

行政院國家科學委員會專題研究計畫 成果報告

總計畫(1)

計畫類別：整合型計畫

計畫編號：NSC92-2219-E-009-003-

執行期間：92年08月01日至93年07月31日

執行單位：國立交通大學電子工程學系

計畫主持人：杭學鳴

計畫參與人員：王俊能,唐之璇,王士豪,彭文孝,李志鴻,李俊毅,黃名彥,蔡家揚

報告類型：完整報告

處理方式：本計畫可公開查詢

中 華 民 國 93 年 12 月 8 日

行政院國家科學委員會補助專題研究計畫 成果報告
 期中進度報告

基於 MPEG 標準之多媒體通訊與串流整合平台及其應用(總計畫)(I)
MPEG Integrated Multimedia Platform and Applications (I)

計畫類別： 個別型計畫 整合型計畫

計畫編號：NSC 92-2219-E-009-003

執行期間：92年8月1日至93年7月31日

計畫主持人：杭學鳴 國立交通大學電子工程學系 教授
共同主持人：王聖智 國立交通大學電子工程學系 副教授
李素瑛 國立交通大學資訊工程學系 教授
蔡淳仁 國立交通大學資訊工程學系 助理教授
蔣迪豪 國立交通大學電子工程學系 副教授
林大衛 國立交通大學電子工程學系 教授

計畫參與人員：王俊能,唐之璇,王士豪,彭文孝,李志鴻,李俊毅,黃名彥,蔡家揚

成果報告類型(依經費核定清單規定繳交)： 精簡報告 完整報告

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

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢

涉及專利或其他智慧財產權， 一年 二年後可公開查詢

執行單位：國立交通大學電子工程學系

中華民國 92年 10月 5日

行政院國家科學委員會專題研究計畫成果報告

基於 MPEG 標準之多媒體通訊與串流整合平台及其應用(總計畫)(I)

MPEG Integrated Multimedia Platform and Applications (I)

計畫編號：NSC 92-2219-E-009-003

執行期限：92 年 8 月 1 日至 93 年 7 月 31 日

主持人：杭學鳴 國立交通大學電子工程學系教授

共同主持人：王聖智，李素瑛，蔡淳仁，蔣迪豪，林大衛 國立交通大學

計畫參與人員：王俊能，唐之璇，王士豪，彭文孝，李志鴻，李俊毅，黃名彥，
蔡家揚 國立交通大學

中文摘要

本群體計畫建構符合 MPEG 標準之網際網路即時(real-time)多媒體傳輸測試平台並在其上開發相關之技術與應用。此測試平台含媒體伺服器、使用端與模擬網路。本計畫以此平台為核心開發各項多媒體傳輸相關技術，如：可調層次式視頻編碼技術，串流多媒體伺服器與資料庫，智財管理與保護等等。目前本團隊將此測試平台提案到 MPEG 標準委員會，已被接受成為 MPEG-21 Part 12: Multimedia Test Bed for Resource Delivery 的委員會草案(Committee Draft)。

以多媒體傳輸測試平台為核心台，上述多媒體傳輸相關技術分別在六個子計畫中進行：(1) 可調層次式視訊編碼技術之前處理與後處理技術研究，(2) 視訊串流伺服器及視訊資料庫整合，(3) MPEG 多媒體傳輸網路機制、協定及模擬測試環境的分析設計，(4) 高等精細可調層次式視頻編碼技術之研究，(5) MPEG 智財管理與保護系統及強韌視訊解碼器之設計與模擬，以及(6) 多點視訊會議技術之研究。整體計畫時程擬為三年，本報告將略敘第一年之各項成果。

關鍵詞：MPEG-4、MPEG-7、MPEG-21、多點視訊會議、串流視訊、多媒體測試平台、精細可調層次式編碼

英文摘要

The goal of this project is to construct a real-time multimedia streaming test bed and develop related technologies and applications. This test bed can be used to verify multimedia coding algorithms and MPEG-21 Digital Item processing schemes. In this project, we provide such a test bed prototype system for the MPEG committee to experiment with various user scenarios. Because of the value of this test bed, it has been accepted as the Committee Draft of MPEG-21 Part 12: Multimedia Test Bed Resource Delivery. As the MPEG standard activity progresses, we will continue working on this test bed for another two years.

Based on this test bed, we develop various types of application technologies and systems, for example, fine granularity scalability codec, streaming video servers and database, intellectual property rights protection and management, and preprocessing and postprocessing algorithms for scalable video streaming. These technologies are studied in 6 sub-projects: (1) Preprocessing and Postprocessing Techniques of Scalable Video Streaming, (2) Video Streaming Server and Video Database Integration, (3) Design and Analysis of MPEG Multimedia Transport Mechanisms, Protocols, and Simulation Test Bed Design, (4) Advanced Fine Granularity Scalability, (5) MPEG IPMP System and Robust Video Decoder Design and Simulation, and (6) Research in Multipoint Videoconferencing Technologies. The entire duration of this project is 3 years. This report contains the research results of the first year.

Keywords: MPEG-4, MPEG-7, MPEG-21, Multi-point Videoconferencing, Video Streaming, Multimedia Test Bed, Fine Granularity Scalable Coding

目錄 Table of Contents

1. 背景與目的.....	4
2. 可調層次式視訊編碼技術之前處理與後處理技術研究 -- 王聖智教授.....	5
3. 視訊串流伺服器及視訊資料庫整合 --李素瑛教授.....	8
4. MPEG 多媒體傳輸網路機制、協定及模擬測試環境的分析設計 -- 蔡淳仁教授	10
5. 高等精細可調層次式視頻編碼技術之研究 -- 蔣迪豪教授.....	13
6. MPEG 智財管理與保護系統及強韌視訊解碼器之設計與模擬 -- 杭學鳴教授.....	15
7. 多點視訊會議技術之研究 -- 林大衛教授.....	18
8. 計畫整體成果與自評.....	22

1. 背景與目的

本整合型計畫希望達成下述共同目標：

1. 本計畫建構符合 MPEG 標準之網際網路即時(real-time)多媒體傳輸測試平台。此測試平台含使用端與模擬網路。在 MPEG-21 應用中，媒體伺服器與編解碼器均為使用端(之部分)。本計畫以此平台為核心開發各項多媒體傳輸相關技術，如：可調層次式視頻編碼技術，串流多媒體伺服器與資料庫，智財管理與保護等等。2002 年 12 月將此測試平台提案到 MPEG 標準委員會，成為 MPEG-21 Part 12: Multimedia Test Bed for Resource Delivery 的工作草案(Working Draft)。接著經過持續努力，這個工作草案於 2003 年 12 月晉升至大會草案(Committee Draft/Proposed Draft of Technical Report)。未來兩年仍需再接再厲，才能成為正式標準文件。
2. 本計畫在上述多媒體傳輸測試平台上整合各子計畫成果，建構應用實例展示，規劃中有即時串流視訊與多點視訊會議兩系統。利用本平台之即時網路特性，驗證所開發之技術在整合環境下之成效。並可依測試數據，改良演算法與系統設計，此功能是一般非即時模擬或單獨個別技術測試所無法達成的。
3. 由於此一多媒體之測試平台可望由 MPEG 標準委員會接納，成為標準參照軟體。因此可提供媒體製作者、媒體伺服器業者、通訊與多媒體服務系統業者、多媒體相關研究單位與學者等，一個實用的、可重複的、易嵌入的實驗與測試平台，作為網際網路應用技術最佳的研發工具。

隨著多媒體標準 MPEG 系列之制定，2.5G 及 3G 無線通訊標準之發表，與日新月異的數位電視，掌上型電腦與行動網路之廣受歡迎，更前瞻的多媒體應用日漸普及。多媒體技術的發達促使資訊的存取與資料的傳遞之相關應用廣泛地發展。多媒體應用在百家爭鳴情況下，卻造成彼此之間在資料傳遞與存取發生互動困難的問題。因此，MPEG-21 制定了多媒體通用存取的架構與法則，不同的消費者群(customers)可以透過談判來連接與互動。因此，在 MPEG-21 應用中基於可調層次式視頻編碼技術之隨選視訊與多點視訊會議之多媒體系統，特別受到產學官各界的青睞；因其能夠契合 2.5G 及 3G 視訊服務的需求與網路變異時之可調層次性。

可調層次式視頻編碼技術已經被 ISO/IEC 組織規劃為 MPEG-21 Part 13: Scalable Video Coding。對於視訊壓縮，*Call for Proposals on Scalable Video Coding Technology* 是這個領域目前最熱門的主題之一。2004 年 3 月 MPEG 第 68 次會議上，很多的公司組織帶著他們的技術參與這場盛大的競賽，以角逐成為 MPEG-21 視訊壓縮技術的標準之起始點。交通大學團隊提出兩項提案，在 14 個付費參與視覺品質評估的候選方案中，除了 Microsoft Research Asia 之外，交通大學是唯一提出兩案的單位。在未來幾年之內，本總計劃所提的整合平台與基於這個整合平台的設計概念所衍生的媒體製作與視訊服務之相關產業、產品、研發等將成為一股風潮。因此，本計劃中欲建構之基於 Scalable Video Coding 的視訊串流系統是今日工業界與學界的趨勢，其所研發原型系統也可帶動產品的開發並增加其產值。

此「基於 MPEG 標準之多媒體通訊與串流整合平台及其應用」整合計畫的分工情形大概如下：

- (一) 可調層次式視訊編碼技術之前處理與後處理技術研究 —— 王聖智教授
- (二) 視訊串流伺服器及視訊資料庫整合 —— 李素瑛教授

- (三) MPEG 多媒體傳輸網路機制、協定及模擬測試環境的分析設計 —— 蔡淳仁教授
- (四) 高等精細可調層次式視頻編碼技術之研究 —— 蔣迪豪教授
- (五) MPEG 智財管理與保護系統及強韌視訊解碼器之設計與模擬 —— 杭學鳴教授
- (六) 多點視訊會議技術之研究 —— 林大衛教授

2. 可調層次式視訊編碼技術之前處理與後處理技術研究 -- 王聖智教授

在這次的進度報告中，我們針對 SNR 可調層次式視訊編碼技術加以研究，藉由適應性調整 macro block 的參數，來最佳化 temporal prediction，以提升視訊的品質。在 FGS 的技術中，因 Robust FGS (H.C. Huang, C.N. Wang, and T. Chiang, “A robust fine granularity scalability using trellis-based predictive leak,” IEEE Trans. Circuits Syst. Video Technol., pp. 372-385, vol. 12, no. 6, Jun. 2002)提供了很有彈性的方法來調整 temporal prediction，因此本研究以 Robust FGS 為基礎，對其參數加以最佳化。實驗證實，本研究所提出的方法可以在不同給定的頻寬區段與不同測試條件下提升視訊的品質，特別是 slow motion 的視訊編碼。

(1) 背景

要將龐大的視訊資料以有限的頻寬傳送，必須仰賴高效率的壓縮技術，因此，在過去的幾十年中，視訊的壓縮技術越來越受到重視。傳統的視訊壓縮技術(如 MPEG1)著重於固定頻寬的視訊壓縮，然而，若要將視訊資料傳輸於網路，則必須進行可調層次式視訊編碼，因為網路的頻寬是動態變化的，在這種狀況下，當要對視訊壓縮時，必須考慮一段 bitrate 區間，而非單點，這類視訊技術普遍稱作可調層次式視訊編碼技術(Scalable Video Coding)。FGS 為 SVC 的解決方案之一，並已列入 MPEG4 標準中。FGS 將視訊資料依重要程度區分為代表核心資訊的 base layer，以及用於增加解析度的 enhancement layer，base layer 是以 DCT-based non-scalable single layer 編碼，而 enhancement layer 則再細分為不同的 bitplanes(如圖 2-1 所示)，當頻寬不足或軟硬體資源不足時，可以僅傳送資料量較少的 base layer 資訊而獲得影像品質較差的視訊資料，而當頻寬充裕或軟硬體資源充足時，則可以將 enhancement layer 中的 bitplane 資料陸續添加進來，以獲得較佳的影像品質。

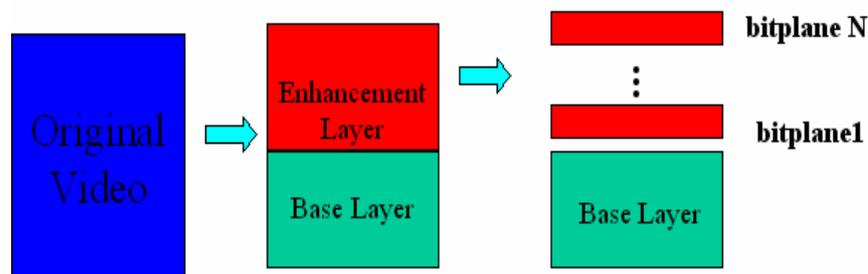


圖 2-1. Fine Granularity

(2) 架構

我們希望視訊編碼在一段 bitrate 區間中最佳化，在這個區間中的所有可能狀況都要

列入考慮。預測編碼 temporal prediction 在不同 bitrate 區間會呈現不同的效應，在 high bitrate 區間，使用越多 EL 的資料來做 temporal prediction 將會有越多的增益，因為大部分的資料都已經傳送到 decoder 端，誤差飄移(error drift)將會很小，相反的在 low bitrate 區間應使用較少 EL 的資料來做 temporal prediction，因接收端收到的資料量並不多。在本研究中，我們試著探討 motion compensation gain 與 error drift 之間的交互作用，並嘗試平衡 motion compensation gain 與 error drift，藉以將視訊編碼在一段區間內最佳化。

給定一 bitrate 區間 PB 及 N 個監測點(checkpoints) $S_1 \dots S_N, S_i < S_j$ if $i < j$ 。在不同的 FGS 方法中，RFGS 提出了一個很有彈性的架構來控制 temporal control，因此，我們的系統是以 RFGS(圖 2-2)為基礎來最佳化 temporal prediction。在 RFGS 中有 α 與 β 兩個參數，其中 β 決定要取幾層 enhancement layer 的 bitplane 來做 motion compensation ($0 \leq \beta \leq \text{maximum bitplane}$)， α 為小於 1 的參數($0 \leq \alpha \leq 1$)來控制 error drift，因此本研究的目標就是找出最佳的 α_{opt} 與 β_{opt} ，藉以達成在 PB 區間中最大的平均編碼效率，可表示成：

$$(\alpha_{opt}, \beta_{opt}) = \arg \max_{\alpha, \beta} \left(\frac{1}{N} \times \sum_{i \in S} (G_i + L_i) \right) \quad (1)$$

在本研究中，我們以 PSNR 作為客觀憑斷視訊影像品質的標準。

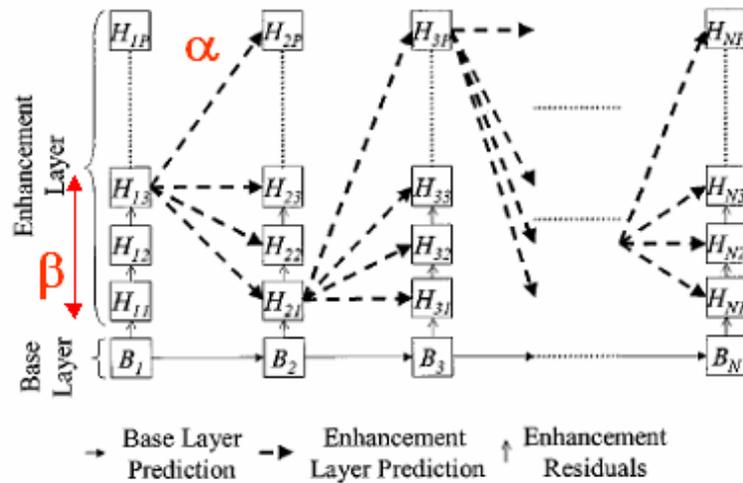


圖 2-2. RFGS 架構與參數

我們提出以 MB 為主的參數調整方式，也就是在同一個 frame 中的每個 MB 都有自己的 α 與 β 。假設一個 frame 包含了 K 個 MBs, MB(1)...MB(K)，且假設 Enhancement Layer 中 maximum bitplane 為 M， α 被 quantized 為 Q steps，則所有可能的 solution space 為 $(K)^{Q+M}$ ，為了在再這麼大 solution space 中找出最佳解，我們先做 initial guess，然後再以迭代(iterative)的方式搜尋最佳解。

我們使用在解碼端得到的平均 PSNR(定義為 APR)來估計 motion compensation gain G and error drift loss L 的淨效應。對於每個 MB(j)， $1 \leq j \leq K$ ，我們先決調整趨勢 $TR(j)$ ， $TR(j)$ 定義為：

$$TR(j) = \begin{cases} 0 & \text{if } \max(\beta_0(j)^+, \beta_0(j)^-) < \beta_0(j) \\ +1 & \text{else if } APR(\beta_0(j)^+) > APR(\beta_0(j)^-) \\ -1 & \text{otherwise,} \end{cases}$$

其中 $\beta_0(j)^+$ and $\beta_0(j)^-$ 代表對第 j 個 MB 的 β 上下移動一個 step 且保持其餘的 MBs 的 β 值不變的狀況下所得到的新 β 組合。接著，定義 Impulse gain likelihood function LG 為：

$$LG(j) = \max(\text{APR}(\beta_0(j)^+) - \text{APR}(\beta_0), \\ \text{APR}(\beta_0(j)^-) - \text{APR}(\beta_0)).$$

函數 LG 提供了一個好的指標，來決定哪個 MB 參數變化應該有較高的優先權。

我們提出了一個 three-step search 演算法，同時將 α 與 β 列入考慮，並且為了權衡時間複雜度與精確度，我們將 iteration 的數目限制在三次，演算法如下：

Phase 1: 為了增廣搜尋區域並避免落入 local minimum，第一階段以 frame-based 的方式決定整張 frame 的 α 與 β (共 $M \times Q$ 測試向量)，設為 $\bar{\alpha}_0$ 與 $\bar{\beta}_0$ 。

Phase 2: 以 $\bar{\alpha}_0$ 與 $\bar{\beta}_0$ 為 initial guess，對 frame 裡的 MB，以 $\alpha_0(j)$ 與 $\beta_0(j)$ 為中心，向上及向下位移一個單位 SP，行成定義區間組合 (a,b), $a \in (\alpha_0(j) - SP, \alpha_0(j), \alpha_0(j) + SP)$, $b \in (\beta_0(j) - SP, \beta_0(j), \beta_0(j) + SP)$ ，對這些組合加以排序，並找出區間中具有最大 APR 的參數組合，來決定 TR and LG 接著以 2.2 節的方法找出 LG', TR', 及 OD, 藉以得到第二 iteration 的最佳參數組合 $\bar{\alpha}_0'$ 與 $\bar{\beta}_0'$ 。

Phase 3: 以 $\bar{\alpha}_0'$ 與 $\bar{\beta}_0'$ 為 initial guess，重複 phase2 的動作，並以產生 peak 的 $\bar{\alpha}_0''$ 與 $\bar{\beta}_0''$ 為 suggested optimal α 與 β 之組合 $\bar{\alpha}_{opt}$ 與 $\bar{\beta}_{opt}$ 。

(3) 實驗結果

為了驗證提出方法的可行性，我們把提出的方法應用在常用的測試視訊，並紀錄其結果。測試條件如下：

- A. $\bar{S} = \{256, 284, 512, 768, 1024, 1536, 2048, 2560\}$ kbps.
- B. GOV = 15 (one I frame followed by 14 P frames).
- C. Video format= CIF, 30 frames per second.
- D. Base layer= 256 kbps, no rate control

測試視訊包含三個 video sequences, News, Foreman, and Stefan，分別代表 slow motion, moderate motion, and fast motion。為了將提出的方法(以 M2 表示)與原來的 RFGS 做比較，我們以實驗的方式得到當 PB=(256k,2560k)時，dominated β 為 (2,3,4)，然後，對這三個 β 找出最佳的 α 配對，分別為 F1(0.9,2), F2(0.8,3), 與 F3(0.8,4)。圖 2-3(a)顯示對 "News" 的實驗結果，與固定參數的 RFGS 比較(F1,F2,F3)，平均 PSNR 增益約為 2dB，證明了我們所提出的方法對於 slow motion sequence 的編碼有相當好的效果。圖 2-3(b)顯示 \bar{S} 中各個不同的監測點的平均 GOP(15 frames) PSNR，對於固定參數的 RFGS 而言，若採用比較多的 temporal prediction，將導致 high rate performance 較好，但 low rate drift error 較大，相反的比較少的 temporal prediction 會導致 low rate performance 較好，但 high rate efficiency 不足。然而，由圖 2-3 (b)可看出，我們所提出的適應性調整方法幾乎在所有監測點都有不錯的表現。

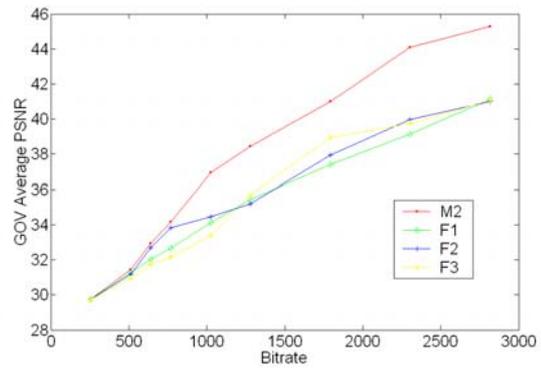
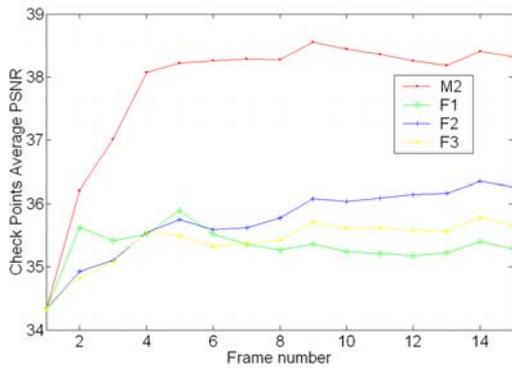


圖 2-3 (a). News sequence (frame v.s. PSNR) 圖 2-3 (b). News sequence (rate v.s. PSNR)

3. 視訊串流伺服器及視訊資料庫整合 --李素瑛教授

本計畫主要目的在於研究與實踐視訊資料庫以及視訊伺服器。在視訊資料庫方面，除了視訊資料內容之分析與高階與語意特徵擷取之外，並結合音訊資料之特徵加上影片內容字幕之偵測及擷取，利用分析所得之語意資訊加以解譯及推測，以偵測影片資料內容之事件，藉此達成高階特徵之擷取並以此為基礎產生既簡短又具語意意義之視訊資料摘要提供視訊資料預覽之用，除此之外在產生影片內容摘要之過程中，同時產生符合 MPEG-7 之低階至高階特徵之描述 (descriptor) 以及描述結構 (description scheme)，以利資料庫索引之建立並藉此建立互動式影片瀏覽系統。

在視訊伺服器方面，我們希望設計支援單點傳送與多點傳送的視訊串流伺服器。在單點傳送視訊串流伺服器方面，主要有兩項任務，一項是傳輸速率控制(rate control)，另一項則是錯誤控制(error control)。在傳輸速率控制中，不僅要根據網路狀況，更要根據視訊資料的重要性來做網路頻寬的最佳使用；在錯誤控制中，傳統的重送機制必須花費較多的時間才能達成錯誤恢復，因此我們希望透過傳輸錯誤更正(forward error correction) 機制，並且搭配不平等的錯誤保護機制(Unequal Error Protection)以降低錯誤恢復所需的時間並且減低頻寬的浪費。

在多點傳送視訊串流伺服器方面，由於多點傳送有效率地使用頻寬，因此我們的重心在於錯誤控制。在多點傳送的錯誤控制中，當封包遺失發生時，有兩個主要的問題需要解決，一個是區域性地錯誤恢復(Local Error Recovery)，一個是 NACK 爆炸(NACK Implosion)的議題，前者是希望能盡量縮短錯誤恢復所需要的時間，後者則避免造成區域性網路壅塞的問題。因此如何達到區域性地錯誤恢復以及避免 NACK 爆炸所造成的區域性網路壅塞便是我們的目標。

(1) 緣由與目的

過去以文字為基礎的資料庫系統已發展了一段時間並且已經相當成熟，但相反地以多媒體內容為索引及抽取依據的資料庫管理系統發展卻相當落後。最近將 meta-data 的觀念延伸至其他應用已經引起廣泛的探討，其應用可包含形形色色且較為複雜的資料型態，如壓縮的影像、audio、video 等等。因此對於大量的音樂資料、新聞、紀錄片、電影以及影像提供充分適當的描述對於不同領域及不同應用的使用者而言相當重要。在

應用上如，數位圖書館、隨選視訊以及網路多媒體傳播等等。

因此 MPEG-7 試著定義多媒體特徵的描述方法使得以內容為基礎的資料索引及資料搜尋更為容易。此外 MPEG-7 包含更多的資料型態並且將內容的描述 (descriptor) 和描述法 (description scheme) 標準化藉以達成能夠描述更多種類的多媒體資料。然而特徵產生之方式並非 MPEG-7 欲標準化的部份，除此之外由於多媒體資料型態及內容的多樣性使得更有效、更具高階意義的特徵亟待產生。另一方面為了提供系統間之相互作用 (interoperability)，特徵抽取以及搜尋引擎兩大議題皆非 MPEG-7 欲標準化之部分。

本計畫主要目的為視訊資料內容之分析與擷取，並利用分析所得之資訊加以推理 (inference)，以偵測視訊資料內容之事件，藉此達成高階特徵之擷取並以此為基礎產生既簡短又具語意意義之視訊資料摘要 (summary)，除此之外在產生視訊內容摘要之過程中，同時產生低階至高階特徵之描述 (descriptor) 以及描述結構 (description scheme) 以符合 MPEG-7 之標準。

(2) 系統設計與實驗結果

1. 傳輸速率控制模組 (Rate Control Module)

在傳輸速率控制模組中，多媒體伺服器便根據多媒體資料的重要性來做相對應的調整，如此不僅可以充分使用現有的網路頻寬，更可以好好利用接收端有限且受限制的緩衝區來達到多媒體品質的最佳化。此模組之輸入包含多媒體資料位元流以及估量的網路頻寬，輸出為經過調整的多媒體資料位元流，核心部分則包含資料傳輸率配置演算法 (Bit-rate Allocation Algorithm)、分層丟棄策略 (Layer Dropping Policies)、虛擬的 FS 緩衝區模型 (Virtual Forward-shifting Buffer Model) 以及洞悉多媒體資料內容的傳輸速率控制器 (Content-aware Rate Controller)

2. 傳輸錯誤更正模組 (Forward Error Correction Module)

錯誤更正模組將輸入的封包依其資料的重要性來決定是否要輸出重複的封包，例如，輸入的封包中，倘若其屬於基本層的資料封包，則其會被複製一份並且輸出；倘若其屬於加強層的資料封包，則其並不會有被複製的動作，而僅僅只是被輸出，因此，由於我們只對基本層做封包保護的動作，所以可以降低頻寬的花費，再者，上述的做法屬於傳輸層次，所以也不會造成編碼端以及解碼端的延遲，最後，即使網路狀況不佳，導致封包遺失率提高，但接收端依然可以根據收到資料的多寡，來做解碼的動作。

3. 網路模組 (Network Module)

在網路模組中，我們使用的是 RTP (Real-Time Transport Protocol)/RTCP (RTP control protocol)，RTP 被用來傳輸多媒體資料，RTCP 則被用來監視資料傳輸的狀況，要注意的是，RTP/RTCP 本身並不提供 QoS 的能力，也就是它不保證資料能正確無誤的被接收，也不保證資料能依序的到達，但是它提供 QoS 相關的資訊，在 RTP 的標頭欄位，其中的序列號碼 (Sequence Number) 可以用來重組接收封包的順序，也可以用來做封包遺失的偵測；時間戳記可以被用來做多媒體資料的同步。另外，RTCP 則會週期性地傳送控制封包，讓通訊的雙方可以知道彼此接收資料的品質狀況。

4. 視覺資料特徵擷取

我們設計了一個以移動物體為基礎之高階視訊內容描述方法 (object-based T2D-histogram descriptor)，此描述方法不僅精簡且在協助視訊片段搜尋上更是有效率，為衡量此描述方法之效能，我們與 MPEG-7 之兩個 motion activity descriptors, Run-Length Descriptor (RLD) 以及 Shot Activity Histogram (SAH) 比較，如圖 3-1 所示：

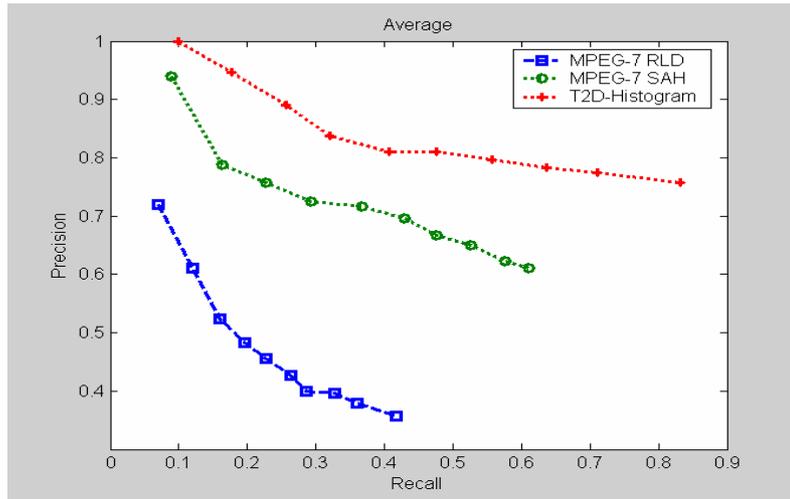


圖 3-1、視訊擷取效能比較

由實驗結果可知，在視訊擷取的準確性方面，我們所提出的 T2D- histogram 比 RLD 好約為 40%，另外亦比 SAH 好將近 20%。

4. MPEG 多媒體傳輸網路機制、協定及模擬測試環境的分析設計 —— 蔡淳仁教授

目前有許多國際組織都在致力於製定公開標準技術的多媒體傳輸服務。除了 3GPP 和 ISMA 外，MPEG 也正在製訂一個通用的多媒體傳輸架構 MPEG-21。MPEG-21 的標準和 3GPP 及 ISMA 所製訂的標準主要差異在於後二者的標準是針對特定的網路架構和客戶端功能而設計，因而在實作上的技術大抵已有定案。而 MPEG-21 則採取較高層、較抽象的架構設計。如何將 MPEG-21 的概念實現則仍是一個十分有潛力的研究題目。另外 MPEG-21 試圖涵蓋的技術範圍也比較大。本計畫的主要目的在研究及設計在 MPEG-21 的架構下適用於 MPEG 多媒體數位內容的封包交換式網路傳輸機制、協定、及模擬測試環境。

(1) 緣由與目的

一個完整的分散式數位多媒體系統含蓋的範圍極廣，包括數位內容的製作、數位資料庫的建立、使用者收費機制、智財權保護機制、媒體傳輸伺服器、應用服務介面，和媒體接收播放器的設計等等。為了能有一個統一的國際標準能達到建構互通的分散式多媒體系統的目的，MPEG 國際標準組織在西元 2001 年開始製訂一個新的國際標準：MPEG-21。

由於整個系統的重點在於能橫跨不同的網路架構和在不同的客戶端設備上（PC、手機、PDA 等等）提供一致而且最高品質的多媒體傳輸播放服務，因此傳輸系統的設計必須能動態的根據不同的平台調整。簡言之，一個數位多媒體傳輸系統的架構必需包含流量控制和容錯機制。另外，依據客戶端的能力來調整媒體資料流品質的能力也是十分重要的。本計畫的主要目的在研究以 MPEG 的多媒體 codecs 及 IETF 的通訊傳輸協定為核心的系統在不同的網路之下的流量控制和容錯機制。另外，本計畫也會發展一個完整的多媒體 IP 網路測試模擬環境以提供多媒體應用系統進行效能分析。這樣的網路模

擬器，對於多媒體傳輸系統的設計，具有極高的價值。特別是多媒體通訊協訂，必須能有相當的強韌性，以對付一個不可靠的通訊管道所有可能產生的資料誤失。在發展系統的過程中，一個可控制的完整的通訊網路模擬器，將可幫助鎖定需要加強的部份。另外，也可以有利於改進現有的通訊協訂，以及幫助分析多媒體網路一些參數的調整對整個傳輸系統設計會產生的影響。

(2) 結果與討論

本計畫主要的重點在於制定 MPEG-21 的多媒體傳輸共通測試平台。目前此平台標準已經成為 Committee Draft (CD)，在短期內就會成為國際標準。在這個整合計畫下，總計畫團隊為 MPEG 所設計的開放原始碼包含了完整的可調式媒體伺服器、網路模擬器、及媒體播放器。詳細的架構請參考 ISO/IEC JTC 1/SC 29, Information Technology – Multimedia Framework (MPEG-21) – Part 12: Test Bed for MPEG-21 Resource Delivery, ISO/IEC TR 21000-12: 2004(E), 2004。

另外，在流量控制和容錯機制的設計方面，為配合 MPEG-21 Scalable Video Codec 的標準制定及 MPEG-21 Digital Item Adaptation 的理念，我們進行了碼率失真最佳化 (rate-distortion optimized) 可調式串流傳輸機制的設計。目前在這方面較知名的是由 P. A. Chou 等人發展的系統 (P. A. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," IEEE Transactions on Multimedia, February 2001)。不過這套方法目前發表的成果以理論分析為主，在實作上有很多細節並沒有提出解決方案，而且在頻寬變化大的網路環境下，串流傳輸最難達到的平滑播放要求也沒有考量。

在可調式位元串流傳輸中，影像資料可以分成好幾次傳送，每次的傳送都可以幫助解碼端得到更接近於原影像資料的重建訊號，因此可調式位元串流的調適 (scalable bitstream adaptation) 設計必須考慮到如下幾點：必須支援多樣化的更新運作(update operations)以產生有效可解碼的串流、將資料刪除時不能違反解碼相關性(decoding dependencies)的原則、允許在各個次元(dimensions)的可調性、對於媒體的特性 (如：碼率、失真率、frame rate、frame size...等)必須提供所有可能的可調適性、針對不同的調適單元(adaptation units)可能必須設計不同的調適決策、對於網路服務品質(quality of service, QoS) 設計所有可能的調適方法。媒體資源的傳遞和調適在可調適的地點 (location of adaptation) 我們可以分成：傳送端驅動調適(sender-driven adaptation)、接收端驅動調適(receiver-driven adaptation)、網路驅動調適(network-driven adaptation)等三個不同的類別來考量。而對於可調式位元串流系統，可以把視訊資料分成一個基礎層和一個或多個加強層，解碼器可以選擇只解碼基礎層以獲得較低品質的重建訊號，或再加上一些加強層以獲得較高品質的重建訊號。

本計畫針對以下傳輸模組進行設計以達到最佳的全維度可調式媒體動態傳輸的效能：

1. 媒體封包相依性控制：媒體封包相依控制 (packet dependency control) 的設計目標是針對提供較高的錯誤抵抗能力 (higher error-resilience) 和消除影像封包的重傳 (re-transmission) 需求。典型的多媒體串流在影像封包之間具有強烈的相依關係，如果其中一個影像封包在傳送過程中丟失，則與這個封包有相依關係且跟隨在後的 frames 在解碼時將可能會受到影響。網路調適性的媒體封包相依控制模組可以用來改善可調式多媒體串流的錯誤抵抗能力和減少延遲 (latency)，在此，可以運用一個樹狀的模型來記錄通道的封包丟失率 (channel loss rate) 和錯誤傳遞 (error propagation) 以達成有效的控制機制。

2. 碼率最佳化傳輸控制：這部份的設計是針對前述 Chou 的方法進行改良。一個多媒體封包傳送的率碼失真最佳化控制架構必需在資料單元群組之間利用解出 Lagrangian cost function of expected rate and distortion 的最小值來有效率的分配時間和頻寬的網路資源。在率碼失真最佳化控制的多媒體串流系統中，決定那些封包在每個可傳送機會是否應該被試著傳遞到客戶端是依據此一封包的截止日期、傳送過程的歷史記錄、通道的統計資料、回饋的資訊、封包間的相依性和如果封包可以在接收端正常解碼之每一封包可以減少的失真度來一起進行評估。
3. 傳輸系統驅動的媒體播放速度控制：在媒體傳輸時封包與封包之間的到達時間是隨機的，或者封包的到達時間雖然是有固定的時間但卻跟隨著連續隨機錯誤 (random errors bursts)。動態媒體播放 (adaptive media playout) 允許我們在媒體 frames 的傳送期間獨立分開的去調整播放的速度，並因而給我們在到達解碼端的截止日期 (arrival deadlines) 之前可設計一些控制機制，例如，以較慢的速度播放這個 frame，並延長這個 frame 後續 frames 的到達截止日期，針對影像而言，frame 週期的調整可以簡單的設計成調整每個 frame 的時間週期，另針對聲音而言，將音訊號處理的過程與時間軸的刻度連結以維持訊號的品質。因此，動態媒體播放技術能被用在多媒體串流系統以減少用戶端受到延遲的影響，用戶端的緩衝器可以用來對抗隨機的封包遺失和延遲，但如果使用較大容量的緩衝器雖可以吸收掉較大的頻寬變動，但也會在客戶端造成較長的播放延遲，動態媒體播放技術將利用可變動的播放速率來緩解客戶端緩衝器溢滿缺空 (buffer overflow and underflow) 的問題，並且降低播放延遲的情形。一個先進的系統會分析在動態媒體播放技術中延遲和緩衝器溢滿缺空的狀況，並調整動態播放的速率以防止緩衝器溢滿缺空的情形發生，其目標是將播放的延遲 (playout delay) 最小化，並且降低因傳輸延遲時間太長而誤以為封包遺失的比率 (late loss rate)。
4. TCP-friendly 流量控制：這個傳輸模組是用來控制 data units 的傳送速率。當 data units 的群組較大時，我們應該限制傳送端的傳輸速率，如此一來傳送端和網路的資源才不會被太多的傳送資料淹沒，為了達到這個目的，一個系統可以採用 equation-based 的 TCP 流量控制機制，目的是讓 UDP-Based 的媒體播放傳輸系統在傳送封包時，不要排擠同時在傳輸的 TCP 頻寬，並可使整個系統長時間的頻寬使用量接近一般的 TCP session 頻寬用量。根據 TCP-friendly 流量控制所量測出來的頻寬，多媒體串流可以被動態地調整傳送速率，並讓傳送速率的變化量儘可能的平緩，以符合多媒體串流的需求。而在網路擁塞情況很嚴重時，不但可以充分利用目前網路上的可用頻寬，並在與 TCP flow 競爭可用頻寬時也會讓 TCP flow 可以較公平的分享可用頻寬。不過一般 TCP-friendly 的機制所慣用的“等差性流量增加，等比性流量減小”演算法有一個很大的問題，就是位元速率會呈現出顯著的振盪特性，它會嚴重地影響多媒體串流的品質一致性。目前我們正針對這個缺點進行改進。

5. 高等精細可調層次式視頻編碼技術之研究 —— 蔣迪豪教授

Recently, MPEG committee announces a call for proposal to gather various technologies on scalable video coding. The report includes the detailed technical descriptions about the proposal from NCTU. In summary, our proposal is an extension of the prior work SRFGS. In our new approach, we utilize the wavelet transform to offer the spatial scalability to SRFGS. In addition, we present a context adaptive binary arithmetic coding for the enhancement-layer bit-plane coding. With the added scalability in spatial dimension, our SRFGS becomes a fully scalable codec that can simultaneously support spatial, temporal and SNR scalabilities. Like SRFGS, we adopt H.264 as the base layer video codec.

(1) Prediction architecture for spatial scalability

To support spatial scalability, we additionally apply discrete wavelet transform (DWT) to the prediction error of each stack. The DWT is conducted before DCT and quantization. We elaborate the details of our modifications for supporting spatial scalability in our proposal to participate the competition of MPEG-21 Part-13 Scalable Video Coding.

A. Macroblock level coding procedure for spatial scalability

The H.264 encoder encodes the picture in a MB-by-MB manner.

B. Wavelet transform and intra prediction at full resolution

Applying the wavelet transform to the prediction error may have problems when any intra MBs exist. As aforementioned, the original H.264 encoder conducts the coding in a MB-by-MB manner. When a MB is coded at intra mode, it is predicted by the neighboring reconstructed pixels. However, in our approach, the reconstructed pixels are not available during the prediction stage. For intra frames, we can simply solve the problem by applying the DWT on the original frame and encoding the new “original” frame, which may be not applicable to the intra MBs in an inter frame. If we apply DWT first and then perform motion compensation on the coefficients of each band separately, we can not remove temporal correlation between DWT bands. To resolve the problem, we partition the coding procedures of inter MBs and intra MBs into two passes. In the first pass, we encode all the inter MBs and apply DWT to both the prediction frame and the prediction error frame. After coded, we add up the reconstructed prediction error image with the prediction frame in the (wavelet) transform domain. In the second pass, the reconstructed frame containing only inter MBs is used for intra prediction. In the first pass, the intra MB has neither inter prediction value nor inter prediction error. Both the prediction frame and the prediction error frame could contain “holes” (missing MBs). The holes actually locate the positions of intra MBs. In the second pass, we perform the intra prediction in wavelet domain. Since each MB contains coefficients of four bands with 1-stage wavelet decomposition, we use the adjacent pixels of corresponding frequency bands for inter prediction. For example, the predictor of LL band should use the adjacent LL band pixels for prediction. With appropriate reference pixels, each band uses the same intra prediction method as in H.264. For simplicity, we disable luma 4x4 intra mode. In addition, there is no “plane” mode in luma 16x16 intra mode and chroma 8x8 intra mode.

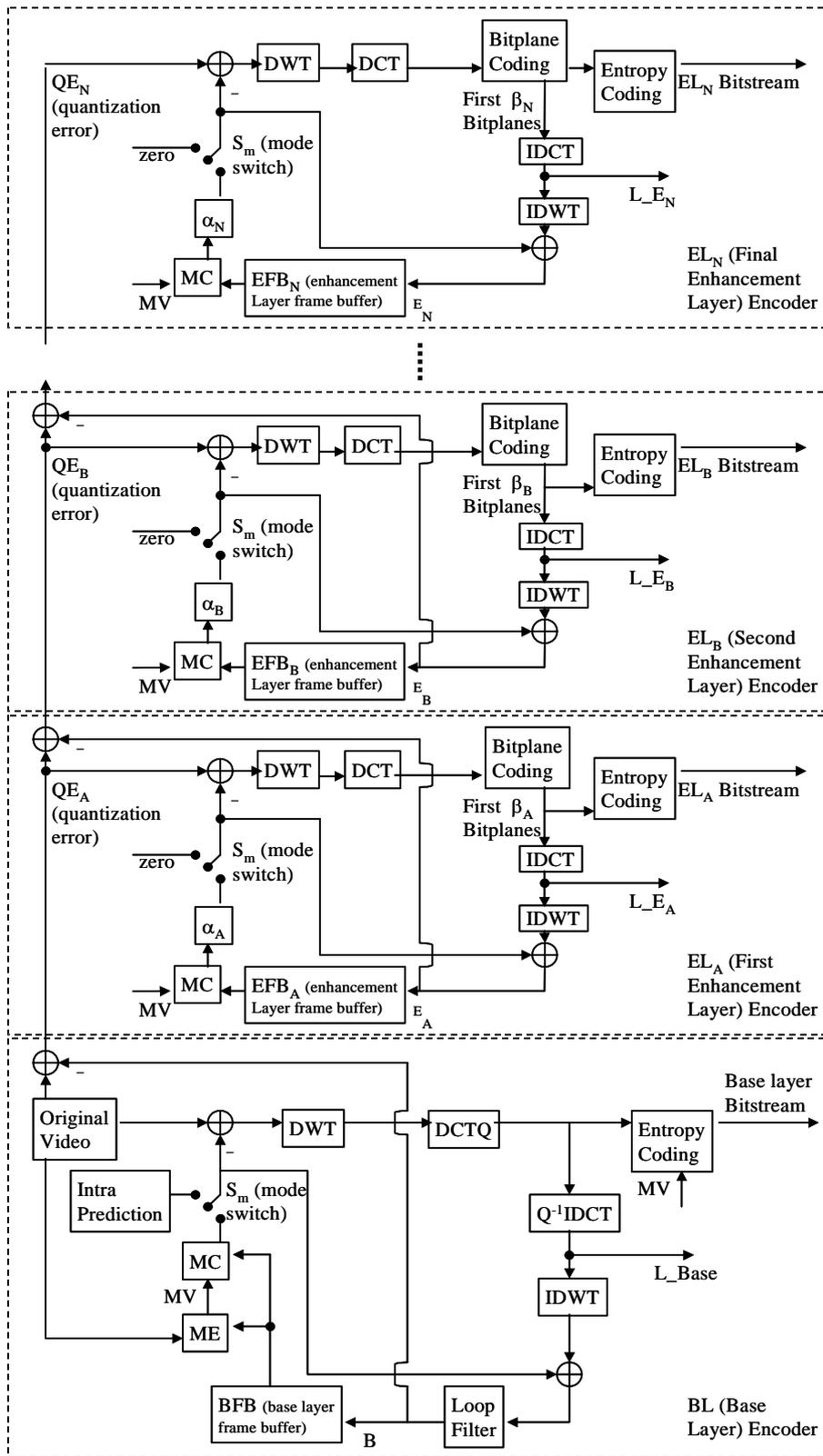


Figure 5-1. Proposed encoder for full resolution picture.

(2) Entropy coding of enhancement-layer

We proposed a context based binary arithmetic coding scheme as the entropy coding for enhancement-layer. Note that the proposed scheme is based on 4x4 integer transform. For the enhancement-layer bit-plane coding, we apply 4x4 integer transform to the wavelet coefficients. Specifically, we partition the bits of transform coefficients into different types. For each type, we design specific context according to different forms of correlations. In the following, we first introduce our bit classification.

A. Bit classification

Different bits of a transform coefficient carry different weights of information. Most significant bit (MSB) represents the dominant energy of a coefficient and least significant bit (LSB) denotes the less important information. Thus, bit classification is required for context design. The bits of each transform coefficient are partitioned into three types including significant bit, refinement bit and sign bit. Additionally, for each bit-plane of a transform block, we define two symbols, End-Of-Significant-Bit-plane (EOSP) and Part_II_ALL_ZERO, to save more bits by grouping the zero bits.

B. Context design

Our context design for each type of bit is mainly based on spatial correlation and energy distribution of transform blocks.

C. Adaptive zigzag scanning and Enhancement-layer coding flow

Different subbands have different energy distributions. The LL band mainly contains low frequency signal while the HH band includes high frequency signal. To account for different energy distributions in different bands, we apply different zigzag scans for different subbands. Also, we propose a bit-plane coding scheme using the raster scan order.

(3) Concluding Remarks

The Robust Scalable Video Coding (RSVC) proposed submitted MPEG committee is a fully scalable video codec that can simultaneously support temporal, spatial and quality scalabilities.

6. MPEG 智財管理與保護系統及強韌視訊解碼器之設計與模擬 -- 杭學鳴教授

過去數年間，由於多媒體通訊突飛猛進，網際網路迅速普及，以多媒體物件製作、處理、編輯及傳輸為目標的 MPEG-4/MPEG-21 標準活動受到大家的關注。我們在此研究計畫的目的為(1)深入研究並模擬 MPEG-4 IPMP 延伸系統，並將其概念實作於 MPEG-21 Test Bed 上，以及(2)畫面間小波視訊編碼 (Interframe Wavelet) 之開發與研究，利用 AVC 中之移動估測方法改善移動補償時間濾波，並提出一個適用於 AVC 移動估測之可調式移動向量技術。

(1) MPEG-4 IPMP 系統的研究

由 1997 年開始，MPEG 組織開始嘗試為 IPMP 訂立標準的公用介面和相關協定，以解決數位化媒體在網路上交易所衍生的問題，目標是把 IPMP 整合到現有的 MPEG 標準之中。其中利用密碼方法或數位浮水印的技術以有效的提高非法剽竊的困難度。目前，類似但簡化之技術已被應用在數位電視系統中。截至本論文撰寫時，MPEG-4 和

MPEG-2 標準中的 IPMP 訂定已經完成，而 MPEG 組織也開始投入更多的資源訂定 MPEG-21 IPMP 標準。

A. IPMP 標準與架構

我們的研究承續之前對於 MPEG-4 系統的了解，本期專注於 MPEG-4 標準 (ISO/IEC 14496-1) 中所提供的 IPMP 架構，並輔以部分 MPEG-21 IPMP 及 MPEG-21 Test Bed，以期對 MPEG IPMP 之運作有更前瞻的認識。IPMPX 架構觀念為虛擬終端機(Virtual Terminal)，與既有的 MPEG-4 系統以 Message 互相溝通。IPMP 虛擬終端主要由兩大概念合成，一為 Message Router (MR)，另一為 Tool Manager (TM)。Message Router 負責將所有的 IPMP Message 傳送至對應的 IPMP Tool 或終端機本身，而接收 message 的一方則根據 Message 內容負責串流的資料處理或控制，例如解碼或是權限控管。Tool Manager 負責 Tool 的建立、消滅、與關聯等功能，當需要時，可由 MPEG-4 系統或 Message Router 發出管理需求。

B. 模擬與實驗

軟體模組的設計上，我們盡可能保留 MPEG 文件上所描述的概念性物件，例如 MessageRouter 與 ToolManager 等等(圖 6-1)。而在 API 的設計上，我們則參考 IM1 IPMPX 的設計，並稍加延伸，例如加上 timestamp 參數以利 terminal 傳遞系統時間給 tool。在實作上，為求與標準格式相容，我們採用 PSL MPEG-2 IPMPX 錯誤! 找不到參照來源。中的 MessageInterface 函式庫產生與解析 ToolMessage。較為特別的是 context 的設計，ToolMessage 利用 context ID 決定 routing 方式，但 IPMPX 標準中並未定義 context 的實際結構，因此我們根據可能的使用情況，將 context 設計成樹狀結構，依照 top、object、elementary stream、IPMPTool 四個階層加以關聯，並自動給予各 context 識別編號。另一個標準中沒有定義的物件為 IPMPFilter，根據相關描述，我們歸納出 IPMPFilter 應該是一個 container，允許至多 256 個 IPMPTool 串接，而資料串流則依序通過各 Tool。

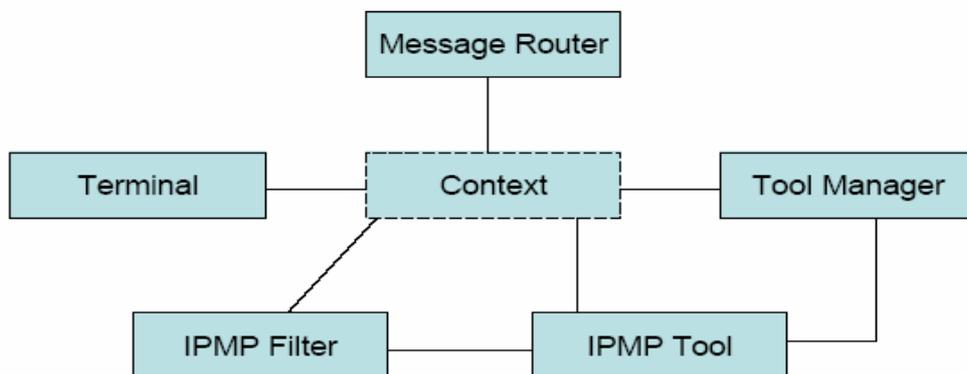


圖 6-1 重新設計的 IPMPX 模組關係圖

最後，我們用一個加密傳送影像的系統展示我們的研究與實作成果。圖 6-2 所示 client 端程式啟動 IPMPX 子系統，並週期性換 key 進行 DES 解密的過程，就使用者而言，這些動作都是隱藏在 virtual terminal 完成，只會看到播放程式播放短片。接著我們刻意在某個時間區間給定錯誤的 key，對 decoder 而言，因為無法經由解密過程取得正確的影

片資料，所以解碼過程發生錯誤，如圖 6-3 左側照片所示。



圖 6-2 程式啟動 IPMPX 與正常解密 圖 6-3 解密用的 key 錯誤造成解碼錯誤

(2) 視訊畫面間小波編碼技術的研究

視訊畫面間小波編碼技術(Interframe Wavelet)是近幾年被提出的壓縮演算法，其優點就在於全可調層次性(Full Scalability) — 同時兼具三種可調層次性(空間可調層次性(Spatial Scalability)，時間可調層次性(Temporal Scalability)，位元率可調層次性(Bit Rate Scalability))，所以利用一次壓縮好的視訊，就可以提供給各種應用。而其壓縮效能在高位元率時跟目前標準制定的先進視訊編碼技術(ISO/IEC 14496-10 Advanced Video Coding, AVC)十分接近。

A. 移動補償時間濾波之改良

本計畫利用 AVC 中之移動估測方法對移動補償時間濾波改善在此移動搜尋的過程中，為了增加搜尋的精確度，我們將參考畫面經填補濾波器(interpolation filter)，進行 1/2 像素與 1/4 像素精確度的移動搜尋。由於搜尋方塊切割較為細緻，也造成了移動向量之位元率過高，因此我們利用相鄰方塊移動向量的相關性，進行預測。我們用被預測方塊的左方、左上方、上方及右上方進行被預測方塊的移動向量預測。藉由細緻的搜尋方塊切割與移動向量預測，我們可以精確有效率的將畫面切割成適當的搜尋方塊，並找到對應的移動向量。圖 6-4 為經過移動搜尋之後的畫面切割圖例。

B. 可調式移動資訊

在傳統的小波轉換中，擁有空間軸、時間軸及畫質之可調式設計，移動資訊在空間軸與畫質可調式設計中是無法切割的。當可允許傳送之位元率太低時，抽取器(Extractor or Puller)很可能會因移動資訊過大，而沒有足夠可分配的位元率進行小波係數切割而失敗。此外，在非常低位元率的情況下，我們會希望在省去部份的移動資訊將較多的位元率分配給小波係數，以換取較好的畫質。因此，在移動估測之後，我們對移動資訊進行切割。



圖 6-4: 經過移動搜尋之後的畫面切割

在 AVC 的畫面間預測中，基本的處理大小為 16×16 之 macroblock。每一個 macroblock 可再細分為 16×16 , 16×8 , 8×16 , 8×8 , 8×4 , 4×8 及 4×4 ，並且所對應到的移動向量具有 $1/4$ 像素之精確度。我們依下列步驟將移動向量做切割。

第一步：進行 16×16 大小之整數像素精確度之移動搜尋，所產生之移動向量為“基本層”。

第二步：進行 16×16 與 8×8 大小之 $1/2$ 像素精確度之移動搜尋。與基本層之差值將被保留編碼，稱“第一加強層”。

第三步：進行所有搜尋方塊大小之 $1/4$ 像素精確度之移動搜尋。將其與基本層與第一加強層之和的差值將被保留編碼，稱“第二加強層”。

第四步：將所有的移動向量階層分別利用 CABAC 編碼。

(3) 成果

這部分成果分成兩部分，其中 IPMP System 在 2004 年 7 月加入 MPEG-21 Multimedia Test Bed 中，已成為 Test Bed 軟體的一部份。Interframe Wavelet 則在 2004 年 3 月與 7 月提案 MPEG 標準組織。3 月之提案為參加 scalable video coding Call-for-Proposal 競賽，在 14 個提案中經視覺主觀評審，成績中等。在國際大企業競爭下，此結果似乎尚可。7 月之提案為參加 Core Experiments，改善目前的 Reference Model。

7. 多點視訊會議技術之研究 -- 林大衛教授

近年來，桌上型視訊會議技術已愈趨實用與可即。然而目前一般的系統仍不具有近似當面開會的視聽感覺。本計畫主旨在研究分散式桌上型多點視訊會議技術，其中特別著重視訊的處理。我們擬在每個會議端點的電腦螢幕上顯示一個虛擬的會議室場景，其中呈現所有其他端點的與會人員。為此，每一端點需先將本地輸入視訊加以分割，取出與會者影像予以編碼，然後傳到其他端點。每一端點也需將所有接收到的視訊予以解碼及

合成。本計畫之研究係建構在 MPEG-4 規範的基礎上，採用個人電腦為實現平台。計畫之研究子題可分四大組：會議系統、網路傳輸、傳送視訊處理、與接收視訊處理，預定以三年時間進行研究。本報告係針對第一年之研究，其中著重視訊分割技術、有效率之 MPEG-4 視訊編碼、及 MPEG-4 規範之了解。在視訊分割方面，我們提出了一些做法，並使用個人電腦實作了一個即時視訊輸入與分割系統。在 MPEG-4 視訊編碼方面，我們使用一個公眾領域的軟體加以改進，在個人電腦上實現了一個較快速的 MPEG-4 視訊編碼器。在其他 MPEG-4 相關規範之了解方面，我們研讀了有關場景合成與訊號傳輸介面的文獻，也試取得相關軟體並檢驗其功能。

(1) 緣由與目的

計畫主旨在研究桌上型多點視訊會議之相關技術，並建構一個實驗性之系統。本計畫之研究標的，可透過圖 7-1 說明之。圖中左方呈示之個人電腦(PC)及螢幕(display)，為每一會議點所使用之設備。螢幕顯示一個虛擬之會議室(virtual conference room)，其中之視訊為所有其他會議點所傳來之與會人員視訊的一個合成(composition)。會議室場景及各人的位置是由主席安排。螢幕上另有二個視窗，即控制台(control panel)與本地視訊之預覽(local preview)。我們發現：最近擬訂的 MPEG-4 標準，其中的若干規範相當適合本研究之所需。例如：其視訊編碼部分容許將視訊分割後再編碼，其資料結構與合成部分定義了一個相當有效率的 BIFS (Binary Format for Scenes)，其網路傳輸部分定義了 DMIF (Delivery Multimedia Integration Framework)等。故本計畫之研究係基於 MPEG-4 之規範。

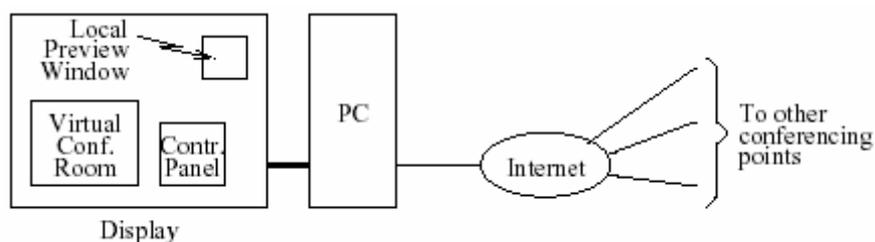


圖 7-1：多點視訊會議系統架構示意圖

本計畫之研究子題可分為四個群組，即會議系統、網路傳輸、傳送視訊處理、及接收視訊處理，預定以三年時間進行相關研究。本報告係針對第一年之研究，其中重點為第三群組子題之深入研究，以及第二與第四兩群組子題之初步探討。申言之，在第三群組子題，即傳送視訊處理方面，我們研究了視訊分割技術與有效率之 MPEG-4 視訊編碼。關於視訊分割，我們提出了一些適用於會議型態影像的視訊分割方法，並使用個人電腦實作了一個簡單的即時視訊輸入與分割系統，目前在繼續改進其運算速度中。關於 MPEG-4 視訊編解碼，我們使用一個公眾領域的軟體加以改進，在個人電腦上實現了一個較快速的 MPEG-4 視訊編碼器。而在第二及第四群組子題，即網路傳輸與接收視訊處理兩方面，我們研讀了有關 MPEG-4 場景合成與訊號傳輸介面之規範的文獻，也試取得相關軟體並檢驗其功能。在第二年度之研究中，我們將較著重第二及第四群組子題的探討，也擬開始第一群組子題，即會議系統之研究。上述視訊編碼器改進過程中所累積的經驗，可為第四群組子題中，視訊解碼器實現之參考。

(2) 結果與討論

A. 視訊分割：更精確界定物件邊界及將分割結果用於場景合成之研究

本研究所提出的視訊分割演算法架構如圖 7-2 所示。其中“Edge Detection”使用 Canny edge detector；“Change Detection”係將兩畫框(frame)間相異程度較高的像素取

出稱為 changed pixels；“Forward Tracking”及“Backward Validation”係使用階層式運動估計，等效之搜尋範圍為正負 14 像素，複雜度約與全尋法相似。“Video Object Extraction”為最創新之部分，使用形態學式(morphological)之處理，以獲得一相當逼近實際物件邊界之物件草型(object mask)，然後用 Dijkstra 最短路徑演算法 (D. W. Dijkstra, “A note on two problems in connexion with graphs,” *Numerische Mathematik*, vol. 1, pp. 269-271, 1959) 連結草型外緣邊界之「斷裂」之處，以得到最後萃取出之物件。由於 Dijkstra 演算法之複雜度與其所需搜尋的像素數目成平方關係，故當物件草型相當逼近實際物件邊界時，Dijkstra 演算法之複雜度可以降低。

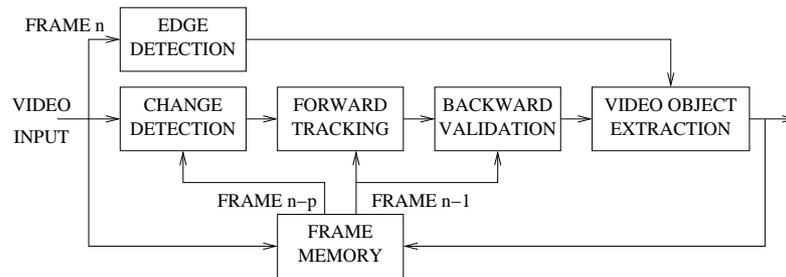


圖 7-2：所提出之視訊物件萃取與追蹤方法之一

以上 video object extraction 方法，在物件形狀高度非凸狀(highly nonconvex)時，特別能顯出其效用。圖 7-3 顯示一些視訊分割結果，其中最後之 Dijkstra 演算法所用之搜尋深度為 $D_w = 5$ 。配以較高速之運動估計法，整個分割法可對 CIF (352x288) 視訊在現有一般個人電腦中達每秒數十張畫框之即時執行速度。

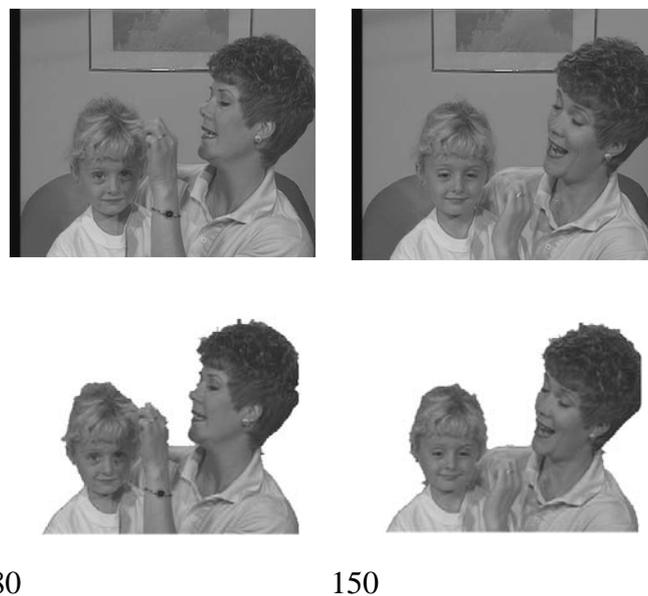


圖 7-3：一些 Mother-and-Daughter 視訊之分割結果。上列：原始圖框；下列：分割出之移動物件。底部數字為圖框序號

B. 視訊分割：使用背景建構法分割視訊以及即時視訊分割系統之實現

在此，我們設計一種方法來收集各個畫框中的背景部份，建構出一個盡量完整的背景圖。然後將現在收到的畫框與背景圖比較，把差異很大的部份取出並做一些修飾，就

可以將移動的前景物件分割出來了。

我們首先分析視訊中的攝影機雜訊量(Camera Noise Estimation)。我們設計了一個二級的方法來估計雜訊的變異數，以減低移動物件對估計精確度的負面影響。其次，我們構建一個暫時性的前景物件草型(temporary foreground mask)。這是使用 Frame Difference、Fill-In、及 Canny Operator 三個功能方塊達成。其中 Frame Difference 取得畫框中變異較大的像素，Fill-in 將像素間的空白處填滿，使其涵蓋移動物件的區域，Canny Operator 及相隨的運算則使區域內縮，使之更接近實際的物件形狀(但可能仍有相當差異)。圖 7-4(a)所示為一個結果的例子。第三，我們構建一個短期的背景(Short-term Background Estimation)。這是透過分析連續六張畫框來達成。如果某一像素值在這些畫框中變化不大，則暫將之算為背景像素。圖 7-4 (b)所示為一個結果的例子。第四，我們使用以上結果來構建一個靜態背景畫面(Stationary Background Buffer)。由於移動物件若是內部的亮度與色彩相當平滑，則在簡單的分析中，有可能被誤判為背景，所以我們使用之前獲得的暫時性前景物件草型來將短期背景像素予以加權，如果其累積之加權值超過某一門檻，再將之放入最終的靜態背景畫面。繼續上例，圖 7-4 (c)所示為根據圖 7-4 (a)所得之權值，其中黑色表示最高的權值(比較可靠的背景部份)，白色表示最低的權值(零)，灰色表示中等的權值。最後分割結果在圖 7-4 (d)。

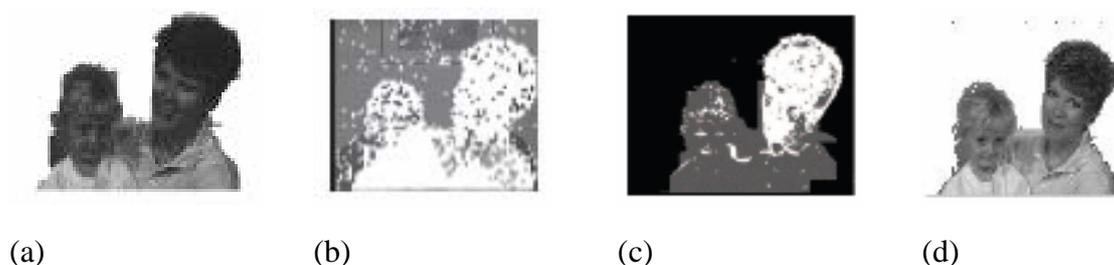


圖 7-4：演算過程中之部分結果

C. 有效率之 MPEG-4 視訊編碼

由於我們考慮使用個人電腦做會議系統的平台，所以需要在其上建構一個有效率的 MPEG-4 視訊編碼器。此處我們使用 Microsoft 公司開發的一個公眾領域的 MPEG-4 視訊編碼軟體，在個人電腦上，試改良程式寫法，並使用 Intel CPU 的 MMX (multimedia extension) 處理器單元以加速其運算。上述編碼軟體係處於 MPEG-4 視訊定義之 Main Profile 和 Simple Scalable Profile 之層次。

我們遂分析 MPEG-4 視訊編碼軟體，找到其中較耗時的部份。然後使用適用的 MMX 處理器指令來加速。上述分析，主要是藉助於 Intel 的 VTune performance analyzer。此軟體工具提供 tree-structured call graph、各函式使用時間百分比及使用之 clockticks 等資料，相當有助於了解程式的瓶頸所在，並可用以比較修改前及修改後之程式運作差異。

程式中第一耗時的，如一般可預期，是運動估計。對有的視訊而言，可佔 90% 以上的運算時間。我們也使用較快速(但效果稍遜)的運動估計法作實驗。實驗顯示 Foreman 視訊的編碼速率，其中原始程式約為每秒 3 張畫框，修改後但使用相同運動估計法(FS, 即 full search)的程式約每秒 7 張，改用較快速的運動估計法(DS = diamond search, NDS = new diamond search, 2DLS = two-dimensional logarithmic search)則可達每秒 20 張。

8. 計畫整體成果與自評

本計畫旨在建構符合 MPEG 標準之網際網路即時(real-time)多媒體傳輸測試平台，並以此平台為核心開發各項多媒體傳輸相關技術，如：可調層次式視頻編碼技術，串流多媒體伺服器與資料庫，智財管理與保護等。參與本項計畫的三位老師杭學鳴、蔣迪豪、蔡淳仁及博士後王俊能，在國科會與業界廠商合作計畫補助旅費情況下參與 MPEG 標準會議，過去一年中，提出 31 件 MPEG 貢獻文件(Contributions)，將上述技術提案至 MPEG 標準組織，有助我國技術之進入國際舞台。參與 MPEG 國際會議之具體成果簡述如下。

蔡淳仁教授擔任 MPEG-21 Part-12 Multimedia Test Bed for Resource Delivery 以及 MPEG-21 Part-13 Scalable Video Coding 之共同編輯者(editor)。交大 MPEG-21 團隊之成員人數每學期約在 12 人左右。該團隊的目標是建構並改進 MPEG-21 Part-12 Multimedia Test Bed。這個團隊合作成果已經在 2003 年 12 月被 MPEG 標準委員會接受為 MPEG-21 Part-12 Multimedia Test Bed for Resource Delivery 的大會草案(Committee Draft)，在 2004 年 7 月 MPEG 會議中成為 FDTR。

蔣迪豪教授擔任 MPEG-4 14496-7 Optimized Reference Software 之共同編輯者(editor)。負責編撰的最佳化 MPEG-4 14496-2 之 Simple Profile Video Codec 已經進入此標準之修正案第三(DAM)階段，成為正式標準之修正案。蔣迪豪教授亦擔任 MPEG-4 Part-2 Video Subgroup 之 Co-Chair，參與 MPEG-4 Part-2 標準書之修訂。並與 MPEG-4 Part-9 之專家合作以完成 Software-Hardware codesign 的平台。

2004 年 3 月 MPEG 標準委員會舉行 scalable video coding Call-for-Proposal 競賽。交通大學團隊提出兩項提案，分別為 S07 -- A Robust Scalable Video Coding Technique，以及 S08 -- A Scalable Video Coding Scheme Based on Interframe Wavelet Technique。在 14 個付費參與視覺品質評估的候選方案中，除了 Microsoft Research Asia 之外，交通大學是唯一提出兩案的單位。我方案品質評估為中上，在世界強隊競爭下，結果尚佳。2004 年 7 月 MPEG 的第 69 次會議上，交通大學團隊繼續提出一項改善 Interframe Wavelet 提案，Response to CE1b in SVC-- SB-Reach Method for Entropy Coding。

此外，本計畫有多項學術與應用技術方面之創新：在視訊串流伺服器、可調層次式視訊編碼技術、多媒體傳輸網路機制、多點視訊會議系統、及視訊分割合成與編解碼等等技術與實作方面的新發現。發表國內外期刊與學術研討會論文多篇，詳見各子計畫報告。

更直接並且對工商業更有價值的貢獻將是人才訓練。同學們在學校階段已熟悉較前瞻的世界標準，畢業後進入產業，直接有助於產業界開發新產品，提昇我國工業技術能力。

綜合評估：本計畫產出相當多具有學術與應用價值的成果，並積極參與國際 MPEG 標準會議，將我人研發成果推廣到國際舞台。此外亦達到人才培育之效，整體成效良好。