

行政院國家科學委員會專題研究計畫成果報告

虛擬與真實混合編碼標準中之即時臉部影像合成技術

Real-time Face Synthesis in a Synthetic/Natural Hybrid Coding (SNHC) Environment

計畫編號：NSC 89-2213-E-002-068

執行期限：88 年 8 月 1 日至 89 年 7 月 31 日

主持人：歐陽明 台灣大學資訊工程所

計畫參與人員：吳賦哲、何建璋、黃建峰、洪振盛、李郁中
台灣大學資訊工程所

一、中文摘要

在 MPEG-4 標準中所使用的物件模型導向壓縮技術(model-based coding)，可將壓縮比再提高，因此已蔚為一股潮流。針對視訊會議等應用，MPEG-4 亦提出了 SNHC (Synthetic/Natural Hybrid Coding)標準，對臉部影像的壓縮方式作特別的分析處理。在物件模型導向壓縮技術中，其處理步驟可大略分為追蹤頭部在三度空間上的運動(3D-head motion estimation)、三維頭部模型重建(3D-head model reconstruction)、合成臉部影像所需的頭部質紋貼圖(texture mapping)。

前一年本計畫已開發出利用追蹤眼睛及嘴巴三個特徵點，即時地(real-time)推算出頭部在三度空間上的運動。在今年的工作中，已完成三維頭部模型的重建與合成臉部影像所需的頭部質紋貼圖，以及他們之間互相的對應。

頭部模型的建立是一件相當不容易的工作，是由於我們對於人類的臉部樣子都相當熟悉，只要有一點點小小的瑕疵都會令人覺得不真實。因此，如何有一種方法能快速而非常真實的建立使用者的三維頭部模型，便成為相當受到重視的課題。本計畫的工作是利用使用者一張正面與兩張側面的照片，來調整一個通用的頭部模型(generic head model)以符合使用者，然後再產生使用者所對應的質紋貼圖，並解決產生質紋貼圖時在接合的地方出現膚色不連續的問題。

關鍵詞：三維頭部模型重建、質紋貼圖、物件導向影像壓縮、MPEG-4 SNHC

Abstract

Head modeling and facial animation is the most difficult task in modeling, because everyone is an expert in understanding delicate facial expressions. The goal of this project is to develop a method to modify a generic head model to fit into a specific person's face, and to get its texture map.

One way to have an "individualized" 3D-head

model is to get the feature points from three photos of a person in the front view and in side views, and use the corresponding information for later modification. The texture map of the 3Dhead model will be a problem if not dealt correctly, since the combined image from two side views and a front view of a person can not be without artifacts. Actually, there will be an obvious line in terms of color discontinuity in the combined texture map. To remove the color discontinuity in the texture map while still maintaining the details, a low pass filter plus a Lapacian operator are used in a multiresolution way. Finally a texture map for an "individualized" 3D-head model can be generated efficiently. Facial animations and morphing can be easily done from this point on.

Keywords: 3D-head model reconstruction, texture mapping, model-based image coding, MPEG-4 SNHC

二、緣由與目的

隨著影像壓縮技術的進步與成熟，目前已可大量地運用這些技術於一般家庭及娛樂等用途上。然而，對於目前主要的遠距傳輸媒介—網際網路(Internet)而言，現行通用的壓縮技術仍面臨一些尚待解決的問題。目前預期為下一代影音壓縮潮流的 MPEG-4，便積極引入了物件模型壓縮方式(model-based coding)的概念於標準制定中。

就視訊電話(video phone)而言，由於前景是人臉，背景是不動的環境，所以應用電腦圖學(computer graphics)與虛擬實境(virtual reality)的技術，配合影像處理(image processing)及電腦視覺(computer vision)，可利用一虛擬的頭部模型(synthetic head model)配合人臉的照片做質紋貼圖(texture mapping)來逼近真實的臉部影像(natural face image)。至於不動的背景則可用一張圖片(texture)貼到一面牆表示。鏡頭(camera)的移動也可用各式矩陣轉換(affine transformation)解決。如此做大約可以比現在的 H.261/H.263 再有效減少 75% 至 90% 的資料量。

三、結果與討論

我們利用一個符合一般平均值的通用三維頭部模型(generic 3-D head model)與一個人的一張正面與兩張側面的照片,藉由人工標明少數特徵點的方式,利用三維模型形變(3D model morphing)的技巧即可得到此人專用的頭部模型。得到這個足夠逼真的人頭模型之後,就需要一個快速而合理的質紋貼圖(texture mapping)方法來真實地呈現人頭模型。

當我們已經有欲建構的人頭模型以及欲貼於其上的一張正面與兩張側面照片時,要做質紋貼圖最直接的方法就是將人頭模型先投影在正面以及兩張側面平面上之後,與三張照片上的每一個點互相對應,就可以知道原先的模型上的每一點應該塗上什麼顏色值。但是這樣一來在三張照片接合的地方將因為拍照時的膚色或是照片光源不同等原因,使得接合的地方出現膚色不連續的情形。為了避免這種情形,我們需先將這三張照片依其相同的特徵線接合之後,利用高斯運算子(Gaussian operator)將接合的地方做模糊化處理,使得膚色看起來較自然連續,並取得一部份高頻的資訊使更影像清晰,再利用質紋貼圖的技巧將照片貼於人頭模型上。

在本計畫結果相關應用方面,本計畫所快速產生之人頭模型,目前已應用於本研究單位之 very low-bit-rate web-enabled synthetic talking head 的系統中,此系統利用 MPEG4 所定義的人臉動畫控制參數(Facial Animation Parameters)來驅動 talking head 的嘴形與臉部表情變化。此 Model-based animation coding 系統,相較於傳統的 frame-based video 系統,由於只需傳送人臉特徵點的變化,而不需要對每個畫面的人臉影像進行壓縮,所需要的網路頻寬極低,可達到 14Kbps (含 8Kbps 的臉部控制資料與 6Kbps 的語音壓縮資料),可使用於一般家用的數據機系統。

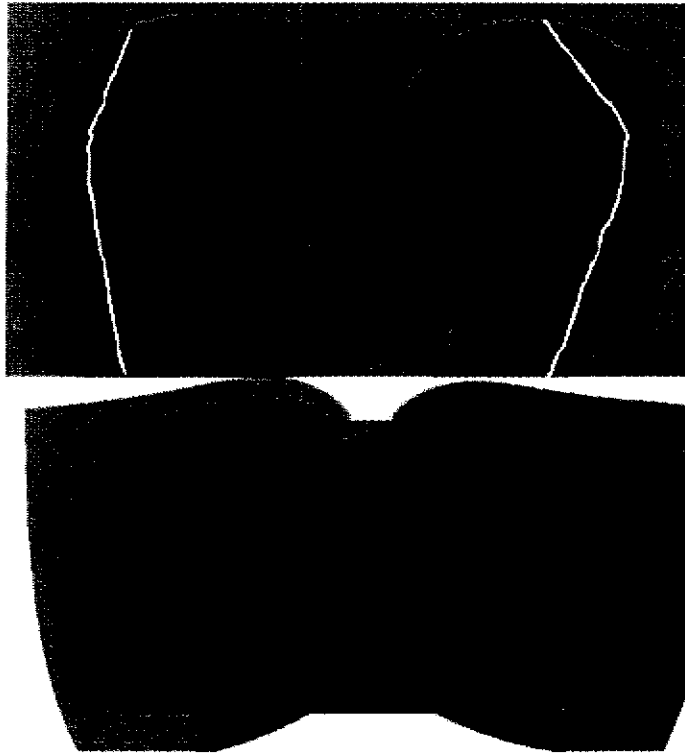
四、計畫成果自評

本計畫之結果除了已應用於 very low-bit-rate web-enabled synthetic talking head 的系統外,並有多篇相關之國際會議論文發表[16,26,27,28]。

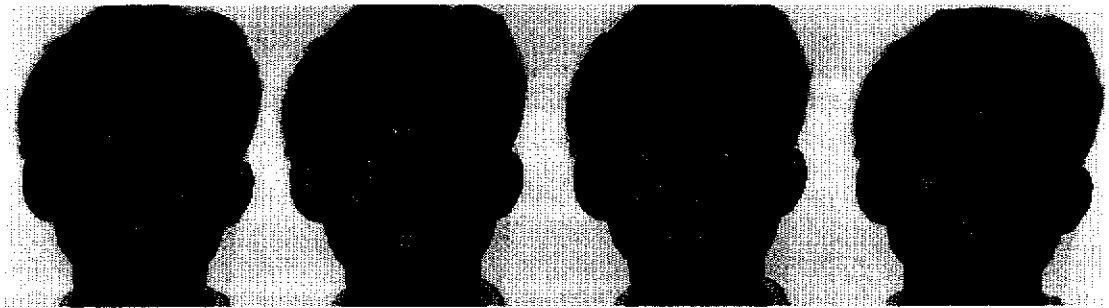
五、參考文獻

- [1]. MPEG4 Systems Group. Text for ISO/IEC FCD 14498-1 Systems, ISO/IEC JTC1/SC29/WG11 N2201, 15 May 1998.
- [2]. J. Ostermann, Animation of Synthetic Faces in MPEG-4, Proc. of Computer Animation, pp.49-51, Philadelphia, Pennsylvania, USA, June 8-10, 1998.
- [3]. Demetri Terzopoulos, Keith Waters. Analysis and synthesis of Facial Image Sequences using Physical and Anatomical Models, IEEE Tran. On Pattern and Machine Intelligence, 15(6), Jun.1993, pp.569-579.
- [4]. Frédéric Pighin, Jamie Hecker, Dani Lischinski, Pichard Szeliski, David H. Salesin. Synthesizing Realistic Facial Expressions from Photographs, Proceedings of ACM Computer Graphics (SIGGRAPH 98), pp. 75-84 Aug-1998.
- [5]. B. Guenter, c. Grimm, D. Wood, H. Malvar, F. Pighin. Making Face, Proc. of Computer Graphics (SIGGRAPH '98), pp. 55-66, Aug. 1998.
- [6]. Won-Sook Lee, Nadia Magnenat Thalmann. Head Modeling from Pictures and Morphing in 3D with Image Metamorphosis Based on Triangulation, Proc. CAPTECH'98, Geneva, pp.354-267, 1998.
- [7]. C. Bregler, M.Covell, M.Slaney. Video Rewrite: Driving Visual Speech with Audio, Proc. SIGGRAPH'97, pp.353-360, 1997.
- [8]. Matthew Brand. "Voice Puppetry", Proc. SIGGRAPH'99, pp.21-28, 1999.
- [9]. Woei-Luen Perng, Yungkang Wu, Ming Ouhyoung. Image Talk: A Real Time Synthetic Talking Head Using One Single Image with Chinese Text-To-Speech Capability. Proc. of PacificGraphics 98, pp. 140-148, Singapore, Oct 1998.
- [10]. I-Chen Lin, Cheng-Sheng Hung, Tzong-Jer Yang, Ming Ouhyoung. "A speech Driven Talking Head Based on a Single Face Image", pp.43-49, Proc. of PacificGraphics'99, Seoul, Oct. 1999.
- [11]. Thaddeus Beier, Shawn Neely. Feature-Based Image Metamorphosis", Proc.of SIGGRAPH 92. Computer Graphics, pp. 35- 42, 1992.
- [12]. Steven M.Seitz, Charles R. Dyer. View Morphing, Proc. SIGGRAPH 96, pp. 21-30.
- [13]. Eric Cosatto, Hans Peter Graf. Sample-Based Synthesis of Photo-Realistic Talking Heads, Proc. of Computer Animation 98, pp. 103-110, Philadelphia, Pennsylvania, June 8-10, 1998.
- [14]. Microsoft Speech Technology SAPI 4.0 SDK, <http://www.microsoft.com/it/projects/sapisdk.htm>
- [15]. Applied Speech Technologies Corporation. <http://www.speech.com.tw>
- [16]. Tzong-Jer Yang, I-Chen Lin, Cheng-Sheng Hung, Chien-Feng Huang and Ming Ouhyoung. Speech Driven Facial Animation, pp. 99-108, Proceedings of Computer Animation and Simulation Workshop'99, Milan, Italy, Sept. 1999.
- [17]. M.Esoher and N.M. Thalmann. Automatic 3D Cloning and Real-Time Animation of a Human Face, Proc. Computer Animation 97, pp.58-66, 1997.
- [18]. P.E Kmon, W.Fresen. Facial Action Coding System: A Technique for the Measurement of Facial Movement, Consulting Psychologists Press, Palo Alto, CA, 1978.
- [19]. S. Morishima, H.Harashima. A Media Conversion from Speech to Facial Image for Intelligent Man-Machine Interface, IEEE J. Selected Areas in communications, 9, pp. 594-600, 1991.

- [20]. M.M. Cohen and D.W. Massaro. Modeling co-articulation in synthetic visual speech. In N.M. Thalmann and D. Thalmann, editors, *Models and Techniques in Computer Animation*. Springer-Verlag, 1993.
- [21]. K. Waters and T. Levergood. An Automatic Lip-Synchronization Algorithm for Synthetic Faces. In *Proceeding of ACM Multimedia*, pp. 149-156, San Francisco, CA, USA, 1994, ACM Press.
- [22]. Gregory M. Nielson. Scattered data modeling, in *IEEE Computer Graphics and Applications*, 13(1), pp.60-70, Jan. 1993.
- [23]. Thomas S. Huang, and Arun N. Netravali. Motion and Structure from Feature Correspondences: A Review, in *Proceedings of the IEEE*, 82(2), pp. 252-268, Feb. 1994.
- [24]. H. Goldstein. *Classical Mechanics*. MA: Addison Wesley, 1980.
- [25]. S. D. Blostein and T. S. Huang. Algorithms for motion estimation based on 3-D correspondences, in *Motion Understanding*, W. Martin and J. K. Aggrawal, Eds. Norewell, MA: Kluwer, 1988.
- [26]. Chien-Feng Huang, I-Chen Lin, Ming Ouhyoung, "High Resolution Calibration of Motion Capture Data for Realistic Facial animation", Vol. 11, No. 9, pp.1141-1150, *Journal of Software*, ISSN 1000-9825, September 2000, China Computer Federation.
- [27]. I-Chen Lin, Chien-Feng Huang, Jia-Chi Wu, Ming Ouhyoung, "A Low Bit-rate Web-enabled Synthetic Head with Speech-driven Facial Animation", *EuroGraphics Workshop on Computer Animation and Sumulation'2000*, Switzerland, Aug. 2000
- [28]. Ming Ouhyoung, I-Chen Lin, David S.D. Lee, "Web-enabled Speech Driven Facial Animation", (invited speaker), *Proc. of ICAT'99* (International Conference on Artificial Reality and Tele-existence), pp. 23-28, Dec., Tokyo, Japan.



圖一。利用正面與側面影像接合人頭材質。左上為原始正面影像，右上圖為原始側面影像，下圖為接合過之人頭材質。



圖二。將合成之人頭模型做誇張的臉部表情變化。



圖三。人臉建構之微調工具。