IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 24, 2019, accepted July 22, 2019, date of publication July 25, 2019, date of current version August 7, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2931012

Cross-Domain Face Sketch Synthesis

MINGJIN ZHANG[1,2], JING ZHANG[1], YUAN CHI[3], YUNSONG LI[1], NANNAN WANG[1], (Member, IEEE), AND XINBO GAO[4], (Senior Member, IEEE)

[1]State Key Laboratory of Integrated Services Networks, School of Telecommunications Engineering, Xidian University, Xi'an 710071, China
[2]Key Laboratory of Spectral Imaging Technology, Chinese Academy of Sciences, Xi'an 710119, China
[3]Science and Technology on Reliability Physics and Application Technology of Electronic Component Laboratory, Guangzhou 510610, China
[4]State Key Laboratory of Integrated Services Networks, School of Electronic Engineering, Xidian University, Xi'an 710071, China

Corresponding authors: Jing Zhang (jingzhang@xidian.edu.cn) and Nannan Wang (nnwang@xidian.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grants 61876142, Grant 61671339, Grant 61772402, Grant U1605252, Grant 61501339, Grant 61432014, Grant 61571345, Grant 61801359, and Grant 61801124, in part by the National Key Research and Development Program of China under Grant 2016QY01W0200, in part by the National High-Level Talents Special Support Program of China under Grant CS31117200001, in part by the Joint Fund of Ministry of Education for Equipment Pre-research under Grant 6141A02033705, in part by the China Post-Doctoral Science Foundation under Grant 2017M623125, in part by the Opening Project of Science and Technology on Reliability Physics and Application Technology of Electronic Component Laboratory under Grant 17D03-ZHD201701, in part by the Open Research Fund of CAS Key Laboratory of Spectral Imaging Technology under Grant LSIT201901W, in part by the Natural Science Basic Research Plan in Shaanxi Province of China under Grants 2017JM6085, Grant 2017JQ6007, and Grant 2018JQ6028, in part by the Young Talent Fund of the University Association for Science and Technology in Shaanxi, China, in part by the Fundamental Research Funds for the Central Universities under Grant XJS17086, Grant XJS17109, and Grant JBX180102, in part by the CCF-Tencent Open Fund, in part by the 111 Project under Grant B08038, in part by the Xidian University-Intellifusion Joint Innovation Laboratory of Artificial Intelligence, and in part by the Light of West China of Chinese Academy of Sciences under Grant XAB2016B23.

ABSTRACT Synthesizing sketches from facial photos is of great significance to digital entertainment. Along with higher demands on sketch quality in a complex environment, however, it has been an urgent issue on how to synthesize realistic sketches with the limited training data. The existing face sketch methods pay less attention to the insufficient problem of the training data, leading to the synthesized sketches with some noise or without some identity-specific information in real-world applications. Target on providing sufficient photo-sketch pairs to meet the demand of users in digital entertainment, we present a cross-domain face sketch synthesis framework in this paper. In the photo-sketch mixed domain, we leverage the generative adversarial network to construct a cross-domain mapping function and generate identity-preserving face sketches as the hidden training data. Combined it with the insufficient original training data, we provide sufficient training data to recover the underlying structures and learn the cross-domain transfer of the high-level qualitative knowledge from the photo domain to the sketch domain by the latent low-rank representation. The qualitative and quantitative evaluations on the public facial photo-sketch database demonstrate that the proposed cross-domain face sketch synthesis method can solve the insufficient problem of the training data successfully. And it outperforms other state-of-the-art works and generates more vivid and cleaner facial sketches.

INDEX TERMS Face sketch synthesis, cross-domain, latent low-rank representation.

I. INTRODUCTION

Face sketch synthesis technique has drawn considerable interest in entertainment [1]. For instance, in many public arenas such as parks, artists draw face sketches for tourists. It takes a lot of time and money. A photo app with a face sketch synthesis technique can give its users the power to synthesize their sketches and show them on the social media network or face sketch wall freely after taking photos immediately (Fig.1). Since there exist various users, the face photos taken by them are different. For instance, some facial photos with the identity-specific information in multiple views are taken under the abnormal lightings. But the dataset for training synthesis model is limited. In most cases, it includes only facial photos and sketches in the frontal view under the normal lighting environment. With the insufficient training pairs, it is hard to recover the underlying structure or construct the mapping model by the existing face sketch

The associate editor coordinating the review of this manuscript and approving it for publication was Peter Peer.

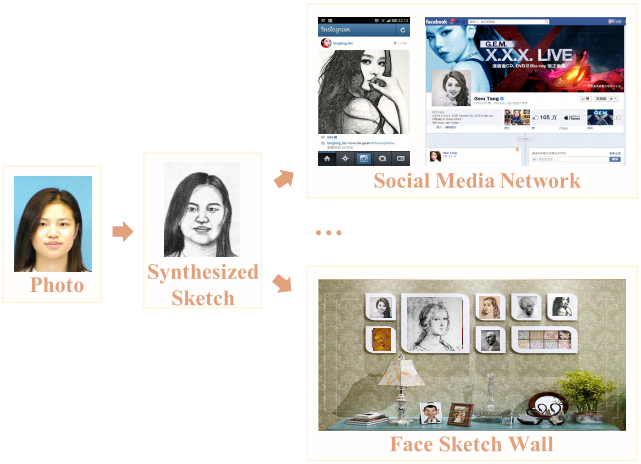


FIGURE 1. Example of face sketch synthesis application.

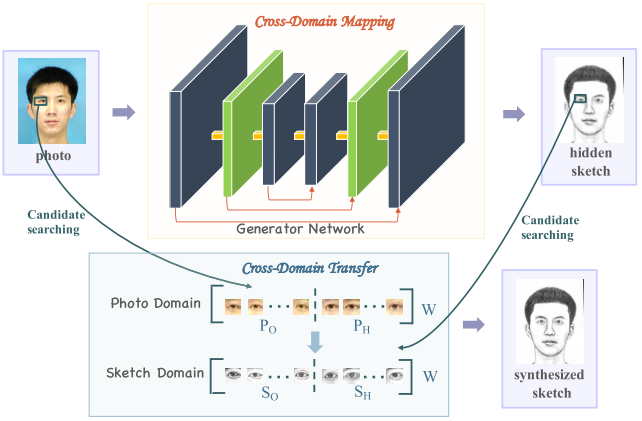


FIGURE 2. Flowchart of the proposed cross-domain synthesis method. In the photo-sketch mixed domain, we leverage the generator network to construct a cross-domain mapping function and generate hidden face sketches which preserve the identity-specific information. Then we select candidates P_O and S_H for the photo and the hidden sketch, respectively. The cross-domain transfer of the high-level qualitative knowledge is learned by LLRR.

synthesis methods. The synthesized sketches may lose some identity-specific information or appear noisy and distorted. Only if we get prepared with the sufficient training photos and sketches can we generate realistic sketches to meet the demands of users in entertainment.

The state-of-art methods pay less attention on solving the insufficient problem of the training data. Only two simple strategies are proposed in the conventional works. On the one hand, some face sketch synthesis methods leverage the manipulation of linear combination to produce a new patch [2]–[4], [6]. But some characteristics are not included in the generated patches resulting from the simple linear manipulation. Only by a nonlinear cross-domain mapping can we synthesize a satisfying new patch. On the other hand, some existing methods [5] utilize an expansion of the searching area from the local to the global for finding more similar and sufficient patches. But the lack of local constraint results in the loss of some structures, such as the bridge of the nose. In summary, the existing synthesis frameworks cannot deal with insufficient problem very well.

Target to synthesize face sketches when training data are insufficient, we present a cross-domain synthesis framework. To build sufficient training data, we learn a nonlinear cross-domain mapping relationship in the photo-sketch mixed domain by the generative adversarial networks (GAN). Then the cross-domain mapping function is transferred from the training data to the test data and the hidden sketches preserving the characteristics of the test photos are generated. Thus, the original insufficient training data and the generated hidden data are concatenated to the sufficient training data. To recover an underly structure and learn a cross-domain transfer of high-level quality knowledge from the photo domain to the sketch domain, we introduce the hidden data to a low-rank representation (LRR) to obtain a latent low-rank representation (LLRR). The proposed cross-domain approach is superior to the current face sketch synthesis methods in two aspects. 1) It is capable of synthesizing more lifelike face sketches when the training data is insufficient in reality. 2) Compared to the existing methods, it produces more satisfactory results in visual effects and objective indices. The flowchart of the proposed cross-domain face sketch synthesis is as shown in Fig.2.

II. RELATED WORK

In this section, we make a discussion on some conventional face sketch synthesis approaches. The state-of-the-arts can be divided into two classes: the shallow learning-based face sketch synthesis methods and the deep learning-based face sketch synthesis methods.

A. SHALLOW LEARNING-BASED METHODS

The shallow learning-based methods transfer a manifold from a photo domain to a sketch domain. It has three subclasses as follows.

The subspace learning-based methods recover the structures of face photos in a lower dimensional space. For instance, the principal component analysis (PCA) is utilized to generate face sketch [7]–[9]. The hair region of the synthesized sketches appears unclear. Since they make an assumption on the whole images that the sketch domain share the same coefficients with the photo domain. Liu *et al.* [2], [11] propose a local linear embedding (LLE)-based [11] method in the patch level. They divide images into patches, given that the linear combination coefficients of photo patches and sketch patches are the same. Furthermore, Song *et al.* [12] shrink the image patch to pixel and present a spatial sketch denoising (SSD)-based method. Even though it is easier to find candidates for the test photo pixel or patch than the whole image when training data are insufficient, these improvements cannot solve the insufficient problem essentially. Moreover, the sparse-based methods pay more attention on recovering the structure in the photo domain and focus less on the transformation between the photo and sketch domain.

The Bayesian inference-based methods construct a synthesized sketch based on the probabilistic graphical model solved by a heuristic algorithm. The heuristic algorithm refers to the expectation maximization algorithm or the alternating minimization algorithm. Gao and Xiao *et al.* [13]–[15] leverage the embedded hidden Markov model (E-HMM) with the Baum-Welch algorithm, Viterbi algorithm and maximum a posteriori criterion to synthesize face sketches. Consider the neighbor relationships of adjacent patches, Wang *et al.* [16] propose a face sketch synthesis method based on a Markov random field (MRF). Due to the superior of the MRF-based model, a range of MRF extended methods are spawning, such as multiple filter and feature-based method [17] and superpixel-based method [18]. Since only one candidate selected from training data is fed into the MRF-based model, this candidate should contain every possible feature. In other words, the training set should have sufficient photos and sketches for candidate selection. Zhou *et al.* [3] apply a weighted MRF method to produce the novel patches by a linear combination. Then it is extended to a Bayesian-based method by Wang *et al.* [19]. These approaches take on the image-to-image translation task as a non-convex optimization issue with the solution of the heuristic algorithm. Once the algorithm gets stuck at local minima, they may not produce enough sufficient training data and undermine the performance of the synthesized sketches.

In the sparse representation-based method, a face photo is often decomposed into a sparse coefficient matrix and a dictionary for the reconstruction of the sketch [21]. Gao and Wang *et al.* [22]–[24], [24] assume that the sparse coefficient matrices in the photo and sketch modalities are the same. And the photo and sketch domains share a similar sparse coding. Although the number of candidates can be fixed adaptively, some satisfying candidates cannot be found in the local searching area when the training data is limited in reality. To overcome these defects, Zhang *et al.* [5], [26], [26] propose another type of sparse representation-based method. It is developed for the convenience of selecting candidates in the global area. To be specific, it utilizes the sparse coefficient of patches, instead of pixel intensity, to select candidates to reduce the computation cost of candidate selection on the whole image. But without the local constraint, the synthesized sketches lose some structures.

B. DEEP LEARNING-BASED METHODS

The deep learning-based methods learn a direct mapping function between the photo modality and the sketch modality. And the knowledge is transferred from the source domain to the target domain. Gatys *et al.* [27] present an artistic style generator. Since the higher layers of the content network overlook the preserving of the detailed pixel information, the delicate sketch style cannot be preserved during transfer. Zhang *et al.* [26] utilize a generative loss to transfer the sketch style from training samples to test samples in mixed photo-sketch domain based on a fully convolutional network (FCN). Even though GAN [28] leverages the generator and discriminator to stylize images, the results still have some noises. Only convolutional layers are stacked in the neural network, resulting in blurred contour and noise. Zhang *et al.* [29] utilize the GAN to map the nonlinear relationship between the high-frequencies of the photos and sketches and propose a dual-transfer method. Some identity-specific information can be transferred from test photos to the target sketches. But some noises appear on the synthesized sketches. Zhang *et al.* [30] improve the GAN-based method by adding a probabilistic graphic model and propose a coarse-to-fine method. Although it can add the delicate details on the coarse sketches, some noises cannot be erased from the coarse sketches. Since the low-rank constraint can remove the noises and uncover the data structure, it will be demonstrated that the proposed method can synthesize the clean sketches with an improved low-rank constraint. Zhang *et al.* [31] present a Markov neural random field (MRNF) face sketch synthesis method. It induces a neural network to the probabilistic graphical model. Clearly, the deep learning-based methods can transfer the knowledge of the test photos under the limited training data, but the final results illustrate that some noises are produced during synthesis.

III. CROSS-DOMAIN FACE SKETCH SYNTHESIS

In the proposed cross-domain synthesis work, the source task is to construct the structure of faces in the photo domain, while the target task is to recover the structure in the sketch domain. But in reality, the training data is not sufficient to learn the model. In other words, the target task has different variables than the source task. The manifold cannot be directly transferred from the source task because the underlying structures of the sketch and photo are different. Thus, the hidden training data should first be generated.

We learn cross-domain transfer function and exploit the characteristics of the test faces by a convolutional neural network. Since the conditional GAN [28] including a generator and a discriminator demonstrates the promising performance for synthesizing images with the characteristics of the test samples, we select it to generate the hidden training data.

For one thing, the generator targets to synthesize face sketches similar to the sketches drawn by artists. The generator network connects 2 strided convolutional layers for downsampling, 5 residual blocks, and 2 fractionally strided convolutional layers for upsampling. For another, the discriminator attempts to discriminate the synthesized sketch and the sketch drawn by an artist. The discriminator architecture is 4 Convolution-BatchNorm-ReLU [44] layers with 64 filters, 128 filters, 256 filters and 512 filters, respectively. The negative example of the discriminator is the pair of the synthesized sketch and the photo. And the positive one is the pair of the sketch drawn by the artists and the photo. Hence, the objective function of the hidden training data generation

optimized alternatively can be written as

$$\begin{aligned}\min_G \max_D \; & V(G,D) \\ &= E_{\mathbf{p}\sim p_{data}(\mathbf{p}),\mathbf{z}\sim p_z(\mathbf{z})}[log(1-D(\mathbf{p},G(\mathbf{p},\mathbf{z})))] \\ &\quad + E_{\mathbf{p},\mathbf{s}\sim p_{data}(\mathbf{p},\mathbf{s})}[logD(\mathbf{p},\mathbf{s})] \\ &\quad + \alpha E_{\mathbf{p},\mathbf{s}\sim p_{data}(\mathbf{p},\mathbf{s}),\mathbf{z}\sim p_z(\mathbf{z})}(\|\mathbf{s}-G(\mathbf{p},\mathbf{z})\|_1) \end{aligned} \tag{1}$$

where G and D represent the generator and discriminator, respectively. $\mathbf{p}$ and $\mathbf{s}$ are the training photo and sketch. $\mathbf{z}$ is a noise term and the distribution $p_\mathbf{z}$ is over the noise $\mathbf{z}$. The distribution p_{data} is over the training face photo-sketch pairs $(\mathbf{p},\mathbf{s})$. α is a balance parameter between the adversarial loss and the L1 loss.

The hidden sketches $\hat{\mathbf{s}}$ generated by the GAN is expressed as:

$$\hat{\mathbf{s}} = G(\mathbf{p}',\mathbf{z}) \tag{2}$$

where $\mathbf{p}'$ is a test photo.

Then the hidden sketch $\hat{\mathbf{s}}$ is divided into M patches $\hat{\mathbf{S}}$ as:

$$\hat{\mathbf{S}} = [\hat{\mathbf{s}}^1,\cdots,\hat{\mathbf{s}}^M]$$

We select K candidates for each patch $\hat{\mathbf{s}}^i$ from training photos, denoted by $\mathbf{P}_H$ and their corresponding candidates from training sketches, denoted by $\mathbf{S}_H$.

$$\begin{aligned}\mathbf{S}_H &= [\mathbf{s}_H^1,\cdots,\mathbf{s}_H^K] \\ \mathbf{P}_H &= [\mathbf{p}_H^1,\cdots,\mathbf{p}_H^K]\end{aligned}$$

And we divide a test photo $\mathbf{p}'$ into M patches $\mathbf{P}'$ and select K candidates for each patch from training photos as $\mathbf{P}_O$. Their corresponding sketch candidates are $\mathbf{S}_O$.

$$\begin{aligned}\mathbf{P}' &= [\mathbf{p}'^1,\cdots,\mathbf{p}'^M] \\ \mathbf{P}_O &= [\mathbf{p}_O^1,\cdots,\mathbf{p}_O^K] \\ \mathbf{S}_O &= [\mathbf{s}_O^1,\cdots,\mathbf{s}_O^K]\end{aligned}$$

Hence, the sufficient candidates $[\mathbf{P}_O,\mathbf{P}_H]$ for test photo patches $\mathbf{P}'$ are achieved.

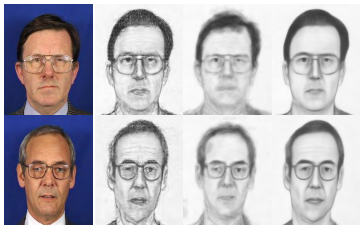
Target to uncover the underlying structure of the test photo patches $\mathbf{P}'$, we induce a low-rank constraint to guide the face sketch synthesis as:

$$\min_{\mathbf{W}} \|\mathbf{W}\|_*, \quad s.t.\mathbf{P}' = [\mathbf{P}_O,\mathbf{P}_H]\mathbf{W} \tag{3}$$

where $\mathbf{W}$ is a weight matrix. With manipulation of the skinny singular value decomposition for the test photo patch matrix $\mathbf{P}'$, the original photo patch matrix $\mathbf{P}_O$ and the hidden photo patch matrix $\mathbf{P}_H$, the above problem can be rewritten as:

$$\min_{\mathbf{W}} \|\mathbf{W}\|_*, \quad s.t.\mathbf{U}\mathbf{\Sigma}\mathbf{V}_{\mathbf{P}'}^T = [\mathbf{U}\mathbf{\Sigma}\mathbf{V}_{\mathbf{P}_O};\mathbf{U}\mathbf{\Sigma}\mathbf{V}_{\mathbf{P}_H}]^T\mathbf{W} \tag{4}$$

where $\mathbf{U}$ is a complex unitary matrix. $\mathbf{\Sigma}$ is a rectangular diagonal matrix. $\mathbf{V}_{\mathbf{P}'}$, $\mathbf{V}_{\mathbf{P}_O}$ and $\mathbf{V}_{\mathbf{P}_H}$ are the complex unitary matrix of the test photo patch matrix $\mathbf{P}'$, the original photo patch matrix $\mathbf{P}_O$ and the hidden photo patch matrix $\mathbf{P}_H$, respectively. Due to the orthogonality of the unitary matrix



(a) (b) (c) (d)

FIGURE 3. Comparison of different components in the cross-domain synthesis approach. (a) Input photos. (b) Results of the GAN. (c) Results of the LRR. (d) Results of the cross-domain synthesis approach.

$\mathbf{U}$ and the rectangular diagonal matrix $\mathbf{\Sigma}$, we can reformulate Eq. (4) as follow.

$$\min_{\mathbf{W}} \|\mathbf{W}\|_*, \quad s.t.\mathbf{V}_{\mathbf{P}'}^T = [\mathbf{V}_{\mathbf{P}_O};\mathbf{V}_{\mathbf{P}_H}]^T\mathbf{W} \tag{5}$$

According to the theoretical results of [35], [36], Eq. (5) has a closed-form solution. And its unique minimizer is:

$$\mathbf{W}^* = [\mathbf{V}_{\mathbf{P}_O};\mathbf{V}_{\mathbf{P}_H}]\mathbf{V}_{\mathbf{P}'}^T \tag{6}$$

Then we obtain the photo patches $\mathbf{P}'$ in line with Eq. (3).

$$\begin{aligned}\mathbf{P}' &= [\mathbf{P}_O,\mathbf{P}_H]\mathbf{W}^* \\ &= [\mathbf{P}_O,\mathbf{P}_H][\mathbf{V}_{\mathbf{P}_O};\mathbf{V}_{\mathbf{P}_H}]\mathbf{V}_{\mathbf{P}'}^T\end{aligned} \tag{7}$$

We can transfer the weight matrix $\mathbf{W}$ from photo to sketch modalities and generate the sketch patches $\mathbf{S}'$, resulting from the satisfying locality aware reconstruction. It is guaranteed by the low-rank constraint in the proposed cross-domain synthesis approach. The target sketch patches can be reconstructed by the source sketch patches corresponded to the test photo patches with the same weight matrix $\mathbf{W}^*$.

$$\begin{aligned}\mathbf{S}' &= [\mathbf{S}_O,\mathbf{S}_H]\mathbf{W}^* \\ &= [\mathbf{S}_O,\mathbf{S}_H][\mathbf{V}_{\mathbf{S}_O};\mathbf{V}_{\mathbf{S}_H}]\mathbf{V}_{\mathbf{S}'}^T\end{aligned} \tag{8}$$

Finally, the generated sketch patches $\mathbf{S}'$ are fused to a clean and vivid target sketch $\mathbf{s}'$ with the characteristics of the test photo.

We compare the different components of the proposed method in Fig.3. The GAN can generate the identity-preserving face sketches, resulting in sufficient training data. But some noises appear on the generated sketches. The LRR has the property of a smoothing image but it loses some characteristics. When we induce the GAN into the LLRR, the synthesized sketches are clean and delicate. It demonstrates that the proposed cross-domain synthesis method can solve the insufficient problem of the training photo-sketch pairs successfully.

As shown in Algorithm 1, the proposed cross-domain face sketch synthesis is summarized.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

We conduct the experiments on the CUHK student database, the AR database [37], and the XM2VTS database [38], as shown in Fig.4, and compare the cross-domain synthesis

Algorithm 1 Cross-Domain Sketch Synthesis From a Face Photo

Input: patches of the training sketch-photo pairs $\mathbf{P}$ and $\mathbf{S}$, a test photo $\mathbf{p}'$, K, α;

Steps:

1. Synthesize a hidden sketch $\hat{\mathbf{s}}$ from $\mathbf{p}'$ according to Eq. (2);

2. Divide $\mathbf{p}'$ and $\hat{\mathbf{s}}$ into patches $\mathbf{p}'^j$ and $\mathbf{s}'^j$, $j = 1, \cdots, M$, respectively;

3. For each $\mathbf{p}'^j$ and each $\hat{\mathbf{s}}^j$, do:

3.1. Find K photo candidates for $\mathbf{p}'^j$ from $\mathbf{P}$ denoted by $\mathbf{P}_O$. Their corresponding sketch patches selected from $\mathbf{S}$ denoted by $\mathbf{S}_O$;

3.2. Find K sketch candidates for $\hat{\mathbf{s}}^j$ from $\mathbf{S}$ denoted by $\mathbf{S}_H$. Their corresponding photo patches from $\mathbf{P}$ denoted by $\mathbf{P}_H$;

3.3. Calculate the parameters $\mathbf{W}^*$ according to Eq. (6);

3.4. Synthesize sketch patches $\mathbf{S}'$ from the observed sketch patches $\mathbf{S}_O$, the hidden sketch patches $\mathbf{S}_H$, and the parameters $\mathbf{W}^*$ according to (8);

4. Reconstruct the target sketch $\mathbf{s}'$ from $\mathbf{S}'$;

Output: a final sketch $\mathbf{s}'$.

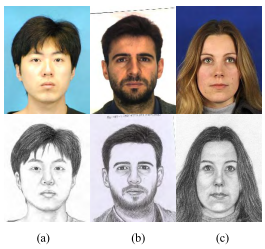


FIGURE 4. Examples of pairs in experiment. (a)-(c) The pairs from CUHK student database [16], AR database [37], and XM2VTS database [38], respectively.

approach with the current approaches including the shallow and deep learning-based approaches.

A. EXPERIMENTAL SETTINGS

The parameters of the cross-domain synthesis approach are set as: the candidates K is set to 15, the image patch is in the size of 19, the overlap size is 14, the size of search region is 5, and the balance parameter α is 100. The momentum parameter β_1, momentum parameter β_2 and learning rate l are set to 0.5, 0.999 and 0.0002. We randomly choose 88, 100 and 100 sketch-photo pairs from the CUHK student database, the AR database, and the XM2VTS database for training the GAN. And the remaining 100, 23 and 195 pairs are for test. For the discriminator, we divide generated and artist-drawn sketches into patches in the size of 70 × 70. We train the GAN on Ubuntu 14.04 system with 12G NVIDIA Titan X GPU with Torch.

B. FACE SKETCH SYNTHESIS

In Fig.5, we demonstrate visual comparisons of the cross-domain synthesis approach with the current approaches. The cross-domain synthesis approach preserves more characteristics, such as the bangs in the fourth and fifth lines, the glasses in the fourth line, and the ear in the first line. The generated sketches by MRF-based method [16] are lack in some characteristics, such as the glasses in the last line, and the ear in the first row. Since only one candidate is leveraged to synthesize a sketch. Once the candidate selected from the insufficient training data excludes the characteristics of the test photo, the synthesized sketch will lose the specific information. The linear combination-based methods can generate some new patches under the experimental environment, but they do not work in complex practical application. The textures generated by the SST-based method are not satisfying, especially in the first three lines. The GAN-based and FCN-based approaches deliver a noisy impact on faces. Although the sketches synthesized by the dual-transfer method preserve the high-frequency detail information, some noises appear on the sketches, such as mouth regions in the first three lines of the Fig.5. The coarse-to-fine method generates subtle details on the synthesized sketches, such as the bang in the fourth line of the Fig.5. But it cannot erase some noises from the coarse sketch to the fine sketch, such as the left ear in the first line. Because the probabilistic graphic model cannot remove the noises well with a simple smooth compatibility function.

C. FACE SKETCH RECOGNITION

We conduct the face recognition experiments from different aspects by the unsupervised method: Eigenface, and the unsupervised method: Fisherface [39] and null-space linear discriminant analysis (NLDA) [40]. Table 1, Table 2 and

TABLE 1. Comparision of recognition accuracy using eigenface(%).

Comparison approaches	Eigenface
LLE-based approach [2]	95.7
MRF-based approach [16]	94.0
MWF-based approach [3]	94.7
SST-based approach [5]	74.3
Bayesian-based approach [4]	95.3
FCN-based approach [27]	82.0
GAN-based approach [28]	94.0
MRNF-based approach [31]	97.0
Dual-transfer method [29]	98.3
Coarse-to-fine method [30]	96.3
Cross-domain-based approach	97.3

TABLE 2. Comparision of recognition accuracy using fisherface(%).

Comparison approaches	Fisherface
LLE-based approach [2]	87.7
MRF-based approach [16]	89.3
MWF-based approach [3]	89.7
SST-based approach [5]	75.0
Bayesian-based approach [4]	91.7
FCN-based approach [27]	85.0
GAN-based approach [28]	89.3
MRNF-based approach [31]	98.7
Dual-transfer method [29]	99.7
Coarse-to-fine method [30]	95.3
Cross-domain-based approach	99.0

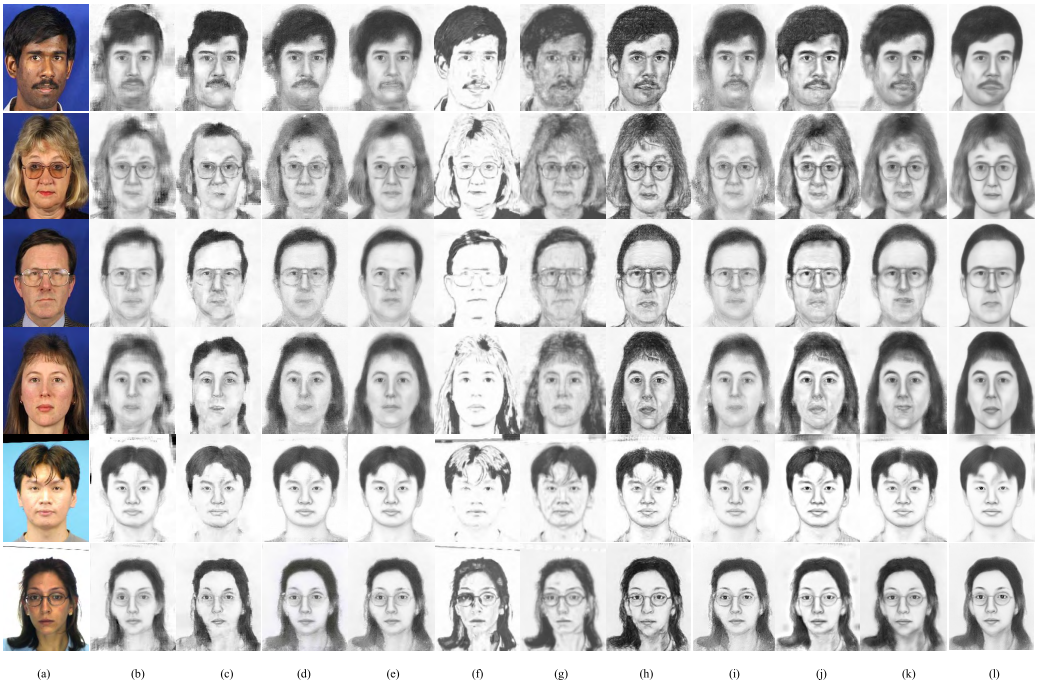


FIGURE 5. Comparison between the cross-domain synthesis approach and the current approaches. (a) Input photos. (b)-(f) Results of the shallow-learning based approaches including the LLE-based [2], MRF-based [16], MWF-based [3], Bayesian-based [4], and SST-based [5] approaches. (g)-(i) Results of the deep-learning based approaches including the GAN-based [28], FCN-based [27], MRNF-based [31] dual-transfer [29] and coarse-to-fine [30] approaches. (j) Results of the cross-domain approach.

TABLE 3. Comparision of recognition accuracy using NLDA(%).

Comparison approaches	Fisherface
LLE-based approach [2]	91.0
MRF-based approach [16]	87.7
MWF-based approach [3]	92.7
SST-based approach [5]	78.0
Bayesian-based approach [4]	97.3
FCN-based approach [27]	100.0
GAN-based approach [28]	99.0
MRNF-based approach [31]	100.0
Dual-transfer method [29]	100.0
Coarse-to-fine method [30]	99.0
Cross-domain-based approach	100.0

Table 3 are concerned with the recognition accuracy rates of the different synthesis approaches by Eigenface, Fisherface, and NLDA, respectively. The recognition accuracy rate using NLDA by the proposed method achieves 100%. The cross-domain synthesis method is the second-best performing approach for recognizing the identity of the synthesized sketches by Eigenface and Fisherface. The dual-transfer method performs better than the proposed method. It lies in the fact that it leverages a GAN to learn the high-frequency information of the target sketch. And the high-frequency information includes most of the identity-specific information, leading to satisfactory recognition performance. The cross-domain synthesis method does not specially design a neural network for producing the identifiable facial information. We can improve the cross-domain synthesis method by drawing on the thought of the dual-transfer method with the high-frequency information. The SST-based approach tries to provide a bigger searching region for candidates, but it ignores the local constraint and loses some components. Thus, its accuracy rates are the lowest in three face recognition indices. The accuracy rates of the linear combination-based methods are lower than those of the cross-domain synthesis approach. Since the proposed nonlinear combination-based method can produce more sufficient candidates than the linear ones. The MRNF-based method reaches the high recognition rate using NLDA. But it gets slight smaller recognition rates using Eigenface and Fisherface for its lack of some hair information.

D. IMAGE QUALITY ASSESSMENT

By following the existing face sketch synthesis approaches for comparison, we utilize the visual information fidelity index (VIF) [42] and the structural similarity index (SSIM) [43] to evaluate the cross-domain synthesis approach quantitatively [39], as shown in Table 4 and 5. The proposed cross-domain approach outperforms the other approaches in the SSIM average value (0.4719). Since the probabilistic graphic model is worse than the latent low rank in removing the noises. Both SSIM and VIF average values of the sketches synthesized by the coarse-to-fine method are lower than those by the proposed method. The dual-transfer method adds the high-frequency information to the final results, resulting in the higher VIF average value. But the noises of its synthesized sketches hamper the SSIM average value (0.4587) of the

TABLE 4. SSIM Values of different approaches.

Comparison approaches	SSIM
LLE-based approach [2]	0.4619
MRF-based approach [16]	0.4282
MWF-based approach [3]	0.4605
SST-based approach [5]	0.4006
Bayesian-based approach [4]	0.4622
FCN-based approach [27]	0.4254
GAN-based approach [28]	0.4118
MRNF-based approach [31]	0.4674
Dual-transfer method [29]	0.4587
Coarse-to-fine method [30]	0.4718
Cross-domain-based approach	0.4719

TABLE 5. VIF Values of different approaches.

Comparison approaches	VIF
LLE-based approach [2]	0.0783
MRF-based approach [16]	0.0693
MWF-based approach [3]	0.0786
SST-based approach [5]	0.0662
Bayesian-based approach [4]	0.0790
FCN-based approach [27]	0.0707
GAN-based approach [28]	0.0736
MRNF-based approach [31]	0.0801
Dual-transfer method [29]	0.0860
Coarse-to-fine method [30]	0.0837
Cross-domain-based approach	0.0838

dual-transfer method. The SST-based approach performs the poorest either on the VIF average value or on the SSIM average value. Since they lost some textures, leading to bad visual effect, and lost some facial structures, resulting in poor structural information. The GAN-based [28] and FCN-based [27] approaches produce some noises in the final results, resulting from the lower VIF average value.

E. COMPUTATIONAL COMPLEXITY

We make a comparison in the computation cost of the cross-domain synthesis method with two representative methods. They are the FCN-based method with the fastest speed and the dual-transfer method with the best performance in the existing face sketch synthesis methods. The FCN-based method is in an end-to-end structure. The average time consumption of it to synthesize one sketch is 0.04s. It is faster than the proposed method, but the synthesis performance is worse than the proposed method. The most time-consuming part of the proposed method is the process of selecting candidates. It is $O(cp^2\ MN)$. c and N represent the number of patches in the searching region and training data, respectively. p^2 denotes the area of the patch. M is the number of patches in the test photo. The time complexity of the dual-transfer method is $O(cp^2\ MN + p^2\ Kmn)$, as listed in [29], where K is the number of candidates. The size of the test photo is $m \times n$. It is slower than the proposed method. In the above, the cross-domain synthesis method achieves the best performance of the visual effects and the objective indices in a mediate manner.

FIGURE 6. Comparison of the cross-domain synthesis approach with the representative approaches on celebrity photos. (a) Input photos. (b)-(c) Results of the shallow learning-based approaches including the MRF-based [16] and Bayesian-based [4] approaches. (d)-(e) Results of the deep learning-based approaches including the GAN-based [28], MRNF-based [31] dual-transfer [29] and coarse-to-fine [30] approaches. (f) Results of the cross-domain synthesis approach.

V. DIGITAL ENTERTAINMENT APPLICATION

In reality, many users take the photo with the different head poses under the varied lighting conditions, while the data for training are the face photos taken in the frontal view under the normal lighting. The complex of the face photos of users is greater than the data for training, resulting in the insufficient training data issue during synthesizing face sketches.

We conduct the experiment on the celebrity face photos downloaded from the internet, and compare the proposed cross-domain approach with the representative approaches. Due to the lack of face sketches on the extended CUHK face sketch database, the model is trained on the CUHK student face sketch database. A robust comparison of the sketches generated by the representative approaches is shown in Fig.6. The cross-domain synthesis approach is the top performer. Our synthesized sketches are more vivid than those generated by the shallow learning-based and deep learning-based approaches. Compared with the MRF-based approach [16] and the Bayesian-based approach [4], the proposed approach does not produce black regions in low-light situations and generates the complete facial components. They are cleaner than those of the GAN-based approach [28] and the MRNF-based approach [31]. The sketches synthesized by the coarse-to-fine synthesis method have extra drop shadows including the first and third lines in Fig.6. In real application, the insufficient problem of training data is recurring often. For instance, the training photos are taken under normal lighting, respectively, while the test photo is taken under the dark side illumination. Some noises appear on the sketches synthesized by the dual-transfer method, such as the right faces in the second line. It demonstrates the dual-transfer method cannot deal with the insufficient issue as well as the proposed method.

VI. CONCLUSION

We propose a cross-domain synthesis approach for synthesizing sketches from photos. We first leverage a GAN in the photo-sketch mixed domain to learn a cross-domain mapping relationship from the insufficient training data. With the cross-domain mapping relationship, the hidden training data is generated as a complement to the sufficient training data. We then utilize an LLRR to recover the underlying structure and transfer the high-level quality knowledge from the photo domain to the sketch domain. Experiment results on the CUHK student face sketch database, AR database and XM2TVS database illustrate the superiority and flexibility of the cross-domain synthesis approach. In the future, we would explore deep multi-task learning [45], [46] the performance of more advanced IQA metrics for synthesis evaluation [47]–[49].

REFERENCES

[1] A. Akram, N. Wang, J. Li, and X. Gao, "A comparative study on face sketch synthesis," *IEEE Access*, vol. 6, pp. 37084–37093, 2018.

[2] Q. Liu, X. Tang, H. Jin, H. Lu, and S. Ma, "A nonlinear approach for face sketch synthesis and recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 1005–1010.

[3] H. Zhou, Z. Kuang, and K.-Y. K. Wong, "Markov weight fields for face sketch synthesis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 1091–1097.

[4] N. Wang, X. Gao, L. Sun, and J. Li, "Bayesian face sketch synthesis," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1264–1274, Mar. 2017.

[5] S. Zhang, X. Gao, N. Wang, J. Li, and M. Zhang, "Face sketch synthesis via sparse representation-based greedy search," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2466–2477, Aug. 2015.

[6] C. Hao, Y. Chen, and E. Wu, "Efficient patchmatch-based synthesis for cartoon animation," *IEEE Access*, vol. 7, pp. 31262–31272, 2019.

[7] X. Tang and X. Wang, "Face sketch recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 50–57, Jan. 2004.

[8] X. Tang and X. Wang, "Face photo recognition using sketch," in *Proc. Int. Conf. Image Process.*, Sep. 2002, pp. 257–260.

[9] J. H. Davis and J. R. Cogdell, "Calibration program for the 16-foot antenna," Elect. Eng. Res. Lab., Univ. Texas, Austin, TX, USA, Tech. Rep. NGL-006-69-3, Nov. 1987.

[10] W. Liu, X. Tang, and J. Liu, "Bayesian tensor inference for sketch-based face photo hallucination," in *Proc. IEEE Conf. Artif. Intell.*, Jan. 2007, pp. 2141–2146.

[11] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[12] Y. Song, L. Bao, Q. Yang, and M. H. Yang, "Real-time exemplar-based face sketch synthesis," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 800–813.

[13] X. Gao, J. Zhong, J. Li, and C. Tian, "Face sketch synthesis using E-HMM and selective ensemble," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 4, pp. 487–496, Apr. 2008.

[14] X. Gao, J. Zhong, D. Tao, and X. Li, "Local face sketch synthesis learning," *Neurocomputing*, vol. 71, nos. 10–12, pp. 1921–1930, Jun. 2008.

[15] B. Xiao, X. Gao, D. Tao, Y. Yuan, and J. Li, "Photo-sketch synthesis and recognition based on subspace learning," *Neurocomputing*, vol. 73, nos. 4–6, pp. 840–852, Jan. 2010.

[16] X. Wang and X. Tang, "Face photo-sketch synthesis and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 11, pp. 1955–1967, Nov. 2009.

[17] C. Peng, X. Gao, N. Wang, D. Tao, X. Li, and J. Li, "Multiple representations-based face sketch–photo synthesis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 11, pp. 2201–2215, Nov. 2016.

[18] C. Peng, X. Gao, N. Wang, and J. Li, "Superpixel-based face sketch–photo synthesis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 2, pp. 288–299, Feb. 2015.

[19] N. Wang, D. Tao, X. Gao, X. Li, and J. Li, "Transductive face sketch–photo synthesis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 9, pp. 1364–1376, Sep. 2013.

[20] R. He, W. S. Zheng, T. Tan, and Z. Sun, "Half-quadratic based iterative minimization for robust sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 2, pp. 261–275, Feb. 2014.

[21] M. Chang, L. Zhou, Y. Han, and X. Deng, "Face sketch synthesis via sparse representation," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 2146–2149.

[22] N. Wang, X. Gao, D. Tao, and X. Li, "Face sketch–photo synthesis under multi-dictionary sparse representation framework," in *Proc. 6th Int. Conf. Image Graph.*, Aug. 2011, pp. 82–87.

[23] X. Gao, N. Wang, D. Tao, and X. Li, "Face sketch-photo synthesis and retrieval using sparse representation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 8, pp. 1213–1226, Aug. 2012.

[24] S. Wang, L. Zhang, Y. Liang, and Q. Pan, "Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2216–2223.

[25] S. Zhang, X. Gao, N. Wang, and X. Li, "Face sketch synthesis from a single photo–sketch pair," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 2, pp. 275–287, Feb. 2015.

[26] S. Zhang, X. Gao, N. Wang, and J. Li, "Robust face sketch style synthesis," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 220–232, Jan. 2016.

[27] L. A. Gatys, A. S. Ecker, and M. Bethge, "A neural algorithm of artistic style," 2017, *arXiv:1508.06576*, [Online]. Available: https://arxiv.org/abs/1508.06576

[28] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," 2017, *arXiv:1611.07004*. [Online]. Available: https://arxiv.org/abs/1611.07004

[29] M. Zhang, R. Wang, X. Gao, J. Li, and D. Tao, "Dual-transfer face sketch–photo synthesis," *IEEE Trans. Image Process.*, vol. 28, no. 2, pp. 642–657, Feb. 2019.

[30] M. Zhang, N. Wang, Y. Li, R. Wang, and X. Gao, "Face sketch synthesis from coarse to fine," in *Proc. AAAI*, Apr. 2018, pp. 7558–7565.

[31] M. Zhang, N. Wang, X. Gao, and Y. Li, "Markov random neural fields for face sketch synthesis," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, 2018, pp. 1142–1148.

[32] V. Larsson and C. Olsson, "Convex low rank approximation," *Int. J. Comput. Vis.*, vol. 120, no. 2, pp. 194–214, 2016.

[33] M. Shao, D. Kit, and Y. Fu, "Generalized transfer subspace learning through low-rank constraint," *Int. J. Comput. Vis.*, vol. 109, no. 1, pp. 374–393, Aug. 2014.

[34] M. Elad and P. Milanfar, "Style transfer via texture synthesis," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2338–2351, May 2017.

[35] C. Liu and S. Yan, "Latent low-rank representation for subspace segmentation and feature extraction," in *Proc. Eur. Conf. Comput. Vis.*, Nov. 2011, pp. 1615–1622.

[36] R. He, T. Tan, and L. Wang, "Robust recovery of corrupted low-rankmatrix by implicit regularizers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 4, pp. 770–783, Apr. 2014.

[37] A. Martinez and R. Benavente, "The AR face database," CVC, New Delhi, India, Tech. Rep. #24, 1998, vol. 24, no. 1, pp. 1–10.

[38] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, "XM2VTSDB: The extended M2VTS database," in *Proc. 2nd Int. Conf. Audio Video-Based Biometric Person Authentication*, Mar. 1999, pp. 72–77.

[39] X. Wu, L. Song, R. He, and T. Tan, "Coupled deep learning for heterogeneous face recognition," in *Proc. 32nd AAAI Conf. Artif. Intell.*, Apr. 2018, pp. 627–634.

[40] L. Chen, H. Liao, M. Ko, J. Lin, and G. Yu, "A new LDA-based face recognition system which can solve the small sample size problem," *Pattern Recognit.*, vol. 33, no. 10, pp. 1713–1726, 2000.

[41] F. Gao, D. Tao, X. Gao, and X. Li, "Learning to rank for blind image quality assessment," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 10, pp. 2275–2290, Oct. 2015.

[42] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

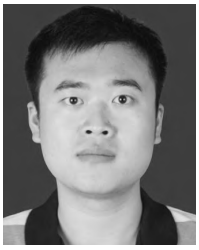
[43] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[44] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: https://arxiv.org/abs/1502.03167

[45] J. Yu, Z. Kuang, B. Zhang, D. Lin, and J. Fan, "iPrivacy: Image privacy protection by identifying sensitive objects via deep multi-task learning," *IEEE Trans. Inf. Forensics Secur.*, vol. 12, no. 5, pp. 1005–1016, May 2017.
[46] J. Yu, C. Hong, Y. Rui, and D. Tao, "Multitask autoencoder model for recovering human poses," *IEEE Trans. Ind. Electron.*, vol. 65, no. 6, pp. 5060–5068, Jun. 2018.
[47] F. Gao and J. Yu, "Biologically inspired image quality assessment," *Signal Process.*, vol. 124, pp. 210–219, Jul. 2016.
[48] F. Gao, Y. Wang, P. Li, M. Tan, J. Yu, and Y. Zhu, "DeepSim: Deep similarity for image quality assessment," *Neurocomputing*, vol. 257, pp. 104–114, Sep. 2017.
[49] F. Gao, J. Yu, S. Zhu, Q. Huang, and Q. Tian, "Blind image quality prediction by exploiting multi-level deep representations," *Pattern Recognit.*, vol. 81, pp. 432–442, Sep. 2018.

MINGJIN ZHANG received the B.Sc. degree in electronic and information engineering from Xidian University, Xi'an, China, in 2011, and the Ph.D. degree in circuits and systems, in 2017. From October 2015 to October 2016, she was a Visiting Ph.D. Student with the University of Technology Sydney, NSW, Australia. She is currently with the State Key Laboratory of Integrated Services Networks, Xidian University. She has published more than ten papers in refereed journals and proceedings, including IEEE TC, IEEE TNNLS, IEEE TIP, AAAI, IJCAI, and so on. Her current research interests include computer vision, pattern recognition, and machine learning.

JING ZHANG received the B.Sc. degree in information engineering from Xi'an Jiaotong University, Xi'an, China, in 2003, and the Ph.D. degree in information and communication engineering, in 2009. From September 2007 to September 2008, she was a Visiting Ph.D. Student with Mississippi State University, USA. She is currently with the State Key Laboratory of Integrated Services Networks, Xidian University. Her current research interests include image processing and high-performance computing.



YUAN CHI received the Ph.D. degree in circuit and systems from Xidian University, Xi'an, China, in 2015. He is currently an Engineer with the Science and Technology on Reliability Physics and Application of Electronic Component Laboratory, the Fifth Electronics Research Institute of Ministry of Industry and Information Technology, Guangzhou, China. His research interests include integrated circuits design, chip reliability, and chip security.

YUNSONG LI received the M.S. degree in telecommunication and information systems and the Ph.D. degree in signal and information processing from Xidian University, China, in 1999 and 2002, respectively. He joined the School of Telecommunications Engineering, Xidian University, in 1999, where he is currently a Professor. He is the Director of the Image Coding and Processing Center, State Key Laboratory of Integrated Service Networks. His research interests include image and video processing, hyperspectral image processing, and high-performance computing.

NANNAN WANG (M'16) 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, 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 with the State Key Laboratory of Integrated Services Networks, Xidian University. He has published more than 50 papers in refereed journals and proceedings, including IEEE T-PAMI, IJCV, AAAI, IJCAI, and so on. His current research interests include computer vision, pattern recognition, and machine learning.

XINBO GAO (M'02–SM'07) received the B.Eng., M.Sc., and Ph.D. degrees in 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 with the Department of Computer Science, Shizuoka University, Shizuoka, Japan. From 2000 to 2001, he was a Postdoctoral Research Fellow with the Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong. Since 2001, he has been with the School of Electronic Engineering, Xidian University. He is currently a Cheung Kong Professor of the Ministry of Education, a Professor of pattern recognition and intelligent system, and the Director of the State Key Laboratory of Integrated Services Networks, Xi'an, China. He has published six books and around 200 technical articles in refereed journals and proceedings. His current research interests include multimedia analysis, computer vision, pattern recognition, machine learning, and wireless communications. He is a fellow of the Institute of Engineering and Technology and a fellow of the Chinese Institute of Electronics. He is on the Editorial Boards of several journals, including *Signal Processing* (Elsevier) and *Neurocomputing* (Elsevier). He served as the General Chair/Co-Chair, the Program Committee Chair/Co-Chair, or a PC Member for around 30 major international conferences.