Self-Shadow and Ambient Occlusion Recovery for Face Images in Face Replacement

†*Shih-Hao Hsiung,* ‡*Hong-Shang Lin,* †*Ming Ouhyoung*
†Dept. of Computer Science and Information Engineering,
National Taiwan University, Taipei
E-mail: {supermfb, ming}@cmlab.csie.ntu.edu.tw
‡Dept. of Electrical Engineering,
National Taiwan University, Taipei
E-mail: {amsd}@cmlab.csie.ntu.edu.tw

ABSTRACT

ABSTRACT

In this thesis, we present a face replacement system, which can simulate the harsh lighting condition such as self-shadow and ambient-occlusion in the target face. In the relighting step, we propose a method that combines the relighting source face and a shadow map to simulate the self-shadow condition. Then in the composition step, we use Drag-and-Drop pasting to better maintain the facial saliencies with optimal composition boundary. Furthermore, we incorporate the mixing gradient in the blending process to preserve the vivid expression details.

Keywords: Face Replacement, Face Relighting, Face Composition.

1. Introduction

In recent years, with the advance of digital camera and camcorder, the public can get more and more digital media than before. An interesting area people usually care about is: editing the digital media to get personalized result. One important issue in such area is face replacement.

The goal of face replacement is to replace any one‘s face to the subject face of images or movies in existence, and the replaced movies or images look realistic and natural. It is already widely used by visual effects in film industry.

However, the traditional face replacement methods usually assume normal lighting conditions in target images. So the harsh lighting conditions like ambient occlusion and self-shadows will not be recovered well.

In this paper, we aim to handle the problems for harsh lighting condition. In the relighting step, we still use spherical harmonics function to generate the basic relighting image for source face. Moreover, we also estimate the position of the principle light source to generate the corresponding shadow map. Then we combine the spherical harmonics relighting image and the shadow map to simulate the hard self-shadow in the target face. In the composition step, we apply Drag-and-Drop pasting to do seamless composition, so the facial saliencies near the boundary can be better maintained. To make the facial expression more natural, we apply the mixing gradient when we blend the source face and the target face. The wrinkles and folds in target face can thus be preserved.

2. Related Work

2.1. Face Replacement

Face replacement can be categorized in image-based and model-based methods. The difference between them is whether the 3D model of the target face is synthesized.

One of image-based methods is [3]. [3] uses a large database of face images derived from the internet. They select some candidates in the database, adjusting the lighting condition and skin color of faces to blend with the target images. This method does

not estimate pose on 3-dimension, and it tends to replace the target faces with similar ones in the database, not with arbitrary user-input face. The quality is thus dependent on the candidate selection. [6] provides a 3D morphable model based method to reconstruct face model from single image. And they also estimate rendering parameters such as camera calibration, illumination, etc. [4] records additional dataset of 35 static 3D laser scans which form the vector space of mouth shapes and facial expressions to capture the mouth movement. [5] apply the estimated rendering parameters of the target face to the synthesis source face model for face exchanging. It also estimates one principle direct light. However, the way of light source estimation is different from our method. And the system in [5] produce artifacts when composite the faces since it does not consider the difference between skin colors. [8] presents a system that replaces the target face in a video. They clone the expression of the target face at each frame to the synthesis source face model with a 3D face expression database, and enforce the temporal coherence for illumination and face poses. Our face replacement framework is similar to those model-based methods, while it focuses on the improvement of two aspects: illumination simulation, and seamless composition, and expression detail preservation. The replacement results are more natural and vivid.

2.2. Pose Estimation

Pose estimation in face replacement is required. The techniques can also be categorized into image-based and model-based methods.

Image-based methods like [15] consider this problem as a classification problem. They extract distinctive facial saliencies from Haar-like features to determine the optimal pose range with classifiers. Other image-based techniques are based on facial characteristics, and they often assume that the head pose is closed to a frontal pose. [7] uses the colors on skin region and hair region of heads to estimate head poses. All of the above methods need uniform lighting condition on the face, because image-based techniques rely on color information.

Most model-based techniques start from the morphable model. [5] and [4] estimate lighting condition and pose estimation at the same time to handle different illumination. Instead of using color information, [10] extracts contours and feature points of the target face. Those features are mapped onto the source face model and the temporal coherence of head pose is considered between near frames. The problem of using feature points only is that the facial saliency is not preserved well after face replacement. We synthesize the source face by the morphable model in our system, and then we combine intensity and feature information to estimate pose of the face model. The facial saliencies are better preserved while the pose is accurately estimated in our system.

2.3. Relighting

Instead of directly estimating positions and directions of light sources for the target face, most traditional techniques for face replacement use finite linear subspace to embed the lighting condition in function basis and generate the corresponding relighting image. [1] and [17] suppose that the head is a convex Lambertian object under distant and isotropic light, and they approximate the lighting conditions with spherical harmonics. To handle objects which have different materials in different area, [12] proposes a ratio image technique to remove material dependency in the radiance environment map. [16] and [2] design a framework to refine the surface normal and albedo for using spherical harmonics. [2] tries to segment human face into several parts, and they iteratively computes surface normal, albedo, and lighting. [18] incorporates spatial coherence constraint of skin colors into their energy function.

In our system, we refer [18] to get precise albedo and spherical harmonics coefficients. Furthermore, we combine the shadow map and spherical harmonic relighting image to simulate the hard self-shadow in target faces.

2.4. Seamless Composition

When pasting other objects to the target, the intensity variation near the boundary must be smooth to make results more natural. [13] optimizes the pasting boundary to both minimize color variation and the replacing area. [9] aims to estimate the opacity for each pixel in the foreground of the image. Both of [9] and [13] cannot handle the case with very different skin colors. [14] proposes a blending method

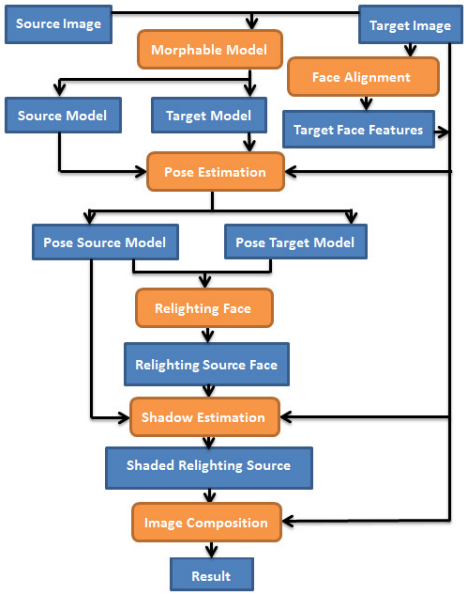


Figure 1: System Pipeline

to overcome this problem. It keeps gradient variation on the source image and enforces colors of target face in the boundary condition. Blending two images is by the Poisson equation. The quality of results from [14] is strongly depend on the user-specified destination area. [11] proposes an object function to compute an optimal boundary. They also construct a blended guidance field to incorporate the object‘s alpha matte to faithfully preserve the characteristics near boundaries of object‘s.

We use the Drag-and-Drop pasting [11] to better preserve natural facial saliencies in the target face, as well as the ambient occlusion.

3. System Overview

Our input consists of a source image and a target image. The two image do not need to have the similar illumination and head pose, and characters in the images can be also different people with dissimilar skin colors, as well as facial expressions. Our processing pipeline is shown in Figure 1. Finally, our system outputs the blending face. The source and target head models are synthesized by the morphable model [6] first. And then we detect facial features and silhouette in the target face in the face alignment module [19]. The positions of eyes, nose, mouse and the region of the face in the image will be detected. In the pose estimation module, we will estimate the suitable poses for the source face model and the target face model to fit the facial features of each other. In the face relighting module, the albedo, normal, and the lighting condition are estimated with spherical harmonic function, and then the corresponding relighting image for the source face is generated with the ratio image technique. To better simulate the shadows on the target face, we will estimate the principle light source with the relighting source image, the shadow map on the source face, and the target face in the shadow estimation module. After we generate a shadowy-relighting source face, it will be blended into the target image in the image composition module. To make blending results more seamless and the skin colors are more natural, we use the concept of Drag-and-Drop pasting to get a seamless and natural skin color result. And the mixing gradient is utilized to keep the facial expression wrinkles in both source and target image.

4. Pose Estimation

We combine the intensity coherence and geometry coherence to estimate the head pose. The intensity coherence minimizes color difference between two areas, one is the face region in the target image, and the other one is the projection area of source model.

The geometry coherence prevents the error brought by incorrect colors under different lighting condition or skin colors. We add constraints with the positions of some facial features, assuming the positions will not change too much through different facial expressions. We choose 15 features for the target face from face alignment result [19]. Four are the corners of eyes, eight are surrounding nose, two are corners of mouth, and one is in the chin. More features are allowed but it takes more time to track. The corresponding facial features for the source face can be obtained from the pre-marked 3D vertices in the data set of 3D morphable model. The geometry constraint enforces that the positions of the projected facial features of the source face are close to the ones of the target face.

We combine the intensity term and geometry term to guarantee precise pose estimation results.

We exploit to minimize the energy function:

$$E = \sum_{x,y} \left(I_{target}(x,y) - I_{model}(x,y)\right)^2 + \sum_i \left(\begin{pmatrix} u_i \\ v_i \end{pmatrix} - \begin{pmatrix} x_i \\ y_i \end{pmatrix}\right)^2$$

, where I_{target} is color of target image. I_{model} is the image by projected source model to the image plane with the estimated pose parameters . The feature term can be set 2-norm distance in pairs of target image(x_i, y_i) and source model feature points(u_i, v_i).

5. Illumination

Because the relighting with spherical harmonics cannot simulate hard self-shadow very well, the replacement will appear unnatural if the source face and target face have very different illumination condition.

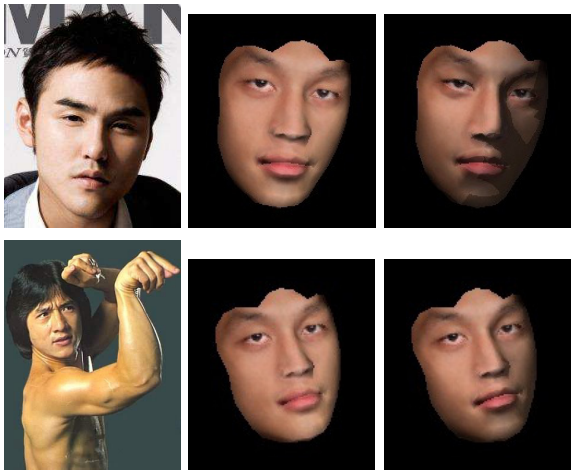
Because the deep self-shadow on the target face is often generated from a single principle light or several lights in the closed location, we can approximate the lighting condition with a small number of light sources. In this paper, currently we only estimate one principle light. We believe this is enough for common scenes by arguing that: When the scene has more than one light source and those lights locate different position, the shadow will be interactively influenced by different lights. And the edge of shadows becomes softer or close to the ambient occlusion effect, which can be approximated well by spherical harmonics.

Based on the above reasoning, we combine the spherical harmonica results and estimated shadow map generated from the light source to simulate harsh lighting conditions.

In Section 5.1, we briefly explain the method for albedo and face normal estimation with spherical harmonic functions. In Section 5.2, we introduce our light source estimation method, and the corresponding shadow map generation.

5.1. Robust Albedo Relighting

Using spherical harmonics function to synthesize illumination model, the average skin color is taken as the initial albedo. In the face with cast shadow and saturate region, the average skin color cannot approximate the true albedo. In our work, we adopt the algorithm proposed in [18] to refine albedos. The energy function consists of the data term and smoothness term. Assuming neighboring pixels with similar colors have similar albedos, a spatial constraint is incorporated into the smoothness term. The simultaneous over-relaxation approach is utilized to minimize this energy function.



(a) (b) (c)

Figure 2: Shadow map generation. (a) Target face (b) Synthesized source face model with same pose(c) After adding self-shadow on the face model.

5.2. Self-Shadow in Face Image

First, the initial position of the light source is specified by users or the default value. Given the position of the light source, we can use volume shadow method to generate the corresponding shadow map. We propose an energy minimization framework to estimate the optimal position of the light source. Given the shadow map and relighting source image, the goal is to minimize the Euclidean distance over all pixels and color channels between the target image and shaded relighting image, which is combination of the relighting image and the shadow map. The energy function can be written as:

$$E = \sum_{x,y} \left(S(l_p)(x,y) - I_{original}(x,y) + I_{relight}(x,y) - I_{target}(x,y)\right)^2$$

,where l_p is the lighting position. $S()$ is the

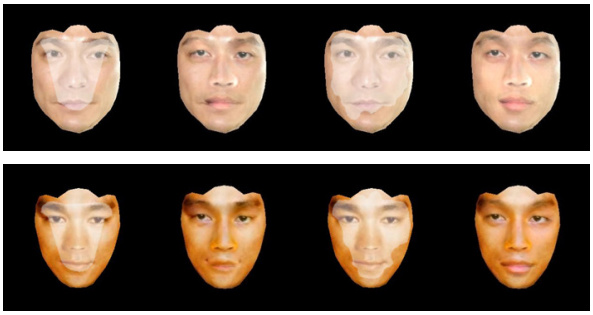
lighted model with the shadow which is generated by l_p. $I_{original}$ is the same light but no shadow. $I_{relight}$ is relighting source image by spherical harmonica method. I_{target} is the target face. The shadow map is as Figure 2. Since the energy function is hard to analyze from partial derivation, we choose NM simplex (Nelder-Mead simplex algorithm) to minimize it. Nelder-Mead algorithm extrapolates the behavior of the energy function by arranging parameters as simplex. In each iteration step, it chooses to replace one of the simplex with a new simplex, if new simplex is unimodal and it can make function smoother than current simplex. On the other hand, if new simplex cannot get much better than current simplex, it will step across a valley. The simplex converges towards a better result.

The NM simplex multidimensional minimization algorithm has been provided in GNU Scientific Library. We need to set the formula of the function, initial guess, and size of the initial trial steps to minimize the energy function. In our experiment, users just need to specify a rough light source position. As long as the initial shadow from user-specified light source has partial overlapping region with the true shadow, the optimal result solved by minimizing our energy function will be good.

5.3. Optimal Boundary

Although shaded relighting image simulates the hard self-shadows of target image well, there are still some artifacts when we blend it to the target face. Since the shaded relighting source face may not have fully consistent shadow position or ambient occlusion to the target face, the pasting area must be well determined to make seamless composition. In this paper, we use the concept from Drag-and-Drop pasting to estimate optimal boundary of pasting area. The shadows and ambient occlusions of target face and source face will blends well with the optimal boundary.

How can we find the optimal boundary? First, the replaced area is restricted to be smaller than both the source face area and target face area in the images. Second, the boundary cannot across both the facial features of the source face and the target face. Face alignment is utilized again to determine the face area and positions of facial features. Finally, the optimal boundary is solved with the shortest path algorithm [11].



(a) (b) (c) (d)

Figure 3: We show the difference between No optimal boundary results and results with optimal boundary. Light white area is replacement region. (a) The boundary surround facial features. (b) Results with general boundary. (c) Optimal boundary. (d) Results with optimal boundary.

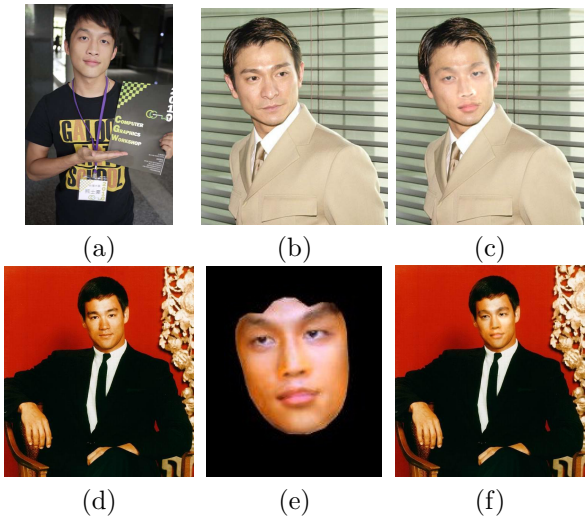
6. Target Face Geometry Recovery

Complicated facial geometry like wrinkle, scruff, pock or scar, is hard to clone from the target face to the source face via 3D geometry modification, because such methods rely on the quality of the synthesis model and registration accuracy between two faces. Instead of modifying 3D geometry, we regard those facial saliencies as textures. All we need to do is selectively retain those textures when we composite the two faces.

Those textures often have obvious edge and color variance. So we can detect and preserve those regions with larger gradients. Since we want the replaced result not only preserves the features of the target face but also has the features in the source face, source face gradients will be compared with target face gradients. Then the stronger gradient variance will be used as the guidance field of Poisson methodology. This method is called mixing gradients, using the following guidance field:

$$v_{pq} = \begin{cases} f_p^* - f_q^* & \text{if} |f_p^* - f_q^*| > |g_p - g_q|, \\ g_p - g_q & \text{otherwise,} \end{cases}$$

, where v_{pq} is guidance field of pixel (p, q) for poisson image editing. $(f_p^* - f_q^*)$ and $(g_p - g_q)$ are gradients of pixels (p, q) in the source and target image. Gradients are avoided to mix on regions contain fa-



(a) (b) (c) (d) (e) (f)

Figure 4: Results of normal lighting condition. (a) The target image of Andy Lau. (b) Face replaced result. (c) The target image of Bruce Lee. (d) Unusual skin color on relighting source image. (e) Face replaced result.

(a) (b) (c)

Figure 5: Hard self-shadow simulation. (a) The target image of Ethan Ruan. (b) The blending face without adding shadow. (c) The blending face with adding shadow.

(a) (b) (c)

Figure 6: Case with strong ambient occlusion and complicate facial geometry (a) The target image of Jackie Chan. (b) Blend faces without mixing gradient. (c) Blend faces with mixing gradient.

cial saliencies of two faces. Those regions are only replaced by the source face. If we do not enforce this constraint, those regions may generate ghost or lost the characteristic of source face.

7. Experimental Results

We experiment our system for several characters. To demonstrate that replacing shadowy face gets better results by our system, we select the faces which contain self-shadow and ambient occlusion from public web pages. And we show the result of each input with original image, relighting face, optimal replaced region, and final blending result.

In Figure 4 (a) and (b), we replace Subject A's face into Andy Lau. This image does not have complicated illumination environment, so there just have a little saturation region and no wide-region shadow. Because of slight changes in light and shadow, the generated face replacement result looks realistic in this case. We replace subject A's face into Bruce Lee in Figure 4 (a) and (d). Figure 4 (d)has similar lighting conditions as Figure 4 (b), with slight changes in light and shadow, but the spherical harmonic relighting image has unusual skin color in the saturated region Figure 4 (e). We suppose that this situation results from relative pale skin color of target face, but it still has similar illumination model. When we blend two faces by seamless composition, it can adjust the color tone of relighting result. We can still generate realistic result.

In Figure 5, we replace the movie star Ethan Ruan. This target image has wide self-shadow region. We compare the replaced result which using spherical harmonics relighting only. As mentioned above, we not only simulate self-shadow well but also maintain ambient occlusion effect on the target face.

In Figure 6, we replace subject A's face into the movie star Jackie Chan. The target face has obvious wrinkles on the eyebrows and ambient occlusion on his forehead. We compare the results whether using mixing gradients or not. The figure combines without mixing gradients, so it has the smooth area

on the eyebrows. On the other hand, the mixing gradients result preserve wrinkles on the eyebrow, and the facial expression make character more lifelike.

8. Conclusion and Future Work

In this paper, we present a system to semi-automatically replace faces from source image to the target image. Our system can handle harsh lighting condition and facial expression details. First, we combine face alignment and Drag-and-Drop pasting to find the optimal pasting boundary on the target face, thus produce a seamless blending results while maintain the facial saliencies and ambient occlusion parts. Second, we apply mixing gradient to the blending process, preserve both facial expression details on the source face and the target face. Third, we combine the spherical harmonic relighting image and the shadow map to simulate deep self-shadow on the target face. The corresponding shadow map is generated from our principle light source estimation. Our face replacement system still has limitations. First, the system relies on face alignment results. The facial features region detection from face alignment will affect the quality of pose estimation and composition boundary. Second, we synthesize the source face model by morphable model. The reconstruction accuracy strongly depends on the 3D model data set. Usually the morphable model can not reconstruct complicate facial expression details due to the limitation of data sets. So the reconstructed model may lose important information of the source face at the model synthesis stage. After blending the source face and the target face by mixing gradient, the results may be too similar to target face. Moreover, though the face relighting can simulate the illumination and shadow, it quite depends on geometry of model. The error of 3D reconstruction by morphable model will influence the correctness of illumination model.

Our system is not suited for certain cases. For example: wearing glasses, hair bangs, or other object occlusion on the face. Under above cases, it will usually generate visual artifacts in our results.

There are two avenues for our future work. First, we want to replace faces in video clips. We want to explore the temporal coherence of self-shadow and ambient occlusion. Second, we plan to use the Kinect to recover more precise face geometry to reduce the error in the face model synthesis module.

REFERENCES

[1] R. Basri and D. Jacobs. Lambertian reflectance and linear subspaces. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(2):218 – 233, feb 2003.

[2] S. Biswas, G. Aggarwal, and R. Chellappa. Robust estimation of albedo for illumination-invariant matching and shape recovery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31:884–899, 2009.

[3] D. Bitouk, N. Kumar, S. Dhillon, P. N. Belhumeur, and S. K. Nayar. Face Swapping: Automatically Replacing Faces in Photographs. *ACM Trans. on Graphics (also Proc. of ACM SIGGRAPH)*, Aug 2008.

[4] V. Blanz, C. Basso, T. Vetter, and T. Poggio. Reanimating faces in images and video. In P. Brunet and D. W. Fellner, editors, *EUROGRAPHICS 2003 (EUROGRAPHICS-03) : the European Association for Computer Graphics, 24th Annual Conference*, volume 22 of *Computer Graphics Forum*, pages 641–650, Granada, Spain, 2003. The Eurographics Association, Blackwell.

[5] V. Blanz, K. Scherbaum, T. Vetter, and H.-P. Seidel. Exchanging faces in images. In M.-P. Cani and M. Slater, editors, *The European Association for Computer Graphics 25th Annual Conference EUROGRAPHICS 2004*, volume 23 of *Computer Graphics Forum*, pages 669–676, Grenoble, France, 2004. Blackwell.

[6] V. Blanz and T. Vetter. A morphable model for the synthesis of 3d faces. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '99, pages 187–194, New York, NY, USA, 1999. ACM Press/Addison-Wesley Publishing Co.

[7] Q. Chen, H. Wu, T. Fukumoto, and M. Yachida. 3d head pose estimation without feature tracking. *Automatic Face and Gesture Recognition, IEEE International Conference on*, 0:88, 1998.

[8] Y.-T. Cheng, V. Tzeng, Y. Liang, C.-C. Wang, B.-Y. Chen, Y.-Y. Chuang, and M. Ouhyoung. 3d-model-based face replacement in video. In *SIGGRAPH '09: Posters*, SIGGRAPH '09, pages 29:1–29:1, New York, NY, USA, 2009. ACM.

[9] Y.-Y. Chuang, B. Curless, D. H. Salesin, and R. Szeliski. A bayesian approach to digital matting. In *Proceedings of IEEE CVPR 2001*, vol-

ume 2, pages 264–271. IEEE Computer Society, December 2001.

[10] P. Fitzpatrick. Head Pose Estimation Without Manual Initialization.

[11] J. Jia, J. Sun, C.-K. Tang, and H.-Y. Shum. Drag-and-drop pasting. *ACM Transactions on Graphics (SIGGRAPH)*, 2006.

[12] R. Kumar, M. Jones, and T. K. Marks. Morphable reflectance fields for enhancing face recgonition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.

[13] V. Kwatra, A. Schödl, I. Essa, G. Turk, and A. Bobick. Graphcut textures: image and video synthesis using graph cuts. *ACM Trans. Graph.*, 22:277–286, July 2003.

[14] P. Pérez, M. Gangnet, and A. Blake. Poisson image editing. In *ACM SIGGRAPH 2003 Papers*, SIGGRAPH '03, pages 313–318, New York, NY, USA, 2003. ACM.

[15] M. B. Vatahska, T. and S. Behnke. Feature-based head pose estimation from images. pages pp.330–335., 2007.

[16] Y. Wang, Z. Liu, G. Hua, Z. Wen, Z. Zhang, and D. Samaras. Face re-lighting from a single image under harsh lighting conditions. In *CVPR'07*, pages –1–1, 2007.

[17] Z. Wen, Z. Liu, and T. Huang. Face relighting with radiance environment maps. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 2, pages II – 158–65 vol.2, june 2003.

[18] M. H. Xuan Zou, Josef Kittler and J. R. Tena. Robust albedo estimation from face image under unknown illumination. In *Proc. SPIE 6944, 69440A*, page doi:10.1117/12.778599, 2008.

[19] Y. Zhou, L. Gu, and H.-J. Zhang. Bayesian tangent shape model: Estimating shape and pose parameters via bayesian inference. In *Proceedings of the 2003 IEEE computer society conference on Computer vision and pattern recognition*, CVPR'03, pages 109–116, Washington, DC, USA, 2003. IEEE Computer Society.

Figure 7: More results.