Face Sketch Synthesis System Utilizing Both Global Structure and Local Detailed Texture Components

Ching-Ting Tu and Yu-Hsien Chan

Tamkang University, New Taipei City 25137, Taiwan, R.O.C.

ctu@mial.tku.edu.tw and worstfish61@gmail.com

Abstract

This study developed a facial sketch synthesis system based on the two-dimensional direct combined model (2DDCM) approach that employs a collection of photo/sketch pairwise training samples. The proposed synthesis framework addresses the following key issues. First, we directly combine each photo and sketch pairwise sample in a concatenated form in order to completely preserve their relationship. Second, photo and sketch images are formed as two-dimensional matrices instead of vectors in order to preserve the facial geometry. Third, both the global facial geometry and the local detailed textures are included in the proposed synthesis framework. Experiments demonstrate our approach can synthesize high quality reconstructed facial sketches from given unseen photos.

Keywords: two-dimensional direct combined model; facial sketch synthesis; eigenspace.

1. Introduction

In recent years, facial sketch synthesis is usually applied in law enforcement and digital entertainment. For example, witness provides the appearance of the suspect, and the artist draws a corresponding sketch for visualization. Also, the sketch can also be used in digital entertainment, when we can use facial sketches to get others attention. There are two types of synthesis approaches to generate facial sketches: the image-based approach and the exemplar-based approach. Image-based methods [2, 5] generate special strokes by using image processing approaches. For example, framework in [2] generates

sketch by applying bilateral filter to the color distribution of input photo; Work [5] combines the properties of image tone and sketch stroke to synthesize facial sketches. However, the synthesized results produced by these kind of approaches are without particular drawing styles as the sketches drawn by the talented artist.

Exemplar-based methods, on the other hand, require a large number of training pairwise photo and sketch samples. These training samples are then used to learn how to generate (or exaggerate) a correct sketch for an unseen input photo. A typical kernel learning method is non-parametric-based Markov random field (MRF) [9, 11, 12]. For example, work [1] used the E-HMM approach to construct the relation between photo and sketch, where a human face is divided into five parts, namely, forehead, eye, nose, mouth, and chin, in order to preserve detailed textures. Work [11] used the MRF graph to establish the correlation between the photo and sketch and constrains the smoothness between the neighborhood patches of the synthesized sketch. Work [12] was also based on the MRF approach, but this framework solved the varied light and pose changing conditions by using difference-of-Gaussian (DoG) representations. Work [9] integrated the parametric-based approach in the MRF structure.

These exemplar-based frameworks provided high quality synthesized sketch, however, the synthesized result are sensitive to the number of training samples. For example, synthesized results don't contain the personal specific characteristics of the corresponding input images since these geometry characteristics are not provided in the training dataset. In this study, a facial sketch synthesis system is proposed based on the parametric-based 2DDCM approach [8]. Different from the MRF-based framework, facial sketches produced by the 2DDCM approach are the

Fig. 1. Proposed 2DDCM-based synthesis framework; the global facial sketch synthesis module and the local detailed texture synthesis module, and the enhancement module.

linear combination of the training set results. Consequently, the appearances of synthesized sketches are not limited by that of training samples. Furthermore, both facial structure and local detail textures are considered in the proposed framework.

2. FACIAL SKETCH SYNTHESIS FRAMEWORK

The proposed framework is illustrated in Figure 1, where it composes two modules, i.e., the global and local component synthesis modules. The first module learns the global facial geometry correlation between photo and sketch by using 2DDCM method [8] from N pairwise photo/sketch samples, $\{(P_i, S_i)\}_{i=1}^N$. The significant features of their correlation are extracted by the 2DDCM models, U^g , where its 2DDCM transformation from photo to sketch, T_{PS}^g , is then constructed. Note that, such global 2DDCM model captures the global facial geometry, but less address the local detailed textures.

In the second module, the 2DDCM approach is applied in the patch-based manner. The goal of this module is to estimate the high frequency components lost in the global module. In the synthesis process, the photo residual of the current test P_t , P_t^l , is reconstructed first. P_t^l is the photo high frequency component lost in the global module, which is then used to predicted the sketch lost, S_t^l , by applying the local patch-based 2DDCM approach. To be specified, P_t^l is decomposed as several overlapped patches, each patch searches the corresponding K-NN pairwise patch in the training dataset, $\{(P_i^l, S_i^l)\}_{i=1}^N$. These

similar pairwise samples are used to construct the local 2DDCM models, U_p^l , to infer the corresponding local patch in S_t^l . Combine S_t^l with the global synthesized result S_t^g , the system output is S_t^* .

2.1 Global Facial Component Synthesis

We adopt the 2DDCM approach [8] for learning the correlation between two related but different classes, i.e. the class of facial photo and the class of facial sketch. It is because the 2DDCM approach has the ability to synthesize sketches through a geometry-preserved representation. Given a training dataset of N photo and sketch pairwise samples $\{(P_i, S_i)\}_{i=1}^N$, where each pairwise sample, (P_i, S_i) , is represented as a two-dimensional combined matrix form A_i :

$$A_i = \begin{bmatrix} A_i^p \\ A_i^s \end{bmatrix} = \begin{bmatrix} P_i & S_i \\ S_i & S_i \end{bmatrix}, \quad (1)$$

where $P_i \in \mathbb{R}^{m \times n}$ and $S_i \in \mathbb{R}^{m \times n}$ (m is the image height and n is the image width), and thus $A_i \in \mathbb{R}^{2m \times 2n}$.

Accordingly, the covariance matrices of $\{A_i\}_{i=1}^N$ can be calculated as:

$$C_U = \sum_{i=1}^N \frac{1}{N} (A_i - \bar{A})^T U U^T (A_i - \bar{A}) \quad (2)$$

, where $\bar{A} = [(\bar{A}^p)^T (\bar{A}^s)^T]^T$ is the mean matrix of the training samples; (2) is the covariance of all column patterns projected onto a subspace of U . The objective function of 2DDCM approach is to find the subspace U that

Fig. 2. Residual sketches obtained using the difference between the ground truth sketches and the global synthesized results.

Fig. 3. Illustration of generation process for local component of image, I^l , and local component of sketch, S^l , using 2DDCM-based bidirectional transformations, i.e. T_{PS}^g and T_{SP}^g .

maximizes C_U , where U can be solved by applying singular value decomposition (SVD) process on C_U (i.e., as performed in [8], albeit using the vertical column subspace only, for the sketch synthesis case).

The role of $U = [U_S^T \ U_P^T]^T$ is the significant pairwise features. Based on such basis (subspace U), we have the 2DDCM transformation from photo to sketch:

$$S_U = \bar{A}_S + U_S (U_P^T U_P + \lambda_U \Lambda_U^{-1})^{-1} U_P^T (P - \bar{A}_P) \quad (3)$$

, where $\lambda_U = 0.1$ is a user-predefined parameter; Λ_U is diagonal matrix of eigenvalues of U . Here, we denote such transformation as T_{PS}^g , which means the 2DDCM transformation from photo to sketch processed at the face level for the global appearance transformation.

2.2 Local Facial Component Synthesis Module

Fig. 4. Residual sketches generation based on the 2DDCM approach by using different number of K in the proposed local model. **Upper:** $K=30$; **Lower:** $K=300$.

In Fig. 2, four example subjects are provided to illustrate the differences between the global synthesized result and the corresponding ground-truth sketch. Such differences are the local detailed textures, i.e. the high frequency component, that lost in the global 2DDCM module.

These high-frequency components are then estimated according to the following observations; first of all, since the global 2DDCM model U^g directly concatenates photo and sketch in a combined matrix in (1), the synthesized global photo P^g of the input photo P produced by the 2DDCM bi-transformation (i.e. the transformation from photo to sketch, T_{PS}^g , followed by its inverse transformation from sketch to the photo, T_{SP}^g) should be equal to the input photo P , in the ideal case. However, there is a difference between P and P^g , such difference is denoted as the photo residual, P^l . The sketch residual, S^l , can be generated using a similar approach. The N training pairwise local component samples, $\{P_i^l, S_i^l\}_{i=1}^N$, are produced by the 2DDCM bi-transformation as the process illustrated in Fig. 3, which are then used for learning the local component correlation between photo and sketch.

Different from the synthesizing process of the global module that the 2DDCM transformation is performed at the face level. The estimation process of S^l is performed in local patch-based manner. First of all, the photo residual of the system input P_i^l are decomposed into NP patches to search the corresponding K -NN pairwise patch samples from the training dataset, $\{P_i^l, S_i^l\}_{i=1}^N$, for modeling the local residual correlation between photo and sketch. For example, for the p 'th patch $P_{i,p}^l$, its local 2DDCM transformation $T_{PS,p}^l$ is trained by these K similar training samples. The

Fig. 5. Example synthesized results of female subjects produced by the proposed framework. (a) Input photos; (b) ground-truth sketches; (c) synthesized result obtained from the global module; (d) synthesized result obtained from the proposed all two modules.

Fig. 6. Example synthesized results of male subjects produced by the proposed framework. (a) Input photos; (b) ground-truth sketches; (c) synthesized result obtained from the global module; (d) synthesized result obtained from the proposed all two modules.

corresponding sketch residual patch $S_{t,p}^l$ is then synthesized by applying $T_{PS,p}^l$.

Note that, in order to save the computation time, the searching process is done within a searching window size at the correspond position of the p 'th patch, where each training sample may provide more than one candidates for

training the transformation $T_{PS,p}^l$.

Fig. 4 is the synthesized result of S_t^l by using two different K -values for the local 2DDCM transformation learning process. It can be observed that, when $K=300$, the synthesized results are blurred, but preserve the correct local geometry. On the other hand, for the case of $K=30$, the synthesized results contain more detailed textures, but along with the noise, which makes unexpected lines or spots. In this study, we set the value of K as 80.

3. Experiment Results

We validated our method using the CUHK database; there are 188 photo-sketch pairwise samples, where 88 samples are selected for the training purpose while the rest 100 subjects are used for the testing purpose. Example synthesized results obtained from the proposed framework are shown in Fig 5 and Fig. 6. It can be seen that the synthesized results obtained using the global module preserve the facial geometry (Figs. 5(c) and 6(c)). Figures 5(d) and 6(d) show the synthesized results obtained from the local module (i.e., the global module followed by the local module). Compared with the synthesis results obtained from the global module only, those obtained from the global and local modules contain more distinct sketch lines and curved strokes.

Figure 7 shows the synthesis results obtained using various existing kernel methods, namely the Eigen-transformation method in [6], the one-dimensional canonical correlation analysis (1DCCA) method in [10], the two-dimensional CCA (2DCCA) method in [3], the direct combined model (1DDCM) in [7], and the two-dimensional DCM method (2DDCM) in [8]. It is observed that the synthesized results obtained using the 2DDCM approach and 2DCCA approach are qualitatively better than the others. Moreover, the images synthesized using the 2DDCM method (used in the present study) are clearer than those obtained using the 2DCCA approach. Notably, the synthesized results obtained using the 1D-based methods are less recognizable as the input image since the subject-specific geometry details are lost in the synthesis process.

Figure 8 shows the synthesis results obtained from the proposed framework for the non-profile images considered in [13]. Note that the framework was once again trained using the CUHK dataset in the first experimental setting case. As shown, the sketches synthesized using the global module contain obvious facial traces in the central image

Fig. 7. Synthesized results obtained using various exemplar-based methods. (a) input photos; (b) ground truth sketches; (c) canonical correlation analysis (1DCCA) [10]; (d) direct combined model (1DDCM) [7]; (e) two-dimensional CCA (2DCCA) [3]; (f) two-dimensional DCM (2DDCM) [8]; (g) two-step based approach [4]; (h) Eigen-transformation method [6].

Fig. 8. Synthesized results obtained for images used in Ref [13] when using CUHK training dataset in the first case for training purposes. First row: input images; second row: synthesized results from global module, and the third row: synthesized results obtained by the local 2DDCM in the local module.

region (see the images presented in the second row of the figure). The 2DDCM stage of the local module eases this problem and generates more clear and distinct sketch textures (see the third row in the figure).

4. Conclusion

We have developed a new facial sketch synthesis framework based on a training dataset. This work was motivated by the inefficiencies and unsatisfactory results of

representing two related images by two individual 1D vectors. Especially, both the global and local components are included in the final synthesized results. Compared with existing approaches, the proposed approach that embeds the facial geometry patterns into the 2D combination subspace improves the facial sketch synthesis technology.

Acknowledgment

C.-T. Tu's research is sponsored by MOST Grant

References

- (1) X. Gao, J. Zhong, J. Li and C. Tian, "Face sketch synthesis using E-HMM and selective ensemble," *IEEE Trans. Circuits Syst. Video Technol.*, 2008, vol.18, pp.487-496.
- (2) E. S. L. Gastal and M. M. Oliveira, "Domain transform for edge-aware image and video processing," *ACM Trans. Graph.*, 2011, vol.30, no.4, pp.69:1-69:12.
- (3) L. S. Ho and C. Seungjin, "Two-dimensional canonical correlation analysis," *IEEE Transactions on Signal Processing Letters*, 2007, pp.735-738.
- (4) C. Liu, H. Shum and C. Zhang, "A two-step approach to hallucinating faces: global parametric model and local non-parametric model," *Proc. Int'l Conf. Computer Vision (ICCV)*, 2001, vol.1, pp.192-198.
- (5) C. Lu, L. Xu, and J. Jia, "Combining sketch and tone for pencil drawing production," in *Proceedings of the Symposium on Non-Photorealistic Animation and Rendering (NPAR '12)*, 2012, pp.65-73.
- (6) X. Tang and X. Wang, "Face sketch synthesis and recognition," *Proc. IEEE Int. Conf. Comput. Vis.*, 2003, pp.687-694.
- (7) C. T. Tu and J. J. Lien, "Automatic location of facial feature points and synthesis of facial sketches using direct combined model," *IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2010, Vol.40, No.4, pp. 1158-1169.
- (8) C. T. Tu, K. H. Lin, and T. L. Liu, "A two-dimensional direct combined model for facial hallucination," *IAPR International Conference on Machine Vision Application*, 2013, pp.149-152.
- (9) X. Wang and X. Tang, "Face photo-sketch synthesis and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2009, vol.31, no.11, pp.1955-1967.
- (10) W. Yang, D. Yi, Z. Lei, J. Sang, and S. Li, "2D-3D face matching using CCA," *IEEE Conference on Automatic Face Gesture Recognition*, 2008, pp.1-6.
- (11) H. Zhou, Z. Kuang and K. Wong, "Markov weight fields for face sketch synthesis," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp.1091-1097.
- (12) W. Zhang, X. Wang, and X. Tang, "Lighting and pose robust face sketch synthesis," *ECCV*, 2010, pp.420-433.
- (13) S. Zhang, X. Gao, N. Wang, J. Li, and M. Zhang, "Face

sketch synthesis via sparse representation-based greedy search," *IEEE Transactions on Image Processing*, 2015, vol. 24, no. 8, pp. 2466-2477.