

# 行政院國家科學委員會專題研究計畫 成果報告

## 子計畫五：MPEG 智財管理與保護系統及強韌視訊解碼器之設計與模擬(II)

計畫類別：整合型計畫

計畫編號：NSC93-2219-E-009-021-

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

計畫參與人員：張峰誠 蔡家揚 洪朝雄 徐漢光 呂家賢 陳旻弘

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 標準之多媒體通訊整合平台及其應用(2) —子計畫五：  
MPEG 智財管理與保護系統及強韌視訊解碼器之設計與模擬(2/2)  
**MPEG IPMP System and Robust Video Decoder  
Design and Simulation (2/2)**

計畫類別： 個別型計畫  整合型計畫

計畫編號： NSC 93-2219-E-009-021

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

計畫參與人員：張峰誠 蔡家揚 洪朝雄 徐漢光 呂家賢 陳旻弘

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  
MPEG 智財管理與保護系統及強韌視訊解碼器之設計與模擬(2/2)  
MPEG IPMP System and Robust Video Decoder  
Design and Simulation (2/2)

計畫編號：NSC 93-2219-E-009 -021

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

計畫參與人員：張峰誠 蔡家揚 洪朝雄 徐漢光 呂家賢 陳旻弘

國立交通大學電子研究所研究生

## 中文摘要

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，與(2)用於影像壓縮之方向性多重解析度轉換和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。在位元率控制法部分，利用人類視覺系統屏蔽效應(masking effect)等特性，我們提出了一個算出視覺權重的方法，可以在同樣的傳輸速率下達成較佳的視覺品質，雖然其訊號雜訊比(PSNR)較低。在小波轉換編碼的空間轉換部分，Minh N. Do 提出了輪廓轉換(contourlet)建立新的圖片表示方法。我們用 Minh N. Do 的輪廓轉換放入 MPEG Wavelet Coding 中，並取代掉原本壓縮器常使用的小波轉換。在低位元比率時(0.625% 和 0.9375%)比原本的方法要來的好，尤其當圖片包含較多的線條成分時，視覺品質的改善更加明顯。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，我們藉著位元平面上能量聚集的現象讓提升編碼效率，因而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。在低位元比率時，在第一、第二有較明顯節省位元減少比例。

**關鍵字：**畫面間小波視訊編碼、人類視覺系統、位元率控制、方向性多重解析度轉換、算數編碼。

## 英文摘要

Interframe wavelet video coding algorithm has received much attention lately because of its high coding efficiency and flexible temporal and spatial scalability. It can achieve three types of scal-

ability: SNR, temporal and spatial, in one single bit-stream. Our interframe wavelet scheme has been submitted to the MPEG committee in response to the scalable video coding (SVC) Call-for-Proposal. This year, we continue participating in the discussion of MPEG Ad Hoc Group and refinement of the MPEG interframe wavelet reference model. In 2005, we also propose algorithms to join the MPEG SVC Core Experiments. In this project, we have two research topics as follows: (1) HVS (human visual system)-based rate control algorithm for interframe wavelet video coding, and (2) directional multiresolution transform and block-based bitplane arithmetic coding for image compression. The project goal is to improve the coding efficiency of using spatial transform, entropy coding, and rate control techniques in the interframe wavelet video coding framework.

In the HVS-based rate control algorithm part, we use the human perceptual properties such as masking effect to propose a method that transforms the ordinary “mathematical distortion” to the “visual distortion”. Our simulations indicate that the reconstructed frames may have lower PSNR but higher visual quality. In the wavelet transform part, Dr. Minh Do suggested a contourlet transform, which has a better visual quality in image representation. We replace the ordinary separable wavelet transform by the contourlet representation in the interframe wavelet system and, at lower bit rates, this new representation demonstrates better subjective image quality. In the entropy coding part, we observed the clustering phenomenon in the bit planes. Thus, we proposed an SB-reach plane method to improve the coding efficiency. For particularly bit planes one and two, this technique offer more significant bit savings o in our simulations.

**Keywords:** interframe video coding, human visual system, rate control, directional multiresolution transform, and arithmetic coding.

## 目錄

中文摘要.....	II
英文摘要.....	II
目錄.....	IV
報告內容.....	1
第一部份 以人類視覺系統為基礎的位元控制法.....	1
A. 前言.....	1
B. 研究目的.....	1
C. 文獻探討.....	1
D. 研究方法.....	5
E. 結果與討論.....	8
第二部份 用於影像壓縮之方向性多重解析度轉換和區塊位元平面算數編碼.....	11
A. 前言.....	11
B. 研究目的.....	11
C. 文獻探討.....	11
D. 研究方法.....	13
E. 結果與討論.....	16
參考文獻.....	19
計畫成果自評.....	20
Publications.....	20
MPEG Standard Contributions .....	20

## 報告內容

### 第一部份 以人類視覺系統為基礎的位元控制法

#### A. 前言

在多媒體傳輸的應用中，不同的接收者會有不同的承受量，故可調整性(scalability)在今日的多媒體傳輸中是一個重要的特性。用於畫面之間的小波轉換編碼(Interframe Wavelet Video Coding)是一個新的視訊編碼方式且能提供良好的可調整性。因此這個編碼方式在近年來受到不少矚目，而且已經有很多的研究和改良來增進它的效能。

在很多環境下，人眼都是視訊品質的最後判斷所在。然而，在設計視訊編碼時要考慮人類視覺的效應卻很困難。我們必須要能把客觀的“數學上的不同”轉換成主觀的“視覺上的不同”，也就是說，我們必須要把普通的“量化錯誤”轉換成“人類視覺上的加重錯誤”。因此，本計畫此部份就人類視覺效應之特性，進行可調整性編碼傳輸時之位元控制之研究。

#### B. 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。

#### C. 文獻探討

##### C.1 位元控制法 (Rate Control)

JPEG2000 使用了“嵌入式位元流作最佳的截斷的嵌入式區塊編碼(Embedded Block Coding with Optimized Truncation of the embedded bit-stream, EBCOT) “[1]。此編碼方法能把每個次頻帶(subband)分成許多小編碼區塊(code-block)。假設每個區塊內有  $m$  個截斷點(truncation point)，對某個截斷點  $n$  而言，要編碼某個區塊  $i$  所到此截斷點所需要的位元率為  $R_i^n$ ，且到此截斷點所造成的失真為  $D_i^n$ 。則全部的區塊所造成的全部的失真

$D = \sum_i D_i^n$  且全部的位元率  $R = \sum_i R_i^n$ 。我們希望在每個區塊裡能夠找到一個截斷點，

而且此組截斷點  $\{n_i^\lambda\}$  能夠達成(1)全部的位元率要能小於所指定的位元率且能達成整體的失真最小。或(2)全部的失真要能小於所指定的失真且能達成全體的位元率最小。也就是說，要找到一組  $\{n_i^\lambda\}$  令  $(D(\lambda) + \lambda R(\lambda)) = \sum_i (D_i^{n_i^\lambda} + \lambda R_i^{n_i^\lambda})$  最小。不過通常截斷點大都已经被事先指定了，故我們要找的是  $\lambda$ 。

我們要先知道在每個區塊  $i$  的截斷點  $j$  的斜率(slope)  $\frac{\Delta D_i^j}{\Delta R_i^j} = \frac{D_i^{j-1} - D_i^j}{R_i^j - R_i^{j-1}}$ ，其中  $R_i^j$  是

從區塊  $i$  的開始到截斷點  $j$  所累積位元長度(bit length)而  $D_i^j$  是區塊  $i$  的開始到截斷點  $j$  所累積失真。只要某個截斷點的斜率大於或等於  $\lambda$ ，我們就會把這個截斷點的資料送去封裝(package)，最後我們會檢討全部封裝後的位元率是否小於或等於所要求的位元率。如果封裝後的位元率太大，我們就要增加  $\lambda$ ；不然就要減少  $\lambda$ ，直到找到最適合的  $\lambda$  為止[1]。

北京微軟研究中心(Microsoft Research Asia)提出了一套稱為擁有良好的截斷的三維嵌入式次頻帶編碼(Three-Dimensional Embedded Subband Coding with Optimized Truncation, 3-D ESCOT)，它是把EBCOT應用到三維的資料(2個空間軸和1個時間軸)[7]，這編碼方法中，每個位元平面(bit-plane)有三個截斷點。

## C.2 對比敏感度 (Contrast Sensitivity)

人類的視覺對亮度的相對性比對亮度的覺對性還敏感，但因為自然的影像的複雜性，適合所有的狀況的共通的敏感度的定義並不存在。目前最通用於複雜的影像的敏感度

定義為區域頻帶限制敏感度(local band limited contrast)  $C_i(x, y) = \frac{BP_i(x, y)}{LP_i(x, y)}$  [2]，其中

$C_i(x, y)$  為位於影像中的位置  $(x, y)$  的對比敏感度， $BP_i(x, y)$  為把位在  $(x, y)$  的影像通過一個頻帶  $i$  的值，而  $LP_i(x, y)$  為所有的影像通過頻帶  $i$  之下的頻帶的值。

## C.3 屏蔽效應 (Masking Effect)

如果一個訊號本身可以被人眼所看到，會因為其他的訊號的存在而無法看到，也就是此訊號的視覺界限(visual threshold)增加了，如圖 1 所示，這稱為屏蔽效應(masking effect)[3]。

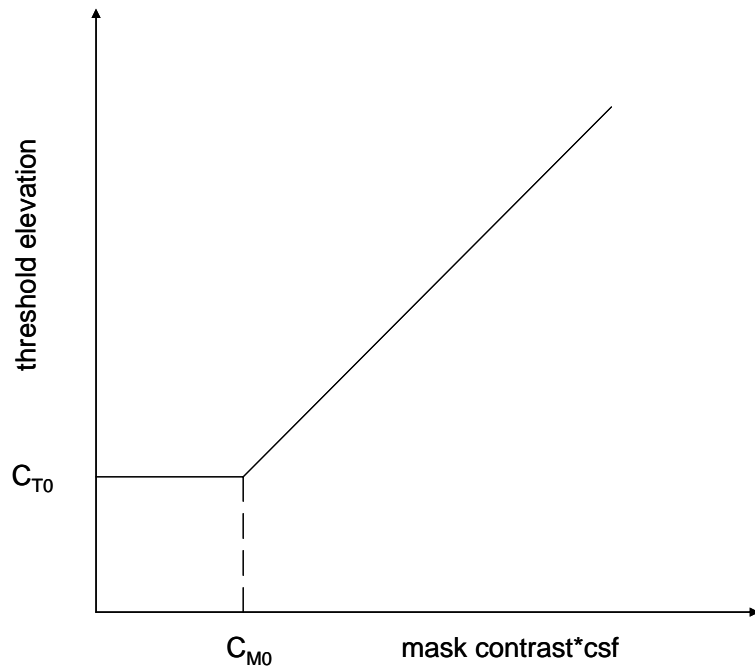


圖 1：屏蔽效應

#### C.4 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) $\lambda$  以及方位(orientation)  $\theta$  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  $a = 0.495$ ， $k = 0.466$ ， $f_0 = 0.401$ 。  $g_\theta$  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[4]。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  $\lambda$  的小波轉換所對應的空間上的頻率  $f$  為  $f = r * 2^{-\lambda}$ 。圖 2 是在  $r = 16$  的狀況下所算出來的各次頻帶的亮度成份的正好可被注意到的失真。



(3, LL) 2.0 0.663	(3, LH) 2.0 0.835	(2, LH) 4.0 1.444	(1, LH) 8.0 3.034
(3, HL) 2.0 0.835	(3, HH) 2.0 1.359		
(2, HL) 4.0 1.444		(2, HH) 4.0 2.804	(1, HH) 8.0 7.027
(1, HL) 8.0 3.034			

(level, orientation)  
spatial frequency  
minimum threshold

圖 2：各次頻帶的亮度成份的正好可被注意到的失真

為了要把之前所算出來的結果用在人類視覺系統(human visual system, HVS)上，我們要把所量測到的正好可被注意到的失真從空間範圍轉換到小波範圍，也就是說，我們要量測出最小的正好可以產生出正好可被注意到的失真的小波係數。我們有“最差的狀況”的公式來計算，也就是  $t_{JND}(\lambda, \theta) = \frac{y(\lambda, \theta)}{i_\theta \cdot p_l^{2(\lambda-1)}}$ 。其中  $i_\theta$  是  $p_l^2$ ， $p_h^2$ ，或  $p_l p_h$  分別對應到 LL，HH，或 LH/HL 等次頻帶。 $p_l$  是低通合成濾波器(low pass synthesis filter)的最大係數，而  $p_h$  是高通合成濾波器(high pass synthesis filter)的最大係數[4][5]。圖 3 是把圖 2 用 Daubechies 9/7 合成濾波器所得到的結果，其中  $p_l = 1.115087052456994$  而  $p_h = 0.6029490182363579$ 。

(3, LL) 2.0 0.345	(3, LH) 2.0 0.803	(2, LH) 4.0 1.727	(1, LH) 8.0 4.513
(3, HL) 2.0 0.803	(3, HH) 2.0 2.419		
(2, HL) 4.0 1.727		(2, HH) 4.0 6.204	(1, HH) 8.0 19.329
(1, HL) 8.0 4.513			

(level, orientation)  
spatial frequency  
 $t_{JND}$

圖 3：各頻帶的亮度成份在小波範圍的正好可被注意到的失真

### C.5 視覺失真量度(Perceptual Distortion Metric)

假設一張影像經過小波轉換後， $e(\lambda, \theta, x, y)$ 位於次頻帶 $(\lambda, \theta)$ 裡的位置 $(x, y)$ 的量化錯誤(quantization error)，而  $t_{JND}(\lambda, \theta, x, y)$ 是位於此點的正好可被注意的失真，則在此點可以

$$\text{看到失真的機率為 } p_{(\lambda, \theta, x, y)} = 1 - \exp\left(-\left|\frac{e(\lambda, \theta, x, y)}{t_{JND}(\lambda, \theta, x, y)}\right|^{\beta_b}\right) [6]。$$

## D. 研究方法

### D.1 轉換斜率的表示法

一般來說，每個截斷點的斜率  $\frac{\Delta D_i^j}{\Delta R_i^j}$  的值都很大而且他們之前的差距也都很大，我們改

用  $\left\lceil 0.5 + 2^9 * \log_2\left(\frac{\Delta D_i^j}{\Delta R_i^j}\right) \right\rceil$  來表示個截斷點的斜率，如此可把它們的值和之間的差距縮

小。重要的是，它們的排列順序不會改變，也就是說，它們在作位元控制時被封裝的

順序都還是一樣的。

## D.2 次頻帶內的權重要素(Intra-Subband Weighting Factor)

次頻帶內的權重要素是用來決定在同一個次頻帶內的所有的截斷點的視覺權重要素。圖 3 中的斜率經由實驗證明為 1 且  $C_{T0} = C_{M0} = 1$ [3]。如果把圖 3 的橫軸和縱軸都對除以相對應的某個頻率的對比度方程式(contrast sensitivity function)，此圖可以用在全部的頻率 [3]。而除以相對應的某個頻率的對比度方程式相當於乘上由  $\log(y) = \log(a) + k \cdot (\log(f) - \log(g_\theta f_0))^2$  所算出來的  $y$ ，如此我們可以得到各次頻帶的屏蔽效應的圖。

接下來要知道在某個次頻帶內的某個像素的對比(contrast)，我們可以用之前所提出的區域頻帶限制敏感度來算出對比，在這裡我們作了一些修正，如果我們假設  $C_i(x, y) = \frac{BP_i(x, y)}{LP_i(x, y)}$  中的  $LP_i(x, y)$  是對整張影像的每個像素而言都是一樣的且等於最低層的次頻帶的系數(相當於圖 2 的(3,LL)這個次頻帶)的平均值，我們可以找出在整張影像

上位於位置  $(i, j)$  的像素的對比，也就是對比  $c(i, j) = \left| \frac{C(i, j)}{E(C_{lowest-spatial-subband})} \right|$ ，其中

$E(C_{lowest-spatial-subband})$  是最低層次頻帶的系數平均， $C(i, j)$  此位置的小波系數[5]。如此我們可以找到全部的次頻帶上的全部的像素的對比，並可以藉由此次頻帶的屏蔽效應的圖找出此次頻帶上每個像素的正好可被注意到的失真。根據實驗的結果，我們選擇次頻帶中所有的像素的對比的中間值所對應到的正好可被注意到的失真來當做此次頻帶的正好可被注意到的失真  $T(\lambda, \theta, C_{middle})$ ，接下來我們要把此失真從空間範圍轉換到小波範圍，也就是  $t_{JND}(\lambda, \theta, C_{middle}) = \frac{T(\lambda, \theta, C_{middle})}{i_\theta \cdot p_i^{2(\lambda-1)}}$ 。

我們把前面的視覺失真的公式  $p_{(\lambda, \theta, x, y)} = 1 - \exp\left(-\left|\frac{e(\lambda, \theta, x, y)}{t_{JND}(\lambda, \theta, x, y)}\right|^4\right)$  在此作一些修

正，也就是把  $t_{JND}(\lambda, \theta, x, y)$  換成  $t_{JND}(\lambda, \theta, C_{middle})$ 。一般而言，人眼的視角可達 180 度到 160 度，但人眼真正會注意的部分大概只佔了 2 度的視角，假設有個區域在以  $(n1, n2)$  為中心且佔了 2 度左右的視角，在這個區域會看到失真的機率為  $P_{F(n1, n2)} = 1 - \prod_{(\lambda, \theta, x, y) \in F} (1 - p_{(\lambda, \theta, x, y)})$  [6][8]。實驗發現在一般看電腦螢幕的狀況下，這個區域

大概是一個 3D-ESCOT 的編碼區塊的大小，故我們可以得到看到失真的機率為

$$P_{F(n1,n2)} = 1 - \exp\left(-\left(\frac{\sum_{(\lambda,\theta,x,y) \in F} |e(\lambda,\theta,x,y)|^4}{(t_{JND}(\lambda,\theta,C_{middle}))^4}\right)\right)。$$

要算出每個編碼區塊的截斷點的次頻帶內的權重要素，現在是一個編碼區塊為處理單位，假設現在是在次頻帶 $(\lambda,\theta)$ 中的第 $z$ 個編碼區塊，經由實驗發現，在這個編碼區塊內的某個截斷點的次頻帶內的權重要素 $w_l$ 為

$$P_{F(n1,n2)} = 1 - \exp\left(-\left(\frac{\sum_{x=0}^{W(\lambda,\theta,z)-1} \sum_{y=0}^{H(\lambda,\theta,z)-1} |e(\lambda,\theta,z,x,y)|^4}{H(\lambda,\theta,z) * W(\lambda,\theta,z) * (t_{JND}(\lambda,\theta,C_{middle}))^4}\right)\right)。$$

其中 $H(\lambda,\theta,z)$ 是此編碼區塊 $W(\lambda,\theta,z)$ 的長度而是次編碼區塊的寬度。而 $e(\lambda,\theta,z,x,y)$ 是在次頻帶 $(\lambda,\theta)$ 內的第 $z$ 個編碼區塊的位置 $(x,y)$ 上對此截斷點的失真。

### D.3 次頻帶間的權重要素(Inter-Subband Weighting Factor)

同一個截斷點在不同的次頻帶內是有不同的視覺權重，這就是次頻帶間的權重要素，

$$\text{實驗發現次頻帶間的權重要素 } w_2 = 1 + \frac{\exp\left(\frac{t_{JND-lowest-spatial-subband}}{t_{JND}(\lambda,\theta,0)}\right)}{10}，\text{ 其中}$$

$t_{JND-lowest-spatial-subband}$  是最低層的次頻帶的正好可被注意到的失真，在圖 3 中，

$$t_{JND-lowest-spatial-subband} = t_{JND}(3,LL,0) = 0.345。$$

### D.4 位元控制

先把所有的截斷點的斜率都算出來，然後用 C.1 所講的式子作轉換，把轉換出來的斜率，乘上由 C.2 和 C.3 所算出來的權重要素，再去作位元控制。

## E. 結果與討論



(a)原本, 訊號雜訊比 31.19dB, 位元流大小 4371 位元組



(b)權重, 訊號雜訊比 30.98dB, 位元流大小 4317 位元組

圖 4：測試影像一在每秒 1000K 個位元的傳輸速率



(a)原本, 訊號雜訊比 26.85dB, 位元流大小 4687 位元組



(b)權重, 訊號雜訊比 26.67dB, 位元流大小 4272 位元組

圖 5：測試影像二在每秒 1000K 個位元的傳輸速率

圖 4 和圖 5 的傳輸速率是假設解碼端每秒可以解碼 30 張的狀況下。從圖 4 和圖 5 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。

因為我們把每個截斷點的斜率轉換成視覺失真的斜率，換句話說，我們把“數學上的量化的錯誤”轉換成“視覺上的權重的失真”，因此最後解碼出來的影像雖然有較佳的視覺效果，但不並會有較高的訊號雜訊比，因為訊號雜訊比是不考慮人眼視覺的特性，只是考慮到量化的錯誤，也就是單純的數學上的差。

在把每個截斷點的斜率都乘上權重後，我們可以發現，越接近最小重要性位元 (least significant bit, LSB) 的位元平面的截斷點的權重都很低，如此，視覺權重會把接近最大重要性位元 (most significant bit, MSB) 的位元平面的截斷點和越接近最小重要性位元的位元平面的截斷點之間的差距拉開，也就是說，如果我們要封裝乘上權重後的越接近最小重要性位元的位元平面的截斷點的資料，我們要更大的傳輸速率才行。因此，乘上權重後所封裝的資料就比較少。

在本計畫中，我們提出了一個算出視覺權重的方法，可以在同樣的傳輸速率下達成較佳的視覺品質，但並不會有較高的視覺雜訊比，這也表示說有較高的視覺雜訊比的影像並不表示說有較佳的視覺品質。

## 第二部份 用於影像壓縮之方向性多重解析度轉換和區塊位元平面算數編碼

### A. 前言

在圖片和影片壓縮的領域中，空間轉換扮演著重要的角色。藉著空間轉換，圖片就能夠被轉換到頻域。對於一般常看到的自然影像，能量多半會聚集在某些區域的頻段上，這些頻段一般是低頻頻段。由於這樣能量聚集的特性，所以使的壓縮的效率能夠提升。在理論上 FFT、KLT、DCT 和小波轉換是常見的空間轉換方法，然而實際的編碼器要考慮到效能和實現困難度的取捨，因此 JPEG、MPEG-1、MPEG-2 和 H.264 是採用 2D-DCT。在 JPEG2000 和畫面間小波轉換編碼則是使用小波轉換。經過空間轉換後之影像，必須搭配有效率的熵(Entropy)編碼器，才可發揮壓縮效能，本計畫部份主要針對小波轉換編碼的空間轉換與熵編碼器進行研究。

### B. 研究目的

小波轉換的優點在於能有效的表示點和點的變化，由於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，在幾年前 Candes and Donoho[9]發表了新的空間分析方法，此方法叫做曲線轉換 (curvelet)，藉著這個方法來趨近二維曲線的函數。藉著曲線轉換的激發，Minh N. Do [10] 提出了輪廓轉換(contourlet)建立新的圖片表示方法。

我們的研究會用 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。而在熵編碼方面，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。

### C. 文獻探討

#### C.1 輪廓轉換

由於本研究的主要部分是使用 Minh N. Do 提出的輪廓轉換，因此文獻討論會圍繞在輪廓轉換以及此轉換所會運用的相關知識。



輪廓轉換由 Minh N. Do 提出的，此轉換綜合了曲線轉換和次頻域分解的優點的優點。他對圖片的分解方法可以粗略的分成兩個步驟：(1) 大區域多重解析度轉換與 (2) 局部區塊方向性轉換，第一個步驟的目的事為了邊界偵測並且同時把類似小波轉換的方法應用到影像分析，第二個步驟的目的是利用局部區塊方向性轉換來處理物件輪廓的區塊。

前述的兩個步驟在實作上分別使用 Burt 和 Adelson [11]提出的拉氏金字塔分解以及 Bamberger 和 Smith 提出的二維方向性濾波頻帶，下面分別簡述這兩個參考文獻。

拉氏金字塔分解可以達到多重解析度的分解，此分解可以使圖片分成高低頻成分，低頻部分圖片由原始圖片產生，並且同時執行下採樣，至於經過反向預測的低頻圖片和原始圖片的差異部分就形成高頻成分圖片。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圖 6 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。

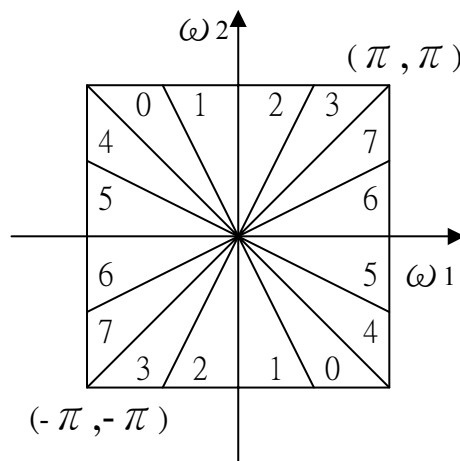


圖 6：二維方向性濾波頻帶對頻譜的分解示意圖

## C.2 內嵌式零區塊編碼

內嵌式零區塊編碼(Embedded zero block coding, EZBC)可用於連續小波轉換後的係數 [15]。內嵌式編碼方法藉著兩個強大的內嵌是編碼技巧--階層分割(set partitioning)和內容模型(context modeling)，達到低計算量但有效率壓縮。從小波係數建立的四元樹節點 (quad-tree nodes)有極大的統計相關性，利用此特性可以達到高效率的壓縮。

## D. 研究方法

### D.1 方向性多重解析度轉換

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，圖 7 是輪廓轉換的區塊示意圖。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。

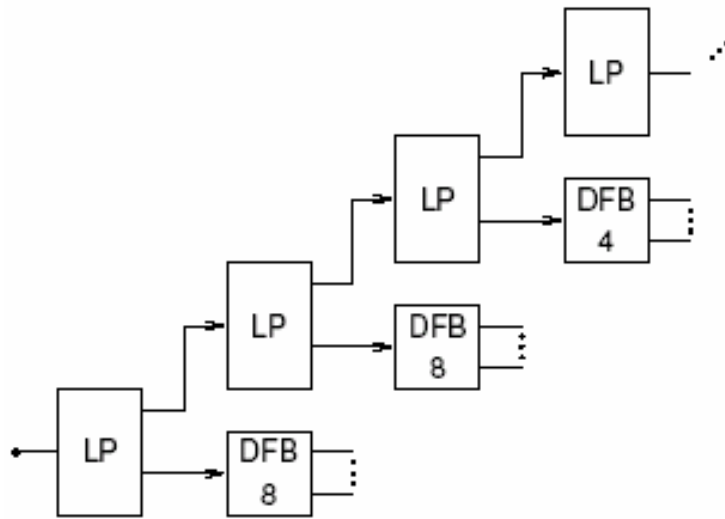


圖 7：輪廓轉換區塊示意圖

- 拉氏金字塔分解

圖 8 為拉氏金字塔分解的架構圖，

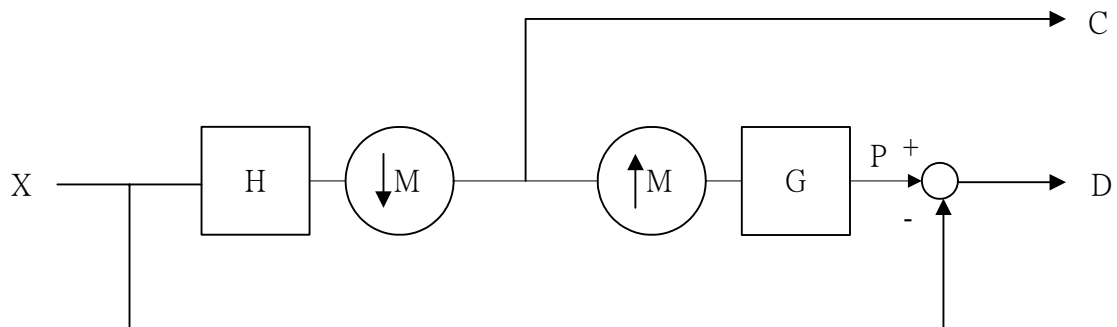


圖 8：拉氏金字塔分解的分析端，C 為拉氏金字塔分解的低頻訊號，D 為高頻成分，也就是輸入圖片 X 和 C 之間的差異處。

圖 9 為相對應的合成端，H 和 G 是使用 9/7 Daubechies 濾波器[12]

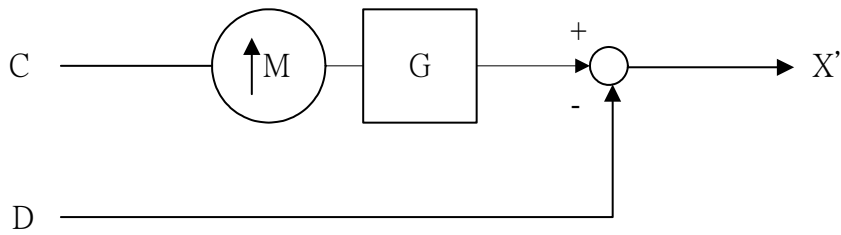


圖 9：拉氏金字塔分解的合成端，X' 為重建圖片

● 二維方向性濾波頻帶

二維方向性濾波頻帶主要是由扇形濾波器和 quincinx 次取樣所組成，在我們實際現實時，扇形濾波器利用 Phoong 所提出的方法[13]，再做頻域調變；quincinx 次取樣主要是參考[14]。圖 10 為四個方向的二維方向性濾波頻帶的架構圖。

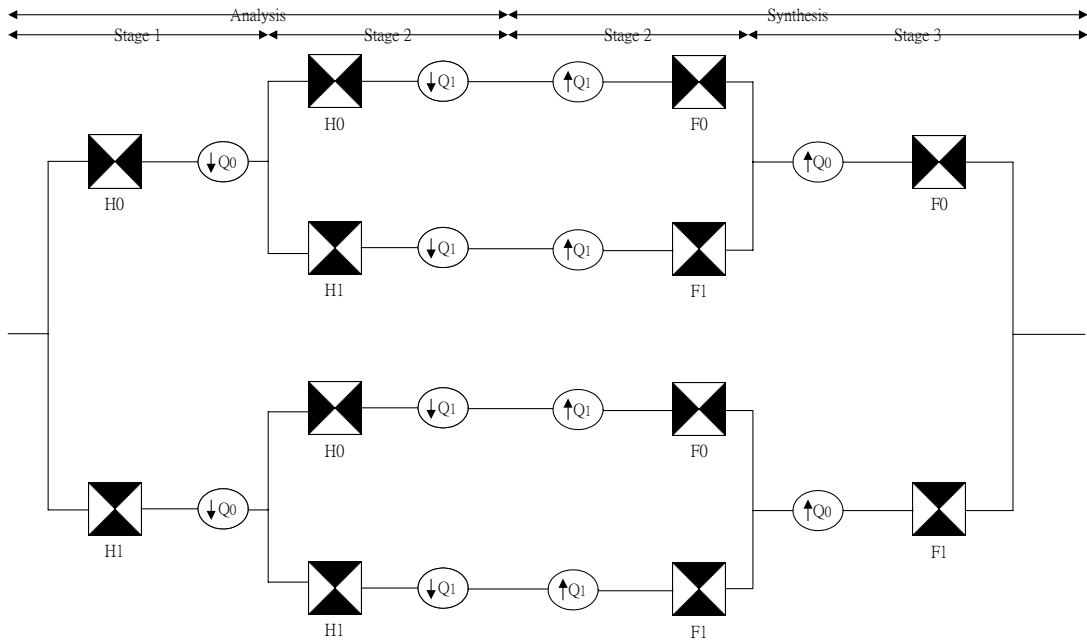


圖 10：四個方向的二維方向性濾波頻帶之架構圖，分別包括分析端和合成端

D.2 有效位元整併法

為了達成研究目的，我們提出並說明 significant 位元整併法。在 MPEG Wavelet Coding 參考軟體使用的三維嵌入式塊最佳截斷編碼(3D EBCOT)，其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)，圖 11 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。

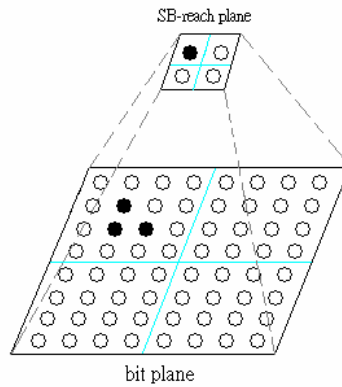


圖 11：整併平面參數對應 4X4 位元平面區塊

修正過後的編碼過程如下：如圖 12 所示，我們首先建立有效位元整併面(SB-reach plane)直到有效位元整併深度(SB-reach depth)之內的位元面。接著對整併面編碼，編碼方式包括 Significant Propagation pass 和 Normalization pass，這些步驟跟核心實驗軟體中的方法是一樣的，若是整併面在之前的位元面就已經 significant，其相對應的整併面參數必為 1，但此資訊因之前已經被傳遞出去，所以不需要再編碼。

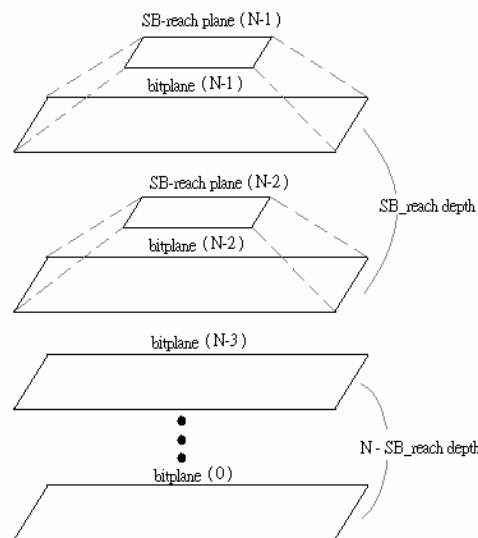


圖 12：『有效位元整併深度』之說明

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圖 13

說明。位元面的編碼步驟和核心實驗軟體中是一樣的，圖 14 是一個簡單說明我們編碼方式的例子，此例子中的位元會被送到算數編碼器作編碼。

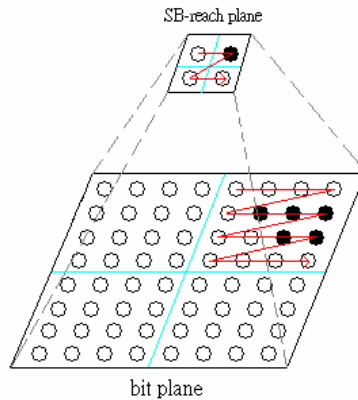


圖 13：有效位元整併面與對應位元面之編碼過程

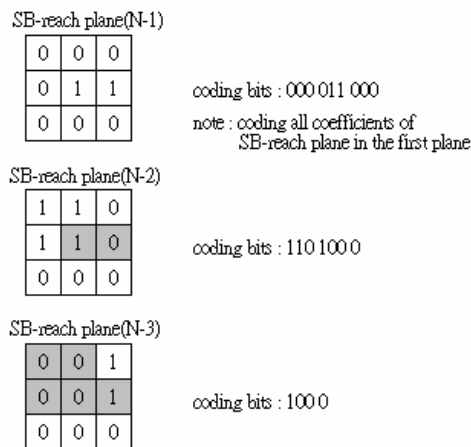


圖 14：我們提出編碼步驟一例

根據上面所提的方法，我們嘗試所有不同的對應區塊和整併深度，挑選出效率最好的配對，此配對關係將會被編碼以供解碼器使用。

## E. 結果與討論

在方向性多重解析度轉換研究方面，表 1 和表 2 分別是測試圖片為 barbara 和 fingerprint 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 1：測試圖片 barbara 的 PSNR 比較表

barbara

256(k byte)

512(width)

512(height)

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

表 2：測試圖片 fingerprint 的 PSNR 比較表

fingerprint

256(k byte)

512(width)

512(height)

ratio(%)	PSNR		
	MSSVC	JPEG2000	MSSVC_MDT
0.625	17.26	18.26	18.08
0.9375	18.79	19.41	19.68
1.25	20.62	20.75	20.53
5	27.48	26.63	25.53
10	31.62	30.25	29.41

MSSVC\_MDT 在低位元比率時 (0.625% 和 0.9375%) 有比原本的方法要來的好，尤其當圖片包含較多的線條成分時，例如：fingerprint，PSNR 和視覺品質的差異會更加明顯，然而在 PSNR 比較表也可以看到當位元比率高時，MSSVC\_MDT 沒辦法和原方法或是 JPEG2000 一樣好。

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表 3 到表 5，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。

表 3：節省位元數(百分比) FOREMAN 與 BUS (H frames at temporal levels 1 and 2).

Cumulative bitplane	FOREMAN				BUS			
	LL	LH	HL	HH	LL	LH	HL	HH
2	0.22%	0.17%	0.27%	0.18%	-1.86%	-0.63%	-0.59%	-0.17%
3	0.67%	0.45%	0.51%	0.37%	0.36%	0.51%	0.30%	0.45%
4	0.46%	0.25%	0.28%	0.23%	0.18%	0.23%	0.17%	0.22%

表 4：節省位元數(百分比) (H frames at temporal levels 3 and 4).

Cumulative bitplane	FOREMAN				BUS			
	LL	LH	HL	HH	LL	LH	HL	HH
2	1.04%	1.04%	1.11%	0.91%	-0.71%	-0.32%	-0.49%	0.10%
3	1.47%	1.26%	1.21%	1.20%	0.53%	0.61%	0.29%	0.68%
4	0.81%	0.66%	0.59%	0.83%	0.27%	0.30%	0.14%	0.37%

表 5：節省位元數(百分比) (at the bottommost temporal level).

Cumulative bitplane	FOREMAN				BUS			
	LL	LH	HL	HH	LL	LH	HL	HH
2	-0.05%	0.91%	0.39%	0.97%	0.31%	0.28%	-0.06%	0.05%
3	0.13%	0.13%	0.99%	1.03%	0.24%	0.47%	0.24%	0.61%
4	0.14%	0.80%	0.67%	0.68%	0.15%	0.25%	0.10%	0.30%

## 參考文獻

- [1] D. Taubman, “High performance scalable image compression with EBCOT”, *IEEE Transactions on Image Processing*, vol. 9:7, pp. 1158-1170, July 2000.
- [2] E. Peli, “Contrast in complex images”, *J. Opt. Soc. Amer. A*, vol. 7, pp. 2032-2039, October 1990.
- [3] S. Daly, “The visible difference predictor: An algorithm for the assessment of image fidelity”, in *Digital Images and Human Vision*, A. B. Watson, Ed. Cambridge, MA: MIT Press, 1993, pp 176-206.
- [4] A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor, “Visibility of wavelet quantization noise”, *IEEE Transactions on Image Processing*, vol. 6, no. 8, pp. 1164-1175, August 1997.
- [5] A. P. Bradley, “A wavelet visible difference predictor”, *IEEE Transactions on Image Processing*, vol. 8, no. 5, May 1999.
- [6] Z. Liu, L. J. Karam, and A. B. Watson, “JPEG2000 encoding with perceptual distortion control”, *IEEE Transactions on Image Processing*, vol.1, 14-17, pp. I-637-40, September 2003.
- [7] J. Xu et al, “3D subband video coding using Barbell lifting”, ISO/IEC JTC1/SC29/WG11, MPEG2004/M10569/S05, Munich, March 2004
- [8] I. Hontsch and L. J. Karam, “Adaptive image coding with perceptual distortion control”, *IEEE Transactions on Image Processing*, vol. 11, no. 3, pp. 213-222.
- [9] E. J. Candès. “Ridgelets: Theory and Applications”, PhD thesis, Department of Statistics, Stanford University, 1998.
- [10] M. N. Do, “Directional Multiresolution Image Representations”, Ph.D. Thesis, Department of Communication Systems, Swiss Federal Institute of Technology Lausanne, November 2001.
- [11] P. J. Burt and E. H. Adelson. “The Laplacian pyramid as a compact image code”, *IEEE Transactions on Communications*, vol. 31:4, pp 532 – 540, April 1983.
- [12] A. Cohen, I. Daubechies, and J.-C. Feauveau. “Biorthogonal bases of compactly supported wavelets”. *Commun. on Pure and Appl. Math.*, vol.45, pp 485 – 560, 1992.
- [13] S.-M. Phoong, C. W. Kim, P. P. Vaidyanathan, and R. Ansari. “A new class of two-channel biorthogonal filter banks and wavelet bases”, *IEEE Transactions on Signal Processing*, vol. 43:3, pp 649 – 665, Mar. 1995.
- [14] M. Vetterli. “Multidimensional subband coding: Some theory and algorithms”, *Signal Processing*, vol. 6:2, pp 97 – 112, Feb. 1984.
- [15] S.T. Hsiang and J.W. Woods, “Invertible three-dimensional analysis/synthesis system for video coding with half-pixel-accurate motion compensation”, *SPIE Conference on Visual Communication and Image Processing*, vol. 3653, Jan 1999, pp537-546



## 計畫成果自評

本計畫有以下幾類成果。第一類為 MPEG-4 IPMP System 與 Interframe Wavelet 所發展出的技術、經驗及成品與國際 MPEG 標準直接相關，極具實用價值，可促進國內工業研發技術開發。第二類為將上述技術提案至 MPEG 標準組織，有助我國技術之進入國際舞台，Interframe Wavelet 在 2004 年 3 月與 7 月提案參加 MPEG 標準組織 scalable video coding Call-for-Proposal 競賽。2005 年我們持續提案參加 MPEG 標準組織 scalable video coding 的 Core Experiments。第三類為計畫執行過程所獲得之研究成果論文四篇，已發表於國內外學術會議。其四，參與計畫之同學可獲得國際多媒體最先進的 MPEG-4 與 MPEG-21 相關技術及多媒體系統設計經驗，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：本計畫產出相當多具有學術與應用價值的成果，特別是直接參與國際標準會議，在國際上展示成果。並培育高科技人才培育，整體成效良好。已發表學術論文四篇，碩士學位論文三冊，以及三篇 MPEG 標準提案如下表。

## Publications

- (1) F.-C. Chang, H.-C. Huang and H.-M. Hang, “Combined Encryption and Watermarking Approaches for Scalable Multimedia Coding,” *Pacific Rim Conference on Multimedia 2004*, Tokyo Japan, Dec. 2004. [NSC-92-2219-E-009-008]
- (2) F.-C. Chang and H.-M. Hang, “A Relevance Feedback Image Retrieval Scheme Using Multi-Instance and Pseudo Image Concepts,” *Storage and Retrieval Methods and Applications for Multimedia 2005*, San Jose, USA, Jan. 2005. [NSC 91-2219-E-009-041]
- (3) H.-K. Hsu and H.-M. Hang, “Image compression using directional multiresolution transform,” in *2005 Conf. on Computer Vision, Graphics, and Image Processing*, Taipei, Taiwan, Aug. 2005. [NSC-93-2219-E-009-021]
- (4) C.-H. Hong, and H.-M. Hang, “An HVS-based rate control algorithm for wavelet image coding,” in *2005 Conf. on Computer Vision, Graphics, and Image Processing*, Taipei, Taiwan, Aug. 2005. [NSC-93-2219-E-009-021]
- (5) Chao-Hsiung Hong 洪朝雄, *HVS-based Rate Control Algorithm for Interframe Wavelet Video Coding*, MS Thesis, NCTU, June 2005.
- (6) Han-Kuang Shu 徐漢光, *Directional Multiresolution Transform and Block-Wise Bitplane Arithmetic Coding for Image Compression*, MS Thesis, NCTU, July 2005.
- (7) Chia-Yang Tsai 蔡家揚, *Enhanced Motion Compensated Temporal Filtering for Interframe Wavelet Video Coding*, MS Thesis, NCTU, June 2005.

## MPEG Standard Contributions

1. H.-M. Hang, “ISO/IEC JTC1/SC29/WG11 M11482: AHG Report on Further Exploration in Wavelet Video Coding,” January 2005 (71st, Hong Kong, China)
2. 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)

3. 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)