

國立交通大學

電機與控制工程學系

碩士論文

基於類神經網路與使用全域與運動特徵之人形



NEURAL NETWORK-BASED HUMAN
RECOGNITION SYSTEM USING GLOBAL
AND MOTION FEATURES

研究 生：徐有德

指導教授：林進燈 博士

中華民國 九十五 年 七月

基於類神經網路與使用全域與運動特徵之人形辨識系統
NEURAL NETWORK BASED HUMAN RECOGNITION
SYSTEM USING GLOBAL AND MOTION FEATURES

研 究 生：徐有德

Student : Yu-Te Hsu

指導教授：林進燈 博士

Advisor : Dr. Chin-Teng Lin

國立交通大學

電機與控制工程學系



Submitted to Department of Electrical and Control Engineering

College of Engineering and Computer Science

National Chiao Tung University

in Partial Fulfillment of the Requirements

for the Degree of Master

in

Electrical and Control Engineering

July 2006

Hsinchu, Taiwan, Republic of China

中華民國 九十五 年 七 月

基於類神經網路與使用全域與運動特徵之人形 辨識系統

學生：徐有德

指導教授：林進燈 博士

國立交通大學電機與控制工程研究所

中文摘要

此論文提出了靜態辨認與動態追蹤的方法，可用來偵測一場景中的移動物體，何者為人，何者非人。在靜態的辨別方面，不同於一般常用的基於樣板和外型的方法，我們採用了主成份分析法（Principle Component Analysis）去分析每一張圖形的全域特徵（Global Features），選取足以代表圖形大多數資訊的特徵向量，將圖形經過降維處理後，利用類神經網路的方式訓練。和之前使用特徵人臉（Eigenface）偵測臉部的方式最大的不同在於，特徵人臉的方法使用的只有臉部的圖形，而本實驗需要同時將人與非人的圖形同時選擇，在 PCA 中一起做特徵萃取的動作，才能達到效果。最後利用將結果送進一個多層級的倒傳遞類神經網路（Multilayer Back-propagation Neural Network），其輸出結果即顯示此移動物體為人或非人。利用此方法，在人的辨認度上可以達到 96%，在非人的辨認度上可以達到 91%。

在動態的部份，本實驗可以利用運動的特徵（Motion Feature）分辨出同一個場景中移動的人以及人形看板。首先我們將人的移動分為兩部份來處理，一為側向，一為直向，兩種在移動的特徵有非常大的不同。側向的移動我們利用寬度變化的特定頻率為特徵做判別，直向的移動則是利用特定部位的質心移動方式作為判定特徵。利用此方法在實際的運用上，可達到不錯的成效。

Neural Network Based Human Recognition System Using Global Features and Motion Features

Student: Yu-Te Hsu

Advisor: Dr. Chin-Teng Lin

Department of Electrical and Control Engineering
National Chiao Tung University

Abstract

In this thesis, we provide static recognition and dynamic tracking to detect the moving objects in a scene and then judge which is human and which is not. In static recognition, unlike prevalent template-based or shape-based methods, we adapt Principle Component Analysis to analyze the global features of each segmented image and select the eigenvectors which could represent most information of segmented image. After multiplied with the principle components matrix, the dimension of each segmented image is reduced. Then we send it to neural network to be trained. The most different part between Eigenface and our method is that, Eigenface method uses only human face images to be computed their principle components in PCA, but ours selects all human and non-human images to be computed at the same time. This could increase the complexity of computation, but can have better performance. Then we apply multilayer Back-propagation neural network to train the training images, the result shows this moving object is human or non-human. Using our testing samples to evaluate the accuracy rate, we have 96% in human recognition, and 91% in non-human.

In dynamic tracking, our experiment uses motion features to recognize moving person and human-shaped board in the same scene. We roughly separate human movement into two styles – lateral and frontage. The motion features within these two styles have significant difference. In lateral, we use the specific frequency generated by the width changing while human is walking. In frontage, we use gravity movement in specific part of human body to be the motion feature. It has fine performance while we apply these methods in practical situation.

致謝

本論文的完成，首先要感謝指導教授林進燈博士這兩年來的悉心指導，讓我學習到許多寶貴的知識，在學業及研究方法上也受益良多。另外也要感謝口試委員們的建議與指教，使得本論文更為完整。

其次，感謝資訊媒體實驗室的學長鶴章及剛維、世安、立偉、子貴以及建霆，以及學姊 Linda，同學家昇、峻谷以及庭瑋的相互砥礪，及學弟育弘、亞書、肇廷、德瑋以及智文在研究過程中所給我的鼓勵與協助。

感謝我的父母親對我的教育與栽培，並給予我精神及物質上的一切支援，使我能安心地致力於學業。此外也感謝弟妹對我不斷的關心與鼓勵。

謹以本論文獻給我的家人及所有關心我的師長與朋友們。



目 錄

中文摘要	ii
英文摘要	iii
致 謝	iv
目 錄	v
表目錄	vi
圖目錄	vii
第一章 引言	1
第二章 相關研究	3
第三章 靜態特徵之人與非人辨認	6
3.1 Off-line 訓練.....	6
3.1.1 移動物體的擷取與正規化	7
3.1.2 主成份的選取.....	14
3.1.3 倒傳遞類神經網路訓練.....	18
3.2 On-line 偵測	20
第四章 人形動態分析	22
4.1 側面動態判斷.....	23
4.2 正面動態分析.....	26
第五章 實驗結果	30
5.1 靜態人與非人之辨認.....	30
5.1.1 Off-line 人與非人之辨認.....	30
5.1.2 On-line 人與非人之偵測	32
5.2 動態人與非人之辨認.....	39
5.2.1 動態側面資訊.....	39
5.2.2 動態正面資訊.....	43
第六章 結論與未來展望	46
參考文獻	48

表 目 錄

表 2.1：偵測人的方法	3
表 5.1：訓練資料的正確率	31
表 5.2：測試資料的正確率	31



圖目錄

圖. 3-1 : Off-line 訓練的流程圖	7
圖. 3-2 : 移動物體的擷取與正規化	8
圖. 3-3 : 顏色轉換	9
圖. 3-4 : 使用均值平滑濾波器的結果	10
圖. 3-5 : 圖形經過膨脹後的結果	12
圖. 3-6 : 連通成份標記示意圖	13
圖. 3-7 : 像素重新排列示意圖	15
圖. 3-8 : 特徵向量所包含的圖形資訊	17
圖. 3-9 : 人的訓練影像範例	19
圖. 3-10 : 非人的訓練影像範例	19
圖. 3-11 : On-line 偵測的流程圖	21
圖. 4-1 : 動態判斷流程圖	23
圖. 4-2 : 側面行走的連續影像	24
圖. 4-3 : 寬度的時域變化情形	25
圖. 4-4 : 圖 4-3 的FFT轉換的能量頻譜	25
圖. 4-5 : 側面行走之另一範例	26
圖. 4-6 : 人正面行走之連續影像	27
圖. 4-7 : 紅色矩形處的腳部移動	27
圖. 4-8 : 質心的位置移動	28
圖. 4-9 : 正面行走另一範例	29
圖. 5-1 : 觀測場景	32
圖. 5-2 : 程式執行介面	34
圖. 5-3 : 人與車的判別情形	35
圖. 5-4 : 人以不同方向行走	36
圖. 5-5 : 人背著背包與提著物品	36
圖. 5-6 : 多人與非人的移動物體	37
圖. 5-7 : 誤判 (一) 移動物體過於靠近	38
圖. 5-8 : 誤判 (二) 遮蔽情況判為非人	38
圖. 5-9 : 誤判 (三) 遮蔽情況判為人	39
圖. 5-10 : 動態前景、背景與前處理結果	40
圖. 5-11 : 人側向移動的處理結果	41
圖. 5-12 : 人形看板的處理結果	42
圖. 5-13 : 動態正面前景、背景與前處理結果	43
圖. 5-14 : 人正面處理結果	44
圖. 5-15 : 人形看板正面處理結果	45

第一章

引言

移動物體，尤其是人或者是交通工具的偵測，已被大量的使用在監視系統中。近幾年來，在對於人的偵測方面，許多利用監督式的學習分類法(Supervised Classification)也已經被提出。大致上來說，這些方法都可被分成兩個步驟：(一)移動物體的分離，與(二)移動物體的辨認。當一個移動物體進入所觀測的場景後，系統首先會偵測到此移動物體在畫格中位置，再將其從畫格中分離處理。最後，辨認此物體為人或非人。

本系統主要來說，分別使用靜態辨認與動態追蹤的方式，對於移動物體作一判斷。在靜態辨認方面，我們對於每個單一畫格都做處理，而不使用動態的特徵。在動態追蹤方面，我們對於每個移動物體，累積一定的資訊之後，才加以做出判斷。一般說來，人的判斷有時會比判斷其它的物體要難，原因是因為人為一個典型的非剛性物體，尤其是當我們同時考慮到人的大小、衣著顏色，形體等，其多變性造成許多辨別上的困難。

在本實驗系統的靜態辨認部份，可分為 Off-line 的訓練以及 On-line 的偵測。在 Off-line 的訓練部份，我們可將步驟劃分為三個：(一)移動物體的擷取及正規化，(二)主成份的選取，以及(三)倒傳遞類神經網路的部份。首先我們將場景中的移動物體與畫面分離，經過一些影像的前處理後，將其框架出來，然後將表現移動物體的影像正規化為 100×50 的大小，也就是說，每個移動物體包含了 5000 個像素。之後將像素重新排列，交由 PCA 去作主成份的選取，目的是能夠選取出能夠顯示出每張移動物體影像大部分資訊的特徵向量。等到每一張圖經過主成份分析法之後，能夠用較少的維度表現，我們將其送入倒傳遞類神經網路做訓練，目的是得到一個權重矩陣。

在 On-line 辨認的部份，取出移動物體的方式和影像前處理的部份，與

Off-line 訓練相同。所不同的是，移動物體的影像取出和正規化之後，只需與主成份矩陣和權重矩陣相乘，依照輸出值來決定此移動物體為人或非人。此實驗系統的好處在於：(一) 不需利用人形的模型樣板，(二) 對於人穿著的樣式、顏色、人行走的方向和姿勢，不需要嚴格限定，以及 (三) 本系統在靜態辨認部份對於人與非人有很高的辨識率，在人的部份，可以達到 96%，非人則達到 91%。

在動態追蹤的部份，我們首先分析人在以橫向行走與直行的特徵，利用此特徵做分析。實驗的進行方式是用一人形的看板（在靜態時會被辨認成人），靜止或是緩緩移動，旁邊有一實際的真人在觀測場景中移動著。經過一段時間的動態資訊累積之後，我們可以分辨出何者為人，何者為人形看板。

在章節的安排上，第二章為對於人與非人的相關研究。第三章敘述靜態辨認所使用的方法與技術。第四章則為動態追蹤所使用的方法。第五章則是實驗結果。第六章的部份則對論文作一結論，並對未來可以加以增進的部份作一說明。



第二章

相關研究

人的辨認，從以前到現在，一直是一個很熱門的議題，許多的方法也被提出來並行之已久。大抵來說，人的辨識流程可分為兩個部份：(一) 移動物體的分離，以及 (二) 物體的辨認。為了達到此二目標的使用方法，大致上可分成表 2-1 所示的幾種。

表 2.1：偵測人的方法

目標	方法		
移動物體的分離	Optical Flow	Stereo Vision-based	Temporal Differencing
物體的辨認	Shape-based	Motion-based	Multi-cue-based

移動物體的分離目的在找出連續畫格中的移動物體。光流(Optical Flow)[1] 可以被用來偵測單一移動物體的移動情形。在[2]和[3]中，光流的技術被使用在交通工具的偵測上。但使用此技術偵測移動物體的缺點就是其計算的複雜度相當高，若是想要將系統運用在即時(real time)的行人偵測上，運算速度會是一個相當重要的考量。此外，光流的技術在對於非剛性物體的移動偵測並不適用。根據這幾種原因，我們並不使用此方法來偵測人。

參考資料[4]使用立體視覺(Stereo Vision-based)的技術來作為人的偵測之用。Zhao et. al. [5]提出了以立體的方式為基礎之分割物體的方法，並利用類神經網路對物體做辨認。這種方式有幾種好處：(一) 它可以偵測畫面中的移動以及靜止的物體，(二) 並不需要複雜及大量的運算，以及 (三) 對於光度的變化不敏感。雖然此方法有這些好處，但在使用時，需要不只一台相機或攝影機，而且只能使用在較近距離的偵測上。尤其困難的是，在影像的處理上，要找出左

右兩邊互相對應的影像。

Smith et. al. [6]利用固定背景作為移動物體的分離。這種方法最大的問題在於光度的變化以及攝影機的晃動。[7]以及[8]提出了一些方法對於畫格中的一些小變化做一些技術上的補償，但對於大範圍的變動或是突然間的光度變化，仍舊是此方法無法克服的部份。

將移動物體從畫格中分離出來後，下一部的動作就是對其做辨認。因為辨認的技術方法不同，所選取的特徵也有所不同。以形狀為基礎(Shape-based)的技術利用一些外型上的特徵作為物體的辨認方式。此方法的好處在於，不論物體為靜止或是移動中，都能夠被辨認出來。但是用在人的偵測上，因為人在移動上的外形改變，會發生判斷上的困難。Rohr [9]利用手繪的人的模型偵測畫面中的人，Gavrila et. al. [10]則是利用距離的轉換用在即時的行人偵測，此方法對於物件形體的小幅變動不受影響。在參考資料[11]中，則是利用人在行走時的週期現象做判斷的依據。在[12]及[13]中，利用了SVM(Support Vector Machine, 支援向量機器)以及小波的特徵點選取的搭配，對人形能有不錯的判斷效果，然而，缺點就是在處理上相當耗費時間。Toth et. al. [14]使用傅立葉描述子來偵測人形，可以不受到大小和位置的影響。缺點是在描述人的外形時，若有小地方的缺陷沒有被填補到，則描述子的位置會有大幅度的改變，造成判斷上的錯誤情況。

在利用運動特徵(Motion-Based)的行人偵測上，Heisele et. al. [15]利用時間延遲的類神經網路去分辨人小腿部份的運動情況。在[16]及[17]中，提供了行走時具有節奏這個特性來分辨移動物體。不過此方法具有幾種先天上的缺點：(一) 人的雙腳為可視，(二) 此方法對於畫面中靜止的人形無法辨識，(三) 當人以較為複雜的方式移動時，如跳躍等，則無法偵測，以及(四) 連續畫格必須到達一定的數量以上，才能累積足夠的資訊做判斷。

有些系統則會使用不只一種的特徵來作判斷。在[4]中，形狀、膚色以及臉部偵測這些特徵被用來偵測畫面中的人物。然而，臉部以及膚色的資訊對於人的

偵測來說是不穩定且不可靠的。臉部偵測只有在行人是正面面對攝影機時才能發生作用，而膚色資訊只有當行人靠近攝影機時，才能發揮一定效果，如果距離較遠以致於膚色的資訊很少時，此特徵並無法作為判斷的依據。

本論文對於人形的偵測使用兩種不同的方法。當靜態時，我們採用單張影像辨認的方式，而動態的狀況下則分別處理人形以側面或正面行走時的不同特徵作為判斷的依據。在移動物體的分離上，我們使用時間差異法(Temporal Differencing Method)再搭配一些影像處理的技術。在物體的辨認方面，靜態時我們利用主成份分析法(Principle Component Analysis, PCA)做人與非人的特徵選取，之後利用倒傳遞類神經網路(Back-Propagation Neural Network)當作分類器，對移動物體做辨認的工作。動態方面，我們在側向移動時使用寬度變化的頻率資訊作為判斷依據，而正向移動時則以特定區域的質心變化情形作為特徵判斷。其技術方法在第三章與第四章中有詳細的介紹。



第三章

靜態特徵之人與非人辨認

在本章裡，我們說明對於靜態的人形辨認所採用的方法。我們將內容分為 Off-line 的訓練跟 On-line 的辨認。Off-line 訓練的目標是選取一些具有代表性的特徵向量，即圖形主成份 (Principle Components) 的構成，並在經過倒傳遞網路訓練之後，利用其所產生的權重矩陣，可在 On-line 的過程中經過一些計算，判定移動物體是人或非人。

3.1 Off-line 訓練

圖 3-1 利用流程圖簡短說明 Off-line 訓練的流程圖。首先，我們將影像中所有移動物體中個別擷取出來，並將其正規化為 100×50 的大小，選取 2000 張人及 800 張非人的圖形作為訓練的樣本。之後利用主成份分析(Principle Components Analysis)的方法，得到這些訓練樣本的全域特徵(Global features)。選取具有代表性的特徵向量，通常稱其為主成份，將訓練圖形與此主成份矩陣相乘，於是，每個圖形可用一個特徵向量(Feature Vector)表示。將所有訓練圖形個別表示為其特徵向量之後，利用倒傳遞網路做訓練。當均方根誤差(Mean Square Error)收斂後，會產生一個權重矩陣，將此矩陣做儲存。

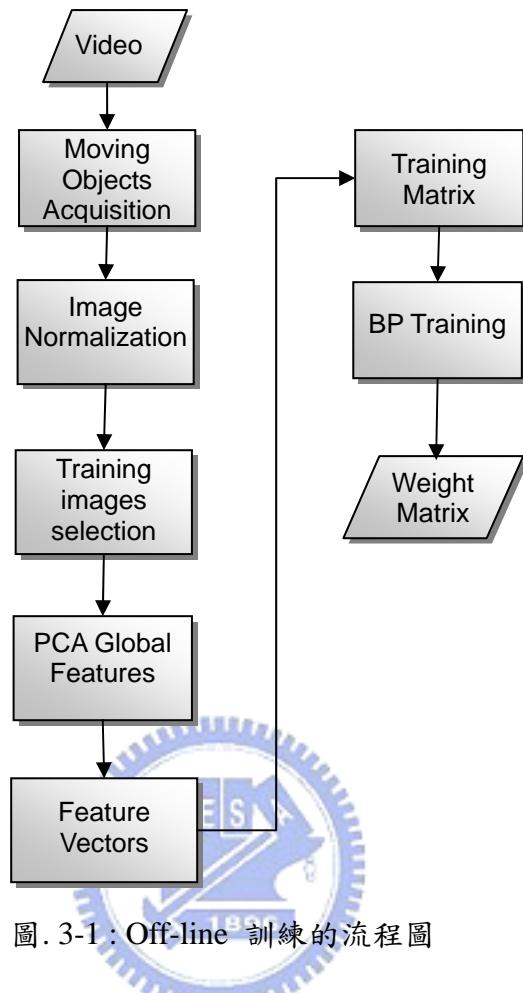


圖. 3-1 : Off-line 訓練的流程圖

本流程大約可分為三方面，(1) 移動物體的擷取及正規化，(2) 主成份的選取，以及 (3) 倒傳遞類神經網路的訓練。在 3.1.1 中，我們對於移動物體的擷取與正規化作說明，3.1.2 將介紹主成份的選取方式，3.1.3 則是說明本實驗利用倒傳遞網路的方法。

3.1.1 移動物體的擷取與正規化

圖 3.2 用流程圖對於移動物體的擷取與正規化做了一簡短的表示。不論是在 Off-line 訓練或是 On-line 偵測時，這個步驟都相同。首先將彩色的影像轉成灰階，再利用時間差異法(Temporal Differencing Method)將移動物體擷取出來，之後用均值平滑濾波器(Mean Filter)消除一些小雜訊。為了完整地取出移動的物體，首先將與背景相減後的影像做二值化(Binarize)的處理，做兩次膨脹(Dilation)

之後，利用連通成份(Connected Components)原理，避免單一移動物體被分割的情況。用一矩形的方框將移動物件標示起來之後，為了配合系統需要及考量到處理時間上的問題，計算矩形的面積大小，必須大於一個門檻值，才會做接下來的正規化處理。若是小於此門檻值，則將此移動物體忽略不做處理。正規化之後的圖形，再送至下一步驟的 PCA 做分析。

這整個流程所使用的技術方法，我們在此作一簡短的說明。

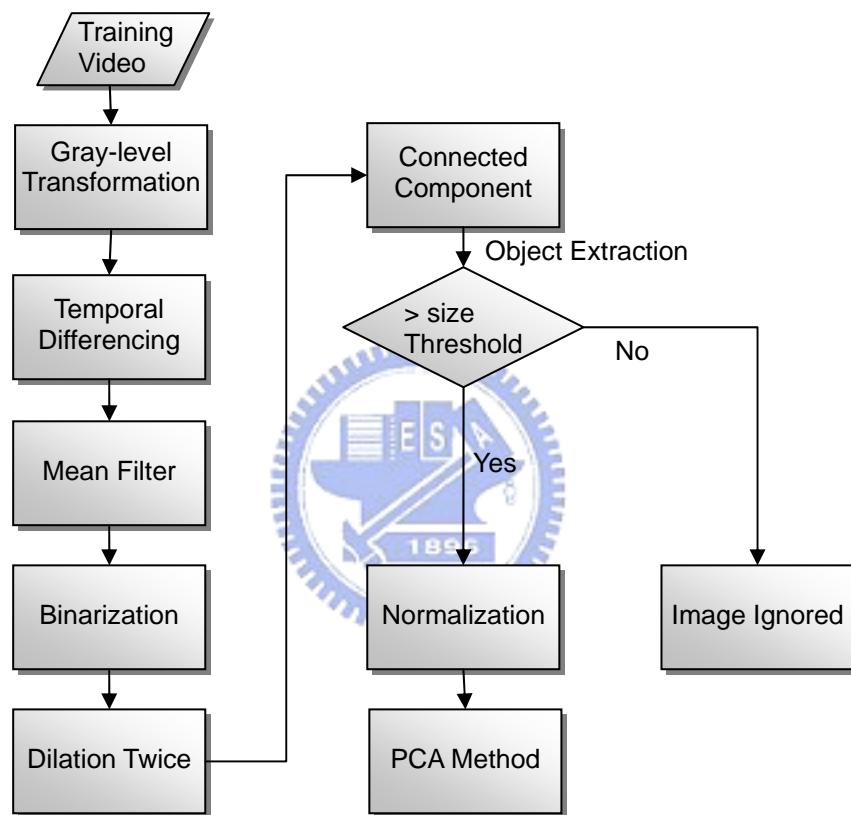


圖. 3-2：移動物體的擷取與正規化

我們將所拍攝的影像以每秒 30 張畫格的方式讀取進來，並將其從彩色轉成灰階做處理。其原因有二：

一、提高人行辨識的正確率

在電腦視覺中，每個彩色的像素包含了 3 個資料。因為本系統為人和非人的辨識，但是因為人衣著的多樣化，其所包含的顏色相當多，轉成灰階之後，顏色資訊的重要性可被降低。在圖 3-3 中，明顯的我們

可以發現左邊的兩張圖在顏色方面明顯的不同，但轉成灰階值後，都如同最右邊灰階的顏色。



圖. 3-3：顏色轉換

二、灰階包含了較少的資料量

如之前所述，每個彩色的像素包含了 3 個資料，灰階只包含了一個，對於計算量的降低有一定的幫助。

電腦對於影像顏色，基本上是以 RGB 的方式讀取進來的。我們將色彩空間從 RGB 轉成 YUV，因為 YUV 在數位色彩中是最常被使用的格式。Y 包含了明暗的資訊，U 和 V 則是色彩的資訊。從 RGB 轉換成 YUV 的過程中，我們使用了如(3.1)所示的轉換矩陣。

$$\begin{vmatrix} Y \\ U \\ V \end{vmatrix} = \begin{vmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & -0.100 \end{vmatrix} \begin{matrix} R \\ G \\ B \end{matrix} \dots \quad (3.1)$$

因為 Y 只包含了明暗的特徵，所以可被視為灰階的一種，所以我們只選取 Y 作為之後所使用的資訊。

為了要得到畫面中所有移動中的物體，我們採用了時間差異法（Temporal Differencing Method）在場景中切割出移動的物件。以此做法為基本的話，我們必須先建立一個靜止的背景，之後的畫格與此背景做相減，其灰階值的絕對值差異若大於一個臨界值，就是畫面中可能的移動物體，在此我們使用了如(3.2)的方法。根據此方法，當新畫格與背景的差異度大於一門檻值之後，我們保留新畫格

的灰階值，若是小於此門檻值，則將其變化度視為零。

$$H(x) = \begin{cases} 0, & \text{if } T(x) < \text{threshold} \\ T(x), & \text{otherwise} \end{cases} \quad \dots \dots \dots \quad (3.2)$$

在此， $T(x)$ 為新畫格中像素的灰階值，而 $H(x)$ 為經過門檻值篩選後的結果。

但因為畫格中可能存在著一些小雜訊，例如樹枝及樹葉輕微的搖晃，經過時間差異法的處理之後，會在圖上產生細碎的小點，這時，雜訊的消去就是一項很重要的工作了。我們使用均值平滑濾波器(mean filter)來作雜訊消去的工作。此方法的介紹如下所示：

1. 假設有個 3×3 的影像，其中的各個像素位置用 1, 2, 3……標示著，我們要決定的是像素 5 的值。

1	2	3
4	5	6
7	8	9



2. 像素 5 的值是周圍八個像素值的平均。

經過此步驟處理後，許多由搖晃的樹葉產生的雜訊可被消除掉，圖 3.4 為結果。

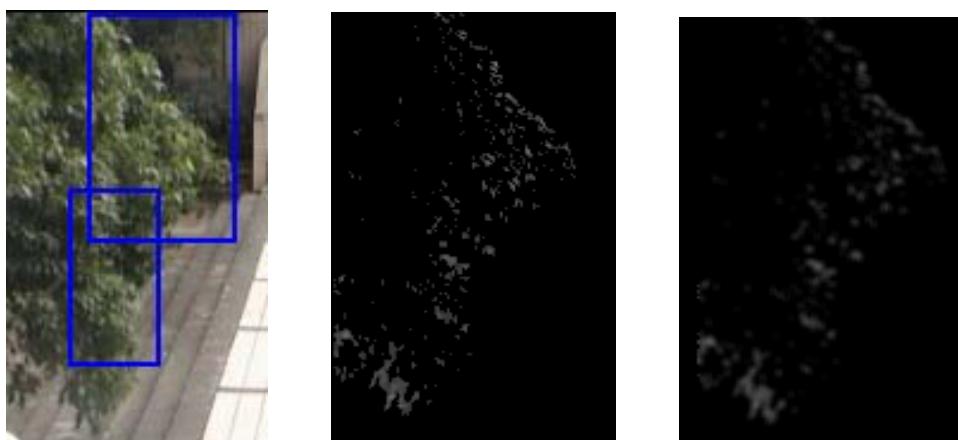


圖. 3-4：使用均值平滑濾波器的結果

最左方的圖中有兩個藍色的矩形框住之處，表示僅使用之前的方法無法消除這些劇烈的搖晃。中間的圖形是僅使用時間差異法所得到的結果，最右方的圖形是經過均值平滑濾波器處理後的圖形。可以很明顯的發現經過處理後，許多雜訊點可被消去，但同樣的，畫面會變得較為模糊。

經過均值平滑濾波器的處理後，為了從畫面中挖出完整的人形，我們必須使用一些處理方式。首先將經過濾波器處理的圖形二值化，凡是經過時間差異法之後有值的點，我們將其表為 255 (白色)，其餘的仍舊為 0 (黑色)。

使用時間差異法除了會產生一些小雜訊之外，主要的缺點還有，許多與背景灰階值相近的顏色再經過相減後，會容易被分成碎片的情況，怎樣將這些片段結合起來，以便在取出人形上夠較為完整，變得相對重要。在此我們使用膨脹 (Dilation)的方式。膨脹的技術方法簡述如下：

1. 假設有個 3×3 的二值化影像，其中的各個像素位置用 1, 2, 3……標示著，我們要決定的是像素 5 的值，其中，周圍的像素值不是 1 (白色)，就是 0 (黑色)。

1	2	3
4	5	6
7	8	9



2. 若是像素 5 周圍的八個像素有一個像素的值為 1，則像素 5 的值也為 1，否則則為 0。

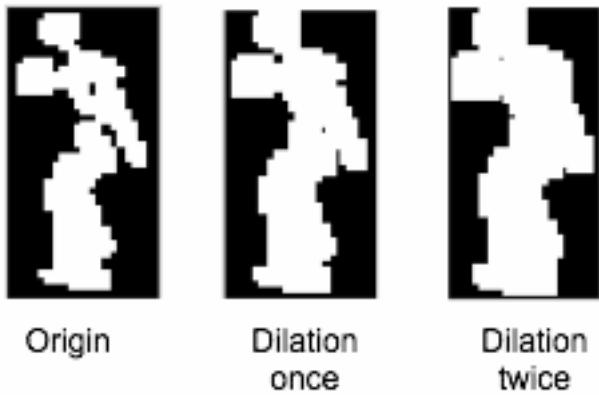


圖. 3-5：圖形經過膨脹後的結果

圖 3-5 為一張二值化的影像分別經過一次及兩次膨脹後的圖形。經過兩次膨脹後的圖形確實在軀幹內許多空洞的部份都能有所補強。為了取得較為精確的結果，我們將畫面經過兩次的膨脹結果作為接下來連通成份處理的使用。

以上所敘述的方法都是在處理整張畫格，現在我們需要做的是將畫格裡可能的移動物體標上標記，以便之後的正規化程序。我們採取了連通標記（connected components）的技術來完成此項動作。它可以將屬於同一群的物件給予同一個標記號碼，之後我們必須利用以這個標記為基準，對移動物體從畫格當中擷取出來。

基本上，這個連通標記的方法，我們是以參考資料[19]為基礎發展的。此參考資料對於連通成份標記的演算法提供了一個有效率的方法。不過跟參考資料不同的是，我們使用 8 連通性的處理，而不是 4 連通性的方式。基本的觀念如下所示：

1. 假設有個 3×3 的影像，中央的像素 C 的值為 1 (白色)，同時也被標記為 1。若是它周圍八個位置有任何一個的值也為 1，則也會被標記成 1。

	C	

3. 圖 3-6 為一範例。假設有一 8×8 的圖像，黑點的地方表示其像素值為 1。

經過前述的方法做連通成份的處理後，標記的結果會如右方所顯示。

•			•	•			
	•		•		•		
	•	•	•				
	•		•				
			•		•		
	•			•			
					•	•	

1			1	1			
	1		1		1		
	1	1	1				
	1		1				
					2	2	
					3	2	
						2	2

圖. 3-6：連通成份標記示意圖

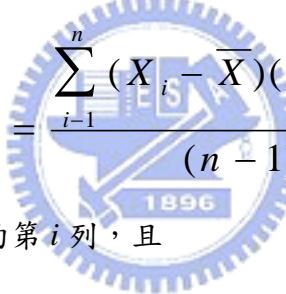
我們挑出每個標記最左方、最右方、最上方及最下方的點，然後利用最左方與最上方組合成一個新座標，最右方與最下方組成另外一個，以這兩點為基礎，畫出一個矩形，如此一來，此矩形就會包含了所有連通成份標記相同的區域，也就是畫面中屬於同一部份的移動物體。有的時候一些小地方的移動也會被標記出來，例如無法消去的樹葉移動，空中飛舞的碎屑等等，所以我們設計一個門檻值，矩形的面積必須要大於此一門檻值，才會經過計算，若是小於此值的話，則不做處理。此種做法一方面可降低電腦的運算量，另一方面也提供了另一層的雜訊消除。

當移動物體繼續移動時，在不同的畫格上會有不同的大小，所以我們必須將矩形所框架出來的移動物體大小做一個正規化，才不會受到移動物體位置的影響。在此考慮到計算量的大小與全域特徵(Global Feature) 的表示，我們選取 100×50 為正規化之後的圖形大小。在長度方面為 100，寬度方面為 50。也就是說，每一個選取出來的圖像會含有 5000 個像素。我們選取 2000 張人的影像以及 800 張非人的影像進入接下來的主成份分析(Principal Components Analysis, PCA)，將每張影像的全域特徵都使用其方法來選取主成份。

3.1.2 主成份的選取

主成份分析法用於資料的分析已經有一段時間了。主要是用來將多個變量經過線性變換以選出較少個數重要變量的方法。其中每個變量在數學上的意義 j 為與其他變量正交，也就是不具有交互相關性。於是，我們可以用此方法將原本包含大量變數的資料用較少的主成份所表示，不但可以達到降維的作用，同時也能對於一些雜訊（比較不重要的成份）有所過濾。

使用主成份分析的方法，事實上就是在利用奇異值分解法(Singular Value Decomposition) 解出資料的相關變異矩陣(Covariance Matrix) 的特徵值與特徵向量。參考資料[20]介紹基本的數學概念。假設有一個一維的矩陣 X ，計算它的相關變異就等於計算它的自身變異量。在此簡短說明[20]所提及的方式。

$$\text{cov}(X) = \text{var}(X) = \frac{\sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})}{(n-1)} \quad \dots \quad (3.3)$$
在此， X_i 表示矩陣 X 的第 i 列，且

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \quad \dots \quad (3.4)$$

於是，此矩陣的自身的相關變異矩陣為

$$C^{n \times n} = (c_{i,j}, c_{i,j} = \text{cov}(Dim_i, Dim_j)) \quad \dots \quad (3.5)$$

在此， $C^{n \times n}$ 表示這是一個具有 n 行跟 n 列的矩陣。 Dim_x 表示第 x 維。

一旦我們得到了一個方形的相關變異矩陣後，即可計算其特徵值與特徵向量。這些特徵向量彼此互相垂直，但更重要的，它們同時說明了資料量的一些資訊。我們將這些特徵向量與其所對應的特徵值按照大小順序排列，較大的特徵值所對應的特徵向量包含了資料集中較為重要的資訊。於是，我們可以捨棄掉一些包含比較少資訊，也就是較不重要的成份，並且不會失去原資料集的大部分

資訊。假設一個資料集最初為 n 維，於是可計算出 n 個特徵值與特徵向量。若是我們只選取 p 個特徵向量做為主成份，則原本 n 維的資料，經過處理後，可用 p 維來表示。

以下簡述本實驗所用的方法：

1. 取得 M 張訓練影像，分別標示為 I_1, I_2, \dots, I_M ，其中，每張影像的大小皆為 $N \times N$ 。
2. 將每一張影像 I_i 的像素重新排列為 $N^2 \times 1$ 的向量 Γ_i ，如圖 3-7 所示，因為有 M 張影像，則會產生一個 $N^2 \times M$ 的矩陣。

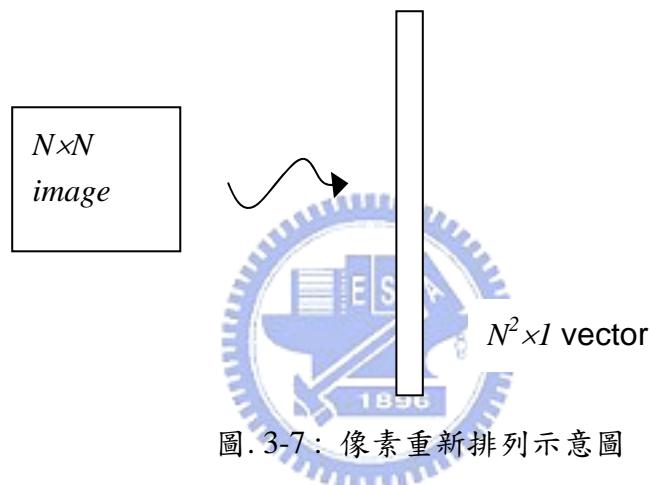


圖 3-7：像素重新排列示意圖

3. 計算平均向量 Ψ

$$\Psi = \frac{1}{M} \sum_{i=1}^M \Gamma_i$$

4. 將每個 Γ_i 與平均向量 Ψ 做相減，得到的結果為 Φ_i

$$\Phi_i = \Gamma_i - \Psi$$

5. 計算其相關變異矩陣 C

$$C = \frac{1}{M} \sum_{n=1}^M \Phi_n \Phi_n^T = AA^T \quad (N^2 \times N^2 \text{ matrix})$$

where $A = [\Phi_1 \Phi_2 \dots \Phi_M]$ ($N^2 \times M$ matrix)

6. 計算 AA^T 的特徵向量 u_i ，基本上會產生 N^2 個特徵向量，將其從對應的

特徵值由大到小排列。

7. 保留對應前 K 個最大的特徵值的特徵向量。
8. 將訓練影像以這些特徵向量為基底加以表示：

每張影像經過與平均向量相減之後的 Φ_i ，可以此 K 個特徵向量作一線性組合來表示。

$$\Phi_i = \sum_{j=1}^K w_j u_j, \quad (w_j = \Phi_i^T u_j)$$

經過上述的步驟後，每一張影像可被表示成一個 $1 \times K$ 的向量。因為我們選取了 M 張影像，所以將其排列一個 $M \times K$ 的矩陣。在其後的偵測及辨認部分，可使用歐基里德距離(Euclidean Distance)或是 Mahalanobis 距離。如果給予一個不知名的影像 Γ ，第一步我們計算 $\Phi = \Gamma - \Psi$ ，然後計算 $\hat{\Phi} = \sum_{i=1}^K w_i u_i$ ($w_i = u_i^T \Phi$)，最後則是計算 $e_d = \|\Phi - \hat{\Phi}\|$ 。若是 $e_d < T_d$ ，則 Γ 為所需影像。

然而，此方法具有一些嚴重的缺點，例如背景的複雜度、光線的條件以及臉部的方向性，這會導致許多偵測錯誤的情況發生。為了增加其辨別的正確率，許多額外的特徵必須加入作為判斷的依據，例如皮膚顏色或是人形樣板等等。但是在辨別人形上，這些特徵都無法使用。因為人穿著的多樣性加上我們攝影機架設的位置離人有一段距離，所以人體身上所能觀測到的皮膚顏色範圍可能不大。再者，人為一個非剛性的物體，其正面、側面或是斜面都有不一樣的形狀，利用樣板判斷容易會有誤判的情況產生。

參考資料[24]把訓練的圖像分成人跟車，分別做出不同的子空間圖形，在用這些圖形作為訓練依據。我們觀察其實驗結果可發現其所偵測的環境大致單純，可能出現的移動物體也只有人跟車而已，而且其攝影機所架設的位置離所觀測處有一段距離，所以基本上人跟車子在畫面中的大小就差很多，也就是說，如果單純的利用大小的判斷，其實就應該會得到相去不遠的結果。然而，在我們所觀測的環境中，首先，我們離被觀測物近得多，所以人、狗以及移動樹葉的大小其實

相去不遠，而且人在靠近或是遠離攝影機移動的大小變動的很劇烈。再者，畫面中可能含有的移動物體不單單只有人，其他如狗、車輛、飛舞的碎屑或是搖晃的樹枝等等，都可能充斥在畫面當中，如果對每一個移動的物體都要做出一個特徵子空間的話，並不是一個聰明的做法。於是，我們利用 PCA，但採取不同的方式選取其主成份。

首先，我們將前一節所提到的 2000 張人和 800 張非人的影像按照前述的方式重新排列。按照步驟計算其平均向量，之後求出變異矩陣的特爭執與特徵向量。

得到特徵向量之後，我們必須選出足以代表影像大部分資訊的主成份。理論上來說，越多的特徵向量包含了越多資訊，判斷會越準確，但同樣的，會造成計算上的負擔。於是，主成份個數的選取變得必須仔細挑選。

圖 3-8 表示了特徵向量所包含的圖形資訊，由圖上可以看出，60 個特徵向量可以表示圖形 80% 的資訊，89 個可以表示 85%，111 可表示 87.5%，而 143 個可以表現 90% 的資訊。因為通常說來，人和非人的圖像相當不同，而且，經過我們試驗的結果（表 5-1、表 5-2），89 個主成份和 111 個主成份所得出的效果幾乎相同，而 60 個主成份就相對較差，所以我們選取 89 個特徵向量作為我們表示的主成份，同樣也可以降低計算上的負擔。

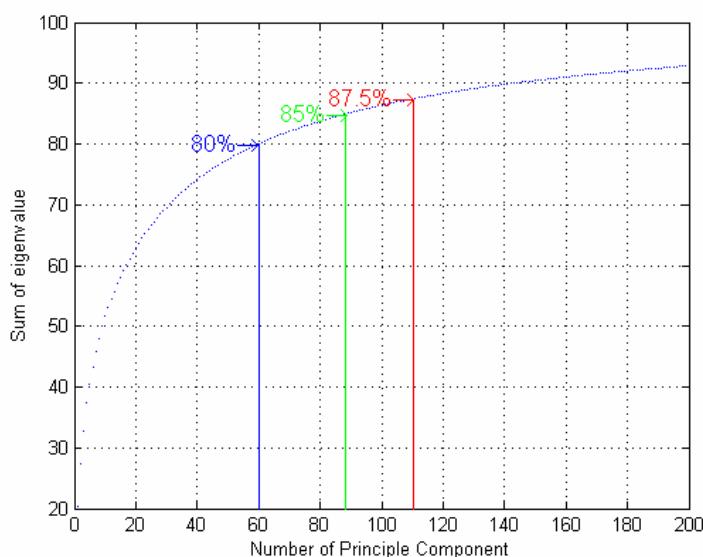


圖. 3-8：特徵向量所包含的圖形資訊

當主成份確定之後，每一張經過正規化的影像，都可以表示為這些主成份的線性組合。

如同之前所提過的，每張影像可表示為的 1×89 向量，既然我們有 2800 張的影像，我們於是可將其表示為一個 2800×89 的矩陣。接下來為了增加準確度，我們將這個矩陣利用倒傳遞網路做訓練。將原先是人的影像標示為 $[0\ 1]$ ，非人的影像標示為 $[1\ 0]$ 。當此工作完成之後，即可將這些資料送進下一步驟做訓練。

3.1.3 倒傳遞類神經網路訓練

為了增加準確率，我們使用了類神經網路來代替歐基里德距離的判斷。這種方法可以對輸入和輸出之間產生較為精確的關係，對於輸出判定的正確率也會有所提高。在此，我們採用了倒傳遞類神經網路(Back-propagation Neural Network, BP)作為分類器。這種分類器可決定高維度的判定邊界(decision boundary)，以及，相較於其他類神經分類器而言，用較少的訓練資料就能進行學習，而且準確率也不會太差。

本論文所使用的倒傳遞類神經網路具有三層的結構，一為輸入層，一為隱藏層，另一層為輸出層。輸入層的數目和主成份選取的數目相同，為 89 個，輸出層為 2 個（分別為 $[1\ 0]$ 以及 $[0\ 1]$ ），至於隱藏層數目的選取，我們依據的是參考資料[29]所建議的方式。其計算方式如下：

$$(1) \text{Hidden Nodes} = \frac{\text{Input Nodes} + \text{Output Nodes}}{2}$$

$$(2) \text{Hidden Nodes} = \sqrt{\text{Input Nodes} \times \text{Output Nodes}}$$

這兩種建議方式適用於不同的情況。若是問題雜訊高，則隱藏層單元數目宜少。若是問題複雜性高，則隱藏層數目宜多。在我們的資料中，大抵來說只有「人」以及「非人」兩種情況，所以問題的複雜性並不高，於是我們選取(2)的方式。隱藏層的節點數於是為 $\sqrt{89 \times 2} = 13.34$ ，我們選取 14 作為節點數。

為了增加輸出判斷的精準，輸出層選擇兩個，分別為 $[0\ 1]$ 表示人， $[1\ 0]$

表示非人。這樣做可以增加判斷的準確。人跟非人的判別都要在輸出的其中一個數字由 0 變為 1 (或是接近 1) 及由 1 變為 0 (或接近 0) 的情況下完成，條件較為嚴苛。

我們選取 2000 張人的影像以及 800 張非人的影像做訓練。其中人的訓練影像包含了不同的衣著顏色、性別、行走的姿勢和方向以及拿著背包或是手提包的影像。非人的部份則包括了搖晃的樹葉、狗以及車輛。其中一部份的訓練範例在圖 3-11 與圖 3-12。



圖. 3-9：人的訓練影像範例

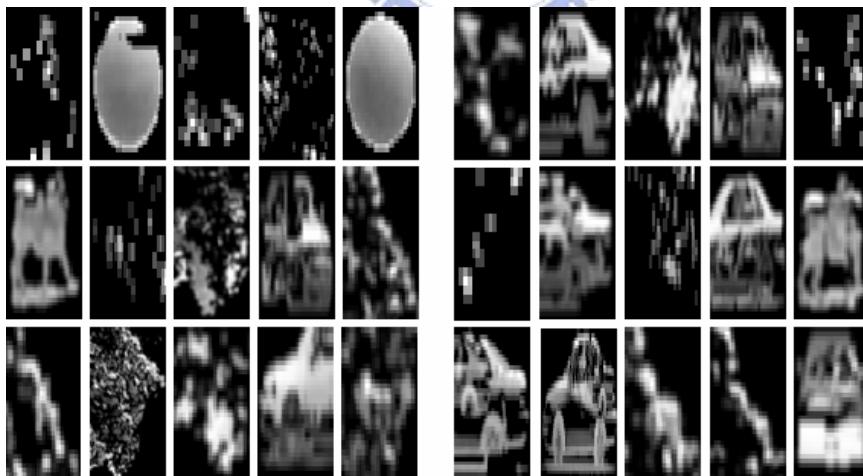


圖. 3-10：非人的訓練影像範例

經過倒傳遞類神經網路訓練後，會產生一個權重矩陣，此舉陣描述了輸入和輸出之間的關係。將之前所提及的平均向量 \bar{w} 和所選取的主成份 C 以及此權重矩陣儲存起來，作為 On-line 的人與非人之判斷使用。到此，Off-line 的訓練

就告一個段落。

3.2 On-line 偵測

當一新的連續影像輸入時，我們首先要做的工作，是將此畫格中的移動物體給分隔出來，並將其大小正規化為 100×50 。這個步驟和之前 Off-line 所使用的方法及步驟是一樣的。

在本論文中，我們對於輸入的連續畫格式以單張的方式做處理，畫格中每個滿足之前所述條件的移動物體，都會分別拿來作判斷。前一節中所提到的被儲存的資料，分別為平均向量 Ψ 和所選取的主成份 C 以及此權重矩陣 W ，在此會被用來與新的輸入移動物體的影像做處理。

首先，有一張經過正規化的影像，大小為 100×50 。其處理步驟如下：

1. 依照 3.1 的方式將此影像的像素重新做排列，使其成為一個 5000×1 的向量 Γ 。

2. 與平均向量做相減，結果為 Φ

$$\Phi = \Gamma - \Psi$$

3. 用所得到的結果與主成份向量 C 做相乘， C 為一 5000×89 的矩陣，其中，5000 表示之前的圖形的總像素，89 為所選取的主成份數目。其結果 F 為一 1×89 的向量，我們稱其為圖形的特徵向量(Feature Vector)，也就是這些特徵向量的線性組合的係數。

$$F = \Phi^T C$$

4. 我們利用得出來的 F 與權重矩陣 W 作相乘，所得到的結果 R 將是一個 1×2 的向量。如果輸入的影像是人，那麼結果將會貼近 $[0 \ 1]$ ，如果是非人的話，會貼近 $[1 \ 0]$ 。

$$R = FW$$

我們利用圖 3-13 對此程序簡短的畫出一流程圖。

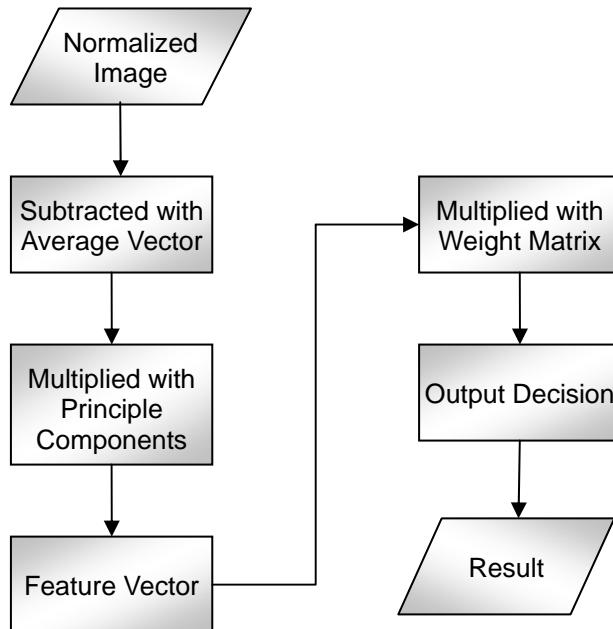


圖. 3-11 : On-line 偵測的流程圖



在此還有一件值得討論的事情，因為經過與權重矩陣相乘之後的結果並不一定剛好為 [0 1] 或是 [1 0]，有的時候只會貼近這個數字而已，例如 [0.02 0.98] 可能表示人的影像，或是 [0.99 0.01] 表示非人的影像。有的時候會產生像 [0.39 0.99] 或是 [0.14 0.88] 的數字，於是，我們就必須訂立一個門檻值，當雙方都滿足此門檻值得條件之後，才正式判定此移動物體為人或非人。例如，如果我們將門檻值設為 0.1 及 0.9，小於 0.1 的數字我們當成 0，而大於 0.9 的數字我們會當成 1。

第四章

人形動態分析

對於靜態的人與非人的辨認，我們是利用單張的畫格分別來作辨認的，並沒有用到動態的資訊。在此一節中，我們利用動態的資訊對於一特殊情況做判斷。前提假設是在一個場景中，存在著一個人形看板以及移動中的人，在特殊的情況下，人形看板也會做移動。如果只是利用單張畫格做判斷的話，這個人形看板也會被分類成人。所以，如何運用動態的資訊，對於人的行為做進一步的分析，為此章節的主要課題。

人的運動行為，在大多數的情況之下，最為顯著的部份，就是其運動具有週期性。參考資料[25]利用 2D 的晶格結構(Lattice Structure)和點陣圖(Dot Matrix)做為運動相似度的測量。我們並不需要使用如此複雜的方式，而利用另外一種較為單純，且具有可靠度的方式判斷人的運動。

基本上，我們將人行走的樣式分為兩種：正面及側面。對於正面和側面，我們有不同的判別方式。在參考資料[15]中，對於側面行走的人，是利用寬度的變化作為判斷依據的。而正面行走的判斷就頗具難度。在參考資料[16]中，提出了一個計算像素數量的方式作為判斷的依據，但對於本實驗的基本架構，並不適用其方法。但是，我們仍舊將人的行走分為正面和側面來作動態的判斷，與畫面中存在的人形看板做區分。

圖 4-1 為動態判斷的流程圖。主要可以分為三個部份：(一)人偵測器(Human Detector)，(二)追蹤器(Tracker)以及(三)程序協調器(Process Coordinator)。一但人偵測器偵測到被辨認為「人」的影像，程序協調器會為每一個辨認成人的影像建立一個追蹤的程序，並且將此影像之後連續數個的一些特徵資訊儲存起來。經過一段時間之後，追蹤器會分析在這段時間內所追蹤的影像的資訊，並決定這個影像在動態特徵的條件之下是否為人。一但判斷過

後，此追蹤程序也自動結束。

接下來，我們就側面的情況先作一介紹，之後則是正面的情況。

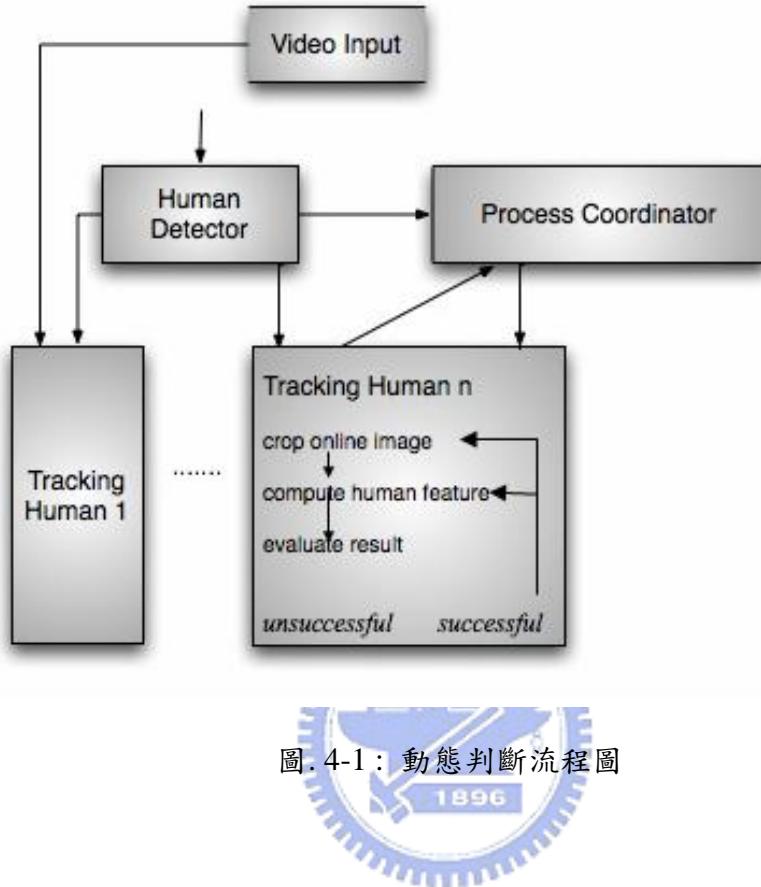


圖. 4-1：動態判斷流程圖

4.1 側面動態判斷

之前我們對於人形的判斷，會先利用許多前處理的技術將移動物體以矩形的方式框架出來。於是，在框架出來的影像尚未經過正規化之前，人在側面行走的時候，因為手跟腳的擺動，在框架時會造成矩形寬度的改變。圖 4-2 為一個側向行走的人在連續二十九個畫格中，寬度的像素數目變化情形。在第一個畫格中，這個人準備進行跨步的動作，第九個畫格的時候，此人的框架有最大的寬度，同時他的右腳在前，左腳在後。從畫格十三到畫格二十一，此人跨出左腳，並在畫格二十五時達到最大的寬度。在大多數的情況下，如果框架的好的話，這種變化會存在著週期性。因為每個不同的人在行走時會有不同的節奏，而我們並不需

要去規定在某一個特殊的頻率範圍才為人，而是要找出這種寬度的變化是否存在一個頻率。如果是，則此移動物體為人，否則則為人形看板。

我們假設人在一秒鐘之內可以完成一部的步伐，因為本實驗所使用的攝影機一秒鐘擷取三十張畫格，因此我們以三十個畫格中寬度像素的變化量做頻率判斷。圖 4-3 為寬度的像素在時域的變化，可以由圖中看出其近似一個弦波。圖 4-4 為圖 4-3 的快速傅立葉轉換(Fast Fourier Transformation)圖，利用此傅立葉轉換圖可以得到圖 4-3 的能量頻譜(Power spectrum)，觀察此變化在何頻率中具有較大的能量，即表示此人應該是以此頻率在觀測場景中行走著。有的時候因為一些雜訊或是矩形框架的問題，在寬度的像素數目變化上並不會相當精準，也就是說，可能在其他的頻率的地方也會產生一些能量，不過與主頻率的能量比較起來，在能量高低上會有顯著的差異，我們以圖 4-4 做說明。因為經過 FFT 轉換的能量頻譜是對稱的，我們僅需觀察前半部份。在圖 4-4 的前半部份中，我們看到在頻率約為 2Hz~3Hz 的地方，有一個明顯的峰值，在 3Hz~4Hz 中，有第二個小峰值。此兩峰值在能量頻譜上的大小差異十分顯著，於是我們可以認定此人行走的主要頻率是在最大的峰值所對應的頻率上，而其次的峰值則可能是雜訊的影響。

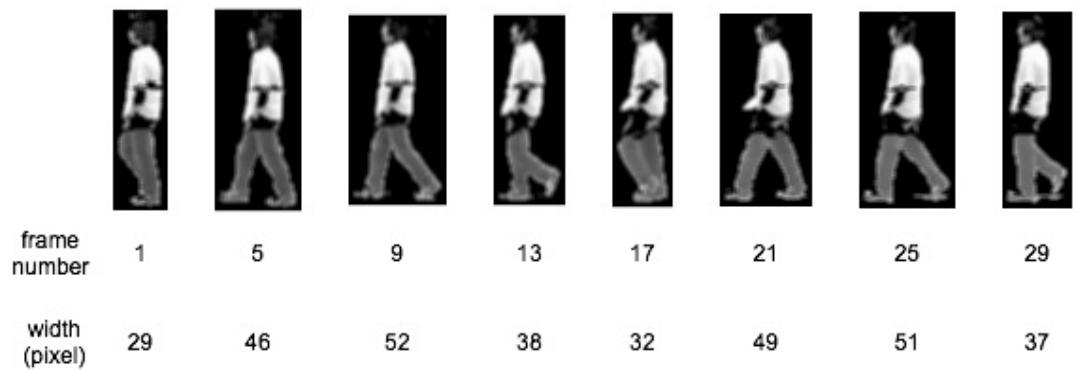


圖. 4-2：側面行走的連續影像

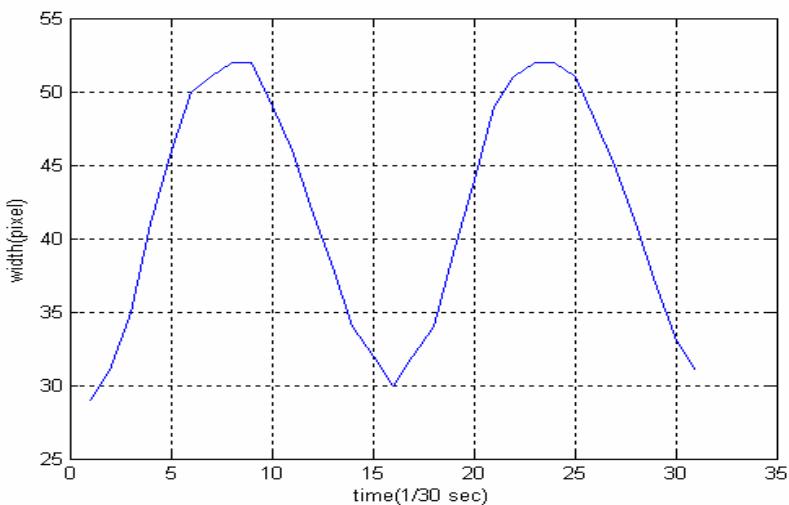


圖. 4-3：寬度的時域變化情形

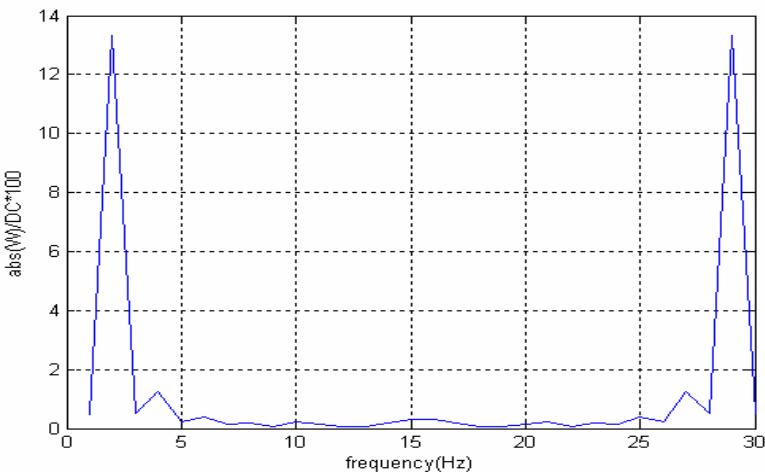
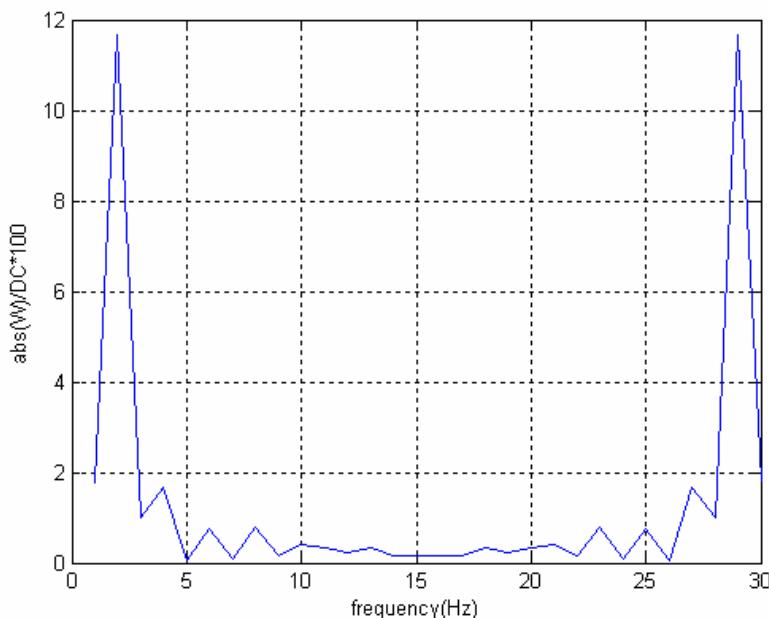
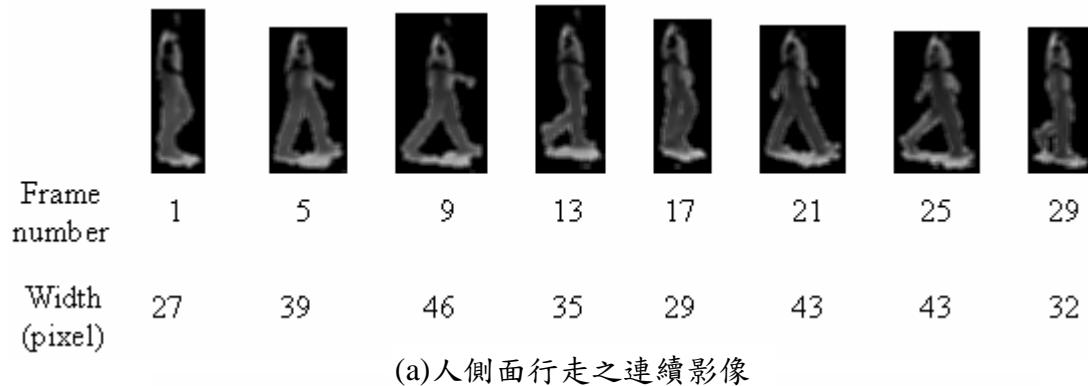


圖. 4-4：圖 4-3 的 FFT 轉換的能量頻譜

接下來的範例我們使用以不同方向，但仍是側面行走的人形做判斷，看是否也同樣具有之前所述的行為。圖 4-5 的人形以左向右移動，在其中幾張畫面中，因為頭髮的顏色與背景過於相似，以致於使用時間差異法之後，其部份在連通成份的技術中無法被框架出來，但是我們在此並不使用高度變化資訊，而只在乎寬度的改變，所以頭部無法框架的部份並不會造成影響。此連續影像的寬度變化情形經過 FFT 轉換之後的能量頻譜如圖 4-5(b)所示。和前述的情形一樣，會發生在某個頻率中具有高能量，並與其他頻率的能量有顯著的差異。



(b)寬度變化 FFT 轉換之能量頻譜

圖. 4-5：側面行走之另一範例

4.2 正面動態分析

正面的人行走判斷相對於側面來說較為困難。在參考資料[9]中，提供了一個以姿勢的樣本為基礎(Sign Pattern-Based)做正面的人和交通工具之分辨判斷。然而，它所提供的方法，在移動物體的分離方面就和本實驗系統不同。再者，參考資料[9]中，所判斷的人身上並沒有背包或是其他會影響寬度的物體存在，且本實驗攝影機所架設的位置並不如其這樣靠近觀測物體，所以其判斷方式並不適合用在本實驗系統上。有些論文利用移動物體整體的質心在整個畫面上的移動

方式作為判斷條件。但本實驗所要解決的問題是，如果有人拿著人形的看板，躲在在其後頭進行移動，我們還是要判斷其為非人；所以此方式並不適合。於是我們觀察人以正面行走時，發生變化的地方，並試著利用此特徵做判斷依據。

圖 4-5 為一正面遠離攝影機的人的影像，尚未經過正規化。在連續的 29 個畫格中，其寬度與長度並沒有明顯的變化。又因為衣著的關係，也無法利用骨架 (skeleton) 的條件做判斷。此連續影像中，有變化的地方只有手部和腳部。但若是人的手上或肩上背著背包或是提著手提包，則手部也無法用作辨認依據。於是我們將特徵的選取重點放在腳部的地方。

圖 4-6 為圖 4-5 經過正規化成大小為 100×50 之後的情形，我們特別將重點放在長度 70~90 的地方。因為攝影機架設角度的關係，人在行走時前後腳交錯，會使得此部份的質心有所偏移。選擇 70 的原因是經過正規化之後的圖形，此數字約略在人膝蓋的位置。選擇 90 的因素是因為儘量要避免陰影的影響，同時也只要觀測小腿在影像之中的位置，如圖 4-7 用紅色矩形框起來之處。

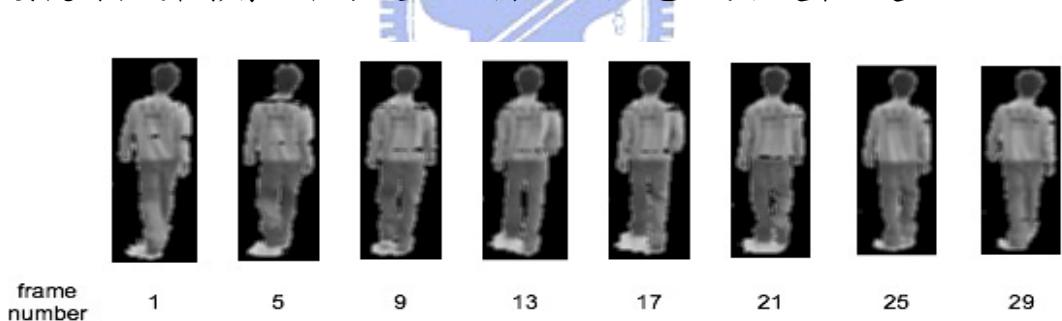


圖. 4-6：人正面行走之連續影像



圖. 4-7：紅色矩形處的腳部移動

圖 4-8 為圖 4-7 紅色矩形範圍中，連續影像的質心移動情形。我們選取了

連續的 30 張圖。由圖 4-7 中可以發現，影像中的人先跨出左腳，於是顯示在圖 4-8 中，質心的位置向左方移動，在圖中的位置就是向原點方向。之後跨出右腳時，質心的位置又會往右移。一些稍微不規則的變化，可能是無法消去的陰影部份或是腳的位置所造成的影響，不過大抵上，質心的變化是依照此趨勢在做移動的。而人形看板不論是靜止或是緩緩移動的情況，因為其形狀固定，所以質心並不會發生偏移的情形。依照此特徵，我們可以判定此移動物體為人或是非人。理論上，如果此人以正常的方式行走，且能夠完全消除陰影及其他雜訊的干擾的話，此移動方式應該也會有一固定的頻率。但在現實中，不論是人的跨步或是干擾，幾乎不可能完全達到理想狀況的要求，所以經過 FFT 的頻譜轉換，產生最大峰值的情況並不若橫向行走時的寬度變化那樣明顯，所以這裡，我們並不採用 FFT 的方式加以分析，而只就質心的變化方式作為分析的條件。

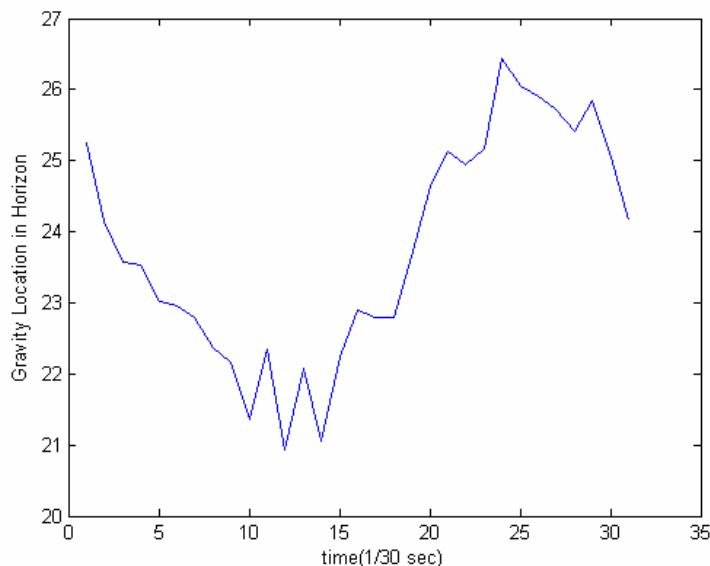
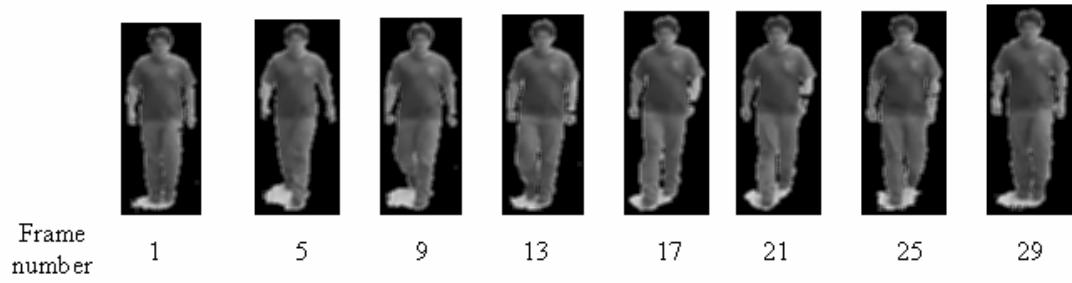


圖 4-8：質心的位置移動

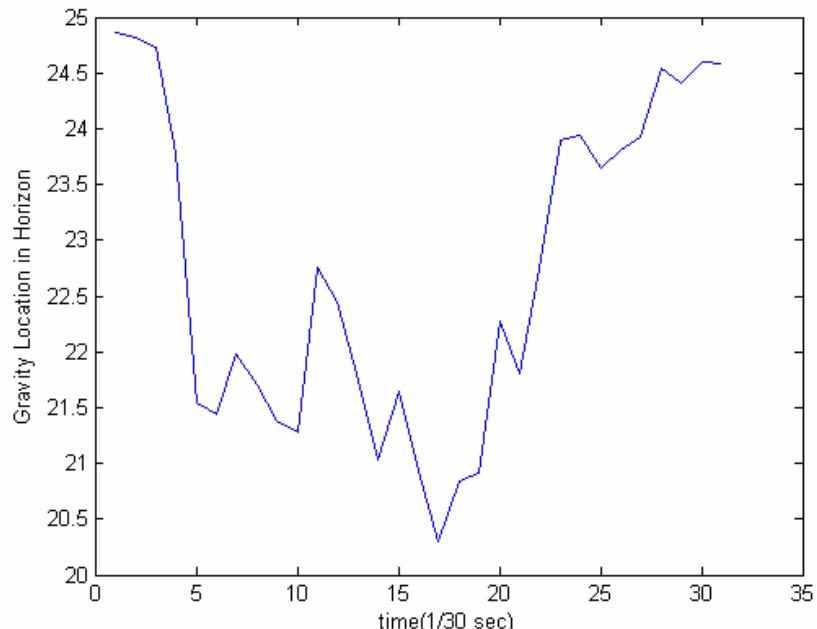
圖 4-9 為另外一個例子，和圖 4-7 不同的地方在於，圖 4-7 中的人以遠離攝影機的方向行走，而圖 4-9 為面向攝影機行走。在圖 4-9 中，(a)為正面向攝影機移動的人影像，(b)為其選定區域的質心移動情形。質心的移動仍舊隨著左右腳往復性的運動而往同樣的方向偏移。

仔細觀察圖 4-8 以及圖 4-8 中(b)的部份，雖然兩者的變化性沒有一固定的

頻率，但是基本上，只要人移動的方式正常的話，質心位置還是會有隨著步伐回到和原先起始位置差不多相同之處的特性。



(a) 人正面行走之連續影像



(b) (a)腳部質心的移動情形

圖. 4-9：正面行走另一範例

第五章

實驗結果

本章節將實驗結果分成靜態與動態兩方面做介紹。在 5.1 當中，先分別介紹靜態的 Off-line 的訓練結果和 On-line 辨認的部份，接著 5.2 則說明了動態追蹤的實驗結果。

5.1 靜態人與非人之辨認

本系統利用影像的靜態特徵，可做人與非人之辨認。首先在 5.1.1 中，我們對於 Off-line 時所選取的主成份數量的不同，在辨別上有不同的效果，說明選取此數量的理由。接著 5.1.2，我們介紹將此方法運用在實際狀況的辨別結果。

5.1.1 Off-line 人與非人之辨認



我們選取了 2000 張人的影像跟 800 張非人的影像作為訓練，同時，我們也選取了 215 張人的影像以及 90 張非人的影像作為測試，比較在不同的主成份數量下以及不同的門檻值設定下，其辨識的正確率。如果選取的主成份數量少，雖然在計算上較為省時，但因為所含的圖形資訊少，可能會失去其判斷的準確性。而選取的主成份數量多，雖然在判斷上較為準確，卻必須花掉較多的計算時間。而較高的門檻值會對判斷的準確率有提升，但對於一些較為模糊的人形影像，則會有判成非人的可能。所以我們必須選取適合的主成份個數以及門檻值。

表 5-1 和表 5-2 分別說明了訓練的樣本和測試的樣本在不同的主成份數量和門檻值下的正確率。由於理想的輸出數字為 0 或 1，經由不同的測試結果之後，在靠近 1 的地方，我們給予三個門檻值做測試：0.95、0.9 和 0.85，而觀察輸出的結果，在靠近 0 的輸出值再這三個門檻值下，多低於 0.1，所以我們就設定一方必須低於 0.1，另一方則用此三個門檻值來作比較。另外，主成份的數量選取，

我們分別選了足以代表圖像 80% 資訊的 60 個，85% 資訊的 89 個，以及 87.5% 資訊的 111 個，而這些不同的門檻值在經過訓練之後，其均方誤差都要達到收斂的狀況才行。在表 5-1 和 5-2 中，最左方的三個數字代表門檻值，上方的數字代表了主成份數目的選取，用訓練資料和測試樣本分別做判斷正確率的測試，用此結果作為系統使用的依據。

表 5.1：訓練資料的正確率

	人： 2000 張影像			非人： 800 張影像		
	60	89	111	60	89	111
0.95	96%	100%	99.95%	100%	100%	100%
0.9	96.35%	100%	100%	100%	100%	100%
0.85	96.7%	100%	100%	100%	100%	100%

表 5.2：測試資料的正確率

	人： 215 張影像			非人： 90 張影像		
	60	89	111	60	89	111
0.95	92.6%	96.3%	95.81%	94.44%	90%	91.11%
0.9	93%	96.3%	95.81%	95.56%	91.11%	91.11%
0.85	94.4%	97.2%	96.74%	96.67%	92.22%	91.11%

我們先觀察不同主成份數量的結果，發現在對於人的判斷上，60 個跟 89 個不同的主成份數量對判斷的結果會有一定程度的差距，而 89 個跟 111 個幾乎沒有差距，所以我們選擇 89 個主成份作為系統的使用。接下來觀察門檻值的部份，因為不論 0.85、0.9 或是 0.95 都可以達到一定的效果，相距其實不遠，所以我們先選擇 0.9 作為實驗的門檻值數字。

得到這些數字之後，我們在下一節中利用產生的結果在實際的情況下作測試。

5.1.2 On-line 人與非人之偵測

在本實驗中，我們選用國立交通大學校內的幾個地方作為實驗的場景。本實驗使用的硬體設備為 Pentium 4 CPU 2.60Hz，作業系統為 Microsoft Windows XP。本實驗主要的軟體開發平台為 Borland C++ 6.0，除了求取特徵向量的部份是用 MATLAB 7.0，以及倒傳遞類神經網路的訓練是用 Visual C++ 6.0 的部份完成的。我們將攝影機架設在交通大學學校內部的幾個場景，如圖 5-1 所示。此實驗系統可以分辨在畫面中的人與非人之情況，並且進一步加以計數。

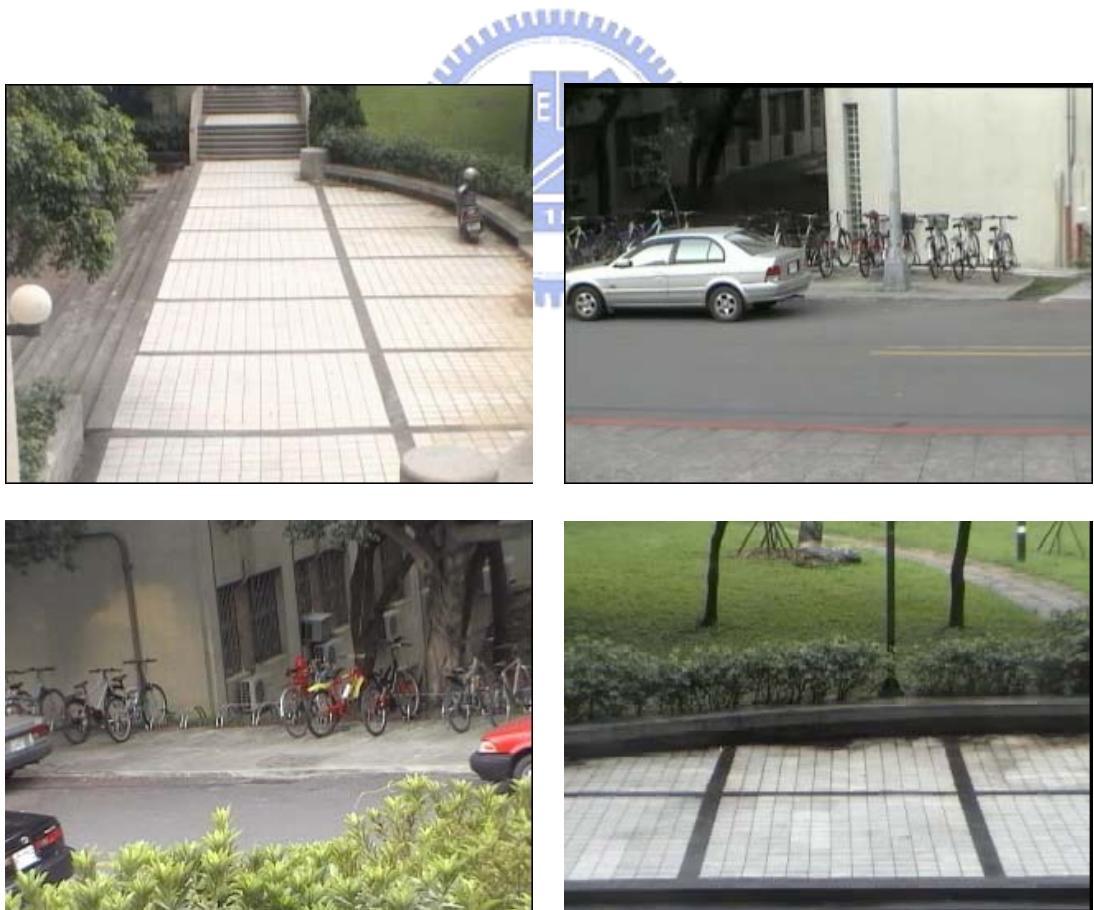


圖. 5-1：觀測場景

首先，我們在每個場景中錄製一段影片，將此影片檔以每秒 30 個畫格的方

式讀進檔案中。選定一個空景為背景之後，以時間差異法擷取其中的移動物體，再經過一些平滑濾波、連通成份的處理之後，正規化每個符合條件的個別物體影像至 100×50 的標準大小，之後與主成份矩陣相乘，得到此影像的全域特徵向量後，再與權重矩陣做相乘，於是得到一個 1×2 的輸出矩陣。

在此實驗中，我們將與前景與背景之灰階值差異大於 35 取出，而經過連通成份處理後的影像大小必須大於 1000 這個門檻值才會被取出，做接下來的判斷動作。判斷為「人」的門檻值設為 [0.1 0.9]。

此程式的執行界面如圖 5-2 所示，我們將其作一基本的介紹。

Region 1：此區域顯示原始的影像。在此區域當中，會有紅色及藍色的矩形將移動物件框架出來。藍色的表示被判定為「非人」的物體，紅色的表示被判定為「人」的物體。



Region 2：利用時間差異法及經過均值平滑濾波器處理後的結果會顯示在此區域。為了方便起見，此區域只用紅色的矩形框出判定為人的移動物體。

Region 3：被判定為「人」的物體經過正規化成 100×50 之影像會顯示在此區域。

Region 4：此區域的最上方顯示了移動物體的總數，判定為「人」及「非人」之個數。Detection Size 的地方表示了經過連通處理的移動物體影像必須大於此值，才會做接下來判斷的動作。

在 Noise Elimination 的地方，我們都將其設定在「Average」之處，也就是都利用之前所敘述的均值平滑濾波器作一雜訊消除的動作。

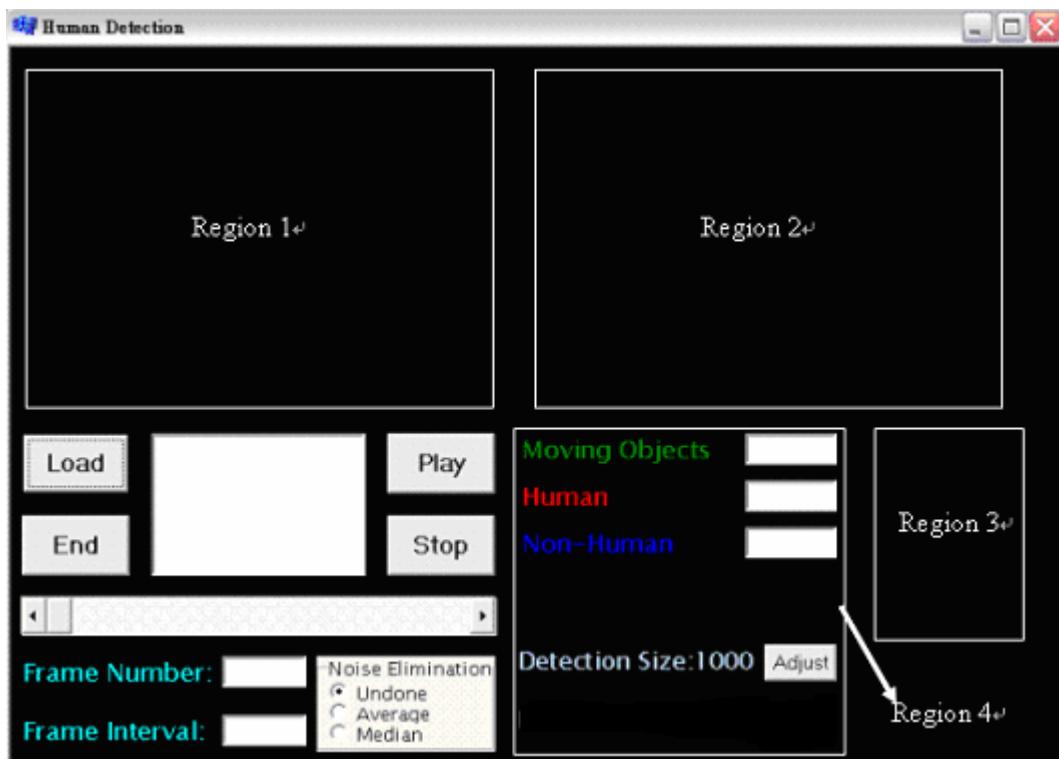


圖 5-2：程式執行介面

在我們所觀測的場景當中，人的影像包含了不同樣式的衣著、不同性別、行走的姿勢與方向、背著背包或是提著手提包的情況。而非人的影像則包含了狗、移動的樹葉、球以及車輛。

圖 5-3 說明了當一個畫面中有人與非人的物體同時存在的話，本系統可以將其做適當的分類判別。在圖中，我們有一個人以側面由左到右行走，有一台車也是從左方到右方開過。於是，在系統上就會顯示出有兩個移動物體在畫面當中。其中人的部份用紅色的矩形框住，車子用藍色的矩形框住。而在 Region 2 的地方，就只有判定為「人」的移動物體用紅色的矩形框住。在 Region 4 的地方，顯示了判定為人的影像經過正規化處理之後的情形。

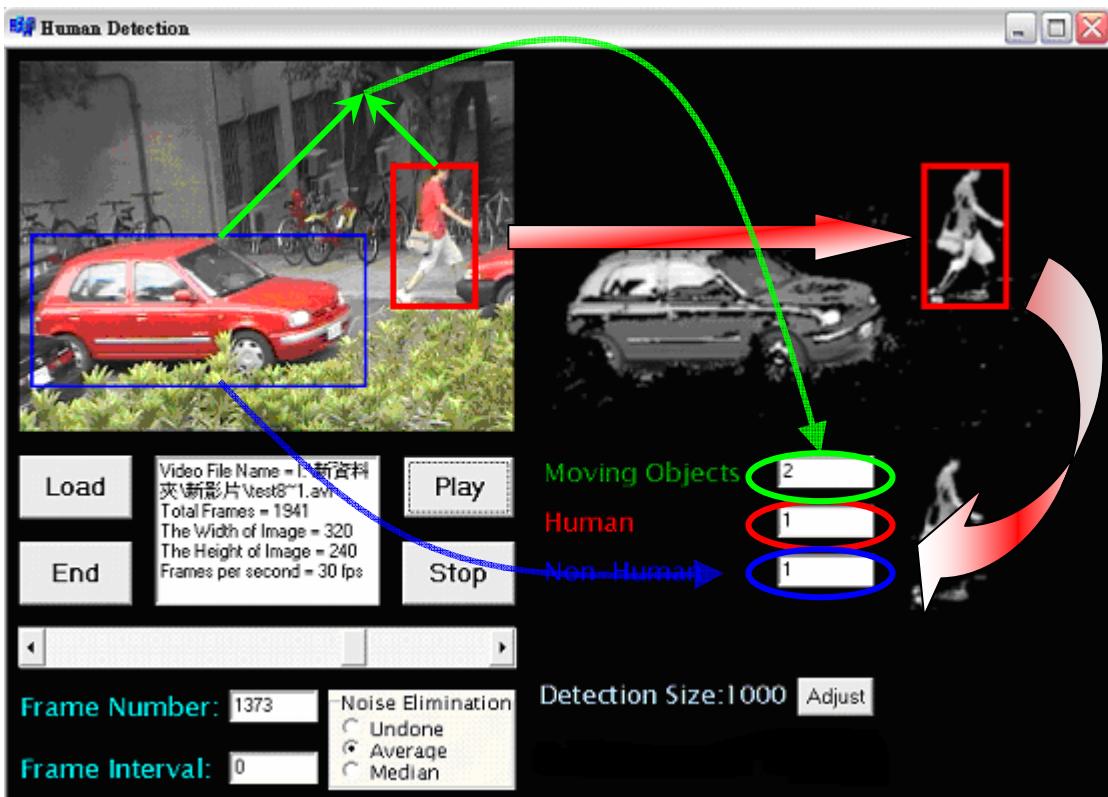


圖. 5-3：人與車的判別情形

同樣的，對於人以不同的方向行走，也能正確的判斷。在圖 5-4 中，兩個人從不同的方向走來，系統判別的結果也是正確的。圖 5-5 裡，有一個人手拿著雨傘，另外一個人則是背著背包，在我們的實驗結果中，也能夠正確的判別出來。

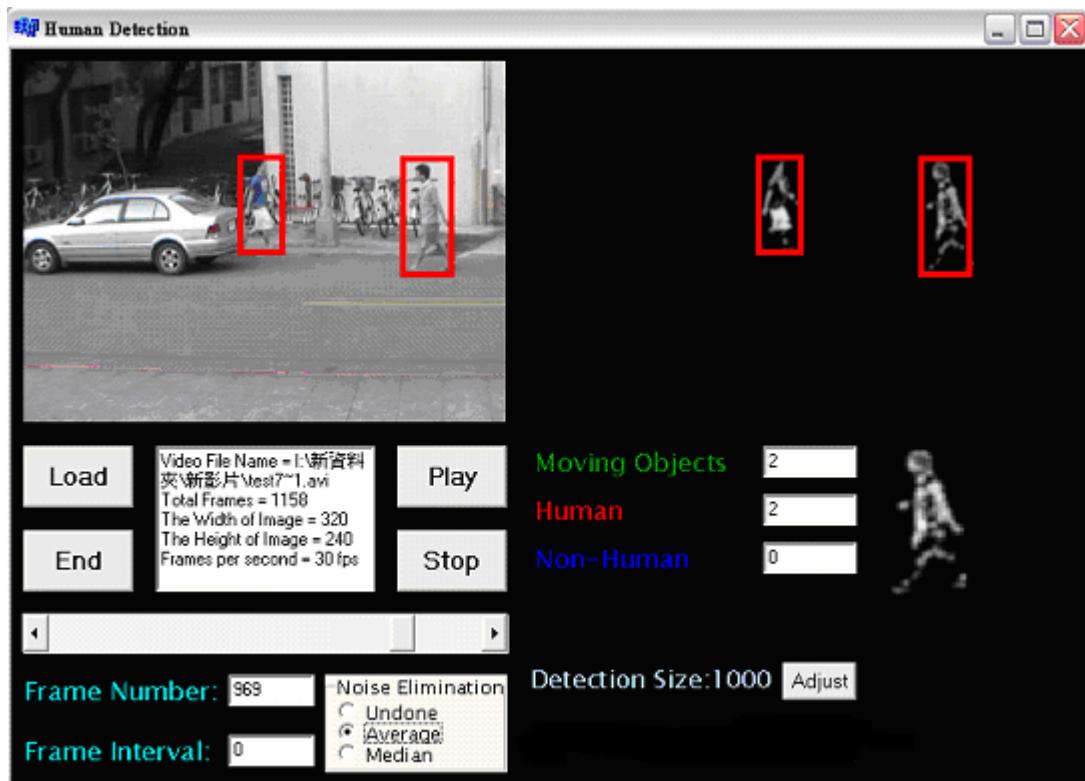


圖. 5-4：人以不同方向行走



圖. 5-5: 人背著背包與提著物品

圖 5-6 中，左方的樹葉搖晃的很劇烈，於是在經過時間差異法以及均值濾

波器的處理之後仍舊不能完全消去其雜訊，造成畫面中人與非人的個數皆超過一個，但是個別移動物體的正規化圖形經過與主成份還有權重矩陣相乘之後，得到的判別結果仍是正確的。

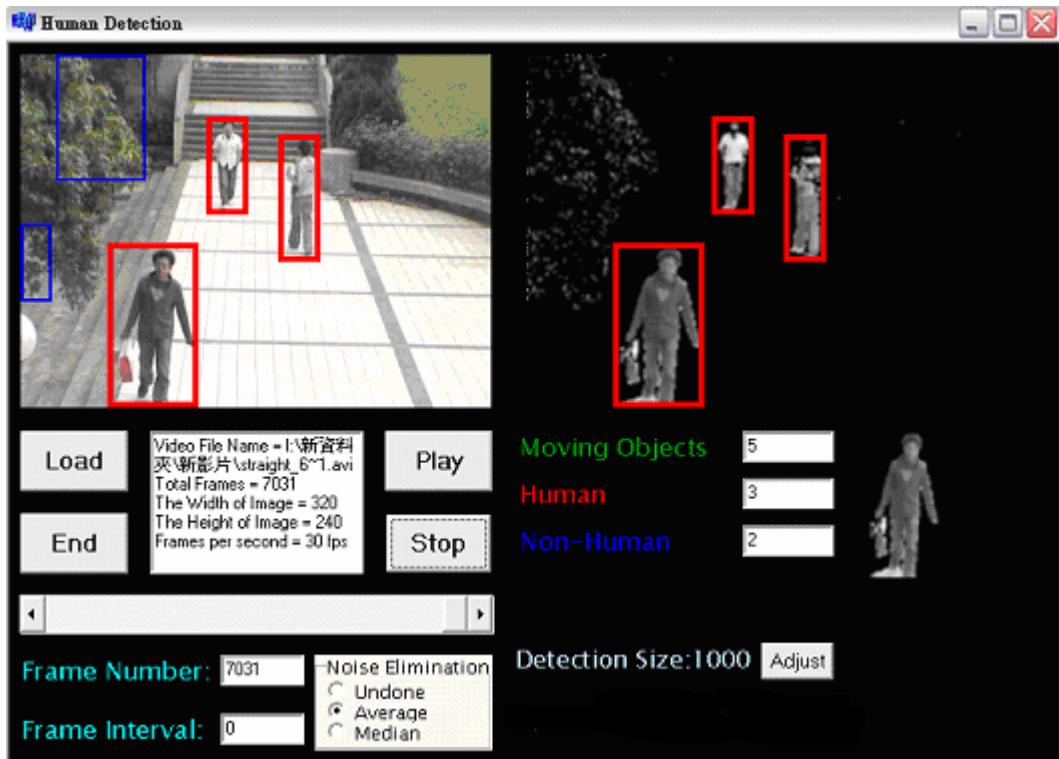


圖. 5-6: 多人與非人的移動物體

但本實驗系統對於兩種情況的發生，則會有誤判的情形發生。一種適當兩個移動物體過於靠近時，另外一種則是有遮蔽的情況發生。圖 5-7 中，兩位正面行走的人因為過於靠近，使得背包的部份在我們的技術處理中連在一起，所以被當成同一個移動物體，於是誤判的情況發生。圖 5-8 和圖 5-9 則是說明當遮蔽發生時，本實驗系統所判斷的情況。圖 5-8 發生遮蔽時，其結果被判斷成非人。而圖 5-9 則被判斷成人。

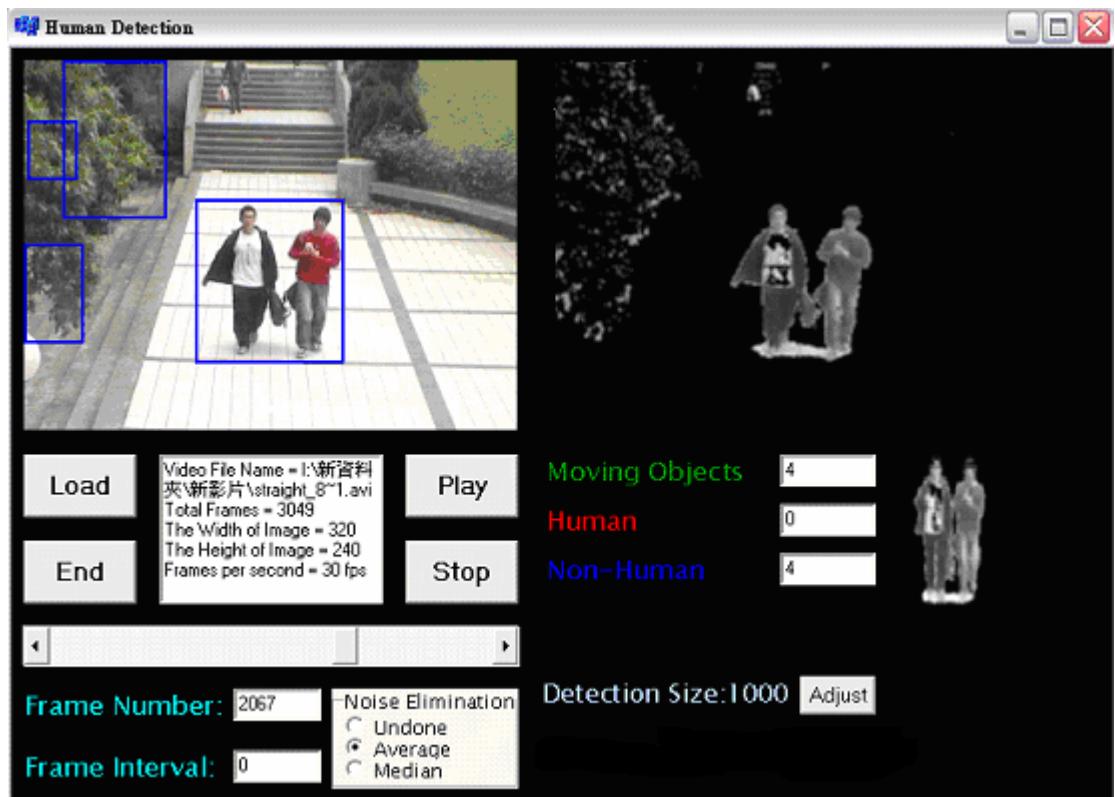


圖. 5-7: 誤判（一）移動物體過於靠近

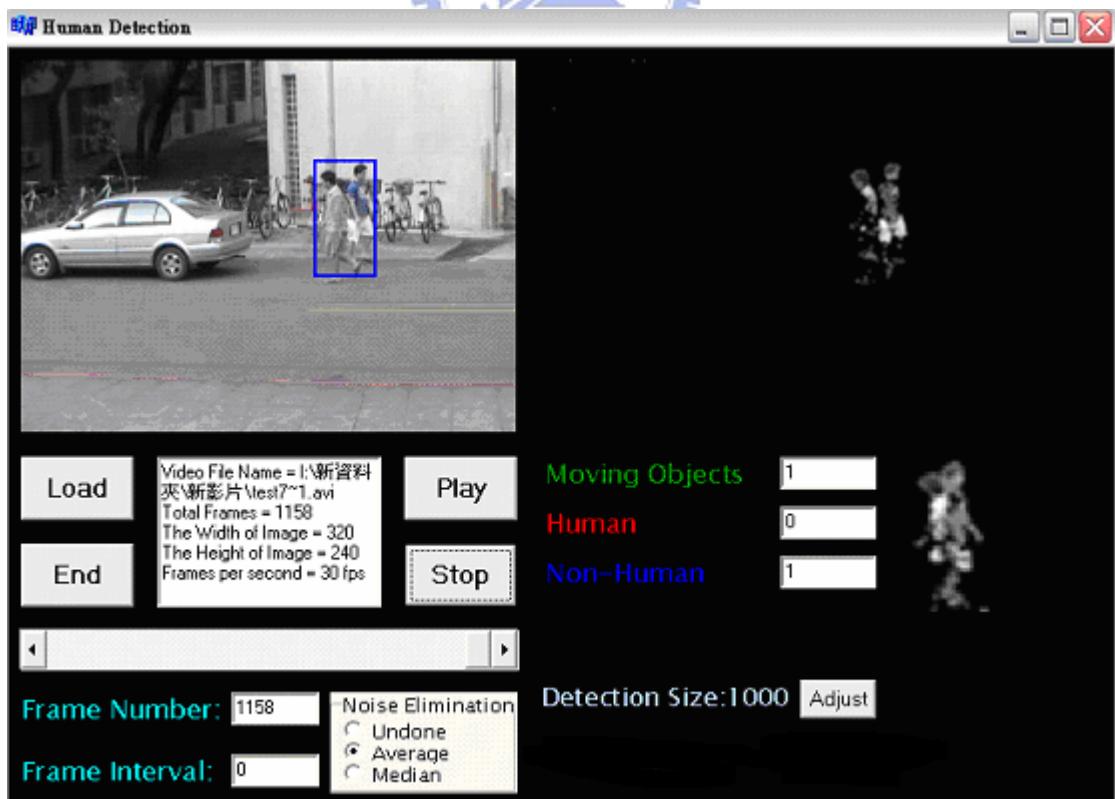


圖. 5-8: 誤判（二）遮蔽情況判為非人

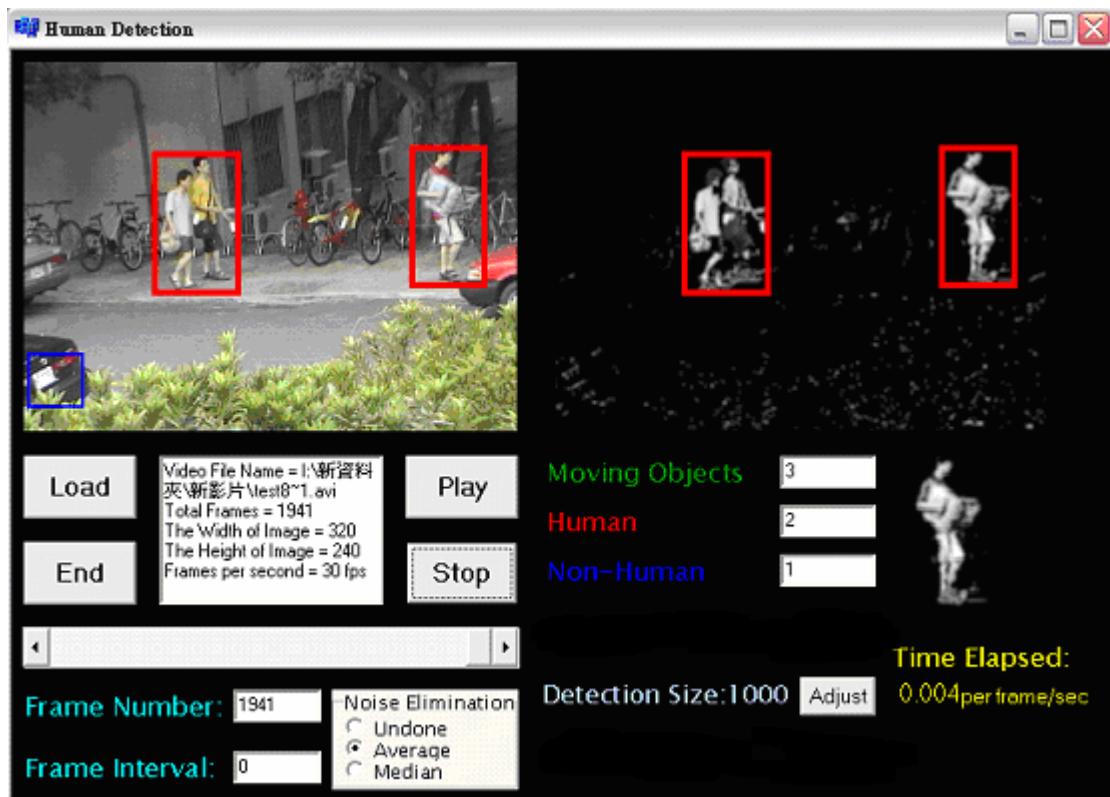


圖. 5-9: 誤判（三）遮蔽情況判為人

5.2 動態人與非人之辨認

在靜態的辨別中，若存在著外型和人相似的看板，則靜態的判斷會將其辨別為人，而不是非人。於是，在本章節中，我們根據第四章所提出的方法，因為在人以側面及正面的移動，我們所利用的判斷特徵不同，於是將實驗結果分成動態正面與動態側面二部份做介紹。

5.2.1 動態側面資訊

側面的資訊，如同 4.1 節所說的，是以人移動時，寬度變化的頻率來作判斷。在實驗中，我們首先選取背景的圖片，之後將出現在畫面中的人形看板與移動中的人和背景相減，同樣經過均值濾波器與兩次膨脹的前處理，再利用所得到的座標值擷取出灰階影像中的人形以及人形看板。這些步驟以及方法都和靜態辨

認的影像前處理方式相同，只是不需要經過正規化的動作。圖 5-10 為觀測場景的背景圖、前景圖，以及相減之後的灰階結果。

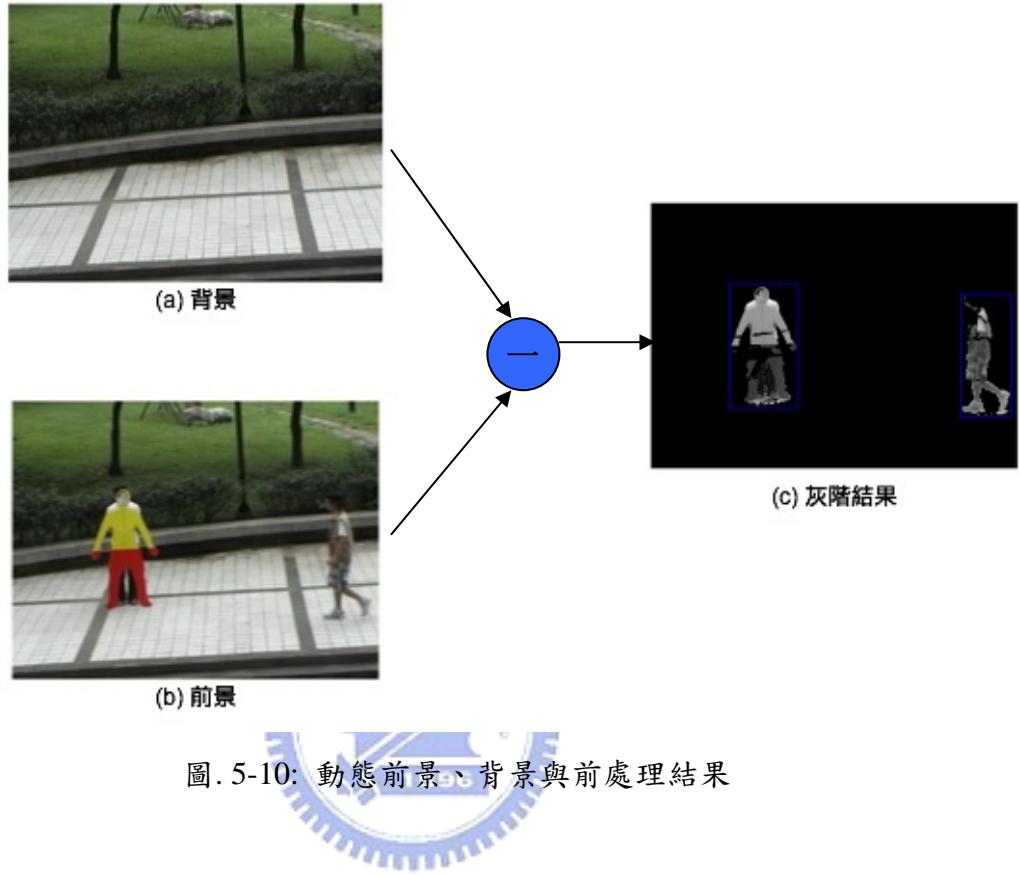
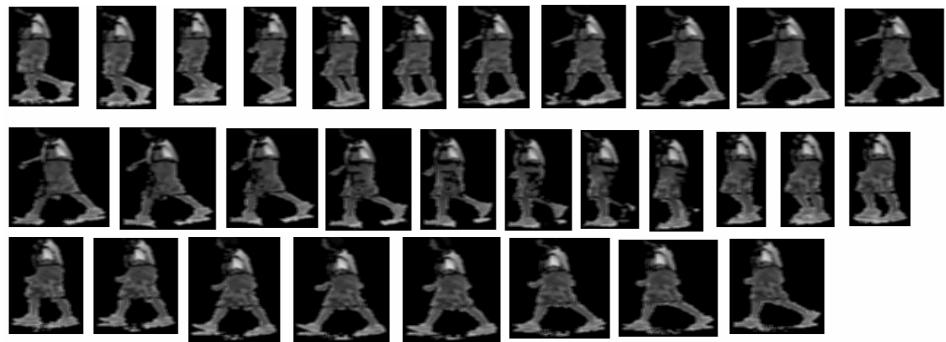


圖 5-10: 動態前景、背景與前處理結果

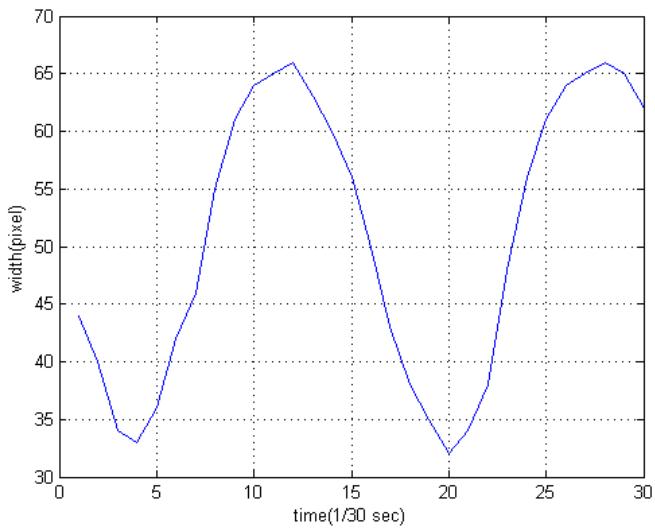
同樣的，我們利用 30 個，也就是一秒鐘的連續畫格，從中擷取出人與人形看板，觀察其寬度變化的頻率性。圖 5-11 中，(a)為連續 30 張移動的人的分離圖形，我們可以看出其寬度變化非常顯著。(b)為寬度變化在時域上的表現情形。(c)為將寬度變化做 FFT 轉換之後的能量頻譜。和 4-2 節所使用的判斷特徵相同，因為人運動時的週期往復性，其寬度變化呈現增減的情況，會造成在某個頻率上具有較大的能量，所以產生一峰值，而且與其它的峰值大小相差甚大，於是此為人的移動。

圖 5-12 中，(a)為連續 30 張人形看板的影像，(b)為其寬度變化在時域上的表現。(c)為將寬度變化做 FFT 轉換之後的能量頻譜。觀察(b)的情況，發現只有剛開始時，其寬度的像素值從 60 變為 61，之後就沒有變化的情形。所以在能量的頻譜上，產生的結果為一直線，就算有一些雜訊的干擾，在能量頻譜上的相

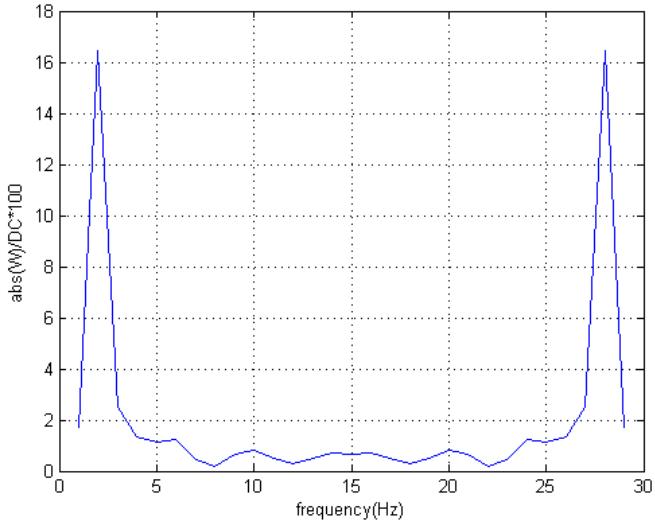
差結果也不會很大。利用此一方式，確實能夠準確的判斷在側向移動時，人與人形看板的分別。



(a) 30 張人移動的連續影像

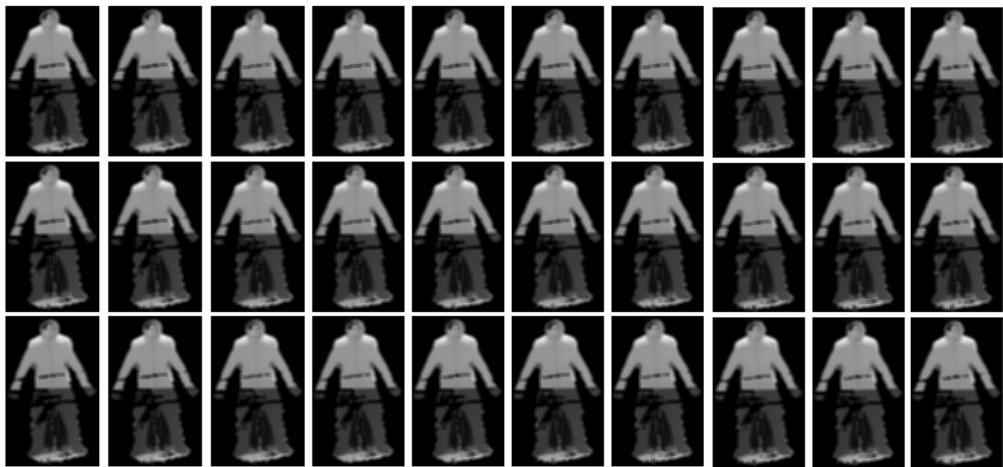


(b) 寬度的時域變化

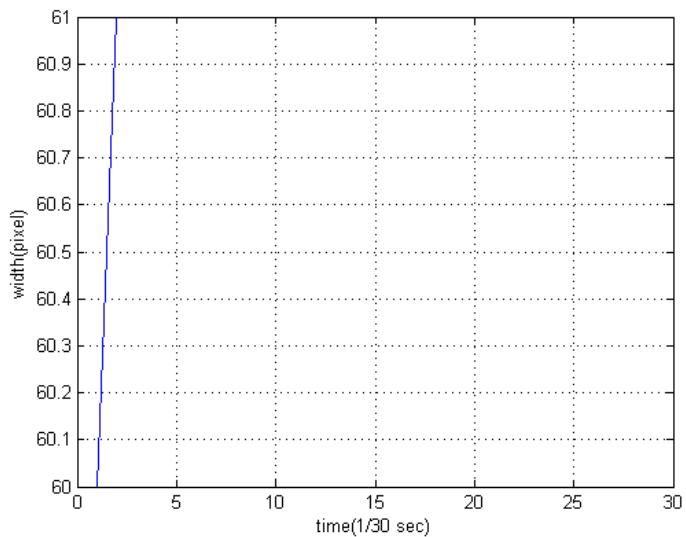


(c) 時域變化的能量頻譜

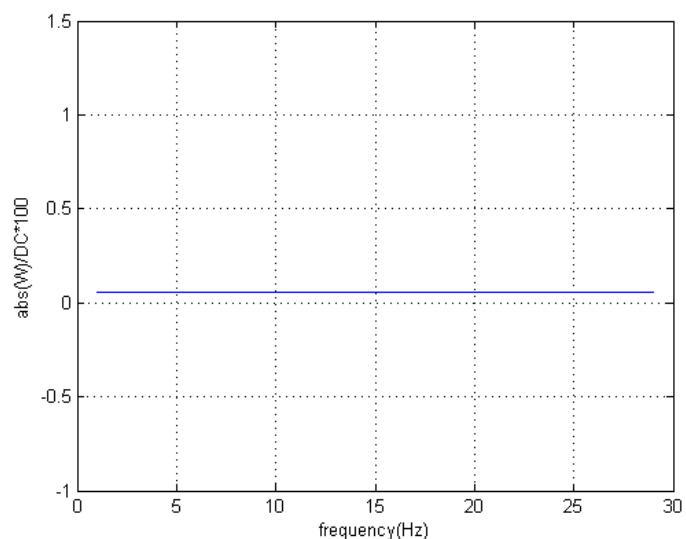
圖. 5-11: 人側向移動的處理結果



(a)人形看板 30 張連續圖像



(b)人形看板寬度之時域變化



(c)時域變化的能量頻譜

圖. 5-12: 人形看板的處理結果

5.2.2 動態正面資訊

本實驗的操作方式，是假設我們所觀測的場景中，有一移動中的人，以及一個緩緩移動的人形看板。可能發生的情況是人形看板底下架設軌道，以滑行的方式朝向或是遠離攝影機。因為器材取得不易，所以我們使用的方式是以人拿著人形看板，將自己的身體隱藏在其後，儘量以不晃動的方式讓人形看板移動。這個實驗的目的主要是在找出當人以正面移動時所具有的運動特徵。圖 5-13 為觀測場景的背景圖、前景圖以及相減之後的灰階輸出結果。

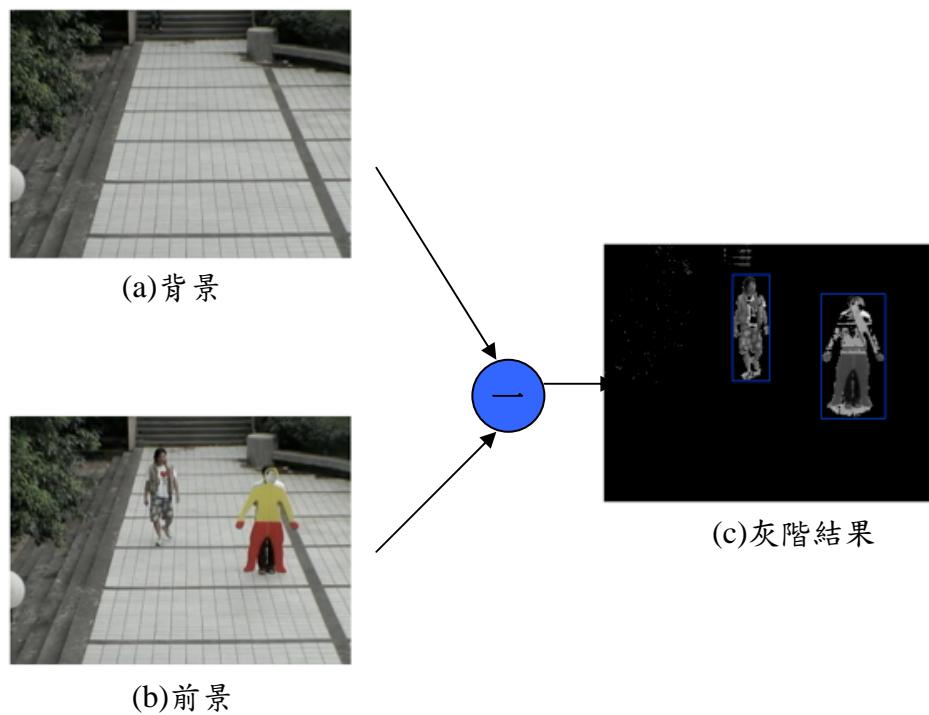


圖. 5-13: 動態正面前景、背景與前處理結果

我們同樣利用連續 30 張的畫格，將其中的移動物體分離出來，從中擷取移動中的人以及人形看板，分析其運動特徵。觀察的重點在 4.2 中提到，是經過正規化的圖形，其長度 70 至 90 的地方。圖 5-14 中，(a)為正面行走的人的分離影像，經過正規化處理後的結果。因為此人為面對攝影機的方向走來。起始的位置是左腳在前，右腳在後，之後跨出右腳。就本實驗方法，質心會向左邊，也就是

離原點較近的地方作移動，之後跨出左腳時，質心再往遠離原點的方向移動。圖 5-14(b)中確實顯示了此現象。圖 5-15 為有人拿著看板緩緩向前移動的連續 30 張經過正規化的分離影像。因為看板外型的關係，背後拿著的人的雙腳在細部移動的過程中，造成質心會發生小小的變化情形。理論上來說，如果只有看板在移動的話，而不考慮突然間的光度變化，則看板紅色區域的質心應該不會發生變化。但在同樣量測的刻度座標上來看，雖然腳步的細碎移動影響了結果，但變化的幅度仍舊較一般人以正常步伐行走時要來的小，也證明了此運動特徵是可以用來作判別的，而且不受人是否提著手提包或是會影響寬度資訊的物體的影響。

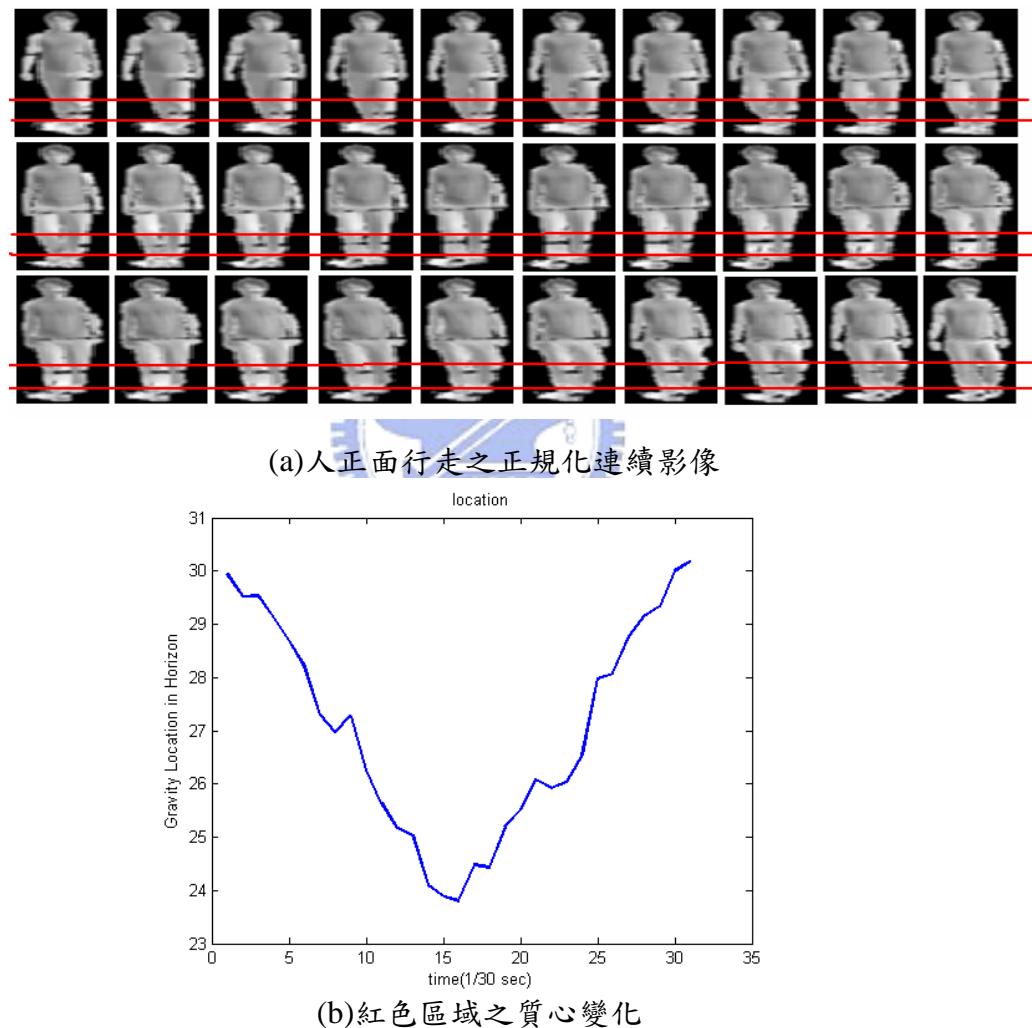
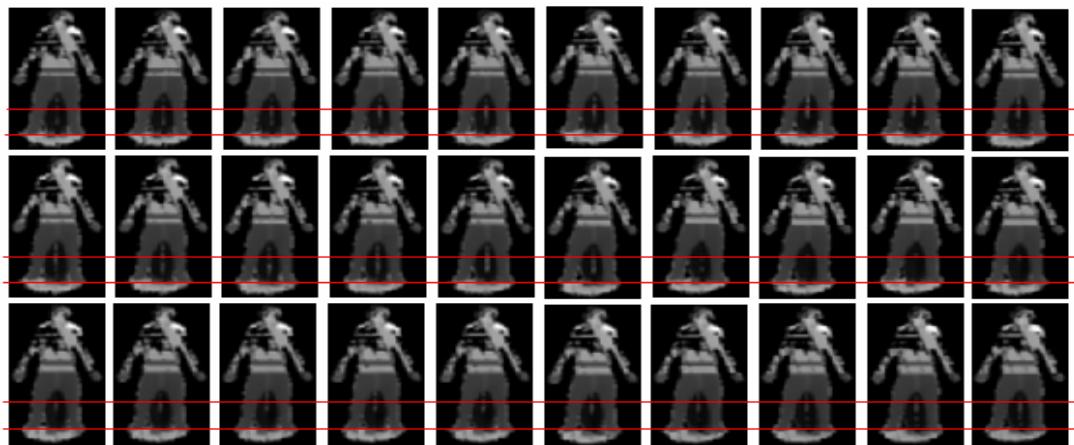
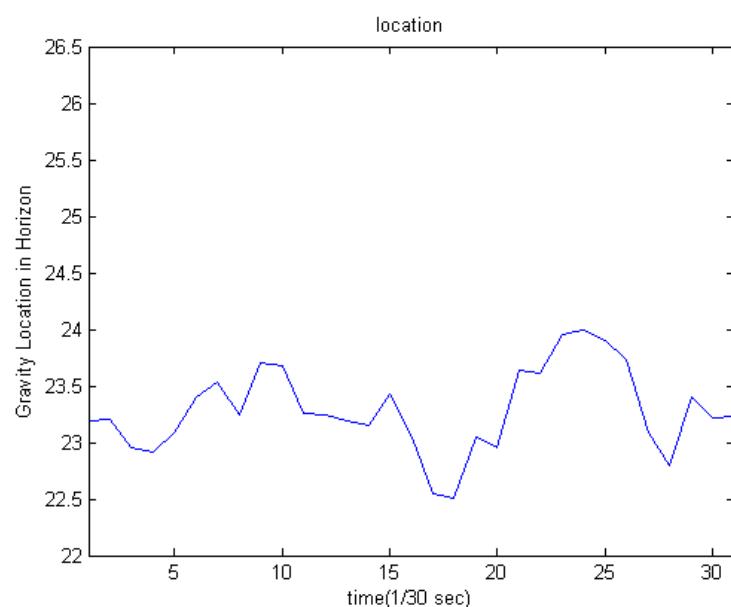


圖. 5-14: 人正面處理結果



(a)人形看板正面之正規化連續影像



(b)紅色區域質心移動情形

圖. 5-15: 人形看板正面處理結果

第六章

結論與未來展望

本論文對於人與非人之靜態辨認及動態追蹤提出了新的方法作為判斷依據。在靜態辨認的部份，可分為移動物體的分離以及人與分人的辨認。在移動物體的分離這個步驟，我們使用了時間差異法以及一些影像雜訊消除的前處理，在人的辨認這個步驟，將移動物體的全域特徵利用主成份分析法選取具有代表性的部份，經過與主成份矩陣相乘之後，每個影像經過降維的資料會被送進倒傳遞類神經網路做訓練，以輸出結果做為人與非人之判斷。

在動態追蹤的部份，我們在橫向利用寬度變化的頻率，在直向利用特定位置的質心移動作為特徵，用以分辨人與人形紙板（非人）的判斷依據。

靜態辨認時的主成份個數，依照我們分別取可以代表所有人與非人圖形 80%、85% 和 87.5% 的資訊，分別是 60、89 及 111 個。經由正確率以及運算速度的考量，我們選取 89 個作為主成份的個數。在正確率的辨識上，如果將輸出的門檻值定為 0.1 和 0.9 的話，人的判斷部份可以達到 96.3%，對於非人的辨識率則有 91.11%。

動態追蹤的地方，我們利用連續多個畫格寬度變化的頻率性以及特定區域質心移動的情況為特徵作判斷。由實驗結果可以看得出來，在不發生交錯的情況下，都能夠正確的判斷。

本實驗有以下幾個特點：

(一) 本實驗選取了 2000 張的人形影像跟 800 張的非人影像作訓練，也就不需要先前的樣板或是人形的模型。

(二) 在特徵的選取上，我們是利用全域特徵(Global Features)取代其它例如 Haar-like 的特徵，在判斷的準確率上相對地提高。

(三) 對於主成份分析法(PCA)的使用，，將人與非人的影像同時選取主

成份，在較為複雜的環境中對於人與非人也能夠準確地做判斷。

(四) 在動態追蹤的人辨認方面，提出了新的想法，可利用在未來更進一步的研究當中。

本實驗系統的應用層面廣泛。在防盜系統的使用中，我們可以利用此系統偵測小偷的闖入，並發出警報警示。在對於人數需要做控制的地分，本系統也能夠計算出場景中的人數多寡。

雖然本實驗系統在辨別人與非人上的效果良好，但仍舊存在著一些需要克服的缺點。我們的攝影機是以固定的方式架設著。雖然對於一些輕微的樹葉晃動以及陰影的雜訊可以經過一些前處理加以過濾，然而，攝影機的輕微晃動或是所觀測場景發生改變將會產生嚴重的影響。此缺點可利用定期的背景更新作為解決的方式。另外，因為我們所選取的訓練樣本都是人以側向或是直向方式行走的樣本，對於觀測場景中若是存在著人以爬行或是半蹲等其他姿勢的話，則辨別不出來。一個改進的方式是可以將此圖像加入訓練樣本中，不過可能會與其他非人的影像，如狗等發生形體相近的情況，而增加判別的困難度。當人發生重疊或是彼此距離接近時會發生誤判的情況，可能就需要一些形態分析才能加以解決。

在動態的判斷方面，一但發生重疊的情形，則連通成份的技術會將重疊的兩個物體當成同一個，以致於發生無法判斷的情形，要等到重疊的物體分離之後，才能重新做判斷。

參考文獻

- [1] P. J. Burt, and J. R. Bergen, "Object tracking with a moving camera: An application of dynamic motion analysis," in *Proceedings IEEE Conference Workshop Visual Tracking*, 1989, pp. 2-12.
- [2] P. H. Batavia, D. E. Pomerleau, and C. E. Thorpe, "Overtaking vehicle detection using implicit optical flow," in *1997 IEEE Conference on Intelligent Transportation System*, 1997, pp. 729-734.
- [3] W. J. Gillner, "Motion based vehicle detection on motorways," in *Proceedings of the Intelligent Vehicles '95 Symposium*, 1995, pp.483-487.
- [4] T. Darrell, G. Gordon, M. Harville, and J. Woodfill, "Integrated person tracking using stereo, color, and pattern detection," in *1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1998, pp. 601-608.
- [5] L. Zhao, and C. E. Thorpe, "Stereo- and neural network-based pedestrian detection," *IEEE Trans. Intelligent Transportation Systems*, vol. 1, pp. 148-154, Sept. 2000.
- [6] C. E. Smith, C. A. Richards, S. A. Brandt, and N. P. Papanikolopoulos, "Visual tracking for intelligent vehicle-highway systems," *IEEE Trans. Vehicular Technology*, vol. 45, no. 4, pp. 744-759, Nov. 1996.
- [7] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "Pfinder: real-time tracking of the human body," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, pp. 780-785, July 1997.
- [8] J. Segen, and S. Pingali, "A camera-based system for tracking people in real-time," in *Proceedings 13th International Conference on Pattern Recognition*, 1996, pp.63-67.
- [9] K. Rohr, "Toward model-based recognition of human movements in image sequences," *Computer Vision, Graphics and Image Processing, Image Understanding*, vol.59, no.1, pp.94-115, Jan. 1994.

- [10] D. M. Gavrila, and V. Philomin, “Real-time object detection for ‘smart’ vehicles,” in *The Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 1, 1999, pp. 87-93.
- [11] A. Broggi, M. Bertozzi, A. Fascioli, and M. Sechi, “Shape-based pedestrian detection,” in *Proceedings of the IEEE Intelligent Vehicles Symposium*, 2000, pp. 215-220.
- [12] M. Oren, C. Papageorgiou, P. Sinha, E. Osuna and T. Poggio, “Pedestrian Detection Using Wavelet Templates,” in *Proceedings of Conference on Computer Vision and Pattern Recognition*, 1997.
- [13] C. Papageorgiou, and T. Poggio, “Trainable pedestrian detection,” in *1999 International Conference on Image Processing*, vol.4, 1999, pp. 35-39.
- [14] D. Toth and T. Aach, “Detection and Recognition of Moving Objects using Statistical Motion detection and Fourier Descriptors,” in *12th International Conference on Image Analysis and Processing*, pp. 430-435, 2003.
- [15] B. Heisele and C. Wohler, “Motion-Based Recognition of Pedestrians,” in *International Conference on Pattern Recognition*, vol.2, pp. 1325-1330, 1998.
- [16] H. Mori, N. Charkari and T. Matsushita, “On-Line Vehicle and Pedestrian Detections Based on Sign Pattern,” in *IEEE Trans. On Industrial Electronics*, vol. 41, No.4, 1994.
- [17] S. Yasutomi, and H. Mori, “A method for discriminating of pedestrian based on rhythm,” in *IROS '94. Proceedings of the IEEE/RSJ/GI International Conference on Intelligent Robots and Systems '94. 'Advanced Robotic Systems and the Real World'*, vol. 2, 1994, pp. 988 –995.
- [18] A. Utsumi and N. Tetsutani, “Human Detection using Geometrical Pixel Value Structures,” in *Proceeding of the 5th IEEE international Conference on Automatic Face and Gesture Recognition*, 2002.
- [19] L. D. Stefano and A. Bulgarelli, “A Simple and Efficient Connected Components Labeling Algorithm,” in *International Conference on Image Analysis and*

Processing, 1999.

- [20] Lindsay I Smith, “A tutorial on Principle Components Analysis,” http://csnet.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf
- [21] M. Turk and A. Pentland. “Face Recognition Using Eigenfaces,” in *Proceedings of IEEE Computer Vision and Pattern Recognition*, pp. 586--590, Maui, Hawaii, Dec. 1991.
- [22] M. Turk and A. Pentland, “Eigenfaces for Recognition,” *Journal of Cognitive Neuroscience*, vol.3, no. 1, pp. 71-86, 1991.
- [23] P. Viola and M. Jones, “rapid object detection using a boosted cascade of simple features,” *IEEE Conf. on Computer Vision and Pattern Recognition*, 2001.
- [24] O. Javed, S. Ali and M. Shah, “Online Detection and Classification of Moving Objects Using Progressively Improving Detectors,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol.1, pp. 696-701, 2005.
- [25] R. Culter and L. Davis, “Robust Real-Time Periodic Motion Detection, Analysis, and Applications,” in *IEEE Trans. On Pattern Analysis and Machine Intelligence*, vol. 22, No. 8, 2002.
- [26] P. Viola, M. Jones and D. Snow, “Detecting Pedestrians Using Patterns of Motion and Appearance,” in *Proceedings of the International Conference on Computer Vision*, 2003.
- [27] R. Polana, and R. Nelson, “Detecting activities,” in *1993 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1993, pp. 2-7.
- [28] D. Zhang and G. Lu, “Content-Based Shape Retrieval Using Different Shape Descriptors: A Comparative Study,” *IEEE International Conference on Multimedia and Expo*, 2001.
- [29] 葉怡成，類神經網路模式應用與實作，儒林出版社，2002。
- [30] C. Curio, J. Edelbrunner, T. Kalinke, C. Tzomakas, and W. von Seelen, “Walking

pedestrian recognition," *IEEE Trans. Intelligent Transportation Systems*, vol. 1, issue 3, pp.155-163, Sept. 2000.

- [31] H. Roh, S. Kang and S-W Lee, "Multiple People Tracking Using an Appearance Model Based on Temporal Color," in *15th International Conference on Pattern Recognition (ICPR'00)* - Volume 4, 2000.
- [32] D. Zhang and G. Lu, "A Comparative Study of Fourier Descriptors for Shape Representation and Retrieval," *the 5th Asian Conference on Computer Vision*, Jan. 2002.
- [33] I. Haritaoglu, D. Harwood and S. Davis, "W⁴: Real-Time Surveillance of People and Their Activities," *IEEE Trans. On Pattern and Machine Intelligence*, Vol.22, No.8, Aug. 2000.
- [34] I. Haritaoglu, D. Harwood and S. Davis, "Backpack: Detection of People Carrying Objects Using Silhouettes," *IEEE International Conference on Computer Vision*, pp. 102-107, 1999.
- [35] I. Haritaoglu, D. Harwood and S. Davis, "Hydra: Multiple People Detection and Tracking Using Silhouettes," In Second Workshop of Visual Surveillance at CVPR, pages 6-13, 1999.
- [36] T. Zhao and R. Nevatia, "Tracking Multiple Human in Complex Situations," *IEEE Trans. On Pattern and Machine Intelligence*, vol. 26, No. 9, Sep, 2004.
- [37] S. J. McKenna, S. Jabri, Z. Duric, H. Wechsler and A. Rosenfeld, "Tracking Groups of People," *Comput. Vision Image Understanding*, no. 80, pp. 42--56, 2000.
- [38] Y. Kuno, T. Watanabe, Y. Shimosakoda and S. Nakagawa, "Automated detection of Human for Visual Surveillance System," in *IEEE Proceedings of ICPR*, 1996.
- [39] M. Riesenhuber and T. Poggio, "Models of Object Recognition," in *Nature Neuroscience Supplement*, vol. 3, Nov, 2000.

- [40] I. Sekita, T. Kurita and N. Otsu, “Complex Autoregressive Model for Shape Recognition,” in *IEEE Trans. On Pattern Analysis and Machine Intelligence*, vol. 14, No4, Apr, 1992.
- [41] H. Kauppinen, T. Seppanen and M. Pietikainen, “An Experimental Comparision of Autoregressive and Fourier-Based Descriptors in 2D Shape Classification,” in *IEEE Tran. On Pattern Analysis and Machine Intelligence*, Vol. 17, No. 2, Feb, 1995.
- [42] L. Wang, T. Tan, W. Hu and H. Ning, “Automatic Gait Recognition Based on Statistical Shape Analysis,” in *IEEE Trans. On Image Processing*, vol. 12, No. 9, Sep, 2003.

