

# 行政院國家科學委員會專題研究計畫成果報告

計畫名稱：DCT-領域上之多點連續視訊會議系統結合器物

計畫編號：NSC89-2213-E-002-018

執行期限：88年8月1日至89年7月31日

計畫主持人：吳家麟 教授

執行機構：國立台灣大學資訊工程學系

## 一、中文摘要

視訊會議系統在可預見的未來，將成為人們最主要的通訊方式之一。目前的視訊會議系統主要的問題之一，即在於只能支援點對點的通訊。本計畫旨在研發一種定義在離散餘旋轉換領域上的視訊結合方法，可支援至多六名使用者的多點連續視訊會議。此方法是在 H.261 或 H.263 國際視訊編碼標準的巨集區塊單元之上，將多個編碼後的視訊資料結合成一個合法的編碼後視訊位元流，由單一解碼器完成解碼工作。本計畫涵蓋了三個主要的視訊結合技術課題，包括視訊同步機制，視訊框架結合方法，以及延遲累積的克服技術。首先使用重新對映時間上的參考來達到視訊的同步，接著將同步後的視訊框架在視訊編碼標準中所定義的巨集區塊單元作結合。針對視訊框架結合時因巨集區塊位於不同的區塊層所造成的問題（包含位移補償及量化因子設定）。我們使用重新計算位移以及協議訊號的方式來解決。經由對延遲累積的

分析，我們修正視訊同步的方法，並且證明了本計畫提出的方法會將延遲累積限制在可接受的範圍內。最後，我們已實作出一套至多同時可以支援六名使用者參與視訊會議之視訊結合器。詳見[16]

關鍵詞：視訊會議，H.261，H.263，視訊結合器

Abstract

Video conference system will become one of the major approaches of our communication in the predictable future. However, one of the most important problems of current video conference system is that only point to point communication is supported. In this project, we try to derive a video combining scheme in DCT domain to support multipoint continuous presence video conference up to six users. This scheme combines several coded video streams into single legal coded video stream, based on macroblocks of H.261 or H.263 video coding standard.

This project contains three primary technical issues, including frame synchronization, frame combining scheme, and accumulation of delay. First, we remap temporal reference to synchronize video stream, and combine those synchronized video frames. We use motion vector recomputation and handshaking signal to solve the problems including motion compensation and quantization, which are caused by macroblocks come from different GOB's when video combining is performed. According to the analysis of accumulation of delay, we modified the method of video synchronization, and proved that accumulation of delay can be restricted with an acceptable range. Finally, we have implemented a video conference system combiner supporting up to six users [16].

Keywords : Video Conference , H.261 , H.263 , Video Combiner

## 二、緣由與目的

由於電腦技術在資料壓縮技術，電腦網路技術，電腦硬體技術等等方面均突飛猛進，使得電腦通訊不在侷限於傳統文字或語音資料的傳輸，而是可以多媒體資料(包括文字，影像，動畫，圖形，視訊，聲音等)的形態來進行，視訊會議系統正是其中一項重要的應用。由於人類對傳統語音會議不再感到滿足，而且電腦價格的持續下降使得視訊會議將越來越重要且越來越受歡迎。目前已有許多的視訊會議的相關標

準規格(如 H.324)及產品問世，但是這些產品與規格其實僅俱點對點(point-to-point)的視訊會議功能，也就是僅具視訊電話(video phone)的功能，因此發展多點(multipoint)視訊會議就成為下一代視訊會議的重要研究方向。多點視訊會議通常需有一個多點控制單元(Multipoint Control Unit)，此多點控制單元接收來自各個不同參與視訊會議者的視訊訊號，然後提供一個相同的輸出視訊訊號給與所有參加視訊會議的使用者，所處理的輸出與輸入的視訊訊號通常是經壓縮過後的數位資料形式。本計畫的主旨即在研發多點視訊會議之多點控制單元的相關技術上。

多點視訊會議可分為連續呈現(continuous presence)與交互呈現(switched presence)兩種形式。交互呈現的多點視訊會議為多人同時參與但是在同一時間卻僅有一名使用者為主動(active)狀態，而只有狀態為主動的使用者畫面會呈現在所有參與會議者的螢幕上，因此雖然所有的使用者均將其視訊訊號送至多點控制單元，但是多點控制單元卻僅選擇其中一個視訊訊號傳送給所有的與會者。換言之，此種方式並沒有視訊結合的功能。交互呈現的多點視訊會議顯然並不很理想，因為同一時間僅有一名使用者可以呈現在畫面中，而其他與會者僅能觀看這個在主動狀態的使用者的畫面，卻無法立即的進行溝通。

而連續呈現的多點視訊會議中，所有的與會者的視訊畫面均會同時呈現並輸出視訊訊號，此時，多點控制單元接收來自所有與會者的使用者的輸

出視訊訊號,然後將這些視訊資料進行結合,產生一個經過結合所有與會者視訊資料的訊號,再同時傳送給所有與會者,所有的與會者接收的均是相同的視訊訊號。因此多點連續呈現的視訊會議比起多點交互呈現的視訊會議,更能模擬真實的會議情況。

對於多點連續視訊會議系統中的多點控制單元的實現方法,包括了離散餘弦轉換領域上(DCT-domain)的視訊結合,以及像素領域上(pixel-domain)的視訊訊號結合兩種。

像素領域上的視訊訊號結合,先由多點控制單元接收來自所有與會者所傳來的經壓縮過的視訊訊號,接著由多點控制單元將其完全解碼,也就是解碼至像素領域然後進行結合,再經壓縮後傳回給所有與會者。像素領域上的視訊結合之優點是:若要對個別的輸入影像作處理較具有彈性,而且甚至可以允許不同的與會者使用不同的壓縮標準,只要多點控制單元裡包含了各個不同的壓縮標準的程式庫,即可以針對不同的使用者提供不同的壓縮解壓縮服務或提供不同的位元速率(bit-rate),然而其缺點則是系統複雜度太高,而由於複雜度高的原因,此法也會造成較長的延遲時間。

而離散餘弦領域上的視訊訊號結合,則僅部分解碼後便進行視訊訊號結合,然後再編碼傳送至所有與會者。由於僅有部分解碼的動作,因此優點是複雜度較低,延遲時間也較短。然而缺點相對的是比較沒有彈性。但目前視訊會議系統多半按照 H.261 或 H.263 視訊壓縮

標準來達到,因此,在 DCT-domain 上來進行視訊結合的重要性與日漸增,這也是本計畫規劃時的主要考量。

### 三、結果與討論

本計畫所研發的多點連續視訊會議系統結合器,主要是能接受來自四個至六個終端系統所傳送經過 H.261 或 H.263 標準壓縮的位元流(bitstream)資料。每個位元流資料皆是以 QCIF 的圖形格式存在,然後經過離散餘弦領域上之視訊結合方法結合成單一的合於 CIF 圖形格式的資料流,提供給各終端系統。

DCT-領域上的視訊會議系統結合器依其定義即是不會將壓縮過的資料(也就是位元流)完全解壓還原至像素領域(pixel-domain)再作結合。而是將壓縮過的位元流資料直接作結合,或是僅將壓縮過的位元流資料作部份的解壓(partial decode)而後進行結合。

圖 1 為 H.263 支援的其中兩種的圖形格式,左邊為 QCIF 的圖形格式(176x144 pixels),右邊為 CIF 的圖形格式(352x288 pixels)。在 H.263 視訊壓縮標準中,將 QCIF 的圖形格式分為 9 個區塊群(GOB),每個區塊群分為 11 個巨集區塊(Macroblock),並且將 CIF 的圖形格式分為 18 個區塊群,每個區塊群包含 22 個巨集區塊。由於 QCIF 圖形格式與 CIF 圖形格式的區塊群大小不同而巨集區塊的大小則是相同的,因此要將數個 QCIF 的圖形格式經由視訊會議系統結合器結合成一個 CIF 的圖形格式,必須要在巨集區塊層甚或巨集區塊以下的層級來作結合。為了系統複

雜度以及實際需求的考量,我們選擇在巨集區塊層來作視訊結合的操作。由於 H.261 標準中的圖形格式與 H.263 十分類似,在此就不再贅述。

在一個多點連續的視訊會議系統中,將畫面適當的作分割是必要的,由於 CIF 圖形格式恰好為 QCIF 圖形格式的四倍大小。因此,我們的視訊結合器提供至少四個與會者同時進行視訊會議並將畫面平均分割四等份是明顯且合理的安排,見圖 2。但是由於在視訊會議上的實際需求,使用者常常希望能有額外的空間來存放以及支援共享白板或是共享文件以便利會議的進行。在這種情況下,若將畫面分割成四等分而其中一等分放置共享白板或是共享文件,則餘下的空間僅能支援至多三個使用者參與會議,但是在許多實際的情況下經常會需要超過四個以上的使用者參與視訊會議,因此如何支援四個使用者甚或超過四個使用者參與的視訊會議便成為一個重要而有趣的課題。

本計畫所研發的視訊結合器具有能提供至多六個參與者參加同一視訊會議的能力。而五個參與者及六個參與者參加時的畫面分割分別如圖 3 及圖 4。

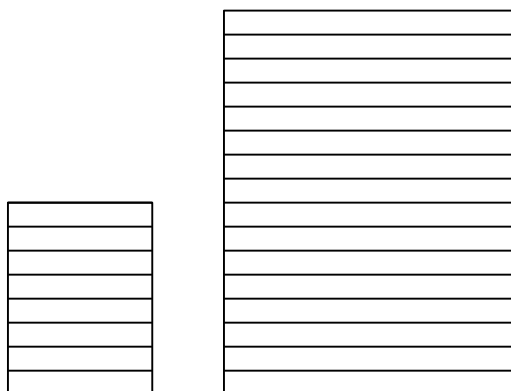


圖 1. H.263 中之 GOB 架構圖

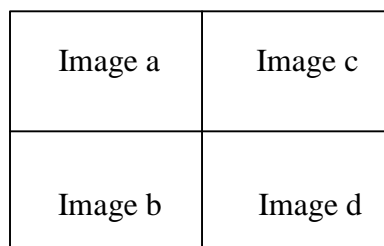


圖 2. 4 名使用者之畫面分割圖

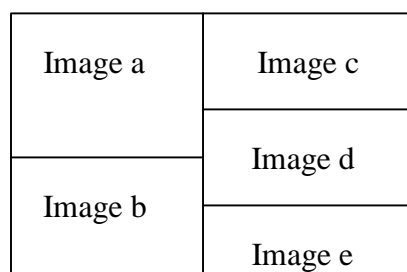


圖 3. 5 名使用者之畫面分割圖

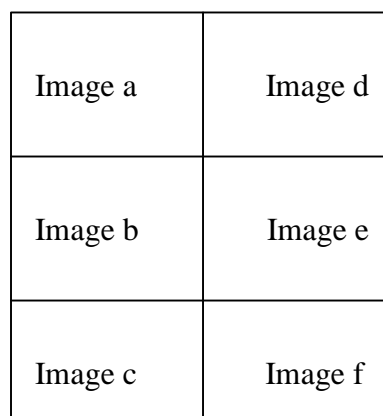


圖 4. 6 名使用者之畫面分割圖

由於不同的與會者間使用不同的終端系統。因此,也可能會具有不同的視訊框架速率處理能力(frame-rate capability)。所謂的視訊框架處理能力指的是某個終端系統最多每秒所能處理的視訊框架數,而非所處理的平均視訊框架數,這裡所稱的處理包含使用 H.263 編解碼器(codec)所提供的編碼與解碼,因此,我們稱一個終端系統具

有  $N$  frames/s 的視訊框架處理能力,表示此終端系統每秒可以編碼與解碼的視訊框架速率可以是不固定的,但是卻必須小於或等於  $N$  frames/s。舉例來說,當具有不同框架速率處理能力的終端系統要連結在一起進行視訊會議時,為了要保證所有的終端系統都能夠正確的運作,我們打算在系統連結開始時,經由一個協議的訊號(handshaking)來協調各個終端系統,使得所有終端系統在一個相同的視訊框架處理能力下工作。

雖然所有的終端系統以相同的視訊框架處理能力來運作,並不就表示所有的終端系統每秒所編碼與解碼的視訊框架數都會相同。這是因為視訊框架處理能力指的僅是此終端系統的最大能力,所以對於每個終端系統來說還是可能會有不同的視訊框架速率(frame-rate)。這代表著每個終端系統每秒所編碼的視訊框架數會不相同,但是由於 H.263 視訊框架交互參考編碼(inter-frame coding)的特性,每一個視訊框架均不可以被丟棄。一旦丟棄了某個視訊框架則往後的解碼過程所參考的視訊框架將會錯誤,因此將會得到一個錯誤的解碼結果。換言之,我們必須設計一個多工處理(multiplexing)的機制,使得在做視訊會議系統結合器的視訊框架結合過程,能夠不捨棄任何的視訊框架,而又可以達到視訊框架的同步化。

視訊結合器要將視訊資料在 DCT-domain 作結合,最大的問題來自 H.263 位移補償的方式,由於 H.263 的位移補償是採用差別位移向量的方式

(differential motion),因而使得位於視訊框架結合器的部份,產生了位移補償計算錯誤的情形。此外,由於 H.263 的量化因子可在多處改變,也造成邊界處量化因子計算錯誤的情形。本計畫針對上述兩種可能出現的計算錯誤,進行位移量重算,以達成對位移補償計算錯誤的修正工作。

延遲(delay)是影響視訊會議品質最重要的因素之一。不但直接影響了使用者對視訊會議的感知品質(perceptual quality),嚴重的延遲累積將會使得視訊會議的性能(performance)完全無法被與會者所接受。由於視訊會議的視訊資料是經過壓縮的資料,因此來自於編碼及解碼過程延遲,以及實作編碼器與解碼器時必備的緩衝器(buffer)延遲是無法避免的,稱之為固有的(inherent)延遲。

在進行多點連續視訊會議時,除了固有的延遲之外,仍會有來自於其他原因的延遲。例如在視訊訊號結合時多點控制單元所產生的非固有延遲,以及在網路傳輸時緩衝器控制不當所導致的嚴重延遲累積等。

由於非固有延遲發生的原因主要來自於視訊結合器等待資料的輸入所造成的緩衝器空溢及滿溢,和延遲的累積,本計畫提出 (i) 避免視訊結合器的等待及 (ii) 避免延遲累積等兩種方法來解決延遲所造成的問題,詳見[16]。

#### 四、計畫成果自評

##### 1. 本計畫之具體成果如下

(一)提出解決視訊結合技術所牽涉到

的三項難題；

即

- (a)視訊同步機制
- (b)視訊框架適切之結合方法
- (c)克服延遲累積的方法

(二)，實作一套視訊結合系統，提供至多可同時支援六名使用者參與的視訊會議系統。多媒體個人電腦是我們系統實作的平台。

2. 本計畫所提出之視訊結合器，不但有學術研究價值，對日漸重要的視訊會議工業或視訊安全監測系統均有直接的助益。

3. 參與本計畫之人，獲得以下的訓練，

- (a)視訊標準 H.261 及 H.263 相關之技術開發
- (b)視訊結合器所遭遇困難之克服
- (c)視訊信號與網路 Programming 相結合的知識與實作經驗
- (d)壓縮核心層與網路層同步機制之整合技術

圖五所示將是視訊結合系統完成後在各使用者終端器上可看到之畫面(此列係以同時提供五人參與會議為例)。

本計畫除系統實作外，並有 1 篇學術論文發表於[16]。



圖 5. 五名與會者的畫面

## 五、參考文獻

- [1] ITU-T Study Group XV-Report R93, Draft Recommendation H.231, "Multipoint control units for audiovisual systems using digital channels up to 2 Mb/s," May 1992
- [2] ITU-T Study Group XV-Report R94-E, Draft Recommendation H.243, "Procedures for establishing communication between three or more audiovisual terminals using digital channels up to 2 Mb/s," May 1992
- [3] ITU-T Study Group XV-Report R95, Draft Recommendation H.261, "Video codecs for audiovisual services at p x 64 kb/s," May 1992
- [4] S. M. Lei, T. C. Chen., and M.T. Sun, "Video bridging based on H.261 standard," IEEE Trans. Circuits Syst. Video Technol., vol. 4, pp.425-437, Aug. 1994
- [5] S. Oka and Y. Misawa, "Multipoint

- Teleconference Architecture for CCITT Standard Videoconference Terminals, "Visual Communications and Image Processing '92,P.Maragos,Ed.,Proc. SPIE 1818, pp.1502-1511, Nov. 1992
- [6] S. F. Chang, W. L. Chen,and D. G. Messerschmitt, "Video Compositing in the DCT Domain, " IEEE Workshop on visual Signal Processing and Communications, Raleigh, NC, pp.138-143, Sept. 1992
- [7] CCITT Study Group XV-Report XV-Report R 95,Recommendation H.221, Frame Structure for a 64 to 1920 kbit/s channel in Audiovisual Teleservices, May 1992
- [8] C. T. Chen and Andria Wong, "A Pseudo Fixed-Bitrate Video Coding Algorithm Using Self-Governing Rate Buffer Coding Strategy, "IEEE trans, Image Processing, vol.2,no. 1,pp.50-59, Jan. 1993
- [9] K. H. Tzou, T. C. Chen, and F. E. Fleischer, "An Intelligent Rate Control for an Intrafield Subband HDTV Coder, "Proc. Int'l Symp. Of Circ. And Sys., Singapore, Jun. 1991
- R. C. Lau, P. E. Fleischer, and S. M. Lei, "Receiver Buffer Control for Variable Bit-rate Real-time Video, " Proc. OfInt'l conf. On Communications '92, vol.1,pp.544-550,Chicago,IL,Jun.1992
- [10] CCITT Study Group XV- Document 525, Description of Reference Model 8 (RM8), Jun. 9,1989
- [11] ISO- IEC/JTC1/SC29/WG11,MPEG II, Test Model 5, Apr. 1993
- [12] Bellcore, "Some comments on Video Combining from 4 QCIFs to CIF," ITU-T-SG15 Contribution, Document AVC-560, Brussels, Sept. 2, 1993
- [13] H. Song, et.al., "Real-time motion-based H.263+ frame rate control," proc. SPIE on Visual Communication and Image Processing '99, [3653-02], Jan. 1999. T. Y. Kuo, et.al., "Improved H.263 video codec with motion based frame interpolation," Proc. SPIE on Visual Communication and Image Processing '99, [3653-03], Jan. 1999.
- [16]Da-Jin Shiu,Chia-Chiang Ho, and Ja-Ling Wu, "ADCT-domain H.263 Based Video Combiner for Multipoint Continuous Presence Video Conference," IEEE International Conference on Multimedia and Computing Systems, 1999, pp.77-81.