FACE SKETCH-PHOTO SYNTHESIS BASED ON SUPPORT VECTOR REGRESSION

Jiewei Zhang[1], Nannan Wang[1], Xinbo Gao[1], Dacheng Tao[2], Xuelong Li[2]

1. School of Electronic Engineering, Xidian University, Xi'an 710071, Shaanxi, P.R. China
2. Center for OPTical IMagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, P. R. China

ABSTRACT

The existing face sketch-photo synthesis methods trend to lose some vital details more or less. In this paper, we propose a novel sketch-photo synthesis approach based on support vector regression (SVR) to handle this difficulty. First, we utilize an existing method to acquire the initial estimate of the synthesized image. Then, the final synthesized image is obtained by combining the initial estimate and the SVR based high frequency information together to further enhance the quality of synthesized image. Experimental results on the benchmark database and our new constructed database demonstrate that the proposed method can achieve significant improvement on perceptual quality. Moreover, the synthesized face images can obtain higher recognition rate when used in retrieval system.

***Index Terms*—** Face sketch-photo synthesis, high-frequency information, support vector regression.

1. INTRODUCTION

Automatic face retrieval systems have been widely used in law enforcement in recent years. But sometimes face photos of suspects are not available and it is possible to search a photo in the photo database using a corresponding sketch drawn by an artist. However, most of existing face recognition algorithms can not be applied directly due to the heterogeneity between sketches and photos. Then this problem can be overcome by face sketch-photo synthesis. The synthesized images have two modalities: pseudo-sketch and pseudo-photo. A pseudo-sketch synthesized from an input photo is obtained to narrow down the difference between a photo and its corresponding sketch, and we can also obtain a pseudo-photo synthesized from an input sketch. Apart from law enforcement, pseudo-sketch/pseudo-photo with good perceptual quality can be applied to entertainment.

The synthesis methods go through linear to nonlinear approaches. Tang et al. proposed an eigentransformation algorithm [1] to synthesize a pseudo-sketch under a linear assumption. Then Liu et al. generalized the linear mapping relationship between photos and sketches and a local linear embedding (LLE) based method was introduced in [2], in which nonlinearity was approximated by local linearity. After that, several nonlinear approaches are introduced. Gao et al. proposed a novel sketch synthesis algorithm using embedded hidden Markov model (E-HMM) [3, 4], which was a real nonlinear method, and the E-HMM based photo synthesis method was proposed in [5] by Xiao et al.

Although the existing methods have made some progress, most of them trend to lose some vital details which will influence the perceptual quality and face recognition rate, so a two-step method using SVR is proposed in this paper. The first step is to get an initial estimate using an existing approach. Subsequently, a SVR based model is applied to synthesize the high-frequency information corresponding to the initial estimate. Then the two parts are fused together. In the experiments, a sparse representation based algorithm [6] is used to do face recognition. Both the recognition rate and perceptual quality are superior on the benchmark database and our new database.

2. RELATED MATHEMATICAL MODEL

To be better understanding the proposed approach, a brief introduction to SVR [7] is given in this section. Consider a set of training data

$$\{(x_i, y_i) \mid x_i \in R^n, y_i \in R, i = 1,\dots, N\}, \quad (1)$$

where x_i is a point from the input space R^n and y_i is the corresponding output from the space R . All we need do is to find a function with the least deviation between $f(x_i)$ and y_i , $i = 1,\dots, N$.

Suppose $f(x)$ takes the following form:

$$f(x) = w \cdot \phi(x) + b, \quad (2)$$

where $w \in R^n, b \in R$, and ϕ is a nonlinear function to tran-s form x to a high-dimensional space. One has to find the v-alue of w and b to minimize the regression risk error,

$$E(w,b) = C\sum_{i=0}^{N} L(y, f(x)) + \frac{1}{2}\|w\|^2, \quad (3)$$

where $E(w,b)$ is a cost function. $L(y, f(x))$ takes the form

$$L(y, f(x)) = \begin{cases} 0, & \text{if } |y - f(x)| \le \varepsilon \\ |y - f(x)| - \varepsilon, & \text{otherwise} \end{cases} . \quad (4)$$

The solution of the minimization problem (3) is as follows:

$$\begin{aligned} f(x) &= \sum_{i=1}^{N} (\alpha_i - \alpha_i^*)(\phi(x_i) \cdot \phi(x)) + b \\ &= \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) K(x_i, x) + b \end{aligned} . \quad (5)$$

where $K(\cdot)$ is a kernel function, α and α^* are Lagrange multipliers. There are a lot of other SVR types and kernel functions in [7].

3. FACE SKETCH-PHOTO SYNTHESIS USING SVR

3.1. Framework of the proposed method

As a classical regression method, SVR can effectively express the high-frequency mappings between samples and object values [8]. Thus, a SVR based face sketch-photo synthesis method is proposed to compensate the details. In the following we take the synthesis process of pseudo-sketch for instance, the synthesis process of pseudo-photo can be obtained just by switching roles of sketches and photos in pseudo-sketch synthesis process. The proposed algorithm is divided into two steps. The first step is to obtain an initial estimate by any existing method such as [2, 3, 4, 5]. Subsequently, the SVR based high-frequency information synthesis approach is given to enhance the initial estimate. Then the two parts are combined together to achieve the final pseudo-sketch. A general framework of the proposed algorithm is shown in Fig. 1.

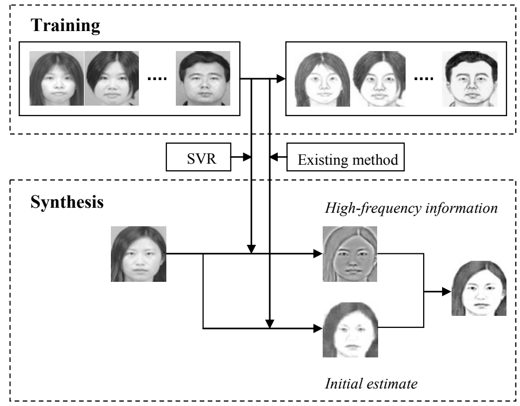

Fig. 1. Framework of the proposed method

3.2. Learning the high-frequency information

Given the initial estimate of the sketch has been obtained, the detail of generating high-frequency information is described in this subsection. First, the sketch-photo training data set can be expressed as the following form:

$$\{(P_i, S_i) \mid P_i \in R^n, S_i \in R, i = 1,..., K\} . \quad (6)$$

Considering the complexity of face structure, sketches and photos are divided into T small overlapped image patches in the same way, which are denoted as P_i^t and $S_i^t, t = 1,...,T$. A clustering process is proposed before the training stage. We arrange the first order gradients and second order gradients of patches P_i^t in horizontal and vertical direction into a vector and take the vector as the feature of clustering. The total $K*T$ patches of the photos are clustered into M set $P_m, m = 1,...,M$ using K-means clustering algorithm and there are also M corresponding sketch sets S_m, where P_m is the set consisting of patches $P_m^r, r = 1,..., N_m$ and S_m is the set consisting of patches $S_m^r, r = 1,..., N_m$. D_m^r is defined as the central pixel of the patch S_m^r and let the real number $H_m^r = D_m^r - \text{mean}(S_m^r)$. In the training stage, a SVR model is trained between each P_m and S_m pair. For learning high-frequency information, we take the vector of the mean of P_m^r subtracted by P_m^r as the input and the output is the corresponding H_m^r in H_m. Therefore, we will have M SVR models. Then in the synthesis stage, given a photo $P_j, j = 1,..., L$, we can obtain patches $P_j^t, t = 1,...,T$. Then for each P_j^t we choose a SVR model generated in the training stage through nearest neighborhood, and we will obtain the corresponding H_j^t. The final pseudo-sketch is the sum of the initial approximate and the high-frequency information H_j. The structure of generating high-frequency information is shown in Fig. 2.

4. EXPERIMENTS

4.1. Sketch-photo synthesis

Most of the existing algorithms conduct experiments on the Chinese University of Hong Kong (CUHK) student database which includes 188 sketch-photo pairs [9]. However, all sketches on the benchmark CUHK database are drawn by only one artist. A new sketch-photo database named VIPS including 200 face photos in total from the FERET database [10], the FRAV2D database [11] and the Indians face database [12] is constructed. There are five face sketches drawn by five different artists corresponding corresponding to each photo. Some examples of VIPS database are shown in Fig. 3. All face sketches and photos are in a frontal view, under normal lighting condition, and with a neutral expression, and they are cropped and normalized in geometric and grayscale. In our experiments, the size of all images is 64×64. For LLE method, images are divided into patches with the patch size 9×9 and the overlapping size 5×5. In the stage of generating high-frequency information, the patch size is 7×7 with an overlap 6×6. The cluster number M is set to 25 for the

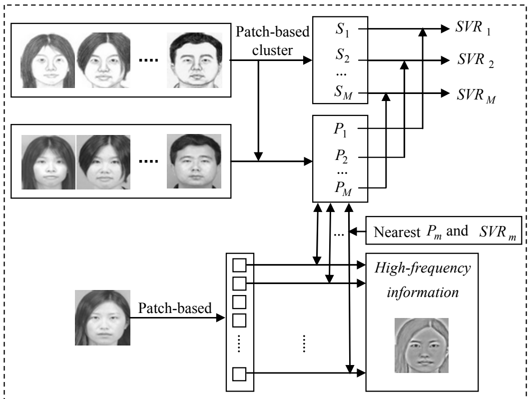

Fig. 2. The structure of generating high-frequency information

CUHK student database and 30 for the VIPS database. For SVR, the default parameters are used in LIBSVM [13]. By adding 3 pixels at each four edges, the number of patches T for each image is 4096, thus we have 4096 photo patches P_i^t corresponding to 4096 sketch central pixel D_i^t. Fig. 4 and Fig. 5 show the synthesized pseudo-sketches and pseudo-photos on the CUHK student database. Among them the (c) is synthesized by LLE in [2] and the (d) is synthesized by E-HMM with an improved form in [4]. The pseudo-sketches and pseudo-photos generated on VIPS database are presented in Fig .6 and Fig .7 respectively. The (e) is synthesized by the proposed SVR-LLE which is combination of SVR for high-frequency and LLE for initial estimate. The (f) synthesized by SVR-E-HMM is similar. As is shown in Fig 4, 5, 6 and 7, we can find that the (e) and (f) based on SVR are better than (c) and (d) respectively in perceptual quality. The proposed method based on SVR has a significant improvement on details.

4.2. Face sketch-photo recognition

In this subsection, we give a comparison of the proposed method and corresponding existing methods on face sketch-photo recognition with a sparse representation based recognition method. Experiments are conducted on two face sketch-photo retrieval systems. One system using pseudo-sketch takes pseudo-sketches as the dictionary in the recognition stage and the corresponding face sketches are the testing images. The other system using pseudo-photo takes photos as the dictionary in the recognition stage and testing images are the corresponding pseudo-photos. Table 1 presents face recognition results on CUHK student database using pseudo-sketch and pseudo-photo which are generated by different synthesis methods. The number of training images on CUHK student database is 88 and the number of testing images is 100. The results in Table 2 and 3 are conducted on VIPS database using pseudo-sketch and pseudo-photo respectively. On VIPS database, if the image number for training per person is 1, the final result is the

Fig. 3. (a) Original photo. (b) Sketch database A. (c) Sketch database B. (d) Sketch database C. (e) Sketch database D. (f) Sketch database E.

Fig. 4. (a) Original photo. (b) Original sketch. (c) Pseudo-sketch generated by LLE. (d) Pseudo-sketch generated by E-HMM. (e) Pseudo-sketch generated by SVR-LLE. (f) Pseudo-sketch generated by SVR-E-HMM.

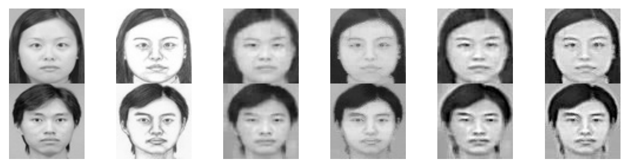

Fig. 5. (a) Original photo. (b) Original sketch. (c) Pseudo-photo generated by LLE. (d) Pseudo-photo generated by E-HMM. (e) Pseudo-photo generated by SVR-LLE. (f) Pseudo-photo generated by SVR-E-HMM.

Fig. 6. (a) Original photo. (b) Original sketch. (c) Pseudo-sketch generated by LLE. (d) Pseudo-sketch generated by E-HMM. (e) Pseudo-sketch generated by SVR-LLE. (f) Pseudo-sketch generated by SVR-E-HMM.

Fig. 7. (a) Original photo. (b) Original sketch. (c) Pseudo-photo generated by LLE. (d) Pseudo-photo generated by E-HMM. (e) Pseudo-photo generated by SVR-LLE. (f) Pseudo-photo generated by SVR-E-HMM.

mean of recognition results on five databases. If the training image per person is 5, 500 pseudo-sketches are used for dictionary in the recognition stage and 500 sketches are used for testing. As is shown in Table 1, 2, 3, the proposed methods based on SVR have a higher recognition rate than the compared ones.

Table 1. Face recognition on CUHK student database using pseudo-sketch and pseudo-photo

Image type	Pseudo-sketch	Pseudo-photo
LLE	1	0.82
SVR-LLE	1	0.96
E-HMM	1	1
SVR-E-HMM	1	1

Table 2. Face recognition on VIPS database using pseudo-sketch

Training image per person	1	5
LLE	0.902	0.904
SVR-LLE	0.916	0.918
E-HMM	0.92	0.938
SVR-E-HMM	0.928	0.94

Table 3. Face recognition on VIPS database using pseudo-photo

Training image per person	LLE	SVR -LLE	E-HMM	SVR -E-HMM
1	0.882	0.89	0.915	0.92

5. CONCLUSION

This paper presents a novel approach based on SVR to synthesize the pseudo-sketch/pseudo-photo. To improve the perceptual quality and recognition rate, we propose to obtain an initial estimate firstly, and then a SVR-based high-frequency information synthesis method is applied. We also build another database to test the performance of the algorithms. Experimental results show the effectiveness of the proposed method both on the CUHK student database and VIPS database. Future work will focus on developing a more efficient feature for clustering in generating high-frequency information owning to its importance.

6. ACKNOWLEDGEMENT

The presented research work is supported by the National Basic Research Program of China (973 Program) (Grant No.2011CB707100), the National Natural Science Foundation of China (Grant Nos. 60832005, 41031064, and 61072093), the Ph. D. Programs Foundation of Ministry of Education of China under Grant 2009203110002, the Key Science and Technology Program of Shaanxi Province of China under Grant 2010K06-12, the Natural Science Basic Research Plan in Shaanxi Province of China under Grant 2009JM8004.

7. REFERENCE

[1] X. O. Tang and X. G. Wang, “Face photo recognition using sketch,” in Proc. IEEE Int. Conference on Image Process, Rochester, NY, USA, pp. 257-260, 22-25 Sep.2002.

[2] Q. S. Liu and X. O. Tang, “A nonlinear approach for face sketch synthesis and recognition,” in Proc. IEEE Int. Conference on Computer Vision, San Diego, CA, pp. 1005-1010, 20-26 Jun. 2005.

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

[4] X. Gao, J. Zhong, D. Tao and X. Li, “Local face sketch synthesis learning,” Neurocomputing, vol. 71, no. 10-12, pp. 1921-1930, Jun. 2008.

[5] B. Xiao, X. Gao, D. Tao, and X. Li, “A new approach for face recognition by sketches in photos,” Signal Processing, vol. 89, no. 8, pp. 1531-1539, Aug. 2009.

[6] J. Wright, A. Y. Yang, A. Ganesh, S. Sastry, and Y. Ma, “Robust face recognition via sparse representation,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 31, no. 2, pp. 210-227, Feb. 2009.

[7] V. Vapnik, The nature of statistical learning theory, Springer Verlag, New York, 1995.

[8] K. S. Ni and T. Q. Nguyen, “Image superresolution using support vector regression,” IEEE Trans. Image Processing, vol. 16, no. 6, pp. 1596-1610, Jun. 2007.

[9] X. G. Wang and X. O. Tang, “Face photo-sketch synthesis and recognition,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 31, no. 11, pp. 1955-1967, Nov. 2009.

[10] P. J. Phillips, H. Wechsler, J. Huang and P. Rauss, “The FERET database and evaluation procedure for face recognition algorithms,” Image and Vision Computing, vol. 16, no. 5, pp. 295-306, Apr. 1998.

[11] A. Serrano, Martin de Diego, C. Conde, E. Cabello, L. Shen and L. Bai, “Influence of wavelet frequency and orientation in an SVM-based parallel Gabor PCA face verification system,” in Proc. Conference on Intelligent Data Engineering and Automated Learning, Birminhham, UK, pp. 219-228, 16-19 Dec. 2007.

[12] Vidit Jain and Amitabha Mukherjee, The Indian Face Database, http://vis-www.cs.umass.edu/~vidit/IndianFaceD-ase/,2002.

[13] C. C. Chang and C. J. Lin, LIBSVM: a library for support vector machines, 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.