

DT
448
0143.2
100
225288

使用 Fisher 線性判別分析之快速行人偵測系統

Fast Pedestrian Detection System using Fisher Linear
Discriminant Analysis

研究生：顏志鴻
指導教授：郭英哲 博士



A Thesis
Submitted to
Institute of Industrial Design
National Chin-Yi University of Technology
in Partial Fulfillment of the Requirements
for the Degree of
Master of Design
In

Electrical Engineering
國立勤益科技大學圖書館



225288

July 2011



Taiping, Taichung, Taiwan, Republic of China
中華民國一〇〇年七月

國立勤益科技大學

博碩士論文全文上網授權書

(提供授權人裝訂於紙本論文書名頁之次頁用)

本授權書所授權之論文為授權人在國立勤益科技大學
電機工程系 控制 組 99 學年度第 2 學期取得碩士學位之
論文。

論文題目：使用Fisher線性判別分析之快速行人偵測系統
指導教授：郭英哲

■ 同意

本人具有著作權之論文全文資料，非專屬、無償授予本人畢業學校
圖書館，不限地域、時間與次數，以微縮、光碟或數位化等各種方
式重製與利用，提供讀者基於著作權法合理使用範圍內之線上檢
索、閱覽、下載及列印。

論文全文上載網路公開之範圍及時間：

校內區域網路	■ 中華民國 101 年 8 月 22 日公開
校外網際網路	■ 中華民國 102 年 8 月 22 日公開

授權人：顏志鴻

簽名：顏志鴻

中 華 民 國 100 年 8 月 22 日

國家圖書館 博碩士論文電子檔案上網授權書

本授權書所授權之論文為授權人在國立勤益科技大學電機工程系 99 學年度第 2 學期取得碩士學位之論文。

論文題目：使用Fisher線性判別分析之快速行人偵測系統
指導教授：郭英哲

茲同意將授權人擁有著作權之上列論文全文（含摘要），非專屬、無償授權國家圖書館，不限地域、時間與次數，以微縮、光碟或其他各種數位化方式將上列論文重製，並得將數位化之上列論文及論文電子檔以上載網路方式，提供讀者基於個人非營利性質之線上檢索、閱覽、下載或列印。

上列論文為授權人向經濟部智慧財產局申請專利之附件或相關文件之一（專利申請案號：_____），請於 _____ 年 _____ 月 _____ 日後再將上列論文公開或上載網路。

因上列論文尚未正式對外發表，請於 102 年 8 月 22 日後再將上列論文公開或上載網路。

授權人：顏志鴻

親筆簽名及蓋章：顏志鴻

民國 100 年 8 月 22 日

電話：_____ 傳真：_____
聯絡地址：高雄市三民區鼎金後路532巷4號2F
E-Mail：likaobear@gmail.com

國立勤益科技大學
研究所碩士班
論文口試委員會審定書

本校電機工程系碩士班 顏志鴻 君

所提論文 使用 Fisher 線性判別分析之快速行人偵測系統

合於碩士資格水準，業經本委員會評審認可。

論文口試委員會：

召集人：郭景文

委員：陳永隆

郭景文

指導教授：郭景文

主任：姚賀騰

中華民國一〇〇年七月

摘要

近年來，行人偵測系統之研究多以影像處理技術來辨識影像中之行人位置。由於所採用的影像處理演算法需要大量的運算或需建立學習資料庫，因此現行之以影像為基礎的行人偵測系統在偵測常見的 320×240 解析度的影像時，通常需要耗費數秒來完成。因此，本論文提出了一個以單一攝影機擷取影像並以 Fisher 線性判別分析為基礎的快速行人偵測系統，其能夠即時正確執行行人偵測。

本研究之行人偵測系統以 Haar-like 特徵與共變異矩陣描述子作為特徵萃取的方式，並且整合積分影像的概念，使得特徵的計算能夠更快的進行。有別於一般使用分類器學習演算法的行人偵測系統，此行人偵測系統藉由 Fisher 線性判別分析的結果做為行人偵測系統判別的分類器，以簡單矩陣運算之判別方式可有效的縮減分類時所需要的時間。並且依據兩種不同的特徵，設計了兩層的 Fisher 分類器，前級使用 Haar-like 特徵的 Fisher 分類器能夠快速的在影像中篩選出行人可能存在的候選區域，而後級使用共變異矩陣描述子的 Fisher 分類器能夠針對這些候選區域再進一步的驗證是否真的有行人存在於該區域。

實驗結果顯示本研究之行人偵測系統在攝影鏡頭的解析度為 320×240 每秒 30 幀的影像資訊串流下，可以接近即時的處理，並且在複雜的背景下能有效率且正確的標示出行人所在的位置。

關鍵字：行人偵測、Fisher 線性判別分析、Haar-like 特徵、共變異矩陣描述子。

Abstract

In recent years, the pedestrian detection system usually uses a camera to capture the image of pedestrians and allocate the pedestrians in the image with the image processing technologies. But the adopted image processing algorithm in recent researches is time consumption and needs to set up a huge database of pedestrian images. It makes the detection system hard to be implemented. Therefore, this paper proposed a fast pedestrian detection system with single video camera to detect pedestrians in the image correctly and effectively.

The proposed detection algorithm employs Haar-like features and covariance matrix descriptor to describe pedestrians. A technique employing the integral image is also applied in order to improve the efficiency of both the computations of the Haar-like feature and covariance matrix descriptor. Our proposal is based on the Fisher linear discriminant analysis (FLDA) which can classify pedestrians rapidly. In order to further accelerate the proposed detector, the two-layer cascade classifiers are adopted. The former (the classifier based on Haar-like feature) can quickly select candidate regions where pedestrians may present, and the latter (the classifier based on covariance matrix descriptor) can accurately determine whether there are pedestrians in the candidate regions or not. If a region passes the two-layer cascade classifiers, it would be sure to contain the pedestrians.

For the experiments, the test set of video sequences is taken from real scenes, using the video stream in 30 frames per second sized 320×240 pixels resolution. The results of experiments show that the system is fast and could be used in real time.

Keyword: Pedestrian detection, FLDA, Haar-like feature, covariance matrix descriptor.



誌謝

這篇論文的完成首先要感謝的是指導教授郭英哲老師，在大學的專題生時期以及這兩年的碩士生涯中悉心的指導，自由但不放任的管理風格使我在研究上有許多自由發展的空間，舒適的研究環境及豐富的資源使我能夠更專注於研究中並順利的取得碩士學位。在此，同時也感謝三位口試委員暨南大學的郭耀文教授、台中技術學院的陳永隆教授與本校的陳文淵教授對本篇論文的指正與建議，使得本論文更臻於完善。

感謝實驗室的林華峰大學長在我接觸嵌入式系統時的指導與幫助，使我能夠快速的熟悉嵌入式平台，讓日後的大學專題、研究所與老師一同進行的計畫甚至是找工作都非常的順利。感謝李彥鋒學長所留下來的 PPT 範本還有實驗室 Logo，最受用的還是學長那精益求精的做事態度，讓我受益良多。感謝蔡政育學長在我進碩士班前給的建議與心得分享。感謝實驗室的學長陳璿文、徐溫嶺，於碩一期間在專業領域上的細心指導與做實驗的心得討論。感謝實驗室同學張為彥在學業上的切磋與建議，以及這兩年來提醒我許多重要事件的時程。感謝實驗室碩一學弟揚子儀這兩年來協助打理實驗室中大大小小的事務。感謝隔壁實驗室的同學張世平、賴俊源，學弟陳建旻、游宏裕，樓上實驗室的張維仁、鄭亞倫，學弟莊倍嘉、魯浚璋，三步五時的聚在一起聊天、吃飯、宵夜，也在課餘的時間打羽球、魔獸，讓生活增添不少樂趣。

最後要感謝我的母親在我碩士兩年來的支持與鼓勵，讓我無後顧之憂的完成碩士學位。感謝我的妹妹在我不常回家的兩年中總是盡量的抽空回家陪伴母親照看家裡的情況。

顏志鴻 謹誌 中華民國一〇〇年七月二十二日

目錄

摘要	i
Abstract	ii
誌謝	iv
圖目錄	vii
表目錄	ix
第一章 緒論	1
1.1 研究背景	1
1.2 研究動機與目的	2
1.3 系統架構	3
1.4 論文貢獻	5
1.5 論文架構	5
第二章 相關研究	6
2.1 概述	6
2.2 特徵萃取方法探討	7
2.2.1 Haar-like 特徵	9
2.2.2 梯度方向直方圖	9
2.2.3 共變異矩陣描述子	11
2.3 分類器架構	12
2.3.1 AdaBoost	12
2.3.2 支援向量機	13
2.4 行人偵測之系統架構探討	15
第三章 Haar-like 特徵	18
3.1 概述	18
3.2 特徵模板	18
3.3 偵測窗中的特徵模板	20
3.4 積分影像	21
3.5 利用積分影像計算 Haar-like 特徵	23
第四章 共變異矩陣描述子	26
4.1 概述	26

4.2 共變異矩陣	26
4.3 用於行人偵測的特徵向量(隨機向量 X).....	27
4.4 共變異矩陣描述子	28
4.5 利用積分影像計算共變異矩陣描述子	31
第五章 Fisher 線性判別分析	34
5.1 概述	34
5.2 Fisher 線性判別分析	36
5.3 Fisher 分類器	39
第六章 行人偵測系統	41
6.1 概述	41
6.2 行人影像資料庫	41
6.3 特徵萃取及萃取區域選擇	44
6.3.1 Haar-like 特徵之萃取範圍選取	44
6.3.2 共變異矩陣描述子之萃取範圍選取	49
6.4 Fisher 分類器的訓練流程	50
6.4.1 基於 Haar-like 特徵的 Fisher 分類器之訓練流程	50
6.4.2 基於共變異矩陣描述子的 Fisher 分類器之訓練流程	52
6.5 串級的行人偵測系統	56
6.5.1 以 Haar-like 特徵為基礎的候選區域篩選模組	57
6.5.2 以共變異矩陣描述子為基礎的行人分類模組	58
6.5.3 偵測結果之標示區域的整併	60
第七章 實驗結果與分析	62
7.1 概述	62
7.2 行人偵測系統效能評估實驗	62
7.3 系統實際運行測試	66
第八章 結論及未來展望	70
8.1 結論	70
8.2 未來展望	70
參考文獻	72

圖目錄

圖 1.1 駕駛者補助系統(DAS)的架構	2
圖 1.2 行人偵測系統架構	4
圖 2.1 模板識別技術流程	6
圖 2.2 將小方格的梯度方向轉換成 9 維的特徵值	10
圖 2.3 特徵空間中分隔兩個群組的超平面[14]	14
圖 2.4 利用核函數將資料映射到高維度空間找出超平面[14]	15
圖 2.5 Paisitkriangkrai 所做的行人偵測系統效能評比[6]	16
圖 3.1 常見的 Haar-like 特徵模板	18
圖 3.2 人臉偵測的 Haar-like 特徵模板[15]	19
圖 3.3 所有雙矩形特徵在 3×3 偵測窗中的形態	20
圖 3.4 以兩個點來表示影像中的一個矩形區域	22
圖 3.5 一張 5×5 灰階影像之積分影像的計算	23
圖 3.6 參考四個點計算矩形區域 D 中的像素值加總	23
圖 3.7 參考六個點計算雙矩形 Haar-like 特徵	24
圖 4.1 原始影像轉換至各個特徵影像分量	28
圖 4.2 共變異矩陣 C_R 所描述之矩形區域	29
圖 4.3 利用積分影像計算共變異矩陣描述子	33
圖 5.1 若將樣本群任意的投影到 X_1 軸或 X_2 軸則無法完整分離	35
圖 5.2 經由最佳投影向量投影後的樣本群	35
圖 6.1 INRIA 資料庫正樣本訓練集的部分影像	42
圖 6.2 INRIA 資料庫負樣本訓練集的部分影像	43

圖 6.3 行人特徵影像模型	45
圖 6.4 垂直、水平以及對角方向的濾波器	45
圖 6.5 行人影像模型之垂直、水平以及對角梯度強度影像	46
圖 6.6 二值化後的梯度強度影像	46
圖 6.7 本論文所使用的三種特徵模板	47
圖 6.8 本論文使用的 5 個垂直雙矩形特徵	47
圖 6.9 本論文使用的 6 個水平雙矩形特徵	48
圖 6.10 本論文使用的 12 個四矩形特徵	48
圖 6.11 共變異矩陣描述子的萃取範圍	49
圖 6.12 基於 Haar-like 特徵的 Fisher 分類器之訓練流程	51
圖 6.13 正樣本的共變異矩陣描述子計算流程	53
圖 6.14 負樣本的共變異矩陣描述子計算流程	54
圖 6.15 基於共變異矩陣描述子的 Fisher 分類器訓練流程	55
圖 6.16 以偵測窗擷取輸入影像的資訊	56
圖 6.17 以 Haar-like 特徵為基礎的候選區域篩選模組	57
圖 6.18 以共變異矩陣描述子為基礎的行人分類模組	59
圖 6.19 偵測結果之標示資訊的合併流程	60
圖 7.1 基於 Haar-like 特徵的 Fisher 分類器性能曲線	63
圖 7.2 本文之 Fisher 分類器及[3]的 AdaBoost 分類器之性能比較	64
圖 7.3 行人偵測系統在一般情況的偵測結果	67
圖 7.4 行人偵測系統在光線不足時的偵測結果	68
圖 7.5 行人偵測系統在光線不均勻時的偵測結果	68
圖 7.6 行人偵測系統在行人重疊及靠近時的偵測結果	69

表目錄

表 7.1 行人偵測系統效能評估實驗之運行環境.....	65
表 7.2 比較特徵萃取方式與分類器搭配之速度.....	65
表 7.3 系統實際運行結果之實驗環境.....	66
表 7.4 行人偵測系統執行時間.....	66



第一章 緒論

1.1 研究背景

根據世界衛生組織(World Health Organization)的報告[1]，全世界每年交通意外造成了近 120 萬件死亡事件及 5,000 萬件損害。在所得較低的國家，主要的死亡事件並非車輛駕駛人，反而是行人、自行車駕駛、及其他小型車乘客。在高所得的國家行人的死亡交通事故雖相對較低，但也耗費很大的社會及經濟成本。分析其肇事原因，百分之九十七以上是駕駛不當所致。這些交通事故不僅造成個人及家庭的負擔，更是社會有限醫療資源及國家生產力的損失。有鑑於此，近年來有越來越多產學專家進行研究如何增進「行」的安全。除此之外，也希望提升「行」的效率及便利性。

智慧型運輸系統(Intelligent Transportation System, ITS)利用先進的電子、通信、資訊與感測等應用技術，整合人(駕駛、行人、系統管理人)、路、車輛的管理機制。利用即時資訊提高運輸系統的安全與道路運輸效率，同時也減少交通對環境的負面影響。達到 (1)減少交通事故，提升行車安全。(2)減少空氣、噪音污染，提高能源使用效率。(3)降低交通擁擠，提高運輸機動性。(4)促進相關產業發展，增加就業機會等目標。

而駕駛者輔助系統(Driver Assistance System, DAS)即以加強車輛行駛安全，減少交通事故為目的。如圖 1.1 所示，利用安裝於車上的各種感測器(通常使用攝影機、雷達、紅外線)，收集車內與車外的環境數據資料，再經由資訊的處理核心，讓駕駛者能在最快的時間內察覺可能發生的潛在危險，達到危險規避確保行駛安全，進而避免意外事故發生。而行人偵測(pedestrian detection)是電腦視覺研究領域中一項關鍵的技術，其在智慧型車輛、監測系統與先進機器人中都扮演著重要的角色，尤其是在智慧型車輛的駕駛者輔助系統中，行人偵測是繼車輛偵測之後的

另外一個重要且急需發展的環節。

行人偵測在駕駛者輔助系統應用中，近年來的研究多以攝影機擷取車輛前方影像，進而以影像處理技術辨識影像中之行人位置。由於所採用的影像處理演算法則需要大量的運算量或需建立學習資料庫，因此多以個人電腦來實作完成。本論文考量行人偵測系統在未來產品化可行性，嘗試降低行人偵測系統運算過程的複雜度以達到更快的值行速度，目的是希望在未來能夠以嵌入式系統來實現行人偵測系統。除可明顯降低成本、大幅縮小體積，並擁有友善的觸控式圖形化操作介面，也能夠與車內多媒體視聽娛樂系統及全球定位系統(Global Positioning System, GPS)整合。



圖 1.1 駕駛者補助系統(DAS)的架構

1.2 研究動機與目的

儘管地球即將面臨能源耗盡的危機，但汽車的數量仍以驚人的速度成長。汽車數量的成長所帶來的影響，除了考驗運輸系統外，交通事故的增加更是不可忽視的。車輛駕駛人在車輛行駛中常常會因為來自車內或是車外的一些干擾，如音樂、交談或是車外的廣告看板等，將視線由正前方移開。如果此時有突發事件發

生，駕駛人通常無法有效的做出正確的判斷，以致交通事故發生。駕駛者輔助系統的主要功能，就是在突發事件發生前便主動的追蹤可能引發事件的因素，在找到引發因素時事先提醒駕駛人避開可能發生的突發事件。

現行基於電腦視覺的行人偵測系統的研究[2-6]，在其實驗結果中皆擁有不錯的辨識率，但即使在 PC 上運作，它們在偵測 320×240 像素的影像時大多需要數秒鐘甚至是數分鐘才能完成，在未來若想將其置於系統資源有限的車用嵌入式系統上，必定要大大的提升其運算效率才有可能實現。提升基於電腦視覺的行人偵測系統的運算速度不外乎提升硬體效能或者是降低演算法的複雜度，若是以『在未來能夠以嵌入式系統來實現行人偵測系統』為目標的話，降低演算法的複雜度似乎較為可行，因此降低基於電腦視覺的行人偵測演算法的複雜度以及組合出更有效率的行人偵測演算法，是本論文中的重要議題。

1.3 系統架構

本論文所提出的行人偵測系統主要由兩個模組所構成 (1) 候選區域篩選模組、(2) 行人分類模組，候選區域篩選模組最主要的工作為掃描整張輸入影像並找出行人可能存在的候選區域，隨後再由行人分類模組對這些候選區域進行更精確的分類辨別，如圖 1.2 所示。

(1) 候選區域篩選模組：

- a. Haar-like 特徵計算單元：功能為萃取輸入影像的數個 Haar-like 特徵[7-9]，並且將這些特徵組合成一個向量供分類器使用。
- b. Fisher 分類器：由 Fisher 線性判別分析[10]之結果所構成的分類器，其在候選區域篩選模組主要的功能是分析 Haar-like 特徵計算單元所傳來的特徵值，並且標示出影像中行人可能存在的候選區域。

(2) 行人分類模組：

- a. 共變異矩陣描述子計算單元：功能為將候選區域中的的影像資訊轉換為共變異矩陣描述子(covariance matrix descriptor)[11]供分類器使用。
- b. Fisher 分類器：由 Fisher 線性判別分析[10]之結果所構成的分類器，其在行人分類模組主要的功能是分析共變異矩陣描述子計算單元所傳來的特徵值，並且進一步的判斷從候選區域篩選模組傳送來的候選區域是否真的有行人存在。

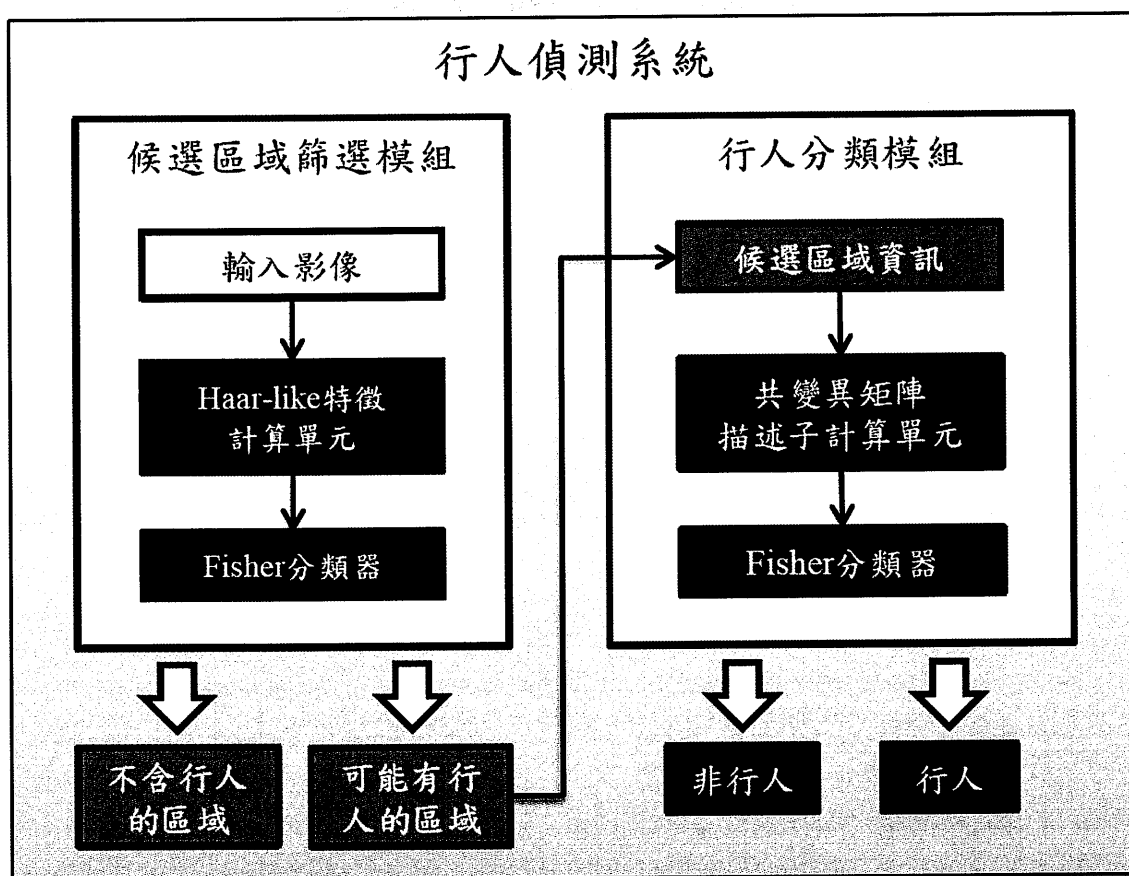


圖 1.2 行人偵測系統架構

1.4 論文貢獻

本論文利用單一個 CCD 攝影鏡頭實現基於影像處理的快速行人偵測系統，並且以『在未來能夠以嵌入式系統來實現行人偵測系統』為目標，盡可能的降低基於影像處理之行人偵測演算法的複雜度。本行人偵測系統先以 Haar-like 特徵做為候選區域篩選模組的特徵萃取方法，使分類器得以快速的找出行人可能存在的候選區域，再使用強健性極高的共變異矩陣描述子使分類器能夠精準的辨識這些候選區域是否有行人存在，並且結合積分影像(Integral Image)的概念使得 Haar-like 特徵及共變異矩陣描述子的計算能夠更有效率的進行，最後以 Fisher 線性判別分析取代傳統的分類器演算法使得特徵的判別能夠以更快的速度進行。

1.5 論文架構

本論文共分為七個章節，第一章為緒論，主要是描述目前在基於電腦視覺的行人偵測的環境以及背景，並且介紹系統架構、論文貢獻...等。第二章為相關研究，主要是針對行人偵測的方法進行分析以及探討。而第三章至第五章為本論文所使用的特徵萃取演算法以及分類器演算法的詳細介紹。第六章為行人偵測系統也是本論文的核心，其詳細的介紹行人偵測系統的訓練以及辨識流程。第七章為實驗結果與分析，詳細的說明本論文的行人偵測系統之結果。第八章為本論文的結論。最後是本論文的參考文獻。

第二章 相關研究

2.1 概述

目前在視覺為基礎的行人偵測相關研究大多是以模板識別(pattern recognition)技術為主要的偵測方法，其方法大致上可分為兩個最主要的步驟：(1)特徵萃取(feature extraction)。(2)分類器架構(classifier construction)。如圖 2.1 所示，在特徵萃取中主要的特徵值皆萃取自大量的樣本，而這些特徵值將會被使用來進行分類器的學習與訓練，並且在進行測試時，訓練完成的分類器將會掃描整張輸入影像，通常這樣的過程皆能有效的在影像中辨識出特定的物件，因此模板識別技術被普遍的應用在各個領域的應用中，像是人臉偵測[7][8]、行人偵測[2-6]以及物件偵測與分類[9][10]。因此，本章節將以特徵萃取的方法以及分類器的架構兩個大方向來進行相關研究及文獻的探討。

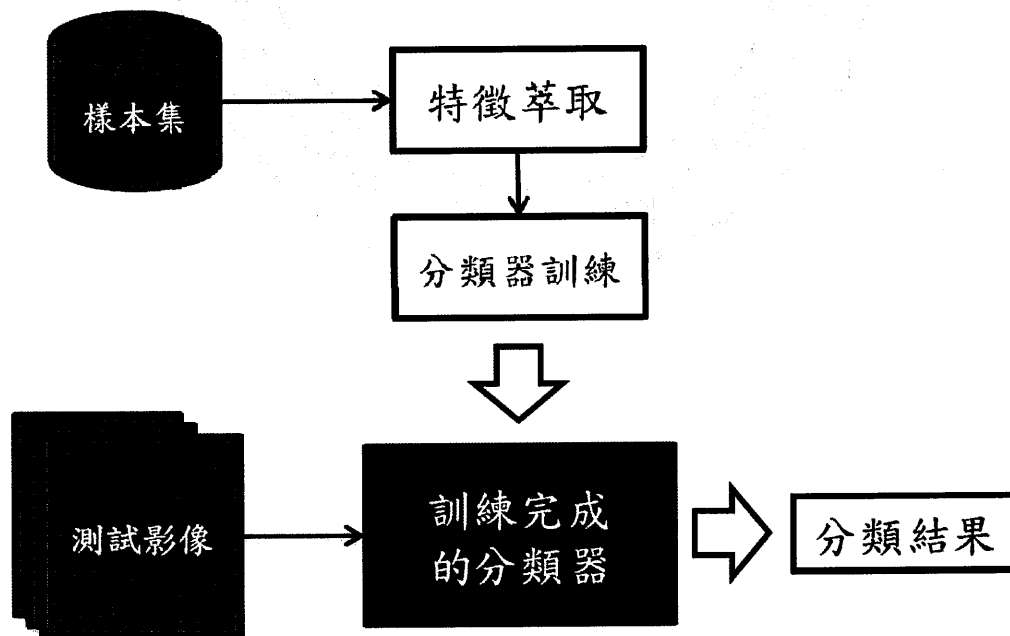


圖 2.1 模板識別技術流程

2.2 特徵萃取方法探討

在模板辨識技術中行人偵測要比一般的物件偵測(例如：車輛、人臉...等)困難得多，主要的原因為：(1)人體並非剛體物件，其會隨著姿勢的改變使得外觀有非常廣泛的變化、(2)人們的衣著可能是任何的顏色以及任何形式的紋理，沒有任何的規則可循、(3)人們所處的環境通常伴隨著凌亂的背景。因此實現以電腦視覺為基礎的行人偵測系統其首要的挑戰為『該如何有效的描述行人在影像中獨有的特徵』。

而這些問題在眾多文獻中都有提及該如何在影像中找行人獨有的特徵，其依據特徵的計算方式大致上可以分成(1)移動基礎(motion-based)、(2)形狀基礎(shape-based)：

- (1) 移動基礎：在移動基礎的特徵計算方面需要使用一些影像處理的前級處理技術來找出行人在影像中的位置，例如：在影像串流中將目前的影像與前一張影像以減算法計算兩張影像之間的差異找出移動中的行人。通常這樣的方法都會配合其它的前級處理技術來使用，像是形態學(morphology)影像處理中的侵蝕(erosion)與膨脹(dilation)以及影像分割技術中的邊緣檢測器(edge detector)等。本論文的行人偵測系統其最終目的是應用於車輛上的駕駛者輔助系統，而在車輛上除了行人之外，背景亦會隨著車子的移動不斷的改變，以前級處理技術來說根本無法萃取出有效的資訊，因此移動特徵不適用於本論文所提出的行人偵測系統中。
- (2) 形狀基礎：形狀基礎的特徵計算方式通常是跳過影像的前級處理過程，直接使用特定的演算法對行人的人體或姿勢的影像資訊做編碼，像是[2]中的梯度方向直方圖(Histogram of Oriented Gradient, HOG)、[3]中的共變異矩陣特徵(covariance feature)...等。即使背景不斷改變的情況下也不影

響其萃取特徵的能力，因此本篇論文所提出的行人偵測系統所採用的特徵萃取方式皆為形狀特徵。

特徵依其測量的規模又可分為(1)全域特徵(global feature)以及(2)區域特徵(local feature)，全域特徵與區域特徵最大的不同在於全域特徵是對整張輸入影像做運算，而區域特徵則是對輸入影像中某些特定的子區域做運算：

- (1) 全域特徵(global feature)：全域特徵的特徵萃取範圍通常是一整張影像，其較具代表性的例子為主成份分析(Principal Component Analysis, PCA)，因為 PCA 能夠有效的壓縮資料的維度，所以被廣泛的應用在各個領域中。但全域特徵使用於行人偵測並不能夠有效的描述行人影像的資訊，因為全域特徵在物件的外觀、姿勢以及光線有較大的變化時便會失去強健性，以致無法萃取出有效的資訊。
- (2) 區域特徵(local feature)：區域特徵的特徵萃取範圍通常是影像中目標物件的一小部分，例如：行人的全身、上半身、下半身、上肢以及腳...等。且區域特徵對於物件的外觀、姿勢以及光線的改變不會像全域特徵那樣的敏感，主要是因為區域特徵是萃取自原始影像中的小區塊，而屬於區域特徵的特徵萃取方式有：梯度方向直方圖[2]、區域共變異矩陣(region covariance)[11]以及[7-9]中的小波係數(wavelet coefficient)...等。

綜上所述在特徵萃取方法探討的這個章節中，將會以形狀以及區域的特徵萃取方式做為主要的探討方向，其包括 Haar-like 特徵、HOG 以及共變異矩陣描述子。

2.2.1 Haar-like 特徵

Haar-like 特徵[8]最早由 Viola 與 Jones 在 2001 年提出，並在 2004 年將其應用於人臉偵測中[7]，此特徵萃取方法有著計算速度快以及精確度相對較高的優點，因此後來被廣泛的應用在電腦視覺領域中的特徵描述方法，雖然此方法可以單獨的使用於模板識別，像是人臉與行人偵測[9]、人臉偵測[7]，但用於行人偵測中通常是將其使用於系統的前級，並且配合其他的演算以增加辨識的成功率以及速度，像是[3]中所提到行人偵測系統的架構。

Haar-like 特徵是一種簡單的矩形特徵，以雙矩形的 Haar-like 特徵為例，其定義為計算影像中相鄰矩形面積內的像素值總和之差值，並能夠以任何大小位置放置於原始影像中，此外 Viola 與 Jones 也定義了三矩形以及四矩形的 Haar-like 特徵[7]。每種不同類型的 Haar-like 特徵即代表影像中存在(或不存在的)特點，像是邊緣(edge)或紋理(texture)的變化等。Haar-like 特徵萃取的步驟以及細節在第三章會有詳盡的說明。

2.2.2 梯度方向直方圖

梯度方向直方圖(Histogram of Oriented Gradient) [2]是由 Navneet Dalal 與 Bill Triggs 在 2005 年所提出，梯度方向直方圖顧名思義就是使用影像的梯度資訊配合統計學中的直方圖來描述一張影像中的資訊，其計算步驟為：

1. 將影像分成小範圍的 8×8 方格(cell)[2]。
2. 然後以一個小方格為單位，計算方格中每個像素的梯度方向(orientation)以及梯度強度(magnitude)，如(2.1)至(2.4)所示。

$$dx = I(x + 1, y) - I(x - 1, y) \quad (2.1)$$

$$dy = I(x, y + 1) - I(x, y - 1) \quad (2.2)$$

$$m(x, y) = \sqrt{dx^2 + dy^2} \quad (2.3)$$

$$\theta(x, y) = \tan^{-1} \left(\frac{dy}{dx} \right) \quad (2.4)$$

其中 $m(x, y)$ 與 $\theta(x, y)$ 分別為原始影像 $I(x, y)$ 的梯度強度與梯度方向。

- 將小方格內所有的梯度方向以 9 個基礎方向表示，並以直方圖統計後轉化成一個 9 維的特徵值。如圖 2.2 所示。

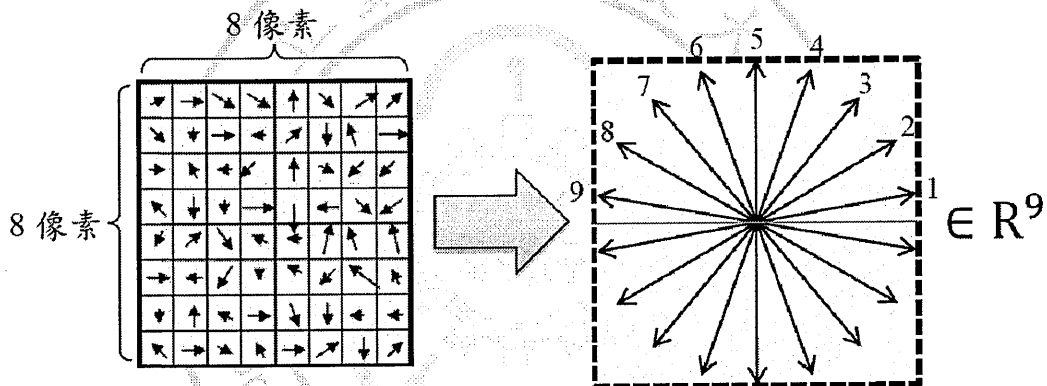


圖 2.2 將小方格的梯度方向轉換成 9 維的特徵值

- 以每 4 個小方格聯合成一個區間(block)，並且為了提高精確度將其做對比度正規化(contrast-normalized)的計算，最後取得 36 維的梯度方向直方圖(HOG)描述子。

由於梯度方向直方圖是以影像的局部範圍(小方格, cell)為運算的單位，因此與其它的特徵描述法相比的話，其對於物件外觀、姿勢以及光線的改變都能維持著

較好的強健性，因為這些改變只會出現在較大的範圍中。而且只要行人保持著立姿，HOG 描述子甚至可以容許行人有一些不同的肢體動作，這些動作對於 HOG 描述子來說都只是細微的改變並不影響偵測結果，因此 HOG 可以說是用於行人偵測最理想的特徵描述方法。

2.2.3 共變異矩陣描述子

共變異矩陣描述子(covariance matrix descriptor) 又稱之為區域共變異數(region covariance)[11]是一種區域性的描述子(region descriptor)，由 Tuzel 與 Porikli 等人在 2006 年提出並且應於物件偵測(object detection)以及紋理辨識(texture classification)，其概念為使用共變異矩陣來描述一個感興趣區域內的影像資訊，任何大小的感興趣區域其輸出的結果將會是一個固定大小的 $n \times n$ 矩陣，而該矩陣內所帶有的資訊則是感興趣區域內之像素的 n 種特徵影像之間的交互關係，且這些特徵影像可以是像素的 x 座標、 y 座標、一階偏導數、二階偏導數、梯度方向、梯度強度...等。

由於共變異矩陣描述子是以感興趣區域的數個特徵影像為特徵萃取的基礎，並且也是以影像的局部範圍為描述對象，故此特徵描述方法對於物件外觀、姿勢以及光線的改變都能維持相當好的強健性。雖然其在[3]的實驗結果中，準確度略低於 HOG 描述子，但因為共變異矩陣描述子可以藉由積分影像的概念大大的增加其運算的速度，因此若考慮行人偵測之特徵描述方法的運算速度以及強健性，那麼共變異矩陣描述子是非常適合的選擇。共變異矩陣描述子的計算步驟以及細節在第四章會有詳盡的說明。

2.3 分類器架構

在行人偵測中常用的分類器架構主要有兩種，AdaBoost[12]以及支援向量機 (Support Vector Machine, SVM) [13]。

2.3.1 AdaBoost

AdaBoost 為 Adaptive Boosting 的簡稱，是基於 Boosting 的概念延伸出的演算法之一，由 Yoav Freund 與 Robert Schapire[12]於 1997 年提出，其概念為藉由聯合一定數量的弱分類器來構成一個有良好決策效能的強分類器，只要分類器的正確率達到 50%以上便是一個可以使用的弱分類器。

AdaBoost 演算法的流程如下所示：

1. 已知 n 個訓練樣本為 $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ ，其中 $X_i \in \bar{X} = [F_1, F_2, \dots, F_m]$ 為第 i 個訓練樣本(training simple)且每個樣本中有 m 個特徵，而 $Y_i \in \{+1, -1\}$ 為第 i 個訓練樣本的標籤(label)。
2. 初始化權重 $D_t(i) = \frac{1}{n}$ ，其中 $D_t(i)$ 為第 t 個弱分類器於第 i 個訓練樣本的權重。
3. 開始 AdaBoost 的訓練迭代迴圈：

for $t = 1$ to T

a. 利用已知的 $D_t(i)$ 來訓練一個弱分類器 $h_t(X) : X \rightarrow \{+1, -1\}$ 。

b. 計算 h_t 中每種特徵檢測方式 F_j 之於所有樣本的錯誤率：

for $j = 1$ to m

$$\varepsilon_j = \sum_{i=1}^n D_t(i) |h_t(X_i, F_j) - Y_i|$$

end

c. 從 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m$ 中找出最小的就是本次訓練的最小錯誤值 E_t 。

d. 更新權重：

$$D_{t+1}(i) = D_t(i) \exp(-\alpha_t y_i h_t(x_i))$$

$$\text{其中 } \alpha_t = \frac{1}{2} \ln \left[\frac{1-E_t}{E_t} \right]。$$

e. 正規化權重：

$$D_{t+1}(i) = \frac{D_{t+1}(i)}{\sum_{k=1}^n D_{t+1}(k)}$$

end

傳回最終弱分類器 h_t 以及各分類器的權重 α_t ，其中 $t = 1, 2, \dots, T$ 。

4. 最終的強分類器 $H(X)$ 為：

$$H(X) = \begin{cases} 1 & \sum_{t=1}^T \alpha_t h_t(X) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ 0 & \text{otherwise} \end{cases}$$

2.3.2 支援向量機

支援向量機(Support Vector Machine, SVM)[13]由 Cortes 與 Vapnik 於 1995 年提出，它的基本概念為找出一個超平面(hyper-plane)分隔兩群不同的資料集，並且使該平面與資料群的邊界(margin)有著最遠的距離。以圖 2.3(A)的 H_1 、 H_2 以及 H_3 三個超平面中，只有 H_1 以及 H_2 可完整切割兩群不同的資料集，其中又以 H_2 這個平面與資料群的邊界有著最遠的距離，而 SVM 演算法的功能就是先找出資料群的邊界，如圖 2.3(B)所示，上面的虛線即為資料群的邊界又稱為支援超平面(support hyper-plane)，且這兩個超平面必定為平行，而在兩個超平面中間的便是分隔兩群不同資料的最佳分離超平面(optimal separating hyper-plane)，因此只需要判斷新的資料點落在最佳分離超平面的哪一邊，便能得知它是屬於哪一個類別的資料。

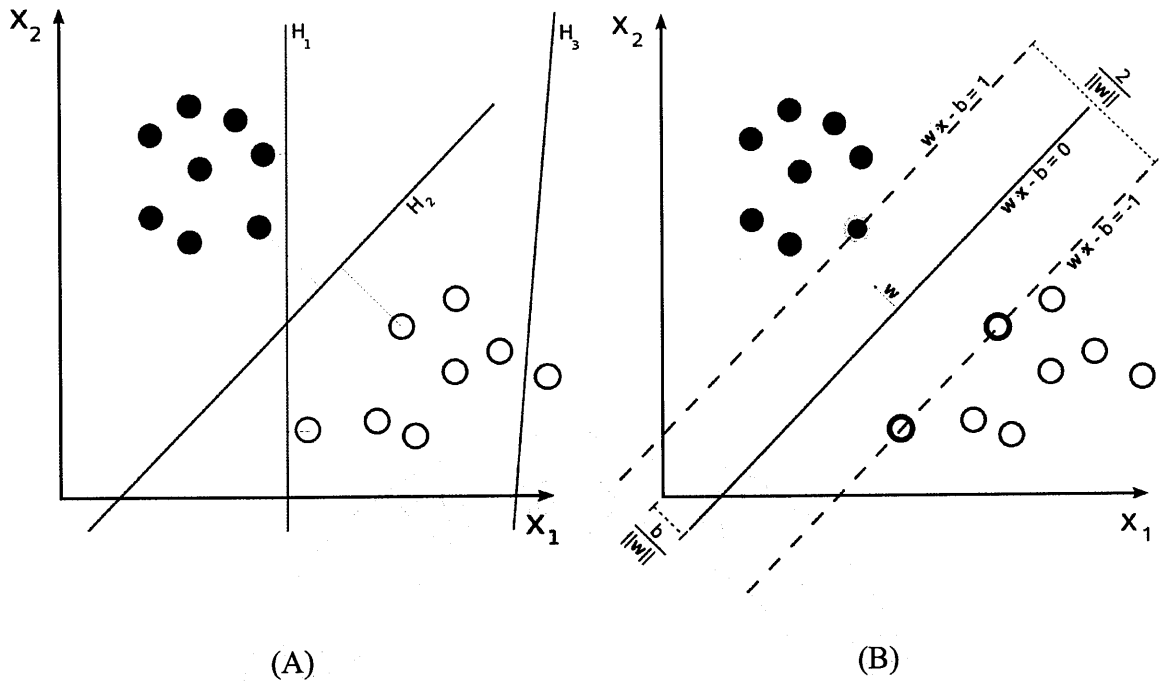


圖 2.3 特徵空間中分隔兩個群組的超平面[14]

以上所敘述的是 SVM 在解決線性資料的分類問題，而大部分所遇到的問題都是非線性的問題，而解決的方式則是利用核函數(kernel function)將資料映射到更高維度的空間或者是特徵空間(feature space)去找出超平面，如圖 2.4 所示。

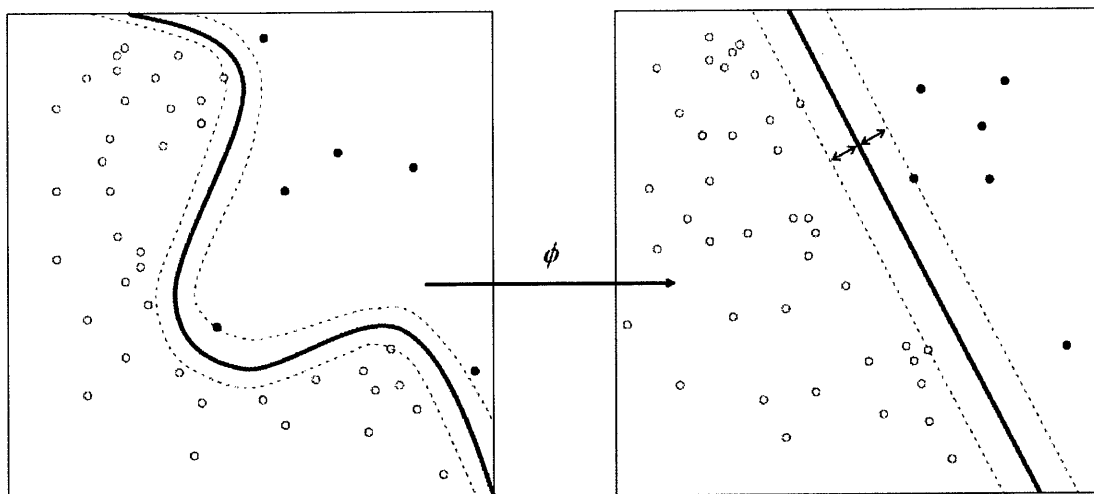


圖 2.4 利用核函數將資料映射到高維度空間找出超平面[14]

2.4 行人偵測之系統架構探討

現行的行人偵測系統[2-6]在其實驗結果中皆有相當好的辨識率，但在偵測 320 x240 大小的影像大多需要數秒鐘甚至是數分鐘才能完成，這對於需要即時運算的應用(例如：車載駕駛輔助系統的行人碰撞預警)是一個相當大的問題。

然而在行人特徵描述方法的部分，目前是以 Tuzel 等人[11]所提到的共變異矩陣搭配積分影像的方法最為適合，因為經過 Paisitkriangkrai[3]的實驗結果分析，共變異矩陣描述子對於行人外觀、姿勢以及光線的改變都能維持相當好的強健性，雖然其精確度與 Paisitkriangkrai[6]使用共變異矩陣描述子配合 AdaBoost 的偵測效能略低於梯度方向直方圖(HOG)描述子配合 SVM 的偵測效能，如圖 2.5[6]所示，但是其在低解析度的情況下仍保有著一定的強健性，不像 HOG 描述子當解析度過低時其精確度就開始下滑了，此外更重要的一點是共變異矩陣描述子可以使用積分影像來大大的增加其運算速度，因此本篇論文決定以共變異矩陣描述子做為行人偵測系統的特徵萃取方法。

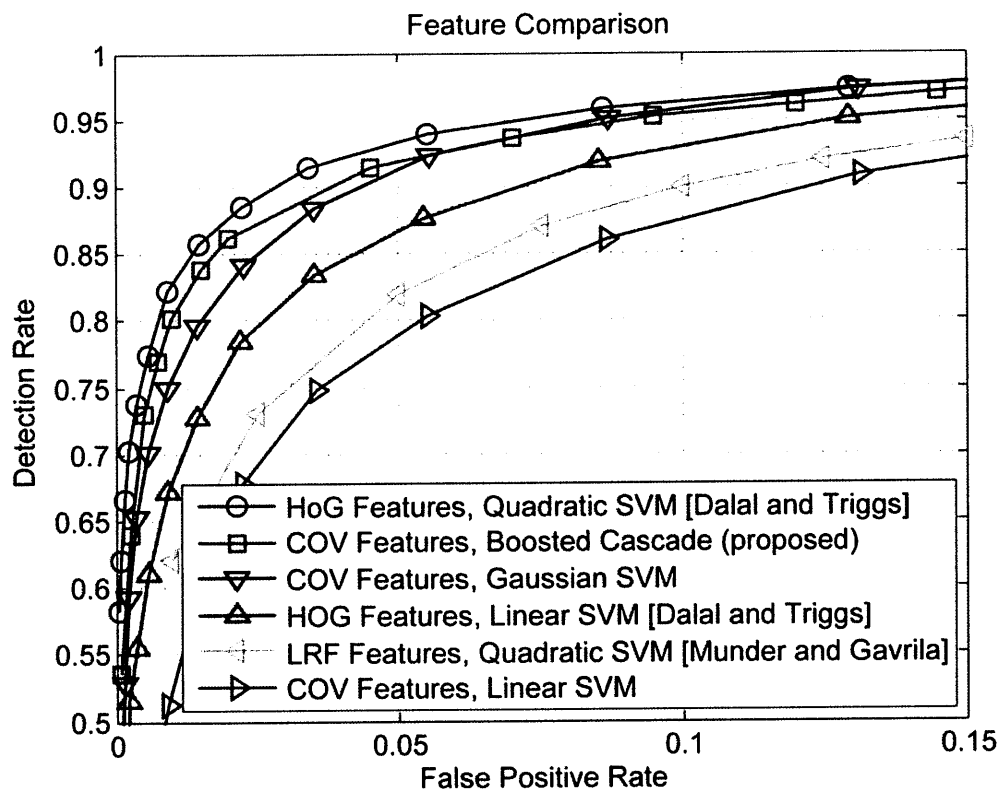


圖 2.5 Paisitkriangkrai 所做的行人偵測系統效能評比[6]

而為減少強健描述子(共變異矩陣描述子以及 HOG 等)的運算次數，大部份的做法都是在行人分類器 (pedestrian classifier) 的前級加上行人偵測器 (pedestrian detector)，目的在於使用前級的行人偵測器快速的找出行人可能存在的候選區域，並且過濾掉大部分的無行人區域以減輕後級行人分類器的負擔，進而提升整個行人偵測系統的運行效率。而大部份的文獻皆使用 Haar-like 特徵做為前級行人偵測器的特徵描述方式，像是[3]中所提到的雙層行人偵測系統架構，因此本論文也將採用以 Haar-like 特徵為基礎的前級行人偵測器。

因此在特徵值描述方法方面已經使用了相較之下比較快速的方法，而在經由實驗測試後發現行人偵測系統運算所花費的時間有一部分是在於分類器演算法的分類過程過於複雜，造成行人偵測系統運行效率下降。而由於 AdaBoost、SVM 與

類神經網路等機器學習演算法因效果卓越因此被普遍的應用，不過它們分類過程的演算法因較為複雜故在運作時花費的時間也較長。因此本論文以線性判別分析的方式採用 Fisher 線性判別分析[10]來做為行人偵測的分類依據。

Fisher 線性判別分析嚴格說來不算是分類器學習演算法的一種，它原本的作用是使用於補助數種特徵相近的資料集(例如：花卉的分類...等)辨識，Fisher 線性判別分析可以讓這些資料擁有最佳的鑑別度，其運作方式是將這類的資料集投影到擁有最佳鑑別度的一維空間，且使這些資料在新的子空間將擁有最大類別間距 (between classes distance)與最小的類別內距(within classes distance)，而通常這些一維的資料會再使用其他的分類器演算法再做一次訓練，而 Paisitkriangkrai[3]等人也是以上述的方式來辨別行人，但是本研究認為行人偵測在分類時只有兩種類別(行人與非行人)，因此使用 Fisher 線性判別分析來使兩個特徵集(正樣本集與負樣本集)擁有最大類別間距與最小類別內距，如此就足以用來判別行人了。因此本論文以 Fisher 線性判別分析來取代以往的分類器學習演算法做為我們行人偵測系統的分類器，在實驗結果中以 Fisher 線性判別分析為基礎的分類器確實能夠以犧牲非常少量的準確度(0.1%~0.5%)做為代價換取近兩倍以上的分類速度。

第三章 Haar-like 特徵

3.1 概述

Haar-like 特徵是一種用於物件辨別的數位影像特徵，Haar-like 特徵的萃取方式是以影像中某個矩形區域為主，此矩形區域又可稱為矩形特徵(rectangular feature)，又因為此矩形特徵相似於小波轉換(wavelet transform)中的 Haar 小波函數之結構因而得名，最早由 Viola 與 Jones[8]提出且將其應用於人臉偵測[7]。

Haar-like 特徵會先選定偵測窗(detection window)中某個位置的相鄰矩形區域，然後將其相鄰矩形區域內的像素值加總，並且計算它們之間的差做為特徵值。由多個位置、大小與類型皆不相同之矩形特徵所萃取出的特徵值組成的 n 維資料，將可使用於分類器的訓練與測試。

3.2 特徵模板

每個 Haar-like 特徵都是由相鄰的矩形組合而成，而這些矩形的組合我們稱之為特徵模板(feature pattern)，特徵模板皆由兩個或數個大小相同的矩形比鄰組合而成。特徵模板內會有白色與灰色兩種矩形區塊，而每一種特徵模板之特徵值的定義為白色區塊內像素值的加總減去灰色區塊內像素值的加總，較常見的特徵模板如圖 3.1 所示。

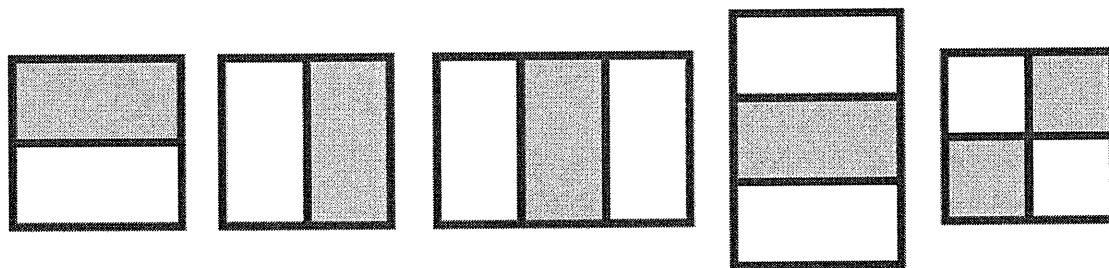


圖 3.1 常見的 Haar-like 特徵模板

這些常見的特徵模板依其內所含的矩形個數可分為三類：(1)雙矩形特徵、(2)三矩形特徵與(3)四矩形特徵，經由計算所得的特徵值可以用來描述偵測對象：

- (1) 雙矩形特徵(two-rectangular feature)：計算白色矩形區域內之像素值加總減去灰色矩形區域內之像素值加總的值。
- (2) 三矩形特徵(three-rectangular feature)：計算兩旁白色矩形區域內之像素值加總然後減去中間灰色矩形區域內之像素值加總的值。
- (3) 四矩形特徵(four-rectangular feature)：計算對角的白色矩形區域內之像素值加總然後減去另外一個對角的灰色矩形區域內之像素值加總的值。

除此之外依照各種不同的應用衍生出許多不同樣式的模板，如圖 3.2 所示，此為 Haar-like 特徵用於人臉偵測的模板，與常見的特徵模板相同，它們皆由許多大小不同的相鄰矩形組合而成，唯一的原則是模板中的白色區塊的面積與灰色區塊的面積在經過加權後必須完全相同。

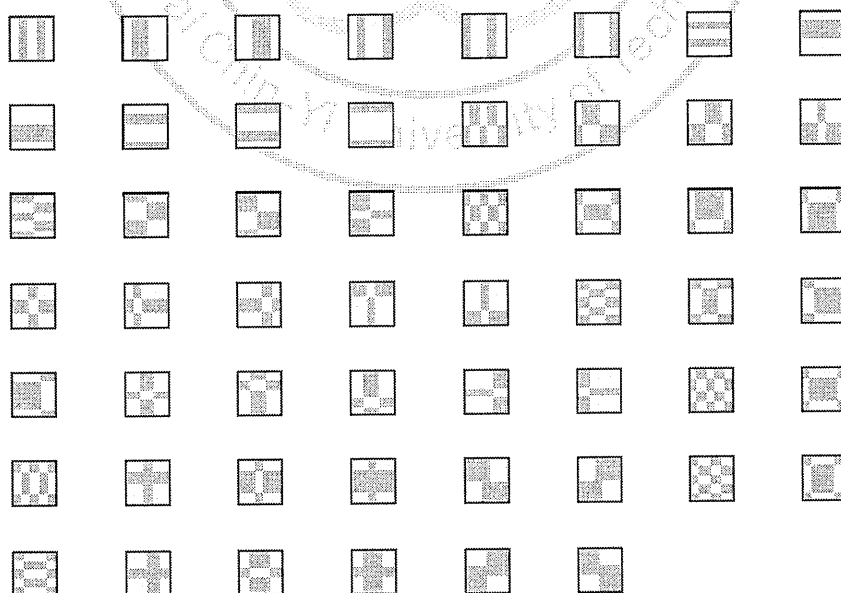


圖 3.2 人臉偵測的 Haar-like 特徵模板[15]

3.3 偵測窗中的特徵模板

特徵模板在偵測窗中可以被放置在任意位置、可以是任意大小並且相鄰矩形的排列方式垂直或水平皆可，每一種不同的形態都是一個 Haar-like 特徵。在此以雙矩形 Haar-like 特徵(two-rectangular feature)在一個 3×3 大小的偵測窗中的形態為例子，雙矩形 Haar-like 特徵依其位置、大小與相鄰矩形的垂直或水平組合變化，在 3×3 的範圍中總共存在著 24 種不同的形態，換句話說即是在 3×3 大小的偵測窗中總共有 24 個 Haar-like 特徵，如圖 3.3 所示。

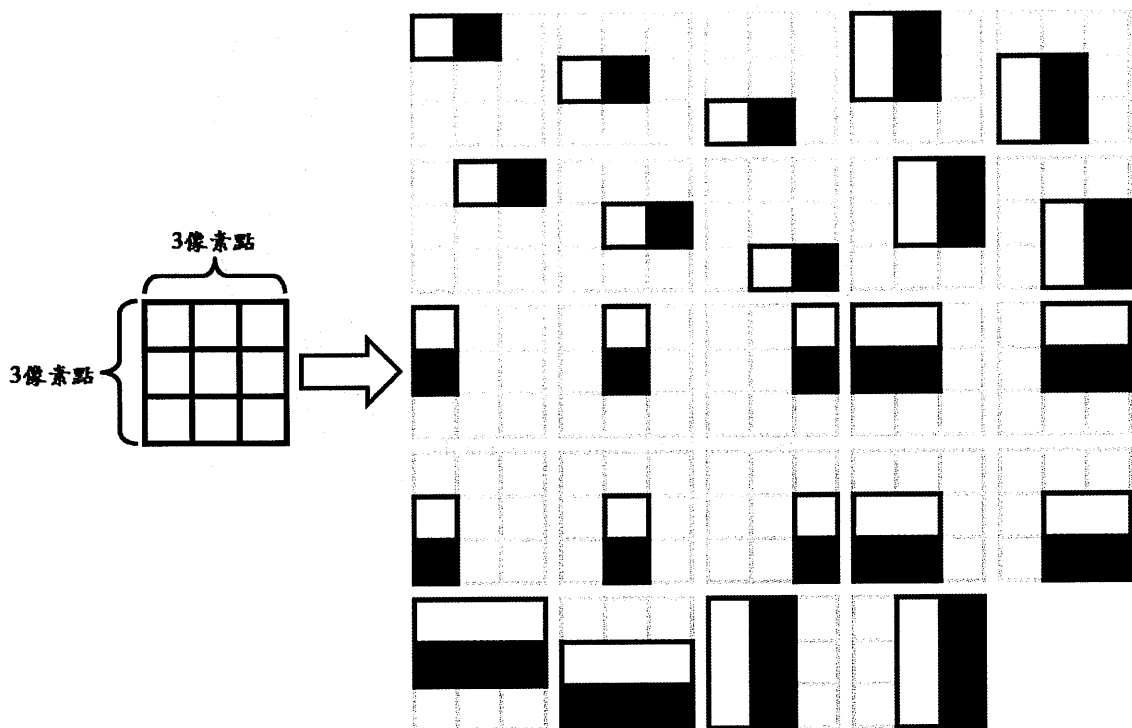


圖 3.3 所有雙矩形特徵在 3×3 偵測窗中的形態

由以上的例子可以看出雙矩形 Haar-like 特徵在 3×3 的範圍中萃取特徵值時，有許多模板所涵蓋的矩形區域是重疊的，而萃取 Haar-like 特徵時往往需要統計某

個矩形區域之像素值加總，因此若使用一般的方法計算其加總將會使模板間重疊的區塊再三的被統計，除此之外隨著特徵模板涵蓋的範圍增加其統計像素值加總的計算時間也會隨之拉長，如此一來將會浪費非常多的時間在統計像素的加總上。 3×3 偵測窗中模板間的重疊區塊已經非常多了，更何況是使用於行人偵測系統的 64×128 的偵測窗。

所幸這些問題可以利用積分影像(integral image)的方法讓 Haar-like 特徵在計算時可以避免因為模板涵蓋區塊重疊所造成的重複統計像素值加總問題。除此之外無論模板涵蓋的範圍大小，只要利用積分影像的方法都可在常量的時間內迅速的求得像素值的加總。

3.4 積分影像

積分影像(integral image)依其演算法的計算結果又可稱為加總面積表(summed area table)，它能夠快速且有效的計算一個集合中的矩形區塊之子集合的加總，於 1984 年首次使用於電腦視覺的領域中，但並沒有被廣泛的應用，直至 2004 年 P. Viola 與 M. J. Jones 將此演算法使用於人臉偵測[7]的架構中才被廣泛的應用。

影像中的一個矩形的區域通常可藉由該區域的左上角的點 (x', y') 與右下角的點 (x, y) 來表示，如圖 3.4 所示。而一個積分影像 ii 在點 (x, y) 的位置的值為原始影像 i 中的點 (x', y') 至點 (x, y) 所標示的矩形面積內所有像素值的加總，其定義如(3.1)所示：

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y') \quad (3.1)$$

其中 i 為原始影像且 $i(x', y')$ 為原始影像點 (x', y') 處的值。假設以灰階影像為例子，

那麼這些 $i(x',y')$ 位置上的值都是介於0~255之間的純量。

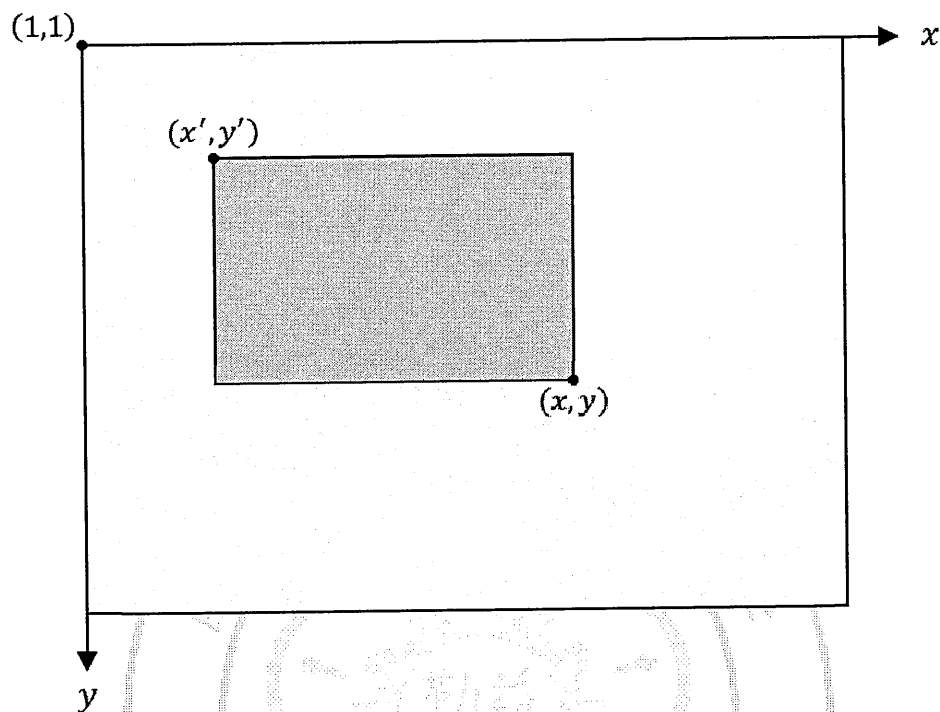


圖 3.4 以兩個點來表示影像中的一個矩形區域

而一張原始影像 $i(x,y)$ 的積分影像 $ii(x,y)$ 可以藉由(3.2)與(3.3)兩個式子所構成的一組遞迴方程式求得：

$$s(x,y) = s(x,y-1) + i(x,y) \quad (3.2)$$

$$ii(x,y) = ii(x-1,y) + s(x,y) \quad (3.3)$$

其中 $s(x,y)$ 為原始影像 $i(x,y)$ 之 y 方向之像素值累加的和，並且定義 $s(x,-1) = 0$ 與 $ii(-1,y) = 0$ 。以一張 5×5 的灰階影像作為例子，依照(3.2)與(3.3)之遞迴方程式組計算其積分影像，如圖 3.5 所示。

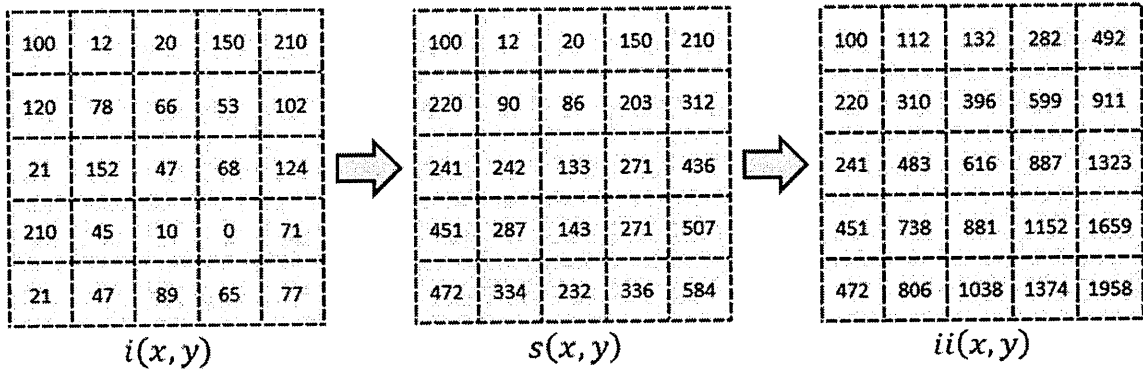


圖 3.5 一張 5×5 灰階影像之積分影像的計算

由上面的例子可以看出，一張 $M \times N$ 的影像只需要迭代 $M \times N \times 2$ 次即可得到該影像之積分影像。

3.5 利用積分影像計算 Haar-like 特徵

如圖 3.6 所示，在求得一張影像的積分影像之後，就可以藉由參考積分影像上矩形區域 D 的四個點 (ii_1, ii_2, ii_3, ii_4) 的值來計算該矩形區域中之像素值加總。

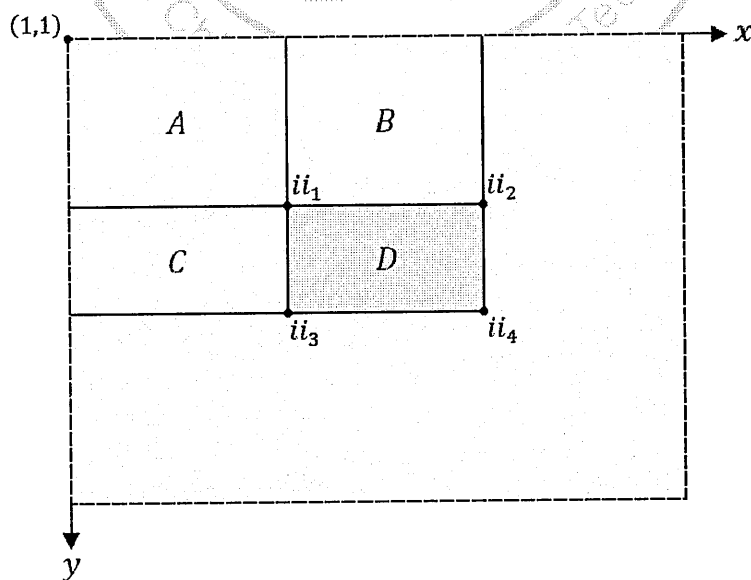


圖 3.6 參考四個點計算矩形區域 D 中的像素值加總

其中：

ii_1 為矩形區域 A 中的像素值加總。

ii_2 為矩形區域 $A + B$ 中的像素值加總。

ii_3 為矩形區域 $A + C$ 中的像素值加總。

ii_4 為矩形區域 $A + B + C + D$ 中的像素值加總。

因此：

矩形區域 D 中之像素值加總 = $ii_4 + ii_1 - (ii_2 + ii_3)$ 。

利用以上的概念就可以利用積分影像計算原始影像中任何位置以及任何大小的矩形區塊之像素值加總。以 Haar-like 特徵的雙矩形特徵(two-rectangular feature) 模板之特徵值計算為例，由白色矩形區域 A 之像素值加總減去灰色矩形區域 B 之像素值加總，如圖 3.7 所示：

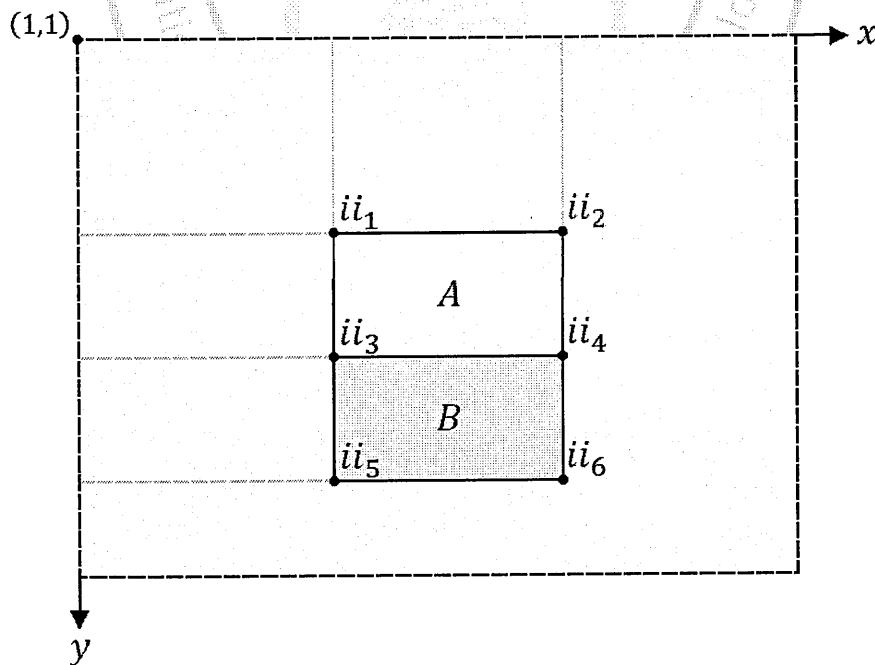


圖 3.7 參考六個點計算雙矩形 Haar-likers 特徵

其中：

$$\text{矩形區域}A\text{像素值加總} = ii_4 + ii_1 - (ii_2 + ii_3)$$

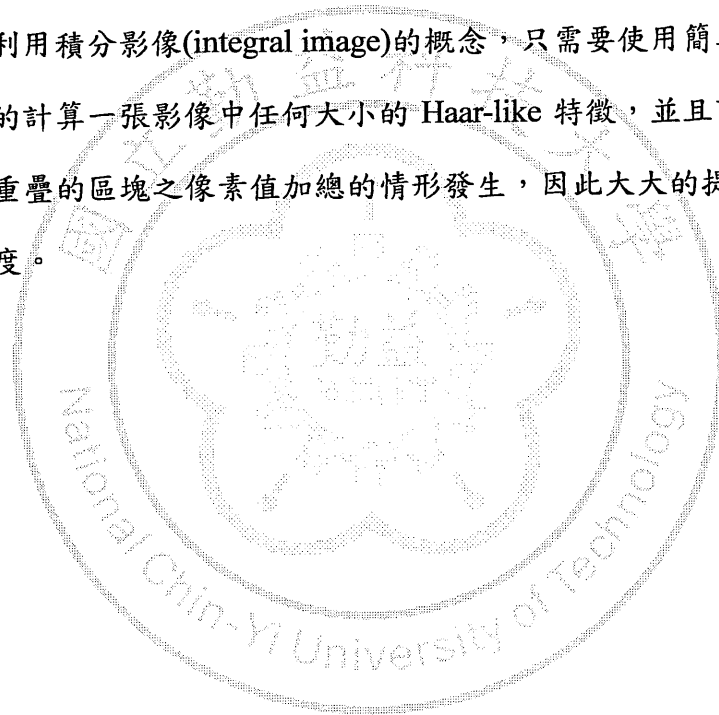
$$\text{矩形區域}B\text{像素值加總} = ii_6 + ii_3 - (ii_4 + ii_5)$$

$$\text{雙矩形 Haar-like 特徵值} = \text{矩形區域}A\text{像素值加總} - \text{矩形區域}B\text{像素值加總}$$

故此雙矩形 Haar-like 特徵之值為：

$$(ii_4 - ii_3) - (ii_2 - ii_1) + (ii_4 - ii_3) - (ii_6 - ii_5)$$

由此可見利用積分影像(integral image)的概念，只需要使用簡單的加減法運算就可以有效率的計算一張影像中任何大小的 Haar-like 特徵，並且可避免重複的計算特徵模板間重疊的區塊之像素值加總的情形發生，因此大大的提高了 Haar-like 特徵的運算速度。



第四章 共變異矩陣描述子

4.1 概述

共變異矩陣描述子(covariance matrix descriptor)又稱之為區域共變異數(region covariance)，由 Tuzel 與 Porikli 等人在 2006 年提出並且應於物件偵測(object detection)以及紋理辨識(texture classification)[11]。共變異矩陣描述子是藉由計算原始影像之數種特徵影像之間的共變異矩陣來描述這些特徵影像間之交互關係，而特徵影像包含像素的 x 座標、 y 座標、一階偏導數、二階偏導數、梯度方向、梯度強度...等。共變異矩陣描述子的計算區域通常都只是影像中小範圍的感興趣區域，因此共變異矩陣描述子是一種區域性的描述子(region descriptor)。

4.2 共變異矩陣

共變異矩陣(covariance matrix)又可以稱為隨機向量 X 的變異數(Variance of random vector X)，它是由隨機向量 X 中的各個分量之間的共變異數所組成的矩陣。設 d 維隨機向量 $X = [x_1, x_2, x_3, \dots, x_d]^T$ ，則其共變異矩陣 C 的第 (i, j) 個元素的定義如(4.1)所示：

$$C_{i,j} = cov(x_i, x_j) = E[(x_i - \mu_i)(x_j - \mu_j)] \quad i, j = 1, 2, 3, \dots, d \quad (4.1)$$

其中 μ_i 為隨機向量 X 中第 i 個分量的平均值(期望值)，也就是說 $\mu_i = E(x_i)$ 。而隨機向量 X 的共變異矩陣定義如(4.2)所示：

$$C = E[(X - E[X])(X - E[X])^T] \quad (4.2)$$

共變異矩陣中的每一個元素是 d 維隨機向量 X 的不同分量之間的共變異數，共變異數為變數之間的二階統計特性，若 d 維隨機向量 X 分量之間的相關性很小，那麼所得到的共變異數矩陣幾乎是一個對角矩陣(diagonal matrix)，因此共變異矩陣中的元素不是樣本之間的共變異數。

4.3 用於行人偵測的特徵向量(隨機向量 X)

共變異矩陣中的每一個元素是 d 維隨機向量 X 的不同分量之間的共變異數，若欲藉由共變異矩陣來描述影像中的資訊，就必須先定義 d 維的隨機向量 X 中各個分量所對應的特徵，例如：色彩空間分量、邊緣資訊、梯度方向、濾波器響應...等。

設 F 為 $W \times H \times d$ 的特徵影像，其中 W 為原始影像的寬、 H 為高，而 d 為特徵影像的種類，換句話說 d 就是隨機向量 X 中的分量個數，如(4.3)所示：

$$F(x, y) = \phi(I, x, y) \quad (4.3)$$

其中 ϕ 為原始影像 I 的特徵影像分量的轉換函式，即色彩空間分量、邊緣資訊、梯度方向、濾波器響應...等。

在本篇論文中，針對行人偵測所制定的特徵影像分量的轉換函式，其定義如(4.4)所示：

$$F(x, y) = \left[x \quad y \quad |I_x| \quad |I_y| \quad \sqrt{I_x^2 + I_y^2} \quad |I_{xx}| \quad |I_{yy}| \quad \tan^{-1} \frac{|I_x|}{|I_y|} \right]^T, \quad (4.4)$$

$$|I_x| = \left| \frac{\partial I(x, y)}{\partial x} \right|, |I_y| = \left| \frac{\partial I(x, y)}{\partial y} \right|, |I_{xx}| = \left| \frac{\partial^2 I(x, y)}{\partial x^2} \right|, |I_{yy}| = \left| \frac{\partial^2 I(x, y)}{\partial y^2} \right|$$

其中 x 與 y 為原始影像 I 的像素點座標位置， I_x 與 I_y 為原始影像 I 的水平與垂直方向的一階偏導數，其影像濾波器的形態為 $[-1 \ 0 \ 1]$ ； I_{xx} 與 I_{yy} 為原始影像 I 的水平與垂直方向的二階偏導數，其影像濾波器的形態為 $[-1 \ 2 \ -1]$ ， $\sqrt{I_x^2 + I_y^2}$ 為原始影像 I 的梯度強度， $\tan^{-1} \frac{|I_x|}{|I_y|}$ 為原始影像 I 的梯度方向。如圖 4.1 所示：

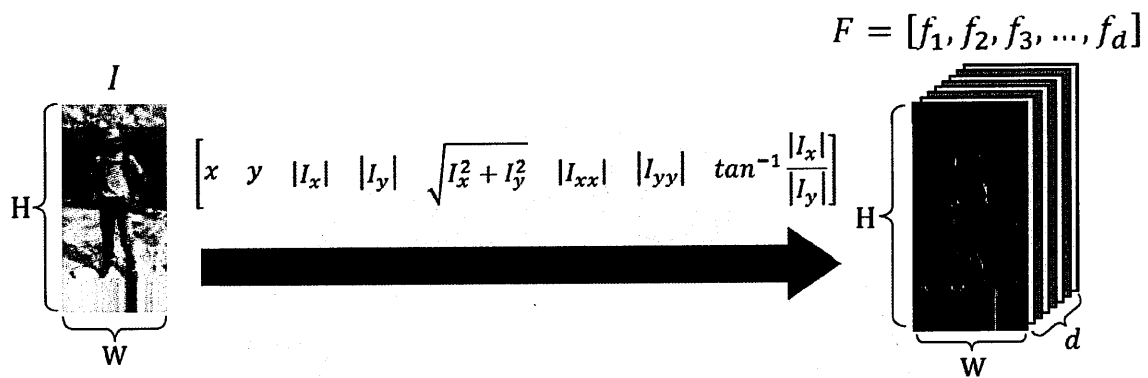


圖 4.1 原始影像轉換至各個特徵影像分量

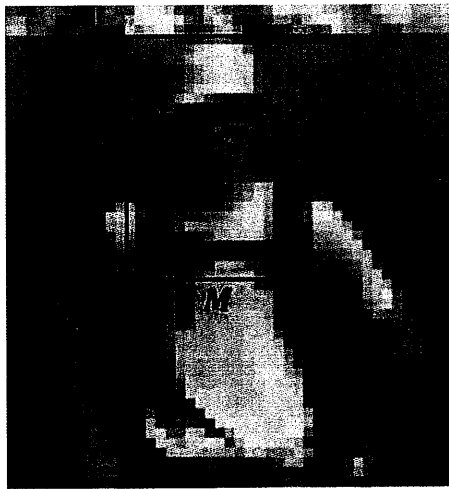
4.4 共變異矩陣描述子

與第三章的 Haar-like 特徵的萃取相似，都是以偵測窗中的矩形子區域為基礎，其不同之處在於 Haar-like 特徵的萃取結果是一個純量，而共變異矩陣描述子的萃取結果是一個共變異矩陣，又因為此共變異矩陣描述子是描述影像中的一個矩形區域內各特徵影像分量之間的關係，因此又將它稱為區域共變異矩陣(region covariance matrix)。

區域共變異矩陣的計算皆建立在 4.2 節所提到的特徵影像 F 上，設一個矩形區域 $R \in F$ ，令 $\{z_k\}_{k=1,2,3,\dots,MN}$ ，則矩形區域 R 的共變異矩陣 C_R 的定義如(4.5)所示：

$$C_R = \frac{1}{MN - 1} \sum_{k=1}^{MN} (z_k - \mu)(z_k - \mu)^T \quad (4.5)$$

其中 M 與 N 分別為矩形區域 R 的寬與高，而 MN 則是矩形區域 R 中的像素個數總量， z_k 為矩形區域 R 中第 k 個像素上特徵值所組成的向量， μ 為矩形區域 R 中各個特徵影像分量的平均值所組成的向量，如圖 4.2 所示。



$$\mu = [E[f_1], E[f_2], E[f_3], \dots, E[f_d]]^T$$

$$z_k = [f_{1k}, f_{2k}, f_{3k}, \dots, f_{dk}]^T$$

$$\text{其中 } k = 1, 2, 3, \dots, MN$$

圖 4.2 共變異矩陣 C_R 所描述之矩形區域

令 $\mu = [E[f_1], E[f_2], E[f_3], \dots, E[f_d]]^T = [\mu_1, \mu_2, \mu_3, \dots, \mu_d]^T$ ，並參考(4.1)將 C_R 展開，如(4.6)所示：

$$C_R = \frac{1}{MN - 1} \times \begin{bmatrix} \sum_k (f_{1k} - \mu_1)(f_{1k} - \mu_1) & \sum_k (f_{1k} - \mu_1)(f_{2k} - \mu_2) & \cdots & \sum_k (f_{1k} - \mu_1)(f_{dk} - \mu_d) \\ \sum_k (f_{2k} - \mu_2)(f_{1k} - \mu_1) & \sum_k (f_{2k} - \mu_2)(f_{2k} - \mu_2) & \cdots & \sum_k (f_{2k} - \mu_2)(f_{dk} - \mu_d) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_k (f_{dk} - \mu_d)(f_{1k} - \mu_1) & \sum_k (f_{dk} - \mu_d)(f_{2k} - \mu_2) & \cdots & \sum_k (f_{dk} - \mu_d)(f_{dk} - \mu_d) \end{bmatrix} \quad (4.6)$$

由(4.6)可以看出無論多大的 $M \times N$ 矩形區域 R 在計算其共變異矩陣描述子之後，皆被表示成一個 $d \times d$ 的共變異矩陣 C_R ，可以達到影像資訊精簡化的效果，且 C_R 為對稱矩陣(symmetric matrix)，因此只需要紀錄其上三角矩陣的變數所構成的向量即可。

4.5 利用積分影像計算共變異矩陣描述子

與計算 Haar-like 特徵的情況類似：(1)需要經常計算某區域像素值之平均值(加總)、(2)有矩形區域(特徵模板)重疊的可能...等，而使用共變異矩陣描述子的好處在於儘管其運算過程複雜，但同樣的能夠利用積分影像(Integral Image)的概念大大的提升其實際應用時的運算效率。

為了利用積分影像來計算共變異矩陣描述子，必須配合共變異矩陣的矩陣運算特性，參考(4.5)以及(4.6)的結果，定義最終的共變異矩陣之第 (i, j) 個元素如式子(4.7)所示：

$$C_R(i, j) = \frac{1}{MN - 1} \sum_{k=1}^{MN} (z_k(i) - \mu(i))(z_k(j) - \mu(j)) \quad (4.7)$$

將(4.7)乘開後整理後如(4.8)所示：

$$C_R(i, j) = \frac{1}{MN - 1} \left[\sum_{k=1}^{MN} z_k(i)z_k(j) - \frac{1}{MN} \sum_{k=1}^{MN} z_k(i) \sum_{k=1}^{MN} z_k(j) \right] \quad (4.8)$$

其中 $z_k(i)$ 為第 i 個特徵影像分量中第 k 個像素點的值。

由(4.8)得知若要計算已知矩形範圍 R 的共變異矩陣，則必須先求出所有特徵影像分量 $\{z_k(i)\}_{i=1,2,3,\dots,d}$ 的積分影像 P ，以及任意兩個特徵影像分量相乘 $\{z_k(i)z_k(j)\}_{i,j=1,2,3,\dots,d}$ 的積分影像 Q ，如(4.9)與(4.10)所示：

$$P(x, y, i) = \sum_{x' \leq x, y' \leq y} F(x', y', i) \quad (4.9)$$

$$Q(x, y, i, j) = \sum_{x' \leq x, y' \leq y} F(x', y', i)F(x', y', j) \quad (4.10)$$

由(4.9)與(4.10)的結果可將 P 積分影像表示成 d 維的向量 $P_{x,y}$ ，而 Q 積分影像可以表示成 $d \times d$ 的矩陣 $Q_{x,y}$ ，如(4.11)與(4.12)所示：

$$P_{x,y} = [P(x, y, 1) \quad \dots \quad P(x, y, d)]^T \quad (4.11)$$

$$Q_{x,y} = \begin{pmatrix} Q(x, y, 1, 1) & \dots & Q(x, y, 1, d) \\ \vdots & \ddots & \vdots \\ Q(x, y, d, 1) & \dots & Q(x, y, d, d) \end{pmatrix} \quad (4.12)$$

將(4.11)及(4.12)的結果代回(4.8)，將可以得到從原點 $(1,1)$ 至點 (x,y) 所標示的矩形範圍 $R(1,1;x,y)$ 之共變異矩陣描述子 $C_{R(1,1;x,y)}$ ，如(4.13)與圖 4.2 所示：

$$C_{R(1,1;x,y)} = \frac{1}{n-1} \left[Q_{x,y} - \frac{1}{n} P_{x,y} P_{x,y}^T \right] \quad (4.13)$$

其中 $n = x \cdot y$ 。最後，與計算 Haar-like 特徵相同，只需要參考特徵影像中的四個點，便可以透過簡單的四則運算得到任意位置以及大小之矩形區域 $R(x,y;x'',y'')$ 的共變異矩陣描述子 $C_{R(x,y;x'',y'')}$ ，如(4.14)與圖 4.2 所示：

$$C_{R(x,y;x'',y'')} = \frac{1}{n-1} \left[Q_{x'',y''} + Q_{x,y} - Q_{x'',y} - Q_{x,y''} - \frac{1}{n} (P_{x'',y''} + P_{x,y} - P_{x,y''} - P_{x'',y}) (P_{x'',y''} + P_{x,y} - P_{x,y''} - P_{x'',y})^T \right] \quad (4.14)$$

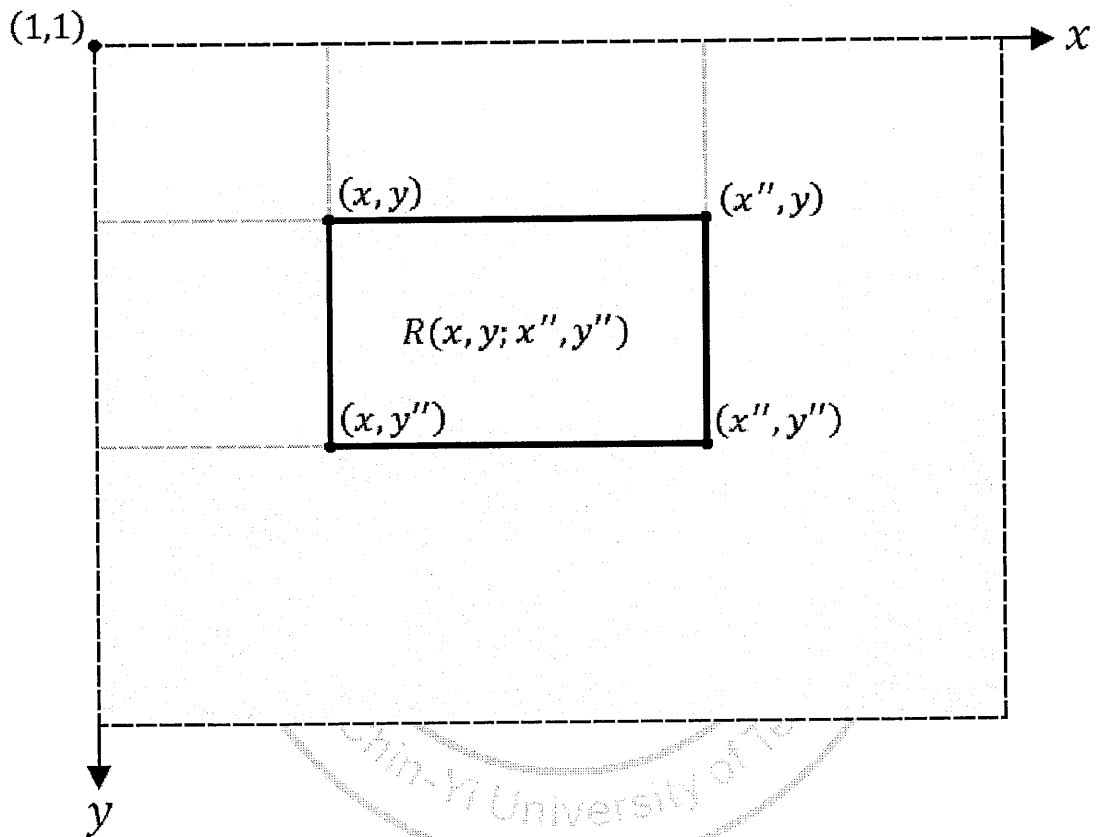


圖 4.3 利用積分影像計算共變異矩陣描述子

第五章 Fisher 線性判別分析

5.1 概述

Fisher 線性判別分析(Fisher Linear Discriminant Analysis, FLDA)是 1936 年由 Fisher [10]所提出的一種劃分群體的技術，於 1996 年被 Belhumeur [16]引入模板辨識與人工智慧的領域，也因此成為模板識別(pattern classification)的經典演算法之一。FLDA 的基本概念是將高維度的模板樣本投射到擁有最佳鑑別度的一維(1-D)空間，且在投影後可以保證模板樣本在新的子空間擁有最大的類別間距(between classes distance)以及最小的類內間距(within classes distance)，如此一來模板樣本在該空間將擁有最佳的可分離性。

若把多維度的向量投影到一維的空間，就可以將其壓縮成一維的純量，但若是毫無根據的將高維度空間中的樣本投影至任意的一維空間，將會使不同類別的樣本變得難以區分。如圖 5.1 所示，將樣本群投影至 X_1 軸或 X_2 軸皆無法將正樣本與負樣本區分開來，但若是經過 FLDA 對樣本群進行計算後求得最佳投影向量便能夠使樣本群投影至最佳的投影軸上，並且擁有最佳的可分離性，如圖 5.2 所示。因此，如何找到樣本群的最佳投影向量 W_{opt} ，使樣本群在新的一維子空間擁有最佳的可分離性，這正是 FLDA 所要解決的問題。

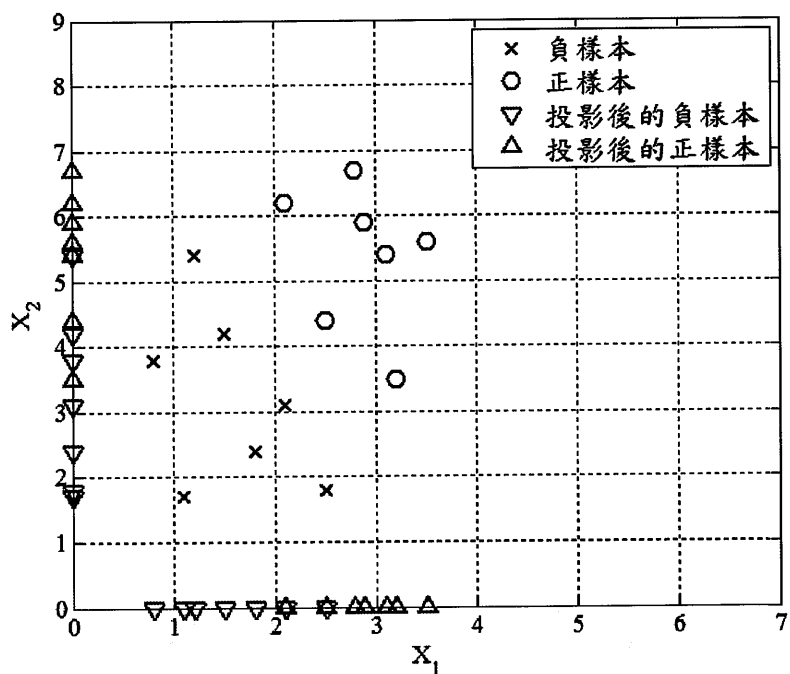


圖 5.1 若將樣本群任意的投影到 X_1 軸或 X_2 軸則無法完整分離

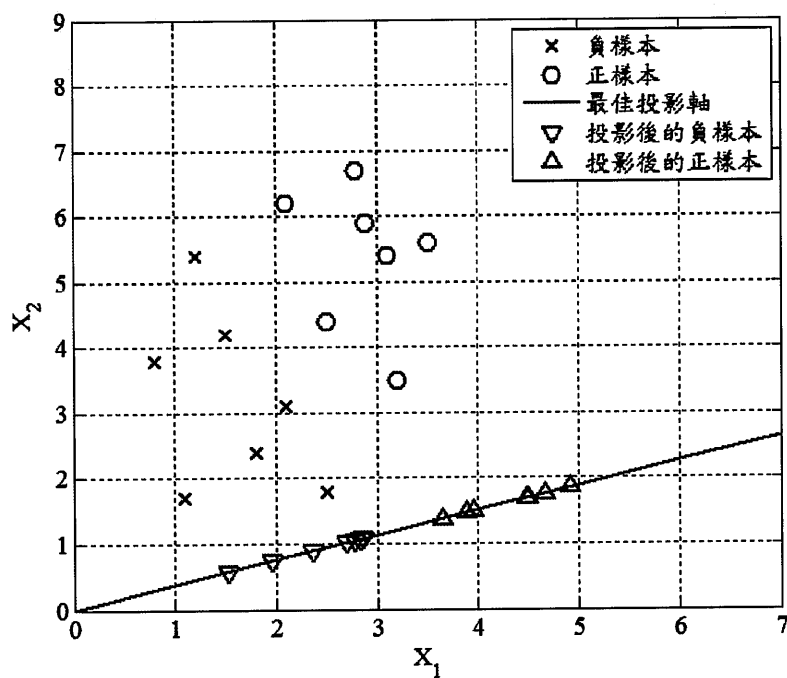


圖 5.2 經由最佳投影向量投影後的樣本群

5.2 Fisher 線性判別分析

假設在一個 R^N 的空間有 m 個樣本分別為 x_1, x_2, \dots, x_m ，即每一個樣本 x_k 都是一個 N 維的向量，並且定義 n_i 用來表示屬於第 i 個類別的樣本個數，設 R^N 的空間中總共有 c 個類別的樣本，則 $n_1 + n_2 + \dots + n_i + \dots + n_c = m$ 。為了找出最佳的投影向量則必須計算各類樣本的期望值(平均值)、樣本類間離散度矩陣(between classes scatter matrix) S_B 以及總類內離散度矩陣(within classes scatter matrix) S_W ，其定義如(5.1)與(5.2)所示：

$$S_B = \sum_{i=1}^c n_i (u_i - \bar{x})(u_i - \bar{x})^T \quad (5.1)$$

$$S_W = \sum_{i=1}^c \sum_{x_k \in \text{class } i} (u_i - x_k)(u_i - x_k)^T \quad (5.2)$$

其中 \bar{x} 為所有樣本的平均值，如(5.3)所示：

$$\bar{x} = \frac{1}{m} \sum_{i=1}^m x_i \quad (5.3)$$

而 u_i 為第 i 個類別中的樣本之平均值，如(5.4)所示：

$$u_i = \frac{1}{n_i} \sum_{x_k \in \text{class } i} x_k \quad (5.4)$$

既然要使用 FLDA 做為一個分類器演算法的基礎，那麼當然會希望樣本集投影至一維的子空間後，各個類別的樣本可以盡可能的分開，因此不同類別之樣本間的平均值應當是相差越大越好，即樣本類間離散度矩陣 S_B ；同時希望相同類別樣本之分佈情形盡量的密集，即總類內離散度矩陣 S_W ，所以在此引入 Fisher 線性判別表示式，如(5.5)所示：

$$J_{fisher}(\phi) = \frac{\phi^T S_B \phi}{\phi^T S_W \phi} \quad (5.5)$$

其中 ϕ 為任一的 n 維向量。FLDA 就是選取令 $J_{fisher}(\phi)$ 達到最大值的向量 ϕ 做為最佳的投影向量，其物理意義即是投影後的樣本擁有最大的類別間距以及最小的類別內距。隨後將(5.1)與(5.2)代入(5.5)得到：

$$J_{fisher}(\phi) = \frac{\sum_{i=1}^c n_i \phi^T (u_i - \bar{x})(u_i - \bar{x})^T \phi}{\sum_{i=1}^c \sum_{x_k \in class_i} \phi^T (u_i - x_k)(u_i - x_k)^T \phi} \quad (5.6)$$

由(5.6)可以設矩陣 $R = \phi^T (u_i - \bar{x})$ ；其中 ϕ 可以將它看成是一個空間，換句話說即 $\phi^T (u_i - \bar{x})$ 是 $(u_i - \bar{x})$ 構成的低維度空間的投影，而 $\phi^T (u_i - \bar{x})(u_i - \bar{x})^T \phi$ 也可以將其表示成 RR^T 。若樣本為行向量(column vector)時， RR^T 即表示 $(u_i - \bar{x})$ 在 ϕ 空間之幾何距離的平方。所以(5.6)的 Fisher 線性判別表示式的分子即為樣本在投影空間 ϕ 中的類間幾何距離的平方和。同理也可得知分母即為樣本在投影空間 ϕ 中的類內幾何距離平方和。因此分類的問題就轉變為：『找一個低維度的空間使得樣本投影到該空間時，投影後的類間距離平方和與類內距離平方和之比為最大。』即使樣本擁有最佳的可分離性。由上述的概念，即透過最佳化下列的標準函數(5.7)找到由一組最佳鑑別向量所構成的投影向量 W_{opt} ：

$$W_{opt} = \arg \max \frac{|W^T S_B W|}{|W^T S_W W|} = [w_1, w_2, \dots, w_n] \quad (5.7)$$

而當 S_W 為非奇異矩陣(nonsingular matrix)時，則(5.8)成立：

$$S_B \phi = \lambda S_W \phi \Rightarrow S_W^{-1} S_B \phi = \lambda \phi \quad (5.8)$$

由(5.8)又可推導出(5.9)：

$$S_B \phi_i = \lambda_i S_W \phi_i \quad (5.9)$$

又因為 $W = [\phi_1, \phi_2, \dots, \phi_d]$ ，將(5.9)代入(5.7)可得：

$$\begin{aligned} S_B \phi_1 &= \lambda_1 S_W \phi_1 & \phi_1^T S_B \phi_1 &= \phi_1^T \lambda_1 S_W \phi_1 \\ \vdots &= \vdots & \Rightarrow \vdots &= \vdots \\ S_B \phi_d &= \lambda_d S_W \phi_d & \phi_d^T S_B \phi_d &= \phi_d^T \lambda_d S_W \phi_d \end{aligned} \quad (5.10)$$

將(5.7)分解可得：

$$\max \frac{|\phi_i^T S_B \phi_i|}{|\phi_i^T S_W \phi_i|} = \max \frac{|\phi_i^T \lambda_i S_W \phi_i|}{|\phi_i^T S_W \phi_i|} = \max \frac{\lambda_i |\phi_i^T S_W \phi_i|}{|\phi_i^T S_W \phi_i|} = \max \lambda_i \quad (5.11)$$

因此由(5.11)的意義來看，欲取得最佳的投影向量只需要找出最大的特徵值(Eigenvalue) $\max \lambda_i$ 即可。整理上述的 FLDA 的推導過程可得到結論：最佳的投影向量 W_{opt} 為 $S_W^{-1} S_B$ 的最大特徵值 $\max \lambda_i$ 所對應之特徵向量(Eigenvector)。因此 W 的

第 $\arg \max \lambda_i$ 行之行向量 (column vector) 即為最佳的投影向量 W_{opt} ，且最佳投影向量的個數 $d \leq c - 1$ 。

5.3 Fisher 分類器

最終的 Fisher 分類器 $F(x)$ 如(5.12)所示：

$$F(x) = \begin{cases} \text{class}_1, & (W_{opt})^T x > \theta \\ \text{class}_2, & (W_{opt})^T x < \theta \end{cases} \quad (5.12)$$

其中 θ 為分類器的閾值。欲找出分類器的閾值 θ ，則必須要先計算各類別的樣本投影至一維子空間後的平均值。首先將樣本投影至一維的子空間如(5.13)所示：

$$y = (W_{opt})^T x \quad (5.13)$$

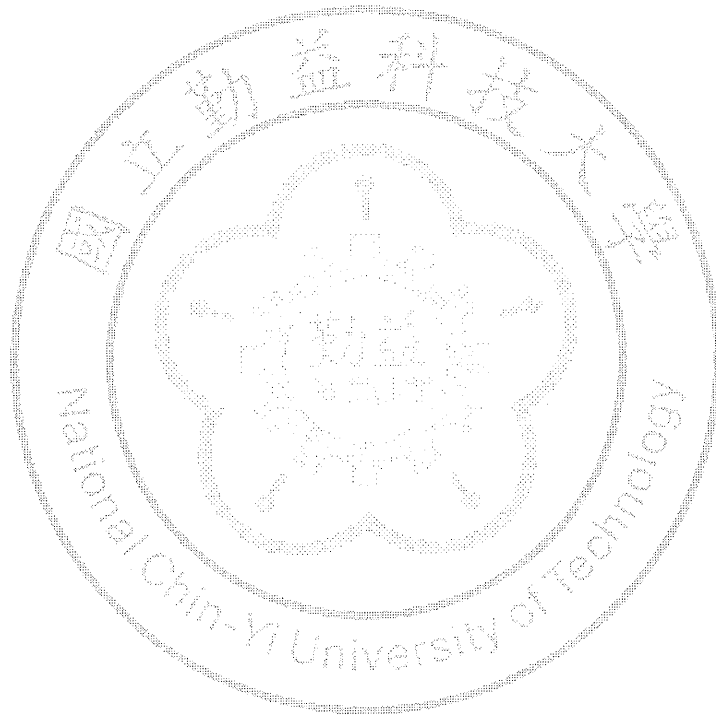
隨後計算各類樣本投影至一維子空間後的平均值，如(5.14)所示：

$$\tilde{u}_i = \frac{1}{n_i} \sum_{y \in \text{class}_i} y \quad i = 1, 2 \quad (5.14)$$

本篇論文主要是針對行人偵測，因此只有兩個類別：行人、非行人，因此 $i = 1, 2$ 。

最後透過(5.15)求得投影空間中的分類閾值 θ ：

$$\theta = \frac{n_1 \tilde{u}_1 + n_2 \tilde{u}_2}{n_1 + n_2} \quad (5.15)$$



第六章 行人偵測系統

6.1 概述

為了實現快速的行人偵測系統，本篇論文的行人偵測系統中採用了 Haar-like 特徵與共變異矩陣描述子做為特徵萃取方式，並且使用 Fisher 線性判別分析針對兩種不同的特徵萃取方式設計兩層 Fisher 分類器來偵測影像中的行人。影像串流資訊每個幀都會先經由以 Haar-like 特徵為基底的候選區域篩選模組迅速的挑選出可能有行人存在的候選區域，再由共變異矩陣描述子為基底的 Fisher 分類器對這些候選區域做進一步的驗證，通過這些檢測的子窗口即是行人所在的位置。

本章將對於本論文所提出的快速行人偵測系統做詳盡的介紹，其中包含：使用的資料庫、特徵的萃取設置及結果、Fisher 分類器的訓練流程、分類器的偵測流程...等。

6.2 行人影像資料庫

本論文所使用的行人影像資料庫為 INRIA[17]的行人影像資料庫，其訓練集裡含有 2,416 張的正樣本資料以及 1,218 張的負樣本資料；測試集中含有 1,126 張的正樣本資料以及 453 張的負樣本資料。

其中，訓練集的正樣本主要由街道上步行的行人、騎乘腳踏車的行人以及其鏡像所構成，訓練集的正樣本影像大小均為 96×160 ，其中行人的影像的平均的寬約為40個像素，高約為100個像素，如圖6.1所示：

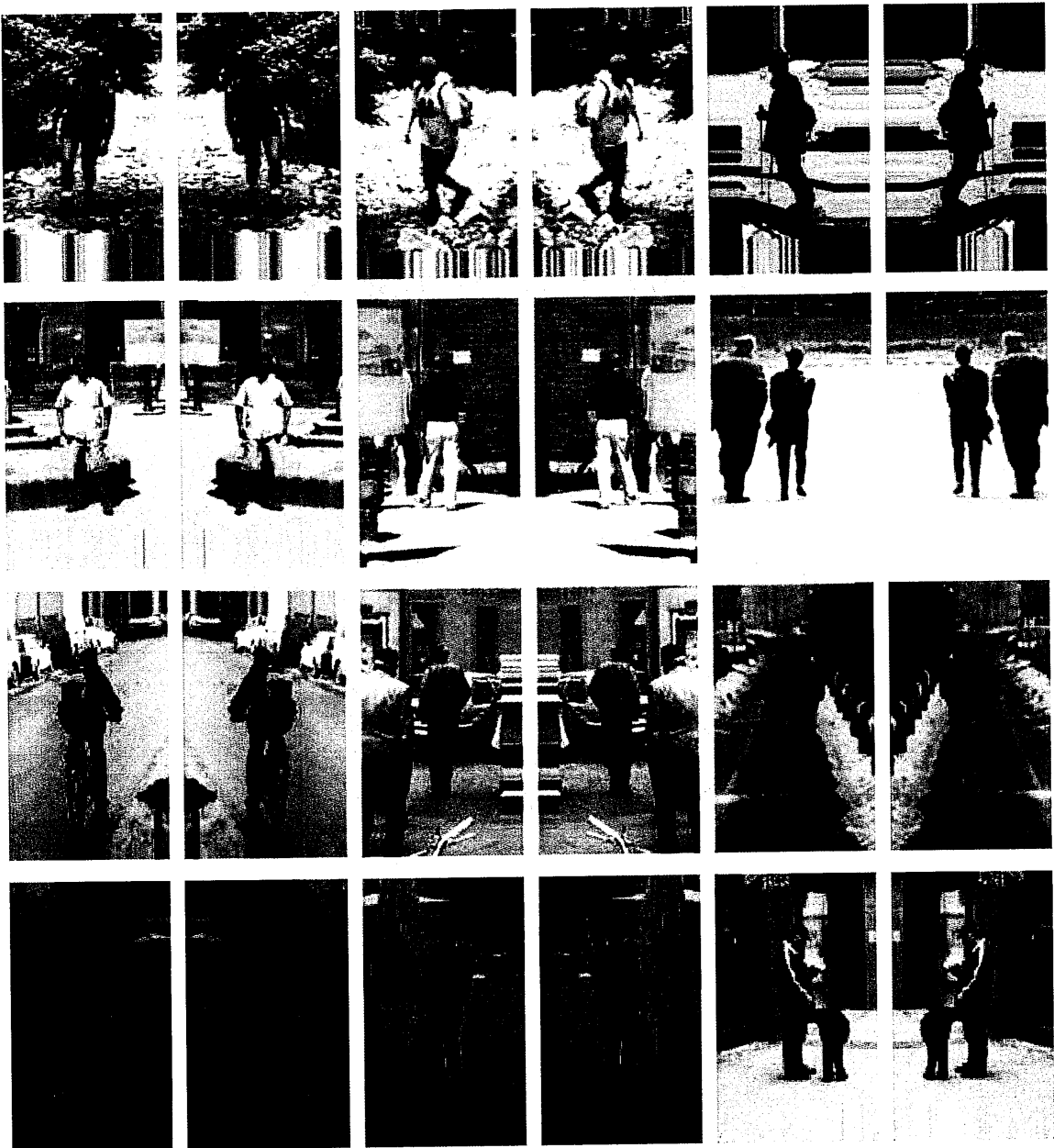


圖 6.1 INRIA 資料庫正樣本訓練集的部分影像

訓練集中的負樣本則是以不包含行人的街景為主，大部分的負樣本大小皆為 320×240 ，但也有少部分是其他的大小，像是 640×480 以及 248×373 等，如圖 6.2 所示：



圖 6.2 INRIA 資料庫負樣本訓練集的部分影像

6.3 特徵萃取及萃取區域選擇

特徵萃取是本行人偵測系統最基礎也是最重要的一個環節，由於 Haar-like 特徵與共變異矩陣描述子的特徵值都是用於描述影像中多個感興趣之矩形區域，並將這些資訊組成一連串的资料供分類器訓練以及測試，因此這些區域的位置、大小以及涵蓋的範圍...等，對於行人偵測系統的效能都有著莫大的影響，若特徵萃取的位置與範圍挑選得宜，則經由 FLDA 所產生的 Fisher 分類器之效能也會更加的理想。

6.3.1 Haar-like 特徵之萃取範圍選取

為了能夠有效率的找出有效的 Haar-like 特徵萃取範圍，本論文將藉由建立行人特徵影像模型來輔助 Haar-like 特徵萃取範圍的選取。由於本行人偵測系統在偵測時是以偵測窗為單位，一次進行一個偵測窗的辨識，且偵測窗皆以 64×128 的比例為基礎來進行縮放，因此在選擇 Haar-like 特徵的萃取區域之前，會先將原始大小為 96×160 之 INRIA 訓練集正樣本影像裁剪至 64×128 的大小，以確保已被記錄的 Haar-like 特徵模板之位置、大小以及縮放倍率在訓練時不會產生誤差。因為不需要彩色影像的資訊，故樣本的預處理步驟為：(1)轉換成灰階影像、(2)做直方圖等化的處理。隨後對 2,416 個 64×128 已預處理完成的正樣本取平均值，最終得到一張特徵影像，如圖 6.3(a)所示。隨後將 64×128 的特徵影像分割成 8×16 個 8×8 的小方格(square-cell)，並且將各個方格內的像素取平均值即是行人特徵影像模型，如圖 6.3(b)所示。



圖 6.3 行人特徵影像模型

隨後為了找出圖 6.3(b)行人影像特徵模型中的關鍵 Haar-like 特徵，我們將使用圖 6.4 中的三種濾波器對行人特徵影像模型做濾波，以找出垂直方向、水平方向以及對角線方向的顯著特徵。

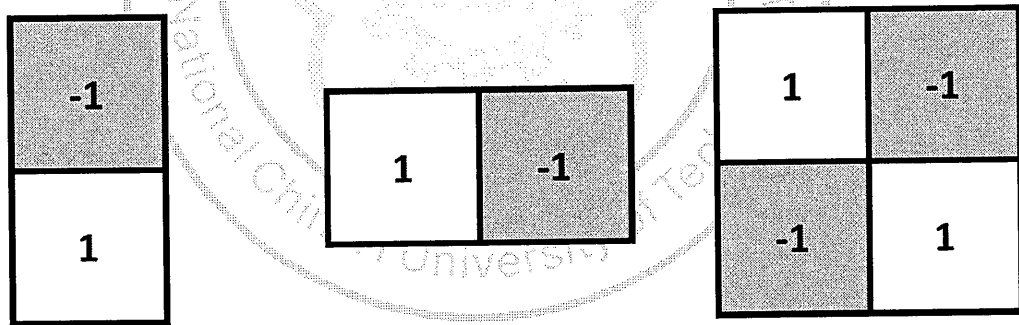


圖 6.4 垂直、水平以及對角方向的濾波器

經過垂直方向、水平方向以及對角方向濾波之後的行人影像特徵模型如圖 6.5 所示，由左至右為：垂直梯度變化、水平梯度變化、對角梯度變化，其中顏色較深的小方格表示行人影像特徵模型該處梯度變化大，反之顏色較淺的小方格則表示行人影像特徵模型該處梯度變化較小。

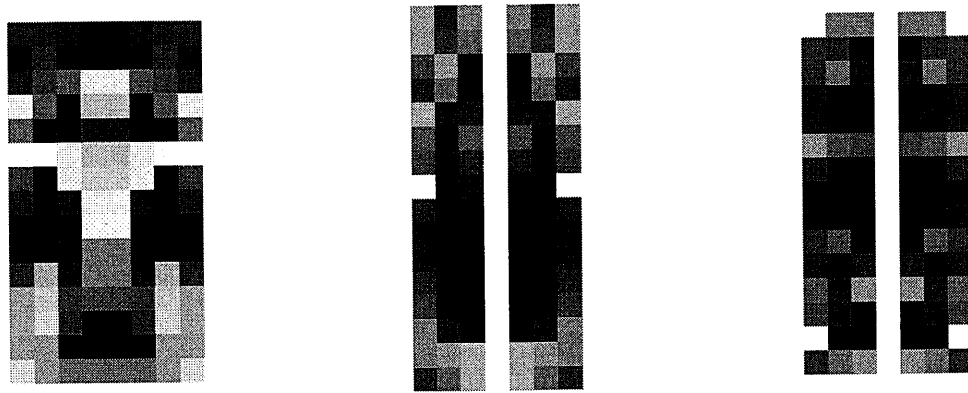


圖 6.5 行人影像模型之垂直、水平以及對角梯度強度影像

為了能夠準確的找出行人影像模型中關鍵的特徵，則必須要將圖 6.5 的梯度強度影像做影像二值化的動作，而二值化的方法為奧茲二值化法(Otsu's method)，其結果如圖 6.6 所示：

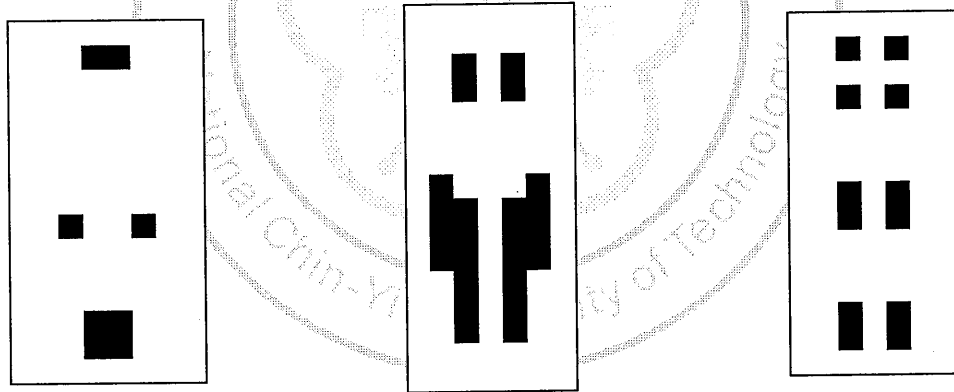


圖 6.6 二值化後的梯度強度影像

由圖 6.6 便可看出行人影像模型中的關鍵 Haar-like 特徵之位置以及大小，而本論文所選用的 Haar-like 特徵模板共有三種：(A)垂直雙矩形特徵模板、(B)水平雙矩形特徵模板、(C)四矩形特徵模板，如圖 6.7 所示。

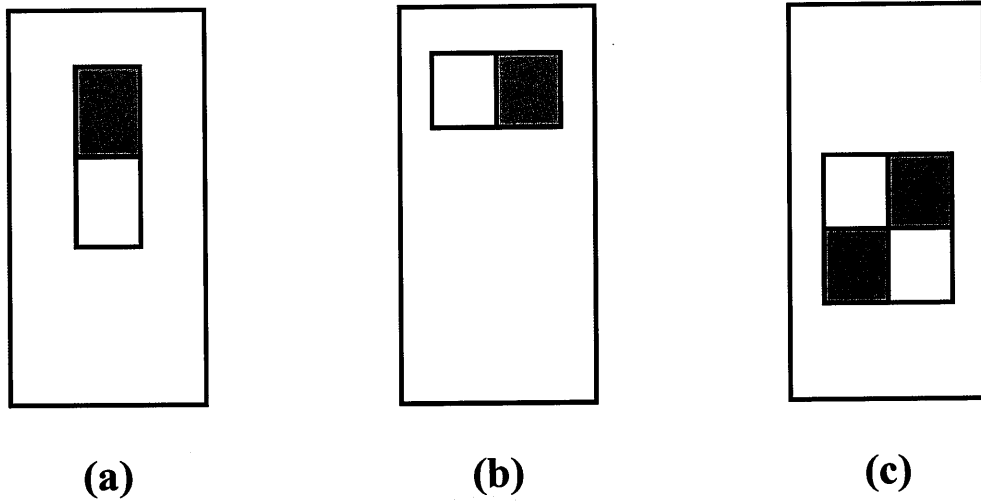


圖 6.7 本論文所使用的三種特徵模板

本論文最終的 Haar-like 特徵萃取範圍為：(A) 5 個垂直雙矩形特徵、(B) 6 個水平雙矩形特徵、(C) 12 個四矩形特徵，如圖 6.8 至圖 6.10 所示：

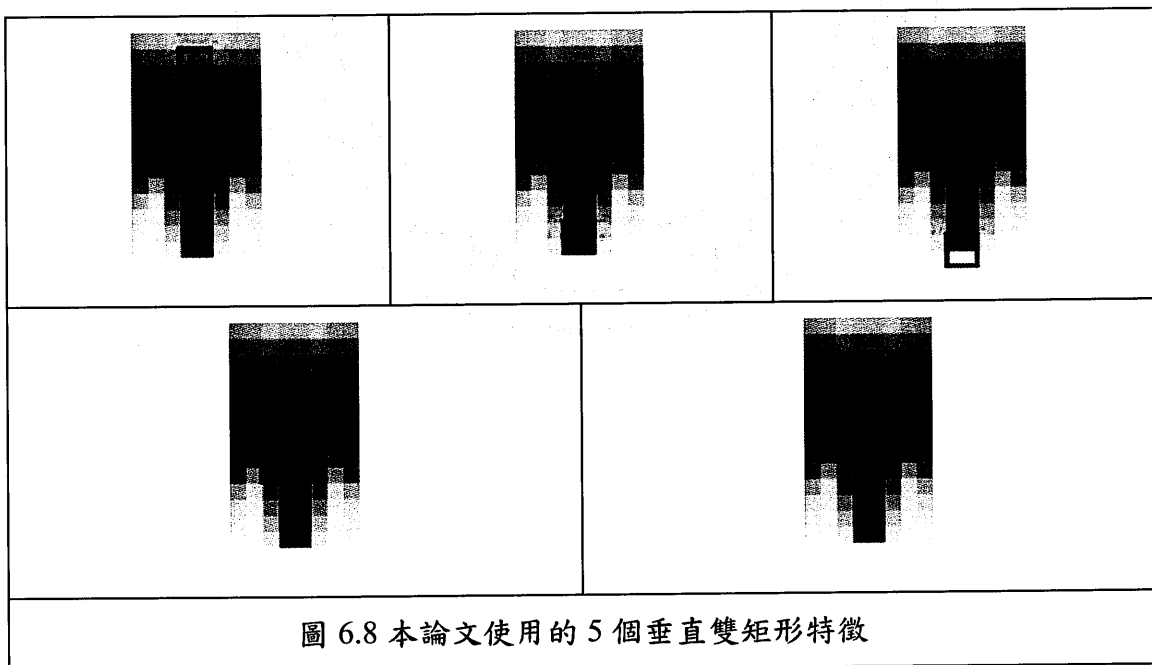
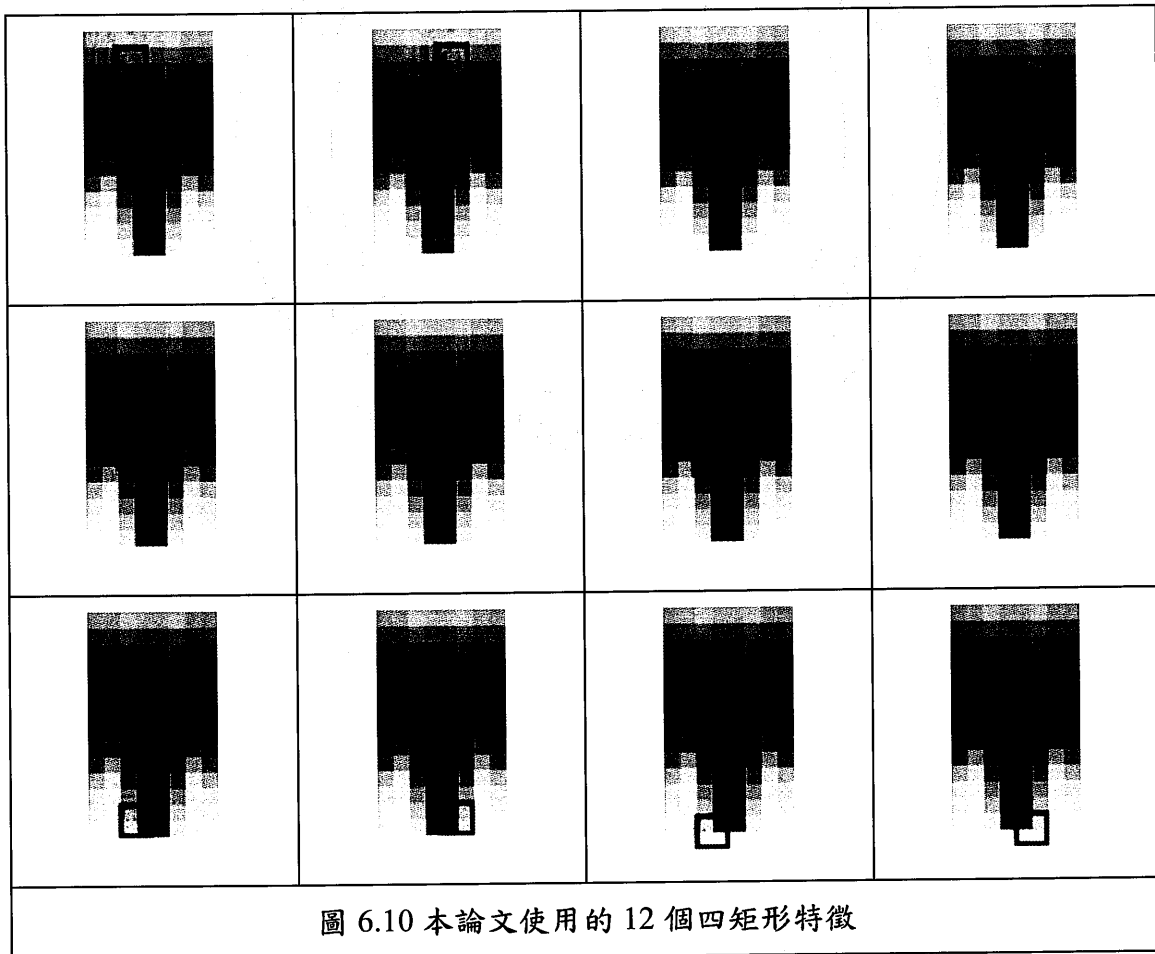
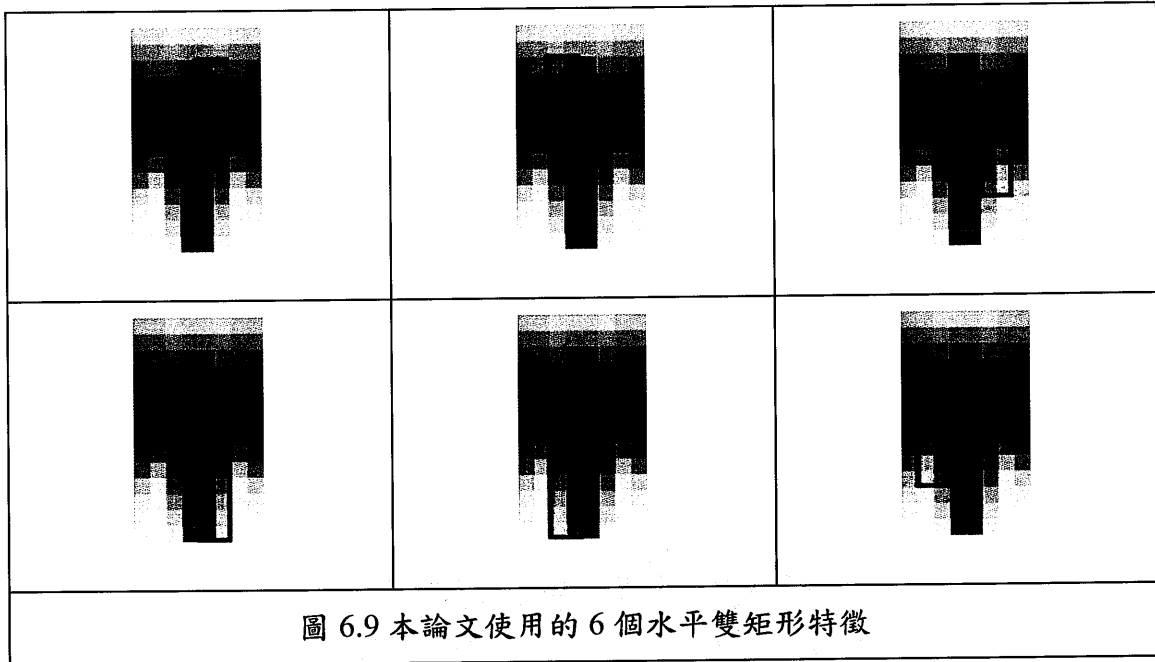


圖 6.8 本論文使用的 5 個垂直雙矩形特徵



6.3.2 共變異矩陣描述子之萃取範圍選取

共變異矩陣描述子是由共變異矩陣萃取而成，這些共變異矩陣紀錄著許多特徵影像之間的交互關係，而這些特徵影像即為輸入影像之色彩空間分量、邊緣資訊、梯度方向、濾波器響應...等，而這樣子由許多特徵影像所組合而成的資訊稱之為基於外觀(appearance-based)的資訊，也就是說共變異矩陣描述子是一種基於外觀的特徵檢測器。

基於外觀的特徵檢測器在檢測一個行人影像時，其檢測的方式與我們人類在判斷一個眼前的物體是否為行人時相似，其相似之處在於都是藉由分析各種不同的特徵所結合而成的外觀資訊來判斷。而兩者不同的地方只在於，基於外觀的特徵檢測器所得到的外觀資訊需要靠分類器來做判斷，而人類則是使用腦將眼前訊息整合成外觀資訊直接做判斷。

因此在選擇共變異矩陣描述子的特徵萃取範圍時，可以使用比較直觀的方式直接使用行人外觀上比較顯著的部位，例如：全身、頭部與軀幹的組合、頭部與肩膀的組合、上肢與軀幹的組合以及雙腳...等，本論文所選取的共變異矩陣描述子的特徵萃取範圍如圖 6.11 所示：



圖 6.11 共變異矩陣描述子的萃取範圍

6.4 Fisher 分類器的訓練流程

本行人偵測系統的訓練以及辨識過程皆以偵測窗所擷取的影像大小為基礎，並且採用了兩種不同的特徵粹取方式：Haar-like 特徵和共變異矩陣描述子，其中 Haar-like 特徵由圖 6.8 至圖 6.10 中所標示的區域在一個偵測窗中萃取特徵值之後構成一個 23 維的向量。而共變異矩陣描述子由圖 6.11 中所標示的區域在一個偵測窗中萃取特徵值之後構成 5 個 36 維的向量，也因此訓練 Fisher 分類器時需要使用兩種不同的方式來進行。

6.4.1 基於 Haar-like 特徵的 Fisher 分類器之訓練流程

基於 Haar-like 特徵的 Fisher 分類器之訓練流程主要分成兩個部分，第一部分是 Haar-like 特徵萃取、第二部分是 Fisher 線性判別分析運算。基於 Haar-like 特徵的 Fisher 分類器之訓練流程主要步驟如下(圖 6.12)所示：

1. 計算從樣本集取出的正樣本影像以及負樣本影像之積分影像。
2. 在正樣本的積分影像中行人所在的位置擷取一個 64×128 大小的積分影像；在負樣本的積分影像中的任意位置隨機擷取一個 64×128 大小的積分影像。
3. 使用圖 6.8 至圖 6.10 中的 Haar-like 特徵模板在 64×128 的積分影像上計算該影像的 Haar-like 特徵值，因為有 23 個 Haar-like 特徵模板，故特徵值是一個 23 維的向量。
4. 重複步驟 1. 到步驟 3.，直至蒐集到所需的樣本數量為止。
5. 最後將所有 23 維的 Haar-like 特徵值做 Fisher 線性判別分析運算，最後得到一個最佳投影向量 W_{opt} 以及分類器閾值 θ 。

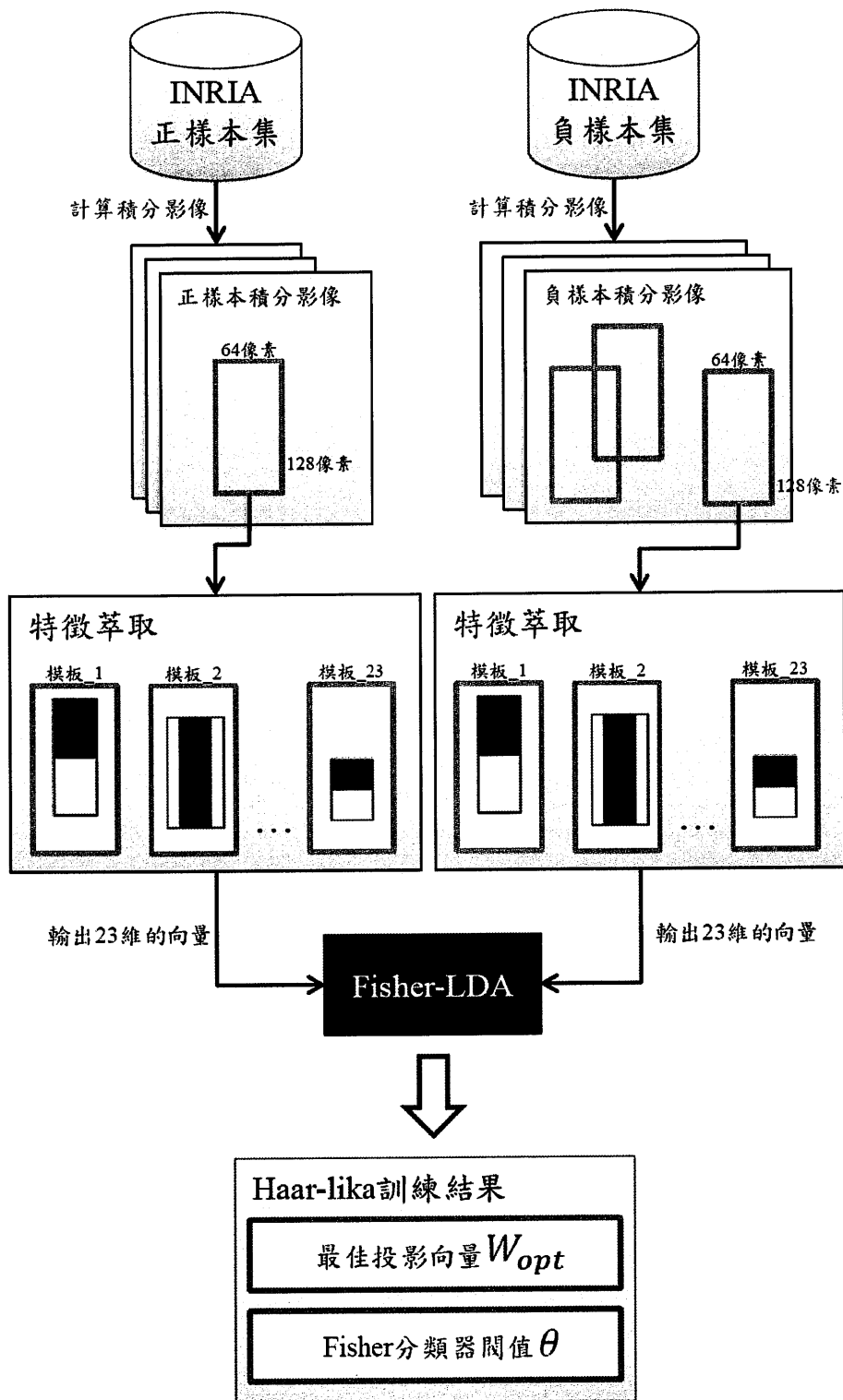


圖 6.12 基於 Haar-like 特徵的 Fisher 分類器之訓練流程

6.4.2 基於共變異矩陣描述子的 Fisher 分類器之訓練流程

基於共變異矩陣描述子的 Fisher 分類器，其訓練流程與基於 Haar-like 特徵的 Fisher 分類器不同的地方在於，前者每一個偵測窗中的資訊是 23 個 Haar-like 特徵模板(特徵萃取範圍)所組合而成的 23 維向量來表示，而後者則是每一個特徵萃取範圍皆能產生一個 36 維的向量，因為本論文採用 5 個共變異矩陣描述子萃取範圍，如圖 6.11 所示，故最終每一個偵測窗中的資訊會以 5 個 36 維的向量來表示。

基於共變異矩陣描述子的 Fisher 分類器之訓練流程主要分成兩個部份：第一部分是共變異矩陣描述子的計算、第二部分是 5 個 Fisher 線性判別分析運算。基於共變異矩陣描述子的 Fisher 分類器訓練流程如下列以及圖 6.13、6.14、6.15 所示：

1. 在正樣本的原始影像中行人所在的位置擷取一個 64×128 大小的影像；在負樣本的原始影像中的任意位置隨機擷取一個 64×128 大小的影像。
2. 藉由 4.4 式將上個步驟所取得之 64×128 大小的原始影像轉化成特徵影像，其中包含原始影像各個像素的： x 座標、 y 座標、 x 方向的一階偏導數絕對值 $|I_x|$ 、 y 方向的一階偏導數絕對值 $|I_y|$ 、 x 方向的二階偏導數絕對值 $|I_{xx}|$ 、 y 方向的二階偏導數絕對值 $|I_{yy}|$ 、梯度強度 $\sqrt{I_x^2 + I_y^2}$ 以及梯度方向 $\tan^{-1} \frac{|I_x|}{|I_y|}$ 。
3. 接著再以第四章所敘述的方法計算圖 6.11 所標示之共變異矩陣描述子萃取 5 個範圍中的共變異矩陣，每個萃取範圍之計算結果皆為一個 8×8 的對稱矩陣，因此只需要取矩陣的上三角矩陣之 36 個元素，並且由左而右由上而下的依序排列成一個 36 維的向量即為該範圍的共變異矩陣描述子，因此每個偵測窗會有 5 個共變異矩陣描述子來表示其影像資訊。
4. 重複步驟 1. 至步驟 3.，直至蒐集到所需的樣本數量為止。
5. 最後將各萃取範圍所對應的共變異矩陣描述子分別進行 Fisher 線性判別分析運算，最後會得到 5 個不同的最佳投影向量 W_{opt} 以及分類器閾值 θ 。

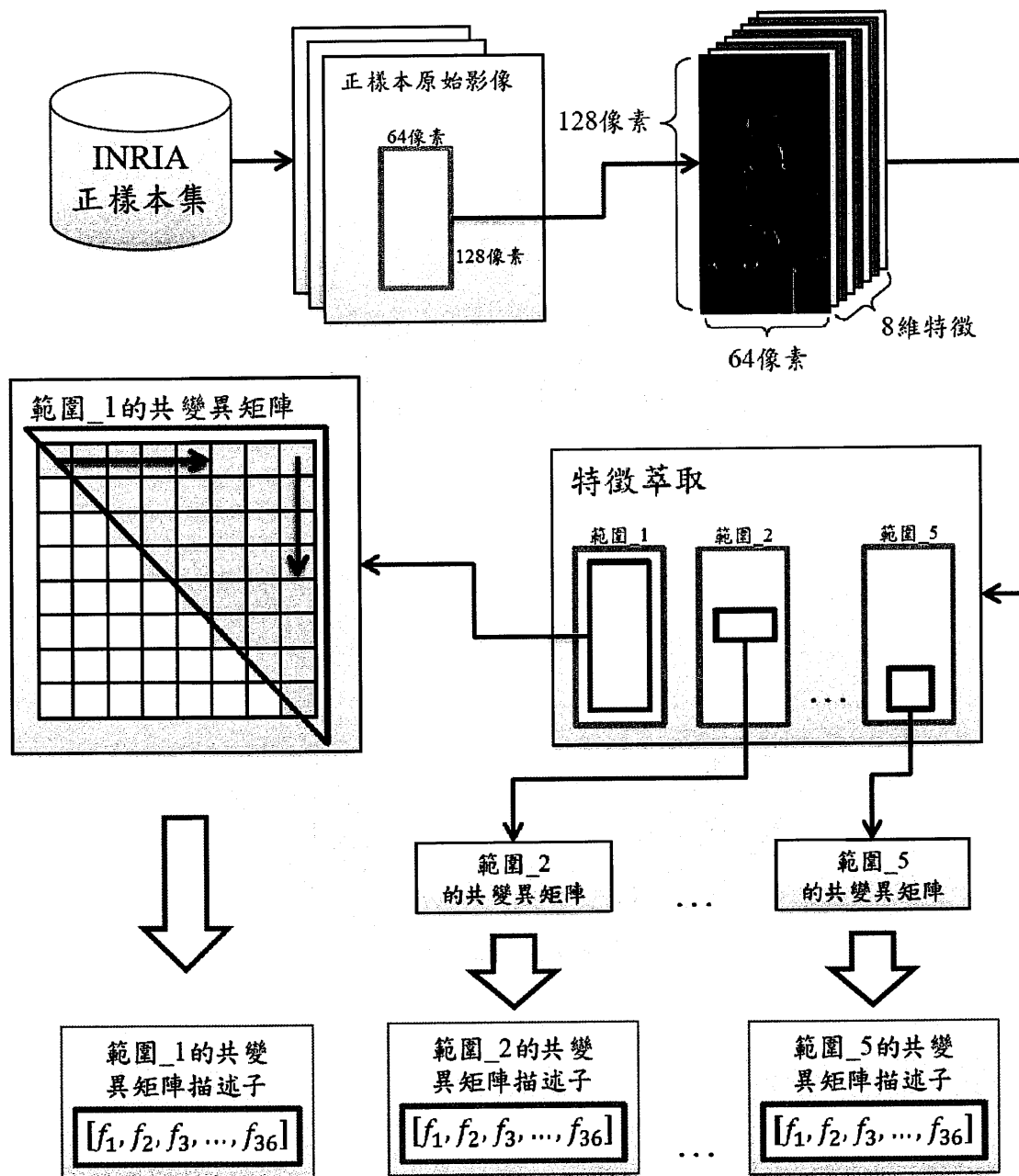


圖 6.13 正樣本的共變異矩陣描述子計算流程

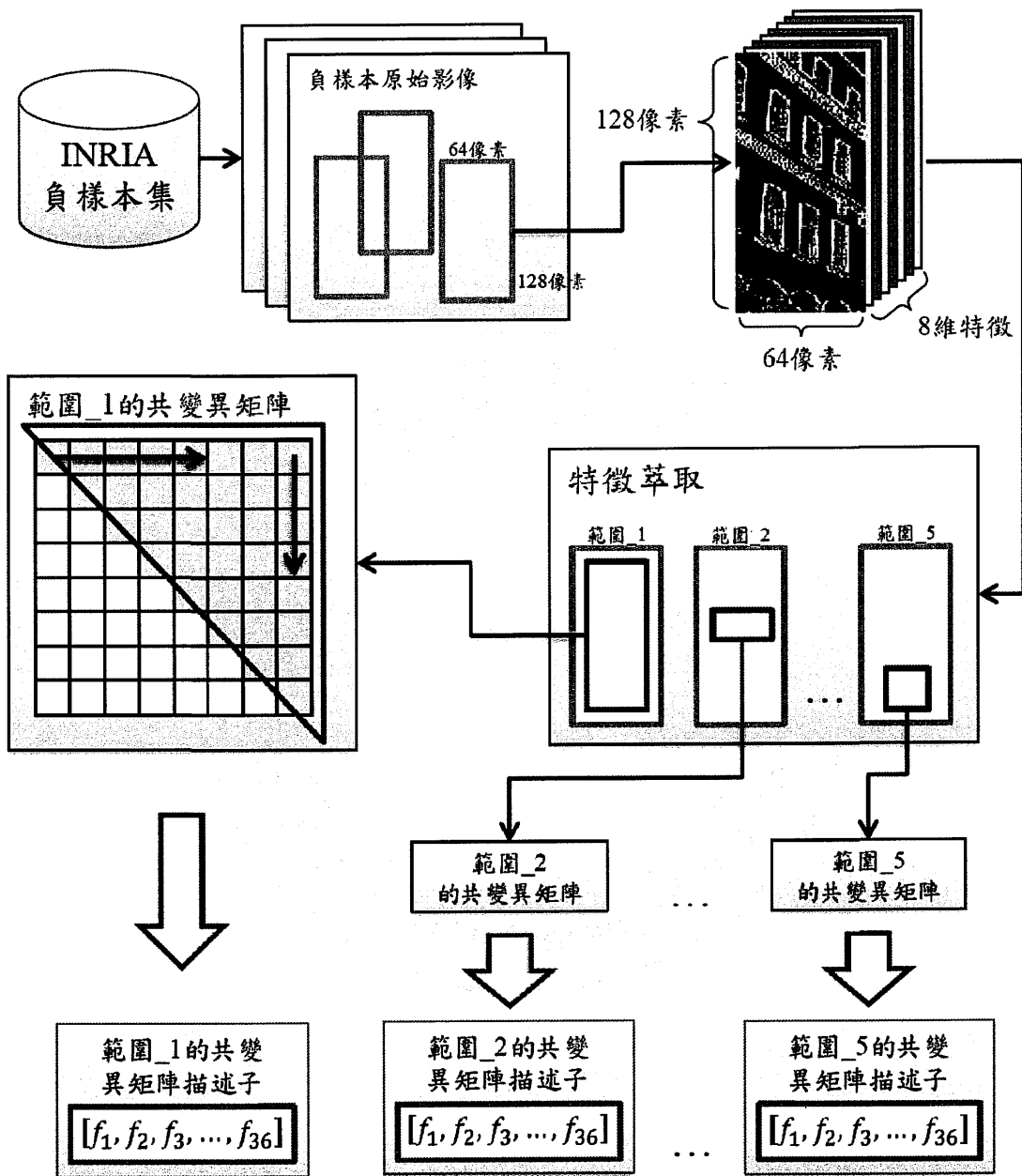


圖 6.14 負樣本的共變異矩陣描述子計算流程

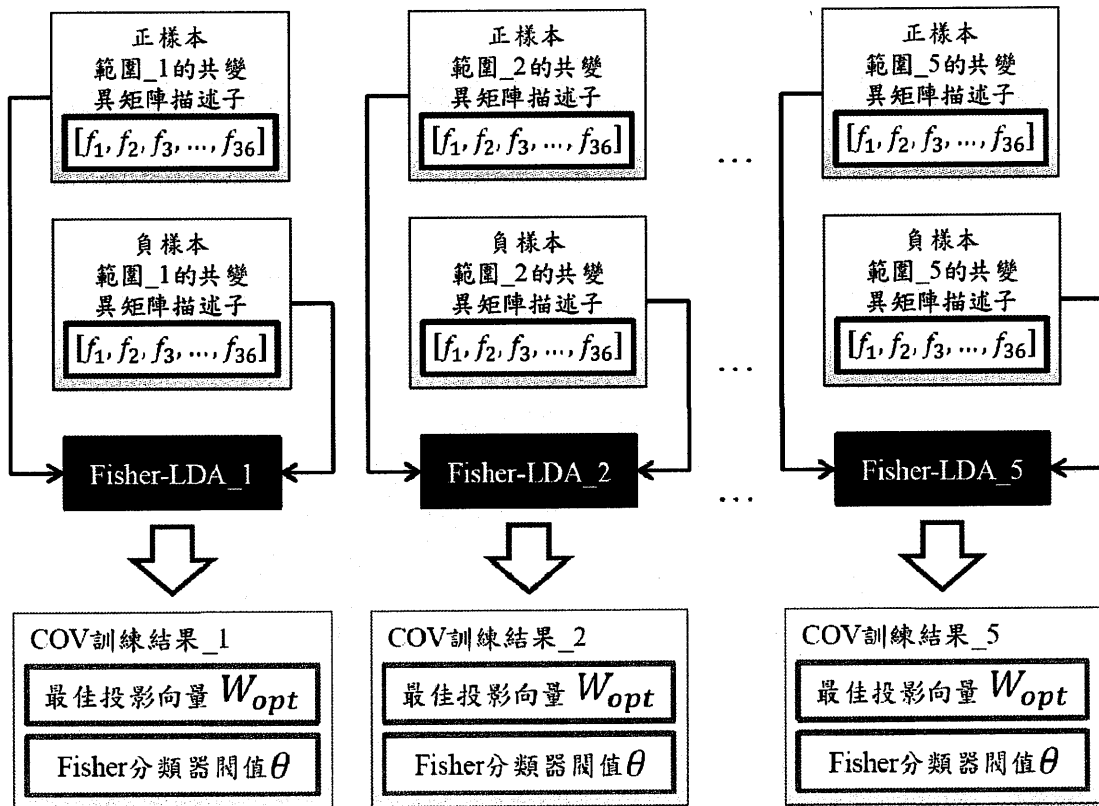


圖 6.15 基於共變異矩陣描述子的 Fisher 分類器訓練流程

6.5 串級的行人偵測系統

為了使行人偵測可以更有效率的進行，本行人偵測系統採用串級的偵測流程，其主要是由兩層 Fisher 分類器所構成的模組串接而成的雙層行人偵測系統：

1. 以 Haar-like 特徵為基礎的 Fisher 候選區域篩選模組。
2. 以共變異矩陣描述子為基礎的 Fisher 行人分類模組。

本行人偵測系統並非將整張影像傳送給 Fisher 分類器做判別，而是採用以偵測窗為基礎的偵測方式使偵測窗滑動於輸入影像之上，並將偵測窗的範圍資訊(大小以及位置)傳送給行人偵測系統的候選區域篩選模組，再由候選區域篩選模組取得該範圍內的影像，如圖 6.16 所示。偵測窗的移動方式為由輸入影像的左上至右下，每次向右或向下移動 1/4 個偵測窗，偵測窗大小的初始值為寬 40 像素、高固定為寬的兩倍，將輸入影像掃描一次後偵測窗即放大一次，並以偵測窗的寬為基礎每次遞增 8 個像素，直到偵測窗的寬為 64 像素為止，因此每張輸入影像皆以不同大小的偵測窗掃描四次。

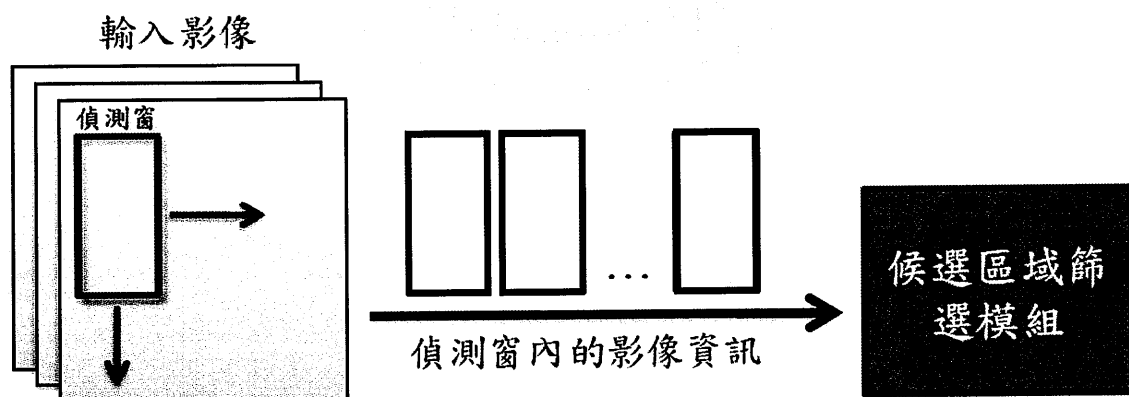


圖 6.16 以偵測窗擷取輸入影像的資訊

6.5.1 以 Haar-like 特徵為基礎的候選區域篩選模組

候選區域篩選模組位於行人偵測系統的前級，主要的任務是快速的找出影像中行人可能存在的候選區域，其偵測的步驟如下列以及圖 6.17 所示：

1. 依據偵測窗的範圍資訊擷取輸入影像一小部分的影像。
2. 以 Haar-like 特徵計算單元計算特徵值。Haar-like 特徵計算單元會先計算步驟 1. 所取得之影像大小，隨後將各個 Haar-like 特徵模板依據該影像的大小作等比例的縮放，再計算出每個 Haar-like 特徵模板的特徵值，並將其構成一個 23 維的向量輸出。
3. 利用章節 6.4.1 所訓練出來的 Fisher 分類器做判斷，若非行人則丟棄偵測窗的範圍資訊，若為行人則將偵測窗的範圍資訊送至行人分類模組。

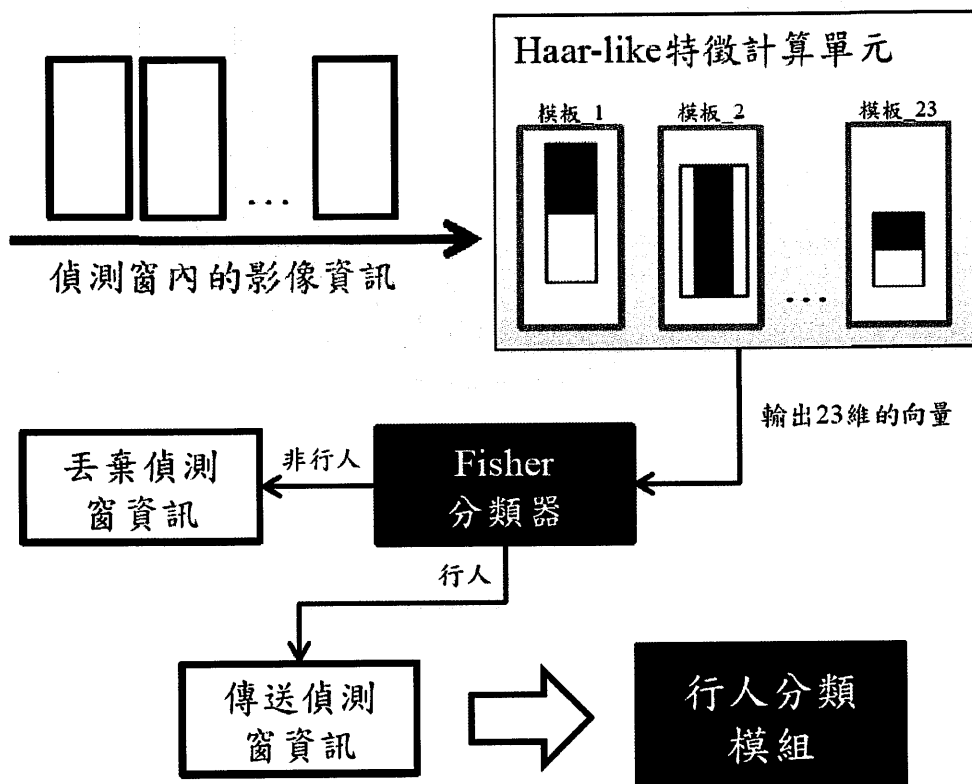


圖 6.17 以 Haar-like 特徵為基礎的候選區域篩選模組

6.5.2 以共變異矩陣描述子為基礎的行人分類模組

行人分類模組位於行人偵測系統的後級，主要的任務是針對已通過候選區域篩選模組之候選區域進一步的以更精確的方法進行檢測，其偵測的步驟如下(圖 6.18)所示：

1. 將原始影像轉換成特徵影像。
2. 根據候選區域篩選模組所傳來的偵測窗資訊，從步驟 1. 所產生的特徵影像中擷取對應範圍的特徵影像資訊。
3. 以共變異矩陣計算單元計算偵測窗中各個子區域的共變異矩陣。
4. 將偵測窗中各個子區域的共變異矩陣取上三角矩陣，並且以由左至右、由上而下的方向蒐集上三角矩陣中的元素，構成 36 維的共變異矩陣描述子。
5. 將得到的共變異矩陣描述子分別以章節 6.4.2 所訓練出來的 5 個 Fisher 分類器做判斷。
6. 將 5 個 Fisher 分類器的判斷結果進行投票。每個分類器在訓練時會依其準確度給與權重，而所有分類器之權重值的加總為 1，而在分類時當權重值總和超過 0.5 即判斷為該區域有行人存在。
7. 將已確定有行人存在的區域之座標、寬以及高輸出至紀錄偵測結果的陣列中，該陣列內的資料將用於標示行人的用途。

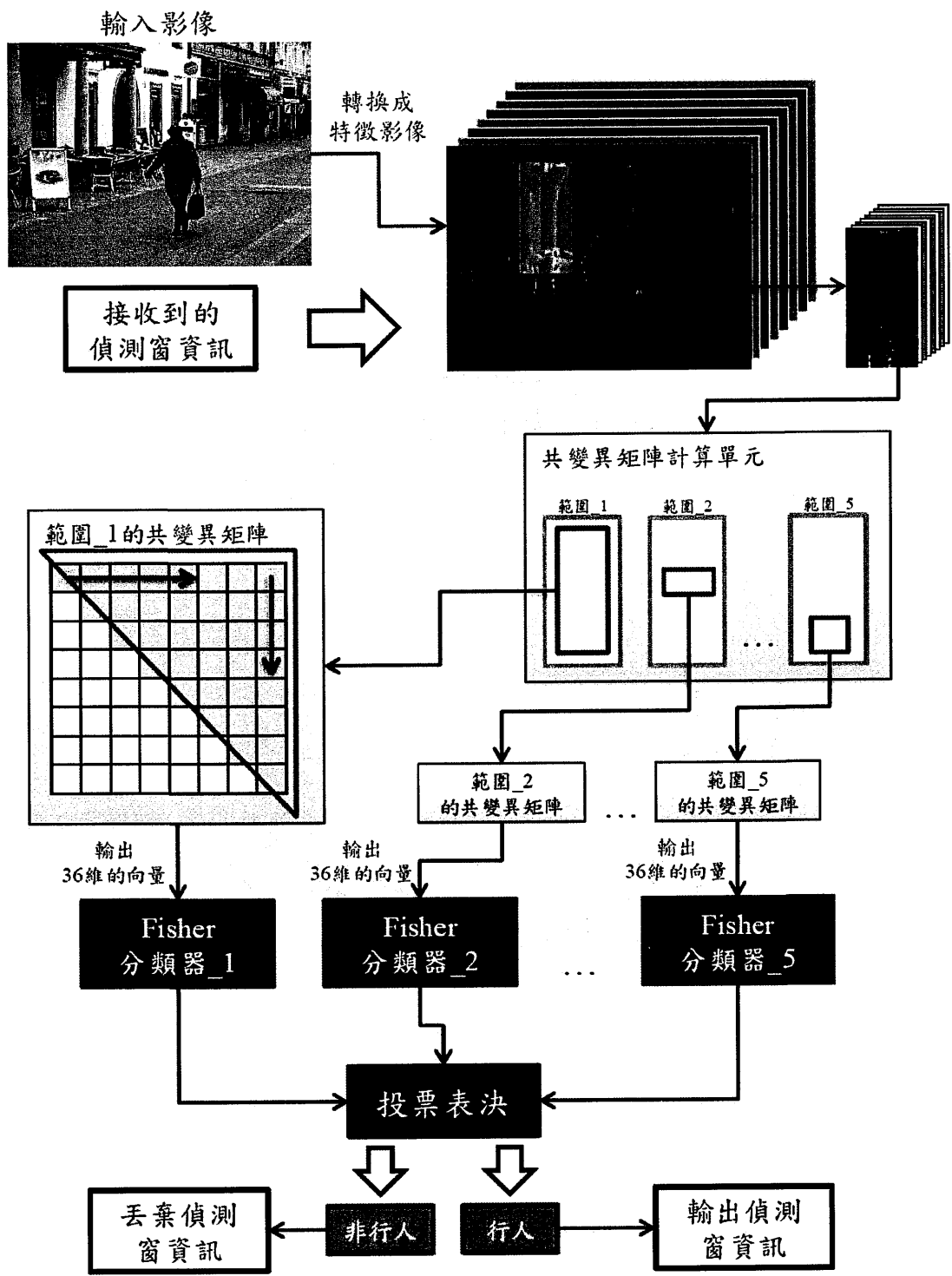


圖 6.18 以共變異矩陣描述子為基礎的行人分類模組

6.5.3 偵測結果之標示區域的整併

當一個影像經過行人偵測系統的全域檢測後，會將通過檢測的偵測窗之座標、寬以及高都記錄下來，並且作為標示行人所在位置的依據。但是因為檢測器敏感度的問題，若這些標示資訊未經過適當的整理與合併，直接對原始影像進行標示的話，那麼行人影像的附近通常都會有多個重疊的標示結果，而造成觀測的困難。因此對這些標示資訊進行適當的整理與合併是必須的。其合併的流程如圖 6.19 所示：

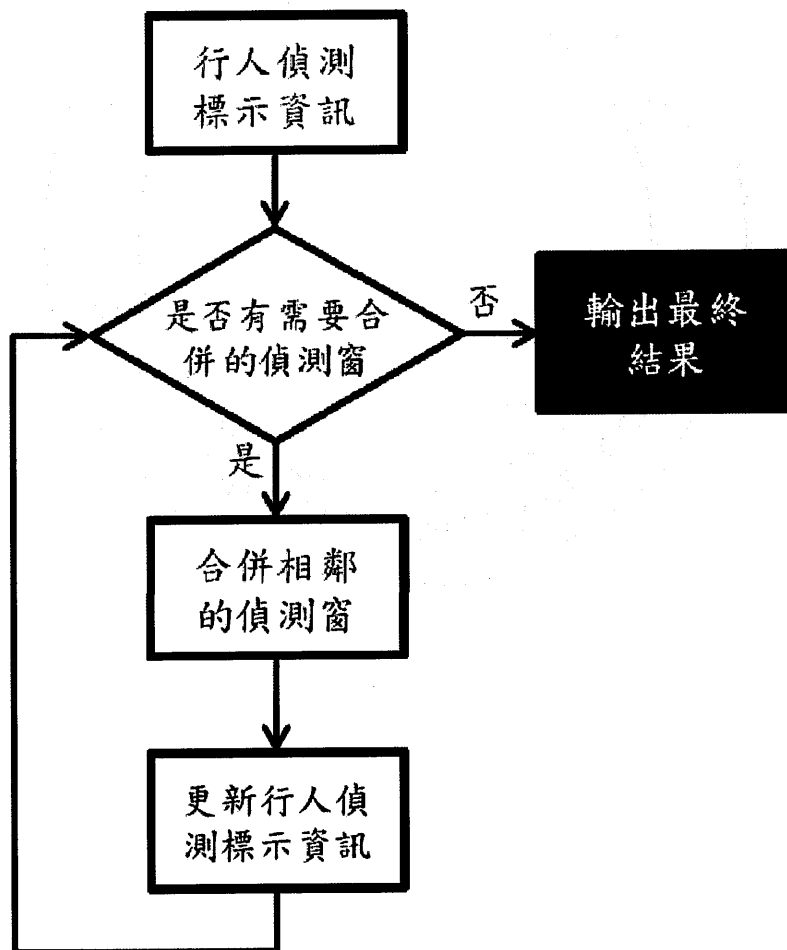
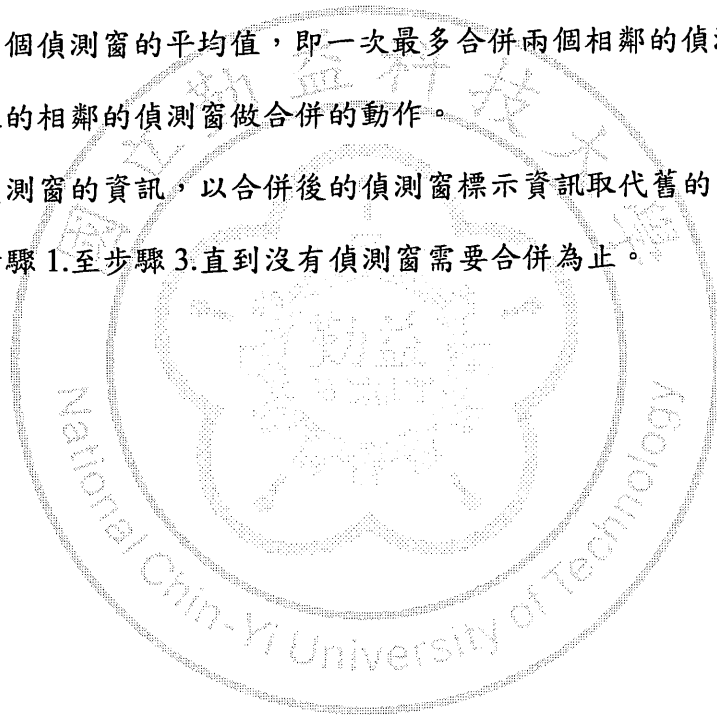


圖 6.19 偵測結果之標示資訊的合併流程

其合併的詳細步驟如下列所示：

1. 判斷紀錄行人偵測結果之陣列中的偵測窗資訊在兩兩互相配對之組合中
有沒有需要合併的偵測窗。根據測試的結果當兩個偵測窗重疊的面積超過
30%時將其合併會有較好的標示結果。
2. 若無需合併則輸出最終結果在影像上標示出行人所在的位置；若需要合併
則將偵測窗的座標、及寬與高取平均值存在新的偵測結果陣列，每次最多
計算兩個偵測窗的平均值，即一次最多合併兩個相鄰的偵測窗，不會對兩
個以上的相鄰的偵測窗做合併的動作。
3. 更新偵測窗的資訊，以合併後的偵測窗標示資訊取代舊的。
4. 重複步驟 1.至步驟 3.直到沒有偵測窗需要合併為止。



第七章 實驗結果與分析

7.1 概述

本論文將以兩個方式呈現所提出之行人偵測系統的偵測效能，首先以本論文所提出的行人偵測系統與[3]中的行人偵測系統做效能以及執行速度的比較，再來會以行人偵測系統實際運行的截圖呈現實驗結果。

7.2 行人偵測系統效能評估實驗

在行人偵測系統效能評估實驗中，主要是以本文的行人偵測系統與[3]中的行人偵測系統做比較，在[3]的實驗中已經將現今用於行人偵測的特徵萃取方法(Haar-like 特徵、Region Covariance、HOG、LRF)以及分類器架構(AdaBoost、SVM、類神經網路)做了詳細的交叉比對報告，且進而提出了最有效率的分類器架構，因此是本論文的主要的參考文獻以及本論文所提出的行人偵測系統之效能的參考對象。

本實驗使用 INRIA 的行人影像資料庫，首先利用訓練集的正樣本資料以及由負樣本訓練集隨機抽樣的資料對分類器做訓練，再來在訓練完成後對測試集的正樣本資料以及由負樣本測試集隨機抽樣的資料做測試以蒐集行人偵測系統之效能的數據。實驗步驟為：(1)訓練、(2)測試、(3)數據紀錄。每次的實驗皆會紀錄三筆數據：(1)偵測率、(2)誤判率、(3)總錯誤率。每次的實驗皆從 INRIA 正樣本訓練集隨機抽出 2,000 筆資料，同樣的也從 INRIA 負樣本訓練集隨機抽出 2,000 筆負樣本資料，並且為了公平起見一組樣本將同時用於 Fisher 分類器以及 AdaBoost 分類器的訓練。

在經過 2,000 次的實驗，基於 Haar-like 特徵的 Fisher 分類器所得到的測試數據，如圖 7.1 所示：

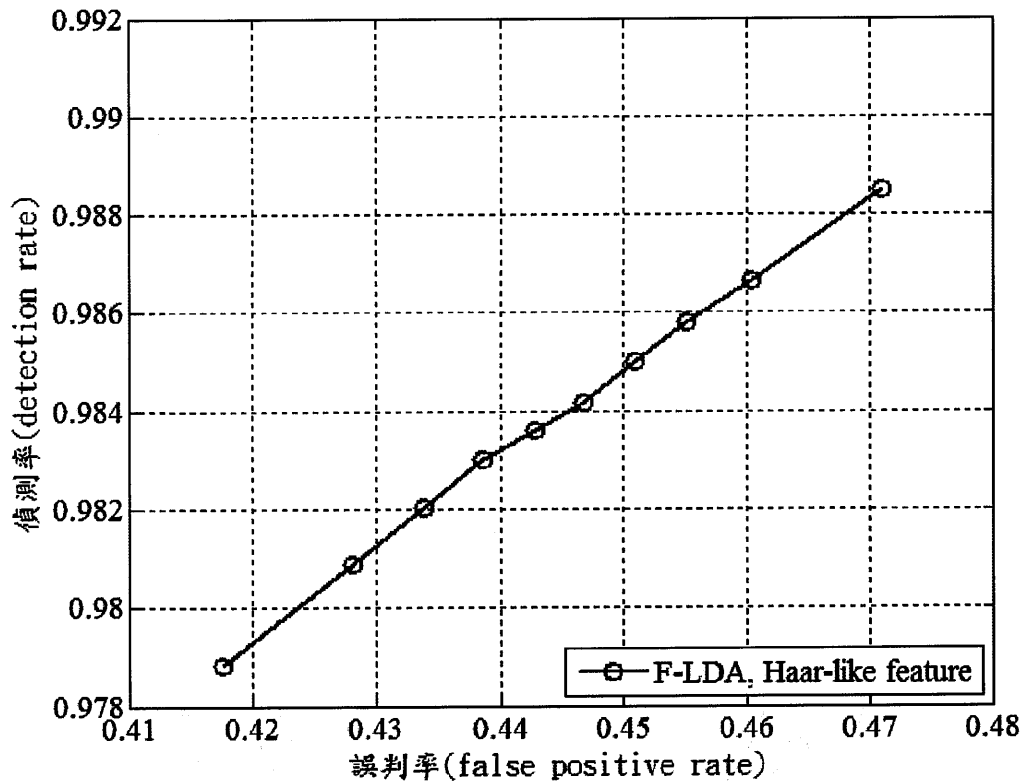


圖 7.1 基於 Haar-like 特徵的 Fisher 分類器性能曲線

前級分類器其主要的功能為快速的篩選出行人可能存在的候選區域，為了能夠盡量的讓所有可能為行人的影像通過檢測因此其目標偵測率通常都會設在 0.99 以上，為了能夠剔除掉大部份之不可能為行人的區域因此目標誤判率大約在 0.5 以下即可，故在此不做前級分類器的性能比較。由圖 7.1 可以看出本文的前級分類器的偵測率(detection rate)為 0.978 至 0.989，而誤判率(false positive rate)為 0.42 至 0.47 左右。

在經過 2,000 次的實驗，基於共變異矩陣描述子的 Fisher 分類器以及基於共變異矩陣描述子的 AdaBoost 分類器所得到的測試數據，如圖 7.2 所示：

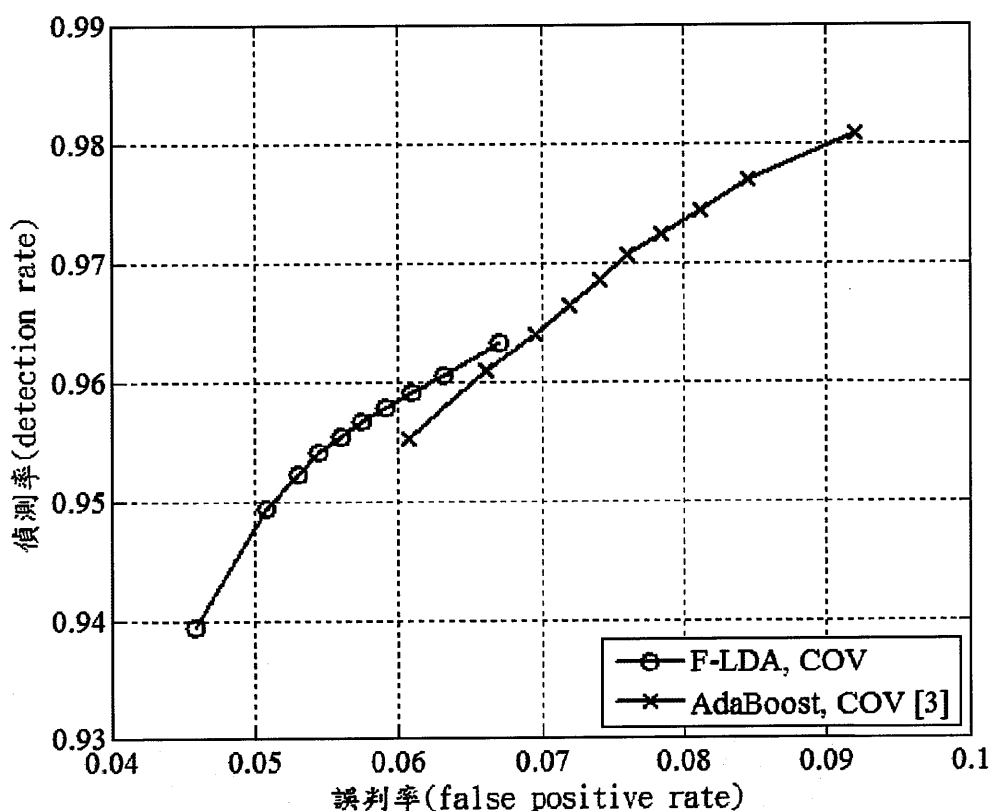


圖 7.2 本文之 Fisher 分類器及[3]的 AdaBoost 分類器之性能比較

在圖 7.2 中可以看出基於共變異矩陣描述子的 Fisher 分類器其偵測率落在 0.94 至 0.964 之間，而誤判率(false positive rate)落在 0.046 至 0.068 之間；基於共變異矩陣描述子的 AdaBoost 分類器其偵測率在 0.955 至 0.981 之間，而誤判率(false positive rate)在 0.061 至 0.092 之間，以圖 7.2 的曲線圖來說，曲線的位置越往左上方代表該分類器的效能越好，雖然 AdaBoost 分類器的偵測率可以高達 0.981，這是 Fisher 分類器所無法到達的，不過由兩條曲線的分佈方式可以看出，這兩種分類器的效能可以說是非常的接近。

再以分類器的運行速度來看，首先是本實驗之行人偵測系統的運行的環境，如下表 7.1 所示：

表 7.1 行人偵測系統效能評估實驗之運行環境

CPU	Intel Core Duo 1.83GHz
記憶體	2GB
使用的程式語言	MATLAB

而表 7.2 是 Haar-like 特徵及共變異矩陣描述子搭配不同分類器時的偵測速度：

表 7.2 比較特徵萃取方式與分類器搭配之速度

	偵測窗/秒
Haar-like, F-LDA	2640
Haar-like, Adaboost	461
COV feature, F-LDA	546
COV feature, Adaboost	260

由表 7.2 可以看出以 Fisher 線性判別分析為基底的分類器不管是使用 Haar-like 特徵或是共變異矩陣描述子，其偵測速度均比使用 AdaBoost 的分類器快上許多，但由圖 7.1 與 7.2 可以看出兩個分類器皆以極少量的精確度做為代價換取更快的執行速度。

此外因為是使用 MATLAB 建構此行人偵測系統，故運行速度會比使用 C 語言或者 C++ 建構的行人偵測系要來得慢，若以 C 語言或 C++ 來實踐本行人偵測系統的話必定能更進一步的提升其運行速度。

7.3 系統實際運行測試

系統在實際運作時因考慮到需要模擬攝影鏡頭在車輛上隨著車輛移動的情況，故在此是在筆記型電腦前方架設一個網路攝影鏡(Web Camera)來做為行人偵測系統的視訊輸入來源，本實驗之行人偵測系統運行的環境如表 7.3 所示：

表 7.3 系統實際運行結果之實驗環境

CPU	Intel Core Duo 1.83GHz
記憶體	1GB
使用的程式語言	MATLAB
網路攝影鏡頭	Logitech Webcam C500
影像串流輸入速度	320 × 240 像素 每秒 30 幀

本行人偵測系統實際運行速度如表 7.4 所示：

表 7.4 行人偵測系統執行時間

行人偵測平均花費時間	0.19 ± 0.02 秒/幀
------------	-----------------

行人偵測系統運行的部分結果如圖 7.3 至圖 7.6 所示，所有偵測結果皆為攝影鏡頭移動的狀況下拍攝所呈現的偵測結果，其中圖 7.3 為一般的情況下標示出的行人。圖 7.4 為行人在陰影中，光線不足的情況所標示出的行人。圖 7.5 為行人在日光燈所造成之光線不均勻的情況下所標示出的行人。圖 7.6 為行人重疊時所標示出的行人。

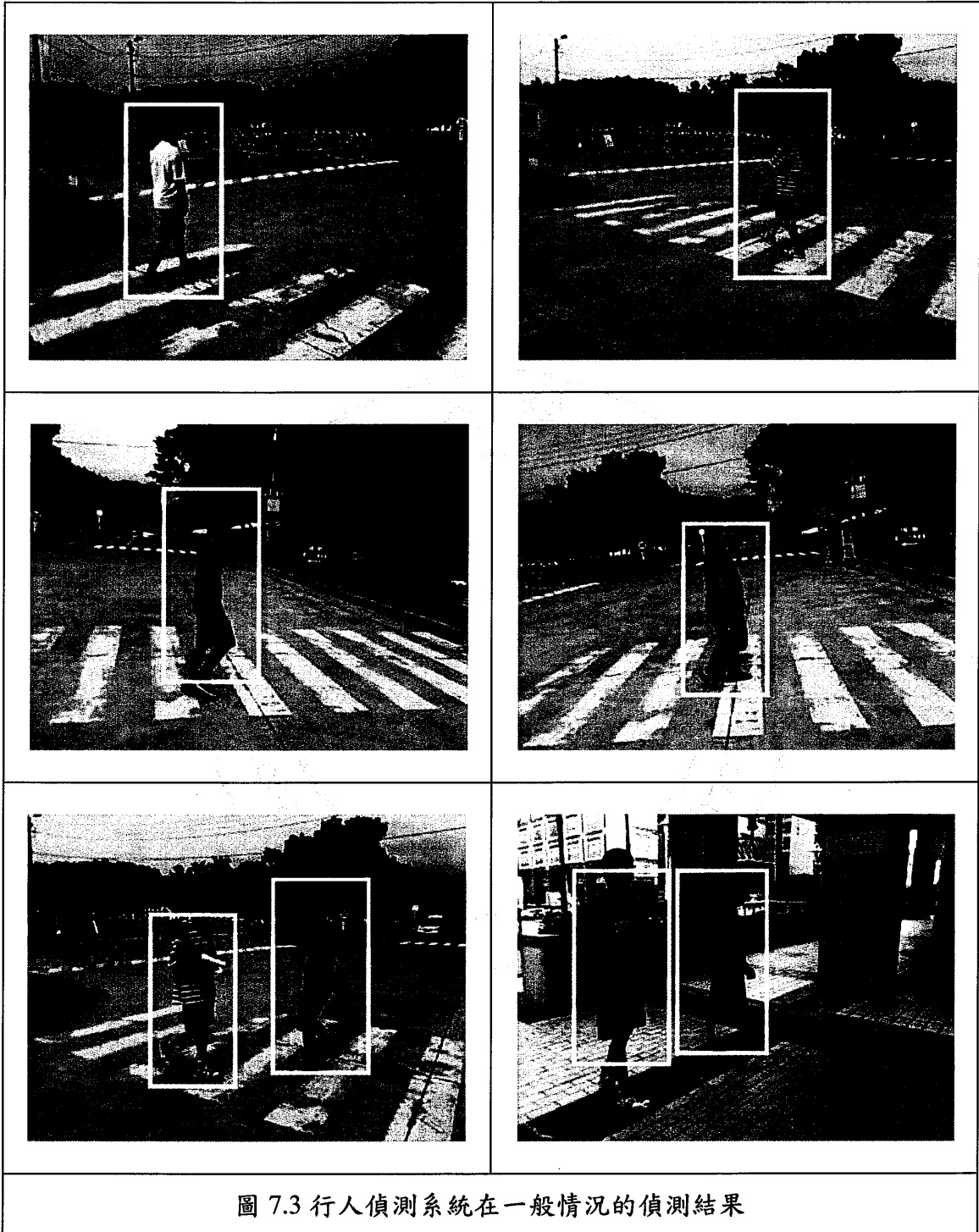


圖 7.3 行人偵測系統在一般情況的偵測結果

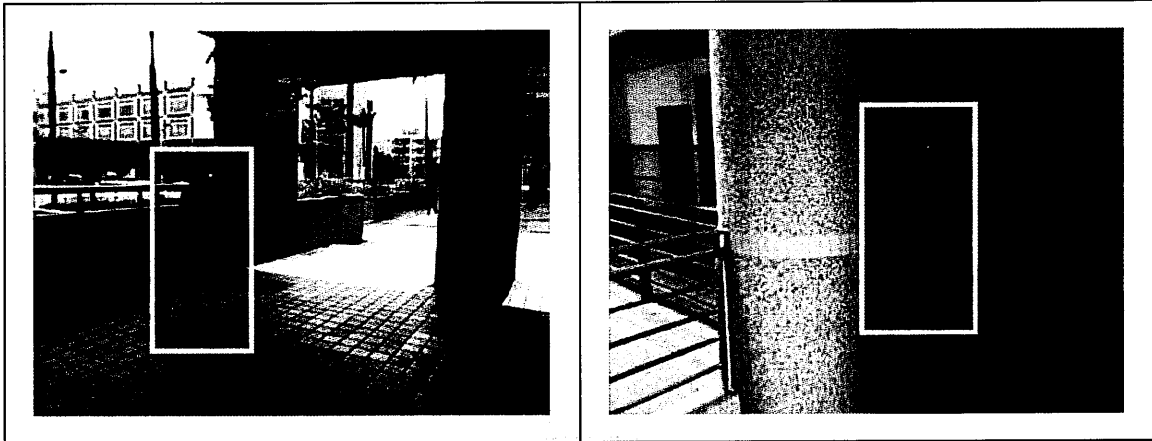


圖 7.4 行人偵測系統在光線不足時的偵測結果

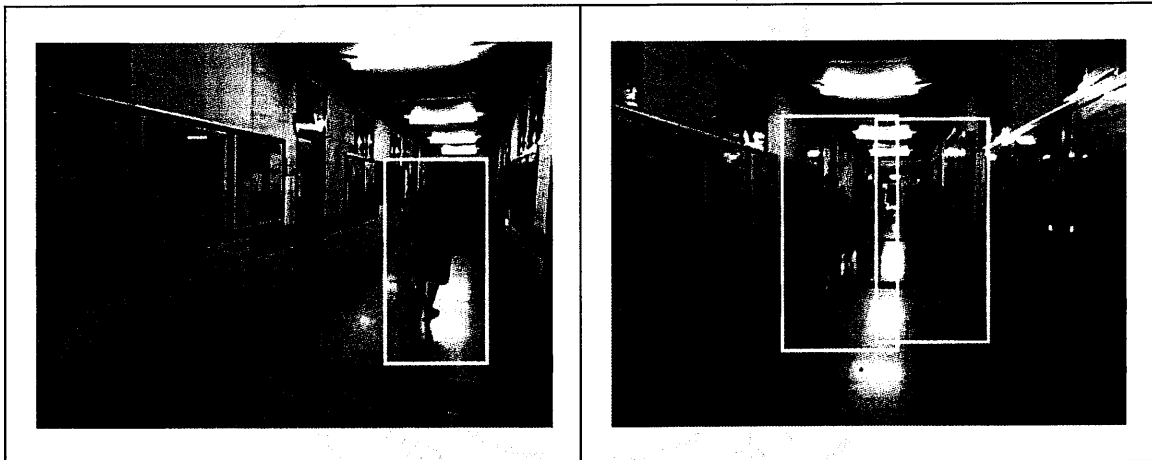


圖 7.5 行人偵測系統在光線不均勻時的偵測結果

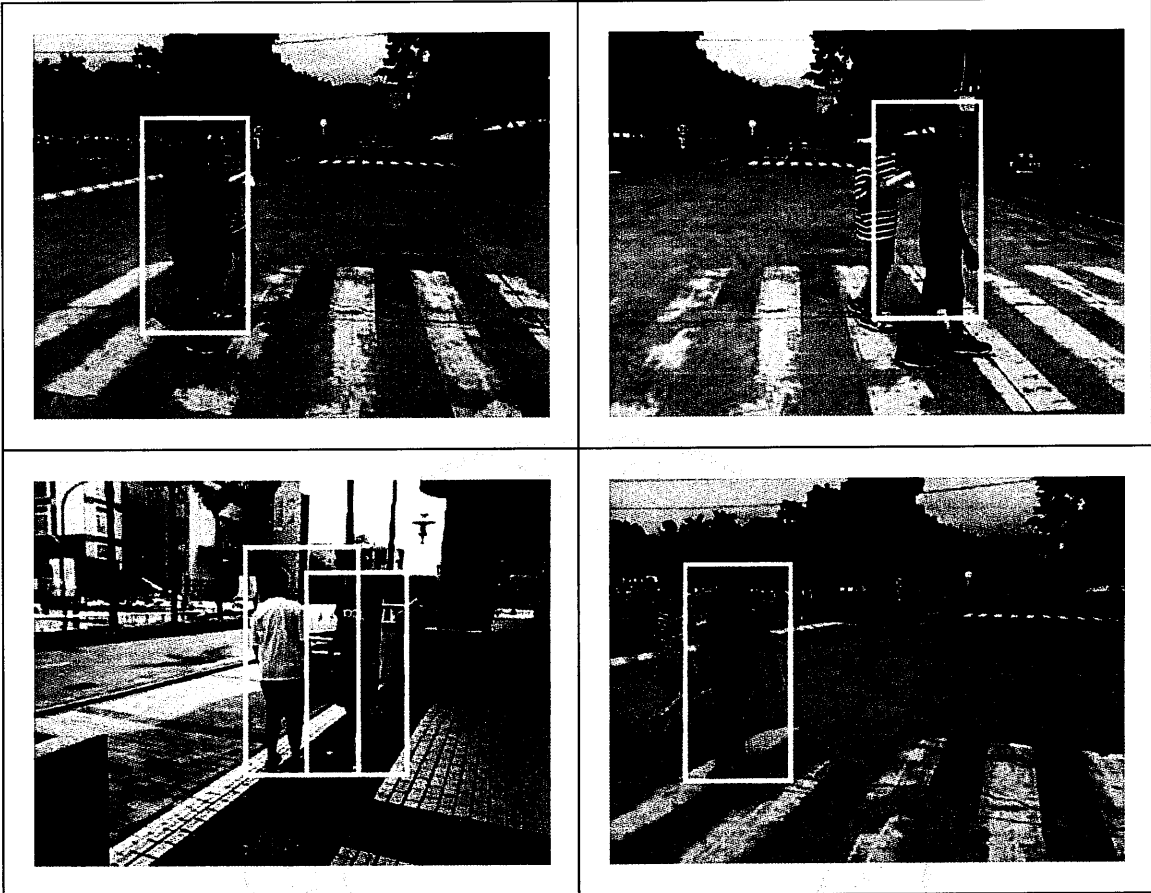


圖 7.6 行人偵測系統在行人重疊及靠近時的偵測結果

第八章 結論及未來展望

8.1 結論

本論文提出一套以 Fisher 線性判別分析為基礎的快速行人偵測系統架構，系統的功能為在攝影鏡頭也處於移動的狀態下，即時的偵測出現在影像中的行人並且將其標示於影像中。

本論文以兩個主要的部分建構出行人偵測系統，前級的候選區域篩選模組以及後級的行人分類模組。候選區域篩選模組最主要的功能是快速的在輸入影像中找出行人可能存在的候選區域，因此本論文採用 Haar-like 特徵萃取方式配合 Fisher 線性判別分析訓練出一個 Fisher 分類器，其偵測速度比採用 AdaBoost 演算法的分類器快了將近 6 倍；行人分類模組最主要的功能是對候選區域篩選模組所傳來的影像資料再做進一步的確認，本論文採用共變異矩陣描述子的特徵萃取方式配合 Fisher 線性判別分析訓練出 5 個 Fisher 分類器構成行人類模組，其分類速度比採用 AdaBoost 演算法的分類器快了 2 倍，並在實驗結果中證明 Fisher 分類器與 AdaBoost 的分類器之效能相去不遠，其效能在理想的條件下甚至比採用 AdaBoost 演算法的分類器還好上一些。

而在實際的測試實驗中顯示，本論文所提出的行人偵測系統在對於行人外觀、姿勢以及光線的改變也都能維持著良好的強健性，而偵測窗整併的演算法採用遞迴的方式逐漸的將過度重疊的偵測窗以位置、寬以及高的平均兩兩合併，因此可以更正確的標示出行人所在的位置而不會產生標示歪斜的問題。

8.2 未來展望

雖然本論文所提出之行人偵測系統對於光線的變化擁有著一定的強健性，但是在光源微弱時，就無法對行人所在的位置進行偵測。因為 Haar-like 特徵為計算

影像中之相鄰矩形區內像素值總和的差值，因此在光源不足的情況下該差值是一個趨近於 0 的值；而在共變異矩陣描述子的方面，因為其以特徵影像為計算的基礎，但是特徵影像中的一階偏導數、二階偏導數、梯度強度、梯度方向...等均需要計算相鄰像素之間的像素值之落差，在光源微弱時，影像中像素值的差異均趨近於 0，因此連共變異矩陣描述子也無法在低光源的時候正確的描述行人。而在未來可以考慮加入紅外線攝影鏡頭以偵測低光源的行人影像以及加強在光源充足時的偵測正確率。

目前所採用的偵測窗合併演算法因為是使用遞迴的方式將過度重疊的偵測窗合併，雖然能有效的提升標示的精確度，但也因此拖累了系統的運行速度，其計算的時間約占總偵測運算時間的 30%，其主要原因是行人偵測系統在掃描整個影像之後，因為偵測器敏感度的問題會在行人的附近留下許多重疊的標記，而重疊的標記若數量過於龐大會造成式窗整併的遞迴次數增加，因此若在未來可以使行人偵測系統的判斷更為精準或是找到更有效率的偵測窗整併方式，相信可以更進一步的提升此行人偵測系統在實際運行的效率。

參考文獻

- [1] M. Peden, R. Scurfield, D. Sleet, D. Mohan, A. A. Hyder, E. Jarawan, and C. Mathers, "World report on road traffic injury prevention," *Geneva, Switzerland: World Health Organization*, 2004.
- [2] N. Dalal, B. Triggs, "Histograms of Oriented Gradients for Human Detection," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 886–893, 2005.
- [3] S. Paisitkriangkrai, C. Shen, and J. Zhang, "Fast pedestrian detection using a cascade of boosted covariance features," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 8, 2008.
- [4] S. Paisitkriangkrai, C. Shen, and J. Zhang, "Performance evaluation of local features in human classification and detection," *IET Comput. Vis.*, vol. 2, no. 4, pp.236-246, 2008.
- [5] O. Tuzel, F. Porikli, and P. Meer, "Human detection via classification on Riemannian manifolds," *IEEE Computer Society International Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2007.
- [6] S. Munder, D.M. Gavrila, "An Experimental Study on Pedestrian Classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, November 2006.
- [7] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.
- [8] P. Viola and M. J. Jones, "Rapid object detection using a boosted cascade of simple

- features,” *IEEE Conf. on Computer Vision and Pattern Recognition*, Kauai, Hawaii, 2001.
- [9] C. P. Papageorgiou, M. Oren, and T. Poggio, “A general framework for object detection,” *In International Conference on Computer Vision*, 1998.
- [10] R. A. Fisher, “The use of multiple measures in taxonomic problems,” *Ann. Eugenics*, vol. 7, pp. 179 - 188, 1936.
- [11] O. Tuzel, F. Porikli, and P. Meer, “Region covariance: A fast descriptor for detection and classification,” *Proc. Eur. Conf. Comput. Vis.*, vol. 2, pp. 589–600, Graz, Austria, May 2006.
- [12] Y. Freund and R. E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” *Journal of Computer and System Sciences*, 55(1):119–139, August 1997.
- [13] C. Cortes and V. Vapnik, "Support-Vector Networks", *Machine Learning*, 20, 1995.
- [14] Wikipedia Encyclopedia. “Support vector machine, the free encyclopedia,” Jul. 2008, from http://en.wikipedia.org/wiki/Support_vector_machine.
- [15] C. Yuriy, “Ultra Rapid Object Detection in Computer Vision Applications with Haar-like Wavelet Features,” from http://www.codeproject.com/KB/audio-video/haar_detection.aspx
- [16] P. Belhumeur., J. Hespanha and D. Kriegman, “Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection,” *Proc. Fourth European Conf. on Computer Vision*, vol. 1, pp. 45-58, 1996.
- [17] <http://pascal.inrialpes.fr/data/human/>

- [18] Xian-Bin C., Hong Q., "A Low-Cost Pedestrian-Detection System With a Single Optical Camera," *IEEE Transaction on Intelligent Transportation System*, vol. 9, no. 1, March 2008.
- [19] Yanwei P., Yuan Y., Xuelong Li, "Gabor-Based Region Covariance Matrices for Face Recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 7, July 2008.
- [20] F. Porikli, O. Tuzel, P. Meer, "Covariance tracking using model update based on lie algebra", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 728–735, 2006.
- [21] W. Förstner, B. Moen, "A Metric for Covariance Matrices," *Technical report, Dept. of Geodesy and Geoinformatics, Stuttgart University*, 1999.
- [22] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, K.-R. Müller, "Fisher discriminant analysis with kernels," *IEEE International Workshop on Neural Networks for Signal Processing IX*, Madison, pp. 41–48, USA, August, 1999.