

What's fun in EE

臺大電機系科普系列



動畫狂想曲：電影「阿凡達」的製作及 人臉電腦繪圖技術突破

鄭亦珊、馬萬鈞、林宏祥、熊士豪、歐陽明
臺大資訊工程系及資訊網路與多媒體研究所

【本文作者歐陽明教授為臺大資工系教授／臺大電機系 1981 年畢業系友】

摘要

在本篇文章中，我們將介紹實驗室裡有關虛擬實境的技術，透過電腦所製作出來的 3D 模型，加上動態捕捉等技術，將人臉建模；在這其中我們有一項新的突破，亦即於皮膚的毛孔、皺紋，以及透光度的呈現上均有很好的效果，故在模擬真實的人臉時，才能讓人無法分辨出真偽，這些技術成功地套用在現今的 3D 電影跟動畫。

引言

近年來的電影越來越多運用到 3D、虛擬場景的技術，尤其在去年的阿凡達電影中，不僅集結了特效，更在虛擬空間中建構了似真的人臉，在皮膚、臉部紋理都有很好的效果，但這是怎麼做的呢？又有甚麼是有別於之前電影所應用的技術嗎？以下我們將一一對此做介紹。

一、如何在虛擬的場景中置入現實的物體或人物達到擬真的效果

傳統在 2D 上我們運用將前景、背景做分離這個技術，融合前景到另一個背景中；現今有 3D 模型可以直接透過影像生成出我們要的場景，惟生成的品質好壞則有所不同。

1. 電影特效的介紹

在這邊我們可以將電影特效大致上分為兩部分來講解。

- I、從早期的將前景、背景做分離，再將人物或前景的物體置入新的背景中；但如今的場景，皆已可藉由 3D 模型去生成影像。
- II、故 3D 模型所生成影像的品質好壞，則有其差異所在。



臺灣大學電機工程學系

10617 台北市 大安區 羅斯福路四段一號

Email: dept@cc.ee.ntu.edu.tw

http://www.ee.ntu.edu.tw/





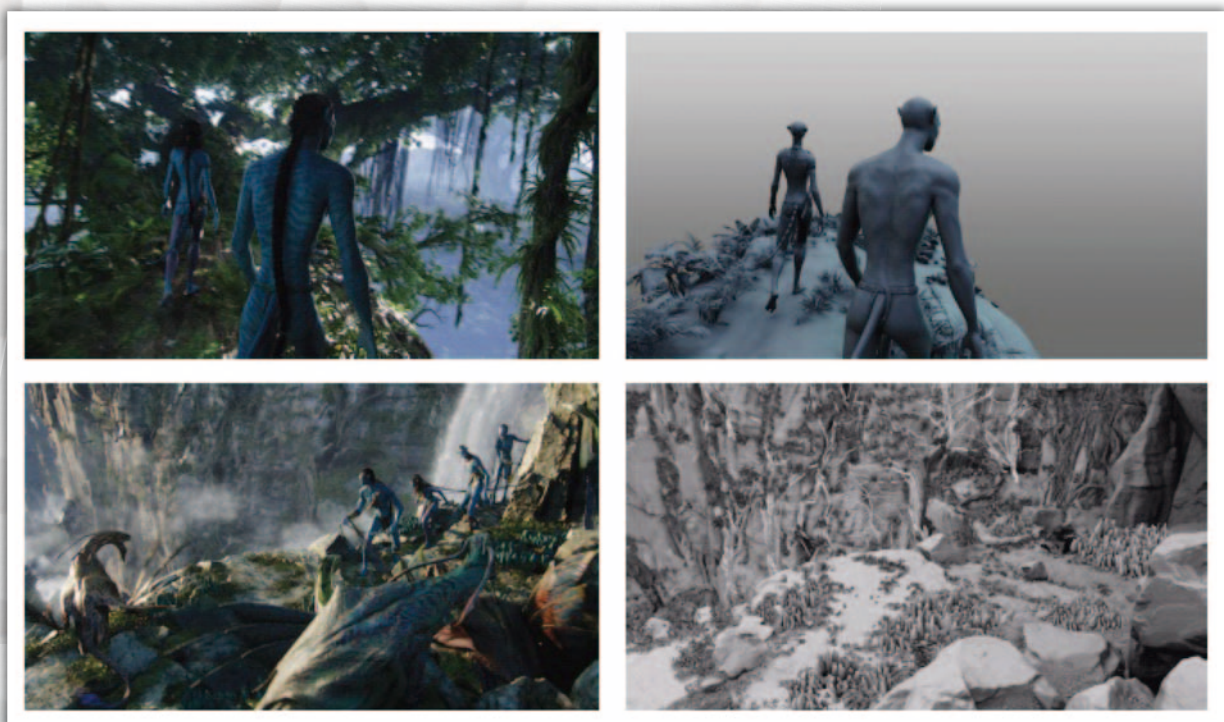
在近代電影中，常常可以在影片中看到極壯觀的場景、或是很多不太可能由真實人物在現場演出的場景，如：在空中飛翔、人在外太空、或潛入很深的深水之下 … 等；在實際上拍攝的時候，我們可能無法找到這種景觀，或無法在那樣的環境下進行拍攝；此時就需要對影片進行後製，我們稱此為一種電影特效，亦即一種虛擬實境（Virtual Reality）。

我們經常可於電影花絮中，看到演員站在以藍色或綠色的布為背景的空間中，進行拍攝；這樣做的目的，在使前景角色和背景之間產生顏色差異，經由後製，可以將綠幕（或藍幕）的部分去除，再透過電腦圖形技術，可以將特效場景與拍攝人物合成在一起，得到彷彿置身於真實場景的效果。

隨著電腦技術進步，早就已經可以透過一些 3D 模型生成軟體，如：3dmax、Maya … 等，生成 3D 模型，且將之置入所想要的場景中；但根據經驗來說，我們知道這是非常耗費人力及時間的；故在近期的電影中，採用多張影像去建出 3D 模型（成品如圖一、二），尤其在圖二中，我們可以知道，經由 3D 模型去生成影像的品質已經越來越好！那這些是透過甚麼技術而達到呢？在之後的文章中，我們會一一詳述之。



圖一



圖二 此為電影阿凡達中的場景之一，其背景即是透過 3D 模型所生成的影像。





2. 如何塑造人臉

在電影阿凡達中，我們可以看到納美人臉部上擁有許多豐富的表情，無論是喜悅、憤怒、悲傷、齜牙裂嘴，更甚於是惶恐、無辜 … 等，諸如人類才有的細微表情，都被栩栩如生的呈現出來。大家都知道，真實世界中沒有所謂的納美人，他們都是由電腦所製作出來的 3D 模型，來模擬真實的人臉（如圖三所示）。

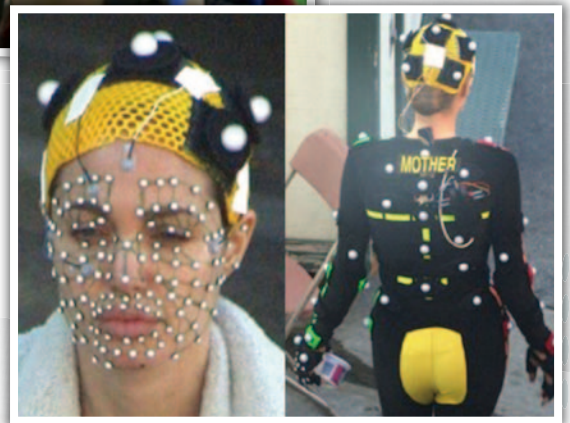


圖三

二、但這些擬真的表情是如何從人臉模擬至 3D 模型上呢？

其實這樣的技術早在《貝武夫：北海的詛咒》中即被大量地使用，在貝片中，無論是安潔莉娜裘莉（女妖），抑或是雷溫斯頓（貝武夫）均是 3D 虛擬出來的角色，我們可以看到貝片中人物的臉部外形、表情、動作，其實都與真實人物相去不遠。

要取得與真實人物相似的結果，主要是讓演員穿上佈滿感應球的服裝，並且在臉部依照肌肉及五官的位置標記適量的感應點（如圖四所示），再藉由攝影機拍攝下來的多張圖片，去推估每個感應元件的 3D 座標。



圖四

將那些 3D 座標對應至 3D 數位演員模型上，便可驅動數位演員的臉部表情跟動作；這樣的技術被稱為 3D 動態捕捉技術。

要做到近真實臉部表情的模擬，最主要的問題在於如何精確地推估出臉上各個感應點的 3D 座標，而這些感應點在此我們又稱作特徵點。推估 3D 座標有很多種做法，常用的做法是藉助多張不同角度的圖片，求出不同圖片之間，特徵點的彼此對應關係。我們的實驗室很早就正在進行相關的研究與應用。在 2000 年的時候，我們設計了一套可以讓虛擬人臉演講的系統，將其做成網頁瀏覽器的外掛程式。這套系統會針對影片的演講內容做分析，藉由得到的發音特徵，來驅動虛擬人臉做出相對應的嘴型跟表情，並且可以即時呈現出來（如圖五）。



圖五



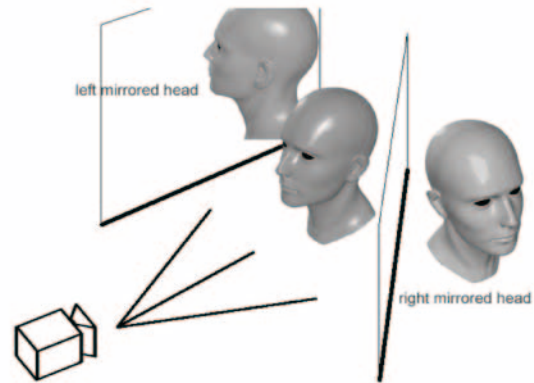
在當時，我們就利用了前文所述的動態捕捉技術，用了八台攝影機去做同步拍攝的工作，進一步建立虛擬角色的表情資料庫。然後配合輸入語音取得適當的資料，來操控我們的「演講者」。

我們對表情模擬的重點，主要在處理臉部肌肉及五官位置的改變而形成不同表情的變化。因此，在設置特徵點時，我們會著重在額頭、眉邊、雙頰、嘴唇附近為主，這些位置皆是呈現表情時臉部主要變化的部分。當然，若希望要求更精密的表情變化，則會散布更多且適量特徵點在臉上（如圖六所示），但通常特徵點的數量是有所限制的，太密集的特徵點，反而會因為要處理在不同角度攝影機的特徵點的對應關係，造成計算上或是人為處理上的困難，甚至對應上的誤判，這些錯誤會導致推估空間中位置的誤差或計算出不正確的空間位置。



圖六

在 2002 年，我們提出了另外一套設計，主要的改良在於簡化了拍攝所需要的硬體設備，卻依舊可以維持一樣的模擬精準度。這裡的基本概念，是藉由鏡面的反射，進而得到不同角度物體的影像（如圖七所示）。



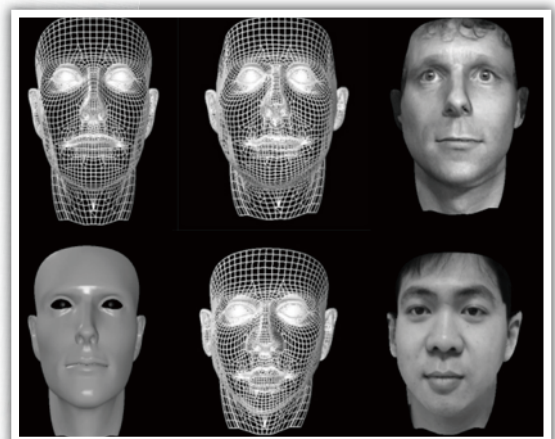
圖七 左圖中，除了中間為真實人臉外，其餘左右兩側皆為鏡子反射的影像。

◎ 接下來我們將更深入地說明擬真人臉的作法。

3. 表情資料庫

有了物體不同角度的影像以及攝影機資訊，我們便只須針對人臉上的特徵點找到鏡中人臉反射影像相對應的特徵點，即可藉由投影公式計算此特徵點在 3D 空間中的位置；而隨著我們的演員做出不同的表情變化，臉上特徵點所對應到的 3D 座標也會產生變化，這些 3D 座標的變化就代表了表情變換跟頭部的動作。（如圖八所示）

我們想要單純抽取出人臉部的表情變化，卻不太可能在拍攝的過程當中，限制演員完全不能移動頭部。所以我們必須從圖片中估計出頭部移動量，並將其移除，才會得到我們真正想要表情變化的參數。最簡單的方式便是在標記人臉部特徵點時，部分選取的點與表情無關，只與頭部的移動相關。



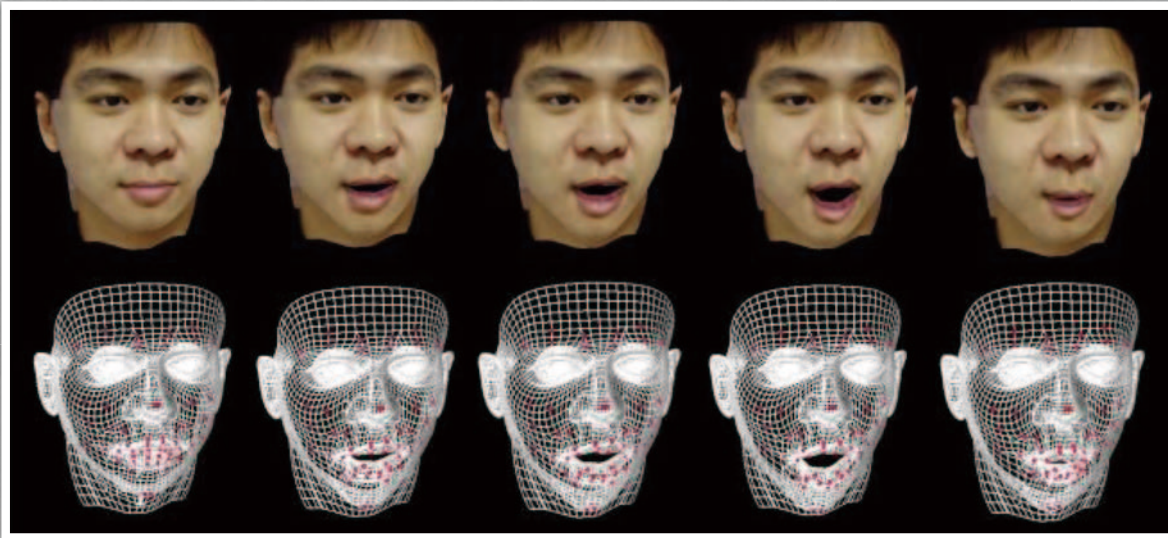
圖八



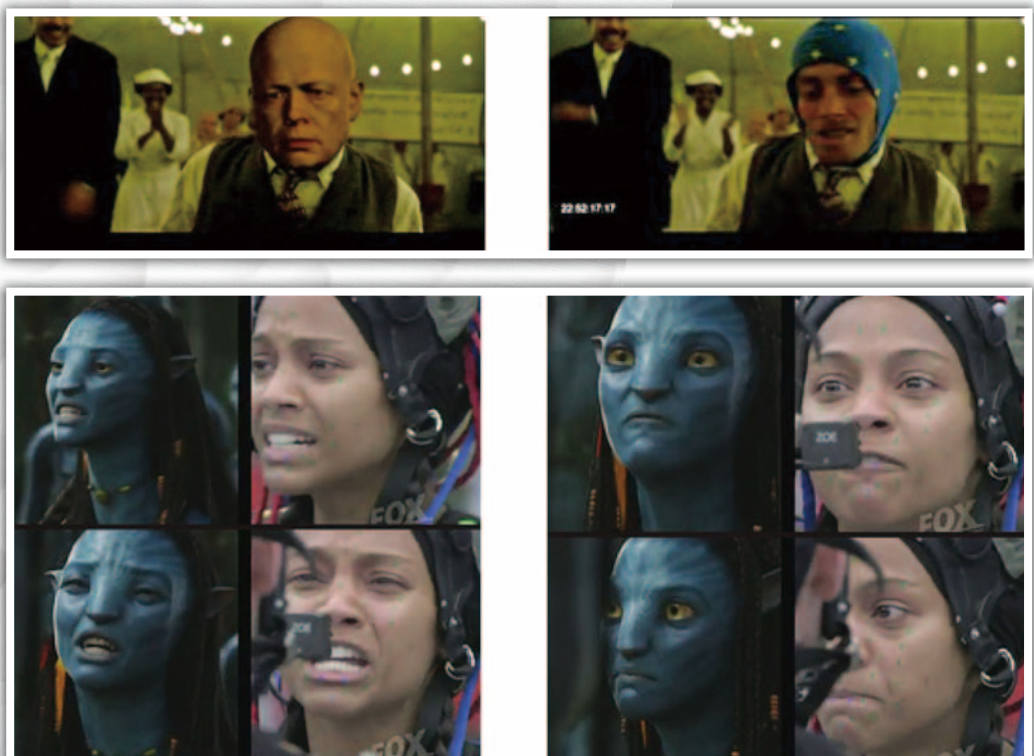


基本上可以選擇靠近耳朵的臉頰部分，抑或是直接在耳朵上標記特徵點，當我們最後抽取影像表情特徵點位置時，即可將所有 3D 特徵點座標先移回坐標系的基準，再取表情特徵點位置，此處取得的位置即不受頭部移動的影響，而是單純表情參數。最後我們將這些表情特徵點位置的變化套用在我們所生成的數位演員上相對應的點，則這些數位演員便可以產生出各種近似真人的表情了（如圖九所示）。

而這些表情參數並非只可以應用在原始的角色臉上，亦可以對應到不同數位演員身上，如：魔戒的咕魯、班傑明、納美人 … 等都是很好的例子（如圖十所示）。



圖九



圖十



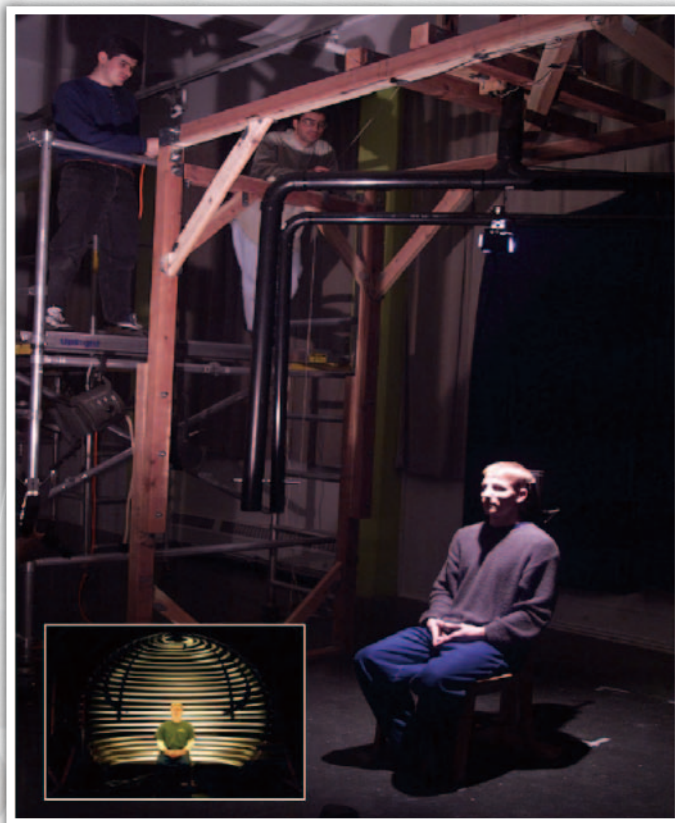


4. 特徵點用於人臉建模

當 3D 模型的長相跟參與拍攝的演員長相不一樣。這個時候我們可以拿捕捉到的特徵點，對該臉部模型直接做變形。主要利用的是臉頰、眉角、耳朵、嘴部周邊等有標記處的座標。使得此臉部模型上相對於我們所標記出的臉部特徵頂點能盡量靠近我們所求出的 3D 特徵點座標。經過此變形後的人臉模型將會相似我們影像中的人物。至此，我們已經有一個外型相似於拍攝人物的數位演員。

5. 人臉建模 – 通過圖寧測試 (Turing test)

前面提到，利用捕捉的特徵點讓既有的 3D 模型變形，以生成外型相似於拍攝人物的數位演員。然而，如此產生的數位演員模型，並無法完整模擬真實演員皮膚紋路細節，以及在不同照明條件和場景下所導致臉部膚色的反射結果，這將使得數位演員在影片中不夠真實或是與影片場景格格不入的結果。對於日益挑剔的廣大電影客群而言，這是越來越不被允許的事情。那是否還有其他的方法可以生成較為精細的臉部模型呢？事實上是有的，只是大多數的做法都需要花費很多時間及昂貴的成本，要不然就是生成結果不夠精確。



圖十一

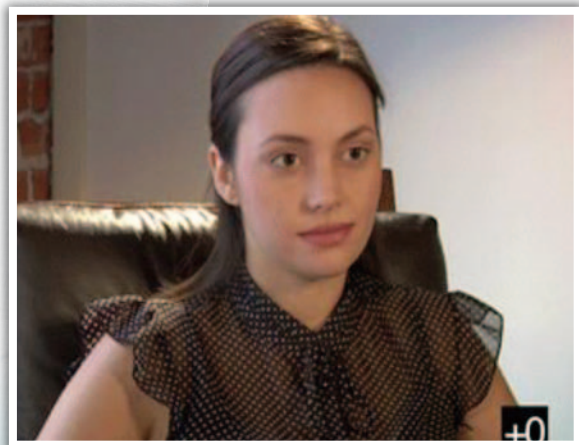
經過了幾年的努力，利用 Light stage 做人臉建模的技術漸趨成熟。

2008 年馬萬鈞博士參與的 Emily Project，成功地通過了圖寧測試，表示人臉建模在視覺上已經達到一般人眼無法分辨真偽的水準。

當時我們播放了一段女演員說話的影片（如圖十二所示），並且請大家指出哪些片段是電腦重建出來的人臉，哪些是真人。在播放完以後，幾乎所有的人都無法指出來。

在 2000 年時，美國柏克萊大學 Paul Debevec 博士的博士論文，提出了一種名為 Light stage 的攝影裝置（如圖十一所示）。Light stage 在拍攝人臉的時候，會加入許多經過設計的光照條件，例如：不同光源位置或者改變光射出的樣式。如此的光照條件使得人臉的每個細部都能夠具有明顯強烈的特徵；不只如此，這些特殊的光照條件也有助於影像間做出精確且大量的對應，精細的程度甚至可以使各別影像中每個像素在其餘影像中找到相對應關係的像素，這樣的結果可以使我們在空間中計算出成千上萬的 3 維座標點，我們稱為 Point cloud。這跟動態捕捉技術只能捕捉到少數的感應點，產生幾個點是不一樣的。這項技術的提出有助於重建精確的人臉模型。

Paul Debevec 博士之後到南加大任教，繼續跟他的學生改良 Light stage。我們實驗室的馬萬鈞博士過沒多久也參與這個團隊，合作研究虛擬人臉建模。



圖十二

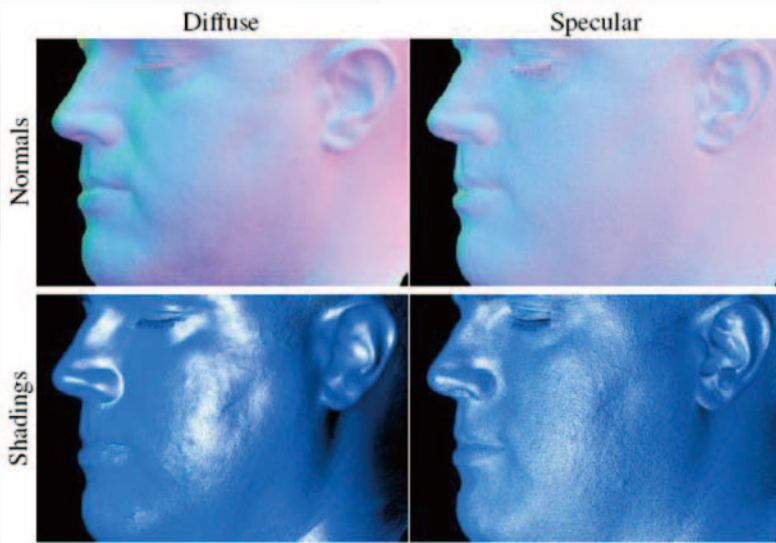




因為電腦重建出來的 3D 模型，在皮膚的毛孔、皺紋，以及透光度的呈現上已經有了非常真實的成果，以致一般人無法分辨真偽（如圖十三所示）。



圖十三



圖十四

達到這樣的技術突破，背後的概念是甚麼呢？我們可以將人臉表面的材質想成兩層。一層是粗糙皮膚，一層是油脂皮膚。粗糙皮膚對入射光的反應是散射反射，這種反射強度與觀測者所在的位置無關。油脂皮膚對入射光的反應是似鏡面反射，與觀測者的位置有關。我們可以在拍攝人臉的時候，同時對人臉打上不同的光照，並利用硬體設備，將反射到攝影機的反射光的來源分成散射反射跟似鏡面反射，就可以分別做出粗糙皮膚跟油脂皮膚的材質以及幾何性質（如圖十四）。

最後再將這兩種材質的皮膚合成，所做出來的人臉模型就非常真實。

結論

在本文中我們介紹了在阿凡達製作中所用到與我們實驗室的相關技術，並提出了對電影製作的貢獻之處，尤其在人臉模型上已經可以通過圖寧測試（Turing test），這代表我們所建出來的 3D 臉模型已可達到幾可亂真的程度。我們相信這樣的發展是一大突破，也希望這樣的視覺技術，未來能夠在虛擬實境的領域開發出更多應用。

參考文獻

1. Wan-Chun Ma, Andrew Jones, Jen-Yuan Chiang, Tim Hawkins, Sune Frederiksen, Pieter Peers, Marko Vukovicz, Ming Ouhyoung, Paul Debevec, "Facial Performance Synthesis using Deformation-Driven Polynomial Displacement Maps," ACM Transaction on Graphics, Vol. 27, No.5, Vol. 27, pp. 121:1-10, (Proceedings of SIGGRAPH ASIA 2008), Dec. 2008.
2. Michael I-Chen Lin, Jeng-Sheng Yeh, Ming Ouhyoung, "Extracting 3D facial animation parameters from multi-view video clips," pp. 72-80, IEEE Computer Graphics and Applications, Vol. 22, No. 6, Nov., 2002.
3. I-Chen Lin, Chien-Feng Huang, Jia-Chi Wu, Ming Ouhyoung, "A Low Bit-rate Web-enabled Synthetic Head with Speech-driven Facial Animation," Proc. of Eurographics Workshop on Computer Animation and Simulation 2000, pp. 29-40, (ISBN 3-211-83549-0), Interlaken, Switzerland, Aug. 2000.

