

行政院國家科學委員會專題研究計畫 期中進度報告

新一代視訊壓縮技術--可調整性編解碼器設計之研發(2/3)

計畫類別：個別型計畫

計畫編號：NSC92-2213-E-002-023-

執行期間：92年08月01日至93年07月31日

執行單位：國立臺灣大學資訊工程學系暨研究所

計畫主持人：吳家麟

計畫參與人員：童怡新，林育慈，劉錦昕，楊雅婷，黃義欽，陳宇皓

報告類型：精簡報告

報告附件：出席國際會議研究心得報告及發表論文

處理方式：本計畫可公開查詢

中 華 民 國 93 年 5 月 31 日

新一代視訊壓縮技術 可調整性編碼器設計與研發(II)

國科會補助專題研究計畫

計畫編號：NSC 92-2213-E-002-023

全程計畫：民國 91 年 08 月 01 日至民國 94 年 07 月 31 日

本年度計畫：民國 92 年 08 月 01 日至民國 93 年 07 月 31 日

計畫主持人：吳家麟 台灣大學資訊工程學系教授

一、摘要 (中文摘要)

隨著數位科技的發展，視訊串流傳輸(Video Streaming)普遍且廣泛地被大眾所使用。在不同環境下通用的存取(Universal Access)相同的視訊服務變得越來越重要且實用。在這種趨勢的促使下，新一代的視訊壓縮將不僅擁有傳統壓縮標準的壓縮效率(Efficiency)，還必需同時對不同的使用環境具高度適應性，包括對數位頻寬需求的可調適性(Bitrate Scalability)以及對串流傳輸中資料錯誤具抵抗(Error Resilience)等能力。在這些新的功能性當中，資料流的可擴充性可使壓縮處理過後的單一資料流可依使用上不同的需求，經由適當的調整參數，擷取部分資料以應用在不同的環境之下，這樣的功能符合異質網路以及分頻網路的頻寬變異的要求。

在本項計畫中，預定以新一代的國際視訊標準 MPEG-4 為基礎，設計一同時具備多重可擴充性的視訊壓縮框架架構，在這種架構下，將可完成下列多種性質的可擴充性供使用上調整之用。1)資料量的可調整性可適應頻寬多變的傳輸環境。2)空間與時間解析度上的可調整性對具不同解析度的顯示裝置提供最有效的利用。3)畫質失真度(Signal-To-Noise Ratio, SNR)的可調整性將有利於視訊的漸進式顯示。4)計算能力的可調整性能讓配備不同速度處理器或分時多工環境下的處理器均能依其能力完成不同層次的解碼工作。此外，為達

成串流資料量的細微可調整性以及維持分層壓縮的效率與可容錯性，在層次壓縮中架構中同時採用位元平面壓縮(Bitplane Coding)技巧以及漸進式預測(Progressive Prediction)結構。為達成壓縮後資料流需儘量滿足最佳的漸進式畫質改善之特質，將考慮人類視覺相關之研究，以求設計出一依媒體(Content)內容及性質而自動調整的最佳化壓縮參數設定。本計畫將同時製作可擴充性資料流串流系統以及高層次可擴充性資料流轉換工具。

(英文摘要)

Video streaming is now in widespread use, and universally accessing the same video content under different scenarios becomes more and more important and practical in the market. Such a demand leads the modern codec design to possess multi-functionalities. The codec designers need to care about not only efficiency, but also error resilience and scalability simultaneously. The new functionality, scalability, enables a single video stream to be scaled up and down such that it can be applicable to different environments.

In this project, we try to devise a hybrid-scalable framework for designing modern video codecs, such that the following scalabilities can be achieved. 1) Rate scalability, it enables coded streams to serve varying-bandwidth clients. 2) Spatial and temporal scalabilities, they benefit contents to

represent in different resolutions. 3) SNR scalability, it gifts the capability of progressive display. 4) Computational scalability, it makes video play well under different computation-power machines or multitasking environments. Moreover, we adopt some latest coding techniques, such as fine granularity scalability, progressive prediction structure, and coding tools provided by MPEG-4, which enhances the underlying coding capability. Taking advantages of hybrid sourcing data reduction and quality control in the layered coding, the proposed scalable system can produce a highly rate scalable and hybrid-functional stream offline. The generated scalable stream can be used to serve all types of clients, especially suitable for bandwidth-varying Internet and multitasking PC.

After building the hybrid scalable framework, we next introduce human perceptual criterion to advice the whole framework. The scalable system inevitably shows the content in distinct quality. It is desired that any enhancement (no matter it is bandwidth, computation, or any constraint releasing) should be used to improve the video quality most. The quality here is of course the received feeling assessed by human being. Thus, we want to seek perceptual rules or algorithms for variable frame rate, resolution and bitrate encoding. Four important issues are disserted here: 1) choosing low-frame-rate but smooth video representation by adopting frame classification and selection, 2) using suitable spatial-temporal decomposition which is coincident with human vision, 3) evaluating masking effects by just-noticeable-distortion (JND) metric and 4) defining suitable properties for each layer and their coding orders to get best quality improvement while moving to the next layer. Considering all these ideas at the same time, the input video source is first analyzed to select representative frames for low-frame-rate coding, and to decide its possible

spatial scalability. Analyzing the video content and applying some observations of visual experimental results helps us to conduct this process. Then, the spatial-temporal-decomposition module is operated with the human perception guided. Next, before going on operating layer-coding modules, decomposed frames are further analyzed to find its perceptual weighting map (frame masking map). These maps together with layered parameters are used as the stimulus to operate the layered encoders. Through these means, the proposed framework is optimized by human perception.

二、計畫緣由與目的

在二十世紀末最後幾年當中，要說到對人們生活影響最大的科技變化莫過於網路與通訊的快速發展與普及。這些年當中，我們看到了網際網路刮起的旋風，行動通訊對人與人的溝通模式所造成的改變，以及寬頻網路的盛行所帶來的多樣化資訊服務，這種種變化，都對整體人類生活產生巨大的衝擊。今日，我們可以隨時隨地廉價且輕易地使用到網路的資源，享受它帶來的便利。而隨著各種網路服務的興起，我們可以預見新一代的網路視訊需求即將產生，而這股強大的需求將促使傳統視訊壓縮更進一步發展與重新設計，以符合現實世界中多變的網路使用者所處的異質環境。

數位壓縮標準的制定

過去十多年當中，數位視訊的先驅持續不斷地對視訊壓縮做研發與改進，我們看到越來越多的壓縮技術在學界與業界中被提出與採用。這些壓縮技術有效地消除連續視訊畫面中的重複性，以達到高壓縮比低失真的目的。檢視這些普遍被大多數人所接納採用的國際壓縮標準，可以發現它們均成功地整合多項壓縮技術(Hybrid Coder)，以提供高效率數位視訊的壓縮與儲存方式。國際標準制定組織(International Standardization Organization, ISO)以及國際電信聯盟(International Telecommunication Union, ITU)為廣播電視與視

訊通訊產業，成功地制定與推廣了 VCD(MPEG-1)、DVD(MPEG2)以及 H. 261 與 H. 263 等壓縮標準。這些數位視訊標準的制定，讓數位視訊正式地取代以錄影帶為主的類比視訊，並結合數位網路傳輸與數位儲存方式技術，帶領人類生活進入真正的數位時代。今日，真正的數位時代已經來臨，隨著網路普遍化帶來的衝擊，單一功能性的傳統視訊壓縮標準已不復現實需求，視訊壓縮將融入更多更實用的性質以符合網路應用多樣化的需求，邁向實現多功能性的新一代壓縮方式。

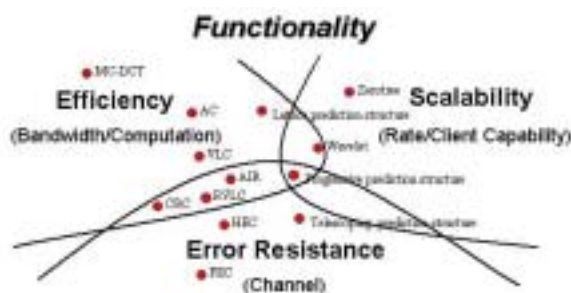


圖 1：新一代視訊資料壓縮所需具備的三大功能性以及現今壓縮技巧與此三項功能性的關係。

數位視訊壓縮的演變

視訊壓縮的原意在於減少數位視訊巨大的資料量，因此“效率”是壓縮所欲達成的最基本也是最重要的功能。如何在合理且受限制的有限運算能力下，將巨量的視訊數位資料減少成現有數位傳輸與儲存媒介所能使用的頻寬是傳統視訊壓縮所欲解決的問題。換句話說，傳統的數位壓縮便是在找出運算效率與壓縮效率兩者與現實環境中的運算晶片與儲存技術的單位價格所做出的妥協點。之後，然而隨著通訊技術與資訊產業迅速地發展，視訊會議的需求建立起通訊與視訊壓縮的橋樑。環境上的改變帶來新的考量因素，通訊當中所發生的些許資料錯誤率對高壓縮率的影像壓縮而言，將造成一連串嚴重的視覺影像失真，對視訊的使用上造成極大的困擾。因此，視訊壓縮的“抗錯誤能力”成為此壓縮標準是否適用在通訊環境上的一大指標。許多較後期的研究均花心力在設計新的演算法或增添資料流同步機制，以便在壓縮與解壓縮流程中，在

發生資料錯誤時能夠有一套自我修正的機制，而將通訊錯誤所造成的影響降到最低。較新的壓縮標準如 H. 263+以及 MPEG-4 均包含了這些技術的初步成果，使這些標準在通訊環境中也能有不錯的表現。

然而環境持續不斷的改變，通訊蓬勃地發展，促成了異質網路的連結；異質網路的連結使得不同的裝置與配備無論是桌上型個人電腦、筆記型電腦、個人數位助理以至於行動電話，都能夠存取最大的資訊叢集。網際網路，這項變化促成了多領域間的連結，並對資訊的流通性有莫大的助益。而對視訊服務而言，更具大的挑戰卻也隨這個改變同時產生。異質網路的成功連結後，視訊服務將可能被不同的裝置配備透過不同的實體網路所使用 (Universal Access)，而這些不同環境下的使用者，卻同樣地對視訊服務有同樣的渴望。他們希望能收看到最新最即時的新聞影片或者某部熱門的電影。但由於這些使用者的網路頻寬、顯示裝置的解析度、使用裝置的運算能力等均可能有不同程度的差異，視訊服務的提供者很難有辦法對這些使用者同時服務，並做到人人滿意的地步。以傳統的壓縮技術而言，要解決這樣的問題，伺服器必需事前統計歸納使用者的種類，對不同環境參數的組合分別產生其相對應的資料流，或做即時的重新壓縮處理 (Transcoder)，才能滿足這些使用上的需求。然而前者需要用大量的儲存空間換取各式各樣不同需求的視訊版本，而後者雖可即時轉換成符合使用要求的資料流但卻需要伺服器端配備有大量的處理器以完成即時的多版本重新壓縮，這對於同時可服務的使用者數量產生限制，並缺乏動態可擴充性。因此，兩者均非最好的解決方法。視訊壓縮的研發者試圖尋找一新的解決方法，兼具上述兩者之優點，能以一事先處理好的分層壓縮的資料流儲存於伺服器端，並提供相其對應的資料屬性。如此一來，伺服器便可依個別使用者的需求與當時的頻寬，截選全部資料流中的部分，傳輸給使用端，而使用者更可依其當時所擁有的運算能力，做部分或全部資料流解碼，達成使用單一資料流卻可服務多數使用者的要求。因此，新的壓縮標準在設計時必

需顧及到此“可擴充性”，如此一來，才能讓視訊壓縮可順利地應用的範圍擴及到異質網路與網際網路。

本計畫欲實作一可同時完成前述之多種可擴充性資料壓縮的系統，用以推廣網路視訊之普及並促進國內產業之升級。此一系統必須能同時支援多種類的實體網路與多樣化的客戶端使用配備，因此支援高層次且多類別的編解碼功能是本計畫的核心且首要之務。

三、研究方法及成果

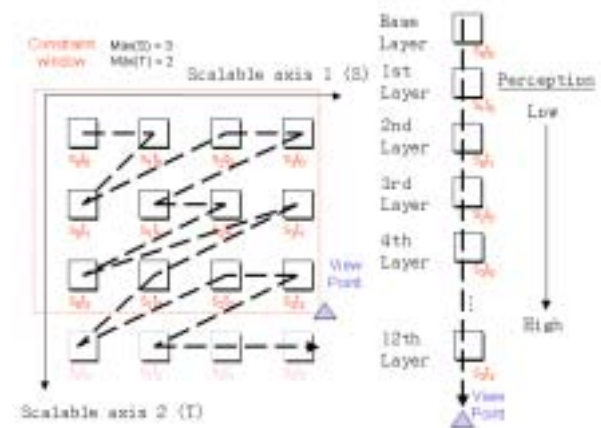
本計畫預計以三年時間完成，第一年著重於整體框架架構之設計與各技術項目模組之實作，第二年將著重於整合各項技術，完成實用之應用程式之雛形，而第三年將建立衡量通用可擴充性的工具，並在不同平台上微調各技術模組以發揮最大之效益。目前本計畫接近第二年之完成階段，預期完成之目標均接近達成階段，以下簡介前兩年計畫之成果。

1.完成高層次複合可擴充性壓縮框架架構之設計。

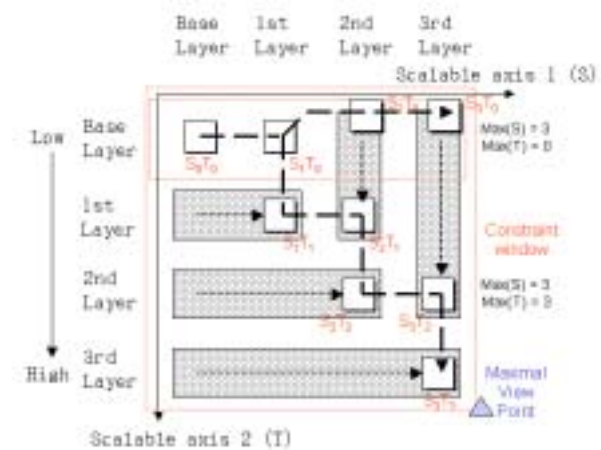
在結合多種性質的可擴充性壓縮時，維持優越的壓縮效率將變為更加的困難。理想狀況下的格狀視訊擴充結構雖然能滿足所有不同使用者在各種環境下對視訊畫質參數做任意調整之需求，但卻不可避免地將喪失較多的壓縮效率。因此，我們提出一個以階層樹狀結構為階層擴充順序之依據的視訊壓縮方式，用以改善格狀視訊擴充結構所遭遇的缺點。這種樹狀結構擴充方式，雖然不像格狀擴充方式可以滿足所有不同種類的使用需求(如圖 2 所示)，但在適當的設計下，它依舊可以被最佳化滿足設計中有限種類的使用族群。除此之外，藉由所設計的階層樹狀結構，我們同時提出一個可根據使用者環境參數的限制，而在瀏覽樹狀結構(Tree Traversal)的過程中決定一條路徑以做為傳輸流程表(Schedule)之演算法，用以實做伺服器端之傳輸機制。除此之外，適當地在樹狀結構中的末端階層採用位元平面壓縮技術，並動態地選擇樹狀圖中可選擇的路徑，將能同時達到大範圍且細微資料量可調整性的要求。

2.實作高層次複合可擴充性壓縮之解碼器。

我們結合 MPEG-4 所提供的兩個有關可擴充性視訊壓縮的工具集(Profile) Simple Scalable Profile 以及 Streaming Video Profile，以實做一結合兩者優點並根據階層樹狀結構之設定所完成的一可隨使用族群調整的通用可擴充性壓縮方式。實驗中實際的樹狀結構如圖 3 所示，利用這樣一個階層架構以同時服務假設的五類型用戶為主要目標，並可依所擁有的資源做可擴充性的調整。



(a)



(b)

圖 2：可擴充性視訊壓縮中之層次分割方式(a) 格狀擴充方式(b) 階層樹狀結構

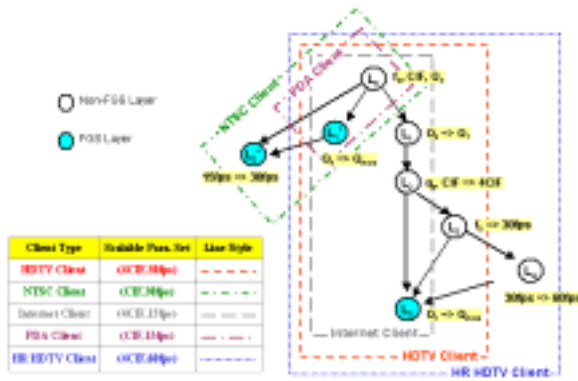


圖 3：實做系統中的視訊擴充結構圖。

3. 人類視覺現象之探討，並研究其與數位視訊壓縮參數之相關性。

無論是傳統視訊壓縮或可擴充性視訊壓縮，視訊解碼後最後的觀察者是人類的雙眼，因此，在壓縮視訊的過程中，當欲決定所欲處理視訊的壓縮參數時應以改善最後的視訊觀察結果為依歸。我們根據一些觀察到的人類視覺現象，制定出能改善畫質的壓縮策略，以便在階層壓縮的過程中，決定更好的壓縮參數，以有效地使用有限的頻寬。這些相關的參數包括有根據視訊區域的特性，評估人類視覺的區分敏感程度(見下一節)、視訊時間解析度之重要性(見下兩節)、不同頻率訊號的重要性等。

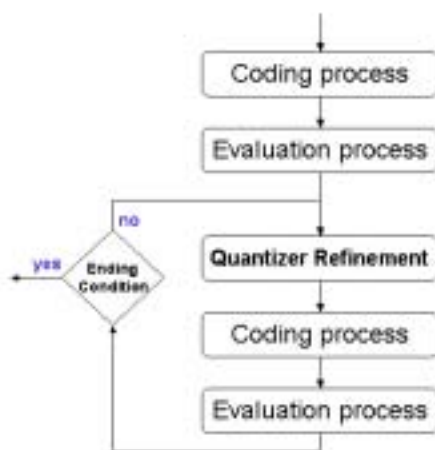


圖 4：利用遞迴演算步驟，逐步改善影響視訊區域畫質參數之機制。

4. 人類視覺為導向的區域畫質分析工具。

人眼對視訊訊號有所謂的遮蔽效應(Masking)與促進效應(Facilitation)。也就是說，一個視訊訊號可能因為周圍的訊號具有某些特質，而減弱了或增強了這個訊號對雜訊的抵抗能力。這種現象將使得不同種類與區域的視訊，對失真現象(Distortion)有不同的容忍能力。視訊壓縮必須將這種現象考慮在決定壓縮參數的過程當中。在計劃中，我們設計一迴圈式的演算法，並利用 Sarnoff 所提供的相對閾衡量工具(Just-Noticeable-Difference Metric)來尋找一較佳的區域畫質參數之表現方式。圖四簡示其主要構成方塊圖。

5. 人類視覺為導向的視訊連續性及重要性之分析工具。

人眼對視訊空間與時間解析度的觀察能力會因視訊內容的特性而有所不同。因此我們根據視訊內容的平穩度(Smoothness)來決定基層資料流的視訊格數以及畫框其平均所需畫質，進而設計一以視訊特色為基礎之動態視訊格數(Dynamic-Frame-Rate Encoding)壓縮方式。圖 5 簡示依據畫框平穩度決定時間解析度壓縮參數之主要方塊圖。

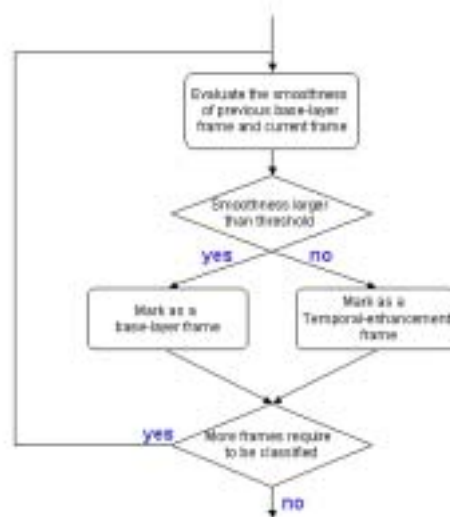


圖 5：配合判斷輸入視訊的平穩度以決定時間解析度相關的壓縮參數之方塊圖。

6. 整合視訊時間與空間重要性之高層次複合可擴充性壓縮程式之雛形。

結合上述兩項壓縮策略，以及通用可擴充性樹狀壓縮結構，我們提出一套整合這些技術之自動運作的壓縮方式，用以將現存的視訊轉換成可串流的通用可擴充性視訊資料流。如圖六所示，傳統資料流之解碼程式將用以導入現存之高畫質視訊資料，並讀入可用之相關視訊特徵參數，這些現有特徵參數之不足之處，將可利用視訊分析工具加以補足，這些特徵參數將可送入通用可擴充壓縮系統，產生高可擴充性之資料流，最後透過系統層，將這可些資料流以及適用於串流使用之特徵參數一同儲存於標準之 MPEG-4 檔案格式中。透過藉由這套系統，可以減輕視訊服務端對於耗時的壓縮過程所支付的成本，以期能以較低的成本製作出高效率的視訊資料流，進而提昇此系統之實用性。

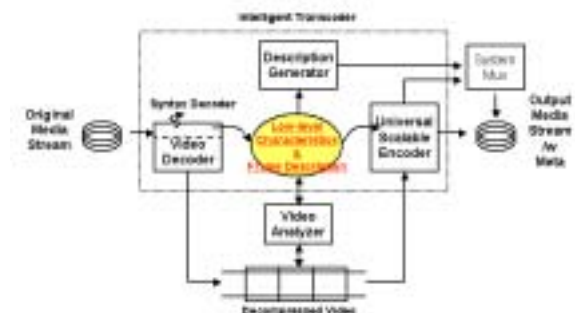


圖 6：通用可擴充性壓縮之壓縮系統之方塊圖。

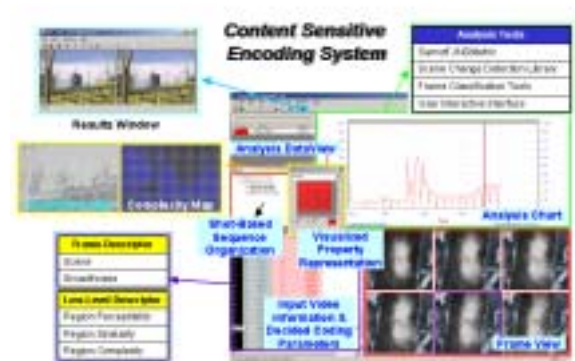


圖 7：通用可擴充性壓縮之使用者介面以及其功能與技術之示意圖。

7. 視訊特徵視覺化之顯示與互動之使用者介面。

視覺相關的畫質評估最終的觀察者是人眼，因此在

選定視覺分析參數後，我們需要一個視覺化的工具，將這些參數用符合觀看者感覺的有效方式顯示，如此一來，藉由使用者的干預，我們才能有機會找到較佳的自動參數決定函數。同時對自動函數不足之處，也提供應用程式一個手動補救的機會。本系統提供一個以 MPEG-4 最小壓縮單元 (Macroblock) 為顯示單位的複雜度及壓縮畫質參數的顯示設定介面，如圖 7 所示。

8. 以 IP 網路串流架構為基礎之可擴充性資料流視訊傳輸設計

通用可擴充視訊流可適用於多種異質網路以及頻寬頻寬變異之環境，利用 MPEG-4 檔案格式所支援的多層資料流，本計畫將設計相對應之串流伺服器串流架構將採符合 MPEG-4 第八部分之 MPEG-4 在 IP 網路上傳送之應用 (MPEG-4 over IP Networks)。同時，為達成對客戶端的適應，客戶端的設定與頻寬的偵測將在連線建立時與連線當中持續被偵測，以達到將在使用者異質環境與變異頻寬下最佳的顯示結果。整合數位視訊內容的產生、儲存、傳輸以及消費的預定架構以及技術內容如圖八所示。

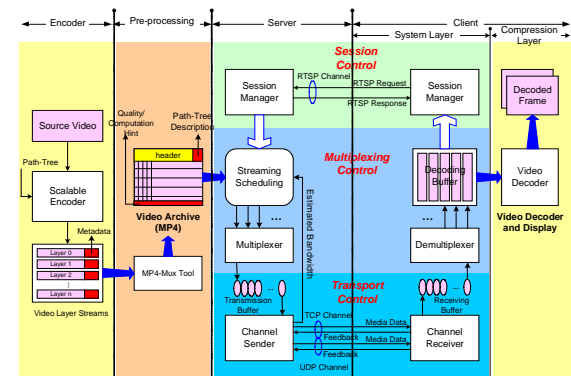


圖 8：通用可擴充性壓縮之串流伺服器與內容資料產生與消耗之流程示意圖。

四、結論與討論

視訊的可擴充性是解決使用端異質性的有效方法。近幾年來，在壓縮技術的逐步改善與運算能力所需要的單位成本越來越低等的客觀環境變化下，我們有理由相信，當客戶端異質性變化的曲線與可擴充性視訊解碼的成本達到某種平衡後，可擴充性視訊資料流將成為熱門的技術話題。以現今的

環境為例，隨選視訊在網際網路與個人電腦平台上的使用逐漸普及，隨著處理器的運算能力大幅提昇，我們可以大膽預期可擴充性視訊壓縮將首先被網際網路與個人電腦所採用。

可擴充性視訊壓縮的研究隨著新的壓縮技巧的發明終將愈趨成熟，本計劃在整合新的壓縮方式並改善可擴充性壓縮的同時，也針對人類視覺對視訊訊號感受能力相關研究做進一步地討論，因為我們知道，一個好的壓縮結果，一定是要了解觀察者的感受，並將視訊壓縮參數做最適應觀察者的調整，以達到理想的呈現效果。

目前，MPEG 正在著手制訂一可擴充壓縮方式，作為將來網際網路上的可擴充壓縮視訊媒體壓縮方式，本計畫也將持續注意相關之發展，進而尋求互補長短之可能改善方式。

五、相關論文之發表狀況

1. Yi-Shin Tung, Po-Kang Hsiao, Ja-Ling Wu, and Chi-Hui Huang, "A Highly Scalable Video Service based on MPEG-4 Spatial-Temporal-FGS Scalable Codec and Hybrid TCP/UDP Transmission," *Proc. of 3rd Workshop and Exhibition on MPEG-4*, 2002.
2. Yi-Shin Tung, Ja-Ling Wu, Wen-Huang Cheng and Ting-Jian Pan, "Encoding Strategies for Realizing MPEG-4 Universal Scalable Video Coding," *Proc. of SPIE Visual Communications and Image Processing*, 2003. (VCIP2003)
3. Yi-Shin Tung, Ja-Ling Wu and Po-Kang Hsiao, "The Design and Implementation of an MPEG-4 Universal Scalable Video Codec Based on Configurable Layered Path-Tree Structure," submitted to *IEEE Trans. on Circuits and Systems for Video Technology*.
4. Sheng Ho Wang, Sung-Wen Wang, Yi-Chin Huang, Yi-Shin Tung, Ja-Ling Wu, "Boundary-energy sensitive visual de-blocking for H.264/AVC coder," *49th SPIE Annual Meeting*, 2004.
5. Sung-Wen Wang, Ya-Ting Yang, Chia-Ying Li, Yi-Shin Tung, Ja-Ling Wu, "An optimization of H.264/AVC baseline decoder on low-cost TriMedia DSP processor," *49th SPIE Annual Meeting*, 2004.

六、參考文獻

- [1] "Information Technology – Coding of Audio-Visual Objects," *ISO/IEC 14496-2:2001*, Second Edition, International Organization for Standardization, 2001.
- [2] "Streaming Video Profile," *ISO/IEC 14496-2:2001/Amd 2:2002*, International Organization for Standardization, 2002.
- [3] J. Lubin, "A HUMAN VISION SYSTEM MODEL FOR OBJECTIVE PICTURE QUALITY MEASUREMENTS," *Proc. of IEEE International Broadcasting Convention*, pp. 12-16, September 1997.
- [4] "JNDmetrix-IQTM Reference Guide, Version 7.6", Sarnoff Corporation, May 2001.
- [5] "Scalable Video Model Version 1.0," *ISO/IEC JTC1/SC29/WG11 Tech. Contribution Doc. 6372*.
- [6] 其他參考文獻請參考計畫書 (NSC 91-2213-E-002-089)。