行政院國家科學委員會專題研究計畫 成果報告

使用空間與時間關聯預測之H.264/AVC快速Intra/Inter模式決定系統
研究成果報告(精簡版)

計 畫 類 別 ：個別型
計 畫 編 號 ：NSC 98-2221-E-216-034-
執 行 期 間 ：98年08月01日至99年07月31日
執 行 單 位 ：中華大學資訊工程學系

計 畫 主 持 人 ：連振昌
共 同 主 持 人 ：周智勳

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 99年10月27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

（使用空間與時間關聯預測之 **H.264/AVC** 快速 **Intra/Inter** 模式決定系統）

計畫類別：■個別型計畫　□整合型計畫
計畫編號：**NSC 98-2221-E-216 -034**
執行期間：98/08/01 ～ 99/07/31

執行機構及系所：中華大學資訊工程學系

計畫主持人：連振昌
共同主持人：周智勳
計畫參與人員：余冠霖、黃耀德

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

本計畫除繳交成果報告外，另須繳交以下出國心得報告：
□赴國外出差或研習心得報告
□赴大陸地區出差或研習心得報告
■出席國際學術會議心得報告
□國際合作研究計畫國外研究報告

處理方式：除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

中華民國九十九年十月二十七日

一、 前言

在 H.264/AVC 編碼標準[1]，對於 Inter frame 運動向量預估模式、entropy coding、多張參考 frame 之 Inter frame 編碼、I-frame 的 Intra frame 預測編碼模式及更精準的次像素運動向量預估模式微調等，提供很大的彈性以獲得最佳的 R-D cost。對於 Inter frame 運動向量補償編碼中，最主要的運算落在每一個方塊內，必須從 4×4 到 16×16 的七種動態預估模式中找尋最佳的動態補償誤差值(motion compensation error)。對於 Intra frame 預測模式編碼中，8×8 色度(Chroma)區塊與 16×16 亮度(Luma)區塊會有四種預測模式，4×4 亮度區塊會有八個方向及一個 DC 的預測模式。因此，我們可以發現在 JM-14.0 參考軟體中，高複雜的 Intra/Inter frame 完整搜尋方式造成編碼過程非常耗時。所以，許多的方法應運而生，如 SAD、相同區塊的分析、邊緣偵測等都是為了抉擇最佳的動態預估模式。然而，這些的運用都需要在動態補償過程中，花費額外的時間與資源。在此計畫中，我們所發展之技術不需要額外的影像處理，就可以達成的快速視訊壓縮編碼。本計劃構思利用當前畫面和參考畫面間，分析 Intra/Inter 預測模式空間與時間的關聯性，找出在空間和時間上,每一個 macroblock 預估模式的關聯性後,建構快速模式決定的方法。而且，在模式處理過程中，採取漂移補償來避免誤差的累積。目前於實驗數據中，可以發現總運算量的時間可以節省約百分之七十,傳輸率(bit rate)方面只增加百分之三點七，平均的 PSNR 大約只有損失 0.08 dB。

二、 研究目的

新一代的視訊編碼標準 H.264/AVC[1]是 Joint Video Team (JVT) [1]所發展的低位元率壓縮技術。在 H.264/AVC 編碼標準[1]，對於 Inter frame 運動向量預估模式、entropy coding、多張參考 frame 之 Inter frame 編碼、I-frame 的 intra frame 預測編碼模式及更精準的次像素運動向量預估模式微調等，提供很大的彈性以獲得最佳的 R-D cost。特別是在七種運動向量預估模式從 4×4 到 16×16 模式在每個 macroblock 找最小運動補償誤差值。在當前的 JVT 參考軟體中，七種模式(各式各樣的 block sizes)被應用於進行運動補償過程，此類的最佳化的 R-D cost 定義如(1)

$$J_{Mode} = D + \lambda_{Mode} \times R \tag{1}$$

在式子中 D 表示 macroblock 運動補償誤差值，λ_{Mode}是 Lagrange multiplier，R 表示要求的位元率。為了得到最佳的運動向量估計模式，我們必須在一些費時的過程，例如包含運動向量估計，DCT 轉換，量化器，計算每一個運動向量估計模式的 R-D cost。

另外，利用七種運動向量估計模式從 4×4 到 16×16 (見 Fig. 1)去決定在每個 macroblock 最佳的運動向量估計模式。對 intra 預測模式，利用在 Fig. 2 中的四種預測模式去編碼 8×8 chroma 與 16×16 luma blocks 及利用在 Fig. 3 中的八種預測方向和一種 DC 預測去編碼 intra 4×4 luma block 。然而，在參考軟體 JM-14.0[6]使用全域搜尋法(Full Search)會需要大量的計算量，使得此技術變的無效率。

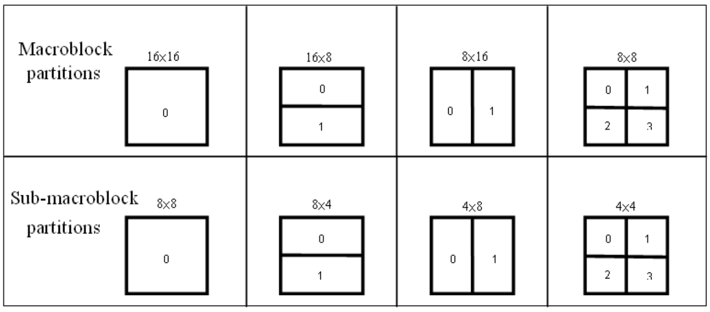


Fig. 1 Macroblock and Sub-macroblock partitions in H.264

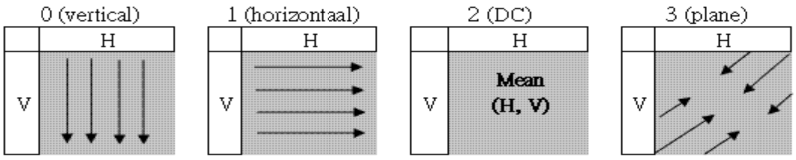


Fig. 2 Prediction modes in Intra 16×16 luma and Intra 8×8 choma blocks.

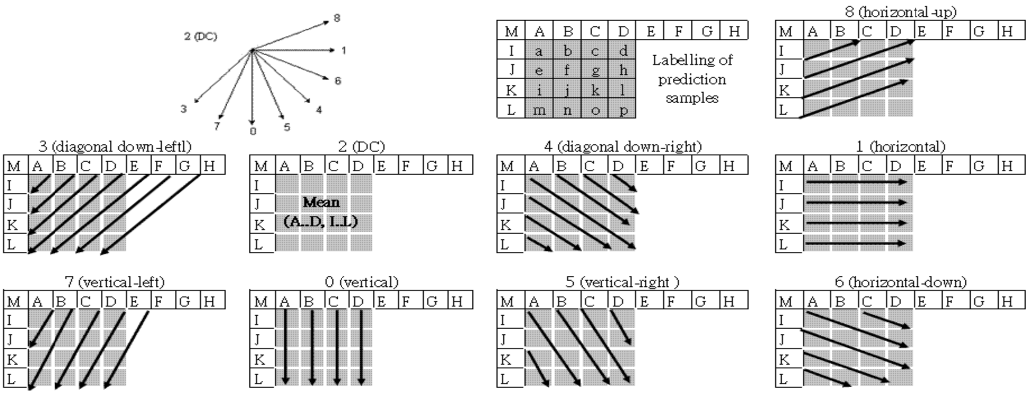


Fig. 3 Prediction modes in Intra 4×4 luma block.

三、 文獻探討

近年來，許多研究[7-14]提出關於快速模式的決定方法，在文獻[7]，每個 macroblock 應用目前的 frame 和前一張 frame 之間的 SAD (sum of absolute difference)，去估計哪一種的 block 適合於運動向量估計過程。在文獻[7]的方法如下所述，第一，計算 SAD 和事先定義的 threshold 比較，如果 SAD 比 threshold 值小，則運動向量估計模式將從 16×16，16×8 和 8×16 block 類型之中選擇。否則，運動向量估計模式將會由估計所有七種 block 類型去決定。在文獻[8]，邊緣偵測[9]被應用於將 homogeneous 和 non-homogeneous 的區域分類。如果 macroblock 屬於 homogeneous 的區域則運動向量估計模式 4,5,6 和 7(8×8,8×4,4×8 和 4×4) 會被移除掉。否則，macroblock 會被分成許多大小 8×8 的 sub-macroblock 且分析每一個 sub-macroblock 是不是屬於 homogeneous 的區域。如果每一個 8×8 的 sub-macroblock 是 homogeneous 則運動向量預估模式 5,6 和 7(8×4,4×8 和 4×4)

會被移除掉。直到所有的 sub-macroblock 都被計算過這個程序才停止。最後，rate-distortion 最佳化被利用於去決定哪個是最好的運動向量預估模式。在文獻[9]，Pan 等應用邊緣方向值方圖去降低 intra 預測模式的決定的計算成本。在文獻[10]，Zhu 等應用低解析度的影像和邊緣偵測去決定運動向量預估模式 (8×4,4×8 和 4×4)。在文獻[11]，Meng 等應用兩個方法降低模式決定的複雜度。第一，marroblock decimation 被利用於估計模式的最小的粗略 R-D cost。第二，如果 R-D cost 比事前定義的 threshold *T* 來的小，則我們選擇這個模式和這它的兩個鄰近的模式當作候選的模式。

對於 H.264/AVC 的 intra coding，文獻[12]Cheng 等應用三個步驟階層搜尋方法去預測 4×4 intra 預測方法。第一，三個模式 0, 1 和 2 根據 R-D cost 被計算。第二，基於步驟一的作法估計，兩個模式求一組(5,7)和(6,8)的值。最後，分析四組(4, 5) ,(3, 7), (4, 6) 和 (6, 8)的 R-D cost 去選擇最佳的 intra 預測模式。在每個步驟，獲得最小的 R-D cost 決定下一個去計算，且它會降低計算 cost 從 9 模式到 6 模式。在文獻[13]，Hwang 等應用時間的資訊去預測 intra 模式在 P-frames。對每一個 block 在同一個位置 previous 預測模式且它鄰近的 block 被當作候選預測模式去減少重要的搜尋方法。在文獻[14]， Lien 等不只應用時間的資訊也利用空間的相關性去發展新的快速模式決定方法利用空間-時間預測模式。空間-時間相關性在參考 frame 和鄰近 macroblocks 之中被分析去發展快速模式決定方法不利用額外的影像處理。此外，drift compensation 的概念被採用避免在模式決定的過程累積的誤差值。

在本計劃，不但利用在參考 frame 和鄰近 macroblocks 之中時間-空間的關聯性開發快速模式決定(運動估計)方法，也使用 P-frame 的 intra 預測模式。這個新的想法不需要額外的影像處理且這個新的想法可以比[14]省更多的 cost。除此之外，此計劃有別於文獻[14]，我們除了考慮 interframe coding 也同時考慮 intraframe frame coding. 也就是較完整的研究如何改善 H.264/AVC 的編碼效能。根據比較 JM-14.0 參考軟體，實驗結果顯示總共的計算 cost 可減少 70%以上，總共的 bitrate 只增加 3.7%而且平均的 PSNR 只有下降大約 0.08dB。

四、 研究方法

4.1 H.264/AVC 以及 JM-14.0 參考軟體

近年來，新的視訊編碼標準 Advanced Video Coding (AVC)被提出到 ISO/IEC 國際標準 ISO/IEC 14496-10 視訊編碼和 ITU-T H.264[1]。H.264/AVC 編碼的主要目的是要為了改善壓縮效率、網路適應性、抗錯能力與互動式和無互動式視訊應用。為了獲得最佳的 R-D cost，在 H.264/AVC 編碼標準中，對於運動估計模式有很大的彈性，提供多張參考 frame、多種 I-frame 的 intra 預測模式、更精準的次像素運動向量預估模式微調、entropy coding…等。H.264/AVC 編碼架構如 Fig.4 中所顯示。

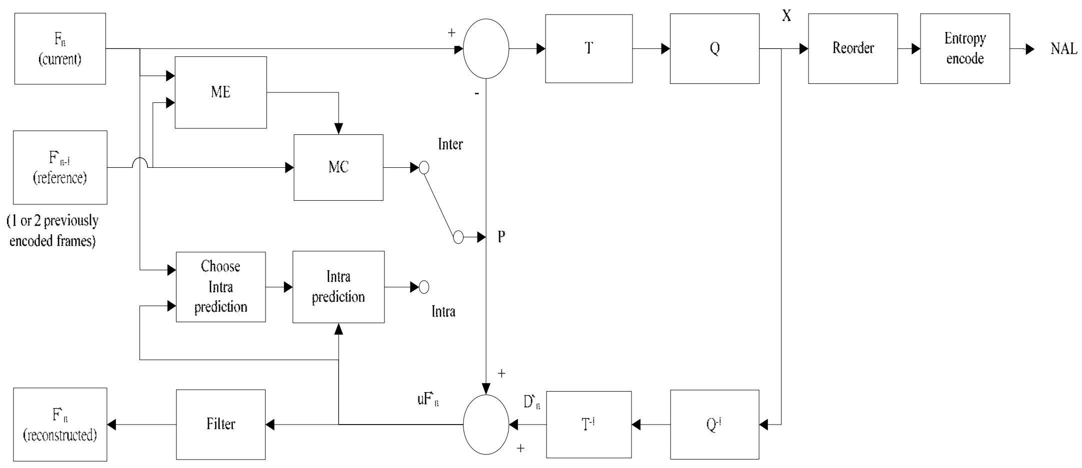


Fig. 4 The structure of H.264 Encoder

4.1.1 H.264/AVC Baseline 組態

概括來說，H.264/AVC 有四種編碼器組態：baseline、main、extended 和 fidelity range extensions(FREXT)[16]。在本計劃中，我們只利用 baseline 組態來發展快速模式決定方法對於 Intra/Inter frame 編碼。在 baseline 組態中，只支援 I-和 P-slices。I-slices 由 intra-coded macroblock 構成，而 P-slices 是由 intra-coded，inter-coded 和 skipped macroblock 構成。在 P-slices 的 inter-coded macroblock 是經由多重參考 frame(先前編碼的畫面)的運動補償處理編碼而來的。接著冗餘資訊用 4×4 的整數轉換器[14]轉換和量化。轉換係數是藉由 context-adaptive variable length coding scheme (CAVLC)[17-18]編碼。

在傳統的視訊編碼標準中，例如，Mpeg-II，參考 frame 為了編碼 P-frame 是固定的當作前面的 frame。不過，在 H.264/AVC 中參考 frame 可能是許多前面的 frame。另外，運動補償預測處理的 block size 不再限制為 16×16。在 H.264/AVC 中，block size 可能從 16×16 到 4×4(Fig.1)的變化。總共有七種 block sizes 可以用在運動補償預測的處理。因此，運動向量將會在每一個 macroblock 和 sub-macroblock 內來做計算。如果編碼系統選擇大的預測尺寸(16×16, 16×8 或 8×16)則較少的位元數來編碼運動向量。相反的，如果邊編碼系統選擇小的預測尺寸(8×8, 8×4,4×8 或 4×4)則編碼運動向量要求較大的位元數。

對於 intra 預測編碼系統，macroblocks I16MB(16×16)和 I8MB(8×8)被四種 intr 預測模式(見 Fig. 2)包含模式 0(垂直)，模式 1(水平)，模式 2(DC)，模式 3(平面)來編碼。在 Fig. 3 有一些範例對 4×4 block 的 intra 編碼的說明。如果選擇模式 1，pixels a, b, c 和 d 會被鄰近的 pixel I 預測到；pixels e, f, g 和 h 會被鄰近的 pixel J 預測到，諸如此類。如果選擇模式 4，pixel c 會被(A/4 + B/2 + C/4)的四捨五入後的數值預測到，pixel d 會被(B/4 + C/2 + D/4)的四捨五入後的數值預測到諸如此類。如果選擇模式 2，所有的 pixels 會被(A+B+C+D+I+J+K+L)/8 的數值預測到。

因此，在 H.264/AVC 中 intra 預測模式的選擇是一個非常耗費時間的過程。luma 和 chroma 預測模式在一個 macroblock 的搜尋數量計有 M8×(M4×16+M16) 和 592(4×(16×9+4))，被需求去找最好的 RDO 模式。

4.1.2 The JM-14.0 編碼結構

在本計劃中，利用 H.264/AVC 參考軟體 JM-14.0[6]去發展快速模式決定方法。Fig. 5 顯示 JM-14.0 編碼器的主要部份的流程圖。在 JM-14.0 參考軟體中，configuration (encoder_baseline.cfg)被初始化在 main”function 的開頭。然後，frame_picture function 用來定義 interlaced 或是 progressive 的視訊型態。在 field_picture function 中，PartitionMotionSearch function 用來分割 frame 成為 macroblock，encode_one_macroblock function 用來編碼每一個 macroblock。最後 BlockMotionSearch function 選擇運動估計模式用於運動補償編碼。

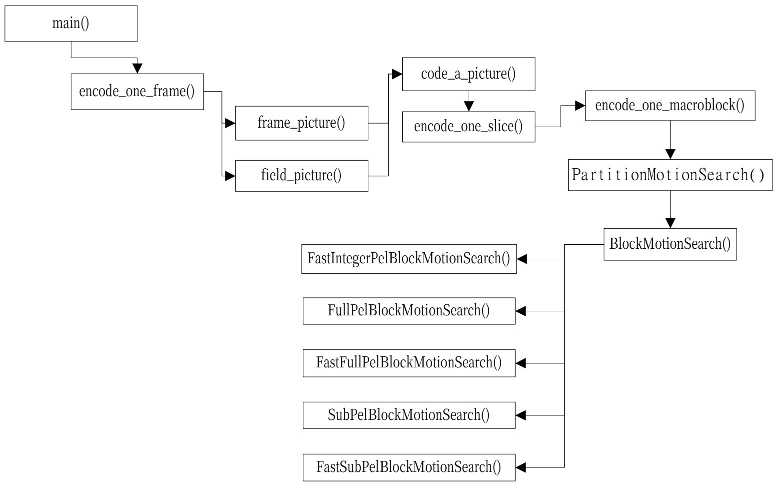


Fig. 5 Flowchart of the JM-14.0 encoder.

Fig. 6 顯示在參考軟體 JM-14.0 中再每一個 macroblock 運動估計模式的決定方法與視訊壓縮編碼的 rate distortion 最佳化的流程。

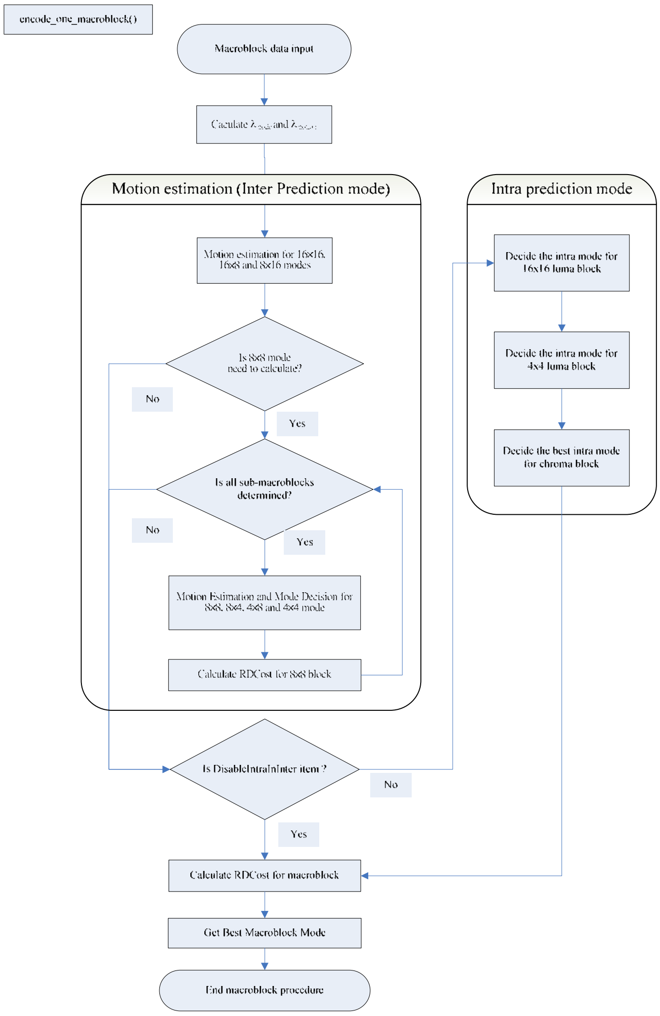


Fig. 6 The flowchart of function “encode_one_macroblock”.

4.2 快速模式決定方法

在參考軟體 JM-14.02 中用全域搜尋處理(16×16, 8×8,…, 4×4)去決定估計模式，如此會使編碼過程很沒有效率。在本計劃，我們所提出的快速模式決定方法，首先會先分析空間與時間的關聯性，找出在空間和時間每一個 macroblock 估計模式的關聯性後，建構快速模式決定的方法。

4.2.1 空間-時間模式關聯性

經由仔細的觀察在 JM-14.0 參考軟體裡模式決定過程，一個 macroblock 的運動估計模式與上一張的參考 frame(多重參考 frame)中 macroblock 鄰近的相同位置的模式有很高的相關性。在 Fig.7 中描述模式關聯性

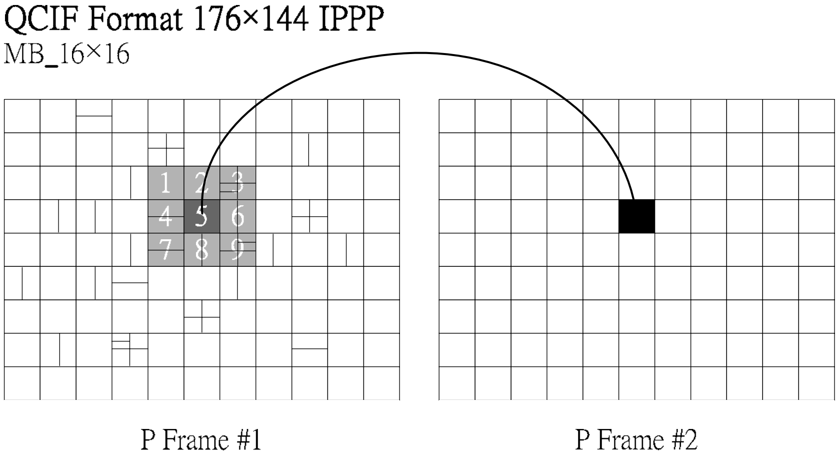


Fig. 7 The mode correlation for a macroblock between adjacent frames.

在 P-frame #2 圖中所示的 macroblock，所採用之運動向量預估模式的機率會近似於在 P-frame #1 中以此 macroblock 位置為中心的鄰近 macroblock (在 Fig.7 P-frame #1 的編號 1~9) 所採用之運動向量預估模式。根據 JM 參考軟體，運動向量預估模式關聯性之分析列在 Table 1。在這裡，我們分析運動向量預估模式：16×16, 16×8,8×16, 8×8 和 skip 模式之模式關聯性。

Table 1. Probabilities for the mode of a macroblock on P-frame #2 is similar to the modes of the neighboring macroblocks centered at the same position on P-frame #1.

Sequences	QP28	QP32	QP36	QP40
Akiyo	95.11%	96.02%	97.22%	98.26%
Container	93.35%	94.88%	96.40%	97.55%
Hall_Monitor	95.28%	96.62%	97.16%	97.51%
Moth&Daug	91.77%	93.59%	95.11%	96.42%
News	92.90%	93.63%	94.52%	95.73%
Salesman	94.63%	95.10%	96.00%	97.31%
Carphone	88.54%	90.16%	92.32%	94.40%
Coastgrd	86.05%	87.31%	89.17%	92.70%
Foreman	87.28%	88.29%	89.22%	90.90%

在 Table 1 中，九個長度 200 frame 的 video sequences 被使用在分析模式關聯性。此外，量化的 step size 分別設定為 QP=28, 32, 36 和 40。對於用 QP=28 的 Foreman video sequence，macroblock 的估計模式的可能性是接近於前一張參考 frame 中鄰近的 macroblock 相同位置的中心附近的估計模式是 87.28%。從 Table 1 的分析結果中，觀察到模式關聯性對於在兩個相鄰的 frame 的 macroblock 是很高的。另外，對於模式關聯性的亮個額外的實驗線顯示在 Table 2 和 3 中。在 Table 2 和 3 中，skip 和 16×16 模式被當

程式同一類。Table 2 和 3 說明 macroblock 的估計模式相似於第一名和前二名分別在前一張參考 frame 的鄰近 macroblock 之估計模式機率。

Table 2. Probability that the estimation mode of a macroblock is similar to the top one mode sorted from the neighboring macroblocks centered at the same position on the previous frame.

Sequences	QP28	QP32	QP36	QP40
Akiyo	90.10%	93.78%	96.76%	98.57%
Container	89.18%	93.38%	96.46%	98.49%
Hall_Monitor	91.58%	92.65%	93.74%	95.60%
Moth&Daug	78.31%	87.71%	93.86%	97.30%
News	81.11%	85.39%	90.04%	93.76%
Salesman	85.52%	87.99%	92.64%	96.69%
Carphone	64.70%	76.14%	85.99%	92.74%
Coastgrd	53.23%	65.31%	78.79%	89.42%
Foreman	54.29%	63.22%	74.78%	84.00%

Table 3. Probability that the estimation mode of a macroblock is similar to the top two modes sorted from the neighboring macroblocks centered at the same position on the previous frame.

Sequences	QP28	QP32	QP36	QP40
Akiyo	93.48%	95.74%	97.76%	98.94%
Container	93.08%	95.37%	97.30%	98.70%
Hall_Monitor	96.01%	96.48%	96.41%	97.22%
Moth&Daug	86.42%	92.06%	95.95%	97.85%
News	87.86%	90.41%	93.16%	95.51%
Salesman	91.85%	92.35%	95.19%	97.80%
Carphone	78.22%	84.90%	90.51%	94.81%
Coastgrd	73.61%	79.11%	86.36%	93.32%
Foreman	72.11%	77.30%	84.31%	89.66%

在 Table1,2 和 3 中觀察，我們發現到 macroblock 的運動估計模式與前一張參考 frame 的鄰近 macroblock 位置具有很高的關聯性。根據以上的模式相關性分析，我們提出一個新的快速模式決定的方法去改善在 H.264/AVC 編碼系統中模式決定處理的計算量。

4.2.2 使用空間-時間模式關聯性的快速 Intra/Inter 模式決定

對於每一個 GOP，我們使用 JM 參考軟體中之全域搜尋方法來決定第一個 P-frame 中每一個 macroblock 所使用之運動向量預估模式。對於後面之 P-frame，我們會利用前畫面已知運動向量預估模式預測每一個 macroblock 所使用之運動向量預估模式。快速模式決定方法的演算法在下列作描述。

步驟 1. macroblock 追蹤

對每一個 macroblack 找出運動向量預測模式之統計直方圖，為了精準求出每一個 macroblack 之統計直方圖，我們需要找到這個 MB 在前一個 frame 的所在位置。相鄰兩畫面同一個 MB 的位移關係可由前一個 frame MB 所在位置鄰近方塊之權重運動向量計算之，如方程式(2)所示。Fig. 8 說明了此追蹤程序。

$$MV(x,y) = \Big(\sum_{i=x-1}^{i=x+1}\sum_{j=y-1}^{j=y+1}\mathbf{m}(i,j)\Big)/T \qquad (2)$$

這裡 *MV* 是指預測運動向量，x 和 y 是指目前 frame 的 block 的座標，m(i, j)是在前一張 frame 的 block(i, j)的運動向量。Fig. 8 說明估計運動向量的方法。運動向量的精確估計可能被計算利用權重的每一個 block 根據面積比率的 block 到 4×4block。舉個例，Table 4 說明顯示在 Fig. 8 中每一個 macroblock 內之 block 相對於 4×4 block 的面積比率。在 Eq. (2)至中數值 *T* 是觀察到的所有權重 factors 的總和。然而，當 MV 的數值不是整數時，我們選擇最接近的 macroblock 如 Fig. 9 所示，去計算模式的統計直方圖。

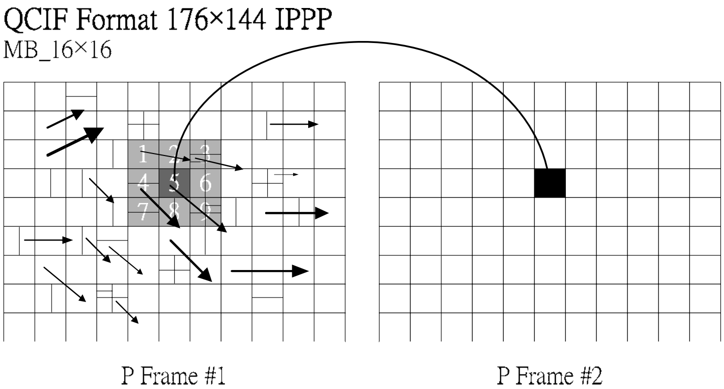


Fig. 8 Distribution of motion vectors for a macroblock on P-frame #2.

Table 4 The number of the available motion vectors in Fig. 8

No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8	No.9	T
16	0	4	8	16	0	0	0	0	44

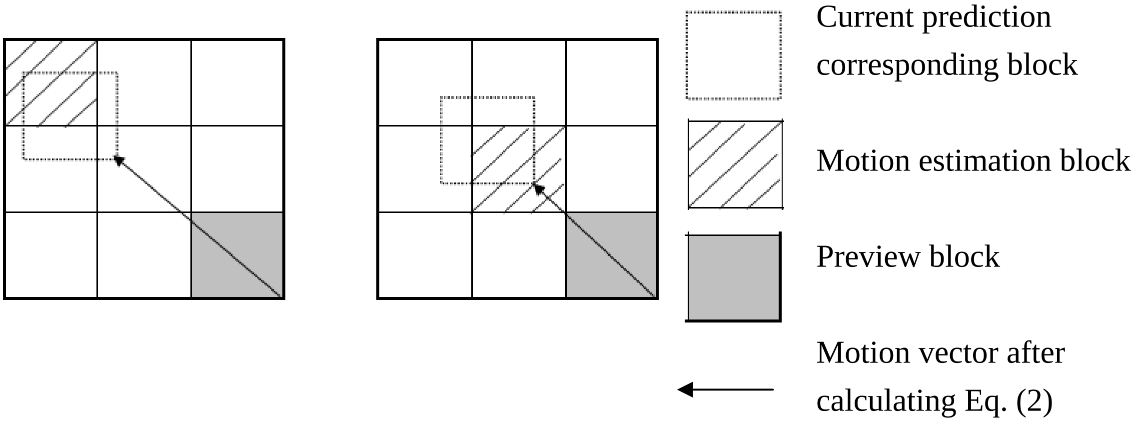


Fig. 9 The prediction block from previous block.

步驟 2 模式統計直方圖的計算

一但畫面中的每一個macroblock追蹤後，我們可以發現目前畫面(current frame)中追蹤過的macroblock的位置如 Fig.10。每一個追蹤過的macroblock的模式統計直方圖，可以靠目前畫面(current frame)追蹤過的macroblock位置為中心的鄰近macroblock的模式數量獲得。而為了降低運算量，Fig. 1中所有的模式依照block的大小歸類成五個種類(列於Table 5)，省略計算運算模式統計直方圖，我們用整理種類統計直方圖來加以替換。根據整理出的種類統計直方圖，我們選擇在最高兩個種類的block形式，當作模式決策程序的候補。

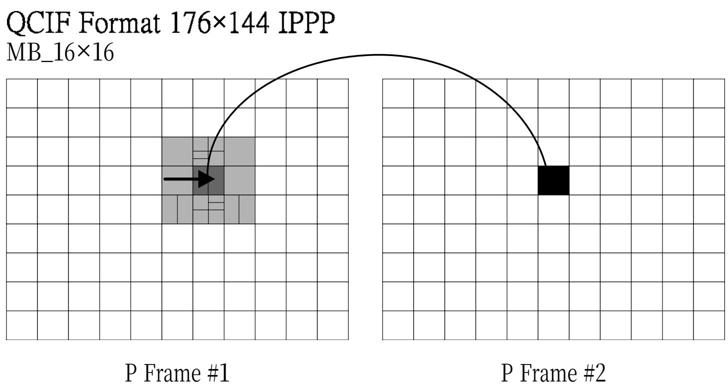


Fig. 10 The mode histogram for each tracked macroblock (P-frame #2) may be obtained by calculating the number of each mode among the neighboring macroblocks centered at the position of the tracked macroblock on previous frame (P-frame #1).

Table 5 Five categories for the block mode classification.

Mode Category	
1	SKIP / 16×16
2	16×8 / 8×16
3	8×8
4	8×4 / 4×8
5	4×4

步驟 3 漂移補償

為了預防在模式決策和運動向量預測程序時的產生的失真累積補償現象，當 macroblock B_{ij} 的 R-D cost 大於一開始定義的臨界值 T_{ij} 時，候選種類必須重新尋找。首先我們紀錄包含於 JM 模式決策中(使用 full search)，第一個 P-frame 中每一個 macroblock B_{ij} 的 R-D cost C_{ij}。接著每一個 macroblock 的臨界值 T_{ij} 設為αC_{ij}，對於連續 P-frames 每一個 macroblock 的 R-D cost 需跟 T_{ij} 作比較。假如 macroblock B_{ij} 的 R-D cost 大於 T_{ij}，模式決策程序將由以下規則重新尋找：：

1. 用較小的 block 組成當成新的候選模式類別置換較大的 block 組成模式類別。但是新的被選擇的模式類別不應該和已被挑選的模式類別一樣。
2. 如果候選模式類別不能更新成較小的 block size，換言之，候選模式類別被類別#4 和#5 組成，那麼候選模式類別要停止更新。
3. 一旦決定了每一個 macroblock 的運動估計模式，分割系統和相對應的運動 向量會被紀錄。

舉例來說，如果在 Fig.11 中的 P-frame#2 的 macroblock#2 的 R-D cost 大於 Fig.11 中的 P-frame#1 的 macroblock#1 的 R-D cost，那麼 P-frame#3 的 macroblock#3 的候選模式將會被決定藉由以上的規則。

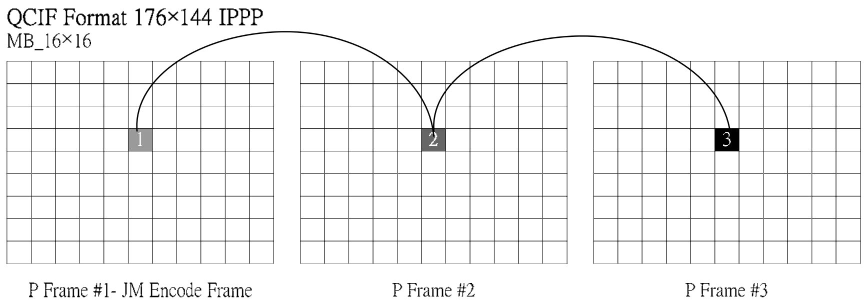


Fig. 11 The drift compensation process.

步驟 4 快速 Intra 預測

在 H.264/AVC 中，intra 預測模式決定是很複雜且非常耗時的。在此計劃有別於文獻[14]，我們除了考慮 interframe coding 也同時考慮 intraframe frame coding. 也就是較完整的研究如何改善 H.264/AVC 的編碼效能。在這裡，每一個 GOP 內，JM 完全模式搜尋被應用在第一個 P-frame 作全模式之搜尋，並且對於每一個 macroblock 的 intra 編碼所決定的模式要被紀錄起來。在後繼的 P-frame 中被追蹤到的 macroblock 所可能採取的 Intra 預測編碼模式，即可以參考前一 P-frame 所記錄的 Intra 預測編碼模式。本計劃所提出的模式決定方法可應用於 intra 16×16 預測模式和 intra 4×4 預測模式兩者。在 JM 完全模式搜尋方法，模式估計的數量大約是 592 = 4×(16×9+4)。然而，利用空間-時間預測系統，模式決定的數量可以減少成 17=1×(16+1)。

步驟 5 模式紀錄

一旦在每一個 macroblock 所決定之 Inter/Intra 預測編碼模式，macroblock 分割圖，和相對應的運動向量都需要記錄起來。

當我們想要在 P-frame 中編碼 macroblock，將會需要做兩個重要的步驟(intra 預測和 inter 預測)。之後每一個 R-D cost 的計算，我們將得到最好的預測模式。intra / intera 模式決定方法利用空間-時間關聯性的 block diagram 說明在 Fig. 12。它有兩個部份去作這個演算法，一個是快速 inter 預測模式和另一個是 intra 預測模式，兩者都被利用於降低預測

模式計算複雜度，所以效率可以被大幅提昇。

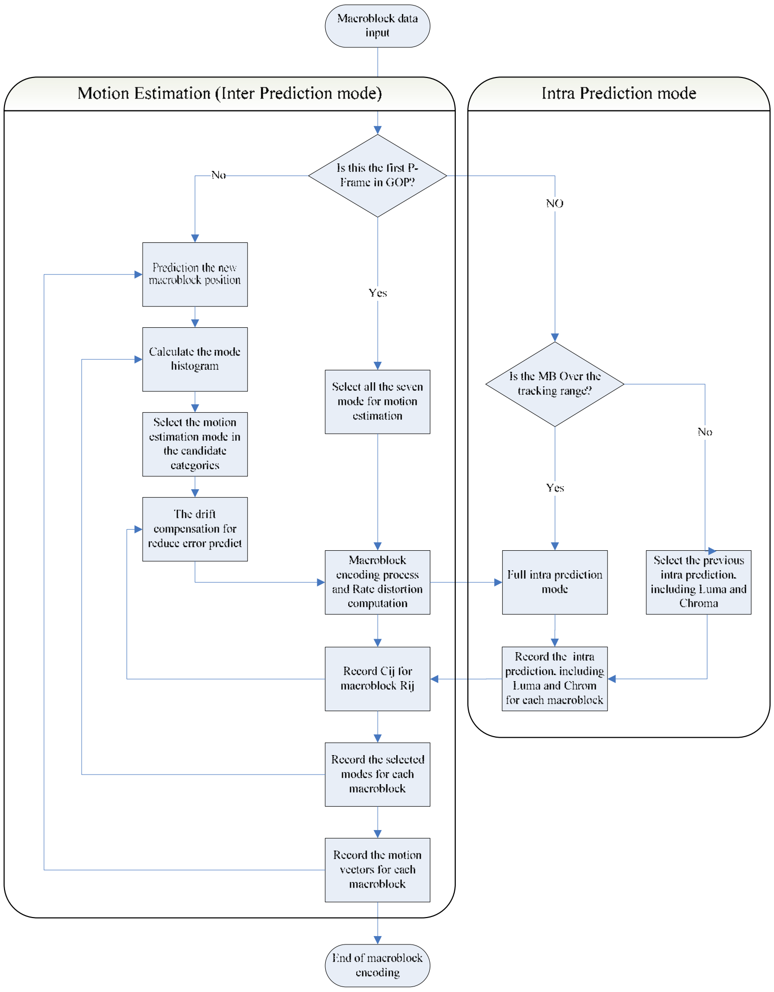


Fig. 12 The block diagram of fast Intra/Inter mode decision using spatial-temporal mode correlation.

五、結果與討論

5.1 針對固定 QP 參數的效率和 PSNR 的分析

在這個章節中。說明針對固定 QP 參數的效率和 PSNR 的分析。在 H.264/AVC 中 QP 參數被固定為 28, 32,36,40 在這方面對以上作分析。利用 E q. (3)分析計算率和利用 E q. (4)分析傳送速率。效率分析被執行經由計算保留的時間從我們提出的方法到 JM 全域搜尋的方法而且說明在 Table 6。明顯的利用空間-時間關聯性的模式決定的所有解法可以降低計算時間大約 70%對 9 段 video sequences。PSNR 的分析說明在 Table 7。快速模式決定方法的平均 PSNR 降低大約 0.08 dB。bit-rate 分析說明在 Table 7。他顯示我們提出的方法至增加大約 3.7%。總共的編碼 cost 降低計算時間大約 70%。一般而言，效率改善的程度對於 high motion 的 video sequences (Carphone, Coastgrd, Foreman)會比 low motion (Akiyo, Container, Hall_Monitor, Moth&Daug)來的小。

$$\Delta T = \frac{Time[JM] - Time[proposed]}{Time[JM]} x100\% \quad (3)$$

$$\Delta R = \frac{BitRate[proposed] - BitRate[JM]}{BitRate[proposed]} x100\% \quad (4)$$

Table 6 The motion estimation speed up analysis of proposed and Lien's method.

Moiton Estimation	Video Clip	Type	28	32	36	40
A	akiyo	Proposed method	73.0%	73.7%	72.5%	71.2%
		Lien's method [14]	74.5%	73.1%	72.0%	71.2%
	container	Proposed method	73.7%	73.8%	73.6%	74.7%
		Lien's method [14]	73.1%	72.7%	73.7%	74.7%
	hall monitor	Proposed method	72.3%	72.4%	70.4%	71.4%
		Lien's method [14]	74.9%	73.4%	72.6%	71.1%
	Moth&Daug	Proposed method	71.1%	71.4%	71.1%	69.8%
		Lien's method [14]	71.3%	71.1%	71.0%	68.3%
B	news	Proposed method	72.5%	71.0%	71.8%	72.4%
		Lien's method [14]	70.3%	68.7%	70.6%	71.7%
	salesman	Proposed method	72.2%	72.2%	71.1%	70.6%
		Lien's method [14]	72.5%	71.2%	70.2%	71.0%
C	carphone	Proposed method	71.7%	70.4%	70.7%	70.5%
		Lien's method [14]	70.9%	70.6%	70.1%	70.7%
	coastgrd	Proposed method	71.5%	71.9%	72.2%	71.2%
		Lien's method [14]	70.8%	71.2%	71.6%	72.7%
	foreman	Proposed method	70.8%	70.1%	70.1%	70.1%
		Lien's method [14]	69.5%	70.3%	71.0%	69.8%

Table 7 The PSNR (dB) analysis of proposed and Lien's method.

Y PSNR	Video Clip	Type	28	32	36	40
A	akiyo	Proposed method	-0.07	-0.12	-0.1	-0.07
		Lien's method [14]	-0.07	-0.12	-0.1	-0.07
	container	Proposed method	-0.08	-0.11	-0.11	-0.09
		Lien's method [14]	-0.08	-0.11	-0.1	-0.09
	hall monitor	Proposed method	-0.07	-0.09	-0.1	-0.08
		Lien's method [14]	-0.05	-0.09	-0.09	-0.08

	Moth&Daug	Proposed method	-0.06	-0.06	-0.05	-0.02
		Lien’s method [14]	-0.07	-0.06	-0.04	-0.01
B	news	Proposed method	-0.09	-0.1	-0.11	-0.09
		Lien’s method [14]	-0.08	-0.1	-0.1	-0.09
	salesman	Proposed method	-0.07	-0.08	-0.06	-0.04
		Lien’s method [14]	-0.06	-0.07	-0.06	-0.04
C	carphone	Proposed method	-0.1	-0.14	-0.11	-0.13
		Lien’s method [14]	-0.1	-0.12	-0.09	-0.11
	coastgrd	Proposed method	-0.02	-0.02	-0.05	-0.06
		Lien’s method [14]	-0.02	-0.02	-0.05	-0.06
	foreman	Proposed method	-0.11	-0.15	-0.14	-0.1
		Lien’s method [14]	-0.1	-0.14	-0.12	-0.11

Table 8 The bit-rate analysis of proposed and Lien’s method.

Total BitRate	Video Clip	Type	28	32	36	40
A	akiyo	Proposed method	3.7%	0.8%	0.0%	-0.2%
		Lien’s method [14]	3.8%	0.8%	0.0%	-0.2%
	container	Proposed method	3.8%	0.9%	-0.3%	-0.9%
		Lien’s method [14]	3.5%	0.9%	-0.1%	-0.9%
	hall monitor	Proposed method	6.9%	6.3%	4.5%	3.0%
		Lien’s method [14]	6.9%	6.2%	4.5%	2.7%
	Moth&Daug	Proposed method	5.1%	3.4%	1.7%	2.0%
		Lien’s method [14]	4.8%	2.9%	1.7%	2.0%
B	news	Proposed method	6.8%	5.5%	3.9%	2.3%
		Lien’s method [14]	6.2%	4.7%	3.0%	1.8%
	salesman	Proposed method	6.5%	3.6%	1.4%	0.1%
		Lien’s method [14]	6.4%	3.4%	1.4%	0.3%
C	carphone	Proposed method	6.9%	4.1%	2.0%	1.1%
		Lien’s method [14]	6.5%	4.0%	1.9%	1.0%
	coastgrd	Proposed method	7.7%	7.5%	4.8%	2.5%
		Lien’s method [14]	7.7%	7.6%	4.7%	2.6%
	foreman	Proposed method	10.1%	7.3%	5.1%	3.5%
		Lien’s method [14]	9.7%	6.8%	4.6%	2.7%

Table 9 The total encoding cost speed up analysis of proposed and Lien’s method.

Total encoding	Video Clip	Type	28	32	36	40
A	akiyo	Proposed method	74.3%	73.8%	73.4%	72.9%
		Lien's method [14]	40.7%	42.1%	43.7%	44.3%
	container	Proposed method	72.0%	72.3%	73.1%	73.5%
		Lien's method [14]	41.8%	42.9%	44.9%	46.4%
	hall monitor	Proposed method	73.6%	73.1%	72.4%	72.2%
		Lien's method [14]	39.6%	40.4%	42.4%	43.5%
	Moth&Daug	Proposed method	70.5%	70.5%	70.5%	70.3%
		Lien's method [14]	44.7%	45.7%	45.7%	44.3%
B	news	Proposed method	72.0%	71.9%	71.8%	71.6%
		Lien's method [14]	41.1%	42.5%	44.2%	45.7%
	salesman	Proposed method	73.1%	72.6%	72.5%	71.9%
		Lien's method [14]	41.0%	44.0%	46.3%	46.8%
C	carphone	Proposed method	67.6%	68.0%	68.3%	68.4%
		Lien's method [14]	45.7%	46.7%	47.1%	46.5%
	coastgrd	Proposed method	66.8%	67.3%	68.2%	68.8%
		Lien's method [14]	51.4%	52.1%	52.7%	52.7%
	foreman	Proposed method	66.5%	66.7%	66.8%	66.8%
		Lien's method [14]	48.4%	49.1%	50.1%	50.0%

5.2 The Rate-Distortion Analyses

在章節 4.2，rate-distortion 的分析針對 bit-rate 100K 到 2500K bit/sec 作說明。rate-distortion 的分析被執行在下列兩個環境設定：(1)所有的運動估計模式在 JM-14.0 中被利用，(2)利用 Len 的方法。從 Fig. 13 到 Fig. 15，模擬結果顯示 PSNR 數值接近從 JM-14.0 參考軟體獲得的最佳數值。

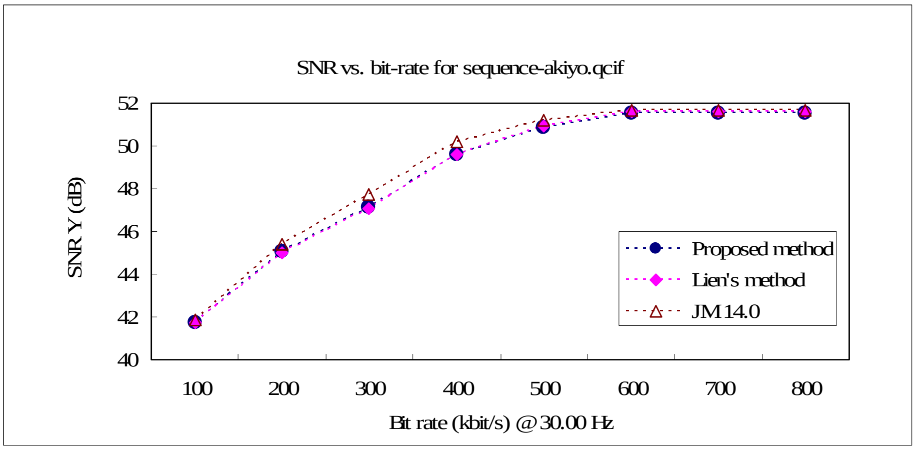


Fig. 13 Rate-distortion curves for Akiyo video sequence obtained by JM-14.0, our proposed method, and Lien's method.

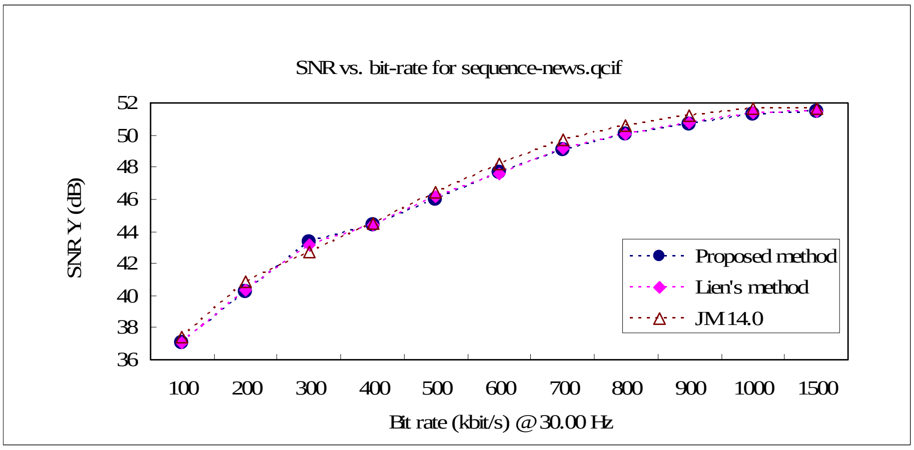


Fig. 14 Rate-distortion curves for Foreman video sequence obtained by JM-14.0, our proposed method, and Lien's method.

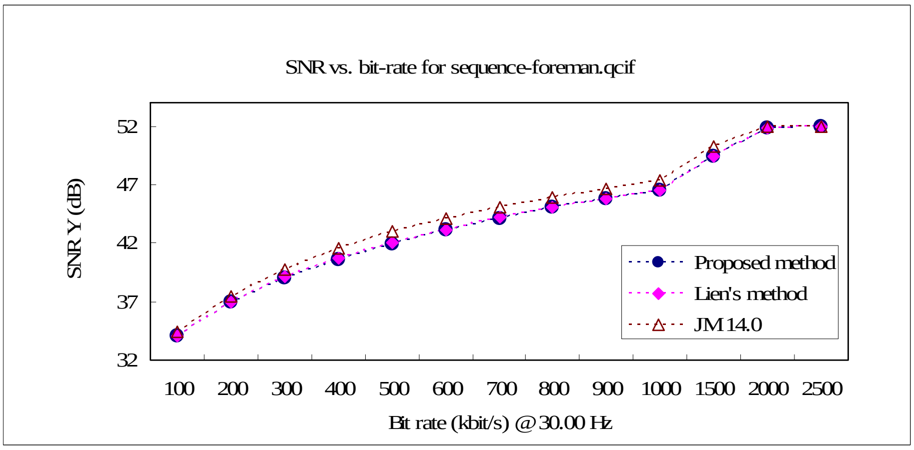
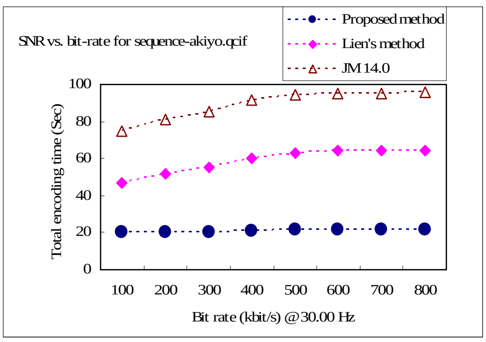
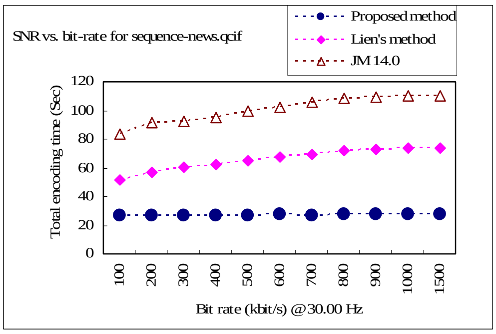


Fig. 15 Rate-distortion curves for News video sequence obtained by JM-14.0, our proposed method, and Lien's method.

(b)



(a)



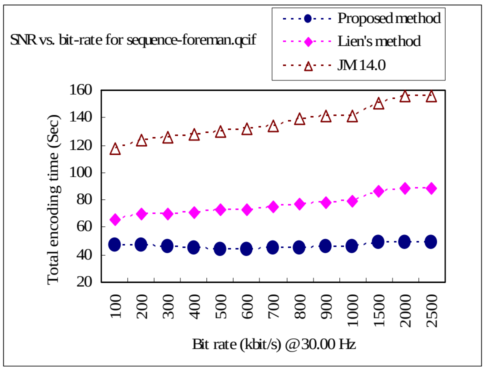
(c) 

Fig. 16 The computation time of the total encoding cost for (a)Akiyo, (b)News and (c)Foreman video sequences using the JM-14.0 reference software, our proposed method, and Lien's method.

我們可以從 Fig.16 看見，我們提出的所以模式覺得的解法對於壓縮效率可以接近大約 3 倍的改善。在最好 bit rate 上我們提出的方法還比”Lien 的方法”快。最後，我們可以發現空間-時間關聯性不只是對運動估計有效也適合用於 P frame 的 intra 預測。運動估計和 intra 預測都有計算時間上的速度提昇。所有的 bit rate 也沒有增加太多，而且 PSNR 也大多相同於” Lien 的方法 ”。

六、 參考文獻

[1] “Draft ITU-T recommendation and final draft international standard of joint video specification (ITU-T Rec. H.264/ISO/IEC 14486-10 AVC,” in Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, JVT-G050, March, 2003.

[2] Iain E. G. Richardson, *H.264 and MPEG-4 video compression*, John Wiley & Sons, 2003.

[3] ITU-T Recommendation H.263, “Video coding for low bit rate communication,” version 3, Nov. 2000.

[4] ISO/IEC 14496-2: “Information technology-coding of audiovisual objects-part 2: visual,” Geneva, 2000.

[5] J. Ostermann, J. Bormans, P. List, D. Marpe, M. Narroschke, F. Pereira, T. Stockhammer, T. Wedi, “Video coding with H.264/AVC: tools, performance, and complexity,” *IEEE Circuits and Systems Magazine*, vol. 4, no. 1, pp. 7 - 28, March, 2004.

[6] Joint Video Team (JVT) Reference software JM-14.0, May, 2008, http://iphome.hhi.de/suehring/tml/

[7] X. Jing and L. P. Chau, “An efficient inter mode decision approach for H.264 video coding,” *IEEE International Conference on Multimedia & Expo, ICME2004*, Taipei, Taiwan, June 2004.

[8] D. Wu, S. Wu, K. P. Lim, Pan Feng, Z. G. Li, C. C. Ko, “Fast inter mode decision with adaptive thresholds for H.264 encoding,” Accepted by 8th International Symposium on Consumer Electronics (ISCE'2004), Reading, UK, 1-3 Sept. 2004.

[9] 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 (JVT_G013) MPEG2003, Pattaya, Thailand, March, 2003

[10] D. Zhu, Q. Dai and R. Ding, “Fast inter prediction mode decision for H.264,” *IEEE the fifth International Conference on Multimedia and Exposition* (ICME 2004), Taipei, June

28, 2004.
[11] B. Meng, O.C. Au, "Fast intra-prediction mode selection for 4A blocks in H.264," *IEEE International Conference on Acoustics, Speech, and Signal Processing* (ICASSP 2003), vol. 3, pp. 389-92, April, 2003.
[12] C.-C. Cheng, T.-S. Chang, "Fast three step intra prediction algorithm for 4/spl times/4 blocks in H.264," *IEEE International Symposium on Circuits and Systems* (ISCAS 2005), vol. 2, pp. 1509 - 1512, May, 2005.
[13] M.-C. Hwang, J.-K. Cho, J.-S. Kim, J.-H. Kim, S.-J. Ko, "Fast Intra Prediction Mode Selection Scheme Using Temporal Correlation in H.264," *IEEE International Region 10 Conference*, pp. 1-5, Nov., 2005.
[14] C.-C. Lien, C.-P. Yu, "A Fast Mode Decision Method for H.264/AVC Using the Spatial-Temporal Prediction Scheme," *International Conference for Pattern Recognition* (ICPR 2006), vol. 4, pp. 334 – 337, 2006
[15] P. Yin, A. Vetro, H. Sun, B. Liu, "Drift compensation for reduced spatial resolution transcoding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 11, pp. 1009-1020, Nov. 2002.
[16] G. J. Sullivan, P. Topiwala, and A. Luthra, "The H.264/AVC advanced video coding standard: Overview and introduction to the fidelity range extensions," *SPIE Conference on Applications of Digital Image Processing*, Aug. 2004.
[17] H.264/AVC Part 10 Tutorials, "H.264 variable length coding," Oct. 2002, http://www.vcodex.com/h264.html
[18] G. Bjontegaard and K. Lillevold, "Context-adaptive VLC coding of coefficients," Fairfax, VA, JVT-C028, May, 2002.

七、　附錄

部分成果已發表在

Cheng-Chang Lien, Chih-Chang Chen and Yang-Kai Chang, “Fast Intra/Inter Mode Decision for H.264/AVC Using the Spatial-Temporal Prediction Scheme,” IEEE Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing in conjunction with the UIC’09 and ATC’09 conferences, Brisbane Australia, July 7- 9, 2009, pp. 194-199. **(EI)**

Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing

Fast Intra/Inter Mode Decision for H.264/AVC Using the Spatial-Temporal Prediction Scheme

Cheng-Chang Lien, Chih-Chang Chen and Yang-Kai Chang
Department of Computer Science and Information Engineering,
Chung Hua University, Hsinchu, Taiwan, R.O.C.
E-mail: cclien@chu.edu.tw

Abstract

In the H.264/AVC coding standard, there are seven motion estimation modes from 4×4 to 16×16 are used to find the minimum motion compensation error for each macroblock in the inter frame coding. For intra prediction coding, four prediction modes are used to encode the 8×8 chroma and 16×16 luma blocks and eight prediction directions and one DC prediction are used to encode the intra 4×4 luma block. In this study, the spatial-temporal correlations between the current frame and the reference frame are analyzed to develop a fast mode decision method for inter/intra frame encoding in which no extra image processes are used. The experimental results show that the total computation cost can be reduced above 70%, the total bit rate just increase less than 3.7 % and the average PSNR is only dropped about 0.08 dB.

Keywords: H.264, JM-14.0, mode decision, intra prediction, spatial-temporal correlation, R-D Cost.

1. Introduction

Recently, the new video coding standard H.264/AVC [1] is proposed by the Joint Video Team (JVT) to develop the new low bit-rate video compression technology. In the H.264/AVC coding standard, there are seven motion estimation modes from 4×4 to 16×16 are used to find the minimum motion compensated error for each macroblock. In the current JVT reference software [2], seven modes (the various kinds of block sizes) are applied to perform the motion compensation process such that the R-D cost defined in Eq. (1) is optimized.

$$J_{Mode} = D + \lambda_{Mode} \times R, \qquad (1)$$

where D denotes the motion compensated error of a macroblock, λ_{Mode} is the Lagrange multiplier and R represents the demanded bit-rate. In order to obtain the optimum motion estimation mode, we must calculate the R-D cost for each motion estimation mode in which some time-consuming processes, e.g., the motion estimation, DCT transformation and quantization are involved. In addition, seven motion estimation modes from 4×4 to 16×16 (see Fig. 1) are used to determine the optimum motion estimation mode for each macroblock. For intra prediction coding four prediction modes shown in Fig. 2 are used to encode the 8×8 chroma and 16×16 luma blocks, eight prediction directions and one DC prediction shown in Fig. 3 are used to encode the intra 4×4 luma block. Hence, the high computation cost for the full search method used in the reference software JM-14.0 [2] make the encoding process inefficient.

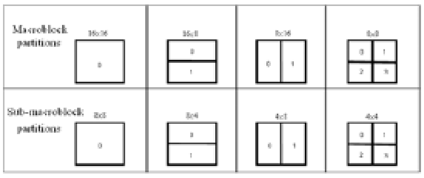


Fig. 1. Macroblock and Sub-macroblock partitions in H.264.

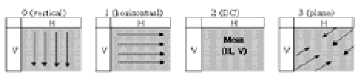


Fig. 2. Prediction modes in intra 16×16 luma and intra 8×8 choma blocks.

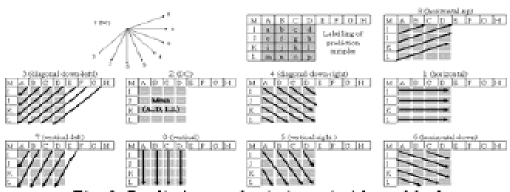


Fig. 3. Prediction modes in intra 4×4 luma block.

978-0-7695-3737-5/09 $25.00 © 2009 IEEE
DOI 10.1109/UIC-ATC.2009.71

Recently, many researches [3-10] addressed on the fast mode decision methods are proposed. In [3], the SAD (sum of absolute difference) between the current frame and previous frame for each macroblock is calculated to evaluate which kind of block size is appropriate for the motion estimation process. In [4], the edge detection is applied to classify the homogeneous and non-homogeneous regions and then the rate-distortion optimization is used to determine the motion estimation mode. In [5], Pan et al. applied the edge direction histogram to reduce the computation cost of the decision of intra-prediction modes. In [6], Zhu et al. applied the low-resolution image and edge detection to determine the motion estimation modes (8×4, 4×8 and 4×4). However, the efficiency of the motion compensation process will be reduced by the extra image processes for determining the appropriate motion estimation mode. In [7], Meng et al. applied two methods to reduce the mode decision complexity. First, macroblock decimation is used to estimate the mode with minimum rough R-D cost. Second, if the R-D cost is smaller than a predefined threshold, we can choose this mode and its two neighboring modes as the candidate modes.

For intra coding in H.264/AVC, Cheng et al. [8] applied three step hierarchal searching method to predict 4×4 intra prediction modes. In each step, the minimum RD-cost is acquired to decide the next one to compute and it can reduce the computation cost from 9 modes to 6 modes. In [9], Hwang et al. applied the temporal information to predict the intra mode on P frames. For each block the previous prediction modes at the same position and its neighboring blocks are regarded as the candidate prediction modes to reduce the mode searching significantly.

In this paper, the spatial-temporal correlations among the reference frames and neighboring macroblocks are not only used to develop a fast mode decision (motion estimation) method, but also used in the intra prediction on P frames. The proposed method doesn't need extra image processes and can save more computation cost than the previous work [10]. By comparing with JM-14.0 reference software, the experimental results show that the total computation cost may be reduced above 70%, the total bit rate just increase than 3.7 % and the average PSNR is only dropped about 0.08 dB.

2. The fast mode decision method

The full searching process (16×16, 16×8,⋯, 4×4) for determining the estimation mode in the reference software JM-14.0 will make the encoding process inefficient. In this section, the spatial-temporal mode correlations among the spatial and temporal macroblocks are analyzed to develop the fast mode decision method.

2.1 The spatial-temporal mode correlation

By the careful observation of the mode decision process in JM-14.0 reference software, the motion estimation mode of a macroblock is highly correlated with the modes of the macroblocks neighboring to the same position on the previous reference frames (multiple reference frames). The mode correlation is described in Fig. 4.

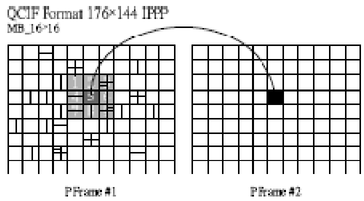


Fig. 4. The mode correlation for a macroblock between adjacent frames.

The probability that the mode of a macroblock on P-frame #2 is similar to the modes of the neighboring macroblocks centered at the same position on P-frame #1 (The number 1~9 of P-frame #1 in Fig. 7) is high. Based on the JM reference software, the mode correlation is analyzed and listed in Table 1. Here, the estimation modes including 16×16, 16×8, 8×16, 8×8 and skip mode are used to analyze the mode correlation.

Table 1. The probabilities for the mode of a macroblock on P-frame #2 is similar to the modes of the neighboring macroblocks centered at the same position on P-frame #1.

Sequences	QP28	QP32	QP36	QP40
Akiyo	95.11%	96.02%	97.22%	98.25%
Container	93.35%	94.88%	96.40%	97.55%
Hall_Monitor	95.28%	96.62%	97.16%	97.51%
Moth&Daug	91.77%	93.59%	95.11%	96.42%
News	92.90%	93.63%	94.52%	95.73%
Salesman	94.63%	95.10%	96.00%	97.31%
Carphone	88.54%	90.16%	92.32%	94.40%
Coastgrd	86.05%	87.31%	89.17%	92.70%
Foreman	87.28%	88.29%	89.22%	90.90%

In Table 1, nine video sequences with length of 200 frames are used to analyze the mode correlation. Furthermore, the quantization step size is set as QP=28, 32, 36 and 40 respectively. For the Foreman video sequence with QP=28, the possibility that the estimation mode of a macroblock is closed to the estimation modes of the neighboring macroblocks centered at the same position on the previous frame is

87.28%. From the analysis results in Table 1, it was observed that the mode correlation for a macroblock between adjacent frames is high. Furthermore, two additional experiments for the mode correlation analysis are given in Tables 2 and 3. In Table 2 and 3 the skip and 16×16 modes are regarded as the same class. Table 2 and 3 illustrate the probabilities that the estimation mode of a macroblock is similar to the top one and two estimation modes among the neighboring macroblocks centered at the same position on the previous frame respectively.

Table 2. The probability that the estimation mode of a macroblock is similar to the top one mode sorted from the neighboring macroblocks centered at the same position on the previous frame.

Sequences	QP28	QP32	QP36	QP40
Akiyo	90.10%	93.78%	96.76%	98.57%
Container	89.18%	93.38%	96.46%	98.49%
Hall_Monitor	91.58%	92.65%	93.74%	95.60%
Moth&Daug	78.31%	87.71%	93.86%	97.30%
News	81.11%	85.39%	90.04%	93.76%
Salesman	85.52%	87.99%	92.64%	96.69%
Carphone	64.70%	76.14%	85.99%	92.74%
Coastgrd	53.23%	65.31%	78.79%	89.42%
Foreman	54.29%	63.22%	74.78%	84.00%

Table 3. The probability that the estimation mode of a macroblock is similar to the top two modes sorted from the neighboring macroblocks centered at the same position on the previous frame.

Sequences	QP28	QP32	QP36	QP40
Akiyo	93.48%	95.74%	97.76%	98.94%
Container	93.08%	95.37%	97.30%	98.70%
Hall_Monitor	96.01%	96.48%	96.41%	97.22%
Moth&Daug	86.42%	92.06%	95.95%	97.85%
News	87.86%	90.41%	93.16%	95.51%
Salesman	91.85%	92.35%	95.19%	97.80%
Carphone	78.22%	84.90%	90.51%	94.81%
Coastgrd	73.61%	79.11%	86.36%	93.32%
Foreman	72.11%	77.30%	84.31%	89.66%

From the observation in Tables 1, 2 and 3, we found that the motion estimation mode of a macroblock is highly correlated with the motion estimation modes of the macroblocks neighboring to the same position on the previous reference frame. According to the above mode correlation analysis, we propose a new method of fast mode decision to improve the computation cost of the mode decision process in the H.264/AVC encoding system.

2.2 Fast intra/inter mode decision using spatial-temporal mode correlation

For each GOP, the mode decision process for first P-frame is determined by using the full search method in JM reference software and the determined estimation modes for each macroblock are used to predict the modes for next P-frame. The algorithm of the fast mode decision method is described as follows.

2.2.1 Tracking of the macroblock

To find the mode histogram for each macroblock, the corresponding position on previous frame for each macroblock should be tracked. The position of each macroblock on the current P-frame is tracked with the weighted motion vectors of the macroblocks neighboring to the same position on the previous reference frame shown in Fig. 5. The weighted motion vector is calculated as

$$MV(x,y) = \left(\sum_{i=x-1}^{i=x+1}\sum_{j=y-1}^{j=y+1} m(i,j)\right)/T, \qquad (2)$$

where MV is the predicted motion vector, x and y denote the block coordinates of current frame and $m(i,j)$ is the motion vector of block (i,j) on previous frame. Fig. 5 illustrates the method of estimating the motion vector. The precise prediction of the motion vector may be calculated by weighting each block according to the area proportion of the block to 4×4 block. For example, Table 4 illustrates the area proportion of each block to 4×4 block for the macroblock shown in Fig. 5. The value of T in Eq. (2) is obtained by summing all the weighting factors. However, when the value of the MV is not integer, we select the nearest macroblock shown in Fig. 6 to calculate the mode histogram.

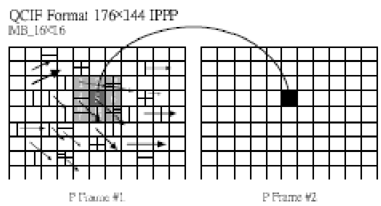


Fig. 5. Distribution of motion vectors for a macroblock on P-frame #2.

Table 4. The number of the available motion vectors in Fig. 8.

No. 1	No. 2	No. 3	No. 4	No. 5	No. 6	No. 7	No. 8	No. 9	Total
16	0	4	8	16	0	0	0	0	44

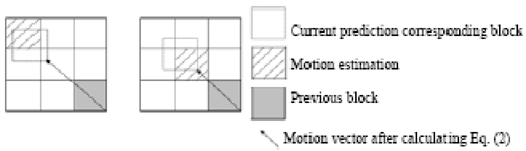


Fig. 6. The prediction block from previous block.

2.2.2 Calculation of the mode histogram

Once each macroblock on current frame is tracked, we may find the position of the tracked macroblock on previous frame shown in Fig. 7. Then, the mode histogram for each tracked macroblock may be obtained by calculating the number of each mode among the neighboring macroblocks centered at the position of the tracked macroblock on previous frame. However, to reduce the computation, all the modes shown in Fig. 1 are classified into five categories listed in Table 5 according to their block size. Then, instead of calculating the mode histogram, the category histogram is calculated and sorted. Based on the sorted category histogram we select the block modes in the top two categories as the candidate modes for the mode decision process.

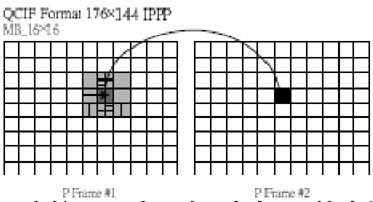


Fig. 7. The mode histogram for each tracked macroblock (P-frame #2) may be obtained by calculating the number of each mode among the neighboring macroblocks centered at the position of the tracked macroblock on previous frame (P-frame #1).

Table 5. Five categories for the block mode classification.

	Mode Category
1	SKIP / 16×16
2	16×8 / 8×16
3	8×8
4	8×4 / 4×8
5	4×4

2.2.3 Drift compensation

In order to prevent the drift phenomenon in the mode decision and motion estimation processes, the candidate categories need to be refined when the R-D cost for a macroblock is larger than predefined threshold. The mode decision process is refined as the following rules.

Firstly, to replace the mode category that composed of the largest block with the next one that composed of smaller blocks as the new candidate mode category. But, the new chosen mode category should not be the same as the one in the candidate mode categories. Secondly, if the candidate mode categories can't be updated with the one with smaller blocks, i.e., the candidate mode categories are composed of categories #4 and #5, then the candidate mode categories stop updating. Thirdly, once the motion estimation modes in each macroblock are determined, the partition scheme and corresponding motion vectors are recorded.

For example, if the R-D cost of macroblock #2 on the P-frame #2 is greater than the R-D cost of macroblock #1 on P-frame #1 in Fig. 8, then the candidate modes for the macroblock #3 on P-frame #3 will be determined by the above rules.

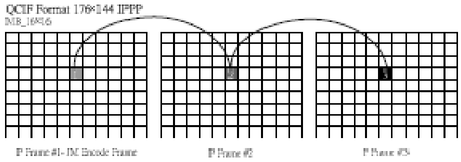


Fig. 8. The drift compensation process.

2.2.4 Fast intra prediction

In H.264/AVC, the mode decision in intra prediction is very complex and time consuming. Here, the JM full mode searching is applied to the first P frame for each GOP, and then the determined mode for intra coding of each macroblock is recorded. The tracked macroblock in the successive P-frame can refer to the determined mode of the corresponding macroblock in previous frame. The proposed mode decision method is applied to both the intra 16×16 prediction modes and intra 4×4 prediction modes. In the JM full searching method, the number of mode estimation is about 592=4×(16×9+4). However, using the spatial-temporal prediction scheme, the number of mode decision can be reduced to 17=1×(16+1).

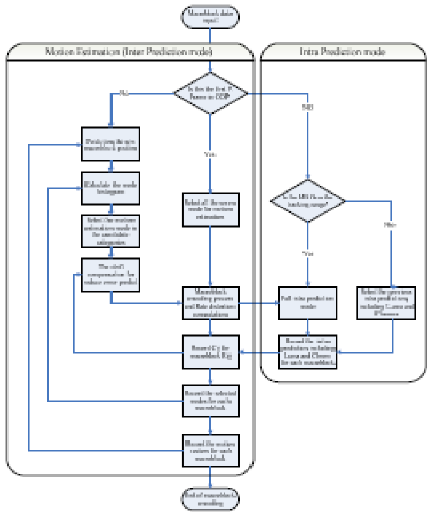


Fig. 9. The block diagram of fast intra/inter mode decision using spatial-temporal mode correlation.

2.2.5 Mode recording

Once the motion estimation modes within each macroblock are determined, the partition scheme, intra prediction modes, and corresponding motion vectors are recorded. After the calculation of each RD-cost, we will get the best prediction mode. The block diagram of intra/inter mode decision method using the spatial-temporal correlation is illustrated in Fig. 9. It consists of two parts: one is the fast inter prediction mode decision process and another is fast intra prediction mode decision process.

3. Experimental results

Here, all the efficiency and rate-distortion analyses are constructed on the basis of JM-14.0 reference software and the simulation environments are defined as follows;

(1) The length of video frames for the simulation is 200.
(2) The number of reference frames is 5.
(3) The search range for the motion estimation is 32 pixels.
(4) The Hadamard transform for encoding DC components is used.
(5) The rate-distortion optimization is applied.
(6) The length of GOP is 13.
(7) Nine video sequence of QCIF format are used as the testing video (Low motion including akiyo, container; Regular motion including Hall_Monitor, Mother and daughter, News and Salesman; High motion including Carphone, Coastgrd and Foreman).
(8) Disable the "FastCrIntraDecision".

Based on the above environment setting, the efficiency and rate-distortion analyses for the proposed system are illustrated.

3.1 The efficiency and PSNR analyses for fixed QP parameters

In this section, the efficiency and rate-distortion analyses for fixed QP parameters are illustrated. Here the QP parameters in H.264/AVC are fixed as 28, 32, 36 and 40 respectively for the above analyses. The computation efficiency is analyzed with Eq. (3) and the transmission rate is analyzed with Eq. (4). The efficiency analysis is performed by computing the saving time for the proposed method to the JM full searching method and illustrated in Table 6. It is obviously that the mode decision using the spatial-temporal correlation can reduce the computation time about 70% for the listed video sequences. The PSNR analysis is illustrated in Table 7. The average PSNR of the fast mode decision method is decreased about 0.08dB. The bit-rate analyses are illustrated in Table 8. It is shown that the proposed method is only increased about 3.7%. In Table 9, we show that the computation time can be reduced about 70% with the proposed method. In general, the degree of the efficiency improvement for the video sequences with high motion (Carphone, Coastgrd and Foreman) is less than the one with low motion (Akiyo, Container, Hall_Monitor and Moth&Daug).

$$\Delta T = \frac{Time[JM] - time[proposed]}{Time[JM]} \times 100\% \quad (3)$$

$$\Delta R = \frac{BitRate[proposed] - BitRate[JM]}{BitRate[proposed]} \times 100\% \quad (4)$$

Table 6. The motion estimation speed up analysis for the proposed method and the method in [10].

	Video Clip	Type	28	32	36	40
A	akiyo	Proposed method	73.0%	73.7%	72.5%	71.2%
		method in [10]	74.5%	73.1%	72.0%	71.2%
B	salesman	Proposed method	72.2%	72.2%	71.1%	70.6%
		method in [10]]	72.5%	71.2%	70.2%	71.0%
C	foreman	Proposed method	70.8%	70.1%	70.1%	70.1%
		method in [10]]	69.5%	70.3%	71.0%	69.8%

Table 7. The PSNR (dB) analysis for the proposed method and the method in [10].

	Video Clip	Type	28	32	36	40
A	akiyo	Proposed method	-0.07	-0.12	-0.1	-0.07
		method in [10]]	-0.07	-0.12	-0.1	-0.07
B	salesman	Proposed method	-0.07	-0.08	-0.06	-0.04
		method in [10]]	-0.06	-0.07	-0.06	-0.04
C	foreman	Proposed method	-0.11	-0.15	-0.14	-0.1
		method in [10]]	-0.1	-0.14	-0.12	-0.11

Table 8. The bit-rate analysis for the proposed method and the method in [10].

	Video Clip	Type	28	32	36	40
A	akiyo	Proposed method	3.7%	0.8%	0.0%	-0.2%
		method in [10]]	3.8%	0.8%	0.0%	-0.2%
B	salesman	Proposed method	6.5%	3.6%	1.4%	0.1%
		method in [10]]	6.4%	3.4%	1.4%	0.3%
C	foreman	Proposed method	10.1%	7.3%	5.1%	3.5%
		method in [10]]	9.7%	6.8%	4.6%	2.7%

Table 9. The speed up analysis for the proposed method and the method in [10].

	Video Clip	Type	28	32	36	40
A	akiyo	Proposed method	74.3%	73.8%	73.4%	72.9%
		method in [10]]	40.7%	42.1%	43.7%	44.3%
B	salesman	Proposed method	73.1%	72.6%	72.5%	71.9%
		method in [10]]	41.0%	44.0%	46.3%	46.8%
C	foreman	Proposed method	66.5%	66.7%	66.8%	66.8%
		method in [10]]	48.4%	49.1%	50.1%	50.0%

3.2 The rate-distortion analyses

In this section, the rate-distortion analyses from bit-rate 100K to 2,500K bits/sec are illustrated. From Fig. 10 to Fig. 12, the simulation results show that the PSNR value is closed to the optimal value obtained from the JM-14.0 reference software.

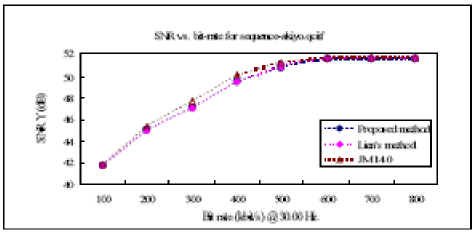


Fig. 10. Rate-distortion curves for Akiyo video sequence obtained from JM-14.0, our proposed method and method in [10].

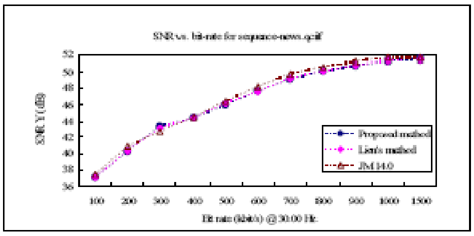


Fig. 11. Rate-distortion curves for Foreman video sequence obtained from JM-14.0, our proposed method and method in [10].

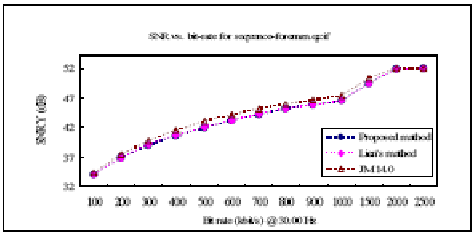


Fig. 12. Rate-distortion curves for News video sequence obtained from JM-14.0, our proposed method and method in [10]

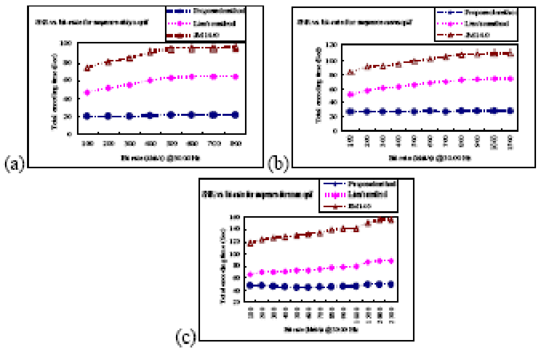


Fig. 16. The computation time of the inter/intra encoding process for (a)Akiyo, (b)News and (c)Foreman video sequences using the JM-14.0 reference software, our proposed method and method in [10].

In Fig. 16, the mode decision of our purposed can improve the compression efficiency significantly. Finally, we can find the spatial-temporal correlation not only improve the efficiency of the motion estimation process, but also reduce the computation complexity on the intra prediction process on the P frames.

4. Conclusion

In this paper, the spatial-temporal correlations between the current frame and the reference frame are considered to develop a fast mode decision method in which no extra image processes are used. Furthermore, the concept of drift compensation is adopted to avoid the error accumulation phenomenon during the mode decision process. The experimental results show that the total computation cost may be reduced about 70% and average PSNR is only dropped about 0.08 dB.

5. Reference

[1] "Draft ITU-T recommendation and final draft international standard of joint video specification (ITU-T Rec. H.264/ISO/IEC 14486-10 AVC)," *JVT-G050*, 2003.

[2] Joint Video Team (JVT) Reference software JM-14.0, 2008, http://iphome.hhi.de/suehring/tml/.

[3] X. Jing and L. P. Chau, "An efficient inter mode decision approach for H.264 video coding," *ICME2004*, 2004.

[4] D. Wu, S. Wu, K. P. Lim, Pan Feng, Z. G. Li and C. C. Ko, "Fast inter mode decision with adaptive thresholds for H.264 encoding," *ISCE2004*, 2004.

[5] 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 (JVT_G013) MPEG2003*, 2003.

[6] D. Zhu, Q. Dai and R. Ding, "Fast inter prediction mode decision for H.264," *ICME*, 2004, vol. 2, pp. 1123-1126.

[7] B. Meng and O. C. Au, "Fast intra-prediction mode selection for 4A blocks in H.264," *ICASSP 2003*, vol. 3, pp. 389-92, 2003.

[8] C. C. Cheng and T. S. Chang, "Fast three step intra prediction algorithm for 4/spl times/4 blocks in H.264," *ISCAS 2005*, vol. 2, pp. 1509-1512, 2005.

[9] M. C. Hwang, J. K. Cho, J. S. Kim, J. H. Kim and S. J. Ko, "Fast Intra Prediction Mode Selection Scheme Using Temporal Correlation in H.264," *IEEE International Region 10 Conference*, pp. 1-5, 2005.

[10] Cheng-Chang Lien and Chung-Ping Yu "A Fast Mode Decision Method for H.264/AVC Using the Spatial-Temporal Prediction Scheme," *Journal of Information Science and Engineering*, Vol. 23, No. 6, Nov. 2007, pp. 1911-1921.

國科會補助專題研究計畫項下出席國際學術會議心得報告

日期： 99 年 10 月 27 日

計畫編號	NSC 98-2221-E-216 -034		
計畫名稱	使用空間與時間關聯預測之 H.264/AVC 快速 Intra/Inter 模式決定系統		
出國人員姓名	連振昌	服務機構及職稱	中華大學資訊工程學系
會議時間	2009 年 9 月 12 日 至 2009 年 9 月 14 日	會議地點	日本 京都
會議名稱	(中文)IEEE 智慧訊息隱藏暨多媒體訊號處理國際會議 (英文) IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing 2009 (IIH-MSP 2009) **(EI)**		
發表論文題目	(中文)使用多模式多目標物追蹤之人員計數系統 (英文) People counting using multi-mode multi-target tracking scheme		

一、 參加會議經過

這一次參加之會議發表論文一篇，題目為 People counting using multi-mode multi-target tracking scheme。論文主題為提出一項多模式目標物追蹤技術，並另用此技術發展 front view 之人員計數技術。會議所報告之 SESSION 為

Session A01：Multimedia Signal Processing for Intelligent Applications

A01-01【IIH-MSP-2009-IS32-01】
People Counting Using Multi-Mode Multi-Target Tracking Scheme
Cheng-Chang Lien, Ya-Lin Huang, and Chin-Chuan Han
Chung Hua University, Taiwan, R.O.C.

A01-**02【IIH-MSP-2009-IS32-02】**
Human Behavior Description Model Based on Action Recognition
Fang-Hsuan Cheng and Cheng-Yuan Chang
Chung Hua University, Taiwan, R.O.C.

A01-**03【IIH-MSP-2009-IS32-03】**
A New Feature Integration Approach and Its Application to 3D Model Retrieval
Jau-Ling Shih, Chang-Hsing Lee, Chih-Hsun Chou, and Yu-Cheng Chang
Chung Hua University, Taiwan, R.O.C.

A01-**04【IIH-MSP-2009-IS32-04】**
Modulation Spectral Analysis of Static and Transitional Information of Cepstral and Spectral Features for Music Genre Classification
Chang-Hsing Lee, Hwai-San Lin, Chih-Hsun Chou, and Jau-Ling Shih
Chung Hua University, Taiwan, R.O.C.

A01-**05【IIH-MSP-2009-IS32-05】**
Face Image Super-Resolution Using Two-dimensional Locality Preserving Projection
Yuan-Kai Wang and Cai-Ren Huang
Fu Jen University, Taiwan, R.O.C.

A01-**06【IIH-MSP-2009-233】**

Random Number Generator Designed by the Divergence of Scaling Functions Jengnan Tzeng, I-Te Chen, and Jer-Min Tsai
National Chengchi University, Taiwan, R.O.C.

A01-**07【IIH-MSP-2009-292】**

A Provably Secure Two-Party Attribute-based Key Agreement Protocol
Hao Wang, Qiuliang Xu, and Tao Ban
Shandong University, China.

A01-**08【IIH-MSP-2009-317】**

A Secure Steganography: Noisy Region Embedding
Yifeng Lu, Xiaolong Li, and Bin Yang
Peking University, China.

於會議中也日本學者交換意見。IIHMSP 會議是由 IEEE 協會主辦，會議所發表之論文也會被 EI 所索引，以增加論文之能見度。這一次 IIHMSP2009 會議之 Program 及參與會議之圖像記錄列表如下：

Sept. 11, 2009		
Time	Schedule	Place
~~14:00--17:00~~ 15:00--20:00	Registration	Lobby of the sixth floor in MIELPARQUE, Kyoto

Sept. 12, 2009			
Time	Pragram	Place	Notes
8:15--9:00	NULL	NULL	Registration Lobby of the sixth floor in MIELPARQUE, Kyoto
9:00--9:15	Opening	Conference Room A	
9:15--10:15	Keynote speech	Conference Room A	
10:15--10:30	Coffee Break	Conference Room A	
10:30--12:30	Session A01--A05	Conference Room A--E	
14:00--16:00	Session A06--A10	Conference Room A--E	
16:00--16:20	Coffee Break	Conference Room A	
16:20--18:20	Session A11--A15	Conference Room A--E	
19:00--20:30	Receiption Party	the seventh floor in MIELPARQUE, Kyoto	

Sept. 13, 2009			
Time	Pragram	Place	Notes
8:15--9:00	NULL	NULL	Registration Lobby of the sixth floor in MIELPARQUE, Kyoto
9:00--10:00	Keynote speech	Conference Room A	
10:00--10:20	Coffee Break	Conference Room A	
10:20--12:20	Session B01--B05	Conference Room A--E	
14:00--16:00	Session B06--B10	Conference Room A--E	
16:00--16:20	Coffee Break	Conference Room A	
16:20--18:20	Session B11--B15	Conference Room A--E	
19:00--21:00	Banquet Party	Genji Ballroom of Hotel Granvia Kyoto (3F)	

Sept. 14, 2009			
Time	Pragram	Place	Notes
8:15--9:00	NULL	NULL	Registration Lobby of the sixth floor
9:00--10:00	Keynote speech	Conference Room A	

10:00--10:20	Coffee Break	Conference Room A	in MIELPARQUE, Kyoto
10:20--12:20	Session D01 and C01--C03	Conference Room A--D	
14:00--16:00	Session C04--C06	Conference Room B--D	
16:00--16:20	Coffee Break	Conference Room A	
16:20--18:20	Session C07 and P01--P02	Conference Room B--D	

二、 與會心得

每一次國際會議都會發現許多新技術的發展與應用，以及了解多媒體領域發展的新趨勢。因此研究要與國際接軌，參加國際學術研討會是必要的。

三、考察參觀活動(無是項活動者略)

無此項活動

四、建議

無

五、攜回資料名稱及內容

攜回資料為會議議程及論文光碟片。

六、其他

無

國科會補助計畫衍生研發成果推廣資料表

日期 2010年10月26日

國科會補助計畫	計畫名稱: 使用空間與時間關聯預測之H.264/AVC快速Intra/Inter 模式決定系統 計畫主持人: 連振昌 計畫編號: 98 -2221-E -216 -034 - 學門領域: 影像處理
研發成果名稱	(中文) “Fast Intra/Inter Mode Decision for H.264/AVC Using the Spatial-Temporal Prediction Scheme,” IEEE Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing in conjunction with the UIC’09 and ATC’09 conferences (英文) “Fast Intra/Inter Mode Decision for H.264/AVC Using the Spatial-Temporal Prediction Scheme,” IEEE Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing in conjunction with the UIC’09 and ATC’09 conferences
成果歸屬機構	中華大學 ; 發明人(創作人): 連振昌,陳志昌
技術說明	(中文)我們可以發現在JM-14.0參考軟體中，高複雜的Intra/Inter frame完整搜尋方式造成編碼過程非常耗時。所以，許多的方法應運而生，如SAD、相同區塊的分析、邊緣偵測等都是為了抉擇最佳的動態預估模式。然而，這些的運用都需要在動態補償過程中，花費額外的時間與資源。在此計畫中，我們所發展之技術不需要額外的影像處理，就可以達成的快速視訊壓縮編碼。本計劃構思利用當前畫面和參考畫面間，分析Intra/Inter預測模式空間與時間的關聯性，找出在空間和時間上，每一個macroblock預估模式的關聯性後，建構快速模式決定的方法。而且，在模式處理過程中，採取漂移補償來避免誤差的累積。目前於實驗數據中，可以發現總運算量的時間可以節省約百分之七十，傳輸率(bit rate)方面只增加百分之三點七，平均的PSNR大約只有損失0.08 dB。 (英文)The high computation cost of the full search method in the reference software JM-14.0 make the encoding process inefficient. Therefore, the methods of applying the SAD (sum of absolute difference), homogeneous region analysis, and edge detection are developed to determine the optimum motion estimation mode. However, the additional computation cost of the image processing will reduce the efficiency of the motion compensation process. In this paper, the spatial-temporal correlations between the current frame and the reference frame are analyzed to develop a fast mode decision method for Inter/Intra frame encoding in which no extra image processes are used. Furthermore, the concept of drift compensation is adopted to avoid the error accumulation phenomenon during the mode decision process. The experimental results show that the total computation
產業別	研究發展服務業
技術/產品應用範圍	視訊壓縮相關產品
技術移轉可行性及預期效益	於視訊串流之技術移轉具可行性，預期效益在於大幅加速H.264壓縮速度。

註：本項研發成果若尚未申請專利，請勿揭露可申請專利之主要內容。

98年度專題研究計畫研究成果彙整表

計畫主持人：連振昌			計畫編號：98-2221-E-216-034-				
計畫名稱：使用空間與時間關聯預測之H.264/AVC快速Intra/Inter 模式決定系統							
成果項目			量化			單位	備註（質化說明：如數個計畫共同成果、成果列為該期刊之封面故事...等）
			實際已達成數（被接受或已發表）	預期總達成數(含實際已達成數)	本計畫實際貢獻百分比		
國內	論文著作	期刊論文	0	0	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	0	0	100%		
		專書	0	0	100%		
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（本國籍）	碩士生	2	2	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		
國外	論文著作	期刊論文	0	1	0%	篇	準備投期刊中
		研究報告/技術報告	0	0	100%		
		研討會論文	1	1	100%		’Fast Intra/Inter Mode Decision for H.264/AVC Using the Spatial-Temporal Prediction Scheme,’ IEEE Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing in conjunction with the UIC’09 and ATC’09 conferences, Brisbane Australia, July 7- 9, 2009, pp. 194-199. (EI)

		專書	0	0	100%	章/本	
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（外國籍）	碩士生	0	0	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		

其他成果 (無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。)	已發表 EI 研討會論文 Cheng-Chang Lien, Chih-Chang Chen and Yang-Kai Chang, ’Fast Intra/Inter Mode Decision for H.264/AVC Using the Spatial-Temporal Prediction Scheme,’ IEEE Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing in conjunction with the UIC’09 and ATC’09 conferences, Brisbane Australia, July 7- 9, 2009, pp. 194-199. (EI)

	成果項目	量化	名稱或內容性質簡述
科教處計畫加填項目	測驗工具(含質性與量性)	0	
	課程/模組	0	
	電腦及網路系統或工具	0	
	教材	0	
	舉辦之活動/競賽	0	
	研討會/工作坊	0	
	電子報、網站	0	
	計畫成果推廣之參與（閱聽）人數	0	

國科會補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

■達成目標

□未達成目標（請說明，以 100 字為限）

□實驗失敗

□因故實驗中斷

□其他原因

說明：

2. 研究成果在學術期刊發表或申請專利等情形：

論文：■已發表 □未發表之文稿 □撰寫中 □無

專利：□已獲得 □申請中 ■無

技轉：□已技轉 □洽談中 ■無

其他：（以 100 字為限）

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）（以 500 字為限）

H.264 視訊壓縮快速演算法的研究一直在各項串流應用上扮演極重要的目的，例如安全監控之視訊流傳輸應用。本計劃，考慮在目前 frame 和參考 frame 之間的空間-時間的關聯性去發展快速模式決定方法，他不需要額外的影像處理。此外，漂移補償的概念被採用去避免在模式決定處理期間的誤差值累積現象。在 P-frame 中的 intra 預測模式也是用於這種方法，而且我們可以根據這種架構去發展演算法。不需要有額外的計算，可以得到較好的結果。實驗結果顯示總共的計算 cost 可以降低大約 70%而且平均的 PSNR 只掉下的大約 0.08dB。