

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—子計畫三： 比例式視訊編碼技術及視訊通訊終端機技術之研究 (2/3) Research in Scalable Video Coding Techniques and Visual Communication Terminal Technologies (2/3)

計畫編號：NSC 90-2213-E-009-140

執行期限：90 年 8 月 1 日至 91 年 7 月 31 日

主持人：林大衛 交通大學電子工程學系 教授

計畫參與人員：詹益鎬、周依祥、洪崑健、陳彥福、郭沛昫

交通大學電子工程學系 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，主要為物件域之比例式編碼；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，我們研究視訊內容分割法，迄今已提出數項分割技術。模擬結果顯示：萃取得之物件符合人類常態知覺。現正繼續研究改進視訊分割技術。在網際網路視訊終端機方面，我們過去已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，其中的視訊編碼採用 H.263 標準。目前在持續改進此系統之功能，並進行 MPEG-4 simple profile 及 fine-grain scalable 即時編解碼之實作。

關鍵詞：物件域比例式視訊編碼、視訊分割、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, primarily object-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 research into methods for video content segmentation. We have proposed several segmentation techniques. Simulation results show that the extracted objects are in accord with common sense of human perception. Presently, we are continuing the research and improvement of video segmentation techniques. Concerning internet visual communication terminals, we have previously developed a point-to-point

video coding and transmission system employing personal computers and digital signal processors as the platform. The video codec in the system employs the H.263 standard. Presently, we are improving the functionalities of the system, and we are working on the real-time implementation of MPEG-4 simple profile and fine grain scalable codecs.

Keywords: Object-Oriented Scalable Video Coding, Video Segmentation, 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，二是即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，主要是物件域之比例式編碼；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

MPEG-4 的一項重要創新，是採用物件導向型式(object-oriented)的視訊編碼，以求能對畫面做更有彈性的編輯等處理。因此，一個很重要的課題就是視訊區域的分割，而此分割結果最好在人類常態視覺的角度看來是有意義的。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。

視訊區域的分割方法，約可分為二大途徑。一是對畫面做直覺的運動與紋理(texture)分析；二是使用機率模型，即馬可夫隨機場及貝氏估測等數學理論，來設計分割方法。兩途徑之間，難斷優劣。本計畫兼考慮此二途徑。以下分別討論之。

1. 以直覺分析為基礎的視訊分割

此類分割方法，通常包括四個基本功能方塊，即紋理分析、運動分析(運動估計)、初始分割、及區域追蹤。我們提出了幾個分割方法，其基本架構亦如是[1]-[4]。圖一呈現我們發表於[2]的架構，我們將較仔細的介紹此架構。

初始分割(initial segmentation)是針對第一張畫面而做，其方法約可分為二基本型，一是根據紋理分析，二是根據運動分析。圖一所示是採用第一型。(在[3]中我們考慮第二型，而在[4]中我們則考慮此二型的結合。)由於自然物件常含有一塊或或多塊平滑的區域，因此我們針對每張畫面，先找到若干平滑的區域，以作為區域分割與追蹤的基礎，就是圖中所示的 seed-area identification。如果是第一張畫面，我們接著就做習見的 region growing 和 region merging，完成初始分割。

如果是第二張或後續的畫面，我們就對每一個分割出的區域算出其 dense motion field，再將相鄰而具相似運動的區域合併。其中在 dense motion field 的計算方面，為了能降低其運算量，我們是先找到區域邊界上每一像素的位移向量，再從這些向量中來尋找該區域中各像素的最佳位移向量。這就是圖一中的 motion-based segmentation。

接著，我們將各視訊區域依其位移向量投射至下一畫面上。投射的區域和下一張中的 seed areas 相重疊部分重新定義為下一張畫面的 initial seed areas。Initial seed area 必須大於某下限，且其中像素之運動補償預測誤差(motion-compensated prediction error)必須小於某上限，否則不算做可接受的 initial seed area。接下來就由被接受的 initial seed areas 做 region growing 及 region merging，完成下一畫面的分割。最後再將前後畫面中能對應的區域連起來，達成物件的追蹤。

圖二以 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。圖三顯示在若干畫面中所分割出並追蹤而得之「人身與球拍」區域。

視訊分割有兩大困難，一是如何能在一連串的畫面中保持精確而前後一致的物件邊界認定，二是如何使計算量不致太大。以上方法對此兩議題均有其設計，但仍未盡理想。針對第一個議題，我們進行了一些研究。文獻[3]是一些稍早的成果。在研究過程中，我們亦發現在視訊分割文獻中未有用 Salesman 視訊為例者。經試對之進行分割，發現難度較其他若干習用視訊為高。經研究，發現若根據運動分析做初始分割，並針對移動區域邊緣明暗對比較強的部分作進一步的檢視，以確認其運動究竟與該邊緣那一側的物件較符合，則可將 Salesman 視訊做較好的分割。其對 Salesman 及 Akiyo 兩視訊之部分分割結果示於圖四及五。

最後，我們考慮結合紋理及運動分析來做初始分割[4]。其整個視訊分割的演算法可視為三層次的結合，即像素層(進行紋理與運動分析)、前景層(將視訊中移動的區域分割出來)、及重疊層(將前景中不同的物件分割出來)。可達到不錯的結果。圖六呈現其對 Salesman 之部分分割結果。

2. 以機率模型為基礎的視訊分割

此類分割方法，通常使用疊代(iterative)計算，因此計算量很高。

考慮前後二張畫面。若是一些相鄰的像素具相似的運動，其附近地區又具相似的紋理，則它們應有頗高的機率是屬於同一物件。以機率模型為基礎的分割法就是將這一直覺的觀察寫成一個機率模型。其中幾個主要參數是 line field (兩個相鄰視訊區域之邊界)，motion field (像素或視訊區域之運動狀況)，及 occlusion field (因物體移動而使背景被遮蓋之地區) [5]。針對一個畫面，若是我們對其 line field, motion field, 及 occlusion field 已有一個初步的估計，則可以將此估計套入機率模型，並試修改之以提高其機率。使用疊代計算的地方，就在這些 fields 的反覆修改與機率的反覆計算。

以上機率模型的訂定，雖然有理論的基礎，但在實際應用上，還是要靠經驗與直覺。基本上，此一機率模型使用到一個觀念，就是一個物件內(line field 所圍起來的

區域)的附近地區,其紋理及運動應具相當的相似性。一個簡單而合理的假設,就是令其機率隨紋理及運動的均方誤差大小而呈高斯分佈。這就是所謂馬可夫隨機場模型。

之所以使用疊代方式進行計算,是因以上馬可夫隨機場機率的最大化,並無 closed-form 解。為使疊代次數不要太過巨大,初步的 line field, motion field, 及 occlusion field 之估計,最好已相當合理(就人類主觀視覺而言)。

為求方便,我們使用 JSEG [6]做單一畫面之初步分割,並用區塊比對的方法計算初步的像素運動場,然後以疊代方式進行分割之修正。由於 JSEG 達成之初步分割未盡理想,之後的疊代計算之設計也頗費工夫。圖七顯示對 Table Tennis 視訊作分割的部分結果。可見經我們的疊代計算後,可使初始分割不宜之處有所改善,惟目前之成效尚不如前節之直覺分析設計。較完整之討論可參[7]。

B. 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視訊通訊的視訊壓縮與傳輸終端系統。本項研究亦分兩子題,一是既有 H.263 編解碼與傳輸系統的改進,二是 MPEG-4 編解碼與傳輸系統的實作。

1. H.263 編解碼與傳輸系統之改進

本系統結構如圖八所示。傳輸端的個人電腦是 server,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插卡之需求,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 H.263,但予以簡化以利即時實現。聲訊以外之系統功能,大體上可參[8]。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。

前段所述的系統架構,是本年度以前就已完成的,但我們持續在改進其功能,例

如:配合新個人電腦硬體與新的攝影機之電腦介面所做的系統更新。過去的攝影機可以支援 sub-QCIF 畫面(128x96),但新的攝影機只支持 QCIF (176x144)以上的畫面。這雖只是簡單的參數更改,但卻花了我們許多時間來解決個人電腦與數位訊號處理器插板間的緩衝器設定問題。

目前新系統之攝影機輸入速度約為每秒 19 張 QCIF 畫面,編碼速度則約為每秒 11 張 QCIF 畫面。與過去 sub-QCIF 編碼速度(約每秒 20 張)相較,約為 1/2,符合畫面大小的比例。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,故整體編碼與傳輸速度約為每秒 11 張畫面。

我們持續在探討改進系統的功能,如:研究增進編碼速度的方法、試圖能編解更大畫面的視訊(CIF 或更大)、試圖加入一項當初刪除的 H.263 rate control 功能、及試圖增強聲視訊的同步功能。預期在下年度有更進一步的成果。

2. MPEG-4 編解碼與傳輸系統之實作

此為本年度開始的研究項目,預期下年度將有較完整的成果。此研究分成二部分:一是 simple profile codec 之實作,二是 fine-grain scalable (FGS) codec 之實作。後一功能使網路節點可以調整視訊傳輸速率以彈性因應網路擁塞。兩個 codec 之實作均採用外間既有之模擬編解碼軟體為藍本,其中 simple profile codec 將獨立運作,而 FGS codec 則將與前節之 H.263 codec 形成一整個 scalable codec。

兩個 codec 現均建構在如前節所述的系統架構下。軟體的實作正穩定進展中。

四、計畫成果自評

計畫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:若干成果已發表於國內與國際學術會議,其他若干則在撰稿投稿過程中(含期刊與會議)。

綜合評估:本計畫獲得一些具學術與應用價值的成果,並建立了一個實驗系統,亦達人才培育之效。自評為「佳」。

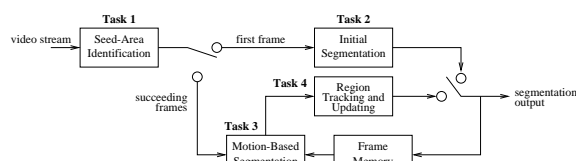
五、參考文獻

[1] Y.-H. Jan and D. W. Lin, "A method for video segmentation based on object

tracking,” in *Proc. Int. Symp. Commun.*, paper 10.4, Tainan, Taiwan, ROC, Nov. 2001.

- [2] Y.-H. Jan and D. W. Lin, “Image sequence segmentation via heuristic texture analysis and region tracking,” in *SPIE vol. 4671, Visual Commun. Image Processing*, pt. 2, pp. 543-551, Jan. 2002.
- [3] Y.-H. Jan and D. W. Lin, “Extraction of video objects by combined motion and edge analysis,” in *Proc. IEEE Int. Symp. Circuits Syst.*, pp. V-677—V-680, May 2002.
- [4] Y.-H. Jan and D. W. Lin, “A three-tier algorithm for video segmentation with extraction of overlaying objects,” in preparation.
- [5] A. M. Tekalp, *Digital Video Processing*. Prentice Hall, 1995, ch. 8.
- [6] <http://vision.ece.ucsb.edu/segmentation/JSEG>.
- [7] Y. Chou, “Video segmentation via iteratively enhanced spatial-temporal analysis,” M.S. thesis, Dept. Electronics Eng., National Chiao Tung University, June 2002.
- [8] 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.

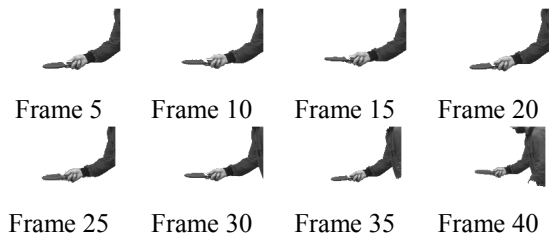
六、圖表



圖一：直覺分析視訊分割法之一



圖二：Table Tennis 視訊前 35 張畫面中桌球的位置(分割法一)



圖三：Table Tennis 視訊「人身與球拍」區域在前 40 張畫面中之分割與追蹤結果(分割法一)



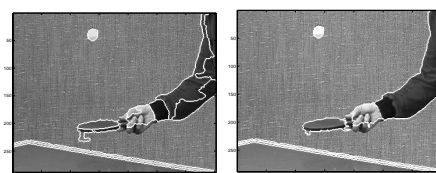
圖四：Salesman 視訊分割結果(分割法二)



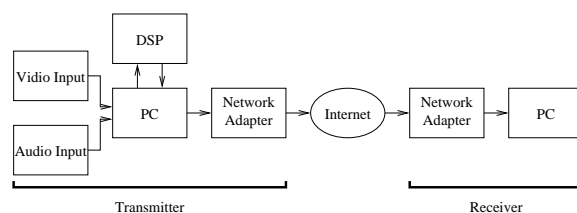
圖五：Akiyo 視訊分割結果(分割法二)



圖六：Salesman 視訊分割結果(分割法三)



圖七：機率模型視訊分割之部分結果(左：初始分割；右：疊代計算之後)



圖八：網路視訊通訊系統架構