

國立中央大學

資訊工程研究所

碩士論文

高速公路上之道路牌文字偵測

Text Detection of Road Signs on Highway

研究生：曾韻榮

指導教授：范國清 博士

中華民國九十五年六月



國立中央大學圖書館 碩博士論文電子檔授權書

本授權書所授權之論文全文電子檔，為本人於國立中央大學，撰寫之碩/博士學位論文。(以下請擇一勾選)

- (☒)同意 (立即開放)
(☐)同意 (一年後開放)，原因是：_____
(☐)同意 (二年後開放)，原因是：_____
(☐)不同意，原因是：_____

以非專屬、無償授權國立中央大學圖書館與國家圖書館，基於推動讀者間「資源共享、互惠合作」之理念，於回饋社會與學術研究之目的，得不限地域、時間與次數，以紙本、微縮、光碟及其它各種方法將上列論文收錄、重製、公開陳列、與發行，或再授權他人以各種方法重製與利用，並得將數位化之上列論文與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。

研究生簽名：____曾韻榮_____

論文名稱：高速公路上之道路牌文字偵測

指導教授姓名：____范國清_____

系所：____資訊工程____所 ☐博士 ☒碩士班

學號：____93522033_____

日期：民國95年6月29日

備註：

1. 本授權書請填寫並親筆簽名後，裝訂於各紙本論文封面後之次頁（全文電子檔內之授權書簽名，可用電腦打字代替）。
2. 請加印一份單張之授權書，填寫並親筆簽名後，於辦理離校時交圖書館（以統一代轉寄給國家圖書館）。
3. 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，應依著作權法相關規定辦理。

Abstract

With the advancement of scientific technologies, the cost of digital camera decreases rapidly. It is a trend to improve and uplift the living quality of people using image processing techniques. Automatic detection of text from video is one of the applications which is an essential task for understanding and indexing of video. In this thesis, a driver assistant system is designed by automatic detection of text on road sign (guid signs and limit signs) on highway to provide drivers information for navigation, such as location, direction, and speed limit. It may also alleviate the load of driver who may lose his/her attention looking at road signs while focusing on driving. For 2008 Olympic in Beijing, there will be many foreigner visiting China and not all of them understand Chinese language. Hence, the translation of text on road sign is another goal that can be accomplished.

In this thesis, a set of feature is devised to detect road signs. The proposed system consists of three modules. The first module finds the constituting colors of road signs using the color transform model and locates road sign candidates. In the second module, affine transformation is performed to restore road signs which are captured by camera in different positions to let every road sign seems to be vertical to the camera optical axis. Moreover, affine transformation can improve the accuracy in detecting texts embedded in road signs. As to the third module, it performs the task of detecting texts on road signs. The method we adopt is canny edge detector to obtain clearer edge information.

Experiments were conducted on a variety of situations. 20 video sequences (sunny*10 and cloudy*10) including light variations and straight or cursive road conditions were tested to verify the validity of the proposed method. The recall and precision rates in locating road sign are 91.1% and 80.8%, respectively. The recall and precision rates in detection text are 93.6% and 88.0%.

摘要

自動地對影片中的文字做偵測，對於影片的檢索和瞭解是不可或缺的工作。而本論文主要是針對輔助駕駛系統做應用，利用自動對高速公路牌（指示牌、速限牌）上的文字做偵測，提供駕駛者在高速公路上的導航，例如：現在駕駛者所處的位置、方向、現在應該保持的速度。讓駕駛者可以專心地開車，而不會因為為了看道路上的路牌而造成駕駛不當，或者因為超速而造成車禍；且為了響應 2008 年北京奧運，在未來工作方面，除了對高速公路牌上的文字自動辨識之外，可以進階地將所攝影機擷取到的文字，自動地翻譯成各國語言，使各國來的人士能容易地了解各個路牌所代表的意思。

本論文提出一個對擷取道路牌上強而有效的新特徵，使用顏色的資訊來對高速公路上的道路牌做定位，並結合邊的資訊，自動偵測道路牌上的文字。主要的架構可分為三個，第一個步驟，利用顏色和不同物質具有不同傳導係數的特性，配合類神經網路的訓練，將路牌和其他的物質分開；第二個步驟，加入仿射（affine）矯正，將照相機在不適當的拍攝角度，所拍得的變形道路牌復原回來，使道路牌上的文字正對著相機，增加框選文字的正確性；第三個步驟，利用 Canny 邊緣偵測來取得邊的資訊，藉此在每一個道路牌上框出文字的候選區。

在實驗的部分，我們希望本論文提出的方法能適用在大多的情況下，所以從多段影片之中擷取出 20 段影片，其中包括了晴天、多雲和筆直的或彎曲的道路情況。在道路牌定位的部份，檢出率(recall)和精確率(precision)分別為 91.1%及 80.8%，在文字偵測上則是 93.6% 和 88.0%。

誌謝

本論文承蒙指導教授 范國清博士兩年來在學業上的悉心指導與督促，使論文得以順利完成，在此致上最誠摯的謝意。感謝口試委員黃興燦、王元凱、林啟芳與陳淑媛博士在口試時給予論文上的寶貴意見與指導，使本論文得以更臻完善。

在這裡也要感謝實驗室博士班的學長，碩士班學長、同學及學弟妹們的幫忙，尤其是呂信德及蔡洛緯學長在研究過程中給予珍貴的指導及建議，在日常生活中也經常給予鼓勵與協助，使我兩年的實驗室生涯過得如此的充實。

最後，感謝我的家人及女友，在這兩年的求學路上默默的給予我支持與鼓勵，讓我得以專心投入研究而無其他顧慮。

付梓在即，僅以最誠摯的心感謝所有愛護我、關心我的人，以此論文與他們一起分享。

目錄

Abstract.....	i
摘要.....	iii
附圖目錄.....	vii
表格目錄.....	ix
第一章 緒論.....	1
1.1 研究動機.....	1
1.2 相關研究.....	2
1.3 系統流程.....	6
1.4 論文架構.....	7
第二章 Karhunen-Loève 轉換.....	8
2.1 Karhunen-Loève 轉換.....	8
2.2 放射狀基底函數網路.....	12
2.2.1 網路架構.....	12
2.2.2 訓練神經網路.....	13
第三章 道路牌的定位.....	15
3.1 道路牌定位.....	18
3.2 候選區域的篩選.....	22
第四章 仿射校正.....	24
4.1 候選區域的分析.....	25
4.1.1 區域填充.....	26
4.1.2 參數的取得.....	28
4.2 轉換參數的估計與校正.....	31
第五章 文字的偵測.....	35
5.1 邊緣偵測之方法探討.....	36
5.2 Canny 邊緣偵測.....	39

第六章 實驗結果與討論	43
6.1 實驗資料.....	43
6.2 在各環境下之實驗結果	44
6.3 錯誤分析及討論.....	49
第七章 結論與未來工作	53
7.1 結論.....	53
7.2 未來工作.....	54
參考文獻.....	56

附圖目錄

圖 1.1 肇事原因分析圖	2
圖 1.2 (a)不同的字形、大小和位置；(b)因為玻璃遮蔽而造成的模糊； (c)水平和垂直傾斜；(d)光線影響	3
圖 1.3 系統流程圖	7
圖 2.1 (a)原本的三維樣本 (b)降低樣本維度到二維	9
圖 2.2 放射狀基底函數網路架構圖	13
圖 3.1 (a)邊緣偵測前 (b)邊緣偵測後	15
圖 3.2 道路牌定位流程圖	16
圖 3.3 (a)在德國高速公路上的藍色指示牌；(b)在台灣高速公路上的 綠色指示牌；(c)將(b)轉成 HIS 色彩空間後所偵測出的結果	17
圖 3.4 道路顏色的分佈和 u, v 特徵平面	19
圖 3.5 車輛偵測	20
圖 3.6 道路牌之 (a)訓練樣本；(b)非道路牌之訓練樣本；(c)將訓練樣 本同時投影在 u, v 平面上的結果	20
圖 3.7 道路牌候選區	21
圖 3.8 候選區域篩選之後的結果	23
圖 4.1 不與相機光軸垂直的道路牌	24
圖 4.2 (a)不與相機光軸垂直的道路牌, (b)仿射校正流程圖	25
圖 4.3 (a)集合 A ；(b) A 的補集；(c)結構元素 B ；(d)邊界內側的起始點 p ；(e)~(h)各為各步驟疊代的結果； (i)由(a)和(h)的聯集	27
圖 4.4 (a)區域填充前；(b)區域填充後	27
圖 4.5 (a)8 方向鍵的方向數；(b)將重新取樣網格重疊上的數位邊界；(c) 重新取樣的結果；(d)8 方向鍵碼	30
圖 4.6 (a)尋找最遠的點 v_f ；(b)將 v_i, v_j, v_f 變成 $v_i v_f$ 和 $v_f v_j$	30
圖 4.7 (a)候選區域分析前；(b)候選區域分析後(綠色部分為輪廓) ...	31

圖 4.8 傾斜道路牌之視意圖	32
圖 4.9 以簡單的圖形為例，其中包括了垂直傾斜、水平傾斜、垂直加上水平傾斜，多個物體的仿射校正。	33
圖 4.10 (a)實際地對道路牌做仿射校正之前；(b)實際地對道路牌做仿射校正之後	34
圖 5.1 文字偵測流程圖	35
圖 5.2 邊緣的定義	36
圖 5.3 Sobel Operator	37
圖 5.3 二維的 LOG 函數圖形	38
圖 5.4 邊緣偵測結果(a)原始影像；(b)Sobel；(c)Canny；(d)LOG	39
圖 5.5 道路牌上文字偵測結果	42
圖 6.1 Canon PowerShot A95.....	43
圖 6.2 在晴天的直線道路上 (a)原圖；(b)~(e)不同道路牌定位結果；(f)道路牌文字偵測後的結果	45
圖 6.3 在晴天的彎曲道路上 (a)原圖；(b)~(e)對同一道路牌做定位；(f)道路牌文字偵測的結果	46
圖 6.4 在多雲的直線道路上 (a)原圖；(b)~(e)對同一道路牌做定位的結果；(f)道路牌文字偵測的結果	47
圖 6.5 在多雲的彎曲道路上 (a)原圖；(b)~(e)對同一個道路牌做道路排定位的結果；(f)道路牌文字偵測的結果	48
圖 6.6 道路牌定位上之錯誤例子 (a)車子擋風玻璃所造成的影響模糊 (b)車子顏色所造成的誤判.....	51
圖 6.7 文字偵測上之錯誤例子 (a)車子擋風玻璃所造成的影響模糊 (b)道路牌上的箭頭也具有強烈的邊緣資訊	52
圖 7.1 對於同一個道路牌，邊緣資訊的比例(從遠到近) (a) 0.417 (b) 0.351 (c) 0.183.....	55

表格目錄

表 6.1 Canon PowerShot A95 之規格	44
表 6.2 道路牌定位及文字偵測之分析統計表	49
表 6.3 邊緣資訊在不同天氣下的比較	51

第一章

緒論

1.1 研究動機

隨著科技的日益進步，數位相機的價錢越來越便宜，所以我們對於如何利用影像處理方式來改善生活有很大的興趣。本論文主要是針對輔助駕駛系統做應用，自動地對高速公路上指示牌、速限牌、告示牌的文字做偵測，提供駕駛者在高速公路上的導航，例如：現在駕駛者所處的位置、方向、現在應該保持的速度；再則也可以結合其他的導航系統，如全球導航系統(GPS)，以降低駕駛人開車時精神上的負擔。發展智慧型運輸系統(Intelligent Transportation System, 簡稱 ITS) 其中一個重要原因就是想要減少交通事故的產生，由國家高速公路局的統計 [2] 可以知道，從民國六十四年到民國九十三年，「駕駛不當」皆佔有在肇事原因中很大的比例，因此我們可以了解到駕駛不當的嚴重性，如圖 1.1 所示。駕駛員開車不當造成的交通意外，主要是受前車變道、減速等影響，駕駛員盲目轉向，車輛失控撞擊護欄或防撞墩。所以若我們能輔助駕駛人行車，讓駕駛者可以專心地開車，減少駕駛者因為看道路上的路牌而造成駕駛不當的危險產生。

在 2008 年中國於北京將舉辦奧運，到時會有世界各地的人士到中國去，他們可能想自己使用汽車作為代步，並利用高速公路來到達目的地，但是可能因為語言的關係而看不懂道路牌上所代表的意思，使他們有所卻步；相同地，台灣是個美麗的寶島，常常吸引各世界各

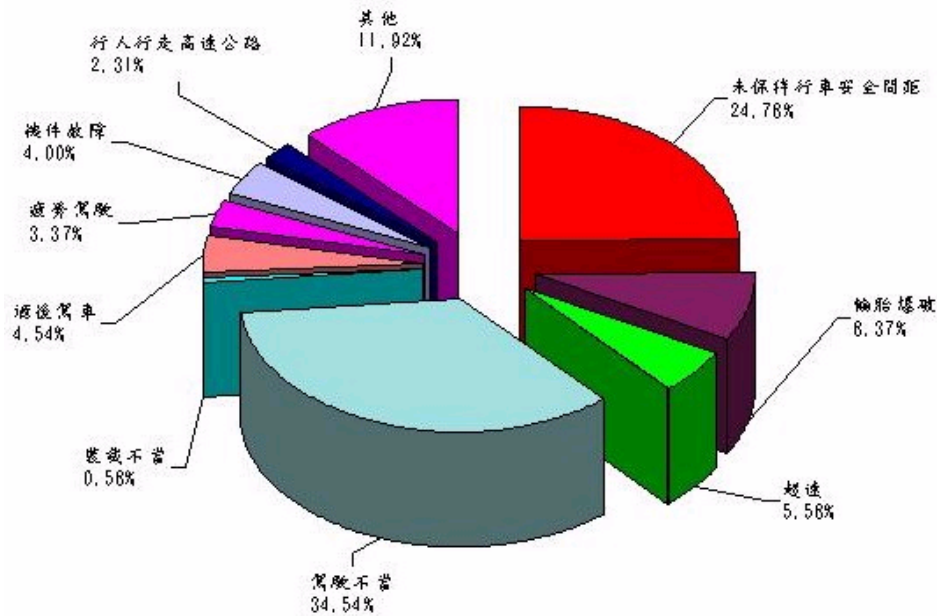


圖 1.1 肇事原因分析圖

(資料來源: http://www.freeway.gov.tw/11_92_05_02.asp)

地的人士來觀光旅遊，若因為語言的關係使他們沒辦法上高速公路的話，難免有缺憾；因此在未來工作方面，除了對高速公路牌上的文字做自動辨識之外，可以進階地將所擷取到的文字，自動地翻譯成各國語言，使各國來的人士能容易地了解各個路牌所代表的意義，幫助他們順利地到達目的地。

1.2 相關研究

文字對於瞭解影像和影片有相當程度的幫助，而如何將文字從影像和影片中擷取出來是許多學者專家所關切的，相關的研究也有許多[3]。而在影片中的文字上，大致上可以分為兩種：圖像中的文字(graphic text)、自然景象中的文字(scene text)。所謂圖像中的文字所指的是在影片被抓取後，才把文字放在上面，例如：新聞影



(a)



(b)



(c)



(d)

圖 1.2 (a)不同的字形、大小和位置；(b)因為玻璃遮蔽而造成的模糊；(c)水平和垂直傾斜；(d)光線影響

片、電影字幕；而自然景象中的文字，則是在影片抓取前就已經存在，它可能是某個物體的一個部份。在一般的假設下對於圖像中的文字，文字的平面是垂直於相機的光軸 [4], [5]；相對地，自然影像中的文字並不一定符合這個假設，所擷取到的文字平面對於相機的光軸來說，存在著各種角度，由這點我們知道，對於自然影像的文字偵測會困難許多，之前對於此問題的相關研究也比較少。

自然影像中的文字存在著許多挑戰。最主要的是在於多變化的

文字，它可以有多種的字型、大小、方向和位置，如圖 1.2(a) 所示。它也可能被其他的物體遮蔽、扭曲，或是因為移動而使它模糊，如圖 1.2(b) 所示。自然影像中的文字是存在於三度空間中，所以他也存在著垂直、水平等傾斜，如圖 1.2(c) 所示。它會受到背景複雜的影響，更會受到周圍光線強弱、陰影而有所影響，如圖 1.2(d) 所示。

對於擷取自然影像中文字的相關研究，例如：一般的招牌、鈔票、路牌等，都包括在其內。自動地對文字做偵測和辨識在應用上可以幫助到社會上較弱勢的族群，J. Pierre [6] 介紹以個人行動助理 (PDA) 為基礎的系統如何幫助視障者，利用了紋理的資訊將文字的特徵擷取出來，並用 K-means 做特徵上的分類，對書本、雜誌、鈔票上做文字偵測。N. Ezaki [7] 也利用一台個人行動助理再加上一台相機，搭配 morphology operation、Sobel 邊緣偵測、與邊相反的特徵、顏色等四種方式，對自然環境中招牌的文字做偵測，來幫助視障者了解自己所在的位置。

在道路牌定位方面，W. Wu [20] 運用了 HIS 的顏色模型配上高斯混合模型 (Gaussian mixture modal) 來擷取屬於道路牌的特徵，再假設由特徵所形成的平面會與相機的光軸垂直，來篩選所得到的道路牌後選區；但是在現實生活上，我們知道道路牌所形成的平面卻不會永遠地與相機的光軸互相垂直。G. Loy [21] 則是利用道路牌的形狀資訊來定位道路牌，將所得到的邊，用其角度和方向，來判斷道路牌的所在位置。也有些研究基於顏色的特徵易受光線的影響，所以是使用灰階或是二值化後影像裡找白色的部份，再利用圓形、方形的形狀特徵來定位出道路牌的位置 [22]。對於形狀的特徵，也可以用距離轉換 (distance transform)，搭配樣板比對 (template matching) 來做定位 [23]。C. Y. Fang [8] 介紹利用兩個類神經網路對一連串

的影像做顏色和邊的擷取和訓練，定出路牌候選區的位置，之後再利用卡曼濾波器做其做追蹤，但是沒有包括指示牌。

若想將照相機在不適當的拍攝角度，所拍得的變形路牌復原回來，使其文字好像永遠都正對著相機，就需要做仿射轉換（affine transformation）的計算。X. Chen [9] 在得到了文字之後，對於文字的分佈做了更進一步地分析（layout analysis），將其分成了兩大種類的特徵：內在（intrinsic）和外在（extrinsic）的特徵。內在的特徵包括了顏色、字型、對比，這些特徵不會受到相機拍攝的位置、角度而有所改變；而外在的特徵則包括了字型的大小、招牌的形狀，這些特徵是會受到相機位置和拍攝角度的影響。這些外在的特徵可以對仿射轉換的參數提供一些資訊，知道字的方向，進而求得文字所存在的平面。假設我們知道了在相機的座標軸下，文字所屬平面的法線，我們就可以推得兩者之間的轉換關係，配合內插法，將變形的路牌還原回來。

有面積分析和邊緣分析兩種方式能將影像中的文字偵測出來。所謂的面積分析，是指從一塊面積搜尋如紋理或顏色的特徵。離散餘弦轉換（Discrete cosine transform）和小波轉換（wavelet transform）常被用來做面積的分析[10],[15]。將離散餘弦轉換用在面積分析上最主要的優點的是其係數能直接從 JPEG 或 MPEG 的影像中擷取出來；而小波轉換則是能提供較多穩定的特徵，而它的缺點是在於易受到光線和大小改變的影響。另一方面，邊緣分析能在自然環境中的影像提供穩定的特徵，較不受到光線的影響 [11]。

1.3 系統流程

在系統架構上，本論文包含了三個模組，分別為道路牌偵測 (Locating of Road Signs)、仿射校正(Affine Rectification)、文字偵測(Text Detection)。模組間的關係如圖 1.3 所示。

- 道路牌偵測：利用特定的顏色轉換函式，將影像中的每一個像素 (pixel) 投影到新的特徵空間上，再利用類神經網路作為分類器，來分開道路牌以及非道路牌。
- 候選區域的驗證：利用道路的幾何限制以及邊緣資訊的含量，去除不可能為道路牌的候選區，節省後續所需要的處理時間，提升整體系統的效率。
- 仿射校正：傾斜的道路牌會造成偵測文字和辨識文字的正確率降低，如何對傾斜的道路牌做校正，也是我們所關切的問題。
- 文字偵測：我們已經將偵測文字的範圍縮小到每一個路牌候選區，只需要對候選區藉由 Canny Edge Detector 來得到邊的資訊，就可以將文字偵測出來，並藉由幾何限制來篩選文字候選區域。

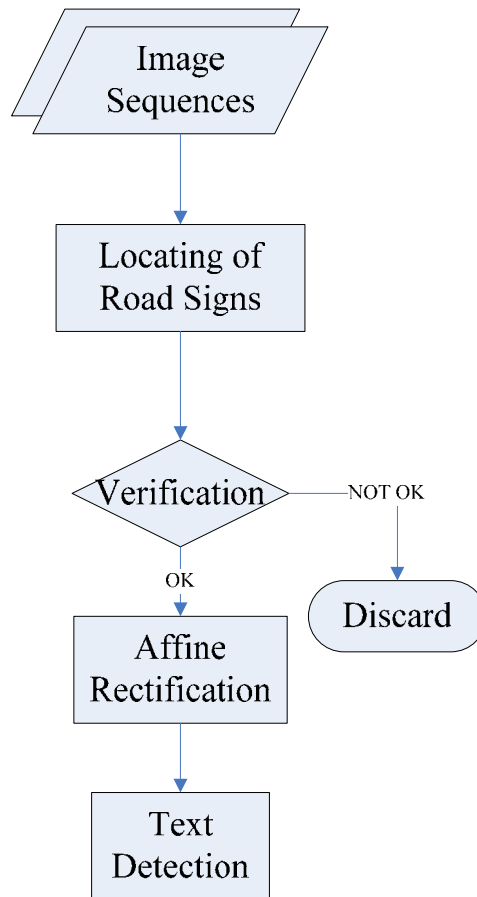


圖 1.3 系統流程圖

1.4 論文架構

本論文總共分為七章，第一章是敘述本篇論文的研究動機、過去的相關研究以及系統流程與論文架構。第二章介紹我們所用來擷取特徵和判斷特徵的方法。第三章詳細述說我們如何偵測道路牌。第四章會解釋我們如何只用一張影像來做仿射校正。第五章會對一般的邊緣偵測作一個比較，來說明為什麼會選擇用 Canny Edge Detector。第六章是實驗結果的部份，一開始會先對實驗資料和環境做些說明，然後會列出在不同天氣條件下的實驗成果，並且對錯誤的地方加以討論。最後一章則為本篇論文的結論和未來工作。

第二章

Karhunen-Loève 轉換

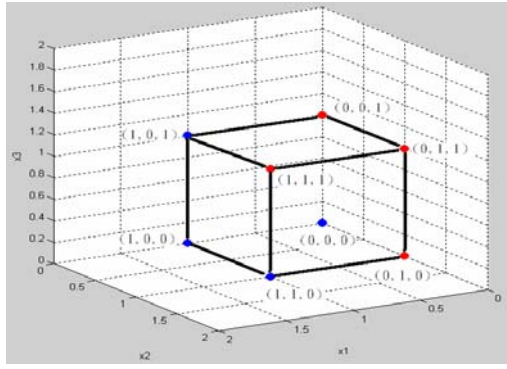
與

放射狀基底函數網路

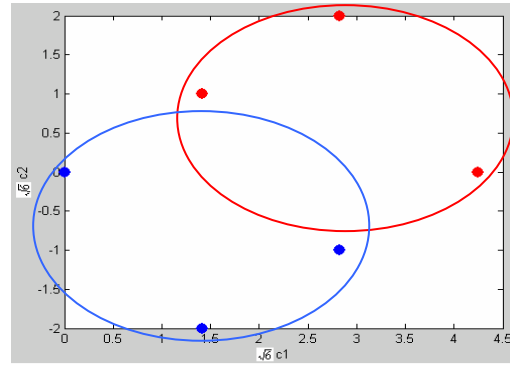
2.1 Karhunen-Loève 轉換

在做完影像的前處理後，緊接著我們要做的是特徵的擷取和辨識。特徵擷取的目的是在於尋找資料間重要的資訊，在不破壞資料原本的特性下，找到足以區分不同資料的特徵。而降維的動作常跟隨著特徵的擷取，用來減輕計算量。

人類不容易瞭解在高維度下資料的密集度和形狀，透過降低維度的動作，我們可以將高維度的資料轉成二維的資料，來幫助人類了解資料的密集度和形狀，請參考圖 2.1 [12]。Karhunen-Loève 展開式可以用來特徵化我們所針對的樣本，利用訓練樣本之間彼此的差異性來求得特徵空間 (eigenspace)，特徵空間是由許多獨立的特徵向量 (eigenvector) 所構成，它的線性組合可以完整地表現原始的訓練樣本，而每一個訓練樣本都是特徵空間中一個投影點。



(a)



(b)

圖 2.1 (a)原本的三維樣本 (b)降低樣本維度到二維

接下來，更詳細介紹如何用 K-L 展開式來建構特徵空間。 假設某一個 m 維的訓練樣本 P ，經由 $m \times m$ 維之 T 矩陣轉換後可得 $Q = (q_1, q_2, \dots, q_m)^T$ ，寫成

$$Q = TP \quad (2.1-1)$$

假設 X 為 T 之反矩陣，則

$$\begin{aligned} P &= XQ = q_1 X_1 + q_2 X_2 + \dots + q_m X_m \\ &= \sum_{j=1}^m q_j X_j \end{aligned} \quad (2.1-2)$$

此處 $X = (X_1, X_2, \dots, X_m)$ 。由 (2.1-2) 式 m 個中取出 n 個，以 P^* 來表示，我們將它改寫成

$$\begin{aligned} P &= XQ = q_1 X_1 + q_2 X_2 + \dots + q_n X_n \\ &= \sum_{j=1}^n q_j X_j \end{aligned} \quad (2.1-3)$$

P^* 可以看成是 P 之近似，其平方誤差 ε^2 為

$$\begin{aligned}
\varepsilon^2 &= \left(\sum_{j=n+1}^m q_j X_j \right)^T \left(\sum_{j=n+1}^m q_j X_j \right) \\
&= \sum_{j=n+1}^m q_j^2
\end{aligned} \tag{2.1-4}$$

另外，由 (2.1-2) 式我們可以得到

$$X_j^T P = X_j^T \sum_{i=1}^m q_i X_i = q_j \tag{2.1-5}$$

我們將 (2.1-5) 式代入 (2.1-4)，可以得到

$$\begin{aligned}
\varepsilon^2 &= \sum_{j=n+1}^m q_j^2 = \sum_{j=n+1}^m (X_j^T P)(X_j^T P)^T \\
&= \sum_{j=n+1}^m X_j^T P P^T X_j
\end{aligned} \tag{2.1-6}$$

(2.1-6)式是將單個 m 維的訓練樣本轉換成 n 維特徵空間時的平方誤差。現在將單各訓練樣本推廣為多個訓練樣本。假設有 N 個訓練樣本 $P_i (i=1, 2, \dots, N)$ ，則平方誤差的平均值 $\bar{\varepsilon}^2$ 表示為

$$\bar{\varepsilon}^2 = \sum_{j=n+1}^m X_j^T S X_j \tag{2.1-7}$$

$$S = \frac{1}{N} \sum_{i=1}^N P_i P_i^T \tag{2.1-8}$$

Karhunen 和 Loève 展開向量 X_j ，證明它是矩陣 S 的特徵向量，而且其特徵值為 λ_j ，使得下列式子成立

$$SX_j = \lambda_j X_j \quad (2.1-9)$$

(2.1-7)式可以重寫為

$$\bar{\varepsilon}^2 = \sum_{j=n+1}^m \lambda_j \quad (2.1-10)$$

依特徵值的大小來排列，也就是對特徵向量 X_j ($j = 1, 2, \dots, m$) 來說

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \quad (2.1-11)$$

所以使用 K-L 展開式將 m 維之樣本空間轉換到 n 維之特徵空間，選擇最大的前 m 個特徵值所對應的特徵向量可以使平均誤差的平均值為最小。藉由此方式，根據不同的應用，我們不用取全部的資訊去做辨識，只需要取用前幾個特徵向量就可以兼顧執行效率和正確性。我們藉由計算協方差矩陣 C 的特徵向量來得到轉換矩陣 H 。假設有一個輸入資料集 $X = \{x_1, x_2, \dots, x_n\}$ ，其轉換矩陣的計算如下列演算法：

步驟一：假設 m 代表平均值(mean)，

C 代表的是協方差矩陣(covariance matrix)，則

$$m = \frac{1}{n} \sum_{i=1}^n x_i$$

$$C = \frac{1}{n} \sum_{i=1}^n (x_i - m)(x_i - m)^T$$

步驟二：計算特徵值 d_1, d_2, \dots, d_n ，得到 C 中相對的特徵向量 λ ，並將它們依序排列 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$

步驟三：形成矩陣 $H = [\lambda_1, \lambda_2, \dots, \lambda_n]^T$

步驟四： $y_i = Hx_i$ for $i = 1, 2, \dots, N$

經過轉換之後，特徵的協方差矩陣變成一個對角矩陣(diagonal matrix)，這個矩陣會將輸入的資料投影到子空間中，其軸的方向具有最大的變異。

2.2 放射狀基底函數網路

類神經網路具有學習功能，對於監督式的學習網路，只要標記哪些是正樣本(positive samples)、哪些是負樣本(negative samples)；學習完之後，當有未知的樣本進來，學習網路就可以告訴我們，此樣本是否是我們要的[13]；此外，為了增加系統的效率，我們可以在離線下先將類神經網路訓練好，在線上運作時，就可以直接拿來使用。

2.2.1 網路架構

所謂的“放射狀基底函數網路(radial basis function network, 簡稱為 RBFN)”是一種層狀的類神經網路，網路的輸出可以表示成：

$$\begin{aligned} F(x) &= \sum_{j=1}^J w_j \phi_j(x) + \theta \\ &= \sum_{j=0}^J w_j \phi_j(x) \end{aligned} \quad (2.2-1)$$

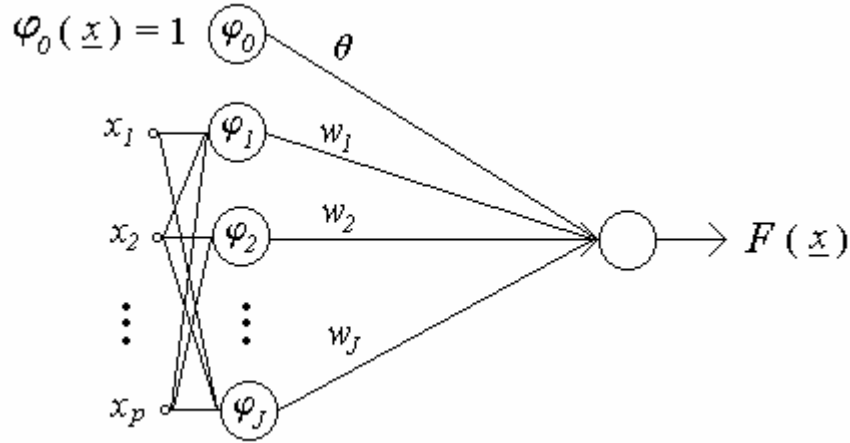


圖 2.2 放射狀基底函數網路架構圖

其中 w_j 代表第 j 個隱藏層類神經元到輸出類神經元的鍵結值， $\theta = w_0$ 代表可調整的偏移量， $\varphi_j(x)$ 代表計算第 j 個隱藏層類神經元輸出值的基底函數，如圖 2.4。而常見的基底函數有下列幾個：

1.
$$\varphi_j(\underline{x}) = \frac{1}{(\|\underline{x} - \underline{m}_j\|^2 + c^2)^{1/2}}$$
2.
$$\varphi_j(\underline{x}) = \exp\left(-\frac{\|\underline{x} - \underline{m}_j\|^2}{2\sigma_j^2}\right) \quad (2.2-2)$$
3.
$$\varphi_j(\underline{x}) = \exp\left(-\frac{1}{2}(\underline{x} - \underline{m}_j)^T \Sigma_j^{-1}(\underline{x} - \underline{m}_j)\right)$$

基本上，上述的基底函數的效能都不會差太多，在本論文中，我們是採用第二個基底函數。

2.2.2 訓練神經網路

假設訓練集中有 N 個輸入-輸出對 $(\underline{x}_1, y_1), (\underline{x}_2, y_2), \dots, (\underline{x}_N, y_N)$ ，用來訓練放射性基底函數網路，使其近似於函數，此種網路的每一層

類神經元，所執行的工作性質都不相同；隱藏層類神經元執行非線性的空間轉換，將輸入向量 $\underline{x}_i = (x_{i1}, \dots, x_{ip})^T$ 轉換成 $\underline{\varphi} = (\varphi_0(\underline{x}_i), \varphi_1(\underline{x}_i), \dots, \varphi_J(\underline{x}_i))^T$ ；而輸出類神經元執行的是，將此轉換後的向量作線性組合，亦即執行 $\sum_{j=0}^J w_j \varphi_j(\underline{x}_i)$ 的計算，因此網路的訓練策略可以分為以下兩種方式：(1)只調整 w_j 和 θ (2)全部參數一起調整。其中第一中方式可以分為 LMS 演繹法則以及虛擬反置矩陣法，但是不管是採用哪一種方法，網路訓練的結果是否能達到我們的要求，其實與放射性基底函數的相關參數之設定很有關係。然而，訓練效果最好的是將全部的參數一起做調整，其基本上就是往均方誤差函數的梯度方向調整，調整公式如下：

$$\underline{w}(n+1) = \underline{w}(n) + \eta(y_n - F(\underline{x}_n)) \cdot \underline{\varphi}(\underline{x}_n) \quad (2.2-3)$$

$$\underline{m}_j(n+1) = \underline{m}_j(n) + \eta(y_n - F(\underline{x}_n)) \cdot w_j(n) \cdot \varphi_j(\underline{x}_n) \cdot \frac{1}{\sigma_j^3} (\underline{x}_n - \underline{m}_j(n)) \quad (2.2-4)$$

$$\sigma_j(n+1) = \sigma_j(n) - \eta(y_n - F(\underline{x}_n)) w_j(n) \varphi_j(n) \frac{1}{\sigma_j^3} \|\underline{x}_n - \underline{m}_j(n)\|^2 \quad (2.2-5)$$

其中 $\underline{\varphi}(\underline{x}_n) = [\varphi_0(\underline{x}_n), \varphi_1(\underline{x}_n), \dots, \varphi_J(\underline{x}_n)]^T$ ， $\varphi_0(\underline{x}_n) = 1$ 以及 $\underline{w}(n) = [\theta(n), w_1(n), \dots, w_J(n)]^T$ 。放射性基底函數網路有快速學習的好處，但是缺點是隱藏層類神經元數目可能會很多，會需要較大的記憶體空間來儲存相關的參數值。

第三章

道路牌的定位

若想對整張影像偵測文字，通常會利用邊緣的資訊，因為邊緣的資訊不容易受到光源的影響。但是整張影像具有邊緣資訊的，卻不一定是文字，對於自然環境的影像更是如此，因為它可能是窗戶、隔音牆等，如圖 3.1。道路牌是設計給每一位駕駛者看的，所以被要求用在道路牌的顏色要對比鮮明、清晰可見；我們就利用此顏色的特點，欲先將道路牌定位，來縮小偵測文字的範圍，增加整體系統的效率。對於道路牌定位的流程圖，請見圖 3.2。

然而我們可以根據不同的應用來轉換到不同的色彩空間，如 XYZ、HIS 和 YIQ 等，來取出較合適的特徵。例如對於尋找人臉在整張影像裡的位置，我們可以利用找膚色區域來達到目的。在找膚色區域之前必須先定義什麼是膚色，在考慮因為環境光線變化對膚色的影響很大，通常會選擇能夠從色彩中分離出亮度因素的色彩空間，所以通常會將色彩空間轉換到 HSV 與 YCbCr 色彩空間中。



(a)



(b)

圖 3.1 (a)邊緣偵測前 (b)邊緣偵測後

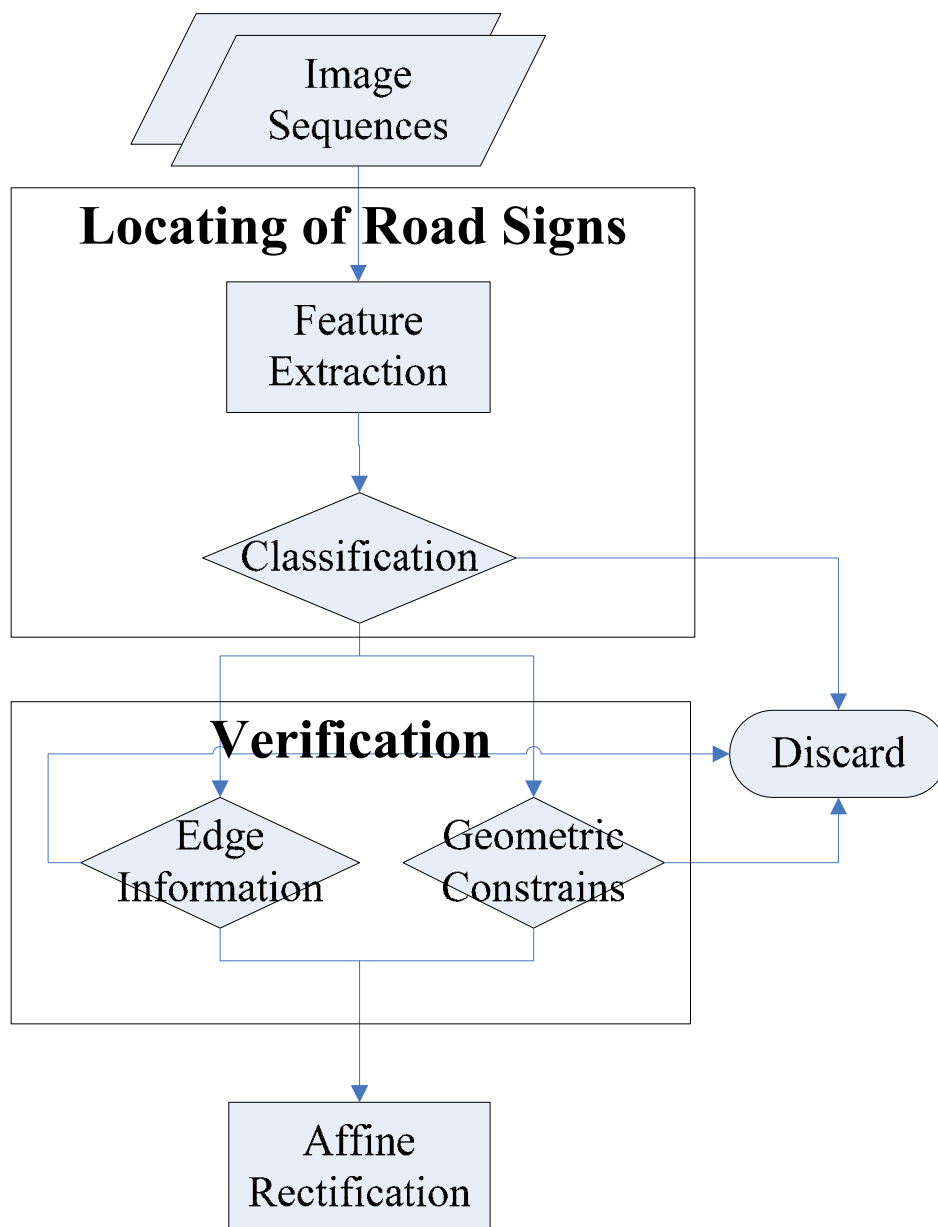


圖 3.2 道路牌定位流程圖



(a)



(b)



(c)

圖 3.3 (a)在德國高速公路上的藍色指示牌；(b)在台灣高速公路上的綠色指示牌；(c)將(b)轉成 HIS 色彩空間後所偵測出的結果

對於道路牌(指示牌、速限牌)，若環境中的雜訊不考慮，我們可以將色彩空間轉換到 HIS(色調、飽和度、強度)色彩空間，利用色調的範圍來偵測紅色和藍色的道路牌。但是這樣並不能符合我們所需要的，因為在一些國家中除了速限牌是紅色和藍色之外，高速公路上的指示牌卻是綠色的並非藍色的，使得在不考慮遮蔽、光線等條件下，環境中就存在很多的雜訊。在台灣的高速公路上，指示牌是綠色的，而高速公路的兩旁時常常都是充滿著山、草、樹木，它們也是綠色的，所以若是我們將色彩空間轉換到 HIS 色彩空間中，並不會得

到任何的好處，如圖 3.3。

3.1 道路牌定位

由上面的結果我們知道，將色彩空間轉換到 HIS 色彩空間中，直接地對高速公路的道路牌做定位並不會得到任何好處，所以在這一節之中，我們將詳細描述如何找到道路牌的顏色。我們從 N 張的道路牌和非道路牌影像中擷取出 R、G、B 的顏色分佈，進而得到相對應的協方差矩陣 Σ 。利用 Karhunen-Loève 轉換，我們得到協方差矩陣 Σ 的特徵值 λ_i 和特徵向量 e_i ($i=1, 2, 3$)。接下就可以定義三個新的顏色特徵如下：

$$C_i = e_i^r R + e_i^g G + e_i^b B \text{ for } i=1, 2, \text{ and } 3, \quad (3.1-1)$$

其中 $e_i = (e_i^r, e_i^g, e_i^b)$ 。Ohta *et al* [14] 介紹了由最大的特徵值所表示的特徵 C_1 ，它被用來做彩色到灰階之間的轉換，換句話說，

$$C_1 = \frac{1}{3}R + \frac{1}{3}G + \frac{1}{3}B \quad (3.1-2)$$

而另外兩個特徵 C_2 和 C_3 與 C_1 互相成直角，其表示如下：

$$C_2 = \frac{R - B}{2} \quad \text{和} \quad C_3 = \frac{2G - R - B}{4} \quad (3.1-3)$$

如此我們可以分別利用 $(1/3, 1/3, 1/3)$ 、 $(1/2, 0, -1/2)$ 和 $(-1/4, 1/2, -1/4)$ 三個向量將每一點像素 (R, G, B) 投影，來得到全部的顏色特徵。Healey [24] 指出同種非傳導性表面 (homogeneous dielectric surfaces) 的顏色，像是道路或雲，它們 R, G, B 的分佈都沿著 (3.1-2) 式中 $(1/3, 1/3, 1/3)$ 這條軸。這代表著同種非傳導性表面上的顏色除了色彩

飽和度的強度之外，並不會隨著時間而有所改變。相對於其他物質的表面，道路的表面更容易被模擬(model)。圖 3.4 道路在(R, G, B)色彩空間之中顏色的分佈，以紅點作表示。在[16]中，Rojas *et al.* 也發現道路的颜色會集中在 $(1/3, 1/3, 1/3)$ 軸周圍的圓柱上；因此，若將所有道路的颜色投影在垂直於 C_1 的平面上，可以讓它們集中在一個小小的圓裡面。只要物體的颜色分佈的特徵向量被找到，物體的颜色就會被投影在特徵空間中，形成一個特別的群集。在[17]，Luo-Wei Tsai 實際地將這項觀察應用在「偵測車輛」中，得到良好的效果，如圖 3.5。綜合以上研究，本篇論文提出的颜色模型(color model)，會將影像中所有的颜色像素轉換到二維的特徵空間中，使其在二維的特徵空間中之分佈範圍能縮小到某一區域，藉此達到道路牌和非道路牌分離之效果。

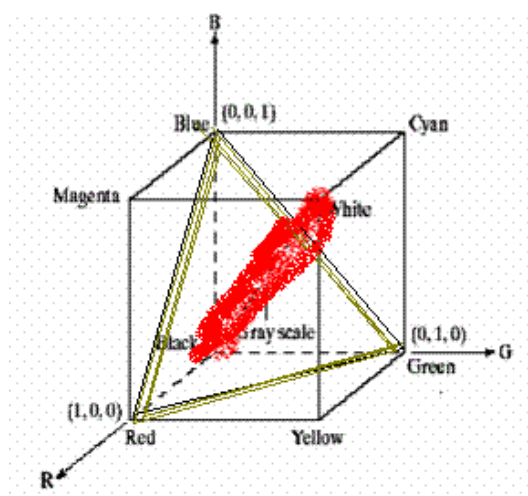


圖 3.4 道路顏色的分佈和 (u, v) 特徵平面

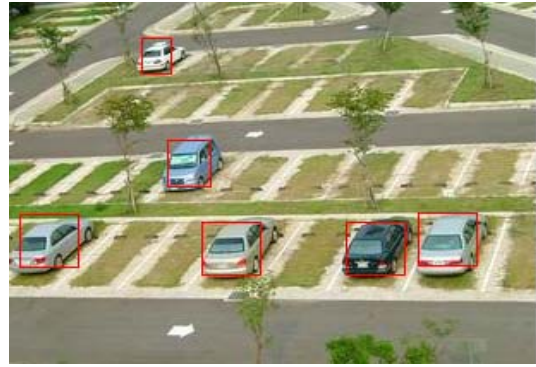
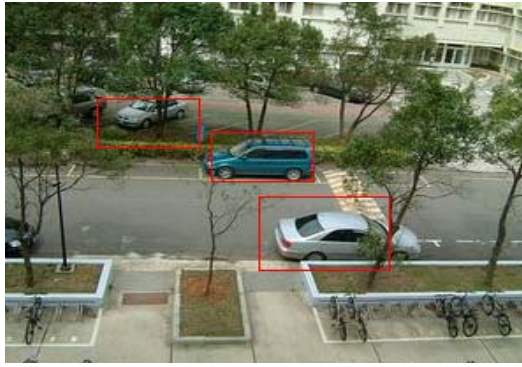


圖 3.5 車輛偵測



(a)



(b)



(c)

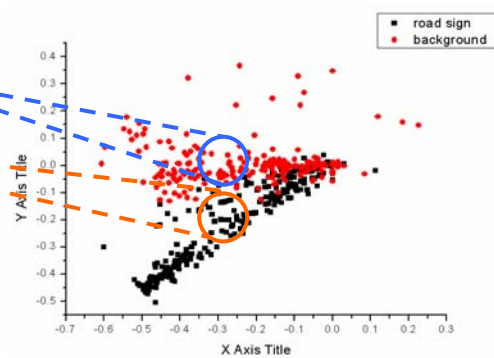


圖 3.6 道路牌之 (a)訓練樣本；(b)非道路牌之訓練樣本；(c)將訓練樣本同時投影在 u, v 平面上的結果



(a)



(b)

圖 3.7 道路牌候選區

以綠色道路牌為例，我們先將綠色的道路牌之訓練樣本，如圖 3.6(a)，利用 Karhunen-Loève 轉換得到三個特徵向量，其由大至小分別為： $(0.49038, -0.156904, -0.352716)$ ， $(0.122318, -0.489759, 0.387924)$ ， $(-0.339572, -0.339225, -0.321204)$ 。我們取出最大的兩個特徵來做為本論文之 (u, v) 顏色平面(color plane)，並且使其滿足與 $(1/3, 1/3, 1/3)$ 互相垂直(orthogonal)條件，因此我們可以得到下列式子：

$$u_p = \frac{2 \times (R_p - B_p)}{5}, \quad v_p = \frac{R_p + B_p - 2G_p}{5} \quad (3.1-3)$$

其中 p 代表的是像素(pixel)。訓練樣本轉換成 u, v 平面後的結果，請參考圖 3.6(c)。再將轉換後的結果作為輸入，來訓練類神經網路，在類神經方面我們是採用放射性基底函數，訓練後將背景濾掉，配合連接連通的方法來標記每一個相連區域，留下道路牌的候選區，請見圖 3.7。

3.2 候選區域的篩選

由 3.1 節所得到的道路牌候選區並不一定全是我們要的，如圖 3.7 所示，少量的樹葉或是車燈等，會造成框選到不正確的候選區。所以我們利用簡單的幾何判斷式作為過濾。首先，先假設候選區的長和寬分別為 H 及 W ，而 E 為候選區內所含有的邊緣資訊的總數。依據此假設，對於影像大小 640×480 ，不滿足下列條件的區域者則會將其濾除：

長寬比例不合： $H \times 5.0 \leq W \text{ or } W \times 5.0 \leq H$

區域過小： $W < 10 \text{ or } H < 10$

邊緣資訊過少： $E / (H \times W) < 0.02$ (3.2-1)

第一個條件是因為每一種道路牌的長寬比例是固定的，在考慮傾斜的情況下，根據我們所能取得的實驗資料做分析，所以訂定長寬比限制在 5.0 以內，若超過此比例，則將其濾除。第二個條件是為了濾除掉樹葉或是車燈等所造成的少量雜訊。最後一個條件是藉由我們所求得的道路牌內文字來取得資訊，文字能產生邊緣資訊的特性，而道路牌內也含有大量文字，所以若候選區內所含有的邊緣資訊不足，則表示此候選區並不含文字或是因為候選區距離相機太遠的機

會很大，所以我們也將其濾除。候選區經過篩選的結果請參考圖 3.8。



(a)



(b)

圖 3.8 候選區域篩選之後的結果

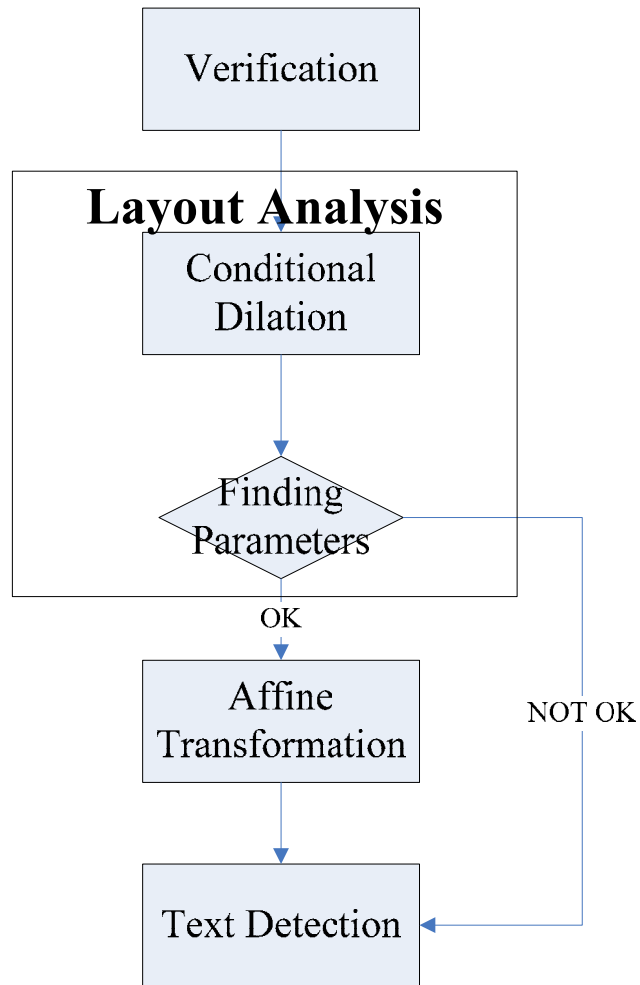
第四章

仿射校正

我們將照相機置於車的右前座並往前拍攝，所以車子的行駛方向是與相機的光軸平行。而在高速公路上，道路並非都是直線的，道路牌可能存在於高速公路的轉彎處，以致於所拍攝的影像中，所擷取出的道路牌並非都是正對相機，換句話說，道路牌的正面不一定與相機的光軸垂直；再加上道路兩旁的速限標誌，其正面也並非正對著車子的行駛方向，所以擷取下來的道路牌會有仿射(affine)校正的問題。仿射校正的問題會增加之後在擷取文字和辨識文字時的困難度，所以在這裡做仿射校正的目的，是在於讓所擷取出來的道路牌能永遠與相機的光軸垂直。然而，我們對於四方形的道路牌，可以利用道路牌的四個端點，來估計仿射校正所需要的參數，但是對於圓形的速限牌，目前無法得到相關的四個端點，所以在這裡我們所著重的部份，是對於四方形的道路牌，如圖 4.1；仿射校正流程圖請見 4.2。



圖 4.1 不與相機光軸垂直的道路牌



(b)

圖 4.2 (a)不與相機光軸垂直的道路牌, (b)仿射校正流程圖

4.1 候選區域的分析

一道路牌上的文字全部是屬於同一平面，候選區域的分析是希望利用此道路牌的特性，來得到仿射轉換所需要的參數。此分析的步驟可以分為兩步：先進行區域填充得到填滿後的道路牌區塊，再利用尋找輪廓的方式來得到我們仿射校正所需要的四個頂點。

4.1.1 區域填充

以圖 4.3 為例， A 代表一個集合，它含有一個子集合，該子集合是經過放射狀基底函數之後所產生的結果，從邊界內側任選一點 p 就可以開始區域填充的程序，其方法如下：

$$X_k = (X_{k-1} \oplus B) \cap A^c \quad k=1,2,3,\dots \quad (4.1-1)$$

其中 $X_0 = p$ ，而 B 集合如圖 4.3(c)。第 k 次疊代時 X_{k-1} 會被 B 集合向 4 聯通(4-connected)方向膨脹，不過因為有跟 A^c 取交集，所以填充不至於擴大到整個區域。如果不再變化則演算法終止($X_k = X_{k-1}$)，則此時 A 的區域填充 $R(A)$ 為

$$R(A) = X_k \cup A \quad (4.1-2)$$

此演算法可以用於任一有限數目的子集合，只要對每個子集合分別求出邊界與給定邊界內側的一點，就可以用遞迴方式的區域填充完成每一個子集合。雖然用遞迴的方式來進行區域填充其原理和作法都很簡單，但是由於多次的遞迴會造成耗時、耗記憶體，效率並不高。所以為了減少遞迴方式所產生的負擔，我們在這裡採用的是區域填充掃描線算法。其基本的流程如下：當給定開始點 (x,y) 時，首先填充位於給定區域的一個區段開始點所在的掃描線，然後確定位於給定區域內的區段與這一區段相連通的上、下兩條掃描線，並依次保存下來。反覆此過程，直到填充結束。可以由下列四個步驟完成：

- (1) 初始化：將堆疊設為空。將開始點 (x,y) 丟入堆疊內。
- (2) 取堆疊上的一元素：如果堆疊為空的，則結束掃描。否則取出堆疊裡最上面的元素 (x,y) ，以 y 當作現在要做的掃描線。

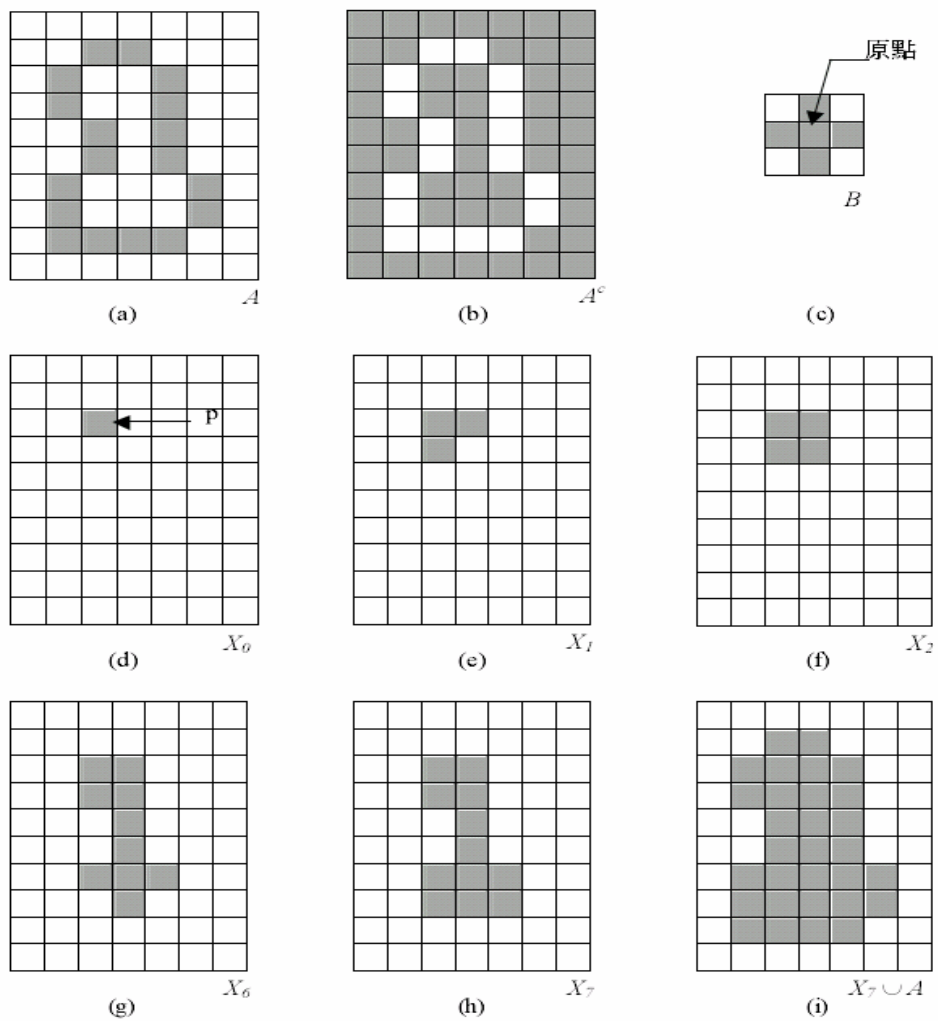


圖 4.3 (a)集合 A ; (b) A 的補集 ; (c)結構元素 B ; (d)邊界內側的起始點 p ; (e)~(h)各為各步驟疊代的結果 ; (i)由(a)和(h)的聯集



(a)



(b)

圖 4.4 (a)區域填充前 ; (b)區域填充後

- (3) 做填充並確定開始點所在的區段：由開始點 (x, y) 出發，沿著現在要處理的掃描線方向，向左、右兩個方向進行填充，直到邊界。分別記錄區段的左、右端點 x_{left} 和 x_{right} 。
- (4) 確定新的開始點：在區段 x_{left} 和 x_{right} 之中檢查與現在掃描線 y 上、下相鄰的兩條掃描線的像素。如果存在非邊界，未填充的像素，則把每一區間的最右邊的像素作為開始點丟入堆疊之中，返回第二步。

對於掃描線填充法，如上述可以知道對於每一個等待填充的區段，只需要丟入到堆疊一次；而在遞迴的方式之中，每一個像素都需要丟入到堆疊之中。因此，掃描線填充法能改善區域填充的效率。其結果如圖 4.4。

4.1.2 參數的取得

在尋找輪廓之前，我們會先用 Canny Edge Detector 尋找邊緣的資訊，再做侵蝕(Erosion)與膨脹(Dilation)來將雜訊清除掉。Canny Edge Detector 的部份，我們於下一章會有更詳盡的說明；而對於侵蝕與膨脹的方法如下：

- (1) 侵蝕：

$$A \ominus B = \bigcap_{b \in B} (A)_{-b} \quad (4.1-3)$$

意思是將 A 集合向內平移 B 集合中每一個元素的距離，然後取出這些集合的交集，所以侵蝕會使 A 集合縮小，且縮小的範圍是由 B 集合之結構所決定。

(2) 膨脹：

$$A \oplus B = \bigcap_{b \in B} (A)_b \quad (4.1-4)$$

意思是將 A 集合向外平移 B 集合中每一個元素的距離，然後把這些集合聯集起來，所以膨脹會使 A 集合擴大，且擴大的範圍由 B 集合之結構所決定。

至於尋找輪廓的方式，其逼近的方法我們是採用鏈碼(chain code)來壓縮水平、垂直、對角的分割，換句話說，就是只留下末端的像素點位置，並且將所有的輪廓存在一個串列中。其簡易的步驟如下：

(1)藉由較大的格子(grid)來對於邊界重新取樣(Re-sample)。

(2)將邊界的點分配給格子的頂點，用來逼近原始邊界。

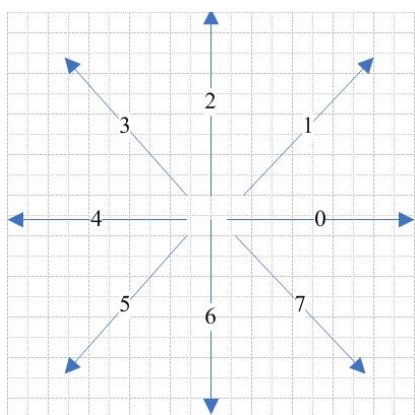
(3)利用 8-code 來連接相鄰的邊界點。

請參考圖 4.5 [18]。在尋找輪廓之後，利用 Douglas-Peucker Algorithm，來減少代表的點數，演算法如下：

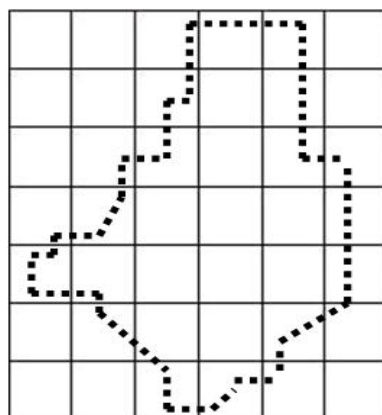
(1) 尋找與 v_i, v_j 距離 d 最遠的點 v_f ，如圖 4.6(a)。

(2) 如果 $d < \varepsilon$ ，則使得 v_i, v_j 連線 $\overline{v_i v_j}$ ；否則將 v_i, v_j, v_f 變成 $\overline{v_i v_f}$ 和 $\overline{v_f v_j}$ ，如圖 4.6(b)。回到第一步。

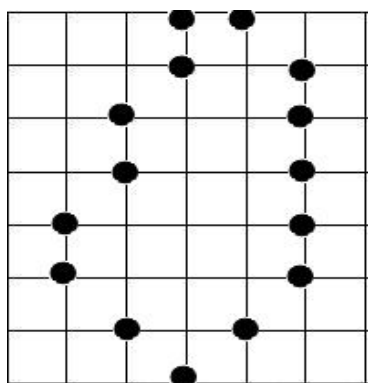
因為我們的目的是找四方形的道路牌，進而對它們做仿射轉換，所以最後留下只具有四個頂點的輪廓，得到仿射轉換所要的四個頂點，如圖 4.7。如果沒有得到四個頂點的候選區，則不做仿射轉換。



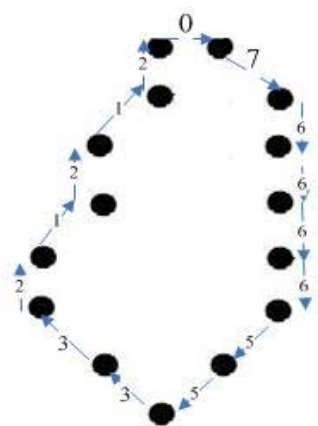
(a)



(b)

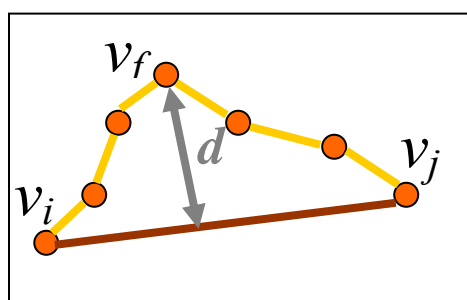


(c)

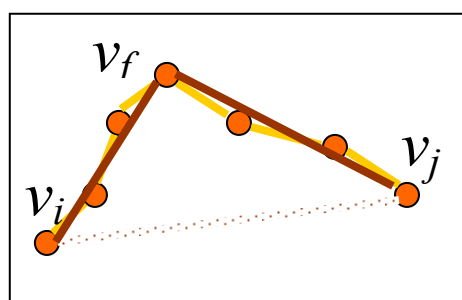


(d)

圖 4.5 (a)8 方向鍵的方向數;(b)將重新取樣網格重疊上的數位邊界;
(c)重新取樣的結果;(d)8 方向鍵碼



(a)



(b)

圖 4.6 (a)尋找最遠的點 v_f ;(b)將 v_i, v_j, v_f 變成 $\overline{v_i v_f}$ 和 $\overline{v_f v_j}$



(a)



(b)

圖 4.7 (a)候選區域分析前；(b)候選區域分析後(綠色部分為輪廓)

4.2 轉換參數的估計與校正

自然環境的道路牌是存在於三度空間中，我們所擷取出來的道路牌不一定只含有水平傾斜，可能還會具有垂直傾斜。所以我們欲用仿射轉換矩陣來解決這個問題。藉由上節所得到仿射轉換所需要參考的四個頂點之後，我們開始估計轉換後所要對應的四個頂點座標，藉由前後共八個點來尋找仿射校正的轉換矩陣。假設我們由上節方法所得到的四個頂點為 $(x_i, y_i), i = 0, 1, 2, 3$ ；而所要估計的四個頂點為 $(x'_i, y'_i), i = 0, 1, 2, 3$ ，其中頂點排列的順序為順時針，如圖 4.8。則估計的步驟如下：比較候選區左右兩邊的 y 座標的差異值，取較大的一邊的上頂點為基準點，利用基準點來訂定其他三個頂點的座標。取 (x_0, y_0) 為基準點，則其他三個頂點的座標為 $(x_1, y_1) = (x_1, y_0)$ ， $(x_2, y_2) = (x_1, y_3)$ ， $(x_3, y_3) = (x_0, y_3)$ 。而仿射轉換的公式如下

$$\begin{pmatrix} t_i x'_i & t_i y'_i & t_i \end{pmatrix}^T = M^* \begin{pmatrix} x_i & y_i & 1 \end{pmatrix}^T, \quad \text{其中} \quad (4.2-1)$$

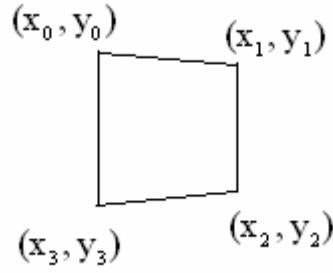


圖 4.8 傾斜道路牌之視意圖

$$M = \begin{bmatrix} m0 & m1 & m2 \\ m3 & m4 & m5 \\ m6 & m7 & m8 \end{bmatrix}$$

(x_i, y_i) 為原圖的頂點， (x'_i, y'_i) 為校正後的頂點， $i=0,1,2,3$ ，而 M 為一個 3×3 的轉換矩陣，且 $m8=1$ 。以圖 4.7(b) 其中一個道路牌為例，經由候選區域的分析，我們可以得到道路牌的四個頂點座標，以左上角頂點為起點，其四個頂點座標以順時鐘方向分別為(515, 311)、(601, 317)、(600, 136)、(507, 109)，假設經過變形校正後，四個頂點的座標分別為 (515, 311)、(601, 311)、(601, 109)、(515, 109)，利用(4.2-1)式推得關係式，解開關係式我們就可以得到， $t_0 \approx 0.60$ ， $t_1 \approx 0.54$ ， $t_2 \approx 0.58$ ， $t_3 \approx 0.65$ ，而所得到轉換矩陣 M 的參數分別為(由左至右，由上到下)：0.21089、(-0.13089)、241.19556、(-0.24692)、0.58404、132.1902、(-0.00065)、(-0.00021)、1.00000。配合內插法，其實作的結果，請見圖 4.9 及圖 4.10。

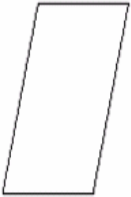



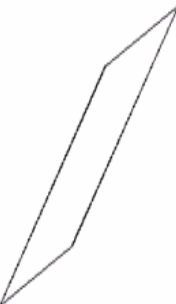
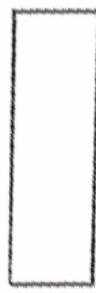

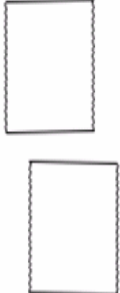
名稱	Before Affine Transform	After Affine Transform
水平 15 度		
垂直 -30 度		
水平 45 度 + 垂直 30 度		
兩平行		

圖 4.9 以簡單的圖形為例，其中包括了垂直傾斜、水平傾斜、垂直加上水平傾斜，多個物體的仿射校正。



(a)

(b)

圖 4.10 (a)實際地對道路牌做仿射校正之前；(b)實際地對道路牌做仿射校正之後

第五章

文字的偵測

若把文字當作是前景，則道路牌不含文字的部份可視為均勻的背景。此時我們可以將文字當作是特殊的紋理，利用紋理的特性來擷取出道路牌上文字；但是道路牌是存在於自然環境中，在不同的地方不一定具有相同的光線強度，所以光線的問題必須考慮。在本論文中，我們對於文字偵測所採取的方式是利用邊緣的資訊，邊緣的資訊不易受光線的影響，所以符合我們對於偵測道路牌上文字上的需要。文字偵測的流程圖請見圖 5.1。

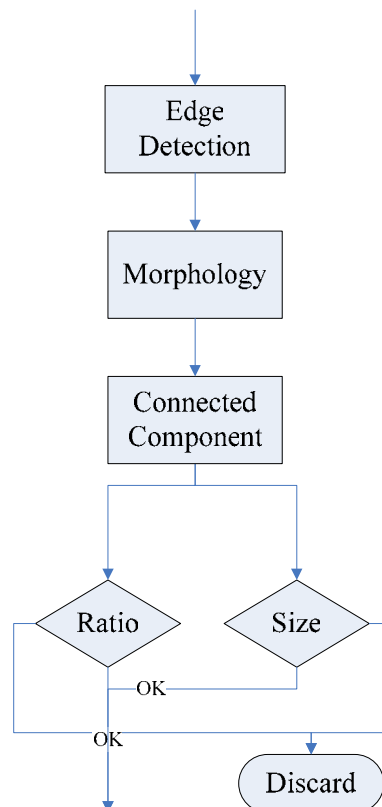


圖 5.1 文字偵測流程圖

5.1 邊緣偵測之方法探討

關於“邊緣”的定義，它存在於兩個具有不同灰階性質的區域之間的邊界，請參考圖 5.2[18]。在影像處理方面，要萃取出邊緣資訊的方法有很多，主要可分為一階導數運算式和二階導數運算式。常見的方式中，一階導數運算式的有索貝爾邊緣偵測（Sobel Edge Detection）、卡尼邊緣偵測（Canny Edge Detection），主要是用局部的極值來找邊緣區域；而二階導數運算式的有高斯拉普拉斯邊緣偵測（Laplacian of Gaussian Edge Detection，簡稱 LOG），則是利用過零點（zero crossing）的特性來估計邊緣區域。下面我們會對上面三種常見的邊緣萃取方法做個說明，並從中選擇出適合我們的方法。

(1.) Sobel 邊緣偵測

邊緣偵測可以看成高通濾波器，利用微分的特性，濾除低頻的訊

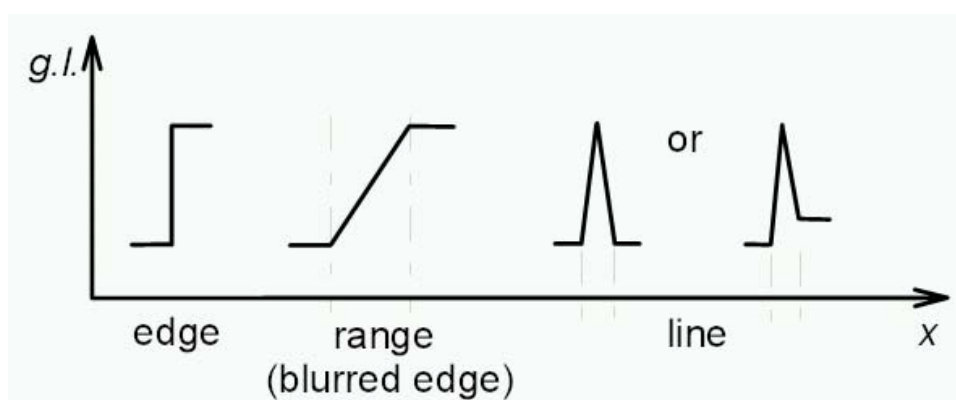


圖 5.2 邊緣的定義

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

圖 5.3 Sobel Operator

號，保留了變化劇烈的高頻資訊。Sobel 邊緣偵測是利用水平、垂直和對角線的遮罩（mask）對影像做摺積（convolution），藉此來得到梯度（gradient）分量，萃取出我們要的邊緣；相對於Prewitt運作元，用 Sobel 運作元所得到的影像較為平滑，減少了雜訊的影響。圖 5.3 為三種 3×3 的 Sobel Operator。

(2.) Canny 邊緣偵測

此方法由作者 Canny 所提出[19]，是最佳邊緣偵測之一，它能把雜訊所產生的假邊界點有效地消除。雖然整個計算過程較其他方法複雜，但是所產生的結果也比許多方法更貼近使用者的需求。以下是 Canny 邊緣偵測所奉的三個基準：

- 良好的偵測（good detection）：將所有具有灰階變化的邊線找出，且不具有假邊界點存在。
- 良好的定位（good location）：使得被偵測出的邊界位置能夠貼近真實的邊界位置，距離越小表示定位越準確。
- 單一響應的限制（single response constraint）：為了解決邊界受到雜訊的影響而有的多響應，Canny 只允許有單一響應，來找到更準確的邊界。

(3.) LOG 邊緣偵測

拉普拉斯運作元為二階導數運算式，相較於一階導數運算式，利用過零點的特性來估計邊緣區域的拉普拉斯運作元可以得到較強的邊緣響應，但是對於雜訊會比較敏感，所以在使用上通常會配合低通濾波來運作。LOG 邊緣偵測及結合了拉普拉斯運作元和高斯 (Gaussian) 濾波器，其結合後的公式如下：

$$LoG(x,y) = \frac{x^2 + y^2 - 2\sigma^2}{\sigma^4} e^{-(x^2+y^2)/2\sigma^2} \quad (5.1-1)$$

其中 σ 為標準差；而二維的 LOG 曲線圖請見圖 5.3。圖 5.4 為 Sobel、Canny、LOG 邊緣偵測後的例子，其中梯度門檻值 (threshold) 是用 Otsu[1] 的方法自動地作決定。

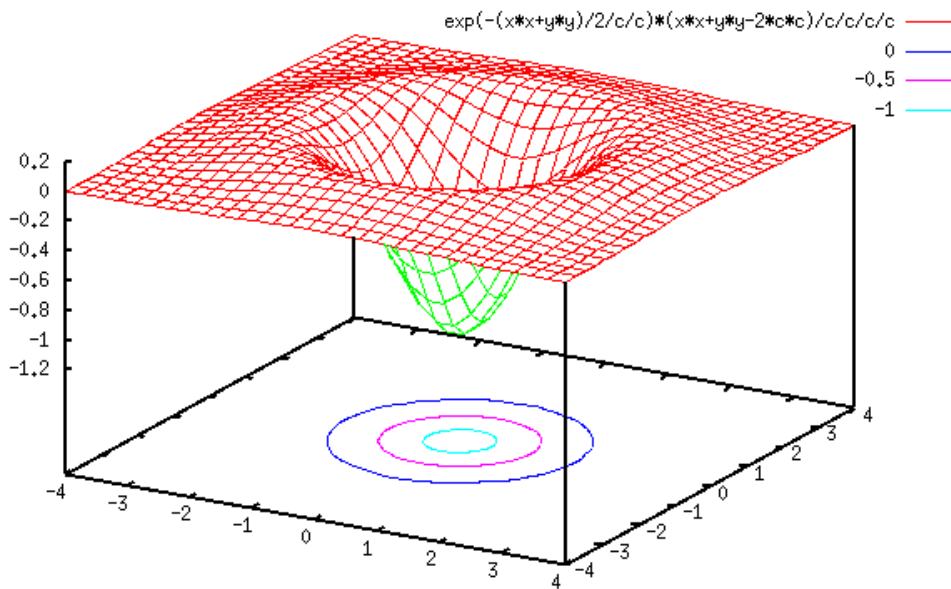


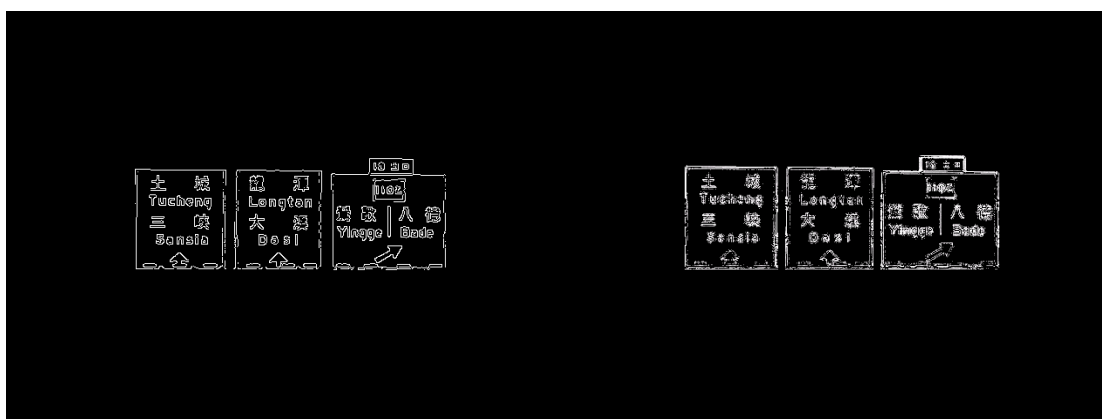
圖 5.3 二維的 LOG 函數圖形

(資料來源: <http://fourier.eng.hmc.edu/e161/lectures/gradient/node10.html>)



(a)

(b)



(c)

(d)

圖 5.4 邊緣偵測結果(a)原始影像；(b)Sobel；(c)Canny；(d)LOG

5.2 Canny 邊緣偵測

由上一節的實作結果我們可以清楚地看到，Sobel 運算元和 LOG 運算元都具有不錯的邊緣萃取能力，但是仔細地看影像被處理後的結果，我們會發現所得到的邊緣資訊仍然相當的粗略。雖然我們可以利用文字具有大量邊緣資訊的特點來偵測文字，但是若萃取出來的資訊含有許多無意義的邊緣，配合以相連成份為基礎的方法，將會使得結果不如我們所能預期的。綜合以上對於萃取邊緣資訊方法的說明

與討論，本論文採用的方法為 Canny 邊緣偵測，因此下面我們再進一步說明它的演算法。

其方法大致分為四個步驟：高斯模糊、計算梯度值與角度、非最大值抑制法（non-maximum suppression）以及遲滯性界定（Hysteresis Thresholding），演算法如下：

(1.) 使用高斯模糊來消除影像中的雜訊，一維高斯函數如下：

$$G(x) = e^{-(x^2 / 2\sigma^2)} \quad (5.2-1)$$

(2.) 利用一階導數運算元來得到梯度影像，並紀錄各像素所對應之梯度值和梯度方向，而公式(5.2-1)的一階導數為：

$$G'(x) = -\frac{x}{\sigma^2} e^{-(x^2 / 2\sigma^2)} \quad (5.2-2)$$

雖然可以用二維的方式做處理來取得梯度值，但是由於二維的摺積運算量大，效率並不好，所以在實作時分為 x 和 y 兩個分量做處理。

(3.) 在梯度影像中，對於每一個被找到的邊界點作梯度方向的搜尋，與其方向上所有的像素點作比較，如果某像素為該梯度方向上所有梯度值之中的最大值，則將這個像素點標記為邊界的候選點，否則標記為非邊界點，此方法被稱為非最大值抑制法，整張影像經過此處理，會得到一個邊緣細化的影像。

(4.) 接下來，為了補救邊緣斷裂問題和刪除不正確的邊界點，Canny 給定了兩個門檻值 T_H 和 T_L (其中 $T_H > T_L$) 來增加整個邊緣偵測

上的彈性。對於由第三步驟所得到之各邊界點候選點，若梯度值大於 T_H 則判定邊界點；對於梯度值介於 T_H 和 T_L 之間的點，若八個鄰域中包括了一個以上的邊界點，就會把此點補救回來，但是若梯度值小於 T_L 我們就確定此點為非邊界點。

由邊緣偵測處理後所產生的影像為一張二值化的影像，對於所得到的影像做侵蝕和膨脹，來解決像中文字裡常發生同一個字被分成多個區塊的情形；再配合連接連通（Connected Components）的方法來標記每一個相連的區塊，藉此來得到文字區塊。假設文字候選區的長和寬分別為 H 及 W ，若加上下列兩個限制條件：

長寬比例不合： $H \times 8.0 \leq W$ or $W \times 8.0 \leq H$

區域過小或過大： $H \times W < 30$ or $H \times W > 2500$ (5.2-3)

我們可以得到更正確的文字區塊。實作的結果請參考圖 5.5。



(a)



(b)

圖 5.5 道路牌上文字偵測結果

第六章

實驗結果與討論

6.1 實驗資料

我們透過在高速公路上所拍得的多段影片來對整個系統做評估。在拍攝的相機方面，採用 Canon PowerShot A95（圖 6.1），其詳細的規格請見表 6.1。將相機放置於座車的右前座，拍攝的角度為道路和天空各為一半。因為我們希望本論文提出的方法能適用在大多的情況下，所以從多段影片中擷取出 20 段影片來做實驗，其中包括不同的道路情況(直的、彎的)和不同的天候條件(晴天、多雲)。在這 20 段影片之中總共含有 63 個道路牌(指示牌、告示牌和速限牌)，取出每一個道路牌最近的連續 5 張頁框來做測試，總共測試了 315 個道路牌；在偵測文字的部分，總共含有 5182 個字，這些字裡包括了中文字、英文字(例如:A, B, ..., a, ...)以及數字(例如:0, 1, 2, ...)。系統上，我們是用 C++來實作，測試在 3.0-GHz Pentium IV 的個人電腦上。



圖 6.1 Canon PowerShot A95

表 6.1 Canon PowerShot A95 之規格

感光元件總像素	530 萬
感光元件有效像素	500 萬
感光元件尺寸	1/1.8 英寸 (CCD 感光元件)
鏡頭性能	f=7.8-23.4mm；5 組 6 片，2 片非球面
光學變焦	3 倍
最高解析度	2592×1944

6.2 在各環境下之實驗結果

下面將對做完道路牌定位、仿射校正和文字偵測結果用四段影像序列作呈現。圖 6.2 為在晴天的直線道路上，對於右邊的道路牌做此三各步驟之後的結果。圖 6.3 為在晴天的彎曲道路上，對在道路上方的道路牌做定位、仿射校正和文字偵測。圖 6.4 為在多雲的天氣下，對直線道路右邊的道路牌做此三各步驟之後的結果。圖 6.5 則是在多雲的天氣下，對道路上面的道路牌做此三各步驟之後的結果。對於道路牌偵測，影像大小為 320×240，系統處理速度上可以達到約 20 張/秒，而對於影像大小為 640×480，則是 4 ~ 5 張/秒；若再加上仿射校正及文字的偵測，對於影像大小為 640×480，處理速度為 1 ~ 2 張/秒。表 6.2 為針對 20 段影片統計分析後的結果。評估的標準上，我們使用大多數研究所用的檢出率(recall)及精確率(precision)，其公式如下

$$\text{檢出率: } \frac{\text{correct}}{\text{correct} + \text{missed}} \quad (6.2-1)$$

$$\text{精確率: } \frac{\text{correct}}{\text{correct} + \text{false}} \quad (6.2-2)$$



(a)



(b)



(c)



(d)



(e)



(f)

圖 6.2 在晴天的直線道路上 (a)原圖;(b)~(e)不同道路牌定位結果;
(f)道路牌文字偵測後的結果



(a)



(b)



(c)



(d)



(e)



(f)

圖 6.3 在晴天的彎曲道路上 (a)原圖;(b)~(e)對同一道路牌做定位;
(f)道路牌文字偵測的結果



(a)



(b)



(c)



(d)



(e)



(f)

圖 6.4 在多雲的直線道路上 (a)原圖；(b)~(e)對同一道路牌做定位的結果；(f)道路牌文字偵測的結果



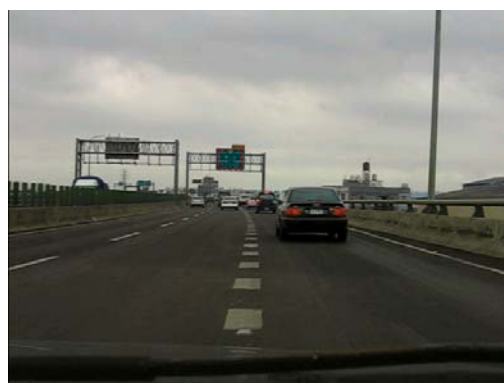
(a)



(b)



(c)



(d)



(e)



(f)

圖 6.5 在多雲的彎曲道路上 (a)原圖；(b)~(e)對同一個道路牌做道路排定位的結果；(f)道路牌文字偵測的結果

表 6.2 道路牌定位及文字偵測之分析統計表

	Numbers	False Negative
Sign Location	315	28
Text Detection	5182	332

False Positive	Correct	Recall	Precision
68	287	91.1%	80.8%
664	4850	93.6%	88.0%

其中 correct 為正確地被偵測到的區塊數目，missed 為應該被偵測出來卻沒有被偵測到的區塊數目，而 false 為不應該被偵測出來卻被偵測到的區塊數目；在文字偵測的部份，主要是針對已經定位出來的道路牌來進行偵測，對於定位錯誤的候選區，我們將其計算在 False Positive 之中。

當檢出率為 100%時，代表著所有的文字區塊都被找到；當精確率達到 100%時，則表示所有被找出來的區塊都是文字區塊。

6.3 錯誤分析及討論

根據 20 段影片，針對錯誤的地方我們分為三個部份做分析及討論。第一部份，造成道路牌定位出問題的原因如下：因為車子的震動或玻璃上的污漬而造成影像模糊，再則在某些條件下可能因為車子的顏色而造成判斷上的錯誤，如圖 6.6。因為在高速公路上道路牌

和照相機之間只存在著相對的運動和互相靜止的狀態，而照相機與其他車輛之間卻另外含有同一方向運動的特性，所以針對因車輛顏色而造成的定位錯誤，希望能藉由追蹤的方式，來減少這種錯誤的產生。

第二部份，造成偵測文字出問題的原因，一樣地也跟影像模糊有關，影像的模糊使得邊緣資訊減少或增加，對於文字偵測上的正確率有一定的影響；指示牌上的標誌，如指示箭頭，具有強烈的邊緣資訊；另外，高速公路上方支撐道路牌的鐵架，也會擋住部分的文字。雖然我們可以在遠方就可以得到道路牌的位置，但是道路牌上的字卻是很模糊的，這樣會使得邊緣資訊全部聚在一起，造成文字偵測上的錯誤，如圖 6.7。而對於仿射校正，若候選區的四個點被遮蔽或是額外的雜訊，會使得系統無法正確地取出轉換時所需要的參數，所校正後的結果也就無法預期。

最後一部份，針對不同的天氣、同一路段，我們對邊緣資訊做分析及討論，由表 6.3 我們可以知道邊緣資訊在晴天比在多雲的天氣更易顯現，因為對於檢出率在晴天下為 96.3%，而多雲的天氣下則是 91.4%，所以我們可以在晴天得到更準確的文字區塊；針對精確度（在多雲下為 92.4%，而晴天的天氣下則是 84.0%），我們知道晴天比多雲的天氣擁有更多的雜訊，進一步研究資料，我們發現道路牌的表面上也含有邊緣的資訊，因此影響到文字偵測後的結果。

表 6.3 邊緣資訊在不同天氣下的比較

	Numbers	False Negative
Sunny day	2557	95
Cloudy day	2625	237

False Positive	Correct	Recall	Precision
468	2462	96.3%	84.0%
196	2388	91.4%	92.4%



(a)



(b)

圖 6.6 道路牌定位上之錯誤例子 (a)車子擋風玻璃所造成的影響模糊 (b)車子顏色所造成的誤判



(a)



(b)

圖 6.7 文字偵測上之錯誤例子 (a)車子擋風玻璃所造成的影響模糊
(b)道路牌上的箭頭也具有強烈的邊緣資訊

第七章

結論與未來工作

7.1 結論

隨著科技的日益進步，數位相機的價錢越來越便宜，所以如何利用影像處理方式來改善生活，使我們有很大的興趣。在自然的環境中，雖然要考量的因素很多，但是也蘊含著大量的資訊的裡頭。道路牌就是一個在自然環境中擁有豐富資訊的例子，而本論文主要是針對輔助駕駛系統做應用，自動地對高速公路上道路牌的文字做偵測，提供駕駛者在高速公路上的導航。對於道路牌的定位，我們提供一種新的特徵空間，從影片中來尋找道路牌的候選區；再藉由簡單的幾何判斷式和邊緣資訊的含量作為過濾。

而對於道路牌的正面不一定與相機的光軸垂直的問題，也提供了對道路牌分析及轉換的方法，讓所擷取出來的道路牌能與相機的光軸垂直，使傾斜的道路牌不會影響到偵測文字和辨識文字的正確率。文字偵測部分，對每一個候選區利用 Canny 邊緣偵測把雜訊所產生的假邊界點有效地消除，所產生的結果比許多其他邊緣偵測的方法更貼近使用者的需求。最後，在實驗結果的部份也顯示我們所提出的方法，在不同的天氣(晴天、多雲)及道路(直的、彎的)的條件下也擁有著不錯的效果。

7.2 未來工作

在未來工作上，我們可以分為三個部份做努力，也許能得到更好的結果，此三個部份如下：

1. 對於道路牌定位：

- (1) 用好一點的錄影設備，讓取出來的影像更好。
- (2) 利用追蹤(tracking)來增加效率，也可以拿來作 verify 之其中一個方法，增加「道路牌定位」的精確度。
- (3) 不只是用顏色的資訊，我們也可以加入邊緣的資訊，以並行的方式來做道路牌定位。

2. 對於仿射校正：

- (1) 當牌子遇到遮蔽但文字皆無遮蔽時，藉由分析文字的排列，來得到仿射校正所需要的參數。

3. 對於偵測文字：

- (1) 統計，用密度，來當作 text 之後的 verify，增加「偵測文字」的精確度(precision)。
- (2) 結合其他的方法，如紋理的資訊，來提升整體的正確率。

由實驗結果我們可以知道，雖然做完傾斜校正後，我們可以得到正面的影像來幫助我們提昇偵測文字的精確度，但是在平均的處理速度上卻面臨著考驗；其實，我們可以不用每一張頁框都去做道路牌定位、傾斜校正和文字偵測等三個步驟；在未來工作裡，我們可以藉由

影像的分析來決定要哪一張頁框才做以上的步驟，利用此方法就可以大大提升整體系統的效率。而對於影像上初步的分析，我們發現對於同一個道路牌，邊緣資訊的含量相對於整張道路牌有減少的趨勢，如圖 7.1。所以如果我們一開始在遠方定位出道路牌後，我們就可以利用追蹤的技術來節省做其他步驟所需要耗費的時間，直到道路牌大到超過某一閾值，或邊緣資訊比例的含量小於某一閾值的時候，我們再做完之後的步驟即可。

本論文的研究其基本的概念應該可以用在對影片中文字偵測的相類似的議題上，而對於影片中文字偵測上尋找更強而有利的特徵也是令人興奮的研究。



(a)



(b)



(c)

圖 7.1 對於同一個道路牌，邊緣資訊的比例(從遠到近) (a) 0.417 (b) 0.351 (c) 0.183

參考文獻

- [1] N. Otsu, “A threshold selection method from gray level histogram,” *IEEE Trans. On Systems, Man, and Cybernetics*, SMC-8, pp.62-66, 1978.
- [2] http://www.freeway.gov.tw/11_92_05_02.asp
- [3] K. Jung, K. In Kim and A. K. Jain, “Text information extraction in images and video: a survey,” *Pattern Recognition*, vol.37 pp.977-997, 2004.
- [4] R. Lienhart and A. Wernicke, “Localizing and segmenting text in images, videos and web pages,” *IEEE Trans. Circuits Syst. Video Technol.*, vol.12, no. 4, pp. 256-268, Apr. 2002.
- [5] M. R. Lyu, J. Song and M. Cai, “A comprehensive method for multilingual video text detection, localization, and extraction,” *IEEE Trans. Circuits Syst. Video Technol.*, vol.15, no. 2, Feb. 2005.
- [6] J. P. Peters, C. Thillou and S. Ferreira, “Embedded reading device for blind people : a User-Centered Design,” in *Proceeding of Imagery Pattern Recognition Workshop* , 1550-529/04, 2004.
- [7] N. Ezaki, M. Bulacu and L. Schomaker, “Text detection from natural scene images : towards a system for visually impaired persons,” in *Proceeding of International Conference on Pattern Recognition*, 2004.
- [8] C. Y. Fang, S. W. Chen and C. S. Fuh, “Road-sign detection and tracking,” *IEEE Trans. Vehicular Technol.*, vol. 52 no.5, Sep. 2003.
- [9] X. Chen, J. Yang, J. Zhang, A. Waibel, “Automatic detection and recognition of signs from natural scenes,” *IEEE Trans. Image Processing*, vol. 13, 2004.

- [10] H. Li, D. Doermann, and O. Kia, "Automatic text detection and tracking in digital video," *IEEE Trans. Image Proceeding*, vol.9, pp.147-156, 2000.
- [11] Y. Zhong, K. Karu, and A. K. Jain, "Locating text in complex color images," *Pattern Recognition*, vol.28, no.10, pp. 1523-1536, 1995.
- [12] J. T. Tou, R. C. Gonzalez, *Pattern Recognition Principles*, pp.271 – 283, 1974.
- [13] 蘇木春, 張孝德, 機器人學習類神經網路, 模糊系統以及基因演算法則, 全華科技圖書股份有限公司, 2003 年 12 月二版六刷.
- [14] Z. Sun, G. Bebis, and R. Miller, "On-road vehicle detection using optical sensors: a view," in *Proceeding of IEEE Intelligent Transportation Systems Conference*, pp.585-590, Washington, D.C., USA, Oct. 3-6, 2004.
- [15] Y. K. Lim, S. H. Choi, S. W. Lee, "Text extraction in MPEG compressed video for content-based indexing," in *Proceeding of 15th ICPR*, vol. 4, 2000, pp.409-412.
- [16] J. C. Rojas and J. D. Crisman, "Vehicle detection in color images," in *Proceeding of IEEE Conference on Intelligent Transportation System*, pp.403-408, 1997.
- [17] L. W. Tsai, J. W. Hsieh, and K. C. Fan, "Vehicle detection using normalized color and edge map," *IEEE International Conference On Image Processing*, vol.2, pp.598-pp.601, 2005.
- [18] R. C. Gonzalez, R. E. Woods, *Digital Image Processing*, 2nd edition, Prentice-Hall , Upper Saddle River, NJ, 2002.
- [19] J. F. Canny, "A computation approach to edge detection," *IEEE Trans. On Pattern Analysis and Machine Intelligence*, 1986, vol. 8, pp.679-pp.698.

- [20] W. Wu, X. Chen, J. Yang, "Incremental detection of text on road signs from video with application to a driving assistantsystem," in *ACM Multimedia (MM '04)*, New York, USA, 2004.
- [21] G. Loy, N. Barnes, "Fast shaped-based road sign detection for a Driver Assistance System," in *Proceeding of IROS 2004*, 2004.
- [22] J. Miura, T. Kanda, Y. Shirai, "An active vision system for real-time traffic sign recognition," *Proceedings of the IEEE Intelligent Vehicle Symposium*, Dearborn, MI, pp. 52-57.
- [23] M. Steck, R. Müller, "Road sign recognition - RSR200," Computer Perception Group, School of Engineering and Information Technology - Berne University of Applied Sciences, Germany, 2003.
- [24] G. Healey, "Segmenting images using normalized color," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 22, pp. 64-73, 1992.