Synthesis of Facial Images Based on Relevance Feedback

Caie Xu
University of Yamanashi
Yamanashi Japan
g17dhl01@yamanashi.ac.jp

Shota Fushimi
University of Yamanashi
Yamanashi Japan

Masahiro Toyoura
University of Yamanashi
Yamanashi Japan
mtoyoura@yamanashi.ac.jp

Jiayi Xu
Hangzhou Dianzi University
Hangzhou China
xujiayi@hdu.edu.cn

Honglin Li
University of Yamanashi
Yamanashi Japan
lihonglin79@qq.com

Xiaoyang Mao
University of Yamanashi
Yamanashi Japan
mao@yamanashi.ac.jp

***A Abstract*—We propose a dialogic system based on a relevance feedback strategy that allows for the semiautomatic synthesis of a facial image that only exists in a user's mind. The user is presented with several facial images and judges whether each one resembles the face that he or she is imagining. Based on the feedback from the user, a set of sample facial images are used to train an Optimum-Path Forest classifying the relevance of facial images. An interpolation method is then employed to synthesize new facial images that closely resemble the imagined face. A series of experiments are conducted to evaluate and verify the effectiveness and efficiency of the proposed technique.**

Keywords—face image synthesis; relevance feedback; optimum-path Forest

I. INTRODUCTION

Face recognition technique benefits a wide range of applications, especially in assisting law enforcement [1]. For instance, automatic face retrieval from police dataset can help the police seek and identify suspects quickly. However, in many situations, the face images of suspects are not available. If there were a technique for computer to create physical images of the faces that people picture in their minds, it would be possible to develop more accurate caricatures of criminals.

Our study proposes a semiautomatic method of synthesizing facial image in a user's mind through relevance feedback (a form of dialogic searching). A set of sample facial images are used to train an Optimum-Path Forest (OPF) classifying the relevance of facial images based on users' feedbacks. Based on the trained OPF, new facial images that closely resemble the imagined facial image are created.

The remainder of the paper is organized as follows: Section II reviews the related works. Section III presents the details of the proposed method. Section IV describes the preparing of the training dataset. Section V demonstrates some results and describes the evaluation experiment.

II. RELATED WORK

While face recognition has been one of the most active research field of computer vision, there are few studies have been conducted on the synthesis of facial images.

W. Di and D. Qionghai [2] and N. Wang *et al.* [3][4] proposed methods for synthesizing facial images based on sketches. The way introduced by W. Di, and D. Qionghai searches a facial image database for the parts with the highest degree of resemblance and then patches the results together to form a facial image [2]. N. Wang *et al.* presented Bayesian face sketch synthesis methods [3][4], which uses the spatial neighboring constraint in both the neighbor selection and weight computation. However, both methods need to draw sketch that not everyone is blessed with this gift.

Face hallucination is the technique for synthesizing high resolution facial images from low resolution ones. The method by C. Liu *et al.* [5] assumes that low resolution images still manage to reveal general facial characteristics and uses a high resolution sample database for obtaining the information required for the synthesis. Recently, advanced approach using deep learning technology has been proposed for hallucinating faces of unconstrained poses [6]. While hallucination method requires low resolution images to infer the information for reconstructing the high resolution images, our technique uses relevance feedback to acquire the information required for synthesizing the face a user images.

III. PROPOSED METHOD

As Figure 1 shows, extracting primary feature, training OPF based on relevance feedback and synthesizing facial images are three major components of the system.

In our study, we used 1,000 images for building the training database. These images are converted to feature space for training an OPF classifier. The ultimate purpose of our method is not to classify those sample facial images or to retrieve a particular face from database, but to synthesize a new image resembling a face in the user's mind. The trained OPF is used to define the positions in feature space corresponding to the desired facial images.

To train the OPF, the system defines an initial classification boundary by letting the user to evaluate an initial dataset consisting of facial images of different sex and ages. Then, the system shows the user multiple unevaluated images (have not been judged) that lie near the classification boundary and has the user label them according to whether they resemble or do not resemble the face in his or her mind. Based on these feedbacks, the system updates the classification boundary.

Then, the system finds a position farthest from the classification boundary on the positive side, and produces the synthesized image. If the result satisfies the user, the whole process is complete; otherwise, the user repeats the labeling process on unlabeled cases near the classification boundary.

A. Constructing feature space

Various feature representations have been studied in the context of face recognition in the past few decades. Recent research results have demonstrated that deep learning can be used to learn the facial representation which is effective for both face identification and verification [7]. However, since our purpose is to synthesize a target facial image, we need a feature representation not only can discriminate faces but from which a facial image can be generated. For this purpose, we use the pixel level image feature used in Face Hallucination method [5].

The first step of the Face Hallucination method is to prepare a database of high resolution facial images. The method then separates high resolution facial image I_H into a global face image I_H^g, which expresses the overall feature of the image, and local face image I_H^l.

- $$I_H = I_H^g + I_H^l \qquad (1)$$

Face Hallucination creates a global face image feature vector space by applying Principal Component Analysis (PCA) to the facial images in the database and finding the principal components with large eigenvalues. Formula (2) expresses global face image I_H^g in terms of global face feature space basis B, coordinate vector X, and average facial image μ

- $$I_H^g = BX + \mu \qquad (2)$$

Our study uses the global feature space as the search space for facial image synthesis. We apply PCA to 1000 sample facial images to define the average facial image μ and basis B of global face feature space. Given a coordinate vector X, a facial image can be generated with equation (2). Our method use trained OPF to locate the coordinates of the image that best matches the corresponding face in mind.

B. Training Optimum Path Forest based on relevance feedback

Relevance feedback, a process that shows synthesis results to the users and updates classifiers based on users' feedback, is often used in image retrieval. Several researchers have proposed methods that employ various classifier types and reuse past classification result to obtain good results based on relatively minimal amounts of feedback [8].

Our study used Optimum Path Forest (OPF) [9,10,11] for classification. OPF works by modeling the classification as a graph partition in a given feature space. It starts as a complete graph, whose nodes represent the feature vectors of all images in the database. All pairs of nodes are linked by arcs which are weighted by the distances between the feature vectors of the corresponding nodes (referred as costs here and after). Given a set of training nodes, a minimum spanning tree (MST) can be generated from the complete graph. Then the adjacent training nodes are marked as prototypes if they belong to different classes, which are relevant and irrelevant in our case. The partition of the graph is carried out by the competition process among prototypes, which offer optimum paths to the remaining nodes of graph. The optimum paths from the prototypes to the other samples are computed by the algorithm of the image foresting transform (IFT), which is essentially Dijkstra's algorithm modified for multiple sources and more general path-value functions. At last, all the non-prototypes are connected with a prototype directly or indirectly with the minimum costs. With the prototypes as the roots and the non-prototypes as the intermediate and terminal nodes, the optimum trees are built, which constitute the OPF. OPF is very important in a prototype based on RF approach that generates results in a dialogic fashion. It shows good performance for samples represented in complex and high dimensional feature space. As depicted by Figure 1, the OPF is trained based on users' relevance feedbacks in the following 4 steps:

Step 1: The system presents the user 5 male facial images and 5 female facial images of different ages and waits for the user to select one he/she thinks to be closest to the face in his/her mind. Since none of those 10 images may resemble the target face, the user is likely to select the image that similar to what they desire according to sex and age, which acts as the initial classification boundary.

Step 2: Four images closest to the user selected facial image in feature space are returned to the user. User starts to evaluate and labels the images with positive (○) or negative (X), which serve as the prototype for the OPF classifier. This evaluation phase ends up if the user is satisfied with at least one of the 4 facial images.

Step 3: An OPF classifier is built based on this set. Then, we use the OPF classifier to divide the unlabeled images of the database into two classes, relevant and irrelevant.

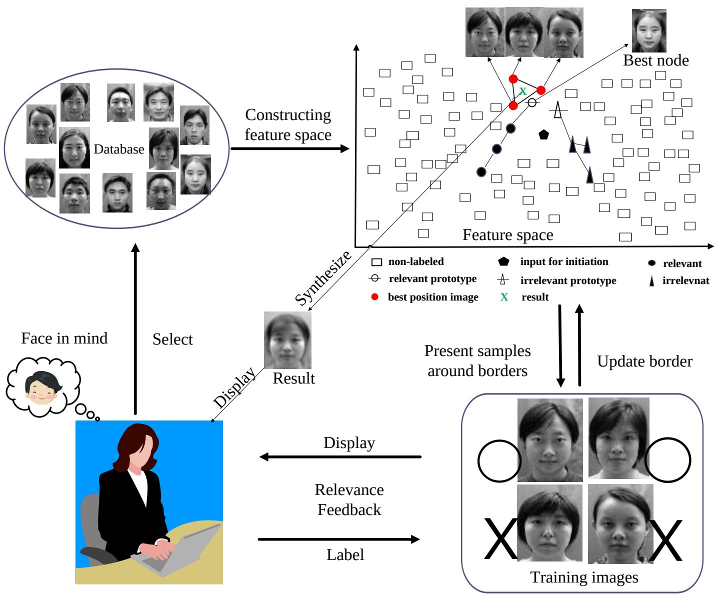


Figure 1: Overview of the proposed system

Step 4: Four border nodes are selected and the corresponding images are presented to the user. The user evaluates and labels the images with positive (○) or negative (X), and the new marked training images will constitute and replace the former training samples to rebuild new OPF classifier.

At every iteration before Step4, best positive nodes located far from the negative prototype and close to the positive prototype are selected and used to create the resulting facial image. If the user is satisfied with the created image, the whole relevance feedback procedure ends.

C. *Synthesizing virtual facial images using interpolation*

The traditional relevance feedback approach is for searching actual images in the given database. However, in our study, synthesizing non-existent facial images is the ultimate goal. In principle, any point near the best positive node should correspond to a desired facial image.

As a practical solution, we propose to select the top k (in this paper k=3) best positive nodes as shown in Figure 1 and calculate the result according to the following formula (3)

- $$x = \sum_{i}^{k} w(x_i) x_i \Big/ \sum_{i}^{k} w(x_i) \qquad (3)$$

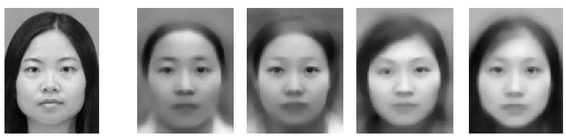
Here, x and x_i (i=0,1,2) are the coordinate vectors of the resulting facial images and the 3 best positive images, respectively. $w(x_i)$ is the weight assigned to x_i based on the distance given by the classifier. In this paper, the value of $w(x_i)$ is assigned using two kinds of weights which are average weight and disproportionately weight. The average weight means all 3 images with equal weight. The disproportionately weight is inversely proportional to its distance from each of the 3 images.

IV. Database

For our sample image set, we used 1,000 images of Asian faces from the CAS-PEAL database [12] and Cartoon Face database [13]. We made all the images monochrome, and set the size to 96 × 128. The database comprised only frontal facial images, but the positions and sizes of the faces differed. We aligned, resized, and trimmed the images so that the points between the eyes in the facial images would all line up. Our study is concerned only with general facial features, so we used a low resolution of 96 × 128 for all the images.

V. Experiment

A prototype system for synthesizing faces in mind has been built based on our proposed method and experiments were conducted to validate the effectiveness of the proposed method. We recruit 10 volunteers (7 male and 3 female university students in their 20s). The subjects are asked to perform two different kinds of task. In the first task, the subject is given a facial image as a reference and asked to generate a facial image similar to the reference image. In the second task, the subjects are asked to imagine a face and create a facial image similar to the one in his/her mind.



(a)Reference (b) 1st (c) 4th (d) 10th (e) 14th

Figure 2: Results at nth iteration during the relevance feedback procedure

A. *Task 1：Synthesizing a facial image by referring to a given facial image*

Figure 2 shows the synthesized results during the relevance feedback process. 2(b) ~ 2(e) are the output image after first, fourth, tenth, fourteenth iterations of our relevance feedback algorithm. We can see the resulting facial image becomes more resemblance to the reference image gradually when the number of iteration increases.

In Figure 3, the final results (user satisfied faces) for 2 reference facial images are shown. Synthesized image 1 has the similar face shape as the reference face and presents a similar side hairstyle, although the eyes opened wider than the reference image. Synthesized image 2 presents similar hairstyle as reference image. However, it has wider foreheads with less frontal hair then the reference image and has a mole at the lower left of the nose. Such detailed information cannot be synthesized with the proposed method. Manual attachment of the mole can easily improve the quality of synthesized image.

(a)Reference image 1 (b)Synthesized image 1 (c)Reference image 2 (d)Synthesized image 2

Figure 3: Results using reference images

B. *Task 2: Subjective grade of synthesized faces*

In this experiment, we ask subjects to imagine a face, and let them to use the system to create the images accordingly. Then, the subjects are asked to grade the similarity of the resulting images to their imagined images with a five-point system (1: No resemblance; 2: very weak resemblance; 3: Neither weak nor strong resemblance; 4: Somewhat strong resemblance; 5: strong resemblance).

Figure 4 shows the final scores of this evaluation. The average final score was 4.0. Figure 5 shows the number of iterations it took to reach the final result and the corresponding number of subjects, with the number of iterations on the x-axis and the number of subjects who reached their final results at that number of iterations on the y-axis. On average, it took 10.9 iterations to reach the final results. Our method thus earned rating of 4 or higher from the majority of experiment participants and produced final results in a feasible number of iterations.

VI. CONCLUSION AND FUTURE WORK

In this paper, we propose a method for the semiautomatic synthesis of a facial image in a user's mind. By training an OPF based on the user's feedback, we successfully create synthesized images that resembled the facial images that users had in mind. Currently, the face image is aligned only by eye distance and caused the blurring. In the future, we will design more robust alignment algorithm to replace current approach. Also, we will conduct more experiments to verify the system performance.

ACKNOWLEDGEMENT

This work was supported by JSPS KAKENHI Grant Number 17H00737 and the Public Projects of Science Technology Department of Zhejiang Province, China (No. 2016C33G2040006/2016C33012).

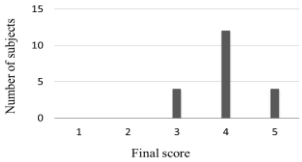


Figure 4: Scores of final results

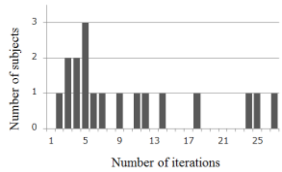


Figure 5: Number of iterations required for achieving satisfied result

REFERENCES

[1] X. Wang, and X. Tang, "Face Photo-Sketch Synthesis and Recognition," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 31, no. 11, pp. 1955-1967, 2009.

[2] D. Wu, and Q Dai, "Sketch Realizing: Lifelike Portrait Synthesis from Sketch," Computer Graphics International Conference, pp. 13-20, 2009.

[3] Q. Liu, X. Tang, H. Jin, H. Lu and S. Ma, "A nonlinear approach for face sketch synthesis and recognition," IEEE Computer Society Conference, vol. 1, pp. 1005-1010. 2005.

[4] N. Wang, X. Gao, L. Sun and J. Li, "Bayesian Face Sketch Synthesis," IEEE Transactions on Image Processing, vol. 26, no 3, pp. 1264-1274, 2017.

[5] C. Liu, H. Y. Shum, W. T. Freeman, "Face Hallucination: Theory and Practice," International Journal of Computer Vision, vol. 75, no. 1, pp. 115-134, 2007.

[6] S. Zhu, S. Liu, C. Loy, and X. Tang, "Deep Cascaded Bi-Network for Face Hallucination," European Conference on Computer Vision, pp. 614-630, 2016.

[7] Y. Sun, Y. Chen, X. Wang and X. Tang, "Deep learning face representation by joint identification-verification," Advances in Neural Information Processing Systems, pp. 1988-1996, 2014.

[8] I. Ruthven, M. Lalmas, "A survey on the use of relevance feedback for information access systems," The Knowledge Engineering Review, vol. 18, no. 2, pp. 95–145, 2003.

[9] A. T. Silva, A. X. Falcao, L. P. Magalhaes, "Active learning paradigms for CRIR systems based on optimum-path forest classification," Journal of WSCG, vol.18, no. 1-3, pp. 73-80, 2010.

[10] J. P. Papa and A. X. Falcao, "Optimum-path forest: a novel and powerful framework for supervised graph-based pattern recognition techniques," Institute of Computing University of Campinas, pp. 41-48, 2010.

[11] H. L. Li, M. Toyoura, K. Shimizu, W. Yang and X. Y. Mao, "Retrieval of clothing images based on relevance feedback with focus on collar designs," Visual Computer, vol. 32, no. 10, pp. 1351-1363, 2016.

[12] H. L. Li, G.H. Liu, and K. N. Ngan, "Guided Face Cartoon Synthesis," IEEE Transactions on Multimedia, vol. 13, no. 6, pp. 1230-1239, 2001.

[13] W. Gao, B. Cao, S. G.Shan, X. L. Chen, D. L. Zhou, X. H. Zhang and D Zhao, "The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations," IEEE Transactions on Systems, Man, and Cybernetics (Part A), vol. 38, no. 1, pp. 149-161, 2008.