

行政院國家科學委員會專題研究計畫 成果報告

總計畫(II)

計畫類別：整合型計畫

計畫編號：NSC93-2219-E-009-016-

執行期間：93年08月01日至94年07月31日

執行單位：國立交通大學電子工程學系暨電子研究所

計畫主持人：杭學鳴

共同主持人：蔡淳仁，蔣迪豪，林大衛，李素瑛，王聖智

計畫參與人員：唐之璇，唐之璋，楊政翰，陳繼大，施瑛姿，徐漢光，鍾國洋，
蔡崇諺，蘇郁淵，丁健文，陳勇竹，吳和璋

報告類型：完整報告

處理方式：本計畫可公開查詢

中 華 民 國 94 年 10 月 31 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)(II)
MPEG Integrated Multimedia Platform and Applications (II)

計畫類別： 個別型計畫 整合型計畫

計畫編號：NSC 93-2219-E-009-016

執行期間：93 年 8 月 1 日至 94 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 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：唐之璇，唐之瑋，楊政翰，陳繼大，施瑛姿，徐漢光，鍾國洋，蔡崇諺，蘇郁淵，丁健文，陳勇竹，吳和璋

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 94 年 10 月 30 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)(II)

MPEG Integrated Multimedia Platform and Applications (II)

計畫編號：NSC 93-2219-E-009-016

執行期限：93 年 8 月 1 日至 94 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：唐之璇，唐之瑋，楊政翰，陳繼大，施瑛姿，徐漢光，鍾國洋，蔡崇諺，蘇郁淵，丁健文，陳勇竹，吳和璋 國立交通大學研究生

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 的技術報告類標準文件。

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 多媒體傳輸機制及通訊協定在嵌入式行動平台上的分析設計，(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，本報告將略敘第二年之各項成果。

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 Technical Report of MPEG-21 Part 12: Multimedia Test Bed Resource Delivery.

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 and Protocols for Embedded Mobile Platforms, (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 second 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 多媒體傳輸機制及通訊協定在嵌入式行動平台上的分析設計——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	14
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	17
7. 多點視訊會議技術之研究 -- 林大衛教授.....	20
8. 計畫整體成果與自評	23

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)。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)。目前本案已正式成為 MPEG 技術報告類標準文件: 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.*
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被 ISO/IEC 組織規劃為 MPEG-21 Part 13: Scalable Video Coding (SVC)。後來轉為 MPEG-4 part 10 (ITU H.264) Advanced Video Coding 的 extension。2004 年 3 月 MPEG 第 68 次會議上評估世界多家知名研究機構的提案，以角逐成為 SVC 視訊壓縮標準之起始點。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。之後在 2005 年我們持續提案參加 MPEG SVC 的 Core Experiments。在未來幾年之內，本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 多媒體傳輸機制及通訊協定在嵌入式行動平台上的分析設計 —— 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，我們提出了兩項成果。一是以影像內容為基礎的視訊對比加強技術，我們稱之為「以區域資訊為基礎的對比加強」。另一項是一種彈性的編碼架構，我們稱之為「頻率資訊選擇技術」。在「以區域資訊為基礎的對比加強」方面，我們根據影像內容分析的結果，產生相對應的亮度調整機制。運用此項技術，我們可以用來改善視訊品質，並達到特定應用之目的。在「頻率資訊選擇」方面，我們以 MPEG4 視訊編碼規格中的 SNR 可調層次式編碼技術為基礎，提出一種更具彈性的架構，可針對不同頻段的資訊個別調整。同時我們設計視覺實驗，根據測試結果歸納出各個頻段之間適當的調整方式。

A. 以區域資訊為基礎的對比加強技術

(1) 背景

本專題中將特別研究視訊資料的對比處理。給一影像，我們可以統計出該圖對應的亮度值柱狀圖來。之後，我們希望藉著調整此柱狀圖而得到一張亮度對比較高的輸出影像。而柱狀圖均等化技術的基本概念便是分配較多的亮度範圍給柱狀圖中出現頻率較多的亮度值，以期達到對比加強的目標。本次研究提出一新的對比加強作法。我們將以柱狀圖均等化的作法出發，以一全新的觀點來重新檢視柱狀圖均等化法的作法。並以此觀點出發，發展一套全新的對比加強方式，並藉此改善所接收到的視訊品質。

(2) 演算法說明

我們提出一以區域資訊為基礎的對比加強演算法，以一個大於 1×1 的視窗來作為分析的基準，並根據分析的結果產生一新的柱狀圖資訊，用來進行對比加強。首先我們定義一 $n \times n$ 大小的視窗作為分析的基準。另外我們提出兩個函數 f_1, f_2 ，並以此分析視窗內亮度值的特性。其中 f_1 主要是分析視窗中像素亮度值之平均趨勢，而 f_2 主要是分析視窗中像素亮度值的變異度資訊。我們累計 (f_1, f_2) 的出現頻率。當整張影像掃描完畢後，我們會得到一二維的 (f_1, f_2) 機率分佈函數， $\text{Dist}(f_1, f_2)$ ，我們稱之為**影像特徵分布函數**。函數統計了此張影像中亮度值的局部平均值與局部變異度的分佈情形。

在本篇報告中，我們選用下列函數作為分析的基準(假設我們使用 3×3 的視窗)，其中

$$f_1(x) = \frac{1}{9} \sum_{i=1}^9 I_i$$

$$f_2(x) = \text{Max}\{I_i\} - \text{Min}\{I_i\} \quad i = 1, \dots, 9$$

而 I_i 表示影像亮度資訊。

(3) 實驗結果

圖 2-1 為所提方法和柱狀圖均等化的比較，我們可以清楚的觀察到，藉由不同參數的設定，我們所提方法能提供較傳統作法更為彈性的調整機制，且畫面較為自然。此外，若測試影像序列受到高斯雜訊的干擾，將此序列經由傳統的作法處理後，將會出現雜訊過度加強的效應(見圖 2-1(c))，然而，經由所提方法的處理，則不僅能夠達到適當的對比加強，雜訊也不會比原圖明顯(見圖 2-1(b))。

(4) 結論

在本研究中，我們提出了一個全新的概念來進行影像對比加強處理，我們從影像分析的角度出發，以兩個分析函數來分析影像中的特徵，並根據求出影像的特徵分布函數，推出符合影像特性的亮度轉換函數。另外，所提出方法也成功將對比加強中常遇到的雜訊加強和過度加強的問題作一處理。

B. 頻率資訊選擇技術

(1) 背景

可調層次式視訊編碼技術(Scalable Video Coding)。FGS 為 SVC 的解決方案之一，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，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。

目前的影像/視訊編碼皆有保留針對影像/視訊內容可做的調整度，其方法大多是在 DCT quantization table 作調整。SNR-scalability 中，base layer 也具有可調整的 Q-table，但是 enhancement layer 並無類似的能力。另外一種可調層次式編碼-空間可調層次式(Spatial-scalability)，其 enhancement layer 具有頻率調節的能力，但是其編碼效率下降，運算複雜度提升。本報告研究的目的，便在於提供一個具有頻寬調變與頻率調變能力的編碼架構，其編碼效率與運算複雜度接近 baseline FGS。

(2) 架構

Frequency Information Selection (FIS)架構在 SNR-scalability 之上，亦分成基本層與加強層，而基本層與加強層再分別分成三個頻段，由低頻成分到高頻成分，分別為 L0、L1 及 L2，如圖 2-2 所示。



(a) 原始影像



(b) 所提方法，其中結合常數 $k=1$ ，調整常數 $M_0=3$



(c) 柱狀圖均等化

圖 2-1 處理結果比較

FIS 在傳輸時，比 SNR-scalability 多了一個維度-頻率維度。SNR-scalability 可以依照給定頻寬傳送對應大小的加強層；FIS 亦能根據頻寬調整資料量，在某個給定的頻寬，FIS 還能調整不同頻段之間的比例，以最適合目前視訊內容的方式傳送，如圖 2-3 所示。

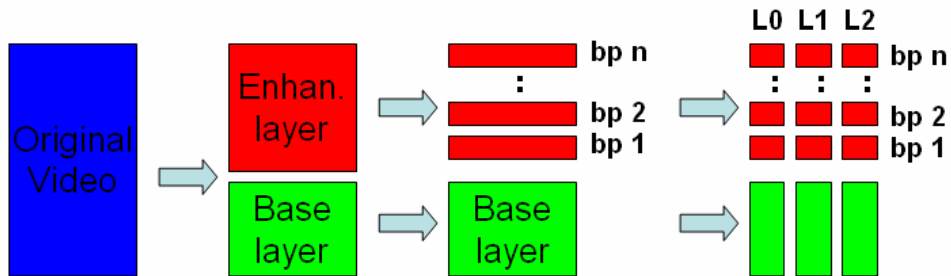


圖2-2. FIS整體架構圖

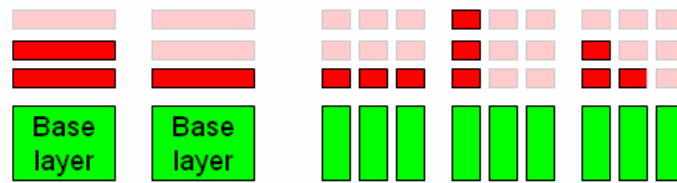


圖 2-3. SNR-scalability 與 FIS 串流示意圖

FIS 採用如圖 B.4 的方式將 8×8 離散餘弦轉換係數分成三個頻段，同時也將係數量化分離為基本層與加強層，加強層再進一步分成數個位元層，如圖 2-4。加強層中，L0 改用霍夫曼編碼(Huffman Coding)，L1 與 L2 各自採用新設計的編碼表。

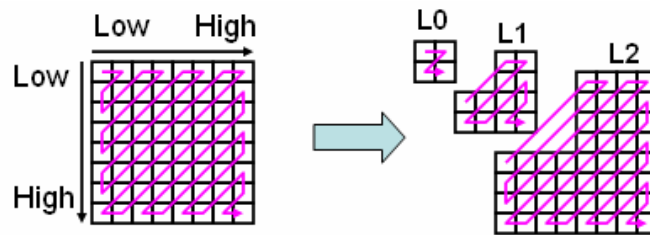


圖 2-4. 8×8 離散餘弦轉換係數分頻段與 Split-Zigzag 掃描

以我們所做的視覺實驗，可歸納出如下的結論：

- 資料量低時(平均兩個位元層以下)，大抵上低頻資訊比較重要
- 資料量高時(平均三個位元層以上)，額外的高頻資訊將獲致較大的品質改善
- 不同的測試者的測試結果具有相當的差異
- 平均位元層介於 2~3 之間時，改善較明顯，不同測試者的結果較一致

3) 實驗結果

我們以 FIS 與 SNR-Scalability 對 foremen 和 akiyo 選定相同參數編碼，並選擇數個位元率解碼，FIS 採用「只取 L0」的選擇機制。「只取 L0」對於 foreman 的效果特別好，加強層位元率由 256kbps 到 768kbps 間，所有受測者一致認為 FIS 提供了明顯的品質提升，由於在每個 IPB 週期的開始；也就是當區塊效用嚴重的時候，加強低頻資訊能明顯提升 foreman 的視覺品質。

(4) 結論

在這部份的研究中，我們提出了一個架構在 SNR-Scalability 上的編碼方式-FIS，可以獨

立調變三個頻段的資料量；藉著這樣的自由度，依照視訊內容選擇傳輸的優先次序，可以在不同頻寬下對視訊品質作進一步的最佳化。對於區塊效應明顯的情形，簡單的選擇機制便能帶來明顯的品質改善。

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

隨著網際網路規模的不斷擴大以及資訊科技進步快速，愈來愈多的多媒體資料以數位的形式藉網路為媒介廣泛地傳播著。在傳輸多媒體資料的過程中，有網路頻寬、封包遺失以及封包延遲的限制，但是現今的網路並沒有提供 QoS (Quality of service) 的能力，來保證多媒體資料的傳輸，因此，如何設計一些機制來確保多媒體資料的傳輸便面臨很多的挑戰。

資訊種類之多元以及資訊量之豐富對於使用者來說哪些是有用的資訊及哪些是可忽略的資料自然成為一個重要的議題。因此對於使用者而言，在眾多資料之中如何有效且快速地獲得真正所要的內容便是我們所要努力的目標。然而除了提供快速有效的查詢方式之外，還必須讓使用者同時享有自然且友善的查詢環境，而利用資料內容為索引及查詢 (Indexing & Retrieval) 為依據的方式(Content-Based Access) 正能夠滿足此需求。因為所謂 Content-Based Access 即是以資料內容之特徵當成索引來建立資料庫，因此特徵描述之充分與否則直接影響查詢結果之優劣。而傳統資料內容之描述可為某些關鍵字、屬性的集合、或是某些以文字為主的抽象表示法。

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之多媒體資料。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。最後, 我們希望分別實作出支援單點傳送與多點傳送的視訊串流伺服器以及視訊資料庫, 最後再將它們整合。

(2) 系統設計與實驗結果

◆ 支援多點傳送的視訊串流伺服器

1. 應用層通訊協定設計 (Application-layer Protocol Design)

在多點傳送的視訊串流伺服器中, 由於結合 FGS 適合在網路上傳送的特性, 伺服器只需要將多媒體資料分成數個 channel 並把它們送到網路上即可, 而且只需送一份資料, 然而在這樣的模式下, 一樣會發生封包遺失的情況, 因此, 必須加入錯誤恢復的機制, 在這個計劃中, 我們設計一應用層通訊協定稱 SRTP (Scalable Real-time transport protocol) 以提供接收端錯誤恢復的能力。

2. 接收端緩衝區管理模組 (Buffer Management Module)

為了降低接收端封包延遲以及封包延遲變動的影響, 必須在接收端配置緩衝區以讓接收端的多媒體品質保持平順的狀態, 因此, 如何設計一個緩衝區, 使得它具有快速存取、有效的空間利用以及支援封包遺失的恢復能力, 便是一個很重要的研究議題。因此我們提出一個新的接收端緩衝區結構, 這個結構包含緩衝區索引表、指標以及資料緩衝區三個部分。緩衝區索引表屬於一個雜湊表 (Hash Table), 被用來紀錄每張畫面在資料緩衝區的起始位置以及快速地將資料存放至資料緩衝區中; 資料緩衝區為多媒體資料真正儲存的地方, 而且還儲存有關多媒體資料封包的資訊, 例如, 我們設計的 SRTP 中的標頭資訊, 以讓該緩衝區的設計支援快速的錯誤偵測。

3. 錯誤恢復模組 (Packet-loss Recovery Module)

在多點傳送的模式下, 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, 我們提出的錯誤恢復模組也正朝這兩個方向努力, 其中包含 (1) 錯誤恢復啟動管理 (Activating Loss Recovery Manager)、(2) 服務品質監視器 (QoS Monitor)、(3) 虛擬群組管理 (Virtual Group Manager)、(4) 錯誤恢復管理 (Recovery Manager) 四個部分來監控封包發生遺失時以使服務品質獲的確保。

◆ 視訊影片資料庫

視訊影片資料庫的有效性依靠於是否能夠容易找到有興趣的視訊影片。為了提供能根據視覺的內容來探索、瀏覽和擷取影片的能力, 在視訊影片資料庫中有效的偵測相似影片的技术成為在多媒體與資料庫相關領域中的一個最重要與富有挑戰性的議題。我們在視覺特徵方面, 提出了擷取較特別之特徵如 **edge**, 來建構一個新的時間化之邊緣型態表示法以用來偵測影片的相似度。

為了減少計算複雜度，我們使用影片中的關鍵畫面來做比對。利用一個以 GOP 為基礎的方法來做場景變換偵測，並將影片切割為許多片段。從已切割好的影片片段中選出 I-frame 作為其關鍵畫面。最後，使用一個結合梯度能量直方圖、最高邊緣密度垂直切片變化與區域逆機率差異矩的混合邊緣特徵表示法來計算影片間的相似度。圖 3-1 為我們所設計創新的 temporal edge pattern descriptor 產生過程。產生步驟包含(a) Edge Detection; (b) 計算 Gradient Energy Histogram; (c) Dense Region Extraction; (d) 計算 HDVS Variance。

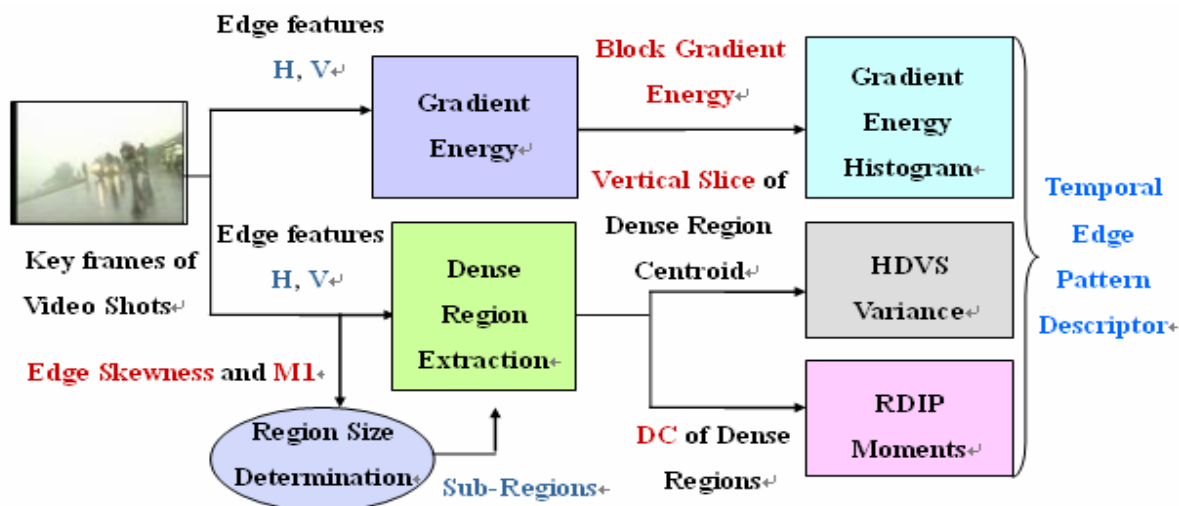


圖 3-1、temporal edge pattern descriptor 產生

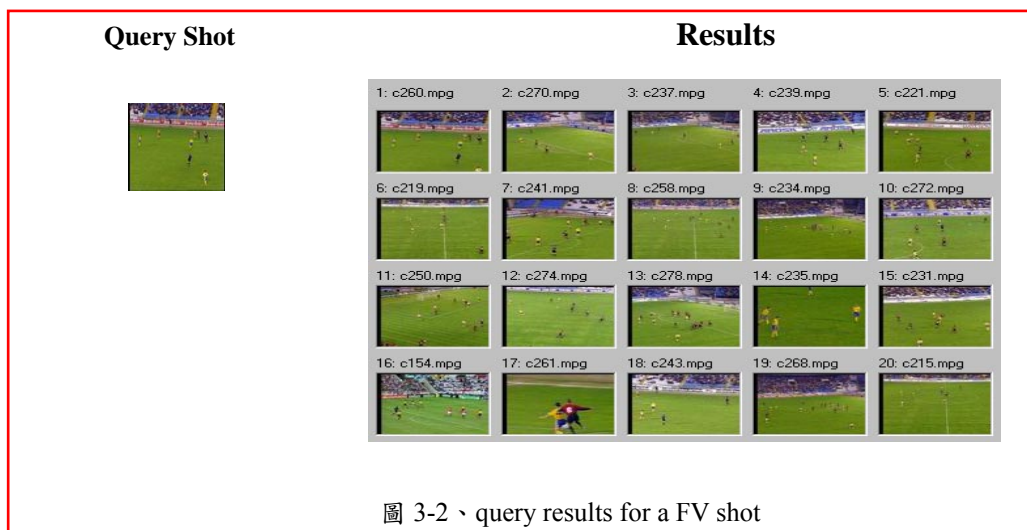


圖 3-2、query results for a FV shot

從影片查詢實驗結果可以看出，不管全景影片(如圖 3-2 所示)或特寫主播片段，我們所提出用來查詢相似影片的特 temporal edge pattern descriptor，以及設計的視訊資料庫查詢架構相當實用。查詢各種片段所得到相似影片的結果非常好，可達到接近 90%的準確率。

4. MPEG 多媒體傳輸機制及通訊協定在嵌入式行動平台上的分析設計——蔡淳仁教授

本子計畫除了延續第一年的工作，協助 MPEG 標準制定(Test Bed for MPEG-21 Resource Delivery 已正式成為國際標準 ISO/IEC TR 21000-12)，也根據這個 test bed 的架構，進

行了兩個新的研究方向。首先是設計了一套 streaming 的流量及容錯控制機制，能夠把傳輸頻寬和封包丟失的影響考量在系統中，以到碼率最佳化 (Rate-Distortion Optimized) 的串流傳輸效果。另一個研究的方向則是設計了以 wavelet 為基礎之 scalable video codec 的流量控制機制。這部份的研究主要是在給定頻寬要求的前題下達到碼率失真最佳化的位元流(bitstream)切割，以供後端的串流傳輸系統進行封包包裝及傳輸的工作。而這邊所設計出來的流量控制(rate control)機制和其它系統的主要改進在於我們加入了 multiple adaptation 的考量，這對未來嵌入式行動多媒體系統 (如 P2P 多媒體傳輸) 的應用將有十分大的幫助。

(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。

為了在頻寬和封包漏失率隨時在改變的環境下給使用者最好的串流傳輸品質，許多人嘗試對傳輸的資料進行碼率失真最佳化(Rate-Distortion)分析，然後根據分析結果來傳輸視訊串流。過去所提出的碼率失真最佳化的網路傳輸的流量控制機制裏，最大的問題在於封包漏失所造成的失真(distortion)沒有辦法量化。許多現有系統的做法在封包漏失所造成的失真的估測都十分不切實際 (或完全忽略了)。我們在本年度的研究中提出了一個十分有創見的方法，可以合理的把封包失真量化。

要達到碼率失真最佳化的網路串流傳輸，除了要根據網路頻寬和封包漏失率來估測出最好的串流資料量，還必須能在算出的串流資料量限制下，從 scalable video 的位元流中，抽出最好的子位元流。所以我們針對這方面的需求，設計出了一個快速收斂的小波轉換視訊壓縮法(wavelet video coding)的流量控制機制，另外，更進一步設計了可以有效進行多次碼率最佳化的子位元流抽取的機制。

我們所設計出的碼率最佳化串流傳輸機制，也在總計畫為 MPEG 所設計出的多媒體傳輸共通測試平台上實作測試。目前此平台已經成為 MPEG 國際標準(ISO/IEC JTC 1/SC 29 Part 12: Test Bed for MPEG-21 Resource Delivery, 2004)。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。詳細的架構請參考上述標準文獻。

(2) 結果與討論

本計畫主要的重點在於設計以 MPEG 技術為主的多媒體串流傳輸系統。在流量控制和容錯機制的設計方面，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 ("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。而且這個系統有兩大缺點。首先是 Chou 使用封包漏失率來代表碼率失真最佳化分析中的失真。這是很不實際的做法。其次，他用來降低失真的方法則是預先重傳封包 (非 ARQ)，這也是很沒有效率的。

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。

在本計畫的碼率失真最佳化串流傳輸系統中，我們把封包漏失所造成的失真，轉化為不同程度的 FEC 保護所造成的失真。舉例而言， 10^{-3} 的封包漏失率造成的失真，就相當於 10^{-3} 的 FEC 的 error protection 導致 data rate 降低所造成的失真。整個系統可以分成兩大部份：

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. 碼率失真最佳化傳輸控制：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碼率及失真的 Lagrangian cost function 來算出最小值的解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，決定每一個封包的 interleaving FEC 的保護程度。而這個程度則是依據此一封包的截止期限、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、以及這段資料本身的 source coding 的碼率失真分析來一起進行評估。(詳見下節)
3. 在快速收斂的小波轉換視訊壓縮法的流量控制機制方面，有別於一般系統慣用的查表法，我們採用的技巧是設計了一個更有效的雙參數視訊資料 R-D model 來進更最佳的子位元流切割點的快速搜尋。也因為我們所提出的視訊資料 R-D model 更精簡也更有效率，我們可以把它隨著抽取出的位元流一起傳輸到接收端以進行再一次的碼率失真最佳化子位元流切割。

(3) Adaptive Interleaved Forward Error Correction 系統與實驗

We have added Reed-Solomon coding modules, an interleaver module, and a de-interleaver module to the original MPEG-21 Multimedia test bed. The system diagram is shown in Fig. 4-1. In the experiments, the CIF version of the FOREMAN sequence is used. The sequence is encoded using ISO/IEC 14496 (MPEG-4) visual reference software (Microsoft-FDAM1-2.5-040207) at 10 frames per second. The coding mode is one I-frame followed by nine P-frames at 10 frames per second.

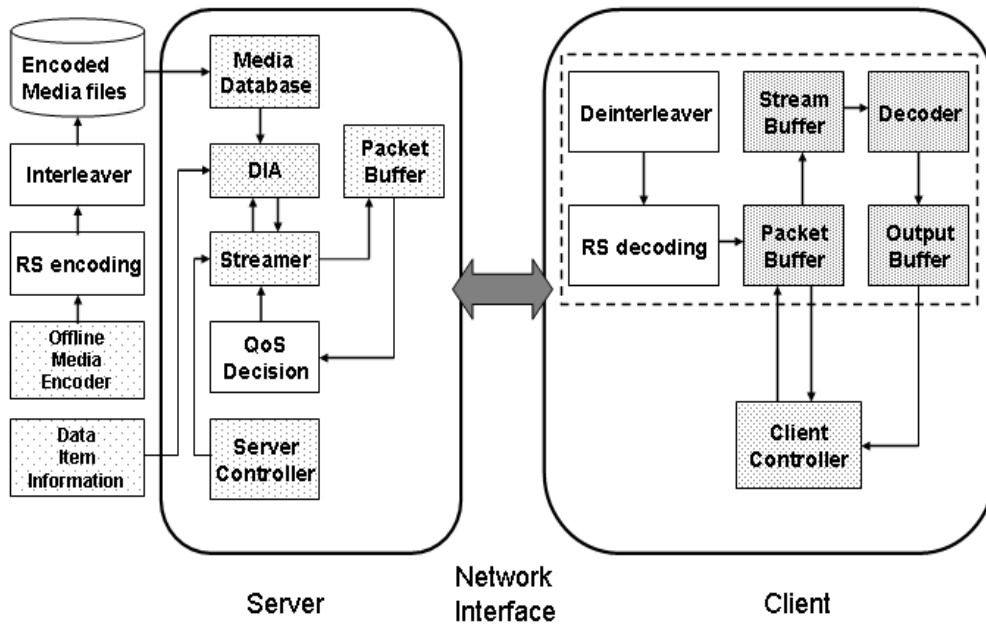


Fig. 4-1. Architecture of the proposed system

Figure 4-2(a) presents the PSNR performance of the streaming system under a variable bandwidth scenario, ranging from 68kbps to 240kbps. When the transmission rate fluctuates significantly, e.g., from the 40th to the 80th frame, the proposed system can adapt the bandwidth changing quickly and reduce the degree of quality variation. Fig. 4-2(b) shows the bit rate variation of the FEC protected base-layer and enhancement-layer bitstream under the variable bandwidth condition. For example, when the transmission rate changes from 180kbps to 116kbps, the system performs dynamic rate allocation to add more FEC protection on the base-layer. Consequently, most of the enhancement-layer bitstream is truncated during the period.

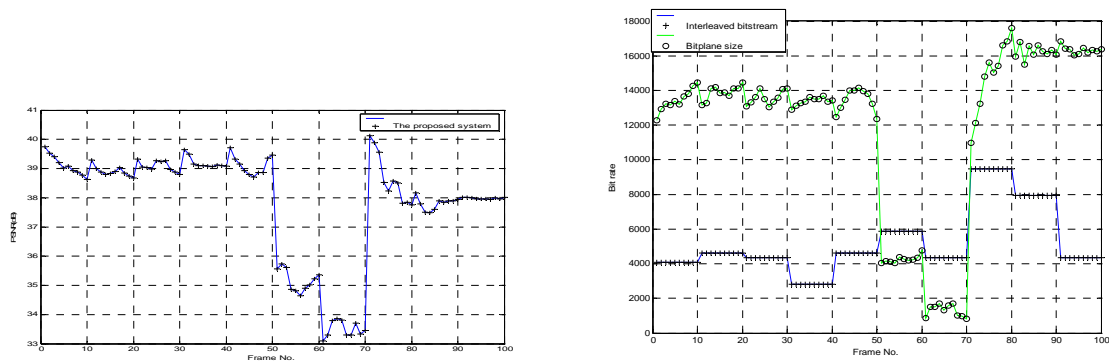


Fig. 4-2. (a) PSNR under variable transmission rates. (b) Bitrates for the FEC-protected base-layer and the enhancement-layer bitstreams.

5. 高等精細可調層次式視頻編碼技術之研究 —— 蔣迪豪教授

在 MPEG 可調視訊編碼 (Scalable Video Coding) 標準的規格要求中，感興趣區域 Regions-of-interest (ROI) 的品質加強是一個重要的功能。對於可調視訊編碼的應用而言，解碼端希望在增強層遺失時能夠在感興趣的區域中提供較好的品質。然而，目前可調視訊編碼標準的參照軟體 JSVM1.0 並沒有提供此一功能。因此在本計劃我們開發了一個可提供任意形狀感興趣區域和優雅品質調整的演算法。

(1) Background

The ROI functionality is commonly supported in scalable video/image coding. In MPEG-4 FGS, the ROI functionality is enabled by a selective bit-plane shifting scheme. Different from MPEG-4 FGS, in SVC the bit-planes are replaced with FGS layers which are produced by successive quantization of inter-layer prediction residue. Each FGS layer can be deemed as a group of multiple bit-planes. However, these bit-planes are coded by a cyclical block coding instead of traditional bit-plane coding. In SVC, the coding of a FGS layer is partitioned into the significant and refinement passes. The significant pass first encodes the insignificant coefficients which show zero values in the subordinate layers. After that, the refinement pass refines the remaining significant coefficients ranging from -1 to +1. Particularly, the significance pass coding is performed in a cyclical manner. On the other hand, the refinement pass coding is conducted subband by subband.

(2) Method -- Prioritized Cyclical Block Coding

To provide the ROI functionality based on cyclical block coding, we proposed to code each block unequally in a coding cycle. For illustration, Fig. 5-1 and Fig. 5-2 show the differences between the cyclical block coding and our prioritized coding scheme. For simplicity, we use the notation of (EOB, Run, Level) symbol to represent the coefficients of a block that are to be coded in a coding cycle of significance pass. In Fig. 5-1, the cyclical block coding equally encodes each block with one (EOB, Run, Level) symbol (or one refinement symbol) in every cycle. However, as shown in Fig. 5-2, to offer the ROI functionality, we should encode the blocks in the ROIs with more symbols by enabling the coding prior to the blocks which are outside the ROIs. We can further extend this concept to enable the coding of different ROIs at different cycles to have graceful selective enhancement. For example in Fig. 5-2, we have selected two ROIs. The coding of the highest priority region ROI 1, which includes Block 0, is enabled prior to the rest of blocks. After two coding cycles, the coding of ROI 2 which is of lower priority is activated subsequently. In particular, the coding of ROI 2 is started before the coding of background region, i.e., Block 2 and Block 3. With such prioritization, the blocks in the ROIs will be coded with more symbols at the end of each cycle. When the enhancement-layer is truncated, the blocks in the ROIs will be firstly decoded and updated.

For specifying the priority of a ROI, we use the number of shifting cycles which is defined as the number of coding cycles in the ROI before the coding of background region. For example, the number of shifting cycles of ROI 1 in Fig. 5-2 is 2 which means that the Block 0 has to be coded with 2 cycles before the coding of background region. Higher number of shifting cycles means higher coding priority.

(3) Experiment

For the experiment, we implement our prioritized cyclical block coding based on JSVM1.0. Specifically, we use the Foreman sequence as an example to demonstrate the functionality of ROI. We compare the regional PSNR and the subjective quality between the schemes with ROI and without ROI.

Fig. 5-3 shows the regional PSNR comparison of different ROI coding modes. The legend SEN denotes the performance of our prioritized cyclical block coding. Specifically, the num-

ber n represents the foreground shifting factors. In addition, SE_n _Remap shows the performance with both the prioritized coding and layer remapping. As shown, enabling the ROI functionality can dramatically increase the PSNR of foreground region. Without the layer remapping, the PSNR of foreground region is the same as the original cyclical block coding at the end of FGS layer 1 (around the bit rate of 700kbits/s). However, with the layer remapping, we can have ROI functionality over the entire bit rate ranges. Particularly, the curve of JSVM1_Remap shows that simply enabling the layer remapping without our prioritized cyclical block coding is not sufficient to offer the ROI functionality over the entire bit rate range.

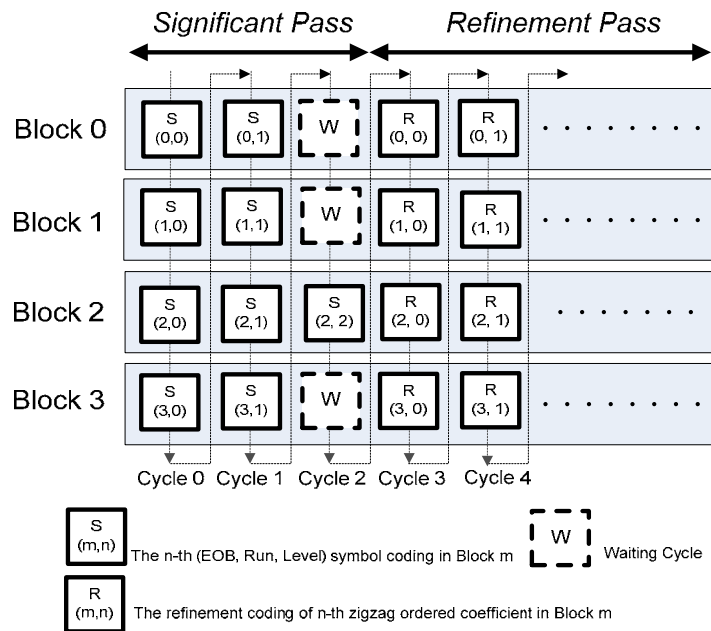


Figure 5-1: Cyclical block coding.

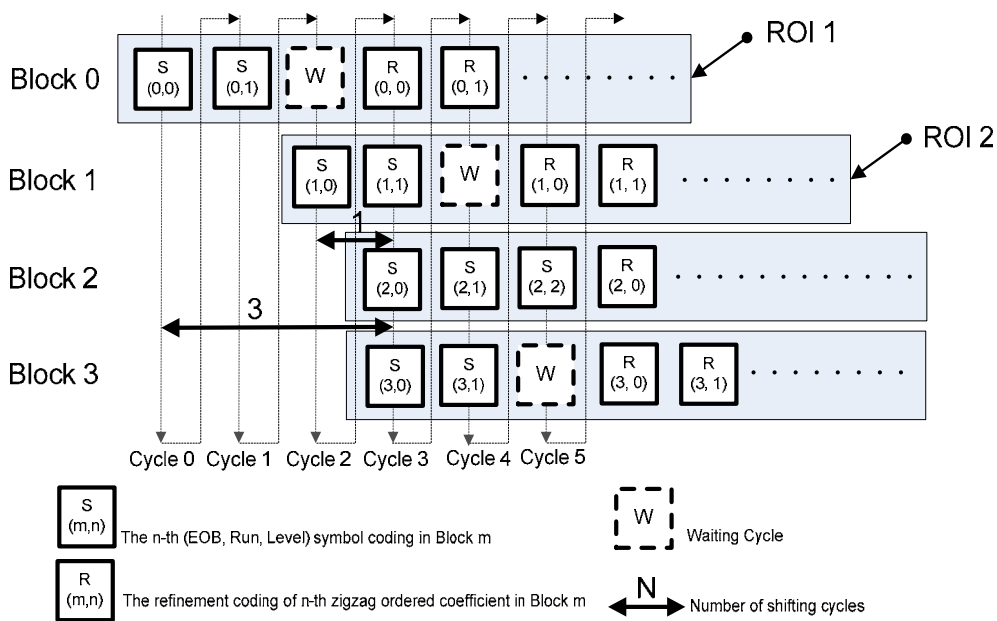


Figure 5-2: Prioritized cyclical block coding for the ROI functionality.

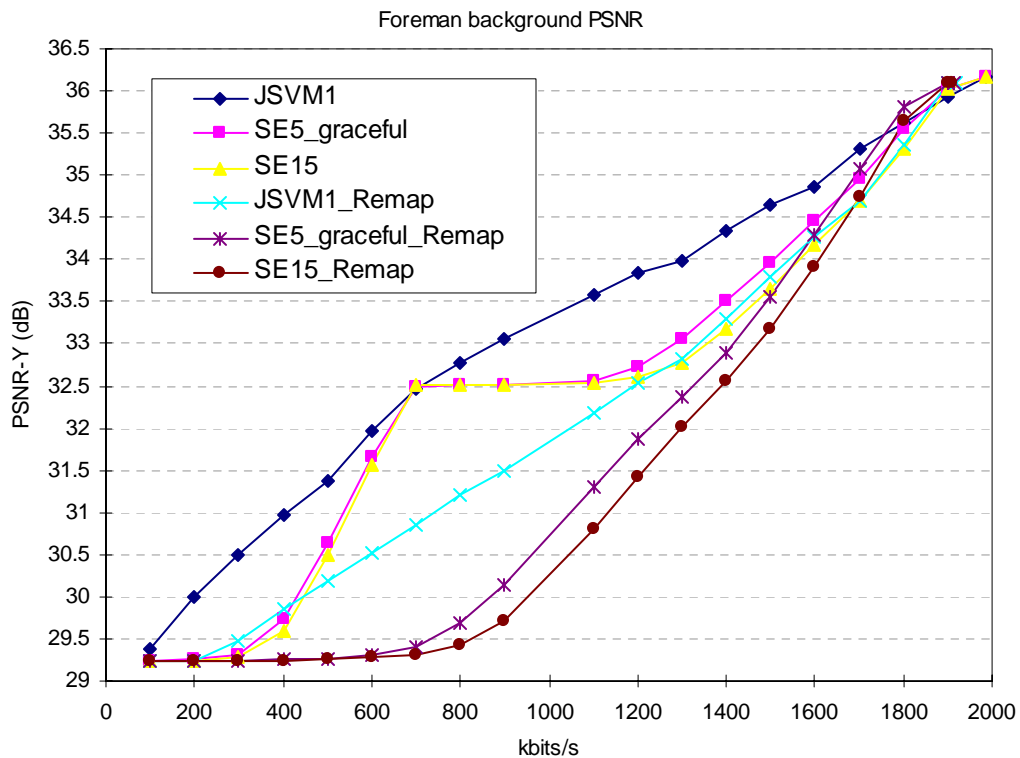
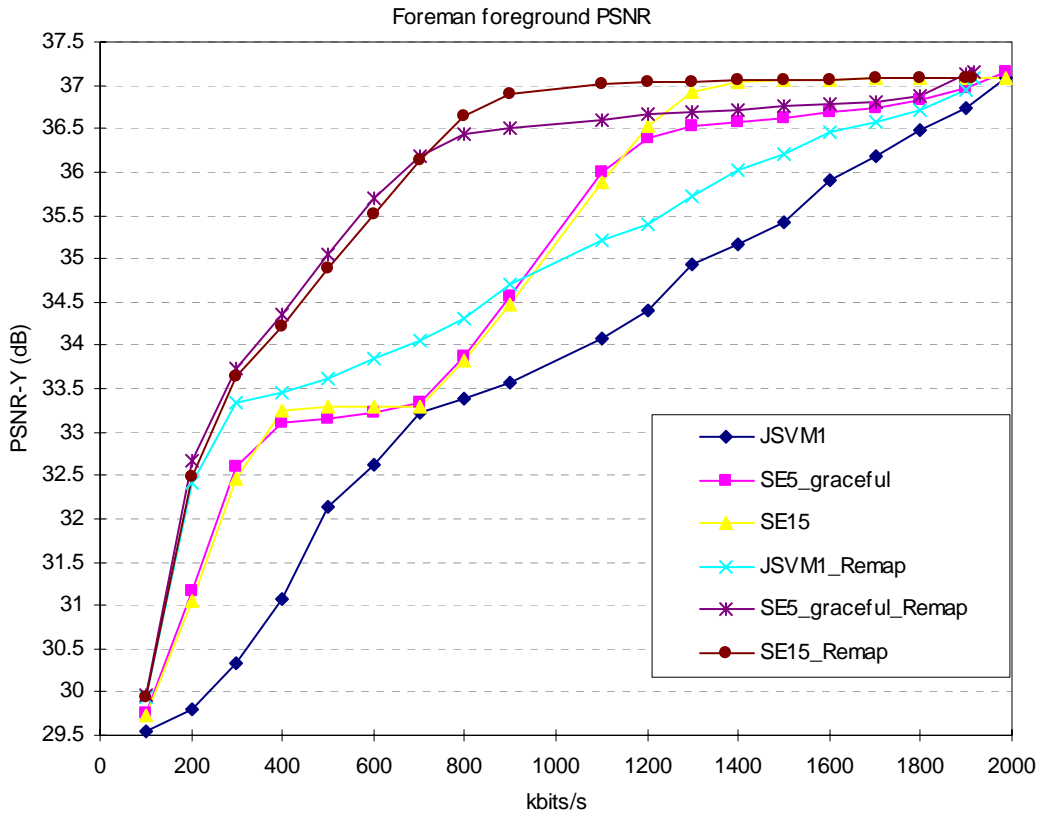


Figure 5-3: Regional PSNR comparison.

(4) Conclusions

In this project, we use the cyclical block coding in JSVM1.0 to develop a graceful and arbitrary shape region-of-interest (ROI) functionality for SVC. To enable the ROI functionality,

we additionally introduce 2 syntax in the FGS slice header, 3 syntax at the FGS layer level, and one syntax at MB level. Experimental results show that the proposed ROI functionality can significantly improve the subjective quality while maintaining the coding efficiency and scalability. Thus, we suggest the group to look into this technology and consider including ROI functionality in the SVC.

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

畫面間小波視訊轉換(Interframe Wavelet Video Coding)由於擁有良好的壓縮比以及多元的可調性近來備受矚目。它可以做到三種不同的可調性：1)資料量的可調整性、2)時間解析的可調整性、3)空間解析的可調整性。畫面間小波視訊轉換初步研究成果在 2004 年 3 月與 7 月提案參加於 MPEG 標準組織 scalable video coding Call-for-Proposal 競賽，之後仍參與 MPEG Ad Hoc Group 進行討論。2005 年我們持續提案參加 MPEG 標準組織 scalable video coding 的 Core Experiments。在本計畫中，主要有兩個研究主題 (1)以人類視覺系統為基礎的位元率控制法，(2)用於影像壓縮之方向性多重解析度轉換，和(3)區塊位元平面算數編碼。計畫目標為針對畫面間小波視訊轉換之空間轉換、熵編碼器、與位元率控制等進行改良。

A. 以人類視覺系統為基礎的位元控制法

(1) 背景與研究目的

在畫面間小波視訊轉換(Interframe Wavelet Video Coding)位元控制法(rate control algorithm)中，每個在用於畫面之間的小波轉換編碼的截斷點(truncation point)都有自己相關聯的失真(distortion)和位元長度(bits length)。而每個截斷點的斜率(slope)就是把失真的差異(distortion difference)除以位元差異(bit difference)所得到的商。在最佳化理論中(optimization theory)，擁有較高斜率的截斷點有較高的優先權被傳送。在這裡我們提出一個方法，就是說我們把每個截斷點的斜率乘上一個由人類視覺系統算出來的權重。故這個經過視覺加重的斜率會成為位元控制法中判斷的標準。我們的模擬會指出最後的重建影像有較低的最高訊號雜訊比(PSNR)和較佳的視覺品質。

(2) 方法與實驗

正好可被注意到的失真(Just Noticeable Distortion) -- 把一張影像經過小波轉換(wavelet transform)後，每個次頻帶可用層次(level) λ 以及方位(orientation) θ 表示，而每個次頻帶的亮度成份(luminance component)的正好可被注意到的失真(just noticeable distortion) y 可以用 $\log(y) = \log(a) + k \cdot (\log(f) - \log(g_\theta f_0))$ 來算出，其中 $a = 0.495$ ， $k = 0.466$ ， $f_0 = 0.401$ 。 g_θ 為 1.501，1，或 0.534 分別對應到 LL，LH/HL，或 HH 次頻帶。 f 是空間上的頻率(spatial frequency)且在不同的狀況下有不同的數值。在一般看電腦螢幕的狀況下，展示解析度(display resolution) r 為 16 像素/度，而每一層次 λ 的小波轉換所對應的空間上的頻率 f 為 $f = r * 2^{-\lambda}$ 。

(3) 結果與討論

依照正好可被注意到的失真及其他視覺特性，我們可推導出一個次頻帶內所有截斷點的視覺權重要素。然後依據截斷點加權後失真和位元關係，決定其位元分配。從圖 6-1 可

以看出，乘上權重後的影像在視覺效果比原本的好，尤其是在大面積的平坦的區域上可以另其看起來更為平滑；但有時候會另影像中的線條的部份更模糊，尤其是在較低的傳輸速率上，因為原本的權重就是把低頻的權重加大，把高頻的權重減輕，原因在於人眼對低頻的訊號比高頻的訊號敏感。



(a)原本, PSNR=31.19dB, 位元流大小 4371 位元組; (b)權重, PSNR=30.98dB, 位元流大小 4317 位元組

圖 6-1：測試影像一在每秒 1000K 個位元的傳輸速率

B. 用於影像壓縮之方向性多重解析度轉換

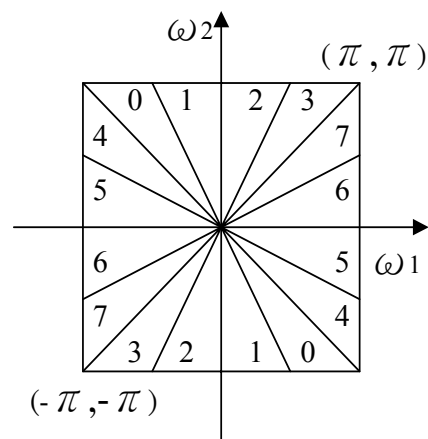
(1) 背景與研究目的

小波轉換在一維方向有不錯的趨近特性，所以在訊號處理的領域中常常被使用，然而在二維方向的表現卻不如一維方向。二維分別(separable)小波轉換雖然可以處理點的變化，但對於線和曲線的變化卻不能發揮，因此，Candes and Donoho 發表了新的空間分析方法，叫做曲線轉換 (curvelet)。其後 Minh N. Do 提出了輪廓轉換(contourlet)建立新的圖片表示方法。我們的研究用 Minh N. Do 的輪廓轉換放入 MPEG Wavelet Coding 參考軟體中，並且取代掉原本壓縮器常使用的小波轉換，來檢視輪廓轉換是否能夠彌補原本小波轉換不足之處。

(2) 方法與實驗

二維方向性濾波頻帶基本上會把頻譜分解成像風扇的樣子如右圖所示，每一個分解區域會對應到一個次頻帶，在實現的過程中我們會使用樹狀架構，頻帶分解的個數是依照此樹狀架構的層數來決定。

輪廓轉換包含方向性和多重解析度的分解，多重解析度的分解先將圖片分成高頻成分和低頻成分圖片，多重解析度分解可以一直遞迴應用在低頻部分圖片，經過多重解析度分解後產生的高頻成分圖片會接著做方向性分解。前述的兩項分解在實作時主要是使用拉氏金字塔分解和二維方向性濾波頻帶，因此在研究步驟上要先建立這兩項分解的模組。



(3) 結果與討論

在方向性多重解析度轉換研究方面，表 6-1 測試圖片 barbara 的壓縮結果比較，MSSVC(Microsoft 的 MPEG Wavelet Coding Software)為原參考軟體的結果，MSSVC_MDT 為加入輪廓轉換後的結果。由於 JPEG2000 大家熟悉的壓縮方法，我們也將 JPEG2000 包括進來當作 PSNR 比較的參考。

表 6-1：測試圖片 barbara 的 PSNR 比較表

ratio(%)	PSNR		
	MSSVC	JPEG2000	MSSVC_MDT
0.625	21.41	22.44	22.43
0.9375	23	23.33	22.8
1.25	23.89	23.97	23.86
5	30.85	29.56	28.18
10	35.5	33.97	31.89

C. 區塊位元平面算數編碼

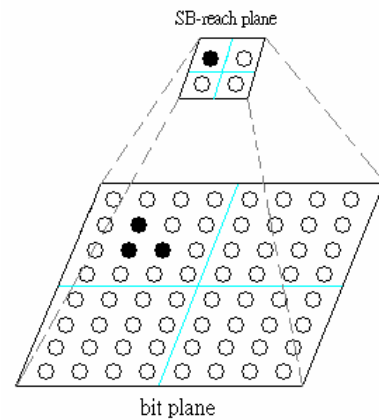
(1) 背景與研究目的

在熵編碼方面，由於參考軟體所使用的三維嵌入式塊最佳截斷編碼沒有利用到能量聚集的現象，因此我們希望能藉著位元平面上能量聚集的現象讓編碼效率能夠提升。

(2) 方法與實驗

小波係數之初始設定為『非有效性的』(insignificant)，當係數的第一個非零位元被找到時，就將其設定為『有效』(significant)，此第一個非零 significant 位元即是係數的『有效位元』(Significant Bit, SB)，我們在每一個位元面建立另一個位元面叫做有效位元整併面

(SB-reach plane)，如右圖所示。若在此位元面上的一個數值代表其 $N \times N$ 平方區塊的係數位元面。當對應區塊越大有效位元整併平面越小，當對應區塊中有一個係數變成 significant，有效位元整併面(SB-reach plane)的數值會被設為 1，反之，當整併面的參數為 0 代表其相對應區塊全部為 0。



編碼完整併面後，我們對原來的位元面編碼。如果之前整併面的數值為 0，其位元面上相對應區塊就都不需要編碼，若整併面上的參數為 1，其位元面編碼過程在**錯誤! 找不到參照來源**。說明。位元面的編碼步驟和核心實驗軟體中是一樣的，

(3) 結果與討論

在熵編碼方面，我們提出有效位元整併方法(SB-reach plane)來提升三維嵌入式塊最佳截斷編碼，位元率的節省百分比在**錯誤! 找不到參照來源**。到**錯誤! 找不到參照來源**。，在這些表顯示從第一層位元平面到目前指定位元面累加節省的總位元數。舉例來說，累加到第三層位元面表示第一、第二和第三位元面節省位元數，而若為負百分比表示位元率比原來的的方法還要多，表上也顯示出空間上每個次頻帶的節省位元數。在低位元比率時，在第一、第二有較明顯節省位元減少比例。

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，但也探討系統的整合。我們擬想中的系統，是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與會人員的合成影像。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，然後併同聲訊傳到其他端點。每一端點也需將所有接收到的視訊及聲訊予以解碼及合成。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。其中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。本計畫原定以三年時間進行研究。本報告係針對第二年之研究，其中著重視訊分割技術、傳送端系統整合技術、及接收端系統整合技術。在視訊分割方面，我們提出了一些演算法。在傳送端系統整合方面，我們將過去已實現的視訊分割法、加速後的 MPEG-4 視訊編碼程式、一個 MPEG-4 聲訊編碼程式、以及一個 RTP 網路傳輸程式，在個人電腦上加以整合，形成了一個傳送端系統的雛型。在接收端系統整合方面，我們則將一個 RTP 網路傳輸程式、一個 MPEG-4 視訊解碼程式、以及一個 MPEG-4 聲訊解碼程式，在個人電腦上加以整合，並設計了一個多視訊結合的方法，將以上組合成了一個接收端系統的雛型。

(1) 緣由與目的

本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。本計畫擬想中之視訊會議系統，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可考慮由一位與會者統一安排。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)。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)。

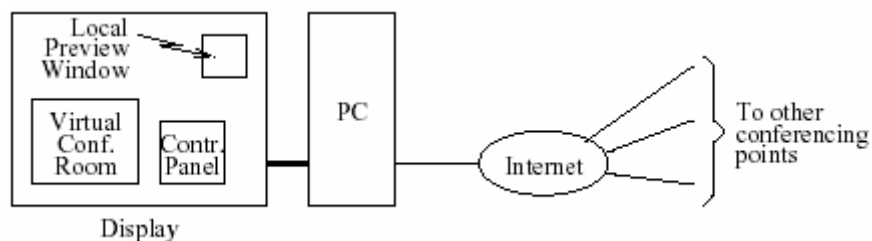


圖 7-1：多點視訊會議系統架構示意圖

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，原定是以三年時間進行相關研究。本報告係針對第二年之研究，其中重點在第三群組中的視訊分割技術，以及第一群組的傳送端系統整合技術與接收端系統整合技術。在視訊分割方面，我們提出了一些演算法。在傳送端系統整合方面，我們將過去已實現的視訊分割法、加速後的 MPEG-4 視訊編碼程式、一個 MPEG-4 聲訊編碼程式、以及一個 RTP 網路傳輸程式，在個人電腦上加以整合，形成了一個傳送端系統的雛型。在接收端系統整合方面，我們則將一個 RTP 網路傳輸程式、一個 MPEG-4 視訊解碼程式、以及一個 MPEG-4 聲訊解碼程式，在個人電腦上加以整合，並設計了一個多視訊結合的方法，將以上組合成了一個接收端系統的雛型。

(2) 結果與討論

a. 視訊分割演算法之進一步研究與結果

視訊分割有兩大核心議題，一是物件邊界之精確界定，二是運算複雜度。一個常見的設計是先透過時空域分析來獲取有興趣之物件的大略位置與形狀，然後試著對物件的邊界做更精確的界定。關於後者，有幾個常見得作法，分別是曲線演化(contour evolution)、分水嶺分析(watershed analysis)、及銳緣連結(edge linking)。我們考慮最後這種途徑，並提出幾個有效率又相當準確的方法。

第一個方法是建立在兩個簡單的想法上。想法之一是使用物件已知的大略位置與形狀來減少需要搜尋的像素區域，之二是有效地使用搜尋區域內所已經偵測到的銳緣。圖 7-2 顯示這個方法的程序。其中 CDM Round-Out 是檢視上述已經大略定出位置及形狀的物件，其中若有中空之處，或是有一二個像素寬的隙縫，就將之填滿，以形成一個實心體。Boundary Tightening 是假設實心體最靠外的銳緣就是物件的真實邊界，因此將實心體的外圍內縮，盡量刪去這些靠外的銳緣之外的像素。但是，使用銳緣偵測方法所找到的銳緣，常有斷裂之處，不能完整定義物件邊界，所以還要經過 Shortest-Path Search 來把斷裂的地方接起來。Shortest-Path Search 的複雜度和所需搜尋的範圍有關。經過前面兩步驟的處理，可以降低搜尋範圍，所以可以降低複雜度。實驗結果也顯示以上方法可以獲得相當準確的物件邊界。

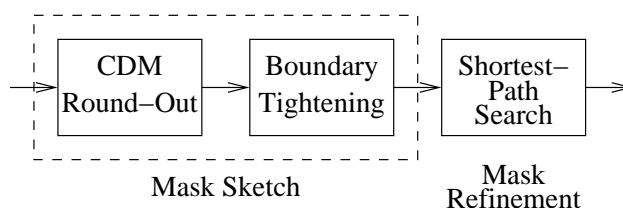


Fig. 7-2: One of the proposed methods for refinement of video segmentation.

第二個方法也是採用銳緣連結，但是用了一個自創的、比上述方法中的 Shortest-Path Search 更簡化的銳緣連結方法。實驗結果顯示，在一個具備 1.8 GHz Pentium CPU 的個人電腦上，使用一個尚未優化的程式，對 CIF 視訊做分割，平均每畫面只需約 30-40 ms。因此本方法適用於即時桌上型視訊會議。

b. 傳送端系統整合技術之研究與成果

如前述，我們考慮建構一個視訊會議系統，其中與會者的影像在接收端要在解碼後合成一個虛擬的會議場景。因此，我們需要在各個傳送端將當地與會者的影像分割出來，然後編碼傳送出去。此外，各傳送端也要將當地的音訊加以編碼送出。MPEG-4 的若干規範，相當適合這樣的應用。我們所建構的傳送端雛形系統結構如圖 7-3 所示。使用一個具備 Intel Centrino Pentium M 1.5 GHz CPU 及 512 MB DDR RAM 的個人電腦，在 Microsoft Windows XP Professional Version 2002 作業系統下所作的實驗結果，顯示我們所建構之尚未優化的程式可以達到每秒約 10.7 張 CIF 畫面的處理速度。

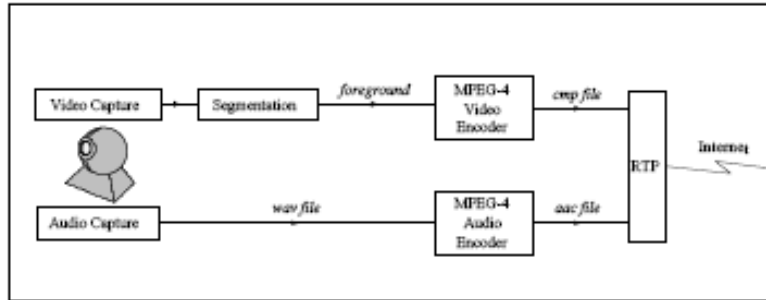


Fig.7-3: Architecture of videoconference transmitter system.

c. 接收端系統整合技術之研究與成果

本節所述的研究，是與上節互相配合，但分別進行。在上述視訊會議系統的接收端，需要將各地傳來的視訊和音訊個別加以解碼，然後組合並輸出。其中視訊畫面的組合很費工夫，也是本研究一個重點。我們所設計的系統，架構如圖 7-4 所示。為了簡單，我們暫時假設所有視訊的訊框率(frame rate)皆相同。(這不是太嚴重的限制。)我們在程式中，任意令一個接收到的視訊主控解碼與輸出的時序。因為音視訊來源的數目可能會隨時間而變化，主控的視訊也可能退出會議。所以我們在程式中設計了一個簡單的機制來處理這個情況，讓程式可以任意選用令一個存在的視訊取得主控地位。實驗顯示，系統的處理速度在只有一個 CIF 視訊時約為每秒 11.2 畫面，二個視訊時約每秒 5.2 畫面，三個視訊時則約每秒 3.6 畫面。因此，處理速度約與視訊數目成反比。

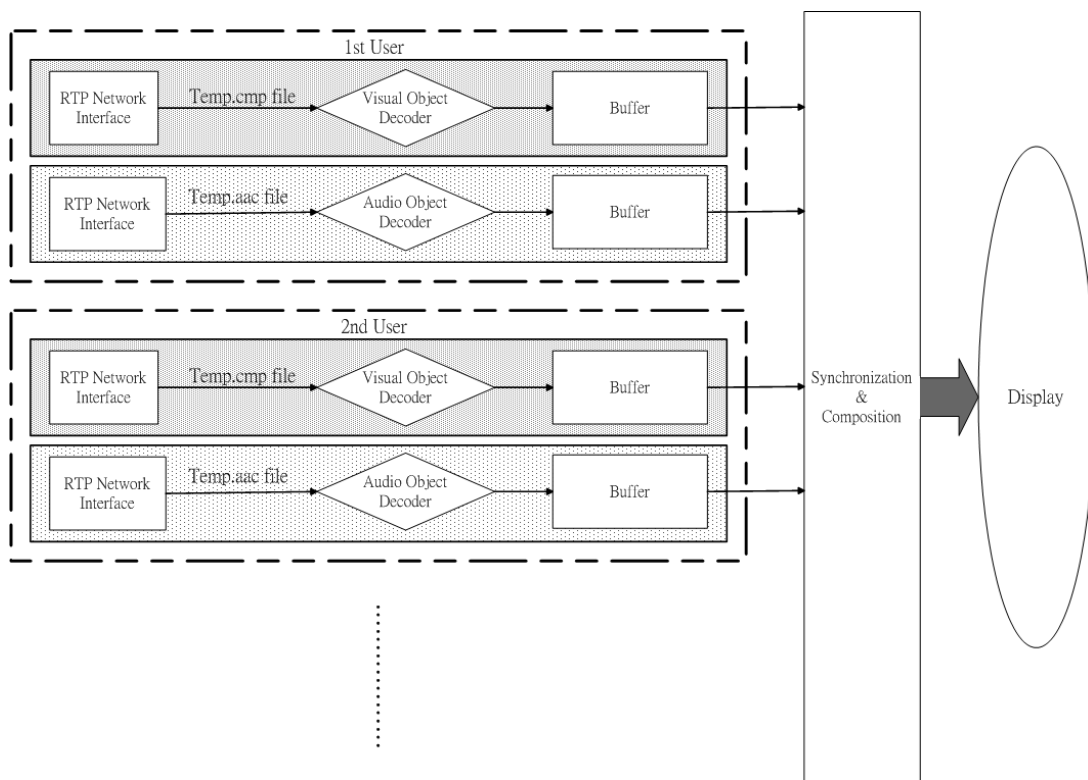


Fig. 7-4: Structure of the proposed videoconference receiver.

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、蔡淳仁及博士生彭文孝，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 9 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 之共同編輯者(editor)。交大 MPEG-21 團隊之成員人數約在 10 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 2004 年被 MPEG 標準委員會接受為技術報告類標準文件: 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.

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 亦已完成，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，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，交通大學團隊提出一項改善 Interframe Wavelet 提案，Response to CE1b in SVC-- SB-Reach Method for Entropy Coding。2005 年 4 月 MPEG 的第 72 次會議，交通大學提出另一項改善 Part 10 Scalable Video Coding (SVC)提案，Adding Selective Enhancement Functionality to Scalable Video Coding，可加強 SVC 的 Region-of-Interests (RoI) 能力。

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。

2003.8 – 2004.6 本研究群發表之 MPEG 標準文件

1. F.-C. Chang, C.-P. Ho, C.-Y. Tsai, C.-H. Li, J.-D. Cheng, W.-C. Chang, Y.-T. Shih, C.-L. Lin, C.-H. Lu, C.-C. Cheng, C.-Y. Liu, J.-C. Ma, K.-C. Lee, and C.-J. Tsai, "ISO/IEC JTC1/SC29/WG11 M12373: Update to the FGS-Based Multimedia Resource Delivery

- Test Bed Software”, July 2005 (73rd, Poznan, Poland)
2. C.-J. Tsai, “ISO/IEC JTC1/SC29/WG11 M12074: Suggestions on the direction of VCTR”, April 2005 (72nd, Busan, Korea)
 3. T. Chiang, Y-S. Tung, C.-N. Wang, “ISO/IEC JTC1/SC29/WG11 M11785: AHG on Maintenance of MPEG-4 Visual related Documents, Reference Software and Conformance”, April 2005 (72nd, Busan, Korea)
 4. W.-H. Peng, Tihao Chiang and H.-M. Hang, “ISO/IEC JTC1/SC29/WG11 M11914: Adding Selective Enhancement Functionality to Scalable Video Coding,” April 2005 (72nd, Busan, Korea)
 5. T. Chiang, Y-S. Tung, C.-N. Wang, “ISO/IEC JTC1/SC29/WG11 M11475: AHG on Maintenance of MPEG-4 Visual related Documents, Reference Software and Conformance”, Jan 2005 (71th, Hong Kong, China)
 6. H.-M. Hang, “ISO/IEC JTC1/SC29/WG11 M11482: AHG Report on Further Exploration in Wavelet Video Coding,” January 2005 (71st, Hong Kong, China)
 7. Y.-S. Wang, C.-N. Wang, T. Chiang and J.-R. Ohm, “ISO/IEC JTC1/SC29/WG11 N6710, MPEG-4 Visual: List of Problems Reported, v.14”, Oct 2004 (70th, Palma de Mallorca, Spain)
 8. T. Chiang, Y-S. Tung, C.-N. Wang, “ISO/IEC JTC1/SC29/WG11 M11156: AHG on Maintenance of MPEG-4 Visual related Documents, Reference Software and Conformance”, Oct 2004 (70th, Palma de Mallorca, Spain)
 9. Y.-S. Tung, C.-N. Wang, T. Chiang, J.-R. Ohm, “ISO/IEC JTC1/SC29/WG11 M11327: MPEG-4 Visual: List of Problems Reported, v.13.1”, Oct 2004 (70th, Palma de Mallorca, Spain)