

行政院國家科學委員會專題研究計畫 成果報告

人體運動之智慧型多相機監視系統(1)

計畫類別：個別型計畫

計畫編號：NSC91-2213-E-009-092-

執行期間：91年08月01日至92年07月31日

執行單位：國立交通大學資訊工程研究所

計畫主持人：陳稔

報告類型：精簡報告

報告附件：出席國際會議研究心得報告及發表論文

處理方式：本計畫涉及專利或其他智慧財產權，2年後可公開查詢

中 華 民 國 92 年 11 月 10 日

行政院國家科學委員會補助專題研究計畫成果報告

人體運動之智慧型多相機監視系統

計畫類別：個別型計畫 整合型計畫

計畫編號：

執行期間：91年8月1日至92年7月31

計畫主持人：陳 稔 國立交通大學資工系教授

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：交通大學資訊工程學系

中 華 民 國 九 十 二 年 十 一 月 七 日

人體運動之智慧型多相機監視系統

Intelligent Multi-Camera Surveillance System for Human Motion

計畫編號：NSC 91-2213-E-009-092-

執行期限：91 年 8 月 1 日至 92 年 7 月 31 日

主持人：陳稔 國立交通大學資工系教授

E-mail: zchen@csie.nctu.edu.tw Tel: 03-5731875

1. 摘要

本計畫提出一套智慧型多相機監控系統，各相機皆具有 pan/tilt/zoom 的功能。系統可透過網際網路取得相機影像，對場景中的移動人物進行偵測與追蹤，在事件發生時，可控制其中一部特寫用的相機，對該事件發生的位置拍攝局部近照。這形同以專人用人眼做監控工作，且由於使用網際網路作為傳輸與控制的媒介，可突破時間與空間的限制，達成隨時隨地的監控效果。

關鍵詞：智慧型視訊監視系統、相機校正、場景變動偵測與追蹤、事件偵測

Abstract

In this research we construct an intelligent video surveillance system containing several networked TV cameras with pan/tilt/zoom capability. Through the internet, the system can remotely detect and track moving objects in the scene. When an event happens, the system can control one of the cameras to take a close-up shot of the event. As a result, the traditional manual visual alert can be replaced by the system. Also, with the aid of the internet, the system works at any time and any place.

Keywords: Intelligent video surveillance system, Camera calibration, Change detection and tracking, Event detection.

2. 緣由與目的

視覺監視系統 (Vision-based surveillance system) 常見於目前社區安全監控室，保全公司監控室，道路交通安全監控室，以及銀行大廳及提款機等的閉路攝影機 (CCTV) 監控。此種視覺監視系統大致可分為三代 [1]。第一代 (1960-1980) 是以 analog 形態為主，感測輸入端的 analog picture 經由 analog transmitter 及 receiver 或 coaxial

ohm cable network 傳到接收端，再由顯示器播放，以人眼監視；儲存時以 analog video tapes 儲存。由於 video 信號未做壓縮，以及以人眼監控，故效率不高。第二代 (1980-2000) 由於低成本的計算機出現，video signal 開始可做數位化壓縮及處理，而且開始有較大頻寬的 communication networks 可以傳輸。到了第三代 (2000 之後) 更多樣化更寬頻的網路如 cable modem, X-DSL modem, 以及光纖等有線網路以及 IEEE 802.11 等無線網路的出現，使得未來 WAN/LAN 傳送影像更快更方便。全數位化 (full digital) 的視覺監視系統即將來臨。此時，高度自動化的智慧型視覺監視系統可提供給人類更安全，更便利的生活品質。

目前國內有廠商提供視訊監視系統，但多半是硬體設備，可搭配的應用軟體有限。國外如歐美，日本則有些大學及公司研究單位在研究新穎的系統，例如微軟公司的 EasyLiving，美國 CMU 大學的 DARPA VSAM project [2]，馬里蘭大學的 W⁴ [3]，麻省理工的 Pfinder [4]、LOTS [5]、TI [6]、SRI [7]。此外，此方面的學術會議及專輯逐年受到重視。

本計畫的目的是希望使用已有的硬體與技術，設計一套監控系統，具備有即時 (real-time)、韌性 (robust)、精確 (accurate)、自動 (automatic)、智慧 (intelligent)、網路化 (networked) 等特質。能對固定場景中的人物進行追蹤、拍攝特寫鏡頭並偵測簡易事件的發生。監控系統具備這些能力後，可以取代絕大部份人工監視作業，達到無時間限制 (隨時)、無空間隔離 (隨地) 的監視。此種系統未來應用將更蓬勃發展。

3. 研究方法

為達成上述目標，我們架設了一套可透過網際網路控制的多攝影機系統，每部攝影機並具有 pan/tilt/zoom/auto-focus 的能力。初期我們僅使用其中的二部攝影機，一部用來監控大範圍的全域場

景，另一部則根據系統的指示，對人物或發生的事件拍攝特寫鏡頭。

監控系統的主要流程可分為以下幾個步驟：

- 相機校正(Camera calibration)：用以建立場景所在的世界座標系(World coordinate system)與影像座標系(Image coordinate system)之間的對應關係。
- 場景變動偵測(Change detection)：偵測場景中新植入的物件，亦即將攝影機影像中的前景(人或任何需要監控的事物)與背景(固定不變的場景)分離。將前景找出後，便可交由後續的步驟作進一步的處理。
- 追蹤與事件偵測(Tracking and event detection)：對場景中變動的物件繪出其移動軌跡，並將有特殊事件發生的影像標示出來。
- 拍攝特寫鏡頭：改變特寫用的攝影機之 pan/tilt/zoom 等參數，對場景中的特殊事件或人物拍攝特寫畫面。

以下分各小節將對上述四個步驟作更詳細描述。

3.1 相機校正

3.1.1 求取內部參數(Intrinsic Parameter) K

傳統的相機內部參數求取是先以預先設計好的校正物(Calibration object)估算出相機的投影矩陣 $H_{3 \times 4}$ ，再從其中分解出相機內部參數矩陣 $K_{3 \times 3}$ 。但這個方法的缺點在於需有定位精準的校正物的協助才能算出準確的投影矩陣 $H_{3 \times 4}$ ，也才能確保之後算出的相機參數的準確度。需要額外的校正物違背我們希望整個流程自動化的目標。我們希望能利用已有的影像內的資訊直接求得相機校正矩陣。

在 Hartley [8] 所提出的方法中，使用三張不同角度、固定 COP 拍攝的影像，求得兩個平面投影轉換矩陣(Homography)，即可推出相機的內部參數矩陣 K。如下式：

$$P_j \cdot K \cdot K^T = K \cdot K^T \cdot (P_j)^{-T}$$

其中 P_j 為任兩張影像間的平面投影轉換矩陣，代入兩個以上的 P_j 解線性方程組便能得到 $K \cdot K^T$ ，再根據 K 為上三角矩陣的性質，便能求得相機的各項內部參數。

我們使用 Gabor filter 自動選取點對應，剔除

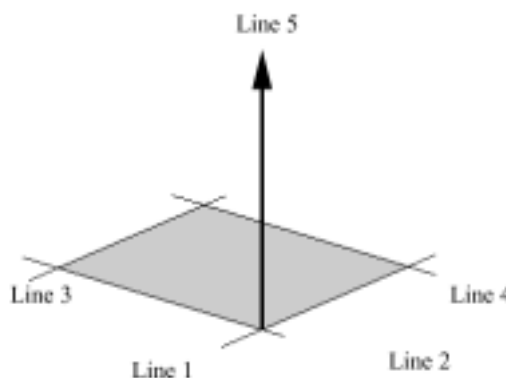
Outliers 的干擾後，再搭配使用 Levenberg-Marguardet 非線性遞迴演算法[9]可自動化求得最佳的平面投影轉換矩陣，如此可增加上述方法求得 K 的準確度。

另外，由於攝影機具備調整 zoom lens 的功能，我們必須對每一個可能的 focal length 事先求得其對應的內部參數 K 並記錄成表格，系統實際運作時，直接查表即可，不需要重複進行上述運算。

3.1.2 相機旋轉矩陣 R

在前一節求得相機內部參數矩陣 K 後，我們即可利用該矩陣將座標系由 Image coordinate system (簡稱 ICS) 轉到 Camera coordinate system (簡稱 CCS)，若再配合事前定義好的 World coordinate system (簡稱 WCS)，我們即可定義出 WCS 和 CCS 間的轉換關係。以下是我們的做法：

Step 1: 在場景地面上標示出四條直線，Line 1、Line 3 為 WCS 下互相平行的直線，Line 2 和 Line 4 為另一組平行的直線，如 (圖 1) 所示。其中 Line 1、Line 2 和其外積而成的第三個方向(Line 5)為我們定義的 WCS。



(圖 1) 定義 WCS 示意圖

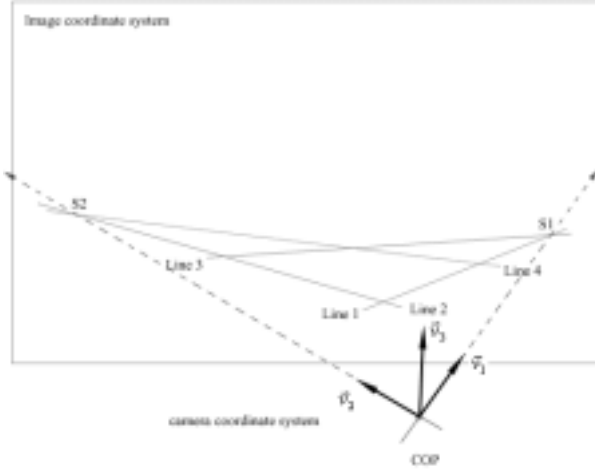
Step 2: 拍攝含有校正物的場景，並使用上一節描述的方法求得相機內部矩陣 K。

Step 3: 因為我們採取的相機模型為透視投影(perspective projection)，所以 WCS 中兩條平行的直線會在相交於 ICS 中無限遠的某一點(Vanishing point)，所以我們在影像上的 Line 1 及 Line 3 上取數個點，算出其最逼近的直線方程式，並求其交點 S_1 。

Step 4: 利用 K 將 S_1 由 ICS 轉至 CCS，則 COP 至 S_1 的方向為即為 CCS 中 Line 1 的方向，令其上的單位向量為 \hat{v}_1 。

Step 5: 重覆以上步驟求得 CCS 中 Line 2 的方向，令其為 \vec{v}_2 。

Step 6: 令 $\vec{v}_3 = \vec{v}_1 \times \vec{v}_2$ 。若 $\vec{v}_1 \cdot \vec{v}_2 \neq 0$ ，令 $\vec{u}_2 = \vec{v}_3 \times \vec{v}_1$ 。Step 3~6 示意圖如 (圖 2)。



(圖 2) 利用消失點求取旋轉矩陣

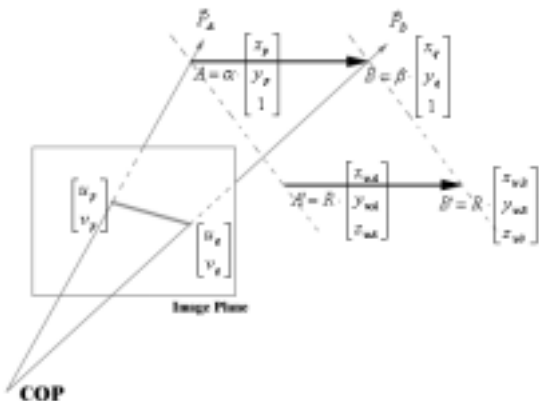
Step 7: 則 CCS 與 WCS 間存在下列關係。

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = R \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} + \vec{t}_{3 \times 1}, \text{ 其中 } R = \begin{bmatrix} \vec{v}_1 & \vec{u}_2 & \vec{v}_3 \end{bmatrix}$$

如此,我們即估算出 WCS 與 CCS 間的旋轉矩陣 R。

3.1.3 相機位移向量 $\vec{t}_{3 \times 1}$

在算出 WCS 與 CCS 相對旋轉矩陣 R 後,若直接將 WCS 的座標乘上 R,則和實際的 CCS 座標間相差一個位移 $\vec{t}_{3 \times 1}$ 。但線段長度和實際 CCS 內的線段長度相度且具平行的關係(因整體皆差一個位移)。而實際 CCS 的座標可以 ICS 座標投過相機校正矩陣 K 得到。利用已知一條線段的長度,我們試著估算出 WCS 與 CCS 間真正的位移值。



(圖 3) 求取位移向量

如 (圖 3) 所示, $A = \alpha \cdot K \cdot \begin{bmatrix} u_p \\ v_p \\ 1 \end{bmatrix} = \alpha \cdot \begin{bmatrix} x_p \\ y_p \\ 1 \end{bmatrix}$ 、

$B = \beta \cdot K \cdot \begin{bmatrix} u_q \\ v_q \\ 1 \end{bmatrix} = \beta \cdot \begin{bmatrix} x_q \\ y_q \\ 1 \end{bmatrix}$ 分別為影像上兩點 $\begin{bmatrix} u_p \\ v_p \end{bmatrix}$

和 $\begin{bmatrix} u_q \\ v_q \end{bmatrix}$ 透過 K 轉到到 CCS 上的座標。因透過 K 將

座標由 ICS 轉到 CCS 會差一個倍數(scale factor),

所以我們在前頭分別乘上 α 和 β 。而 $A' = R \cdot \begin{bmatrix} x_{wA} \\ y_{wA} \\ z_{wA} \end{bmatrix}$,

和 $B' = R \cdot \begin{bmatrix} x_{wB} \\ y_{wB} \\ z_{wB} \end{bmatrix}$ 分別為 WCS 中兩點 $\begin{bmatrix} x_{wA} \\ y_{wA} \\ z_{wA} \end{bmatrix}$, 和

$\begin{bmatrix} x_{wB} \\ y_{wB} \\ z_{wB} \end{bmatrix}$ 透過旋轉矩陣 R 轉到 CCS (差一個位移)的座

標。由以上假設我們可得到兩組向量

$$\vec{AB} = \begin{bmatrix} \beta \cdot x_q - \alpha \cdot x_p \\ \beta \cdot y_q - \alpha \cdot y_p \\ \beta - \alpha \end{bmatrix} \text{ 和 } \vec{A'B'} = R \cdot \begin{bmatrix} z_{wB} - z_{wA} \\ z_{wB} - z_{wA} \\ z_{wB} - z_{wA} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}$$

且 $\vec{AB} \parallel \vec{A'B'}$ 、 $\|\vec{AB}\| = \|\vec{A'B'}\|$ 。將此二條件代入,即可求得 α 與 β 的值。

有了 α 和 β 的值之後,我們便能求出 $\begin{bmatrix} x_{wA} \\ y_{wA} \\ z_{wA} \end{bmatrix}$ 在

CCS 中實際的座標值 $A = \alpha \cdot K \cdot \begin{bmatrix} u_p \\ v_p \\ 1 \end{bmatrix} = \alpha \cdot \begin{bmatrix} x_p \\ y_p \\ 1 \end{bmatrix}$, 而

它和 $R \cdot \begin{bmatrix} x_{wA} \\ y_{wA} \\ z_{wA} \end{bmatrix}$ 的差值即為 WCS 和 CCS 間的位移

量。

3.2 場景變動偵測

智慧型視訊監控系統最根本的一個步驟,就是將攝影機攝得的影像之前景(人或任何需要監控的事物)與背景(固定不變的場景)分離。為了提高監控系統的適應性及容錯能力,變動偵測(Change detection)必須具備以下幾點特點:

- (i) 抗噪能力: 系統必須能容忍現有的低解析度 NTSC 攝影機所帶來的無可避免的

影像雜訊。

- (ii) 重複移動的微小變化(repetitive small motion)，如隨風搖曳的樹葉，應被視為靜止的背景。
- (iii) 光源的照明度變化：真實的場景會因為光照的不同，使得不同時間所攝得的攝影機影像在亮度上有所差異。
- (iv) 陰影：人物在場景中移動時，會因為光照的影響而產生陰影，系統必須能自動將這些陰影排除，僅對實際的人物作處理。
- (v) 偵測的人體區塊盡可能不要破碎：偵測出的人體區塊越完整，對後續的應用，如活動模式的分析與辨識，就更有利。

在本計畫中我們混和使用 Pixel-based statistical model [4][10]及 Block-based vector model [11] 以達成上述功能。

首先，我們必須取得一系列的 Image，一開始時沒有任何的 Foreground 只是很單純的 Background，我們將這些純 Background Frame，當成我們一開始的 Background statistical model，接著 Object 開始進入 Scene，我們將得到帶有 Foreground 的 Frame，接下來開始進行 Background Subtraction。

我們首先引用 L. Davis 等人所提出的 Non-parametric model [12] 來作初步的前景與背景分離。方法如下：

3.2.1 初步分析前景與背景

利用統計的方法粗略的把前景與背景分開。假設 x_1, x_2, \dots, x_t 為一 Pixel 最近幾個 Samples 之 Intensity Values，我們利用以下公式計算該 Pixel 在 Time t 的 Intensity Values x_t 的機率值

$$\Pr(x_t) = \frac{1}{N} \sum_{i=1}^N \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x_t - x_i)^T \Sigma^{-1} (x_t - x_i)}$$

$$\text{其中 } \Sigma = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix}$$

假設 RGB 每一個 color channel j ，都有一個不同的 Kernel Bandwidth σ_j^2 且它們之間是相互獨立的，原式可以化簡成

$$\Pr(x_t) = \frac{1}{N} \sum_{i=1}^N \prod_{j=1}^d \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{1}{2} \frac{(x_{tj} - x_{ij})^2}{\sigma_j^2}}$$

而假如 $\Pr(x_t) < \text{th1}$ ，則我們就說該 Pixel 在 Time Frame t 時視為 Foreground Pixel。我們在定 th1 必須考量到下列因素：

1. 必須決定到底停留多久才叫 Background。
2. 在定 th1 的時候，必須讓 Background Model 能夠容許些微的變化，而不是只要有一小點的變化，就立刻被判定成 Foreground。
3. 訂立 th1 準則：停留的時間太長或太短均不應該被視為 Foreground Pixel。

但是，我們會發現有些 Pixel 只是很細微的來回擺動(譬如說受到樹葉飄動的影響)，不是我們想要的 Foreground，我們可以利用下一小節做 False Detection 的修正。

3.2.2 修正 False Detection

為了解決上述重複細微的變動，就被判定成 Foreground Pixel，採用以下方法作修正：

$$\text{定義： } P_N(x_t) = \max_{y \in N(x)} p(x_t | B_y)$$

x_t : a certain pixel has intensity x_t at time t

y : x 的 neighborhood

B_y : x 的 neighborhood 中為 background

$$\text{原式可改寫 } = p(x_t | y_{x, \max})$$

$$= \max \{p(x_t | y_1), p(x_t | y_2), \dots, p(x_t | y_{N(x)})\}$$

上面的式子簡單的說，就是對於 Foreground Pixel x 而言，其 Neighborhood 為 Background 的 Pixel 中，找出其 Probability 最大的，而且不僅要最大，還要大於 th2 這是因為若 Foreground Pixel x 原本是 Background，則該點與鄰近亦為 Background 的點 Probability 會很相近。

$$\text{定義： } P_C = \prod_{x \in C} P_N(x), \text{ C: Foreground Connect}$$

Component

$$= P_N(x | x \in C) * P_N(x_1 | x_1 \in C) * P_N(x_{|C|} | x_{|C|} \in C), |C| = \text{C 的個數}$$

$$= P_N(x | y_{x, \max}) *$$

$$P_N(x_1 | y_{x_1, \max}) * P_N(x_{|C|} | y_{|C|, \max})$$

上面的式子簡單的說，就是把與 Pixel x 相連的 Foreground Pixel 之 P_N 相乘。假如 P_C 大於某個定值 th3 ，表示他鄰近的 Pixel 也受到了其鄰近 Background Pixel 的影響，則 Pixel x 更肯定是 Background Pixel。反之假如沒有大於 th3 ，表示他的確是 Foreground，只是恰巧與其鄰近 Background Pixel 的顏色相近，

造成 Pixel x 在 P_N 階段被視為 Background, 但是在 P_C 階段時, 雖然 Pixel x 的 P_N 很大, 但他周圍鄰近的 Foreground Pixel 的 P_N 均很小, 所以 P_C 值不會超過我們所給定的 th3, 因此仍會被判定成 Foreground, 不會受到 P_N 的影響。

接下來, 我們就要處理有關陰影的問題了。在監控室外場景時, 在不同的時刻光照的角度也會跟著改變, 隨之衍生的問題是因不同光照角度場景中的物體投射的陰影也會跟著改。在進行場景變化偵測時這些陰影會被一併抓出, 但實際上這些陰影應該算是場景的一部份。故我們希望能排除陰影變化的影響。另一方面, 若能移除陰影的變化, 也能增進目標物追蹤的準確度。

3.2.3 去除陰影及光照變化的影響

陰影的形成, 我們可以視為在陰影所在區域光照度減弱的結果, 而在白天的室外場景, 主要的光源來自於太陽光, 因為太陽光可以視為平行光, 所以在陰影形成的範圍內的光照度會呈現等量的減弱, 換句話說, 在陰影的範圍內只有一致的光照量變化。對於這個部分我們將採取利用 Block-Based 的方式來解決, 而不用 Pixel-Based 的方式。

我們改良 E.Durucan 等人在[11]中提出的 Vector model, 作法如下。

首先定義 Vector model: 將影像上的每一點, 以其周遭 $n \times n$ 大小範圍內的像素的亮度值排列而成的 vector 代表之。一個 3×3 的例子如下式:

$$\begin{array}{|c|c|c|} \hline I_{1,1} & I_{1,2} & I_{1,3} \\ \hline I_{2,1} & I_{2,2} & I_{2,3} \\ \hline I_{3,1} & I_{3,2} & I_{3,3} \\ \hline \end{array} \quad V_{2,2} = \begin{bmatrix} I_{1,1} \\ I_{1,2} \\ I_{1,3} \\ I_{2,1} \\ I_{2,2} \\ I_{2,3} \\ I_{3,1} \\ I_{3,2} \\ I_{3,3} \end{bmatrix}$$

跟據光照的理論, 一物體上某一點的亮度 (intensity) I_o 大致上可以視為光照量 I_i 和物體表面反射係數 S 的乘積[13], 即

$$I_o = I_i * S$$

今在兩不同的光照條件下

$$I_{o1} = I_{i1} * S \quad \text{及} \quad I_{o2} = I_{i2} * S$$

對同一物體上同一點而言, 反射係數 S 不變的話, 我們可以導出下面的關係

$$\frac{I_{o1}}{I_{o2}} = \frac{I_{i1}}{I_{i2}}$$

所以說該點的亮度值與光照量成等比例。若我們同時考慮兩張相同場景對應的 window 內所有像素的亮度值的話, V_1 為代表第一個 window 的向量, V_2 為代表第二個 window 的向量, 若場景僅僅是光照量發生變化, 根據上面的關係我們可以寫出下列式子

$$V_1 = \begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_8 \\ I_9 \end{bmatrix}, \quad V_2 = \begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_8 \\ I_9 \end{bmatrix} = \begin{bmatrix} \alpha \cdot I_1 \\ \alpha \cdot I_2 \\ \vdots \\ \alpha \cdot I_8 \\ \alpha \cdot I_9 \end{bmatrix} = \alpha \cdot \begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_8 \\ I_9 \end{bmatrix}$$

我們可以看出若兩張影像對應的 window 內僅有光照度發生改變的話, 所代表的兩個 vector 以向量的角度來看則呈現一平行的關係。

所以, 要決定兩張影像中的同一點是否有變化, 可藉由檢查對應的兩個 vectors 是否為平行。若兩點的值不相等但兩 vectors 具有平行關係, 表示該點週圍其他點的值也呈等倍數的增減, 即意味著可能是整個環境的光照度發生變化, 該點可以視為沒有變化, 可歸類為背景。相反的, 若兩向量非平行, 即代表該 $n \times n$ 範圍內的值的變化並不一致, 也就是說, 有可能是外物所在的位置。

檢查平行的方法是利用向量內積的性質。我們先將兩向量各別取其同方向的單位向量, 之後取其內積, 內積所得值再利用 \cos 函數即可求得兩向量間的夾角。

整合 Non-Parametric statistical model 與 Vector model 之步驟為:

STEP.1: 先用 Non-parametric model 作初步的判斷找出 Foreground pixel 及各 Pixel 之 Background model's distribution。

STEP.2: 將 STEP1 所得之 Foreground Pixel, 使用 Vector Model 的方法, 判斷該 Pixel 是否為 Shadow 或光影所造成的, 若是, 則將該 Pixel 改為 Background。(用 STEP.1 中所估得的 Short-Term Background Model Distribution 之 Mean Value, 作為 Vector Model 中的 Background Pixel Value)

最後一步, 就是我們 Background 必須做更新, 來保持我們偵測出來前景的正確性。

3.2.4 更新 Background Model 的方法

(Connected component), 以此將所有的前景分類成以區塊為單位。

- b. 在視覺監視系統中我們有興趣的是人體的追蹤，而在影像偵測中，所偵測出來的人顯少會有分成兩截，因此偵測出來的前景所構成的區塊必定不會太小塊。太小塊的區塊為之前影像偵測的不完善而造成的誤差，因此必須將小塊的區塊去除掉，留下大塊的區塊即為所需要做追蹤之物體。

步驟 2：物體之追蹤。

將影像中屬於前景的區塊切出後，必須對各個區塊作連續的追蹤，並將零散的區塊整合為完整可代表人體的區塊。我們採用區塊的色彩特徵與空間關連性，決定連續影像中的區塊是否為同一體，再對同一區塊追蹤一段時間後，便能分析出其運動特性(移動方向、速度等)。

當物體尚未完全進入或離開場景，此時物體不完全，因此暫時不追蹤物體。而判斷物體是否完全進入或離開場景可用影像偵測所之結果來決定，若區塊落在場景邊界，且經過連續偵測時區塊面積有持續變大或變小表示物體正在進入或離開場景。

物體重心(centroid)的移動軌跡可視為是物體移動之軌跡，因此計算出所有區塊的重心位置，再到原圖取得區塊之色彩，然後根據區塊重心之位置及移動方向和色彩特徵來偵測物體移動之軌跡。而在影像偵測中難免會有因偵測誤差而使得同一個物體被分成兩個區塊，但基本上，只要是同一個物體，其移動方向、速度就會相同，若兩個以上的重心其移動方向、速度差不多，在空間中的相對位置相距不大且幾乎固定不變，則可視為同一物體而將之群組化(Grouping)。

步驟 3：事件偵測。

對於事件之偵測，我們可利用群組化之後區塊的大小及個數來判斷事件的發生，通常只要有事件的發生通常都會造成區塊個數的增加或減少，或是區塊大小的改變。如實驗成果中，人提著袋子，所以偵測出來群組化之後的塊區是整個人提著袋子的區塊，當人將袋子放下後，袋子和人分開，區塊會由一個變成二個，一個是人的區塊、一個是袋子

的區塊，且區塊大小變小，我們即可判斷出有事件發生。並透過下一節的方法對有興趣的區塊(如掉落的袋子)拍攝特寫近照。

3.4 拍攝特寫鏡頭

我們使用全域攝影機監控整個場景，透過 3.2、3.3 節的方法，可偵測出特殊事件發生在 ICS 中的位置。假設事件皆發生在 WCS 中某一事前定義的平面上(例如：地平面， $Z_{WCS}=0$)，由於攝影機已校正完成，故只需將事件發生點的投影射線(Projection ray)與該平面相交，即可求得事件發生在 WCS 的實際座標位置。

同理，將特寫攝影機的 COP 與事件發生點連線後，即可得到特寫攝影機新的光軸向量，求此向量與舊光軸的夾角，便能得之該如何調整 pan/tilt 參數，方能攝得正確的特寫鏡頭。

4. 實驗結果

4.1 拍攝影像

分別在戶外取景(場景包含細微變化的物體)，及室內取景，並作簡易追蹤。實驗的目標為：
(1) 展現 3.2.2 節中 P_c 的效果：對於重複變化的場景(如樹葉)，能夠有效的去除，只留下我們真正關心的前景。

(2) 主角提著一件東西，走到場景中央後丟下(讓該物體繼續靜置在場景中，我們希望該物體在經過一段合理的時間後，會融入 Background Model 中)，之後主角繼續往前走。

(3) 場景亮度突然改變(此項實驗可在室內進行，因為我們注重的是場景特殊事件的發生，因此我們可以找室內進行實驗，排除其他影響實驗的因素)。

4.2 實驗成果

● Change Detection

這部分我們共有兩組實驗：

第一組為戶外取景(含有一個 model，該 model 手中提著一個物體，在場景中移動，該 model 走到場景中央時把該物體放下，繼續前進，我們希望能夠偵測出移動的物體，而該物體在被放下後，一開始時候該物體被偵測為前景，後來經過一段時間後，由於該物體待的時間夠久了，足以影響 Background Model 所以後來就融入 Background 中，而消失不見)，見(圖 5)。

在(圖 5)中，A1~A5 為一序列的原始影像，

B1~B5 是將 A1~A5 做初步判斷所得到的影像(因為圖形縮小,所以需放大才能看的較為清楚),C1~C5 是將 B1~B5 所得出的結果再做去除 Repeat Motion 而得出到的影像,D1~D5 是將 C1~C5 所得出的結果再做去除陰影、光線變化而得出的最終結果的影像。

我們可以從 A1 很清楚的看到一開始的場景只有背景,而我們偵測出來的最終結果 D1 確實沒有任何前景的部份,而隨著人探著袋子走進來,我們也能很清楚的偵測到人提著袋子的部份,到了 A5,人已經走出場景,且袋子也在場景中放了一段時間,所以袋子已經慢慢的被 update 成背景的一部份,因此在 D5 所得出來的結果中沒有任何前景的部份。

第二組實驗為室內取景(基本的情況和第一組實驗大致相同,只要差別為我們希望實驗看看當有特殊事件發生(譬如我們將場景的燈光突然整個打亮或打暗)時我們的程式是否依舊可以偵測出移動的物體,為了專注於我們的焦點,也就是特殊事件的發生,我們選擇室內沒有樹葉的飄動來進行取景,以增加實驗的速度),見(圖 6)。

在(圖 6)中,A1~A4 為一序列的原始影像,B1~B4 是將 A1~A4 做初步判斷所得到的影像,C4~C4 是將 B1~B4 所得出的結果再做去除 Repeat Motion 而得出到的影像,D1~D4 是將 C1~C4 所得出的結果再做去除陰影、光線變化而得出的最終結果的影像。

我們可以從 A1 很清楚的看到一開始的場景只有背景,而我們偵測出來的最終結果 D1 確實沒有任何前景的部份,隨著人提著袋子走進來,我們也能很清楚的偵測到人提著袋子的部份,而到了 A3 時,燈光突然變暗,使得在初步判斷出來的結果產生極大的誤差,但因為是整個場景的光線都變暗,所以經過 D3 處理光照度變化後的結果我們可以發現絕大部份因光線變化所產生的前景已被去掉。圖 A4 中人已經走出場景,光線也持續暗了一段時間,所以我們可以在 B4 中發現因光線變化而產生的判斷誤差已減少許多,此因光線的變化已經慢慢的被 update 成背景的一部份。

● Tracking and Event Detection

以戶外取景為代表(含有一個 model,該 model 手中提著一個物體,在場景中移動,該 model 走到場景中央時把該物體放下,繼續前進,我們希望能夠 detect 出移動的 object,而該物體在被放下後,一開始時候該物體被偵測為前景,後來經過一

段時間後,由於該物體待的時間夠久了,足以影響 Background Model 所以後來就融入 Background 中,而消失不見)。(圖 7)所展示的便是 Tracking 所得的移動軌跡。

(圖 7-a)中,欲偵測的物體剛進入場景,此時人提著袋子走進場景,人和袋子是相連的,所以在影像偵測後只有一個區塊,因此重心也只有一個。

(圖 7-b)中,此人提著袋子走了一段距離,此時人仍然提著袋子直線走了一段距離,但因每次影像偵測所測出的前景部份並不完全一樣,所以重心的位置會略有化。

(圖 7-c)中,人將袋子放下後繼續往下走,此時人與袋子已分開,所以在影像偵測中會得到二個區塊,所以重心的位置會有二個,加上區塊面積變小,由此可知有事件發生,此二區塊即為人和袋子。

(圖 7-d)中,人走出了場景,而袋子繼續留在場景中,因人已出了場景,且袋子在經過一段時間後即變為背景,所以偵測出來沒有前景,因此沒有東西可以追蹤,Tracking 的軌跡亦不會有變動。

另外,當偵測到有事件發生時,改用特寫的攝影機,將之對準特殊事件發生的地方拍攝一特寫畫面,(圖 8)即為人將袋子放下,系統判斷有特殊事件發生後所拍攝的特寫畫面。



(圖 8) 事件發生時的特寫畫面

5. 結論

本計畫利用現有的數位攝影機、pan/tilt/zoom 裝置等硬體,架設一套可透過網路控制的智慧型多相機監控系統,並發展相關的演算法及軟體,包括:相機校正、場景變動偵測、移動物體追蹤、簡易事件的偵測,及拍攝特寫鏡頭等功能,且具有全自動化、容錯性高的特色。

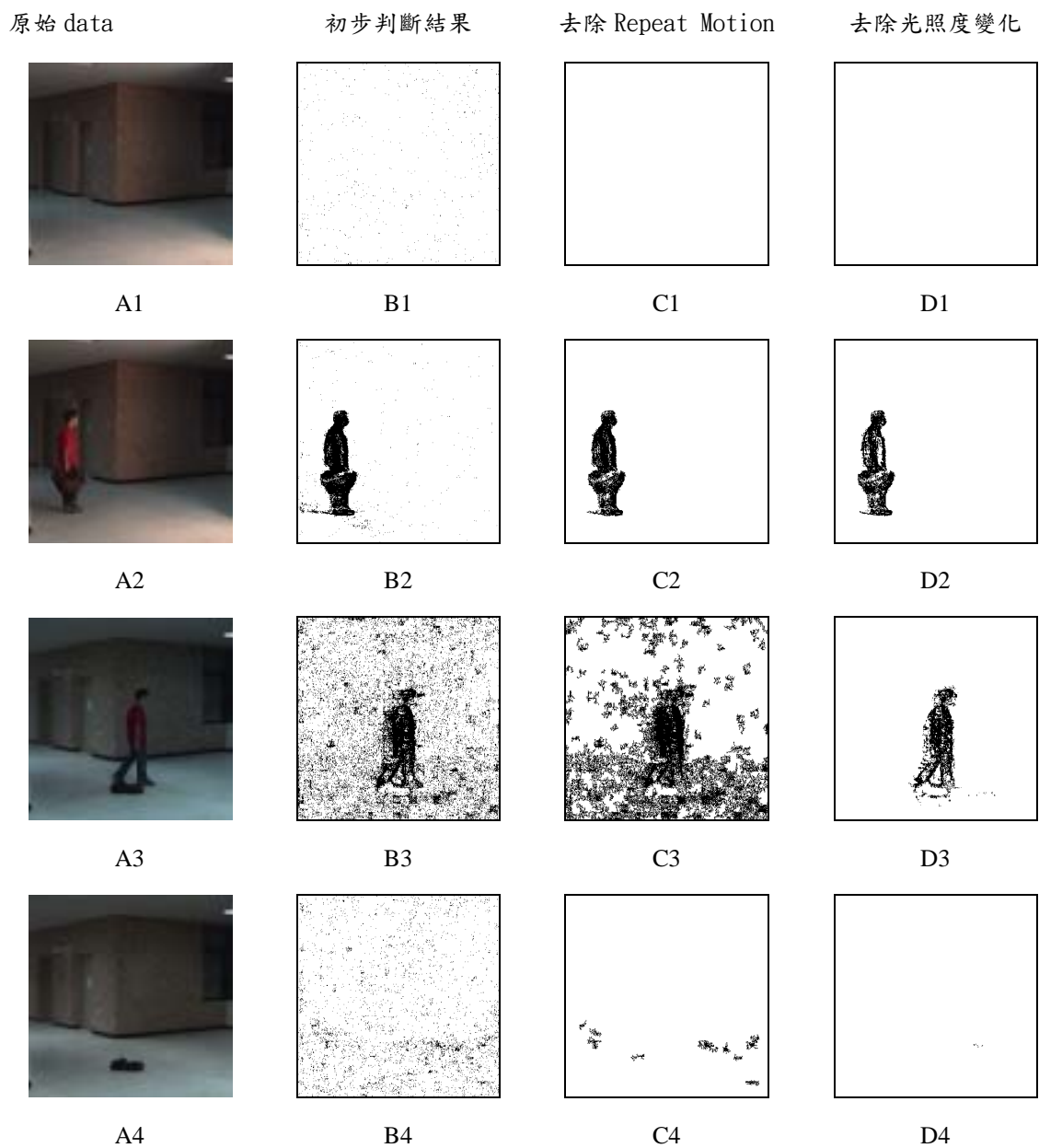
系統各部分模組的功能雖不盡完善,但已展現了智慧型視訊監控系統未來發展的可行性,若能將其擴充,將能大幅減低傳統監視設備所需的人力成本。

7. 参考文献

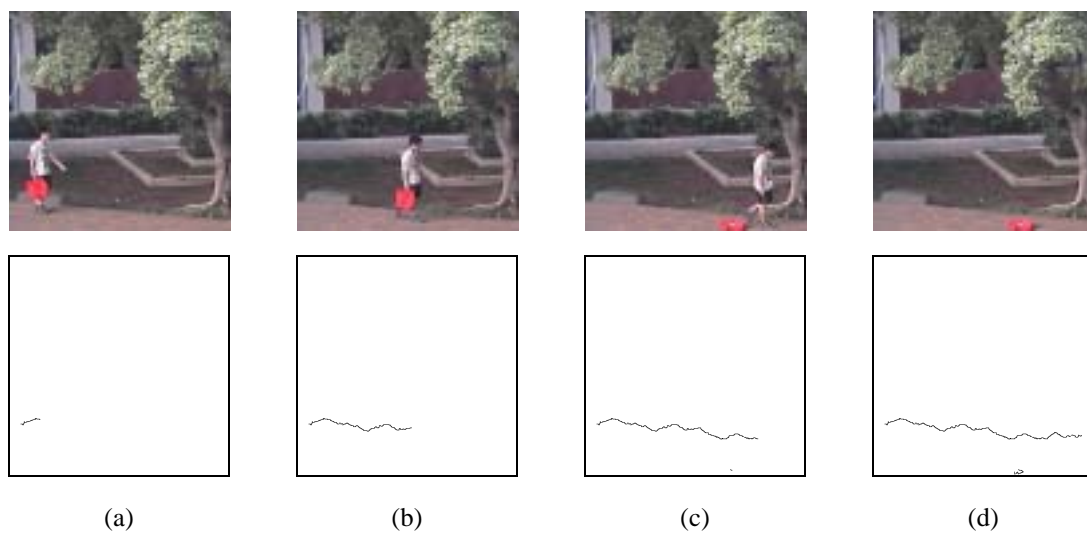
- [1] Special issue on video communications, processing, and understanding for third generation surveillance systems, Proc. IEEE, vol.89, No.10, October 2001.
- [2] R.T. Collins, A.J. Lipton, and T. Kanade, "A system for video surveillance and monitoring", 1999.
- [3] I. Haritaoglu, D. Harwood, and L.S. Davis, "W⁴: Real-time surveillance of people and their activities", IEEE TPAMI, Vol.22, No.8, 2000, pp.809-830.
- [4] C.R. Wren, A. Azarba yejani, T. Darrell, and A.P. Pentland, "Pfinder: real-time tracking of the human body", IEEE TPAMI, Vol.19, No.7, 1997, pp.780-785.
- [5] T. Boult, "Frame-Rate Multibody Tracking for Surveillance," Proc.DARPA Image Understanding Workshop, 1998.
- [6] T. Olson and F. Brill, "Moving Object Detection and Event Recognition Algorithms for Smart Cameras," Proc.DARPA Image Understanding Workshop, pp. 159-175, 1997.
- [7] D. Beymer and K. Konolige, "Real-Time Tracking of Multiple People Using Stereo," Proc. IEEE Frame Rate Workshop, 1999.
- [8] R. Hartley, "Self-Calibration of Stationary Camera", Internation Journal of computer Vision 22(1), pp.5-23, 1997.
- [9] W.H.Press et al., "NUMERICAL RECIPES IN C: THE ART OF SCIENTIFIC COMPUTING", second edition, Cambridge Univ. Press, Cambridge, England, 1992.
- [10] T. Aach et al., "Statistical model-based change detection in moving video," Signal Process., vol.31, pp.165-180, 1993.
- [11] E. Durucan and T. Ebrahimi, "Change Detection and Background Extraction by Linear Algebra," Proceedings of the IEEE, vol.89, no.10, pp.1368-1381, Oct. 2001.
- [12] A. Elgammal, D. Harwood, L. Davis, "Non-parametric Model for Background Substraction", IEEE 7th International Conference on Computer Vision, Sep. 1999.
- [13] K. Skifstad and R. Jain, "Illumination independent change detection for real world image sequences," CVIP, vol.46, no.3, pp.387-399, 1989.



(圖 5) Change Detection 第一組實驗結果(戶外場景)



(圖 6) Change Detection 第二組實驗結果(室內場景、Event 發生)



(圖 7) Tracking 的實驗結果