proposed face sketch recognition method is shown in Figure 1.

4. Experiments and Results

4.1. Datasets

Two public available datasets are adopted to assess the recognition effect of the proposed approach: the CUHK dataset (CUFS) and the CUFSF dataset [19]. The CUFS dataset is composed of facial images from 3 datasets: the CUHK student dataset (188 subjects), the AR dataset (123 subjects), and the XM2VTS dataset (295 subjects). The CUFSF dataset has 1194 subjects of FERET dataset. In CUFS and CUFSF datasets, there are one photo image and one sketch image for each subject. All the face photo and sketch images are normalized and cut to with size of 250×200. The first two rows of Figure 3 shows the examples of photo-sketch pairs from these two datasets. The first three columns are face images of CUFS and the final column are face images of CUFSF. The first row are the original face sketches and the second row are the face photos. The final row are the generated sketches.

4.2. Experiments

First. we evaluated the recognition effect between adopting face photo images as gallery and adopting the generated sketch images as gallery. Figure 4 shown the comparison results, from it we can see that using the synthesized sketch images as gallery obtained better recognition performance. The results indicated that modality difference of the face photo and sketch images was narrowed down after synthesizing face photos to sketches.

We extracted last several layers from the VGG-face networks as the deep face features for recognition. To compare the performance by different layers, the experiment on CUHK student sub-dataset was conducted with same distance measurement. The result shown in Figure 5 indicated that the 32nd layer obtains the highest recognition accuracy. Thus, the feature vectors extracted from 32nd layer were set as the face representation in this paper.

Three existing approaches are compared to assess the recognition effect of the proposed approach, namely the HOG based approach, the DCP based approach [20] and the Light-CNN based approach [21]. Figure 6 (a) and (b) shown the comparison results on the CUFS and CUFSF datasets. Table 1 shown the rank-1, rank-5 and rank-10 recognition accuracies on CUFS and CUFSF datasets. From them, it can be seen that the proposed approach achieves the best recognition performance, compared to other methods. For the CUFS dataset, even though it is easy for other approaches to obtain high recognition performance, but our approach is the quickest one achieving 100% accuracy. The CUFSF dataset is more difficult due to the face photo images are taken under various illumination conditions. In addition, the face sketches exist shape deformation compared with the photos in CUFSF dataset. But the proposed approach still achieves the best recognition accuracy. In summary, the proposed face sketch to photo recognition approach is effective and practical.

5. Conclusions

In this paper, we presented a face sketch-to-photo recognition approach based on deep feature representation. The face photos were synthesized to face sketches to decrease the modality gap of the face photo images and sketch images. After that, the VGG-Face network was employed to extract feature representation for each face image. With the extracted face features, the face similarity can be measured directly. Experimental results on the CUFS and CUFSF datasets indicated the superiority of the proposed approach. The proposed approach in this paper is our preliminary research and only the pre-trained VGG-Face network was adopted. In the future, we plan to develop high-performance networks for face photo to sketch recognition.

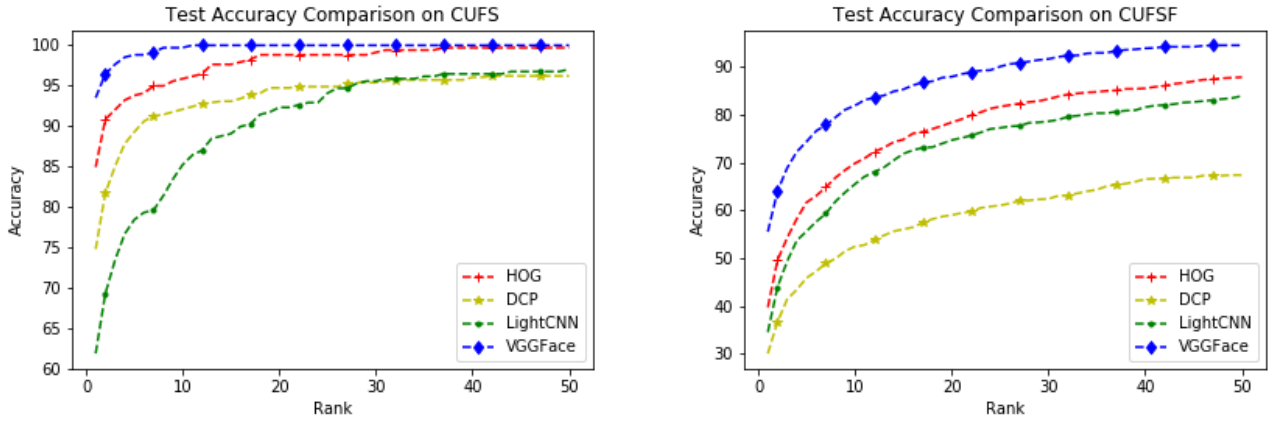

Figure 6. The comparison of recognition accuracies by different approaches on the CUFS and CUFSF datasets.

Table 1. Recognition accuracy (%) by different methods.

Methods		HOG	DCP	LightCNN	Proposed
CUFS	Rank-1	84.91	74.75	61.83	**93.49**
	Rank-5	93.79	89.27	78.40	**98.81**
	Rank-10	95.86	92.08	85.21	**99.70**
CUFSF	Rank-1	39.72	30.08	34.53	**55.51**
	Rank-5	61.65	45.87	55.61	**74.36**
	Rank-10	69.81	52.43	6.546	**81.89**

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This research was supported by "Research Base Construction Fund Support Program" funded by Chonbuk National University in 2018. This research was also supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2018-2015-0-00378) supervised by the IITP (Institute for Information & communications Technology Promotion). This research was also supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (GR 2016R1D1A3B0393 1911).

References

[1] M. Zhang, J. Li, N. Wang, and X. Gao, "Recognition of facial sketch styles," *Neurocomput.*, vol. 149, pp. 1188-1197, 2015.

[2] N. Wang, X. Gao, L. Sun, et al., "Anchored Neighborhood Index for Face Sketch Synthesis," *IEEE Trans. Circuits Syst. Video Technol*, 2017.

[3] D. Lin and X. Tang, "Inter-modality face recognition," *in Proc. 9th Eur. Conf. Comput. Vis.*, 2006, pp. 13–26.

[4] Z. Lei and S. Li, "Coupled spectral regression for matching heterogeneous faces," *IEEE Conf. Comput. Vis. Pattern Recog.*, 2009, pp. 1123–1128.

[5] Z. Lei, C. Zhou, D. Yi, A. Jain, and S. Li, "An improved coupled spectral regression for heterogeneous face recognition," *in Proc. Int. Conf. Biomed.*, 2012, pp. 7–12.

[6] A. . Sharma and D. Jacobs, "Bypass synthesis: PLS for face recognition with pose, low-resolution and sketch," *in Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2011, pp. 593–600.

[7] A. Mignon and F. Jurie, "CMML: A new metric learning approach for cross modal matching," *in Proc. Asian Conf. Comput. Vis.*, 2012.

[8] M. Kan, S. Shan, H. Zhang, S. Lao, and X. Chen, "Multi-view discriminant analysis," *in Proc. 12th Eur. Conf. Comput. Vis.*, 2012, pp. 808–821.

[9] B. Klare, Z. Li, and A. Jain, "Matching forensic sketches to mug shot photos," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 639–646, Mar. 2011.

[10] T. Ojala, M. Pietik€ainen, and T. M€aenp€a€a, "Multiresolution grayscale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002

[11] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[12] W. Zhang, X. Wang, and X. Tang, "Coupled information-theoretic encoding for face photo-sketch recognition," *in Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2011, pp. 513–520.

[13] H. Galoogahi and T. Sim, "Face sketch recognition by local radon binary pattern," *in Proc. 19th IEEE Int. Conf. Image Process.*, 2012, pp. 1837–1840.

[14] H. Galoogahi and T. Sim, "Inter-modality face sketch recognition," *in Proc. IEEE Int. Conf. Multimedia Expo*, 2012, pp. 224–229.

[15] Z. Lei, D. Yi, and S. Li, "Discriminant image filter learning for face recognition with local binary pattern like representation," *in Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2012, pp. 2512–2517.

[16] A. Alex, V. Asari, and A. Mathew, "Local difference of gaussian binary pattern: Robust features for face sketch recognition," *in Proc. IEEE Int. Conf. Syst. Man Cybern.*, 2013, pp. 1211–1216.

[17] O. M. Parkhi, V. Andrea, and Z. Andrew, "Deep face recognition," British Machine Vision Conference, vol. 1, no. 3, 2015.

[18] C. Peng, X. Gao, N. Wang N, et al., "Face recognition from multiple stylistic sketches: Scenarios, datasets, and evaluation," *Pattern Recognition*, 2018, 84: 262-272.

[19] Wang X, Tang X. Face photo-sketch synthesis and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(11): 1955-1967.

[20] Ding, C, Choi J, Tao D, Davis L S. Multi-directional multi-level dual-cross patterns for robust face recognition. *IEEE Trans Pattern Anal Mach Intell*, 2016, 38:518-531.

[21] Wu X, He R, Sun Z, Tan T. A light CNN for deep face representation with noisy labels. *IEEE Trans Inf Forensics Security*, 2018, 13:2884-2896.