

基於兩級式類神經網路分類器之快速立體視訊

編碼技術

陳韋志、徐國峰、江瑞秋

國立中正大學電機工程研究所

E-mail: {wcchen, kfhsu, jcchiang}@samlab.ee.ccu.edu.tw

摘要

立體視訊能夠讓人們感知到場景的深度變化，宛如置身其中而得到較佳的視覺饗宴。然而立體視訊的資料龐大，因此近年來立體視訊的壓縮技術引起許多學者的關注。相較於一般視訊(單通道)的壓縮，立體視訊壓縮之難處在其較複雜的參數抉擇以確保較佳的編碼效益，為了解決上述問題，我們使用 JMVM 作為立體視訊編碼平台，左眼視訊採用傳統時間域的預測編碼，右眼視訊則採用時間域以及空間域的混合預測編碼，並採用一個階層式兩級類神經分類器以進行快速模式運算，而達到快速編碼的目的。我們所提出階層式兩級類神經分類器的第一級類神經分類器用以決定 Macroblock (MB) 的區塊分割 (block partition) 方式；第二級則從 forward/backward 時間域以及 inter-view (disparity, 視差) 空間域中挑選出最有可能的參考來源，而此二級類神經分類器的輸入特徵皆由分析 inter-frame 與 inter-view 而得。此外，我們所提出的架構可再加入常用的快速移動向量搜尋演算法以達到更佳的加速效果。在實驗結果中，我們提出的演算法最多能夠節省 95% 的整體編碼時間，且只有些微的編碼效益衰減，而 bitrate (位元率) 的增加也在可接受的範圍內 (5.38%)。

一、簡介

隨著 3D 顯示器的快速發展以及高效率的視訊壓縮技術的開發，家用影音設備所能提供的娛樂性與日俱增。其中 3D 顯示器的應用中又以 3DTV 最具潛力。許多 LCD 面板廠商已積極投入立體顯示器市場，而在國外也有不少進行的 3DTV 相關計劃，如日本在 2007 年 12 月開始使用數位衛星進行 3DTV 廣播，在歐洲也有 ATTEST (Advanced Three-dimensional Television System Technologies)

[1] 計畫，以及韓國的 3DMB 系統等。

目前常見的 3D 顯示器輸入格式有 2 種，第一種為 2D+depth 格式，包含了一組彩色影像以及其對應的深度影像序列。一般而言，因深度影像的解析度低於彩色影像，其所需的傳輸頻寬相對較少，藉由 depth image based rendering (DIBR) 的技術，目前已經能夠將輸入的彩色及深度資訊，透過 ASIC 晶片實現即時的多視域影像合成 (如利浦公司所開發的 42 吋裸眼立體顯示器)。但這種顯示方式最主要的挑戰在於，經由合成產生的新視角影像有 occlusion (遮蓋效應) 的問題，且其影像品質與深度資訊的準確度有很大的關係，因此如何得到正確的深度資訊是此種 2D+depth 立體撥放格式上一個重要的課題。另一類的 3D 顯示器的輸入方式為兩個通道的彩色視訊 (左眼與右眼視訊)，此種視訊可以由兩架平行擺設的攝影機擷取，相較於 2D+depth 的格式，此種方式只適用於立體視訊之撥放，無法應用於多視域的影像合成，但其立體品質較 2D+depth 的立體格式來的好，目前為止，這兩種格式的顯示器皆有廠商使用，不過市面上常見的商品則以兩通道彩色視訊居多。

在本篇論文中，我們將探討如何針對兩通道彩色視訊格式的立體視訊進行快速壓縮，此種格式的資料量為一般 2D 視訊的 2 倍，直觀的方式可採用一般的編碼標準如 H.264/AVC，針對兩個通道的視訊各別編碼，但由於兩個視訊之間尚存在不少冗餘資訊，而無法達到好的壓縮效果。因此，如何發展有效率的立體視訊編碼技術是越來越受注目的議題。

立體視訊的壓縮標準最早為 MPEG-2 Multi-view profile 所製定 [2]，其中將左眼視訊以一般移動向量補償的方式進行預測編碼，而右眼除了參考時間上的資訊外，也將解碼完的左眼影像視為預測源 (disparity estimation, DE [3-4])，有不少的研究注重在使用兩視域間的相關性來增加高移動量區域的預測準確度 [5-7]。

H.264/AVC 是目前最熱門的視訊壓縮標準之一，它包含了許多不同於以往壓縮標準之特色，如可變大小的區塊選擇 (variable block size)、多畫面的參考架構 (multiple reference frame)、1/4 像素之精確度的移動向量 (motion vector, MV) 以及 rate-distortion optimization (RDO) [8] 等，其中又以模式選擇 (mode decision) 為 H.264/AVC 與其它編碼標準最大的區別所在。在模式選擇中會決定每個 MB 的最佳切割方式、最佳參考來源與最佳移動向量。由於需計算各種可能的 MB 移動向量與切割組合，模式選擇的運算複雜度相當高 [8-9]，因此有不少研究提出應

用於 H.264/AVC 標準的快速編碼演算法[10-13]。

在 H.264/AVC 的編碼模式中主要可以分為兩種：intra mode 與 inter mode，其中又以 inter mode 的模式運算最耗時，然而 inter mode 中的 SKIP mode 具有運算複雜度低以及其在大的量化參數 (quantization parameter、QP) 時有較大比重的特性[11-13]，因此許多研究提出以先檢驗 SKIP mode，再依序測試其他較耗時的模式之 early termination 演算法，此種演算法多是以一些觀查的結果來製定 early termination 的條件或閾值，並依可能選擇的模式順序與閾值比較以確保編碼之效率。然而此種方式難以製定一個適合用於立體視訊的條件，因此我們欲提出一種適用於 H.264/AVC 立體視訊上的快速模式選擇演算法。

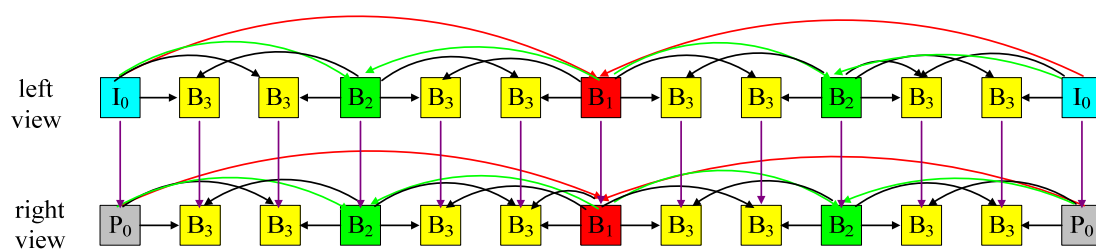
為了降低整體模式選擇與移動搜尋 (motion estimation, ME) 或視差搜尋 (disparity estimation, DE) 的運算複雜度，已有許多的快速編碼演算法被提出 [14-18]，這些方法共同的特色為利用移動向量與視差向量間的關係，以降低 ME/DE 的搜尋範圍，或是將 primary view 所編碼完的模式用來預測 secondary view 的模式。文獻[14-15]中提出依據一些合理的預測方式節省非必要的 ME 與 DE 運算，在文獻[16]中，作者提出 DTM (disparity to motion) 與 MTD (motion to disparity) 演算法，用以降低移動向量或視差向量之運算，在 DTM 中藉由先以 full search 方式計算視差向量，並用以預測移動向量，再以區域搜尋的方式做修正，而 MTD 也是類似的方式。在文獻[17]中，作者提出一個多視域視訊編碼的快速模式決定演算法，在右眼影像的 MB 編碼模式可使用相對應的左眼影像的 MB 模式來預測，藉由此種方式可以降低大量的 ME 搜尋點數，並只有減少些微的 PSNR (peak signal to noise ratio)，不過作者並沒有明確的說明整體時間的加速情況。另外，基於對 MB 觀察發現，高移動量的 MB 傾向於使用視差方向的影像作為預測源，在文獻[18]中提出另一種以移動物體為主的快速模式運算，作者藉由計算視差向量來區分前景與背景物體，將前景與背景的 MB 各別使用視差方向影像與時間域方向影像進行預測編碼。

本論文提出一個新的模式決定方式，適用於 H.264/AVC 為基礎的立體視訊編碼架構。藉由兩級式類神經網路分類器以及快速移動向量搜尋演算法，能夠有效達到快速編碼的目的，且不降低編碼效率。在第二章我們將介紹以 H.264/AVC 為基礎的立體視訊編碼架構；第三章將敘述一些對於編碼模式之實際觀察與分析，而提出的類神經分類器演算法以及實驗結果將在第四章與第五章做介紹，最

後在第六章將為此論文做總論。

二、以H.264/AVC為基礎的立體編碼架構

JMVM (Joint Multi-view Video Model) [19] 是一個以 H.264/AVC 視訊編碼標準為基礎，而適用於多視域視訊編碼的平台，已於2008完成 final draft。圖一為使用 JMVM 進行兩個視域編碼架構圖。其中，左眼視訊使用 hierarchical B pictures (HBP) 架構進行時間域上的預測編碼，而右眼視訊的編碼，除了時間軸上的向前/後畫面參考外，尚可由 inter-view 方向(左眼)進行預測。



圖一、兩通道立體視訊編碼架構 (group of picture, GOP =12)[19].

為了將 H.264/AVC 的編碼架構延伸至多視角視訊之編碼，在編碼右眼視訊時，解碼完之左眼視訊將作為參考影像並放置於時間域向前的緩衝區上，如圖二所示。因此在 JMVM 的架構上，右眼的 B frames 允許 5 種不同的預測方向，包含了“forward motion”、“backward motion”、“left-view disparity”、“forward motion + backward motion”以及“disparity + backward motion”。

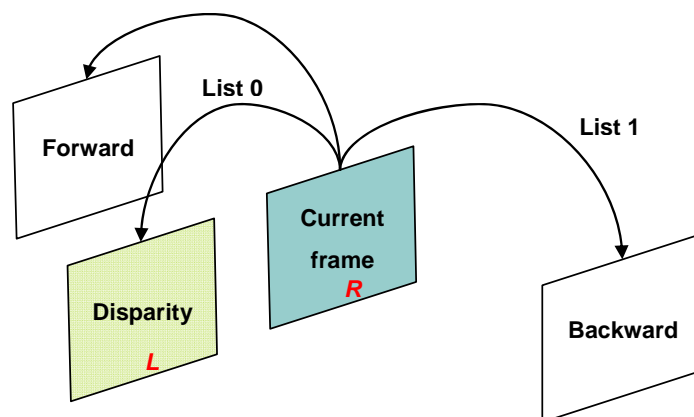
JMVM 所使用的模式決定方式與 H.264/AVC 相似，在一個固定的量化參數 (quantization parameter, QP) 之下，一個 MB 的最佳模式是以能產生最小的 Lagrange cost 來決定，

$$J(s, c, MODE | QP, \lambda_{MODE}) = SSD(s, c, MODE | QP) + \lambda_{MODE} R(s, c, MODE | QP), \quad (1)$$

$$\lambda_{MODE} = 0.85 \times 2^{(QP-12)/3}, \quad (2)$$

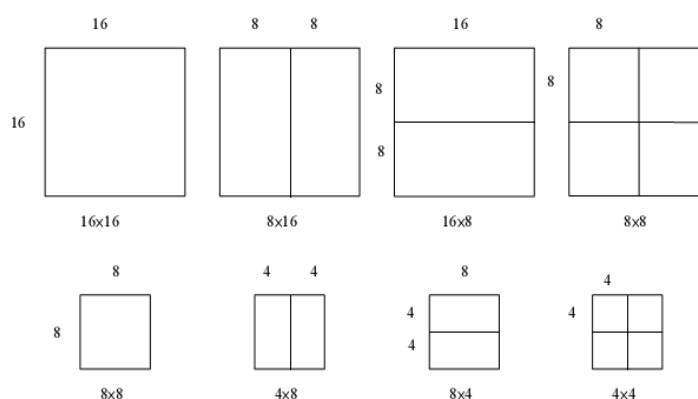
其中 $MODE$ 表示所採用的模式， s 與 c 各為原始與重建後(根據所採用的模式與 QP)的 MB 資料， λ_{MODE} 為 Lagrangian multiplier， SSD 為 s 與 c 之誤差平方和， R 為編碼 c 所需的位元量，包含了檔頭、移動向量以及殘餘值 (residual)。然而立體視訊的編碼多了視域間的參考方向，使得模式決定的複雜度遠高於傳統

H.264/AVC，也因此需花費大量的運算時間。



圖二、右眼視訊編碼時可能的參考來源

在 H.264/AVC baseline profile 中，僅有一個參考畫面之狀況下，一個 inter-coded MB 的可能模式有259種 (3+256)，前面3種可能的模式為 16x16、16x8 以及 8x16，而 8x8 模式可以有4個 sub-block partition，使得一個 MB 可能的組合數為 $4^4 = 256$ 種，如圖三所示。當使用 JMVM 進行立體視訊編碼時，右眼影像可以有一個以上的參考方向，由於可能的參考方向組合為5種，一個MB 可能的模式則有160055 ($160055 = 5 + 5^2 + 5^2 + 20^4$) 種可能。其中16x16 有5種可能方向， 5^2 為 16x8 的可能方向， 5^2 為 8x16 的可能方向， $(5+5+5+5)^4$ 為 8x8 的可能方向，值得注意的是模式決定所需運算的模式由 259 種增加至 160055 種，顯現出立體視訊編碼需要龐大的運算量。



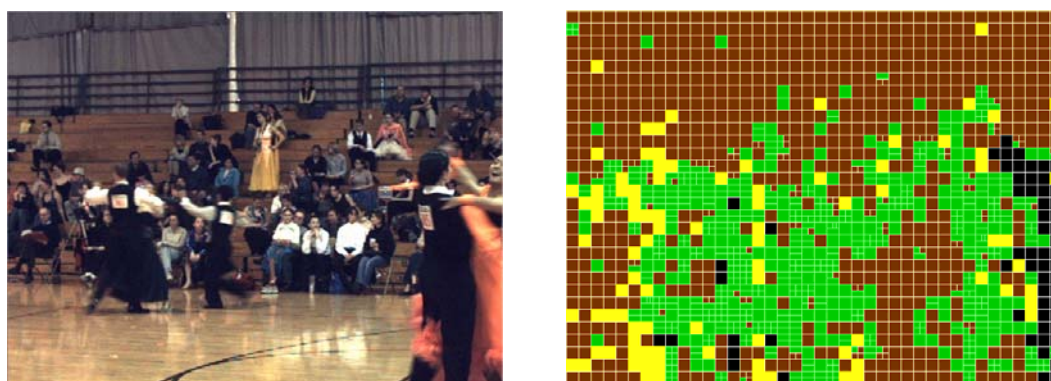
圖三、H.264/AVC 所支援的多種不同大小區塊選擇

三、對於編碼模式分佈之觀察

我們發現，模式分佈與視訊內容(如畫面中的移動量)以及編碼時所使用的參數如 QP、時間上的階層以及參考架構等息息相關，透過對模式分佈的觀察將有助於設計一個有效率加速演算法。接下來，我們將一一介紹這些會影響模式分佈的因子。

A. 模式分佈 vs. 移動程度

立體視訊編碼不同於單一通道的視訊編碼，視域間的冗餘資訊需要有效的消除，以達到較佳的編碼效益。一般而言，視訊中的背景區域以及低移動量的區域可以藉由時間上的預測來消除冗餘資訊，而高動態的區域(如前景區域)則較傾向於使用視差方向的預測。圖四 (b) 表示了不同區域所使用的編碼模式，其中舞者部分的區域(高動態區)大多採用視差方向的預測(黃色區塊)。



(a) 原始視訊

(b) 模式分佈

- : Disparity
- : Temporal inter
- : Direct (temporal)
- : Intra
- : Disparity + backward

圖四、“Ballroom”測試序列的各種模式分佈

B. 模式分佈 vs. 量化參數

由公式(1)(2)可以清楚的看出 MB 編碼區塊的大小(模式)與 QP 有關，表一呈現了數個測試視訊使用 JMVM 編碼平台在全模式選擇(考慮所有可能的區塊切割與參考方向)下，6種不同的區塊切割(Direct、16×16、16×8、8×16、8×8以及其 sub-blocks、Intra)在不同 QP 設定時的統計性質。由表中可得知在 QP 較大

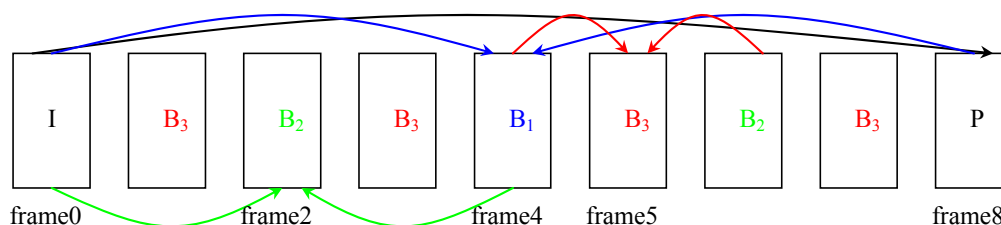
時，最有可能選擇 Direct 與 16x16 模式，這可由公式(2)來解釋。當 QP 大時， λ_{MODE} 相對較大而使得 bitrate R 主導了整個 RD cost，因此 bitrate 需求較少的 Direct 與 16x16 模式較有可能被選到。相對的，當 QP 較小時，模式分佈則較平均 (由於 λ_{MODE} 相對較小使得主導 RD cost 的項次為 distortion)，因此快速模式決定需依據不同的 QP 做最佳化的設計。

表一、不同QP下的模式分佈

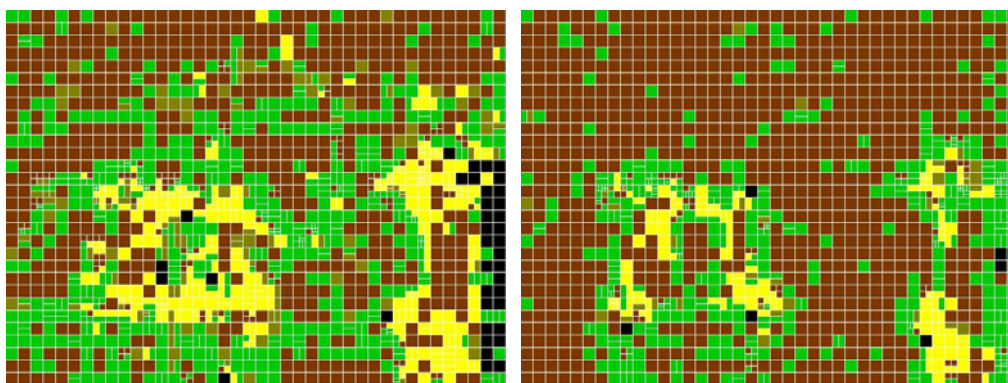
	QP=16	QP=28	QP=40
Direct	26.60%	64.59%	79.81%
16x16	19.06%	21.37%	14.71%
16x8	10.17%	4.86%	2.14%
8x16	9.31%	4.30%	1.97%
8x8	20.76%	2.98%	0.42%
Intra	14.10%	1.90%	0.95%

C. 模式分佈 vs. 時間層級

由於 JMVM 使用“ hierarchical B”的參考架構，每個畫面與其參考畫面在時間域上的距離與其層級有關，舉例而言，當 GOP 大小為 8 時，如圖五 (a) 所示，frame 4、frame 2 與 frame 5 (各被標記為時間階層中的 B₁、B₂ 與 B₃) 和其參考畫面的距離各為 4、2、1。經由實驗結果，我們觀察到越高階層的畫面(如 B₁) 會傾向於以視差方向來預測，這是由於其時間上的參考畫面距離較遠使得其相關性減少，圖五 (b) 與 (c) 印證了這個現象，在 frame 4 中以視差方向預測的 MB 數量大於 frame 5 中的個數。



(a) GOP為8時的temporal hierarchy



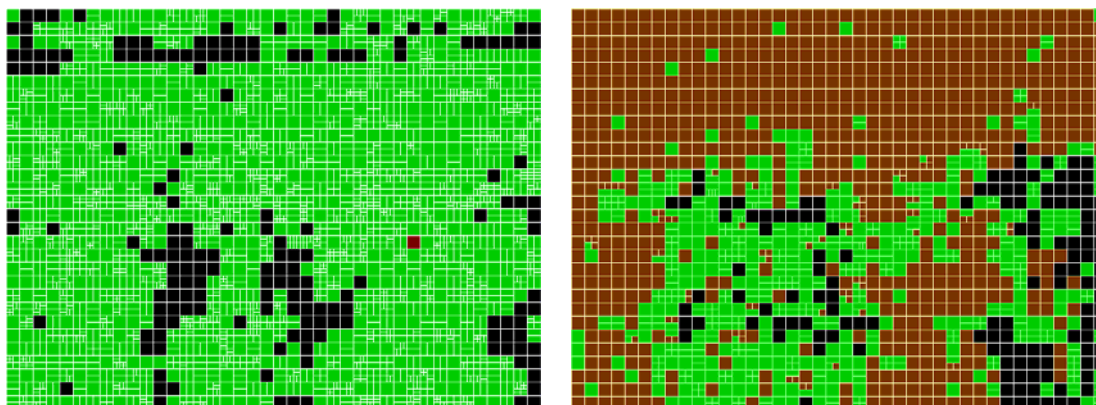
(b) “Ballroom” frame 4

(c) “Ballroom” frame 5

圖五、不同時間層級影像的模式分佈

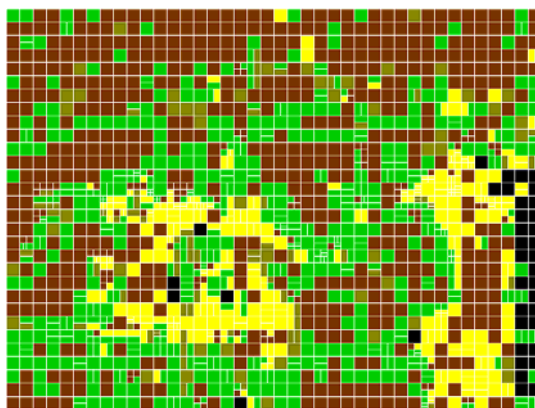
D. 區塊大小 vs. 參考架構

為了驗證不同的參考架構對 MB 大小的影響，圖六顯示了三個實驗的結果，其中(a)為 H.264/AVC baseline profile 使用 IPPP 編碼架構，(b)為 H.264/AVC main profile 使用 IBBB 編碼架構，而(c)為 JMVM 使用 IBBB 編碼架構 (使用 GOP12)下的模式分佈狀況，其中測試的視訊為“Ballroom”，而QP為16。在 H.264/AVC baseline profile 中，我們僅使用向前的移動向量預測，而在 H.264/AVC main profile 中我們使用了向前與向後的移動向量預測。由圖六 (a) 與 (b) 中可以明顯看出小區塊的數量 (如 4x8、8x4 與 4x4) 在 baseline profile 中較多，而在 main profile 中則減少了。另外由圖六 (b) 與 (c) 可以發現，使用時間方向進行預測的小區塊會被使用視差方向預測的大區塊取代，這些行為顯示了某些小區塊或許可以採用較大的區塊分割取代，並且不會對編碼效率有太大的影響。



(a) H.264 baseline profile

(b) H.264 main profile



(c) JMVM

圖六、“Ballroom” 測試序列在 QP = 16 不同參考架構下的模式分佈

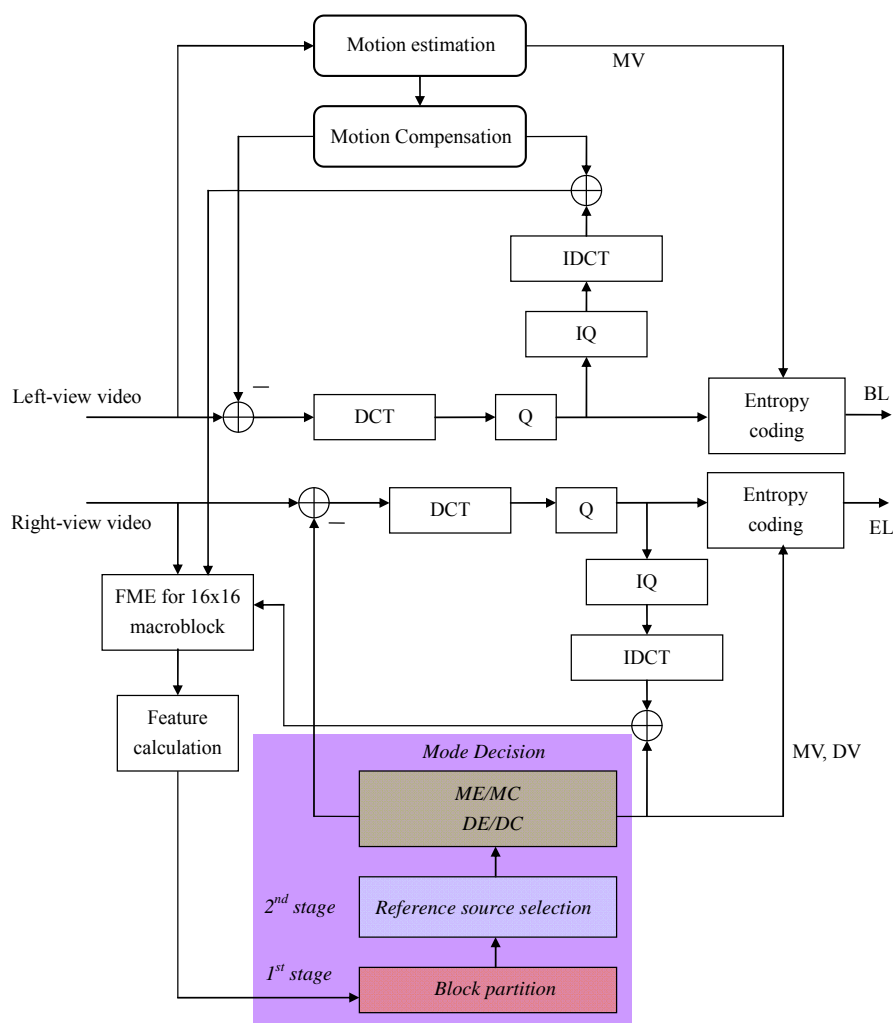
四、本論文提出之類神經網路架構

根據上一章節對編碼模式之性質觀察，本論文提出一個可用於立體視訊編碼快速模式決定之兩極階層式類神經網路分類器。第一級分類器選擇較佳的區塊大小，而第二級分類器則針對預測來源做選擇，此兩級分類器皆使用立體視訊上時間域或空間域所取出之特徵。此外本論文所提出之快速模式演算法尚能結合一般常用之快速移動向量搜尋法 (如 JMVM 所使用的 diamond search)，達到更進一步的加速效果。圖七為本論文提出的編碼架構。

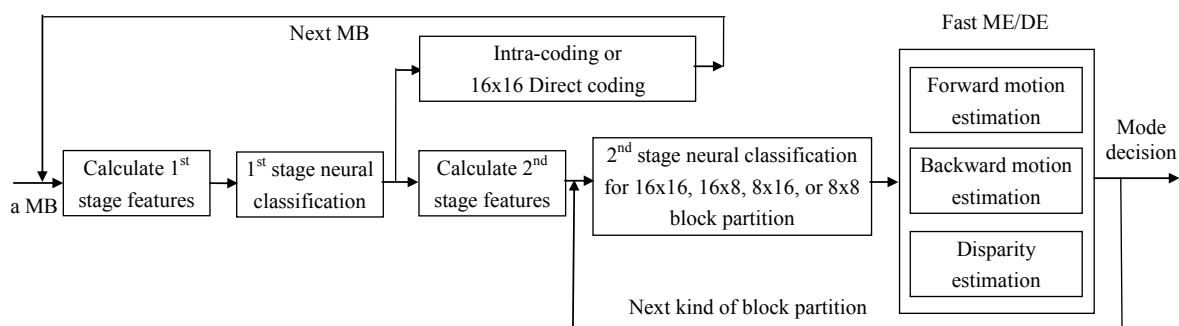
由先前討論，我們知道在 JMVM 平台上對於右眼的每個 MB 總共有 160050 種可能的模式，而這些模式可將其對應至兩級式的階層架構：

- (1) 第一級：不同大小之區塊切割，分為 16x16 Direct、16x16、16x8、8x16、8x8 以及 Intra。
- (2) 第二級：參考來源之方向，包含 forward motion、backward motion 以及 disparity。

在經過第二級的類神經分類器之後，被分類為 8x8 的區塊將進一步測試 4x8、8x4 以及 4x4 分割的組合，圖八顯示了我們所提出的演算法之處理流程。



圖七、本論文提出的應用於立體視訊編碼快速模式決定之架構圖



圖八、應用於立體視訊編碼之階層式兩級類神經網路分類器

A. 第一級分類器

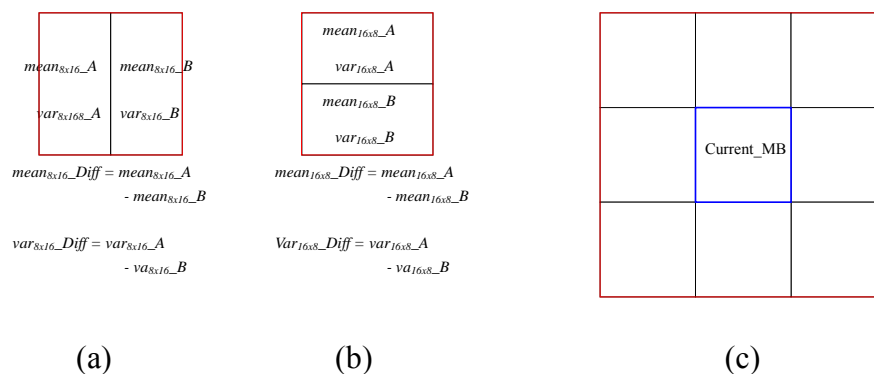
一般而言，區塊大小的決定與影像內容的特性關係密切，特別是材質資訊影響為最。舉例來說，在材質平滑的區域有較大的可能是使用大的區塊模式來進行編碼，如 Direct/Skip 或 16x16 inter modes。相對的，小區塊模式如 8x8、8x4、

4x8、4x4 則較常使用於編碼材質較複雜的區域。另一個主要影響區塊切割的因素為畫面的移動量，通常高移動量的區域為了能更精確的預測，會使用較小的區塊切割。

我們首先計算各別經由 forward-motion 以及 disparity 方向經移動補償 (motion compensation, MC) / 視差補償 (disparity compensation, DC) 後之 MB 殘餘值 $D_{t-h,t}$ 與 $D_{d,t}$ ，在此使用的是快速 ME/DE 搜尋法以降低運算時間，其中 d 表示以 disparity 方向為參考方向， t 與 $t-h$ 代表不同的時間點，而時間距離 h 是依據現在的畫面其在參考架構的層級決定。舉例來說，對一個 B₁ 層級的畫面， h 的值為 GOP 大小的一半 (請參照圖五 (a))。為了消去參考來源對於決定區塊切割的影響， $D_{t-h,t}$ 與 $D_{d,t}$ 中 L_0 -norm 較小的，將用來計算輸入類神經分類器的特徵，這邊要注意的是 $D_{t-h,t}$ 與 $D_{d,t}$ (參照圖七中的“FME for 16x16 macroblock” 區塊) 可以再次使用於 16x16 inter 與 disparity 這些模式 RD cost 的計算上。

以下為第一級類神經網路所使用的 9 個影像特徵，皆取自 $D_{t-h,t}$ 與 $D_{d,t}$ 中 L_0 -norm 較小者：

1. MB 的平均值與變異數
2. MB 中 8x16 大小的左、右區塊之區塊平均值與變異數之差值 $mean_{8x16_Diff}$ 與 var_{8x16_Diff} ，如圖九 (a) 所示
3. MB 中 16x8 大小的上、下區塊之區塊平均值與變異數之差值 $mean_{16x8_Diff}$ 與 var_{16x8_Diff} ，如圖九 (b) 所示
4. 高殘餘值像素(與一個先前定義之閾值相比)的個數在目前 MB 中所佔的百分比
5. 目前 MB 與其周圍8個相鄰 MB 的平均值與變異數之差值，如圖九 (c) 所示



圖九、一級類神經網路分類器之特徵計算

請注意圖九 (a) 與 (b) 可以視為兩種適用於驗證 8x16 與 16x8 區塊分割的樣版。

上述的特徵將用於輸入第一級類神經網路分類器，我們使用的一級類神經網路分類器為三階層式的 back-propagation neural network (BPNN)，其中輸入層 (input layer)、隱藏層 (hidden layer) 以及輸出層 (output layer) 的神經元個數分別為 9、9、6。所有輸入的特徵值皆正規化至 0.0~1.0 的範圍。經由足夠的訓練過程之後，分類器會依據欲編碼之 MB 其輸入的特徵向量，輸出 6 個值域在 0.0~1.0 的數值，分別代表每種可能模式的分數。我們使用的策略為以輸出分數的高低篩選可能的編碼模式個數 K_1 ，使用者可依不同的編碼品質與速度需求使用不同的 K_1 ，無疑地，較小的 K_1 值將可以大幅的加速模式決定程序，但是也將增加 bitrate (由於省略了一些模式運算使得預測編碼的效率降低)。反之若是使用較大的 K_1 ，編碼時間的減少量較低，但能夠維持較好的編碼效率 ($K_1 = 6$ 等同 JMVM 平台的全模式選擇)。為了在編碼效率與時間上取得平衡，我們使用一個適應性選擇 K_1 的演算法，將輸出的分數與一個由經驗得到的閾值比較。當輸出的分數高於閾值則其對應的區塊切割則列為可能的模式之一並接續後面的運算，請注意此 6 種區塊分割所需接著進行的運算時間不盡相同，其中“16x16 Direct”與“Intra”不需再進行第二階段的類神經分類以及移動向量/視差向量之運算，然而“8x8”需要再進一步計算 8x4、4x8 以及 4x4 的子區塊分割模式。

B. 第二級分類器

根據先前的討論，立體視訊編碼的參考來源相較於傳統 H.264/AVC 複雜許多，第二級類神經網路的方式類似第一級，採用經 MC/DC 的 MB 殘餘值，計算其特徵，包括 $D_{t-h,t}$ 、 $D_{t,t+k}$ 與 $D_{d,t}$ ($D_{t,t+k}$ 為 backward motion 的參考來源)。

與第一級類似，我們計算 MB 殘餘值的區域平均值與變異數，在此的區域為經由第一級類神經分類器的結果得到的區塊 (兩級形成一階層架構，並非單純串聯)。舉例來說，假設 16x8 的區塊為第一級 K_1 個候選模式之一，則第二級計算區域平均值與變異數時即使用 16x8 的區塊計算。下面列出第二級類神經網路使用的 6 個特徵。

1. $D_{t-h,t}$ 的區域平均值與區域變異數

2. $D_{t,t+k}$ 的區域平均值與區域變異數

3. $D_{d,t}$ 的區域平均值與區域變異數

第二級類神經網路使用的輸入層、隱藏層以及輸出層的神經元個數分別為6、5、3，如第二章所提及的，JMVM 平台允許 5 個參考來源 (包括 3 個方向的組合)，但在此我們只將來源分為 3 類，分別是 forward motion、backward motion 以及 disparity 方向。同樣地我們從輸出的分數中挑選 K_2 個 (至多為3) 可能的參考來源，並以其組合作為可能預測方向。舉例來說，若是“disparity”與“backward motion”為篩選後的輸出，則我們會測試“disparity”、“backward”以及“disparity + backward”三個參考來源。

與第一級類似，第二級也採用一個適應性 K_2 選擇演算法，首先將輸出的分數以一個距離閾值進行分群 (採用 nearest neighbor distance [20])，分為1至3群，只有輸出分數屬於平均分數最高的一群為所求。

C. 快速移動搜尋

經過本論文提出的兩級類神經分類器處理後的結果，可搭配現有的各種快速演算法 (如three-step search、diamond search 或其它 JMVM 所採用之方法) 進行快速移動向量/視差向量之運算，以達到更快速之模式決定流程。在本論文的實驗中使用的是 JMVM 平台所支援的快速移動搜尋 (fast motion estimation, FME) 演算法。

五、實驗結果

我們共使用了6組立體視訊進行測試，每一組立體視訊的大小為 640x480，共 100組畫面，其中“Ballroom” (物體高移動量)、“Race1” (物體與攝影機皆移動)、“Exit”以及“Vassar” (物體低移動量)為使用一維平行擺設之攝影機陣列所拍攝的多視域視訊 (只取其中相鄰的兩個視域進行實驗)，而“Soccer2” (物體與攝影機皆移動)與“Puppy” (物體低移動量)為立體拍攝的立體視訊對。

我們使用的測試平台為 JMVM8.0，並以其實現 full-mode decision 的結果作為比較的對象，下面將針對右眼視訊編碼的位元率 (bitrate)以及整體編碼時間進行分析討論。表二列出了我們實驗環境使用的參數，其中考慮攝影機大部分為水

平架設，因此垂直方向的視差搜尋範圍的設定較小 $([-8, 8])$ 。而“Direct”模式在此設定其參考方向為“forward + backward motion”。根據第三章 B 與 C 節的觀察，模式分佈隨著 QP 的不同與時間上的層級而變化，因此我們在進行類神經分類器的訊練時依據不同的 QP (QP = 16、28以及40) 以及不同時間層級影像 (B_1 、 B_2 以及 B_3)進行訊練。

表二、實驗所使用的立體視訊編碼參數

Frame format	4:2:0
GOP size	12 frames
Number of prediction frames for B frames	3 frames (forward, backward and disparity direction)
Disparity search range	$X \in [-16, 16]$ pixels, $Y \in [-8, 8]$ pixels
Direct mode	Forward motion + backward motion
Motion search range	$[-16, 16]$ pixels
RDO	On
Deblocking filter	On
1/4 Sub-pixel search	On
Quantization Parameters	16, 28, 40

A. 單一級類神經網路的編碼效益分析

在此實驗中，我們測試了單純使用各個不同的方式進行加速，除了比較本論文提出的類神經分類器之外(標示為 NN1 與 NN2)，也比較使用機率統計 (Probability-based)的方式進行第一級(區塊切割)的加速。機率統計為依據 6 種區塊切割的機率統計性質，固定選擇前 K_1 種區塊切割。表三 (a) 至 (e) 顯示了“probability-based ($K_1=3$)”，NN1 (fixed $K_1=3$)，NN1 (adaptive K_1) 以及 NN2 的編碼效能，表中分別列出了相對於 JMVM full-mode decision 的 Δ PSNR、 Δ bitrate 以及 TS (Time saving)。

其中“probability-based”方法對於任何的 MB 皆選擇固定的切割模式(以表 1 為例，當 QP=28、40時固定選擇 Direct、16x16以及16x8)，由表三(a)可觀察到，在中低位元率的情況下可以節省74%的時間，然而在高位元率時效能驟降，只有約23%。值得注意的是，此種方式導致位元率增加頗多，特別是“Ballroom”的測試序列。

除了表三 (b) 與 (c) 顯示的資料外，圖十繪出了本論文使用的 NN1 與 Probability-based 在不同的 K_1 ($K_1 \in \{2,3,4,5\}$) 之下其 TS 與增加的位元率之比

較，我們使用了“Exit”測試序列，搭配QP=28，由圖中可以看出NN1與Probability-based最佳的 K_1 值各別為3與4。另一方面，NN1 with adaptive K_1 (相當於 $K_1=2.27$)的方式與NN1固定 $K_1=3$ 的效能相近。

為了驗證本論文提出的NN1的正確性，表四列出了輸出分數的Top 3與Top 1和真正的區塊切割(經由full-mode search得到)的hit rate統計，可以發現在QP=28、40時平均top-3的hit rate高達90%。雖然在QP=16時的hit rate稍低，但仍維持在60%，且這邊的hit rate已考慮到相鄰模式的選擇錯誤而會有錯誤漫延的狀況。而由表三(b)與(c)的細節可以看出的適應性 K_1 表現優於固定 $K_1=3$ ，表三(d)中可以看出NN2能夠節省最多71%的運算時間，少於NN1所能節省的最大時間。

B. 階層式兩級類神經網路的編碼效益分析

表五顯示了NN1+NN2的實驗結果，可以看出在大QP的狀況下最多能夠節省89%的運算時間，且位元率增加在6%以內。對於中高位元率的應用上，將NN1與NN2整合的結果能進一步節省5%~12%的運算時間，且位元率的增加小於3%。值得注意的是，在QP=16的狀況之下依然能夠節省約60%的運算時間。

C. 應用快速移動搜尋於階層式兩級類神經網路的編碼效益分析

表六顯示了整合NN1+NN2+FME的編碼架構其效能，在QP=16的狀況下能達到節省82%的編碼時間，而在QP=40時更能達到節省95%的編碼時間，而位元率最多增加5.38%。

表三、不同編碼架構於不同QP下之效能分析

(a) Probability-based

sequence	QP=16			QP=28			QP=40		
	Δ PSNR (dB)	Δ bitrate (%)	TS (%)	Δ PSNR (dB)	Δ bitrate (%)	TS (%)	Δ PSNR (dB)	Δ bitrate (%)	TS (%)
Ballroom	-0.11	1.33	23.03	-0.08	6.79	74.40	-0.08	4.66	73.65
Exit	-0.14	-0.21	23.08	-0.04	4.93	74.03	-0.05	1.96	73.89
Vassar	-0.04	-0.29	22.77	-0.01	0.50	74.04	-0.01	0.61	73.70
Soccer2	-0.10	0.08	23.47	-0.05	4.25	73.99	-0.05	1.04	73.42
Race1	-0.06	1.73	23.28	-0.04	1.68	73.92	-0.03	-1.53	73.72
Puppy	-0.05	-0.06	22.67	-0.01	0.11	74.20	-0.01	0.02	73.50

(b) NN1 with a fixed $K_1=3$

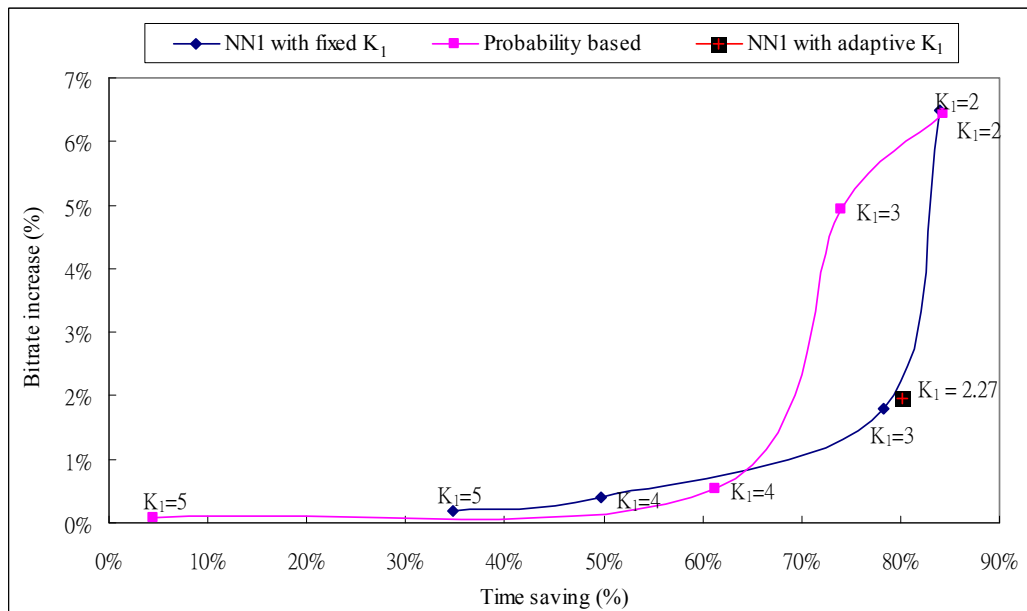
sequence	QP=16			QP=28			QP=40		
	Δ PSNR (dB)	Δ bitrate (%)	TS (%)	Δ PSNR (dB)	Δ bitrate (%)	TS (%)	Δ PSNR (dB)	Δ bitrate (%)	TS (%)
Ballroom	-0.06	1.21	28.31	-0.06	4.26	74.80	-0.07	4.51	79.97
Exit	-0.09	0.29	32.43	-0.03	1.79	78.36	-0.03	0.77	80.92
Vassar	-0.04	0.08	23.63	-0.01	0.21	77.11	-0.01	3.44	78.14
Soccer2	-0.08	0.82	27.99	-0.04	1.55	74.37	-0.04	2.40	79.09
Race1	-0.02	3.53	24.07	-0.05	3.93	68.09	-0.02	-0.07	77.29
Puppy	-0.04	0.76	29.92	-0.01	0.32	78.23	-0.02	0.43	75.41

(c) NN1 with adaptive K_1

sequence	QP=16				QP=28				QP=40			
	Δ PSNR (dB)	Δ bitrate (%)	TS (%)	Avg. K_1	Δ PSNR (dB)	Δ bitrate (%)	TS (%)	Avg. K_1	Δ PSNR (dB)	Δ bitrate (%)	TS (%)	Avg. K_1
Ballroom	-0.06	1.28	29.18	2.96	-0.06	4.41	76.01	2.33	-0.07	4.51	79.97	2.51
Exit	-0.09	0.36	33.16	2.93	-0.03	1.95	80.11	2.27	-0.03	0.77	80.92	2.52
Vassar	-0.04	0.08	24.69	2.98	-0.01	0.18	78.90	2.45	-0.01	3.44	78.14	2.51
Soccer2	-0.08	0.87	29.66	2.97	-0.05	1.46	76.31	2.61	-0.04	2.40	79.09	2.31
Race1	-0.02	4.10	25.91	2.85	-0.05	3.32	70.55	2.50	-0.02	-0.07	77.29	2.69
Puppy	-0.04	0.84	31.22	2.92	-0.01	0.41	79.57	2.43	-0.02	0.43	75.41	2.72

(d) NN2

sequence	QP=16			QP=28			QP=40		
	Δ PSNR (dB)	Δ bitrate (%)	TS (%)	Δ PSNR (dB)	Δ bitrate (%)	TS (%)	Δ PSNR (dB)	Δ bitrate (%)	TS (%)
Ballroom	-0.02	0.74	38.82	-0.01	0.61	47.81	-0.01	0.14	56.18
Exit	-0.01	1.00	40.57	-0.01	0.72	54.71	-0.01	0.25	67.54
Vassar	-0.02	0.46	40.28	0.00	-0.04	55.13	0.00	-0.19	71.27
Soccer2	-0.01	0.55	29.36	-0.02	2.11	39.35	-0.01	1.21	53.30
Race1	-0.01	0.76	21.55	-0.02	1.18	32.22	-0.03	1.85	34.46
Puppy	-0.01	0.94	37.17	-0.01	-0.08	63.26	0.00	-0.16	69.14



圖十、“Exit”序列在 QP=28 時，不同的 K_1 值下其 TS 與 位元率之比較

表四、第一級分類器的hit rate表現

sequence	QP=16		QP=28		QP=40	
	Top 1	Top 3	Top 1	Top 3	Top 1	Top 3
Ballroom	30%	65%	58%	84%	75%	93%
Exit	23%	59%	73%	94%	90%	98%
Vassar	35%	73%	78%	97%	88%	99%
Soccer2	23%	58%	55%	88%	75%	95%
Race1	21%	50%	54%	86%	71%	93%
Puppy	27%	64%	87%	99%	88%	99%

表五、階層式兩級快速編碼架構的效能分析

sequence	QP=16			QP=28			QP=40		
	Δ PSNR (dB)	Δ bitrate (%)	TS (%)	Δ PSNR (dB)	Δ bitrate (%)	TS (%)	Δ PSNR (dB)	Δ bitrate (%)	TS (%)
Ballroom	-0.09	2.28	55.88	-0.08	5.67	84.42	-0.09	5.27	85.69
Exit	-0.11	1.49	60.94	-0.05	3.08	88.13	-0.05	1.63	88.49
Vassar	-0.06	0.68	57.38	-0.02	0.29	88.06	-0.01	3.64	88.59
Soccer2	-0.10	1.56	42.50	-0.09	4.55	83.32	-0.05	4.23	86.14
Race1	-0.04	5.37	44.05	-0.07	4.56	75.37	-0.05	1.29	82.30
Puppy	-0.05	3.01	60.90	-0.02	0.16	89.67	-0.02	-0.19	88.00

表六、應用快速移動搜尋於階層式兩級快速編碼架構的效能分析

sequence	QP=16			QP=28			QP=40		
	Δ PSNR (dB)	Δ bitrate (%)	TS (%)	Δ PSNR (dB)	Δ bitrate (%)	TS (%)	Δ PSNR (dB)	Δ bitrate (%)	TS (%)
Ballroom	-0.09	2.30	78.00	-0.08	5.38	91.85	-0.08	4.56	93.30
Exit	-0.12	1.59	80.68	-0.06	3.19	92.45	-0.07	1.51	94.71
Vassar	-0.06	0.72	79.90	-0.02	-0.13	94.34	-0.01	2.90	94.87
Soccer2	-0.09	1.39	75.68	-0.09	4.04	91.65	-0.06	3.89	93.24
Race1	-0.03	5.36	71.50	-0.06	4.07	87.52	-0.08	1.42	91.19
Puppy	-0.05	3.00	82.62	-0.03	0.20	95.42	-0.03	1.11	94.65

六、結論

本研究的動機在於使用 H.264/AVC 架構的立體編碼系統時，其模式決定由於多個參考來源以及 hierarchical B-frames 的架構，使得運算複雜度過高難以符合實際應用的需求，加上傳統應用於 H.264/AVC 的快速模式運算無法直接應用於立體編碼上，因此我們提出了一個階層式兩級類神經網路分類器，包含了區塊切割與來源選擇，用以進行立體編碼之快速模式運算，藉由調整 NN1 與 NN2 的閾值參數能夠達成可變的運算複雜度與編碼效益。

由實驗結果驗證了我們提出的 NN1+NN2+FME 架構，在中低位元率時可以節省95%的編碼時間，且僅有些微的位元率增加，本論文提出的方式目前是以3DTV廣播的應用為主，但未來也可以應用至多視域視訊編碼之加速。

七、參考文獻

- [1] C. Fehn, "A 3D-TV System Based on Video Plus Depth Information," *Proc. of 37th Asilomar Conference on signals, systems and Computers*, Vol. 2, pp. 1529-1533, Nov. 2003.
- [2] ISO/IEC IS 13818-2 AMD3: MPEG-2 Multi-view Profile, Sep. 1996
- [3] J. Sun, N.-N. Zheng, and H.-Y. Shum, "Stereo Matching using Belief Propagation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 25, No.7, pp. 787-800, 2003.
- [4] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A Comparison and Evaluation of Multi-view Stereo Reconstruction Algorithms," *Proc. of IEEE Int'l Conf. on Comp. Vision and Pattern Recognition*, pp. 519-528, 2006.
- [5] E. Martinian, A. Behrens, J. Xin, A. Vetro and H. Sun, "Extension of H.264/AVC for Multi-view Video Compression," *Proc. of IEEE Int'l Conf. on Image Processing*, pp. 2981-2984, 2006.
- [6] P. Merkle, K. Müller, A. Smolic and T. Wiegand, "Efficient Compression of Multi-view Video Exploiting Inter-view Dependencies Based on H.264/MPEG4-AVC," *Proc. of IEEE Int'l Conf. on Multimedia & Expo*, pp. 1717-1720, 2006.
- [7] R. S. Wang and Y. Wang, "Multi-view Video Sequence Analysis, Compression, and Virtual Viewpoint Synthesis," *IEEE Trans. on Circuits and Systems for Video Technology*, Vol.10, No. 3, pp. 397-410, April 2000.
- [8] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini and G. J. Sullivan, "Rate-constrained Coder Control and Comparison of Video Coding Standards," *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 13 , No. 7, pp. 688-703, July 2003.

- [9] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 13, No. 7, pp. 560-576, July 2003.
- [10] F. Pan, X. Lin, S. Rahardja, K. P. Lim, Z. G. Li, D. Wu, and S. Wu, "Fast Mode Decision Algorithm for Intra Prediction in H.264/AVC Video Coding," *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 15, No. 7, pp. 813-822, July 2005.
- [11] C. Grecos, and M. Yang, "Fast Mode Prediction for the Baseline and Main Profile in the H.264 Video Coding Standard," *IEEE Trans. on Multimedia*, Vol. 8, No. 6, pp. 1125-1134, Dec. 2006.
- [12] D. Wu, F. Pan, K.P. Lim, S. Wu, Z. G. Li, X. Lin, S. Rahaedja, and C. C. Kuo, "Fast Inter-mode Decision in H.264/AVC Video Coding," *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 15, No. 7, pp. 953-958, July 2005.
- [13] L.-J. Pan, and Y.-S. Ho, "Fast mode Decision Algorithm for H.264 Inter-prediction," *Electronics Letters*, Vol. 43, No. 24, pp.1351-1353, Nov. 2007.
- [14] Y. Kim, J. H. Lee, C. Park and K. Sohn, "MPEG-4 Compatible Stereoscopic Sequence Codec for Stereo Broadcasting," *IEEE Trans. on Consumer Electronics*, Vol. 51, No. 4, pp. 1227-1236, 2005
- [15] L.-F. Ding, S.-Y Chien, Y.-W. Huang , Y.-L. Chang, L.-G. Chen, "Stereo Video Coding System with Hybrid Coding Based on Joint Prediction Scheme," *Proc. of IEEE Int'l Sym. on Circuits and Systems*, Vol. 6, pp.6082- 6085, 2005.
- [16] P. Lai and A. Ortega, "Predictive Fast Motion/Disparity Search for Multi-view Video Coding," *Proc. of SPIE*, Vol. 6077: *Visual Communication and Image Processing*, 2006.

- [17] L.-F. Ding, T.-K. Tsung, S.-Y. Chien, W.-Y. Chen, and L.-G. Chen, "Content-aware Prediction Algorithm with Inter-view Mode Decision for Multi-view Video Coding," *IEEE Trans. on Multimedia*, Vol. 10, No. 8, pp. 1553-1564, 2008.
- [18] S.-Y. Lee, K.-M. Shin, and K.-D. Chung, "An Object-based Mode Decision Algorithm for Multi-view Coding," *Proc. of IEEE Int'l Sym. on Multimedia*, pp. 74-81, 2008.
- [19] ISO/IEC JTC1/SC29/WG11/JVT-AA207, Joint Multi-view Video Model (JMVM) 8.0, Geneva, Switzerland, April, 2008.
- [20] R.O. Duda, P.E. Hart and D.G. Stork, *Pattern Classification*, Wiley-Interscience Publication, 2nd Edition, 2001.