

# 行政院國家科學委員會專題研究計畫 成果報告

## 子計畫一：可調層次式視訊編碼技術之前處理與後處理技術 研究(1)

計畫類別：整合型計畫

計畫編號：NSC92-2219-E-009-004-

執行期間：92年08月01日至93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：王聖智

報告類型：完整報告

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 10 月 4 日

# 可調層次式視訊編碼技術之前處理與後處理技術研究

計畫編號：NSC-92-2219-E-009-004

執行期限：92 年 8 月 1 日至 93 年 7 月 31 日

主持人：王聖智 (交通大學電子工程系副教授)

計畫參與人員：郭倫嘉、黃至治 (交通大學電子所研究生)

## 一、中文摘要

在這次的進度報告中，我們針對 SNR 可調層次式視訊編碼技術加以研究，藉由適應性調整 macro block 的參數，來最佳化 temporal prediction，以提升視訊的品質。在 FGS 的技術中，因 Robust FGS[1]提供了很有彈性的方法來調整 temporal prediction，因此本研究以 Robust FGS 為基礎，對其參數加以最佳化。實驗證實，本研究所提出的方法可以在不同給定的頻寬區段與不同測試條件下提升視訊的品質，特別是 slow motion 的視訊編碼。

**關鍵詞：**SNR 可調層次式視訊編碼，temporal prediction 控制。

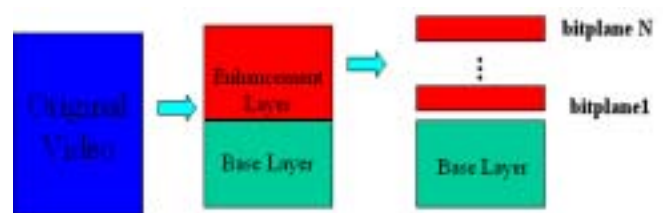
This research proposed an algorithm to adaptively control the temporal prediction for the enhancement layer in FGS coding. We develop our method based on the framework of RFGS, which contains two parameters,  $\alpha$  and  $\beta$ . To represent different degrees of motion activity in different portions of a frame, these two parameters are adjusted based on macro blocks. A three-step search is proposed to choose optimal parameters for each macro block. The experimental results show that the proposed method outperforms the RFGS with fixed parameters, especially for slow motion video sequences.

**Keywords :** SNR Scalability, temporal prediction control.

## 二、進度報告

### (1) 背景

要將龐大的視訊資料以有限的頻寬傳送，必須仰賴高效率的壓縮技術，因此，在過去的幾十年中，視訊的壓縮技術越來越受到重視。傳統的視訊壓縮技術(如 MPEG1)著重於固定頻寬的視訊壓縮，然而，若要將視訊資料傳輸於網路，則必須進行可調層次式視訊編碼，因為網路的頻寬是動態變化的，在這種狀況下，當要對視訊壓縮時，必須考慮一段 bitrate 區間，而非單點，這類視訊技術普遍稱作可調層次式視訊編碼技術(Scalable Video Coding)。FGS 為 SVC 的解決方案之一，並已列入 MPEG4 標準中[2]。FGS 將視訊資料依重要程度區分為代表核心資訊的 base layer，以及用於增加解析度的 enhancement layer[3]，base layer 是以 DCT-based non-scalable single layer 編碼，而 enhancement layer 則再細分為不同的 bitplanes(如圖一所示)，當頻寬不足或軟體資源不足時，可以僅傳送資料量較少的 base layer 資訊而獲得影像品質較差的視訊資料，而當頻寬充裕或軟體資源充足時，則可以將 enhancement layer 中的 bitplane 資料陸續添加進來，以獲得較佳的影像品質。



圖一. Fine Granularity Scalability 示意圖

對 baseline FGS 而言，雖然可支援 scalable coding，但因為 enhancement layer 是獨立編碼，沒有利用預測編碼(temporal prediction)的技術，以 motion compensation 來增加編碼效率所，以與 single layer coding 比起來，編碼效率並不理想，研究者提出了很多方法來增加編碼的效率

[1],[4]-[13]。既然我們希望視訊編碼在一段 bitrate 區間中最佳化，在這個區間中的所有可能狀況都要列入考慮。預測編碼 temporal prediction 在不同 bitrate 區間會呈現不同的效應，在 high bitrate 區間，使用越多 EL 的資料來做 temporal prediction 將會有越多的增益，因為大部分的資料都已經傳送到 decoder 端，誤差飄移(error drift)將會很小，相反的在 low bitrate 區間應使用較少 EL 的資料來做 temporal prediction，因接收端收到的資料量並不多。在本研究中，我們試著探討 motion compensation gain 與 error drift 之間的交互作用，並嘗試平衡 motion compensation gain 與 error drift，藉以將視訊編碼在一段區間內最佳化。

## (2) 架構

### 2.1 問題呈現

給定一 bitrate 區間 PB 及 N 個監測點(checkpoints)  $S_1 \dots S_N$ ,  $S_i < S_j$  if  $i < j$ 。如前節所述，temporal prediction 在不同 PB 區間會呈現不同的效應，對於 high  $S_i$  而言，使用越多 EL 的資料來做 temporal prediction，將會產生越多的增益  $G_i$  且較少的誤差飄移  $L_i$ ，相反的，對於 low  $S_i$  而言，使用越多 EL 的資料來做 temporal prediction，將會造成較少的增益  $G_i$  且較多的誤差飄移  $L_i$ 。既然我們想要在區間 PB 中最佳化視訊品質，在  $S$  中所有的監測點都要考慮。在不同的 FGS 方法中，RFGS 提出了一個很有彈性的架構來控制 temporal control，因此，我們的系統是以 RFGS(圖二)為基礎來最佳化 temporal prediction。在 RFGS 中有  $\alpha$  與  $\beta$  兩個參數，其中  $\beta$  決定要取幾層 enhancement layer 的 bitplane 來做 motion compensation ( $0 \leq \beta \leq \text{maximum bitplane}$ )， $\alpha$  為小於 1 的參數 ( $0 \leq \alpha \leq 1$ ) 來控制 error drift，因此本研究的目標就是找出最佳的  $\alpha_{opt}$  與  $\beta_{opt}$ ，藉以達成在 PB 區間中最大的平均編碼效率，可表示成：

$$(\alpha_{opt}, \beta_{opt}) = \arg \max_{\alpha, \beta} \left( \frac{1}{N} \times \sum_{i \in S} (G_i + L_i) \right). \quad (1)$$

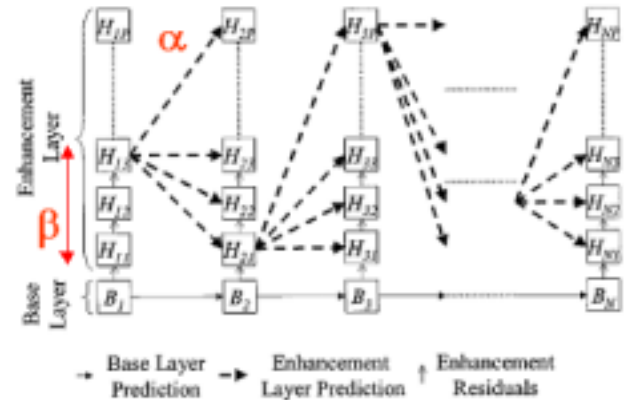
在本研究中，我們以 PSNR 作為客觀憑斷視訊影像品質的標準。PSNR 定義為：

$$PSNR = 20 \times \log_{10} \left( \frac{255}{MSE} \right) \quad (2)$$

其中 MSE 為解碼端接收視訊( $f'$ )與原來的視訊 ( $f$ ) 之 mean square error。

$$MSE = \sum \frac{(f(i, j) - f'(i, j))^2}{N^2} \quad (3)$$

所以本研究的目的就是找出最佳的  $\alpha_{opt}$  與  $\beta_{opt}$  來使得  $S$  中平均 PSNR 達到最大值。



圖二. RFGS 架構與參數

### 2.2 MB-based 參數選擇

在 RFGS 原本的架構中，參數  $\alpha$  與  $\beta$  是以 frame 為單位，整個 frame 指定相同的  $\alpha$  與  $\beta$ ，但在同一個 frame 中，有些部分動的快，有些動的慢，因此，對整個 frame 指定相同的參數並不適合，若能針對不同的區塊適應性的設定不同的參數，應可增加編碼的效率。以 CIF 格式而言，每個 frame 有 396 個 Macro Block，如此的解析度可允許每個 MB 大致對應到影像中不同的部分，因此，我們提出以 MB 為主的參數調整方式，也就是在同一個 frame 中的每個 MB 都有自己的  $\alpha$  與  $\beta$ 。假設一個 frame 包含了 K 個 MBs, MB(1)...MB(K)，且假設 Enhancement Layer 中 maximum bitplane 為 M， $\alpha$  被 quantized 為 Q steps，則所有可能的 solution space 為  $(K)^{Q \times M}$ ，為了在再這麼大 solution space 中找出最佳解，我們先做 initial guess，然後再以迭代(iterative)的方式搜尋最佳解。我們使用在解碼端得到的平均 PSNR(定義為 APR)來估計 motion compensation gain G and error drift loss L 的淨效應。對於每個 MB(j),  $1 \leq j \leq K$ ，我們先決調整趨勢  $TR(j)$ ， $TR(j)$  定義為：

$$TR(j) = 0 \quad \text{if} \quad \max(\beta_0(j)^+, \beta_0(j)^-) < \beta_0(j)$$

$$\begin{aligned} &+1 \quad \text{else if } \text{APR}(\beta_0(j)^+) > \text{APR}(\beta_0(j)^-) \\ &-1 \quad \text{otherwise,} \end{aligned} \quad (4)$$

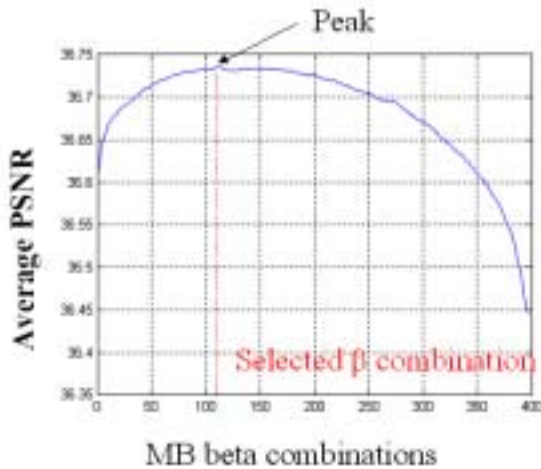
其中  $\beta_0(j)^+$  and  $\beta_0(j)^-$  代表對第  $j$  個 MB 的  $\beta$  上下移動一個 step 且保持其餘的 MBs 的  $\beta$  值不變的狀況下所得到的新  $\beta$  組合，接著，定義 Impulse gain likelihood function LG 為：

$$\begin{aligned} \text{LG}(j) = &\max(\text{APR}(\beta_0(j)^+) - \text{APR}(\beta_0), \\ &\text{APR}(\beta_0(j)^-) - \text{APR}(\beta_0)). \end{aligned} \quad (5)$$

函數 LG 提供了一個好的指標，來決定哪個 MB 參數變化應該有較高的優先權，對於具有較大 LG 值的 MB 而言，因其對 PSNR 增益的效果較好，所以應具有較高的改變優先權，因此我們對 LG 以 descending order 加以排序， $\text{LG}'(p) > \text{LG}'(q)$  if  $p < q$ 。基於  $\text{LG}'$ ，我們可得一個將 MB 重新排序的 prioritized sequence OD，並將對應之 TR 改為  $\text{TR}'$ ，接著，我們依照 MB 在 OD 的先後次序依序的改變  $\beta$  值，可得測  $K$  組試向量  $\beta_1', \dots, \beta_K'$ ：

$$\begin{aligned} \beta_u(\text{OD}(j')) &= \beta_0(j') + \text{TR}'(\text{OD}(j')) * \text{SP} \quad \text{if } j' > u \\ \beta_u(\text{OD}(j')) &= \beta_0(j') \quad \text{otherwise,} \end{aligned} \quad (6)$$

其中  $1 \leq u \leq K$  and  $1 \leq j' \leq K$ 。SP 為單位增量 step，在本研究中設為 1，接著，量測各測試向量的 APR，找出產生最大 APR 的測試向量  $\beta_0'$ ，以  $\beta_0'$  為這個 iteration 的最佳參數組合。接著，以  $\beta_0'$  為下一階段的 initial guess，可以迭代的方式逐一找出接下來每個 iteration 的最佳參數組合。



圖三. 最佳參數組合搜尋

### 2.3 三階段搜尋(Three Step Search)

基於 2.2 節提出的方法，我們提出了一個 three-step search 演算法，同時將  $\alpha$  與  $\beta$  列入考慮，並且為了權衡時間複雜度與精確度，我們將 iteration 的數目限制在三次，演算法如下：

Phase 1: 為了增廣搜尋區域並避免落入 local minimum，第一階段以 frame-based 的方式決定整張 frame 的  $\alpha$  與  $\beta$  (共  $M \times Q$  測試向量)，設為  $\alpha_0'$  與  $\beta_0'$ 。

Phase 2: 以  $\alpha_0'$  與  $\beta_0'$  為 initial guess，對 frame 裡的 MB，以  $\alpha_0(j)$  與  $\beta_0(j)$  為中心，向上及向下位移一個單位 SP，形成定義區間組合 (a,b),  $a \in (\alpha_0(j) - \text{SP}, \alpha_0(j), \alpha_0(j) + \text{SP})$ ,  $b \in (\beta_0(j) - \text{SP}, \beta_0(j), \beta_0(j) + \text{SP})$ ，對這些組合加以排序，並找出區間中具有最大 APR 的參數組合，來決定 TR and LG 接著以 2.2 節的方法找出  $\text{LG}'$ ,  $\text{TR}'$ ，及 OD，藉以得到第二 iteration 的最佳參數組合  $\alpha_0'$  與  $\beta_0'$ 。

Phase 3: 以  $\alpha_0'$  與  $\beta_0'$  為 initial guess，重複 phase2 的動作，並以產生 peak 的  $\alpha_0'$  與  $\beta_0'$  為 suggested optimal  $\alpha$  與  $\beta$  之組合  $\alpha_{opt}$  與  $\beta_{opt}$ 。

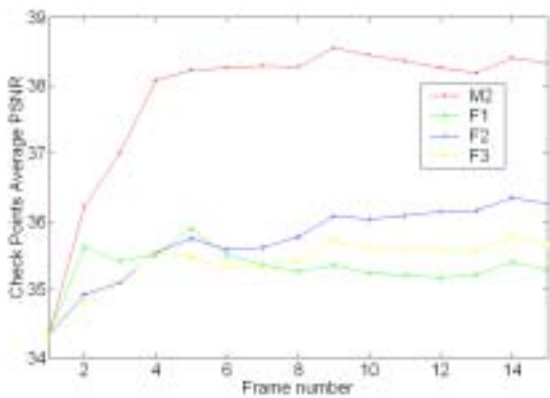
### (3) 實驗結果

為了驗證提出方法的可行性，我們把提出的方法應用在常用的測試視訊，並紀錄其結果。測試條件如下：

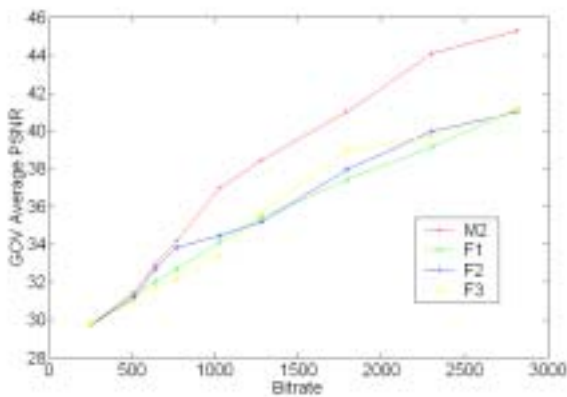
- A.  $S^p = \{256, 284, 512, 768, 1024, 1536, 2048, 2560\}$  kbps.
- B. GOV = 15 (one I frame followed by 14 P frames).
- C. Video format= CIF, 30 frames per second.
- D. Base layer= 256 kbps, no rate control

測試視訊包含三個 video sequences, News, Foreman, and Stefan，分別代表 slow motion, moderate motion, and fast motion。為了將提出的方法(以 M2 表示)與原來的 RFGS 做比較，我們以實驗的方式得到當  $\text{PB}=(256\text{k}, 2560\text{k})$  時，dominated  $\beta$  為 (2,3,4)，然後，對這三個  $\beta$  找出最佳的  $\alpha$  配對，分別為 F1(0.9,2), F2(0.8,3), 與 F3(0.8,4)。圖四 (a) 顯示對 "News" 的實驗結果，與固定參數的 RFGS 比較 (F1,F2,F3)，平均 PSNR 增益約為 2dB，證明了我們所提

出的方法對於 slow motion sequence 的編碼有相當好的效果。圖四(b)顯示  $S$  中各個不同的監測點的平均 GOP(15 frames) PSNR，對於固定參數的 RFGS 而言，若採用比較多的 temporal prediction，將導致 high rate performance 較好，但 low rate drift error 較大，相反的比較少的 temporal prediction 會導致 lowrate performance 較好，但 highrate efficiency 不足。然而，由圖四(b)可看出，我們所提出的適應性調整方法幾乎在所有監測點都有不錯的表現。

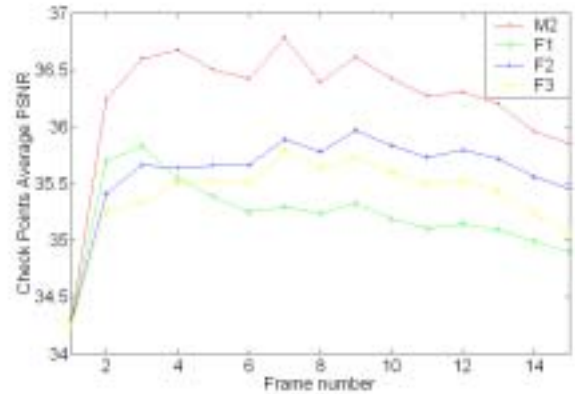


圖四(a). News sequence (frame v.s. PSNR)

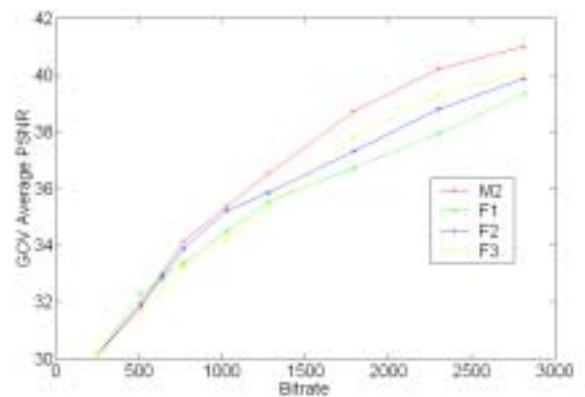


圖四(b). News sequence (rate v.s. PSNR)

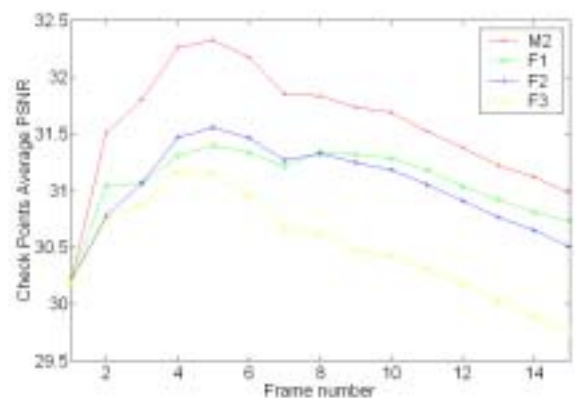
圖五 (a)(b) 與圖六 (a)(b) 顯示對 video sequence "Forman" and "Stefan"的實驗結果，雖然本研究提出的方法似乎對 moderate motion 及 fast motion 的增益沒有 slow motion 大，但仍有一定程度的提升，約 0.7-0.3dB。



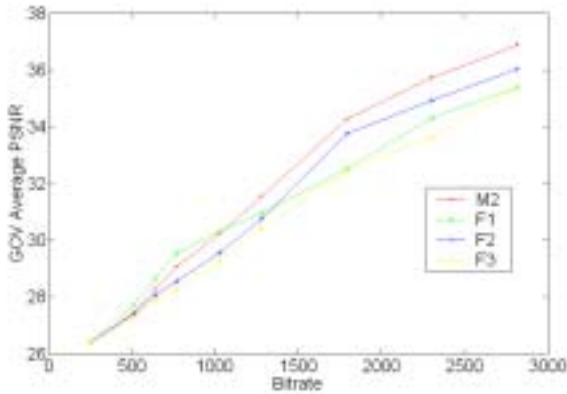
圖五(a). Foreman sequence (frame v.s. PSNR)



圖五(b). Foreman sequence (rate v.s. PSNR)



圖六(a). Stefan sequence (frame v.s. PSNR)

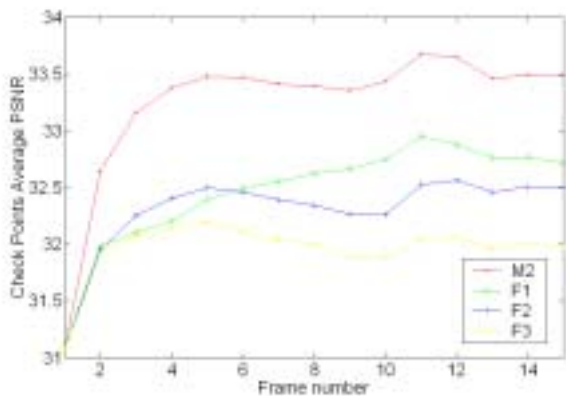


圖六(b). Stefan sequence (rate v.s. PSNR)

為了驗證本研究所提方法之通用性，我們改用不同的測試條件：

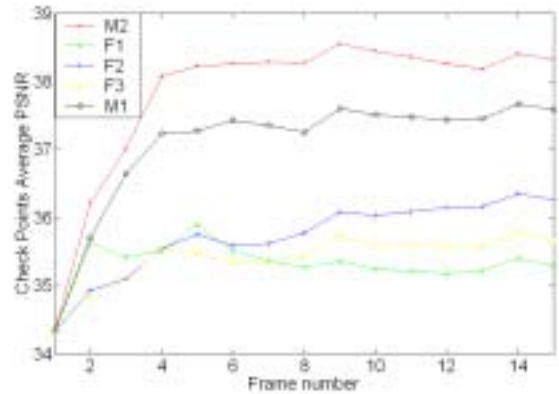
- A. PB=(128,256) kbps.
- B. Video format= CIF, 15 fps.
- C. Base layer=256 kbps, 有 rate control.

我們以”foreman”進行測試，結果如圖七，由圖七可知，我們所提的方法表現依然優於固定參數的 RFGS。此外，這個實驗也證實了我們所提的方法可以與 Rate control 並用，更增加編碼的效率。不同的測試結果列於表一。



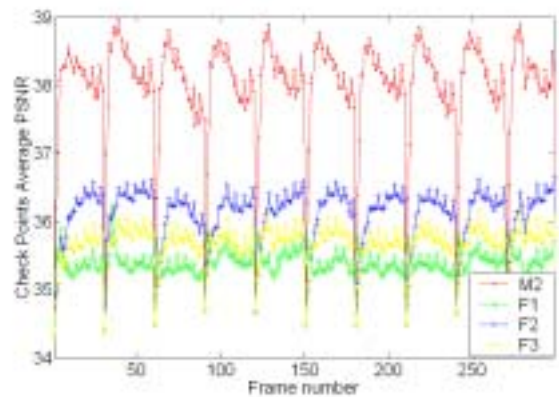
圖七.提出方法在不同測試條件下之結果

若將 leaky factor  $\alpha$  設為 1 (以 M1 表示)，則只針對  $\beta$  做最佳化。圖八顯示 M2 與 M1 的比較，由圖可知，對兩個參數同時做最佳化的效果會優於只對  $\beta$  做最佳化。另一方面 M1 的 PSNR 表現仍明顯優於固定參數的 RFGS，這個結果也顯示了我們所提的方法可以應用到不同的 FGS 架構，像 MC\_FGS[4]。



圖八. M1 與 M2 之比較

為了進一步驗證本研究所提的方法的可行性，我們對 news 的整個 sequence 加以測試 (GOV=30)，結果如圖九。實驗結果證實，本研究所提的方法在不同的 GOV 區間均有不錯的表現，平均的 PSNR 增益為 2dB。



圖九. PSNR performance for entire sequence of News (300 frames)

#### (4) 結論

在這次的進度報告中，我們提出了一個架構可適應性調整 macro block 的參數來平衡 motion compensation gain 與 error drift loss，藉以最佳化 temporal prediction 以及提升可調式視訊的品質。因 Robust FGS 提供了很有彈性的方法來調整 temporal prediction，因此本研究以 Robust FGS 為基礎，對其參數  $\alpha$  與  $\beta$  加以最佳化。實驗證實，本研究所提出的方法可以在不同的頻寬區段與不同測試條件下提升視訊的品質，對於 slow motion 的視訊而言，PSNR 增益可達 2dB 對 moderate motion 與 fast motion 的視訊則約 0.7-0.3dB。

## (5) 文献

- [1] H.C. Huang, C.N. Wang, and T. Chiang, "A robust fine granularity scalability using trellis-based predictive leak," *IEEE Trans. Circuits Syst. Video Technol.*, pp. 372-385, vol. 12, no. 6, Jun. 2002.
- [2] *Streaming Video Profile-Final Draft Amendment (FDAM 4)*, MPEG01/N3904.
- [3] W. Li, "Overview of fine granularity scalability in mpeg-4 video standard," *IEEE Trans. Circuits Syst. Video Technol.*, pp. 385-398, vol. 11, no. 3, March 2001.
- [4] M. van der Schaar and H. Radha, "Motion-compensation based fine-granular scalability (MC-FGS)," ISO/IEC JTC1/SC29/WG11, MPEG00/M6475. Oct. 2000.
- [5] F. Wu, S. Li, YQ Zhang, "A framework for efficient progressive fine granularity scalable video coding," *IEEE Trans. Circuits Syst. Video Technol.*, March 2001.
- [6] S.R. Chen, C.P. Chang, and C.W. Lin, "MPEG-4 FGS coding performance improvement using inter-layer prediction," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, May 2004, Montreal, Canada.
- [7] Y. Liu, Z. Li, Paul Salama, and E. J. Delp, "A discussion of leaky prediction based scalable coding," in *Proc. IEEE Int. Conf. on Multimedia and Expo (ICME)*, pp. 565-568, vol. 2, July 6-9, 2003, Baltimore, MD.
- [8] C.C Ho, J.L. Wu and S.W. Wang, "A user adaptive perceptual rate control scheme for FGS videos," in *Proc. IEEE Conf. Consumer Electronics 2003 (ICCE'03)*, pp. 42-43, June 2003, LA, USA.
- [9] W.N. Lie, M.Y. Tseng, and I.T. Ting, "Constant-quality rate allocation for spectral fine granular scalable (SFGS) video coding," in *Proc. IEEE Int. Sym. on Circuits and Systems (ISCAS)*, pp. 800-803, May, 2003.
- [10] T. Kim and M. Ammar, "Optimal quality adaptation for MPEG-4 fine-grained scalable video," in *Proc. IEEE INFOCOM*, pp. 641-651, 2003.
- [11] J. Zhou, H. Shao, C. Shen, M.T. Sun, "FGS enhancement layer truncation with minimized intra-frame quality variation", *IEEE International Conference on Multimedia and Expo (ICME)*, vol. 2, pp. 361-364, July 2003.
- [12] K. Ugur, P. Nasiopoulos and R. K. Ward, "Extremely fast selective enhancement method for fine granular scalable enabled H.264 video," in *Proc. IEEE Canadian Conference on Electrical and Computer Engineering*, pp. 1103-1106, Montreal, May 2003.
- [13] X.K. Yang, W.S. Lin, Z.K. Lu, E.P. Ong, and S.S. Yao, "On incorporating just-noticeable-distortion profile to motion-compensated prediction for video compression," in *Proc. IEEE Conf. Image Processing (ICIP)*, pp. 833-836, 2003.

Video sequence	Test condition	Average PSNR gain
Akiyo	GOV=15,fps=30, Base layer=256k, Target rate=[256k,2560k]	1.56 dB
Foreman	GOV=15,fps=30, Base layer=256k, Target rate=[256k,2560k]	0.65 dB
Crew	GOV=30,fps=15, Base layer=256k, Target rate=[384k,750k]	0.73 dB
Stefan	GOV=15,fps=30, Base layer=256k, Target rate=[256k,2560k]	0.29 dB
News	GOV=15,fps=30, Base layer=256k, Target rate=[256k,2560k]	1.91 dB
Foreman	GOV=30,fps=30, Base layer=128k, Target rate=[128k,256k]	0.77 dB
Paris	GOV=15,fps=30, Base layer=256k, Target rate=[256k,2560k]	1.27 dB

表 1. 實驗結果