Synthesising Images of Imagined Faces Based on Relevance Feedback

Caie Xu[1], Shota Fushimi[1], Masahiro Toyoura[1], Jiayi Xu[2]and Xiaoyang Mao [1]

[1]University of Yamanashi, Yamanashi, Japan
[2]HangzhouDianzi University, Hangzhou, China
mao@yamanashi.ac.jp

Abstract. In this paper, we propose a user-friendly system that can create a facial image from a corresponding image in the user's mind. Unlike most of the existing methods, which require a sketch as input or the tedious work of selecting similar facial components from an example database, our method can synthesise a satisfying result without questioning the user on the explicit features of the face in his or her mind. Through a dialogic approach based on a relevance feedback strategy to translate facial features into input, the user only needs to look at several candidate face images and judge whether each image resembles the face that he or she is imagining. A set of sample face images that are based on users' feedbacks are used to dynamically train an Optimum-Path Forest algorithm to classify the relevance of face images. Based on the trained Optimum-Path Forest classifier, candidate face images that best reflect the user's feedback are retrieved and interpolated to synthesise new face images that are similar to those the user had imagined. The experimental results show that the proposed technique succeeded in generating images resembling a face a user had imagined or memorised.

Keywords: Face image synthesis; Relevance feedback; Optimum-Path Forest.

1 Introduction

Face image synthesis has potential applications in public safety, such as video surveillance and law enforcement. For example, creating a portrait of a suspect from an eyewitness can greatly help the police identify criminals. Also, a similar technique can be used for giving concrete form to imagined ideas of romantic 'types' and translate other imagined faces into explicit images. However, drawing an image based on descriptions of what is in one's mind is not an easy task for the majority of people. Although the montage approach to face image synthesis [1] allows users to create face images by selecting face components, it involves the time-consuming task of choosing the right parts from a wide array of options. It is known that the composition of face parts is a more important factor in the perception of a face than the individual parts [2]. However, it can be very difficult to adjust the positions of individual parts to achieve a desired composition. Several methods have been developed for synthesising

face images according to sketches [3]. Such methods, however, require the user to provide a sketch, which is not always a possibility.

Motivated by the above mentioned potential applications and the limitation of current face image synthesis technologies, we aim to develop a novel system that can generate an image of a face from a user's imagination and memory through some simple user interactions. In the proposed system, a set of example images are used to train an Optimum-Path Forest (OPF) algorithm to classify the face images based on their relevance to the face in the user's mind. We favour OPF over other classification algorithms in its fast, simple, multi-class, parameter independent, and not making any assumption about the shapes of the classes [16] .The training process is conducted through a relevance feedback approach. All the user must do under this method is to indicate whether the image of the face shown bears a general resemblance to the face that he or she is imagining, thereby eliminating the need to evaluate individual parts and features separately (as is the case with the montage approach) or visualise or verbalise specific characteristics (as is the case with caricatures).

The remainder of the paper is arranged as follows. Section 2 reviews the related works. Section 3 describes the algorithm in detail. In Section 4, the experimental results are demonstrated and discussed.

2 Related Works

Although face recognition is one of the most active research fields in computer vision, to the best of the authors' knowledge, there are few studies that have been conducted on the synthesis of face images. With the montage approach to face image synthesis [2], the user looks through a database of various face components (e.g., eyebrows, eyes, noses, mouths, etc.) for the ones that most closely match the image in his or her imagination or memory. The selected features are then synthesised into a face image. The process requires the user to search for each part separately and make isolated judgments on resemblance; the user looks only at the eyes when looking for the eyes that approximate those of the face in his or her mind. Again, this is another challenging task. Finding the ideal combination of parts can take a considerable amount of time, as well. E-FIT [1], a montage synthesis system that facilitates the creation of 3-D, computer generated) faces, narrows down the search range by age and sex and lets the user make post-synthesis tweaks to facial feature sizes and positions to make the final face models more accurate. However, the effectiveness of E-FIT in generating a face model also depends on the user's past experience with modelling and sensitivity to various face features.

Wu and Dai [3] present method for synthesising face images according to sketches. By querying a face image database using different parts of a face sketch, the corresponding face parts with the highest degrees of resemblance are patched together to form a final image. Users can adjust the size, shape and colour of face parts to make the resulting face accurate. However, these methods require the user to draw a sketch, a talent that not everyone has.

Kurt et al. [4] proposed a semiautomatic method that uses a genetic algorithm to update feature parameters to synthesize a face image. Since their technique uses the AAM (Active Appearance Model) [4] for modelling and synthesizing face images, facial features in areas with no high frequency information cannot be captured.

The face hallucination technique [5, 6] uses information from a face image database to synthesise high-resolution images from low-resolution images. One potential application of this method is to synthesize high-resolution images from the grainy, low-resolution images captured by surveillance cameras. The image database is used to compute probable high-resolution features from the low-resolution images. Most recently, deep learning based techniques have been combined with face hallucination [7], making it possible to generate high-resolution images from images with very low-resolution and unconstrained pose.

However, prior methods based on feature mapping and deep learning could not be employed to estimate facial features in the absence of a reference image. For example, sketch based method heavily relies on a sketch face image; face hallucination method requires a low-resolution image as the input. Our method can generate face images that are satisfactory to the user demands without needing to seek clues from a reference image.

Our system uses an active learning scheme to narrow the gap between low-level image features and high-level semantic understanding. Recently, active learning algorithms combining conventional machine learning techniques with relevance feedback have been attracting large attentions. For example, in Content-Based Image Retrieval (CBIR) systems, Support Vector Machine (SVM) based active learning schemes are used for efficient image data clustering. Liu *et.al* [8] presented a SVM based relevance feedback technique for image retrieval on small database. Wang *et.al* [9] combined a few one-class SVM classifiers to boost the retrieval performance. Wang *et.al* [10] introduced a Neural Network (NN) based method for CBIR and evaluated their algorithm on a database of 2,000 images. However, with the growing sample data, SVM [16] and NN algorithm become less efficient than Random Forest (RF) and Optimum-Path Forest methods (OPF) [18] in handling multi-class classification. Fu and Qiu [11] developed a RF based image retrieval framework and examined their system in image-based and keyword-based image retrieval scenarios. The RF was generated based on semantic similarity measure. Although RF runs quite efficient, if the sample distribution is uneven, the classification result is unreliable. Based on these observations, we employ an OPF based relevance feedback technique.

3 Proposed Method

As depicted in Figure 1, the proposed system includes three major components: extracting primary features, training an OPF classifier based on relevance feedback and synthesising face images that do not already exist in the database.

In our study, we used 1,000 sample images in the training database. These images were converted to a feature space for training an OPF algorithm to classify whether a face image resembles the face in users' minds based on their relevance feedback. The

ultimate purpose of our method is not to classify those sample face images or to retrieve a particular face from these sample face images but to synthesise a new image resembling the face in the user's mind. The trained OPF classifier defines the positions in the feature space that correspond to the desired face images.

To train the OPF classifier, the system defines an initial classification boundary by letting users evaluate an initial dataset consisting of face images of different sexes and ages. Then, the system shows the user multiple unevaluated images (i.e., cases that have not been judged by the user to resemble or not resemble the picture in his or her mind) 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 labels, the system updates the classification boundary.

Then, the system interpolates *K* cases in the positions farthest from the classification boundary on the positive side and produces the final synthesis. If the results satisfy the user, the search process is complete; otherwise, the user repeats the labelling process on unlabelled cases near the classification boundary.

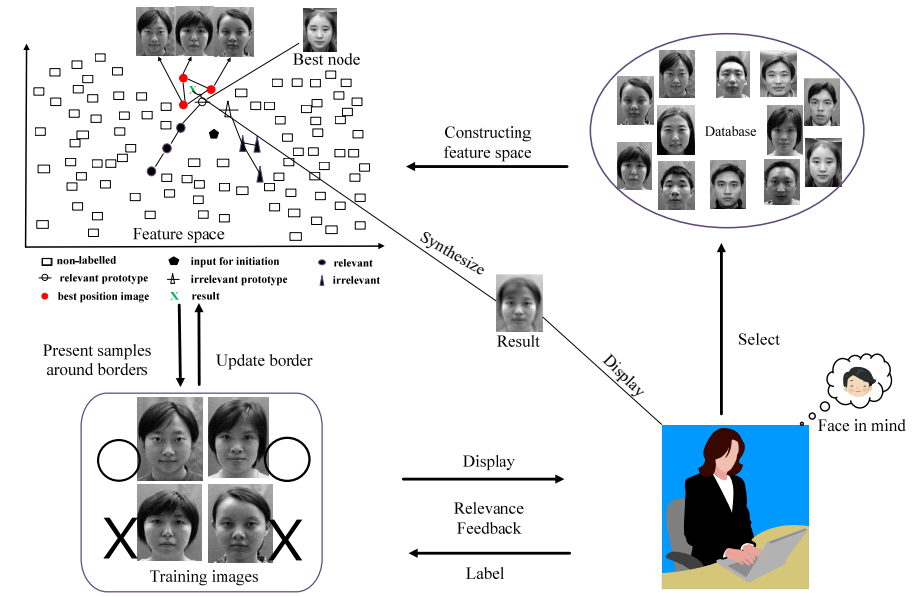


Fig. 1. Overview of the proposed system

3.1 Constructing the 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 face representation, which is effective for both face identification and verification [12, 13]

However, since our purpose was to synthesise a target face image, we needed a feature representation that could not only discriminate faces but could also be used to generate a face image. The feature vector space needed to be compact enough to al-

low for the interactive relevance feedback process. For this purpose, we used the pixel-level image feature used in the face hallucination method [5].

The basic idea is to separate a face image I into a global image I_g, which expresses the overall features of the image, and a local image I_l, which expresses the detailed face features.

$$I = I_g + I_l \ . \tag{1}$$

While the local image adds the details of the face, global images comprise information required for distinguishing between individuals. A feature vector space of global images can be constructed by applying principal component analysis to the face images in the database and finding the principal components with large eigenvalues. Formula (2) expresses a global image I in terms of the basis B of a global feature space, a coordinate value X and an average face image μ:

$$I = BX + \mu \ . \tag{2}$$

Our study uses the global feature space as the search space for locating the coordinates of the image that best matches the corresponding face in mind.

3.2 Training the Optimum-Path Forest Classifier Based on Relevance Feedback

Relevance feedback, a process that shows synthesis results to the users and updates classifiers based on user feedback, is often used in image retrieval with specific themes, such as oceans, cats or sunsets. Several researchers have proposed methods that employ various classifier types and reuse past classification results to obtain good results based on relatively minimal amounts of feedback [14, 15, 16].

Our study used the OPF [16, 17, 18] for classification. The OPF works by modelling 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 that are weighted by the distances between the feature vectors of the corresponding nodes (referred to as costs hereafter). Given a set of training nodes, a minimum spanning tree can be generated from the complete graph. Then, the adjacent training nodes are marked as prototypes if they belong to different classes. We used two classes: relevant and irrelevant. The partition of the graph is carried out by the competition process among the prototypes, which offer optimum paths to the remaining nodes of the graph. The optimum paths from the prototypes to the other samples are computed by the image foresting transform algorithm, which is essentially Dijkstra's algorithm modified for multiple sources and with more general path-value functions. Finally, all of the non-prototypes are directly or indirectly connected with the prototype that has the minimum cost. With the prototypes as the roots and the non-prototypes as the intermediate and terminal nodes, the optimum trees are built, which constitutes the OPF. OPF performs well with samples represented in a complex and high-dimension feature space. Because of this, OPF is

very important in systems that are based on the relevance feedback approach and generate results in a dialogic fashion.

Figure 2 shows an example of a classification with OPF. A minimum spanning tree is first constructed from all the samples. Then, the user labels some selected samples as positive (○) or negative (×). We thus focus on the paths that bridge positive and negative samples. The nodes bridged by the paths are called prototypes, which are represented as green dots. All other unlabelled nodes whose parent is a positive prototype are labelled as positive, and the ones whose parent is a negative prototype are labelled as negative. The nodes next to the prototypes are called border nodes as indicated by the red dots. The node located the farthest from the negative prototype and closest to the positive prototype (depicted by the purple node in the figure) is selected as the best positive sample.

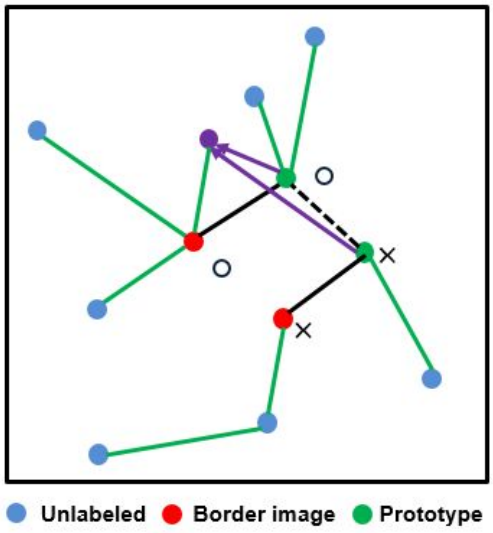


Fig. 2. Overview of the Optimum-Path Forest algorithm

As depicted in Figure 1, the OPF is trained based on users' relevance feedbacks in the following four steps:

1. The system presents the user with five male face images and five female face images of different ages and waits for the user to select one he or she thinks to be closest to the face in his or her mind. Since none of those 10 images is likely to resemble the target face, the user will select the image that is the most similar to what they are imagining according to sex and age, which acts as the initial classification boundary.
2. The four images closest to the user's selected face image in the feature spaces are returned to the user. The user evaluates and labels the images as positive (○) or negative (×), which serve as the prototypes for the OPF classifier. This evaluation phase ends if the users are satisfied with at least one of the four face images.
3. An OPF classifier is built based on this set as illustrated in Figure 2. Then, the

OPF classifier divides the unlabelled images of the database into two classes: relevant and irrelevant.

4. Four border nodes are selected, and the corresponding images are presented to the users. The user evaluates and labels the images as positive (○) or negative (×), and the new marked training images constitute and replace the former training samples to build a new OPF classifier.

At every iteration before step 4, the best positive nodes located the farthest to negative prototype and the closest to positive prototype are selected and interpolated to create the resulting face image being presented to the user. If the user is satisfied, the whole relevance feedback procedure ends.

When selecting the border nodes and the best positive nodes, we compare the costs of paths from all non-training nodes to all relevant and irrelevant prototypes. The training samples are the four images which belong to the relevant class and have the smallest ratio between the cost to the relevant prototypes and the cost to the irrelevant prototypes. The best positive nodes are those that belong to the relevant class and with the largest ratio between the cost to the relevant prototypes and the costs to the irrelevant prototypes.

In our implementation, the cost of the arc connecting two adjacent nodes of the OPF feature space is calculated with the L2 norm. Assuming there are k number of relevant prototypes and m number of irrelevant ones represented as p_i (i = 1, 2, … k) and q_j (j = 1, 2, … m) respectively, we consider $k \times m$ pairs of (p_i , q_j) in computing the ratio of the path costs to the relevant and irrelevant prototypes. Let $CR_{U \to pi}$ represent the cost of the path from the non-training sample U to the relevant prototype p_i , and let $CI_{U \to qj}$ represent the cost of the path from U to the irrelevant prototype q_j . $Relevance_{U \to (pi,qj)}$ which represents the ratio of $CR_{U \to pi}$ to $CI_{U \to qj}$, is computed as Formula (3):

$$Relevance_{U \to (pi,qj)} = \left(\frac{CR_{Ur} \to p_i}{CI_{Ur} \to q_j}\right). \tag{3}$$

In the traditional relevance feedback based image retrieval, the final result is the positive case in the position farthest from the classification boundary. To establish the classification boundary correctly, the image shown to the user for feedback must lie near the classification boundary. OPF based retrieval thus requires an initial classification boundary that sits relatively close to the positive case. Our study satisfied this requirement by gathering age and sex input information at the beginning of the process.

3.3 Synthesising Virtual Face Images Using Interpolation

The traditional relevance feedback approach is designed for searching actual images in a given database, making it impossible to synthesise non-existent face images. By synthesising images, however, it is possible to obtain the desired outcomes with a limited number of samples. Our study thus proposes a process of synthesizing face images that do not exist in the database by interpolating multiple positive images in positions far

away from the classification boundary. In principle, any point near the best positive node (i.e., the node that belongs to the relevant class and with the largest ratio between the cost to the relevant prototypes and the costs to the irrelevant prototypes) should be a desired face image.

As a practical solution, we select the top k (k = 3 in the current implementation) best positive nodes as shown in Figure 1 and calculate the result according to the following Formula (4):

$$x = \sum_{i}^{k} w(x_i)x_i / \sum_{i}^{k} w(x_i) . \quad (4)$$

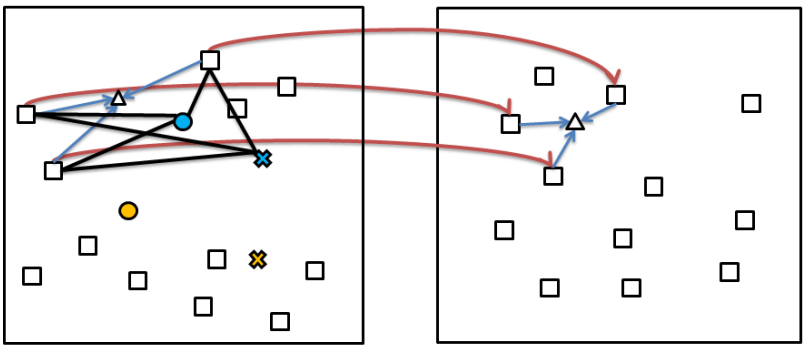
Here, x and x_i (i = 0, 1, 2) are the feature vectors of the resulting face images and the 3 best positive images, respectively. The weight assigned to x_i is $w(x_i)$, which is based on the distance given by the classifier. In the current implementation, $w(x_i)$ is assigned the average weight, which means all 3 images have equal weight.

3.4 Registration by Eyes and Mouth

The sample images in the training database need to be aligned in order to create face images without blurring. In cases where the same images were aligned only by one single registration point when synthesising new face images by interpolating several face images, the system was prone to blurring portions of the face away from the registration point due to the inherent individual variations among these different faces. Figure 4(a) and (b) show the results generated with the images were aligned by eye position and mouth position only, respectively. We can see that areas far from the registration areas are severely blurred.

To solve this problem, we built two image databases from the same source database: one composed of face images aligned by the eyes and the other composed of face images aligned by the mouth. In order to synthesise a clear face image, a group of images from the eye-aligned database and the corresponding images from the mouth-aligned database are used. More specifically, three procedures are carried out: first, a candidate image in the eye-aligned face feature space is synthesised; then, another face image in the mouth-aligned space is synthesised; by blending the two images, a clear composited face image is produced.

As the system makes it possible to obtain the same face from both databases, we only need to perform the relevance feedback process with one of the two databases to build the OPF for both databases. Figure 3 illustrates the integration between the feature spaces of the two databases. When selecting the three highest-ranking coordinates in the feature space defined by the images aligned by eyes, for example, the one-to-one correspondence between the two spaces means that we can obtain the corresponding three highest-ranking coordinates in the other feature space built from the examples aligned by the mouth. As the arrows in Figure 3 show, the system thus enables coordinate matching across the two spaces. Thus, we can synthesise two face images by interpolating the three highest-ranking coordinates in the two spaces, respectively.



(a) Eye-aligned (b) Mouth-aligned

Fig.3. Image correspondence between eye-aligned space and mouth-aligned space. In the eye-aligned feature space shown in (a), the yellow ○ and × represent user-labelled images, and the blue ○ and × represent the positive and negative prototypes. The □'s connected to the prototypes by black lines are the three best positive images. The △ represents the generated virtual image interpolated using the three best positive images. The correspondence between the best positive images in the eye-aligned and mouth-aligned spaces are illustrated with red lines. The interpolated virtual image in the mouth-aligned space is shown with a △.

To fuse the two face images computed from the separately aligned spaces (i.e., the images represented by △ in Figure 3[a] and [b]) and form a new image with clear face components, α blending is used, as given by Formula (5):

$$I=\alpha I_e+(1-\alpha)I_m. \tag{5}$$

I_e and I_m are the images from the eye-aligned and mouth-aligned spaces, respectively, and α is the blending weight. We set the value of α to 1 in the area above the eyes, set the value of αto 0 in the area below the mouth, and changed the α value in the area between the eyes and the mouth in linear interpolation. The blended image is further filtered with a bilateral filter to decrease the degree of edge blur.

Figures 4(a), (b) and (c) show the resulting images synthesised in eye-aligned space, mouth-aligned space and by α blending the former two images, respectively. We can see in Figure 4(a) and (b) that areas far from the registration areas are severely blurred, while in Figure 4(c), such flaws are alleviated.

(a)Image generated in eye-aligned space (b) Image generated in mouth-aligned space (c) Image generated via α blending

Fig. 4. Comparison between single point registered faces and face obtained by α blending.

4 Experiment and Discussion

4.1 Database

For our sample image set, we used 1,000 images of Asian faces from the CAS-PEAL database [19] and the Cartoon Face database [20]. We made all the images monochrome, and set the resolution to 96 × 128. The database comprised only frontal face images, but the positions and sizes of the faces differed. We resized and cropped the images. Then we created two databases that were aligned by eye positions and mouth positions, respectively. Our study was concerned only with general face features, so we used a low resolution of 96×128 for all the images. We set the images to monochrome to prevent colours not found in the original cases from appearing when the system interpolate multiple colour images.

Each dimension of the feature space corresponds to a pixel of the face images. Therefore, the feature vector has 12,288 (96 × 128) dimensions. Based on a primary component analysis, we used the 80 dimensions with the highest eigenvalues as our global face feature space, which provided a cumulative contribution ratio above 80%.

4.2 Experiments

To validate the effectiveness of the proposed method, we had 12 subjects (all university students in their 20s, 11 of whom were male and 1 of whom was female) attend our experiments. During the experiments, we asked the subjects to ignore hairstyles when creating and evaluating the face images because the significant differences in hairstyles among the images in the database led to blurred hair in all the generated images.

We conducted the following three experiments to determine whether the subjects could create satisfactory face images using the system and how long (in terms of time and iteration count) this process would take.

Creating Imagined Face Images
In this experiment, we had each subject imagine a face and let them use the system to create a similar image. Figure 5 shows the created images based on the subjects' imagined faces. In section 4.3, we will evaluate how these created face images resembled the imagined faces.

Creating Face Images Based on Briefly Presented Reference Images
In this experiment, we presented a reference face image that did not exist in the database to each subject for 3-4 seconds and asked the subject to create a face image resembling the reference image to validate whether the system enabled the user to synthesize an image from his or her memory. Such a situation is similar to the case where an eyewitness has seen a criminal's face for a very short time and tries to reconstruct the face image based on his or her rough impression and memory. Figure 6 shows the face images that the subjects saw for 3-4 seconds and the corresponding generated face images. As can be seen from the figures, the resulting images capture some ma-

jor features of the reference faces, such as the overall shape of faces and the relatively small sizes of the eyes.

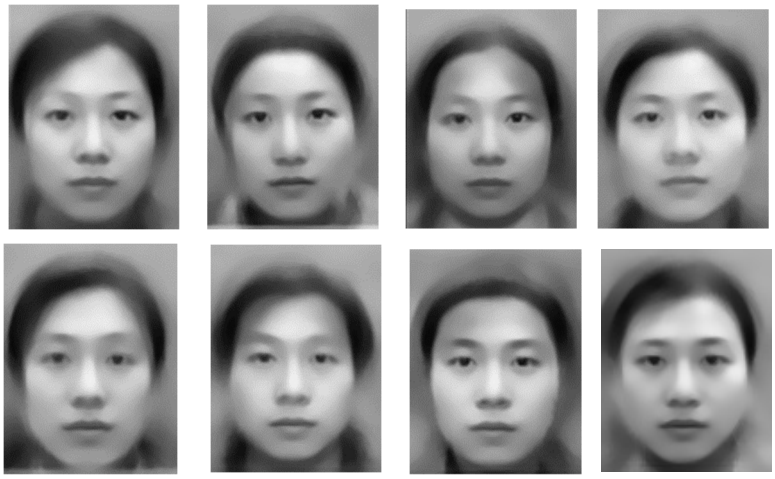


Fig. 5. Images created based on the subjects' imagined faces

Reference image A Created image of A Reference image B Created image of B

Fig.6. Reference images how n for 3-4 seconds and the corresponding created images

Creating Face Images Based on the Reference Images Presented During the Entire Process

For a more objective validation, we conducted a third experiment that presented the subjects with a reference image for the entire duration of the process until they reached a result they found satisfactory. Figure 7 shows two examples of the results. The resulting images maintain a basic consistency with their respective target images.

Reference image A Created image of A Reference image B Created image of B

Fig. 7. Reference images shown during the entire experiment and the corresponding created images

Fig. 8. Changes of the synthesised result over the process

Figure 8 illustrates how the resulting images actually changed over the process. The target image and the resulting image were noticeably different at first. As the subject went through iterations of the process, the face in the resulting image gradually came to more closely resemble the reference one.

4.3 Evaluation

Evaluation Based on Subjective Scoring

In the three experiments mentioned above, we also asked the subjects to score the results on a five-point scale (1: No resemblance; 2: Very weak resemblance; 3: Neither weak nor strong resemblance; 4: Somewhat strong resemblance; 5: Strong resemblance).

Figure 9 shows the average scores, in which the scores of three experiments are very similar. The average score for all three experiments came to 3.833. Many of the subjects declared that they were satisfied once the created images began to bear a somewhat resemblance to the target faces. Figure 10 shows the times (in seconds) that it took the subjects to arrive at satisfactory results.

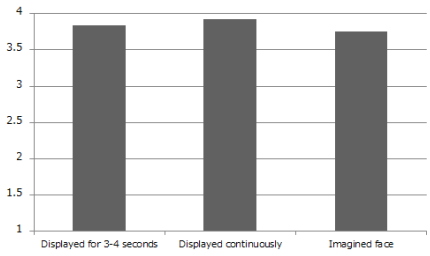


Fig. 9. Average final scores

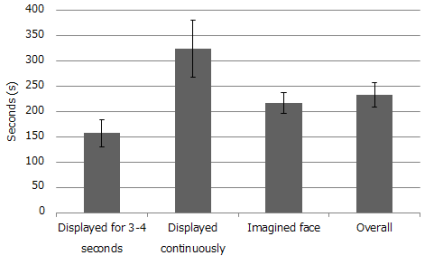


Fig. 10. Average time to final results (in seconds)

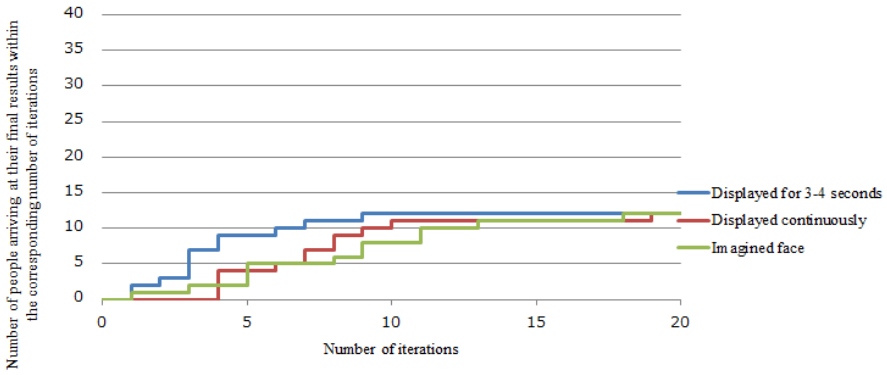


Fig. 11. Iteration numbers of each experiment. The vertical axis of the graph represents the number of subjects who reached their final results with the number of iterations shown on the horizontal axis.

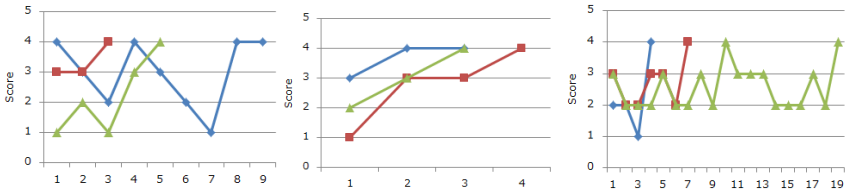


Fig. 12. Changes in scores during the relevance feedback process: (a) Creating an imagined face image, (b) creating a face image based on briefly (3–4s) presented reference images, and (c) creating a face image based on the reference images presented during the entire process.

Figure 11 shows the number of iterations that it took the subjects to arrive at satisfactory results. Figure 12, meanwhile, illustrates the changes in scores for three subjects during the relevance feedback process. Each line represents a single iteration of the process by an individual subject. On average, it took 6.5 iterations for the subjects to arrive at the final results.

Evaluation Based on Matching Test

To evaluate the effectiveness of our method, we conducted a matching test by letting a group of participants creating face images using the system from reference images, and then having another group of participants match the generated images with their reference images. In this test, 10 peoples of different age(5 in 20s, 1 in 30s, 3 in 40s and 1 in 50s) and gender(6 male and 4 female) were asked to synthesize face image, while another 13 peoples of different age(9 in 20s, 4 in 40s and 1 in 50s) and gender(10 male and 4 female) were asked to attend the matching test relating the synthesize image to the right reference image. We performed the test as the following 2 steps.

Face image generating step: we randomly choose 20 face images (12 female and 8 male) from the test image database as reference images. Each of the 10 subjects

participated the face image generation was given 2 different reference images randomly selected from these 20 images and be asked to synthesize 1 face based on each reference image. Therefore, we obtained 20 synthesized face images generated from the 20 difference reference images.

Image matching step: For each of the 13 subjects participated the image matching test, we randomly divided the 20 pairs of synthesized image and references image into 10 groups. Each group contains 2 pairs of synthesized image and references image of the same gender. Thus, we have 6 female pairs and 4 male pairs. Then, the 10 groups were shown to the subject one by one, and for each group the subject was asked to match between the generated image and the reference image. Since each of the 13 subjects performed the matching task for 10 groups, the total number of trial was 130. Out of which, 100 trials gave a correct matching result. A binominal test showed that the generated images were correctly matched to their corresponding reference image at a significance level above 99%. The result demonstrates that our system can generate images resembling the reference images.

4.4 Discussions

The results of the experiment reveal several findings. When we displayed the reference images for 3-4 seconds and then had the subjects create their images without being able to see the original references, the subjects took fewer iterations and less time to arrive at their results than they did when the reference image was presented throughout the whole process. This is likely because the images were only visible to the subjects for a matter of seconds, which made it hard for the subjects to establish a clear, accurate mental picture of the target face for comparison. Thus, the system-generated images probably created a slight recognition bias in the subjects' minds, leading them to converge on their final results relatively quickly. The subjects' evaluations of the resulting images showed several interesting trends, as well. In many cases, the evaluation scores remained relatively constant for several iterations before eventually increasing. This is because the experiment used the three highest-ranking results. Even if we were to have shuffled the rankings of the three images, their relative mutual similarity and central position would have resulted in the same generated face and produced the same score.

Some users reported that sometimes they were satisfied with most parts of the generated face after a few iterations, but unsatisfied with one particular part. The users then continued the iteration process in anticipation of getting a better result for that particular part. But unfortunately, they obtained a globally worse image with other satisfying parts became less satisfied. Although allowing users to evaluate and control the face as a whole is an advantage of our method over the component-based approaches like montage system, it is desirable to improve our system by allowing users to locally adjust individual parts.

Another drawback of current implementation lies in our feature representation. We employed a global feature space based on PCA which fail to capture the personal detail well, causing the generated face quite similar to the average face. As a major future work, we will explore better feature representation including using Convolutional Neural Network.

5 Conclusion and Future Work

In this paper, we proposed a method for the semiautomatic synthesis of a face image from a user's imagination. By training an OPF based on the user's feedback, our system successfully creates synthesised images that resembled the face images that users had in mind. One potential avenue for future work related to this paper is to explore other feature representations, such as the Convolutional Neural Network.

6 Acknowledgement

This work was supported by JSPS KAKENHI (Grant No. 17H00737) and the Public Projects of Zhejiang Natural Science Foundation Province, China (Grant No. LGF18F020015).

References

1. "E-FIT," http://www.visionmetric.com/
2. Wang, X., Tang, X.: Face Photo-Sketch Synthesis and Recognition. IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 31, no. 11, pp. 1955-1967 (2009).
3. Wu, D., Dai Q.:Sketch Realizing: Lifelike Portrait Synthesis from Sketch. Computer Graphics International Conference, pp. 13-20 (2009).
4. Cootes, T., Edwards, G., Taylor. C.: Active Appearance Models. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 6, pp. 681-685 (2001).
5. Liu, C. Shum, H. Freeman, W.: Face Hallucination: Theory and Practice. International Journal of Computer Vision, vol. 75, no. 1, pp. 115-134 (2007).
6. Gao, X., Yang, J., Lai, Z., Huang, P., Jiang, J., Gao, H., Yue, D.: Nuclear Norm Regularized coding with Local Position-Patch and Nonlocal Similarity for Face Hallucination. Digital Signal Processing, vol. 64, 107-120(2017).
7. Zhu, S., Liu, S., Loy, C., Tang, X.: Deep Cascaded Bi-Network for Face Hallucination. Computer Vision and Pattern Recognition, pp. 614-630 (2016).
8. Liu, R., Wang, Y., Baba, T., Masumoto, D., Nagata, S.: SVM-based active feedback in image retrieval using clustering and unlabeled data. Pattern Recognition, vol. 41, no. 8, pp. 2645-2655 (2008).
9. Xiang, Y., Yang, H., Li, Y., Chen. J.: A new SVM-based active feedback scheme for image retrieval. Engineering Applications of Artificial Intelligence, vol. 37, pp. 43-53 (2015).
10. Wang, B., Zhang, X., Li, N.: Relevance Feedback Technique for Content-Based Image Retrieval using Neural Network Learning. International Conference on Machine Learning and Cybernetics, (2006).
11. Fu, H., Qiu, G.: Fast Semantic Image Retrieval Based on Random Forest，International Conference on Multimedia, pp. 909-912 (2012).
12. Sun, Y., Chen, Y., Wang, X., Tang, X.: Deep Learning Face Representation by Joint Identification-Verification. Advances in Neural Information Processing Systems, pp. 1988-1996 (2014).
13. Sun, Y., Wang, X., Tang, X.: Deep Learning Face Representation from Predicting 10,000 Classes. Computer Vision and Pattern Recognition, pp. 1891-1898 (2014).

14. Ruthven, I., Lalmas, M.: A Survey on The Use of Relevance Feedback for Information Access Systems. The Knowledge Engineering Review, vol. 18, no. 2, pp. 95–145 (2003).
15. Li, H., Toyoura, M., Shimizu, K., Yang, W., Mao, X.: Retrieval of Clothing Images Based on Relevance Feedback with Focus on Collar Designs. Visual Computer, vol. 32, no. 10, pp. 1351-1363 (2016).
16. Silva, A., Falcao, A., Magalhaes, L.: Active Learning Paradigms for CRIR Systems Based on Optimum-Path Forest Classification. Journal of WSCG, vol.18, no. 1-3, pp. 73-80 (2010).
17. Papa, J., Falca, A.: 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).
18. Papa, J., Falcao, A., Suzuki, C.: Supervised Pattern Classification Based on Optimum-Path Forest. Imaging Systems and Technology, vol, 19, Issue 2, pp. 120-131 (2009)
19. Li, H., Liu, G., Ngan, K.: Guided Face Cartoon Synthesis. IEEE Transactions on Multimedia, vol. 13, no. 6, pp. 1230-1239 (2001).
20. Gao, W., Gao, B., Shan, S., Chen. X., Zhou, D., Zhang, X., Zhao, D.: The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations. IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 38, no. 1, pp. 149-161 (2008).