行政院國家科學委員會專題研究計畫　成果報告

擷取與合成細部表情於三維擬真人臉動畫之研究

計畫類別：個別型計畫
計畫編號：NSC94-2213-E-009-151-
執行期間：94 年 08 月 01 日至 95 年 07 月 31 日
執行單位：國立交通大學資訊科學學系(所)

計畫主持人：林奕成

報告類型：精簡報告

處理方式：本計畫可公開查詢

中　華　民　國 95 年 10 月 30 日

行政院國家科學委員會補助專題研究計畫成果報告

擷取與合成細部表情於三維擬真人臉動畫之研究

Extracting and Synthesizing Detailed Expressions for Realistic 3D Facial Animation

計畫類別：☑ 個別型計畫　　□ 整合型計畫

計畫編號：NSC 94-2213-E-009 -151

執行期間： 94 年 8 月 1 日 至 95 年 7 月 31 日

計畫主持人：林奕成　　國立交通大學資訊工程系助理教授

成果報告類型(依經費核定清單規定繳交)：☑精簡報告　□完整報告

本成果報告包括以下應繳交之附件：

□赴國外出差或研習心得報告一份

□赴大陸地區出差或研習心得報告一份

□出席國際學術會議心得報告及發表之論文各一份

□國際合作研究計畫國外研究報告書一份

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢

□涉及專利或其他智慧財產權，□一年□二年後可公開查詢

執行單位：：國立交通大學資訊工程學系(所)

中　華　民　國　九 十 五 年 十 月　三 十 日

擷取與合成細部表情於三維擬真人臉動畫之研究

Extracting and Synthesizing Detailed Expressions for Realistic 3D Facial Animation

計畫編號：NSC 94-2213-E-009 -151

執行期間： 94 年 8 月 1 日 至 95 年 7 月 31 日

計畫主持人：林奕成　　國立交通大學資訊工程系助理教授

中文摘要

三維人臉動畫一直是電腦圖學界中頗受注目且重要的研究課題。如今，藉由動態捕捉技術的輔助，人臉上的特徵點動作可由追蹤貼在其上的標記點來估算。將之應用於動畫上，產生人臉動畫的時程可因此大幅縮短。現有的動態捕捉技術雖能掌握特徵點的運動軌跡，但無法擷取人臉表面細微的變化，如皺紋、酒窩等。日常生活中，我們常藉由這些細微的部分來判斷他人的情緒，如能逼真地合成出這些細微表情變化，對於人臉動畫擬真度可大為提升。

在此計畫中，我們針對此關鍵的問題，提出結合擷取人臉細部表情的方法以應用於三維人臉表情編輯。首先，由受測者臉部的標記點，以立體電腦視覺配合通用人臉模型（generic face model）形變的方式，估計出受測者的初步幾何臉部模型。在已校正光源的環境中，採用 Lambertian 等的光照模型，人臉表面的細部法向量變化可以由影像的色澤變化來計算出。

藉由收集多個不同表情之細部臉部表面，我們將這些樣本形成一由特徵點群位置來控制之臉部表情空間。當使用者調整臉部模型的特徵點，我們的系統將利用最佳化的方式，推估最適當的臉部細部表情變化。現有的動態捕捉系統亦可利用本研究的技術，使其增加展現人臉細部表情的能力。

關鍵字：三維人臉動畫、動態捕捉技術、細部人臉表情合成、表面重建。

Abstract

3D facial animation has been an attractive and important topic in computer graphics for decades. Nowadays, with the help of facial motion capture techniques, recording motion of conspicuous makers stuck on subjects' faces, the progress of generating facial animation is dramatically speeded up. However, existing facial motion capture techniques can only track motion on these feature points, but they are incapable of capturing variations on other facial details, such as wrinkles or creases. For realistic facial animation, faithful synthesis of these detailed expressions is quite critical, since we people usually take these as cues to recognize others' emotions or feelings.

To tackle this key problem, in this project, we propose developing an approach that is able to extract and synthesize facial detailed expressions for 3D facial animation. We propose estimating facial detailed undulation by combining a stereo geometry reconstruction and a shape-from-shading approach. By analyzing the markers in two views, we can reconstruct the markers' 3D structure. After deforming a generic model according to the markers' 3D positions, an approximate 3D facial surface can be estimated. Given a pre-calibrated condition with the Lambertian lighting model, we evaluate the detailed normal variations on a facial surface point from intensity variations in video clips.

By collecting facial detailed samples, an expression space controlled by feature points can be constructed.

After analyzing various wrinkle surfaces, we can propose a synthetic facial model with detailed expressions. While users adjust feature points of a face model, our system will evaluate the most appropriate facial details by an optimization method. With the proposed technique, the existing facial motion capture system can also be enhanced.

Keywords: 3D facial animation, motion capture, detailed facial expression, surface reconstruction.

1. Introduction

Facial animation has been an attractive research topic for a long time. From movies, games, to virtual communication interfaces, nowadays, these research works come to our daily lives. Example-based and physical-based approaches are two major approaches to drive facial models. Example-based approaches usually animate faces according to motion capture (Mocap) data or synthesize novel expressions from images or video samples.

Motion capture techniques estimate motion trajectories of feature points on real persons. Subjects have to be pasted a few markers on the faces. 3D positions of markers can be reconstructed by stereo vision or electromagnetic positioning techniques. The estimated motion trajectories can then be used to driven face meshes. This approach, however, can't acquire facial details like wrinkles. In order to generate facial details, a lot of labor-intensive post-processing works are required. Artists have to adjust meshes or textures to mimic wrinkles according to different expressions. Another appearance-based approach is image-based synthesis which animates faces from sets of image or video samples. Facial details are usually well preserved in this approach, but it is difficult to relight the model and view directions are limited.

The goal of the proposed work is to build an expression edition system that can synthesize not only geometry approximation but surface details. To avoid complex simulation, we adopt to capture face details from real persons. We also observed that variations of surface details are highly related to adjacent feature points. Therefore face details can be synthesized according to positions of feature points.

2. Related work

Many researches about facial animation have been proposed these decades. Video Rewrite [C. Bregler 97] synthesized new movie sequences from existing footages. The results are visually convincing but relighting and changing view directions are difficult. In 2001, Z. Liu and Y. Shan proposed expression ratio images [Z.Liu 01]. They took the advantage of the property that expression ratios of subjects are the same under the Lambertian lighting model. Therefore, the expression in a source image can easily be retargetted to a new face by scaling. Like other image-space approach, Lighting and view directions of their work are limited.

V. Blanz et al. [99 and 03] used a large amount of scanned faces to build a morphable model. They assumed that human faces can be synthesized by convex combination of prototypes. By minimizing intensity differences of an input image and the projection of the morphable model, a target 3D model can be reconstructed. In the synthetic step, users can also adjust corresponding weight of each scan for exaggerated results.

Zhang et al. [Zhang 06] proposed a geometry-driven approach to synthesize expressions of a particular subject. They assumed the expressive textures on a face are related to geometric variation. Their system can generate corresponding textures when users adjust geometric features. This concept is similar to our proposed work. But we focused on extraction, synthesis, and editing of 3D faces.

To reliably estimate facial details, we use shape-from-shading techniques. Shape-from-shading (SFS) is a kind of surface reconstruction technology that can reconstruct depth from a single image. Readers can refer to the literature [R. Zhang 1999] for details.

Among various types of approaches, Horn [Horn 1990] proposed that the surface normals can be recovered from the intensity variations of an image. He took an optimization method that iteratively minimized errors. H. Fang et al. [H. Fang 2004] proposed a tool to apply textures in a photograph. First, the light direction in the image was approximated iteratively. Then, they utilized a SFS approach to evaluate the surface undulation for further texture mapping.

3. The proposed method

The flow chart of our system is depicted in figure 1. First, we acquire sample images of a particular subject from three different views (front, left-side and right-side). The left and right views are used to reconstruct 3D positions of feature points. The frontal views are used to estimate normals of face details by shape-from-shading (also known as photoclino-metry).

Before synthesizing any expressions, we have to fit a generic model into our subject and retargeting the feature points which are captured from stereoposis. In the synthesis step, we analyze the correlation between feature points and facial details. Therefore, we can synthesize novel expressions according to various feature configurations. We form the synthesis of novel expressions as an optimization problem. We evaluate the faithfulness by an objective function with inequality constraints. By iteratively minimizing this objective function, we can find the best approximation of the target expression. Finally, we represent each novel expression as a special texture (known as Normal difference map which will be explained later.) and use programmable shaders for real-time rendering.

3.1 Acquisition of detailed 3D expression

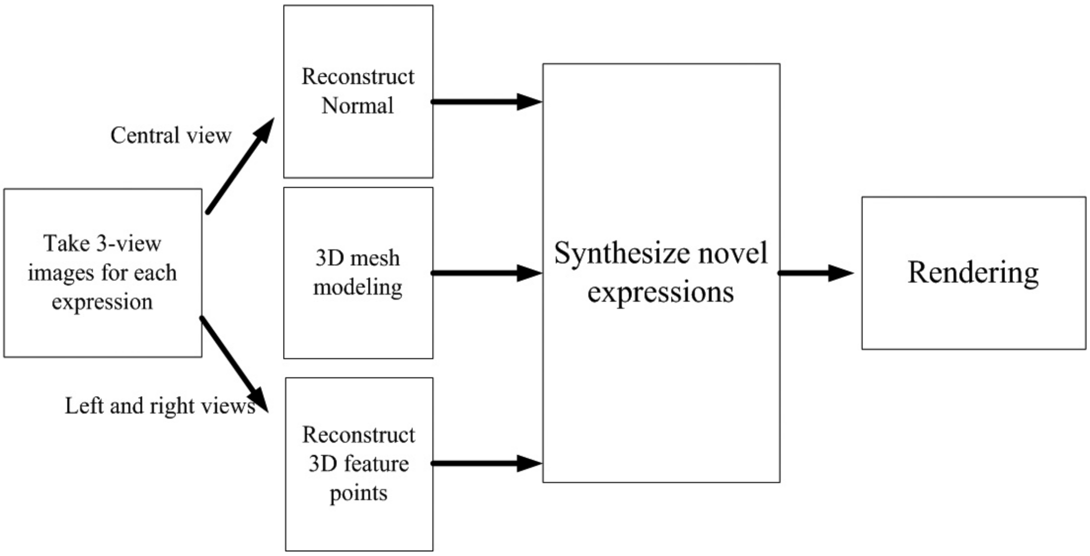

Figure 1: The system overview

We assumed that the color of the skin is almost uniform. Based on this assumption, the variations of skin color in the image is due to variations of the incident light directions.

As shown in Fig. 2, let S be the unit vector of the light source direction. To evaluate the surface normal of a pixel in the image, first, we have to estimate a projection vector G_{xy} as shown in Eq.1.

$$G_{xy} = \nabla I_{xy} - (\nabla I_{xy} \cdot S)S \tag{1}$$

where $\nabla I_{xy} = (\partial I_{xy} / \partial x, \partial I_{xy} / \partial y, 0)$ is the image gradient and G_{xy} means the projection of vector ∇I_{xy} to the plane perpendicular to S. The cosine of the angle between the surface normal and the incident light direction can be evaluated as follows:

$$C(x, y) = (I_{xy} - I_{\min}) / (I_{\max} - I_{\min})$$

where $I_{\min}$, the darkest intensity value, implies ambient light in the scene and the brightest value, $I_{\max}$, implies the intensity when a pixel faces the light source.

The normal N_{xy} can be estimated as Eq. 2.

$$N_{xy} = C(x, y)S + S(x, y)G_{xy} / \|G_{xy}\| \tag{2}$$

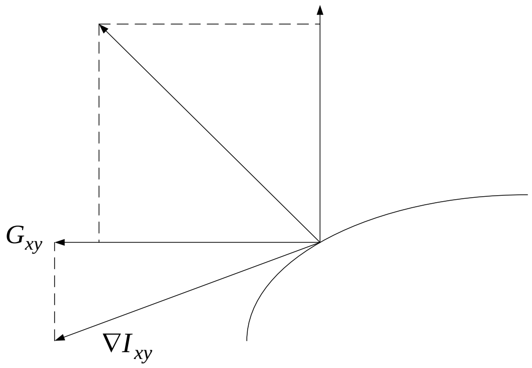

Figure 1: We project the gradient of image to the plane perpendicular to the incident light vector. The normal of the pixel can be calculated through cosine and sine function estimated from the image.

After reconstructing the normal of the subject, we would like to get the normal difference map which represents the normal variations due to a facial expression. The image I_n is indicated the normal map of the neutral face and the image I_e indicated the normal map of an expression e.g. raising eyebrow, laughing, etc.

Under our uniform skin color assumption, we will have "fault" normals caused by eyebrow, acne, color markers, etc. in normal maps (as shown in Fig.3(a)(b)). Notwithstanding, these fault normals will be eliminated by the subtracting procedure.

(a) (b) (c)

Figure 3: (a) is the normal map of the expression. (b) is the normal map of the neutral face. And (c) is the difference map by subtracting (b) from (a).

3.2 Synthesizing Novel Expressions

We assume that normal variation is highly related to the movement of control points. With such assumption, we can analyze the correlation between wrinkle variations and control points such that each new expression can be synthesized by the configuration of control points. Furthermore, since it's infeasible to record all kinds of expressions, we select a set of representative expressions to form an expression vector space. Given a configuration of control points, the corresponding details can then be evaluated.

To form such an expression space, we extend Zhang's method [Zhang 2006]. Each expression is represented as $E_i = (G_i, I_i)$ where E_i, G_i represents an expression, and geometry respectively. $I_i = (N_i, T)$ where N_i is the surface normal map and T is the face texture. Let $H(E_0, E_1, ,..., E_m)$ be the space of all possible convex combinations of these examples.

$$H(E_0, E_1, ..., E_m) = \left\{ \left(\sum_{i=0}^{m} ciGi, \sum_{i=0}^{m} ciIi \right) \middle| \sum_{i=0}^{m} ci = 1, c0, ..., cm \geq 0 \right\}$$

Our object is to synthesize new normal difference map, and a new normal difference map can be calculated by convex combination as follow:

$$N_{(new)} = \sum_i c_i N_i$$

where *N(new)* is novel normal difference map.

Let G^R denote the new positions of feature points and G_i^R denote the sample expressions in our database. The combination coefficients c_i can be found by project G^R into the convex hull of $G_0^R ... G_m^R$. Thus, we estimate c_i by an optimization approach:

$$\text{Minimize} \left(G^R - \sum_{i=0}^{m} c_i G^R \right)^T \left(G^R - \sum_{i=0}^{m} c_i G_i^R \right),$$

Subject to: $\sum_{i=0}^{m} c_i = 1, c_i \geq 0$ for i = 0,1,…,m,

The objective function of the optimization problem above can be rewrite as:

$$C^T g^T gC - 2G^{R^T} gC + G^{R^T} G^R \qquad (3)$$

where $g = (G_0^R, G_1^R, ..., G_m^R)$, $C = (c_0, c_1, ..., c_m)$. This optimization is a positive semi-definite quadratic programming problem with linear constraints. The problem can be solved by various methods such as interior-point method or active set method. In our approach, we used the active set method to solve this optimization problem.

4. Result and conclusion

We propose a 3D expression editing tool where facial details can be synthesized according to control points. Since we use the variations of normals to represent face details, the relighting problem is overcome and realism can be improved in many interactive applications. For example, this approach can make virtual faces more expressive in real-time application like game or telecommunication. For synthesizing novel expressions by manual works, editing feature points to drive wrinkles is very intuitive and user-friendly. Moreover, we can retarget the synthetic results to another subjects or even animals. There are two disadvantages in the current system. In normal recovery phase, the recovered normals are usually with unavoidable noise and it may lead odd synthetic wrinkles. We will use bilateral filters in the future. Second, currently, our system doesn't take into account temporal consistency. We will include temporal penalty terms for animation.

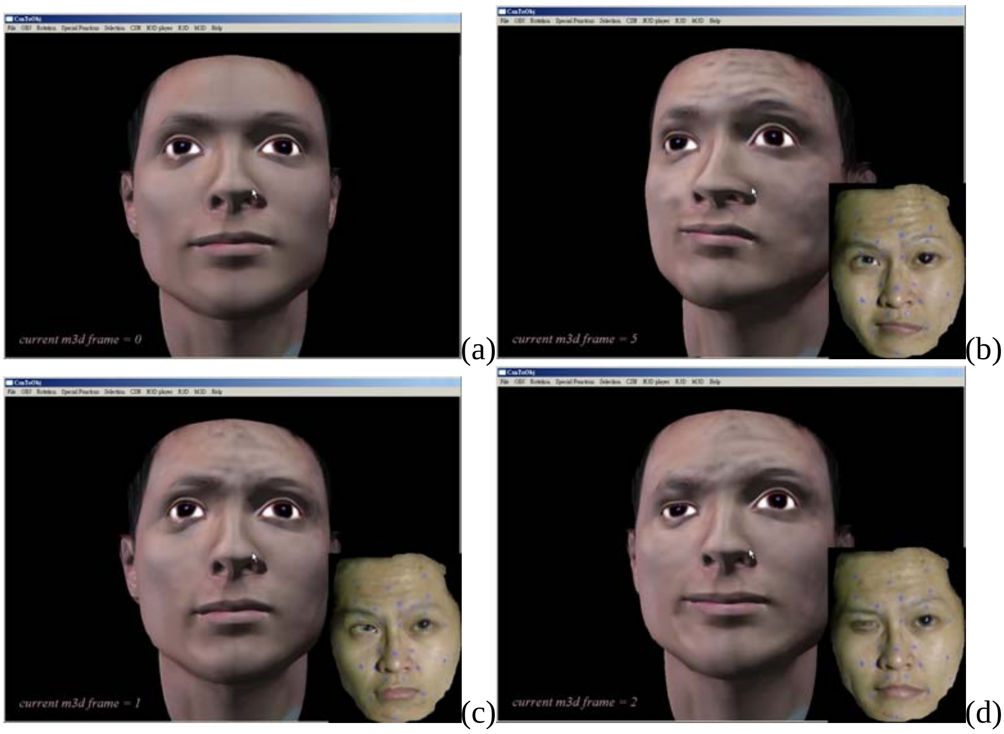

Figure 4. (a) the neutral face (b)(c)(d) synthetic expressions according to configurations of feature points

5. Publications

- ☐ 莊棨元 碩士論文 "Synthesis of 3D Detailed Facial Expression"。
- ☐ "Acquisition and synthesis of 3D detailed expressions", prepared for submission.

預期工作項目	實際工作成果	說明
由影像擷取三維人臉細部表情參數之技術	利用shape-from-shading技術,由臉部反射光澤變化估計細部表面起伏	符合。
細微表情擷取技術與臉部動態捕捉技術之結合	將所獲得之細部臉部資料形成向量空間,輸入特徵點位置,可利用最佳化方式尋找最適合的細部表情組合。	符合。
合成具細微表情之擬真三維人臉動畫。	完成一表情編輯系統,可由使用者輸入特徵點或是提供特徵位置,可應用於動畫產生。	符合。

附件二

可供推廣之研發成果資料表

□ 可申請專利　☑ 可技術移轉　日期：95 年 10 月 30 日

國科會補助計畫	計畫名稱：擷取與合成細部表情於三維擬真人臉動畫之研究 計畫主持人：林奕成　國立交通大學資訊工程系助理教授 計畫編號：NSC 94-2213-E-009 -151　學門領域：計算機
技術/創作名稱	擷取與合成細部表情於三維擬真人臉動畫
發明人/創作人	林奕成、莊榮元、林昭自 、羅永盛
技術說明	中文： 三維人臉動畫一直是電腦圖學界中頗受注目且重要的研究課題。如今，藉由動態捕捉技術的輔助，人臉上的特徵點動作可由追蹤貼在其上的標記點來估算。現有的動態捕捉技術雖能掌握特徵點的運動軌跡，但無法擷取人臉表面細微的變化，如皺紋、褶紋、酒窩等。在此計畫中，我們針對此關鍵的問題，提出結合擷取人臉細部表情的方法以應用於三維人臉表情編輯。首先，由受測者臉部的標記點，以立體電腦視覺配合通用人臉模型（generic face model）形變的方式，估計出受測者的初步幾何臉部模型。在已校正光源的環境中，採用 Lambertian 等的光照模型，人臉表面的細部法向量變化可以由影像的色澤變化來計算出。 藉由收集多個不同表情之細部臉部表面，我們將這些樣本形成一由特徵點群位置來控制之臉部表情空間。當使用者調整臉部模型的特徵點，我們的系統將利用最佳化的方式，推估最適當的臉部細部表情變化。 現有的動態捕捉系統亦可利用本研究的技術，使其增加展現人臉細部表情的能力。此研究亦可增快人臉動畫的製作速度。

	英文： 3D facial animation has been an attractive and important topic in computer graphics for decades. Nowadays, with the help of facial motion capture techniques, recording motion of conspicuous makers stuck on subjects' faces, the progress of generating facial animation is dramatically speeded up. However, existing facial motion capture techniques can only track motion on these feature points, but they are incapable of capturing variations on other facial details, such as wrinkles or creases. For realistic facial animation, faithful synthesis of these detailed expressions is quite critical, since we people usually take these as cues to recognize others' emotions or feelings. To tackle this key problem, in this project, we propose developing an approach that is able to extract and synthesize facial detailed expressions for 3D facial animation. We propose estimating facial detailed undulation by combining a stereo geometry reconstruction and a shape-from-shading approach. By analyzing the markers in two views, we can reconstruct the markers' 3D structure. After deforming a generic model according to the markers' 3D positions, an approximate 3D facial surface can be estimated. Given a pre-calibrated condition with the Lambertian lighting model, we evaluate the detailed normal variations on a facial surface point from intensity variations in video clips. By collecting facial detailed samples, an expression space controlled by feature points can be constructed. While users adjust feature points of a face model, our system will evaluate the most appropriate facial details by an optimization method. With the proposed technique, the existing facial motion capture system can also be enhanced. It will substantially improve the generation speed of 3D facial animation.
可利用之產業 及 可開發之產品	數位內容產業、電腦動畫製作、電腦遊戲
技術特點	☐ 由真實人臉部影像推估細部表面起伏（如皺紋等） ☐ 利用特徵點控制人臉細部表情合成系統
推廣及運用的價值	可應用於三維人物動畫與遊戲等製作

※ 1.每項研發成果請填寫一式二份，一份隨成果報告送繳本會，一份送 貴單位研發成果推廣單位（如技術移轉中心）。

※ **2.本項研發成果若尚未申請專利，請勿揭露可申請專利之主要內容。**

※ 3.本表若不敷使用，請自行影印使用。