國立交通大學

電控工程研究所

碩士論文

應用於機器人之基於影像人員活動偵測

Image-Based Human Activity Detection for Robotic Applications

研 究 生：陳維峻

指導教授：宋開泰 博士
羅佩禎 博士

中 華 民 國 九十九 年 十一 月

應用於機器人之基於影像人員活動偵測

Image-Based Human Activity Detection for Robotic Applications

研 究 生：陳維峻　　Student: Wei-Jyun Chen

指導教授：宋開泰 博士　　Advisor: Dr. Kai-Tai Song

羅佩禎 博士　　Dr. Pei-Chen Lo

國 立 交 通 大 學

電控工程研究所

碩 士 論 文

A Thesis

Submitted to Institute of Electrical Control Engineering

College of Electrical and Computer Engineering

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of Master

in

Electrical Control Engineering

November 2010

Hsinchu, Taiwan, Republic of China

中 華 民 國 九 十 九 年 十 一 月

應用於機器人之基於影像人員活動偵測

學生：陳維峻　　指導教授：宋開泰 博士
羅佩禎 博士

國立交通大學電控工程研究所碩士班

摘要

本論文之主要目的在研究以影像處理偵測人員之活動狀況，藉由單眼視覺攝影機擷取到的影像資訊，得以在環境中尋找人員的存在，並且完成五種人體姿態的辨識；在完成五種人體姿態後，藉由環境位置與姿態和停留時間的組合分析人員在家中常見的行為。本系統分為三個部份，分別為人員偵測、姿態辨識及活動偵測。人員偵測利用方向梯度直方圖(Histogram of Oriented Gradient, HOG)做為特徵，搭配支持向量機(Support Vector Machine, SVM)分類器來完成人員偵測。姿態辨識使用星狀骨架(Star Skeleton)做為特徵，搭配隱藏式馬可夫模型(Hidden Markov Model)訓練及辨識出站立、行走、蹲、坐及躺五種姿態。而在活動偵測設計中，本論文發展一套方法，利用環境物體特徵自動調整機器人移動後之環境邊界，得以偵測出人員目前所在的環境區域，再由人員在畫面中的位置與停留時間、姿態及偵測出人員所在的環境區域的組合，使用有限狀態機(Finite State Machine)辨識出在不同環境中可能的人員活動。經由實驗驗證人員偵測辨識率達95.33%，五種姿態之平均辨識率可達 94.8%，另外，在不同的機器人視角下可以準確的辨識人在環境中的位置及其所對應的行為。

Image-Based Human Activity Detection for Robotic Applications

Student: Wei-Jyun Chen Advisor: Dr. Kai-Tai Song

Dr. Pei-Chen Lo

Institute of Electrical Control Engineering
National Chiao Tung University

ABSTRACT

The main purpose of this thesis is to develop a vision-based human activity detection system employing a monocular camera. This system can be used for human-robot interaction in a home setting to provide service to people. A method is proposed to detect a human in acquired image frames. Five human poses are then recognized. The human activity detection system was designed by combining information from human location, human pose and the stay time. An environmental boundary detection method is proposed to determine the location of a human in the environment. This method uses features in the environment to automatically set environmental boundary, such that human location in the environment can be obtained. Satisfactory experimental results have been obtained with human detection rate of 95.33%. The pose recognition rate of five poses (standing, walking, sitting, squatting and lying) is 94.8%. Experimental study validates the performance of the developed method for human activity detection from different view angles.

致謝

首先感謝我的指導教授宋開泰博士，感謝他在專業上的指導，以他豐富的學識與經驗，配合實務的應用，使得此論文得以順利完成。

感謝學長 孟儒、嘉豪、格豪及學姊巧敏在理論及實作上的指導，並感謝與我共同努力的同學 信毅、傑巽、宗暘、哲豪及奕彣彼此之間相互的鼓勵與提攜，以及學弟 建宏、仕晟、上畯及、碩成、家昌和章宏在平日事務上的幫忙及繁忙生活中帶來的樂趣。感謝我的朋友 松明、施暹、東笠、慧雯、姿亘、欣榮等…，讓我的研究生涯中還不忘對人生探討許多話題，也讓我的生活更豐富且多采多姿。

感謝我的女友 紜鈴，總是給我源源不斷的動力及生活上的加油打氣讓我在低潮難過時可以傾訴；在開心時可以分享，我很感謝她的陪伴，使我在研究路途上不孤單且動力滿滿。

最後，特別感謝我最愛的父母，由於他們辛苦的栽培及生活上對我的關懷與照料，有他們全力的支持我就讀碩士學位，使我得以順利的完成此碩士論文的研究。

目錄

圖目錄

表目錄

第一章 緒論

1.1 研究動機

隨著科技的迅速發展，如今電腦的運算能力大幅提升，使得近幾年來機器人產業在技術及應用方面進步許多，不論在工廠自動化、展場導覽、居家看護、娛樂及保全等方面都可以看到智慧型機器人的使用。相較早期偏向於工業、太空、軍事等，代替人去做一些反覆性或高危險的事情，現今國內外企業及實驗室不斷的開發研究，使機器人可以在生活上替我們帶來一些便利性及娛樂性，因此機器人也漸漸的走入我們的生活之中。

機器人的影像辨識系統是機器人最重要的感測來源之一，這會直接影響到機器人的智慧型功能，一個具有高度辨識能力的機器人可以偵測到更多更準確的環境資訊，進而可以從這些資訊中做出最正確的決策。機器人與人互動(Human Robot Interaction, HRI)在近幾年來吸引到很多的注意及研究[1]，像日本 HONDA 公司的 ASIMO[2]能夠執行例如攜帶托盤和推車等任務能力。未來在家庭環境中透過機器人的協助可以達到一些生活上的便利，因此在這樣的應用情境中機器人必須知道人員在環境中行為活動及所代表的意義，而可以進一步和人互動。

1.2 相關研究

很多文獻都致力於人員活動偵測(Human Activity Detection)，Fujiyoshi and Lipton[3]提出利用星狀骨架的方式去描述人體的動作，而陳宣勝[4]將星狀骨架搭配了隱藏式馬可夫模型(Hidden Markov model, HMM)的方式去訓練一系列的人員姿態。Li and Aloimonos[5]也提出結合許多不同的人體姿態的組合來了解人的行為活動。在[6,7]中提出使用特徵比對的方式達到人類活動的辨識。由於一般人的活動會由很多不同的姿態所組成，因此，學習機制就被應用到人員活動的辨識程序當中。像 Carter 等人[8]就結合貝式定理(Bayesian)和馬可夫鍊(Markov chain)來辨識人員活動。如圖 1.1 示，Kellokumpu 等人[9]提出了利用人體的姿勢透過

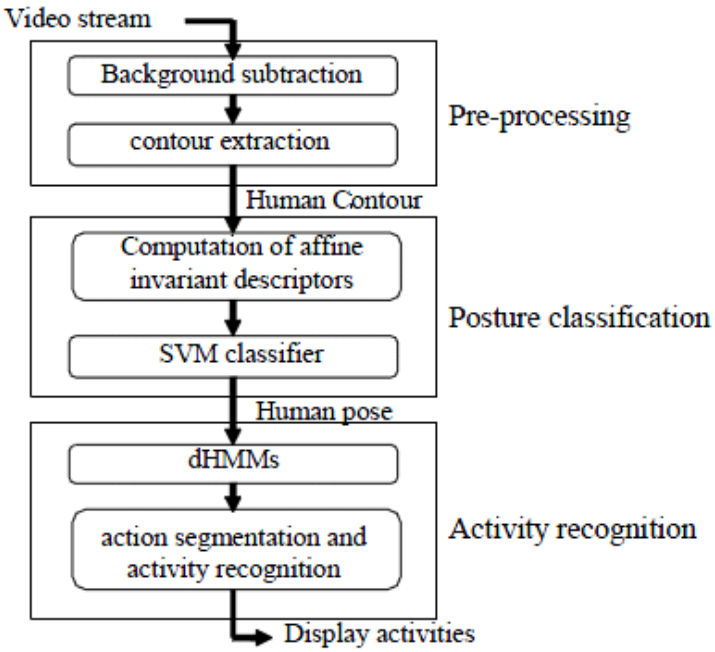


圖 1.1 人員活動辨識流程圖[9]

隱藏式馬可夫模型的方式來訓練達成活動辨識的目的。大多的結果都只有涉及到人員身體的姿態分析，例如，站，走，坐…等，然而姿態雖然是人員行為相當重要的依據之一，但並不足以描述人員活動。如圖 1.2 所示，Wang *et al.*[10]提出一個人的行為活動分析系統需要兩個低階的處理：人員偵測和追蹤，以及一個高階的處理過程：人員行為活動的了解。人員活動通常包含一系列的運動過程所組成，而每一個運動過程都有基本且完整的動作。由於人的身體是由關節的許多自由度去表示當前的動作，所以當只用 2-D 的影像是很難表達人的姿態是什麼，所以給一連串的動作會比較容易定義出人的姿態與活動。但是，要整合從影像中提取到的訊息，最重要的是要找到一種模式，可以有效的定義人員的活動。有限狀態機(Finite State Machine, FSM)[11]就是一個很好的方式，它可以整合在影像中所得到的資訊，將這些資訊透過自定的狀態組合找出合適的人員活動。人員的活動除了人體本身的特徵辨識之外，所處的環境位置不同，每個姿態所代表的意義也會不一樣。如圖 1.3 所示，Zhou 等人[12]提出利用單一固定式攝影機，定出環境每個區塊的位置，藉以監控人在環境中的活動，所以可以透過人所處的位置去推測人員目前的活動是什麼。

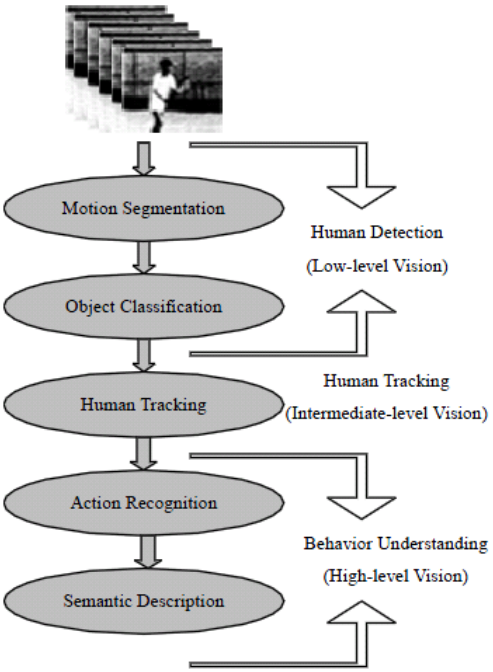


圖 1.2 Wang 等人所提出的人員行為活動辨識所需的過程[10]

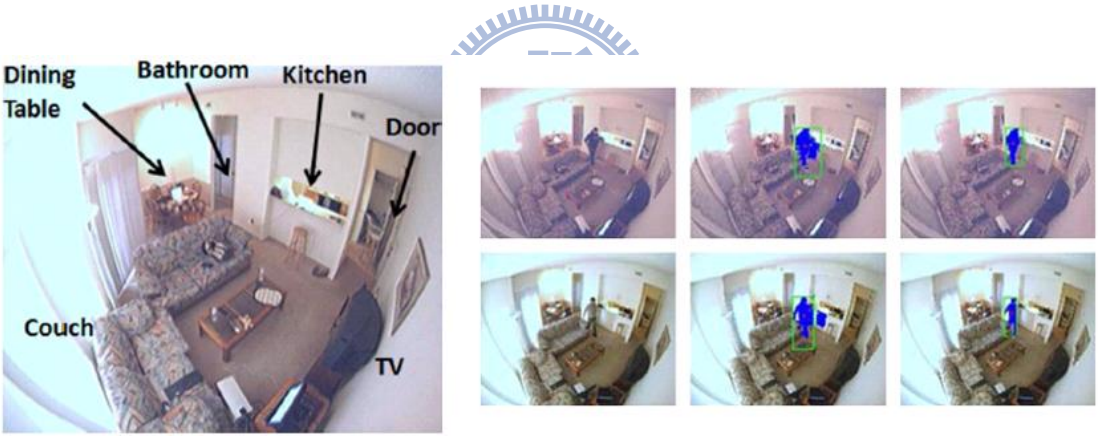


圖 1.3 Zhou 等人提出利用環境位置定位去監控人員活動[12]

如圖 1.4 和圖 1.5 所示，Shiomi 等人[13]設計機器人在購物商場之應用，他們利用雷射掃描儀去估測出人的位置及其所處的環境，此系統可以辨識出人員快步走、一般速度的行走、排徊以及停留，主要就是要讓機器人可以知道目前商場中的客人的狀態是什麼，讓機器人可以對人做出適當的回應和互動。若是快步行走，就可以知道客人可能是在趕時間，機器人不會去做打擾；若徘徊或停留，客人可能就是在等待或是不知道方向，此時機器人就可以去做慰問或引導的動作。

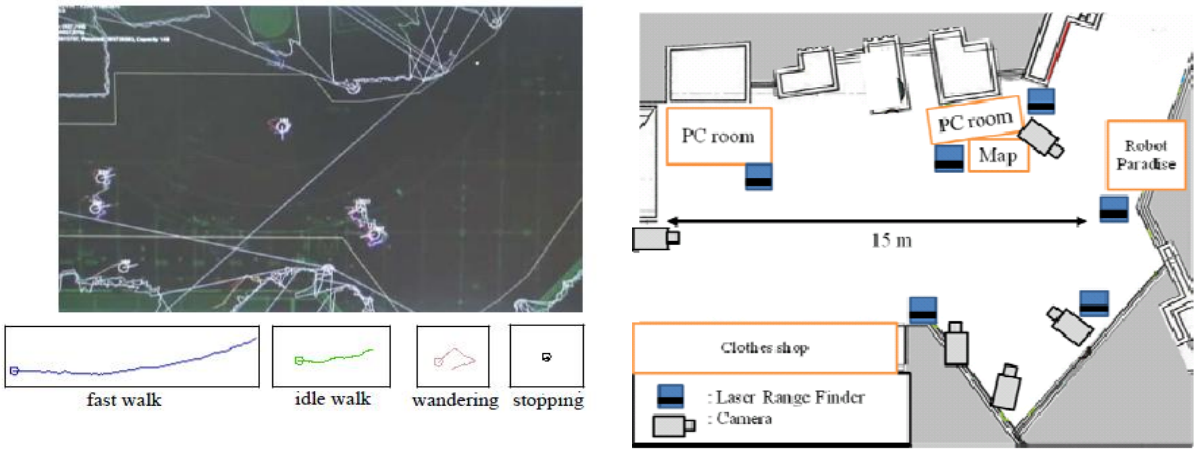


圖 1.4 賣場環境設定以及人員的活動偵測[13]

圖 1.5 賣場機器人與人互動[13]

由上面的相關研究可以知道，我們要辨識環境中人員的活動除了需要知道人員所在位置之外，必須先將人找出來，再對人做姿態的辨識。因此首先要有一套人員偵測和姿態辨識系統。在人員偵測的設計上，我們需要一個穩固的特徵萃取方法。在過去的研究中，特徵萃取主要可以分為邊緣方向直方圖(Edge oriented histogram) [14]、基於小波轉換法(Wavelet based detector)、尺度不變特徵轉換法(SIFT- descriptors) [15]和外形上下文(Shape contexts) [16]。

Viola 與 Jones 提出利用矩形特徵來進行人臉偵測，可得到高效率與高偵測率的結果 [17]，其也將該方法應用在行人偵測上，所使用的矩形特徵如圖 1.6 所示，然而由於行人的姿勢變異程度大，使得該方法偵測的正確率並無偵測人臉時來的高，針對此一問題，他們加上了行人運動(Motion)的資訊[18]，先計算前後

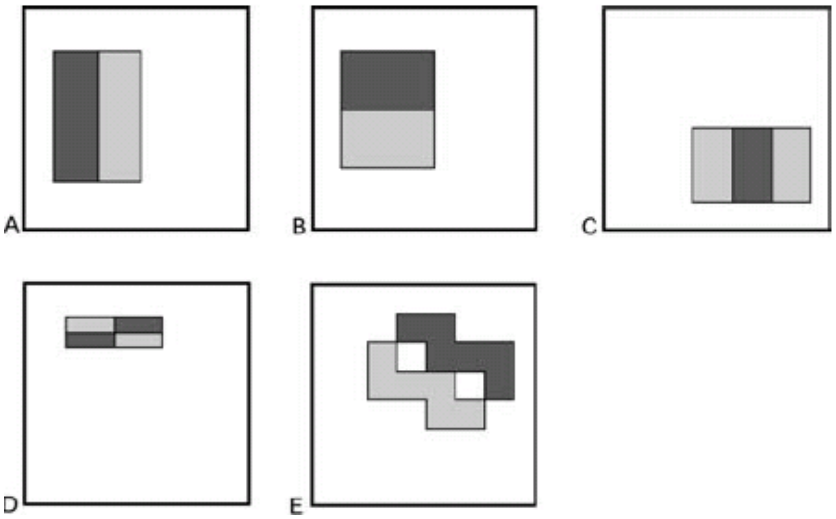


圖 1.6 在偵測行人時所使用的矩形特徵[18]

張影像在對不同方向做位移後的差值影像，再用距形特徵來抽取出行人運動的資訊，如此一來，大大地提昇了行人偵測的準確度。對行人而言，由於所穿著的衣服顏色變化大且環境光線不穩定，因此，顏色不是一個很穩定的特徵，相較於顏色，邊緣(Edge)資訊是一較為穩定的特徵。Gavrila [19] 提出利用邊緣資訊來做行人偵測。首先，如圖 1.7(c)所示，先找出輸入影像上的邊緣位置圖，再計算出該影像上所有像素與影像上邊緣的距離轉換(Distance Transform)影像，如圖 1.7(d)所示。其轉換方法為計算每一個像素與最近影像邊緣像素的距離，接下來藉由比對資料庫中行人的邊緣圖與經距離轉換 後的影像算出其斜面距離(Chamfer Distance)，最後以一臨界值來判斷輸入的影像何處含有行人。但是，由於 Gavrila 所提出的方法，也需要利用到環境的邊緣資訊，當環境光線不穩定或是有雜訊時，此方法就會降低準確度。Dalal 和 Triggs[20]提出利用方向梯度直方圖(Histogram of Oriented Gradient, HOG) 來萃取人體特徵，此特徵也是一種利用邊緣資訊做為特徵，但特別的是 HOG 是利用局部的特徵向量強度及方向去做統計，故有較高的準確率，此方法在第二章中會詳細說明。

在人員的姿態辨識上，大略分成三種方法：基於移動歷史影像法（Motion History Image, MHI）、基於輪廓（Silhouettes-based）與基於隱藏式馬可夫模型

法（Hidden Markov Model, HMM）。Davis[24]提出了移動歷史影像（Motion History Image, MHI）去識別人類的動作，它是基於移動歷史影像之辨識，藉由匹配基於瞬間的特徵來統計以達成辨識，把當時該點變化的持續時間給記錄下來，但 MHI 會容易受到雜訊物件動作的時間間隔影響，藉由圖 1.8 所示的揮手締造出歷史影像，再以歷史影像做移動梯度強度雷達圖來辨識人的姿態。

基於輪廓法之辨識，如圖 1.9 所示，Haritaoglu et al.等人[25]提出用黑色輪廓辨識方法，他們使用名為"Ghost"的單眼系統，即時的去找到從影像得到身體部位的標籤，Ghost 是一個在單色的影像裡去偵測人類身體部位的即時系統，它是基於輪廓的身體模型去決定身體部位的位置，當人類在做動作時就利用 Ghost 去預測六個身體部位的位置（頭部、雙手、雙腳和軀幹）。

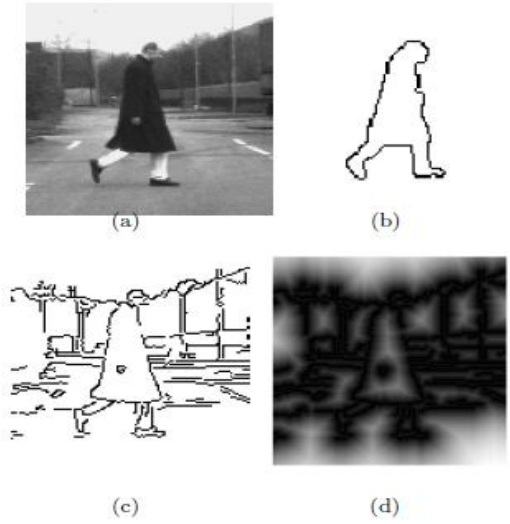


圖 1.7 (a)原始輸入影像；(b)資料庫中的行人邊緣影像；(c)輸入影像的邊緣影像；(d)影像(c)經過距離轉換後所得到的影像(DT Image)[19]

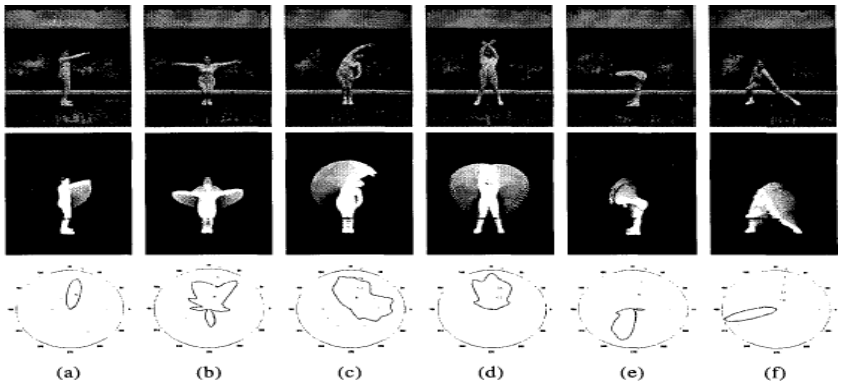


圖 1.8 不同姿態下之移動梯度強度展示圖[24]

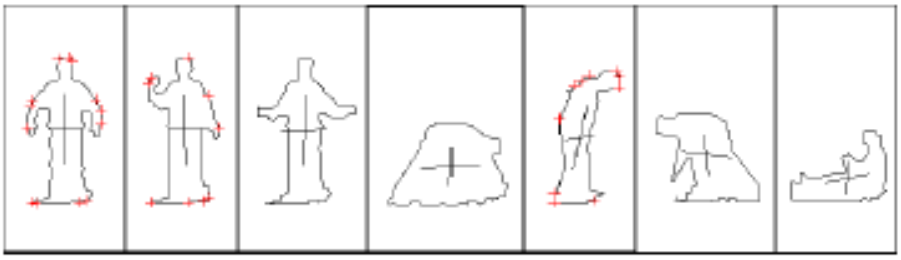


圖1.9 利用輪廓在不同姿態下之辨識結果[25]

Keiichi *et al.*等人[26]提出用短期動作(Short-term motion)和長期動作(Long-term motion)的隱藏式馬可夫模型來做辨識，所謂短期動作的的意思就是人類只做幾秒鐘的動作而已，例如揮手、舉手、點頭或搖頭等都是屬於短期動作，而長期動作就是人類做數十秒甚至是數分鐘的動作，而要分辨這些短期和長期動作就是使用HMM，但是在做辨識之前必須先對人類做部位的偵測，在人員的頭和雙手，使用特徵點偵測（人臉偵測、膚色偵測及點追蹤）來辨識人類的頭和左右手，且HMM有已經被定義並且訓練好的短期動作資料庫，根據人類的動作直接與資料庫進行比對，比對匹配的即是辨識出來的動作，同理，長期動作也是一樣，因為它是由一連串的短期動作所組成，也是有一組資料庫，看人類做什麼動作就直接跟資料庫比對，比對匹配的即是辨識出來的動作。

1.3 問題描述

由以上的文獻研究中得知人員活動會與人的姿態和所處的位置有關，所以在本論文將人員活動定義為，由人、人體姿態、和所在位置所組成，例如：“他走進房間”，“他”就代表人，“走”就代表人體姿態，“房間”就是所在位置。故本論文發展一套人員活動偵測系統，包含人員偵測、姿態辨識以及人員活動偵測設計三個部份。

人員偵測系統就是可以在環境中準確找到人，而一般人員偵測系統會使用膚色找出頭部及其四肢，但是在家庭環境或其他環境中，背景多變、光源較有不固定之變化且人員所穿著的衣物顏色皆會影響到經由顏色判別人形的準確率，故要找到一個能夠對抗這些問題的人員辨識方法。

姿態辨識部份，常見的方法有，在人身上佈置感測器、透過膚色找尋四肢及頭部來建立人體 3D 模型或多攝影機來抓取不同角度下之人體姿態[30,31]，但對一般使用者來說，在身上佈置感測器並不是一個較合適且自然的方式，人體 3D 模型要透過膚色或特定顏色來定出四肢故使用上較不具一般性，而多攝影並不適用於機器人與人互動上。因此，希望能透過單一視覺攝影機且不用依靠顏色及感測器的特徵萃取來達到人體姿態辨識部份。

本論文主要貢獻在活動偵測設計方面，因為要應用於一般家庭之中，所以需要人員的姿態和人員所在位置以便推測人員可能的活動。且應用於機器人上，而機器人是一個移動式的平台，所以希望能有一套自動環境邊界的產生，達到辨識人員位置的目的，並結合由人員偵測和姿態辨識得到目前人員所在位置、停留時間及目前人員姿態組合來達成人員的活動偵測。

1.4 系統架構

本論文系統架構如圖 1.10 所示，由攝影機取到影像後，透過環境邊界設定的運算，算出目前影像平面的環境邊界；接著影像會進入人員偵測系統，若有辨識到人員，則輸出環境中人員停留的時間並且利用環境邊界判斷出人員目前所處的環境位置，並會輸出人形的剪影供姿態辨識使用。姿態辨識系統利用人形剪影辨識出人員在影像中的動作，根據目前人員在影像平面中的動作組合，辨識出人員的姿態，結合目前人員所在位置、停留時間以及人員姿態，透過本論文設計的有限狀態機去判定出人員的活動。

1.5 章節說明

本論文一共分成六章，第一章介紹研究動機與目的，並且概略介紹所設計之系統架構。第二章說明利用HOG特徵萃取達成人員偵測的方法。第三章則是介紹利用星狀骨架找出人體的姿態。第四章介紹本論文所設計的人員活動偵測方法，包含如何自動產生環境邊界。第五章為實驗結果，驗證整體設計上的可行性。第六章為結論與未來展望。

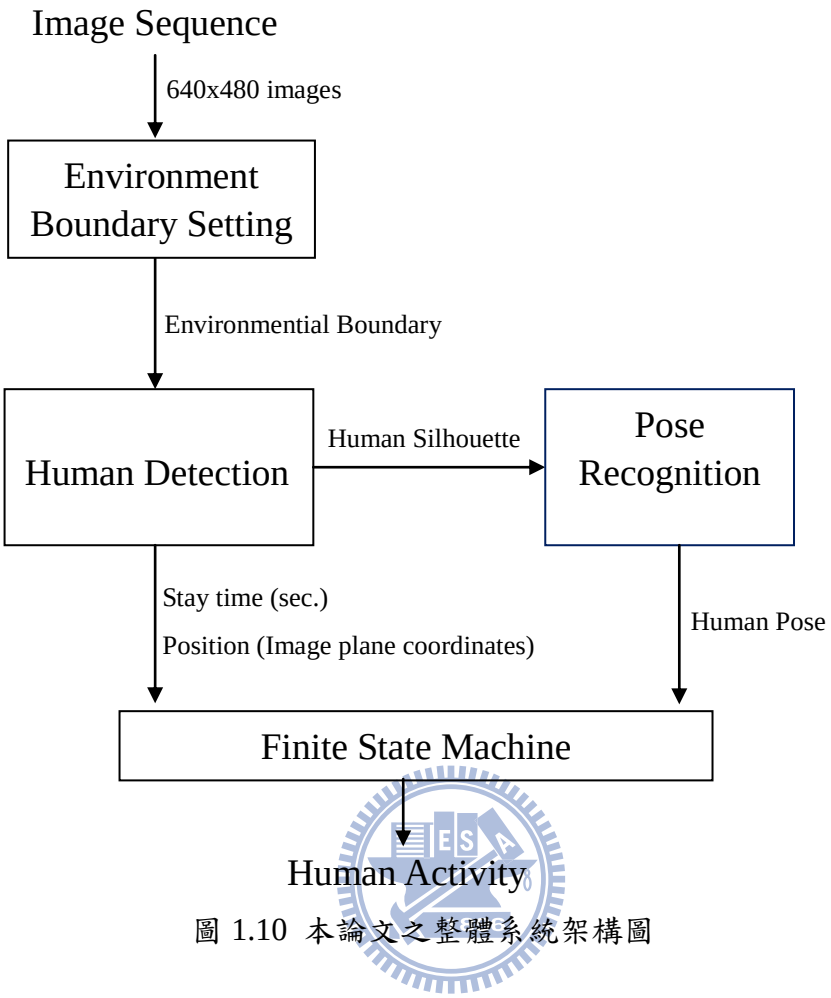


圖 1.10 本論文之整體系統架構圖

第二章 人員影像偵測與辨識

在本論文研究中，我們期望找出攝影機畫面中人體所在的位置及姿態來達成活動偵測的目的，因此首先便是要在畫面中判斷人是否存在。利用影像來偵測人是否存在畫面中是一項挑戰，因為人有很多不一樣的姿態且環境中存在著許多複雜背景和多變的光線。因此， Dalal 和 Triggs[20]提出利用方向梯度直方圖(Histogram of Oriented Gradient, HOG) 來萃取人體特徵，再透過線性支持向量機(Linear Support Vector Machine, LSVM)的方式去辨識是否為人，在他們的實驗中有比較方向梯度直方圖(HOG)、小波轉換(Wavelet)、尺度不變特徵轉換(Scale-invariant feature transform, SIFT)以及形狀上下文(Shape contexts)這幾種特徵描述方法的優劣，如圖 2.1 所示，也證明了 HOG 辨識的準確率優於其他特徵萃取的方法。

2.1 人員偵測架構圖

本論文之人員偵測系統架構，如圖 2.2 所示。以攝影機取像後，會先經過背景移除將所需要的前景取出來，接著會將前景利用像素投影的方式將感興趣的區域抓出來，再透過 HOG 將特徵點取出，最後利用 SVM 分類器去辨識是否為人形。

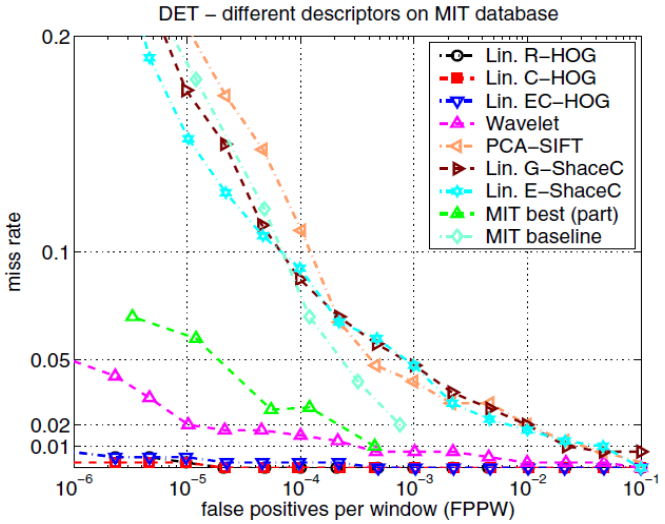


圖 2.1 各種人員偵測方法的效能比較圖[20]

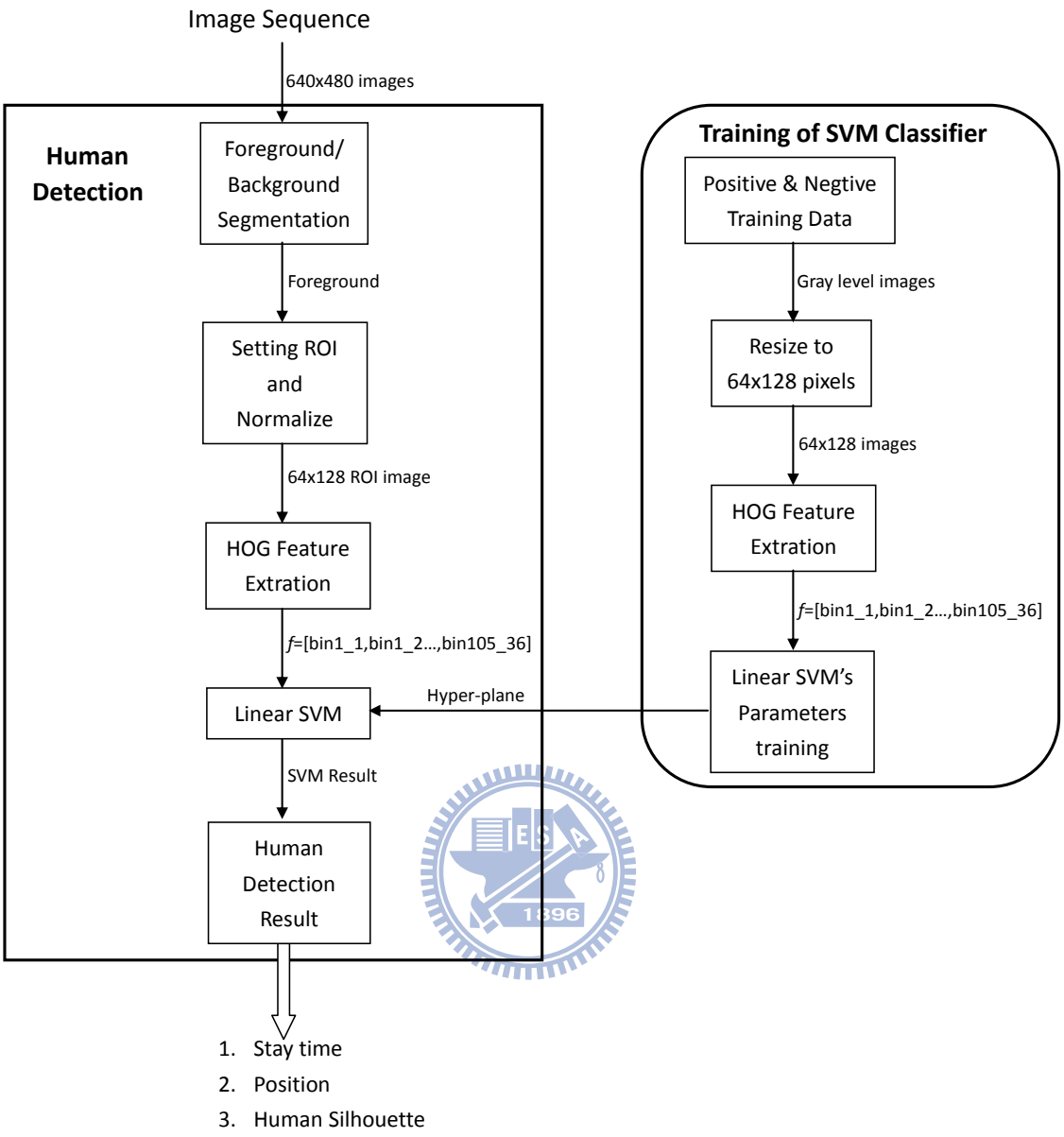


圖 2.2 人員偵測架構圖

2.2 辨識區域選取

首先，先在靜止畫面中存取一張背景圖片，再將之後得到的連續影像與此背景圖片做影像相減，可以得到二值化的移動物體前景，如圖 2.3(b)所示，接著對此張前景圖做水平方向及垂直方向的影像投影，可以得到初始的辨識區域 Xtart1、Xend1、Xstart2、Xend2、Ystart1 和 Yend1，如圖 2.3(c)所示，但由於在畫面中可能同時出現兩個以上的移動物體，若只做一次的投影所框選區域 Ystart1 及 Yend1 會選最高及最低的部份，就會造成個別的框選區域有留白的情況，如圖

2.3(d)，所以我們必須對各框選區域再做一次垂直方向的投影來分出 Ystart1、Yend1、Ystart2 及 Yend2，最後就可以找出目標物一的(Xstart1,Ystart1)、(Xend1,Yend1)及目標物二的(Xstart2,Ystart2) 、(Xend2,Yend2)，如圖 2.3(e)所示，就可以找出欲辨識的區域(ROI)，接著就可以對欲辨識區域萃取特徵。

2.3 Histogram of Oriented Gradient 特徵萃取

由 Navneet Dalal 的博士論文中[21]，可以得知 HOG 是一套俱有抗多姿態、複雜背景、不固定光源且穩定的特徵萃取演算法，圖 2.4 為 HOG 演算法的整體架構。當我們得到欲辨識的區域後，即為圖 2.4 中的辨識視窗(detection window)，就會對每一個欲辨識的區域中的像素(pixel)做梯度運算，梯度運算會包含梯度方向和梯度大小，對每一個細胞(cell)做梯度方向及強度的統計，再來對每一個區塊(block)裡面的梯度強度做正規化(normalization)，當整個辨識區域都做完時會得到一個向量式來表示它的特徵，下面會依據此演算法流程做詳細介紹。

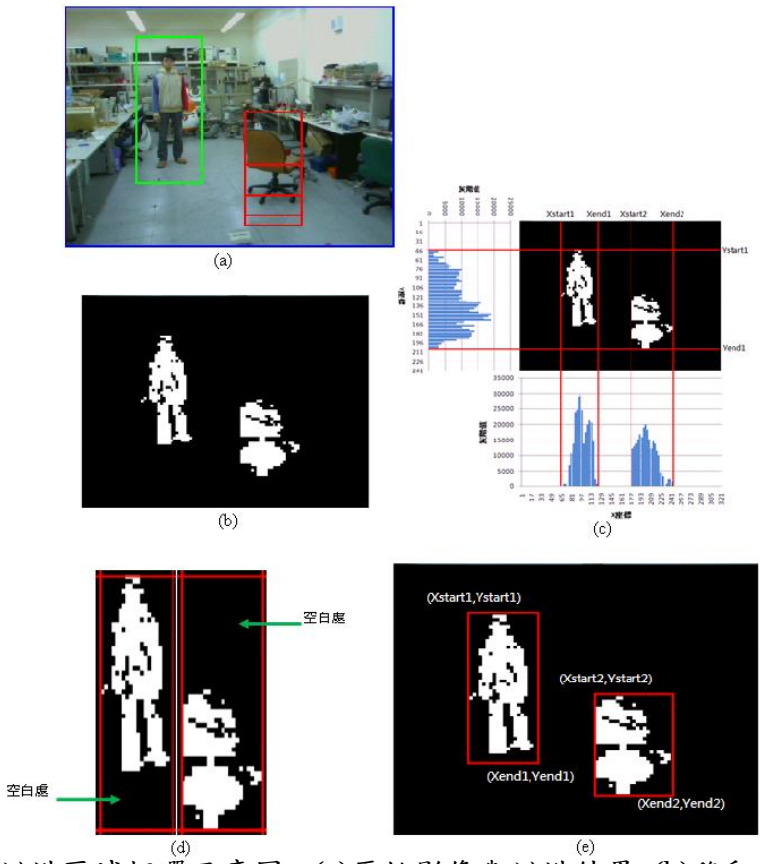


圖 2.3 辨識區域框選示意圖。(a)原始影像與辨識結果 (b)移動目標物前景 (c)第一次的影像投影 (d)框選區域留白 (e)第二次的影像投影

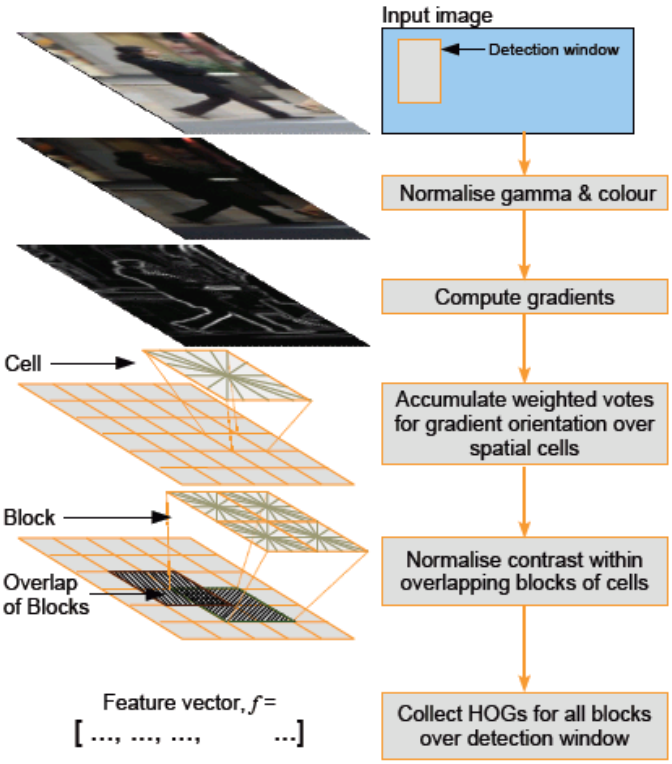


圖 2.4 方向梯度直方圖演算法整體架構圖[21]

2.3.1 辨識視窗大小

由於我們取出的欲辨識區塊的大小都是不固定的，但這邊所使用的 HOG 特徵萃取方式必須要使最終特徵向量的維度相等，故在這邊要將所有的辨識區塊正規化成相同的辨識視窗大小，根據圖 2.5 中所測試的三種辨識視窗大小，其中以 64pixels*128 pixels 的辨識率最佳，故我們在取特徵時會將欲辨識區域都正規化成 64pixels*128 pixels 大小的辨識視窗來做萃取。

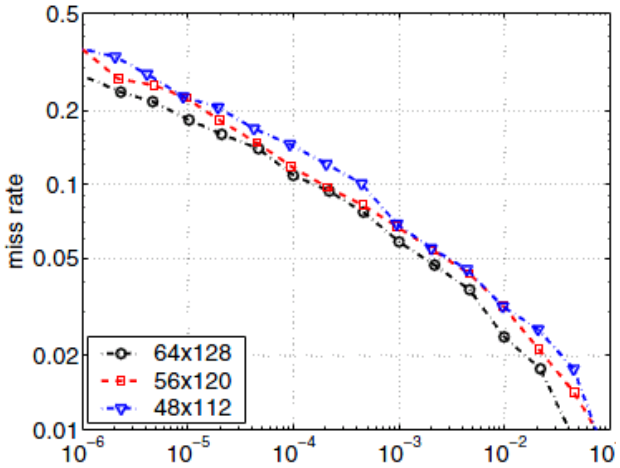


圖 2.5 不同辨識視窗大小之效能[20]

2.3.2 梯度運算

在運算每個像素的梯度強度和方向時，Dalal 測試了許多不同的遮罩性能[21]，其中包含了 uncentred [-1,1]，centred [-1,0,1]，cubic-corrected [1,-8,0,8,-1]，3*3 Sobel masks 和 2*2 diagonal ones $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$, $\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$。比較準確率如表 2.1，可以看到 1-D centred [-1,0,1]的效能是最佳的，故我們令水平遮罩 G_h = [-1,0,1]垂直遮罩 G_v = $[-1,0,1]^T$，使用這兩個遮罩我們可以在點(x,y)得到水平像素差分(horizontal difference) $d_h(x,y)$和垂直像素差分(vertical difference) $d_v(x,y)$，而(x,y)的梯度強度為 mag(x,y)，梯度方向為 θ(x,y)，其中

$$\mathrm{mag}(x,y) = \sqrt{d_h(x,y)^2 + d_v(x,y)^2} \tag{2-1}$$

$$\theta(x,y) = \tan^{-1}\left(\frac{d_v(x,y)}{d_h(x,y)}\right) \tag{2-2}$$

表 2.1 不同梯度遮罩之效能比較表[21]

Mask Type	1-D centred	1-D uncentred	1-D cubic-corrected	2 × 2 diagonal	3 × 3 Sobel
Operator	$[-1,0,1]$	$[-1,1]$	$[1,-8,0,8,-1]$	$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$, $\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$, $\begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$
Miss rate at 10^{-4} FPPW	11%	12.5%	12%	12.5%	14%

2.3.3 特徵點統計

由梯度運算我們已經知道辨識視窗中所有像素的梯度大小和方向，接著將辨識視窗影像分成大小為 8*8 像素且互不重疊的細胞(cell)，如圖 2.6(a)所示，由於梯度方向相差 180 度可視為同一方向來統計，因此將每個細胞依梯度方向在 0^o 到 180^o 分成 9 個方向的統計箱(bin)，也就是 0^o 到 20^o 為 bin1，20^o 到 40^o 為 bin2 依此類推到 bin9，每個細胞內所有像素分別對其所屬的方向統計箱做投票統計，所投的票數為該像素的邊緣強度，這九個方向的資訊可用 9 維的向量來代表，如圖 2.6(b)所示，最後，區塊(block)用其內 4 個細胞方向的統計箱來描述訓練影像在該位置的局部邊緣資訊，可以 36 維向量代表，如圖 2.6(c)所示。

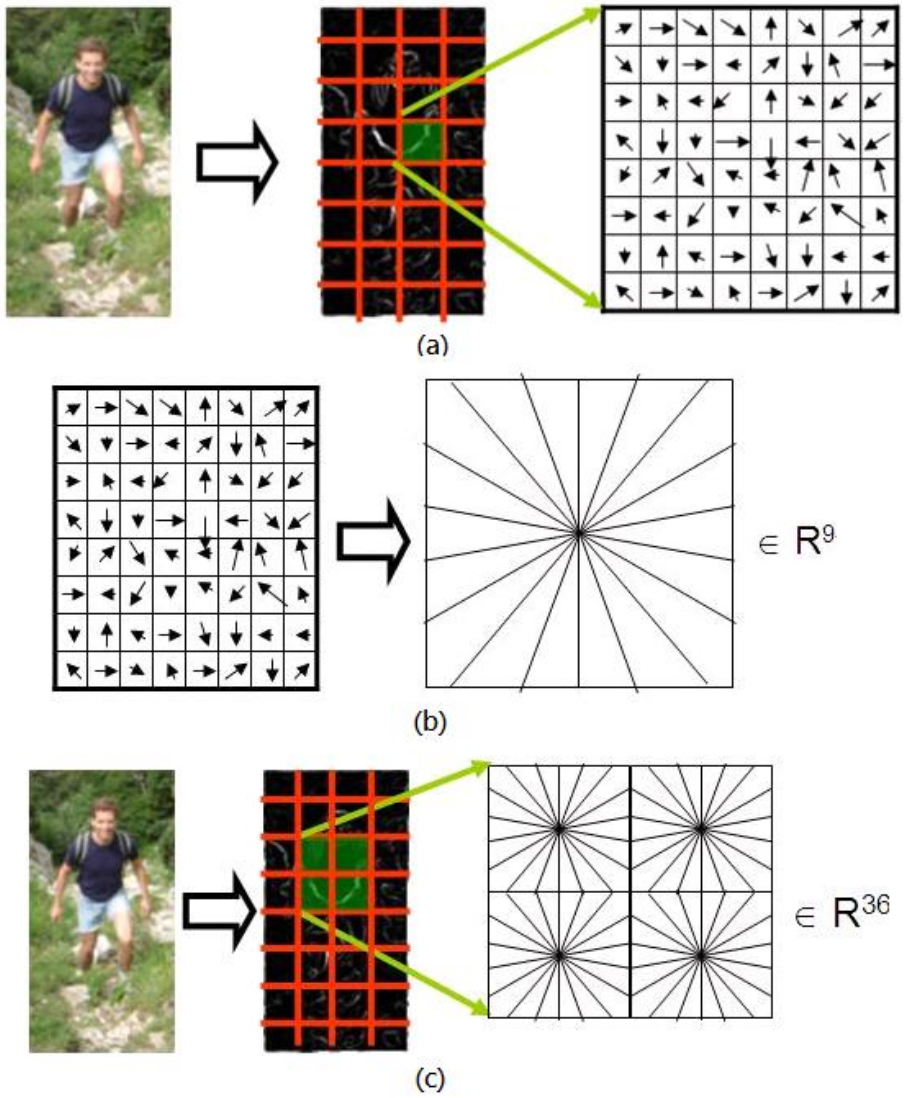


圖 2.6 特徵點統計方式示意圖。(a)圖片分割成細胞(cell)，對 cell 內的像素做梯度運算 (b)將 cell 內的像素分成 9 個方向的統計箱(bin)做投票統計 (c)由 4 個 cell 組成一個區塊(block)，一個 block 可由 36 維的向量表示。[22]

2.3.4 區塊正規化(Block normalization)

將每個區塊(Block)得到的 36 個向量資訊做正規化(normalization)，Dalal[20]有比較過各種常見的方法，例如 L2-*norm*、L2-*Hys*、L1-*sqrt* 和 L1-*norm* 等等，其中，L2-*norm*、L2-*Hys*、L1-*sqrt* 的性能相較於其他方法都比較好，見圖 2.7，故在這邊我們使用 L2-*norm* 來做正規化。L2-*norm*(式 2-3)，中 v 為未正規化的特徵向量，$\|v\|_k$ 為特徵向量的 k-*norm*，ε 為很小的常數。

$$v \rightarrow v/\sqrt{\|v\|^2 + \epsilon^2} \tag{2-3}$$

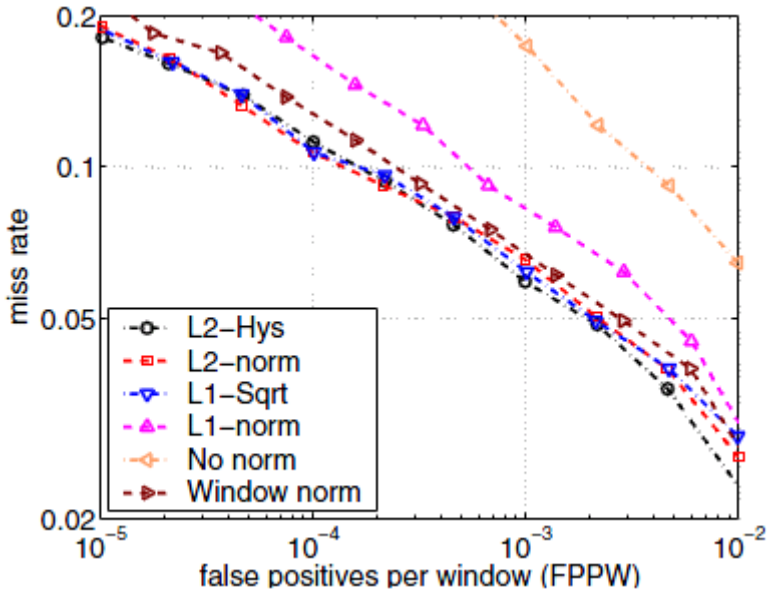


圖 2.7 各種不同正規化法的效能比較圖[20]

2.3.5 HOG 特徵描述

將區塊正規化後，每個區塊和前一個區塊重疊一個細胞來選取特徵，如圖 2.8 所示，一個辨識視窗(detection window)會有 7*15 個區塊，每個區塊有 36 個特徵，故一個辨識視窗會有 3780 個特徵，結果如圖 2.9 所示，最後，要將 3780 個特徵表示成特徵向量的方式來計算(式 2-4)，其中 bin1_1 代表第 1 個 block 中的第 1 個角度值所統計的梯度強度，bin105_36 代表第 105 個 block 中的第 36 個角度值所統計的梯度強度。

$$f = [\text{bin1\_1}, \text{bin1\_2}, \ldots, \text{bin1\_36}, \ldots, \text{bin105\_36}] \tag{2-4}$$

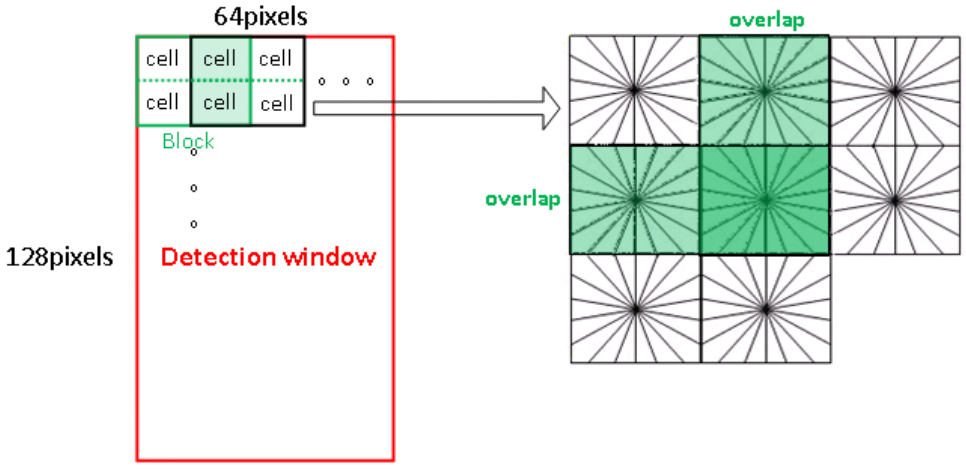


圖 2.8 區塊重疊取樣示意圖

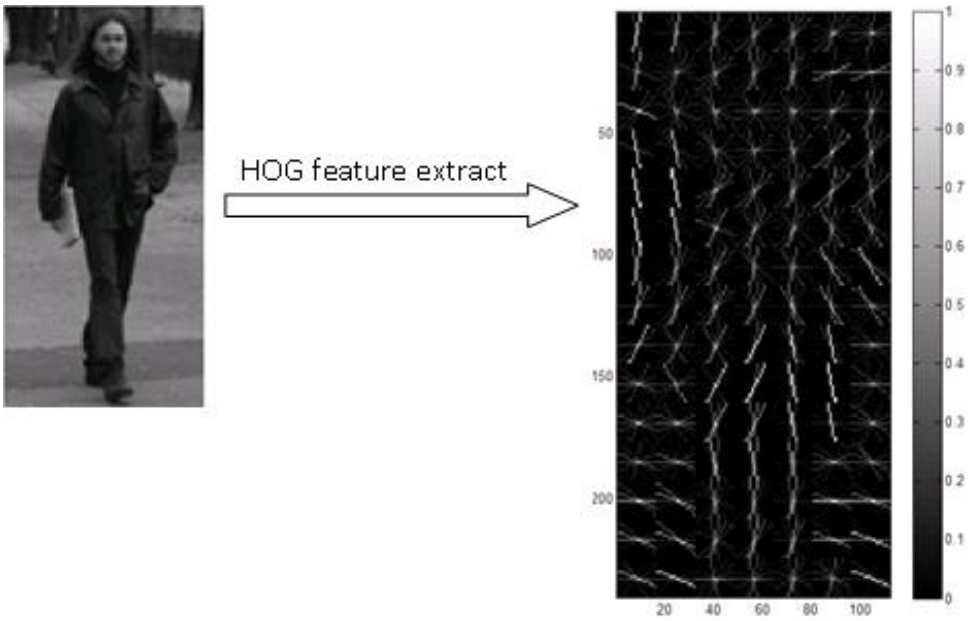


圖 2.9 方向梯度直方圖特徵萃取結果圖

2.4 SVM 分類器

對於一群資料而言，有時候我們會希望依據資料的一些特性來將這群資料分為兩群。而就資料分群而言，我們已知有一些效果不錯的方法。例如：最近鄰居(Nearest Neighbor)、類神經網路(Artificial Neural Networks)和決策樹(Decision Tree)等等方式，而如果在正確的使用的前提之下，這些方式的準確率相去不遠，然而，SVM 的優勢在於使用上較為容易。

2.4.1 SVM 分類器介紹

根據文獻[23]之說明，如圖 2.10 所示，支持向量機將向量映射到一個更高維的空間裡，在這個空間里建立有一個最大間隔超平面。在分開數據的超平面的兩邊建有兩個互相平行的超平面。分隔超平面使兩個平行超平面的距離最大化。假定平行超平面間的距離或差距越大，分類器的總誤差越小。

由圖2.10，實線為需要找出的超平面(Hyper-plane)，而將H1與H2稱之為支持超平面(Support Hyper-planes)，而我們希望能夠找出最佳的分類超平面(Classification Hyper-plane) 使兩Support Hyper-planes之間有最大的邊界(margin)。其計算過程如下：

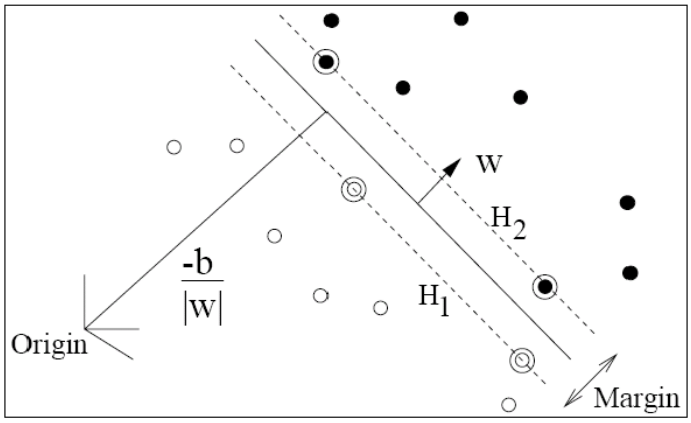


圖 2.10 SVM 原理解說圖[23]

將 Classification Hyper-plane(實線)定義為 $w^T x = - b$ 即 $w^T x + b = 0$。

因此，可以把 Support Hyper-plane(H1、H2)寫為：

$$H1：w^T x + b + \delta \tag{2-5}$$

$$H2：w^T x + b - \delta \tag{2-6}$$

而吾人可以利用一個常數將 w、b 與δ做縮放(Scaling)，因此可以把上兩式重寫為：

$$H1：w^T x + b = 1 \tag{2-7}$$

$$H2：w^T x + b = -1 \tag{2-8}$$

而同時，從 H1 到原點的距離為：|1-b|/ ‖ w ‖ ；H2 到原點的距離為：|-1-b|/ ‖ w ‖ ，因此 H1 與 H2 之間的距離為：2/ ‖ w ‖ 。利用以上的等式，而對於在 $\mathbf{R}^d$ 的空間來說，資料點必須滿足：

$$w^T x + b \geq 1 \, , if \; y_i = 1 \tag{2-9}$$

$$w^T x + b \leq -1 \, , if \; y_i = -1 \tag{2-10}$$

可以將以上兩式改寫為：

$$y_i (w^T x + b) \geq 1 \tag{2-11}$$

同時，因為希望兩個Support Hyper-plans之間距離為最大，因此希望2/ ‖ w ‖ 為 Maximize，亦即：Minimize(‖ w ‖ /2)。

已知訓練資料集合：$\{x_i,y_i\}$,i=1,2,…,n，其中$x_i \in \mathbf{R}^d$, $y_i \in \{1,-1\}$，我們希望利用訓練資料找出一最佳超平面H，以利將未知的x_i歸類。綜合以上，可以在滿足(式2-12)的情況下利用訓練資料找出w^T和b，即得到最佳的H。

$$\begin{cases} \min_{w,b} \frac{1}{2} w^T w \\ y_i\left(\left(w^T x_i\right)+b\right) \geq 1 \end{cases} \tag{2-12}$$

2.4.2 訓練資料庫說明

在訓練資料庫方面，如圖2.11，本論文使用INRIA Person dataset以[32]及MIT Pedestrian Database[33]取出了各一千筆的人形與非人形的訓練資料，來訓練出一組SVM分類器，再使用同資料庫中另外各一千筆的人形和非人形的測試資料來測試此分類器的準確度，其準確度如表2.2所示，人形測試資料中的準確率為96.8%，而非人形誤判為人形的誤判率只有1.5%，所以利用此訓練資料所訓練出來的SVM分類器的效能是可以被使用的。

表 2.2 SVM 對測試資料庫測試之準確率

測試資料	人形	非人形	準確率
人形1000筆	968	32	96.8%
非人形1000筆	15	985	98.5%

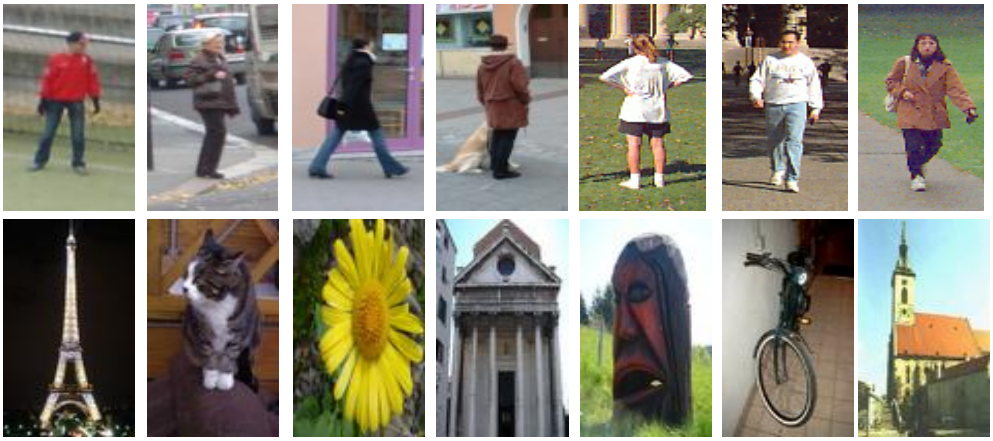


圖 2.11 INRIA person dataset 及 MIT Pedestrian Database 部份人形與非人形影像

2.5 討論

本章討論有關人員偵測的方法，可分為兩部份，一為人員偵測辨識，另一為SVM分類器訓練，利用已知的人形及非人形的資料庫，資料庫中影像的大小皆為64pixels*128pixels，將所需要的SVM分類器訓練出來，接著，利用在2.3節當中提到的人員辨識區域框選方法，將欲辨識的區塊抓出來，然後正規化成64pixels*128pixels大小，取出其方向梯度直方圖特徵，最後利用已訓練好的SVM分類器去辨識是否為人形。

將人辨識出來之後，根據物體運動連續性，前後兩張影像之移動物體區塊移動小於50pixels，就視為同一物體，並且對移動後的區塊持續做人員偵測，若移動後的區塊小於50pixels且偵測為人形，就視為同一個人，吾人即可以對目標人員做持續的追蹤，得到此人在畫面中停留的時間及在影像平面上的座標。

第三章 人體姿態辨識

在第二章我們討論了如何在環境中辨識人員及其所在的位置，本章的目的是要將環境中人的姿態辨識出來。本研究參考 Fujiyoshi and Lipton[3]所提出的星狀骨架來對人類姿勢做代表性的描述以及陳宣勝[4]所提出利用隱藏式馬可夫模型(Hidden Markov Model,HMM)訓練骨架的組合來辨識姿態。星狀骨架是一種藉由連結物件中心到物件輪廓突出點的快速骨架技巧，因為頭和四肢經常是人形狀的突出點，所以辨識的特徵被定義為星狀向量描述。接著，將這些動作視為沿著時間的一連串星狀骨架，再轉換為特徵向量序列，最後，利用隱藏式馬可夫模型來計算出最符合資料庫中的姿態。

3.1 姿態辨識架構

圖 3.1 顯示本論文之姿態辨識系統架構圖，人體姿態辨識大致可以分成特徵萃取、特徵編碼比對及動作序列辨識三個部份。此系統可以辨識出站、走、坐、蹲、躺這五種姿態，在一開始的時候先對框選出的人形，依照寬長比亦即若 W/H>1.5 就辨識成躺的姿態，若 W/H≦1.5 就會進入特徵萃取部份。

在特徵萃取的部份，因為人員偵測的時候可以得到人形的剪影，在此就針對剪影做輪廓的萃取，而在萃取的過程中使用星狀骨架的技術來描述此人形的輪廓。這些被萃取到的星狀骨架會表示成特徵向量的方式作為之後動作序列辨識的資訊。在特徵萃取之後，會使用到向量量化來將特徵向量對編碼序列做比對的動作。吾人建立一套人員動作骨架編碼書，其中包含了每個動作的特徵向量所對映的編碼以及這些編碼的序列所組成的欲辨識的四種姿態。當人員動作的星狀骨架特徵向量被萃取後，就會利用這些特徵向量的方向和大小和已建立的編碼書中的動作去做映對，找出一個最相似的動作編碼，最後就會輸出一個動作編碼的序列。動作序列辨識包含兩方面：訓練和辨識。在訓練方面，本論文先收集一群已標示過的動作編碼序列來建立馬可夫模型，用以辨識出站、走、坐、蹲四種常見人體姿態。

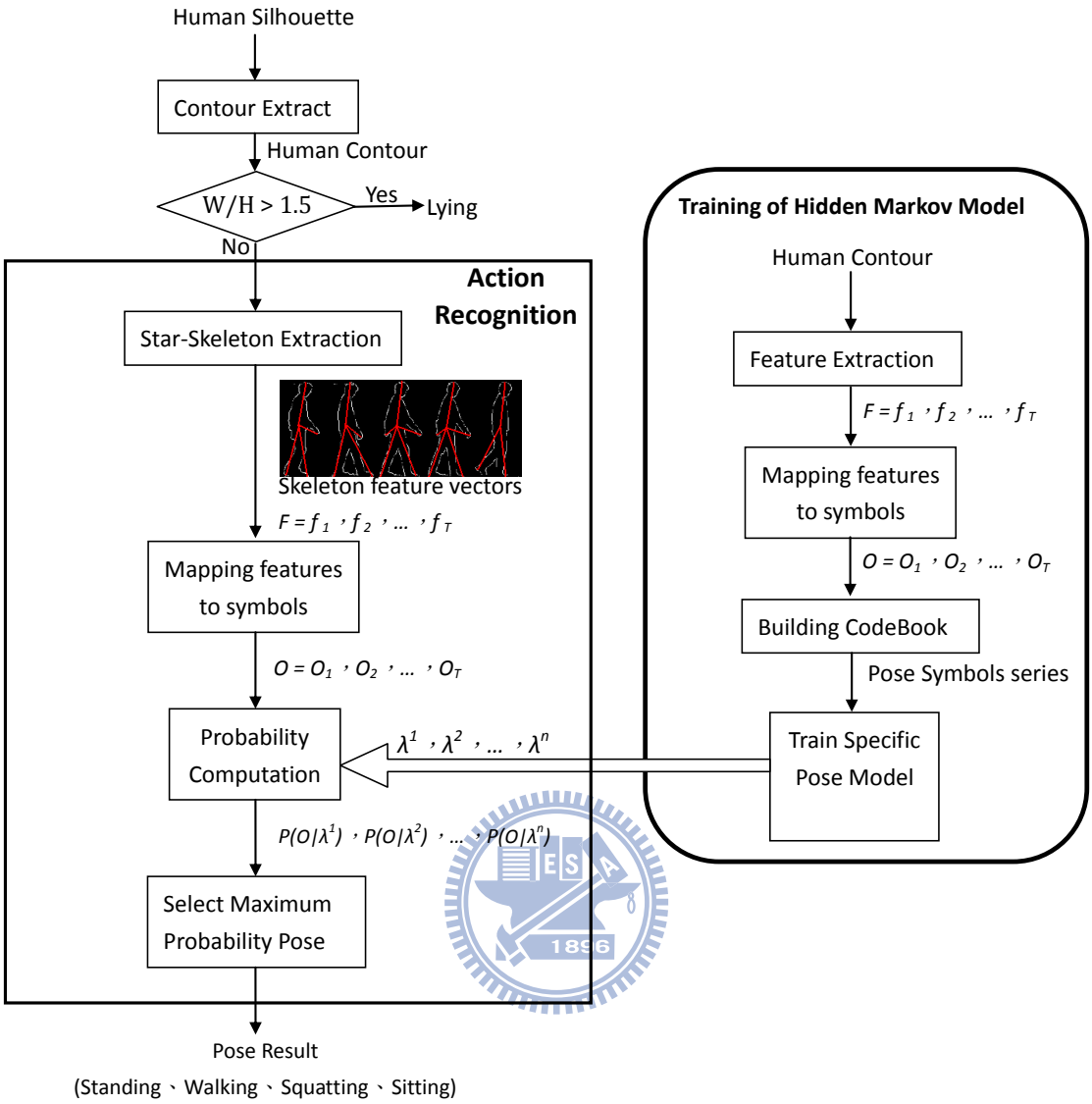


圖 3.1 人體姿態辨識架構圖

針對各種不同姿態的訓練樣本，分別建立出每種姿態的隱藏式馬可夫模型；辨識方面，當有一段測試的人體動作資料想要辨識為何種姿態時，分別到個別姿態所代表的隱藏式馬可夫模型中計算其機率值，由個別姿態的隱藏式馬可夫模型中計算出越高的機率值，則代表這段測試的動作資料越接近此種姿態的隱藏式馬可夫模型。

3.2 隱藏式馬可夫模型

以下有關隱藏式馬可夫模型之說明為參考文獻[4,27,34]中之描述所撰寫。隱藏式馬可夫模型通常使用在語音辨識以及手勢辨識上，且都已經有相當不錯的研

究成果。因此，本論文將利用隱藏式馬可夫模型的特性及其估測模型之優點將其套用在以人體動作為基礎的姿態辨識上，下面將針對隱藏式馬可夫模型做簡單的介紹。

隱藏式馬可夫模型是由某個狀態(state)轉換到另一個狀態的機率所組合而成，隨著時間的前進，狀態的轉換是以一種隨機(stochastically)的方式做改變。像馬可夫模型，狀態在任何時間點只與此狀態的上一個時間點的狀態有關。一個符號(symbol)的產生是由隱藏式馬可夫模型根據狀態轉換機率所推測而得，此推測過程是不能直接觀察到的，只有透過另一群推測過程所產生的可見符號序列所觀察到。在 Rabiner [27]對隱藏式馬可夫模型說明中，說明 HMM 需要兩個參數模型(狀態個數及輸出符號個數)以及三個機率量測(狀態轉換機率、輸出符號機率以及初始狀態機率)，在說明方法之前以下簡單描述參數定義：

T：可見序列長度。

Q = $\{q_1,q_2,\ldots,q_N\}$：狀態(state)集合。

N：模型內的狀態(state)數量。

V = $\{v_1,v_2,\ldots,v_M\}$：輸出符號(symbol)集合。

M：可見的輸出符號(symbol)數量。

A = $\{a_{ij} \mid a_{ij} = P_r(S_{t+1} = q_j \mid s_t = q_i)\}$：狀態轉換機率。

B = $\{b_{jk} \mid b_{jk} = P_r(v_k \mid s_t = q_j)\}$：符號輸出機率。

$\pi = \{\pi_i \mid \pi_i = P_r(s_1 = q_i)\}$：初始狀態機率。

$\lambda = \{A,B,\pi\}$：HMM 機率參數集合。

S = $\{s_t\}$, $t = 1,2,\ldots,T$：狀態 s_t 為時間 t 的狀態。

O = $O_1,O_2,\ldots,O_T$：可見符號序列。

圖 3.2 說明了一個 HMM 的概念和狀態轉換圖，在此例子中存在三個狀態(q_1, q_2, q_3)，每一條指向線代表一個狀態轉換到另外一個狀態，而 a_{ij} 表示從狀態 q_i 轉換到狀態 q_j 的機率。在 HMM 中的每一個狀態會隨機輸出一個符號，其中，狀態

q_j 所輸出符號 v_k 的機率值表示為 b_{jk}。若有 N 個狀態以及 M 個輸出符號，則此輸出轉換矩陣就會是一個 N*M 大小的矩陣。而 HMM 的初始狀態也會由初始狀態機率 π 來決定，所以一個 HMM 會由三個矩陣所組成：狀態轉換機率矩陣 A、輸出符號機率矩陣 B 以及初始狀態機率矩陣 π。

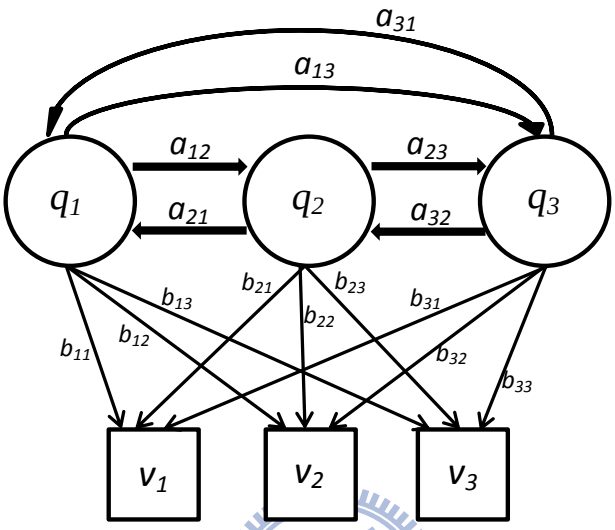


圖 3.2 隱藏式馬可夫模型概念圖

為了要辨識可見的符號序列 O 是屬於哪一個姿態，我們對欲辨識的四種姿態建立各別的 HMM 後，將此序列分別比對各姿態 HMM 的參數 λ_i，其中 $\lambda_i = \{A_i, B_i, \pi_i\}$, i = 1,2,3,4.，找出最符合的姿態。換句話說，當符號序列還不知是哪一姿態時，我們會計算 $Pr(O \mid \lambda_i)$，然後選出機率最高 HMM 所代表的姿態。當我們給了一個可見符號序列 $O = O_1, O_2, \ldots, O_T$ 以及 HMM 的 λ_i，最重要的就是要如何計算 λ_i 會產生此序列的機率 $Pr(O \mid \lambda_i)$是多少。

我們考慮一個固定的狀態序列 $Q = q_1, q_2, \ldots, q_T$，其中 q_1 為初始狀態，在一個特定狀態序列 Q 的觀測序列機率為：

$$P(O|Q,\lambda) = \prod_{t=1}^{T} P(o_t|q_t,\lambda) = b_{q_1 o_1} \times b_{q_2 o_2} \cdots b_{q_T o_T} \tag{3-1}$$

而狀態序列的機率為：

$$P(Q|\lambda) = \pi_{q_1} a_{q_1 q_2} a_{q_2 q_3} \cdots a_{q_{T-1} q_T} \tag{3-2}$$

所以我們可以算得此 HMM 會產生此可見符號序列 O 的機率為：

$$P(O|\lambda) = \sum_{all\ Q} P(O|Q,\lambda)P(Q|\lambda)$$
$$= \sum_{q1\cdots qT} \pi_{q_1} b_{q_1 o_1} a_{q_1 q_2} b_{q_2 o_2} \cdots a_{q_{T-1} q_T} b_{q_T o_T} \quad (3\text{-}3)$$

之後，對每個姿態的 λ 都做式(3-1)到式(3-3)的運算，就可以求出此序列 O 最有可能的姿態。最後，根據人體星狀骨架的辨識可以判對出影片中每張影像所屬的動作，將每一個動作都以一個字元符號(Symbol)來編碼，即本論文中編碼書中的動作編碼(Motion No.)。因此，一段影片中連續的影像動作的變化可用字元符號編碼的變化代表之，下一節我們將要介紹如何萃取出辨識人體動作的星狀骨架特徵，來作為訓練的樣本。

3.3 特徵萃取

本論文採用星狀骨架特徵，此為參考陳宣勝所使用之方法[4]。人體姿態是由一連串的動作序列所組合而成的，而描述這些動作的方法就是利用人體的輪廓形狀。然而，當周圍的其他邊緣和人體動作邊緣很相似時，利用整個人體的輪廓去描述人體動作的效能是很差的，雖然有像主成份分析(principle component analysis, PCA)的方式去去除一些冗餘的特徵點，但是整體的計算量還是非常的大。另一方面，一些簡單的資訊像是寬度和高度可能可以粗略的描述一些動作，但光有這樣的資訊還是不夠準確描述一些相似的動作，而星狀骨架就是一種典型的特徵萃取方式，來描述人體的動作，它具有簡單、即時且穩固的特點，因此本論文選用星狀骨架來當作動作描述的特徵。

3.3.1 星狀骨架特徵描述

將從人體中心點到人體輪廓邊緣局部的向量極值定義為特徵向量，稱之為星狀向量[4]。人體頭部、兩手、及兩腿是在人體輪廓中常見的突出點，因此它們可以適當的描述人體形狀的資訊。因為它們通常是星狀骨架中局部向量的最大

值，本論文定義此向量為五維向量。但有些動作像是兩腳重疊或一手被遮蔽，此時的星狀骨架向量維度就會低於五維，就會有零向量加入星狀骨架的向量描述當中。同樣地，有可能會取到超過五個以上的極值，此時，可以調整低通濾波器來減少人體動作局部極值，控制在五個以內的顯著的極值即可。星狀骨架的概念就是由人體輪廓的中心點連接到四肢所組成的特徵向量，為了找出人體輪廓的四肢，必須計算從人體中心到輪廓的每個點的距離，而四肢就會位於這些距離值的局部極值當中。當中可能會因為雜訊增加了定出四肢的難度，所以可以使用一些平滑濾波器或是低通濾波器讓距離值的訊號變的平滑，以利準確的找出極值。星狀骨架就是將這些極值點連接到人體中心所建構出來的，星狀骨架的處理流程，如圖 3.3 所示，而點 A、B、C、D、E 就是這些距離值的局部極值。

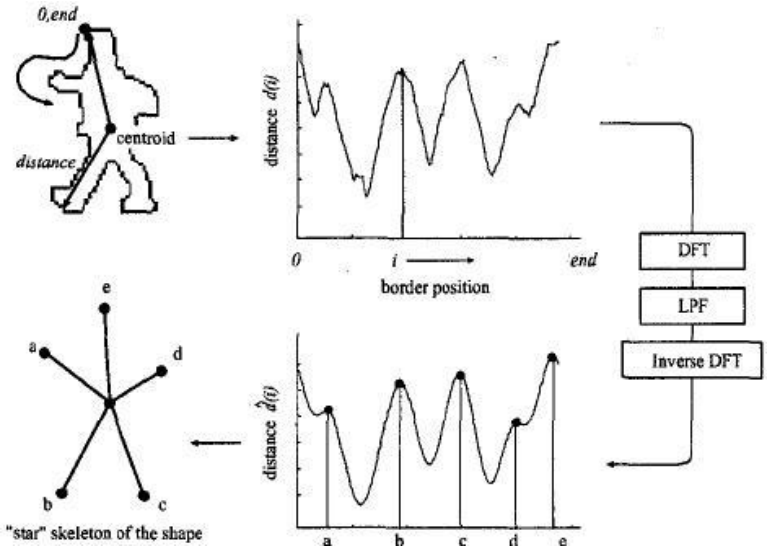


圖 3.3 星狀骨架流程圖[3]

星狀骨架演算法整理如下

輸入：人員輪廓(Human contour)

輸出：星型骨狀圖(A skeleton in star fashion)

1. 計算人體輪廓的中心點(X_c,Y_c)

$$X_c = \frac{1}{N_b}\sum_{i=1}^{N_b} X_i \tag{3-4}$$

$$Y_c = \frac{1}{N_b}\sum_{i=1}^{N_b} Y_i \tag{3-5}$$

N_b是輪廓邊緣的 pixel 數量，而(X_i, Y_i)是每一個輪廓邊緣 pixel 的 x 座標和 y 座標。

2. 計算每個輪廓邊緣(X_i, Y_i)到中心點(X_c, Y_c)的距離d_i

$$d_i = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2} \tag{3-6}$$

將d_i表示成一個一維的向量 d(i) = d_i。

3. 將距離訊號平滑化 d(i) to $\hat{d}(i)$，在頻域中使用低通濾波器來濾除雜訊將距離訊號平滑化。

4. 找出$\hat{d}(i)$最大值的點，這些點就代表人體的頭部和四肢，再分別將這五點與中心點連起來，由(式 3-7)差分方程式的 zero-crossings 找出$\hat{d}(i)$局部極值的點。

$$\delta(i) = \hat{d}(i) - \hat{d}(i-1) \tag{3-7}$$

3.3.2 星狀骨架特徵定義

在本研究當中是利用星狀骨架的特徵向量去描述人體動作，藉由一連串的人體動作透過已訓練好的隱藏式馬可夫模型來計算出最有可能的姿態是什麼。作為一個能夠被訓練且辨識的特徵，此特徵向量的維度就必須固定。因此星狀骨架的特徵向量維度在本研究中定義為五維，因為頭、雙手和雙腳是最常見的局部突出點，也是判斷一個動作最關鍵的部位。如果人體的動作經過星狀骨架的萃取後，可能會因為一些雜訊的因素超過五維，此時就會利用低通濾波器將此星狀骨架的維度降到五維；同樣地，當特徵萃取後也可能因為腳的重疊或手被遮蔽使的維度不到五維，這時就會加入零向量把維度補到五維。由於使用的特徵是向量的形式，所以向量的絕對值大小也會因為人形的大小和形狀有所不同，因此要透過正規化得到相對分佈的特徵向量。要達到此目的可以將特徵向量的 X 分量除以人形的寬度，Y 分量除以人形的高度，可以將 3-6 中的 x_i-x_c 改寫成$(x_i$-$x_c)/H_w$，y_i-y_c 改寫成$(y_i$-$y_c)/H_h$，如 3-8 所示，其中 H_w 為人形的寬度，H_h 為人形的高度。

$$d_i = \sqrt{(\frac{x_i - x_c}{H_w})^2 + (\frac{y_i - y_c}{H_h})^2} \quad (3\text{-}8)$$

3.4 特徵比對

知道如何萃取星狀骨架與定義之後，本節要說明如何將星狀骨架做量化，建立姿態的編碼書，以及如何將從影像中得到的人類動作骨架和編碼書中的動作做比對。要讓 HMM 能夠辨識一般時序的影片，就需要將萃取的骨架特徵序列轉換成動作編碼序列，所以在本論文就使用向量量化的技術來達成。

3.4.1 向量量化

在向量量化中，代號群 $g_j \in R^n$，而此代號群代表了在 R^n 特徵空間中的中心群集。而一個動作的特徵向量就會有指定一個代號來表示，在此稱這步驟為編碼，而所需要的動作的編碼就會組成一本自定的編碼書。因此，編碼書的大小跟 HMM 能輸出的編碼數量是一樣的。每一個動作的骨架特徵向量 f_i 都會被對應到和編碼書中最相近的特徵向量 v_j 如 3-9 所示，然後就會輸出此編碼書中特徵向量所代表的編碼，其中 d(x,y)為 x 向量及 y 向量之間的距離。

$$f_i \rightarrow g_j, if\ j = \text{argmin}_j d(f_i, v_j) \quad (3\text{-}9)$$

3.4.2 編碼書建立

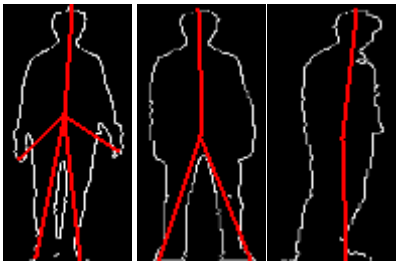
在本論文中，要辨識人員在家中常見的五種姿態，站、走、坐、蹲和躺，但是躺在圖 3.3 的姿態辨識流程圖中可以看到，我們預先辨識了躺的姿態，所以這邊我只建立了站、走、坐和蹲這四種姿態的編碼書。而在編碼書中的這四種姿態都有它所代表的動作序列所組成，而這些動作的星狀骨架特徵向量也都有指定一個代碼來表示，如圖 3.3 所示，即本論文所建立的編碼書。

3.5 結論與討論

由本章可以知道，本研究是利用星狀骨架去描述人體動作，包含頭部以及四肢的特徵，再由一些動作的組合組成我們需要辨識的姿態，將這些姿態組合所代

表的編碼序列建立編碼書，最後，利用隱藏式馬可夫模型去建構站、走、蹲、坐，四種姿態的模型。當輸入一連串的動作時，會比對在編碼書中所代表的編號，轉換成編碼序列輸出，將者些編碼序列代入由四種姿態所建立其個別的馬可夫模型當中，算出最有可能的姿態。

1. Stand

Star-Skeleton features	
Motion No.	1 2 3
Pose	Stand

2. Walk

Star-Skeleton features	
Motion No.	4 5 6 7 8 9 10 11 12
Pose	Walk

3. Squat

Star-Skeleton features	
Motion No.	13 14 15 16 17 18 19 20 21
Pose	Squat

圖 3.4 各姿態所包含的動作編號編碼

4. Sit

Star-Skeleton features								
Motion No.	22	23	24	25	26	27	28	29
Pose	Sit							

圖 3.4(續) 各姿態所包含的動作編號編碼書

第四章 活動偵測設計

由第二章的人員偵測和第三章的姿態辨識，我們已經可以得到人員在影像平面上的位置座標、停留時間以及當前的姿態。本章將介紹如何把這些資訊搭配人員所處的環境位置，利用有限狀態機(Finite State Machine,FSM)來做人員活動的偵測。在本論文中的活動偵測狀態機，是結合人員所在的環境位置、停留時間及其目前姿態當作狀態因子，在 4.1 節中，會先介紹本論文所提出一套設定環境邊界的方法。4.2 節中會介紹在本論文中為活動偵測所設計的狀態機。

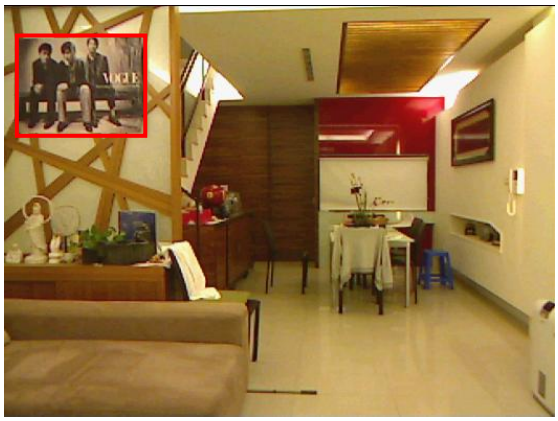
4.1 環境位置邊界設定

由於本論文所提出的人員活動偵測，需要利用目前人員所在的環境位置來做為辨識的依據。例如：人員在「客廳」看電視，客廳就是所處的環境位置。而當機器人所看到的畫面可能同時存在兩種環境，例如餐廳和客廳，此時就需要一個環境邊界去分辨餐廳和客廳的分界處，即環境邊界。

如圖 4.1 顯示，當機器人移動一段距離後，在影像平面上的邊界也會跟著移動，例如選定圖 4.1 紅框處之目標物，當影像平面上目標物之影像放大後代表機器人往前面的方向移動，相對的環境的邊界在影像平面上會往下移。本論文提出一種方法，利用辨識環境目標物的方式去自動訂出環境邊界。

本論文所提出的方法需要先設定一個目標物，當辨識出影像畫面中之目標物後，可以透過單應矩陣(Homography)的轉換關係找出當下目標物外框在影像平面上的位置。由於目標物和環境邊界在環境中是固定不會移動的，在影像平面上環境目標物和環境邊界也會存在一個相對應的關係，所以我們可以利用此特性求得當機器人在不同遠近下，目標物和環境邊界在影像平面上的相對關係(在影像平面上目標物放大縮小與環境邊界前後移動的關係，詳細會在 4.1.3 小節中說明)。當我們得到上述的關係式之後，在實際測試畫面中，當機器人辨識出目標物在影像平面上的位置後就會根據此關係式產生一條環境的邊界。

(a)



(b)

圖 4.1 影像平面移動前後所改變的目標物大小和環境邊界位置(a)移動前 (b)攝影機往前移動後

4.1.1 環境邊界設定流程圖

環境邊界設定之步驟說明如圖 4.2 及圖 4.3 所示，其流程圖如圖 4.4 所示，包含目標物體辨識和邊界標定兩個部份。目標物辨識部份，當影像從攝影機擷取之後，透過加速強健特徵點演算法(Speed Up Robust Feature, SURF)特徵點比對的方式將我們欲辨識的目標物找出來。因為目標物的特徵點分布可能在影像平面上的任何位置，於是為了得到穩定的參考點，我們將比對成功的特徵點利用 homography 的方式去算平面轉換矩陣，因為這個平面轉換矩陣包含著縮放及旋轉因子，所以我們可以由這個平面轉換矩陣框出目標物在影像平面中的區域。由於在目標物辨識中已經得到影像平面中目標物的區域，所以在邊界標定部份，要先計算出目標物與環境邊界的關係式。當攝影機移動時，就可以透過此關係式來估測出環境邊界。

環境邊界設定流程步驟：

Step1) 將輸入的影像畫面利用 SURF 找出欲辨識物體與資料庫的 Matching Point。

Step2) 將比對成功的特徵點利用 Homography 找出目標物在影像平面四個角的座標。

Step3) 利用目標物與環境邊界的相對關係式，找出目前影像畫面中的環境邊界。

圖 4.2 環境邊界設定流程步驟(辨識)

目標物與環境邊界關係式計算步驟：

Step1) 將攝影機擺到離邊界最近的位置(即再靠近邊界就會消失)，紀錄目標物四個角的座標以及環境邊界的位置。

Step2) 將攝影機擺到離邊界最遠的位置(即能辨識到目標物的最遠距離)，紀錄目標物四個角的座標以及環境邊界的位置。

Step3) 將攝影機擺到中間的位置，紀錄目標物四個角的座標以及環境邊界的位置。

Step4) 利用上面三步驟去找出目標物與環境邊界的關係式。

圖 4.3 環境邊界設定流程步驟(目標物與環境邊界關係式計算)

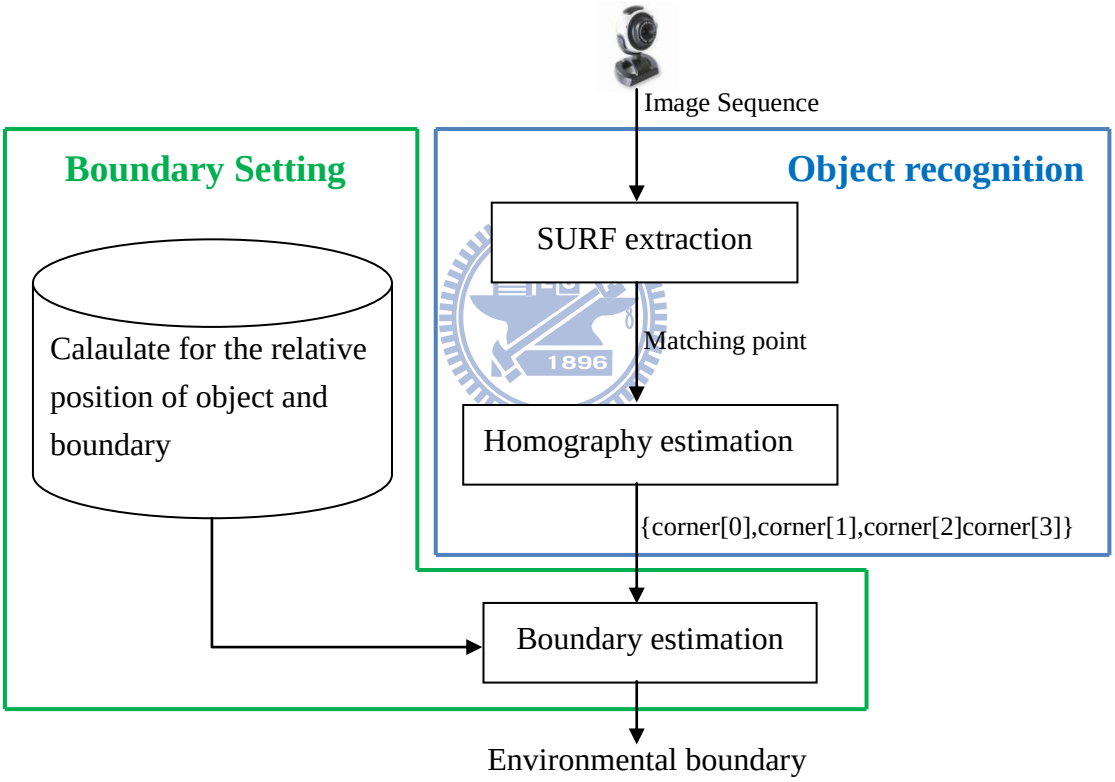


圖 4.4 邊界設定流程圖

4.1.2 環境物體辨識

要從攝影機擷取影像中找出我們所要的目標物，需要有一個辨識物體的演算法，在本論文中選用 SURF[28]來當做特徵擷取的方法，SURF 和 SIFT[15]一樣具有尺度不變且抗旋轉的優點，但 SURF 的運算上較 SIFT 快速。由 SURF 擷取

出特徵點後，再利用最近鄰居演算法（Nearest neighborhood algorithm）找到與資料庫比對成功的目標物特徵點，利用比對成功的特徵點透過 Homography 找出目標物在影像平面上的位置。

4.1.2.1 SURF 特徵擷取

SURF [28]為 Bay *et al.*所提出的一種新的尺度不變且抗旋轉的特徵點偵測及描述法。這個方法的設計概念在於發展出一套重複性、獨特性以及強健性優於現存方法的特徵點擷取演算法，且能夠有更快的運算速度。在計算過程中，有使用到積分影像(Integral image)的運算，它是讓 SURF 能快速計算的關鍵[29]。圖 4.5 即為資料庫影像利用 SURF 所擷取到的特徵點。

4.1.2.2 環境目標物特徵點比對

當計算出影像平面中的特徵點後，便可以透過與資料庫中所儲存的特徵點資訊進行比對，藉此得以判斷當前影像中是否存在所需的目標物。本論文利用最近鄰居演算法（Nearest neighborhood algorithm），如 (4-1)式，其中$Des_c(i)$代表當前影像中該特徵描述向量的第i個元素，$Des_d(i)$代表資料庫影像中該特徵描述向量的第i個元素。藉由比較當前影像中所有特徵點描述向量與資料庫中現存的特徵點描述向量，尋找其兩點距離最短的，則其本身則可能為相同的特徵點。

$$d = \left(\sum_{i=1}^{128} (Des_c(i) - Des_d(i))^2 \right)^{1/2} \quad (4\text{-}1)$$

圖 4.5 被偵測到的特徵點

4.1.2.3 Homography

單應矩陣(Homography)[35]的功用在於找到兩個影像平面中，點跟點之間的對應關係。在影像平面上，一組對應的特徵點之間，存在著一種線性變換的關係，而Homography定義為其中一個影像平面上的點P_a轉換到另一個影像平面上的點P_b之間的線性轉換。Homography 是由一3*3的非奇異矩陣(Non-singular)矩陣所決定，因為具有縮放因子ω'的關係，能夠反映出目標物在影像平面上，跟資料庫影像相比，尺度大小的變化倍率。它具有8個自由度(degree of freedom)，因為平面上一個點(x, y)具有兩個自由度，所以決定一個Homography至少需要四點以上的對應關係。

$P'_b=H_{ab}P_a$，其中P_a，$P_b \epsilon P^2$， H為3*3的Homography matrix

$$\mathrm{H}_{ab}=\begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \tag{4-2}$$

$$p'_b = H_{ab}p_a = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}\begin{bmatrix} x_a \\ y_a \\ 1 \end{bmatrix} = \begin{bmatrix} \omega' x_b \\ \omega' y_b \\ \omega' \end{bmatrix} \tag{4-3}$$

$$x_b = \frac{h_{11}x_a + h_{12}y_a + h_{13}}{h_{31}x_a + h_{32}y_a + h_{33}}，y_b = \frac{h_{21}x_a + h_{22}y_a + h_{23}}{h_{31}x_a + h_{32}y_a + h_{33}}，\omega' = h_{31}x_a + h_{32}y_a + h_{33} \tag{4-4}$$

轉換矩陣H共有9個未知數，由於齊次座標轉換有比例相等的關係，於是假設轉換矩陣H中的某一個未知數固定(如: h_{33}=1)，如此一來待解的參數變成8個，而平面上一組對應點可提供2個線性獨立的方程式，因此8個未知數需要4組對應點提供8個線性獨立的方程式，而這4組對應點是從所有的對應點中任意選出的，所以要得到一個轉換矩陣至少要4組對應點。當求得轉換矩陣後，可算出每個對應點的誤差，因此符合最小誤差的對應矩陣極為我們所需要的對應矩陣。我們可以將

(4-4)改寫成(4-5)。

$$\begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x'_1 x_1 & -x'_1 y_1 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -y'_1 x_1 & -y'_1 y_1 \\ x_2 & y_2 & 1 & 0 & 0 & 0 & -x'_2 x_2 & -x'_2 y_2 \\ 0 & 0 & 0 & x_2 & y_2 & 1 & -y'_2 x_2 & -y'_2 y_2 \\ x_3 & y_3 & 1 & 0 & 0 & 0 & -x'_3 x_3 & -x'_3 y_3 \\ 0 & 0 & 0 & x_3 & y_3 & 1 & -y'_3 x_3 & -y'_3 y_3 \\ x_4 & y_4 & 1 & 0 & 0 & 0 & -x'_4 x_4 & -x'_4 y_4 \\ 0 & 0 & 0 & x_4 & y_4 & 1 & -y'_4 x_4 & -y'_4 y_4 \end{bmatrix} \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \end{bmatrix} = \begin{bmatrix} x'_1 \\ y'_1 \\ x'_2 \\ y'_2 \\ x'_3 \\ y'_3 \\ x'_4 \\ y'_4 \end{bmatrix} \quad (4\text{-}5)$$

當我們算出轉換矩陣之後，就可以經由(4-6)來驗證特徵點之間的轉換關係跟此轉換矩陣的關係是否一致。將資料庫影像中的特徵點代入(4-10)中，可以得到資料庫中的特徵點對應在目前影像平面上的特徵點位置。

$$p_b = p'_b / \omega' = \begin{bmatrix} x_b \\ y_b \\ 1 \end{bmatrix} \quad (4\text{-}6)$$

因為homography具有對應旋轉及尺度變化的特性，所以即使目標物在不同距離或是不同角度，我們都可以準確的框出目標物在目前影像平面上的位置。如圖4.6所示，左邊是資料庫影像，首先藉由所擷取到的特徵點，求得資料庫影像跟目前影像平面之間的轉換矩陣，接著透過轉換矩陣H，我們可以得到目標物在目前影像平面(image plane)中的位置。

4.1.3 環境物體與邊界相對關係之訓練

目前已經可以將環境中欲辨識的物體找出，如圖4.7所示，畫線的部份即為影像中目標物與資料庫影像比對成功的點，而影像平面上目標物所框選的區域就是透過homography的轉換關係所求得目標物在影像平面上的四個角座標，現在要透過目標物去辨識環境的邊界，所以必須先定義環境邊界與目標物的關係。

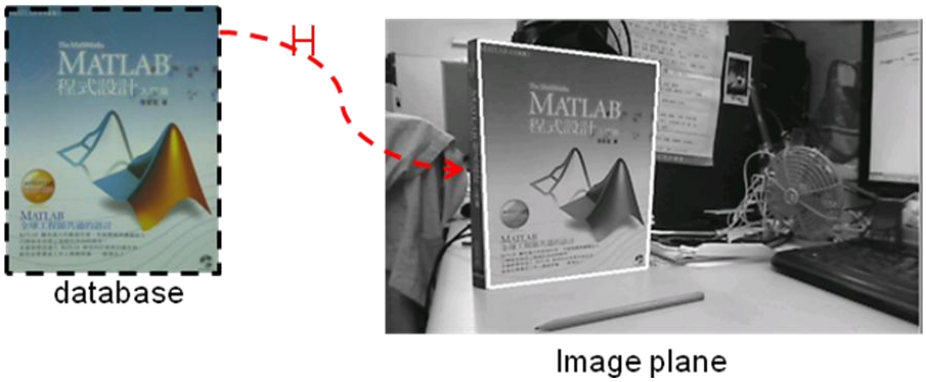


圖 4.6 透過轉換矩陣 *H* 得到資料庫物件在目前影像平面中之位置

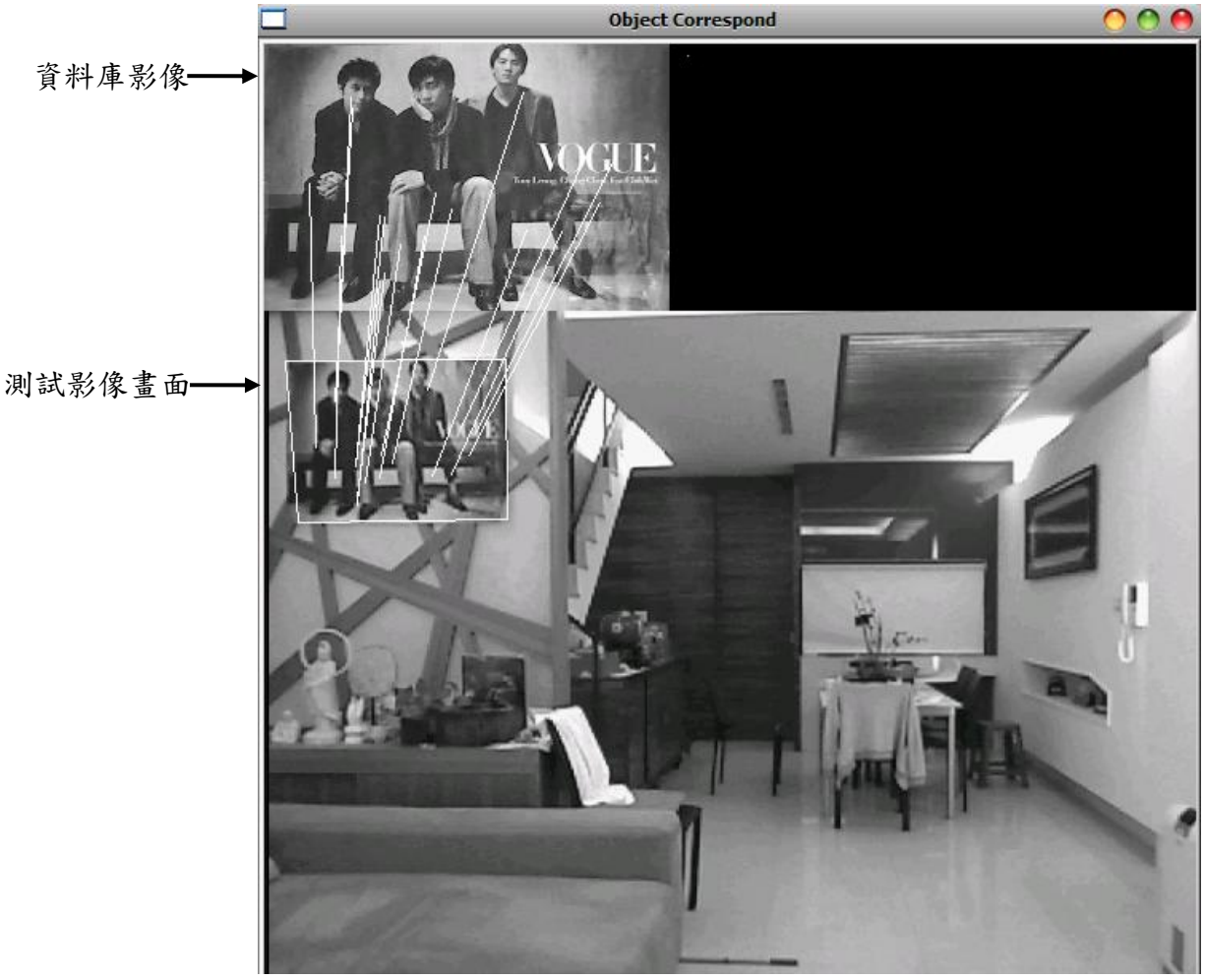


圖 4.7 目標物與資料庫特徵比對結果及目標物在影像平面上被框選的位置

在本研究中，目標物有一些預設的條件存在：1.面積要夠大、2.平面的物體、3.平行於地面。首先，先定義目標物和環境邊界的一些相關參數，如圖4.8所示。

0，1，2，3代表目標物四個角點的編號，$m_{boundary}$為環境邊界的斜率。

$$d_1 = C_r\_y_3 - C_r\_y_0 \tag{4-7}$$

其中$C_r_y_3$即為3號角點的y座標，$C_r_y_0$即為0號角點的y座標。
目標物下底線的斜率為：

$$m_{23} = \frac{C_r\_y_2 - C_r\_y_3}{C_r\_x_2 - C_r\_x_3} \tag{4-8}$$

其中，$C_r_y_2$為2號角點的y座標，$C_r_y_3$為3號角點的y座標，$C_r_x_2$為2號角點的x座標，$C_r_x_3$為3號角點的x座標。

由圖4.8所示，環境邊界由一個參考點(B_r_x,B_r_y)和斜率$m_{boundary}$所形成，由於環境目標物在現實環境中是平行於邊界的，所以在影像平面上我們將邊界的斜率設定與目標物下底線斜率相同，即$m_{boundary} = m_{23}$ 。接下來就要將環境邊界的參考點找出，而環境邊界參考點的x,y座標定義如下：

$$B_r\_x = C_r\_x_0 \tag{4-9}$$

$$B_r\_y = C_r\_y_0 + d_2 \tag{4-10}$$

其中， B_r_x為環境邊界參考點的x座標，B_r_y為環境邊界參考點的y座標，$C_r_x_0$為0號角點的x座標，$C_r_y_0$為0號角點的y座標。

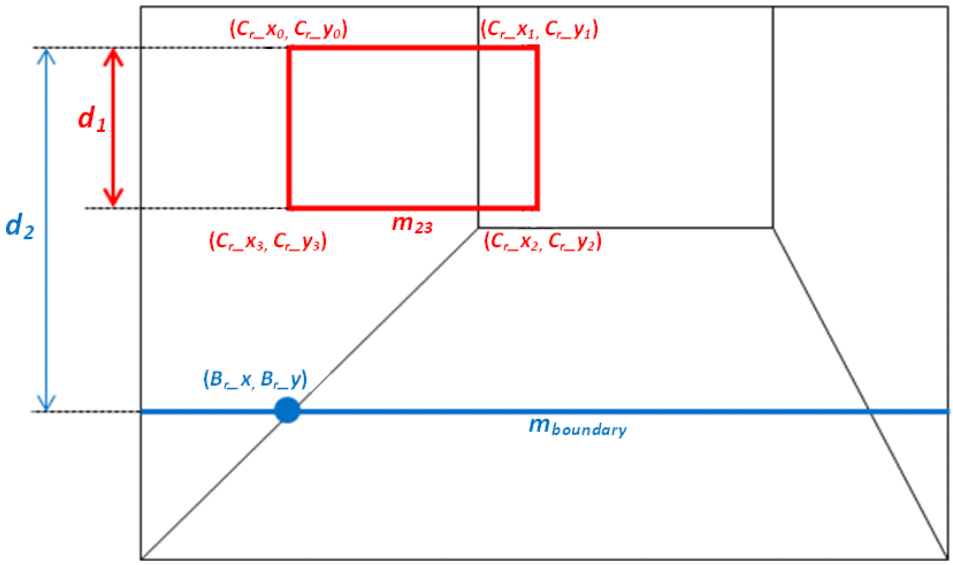


圖 4.8 目標物及環境邊界參數

在這邊我們定義環境邊界參考點的x座標和目標物0號角點的x座標相同如(4-9)所示，而環境邊界參考點的y座標定義為0號角點的y座標加上d_2如(4-10)所示。我們知道，當鏡頭往前移動時，目標物會變大(d_1變大)，環境邊界會往下方移動(d_2變大)；當鏡頭往後移動時，目標物會變小(d_1變小)，環境邊界會往上方移動(d_2變小)。所以我利用最遠、中間、最近三種距離(在這邊以示意圖4.9表示)來求得d_1,d_2的關係式，藉以求得環境邊界(boundary)的y座標。而最遠指的是能夠辨識到目標物的最遠距離，最近就是環境邊界已經要消失不見的距離，而中間就是在兩者之間的距離。在最遠的距離時可以得到$d_1 = 71$、$d_2 = 257$，在中間距離時可以得到$d_1 = 93$、$d_2 = 341$，在最近距離時可以得到$d_1 = 114$、$d_2 = 449$。故假設d_1和d_2存在一個二元一次方程式的關係式如(4-11)所示。

$$d_2 = \alpha * {d_1}^2 + \beta * d_1 + \gamma \tag{4-11}$$

將上述所求得的三組d_1、d_2值代入(4-11)，求得α=0.0308、β=-1.2341、γ=229.3572，如(4-12)所示。

$$d_2 = 0.0308 * {d_1}^2 - 1.2341 * d_1 + 229.3572 \tag{4-12}$$

總結以上，當測試影像輸入後，會先將目標物的座標在影像平面中找出來，接著可以利用上述的方式將環境邊界的參考點座標(B_r_x,B_r_y)和斜率$m_{boundary}$找出，就可以自動調整出目前環境邊界的位置。如圖4.9所示，在不同距離下所自動調整的環境邊界。

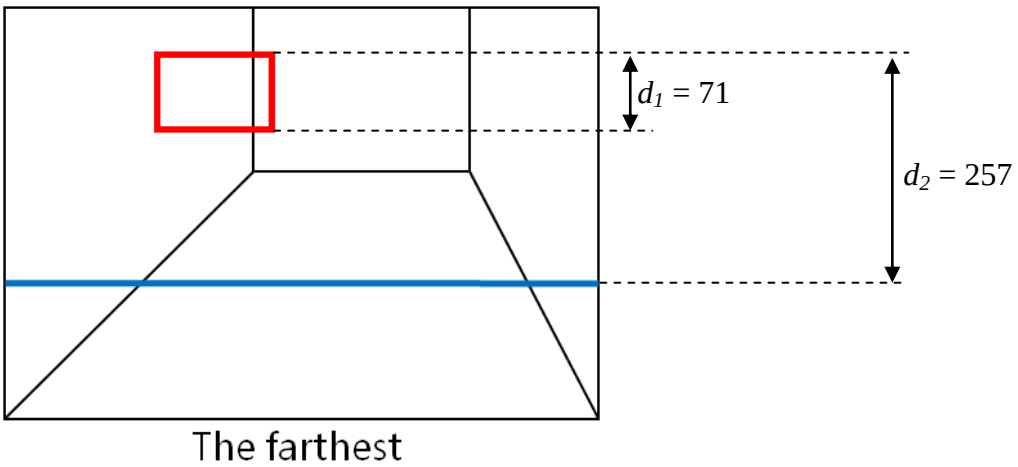


圖 4.9 以三種不同距離訓練 d_1 與 d_2 關係式之示意圖

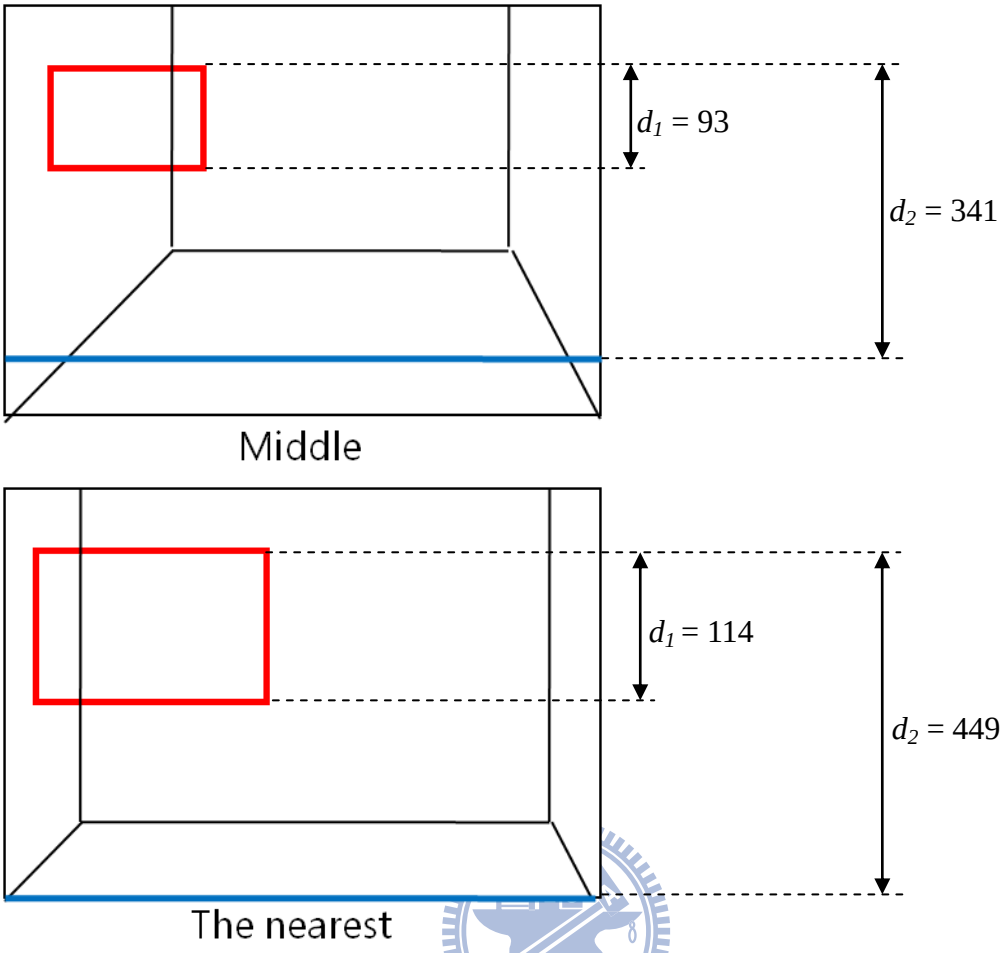


圖 4.9(續) 以三種不同距離校正其 d_1 與 d_2 關係式之示意圖

圖 4.10 在不同距離下所自動調整的環境邊界

4.2 活動判定之有限狀態機設計

由上節中，已經可以自動調整環境的邊界，所以在攝影機移動過後的狀況下，也可以判別出邊界位置，進而得知人員目前所在的環境是哪裡。再由第二章得到的人員停留時間和第三章得到的人員目前的姿態，我們可以將這些資訊透過一種狀態組合的判斷方式，來判定出目前人員的行為可能是什麼。在本研究中透過有限狀態機(Finite State Machine,FSM)的方式，結合所有資訊來做人員活動的判斷。本論文所設計的FSM如圖4.11所示，當人員一被偵測到時，就會利用環境邊界來判定人員在哪一個環境中，在本論文中設定兩個環境：餐廳和客廳，當剛開始處的環境時就會有走入此環境的活動判定，而在所處環境中超過10frames之後就會判定該人員是在此環境當中。例如，當人員由餐廳走入客廳時，活動偵測就會判定為：走入客廳，當人員待在客廳超過10frames後就會判定為：待在客廳。當人員所處的環境判定了之後，就會對人員的不同姿態和所停留的時間做出不同的活動判定。例如：若目前人員判定為在客廳中，此時人員的姿態是：站，就會判定為站在客廳裡，若人員的姿態是：坐，就會判定為坐在客廳裡。而在餐廳中的活動就如圖4.11所示，在此就不多加舉例說明。這邊有一個特殊的人員活動：跌倒，跌倒的判定是，當人員的姿態由站→躺或由走→躺，就會被判定為跌倒，而當時處在哪一個環境當中，就會判定為在哪裡跌倒。若是由坐→躺或蹲→躺，就只會被判定為躺。接著，在第五章的實驗結果說明中，會有實際的測試影片，測試本方法之可行性。

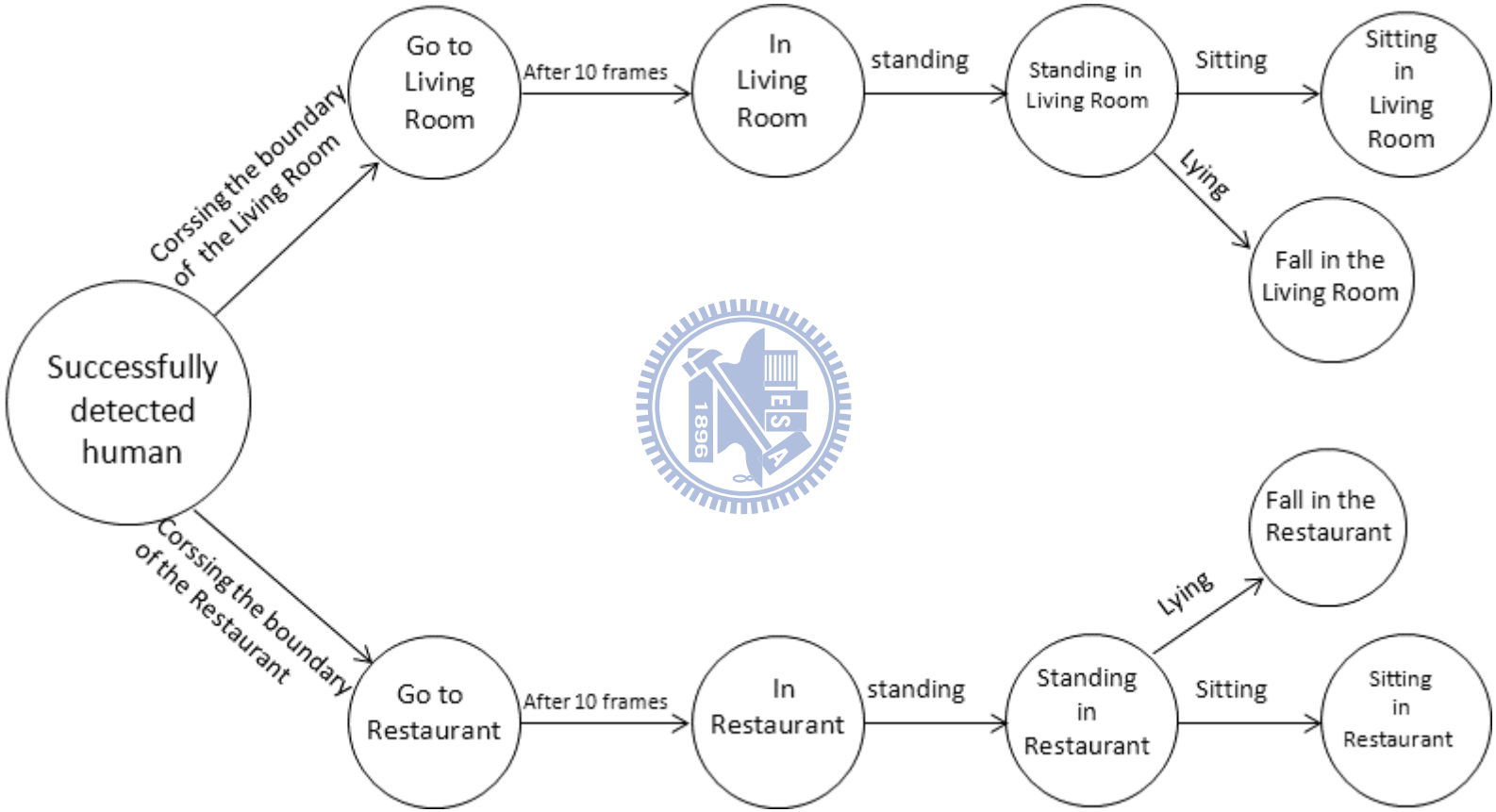


圖4.11 人員活動偵測判定之有限狀態機

第五章 實驗結果

由第四章，人員活動偵測設計可以知道，本論文所提出的人員活動辨識方法，是需要結合人員偵測、姿態辨識以及環境位置辨識，再透過狀態機去偵測的。所以要能夠使活動偵測的準確率高，也必須仰賴人員偵測及姿態辨識系統的準確率。所以本章的實驗結果會分成三個部份去測試，5.1 節中的人員偵測辨識結果，測試在環境中辨識到人的準確率。5.2 節中的人體姿態辨識結果去測試五種姿態的準確率。最後，在 5.3 節中會呈現人員活動偵測的實驗結果。

5.1 人員偵測辨識結果

人員偵測的部份，分別對人形不同角度各取 300 張(0^o、45^o、90^o、135^o、180^o、225^o、270^o、315^o)測試人員偵測準確度，如圖 5.1 到圖 5.8 所示。藉以希望人員在不同面向於攝影機時都可以被成功的辨識。

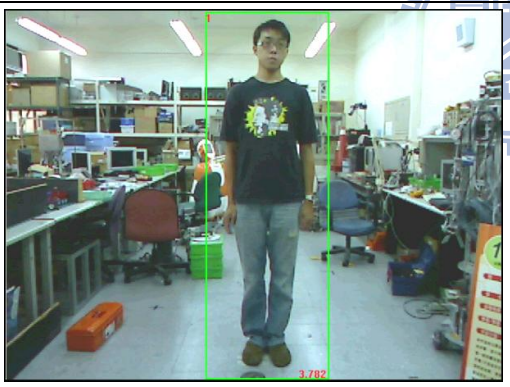
0°	Right	Wrong	Detection rate
	297	3	99%

圖 5.1 人員偵測在 0 度角時的準確率

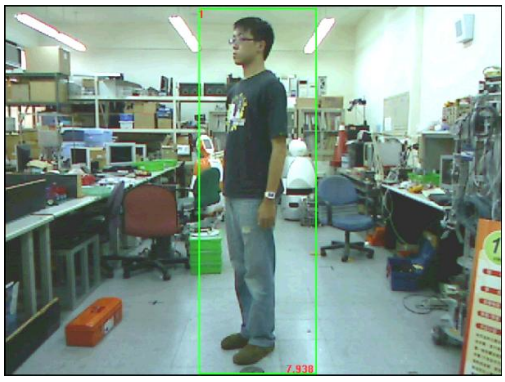
45°	Right	Wrong	Detection rate
	282	18	94%

圖 5.2 人員偵測在 45 度角時的準確率

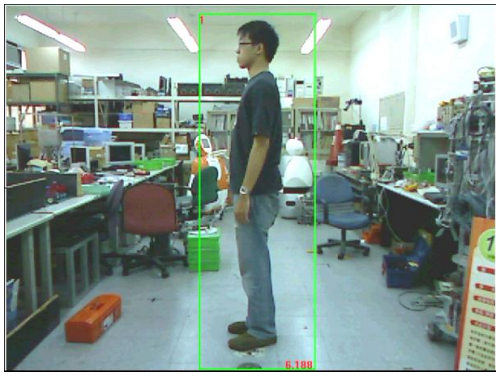
90°	Right	Wrong	Detection rate
	272	28	90.66%

圖 5.3 人員偵測在 90 度角時的準確率

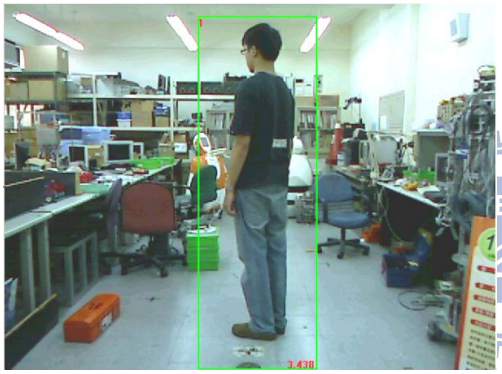
135°	Right	Wrong	Detection rate
	286	14	95.33%

圖 5.4 人員偵測在 135 度角時的準確率

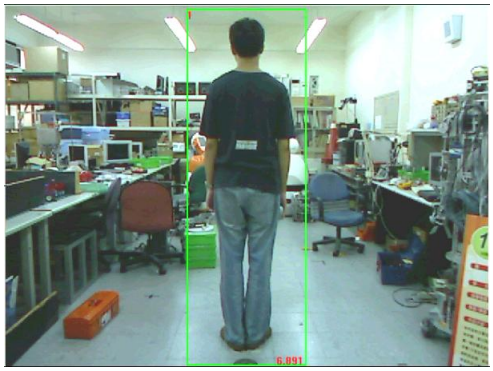
180°	Right	Wrong	Detection rate
	293	7	97.66%

圖 5.5 人員偵測在 180 度角時的準確率

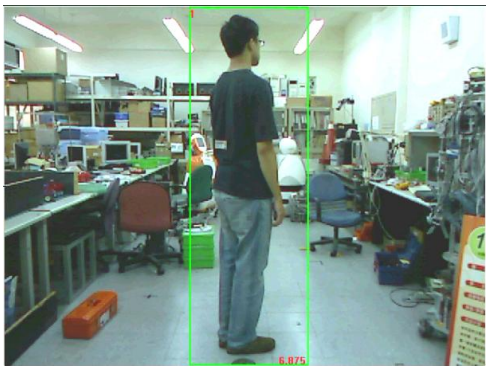
225°	Right	Wrong	Detection rate
	291	9	97%

圖 5.6 人員偵測在 225 度角時的準確率

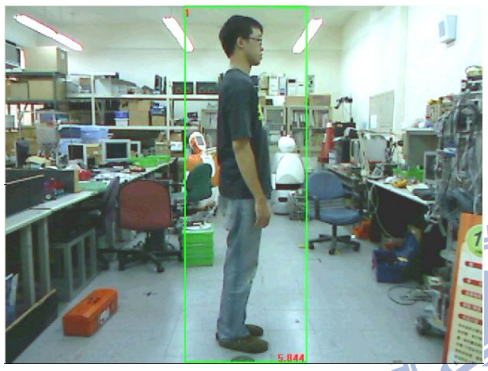
270°	Right	Wrong	Detection rate
	288	12	96%

圖 5.7 人員偵測在 270 度角時的準確率

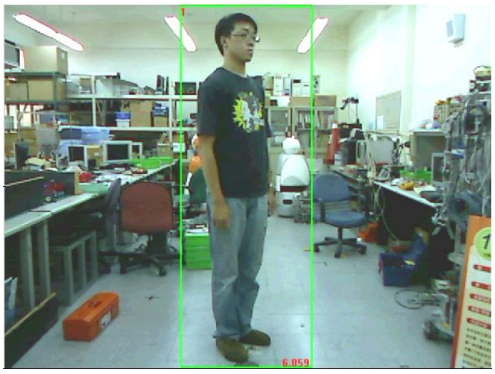
315°	Right	Wrong	Detection rate
	279	21	93%

圖 5.8 人員偵測在 315 度角時的準確率

上面實驗結果中，最低辨識率為 90 度的 90.66%，最高的辨識率為 0 度的 99%，平均辨識率也可達 95.33%。經過實驗測試，HOG 的特徵萃取加上 SVM 的分類器對人員偵測來說是可行且效果良好。

5.2 人體姿態辨識結果

人體姿態部份，對不同距離下(2.5m、3m、3.5m、4m、4.5m、5m)之站、走路、蹲、坐以及躺，這五種人體姿態辨識成功率做測試，如表 5.1 到表 5.6 所示。由於人員在環境中走動，對攝影機的距離都不相同，藉此測試在距離攝影機不同距離下是否能夠辨識成功。圖 5.9 為實驗的環境以及各姿態的序列圖，依序是站→走→坐→蹲→躺。

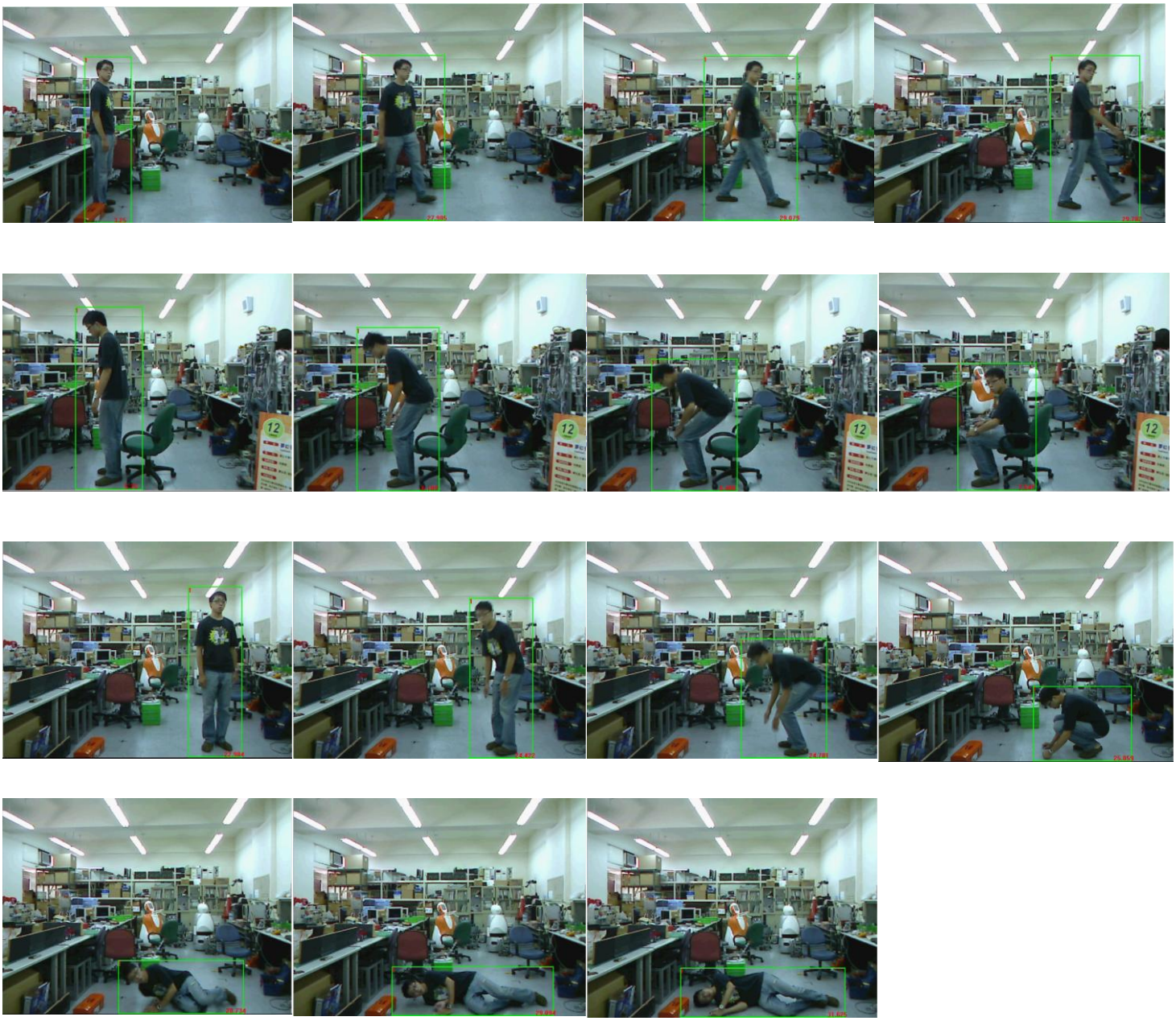


圖 5.9 為實驗的環境以及各姿態的序列圖

表 5.1 距離攝影機 2.5m 之人體姿態辨識結果

Actual Action / Estimated Action	Standing	Walking	Squatting	Sitting	Lying
Standing	47	2	0	0	0
Walking	3	48	1	0	0
Squatting	0	0	46	4	0
Sitting	0	0	3	46	0
Lying	0	0	0	0	50
Recognition rate	94%	96%	92%	92%	100%
Average Recognition rate	94.8%				

表 5.2 距離攝影機 3m 之人體姿態辨識結果

Actual Action / Estimated Action	Standing	Walking	Squatting	Sitting	Lying
Standing	49	1	0	0	0
Walking	1	48	1	0	0
Squatting	0	1	45	3	0
Sitting	0	0	4	47	0
Lying	0	0	0	0	50
Recognition rate	98%	96%	90%	94%	100%
Average Recognition rate	95.6%				

表 5.3 距離攝影機 3.5m 之人體姿態辨識結果

Actual Action / Estimated Action	Standing	Walking	Squatting	Sitting	Lying
Standing	50	3	0	0	0
Walking	0	47	0	1	0
Squatting	0	0	45	6	0
Sitting	0	0	5	43	0
Lying	0	0	0	0	50
Recognition rate	100%	94%	90%	86%	100%
Average Recognition rate	94%				

表 5.4 距離攝影機 4m 之人體姿態辨識結果

Actual Action / Estimated Action	Standing	Walking	Squatting	Sitting	Lying
Standing	45	3	0	0	0
Walking	5	47	2	0	0
Squatting	0	0	46	3	0
Sitting	0	0	2	47	0
Lying	0	0	0	0	50
Recognition rate	90%	94%	92%	94%	100%
Average Recognition rate	94%				

表 5.5 距離攝影機 4.5m 之人體姿態辨識結果

Actual Action / Estimated Action	Standing	Walking	Squatting	Sitting	Lying
Standing	48	2	0	0	0
Walking	2	48	1	0	0
Squatting	0	0	44	3	0
Sitting	0	0	5	47	0
Lying	0	0	0	0	50
Recognition rate	96%	96%	88%	94%	100%
Average Recognition rate	94.8%				

表 5.6 距離攝影機 5m 之人體姿態辨識結果

Actual Action / Estimated Action	Standing	Walking	Squatting	Sitting	Lying
Standing	44	1	0	0	0
Walking	4	49	0	0	0
Squatting	0	0	49	3	0
Sitting	2	0	1	47	0
Lying	0	0	0	0	50
Recognition rate	88%	98%	98%	94%	100%
Average Recognition rate	95.6%				

由本實驗可以得知各姿態在不同距離下有 86%到 100%的準確率，總平均準確率也可達 94.8%，其中躺的姿態在不同距離下都是 100%的準確率，因為我將躺的姿態藉由人體比例關係求出，以此實驗結果證明是可行的。

5.3 活動偵測結果

由於活動偵測需由人員偵測和姿態辨識所提供的資訊來做為判斷的依據，所以由前兩節的測試結果可以得知人員偵測和人體姿態都能夠有良好的辨識率。在活動偵測的結果方面，實際測試了一段居家生活情境的影片來呈現環境邊界的自動設置、人員所停留的時間、人員所處的位置及其姿態，驗證 FSM 內所設計得幾種來人員活動，圖 5.10 為程式右側顯示部份說明圖，辨識範圍可以設定辨識的畫面長寬，在本實驗 640x480 的畫面中設定為全畫面辨識(寬 0~640pixel、長 0~480pixel)。程式參數部份，SVM 閥值即為當 HOG 特徵透過 SVM 計算出的值大於此閥值就判定為人，在本實驗中設定為 0，而 HOG Value 就是目前畫面下的人員利用 HOG+SVM 算出來的值。人員計數可以得知目前程式有多少人次進出畫面。本實驗主要結果呈現在人員姿態和人員活動的顯示部份。

當攝影機固定後(圖 5.11)，開始對目標物作辨識(圖 5.12)，取十次成功辨識到的目標物座標取平均，得到最後固定的目標物座標以計算出的環境邊界(圖 5.13)，接著就開始對人員做偵測。人一開始走入餐廳(圖 5.14)，連續 10frames 後會得到目前人員在餐廳裡(圖 5.15)，當人的區域超過了環境邊界下方後就代表走入客廳中(圖 5.16)，連續 10frames 會得到目前人員在客廳(圖 5.17)，之後辨識出人員是坐的姿態且待在客廳，就會偵測出人員坐在客廳的活動(圖 5.18)。當攝影機移動一段距離後，一樣會辨識出目標物且計算新的環境邊界(圖 5.19)。人由新的邊界上方走入邊界下方時就會有走入客廳的活動偵測，而由圖 5.20 和圖 5.21 可以看出新的邊界是較舊的邊界靠近影像平面下方。接著人員會走入餐廳並且坐在餐廳裡(圖 5.22)。最後，人由站的姿態直接轉成躺的姿態(圖 5.23)、(圖 5.24)，系統會判定人是跌倒的並且位於餐廳中，就會偵測到在餐廳中跌倒的人員活動。

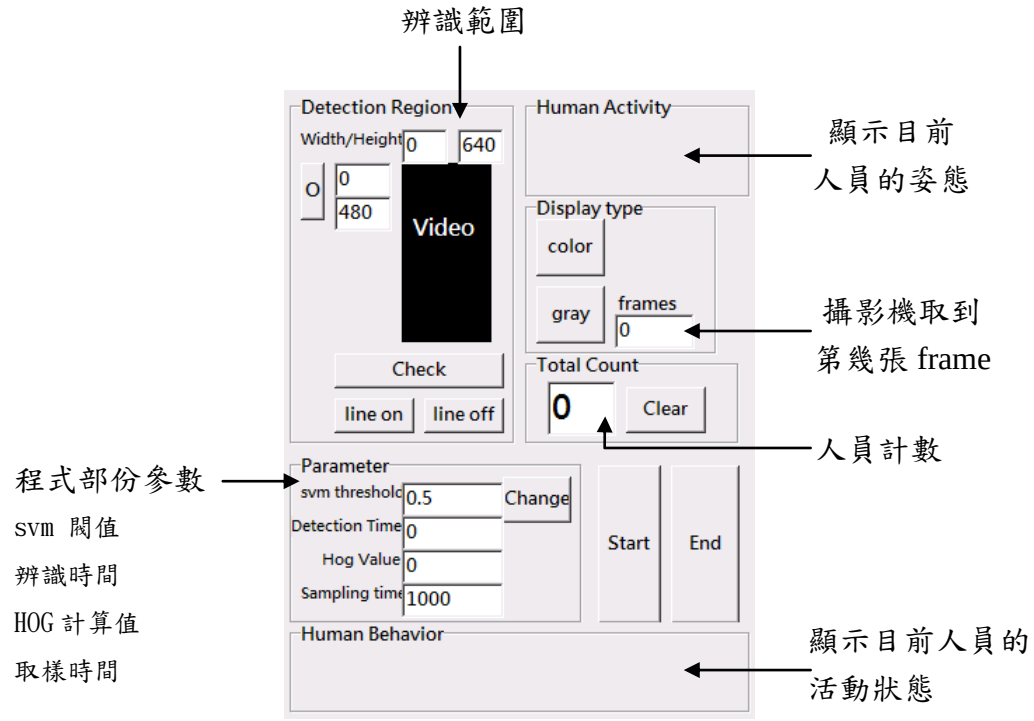


圖 5.10 程式右側顯示部份說明圖

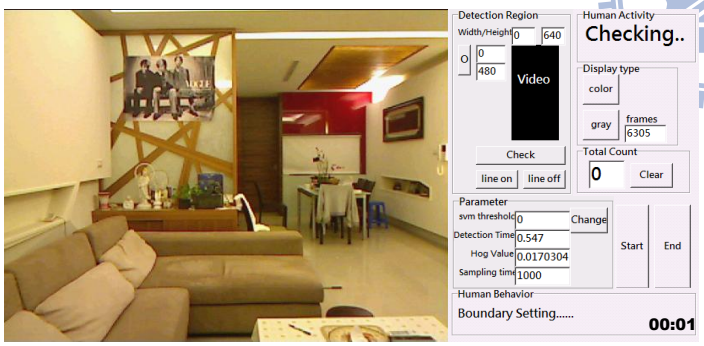


圖 5.11 攝影機固定

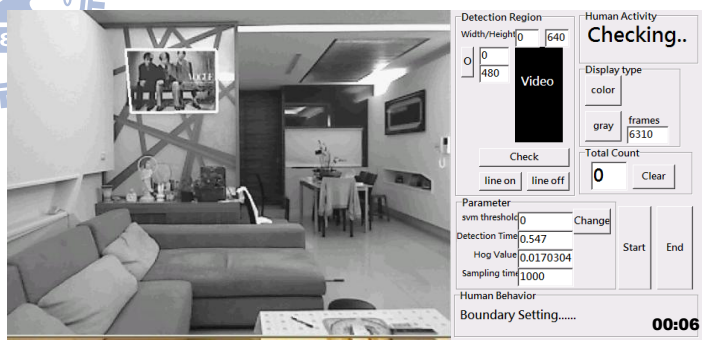


圖 5.12 辨識目標物及取其座標

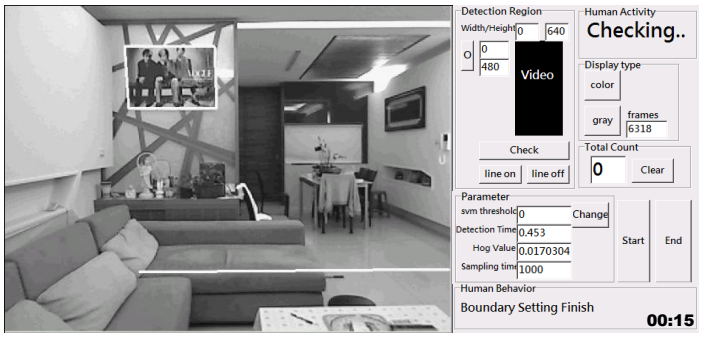


圖 5.13 由目標物座標計算出環境邊界

圖 5.14 人員走入餐廳

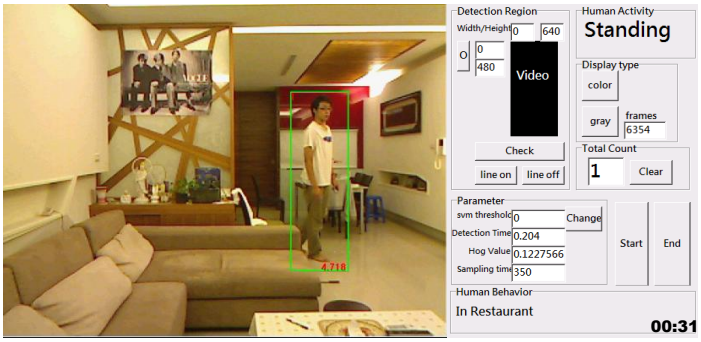


圖 5.15 人員在餐廳裡

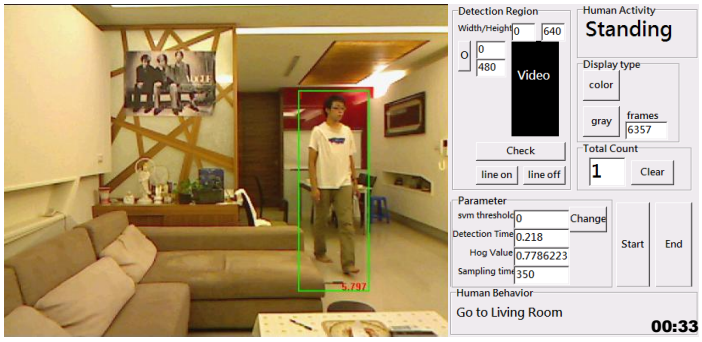


圖 5.16 人員走入客廳

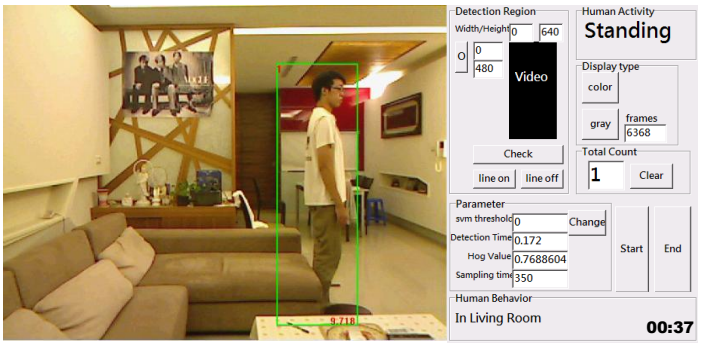


圖 5.17 人員在客廳裡

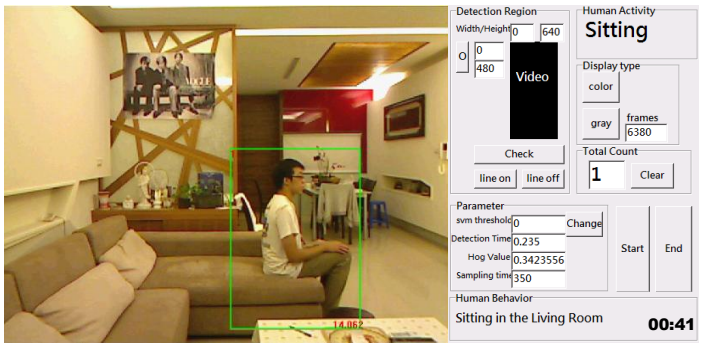


圖 5.18 人員坐在客廳裡

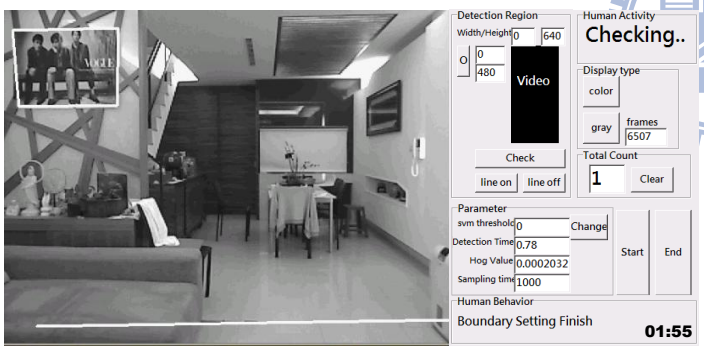


圖 5.19 攝影機移動後之目標物區域及計算出的新環境邊界

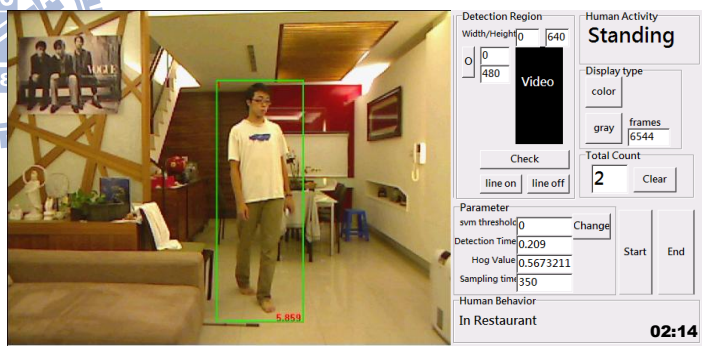


圖 5.20 人員在餐廳裡

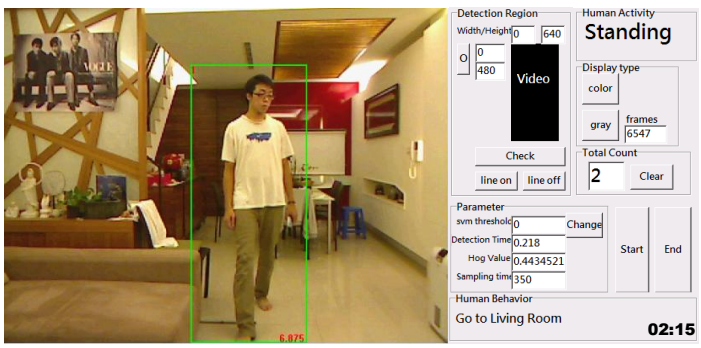


圖 5.21 人員走入客廳

圖 5.22 人員坐在餐廳裡

圖 5.23 人員在餐廳裡

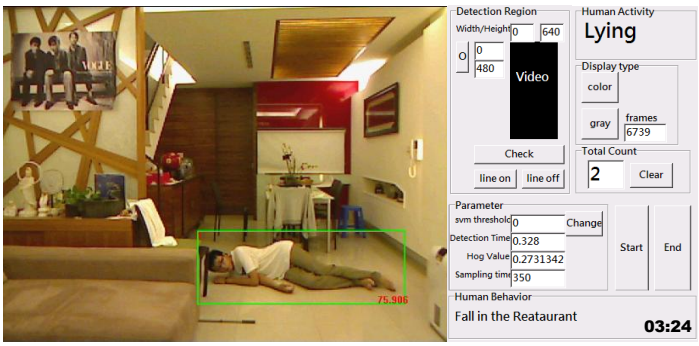


圖 5.24 人員在餐廳跌倒

第六章 結論與未來展望

6.1 結論

本論文提出了一套設計方法，利用人員所在的環境區域、人員姿態以及停留時間的組合，來達成人員活動偵測的目的。因為必須結合人員偵測和姿態辨識所提供的條件，所以這兩部份的準確率也相當重要。本論文使用 HOG 特徵萃取加 SVM 分類器去達成人員偵測，在不同的角度下平均辨識率可達 95.33%；姿態辨識使用星狀骨架的特徵加上 HMM 的訓練器來達成，可辨識站、走、坐、蹲、躺五種姿態，在不同的距離下平均辨識率可達 94.8%。

本論文之設計主要為應用於機器人上，所以開發了一套可以自動產生及調整邊界的方法，以辨識人員所在的區域。這套方法需要透過環境目標物的辨識來找出影像平面上目標物與環境邊界的相對關係。由於目標物的辨識是透過 SURF 和 Homography 找到其在影像平面上的位置且環境邊界在本論文是定義與目標物的下框線平行，故目標物的限制有：1.面積要夠大(能夠比對到特徵點)、2.必須是平面的物體(homography 需要平面的物體來做座標轉換)、3.平行於地面。

在實驗中，本論文以一張全開尺寸的海報(80cm*120cm)當做目標物，經由實驗測試，攝影機距離目標物 5 公尺以內都可以成功辨識到目標物。

在人員活動偵測的實驗中，由 5.3 節中的實驗結果可以得知，攝影機往前移動一段距離後，在目標物能夠成功被辨識到的前提下，可以產生出一條環境邊界。利用環境邊界得知人員所在區域後，結合人員姿態以及停留時間透過本論文設計的有限狀態機達成偵測人員活動的目的。

6.2 未來展望

由於本論文由人員偵測、姿態辨識及活動偵測所組成，而人員偵測及姿態辨識包含相當多的特徵萃取和分類器，且每個偵測的步驟都有獨立的特徵去完成，所以使得程式的運算量大。希望未來可以簡化演算法及程式的運算時間，使得更

符合應用於機器人於實際生活當中。

在活動偵測方面，本論文是透過辨識環境目標物來產生環境邊界，而目標物需要是平面、平行於地面且能夠被辨識得到單一物體。因為機器人在環境中的位置並不固定，也不能保證每次都能夠辨識得到我們所設定的目標物，所以未來希望能夠利用機器人在環境中的定位及其視角的關係，找出更多能夠辨識的環境物體，來增加其穩固性與實用性。

在機器人應用方面，要偵測出人員活動的目的就是希望未來能夠應用在家庭中機器人與人互動上；當機器人知道人員的活動狀態，能夠給與適當的回應，例如：人在客廳吃飯，機器人就可以幫忙端茶水給該人員。由於本論文所提出的是一套透過人員所處的環境區域為主軸，以及利用有限狀態機的方式來達成人員活動偵測，在許多人員活動更細部的動作辨識尚未能達成，所以希望能夠加入人類細部動作分析，例如：當機器人知道人員在客廳中，就可以移動到較靠近人員的地方去辨識該人員的細部動作，藉以達成更準確的人員活動的辨識。另外，在機器人與人互動的過程中，讓機器人知道該人員的身份也是相當重要的一個因素，未來可以將本實驗室開發的人臉辨識系統融入在本系統當中，期望機器人能夠針對每個不同的家庭成員有不一樣的互動。

參考文獻

[1] http://www.hriweb.org/

[2] http://world.honda.com/ASIMO/

[3] H. Fujiyoshi and A.J. Lipton, "Real-Time Human Motion Analysis by Image Skeletonization, " *Proc. IEEE Workshop Applications of Computer Vision*, Princeton, New Jersey, 1998, pp.15-21.

[4] 陳宣勝, "使用星狀骨架作人類動作自動辨識" 國立交通大學資訊工程研究所碩士論文, 2006.

[5] Y. Li and Y. Aloimonos, "The Action Synergies: Building Blocks for Understanding Human Behavior, " *International Conference on Affective Computin and Intelligent Interaction*, Amsterdam, The Netherlands, 2009, pp.1-7.

[6] H. Miyamori and S.I. Iisaku, "Video Annotation for Content-Based Retrieval Using Human Behavior Analysis and Domain Knowledge, " *IEEE International Conference Automatic Face and Gesture Recognition*, Grenoble, France, 2000, pp.320-325.

[7] T. Zhao and R. Nevatia, "Tracking Multiple Humans in Complex Situations, " *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No.9, 2004.

[8] N. Carter, D. Young and J. Ferryman, "A Combined Bayesian Markovian Approach for Behavior Reocgnition, " I*EEE International Conference on Pattern Recognition*, Hong Kong, China, 2006, pp.761-764.

[9] V. Kellokumpu, M. Pietikainen and J. Heikkila, "Human Activity Recognition Using Sequences of Postures, " *Proc. IAPR Conference on Machine Vision Applications*, Tsukuba Science City, Japan, 2005, pp.570-573.

[10] L. Wang, W. Hu and T. Tan, "Recent Developments in Human Motion Analysis, " *Pattern Recognition*, Vol. 36, No.3, 2003, pp.585-601.

[11] http://en.wikipedia.org/wiki/Finite-state_machine

[12] Z. Zhou, X. Chen, X. Han, J. Keller and Z. He, "Activity Analysis, Summarization and Visualization for Eldercare, " *IEEE Transactions on Circuits and System for Video Technology*, Vol. 18, No.11, 2008, pp.1489-1498.

[13] M. Shiomi, T. Kanda, D.F. Glas, S.Satake, H. Ishiguro and N. Hagita, “Field Trial of Networked Social Robots in a Shopping Mall, ” *IEEE/RSJ International Conference on Intelligent Robots and Systems*, St. Louis, USA, 2009, pp.2846-2853.

[14] W.T. Freeman and M. Roth, “Orientation Histogram for Hand Gesture Recognition, ” *IEEE International Workshop on Automatic Face and Gesture Recognition*, Zurich, Swizerland, 1995, pp.296-301.

[15] D.G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints, ” *International Journal of Computer Vision*, Vol. 60, No.2, 2004, pp.91-110.

[16] S. Belongie, J. Malik and J. Puzicha, “Matching Shapes, ” *IEEE international Conference on Computer Vision*, Vancouver, Canada, 2001, pp.454-461.

[17] P. Viola and M.J. Jones, “Robust Real-Time Face Detection, ” *International Journal of Computer Vision*, Vol. 52, No.2, 2004, pp.137-154.

[18] P. Viola and M.J. Jones, “Detecting Pedestrians Using Pattern of Motion and Appearance, ” *International Journal of Computer Vision*, Vol. 63, No.2, 2005, pp.153-161.

[19] D. Gavrila, “Pedestrian Fetection from a Moving Vehicle, ” *European Conference on Computer Vision*, Dubin, Ireland, 2000, pp.37-49.

[20] N. Dalal and B. Triggs, “Histogram of Oriented Gradients for Human Detection, ” *IEEE Conference on Computer Vision and Pattern Recognition*, San Diego, CA, USA, 2005, pp.886-893.

[21] N. Dalal, "Finding People in Images and Videos," Ph.D. dissertation, Institut National Polytechnique de Grenoble, Grenoble, France, 2006.

[22] http://vbie.eic.nctu.edu.tw/technical.php?index=57

[23] http://en.wikipedia.org/wiki/Support_vector_machine

[24] J.W. Davis, “Hierarchical Motion History Images for Recognizing Human Motion, ” *IEEE Workshop on Detection and Recognition of Events in Video*, Vancouver, BC, Canada, 2001, pp.39-46.

[25] I. Haritaoglu, D. Harwood and L.S. Davis, “Ghost:A human Body Part Labeling System Using Silhouettes, ” *IEEE International Conference on Pattern Recognition*, Brisbane, Australia, 1998, pp.77-82.

[26] K. Keiichi, T. Tomonaka, S. Shiotani, Y. Koketsu and M. Iehara, “Recognizing

Human Behaviors with Vision Sensors in Network Robot System, ” *IEEE International Conference on Robotics and Automation*, Orlando, Florida, USA, 2006, pp.1274-1279.

[27] L.R. Rabiner, “A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, ” *Proceedings of The IEEE*, Vol. 77, No. 2, 1989, pp.257-286.

[28] H. Bay, T. Tuytelaars and L.V. Gool, “SURF: Speeded Up Robust Features, ” *in Proc. Of the 9th European Conf. on Computer Vision*, Graz, Austria, 2006, pp.404-417.

[29] P. Viola and M.J. Jones, “Rapid Object Detection using a Boosted Cascade of Simple Features, ” *in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Kauai, Hawauu, USA, 2001, Vol. 1, pp.511-518.

[30] 賴裕宏,”基於影像處理之人體姿態辨識” 國立交通大學電機學院電機產業研發碩士班碩士論文, 2009.

[31] K. Yamauchi, B. Bhanu and H. Saito, “Recognition of Walking Humans in 3D : Initial Results, ” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Miami Beach, Florida, 2009, pp.45-52.

[32] http://pascal.inrialpes.fr/data/human/

[33] http://cbcl.mit.edu/software-datasets/PedestrianData.html

[34] http://en.wikipedia.org/wiki/Hidden_Markov_model

[35] S. Benhimane and E. Malis, “Homography-Based 2D Visual Tracking and Servoing, ” *International Journal of Robotics Research*, 2007, pp.661-676.