FACIAL IMAGE COMPOSITION BASED ON ACTIVE APPEARANCE MODEL

Hong-Xia Wang, Chunhong Pan, Haifeng Gong, Huai-Yu Wu,

National Laboratory of Pattern Recognition, Institute of Automation,
Chinese Academy of Sciences, Beijing 100080, P.R. China
{hxwang, chpan, hfgong, hywu}@nlpr.ia.ac.cn

ABSTRACT

In this paper, based on Active Appearance Model (AAM), we present an easy-to-use framework for facial image composition, which can automatically exchange the source image's face or facial features onto the target image. The manual interaction is simple and the user only needs to input semantic information of ROI (region of interest) to be exchanged, such as 'face' or 'eyes'. Our framework mainly consists of two steps: model fitting and component compositing. Model fitting is designed to interpret each input image and obtain a synthesized model face of the image. Then by using component compositing, visual pleasing result is generated by solving Poisson equation with the boundary condition, produced automatically from model fitting. Furthermore, we propose a solution for eliminating the artifacts when part of the target face is occluded by hair, glasses, etc. The visually satisfactory results demonstrate the effectiveness of our facial image composition system.

Index Terms— Image composition, Active appearance model, Poisson equation, Image matting, Thin-plate splines

1. INTRODUCTION

Image composition is one of the key topics in image editing and is widely used in the field of entertainment and film. Generally, it can be divided into two categories. The first one is to paste an object or a region from the source image onto the target image. The second one is to piece together image patches from a single or multiple images to generate a new picture. Below, we give a brief overview of works related to image composition.

Image matting: Image matting [1] is a common way to extract an object from a source image and then paste it onto a target image with an alpha channel. Most existing matting methods require the input image to be accompanied by a trimap, used to estimate foreground colors and alpha channel. For good results, the unknown regions in trimap must be as small as possible, which needs careful user interaction.

This research is sponsored by Natural Science Foundation of China (NSFC No. 60675012).

Poisson image editing: Poisson image editing [2], is an effective approach for seamless image composition. By solving Poisson equation with a guidance field and a user-specified boundary, Poisson image editing can seamlessly blend the colors of both input images without visible discontinuities around the boundary. For different tasks (e.g. seamless cloning, texture flattening and seamless tiling), different guidance fields and boundary conditions are exploited to solve the Poisson equation. However, the effectiveness of this technique greatly depends on the quality of boundary provided by the user.

Interactive digital photomontage: Interactive digital photomontage [3] is an interactive, computer-assisted framework for combining parts of a set of photographs into a single composite picture. This framework makes use of two techniques: graph cut and Poisson editing. First, the user needs to draw a number of strokes to initialize the graph cut. Graph cut is used to choose good seams within constituent images so that they can be combined as seamlessly as possible. Then Poisson editing is used to further reduce remaining visible artifacts.

As discussed above, it can be seen that for most image composition techniques, tedious interactions are required. In this paper, we present a framework that can exchange faces or facial features automatically. The user only needs to input semantic information of ROI, such as simple words 'face', 'eyes', etc. Our framework consists of three modules: model fitting, component compositing and occlusion processing. Specifically, first, based on Active Appearance Model [4], our method interprets each input face image and obtains the shape and texture information of each image. Then, a seamless composition is generated by solving Poisson equation with the boundary condition, produced automatically from model fitting. In addition, we also propose a solution for eliminating the artifacts if part of the target face is occluded.

In Section 2, we briefly summarize the three modules of the framework. Section 3 presents our experimental results. Section 4 concludes the paper and discusses our future work.

2. FRAMEWORK OVERVIEW

Fig.1 shows the flow chart of our framework. Given two face images (grey-level or color images) and the semantic information of ROI (e.g. words like 'face', 'eyes'), the goal of our

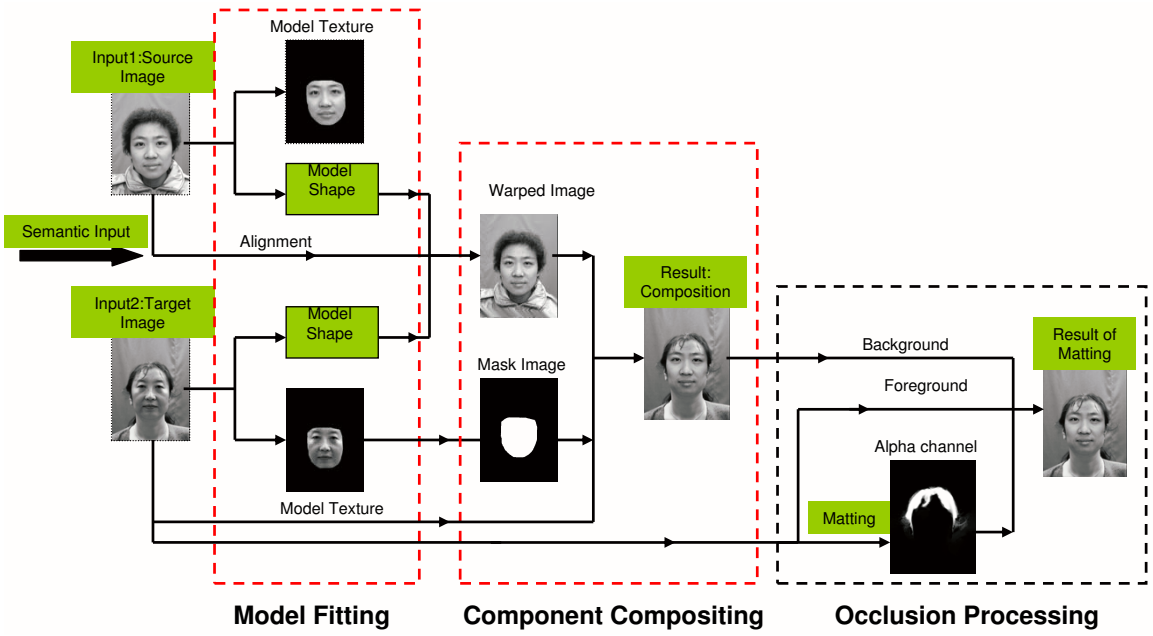

Fig. 1. The flow chart of the framework. Three modules: Model Fitting , Component Compositing and Occlusion Processing.

framework is automatically to paste the source image's ROI onto the target image. The first step is model fitting. Based on Active Appearance Model, each input face image is interpreted and the shape and texture information of the image are obtained from the synthesized model face. With the information obtained from model fitting, the boundary of ROI is automatically produced and the source image is aligned with the target image. The goal of the alignment here is to ensure the position of the features in the source and target images more accurate. Then, the composition result is generated based on Poisson equation. Finally, if occlusion exits, we solve the problem by using the third module (image matting) to eliminate the artifacts.

2.1. Model Fitting

Active Appearance Model (AAM) has been widely used for face image processing, such as face detecting, face tracking, etc. It contains a statistical model of the shape and grey-level appearance. Given a good enough training set, the appearance model can generalize to almost any face, potentially giving a full photo-realistic approximation.

We select 280 face images as the training data from the CAS-PEAL-R1 [5], each labelled with 87 landmark points at key positions to outline the main features. The shape model is generated by representing each set of landmarks as a vector $\mathbf{x}$ and applying a principal component analysis (PCA) to the data. To build a statistical model of grey-level appearance, we use the technique developed by Bookstein [6], which is based on thin-plate splines, to warp each face image so that its control points well match the mean shape. Fifteen landmarks are used to deform the face images. These landmarks are selected from 87 landmarks of the shape model. Then we sample the grey level values $\mathbf{g}$ from shape-normalized faces and obtain a texture model by applying PCA. Since there exits correlations between the shape and texture parameters, we apply a further PCA to the concatenated parameter vector, to obtain an active appearance model:

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{Q}_s \mathbf{c} \quad (1)$$

$$\mathbf{g} = \bar{\mathbf{g}} + \mathbf{Q}_g \mathbf{c} \quad (2)$$

where $\bar{\mathbf{x}}$ is the average shape, $\bar{\mathbf{g}}$ is the mean texture, $\mathbf{c}$ is a vector of appearance parameters controlling both the shape and grey-level of the model, $\mathbf{Q}_s$ and $\mathbf{Q}_g$ map the value of $\mathbf{c}$ to changes in the shape and shape-normalized grey-level data.

To explain the 98% of the observed variation of the 280 examples, our facial model contains 66 shape parameters, 168 texture parameters and 121 combined parameters. And our model uses 41125 pixels to make up the face patch.

Then, we interpret a new image with the appearance model. It is actually an optimization problem to minimize the difference between the synthesized image obtained from the AAM and the new image. An iterative algorithm is used to solve the optimization problem. At each iteration, the model and pose parameters are adjusted by the difference between the synthesized image and the new image. Therefore, before optimization, we use the method of Fixed Jacobian Matrix Estimate [7] to learn the relationship between the difference and the error in the model model parameter from the training images. For each new image, the search is started with the mean model displaced from the true face center.

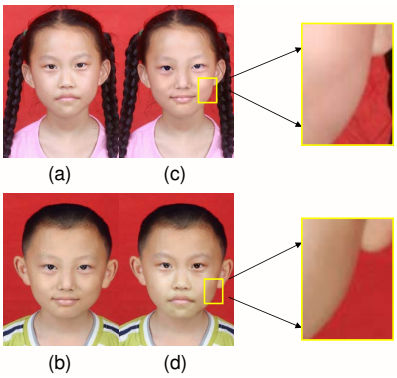

Fig. 2. The composition result of two color images. (a) and (b) are two input images, (c) and (d) are the composition results generated by exchanging each face. The yellow rectangular shows the boundary of the editing region. The images in the right column are the zoom-in views. This result is obtained by our system without any user interaction.

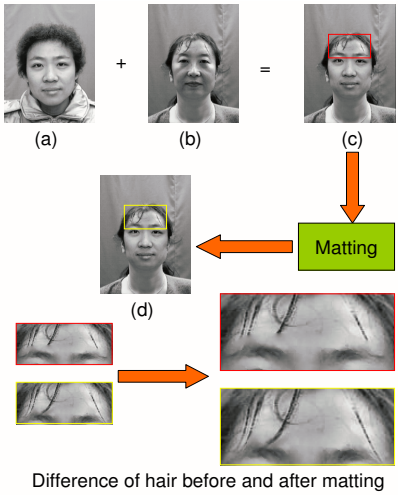

Fig. 3. The hair occlusion problem. (c) and (d) are results before and after matting. The bottom row shows the zoom-in views of the results.

2.2. Component Compositing

In the second module of our framework, we exploit the Poisson equation to perform the seamless component compositing. The difference between the Poisson image editing [2] and our method is that we obtain the boundary of ROI automatically from the facial model. Since the boundaries of different regions are obtained by different methods, we take the face and eyes for instance to describe how to get the boundary information.

After model fitting, we obtain both the shape location and texture patch of each input image. If the whole face is just the ROI, the boundary information can be produced from the model patch of the target image directly. We take the boundary information as a mask image. In the mask image, the pixel values belonging to the face are set to 255, and other pixels are set to 0 (see the Fig.1). However, If part of the face (such as 'eyes') is the ROI, we should use the shape information of the model. Every training image is marked with 87 points and the order of the points is fixed, i.e., the IDs of the points representing the eyes region are known. So with such information, we can get a mask image of eyes from the shape information of the model. Moreover, with the shape information, the source image is well aligned with the target image in order to ensure the position of the features in the source and target image are more accurate.

By solving the Poisson equation with the boundary condition of ROI, the composition result is generated by blending the colors of the warped source image and the target image. Let g be the warped source image, f^* be the target image, Ω be the ROI in the mask image and f be result image. The pixel values out of the region Ω in the result image f are equal to the target image. The guided interpolation of f over Ω is to make its gradient field close to the gradient field of g defined as the solution of the minimization problem:

$$\min_f \iint_\Omega \parallel \nabla f - \nabla g \parallel^2 \; with \; f|_{\partial\Omega} = f^*|_{\partial\Omega} \qquad (3)$$

where $\nabla . = [\frac{\partial}{\partial x}, \frac{\partial}{\partial y}]$ is the gradient operator.

The problem can be solved very efficiently by a number of well developed Poisson solvers. For the composition of color images, the three color channels can be computed separately. Considering the color continuities in the composition image, we scale the gradient field with a factor d. Fig.2 shows the composition results of two color images. From the zoom-in views at the right column, we can observe that there are no color discontinuities around the boundary of the face in the result images.

2.3. Occlusion Processing

The occlusion will cause obvious artifacts in composition result —when part of the target face is occluded by an object, such as hair, glasses. Take the hair occlusion for instance. As illustrated in Fig.3, part of the face in the target image Fig.3(b) is occluded by the fringes, therefore, in the composition image Fig.3(c), the fringes are not continuous and the result looks unpleasing. Based on image matting, our method successfully deals with this problem, and obtains satisfactory results.

As we known, an image can be composed with the foreground, the background and alpha channel. Image matting is used to extract an object from a source image and then naturally paste it onto a target image using an alpha channel. For instance, we take the hair of the target image Fig.3(b) as foreground and use the matting algorithm to extract corresponding alpha channel (see Fig.1). The composition im-

age Fig.3(d) is generated by combining the alpha channel, the foreground hair in the target image Fig.3(b) and the background region (except the hair) in the composition result Fig.3 (c). In order to reduce manual interactions, we use the matting method presented by Anat Levin [8]. This method only needs a small amount of user strokes and can obtain a globally optimal alpha matte. The zoom-in views at the bottom row in Fig.3 compare the results in the images before and after the matting process, and demonstrate that our method can solve the occlusion problem sucessfully.

3. EXPERIMENTAL RESULTS

We apply our method on a variety of input images, and achieve visually pleasing results. Fig.1 shows the flow chart of our framework. Given two face images and semantic information of ROI, our method can automatically generate the satisfactory composition results. Fig.2 illustrates the composition of two color images. Thanks to the accurate location of model, no color discontinuity exits around the boundary of face. Occlusion problem is also well solved, as shown in Fig.3. Fig.4 shows a set of composition results. The first row of images are the input face images and the second row are the results of composition. These images come from [3], in which a amount of user strokes are required for composition, while our system do not need any manual interactions and the results are still perfect.

4. CONCLUSION AND DISCUSSION

In this paper, we propose a framework to perform seamless facial image composition without tedious user interaction. Our system is easy-to-use and the only interaction is to input the semantic information of ROI. Our method makes use of the AAM and Poisson equation to exchange faces or facial features and also addresses the occlusion problem.

We suggest a few future directions: 1) the AAM can be replaced by a high resolution grammatical model [9]. This model achieves nearly lossless representation of high resolution image, and thus we can use the model patch instead of the source image to solve Poisson equation. 2) Our system can not deal with images with large-scale viewpoint. One possible solution is to get the front picture of input image, then project the composition result to original viewpoint.

5. REFERENCES

[1] Y. Chuang, B. Curless, B. Salesin, and R. Szeliski, "A bayesian approach to digital matting," in *Proceedings of CVPR*. IEEE, 2001, vol. 2, pp. 264–271.

[2] P. Perez, M. Gangnet, and A. Blake, "Poisson image editing," in *Proceedings of ACM SIGGRAPH*. IEEE, 2003, pp. 313–318.

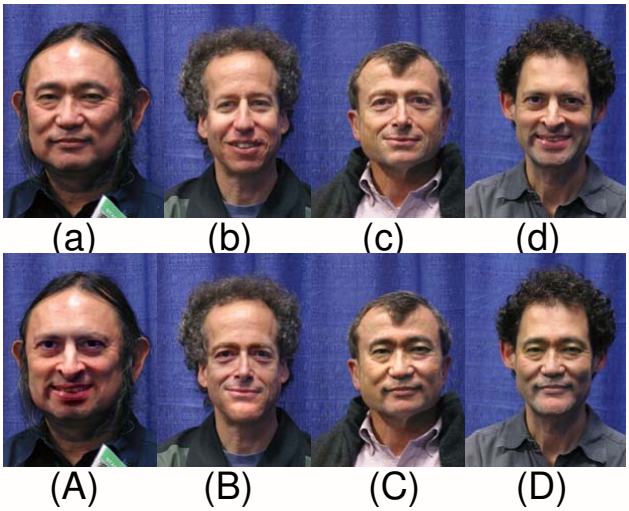

Fig. 4. The results of composition by changing faces. (A) is the combined result of (a) and (d); (B) is the combined result of (b) and (c); (C) is combined result of (c) and (a); (D) is combined result of (a) and (d).

[3] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen, "Interactive digital photomontage," in *Proceedings of ACM SIGGRAPH*. IEEE, 2004, vol. 3, pp. 294–302.

[4] T. Cootes, G. Edwards, and C. Taylor, "Active appearance models," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 681–685, 2001.

[5] Wen Gao, Bo Cao, Shiguang Shan, Xilin Chen, Delong Zhou, Xiaohua Zhang, and Debin Zhao, "The cas-peal large-scale chinese face database and baseline evaluations," *IEEE Transaction on System Man, and Cybemetics*, 2004.

[6] Fred L.Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 567–585, 1989.

[7] M. Stegmann, B. Ersboll, and R. Larsen, "Fame–a flexible appearance modeling environment," *IEEE Transaction on Medical Imaging*, vol. 22, pp. 1319–1331, 2003.

[8] A. Levin, D. Lischinski, and Y. Weiss, "A closed form solution to natural image matting," in *Proceedings of CVPR*. IEEE, 2006, vol. 1, pp. 61–68.

[9] Zijian Xu, Hong Chen, and Song-Chun Zhu, "A high resolution grammatical model for face representation and sketching," in *Proceedings of CVPR*. IEEE, 2005, pp. 264–271.