J Inf Process Syst, Vol.15, No.6, pp.1449~1461, December 2019
https://doi.org/10.3745/JIPS.02.0125

ISSN 1976-913X (Print)
ISSN 2092-805X (Electronic)

JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

Face Sketch Synthesis Based on Local and Nonlocal Similarity Regularization

Songze Tang*, Xuhuan Zhou*, Nan Zhou*, Le Sun**, and Jin Wang***

Abstract

Face sketch synthesis plays an important role in public security and digital entertainment. In this paper, we present a novel face sketch synthesis method via local similarity and nonlocal similarity regularization terms. The local similarity can overcome the technological bottlenecks of the patch representation scheme in traditional learning-based methods. It improves the quality of synthesized sketches by penalizing the dissimilar training patches (thus have very small weights or are discarded). In addition, taking the redundancy of image patches into account, a global nonlocal similarity regularization is employed to restrain the generation of the noise and maintain primitive facial features during the synthesized process. More robust synthesized results can be obtained. Extensive experiments on the public databases validate the generality, effectiveness, and robustness of the proposed algorithm.

Keywords

Face Sketch Synthesis, Local Similarity, Nonlocal Similarity, Patch Representation

1. Introduction

In real world, it is not easy to directly obtain a frontal face image of the criminal suspect in an actual investigation. The suspect often intentionally cover their face in a surveillance video [1,2]. However, an artist can draw a sketch for us according to information in the video surveillance. This sketch may then serve as a substitute for identifying the suspect. Face sketch synthesis, which refers to transformation of a face photo into a sketch, has recently been used. We roughly divide face sketch synthesis methods into two categories: image-based and patch-based.

1.1 Prior Works

Image-based methods treat the input photo image as a whole, and produce a sketch image with some models. Wang et al. [3] converted greyscale images to pencil sketches, in which the pencil strokes adhered to the image features. Li and Cao [4] proposed a simple two-stage framework for face photo-sketch synthesis. These methods did not mimic well a sketch style. Since the breakthrough of deep learning [5], it has been achieved great attention in image processing problems [6-8]. Fully convolutional network

Manuscript received May 21, 2018; first revision August 19, 2019; accepted October 14, 2019.

Corresponding Author: Songze Tang (ts198708@163.com)

* Dept. of Criminal Science and Technology, Nanjing Forest Police College, Nanjing, China (ts198708@163.com, 416126986@qq.com, nudge@163.com)

** School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, China (sunlecncom@163.com)

*** School of Computer & Communication Engineering, Changsha University of Science & Technology, Changsha, China (jinwang@csust.edu.cn)

(FCN) was first introduced to learn some mapping functions from photos to sketches [9]. Recently, the generative adversarial network (GAN) [10] has been attracting growing attention. To infer photo-realistic natural images, Ledig et al. proposed a perceptual loss function that consisted of adversarial loss and a content loss [7]. Based on the GAN method, a class of loss functions were designed to generate images [11] with perceptual similarity metrics. Wang et al. employed a back projection strategy to improve the final synthesized performance further [12].

Different patch-based methods have been proposed. They are divided into three categories, i.e., subspace learning based methods [13-17], sparse representation based methods [18-23], and Bayesian inference based methods [24-28].

The subspace learning framework mainly includes linear subspace-based methods and nonlinear subspace-based methods. The seminal work in linear face sketch synthesis was the eigen-transformation method [13]. Considering the complexity of human faces, a linear relationship may not always hold, thus Liu et al. [14] proposed to characterize the nonlinear process of face sketch synthesis according to the concept of locally linear embedding (LLE) [15]. Inspired by the image denoising method, Song et al. [16] explored the K surrounding spatial neighbors for face sketch synthesis. Instead of searching for neighbors online, Wang et al. [17] randomly sampled the similar patches offline, and used them to reconstruct the target sketch patch. It was named Fast-RSLCR.

Due to the great success of the sparse representation in many image processing problems [18-20]. Chang et al. [21] incorporated it into face sketch synthesis. To mitigate the handicap in the face image retrieval process, a two-step framework was proposed by Gao et al. [22]. They obtained a coarse estimation using neighbor selection, and then enhanced the definition of the initial estimate by sparse representation. The above methods assumed that the photo patches and the corresponding sketch patches had the same sparse representation coefficients. Actually, the relationships between different styles of images are complex. The assumptions are not always sufficient. Wang et al. [23] proposed to learn a map of the sparse coefficients between the sketch patch and the corresponding photo patch.

To take the constraints between neighboring image patches into consideration, Bayesian inference methods explore the constraints between neighboring image patches. Wang and Tang [25] considered the relationship at different scales, and they called their method the multi-scale Markov random fields (MRF) method. This method generated facial deformations. To address this issue, Zhou et al. [26] proposed a Markov weight field (MWF) method by embedding the LLE idea into the MRF model. Because lighting and pose variations often appear, neighbor selection is not robust. Peng et al. [28] adaptively represented an image patch by multiple features to improve the robustness.

1.2 Motivation and Contributions

As is generally understood, face images have strong structural similarity in local regions (the mouth is not similar to the nose) [29], which means that similarity is low between photo patches and sketch patches that do not correspond to the same small region. Thus, a local similarity constraint is employed to search for the best matching neighbor patches from the training sets. In addition, inspired by the nonlocal denoising method [30], a nonlocal similarity regularization is also introduced to further improve the sketch synthesis quality.

The main contributions of our work can be summarized as follows:

- We impose local similarity constraints on the selection of similar patches. This improves the quality of the synthesized sketches by discarding dissimilar training patches.

- In addition, taking into consideration the redundancy of image patches, a global nonlocal similarity regularization is employed to inhibit the generation and maintain primitive facial features during the synthesis process. Thus, more robust synthesized results can be achieved.

2. Related Work

Given a training set with M face photo-sketch pairs, we divide each training image into N small overlapping patches. Let $\mathbf{X}$ and $\mathbf{Y}$ be an input test photo and an estimated sketch image, which are divided into N overlapping patches $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\}$ and $\{\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_N\}$, respectively, in the same way. For each test photo patch $\mathbf{x}_i$, we reconstruct it using the K nearest photo patches $\mathbf{P}_{i,K} = \{\mathbf{p}_{i,k}\}_{k=1}^{K}$ from the training dataset with corresponding weight vector $\mathbf{w}_{i,K} = \{w_{i,k}\}_{k=1}^{K}$. Thus, the corresponding sketch image patch $\mathbf{y}_i$ can be synthesized by the corresponding K nearest sketch patches $\{\mathbf{s}_{i,k}\}_{k=1}^{K}$ with the above obtained weight vector $\mathbf{w}_{i,K}$.

2.1 LLE

For a test photo patch $\mathbf{x}_i$, the training set was searched for the K nearest photo patches by Euclidean distance. The combination weight was then achieved according to LLE.

$$\underset{w_{i,k}}{arg\min} \left\| \mathbf{x}_i - \sum_{k=1}^{K} w_{i,k} \mathbf{p}_{i,k} \right\|_2^2, \ s.t. \sum_{k=1}^{K} w_{i,k} = 1 \tag{1}$$

where $w_{i,k}$ represents the linear combination weight for the *k*-th photo patch $\mathbf{p}_{i,k}$. We can rewrite (1) as

$$\underset{\mathbf{w}_{i,K}}{arg\min} \left\| \mathbf{x}_i - \mathbf{P}_{i,K} \mathbf{w}_{i,K} \right\|_2^2, \ s.t. \mathbf{1}^T \mathbf{w}_{i,K} = 1 \tag{2}$$

After the combination weight is obtained from (2), the target sketch patch $\mathbf{y}_i$ can be synthesized as:

$$\mathbf{y}_i = \sum_{k=1}^{K} w_{i,k} \mathbf{s}_{i,k} \tag{3}$$

When we generated all the sketch patches from (3), a whole sketch $\mathbf{Y}$ can be assembled by averaging overlapping pixel values.

2.2 MWF

Taking the relationship between adjacent sketch patches into account, Zhou et al. [26] proposed an MWF method, which introduced a linear combination into the MRF model. This is equivalent to minimizing the following cost function.

$$\mathop{arg\min}_{\mathbf{w}_{i,K}} \sum_{i=1}^{N} \left\| \mathbf{x}_i - \mathbf{P}_{i,K} \mathbf{w}_{i,K} \right\|_2^2 + \lambda \sum_{(i,j)\in \mathrm{Ne}} \left\| \mathbf{O}_i^j \mathbf{w}_{i,K} - \mathbf{O}_j^i \mathbf{w}_{i,K} \right\|_2^2, \ s.t. \mathbf{1}^T \mathbf{w}_{i,K} = 1, w_{i,k} \geq 0 \tag{4}$$

where $(i,j)\in \mathrm{Ne}$ represents the *i*-th and *j*-th patches are neighbors. $\mathbf{O}_i^j$ is a matrix with the column $\mathrm{o}_{i,k}^j$, which denotes the overlapping area of the *k*-th candidate for the *i*-th sketch patch and the *j*-th patch. λ is a balancing parameter between the two terms.

2.3 Fast-RSLCR

The abovementioned methods search for nearest neighbors online, thus, the time consumption of testing becomes significantly slower. Wang et al. [17] randomly sampled patches offline. A locality constraint was imposed to regularize the reconstruction weights.

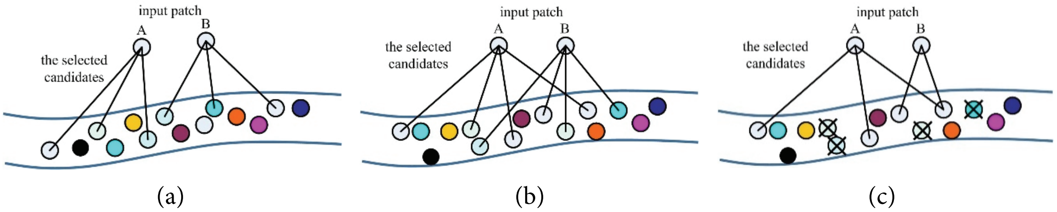
$$\mathop{arg\min}_{\mathbf{w}_{i,N}} \left\| \mathbf{x}_i - \mathbf{P}_{i,N} \mathbf{w}_{i,N} \right\|_2^2 + \lambda \left\| \mathbf{d}_i \cdot \mathbf{w}_{i,N} \right\|_2^2, \ s.t. \mathbf{1}^T \mathbf{w}_{i,N} = 1 \tag{5}$$

where $\mathbf{d}_i = \left\| \mathbf{x}_i - \mathbf{p}_i \right\|_2, (1 \leq i \leq N)$ measures the distance between $\mathbf{x}_i$ and $\mathbf{p}_i$. λ is a balancing parameter. The Fast RSLCR method speeds up the synthesis process. However, more candidate patches were allowed to be sampled. This method has two shortcomings: one is that the discriminability of the synthesized sketch is reduced, and the other is that it increases the spatial complexity.

3. Face Sketch Synthesis Based on Local and Nonlocal Similarity Regularization

3.1 Adaptive Regularization by Local Similarity

As mentioned, the LLE method [15] calculates the linear combination coefficient vector without using the appropriate constraints, thus, biased solutions can be obtained easily. Due to the large quantization errors, similar test patches may have different content (Fig. 1(a)). In Fast-RSLCR, each test patch is more accurately represented by capturing the correlations between different training patches. However, not all patches play a positive role in the final results of the face sketch synthesis. A greater number of patches contributing to the final synthesized result implies lower discriminability, as shown in Fig. 1(b). Therefore, we introduce a local similarity regularization to the neighbor selection, which leads to (i) a stable solution, and (ii) discriminant synthesized results (Fig. 1(c)).



(a) (b) (c)

Fig. 1. Comparison between LLE (a), Fast-RSLCR (b), and the proposed (c).

In our local similarity constraint model, we consider only the most relevant patches in the training set as effective samples. For each patch $\mathbf{x}_i$ in the test photo, the optimal weights are obtained by minimizing the local similarity regularized reconstruction error:

$$\underset{\mathbf{w}_{i,K}}{arg\min}\left\|\mathbf{x}_i-\mathbf{P}_{i,K}\mathbf{w}_{i,K}\right\|_2^2+\lambda_1\left\|\mathbf{D}_{i,K}\mathbf{w}_{i,K}\right\|_2^2,\ s.t.\,\mathbf{1}^T\mathbf{w}_{i,K}=1 \tag{6}$$

where $\mathbf{D}_{i,K}=\begin{bmatrix} d_{i,1} & & & 0 \\ & d_{i,2} & & \\ & & \ddots & \\ 0 & & & d_{i,K} \end{bmatrix}$, $d_{i,j}=\exp\left(\frac{\left\|\mathbf{x}_i-\mathbf{p}_j\right\|_2^2}{\sigma}\right)$ represents the entries of the exponential locality adaptor and σ is a positive number. The K sampled training photo patches constitute the matrix $\mathbf{P}_{i,K}$. $\mathbf{w}_{i,K}$ is the weight representation. λ_1 is a balancing parameter. To preserve the data structure, the exponential function is used to improve representation. Because $d_{i,j}$ grows exponentially with $\left\|\mathbf{x}_i-\mathbf{p}_j\right\|_2^2/\sigma$, the exponential locality adaptor will be quite large when $\mathbf{x}_i$ and $\mathbf{p}_j$ are far apart. This property is useful when we want to stress the importance of data locality (Because $d_{i,j}$ is the weight of $w_{i,j}$ in (6), a large value of $d_{i,j}$ causes $w_{i,j}$ to be small).

To determine the solution $\mathbf{w}_{i,K}$ in (6), we consider the Lagrange function $L\left(\mathbf{w}_{i,K},\lambda_1,\beta\right)$, which is defined as

$$L\left(\mathbf{w}_{i,K},\lambda_1,\beta\right)=\left\|\mathbf{x}_i-\mathbf{P}_{i,K}\mathbf{w}_{i,K}\right\|_2^2+\lambda_1\left\|\mathbf{D}_{i,K}\mathbf{w}_{i,K}\right\|_2^2+\beta\left(\mathbf{w}_{i,K}\mathbf{1}^T\text{-}1\right) \tag{7}$$

(7) can be reformulated as

$$\begin{aligned} L\left(\mathbf{w}_{i,K},\lambda_1,\beta\right)&=\left\|\mathbf{x}_i\mathbf{1}^T\mathbf{w}_{i,K}-\mathbf{P}_{i,K}\mathbf{w}_{i,K}\right\|_2^2+\lambda_1\left\|\mathbf{D}_{i,K}\mathbf{w}_{i,K}\right\|_2^2+\beta\left(\mathbf{1}^T\mathbf{w}_{i,K}\text{-}1\right) \\ &=\mathbf{w}_{i,K}{}^T\left(\mathbf{x}_i\mathbf{1}^T-\mathbf{P}_{i,K}\right)^T\left(\mathbf{x}_i\mathbf{1}^T-\mathbf{P}_{i,K}\right)\mathbf{w}_{i,K}+\lambda_1\mathbf{w}_{i,K}{}^T\mathbf{D}_{i,K}{}^T\mathbf{D}_{i,K}\mathbf{w}_{i,K}+\beta\left(\mathbf{1}^T\mathbf{w}_{i,K}-1\right) \\ &=\mathbf{w}_{i,K}{}^T\mathbf{Z}\mathbf{w}_{i,K}+\beta\left(\mathbf{1}^T\mathbf{w}_{i,K}-1\right) \end{aligned} \tag{8}$$

where $\mathbf{1}$ is a column vector where all values are equal to 1. $\mathbf{Z}=\left(\mathbf{x}_i\mathbf{1}^T-\mathbf{P}_{i,K}\right)^T\left(\mathbf{x}_i\mathbf{1}^T-\mathbf{P}_{i,K}\right)+\lambda_1\mathbf{D}_{i,K}{}^T\mathbf{D}_{i,K}$. By setting $\frac{\partial}{\partial\mathbf{w}_{i,K}}L\left(\mathbf{w}_{i,K},\lambda_1,\beta\right)=0$ and $\frac{\partial}{\partial\lambda_1}L\left(\mathbf{w}_{i,K},\lambda_1,\beta\right)=0$ according to (8), we obtain the combination weight.

$$\mathbf{w}_{i,K}=\frac{\mathbf{Z}^{-1}\mathbf{1}}{\mathbf{1}^T\mathbf{Z}^{-1}\mathbf{1}} \tag{9}$$

Then the sketch patch $\mathbf{y}_i$ can be synthesized by (3) with the weight in (9). Finally, a whole sketch $\mathbf{Y}$ can be assembled by averaging overlapping pixel values.

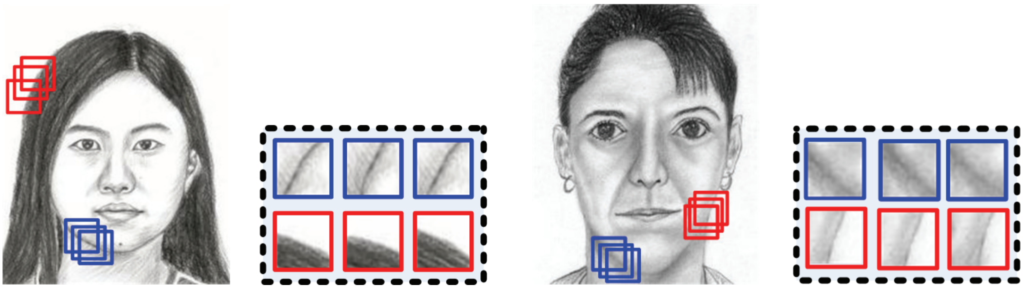


Fig. 2. Nonlocal similarity in the sketch images.

3.2 Adaptive Regularization by Nonlocal Similarity

The local context constraint model exploits local geometry in data space. There are also many repetitive patterns throughout a sketch image, which is quite helpful for improving the quality of final sketch images [31,32], as shown in Fig. 2. Therefore, we explore nonlocal self-similarity. Generally, for each extracted patch $\mathbf{y}_i$ from the sketch image $\mathbf{Y}$, we search for its L similar patches $\left\{\mathbf{y}_i^l\right\}_{l=1}^L$ in $\mathbf{Y}$. Then, there is a following linear relationship between $\mathbf{y}_i$ and $\left\{\mathbf{y}_i^l\right\}_{l=1}^L$

$$\mathbf{y}_i=\sum_{l=1}^{L}\mathbf{y}_i^l b_i^l \tag{10}$$

The nonlocal similarity weight b_i^l is inversely proportional to the distance between patches $\mathbf{y}_i$ and $\mathbf{y}_i^l$ in (10) and value is calculated as

$$b_i^l=\exp\left(-\left\|\hat{\mathbf{y}}_i\text{-}\hat{\mathbf{y}}_i^l\right\|_2^2/h\right) \tag{11}$$

where h is a pre-determined control factor of the weight. Let $\mathbf{b}_i$ be the column vector containing all weights b_i^l and $\boldsymbol{\beta}_i$ be the column vector containing all $\mathbf{y}_i^l$. (11) can be rewritten as:

$$\mathbf{y}_i=\mathbf{b}_i^T\boldsymbol{\beta}_i \tag{12}$$

By incorporating the nonlocal similarity regularization term (12) into patch aggregation, we obtain:

$$\mathbf{Y}^*=\arg\min_{\mathbf{Y}}\left\{\sum_i\left\|\mathbf{R}_i\mathbf{Y}-\mathbf{y}_i\right\|_2^2+\lambda_2\sum_i\left\|\mathbf{y}_i-\mathbf{b}_i^T\boldsymbol{\beta}_i\right\|_2^2\right\} \tag{13}$$

where $\mathbf{R}_i$ is to extract a patch from an image. (13) can be rewritten as

$$\mathbf{Y}^*=\arg\min_{\mathbf{Y}}\left\{\sum_i\left\|\mathbf{R}_i\mathbf{Y}-\mathbf{y}_i\right\|_2^2+\lambda_2\sum_i\left\|\left(\mathbf{I}-\mathbf{B}\right)\mathbf{Y}\right\|_2^2\right\} \tag{14}$$

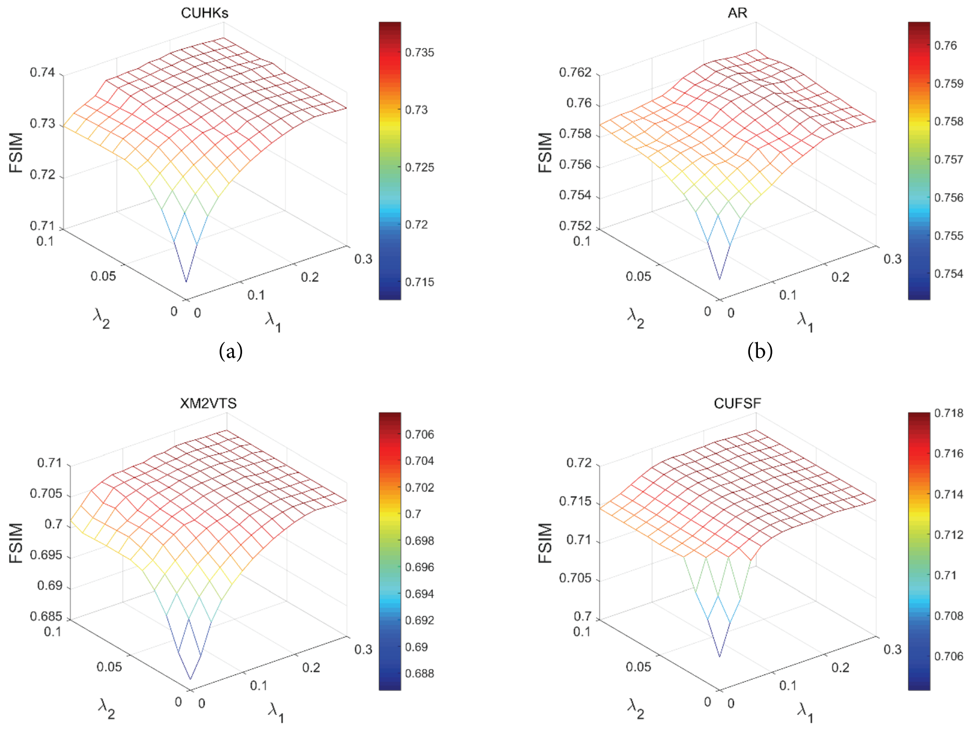
where $\mathbf{I}$ is the identity matrix and $\mathbf{B}(i,j)=\begin{cases} b_i^l, & \textit{if } \mathbf{y}_i^l \textit{ is an elment of } \boldsymbol{\beta}_i, b_i^l\in\mathbf{b}_i \\ 0, & \textit{otherwise}\end{cases}$. Now, we can easily get the final synthesized image

$$\mathbf{Y}^* = \left\{ \sum_i \mathbf{R}_i^T \mathbf{R}_i + \lambda_2 (\mathbf{I}-\mathbf{B})^{-1} (\mathbf{I}-\mathbf{B}) \right\}^{-1} \left(\sum_i \mathbf{R}_i^T \mathbf{y}_i \right) \tag{15}$$

4. Experimental Results and Analysis

4.1 Database Description

We validate our method on the Chinese University of Hong Kong (CUHK) face sketch database (CUFS) [25] and the CUHK face sketch FERET database (CUFSF) [33]. The CUFS database contains three sub-datasets, i.e., the CUHK student (CUHKs) database, the AR database [34], and the XM2VTS database [35]. In the CUHKs database, 88 photo-sketch pairs were constructed a training set, and the remaining 100 pairs were used for testing. In the AR database, 80 pairs were randomly selected as the training set and the rest were used as test cases. As to the XM2VTS database, the training set had 100 pairs. There are 1,194 photo–sketch pairs in the CUFSF database [36]. The 250 pairs were randomly selected to construct the training set, and the remaining 944 pairs were used as test cases. All face images were cropped to 250×200 pixels.



(a)

(b)

(c)

(d)

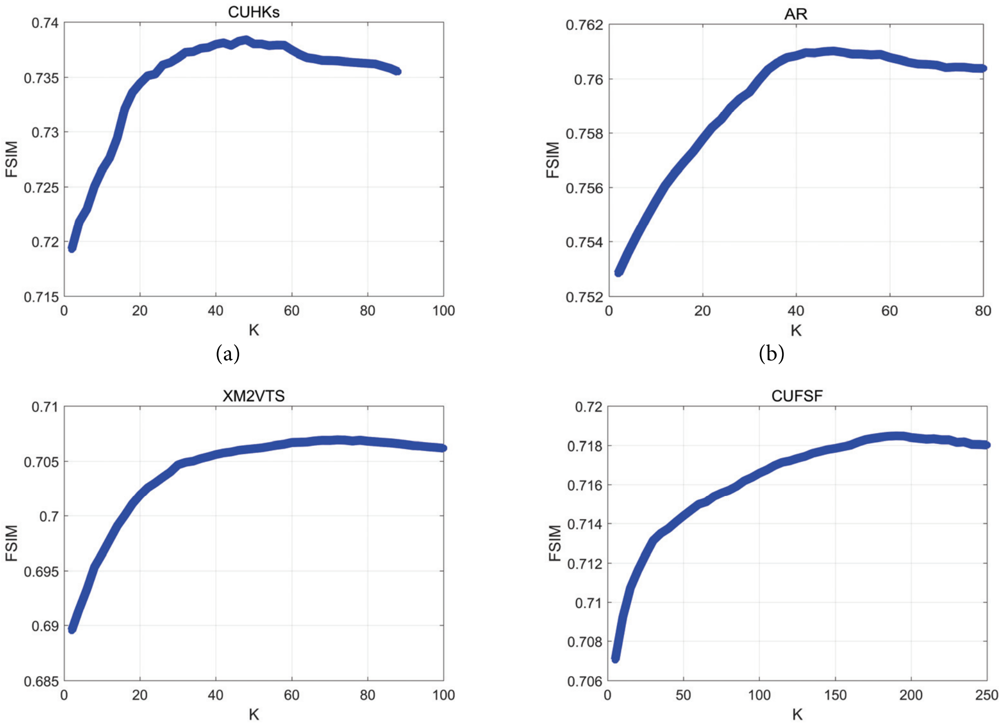
Fig. 3. FSIM as a function of the regularization parameters λ_1 and λ_2 on different datasets: (a) CUHKs, (b) AR, (c) XM2VTS, and (d) CUFSF.

The proposed method was compared with some related methods, including the LLE [14], MWF [26], Fast-RSLCR [17], FCN [9], and BP-GAN [12]. The feature similarity index metric (FSIM) [37] was adopted as the evaluation criterion to estimate the quality of final synthesized sketches.

4.2 Discussion on the Parameters

4.2.1 The influence of different regularization parameters

Our algorithm has two free-regularization parameter λ_1 and λ_2 , which balance the different contribution of regularization terms. A set of parametric experiments are performed to validate the effectiveness of the proposed regularization terms. We carefully tune the local similarity parameter λ_1 (from 0 to 0.3 with step size of 0.02) and the nonlocal similarity parameter λ_2 (from 0 to 0.1 with step size of 0.01). Fig. 3 shows the surfaces of FSIM variations. It can be clearly observed that the synthesized performance are stable in terms of FSIM with regards to $\lambda_1 \in [0.18, 0.26]$ and $\lambda_2 \in [0.05, 0.07]$.



(a)

(b)

(c)

(d)

Fig. 4. FSIM score of the different databases with different numbers of the nearest patches: (a) CUHKs, (b) AR, (c) XM2VTS, and (d) CUFSF.

4.2.2 The influence of nearest neighbor number

The synthesized performance generated by the proposed method correlated with the number of the nearest neighbors. We conduct experiments on the four mentioned databases by changing the value of *K*. The curved lines of the FSIM values plotted against the number of training patches are shown in Fig.

4. When the value of *K* is equal to the number of training photos, our proposed method does not obtain the best performance. As shown in Fig. 4, the values of the FSIM increase steadily with the increase in the number of nearest neighbors. Nonetheless, after the nearest patch number reaches a suitable value, the performance of our proposed method remains constant. We also note that performance shows a descending trend with the increasing value of *K* (for a value of larger than 80% of the number of training photos). In view of this, to achieve the optimal or nearly optimal performance, we recommend setting *K* = 80% of the number of training photos.

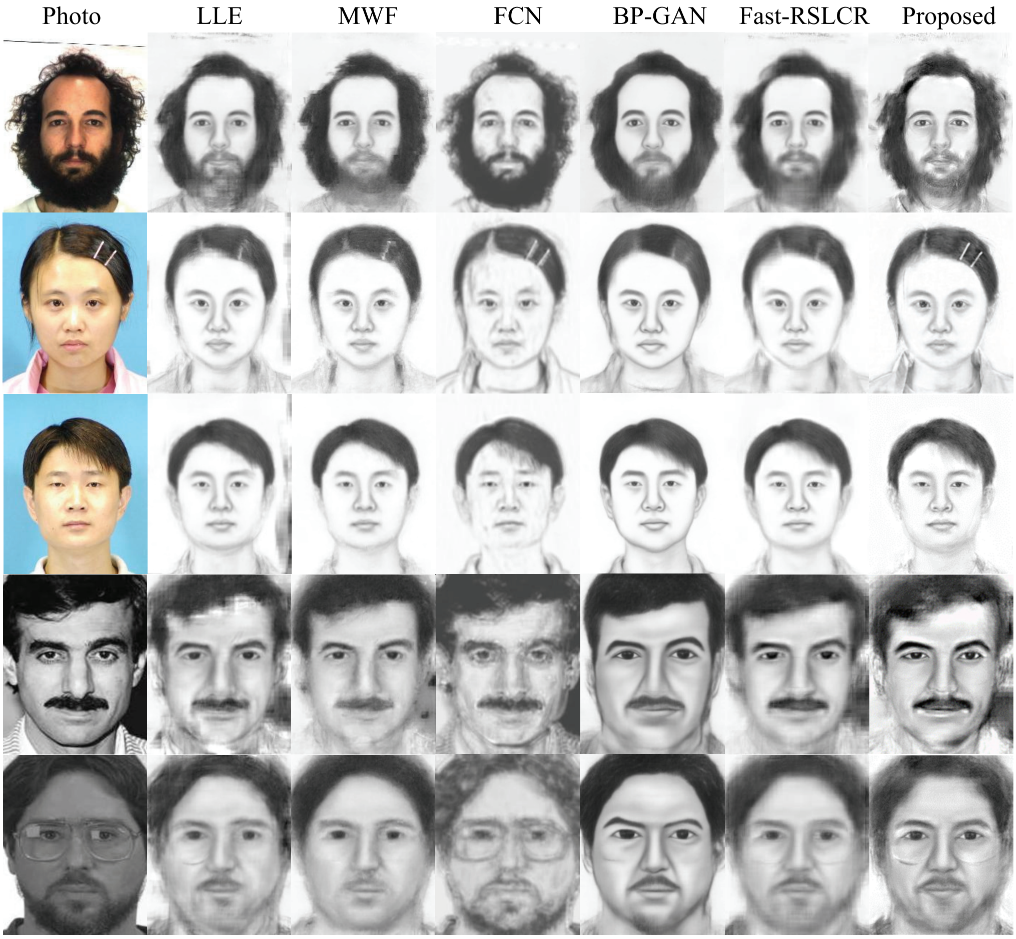


Fig. 5. Synthesized sketches on different databases by LLE, MWF, FCN, BP-GAN, Fast-RSLCR, and the proposed method, respectively.

4.3 Face Sketch Synthesis

Fig. 5 presents some synthesized sketches using different methods on the abovementioned databases. Generally speaking, the proposed method generates much more detail in comparison with the other five popular methods. In the AR database, we note that the LLE, MWF, and Fast-RSLCR methods produce very smooth sketches (the first row in Fig. 5). For example, they did not generate some details, such as hair. Then we compared the synthesized sketches on the CUHKs database. As shown in the remaining results of Fig. 5, the proposed method presents much better synthesized performance than other patch-based methods. Textures (e.g., hair regions) are synthesized successfully. FCN can produce some details, but some distortions are shown in the results. The results for BP-GAN look very well, but some details are also missing, such as the hairs of the first and second person. As shown in the results, our approach predicted unusual features well, while the comparison methods tended to smooth these regions. This illustrates the robustness and effectiveness of the proposed method.

To investigate the robustness of the proposed methods against the complex illumination, we compared the synthesis results on the CUFSF database using different methods, as shown the last two rows in Fig. 5. Our proposed method generated competitive results with more facial detail. Table 1 presents the average FSIM comparisons of different methods.

Table 1. Average FSIM scores of different methods on different databases

	LLE	MWF	FCN	BP-GAN	Fast-RSLCR	Proposed
CUFS	0.7180	0.7294	0.6936	0.6899	0.7142	0.7352
CUFSF	0.7043	0.7029	0.6624	0.6814	0.6775	0.7181

Table 2. Time consuming (in second) on different databases

	LLE	MWF	FCN	BP-GAN	Fast-RSLCR	Proposed
CUFS	758.95	21.48	0.019	339.35	3.34	10.83
CUFSF	1,964.57	53.11	0.028	618.72	5.00	17

4.4 Time Consuming

To compare the time cost of the proposed face sketch synthesized method, we further list the runtimes of our algorithm and the other five competitors on different databases in Table 2. It can be seen that the FCN method has the fastest computation time of less than 0.1. The BP-GAN method has a long processing time due to the neighbor selection process. Our proposed method runs slowly compared with the Fast-RSLCR method. Overall, the proposed method gets the best synthesis results within a moderate time consumption between the comparison algorithms.

5. Conclusion and Future Work

In this paper, we presented a novel face sketch synthesis method using two regularization terms. By incorporating a local similarity regularization term into the neighbor selection, we selected the most relevant patch samples to reconstruct face sketch versions of the input photos, thus generating discriminant face sketches with detailed features. A global nonlocal similarity regularization term was employed to further maintain primitive facial features. The results of thorough experimental testing on public databases demonstrated the superiority of the proposed method over other methods.

Compared with traditional synthesis methods, our novel generative approach retained more detailed information from the photos. However, our inference time was dependent on the amount of training data. Thus, we could incorporate priors into the deep learning method to improve performance and speed up the processing in the future.

Acknowledgement

This paper is supported in part by the National Natural Science Foundation of China (No. 61702269, 61971233, and 61671339), in part by the Natural Science Foundation of Jiangsu Province (No. BK20171074), and the Fundamental Research Funds for the Central Universities at Nanjing Forest Police College (No. LGZD201702).

References

[1] R. G. Uhl and N. da Vitoria Lobo, "A framework for recognizing a facial image from a police sketch," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, 1996, pp. 586-593.

[2] N. Wang, X. Gao, L. Sun, and J. Li, "Anchored neighborhood index for face sketch synthesis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 9, pp. 2154-2163, 2017.

[3] J. Wang, H. Bao, W. Zhou, Q. Peng, and Y. Xu, "Automatic image-based pencil sketch rendering," *Journal of Computer Science and Technology*, vol. 17, no. 3, pp. 347-355, 2002.

[4] X. Li and X. Cao, "A simple framework for face photo-sketch synthesis," *Mathematical Problems in Engineering*, vol. 2012, article no. 910719, 2016.

[5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436-444, 2015.

[6] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on Computer Vision And Pattern Recognition*, Honolulu, HI, 2017, pp. 1125-1134.

[7] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE Conference on Computer Vision And Pattern Recognition*, Honolulu, HI, 2017, pp. 4681-4690.

[8] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295-307, 2015.

[9] L. Zhang, L. Lin, X. Wu, S. Ding, and L. Zhang, "End-to-end photo-sketch generation via fully convolutional representation learning," in *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, Shanghai, China, 2015, pp. 627-634.

[10] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in Neural Information Processing Systems*, vol. 27, pp. 2672-2680, 2014.

[11] A. Dosovitskiy and T. Brox, "Generating images with perceptual similarity metrics based on deep networks," *Advances in Neural Information Processing Systems*, vol. 29, pp. 658-666, 2016.

[12] N. Wang, W. Zha, J. Li, and X. Gao, "Back projection: an effective postprocessing method for GAN-based face sketch synthesis," *Pattern Recognition Letters*, vol. 107, pp. 59-65, 2018.

[13] X. Tang and X. Wang, "Face photo recognition using sketch," in *Proceedings of IEEE International Conference on Image Processing*, Rochester, NY, 2002.

[14] Q. Liu, X. Tang, H. Jin, H. Lu, and S. Ma, "A nonlinear approach for face sketch synthesis and recognition," in *Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, CA, 2005, pp. 1005-1010.

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

[16] Y. Song, L. Bao, Q. Yang, and M. H. Yang, "Real-time exemplar-based face sketch synthesis," in *Computer Vision-ECCV 2014.* Cham: Springer, 2014, pp. 800-813.

[17] N. Wang, X. Gao, and J. Li, "Random sampling for fast face sketch synthesis," *Pattern Recognition*, vol. 76, pp. 215-227, 2018.

[18] S. Tang, L. Xiao, P. Liu, L. Huang, N. Zhou, and Y. Xu, "Pansharpening via sparse regression," *Optical Engineering*, vol. 56, no. 9, article no. 093105, 2017.

[19] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1031-1044, 2010.

[20] R. Qi, Y. Zhang, and H. Li, "Block sparse signals recovery via block backtracking-based matching pursuit method," *Journal of Information Processing Systems*, vol. 13, no. 2, pp. 360-369, 2017.

[21] L. Chang, M. Zhou, Y. Han, and X. Deng, "Face sketch synthesis via sparse representation," in *Proceedings of the 20th International Conference on Pattern Recognition*, Istanbul, Turkey, 2010, pp. 2146-2149.

[22] X. Gao, N. Wang, D. Tao, and X. Li, "Face sketch–photo synthesis and retrieval using sparse representation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 8, pp. 1213-1226, 2012.

[23] S. Wang, L. Zhang, Y. Liang, and Q. Pan, "Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI, 2012, pp. 2216-2223.

[24] X. Gao, J. Zhong, J. Li, and C. Tian, "Face sketch synthesis algorithm based on E-HMM and selective ensemble," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 4, pp. 487-496, 2008.

[25] X. Wang and X. Tang, "Face photo-sketch synthesis and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 11, pp. 1955-1967, 2008.

[26] H. Zhou, Z. Kuang, and K. Y. K. Wong, "Markov weight fields for face sketch synthesis," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI, 2012, pp. 1091-1097.

[27] W. Zhang, X. Wang, and X. Tang, "Lighting and pose robust face sketch synthesis," in *Computer Vision-ECCV2010*. Heidelberg: Springer, 2010, pp. 420-433.

[28] C. Peng, X. Gao, N. Wang, D. Tao, X. Li, and J. Li, "Multiple representations-based face sketch–photo synthesis," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 11, pp. 2201-2215, 2016.

[29] C. Li, S. Zhao, K. Xiao, and Y. Wang, "Face recognition based on the combination of enhanced local texture feature and DBN under complex illumination conditions," *Journal of Information Processing Systems*, vol. 14, no. 1, pp. 191-204, 2018.

[30] A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, San Diego, CA, 2005, pp. 60-65.

[31] Z. Zha, X. Zhang, Q. Wang, Y. Bai, Y. Chen, L. Tang, and X. Liu, "Group sparsity residual constraint for image denoising with external nonlocal self-similarity prior," *Neurocomputing*, vol. 275, pp. 2294-2306, 2018.

[32] D. Liu, B. Wen, Y. Fan, C. C. Loy, and T. S. Huang, "Non-local recurrent network for image restoration," *Advances in Neural Information Processing Systems*, vol. 31, pp. 1673-1682, 2018.

[33] W. Zhang, X. Wang, and X. Tang, "Coupled information-theoretic encoding for face photo-sketch recognition," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI, 2011, pp. 513-520.

[34] A. Martinez and R. Benavente, "The AR face database," 1998; http://www.cat.uab.cat/Public/Publications/1998/MaB1998/CVCReport24.pdf.

[35] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, "XM2VTSDB: the extended M2VTS database," in *Proceedings of the 2nd International Conference on Audio and Video-Based Biometric Person Authentication*, Washington, DC, 1999.

[36] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine intelligence*, vol. 22, no. 10, pp. 1090-1104, 2000.

[37] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: a feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378-2386, 2011.

Songze Tang https://orcid.org/0000-0002-1964-6579

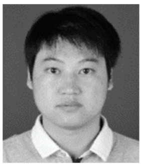
He received his B.S. degree in information and computation science from Anhui Agriculture University, Hefei, China, in 2009, and the Ph.D. degree from the School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing, China, in 2015. In 2016, he joined the Department of Criminal Science and Technology, Nanjing Forest Police College, as a Lecturer. His current research interests include inverse problems in signal representation and face recognition.

Xuhuan Zhou https://orcid.org/0000-0002-9055-9275

She received her B.S. degree in mathematics and applied mathematics from Shandong University, Hefei, China, in 2011, and the Ph.D. degree from the School of Mathematics Science, Zhejiang University, Hangzhou, China, in 2016. In 2016, he joined the Department of Information Science and Technology, Nanjing Forest Police College, as a Lecturer. Her current research interests include inverse problems in image processing, harmonic analysis, and differential equation.

Nan Zhou https://orcid.org/0000-0001-6959-6138

He received the B.S. degree from the School of Chemical Engineering, Anhui University of Science and Technology, Nanjing, China, in 2009, and the Ph.D. degree with the National Key Laboratory of Transient Physics, Nanjing University of Science and Technology, in 2014. He is currently a Lecturer with the Department of Criminal Science and Technology, Nanjing Forest Police College. His research interests include video detection, and image processing.



Le Sun https://orcid.org/0000-0001-6465-8678

He received the B.S. degree from the School of Science, Nanjing University of Science and Technology (NJUST), Nanjing, China, in 2009, and the Ph.D. degree with the School of Computer Science and Engineering, NJUST, in 2014. He holds a post-doctoral position at the Digital Media Laboratory, School of Electronic and Electrical Engineering, Sungkyunkwan University, Korea. He is currently a Lecturer with the School of Computer and Software, Nanjing University of Information Science and Technology. His research interests include hyperspectral image processing, sparse representation, and compressive sensing.

Jin Wang https://orcid.org/0000-0002-6516-6787

He received the B.S. and M.S. degrees from Nanjing University of Posts and Telecommunications, China in 2002 and 2005, respectively. He received Ph.D. degree from Kyung Hee University, Korea in 2010. Now, he is a professor in the School of Computer & Communication Engineering, Changsha University of Science & Technology. His research interests mainly include routing protocol and algorithm design, performance evaluation and optimization for wireless ad-hoc and sensor networks.