

行政院國家科學委員會專題研究計畫 成果報告

總計畫(3/3)

計畫類別：整合型計畫

計畫編號：NSC91-2219-E-009-040-

執行期間：91年08月01日至92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

共同主持人：林大衛，蔣迪豪，王聖智，張文鐘

計畫參與人員：余子瀚，陳俊安，汪秉玉，陳宜妙，王俊能

報告類型：完整報告

處理方式：本計畫可公開查詢

中 華 民 國 92 年 10 月 23 日

MPEG-4 多媒體通訊技術之研究--總計畫(3/3)
MPEG-4 Multimedia Technology (3/3)

計畫類別： 個別型計畫 整合型計畫

計畫編號：NSC 91-2219-E-009-040

執行期間： 91年8月1日至92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
林大衛 國立交通大學電子工程學系 教授
張文鐘 國立交通大學電信工程學系 教授
蔣迪豪 國立交通大學電子工程學系 副教授

計畫參與人員：余子瀚，陳俊安，汪秉玉，陳宜妙，王俊能

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 15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--總計畫(3/3)

MPEG-4 Multimedia Technology (3/3)

計畫編號：NSC 91-2219-E-009-040

執行期限：91 年 8 月 1 日至 92 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, 張文鐘, 林大衛, 蔣迪豪 國立交通大學

計畫參與人員：余子瀚, 陳俊安, 汪秉玉, 陳宜妙, 王俊能 國立交通大學

中文摘要

過去十五年間，多媒體通訊的研究以驚人的速度成長。本計畫的目標在於研究並實現國際標準 MPEG-4 中幾項重要的技術。

MPEG-4 標準頗為龐大複雜，本整合型計畫將就其中之視訊處理、視訊資料編碼與傳輸以及系統等三大部分予以研究，共分五項子計畫：

1. MPEG-4 相關之彩色影像切割技術研究
2. 視訊傳輸的互動迴授型式及錯誤防止
3. 比例式視訊編碼技術及視訊通訊終端機技術之研究
4. MPEG-4 與 MPEG-7 系統之研究
5. 多媒體架構與數位視訊浮水印在網際網路之應用

本報告將略敘第三年之各項成果。

關鍵詞：多媒體通訊, 視訊編碼, MPEG-4, MPEG-7, 多媒體資料庫, 影像切割

英文摘要

Researches in the area of multimedia communications have been advancing at an amazing speed in the past 15 years. The goal of this project is to study and/or implement several important technologies adopted in the MPEG4 standards.

The MPEG-4 standard is rather huge and complicated. This project focuses on three major parts of an MPEG-4 system: (1) video data processing, (2) video data coding/transmission and (3) system issues. We also extend our researches beyond MPEG-4. Particularly, the MPEG-7 and MPEG-21 systems are also investigated. This project consists of 5 sub-projects:

1. Study of MPEG4-related image segmentation techniques,
2. Interactive feedback message and error resilience in video transmission,
3. Research in scalable video coding techniques and visual communication terminal technologies,
4. A study on MPEG-4 and MPEG-7 systems, and
5. Multimedia framework and digital video watermarking for the Internet applications.

Keywords: Multimedia communication, Video coding, MPEG-4, MPEG-7, Multimedia database, Image segmentation.

目錄 Table of Contents

1. 背景與目的.....	4
2. MPEG-4 相關之彩色影像切割技術研究 -- 王聖智教授.....	4
3. 視訊傳輸的互動迴授型式及錯誤防止 -- 張文鐘教授.....	7
4. 比例式視訊編碼技術及視訊終端機技術之研究 -- 林大衛教授.....	10
5. MPEG-4 與 MPEG-7 系統之研究 -- 杭學鳴教授.....	14
6. 多媒體架構與數位視訊浮水印在網際網路之應用 -- 蔣迪豪教授.....	17
7. 計畫成果自評.....	20

1. 背景與目的

本整合型計畫“MPEG-4 多媒體通訊技術之研究”預定為期三年，今年為第三年。近年來，多媒體通訊 (multimedia communications) 有突飛猛進的發展，網際網路 (Internet) 以出乎大家預期的驚人速度開展。不論在國際標準及工業產品上，這幾年都有長足的進步。最近 MPEG4 標準已經成型，逐漸成為多媒體傳輸整體系統的重要標準。鑑於目前整體環境的改變，包括工商業的趨勢和學術產業的進展，我們將繼續研究多媒體通訊相關之技術，將針對 MPEG-4 以及演進中各項 MPEG 標準，進行相關的重要技術研究。

此“MPEG-4 多媒體通訊技術之研究”整合計畫的分工情形大概如下：

(一) MPEG-4 相關之彩色影像切割技術研究

- - 王聖智教授

(二) 視訊傳輸的互動迴授型式及錯誤防止 - - 張文鐘教授

(三) 比例式視訊編碼技術及視訊終端機技術之研究 - - 林大衛教授

(四) MPEG-4 與 MPEG-7 系統之研究

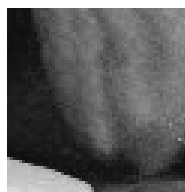
- - 杭學鳴教授

(五) 多媒體架構與數位視訊浮水印在網際網路之應用 - - 蔣迪豪教授

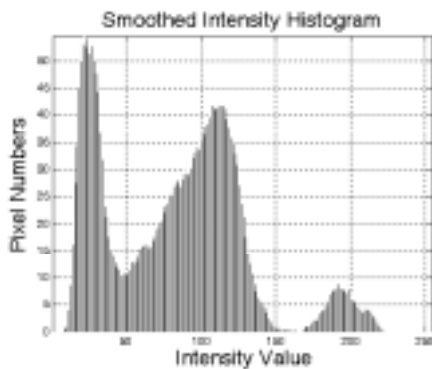
2. MPEG-4 相關之彩色影像切割技術研究 -- 王聖智教授

在本年度的計畫中，我們針對彩色影像切割，提出利用局部影像資料機率分佈圖之分析來偵測均勻區塊邊緣的概念，並依此發展了一套新的彩色影像切割演算法。整個研究包含三個部分：(1)像素可信任度的估測、(2)像素相互之間可信任度的測定以及(3)彩色影像切割的技術。在分析像素可信任度的估測方面，我們討論了單一像素屬於一個均勻區域的可信任度。依此進一步討論像素相互之間可信任度的測定。在影像切割演算法的發展方面，我們提出了一個完整的切割架構，並且得到理想的彩色影像切割結果。

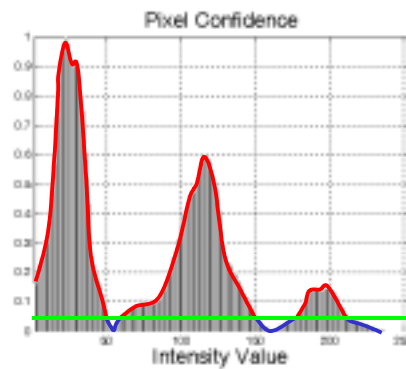
在這份報告中，我們定義了兩種可信任度：(1)像素的可信任度以及(2)像素相互之間的可信任度。像素的可信任度是在分析每個像素屬於一個色彩均勻區域的可信任度高低。另一方面，像素相互之間的可信任度是在測量兩相鄰像素屬於同一個均勻區域的可信度高低。



(a)



(b)



(c)

圖一、 (a) 影像，(b) 影像亮度機率分佈圖平滑化的結果，(c) 像素的可信任度

(1) 像素可信任度的估測

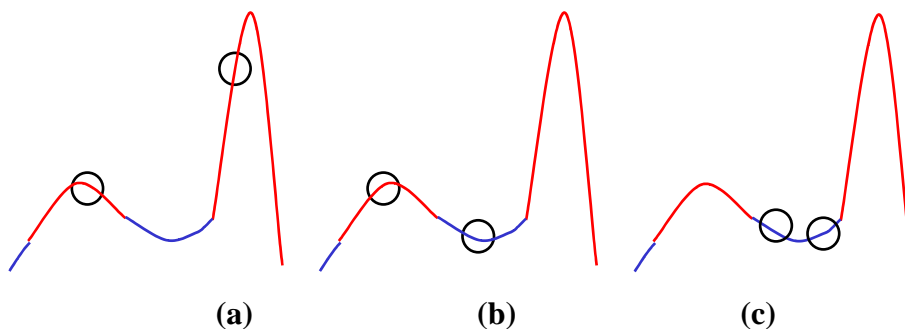
圖一(a)是真實自然影像的一部分，影像的內容包含了一個的均勻背景和三個物件。在影像中，最大物件的亮度值變動範圍是所有物件中最大的，而最大的物件和背景的交流處存在有陡峭的邊界以及平緩的邊界。人類的視覺感知似乎是建立在訊號的比對結果上，因此，在我們所提出的方法中，我們也試著以資料的比對來定義像素的可信任度。圖一(b)是圖一(a)的影像資料的亮度機率統計圖經由平滑化後結果，而平滑化的過程是由Parzen-window 法來實現。假設像素點 Q 的亮度值為 x_Q ，則像素 Q 的可信任度定義為， Q 點的函數值 $p(x_Q)$ 和在 $x = x_Q$ 附近的像素點函數值 $p(x)$ 比較。根據定義，在分佈圖中波峰的位置應該有較高的像素可信任度而波谷的位置應該有較低的像素可信任度。因此，像素可信任度估測的公式如下：

$$C_{pixel}(Q) = \left\{ \left| \frac{dp(x)}{dx} \right| + \left| \frac{d^2 p(x)}{dx^2} \right| \cdot u\left(-\frac{d^2 p(x)}{dx^2}\right) \right\}_{x=x_Q}$$

在機率分佈圖中波谷的地方是一次微分的數值較小而二次微分的數為一正值的位置。因此，根據定義可以得低的像素可信任度。圖一(c)是由圖一(b)機率分佈圖中得到的像素可信任度結果。

(2) 像素相互之間可信任度的測定

在偵測影像中的邊界時，相互之間的可信任度關係也必須考慮，像素相互之間的可信任度是用來確認相連的兩像素點是否屬於同一個均勻區域的可信任度大小。在我們的定義中，只要沒有違反圖二中的三種情況，像素相互之間的可信任度均屬於高，反之為低。如在圖二(a)中，雖然兩個像素的像素可信任度均落於高可信任度區間，但是它們的亮度值之間存在著低可信任度區間，當他們相連時會跨過低可信任度區間，因此它們相互的可信任度為低。而在圖二(b)(c)中，有一個以上的像素落於低可信任度區域的情況發生時，則均屬於低的相互可信任度。

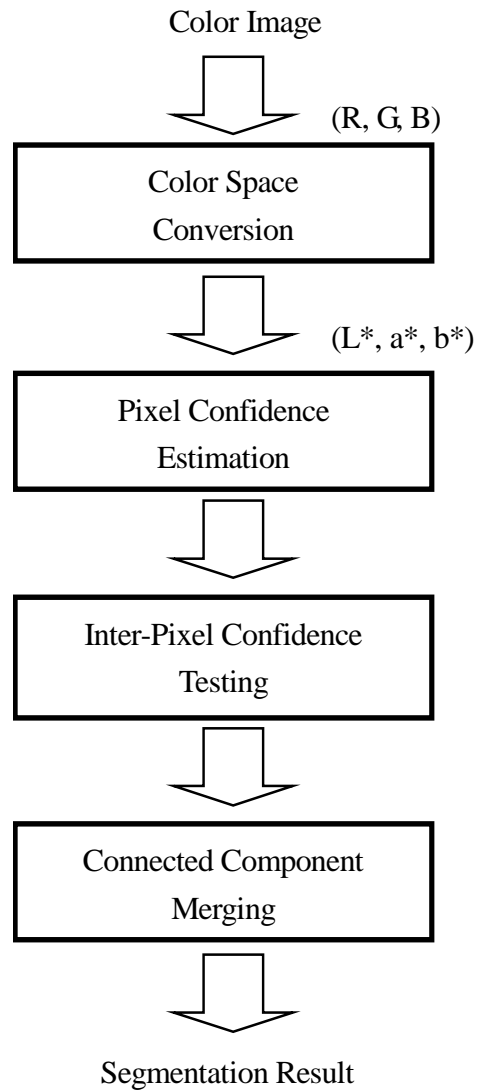


圖二、三種低的相互可信任度的情況

(3) 影像切割演算法的設計

圖三是我們提出的影像切割演算法的流程圖。首先，經由色彩空間的轉換將色彩資訊轉換到 $CIE L^*a^*b^*$ 的色彩空間中，然後個別處理 (L^*, a^*, b^*) 等三個成份。我們將每一

個成份均切成相互重疊且固定大小的數個區域，並在每個區域內統計處理得到平滑化的機率分佈圖。接著估測位於區域中心像素點的可信任度資訊，再測定相互像素的可信任度，然後結合 (L^*, a^*, b^*) 三個成份的資訊得到均勻區域的邊界，最後經由相鄰像素相連的過程而得到最後的切割結果，如圖四。

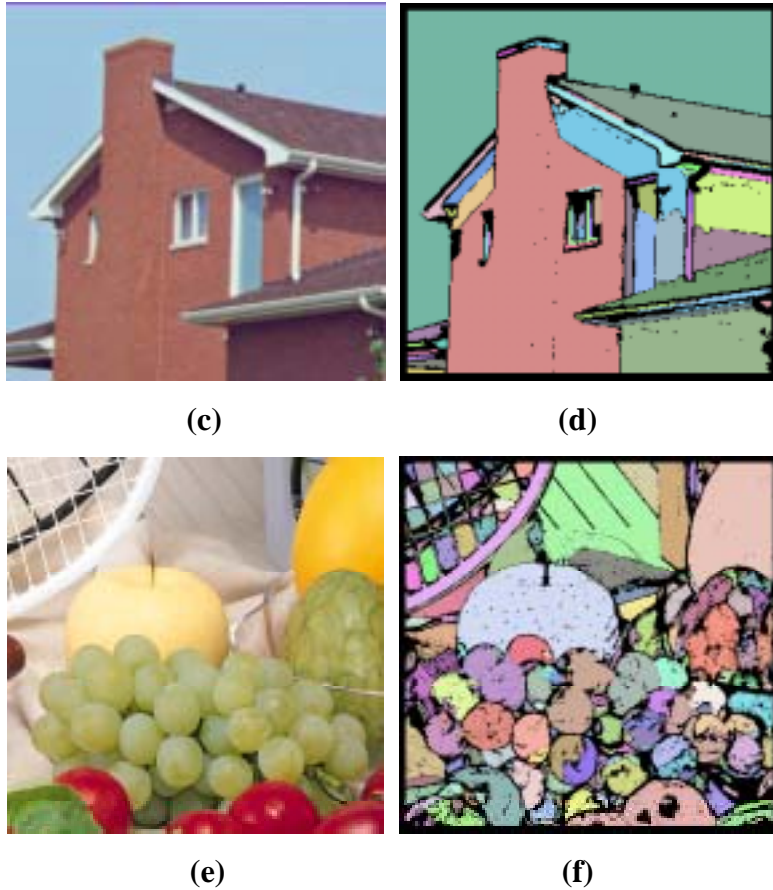


圖三、彩色影像切割演算法的流程圖



(a)

(b)



圖四、(a)(c)(e) 原測試影像 (b)(d)(f) 影像切割結果，相同色彩表示同一區塊

3. 視訊傳輸的互動迴授型式及錯誤防止 -- 張文鐘教授

The primary problem in wireless real time video coding is bit allocation. Secondly is the consumption of the allocated bits. The issue of bit allocation is concerned with the channel estimation, prediction and transmission capacity. The issue of bit consumption is based on an accurate rate-distortion model to derive the quantizer from the allocated bit count. The model parameters have to be adaptive according to the property of the non-stationary signal such that the use of the derived quantizer will result in the same amount of bits as that allocated. This research investigates such a problem and proposes methods to update the model parameter.

Feedback channel with ARQ protocol is commonly used to ensure correct transmission. However this kind of error concealment will increase the transmission burden. The direct consequence is that the effective buffer output rate will decrease due to the retransmission and cause buffer fullness. This requires the source coding rate to be adapted to the buffer condition. With TMN8 rate control, the control mechanism is divided into two levels. The first level is the frame rate control and the second level is the Macro-block rate control. In the frame rate control, the amount of bits to be allocated for a frame is determined based on the buffer condition. To prevent frame skip, the buffer fullness has to be kept under a certain threshold. After the frame rate has been determined, macro-block bit allocation begins. The macro-block bit allocation is proportional to the variance of the motion compensated macro-block. After the bit allocation, rate-distortion model relating the quantization parameter and the bit rate is

used to derive the quantization step size from a given bit rate. In TMN8, the parameters relating to the rate-distortion model are not well estimated such that the deviation between the allocated bit and the resultant coded bit rate very large. In this paper, we discuss some alternative methods for the estimation of these parameters.

The bit allocated by TMN8 to each macroblock is proportional to the product of the standard deviation σ_i and the distortion weighting α_i . Then from the rate-distortion model, $B_i = A \left(K \frac{\sigma_i^2}{Q_i^2} + C \right)$, the quantization value Q can be derived. As shown in Fig.5, the actual coded bits are very different from the predicted bits.

The parameters K and C represent the property of the block that is to be coded. They are a complex function of the image pixel distribution. Since what is concerned is the coding rate, in this paper, a compensation update method is used to derive K and C. When the coded bits and the allocated bits are different, this difference is used to update K and C and try to minimize the difference between bit counts in the next encoding. The parameter C is the overhead rate. In H.263 syntax, some header fields are known before one starts to encode a block. From the header information, the C is modified sequentially block by block.

The estimated K before encoding and updated after encoding based on the actual coded bits, $\hat{K} = \frac{B'_{LC,i} (2QP)^2}{16^2 \sigma_i^2}$. Similarly shown is the estimated C and the updated \hat{C} , where $\hat{C} = \frac{B'_i - B'_{LC,i}}{16^2}$ is the actual header bit count.

Based on the above derivation, we obtain the general update rule.

$$K_i = K_1 \left(1 + \frac{\Delta B_{i-1} \sum_{k=1}^N \sigma_k \alpha_k}{\hat{B} \sum_{k=i}^N \sigma_k \alpha_k} \right).$$

We call this method 1.

A slight modification of the distribution of the difference bit result in the second method.

$$K_i = K_{i-1} \left(1 + \frac{K_1}{K_{i-1}} \frac{\Delta B_{i-1}}{\hat{B}} \frac{\sum_{k=1}^N \sigma_k \alpha_k}{\sum_{k=i}^N \sigma_k \alpha_k} \right)$$

Let us now compare the performance of the proposed two methods with the TMN8. Figure 6 shows the difference between the allocated bit and the actual coded bits. Figure 7 and 8 show the PSNR for each coded frame. From these results we can see that proper estimation of the parameters of the rate distortion model is important in deriving better quantizer step size.

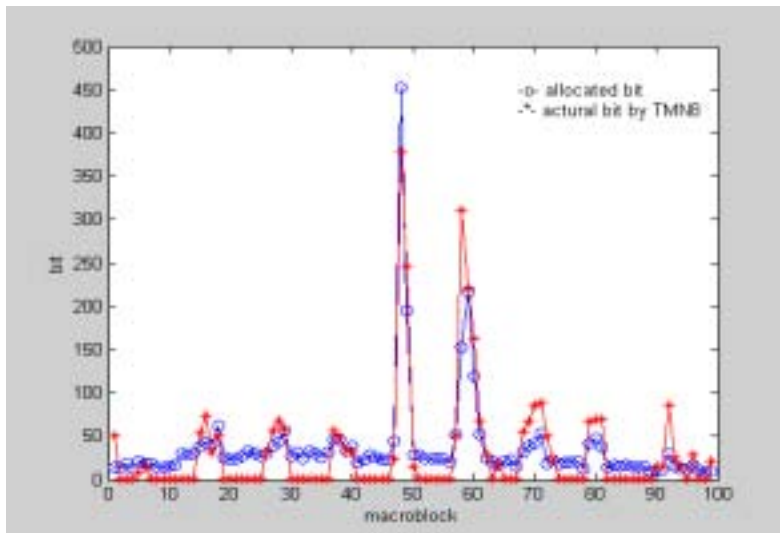


Figure 5. The actual bit rate and the predict bit rate of TMN8

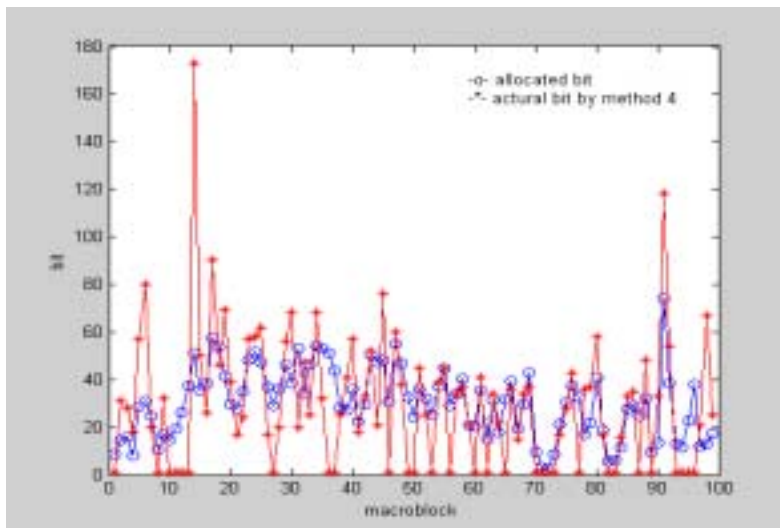


Figure 6. The actual bit rate and the predict bit rate of Method 2

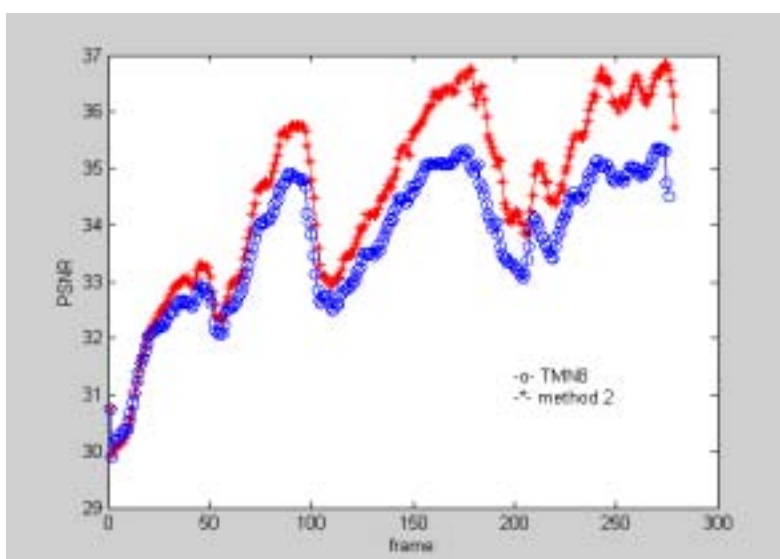


Figure 7. PSNR of TMN8 and Method 1 for salesman

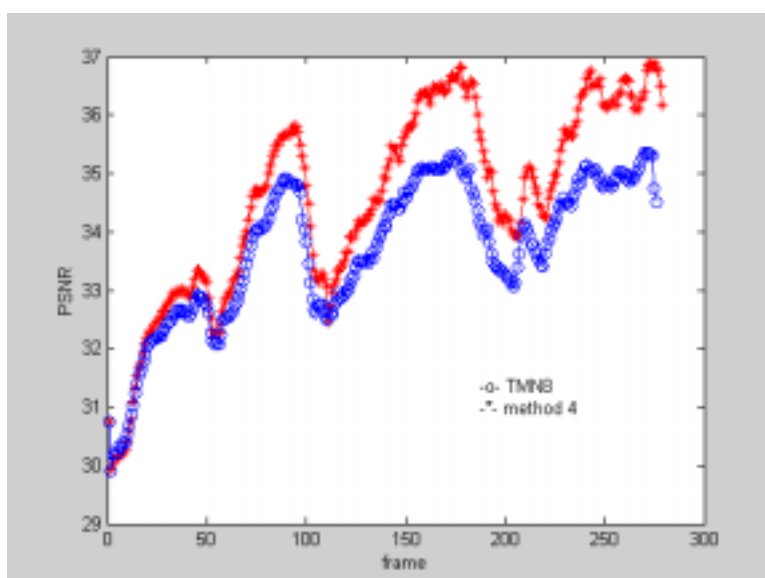


Figure 8. PSNR of TMN8 and Method 2 for salesman

4. 比例式視訊編碼技術及視訊終端機技術之研究 -- 林大衛教授

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相關技術；其二是視訊通訊終端機技術，主要為國際標準視訊編解碼器之實現與相關國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，我們主要研究視訊內容分割法，迄今已提出數項技術。模擬結果顯示：萃取得之物件符合人類常態知覺。現亦在繼續研究改進視訊分割技術。在視訊通訊終端機方面，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。在本計畫中，我們繼續改進此系統之功能，並採用相似的平台進行 MPEG-4 simple profile 及 fine-granularity scalable 即時編解碼之實作。

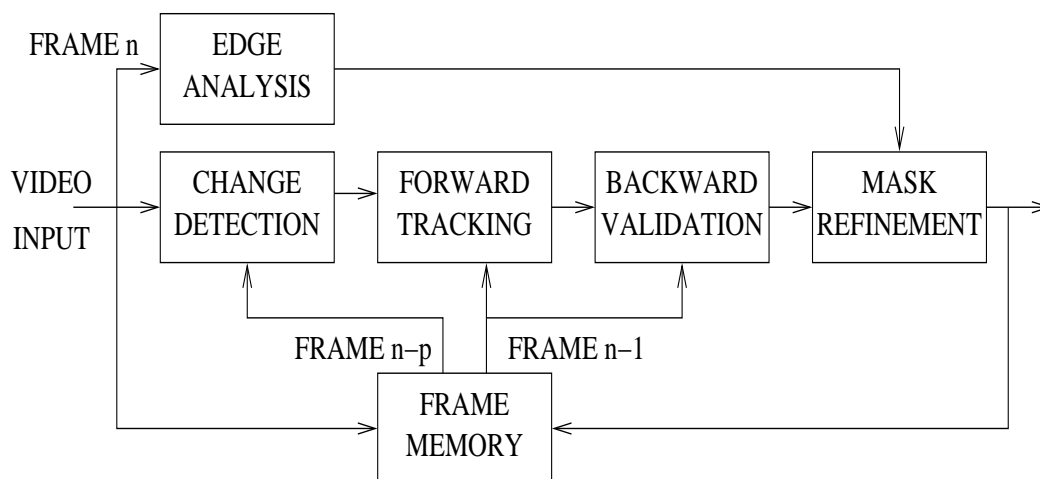
(1) 視訊分割

要達成物件比例式視訊編碼(object-scalable video coding)，一個很重要的課題就是視訊分割，這對自然景像視訊(natural scenes)的編碼而言，尤其為然。由於 MPEG-4 標準中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。如前述，我們在此專注於採用直覺的運動與紋理分析來做視訊分割。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。圖九呈示我們最近提出的一個方法，其中 Edge Analysis 及 Change Detection 屬紋理分析，Forward Tracking 及 Backward Validation 用到運動分析，Mask Refinement 則完成初始分割與區域追蹤。在視訊分割的研究中，兩大議題是物件邊界的精確認定及運算量的降低。此方法在這兩方面都有特別的設計。以下我們就概略介紹此方法。

此方法中的 Edge Detection 目的在於較精確的找到物件的邊界位置。這是因為一般而言，物件的邊界有較大的亮度或色彩變化。我們所用的 Edge Detection 方法為 Canny edge detector。Change Detection 常被用來獲得移動物件的大致位置。我們所使用的 Change Detection 方法與近來若干學術論文所用的方法相似，就是透過 interframe difference 的分析來估計視訊畫面中的攝影機雜訊大小，然後設定一個門檻值，以檢驗 interframe dif-

ference。大於此門檻值的畫面位置就算是 changed，所有算是 changed 之像素就形成移動物件位置的一個粗估。

Forward Tracking, Backward Validation, 及 Mask Refinement 是此方法主要創新之所在，其中又尤以 Mask Refinement 為然。Forward Tracking 是用以估計已分割出來的物件的運動並做粗略的追蹤。由於後續的 Backward Validation 及 Mask Refinement 會更精確的確認物件邊界的位置，所以 Forward Tracking 中的運動估計不必非常精確，也因此可以降低其運算量。我們為此設計了一個特別的運動估計法。Forward Tracking 在跟據所估計得的運動作過初步的物件追蹤後，將其結果與 Change Detection 的結果相結合，作為 Forward Tracking 方塊的輸出。Backward Validation 是將 Forward Tracking 的輸出中，屬於 Change Detection 的結果而不屬於初步物件運動追蹤結果的像素，做反向運動估計，並檢測其是否屬於或鄰接於前張畫面中所分割出來的運動物件。若是，則保留，否則刪去。Mask Refinement 的主要精神，是假設 Backward Validation 的結果中，最外緣的 edge 像素，大多應是物件的邊界所在。透過一些 morphological 處理步驟，我們確認這些邊界像素的位置、針對其斷裂不連續的部分做內插以連接之、並填滿物件的內部。



圖九、直覺分析視訊分割法

實驗顯示此方法可得相當符合主觀視覺的分割結果。圖十呈示對 Mother and Daughter 影像序列進行分割的部分結果。



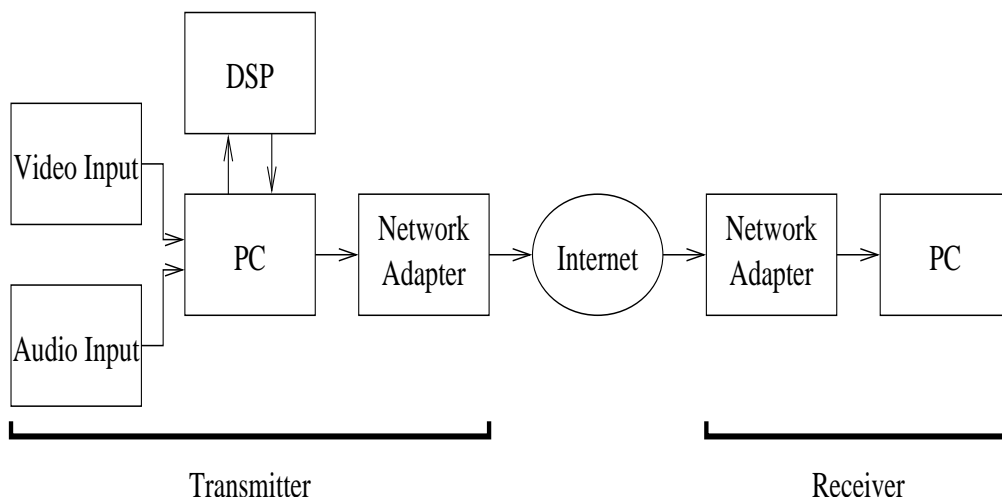


圖十、對 Mother and Daughter 影像序列做分割的部分結果。頂排為原始畫面，中排為分割出之移動前景物件，底排為畫面序號

(2) 視訊通訊終端系統實作之研究

本部分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板來進行軟體視訊編解碼器及視訊壓縮與網路傳輸終端系統之實作。本項研究分兩子題，一是既有 H.263 編解碼與傳輸系統的改進，二是 MPEG-4 軟體視訊編解碼器的實作。以下分別討論之，但重點在第二項，因其為本部分研究之主要項目。

我們之前已經完成一個簡單的 H.263 視訊編解碼與網路傳輸系統。該系統結構如圖十一所示。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，其視訊輸入經個人電腦轉交數位訊號處理器插板做編碼。



圖十一、H.263 視訊編解碼與網路傳輸系統架構

本年的工作主要為系統功能的改進。為免大幅更動系統架構導致意想不到的問題，我們沿用之前使用的數位訊號處理器插板，即 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，但沒有配置所有的功能，使其簡化以利即時實現。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。原始之實現係針對 subQCIF (128x96)之畫面，上年度已改為可處理 QCIF (176x144)畫面，本年則改進為可處理 CIF (352x288)畫面，但編碼速率則成比例下降：subQCIF 每秒約可編 20 張畫面，CIF 則僅 2-3 張。經分析程式，發現

其資料輸出入部分可做一些改進，但對程式加速的幫助極有限。其他改進則尚須做更多分析，才能確定其效用。不過以上經驗將有助於新年度(下年度)之 MPEG-4 研究。

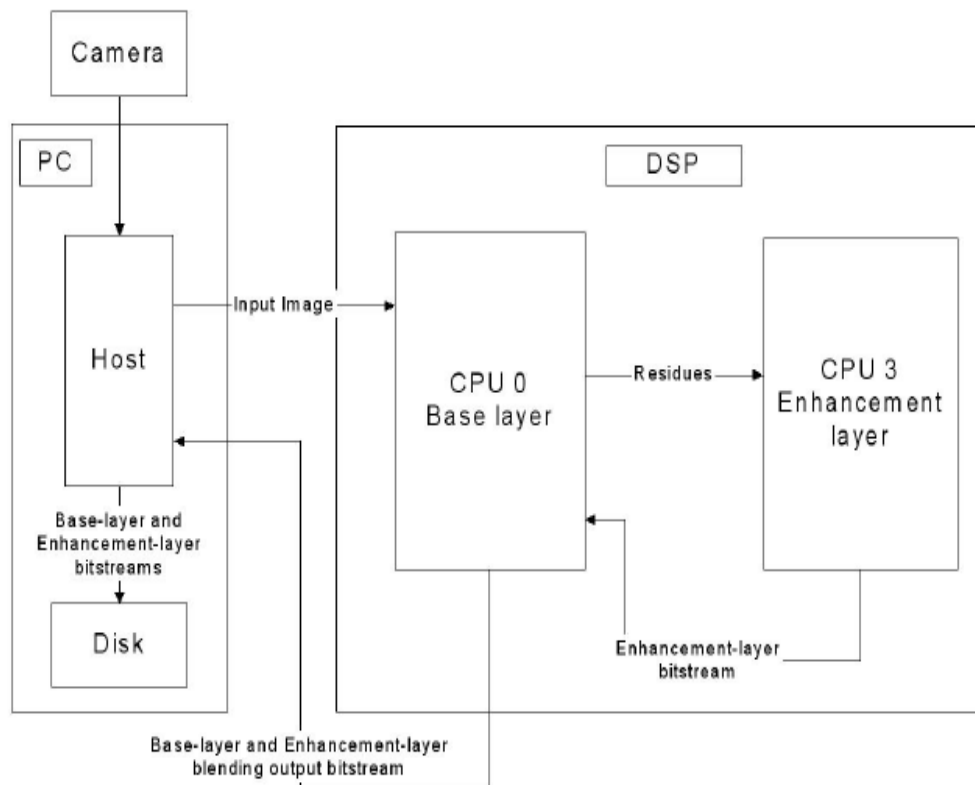
在 MPEG-4 軟體視訊編解碼器部分，我們考慮了其 simple profile 及 FGS 編碼器二者，並分別使用一個數位訊號處理器平台(含個人電腦及數位訊號處理器插板)來實現。所用的數位訊號處理器仍是 TMS320C6201，但插板則為 Innovative Integration 公司的 Quatro62。該插板共裝置四顆 TMS320C6201，但我們的二種編碼器實現則各使用二顆。

(i) MPEG-4 Simple Profile 視訊編碼器

MPEG-4 simple profile 視訊編碼器的實現，係以 MoMuSys C 語言軟體為本，加以修改以適數位訊號處理平台之用。主要工作內容可分程式縮小與程式加速兩方面。實驗顯示編碼速率約為每秒 6-8 張 QCIF 畫面，其 PSNR 值與原始程式相差不遠。

(ii) MPEG-4 FGS 視訊編碼器

我們採用一個既有的 H.263+ 編碼器為 base layer，而 enhancement layer 則使用 MoMuSys 軟體修改而得。其中 base layer 佔一顆數位訊號處理器，而 enhancement layer 則使用另一顆；整個系統架構如圖十二所示。



圖十二、MPEG-4 FGS 軟體視訊編碼器之架構

實現的結果，在沒有省略任何 bitplanes 之情形下，視比較基礎之不同，約可加速為原程式的 2.4-2.7 倍，或 6.4-7.6 倍。編碼速率約每秒 11.5-13.5 張 QCIF 畫面。若省略最後二個 bitplanes，則速率可達每秒 17-19 張畫面。

5. MPEG-4 與 MPEG-7 系統之研究 -- 杭學鳴教授

我們在此研究計畫的目的為(1)深入研究並模擬 MPEG-4 系統，主要為 MPEG-4 IPMP 延伸系統，以及(2)研讀模擬 MPEG-7 系統與製作 MPEG-7 發展平台。

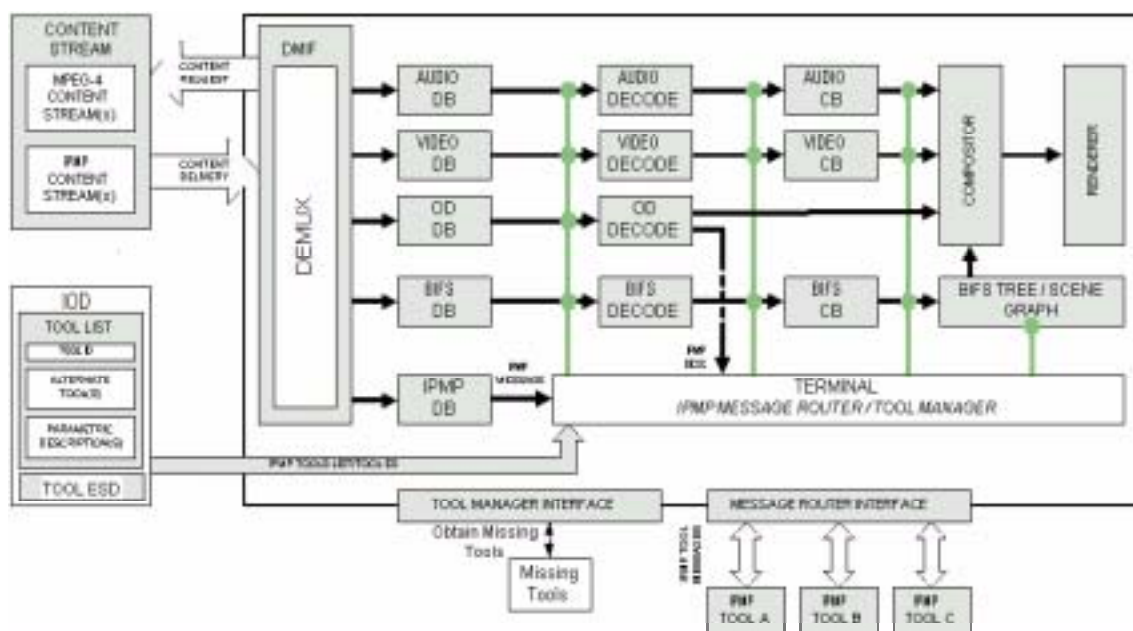
(1) MPEG-4 系統的研究

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，以期對 MPEG IPMP 有更前瞻的認識。

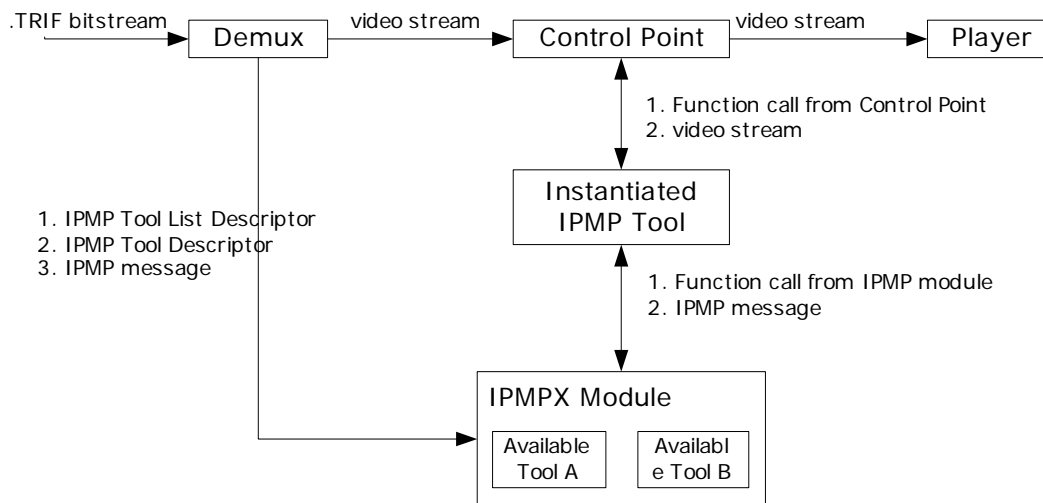
MPEG-4 Systems ver.1 中所定義的 IPMP 架構提供了一個標準化的介面讓 MPEG-4 players 可以使用不同的 IPMP System。MPEG-4/AMD3 提出後，為了區別兩個版本中差異頗大的 IPMP 子系統，前者稱 IPMP Hook，而後者稱為 IPMP Extension (IPMPX)。

IPMP Hook 架構的最大不足之處在於工具間的溝通方式並沒有正式定義，這使得各工具的實作者無法利用他人的實作成果。有鑑於此，IPMPX 架構觀念改為虛擬終端機 (Virtual Terminal)，與既有的 MPEG-4 系統以 Message 互相溝通。如圖十三，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，而各 Tool 則根據 Message 內容負責串流的處理，例如解碼或是權限控制。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。

本模擬重點為 IPMPX，相關的參考軟體不多，一為 Craig A. Schultz 所作，另一個為 MOSES。兩者皆利用 IM1 為 MPEG-4 player。IM1 軟體包括了所有 MPEG-4 System 中標準化的部分。我們分析了 IM1 Core 中有關 IPMPX 的部分，了解 IPMPX 系統的操作模式，並據此設計三個程式，以循序漸進的方式，展示 IPMPX 的運作。圖十四為本展示程式的資料流程圖，包含 IPMP 資訊的 MPEG-4 串流到達後，先經過 Demultiplex 影像資料流往 player 的 control point，而 IPMP 資料則流向 IPMPX 模組，啟動對應的 IPMP Tool 並開始對流經 control point 的影像資料作處理。



圖十三、IPMPX 與 MPEG-4 系統



圖十四、展示程式 Dataflow

研究的最後，我們以三個程式展示 IPMP 模組的啟動、Message 的傳遞、IPMP 資訊的動態更新，以及透過 IPMP 進行權限控管。根據這幾個程式，我們初步驗證 IPMPX 與 MPEG-4 player 搭配，的確可以執行一些電子商務所需的機能。

(2) MPEG-7 系統的研究

MPEG-7 採用 XML 相關技術定義其提供的工具。也就是說，符合 MPEG-7 的 meta-data 格式都是 XML 的 Application。參與 MPEG-7 的專家們除了規範各式各樣的基本 descriptor (簡稱 D) 及 description scheme (簡稱 DS)，同時也提供一套參考軟體供大家驗證其可行性。Reference Software 簡稱 XM，為 eXperimentation Model 的縮寫。主要提供一套 C++ 軟體，展示各 D 及 DS 的可行性與搜尋準確度。

儘管 MPEG-7 XM 已經提供 C++ 軟體供我們研究，但其目的是驗證各個 D 與 DS 是否可行，以及參考演算法是否正確無誤。對於深入研究的需求，例如以圖表方式顯示找出的 D 或 DS 內容，或是資料庫運作的設計，XM 的功能就不符所需。我們根據 MPEG-7 的通用使用狀況，配合研究上可能的需求 (資料庫、網路、圖形化展現等等)，規劃一個軟體平台。這個平台以 Java 語言為基礎，其上建構一個我們設計的 framework，這個 framework 主要由三類抽象元件構成：

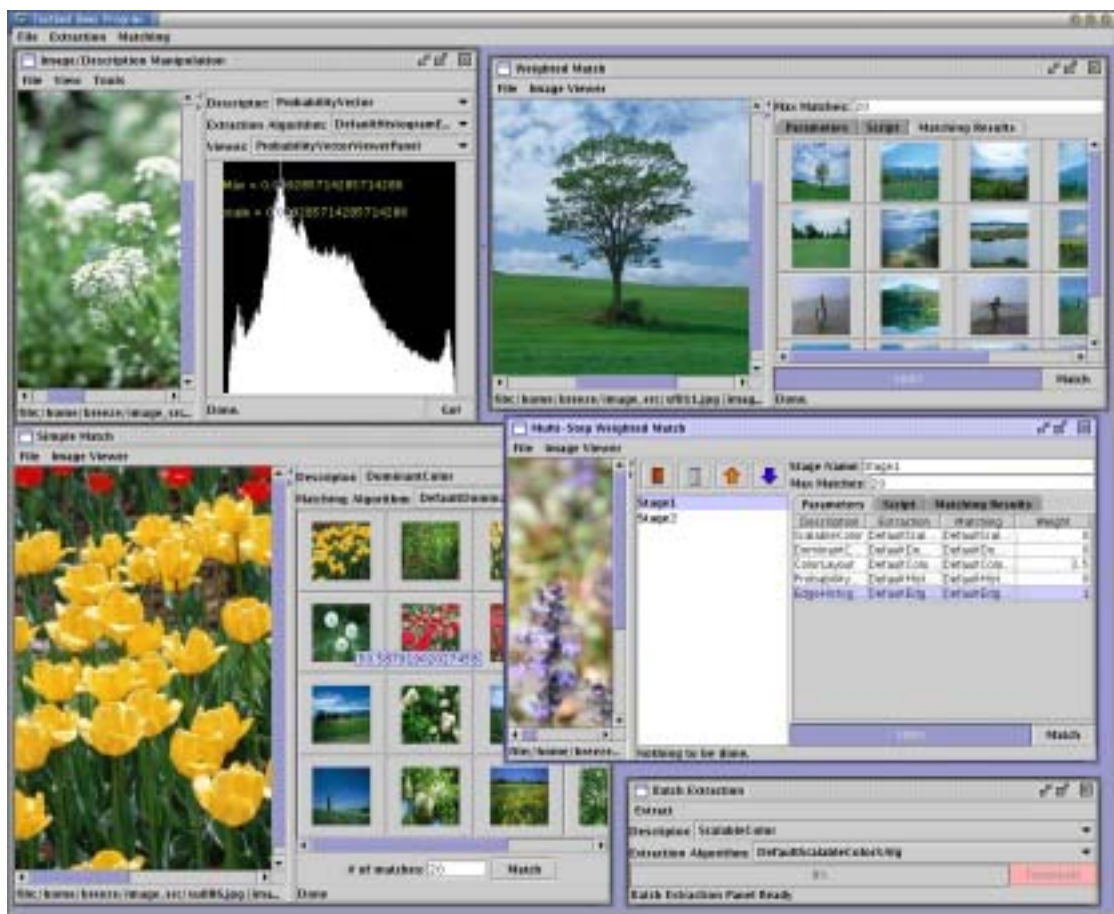
- Data：觀察或處理的主體，可能是媒體資料或 meta-data。
- DataAlg：與 Data 關聯的演算法，就媒體資料而言，為編碼與解碼。若作用主體為 meta-data，則除了編解碼之外，還包括抽取與比對演算法。
- Viewer：用來觀察 Data 的圖形化介面，或是轉換 Data 型式的演算法。

在這個測試平台中，以 Data 為中心，允許有零至多個 DataAlg 或 Viewer 與其相關，因此設計 Component Management Unit 用來維護元件間的關聯性。有了 framework 與 Component Management Unit 後，我們就可依此架構實作不同的媒體、D、DS、以及相

對應的顯示元件。目前在此平台中，我們完成了 Scalable Color、Color Layout、Dominant Color、Edge Histogram 等 meta-data 相關元件，也完成 image 相關元件。

最後，設計一個影像特徵抓取與搜尋原型程式，除了展示基本的 MPEG-7 應用方式 (特徵抓取、儲存、比對、整合)，並提供使用者檢視 descriptor 內容的介面。圖十五所示為利用此軟體平台開發出來的原型程式，在這個程式中，我們除了展示了 Histogram、Color Layout、Edge Histogram、Scalable Color 相關的 feature extraction、feature matching、以及 image/feature viewer，並利用設計的 Component Management 單元展示如何在更動程式下，新增或移除 Concrete Components。

此外，我們在應用程式層次加上額外兩種 matching 方式，稱為 weighted matching 及 multi-step matching，前者可用 weighting factor 方式組合多個 D 或 DS 作為搜尋條件，後者則是進一步將搜尋分成多級，每級皆為 weighted matching，且每級的搜尋範圍限制在前一級的搜尋結果中。實驗結果發現，適當安排各級搜尋條件不僅可以得到較令人滿意的結果，而且搜尋速度與 weighted matching 相差不會太多。除了 weighted matching 類型外，我們也實作了以集合運算結合多種特徵的比對方式，實驗結果顯示在某些狀況下的確可以得到比較好的結果，但缺點是如果集合不夠大，可能無法取到共同交集。



圖十五、軟體平台程式展示

6. 多媒體架構與數位視訊浮水印在網際網路之應用 -- 蔣迪豪教授

MPEG-21 provides a unified solution, Universal Multimedia Access (UMA), for constructing a multimedia content delivery and rights management framework. Based on the concepts of UMA, we build a simplified UMA model on the Internet. In this framework, the source video material is encoded and archived as FGS bitstreams. To support video contents of different formats, we create a transcoder to convert the bitstream from the FGS format to an MPEG-4 simple profile format that fits to the terminal capabilities. Moreover, a novel FGS coding scheme is present to improve coding efficiency and retain robustness of FGS bitstreams for video streaming over Internet. Consequently, the multimedia information can be streamed through the networks without networks jitters and significant quality degradations existed in the current commercial implementations. To have a more strict evaluation methodology according to the specified common conditions for scalable coding, an FGS-based unicast streaming system is used as a test bed of scalability over the Internet.

(1) Architecture of UMA multimedia Delivery System

For achieving Universal Multimedia Access (UMA), we propose a video server that contains the key modules described in MPEG-21. In this model, we combine the tools as referred to MPEG-4 Fine Granularity Scalability (FGS), MPEG-4 Simple Profile, MPEG-7, Digital Watermarking techniques, and Internet protocols.

Based on the concepts of UMA, we build a simplified UMA model on the Internet. To fit with the issue of content adaptation according to terminal capability, we propose a real-time transcoding system that converts the FGS bitstreams into Simple Profile bitstreams. In this framework, the source video material is encoded and archived as FGS bitstreams. To support video contents of different formats, we create a transcoder to convert the bitstream from the FGS format to an MPEG-4 Simple Profile format that fits to the terminal capabilities. As shown in Fig. 16, the proposed system includes five modules. With the FGS bitstreams saved in FGS BitStream Archive module, the proposed system can serve heterogeneous terminals through the Internet. Moreover, according to Internet and Terminal devices capabilities, the Channel Monitor can adapt the different resources to each Terminal. Thus, since the source video is encoded and archived as FGS bitstreams, we can provide various QoS service like SNR scalable video coding schemes.

(2) Efficient FGS-to-Simple Transcoding

To demonstrate the performance of our proposed UMA multimedia delivery system in Fig.16, we build an FGS-to-Simple transcoder. In the proposed system, each sequence is pre-encoded and stored in the FGS Bitstream Archive.

Three methods are considered for comparison:

1. A simple profile encoder using the original video sequence (SP_ME)
2. A cascaded transcoding using a complete FGS enhancement bitstream and motion vectors from the base layer bitstream (SP_MV)
3. An efficient transcoding with complete FGS enhancement video and motion vectors from the base layer bitstream (FGS-to-SP).

The test video sequences, named as Foreman, News, and Container, are in CIF and YUV format. The first frame is coded as an I-VOP and the others are coded as P-VOP's at 30Hz.

For the FGS encoding, the quantization step size (QP) used in the base layer is set at 10 for I-VOP's and 12 for P-VOP's. The MPEG-4 Simple Profile encoder employs constant quantization, where the set of QP used is {5, 7, 14, 21, 28}. As shown in Fig.17, our transcoding schemes (FGS-to-SP) have neglected quality loss in PSNR at low and medium bit-rates and have about 0.5~0.9 dB loss in PSNR at high bit-rate.

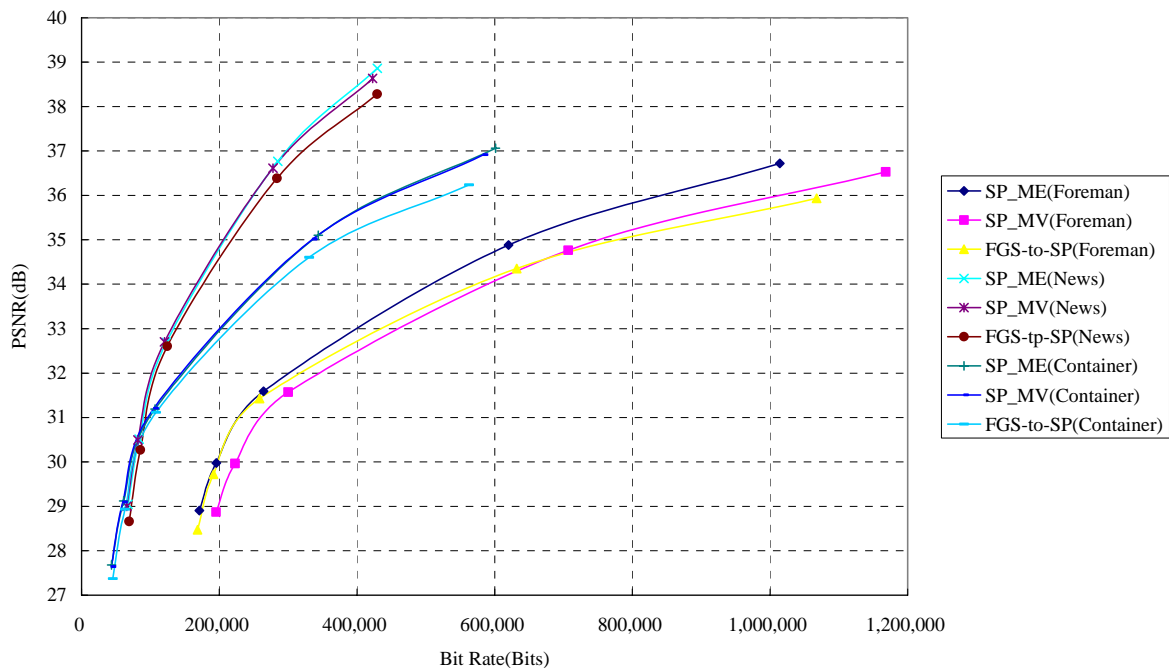


Figure 17. The performance of transcoding with the luminance components of the three video sequences and using various sources of motion vectors and different enhancement information.

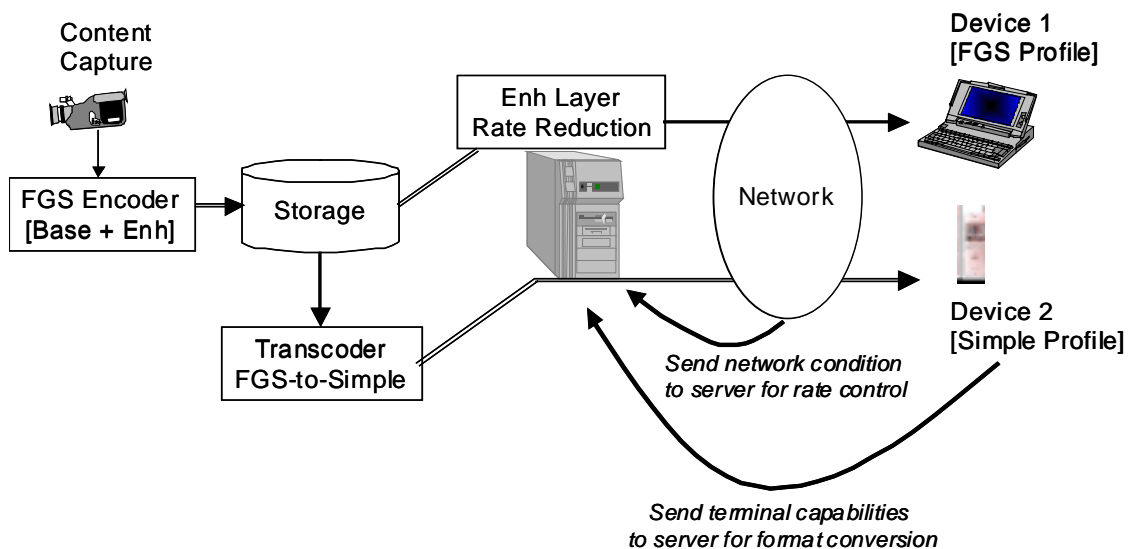
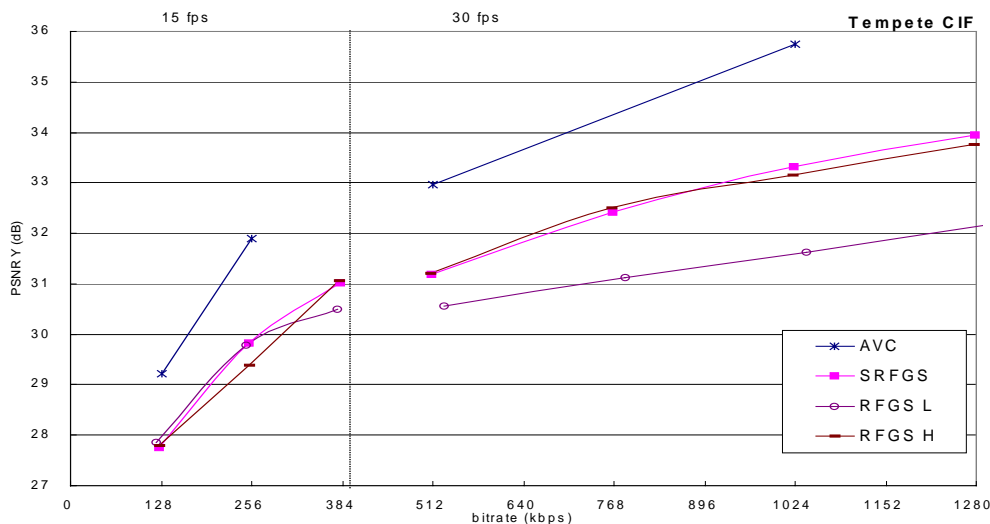


Figure 16. The application scenario of the proposed UMA multimedia delivery system that employs the archived FGS bitstream

(3). FGS Streaming on the Internet

The coding efficiency of the SRFGS is compared with RFGS and MPEG-4 Part-10 Advance Video Coding (AVC). For test conditions, we adopt the testing procedure specified by the MPEG Scalable Video Coding AHG. The sequences including Tempete, Bus and Container in CIF resolution are tested at four bitrates/frame-rates, including 128kbps/15fps, 256kbps/15fps, 512kbps/30fps, and 1024kbps/30fps. The results of AVC use JM42 test model. RD-optimized and CABAC are used. Quarter-pixel motion vector accuracy is employed with search range 32 pixels. Four reference frames are used. Only one I-frame is used at the beginning. The P-period is 3 in both 15fps and 30 fps. For RFGS and SRFGS, the base layer is JM42. The test conditions are the same as AVC except that we have disabled RD-optimized and only one reference frame is used. At 30 fps, the P-period is 6 for Tempete and Container. The P-period is 4 for Bus. At 15 fps, the P-period is 2. The bitplane and entropy coding are identical as the MPEG-4 FGS. In SRFGS, 2 enhancement layer loops are used for Tempete and Bus, and 3 enhancement layer loops are used for Container. A simple frame-level bit allocation with a truncation module is used in the streaming server to obtain the optimized quality under the given bandwidth budget.

The simulation results are shown in Fig. 18. Two RFGS results are shown, one has lower reference bitrate (labeled as RFGS_L) and the others has higher reference bitrate (labeled as RFGS_H). SRFGS has similar performance with RFGS_L at low bitrate, and has 1.7 to 3.0 dB improvement at high bitrate. This is because SRFGS has remove the temporal redundancy at high bitrate while RFGS_L not. As compare with RFGS_H, SRFGS has 0.4 to 1.0 dB improvement at low bitrate. This is because there is more drift error of RFGS_H at low bitrate. At high bitrate, the SRFGS has 0.8 dB improvements at low motion sequence such as Container and has similar performance at high motion sequence, such as Tempete and Bus. This is because at high motion sequence the correlation between successive frames are lower and the improved prediction technique in SRFGS may not help too much. At medium bitrate, SRFGS has 0.15 dB losses than RFGS_H at most. This is because the increased dynamic range and sign bits of each layer in SRFGS slightly degrade the coding efficiency. The above simulation results show that while RFGS can only optimized at one operating point, SRFGS can optimized at several operating point to serve much wider bandwidth with superior performance. Compare with AVC, SRFGS has 0.4 to 1.5 dB loss at base layer. This is because the MV in SRFGS is derived by considering not only the base layer but also the enhancement layer information. Further, the high quality prediction image of B-frame has not totally received at this bitrate. There are 0.7 to 2.0 dB PSNR loss at low bitrate and 2.0 to 2.7 dB loss at high bitrate.



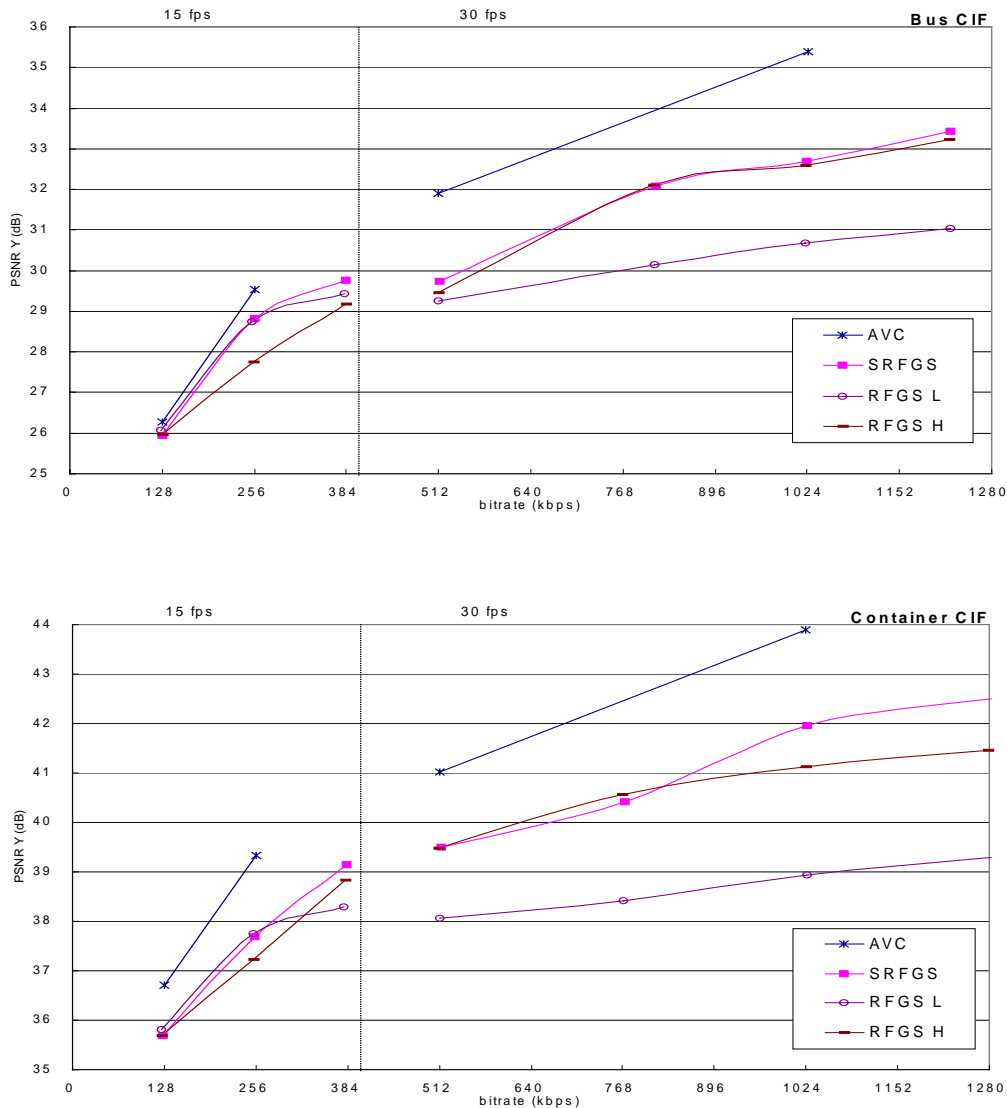


Figure 18. PSNR versus bitrate comparison between SRFGS, RFGS and AVC coding schemes for the Y component

7. 計畫成果自評

本計畫旨在發展與 MPEG-4 相關之各項技術。多媒體通訊(包括數位電視, DVD 等)是近來電子資訊工業的主要趨向之一,也是目前國家政策推展的重要建設目標(為 NII 的主要項目之一)。工商業正使用或未來將使用的標準規格預期將有極大比例與 MPEG-4 標準密切相關,在國科會的支持下,我們較產業早一步對這些規格標準加以探討,發展其中關鍵技術。如上所述,各子計畫研究項目分別獲得相當豐碩成果,發表學術論文,與原訂目標相符。各子計畫的著作發表分別列在個別報告中。

參與本項計畫的三位老師蔣迪豪、杭學鳴、王聖智,並在業界廠商合作計畫補助旅費情況下參與 MPEG 標準會議,並提建議案多項。例如:ISO/IEC JTC1/SC 29/WG 11

14496-2 M8049: A Robust Fine Granularity Scalability (RFGS) Using Predictive Leak , ISO/IEC JTC1/SC29/WG11 M9767: Stack Robust Fine Granularity Scalability 等。目前我們在 MPEG 標準會議進行之主要工作項目有 MPEG-4 Part 7 Optimised Reference Software 以及 MPEG-21 Part 12 Multimedia Test Bed for Resource Delivery。(我們參與 MPEG 標準會議的技術開發與活動經費，亦受到交通大學李立台揚網路研究中心與多媒體標準資源共享等計畫之贊助。)

此外，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。