Markov Random Neural Fields for Face Sketch Synthesis

Mingjin Zhang[†], **Nannan Wang**[†*], **Xinbo Gao**[‡], **Yunsong Li**[†]
[†] State Key Laboratory of Integrated Services Networks,
School of Telecommunications Engineering, Xidian University, Xi'an 710071, China
[‡] State Key Laboratory of Integrated Services Networks,
School of Electronic Engineering, Xidian University, Xi'an 710071, China
{mjinzhang, nnwang}@xidian.edu.cn, {xbgao, ysli}@mail.xidian.edu.cn

Abstract

Synthesizing face sketches with both common and specific information from photos has been recently attracting considerable attentions in digital entertainment. However, the existing approaches either make the strict similarity assumption on face sketches and photos, leading to lose some identity-specific information, or learn the direct mapping relationship from face photos to sketches by the simple neural network, resulting in the lack of some common information. In this paper, we propose a novel face sketch synthesis based on the Markov random neural fields including two structures. In the first structure, we utilize the neural network to learn the non-linear photo-sketch relationship and obtain the identity-specific information of the test photo, such as glasses, hairpins and hairstyles. In the second structure, we choose the nearest neighbors of the test photo patch and the sketch pixel synthesized in the first structure from the training data which ensure the common information of Miss or Mr Average. Experimental results on the Chinese University of Hong Kong face sketch database illustrate that our proposed framework can preserve the common structure and capture the characteristic features. Compared with the state-of-the-art methods, our method achieves better results in terms of both quantitative and qualitative experimental evaluations.

1 Introduction

Face sketch synthesis from photo facilitates the digital entertainment. According to a recent report, the use of "selfie"is increased dramatically – nearly 170-fold in the past 12 mouths. Every second, an Internet user searches a selfie on social media sites, such as Facebook and Instagram. Every three seconds, a user sends its selfie to the blogs. And a considerable portion of these selfies are dealt by the face sketch synthesis technique. With the help of the face sketch synthesis technique, users try to stand out on social media sites or blogs by the vivid synthesized sketches with identity-specific structure. At the same time, they do not want to build a bizarre and motley characters among their friends and family. Thus, the synthesized sketches should reserve the common structure of Miss or Mr Average which refers to the average face of the public. Therefore, in digital entertainment, it is expected to synthesize the face sketches not only with the identity-specific information but with the common information as well.

However, the state-of-the-art methods cannot synthesize delicate face sketches with both the specific and common structures. The shallow learning-based face sketch synthesis methods make an assumption that face sketches and photos share a similar structure in low-dimensional manifolds. Once the photo manifold loses the identity-specific information included in the test photo but excluded in the training data, the sketch manifold will lack the specific structure. On the contrary, the deep learning-based face sketch synthesis methods avoid utilizing the similar assumption and build a direct mapping relationship between photos and sketches. The identity-specific information can be preserved well but some common structures appear unpleasing. Based on the above discussions, we believe that the shallow learning-based face sketch synthesis methods should induce a deep structure, such as the neural network, to ensure that there exists the common information accompanied with the identity-specific information in the synthesized sketches.

For the purpose of achieving this goal, we present a novel face sketch synthesis based on the Markov random neural fields (MRNF). The proposed MRNF-based method has two structures. In the first structure, we focus on how to learn the identity-specific information of the test photo and build a structured regressor between the photo patches and sketch pixels. We formulate the structured regressor in a multivariate Gaussian form and adopt a gradient-based method to solve it. After putting the test photo patches into the learned regressor, we obtain the generated sketch pixels with the identity-specific information. In the second structure, we pay more attentions on how to learn the common information of Miss or Mr Average. This structure takes the fidelity between the test face photo patch and their candidates, the fidelity between the generated face sketch pixel and their candidates, and the com-

[*] Corresponding author: Nannan Wang (nnwang@xidian.edu.cn)

patibility between neighboring sketch patches into account. It can be reformulated as a standard QP problem optimized by a cascade decomposition method. To verify the effectiveness of the proposed method, we conduct the experiments on the public face photo-sketch database and compare with the conventional methods, showing that our MRNF-based face sketch synthesis method achieves a superior performance.

The major contributions of this work are twofold: 1) it induces the neural network to the Markov random fields; 2) The synthesized sketches not only contain the identity-specific information, but also include the common information.

The remainder of the paper is organized as follows. First we present a brief outline of the state-of-art works on face sketch synthesis in Section 2. Section 3 details the proposed face sketch synthesis based on MRNF. This is followed by the experimental results and comprehensive analyses in Section 4. Section 5 gives some concluding remarks.

2 Related Work

In line with the number of the layers used in the model, we classify the existing face sketch synthesis methods into two main categories: shallow learning-based and deep leaning-base methods.

2.1 Shallow Learning-based Methods

There exists only one or no layer in shallow learning-based face sketch synthesis methods. The shallow learning model learns the relationship between the test photo and their candidates selected from the training photos. Due to the similarity assumption, this relationship can be transferred to the target sketch and the corresponding training sketches. The shallow learning-based method can be divided into three groups: subspace learning-based, sparse representation-based, and Bayesian inference-based methods [Wang *et al.*, 2013].

The subspace learning-based methods have no layer and make the assumption the face sketches share the similar projection coefficients with the face photos. Popular approaches include principal component analysis (PCA)-based, local linear embedding (LLE)-based and spatial sketch denoising (SSD)-based methods. They shrink the target object from the whole image level to the patch level, even to the pixel level. Specifically, Tang and Wang [2004] [2002] [2003] present the PCA-based methods and assume that the whole sketches and photos use the same topological structure via PCA. Actually, the mapping relationship between the whole images in the sketch manifold and photo manifold is not linear. Liu *et al.* [2005] [2007] induce the idea of local linear embedding and divide the whole images into local patches, assuming that the sketch and photo patches share the common projection coefficients. Furthermore, Song *et al.* [2014] replace the pixels of the patches and reduce some noise. Although these improvements may help relieve the stress on the strict similarity assumption, they cannot prevent the identity-specific information loss.

The sparse representation-based methods also have no layer and assume that the face sketch patches and photo patches utilize the same sparse representation coefficients [Chang *et al.*, 2010] [Wang *et al.*, 2011] [2012][Gao *et al.*, 2012] .To take the advantage of the sparse property, the sparse representation-based method can fix the number of the candidates adaptively and avoid the redundant or deficient problem of candidates. Nevertheless, once the sketch sparse representation coefficients are not entirely same to the photo sparse representation coefficients and do not match to the sketch patch dictionary very well, both the common and specific structures may lose.

The Bayesian inference-based methods have only one layer and take the neighboring relation of patches into consideration. Wang *et al.* [2009] investigate a MRF-based face sketch model and put the sketch patches, their neighbours and candidates into the probabilistic graphic model. Since the MRF-based method only apply one candidate, it is possible that this candidate is not similar to the test patch when the training data is insufficient. And the synthesized sketches would be lack in the identity-specific information. Hence, Zhou *et al.* [2012] borrow the idea of the weighted combination of candidates and present a weighted MRF (MWF) method. Furthermore, in order to achieve the best candidates and weights, Wang *et al.* [2013] [2017] put forward the alternating MWF-based method between the candidates and weights. And other models are raised to replace the candidate pixels. They extract the multiple features from the sketch and photo patches [Peng *et al.*, 2016], or utilize the super-pixels [Peng *et al.*, 2015] or sparse codes [Zhang *et al.*, 2015] of the image patches. These models may bring some improvements on losing problem of the specific facial structure. Nevertheless, they do not build the direct relationship between photos and sketches and exploit the full identity-specific information.

2.2 Deep Learning-based Methods

The deep learning-based methods are in the multiplayer structures and become the mainstream in a variety of applications, such as heterogeneous image transformation and super-resolution. The deep learning-based face sketch synthesis method pay more attentions on generating the specific structure of the test photo. In the training stage, the deep learning model constructs a direct mapping relationship from the training photos to the corresponding training sketches. The test photo is put into the learned model and the synthesized sketch is then to be the output.

Gatys *et al.* [2017] propose an artistic style generator including a convolutional content network and a convolution style network which transfers the style of the input image to the output. It can deal with the abstract painting style. But the dainty sketch style cannot be transferred well due to the loss of the detailed pixel information in the higher layers. The generative adversarial network (GAN) proposed by Next Isola *et al.* [2014] has two convolutional networks and they are play a two-player min-max game. It can synthesize the face sketches, but the synthesized sketches produce noises and lose some common structures, leading to the poor performance on image quality and face sketch recognition accuracy. And hence Zhang *et al.* [2017] propose a fully convolutional layers (FCN) to improve the performance to a certain extent. But the synthesized sketch gives a blurred and noised visual effect. The reason lies in the fact that the FCN-based method is only stacked by a great deal of convolutional layers and

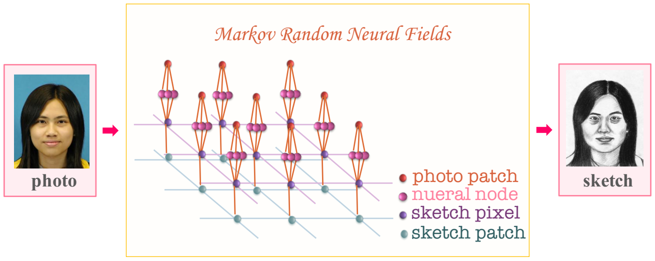


Figure 1: Framework of the proposed face sketch synthesis based on Markov random neural fields. We produce the sketch pixels by the neural network and fuse them into an initial face sketch. Then we divide it into patches and put them into the weighted Markov random network to obtain the final face sketch.

it is difficult to model the mapping relationship between the heterogeneous images: face photos and sketches directly.

3 Face Sketch Synthesis based on Markov Random Neural Fields

Face sketch synthesis focus on how to learn the information from the training photos and the corresponding sketches. Based on the previous discussion, we can safely achieve the conclusion that the learned information should contain two kinds of information: the identity-specific information and the common information. Thus, according to this understanding, we propose the MRNF-based face sketch synthesis including two structures. The first structure pays more attentions on how to generate the specific structures, whereas the second structure concentrates on how to produce the common structures. A flowchart of the proposed face sketch synthesis based on MRNF is shown in Fig.1.

3.1 First Structure

The first structure in an undirected graphical form builds the mapping relationship between the face photo patches and the face sketch pixels and captures the characteristics which is not in the training data but in the test data.

In the training stage, this structure learns the conditional probability of the sketch pixels $\mathbf{Z}$ depending on the training photo patches $\mathbf{X}$. It can be defined as:

$$P(\mathbf{Z}|\mathbf{X}) = \frac{exp(\mathbf{\Omega})}{\int_{-\infty}^{\infty} exp(\mathbf{\Omega}) d\mathbf{Z}} \tag{1}$$

Our potential function $\mathbf{\Omega}$ includes the neural function f_c and neighboring function g_c. It can be written as:

$$\mathbf{\Omega} = \sum_{i=1}^{N}\sum_{c=1}^{C} \alpha_c f_c(\mathbf{Z}_i, \mathbf{X}, \theta_c) + \sum_{(i,j)\in\Xi}\sum_{c=1}^{C} \beta_c g_c(\mathbf{Z}_i, \mathbf{Z}_j) \tag{2}$$

where $i, i = 1, \dots, N$ denotes the ith sketch pixel. $(i,j) \in \Xi$ denotes the ith sketch pixel and the jth are neighbors. $c, c = 1, \dots, C$ denotes the cth neuron in a single layer neural network. α_c and β_c are the parameters of the neural function f_c and neighboring function g_c, respectively. θ_c is the weight vector.

For each neuron c, the neural function f_c means the mapping relationship between the sketch pixels $\mathbf{Z}_i$ and the training photo patches $\mathbf{X}$.

$$f_c(\mathbf{Z}_i, \mathbf{X}, \theta_c) = -(\mathbf{Z}_i - h(\theta_c, \mathbf{X}_i))^2 \tag{3}$$

The sigmoid function $h(\theta_c, \mathbf{X}_i)$ is denoted by:

$$h(\theta_c, \mathbf{X}_i) = \frac{1}{1 + exp(-\theta_c^T \mathbf{X}_i)} \tag{4}$$

so that $0 \leq h(\theta_c, \mathbf{X}_i) \leq 1$. $\mathbf{Z}_i$ is mapped to 0-1 range.

And the neighboring function g_c describes the similarities between the sketch pixel $\mathbf{Z}_i$ and its neighboring sketch pixel $\mathbf{Z}_j$.

$$g_c(\mathbf{Z}_i, \mathbf{Z}_j) = -S_{i,j}(\mathbf{Z}_i - \mathbf{Z}_j)^2 \tag{5}$$

where $S_{i,j}$ is the neighborhood measure. When the sketch pixel $\mathbf{Z}_i$ is the neighbor of the sketch pixel $\mathbf{Z}_j$, there is a connection between the two pixels and the neighborhood measure $S_{i,j}$ is set to 1. On the contrary, when the sketch pixel $\mathbf{Z}_i$ is not the neighbor of the sketch pixel $\mathbf{Z}_j$, there is no connection between them and the neighborhood measure $S_{i,j}$ is set to 0.

$$S_{i,j} = \begin{cases} 1, & |i-j| = 1 \\ 0, & otherwise \end{cases} \tag{6}$$

Above all, for all neurons in the first structure, there exist parameters $\{\alpha, \beta, \theta\}$, where $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_C\}$, $\beta = \{\beta_1, \beta_2, \dots, \beta_C\}$, and $\theta = \{\theta_1, \theta_2, \dots, \theta_C\}$. Thus, we need to estimate the parameters $\{\alpha, \beta, \theta\}$ from the Eq. (1). According to [Baltrusaitis *et al.*, 2014], the Eq. (1) can be reformulated in a multivariate Gaussian form as follow:

$$P(\mathbf{Z}|\mathbf{X}) = \frac{1}{(2\pi)^{n/2}|\mathbf{\Sigma}|^{1/2}} exp(-\frac{1}{2}(\mathbf{Z} - \mu)^T \mathbf{\Sigma}^{-1}(\mathbf{Z} - \mu)) \tag{7}$$

The mean matrix μ of the distribution is expressed as:

$$\mu = \frac{2\alpha^T \mathbf{\Sigma}}{1 + exp(-\theta \mathbf{X})} \tag{8}$$

where $2\alpha^T/(1 + exp(-\theta\mathbf{X}))$ represents the contribution of the neural function f_c while the covariance matrix $\mathbf{\Sigma}$ represents the contribution of the neighboring function g_c. We denote the inverse of the covariance matrix $\mathbf{\Sigma}$ as:

$$\mathbf{\Sigma}^{-1} = 2(\mathbf{A} + \mathbf{B}) \tag{9}$$

The diagonal matrix $\mathbf{A}$ which describes the contribution of the parameter α and the symmetric matrix $\mathbf{B}$ which represents the contribution of the parameter β are listed respectively.

$$\mathbf{A}_{i,j} = \begin{cases} \sum_{c=1}^{C} \alpha_c, & i = j \\ 0, & i \neq j \end{cases} \tag{10}$$

$$\mathbf{B}_{i,j} = \begin{cases} (\sum_{c=1}^{C} \beta_c \sum_{j=1}^{n} S_{i,j}) - (\sum_{c=1}^{C} \beta_c S_{i,j}), & i = j \\ -\sum_{c=1}^{C} \beta_c S_{i,j}, & i \neq j \end{cases} \tag{11}$$

Thus, we can optimize the parameters $\{\alpha, \beta, \theta\}$ by maximizing the log-likelihood of Eq. (7)

$$\max_{\alpha,\beta,\theta} \sum logP(\mathbf{Z}|\mathbf{X}) \tag{12}$$

To estimate the parameters, we apply the partial derivatives of the $logP(\mathbf{Z}|\mathbf{X})$.

$$\frac{\partial logP(\mathbf{Z}|\mathbf{X})}{\alpha_c} = -\mathbf{Z}^T\mathbf{Z} + 2\mathbf{Z}^T\mathbf{H}^T + 2\mathbf{H}^T\mu + \mu^T\mu + tr(\mathbf{\Sigma}) \tag{13}$$

$$\frac{\partial logP(\mathbf{Z}|\mathbf{X})}{\beta_c} = -\mathbf{Z}^T\frac{\partial \mathbf{B}}{\partial \beta_c}\mathbf{Z} + \mu^T\frac{\partial \mathbf{B}}{\partial \beta_c}\mu + tr(\mathbf{\Sigma}\frac{\partial \mathbf{B}}{\partial \beta_c}) \tag{14}$$

$$\frac{\partial logP(\mathbf{Z}|\mathbf{X})}{\theta_{i,j}} = -\mathbf{Z}^T\frac{\partial \mathbf{b}}{\partial \theta_{i,j}} + \mu^T\frac{\partial \mathbf{b}}{\partial \theta_{i,j}} \tag{15}$$

$$\mathbf{H} = \frac{1}{1 + exp(-\theta\mathbf{X})} \tag{16}$$

$$\mathbf{b}_i = 2\sum_{c=1}^{C} \alpha_c h(\theta_c, \mathbf{X}_i) \tag{17}$$

where tr denotes the matrix trace.

In the test stage, the generated sketch pixels $\mathbf{Z}'$ can be inferred from the test photo patches $\mathbf{X}'$ as follow:

$$\mathbf{Z}' = arg\max_{\mathbf{Z}} P(\mathbf{Z}|\mathbf{X}') \tag{18}$$

Using the property of the multivariate Gaussian, the generated sketch pixels $\mathbf{Z}'$ can be solved straightforward and equal to the mean value of the multivariate Gaussian distribution.

$$\mathbf{Z}' = \mu \tag{19}$$

3.2 Second Structure

The first structure produces the vivid characteristic features of the test photo, but it ignores the common structures of the face. Thus, the task of this structure is to synthesize the sketch patches $\mathbf{Y}'$ with the common information from the test photo patches $\mathbf{X}'$ and the generated sketch pixels $\mathbf{Z}'$. Based on Bayesian theorem, we formulate the problem as the probability $P(\mathbf{Y}'|\mathbf{Z}', \mathbf{X}')$.

$$P(\mathbf{Y}'|\mathbf{Z}', \mathbf{X}') = \frac{P(\mathbf{Y}', \mathbf{Z}', \mathbf{X}')}{P(\mathbf{Z}'|\mathbf{X}')P(\mathbf{X}')} \tag{20}$$

$P(\mathbf{Z}'|\mathbf{X}')$ is obtained from the first structure. $P(\mathbf{X}')$ is a normalization term. Hence, only the joint probability $P(\mathbf{Y}', \mathbf{Z}', \mathbf{X}')$ is needed to maximize. Since we can fuse the generated sketch pixels to the initial sketch and combine the $\mathbf{Y}'$ with their K candidates in line with the initial sketch and test photo weighted by ω, the problem can be reformulated to maximize the probability $P(\omega_1, ..., \omega_M, \mathbf{z}'_1, ..., \mathbf{z}'_M, \mathbf{x}'_1, ..., \mathbf{x}'_M)$.

$$\begin{aligned} &\max_{\omega_i} \; P(\omega_1, ..., \omega_M, \mathbf{z}'_1, ..., \mathbf{z}'_M, \mathbf{x}'_1, ..., \mathbf{x}'_M) \\ &\propto \max_{\omega_i} \prod_{i=1}^{M} \Phi(\mathbf{z}'_i, \omega_i) \prod_{i=1}^{M} \Psi(\mathbf{x}'_i, \omega_i) \\ &\quad \prod_{(i,j)\in\Xi} \Upsilon(\omega_i, \omega_j) \end{aligned} \tag{21}$$

The weighted combination of candidate sketch should preserve the identity-specific information of the sketch generated in the first structure and be close to the generated sketch.

$$\Phi(\mathbf{z}'_i, \omega_i) = \exp\{-\|\mathbf{z}'_i - \sum_{k=1}^{K} \omega_i \mathbf{z}'_{i,k}\|^2 / 2\sigma_1^2\} \tag{22}$$

The linear combination of candidate photo should reserve the common information of the public face and be similar to the test photo.

$$\Psi(\mathbf{x}'_i, \omega_i) = \exp\{-\|\mathbf{x}'_i - \sum_{k=1}^{K} \omega_i \mathbf{x}'_{i,k}\|^2 / 2\sigma_2^2\} \tag{23}$$

The overlapping area should be as smooth as possible.

$$\Upsilon(\omega_i, \omega_j) = \exp\{-\|\sum_{k=1}^{K} \omega_{i,k}\mathbf{e}^j_{i,k} - \sum_{k=1}^{K} \omega_{j,k}\mathbf{e}^i_{j,k}\|^2 / 2\sigma_3^2\} \tag{24}$$

where $\mathbf{e}^i_{j,k}$ and $\mathbf{e}^j_{j,k}$ represent the overlapping area of the ith and jth patches for the kth candidate photos. Eq. (21) is equivalent to minimizing the cost function as

$$\begin{aligned} \min_{\mathbf{W}} \quad & \rho\sum_{i=1}^{N} \|\mathbf{z}'_i - \mathbf{Z}'_i\mathbf{W}\|^2 + \tau\sum_{i=1}^{N} \|\mathbf{x}'_i - \mathbf{X}'_i\mathbf{W}\|^2 \\ & + \sum_{(i,j)\in\Xi} \|\mathbf{E}_i^j\mathbf{W} - \mathbf{E}_j^i\mathbf{W}\|^2 \end{aligned} \tag{25}$$

where the parameters $\rho = \sigma_1^2/\sigma_3^2, \tau = \sigma_2^2/\sigma_3^2$. $\mathbf{Z}'_i$, $\mathbf{X}'_i$, $\mathbf{E}_i^j$ and $\mathbf{E}_j^i$ are matrices corresponding to $\mathbf{z}'_i$, $\mathbf{x}'_i$, $\mathbf{e}^j_{i,k}$ and $\mathbf{e}^i_{j,k}$.We can formulate (25) as a standard convex QP problem solved by the cascade decomposition method [Zhou *et al.*, 2012].

4 Experimental Results and Analysis

In this section, the experiments are conducted to verify the effectiveness of the proposed MRNF-based method. We conduct the experiments on the Chinese University of Hong Kong (CUHK) face sketch database (CUFS) [Wang and Tang, 2009]. There exist 606 face sketch-photo pairs including 188 face sketch-photo pairs of the CUHK student database, 123 face sketch-photo pairs of the AR database [Martinez and Benavente, 1998], and 295 face sketch-photo pairs of the XM2VTS database [Messer *et al.*, 1999]. We compare the

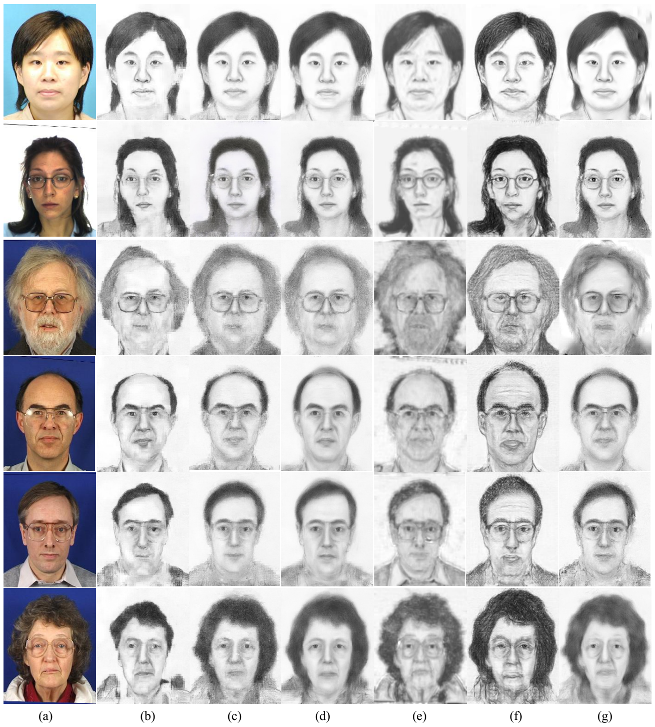


Figure 2: Comparison between the proposed method and the conventional methods for synthesizing sketches on CUFS. (a) Input photos. (b)-(g) Results of MRF-based, MWF-based, Bayesian-based, FCN-based, GAN-based and MRNF-based methods, respectively.

Comparison Methods	Eigenface	Fisherface
MRF-based method	94.0	89.3
MWF-based method	94.7	89.7
Bayesian-based method	95.3	91.7
FCN-based method	82.0	85.0
GAN-based stage	94.0	89.3
Our method	**97.0**	**98.7**

Table 1: Recognition accuracy on CUFS using Eigenface and Fisherface (%)

Comparison Methods	SSIM	VIF
MRF-based method	0.4282	0.0693
MWF-based method	0.4605	0.0786
Bayesian-based method	0.4622	0.0790
FCN-based method	0.4254	0.0707
GAN-based stage	0.4118	0.0736
Our method	**0.4674**	**0.0801**

Table 2: SSIM and VIF values on CUFS

proposed method with the MRF-based method [Wang and Tang, 2009], the MWF-based method [Zhou *et al.*, 2012], the Bayesian-based method [Wang *et al.*, 2017], the FCN-based method [Zhang *et al.*, 2017], and the GAN-based method [Goodfellow *et al.*, 2014].

4.1 Face Sketch Synthesis

In the CUHK student database, 88 pairs are randomly selected to form the training set. And the remaining 100 pairs are used for model test. In the AR database, we choose 100 pairs for model training and the remaining 23 pairs for model test. In the XM2VTS database, we select 100 pairs for training, and the remaining 195 pairs as the test data.

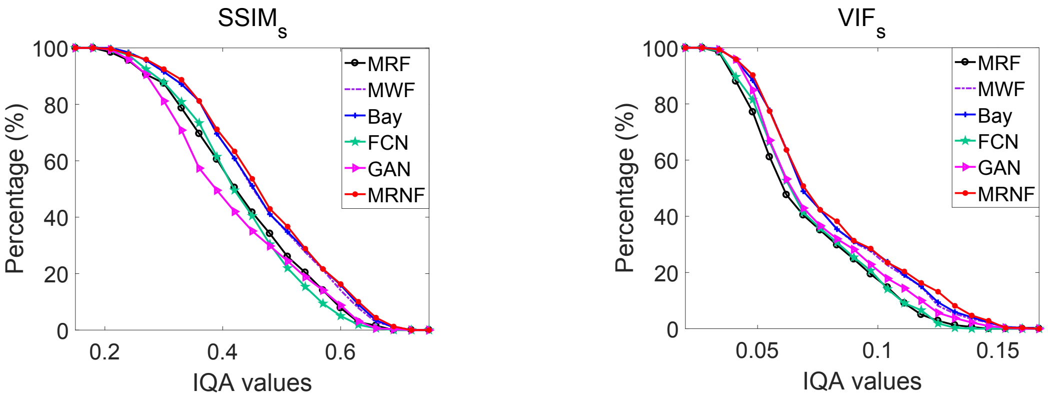
Visual comparisons of the synthesized sketches are shown in Fig.2. The sketches synthesized by the proposed MRNF-based method not only contain the identity-specific information, such as the ear in the first line, the glasses in the last five lines, and the hairstyles in the third and last lines, but also include the common structures, such as the face contours in the second and last two lines, and the eyes under the glasses in the last five lines. The MRF-based method produces the sketches without some common and specific facial structures, resulting from only one candidate to synthesize the sketches. Even though the MWF-based and Bayesian-based methods can produce the new candidates by the weighted combination, either the MWF-based method or Bayesian-based method fails to generate some characteristics which exist only in the test samples, such as the glasses in the last three lines. The sketches synthesized by the FCN-based method appear noisy and blurred. The main rationale behind this is that there exist only a series of convolutional layers to learn the complicated mapping relationship between face photos and sketches. The results of the GAN-based method have some distortions on the facial components, such as the face contours in the second and last two lines, and the eyes in the middle two lines. And the face structures in the last three lines are bigger than that of the corresponding face photos.

4.2 Face Sketch Recognition

In this section, we do the face sketch recognition experiments including Eigenface and Fisherface [Tang and Wang, 2003]. These methods can be viewed as two facets of the performance of the face sketch synthesis methods. The Eigenface is the unsupervised method, whereas the Fisherface is the supervised method.

To recognize the target face between the homogeneous images, we synthesize the face photos in the gallery to sketches and then match the target artist-drawn face sketches. There are 606 pairs in CUFS. They are divided into three parts. The first part has 153 pairs to train the face sketch synthesis models. The other 153 pairs are used to train the Eigenface or Fisherface classifier. The third part including 300 pairs is utilized to calculate the recognition accuracy rate and test the performance of the face sketch synthesis methods.

We compare the recognition accuracy rates of our proposed MRNF-based method with other competitors on CUFS by Eigenface and Fisherface (Table 1). The face recognition accuracy rates of our proposed method are the highest. They are 97.0% and 98.7% respectively. It illustrates our proposed method can synthesize the clean face sketches with the identity-specific and common information of the test photos. The recognition accuracy rates of the shallow learning-based method, such as the MRF-based method [Wang and Tang, 2009], the MWF-based method [Zhou *et al.*, 2012], and the Bayesian-based method [Wang *et al.*, 2017] are lower than



(a)

(b)

Figure 3: Comparison of the SSIM (a) and VIF (b) values for synthesizing sketches on CUFS.

the proposed methods. The reason behinds this is that the results of them lose some specific structures which can be regarded as the identifiable information for every identity. The face sketches synthesized by the deep learning-based methods, such as the FCN-based method [Zhang *et al.*, 2017], and the GAN-based method [Goodfellow *et al.*, 2014] deliver noisy and blunted impacts which hamper the recognition accuracy.

4.3 Image Quality Assessment

In order to evaluate the proposed MRNF-based face sketch synthesis method quantitatively, we induce the full reference image quality assessment (FR-IQA) [Gao *et al.*, 2015] by following the most convention face sketch synthesis methods for comparison. For the CUFS database, the original artist-drawn sketches are regarded as the reference images. And at the same time, we regard the corresponding synthesized photos as the distorted images. In this section, we evaluate performance of the face sketch synthesis methods by two FR-IQA metrics: the visual information fidelity index (VIF) [Wang *et al.*, 2004] and the structural similarity index (SSIM) [Sheikh and Bovik, 2006].

In Table 2, we list the average of the SSIM values and VIF values of the proposed method with the other competitors. It can be seen that the proposed MRNF-based method outperforms the other conventional face sketch methods in the both SSIM average value and VIF average value. The proposed MRNF-based method achieves the SSIM average value as 0.4674 and the VIF average value as 0.0801 while the best state-of-the-art obtains the SSIM and VIF average value as 0.4622 and 0.0790 from the Bayesian-based method.

It can be clearly seen that the curves of the proposed MRNF-based method are higher than the other ones on the horizontal axis in Fig.3. And we can drive the conclusion that both the SSIM values and VIF values of the sketches synthesized by the proposed MRNF-based method are bigger than that of the state-of-art methods. But there is a little visible improvement. Because the existing IQA methods pay less attention on the synthesized sketches. A more rational and applicable objective synthesized IQA framework is needed.

5 Conclusion

In this paper, we propose a face sketch synthesis based on MRNF. The proposed model is composed of two structures. The first structure pays more attentions on generating the identity-specific information and maps the relationship between the face photo patches and sketch pixels. It can be formulated to a multivariate Gaussian problem optimized by a gradient-based method. On the contrary, the second structure focus on producing the common information and regards the face photo patches and the generated sketch pixels as the input and the sketch patches as the output. We reformulate the model as a standard QP problem which can be solved by a cascade decomposition method. Superior visual results with the vivid and clean textures validate the effectiveness of the proposed MRNF-based framework. Compared to the state-of-the-arts, either the recognition accuracy or the quality assessment is the highest one.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (under Grants 61501339, 61772402, 61671339, 61432014, U1605252), in part by the China Post-Doctoral Science Foundation under Grant 2017M623125, in part by Opening Project of Science and Technology on Reliability Physics and Application Technology of Electronic Component Laboratory (under Grant 17D03-ZHD201701), in part by Young Elite Scientists Sponsorship Program by CAST (under Grant 2016QNRC001), in part by Natural Science Basic Research Plan in Shaanxi Province of China (under Grant 2017JM6085 and 2017JQ6007), in part by Young Talent fund of University Association for Science and Technology in Shaanxi, China, in part by CCF-Tencent Open Fund (under Grant IAGR 20170103), in part by the Shaanxi Province Post-Doctoral Science Foundation, and in part by the Fundamental Research Funds for the Central Universities under Grant XJS17109, XJS17086.

References

[Baltrusaitis *et al.*, 2014] T. Baltrusaitis, P. Robinson, and Morency L. Continuous conditional neural fields for structured regression. In *Proc. Eur. Conf. Comput. Vis.*, pages 593–608, 2014.

[Chang *et al.*, 2010] M. Chang, L. Zhou, Y. Han, and X. Deng. Face sketch synthesis via sparse representation. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 2146–2149, 2010.

[Gao *et al.*, 2012] X. Gao, N. Wang, D. Tao, and X. Li. Face sketch-photo synthesis and retrieval using sparse representation. *IEEE Trans. Circuits Syst. Video Technol.*, 22(8):1213–1226, 2012.

[Gao *et al.*, 2015] F. Gao, D. Tao, and X. Li. Learning to rank for blind image quality assessment. *IEEE Trans. Neural Netw. Learn. Syst.*, 26(10):2275–2290, 2015.

[Gatys *et al.*, 2017] L. Gatys, A. Ecker, and M. Bethge. A neural algorithm of artistic style. *arXiv Preprint:1508.06576*, 2017.

[Goodfellow *et al.*, 2014] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Proc. Int. Conf. Neural Information Proc. Syst.*, pages 2672–2680, 2014.

[Liu *et al.*, 2005] Q. Liu, X. Tang, H. Jin, H. Lu, and S. Ma. A nonlinear approach for face sketch synthesis and recognition. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 1005–1010, 2005.

[Liu *et al.*, 2007] W. Liu, X. Tang, and J. Liu. Bayesian tensor inference for sketch-based face photo hallucination. In *Proc. IEEE Conf. Artif. Intell.*, pages 2141–2146, 2007.

[Martinez and Benavente, 1998] A. Martinez and R. Benavente. The AR face database. *CVC Technical Report*, 24, 1998.

[Messer *et al.*, 1999] K. Messer, J. Matas, J. Kittler, and G. Luettin, J.and Maitre. XM2VTSDB: the extended M2VTS database. In *Proc. Int. Conf. Audio and Video-Based Biometric Person Authentication*, pages 72–77, 1999.

[Peng *et al.*, 2015] C. Peng, X. Gao, N. Wang, and J. Li. Superpixel-based face sketch-photo synthesis. *IEEE Trans. Circuits Syst. Video Technol.*, pages 1–12, 2015.

[Peng *et al.*, 2016] 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.*, pages 1–15, 2016.

[Sheikh and Bovik, 2006] H. Sheikh and A. Bovik. Image information and visual quality. *IEEE Trans. Image Process.*, 15(2):430–444, 2006.

[Song *et al.*, 2014] Y. Song, L. Bao, Q. Yang, and M. H. Yang. Real-time exemplar-based face sketch synthesis. In *Proc. Eur. Conf. Comput. Vis.*, pages 800–813, 2014.

[Tang and Wang, 2002] X. Tang and X. Wang. Face photo recognition using sketch. In *Proc. IEEE Int. Conf. Image Process.*, pages 257–260, 2002.

[Tang and Wang, 2003] X. Tang and X. Wang. Face sketch synthesis and recognition. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 687–694, 2003.

[Tang and Wang, 2004] X. Tang and X. Wang. Face sketch recognition. *IEEE Trans. Circuits Syst. Video Technol.*, 14(1):1–7, 2004.

[Wang and Tang, 2009] X. Wang and X. Tang. Face photo-sketch synthesis and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(11):1955–1967, 2009.

[Wang *et al.*, 2004] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.*, 13(4):600–612, 2004.

[Wang *et al.*, 2011] 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.*, pages 82–87, 2011.

[Wang *et al.*, 2012] 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.*, pages 2216–2223, 2012.

[Wang *et al.*, 2013] N. Wang, D. Tao, X. Gao, X. Li, and J. Li. Transductive face photo-sketch synthesis. *IEEE Trans. Neural Netw. Learn. Syst.*, 24(9):1364–1376, 2013.

[Wang *et al.*, 2017] N. Wang, X. Gao, Sun L., and Li J. Bayesian face sketch synthesis. *IEEE Trans. Image Process.*, 26(3):1264–1274, 2017.

[Zhang *et al.*, 2015] S. Zhang, X. Gao, N. Wang, J. Li, and M. Zhang. Face sketch synthesis via sparse representation-base greedy search. *IEEE Trans. Image Process.*, 24(8):2466–2477, 2015.

[Zhang *et al.*, 2017] L. Zhang, L. Lin, X. Wu, S. Ding, and L. Zhang. End-to end photo-sketch generation via fully convolutional representation learning. *arXiv Preprint:1508.06576*, 2017.

[Zhou *et al.*, 2012] H. Zhou, Z. Kuang, and K. Wong. Markov weight fields for face sketch synthesis. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 1091–1097, 2012.