

南台科技大學

資訊工程研究所

碩士學位論文

室內/戶外與建築物/自然風景之影像分類
研究

**A Study on Indoor/Outdoor and
Building/Landscape Digital Image
Classificaiotn**

研 究 生：蔣佳欣

指導教授：陳定宏

中 華 民 國 九 十 五 年 七 月 七 日

南台科技大學

南台科技大學

資訊工程研究所

碩士學位論文

室內/戶外與建築物/自然風景之影像分類
研究

**A Study on Indoor/Outdoor and
Building/Landscape Digital Image
Classificaiotn**

研 究 生：蔣佳欣

指導教授：陳定宏

中 華 民 國 九 十 五 年 七 月 七 日

南台科技大學

博碩士論文授權書

本授權書所授權之論文為本人在南台科技大學 資訊工程系(所)
組 九十五學年度第 二學期取得碩士學位之論文。

論文名稱：室內/戶外與建築物/自然風景之影像分類研究

本人具有著作財產權之論文全文資料(含紙本及電子檔)，授予下列單位：

同意 <u>立即公開</u>	同意 <u>一年後公開</u>	單 位
<input type="checkbox"/>	<input type="checkbox"/>	本人畢業學校圖書館

(上述若勾選一年後公開者，請於繳交紙本論文給本校圖書館時特別聲明，否則一律以立即公開處理)

同意公開	不同意公開	單 位
<input type="checkbox"/>	<input type="checkbox"/>	國家圖書館（若同意公開，請於上傳登錄之後將所印出的授權書送交所屬系所統一寄給國圖；紙本論文則一定要送交教務處註冊組彙轉國圖，不論是否同意公開全文）
<input type="checkbox"/>	<input type="checkbox"/>	行政院國家科學委員會科學技術資料中心 （授權國科會科學技術資料中心者，請個別再寄論文一本至台北市(106-36)和平東路二段 106 號 1702 室）。

為學術研究之目的以各種方法重製得不限地域、時間與次數以微縮、光碟或其他各種數位化方式重製後散布發行或上載網站，藉由網路傳輸，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，惟每人以一份為限。

□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。

(勾選此項者，請於繳交紙本論文給本校圖書館時特別聲明，否則一律以立即公開處理)

上述授權內容均無須訂立讓與及授權契約書。依本授權之發行權為非專屬性發行權利。依本授權所為之收錄、重製、發行及學術研發利用均為無償。上述之欄位若未鈎選，本人同意視同授權。

指導教授姓名：陳定宏

研究生簽名：

(親筆正楷)

學 號：M93G0214

(務必填寫)

電話：0921536928

E-mail : sing-5@yahoo.com.tw

地 址：高雄市新興區忠孝一路76巷30號

日期：中華民國九十五年七月七日

南台科技大學



南台科技大學 碩士論文

室內/戶外與建築物/自然景物之影像分類研究

研究生：蔣佳欣

本論文業經審查及口試合格特此證明

論文考試委員

許永年 李建樹

指導教授 陳定宏

所 長 陳定宏

中 華 民 國 九 十 五 年 七 月 七 日

摘要

以影像內容來進行相片之檢索與分類最近得到廣泛的重視。最常使用的方法是利用相片之低階影像特徵或是語義註釋來分類。然而這些方法都還不能很準確的進行相片場景分類。因此若是加上相片中的所內含的 EXIF 資訊，將可提升分類之準確率。EXIF 所包含的閃光燈 (Flash Fired)、焦距 (Focal Length)、曝光時間 (Exposure Time) 在分辨室內/戶外相片上具有顯著的辨識力。本篇論文中將整合相片的低階影像特徵值與 EXIF 資訊，輔以倒傳遞類神經網路 (Back-propagation neural network) 來進行相片室內/戶外分類。

經過第一步驟戶外/室內的分類後，我們將再針對戶外的相片做進一步的辨別。因為戶外的相片大都包含建築物或是自然風景，而建築物相片相較於自然風景照，其線條構造大都充滿規則性。因此我們將可利用邊緣 (edge) 來作為主要的辨別特徵。首先利用 canny edge method 找出邊緣 (edge)，再計算出邊緣的角度，繪製成邊緣直方圖 (edge hitogram)。因建築物大都由垂直與水平線條構成，所以繪製出的邊緣直方圖，在接近垂直與水平的邊緣角度上，會呈現出雙峰的特性。因此我們可利用此特性來辨別建築物或是自然風景。

Abstract

The content-base image retrieval and classification has obtained more attention for the recent application. The most commonly used methods are applying low level image features or keywords to classify the image contents. However, these methods are not very accurate. In this thesis, we try to utilize the additional information of digital pictures to raise the classificatoin accuracy. The EXIF information such as Flash Fired, Focal Length and Exposure Time, are the embedded information to record the image information during the picture capture moment. With the additional EXIF information and the low level image features, we can easily classify the pictures from the indoor and outdoor scenes. The back propagation neural network here plays a classifier to recognize the indoor and outdoor scenes.

We will proceed the building/landscape classification from the outdoor scene pictures. The most signification feature of a building picture is the amount of the vertical and horizontal edges. First, we use canny edge operator to detect the edges of the picture. Then we calculate the angle of edge and form a edge histogram. The experiment shows the edge histogram of a building picture is bi-modal, i.e. the distribution shape looks like a “W”. The experimental results shows the proposed methods can efficiently classify the digital pictures from the indoor/outdoor and building/landscape scenes.

致 謝

本論文能如期完成，首先要感謝陳定宏教授這兩年來對我的諄諄教誨。由於有老師的悉心指導，不分寒暑假和我們一起做研究，才能讓我的碩士生涯之中，不斷的進步，不斷的努力，讓我們能專心研究，並順利完成研究所有成果。也感謝口試委員孫永年教授與李建樹教授的指導及建議，讓我的論文有所改進。

感謝我的父母及家人，能夠在我求學的過程中，不斷的鼓勵我向上；也感謝研究室的學長、同學及學弟們，因為有你們，才能讓我這兩年生過更加多姿多采。最後也感謝跟我同住兩年的室友良穗、雅婷，你們捐獻給我的相片，讓我的研究有更多的資料可以使用，最後僅此致上最誠摯的謝意與萬分的感激，謝謝幫助與指導過我的各位，謝謝。

目 次

摘 要	i
Abstract	iii
致 謝	iv
目 次	v
圖 目 錄	vii
表 目 錄	ix
第一章 簡介	1
1.1 以關鍵字為基礎的註釋方式	1
1.2 以內容為基礎的註釋方式	2
1.3 論文組織與架構	3
第二章 相片分類的介紹與系統概觀	4
2.1 相片分類－室內/戶外	4
2.1.1 利用 SVM 與 Bayesian network	4
2.1.2 利用 MPEG-7	7
2.2 相片分類－建築物	8
2.3 系統架構	9
第三章 室內/戶外特徵值擷取與演算法	12
3.1 色彩空間	12
3.1.1 RGB 色彩空間	13
3.1.2 HSV 色彩空間	14
3.1.3 YCbCr	16
3.2 LAB 色彩空間	17
3.3 色彩特徵值擷取	18

3.4 亮度特徵值擷取.....	22
3.5 數位相片的 EXIF 資訊.....	22
3.6 類神經網路原理.....	26
第四章 建築物/自然風景特徵值擷取與方法.....	38
4.1 邊緣.....	40
4.1.1 Roberts 邊緣偵測	40
4.1.2 Prewitt 邊緣偵測	41
4.1.3 Sobel 邊緣偵測.....	42
4.1.4 Canny 邊緣偵測	43
4.2 邊緣特徵值擷取.....	44
4.3 辨別建築物的方法.....	49
第五章 實驗結果.....	54
5.1 室內/戶外辨別結果.....	54
5.2 建築物/自然風景辨別結果.....	63
5.3 系統介面.....	69
第六章 結論與未來工作.....	73
參考文獻	75

圖目錄

圖 2.1	線性向量支援機	5
圖 2.2	非線性向量支援機	5
圖 2.3	五種型態的邊緣	8
圖 2.4	系統架構圖	10
圖 2.5	系統流程圖	11
圖 3.1	各國國旗	12
圖 3.2	人眼對顏色的敏感曲線	13
圖 3.3	RGB 色彩空間	14
圖 3.4	彩度-色相表	15
圖 3.5	HSV 色彩空間	16
圖 3.6	LAB 色彩空間	18
圖 3.7	特徵值擷取範圍	19
圖 3.8	藍色特徵值	21
圖 3.9	綠色特徵值	21
圖 3.10	曝光時間在室內戶外的分布情形	24
圖 3.11	生物神經元模型	27
圖 3.12	人工神經元模型	28
圖 3.13	正切雙彎曲轉換函數	29
圖 3.14	對數雙彎曲轉換函數	29
圖 3.15	Back-propagation neural network 架構	32
圖 3.16	類神經網路模型	36
圖 3.17	倒傳遞演算法所使用的特徵值	37
圖 4.1	邊緣比較圖	39

圖 4.2	邊緣角度直方圖	40
圖 4.3	四種邊緣方法比較圖	45
圖 4.4	canny edge threshold 比較圖	47
圖 4.5	邊緣方向	48
圖 4.6	四種邊緣方向	48
圖 4.7	邊緣方向對應圖	49
圖 4.8	角度邊緣直方圖	50
圖 4.9	邊緣圖	51
圖 4.10	建築物定位過程	53
圖 5.1	實驗組室內相片	57
圖 5.2	實驗組戶外相片	58
圖 5.3	對照組室內相片	59
圖 5.4	對照組戶外相片	60
圖 5.5	測試組室內相片	61
圖 5.6	測試組戶外相片	62
圖 5.7	分辨錯誤相片	63
圖 5.8	建築物相片	65
圖 5.9	自然風景相片	66
圖 5.10	辨識錯誤建築物與自然景物相片	66
圖 5.11	錯誤辨別結果解釋圖	67
圖 5.12	建築物辨別定位結果	68
圖 5.13	系統介面圖	69
圖 5.14	辨別室內結果介面圖	70
圖 5.15	戶外/建築物辨識結果介面圖	71
圖 5.16	戶外/自然風景辨別結果介面圖	72

表 目 錄

表 3.1	EXIF 資訊.....	23
表 3.2	焦距在室外戶外的統計	25
表 3.3	閃光燈在室內戶外使用情形	26
表 5.1	訓練組與對照組相片 RGB 色彩空間標準差	55
表 5.2	訓練組與對照組亮度標準差	55
表 5.3	對照組的相片透過倒傳遞網路輸出的正確率	55
表 5.4	測試組相片正確率	56
表 5.5	測試組的相片透過倒傳遞網路輸出的正確率	56
表 5.6	建築物辨別正確率	64



第一章 簡介

近年來，隨著現在資訊科技發達，產生愈來愈多的數位商品，讓人們走到哪都可以隨心所欲的使用。在一切數位化的改變之下，相機也已從傳統利用底片紀錄轉變為僅需使用一張小小的記憶卡就可儲存，儲存的張數甚至比之前的膠捲底片多上好幾十倍。而且在拍照的同時還可以直接觀看影像，拍照後還可立即流覽相片是否拍的完美。因此很多人都利用拍照來替生活做紀錄，而產生大量的數位相片，也造成了管理上的不便。如何將這些龐大的數位相片加以適當的儲存、管理與應用，成為研究上的重要課題[16]。

然而，影像圖片的應用先決需求在於能夠迅速、有效的取得使用者所要的影像或是圖片，因這些使用者的背景不同，所需求的內容與方式也會有很大的不同，因此該如何從龐大的資料庫中找出使用者所需的圖片，也就是最重要的問題所在。早期用來查詢資料庫中的圖片是以文字檢索（Text-based）為主，近年來則以內容檢索（Content-Based）為重心。文字檢索以人類所使用的文字或語言為檢索的依據，屬於高階語意式檢索，而內容檢索則是直接萃取影像本身的特徵（例如color、shape、texture）作為檢索的依據，屬於低階檢索，以下將分別進行說明。

1.1 以關鍵字為基礎的註釋方式

以關鍵字來註釋相片，因其操作簡單所以較為大眾使用。主要是先將影像做分類，或是加上註解或描述性的文字（例如註明影像的名稱、作者、拍攝日期、圖片的內容等）。當使用者需要某張圖片時，即可輸入圖片名稱或是類

別等，當作查詢的依據。這種註釋的好處就是使用人類熟悉的語言與文字，但註釋時需要人工處理較為耗時，另外不同的註釋者，所用的註釋言語不同，也會造成分類上的歧見，使搜尋結果不同，而且在搜尋上往往也會花上相當多的時間來尋找正確的結果。舉例而言，若某甲欲搜尋有關過年的室內相片，但相片註釋者在註釋相片時並沒有清楚的註釋室內或戶外，因此註釋者輸入「過年」此關鍵字時所搜尋出的相片幾乎皆為大家出遊的戶外相片，因此某甲就必須要花費很多的時間利用人工選取的方式，才能找到所需的相片。有鑒於此，能夠自動分類以及自動檢索技術的需求因此產生，也就是所謂的內容檢索。

1.2 以內容為基礎的註釋方式

由於影像圖片快速大量的成長，如用人工處理費力又耗時，外加上人為註釋圖片或是使用者查詢時，均無法避免個人主觀性的偏差，因此從 90 年代開始，許多學者著手發展以圖片內容為基礎的影像檢索系統，稱之為“Content-Based Image Retrieval (CBIR)”[17-22]，對於眾多影像的管理，提供另一種分類方式。內容檢索以相片的低階影像特徵值(low-level image feature)，例如色彩(color)、紋理(texture)、邊緣(edge)、形狀(shape)等來當作檢索的依據。但是使用低階影像特徵值所認為同性質的東西，是無法直接應對到人類的語義；同樣的人類語意認為相同的兩個場景，卻可能具有完全不同的顏色或是紋理等影像特徵，因此可外加上數位相片內所附帶的 EXIF 資訊來加以輔助處理分類的問題。EXIF 資訊如時間戳戳(Timestamps)，曝光時間(Exposure Time)，閃光燈(Flash Fired)使用與否，焦距(Focal Length) ... 等，提供拍照時的基本資訊。因此若能充分利用上述資訊來做分類的特徵值，將可更精確地分類諸如像室內或戶外的分類。

1.3 論文組織與架構

本篇論文在第二章將會介紹目前在相片分類所使用的方法及系統概觀。第三章將會詳細的介紹室內/戶外所需的特徵值、EXIF 資訊與倒傳遞類神經網路。第四章將介紹辨別建築物與自然風景所用的特徵值。第五章將介紹本論文所提出之系統架構，辨別的結果和討論。我們藉由實驗，來驗證所提出的方法是否能達到較好的效能。第六章為結論與未來的發展方向。

第二章 相片分類的介紹與系統概觀

近年來愈來愈多人使用語義來作分類場景的依據[12-15]，在特定的照片裡很多的分類系統都利用影像中的低階特徵值（例如：色彩、紋理、邊緣），大都能分辨出不錯的結果，例如像室內戶外場景的辨別[6-11]、建築物與自然景物的辨別[25-28]亦或是人物與動物的辨別等。以下將針對一些相片分類的種類與方法做大略的介紹。

2.1 相片分類－室內/戶外

在一般隨意拍攝的相片中，所包含的種類非常的廣泛，因此在場景的分類上無法準確的分辨出正確的結果，尤其是只利用相片中的低階特徵值來分類，因此就衍生出相當多的方法來作為分類的依據。在室內與戶外的分類中，大都採用影像中的低階特徵值來作為特徵值。首先藉由 SVM 來擷取色彩特徵值，之後在藉由 Bayesian network 來分類。但除了可用 SVM 來擷取特徵值外，MPEG-7 也提供了擷取特徵值的方法。因此以下將介紹一些不同的分類方法。

2.1.1 利用 SVM 與 Bayesian network

SVM 全名為支援向量機（support vector machine）[56-58]，可分為線性與非線性兩種模式。使用線性的情況如圖 2.1 所示，能將兩類樣本正確的分開，而且分類間隔為最大。SVM 可利用現有的資料做訓練，再利用訓練出來的結果，剔除少部分極端樣本點，選出幾個主要的向量作為整體的資料代表，然後將所選取的向量當成將來測試資料歸類的依據，如方程式（2.1）。

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n), x_n \in R^n, n=1, 2, \dots, i \quad (2.1)$$

$$y_n \in \{+1, -1\}$$

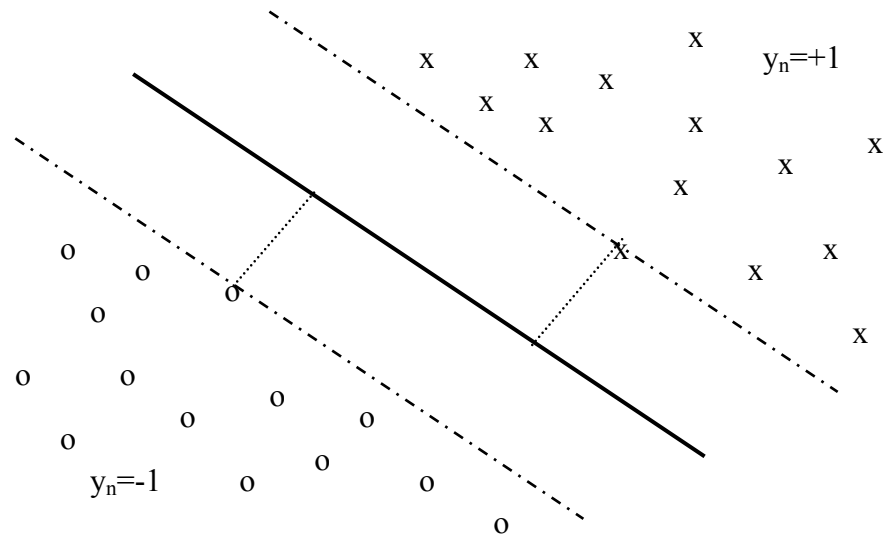


圖 2.1 線性向量支援機

不過有些資料並沒有辦法用線性來做分類，因此必須透過非線性支援向量機來作分類。首先將現有的原始資料轉換至另一個高維度的特徵空間，在高維度的空間內即可用線性的方式來做分類，如圖 2.2 所示。接下來的方法與線性支援向量機相同。

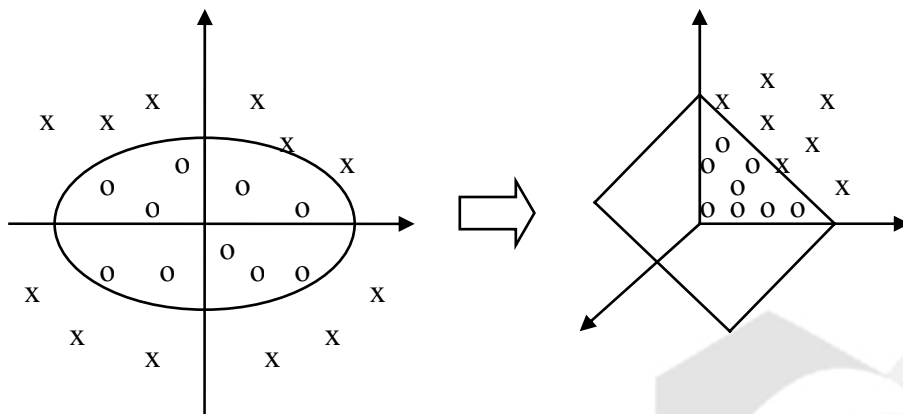


圖 2.2 非線性向量支援機

貝氏網路 (Bayesian network) [54,55]利用條件機率作為預測的機制，可

藉由先前的歷程與經驗，預測所有可能發生狀況的機率。假設現存某個特徵值 A 及某個類別 B，則 $P(A)$ 為該特徵值出現的預測機率， $P(B)$ 為任意亂數取出特徵值落於類別 C 的機率，稱之為事前機率 (prior probability)，則根據條件機率 (conditional probability)，可表示如方程式 (2.2)。

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A|B)P(B)}{P(A)} \quad (2.2)$$

其中 $P(B|A)$ 表示特徵值 A 出現落在類別 B 的機率，也稱為事後機率 (posterior probability)，則 $P(A|B)$ 為落於類別 B，恰巧為特徵值 A 的機率。

Navid et al. [6]，僅利用相片本身的低階影像特徵值 (low-level image feature) 色彩 (color) 與紋理 (texture) 作為分類的依據。利用兩次的 SVM 來分類出 indoor 或 outdoor，先分別擷取色彩與紋理的特徵值，經由 SVM 做大致上的分類，將圖片分成 16 各區塊，再分別計算各特徵值間個別區塊的距離，再經由 SVM 訓練分類結果。

Jiebo & Andreas [7]，利用語義特徵值和相片本身的低階影像特徵值 (low-level image feature) 做為分類的依據。語義的特徵值分為天空 (sky) 跟草地 (grass)，而低階特徵值為色彩 (color) 與紋理 (texture)，使用 Bayesian network 來分類，實驗結果分成兩組，一組是使用 ground truth 的資訊定義天空 (sky) 跟草地 (grass)，另一組是由低階特徵值計算出。

Kane & Savakis [8]，利用語義特徵值和相片本身的低階影像特徵值 (low-level image feature) 做為分類的依據。利用 SVM 來擷取低階特徵值為色彩 (color) 與紋理 (texture)，使用 Bayesian network 來分類，但因 Bayesian network 很可能會錯誤預估機率，因此事先估算可能的錯誤率，加以更正提高準確率。

Boutell & Jiebo [10]，利用拍攝照片的連續時間性，外加上相片本身的低階影像特徵值（low-level image feature）做為分類的依據，藉由 Bayesian 定理來計算出可能發生的機率，先使用 dynamic programming 演算法找出最佳路徑，再利用 Viterbi 演算法來記錄最佳的路徑以提高效率。

Boutell & Jiebo [11]，利用相片本身的低階影像特徵值（low-level image feature）和 EXIF 資訊作為分類的依據。利用 SVM 來擷取低階特徵值為色彩（color）與紋理（texture），再結合曝光時間（Exposure Time）、閃光燈（Flash fired）、目標距離（Subject Distance），最後使用 Bayesian network 來分類。

2.1.2 利用 MPEG-7

MPEG-7[48-53]這套系統中包含了，Systems（系統）、Visual（視覺）…等部份，在Visual中又包含了Color（色彩）、Texture（紋路）、Shape（形狀）…等部分。其中Color又包含Dominant color、Scalable color、Color layout和Color structure。因此底下將針對[4]所提到的特徵值作個概略的介紹。

Color Layout Descriptor（CLD）用來了解在影像裡色彩在空間中散佈的情形，色彩空間必須使用YCbCr。過程可分為四個步驟。

1. 將圖片分割成 8×8 的區塊；
2. 將區塊內的色彩空間平均量化；
3. 將分割後的影像作DCT轉換；
4. 在作zig-zag掃描將結果紀錄下來。

Scalable Color Descriptor（SCD）利用Haar轉換來重新對影像編碼，將結果量化成為256 bins，其色彩空間必須使用HSV。

Edge Histogram Descriptor (EHD) 計算影像邊緣的散佈情形，總共將邊緣分成五種不同的型態， 0° 、 45° 、 90° 、 135° 與沒有方向性的邊緣，如圖 2.3 所示。

Spyrou et al. [9]，利用MPEG-7中的Color Layout Descriptor (CLD)、Scalable Color Descriptor (SCD)、Edge Histogram Descriptor (EHD) 三個Descriptors，來當作分類的特徵值，使用三種方法來做實驗，第一種是直接合併所擷取的特徵值來做為分類的依據，第二種是經由Back-propagation訓練，使用KNN (K-Nearest Neighbor) 來分類，第三種是使用Falcon-ART Neurofuzzy Network 來分類。

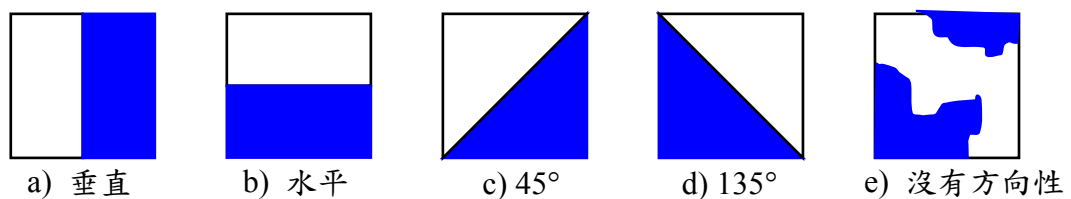


圖 2.3 五種型態的邊緣

2.2 相片分類—建築物

Vailaya et al. [25]開發一個程序來測量顯著的特徵值，根據那個特徵值的內部類別和相關類別距離在平面圖上的分布來分類。他們使用低階特徵值來解決高階分類問題，並以邊緣方向凝聚直方圖 (edge direction coherence histogram) 來辨別都市風景 (cityscape) 和風景圖像 (landscape)。這個特徵值可從隨意的邊緣 (arbitrary edge) 分布中尋找出有具有結構的邊緣 (structure edge)。在圖片中出現人造 (man-made) 的或具有結構性的物體，則在邊緣方向凝聚直方

圖 (edge direction coherence histogram) 裡的尖峰分佈將會較趨近於重大邊緣方向附近，但是在自然影像的邊緣方向凝聚直方圖 (edge direction coherence histogram) 看來趨近於平坦。

Dorado 和 Izquiere [26]，利用 MPEG-7 邊緣直方圖描述器 (edge histogram descriptor) (利用 80-point 直方圖來表示圖片內邊緣方向的分布： 0° 、 45° 、 90° 、 135° 和沒有方向，五種情況) 和局部和整體的邊緣散佈情況來做分類。其方法為先粗略的對應和之後再藉由使用者的相關性回饋，來調整分類的結果。圖片被分割成為 16 個大小相等的子圖片 (sub-image)，再將每個子圖片分割成數個非重複小方形的區塊，並且給予一個指定的數字。再將區塊分成四個子區塊 (sub-block) 和經由 5 過濾器 (filters) 將他們分配到一個相對應的邊緣類別。經由上述邊緣的類別可以了解圖片邊緣分布的情況的。

Iqbal 和 Aggarwal [27,28]，為了偵測較大的人造物體，譬如大廈、橋梁、塔等，根據圖片的感知群組 (連續性、接近度、共線性、共環狀、對稱、平行性)。針對建築物圖片，可分成 3 個種類：建築物 (building)、中型物 (intermediate) 和沒有建築物 (non-building)。而建築物的特徵也可分成三種：“L”型的連接點、“U”型的連接點和“顯著”的平行線段。

2.3 系統架構

本系統的架構如圖 2.4 所示，將相片分類為室內與戶外，接下來可再針對戶外的相片，細分為風景與建築物的相片，其過程如圖 2.5 所示。首先於室內/戶外 (indoor/outdoor) 相片分類問題上，先轉換相片的色彩空間 (color space)，再擷取相片中的低階影像特徵值 (low-level image feature) 與數位相片中所包含的 EXIF 資訊，經正規化後，再放入倒傳遞類神經網路 (Back-propagation neural

network) 進行訓練。經由第一步驟的分類後，再針對戶外的照片作更進一步的分類，將其分類為風景照 (landscape) 及建築物照 (building)。利用 canny edge 擷取相片的邊緣 (edge)，之後再針對垂直的線段加以分析，取出較長的垂直線段作為第一次的分類依據，如達到我們所訂定的門檻值 (threshold) 則判定為建築物的照片。接下來再針對未達到第一次標準的照片做第二次的分類，同樣的如達到我們所訂定的標準則判別為建築物，否則判別為風景照。

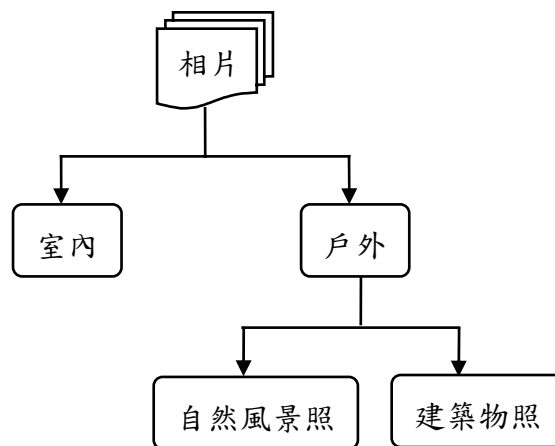


圖 2.4 系統架構圖

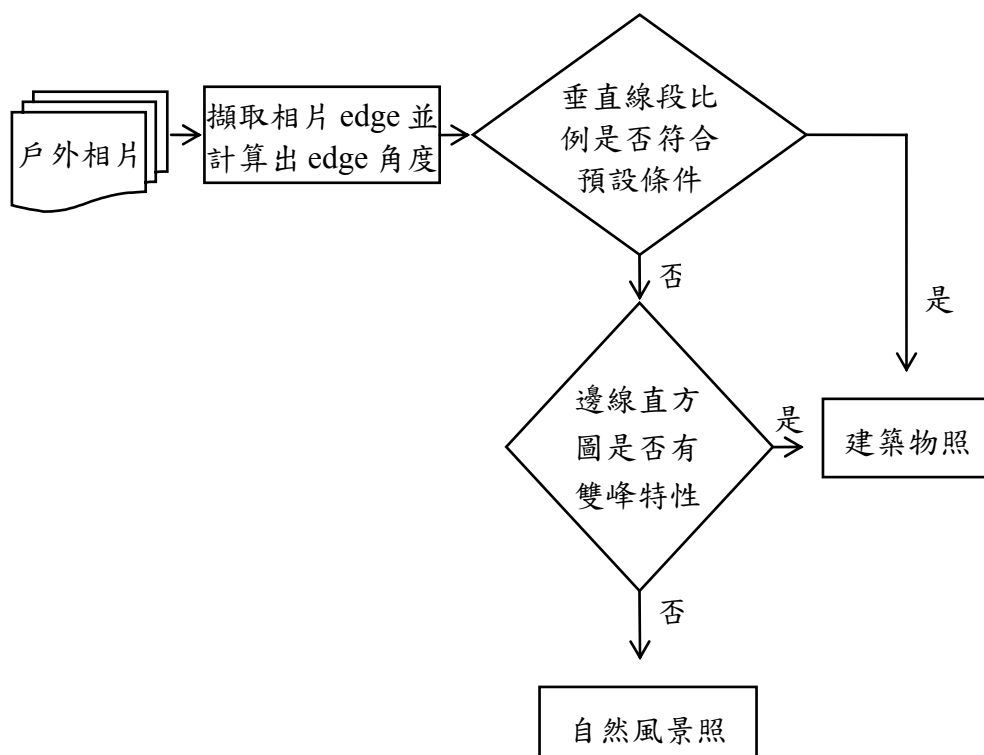
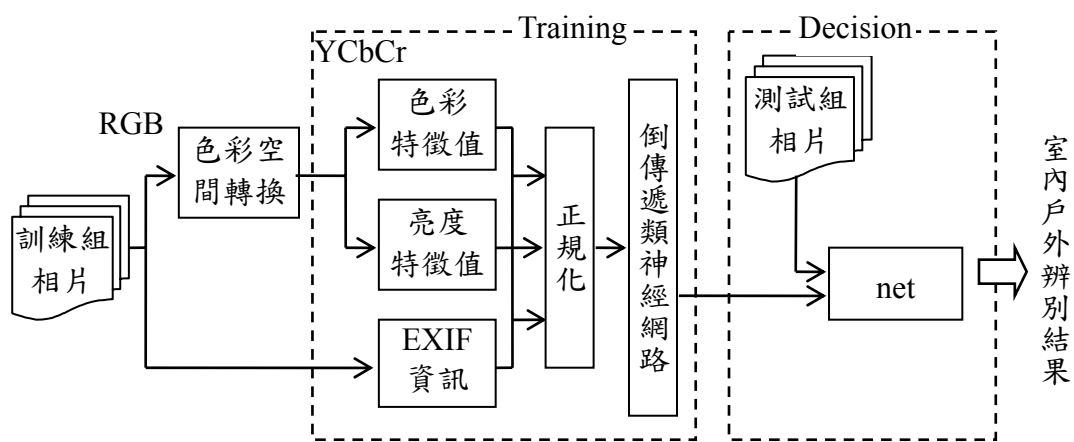


圖 2.5 系統流程圖

第三章 室內/戶外特徵值擷取與演算法

特徵值擷取演算法是一個分類系統相當重要的基礎，良好的特徵值範圍擷取可以提升系統分類的準確率。由於從相片中抽取顏色較為簡單而且與影像大小及位置無關，外加上通常我們看到一張圖片第一個印象就是它的顏色分佈，所以顏色對我們而言是一個很重要的特徵值[1-5, 23,24,59]，可以讓我們很快的就辨別出圖片的差異性。例如圖 3.1 為各國國旗，用顏色我們就可以分辨出各國國旗所代表的國家，如果使用其他的特徵值（Texture、Shape）來分辨的話，一定會無法正確的分辨出結果。



圖 3.1 各國國旗

在分辨室內或戶外時，大部分戶外的相片可能充滿藍色（天空）或綠色（草地），外加上戶外的相片大多光源充足，所以往往會比室內的相片來的更亮，因此亮度在分類室內或戶外的相片上也是非常值得參考的重要資訊。然而要在更近一步針對室外照片辨別城市風景或風景照時，邊緣（edge）的線段將是不可或缺的重要特徵，因此若能善用這些資訊，將可提升分類效率。

3.1 色彩空間

色彩卻是人類視覺系統判斷物體最主要的因素之一，主要是靠光進入眼睛

後產生的知覺反應結果。人類眼睛視網膜中具有兩種感光體，一種稱之為椎狀體（cones），也稱為白晝（photopic）視覺或亮光視覺，因此對色彩很敏感；另一種稱之為桿狀體（rods），也稱為夜視覺（scotopic）或昏暗視覺，桿狀體用來獲得視野中大體的影像，不含彩色視覺，因此對亮度很敏感。人眼中的椎狀體可被分成三個感應類別，大體上相對於紅色、綠色和藍色，圖 3.2 展示出人眼對顏色的敏感曲線。以下將會論述幾個比較常用的色彩空間。

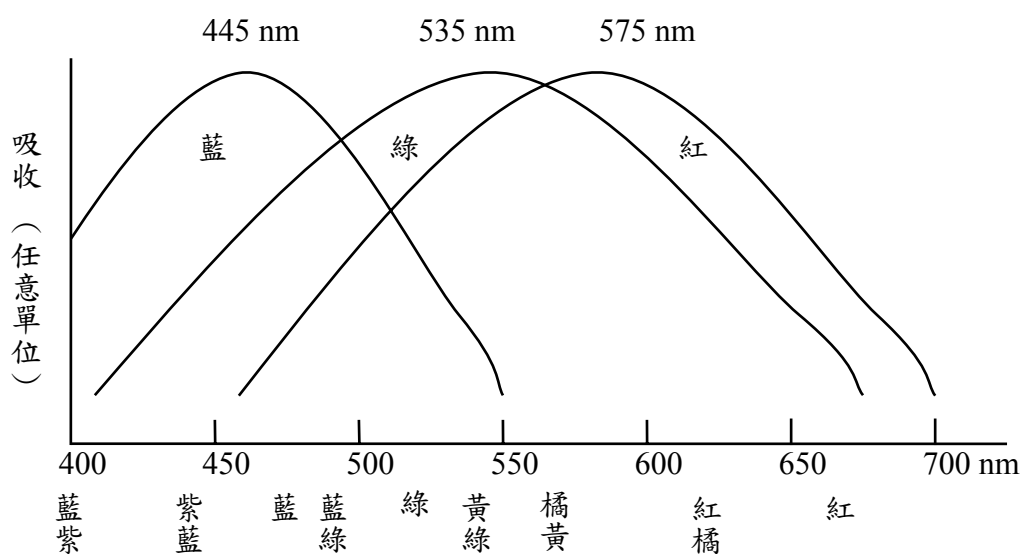


圖 3.2 人眼對顏色的敏感曲線

3.1.1 RGB 色彩空間

RGB 色彩空間—以光的三原色紅、綠、藍為刺激值混合出某一色彩的顯像方式，愈混合後會得到愈明亮的色光，其色彩表示如方程式 (3.1)。整個 RGB 色彩空間可以一個立方體來表示，如圖 3.3 所示。依照一般便利的設定，R、G、B 皆被正規化至 $[0, 1]$ 的範圍之間。若 $(R, G, B) = (0, 0, 0)$ ，則表示黑色；反之若 $(R, G, B) = (1, 1, 1)$ ，則表示白色。而 R、G、B 三值相等時，則會出現介於最低值—黑色—與最高值—白色—之間的灰階值，而灰階值只保留彩色影像的亮度，忽略了色調及飽和度。最簡單的彩色影像轉灰階值方法便是將 R、G、

B 值加總以後，平均即可得到灰階值。

$$C_{\lambda} = r\bar{R} + g\bar{G} + b\bar{B} \quad (3.1)$$

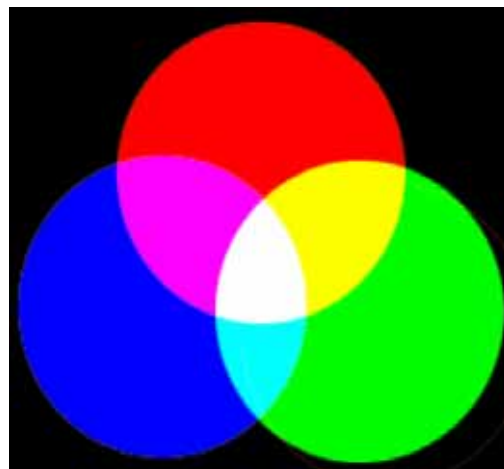
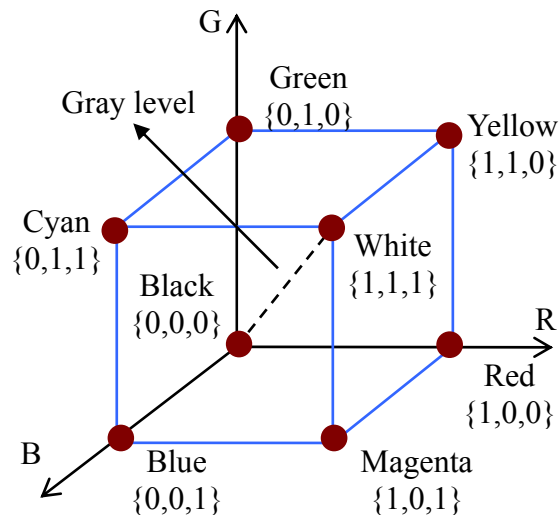


圖 3.3 RGB 色彩空間

3.1.2 HSV 色彩空間

在 RGB 基礎色族內的色域組包括 RGB 空間、HSV 空間和 HLS 空間。HSV 和 HLS 空間則由 RGB 空間轉換而來用以敘述色彩的術語，HSV 代表色相

(Hue)、飽和度 (Saturation) 和明暗度 (Value)，由於每種元素之間的關聯性小，因此其數值不會互相影響。這些都是在色彩學上相當常見而熟悉的名詞，雖然其意義相當接近，在定義上仍有些微差異，但結構是相同的。這三種特性建構成一個三維的六角椎體，如圖 3.5，中心垂直軸為亮度，橫切面愈往上則代表愈亮，每個橫切面都是由不同的色相與飽和度構成的。至於彩度為色彩的純度或飽和度，亦即此色彩中含多少純色之多寡，如圖 3.4。S 與 V 介於[0, 1]的範圍內，H 介於 0°~360°之間，方程式 (3.2) ~ (3.5) 為 RGB 色彩空間的轉換至 HSV 色彩空間的公式：

$$V = \max(R, G, B)$$

$$S = \begin{cases} \frac{V - \min(R, G, B)}{V} & \text{if } V \neq 0 \\ 0 & \text{otherwise} \end{cases} \quad (3.3)$$

$$H = \begin{cases} 5 + B' & \text{if } R = \max(R, G, B) \text{ and } G = \min(R, G, B) \\ 1 - G' & \text{if } R = \max(R, G, B) \text{ and } G \neq \min(R, G, B) \\ 1 + R' & \text{if } G = \max(R, G, B) \text{ and } B = \min(R, G, B) \\ 3 - B' & \text{if } G = \max(R, G, B) \text{ and } B \neq \min(R, G, B) \\ 3 + G' & \text{if } B = \max(R, G, B) \text{ and } R = \min(R, G, B) \\ 5 - R' & \text{otherwise} \end{cases} \quad (3.4)$$

$$\text{let } R' = \frac{V - R}{V - \min(R, G, B)} \quad G' = \frac{V - G}{V - \min(R, G, B)} \quad B' = \frac{V - B}{V - \min(R, G, B)} \quad (3.5)$$

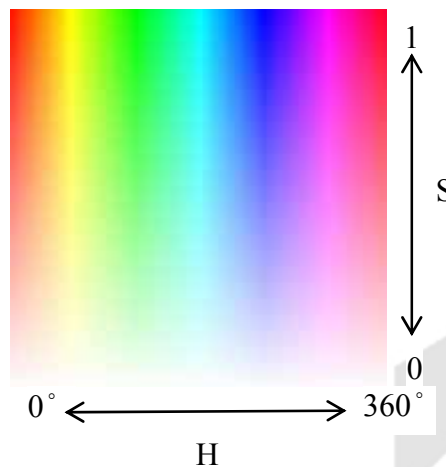


圖 3.4 彩度-色相表

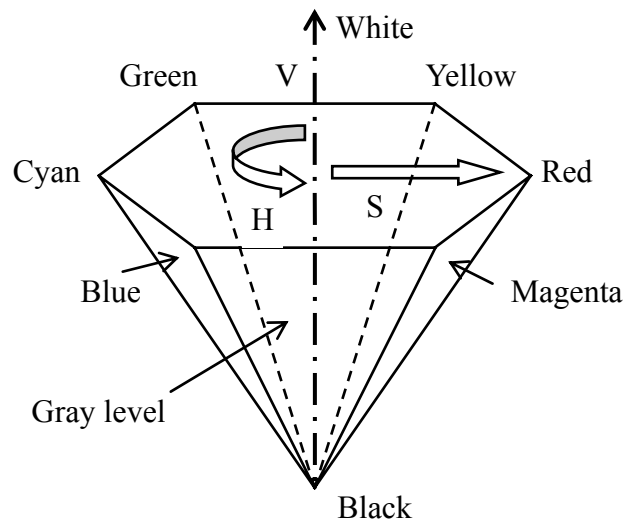


圖 3.5 HSV 色彩空間

3.1.3 YCbCr

YCbCr 色彩空間是修改 YUV 色彩空間所得到的，分為亮度元素 luminance(Y)與 blueness(Cb) 、redness(Cr)兩個彩度元素，對亮度的分離性高，方便與彩度分開操作。根據研究顯示，人眼察覺到的光亮度資訊 (Y) 的 60% ~70% 來自於綠色光，紅色和藍色實際上只是亮度資訊的複製，因此這些重複的資訊可以完全去除掉。最終的結果為圖像可用表徵的亮度與色度來表示。因此系統規定亮度的取值範圍介於 16~235 之間，而 Cb 與 Cr 的取值範圍介於 16 ~240 之間，並規定亮度 0 表示為 128。方程式 (3.6) 為 RGB 色彩空間的轉換至 YCbCr 色彩空間的公式：

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.5 & -0.4187 & -0.0813 \\ -0.1687 & -0.3313 & 0.5 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 0 \\ 128 \\ 128 \end{bmatrix} \quad (3.6)$$

3.2 LAB 色彩空間

LAB 色彩空間是由國際照明協會（CIE）根據人眼視覺的特性，把光線波轉換為亮度與色相的一套描述色彩數據。自然界中的任何顏色都可以用 LAB 色彩空間中來表達，其色彩空間甚至比 RGB 還廣大。此外，此系統是以數字化來描述人類的視覺感應，與設備無關，所以彌補了 RGB 與 CMYK 系統必須依賴設備的缺失。因此，LAB 色彩空間已成為世界各國正式採納、作為國際通用的測色標準，適用於一切光原色或物體色的表示與計算。

LAB 色彩空間是由 CIEXYZ 系統以數學的方式轉換得到，如方程式 (3.7) 所示：

$$\begin{cases} L = 116 \left(\frac{Y}{Y_o} \right)^{\frac{1}{3}} - 16 \\ A = 500 \left[\left(\frac{X}{X_o} \right)^{\frac{1}{3}} - \left(\frac{Y}{Y_o} \right)^{\frac{1}{3}} \right] \\ B = 200 \left[\left(\frac{Y}{Y_o} \right)^{\frac{1}{3}} - \left(\frac{Z}{Z_o} \right)^{\frac{1}{3}} \right] \end{cases} \quad \frac{Y}{Y_o} > 0.01 \quad (3.7)$$

其中 X、Y 與 Z 是物體的三各刺激值； $X_o Y_o Z_o$ 為 CIE 標準照明體的三個刺激值，L 為色彩的亮度，A 與 B 為色彩的範圍。如圖 3.6 所示，+a 表示紅色，-a 表示綠色，+b 表示黃色，-b 表示藍色，而色彩的亮度為 L 用百分數來表示。

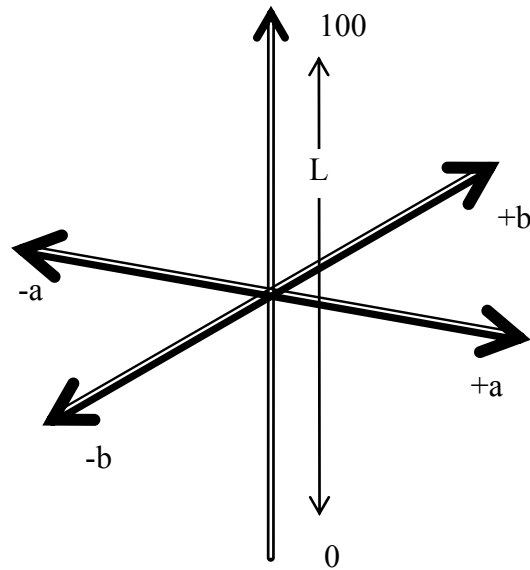


圖 3.6 LAB 色彩空間

3.3 色彩特徵值擷取

戶外相片中常見之主要組成成分大都為天空、草地、海洋、山景等，為了要從相片中找出藍色與綠色的區域。首先，必須先定義出天空、草地、海洋、山景的顏色，利用色彩參數將上述的顏色做具體化描述，因此必須選擇合適的色彩空間來描述顏色的資訊。

在影像處理中，常用的色彩空間如 RGB、HSV、HSI、HSL、YCbCr、CMY 與 LAB 等，由於亮度變化會使同一種顏色在 RGB 色彩空間中呈現不同值，因此，未避免環境亮度變化對藍色與綠色的定義造成影響，必須選擇能從色彩資訊中分離出亮度因素之色彩空間，其中 YCbCr 與 HSV 兩種色彩空間最常使用。

此篇論文中我們所採用的色彩空間為 YcbCr，其原因為 YCbCr 色彩空間對於光線變化的反應比 HSV 色彩空間不敏感。因此可利用方程式 (3.6) 將相片的 RGB 色彩空間轉換成我們所需的 YCbCr 色彩空間。然而為了要辨別出戶外

主要成分之色彩組成，我們首先個別擷取相片中藍天、海洋、草地和山的部分，如圖 3.7 紅色圈選部分，再分別去統計其各自的色彩分布，之後在利用聯集的方式來擷取數值。

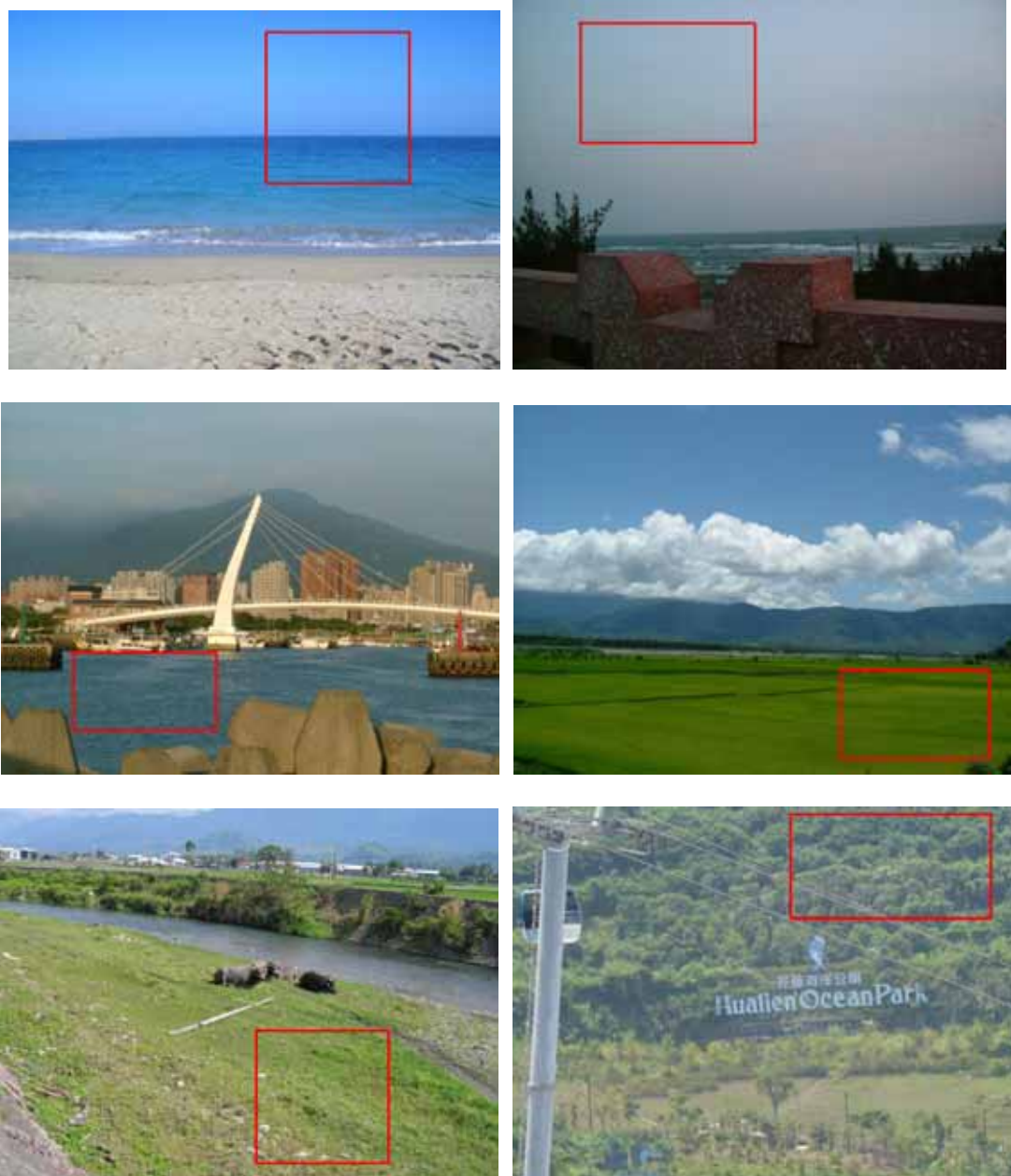


圖 3.7 特徵值擷取範圍

取出樣本之色彩範圍後，針對要訓練、對照和測試的相片，分別計算其藍色、綠色分布，其方法如下：

1. 先將相片的色彩空間轉換成為 YCbCr；
2. 針對相片中的像素，將其每一點逐一與所設定範圍的 Cb、Cr 值比對，紀錄符合範圍值的像素數目；
3. 最後將統計的像素數目結果除以整張相片的像素數目，取其百分比。

經過上述的步驟擷取的結果，綠色與藍色在 Cb 分量和 Cr 分量所佔的值域如下：

藍色

$$128 \leq Cb \leq 157$$

$$89 \leq Cr \leq 125$$

綠色

$$82 \leq Cb \leq 129$$

$$85 \leq Cr \leq 130$$

爲了確認我們定義的色彩空間，可以正確達到擷取的效果，如圖所示，以測試我們所定義的色彩空間是否真正的適用

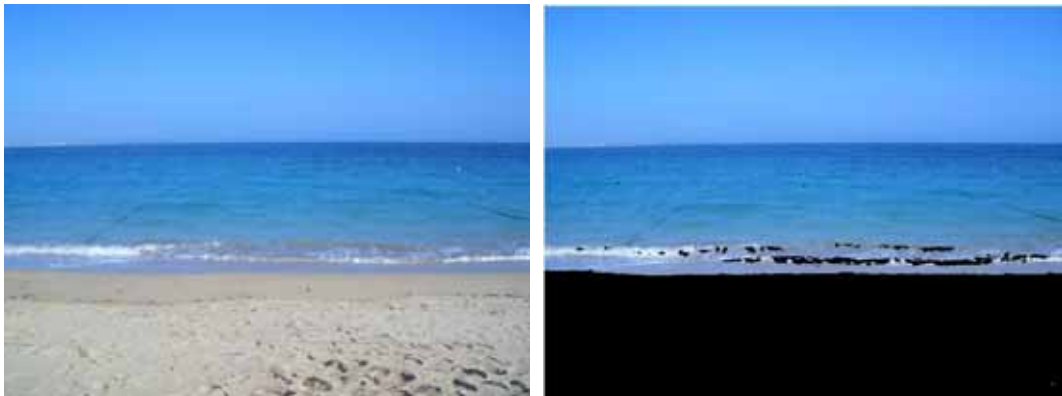




圖 3.8 藍色特徵值



圖 3.9 綠色特徵值

3.4 亮度特徵值擷取

相片拍攝在戶外的亮度跟室內的亮度有所差別。一般來說，室內的相片大多光源不充足，即使經由日光燈的照射，其光源的分布也不均勻，亮度也無法達到戶外的水準。反之，戶外因陽光的照射光源分部較為均勻、光亮，因此對辨別室內戶外也是一個很重要的特徵值。在色彩空間上同樣的也選擇 YCbCr 色彩空間中的 Y (luminance, 亮度) 來作為判別的特徵值，其範圍為 16~235，數值愈大代表愈亮，方法如下：

1. 先將相片的色彩空間轉換成為 YCbCr；
2. 在相片中取出每個像素的亮度，全部累加；
3. 累加後的總亮度值，再除以整張相片的像素數目，計算整張相片的亮度平均值。

3.5 數位相片的 EXIF 資訊

EXIF 資訊最早是由 JEIDA (Japan Electronic Industry Development Association 日本電子工業振興協會) 於 1995 年 10 月制定的，全名為 Exchangeable Image File(可交換圖片文件)。而後於 1997 年 11 月推出 EXIF 2.0 版，最後在 1998 年 6 月推出 EXIF 2.1 升級版，增加了對音頻文件的支持。當 ISO (國際標準組織) 要為 DCF (Design rule for Camera File system) 數位相機檔案系統制定規章的時候，發現 JEIDA 標準涵蓋了數位相片的需求，因此直接引用修改成今日的 EXIF 標準。2002 年 2 月，JEITA (日本電子情報技術產業協會) 公佈了最新草擬 EXIF 2.2 版，其中最大的改革在於將一系列數位相機拍攝數據和列印屬性資料標準化。EXIF 是數位相機在儲存 JPEG 檔案的同時，所

加入的一些資訊，目的是用來記錄攝影時的光圈、快門、ISO、拍攝日期、相機機種等各種和拍攝當時有關的訊息，如表 3.1 所示，以上這些特徵值將可依據不同種類的分類，提供不同程度的參考。

表 3.1 EXIF 資訊

廠牌	SANYO Electric Co.,Ltd.
型號	J4
日期	2006/2/8 上午 02:33:59
曝光時間	0.0316456 s
光圈值	f/7
曝光程式	常規模式
ISO 曝光速度級數	100
亮度值	未知
曝光偏斜值	0
光圈最大值	3
光源	未知
閃光	已閃光 [開啟]
焦距	6 mm
色彩空間	sRGB
X 畫素數	598
Y 畫素數	800
場景類型	照片
自訂修正	正常過程
曝光模式	自動曝光
白平衡	自動白平衡
數位放大倍率	0
相當於 35mm 相機的焦距	38 mm
對比度	標準輸出
飽和度	標準輸出
主體距離範圍	微距
Exif 版本	2.2

對於我們所做的室內或戶外的分類，所需要的特徵值為其中的三項，即曝光時間（Exposure Time）、焦距（Focal Length）和閃光燈（Flash Fired）。我

們將針對這三項加以說明（下列各數據皆由 150 張室內及 150 張戶外的訓練組相片，所統計得到的結果）。

曝光時間 (Exposure Time)：簡單來說就是快門開關一次的時間，曝光愈久感光元件感應到的光量就越大，所拍攝出來的畫面就愈亮（可能過曝），反之則愈暗（能曝光不足）。

一般來說，相片拍攝在戶外的亮度跟室內的亮度有所差別（即使在陰天的情況下），其戶外曝光時間會比在室內來的短，如下圖 3. 10 所示，戶外的相片曝光時間大都集中在 0.02 秒以下。

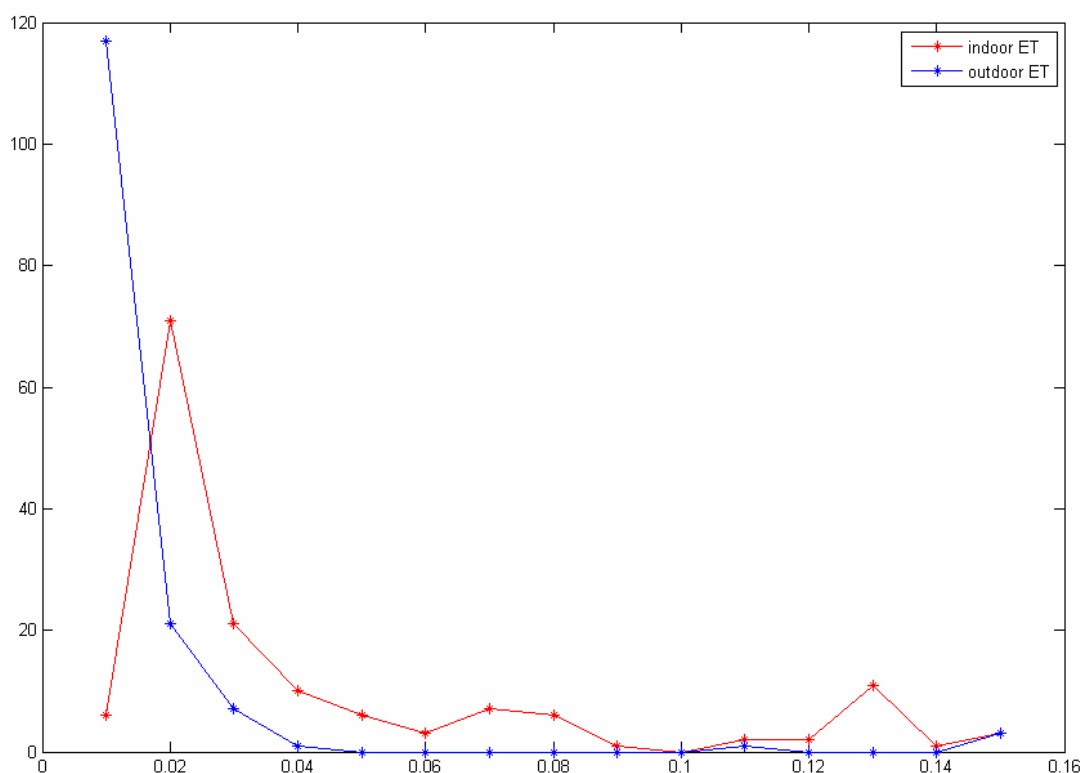


圖 3. 10 曝光時間在室內戶外的分布情形

焦距 (Focal Length)：鏡頭的焦距基本上是指對準無限遠景物時，鏡頭的中軸上清晰影像與菲林平面間的距離。一支鏡頭的焦距，決定被攝主體在菲

林上所形成影像的大小。焦距愈長，所形成的影像便愈大。以一支 100mm 的中距鏡頭為例，它對準無限遠時所形成的影像比在同一部相機上以一支 50mm 的標準鏡頭對準無限遠時所形成的影像要大兩倍，而一支 200mm 的鏡頭，又比一支 100mm 鏡頭所形成的影像大兩倍。目前，焦點距離習慣以 mm(毫米) 計算。

當拍攝的距離不同時焦距的結果也不同，拍攝的主體越遠，則其焦距越大。目前一般市面上的相機在自動調整焦距的情況下，不同廠牌的相機其最低焦距也有所不同。例如像新力 (sony) 的最低焦距為 7.9mm (型號為 W1)，國際牌 (panasonic) 的最低焦距為 5.8mm (型號為 FX7)，佳能 (canon) 的最低焦距為 5.8mm (型號為 E3100)，富士 (FUJIFILM) 的最低焦距為 5.7mm (型號為 FinePix F410)，柯達 (kodak) 的最低焦距為 8mm (型號為 DX4530)，如以 7.9mm 為基準分成 2 類，如表 3.2 所示。室內的相片焦距多為 7.9mm 以下，表示大多為近距離的拍照。

表 3.2 焦距在室外戶外的統計

類別	7.9mm 以下 (包含 7.9mm)	7.9mm 以上
室內	117 張	33 張
戶外	87 張	63 張

閃光燈 (Flash Fired)：閃光燈的主要任務，就是用來補足曝光不足之處的，只要相機的光圈快門可以應付曝光的需要，那也不一定需要閃光燈。許多數位相機有內置的閃光燈。通常有四檔設置：閃光、不閃光、防紅眼閃光和自動閃光，最後一檔是由相機來確定是否需要額外的光線。

由於室內戶外的亮度有所不同，在室內拍攝時，由於環境很暗，因此需要利用閃光燈來加以輔助，所以室內的閃光燈使用頻率比戶外來的高，其結果如表 3.3 所示。

表 3.3 閃光燈在室內戶外使用情形

分類	沒閃光燈	有閃光燈
室內	55 張	95 張
戶外	135 張	15 張

3.6 類神經網路原理

類神經網路 (artificial neural network) 是指模仿生物神經網路的資訊處理系統。較精確的定義為：「類神經網路是一種計算系統，包括軟體與硬體，它使用大量簡單的相連人工神經元模仿生物神經網路的能力。人工神經原是生物神經元的簡單模擬，他從外界環境或者其他人工神經元取的資訊，必加以非常簡單的運算，並輸出其結果到外界環境或者其他人工神經元。」

生物神經網路是由巨量的神經細胞 (neuron)，或稱神經元所組成，神經細胞的形狀與一般的細胞有很大的不同，如圖 3.11 所示，它包括：

1. 神經核 (soma)：核狀的處理系統。
2. 軸索 (神經軸) (axon)：軸索狀的傳輸系統。
3. 樹突 (神經樹) (dendrites)：樹枝狀的輸出入系統。
4. 突觸 (神經節) (synapse)：點狀的連結系統。

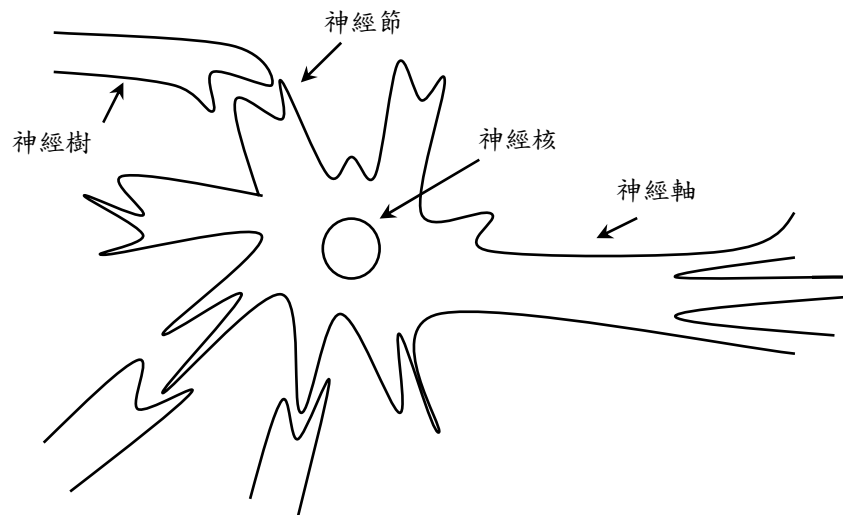


圖 3.11 生物神經元模型

運作過程為由其他神經細胞接受脈波訊號後，神經細胞透過神經樹與神經節，再經過神經核處理，產生一個新的脈波訊號，傳送到所連接的下層神經細胞，一直反覆此動作直至截止。而藉由許多的人工神經細胞（artificial neuron）相連將組成神經網路，人工神經細胞又稱類神經元、處理單元（processing element），如圖 3.12 圖所示，每一個處理單元的輸出以扇形送出，成為其他許多處理單元的輸入。處理單元其輸出值與輸入值的關係式，一般可用輸入值的加權乘積和來表示，如方程式（3.8）。

$$Y_j = f\left(\sum_i W_{ij} X_i - \theta_j\right) \quad (3.8)$$

Y_j ：輸出訊號。

f ：轉換函數（transfer function），是一個將處理單元輸入值的加權乘積和轉換成輸出值的式子。

W_{ij} ：神經節強度，又稱為連結加權值。

X_i ：輸入訊號。

θ_j ：閾值。

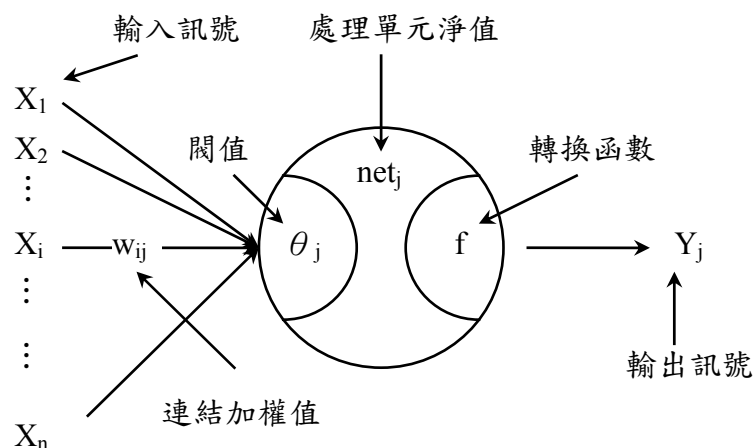


圖 3.12 人工神經元模型

每一個連結上有一個數值的加權值 W_{ij} ，用以表示第 $n-1$ 層的第 i 個處理單元對第 n 層的第 j 個處理單元之影響強度。

常用的轉換函數包括下列幾種：

1. 硬極限轉換函數 (hardlimit transfer function)。
2. 對數雙彎曲轉換函數 (log sigmoid transfer function)。
3. 正切雙彎曲轉換函數 (hyperbolic tangent sigmoid transfer function)。
4. 線性轉換函數 (linear transfer function)。
5. 正線性轉換函數 (positive linear transfer function)。
6. 飽和線性轉換函數 (satlin transfer function)。

在本篇論文中將採用對數正切雙彎曲轉換函數 (hyperbolic tangent sigmoid transfer function)，如圖 3.13 所示；與對數雙彎曲轉換函數 (log sigmoid transfer function)，如圖 3.14 所示兩種轉換函數。以下將大略的介紹。

1. 正切雙彎曲轉換函數 (hyperbolic tangent sigmoid transfer function)

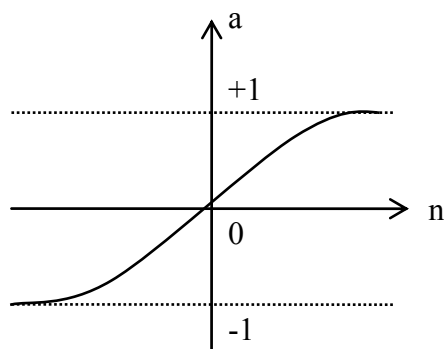


圖 3.13 正切雙彎曲轉換函數

$$Y_j = \frac{1}{1 + e^{-net_j}} \quad (3.9)$$

2. 對數雙彎曲轉換函數 (log sigmoid transfer function)

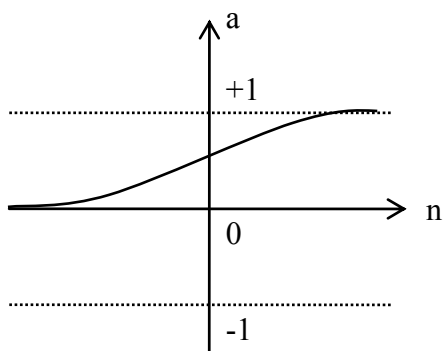


圖 3.14 對數雙彎曲轉換函數

$$Y_j = \frac{e^{net_j} - e^{-net_j}}{e^{net_j} + e^{-net_j}} \quad (3.10)$$

一個類神經網路是由許多個神經元與其連結所組成，並且可以組成各種網路模式 (network model)。其中倒傳遞類神經網路 (Back-propagation neural network, BPN) 模式是目前類神經網路學習模式中最具代表性，應用最普遍的模式。有許多不同的應用實例[34,35]中都用在判別分類上。例如使用倒傳遞類神經網路 error learning 來偵測臉部[36]，利用網路來分類以提升圖片的搜尋率[37,38]，應用在生醫上用來辨別圖片中是否有某種疾病的存在[39,40,41]。

在本論文中，我們利用倒傳遞類神經網路(Back-propagation neural network)來訓練從相片中取出的特徵值，經過多次的訓練來修正權重 (weight)，最後可辨別出結果為室內或戶外。

倒傳遞類神經網路基本原理是利用最陡坡降法(the gradient steepest descent method)的觀念，將誤差函數達到最小化。與感知機(perceptron)做比較，倒傳遞網路作了以下改進：

1. 增加隱藏層，使得網路可表現輸入處理單元間的交互影響。
2. 改用平滑可微分的轉換函數，使得網路可應用最陡坡降法導出修正網路加權值的公式。

倒傳遞類神經網路 (Back-propagation neural network, BPN) 屬於監督式學習網路，包含了輸入層、隱藏層及輸出層，在網路中某一個神經元的的基本架構，每一個神經元的輸出，都乘上其相對應的加權連結值 (weights) 再加總，此相加的總和稱為神經元的淨輸入 (相當於下一層的輸入 $W_i \times$ 上一層輸出)，網路中的每一個神經元都必須先產生淨輸入。

1. 輸入層 (Input Layer)：用以表示網路的輸入變數，其處理單元數目依問題而定，使用線性轉換函數 $f(x) = x$ 。輸入層通常僅用來輸入資料，不對資料做任何處理。
2. 隱藏層 (Hidden Layer)：用以表示輸入處理單元間的交互影響，由數個隱藏層處理單元組成，其單元數並沒有標準方法可以決定，經常需以試驗方式來決定其最佳數目。隱藏層的網路結構，可以有不只一層的隱藏層，也可沒有隱藏層。
3. 輸出層 (Output Layer)：用以表示網路的輸出變數，使用非線性轉換函數 (Nonlinear Transfer Function)。當輸出層單元為單一處理單元時，

通常用於映射輸入值與輸出值間的函數關係，若輸出單元為多個處理單元時，通常用於表示分類問題中每一類別的類別標的。

通常隱藏層定義的數量愈多收斂速度愈慢，但可以達到較小的誤差值，但超過一定的數量後，再增加數量對減少誤差值幾乎沒有多大的幫助，只會增加執行時間。因此為了取得適當的隱藏層數量，其定義可如方程式（3.11）、方程式（3.12）所示。

$$\text{隱藏層數量} = (\text{輸入數量} + \text{輸出數量}) / 2 \quad (3.11)$$

$$\text{隱藏層數量} = (\text{輸入數量} \times \text{輸出數量})^{1/2} \quad (3.12)$$

通常隱藏層層數為一到二層時可以達到最好的收斂效果，太多層或太少層其收斂結果都不好。因為沒有隱藏層將無法與輸入層有交互作用，因而產生較大的誤差，而隱藏層太多會使網路過度複雜，無法收斂。

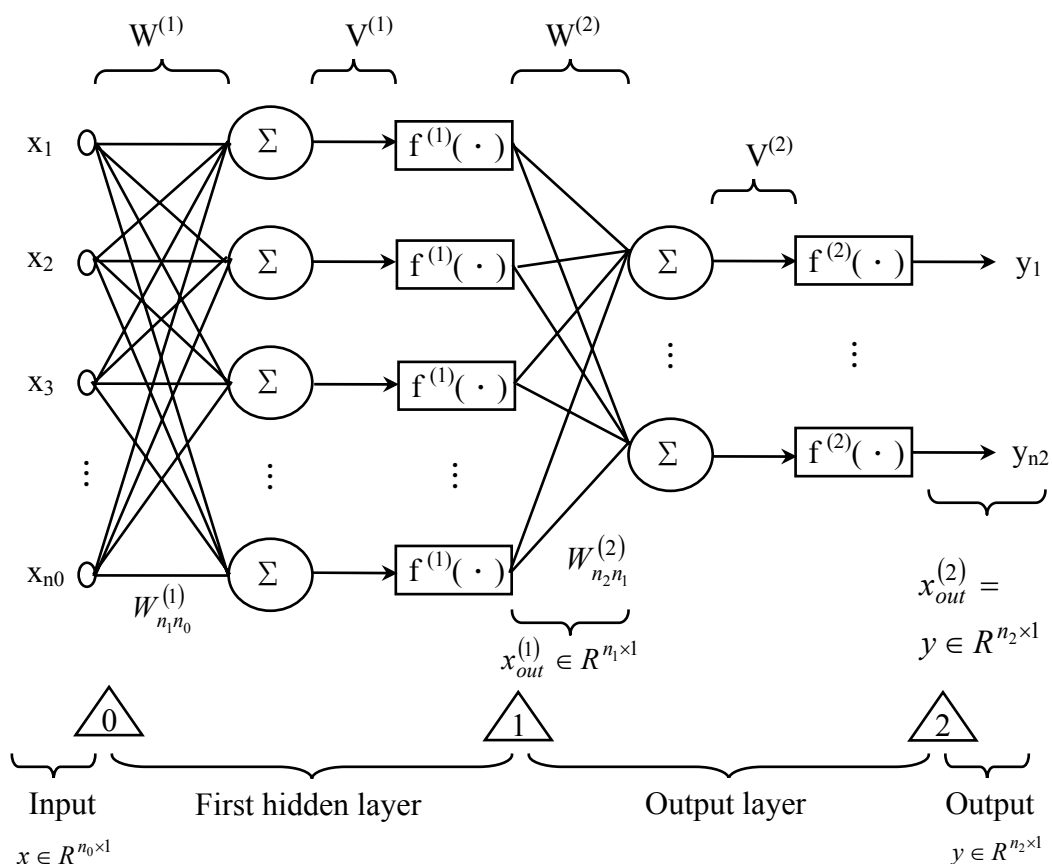


圖 3. 15 Back-propagation neural network 架構

對於多層類神經網路，如圖 3. 15 所示，每一層的輸出將會成為下一層的輸入，如方程式 (3.13)。

$$x_{out}^{(s)} = w^{(s)} \times x_{out}^{(s-1)} \quad (3.13)$$

s：網路中層的數量

倒傳遞類神經網路使用均方誤差做為性能指標，了解訓練出的結果與實際結果的差異，調整網路參數使均方誤差能達到最小，如方程式 (3.14)。

$$E_q = \frac{1}{2} (d_q - x_{out,q}^s)^T (d_q - x_{out,q}^s) \quad (3.14)$$

d_q：網路中第 q 個輸入值所對應的輸出值

使用最陡坡降法（steepest descent），其動態加權連結值（weights）更新如方程式（3.15）。

$$w_{ij}^{(s)}(k+1) = w_{ij}^{(s)}(k) - \alpha^{(s)} \frac{\partial E_q}{\partial w_{ij}^{(s)}(k)} \quad (3.15)$$

$\alpha^{(s)}$ ：第 s 層的學習速率

使用微積分中的連鎖律將方程式（3.14）微分，如方程式（3.16）。

$$\frac{\partial E_q}{\partial w_{ij}^{(s)}} = \frac{\partial E_q}{\partial v_i^{(s)}} \times \frac{\partial v_i^{(s)}}{\partial w_{ij}^{(s)}} \quad (3.16)$$

方程式（3.16）最右邊的分式可改寫如方程式（3.17）。

$$\frac{\partial u_i^{(s)}}{\partial w_{ij}^{(s)}} = \frac{\partial}{\partial w_{ij}^{(s)}} \left(\sum_{h=1}^{n_{s-1}} w_{ih}^{(s)} x_{out,h}^{(s-1)} \right) = x_{out,j}^{(s-1)} \quad (3.17)$$

方程式（3.16）中等號右邊的分式，通常被稱為靈敏度，如方程式（3.18）。

$$\delta_i^{(s)} \triangleq - \frac{\partial E_q}{\partial v_i^{(s)}} \quad (3.18)$$

使用方程式（3.16）、（3.17）與（3.18）改寫方程式（3.15）成為方程式（3.19）。

$$w_{ij}^{(s)}(k+1) = w_{ij}^{(s)}(k) + \alpha^{(s)} \delta_i^{(s)} x_{out,j}^{(s-1)} \quad (3.19)$$

將方程式（3.19）以矩陣的方式表示，如方程式（3.20）。

$$W^{(s)}(k+1) = W^{(s)}(k) + \alpha^{(s)} D^{(s)} x_{out}^{(s-1)} \quad (3.20)$$

$D^{(s)}$ 表示為第 s 層的靈敏度，定義如方程式（3.21）

$$D^{(s)} = - \left[\frac{\partial E_q}{\partial v_1^{(s)}}, \frac{\partial E_q}{\partial v_2^{(s)}}, \dots, \frac{\partial E_q}{\partial v_{n_s}^{(s)}} \right]^T \quad (3.21)$$

將方程式 (3.21) 使用連鎖律可改寫第 s-1 層的靈敏度，如方程式 (3.22)。

$$D^{(s-1)} = - \frac{\partial E_q}{\partial v^{(s)}} = - \left(\frac{\partial v^{(s)}}{\partial v^{(s-1)}} \right)^T \frac{\partial E_q}{\partial v^{(s)}} = \left(\frac{\partial v^{(s)}}{\partial v^{(s-1)}} \right)^T D^{(s)} \quad (3.22)$$

我們可使用 Jacobian 矩陣來推導方程式 (3.22) 兩個靈敏度間的關係，如方程式 (3.23)。

$$\left(\frac{\partial v^{(s)}}{\partial v^{(s-1)}} \right)^T = \begin{bmatrix} \frac{\partial v_1^{(s)}}{\partial v_1^{(s-1)}} & \frac{\partial v_1^{(s)}}{\partial v_2^{(s-1)}} & \dots & \frac{\partial v_1^{(s)}}{\partial v_{n_{s-1}}^{(s-1)}} \\ \frac{\partial v_2^{(s)}}{\partial v_1^{(s-1)}} & \frac{\partial v_2^{(s)}}{\partial v_2^{(s-1)}} & \dots & \frac{\partial v_2^{(s)}}{\partial v_{n_{s-1}}^{(s-1)}} \\ \dots & \dots & \dots & \dots \\ \frac{\partial v_{n_s}^{(s)}}{\partial v_1^{(s-1)}} & \frac{\partial v_{n_s}^{(s)}}{\partial v_2^{(s-1)}} & \dots & \frac{\partial v_{n_s}^{(s)}}{\partial v_{n_{s-1}}^{(s-1)}} \end{bmatrix}^T \quad (3.23)$$

考慮 Jacobian 矩陣中第 i 個、第 j 個元素，如方程式 (3.24)。

$$\begin{aligned} \frac{\partial v_i^{(s)}}{\partial v_j^{(s-1)}} &= \frac{\partial}{\partial v_j^{(s-1)}} \left[\sum_{h=1}^{n_{s-1}} w_{ih}^{(s)} x_{out,h}^{(s-1)} \right] = w_{ij}^{(s)} \frac{\partial x_{out,j}^{(s-1)}}{\partial v_j^{(s-1)}} \\ &= w_{ij}^{(s)} \frac{\partial f(v_j^{(s-1)})}{\partial v_j^{(s-1)}} = w_{ij}^{(s)} f'(v_j^{(s-1)}) = w_{ij}^{(s)} g(v_j^{(s-1)}) \end{aligned} \quad (3.24)$$

將方程式 (3.24) 帶入方程式 (3.23) 改寫成方程式 (3.25)。

$$\begin{aligned} \left(\frac{\partial v^{(s)}}{\partial v^{(s-1)}} \right)^T &= \left[W^{(s)} G(v^{(s-1)}) \right]^T = G(v^{(s-1)}) W^{(s)T} \\ G(v^{(s-1)}) &= \begin{bmatrix} G(v_1^{(s-1)}) & 0 & \dots & 0 \\ 0 & G(v_2^{(s-1)}) & \dots & 0 \\ \dots & \dots & \dots & 0 \\ 0 & 0 & \dots & G(v_{n_{s-1}}^{(s-1)}) \end{bmatrix} \end{aligned} \quad (3.25)$$

最後結合方程式 (3.22) 與 (3.25)，可得到靈敏度公式，如方程式 (3.26)。

$$D^{(s-1)} = G\left(v^{(s-1)}\right) W^{(s)T} D^{(s)} \quad (3.26)$$

網路運作方式分為訓練和測試兩個階段，訓練階段網路必須藉由誤差的回饋與連結權重的調整以記憶學習資料，訓練範例必須具有輸入向量及輸出向量，而輸入及輸出向量之間存在著一種對應關係，其過程如方程式 (3.27)。方程式 (3.28) 為輸出層與之前階層的靈敏度，方程式 (3.29) 為所有隱藏層與之前階層的靈敏度。在測試過程中，輸入向量經由網路的執行便可得到所對應的輸出結果。在本論文所採用的倒傳遞類神經網路，其架構如圖 3.16 所示。

$$W^{(s)}(k+1) = W^{(s)}(k) + \alpha^{(s)} D^{(s)} x_{out}^{(s-1)T} \quad (3.27)$$

$$D^{(s)} = G\left(v^{(s)}\right) \left(d_q - x_{out}^{(s)}\right) \quad (3.28)$$

$$D^{(s-1)} = G\left(v^{(s-1)}\right) W^{(s)T} D^{(s)} \quad (3.29)$$

$\alpha^{(s)}$ ：學習速率

d_q ：期望輸出值



南台科技大學

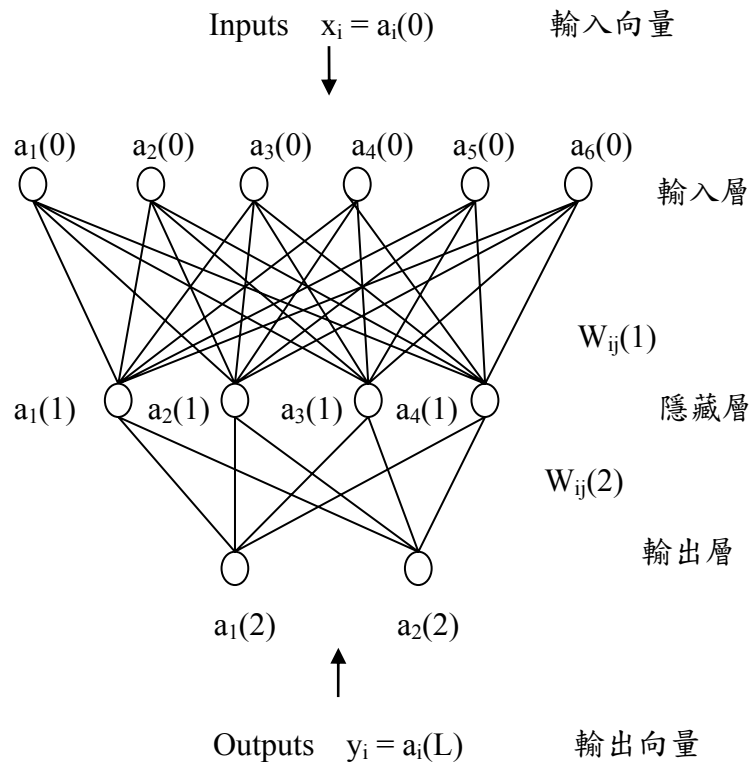


圖 3. 16 類神經網路模型

此學習過程為一次訓練一個訓練範例的方式來進行學習，直至學習完所有的訓練範例，此過程稱為一個學習循環，而網路則是將輸入的結果經過多次的學習循環，直到收斂為止。

我們採用如圖 3. 16 所示之 6 個輸入/3 層架構/2 個輸出之網路；三層架構分別為輸入層、隱藏層與輸出層；六個輸入如圖 3. 17 所示；隱藏層數量由方程式 (3.11) 計算出來，結果為 4 個。訓練階段設定室內的輸出結果為 $\begin{Bmatrix} 1 \\ 0 \end{Bmatrix}$ ，戶外的輸出結果為 $\begin{Bmatrix} 0 \\ 1 \end{Bmatrix}$ 。訓練完後測試階段再使用對照組相片進行測試，測試訓練的目的在於了解此網路是否達到所設定的結果。

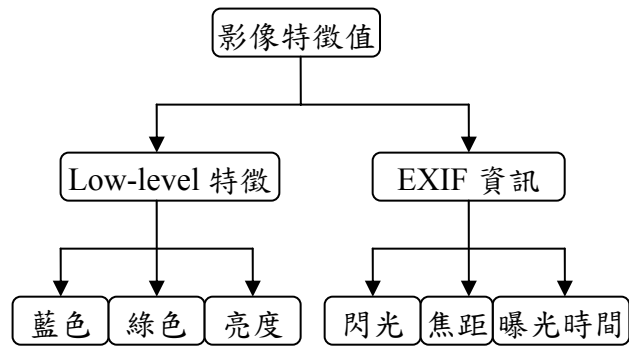


圖 3.17 倒傳遞演算法所使用的特徵值

第四章 建築物/自然風景特徵值擷取與方法

邊緣（edge）對人造物體來說，是一項很重要不可或缺的資訊。因為人造物體所描繪出來的邊緣大都充滿規則性，不會像自然景觀般充滿不規則的線條，如圖 4.1 所示。以建築物為例，最大的特徵就是充滿垂直與水平的線條，因建築物本身的邊緣、窗戶和門大都為四方形所致，而自然景觀大都充滿將近水平或是不規則的線條。因此這些線條、邊緣（edge）就成了為我們判別風景照與建築物照的最大依據。圖 4.2 為自然風景與建築物照的邊緣方向直方圖（edge orientation histogram）。



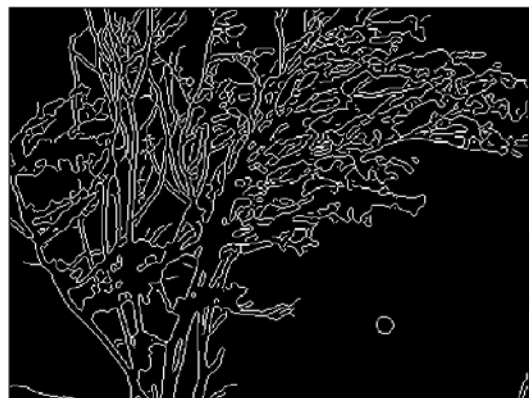
(a)



(b)



(c)



(d)

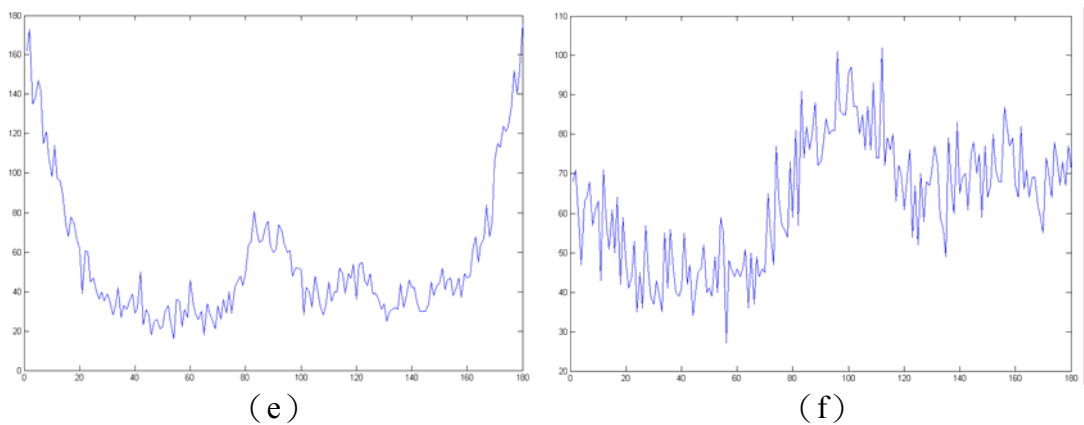
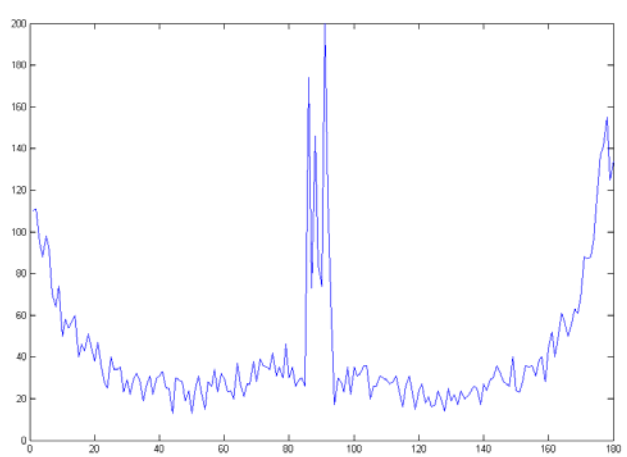


圖 4.1 邊緣比較圖

(a) 人造物體 (b) 自然景物 (c) 人造物體邊緣
(d) 自然景物邊緣 (e) 自然景物邊緣方向直方圖
(f) 自然景物邊緣方向直方圖



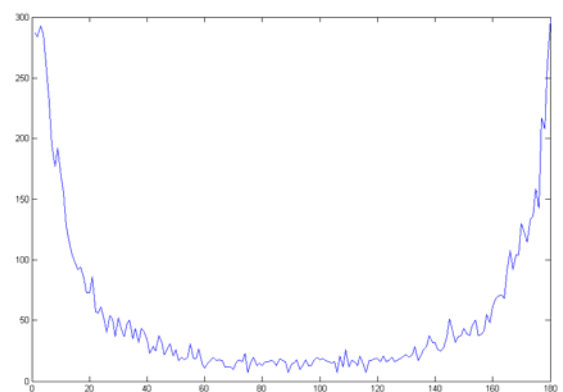
(a)



(b)



(c)



(d)

圖 4.2 邊緣角度直方圖
 (a) 建築物 (b) 邊緣方向直方圖
 (c) 自然風景 (d) 邊緣方向直方圖

4.1 邊緣

邊緣是具有某種影像屬性的細長影像特徵，例如：亮度、色彩、紋理等垂直於圖片邊緣突然的變化。邊緣的偵測對於影像也是相當重要的一環，影像邊緣偵測首度被提出是在 1959 年[32]。之後 Roberts[30]發表第一篇與邊緣偵測有關的 paper，而邊緣偵測的方法大致上可分為一階導函數的局部極值（Sobel filter）與二階導函數的過零點（Laplacian filter）。目前有許多邊緣偵測的方法被提出，如：Sobel edge detector、Canny edge detector、Prewitt edge detector 與 Roberts edge detector 等，各有各的特性與用途，以下將介紹幾個邊緣擷取方法。

4.1.1 Roberts 邊緣偵測

邊緣為影像中鄰近圖素（pixel）值不連續與不平滑的區域集合，邊緣的銳利程度可由影像中圖素（pixel）的梯度（gradient）來決定，梯度 ∇f 可定義如方程式（4.1）。

$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right) \quad (4.1)$$

梯度的大小與方向如方程式（4.2）

$$\begin{aligned} \|\nabla f\| &= \sqrt{\left(\frac{\partial f}{\partial x} \right)^2 + \left(\frac{\partial f}{\partial y} \right)^2} \\ \theta &= \tan^{-1} \left(\frac{\partial f / \partial y}{\partial f / \partial x} \right) \end{aligned} \quad (4.2)$$

Roberts 邊緣[30]偵測是用一個 2×2 的遮罩（mask），對影像進行過濾，

判斷測試的像素為邊線像素或非邊線像素。

1	0
0	-1

0	-1
1	0

Roberts 邊線偵測有兩種形式，可藉由遮罩的水平與垂直向量求得邊線，第一種方式如方程式 (4.3)

$$R(i, j) = \sqrt{(f(i, j) - f(i+1, j+1))^2 + (f(i, j+1) - f(i+1, j))^2} \quad (4.3)$$

第二種方式如方程式 (4.4)，通常都採用此種方式來偵測邊線。

$$R(i, j) = |f(i, j) - f(i+1, j+1)| + |f(i, j+1) - f(i+1, j)| \quad (4.4)$$

4.1.2 Prewitt 邊線偵測

Roberts 邊線偵測方法簡單，運算速度又快，但輸出的結果品質並不是非常的好。因此，1970 年 Prewitt[31] 參考 Roberts 邊線偵測法提出了新的方法。Prewitt 運用兩個具有方向性的 3×3 遮罩 (mask)，一個為水平，另一個則為垂直。

P1	P2	P3
P4	P5	P6
P7	P8	P9

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

P_H

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

P_V

$$\begin{aligned}
 P_H &= -(P1 + P2 + P3) + (P7 + P8 + P9) \\
 P_V &= -(P1 + P4 + P7) + (P3 + P6 + P9) \\
 P &= \sqrt{P_H^2 + P_V^2}
 \end{aligned}
 \tag{4.5}$$

4.1.3 Sobel 邊線偵測

Sobel 邊線偵測法[29]是運用兩個 3×3 大小的遮罩 (mask)，在影像上由左到右，由上到下，對影像進行過濾，判斷測試的像素為邊線像素或非邊線像素。目的在於產生梯度強度影像，使所需處理的資料大為減少，假設預判斷像素 P5 是否為邊線像素， S_A 為列遮罩， S_B 為行遮罩，可透過 S_A 與 S_B 遮罩計算出 S_A 與 S_B 值，利用方程式 (4.6) 可算出 S 值。

S1	S2	S3
S4	S5	S6
S7	S8	S9

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

S_H

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

S_V

$$\begin{aligned}
 S_H &= -(S1 + 2S2 + S3) + (S7 + 2S8 + S9) \\
 S_V &= -(S1 + 2S4 + S7) + (S3 + 2S6 + S9) \\
 S &= \sqrt{S_H^2 + S_V^2}
 \end{aligned}
 \tag{4.6}$$

4.1.4 Canny 邊緣偵測

Canny[33]的偵測方法一共分成四個步驟，步驟如下：

1. 利用高斯濾波器（Gaussian filter）來平滑（smooth）影像；
2. 計算梯度的強度（gradient magnitude）和方向；
3. 利用非最大值抑制演算法（non-maxima suppression algorithm）來尋找梯度的強度；
4. 設置門檻值（threshold）。

Canny edge 主要的原理是使用高斯濾波器（Gaussian）一階導數，其方程式為（4.7），目的是用來降低雜訊所造成的誤差。

$$G_n = \frac{\nabla G}{\partial n} = n \times \nabla G$$

$$n = \frac{\nabla G \times g}{|\nabla G \times g|} \quad (4.7)$$

G：代表高斯濾波器

G_n ：G 在 n 方向的微分

n：梯度（gradient）方向

g：影像函數

經過 smooth 之後產生 J 以水平分量 H 與垂直分量 V 表示，則強度（magnitude）為方程式（4.8）。

$$E_m = \sqrt{H^2 + V^2} \quad (4.8)$$

經由上述找出的邊緣（edge）可能會有重複出現的問題，而使得所得到的邊緣不只 1-pixel 寬，所以使用非最大值抑制演算法（non-maximum suppression algorithm）尋找可能的邊緣位置，先將 E_0 量化為 0° 、 45° 、 90° 與 135° 四個方向，對於每個有非零邊緣量值的像素 M，調查它由邊緣方向決定之兩個鄰近的點，如果兩點的量值皆不大於 M，則將其記為候選。

由 $|\nabla G \times g|$ 計算出邊緣的強度（edge magnitude），最後由設定的門檻值（threshold）剔除掉不必要的候選邊緣，即可求得邊緣圖像（edge map）。

4.2 邊緣特徵值擷取

由圖 4.3 可看出 canny edge 擷取邊緣的效果比 soble edge 及其他兩種的效果來的好的，因此在本篇論文採用 canny edge 來擷取相片的邊緣。

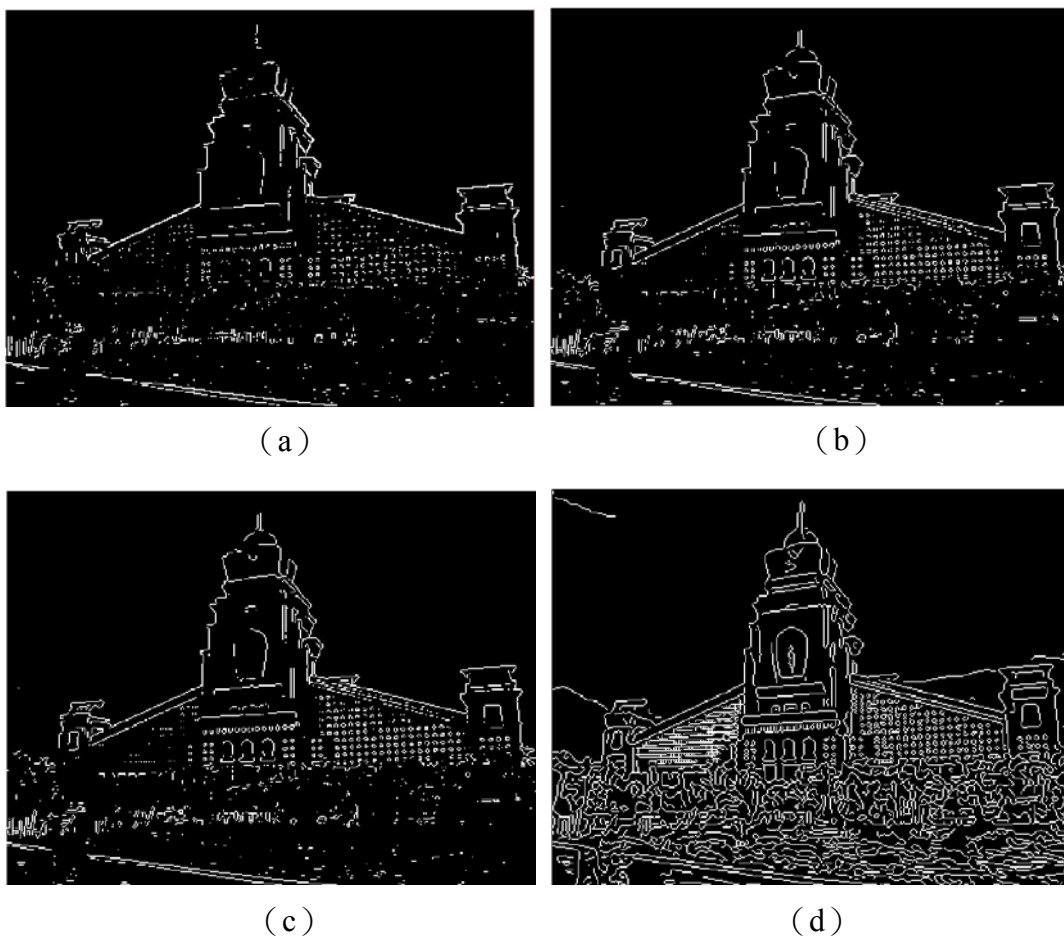
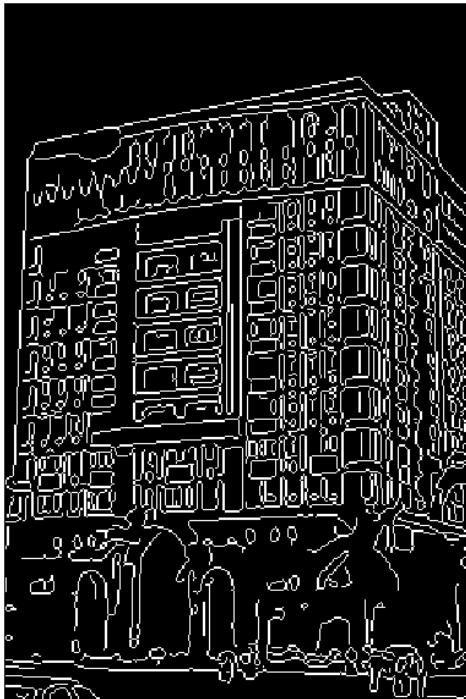


圖 4.3 四種邊緣方法比較圖

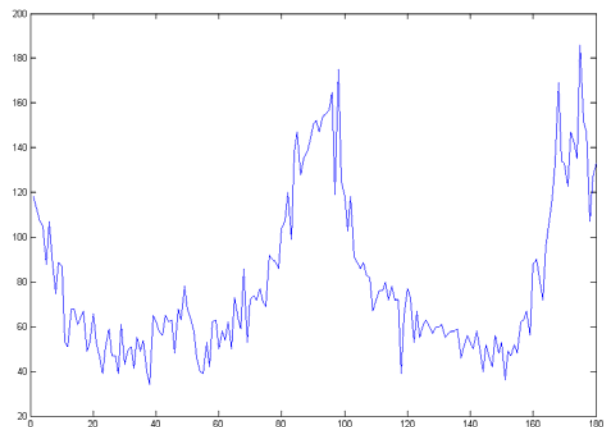
(a) Roberts edge (b) Prewitt edge (c) Sobel edge (d) Canny edge

在 MATLAB 中 canny edge method 可以藉由邊緣的強度 (edge magnitude) 自動決定適當的門檻值 (threshold)，也可由使用者自訂門檻值 (threshold) (條件為 [lowThresh highThresh] 範圍為 0~1 之間， $\text{lowThresh} = 0.4 \times \text{highThresh}$)。為了瞭解 threshold 的設定，對我們所提出建築物邊緣直方圖特有的“W”形狀是否有影響，因此我們針對 threshold 的改變做了個小實驗。圖 4.4 的 (a) 為 MATLAB 自動決定的 threshold [0.031 0.078]、(c) 為我們自訂 threshold [0.01 0.025]、(e) 為我們自訂 threshold [0.048 0.12]、(g) 為我們自訂 threshold [0.08 0.2]。我們可由這個實驗的結果發現 MATLAB 自動決定的 threshold 與人為設定的 threshold，對我們所提出的建築物邊緣直方圖特有的“W”形狀並不會有任何的影響，因此我們將採用 MATLAB 自動決定的 threshold，來計算我們所需

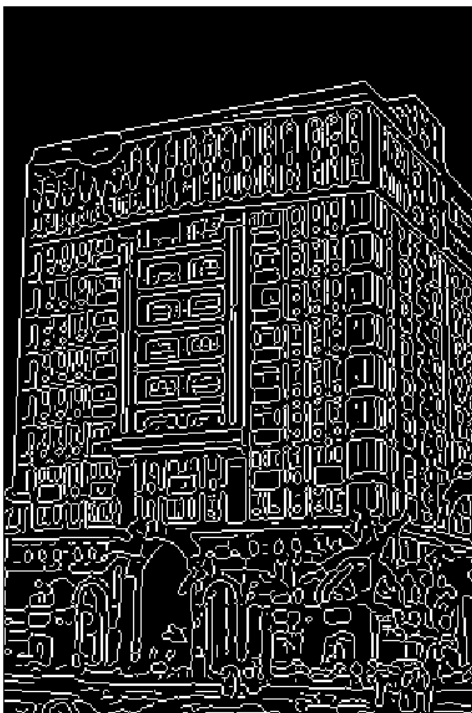
的 edge histogram。



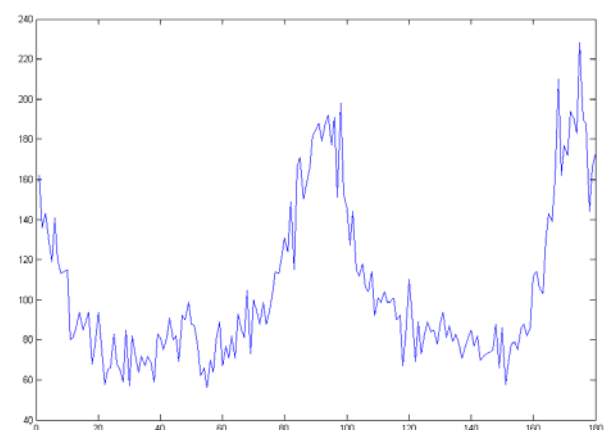
(a)



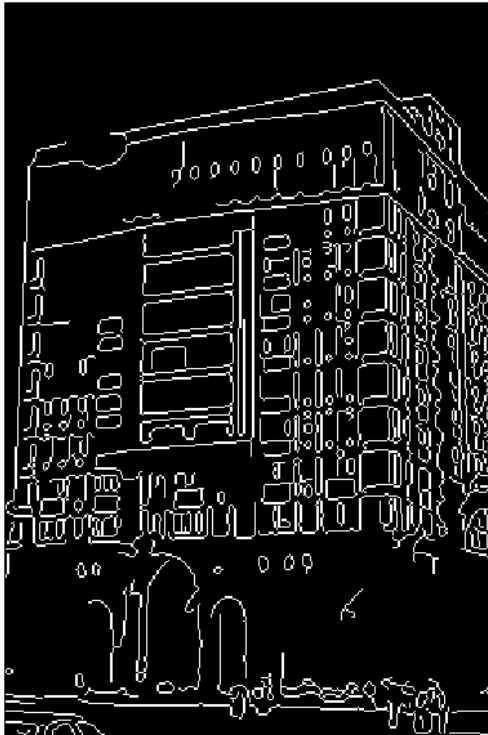
(b)



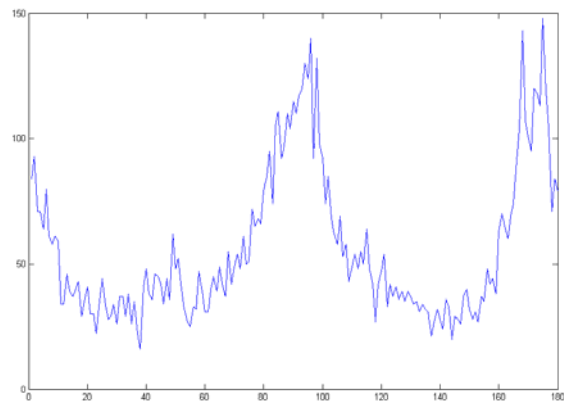
(c)



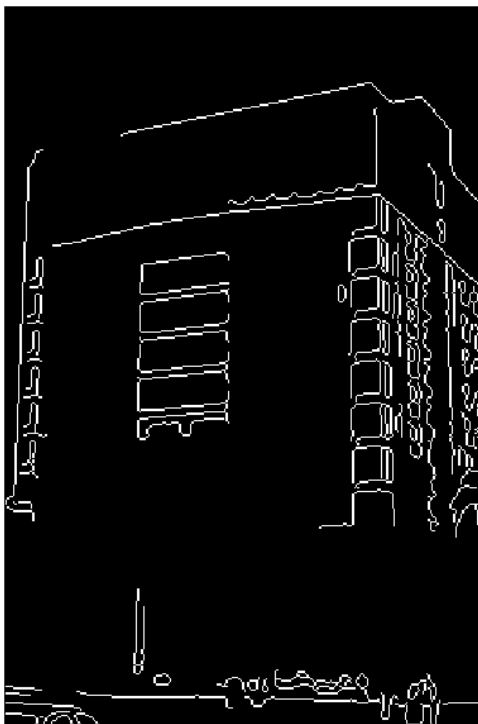
(d)



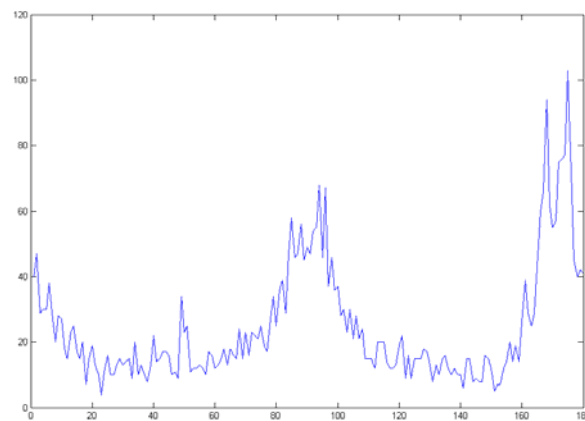
(e)



(f)



(g)



(h)

圖 4.4 canny edge threshold 比較圖

(a) MATLAB 自動決定 threshold (b) 相對應的 edge histogram

(c) 自訂 threshold (d) 相對應的 edge histogram

(e) 自訂 threshold (f) 相對應的 edge histogram

(g) 自訂 threshold (h) 相對應的 edge histogram

確認所用的 threshold 之後，將高斯濾波器平滑後所產生的水平分量與垂直分量利用極座標的方式轉換為我們所需的邊緣角度 (θ)，如圖 4.5 所示，繪製成我們之後所需要的 edge hitogram。

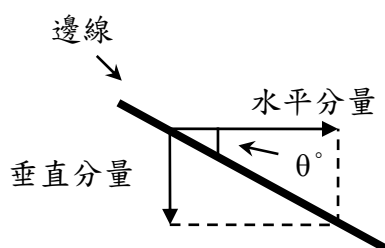


圖 4.5 邊緣方向

建築物的邊緣大都充滿規則性，而且富含垂直與水平線條，如圖 4.8 (b) 所示，因此我們可以把求得的邊緣角度分成主要的四類，如圖 4.6 所示。因此對應到建築物的邊緣直方圖，如圖 4.7 所示，分成五個區間來表示： $p1=[0^\circ-10^\circ]$ 區間的圖素累積數量， $p2=[35^\circ-55^\circ]$ 區間的圖素累積數量， $p3=[80^\circ-100^\circ]$ 區間的圖素累積數量， $p4=[125^\circ-145^\circ]$ 區間的圖素累積數量， $p5=[170^\circ-180^\circ]$ 區間的圖素累積數量。

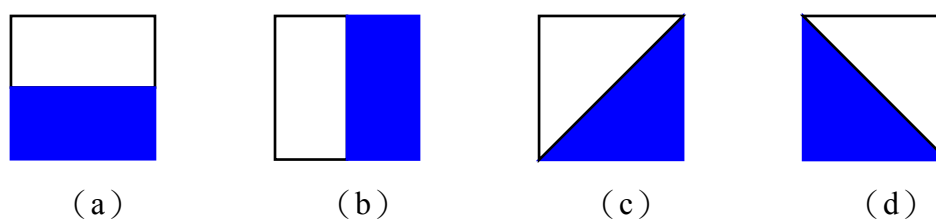


圖 4.6 四種邊緣方向

(a) $\theta \in [0^\circ-10^\circ] \cup [170^\circ-180^\circ]$ (b) $\theta \in [80^\circ-100^\circ]$
(c) $\theta \in [35^\circ-55^\circ]$ (d) $\theta \in [125^\circ-145^\circ]$

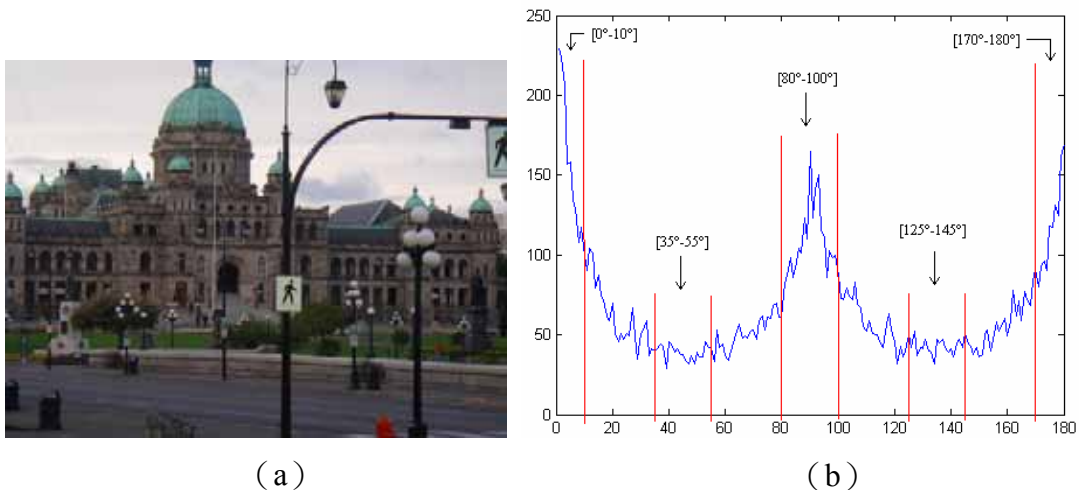


圖 4.7 邊緣方向對應圖
(a) 建築物 (b) 建築物邊緣方向直方圖

4.3 辨別建築物的方法

將判斷是否為建築物的過程分成兩個步驟。首先針對建築物較多的垂直線條作處理，因為建築物雖然充滿垂直的線條，但是自然景物中的樹木也富含垂直線條，如圖 4.8 所示。因此藉由計算出的邊緣角度，訂定 $[80^{\circ}-100^{\circ}]$ 為垂直線段，結果如圖 4.9 (b) 所示，統計 $[80^{\circ}-100^{\circ}]$ 之間的圖素 (pixel) 佔整張相片邊緣的百分比 (α)，隨後再針對先前所訂定的垂直線段作侵蝕 (erosion)，結果如圖 4.9(c)所示，侵蝕的目的在於消除保留垂直線段的雜點與較短線段，其侵蝕的定義如方程式 (4.9)。

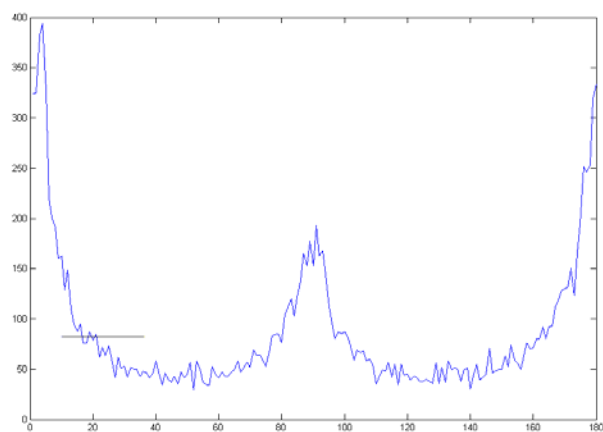
$$A \ominus B = \{z | (B)_z \subseteq A\} \quad (4.9)$$

接著再統計經過侵蝕後剩下的圖素 (pixel) 佔整張相片邊緣的百分比 (β)，若統計過後的百分比大於我們所設定的門檻值 (threshold) α 與 β ($\alpha \geq 0.15$ & $\beta \geq 0.01$)，則辨別為建築物。但是並不是所有的建築物照片都會大於我們所設定的門檻值 (threshold)，也因此可能會產生辨別錯誤的問題，所以我們將在針對不是辨別為建築物的部份，再做更近一步的調整。

先前提到過建築物的特徵大都由水平與垂直線條組合而成，所以在直方圖的 0° 、 90° 與 180° 附近所累積圖素 (pixel) 會比其他角度來的多，因此建築物邊緣直方圖就會產生有如“W”形狀出來，如圖 4.8 (b) 所示。



(a)



(b)



(c) 圖 4.8 角度邊緣直方圖 (d)

(a) 建築物 (b) 建築物邊緣方向直方圖

(c) 自然風景 (d) 自然風景邊緣方向直方圖

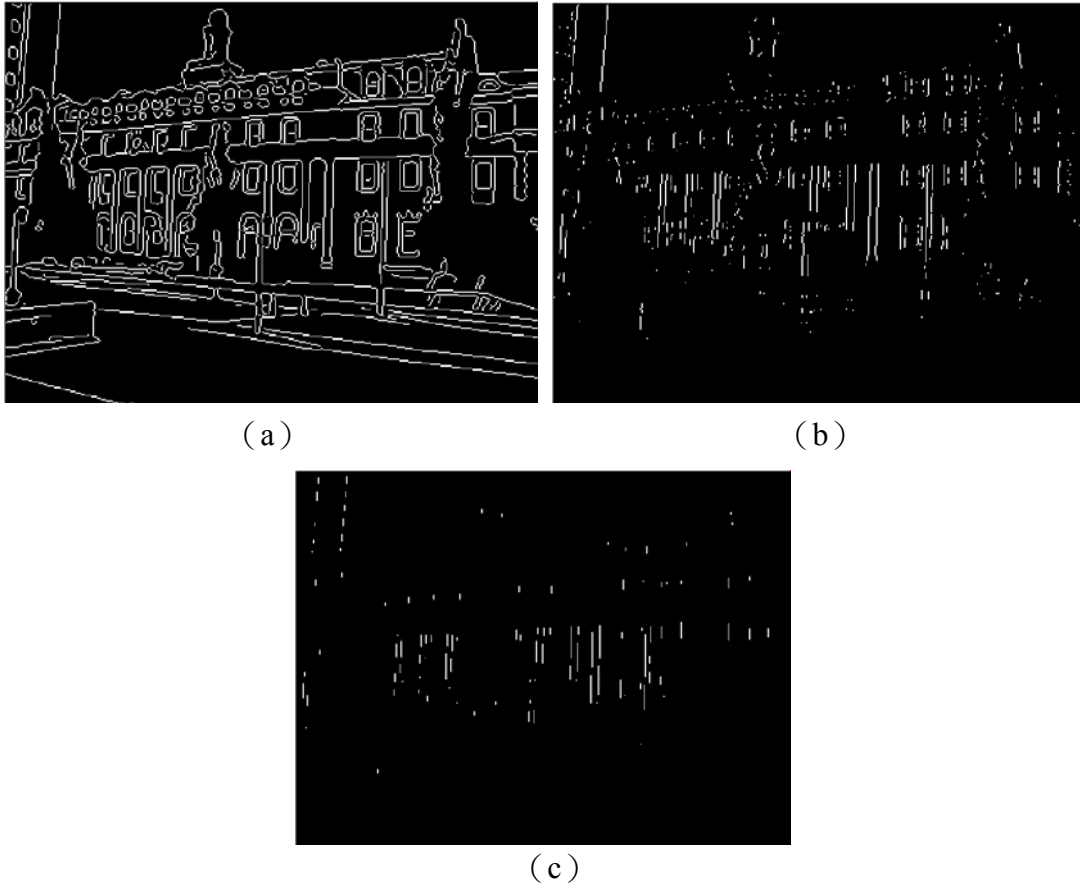


圖 4.9 邊緣圖

(a) 經 canny edge 取出邊緣 (b) 自訂垂直邊緣 (c) 侵蝕後

因此要符合建築物的演算法如下所示：

If sum of vertical line segments $\geq \alpha$ & after erosion $\geq \beta$

This is a Building image.

Else If $p1 > p2$ and $p5 > p4$ and $\text{Max}(p2, p3, p4) = p3$

This is a Building image.

Else

This is a non-building image.

得到判別的結果後，將判別為建築物的相片定位出建築物的所在。其演算

法如下：

1. 將輸入相片量化，如圖 4.10 (b) 所示；
2. 找出先前侵蝕 (erosion) 後的線段集合區域，如圖 4.10 (c) 所示；
3. 利用 k-means clustering 找出區域內最大顏色群集，如圖 4.10 (d) 所示；
4. 依照找出顏色的平均值，設定比對的顏色範圍；
5. 將步驟 4 的結果與步驟 1 的結果作比對，若符合則保留量化後圖片的區塊，圖 4.10 (e) 所示；
6. 取保留區塊的集合組合而成的區域，即為建築物主體，並以框線加以標示，如圖 4.10 (f) 所示。

在此稍微介紹一下步驟3所提及的k-means clustering演算法[42,43]。k-means clustering演算法是分群法中最普遍被應用[44,45,46,47]。利用資料點到群聚中心點的最小平方距離來做分類，經由重複的運算，將多維的特徵空間分類到同性質的類別中。優點為不需大量的輸出節點，因此可大幅度的縮減複雜度。演算法過程如下：

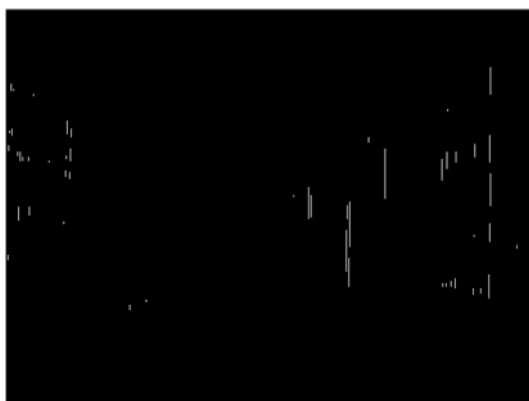
1. 先設定預定要分的群聚數目為A，隨意挑選A個資料當成群聚的中心點C。
2. 新進的資料，利用歐基里德法來計算與各中心點的差距，差距最小者，則將新進的資料分類至該群。
3. 計算後的目標函數若收斂保持不變，則代表分群的結果已經穩定；若沒有，則回到步驟2繼續，直至收斂為止。



(a)



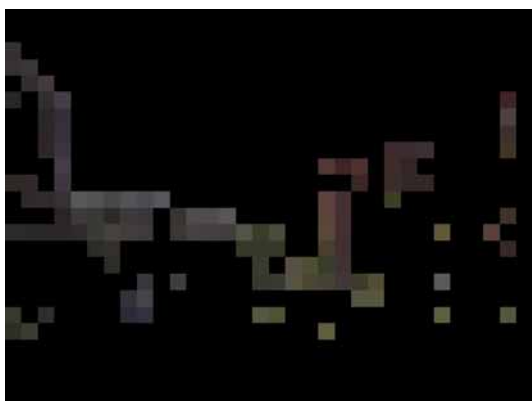
(b)



(c)



(d)



(e)



(f)

圖 4.10 建築物定位過程

(a) 原圖 (b) 量化圖 (c) 侵蝕圖 (d) k-means 結果
(e) 與量化圖比對結果 (f) 建築物定位結果

第五章 實驗結果

5.1 室內/戶外辨別結果

本論文使用倒傳遞類神經網路來做為分類的依據。我們所使用的相片分成三大類，亦即訓練組、對照組與測試組。其中訓練組/對照組用以驗證倒傳遞類神經網路之訓練效度；而測試組相片則用來驗證隨機的相片組合，是否可以達到室內/戶外的辨識效果。

為了驗證倒傳遞類神經網路可以確實辨識出室內/戶外的相片。我們首先從資料庫中選取訓練組與對照組相片各 300 張。訓練組與對照組之相片內容各不相同，然而每組相片均包含 150 張為室內、150 張為戶外相片。而且為了達到實驗結果的可控制性，我們也各自計算出每組相片之 RGB 的色彩標準差和亮度的標準差，結果如表 5.1、表 5.2 所示，標準差公式如方程式 (5.1)。

$$SD = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (5.1)$$

在訓練之前先將各組的相片維度改為 800x600，以利於加快特徵值計算的速度，但這並不會影響該張相片的特徵值的數值，由於特徵值皆以佔圖素的比例來作計算，因此即使放大或縮小相片也不會造成很大的差異。各特徵值的擷取方式如上第三章所說明，將擷取出的特徵值經過前處理正規化為 1 到-1 之間後，再放入倒傳遞類神經網路 (Back-propagation neural network) 進行訓練，重複疊代 40 次之後，確認完全收斂後，以對照組相片進行一致性驗證，確認先前所作之訓練是否正確，其結果如表 5.3 所示。

接下來，我們從資料庫裡隨意的選取 400 張相片（200 張室內、200 張戶外）當成測試組相片。為了驗證辨識的強健性(robustness)，在測試組中，我們甚至選取一些原先並沒有在訓練組和對照組中沒有的相片類型，以測試我們所提出的方法是否真正的適用。其結果如表 5.4 所示，由實驗數據可以證實我們使用方法的效果。

表 5.1 訓練組與對照組相片 RGB 色彩空間標準差

	標準差					
	R		G		B	
	室內	戶外	室內	戶外	室內	戶外
訓練相片	29.65	62.55	27.85	63.32	28.82	64.43
對照相片	30.89	61.78	26.60	65.03	29.83	64.01

表 5.2 訓練組與對照組亮度標準差

	室內亮度標準差	戶外亮度標準差
訓練相片	23.14	61.83
對照相片	24.16	62.54

表 5.3 對照組的相片透過倒傳遞網路輸出的正確率

	亮度	色彩	EXIF	色彩+亮度	亮度+EXIF	色彩+EXIF	亮度+色彩+EXIF
室內	49%	51%	63%	57%	91%	93%	97%
戶外	57%	64%	79%	65%	93%	89%	99%

表 5.4 測試組相片正確率

	正確	錯誤	正確率
室內	167	33	84%
戶外	181	19	91%

表 5.5 測試組的相片透過倒傳遞網路輸出的正確率

	亮度	色彩	EXIF	色彩+亮度	亮度+EXIF	色彩+EXIF	亮度+色彩+EXIF
室內	48%	49%	61%	54%	79%	76%	84%
戶外	43%	60%	77%	64%	84%	86%	91%

我們將部分的實驗結果圖示如圖 5.1、圖 5.2 為實驗組的室內、戶外相片，圖 5.3、圖 5.4 為對照組的室內、戶外相片，圖 5.5、圖 5.6 為在資料庫中隨意選取的測試組室內、戶外相片，其範圍並不包括訓練組與對照組的相片在內。圖 5.7 為測試組相片中判斷錯誤的。

我們嘗試解釋其分類錯誤的原因如下：圖 5.7 左上方與右上方兩張相片，因靠近窗邊，所以相機在抓光源焦點時，可能會以外在光源為主認為光源充足，因此被分類成戶外的相片；圖 5.7 左下方與右下方雖為戶外照片，但因強制把閃光燈關閉，造成相片光源不足，又無明顯戶外特徵（顏色），因此被分類成室內的相片。





圖 5.1 訓練組室內相片



南台科技大學

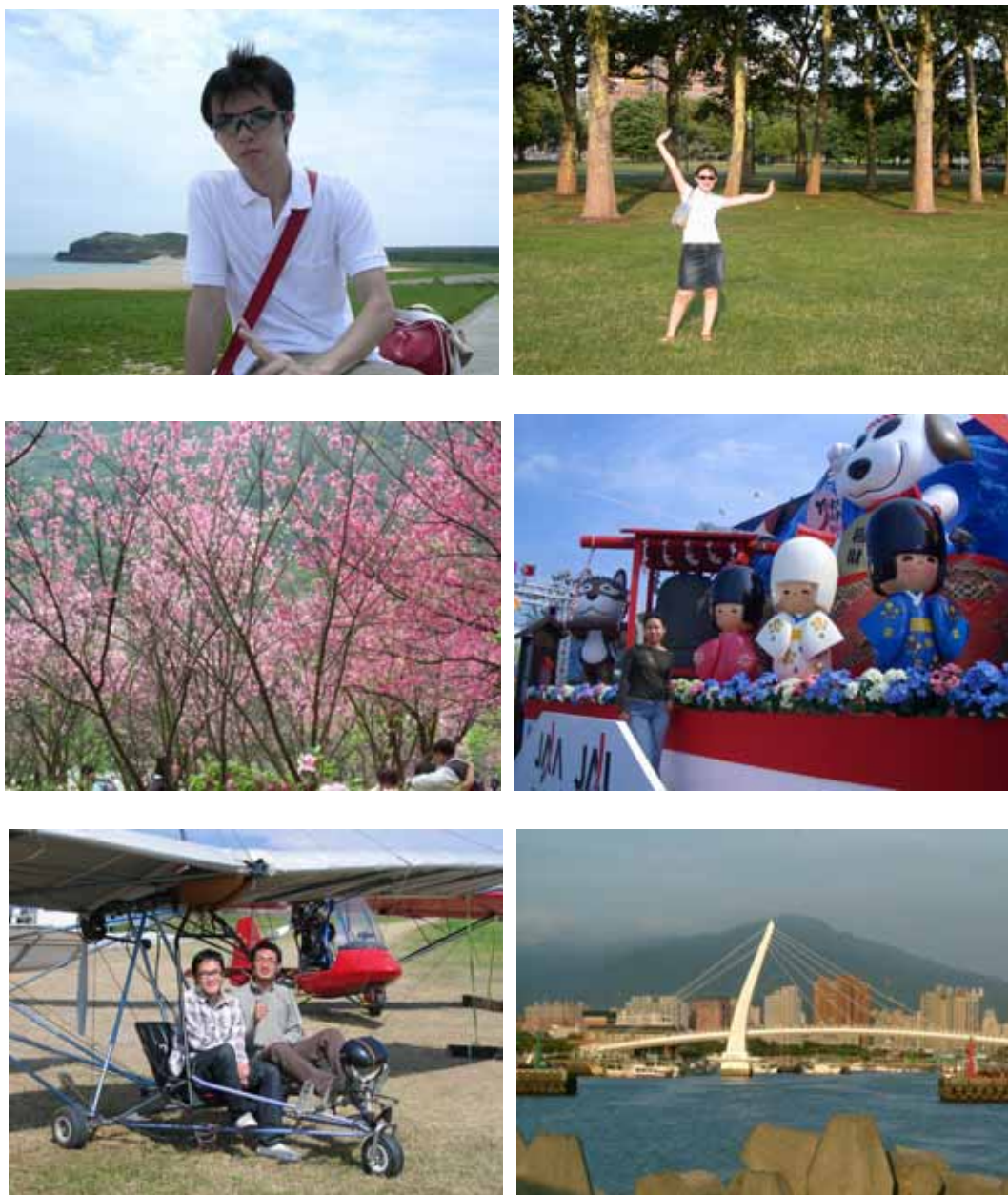


圖 5.2 訓練組戶外相片



南台科技大學



圖 5.3 對照組室內相片



圖 5.4 對照組戶外相片



圖 5.5 測試組室內相片



圖 5.6 測試組戶外相片



圖 5.7 分辨錯誤相片

5.2 建築物/自然風景辨別結果

在本節裡，我們以第四章所介紹的特徵值與方法，來驗證我們所提出的方法。首先，如前一節所描述先將相片維度改為 800×600 ，再藉由第四章所介紹的特徵值與方法，我們由資料庫中隨意選取 200 張相片（100 張建築物照，100 張自然風景照），以測試我們所提出的方法是否真正的適用，其結果如表 5.6 所示，由實驗數據可以證實我們使用方法的效果。

表 5.6 建築物辨別正確率

	正確	錯誤	正確率
室內	82	18	82%
戶外	90	10	90%

我們將結果圖示如圖 5.8 為建築物相片，圖 5.9 為自然風景相片，圖 5.10 辨識錯誤相片。我們嘗試解釋其辨別錯誤的原因如下：圖 5.10 左上方因為相片是由很多的粗大樹木所組成，因此經過侵蝕後仍然存在很強烈的垂直線段訊息，因此超過我們所訂定的門檻值，所以辨別為建築物照，右上方因為相片中包含樹木，樹木本體也是包含豐富的垂直線段，因此原因與左上方相同，所以辨別為建築物照，左下方因為相片中的建築物為“塔”，塔本身建築物所包含的特徵（垂直與水平線段）較少也較短，因此低於我們所設置的門檻值，所以被辨別為自然風景照，右下方的相片中因為建築物是由石頭所建構而成，因此在擷取邊緣時會產生如圖 5.11 的結果，邊緣都呈現不規則形狀，因此辨別為自然風景照。建築物辨識定為結果如圖 5.12 所示。





圖 5.8 建築物相片





圖 5.9 自然風景相片



圖 5.10 辨識錯誤建築物與自然景物相片

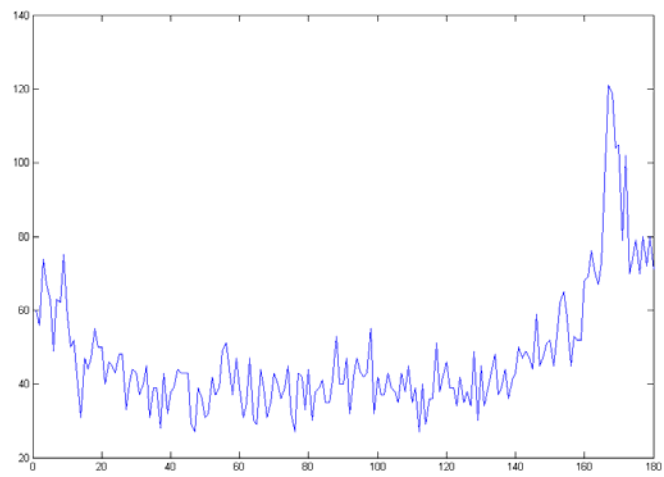
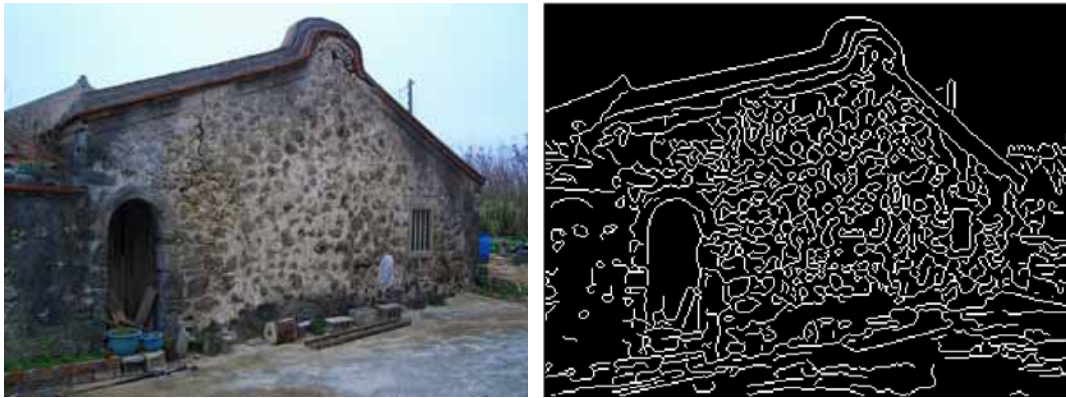


圖 5.11 錯誤辨別結果解釋圖





圖 5.12 建築物辨別定位結果

5.3 系統介面

為了確認是否有達到如表 5. 4 與表 5. 6 所列出的正確率，因此我們採用 MATLAB 撰寫程式，其介面圖如圖 5. 13 所示，圖 5. 14、圖 5. 15 和圖 5. 16 為辨識結果展現方式。

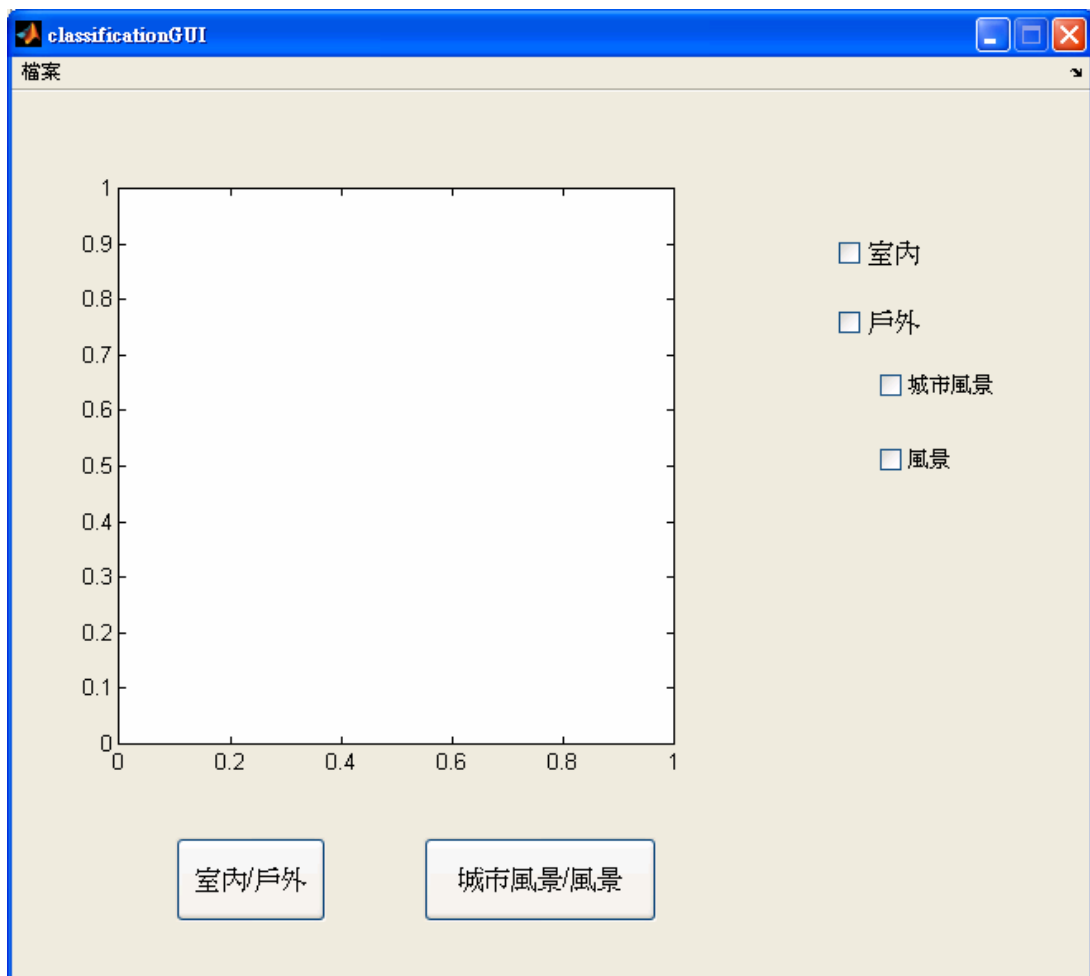


圖 5. 13 系統介面圖



圖 5.14 辨別室內結果介面圖

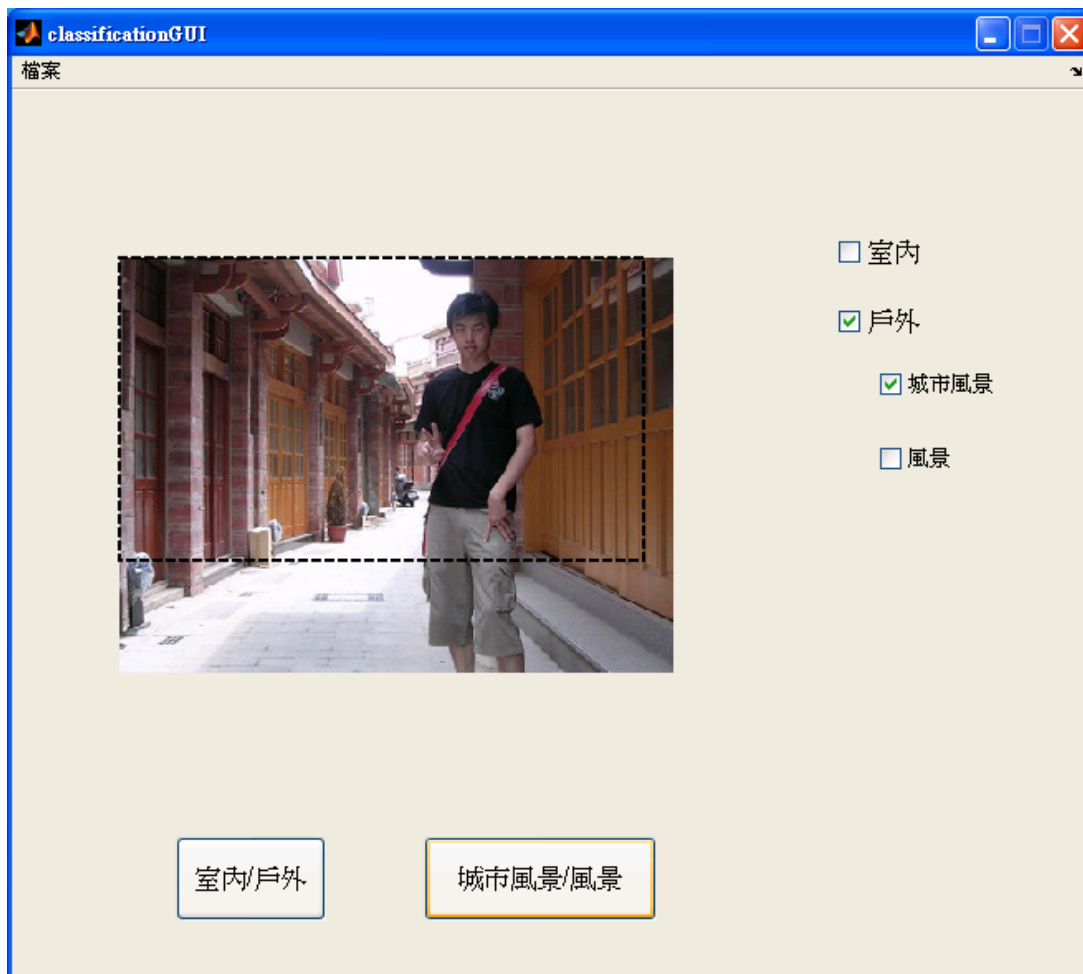


圖 5.15 戶外/建築物辨識結果介面圖



南台科技大學



圖 5.16 戶外/自然風景辨別結果介面圖

第六章 結論與未來工作

在目前商品數位化的時代，很多人都採用相機或是攝影機來記錄生活的過程，因此要如何的管理這些大量的數位相片，就成了我們目前最重要的課題，有鑑於此我們著手室內/戶外與建築物/自然景物的相片分類。

在室內/戶外的相片方面，首先，會想到的特徵值即為色彩。因為戶外的相片很多都是充滿藍天或草地。其次光源也是很重要的特點，因為戶外相片的光源大都比市內來的光亮。接下來再採用數位相片特有的 EXIF 資訊，擷取其中的幾個來作為辨別的依據。特徵值擷取完後再利用倒傳遞網路來訓練，訓練出的結果即為我們的辨別結果。

在建築物/自然景物方面，我們觀察到建築物的邊緣特徵值都較為規律化，而自然景物方面邊緣特徵值大都毫無規則可言，因此我們決定採用邊緣作為我們辨別最主要的特徵值。在尋找辨識建築物的過程中，針對我們所定義的垂直線段作侵蝕，希望能保留較長的線段，作為辨識建築物的第一條件，符合則為建築物；接下來針對第一部分非辨別為建築物的相片，我們藉由邊緣角度所繪製出的直方圖觀察到，直方圖在 0° 、 90° 與 180° 的部份會累積較多的圖素，因為我們先前提到過建築物大都由垂直與水平線段所構成，因此可作為我們辨別建築物的第二個條件，符合則為建築物。之後再藉由 k-means 來定位出建築物的所在。

在室內/戶外辨別方面，未來希望能針對靠窗邊的室內相片，能克服外在光源影響的問題，能更準確的辨別出結果。在戶外相片方面希望能在光源不足卻又刻意關閉閃光燈的情況下，能達到理想的辨識效果。在建築物/自然景物方

面，希望能辨識出垂直與水平線段較不明顯的建築物，建築物的定位能更為準確。

我們可將辨別出的結果儲存於資料庫內，將來以便於系統搜尋，也可提升搜尋的準確率，除此之外，還可在外加其他的註釋分類項目，擴充系統的便利性。由於現今網路發達，因此我們也可將系統網路化，讓使用者可以利用我們的辨別方法，在網路上能很快的尋找到使用者所需要的圖片或相片，增加網路搜尋的便利性。

參考文獻

1. 王朝煌，以區塊的顏色分佈特性為基礎的影像檢索系統，義守大學資訊工程研究所碩士論文，民 92 年。
2. 李宏傑，基於紋理與色彩特徵的影像檢索系統之研究，淡江大學電機工程系碩士論文，民 90 年。
3. 石昭玲，以影像內涵之形狀及色彩為基礎的影像檢索系統，交通大學資訊科學博士論文，民 91 年。
4. 陳志雄，以顏色複雜度與色彩空間分部特徵為基礎的影像查詢系統，朝陽科技大學資訊管理系碩士論文，民 92 年。
5. 曾易聰，利用顏色及材質特徵進行影像檢索及系統設計，台灣大學電信工程學研究所碩士論文，民 89 年。
6. N. Serrano, A. Savakis, and J. Luo, "A Computationally Efficient Approach to Indoor/Outdoor Scene Classification," *In Proc. Int. Conference on Pattern Recognition*, Vol. 4, pp. 146–149, 2002.
7. M. Boutell and J. Luo, "Indoor vs. Outdoor Classification of Consumer Photographs Using Low-Level and Semantic Features," *International Conference on Image Processing*, Vol. 2, pp. 745 – 748, 2001.
8. M.J. Kane and A. Savakis, "Bayesian Network Structure Learning and Inference in Indoor vs. Outdoor Image Classification," *International Conference on Pattern Recognition*, Vol. 2, pp. 479 – 482, 2004.
9. E. Spyrou, H. Le Borgne, T. Mailis, E. Cooke, Y. Avrithis and N. O'Connor, "Fusing MPEG-7 visual descriptors for image classification," *Proc International Conference on Artificial Neural Networks*, pp. 847 – 852, 2005.
10. M. Boutell and J. Luo, "Incorporating Temporal Context with Content for Classifying Image Collections," *International Conference on Pattern*

Recognition, Vol. 2, pp. 947 – 950, 2004.

11. M. Boutell and J. Luo, “Photo Classification by Integrating Image Content and Camera Metadata,” *International Conference on Pattern Recognition*, Vol. 4, pp. 901 – 904, 2004.
12. M. Boutell, J. Luo, and R.T. Gray, “Sunset scene classification using simulated image recomposition,” *International Conference on Multimedia and Expo*, Vol. 1, pp. 37 – 40, 2003.
13. Y. Song and A. Zhang, “Analyzing scenery images by monotonic tree,” *ACM Multimedia Systems Journal*, 2002.
14. M. Szummer and R.W. Picard, “Indoor-outdoor image classification,” *IEEE International Workshop on Content-Based Access of Image and Video Database*, pp. 42 – 51, 1998.
15. A. Vailaya, M. Figueiredo, A. Jain, and H.J. Zhang. “Content-based hierarchical classification of vacation images,” *IEEE International Conference on Multimedia Computing and Systems*, Vol. 1, pp. 518 – 523, 1999.
16. A.C. Loui and A.E. Savakis, “Automatic Image Event Segmentation and Quality Screening for Albuming Applications,” *Proc. Int. Conf. Multimedia and Expo*, 2000.
17. Chang Shih-Fu, W. Chen, H.J. Meng, H. Sundaram, and Zhong Di, “A fully automated content-based video search engine supporting spatiotemporal queries,” *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 8, pp. 602 – 615, 1998.
18. M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Huang Qian, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker, “Query by image and video content: the QBIC system,” *IEEE Computer Magazine*, Vol. 28, pp. 23 – 32, 1995.
19. Li Ze-Nian, R. Zaiane Osmar, and Tauber Zinovi, “Illumination Invariance and Object Model in Content-Based Image and Video Retrieval,” *Journal of Visual Communication and Image Representation*, Vol. 10, pp. 219–244, 1999.

20. R. Smith John, and Chang Shih-Fu, "VisualSEEk: A fully automated content-based image query system," *ACM Multimedia conference*, 1996.
21. M.V. Srinivasan, S. Venkatesh, and R. Hosie, "Qualitative Estimation of Camera Motion Parameters from Video Sequences," *Pattern Recognition*, Vol. 30, pp. 593–606, 1997.
22. Zhang Yue, M.A. Nascimento, and O.R. Zaiane, "Building image mosaics: an application of content-based image retrieval," *International Conference on Multimedia and Expo*, Vol. 3, pp. 317–320, 2003.
23. Rui Yong, S. Huang Thomas, and Chang Shih-Fu, "Image Retrieval: Past, Present, and Future," *International Symposium on Multimedia Information Processing*, 1997.
24. Jr. John M. Zachary, "An Information Theoretic Approach to Content Based Image Retrieval," *Louisiana State University*, 2000.
25. A. Vailaya, A. Jain and H.J. Zhang, "On Image Classification: City Images vs. Landscapes," *Pattern Recognition*, pp. 1921–1936, 1998.
26. A. Dorado and E. Izquierdo, "Exploiting Problem Domain Knowledge for Accurate Building Image Classification," *Proceedings of CIVR 04*, 2004.
27. Q. Iqbal and J.K. Aggarwal, "Applying Perceptual Grouping to Content-based Image Retrieval: Building Images," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 1, 1999.
28. Q. Iqbal and J.K. Aggarwal, "Lower-level and Higher-level Approaches to Content-based Image Retrieval," *IEEE South West Symposium on Image Analysis and Interpretation*, pp. 197–201, 2000.
29. I.E. Sobel, "Camera Models and Machine Perception," *Ph.D. thesis, Stanford University*, pp. 99, 1970.
30. L.G. Roberts, "Machine Perception of Three-dimensional Solids," *Optical and Electro-Optical Information Processing*, pp. 159–197, 1965.

31. J. Prewitt, "Object Enhancement and Extraction," *Picture Process. Psychopict.*, pp.75–149, 1970.
32. B. Julesz, "A Method of Coding TV Signals Based on Edge Detection," *Bell System Tech.*, Vol. 38, pp. 1001–1020, 1959.
33. J. Canny, "A Computational Approach to Edge Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 8, pp. 679–689, 1986.
34. S. Karungaru, M. Fukumi, and N. Akamatsu, "Neural networks and genetic algorithms for learning the scene illumination in color images," *IEEE International Symposium on Computational Intelligence in Robotics and Automation*, Vol. 3, pp.1085–1089, 2003.
35. N. Visen, J. Paliwal, and D.S. Jayas, "Algorithm development for grain kernel identification," *IEEE Canadian Conference on Electrical and Computer Engineering*, Vol. 2, pp.963–967, 2002.
36. J. Nakamura, T. Ikeda, N. Inui, and Y. Kotani, "Learning face marks for natural language dialogue systems," *International Conference on Natural Language Processing and Knowledge Engineering*, pp. 180–185, 2003.
37. M.A. Ibrahim, M.K Arora, S.K Ghosh, and H. Chen, "Approaches to improve accuracy of neural network classification of images dominated by mixed pixels," *IEEE International Geoscience and Remote Sensing Symposium*, Vol. 1, 2004.
38. Han Min, Cheng Lei, and Meng Hua, "Classification of aerial photograph using neural network," *International Conference on Systems*, Vol. 6, pp.6, 2002.
39. K.V. Ramana and S. Khader Basha, "Neural image recognition system with application to tuberculosis detection," *International Conference on Information Technology: Coding and Computing*, Vol. 2, pp. 694–698, 2004.
40. N. Srinivasan, M.T. Wong, and S.M. Krishnan, "A new phase space analysis algorithm for cardiac arrhythmia detection," *International Conference of the IEEE Engineering in Medicine and Biology Society*, Vol. 1, pp. 82–85, 2003.

41. N.A. Wahab, M.A. Wahed, and A.S.A. Mohamed, "Texture Features Neural Classifier of Some Skin Diseases," *IEEE International Midwest Symposium on Circuits and Systems*, Vol. 1, pp. 380–382, 2003.
42. S. Z. Selim and M. A. Ismail, "K-means-type Algorithms: A Generalized Convergence Theorem and Characterization of Local Optimality," *IEEE Trans. Pattern Anal. Machine Intell.*, pp. 81–87, 1984.
43. C. Darken and J. Moody, "Fast Adaptive K-means Clustering: Some Empirical Results," *Proceedings of the International Joint Conference on Neural Networks*, pp. 233–238, 1990.
44. C. Chinrungrueng and C. H. Sequin, "Optimal Adaptive K-means Algorithm with Dynamic Adjustment of Learning Rate," *IEEE Transactions on Neural Networks*, Vol. 6, pp. 157–169, 1995.
45. M. Sarkar and B. Yegnanarayana, "A Clustering Algorithm Using Evolutionary Programming," *IEEE International Conference on Neural Networks*, Vol. 2, pp. 1162–1167, 1996.
46. D. Lee, S. Back, and K. Sung, "Modified K-means Algorithm for Vector Quantizer Design," *IEEE Signal Processing Letters*, Vol. 4, pp. 2–4, 1997.
47. K. Krishna and M. N. Murty, "Genertic K-means Algorithm," *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, Vol. 29, pp. 433–439, 1999.
48. ISO/IEC 15 938-3: MPEG-7 Visual, 2002-03-10
49. M.Abdel-Mottaleb and S. Krishnamachari, "Multimedia descriptions based on MPEG-7: extraction and applications," *IEEE Transactions on Multimedia*, Vol. 6, pp. 459 – 468, 2004.
50. T. Sikora, "The MPEG-7 visual standard for content description-an overview," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 11, pp. 696 – 702, 2001.
51. Chang Shih-Fu, T. Sikora, A. Purl, "Overview of the MPEG-7 standard," *IEEE*

Transactions on Circuits and Systems for Video Technology, Vol. 11, pp. 688 – 695, 2001.

52. B. S. Manjunath, J. R. Ohm, V. V. Vasudevan and A. Yamada, “Color and texture descriptors,” *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 11, pp. 703 – 715, 2001.
53. M. Grgic, M. Ghanbari and S. Grgic, “Texture-based image retrieval in MPEG-7 multimedia system,” *International Conference on Trends in Communications*, Vol. 2, pp. 365 – 368, 2001.
54. I. Beinlich, H. Suermondt, R. Chavez, and G. Cooper, “The alarm monitoring system: A case study with two probabilistic inference techniques for belief networks,” *In Proceedings of the Second European Conference on Artificial Intelligence in Medical Care*, pp. 247–256, 1989.
55. G. F. Cooper and E. Herskovits, “A Bayesian method for the induction of probabilistic networks from data,” *Machine Learning*, pp. 309–347, 1992.
56. M.A. Hearst, S.T. Dumais, E. Osman, J. Platt and B. Scholkopf, “Support vector machines,” *IEEE Intelligent Systems and Their Applications*, Vol. 13, pp. 18–28, 1998.
57. J. Luo, A. Savakis, S. Etz, and A. Singhal, “On the Application of Bayes Networks to Semantic Understanding of Consumer Photographs,” *Proc. ICIP 2000*.
58. F. Jing, M. Li, H.-J. Zhang, and B. Zhang “Support vector machines for region-based image retrieval,” *International Conference on Multimedia and Expo*, Vol. 2, pp. 21–24, 2003.
59. J. Hafner, H. S. Sawhney, W. Equitz, M. Flickner, and W. Niblack, “Efficient color histogram indexing for quadratic form distance functions,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 17, pp. 729–736, 1995.