行政院國家科學委員會專題研究計畫期中進度報告

一個漸進式架構之視訊階層分析

A Progressive Framework for Video Layer Analysis

計畫編號：NSC 94-2213-E-009-132-
執行期限：2005 年 8 月 1 日 至 2006 年 7 月 31 日
主持人：莊仁輝　　國立交通大學資訊工程學系(所)

一、中文摘要

在本報告中，我們提出關於「一個漸進式架構之視訊階層分析」計畫第一年度的具體成果。計畫中第一年度的主要目標為，經由視訊影片的全域運動與區域運動之估測，分析出影片中前景與背景兩個視訊階層。目前均已達成預期之目標，並建構出視訊階層分析系統之雛形。

我們所發展出視訊階層切割方法的核心，是以推估相機自身運動的模型為基礎，計算視訊影像中，不同區塊運動向量的差異，用以切割出與相機自身運動一致的背景部分，以及與相機自身運動相異的前景階層；同時，輔助以初始視訊影像的前景與背景階層標示，逐步推估出整個視訊影片的前景與背景兩個階層的切割結果，以達到漸進式與有效的視訊階層切割之目的。

關鍵詞：視訊階層分析、運動估計

Abstract

In this report, we describe the progress and the accomplishment for the first part of the two-year NSC research project. The primary goal is to segment foreground and background video layers via differentiating local motions of image blocks from global camera motions in a video sequence. Currently, a preliminary video layer analyzing system has been built based on the proposed methods.

The core of the proposed video layer decomposition method is based on a robust estimation of camera motions from a video sequence. Image blocks whose motions are consistent with camera motions are considered as background, while the rest image blocks are labeled as foreground. With the help of manually labeled layer information of the first image frame, more accurate block correspondences for robust estimation of an initial camera motion model can be obtained automatically. Subsequently, by propagating the newly derived layer information, we can establish a progressive framework for decomposing video layers.

Keywords: Video Layer Analysis, Motion Estimation

二、緣由與目的

本計畫第一年的研究，主旨在於建立一個漸進式的分析架構（Progressive Analysis Framework**）；開始時，先由人工標記第一張視訊影像的階層分類，其中包含前景與背景兩個階層，而後應用先前階層標記的資訊，對於連續的視訊影像，逐步估計前後兩張影像間相機的移動模式，並藉由影像區塊運動向量與所估計的相機運動之差異，找出前景與背景區域，如此對視訊影像作逐張、連續的處理，以達到初步區分前景背景階層的效果。**

三、參考文獻

針對攝影機為固定的情況下，切割的方法會比攝影機在運動的情況簡單，最常使用的方法是連續兩張畫面的相減[1]**，用此可以偵測出畫面中有物體運動的區域，做為初步切割的結果。**[2][3]**同樣是利用了兩張畫面間的差異對物體做切割，為了增加可靠性，**[2][3]**會累多張積畫面的差異資訊再做切割。**

在攝影機不為靜止的狀況，對視訊畫

面切割較為困難，[4]使用了物體追蹤與圖形識別的方法來切割運動的物體，此方法的好處是在物體有變形或者遮蔽時，仍然可以有不錯的切割效果，但是第一次步驟必須先用另外的方法做一個初始的切割，之後才可以使用物體追蹤的方式。[4]中並沒有提到一個強固的初始切割方法，而初始切割的好壞大大的影響了後續切割的結果。[5]使用了仿射模型（Affine Model）來估測物體的運動，其做法是每塊區域估計一組仿射參數，之後再合併仿射參數較像的鄰近區域。但是實際合併區域的時候，有可能屬於同一個物體卻不會被合併情況發生，因此造成同一個物體會被切割成許多的區塊。[6]同樣也利用了運動的資訊作為運動物體之切割，為了增加運動資訊的強固性，[6]所提出的方法必須要累積數張畫面間的運動資訊，並求出具有不同運動之物體個數，給定每一個物體設定一組初始的仿射參數，來代表每一個物體。最後，推導出一個機率函數，定義每一個像素屬於某個物體的機率，因此可以把每一個像素分類，對不同的物體作切割。

此外 [7][8] 則是採用正規劃分（Normalized Cuts）方法，這類方法是將影片的像素點是作一個節點（Node），點與點間的連線（Edges）則為彼此的相似度衡量，對這樣的圖形（Graph）找出一個最佳的劃分（Cut），得到視訊影片的分層結果。在 [9][10] 等研究中，則是以每個物件影像像素點的高斯模型著手，建立分層混合與運動後，對應到拍攝影像的顏色機率模型，並用期望值最大化（Expectation Maximization，EM）的方法，求出視訊分層的解。在 Ke 和 Kanade [11]所提出的方法中，是從三維幾何計算的方向入手，應用同形共面（Homographics）的性質，將影像區域作子空間投影與分解，以求出階層分解結果。而 Khan 和 Shah [12] 兩人，則提出了以最大事後機率（Maximum a Posteriori，MAP）估計的方法，先對影片作初始切割，然後再對影片中的每個點，依照初始切割區域的性質，作機率推論，以分出每個點所屬的區域類別。此外，由 Greenspan 等人 [13] 所提出，由學習視訊影像切割區域的混和高斯模型（Mixture Gaussian Model），來分析視訊影像分割，此方法是應用特定模型於影像區域的學習、分析方式。對於上述的研究方法，大都需要對影片作大量的最佳化運算，故整體的計算量較大。

三、研究內容

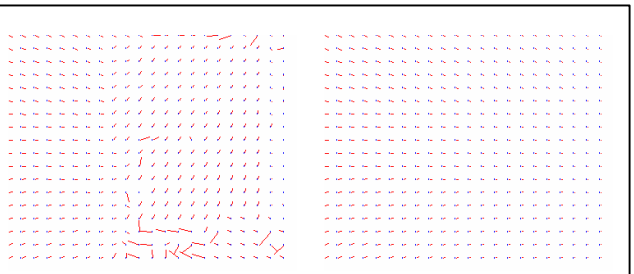
針對建立一個漸進式的視訊階層分析架構之目標，在第一年的研究中，我們著重在下列幾個研究主軸：

1. 估計連續兩張影像間之相機運動

此一步驟，主要是希望透過連續兩張影像 I_t 和 I_{t-1} 之間（如圖 1所示），各個影像區塊區域運動（Local Motion）的計算，與相機全域運動（Global Motion）的估計，得出兩張影像的座標轉換關係；由於在拍攝的過程，攝影機也有可能產生移動，故我們將先從兩張連續影像中，找出兩張影像間，屬於背景部分的特徵或區塊對應關係，以估計出攝影機的運動（如圖 2所示），並應用在其後的步驟中，切割出背景階層。

(a) 處理畫面 I_t　　(b) 參考畫面 I_{t-1}

圖1：視訊影片中的兩張畫面



(a)區塊運動向量　　(b)全域相機運動

圖2：運動向量示意圖

2. 背景階層的分割

圖 3為以區塊為基礎的區域運動向量與全域運動向量之示意圖，每一個區塊皆會有一個全域運動向量與區域運動向量，當兩者的差異越大時，物體不同於攝影機運動的程度越高，也就是說實際上物體本身是在運動的，此時就可被劃分為前景；當兩者的差異越小，物體相似攝影機運動的程度越高，此時真正的物體是靜止不動的，物體看起來會運動的原因是因為攝影機的運動。

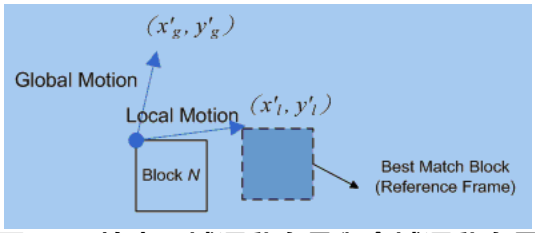


圖3：區塊之區域運動向量與全域運動向量

圖4：初始影像階層標示

3. 初始影像階層標示的使用者介面

由於我們是以相機的運動，來表示影像中背景的運動模式，因此，在估計相機的全域運動時，我們需要盡量選擇屬於背景部分的對應特徵或對應區塊關係，作為估計之用，以較準確的計算出相機運動模型；而由於一開始時，我們並沒有相關前景背景的資訊，因此我們計畫設計一個使用者介面，方便使用者快速的將第一張影像標示出數個階層的介面，其標示結果如圖 4所示，作為初始運動估計所需之先備知識（Prior Knowledge），而後漸進式的分解出後續的視訊階層。

四、結果與討論

針對上述研究主軸，我們分下列幾點說明第一年的研究成果。

1. 發展初始影像階層標示介面

我們透過介面設計，將影像區分成固定大小的區塊，使用者則是可以透過矩形區域的點選，快速的標示屬於不同階層的影像區域，作為初始的階層標記。此一初始的階層資訊，將應用在全域相機運動模型的估測中。

2. 估計連續兩張影像間之相機運動

在作相機全域運動估測時，我們使用了 Tan 等人[14][15]所提出的攝影機運動估測模型，此模型基本假設是攝影機運動只能有放大、縮小、上下或左右傾斜攝影，但是攝影機的中心必須固定於定點不可有平移運動，我們簡稱此模型為 PTZ（Pan、Tilt、Zoom）運動參數模型，我們以此模型來近似一般相機運動模型，其數學式為：

$$x' = \frac{p_1 x + p_2 y + p_3}{p_5 x + p_6 y + 1}, \quad y' = \frac{-p_2 x + p_1 y + p_4}{p_5 x + p_6 y + 1}$$

其中 (x, y) 與 (x', y') 分別為相鄰兩張畫面中具有運動對應關係的兩個點之影像座標，而 $p_1 \sim p_6$ 代表著運動模型的六個參數，此模型之推導可以在[14]中得到，值得注意的是座標（0,0）對應到的是影像畫面的中心。

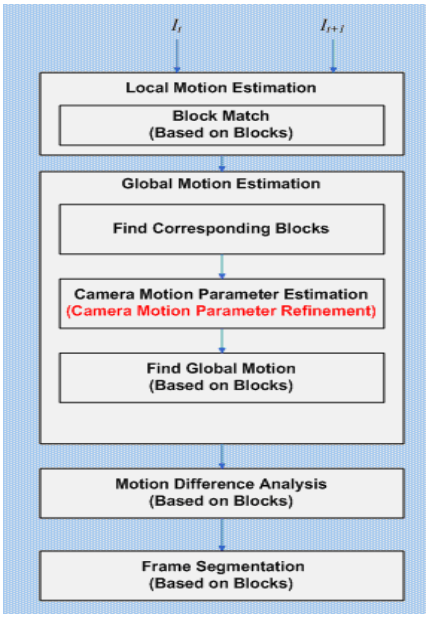


圖5：自動視訊切割架構圖

在估測相機全域運動時，需要用到相對應的(x, y)與(x', y')座標，在我們所提出的方法中，此一對應座標是由影像區塊的區域運動向量所得出，其估計運動模型的架構如圖 5所示。然而由於影像區塊包含了前景與背景部分，因此我們參考了前一張影像已切割出來的階層資訊，選取屬於背景的區塊對應座標，作為初始相機運動模型的估計，而後，在應用如圖 6所示之全域運動估計與偏差參數修正流程，反覆進行更準確的相機運動估計，在此一全域運動估計與偏差參數修正流程中，每次採用估計誤差較小的 35%之對應座標，當作屬於背景的部分，作為相機模型反覆估計的依據，如此可以減少前景對應座標所造成相機運動估計的影響。

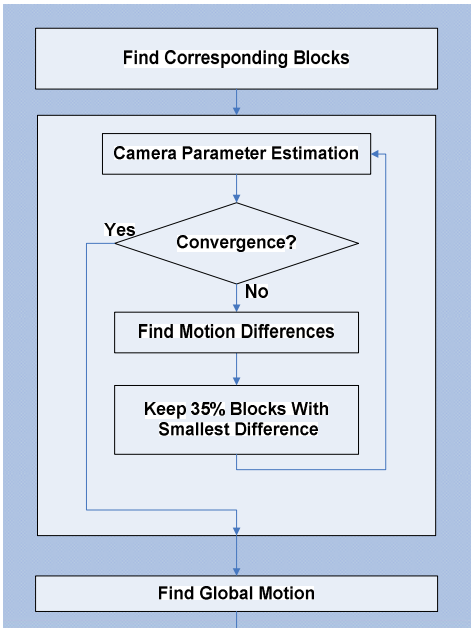


圖6：全域運動估計與偏差參數修正流程

3. 背景階層切割

關於背景切割的部分，我們則是分析影像中，所有區塊的區域運動與全域運動向量差異大小，取臨界值的方式來切割出前景與背景階層，如圖 7所示，我們很容易的可以區分出前景與背景的區塊的分佈，如將臨界值取在 1.5~3.2 之間，我們可以得到如圖 8所示之階層切割結果。

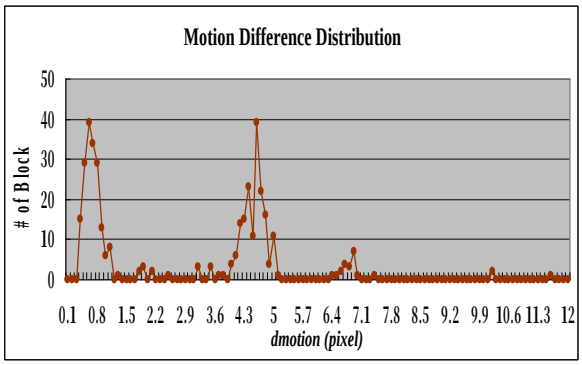


圖7：視訊畫面之區域運動與全域運動差值

(a) 臨界值取在 1.5

(b) 臨界值取在 3.2

圖8：階層切割結果

由圖 7的差值分佈，我們可以觀察到，前景與背景切割的臨界值選取，有一個不小的區間，而不是一個很敏感的範圍，因此我們將在計畫的第二年，朝向以機械學習的方法，自動找出前景與背景分割臨界值的方向進行研究。

四、計畫成果自評

我們以 Mobile 這個視訊影片的階層切割為例（如圖 8所示），當臨界值取在 2 時，共處理了 50 張影像（19800 個 16×16 影像區塊）的視訊階層切割，我們統計出

前景與背景階層切割正確率為 96.07%，而 False Foreground 的比率為 2.17%，False Background 的比率則為 1.76%，此一結果顯示了，以運動特徵作為視訊階層切割之效果顯著，並佐證了行性。

原計畫第一年度預期達成的成果包括：1.友善的初始階層標示介面。2.兩張影像特徵點標定與選取之方法。3.由兩張影像作相機自身運動的估計方法。4.實作以影像比對和貝氏推論為基礎的前景、背景分析系統。5.研究結果將以會議論文之形式作整理與投稿。以上目標的前三項，皆已在第一年度中達成，同時我們也實作出了一個視訊階層分析的雛形系統；而關於第四項的部分，原計畫書內容是規劃以影像疊合比對的方式來切割出視訊階層，經實驗結果分析，採用影像疊合顏色差異的分析方法，不如採用區塊運動模式差異的分析，來得穩定、有效而直接，因此我們在計畫執行時作了修正，實驗結果也顯示了，目前所採用的方法確實可行。對於上述的研究結果，我們目前正在整理撰寫成會議論文，作為投稿發表之用。

五、參考文獻

[1] R. Mech and M. Wollborn, "A Noise Robust Method for Segmentation of Moving Objects in Video Sequences," *Proc. IEEE Int'l Conf. Acoustics, Speech, Signal Processing, ICASSP'97*, Munich, Germany, Apr. 1997, vol. 4, pp. 2657–2660.

[2] R. Castango, T. Ebrahimi, and M. Kunt, "Video Segmentation Based on Multiple Features for Interactive Multimedia Application," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no. 5, Sept. 1998.

[3] S. Y. Chien, S. Y. Ma, and L. G. Chen, "Efficient Moving Object Segmentation Algorithm Using Background Registration Technique," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 7, July 2002.

[4] T. Meier and K. N. Ngan, "Automatic Segmentation of Moving Objects for Video Object Plane Generation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no. 5, Sept. 1998.

[5] J. G. Choi, S. W. Lee, and S. D. Kim, "Spatio-Temporal Video Segmentation Using a Joint Similarity Measure," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, no. 2, April 1997.

[6] R. V. Babu, K. R. Ramakrishnan, and S. H. Srinivasan "Video Object Segmentation: A Compressed Domain Approach," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no 4, pp. 462-474, 2004.

[7] C. Fowlkes, S. Belongie, and J. Malik, "Efficient Spatiotemporal Grouping Using the Nyström Method," *Proc. Int'l Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 231–238, Kauai, Hawaii, 2001.

[8] J. Shi and J. Malik, "Motion Segmentation and Tracking Using Normalized Cuts," *Proc. Sixth IEEE Int'l Conf. Computer Vision*, pp. 1154–1160, Bombay, India, 1998.

[9] B.J. Frey, N. Jojic, and A. Kannan, "Learning Appearance and Transparency Manifolds of Occluded Objects in Layers," *Proc. Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 45–52, Madison, WI, 2003.

[10] N. Jojic and B.J. Frey, "Learning Flexible Sprites in Video Layers," *Proc. Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 199–206, Kauai, Hawaii, 2001.

[11] Q. Ke and T. Kanade, "A Subspace Approach to Layer Extraction," *Proc. Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 255–262, Kauai, Hawaii, 2001.

[12] S. Khan and M. Shah, "Object Based Segmentation of Video Using Color, Motion and Spatial information," *Proc. Conf. Computer Vision and Pattern Recognition*, vol. 2, pp. 746–751, Kauai, Hawaii, 2001.

[13] H. Greenspan, J. Goldberger, and A. Mayer, "A Probabilistic Framework for Spatio-Temporal Video Representation & Indexing," *Proc. Seventh European Conf. Computer Vision*, vol. 4, pp. 461–475, Copenhagen, Denmark, 2002.

[14] Y. P. Tan, S. R. Kulkarni, and P. J. Ramadge, "A New Method for Camera Motion Parameter Estimation," *Proc. IEEE Int'l Conf. Image Processing*, vol. 1, pp. 406–409, Oct. 1995.

[15] Y. P. Tan, D. D. Saur, S. R. Kulkarni, and P. J. Ramadge, "Rapid Estimation of Camera Motion from Compressed Video with Application to Video Annotation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 1, Feb. 2000.