

國立臺灣大學電機資訊學院資訊工程學系

碩士論文

Department of Computer Science and Information Engineering

College of Electrical Engineering and Computer Science

National Taiwan University

Master Thesis

人臉置換下臉部自身陰影及環境光遮擋的修復

Self-Shadow and Ambient Occlusion Recovery for Face Images in

Face Replacement



熊士豪

Hsiung Shih-Hao

指導教授：歐陽明 博士

Advisor: Ouhyoung Ming, Ph.D.

中華民國 100 年 6 月

June, 2011



# 致謝

碩士兩年的時光，即將在這裡劃下一個段落。兩年的時間說長不長，說短不短，回首兩年的時光，這一次不再像過去是以眨眼即逝來形容時間的短暫，因為兩年的累積依舊歷歷在目。很慶幸自己身處在一個充滿溫暖、熱情、且養分充足的實驗室，在這裡所看到、所聽到、所學習、所仿效的都使得自己獲益良多，這樣的成長，不僅是在學術，更在做事情的態度，人生的智慧上有著更深厚的累積。在論文的開始，我要先感謝給予這篇論文莫大幫助的人們，因為有他們的協助和鼓勵，我才得以走到這一步。

首先感謝我的指導教授歐陽明老師，老師給予我在研究上很大的自由度，但在這樣的自由度下卻又不忘給予最有效的建議和適時的約束，還有鼓勵。還要感謝CYY和大炳老師，在每個禮拜與我們meeting過程中，對於我們在學術上的瓶頸和困惑傾囊相授，使我們可以更加廣泛且迅速的吸收更多的新知。感謝實驗室博班的學長們，感謝葉老大，在大小瑣碎事上給予協助，而且永遠義不容辭面帶笑容。感謝金哥，幫我解開許多學術上的困惑並且給我方向。感謝桓桓，沒有你的template和latex教學，要生出這麼精美的論文真是難上加難。感謝魁魁，雖然很像管家婆，但是沒有你在，我想生活將是一片慌亂。感謝winble，總是關心我們在實驗室的生活。感謝Graphic組其他博班學長姊，讓我在實驗室總是有家的感覺。

感謝碩二一起奮戰的夥伴們，一姐、昆哥、星、竺田、孫子、九皮、雅評、江胖、點點、歐維斯、雙胞、一零，沒有戰友和玩伴的碩士生活是苦悶的，因為有大家的陪伴，所以生活才會充滿著點滴，因為有這些點滴，才使得我又能重新注入活力繼續往前走下去，而且走得很勇敢、很充實。當然還有特別感謝我老大，說實話，沒有你的幫忙和協助，我相信不會有這篇論文的產生，你也讓我看到學習了很多，很高興能與你共同奮戰這兩年。還有其他沒提到的碩二同袍、碩一學弟，飽受我和老大摧殘的專題生們，你們在我兩年的生命中，也一樣重要。最後感謝家人和女友，我想有些時候有些事情是需要默默的支持與陪伴的，如果不是你們，我或許不會做得那麼好。我想...這不是我一個人走出來的路，上面也有著你們足跡，才使得得這條人生的路更平整、更寬廣。

兩年的汲取，是不斷的付出與跌倒，才有豐沛的收穫。也因為有你們，才不至於白跌倒，就算真的跌倒了也有你們在旁邊摻著我站起來。一切的一切，只有感謝。

熊士豪 謹誌  
民國一百年七月



# 中文摘要

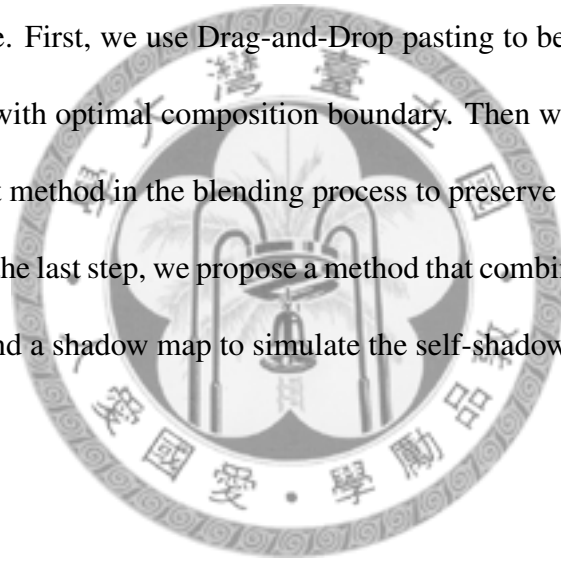
在本論文中，我們提出了一套在任意光照環境下，模擬出人臉陰影(self-shadow)及環境光遮擋(ambient occlusion)的人臉置換系統。藉由結合陰影圖(shadow map)和重新打光(relighting)的臉部影像，我們重新模擬出類似目標人臉的光影情況。並且使用無接縫影像結合演算法(Drag-and-Drop Pasting)找出最佳接縫邊緣，以維持較好的臉部特徵結果。最後，我們在混合兩張人臉時，採用影像強度梯度的混合(mixing gradients)去保存表情細節，使得最後結果與目標人臉的表情更加一致也更加生動。





# 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 image. First, we use Drag-and-Drop pasting to better maintain the facial saliency with optimal composition boundary. Then we incorporate the mixing gradient method in the blending process to preserve the vivid expression details. In the last step, we propose a method that combines the relighting source image and a shadow map to simulate the self-shadow condition in the target face.





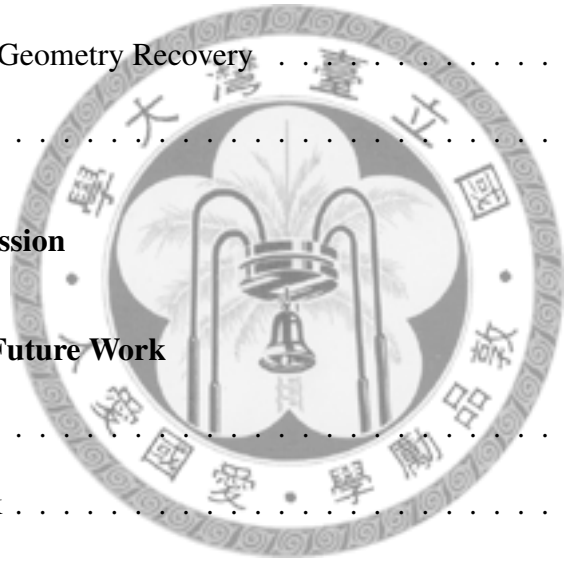


# Contents

致謝	i
中文摘要	iii
Abstract	v
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 System Overview . . . . .	2
1.3 Thesis Organization . . . . .	4
<b>2 Related Work</b>	<b>5</b>
2.1 Face Replacement . . . . .	5
2.2 Pose Estimation . . . . .	6
2.3 Relighting . . . . .	7
2.4 Seamless Composition . . . . .	7
<b>3 Pose Estimation</b>	<b>9</b>
3.1 Estimate pose . . . . .	10
3.2 Overlapping Region . . . . .	14
3.3 Result . . . . .	14



<b>4</b>	<b>Lighting Estimation</b>	<b>17</b>
4.1	Spherical Harmonics Relight . . . . .	19
4.2	Robust Albedo Relighting . . . . .	22
4.3	Self-Shadow in Face Image . . . . .	24
4.4	Result . . . . .	26
<b>5</b>	<b>Face Composition</b>	<b>31</b>
5.1	Poisson Image Editing . . . . .	32
5.2	Optimizing Blending Boundary . . . . .	33
5.3	Target Face Geometry Recovery . . . . .	36
5.4	Result . . . . .	37
<b>6</b>	<b>Result and Discussion</b>	<b>41</b>
<b>7</b>	<b>Conclusion and Future Work</b>	<b>45</b>
7.1	Conclusion . . . . .	45
7.2	Future Work . . . . .	46
	<b>Bibliography</b>	<b>49</b>



# List of Figures

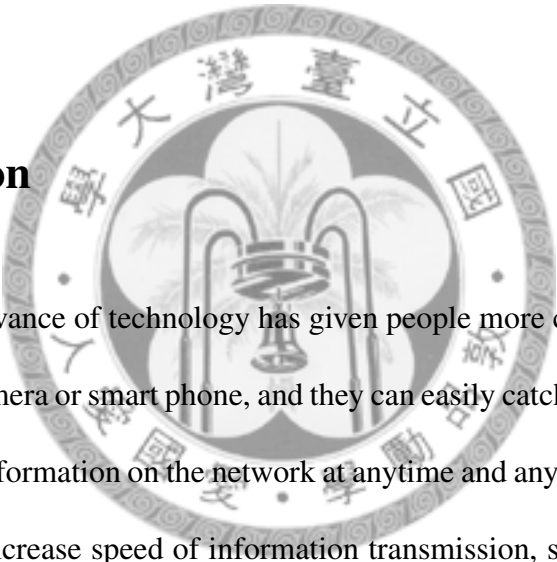
1.1	Pipeline of our face replacement approach. . . . .	3
3.1	To estimate pose in a single image. (a) We translate source model for giving initial guess. (b) Estimation result on the halfway (c) Final result for pose estimation . . . . .	10
3.2	We select the features including four corners of eyes, eight surroundings of nose, two corners of mouth, and one on chin. . . . .	12
3.3	Target face without deciding location of face. (a) Initial guess for Andy Lau (b) Pose estimation result of Andy Lau (c) Initial guess for Jackie Chen (d) Pose estimation result of Jackie Chen . . . . .	14
3.4	Pose estimation result. (a) Initial guess for Bruce Lee (b) Pose estimation result of Bruce Lee (c) Initial guess for Jet Li (d) Pose estimation result of Jet Li . . . . .	15
3.5	Pose estimation result. (a) Initial guess for Jackie Chan (b) Pose estimation result of Jackie Chan (c) Initial guess for Ethan Ruan (d) Pose estimation result of Ethan Ruan . . . . .	15
4.1	Composition results without relighting source face. We show the source face, target face, and composition result. (a) Bruce Lee’s result (b) Jackie Chen’s result (c) Ethan Ruan’s result . . . . .	18
4.2	Spherical harmonic in face images. . . . .	21
4.3	The error function and the corresponding influence function. First row are quadratic function and its influence function. Second row are lorentzian function and its influence function . . . . .	23
4.4	Generating shadow map by the principle light estimation. (a) Target face (b) Synthesized source face model with the same pose(c) After adding self-shadow on the face model. . . . .	25
4.5	The third face is relighting first source face as the same illumination as the second target face, and the red ellipses circle the visible artifacts. . . . .	27
4.6	Delighting results are calculated by that source face divide illumination model. . . . .	28
4.7	First column is target image. Second column is relighting result without robust albedo. Third column is relighting result with robust albedo. . . . .	28

4.8	(a) Delighting results without robust albedo algorithm. (b) (c) (d) (e) Result with different control parameter. . . . .	30
5.1	Replacement areas are surrounding facial features. . . . .	32
5.2	Image composition get unnatural blurring artifacts. . . . .	33
5.3	Left image is relation between original replaced region and region of full face. Right image is enlarge region near the cut. . . . .	34
5.4	Face composition results and corresponding replacement areas. . . . .	37
5.5	Comparing results with optimal boundary or not. First and second column are original results. Third and fourth column are results with optimal boundary. . . . .	39
6.1	Results of normal lighting condition. (b) The target image of Andy Lau. (c) Face replaced result. (d) The target image of Bruce Lee. (e) Unusual skin color on relighting source image. (f) Face replaced result. . . . .	42
6.2	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 . . . . .	42
6.3	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. . . . .	43
6.4	More result. (a) The target image of Takeshi Kaneshiro (b) The final result of Takeshi Kaneshiro (c) The target image of Aaron Kwok (d) The final result of Aaron Kwok (e) The target image of Jet Li (f) The final result of Jet Li (g) The target image of Jacky Cheung (h) The final result of Jacky Cheung (i) The target image of Chow Yun-fat (j) The final result of Chow Yun-fat (k) The target image of Tom Cruise (l) The final result of Tom Cruise (m) The target image of Andy Lau (n) The final result of Andy Lau (o) The target image of Mark Chao (p) The final result of Mark Chao (q) The target image of Rain (r) The final result of Rain (s) The target image of Ekin Cheng (t) The final result of Ekin Cheng . . . . .	44

# Chapter 1

## Introduction

### 1.1 Motivation

The logo of National Sun Yat-sen University is a circular emblem. It features a central design with a sunburst and a book, surrounded by the university's name in Chinese characters: '國立臺灣大學' (National Sun Yat-sen University) at the top and '人文科學' (Humanities and Sciences) at the bottom. The emblem is overlaid on the text of the first paragraph.

In recent years, the advance of technology has given people more convenient life. People almost have digital camera or smart phone, and they can easily catch photo or record video to share their digital information on the network at anytime and anywhere. In addition, the term Web2.0 is also increase speed of information transmission, so people can get more and more digital multimedia on the network than before. The interesting issues are editing and modifying digital multimedia to get personalization results from a great amount of information. One of those issues is face replacement. We replace someone's face to the subject face of target images or videos in existence, and replaced videos or images are realistic and natural. This technology is called face replacement. Face replacement is widely used to visual effects of film industry. For examples, Beowulf, The Curious Case of Benjamin Button, Avatar, The Lord of the Rings, they all use the face replacement skill to edit the character in the movies, so the virtual actor may live in those movies. This technique should be applied to entertainment for general public too, so people can easily

edit images or videos to replace actor's face with their face.

This master thesis is based on the National Science Council (NSC) project. The objective of NSC project is replacing actor's face of old and classical film with others. Those films always have several characteristics. First, we don't know the status of the scene, like illumination. Second, the film only have 2-D information, we cannot estimate position of actor's face easily. Third, because the film has hundreds of thousands of frame, it will spend too much time to modify and edit the film frame by frame. We need to exploit a full or half automatic system to precisely get realistic results for face replacement. In this master thesis, we only aim to a single image, so we don't care about temporal coherence. We focus our problem to the three issues, include of pose estimation, face relighting, and seamless image composition. Our system is compared to previous work. We process self-shadow more effective. The concept of adding shadow map and Drag-and-Drop Pasting is used to improve unnatural artifacts. Our system may not save execution time, but it can automatically replace face from source model to target face and has decent results.

## 1.2 System Overview

In our input, two images may not have the same illumination and head pose, and the characters in the images are also different people with dissimilar skin and facial expression. We design a series of modules to process those situations, as given in Figure 1.1.

The source and target head models are synthesized by morphable model first. Face alignment module detect facial features which are surrounding facial features and silhouette in the target subject face. Those features can help us to identify position of eyes, nose and mouse, and they also can find face region in the target image. Pose estimation module helps us to find the most suitable pose for face composition. It guarantees that

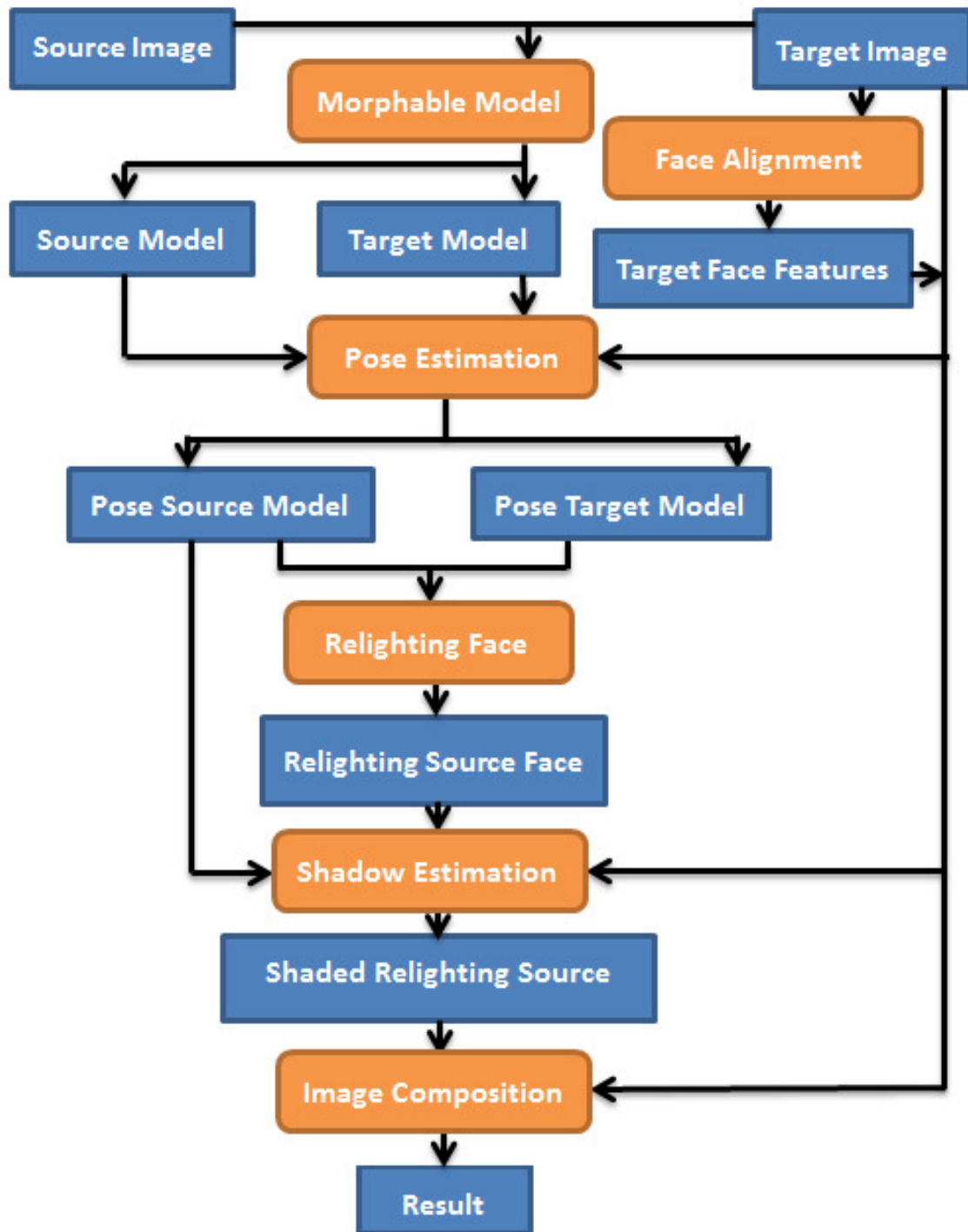


Figure 1.1: Pipeline of our face replacement approach.

the facial features in the source model fit the facial features in the target image. Target and source image are often taken from different scenes, so most of our inputs have different lighting environments. We need to relight the source face to the same lighting condition of the target image. If the target image has hard self-shadow, we estimate the position of the principle light source and add the shadow map to our relighting face, so it can simulate the special lighting condition in the target image. Finally, we use the concept of Drag-and-Drop pasting to get a seamless and natural skin color result. And we use a mixing gradient to keep the facial saliency in both source and target images.

### **1.3 Thesis Organization**

The organization of this paper is as follows: Chapter 2 is the related work about face replacement, pose estimation, relighting, and seamless composition. Chapter 3 presents how we put our source model to the current position. Chapter 4 presents a method for estimating lighting conditions in target images and simulating an illumination model. Chapter 5 introduces a seamless composition algorithm and presents a method to avoid some artifacts. Chapter 6 presents results and discussion. Chapter 7 is the conclusion and proposes some possible work in the future.



# Chapter 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 target face is synthesized by 3D model.

One of image-based methods is [3]. [3] uses a large database of face images which are created from the internet. They select some candidates in the database and computing candidate's pose, lighting, and color, then blending with the target images. This method does not estimate pose on 3-dimension, and they favor to replace the target faces with similar ones in the database. [6] provides a 3D morphable model 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 forming the vector space of mouth shapes and facial expressions in order to capture mouth movement. [5] apply the rendering parameters estimated in the target image to the synthesis source face model for face exchanging. [8] presents a system that replaces the target subject face in a video. They clone the expression of the target video to the

synthesis source face model from a 3D face expression database, and enforce temporal consistency for illumination and face poses. Our face replacement framework is similar to those model-based methods. We focus on the improvement of two aspects: illumination simulation, and seamless composition. And our replace results are more natural and vivid.

## 2.2 Pose Estimation

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

Image-based methods: [16] consider this problem as a classification problem. They extract distinctive facial features from Haar-like features to determine the optimal pose range with classifiers. Other image-based techniques are based on the facial structure, and they often assume the head pose which is closed to a frontal presentation. [7] uses skin region and hair region of heads to estimate head pose. All of 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. [8] projects the face model to the target image, and estimate the optimal head pose by minimizing the difference between the feature position of the projected face model and the target image in the same region. [5] and [4] estimate lighting condition and pose at the same time to handle different illumination. Instead of using color information, [10] extracts contours and feature points of face in the target image. They map those features onto the model and also consider the temporal consistency between near frames. The problem of using feature points only is that the relation between facial form and facial features may be destroyed. We synthesis the source face by the morphable model in our system, and we combine

intensity and feature information to estimate pose. So the facial relation and illumination variation can be both preserved.

## 2.3 Relighting

It is difficult to estimate position and direction of all lights in the single image, so most traditional techniques for face replacement used finite linear subspace to synthesize the relighting image. [1] and [18] suppose that head is a convex Lambertian object under distant and isotropic light, and approximate the lighting conditions with spherical harmonics. To handle objects which may have different materials in different area, [12] proposes a ratio image technique to remove material dependency in the radiance environment map. [17] and [2] design new frameworks to refine the surface normal and albedo for using spherical harmonics. [2] tries to segment human face into several parts, determines surface normal and albedo for each part, and iteratively computes surface normal, albedo, and lighting. [19] incorporates spatial coherence constraint of albedo.

In our system, we refer [19] to get robust albedo and precise spherical harmonic coefficients. Finally we add the shadow map 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] finds the optimizing boundary to minimize color variation and the replacing area. [9] aims to estimate opacity for each pixel of foreground element which is extracted from background image. Both of [9] and [13] cannot handle the case with different skin colors. [15] proposes a blending method to

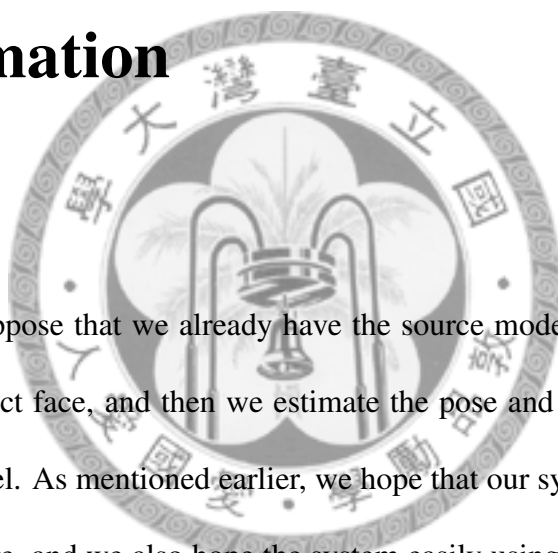
overcome this problems. It keeps gradient variation from the source image and overlap transparent source image into a replaced area. Compositing two images is by the Poisson equation. But the results from [15] too rely on the users-specified destination area. [11] proposes an object function to compute an optimizing boundary. They also construct a blended guidance field to faithfully preserve the object's fractional boundary.

We use the Drag-and-Drop pasting [11] to better preserve natural shading feature in the target face, especially ambient occlusion and other facial saliency.



## Chapter 3

### Pose Estimation



In our system, we suppose that we already have the source model which is synthesized from the source subject face, and then we estimate the pose and illumination model by using the source model. As mentioned earlier, we hope that our system can be applied to the videos in the future, and we also hope the system easily using for users. So the most of our system instructions will be full-automatic to save time for user. The instruction of pose estimation is also need to achieve the goal. If we replace face just in a image, we may adjust the pose of subject face as the same as the target image easily. We only transform the position of the source model to fit the subject face which is in the target image, and the source face is located to the same pose like target face. But we can't do the same thing for videos which are combined hundreds of images at least, we need a half or full automatic algorithm to estimate poses precisely. In the following section, we introduce a semi automatic head pose estimation method, and We combine the intensity coherence and geometry coherence to estimate the head pose.

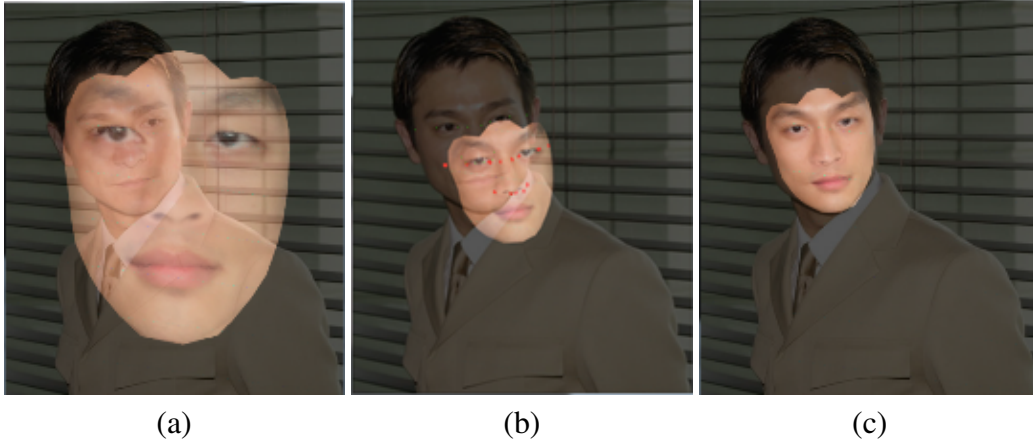


Figure 3.1: To estimate pose in a single image. (a) We translate source model for giving initial guess. (b) Estimation result on the halfway (c) Final result for pose estimation

### 3.1 Estimate pose

Pose estimation is used to find the same pose between a source face model and a target face image. How do the people ensure the same pose between a source face model and a target face image? People try to overlap the source model and the target image, and compare the overlapping region. If the overlapping region has the most similar color and the same facial features, people could usually consider that the source face model has similar pose with the target image. Based on the above ideas, We minimize the Euclidean distance over all pixels and three color channels in the overlap region between  $I_{target}$  and  $I_{model}$ :

$$E_I = \sum_{u,v} (I_{target}(u,v) - I_{model}(u,v))^2$$

, where  $I_{target}$  is color of target image and  $I_{model}$  is the image by projecting source model to the image plane about pose parameters. We use term  $E_I$  to confirm that the result has the most similar color in the overlapping region. Pose parameters include three of rotations and three of the translation. We don't have to estimate scale, because it will be implied at translation of z axis. Since transformation of objects can reach the same result

by moving camera, we just transform model and assume that camera can't move.

$$\begin{bmatrix} x_i' \\ y_i' \\ z_i' \end{bmatrix} = R_\alpha R_\beta R_\gamma \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} + T$$

is the rigid model transformation, where  $\alpha$  is yaw,  $\beta$  is pitch, and  $\gamma$  is roll.  $T$  is translation matrix. We transform model from position  $(x_i, y_i, z_i)$  to new position  $(x_i', y_i', z_i')$ , and then we project model to 2D-plane by using fixed focal length and perspective view.

$$\begin{bmatrix} u_i' \\ v_i' \end{bmatrix} = \begin{bmatrix} \frac{I_{width}}{2} \\ \frac{I_{height}}{2} \end{bmatrix} + f \begin{bmatrix} \frac{x_i'}{z_i'} \\ \frac{y_i'}{z_i'} \end{bmatrix}$$

,  $(u_i, v_i)$  is pixel position in image coordinate.  $I_{width}$  and  $I_{height}$  are size of image.  $f$  is focal length. So projection image can be represented as:

$$I_{model}(u, v) = P(R_\alpha R_\beta R_\gamma M(x, y, z) + T)$$

,  $P$  is perspective function like above equation.  $M(x, y, z)$  is position of model vertex. If both of the source and the target face have the same lighting condition, we might estimate good results by above function. Unfortunately, general scenes mostly have different lighting condition, it can cause different shadow and saturated region from different illumination. Both the shadow and saturated region generate untrue skin color of objects, so the overlapping region with different lighting condition can affect correctness of pose estimation. Although we control lighting condition of source face, we cannot ask target images, even target videos, have frontal uniform lighting condition. We exploit additional

feature term in order to get high accuracy results and avoid above mentioned problem.

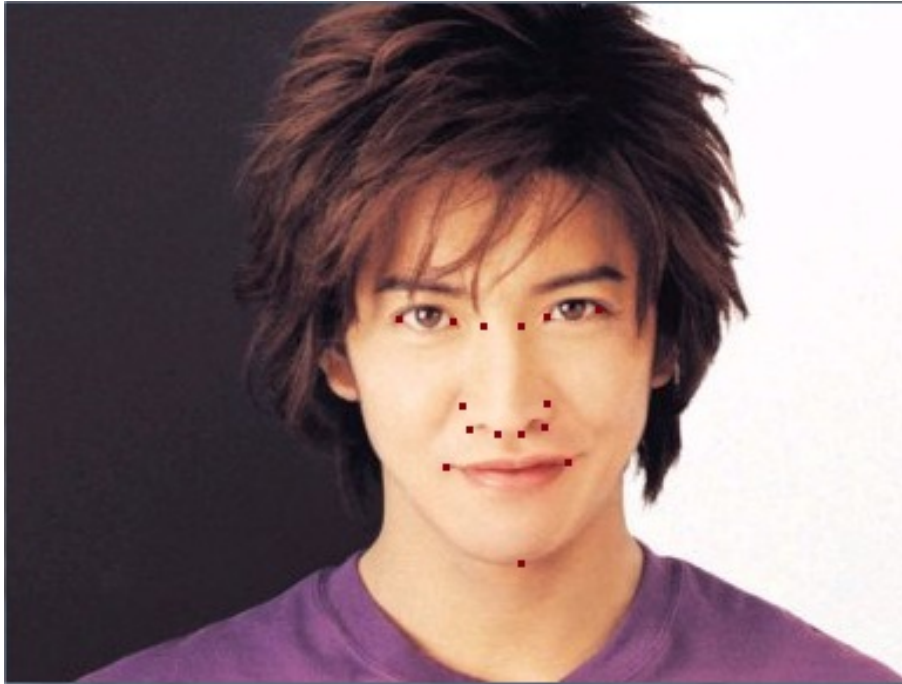


Figure 3.2: We select the features including four corners of eyes, eight surroundings of nose, two corners of mouth, and one on chin.

We select features from the target face by face detection and face alignment [20]. Face detection can find position of subject face in the target image. Face alignment mark points along silhouette and around facial features. Not all of features are selected from face alignment. Some features position will change with different facial expressions. For example, the features at the corners of mouth will rise when people is smile. The features around eyebrows will wrinkle up when people is scowl. Even though silhouette shape is also different between frontal face and side face, silhouette is sharper on side face than on frontal face. We don't want those conditions to affect precision of estimating pose. When we estimate pose, we need nearly consistent features in different facial expression and head pose. So we select feature points which are not easy to change with different facial expressions, include: four corners of eyes, eight surroundings of nose, two corners of mouth, and one on chin (Figure 3.2). After we have selected features points on the target



face, we manually select vertexes on the source model in addition. Those vertexes are one-to-one and correspond to features on the target face. The feature term can be set 2-norm distance in pairs of the target image  $(x_i, y_i)$  and source model feature points  $(v_i, u_i)$ :

$$E_F = \sum_i \left( \begin{pmatrix} x_i \\ y_i \end{pmatrix} - \begin{pmatrix} u_i \\ v_i \end{pmatrix} \right)^2$$

According to our definition of good pose, we need that the overlapping region gets smallest variance of color. So we combine the intensity term and feature term to guarantee precise pose estimation results. We exploit the minimized energy function:

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

We estimate pose parameters by minimizing equation. No matter initial pose is precisely given or not, it can automatically compute result by iteration. Since equation is hard to analysis from partial derivation, we choose NM simplex (Nelder-Mead simplex) method to minimize our error function. NM simplex is a heuristic. Nelder-Mead algorithm extrapolates the behavior of the error function by arranging parameters as simplices. In each iteration step, it chooses to replace one of the simplices with a new simplex, if new simplex can make function smoother than current simplex and is unimodal. On the other hand, if new simplex cannot get much better than current simplex, it will compress the position of simplices. The simplex converges towards a better result. The NM simplex multidimensional minimization method has been provided in GNU Scientific Library. It just needs to set form of the function, initial guess, and size of the initial trial steps to optimize energy function.

## 3.2 Overlapping Region



Figure 3.3: Target face without deciding location of face. (a) Initial guess for Andy Lau (b) Pose estimation result of Andy Lau (c) Initial guess for Jackie Chen (d) Pose estimation result of Jackie Chen

We can confirm overlapping region of source model which is projected region on 2-D plane. We also need to confirm the overlapping region of the target face. If we only decide the overlapping region of the source model, we might get the wrong pose.(Figure 3.3) Since the pose parameters need to get the smallest error in the error function, the pose reduces projection area as less as possible. However, if we only use overlapping region of source model, we could get the pose of side face model to overlap target face 3.3.

The overlapping region of the target face is decided by the face alignment. We take the range surrounding silhouette and eyebrow features to decide the position and the area of the target face. The model projection region and target face region are selected, so we can compute the intensity term for all pixels which exist on the model projection and the target face region.

## 3.3 Result

In figure 3.4, we give two images, Bruce Lee and Jet Li, whose faces have normal frontal lighting, and user give initial pose( figure 3.4(a) (c)). We can see that the source mod-

els are exactly overlapping the target subject face, and the facial features have the same position as each other.

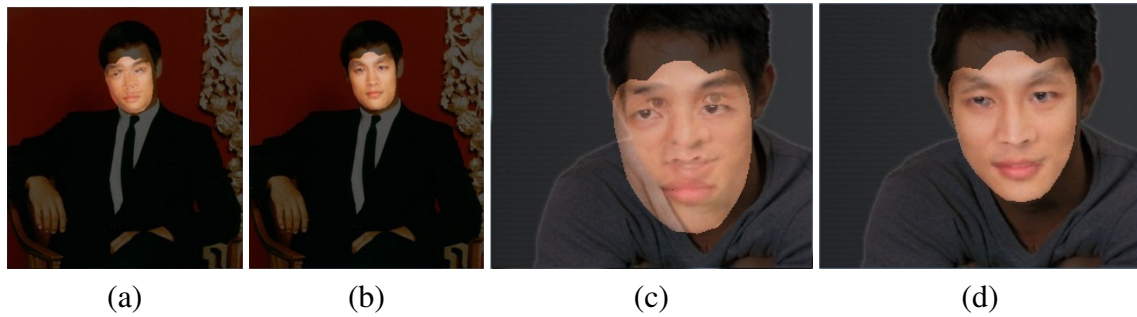


Figure 3.4: Pose estimation result. (a) Initial guess for Bruce Lee (b) Pose estimation result of Bruce Lee (c) Initial guess for Jet Li (d) Pose estimation result of Jet Li

We also test our algorithm for the face which has wide shadow region or saturated region (figure 3.5). Jackie Chan and Ethan Ruan both have shadow resulted in the left lighting. The initial poses are given by the user too, and we can see that the results of both images are mostly accurate.

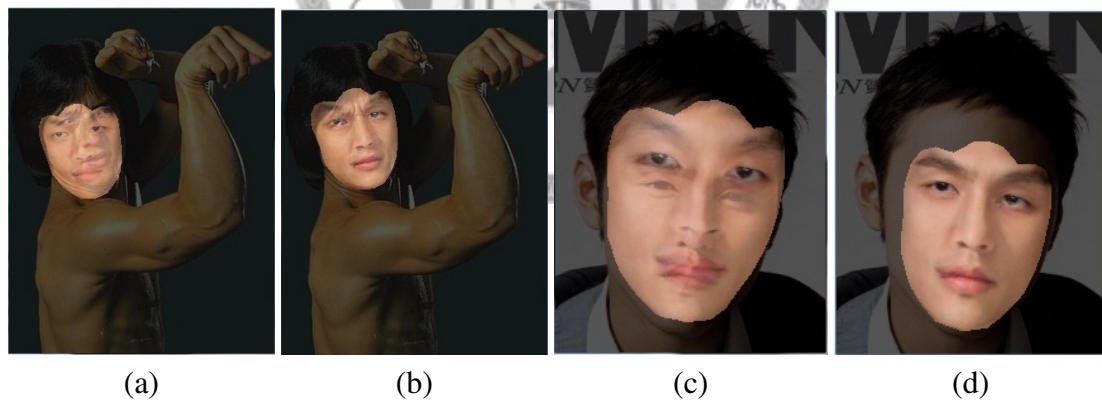
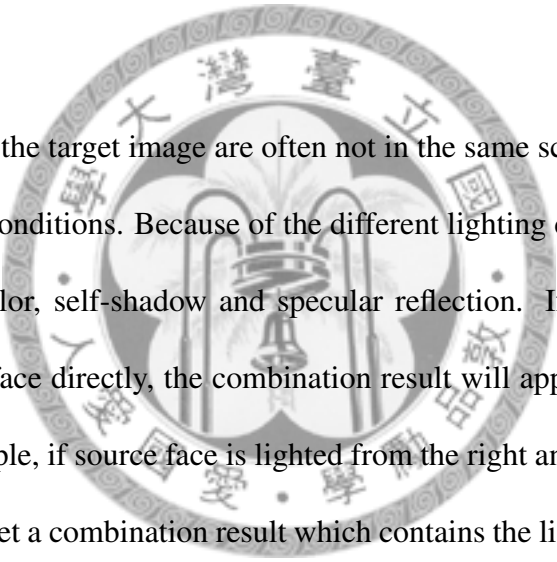


Figure 3.5: Pose estimation result. (a) Initial guess for Jackie Chan (b) Pose estimation result of Jackie Chan (c) Initial guess for Ethan Ruan (d) Pose estimation result of Ethan Ruan



# Chapter 4

## Lighting Estimation



The source image and the target image are often not in the same scene, so they may have the different lighting conditions. Because of the different lighting conditions, people will have different skin color, self-shadow and specular reflection. If we just combine the source and the target face directly, the combination result will appear unnatural artifacts (figure 4.1). For example, if source face is lighted from the right and target face is lighted from the left, we can get a combination result which contains the lighting condition of the source and the target face at the same time. It is not a good result, because we want to keep the lighting condition as the same as the target image. Although the skin color of target face is similar to the source face, different illumination still generates an unnatural result. We propose relighting methods to solve this situation.

But the relighting method by spherical harmonics cannot simulate hard self-shadow very well, we propose a method that combines the spherical harmonics relighting and the principle light source estimation. Although we have the geometry of the source and the target face, it is still too difficult to find all light sources in the scene. Because the deep self-shadow on the target face is often generated from a single principle light or several

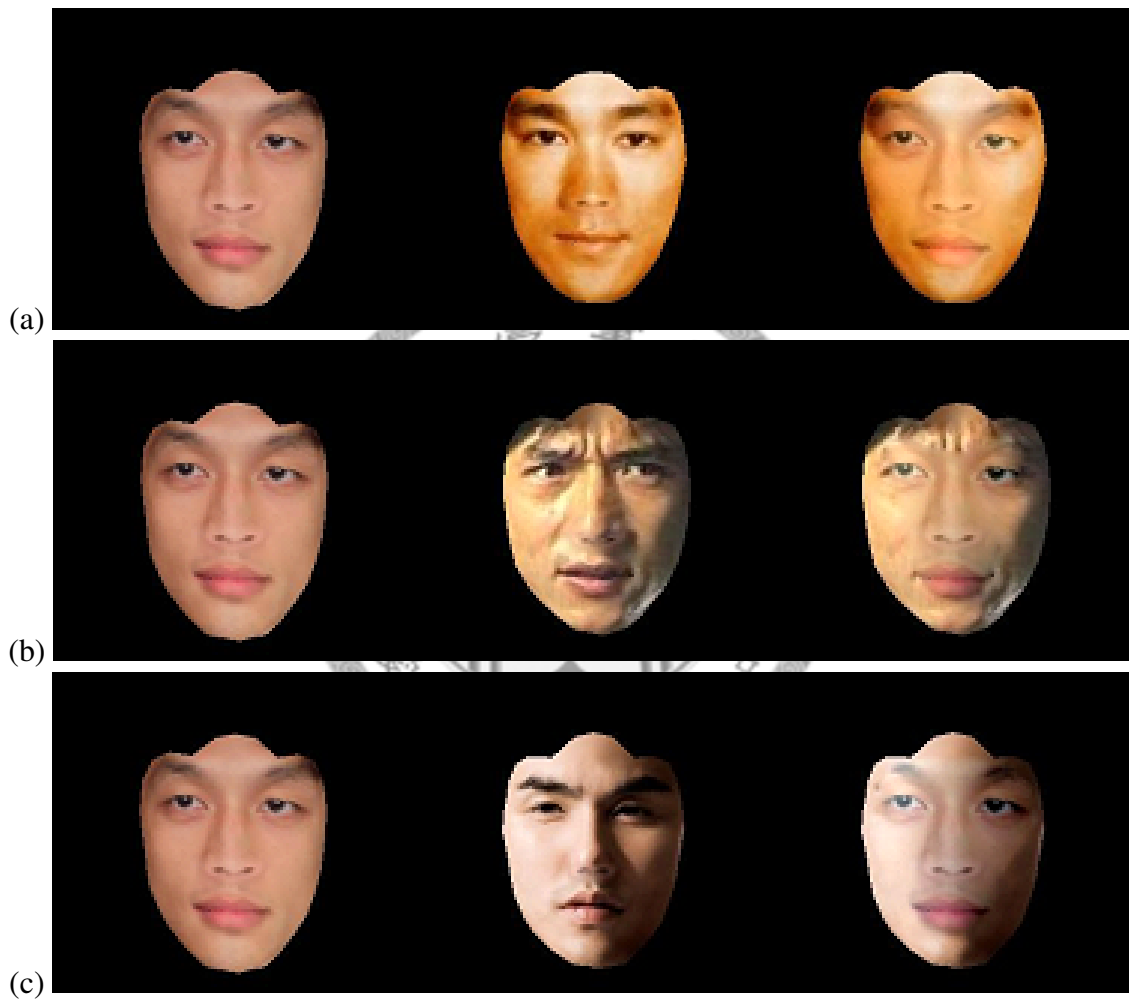


Figure 4.1: Composition results without relighting source face. We show the source face, target face, and composition result. (a) Bruce Lee's result (b) Jackie Chen's result (c) Ethan Ruan's result

lights in the closed location, we can approximate the lighting condition with less number of light sources. In this paper, currently we only estimate one principle light. We believe that 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 its edge becomes more soft or closed 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 principle light source to simulate harsh lighting conditions.

## 4.1 Spherical Harmonics Relight

In general, reflectance field can be separated two parts: illumination model and albedo. Reflectance generally refers to the fraction of incident radiation which is reflected at an interface, so it is usually represented as images. Albedo is also called reflection coefficient, and we can refer albedo to frontal uniform lighting object. Illumination models involve face surface reflectance, surface normal and illumination. It represent that lights achieve practical effect on objects. Relighting algorithm always fixes face surface reflectance and surface normal, and construct different lighting environment by 3D linear subspace combination or using concept of radiance environment map.

Spherical harmonics are a function of spherical coordinate, and it can approximate illumination model effectively by a low-dimensional linear subspace using the first 9 spherical harmonic bases. It has been widely used to construct illumination models. This approach assumes no cast shadow and saturation on objects, distant light source, and that objects are Lambertian convex surface. The Lambertian convex surface has a character-

istic: no matter the incident light arrives at surface in different angles, the luminance is the same to observers. We describe the relation between image and spherical harmonics model by the following equation:

$$I_{face}(\vec{n}) = \rho_{face}(\vec{n}) \sum_{i=0}^9 H_i(\vec{n}) \cdot L_i$$

,where  $I_{face}$  denote the image intensity,  $\vec{n}$  is the surface normal,  $\rho_{face}$  is the surface albedo,  $L_i$  is the lighting coefficient, and  $H_i$  is the spherical harmonic basis as follow:

$$\begin{aligned} H_1 &= \frac{1}{\sqrt{4\pi}}, & H_2 &= \sqrt{\frac{3}{4\pi}} n_z, & H_3 &= \sqrt{\frac{3}{4\pi}} n_y, \\ H_4 &= \sqrt{\frac{3}{4\pi}} n_x, & H_5 &= \frac{1}{2} \sqrt{\frac{5}{4\pi}} (3n_z^2 - 1), & H_6 &= 3\sqrt{\frac{5}{12\pi}} n_x n_z, \\ H_7 &= 3\sqrt{\frac{5}{12\pi}} n_y n_z, & H_8 &= 3\sqrt{\frac{5}{12\pi}} n_x n_y, & H_9 &= \frac{3}{2} \sqrt{\frac{5}{12\pi}} (n_x^2 - n_y^2) \end{aligned}$$

,where  $n_x, n_y, n_z$  is the x, y, z axis component of the surface normal  $\vec{n}$ . According to obtain by Basri and Jacobs[1], the linear combination of the above spherical harmonic bases (Figure 4.2) can be synthesize most arbitrary illumination condition for diffuse objects. Because we don't have the robust albedo, we use the skin detection [14] to get the area of the skin and calculate average value of skin color as albedo  $\rho_{face}$ . Skin color is detected in  $YC_bC_r$  color space. The corresponding skin cluster is given as:  $Y \geq 80, 127 \geq C_b \geq 77, 173 \geq C_r \geq 33$ . We can solve the above equation to get lighting coefficients of the source and the target image.

Finally, we can use quotient image formulation to relight image. Using following equation, we can get the illumination models of the source image and the target image:

$$\begin{aligned} E_{source}(\vec{n}) &= \sum_{i=0}^9 H_i(\vec{n}) L_{i_{source}} \\ E_{target}(\vec{n}) &= \sum_{i=0}^9 H_i(\vec{n}) L_{i_{target}} \end{aligned}$$



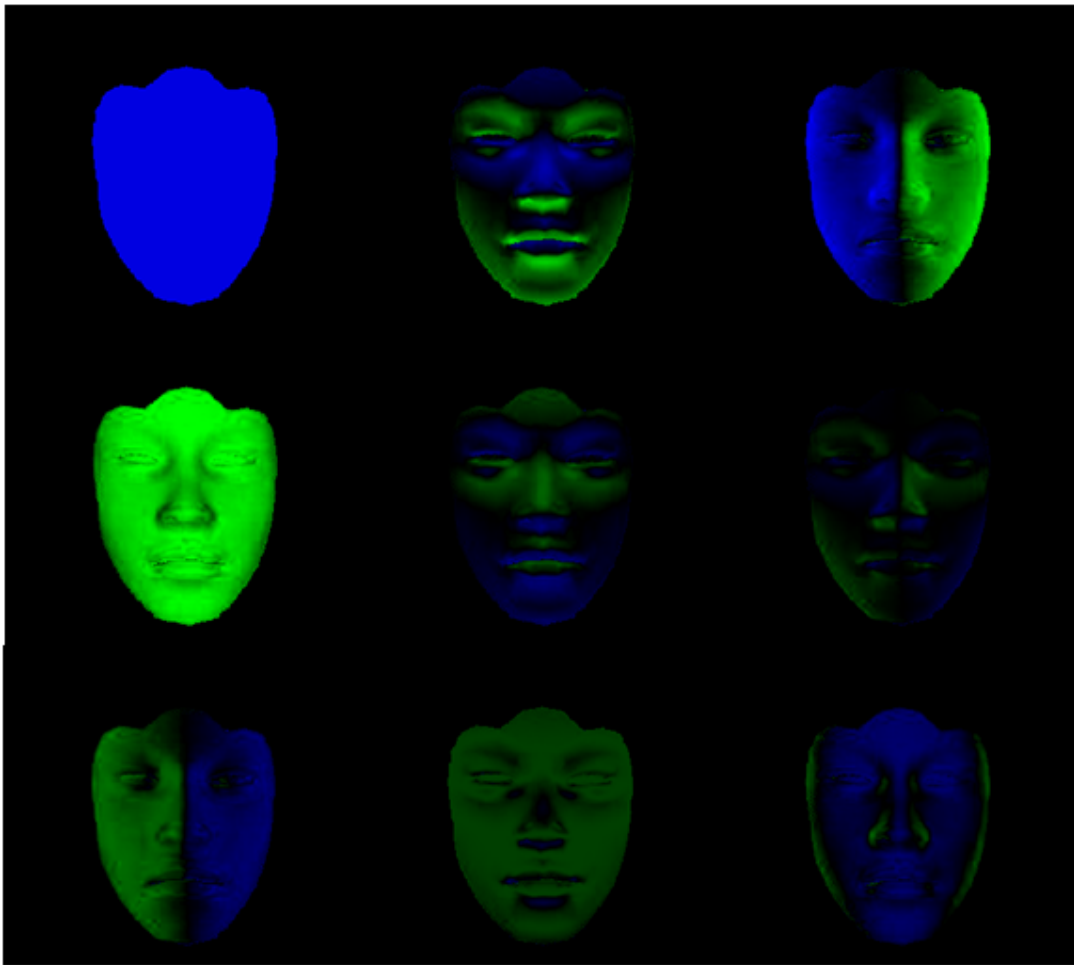


Figure 4.2: Spherical harmonic in face images.

, where  $E_{source}$  and  $E_{target}$  are the illumination models, and  $L_{i_{source}}$  and  $L_{i_{target}}$  are the lighting coefficient. Therefore, we can relight image by:

$$I_{relight}(\vec{n}) = I_{face}(\vec{n}) \times \frac{E_{target}(\vec{n})}{E_{source}(\vec{n})}$$

.When we have the relighting image, we can get more natural face combination results.

## 4.2 Robust Albedo Relighting

There are many methods to construct relighting image, but most of the methods always need robust albedo. Unfortunately, although the best delighting algorithm still could not recover the area which has cast shadow or saturated regions, it is difficult to delight image in hash lighting condition. We implement robust albedo estimation [19] for face relighting algorithm to get more robust albedo and relighting result. However, the first constraint always is conserving data term, so it can guarantee that a simulated face is similar as input data. Second, we know the albedo having the spatial coherence. Most skin color of a pixel in the albedo may have the same color with its neighbour. Adding this constraint to estimate illumination models may get better albedo. We modify the energy function as following:

$$\Gamma(\rho, L) = \sum_{j=1}^m [\lambda_d f_d((I_j - \rho_{face}(\vec{n}_j)) \sum_{i=1}^9 H_i(\vec{n}_j) \cdot L_i), \sigma_d) + \lambda_s [\sum_{n \in g_j} f_s(\rho_{face}(\vec{n}_n) - \rho_{face}(\vec{n}_j)), \sigma_s]]$$

,where  $\lambda_d$  and  $\lambda_s$  are weight for the data conservation term and spatial coherence term,  $g_j$  is the neighborhood of the  $j_{th}$  image pixel. The  $f_D$  and  $f_s$  are Lorentzian functions of the

form:

$$\rho(x, \sigma) = \log\left(1 + \frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right)$$

and  $\sigma_d, \sigma_s$  are parameters that control the shapes of the logarithm functions. Why do not take quadratic function as our error functions? As shown in figure 4.3, the influence function corresponding to the derivative of quadratic function is monotonically increasing function. Because the influence function is sensitive for outlier, we may get the high weight to effect the result. The outlier is generated from high differences between input image and estimation result, but not all of high differences are outliers. Since our initial guess of albedo is average skin color, the lighting estimation result will hold effect of illumination and lose details on the face. Differences could be increased for the location of face details, include: eyebrows, eyes, or mouth, and it will affect albedo estimation. For the purpose of reducing influence of excessive difference from the quadratic function, we take the logarithm function as our new error functions.

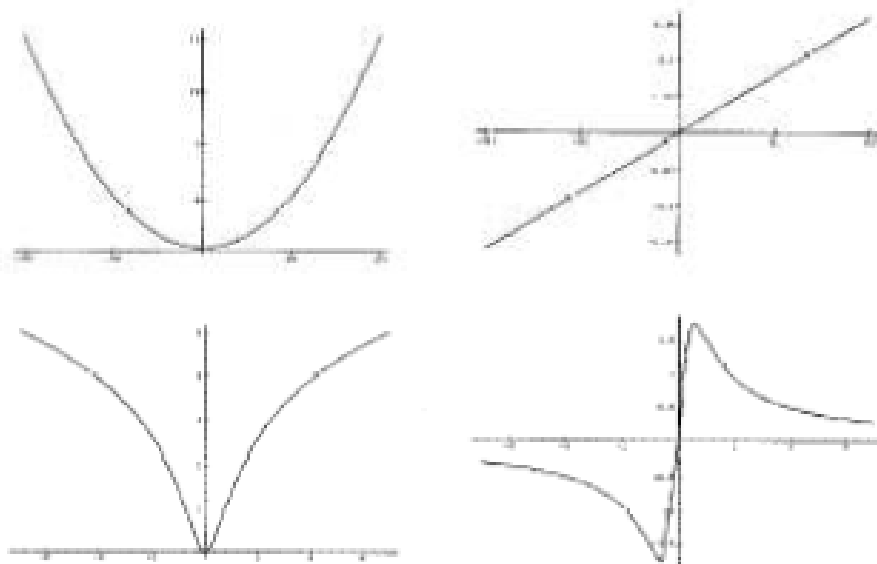


Figure 4.3: The error function and the corresponding influence function. First row are quadratic function and its influence function. Second row are lorentzian function and its influence function

We use the simultaneous over-relaxation approach to minimize this error function.

The initial albedo is still set to the average skin color, and initial lighting coefficients are calculated from the first time least square estimation given the initial albedo. Following is the robust albedo estimation algorithm:

1. Initialize  $\rho_{face}^0$  and  $L_r^0$
2. Compute the derivate of the logarithm function and the upper boundary of its influence function:  $\{\frac{\partial \Gamma}{\partial L_i}\}i = 1 \dots 9, \{\frac{\partial^2 \Gamma}{\partial L_i^2}\}i = 1 \dots 9, \{\frac{\partial \Gamma}{\partial \rho_i}\}i = 1 \dots n, \{\frac{\partial^2 \Gamma}{\partial \rho_i^2}\}i = 1 \dots n, n$  is image size.
3. Refine albedo :  $\rho_i^m = \rho_i^{m-1} - \omega \frac{1}{TL_i^{m-1}} \frac{\partial \Gamma}{\partial L_i^{m-1}}$ , where  $\omega$  is over-relaxation parameter,  $T(\rho_i^{m-1})$  is the upper boundary of the  $\{\frac{\partial^2 \Gamma}{\partial \rho_i^2}\}$ .
4. Refine lighting coefficient:  $L_i^m = L_i^{m-1} - \omega \frac{1}{TL_i^{m-1}} \frac{\partial \Gamma}{\partial L_i^{m-1}}$ , where  $\omega$  is over-relaxation parameter,  $T(L_i^{m-1})$  is the upper boundary of the  $\{\frac{\partial^2 \Gamma}{\partial L_i^2}\}$ .
5. Parameter update by:  $\sigma_d^m = k\sigma_d^{m-1}, \sigma_s^m = k\sigma_s^{m-1}$ , with  $k \in (0, 1)$

In our experiment, the parameter is setting as:  $\omega = 1.995, k = 0.95, \sigma_s = 5, \sigma_d = 10, \lambda_s = 1, \lambda_d = 15$ . And we set the neighbour size for the spatial coherence is  $3 \times 3$ . The control parameters  $\sigma_s, \sigma_d$  decreases to suppress outliers, and  $\lambda_s, \lambda_d$  is the weight for each term.

### 4.3 Self-Shadow in Face Image

Because the pose and geometry of source model are fixed, we only compute the position of principle light source. After we define the position of light source, we also can use simple method to generate fake shadow map. Given the shadow map and relighting image, the goal is to minimize the Euclidean distance over all pixels and color channels between

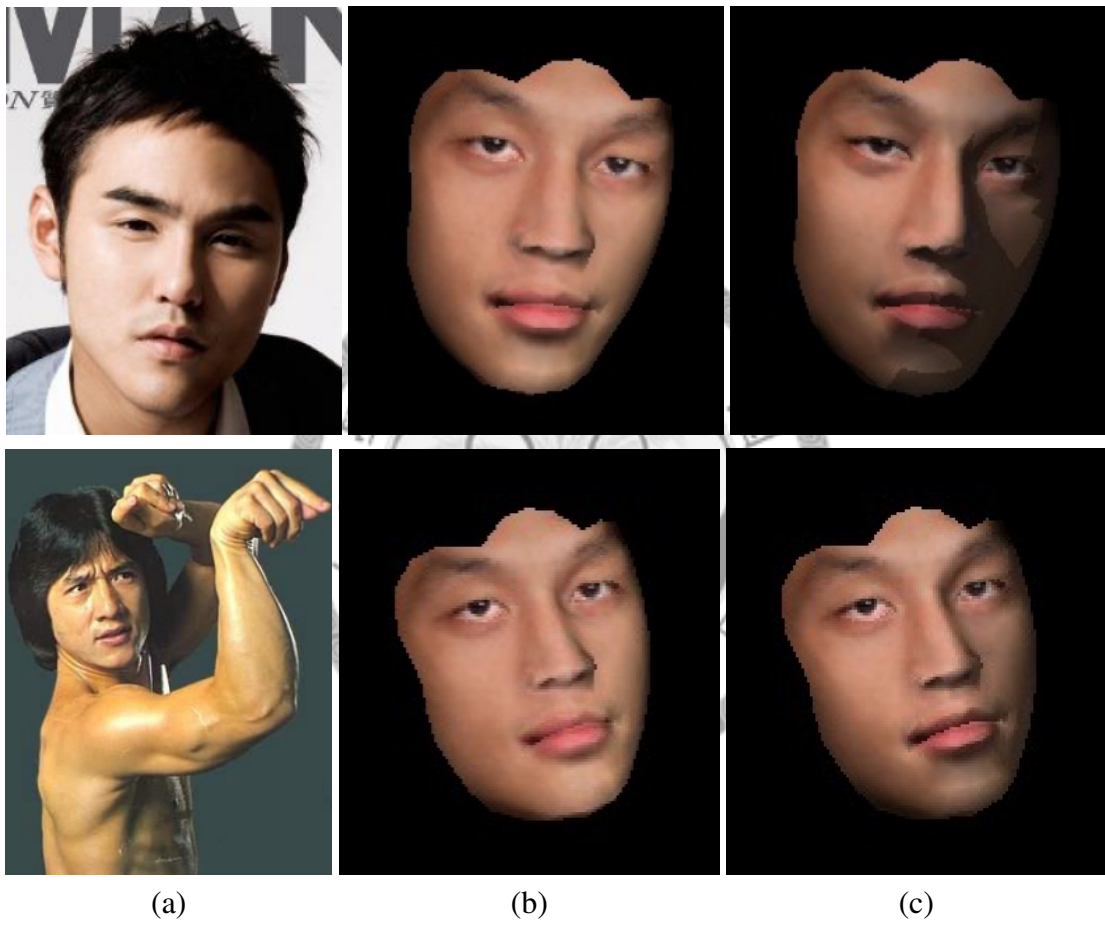


Figure 4.4: Generating shadow map by the principle light estimation. (a) Target face (b) Synthesized source face model with the same pose (c) After adding self-shadow on the face model.

target image and combined shaded relighting image. The energy function can be written as:

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

,where  $l_p$  is the lighting position.  $S()$  is the shaded face which is generated by the light on position  $l_p$ .  $I_{original}$  is a face without shadow and has the same light which locate at the  $l_p$ .  $I_{relight}$  is relighting source image by the spherical harmonics method.  $I_{target}$  is the target face. The shaded face is as Figure 4.4. We estimate position of the light source by minimizing this energy function. Initial lighting position is specified by a user or set to the default value, and then the result is automatically computed. Since the energy function is also hard to analysis from partial derivation, we choose NM simplex (Nelder-Mead simplex algorithm) to minimize it too.

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.

## 4.4 Result

Our relighting algorithm depends on normals of geometry. We use the information of morphable model to calculate normals for all pixels in the region of face. So if the model is similar input subject face, the relighting face will be estimated more realistic. Using morphable model is not sure to synthesis input subject faces well. Since morphable model is symmetry, it is hard to synthesis asymmetric input subject faces. When input face has asymmetry expression, for example: rise up eyebrow in one side and get down another side, it is hard to synthesis face exactly alike. The details of the model cannot perfectly

mapping texture from the source face. It can cause the visible artifacts after relighting. As shown in figure 4.5, after Subject A's is relit to the same illumination environment of Jacky Cheung, the shape of Subject A's eyes and mouse is different to the original input image. The shape is modified as the same as synthetic model, and it even lost original characteristic of Subject A's face.



Figure 4.5: The third face is relighting first source face as the same illumination as the second target face, and the red ellipses circle the visible artifacts.

In section 4.2, we try to estimate robust albedo and estimate precise lighting coefficient. we can see the result at the figure 4.6. First row are the original albedo map. The original albedo map is decided from that initial illumination model divided with input image. The second row is robust albedo estimation results. The robust albedo has smoother result at some region, include of eyes, mouse, and self-shadow region. As shown in figure 4.6, unfortunately, we still cannot recover the region which has more different with realistic albedo. The worse result is generated on the middle image of figure 4.6. In this case, the result loses the information about eyes and eyebrow and still has visible shadow.

Fortunately, when we compare relighting results with robust albedo relighting results (figure 4.7), the latter has better results for recovering natural skin color. But it also needs to adjust parameter for different case, and the robust albedo relighting algorithm also spent more executive time than the general relighting algorithm.



Figure 4.6: Delighting results are calculated by that source face divide illumination model.

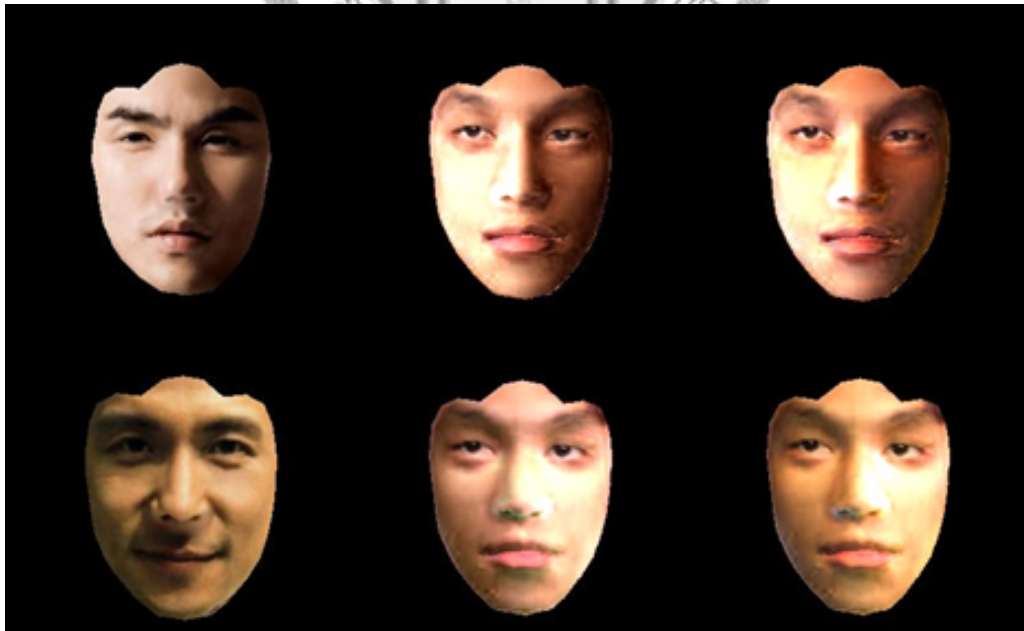
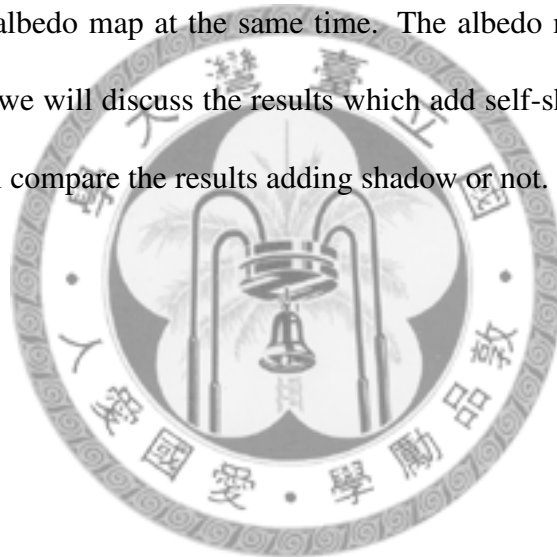


Figure 4.7: First column is target image. Second column is relighting result without robust albedo. Third column is relighting result with robust albedo.



We experiment different parameters for the same character and lighting condition. We set  $\lambda_s$  is bigger than  $\lambda_d$ , so the result will not tend to the initial albedo map which is the average color over the whole face. As figure 4.8(b) and figure 4.8(c), 4.8(c) increase  $\sigma_s$ , and it will get the result more similar as input data; it can clearly show the detail of face. As figure 4.8(b) and figure 4.8(d), when we increase  $\sigma_d$  on 4.8(d), albedo map is segmented to different color block on the face. It looks like a cartoon image, and its relighting result has the wrong skin color. According our experiment, we get a best relighting result by set control parameter  $\sigma_s$  and  $\sigma_d$  to 5 and 10 (figure 4.8e), but we can't get the natural albedo map at the same time. The albedo map is still affected by self-shadow. Finally, we will discuss the results which add self-shadow in face image in chapter 6, and we will compare the results adding shadow or not.



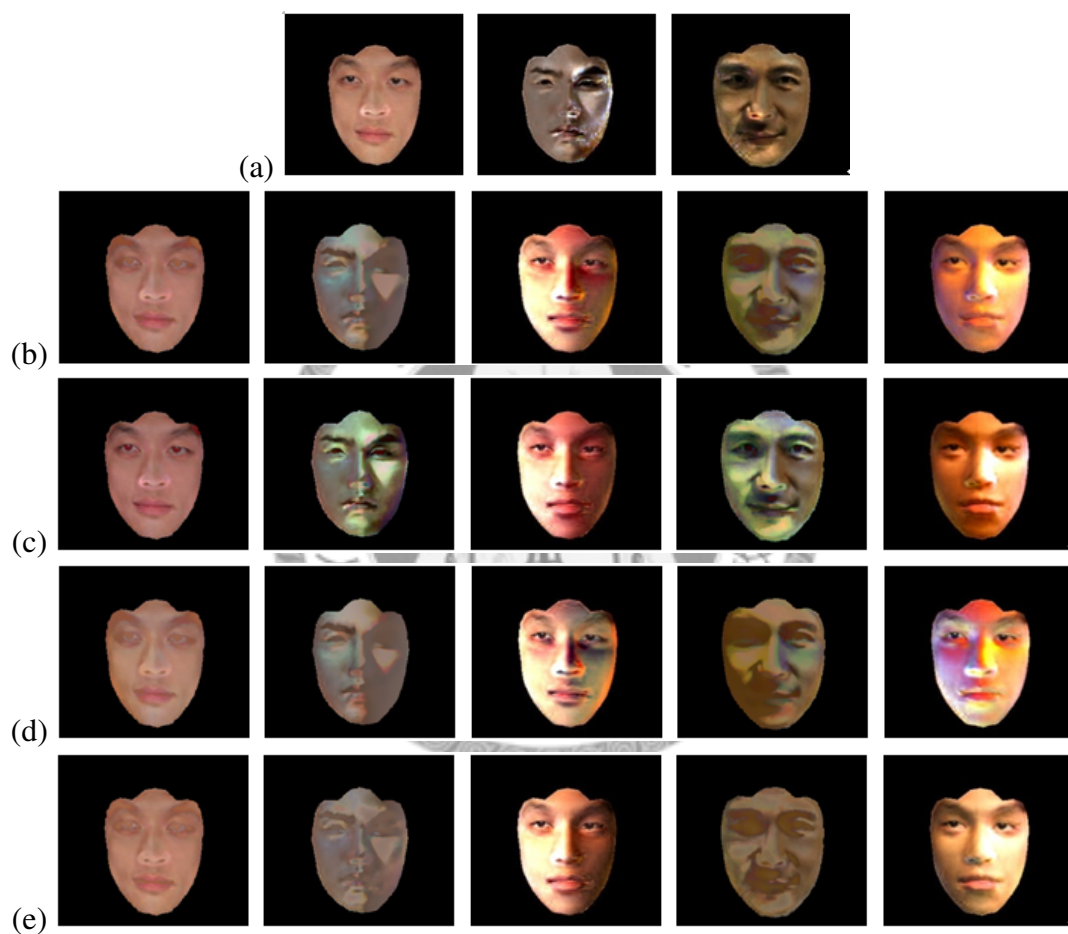


Figure 4.8: (a) Delighting results without robust albedo algorithm. (b) (c) (d) (e) Result with different control parameter.

# Chapter 5

## Face Composition

We shaded the relighting source face in the previous section. Finally, we combine the region from the relighting source face to the target face, and we want to get the realistic result. Although we use the relighting algorithm to make the source model skin color similar to the target image, it still has seam to directly paste the source face to the target face. Therefore, we need to use a seamless composition method to avoid visible artifacts in the final output.

Different people also have different facial form. If we replace the region of the full face, it may modify the facial form of the character in the target image and blend faces with non-face objects. In order that ensuring facial form is consistent with the type of target character's body and combining with the suitable area. We just replace the region  $R_{obj}$  which contains facial features and keeps away from the silhouette. How can we get the region  $R_{obj}$ ? By using the face alignment algorithm, we can get 87 features on the face, 19 of the features are at the silhouette, 20 around the mouth, 16 around the eyes, 12 at the nose, and 20 around the eyebrows. We choose those features which contain facial features that can express emotion. Those features include corners of eyes, corners of the

mouth and side of the eyebrows, and we can get a polygon to surround those features (figure 5.1). This polygon is the replaced region  $R_{obj}$ , and we use a seamless composition algorithm to combine the relighting source image to the target face at the replaced region  $R_{obj}$ .



Figure 5.1: Replacement areas are surrounding facial features.

## 5.1 Poisson Image Editing

A well-known seamless composition algorithm must be the Poisson image editing. It attempts to maintain the intensity differences between the pixel and its neighbor, and spread hue of the target face from the boundary of replacement area. Poisson image editing is using a guidance vector field to solve seamless image composition problem. We set the source guidance vector field  $v = \nabla F$  over replacement region  $R_{obj}$ , and the equation can be written as:

$$\min_F \int_{p \in R_{obj}} |\nabla F - v|^2 dp \quad \text{with} \quad F|_{\partial R_{obj}} = F_{tag}|_{\partial R_{obj}}$$

The solution is obtained by solving Laplace equation:

$$\Delta F = \Delta F_{src} \quad \text{Over} \quad R_{obj} \quad \text{with} \quad F|_{\partial R_{obj}} = F_{tag}|_{\partial R_{obj}}$$

,where  $\Delta$  is the Laplace operator, and  $F$  is the result over  $R_{obj}$ .  $F_{src}$  and  $F_{tag}$  are the source image and the target image, and  $\partial R_{obj}$  is the boundary of the replacement region. We use a least square estimation to solve this equation.

## 5.2 Optimizing Blending Boundary

When we replace the relighting source image to the target image, we only choose region  $R_{obj}$  to be replaced. Although Poisson image editing will get the seamless composition, it may not always produce good results for our case. If the boundary of  $R_{obj}$  intersects with facial saliency, ex: wrinkle, muscle change in different facial expression, scar, self-shadow, it may generate unnatural blurring artifacts. As shown in figure 5.2, the driftwood in the first image is combined to the following image; we can find unnatural blurring artifacts between driftwood and stone on the beach. To eliminate this unnatural blurring error on the face, we need to compute a new curve which still contain facial features, but not intersect with salient structures. This curve is optimal boundary, and we will get better results by this curve.



Figure 5.2: Image composition get unnatural blurring artifacts.

The optimal boundary tries to avoid intersecting facial saliency on both the source and the target face and contains facial features. We use the concept from Drag-and-Drop Pasting to estimate optimal boundary. From Poisson equation, we know that the boundary

condition  $(F_{tag} = F_{src}) | \partial R_{boundary}$  determines the result. When the color of the source and the target face in boundary is very different, it can get unnatural blurring and lost facial details which locate around the boundary. But we always want to preserve those facial saliencies and get more natural results. If we can find fewer color variation cut in the face, we may get more color-smooth result from Poisson image editing and preserve those details.

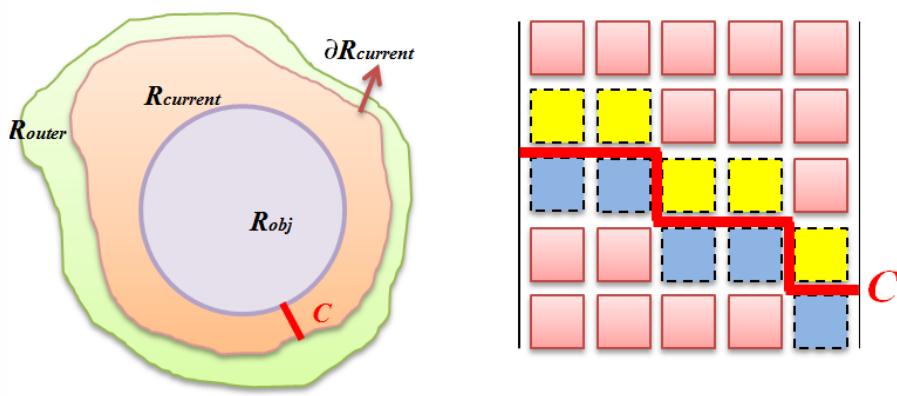


Figure 5.3: Left image is relation between original replaced region and region of full face. Right image is enlarge region near the cut.

We attempt to find the optimizing boundary between the original replaced region  $R_{obj}$  and the region of full face  $R_{outer}$ . We don't find the region smaller than  $R_{obj}$ , because we need to keep facial features in the final replaced region. The optimizing boundary must be color-smooth to avoid unnatural artifacts, so we compute the equation  $(F_{tag} - F_{src}) | \partial R_{optimal}$  to get the less color variance cut. If the color variance of the pixel is smaller, the pixel is more we want. We can describe above concept by the following energy function:

$$E(\partial R, k) = \sum_{p \in \partial R} (F_{tag}(p) - F_{src}(p) - k)^2, \quad s.t. R_{obj} \subset R \subset R_{outer} \quad (5.1)$$

Where  $F$  is the image intensity and  $R$  is the region which we want.  $k$  is average color deviation of the current boundary pixels, and  $k$  can keep that the new boundary and the

previous boundary do not have too much difference. It can be computed by:

$$k = \frac{1}{\partial R} \sum_{p \in \partial R} (F_{tag}(p) - F_{src}(p))$$

It is hard to minimize the equation 5.1, so we transform this minimizing function problem to the shortest path problem. Therefore, we map the variables to the graph. The nodes of graph are the color variance of pixels, and the weight of edges are calculated the average from connected two nodes. As shown in figure 5.3, all yellow pixels on one side of cut C (shown in red line) is computed the shortest path to the blue pixel on the other side of cut C. And we choose the cut C which is shortest straight line segment among all pixels connecting  $R$  and  $R_{obj}$  by computing the Euclidean distance. We can propose an iterative optimization algorithm to get the optimal boundary  $\partial R_{optimal}$ .

1. initialize  $R = R_{outer}$
2. Refine  $k$  by above equation
3. Setting graph's node and edge, the weight of edge is

$$\frac{(F_{tag}(p_1) - F_{src}(p_1) - k) + (F_{tag}(p_2) - F_{src}(p_2) - k)}{2}$$

- .  $p_2$  is the neighbor of  $p_1$ . Given the current  $k$ , and we use the shortest path algorithm to get the optimizing boundary  $\partial R$  between the regions  $R_{current}$  and  $R_{outer}$ .
4. Repeat step 2 and 3 until that the energy is convergent. Then the boundary  $\partial R$  is the optimizing boundary  $\partial R_{optimal}$ .

In the step 4, we finish the iteration by the energy convergence. We define the convergence in two rules. First, difference between the energy of two successive iterations

need to be smaller a threshold, because it confirm that the boundary of two successive iterations do not have too much change. Second, we avoid the final energy being the same as initiative energy, because we don't want the replaced region to be the full face.

### 5.3 Target Face Geometry Recovery

Structure of the face includes contour, facial expression, wrinkle, and other textures. Our system will first maintain silhouette of target face, then we focus on the issue of recovering more complicated facial geometry. Complicated facial geometry like wrinkle, scruff, pock or scar, is hard to clone from target face to 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 features as textures. All we need to do is selectively retain those textures on replaced result. Those textures often have obvious edge and color variance. So we can detect and preserve those regions with larger gradients. Since the replaced result not only preserves the features of target face but also has the features in the source face, we compare source face gradient with target face gradient. Then we define the stronger gradient variance 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 the target face. We use face alignment to label the regions of facial salient which includes eyes, nose, mouth, and avoid mix gradients on those region. Those regions are only replaced by the source facial



features. If we do not enforce this constraint, those regions may generate ghost image or lost the characteristic of source face.

## 5.4 Result

As shown in figure 5.4, using Poisson image algorithm to combine Subject A's face to other actor. The replaced region contains the features which are detected from the face alignment algorithm. We found that the most of face replacement results cover less region; especially on the corners of eyes and the corners of the mouth. The corners of mouse in the target subject faces are still kept on the results after face replacement, so we can see the unnatural hole near the mouth of replaced face. And some actor's eyes are segmented to the boundary of replaced region. If the boundary intersects shadow, it would generate unnatural blurring artifacts.

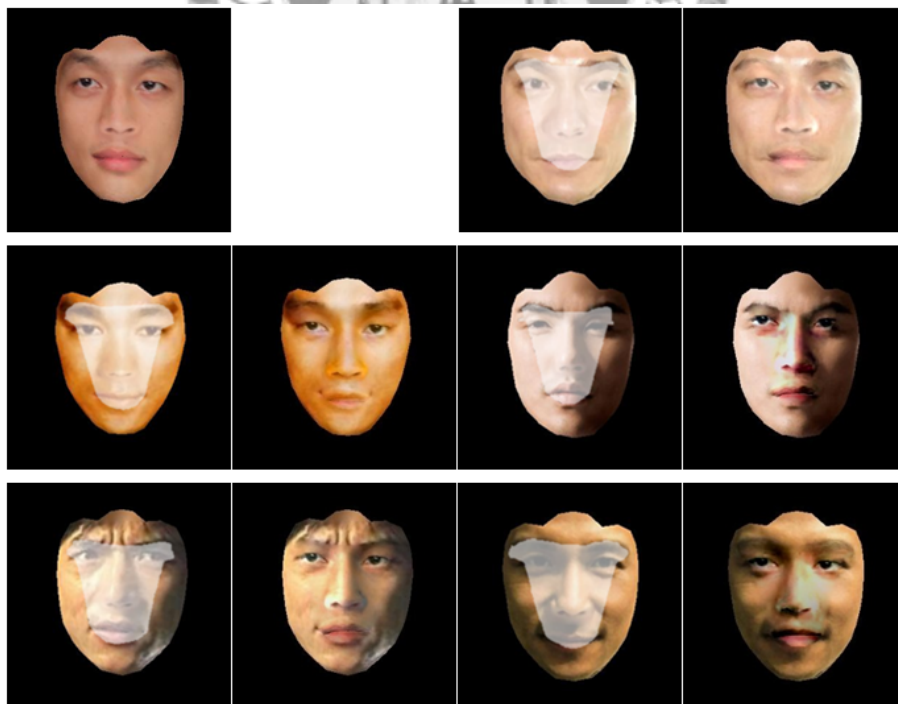


Figure 5.4: Face composition results and corresponding replacement areas.

We refer to Drag-and-Drop composition to find optimal boundary before we imple-

ment Poisson image editing. The replaced region within the optimal boundary almost cover the corners of eyes and corners of the mouth, even though the smoother boundary was found in the self-shadow region(figure 5.5). We can see the result of Ethan Ruan or Jackie Chan. One sides of their face have large shadow area, and their boundaries do not intersect the area which may cause unnatural blurring artifacts. Instead of intersecting the area which has high color variance between the source face and the target face, it intersects color-smooth area. Of course, these results are better than the results which only blend by Poisson image editing. Finally, We will discuss the mixing gradients results in chapter 6 later, and we will compare the results with mixing gradients or not.



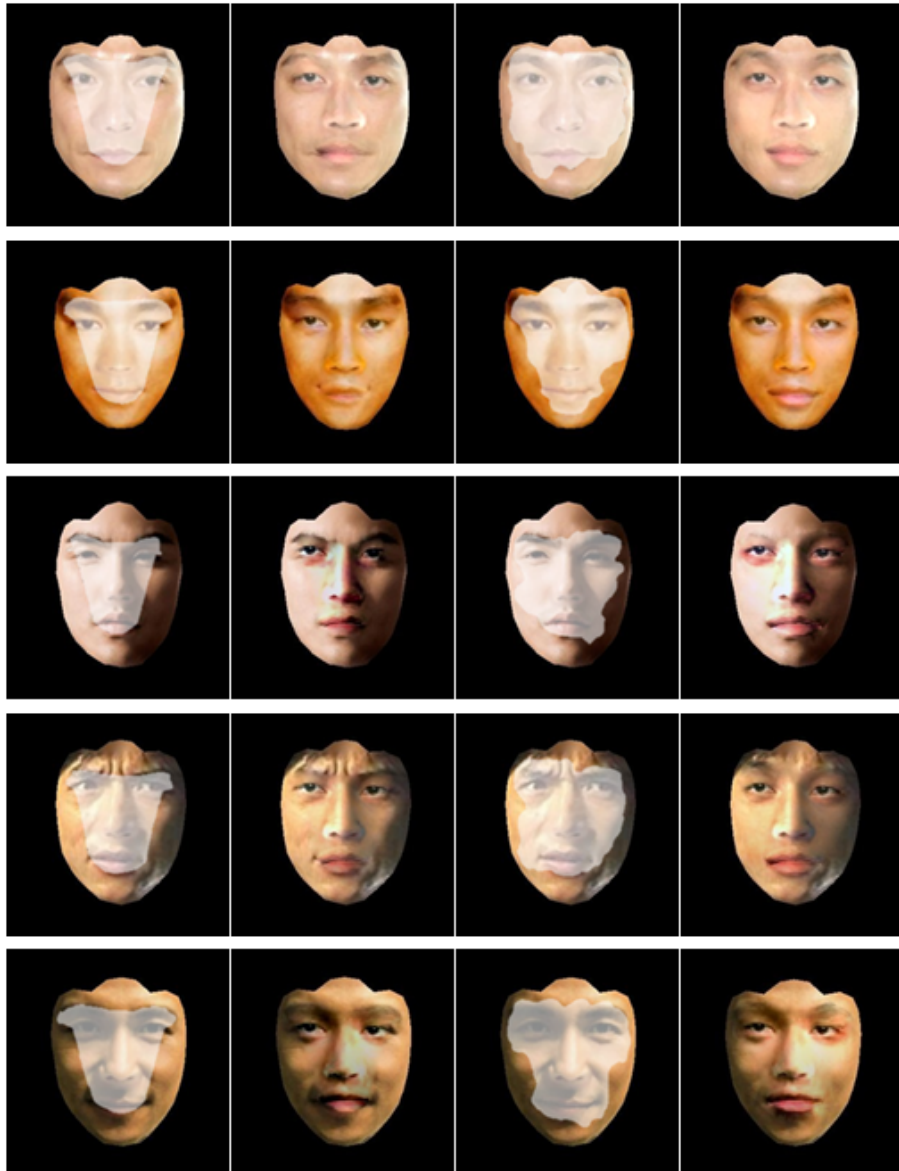


Figure 5.5: Comparing results with optimal boundary or not. First and second column are original results. Third and fourth column are results with optimal boundary.



# Chapter 6

## Result and Discussion

We experiment our system for several characters. To demonstrate that the 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 the original image and Subject A's blending result.

In Figure 6.1 (b) and (c), 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 the light and shadow, the generated face replacement result looks realistic in this case. We replace subject A's face into Bruce Lee in Figure 6.1 (d) and (f). Figure 6.1 (d) has similar lighting conditions as Andy Lau, with slight changes in the light and shadow, but the spherical harmonic relighting image has unusual skin color in the saturated region. 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 the realistic result.

In Figure 6.2, we replace the movie star Ethan Ruan. This target image has wide



Figure 6.1: Results of normal lighting condition. (b) The target image of Andy Lau. (c) Face replaced result. (d) The target image of Bruce Lee. (e) Unusual skin color on relighting source image. (f) Face replaced result.

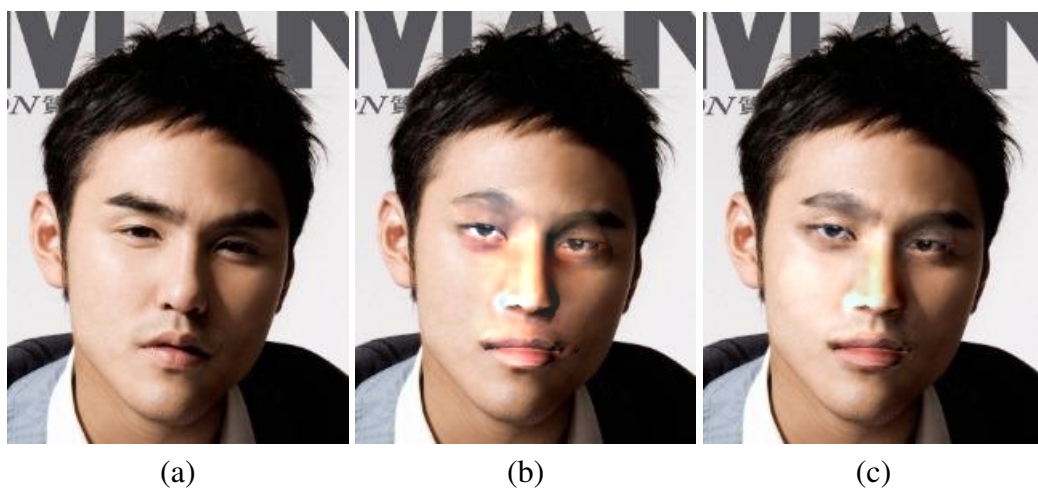


Figure 6.2: Hard self-shadow simulation. (a) The target image of Ethan Ruan (b) The blending face without adding shadow (c) The blending face with adding shadow

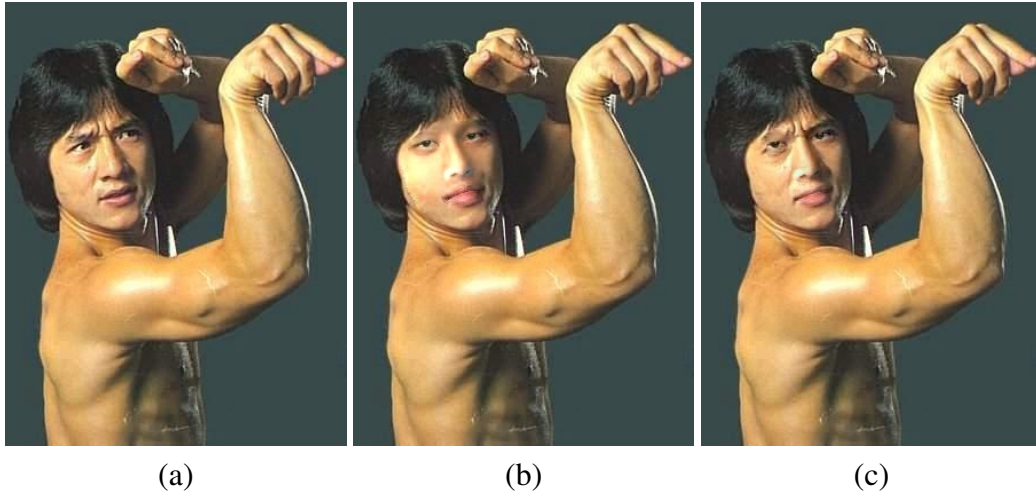


Figure 6.3: Case with strong ambient occlusion and complicate facial geometry (a) The target image of Jackie Chan. (b) Blend faces without mixing gradient. (c) Blend faces with mixing gradient.

self-shadow region. We compare the replaced result which using spherical harmonics relighting only( 6.2 (b)). As mentioned above, we not only simulate self-shadow well but also maintain ambient occlusion effect on the target face.

In Figure 6.3, 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 6.3 (b) 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. Our users feel that figure 6.3 (c) can show more emotion similar to the target face.




Figure 6.4: More result. (a) The target image of Takeshi Kaneshiro (b) The final result of Takeshi Kaneshiro (c) The target image of Aaron Kwok (d) The final result of Aaron Kwok (e) The target image of Jet Li (f) The final result of Jet Li (g) The target image of Jacky Cheung (h) The final result of Jacky Cheung (i) The target image of Chow Yun-fat (j) The final result of Chow Yun-fat (k) The target image of Tom Cruise (l) The final result of Tom Cruise (m) The target image of Andy Lau (n) The final result of Andy Lau (o) The target image of Mark Chao (p) The final result of Mark Chao (q) The target image of Rain (r) The final result of Rain (s) The target image of Ekin Cheng (t) The final result of Ekin Cheng



# Chapter 7

## Conclusion and Future Work

### 7.1 Conclusion

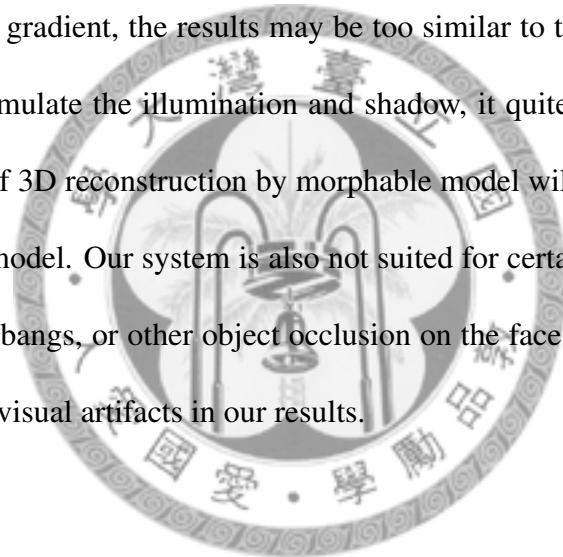
The logo of National Taiwan University is a circular emblem. It features a central design with a stylized tree or plant motif, flanked by two vertical elements that resemble traditional Chinese architectural pillars or a pair of scales. The emblem is surrounded by a decorative border containing the university's name in Chinese characters: "國立台灣大學" (National Taiwan University) at the top and "1946" at the bottom.

In this paper, we present a system to semi-automatically replace faces from the source image to the target image. Our system can handle harsh lighting condition and facial details. To import the model into our system, we use the face alignment algorithm to extract the facial features from the target face. According to the position of those features, we can semi-automatically estimate the head pose for aligning source head to target subject face. 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. We combine the face alignment and Drag-and-Drop pasting to find the optimal pasting boundary on the target face, thus produce a seamless blending result which maintain the facial saliency and ambient occlusion parts. We also apply mixing gradient method to the blending process, and preserve both facial details on the source face and the target face. Our system provides users to replace face in the image, and the source face can be specified by the users. We also reduce the interference

of user, so people can effectively replace arbitrary face by our system.

Our face replacement system still has limitations. The system relies on face alignment results. The feature points and facial saliency region detection from the face alignment, and it will affect the quality of pose estimation and composition boundary.

We synthesize the source face model by morphable model method, and its accuracy strongly depends on the 3D model data set. The morphable model usually cannot reconstruct complicate facial expression details, so the source face model may lose important information at the model synthesis stage. Then after blending the source face and the target face by mixing gradient, the results may be too similar to target face. 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 also 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.



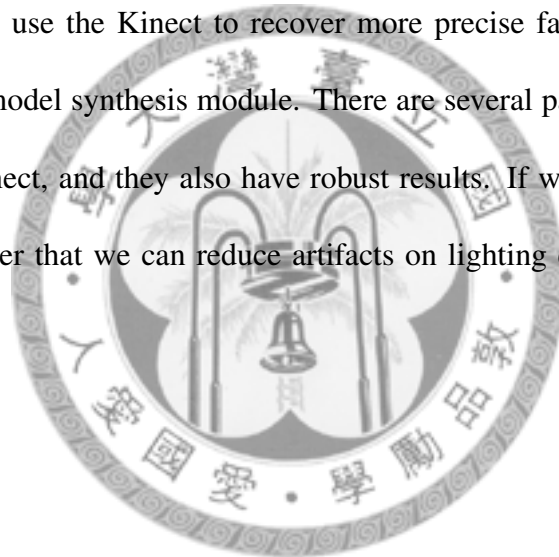
## 7.2 Future Work

There are three avenues for our future work. First, we want to replace faces in video clips, so we need to consider temporal coherence and be allowed less user interference. Both of temporal coherence and less user interference can avoid time-consuming and labor-intensive. We could not manually replace face frame by frame, because it will not only have incoherent result for both head pose and illumination but also spent too much time. If we replace faces in the video, we need a global solution which considers the relation between current frame and previous frame. The relation involves results of face alignment, pose parameters, lighting environment, and replaced region. All of them must

compute temporal coherence solution to avoid that the source face is inconsistent with target character and unstable in the target video.

Second, we estimate the position of principle light to construct shadow map, and the shadow map is only simulate the hard shadow. In our assumptions, we consider that only one principle light generates the hard shadow on the target subject face. If the scene has interaction of several lights to cause soft shadow, we would be hard to synthesis the same lighting condition. We are better to estimate all the light position in practice, and it can synthesis the correct lighting environment.

Third, we plan to use the Kinect to recover more precise face geometry to reduce the error in the face model synthesis module. There are several papers about face model reconstruction by Kinect, and they also have robust results. If we construct face model by Kinect, we consider that we can reduce artifacts on lighting estimation and shadow generation.





# Bibliography

- [1] R. Basri and D.W. Jacobs. Lambertian reflectance and linear subspaces. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(2):218 – 233, feb 2003.
- [2] Soma Biswas, Gaurav Aggarwal, and Rama 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] Volker Blanz, Curzio Basso, Thomas Vetter, and Tomaso Poggio. Reanimating faces in images and video. In Pere Brunet and Dieter 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] Volker Blanz, Kristina Scherbaum, Thomas Vetter, and Hans-Peter Seidel. Exchanging faces in images. In Marie-Paule Cani and Mel 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] Volker Blanz and Thomas 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. In *Proceedings of the 3rd. International Conference on Face & Gesture Recognition*, FG '98, pages 88–, Washington, DC, USA, 1998. IEEE Computer Society.

- [8] Yi-Ting Cheng, Virginia Tzeng, Yu Liang, Chuan-Chang Wang, Bing-Yu Chen, Yung-Yu Chuang, and Ming 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] Yung-Yu Chuang, Brian Curless, David H. Salesin, and Richard Szeliski. A bayesian approach to digital matting. In *Proceedings of IEEE CVPR 2001*, volume 2, pages 264–271. IEEE Computer Society, December 2001.
- [10] Paul Fitzpatrick. Head pose estimation without manual initialization. 2007.
- [11] Jiaya Jia, Jian Sun, Chi-Keung Tang, and Heung-Yeung Shum. Drag-and-drop pasting. *ACM Transactions on Graphics (SIGGRAPH)*, 2006.
- [12] Ritwik Kumar, Michael Jones, and Tim K. Marks. Morphable reflectance fields for enhancing face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.
- [13] Vivek Kwatra, Arno Schödl, Irfan Essa, Greg Turk, and Aaron Bobick. Graphcut textures: image and video synthesis using graph cuts. *ACM Trans. Graph.*, 22:277–286, July 2003.
- [14] T.M. Mahmoud. A new fast skin color detection technique. In *Processing of World Academy of Science, Engineering And Technology Volume 33 ISSN p.2070-3740*, volume 2, pages 525 –528, march 2010.
- [15] Patrick Pérez, Michel Gangnet, and Andrew Blake. Poisson image editing. In *ACM SIGGRAPH 2003 Papers*, SIGGRAPH '03, pages 313–318, New York, NY, USA, 2003. ACM.
- [16] T. Vatahska, M. Bennewitz, and S. Behnke. Feature-based head pose estimation from images. In *Proc. of the IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, 2007.
- [17] Yang Wang, Zicheng Liu, Gang Hua, Zhen Wen, Zhengyou Zhang, and Dimitris Samaras. Face re-lighting from a single image under harsh lighting conditions. In *CVPR'07*, pages –1–1, 2007.
- [18] Zhen Wen, Zicheng Liu, and T.S. 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.

- [19] Miroslav Hamouz Xuan Zou, Josef Kittler and Jose R. Tena. Robust albedo estimation from face image under unknown illumination. In *Proc. SPIE 6944, 69440A*, page doi:10.1117/12.778599, 2008.
- [20] Yi Zhou, Lie Gu, and Hong-Jiang 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.



## Resume

Name:

Shih Hao Hsiung(熊士豪)

E-mail:

supermfb@gmail.com

Education:

[M.S]Department of Computer Science and Information Engineering, National Taiwan University (2009-2011)

[B.S] Department of Computer Science and Information Engineering, National Cheng Kung University (2004-2008) University

[H.S] Kuang Jen Catholic High School. (2000-2003)

Master Thesis:

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

Advisor:

Prof. Ming Ouhyoung

Skill:

Digital Image Processing (DIP), C/C++, Matlab, OpenGL, GNU Scientific Library (GSL), OpenCV, Qt.

Project:

Face replacement project.

Publication:

Hong Shang Lin, Shih Hao Hsiung, Shyh-Kang Jeng, Ming Ouhyoung. Automatic 3D face Reconstruction From Multiple Views Through SIFT Feature. Computer Graphic Workshop (CGW), 2010

Yi Shan Cheng, Wan Chun Ma, Hong Shang Lin, Shih Hao Hsiung, Ming Ouhyoung. 動畫狂想曲—人臉電腦繪圖技術解密. 國科會科學月刊

