

國立交通大學

資訊科學與工程研究所

碩士論文

擁擠環境中的人群計數與追蹤

Human counting and tracking in crowded scene



研究生：陳建榮

指導教授：傅心家 教授

中華民國九十五年七月

擁擠環境中的人群計數與追蹤
Human counting and tracking in crowded scene

研究生：陳建榮

Student : Jian-Rong Chen

指導教授：傅心家 教授

Advisor : Prof. Hsin-Chia Fu

國立交通大學
資訊科學與工程研究所
碩士論文



A Thesis
Submitted to Institute of Computer Science and Engineering
College of Computer Science
National Chiao Tung University
in partial Fulfillment of the Requirements
for the Degree of
Master
in
Computer Science

July 2006

Hsinchu, Taiwan, Republic of China

中華民國九十五年七月

擁擠環境中的人群計數與追蹤

學生：陳建榮

指導教授：傅心家 教授

國立交通大學資訊科學與工程研究所

摘 要

以人工監看方式來進行人潮的管理，不僅費時費力，並且容易發生疏忽。以電腦分析影像來進行自動化的人潮計數與追蹤，除可節省人力的耗損外，更可降低因人為疏忽所造成的危險。但在人潮擁擠的環境下，一般所使用 45 度俯角的攝影方式，會因人在行走交錯時互相遮蔽的效應，而造成人群計數上的困難。為了降低遮蔽效應所造成的影響，本論文採用垂直向下的攝影視角，發展一套可在擁擠環境中進行人群計數與追蹤的方法。

此方法以人頭邊緣灰階梯度方向的放射趨勢為基礎，利用群聚的方式來偵測影像中人頭的位置以及數量，並利用色彩及路徑等特徵，在連續的影像中進行以偵測為基礎的多人移動追蹤。經由實驗發現，本論文的方法不論是在人潮稀疏或擁擠的狀況下，均可達到 80% 以上的正確率。顯示此方法不受遮蔽效應所影響，並能適用於人潮擁擠的環境下。我們將此方法實作為一自動化的人潮監控系統，使其能夠應用在實際環境中。

Human counting and tracking in crowded scene

Student : Jian-Rong Chen

Advisors : Prof. Hsin-Chia Fu

Institute of Computer Science and Engineering
National Chiao Tung University

Abstract

Human monitoring and controlling the crowded situation is not only tedious work but also easy to get mistakes. Automatic head counting and tracking can save the manpower and reduce the chance of human negligence. Because of the occlusion between people, it is difficult to count human by the frontal view. In order to reduce the occlusion effect, we use the overhead view of people to develop a human counting and tracking method in crowded scene.

Based on the radiation of grey-level gradient direction along the human head contour, the method detects human head position in image by clustering. And track multiple people by color and trajectory analysis from the detection results in image sequence. The experimental results presented no matter under sparse or crowded situation, our method can achieve above 80% correction rate. It presents our method doesn't affected by occlusion effect and can be used in crowded scene. We implement our method as an automatic surveillance system and apply it in a real world.

誌 謝

經過反覆的修改以及琢磨，終於將我的論文完成了。雖然比預計的時間要晚了一些，但為了使內容更加的盡善盡美，我想一切都是值得的。兩年的碩士生活雖然辛苦，但也有許多令人難以忘懷的回憶，箇中滋味實在很難用三言兩語來形容，只有自己親身體驗過才會知道。

而這本論文得以順利付梓，首先要感謝的，當然就是我的指導老師傅心家教授。您讓我了解到做研究最重要的事情是必須要有耐心，以及天馬行空般的想像力。並且在您的教導之下，我的簡報製作以及報告技巧方面都有長足的提升。

另外也要感謝實驗室的博士班學長們：永煜、岳宏、柏伸、政龍以及士賢，這兩年來的照顧以及指導，其中又特別感謝政龍學長，在論文的寫作方式以及文句的潤飾上給了我很多的幫助。從你們身上，我學習到許多與以往不同的觀念以及想法，認識你們是我在這個實驗室最大的收穫。

還有同實驗室的好伙伴政邦與富評，一同分享論文寫作過程中的辛苦與快樂。政邦的自信以及對於事物的要求是我前所未見的，是我所學習的目標；而富評也總是很有耐心的回答我各式各樣的數學問題。學弟玉善也總是在實驗室為我們製造歡樂，讓我們能夠暫時的忘卻煩惱。

最後，還要感謝我的母親，讓我能夠衣食無虞；感謝我的兩位哥哥對我的支持與督促；以及感謝我那在天上的父親保佑。當然，還有我那最可愛最漂亮最善良的宇宙無敵超級美少女小甜心，謝謝你一直在我的身邊，陪我度過每一個煩悶與不開心的日子。

目 錄

摘要.....	i
Abstract	ii
誌謝.....	iii
目錄.....	iv
表目錄.....	vi
圖目錄.....	vii
符號說明.....	x
第一章 前言.....	1
1.1 研究動機.....	1
1.2 研究目標.....	2
1.3 章節介紹.....	2
第二章 相關研究.....	3
2.1 以灰階梯度為基礎的人頭偵測.....	3
2.2 移動物體的追蹤.....	5
第三章 以人頭偵測為主的人群計數與追蹤.....	8
3.1 垂直視角的人頭偵測.....	9
3.1.1 前處理：前景偵測.....	9
3.1.2 圓形人頭模型.....	11
3.1.3 梯度源頭的群聚.....	13
3.1.4 隨距離調整的臨界值 T	18
3.2 多人的移動追蹤.....	22
第四章 實驗及結果討論.....	29
4.1 實驗設計.....	29
4.2 人頭偵測.....	30

4.2.1 不同群聚方法的偵測結果.....	30
4.2.2 隨距離調整 T 的偵測結果.....	35
4.2.3 不同亮度環境下的偵測結果.....	39
4.3 多人的移動追蹤.....	40
4.3.1 加入容許偵測遺漏的追蹤結果.....	41
4.3.2 追蹤對於偵測結果的改善.....	43
4.4 處理速度.....	47
第五章 實際應用：自動化的車站人潮監控系統.....	48
5.1 系統架構.....	48
5.2 出入口的人潮分析.....	51
第六章 結論與未來展望.....	53
6.1 結論.....	53
6.2 未來展望.....	54
參考文獻.....	55
附錄 A 影像灰階梯度.....	57

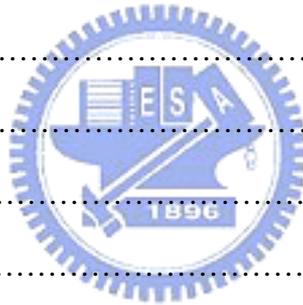


表 目 錄

表 4.1 在各個時段下的不同群聚方式的詳細數值($T=300$).....	34
表 4.2 固定 T 與隨距離調整的 T 在相近 recall 下的比較.....	38
表 4.3 四個鏡頭所拍到影像的亮度資訊.....	39
表 4.4 尖峰時段下四個鏡頭的偵測結果.....	40
表 4.5 不同追蹤方法的連接正確率.....	42
表 4.6 人群移動追蹤結果.....	42
表 4.7 一張影像的平均處理時間.....	47
表 5.1 各個攝影位置的 P_{cr} 值.....	51



圖 目 錄

圖 2.1 橢圓模型的人頭偵測.....	4
圖 2.2 遮蔽效應示意圖.....	5
圖 2.3 以偵測為基礎的移動物體追蹤.....	6
圖 3.1 本論文所提方法的整體流程.....	8
圖 3.2 前景偵測後的結果 (a)原始影像 (b)前景偵測後的影像.....	10
圖 3.3 垂直視角的影像下，人頭邊緣灰階梯度的放射趨勢.....	11
圖 3.4 梯度源頭示意圖.....	12
圖 3.5 理想的梯度源頭集中情況.....	12
圖 3.6 實際的梯度源頭集中情況 (a)灰階梯度向量及其反向 (b)梯度源頭分布圖	13
圖 3.7 梯度源頭的量化問題.....	14
圖 3.8 以不同方式所找出的放射點 (a)只考慮單一點 (b)考慮 3×3 區域.....	14
圖 3.9 固定區域的群聚.....	15
圖 3.10 固定區域式群聚的結果 (a)原始放射點分布 (b)群聚結果.....	15
圖 3.11 固定區域的群聚無法正確群聚的情況 (a)原始放射點分布 (b)群聚結果...	16
圖 3.12 Mean-shift 演算法的示意圖.....	17
圖 3.13 兩種群聚方式結果的比較 (a 固定區域式群聚的結果 (b)Mean-shift 群聚 方式的結果.....	18
圖 3.14 攝影機與人在三度空間的關係.....	19
圖 3.15 攝影機與人頭的座標關係圖.....	19
圖 3.16 人與影像中心的距離與離心率 e 的關係圖 (a) h 與 e 的關係圖 (b) h 與 $1-e$ 的關係圖.....	21
圖 3.17 移動路徑的平滑度分析.....	23

圖 3.18 肩膀部位與移動方向的三度空間關係圖.....	24
圖 3.19 (a) t-1 時刻的偵測結果 (b)t 時刻的偵測結果與可能連接 (c)t 時刻的前景 影像與可能連接.....	27
圖 3.20 多人移動追蹤的結果 (a) t-1 時刻 (b) t 時刻.....	28
圖 4.1 不同群聚方式在離峰時段的 Precision 比較.....	31
圖 4.2 不同群聚方式在離峰時段的 Recall 比較.....	31
圖 4.3 不同群聚方式在離峰時段的 Performance 比較.....	32
圖 4.4 不同群聚方式在尖峰時段的 Precision 比較.....	32
圖 4.5 不同群聚方式在尖峰時段的 Recall 比較.....	33
圖 4.6 不同群聚方式在尖峰時段的 Performance 比較.....	33
圖 4.7 隨距離調整的 T 與固定 T 在離峰時段的 Precision 比較.....	35
圖 4.8 隨距離調整的 T 與固定 T 在離峰時段的 Recall 比較.....	36
圖 4.9 隨距離調整的 T 與固定 T 在離峰時段的 Performance 比較.....	36
圖 4.10 隨距離調整的 T 與固定 T 在尖峰時段的 Precision 比較.....	37
圖 4.11 隨距離調整的 T 與固定 T 在尖峰時段的 Recall 比較.....	37
圖 4.12 隨距離調整的 T 與固定 T 在尖峰時段的 Performance 比較.....	38
圖 4.13 各個鏡頭所拍攝到的影像，a、b、c、d 依序為鏡頭 1~4.....	39
圖 4.14 在離峰時段下加入追蹤與未加入追蹤的偵測 Precision 比較.....	43
圖 4.15 在離峰時段下加入追蹤與未加入追蹤的偵測 Recall 比較.....	44
圖 4.16 在離峰時段下加入追蹤與未加入追蹤的 Performance 比較.....	44
圖 4.17 在尖峰時段下加入追蹤與未加入追蹤的偵測 Precision 比較.....	45
圖 4.18 在尖峰時段下加入追蹤與未加入追蹤的 Recall 比較.....	45
圖 4.19 在尖峰時段下加入追蹤與未加入追蹤的 Performance 比較.....	46
圖 5.1 自動化人潮監控系統架構圖.....	48
圖 5.2 人潮擁擠程度的判斷流程圖.....	50
圖 5.3 進出口人潮分析示意圖.....	51

圖 A.1 (a)原始灰階影像 (b)X 軸上的灰階值分佈 (c)X 軸上的灰階值一階微分.....57

圖 A.2 柏威運算元 (a)水平方向運算元 (b)垂直方向運算元.....58

圖 A.3 索伯運算元 (a)水平方向運算元 (b)垂直方向運算元.....59

圖 A.4 (a)原始影像 (b)水平方向的灰階梯度值(c)垂直方向的灰階梯度值 (d)完整的灰階梯度值.....60



符 號 說 明

B：前景偵測中的背景影像

P：前景偵測中正在處理的影像

R：影像上的人頭半徑

$r(i, j)$ ：在點(i,j)的 3×3 鄰近區域中所包含的梯度源頭個數

λ ：梯度源頭個數的臨界值，大於此值的點即為放射點

T：群內的放射點個數臨界值，大於此值的群即認為是人頭

e：橢圓的離心率

T_L ：隨距離調整的T值中，最低的T值

T_H ：隨距離調整的T值中，最高的T值

H、h：影像上的人頭偵測結果

$smooth_t$ ：移動路徑的平滑度

$color_t$ ：頭髮與衣著色彩相似度

$cover_t$ ：肩膀區域的前景涵蓋率

$P(h_{t-1}, h_t)$ ：偵測結果之間的連接機率

w_{smo} ：移動路徑平滑度在連接機率中的權重

w_{color} ：頭髮與衣著色彩相似度在連接機率中的權重

w_{cover} ：肩膀區域的前景涵蓋率在連接機率中的權重

P_{cr} ：固定區域內人數的臨界值，人數超過此值則認為是擁擠

X：累計擁擠次數的臨界值，累計超過此次數則須通知管理人員

第一章 前言

1.1 研究動機

在人群聚集的公共場合，如車站、機場、賣場與百貨公司等，經常容易出現人潮擁擠的狀況。所謂擁擠，指的是在單位面積內的人數，超過了單位面積內可容納的人數(一般而言為 4 人/平方公尺)，而造成人與人之間的擠壓，使人的身體產生不舒適感。

而在擁擠的狀況之下，因為人群的互相擠壓，就可能造成意外的發生。如車站月台或是電扶梯等，都是非常容易發生意外的地點。是故，此類人群聚集的公共場合，必定需要一套有效的人潮管理方式，來避免意外的發生。

傳統的管理方式通常為指派專任的管理人員，負責區域內人潮的管理與疏散，以防止因人潮過度擁擠所造成的危險。但以人工的方式來管理，除無法顧及車站全面性的安全之外，管理的效率及功能也有限，並有人力的調派與成本上的問題。為了能夠更加有效以及全面性的管理，並減少人力成本，故有監視系統的發明。在欲管理的區域各處與死角設置數個攝影機，將影像訊號傳送至某個定點，並在定點派駐人員監看各攝影機所在地的狀況，即為最普遍的監視系統建置方法。但全天候的監看不僅費時費力，監看人員的注意力也必須相當集中，以便應付隨時有可能發生的危險狀況。有鑑於此，在資訊科技發達的今天，自動化監視系統的想法就應運而生了。

如何以電腦來取代人工，來分析監視攝影機所拍攝到的影像，並且估計人潮的數量以及追蹤其移動，是近年來國內外熱門的研究課題[1-10,12,14]。其中又包括前景(foreground)的偵測、人潮數量的計算、人群所在位置的偵測、人群移動時的追蹤(tracking)等等，均是困難且極具挑戰性的題目。尤其在人潮擁擠的環境之下，人在行走交錯之間會產生互相遮蔽(occlusion)的情形，更增加了人群計

數與追蹤的困難度。

故我們以車站月台的人潮為研究對象，希望透過影像處理的技術，發展一套在人潮擁擠的環境之下，仍可快速且準確的計算影像中所包含人數的方法，並且可透過多張連續影像的資訊，追蹤環境中人群的移動軌跡。

1.2 研究目標

人頭為一個人最為明顯且辨別度高的特徵，故在國內外的研究中[2-8,10]，常使用人頭來做為決定人所在位置的標的。在本論文中，我們將提出一種以人頭偵測(head detection)為主的人群計數與追蹤方法，並具有下列特色：

1. 能夠在複雜且擁擠的環境下，準確的偵測人(即人頭)的所在位置。
2. 能夠抵抗光源的變化，在不同強弱的光源下均可準確偵測。
3. 能夠同時追蹤多人的移動，記錄及分析其移動軌跡。

更進一步的，希望能夠將此方法應用至真實環境中，如在車站、機場、或是百貨公司等地點，建立一自動化的人潮監控系統，使人潮的管理能夠更加的有效率。

1.3 章節介紹

在第二章中，我們將介紹近年來在人群計數與追蹤的相關研究，包括人頭偵測以及移動物體追蹤兩大方面。而在第三章中說明本論文所提出的方法，同樣是分為人頭偵測與人群移動的追蹤兩部分來介紹。第四章則為我們為了測試本方法的效率所進行的實驗，其中包括實驗環境的設定以及實驗結果的討論。第五章為本方法在實際環境中的應用。第六章則為結論與未來的展望。

第二章 相關研究

本章分為兩大部份，第一部分先介紹有關影像處理的背景知識，以及如何將這些技術應用至人群計數及人頭偵測方面。第二部份再介紹連續影像中，移動物體的追蹤的相關研究。

2.1 以灰階梯度為基礎的人頭偵測

如何偵測視訊或影像中人的位置，以及計算人的數量，是近年來國內外熱門的研究課題。其中人頭為最明顯且辨認性高的特徵，故常用來當作人所在位置的標的。底下我們就介紹幾種常見的人頭偵測方法，以及目前人頭偵測方法所遇到的問題。

在人頭偵測方面，常見的做法為利用頭髮及人臉顏色[1,4,10]與利用人頭的形狀[2,3,4,5,6,7,8,9]兩種方式。前者的做法相當直觀，先收集頭髮與人臉的顏色資訊，作為訓練資料建立起頭髮及皮膚的色彩模型(color model)[4,10]。並利用預先建立好的色彩模型，對於影像中的像素進行色彩相似度(likelihood)的計算，以色彩相似度高的地方認為是人頭。對這樣的做法來說，如何建立一個好的色彩模型是最重要的課題，色彩模型的好壞直接影響著偵測的準確度。故為了得到較好的色彩資訊，此方法對於環境的亮度非常敏感，若亮度不足，色彩的資訊不夠時，將會影響到偵測的效果。

另一方面，由於此方法必須建立在影像中的人具有相同的膚色及髮色的假設之上，若有髮色或膚色與預先設定的色彩模型不同的人出現，即無法偵測出來。雖然可以建立多個不同髮色及膚色的模型來解決，但同時也增加了錯誤偵測的機率。並且對於人臉朝向不同的角度時，偵測的效果也會受到影響。如背對攝影鏡頭時，由於沒有人臉膚色的資訊，對於偵測來說將相當的不利。

而利用人頭邊緣的形狀當作特徵來進行偵測的方法，則是利用在 45 度俯角的攝影視角下，人頭的形狀通常為一直立的橢圓形(ellipse)[2,3,7]作為辨認人頭的特徵。做法為先利用灰階梯度(grey-level gradient，詳見附錄 A)或其他的邊緣偵測方式，尋找出物件的邊緣。再與一橢圓形的人頭模型進行比較，找出影像中邊緣形狀接近橢圓的區域，即認為是所謂的人頭。與利用顏色資訊的不同之處為，人頭形狀的資訊不受髮色與膚色的影響(如圖 2.1)。並且不論人臉面向哪一個方向，人頭的形狀均是近似於橢圓形的，故也沒有人臉角度上的問題。除了橢圓的人頭模型之外，另外還有將肩膀的上緣一起考慮進去的 Ω (omega)型人頭模型[8]。



圖 2.1 橢圓模型的人頭偵測

(引用自[3]，原始MPEG影片可於<http://vision.stanford.edu/~burch/>取得)

以形狀做為特徵來偵測人頭，雖然有不受人臉面向影響的優點，但卻容易因為相似形狀的物體，造成錯誤的偵測。或是因為附近背景與人頭的顏色相近，灰階梯度所形成邊緣不夠明顯，導致無法符合預先設定的人頭模型，而沒有準確的偵測出來。雖然可以配合色彩的資訊來加以改善[3,4]，但是色彩模型對於不同髮色與膚色的問題仍然存在，所能增進的效果有限。

不論利用色彩或是形狀的資訊，由於攝影機角度的關係，在不同的位置上的人，在影像上將呈現出不同的大小。故如何對於不同大小的人頭，均能夠準確偵測，是一個重要的課題。

另一個困難的問題是，在以單一攝影機進行拍攝時，位置距離攝影機較遠的人，將會被位置距離攝影機較近的人所遮蔽，使其無法完整的出現在影像中，稱為遮蔽效應(occlusion effect)。圖 2.2 為一個遮蔽效應的示意圖，A、B 為此空間中的兩個人，圖 2.2(a)為這兩個人空間中的位置，而圖 2.2(b)為攝影機以 45

度俯角所拍攝到的影像。由於距離攝影機的遠近不同，在影像中 B 的某個部份被 A 所遮蔽。遮蔽效應增加了人群計數與追蹤的困難性。若是少部分的遮蔽，也許還有機會偵測出來，但若大部分，甚至是完全的遮蔽，就幾乎是無法偵測出來了。而在像車站這種擁擠的環境中，完全的遮蔽勢必是無法避免的。

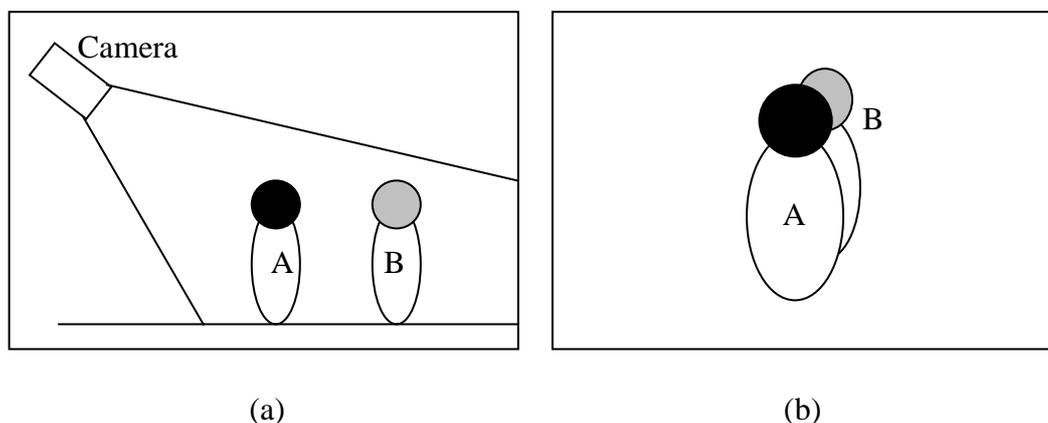


圖 2.2 遮蔽效應示意圖

(a) 空間中攝影機與人的所在位置 (b)以 45 度俯角所拍攝到的影像

對於遮蔽效應，目前尚未有一個很好的解決方式。可能的解決方式，就是將攝影機的角度提高，來減少遮蔽效應的發生[11]。若將攝影機以垂直地面的方式進行攝影，則在一般情況下(空中無遮蔽物時)，均可有效的避免掉人頭被遮蔽的機會。

2.2 移動物體的追蹤

移動物體的追蹤(tracking)主要分成兩大部分。一為利用感測器(sensor)等方式，追蹤物體在 3D 空間的運動。另一即為利用一連串連續的影像中，追蹤物體在影像中的位置。而在本節及本論文中所討論的追蹤即為後者。

一個最簡單的追蹤機制為，選定一個欲追蹤的物體，紀錄此物體之特徵(如形狀、色彩等)，在下一張影像中，以這些特徵來追蹤此物體在影像中的位置。這樣的追蹤機制僅限於特定物體的追蹤，並且欲追蹤的物體個數為已知[1,2,3,4]。

但在許多情況，我們並無法事先得知物體的初始位置，也無法得知物體的個數有多少。在這種情況下，即必須先透過偵測的方式，來取得物體的初始位置以及個數，才有辦法對物體進行追蹤。由於追蹤的目標是經由偵測而來，故稱為以偵測為基礎的追蹤方式(detection-based tracking)[12]。

以偵測為基礎的追蹤機制如下所述：對於每一張影像，均當作獨立的影像來進行偵測。當偵測到欲追蹤的物體出現在影像中時，追蹤隨即開始。經由分析過去與現在的偵測結果之間的關係，來追蹤目標物在影像中的移動。這樣的追蹤機制較適合於場景中同時有多個物體在移動，並且個數及位置均為未知的狀況之下，如我們先前所提到的車站或機場等環境。

物體的追蹤與偵測的關係是相輔相成的。透過追蹤的結果，可以驗證偵測結果的正確性。例如一偵測出的物體，若無法在之後的偵測結果中找到相近的物體，即有可能為一錯誤的偵測。或是在某一張影像中，一個物體未被偵測出來，但經由下一張影像的偵測結果，此物體仍可正確的被追蹤。圖 2.3 為一個以偵測為基礎的移動物體追蹤示意圖。

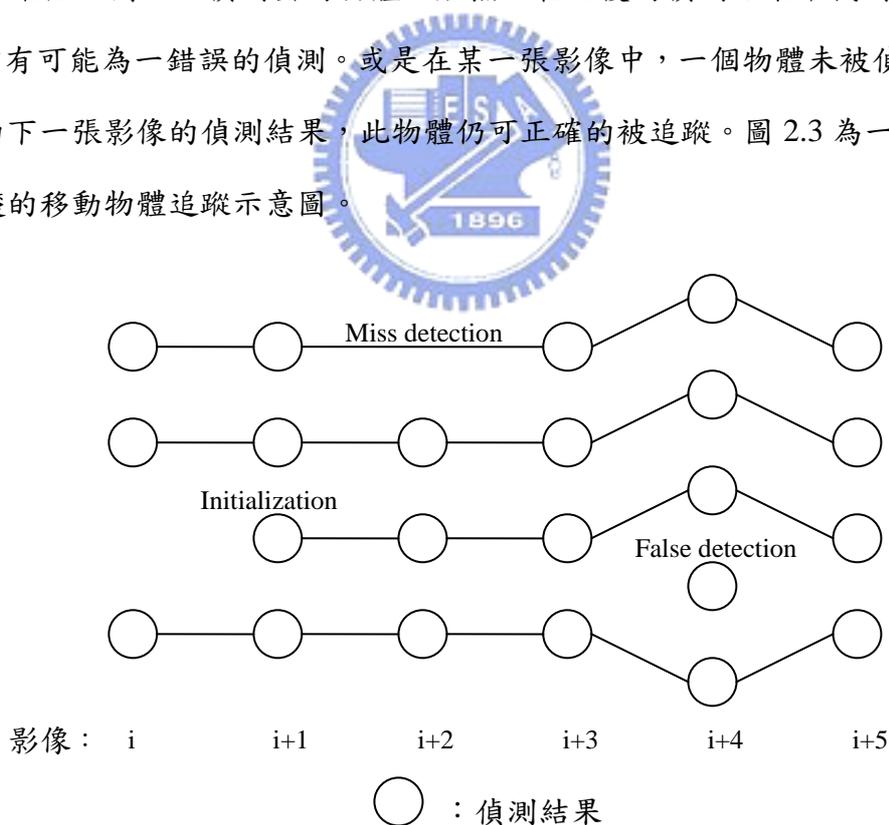


圖 2.3 以偵測為基礎的移動物體追蹤

在移動物體的追蹤中，最常也最廣泛被利用的即為色彩的資訊[1,3,4]。其方式為將欲追蹤的物體的色彩資訊記錄下來，在下一張影像中尋找色彩最為接近的區域或是物體。以人的追蹤來說，就是利用其衣著的色彩。另外物體的外觀形狀與大小[1,2,3]，也是經常用來當作追蹤的依據，通常使用在物體的形狀改變並不劇烈的時候。但在多人的環境中，可能會有兩人穿著相似色彩的衣服，故衣著色彩的資訊並不足夠。同時人在移動時，形狀的改變也相當大，其大小也因為離攝影機的遠近而並非固定。由此可知，人的追蹤與其他物體相比，更加的複雜且困難。

在色彩與形狀的資訊之外，通常也會在加入移動路徑(trajjectory)的分析[12]，若物體的移動路徑不平滑(smooth)或是出現不合理的移動(如瞬間的移動距離過大)，則此路徑即有可能為一個錯誤的追蹤。在移動物體的追蹤中，移動路徑的分析是一個非常重要的動作，其提供了一個在物體本身的特徵之外，客觀的驗證追蹤正確性的方式。



第三章 以人頭偵測為主的人群計數與追蹤

本章將說明本論文所提出的人群計數方法。我們採用垂直視角的方式進行攝影，並以人頭做為計數基準，透過灰階梯度的方向偵測所拍攝到的影像中人的數量以及所在位置。再以偵測的結果為基礎，進一步分析來追蹤人群的移動。本方法的整體流程如圖 3.1 所示，從攝影機擷取影像後，先對影像進行前景偵測的前處理，再針對所偵測出的前景部份，進行人頭的偵測。其中人頭偵測又分為三個步驟，首先計算影像中灰階梯度的方向以及梯度源頭，之後再由梯度源頭中找出較具有參考價值的點，稱為放射點，最後經由群聚(clustering)放射點的方式，找到可能的人頭所在位置。得到人頭偵測的結果之後，再與前一時刻的偵測結果進行比較，連接相似的偵測結果來達到追蹤的目的。最後將目前的偵測結果儲存起來，以供下一時刻的追蹤使用。接下來就針對圖 3.1 的流程中，各個步驟進行詳細的說明。

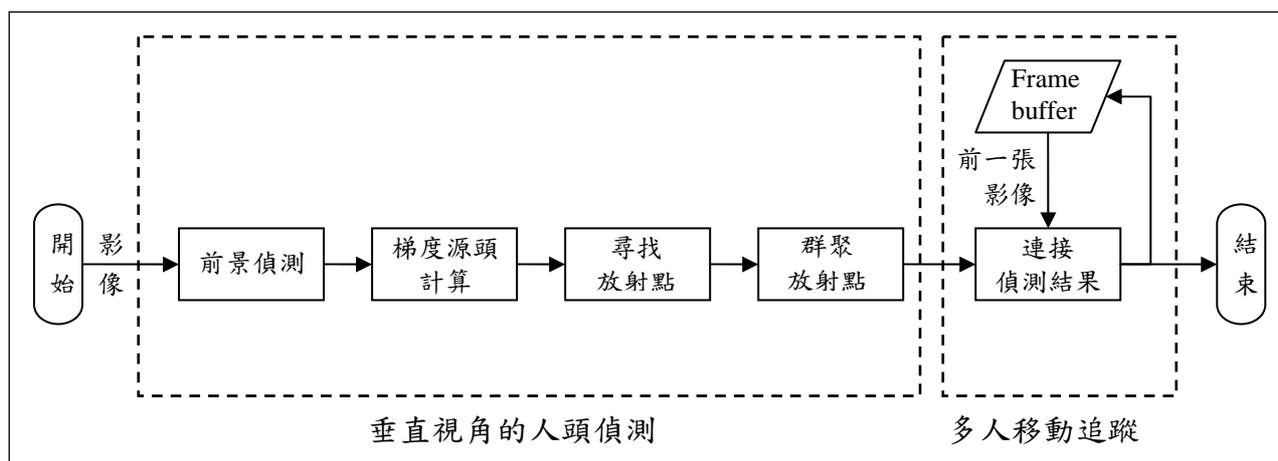


圖 3.1 本論文所提方法的整體流程

3.1 垂直視角的人頭偵測

如同第二章所提到的，在先前的研究中[2,3,4,7,10]，多採用的是 45 度的攝影視角，但在擁擠的環境中，這樣的拍攝方式將產生嚴重的遮蔽效應。為了在擁擠的環境中，依然能夠進行人群的計數與偵測，我們將攝影機架設在天花板，將鏡頭以垂直向下的角度拍攝，藉以降低因為擁擠所產生的遮蔽效應。這樣的視角我們將其稱為垂直視角(overhead-view,[11])。

3.1.1 前處理：前景偵測

前景偵測的目的為去除影像中固定的背景部份，只留下移動的物體，也就是所謂的前景部分[15]。而在本論文中，移動的物體主要就是人。進行前景偵測的優點為，減少背景部份所造成的雜訊影響，並且縮小處理的範圍，加快處理的速度。

為了速度上的考量，我們採用較為簡單的前景偵測方法，稱為背景減除(background subtraction)。我們預先準備一張僅含背景の影像 B，並將要處理の影像 P 的像素，與 B 同位置的像素進行顏色的比較。若同位置的像素顏色相近，則將此在 P 上的像素認為是背景，反之則為前景。

為了降低影像雜訊的影響，我們以 4x4 個像素所組成的區塊為單位，來進行顏色的比較，取代單一像素的顏色比較。並且，為了抵抗光影變化所造成的顏色改變，我們隨著時間對 B 進行調整。調整的方式為利用 P 中被認為是背景的區塊，與 B 中同位置的區塊，以一固定的權重(weight)加成，當作下一時刻的 B 中此區塊的色彩。而其他被認為是前景部分的區塊，則使用前一時刻 B 的色彩。整個背景減除的演算法可以式(3.1)表示。

其中，我們所使用色彩空間為 CIELab 色彩空間，顏色的比較方式則採行 CMC(Color Measurement Committee)於 2000 年所提出的色差(color difference)計算方式。

for 4×4 pixel as a block

$$\begin{cases} P_i(t) \in \text{Background} & \text{if } \text{Diff}(B_i(t), P_i(t)) < \text{Threshold} \\ P_i(t) \notin \text{Background} & \text{Otherwise} \end{cases}$$

Updating

$$\begin{cases} B_i(t+1) = (1 - \alpha)B_i(t) + \alpha P_i(t) & \text{if } P_i(t) \in \text{Background} \\ B_i(t+1) = B_i(t) & \text{if } P_i(t) \notin \text{Background} \end{cases}$$

$B_i(t)$: Background block i at time t

$P_i(t)$: Image block i at time t

α : Learning rate

(3.1)

經過前景偵測後的結果如圖 3.2 所示，圖中白色是判定為前景的部份，黑色則為背景的部份。從圖 3.2 中可以看得出來對於較淺的影子有不錯的去除效果，但對於某些過暗的影子仍然無法去除。然而就人頭偵測來說，如何降低資訊的流失則較為重要，若為了去除較暗的陰影部份，而將顏色比較的臨界值(threshold)提高，可能會造成部分屬於人的區塊也被去除，反而不利於人頭的偵測。

事實上，前景偵測的做法非常多，也各有其優劣。有關前景偵測的相關討論，在[16]中有更詳細的探討。



(a)

(b)

圖 3.2 前景偵測後的結果

(a)原始影像 (b)前景偵測後的影像

3.1.2 圓形人頭模型

在垂直視角的影像中，人頭的形狀通常近似於圓形。由於灰階梯度的方向，具有垂直於邊緣方向的特性(詳見附錄 A)，假設每個人頭的形狀均為正圓形，並且頭髮顏色為深色的話，那麼人頭邊緣的灰階梯度方向，將呈現由頭中心(即圓形的圓心)向外放射的趨勢。圖 3.3 中所表示即為一垂直視角下的人。其中黑色的圓形為人頭，白色的橢圓則為此人的肩膀，箭頭則為其灰階梯度的方向。由圖 3.3 可以很明顯的看出人頭邊緣灰階梯度的放射趨勢。

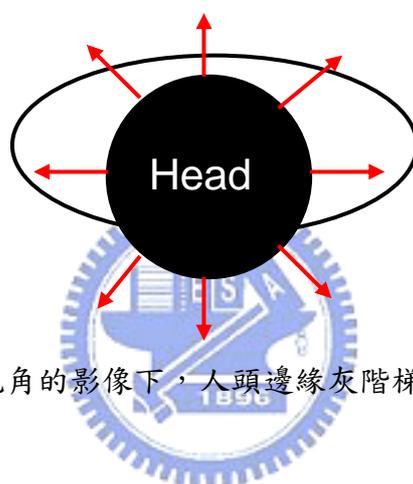


圖 3.3 垂直視角的影像下，人頭邊緣灰階梯度的放射趨勢

因此，若我們能夠找出影像中灰階梯度具有放射趨勢的區域，即可能為人頭的所在位置。那我們該如何找出這樣的趨勢呢？有研究使用樣本比對(template matching)的方式[2,3,5,6,10]，來找出與固定樣本近似的形狀或是特徵。但在垂直視角的影像中，人頭形狀的差異性相當大。不同髮型的人，其頭髮所形成的深色區塊形狀也就不一樣，即使是同一個人，站在不同位置，所形成的人頭形狀也不一樣。在這樣的情況之下，我們很難找到一個可以滿足所有形狀的樣本。若以多個不同樣本來搜尋，速度上又較為緩慢。

從實際的角度來看，雖然人頭的形狀不為正圓，但仍然是一個近似圓形的深色區塊，其邊緣的灰階梯度方向仍然是向外放射的。若我們沿著灰階梯度的方向，反向延伸一固定長度 R (約影像上一個人頭的半徑)，所到達的點我們稱為此梯度的源頭(gradient origin)，其意義為放射出此灰階梯度向量的起點。圖 3.4 說

明了如何找到梯度源頭。

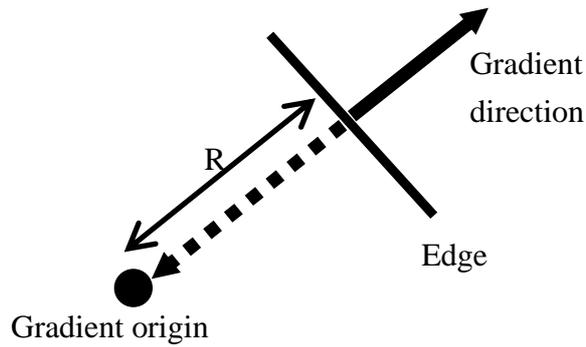


圖 3.4 梯度源頭示意圖

假設人頭為一正圓形，則人頭邊緣的梯度源頭將會集中在圓心，如圖 3.5 所示。但人頭並非為一正圓形，且其大小也各異，故梯度源頭並不會完美的集中在某一點，而是呈現一區域式的集中。圖 3.6 為一實際的梯度源頭集中情況，顏色越淺的點代表其梯度源頭點數越多。故如何找出此一區域式集中的趨勢，為本論文所討論的重點。

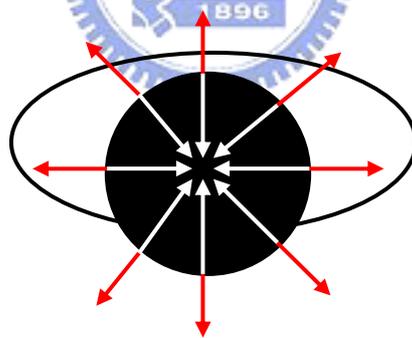


圖 3.5 理想的梯度源頭集中情況

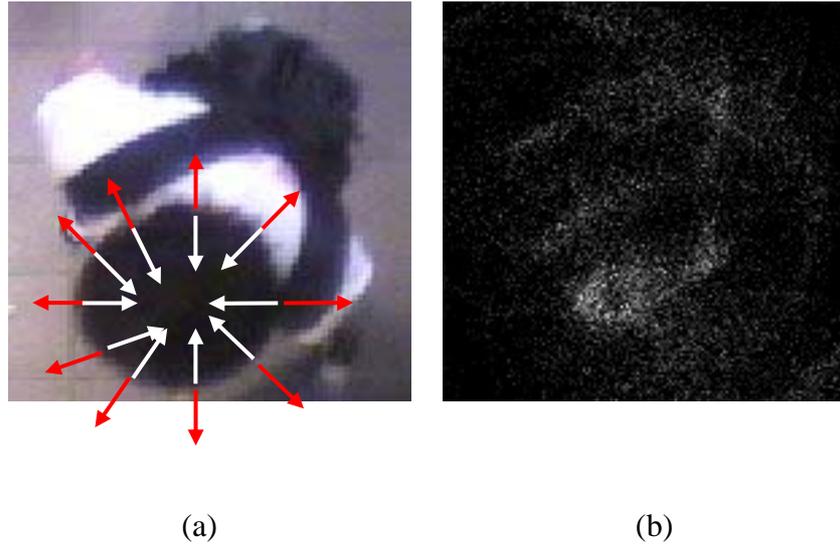


圖 3.6 實際的梯度源頭集中情況

(a)灰階梯度向量及其反向 (b)梯度源頭分布圖

3.1.3 梯度源頭的群聚

若把影像上的梯度源頭，視為一個在二維空間中的資料分布(distribution)。那麼，欲找出梯度源頭集中的區域，即可視為一未知群數的群聚(clustering)問題。故我們可以利用群聚的方式，將離散的梯度源頭加以分群，並藉由計算每一群內的梯度源頭個數，來判斷其是否為源頭。但由於梯度源頭的分布相當複雜，在群聚上相當的困難，故我們在進行群聚之前，必須去除掉較不具參考價值的點，相當於消除雜訊(de-noise)的動作。

而所謂較具參考價值的點，即是放射出較多梯度向量的梯度源頭，我們將其稱為放射點(radiate point)。我們對每一個點，計算其鄰近 3×3 區域中的梯度源頭個數，如式(3.2)的 $r(i, j)$ 。

$$r(i, j) = \sum_{p=-1}^1 \sum_{q=-1}^1 o(i+p, j+q)$$

o : gradient origin

(3.2)

若 $r(i, j)$ 的值大於臨界值 λ ，則將點 (i, j) 判定為一放射點，不足的則當作雜訊去除。之所以使用鄰近 3×3 區域的原因是，由於影像的像素是離散的，經由灰階梯度方向所計算出的梯度源頭，未必會落在影像的像素上，故須將計算出的梯度源頭量化(quantize)至某一像素。圖 3.7 說明了這個情況，圖中白色的點 p 為經由計算得到的梯度源頭位置，其落在 a、b、c、d 四個點的中間。若將 p 量化至鄰近四點中某一點看待，如此即造成了其他三點的誤差，但若考慮 a 的鄰近 3×3 區域，則 p 不論量化至四點中任何一點，均會被包含在此 3×3 區域內。如此便不會有所謂量化的問題，並且更能夠找出具代表性的放射點。以圖 3.6(b) 的梯度源頭分布為例，圖 3.8 為以兩種方式所找出的放射點分布。

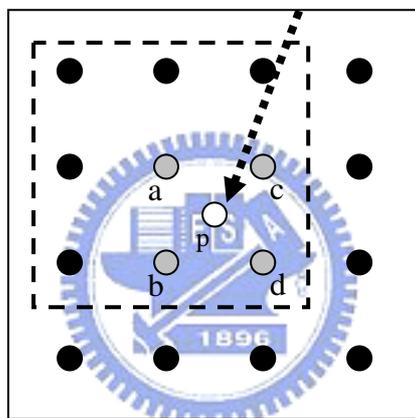


圖 3.7 梯度源頭的量化問題



(a)

(b)

圖 3.8 以不同方式所找出的放射點

(a) 只考慮單一像素 (b) 考慮 3×3 區域

(原始梯度源頭分布請參考圖 3.6(b))

經由以上步驟來找出放射點後，可以看到放射點的分布較原始梯度源頭的分布單純許多，並且在人頭的部份，仍然具有區域式集中的趨勢。故我們以針對放射點的群聚，來取代直接對梯度源頭進行群聚。直覺的群聚方法是，以相鄰的放射點中，梯度源頭個數最多的一點為中心，以人頭半徑 R 為半徑的區域內，所包含的放射點則視為同一群。此群聚方式我們稱為固定區域的群聚 (Fixed-region clustering)，圖 3.9 為此一群聚方式的示意圖，3.10(b) 則為圖 3.10(a) 中的放射點群聚結果。

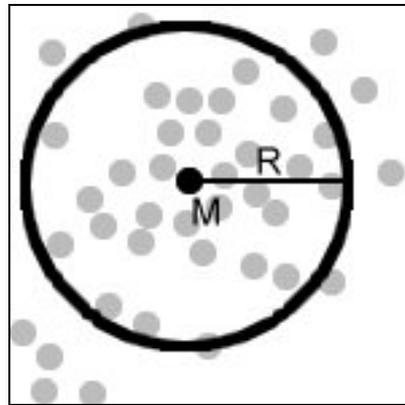


圖 3.9 固定區域的群聚



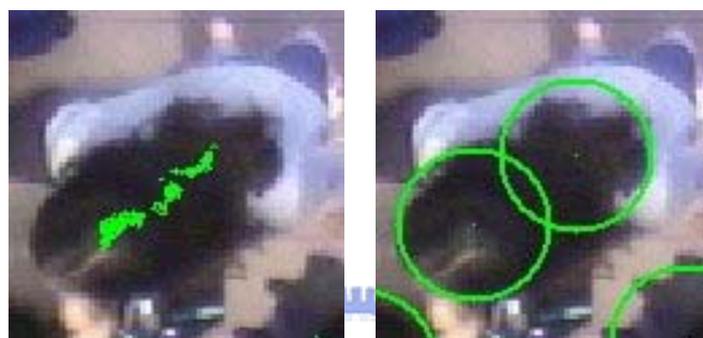
(a)

(b)

圖 3.10 固定區域式群聚的結果

(a)原始放射點分布 (b)群聚結果

但是，這樣的群聚方式只適用於放射點分布近似於圓形，也就是形狀近似於正圓形的人頭。但是，在垂直視角下，並非所有人頭均為正圓形。如圖 3.11(a) 中的人，其髮型為長髮，人頭的形狀則近似於橢圓，其所形成的放射點呈現一長條狀的分布。這樣的情況之下，固定區域的群聚方式，除非使用較大的人頭半徑，否則無法完美的包含屬於此人頭的所有放射點。所群聚出的結果如圖 3.11(b) 所示，放射點被分成兩群，如此一來即有可能因為群內放射點的點數不夠多，不被認為是人頭，造成偵測上的遺漏。



(a) (b)

圖 3.11 固定區域的群聚無法正確群聚的情況

(a)原始放射點分布 (b)群聚結果

針對這種情況，我們使用 Mean-shift 的群聚方式(Mean-shift clustering, [13,14])來解決這個問題。Mean-shift 的群聚演算法為，以任一點為中心，計算其固定區域內的放射點座標的 Mean。在以此 Mean 為中心，重複計算其固定半徑 R 區域內的放射點座標之 Mean，直到 Mean 不再移動，或移動距離小於一臨界值時停止。Mean-shift 演算法的過程可描述如下：

Mean – shift algorithm :

1. For a initial mean M

2. Compute M'

$$M' = \frac{1}{N} \sum_{i=1}^N r_i \quad \text{if } dis(r_i, M) \leq R$$

3. if $dis(M', M) < \varepsilon$, stop

Otherwise, $M = M'$, go to step 2

(3.3)

其中， M, M', r_i 均為空間中的點， R 為用以群聚的區域半徑， ε 為 Mean 移動距離的臨界值。以本論文的情況來說， r_i 即為影像上的放射點， R 則是人頭半徑。圖 3.12 為 Mean-shift 演算法的示意圖。經由 Mean-shift 演算法，我們可以找到此點的鄰近區域中，放射點密度最密集的位置，即所謂的區域最佳解(local optimum)。有關 Mean-shift 演算法的相關證明可以參考[13]。

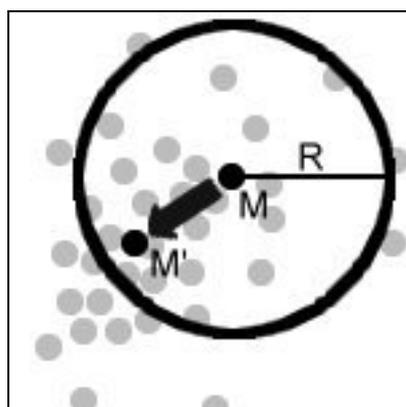


圖 3.12 Mean-shift 演算法的示意圖

雖然 Mean-shift 演算法所找出的只是區域最佳解，但由群聚的觀點來看，此區域最佳解卻正是我們想要的。經由 Mean-shift 演算法所找出的區域最佳解，我們將其視為包含此點的群的中心點。我們對每一個放射點進行 Mean-shift 演算法，並將最後停止在同一點上的放射點視為同一群，以達到群聚的目的。

採用 Mean-shift 演算法進行群聚的好處有二：

1. 對於非正圓形的人頭(如橢圓)，亦可準確的群聚其放射點：Mean-shift 演算法不會受到資料分布形狀的影響，不會發生如固定區域式群聚中，無法正確群聚長條形資料分布的情況。
2. 可以更正確的找到人頭的正確位置：在固定區域的群聚中，中心的決定僅止於區域內放射點所包含的梯度源頭個數最高的位置，但此位置未必為真正人頭的中心點。而 Mean-shift 的群聚方式則為找到放射點密度最高的位置，其位置能夠更接近人頭真正的中心位置。

圖 3.13 為固定區域式群聚與 Mean-shift 的群聚方式所得到結果的比較，可以很明顯的看出 Mean-shift 的群聚方式將放射點群聚為同一群，而不像固定區域式群聚將放射點分割為兩群。

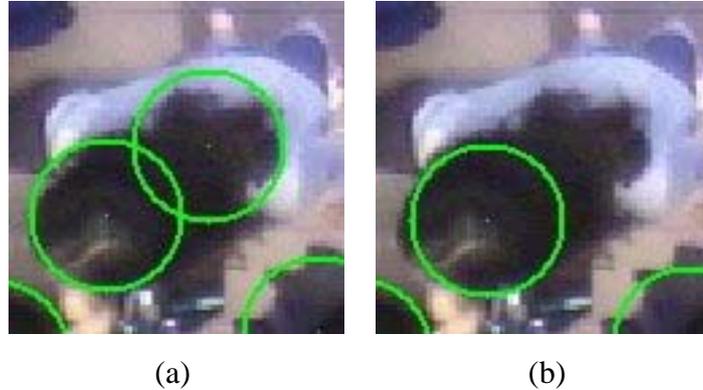


圖 3.13 兩種群聚方式結果的比較

(a)固定區域式群聚的結果 (b)Mean-shift 群聚方式的結果

最後我們定義一放射點個數臨界值 T ，來與每一群所包含的放射點個數做比較，若此群的放射點個數大於 T ，即認為是人頭，小於 T 則不是。



3.1.4 隨距離調整的臨界值 T

對於人頭偵測來說，放射點個數臨界值 T 的決定是一個重要的問題。而思考如何決定 T 的過程中，我們觀察到一個現象，那就是通常當一個人頭越靠近影像的中央時，同一群內的放射點個數也越多。我們猜測其原因為當人的位置越靠近攝影機時，在影像上所形成的形狀會越接近圓形，此時期梯度源頭也會越集中，放射點群聚的效果也會越好。但為了不遺漏掉靠近邊緣的人頭，我們必須使用較低的 T 來將其偵測出來，但如此一來卻也增加了錯誤偵測的機率。故若我們可以找出其集中情況與人所在位置的關係的話，我們就可以依照這個關係來調整(adapting) T 值的大小，以期能達到較低錯誤偵測的效果。

若我們把實際上的人頭形狀看成一直立的橢圓 C ，則在垂直視角下，其在影像上所形成的形狀也會為一橢圓 C' 。而由攝影機的位置做 C 的兩條切線，則

這兩條切線與地板相交兩點，此兩點間的長度即為橢圓 C' 的長軸長度，即如圖 3.14 所示。

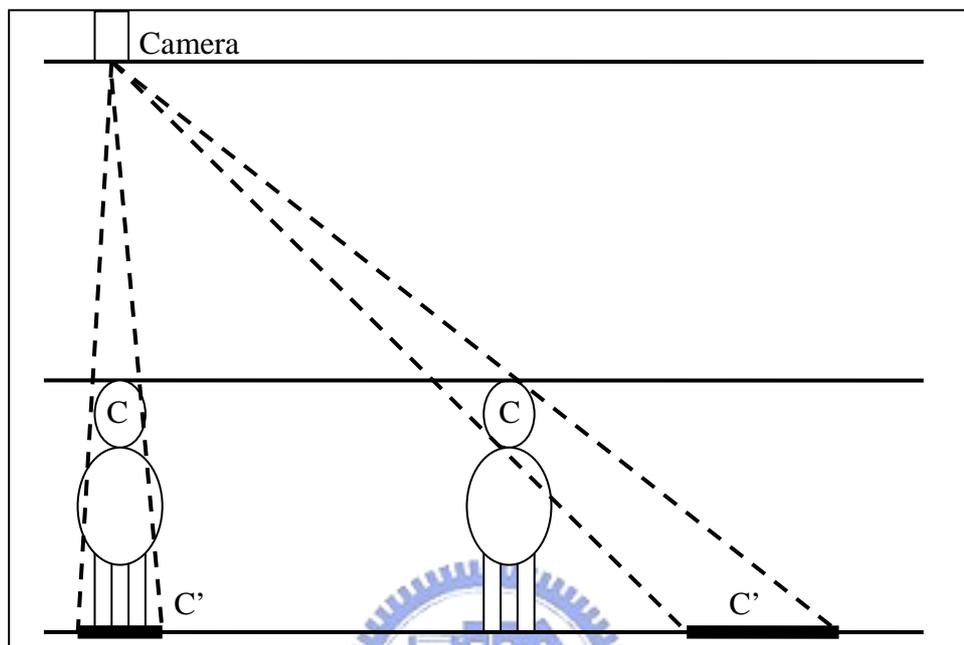


圖 3.14 人的位置與其所形成影像的關係

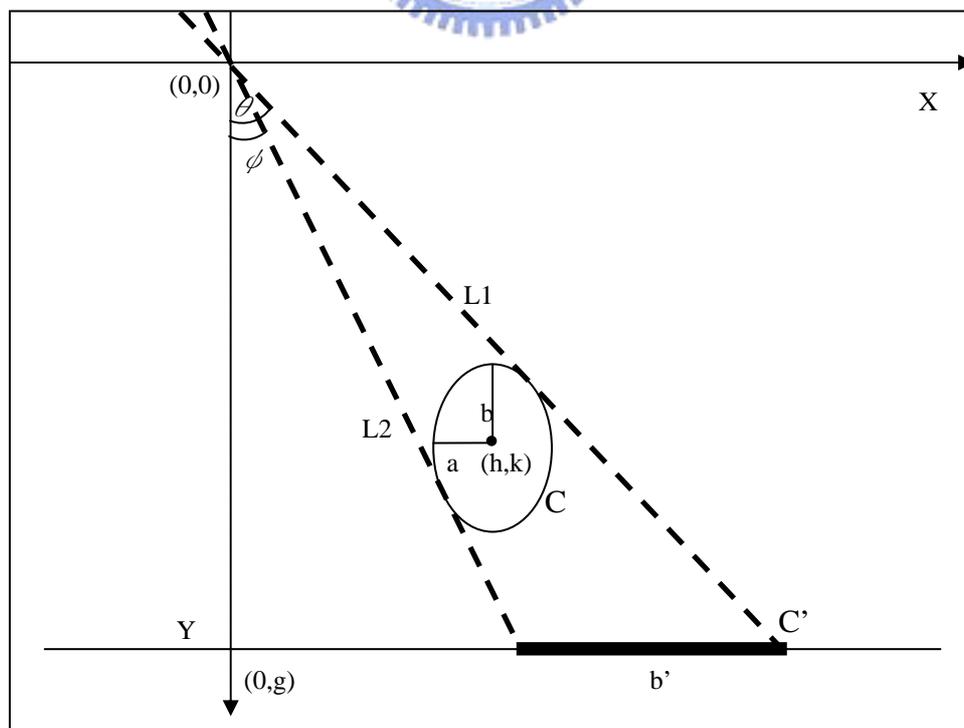


圖 3.15 攝影機與人頭的座標關係圖

我們將圖 3.14 中右邊的人放大來看，並把整個空間視為一座標系，其中 X 軸向右為正，Y 軸向下為正。若攝影機的位置為原點，人頭所形成的橢圓 C 中心位置為 (h, k) ，其短軸為 a ，長軸為 b ，而 L_1 、 L_2 為其通過原點的兩條切線。地板平面為 $y = g$ 的水平線，其與 L_1 、 L_2 的交點形成的線段，即為人頭在影像上所形成的橢圓 C' 的長軸，其長度為 b' ，如圖 3.15 所示。

從圖 3.15 中可以知道， b' 的長度如式(3.4)所示：

$$b' = g(\tan\theta - \tan\phi) \quad (3.4)$$

又 $\tan\theta$ 與 $\tan\phi$ 與 L_1 、 L_2 的斜率(m_{L1} 、 m_{L2})具有倒數的關係，故式(3.4)可改寫為

$$b' = g\left(\frac{1}{m_{L1}} - \frac{1}{m_{L2}}\right) \quad (3.5)$$

而橢圓的中心為 (h, k) ，故我們可令其切線 L_1 、 L_2 為

$$y - k = m(x - h) \pm \sqrt{m^2 a^2 + b^2} \quad (3.6)$$

其切線通過原點，故將 $(0,0)$ 代入，可解出 m

$$m = \frac{kh \pm \sqrt{k^2 h^2 - (h^2 - a^2)(k^2 - b^2)}}{(h^2 - a^2)} \quad (3.7)$$

將式(3.7)代入式(3.5)，整理之後即得到 b'

$$b' = g\left(\frac{2\sqrt{h^2 b^2 + a^2(k^2 - b^2)}}{(k^2 - b^2)}\right) \quad (3.8)$$

由式(3.8)來看，由於人頭的高度以及大小均不變，故 a 、 b 及 k 均固定，則當 h 越大時， b' 也越大，即當人距離中心越遠時，其在影像上所形成的橢圓的長軸也越長。而我們知道當 C' 的形狀為正圓時，其梯度源頭會最集中。故只要知道人頭在不同位置時，在影像上所形成的橢圓與正圓的相似程度，我們就可以計算出適當的 T 值。

橢圓的離心率(eccentricity) e 可以量度橢圓與圓的差異程度，其意義為計算橢圓的焦距(focus)佔長軸(semimajor axis)長度的比例，而離心率的定義如下

$$e = \frac{\text{focus}}{\text{semimajor axis}} \quad (3.9)$$

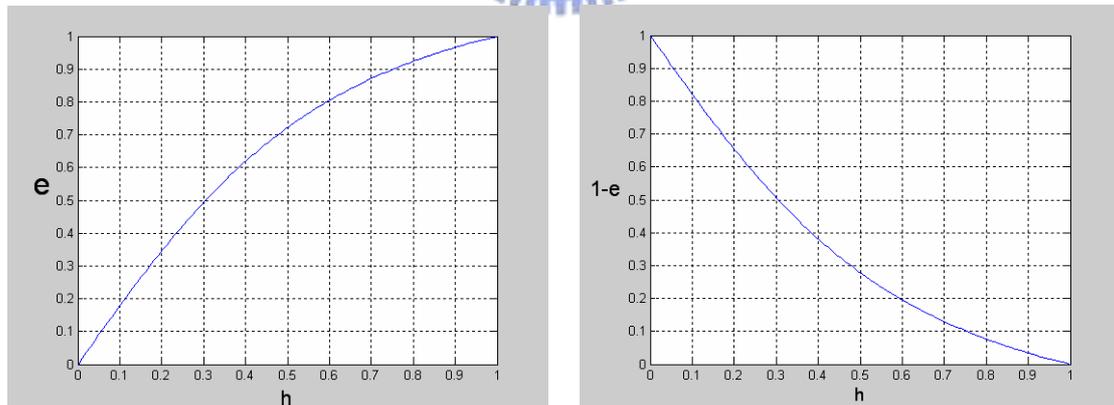
而正圓的焦距為 0，故其離心率为 0，而橢圓的離心率則介於 0 與 1 之間，越接近 0 及表示與正圓越相似。而當人在攝影機正下方時， C' 即為正圓，故我們可將 $h=0$ 代入式(3.8)，即可求出 C' 的短軸長度 a'

$$a' = \frac{2ga}{\sqrt{k^2 - b^2}} \quad (3.10)$$

又橢圓的焦距亦可由長軸與短軸的長度計算出來，故我們可導出離心率與 h 的關係

$$e = \frac{hb}{\sqrt{h^2b^2 + a^2(k^2 - b^2)}} \quad (3.11)$$

由於橢圓的 e 必小於 1，故我們可以用 $1-e$ 來表示橢圓與正圓的相似程度。而 e 與 h 之間的變化關係圖如圖 3.16 所示



(a)

(b)

圖 3.16 人與影像中心的距離與離心率 e 的關係圖

(a) h 與 e 的關係圖 (b) h 與 $1-e$ 的關係圖

依照這個關係，我們定義 T_L 與 T_H 為最低與最高的放射點個數臨界值，將其對應至圖 3.16(b)中 $1-e$ 的 0 與 1，就可以針對位於不同位置的人頭，計算出適當的放射點個數臨界值 T 。

3.2 多人的移動追蹤

經由偵測來找出人在影像中的所在位置後，我們更進一步的希望追蹤人在連續影像中的移動。我們參考 Mei Hail 等人所提出的基於偵測結果的多人移動追蹤的架構[12]，並加以修改，加入符合我們需求的特徵(feature)。關於基於偵測結果的多人移動追蹤的架構，在本論文的第二章中的 2.2 節曾經介紹過，在此就不再多加贅述。本節的重點將著重於我們所使用的特徵及連接機率(connection probability)的計算。

我們的多人移動追蹤演算法為，針對 $t-1$ 時刻影像中欲追蹤的人頭，計算對於 t 時刻影像中所有偵測結果的連接機率，選擇機率最高的進行連接。若連接機率過低，則認為無符合的人頭，即追蹤失敗。而若有兩個人頭連接至同一人頭，則只選擇連接機率較高的進行連接。其中連接機率可以視為前後兩個人頭為同一個人的可能性。我們使用三種特徵來計算連接機率：移動距離與路徑平滑度(trajjectory smoothness)、頭髮與肩膀的衣著色彩、以及肩膀區域的前景涵蓋率(foreground coverage)。接下來我們就針對各個特徵加以說明。

移動距離與路徑平滑度

以人的移動來說，其移動路徑在短時間內，通常不會有太大的變動，即其路徑為平滑的(smooth)。故我們以移動路徑的平滑度，來判斷此移動路徑是否合理。我們以目前的移動方向與前一次移動的方向的夾角大小，來估計其移動路徑的平滑度，如圖 3.17 中所示。圖中 h_t 為在 t 時刻中的人頭位置， $v(h_{t-1}, h_t)$ 則為人頭的移動方向。

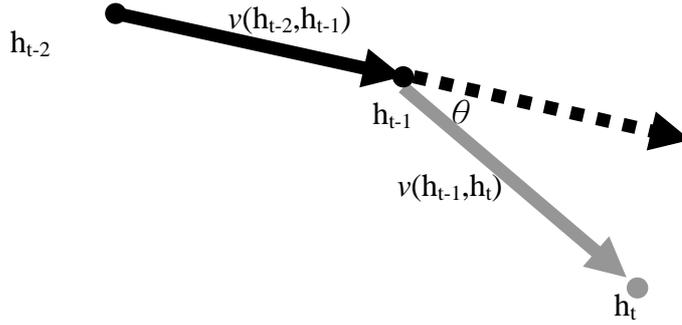


圖 3.17 移動路徑的平滑度分析

移動路徑的平滑度計算如式(3.4)所示，計算夾角 θ 的餘弦值，並將其正規化(normalize)至 0~1 之間。若夾角越大，即移動方向在短時間內的改變越大，則其餘弦值就會越小。

$$smooth_t = \cos \theta = \frac{v(h_{t-2}, h_{t-1}) \cdot v(h_{t-1}, h_t)}{|v(h_{t-2}, h_{t-1})| |v(h_{t-1}, h_t)|} \quad (3.12)$$

由於人未必是持續不斷的移動，可能是停止不動的，或是僅有小幅度的移動。在停止的狀況之下，其移動路徑的改變可能很大(如左右搖晃)，此時其移動路徑的平滑度較不具參考性。故我們亦須考慮其移動距離，來判斷其移動的狀態(moving state)。若移動距離過小，則認為此人為停止(Stop)狀態。若移動距離在適當的範圍內，則認為此人為移動(Move)的狀態。若移動距離過大，超過一合理的範圍，就完全不予考慮。式(3.13)整理了移動距離與狀態之間的關係。

$$Moving\ State = \begin{cases} Stop & \text{if } dis(h_{t-1}, h_t) < Low \\ Move & \text{if } Low \leq dis(h_{t-1}, h_t) \leq High \\ Unreal & \text{if } dis(h_{t-1}, h_t) > High \end{cases}$$

$dis(h_{t-1}, h_t)$: distance between h_{t-1} and h_t

(3.13)

頭髮與肩膀的衣著色彩

色彩為影像中相當重要的特徵。由於我們的人群計數是以人頭偵測為主，故頭髮的顏色是當然的特徵之一。但人的頭髮顏色通常差異性並不大，為了增加其可辨識度，我們必須將人的衣著色彩也考慮進來。

但在垂直視角下，人所在的位置不同，其身體出現在影像中的部位也不同。例如站在距離攝影機較遠的位置，若不考慮遮蔽的問題，其身體與腳的部份均可被攝影到，但若站在攝影機的正下方，則只剩下肩膀的部位能夠被攝影到。唯一在所有位置都會出現的身體部位為肩膀，故我們僅採用肩膀部位的色彩來當作身體衣著的色彩。

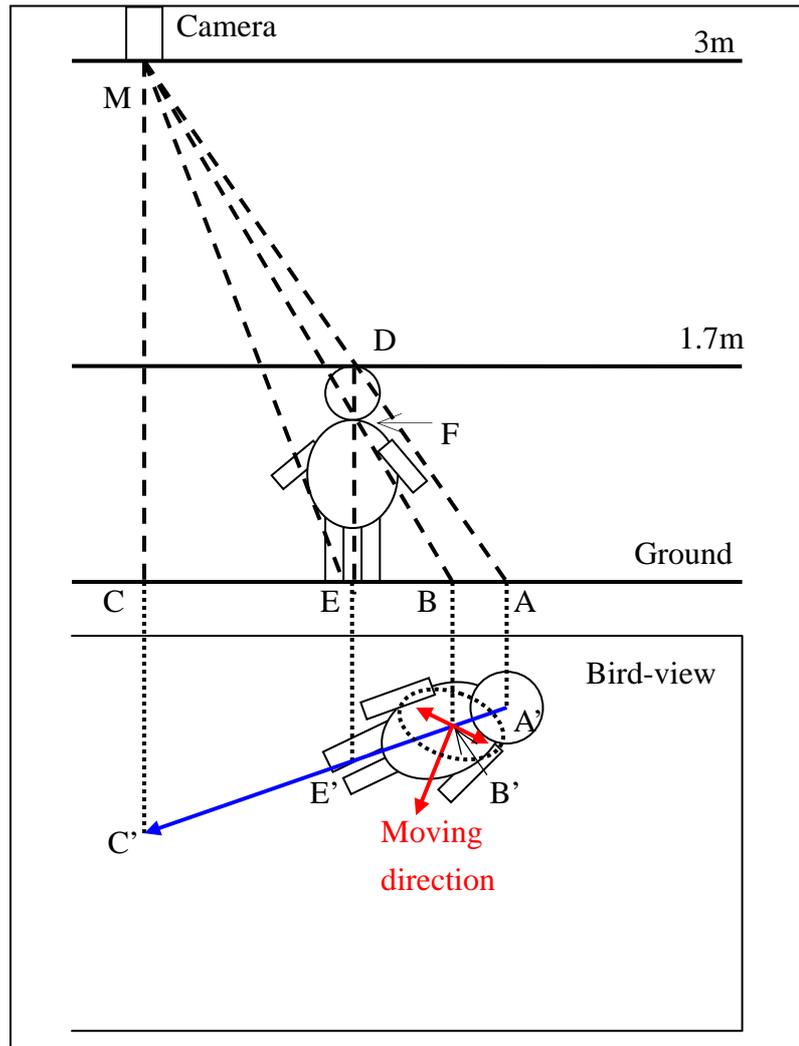


圖 3.18 肩膀部位與移動方向的三度空間關係圖

但是該如何得知肩膀部位的正確位置仍是一個問題。由於攝影機的位置是固定的，故我們可以利用三度空間與影像的對應關係，以及其移動方向來推估肩膀的概略位置。圖 3.18 中，地板的上半部為一個 170 公分高的人與架設在 3 公

尺高天花板的攝影機在三度空間中的位置，下半部則為垂直視角的攝影機所看到的平面影像。由於人所產生的遮蔽效應，原本地板上 A 點的位置被人的頭部(D 點)所遮蔽，故在影像上 A' 點所呈現的即為人的頭部。同理，肩膀的位置遮蔽了地板上的 B 點，即為影像上的 B' 點。E 點為人所站的位置，則三角形 ADE 與三角形 AMC 為相似三角形。故我們可得到如式(3.14)的比例關係：

$$\frac{\overline{AE}}{\overline{AC}} = \frac{\overline{DE}}{\overline{MC}} \quad (3.14)$$

又各點在影像中的比例並不改變，故

$$\frac{\overline{AE}}{\overline{AC}} = \frac{\overline{A'E'}}{\overline{A'C'}} \quad (3.15)$$

將式(3.15)代入(3.14)，即得到式(3.16)

$$\frac{\overline{A'E'}}{\overline{A'C'}} = \frac{\overline{DE}}{\overline{MC}} \quad (3.16)$$

因為 \overline{MC} 、 \overline{DE} 為已知，而 C' 為影像的中心。若人頭位置固定後， $\overline{A'C'}$ 也可經由計算得知，則我們可得知 $\overline{A'E'}$ 在影像中的長度。而肩膀在影像中的位置 B'，也可經由實際上肩膀(F 點)與人的比例來求得，如式(3.17)。

$$\overline{A'B'} = \overline{A'E'} \times \frac{\overline{DF}}{\overline{DE}} \quad (3.17)$$

在一般情況下，我們可以假設人都是面向其移動的方向。若我們以一個橢圓來描述肩膀的形狀，則此橢圓的長軸將會垂直於人的移動方向。故配合人的移動方向，我們可以在 B' 點以一個固定比例的橢圓形來逼近肩膀的概略位置，並記錄此橢圓區域內的色彩資訊來做為人的衣著色彩。

而色彩的比對方式，則是比較在 LCH 的色彩空間下，來計算各個色頻(channel)的分布相似度(distribution similarity)。而計算分布相似度的方式則採用 Fisher 的標準(Fisher criterion)，如式(3.18)。

Fisher criterion :

$$f = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 - \sigma_2^2} \quad (3.18)$$

再將各個色頻的分布相似度加權總和，即為色彩的相似度，如式(3.19)。而為了減低光影變化的影響，給予亮度(Luminance)的色頻較低的權重。

$$color_i = \omega_L f_L + \omega_C f_C + \omega_H f_H \quad (3.19)$$

肩膀區域的前景涵蓋率

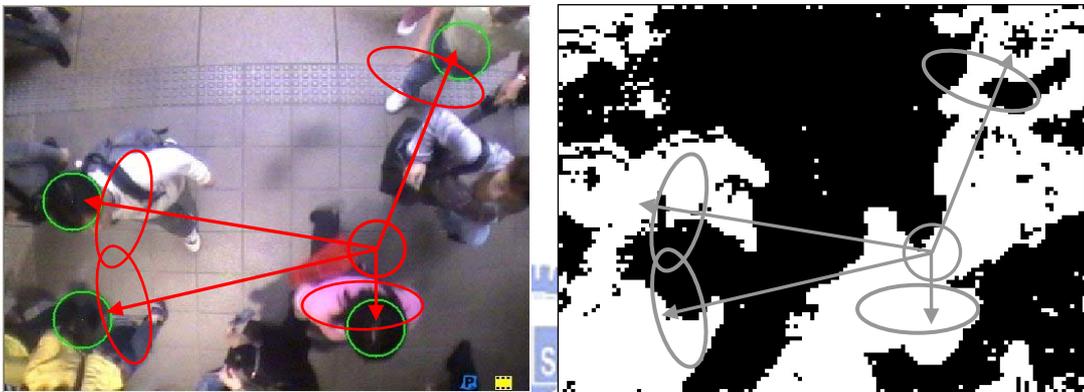
在 3.1.1 中前景偵測所得到的結果，也可以用來輔助追蹤的判斷。依照前景的定義，人的身體部分均屬於前景。若此連接為一個正確的連接，則用以上方式所推估出來的肩膀區域，應該也都屬於前景。所以，計算所推估出的肩膀區域內，屬於前景的像素所佔百分比，可以對肩膀區域的正確性加以確認。而我們將這個百分比稱為肩膀區域的前景涵蓋率，如式(3.20)所示。

$$cover_i = \frac{\text{pixel} \in \text{foreground}}{\text{All shoulder pixel}} \quad (3.20)$$


若所連接的位置並非真正人頭的所在位置，或是其移動的方向不正確，則其區域內前景的涵蓋率將較低，反之則較高。從圖 3.19 中我們可以很明顯的看到，其中圖 3.19(a)為 t-1 時刻的影像，圖 3.19(b)則為 t 時刻的影像，圖 3.19(c)則為 t 時刻的影像經過前景偵測後所得到的前景影像。若以圖 3.19(a)中的人頭 H 為例，在所有可能的連接之中，只有正確的連接所推估出來的肩膀區域均屬於前景，而其餘的連接，其推估出的肩膀區域的前景涵蓋率均較低。



(a)



(b)

(c)

圖 3.19 肩膀區域的前景覆蓋率

(a) t-1 時刻的偵測結果 (b)t 時刻的偵測結果與可能連接 (c)t 時刻的前景影像與可能連接

最後，我們針對每一個可能的連接，分別計算出三種特徵的相似度後，再將其進行加權總和，即為連接機率。如同前面所提到的，若人在停止的狀態下，則不考慮其移動路徑的平滑度。連接機率的計算如式(3.21)所示。

$$P(h_{t-1}, h_t) = \begin{cases} w_{smo} * smooth_t + w_{col} * color_t + w_{cov} * cover_t & \text{if } State = Move \\ w_{col} * color_t + w_{cov} * cover_t & \text{if } State = Stop \end{cases} \quad (3.21)$$

其中，我們以色彩的特徵為主，給予其較重的權重。原因是色彩的相似度代表兩個人頭之間在影像上實際的相似度，而移動路徑的平滑度與前景的涵蓋率則是輔助來確認此路徑的合理性。圖 3.20 為我們的多人移動追蹤結果，其中橢圓

形的部份即為所推估出的肩膀區域，可以看出其與實際的肩膀區域相去不遠。



圖 3.20 多人移動追蹤的結果

(a) t-1 時刻 (b) t 時刻

如同在我們在 2.2 中所提到的，透過追蹤，我們也能對於偵測的結果加以確認與改善。假設影像中偵測出一人頭 H，但前一張影像中，卻沒有任何一個人頭連接至 H。若 H 靠近影像的邊緣，則可能為一剛移動進入畫面的人，否則有可能為一錯誤的偵測。在追蹤過程中，我們即可將此類錯誤的偵測加以去除。

另外，即使在人頭的偵測上發生了遺漏，我們也可以透過連續的影像，使追蹤不至於中斷。若一人頭在 t 時刻的影像發生追蹤失敗，但在 t+1 時刻的影像中卻找到可連接的人頭，則可能在 t 時刻發生了偵測的遺漏。故在追蹤過程中，我們允許一次的追蹤失敗，並將其保留至下一時刻的影像中，來尋找是否有可連接的人頭。如此即使在有偵測遺漏的狀況下，仍可正確的達追蹤的目的，但其缺點為有可能造成錯誤追蹤的延續，故在一次追蹤失敗後，需要以較嚴苛的條件(如較高的連接機率臨界值)來進行追蹤。

第四章 實驗及結果討論

在本章中，針對本論文所提出的人群計數以及追蹤的方法進行實驗。實驗分為垂直視角的人頭偵測與多人的移動追蹤兩大部分來進行。在偵測方面，探討了在不同的群聚方式下人頭偵測的正確率，並且測試在不同的亮度及環境下，對其正確率的影響。而在追蹤方面，除了測試追蹤的正確性之外，也測試了追蹤對於偵測的改善效果。

4.1 實驗設計

為了測試本論文的方法在實際環境中的正確性，我們在某車站月台的四個亮度、環境皆相異的位置裝設攝影機，並且利用這四部攝影機所取得的現場影像來進行實驗。其中鏡頭 1 位在月台的前段人潮較稀少處，鏡頭 2 位在月台中段的電扶梯出入口，鏡頭 3 位在月台邊緣的上下車門口處，鏡頭 4 位在上下車的排隊等候處。

我們所採用的攝影機為照度 0.05Lux，焦距為 1.4mm 的 CCD 攝影機，裝設的高度為 3 公尺，以垂直向下的方式攝影。而用來進行實驗的機器為 P4-3G Hz，2G bytes 記憶體之 PC 一台，所使用的作業系統為 Microsoft Windows XP SP2。

為了測試本論文的方法，在不同人潮的擁擠程度下的偵測效果，我們將分別以交通離峰時段與尖峰時段的影像來進行實驗。我們分別在 2006 年 6 月 1 日星期四 14:00 的交通離峰時段，以及 17:30 的交通尖峰時段收集實驗用的資料。每個時段均以每秒擷取 4 張影像的速度，收集每個鏡頭 2500 張，大約 10 分鐘連續的影像資料，供之後的實驗使用。每張影像的大小為 512x384 像素。

4.2 人頭偵測

在人頭偵測的實驗中，我們採用的人頭半徑 R 為 28 像素，這個半徑同時也是固定區域式群聚以及 Mean-shift 式群聚所使用的區域半徑。我們以 precision 與 recall 做為評量人頭偵測效果的標準。其中 precision 與 recall 的定義如式(4.1)、(4.2)所示，並且以式(4.3)中所定義的 Performance 來評量其整體的偵測結果。

$$Precision = \frac{\text{Correct detected human head}}{\text{All detected result}} \quad (4.1)$$

$$Recall = \frac{\text{Correct detected human head}}{\text{All human head}} \quad (4.2)$$

$$Performance = \frac{2 * Precision * Recall}{Precision + Recall} \quad (4.3)$$

其中當 precision 越高時，代表偵測結果中錯誤偵測(false detection)越少。而 recall 越高時，則代表偵測結果中偵測的遺漏(miss detection)越少。而 performance 的計算為 precision 與 recall 的調和平均數(harmonic mean)，此為一般在資訊檢索(information retrieval)中常見的計算方式，稱為 F-measure。

針對人頭偵測，我們總共進行了三項實驗，分別為：

1. 不同群聚方法的偵測實驗：比較固定區域式群聚方式與 Mean-shift 群聚方式，對於人頭偵測的影響。
2. 固定 T 與隨距離調整 T 的偵測實驗：比較固定 T 與隨距離調整 T ，兩種決定放射點臨界值 T (定義可參見 3.1.3)的方式，對人頭偵測結果的影響。
3. 不同亮度環境下的偵測實驗：以四個不同鏡頭所拍到的影像，進行人頭偵測，比較本論文方法對亮度與環境的適應力。

接下來我們便針對各項實驗的結果進行討論。

4.2.1 不同群聚方法的偵測結果

為了解 3.1.3 中，固定區域式與 Mean-shift 兩種群聚方法，對於偵測效果的影響，我們選定鏡頭 3 的影像，分別使用兩種不同的群聚方法進行人頭偵測。經由人工的統計，在離峰時段的 2500 張影像中，共出現了 588 個人頭，而在尖峰

時段的 2500 張影像中，共出現了 2569 個人頭。圖 4.1 ~ 4.3 為兩種群聚方法在離峰時段下，使用不同的放射點個數臨界值 T ，所得到的 precision、recall，以及整體 performance 的比較。而圖 4.4 ~ 4.6 則為兩種群聚方式在尖峰時段下的比較。

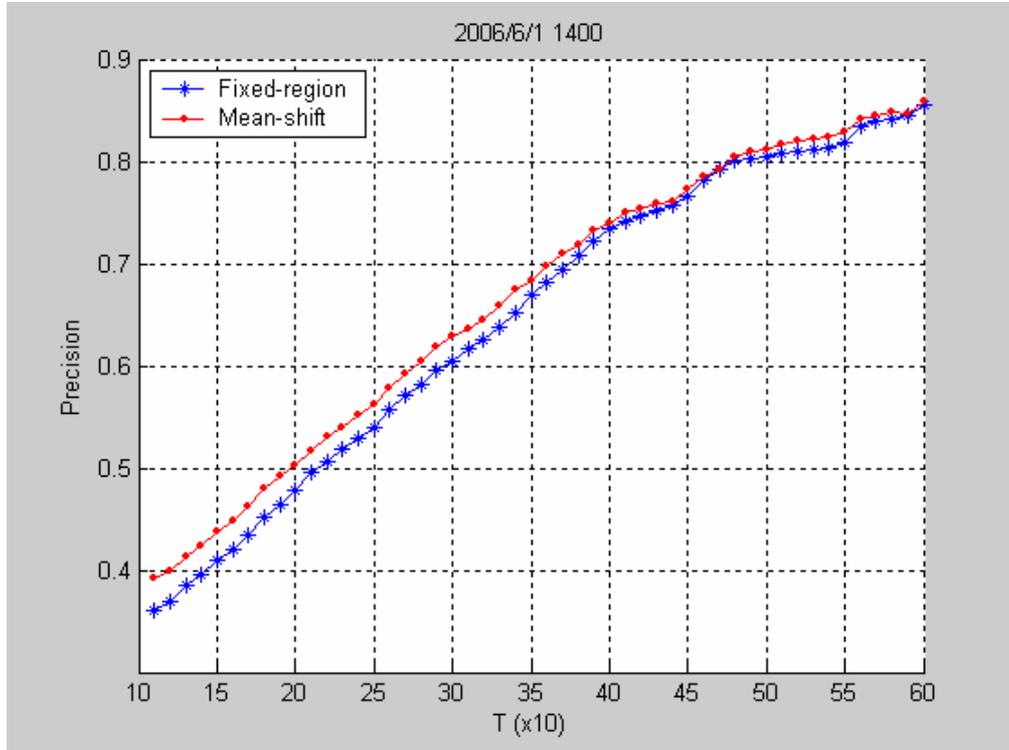


圖 4.1 不同群聚方式在離峰時段的 Precision 比較

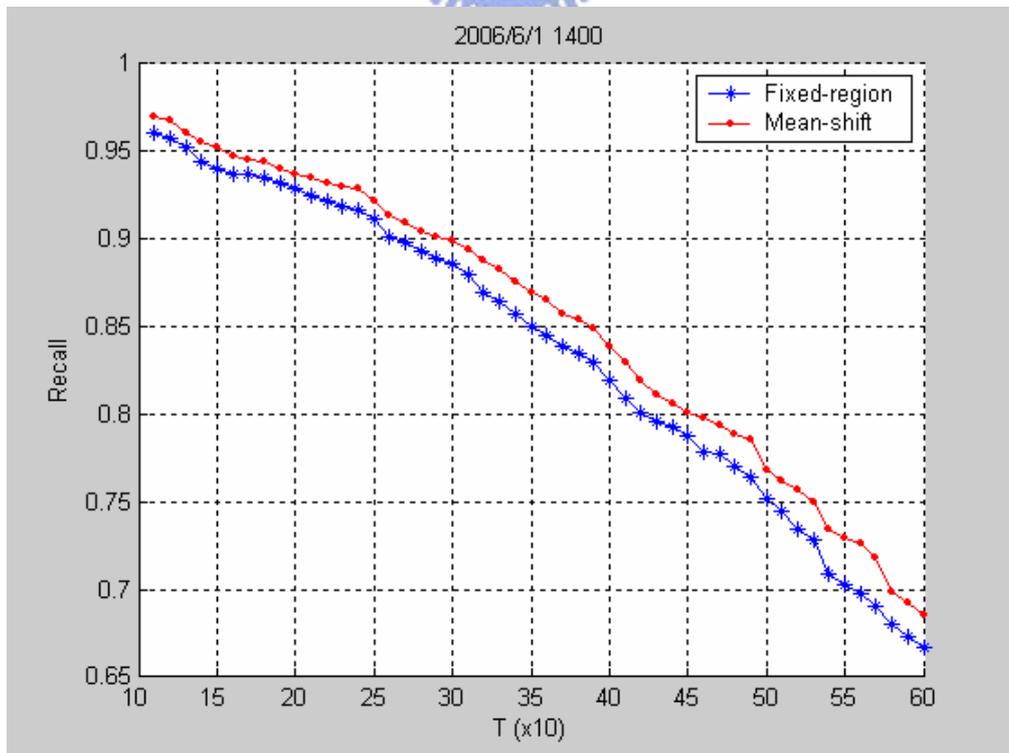


圖 4.2 不同群聚方式在離峰時段的 Recall 比較

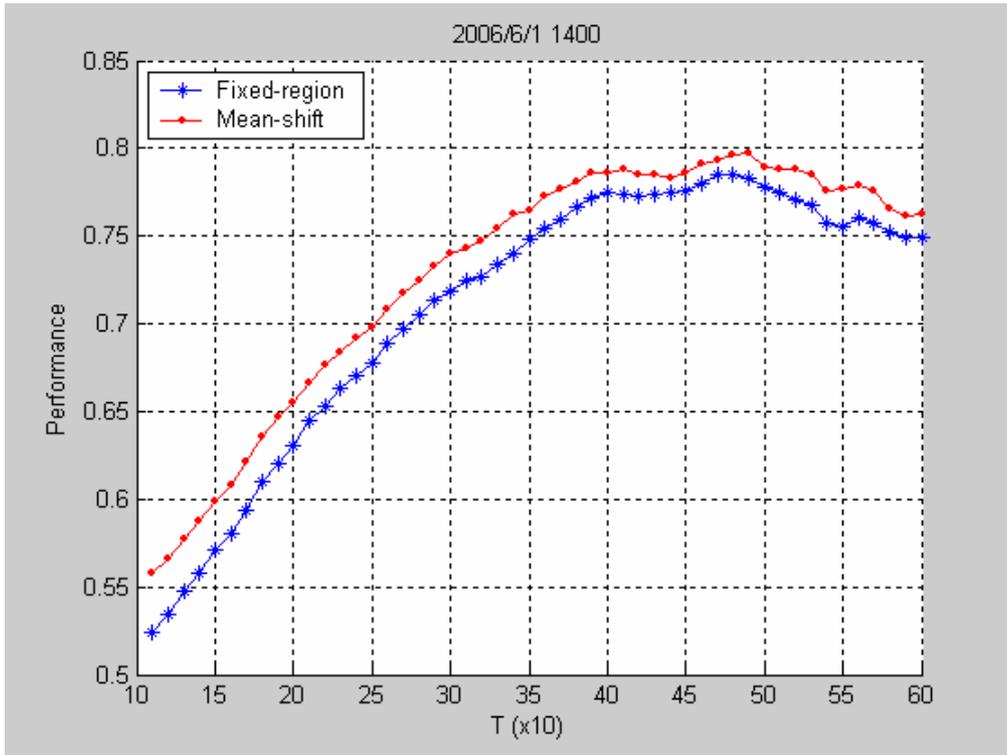


圖 4.3 不同群聚方式在離峰時段的 Performance 比較

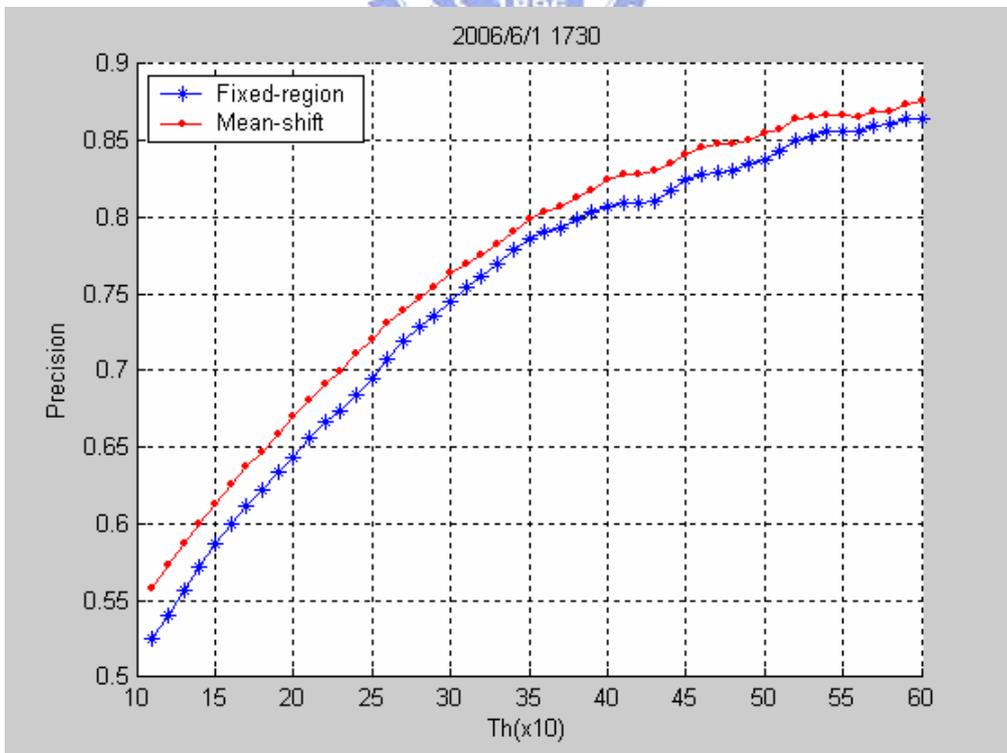


圖 4.4 不同群聚方式在尖峰時段的 Precision 比較

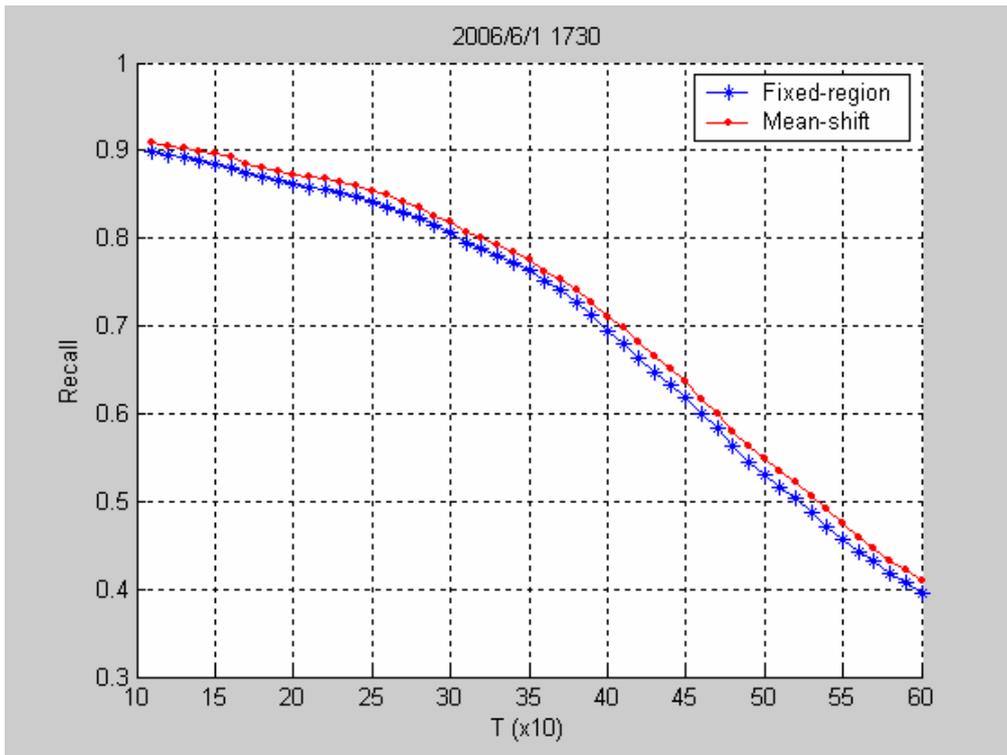


圖 4.5 不同群聚方式在尖峰時段的 Recall 比較

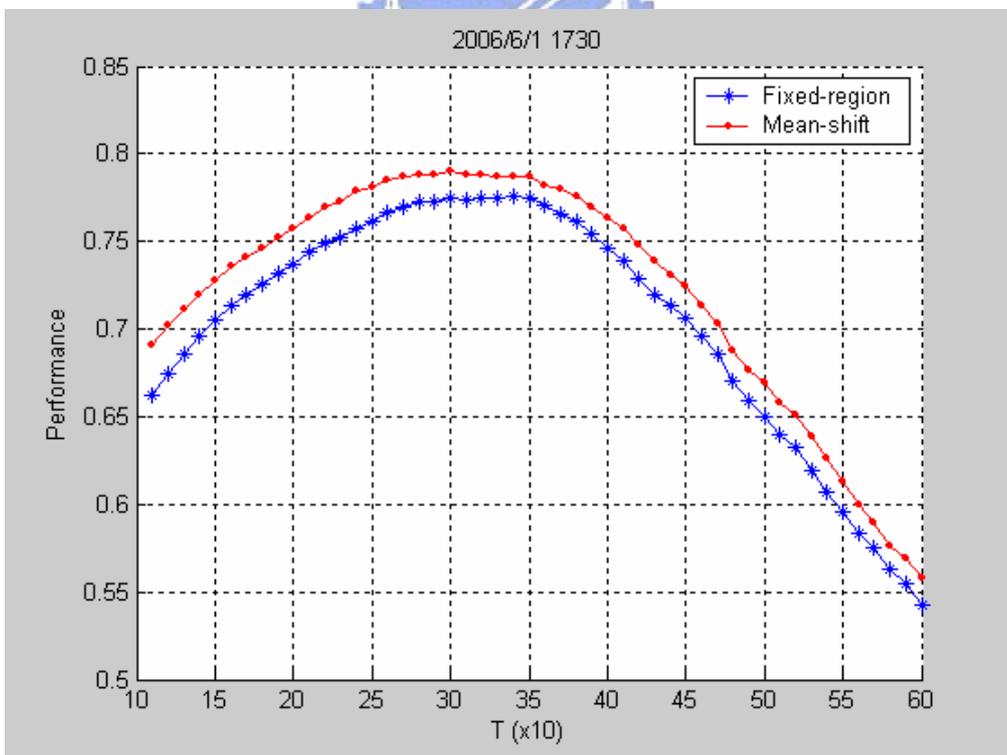


圖 4.6 不同群聚方式在尖峰時段的 Performance 比較

從圖 4.1 ~ 4.6 中，我們可以看到不論是在交通的離峰時段或是尖峰時段下，Mean-shift 的群聚方式在各種 T 值下均得到較固定區域式群聚較好的偵測效果。我們將 T=300 時的各項詳細數值作成表 4.1，來做進一步的比較。

表 4.1 在各個時段下的不同群聚方式的詳細數值(T=300)

Time	Method	Head	Corr.	False	Prec.	Recall	Perfor.
2006/6/1 1400	Fixed-region	588	521	341	0.6044	0.8861	0.7186
	Mean-shift	588	529	313	0.6283	0.8997	0.7399
2006/6/1 1730	Fixed-region	2569	2070	707	0.7454	0.8058	0.7744
	Mean-shift	2569	2103	650	0.7639	0.8186	0.7903

從表 4.1 中我們可以清楚的看到，在相同的 T 值下，Mean-shift 的群聚方式可以較固定區域式的群聚得到較多正確的偵測(correct detection)，並且錯誤的偵測(false detection)也較少。由尖峰時段的測試數據中可以看出的更明顯，Mean-shift 式的群聚較固定區域式的群聚多了 33 個正確的偵測，並且減少了 57 個錯誤偵測。由以上的結果我們可以證實，Mean-shift 式的群聚在人頭偵測上，的確較固定區域式的群聚有較好的效果。

此外，從表 4.1 中，我們可以看到不論哪一種群聚方式，離峰時段的 precision 均較尖峰時段要來得低，而 recall 則較高。造成這種情況的主要因為，由於我們的人頭偵測方式是以「近似圓形的深色區塊」為模型，主要的偵測錯誤來自於地板上深色的影子。但尖峰時段的人潮較為擁擠，反而遮蓋了大部分的地板，而離峰時段的人潮較為稀疏，地板上影子的影響則較為增加，因而造成了離峰時段的偵測錯誤的比率較高。但是尖峰時段的人潮較多，穿深色衣服或是禿頭的人出現的機會也增加，故造成了較多的偵測遺漏，使得 recall 較低。但若以整體 performance 來看，尖峰時段的整體表現甚至較離峰時段還要好，這也說明了本論文的方法在人潮擁擠的環境之下依然適用。

4.2.2 隨距離調整 T 的偵測結果

我們加入 3.1.4 中隨距離調整 T 的方式，來測試其對於偵測結果的影響。我們採用與前一節中相同的影像資料，並且使用 Mean-shift 的群聚方式。為了比較其對於偵測效果的影響，我們將最高放射點個數臨界值 T_H 固定為 400，採用不同的最低放射點個數臨界值 T_L 所得到的偵測效果，來與 T 為固定，且其值與 T_L 相同時的偵測效果來做比較。圖 4.7~4.9 為離峰時段的各項數值，圖 4.10~4.12 則為尖峰時段的各項數值。

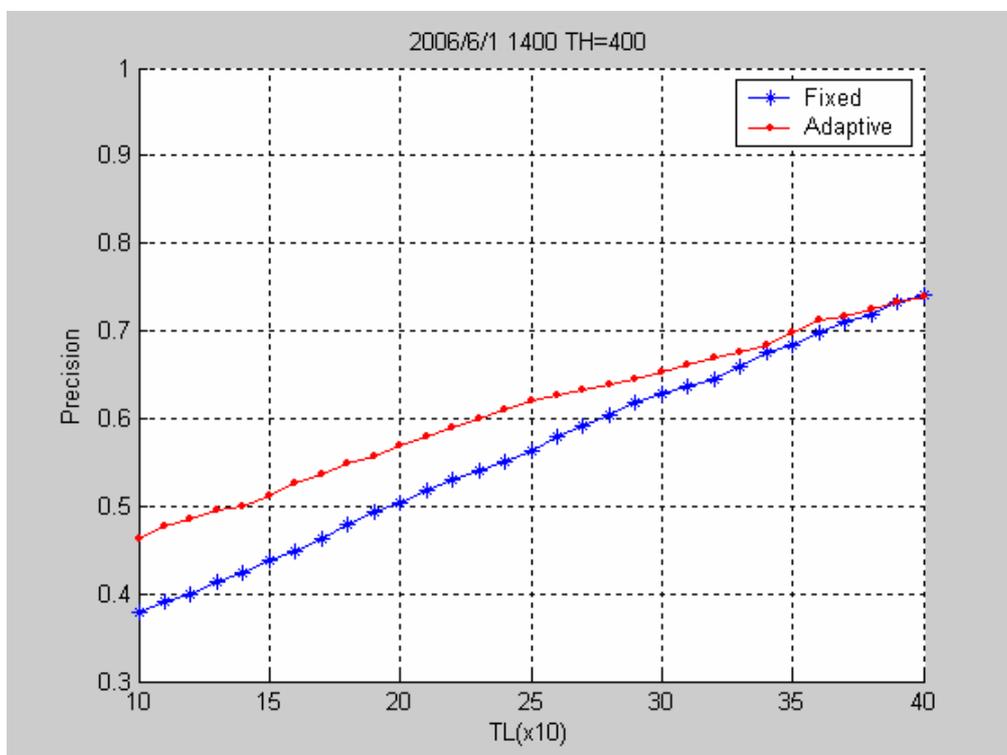


圖 4.7 隨距離調整的 T 與固定 T 在離峰時段的 Precision 比較

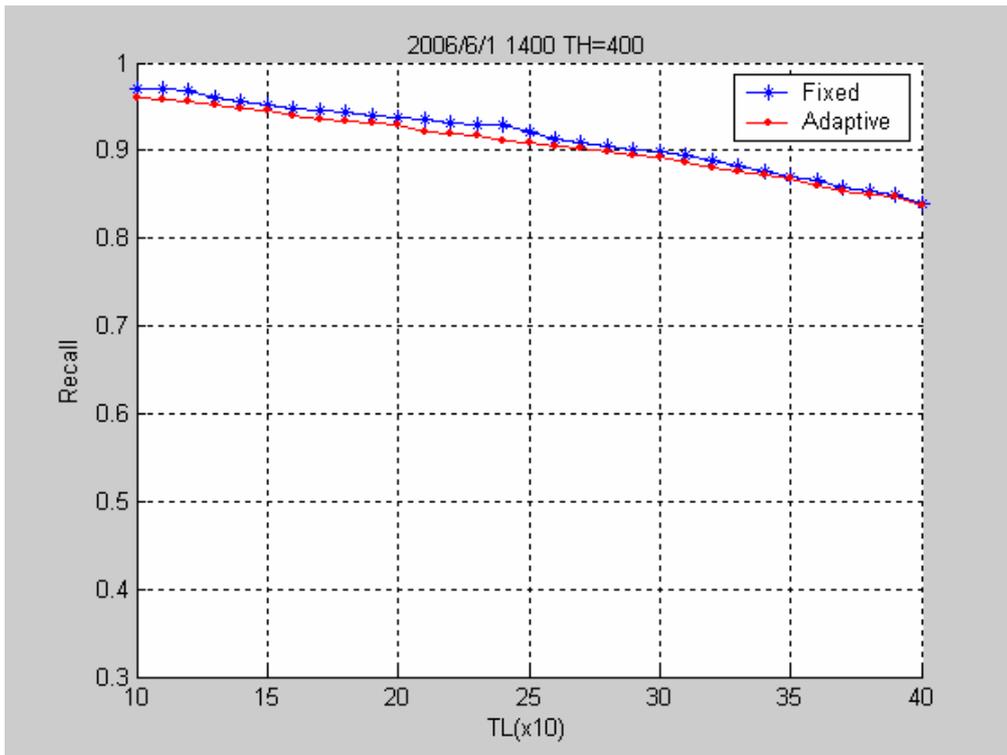


圖 4.8 隨距離調整的 T 與固定 T 在離峰時段的 Recall 比較

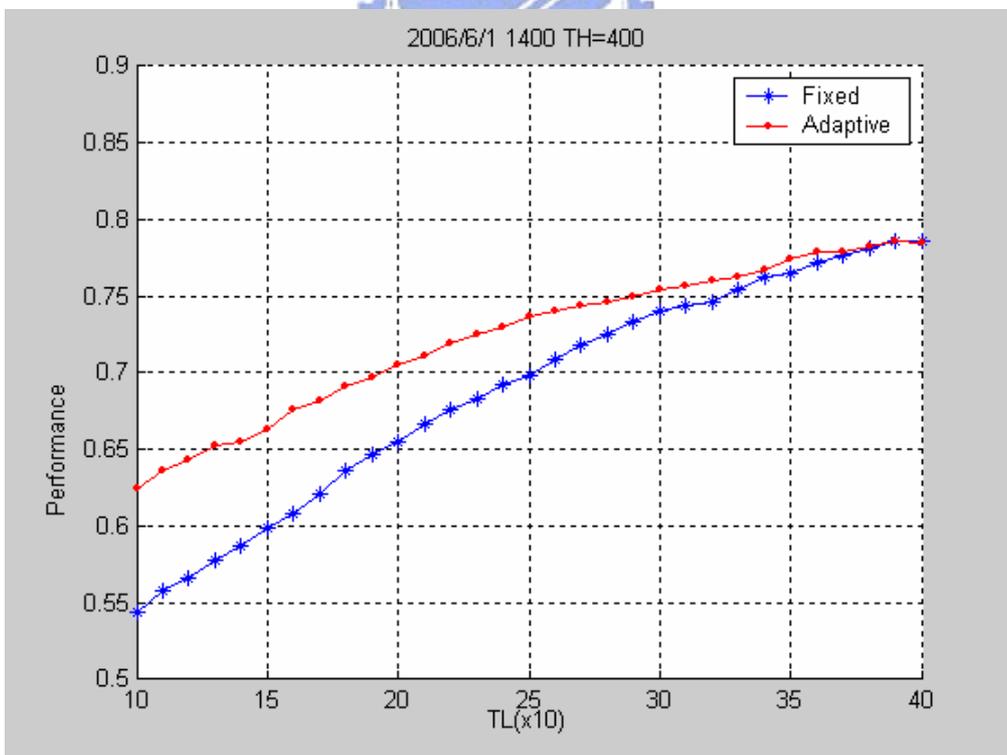


圖 4.9 隨距離調整的 T 與固定 T 在離峰時段的 Performance 比較

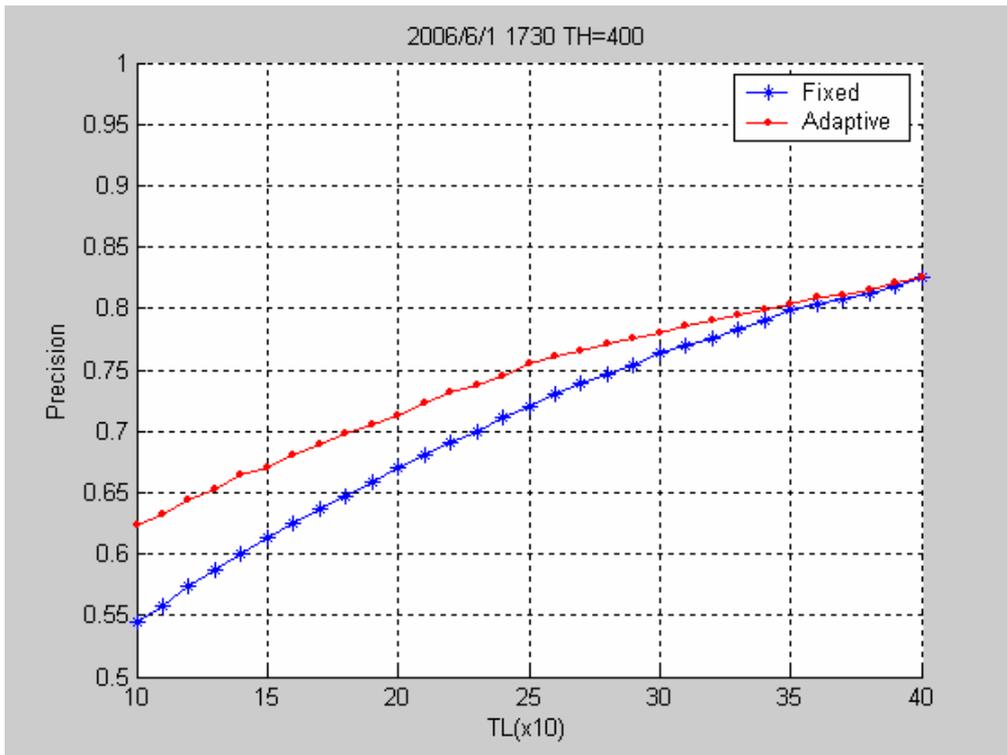


圖 4.10 隨距離調整的 T 與固定 T 在尖峰時段的 Precision 比較

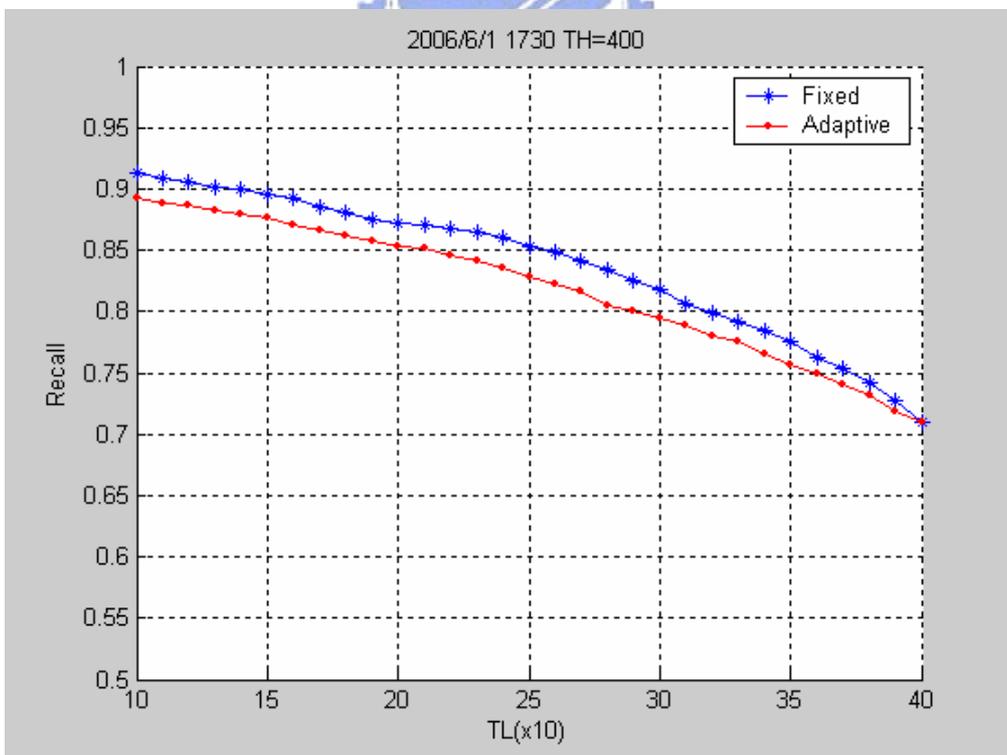


圖 4.11 隨距離調整的 T 與固定 T 在尖峰時段的 Recall 比較

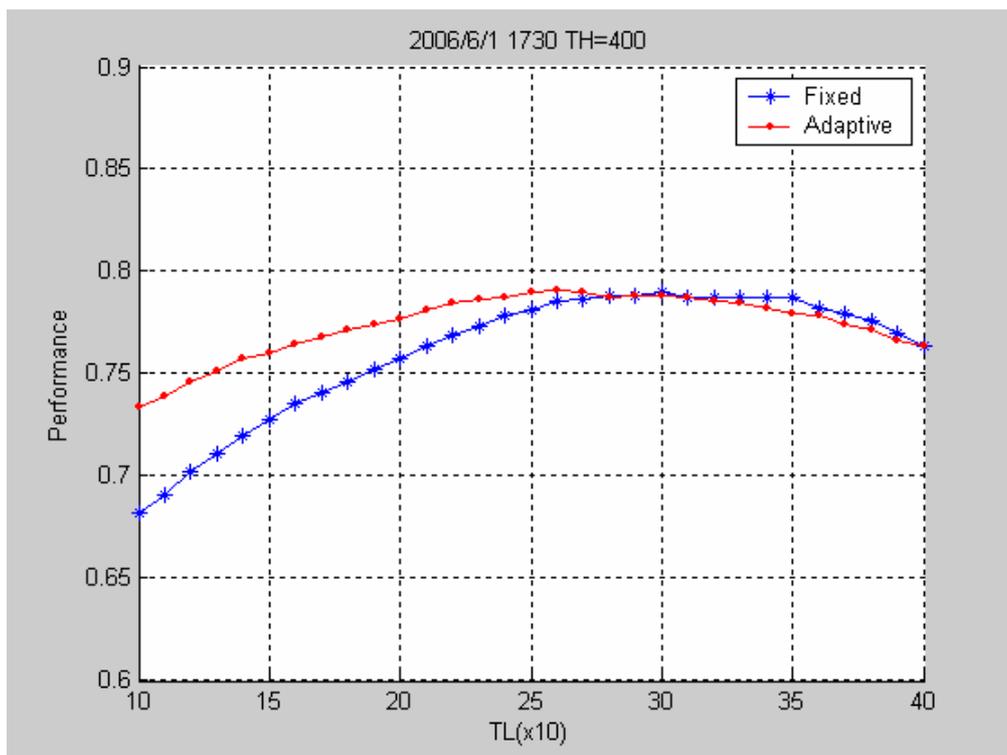


圖 4.12 隨距離調整的 T 與固定 T 在尖峰時段的 Performance 比較

從圖 4.7~4.12 中，我們可以看到在兩個時段中，隨距離調整的 T 的 recall 均較固定 T 有略微的下降，下降幅度約在 3% 以內。但在 precision 方面卻有顯著的提升，最大的幅度可達到 %7。這個現象說明了隨距離調整的 T 的確減少了錯誤偵測的數量。雖然隨距離調整的 T 所能達到的最高 performance，並未較固定 T 要高，甚至在尖峰時段中 $T_L > 300$ 時有些微的下降。但若我們把固定 T=200 時的數值，與隨距離調整的 T 中可以得到相近 recall 的數值做一比較，作成表 4.2。從表 4.2 中我們可以看到，在相近的 recall 下，隨距離調整的 T 的確可以較固定 T 的方式得到較少的錯誤偵測以及較高的 precision。

表 4.2 固定 T 與隨距離調整的 T 在相近 recall 下的比較

Time	Method	T	Head	Corr.	False	Precision	Recall
2006/6/1 1400	Fixed	200	588	551	544	0.5032	0.9381
	Adaptive	L:160 H:400	588	553	496	0.5272	0.9405
2006/6/1 1730	Fixed	200	2569	2240	1105	0.6697	0.8719
	Adaptive	L:160 H:400	2569	2241	1049	0.6812	0.8723

4.2.3 不同亮度環境下的偵測結果

為了了解本論文的方法在各種不同亮度與環境下的偵測效果，我們分別以同一個時段，四個不同位置所拍攝到的影像，來進行人頭偵測的實驗。這四個位置與 4.1 節中介紹的相同，圖 4.13 為各個鏡頭所拍攝到的影像。而表 4.3 為此四個鏡頭所拍攝到的影像亮度資訊，其中影像的亮度為此張影像中所有像素亮度的平均。



圖 4.13 各個鏡頭所拍攝到的影像

a、b、c、d 依序為鏡頭 1~4

表 4.3 四個鏡頭所拍到影像的亮度資訊

Cam	Rank	Average	Std. deviation	Highest	Lowest
1	3	71.51	1.121	75.51	66.66
2	4	67.71	6.373	74.36	57.99
3	2	74.18	3.011	80.83	65.79
4	1	76.41	1.897	81.45	69.19

我們採用 Mean-Shift 的群聚方式，以及隨距離調整的 T 值，其中 $T_L=200$ ， $T_H=400$ ，其餘參數均與之前相同。表 4.4 為尖峰時段下四個鏡頭的偵測結果。

表 4.4 尖峰時段下四個鏡頭的偵測結果

Cam	Rank	Head	Corr.	False	Prec.	Recall	Perfor.
1	3	299	260	112	0.6989	0.8696	0.7750
2	4	3402	2340	919	0.7180	0.6878	0.7026
3	2	2569	2193	884	0.7127	0.8536	0.7768
4	1	759	621	365	0.6298	0.8182	0.7117

從表 4.4 中我們可以看到，本論文的方法在鏡頭 1、3、4 均有 81% 以上的 recall，而鏡頭 2 的 recall 特別低，僅有 68%。推測原因為，在表 4.3 的統計中，鏡頭 2 的影像亮度為 4 個鏡頭中最低，而當亮度過低時，由於人頭與背景之間的邊緣變得不明顯，造成較多的偵測遺漏，故 recall 大幅的下降。並且鏡頭 2 的影像亮度的變化也最大，造成前景偵測上的困難，也會對偵測效果產生影響。

而我們也發現，鏡頭 4 的 precision 也較其他 3 個鏡頭為低。由於鏡頭 4 所在環境中的光線主要來自於一點光源，造成當人通過時，所產生的陰影部分將較一般日光燈光源來得大且暗，我們從圖 4.13(d) 中就可以看到這樣的現象。由於陰影增加的關係，同樣也造成錯誤偵測的增加，故其偵測的 precision 較低。若能使用更精準的前景偵測方法，將此類陰影自前景中去除，應能對其 recall 有所提升。

雖然在各個環境下的偵測結果不盡相同，但以整體 performance 來看的話，本論文的方法均可達到 70% 以上。

4.3 多人的移動追蹤

在多人移動追蹤的實驗中，我們計算連接機率時各個特徵所使用的權重為：移動路徑平滑度 0.25、頭髮與肩膀的衣著色彩 0.5、以及肩膀區域的前景涵蓋率 0.25。我們分別以兩個方面來評量多人移動追蹤的結果。其一是連接的正確

性，我們利用所有連接中是正確連接的比率來評量，稱為連接正確率(connection correction rate)。以及對於通過畫面中的人進行追蹤的 precision 與 recall，其定義如式(4.4)、式(4.5)與式(4.6)所示。而追蹤的整體 performance 則與式(4.3)的定義相同

$$\text{Connection correction rate} = \frac{\text{Correct connection}}{\text{All connection}} \quad (4.4)$$

$$\text{Human tracking precision} = \frac{\text{Correct tracked human}}{\text{All tracking result}} \quad (4.5)$$

$$\text{Human tracking recall} = \frac{\text{Correct tracked human}}{\text{All human}} \quad (4.6)$$

針對多人移動追蹤，我們總共進行了兩項實驗，分別為：

不同群聚方法的偵測實驗：比較固定區域式群聚方式與 Mean-shift 群聚方式，對於人頭偵測的影響。

1. 容許偵測遺漏的追蹤實驗：比較容許偵測遺漏的追蹤結果，與不容許偵測遺漏的追蹤結果的差異。
2. 利用追蹤去除錯誤偵測的實驗：測試追蹤對於去除偵測結果中錯誤偵測的能力。

接下來我們便針對各項實驗的結果進行討論。

4.3.1 加入容許偵測遺漏的追蹤結果

我們分別以不容許偵測遺漏的追蹤方法，與容許偵測遺漏的追蹤方法做一比較，以了解容許偵測遺漏在追蹤上的助益。所選用的資料為與人頭偵測實驗同一天的資料，同樣也以離峰與尖峰時段兩種資料來進行實驗。而所使用的人頭偵測方法均為 Mean-shift 的群聚方式，隨距離調整 T，其值介於 200~400 間。表 4.5 為兩種時段下的，不同追蹤方法的連接正確率，表 4.6 則為其人群移動追蹤的 precision 與 recall。

表 4.5 不同追蹤方法的連接正確率

Time	Method	Connection	Correct	False	Correction
2006/6/1 1400	Not allow	562	427	135	0.7598
	Allow	585	431	154	0.7368
2006/6/1 1730	Not allow	1949	1655	294	0.8492
	Allow	2028	1679	349	0.8279

表 4.6 人群移動追蹤結果

Time	Method	Human	Corr.	False	Prec.	Recall	Perfor.
2006/6/1 1400	Not allow	96	80	15	0.8421	0.8333	0.8377
	Allow	96	82	16	0.8367	0.8542	0.8453
2006/6/1 1730	Not allow	346	280	65	0.8116	0.8092	0.8104
	Allow	346	299	78	0.7931	0.8642	0.8271

在表 4.5 中，我們發現容許偵測遺漏的追蹤方式，雖然其正確連接的數量增加，但是同時也造成錯誤連接的增加，這也就是所謂的錯誤延續(error propagation)。所以在連接正確率的計算上，容許偵測遺漏的追蹤方法，其連接正確率略微的降低。

但是若從追蹤的觀點來看，在表 4.6 中，容許偵測遺漏的方法較不容許偵測遺漏的方法，能夠正確的追蹤出更多的人，得到較高的 recall。這個提升的效果，在偵測遺漏較多，偵測 recall 較低的尖峰時段中(參見表 4.1)更加的明顯，這可以證明容許偵測遺漏的追蹤方法，的確能夠減少偵測遺漏對於追蹤上的影響。但由於錯誤延續的關係，也造成錯誤追蹤的增加，使其 precision 較低。若由整體的 performance 來看，容許偵測遺漏的追蹤方法其整體表現仍是較佳的，並且都超過了 80% 以上。

4.3.2 追蹤對於偵測結果的改善

在追蹤過程中，我們也加入對於錯誤偵測的判斷，並希望能夠藉此來改善人頭偵測的結果。我們以Mean-shift的群聚方式，以及隨距離調整的T值， T_H 固定為 400，測試在不同 T_L 值之下，追蹤對於偵測結果的改善。而所採用的追蹤的方法為容許偵測遺漏的追蹤。圖 4.14~4.16 為在離峰時段的比較，圖 4.17~4.19 則為在尖峰時段的比較。其中各個數據的計算均與 4.2 節中的定義相同。

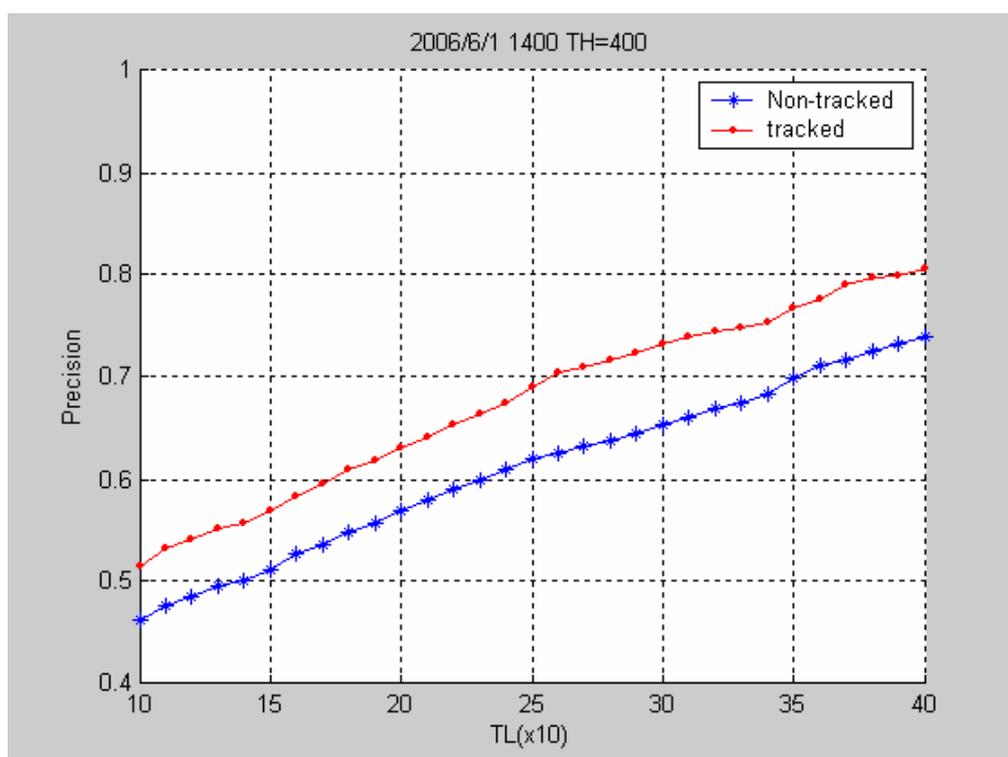


圖 4.14 在離峰時段下加入追蹤與未加入追蹤的偵測 Precision 比較

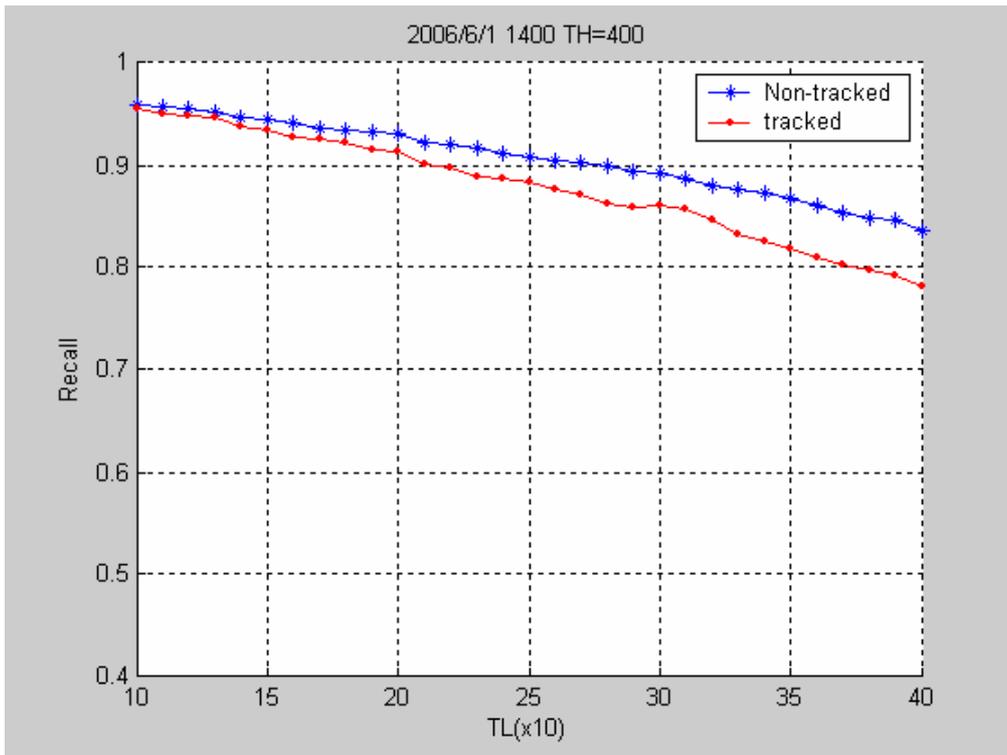


圖 4.15 在離峰時段下加入追蹤與未加入追蹤的偵測 Recall 比較

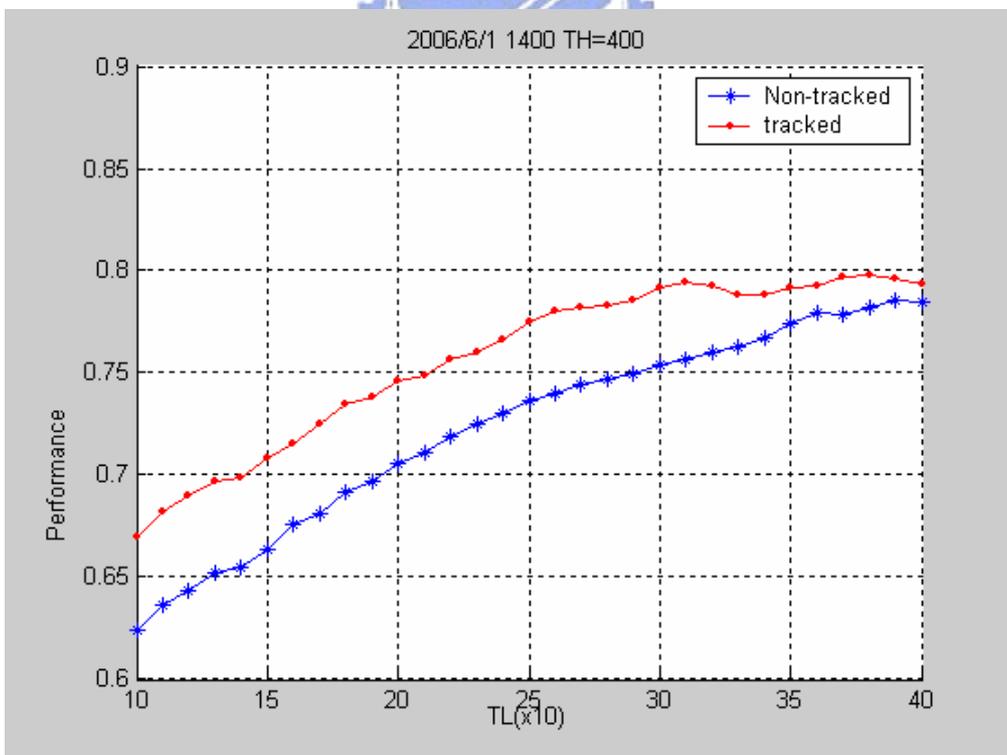


圖 4.16 在離峰時段下加入追蹤與未加入追蹤的 Performance 比較

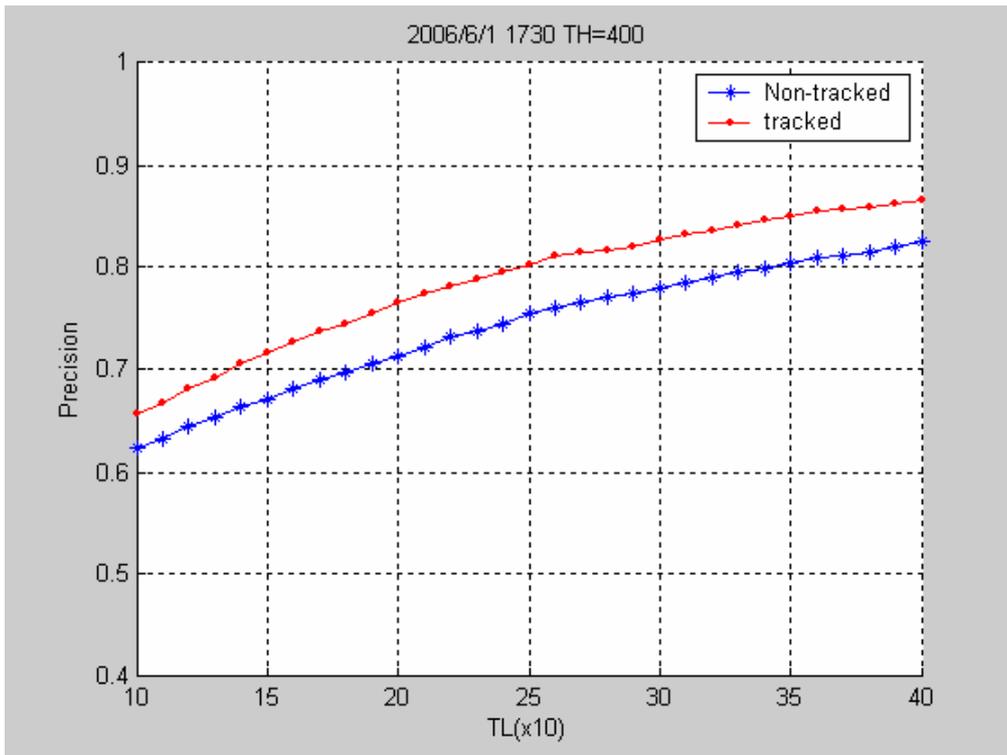


圖 4.17 在尖峰時段下加入追蹤與未加入追蹤的偵測 Precision 比較

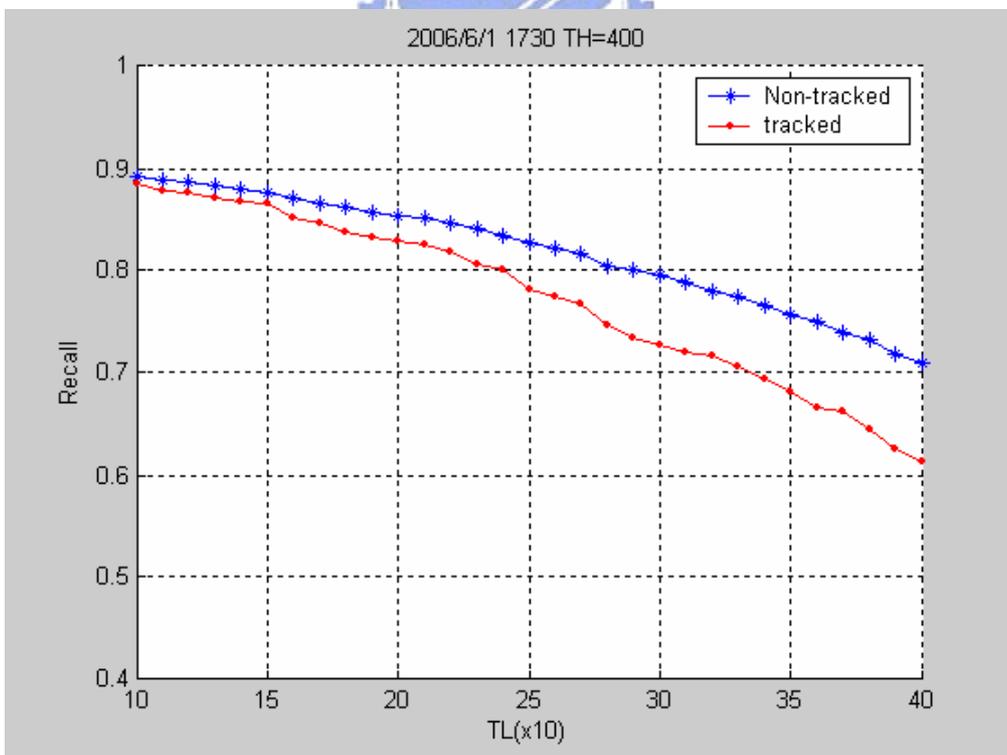


圖 4.18 在尖峰時段下加入追蹤與未加入追蹤的 Recall 比較

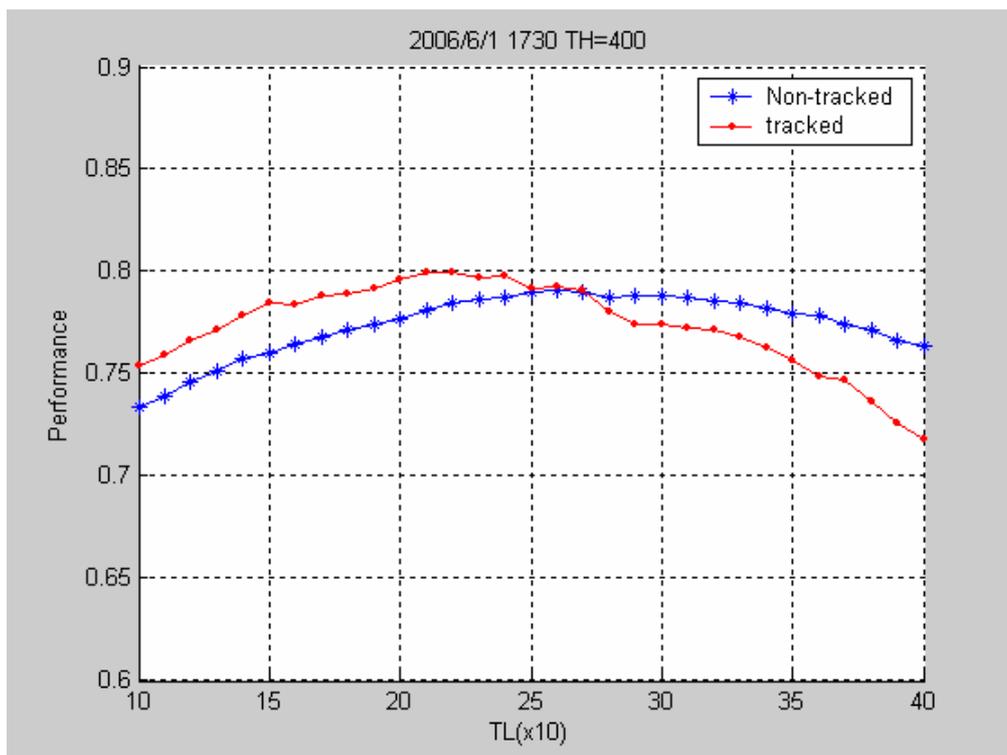


圖 4.19 在尖峰時段下加入追蹤與未加入追蹤的 Performance 比較

從圖 4.14~4.19 可以看出，加入了追蹤之後，人頭偵測的 recall 會下降，而且原本的 recall 越低，其下降的幅度也就越大。造成這種現象的原因是，因為偵測上的遺漏會造成追蹤失敗的發生，使得原本被認為是人頭的位置，因為前後沒有相似的結果與其連接，被認為是錯誤偵測而去除。而當原本的 recall 越低，代表其偵測上的遺漏越多，自然就容易造成追蹤失敗的發生。在圖 4.18 中可以看到，當原本的 recall 小於 80% 時，下降的幅度甚至會大到將近 10%。

而在 precision 方面則是均有提升，並且提升的幅度都差不多，約在 3%~5% 左右，顯示追蹤對於錯誤偵測的去除，有一定程度上的效果。就整體 performance 而言，在離峰時段方面，因為 recall 的下降幅度不大，故仍可維持比未加入追蹤時較好的 performance，而在尖峰時段時，由於 recall 的下降幅度超越了 recall 的提升，而造成整體 performance 的下降。是故雖然追蹤的確可以去除掉部分的錯誤偵測，但在原本的正确偵測太少的情況之下，反而可能造成判斷錯誤，使得許多的正确偵測也被去除，而得到更差的效果。

4.4 處理速度

表 4.7 為本章的實驗中，平均一張影像的處理時間統計。其中包括了讀取與儲存影像的 I/O 時間，以及前景偵測的處理時間。

表 4.7 一張影像的平均處理時間

Data	Only Detection(sec)	Detection & Tracking(sec)
2006/6/1 1400	0.1426	0.1545
2006/6/1 1730	0.1649	0.1865

由表 4.7 中可以看到，若只進行人頭偵測，平均每張所花的時間約為 0.14 ~ 0.16 秒。即使加入追蹤之後，也只需要 0.15 ~ 0.18 秒。由此可以看出本方法的處理速度非常快，每秒可處理約 5 張的影像。



第五章 實際應用：自動化的車站人潮監控系統

在本章中，我們將本論文的人群計數與追蹤方法，實作成為一自動化的車站人潮監控系統，並應用在實際的環境中。除可進行在攝影範圍內的人群計數外，並可透過追蹤來統計出入口的進出人數，進行大區域的人潮分析。

5.1 系統架構

此系統的架構如圖 5.1 中所示，在欲進行人潮監控的地點，以垂直視角的方式架設攝影機。一個或多個攝影機的訊號，連接至中央監控室的 PC，經由我們所設計的人群計數程式，分析各個攝影機所拍攝到的影像。一旦發現人數超過我們所設定的危險值，則發出警告訊息至警報系統。再由警報系統通知車站的管理人員，來進行進一步的處理。以上的過程為全自動化，完全不需人工的介入。

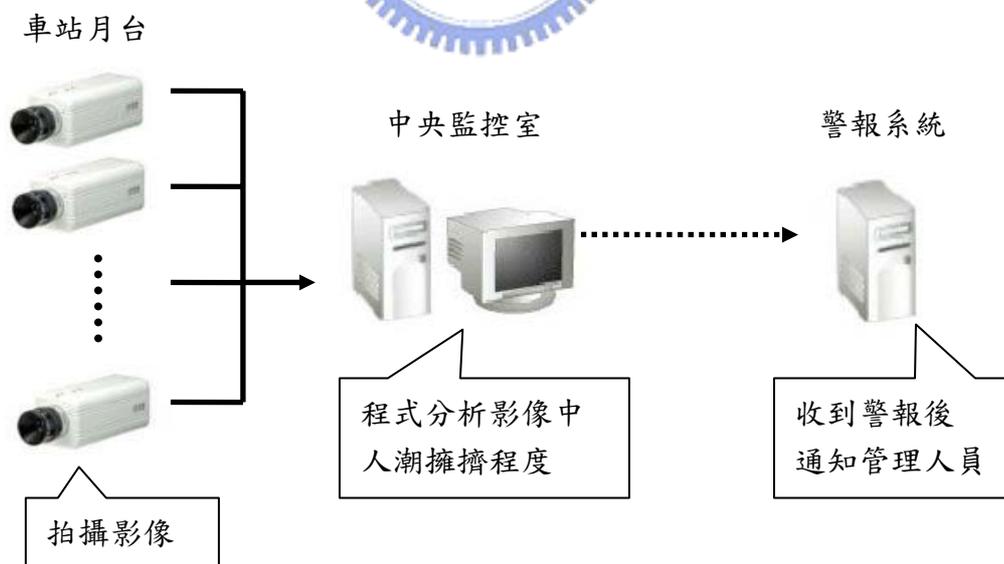


圖 5.1 自動化人潮監控系統架構圖

整個判斷人潮擁擠程度的流程如圖 5.2 所示。每隔固定時間，我們的人群計數程式便擷取攝影機的影像，先經由前景偵測，得到其前景影像。此處所使用的前景偵測方式與第四章中所述不同，而是經由本實驗室改良，更為精確的前景偵測方法，此方法的詳細介紹可參見[16]。得到前景部分的影像後，判斷前景部分在整張影像中所占百分比。若前景部分的百分比小於 70%，則必定不是擁擠的狀態，直接跳至結束。若大於 70%，則對於前景部分進行人頭的偵測。由於人頭偵測只能偵測到頭部出現在影像畫面中的人，對於頭部不在影像中的人則無法偵測到。故人頭偵測部份所得到的人數，我們將其稱為畫面中的人數。

為了處理頭部不在影像畫面中的人，我們將所偵測到的人頭位置，利用三度空間的關係，計算出其可能的身體部份，並將這些部份從前景中扣除。再統計剩餘的前景像素數量，並以一固定比例換算成人數(如 10000 像素/人)。這個換算的步驟，同時也可以減少因人頭偵測上的遺漏，所造成的人數誤差。經由換算前景像素所得到的人數，我們稱為畫面外的人數。

最後，我們將畫面中與畫面外的人數相加，得到一總人數。若此人數超過某一人數臨界值 P_{cr} ，即認為是擁擠，並且累計其連續次數。若一旦判斷為不擁擠，則此次數即歸零，重新開始累計。之所以要累計其次數的原因是，由於列車進站時，可能引發短時間的人潮擁擠，但這樣的人潮可能隨即自月台消失，造成危險的可能性較小。故我們希望預警的是月台上長時間的擁擠，而非短時間內的上下車人潮，避免使管理人員疲於奔命。若擁擠的累計次數超過一定值 X ，即在一段時間內人潮均是擁擠的，則將警告訊息發出至警報系統，再由警報系統通知管理人員，進行進一步的管理或疏散。

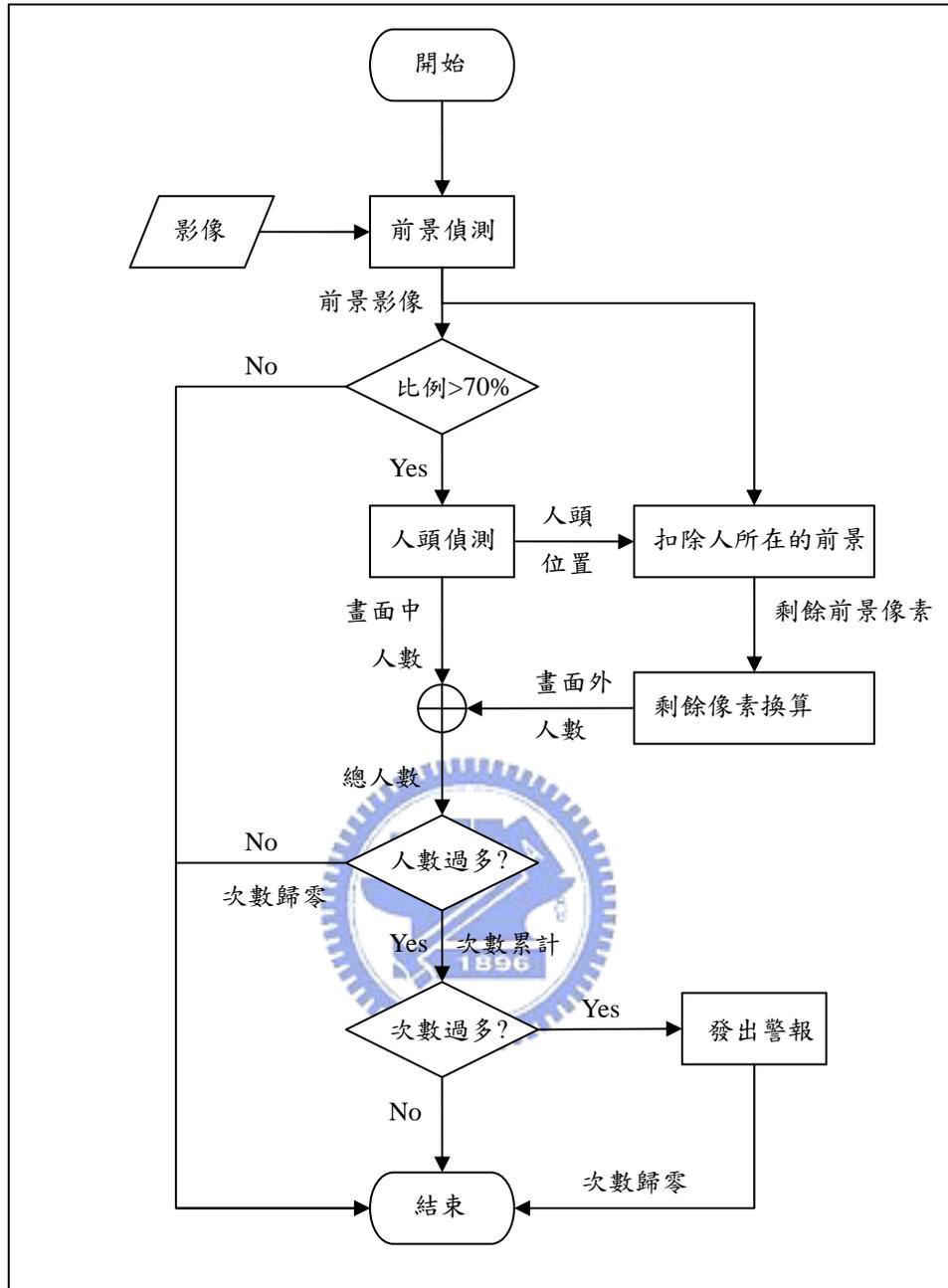


圖 5.2 人潮擁擠程度的判斷流程圖

我們共架設了四台攝影機，而各個攝影機所架設的位置，與本論文 4.1 中所描述的相同。由於每個攝影機所在位置的人潮密集程度不同，故我們對每個攝影機設定不同的 P 值，如表 5.1 所示。我們每 10 秒擷取一張影像，X 則定為 9 次，即一分半鐘。

表 5.1 各個攝影位置的 P_{cr} 值

Camera	P_{cr}
1	12
2	14
3	14
4	14

5.2 出入口的人潮分析

除了可進行攝影範圍內的人群計數之外，我們亦可透過追蹤，來分析欲監控區域內的進出人數，進行大區域的人潮分析。做法為將攝影機設置在與監控區域出入口的位置，如百貨公司電扶梯、列車門口等。則透過追蹤的方式，我們可以知道每個人的移動方向，即可計算進出此區域的人數。圖 5.3 為在進出口進行人潮分析的示意圖。

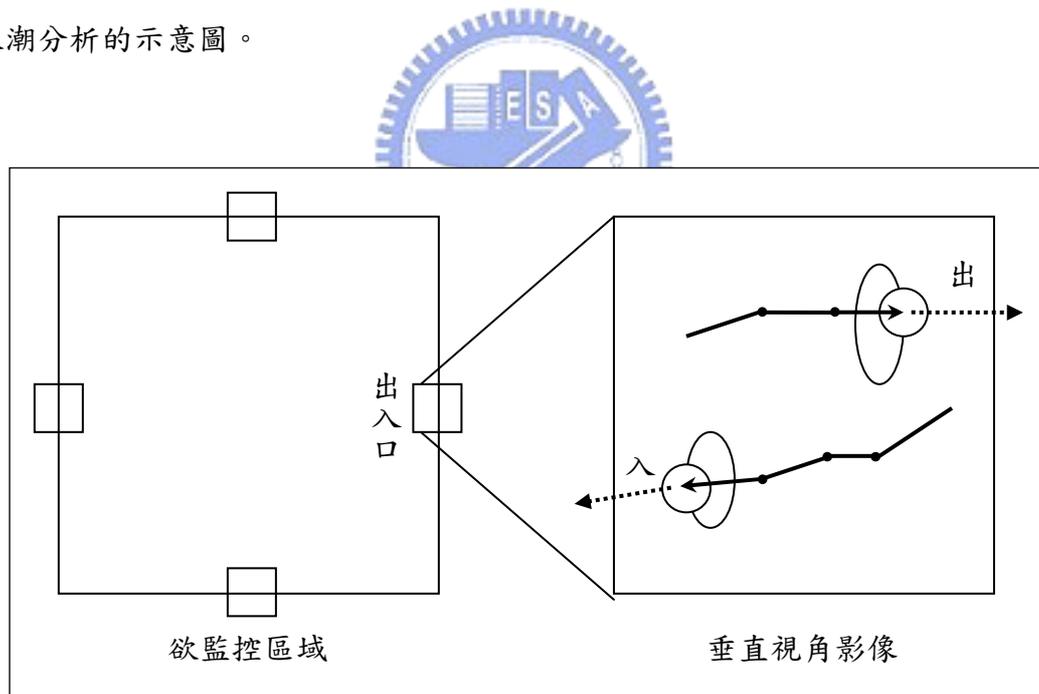


圖 5.3 進出口人潮分析示意圖

在圖 5.3 中，左半部為欲監控區域的鳥瞰圖，右半部則為出入口的攝影機所拍攝到的垂直影像。在此影像中，有兩條我們所追蹤到的路徑，其中一條是朝向區域外的方向，表示此人將自此區域離開，故將區域內人數減去一人。而另一條

的方向則指向區域內，表示此人將進入此區域，故將區域內人數增加一人。如此一來，我們不需要在欲監控區域的每一個地方均設置攝影機，而只要在出入口的部份設置攝影機。則不論欲監控區域的大小，我們均能計算出其任何時刻的人潮數量，並可進行統計，以供日後的管理參考之用。

此外，我們的系統並不只能夠應用在車站的人潮監控之上，亦可應用在許多人群經常聚集的場所，如機場、百貨公司、大賣場等。而所需的設備，只有各出入口的攝影機，以及一台普通的個人電腦，即可進行全自動化的人潮監控。其花費的成本相當低廉，並且可減少相當多的人力成本。



第六章 結論與未來展望

6.1 結論

在車站與賣場等人潮擁擠的環境中，透過單一攝影機的影像來分析人潮，由於人在行走交錯之間會產生互相遮蔽的效應，使得人群計數變得相當困難。為了降低遮蔽效應對於人群計數所造成的影響，本論文採用垂直視角來進行人群的計數與追蹤，利用人頭邊緣灰階梯度方向的放射趨勢，來偵測影像中人頭的位置及數量，並且利用色彩與移動路徑等特徵，進行以偵測為基礎的多人移動追蹤。

透過實驗發現，本論文的方法不論是在人潮稀疏或是擁擠的情況下，80%以上的人均可被正確偵測出來。顯示本論文的方法並不會受到遮蔽效應影響，在人潮擁擠的情況下依然適用。而在人群的移動追蹤方面，利用我們所定義的各項影像特徵，以及加入容許偵測遺漏的能力，85%以上的人均可被成功的追蹤。同時在一定程度的偵測 recall 之下，多人移動追蹤的方法也能夠剔除偵測結果中誤判的部份。

而我們更進一步的配合改良後的前景偵測方法，與本論文的方法整合成一自動化的車站人潮監控系統，在實際車站環境下運作。在系統中，我們以前景部分的像素換算成人數的方式，來逼近實際上的人數，希望能夠減少偵測遺漏所產生的誤差，以及補足僅有身體部份在畫面內的人數。

經由自動化的人潮監控系統，可取代傳統上由管理人員全天候坐在監視器前的監看方式，轉而以不需人工監看，在人潮擁擠時由監控系統自動通知管理人員。除了可將人力做更有效的運用外，也可避免人工監看時，因人員疲累所造成的疏忽。

6.2 未來展望

在本論文的方法中，我們假設人的髮色均為深色。故當禿頭或是白頭髮的人，甚至是戴淺色帽子的人出現，我們的方法可能就無法將其偵測出來。此外，由於我們是以圓形的深色區塊為人頭的模型，所以容易受到人的影子或是其他深色物體的影響，造成偵測上的錯誤較多。未來除了人頭形狀以外，可以考慮利用如紋理(texture)等其他特徵，來輔助人體的偵測，減少錯誤偵測的數量，提升人群偵測與追蹤的正確率。

前景偵測也是一個可加以改善的部份。由於人的偵測是針對前景的部分，若前景偵測的結果能夠更加準確，能夠完整的保留前景的資訊，同時降低背景中的影子以及其他物件的影響，對於人的偵測也會有很大的幫助。

雖然利用垂直視角的攝影，能夠避免掉遮蔽效應所帶來的影響，但是同時也犧牲了人的臉部以及其他身體部位的資訊，使得偵測與追蹤的困難度提高。不論是偵測或是追蹤方面，均還有提升的空間。



參考文獻

- [1] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "Pfinder: real-time tracking of the human body," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 19, Issue 7, pp.780 – 785, July 1997.
- [2] S. Birchfield, "An elliptical head tracker," *Signals, Systems & Computers, Conference Record of the Thirty-First Asilomar Conference*, Volume 2, 2-5, pp.1710 – 1714, Nov. 1997.
- [3] S. Birchfield, "Elliptical head tracking using intensity gradients and color histograms," *Computer Vision and Pattern Recognition, IEEE Computer Society Conference Proceedings*, pp.232 – 237, 23-25 June 1998.
- [4] T. Yang, Q. Pan, J. Li, Y. Cheng, and C. Zhao, "Real-time head tracking system with an active camera," *Intelligent Control and Automation, Fifth World Congress*, Volume 3, pp.1910 – 1914, 15-19 June 2004.
- [5] Y. Ishii, H. Hongo, K. Yamamoto, and Y. Niwa, "Face and head detection for a real-time surveillance system," *Pattern Recognition, ICPR Proceedings*, Volume 3, pp.298 – 301, 23-26 Aug. 2004.
- [6] Y. Ishii, H. Hongo, K. Yamamoto, and Y. Niwa, "Real-time face and head detection using four directional features," *Automatic Face and Gesture Recognition Proceedings*, pp.403 – 408, 17-19 May 2004.
- [7] X. Huang, L. Li, and T. Sim, "Stereo-based human head detection from crowd scenes," *Image Processing International Conference, ICIP*, Volume 2, pp.1353 – 1356, 24-27 Oct. 2004.

- [8] J. Garcia, N. da Vitoria Lobo, M. Shah, and J. Feinstein, "Automatic detection of heads in colored images," *Computer and Robot Vision Proceedings*, pp.276 – 281, 9-11 May 2005.
- [9] T. Zhao, and R. Nevatia, "Bayesian human segmentation in crowded situations," *Computer Vision and Pattern Recognition Proceedings, IEEE Computer Society Conference*, Volume 2, pp.II - 459-66, 18-20 June 2003.
- [10] H. Wu, Q. Chen, and M. Yachida, "Face detection from color images using a fuzzy pattern matching method," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 21, Issue 6, pp.557 – 563, June 1999.
- [11] I. Cohen, A. Garg, and T. S. Huang, "Vision-based overhead view person recognition," *Pattern Recognition Proceedings*, Volume 1, pp.1119 - 1124, 3-7 Sept. 2000.
- [12] M. Hail, A. Setli, W. Hua, and Y. Gong, "A detection-based multiple object tracking method," *Image Processing International Conference, ICIP*, Volume 5, pp.3065-3068, 24-27 Oct. 2004.
- [13] D. Comaniciu and P. Meer. , "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 24, Issue 5, pp.603-619, 2002.
- [14] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence* ,Volume 25, Issue 5, pp. 564-577, 2003.
- [15] R. Cucchiara, C. Grana, M. Piccardi, A. Prati, "Detecting moving objects, ghosts, and shadows in video streams," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 25, Issue 10, pp.1337 – 1342, 2003.
- [16] 聶政邦, "光影變化環境中的背景切割技術", 國立交通大學, 資訊科學與工程研究所, 碩士論文, 2006.

附錄 A 影像灰階梯度

影像灰階梯度(grey-level gradient)為一種常見的影像處理技術，經常被用於影像邊緣偵測(edge detection)以及影像切割(image segmentation)。其概念為利用影像中各像素(pixel)之間的灰階值(grey-level value)變化量，作為像素是否為影像中物體邊緣的依據。而計算變化量的方式，即為對像素灰階值進行一階微分(first derivative)。假設有一影像如圖 A.1(a)，其 X 軸的灰階值分佈為圖 A.1(b)所示，其一階微分的值即為圖 A.1(c)。

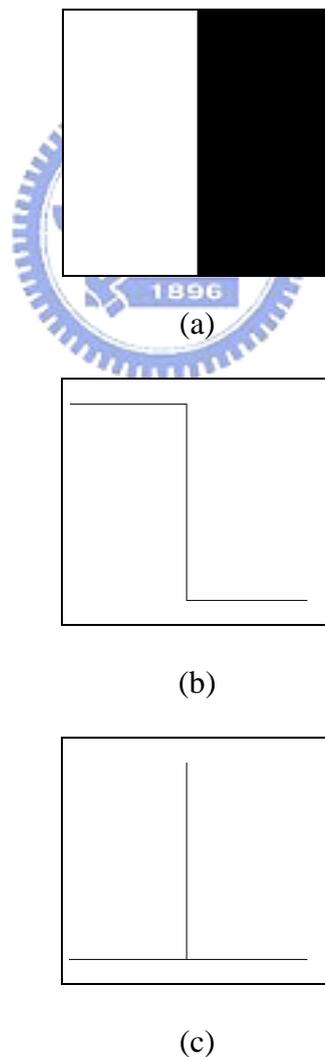


圖 A.1(a)原始灰階影像 (b)X 軸上的灰階值分佈 (c)X 軸上的灰階值一階微分

由圖 A.1(c)中可以發現，落在原影像中黑白交界邊緣處的像素，在對 X 軸的一階微分下得到特別大的值，其餘相同顏色的區塊內的像素對 X 軸的一階微分值則皆為零。故影像上某像素的灰階值變化量，可由其在水平方向(X 軸)與垂直方向(Y 軸)的一階微分值所組成的向量來描述，即為我們所謂的影像灰階梯度。

影像中位於點 (i, j) 的像素，其灰階梯度向量可以由下列式子表示：

$$\nabla f(i, j) = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \quad (\text{A.1})$$

其變化量的大小(magnitude)可以此向量的長度 $mag(i, j)$ 表示：

$$mag(i, j) = |\nabla f(i, j)| = \sqrt{(G_x + G_y)} \approx |G_x| + |G_y| \quad (\text{A.2})$$

其變化的方向(direction)可以此向量與 X 軸的夾角 $\alpha(x, y)$ 表示：

$$\alpha(i, j) = \tan^{-1} \left(\frac{G_y}{G_x} \right) \quad (\text{A.3})$$

由式子(A.1)來看，灰階梯度為點 (i, j) 對 X 方向及 Y 方向的偏微分，此計算方式建立在影像的資料為連續的基礎之上。但在數位影像的領域中，影像資料並不為連續的，而是以離散的方式來表示，我們無法以偏微分的方式來計算。故在離散的資料中，欲計算點 (i, j) 的灰階梯度，可以一 3 乘 3 大小的局部運算元(local operator)來逼近(approximate)，如圖 A.2。

-1	0	1
-1	0	1
-1	0	1

(a)

-1	-1	-1
0	0	0
1	1	1

(b)

圖 A.2 柏威運算元

(a)水平方向運算元 (b)垂直方向運算元

其中圖 A.2(a)為水平方向的運算元，圖 A.2(b)則為垂直方向的運算元，中心點的位置即為點 (i, j) 的所在位置。其概念為以點 (i, j) 及其鄰近各點的灰階差異值來做為點 (i, j) 的灰階變化量，又稱為柏威運算元(Prewitt operator)。設 $g(i, j)$ 為點 (i, j) 的灰階值，其 X 方向與 Y 方向向量的計算相當於下式所示：

$$\begin{aligned} G_x &= (g(i+1, j+1) + g(i+1, j) + g(i+1, j-1)) - (g(i-1, j+1) + g(i-1, j) + g(i-1, j-1)) \\ G_y &= (g(i-1, j-1) + g(i, j-1) + g(i+1, j-1)) - (g(i-1, j+1) + g(i, j+1) + g(i+1, j+1)) \end{aligned} \quad (\text{A.4})$$

而其大小與方向的計算則與式(A.2)、式(A.3)相同。

由於在柏威運算元中，鄰近各點的灰階值變化量與中心點的灰階值變化量均以相同的權重累加，如此容易受到鄰近各點的影響，使得位置靠近邊緣但並非邊緣的點，同樣也得到較大的灰階變化量。為了得到更精確的邊緣位置，之後便有索伯運算元(Sobel operator)的提出。索伯運算元在中心點附近的變化量給予較高的權重(weight)，目的為得到更精確的灰階變化量。



-1	0	1
-2	0	2
-1	0	1

-1	-2	-1
0	0	0
1	2	1

(a)

(b)

圖 A.3 索伯運算元

(a)水平方向運算元 (b)垂直方向運算元

圖 A.3 為索伯運算元，同樣分成水平與垂直兩種運算元，中心點的位置同樣為點 (i, j) 的所在位置。不同點在於中心點附近所使用的權重為鄰近各點的兩倍，將其展開寫成運算式即為：

$$\begin{aligned} G_x &= (g(i+1, j+1) + 2 \cdot g(i+1, j) + g(i+1, j-1)) - (g(i-1, j+1) + 2 \cdot g(i-1, j) + g(i-1, j-1)) \\ G_y &= (g(i-1, j-1) + 2 \cdot g(i, j-1) + g(i+1, j-1)) - (g(i-1, j+1) + 2 \cdot g(i, j+1) + g(i+1, j+1)) \end{aligned} \quad (\text{A.5})$$

由於能夠得到更精確的灰階變化量，索伯運算元被廣泛的使用來計算影像的灰階梯度。灰階梯度的值也成為影像邊緣偵測的特徵之一。由於落在邊緣之上的像素，其灰階梯度值的通常較非邊緣的像素為大，最簡單的方式即為以一臨界值(threshold)來與影像上各點的灰階梯度值做比較，較臨界值大的即被認為是影像中物件的邊緣。而灰階梯度中另一個特徵為灰階梯度的方向，此方向具有垂直於影像邊緣的特性。此特性也為本論文中用以作人頭偵測的主要特徵，我們在第三章中有詳細的說明。

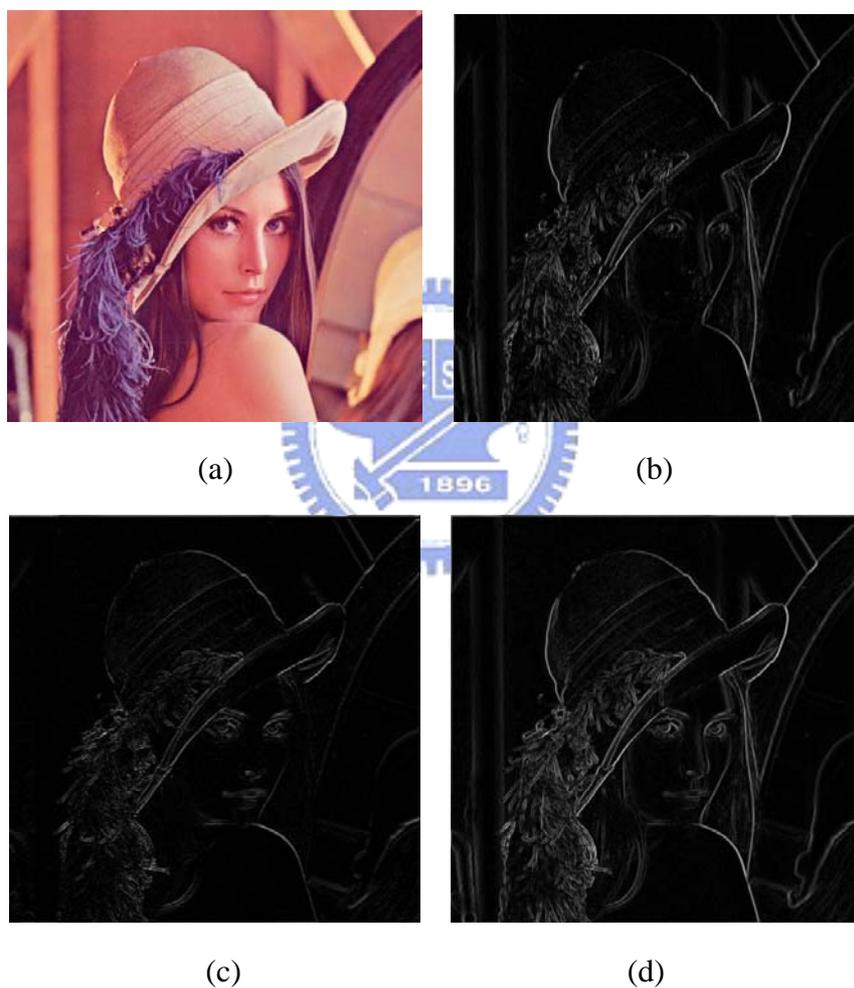


圖 A.4 (a)原始影像 (b)水平方向的灰階梯度值
(c)垂直方向的灰階梯度值 (d) 完整的灰階梯度值