

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

無線串流多媒體之新生代可調式聲視訊編碼技術研究(3/3)  
**Emerging Scalable Audio/Video Coding Technologies for Wireless Streaming Multimedia (3/3)**

計畫類別： 個別型計畫  整合型計畫

計畫編號：NSC 96-2221-E-009-063

執行期間：96年8月1日至97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

計畫參與人員：張峰誠、林鴻志、蔡家揚、陳威年、葉尚諭、賴辰彥  
交通大學電子工程學系 博士後研究員以及研究生

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 97 年 10 月 25 日

# CONTENTS

A. 中文摘要.....	1
B. Abstract (英文摘要).....	2
C. H.264/AVC 編碼之平行化演算法與於 NVIDIA CUDA 之實現.....	4
C.1 研究目的.....	4
C.2 文獻探討.....	4
C.3 研究方法.....	4
C.4 結果與討論.....	6
D. H.264/AVC 可調式編碼版本之快速演算法.....	9
D.1 研究背景與動機.....	9
D.2 編碼器架構介紹.....	10
D.3 相關文獻探討.....	11
D.4 研究分析.....	11
D.5 快速演算法設計.....	13
D.6 實驗結果與討論.....	16
E. 適用於可調式影像/視訊之小波轉換參數訊源模型研究.....	18
E.1 研究目的.....	18
E.2 文獻探討.....	18
E.3 研究方法.....	19
E.3.1 影像與視訊中的小波分析方法.....	19
E.3.2 所提出之機率模型.....	19
E.3.2.1 $\rho$ -GGD source model.....	19
E.3.2.2 Piecewise linear estimation for the shape parameter.....	21
E.3.2.3 Modeling accuracy evaluation.....	22
E.4 結果與討論.....	22
F. 參考文獻.....	25
G. 計畫成果自評.....	28
H. Publications: (含前兩年同一系列計畫產出，於過去一年中發表者).....	29

# 行政院國家科學委員會專題研究計畫成果報告

## 無線串流多媒體之新生代可調式聲視訊編碼技術研究(3/3)

### Emerging Scalable Audio/Video Coding Technologies for Wireless Streaming Multimedia (3/3)

計畫編號: NSC 96-2221-E-009-063

執行期限: 96 年 8 月 1 日至 97 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

計畫參與人員: 張峰誠、林鴻志、蔡家揚、陳威年、葉尚諭、賴辰彥

交通大學電子工程學系 博士後研究員以及研究生

#### A. 中文摘要

多媒體服務 (Multimedia Service) 咸信為寬頻無線網路的最重要推手。在無線寬頻網路上傳輸即時多媒體資訊，面臨許多挑戰，例如封包遺失 (Packet Loss)，網路與接收機非同質性 (Heterogeneity) 等特性。因此，可調式視訊編碼法 (Scalable Video Coding) 目前正熱烈發展中。本計畫之主要目標在可調式聲視訊編碼法技術。本計畫全程為三年，本報告將摘要敘述第三年度的成果，主要有三項：「H.264/AVC 編碼之平行化演算法與於 NVIDIA CUDA 之實現」、「H.264/AVC 可調式編碼版本之快速演算法」、以及「適用於可調式影像/視訊之小波轉換參數訊源模型研究」等。

由於顯示處理器的快速發展，近年來漸漸發展出將顯示處理器應用於非圖形的運算，以輔助中央處理器的模式，此技術通稱為 GPGPU。美國 NVIDIA 公司在 2007 年提出一個全新的顯示處理器架構，其全名為「統一運算單元架構」，簡稱 CUDA，內含數個運算核心 (Multi-cores)。對運算能力要求極高的資料密集型應用程式，提供了具彈性的大型平行運算平台。由於 H.264/AVC 編碼版本極為複雜，因此在此計畫中我們針對 H.264/AVC 編碼系統發展平行化演算法，並且將其重要編碼區塊架構於 NVIDIA CUDA 平台之上，以做為對可調式視訊編碼平行化處理的前期實現。

以往多媒體壓縮多採單一層次編碼，這是因為壓縮標準制定時，即為其預設目標的應用規格。然而在異質環境中，不同節點不僅運算能力差異極大，網路頻寬也可能相差甚遠，造成單一層次壓縮方法面臨應用上的瓶頸。近年來 H.264/AVC 可調式視訊編碼版本發展即是為了解決這方面的問題，然而 H.264/AVC 可調式視訊編碼採多層次編碼，且層與層之間有極為複雜的運算。因此在本計畫中我們針對 H.264/AVC 可調式編碼版本發展快速演算法，可大幅降低運算量，並且維持近似原有的畫質。

除了以 DCT 為編碼基礎的 H.264/AVC 可調式編碼方式外，尚有以小波轉換為基礎的編碼方式。小波視訊編碼具有許多發展潛力，如具有完整之嵌入式編碼資料流以及在 HD 視訊中畫質較好的特色，目前為可調式視訊編碼的重要研究領域。在本計畫中，我們對在影像/

視訊中小波轉換後的參數機率分佈進行模擬分析，提出新型態機率模型。此機率模型較傳統 Laplacian 模型有較高之準確度並且同時具低複雜度。藉由此模型，期望對小波編碼之 R-D 關係有更進一步了解，並將之應用於畫面間小波轉換可調式視訊編碼研究中。

**關鍵詞：**MPEG、H.264、可調式視訊編解碼、多核心架構、平行化運算、快速演算法、畫面間小波轉換視訊

## **B. Abstract (英文摘要)**

Multimedia service is believed to be the driving force for developing wide-band wireless communication systems. However, there are a number of challenges in delivering real-time multimedia over wideband wireless networks such as packet loss and heterogeneous mobile receivers. Thus, a new technology for video representation, the so-called Scalable Video Coding (SVC) standard has been developed in recent years. The goal of this project is to study, design, simulate and implement various streaming audio/video algorithms based on the concept of the scalable video coding. The entire duration of this project is 3 years. We will outline the final results of the third year on this report. There are three major items: "Parallel processing of H.264/AVC and the implementation on NVIDIA CUDA platform", "Fast algorithms for H.264/AVC scalable extension", and "Source modeling of wavelet coefficients for scalable wavelet image/video coding".

Because of the recent development of very powerful general-purpose graphics processor (GPGPU), it is now also used for non-graphics applications that require a large amount of computations. NVIDIA proposed a flexible GPGPU architecture called CUDA. In this project, we suggest an efficient block-level parallel algorithm for the variable block size motion estimation (ME) in H.264/AVC with fractional pixel refinement on the CUDA platform. We decompose the H.264 ME algorithm into 5 steps so that we can achieve highly parallel computation with low external memory transfer rate. Experimental results show that, with the assistance of GPU, the processing time is 12 times faster than that of using CPU only.

The conventional video coding standards adopt the single-layer coding scheme because they are designed for specific applications. However, different clients have different computing capabilities and transmission bandwidth. Thus, the single-layer scheme does not provide satisfactory solution for the heterogeneous network environment. Therefore, H.264/AVC scalable extension is constructed and standardized to solve these problems. The H.264/AVC scalable extension adopts a

multi-layer coding structure and has very complicated prediction methods between layers. Hence, the encoding procedure is highly complicated. To reduce the complexity, we develop fast algorithms for H.264/AVC scalable extension while these schemes produce similar video quality.

In addition to the DCT-based H.264/AVC scalable coding system, the wavelet-based video coding structure produces fully embedded scalable bitstreams and better high-definition video quality. In this project, we first collect wavelet coefficients and propose a new source model after careful analysis. The probability distribution of wavelet coefficients has been previously modeled as Laplacian or generalized Gaussian distribution (GGD). The Laplacian model is simple in calculation but not accurate; on the other hand, GGD is accurate but requires a very complicated modeling procedure. We reformulate the GGD source model and propose a new parameter  $\rho$  in it, where  $\rho$  is the probability of zero-value coefficients. Moreover, a piecewise linear approximation method is proposed to further reduce complexity. Our experiments show that the  $\rho$ -GGD model has high accuracy and consistent performance for modeling the wavelet coefficient pdfs for both spatial 2-D DWT and interframe wavelet video cases.

**Keywords:** MPEG, H.264, Scalable video coding, multi-core architecture, parallel computation, fast algorithm, interframe wavelet video.

## C. H.264/AVC 編碼之平行化演算法與於 NVIDIA CUDA 之實現

### C.1 研究目的

傳統的電腦中央處理器 (CPU) 或是數位訊號處理器 (DSP) 大多為單核架構，對應於視訊的編碼時也都以序列處理 (sequential processing)，因此過往的視訊編碼演算法也都是以序列處理的方式來做編碼。然而 NVIDIA 統一運算單元架構 (CUDA) 則完全的改變了這樣的運算方式與結構。統一運算單元架構是圖形運算處理器 (GPU) 的一種延伸架構，因此也承襲了圖形運算處理器的多核心架構，可同時處理複數筆資料，也就是以平行處理 (parallel processing) 的方式來提升運算效率。

H.264/AVC 視訊編碼是以宏塊 (macroblock) 為基本單位，並且使用動作預測 (motion estimation) 與補償 (motion compensation) 或是幀內預測 (intra-prediction) 的方式來降低宏塊的資訊量，以達到壓縮的效果。其天性非常適合以平行化的方式來做運算，讓不同的宏塊可以同時的做運算，以提升編碼的速度。因此在本計畫中，我們設計出一套適合於統一運算單元架構的平行預算編碼演算法，並且實現於 NVIDIA 所推出之統一運算單元架構處理器上。

### C.2 文獻探討

傳統用圖形運算處理器處理非圖形運算的技術，統稱為 GPGPU (General-purpose computation on GPUs) [1]。過去也有提出將使用 GPGPU 的技術來加速視訊編碼，如 [2][3][4]。其原理也是將不同宏塊的動作預測利用圖形運算處理器的多重核心來平行處理。然而過去的圖形運算處理器，其晶片內部缺乏了分享記憶體 (shared memory) 的機制，因此運算過程中的暫時資料必須存放於顯示卡上的外部記憶體中 (DRAM)。由於視訊壓縮的資料量非常大，因此資料的傳遞成為一個速度提升的瓶頸。

除此之外，也有相關的論文[5]探討幀內預測的平行演算法，其平行化的方式為將宏塊中不同的幀內預測模式 (四種或九種) 同時運算，以提升編碼速度。然而此種方法會沿生出兩個問題：(1) 由於每一個宏塊的預測模式只有四種或九種，但是圖形處理器的核心數最高可以達到上百個，因此平行化的幅度不夠大，因而導致大部分的處理核心仍然是處於閒置狀態 (idle)。(2) 由[6]可得知，不同的幀內預測模式其計算量差異很大，因此不同的模式同時計算的時候，會導致總預測時間會被計算量最大的幀內預測模式給限制，而導致效能的提升也受到限制。

由於統一運算單元架構是全新的處理器架構，其應用方式與特性也和過去的圖形運算處理器有大幅的改變，因此過去所提出之平行演算法亦無法直接應用於統一運算單元架構之上。多數的演算法必須做大幅度的修正，才能符合統一運算單元架構的應用要求。

### C.3 研究方法

在本計畫中，我們首先將 H.264/AVC 中計算量最大的兩個部分：動作預測與幀內預測兩個部分於統一運算單元架構上實現，分別如以下所述：

## □ 幀內預測

由於每個宏塊的幀內預測是使用該宏塊鄰近已編碼完成並解碼回來的宏塊，來做預測，因此宏塊與宏塊間會產生依賴性 (inter-block dependency)，若要每個宏塊平行做運算的話，會導致預測所需的像素 (pixel) 尚未運算完成的情況。為了解決這個問題，我們提出使用原始影像來做幀內預測的演算法，也就是在預測宏塊資料的時候使用原始的像素資料來做計算並且比較每個模式的 SAD (sum of absolute difference)，以找到 SAD 最小的模式。因此每個宏塊便可以同時的做預測並且產生出最好的模式。而最後實際在編碼的時候，再使用重建後的資料來做編碼。如Fig. C-1之流程圖所示。

## □ 動作預測

為了符合統一運算單元架構的特性，我們設計了一套五個步驟的全域搜尋 (full search) 演算法，並且包含了 H.264/AVC 編碼標準中所具備的多重方塊大小動作補償 (variable block size motion estimation)，而動作向量 (motion vector) 精細度達到四分之一像素。五個步驟如Fig. C-2所示。在此演算法之中，我們對於記憶體的配置與使用做了最佳化，讓運算中的暫時資料可以在處理器中的分享記憶體儲存並且重複利用，以節省處理器與外部記憶體間的頻寬使用量。

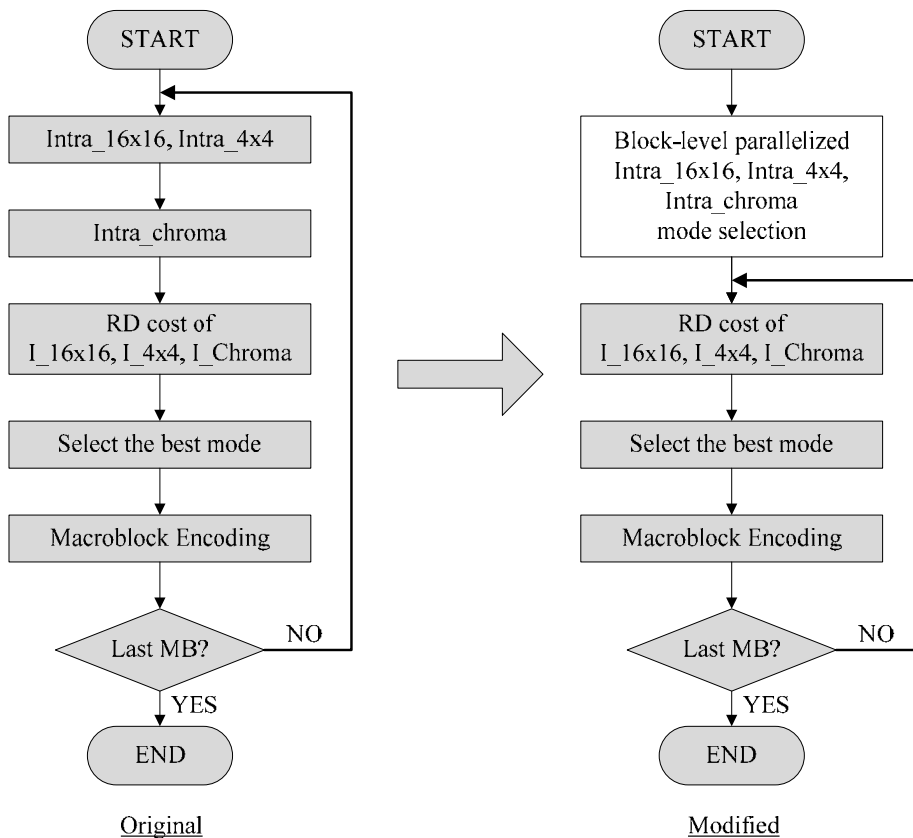


Fig. C-1. 使用原始影像作幀內預測之演算法流程圖

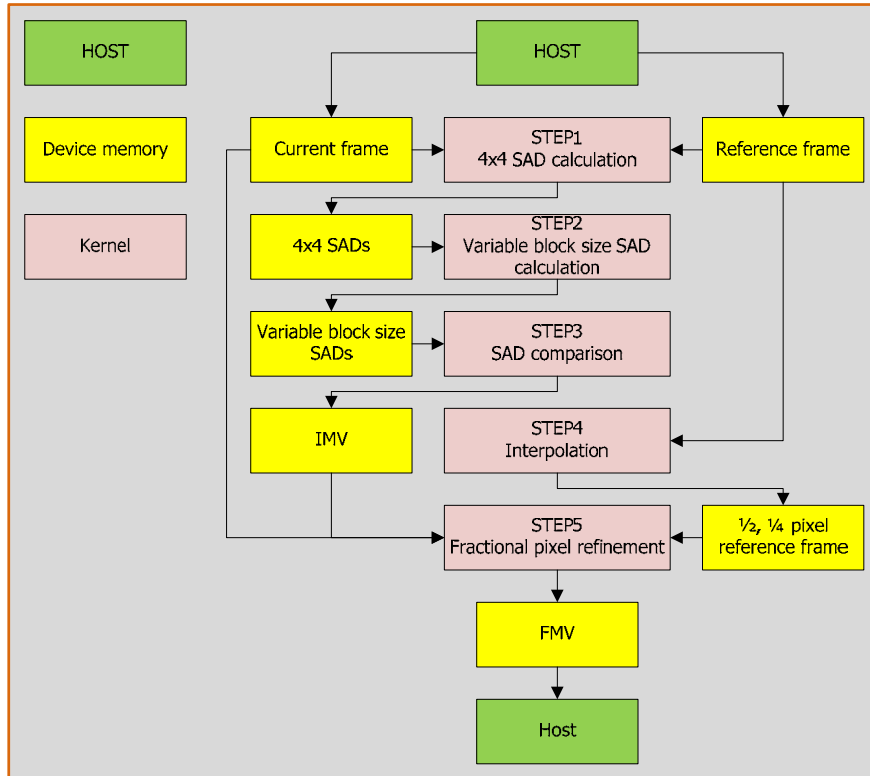


Fig. C-2 五個步驟的動作預測流程圖

## C.4 結果與討論

使用原始影像作幀內預測，雖然可以達成宏塊的平行化，然而卻會因為預測資料與實際資料的不相同，會導致編碼效能下降，如Fig. C-3。由於在一般的應用之下，使用幀內預測與使用動作預測的比例差距很大，大部分的宏塊都會使用動作預測來做編碼，因此使用原始影像作幀內預測的效能下降有限，甚至在大部分的測試影像中，其編碼效能可以逼近使用重建影像作幀內預測的演算法。

而使用平行化的動作預測時，由於每個預測方塊之動作向量預測子（motion vector predictor）需要設置為零，才能消除每個預測方塊之宏塊間依賴性，因此編碼效能也會有些許的下降，如Fig. C-3。

我們所使用的測試平台如Table C-1所示，統一運算單元架構處理器本質上也是一個顯示晶片，因此是建立在個人電腦上，而在需要使用顯示晶片運算的時候，資料會從個人電腦的記憶體中傳送至顯示卡上的記憶體。而運算完畢之後，也是將資料從顯示卡上的記憶體傳送回個人電腦之記憶體中。使用統一運算單元架構來處理幀內預測與動作預測的如Table C-2與Fig. C-4所示，在這兩個部分之中分別都可以達到近 12 倍的加速。另外在幀內預測中若只使用單一個核心來做編碼，其處理時間會比只使用個人電腦來做運算慢了 500 倍，因此只若使用統一運算單元架構卻沒有搭配對應的平行編碼演算法，不但無法有效的做加速，反而會拖慢整體的運算時間。



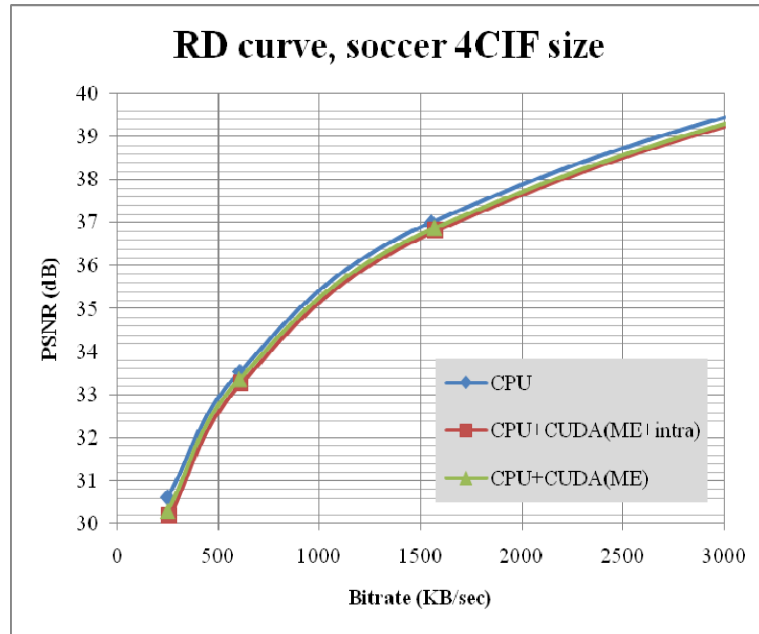


Fig. C-3. 使用平行演算法與原始編碼演算法之效能比較

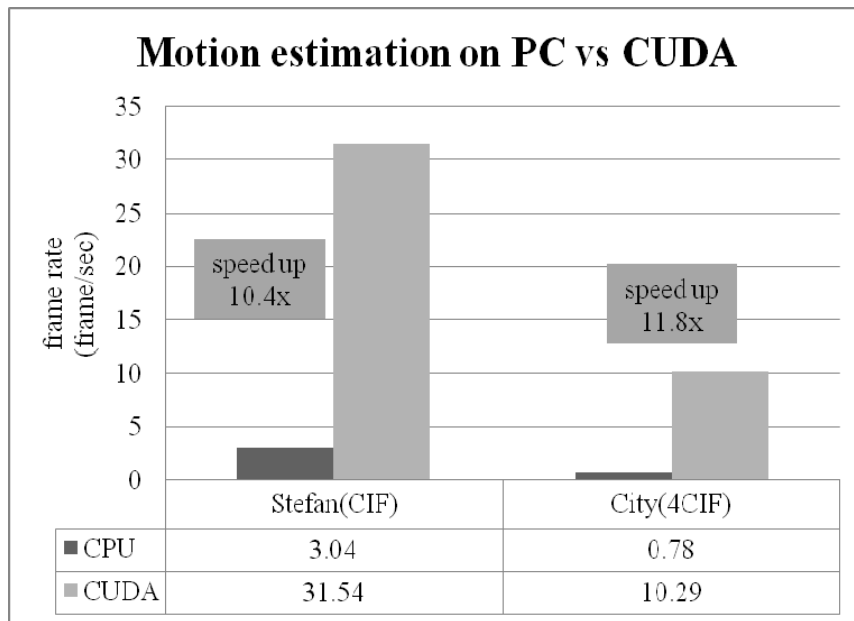


Fig. C-4. 動作偵測於統一運算單元架構之加速效果

Table C-1 統一運算單元架構測試平台

PC	CPU	AMD Athlon 64X2 2.1GHz	
	memory	2,048MB	
	OS	Microsoft Windows XP sp2	
	Compiler	Microsoft Visual Studio 2005	
CUDA	GPU	NVIDIA GeForce 8800GTX	
		Core Clock	575MHz
		Processor numbers	128
		Memory Clock	900MHz
		Memory	768MB
	Memory Bandwidth	86.4(GB/sec)	
	CUDA Toolkit and SDK 1.1		NVIDIA Driver for Microsoft Windows XP with CUDA Support (169.09).

Table C-2 幀內預測於統一運算單元架構之加速效果

	PC (msec)	CUDA (msec)	Speed up	CUDA using one thread(msec)
Intra_16x16	4.06	0.63	6.4x	265
Intra_4x4	15.94	0.92	17.3x	753
Intra_chroma	2.19	0.31	7.1x	93
Total	22.19	1.86	11.9x	1112

## D. H.264/AVC 可調式編碼版本之快速演算法

### D.1 研究背景與動機

如Fig. D-1，為了滿足串流系統在網路上的多樣應用，MPEG 與 ITU 標準小組發展出了一個可調式視訊編、解碼系統(Scalable Video Coding, SVC)[7][8]。其目的是希望能以單一位元流來滿足各種接收機的傳輸頻寬、運算能力與解析度。目前，列入標準的整體架構是德國 HHI 提出以 MPEG-4 14496-10 精進視訊壓縮標準做為延伸的技術。此視訊壓縮標準已於最近制訂完成，以 H.264/AVC 視訊編碼標準[9][10]為基本層架構，可同時提供空間上(Spatial)、時間上(Temporal)、和畫質精細(Quality)可調的特性，其壓縮端參考架構命名為 Joint Scalable Video Model (JSVM)，其編、解碼技術可用在各式不同的設備儀器上，小至手執式之畫面接收器，大至高解析度接收機[11]。

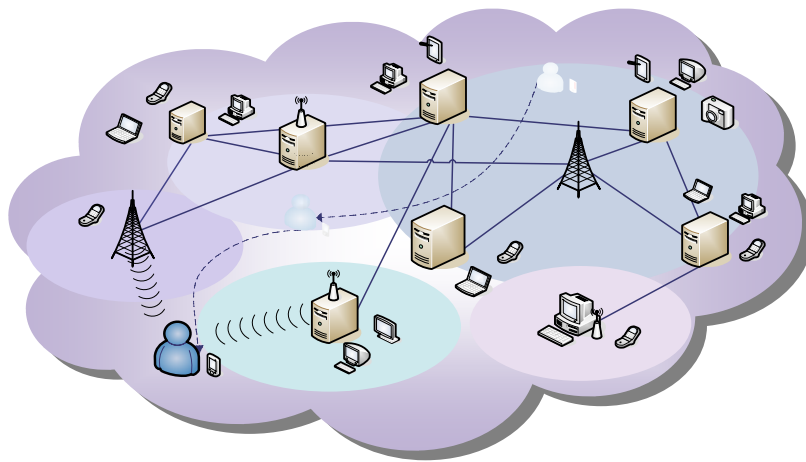


Fig. D-1 可調式視訊編碼應用環境

由 HHI 所提供之標準參考軟體(Reference Software)，稱為 **J**oint **S**calable **V**ideo **M**odel (JSVM)[12]，採用全域搜尋(Exhaustive Search)之方式找尋到最佳之編碼切割模式(Partition Mode)。我們測試兩組典型的視訊影像(FOREMAN 與 CITY)，統計出加強層(Enhancement Layer)的運算複雜度約略是基本層(Base Layer)的 2.5 倍到 3.3 倍，如Table D-1。這代表了當我們的編碼層數增加時，會使得編碼時間大幅度提高，導致可調式視訊編碼在現階段無法被廣泛地應用。因此，在本計畫中，針對目前可調式視訊編碼的運算複雜度，進行相關研究與文獻比較，並說明未來可能的研究方向。

Table D-1. 編碼基本層與加強層之複雜度比較

Setup	$(Qp_B, Qp_E)$	FOREMAN		CITY	
		Ref. = 1	Ref. = 3	Ref. = 1	Ref. = 3
CGS	(40,34)	1:3.08	1:2.49	1:3.22	1:2.52
	(40,30)	1:3.10	1:2.51	1:3.23	1:2.53
	(30,24)	1:3.12	1:2.50	1:3.28	1:2.52
	(30,20)	1:3.13	1:2.51	1:3.29	1:2.52
<b>AVG.</b>		<b>1:3.11</b>	<b>1:2.50</b>	<b>1:3.26</b>	<b>1:2.52</b>

GOP size = 16

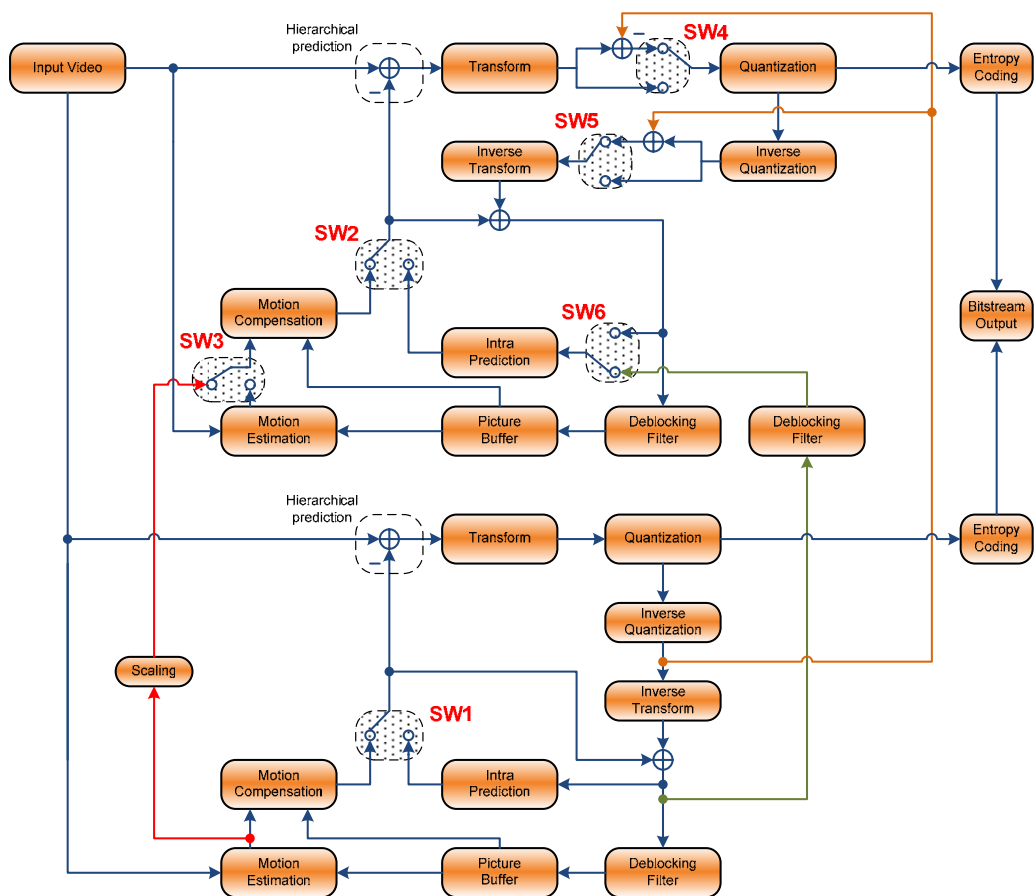


Fig. D-2 JSVM Encoder 架構示意圖

## D.2 編碼器架構介紹

Fig. D-2為一 JSVM 編碼架構之示意圖[13]-[15]，圖中的每一層均可視為一 H.264/AVC 編碼器。為了要增進編碼效率，每一層之間都會利用 Inter-layer Prediction 之機制來使編碼效能改善，詳細說明如下，而

Table D-2說明Fig. D-2中每一個 SW 之功能：

- Inter-layer Intra Texture Prediction：加強層中的宏塊，若其相對應之基本層宏塊是使用內部空間上的預測作編碼時，則可使用 IntraBL 模式作編碼。
- Inter-layer Motion Prediction：加強層之宏塊的切割模式(Partition Mode)、參考畫面索引(Reference Frame Index)與動作向量(Motion Vector)可以利用其相對應之基本層宏塊所擁有的資訊推算得知。
- Inter-layer Residual Prediction：基本層所重建回來之 DCT 數值可以用來減少加強層中編碼所需之位元數。

**Table D-2. Functionality of Switches**

Switch number	Functionality
1,2	Intra/Inter mode selection
3	Inter-layer motion prediction
4	Inter-layer residual prediction (transform domain)
5	Reconstruction at the enhancement layer
6	Inter-layer texture prediction

### D.3 相關文獻探討

發展可調式視訊編碼標準前，已有許多針對 H.264/AVC 編碼標準所設計之快速演算法 [16]-[24]。然而，目前針對可調式視訊編碼標準所設計之快速演算法相關文獻並不多，以下我們分為內部預測(Intra Prediction)與畫面間預測(Inter Prediction)來說明現有文獻之作法及缺失：

- 內部預測(Intra Prediction)：在[25]與[26]兩篇文獻當中，指出在加強層中的 Intra16x16 與 Intra8x8 兩種預測模式在編碼過程中即使被完全的忽略不作測試，並不會會有顯著的編碼效能損失。根據文獻指出，加強層中超過 90%的宏塊均是採用 Intra4x4 與 IntraBL 為其預測模式。再者，由於 Intra4x4 預測模式中有九種預測方向可供選擇，其運算量為所有內部預測模式中最高之模組，所以在[25]中又將此九種預測方向減少剩水平、垂直與 DC 預測模式，已達到複雜度降低之目的。
- 畫面間預測(Inter Prediction)：在[27]與[28]兩篇文獻中，均是利用層狀架構之間的運算相關性，提出了針對可調式視訊編碼之快速演算法，其演算法之主要精神為加強層中之宏塊切割必須比基本層中之宏塊切割模式來的更精細，類似於[25][26]，其內部預測的模式只測試 Intra4x4 與 IntraBL 兩種預測模式。
- 現有方法之缺失與不足：在[25]-[28]數篇文獻中，均只考慮到層狀架構為兩層(一基本層與一加強層)時之情況作快速演算法之設計，並未考慮到當層狀架構超過兩層時，所應該考慮到之問題。再者，在設計演算法上也必須考量到多重參考畫面(Multiple Reference Frames)之需求。雖然是利用層狀架構之間的運算相關性作為減少運算量之依據，而層狀架構之相關性不僅僅只有切割模式的精細度而已，仍有動作向量的相關性尚未被考量到。再者，在[27]與[28]兩篇文獻中提到所設計之演算法必須對於每個宏塊都測試 Intra4x4 與 Inter8x8，事實上，這兩個預測模式是整個編碼器的運算複雜度瓶頸所在，因此，前人之工作所發展之演算法加速程度會有所限制。

### D.4 研究分析

在分析與研究前人之作法後，我們期望能夠設計針對多層層狀架構之編碼系統，盡可能地善用基本層在編碼時所擁有的資訊，藉由層狀架構之間的相關性，達到大量減少編碼時所

需之運算複雜度。統整我們之前的演算法設計[29][29]，以下作了一些數據分析，觀察層狀架構之間所具有之運算相關性。

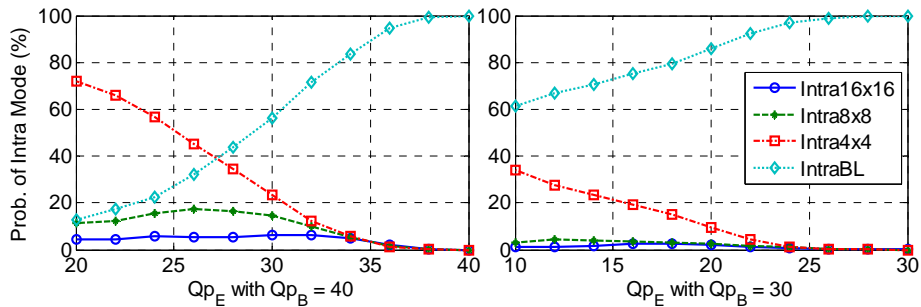


Fig. D-3 Distribution of the Intra Prediction Modes at the Enhancement Layer with CGS Setup

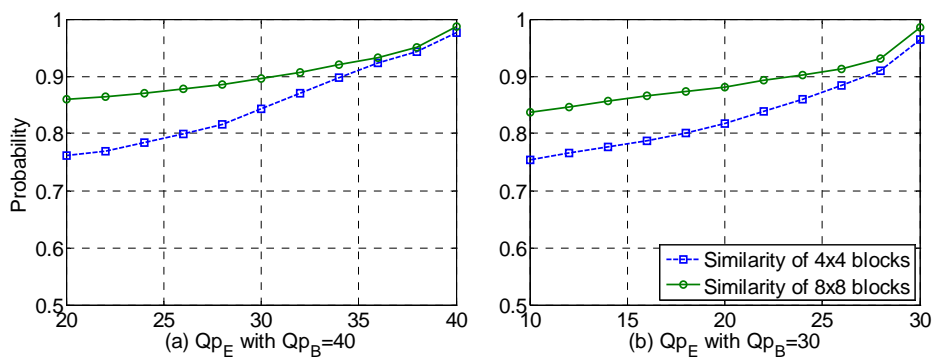


Fig. D-4 Probability Profiles of Similarity for CGS (a) Poor-quality Base Layer (b) High-quality Base Layer

➤ Intra prediction 相關性：

如Fig. D-3所示，加強層中之宏塊所使用之預測模式絕大部分均落在 Intra4x4 與 IntraBL 兩種模式之中。而隨著基本層之畫質變好，使得加強層使用 Intra16x16 或 Intra8x8 之機率大為降低。因此，加強層可以只保留 Intra4x4 與 IntraBL 兩種預測模式，並且維持原有之編碼效能。再者，根據Fig. D-4所示，編碼層與編碼層中相對應的局部區塊所使用的內部預測方向，其相似度在 70% 以上，這項特性可讓加強層中的運算量減少。

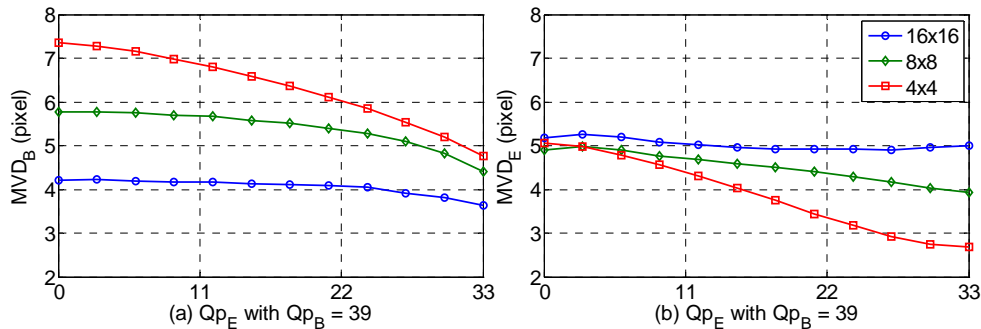
➤ Macroblock partition 相關性，經由分析我們歸納了以下幾點：

- 編碼層與編碼層中相對應的局部區塊，不論基本層與加強層的量化參數差異多大，超過 50% 的比例使用相同的切割模式或是採用 BLSkip 模式。對於 Inter8x8 而言，則是超過了 80% 的比例。
- 除了相同的切割模式與 BLSkip 模式，接著最有可能的兩種切割模式是 Skip 與 Inter16x16，這個觀察與[27][28]兩篇文獻所提的觀點有所不同。
- 如果基本層是用 Skip 模式編碼，則在加強層中所相對應的宏塊只需要測試 Skip 與 Inter16x16。
- 在加強層中，對於最高的兩個 Temporal level 而言，BLSkip 的比例超過 50%。
- 在加強層中，Inter4x4 的比例小於 5%。

- 如果基本層是採用 Inter16x8/Inter8x16 的模式作編碼，則加強層中不太可能使用 Inter8x16/Inter16x8 作為最佳切割模式。

■ 動作向量相關性：

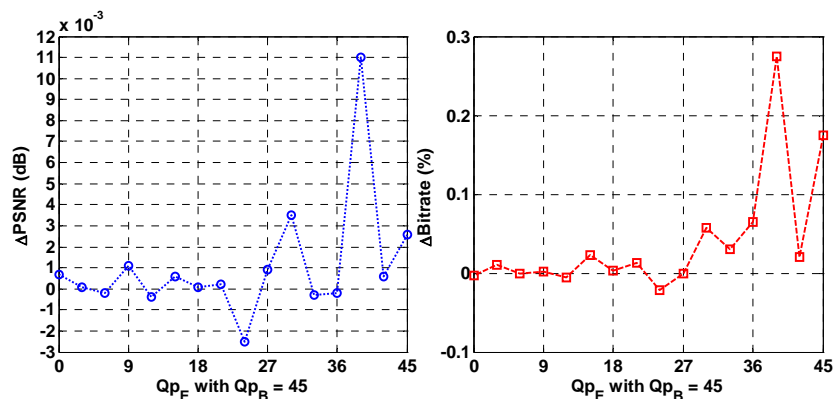
Fig. D-5數據指出，對於 16x16、16x8、8x16 與 8x8 之切割模式，加強層中所用的起始找尋點(Initial search point)為基本層所找尋之最佳動作向量會比標準所制定之動作向量預測值(Motion vector predictor)來的準確。另外，我們可以發現到，平均而言，所找尋出的動作向量與起始找尋點相差不到 10 像素距離，因此在加強層中所設定之搜尋範圍可以比基本層所找尋的範圍來的小。



**Fig. D-5** The motion vector differences using (a) the motion vector of the base layer or (b) the motion vector predictor at the enhancement layer as reference

■ Residual prediction 相關性：

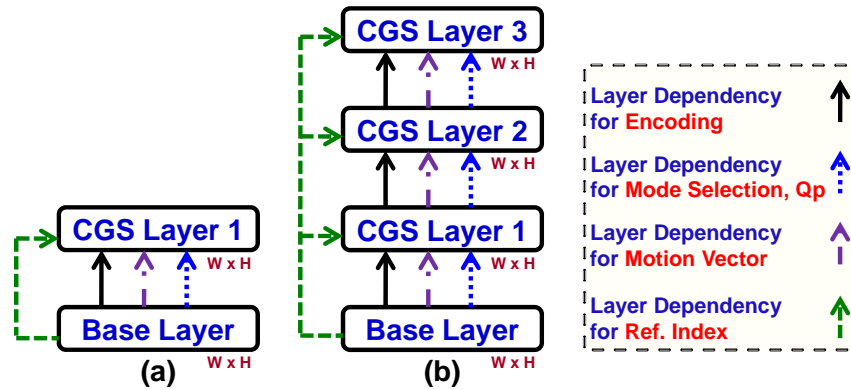
根據Fig. D-6數據，在不同的基本層畫質好壞差異下，啟動 Inter-layer prediction 的機制機率會隨著基本層的畫質變好而提升。然而，我們發現當基本層之畫質不夠好時，啟動 Inter-layer prediction 的機制對編碼效率幾乎沒有改善，因此我們提出在基本層畫質不佳的狀況下，不執行 Inter-layer prediction，以達到減少編碼所需運算量之目的。由其當基本層與加強層之量化參數差異過大時，Inter-layer prediction 的機制對於加強層的編碼效率幾乎沒有改善。



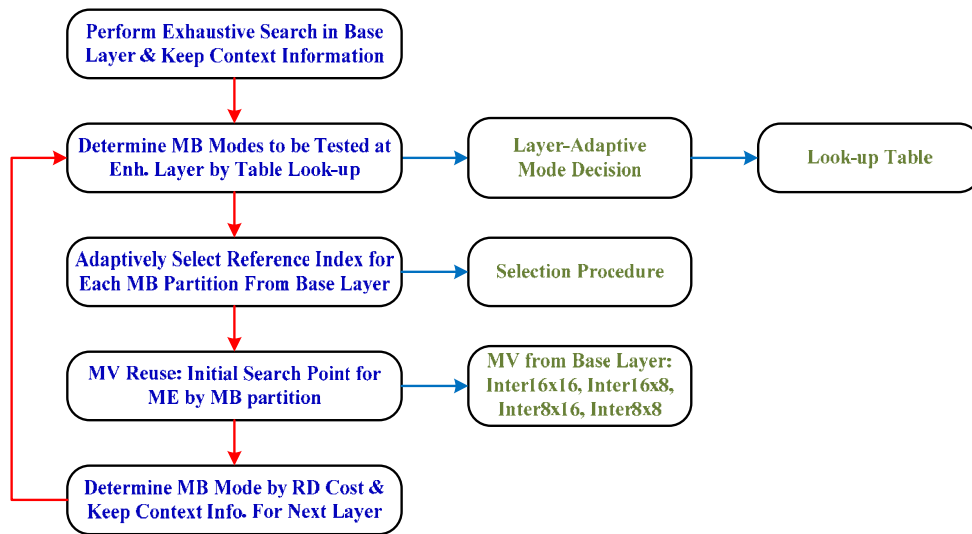
**Fig. D-6** Comparison of Turning off the Residual Prediction in Transform Domain in B Frames

## D.5 快速演算法設計

根據[30]的研究數據分析，我們提出了一套在多層層狀結構之粗略可調特性與時間可調特性的需求設定下，針對 Hierarchical B 畫面之低複雜度宏塊決定模式快速演算法，有效地大量減少內部預測(Intra Prediction)與畫面間預測(Inter Prediction)的龐大運算量，其各編碼層之關聯性，可參考Fig. D-7。我們所提出之演算法流程如Fig. D-8，其中Fig. D-9表示在加強層中模式選擇之流程、Table D-3~Table D-6告知加強層中所需要測試之切割模式、Fig. D-10告知加強層中每一個切割模式所要參照之參考畫面。

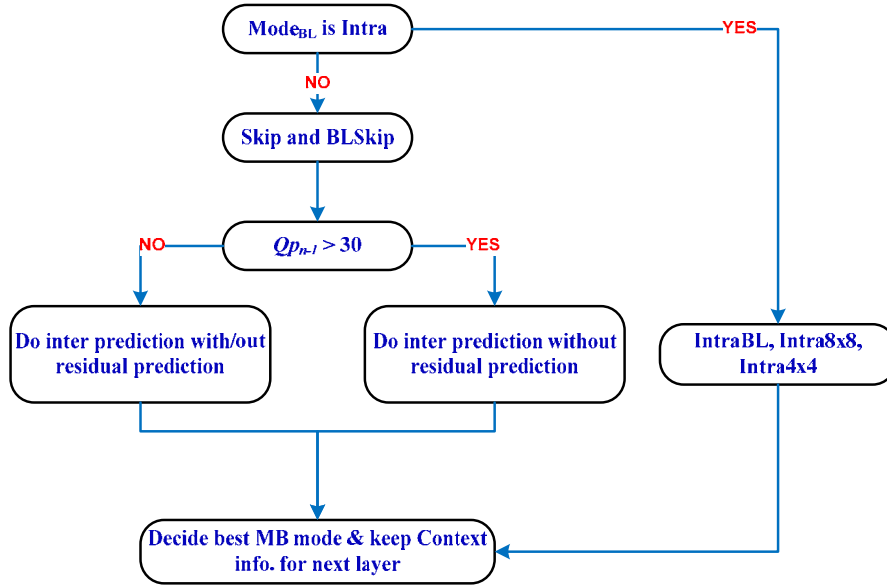


**Fig. D-7** Inter-Layer Dependency Settings of SVC Encoder for CGS (a) 2-Layer Case, (b) 4-Layer Case



**Fig. D-8** Flowchart of Proposed Algorithm





**Fig. D-9** Layer-adaptive Mode Decision

**Table D-3.** Look-up Table for Layer-adaptive Intra Mode Selection

Candidate Modes		Prediction Modes of Reference Layers									
		Intra4x4/Intra8x8									
		0	1	2	3	4	5	6	7	8	
Intra4x4/Intra8x8	0 (V)	●		○			○		○		
	1 (H)		●	○				○		○	
	2 (DC)	○	○	●	○	○	○	○	○	○	
	3 (DDL)				●				○	○	
	4 (DDR)					●		○			
	5 (VR)					○	●				
	6 (HD)					○		●			
	7 (VL)				○				●		
	8 (HU)				○					●	

The candidate mode set is {●} if both the previous two layers are with the same mode; otherwise, the candidate mode set should include {●, ○}.

**Table D-4.** Candidate Modes of Inter Prediction for  $Q_{p_{n-1}} > 30$

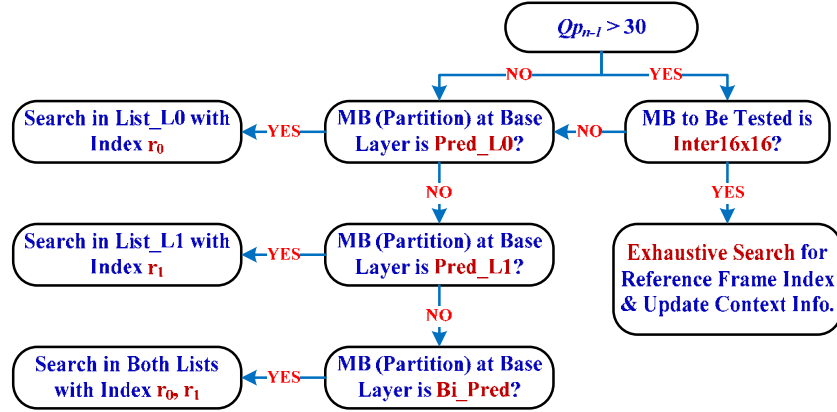
Temporal layer	$T_1 - T_{N-2}$					$T_{N-1} - T_N$				
	Skip	16x16	16x8	8x16	8x8	Skip	16x16	16x8	8x16	8x8
16x16	○	○	○	○	○	○	○	○	○	○
16x8	○	○	○					○		
8x16	○	○		○					○	
8x8					○					○

**Table D-5.** Candidate Modes of Inter Prediction for  $Q_{p_{n-1}} \leq 30$

Temporal layer	$T_1 - T_N$				
	Skip	16x16	16x8	8x16	8x8
16x16		○			
16x8			○		
8x16				○	
8x8					○

**Table D-6.** Candidate Modes of Sub-macroblock of Inter Prediction for all  $Qp$  Values

Temporal layer	$T_1 - T_N$
Skip8x8	○
8x8	○
8x4	-----
4x8	-----
4x4	-----



**Fig. D-10** Layer-Adaptive Selection of Reference Frame Index

## D.6 實驗結果與討論

根據[30]所提出之在多層層狀結構之粗略可調特性與時間可調特性的需求設定下，針對 Hierarchical B 畫面之低複雜度宏塊決定模式快速演算法，我們將之實現在 JSVM 9.11 版之參考軟體上，並且測試了八組測影像，其中四組為 CIF 格式，另外四組是 4CIF 格式，其詳細之編碼參數列表於 Table D-7 中。

在 Table D-8 中，採用  $\Delta$ PSNR、 $\Delta$ Bitrate、TS 與  $TS_E$  四項數據來展現出所提出之演算法的運算複雜度改善與編碼效率差異。實驗數據顯示出所提出之演算法有效地大量減少在多層層狀結構下所需之編碼時間，其減少幅度可以到 84%；另外，在編碼加強層之運算複雜度，平均而言，所提之演算法可以減少 93.4% 運算複雜度，相當於加速編碼加強層之 15 倍速度，並且維持原有之編碼效率。

**Table D-7.** Testing Conditions

Testing Sequences	CIF	AKIYO, FOOTBALL, MOBILE, STEFAN
	4CIF	CREW, HARBOUR, ICE, SOCCER
Encoder Configurations	Software: JSVM_9_11 M.E. Search Range: $\pm 32$ pixels RDO: Enabled GOP Size: 8 Entropy Coding Mode: CABAC Macroblock Adaptive Inter-Layer Prediction: Enabled Machine: Athlon 3800+, 64-bit, Dual-core Processors, 2.0GB RAM with Windows XP	

**Table D-8.** Performance comparisons with JSVM 9.11

Sequences	$\Delta$ PSNR (dB)	$\Delta$ Bitrate (%)	TS (%)	TS <sub>E</sub> (%)
AKIYO	-0.01	0.47	84.2	94.0
FOOTBALL	-0.04	1.51	82.3	93.6
MOBILE	-0.02	0.96	82.4	92.6
STEFAN	-0.01	0.77	82.7	93.3
CREW	-0.01	0.34	83.1	93.2
HARBOUR	0.00	0.23	83.1	93.2
ICE	-0.01	0.58	83.4	93.5
SOCCER	-0.01	0.63	83.5	93.8

未來之方向將著重在空間可調性上的快速演算法設計，希望能夠擴展原有之演算法，延伸至空間可調性上，發展一套可以適用於可調視訊編碼技術標準之低複雜度演算法。

## E. 適用於可調式影像/視訊之小波轉換參數訊源模型研究

### E.1 研究目的

從對訊源的機率模型，可推導出 Rate-Distortion 之良好關係函式，而藉由此關係函式則可進行相關之 Rate-Distortion 分析。因此針對各種壓縮訊源進行準確的機率模型假設是非常重要的。對小波訊號來說，由於經過編碼之小波訊號，具有良好之零值為平均值與左右對稱性，在傳統上以 Laplacian 為機率模型以進行小波訊號之分析，雖然計簡單，但誤差幅度頗大。Laplacian 為廣義高斯分佈之子集合，使用廣義高斯分布可得到較準確之機率模型。然而廣義高斯分布模型的建立時必須以第二階動量(moment)與第四階動量進行參數估計，其複雜度過高；因此在本研究中發展一套以廣義高斯分布為基礎之  $\rho$ -GGD 模型，可以低複雜度準確估測小波訊源之機率分布。

### E.2 文獻探討

在可調式視訊編碼領域中，目前主要有兩大分支：以 DCT 為基礎之 H.264 可調式視訊編碼版本與以小波為基礎之畫面間小波(interframe wavelet)可調式視訊編碼。由於小波訊號本身具有兼具時域與頻域可調解析度的特性，此項特性可充份運用在影像與視訊編碼之中，如可在可調式編碼中調整畫面率與畫面解析度。經過小波轉換後的影像，頻帶(subband)呈現類似階層分布，不同階層展現不同的影像細節；藉由此自然的特性，影像的細節可經過越多層次頻帶訊號的獲取而得到漸近式的改良[31]。

近年來由 Sharpiro [32]， Said 與 Pearlman [33]，以及 Taubman [34] 等人的研究已展現了小波編碼在影像壓縮上的良好表現。而在相關視訊壓縮的進展方面，由於 Ohm [35]，Hsiang 與 Woods [36]，Secker 與 Taubman [37]，以及 Xu *et al.* [38]等人提出了動態補償時間濾波(motion compensated temporal filtering) (MCTF)技術使得小波視訊編碼也得到極大的進展。此外，為了得到適用於內嵌式(embedded)小波影像/視訊編碼架構下最佳的編碼效能，許多 rate-distortion(R-D)分析技巧也被提出[34][39][40]。根據 Shannon 的訊源編碼理論，rate-distortion 函式可經由訊源的機率分布模型推導得到；在過去的文獻中，Laplacian 模型是最被廣泛使用的[41]，而廣義高斯分布模型(Generalized Gaussian distribution)(GGD) [44]也是常被推薦的一種。由於 GGD 有很高的精確度，但要成功建立一個 GGD 模型則有非常複雜的步驟。

在這項研究中，我們引入了一個十分容易測量的參數  $\rho$  至 GGD 模型中， $\rho$  指的是小波參數中的零值機率；此參數為在另一個知名的 R-D 模型中被[42]所提出。借由  $\rho$  值的幫助，我們提出了一個適用於小波參數機率分布的低複雜度  $\rho$ -GGD 訊源模型。為了測量此模型的精確度，我們利用對稱式的(symmetric)Kullback-Leibler divergence [43]來做為測量的標準，並且比較 Laplacian 與  $\rho$ -GGD 之間精確度的差異。在之後的小節中，我們會先針對小波分析的傳統作法討論，並且描述所提出的  $\rho$ -GGD 模型以及對應的低複雜度模型參數求法，最後並展示實驗成果。

## E.3 研究方法

### E.3.1 影像與視訊中的小波分析方法

在 2-D 影像訊號的小波分析中，其實際作法為利用兩次 1-D 的 DWT 進行水平與垂直兩個方向的轉換。每一次的 1-D DWT 都會將 2-D 影像中其一方向的訊號轉換成低通(low-pass)與高通(high-pass)頻帶(subband)，其中的低通頻帶會再經過數次的 DWT 轉換以提高編碼效率。而在畫面間小波視訊編碼中，則是會在時間軸方向進行時間濾波，也就是延著物體動態移動方向進行動態補償時間濾波(MCTF) [35]。經過 MCTF 之後，時間軸上的低通畫面通常會看起來很像數張畫面的平均，而時間軸上的高通畫面則是含有畫面間的差異部份。這些時間軸上低通與高通畫面都會再進一步進行 2-D DWT；因此，再經過畫面間小波視訊轉換後，對於一個 GOP 來說，已經將訊號轉換成了時間-空間頻帶的表現方式。

不管是在影像或是視訊編碼領域，訊源模型一直都是在 R-D 分析中扮演關鍵的角色。小波參數的機率分佈在一些文獻中被用廣義高斯模型(GGD)來模擬[31][44]。然而，為了建立起 GGD 模型，小波訊號的統計變異量與 kurtosis 數值必須被求得以測量 GGD 分佈的型態參數(shape parameter)。

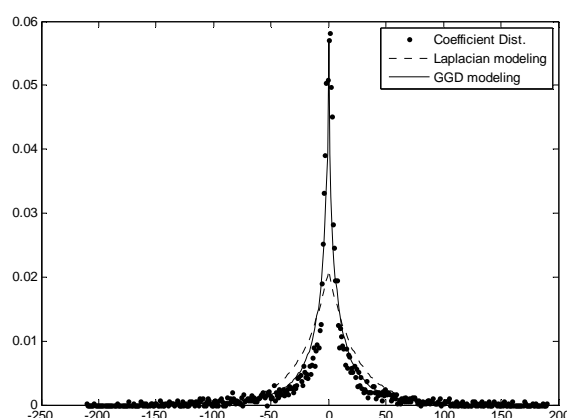


Fig. E-1 An example of wavelet coefficients modeling. (LL-LL-HL subband of image Lena).

然而，由於變異量與 kurtosis 數值是與統計上的二階及四階動量相關，也因此使得建構 GGD 模型變成一個極為複雜的過程。由於複雜度的考量，Laplacian 分佈也常被使用。雖然 Laplacian 訊源模型被廣泛使用，但在許多情況下 Laplacian 模型的不精確情況會發生，如 Fig. E-1 所顯示。因此，我們提出了  $\rho$ -GGD 訊源模型，可達成 GGD 的精確度，但兼具低複雜度的優點。

### E.3.2 所提出之機率模型

#### E.3.2.1 $\rho$ -GGD source model

典型的小波參數的機率分佈具有平均值為零的特性。因此，針對小波參數的廣義高斯模型可設為如下：

$$P_{GGD}(x; \sigma, \nu) = \left( \frac{\nu \eta(\nu, \sigma)}{2\Gamma\left(\frac{1}{\nu}\right)} \right) \exp(-[\eta(\nu, \sigma)|x|]^\nu) \quad (2)$$

其中

$$\eta(\nu, \sigma) = \sigma^{-1} \left[ \Gamma\left(\frac{3}{\nu}\right) / \Gamma\left(\frac{1}{\nu}\right) \right]^{1/2} \quad (3)$$

且  $\sigma$  為小波參數的標準差， $\nu$  為 GGD 模型中的形態參數(shape parameter)， $\Gamma(x)$  則為標準的 Gamma 函數。若令  $\rho$  為小波參數分佈中的零值機率數值。根據(2)， $\rho$  可用以下數學式表示：

$$\rho = P_{GGD}(x=0) = \frac{\nu \eta(\nu, \sigma)}{2\Gamma\left(\frac{1}{\nu}\right)} \quad (4)$$

因此，數學式(2)可以重新寫成以下的  $\rho$ -GGD 表示方法：

$$P_{\rho-GGD}(x; \rho, \nu) = \rho \exp\left(-\left[2\rho \frac{1}{\nu} \Gamma\left(\frac{1}{\nu}\right) |x|\right]^\nu\right) \quad (5)$$

在建立  $\rho$ -GGD 訊源模型的過程中，形態參數  $\nu$  必須要先被求出來。而從數學式(3)與(4)可得知， $\rho$  與  $\sigma$  兩數值的乘積可以下列式子表示：

$$\rho\sigma = \frac{\nu}{2} \left[ \Gamma\left(\frac{3}{\nu}\right) / \Gamma\left(\frac{1}{\nu}\right)^3 \right]^{1/2} = \Phi(\nu) \quad (6)$$

數學式(6)顯示了在  $\rho$ -GGD 模型中， $\rho\sigma$  兩數值的乘積與形態參數  $\nu$  間存在著一個函數的關係，即  $\rho\sigma = \Phi(\nu)$ 。由於  $\rho$  與  $\sigma$  兩個參數非常容易就可以從資料中得到，因此使用此兩數的乘積用來估測  $\Phi(\nu)$  是非常方便的。

從實驗數據中可得知，在影像與視訊的小波轉換參數的分佈中，形態參數  $\nu$  大約介於[0.5, 2.5]這個區間。在Fig. E-2中，實線為在形態參數  $\nu \in [0.5, 2.5]$  之間時  $\Phi(\nu)$  的數值，可以發現  $\Phi(\nu)$  與  $\nu$  在此區間呈現一個逐漸變小且為一對一的關係。因此，在此區間中  $\Phi(\nu)$  的反函數是存在並且也呈現一對一的關係。可利用以上的關係，利用  $\Phi^{-1}(\rho\sigma)$  來估測形態參數  $\nu$ 。

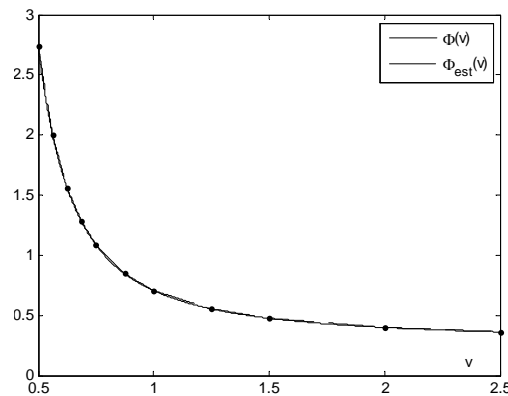


Fig. E-2  $\Phi(v)$  at  $v \in [0.5, 2.5]$  and its piecewise linear approximation.

### E.3.2.2 Piecewise linear estimation for the shape parameter

在Fig. E-2中我們可以發現， $\Phi(v)$ 是呈現一個指數型態變小且平滑的曲線，因此可以用分段式線性(piecewise linear)趨近的方式去模擬。從實驗可得知， $\Phi(v)$ 在  $v \in [0.5, 2.5]$ 之間可以利用分段式線性趨近的方式得到準確的結果。在實際作法上，我們將  $\Phi(v)$ 在  $v \in [0.5, 2.5]$  區間的曲線切割成十個區塊並用十條直線去進行線性趨近。對於每一個落於  $v \in [f_i, f_{i-1}]$ 的曲線來說， $\Phi_{est}(v)$ 為趨近用的直線片段，可表示為：

$$\Phi_{est}(v) = \frac{\Phi(f_i) - \Phi(f_{i-1})}{f_i - f_{i-1}}(v - f_i) + \Phi(f_i), \quad (7)$$

其中  $i = \{1, 2, \dots, 10\}$ ，且

$$\begin{aligned} & \{f_0, f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8, f_9, f_{10}\} \\ & = \{0.5, 0.5625, 0.625, 0.6875, 0.75, 0.875, 1, 1.25, 1.5, 2, 2.5\} \end{aligned} \quad (8)$$

Fig. E-2為利用  $\Phi_{est}(v)$ 趨近  $\Phi(v)$ 的結果，利用  $\Phi_{est}(v)$ 可再進一步求得型態參數的估測值  $v_{est} = \Phi_{est}^{-1}(\rho\sigma)$ ，其中

$$\Phi_{est}^{-1}(\rho\sigma) = \frac{f_i - f_{i-1}}{\Phi(f_i) - \Phi(f_{i-1})}(\rho\sigma - \Phi(f_i)) + f_i \quad (9)$$

且

$$\rho\sigma \in S_i = [\Phi(f_{i-1}), \Phi(f_i)] \quad (10)$$

既然利用了線性估測的方式，可以再更進一步建立數值表格將線性估測中(9)與(10)式中需要的參數值預先計算，以供在實際估測時查閱以降低運算複雜度；表格如Table E-1所示。

Table E-1 Look-up table for shape parameter estimation

$i$	$S_i$	$\frac{\Phi(f_i) - \Phi(f_{i-1})}{f_i - f_{i-1}}$	$\Phi(f_i)$	$f_i$
1	[2.739, 2.000]	-11.810	2.000	0.5625
2	[2.000, 1.563]	-7.005	1.563	0.6250
3	[1.563, 1.281]	-4.506	1.281	0.6875
4	[1.281, 1.089]	-3.080	1.089	0.7500
5	[1.089, 0.848]	-1.926	0.848	0.8750
6	[0.848, 0.707]	-1.126	0.707	1.0000
7	[0.707, 0.555]	-0.610	0.555	1.2500
8	[0.555, 0.476]	-0.314	0.476	1.5000
9	[0.476, 0.399]	-0.154	0.399	2.0000
10	[0.399, 0.363]	-0.073	0.363	2.5000

整體來說，當要建立一個  $\rho$ -GGD 訊源模型時，可利用下列幾個依序步驟完成：

步驟 1: 從小波轉換參數中計算  $\rho$  與  $\sigma$  數值。

步驟 2: 利用  $\rho$  與  $\sigma$  的乘積值對 Table E-1 查表以得到在式子(10)中對應的  $S_i$  與對應到的線性估測用參數。

步驟 3: 利用在步驟 2 所得到的估測參數代入式子(9)以得到  $v_{est}$ 。

步驟 4: 利用式子(5) 以建立起  $\rho$ -GGD 訊源模組。

### E.3.2.3 Modeling accuracy evaluation

為了估計兩個不同的機率分布之間的差異，我們引用了 Kullback-Leibler (K-L) 差異度 (diversity) 的方法，亦稱為差別熵(differential entropy)，其數學式如下：

$$KL(p \parallel q) = \sum_x p(x) \log_2 \left( \frac{p(x)}{q(x)} \right), \quad (11)$$

其中  $p$  為真實的資料機率分佈，而  $q$  則為所建構的機率模型，在此我們利用對稱性的 K-L 差異度[43]進行測量(如(12)所示)。對於此測量而言，K-L 差異度越小代表的是所建構的機率模型越準確。

$$KL(p \parallel q) + KL(q \parallel p) \quad (12)$$

## E.4 結果與討論

利用所提出的  $\rho$ -GGD 訊源模型，我們可以針對影像/視訊的小波轉換參數進行機率分佈的模擬。在我們的實驗設計中，我們將  $\rho$ -GGD 與 Laplacian 兩種訊源模型在不同測試中進行比較。在影像 2-D DWT 轉換時，Daubechies 9/7 互正交(biorthogonal)小波運算子[44]是最常被使用的，也因此我們的實驗中利用此運算子進行小波轉換。Fig. E-3顯示了在"Pepper"這張測試影像的各空間頻帶的小波轉換參數的機率分佈與其模擬。非常清楚的可以發現， $\rho$ -GGD 模型在各頻帶的模擬表現都較 Laplacian 為好。Table E-2則是另一測試影像"Lena"中，利用 K-L 差異度來進行  $\rho$ -GGD 與 Laplacian 兩種模型的準確度比較。整體來說， $\rho$ -GGD 模型在各頻帶中的表現除了在 LH 頻帶外，都較 Laplacian 有非常顯著的改進。

Table E-2. K-L divergences of two source models for the 2-D DWT coefficients for image "Lena".

Band Index Model	HL	LH	HH	LL-HL	LL-LH	LL-HH
Laplacian	0.22	0.07	0.03	0.68	0.52	0.42
$\rho$ -GGD	0.16	0.08	0.03	0.22	0.21	0.20



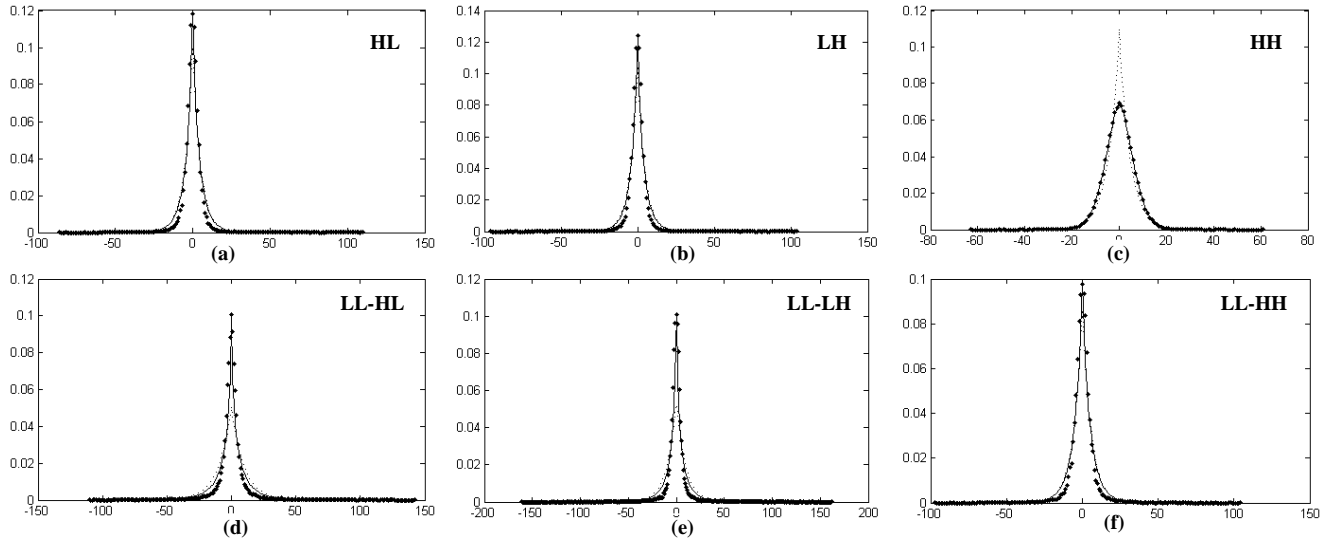


Fig. E-3 The pdfs of wavelet coefficients (dots) and their approximations by Laplacian (dotted line) and the proposed  $\rho$ -GGD (solid line) models in the subbands: (a) HL, (b) LH, (c) HH, (d) LL-HL, (e) LL-LH, (f) LL-HH. The test image is “Pepper”.

而在畫面間小波轉換視訊的情況[38]，時間-空間頻帶的小波轉換參數則是由 MCTF 與 2-D DWT 的過程所共同產生。Table E-3則顯示了“Bus”與“Mobile”兩個 CIF 解析度之測試視訊片段在 GOP=8 與 16 的 K-L 差異度結果。 $\rho$ -GGD 模型顯示出了較 Laplacian 準確的實驗結果。一般來說，拆解次數越多的頻帶訊號越難以分析，但在Table E-3 (b)中我們可以發現即使是在較為深入拆解後的時間頻帶， $\rho$ -GGD 依然展現了準確的模擬準確度。從實驗結果我們可以發現， $\rho$ -GGD 模型不管是在 2-D DWT 或是畫面間小波轉換視訊的兩種情況下，皆有良好的表現。比較起 Laplacian 模型， $\rho$ -GGD 擁有較高的準確度並且在各頻帶中展現了一致的好表現。

Table E-3 K-L divergence comparison of two source models for temporal-spatial subband coefficients. (a) “Bus” with GOP=8. (b) “Mobile” with GOP=16.

Band Index \ Model	Temporal Level 2 (H-frame)				Temporal Level 3 (H-frame)			
	LL	HL	LH	HH	LL	HL	LH	HH
Laplacian	0.89	0.38	0.30	0.08	0.70	0.33	0.27	0.10
$\rho$ -GGD	0.28	0.19	0.09	0.07	0.22	0.14	0.07	0.07

(a)

Band Index \ Model	Temporal Level 4 (H-frame)				Temporal Level 5 (L-frame)		
	LL	HL	LH	HH	HL	LH	HH
Laplacian	0.85	0.33	0.30	0.19	0.47	0.46	0.38
$\rho$ -GGD	0.20	0.05	0.03	0.02	0.06	0.10	0.07

(b)

在本計畫的此項研究中，我們提出了準確且低複雜度的  $\rho$ -GGD 訊源模型，可準確針對影

像/視訊小波轉換參數進行機率分佈的模擬，較傳統 Laplcian 作法準確且可在影像與視訊兩種情況都可得到良好的實驗結果。

## F. 參考文獻

- [1] GPGPU, <http://www.gpgpu.org/>
- [2] C.-Y. Lee, Y.-C. Lin, C.-L. Wu, C.-H. Chang, Y.-M. Tsao, and S.-Y. Chien, "Multi-Pass and Frame Parallel Algorithm of Motion Estimation in H.264/AVC for Generic GPU," in *Proc. IEEE International Conference on Multimedia and Expo, July 2007*, pp. 1603-1606
- [3] C.-W. Ho, O. C. Au, S.-H. Gary Chan, S.-K. Yip, and H.-M. Wong, "Motion Estimation for H.264/AVC Using Programmable Graphics Hardware," in *Proc. IEEE International Conference on Multimedia and Expo, July 2006*, pp. 2049-2052
- [4] Y.-C. Lin, P.-L. Li, C.-H. Chang, C.-L. Wu, Y.-M. Tsao, and S.-Y. Chien, "Multi-Pass algorithm of Motion Estimation in Video Encoding for Generic GPU," in *Proc. IEEE International Symposium on Circuit and Systems, May 2006*, pp. 4451-4454
- [5] J. Rehman, and Y. Zhang, "Fast Intra Prediction Mode Decision Using Parallel Processing," in *Proc. IEEE International Conference on Machine Learning and Cybernetics, August 2005*, pp. 5094-5098
- [6] S. Yang, W. Wolf, and N. Vijaykrishnan, "Power and Performance Analysis of Motion Estimation Based on Hardware and Software Realizations," in *IEEE Trans. Computer*, vol.54, pp.714-726, Jun, 2005.
- [7] T. Wiegand, G. Sullivan, J. Reichel, H. Schwarz, and M. Wien, "Joint Draft 10 of SVC Amendment," *ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT-W201*, 2007.
- [8] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103-1120, 2007.
- [9] T. Wiegand, G. Sullivan, and A. Luthra, "Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC)," *ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT-G050r1*, 2003.
- [10] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560-576, 2003.
- [11] J.-R. Ohm, "Advances in Scalable Video Coding," *Proc. IEEE*, vol. 93, no. 1, pp. 42-56, 2005.
- [12] J. Reichel, H. Schwarz, and M. Wien, "Joint Scalable Video Model JSVM-9," *ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT-V202*, 2007.
- [13] C.-A. Seagall and G.-J. Sullivan, "Spatial Scalability within the H.264/AVC Scalable Video Coding Extension," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1121-1135, 2007.

- [14] H. Schwarz, D. Marpe, and T. Wiegand, "Inter-layer Prediction of Motion and Residual Data," *ISO/IEC JTC 1/SC 29/WG11/M11043*, 2004.
- [15] H. Schwarz, D. Marpe, and T. Wiegand, "Hierarchical B Pictures," *ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT-P014*, 2005.
- [16] Y.-D. Zhang, F. Dai, and S.-X. Lin, "Fast 4x4 Intra-prediction Mode Selection for H.264," *IEEE International Conference*, vol. 2, 27-30 June 2004, pp. 1151-1154.
- [17] F. Pan, X. Lin, R. Susanto, K. P. Lim, Z. G. Li, G. N. Feng, D. J. Wu, and S. Wu, "Fast Mode Decision Algorithm for Intra Prediction in JVT," *ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT-G013*, 2003.
- [18] C.-L. Yang, L.-M. Po, and W.-H. Lam, "A Fast H.264 Intra Prediction Algorithm Using MacroblocK Properties," *International Conference*, vol. 1, 24-27 Oct. 2004, pp. 461-464.
- [19] Z. Zhou and M.-T. Sun, "Fast MacroblocK Inter Mode Decision and Motion Estimation for H.264/MPEG-4 AVC," *International Conference*, vol. 2, 24-27 Oct. 2004, pp. 789-792.
- [20] P. Yin, H.-Y.C. Tourapis, A.M. Tourapis, and J. Boyce, "Fast Mode Decision and Motion Estimation for JVT/H.264," *Proceedings of International Conference on Image Processing*, 2003, pp. 853-856.
- [21] K.-H. Han and Y.-L. Lee, "Fast MacroblocK Mode Decision in H.264," *IEEE Conference*, vol. A, 21-24 Nov. 2004, pp. 347-350.
- [22] C. Grecos and M.-Y. Yang, "Fast Inter Mode Prediction for P Slices in the H.264 Video Coding Standard," *IEEE Transactions*, vol. 51, Issue 2, June 2005, pp. 256-263.
- [23] J. Lee and B. Jeon, "Fast Mode Decision for H.264," *ICME 2004*, vol.2, June 2004, pp. 1131-1134.
- [24] F. Zhang and X. Zhang, "Fast MacroblocK Mode Decision in H.264", *ICSP 2004 7th International Conference*, vol. 2, 31 Aug.-4 Sept. 2004, pp. 1179-1182.
- [25] L. Xiong, "Reducing Enhancement Layer Directional Intra Prediction Modes," *ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT-P041*, 2005.
- [26] L. Yang, Y. Chen, J. Zhai, and F. Zhang, "Low Complexity Intra Prediction for Enhancement Layer," *ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT-Q084*, 2005.
- [27] H. Li, Z.-G. Li, and C. Wen, "Fast Mode Decision for Coarse Grain SNR Scalable Video Coding," *IEEE Int'l Conf. on Acoustics, Speech and Signal Processing*, 2006.
- [28] H. Li, Z.-G. Li, and C. Wen, "Fast Mode Decision Algorithm for Inter-Frame Coding in Fully Scalable Video Coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 7, pp. 889-895, 2006.
- [29] H.-C. Lin, W.-H. Peng, and H.-M. Hang, "A Fast Mode Decision Algorithm with MacroblocK-Adaptive Rate-Distortion Estimation for Intra-Only Scalable Video Coding," *IEEE Int'l Conf. on Multimedia and Expo*, 2008

- [30] H.-C. Lin, W.-H. Peng, H.-M. Hang, and W.-J. Ho, "Layer-adaptive Mode Decision and Motion Search for Scalable Video Coding with Combined Coarse Granular Scalability (CGS) and Temporal Scalability," *IEEE Int'l Conf. on Image Processing*, 2007.
- [31] S. G. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, no. 2, pp. 674-693, July 1989.
- [32] J. M. Shapiro, "Embedded Image Coding Using Zerotrees of Wavelet Coefficients," *IEEE Trans Signal Process.*, vol. 41, pp. 3445-3462, Dec. 1993.
- [33] A. Said and W. A. Pearlman, "A New, Fast and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 243-250, June 1996.
- [34] D. Taubman, "High Performance Scalable Image Compression with EBCOT," *IEEE Trans Image Processing.*, vol. 9, pp.1158-1170, July 2000.
- [35] J.-R. Ohm, "Three-dimensional Subband Coding with Motion Compensation," *IEEE Trans Image Processing.*, vol. 3, no. 5, pp. 559-571, Sep. 1994.
- [36] S.-T. Hsiang and J. W. Woods, "Embedded Video Coding Using Invertible Motion Compensated 3-D Subband Wavelet Filter Bank," *Signal Processing: Image Communications*, vol. 16, pp.705-724, May 2001.
- [37] A. Secker and D. Taubman, "Lifting-Based Invertible Motion Adaptive Transform (LIMAT) Framework for Highly Scalable Video Compression," *IEEE Trans Image Processing.*, vol. 12, no. 12, Dec. 2003.
- [38] J. Xu and *et al.*, "3D Subband Video Coding Using Barbell Lifting," *ISO/IEC JTC1/SC29/WG11 MPEG, M10569*, 2004.
- [39] S. Mallat and F. Falzon, "Analysis of Low Bit Rate Image Transform Coding," *IEEE Trans Signal Process.*, vol. 46, no. 4, pp. 1027-1042, Apr. 1998.
- [40] P.-Y. Chen, J. Li, and C.-C. J. Kuo, "Rate Control for An Embedded Wavelet Video Coder," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, pp. 696-702, Aug. 1997.
- [41] S.R. Smoot and L.A. Rowe, "Laplacian Model for AC DCT terms in Image and Video Coding," *Proc. 9<sup>th</sup> IEEE Image and Multidimension Digital Signal Processing Workshop*, 1996.
- [42] Z. He and S. K. Mitra, "A Unified Rate-Distortion Analysis Framework for Transform Coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 1221-1236, Dec. 2001.
- [43] S. Kullback and R. A. Leibler, "On Information and Sufficiency," *Annals. of Mathematical Statistics*, vol. 22, pp. 79-86, 1951
- [44] R. W. Buccigrossi and E. P. Simoncelli, "Image Compression via Joint Statistical Characterization in the Wavelet Domain," *IEEE Trans Image Processing.*, vol. 8, no. 12, pp. 1688-1701, Dec. 1999.
- [45] M. Antonini and *et al.*, "Image Coding Using Wavelet Transform," *IEEE Trans Image Processing.*, vol. 11, no. 2, pp. 205-221, April 1992.

## G. 計畫成果自評

多媒體服務是寬頻無線網路的最重要應用之一。本計畫重點在先進音、視訊標準編碼的演算法加速與實作，也開發設計適合在無線網路上傳送串流多媒體的視訊編碼法設計。本專題研究將承繼我們過去的經驗與前人的成果，進一步設計發展解決問題方式。所發展出的技術、經驗及成品極具實用價值，可促進國內工業研發技術開發。

參與工作人員(研究生)在學理上習得音訊與視訊編碼技術與國際標準。針對寬頻無線網路，設計開發可調式編碼等演算法，成員得到此課題研究與開發產品的經驗與知識。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，達到人才培育之目的。

綜合評估：研究內容與原計畫進度與內容大致相符，已達成學術研究創新與人才培育之預目標。整體成效良好。研究成果頗具學術與應用價值，承繼之前的同一系列研究項目，已發表五篇學術會議論文，及碩士學位論文二冊如下表。

## H. Publications: (含前兩年同一系列計畫產出，於過去一年中發表者)

- [1] Wei-Nien Chen, 陳威年, "H.264/AVC Motion Estimation Implementation on Compute Unified Device Architecture (CUDA)", MS Thesis, NCTU, June 2008.
- [2] Shang-Yu Yeh 葉尚諭, "Coding Efficiency and Quality Improvement for MPEG Surround Encoding," MS Thesis, NCTU, June 2008.
- [3] C.-C. Liu and H.-M. Hang, "Acceleration and Implementation of JPEG2000 Encoder on TI DSP platform," *IEEE International Conf. on Image Processing '07*, San Antonio, TX, USA, Sept. 2007
- [4] H.-C. Lin, W.-H. Peng, H.-M. Hang, and W.-J. Ho, "Layer-Adaptive Mode Decision and Motion Search for Scalable Video Coding with Combined Coarse Granular Scalability (CGS) and Temporal Scalability", *IEEE Int'l Conf. on Image Processing '07*, San Antonio, TX, USA, Sept. 2007
- [5] H.-C Lin, W.-H. Peng and H.-M. Hang, "A Fast Mode Decision Algorithm with Macroblock-Adaptive Rate-Distortion Estimation for Intra-Only Scalable Video Coding," *IEEE Int'l Conf. on Multimedia and Expo.*, Hannover, Germany, June 2008.
- [6] W-N Chen and H.-M Hang, "H.264/AVC Motion Estimation Implementation on Compute Unified Device Architecture(CUDA)", *IEEE Int'l Conf. on Multimedia and Expo.*, Hannover, Germany, June 2008.
- [7] C.-Y. Tsai and H.-M Hang, "Rho-GGD Source Modeling for Wavelet Coefficients in Image/Video Coding", *IEEE Int'l Conf. on Multimedia and Expo.*, Hannover, Germany, June 2008.