

在棒球視訊中以物件為基礎的視訊檢索與事件偵測

Object-Based Video Retrieval and Event Detection for Baseball Video

杜維昌* 李健銘 張簡大敬

義守大學
資訊工程學系
高雄縣大樹鄉
wcd@isu.edu.tw

Wei-Chang Du Jian-Ming Lee Da-Jing Zhang Jain

Department of Information Engineering
I-Shou University
Kaohsiung County, Taiwan
wcd@isu.edu.tw

Received 20 April 2006; Revised 29 June 2006; Accepted 30 June 2006

摘 要

隨著電腦處理技術與硬體設備的不斷提升，利用數位攝影機拍攝各種場景已不再是那麼遙不可及的事。但由於攝影機只是真實呈現當時場景的影像，缺乏一些高階語意上的資訊，因而仍需依賴人類的認知。為了提升視訊的價值，如何從中自動擷取出一些高階意涵將是必要的過程。在此一研究中，我們以棒球球賽視訊資料為研究標的，針對固定式監視器所拍攝的視訊資料中有效偵測出運動事件。首先，我們考量低解析監視畫面中多移動物件追蹤之情形，利用視訊資料中空間性與時間性特徵作為前景與背景分離的基礎，並結合連續畫面間的匹配與補償，以期可以獲得穩定的追蹤成效。接著，對運動行為與特徵設計相關表達與比對方法，以利於特徵的有效使用。為了進一步擷取高階意涵，開發有關棒球球賽的事件模式，以提供在時間軸上得到較為高階的內涵。

關鍵字：監視視訊、棒球視訊、多物件追蹤、視訊檢索、事件偵測

* 通訊作者

ABSTRACT

In recent few years, the convenience and low cost of obtaining and storing digital video make people eager to have an effective tool for utilizing these data. From the viewpoint of human vision, digital video just keeps track of frames. Since it lacks for semantic information, it is necessary to understand the digital content with human brain. To promote the scope and value of further applications, automatic generation of high-level semantics is necessary. For a general sport video, motion behavior and characteristic of moving objects is an essential feature. In this paper, our goal is to develop an automatic extraction of motion-based feature via a fixed camera. More precisely, we firstly develop the motion-based segmentation and tracking of moving objects. Then, we design the indexing techniques of motion information for further efficient utilization. To demonstrate the effect of extracting motion-based features, we will build a tested platform of motion-based retrieval system. Based on this platform, we will develop an event-based model of baseball video so as to provide high-level captions on the timeline axis.

Keywords: Surveillance video, Baseball video, Object tracing, Video retrieval,
Event detection

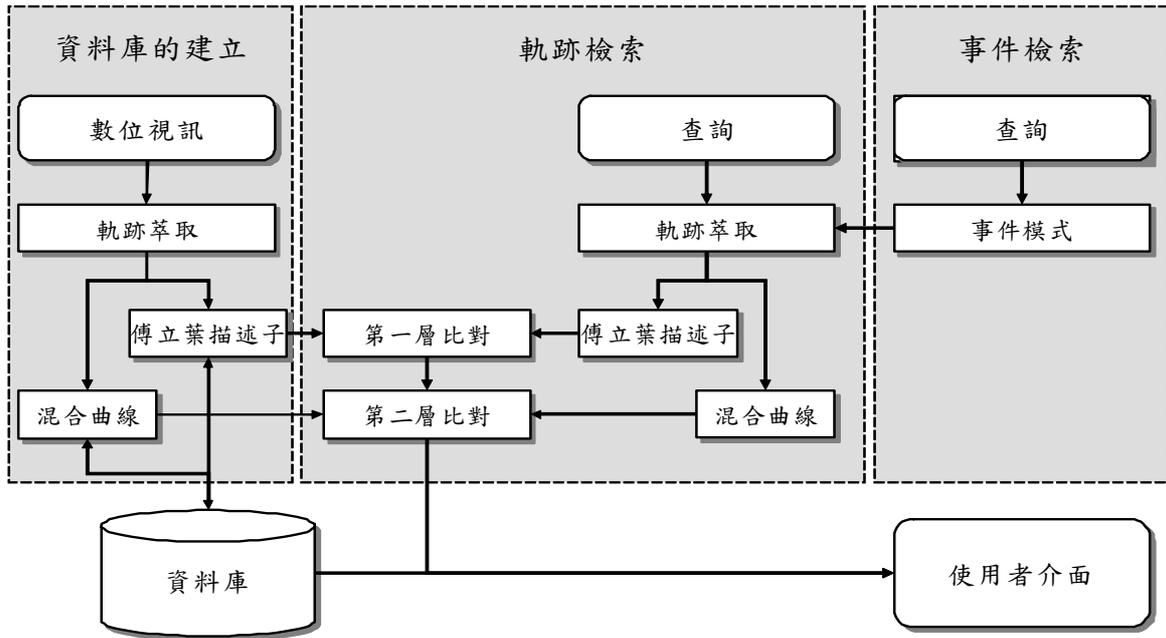
一、簡介

隨著視訊擷取介面和數位攝影機的普及，使得大量的視訊資料得以使用數位方式儲存，因而利用電腦來對視訊資料作自動化管理的構想與需求也因應而生，常見的應用包括監視器系統[23]、隨選視訊[22]、數位圖書館[16]等。一般而言，數位視訊除聲音以外，主要包括一連串的數位影像。以 NTSC 標準為例，視訊是由每秒 30 張畫面所組成，而每張畫面由無數代表顏色的取樣像素所構成。這樣的表示方法僅僅表達視覺效果，因而仍然需要仰賴人類大腦的認知。但在人類思維中，思考的主題經常是較為深沉的語意部分。如何將數位視訊資料轉化為以內涵為基礎的描述方式，於是成為提升電腦智能的重要步驟。自從動畫專家群組制定 MPEG-7 藉以提供影音多媒體內容的描述介面，然而對於描述資料的生成與搜尋技術則急待進一步開發。

由於多媒體視訊資訊用文字的方式不足以充分表達其涵意，取而代之是以內涵為基礎的方式來描述。為了解決此一問題，學者提出各種方法將多媒體資料轉化成以內涵為基礎的描述方式，希望能夠在多媒體資料上自動萃取出符合人類感官的特徵，乃至於高階語意的生成，使電腦對數位多媒體資料的認知能提升至與人腦相近的程度。在影像資料上，其視覺特徵包括了紋理、顏色、外形等，基於視訊資料是由一連串有序的影像所構成，因此視訊資料庫也可以應用影像特徵來建立[28]。另一方面，視訊資料不單只是一張張獨立的影像而已，這些影像之間還具有時間上的相關性，此一性質包括攝影機的運鏡與物件的移動等動態行為。這些動態的特徵對視訊資料而言是相當重要的視覺特徵，同時也是推演出高階語意的重要關鍵，我們可以藉此一動態特徵來對特定類型影片歸納出事件模式。

自從數位內容成為各國未來重點產業，再加上數位攝影機日漸普及，近年來國內外不少研究單位投入視訊內涵的研究。就視訊中的運動內涵來說，主要集中在視訊檢索技術的開發[7] [14] [17] [23]。但由於考量的並非球類運動視訊，假設的運動軌跡與特徵相對來說較為簡易，因而即使並未充分考量高曲率軌跡的速率或方向等物理量，仍可得到不錯的執行成效，但不易推廣到如棒球等之球賽視訊中。在球類視訊上則有一些團隊從事相關的研究，包括在美國 ESPN 頻道播放的 K-Zone 系統[8]。該系統在美國大聯盟球賽中架設固定式攝影機，透過對打擊者與本壘板作定位，以此判別好球袋區域並以繪圖的方式與原始視訊作合成。此外，過去文獻中亦包括以足球賽[20]、網球[21]為主要的運動軌跡偵測與查尋介面，然而並未擴及多球員的追蹤，甚至藉以開發更高階的事件模式，因而尚失了更豐富的球賽內涵。

由於一套視訊檢索系統可涵蓋多個不同層次的議題，而這些議題之間往往存有相依性，因此必須分別從不同因素來加以考量[5] [31]。以視訊資料的動態特徵為例，我們將會面對下列幾個問題：物件動態特徵的表達方式、特徵的比對方法、檢索效率的考量、使用者查詢介面的制定[28]。對於上述問題，在此一研究中，我們將以固定視點之棒球視訊為對象，提出通用性的雙階層比對架構，以尋求在檢索效率與擷取成效間取得較佳的平衡點。在查詢介面上，我們以手繪軌跡作為基本輸入方式，並在其上架構事件



圖一：系統架構圖

式查詢方法，讓使用者能選擇系統管理者預先建立好的事件定義，或是以自行設計的事件來作查詢。基於上述檢索平台，我們將以棒球視訊為例，建立主要的事件模式，並測試相關事件擷取的實際成效，圖一為整體系統架構圖。

二、動態視訊物件分離技術

隨著電腦處理技術提升與網際網路的成熟，多媒體相關應用也隨之蓬勃發展。特別在最近幾年，開發數位視訊高階內涵是目前學術界或產業界重要研究議題之一，而其中移動物件的偵測與分離更是當中重要的關鍵步驟之一，經常運用在視訊監控系統 (video monitor surveillance)[20]、智慧型公路系統 (intelligent-highway system)、入侵監視 (intrusion surveillance)[11] 等系統中。然而面對複雜的場景，一個即時的自動化視訊監控系統可以對移動物件的行為或多物件間的相互關係作有效的描述[1] [3]。以運動視訊 (sport video) 為例，擷取球員間的運動關係有助於事件行為的自動取得。另外，在公路監視系統中，記錄車輛的行進軌跡則有助於自動違規偵測、駕駛行為統計等工作。

一般而言，我們可以利用下列兩項視訊特徵來協助移動物件的分離與偵測工作：第一、空間性視覺特徵：我們可利用前景物件與背景影像在色調上的不連續性，以及物件內部紋理[20] 與顏色[7] 的相依性質，藉此對影像作區域性分割[10]。第二、時間性運動特徵：此類是依據物件在時間軸上變動的位移資訊為基礎[11]，來尋找出前景物件所在的區域。由於兩相鄰畫面間格很短，因而利用兩畫面間的差異是偵測移動物件的重要途徑。

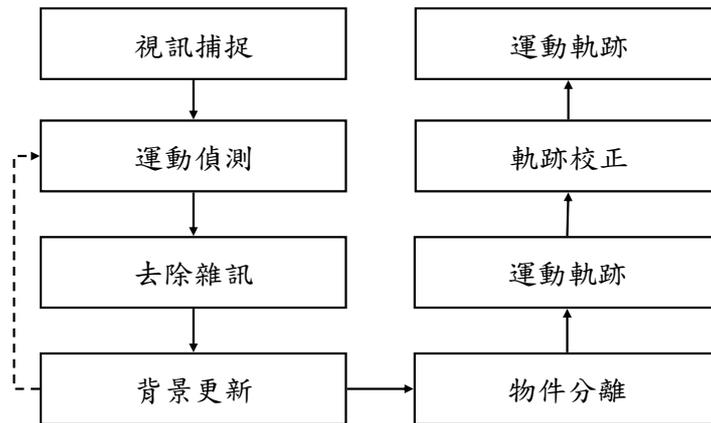
數位視訊資料是由一連串的畫面所組成，播放時為達到人類視覺的順暢度，兩相

鄰畫面被拍攝時間僅相差二十至三十分之一秒，因而差異並不會太大。在視訊資料中，有關運動分離的方法主要包括下列四個主要策略：第一、以顏色為基礎的運動分離技術 [11] [14]：以像素為基礎的分離技術只考慮前後畫面中單一像素在時間軸上的變化，然而由於沒有考慮鄰近像素的變化情況，所以可能會遭遇一些問題。舉例來說：偵測的移動物體內部常有坑洞產生，因而導至無法準確描述一個物件的完整性。此外，整張影像會產生較多的雜訊，在此一情況下，雜訊可能被當成物件，或者確實是物件的部分卻沒被偵測出來。第二、差分影像之運動分離技術 [4] [13]：假設攝影機的位置與角度為固定不動，那麼使用差分影像是典型方法之一。由差分影像中高對比像差的群聚情況來判斷物件的所在位置。第三、以樣板為基礎的運動分離技術 [2] [16]：利用選定的模型套用在時間軸上的影像加以比對，以偵測出與模型相似的區域以此作分離，適合用於一些具有固定結構的物體上。第四、以區域為基礎的運動分離技術 [7]：區域型運動分離技術主要利用影像內色調上的相似程度對影像分割成多個區域。接著，利用相鄰區域的運動向量分佈，進一步合併成單一移動物件。

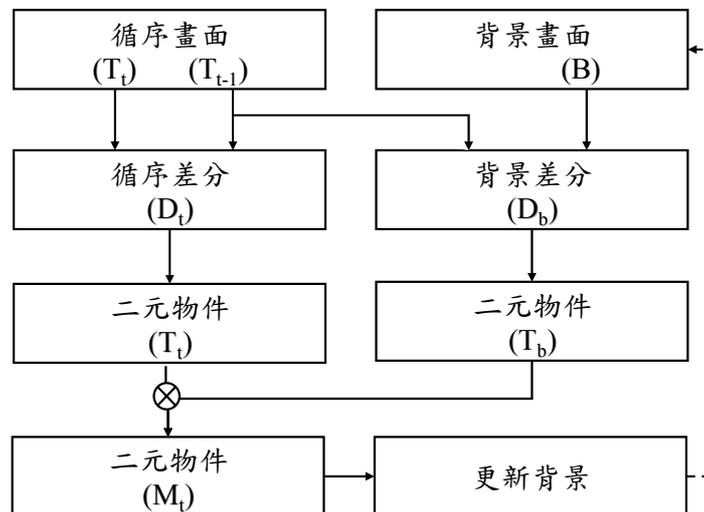
表一：背景分離與循序差分之分離特性

比較項目	背景差分	循序差分
定義	$\Delta T_b = T_t - B$	$\Delta T_f = T_t - T_{t-1}$
背景影像	需要	不需要
偵測物件	較完整	易破碎
雜訊干擾	較多雜訊	較少雜訊
區域描述	停止或行進間物件	行進間物件

在數位視訊中，前景物件可利用不同視訊特徵從背景影像分離出來，主要分成空間性特徵和時間性特徵兩類。在此，我們希望開發具有通用性的偵測與分離技術，在色調上不作任何假設。在實際應用中，物件顏色分佈未必集中，因而單獨依賴物件在空間上的特徵，不足以有效作分離，因此仍需仰賴時間性特徵來協助偵測。一般而言，使用背景差分分離方法由於背景影像無法完全立即反應現行畫面的背景內容，因而偵測的移動物件雖然較為完整，但在背景上容易有較多的雜訊產生。另一方面，使用循序差分因相鄰畫面在色調上較為一致，因而背景出現的雜訊相對較少。但由於物件內紋理或顏色可能相似，因而使用循序差分在物件內部可能產生較多的破洞。在此一研究中，我們將綜合循序差分影像與背景影像差分分離技術的特性，如表一所示。為了達成物件偵測的正確性，採用背景差分分離對物件偵測上較能保有物件的完整性，並且搭配循序差分分離對外環境具有較小雜訊干擾的特性，依此原則將兩者加以整合，將兩者分別選定一適當的濾波器，並搭配一自動背景更新技術，以降低物件偵測的破碎及雜訊干擾的缺失。我們針對此概念設計一個物件追蹤系統，來有效分離出一個或多個移動物件，系統架構如圖二所示。簡略來說，系統架構工作可分成「運動偵測」、「物件分離」與「背景影像更新」，分別說明如下：



圖二：系統架構



圖三：運動偵測流程

2.1 運動偵測(Motion Detection)

對於動態視訊物件分離技術而言，物件運動偵測是首先必須面對的問題，過程中我們希望能從視訊中偵測出顏色變動較大的像素，以作為下一階段物件分離的基礎。對此，我們根據影像差分的概念，依物體運行於畫面中所產生之變動區域為基準，並搭配循序差分影像與背景影像差分分離技術的特性將物件偵測出來。對於運動偵測而言，處理流程如圖三所示，說明如下：首先架設固定的攝影機，拍得的一段視訊畫面 (T_1, T_2, \dots, T_n) ，根據兩張連續性畫面資料 T_{t-1} 和 T_t 與一張背景影像 B ，求得兩者間的差分影像，分別為循序畫面差分影像(frame difference)，亦即 $D_t = T_t - T_{t-1}$ ，與背景差分影像(background image) $D_b = T_t - B$ 。接著，使用一個適當的門檻值來當臨界的條件，大於或等於門檻值者表示在該點像素值變動較為劇烈，因此將其視為移動物件資訊；小於門檻

值者則表示在該點上像素值變動較不劇烈，因此將其視為背景資訊。接著，分別將兩張差分影像 D_t 和 D_b 設計雜訊濾波器，先去除一些雜訊與填補影像所產生的坑洞情況，設計方法如下：

觀察一、由於相鄰兩畫面的時間間距很短，循序差分影像主要偵測物件邊緣與內部明顯紋理，因而容易造成物件破損的問題發生，在步驟中可使用 Dilation 運算來對坑洞作修補。

觀察二、由於拍攝視訊所得畫面並不穩定，可能造成背景差分影像變異過大因而產生較多的雜訊，為了保持物體之完整性並去除雜訊干擾，故選擇中間值濾波器來作雜訊的去除。

廣義來說，我們可將二元物件區分為移動與靜止兩類。若是 $O \in D_b$ 且 $O \notin D_t$ 則可判別為靜止物件；如果 $O \in D_t$ 且 $O \in D_b$ ，則可判別為移動物件。因此，為了偵測移動物體的範圍，將二元影像 D_t 和 D_b 取其交集，亦即 $M_t = D_t \cap D_b$ ，以獲得一張代表移動物件分佈的二元影像。我們基於下列觀察，使用形態學運算來改善 M_t 上物件的完整。為了改善 M_t 上部分雜點及破洞，可採用 Opening 運算來對影像作進一步處理。

2.2 物件分離(Object Segmentation)

根據上節物件偵測流程，可獲得一張物件之二元影像 M_t ，對此影像透過 Opening 運算以除去雜訊干擾，並對破碎物件作填補之動作。接著將修正後影像 M_t' 根據區域成長(Region Growing)演算法[27]，將相連像素作一群組，定義個別區域之像素總數為 $\#M_t'(R_i)$ ， R_i 表示第 i 個區域。然後使用一個適當的門檻值 ε 來當臨界的條件，若是像素總數大於或等於門檻值者則表示在區域相鄰像素值均變動較為劇烈，將此區域判定為物件；反之小於門檻值者則將此區域視為雜訊，有關區域分離定義如下：

$$Rg(R_i) = \begin{cases} \text{object,} & \text{if } \#M_t'(R_i) \geq \varepsilon \\ \text{noise,} & \text{otherwise} \end{cases}$$

依據上述判別準則即可分離出符合之物件區域，接著進而估算個別物件的所在位置，求得個別物件的軌跡特徵，因此我們將可依時間順序求得一連串的重心位置。

2.3 背景影像動更新(Background Updating)

由於外在環境雜訊干擾往往過大，為了增進移動物件偵測的準確性，故我們進一步考量對背景影像作適時更新，以降低外在環境雜訊的干擾。在此，採用的方法如下：首先定義一可調式的背景參數 k ，接著利用目前畫面 T_t 前 k 張影像求出一張參考畫面 F_k 。畫面 F_k 中的像素值是根據前 k 張影像在該位置上像素值序列 $\{T_t(x, y), T_{t-1}(x, y), \dots, T_{t-k}(x, y)\}$ 之中位數(median)來代表該點位置像素 $F_k(x, y)$ 。根據下列式子：

$$B'(x, y) = \begin{cases} B(x, y), & \text{if } |F_k(x, y) - T_{t-1}(x, y)| \geq \varepsilon \\ T_t(x, y), & \text{otherwise} \end{cases}$$

將求得的參考畫面 F_k 跟前一張畫面 T_{t-1} 相減，求得兩者間的差分影像。接著透過一個適當門檻值 ε 來當臨界條件。若差分影像在該點位置上變動劇烈，我們判別它可能是處於物件所在區域。因此，將目前背景所在座標像素值來取代更新的背景畫面。至於變動平緩的位置，猜測可能為背景的部分。因此將目前畫面的該點位置像素值 $T_t(x, y)$ 來取代 $B(x, y)$ 上的像素值，結果可得到更新後的背景影像 B' 。

在求得參考畫面 F_k 中，參數 k 的選擇上具有一些特性。在背景更新過程中，若是參數 k 選擇過小，會使得物件造成破洞的情況產生；若是 k 的選擇過大，由於隨著 k 值的增加，畫面差異度也隨著增加。差異度改變會使得雜訊有隨著 k 增加而有相對增加的趨勢。因此，在變數 k 的選擇過程中，變數 k 太小或太大都是不恰當的。故選擇一個適當的參數值不僅減少雜訊，尚可降低物件破碎的情形發生，以提高移動物體偵測的精確度。

三、多移動物件追蹤技術

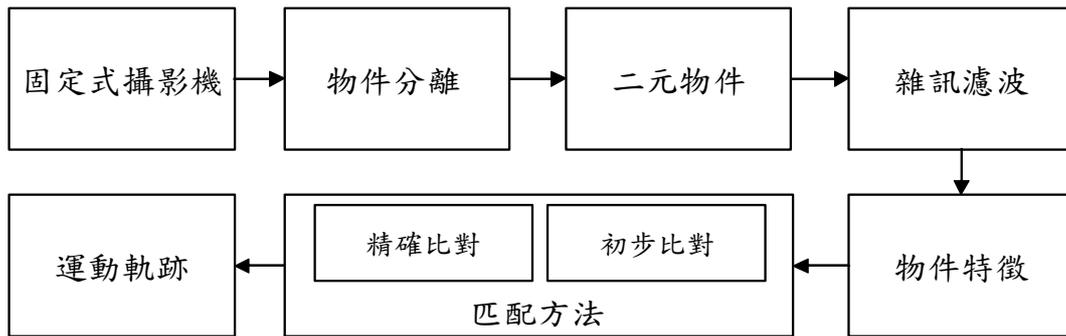
對於物件運動軌跡的建立，我們首先利用上述視訊物件分離技術萃取出個別物件之特徵，接著依物件在時間軸上移動的資訊進一步追蹤其運動軌跡。然而在追蹤過程中，兩個或兩個以上的移動物件可能距離過近或彼此產生遮擋之現象。對此，我們將於本節中作一討論，並且對個別物件之遮擋現象進一步描述與說明。

3.1 多物件匹配

對於多移動物件運動軌跡，有關此偵測流程如圖四所示。物件分離步驟同單一移動物體，可個別求得物件之移動特徵，其特徵包含物件之重心座標 P_p 、區域像素個數 P_s 與色彩強度 P_i 等。依此我們可以獲得兩相鄰畫面物件 O_u 與 O_v 間的相異程度，定義如下：

$$P_{o_u, o_v} = w_p \times \Delta P_p + w_s \times \Delta P_s + w_i \times \Delta P_i$$

其中 w_p 、 w_s 與 w_i 為權重值， ΔP_p 、 ΔP_s 、 ΔP_i 分別為兩物件位置、區域大小、色彩強度之特徵空間歐基里德距離(Euclidean distance)。因此我們可以獲得所有匹配的相似值 P_{o_u, o_v} ，接著將 P_{o_u, o_v} 透過一個適當門檻值 ε 來當臨界條件，小於等於門檻值者則表示在該兩個物件相似程度高，表示可能為同一物件，接著找出與物件相對應的物件編號 u 與 v ，並且紀錄該物件符合之物件總數。最後依據符合總數及對應之編號可獲得一連串之循序編號，故可求得個別之物件移動軌跡。



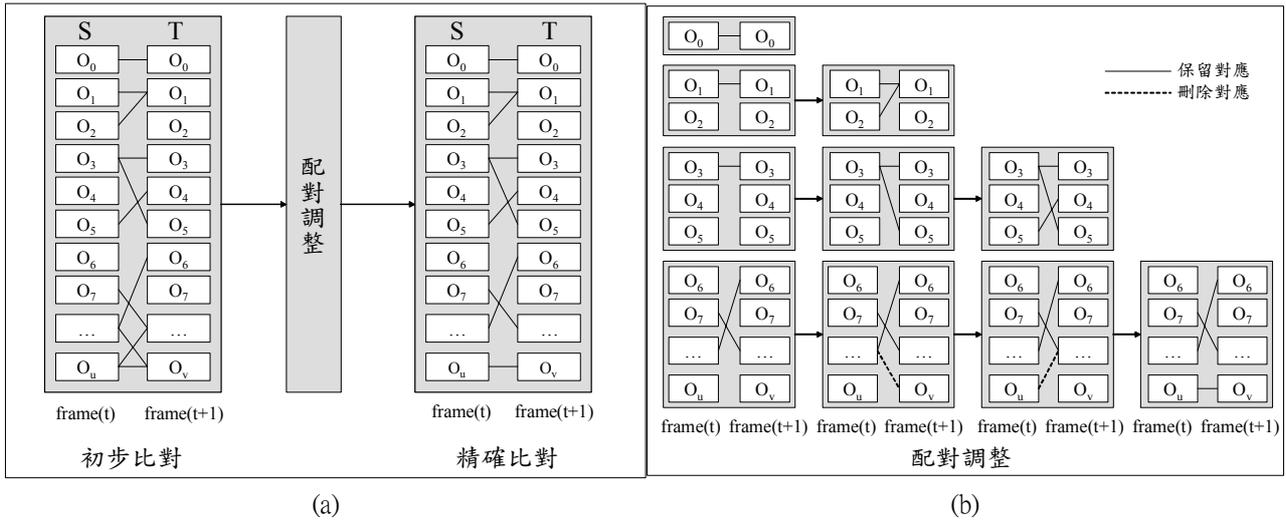
圖四：多物件軌跡偵測流程

3.2 多物件對應關係

在比對過程中，為了適當降低物件比對的不明確，希望透過一雙階層的演算法則來對先前物件集合 S 及目前物件集合 T 作適切的匹配。首先依據擷取出的特徵為基礎，透過初步匹配方式 (*crude matching*) 以求得兩物件 $O_u \in S$ 與 $O_v \in T$ 匹配的相異值 $Dist(S[u], T[v])$ 。接著，依相異程度選擇一個適當門檻值來當臨界條件，小於等於門檻值者則表示該兩個物件相似程度高，表示可能為同一物件。依此找出物件 O_u 所符合之所有相對應物件編號 v 及紀錄符合之物件總數 $DegS[u]$ ，同理可得物件 O_v 可求得符合個數 $DegT[v]$ 。於是依據 $DegS[u]$ 與 $DegT[v]$ 可將其歸類成多種類型，分別代表為分裂(一對多)、合併(多對一)，匹配(一對一)、靜止物件、生成物件或同時為合併與分裂等類別，如表二所示。

表二：比對類型

	DegT[v]=0	DegT[v]=1	DegT[v]≥2
DegS[u]=0	靜止或生成	—	—
DegS[u]=1	—	一對一對應	合併
DegS[u]≥2	—	分裂	分裂或合併



圖五：比對流程結果：(a)比對流程，(b)比對步驟

爲了減少過多的分裂或不必要之合併等缺失，我們根據上節中初步匹配方式(*crude matching*)，將滿足條件的所有配對(或稱對應)表爲 $E \subset S \times T$ 。接著，將集合 E 中每一元素依距離由大至小排序，並依序套用精確匹配(*refined matching*)的演算法則。利用 $DegS[u]=1$ 或 $DegT[v]=1$ 優先準則，來對表二中 $DegS[u] \geq 2$ 且 $DegT[v] \geq 2$ 的類型，亦爲同時具有合併且分裂的模糊情況，適時修正成分裂或合併狀態，以保持其物件比對之穩定性。有關比對流程如圖五所示，圖中由上至下可分成四類，分別爲匹配(一對一)、合併(多對一)、分裂(一對多)和同時爲合併與分裂，接著依序套用演算法則來判別，符合優先準則者將其匹配，亦爲物件符合總數 $DegS[u]=1$ 或 $DegT[v]=1$ ，反之則代表物件具有模糊狀態，故將此匹配去除並適時修正物件之符合總數，亦修正爲 $DegS[u]-1$ 與 $DegT[v]-1$ 。依此準則最後將符合者優先挑選並將匹配後之結果儲放於集合 E' 中，有關匹配演算方法如下所示：

Algorithm Matching(S,T)

```

E=E'=∅;
DegS[1..Leng(S)]=DegT[1..Leng(T)]=0;
for u=1 to Leng(S) do
    for v=1 to Leng(T) do
        if Dist(S[u],T[v])≤ε then
            E=E∪{(u,v)};
            DegS[u]++;
            DegT[v]++;
        end if
    end for v
end for u
while E≠∅ do
    (u,v)=Pick the edge with maximum distance;
    E=E-{(u,v)};
    if (DegS[u]=1 or DegT[v]=1) then
        E'=E'∪{(u,v)};
    else

```

```

        DegS[u]--; DegT[v]--;
    end if
end while
return E';
end Algorithm

```

3.3 多物件運動軌跡的生成

對於多物件遮擋問題，在物件比對過程中由於比對結果不同於單一移動物件，物件符合數不完全為一對一的關係，亦即為 $DegS[u] \neq 1$ 或 $DegT[v] \neq 1$ 。因此物件比對應該對其所處狀態分別適時作修正，根據雙階層配對法則修正後即可獲得個別物件之修正軌跡。

對於物件軌跡修正而言，由於透過差分影像分離技術之概念，我們可依據差分影像中像差的群聚情況來判斷出該物件的所在位置，然而在偵測過程中對於物體本身並非永久保持固定的物件數或持續變動的特性，故在偵測過程中物件可能會有靜止、消失或生成的情況發生。因此我們根據表二及一個適當門檻值來加以區別，若是 $DegT[v] = 0$ 表示其為新增的物件， $DegS[u] = 1$ 及 $DegT[v] = 1$ 表示其比對為一對一，若是 $DegS[u] = 0$ 時表示物件可能處於停止或者消失的狀態，於是紀錄物件 O_u 停留時間 $T(O_u)$ ， u 表示物件編號。接著我們利用一適當的門檻值 ε 作為限制條件來加以區別，若停留時間大於等於門檻值則表示該物件可能為停止或者離開偵測視窗，其值為 $T(O_u) = 0$ ；小於門檻值則表示物件為非靜止狀態，將時間紀錄累加，亦即為 $T(O_u) = T(O_u) + 1$ 。故我們根據 $T(O_u)$ 可分辨出該物件目前是否處於停止或消失狀態，進而求得移動軌跡，對於 $T(O_u)$ 定義如下：

$$T(O_u) = \begin{cases} 0, & \text{if } T(O_u) \geq \varepsilon \\ T(O_u) + 1, & \text{otherwise} \end{cases}$$

對於分裂物件軌跡修正而言，首先以差分影像分離技術為基礎，以獲得該物件之軌跡特徵，接著透過一特徵比對演算法處理，先求得該物件 O_u 符合個數 $DegS[u]$ 及相匹配物件 O_v 的個數 $DegT[v]$ ，然而在偵測過程階段中，我們並不能得知該物體目前是否已為合併之物件，但隨著時間變化透過循序比對的方式，若該物件符合數 $DegS[u] \geq 2$ 時，表示該物體符合數不唯一即物體發生分裂的狀況。從圖中得知該物件於畫面四比對結果 $DegS[u] = 2$ 表示於下階段中此物件將分裂出 $DegS[u] - 1$ 個物件。因此我們應新增 $DegS[u] - 1$ 條新的軌跡，最後再分別依序記錄其分別的軌跡特徵。對於合併物件軌跡修正而言，若目前物件及相對應物件之符合物件數為 $DegS[u] \geq 2$ 及 $DegT[v] = 1$ 狀況時，表示物體之間可能發生合併之狀況。從圖中得知該物件於畫面三時該比對結果為 $DegT[v] = 2$ ，表示於下階段時該物件將與 $DegS[u] - 1$ 個物件合併。因此我們將合併物件加上一個標記，用來表示為合併的物件並且記錄該合併物件之所有滿足匹配之所有物件編號 u 。接著依序比對，若是比對過程中該合併物件 $DegS[u]$ 仍保持為 1 時，表示該物件處於合併狀態並未分離。於是依序將目前匹配之軌跡特徵儲存於合併前個別符合之軌跡序列中，最後即可獲得個別之物件軌跡。

四、運動特徵與檢索方法

視訊物件軌跡的建立，是依據時間軸上連續畫面中的物件，取出一連串的重心座標而成，所得到的軌跡具有位置與時間特徵。在多數的影片格式中，畫面與畫面之間的時間間隔通常是固定的，因此我們常會省略時間軸的表示，而只以 XY 座標平面來表示物件軌跡。假設一物件在第 i 張畫面的座標為 P_i ，若一段視訊有 n 張畫面，則該物件的運動軌跡為 $T = \{P_0, P_1, \dots, P_{n-1}\}$ ，此座標序列又可稱為數位化曲線(digitized curve)。由於軌跡除了位置特徵外，還有方向、速率等其它特徵。在過去文獻中，速率特徵經常被忽略或是獨立於軌跡特徵之外的形式來表示[7]。為了充分運用速率特徵，並與時間上相互對應的位置特徵作緊密結合，我們提出「混合特徵曲線」作為軌跡表示方式。所謂「混合特徵曲線」即是將原本的二維曲線結合其它特徵形成高維度的曲線，那麼物件的軌跡特徵將與其它特徵建立起更緊密的相關性。在本研究中，我們將速率特徵與原來的二維數位化曲線結合而成三維的軌跡曲線，並將其稱為 H-Curve (hybrid curve)。

一般情況下，視訊物件的移動速率並非太快，如此一來在高畫格率的影片下所擷取出的軌跡曲線經常是由高密度點所構成，如果能對曲線作適當的簡化，只保留較重要的特徵點，將可在比對中降低運算量。重複疊代法(Iterative Refinement Method, 簡稱 IRM)[4] 為一多邊形近似法，其作用可以儘量保留高曲率的像素點，簡化後的曲線外形可用來近似原數位曲線。由於計算簡單，同時也容易應用於高維度的曲線上，因此我們利用 IRM 來簡化「混合特徵曲線」。對於一段高維度數位曲線，IRM 的運算步驟如下：

1. 設曲線的首尾兩個點為 P_0 與 P_{n-1} 。
2. 在曲線上所有的點當中，找出距離 $\overline{P_0 P_{n-1}}$ 最遠的點，若其距離大於門檻值 ε ，則記錄該點為近似後的頂點，並設其為 P_s ；若是小於門檻值則停止。
3. 將 P_0 至 P_s 與 P_s 至 P_{n-1} 兩段子曲線分別視為新的曲線，再由步驟一作起。

當運算結束後，首尾兩點與每次疊代中所記錄下來的所有 P_s 點即為簡化後的曲線。步驟 2 中所使用的門檻值 ε 為控制近似程度的參數， ε 設定越大則留下的特徵點數越少，但外形失真度也就越大。IRM 主要應用於外形的近似，並沒有考慮到物件軌跡的時間性質，例如一物體沿著一條直線來回移動，這種情況下 IRM 會將一些重要的軌跡資訊剔除。但一般而言，視訊影片中的物件多屬於自然物體，在這種極端的轉折處，其物體的移動速率通常也會隨著變化，因此加入了速率特徵的 H-Curve 正好可以讓 IRM 偵測出速率變化大的點，因而得以保留原曲線的外形。

在這一節裡，我們將利用物件軌跡來建立軌跡檢索系統。為了在保有精確性的前提下提升比對效率，我們使用雙階層的比對架構。第一層使用傅立葉描述子(Fourier descriptors, 簡稱 FD)的比對方法，以利於快速篩選出候選資料。經由傅立葉轉換所得出的係數中，低頻係數項即代表曲線的大致外形，因此只需比對少數的幾個低頻係數，便能初步判定該軌跡曲線是否吻合，而且每次比對的資料為固定維度的特徵值，所以在資料檢索上可利用高維度索引技巧以得到高效率的搜尋成效。第二層則是以較精確的方式

計算出資料庫軌跡的排名，使用動態規劃法(dynamic programming)計算軌跡與軌跡間在空間上的歐基里德距離(Euclidean distance)，以作為衡量軌跡相似與否的主要依據。

4.1 傅立葉描述子的索引與比對

傅立葉描述子是對一整段軌跡曲線作描述，因而將軌跡曲線作適當分段將有助於兩曲線間的有效比對。首先，定義軌跡分段的規則。基於考量使用者在查詢上的習慣與比對效率問題，所以訂定的規則必須依據高階的意涵以符合事件描述的完整性。在觀察影片中物體的運動行為後可發現，當物體停止移動時，通常為一基本事件的中斷點，因此我們定義若某一段軌跡的速度接近零且持續一小段時間，則可將此一小段軌跡切除，其中速度與持續時間的門檻值應配合不同影片的物件運動特性而作適當的調整。

切割後的軌跡片段可經傅立葉轉換得到頻率係數，但軌跡片段的取樣點數目往往不會一樣。不同數目的資料，在經過傅立葉轉換後得出來的係數也會不同而難以進行比對，所以必須先將所有曲線片段重新取樣成相同點數才行。常見的重新取樣方式，有「等比例時間取樣」及「等比例線段長度取樣」。由於等比例時間取樣會受到物件移動速度的影響，使得取樣後的資料會帶有速度特徵。在考慮到檢索時速度特徵是選擇性條件，因此我們使用等比例線段長度取樣法。假設一曲線片段的長度為 L ，共需取樣 N 個點，則將在曲線上每隔 $L/(N-1)$ 的長度取樣一次。

可應用於傅立葉轉換的外形特徵有複數座標(complex coordinates)、質心距離(centroid distance)、曲率(curvature)及累加角度(cumulative angular)等特徵，其中以質心距離在轉換成 FD 後擁有較強韌的比對效果[30]。質心距離的定義如下：設 (x_c, y_c) 為軌跡的質心座標， $(x(t), y(t))$ 為軌跡的第 t 個座標，則每個座標點的質心距離 $r(t)$ 為

$$r(t) = \sqrt{(x(t) - x_c)^2 + (y(t) - y_c)^2}$$

接著，便可將外形特徵 $r(t)$ 代入離散傅立葉轉換：

$$FD_k = \frac{1}{N} \sum_{t=0}^{N-1} r(t) e^{-i2\pi kt/N}, k = 0, 1, \dots, N-1$$

其中 FD_k 即為傅立葉描述子。由於質心距離特徵是實數值，因此 FD_{N-k} 與 FD_k 是共軛複數，所以只需記錄 FD 的前 $N/2 + 1$ 項。另外，考慮到我們將提供位移與縮放不變量的比對，故必須將 FD 轉成這些不變量。求得方式是對 FD 取絕對值，縮放不變量則可藉由將交流係數除以直流係數而得到：

$$FD_{inv} = \left(\frac{|FD_1|}{|FD_0|}, \frac{|FD_2|}{|FD_0|}, \dots, \frac{|FD_{N/2}|}{|FD_0|} \right)$$

此時， FD_{inv} 即為一組具有位移、縮放不變量特徵。最後，假設兩軌跡曲線的 FD_{inv} 分別為 $(F_Q^0, F_Q^1, \dots, F_Q^{N-1})$ 與 $(F_S^0, F_S^1, \dots, F_S^{N-1})$ ，並且只取前 M 項 FD_{inv} 作比對，那麼我們可定義兩曲線差異度為



圖六：使用傅立葉描述子之比對流程

$$D_{FD} = \sqrt{\sum_{i=0}^M (F_Q^i - F_S^i)^2}, \quad M \leq N-1$$

此乃因為低頻係數項代表了曲線的大致外形，圖六為比對流程。

4.2 混合特徵曲線相似度比對

在特徵相似度比對上，我們以「混合特徵曲線」在空間上的歐基里德距離作為差異度計算方式，假設使用者查詢曲線為 $Q = (P_0^Q, P_1^Q, \dots, P_{m-1}^Q)$ ，資料庫中某段曲線為 $S = (P_0^S, P_1^S, \dots, P_{n-1}^S)$ ，則其差異度可表為

$$D(Q, S) = \max \left\{ \max_{i=0}^{m-1} \min_{j=0}^{n-1} d(P_i^Q, P_j^S), \max_{j=0}^{n-1} \min_{i=0}^{m-1} d(P_i^Q, P_j^S) \right\}$$

其中 $d(P_i^Q, P_j^S)$ 代表 P_i^Q 與 P_j^S 的歐基里德距離。運動軌跡曲線是有序的點集合，上述的差異度計算方式並沒有考慮到匹配的前後次序，如此一來軌跡的比對將與單純的圖形比對無異。例如兩物件所經過的路徑相同，但位移方向卻是完全相反的兩軌跡，在此一比對方法下將分辨不出這種差異，因此必須進一步限制曲線的比對順序。此外，每段視訊資料的特徵數目不一定相等，況且數目相等也不代表曲線 Q 的第 i 點就一定對應到 S 的第 i 點。若要在限制比對順序下套用上述差異度的計算式，則必須要列出兩軌跡的特徵點所有可能的配對，並從這些配對中找出最佳的配對結果。若以一般列舉方式求解將會花費指數次方的計算時間，而動態規劃法則可有效解決此一問題[14] [28]。假設 $DP(Q, S)$ 為最佳解，那麼

$$DP(Q, S) = w(m-1, n-1)$$

其中

$$w(i, j) = \min \left\{ \begin{array}{l} \max \{w(i, j-1), d(P_i^Q, P_j^S)\} \\ \max \{w(i-1, j), d(P_i^Q, P_j^S)\} \\ \max \{w(i-1, j-1), d(P_i^Q, P_j^S)\} \end{array} \right\}, \text{ where } w(k, l) = 0 \text{ if } k < 0 \text{ or } l < 0.$$

運算方式是在 $m \times n$ 晶格點上找出由 $(0,0)$ 走到 $(m-1, n-1)$ 的路徑，且其路徑花費(path cost)要最小。晶格上每個節點 (i, j) 擁有節點花費(node cost)與路徑花費兩個值，其中節



圖七：軌跡曲線處理流程圖

點花費值定為 P_i^O 與 P_j^S 之間的幾何距離，即 $d(P_i^O, P_j^S)$ ；路徑花費則儲存由 $(0,0)$ 走到 (i, j) 的最小花費值，即 $w(i, j)$ ，同時每個節點也都會有一額外的欄位記錄此最小值是由哪個路徑得來。因此表格只要按順序由左至右、由上而下，一直到 $w(m-1, n-1)$ 便可完成計算，最後節點 $(m-1, n-1)$ 的 $w(m-1, n-1)$ 值即為兩軌跡的最佳配對之路徑花費，也就是我們所求的兩軌跡之間的歐基里德距離，其時間複雜度為 $O(mn)$ 。若自節點 $(m-1, n-1)$ 的路徑來源記錄追溯回去，則可得到配對的內容，圖七即為本節所述之混合特徵曲線比對的流程。

五、事件偵測與檢索

在前一節中，我們介紹了雙階層軌跡比對方法，但在實際的應用上，使用者更希望能以高階語意來與電腦溝通，並用以查詢出具有事件意涵的視訊片段。以運動視訊為例，只有少數人會去查詢某球員從畫面上方跑到下方這類單調的畫面，但對於絕大多數的一般使用者而言，比賽中的安打、得分等精彩畫面才是最感興趣的部分。

5.1 單一物件與多物件的事件檢索

視訊的事件有可能只是單一個物件的運動行為，也有可能是多個物件的互動所構成。在上一節所敘述的軌跡比對系統中，顯然可以滿足單一物件運動行為的查詢，我們只需在原有的主系統上，再加上一層作為輸入關鍵字的事件式查詢架構。此子架構對於管理人員或是進階使用者來說，他們可以為一般使用者預先定義每一事件所代表的軌跡關係模型，而一般使用者在查詢時只需簡單選擇檢索的關鍵字，那麼此子架構可將輸入轉換成對應的軌跡，輸入至主系統查詢，藉以達到單一物件的事件檢索方式。

由多個物件的互動所構成的事件上，首先是擴充上述的事件查詢架構，使其能管理包含多個軌跡的查詢，接下來還必須能判別出多個物件間互動關係的差異。每個物件本身皆具有相對於自己的空間與時間關係，而物件間的互動需要有物件間的空間與時間關係。由於軌跡檢索就已含有物件的空間關係，因此我們最後為這層架構加上時間關係的比對功能，那麼即可具備多個物件事件的查詢功能。

5.2 事件檢索系統設計

時間關係的比對上，我們設定每個軌跡都有兩個可供比對的時間點，一為軌跡的起點時間，一為軌跡的終點時間。在關係上，我們設定有「之前」、「同時」、「之後」三種關係，並且可以指定其時間的間隔大小，基本單位為秒。為了使定義能更加明確，我們為系統制定了查詢語法，每一段基本描述式可表為

$obj.tp [op t] rel_op obj.tp [op t] [e:error_time] [l:limit_time]$

分別說明如下：

- *obj* 為輸入軌跡的代號，代表資料庫裡軌跡與 *obj* 相似的物件所構成的集合，同一段描述式內的兩個 *obj* 不得相同。*tp* 為時間點屬性，可為 *start* 或 *end*，*start* 代表起點時間，*end* 代表終點時間。中括號內為選用語法，可以不使用。
- *op* 可為加法(+)或是減法(-)運算子，*t* 為秒數。
- *rel_op* 為關係描述子，可用的關係描述子有大於(>)、等於(=)、小於(<)、大於等於(>=)、小於等於(<=)、不等於(<>)。
- *error_time* 為容許誤差時間，單位為秒，例如在使用等於描述子(=)時可指定容許誤差時間，以避免檢索條件過於嚴苛。*limit_time* 是為使用等於(=)以外的關係描述子時，用於限制最大容許的時間差。以“*a.start* > *b.end*”描述式為例，若是不指定 *limit_time*，則 *a.start* 比 *b.end* 還遲上一個小時這類組合也仍符合描述式，但在一個事件裡，很少有物件互動的時間會相隔這麼長，此時便可設定 *limit_time* 以限制最大容許的時間差。若 *limit_time* 與 *error_time* 同時使用，則最大容許時間差將變為 *limit_time* + *error_time*。
- 描述式可以是複合式的表達方式，描述與描述之間以邏輯運算子結合，可以使用的邏輯運算子為 AND 與 OR，其中 AND 運算優先權高於 OR。若使用了兩個以上的邏輯運算子，可使用小括號來改變運算優先權。

我們舉些查詢範例：假設手繪輸入三個軌跡，分別設定其代號為 *a*、*b*、*c*，若要查詢符合「軌跡像 *a* 的物件，在軌跡像 *b* 物件停止 5 秒後才開始動」的影片片段，則語法為“*a.start* > *b.end* + 5”。同理，要查詢「*a* 開始動之後，*b* 才開始動，且 *a* 與 *c* 同時停止，並容許有前後 3 秒鐘的誤差」，則語法為“*b.start* > *a.start* AND *a.end* = *c.end* e:3”。當使用者輸入了多物件事件資訊並開始檢索後，事件索檢子系統會將每個輸入的軌跡曲線送至主系統查詢，並分別為每個軌跡的查詢回傳結果存成一個集合。當所有集合都收集完畢後，依所輸入的時間關係查詢語法，開始至各個集合內挑選符合語法描述軌跡組合，並將結果資訊回傳給使用者。

接下來說明描述式的時間關係比對與多描述式的邏輯運算方法。假設一描述式 *expr* 裡指定了 Q_a 與 Q_b 兩個使用者手繪輸入的軌跡，且 Q_a 不得等於 Q_b 。假設 Q_a 與 Q_b 於雙階層軌跡檢索分別得出 S_a 與 S_b 的物件集合，則 *expr* 的時間關係比對結果應滿足：(1) $s = \{x_a, x_b\}$ ，其中 x_a 與 x_b 的時間關係符合 *expr* 所述，(2) $x_a \neq x_b$ 。多描述式的邏輯運算，在 AND 運算上，假設有兩單一描述式 *expr_l* 與 *expr_r*，並以 AND 運算子連結成 *expr_l* AND *expr_r* 之複合描述式。設兩軌跡 Q_l 與 Q_r 各自代表在 *expr_l* 與 *expr_r* 中所

指定的使用者輸入之軌跡之一，並參照前述單一描述式比對方法中的定義為 Q_l 、 Q_r 定義出對應的物件軌跡 x_l 與 x_r 。於是， $expr_l$ AND $expr_r$ 的運算結果 $S_{(expr_l, expr_r)}^c$ 應滿足：

- (1) s^c 是由 $S(expr_l)$ 與 $S(expr_r)$ 中各取一組集合，經聯集後所成的一個集合
- (2) 存在至少一組 $Q_l = Q_r$
- (3) 對於所有的 x_l 與 x_r ，如果 $Q_l = Q_r$ ，則 $x_l = x_r$ ；如果 $Q_l \neq Q_r$ ，則 $x_l \neq x_r$ 。

上述定義的 AND 運算，也可將 $expr_l$ 與 $expr_r$ 推廣至複合式描述間的運算。在 OR 運算上，設 $expr_a$ 、 $expr_b$ 、 $expr_c$ 為單一或複合描述式，我們將任何符合

$$expr_a \text{ AND } (expr_b \text{ OR } expr_c) \text{ 與 } (expr_b \text{ OR } expr_c) \text{ AND } expr_a$$

形式的式子展開為

$$(expr_a \text{ AND } expr_b) \text{ OR } (expr_a \text{ AND } expr_c)$$

並為每一個被 OR 運算子分隔的式子個別輸出比對結果給使用者。比對出來的每個一筆結果都是含有多個軌跡的子集合，子集合內所有軌跡中所存在的最早的一個時間點，與最晚的一個時間點所構成的時間區間，便是該筆結果所代表的影片片段。

六、實驗結果

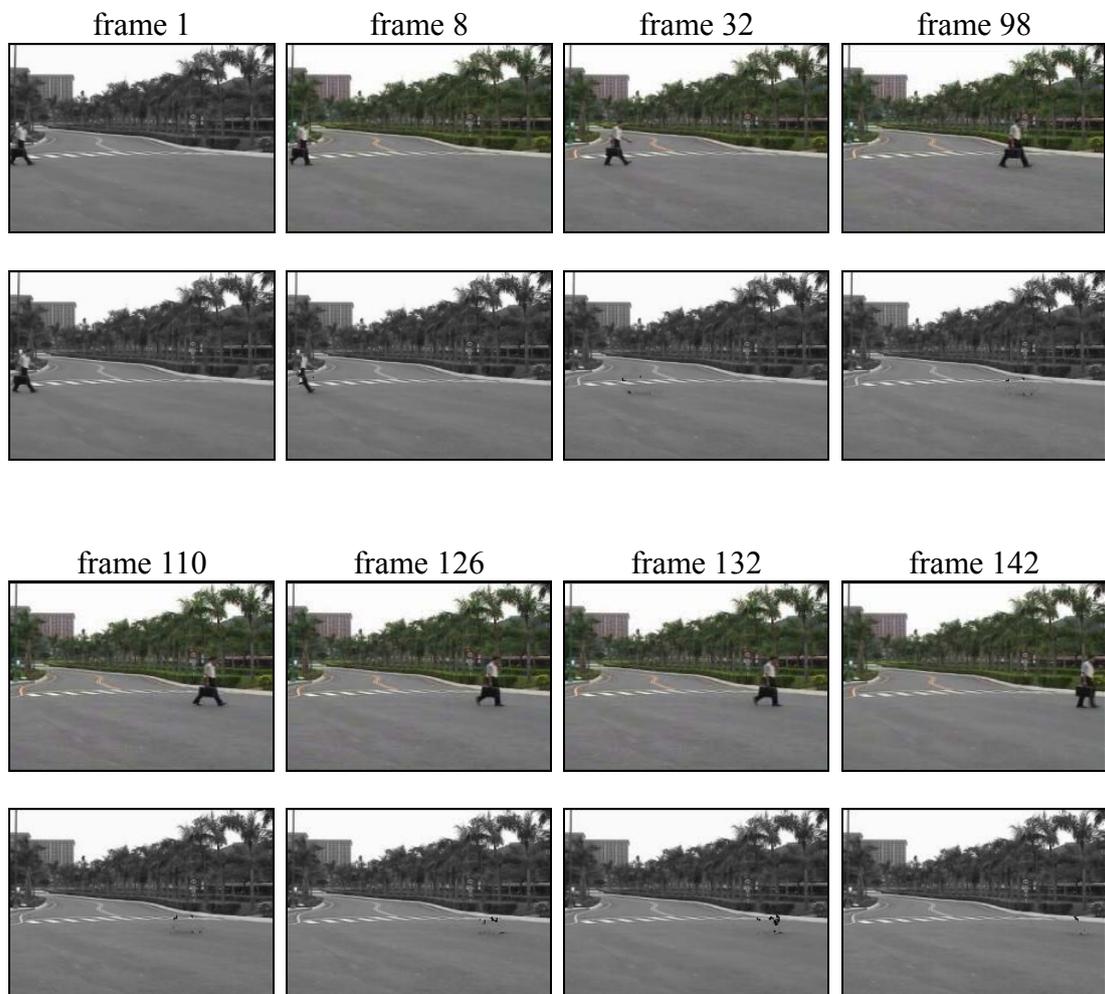
為了驗證上述方法的實際成效，我們分為「多物件追蹤」與「事件檢索與偵測」兩部分來說明：

6.1 多物件追蹤

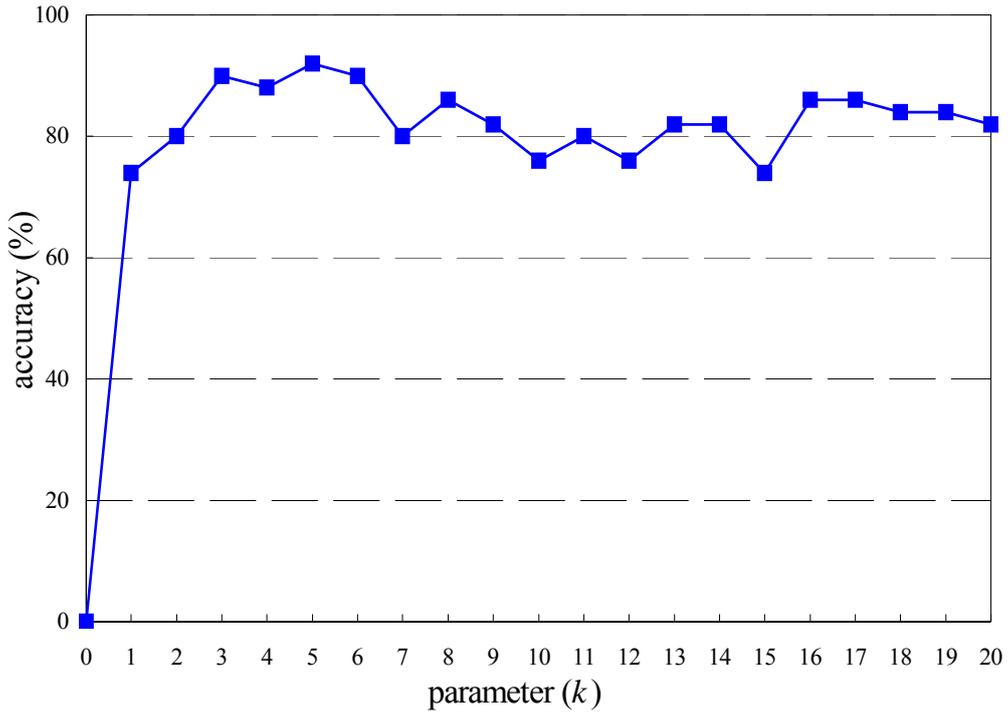
首先，我們架設固定攝影機實際拍得一些視訊資料，視訊測試內容種類包括路上行人、室內外之行進車輛、運動員等，視訊影片取樣為每秒 30 張畫面，接著透過物件偵測流程將物體偵測出來。此外，在偵測過程中我們提出一背景自動更新技術，並配合背景與畫面差分技術，以力求物件擷取的正確性及完整性。為了改善部分雜點及破洞問題，採用形態學上的一些運算或透過濾波器，以維持區域的完整性及去除不必要的雜訊。接著分離出物件並估算其物件在畫面的重心座標，最後獲得物件的運動軌跡，如圖八所示。



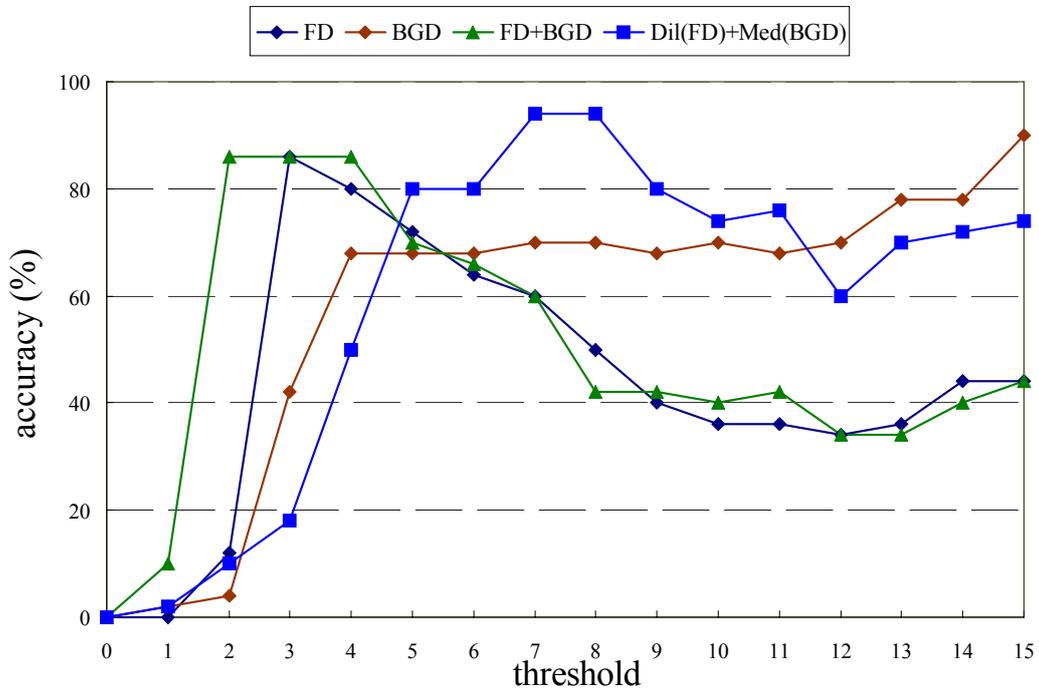
圖八：移動物件偵測與軌跡追蹤



圖九：背景更新後影像：上為不同畫面的影像，下為背景更新後影像



圖十：參數 k 對物件正確分離的關係

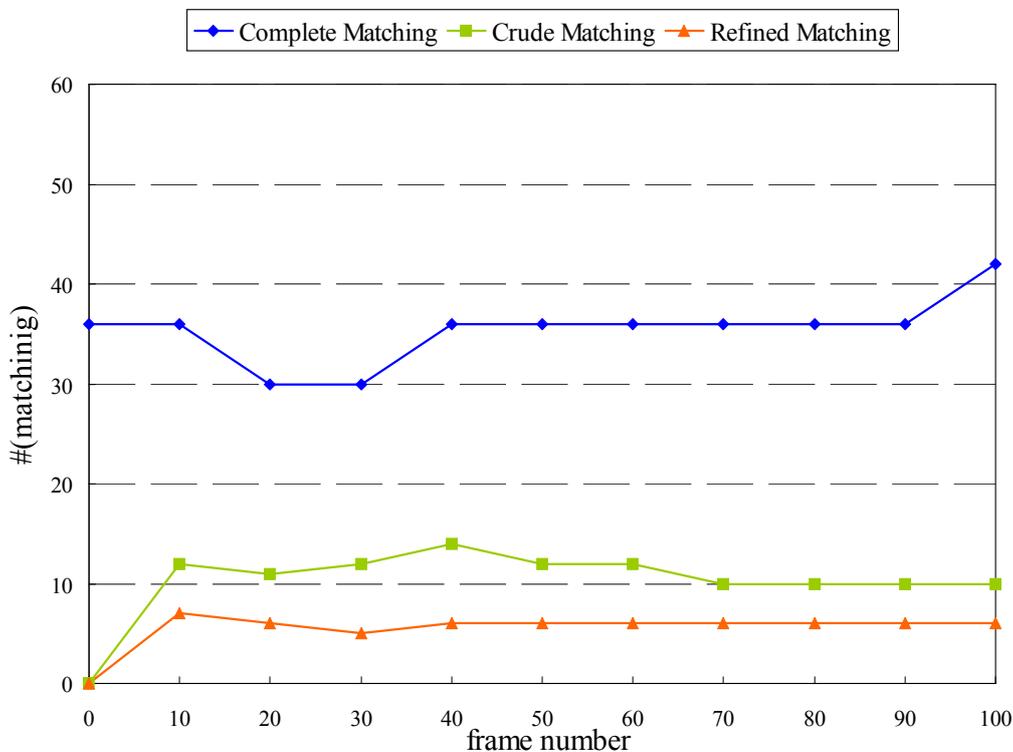


圖十一：不同物件偵測方法的比較

對於背景影像自動更新而言，圖九中即為可調式之背景參數 k 在視訊中不同時刻所得之背景更新影像。為了進一步探討不同 k 值對物件分離的影響，我們展示其分離結果。當 $k = 1$ 時對背景影像更新而言，由於單純只參考前單一影像資訊，可能受其參考資料不足之影響，使得物件完整性喪失而產生破碎，因而造成精確度不佳。反之，若 k 選定過大可能因參考過多資訊，造成更新影像與目前視訊畫面差異過大，導致雜訊生成因而降低準確度。以 $k = 5$ 為例，畫面中不僅保留 $k = 1$ 低雜訊之優點且擁有 $k = 15$ 之物件完整特性。故在此，對參數 k 之選定，應於兩者間尋求一適當值，結果如圖十所示，在 $k = 5$ 時具有較佳的偵測結果。

其次，對於濾波器與物件追蹤上的關係而言，我們比較分離物件的正確率，結果如圖十一所示，其中 FD 表示循序畫面差分影像， BD 為背景差分影像， $Median(BD)$ 表示將背景差分影像透過中間值濾波器運算。在此，依據循序差分影像及背景差分影像的特性，分別選取適當濾波以求得較佳之偵測結果。因 $Dilation$ 運算來對循序差分影像有助於坑洞之修補，中間值濾波器對背景差分影像有利雜訊去除之功效，又具有維持物件之輪廓完整之特性。故適當選取雜訊濾波是有助於提高物件偵測的正確性及穩定性。

根據動態視訊物件分離技術萃取出個別物件之後，依物件於時間軸上移動的資訊進一步地分離出個別移動物件，並且個別求得其移動軌跡特徵，如位置、區域大小、色彩強度之特徵量來相互匹配，尋找出相似的物件。然而於分離過程中，物體彼此間可能因視角不同，彼此間可能發生交錯、遮擋或過近等現象，故我們應該針對此一現象，希

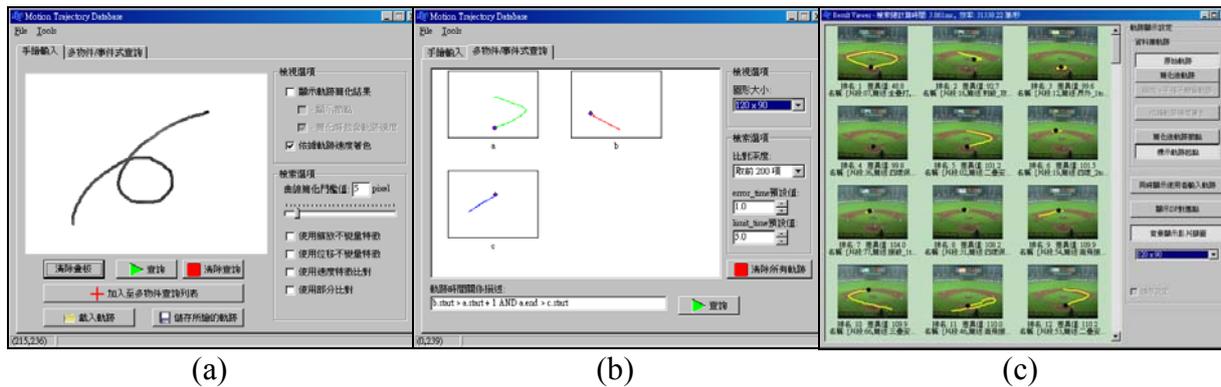


圖十二：不同配對方法之比較

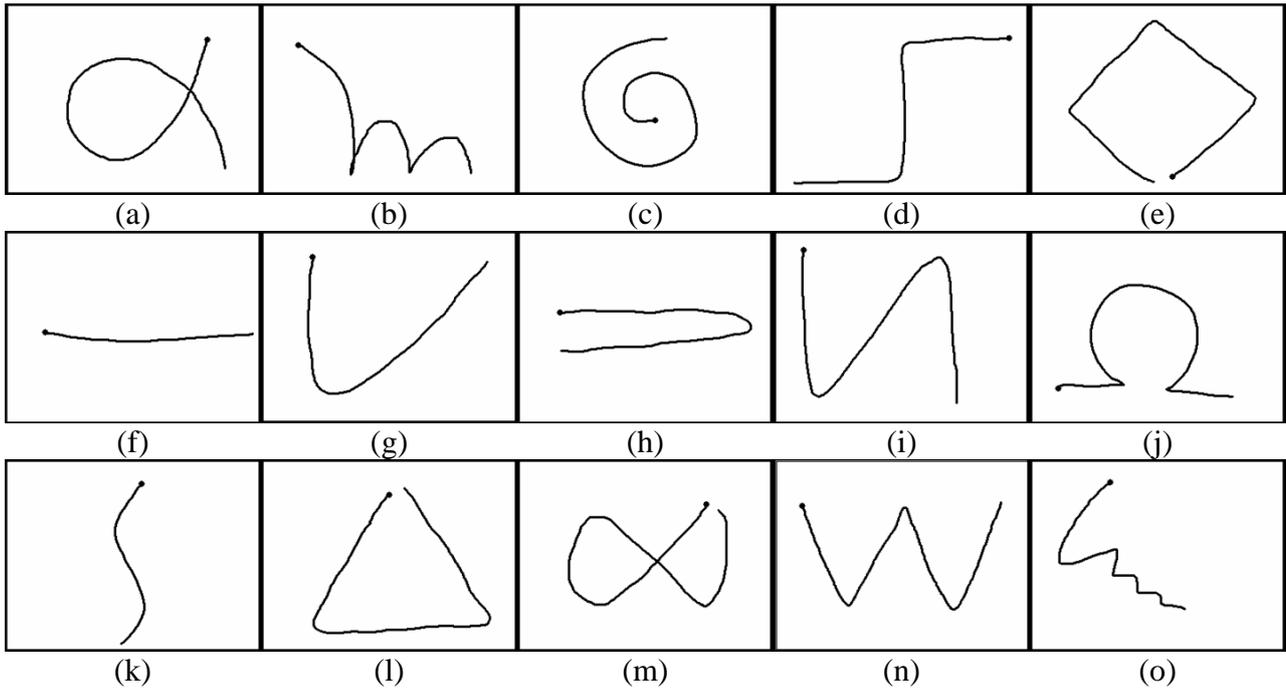
望透過一雙階層演算法則，先依據初步匹配模式(*crude matching*)將比對結果先區分為分裂、合併、靜止或生成等類別，並紀錄目前物件 O_u 及匹配物件 O_v 之個數 $DegS[u]$ 與 $DegT[v]$ ，接著爲了避免物件產生過多的比對，亦防止物件本身同時具有合併且分裂的狀態產生。因此我們透過精確的匹配(*refined matching*)方式將其改善，以期達到物件比對的穩定，最後可獲得物件在該時刻之匹配總個數，有關演算後的結果如圖十二所示。其中最上方表示所有物件皆匹配之結果，中、下方曲線分別代表 *crude matching* 與 *refined matching* 結果。由圖得知，透過 $DegS[u]=1$ 或 $DegT[v]=1$ 的優先條件下可避免物件過多分裂，因此利用精確匹配修正後更能保持物件匹配之穩定性。

6.2 事件檢索與偵測

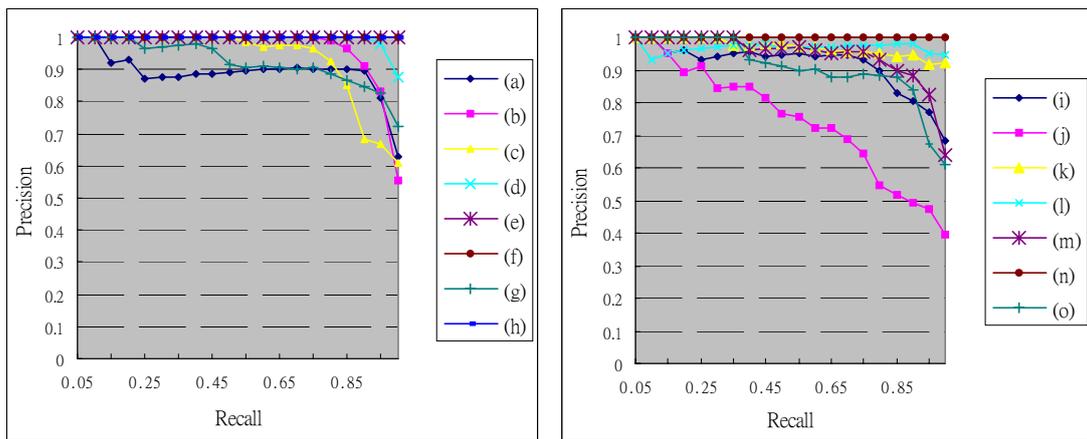
對於上述所提的方法，我們實際在電腦上進行測試。我們首先設計視訊檢索測試平台的兩類查詢介面如圖十三所示：第一類主要供使用者以手繪方式輸入欲查詢的軌跡，第二類則可輸入物件關係描述式以作爲事件檢索的檢索介面。接著，則爲檢索結果瀏覽介面，此介面會將資料庫中的軌跡依比對的相似程度由高至低列出。爲瞭解視訊檢索各種功能實際的擷取成效，我們對個別方法作一測試：



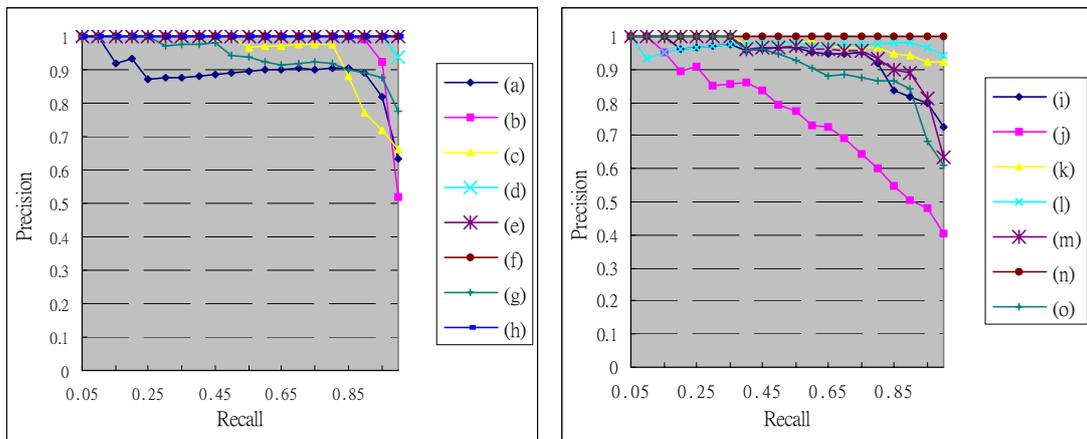
圖十三：使用者輸入介面：(a)手繪輸入查詢介面，(b)事件式查詢介面，(c)檢索結果瀏覽



圖十四：軌跡主要類型

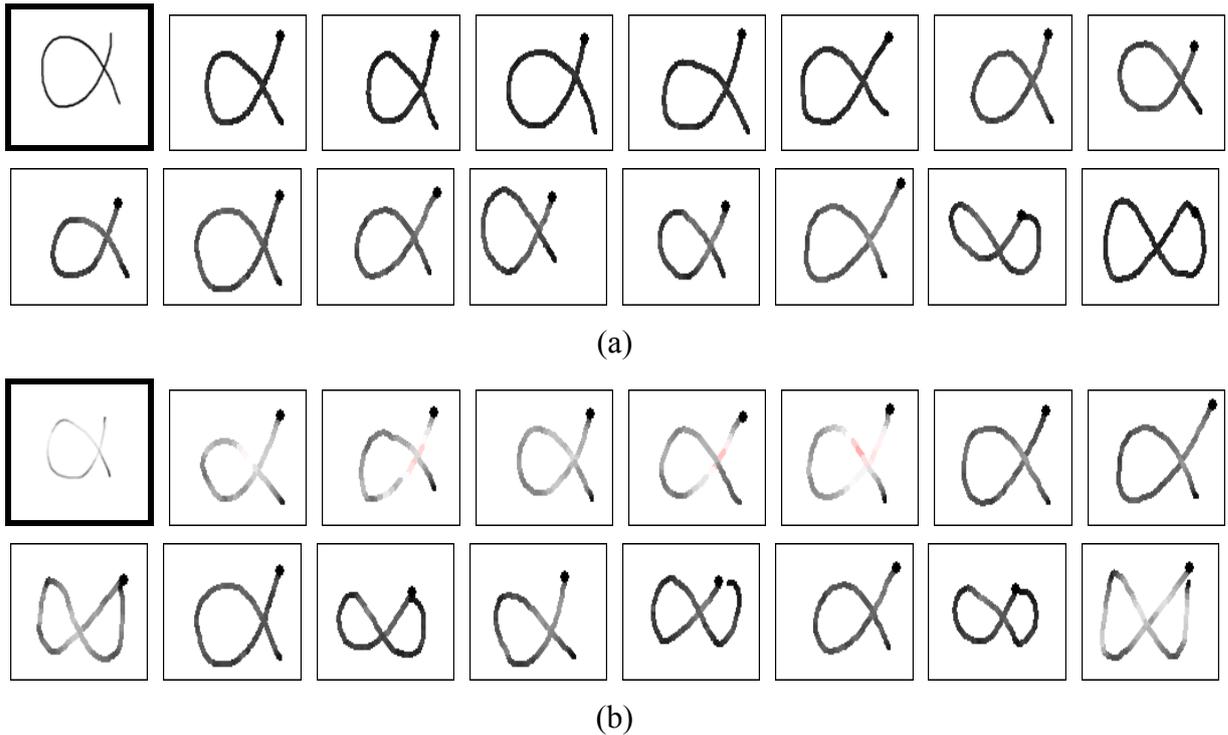


(a)



(b)

圖十五：傅立葉描述子檢索實驗結果：(a) $M = 5$, (b) $M = 10$ 。



圖十六：混合特徵曲線實驗結果，其中粗邊框為查詢軌跡，檢索結果由左至右依序排列：

(a)低移動速率軌跡、(b)高移動速率軌跡

6.2.1 傅立葉描述子檢索測試

在傅立葉描述子的檢索測試中，我們以人工手繪的方式來建立軌跡資料庫，其中軌跡分為 15 個類型，每類 20 筆，總共 300 筆資料以供測試，每類軌跡的外形如圖十四所示。每個類型各檢索 5 次，同時計算取 5 項 ($M = 5$) 與 10 項 FD_{inv} ($M = 10$) 的比對結果，並統計它們在各個回呼率(recall)下的平均準確率(precision)值，圖十五為所得出來的統計圖。我們可以觀察到，相對於只取 $M = 5$ 作比對， $M = 10$ 並不會得到顯著的進步。我們進一步找出在 $M = 5$ ，回呼率為 100% 但準確率未達 100% 的軌跡，與它們在 $M = 10$ 的比對結果作比較，發現 $M = 10$ 只使回呼率為 100% 時的準確率平均上升 1.75% 而已，因此作為第一階層的比對，取 $M = 5$ 已足以表達大致的軌跡。

6.2.2 混合特徵曲線檢索測試

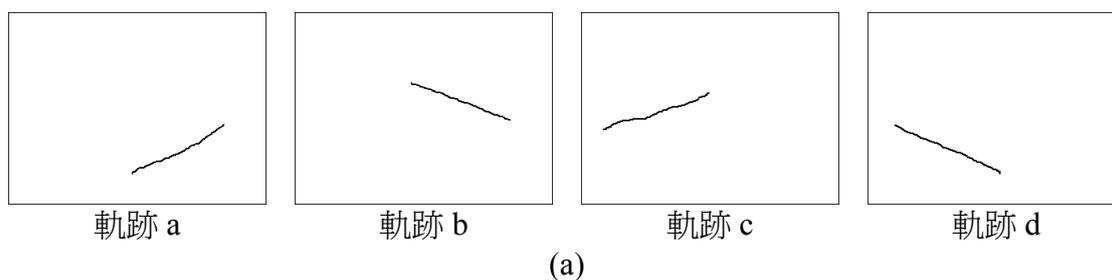
此部分的測試中，我們將曲線簡化門檻值定為 5 像素的誤差容忍值，然後分別手繪輸入同樣外形，但是不同速率的軌跡，藉以同時測試速率特徵檢索的成效。但由於原始軌跡速率的變化幅度不如座標資料來得大，因此我們將所有的軌跡速率值皆乘以 10 倍權重值，以強化速率資訊對檢索的影響力。圖十六即為使用速率特徵比對的檢索結果，圖的左方為輸入的軌跡，右方為比對後排名前 15 的軌跡(由左而右，由上而下)，軌跡顏色由黑—白—紅分別代表速率的慢至快。由圖中可觀察到，速率特徵的使用，明顯反應在比對結果上，部分“∞”形狀的軌跡會被檢索進來則是因為其外形有一定程度的相像，加上速率特徵非常吻合所致。若是不使用速率比對，在回呼率為 100% 的條件下，準確率皆有高於 90% 的成效。

6.2.3 事件檢索測試

在此實驗裡，我們使用實際的影片來作為實驗對象。鑑於運動影片含有較多的事件，而其中棒球比賽又是目前最受國人喜愛的運動球賽之一，因此我們採用棒球影片為例，實地至職業棒球比賽場地，以數位攝影機架設於本壘正後方之看台高處，並固定視點拍攝全球場。然而受限於設備解析度的不足，影片中的物件需以人工方式協助追蹤，同時考慮到我們所提出的檢索方法是屬通用性方法，並無針對特定類型的影片作最佳化。影片裡，可擷取出的物件有攻方球員、守方球員與裁判共三個種類，首先利用球衣顏色[26] 可將畫面上的三類物件區別出來。我們自影片中擷取出的軌跡，共有 121 段，若以打擊者的跑壘動作為主來分類，概略可分出 68 個事件，其中包括安打 24 件、全壘打 1 件、保送 9 件、刺殺出局 13 件、接殺出局 20 件、界外球 1 件。我們將測試部分比對、移動速率比對，以及多物件事件在棒球事件偵測的應用。我們以得分事件來作為偵測目標，由於得分必定是有跑者經由三壘到達本壘，因此便可使用部分比對來檢索出所有會經過這一段路徑的軌跡。我們輸入的軌跡與檢索結果如圖十七所示。在結果中，僅有兩個非得分事件的片段被檢索進來，但其跑者的運動仍完全符合所輸入的軌跡，只是因為同一時間中的打擊者遭刺殺或接殺並達到三人出局，而沒構成得分條件。我們以保送事件來作為偵測目標，被保送的球員通常會從容地跑步至一壘，因此我們可繪入由本壘至一壘的慢速軌跡，並使用速率比對來檢索。輸入的軌跡與檢索結果如圖十七所示，在達到回呼率 100%的範圍中，有 3 個非保送的事件，分別為遭高飛接殺及內野接殺的打擊者及在二壘遭刺殺的跑者，前一者由於在球擊出後即了解自己必定出局，因此放慢速度跑向一壘的移動軌跡也正好符合檢索條件，後兩者則是因為已出局而慢慢地走回球隊休息區，雖然移動路徑與所輸入檢索的軌跡不是很類似，但其慢速移動的速率而使得兩者意外檢索進來。

比對方式	輸入的軌跡	檢索結果	
		正確筆數	準確率(當回呼率為 100%時)
部分比對		13	86.67%
速度比對		9	75%

圖十七：單一移動物件的事件偵測



檢索描述	檢索目標於資料庫中筆數	準確率 (當回呼率為 100%時)
b.start > a.start e:0 l:8	0	
c.start > a.start e:0 l:8	1	failed
d.start > a.start e:0 l:8	1	100%

(b)

圖十八：多物件事件測試：(a)輸入的軌跡與對應的代號，(b)檢索結果

此外，我們以多個在同時期發生的事件來模擬多物件事件測試。這裡以「壘上有跑者時擊出外野高飛球」的事件作為偵測目標。在打擊手擊出外野高飛球並離壘後，位於其他壘上的跑者需在球落地或是被接住後才能離壘，我們以這個規則來嘗試偵測這類特定事件。檢索條件與結果如圖十八，其中 $c.start > a.start$ 此一段描述式並無法檢索出期望的目標影片。檢視資料庫後得知，在目標影片片段中，軌跡 c 所代表的二壘跑者正如同一般情況，預先離壘以拉近與下個壘包的距離，遇到高飛接殺會有回壘動作，然後才再離壘，但此小段影片中的跑者回壘後停留的時間太短，因此軌跡切割方法並沒有偵測到此一時間點，於是回壘與前進下一壘的軌跡連在一起沒分開造成軌跡的起始時間點比打擊者還要早，故無法檢索出來。

七、結論

根據視訊物件分離技術與背景影像重建方法，我們可以有效分離出移動物件與靜止物件。過程中透過一連串畫面資訊，可自動重建出背景影像，以改善傳統背景影像必須事先取得的缺點，並透過影像的即時更新，亦可獲得較佳的分離成效。在偵測過程中，我們根據背景差分分離對物件偵測上較能保有完整性及循序差分分離對外環境具有較小雜訊干擾的特性，依此原則將兩者加以整合，以達到降低物件破碎及雜訊干擾的缺失。此外，我們亦進一步探討多物件在不同畫面間的匹配問題，透過精確匹配演算法 (refined matching) 適時判別相鄰畫面間的物件對應關係，以確保物件追蹤的正確性。由實驗結果中顯示，所提方法可以達到預期的成果。

此外，在此一研究亦探討物件運動軌跡為基礎的視訊檢索與事件偵測方法。在視訊檢索方面，採用質心距離的傅立葉描述子作為資料庫的初步篩選；所提的混合式曲線

用以改善傳統多特徵分離比對，結合動態規劃法可完成兩軌跡的精確比對。結合上述而成的雙階層比對方法，得以在檢索效率與擷取成效間取得較佳的平衡點。在事件偵測方面，則是利用視訊檢索平台開發事件的描述工具。藉由運動軌跡查詢語言，對特定視訊內涵建構事件模式，以期可自動偵測出高階意涵。

爲了驗證所提方法的可行性，我們首先以手繪軌跡資料庫進行測試。實驗結果顯示，對於一般性軌跡仍有不錯的效果。接著，我們以實際棒球視訊爲例進行棒球事件的測試。由於我們所提出的方法是以通用性爲研究目標，然而部份棒球賽裡的物件行爲卻是有著相當的多樣性與不定性，因此只對棒球影片作最低限度的處理下，難免沒辦法對所有的棒球事件都訂定出相對應的查詢模型。但在實驗中得到的結果，仍可看出我們提出的方法在較爲簡易的事件上能有一定程度的適用性。少數無法成功偵測的主要理由是因爲無法單獨使用局部的運動軌跡特徵來充分描述事件的內容，因此未來的研究重點將擴及其它特徵，以進一步提升電腦在視覺的認知能力。

誌 謝

本研究承行政院國科會(NSC 94-2213-E-214-034)、義守大學(ISU 94-01-17)經費補助，特此致謝。

參考文獻

- [1] F. E. Alsaqre and Y. Baozong, "Moving Object Segmentation from Video Sequence," Proceedings of Conference focused on Video/Image Processing and Multimedia Communication, pp. 193-199, July, 2003.
- [2] A. Broggi, M. Bertozzi, A. Fascioli and M. Sechi, "Shape-Based Pedestrian Detection," Proceedings of the IEEE Intelligent Vehicles Symposium, Dearbon, pp. 215-220, October, 2000.
- [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, vol. 12, no. 7, pp. 577-586, 2002.
- [4] Y. S. Cho, S. H. Lee, J. S. Shin and Y. S. See, "Shape Coding Tool: Using Polygonal Approximation and Reliable Error Residue Sampling Method," ISO/IEC JTC1/SC29/WG11 MPEG 95/565, Germany, 1996.
- [5] W. A. Khatib, A. Ghafoor and A. Khokhar, "Trail-based Approach for Video Data Indexing and Retrieval," International Conference on Multimedia Computing and Systems (ICMCS), 1999.
- [6] G. L. Foresti, "A Real-Time System for Video Surveillance of Unattended Outdoor Environments," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 8, no. 6, pp. 697-704, 1998.
- [7] Z. Gu, "Video Database Retrieval Based on Trajectory Analysis," master thesis, Fudan University, 1999.
- [8] A. Guezic, "Tracking pitches for broadcast television," IEEE Computer, vol. 35, no. 3, pp. 38-43, March 2002.
- [9] M. Han, A. Sethiy and Y. Gong, "A Detection-Based Multiple Object Tracking Method," International Conference of Image Processing, NEC Laboratories America, USA, 2004.

- [10] B. Heisele, "Motion-Based Object Detection and Tracking in Color Image Sequences," Fourth Asian Conference on Computer Vision, Taipei, pp. 1028-1033, 2000.
- [11] Y. K. Jung, K. W. Lee and Y. S. Ho, "Content-Based Event Retrieval Using Semantic Scene Interpretation for Automated Traffic Surveillance," IEEE Transactions on Intelligent Transportation Systems, vol. 2, no. 3, pp. 151-163, 2001.
- [12] J. B. Kim and H. J. Kim, "Efficient Region-Based Motion Segmentation for a Video Monitoring System," Pattern Recognition Letter, vol. 24, no. 1, pp. 113-128, 2003.
- [13] J. B. Kim, C. W. Lee, S. W. Hwang and H. J. Kim, "A Real-Time Moving Object Detection for Video Monitoring System," ITC-CSCC, pp. 454-457, 2001.
- [14] J. Z. Li, M. T. Ozsu and D. Szafron, "Modeling of Moving Objects in a Video Database," IEEE International Conference on Multimedia Computing and Systems, pp. 336-343, 1997.
- [15] T. W. Liu, K. Z. Lee and S. Y. Ho, "Automatic Parameter Tuning Technique for Designing Robust Motion Detection Systems," Proceedings of CVGIP, R.O.C., 2004.
- [16] M. R. Lyu, E. Yau and S. Sze, "Video and Multimedia Digital Libraries: A Multilingual, Multimodal Digital Video Library System," Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries, 2002.
- [17] C. Motamed, "Video Indexing Based on Object Motion in Video-Surveillance Context," Conference RIAO, 2000.
- [18] H. Nanda and L. Davis, "Probabilistic Template-Based Pedestrian Detection in Infrared Videos," Proceedings of the IEEE Intelligent Vehicles Symposium, pp. 1-11, 2002.
- [19] A. Naohiro and F. Akihiro, "Detecting Obstructions and Tracking Moving Objects by Image Processing Technique," Electronics and Communication, vol. 82, no. 11, pp. 28-37, 1999.
- [20] J. Perš and S. Kovacic, "Computer Vision System for Tracking Players in Sports Games," Workshop on Image and Signal Processing and Analysis, Pula, Croatia, June 14-15, 2000.
- [21] G. S. Pingali, Y. Jean, and I. Carlbom, "Real time tracking for enhanced tennis broadcasts," in Proc. IEEE Comp. Vision and Patt. Rec. (CVPR), pp. 260-265, 1998.
- [22] L. Rowe, J. Boreczky and C. Eadds, "Indexes for User Access to Large Video Databases," Storage and Retrieval for Image and Video Databases (SPIE), 1994.
- [23] E. Sahouria and A. Zakhor, "A Trajectory Based Video Indexing System for Street Surveillance," IEEE International Conference on Image Processing (ICIP), 1999.
- [24] C. Snoek and M. Worring, "Multimedia Event-Based Video Indexing Using Time Intervals," IEEE Trans. Multimedia, vol. 7, no. 4, pp. 638-647, 2005.
- [25] K. Sobottka, P. Zuber and H. Bunke, "Shape-Based Template Matching for Robust Obstacle Tracking in Low-Resolution Range Image Sequences," International Conference on Systems Research, pp. 105-110, 1999.
- [26] L. Wang, B. Zeng, S. Lin, G. Xu and H. Y. Shum, "Automatic Extraction of Semantic Colors in Sports Video," IEEE International Conference on Acoustic, Speech, and Signal Processing (ICASSP), 2004.
- [27] Y. Wang, R. E. Van Dyck and J. F. Doherty, "Tracking Moving Objects in Video Sequences," Proceedings of Conference on Information Sciences and Systems, vol. 2, pp. 24-29, March, 2000.

- [28] C. Yajima, Y. Nakanishi and K. Tanaka, "Querying Video Data by Spatial-Temporal Relationships of Moving Object Traces," VDB, pp. 357-371, 2002.
- [29] A. Yoshitaka, M. Yoshimitsu, M. Hirakawa and T. Ichikawa, "V-QBE: Video Database Retrieval by Means of Example Motion of Objects," International Conference on Multimedia Computing and Systems (ICMCS), 1996.
- [30] D. S. Zhang and G. Lu, "A Comparative Study on Shape Retrieval Using Fourier Descriptors with Different Shape Signatures," International Conference on Intelligent Multimedia and Distance Education (ICIMADE01), pp.1-9, June 1-3, 2001.
- [31] Y. Zhong, A. K. Jain and M. P. Dubuisson-Jolly, "Object Tracking Using Deformable Templates," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 5, pp. 544-549, 2000.