

使用內容查詢及空間關係之多攝影機視訊保全監控系統

A Multi-Camera Video Surveillance System by Using the Content-based Retrieval and Spatial Relationship

張厥煒
台北科技大學資工系
cwchang@ntut.edu.tw

曹佑羽
台北科技大學資工系
s4598029@ntut.edu.tw

楊棟華
台北科技大學資工系
s4598024@ntut.edu.tw

摘要

傳統的多攝影機保全監控系統，在災害或犯罪事件發生後，想要找出可疑人物，並歸納出行經路線時間，必須經由人工調閱比對事件發生時所錄下的視訊影像，如此一來將花費過多的人力、時間，甚至延誤案情。因此，本論文提出一個適用於多攝影機保全監控系統的空間關係建立與視訊內容檢索搜尋方法。此方法使用藉由關鍵異常畫面擷取(Key Frame Extraction)法，利用異常物體面積、異常物體密度、異常物體寬高比、畫面變動量、畫面清晰度、畫面明亮對比等六種特徵，過濾選取出可明顯表現出異常物體特徵之關鍵異常畫面，儲存於視訊資料庫中。保全管理者可在監控視訊資料的瀏覽過程中，以時間或攝影地點等條件查詢異常事件的畫面，找到可疑並欲搜尋的異常物體後，可經由自行圈選出此異常物體，並以異常物體之物體模型(Object Model)，與監控視訊資料庫中所儲存的，關鍵異常畫面中異常物體的物體模型，計算出物體之間的相似度，以進行比對。搜尋時，可按照相似度排列其結果，並以異常事件發生時間，得知異常物體的行經路徑，以建立出一個完整的多攝影機監控視訊系統資料庫中異常物體搜尋之方法。

關鍵詞：保全監控系統、多攝影機協同事件偵測、物件追蹤、空間關係、視訊內容查詢。

一、前言

由於恐怖攻擊與犯罪率的上升，企業為了防治一般的竊盜行為以及天然災害，紛紛設置各種保全監控系統(Surveillance System)。然而這些監控設備，不但花費成本高，且必須由警衛人員在旁監看，有時易因人為疏失而失去其精確性與即時性。即使今日的數位式第三代保全監視系統，其功能及研發方向也只著重於異常偵測及多點涵蓋面管理等問題。然而，由於欠缺多攝影機彼此在空間上的關係，導致手動追蹤可疑人物時，常有空間迷失的缺點。同時，對於歷史監控畫面的搜尋，也沒有適當的視訊資料庫工具協助。

在多攝影機視訊監控的環境中，有些情形我們會想要在監控畫面中搜尋異常物體，第一種情況是在多攝影機監控環境下，藉由特定嫌疑犯在固定式攝影機中出現的位置，進而得知其行經路線；第二種則是如果可以藉由找出所有特定嫌疑犯出現的畫面，來獲得不同角度的嫌疑犯原貌，以取得可供指認、辨識之特徵。

因此，設計一套可適用於多攝影機之協同偵測與追蹤監控系統平台，建立中央操作控制站機制，負責與多攝影機進行同步協調工作，並以數位影像處理技術，開發新的結構化自動視訊物件偵測、分割、追蹤及儲存等前端處理方法，同時在監控系統背後的歷史視訊資料庫管理，或資料庫畫面的快速檢索等，是目前在學術研究及商業產品上，值得著墨的問題。

本論文的目的即是提出一個多攝影機空間關係建立，以及快速的視訊內容搜尋比對方法。此方法利用圖形理論建立多攝影機網路拓撲空間關係，以特定時間日期範圍、特定空間路線內有異常物體出現的場景段落，進行異常事件的查詢，再以內容為基礎的檢索技術[2] (Content-Based Retrieval)，應用於監控視訊資料庫畫面的檢索，藉由異常物體的顏色特徵等關鍵畫面擷取，找出資料庫中符合條件之異常物體，幫助我們縮小範圍，減少監看大量視訊資料的人力與時間。

二、相關文獻探討

在本論文中將「異常物體」定義為「與背景不同，會移動的夠大顏色區塊」，並將有同一異常物體出現的連續畫面定義為一個「異常事件」。要取得異常物體之特徵，必須先在監控視訊畫面中偵測出移動物體再進一步進行分割、特徵擷取之動作。而今視訊畫面中的移動物體偵測方式主要有三種[3]：(1)背景相減法(Background Subtraction)，此方法是將目前畫面與事先訓練好的背景影像相減，即可將前景區域留下。此方法優點是設計簡單而且計算複雜度低，但容易受到光線變化的影響，而發生誤判導致假警報，故需要良好的背景模型(Background Model)。如 Jabri[6]融合顏色以及邊緣資訊來建立背景模型。(2)時間差異法(Temporal Differencing)，跟背景相減法一樣使用影像相減來留下移動物體資訊，其差別為時間差異法是觀察特定時間內連續的相鄰幅(Frame)變化，以取得移動物體區域，如此一來因畫面間的相隔時間短，較不容易受光線變化影響，但得到的移動物體資訊不若背景相減法完整，如 Lipton[7]便是使用時間差異法，藉由前後張畫面的相減，在經過一個閾值函式(Threshold Function)決定變動區域來取得移動物體區域。(3)光流偵測法(Optical Flow)，此方法是計算光流量變化來取得移動物體資訊。光流偵測法可以在相機移動(Camera Motion)的情況下找到移動物體，

但此方法計算太過繁複，較不適合運用於有即時偵測需求的監控系統上，如 Meyer[8]便是使用光流偵測法來偵測移動物體區域，找出畫面中人物相同區塊的移動情形用來作為步伐的分析。關於異常物體的偵測與切割問題，已經有相當多的文獻探討此問題。因物件切割不是本論文所著重的議題，故本論文中採用最常見也是最容易實現的背景相減方法來偵測移動物體並切割出異常物體。

另一方面，如何達到有效率的檢索、搜尋異常物體資料變成為一個很重要的問題。早期檢索系統，都是使用文字做為輸入檢索條件。隨著多媒體時代來臨，影片、聲音、動畫等多媒體資料普及化，若要對所有資料一一做文字標籤建檔，需要龐大人力，也相當沒有效率。因此，為了解決這個問題，改由使用影像中本身內涵資訊，做直覺式比較，這便是以內容為基礎(Content-based)的影像檢索技術[2]。CBIR(Content-based Image Retrieval)主要的概念是用影像內容的主要特性當特徵來對影像資料本身加以描述，而特徵便成為搜尋影像時的索引條件。例如色彩(Color)、形狀(Sharp)、紋理(Texture)等是常見的低階(Low Level)特徵。理想的特徵對於影像內容必須具有代表性，這樣檢索系統才可以依特徵索引搜尋到符合預期的影像，達到有效的檢索效果。使用關鍵畫面當作影片索引(Query by Key Frame)也是常見的影片檢索方法，如 Chang[3]等人提出一個關鍵畫面擷取演算法(Key Frame Extraction Algorithm)，找出影片中的代表性圖片(Representative Image)當作關鍵畫面，並建立樹狀結構的關鍵畫面階層(Tree-structured Key Frame Hierarchy)，來增加影片檢索速度，此種方法優點是所需儲存資料量較少，但缺點是少了時間性(Temporal)相關特徵。

Content-based Retrieval 技術所面臨的問題，是不同類型的影像包含的特性包羅萬象，很難用特定特徵去代表所有影像。縮小範圍針對特定類型影像分析其特性再選擇特徵，是可行的方式。而在監控視訊

資料庫中，其視訊資料中的異常物體是我們關注的地方。在進行異常物體特徵擷取前，必須先進行的工作就是異常物體偵測。在異常物體從背景切割出來之後，我們需要有縮放不變性質(Scale Invariant)的特徵，並且要可以抵抗在不同角度拍攝下物體相當程度的變形。在物件追蹤相關研究中，以核心為基礎的物件追蹤(Kernel-based Object Tracking, [4])，採用利用物體顏色建立帶有空間位置權重的物體模型，來抵抗因移動造成的物體變形，達到有效追蹤的效果，此一物體模型，便成為我們異常物體特徵選取的最佳引用考量。

此外，關於多攝影機監控系統的研究，Remagnino、Shihab 與 Jones[11]提出一個多代理人架構，對於由多攝影機發出之資訊串流所形成的場景動態融合進行判斷。在此架構下的模組軟體會控制不同的系統元件，並以遞增方式建立場景模型。Black、Makris 與 Ellis[12]針對監控保全應用提出一個新的系統架構。在此架構下，系統具有事件偵測與視訊內容分析之能力。攝影機所拍攝的畫面中，在有部分重疊(Overlapping)之情況下，系統都能夠協調物件的追蹤。

三、系統運作架構與說明

本論文依前述功能需求，建置一個具有多攝影機協調、可疑事件預警和能縮短搜尋關鍵視訊時間的系統，以半自動化監控技術減少監控中心人力，並提昇監控的效能(如：降低回放搜尋時間、準確且有時效的異常預警和可疑物體之特徵擷取等)。本系統平台設計之方法，是要有效的以圖形理論建立多攝影機同步協調機制，並以影像處理技術開發新的結構化自動視訊物件偵測、分割、辨識、追蹤及儲存等前端處理機制。

為達成上述多攝影機協調及建構完整前端處理機制的目的，本系統主要設計分為五個部份，分別為(1)視窗介面管理子系

統、(2)多媒體控制子系統、(3)異常偵測子系統、(4)物件追蹤子系統、(5)攝影機協同子系統及(6)視訊資料庫內容查詢子系統。

視窗介面管理子系統為主功能選單與基本輸出入設定(含即時監控影像畫面顯示、回放影像畫面顯示、滑鼠事件處理、視窗上視訊播放控制功能等)作業，並提供適當的視窗介面執行多媒體控制、異常偵測與攝影機協同之功能。

多媒體控制子系統負責在區域網路環境下與網路攝影機溝通(如：透過 Socket 取得網路攝影機即時影像資料、將影像資料儲存至資料庫、自資料庫中取得回放影像資料等)。

物件追蹤子系統負責偵測、分割、辨識、追蹤目標物體及記錄事件，藉由在事先設定完成的偵測區域(Detection Region)內執行物體移動偵測(Motion Detection)可得知目前物體是否處於某區(Zone)內，一個偵測區域與一個指定的監控空間相關聯，在此空間內可有多台攝影機一起工作。偵測區域的資訊可支援異常偵測子系統進行異常事件偵測和搭配攝影機協同子系統分析目標物體目前於圖形拓撲何處。

異常偵測子系統負責異常事件偵測，當異常事件發生時，異常偵測子系統要負責送出警示信號，異常偵測子系統需與物件追蹤子系統搭配，藉由指定警戒區域(Alert Region)以及判斷物體行動路徑是否符合以區(Zone)為基本單位的條件路徑(Critical Path)等作法，可研判出發生異常事件的情況。

攝影機協同子系統負責建立多攝影機間的協調機制，攝影機協同子系統需與物件追蹤子系統搭配，分析目標物體目前於圖形拓撲何處。其主要工作包含：建立以區為基本單位的圖形拓撲、自動記錄發現物體移動時的時間戳記以及於特定空間下多攝影機監視屬性的變化等。

視訊資料庫內容查詢子系統是儲存並提供在監控資料中，最被關注的異常物體本身視訊畫面的查詢。在監控環境之下，

攝影機通常是固定不動的。因此，我們可以使用固定背景的移動物體偵測方法來偵測異常物體，同時可以記錄異常物體出現的時間及固定位置等相關資訊，記錄有異常物體出現的異常事件資訊，並從異常事件中挑選出清晰、有助於特徵擷取的關鍵異常畫面，再以關鍵異常畫面中異常物體的資訊及特徵做為以後檢索異常物體之依據。在不同攝影機或是不同的位置所觀察到的異常物體，有很大的差異性，唯一不至於變化太大的便是異常物體的顏色，所以在物件追蹤中所使用的能抵抗物體變形，並有縮放不變性質(Scale Invariant)的物體模型是我們所採用的比對模式。故除了異常事件所包含之日期、時間、攝影機編號及位置等資訊之外，再加上異常物體之間的相似度比對，如此便可建立一個可以快速搜尋到重要的異常事件，以及異常物體的監控視訊資料庫系統。

監控攝影機所拍攝儲存之畫面先經過異常偵測，找出異常事件，並加以紀錄下所有異常事件的資訊其中包含起始時間、終止時間、拍攝攝影機編號等。而異常事件通常會包含多張異常畫面，如果每一張畫面都加以儲存，會造成資料庫空間的浪

費以及檢索時間的耗費，故關鍵異常畫面擷取模組可以從異常事件中，挑選出一至數張代表性畫面，我們稱之為關鍵異常畫面，這樣可以有效減少儲存空間的浪費與加快檢索時間。之後特徵擷取模組會再進一步擷取關鍵異常畫面的特徵，成為之後進行相似度比對的依據。至此階段，資料庫中將記錄異常事件以及關鍵異常畫面之相關資訊。

設計後的多攝影機保全中央控制系統程式介面，如圖 1 所示。此多攝影機保全監控系統介面圖，其中包含(1)畫面顯示區域(Camera Display&Control)，可播放即時選擇的現場畫面或點選查詢後的異常歷史畫面、(2)攝影機群組(Group IP)，可增刪攝影機並設定拍攝地點名稱、(3)變動偵測設定(Motion Detection Setup)，設定各項異常警示之臨界值、(4)歷史資料查詢(Historical Record Query)，可依起迄時間、區域地點或特定攝影機、特定事件等查詢並表列、(5)圖形節點關連列(Graph Node Relation List)，可設定不同房間走道間的連通關係、(6)條件路徑設定(Critical Path)，指定需要偵測的條件路徑、(7)異常事件表列 (Event List)，即時以不同顏色燈

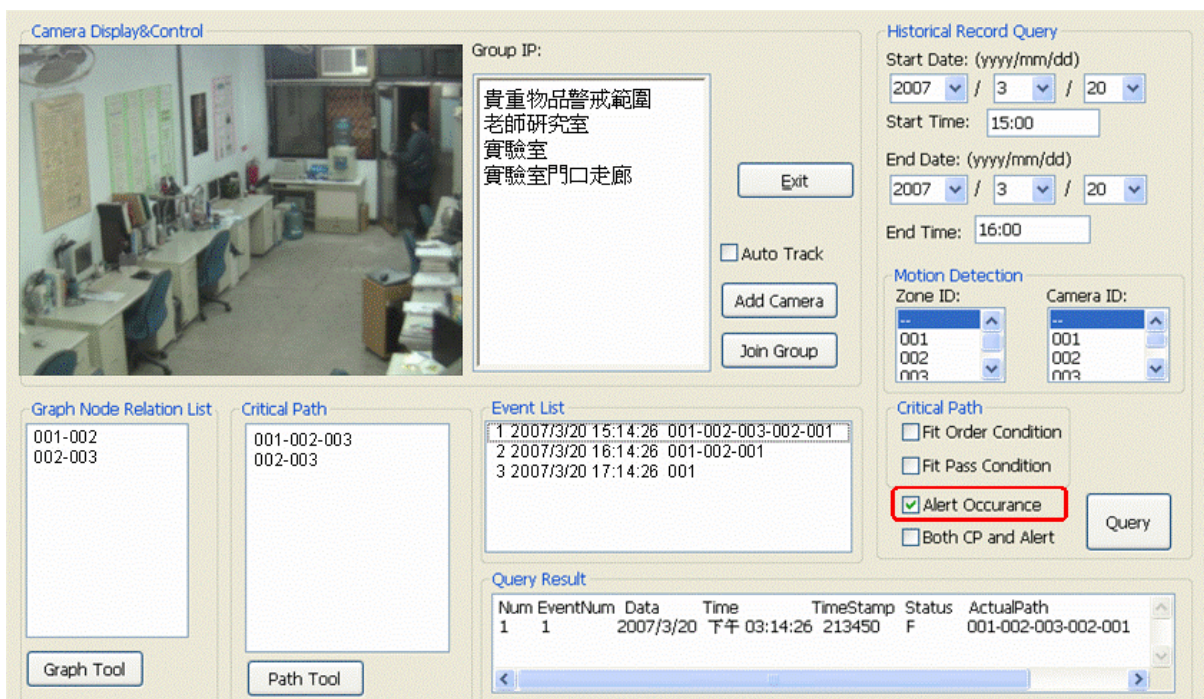


圖 1 多攝影機保全監控系統介面

號顯示監控的結果、(8)查詢結果表列(Query Result)，等介面區域。

其中，「異常偵測模組」使用背景相減方法進行移動偵測，從循序的監控視訊畫面中偵測出有異常物體出現的畫面，再根據有異常物體存在的第一張畫面跟最後一張畫面時間點，當成異常事件的起始時間及終止時間，最後將結構化的異常事件儲存在資料庫中。詳細作法將在第四節中加以描述。

「關鍵畫面擷取模組」會在各個異常事件中選出一張關鍵異常畫面，並將關鍵異常畫面中之異常物體建立特徵索引，成為日後異常物體搜尋之依據。關鍵異常畫面的選取會影響後面異常物體搜尋之效果，故我們希望能找出具代表性且有助於擷取有效特徵的關鍵畫面，本論文以畫面中的(1)異常物體面積、(2)異常物體密度、(3)異常物體寬高比等標準，判斷其為人類的可能性，並加上(4)畫面變動量、(5)畫面清晰度、(6)畫面明亮度對比等三個畫質衡量標準，提出一個關鍵異常畫面擷取演算法，來選出具代表性之關鍵異常畫面。而系統只需記錄關鍵異常畫面中異常物體之資訊，如此便可減少儲存空間的浪費與加快檢索時間。關鍵異常畫面擷取演算法的詳細作法將在第五節中加以說明。

想要在監控視訊資料庫中，快速的搜尋出特定的異常物體出現的所有異常事件，我們需要可以度量出異常物體相似度的特徵與比對方法。根據異常物體在監控畫面中出現的特性，也就是前面所提在不同攝影機觀察到的異常物體形狀與大小存在一定的差異性，所以異常物體的顏色是十分重要的依據。我們所採用的特徵是在物件追蹤中所使用，能抵抗物體變形並有縮放不變性質(Scale Invariant)的物體模型(Object Model)，也就是帶有像素位置權重特性的顏色分佈統計(Color Histogram)，度量相似度的方法採用 Bhattacharyya Coefficient。之後系統便能依此特徵針對不同異常事件所出現的異常物體作相似度比對。特徵擷取與相似度比對將於第六節詳

述。

根據上述模組建構出的系統，異常物體搜尋的操作步驟如下：(1)目標異常物體框選：使用者瀏覽視訊畫面，找到目標異常物體出現的畫面時，使用滑鼠拉曳框選異常物體在畫面中出現的位置與範圍進行搜尋。(2)異常物體特徵擷取與比對：從使用者框選的異常物體擷取出特徵，與視訊資料庫異常事件中關鍵異常畫面的異常物體，擷取出特徵進行比對。(3)相似度比對結果顯示：將異常事件依其異常物體特徵相似度依排序，並依系統所設定的信心閾值當做門檻值顯示出搜尋結果，使用者可點選查詢結果列表或播放以檢視內容。(3)行經路徑判斷：可以依照搜尋結果，按時間以及攝影機位置資訊，推算出異常物體行經路徑。如此一來，系統使用者，便能快速的從監控視訊資料庫中搜尋特定異常物體。

四、異常事件偵測

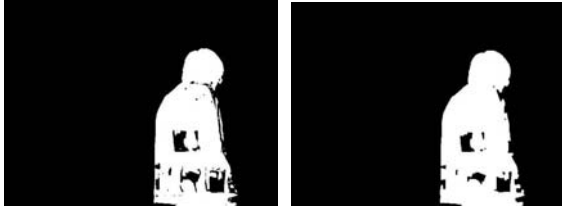
本論文將「異常事件」定義為在監控視訊畫面中有移動物體出現的一個視訊段落(Video Clip)，移動物體也就是相對於背景之前景物體，也就是在此系統中定義的「異常物體」。因此我們使用移動物體偵測，來判斷監控畫面中是否有移動物體出現。由於本論文將重點放在視訊資料庫之檢索，故採用較簡單的背景相減法(Background Subtraction)來偵測移動物體。相較於時間差異法(Temporal Differencing)和光流偵測法(Optical Flow)，背景相減法可以找到比較完整的異常物體資訊，並且沒有太多的計算複雜度。

由於監控環境中系統已事先學習了全彩(RGB 24 Bits)的背景畫面，且本研究中假設背景及光源不會有太大的變化，所以取得背景畫面後便不再做任何更新學習的動作。當目前監控畫面進入系統時，將目前畫面與背景影像相減，每一個相對位置的像素中 RGB 三個成分各自相減後的結果，只要其中之一超過閾值(Threshold)，



(a) 原始影像

(b) 背景影像



(c) 背景相減二值化

(d) 消除雜訊



(e) 填滿區塊

(f) 前景物體及框盒

圖 2 異常物體偵測主要影像處理結果

便視為前景區域(圖 2(c))，之後進行形態學操作(Morphological Operation)以消除雜訊(圖 2(d))。

型態學操作中，我們使用一次斷開運算以及一次閉合運算，用以去除一些雜訊以及填補物件殘缺的部份。

經過形態學操作的運算後，我們將畫面中的最大區塊填滿(圖 2(e))，並畫出可包含最大區塊的最小框盒(Bounding Box)(圖 2(f))，如此一來便取得了畫面中異常物體的位置資訊。當異常物體區域超過一個閾值(Threshold)時，我們便認定有異常發生並記錄下相關資訊。

完成異常物體偵測動作後，我們可以得到一連串有異常物體存在的畫面，將異常物體最先出現及最後出現的時間，視為一個異常事件的起始時間及終止時間，此異常事件的區間便是我們必須記錄於資料庫中的對象。接下來關鍵異常畫面擷取的

動作，則是為此異常事件找出其能成為索引的關鍵異常畫面。

五、關鍵異常畫面擷取

(一) 關鍵異常畫面選取考量因素

關鍵畫面(Key Frame)或稱為代表性圖片(Representative Images)，是指可表示一個場景意境的主要畫面。在 Content-Based Video Retrieval (CBVR) 相關研究中，關鍵畫面擷取(Key Frame Extraction)即是找出代表性畫面當作視訊檔案的索引[2][3]。而在我們的應用中，也是利用這個概念，將關鍵畫面作為視訊資料庫檢索之依據。在本論文的應用中，需考量關鍵畫面中的異常物體是否可擷取出良好之特徵，能幫助系統正確比對異常物體之相似度。因此我們改良李冠穎[1]的方式找出關鍵畫面候選，再利用類似 Zhang[10]的觀念，當前後畫面差異性過大時便視為關鍵畫面。因此本論文提出之關鍵異常畫面擷取演算法，可以依照異常物體在畫面中是否符合清晰度高、物件切割完整等特徵，擷取出良好之關鍵異常畫面。針對異常物體在畫面中的清晰度、總面積以及異常物體是否為假警報(False Alarm)等狀況，我們希望關鍵異常畫面的條件如下：

- (a) 第一是異常物體面積的大小。某些雜訊造成的假警報，以及異常物體只有部份出現在畫面中的情形必須排除，故異常物體的面積大小是我們考量的條件。
- (b) 異常物體與其框盒的面積比，本論文中稱之為異常物體密度，這個值用來確保異常物體切割的完整度。
- (c) 判斷異常物體寬高比在合理範圍內。使用背景相減容易造成的問題是當背景與前景顏色相近時，會造成誤判。此時所得到的異常物體並不完整，根據經驗異常物體的框盒寬高比落於 0.25~0.75 之間時，所取得的異常物體較完整及正確。

(d)畫面變動量的考量。當異常物體出現在監控畫面中時,如果有停留駐足時,我們對這段畫面有較高的興趣,此時前後畫面間的變動量較少,是我們考量的依據。

(e)畫面清晰度的考量。當攝影機對焦不正確時,會造成畫面的模糊不清晰,而畫面清晰與否,邊緣(Edge)強度是很好的參考依據,同一物體的邊緣強度在清晰畫面中會比在模糊畫面中來的強。

(f)畫面明亮度對比的考量。當畫面像素色階分佈越均勻時表示此畫面顏色越佳,畫面顏色越均勻也代表此畫面明亮對比較強烈,也表示畫面中帶有較多的資訊量,故利用此特性作為選取關鍵異常畫面之考量。

以上 a、b、c 三點是考量背景相減後得到異常物體的正確性以及完整性,d、e、f 則是考慮畫面的清晰度及對比。接下來將分別對各項考量,說明如何計算關鍵異常畫面特徵值。

(二)關鍵異常畫面特徵值

1、異常物體面積

異常物體面積就是取切割出之前景區域的面積大小,也就是背景相減再二值化成像素值為 0 與 1 的二值圖後,統計圖中不為 0 的像素總數,計算方法如下,

$$Size(n) = \sum_{x,y} I_{x,y} \quad (1)$$

其中, $Size(n)$ 即是第 n 張畫面中的異常物體面積特徵值, $I_{x,y}$ 表示二值圖中在位置 (x,y) 的像素值。

2、異常物體密度

將上一個異常物體面積特徵值除以其框盒面積,計算方法如下,

$$Density(n) = Size(n) / (Width_n \times height_n) \quad (2)$$

其中, $Width_n$ 為包含異常物體之最小框盒

之寬, $Height_n$ 為包含異常物體之最小框盒之高, $Density(n)$ 即是第 n 張畫面中的異常物體面積特徵值。

3、異常物體寬高比

異常物體寬高比也是使用異常物體的框盒寬高比來計算,計算方法如下,

$$Ratio(n) = \begin{cases} 1 & \text{if } 0.25 \leq (Width_n / Height_n) \leq 0.75 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

其中, $Width_n$ 為包含異常物體之最小 Bounding Box 之寬, $Height_n$ 為包含異常物體之最小框盒之高, $Ratio(n)$ 即是第 n 張畫面中的異常物體面積特徵值。

4、畫面變動量

我們使用邊緣變化率(Edge Change Ration, ECR[9])演算法計算第 n 個畫面與第 $n+1$ 個畫面的邊緣變化來表示畫面變動量。畫面變動量表示異常物體在監控畫面中移動之情形,若變動量小表示物體移動速度慢,此時異常物體在畫面中可能是移動緩慢或靜止的狀態,而我們對異常物體在畫面中停留時的動作、情形感到興趣。故畫面變動量小將是我們考慮的因素之一。

ECR 演算法定義兩個邊緣變化率,其一為離開邊緣像素(Exiting Edge Pixel, Outgoing Edge Pixel),此為在畫面 I_n 邊緣像素,卻不存在於下一個畫面 I_{n+1} 之邊緣像素中,也就是原本屬於 I_n 之邊緣像素,在下一張畫面 I_{n+1} 中消失。其二為進入邊緣像素(Entering Edge Pixel, Incoming Edge Pixel),是在畫面 I_n 邊緣像素不存在,卻存在於下一個畫面 I_{n+1} 之邊緣像素中,也就是原不屬於畫面 I_n 之邊緣像素,再下一張畫面 I_{n+1} 中出現。此離開邊緣像素與進入邊緣像素比例關係,可用來偵測場景內容改變的份量。離開邊緣像素與進入邊緣像素值,需經過畫面邊緣計算流程計算出來,此計算流程步驟如下。

步驟一：

將 RGB 全彩畫面轉成灰階影像，而灰階影像值同等於 YCrCb 色彩空間中 Y 值，也就是明亮度元件 (Luminance Component)。

步驟二：

對兩個連續畫面 I_n 與 I_{n+1} 分別做 Sobel 邊緣偵測 (Edges Detection)，計算畫面中邊緣出現的位置與強度。邊緣偵測處理後，此兩個畫面分別為 E_n^{sobel} 和 E_{n+1}^{sobel} ，每一像素值都被調校在 0 到 255 之間。

步驟三：

是進行 E_n^{sobel} 與 E_{n+1}^{sobel} 兩畫面的二值化，用來保留兩影像中邊緣強度大於事先定義閾值的所有像素，判斷方式如下，

$$E_n^{thresh}(x, y) = \begin{cases} 255, & \text{if } E_n^{sobel}(x, y) \geq T \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

步驟四：

利用形態學 (Morphological Operation) 中的膨脹 (Dilation) 運算將 E_n^{thresh} 和 E_{n+1}^{thresh} 這兩個二值化影像中的邊緣增加厚度，處理後影像為 $E_{n+1}^{dilation}$ 和 $E_n^{dilation}$ 。

此後，利用 E_n^{thresh} 與 E_{n+1}^{thresh} 經公式(5)求得離開邊緣像素影像 X_n^{out} ，式中 AND 與 NOT 為二元影像邏輯運算。同理，進入邊緣像素影像 X_{n+1}^{in} ，也可由 E_{n+1}^{thresh} 與 $E_n^{dilation}$ 依公式(6)求得。

$$X_n^{out} = E_n^{thresh} \text{ AND}(\text{NOT}(E_{n+1}^{dilation})) \quad (5)$$

$$X_{n+1}^{in} = E_{n+1}^{thresh} \text{ AND}(\text{NOT}(E_n^{dilation})) \quad (6)$$

計算出 X_n^{out} 與 X_{n+1}^{in} 影像後，可透過比較離開與進入兩影像的變化量，求得離開與進入邊緣像素比例值，取其兩個比例值最大者為畫面 I_n 之邊緣變化率 $ECR(n)$ ，計算式如(7)。

$$ECR(n) = \max(X_n^{out} / E_n^{thresh}, X_{n+1}^{in} / E_{n+1}^{thresh}) \quad (7)$$

5、畫面清晰度

在影片拍攝中會有模糊及對焦不準確的狀況，此時邊緣強度將是我們找尋對焦準確且清晰的依據。我們計算此場景中所有畫面中框盒涵蓋範圍內的邊緣強度來表示異常物體之清晰度值，計算方式可使用前述 ECR 演算法中，Sobel 邊緣偵測及二值化處理後之畫面 E_n^{thresh} 。如此，在邊緣特徵不明顯像素去除後，累計畫面中每一個像素點的值，再計算框盒中的平均強度即代表畫面清晰度。計算方式如下，

$$Edges(n) = \sum_{x,y} E_n^{thresh}(x, y) \quad (8)$$

其中 x, y 為框盒中之像素座標。

6、畫面明亮度特徵

像素在色階上之統計分佈 (Histogram)，是影像中一項重要的特徵。明亮度像素色階分佈圖可幫助了解影像的明亮度像素色階分佈，也就是明亮度空間特性。根據我們實驗結果，色階分佈上若是能夠越趨於平均，此畫面的整體畫面明亮對比會較強烈。而在我們的應用中，有兩項用處，一是希望所偵測出之異常物體帶有較多的資訊量，以排除框盒只框住物體局部的情形；二是拍攝地點如光源不穩定，如建築物出入口，則我們希望留下明亮對比比較大的異常物體畫面。

同樣的我們在意的範圍只在框盒的範圍裡，故色階分佈計算方法，是利用明亮度分佈機率，計算框盒涵蓋範圍內的熵 (Entropy)，以顏色為灰階值(0~255)之顏色分佈來做計算，式子如下，

$$H(n) = - \sum_x p(x, n) \log_2(p(x, n)) \quad (9)$$

其中 $p(x, n)$ 表示在畫面 n 中灰階值為 x 時之統計機率值。 $H(n)$ 值代表整張圖所需要的位元 (Bit) 數，值越大代表資訊量較大也就是色階分佈較均勻。

根據上述六種特徵值，計算影片畫面

n 的異常物體面積 $Size(n)$ 、異常物體密度 $Density(n)$ 、寬高比 $Ratio(n)$ 、畫面清晰度 $Edges(n)$ 、畫面變動量 $ECR(n)$ 、與畫面明亮對比度 $H(n)$ 。計算完畢取得特徵值後，我們將寬高比 $Ratio(n)$ 之外的特徵值做除以每個特徵值中各自出現的最大值作正規化(Normalization)的動作，讓每一個特徵值都可以介於 0 與 1 之間。

經過上述各種特徵，我們可整理成畫面 n 關鍵值 $KeyFrameScore(n)$ ，公式如下，

$$\begin{aligned} KeyFrame(n) = & w_{Size} \times Size^{Normal}(n) + w_{Ratio} \times Ratio(n) \\ & + w_{Density} \times Density^{Normal}(n) + w_{ECR} \times ECR^{Normal}(n) \\ & + w_{Edges} \times Edges^{Normal}(n) + w_H \times H^{Normal}(n) \end{aligned} \quad (10)$$

其中 w_{Size} 、 $w_{Density}$ 、 w_{Ratio} 、 w_{ECR} 、 w_{Edges} 、 w_H 為各項權重因子。在不同考慮因素下，可調整其權重因子。而權重因子的考量要視使用情境而定，如拍攝物體因距離有 Scaling 的情況，則異常物體面積必須小心考慮，因最大面積的異常物體不見得是完整的。同理，畫面變動量也需視情況而定，在異常物體移動速度不變的情形下，畫面變動量較無考慮價值。故在不同拍攝環境下，需考慮不同的權重因子，已達到較好的關鍵畫面擷取結果。

計算出 $KeyFrameScore$ 後，我們選擇 $KeyFrameScore$ 最高的前百分之二十為關鍵異常畫面候選。但這些關鍵異常畫面間有些十分相似，因為相鄰畫面間異常物體並沒有太大的改變。因此我們將關鍵異常畫面候選按照時間順序排序後，選取第一張當做關鍵異常畫面，再依序檢查直到有一張畫面和目前這張關鍵畫面的異常物體差異超過一個閾值時，就選取這張畫面為關鍵異常畫面，並把此畫面的異常物體當作選取下一張關鍵異常畫面的依據。

六、異常物體特徵相似度比對

在多攝影機的監控視訊影像中，同一個人在不同角度拍攝下的形狀特徵可能有

很大的差異，而不同一個人在相同拍攝角度下形狀特徵卻可能十分相似。因此，本論文使用最直覺的顏色特徵，藉由異常物體的穿著衣物顏色來辨別其在不同攝影機拍攝下的相似度。我們採用使用於物件追蹤(Object Tracking)的物體模型與核心函數。當異常物體在監視畫面中移動時，除了形狀的變形外，其面積也會隨著與攝影機的距離跟著改變。所以我們選擇的特徵需要有縮放不變性質(Scale Invariant)，足以抵抗在不同角度拍攝下物體相當程度的變形與面積的改變。在以核心為基礎的物件追蹤(Kernel-based Object Tracking, [4]) 研究中，使用核心函數(Kernel-function)所建立的物體模型十分符合本論文中異常物體特徵的條件。因此在本論文中我們提出一個以顏色為基礎的物件模型比對方法，來做異常物體的相似度比對，讓使用者可以選擇視訊監控畫面中的異常物體對資料庫中所紀錄的異常物體進行搜尋。

在以核心為基礎中(Kernel-based)的物件追蹤方法中使用核心函數(Kernel Function)建立物體模型特徵，其特性是使用物體中每個像素的顏色作為特徵，並將像素座標位置加入考量，建立一個有顏色權重差異的物體模型。其想法來自於，當觀察畫面中一物體時，會發現越靠近物體中心點的像素穩定性越高，物體邊緣附近像素容易受周遭背景影響造成顏色偏差。我們套用此以核心為基礎的追蹤方法中所用之物體模型，抵抗在不同攝影機的不同角度拍攝下物體的變形與面積的改變，建立出一個越靠近物體中心的像素其色彩資訊的重要性越高的物體模型。

建立這類模型是使用核心函數來建立，核心函數會依據每個像素座標位置與中心點的距離來分配其權重，使得越靠近中央的像素顏色資訊越重要。我們使用核心函數為 Epanechnikov Kernel Function。其特性如下，

$$K(x) = \begin{cases} \frac{1}{2} C_d^{-1} (d+2)(1-\|x\|) & \text{if } \|x\| < 1 \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

其中, d 表空間的維度, 在此為二維影像, 故 $d=2$, C_d 表 d 維空間中, 單位圓的面積, 故在此為 π 。

接著說明如何使用核心函數建立物體模型。假設 $\{x_i\}_{i=1\dots n}$ 代表為目標物體中所有像素點位置集合, 中心點為 y 。另定義函數 $b: R^2 \rightarrow \{1\dots m\}$, $b(x)$ 之值為像素 x 所在位置於模型中的色彩索引值 (Color Index)。物體模型為:

$$q(u) = C \sum_{i=1}^n k \left(\left\| \frac{y - x_i}{h} \right\|^2 \right) \delta[b(x_i) - u] \quad (12)$$

$$C = \frac{1}{\sum_{i=1}^n k \left(\left\| \frac{y - x_i}{h} \right\|^2 \right)} \quad (13)$$

其中, $\delta(x)$ 是 Kronecker delta function, C 表正規化常數, h 表半徑長度。(12)式即代表物體模型分布密度函數。 y 是物體的中心點, h 用來讓計算 x_i 與中心點的距離正規化為最大值為 1。(13)式正規化常數的目的是使 $\sum_{u=1}^m q(u, y)$ 為 1 恆成立。

色彩索引值的多寡和所使用的色彩空間 (Color Space) 會影響物體模型的精確度。RGB 是最常見的色彩空間, 但因為每個像素的 R、G、B 三個元素會同時影響到色彩的明亮度, 因此在向量空間中的辨識度並不算良好。因此我們使用在顏色上辨識度較佳的 HSV 色彩模型來建立物體模型, 並考慮運算量以及資料庫中所需儲存的資料量, 將色度 (Hue) 量化成 16 等分、飽和度 (Saturation) 與亮度 (Value) 量化成四等分。也就是使用 $16 \times 4 \times 4$ 共 256 個 bin 來表示物體模型。

當物體特徵模型建立好後, 我們得到的是一個帶有像素座標位置權重的顏色分佈統計 (Color Histogram)。Bhattacharyya Coefficient 可用來量度兩個物體模型也就是此含像素座標權重值的 Color Histogram 之間的相似度。Bhattacharyya Coefficient 定義如(14)式。

$$\rho(t, q) \equiv \rho[t(u), q(u)] = \int \sqrt{t(u) \cdot q(u)} du \quad (14)$$

其中 $t(u)$ 代表 target 物體模型中第 u 個色彩

索引中的量, $q(u)$ 表物體模型中第 u 個色彩索引中的量。由於圖片為離散訊號, 所以 Bhattacharyya Coefficient 離散形式如式 (15)。

$$\rho(t, q) \equiv \rho[t(u), q(u)] = \sum_{u=1}^m \sqrt{t(u) \times q(u)} \quad (15)$$

Bhattacharyya Coefficient 值越大, 表示兩個物體模型越相似, 反之 Bhattacharyya Coefficient 值越小表示兩個物體模型越不類似。故在模組中藉由 Bhattacharyya Coefficient 計算兩個物體模型的相似度。

七、實驗結果與討論

我們的測試實驗方法, 是在四個地點使用固定攝影機各自拍攝一段時間, 並讓數個測試目標在四個地點輪流出現。拍攝完畢後, 將影片一一經由系統進行異常偵測, 以及關鍵異常畫面擷取和特徵擷取儲存於資料庫中。最後藉由查詢曾在四個地點出現的目標人物, 看是否能藉由本論文提出之方法, 找出衣服顏色相近之異常物體, 以便縮小所需檢視範圍, 快速的在監控視訊資料庫中找到所關注的特定異常物體, 並藉由在攝影機中出現的時間順序找出異常物體行經之路線。此系統開發環境為 Window XP Professional Server Pack 2 作業系統、Intel Pentium4 3.2 GHz 處理器、DDR 1G 記憶體、WD 7200 RPM 250G 硬碟。開發工具是 Microsoft Visual Studio 2005 以及 OpenCV, 使用程式語言為 C++, 資料庫採用 Microsoft Access 2003。實驗影片是 Intel Indeo R5.1 壓縮格式, 解析度是 640×480 , 每秒 30 張畫面。

當使用者瀏覽異常事件後, 想要找出資料庫中某特定異常物體, 可以使用異常物體查詢功能。使用者可以按下去系統介面下方除背景鈕, 讓系統取得對應的背景影像以便減去背景 (結果如圖 4 系統介面左上方畫面), 方便減少背景對特徵擷取時的干擾。使用者勾選系統介面下方異常物體查詢的 Check Box 後, 便可以框選想查



圖 3 搜尋結果輸出介面

詢的異常物體。框選後按下查詢鈕，系統會將查詢結果，以相似度依序排列於查詢結果列表(圖 3 右下方)。而使用者可以選擇信心閾值，將相似度在信心閾值以上的查詢結果留下，此時按下顯示路徑鈕，則系統會依照異常事件發生時間重新排序，並依照攝影機編號推算出異常物體行經路線(圖 3 右方)。

利用相似度比對方法，在下面的例子中我們在框選異常物體並進行查詢後，成功的找出在資料庫中所有該異常物體出現的異常事件。圖 4 (a)是 Query Image，(b)、(c)、(d)、(e)分別是從資料庫中找到最相近的前四個異常事件的異常物體，(b)與(a)是同一段影片中的影像，相似度達 0.945；(c)與(a)相似度為 0.85，用肉眼已可看出在亮度上有所差異；而(d)、(e)除了亮度的關係再加上下半身顏色資訊較少，與(a)相似度已小於 0.8。

我們所使用的相似度比對法有一定的準確度，但使用顏色當作唯一特徵的缺點

是當有不同異常物體穿著相似顏色時，必定會造成混淆。此外，光線也是很重要的影響因素。光線的影響從圖 5 的例子中更可明顯看出，(a)與(b)相似度是 0.87，(b)與(c)則只有 0.76。

八、結論

本論文提出一個運用於保全監控系統的視訊內容檢索搜尋方法架構。此方法使用背景相減(Background Subtraction)方式偵測異常物體出現的事件段落，並藉由關鍵異常畫面擷取(Key Frame Extraction)演算法，利用異常物體面積、異常物體密度、異常物體寬高比、畫面變動量、畫面清晰度、畫面明亮對比等六種特徵，過濾選取出可明顯表現出異常物體特徵之關鍵異常畫面，儲存於視訊資料庫中。保安全管理者可在監控視訊資料的瀏覽過程中，以時間或攝影地點等條件查詢異常事件的畫面，找到可疑並欲搜尋的異常物體後，可經由自行圈選出此異常物體，並以我們設計的



圖 4 相似度比對查詢結果

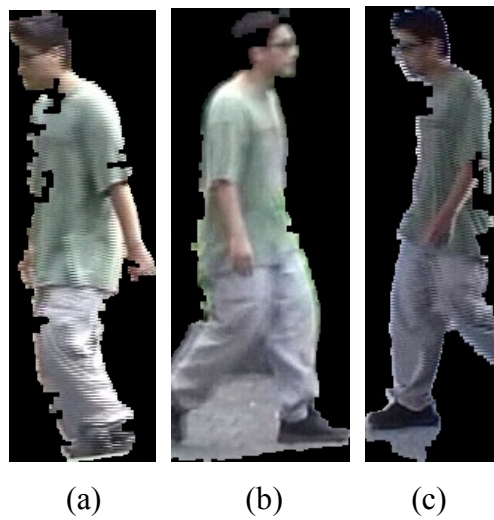


圖 5 受光影影響之相似度比對

使用 Epanechnikov Kernel Function 為核心建立異常物體之物體模型(Object Model)的方法，與監控視訊資料庫中所儲存的，關鍵異常畫面中異常物體的物體模型，計算出 Bhattacharyya Coefficient 來度量異常物體之間的相似度，以進行比對。搜尋時，可按照相似度排列其結果，並以異常事件發生時間，得知異常物體的行經路徑，以建立出一個完整的監控視訊系統資料庫中異常物體搜尋之方法。

本論文提出之方法，經實驗測試證明，可以讓保全管理者有效率的瀏覽查詢監控系統視訊資料庫中的異常事件，異常物體查詢方法也可快速的縮小所需檢視範圍，有效的減少監看錄影畫面時間。

九、參考文獻

- [1] 李冠穎，「視訊片段之代表性圖片選擇法」，國立台北科技大學資訊工程所論文，民國 93 年。
- [2] S. Antani, R. Kasturi, and R. Jain, "A Survey on the Use of Pattern Recognition Methods for Abstraction, Indexing, and Retrieval of Images and Video," *Pattern Recognition*, Vol. 35, No. 4, pp. 945-965, 2002.
- [3] H. S. Chang, S. Sull, S. U. Lee, "Efficient Video Indexing Scheme for Content-Based Retrieval," *IEEE Trans Circuits Sys. Video Techn.*, Vol. 9, No. 8, pp. 1269-1279, 1999.
- [4] D. Comaniciu, V. Ramesh, and P. Meer,

- “Kernel Based Object Tracking,” *IEEE Trans. Pattern Analy. Machine Intel.*, Vol. 25, No.5, pp. 564-557, May 2003.
- [5] W. Hu, T. Tan, L. Wang, and S. Maybank, “A Survey on Visual Surveillance of Object Motion and Behaviors,” *IEEE Transaction on System, Man, and Cybernetics—Part C: Applications and Reviews*, Vol. 34, No. 3, Aug., 2004.
- [6] S. Jabri, and et al., “Detection and Location of People in Video Images Using Adaptive Fusion of Color and Edge Information,” *Proc. Intl. Conf. on Pattern Recognition*, Vol. 4, pp. 627-630, 2000.
- [7] A. J. Lipton, H. Fujiyoshi, and R. S. Patil, “Moving Target Classification and Tracking from Real-time Video,” *Proc. of the 4th IEEE Workshop on Applications of Computer Vision*, pp. 8-14, 1998.
- [8] D. Meyer, J. Denzler, and H. Niemann, “Model Based Extraction of Articulated Objects in Image Sequences for Gait Analysis,” *Proc. IEEE Int. Conf. Image Processing*, pp. 78-81, 1998.
- [9] R. Zabih, J. Miller, and K. Mai, “A Feature Based Algorithm for Detecting and Classifying Scene Breaks,” *Proc. of ACM Multimedia 95*, San Francisco, CA, pp. 189-200, Nov. 1995.
- [10] H. J. Zhang, J. Wu, D. Zhong, and S. W. Smoliar, “An Integrated System for Content-based Video Retrieval and Browsing,” *Pattern Recognition*, Vol. 30, No. 4, pp. 643-658, 1997.
- [11] P. Remagnino, A. I. Shihab and G. A. Jones, “Distributed Intelligence for Multi-Camera Visual Surveillance,” *Pattern Recognition*, Vol. 37, pp.675-689, 2004.
- [12] J. Black, D. Makris and T. Ellis, “Hierarchical Database for Multi-Camera Surveillance System,” *Pattern Analysis Application*, pp. 430-446, 2005.