FACE SKETCH SYNTHESIS USING E-HMM AND SELECTIVE ENSEMBLE

Juanjuan Zhong, Xinbo Gao, and Chunna Tian

Video & Image Processing System Laboratory (VIPSL),
School of Electronic Engineering, Xidian University, Xi'an 710071, China
zhongjuan19821211@tom.com, xbgao@mail.xidian.edu.cn, chnatian@hotmail.com

ABSTRACT

In this manuscript, we propose an automatic sketch synthesis algorithm based on embedded hidden Markov model (E-HMM) and selective ensemble strategy. The E-HMM is used to model the nonlinear relationship between a photo-sketch pair firstly, and then a series of pseudo-sketches, which are generated based on several learned models for a given photo, are integrated together with selective ensemble strategy to synthesize a finer face pseudo-sketch. The experimental results illustrate that the proposed algorithm achieves satisfactory effect of sketch synthesis.

***Index Terms*—** Face recognition, Hidden Markov models, Unsupervised learning, Learning systems, Machine vision

1. INTRODUCTION

An important application of face recognition is the automatic retrieval of suspect's photo from the photo database. Unfortunately, the photo of a suspect is not available in most cases. To deal with such a problem, a simulated sketch is generated by the cooperation of artists and eyewitnesses as a substitute, so that the content-based image retrieval for identification is done in existing photo database to narrow down potential suspects. Thus, sketch-photo recognition, as a new branch of face recognition, comes into being to meet demands of application.

Due to the great geometrical deformations and large difference of texture and grayscale caused by the different generating mechanism and information expressing manner between sketches and photos, most of existing algorithms for face recognition cannot apply to sketch-photo recognition. As a result, the key technology in sketch-photo recognition is how to shrink differences between the two. It is instinctive to change these two kinds of images into the same pattern. Obviously, transformation from photos to sketches is more reasonable than reconstruction of photos from sketches. Hence we focus our interests on synthesizing sketches from photos for automatic sketch-photo recognition.

The research on sketch synthesis is still at its initial stage. For example, Wang *et al* proposed a sketch generation method by simulating pencil [1]; Tang *et al* achieved some good results in the filed of sketch synthesis and recognition [2]-[4]. The disadvantage of both methods above is that a great many of training samples are needed. However, the high cost of sketch acquisition limits the size of face sketch database, which restricts the generalization and applications of face sketch recognition technology.

To this end, a novel face sketch synthesis algorithm is proposed based on machine learning in this manuscript. Let P_i denotes a face photo and S_i denotes the corresponding sketch respectively, and then the key step of sketch synthesis is to learn the nonlinear mapping in each photo-sketch pair, $\mathcal{M}_i : P_i \rightarrow S_i$, which is modeled by the E-HMM in the presented algorithm. For a given face photo P', several pseudo-sketches $S'_i, i = 1, 2, \cdots, n$, which are generated by corresponding models $\mathcal{M}_i$, are fused together to synthesize a finer face sketch S' using selective ensemble strategy.

Thanks to the Key Project of Chinese Ministry of Education under Grant No.104173 and the Program for New Century Excellent Talents in University under Grant NCET-04-0948 for funding. This paper is also supported partially by Xidian Graduate Innovation Project.

2. SKETCH SYNTHESIS BASED ON E-HMM

This section introduces the sketch synthesis algorithm based on E-HMM [5] which models the nonlinear relationship between a photo-sketch pair.

2.1. E-HMM representation for face image

Suppose that the trained E-HMM of the input image has been obtained using the Baum-Welch algorithm. Then, the image sequence is decoded with the embedded Viterbi algorithm so that the optimal state sequence and mixture indices are determined. Since each state has a Gaussian probability density function (PDF), the observation of the most likely output image is equal to the mean vector in the model at the optimal state and index. The photo-sketch pair in Fig.1(a) is used for training and the images in Fig.1(b) are reconstructed from the model. The obtained state sequences are shown in Fig.1(c). It can be seen that the E-HMM has a strong ability to describe face images.

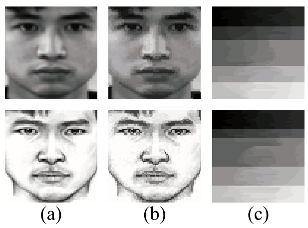

Fig. 1. The reconstructed images and the state sequences based on E-HMM: (a) The original images; (b) The reconstructed images; (c) The state sequences.

2.2. Sketch synthesis based on E-HMM

The proposed sketch synthesis algorithm based on E-HMM is illustrated in Fig.2. Firstly, we obtain the E-HMM of the photo λ_{P_1} and the E-HMM of the sketch λ_{S_1} by conducting joint-training for the given photo-sketch pair (P_1, S_1); then, decode the input photo using the model λ_{P_1}; finally, synthesize the pseudo-sketch S_2 using the decoded sequence and the E-HMM λ_{S_1}. Viterbi decoding reflects the deformation from P_1 to P_2, while partial deformations from P_1 to S_1 are learnt by joint-training. In the following, the sketch synthesis process will be described in details.

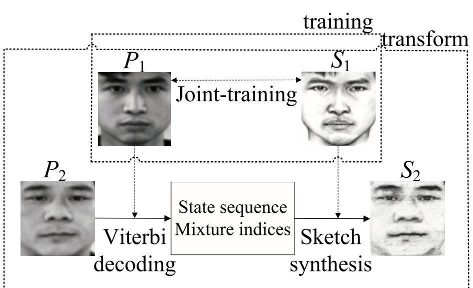

Fig. 2. The framework of sketch synthesis algorithm based on E-HMM.

At first, the joint training of the E-HMMs is introduced to model the nonlinear relationship between a photo-sketch pair. The coupled E-HMMs share a common state transition probability, while they have different mean vectors and covariance matrices in the same state [6]-[7].

Let j, k, l and m be the embedded-state index, the super-state index, the image index ($l = \{0, 1\}$ represents photo and sketch respectively) and the mixture index on the state j. O_{lt} denotes an observed vector of the image l at a pixel t. The output probability of the embedded-state j on the super-state k is given as

$$b_j^{(k)}(\bar{O}_t) = \sum_{m=1}^{N} c_{jm}^{(k)} \mathcal{N}\Big(\bar{O}_t; \mu_{jm}^{(k)}, \Sigma_{jm}^{(k)}\Big)$$

$$= \sum_{m=1}^{N} \left[c_{jm}^{(k)} \prod_{l=0}^{1} \Big\{ \mathcal{N}\Big(O_{lt}; \mu_{jml}^{(k)}, \Sigma_{jml}^{(k)}\Big) \Big\} \right] \tag{1}$$

where

$$\bar{O}_t = \Big[O_{0t}^T, O_{1t}^T\Big]^T \tag{2}$$

$$\mu_{jm}^{(k)} = \Big[\mu_{jm0}^{(k)\,T}, \mu_{jm1}^{(k)\,T}\Big]^T \tag{3}$$

$$\Sigma_{jm}^{(k)} = \begin{bmatrix} \Sigma_{jm0}^{(k)} & \mathbf{0} \\ \mathbf{0} & \Sigma_{jm1}^{(k)} \end{bmatrix} \tag{4}$$

Here, $\mu_{jml}^{(k)}$, $\Sigma_{jml}^{(k)}$, $c_{jm}^{(k)}$ and N represent mean vector, covariance matrix, mixture coefficient and the number of mixtures on the embedded-state j of the super-state k, respectively. We also suppose that $\Sigma_{jml}^{(k)}$ is a diagonal matrix. The coupled model λ is decomposed into two image's models λ_0 and λ_1 using Eqs. (3) and (4).

The coupled E-HMMs of the given photo-sketch pair is trained by using a feature vector, which is composed of five components of the pixel gray value, Gaussian, Laplacian, horizontal derivative and vertical derivative operators.The optimal decoded state sequence $Q = (q_1, q_2, \cdots, q_\tau)$ and the mixture indices $M = (m_1, m_2, \cdots, m_\tau)$ are obtained by the Viterbi decoding algorithm of P_2 using the model λ_{P_1}. Here τ is the number of image pixels. Then the pseudo-sketch $S_2 = [p_1, p_2, \cdots, p_\tau]$ is generated by the decoded sequence using λ_{S_1}. This reconstructing process is summarized as follows. Each pixel of the reconstructed image p_k $(k = 1, 2, \cdots, \tau)$ corresponds to a state index q_k in Q and a mixture index m_k in M which has a Gaussian PDF in λ_{S_1}. Therefore, the feature vector of p_k is equal to the mean vector of this Gaussian distribution, and the gray level of p_k is equal to the first value in the feature vector.

3. THE SKETCH SYNTHESIS ALGORITHM BASED ON E-HMM AND SELECTIVE ENSEMBLE

In order to improve the generalization ability of the proposed sketch synthesis method, the selective ensemble [8] that many could be better than all is introduced as a boosting strategy in this section.

Given a candidate photo P' and a set of photo-sketch pairs $\{(P_1, S_1), (P_2, S_2), \cdots, (P_N, S_N)\}$ for model training, we can establish a model λ_{p_i} for each photo-sketch pair (P_i, S_i) ($i = 1, 2, \cdots, N$), then estimate the similarities $\{P(O_{p'}|\lambda_{p_1}), P(O_{p'}|\lambda_{p_2}), \cdots, P(O_{p'}|\lambda_{p_N})\}$ between the observation sequence $O_{p'}$ of P' and the models $\{\lambda_{p_1}, \lambda_{p_2}, \cdots, \lambda_{p_N}\}$ of all photo-sketch pairs in the training set. Then, choose n models with bigger similarities to generate n pseudo-sketches through the joint-training algorithm presented in Section 2.2. Finally, a better pseudo-sketch is synthesized by fusing n pseudo-sketches with weights which are determined by Eq.(5).

$$w_i = \frac{P(O_p|\lambda_{p_i})}{\sum_{j=1}^{n} P(O_p|\lambda_{p_j})}, \quad i = 1, 2, \cdots, n \tag{5}$$

where $\sum_{i=1}^{n} w_i = 1$.

Fig.3 shows the framework of the proposed selective ensemble of sketch synthesis algorithm.

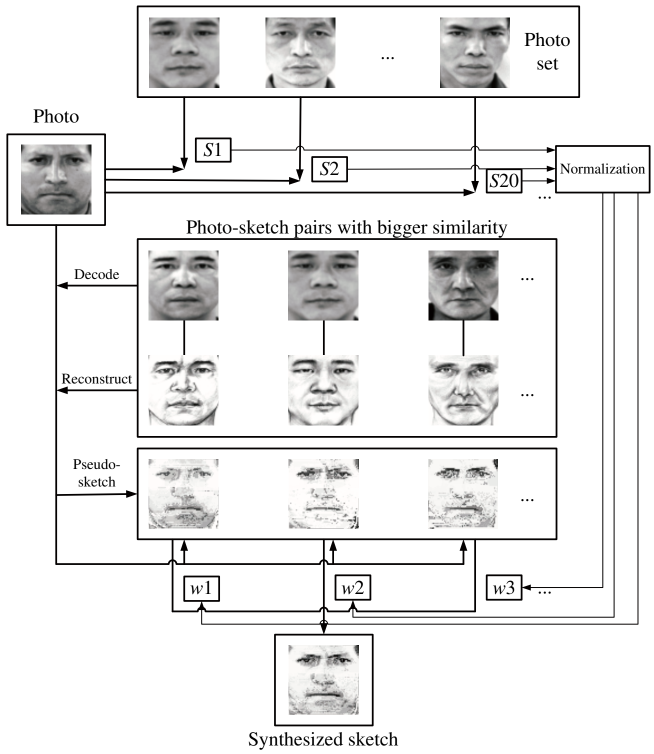

Fig. 3. The framework of the sketch synthesis algorithm based on selective ensemble.

4. EXPERIMENTS AND PERFORMANCE EVALUATION

In order to validate the effectiveness of the proposed sketch synthesis algorithm, two groups of experiments are conducted to compare with the nonlinear approach in [4]. The database containing several pairs of face photos and sketches. Three kinds of pre-processing methods that affine transform, geometric and grayscale normalization, are done firstly. We adopt leave-one-out strategy to leave one photo out as the test sample while the rest photo-sketch pairs in the database act as training samples. The universal image quality index (UIQI) [9] is adopted as an objective standard for sketch quality evaluation. The larger the Q value, the higher the quality of pseudo-sketch.

4.1. Face sketch synthesis

From above analysis, it can be seen that the effect of pseudo-sketches mostly depends on four aspects, the number of models n, the number of Gaussian mixtures in E-HMM m, the state distribution and features used for training E-HMM. It is found that pseudo-sketches would loss some detailed information with too large n, while pseudo-sketches would have too much noise with too small n. As to m, the larger the m, the higher the ability of the E-HMM to describe face images, but the effect of the proposed face synthesis algorithm tends to be stable to some degree when m is big enough. We achieve satisfactory effect with proper n and m, as shown in Fig.4. Fig.4(b) presents the original face sketches drawn by the artist. The pseudo-sketches synthesized with the proposed method are shown in Fig.4(d), while Fig.4(c) illustrates the pseudo-sketches synthesized with the nonlinear approach in [4]. The Q values of the images in Fig.4(a),(c),(d) with respect to (b) are shown in Table 1. We can conclude that the performance of the proposed method is better than that of the nonlinear approach. It is obvious that pseudo-sketches generated by the proposed method are clearer, higher in quality and more similar to the original sketches.

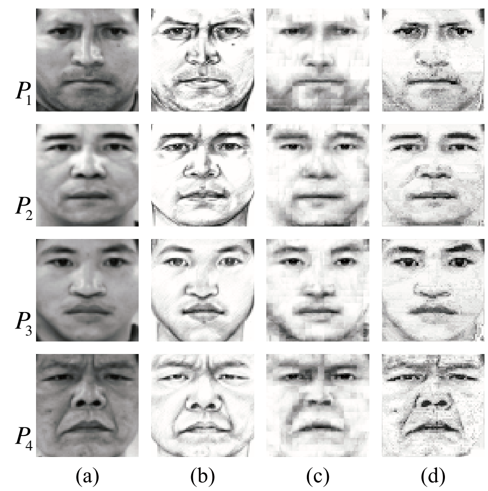

Fig. 4. Comparison of the pseudo-sketch synthesis results: (a) The photo images; (b) The sketches drawn by artists; (c) The pseudo-sketches generated by the nonlinear method; (d) The pseudo-sketches generated by the proposed method.

Table 1. Comparison of Q values of different images under original sketches.

Images	(a) Photos	(c) Pseudo-sketches with the nonlinear method	(d) Pseudo-sketches with the proposed method
P_1	0.5455	0.6346	0.6690
P_2	0.6592	0.7526	0.7632
P_3	0.6372	0.7458	0.7791
P_4	0.5189	0.6489	0.6720

4.2. Sketch-photo recognition

The purpose of sketch synthesis is to improve the recognition rate in sketch-photo retrieval, therefore we can assess the validity of sketch synthesis method with the recognition rate.

For all the photo-sketch pairs in the database, three sets of face image samples are obtained with different methods as follows. The first one is the original photo set; the second one is the pseudo-sketch set generated by the nonlinear approach in [4]; and the last one is the pseudo-sketch set synthesized by the proposed method. The recognition results of the three sets of training samples by Eigenface are shown in Table 2. For the poor matching between photos and sketches directly, the untransformed method only achieve the recognition rate of 19.05%. Obviously, as a result of the better quality of pseudo-sketches, the proposed sketch synthesis algorithm based on E-HMM and selective ensemble has the visible advantage of 20 percent higher recognition rate than the nonlinear approach. Furthermore, Fig.5 shows the variation of sketch-photo recognition rate using the proposed method with the number of models n. The fact that we get worst result when n=1, which means the face synthesis algorithm without selective ensemble is adopted, illustrates the effectiveness of selective ensemble. It is also found that the best performance is achieved when the number of model pairs equals 6 to 10, which is consistent with the idea of selective ensemble again.

Table 2. Comparison on different face sample sets.

Face sample set	Original photo set	Pseudo-sketch set	
		Nonlinear method	Proposed method
Average Q value	0.5311	0.6699	0.6756
Recognition rate	19.05%	71.43%	95.24%

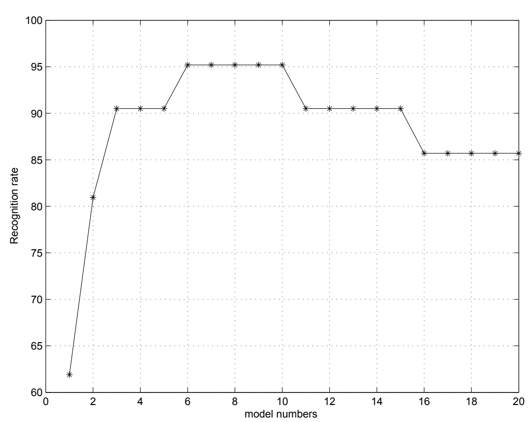

Fig. 5. Comparison of recognition rate with different model numbers.

5. CONCLUSIONS

A novel sketch synthesis algorithm based on E-HMM and selective ensemble is developed in this manuscript. By using E-HMMs for modeling the nonlinear relationships in sketch-photo pairs, a series of pseudo-sketches of the same photo are generated with the selective ensemble strategy which are then fused together for a finer face pseudo-sketch. The experimental results show the effectiveness of the proposed method. It can be used to set up an automatic face sketch recognition system, which can find various applications in the fields of counter-strike, safety guard, and image or video retrieval based on sketches.

Whereas, a clear disadvantage of the E-HMM is that it is hard to learn more complex nonlinear relationship. So, it is necessary to improve the structure of the E-HMM in future study.

6. ACKNOWLEDGMENTS

Thanks must be expressed to the Multimedia Lab of the Chinese University of Hong Kong for providing us with the face photo-sketch images database.

7. REFERENCES

[1] J. Wang, H.J. Bao, and W.H. Zhou, “Automatic image-based pencil sketch rendering,” *Journal of Computer Science and Technology*, vol. 17, no. 5, pp. 347-356, 2002.

[2] X.O. Tang and X.G. Wang, “Face photo recognition using sketch,” in *Proc. Int. Conf. Image Processing*, Sep. 2002, pp. 257-260.

[3] X.O. Tang and X.G. Wang, “Face sketch recognition,” *IEEE Trans. Circuit System and Video Technology*, vol. 14, no. 1, pp. 50-57, 2004.

[4] Q.S. Liu and X.O. Tang, “A nonlinear approach for face sketch synthesis and recognition,” in *Proc. Int. Conf. Computer Vision and Pattern Recognition*, Jun. 2005, pp. 1005-1010.

[5] A.V. Nefian and M.H. Hayes, “Face recognition using an embedded HMM,” in *Proc. Int. Conf. Audio Video Biometric based Person Authentication*, 1999, pp. 19-24.

[6] T. Nagai and T. Nguyent, “Appearance model based face-to-face transform,” in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing*, May 2004, pp. 749-752.

[7] T. Nagai, T. Naruse, M. Ikehara, and A. Kurematsu, “HMM-Based surface reconstruction from single images,” in *Proc. Int. Conf. Image Processing*, Sep. 2002, pp. 561-564.

[8] Z.H. Zhou, J. Wu, and W. Tang, “Ensemble neural networks: Many could be better than all,” *Artificial Intelligence*, vol. 137, no. 1-2, pp. 239-263, 2002.

[9] Z. Wang and A.C. Bovik, “A universal image quality index,” *Signal Processing Lett.*, vol. 9, no. 3, pp. 81-84, 2002.