

運用人體輪廓特徵與靜態動作分類為基礎的 行為分析系統

黃培忠

中原大學資訊工程學系

9477025@cycu.edu.tw

張元翔

中原大學資訊工程學系

changyh@ice.cycu.edu.tw

賴東賢

中原大學資訊工程學系

egria7314@hotmail.com

摘要

隨著科技進步，監控系統已經被廣泛的運用在許多環境。有別於過去單純錄影功能的監控系統，目前研究方向已漸漸邁向智慧型監控。智慧型監控主要是利用電腦視覺的技術，讓監控系統可以自動擷取到人的位置，進而分析其在視訊中的各種行為。本論文的研究目的，即在於利用監控攝影機所拍下的視訊中，偵測其中人類物件的位置，進行追蹤與分析，並將人類物件所表現出的動作分割成多個靜態動作存入資料庫後進行比對，以辨識該種行為屬於何種行為。本系統設計包含擷取視訊物件的前置處理、物件的正規化、人類特徵擷取、建立行為資料庫與人類行為分析等步驟。本系統以在 20 個標準行為所建立出來的行為資料庫為訓練樣本，再經由 32 個未知行為樣本測試，可以成功的分辨出行走和跑步這兩種常見的人類行為，其中行走與跑步所包含 2322 及 610 之靜態動作其判斷率均達到 90% 以上。總而言之，人類的行為是由多個動作所組成的，利用這種方法未來將可以經由各種靜態動作的組合分析出多種的人類行為，將此方法應用在監控系統上將可以應付各種不同行為分析的需求。

關鍵詞：行為分析，監控系統，視訊處理。

Abstract

With the advance of technology, the surveillance system has been widely used in many environments. Computer vision and multimedia

techniques make feasible the development of new and “smart” surveillance systems that are different from traditional systems with only the recording function. In recent research, the surveillance system has been designed to automatically detect the location of human in video and to analyze the human behavior. In this paper, the objective is to identify and track the location of human, to match various static posture states in human behavior with respect to an established database, and to recognize the actual human behavior. The system design includes pre-processing of objects, object normalization, feature extraction of human, and collection of a database of human behavior for further analysis. The system was developed using 20 standard behavior as training samples, and was then evaluated with 32 unknown behavior samples. In addition, 2322 and 610 static poses were also evaluated. Our preliminary results demonstrated a 90% of classification in recognizing the two human behavior, namely the “walking” or “running”. In conclusion, human behavior is composed of many static posture states. Our methods could be used to analyze the combination of these states, and ultimately applied in surveillance systems with the need to recognize various human behavior.

Keywords: behavior analysis, surveillance system, video processing.

1. 前言

隨著科技進步，攝影設備的先進，使得監控系統在日常生活中已被廣泛的應用。只是到目前為止大部分的監控系統都還限於單純的記錄所拍攝的場景。因此要做到可以自動化偵測的智慧型系統變成了目前影像及視訊處理中的一大課題。過去文獻對於不正常的行為辨識，如跌倒、徘徊[1][2]，與靜態動作辨識，如站立、蹲下[3]的相關研究相當多，但對人體一般行為，如行走、跑步、跳躍等反而較有限。然而行走、跑步、跳躍等為人類最平凡的運動行為，如果可以針對人類較為平凡的運動行為加以分析，補足過去只分析不正常人類行為的分析研究，將會使自動化偵測的智慧型監控系統更加完善，進而達到分析所有的人類行為。所以本研究提出一個辨識方法，針對這些平凡的人類行為提供一個有效的解決方案，希望將來能夠使得這些行為分析在監控系統中提供更完整的功能。

人類行為分析是近幾年來相當熱門的研究，近年的研究中，大致上可以分成幾種不同的辨識方法。例如，Hsu等人的研究[4]，是利用三角形分割法將人體分割成多個三角形區塊，然後在連結每個三角形區塊的中心點行成人體骨幹，再與資料庫作尤拉距離的比對來判斷此行為是否符合資料庫中的行為。Haritaoglu等人的研究[5]則是先將人體的中心點找出，接著計算出像素在水平方向與垂直方向的投影量，再依照不同的行為動作(如行走、蹲下、趴下等)分別判斷各種行為投影量與資料庫中的差異，來判斷是否符合該種行為。Cucchiara等人的研究[6]也是先將每個人類物件的垂直與水平方向的投影量算出，然後利用一連串的機率運算計算出每行和列投影量所產生的機率，再與欲偵測視訊物件做比對來達到辨識的效果。

由於人類行為具有高度的重複性，所以本研究會將人類行為先分類成多個簡單的靜態行為，再判斷此一連串的靜態行為是否符合行走、跑步、跳躍等動作的重複性以達到研究目的。本論文先將視訊中物件的行為加以抽取、量化及評估，根據視訊中時間與量化資訊來建立行為資料庫，最後依照欲偵測的行為與資料庫作比對判斷視訊中的物件行為。

2. 研究方法

圖 1 為系統流程圖。在作行為分析之前，系統會將輸入的原始影像經過物件偵測、物件追蹤、正規化、特徵擷取等前置處理，接著將所偵測出來的特徵經過分類所得到的靜態動作狀態的結果儲存於行為資料庫中。而待偵測的未知行為也會經過相同的前置處理，最後與行為資料庫中所儲存的資料加以比對，以判斷該未知行為是屬於何種人類行為。以下就本系統詳述之。

2-1 前景物件擷取與背景更新

在本篇論文中，我們假設所有的視訊(或影像序列)都是由一台固定位置的攝影機所拍攝。當攝影機的位置固定時，擷取前景物件的時候，就可以利用前景減背景的方式來偵測物件位置。假使背景固定不變，該像素位置上並無前景物件出現，在前景減背景的方法中，影像的像素值並不會有太大的變化。反之，若該像素位置上有前景物件出現，則該點的像素值會突然的升高，如圖 2。因此，要擷取前景物件影像，只要針對每個像素隨著時間在前景減背景之後的變化即可判斷該像素上是否有物件出現。

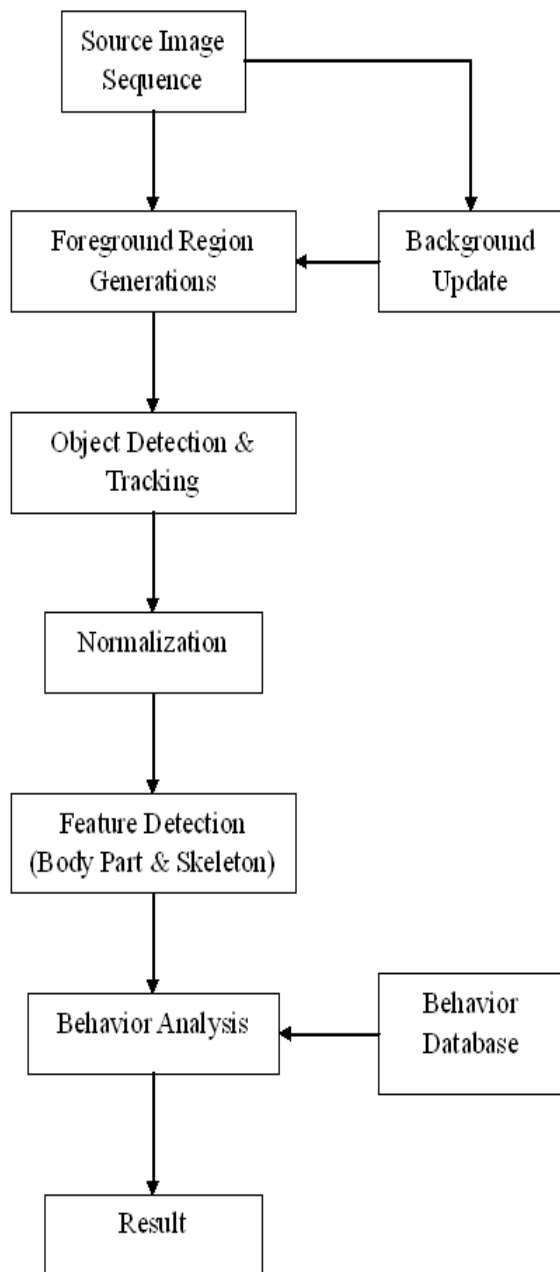


圖 1 系統流程圖。原始影像序列經由物件偵測、物件追蹤、正規化、特徵擷取和行為分析，將分析出來的行為與行為資料庫中之行為做比對。

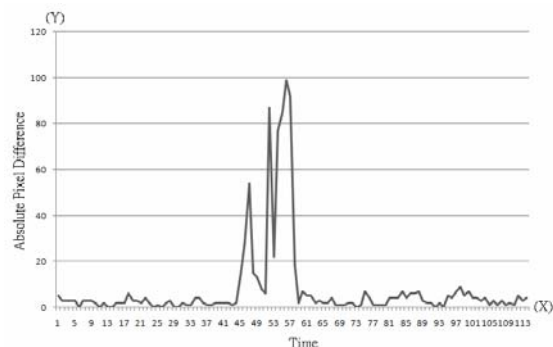


圖 2 影像中像素與時間的關係圖。圖中 X 軸代表時間，Y 軸代表前景減背景的絕對值。當前景與背景的像素差突然升高時代表該像素為移動物件。

其判斷法則為：

$$O_n(x, y) = \begin{cases} 0 & \text{if } |I_n(x, y) - B_n(x, y)| \leq T_1 \\ 1 & \text{if } |I_n(x, y) - B_n(x, y)| > T_1 \end{cases} \quad \dots\dots\dots(1)$$

O_n 為第 n 張影像中的前景物件，若 $O_n(x, y) = 1$ 代表第 n 張影像中被偵測為移動物件的像素，反之 $O_n(x, y) = 0$ 代表所偵測的像素為背景。 I_n 為第 n 張原始影像， B_n 為第 n 張背景影像。 T_1 為一臨界值(Threshold)。

由於在影像中，背景像素會因光線或風等因素而有所改變，所以必須要定時的更新背景影像才可以確保系統的正確性。在背景更新的研究中，若是突然的強光，或是突然有雲層遮住太陽時所產生的大量陰影，將會使背景更新系統無法在短時間內適應如此大的變化，造成更新失敗。此外，同一地點長時間滯留的物件也會使更新系統誤以為該物件為背景而造成失敗的結果。為了解決此問題，我們只針對沒有被視為前景的像素位置作背景的更新。在此我們定義 W_n 為背景影像中沒有被視為前景物件的像素值，其計算公式為：

$$W_n(x, y) = \begin{cases} B_n(x, y) & \text{if } O_n(x, y) = 0 \quad \dots\dots\dots(2) \\ 0 & \text{if } O_n(x, y) = 1 \end{cases}$$

假使 (x, y) 為背景則 $W_n(x, y)$ 將保留原始背景的

像素值， (x, y) 為前景物件時 $W_n(x, y)$ 將設為 0。由於在作背景更新時，前景物件的像素值將會嚴重影響背景更新的正確性，所以我們只利用不包含前景物件的像素作計算。若目前的背景影像與上一張背景影像的像素相差很大，如下公式：

$$|B_n(x, y) - B_{n-1}(x, y)| > T_2 \quad \dots\dots\dots (3)$$

則只針對目前影像與前張背景影像做更新如下：

$$B_n(x, y) = \alpha W_{n-1}(x, y) + (1 - \alpha) W_n(x, y) \quad \dots\dots\dots (4)$$

$\alpha \in (0, 1)$ 代表一個背景的更新參數，此參數會隨著 $B_n(x, y) - B_{n-1}(x, y)$ 的大小而做改變， $B_n(x, y) - B_{n-1}(x, y)$ 越大則 α 越小。假使目前背景影像與上一張背景影像的像素相差不大，如下：

$$|B_n(x, y) - B_{n-1}(x, y)| < T_2 \quad \dots\dots\dots (5)$$

就利用前 k 張影像的資訊來作背景更新，其公式如下：

$$B_n(x, y) = \frac{W_{n-k}(x, y) + \dots + W_{n-1}(x, y) + W_n(x, y)}{k + 1} \quad \dots\dots\dots (6)$$

T_2 為一個判斷是否要更新背景的臨界值且 $0 < T_2 < T_1$ 。

2-2 物件偵測及追蹤

在偵測到所有影像中的移動物件後，我們將經過一次的型態學閉合(Morphological Closing)[8]將鄰近分散的物件連結起來，然後再利用兩次的中值濾波(Median Filter)[8]來消除雜訊，並且利用連通元標記法(Connected Component Labeling)[8]標示所偵測到的物件。物件的重心位置(Centroid)，面積(Area)和像素強度(Intensity)為一個移動物件的最明顯的三個特徵。在做物件追蹤時，採用物件的面積，重心位置和像素強度這三個特徵來判定是否為同一物件。假設每個物件所包含的像素都為一單位

面積， A_n^i 代表第 n 張影像中第 i 個物件所佔的總面積，其計算公式為：

$$A_n^i = \sum_{(x, y) \in O_n^i} 1 \quad \dots\dots\dots (7)$$

第 n 張影像中的第 i 個物件的重心位置定義為 $C_n^i = (X_n^i, Y_n^i)$ ，則可以經由下列公式計算：

$$X_n^i = \frac{1}{A_n^i} \sum_{x \in O_n^i} x \quad \dots\dots\dots (8)$$

$$Y_n^i = \frac{1}{A_n^i} \sum_{y \in O_n^i} y \quad \dots\dots\dots (9)$$

在像素強度這方面，由於同一物件上的像素強度必定相同，所以採取該物件範圍內的平均像素強度來作為追蹤的依據。 AI_n^i 為第 n 張影像中第 i 個物件的平均像素強度， $I_n(x, y)$ 代表第 n 張影像在 (x, y) 位置上的像素值，則 AI_n^i 可以經由下列公式計算：

$$AI_n^i = \frac{1}{A_n^i} \sum_{(x, y) \in O_n^i} I_n(x, y) \quad \dots\dots\dots (10)$$

由於本系統是採取每秒鐘 50 張影像的頻率來作分割，因此每張影像中，物件的移動位置並不會太大，針對這一點，我們將採用重心位置的移動量來判斷該物件是否為同一物件，假設 TrackCent 為利用重心位置判斷是否為同一物件的結果，則其計算方法如下：

$$TrackCent = \begin{cases} 1 & \text{if } \left(\sqrt{(X_n^i - X_{n-1}^i)^2 + (Y_n^i - Y_{n-1}^i)^2} \leq T_3 \right) \\ 0 & \text{if } \left(\sqrt{(X_n^i - X_{n-1}^i)^2 + (Y_n^i - Y_{n-1}^i)^2} > T_3 \right) \end{cases} \quad \dots\dots\dots (11)$$

T_3 為利用兩個物件的重心位置來判斷是否為同一物件的臨界值。

因為所判斷的為同一物件，這兩個物件所包含的面積必定也會接近，因此利用面積大小可以作為物件追蹤的另一項依據。假設 TrackArea 為利用面積判斷是否為同一物件的結果，則其計算方式如下：

$$TrackArea = \begin{cases} 1 & \text{if } (A_n^i - A_{n-1}^i \leq T_4) \quad \dots\dots\dots (12) \\ 0 & \text{if } (A_n^i - A_{n-1}^i > T_4) \end{cases}$$

T_4 為利用兩個物件的面積來判斷是否為同一物件的臨界值。

同一物件上所含的像素顏色也必定相似，因此第三個判斷是否為同一物件的方法極為判斷物件上的平均像素。倘若前後兩個物件所包含的平均像素值接近的話，我們也將其視為同一物件。

$$TrackIntensity = \begin{cases} 1 & \text{if } (I_n^i - I_{n-1}^i \leq T_5) \\ 0 & \text{if } (I_n^i - I_{n-1}^i > T_5) \end{cases} \dots\dots\dots (13)$$

T_5 為利用兩個物件的像素強度來判斷是否為同一物件的臨界值。

最後將同時整合以上三種因素來做最終視訊物件的追蹤結果。

$$(TrackCent \cap TrackArea \cap TrackIntensity) = True \dots\dots\dots (14)$$

倘若上述式子(14)成立則可以判定 O_n^i 與 O_{n-1}^i 為同一物件。

2-3 正規化

由於攝影機的位置是固定的，因此人類物件在視訊中的大小會隨著人與攝影機的相對位置的遠近而有所不同。為了使後續行為分析的處理可以更精確，在擷取所有特徵之前，先對所擷取到的物件作正規化，使每張連續影像在判別其行為時不會因為人物大小的差異而產生錯誤。

正規化分成兩個步驟，第一個步驟是先依照視訊中因為位移所造成的大小誤差做正規化。第二個步驟，將針對攝影機不同高度的俯角所產生的相對距離作正規化。在監控系統中，會因為人行進的方向以及攝影機的高度產生兩種不同的視角。在第一階段的正規化時，先假設攝影機的高度與我們眼睛的高度相同，那麼因人行進的方向所產生的不同視角就如圖 3 所示。

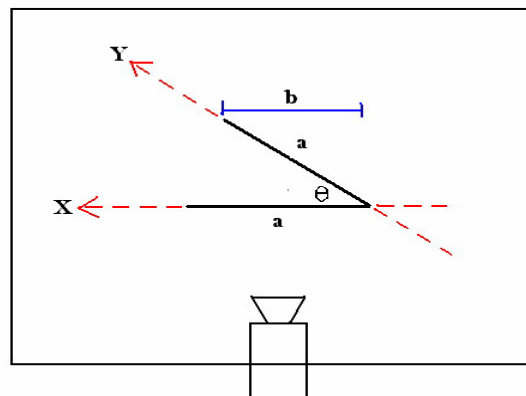


圖 3 攝影機與物件移動方向的俯視圖。X 為攝影機視角垂直線，Y 代表物件實際的行進方向，a 代表物件的原始寬度，b 代表物件再 Y 方向移動時投影在攝影機上的寬度。

X 代表攝影機視角垂直線，而 Y 代表物件實際的行進方向， a_n 為第 n 張影像中物件實際的寬， b_n 代表在第 n 張影像中所偵測到的物件寬。由於攝影機所拍攝到的影像寬度為該物件垂直投影的長度。假設 θ_n 為第 n 張影像中 X 與 Y 路徑的夾角，則物件的實際寬度 a_n 可表示為：

$$a_n = \frac{b_n}{\cos \theta_n} \dots\dots\dots (15)$$

其中 θ 為物件在影像中上下偏移的角度。假設每張影像物件往上(下)偏移了 d 個單位。則 θ 可表示為：

$$\theta_n = \tan^{-1} d_n \dots\dots\dots (16)$$

d_n 代表第 n 張影像中物件在垂直方向所移動的距離。若 f_n^i 代表第 n 張影像中的第 i 個物件在 Y 軸的座標值。因此可以得知：

$$d_n = f_n^i - f_{n-1}^i \dots\dots\dots (17)$$

由於人在行走時中心位置會隨著身體震盪而有些許的上下起伏，故若 d 小於一臨界值，則將視為 $d=0$ 。

因為攝影機是固定位置，所以若人與攝影機的相對位置越遠，所攝得的物件越小，反之越大。

因此在正規化的第二步驟中，把不同大小的物件，轉換成相同大小以避免後續處理時所產生太大的誤差。假設 P_n 為 B_n 正規化後的影像則，

$$P_n = \{p(x, y) \in \{0, 1\}, x \in [0, B_x * r - 1], y \in [0, B_y * r - 1]\} \quad \dots\dots\dots(18)$$

其中 r 為物件的放大率。假設我們欲將所有物件正規化為高度 H 大小的物件，且 B_y 代表放大前物件的原始高度，則

$$r = \frac{H}{B_y} \quad \dots\dots\dots(19)$$

2-4 特徵擷取

在這篇論文中所提出行為分析的方法是利用偵測人體上固定的特徵點在時間上所造成的重複性運動來計算。因此我們必須從人體影像中擷取相關的特徵點，如人體的四肢，軀幹，頭等位置，且所取得特徵點的正確性會直接影響到最後分析行為的結果。為了可以正確的擷取出正確的特徵點，採用兩個步驟分別對人體輪廓做分析：偵測人體骨幹和擷取特徵點。首先試圖找出人體的骨幹，如圖 4。接著利用所尋找出來的骨幹尋找出現在骨幹末端的點就是所需要的特徵點(如頭、手、腳的位置)，如圖 5。

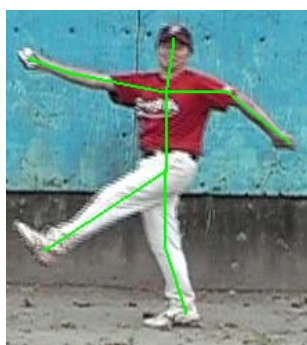


圖 4 人體的骨幹分佈圖。圖中綠色所表示的即為人體的骨幹分佈。

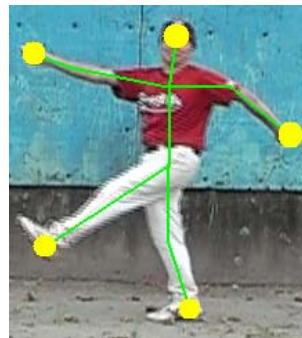


圖 5，人體的特徵點。圖中黃色的點為搜尋骨幹所有末端點後所偵測出頭、手、腳等特徵點。

在取得人體骨幹的方法中，採用 Hsu 等人所提出的方法[1]，先將人體的輪廓劃分成多個三角形，接著再連接所有三角形的中心位置以取得人體骨幹。所有取得的骨幹，其端點必定會落在人體的雙手、雙腳以及頭部。因此依據此骨幹做為路徑，連結骨幹的三角形切割的中心點進行深度優先搜尋(Depth-First Search, DFS)[9]的搜尋。每當搜尋到盡頭時即表示該位置為人體的特徵點位置。

由於在利用 DFS 的方法找到的五個特徵點(頭、左手手掌、右手手掌、左腳腳踝、右腳腳踝中)，系統利用這五個特徵點的相對位置來判定腳踝位置。腳踝的位置必定是在這五個點的最底部，除非倒立、蹲下和趴下等非一般性的人體行為。所以在取得特徵位置時，採用五個特徵點中最低的兩個特徵，該特徵點即為雙腳腳踝的位置。在取得雙腳腳踝後，再套用物件追蹤的方法分別對兩隻腳踝作追蹤以確保往後的特徵點中右腳腳踝(左腳腳踝)不會突然變成左腳腳踝(右腳腳踝)而造成系統分析上的錯誤。

2-5 行為分析

2-5-1 靜態動作分類

靜態動作的行为中，我們將分成兩部份來分

類。第一類為外型輪廓可以明顯分辨的動作，如彎腰、蹲下等。這一類的動作分類我們將人體物件的重心位置找出，利用經過重心位置的垂直軸與水平軸作量化，如圖 6。

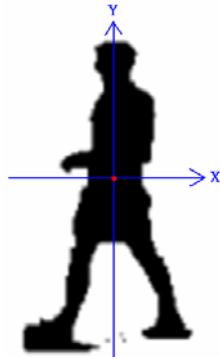


圖 6 利用物件重心位置找出垂直軸與水平軸。

量化時我們採用相加和相減兩種方式做比對。相加的量化即為將中心軸左右兩邊屬於該物件的像素相加，若 $P_1 = \{P_1H_n^{h+}, P_1H_n^{v+}, P_1H_n^{h-}, P_1H_n^{v-}\}$ 代表動作 P_1 在第 n 張影像的量化結果，其中：

$P_1H_n^{h+}$ 為第 P_1 種動作在第 n 張影像中的水平相加量化，

$P_1H_n^{v+}$ 為第 P_1 種動作在第 n 張影像中的垂直相加量化，

$P_1H_n^{h-}$ 為第 P_1 種動作在第 n 張影像中的水平相減量化，

$P_1H_n^{v-}$ 為第 P_1 種動作在第 n 張影像中的垂直相減量化，

若 $I_n^i(x, y)$ 為第 n 張影像中第 i 個物件在 (x, y) 的像素值，則 $P_1H_n^{h+}$ 、 $P_1H_n^{v+}$ 、 $P_1H_n^{h-}$ 、 $P_1H_n^{v-}$ 可分別由下列式子表示：

$$P_1H_n^{h+}(y) = \sum_{x=0}^{Width/2} I_n^i(x, y) + \sum_{x=Width/2}^{Width} I_n^i(x, y) \dots \dots \dots (20)$$

$$P_1H_n^{v+}(x) = \sum_{y=0}^{Height/2} I_n^i(x, y) + \sum_{y=Height/2}^{Height} I_n^i(x, y) \dots \dots \dots (21)$$

$$P_1H_n^{h-}(y) = \left| \sum_{x=0}^{Width/2} I_n^i(x, y) - \sum_{x=Width/2}^{Width} I_n^i(x, y) \right| \dots \dots \dots (22)$$

$$P_1H_n^{v-}(x) = \left| \sum_{y=0}^{Height/2} I_n^i(x, y) - \sum_{y=Height/2}^{Height} I_n^i(x, y) \right| \dots \dots \dots (23)$$

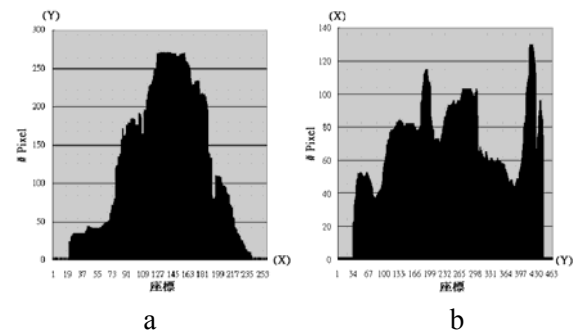


圖 7 經過相加計算的量化結果，其中 X 軸代表中心軸對應到物件的位置，Y 軸代表量化的值。a. 為水平量化 b. 為垂直量化。

以圖 6 的物件為例，所計算出來的結果如圖 7 和圖 8。相加的量化結果可以表現出該物件的分佈範圍，X 軸代表分佈的範圍，Y 軸代表分佈的量。相減量化代表的是物件的對稱性，如果 Y 軸所呈現的值越小，表示該物件在 X 軸位置上所呈現的對稱性越高。

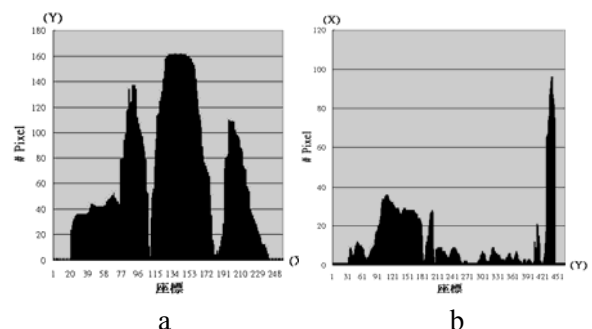


圖 8 經過相減計算的量化結果，其中 X 軸代表中心軸對應到物件的位置，Y 軸代表量化的值。a. 為水平量化 b. 為垂直量化。

在動作分類之前，我們針對每個不同的動作進行訓練，若欲分類的連續影像中有與訓練出來的特定動作相符，則判定它們為同一種動作狀態。

假設訓練的資料有 k 個不同的動作，每個動作

有 n 個影像，則其中每一個訓練後的動作 $TP_i = \{TP_i H^{h+}, TP_i H^{v+}, TP_i H^{h-}, TP_i H^{v-}\}$ 可以由下列式子所表示：

$$TP_i H^{h+} = \frac{1}{n} \sum_{i=0}^{n-1} P_i H_i^{h+} \dots\dots\dots(24)$$

$$TP_i H^{v+} = \frac{1}{n} \sum_{i=0}^{n-1} P_i H_i^{v+} \dots\dots\dots(25)$$

$$TP_i H^{h-} = \frac{1}{n} \sum_{i=0}^{n-1} P_i H_i^{h-} \dots\dots\dots(26)$$

$$TP_i H^{v-} = \frac{1}{n} \sum_{i=0}^{n-1} P_i H_i^{v-} \dots\dots\dots(27)$$

若 S_i 代表第 i 個動作的比對數據， PH_n^{h+} 、 PH_n^{v+} 、 PH_n^{h-} 、 PH_n^{v-} 代表未知動作的量化結果，為了避免數值過大，將比對數值取 $-\log$ 後如下式：

$$S_i = -\log \left[\frac{(PH_n^{h+} - TP_i H^{h+})^2 + (PH_n^{v+} - TP_i H^{v+})^2}{(PH_n^{h-} - TP_i H^{h-})^2 + (PH_n^{v-} - TP_i H^{v-})^2} \right] \dots\dots\dots(28)$$

假使所計算出來的 S_i 大於臨界值，則可以判定此動作 P 和動作 TP_i 可歸類為同一種靜態動作狀態。

第二類的方法利用人體四肢的相對位置來做靜態動作的分類，例如，人在行走的時候所會產生一連串的靜態動作就是跨步、雙腳張開、收腳和雙腳合併這四種狀態。判斷此類行為的靜態動作，我們是利用人雙腳所產生的夾角角度來判斷此時的靜態動作視為跨步還是兩腳合併。如圖 9，圖中的夾角 α 角所代表的是在某個行走的過程中雙腳張合時所產生的夾角，它會因為每個靜態動作的不同而產生不同的角度。

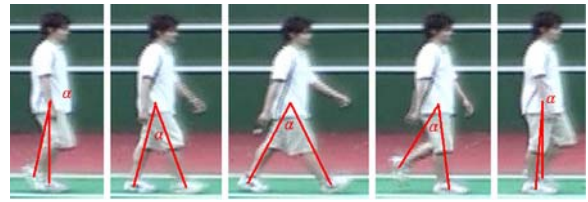


圖 9 連續影像中行走的行為，雙腳將會依照不同狀態所呈現出不同的夾角。

我們將利用兩個臨界角 $\delta_{\min}$ 和 $\delta_{\max}$ 將一連串的動態行為劃分為雙腳合併、雙腳張開、縮腳和跨步四個靜態動作，假設 $s = 1$ 代表雙腳合併， $s = 2$ 代表雙腳張開， $s = 3$ 代表縮腳， $s = 4$ 代表跨步則判斷式如下：

$$\begin{aligned} \text{if } \alpha_i &\geq \delta_{\max} && \text{than } s = 1 \\ \text{if } \alpha_i &\leq \delta_{\min} && \text{than } s = 2 \\ \text{if } \alpha_i &< \alpha_{i-1} && \text{than } s = 3 \\ \text{if } \alpha_i &> \alpha_{i-1} && \text{than } s = 4 \end{aligned} \dots\dots\dots(29)$$



圖 10 人類物件之雙腳與重心的夾角計算。

在計算 α 角的方法中，我們是採用物件的重心位置和兩腳這三個點所形成的向量夾角來計算。如圖 10，假設物件的重心位置為 G ，兩腳的座標點是 F_1 和 F_2 。則 α 可以利用下列式子計算所得：

$$\alpha = \cos^{-1} \left(\frac{(F_1 - G) \cdot (F_2 - G)}{\|F_1 - G\| \times \|F_2 - G\|} \right) \dots\dots\dots(30)$$

其中， G 、 F_1 、 F_2 都是在二維平面上的座標點：

$$G = \{(x, y) \in \mathcal{R}, x \in [0, Width - 1], y \in [0, Height - 1]\} \dots\dots\dots(31)$$

$$F_1 = \{(x, y) \in \mathcal{R}, x \in [0, Width - 1], y \in [0, Height - 1]\} \dots\dots\dots(32)$$

$$F_2 = \{(x, y) \in \mathcal{R}, x \in [0, Width - 1], y \in [0, Height - 1]\} \dots\dots\dots(33)$$

2-5-2 行為資料庫

在行為資料庫中，主要會將所要訓練的行為依照不同的靜態動作狀態做分割，設計成有限狀態機(Finite-State Machine)，作為後續行為分析的依據。

由於不同的行為會由各種靜態動作狀態的不同組合所構成，以行走的行為為例。假設 S_1 代表雙腳合併， S_2 代表雙腳張開， S_3 代表縮腳， S_4 代表跨步，若此人的動態行為中出現 $S_1 \rightarrow S_4 \rightarrow S_2 \rightarrow S_3$ 的靜態動作分割則可以判定此一連串靜態動作所組成的行為是行走行為，如圖 11。

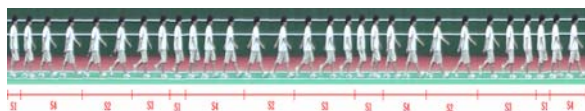


圖 11 此圖是行走的連續影像，以每秒 25 個 Frame 擷取下來。結果從左邊到右邊其中靜態動作會依序 $S_1 \rightarrow S_4 \rightarrow S_2 \rightarrow S_3$ 的順序重複。

由於行走行為是一個不斷重複的行為模式，因此，在設計行為資料庫時，可以將行走的行為設計為圖 12 的狀態圖。

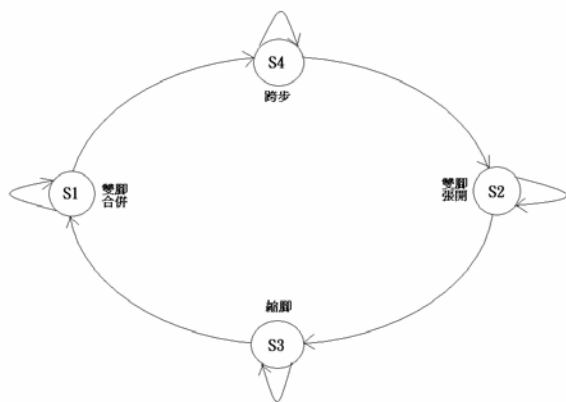


圖 12 行走行為的判別狀態，不論一開始所偵測到的狀態為何者，只要進入了該狀態機，若能符合所有條件產生一個循環，則可確定該

連續影像為行走行為。

然而人類的多種行為中，並不是只有行走的行為會依上述順序，例如跑步也會有雙腳合併、跨步、雙腳張開、縮腳等規律性的動作產生，圖 13 為一個跑步行為的連續影像，其靜態動作分類後的結果也是 $S_1 \rightarrow S_4 \rightarrow S_2 \rightarrow S_3$ 的順序。



圖 13 此圖是跑步的連續影像，以每秒 25 個 Frame 的速率擷取出來。

因此在建立行為資料庫時，還另外增加了人體物件重心座標(C_n^i)，移動速率(R_p^s)等特徵。人

體的重心座標可以依照第二章中的公式(8)(9)來計算。而移動速率則是與每個狀態所佔的影像數量有關。

假設此動作 p 在第 s 個靜態動作狀態的速率 R_p^s

可表示為：

$$R_p^s = \frac{1}{\sum_{p \in s} 1} \dots \dots \dots (34)$$

即使是相同的標準行為動作，在多個相同行為的連續影像中所分割出來的靜態動作狀態有些可能會因為行為進行的速度不同而產生有些靜態動作狀態無法被分割出來，為了提高分析的成功率，我們在訓練時分別依照每個靜態動作的分割所出現的機率來設定權重，在每次訓練的過程中都或出現的靜態動作狀態所得到的權重最高，若出現次數太少的狀態則可以忽略。假設 B_k^n 代表所要訓練的行為，其中 n 代表有 n 個相同的行為，每個相同行為各可以分類成 k 個狀態，則所訓練出的結果中第 j 個狀態權重

TB_j 可以由下式計算：

$$TB_j = \frac{\sum_{i=0}^{n-1} |B_j^i|}{n} \quad \dots\dots\dots(35)$$

其中 $i=0\dots n-1$, $j=1\dots k$, 再配合每個狀態中的重心位置和移動速率, 由於行為資料庫中是以每個靜態動作狀態為單位, 所以在取重心位置時, 將取屬於該狀態的數個連續影像的平均值, 而移動速率則是取一個狀態中所包含的連續影像數。假設 C_k^n 代表所要訓練的行為中, 每個狀態的平均重心; R_k^n 代表所要訓練的行為中每個狀態的移動速率。 n 代表有 n 個相同的行為, 每個相同行為各可以分類成 k 個狀態, 則訓練出來每個狀態的重心位置 TR_j 移動速率如式(36)(37)：

$$TC_j = \frac{\sum_{i=0}^{n-1} C_j^i}{n} \quad \dots\dots\dots(36)$$

$$TR_j = \frac{\sum_{i=0}^{n-1} R_j^i}{n} \quad \dots\dots\dots(37)$$

2-5-3 動態行為分析

在做動態行為分析時, 我們會將未知行為的連續影像經過物件擷取、特徵擷取和動態動作分類等前置處理後所得到的分類狀態順序與行為資料庫作比對。

在行為資料庫中所儲存的是標準行為的各種靜態動作狀態, 每個連續的狀態都將會形成一個循環, 例如走路, 則會形成跨步、雙腳張開、合腳、雙腳合併的循環; 如跳躍, 則會形成屈膝且重心下移、重心上移、重心回到原點等狀態。而在做行為分析時, 我們將分成狀態、重心和移動速率這三個部份來分析。我們將依序比對未知行為是否可以符合行為資料庫中的標準行為所產生的有限狀態、重心位置和移動速率, 並且依照每個狀態的權重來計算這個未知行為和標準行為的相似程度。若 S_i 、 T_i 分別代

表未知行為和行為資料庫中的標準行為中第 i 個狀態, 則其狀態相似值 $Sim_{State}(S, T)$ 、重心位置相似值 $Sim_{Cent}(S, T)$ 和移動速率相似值 $Sim_{Rate}(S, T)$ 可以由下式表示：

$$Sim_{State}(S, T) = \sum_{S_i=T_i} TB_i \quad \dots\dots\dots(38)$$

$$Sim_{Cent}(S, T) = \sum_{|SY_i-TY_i|<T_6} TC_i \quad \dots\dots\dots(39)$$

$$Sim_{Rate}(S, T) = \sum_{|RS_i-RT_i|<T_7} TR_i \quad \dots\dots\dots(40)$$

其中 SY_i 、 TY_i 代表未知行為和標準行為中重心位置的 Y 方向的值; RS_i 、 RT_i 為未知行為和標準行為的移動速率; T_6 、 T_7 分別為判斷重心位置和移動速率是否相同的臨界值。最後將依照這三樣相似值所佔的重要性決定此未知行為是否與標準行為相同。

$$Sim(S, T) = \alpha Sim_{State}(S, T) + \beta Sim_{Cent}(S, T) + \gamma Sim_{Rate}(S, T) \quad \dots\dots\dots(41)$$

$Sim(S, T)$ 代表最終的相似值, 而 α 、 β 、 γ 分別代表各相似值所佔的權重, 且 $\alpha + \beta + \gamma = 1$; $0 \leq \alpha, \beta, \gamma \leq 1$ 。

3. 研究結果

在本實驗中, 我們取跑步、行走兩種行為各 10 個標準行為的連續影像作訓練, 並且包含跑步和行走各 16 個待測的行為做測試。在 20 個的跑步和行走的行為所組成的有限狀態機中, 各有四個靜態動作狀態。各 16 個行走和跑步的待測行為, 經過每秒 50 個影像的分割一共可以得到 2322 與 610 個連續影像, 經過靜態動作分割後, 依照每個狀態所得到的正確率如表 1、表 2 所示：

表 1 行走的靜態動作分割結果

	實際結果數量	偵測成功數量	成功率(%)

S1(雙腳合併)	617	572	92.7
S2(雙腳張開)	609	597	98.0
S3(縮腳)	543	535	98.5
S4(跨步)	553	547	98.9

的分割出每一個靜態行為狀態，所以相對的行為分析的結果也相當高，而在跑步的行為方面，雖然靜態行為狀態所分析的結果也正確，但是物件的移動速率將會影響行為分析的結果。如慢跑，若跑的速度慢到與走路接近時，分析的結果將會造成錯誤。而走路的行為也亦然，若行走的速度過快，使其移動速率接近跑步時，也會造成分析上的錯誤。

表 2 跑步的靜態動作分割結果

	實際結果數量	偵測成功數量	成功率(%)
S1(雙腳合併)	184	163	88.6
S2(雙腳張開)	160	151	94.3
S3(縮腳)	131	129	98.5
S4(跨步)	135	134	99.2

由靜態動作狀態的測試結果得知，大部分的連續影像都可以精準的分類到正確的靜態動作狀態，在狀態交替時的連續影像才會產生少部分的分類錯誤。而將這些分類後的靜態狀態套入有限狀態機中，也都可以依照行為資料庫中所訓練得到的順序，成功的將未知的行為分析出來。表 3 為 32 個待測行為所分析出來的結果，欄代表實際的行為，列代表系統判斷的結果。

表 3 行為資料庫實驗結果。欄代表實際的行為，列代表系統判斷的結果。

判斷結果 \ 實際行為	行走	跑步
行走	15	2
跑步	1	13

結果顯示，在行走的行為中，能夠以非常精確

4. 結論與未來展望

本系統在固定攝影機的情況下，將監控系統所攝影的視訊影像以每秒 50 個連續視訊影做分割，可以成功的分析出視訊中人類的行為高達 90%。然而，人類的行為是由多個靜態動作所組成的，利用這個方法套用在其他人類行為模式上將可以組合出多種的行為資料庫，以便將來應用在各種監控系統所需要的各種需求。

未來將可針對靜態動作分割的部分增加手掌、手肘和膝蓋等特徵判斷，使判斷行為時可以同時利用手腳中多個特徵點加以分析。靜態動作主要是利用身體中各個特徵點的相對位置來判斷，因此若可以增加手腳等多個特徵判斷，將可以分割出更多細微的靜態動作。如在跑步和行走的差異上，行走時，右手和左腳會同時向前，同手同腳表示不正確的行走行為。跑步時一般的習慣會將雙手會彎曲置於胸旁，若可以增加這些靜態動作，可以有效的降低跑步和行走兩者因移動速度上的模糊所造成的錯誤分析。

致謝

本論文作者感謝中央研究院廖弘源博士在本研究過程中提供協助。

參考文獻

- [1] K. Hara, T. Omori and R. Ueno, "Detection of Unusual Human Behavior in Intelligent House," *IEEE Proceedings on Neural Networks for Signal Processing*, pp.697-706, 2002.
- [2] L. Panini, R. Cucchiara, "A Machine Learning Approach for Human Posture Detection in Domotics Applications," *IEEE Proceedings on Image Analysis and Processing*, pp.103-108, 2003.
- [3] T. Harada, T. Mori, T. Sato, "Human Posture Probability Density Estimation Based on Actual Motion Measurement and Eigenpostures," *IEEE International Conference on Systems, Man and Cybernetics*, vol. 2, no.10-13, pp.1595-1600, 2004.
- [4] Y. T. Hsu, J. W. Hsieh, H. Y. and, Liao, "Unusual Behavior Analysis and Its Application to Surveillance System," *IPPR Trans. on Computer Vision, Graphics and Image Processing*, pp. 988-995, 2006.
- [5] I. Haritaoglu, D. Harwood, and L. S. Davis, "W⁴ : Real-Time Surveillance of People and Their Activities," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, No. 8, pp. 809-830, August 2000.
- [6] R. Cucchiara, C. Grana, A. Prati, and R. Vezzani, "Probabilistic Posture Classification for Human-Behavior Analysis", *IEEE Trans. on Systems, Man and Cybernetics, Part A*, vol. 35, no. 1, pp. 42-54, January 2005.
- [7] G. Li, J. Zhang, H. Lin, D. Tu and M. Zhang, "A Moving Object Detection Approach Using Integrated Background Template for Smart Video Sensor," *IEEE Trans. on Instrumentation and Measurement*, vol. 1, no.18-20, pp. 462-466, May 2004.
- [8] R. C. Gonzalez, R. E. Woods, *Digital Image Processing*, 2nd Edition, Prentice Hall, 2002.
- [9] C. E. Leiserson, T. H. Cormen, R. L. Rivest and C. Stein, *Introduction to Algorithms*, 2nd Edition, McGraw-Hill Book Company, 2001.