

行政院國家科學委員會補助專題研究計畫成果報告

MPEG-4 多媒體通訊技術之研究—子計畫四： 比例式視訊編碼技術及視訊通訊 終端機技術之研究 Research in Scalable Video Coding Techniques and Visual Communication Terminal Technologies

計畫類別： 個別型計畫 整合型計畫
計畫編號：NSC 89 - 2213 - E - 009 - 233
執行期間：89年8月1日至90年7月31日

計畫主持人：林大衛 交通大學電子工程系所 教授

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

執行單位：國立交通大學電子工程系所

中 華 民 國 九 十 年 七 月 三 十 一 日

行政院國家科學委員會專題研究計畫成果報告

MPEG-4 多媒體通訊技術之研究—子計畫四：

比例式視訊編碼技術及視訊通訊終端機技術之研究

Research in Scalable Video Coding Techniques and Visual Communication Terminal Technologies

計畫編號：NSC 89-2213-E-009-233

執行期限：89 年 8 月 1 日至 90 年 7 月 31 日

主持人：林大衛 交通大學電子工程系所 教授

計畫參與人員：詹益鎬、陳岱樺、周依祥 交通大學電子工程系所 研究生

一、摘要

本計畫從事兩方面之研究：其一是比例式視訊編碼法，含物件域及像素域之比例式編碼；其二是視訊通訊終端機技術，主要為網際網路視訊通訊終端系統之實作。在物件域之比例式編碼方面，我們提出一個視訊內容分割法。模擬結果顯示：其萃取得之前景物件符合人類常態知覺。在像素域之比例編碼方面，我們考慮 H.263+ 型式之空域比例式編碼，提出一個位元分配的方法，並以計算機模擬檢視其效能。在網際網路視訊終端機方面，我們發展了一個以個人電腦及數位信號處理器為平台的點對點視訊編碼與傳輸系統。

關鍵詞：物件域比例式視訊編碼、視訊分割、像素域比例式視訊編碼、網際網路視訊通訊終端機

Abstract

This project conducts research in two subject areas. The first is scalable video coding, including object-oriented and pixel-oriented scalable coding; and the second is visual communication terminal technologies, primarily the implementation of an internet visual communication terminal system. In object-scalable coding, we propose a video content segmentation method. Simulation results show that the extracted foreground objects accord with common sense of human perception. In pixel-scalable coding, we consider H.263+-type spatial-scalable coding. We present a bit allocation method and perform computer simulation to examine its performance. Concerning internet visual communication terminals, we develop a point-to-point video coding and transmission

system employing personal computers and digital signal processors as the platform.

Keywords: Object-Oriented Scalable Video Coding, Video Segmentation, Pixel-Oriented Scalable Video Coding, Internet Visual Communication Terminal

二、計畫緣由與目的

視訊通訊領域在近數年來有兩個重要的發展方向：一是視訊壓縮與傳輸之理論與技術上的創新與進步，二是真時及儲存式視訊通訊系統的實用化。本計畫兼考慮此兩方向，一方面研究比例式視訊編碼法，含物件域及像素域之比例式編碼，另一方面研究視訊通訊終端機技術，主要為網際網路視訊通訊終端系統之實作。在物件域之比例式編碼方面，我們研究能符合人類常態知覺的視訊內容分割法。在像素域的比例式編碼方面，我們考慮比例式編碼下的位元分配方法。在網際網路視訊通訊終端機方面，我們發展一個以個人電腦及數位信號處理器為平台的點對點視訊編碼與傳輸系統。在下節中，我們分別描述以上各課題的研究與結果。

三、結果與討論

A. 視訊分割

在進行 MPEG-4 的物件式視訊編碼之前，首先應對視訊畫面進行有意義的物件分割。由於 MPEG-4 對於物件的定義及視訊分割的方式均無明確的規範，因此就留給研究者極大的餘裕。

一般說來，傳統靜態影像分割的結果通常具有不錯的影像邊界，但是物件內的區域往往過度分割。相對的，運動物件估測有利於同一物件的取出，但是由於前後畫面的移動區域必定發生的遮蓋現象，因此較不易判定準確的區域邊界。而視訊中所

含的雜訊亦影響位移向量估計的準確度。因此我們試將傳統影像分割和運動物件估測的優點混合，以取出視覺上合理的前景物件(foreground objects)，並進而追蹤該前景物件投射到下一張畫面中的約略位置，然後再次重整影像分割，如此重複進行下去。我們並更進一步將前後張畫面的組合物件進行時間上的串聯，以監督物件隨時間而進行的變化情形。

我們所提出的視訊區域(物件)分割方法如圖一所示，共分初步影像分析、初步影像分割、移動估計與運動區域整合、及區域追蹤與更新等四個階段。其中第二階段僅對第一張畫面執行，而第三與第四階段則對所有後續畫面執行。

在第一階段的初步影像分析，首先針對畫面的亮度平均值將影像以區域成長的方式分割成若干亮或暗的區域，然後分別對各區域計算其亮度梯度的平均值。將各區域內梯度值低於平均值的接鄰點合併一起並予以編號，稱為種子區域。其餘位置的相鄰像素亦合併一起，編為未定區域。

第二階段的初步影像分割包含區域成長與區域合併兩步驟。區域成長係檢視所有與種子區域相接之未定區域像素，若最接近的配對距離符合成長條件，則該配對之種子區域允許成長，並且重複檢視此一步驟直到最接近的配對不符合成長條件為止。之後，區域合併將區域成長完成的結果中，低於某一面積的區域合併至相近的較大區域。

第三階段的移動估計部分是以前面得到的分割區域為本，求每點之前向位移向量。該區域內若有超過某比例的面積有移動，則稱該區域為移動區域。最後將相鄰的移動區域整合起來以形成移動物件。此移動物件可作為 MPEG-4 所定義之視訊物件(video object, VO)。

第四階段的物件追蹤與更新，是將前階段所得之視訊物件依其位移向量投射至下一時間之畫面上。投射的區域和下一張種子區域相重疊部分重新定義為下一張畫面的初步種子區域。由此亦可檢知遮蓋與顯露的區域。初步種子區域內若含遮蓋區域與移動補償誤差較大之像素，則均將其從初步種子區域內剔除。另外，投射結果若發生一個區域分散成數個區域時，則只取最大且其面積在某水準以上者。接下來重新分割的程序和第二階段相同。在前後畫面的物件中，由前張之物件投射至重分割

的後畫面，若某重分割區域的被投射面積超出總面積的 0.5，則該區域被視為前物件之繼承。如圖二所示，物件 R 顯示區域隨時間而可能的變化。上述程序可達物件追蹤的目的。

圖三以 176x120 大小的 Table Tennis 視訊為例，顯示本演算法所獲得之最前 35 畫面中桌球的移動。圖四顯示人身與球拍的位置。相關研究成果可參[1], [2]。

B. 像素域比例式編碼

視訊之多層編碼(multiplayer coding)可用於非同質網路之多播傳輸，以依據不同用戶之選擇來提供不同之視訊品質，或在網路傳輸功能隨時變異時，動態調整傳輸速率。而像素域之比例式視訊編碼為達成多層編碼之一主要途徑。一般而言，像素域之比例式編碼可分空域、時域、及訊雜比域三型，我們考慮空域型。我們探討二種不同的編碼器架構，一是次頻帶架構，二是 H.263+ 架構。

在次頻帶編碼器架構下，我們考慮用小波濾波器將視訊在空域做橫向及縱向的高低頻濾波，然後經下取樣而獲得四個大小為原來畫面四分之一的視訊信號。再對此四個信號個別以 H.263+ 方式進行編碼。如此，依所接收到的次頻帶數目的多寡，可得幾個不同解析度的視訊信號。我們所使用的小波濾波器是 Daubechies (9,7) 濾波器。目前我們所獲得的編碼結果並不理想，推測其原因主要是由於三個含高頻成份的次頻帶信號須做增益及偏化調整 (scaling and bias adjustment)，而我們在這方面之參數選擇不盡理想所致。後續的研究可考慮先在此方面改進。

在 H.263+ 編碼器架構下，我們研究以率差方式(rate-distortion)對不同解析度之畫面一併做位元分配。在計算機模擬中，我們考慮二層編碼，其中基本層為 QCIF 大小，增強層為 CIF 大小。我們採用英屬哥倫比亞大學的 tmn-3.2.0 編碼軟體加以修改，使得能進行二層聯合率差編碼。(原軟體在增強層僅能使用一個固定的量子化步階。)為求簡便，我們依基本層及增強層在傳輸中所使用到的比例來訂定一個平均的誤差值，而編碼器則在符合位元率的限制下，選取基本層及增強層的量子化步階以達到最小的平均誤差值。由於兩層編碼的率差函數不為獨立，故我們提出一種估計增強層率差函數的方法。

我們考慮每層每畫面用一個量子化步階的情況,也考慮每 GOB (group-of-blocks) 可變其量子化步階的情況。這兩種情況下,量子化步階的組合數有極大的不同;前者為 32^2 ,而後者為 32^{27} 。我們使用 Lagrange multiplier 方法來簡化運算。此外,由於前後相隨的畫面之率差關係具相關性,故其最佳之量子化步階亦可能相當接近;我們據此可降低所搜尋之步階數目。

模擬結果顯示,以 GOB 為單位來決定量子化步階,雖然彈性大得多,其效能未必比以畫面為單位之情況為佳。這與我們從前另一個研究的結果相同[3]。在本研究中,推測可能是因為以 GOB 為單位之方法,在率差函數的估計上較不準確之故。圖五是一個典型的編碼效能表現。

本研究之其他結果可參[4]。

C. 網路視訊通訊終端系統實作與研究

本研究主要係使用個人電腦及其上裝置之數位訊號處理器插卡來實作一個可在網際網路上進行點對點視訊通訊的視訊壓縮傳輸終端系統。系統結構如圖六所示。

傳輸端的個人電腦是 server,而接收端的則為 client。接收端不須數位訊號處理器,由個人電腦逕行做視聲訊的解碼與播放。傳輸端的個人電腦,為配合數位訊號處理器之需求,係採用 Windows NT 作業系統。其視訊輸入經個人電腦轉交數位訊號處理器插卡做編碼。目前我們使用的數位訊號處理器插卡為 Blue Wave Systems 的 PCI/C6600,其上裝置 Texas Instruments 的 TMS320C6201 定點數位訊號處理器二顆,工作速率為 200 MHz。但我們的視訊編碼器僅用其中一顆。編碼方法為簡化之 H.263。其基本程式可參[5],但我們做了一些除誤及改進。聲訊部分,未做壓縮,僅由個人電腦將之與壓縮後的視訊組成封包,交由網路卡透過 UDP 規約傳出。聲訊以外之系統功能,大體上可參[6]。

就視訊部分而言,其視訊輸入與編碼速度目前均約為每秒 20 張 sub-QCIF 畫面 (128x96)。由於二者是在不同的處理器(視訊擷取插卡及數位訊號處理器插卡)上分別進行,個人電腦以 multi-thread 的方式來控制,故整體處理速度亦為每秒約 20 張畫面。我們持續在探討改進視訊編碼與傳輸功能的途徑,包括更大的畫面(QCIF, CIF, 或更大)、更快的編碼速度(考慮用 DMA 及 TMS320C6201 發展環境所支援的其他

途徑)、及更適用的傳輸規約(如 RTP)等等。在視訊輸入方面,亦考慮改用更新的擷取卡與軟體。

在聲訊方面,我們發現,因為所用之電腦軟體的功能之故,若我們將視聲訊數據合併傳輸,其中聲訊所用的緩衝器大小就需要依電腦處理速度來調整,否則會影響視聲訊品質。這是我們當初沒有預料到的結果。因此,未來可以考慮將視訊與聲訊分開傳送,在接收端增加一個做二者同步的功能。

關於視訊傳輸,除了嘗試真時編碼與傳輸的實作外,我們過去曾研究傳輸誤差處理的方法。例如:我們曾提出一個稱為位移向量配對的方法來使得接收端可做位移向量傳輸錯誤的檢測[7],但其錯誤修正的方法尚未斟完善。因此,我們又企求改進其錯誤修正的方法,結果見圖七[8]。我們擬在此方面有繼續之探討。

四、計畫成果自評

研究內容與原計畫相符程度:85%。

達成預期目標情況:技術上之創新、實驗系統之建立、人才培育。

成果之學術與應用價值等:若干成果已投稿於國際或國內學術會議。有的成果擬在進一步研究後投稿於期刊,或亦申請專利。

綜合評估:本計畫獲得了一些或具學術價值、或具應用價值的成果,並建立了一個實驗系統,亦達成人才培育之效。成果尚佳。

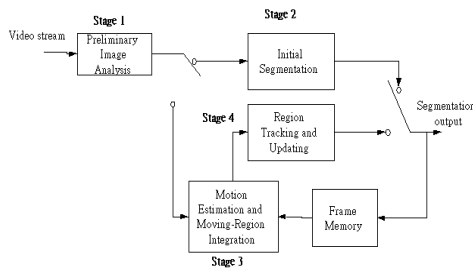
五、參考文獻

- [1] Y.-H. Jan and D. W. Lin, "A method for video segmentation based on object tracking," submitted to *Int. Symp. Commun.*, June 2001.
- [2] Y.-H. Jan and D. W. Lin, "Image sequence segmentation via heuristic texture analysis and region tracking," submitted to *Visual Commun. Image Processing 2002*, June 2001.
- [3] C.-W. Hung and D. W. Lin, "Towards jointly optimal rate allocation for multiple videos with possibly different frame rates," in *Proc. IEEE Int. Symp. Circuits Syst.*, vol. II, pp. 13-16, May 2000.
- [4] D.-H. Chen, "Bit allocation for scalable video coding," M.S. thesis, Dept. of

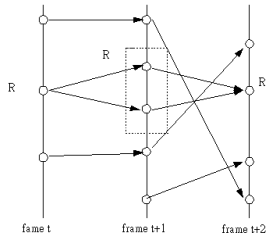
Electronics Engineering, National Chiao Tung University, July 2001.

- [5] S.-W. Chen and D. W. Lin, "H.263 video codec implementation on a TMS320C62xx digital signal processor," in *Proc. Workshop on Consumer Electronics*, pp. 1-4, Taipei, Oct. 1999.
- [6] J.-R. Wu and D. W. Lin, "DSP-based realtime video encoding and transportation for videoconferencing system," in *Proc. Workshop on Consumer Electronics*, pp. 181-184, Taipei, Oct. 2000.
- [7] Y.-L. Chen and D. W. Lin, "Error control for H.263 video transmission over wireless channels," in *Proc. IEEE Int. Symp. Circuits Syst.*, paper no. MPA13-7, June 1998.
- [8] T.-S. Tu and D. W. Lin, "A study on error-resilience techniques for wireless transmission of H.263 video," in *Proc. Workshop on Consumer Electronics*, pp. 135-140, Taipei, Oct. 2000.

六、圖表



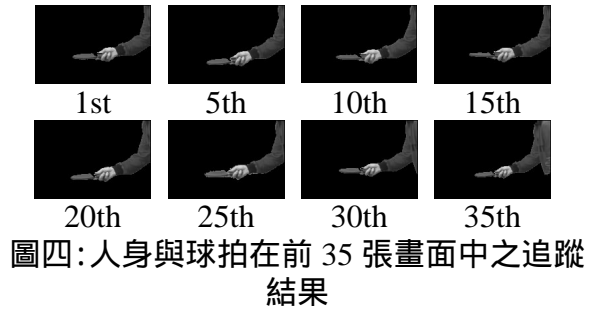
圖一：視訊分割方法



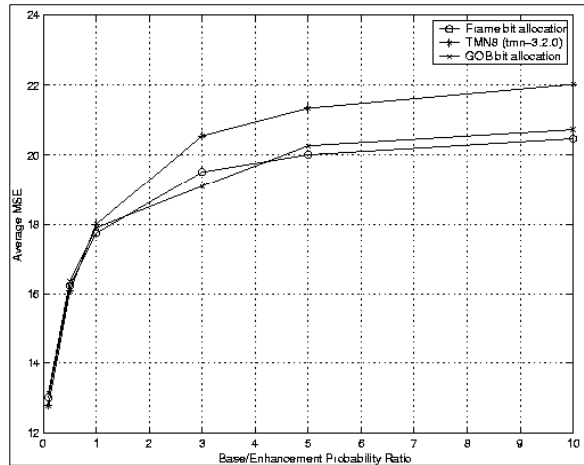
圖二：追蹤路徑示意



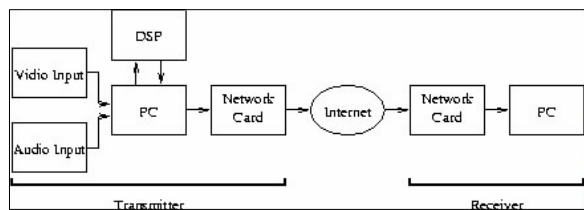
圖三：Table Tennis 前 35 張畫面中桌球的位置



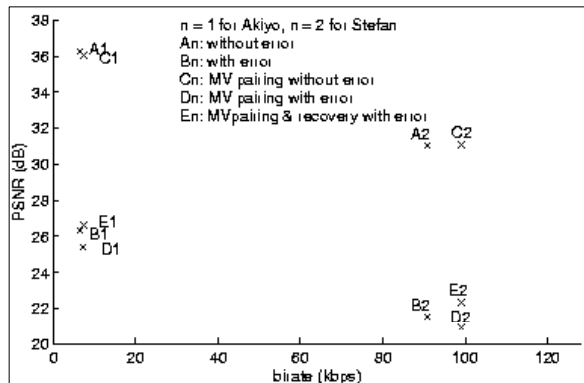
圖四：人身與球拍在前 35 張畫面中之追蹤結果



圖五：典型之空域比例式編碼之不同位元分配方法之編碼效能



圖六：網路視訊通訊系統架構



圖七：移動向量配對及其錯誤修正之效能