行政院國家科學委員會專題研究計畫成果報告

※※※※※※※※※※※※※※※※※※※※※※※※※※※
※ ※
※ **近** MPEG-4 **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 an MPEG-4 Like ※
※ Synthetic/Natural Hybrid Coding (SNHC) Environment ※
※ ※
※※※※※※※※※※※※※※※※※※※※※※※※※※※

計畫編號：NSC 88-2213-E-002-050

計劃期限：民國 87 **年** 08 **月** 01 **日至民國** 88 **年** 07 **月** 31 **日**

計劃主持人　：歐陽明　教授
共同主持人　：無

處理方式：╎ 可立即對外提供參考
- 一年後可對外提供參考
- 兩年後可對外提供參考
(必要時，本會得延展發表時限)

執行單位：國立臺灣大學資訊工程學系
通訊與多媒體實驗室

中華民國八十八年八月

(1) 計劃摘要：中文部份

近幾年來，由於影像壓縮技術(H.261, H.263, MPEG-1, MPEG-2)已漸漸成熟，相關影音產品在消費性市場的應用範圍愈來愈廣。然而，對一般的使用需求而言，現有的壓縮技術仍無法滿足低資料傳輸頻寬 (low bit-rate transmission)的要求。現有的資料傳輸環境，多以電話線或網際網路(Internet)為主。前者的頻寬目前多半為33.6k bits/sec – 56k bits/sec, 而後者的頻寬容量雖大，但使用者眾多, 在資源有限的情形下，每人所能獲得的頻寬亦不高。因此，如果想透過網際網路進行影音資料的傳輸，必須再大幅降低壓縮後的資料量。

傳統的資訊理論壓縮方式(information-theoretic coding)在壓縮比上已到達一個瓶頸，而最近在MPEG-4標準中所使用的物件模型導向壓縮技術(model-based coding)，則可將壓縮比再提高，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。其處理步驟可大略分為臉部分析(face analysis)與臉部合成(face synthesis)兩部份。其中在臉部合成(face synthesis)部份，牽涉到如何自動地(automatically)、即時地(in real-time)以有限的資料合成各種角度的臉部表情。此部份的先前研究成果多必須依賴人工或耗費較多時間才能達成，對於即時性與自動化的要求則尚未有較成熟的成果，而MPEG-4 SNHC標準也尚未制定完成，即使制定完成，在技術方面也不會加以規範，因此，目前全世界均開始投入此一方面的研究。

此項技術臉部合成(face synthesis)不僅可應用在視訊會議(video conference)等即時遠距溝通上，也可應用在其他影音資料的傳送與儲存上。以多媒體電子郵件(multimedia E-mail)為例，藉由降低影音資料量，可在現有的電子郵件協定之限制下，傳送更多、更長時間的影音資訊；再以影音資料庫為例，應用此一技術，也可降低儲存空間的需求。本計劃將針對即時臉部合成的技術進行研究，並實作一系統以驗証所提出的方法確可達成即時性與自動化的要求。

(2) 計劃摘要：英文部份

In recent years, because of the success of video coding techniques (e.g., H.261, H.263, MPEG-1, and MPEG-2), more and more AV (audio/video) products rolled out. Nevertheless, due to limited network bandwidth, these video coding techniques still cannot be fully applied to applications such as distant video communication. At this moment, most home users connect to the Internet via telephone line, with a 36.6K or 56K modem; Even for a user that uses LAN to connect to the Internet, because of large volume of users, one can really use only fractions of network bandwidth. Therefore, if we want to transmit audio and video data via the Internet, we have to further reduce data bit-rate.

The model-based coding technique adopted by MPEG-4 is a potential solution to this problem. Compared to information-theoretic coding methods, model-based coding method offers a higher compression rate, thus reduces the data bit-rate. For applications, such as video conference, MPEG-4 proposes SNHC (Synthetic/Natural Hybrid Coding) standard, which is specifically for face (and body) image compression. SNHC can be divided into two parts: analysis and synthesis. In the face synthesis part, we have to synthesize a face image automatically from limited source images in real-time. Previous works usually require additional user interactions, or are time-consuming, furthermore, how to synthesize a face image automatically in real-time is still an open problem. Even in MPEG-4 SNHC, the exact solution hasn't been proposed, and won't be defined until October 1998. Because of its high potential value, many researchers around the world are heavily involved in this area.

Besides applications in distant video communication, this technique can also be applied to applications for databases and off-line transmissions of video data. For example, with the same constraint of disk storage space introduced by E-mail protocol, more video data can be delivered using this model-based coding method. For a video database, more video data can be stored within the same storage size. In this project, we will investigate model-based coding methods, and implement a real-time face synthesis system to show the result of the work.

(3) 本計劃之背景、目的及重要性

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。對大部份網際網路使用者來說，不論是透過電話線(以 33.6k~56k bits/sec modem 連接)或 LAN(Local Area Network)連上網際網路，都必須面臨頻寬有限所造成的低傳輸速率問題。網路頻寬的大幅躍進是解決方法之一，資訊再進一步壓縮則是解決方法之二，並且相輔相成。

現行通用的壓縮技術，如 H.261、H.263、MPEG-1 及 MPEG-2 等，皆是基於"資訊理論" (information theory)的壓縮方式，對於影像資料的壓縮，並不考慮其中所含的影像是否有特殊的模式(model)可供進一步分析處理之用。目前逐漸盛行的物件模型壓縮方式(model-based coding)，則針對特定用途的影像資料，分析其固定模式，並利用分析結果作資料的進一步壓縮。

在視訊影像上也存在著類似的模式。如果以單純的視訊電話(video phone)來說，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。據研究估計，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。

在 MPEG-4 SNHC(Synthetic/Natural Hybrid Coding)中，針對人臉及軀體的靜態、動態特性作分析，抽取出具代表性的特徵點，稍後再利用這些特徵點合成所需的人臉或軀體動作。在人臉合成部份，如何將壓縮後的資料重新復原為壓縮前的影像，牽涉到利用三維電腦圖學(3D computer graphics)對影像或三維模型(3D model)作各種轉換或處理。

本計劃的目的便在於繼續先前的研發成果，特別針對自動的(automatic)、即時的(real-time)臉部影像合成(face synthesis)部份，提出一可行的技術，並應用於實際的系統中。

由於 MPEG-1 及 MPEG-2 的成功，可預期在效能上有更佳表現的 MPEG-4 亦會在未來的影音標準上佔有相當大的地位。目前全世界各大研究群(如 Swiss Federal Institute of Technology, EPFL)，及電腦、消費性電子廠商已開始投入經費與人力於此一研究領域，以期居於領導地位。此項自動的、即時的臉部影像合成技術，對與 MPEG-4 SNHC 類似的壓縮方式而言，佔有相當重要的地位。如果欠缺此部份的技術，便無法達到以物件模型(model-based)進行壓縮的目的。而且此項技術不僅可應用在 MPEG-4 SNHC 上，也可在諸如視訊資料庫(video database)等視訊存取相關的應用上發揮其降低資料量的優點。

(4) 研究方法及進行步驟

如何由特徵三角形推算出三維的運動方向，可利用求下列方程式之極小值的方法獲得一近似值：

$$\sum_{i=1}^{3}\left(\left|X_i^{'}-X_i\right|+\left|Y_i^{'}-Y_i\right|\right),\ \text{where } X_i=D\frac{x_i}{z_i},\ Y_i=D\frac{y_i}{z_i},\ \text{and } \begin{pmatrix}x_i^{'}\\ y_i^{'}\\ z_i^{'}\end{pmatrix}=R_zR_yR_x\cdot\begin{pmatrix}x_i\\ y_i\\ z_i\end{pmatrix}+T,\ i=1..3$$

其中， (X_i, Y_i)**及** (X_i', Y_i')**分別是三維座標點** (x_i, y_i, z_i) **及** (x_i', y_i', z_i')**的透視投影點，而** (x_i', y_i', z_i') **則為** (x_i, y_i, z_i)**經過某種旋轉及平移後所獲得的新座標。同樣地，** R_x**、** R_y**、** R_z **分別為繞** X **軸、**Y **軸、**Z **軸旋轉的矩陣，而** T **則為平移矩陣。** D **為相機**(camera)**與投影面**(projection plane)**間的距離。**

此方程式的意義在於求得一組轉換矩陣(transformation matrix)**，使得轉換前的投影座標點** (X_i, Y_i)**，經過此一轉換矩陣的作用後，可以逼近已知的轉換後投影座標點** (X_i', Y_i')**。目前利用**(iterative)**的方法，可以在極短的時間內求得一組可能的解，其結果如圖** 1 **所示，上面五個三角形代表真實臉部旋轉或移動後的臉部朝向(或鼻尖方向)，下面五個三角形則為推算出的結果，可發現其臉部朝向非常接近。**

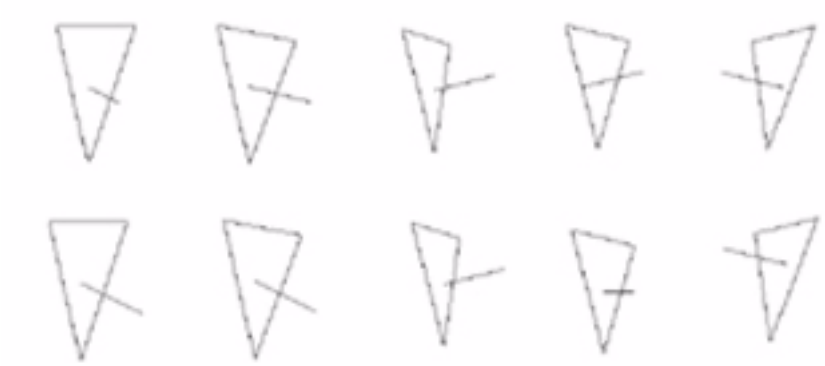

圖 1

(5) 研究成果

預期此一模組完成後，對於視訊溝通(video communication)**及其他視訊資料的壓縮應用上，會是一關鍵性的技術突破，對產業界而言，則可據以開發出技術領先的視訊應用產品(如** video phone、video conference **等)。而參與計劃的人員，由於其所掌握的為未來** MPEG-4 **等國際標準所採用的關鍵技術，因此可協助業界在未來的** MPEG-4、MPEG-7 **及其他以物件模型壓縮方式**(model-based coding)**為主的應用上，及早跨越關鍵門檻，以更快的開發速度取得領先地位。**

本計劃今年完成一即時的(real time)**、自動的**(automatic)**臉部合成**(face synthesis)**模組，其含括之功能有：**

1. **頭部的三維運動**(3D motion)**即時的偵測－能追蹤使用者頭部的三維動作。使用者不需配戴標示記號。我們可以在一人臉影像上，一開始以手動的方式，點三個點。利用眼睛及嘴巴所形成的二維特徵三角形(或其他可利用的特徵資訊)，推算出其相對應的三維運動方向。之後便利用影像比對**(template matching)**的技巧，持續追蹤這三點的位置。而透過這三點的形狀大小的改變，可以推算出頭部在三維空間的運動。此推算過程必須符合即時性及自動化的要求，並必須能找到一組可能解。**

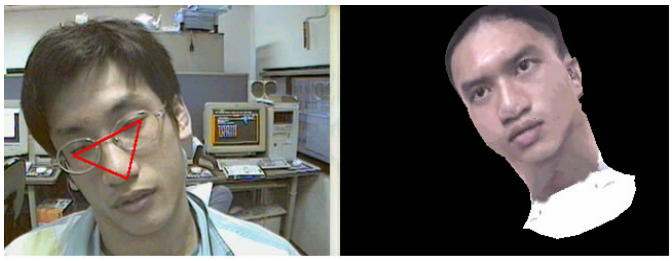

圖 2.**右圖之人頭模型隨著左圖真人而擺動**

2. **擬真的**(realistic)**自動臉部合成技術**(automatic face synthesis)**。取得此有限的臉部影像資訊，合成出的人臉影像。可以利用** MPEG-4 **中有關** facial animation **的方法，也就是** MPEG-4 **中所定義的** Facial Animation Parameters (FAPs) **和** Facial Definition Parameters (FDPs) **來讓** 3D **人頭作各種表情的變化，再加上臉部表情的貼圖，而合成新的臉部影像。**

原圖　　微笑　　生氣

圖 3.**臉部影像合成**

(6) 附錄 (重要之相關發表文獻摘錄)：

1. I-Chen Lin, Chen-Sheng Hung, Tzong-Jer Yang, Ming Ouhyoung, "A Speech Driven Talking Head System Based on a Single Face Image", to appear in Pacific Graphics'99, Oct., Seoul, Korea.
2. Tzong-Jer Yang, I-Chen Lin, Cheng-Sheng Hung, Chien-Feng Huang, Ming Ouhyoung, "Speech Driven Facial Animation", to appear in Computer Animation and Simulation Workshop'99, EuroGraphics, Sept., Millan, Italy.
3. Chien-Feng Huang, Chen-Sheng Hung, I-Chen Lin, Tzong-Jer Yang, Ming Ouhyoung, "A Real-time Model-Based Virtual Phone", to appear in CAD/Graphics'99, Nov., Shainhai, China.
4. W. Perng, Y. Wu, M. Ouhyoung,"Image Talk: A Real Time Synthetic Talking Head Using One Single Image with Chinese Text-to-Speech Capability", pp.140-149, Proc. of Pacific Graphics'98, Sungapore, Oct. 1998.
5. Tzong-Jer Yang, Fu-che Wu, Ming Ouhyoung, "Real-time 3D Motion Estimation in Facial Image Coding", pp. 50-51, Proceeding of International Conference on Multimedia Modeling'98 (MMM'98), Switzland, Oct. 1998.
6. Fu-che Wu, Ming Ouhyoung, "Automatic Feature Extraction and Face Synthesis in Facial Image Coding", to appear in International Symposium of Consumer Electronics'98 (ISCE'98), Taipei, Taiwan.