

行政院國家科學委員會補助專題研究計畫成果報告

以物件為基礎之運動估計及編碼之研究

計畫類別： 個別型計畫 整合型計畫

計畫編號：NSC 89-2213-E-009 - 144 -

執行期間：89年8月1日至 90年7月31日

計畫主持人：薛元澤 教授

共同主持人：

本成果報告包括以下應繳交之附件：

赴國外出差或研習心得報告一份

赴大陸地區出差或研習心得報告一份

出席國際學術會議心得報告及發表之論文各一份

國際合作研究計畫國外研究報告書一份

執行單位：交通大學資訊科學系

中華民國 90 年 10 月 20 日

行政院國家科學委員會專題研究計畫成果報告

以物件為基礎之運動估計及編碼之研究

A study on object-based motion estimation and coding

計畫編號：NSC 89-2213-E-009-144-

執行期限：89年8月1日至90年7月31日

主持人：薛元澤 教授 交通大學資訊科學系

計畫參與人員：林柏青 彭紹貞 交通大學資訊科學系

一、中文摘要

為了要能精確的找出視訊畫面中的物件，以便就個別的物件進行運動估計和編碼，在這篇報告中提出了一個依據視訊中物件移動等訊息，自動將物件從視訊畫面中分割出的演算法。這個具記憶性的演算法能參照過去數張畫面的分割結果，並依此來得到一個穩定的分割結果。

關鍵詞：MPEG-4、物件、分割、記憶性

Abstract

To locate objects in the video sequences precisely so that each object can be motion estimated and coded, respectively, this report proposes an algorithm to automatically extract objects based on motion information. This algorithm with memory can refer to segmentation results from past frames from which a stable segmentation can be derived.

Keywords: MPEG-4, object, segmentation, memory

二、緣由與目的

隨著 MPEG-4 在 1999 年定案，可以預期的是，MPEG-4 將如同過去 MPEG-1 及 MPEG-2 一樣，會有重要的應用。但 MPEG-4 與過去 MPEG-1 及 MPEG-2 不同的是，它是一個以物件為基礎的標準，可以針對個別的物件(object)進行編碼及處理。因此我們對於個別物件的編碼及處理感到很有興趣。

其中個別物件的編碼及運動估計，為顧及各項實作成果相容性的問題，這些規格多已有標準化的做法[1]。但我們也發現了如何從視訊影像中產生個別的物件(通常稱之為物件視訊平面 (Video Object Plane, 或簡稱 VOP)來說，這仍是個開放的議題，過去的方法主要分為半自動產生和自動產生兩種方式，這兩者的差別主要在於產生的過程中是否有人為的參與。而其中自動產生的方法由於不需人為的參與，固然在應用上限制較大，但同時也可以減少人力的負擔，而過去也曾有若干學者提出相關的演算法。其中主要的精神在於利用通常有興趣的物件會移動的前提來當作一個重要的分割線索[2-5]，以獲取一個有意義的物件。然而，光是運動的訊息尚不足以精確地劃分出物件的輪廓，所以可再輔以影像處理中空間分割的方式[6]，以得到較佳的效果。

然而就實地的觀察中，我們發現單是用上述的方式所產生的結果很不穩定。一

個物件可能在某些畫面中暫時不移動，或是只有其中部份有移動。就空間分割的部份而言，即使是連續兩張影像分割出來的區域個數及形狀都有相當的差異，因此我們在這篇報告中提出一個演算法來解決上述的這些問題。

Mech 跟 Wellborn[5]提出參考近幾張畫面每個點隸屬 VOP 的情形來減小每張畫面中運動的差異。這個方法雖然簡單而有效，但因為沒有使用空間分割的方式，使得其為了改善物件的輪廓的精確度而必須要做多次後續的影像處理。另外，Kim 等人[6]提出了結合空間分割的方法，但是其所採用的記憶方式需要做分割區域的配對和映射，而空間分割後的區域往往是不規則形狀且前後張畫面區域的大小及個數不一，這使得實作上較為困難，執行時也需耗費較多的時間。

我們提出的方法主要的精神在於保留記憶方法的有效性和簡單性，但仍可以利用到空間分割的優點。在報告的第三部份，我們將介紹產生的方法及結果。

三、方法與結果。

我們的做法是去估計並記憶每一分割區域有運動的機率，並以此記憶協助往後的物件分割。對於每一個區域，我們記憶的方式如下列的式子：

$$MEM(p) \leftarrow$$

$$\begin{cases} p_i & \text{for all } p \in R_i \text{ if } R_i \text{ is in the VOP} \\ MEM(p) * p_i & \text{for all } p \in R_i \text{ if } R_i \text{ is in the background} \end{cases}$$

其中 p_i 所代表的即為區域 R_i 隸屬於 VOP 的機率。另我們可以利用統計學中假定測試的觀念，來判定每張影像中的點是否有發生運動，並以一張稱之為變動偵測遮罩

(Change Detection Mask, 簡稱 CDM) 的二元影像來表示。若用假定測試得到的影像為 CDM_i ，我們把 CDM_i 套至 MEM(即利用記憶的內容)，可得到一個新的 CDM_s ，套用的方法如下：

$$CDM_s(p) = \max\{CDM_i(p), MEM(p)\}, \forall p$$

最後，可以用如下的式子來判定一個區域是否隸屬 VOP，

$$(1) \sum_{p \in R} CDM_s(p) / (\text{number of } p \text{ in } R) > T_1, \text{ or}$$

$$(2) \text{令 } r = \sum_{p \in \partial R} CDM_s(p) / (\text{number of } p \text{ in } \partial R) > T_2$$

，其中 ∂R 代表區域邊界的點，且

$$\sum_{p \in R} CDM_s(p) / (\text{number of } p \text{ in } R) > 1 - r.$$

簡而言之，一個區域中如果有足夠多的點有運動，那麼此區域即被判定為 VOP 的一部份；另外，若邊界中有足夠的點有運動，且區域中的點也有一定程度的運動，那麼也可以將此區域判定為 VOP 的一部份。後者主要是基於一個均勻材質的區域即使有運動，單單去量測區域內各點的差異仍會很小，因而容易被誤判為沒有移動。而區域的邊界因在灰階上通常有頗大的落差，只要有移動就能使連續兩張影像產生較大的差距，可充分反應出有運動的情形，因此一併列入考慮。

估計每個區域隸屬 VOP 的機率 p_i 可由 CDM 的值除以該區域的點數來得到。為求精確，我們可同時將只含區域邊界部份及不考慮先前畫面的 CDM(即 CDM_i) 而除得的值給若干權重取平均來得到最後的機率估計。

為了再提高分割的正確性，我們也設計的如下的機制：

- (一) 若一個區域鄰近顏色相近的點若大部份皆隸屬於 VOP，則此區域亦屬於 VOP。這樣的步驟可反覆若干次，直到沒有新的區域因此被劃為 VOP 為止，即收斂的情形。
- (二) 有些背景區域(非 VOP 的區域)中的若干點有不斷的運動，但因只佔背景區域的一小部份而跟著被判為背景。這種情形下，只要它們(1)最近有成為 VOP 的記錄、(2)最近仍有運動的情形，我們仍然給那些點有機會變成 VOP 的一部份。
- (三) 可以在針對產生的 VOP 之輪廓透過濾波器來修飾，以減少錯誤及使其輪廓更加平滑。我們採用的有(1)長條的濾波器，即若濾波器的兩端皆為黑點(白點)，則濾波器涵蓋的點皆為黑點(白點)。這可減少空間分割時部份突出或凹陷的區域。(2)形態眾數濾波器[7]，即一方形濾波器中，若其涵蓋的點中大部分為黑點(白點)，則其中心區域修改為黑點(白點)。

我們利用 352x288，300 張畫面 CIF 格式的視訊檔進行測試。測試的視訊檔包括常見的 Akiyo 及 Mother & Daughter，其結果如右附圖一、二：

四、結論

儘管在某些特定的影片當中，自動化的分割物件方式可以得到不錯的結果。但對任意影片情形仍有很大的研究空間。因



圖一、Mother & Daughter



圖二、Akiyo

此結合人為的處理仍是個不錯的方向。此外，目前大部份的視訊檔都是以壓縮的形式儲存，其資料量較解壓縮後播放的資料量小很多。因此，如何在壓縮檔中直接進行 VOP 的分割亦可成為未來一個研究的方向。

五、參考文獻

- [1] Touradj Ebrahimi and Caspar Horne, "MPEG-4 Natural Video Coding An Overview," *Signal Processing: Image Communication*, vol. 15, issue 4, pp. 365-385, Jan. 2000.
- [2] P. Salembier and F. Marqués, "Region-based Representations of Image and Video: Segmentation Tools for Multimedia Services," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 8, pp. 1147-1169, Dec. 1999.
- [3] Thomas Meier and King N. Ngan, "Video

Segmentation for Content-Based Coding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 8, pp. 1190-1203, Dec. 1999.

[4] Tai Aach, André Kaup and Rudolf Mester, “Statistical model-based change detection in moving video,” *Signal Processing*, vol. 31, issue 2, pp. 165-180, March 1993.

[5] Roland Mech and Michael Wollborn, “A noise robust method for 2D shape estimation of moving objects in video sequences considering a moving camera,” *Signal Processing*, vol. 66, issue 2, pp. 203-217, April 1998.

[6] Munchurl Kim et al., “A VOP Generation Tool: Automatic Segmentation of Moving Objects in Image Sequences Based on Spatio-Temporal Information,” *IEEE Trans. Circuit Syst. Video Technol.*, vol. 9, no. 8, pp. 1216-1226, Dec. 1999.

[7] Al Bovik, *Handbook of Image & Video Processing*, Academic Press, 2000.