

國立政治大學統計學研究所

碩士學位論文

貓狗影像辨識之特徵萃取

Feature extraction in dogs and cats image recognition



指導教授：薛慧敏 博士

研究生：鍾立強 撰

中華民國 一百零五 年 七 月

摘 要

近年來，很多要求高安全性的網站都使用扭曲變形的英文或數字字串作為辨識碼，以避免網站或系統受到大量暴力的攻擊。微軟公司則於 2007 年提出以貓狗影像的新辨識碼系統—Asirra。對於電腦而言，貓狗影像辨識較字串更為困難。本研究主要針對 Asirra 的影像資料試圖建構出貓狗影像自動辨識法，藉此來了解此辨識碼系統的有效性。已知影像包含大量雜訊，若使用原始資料則計算困難而且辨識效果差，所以萃取關鍵特徵為重要的研究課題。本文考慮方向梯度直方圖法 (Histograms of Oriented Gradients, HOG) 以及主成分分析 (Principal Components Analysis, PCA) 來篩選重要變數。我們將運用挑選出的特徵建立支持向量機 (Support Vector Machine, SVM) 分類器。在實證分析中，我們發現結合此兩種特徵萃取法，除了能夠大幅降低運算時間，也能得到良好的預測正確率。

關鍵字：Asirra、機器學習、影像辨識、方向梯度直方圖、主成分分析。

Abstract

In recent years, many websites, which requires a high standard of security, use CAPTCHA to avoid mass and brutal attacks from hackers. The CAPTCHA considers the use of strings of twisted and deformed English letters or numbers as an identification code. In 2007, the company Microsoft proposed a new image-based recognition system-Assira, which uses dogs and cats images as an identification code. Dogs and cats image recognition is not more difficult than strings of letters or numbers recognition for human, but is more challenging for computers. In this paper, we aim to develop a classification method for images from Asirra. An image is represented by an enormous number of pixels. Only few pixels carry important feature information, most pixels are noise. The abundance of noise leads to computational inefficiency, and even worse, may results in inaccurate recognition. Therefore, in this problem feature extraction is an essential step before a classifier construction. We consider HOG (Histograms of Oriented Gradients) and PCA (Principal Components Analysis) to select important features, and use the features to construct a SVM (Support Vector Machine) classifier. In the real example, we find that combining the two feature detection methods can dramatically reduce computational time and have satisfactory predictive accuracy.

Keyword : Asirra 、 machine learning 、 image recognition 、 Histograms of Oriented Gradients 、 Principal Components Analysis 。

目次

| | |
|----------------------------|----|
| 摘 要..... | I |
| Abstract..... | II |
| 表次..... | IV |
| 圖次..... | V |
| 第一章 緒論 | 1 |
| 第二章 貓狗特徵萃取與辨識 | 4 |
| 一、數位影像..... | 4 |
| 二、方向梯度直方圖..... | 6 |
| 1. 影像灰階化..... | 6 |
| 2. 統一圖像大小..... | 7 |
| 3. 計算梯度..... | 8 |
| 4. 繪製方向梯度直方圖..... | 11 |
| 5. 區塊正規化及 HOG 特徵值..... | 13 |
| 三、主成分分析..... | 16 |
| 四、支持向量機分類器..... | 19 |
| 第三章 實證分析 | 20 |
| 一、 貓狗特徵萃取的辨識結果..... | 22 |
| 二、 針對貓狗正臉特徵萃取的辨識結果..... | 30 |
| 第四章 結論與建議 | 38 |
| 參考文獻..... | 39 |

表次

| | |
|---|----|
| 表 3-1：5000 張影像以 PIXEL 法之十折交互驗證的預測結果..... | 25 |
| 表 3-2：5000 張影像以 PCA 法選取特徵之十折交互驗證的預測結果 | 25 |
| 表 3-3：5000 張影像以 HOG 法選取特徵之十折交互驗證的預測結果 | 26 |
| 表 3-4(a)：5000 張影像以 HOG+PCA 法選取特徵之十折交互驗證的預測結果 (Linear)..... | 26 |
| 表 3-4(b)：5000 張影像以 HOG+PCA 法選取特徵之十折交互驗證的預測結果 (Polynomial)..... | 26 |
| 表 3-4(c)：5000 張影像以 HOG+PCA 法選取特徵之十折交互驗證的預測結果 (Radial basis)..... | 26 |
| 表 3-4(d)：5000 張影像以 HOG+PCA 法選取特徵之十折交互驗證的預測結果 (Sigmoid)..... | 26 |
| 表 3-5：400 張影像以 PIXEL 法之十折交互驗證的預測結果..... | 35 |
| 表 3-6：400 張影像以 PCA 法之十折交互驗證的預測結果 | 35 |
| 表 3-7：400 張影像以 HOG 法之十折交互驗證的預測結果 | 36 |
| 表 3-8：400 張影像以 HOG+PCA 法之十折交互驗證的預測結果 | 37 |

圖次

| | |
|---|----|
| 圖 2-1：八位元像素光暗層次圖 0~255 | 5 |
| 圖 2-2：RGB 影像組成 | 5 |
| 圖 2-3：原始、灰階、壓縮後的貓狗圖像 | 7 |
| 圖 2-4：梯度計算範例 | 9 |
| 圖 2.5 (a) 貓原始影像，(b)貓影像的梯度強度圖，(c)貓影像的梯度方向圖。(d) 狗 原始影像，(e)狗影像的梯度強度圖，(f)狗影像的梯度方向圖 | 10 |
| 圖 2-6：方向梯度直方圖 | 12 |
| 圖 2-7：格子示意圖 | 12 |
| 圖 2-8：區塊示意圖 | 13 |
| 圖 2-9：區塊正規化中符號定義範例 | 15 |
| 圖 2-10：主成分幾何意義示意圖 | 16 |
| 圖 2-11：核函數投影後資料線性可分示意圖 | 19 |
| 圖 3-1：資料集中部分貓狗圖像 | 21 |
| 圖 3-2：符合正面臉的狗影像標準示意圖；符合正面臉的貓影像標準示意圖 ... | 32 |
| 圖 3-3：上方為針對正臉挑選後部分貓影像示意圖；下方為針對正臉挑選後部分 狗影像示意圖 | 33 |
| 圖 3-4：上方為交叉驗證中被錯誤分類的貓影像；下方為被錯誤分類的狗影像 | 34 |

第一章 緒論

人類為了將觀察到的人事物景像紀錄起來，便有了影像的發明。影像除了可以提供視覺上的直接效果，並且可以表達文字難以敘述的抽象事物。隨著科技的發展，影像的產生、取得與儲存越來越容易，請見 Schwartz (1995)。傳統統計分析的資料型態為量化數字，近年來文字辨識是熱門研究課題。我們相信在不久的未來，影像也將成為另一種型態的資料。故本研究嘗試伸出觸角到影像資料，我們的樣本資料是數千張的貓狗影像，我們試著將統計上的分類問題應用在動物影像自動辨識上。

在分類問題上，近 20 年來的機器學習 (Machine learning) 領域已有很多成熟、有效的工具被發展，見 Domingos (2012)。這些工具利用有效率的計算以分析大量資料中的特定結構或規律，並訓練成一分類器以預測新資料的類別。其中支持向量機 (Support Vector Machine, 簡稱 SVM) 為機器學習中常見用於監督式學習 (Supervised learning) 的算法，最早由 Cortes 和 Vapnik (1995) 提出。支持向量機的用途主要為分類及迴歸問題，且在分類問題上多優於其他監督式學習法，故被廣泛應用在各領域，也是此篇研究所使用的分類器方法。

在科技進步的時代中，各項相關領域技術發展快速，電腦等自動化機械的產品都能經由程式及計算自動辨認出影像中的資訊，因此在 2003 年，Ahn 等人提出全自動區分電腦和人類的公開圖靈測試 (Completely Automated Public Turing test to tell Computers and Humans Apart, 簡稱 CAPTCHA)。CAPTCHA 就是現今電腦上經常可見的驗證碼。驗證碼為由電腦自動產生一串英文或數字結合的文字串影像，其目的是希望產生的影像是人類透過肉眼能夠輕鬆辨識，對電腦卻是難以或無法辨識的影像。因此運用驗證碼可以阻擋惡意的自動化軟體對於網頁大量、頻繁的攻擊，常見於安全要求較高的交易系統。為了使這些自動化軟體更加難以辨認，驗證碼中的文字串往往會被加以扭曲、變形、分割等來增加學習規律的困難度，請參考 Huang 等人(2010)。

但 CAPTCHA 近年來也逐步被駭客破解，此導致越來越多使用者對系統的網路安全的疑慮加深。故在 2007 年微軟 (Microsoft) 的研究中，提出了 CAPTCHA 的改良-Asirra (Animal Species Image Recognition for Restricting Access, 簡稱 Asirra)，見 Elson 等人 (2007)。在精進的 Asirra 中，電腦從大型的貓狗影像資料庫中產生 12 張貓狗影像，使用者必須分辨出影像中哪幾張為貓、哪幾張為狗。與 CAPTCHA 比較起來，Asirra 有更多優勢。首先它採用了我們生活常見的寵物—貓與狗，一般人都可以輕鬆的辨別貓狗之間的差別。而且相較於繁雜且扭曲的數字文字串，電腦使用者將更願意回答 Asirra 產生的問題。另一方面，對電腦而言，辨識外表四肢非常相近的貓狗將比辨識文字困難許多。而且人類在貓狗辨識上較不會受背景所影響，但對於自動化辨識系統來說，背景將摻雜了更多的雜訊，使得辨識更加困難。因此為了挑戰 Asirra，本篇研究希望發展一套自動化辨別貓狗影像的方法。

由於每張影像由數以十萬、百萬計的像素 (Pixel) 所組成，在統計的角度上，每一張影像是一筆觀測值，像素則可視為是變數。故影像資料含有大量的變數，過多的變數將使得分析計算效率差，若多數變數為雜訊則將掩蓋主要的資訊，使得分析結果不佳。因此在眾多繁雜的變數中，萃取出重要特徵或選取有代表性的變數為重要且必要的程序，我們的研究主題便聚焦在貓狗影像辨識問題的特徵萃取上。

在影像上常用於萃取描述特徵的方法如尺度不變特徵轉換 (Scale-invariant feature transform, 簡稱 SIFT) 與加速穩健特徵 (Speeded Up Robust Features, 簡稱 SURF)，見 Lowe (2004) 與 Bay 等人 (2006)。而本論文使用於影像特徵萃取法為由 Dalal 及 Triggs 於 2005 年提出的方向梯度直方圖 (Histograms of Oriented Gradients, 簡稱 HOG)，並針對行人的影像萃取外型特徵。此法計算影像在每個位置的梯度，並根據其方向建立強度直方圖。此直方圖能夠描述行人的邊緣特徵，故能萃取相關資訊。此研究將採用方向梯度直方圖法在貓狗的影像上，以萃取出貓狗的外型邊緣特徵。另一方面，我們也將考慮統計上常用於降低維度與

萃取重要特徵的主成分分析法 Jolliffe (1986)。已知，主成份分析法主要篩選出具有解釋資料變異能力的變數線性組合，我們將探索這些主成份在此影像辨識問題上是否能夠提供關鍵資訊。

以下為本篇論文的文章結構：第二章為研究方法，我們將說明貓狗影像的特徵萃取及分類器建立。在特徵萃取上，我們將介紹梯度方向直方圖法以及主成份分析法。另外，我們也將簡短介紹支持向量機分類器。在第三章中，我們將考慮不同的特徵萃取方法。首先在第一個實驗中，這些方法將被直接運用在 Asirra 部分資料集中，我們根據交叉驗證的結果來比較這些方法的表現。在第二個實驗中，我們人工挑選出資料中那些能夠清楚呈現貓狗正臉的影像，同樣地考慮不同特徵萃取方法以及支持向量機分類器，我們再次評估這些方法的優劣。我們也將比較兩個實驗的結果來瞭解圖像事前處理的重要性。第四章為結論與未來研究方向。



第二章 貓狗特徵萃取與辨識

在這篇研究論文裡，我們的研究目的是發展辨別貓狗影像的方法。在這一章中，我們將先介紹資料也就是數位影像。由於影像由數十萬至百萬像素所組成，也就是變數眾多，所以變數縮減是重要研究課題。而且這些像素變數間在空間位置上呈現複雜相關性，我們希望在這些變數間萃取出對貓狗辨識有意義的重要特徵，所以我們將在這一章中介紹兩種特徵萃取的方法。最後我們也將介紹我們所運用的分類器方法——支持向量機。

一、數位影像

數位影像 (Digital image) 為由數值表示的像素所組成的影像。將影像放大來看，可以發現影像是由細小的格狀點所組成的，這些點稱為像素 (Pixel)，像素乃構成影像最基本的單位。一般像素以八位元 (8-Bit) 來表示，其中位元代表二進位數值(Binary Digit)，所以八位元則表示有 256 (2 的八次方) 種結果。一般灰階影像 (Grayscale image) 以數值 0~255 來呈現單位像素，其數值則反應光的強度，當數值越趨近於 0，則該像素點愈接近黑 (暗) 色；當數值越高、趨近於 255，則表示該像素點越接近白 (亮) 色，見圖 2-1。

當影像中包含顏色資訊則稱為彩色影像 (Color image)。有別於灰階影像，只由單一數值代表該像素點的亮度，彩色影像的像素點為三維度或四維度的向量。常見的彩色影像類型如 RGB (Red, Green, and Blue) 與 CMYK (Cyan, Magenta, Yellow, Key) 等。其中 RGB 影像最為常用，由紅綠藍三種顏色所組成。由於人類在顏色的感知中對於紅綠藍三種顏色較為敏感，它們被定義為光的三原色，彩色影像由這些顏色色板重疊而成，見圖 2-2。舉例來說，若某像素點為 $(R, G, B) = (246, 174, 52)$ ，代表該由紅色色板 246，綠色色板 174，藍色色板 52 重疊而成的像素值，本篇論文採用的資料影像皆為 RGB 類型，對於更詳細的了解 RGB 空間，見 Süssstrunk 等人 (1999)。

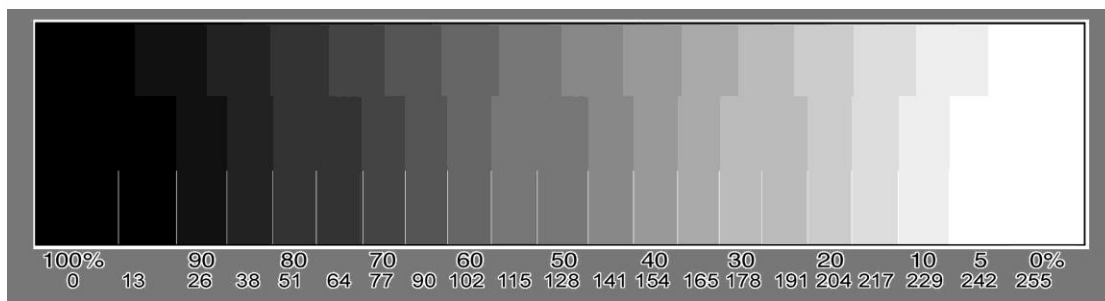


圖 2-1：八位元像素光暗層次圖 0~255(引自 <https://www.inksupply.com/utez.cfm>)。

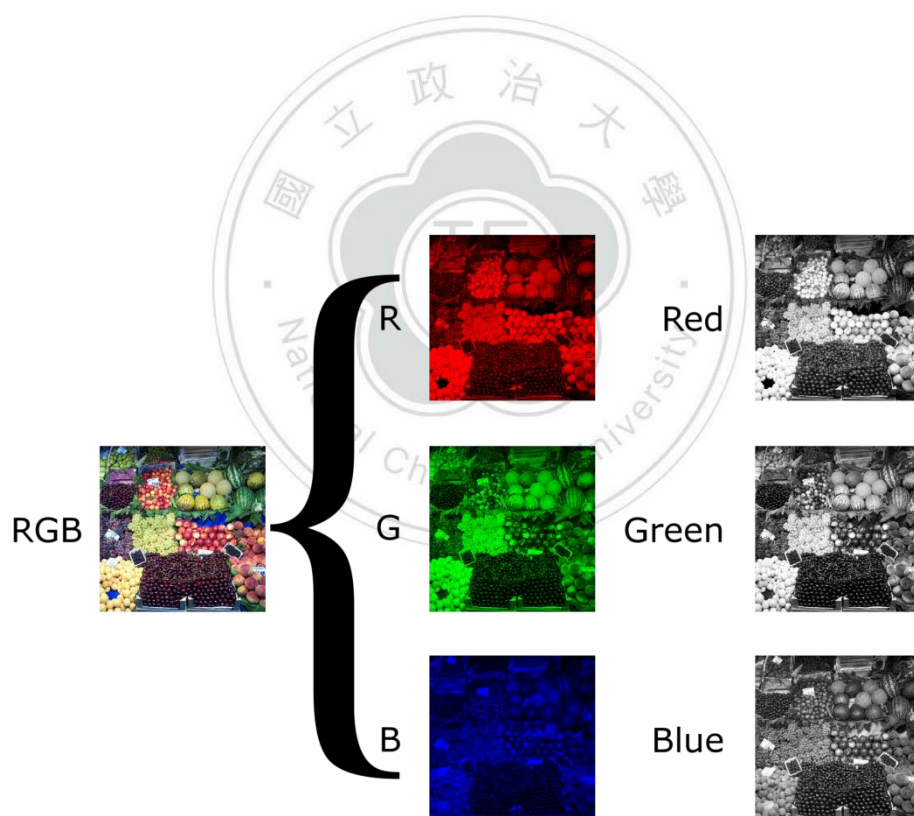


圖 2-2：RGB 影像組成(引自 <https://en.wikipedia.org/wiki/Grayscale>)。

二、方向梯度直方圖

針對貓狗影像的特徵萃取，本研究考慮採用方向梯度直方圖法 (Histograms of Oriented Gradients, HOG)。此法由 Dalal 及 Triggs 於 2005 年提出，主要目的為萃取影像中的物體邊緣 (edge) 特徵。已知在物體的邊緣其鄰近影像的變化大，所以他們運用梯度 (gradient) 反映鄰近影像的變化以偵測邊緣。在此法中，影像被切割成數區段，區段內的像素依照其梯度方向彙整建構出直方圖，當作描繪影像邊緣趨勢的特徵。原作者將此法運用在行人的影像中，我們則採用此法來萃取貓狗影像的特徵，詳細步驟如下：

1. 影像灰階化。
2. 統一影像大小。
3. 計算梯度及梯度方向。
4. 繪製方向梯度直方圖。
5. 區塊正規化及 HOG 特徵值。

以下我們介紹各個步驟的細節。

1. 影像灰階化

由於 HOG 是針對邊緣的特徵萃取方法，著重於物體的形狀以及物體與背景的明暗對比差異，則顏色提供較少資訊，再加上不考慮圖像色彩能大幅降低運算時間，故將彩色影像轉換成黑白影像。對於一張 64×64 像素 RGB 彩色影像來說，原始影像由三個 64×64 的像素矩陣所組成，包括紅色模版的像素、綠色模版的像素以及藍色模版的像素相加而成。在固定位置下，我們採用 Matlab 中灰階化指令內的公式做為灰階化的標準公式，令 Y 為灰階影像的像素值，則

$$Y = 0.299R + 0.587G + 0.114B \quad (2-1)$$

其中 R 代表紅色模版的像素、G 代表綠色模版的像素而 B 代表藍色模版中的像素，灰階化後的貓狗影像可見圖 2-3。

2. 統一圖像大小

由於微軟的 Asirra 提供的資料集中的每張貓狗影像大小並不一致，為了確保每張影像都擁有相同的像素大小，或相同的變數數量，故將所有貓狗影像壓縮成相同的大小。已知當影像經過壓縮後，將造成原影像的資訊損失，以及增加額外雜訊。所以能夠保留原像素越多，後續處理分類問題的結果一般會越好。Dalal 及 Triggs (2005) 在他們的行人影像研究中，考慮影像壓縮至 64×128 像素。由於人是屬於高、長的站立生物，故採長方形為合理選擇。在此研究中，由於貓狗長寬差異較小，故我們將資料集中的影像統一壓縮成 64×64 像素的正方形影像。請見圖 2-3。

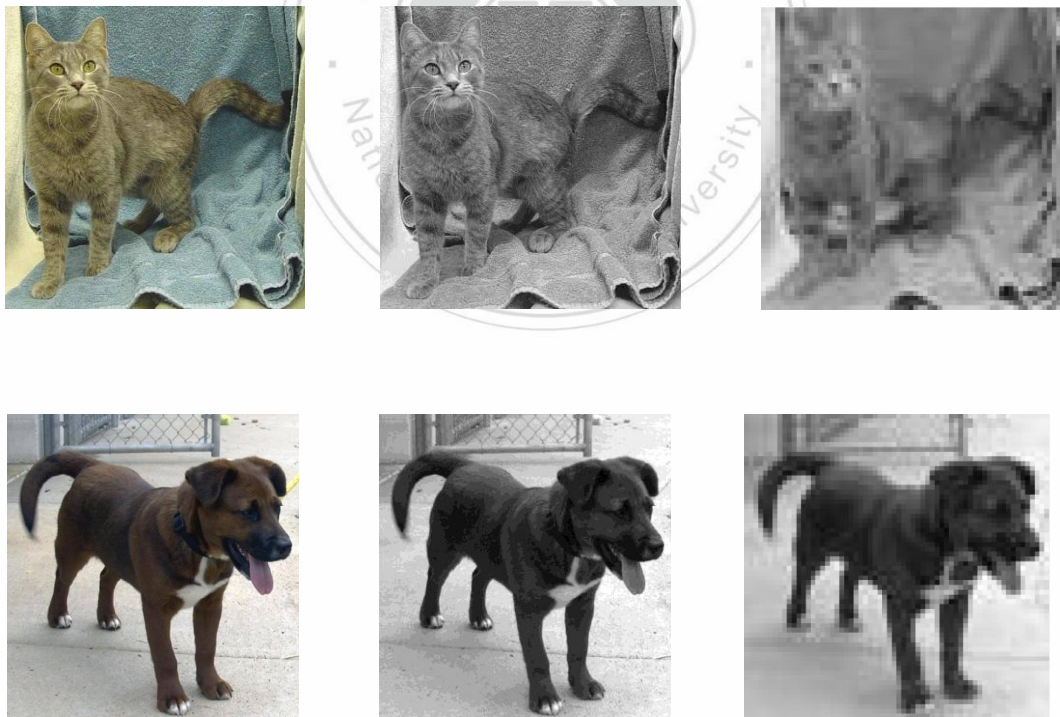


圖 2-3：上方為資料集中貓的影像，下方為狗的影像。由左至右為原影像、灰階化的影像與壓縮過後的影像。

3. 計算梯度

影像梯度定義為在某個位置上像素值的變化量。我們可以想像成站在樓梯中的一個階梯，我們往前踏一步與往後踏一步可以知道高度的上升與下降。換句話說，在一個二維的影像中，我們利用某個像素點中橫軸與縱軸的前後像素點就可以輕鬆得到此點的梯度變化量。

首先令 $I(x,y)$ 為橫軸座標 x 與縱軸座標 y 位置的像素，定義 $f_u(x,y)$ 為橫軸方向的梯度值， $f_v(x,y)$ 代表縱軸方向的梯度值，則

$$f_u(x,y) = \frac{I(x+1,y)-I(x-1,y)}{2}, \quad f_v(x,y) = \frac{I(x,y+1)-I(x,y-1)}{2}。$$

另外，定義梯度強度(magnitude)如下，

$$M(x,y) = \sqrt{f_u(x,y)^2 + f_v(x,y)^2};$$

則梯度方向(orientation)為角度 $\theta(x,y) = \tan^{-1} \frac{f_v(x,y)}{f_u(x,y)}$ ，其值介於 $-90^\circ \sim 90^\circ$ 間。

為了能夠更加的理解，我們以圖 2-4 中一特定部位的像素作為計算梯度範例，其中右圖包含了 3×3 格的像素。則針對中心點 P ，其橫軸、縱軸方向的梯度值分別為

$$f_u(x,y) = \frac{I(x+1,y) - I(x-1,y)}{2} = \frac{81 - 180}{2} = -49.5;$$
$$f_v(x,y) = \frac{I(x,y+1) - I(x,y-1)}{2} = \frac{48 - 185}{2} = -68.5。$$

則梯度強度與梯度方向分別為：

$$M(x,y) = \sqrt{\frac{(81 - 180)^2}{2^2} + \frac{(48 - 185)^2}{2^2}} \doteq 119.5199$$
$$\theta(x,y) = \tan^{-1} \frac{(48 - 185)/2}{(81 - 180)/2} \doteq 54.1471^\circ$$

當每一個位置的梯度資訊被計算出來後，為了理解梯度強度與方向在影像上的意義，我們得以將這些值以影像呈現，見圖 2-5。圖 2.5(a)與(d)為貓、狗原始影像，(b)(e)為這些影像相對應的梯度強度圖，(c)與(f)則為梯度方向圖。在梯度強度圖(b)(e)中越接近白色(越亮)的部分表示該地區梯度變化較大。而梯度方向圖(c)(f)由藍色的方向箭頭所組成，其箭頭越深，也代表該區域的影像梯度越強。從中我們發現梯度強度圖(e)可以清楚描述出狗的邊緣;但在(b)圖中，由於原影像(a)貓的毛色與背景顏色的明亮度相似，而且毛皮與毛巾上分別有很多斑紋與皺褶，導致影像產生許多外型邊緣以外的雜訊，故模糊了邊緣特徵。我們在梯度方向圖上也得到類似的結果。故當原始影像中的物體與背景區分越清楚，以及物體的顏色越一致，則梯度資訊將能提供越清楚的邊緣特徵。

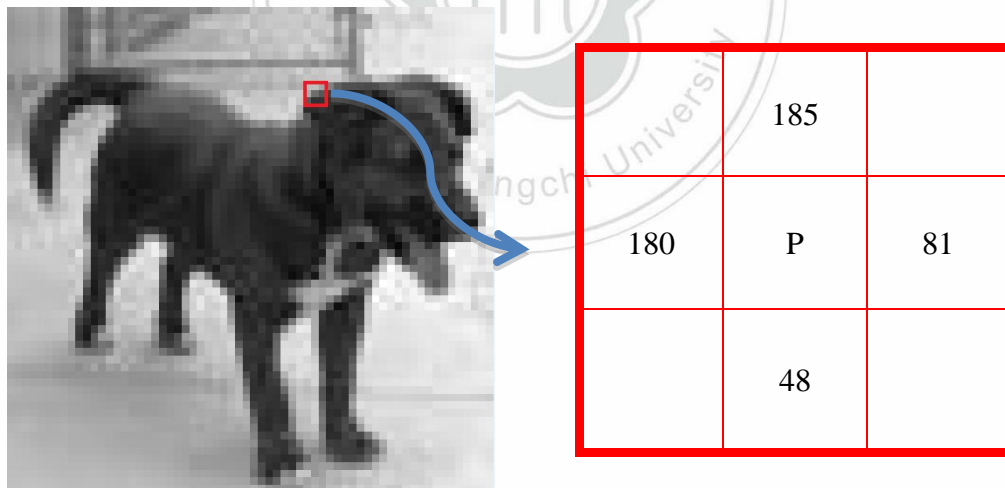


圖 2-4：梯度計算範例。



(a)原始影像-貓。



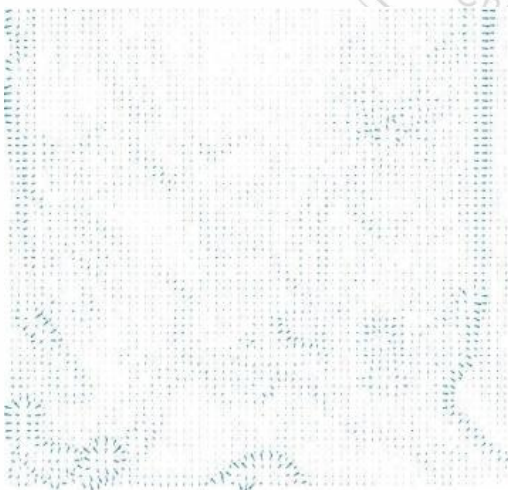
(d)原始影像-狗。



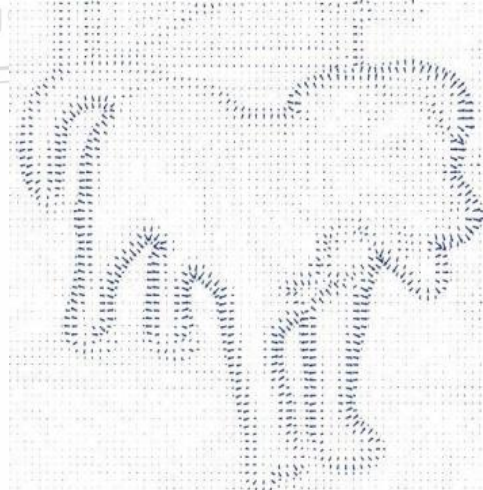
(b)梯度強度圖-貓。



(e) 梯度強度圖-狗。



(c)梯度方向圖-貓。



(f) 梯度方向圖-狗。

圖 2.5：(a)貓原始影像，(b)貓影像的梯度強度圖，(c)貓影像的梯度方向圖。

(d)狗原始影像，(e)狗影像的梯度強度圖，(f)狗影像的梯度方向圖。

4. 繪製方向梯度直方圖

在這個階段中，我們將利用取得的梯度資訊，彙整臨近區域的結果，以梯度強度當作權重，繪製梯度方向的直方圖。這個直方圖將呈現影像上各部位鄰近區域的梯度方向的分佈情況。

首先將影像分割成大小相同的正方格子 (cell)，每一個格子由 $K \times K$ 個像素所組成，見圖 2-8。參考 Dalal 及 Triggs (2005) 的實驗結果，本文考慮每個格子由 8×8 像素組成。每個格子中的 64 個像素中將包含 64 對梯度強度與梯度方向 (M, θ) 的結果。首先為了計算方便，故利用 \tan 的週期特性根據(2-2)式將像素的梯度方向的範圍由 $(-90^\circ \sim 90^\circ)$ 調整為 $(0^\circ \sim 180^\circ)$ 。

$$\theta(x, y) = \begin{cases} \theta(x, y) + \pi, & \text{if } \theta(x, y) < 0 \\ \theta(x, y), & \text{if } \theta(x, y) \geq 0 \end{cases} \quad (2-2)$$

接著我們將梯度方向分為 9 個區間: $0^\circ \sim 20^\circ$ 、 $20^\circ \sim 40^\circ$ 、...、 $160^\circ \sim 180^\circ$ 。利用雙線性內插法(Bilinear interpolation)統計各區間的強度直方圖。假設一像素的梯度強度為 M ，方向為 θ 。令 $\theta_1 \leq \theta \leq \theta_2$ ，其中 θ_1 、 θ_2 為角度 θ 鄰近兩區間的中心點，則將該像素的梯度強度 M 依據角度間的距離分配至兩區間。即將強度 $\frac{\theta_2 - \theta}{\theta_2 - \theta_1} M$ 分至 θ_1 所在的區間，強度 $\frac{\theta - \theta_1}{\theta_2 - \theta_1} M$ 分配至 θ_2 所在的區間。

也就是說，將格子內各像素的梯度強度依據梯度方向，分配兩鄰近方向區間內。舉例來說，若某個像素的梯度強度為 $M=50$ ，方向為 $\theta = 25^\circ$ ，相鄰兩區間內的中心點為 $\theta_1 = 10^\circ$ 、 $\theta_2 = 30^\circ$ 。則將強度 $50 \times \frac{(30-25)^\circ}{20^\circ}$ 為權重分配至 $0^\circ \sim 20^\circ$ 的區間(中心點為 10°)；強度 $50 \times \frac{(25-10)^\circ}{20^\circ}$ 為權重分配至 $20^\circ \sim 40^\circ$ 的區間(中心點為 30°)。所以格子內 64 格像素的梯度強度都能藉由雙線性插值分配至 9 個梯度方向區間中並繪成直方圖，見圖 2-6。故每個格子都對應一張包含 9 個區間的直方圖，換句話說，直方圖上的 Y 值也就是每一個梯度方向區間的累積梯度強度為這個格子內反映邊緣的特徵變數。

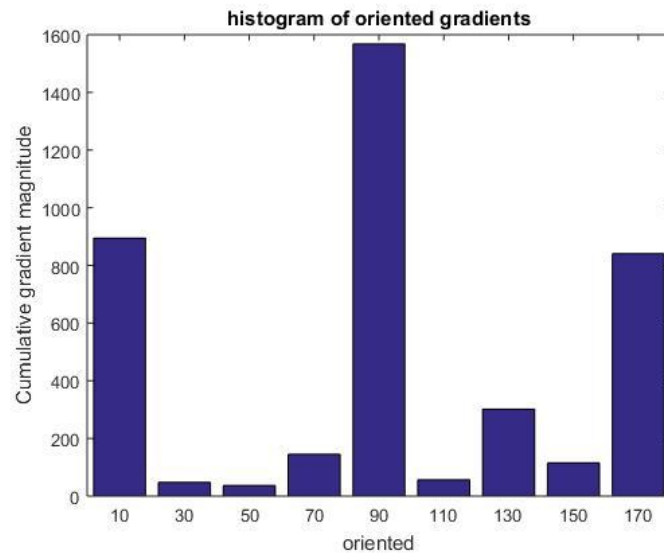


圖 2-6：方向梯度直方圖。



圖 2-7：格子示意圖。

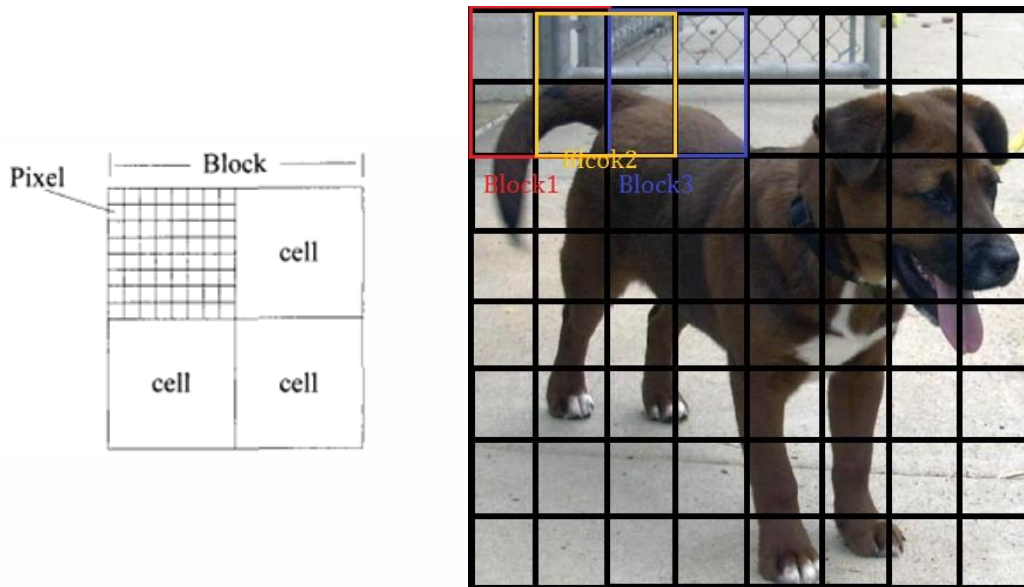


圖 2-8：區塊示意圖(左圖引自 <https://read01.com/On3K5B.html>)。

5. 區塊正規化及 HOG 特徵值

當繪製完成各格的方向梯度直方圖後，我們接著定義區塊 (block)。區塊由數個鄰近的格子所組合而成。常見的區塊為由 $L \times L$ 個格子組合的方型區塊，另外也可以考慮圓形或不規則形。此篇論文採用較為簡單的方型區塊，且 $L=2$ 。Dalal 及 Triggs (2005) 提出可以重疊取區塊，更可以視情況採取不同重疊率。在運算效率的考量下，我們參考 Dalal 及 Triggs (2005) 的結果，本文採用方型的區塊，每一區塊由 2×2 的格子所組成，每一格則由 8×8 像素組成，區塊間的重疊率為 50%，見圖 2-10。故一張 64×64 像素的圖將由 8×8 格子組成， 2×2 格子以 $1/2$ 互相重疊的方式組成各個區塊，總共包括 $(8 - 1) \times (8 - 1) = 49$ 個區塊，見圖 2-9。

我們以一個區塊為單位，每個區塊中包括四個格子，每一個格子皆有對應的九個方向梯度區間的強度直方圖，將這些直方圖上的強度值按照位置順序排列成一 1×36 列向量。給定 $i, j=1, \dots, 8$ ，定義 $V_{ij}(k)$ 為第 i 列第 j 行的格子的方向梯度直方圖裡第 k 個區間的強度，則 $k=1, \dots, 9$ 。則令 C_{ij} 為該直方圖中所有區間強度所組成的向量，

$$C_{ij} = \{V_{ij}(1), V_{ij}(2), \dots, V_{ij}(9)\}。$$

則令 B_{ij} 為第 i 列第 j 行的區塊內四個格子的方向梯度直方圖的區間強度所組成的向量：

$$B_{ij} = \{C_{ij}, C_{i(j+1)}, C_{(i+1)j}, C_{(i+1)(j+1)}\}, i, j = 1, \dots, 7。$$

我們接著對每一個區塊的向量 B_{ij} 採取 \mathcal{L}_2 - norm 正規化。也就是將 B_{ij} 向量裡每個元素除以區塊裡所有向量的平方和相加且開根號，見(2-3)。其中 ε 為一接近且大於 0 的數，本篇實證分析中令 $\varepsilon = 0.01$ ，其設置目的是防止分母為 0 的情形發生。

$$\widetilde{B}_{ij} = \frac{B_{ij}}{\sqrt{\|B_{ij}\|_2^2 + \varepsilon}}。 \quad (2-3)$$

而正規化的目的是為了減少影像中某些部分可能因光線造成物體的表面陰影所導致的強烈梯度變化，藉由正規化區塊可以降低各區塊內的光線對比，來抑制非邊緣提供的梯度資訊。

經過正規化後，每張 64×64 像素影像的 HOG 特徵量將包括 1764 (49 個 BLOCK \times 4 個 CELL \times 9 個方向) 個元素或變數，其主要描述影像中物體的邊緣特徵。

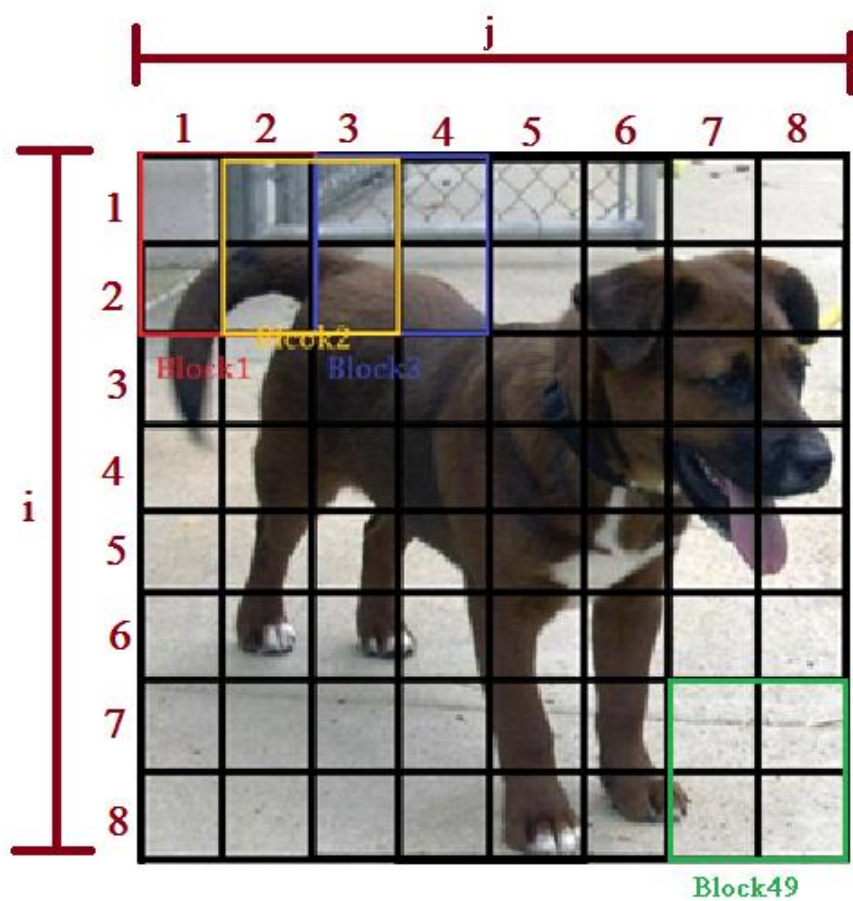


圖 2-9：區塊正規化中符號定義範例

三、主成分分析

由於經過上一節介紹的 HOG 方法處理後，可萃取出 1764 個變數，變數數量仍然眾多，故我們考慮採用統計上常用的主成分分析法 (Principal Components Analysis, 簡稱 PCA) 來達到變數縮減的目的。主成分分析主要是將原先變數透過線性組合後得到新變數，在所有組合中尋找擁有較大的變異數的新變數，這些新變數稱為主成分。其幾何意義便是將變數經過基底變換或是轉座標軸，我們希望在這個新的座標軸上資料能夠以最大的分散程度呈現。當資料有高的分散程度，則對觀測值間的差異提供較多的資訊。為了獲取不同面向的資訊，主成分間彼此線性獨立。經過轉軸後，我們捨棄掉變異能力較低的主成分，只萃取重要的主成分來達到維度縮減的目的，其細節請見 Smith (2002) 及 Rencher (1995)。

舉例來說見圖 2-10 中，假設原先兩變數的座標軸為黑色線，經由線性組合後得到兩個正交的紅色線的 y_1 軸及藍色線的 y_2 軸。由圖形上可明顯看出，資料點在第一主成分 y_1 上的投影有較大變異程度，也就是此變數對不同資料點有較高的分辨能力。

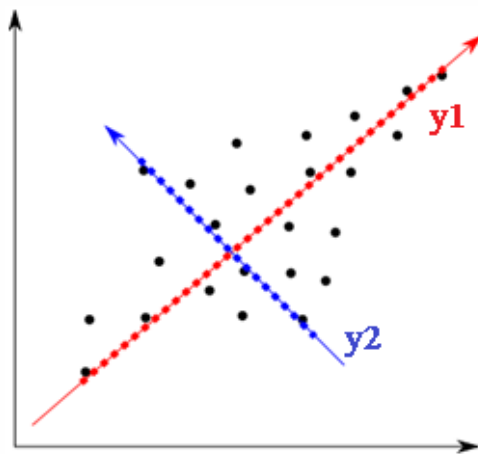


圖 2-10：主成分幾何意義示意圖(引自

<http://www58.atwiki.jp/dooooornob/pages/39.html>)

假設原樣本資料有 p 個變數 $X = (X_1, X_2, \dots, X_p)^T$ ，其共變異數矩陣為 $\text{Var}(X)=S$ 。令 $Y = (Y_1, Y_2, \dots, Y_p)^T$ ，其中 Y_j 為原變數的線性組合，即

$$Y_j = a_j^{*T} X, \quad a_j^* = \begin{pmatrix} a_{1j}^* \\ a_{2j}^* \\ \vdots \\ a_{pj}^* \end{pmatrix}, \quad j = 1, \dots, p。$$

令 $A = (a_1^* \dots a_p^*)$ ，則 $Y=AX$ ，且其共變異數矩陣為 $S_Y = ASA^T$ 。我們的目的是找尋使 $\text{Var}(Y_j)$ 最大化的線性組合。已知當隨機變數乘上一常數，則其變異數將乘上此常數平方，故若未加限制式，則此最佳的線性組合不存在（係數趨近於無限），所以我們考慮加上限制式 $a_j^{*T} a_j^* = 1, j = 1, \dots, p$ 。另一方面，這些 Y_j 之間應彼此無關，故 S_Y 為對角線矩陣，即

$$S_Y = ASA^T = \begin{pmatrix} s_{Y1}^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & s_{Yp}^2 \end{pmatrix}。$$

也就是我們希望找到矩陣 A 使得將原共變異數矩陣 S 對角線化。則根據特徵值分解 (eigenvalue decomposition, 簡稱 EVD)，我們即可得知

$$ASA^T = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)，$$

其中 λ_i 為共變異數矩陣 S 的特徵值 (eigenvalue)，且矩陣 A 內的第 i 個行向量 a_i^* 即為相對應的特徵向量 (eigenvector)。則在此結果下，第 i 個主成分的變異數等於特徵值 λ_i ，即 $s_{Yi}^2 = \lambda_i$ 。由於變異數程度反映主成分的資訊量，故一般將特徵值由大自小排序，則 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ 。

故我們只要對原先變數的共變異數矩陣 S 作特徵值分解，便可以得到 p 個主成分 Y_1, Y_2, \dots, Y_p ，其中第 j 個主成分可表示為：

$$Y_j = a_{1j}X_1 + a_{2j}X_2 + \dots + a_{pj}X_p$$

係數 (a_{1j}, \dots, a_{pj}) 為共變異數矩陣 S 的第 j 個特徵值 λ_j 對應的特徵向量，而且

$$\text{Var}(Y_j) = \lambda_j, \text{cov}(Y_j, Y_{j'}) = 0, j \neq j'。$$

當母體共變異數矩陣 S 未知時，則可以樣本共變異數矩陣替代估計。

當資料中的變數個數高時，很多時候大部分的變異都可由少數的主成分所解

釋，所以我們可根據解釋變異程度來選取重要的主成分。給定主成分個數 k ,

$1 \leq k \leq p$, 則定義前 k 個主成分之解釋變異比例為

$$\frac{\lambda_1 + \lambda_2 + \cdots + \lambda_k}{\lambda_1 + \lambda_2 + \cdots + \lambda_p} \in [0,1]。$$

實務上，通常我們是先給定一解釋變異比例最低水準，來決定最終選取的主成分個數。選取較多主成分，則通常分析結果較為準確，但需要更多計算時間；相反地，變數若經過大幅度刪減，則雖然可節省分析時間，但結果通常較差。常見要求的比例水準為 70%，80% 或 90%，研究人員仍須因應實驗嘗試不同結果。



四、支持向量機分類器

我們考慮的貓狗影像首先經由前述的方向梯度直方圖法以及主成分分析萃取出特徵變數，接著我們運用支持向量機 (Support Vector Machine, SVM) 對貓狗的分類問題建立分類器。

SVM 主要是找到一個超平面，使相異類別的資料集中最靠近此超平面的訓練資料點之距離最遠，則此超平面便是最佳的分類邊界。詳細可以參考 1995 年由 Cortes 和 Vapnik 提出的 Support-Vector Networks。本篇論文採用 MATLAB 中的 LIBSVM 工具箱來建立 SVM 分類器，見 Chang and Lin (2001)，且 LIBSVM 提供了多種的參數選擇，因此將參數調整為最佳能使建立 SVM 的效率提升，包括計算時間，分類正確率等。而 SVM 也編譯成不同程式語言的版本供使用。

SVM 除了可應用在線性空間下可被完全分開的資料，當資料集本身不屬於線性可分時，則嘗試將資料集經由核函數 (Kernel function) 來映射至其他更高維度的空間，若資料經過映射後成為線性可分時，則可解決分類的問題，見圖 2-11，其中 ϕ 為核函數。在下一章的實證分析中，本篇論文考慮 LIBSVM 中內建的四種核函數: Linear、Polynomial、Radial basis function、Sigmoid。我們將透過交叉驗證 (cross validation, CV) 來評估各個核函數在此圖像資料分類問題上的表現。

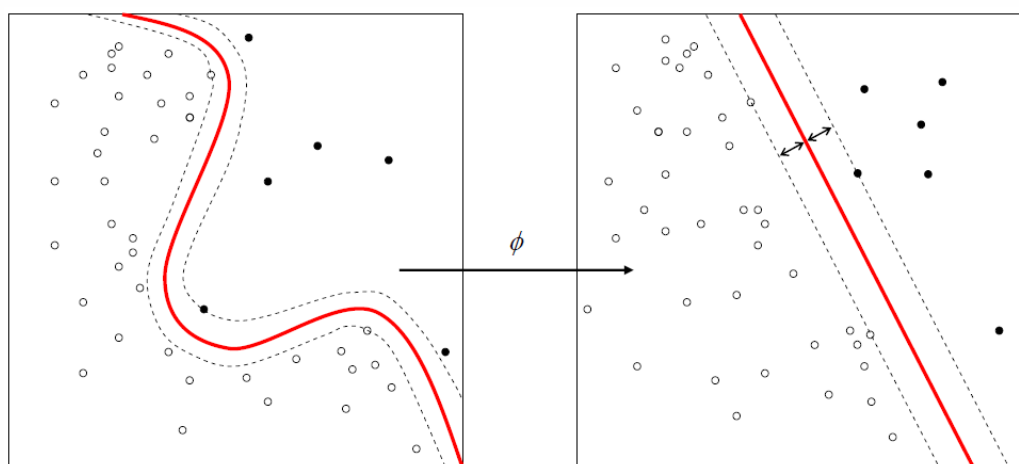


圖 2-11：核函數投影後資料線性可分示意圖

(引自 https://en.wikipedia.org/wiki/Support_vector_machine)

第三章 實證分析

在本章中，我們應用前一章所介紹的方向梯度直方圖來萃取影像中有關邊緣的特徵，以及統計上常用的變數萃取法——主成份分析法在貓狗資料上，之後利用這些特徵，建立支持向量機分類器，以達到貓狗分類辨別。我們首先介紹資料集的影像。在 3-1 節中，則將比較以不同特徵萃取程序所建立的支持向量機分類器，我們將利用十折交叉驗證 (10-fold) 以避免過度配適 (over-fitting) 所產生過度樂觀的結果。我們將整體資料均分為十部分，在每一次實驗中，十分之九的資料集當作訓練集 (training data set) 來萃取特徵以及建立分類器，再運用獲得的分類器來預測剩餘十分之一的測試集 (testing data set) 資料的動物類別。經過重複十次的實驗，我們彙整這些交叉驗證的結果，計算預測的平均正確率，以評估及比較這些方法。我們發現在資料集中的原始影像過於雜亂，部分提供動物正面清楚影像，但在很多影像中，動物為側像或坐臥姿勢，這嚴重影響辨識工作的結果。所以在 3-2 節中，我們將挑選正面清楚影像進行辨識。

本篇研究論文針對微軟的 Asirra 中所提供的 2500 張貓及 2500 張狗共五千張影像做為研究對象，見圖 3-1。由圖 3-1 可得知資料集中的貓和狗影像為一般日常的生活照，例如在草原上玩耍、睡覺、被主人抱著等。這些影像全為 RGB 圖像，但大小並不統一，像素大小的範圍約為 $100 \times 100 \sim 500 \times 500$ 像素。我們首先將影像灰階化並統一壓縮為 $64 \times 64 = 4096$ 個像素大小。

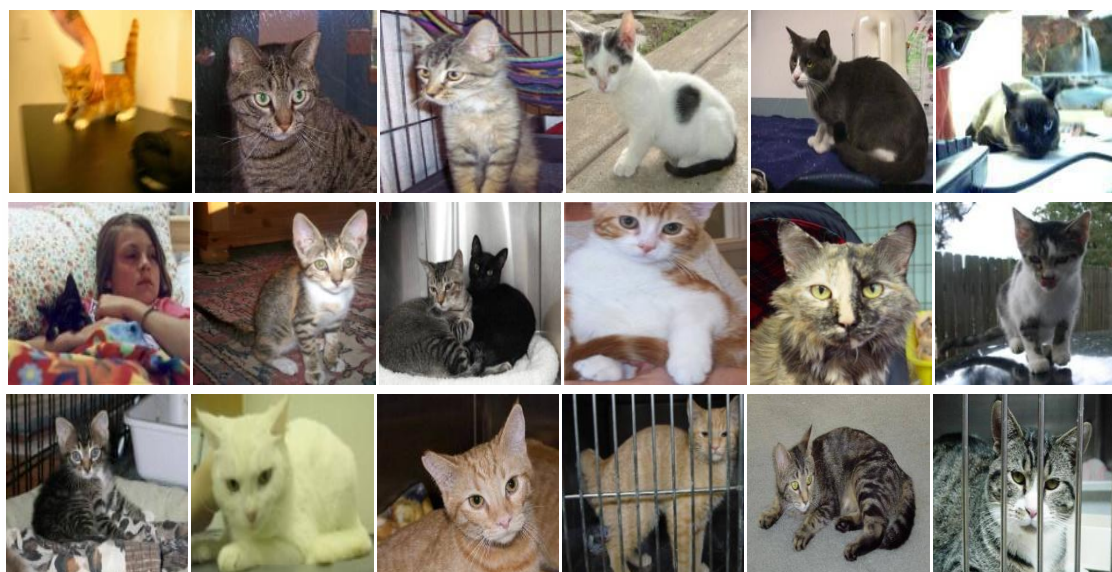


圖 3-1：資料集中部分貓狗影像。

一、 貓狗特徵萃取的辨識結果

我們考慮以下四種不同特徵萃取法：

方法一、 Pixel 法：在此法中，我們不做任何特徵萃取而直接採用所有 4,096 個像素來建立分類器。

方法二、 PCA 法：此法考慮運用主成份分析法以萃取出解釋變異大的主成分。

方法三、 HOG 法：此法運用方向梯度直方圖法以萃取物體邊緣共 1,764 個特徵。

方法四、 HOG + PCA 法：此法考慮先運用方向梯度直方圖法，再從所萃取出的邊緣特徵利用主成份分析法挑選出解釋變異的主成分。

經過特徵萃取後，我們利用這四種方法所選出的變數建立支持向量機分類器。我們將考慮包括 linear, polynomial, radial basis 與 sigmoid 四種核函數。

表 3-1 至 3-4 為四種方法的結果，包括十次交叉驗證的平均預測正確率，預測正確率表示原預測資料中的類別經過訓練資料建立的分類器後預測類別相等，佔所有預測中預測相等的比例；正確率的標準差；交叉驗證所總運算時間（單位為秒），以及建立支持向量機分類器所需之平均迭代次數。在方法二與方法四運用主成份分析法時，我們考慮 60%, 70%, 80%, 90% 的解釋變異比例，表中也提供相對應所選出的主成份平均個數。另一方面，在方法四中我們同時也考慮選取前 100、300、500 或 1000 個主成份。

表一為 PIXEL 方法的預測結果，我們採用的核函數為 linear 與 polynomial，因為 radial basis 與 sigmoid 在 PIXEL 方法中並不能正確建立分類器，故不適用。已知若不考慮任何資訊，而以均等機會隨機預測二元結果，則其預測正確率為 50%。故 50% 正確率可視為是分類器最低的基準。我們由表 3-1 發現，無論採用何種核函數，採用所有像素所建立的分類器的平均預測正確率皆很靠近 50%，

而且預測正確率的標準差接近 0，這表示 10 次的預測正確率皆一致性的偏低，也就是說這個方法並未能有效的獲得影像中有用的資訊來建立分類器。另一方面，此法的迭代次數與所需時間過高，代表在建立支持向量機分類器的過程中，難以找出最佳的分類平面來建立好的分類器。故此法顯示，當原始資料過於複雜、繁複，如此處的貓狗影像，若不採取任何特徵萃取或變數刪減，則後續將導致難以建立好的分類器。

表 3-2 列出以主成份分析法選取解釋變異能力高的主成份，並以此來建立分類器的預測結果。首先我們從選出的主成份個數可得知，雖然原始影像包含超過四千個像素，但卻由不到 5% 的變數解釋 90% 的變異。從分類正確率來看，此法非常不理想，與第一個方法的正確率比較相差無幾，甚至較低，訓練所需時間也大幅提升。甚至計算過程中的迭代次數多數都到達 LIBSVM 中預設的上限 10000000，這代表在支持向量機法中搜尋最佳分類平面的最佳化問題，無法得到收斂的結果。其中僅在當採用 Polynomial 核函數、考慮 80% 與 90% 的解釋成份時，產生了稍好的預測結果，其他都非常的差。故直接採用主成份分析所挑選出的主成份並無法提供好的辨識特徵。故當變數具備統計上的高解釋變異能力，並無法對圖像辨識問題提供有效資訊。

第三個方法為方向梯度直方圖 HOG 法，我們將原始影像像素經由此法萃取物體邊緣特徵，假設每 8×8 個像素為一個格子，每 2×2 個格子為一區塊，區塊間有 50% 的重疊率，則灰階化後的 64×64 像素能夠萃取出 1,764 個特徵變數。後續並同樣地建立支持向量機分類器，此時我們增加了 radial basis 及 Sigmoid 兩種類型的核函數，其分析結果列於表 3-3。相較於前兩個方法，當採用 Linear 和 Polynomial 核函數時，此法的正確率可提升至 70% 以上，且訓練所需時間大幅降低，正確率的標準差則維持穩定。但採用 radial basis 與 sigmoid 核函數時，則預測結果仍然表現不佳。故我們發現核函數對分類器的表現影響相當大。當採用適當核函數時，運用方向梯度直方圖法來萃取特徵能夠增進貓狗之間的辨識。

由於透過 HOG 法後可獲得 1764 個特徵，數量仍然偏高，所以我們考慮之後採用主成份分析篩選重要邊緣特徵，其分類器的結果整理於表 3-4。我們發現和表 3-3 相較之下，增加主成份分析導致正確率上些微減損，但大幅度縮短計算時間。例如在 90%解釋比例下，由 1,764 個特徵中保留約 1/5 個（354 個）主成份，在 Linear 和 Polynomial 核函數下的平均正確率由 71.22%，75.46%修正為 72.42%，73.24%，略有增減。但總計算時間則由 650 秒、646 秒降為 289 秒與 214 秒。另一方面，方法三 HOG 法若採用 radial basis 與 sigmoid 核函數，已知其預測結果不佳。但若我們額外採用主成份分析，並做較大幅度刪減變數，也就是採用較低的解釋變異比例時，將改善這兩個分類器的結果。故我們得知，若採取較消極的變數刪減時，使得資料保留較多雜訊將則影響辨識的結果。



表 3-1：5000 張影像以 PIXEL 法之十折交互驗證的預測結果

| 核函數 | 平均正確率 | 標準差 | 總計算時間 (秒) | 平均 迭代次數 |
|------------|--------|-------|--------------|------------|
| Linear | 50.72% | 1.70% | 2831 | 2928000 |
| Polynomial | 56.84% | 1.79% | 1442 | 117426 |

表 3-2：5000 張影像以 PCA 法選取特徵之十折交互驗證的預測結果

| 核函數 | 解釋變異 比例 (變數個數) | 平均正確率 | 標準差 | 總計算時間 (秒) | 平均 迭代次數 |
|------------|----------------------|--------|-------|--------------|------------|
| Linear | 60%(10) | 50.28% | 3.49% | 4989 | 10000000 |
| | 70%(19) | 49.64% | 2.47% | 5352 | 10000000 |
| | 80%(47) | 51.62% | 3.18% | 5479 | 10000000 |
| | 90%(155) | 50.06% | 2.64% | 5590 | 10000000 |
| Polynomial | 60%(10) | 48.10% | 1.14% | 4843 | 10000000 |
| | 70%(19) | 49.16% | 2.20% | 5539 | 10000000 |
| | 80%(47) | 53.48% | 2.49% | 637 | 266005 |
| | 90%(155) | 55.30% | 2.46% | 518 | 52462 |

表 3-3：5000 張影像以 HOG 法選取特徵之十折交互驗證的預測結果

| 核函數 | 平均正確率 | 標準差 | 總計算時間 (秒) | 平均 迭代次數 |
|--------------|--------|-------|--------------|------------|
| Linear | 71.22% | 1.46% | 650 | 131507 |
| Polynomial | 75.46% | 1.54% | 646 | 13089 |
| Radial basis | 51.12% | 0.83% | 787 | 5057 |
| Sigmoid | 50.04% | 0.13% | 784 | 2250 |

表 3-4(a)：5000 張影像以 HOG+PCA 法選取特徵之十折交互驗證的預測結果

| 核函數 | 解釋變異 比例 (變數個數) | 平均正確率 | 標準差 | 總計算時間 (秒) | 平均 迭代次數 |
|--------|----------------------|--------|-------|--------------|------------|
| Linear | 60%(106) | 72.14% | 1.65% | 156 | 104234 |
| | 70%(155) | 72.38% | 1.20% | 194 | 126799 |
| | 80%(228) | 72.80% | 1.58% | 225 | 153507 |
| | 90%(354) | 72.42% | 1.74% | 289 | 209743 |
| | 前 1000 個 主成分 | 70.90% | 1.72% | 455 | 158611 |
| | 前 500 個 主成分 | 70.98% | 1.55% | 362 | 235696 |
| | 前 300 個 主成分 | 72.68% | 1.42% | 269 | 176871 |
| | 前 100 個 主成分 | 72.06% | 1.95% | 130 | 85253 |

表 3-4(b)：5000 張影像以 HOG+PCA 法選取特徵之十折交互驗證的預測結果

| 核函數 | 解釋變異 比例 (變數個數) | 平均正確率 | 標準差 | 總計算時間 (秒) | 平均 迭代次數 |
|------------|----------------------|--------|-------|--------------|------------|
| Polynomial | 60%(106) | 73.06% | 1.54% | 111 | 8426 |
| | 70%(155) | 73.40% | 1.04% | 126 | 8133 |
| | 80%(228) | 73.64% | 1.44% | 166 | 8200 |
| | 90%(354) | 73.24% | 1.86% | 214 | 8371 |
| | 前 1000 個 主成分 | 72.24% | 1.72% | 444 | 8514 |
| | 前 500 個 主成分 | 72.58% | 1.55% | 249 | 8395 |
| | 前 300 個 主成分 | 73.32% | 1.86% | 180 | 8278 |
| | 前 100 個 主成分 | 72.56% | 1.37% | 103 | 8505 |

表 3-4(c)：5000 張影像以 HOG+PCA 法選取特徵之十折交互驗證的預測結果

| 核函數 | 解釋變異 比例 (變數個數) | 平均正確率 | 標準差 | 總計算時間 (秒) | 平均 迭代次數 |
|--------------|----------------------|--------|-------|--------------|------------|
| Radial basis | 60%(106) | 64.06% | 1.78% | 117 | 6404 |
| | 70%(155) | 61.20% | 2.62% | 140 | 5794 |
| | 80%(228) | 58.04% | 1.83% | 159 | 5305 |
| | 90%(354) | 54.24% | 1.39% | 212 | 5252 |
| | 前 1000 個 主成分 | 51.14% | 0.79% | 460 | 5057 |
| | 前 500 個 主成分 | 52.30% | 0.95% | 271 | 5084 |
| | 前 300 個 主成分 | 55.52% | 1.65% | 180 | 5322 |
| | 前 100 個 主成分 | 64.68% | 1.87% | 113 | 6369 |

表 3-4(d)：5000 張影像以 HOG+PCA 法選取特徵之十折交互驗證的預測結果

| 核函數 | 解釋變異 比例 (變數個數) | 平均正確率 | 標準差 | 總計算時間 (秒) | 平均 迭代次數 |
|---------|----------------------|--------|-------|--------------|------------|
| sigmoid | 60%(106) | 62.48% | 2.31% | 80 | 1552 |
| | 70%(155) | 62.82% | 2.40% | 97 | 1699 |
| | 80%(228) | 63.22% | 1.70% | 117 | 1614 |
| | 90%(354) | 64.44% | 2.22% | 147 | 1650 |
| | 前 1000 個 主成分 | 64.22% | 2.08% | 293 | 1663 |
| | 前 500 個 主成分 | 64.08% | 2.73% | 170 | 1487 |
| | 前 300 個 主成分 | 63.20% | 2.90% | 150 | 1557 |

二、 針對貓狗正臉特徵萃取的辨識結果

在上節中，我們發現影像資料中的雜訊量大，經過特徵萃取仍然無法有效提升預測正確率，所有方法的正確率皆不超過 75%。我們觀察到資料集中貓狗的影像是非常的雜亂，貓狗的姿勢不一，有正面、有側面或背面，甚至有些貓狗在影像中僅佔非常小部分不清楚。這些是造成影像資料中雜訊量大的原因，所以導致分類效果不佳的原因。

在這一節中，為了減少資料中的雜訊，我們將對圖片做人工挑選，從各 2500 張貓與狗影像中挑出各 200 張影像，共 400 張，見圖 3-2 與圖 3-3，我們挑選的準則為動物的臉部正面大範圍的呈現在整張影像上，也就是資料中將單純提供動物臉部附近的影像。我們將採取前一節介紹的四種方法在這 400 張貓狗影像上，此時的研究目的為當給定動物臉部影像，研究這四種特徵萃取方法所建立的分類器能否有效辨識貓狗。

四個方法的分析結果分別列於表 3-5 至表 3-8。首先，從表 3-5 的結果看來，當直接採用原始影像的 4096 個像素時，相較於未事前處理影像的結果(表 3-1，平均正確率為 50.72%，56.84%)，在人工挑選的圖像上其平均正確率提高為 65.5%，71.25%。不過由於樣本影像數的減少(從 5000 張圖像減少為 400 張)，標準差因此增加。當資料純粹提供臉部影像時，的確提供了較好的辨識依據。若我們進一步採用主成份分析選取據統計意義的重要變數，如方法二 PCA，則根據表 3-6 的結果，雖然跟表 3-5 比較起來，運算時間可大幅減低，不過預測正確率卻不甚理想，比採用完整資料差相當多。且從迭代次數多數達到上限看來，此時在建造分類器過程中的最佳化求解並未得到收斂結果。

HOG 法的結果整理在表 3-7。我們發現此時此法能夠充分提供貓狗辨識的資訊。在利用 Polynomial 核函數時，分類正確率高達 89.25%，且所需的訓練時間非常短。不過當 radial basis 與 sigmoid 核函數來說，如同前一節的分析結果，其分類器的表現皆不佳。圖 3-4 在採用 Polynomial 核函數下，十次交叉驗證裡被錯誤分類錯誤的影像。我們首先觀察狗的部分，綜觀圖 3-3 中貓的影像，大部分的貓有尖耳朵，而在圖 3-4 中被分類錯誤的狗大部分的耳朵呈現下垂，我們推測狗的垂耳部分較尖，容易被誤判成貓類。在貓的部分，由圖 3-3 中可發現狗的嘴型明顯比貓長且突出，觀察圖 3-4 中被分類錯誤的貓影像，大部份的貓嘴巴與脖子有較明顯的邊緣，故容易被誤判為狗的長嘴。

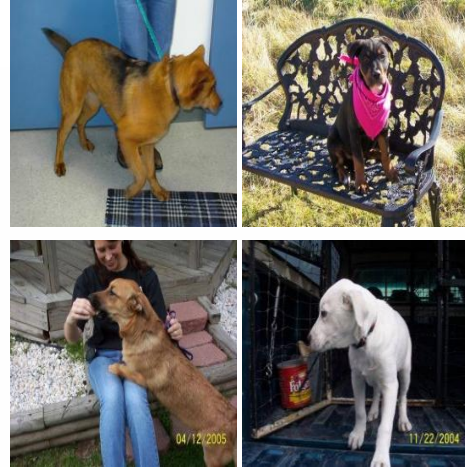
第四個方法 HOG+PCA 法是將前述偵測所得的梯度方向直方圖特徵，在經過主成份分析做變數篩選後的結果，見表 3-8。我們發現原本表現不佳的 Radial basis 與 sigmoid 核函數，在此處有了明顯的改進。除了 radial basis 核函數，解釋變異比例 80%與 90%的組合以外，其他的方法的分類器之平均正確率都提升至約 85%。另外，由於此處僅有 400 張影像，所以透過主成分分析可減少的分析時間並不明顯。但當資料量大時，則預期可節省相當的時間。在此法中，我們發現當主成份分析的解釋變異比例設為 70%時，其相對應的分類正確率普遍較高。

符合正面臉標準



○

不符合正面臉標準



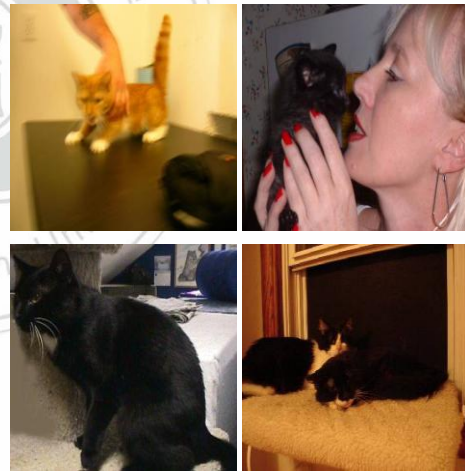
×

符合正面臉標準



○

不符合正面臉標準



×

圖 3-2：符合正面臉的狗影像標準示意圖；符合正面臉的貓影像標準示意圖。



圖 3-3：上方為針對正臉挑選後部分貓影像示意圖；下方為針對正臉挑選後部分狗影像示意圖。

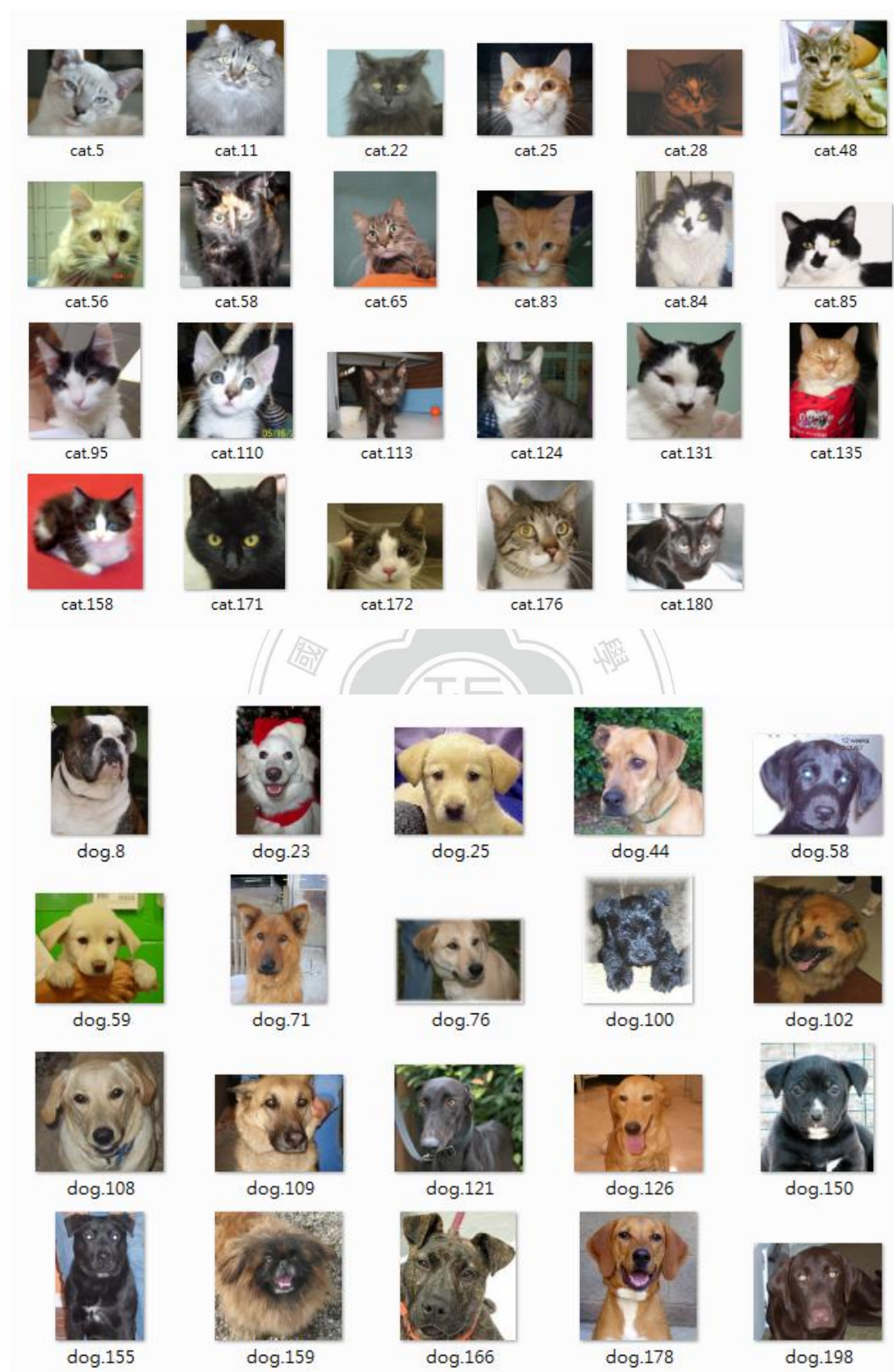


圖 3-4：上方為交叉驗證中被錯誤分類的貓影像；下方為被錯誤分類的狗影像。

表 3-5：400 張影像以 PIXEL 法之十折交互驗證的預測結果

| 核函數 | 平均正確率 | 標準差 | 總計算時間 (秒) | 平均迭代次數 |
|------------|--------|-------|--------------|--------|
| Linear | 65.50% | 6.65% | 10.00 | 5543 |
| Polynomial | 71.25% | 6.03% | 9.50 | 2410 |

表 3-6：400 張影像以 PCA 法之十折交互驗證的預測結果

| 核函數 | 解釋變異 比例 (變數個數) | 平均正確率 | 標準差 | 總計算時間 (秒) | 平均 迭代次數 |
|------------|----------------------|--------|--------|--------------|------------|
| Linear | 60%(9) | 52.75% | 4.92% | 435 | 10000000 |
| | 70%(16) | 49.50% | 4.68% | 443 | 10000000 |
| | 80%(34) | 53.25% | 10.74% | 457 | 10000000 |
| | 90%(80) | 59.50% | 10.53% | 460 | 10000000 |
| Polynomial | 60%(9) | 53.00% | 8.72% | 313 | 10000000 |
| | 70%(16) | 60.75% | 7.36% | 9.45 | 6549 |
| | 80%(34) | 58.50% | 8.83% | 9.24 | 2138 |
| | 90%(80) | 61.25% | 8.60% | 9.35 | 1700 |

表 3-7：400 張影像以 HOG 法之十折交互驗證的預測結果

| 核函數 | 平均正確率 | 標準差 | 總計算時間 (秒) | 平均 迭代次數 |
|--------------|--------|--------|--------------|------------|
| Linear | 84.50% | 3.69% | 3.75 | 1081 |
| Polynomial | 89.25% | 2.37% | 4.54 | 591 |
| Radial basis | 57.75% | 16.35% | 5.73 | 488 |
| sigmoid | 53.75% | 11.13% | 5.79 | 180 |



表 3-8：400 張影像以 HOG+PCA 法之十折交互驗證的預測結果

| 核函數 | 解釋變異 比例 (變數個數) | 平均正確率 | 標準差 | 總計算時間 (秒) | 平均 迭代次數 |
|--------------|----------------------|--------|--------|--------------|------------|
| Linear | 60%(62) | 86.75% | 2.65% | 2.98 | 4530 |
| | 70%(87) | 85.50% | 3.50% | 3.03 | 4607 |
| | 80%(122) | 85.50% | 2.58% | 3.27 | 3808 |
| | 90%(178) | 83.75% | 4.89% | 3.48 | 2459 |
| Polynomial | 60%(62) | 85.50% | 2.30% | 3.08 | 763 |
| | 70%(87) | 85.75% | 2.37% | 3.07 | 774 |
| | 80%(122) | 84.75% | 3.81% | 3.37 | 761 |
| | 90%(178) | 84.75% | 4.32% | 3.38 | 742 |
| Radial basis | 60%(62) | 86.50% | 3.76% | 3.03 | 426 |
| | 70%(87) | 86.60% | 5.83% | 3.36 | 521 |
| | 80%(122) | 59.50% | 15.58% | 3.27 | 511 |
| | 90%(178) | 58.00% | 16.91% | 3.55 | 512 |
| sigmoid | 60%(62) | 83.75% | 5.17% | 2.99 | 126 |
| | 70%(87) | 85.00% | 3.54% | 2.91 | 126 |
| | 80%(122) | 82.75% | 4.63% | 2.97 | 126 |
| | 90%(178) | 83.25% | 3.55% | 3.12 | 120 |

第四章 結論與建議

本論文應用方向梯度直方圖與主成份分析在萃取貓狗影像上的特徵，並將偵測到的特徵建立支持向量機分類器以辨識貓狗。在我們的實證分析中發現方向梯度直方圖所萃取出的邊緣特徵能夠提供有用的資訊，當邊緣特徵過多時，則主成份分析能夠有效篩選重要變數。但當原始影像中充斥太多雜訊時，則特徵萃取的作用有限。過多的雜訊將使得支持向量機分類器的建立產生困難，或者預測的效果不佳。

通常動物的臉對動物辨識提供了有效的資訊，為了增進預測能力，可考慮在分析中加入『臉辨識』處理程序，以降低其他雜訊。在本文的實證中採用人工挑選具備清晰的貓狗臉部的影像，我們發現經過該處理，再採用梯度方向直方圖與主成份分析進行兩步驟特徵萃取，可增進分類正確率至 85% 以上。

本文的重點雖在特徵萃取，但我們發現『臉辨識』將能大大助益後續動物辨識。如何針對貓狗或動物影像進行『臉辨識』將是有用的研究問題。另外其他的事前處理如有關背景前景的區隔等，也是重要課題。人類及動物的臉部辨識研究請見 Kim 等人(2003)及 Burghardt 與 Calic (2006)。

除了物體邊緣外，影像上還有其他重要特徵，例如花色、材質等。如何透過資料分析萃取出這些重要特徵以應用在分類問題為未來可能研究方向。另外在機器學習領域中已發展眾多分類法，希望未來能夠有效結合不同種類的方法，發展出更好，運算時間越低，辨識成功率越高的辨識系統，來增加自動辨識系統的穩定性以及快速性，以及同時能夠處理及辨識多種類的動物。

最後對於 Assira 的辨識碼，從觀察實證分析的結果，能夠發展出有效的自動辨識法，不需要人工辨識，藉由來電腦就能克服辨識貓狗影像，而除了特徵萃取，影像的事前處理也是一門重要的課題。所以相信在未來，有各種類似的影像與辨識問題，隨著許多資訊工具的多樣化，未來的挑戰都能經由時間一一被克服。

參考文獻

- A. Rencher, (1995), "*Methods of multivariate analysis*," New York : John Wiley.
- C.-C. Chang and C.-J. Lin, (2001), "LIBSVM: a library for support vector machines," Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- C. Cortes and V. Vapnik, (1995), "Support vector networks," *Machine Learning*, 20, 273–297.
- D. L. Schwartz, (1995), "Reasoning about the referent of a picture versus reasoning about the picture as the referent: An effect of visual realism," *Memory and Cognition*, 23, 709–722.
- D. G. Lowe., (2004), "Distinctive image features from scale-invariant keypoints," *IJCV*, 60, 91-110.
- H. Bay, T. Tuytelaars, and L. Van Gool, (2006), "Surf: Speeded up robust features," In *European Conference on Computer Vision*.
- I. Kim, J. H. Shim, and J. Yang, (2003), "Face detection," Stanford University, Tech. Rep. , eE368 Final Project Report.
- I. T. Jolliffe, (1986), "Principal Component Analysis," Springer-Verlag, New York.
- J. Elson, J. R. Douceur, J. Howell and J. Saul, (2007), "Asirra: a CAPTCHA that exploits interest-aligned manual image categorization," *Proceedings of the 14th ACM conference on Computer and communications security*, Alexandria, Virginia, USA .
- L. von Ahn, M. Blum, N. J. Hopper and J. Langford, (2003), "CAPTCHA: using hard AI problems for security," In *Lecture notes in computer science*, Berlin: Springer, 294–311.

- L.I. Smith, (2002), “A tutorial on Principal Components Analysis,” Cornell University, USA.
- N. Dalal, B. Triggs, (2005), “Histograms of Oriented Gradients for Human Detection,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2, 886-893.
- P. Domingos, (2012), “A few useful things to know about machine learning,” Commun. ACM. 55, 78–87.
- S.-Y. Huang, Y.-K. Lee, G. Bell, Z.-H. Ou, (2010), “An efficient segmentation algorithm for CAPTCHAs with line cluttering and character warping,” Multimedia Tools and Applications, 48, 267-289.
- S. Süsstrunk, R. Buckley and S. Swen, (1999), “Standard RGB color spaces,” Proc. IS T/SID 7th Color Imaging Conf., 127-134.
- T. Burghardt and J. Calic. , (2006), “Analysing animal behaviour in wildlife videos using face detection and tracking,” IEEE Proceedings - Vision, Image, and Signal Processing, 153, 305-312.

統國

立
計

政

研

治

究

大

所學

碩
士
論
文

貓
狗
影
像
辨
識
之
特
徵
萃
取

鍾
立
強

撰

一
百
零
五
年
七
月

