

國立交通大學

電機學院 電機與控制學程

碩士論文

生活照片之人物分類系統研究

A Study on Face Recognition System of Photographs

研究生：程煌瑞

指導教授：林昇甫 博士

中華民國一百年七月

生活照片之人物分類系統研究  
A Study on Face Recognition System of Photographs

研 究 生：程 煌 瑞

Student：Huang-Jui Cheng

指導教授：林 昇 甫 博士

Advisor：Dr. Sheng-Fuu Lin

國 立 交 通 大 學  
電機學院 電機與控制學程  
碩 士 論 文

The logo of National Chiao Tung University is a circular emblem. It features a gear-like outer ring. Inside the circle, there is a stylized building or structure. The year '1896' is inscribed at the bottom of the inner circle. The text 'A Thesis' is written across the middle of the logo.

A Thesis  
Submitted to College of Electrical and Computer Engineering  
National Chiao Tung University  
in partial Fulfillment of the Requirements  
for the Degree of  
Master of Science  
in  
Electrical and Control Engineering  
July 2011  
Hsinchu, Taiwan, Republic of China

中 華 民 國 一 百 年 七 月

# 生活照片之人物分類系統研究

學生：程煌瑞

指導教授：林昇甫 博士

國立交通大學 電機學院 電機與控制學程 碩士班

## 摘 要

本論文採用膚色偵測法找出彩色生活照片可能的人臉位置，透過賈伯小波(Gabor wavelet)抽取特徵進行類神經網路訓練，以判斷是否為人臉。在人臉辨識階段，先使用主動式外觀模型(active appearance model)和可導引濾波器(steerable filter)進行人臉正規化，接下來採用稀疏編碼(sparse coding)演算法，在五個訓練樣本下，生活照人臉辨識率可達 80%，使用其他正面人臉資料庫(AR 資料庫)辨識率更可高達 98%。並提出直方圖統計法來減少稀疏編碼的權重數目為原來的 60%，除了降低系統運算量，同時特徵向量仍然具有代表性。整體而言，可適用於家庭數位相簿管理，或數位相框之分類系統。

# **A Study on Face Recognition System of Photographs**

**Student : Huang-Jui Cheng**

**Advisor : Dr. Sheng-Fuu Lin**

**Degree Program of Electrical and Computer Engineering**

**National Chiao Tung University**

The logo of National Chiao Tung University is a circular emblem. It features a gear-like outer border. Inside, there is a stylized building or structure with the letters 'ES' and 'A' on it. Below this, the year '1896' is inscribed. The word 'ABSTRACT' is overlaid in a bold, serif font across the center of the logo.

## **ABSTRACT**

This thesis adopts skin-color model to find the candidate face region, then Gabor wavelets transformation is adopted to extract the entire face features. Afterward, neural network is trained to determine whether the candidate region is a human face or not. Finally, this thesis adopts active appearance model and steerable filter to normalize all faces for face recognition. Then this thesis implements sparse coding algorithm with 5 training faces to increase the face recognition rate up to 80% for photographs, and for frontal face of AR database also increases by 98%. Furthermore, this thesis proposes using histogram method to reduce 60% of sparse coding needed which also reduces the amount of system computational cost, and then the features are still representative. As a whole, this system is suitable for digital media classification of family photograph albums or digital photograph frames.

## 誌 謝

首先要感謝指導教授 林昇甫博士細心的指導，除了指正學生論文的謬誤之處，還有生活處事上的叮嚀，讓學生獲益良多。有了 林教授的體諒跟鼓勵，學生才可以兼顧工作和學業，完成此論文。同時也感謝口試委員 蘇建焜教授跟 陳肇業教授的提問與寶貴意見，修正學生論文錯誤的地方。

也要感謝 九暘電子主管 陳忠正副總的體諒，同意讓學生在工作之餘繼續進修。感謝父母親的養育及栽培，還有女友 蔡佳偉在精神上的支持，即使工作再忙再累，也讓學生能夠堅持到底完成論文。

還要感謝公司同仁提供的生活照片，讓學生的測試工作可以順利進行。感謝曾經幫助學生的親朋好友們，謝謝您們。



# 目 錄

摘 要 .....	i
ABSTRACT .....	ii
誌 謝 .....	iii
目 錄 .....	iv
圖 目 錄 .....	vi
表 目 錄 .....	viii
第一章 緒論 .....	1
第二章 相關知識及理論 .....	6
2.1 人臉位置偵測 .....	7
2.1.1 膚色位置偵測 .....	7
2.1.2 人臉特徵抽取 .....	9
2.1.3 類神經網路 .....	10
2.2 人臉正規化 .....	11
2.2.1 主動式外觀模型演算法 .....	11
2.2.2 光線明亮度正規化 .....	14
2.3 人臉辨識 .....	23
第三章 人臉辨識系統 .....	27
3.1 人臉位置偵測方法 .....	27
3.1.1 多重尺寸視窗搜尋 .....	29
3.1.2 訓練類神經網路 .....	29
3.2 光線明亮度正規化濾波器 .....	30
3.3 稀疏編碼權重更新疊代演算法 .....	30
第四章 實驗結果與分析 .....	34
4.1 人臉偵測的實驗結果 .....	34
4.2 特徵人臉辨識率 .....	38
4.3 稀疏編碼人臉辨識率 .....	42
4.4 明亮度正規化對於辨識率的影響 .....	49
4.5 可導引濾波器正規化的人臉辨識率 .....	50

4.6 人臉解析度對於辨識率的影響 .....	56
4.7 生活照片人臉辨識率分析 .....	56
4.8 實驗結果與討論 .....	57
<b>第五章 結論與未來展望 .....</b>	<b>59</b>
<b>參 考 文 獻 .....</b>	<b>60</b>



# 圖 目 錄

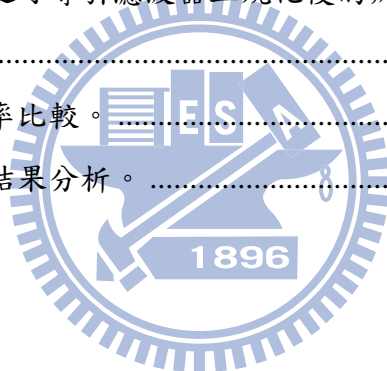
圖 1.1 人臉資料庫建構圖。 .....	1
圖 1.2 人臉辨識系統架構圖。 .....	1
圖 2.1 典型生活照，(a)偵測人臉位置；(b)擷取人臉外形；(c)人臉正規化。 .....	8
圖 2.2 人臉與非人臉分類之類神經網路架構圖。 .....	11
圖 2.3 區塊三乘三像素值計算範例。 .....	16
圖 2.4 計算x方向梯度 $G_x$ 範例，(a) $f(x, y)$ 像素點；(b) $f(x, y)$ 往x方向位移兩個像素； (c) $f(x, y)$ 在x方向最右邊兩列像素為零；(d)梯度 $G_x$ 為(b)減掉(c)。 .....	17
圖 2.5 計算y方向梯度 $G_y$ 範例，(a) $f(x, y)$ 像素點；(b) $f(x, y)$ 往y方向位移兩個像素； (c) $f(x, y)$ 在y方向最下面兩行像素為零；(d)梯度 $G_y$ 為(b)減掉(c)。 .....	18
圖 2.6 五乘五遮罩矩陣計算範例，(a)先取一個五乘五區域矩陣；(b)計算後求得的遮罩矩陣。 .....	20
圖 2.7 六乘六矩陣之ZigZag(蜿蜒)排列範例。 .....	22
圖 2.8 (a)待測人臉；(b)訓練樣本；(c)稀疏編碼計算二十個係數；(d)重建後的人臉。 .....	24
圖 3.1 人臉位置偵測流程圖。 .....	28
圖 3.2 平均人臉。 .....	29
圖 3.3 最左邊是原始人臉，其他經過明亮度濾波器處理後的結果，由左而右分別是SSR，MSR，ASSR，HOMOF，SSSQI，MSSQI，DCT和SF。 .....	30
圖 3.4 選擇權重位置比較圖，(a)統計法 1；(b)統計法 2；(c)門檻法。 .....	31
圖 3.5 兩種權重選擇統計法在辨識率上面的比較，(a)由上往下取樣；(b) 由下往上取樣。 .....	32
圖 4.1 生活照片人臉偵測的部份實驗結果，(a)正面人臉；(b)有遮掩物的正面人臉；(c)有明亮度不均勻的正面人臉；(d)有大角度變化的人臉。 .....	35
圖 4.1(續) 生活照片人臉偵測的部份實驗結果。 .....	36
圖 4.2 (a)中間人臉因為眼鏡反光大及仰角太大而無法偵測；(b)右邊人臉因為仰角太大而無法偵測；(c)中間後面人臉因為遮掩物而無法偵測；(d)下面人臉因為光線不均勻而無法偵測；(e)因為傾斜角度太大及照明問題而無法偵測。 .....	38
圖 4.3 兩種取樣訓練之特徵人臉辨識率比較，(a)由上往下取樣；(b) 由下往上取樣。 .....	40
圖 4.4 生活照特徵人臉兩個訓練樣本。 .....	40



圖 4.5 AT&T ORL 資料庫特徵人臉三個訓練樣本。 .....	41
圖 4.6 AT&T ORL 資料庫特徵人臉五個訓練樣本。 .....	41
圖 4.7 (a)從AR資料庫任意取出 250 張人臉做實驗；(b)從生活照資料庫任意取出 250 張人臉做實驗；(a)(b)的第一列到第十列分別代表每一個類別的人臉，每一類別有十張人臉，其餘十五列共 150 張人臉代表未知人臉類別。 .....	44
圖 4.8 (a)(b)訓練樣本分別代表AR資料庫每一類別的編號 1~2 人臉，和 10~9 人臉；(c)訓練樣本代表未知人臉類別編號 1~50 人臉；(a)跟(b)差異在於明亮度，表情差異不大；(c)包含明亮度高與明亮度低的訓練樣本。 .....	45
圖 4.9 AR 資料庫經過多尺寸視網膜正規化後得到的訓練樣本。 .....	46
圖 4.10 (a)生活照資料庫的訓練樣本每一類別編號 1~4 人臉；(b)未知人臉類別的訓練樣本編號 1~50 人臉；(a)(b)包含明亮度高與明亮度低的訓練樣本。 .....	47
圖 4.11 生活照經過多尺寸視網膜正規化後得到的訓練樣本。 .....	48
圖 4.12 多尺寸視網膜正規化辨識成功的例子，最左邊是原始人臉，左二是經過正規化的人臉，右二是稀疏編碼重建的人臉，最右邊是最靠近類別的人臉。 .....	49
圖 4.13 多尺寸視網膜正規化辨識失敗的例子，最左邊是原始人臉，左二是經過正規化的人臉，右二是稀疏編碼重建的人臉，最右邊是最靠近類別的人臉。 .....	49
圖 4.14 (a)八種明亮度處理演算法之辨識率曲線比較圖；(b)最左邊對應的是Original為原始人臉，最右邊對應的是SF為可導引濾波器處理後的人臉，由左而右分別是SSR，MSR，ASSR，HOMOF，SSSQI，MSSSQI和DCT。 .....	50
圖 4.15 兩種取樣訓練之可導引濾波器(SF)與多尺寸視網膜(MSR)辨識率曲線比較圖，(a)由上往下取樣；(b) 由下往上取樣。 .....	52
圖 4.16 可導引濾波器辨識失敗的例子，最左邊是原始人臉，左二是經過正規化的人臉，右二是稀疏編碼重建的人臉，最右邊是最靠近類別的人臉。 .....	53
圖 4.17 可導引濾波器辨識成功的例子，最左邊是原始人臉，左二是經過正規化的人臉，右二是稀疏編碼重建的人臉，最右邊是最靠近類別的人臉。 .....	53
圖 4.18 生活照經過可導引濾波器正規化後的訓練樣本。 .....	54
圖 4.19 AR 資料庫經過可導引濾波器正規化後的訓練樣本。 .....	55
圖 4.20 生活照片之人物分類系統說明，(a)人臉偵測結果，(b)擷取的人臉外形，(c)主動式外觀模型正規化的人臉，(d)最右邊人臉是男生錯誤的分類結果，(e)最右邊人臉是女生正確的分類結果，(f)取自資料庫的人臉當做辨識結果。 .....	58

# 表 目 錄

表 4.1 生活照片人臉偵測結果分析。 .....	37
表 4.2 由上往下取樣訓練之特徵人臉辨識率在AT&T ORL資料庫跟生活照片的比較。 .....	39
表 4.3 由下往上取樣訓練之特徵人臉辨識率在AT&T ORL資料庫跟生活照片的比較。 .....	40
表 4.4 由上往下取樣訓練之稀疏編碼辨識率在AR資料庫跟生活照片的比較。 .....	42
表 4.5 由下往上取樣訓練之稀疏編碼辨識率在AR資料庫跟生活照片的比較。 .....	42
表 4.6 由上往下取樣訓練之可導引濾波器正規化後的辨識率在AR資料庫跟生活照片的 比較。 .....	51
表 4.7 由下往上取樣訓練之可導引濾波器正規化後的辨識率在AR資料庫跟生活照片的 比較。 .....	51
表 4.8 人臉解析度之辨識率比較。 .....	56
表 4.9 生活照片人臉辨識結果分析。 .....	57



# 第一章 緒論

本論文研究之動機，主要是想以圖找圖，用已知圖形尋找相似圖形，以市面上產品應用狀況來看，數位相框之生活照人物分類，或搜尋及播放相同人物的生活照，都是一個可以著力的地方，只要分類正確，接下來的操作功能就可以依照使用者的需求來達成。

人臉辨識是人類與生俱有的能力，透過眼睛擷取清晰的人臉畫面，經由大腦訓練記憶，爾後再從記憶資料區塊找出匹配的人臉，完成辨識。人腦辨識需要記憶，電腦視覺辨識更是需要，所以本系統也必須在辨識前先經過訓練和記憶的程序。

生活照之人臉辨識和照片分類，必須先建立資料庫(圖 1.1)，以提供辨識系統的訓練樣本，首先偵測生活照片人臉的位置，經過人臉擷取正規化，以手動方式分類每個人物的人臉照片。接下來才是人臉辨識工作(圖 1.2)，也是先找出測試生活照片裡面的所有人臉，經過人臉正規化跟特徵抽取，進一步比對出最接近資料庫的哪一個人臉，最後就可以記錄資料庫每個人物出現在哪幾張生活照片裡。

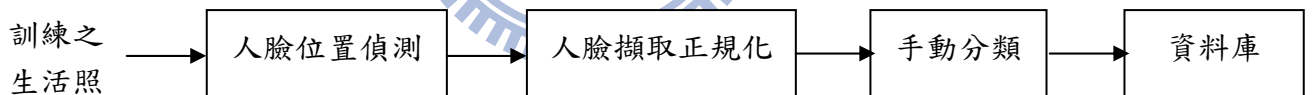


圖 1.1 人臉資料庫建構圖。

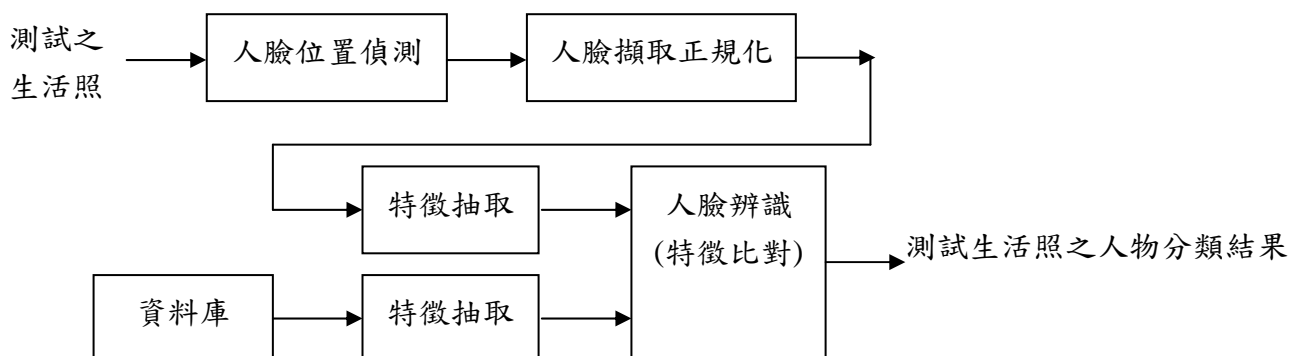


圖 1.2 人臉辨識系統架構圖。

人臉位置偵測目前有統計模型法[1]、類神經網路法[2]、支持向量機分類法[3]、推昇(boost)分類法[4]、眼睛偵測法[5]和膚色偵測法[6]等等方法。主要可分成模型基礎法和膚色形狀基礎法兩大方法，代表性的有，S. Birchfield [7]提出橢圓追蹤法，使用梯度跟膚色兩種特徵模型來偵測人臉位置。L. Wiskott 等四人[8]提出彈力束圖形配對法(Elastic Bunch Graph Matching, EBGM)，透過眼睛、鼻子、嘴巴和輪廓由一系列小波成份組成，可以計算人臉旋轉角度，正規化到正面人臉後再做人臉辨識。T. F. Cootes 等四人[9]提出主動式形狀模型法，使用手動描繪特徵點形成分佈模型來當作訓練樣本，當待測樣本如手掌形狀改變或人臉旋轉角度改變，特徵點仍然可以掌握待測樣本主要的形狀特徵。之後，T. F. Cootes 等三人[10]提出主動式外觀模型(active appearance model)演算法，包含形狀統計模型以及人臉的灰階外觀，疊代計算訓練樣本跟合成影像之間的殘餘值(residual errors)直到最小值，即可得到良好的人臉外形定位效果。D. Maio 和 D. Maltoni 兩人[11]發表粗略跟精細兩階段統計偵測人臉位置。H. S. Lee 和 D. Kim 兩人[12]發表膚色跟臉形兩種分別偵測和追蹤人臉的方法。M. N. Francesc 等三人[13]提出一個費雪彩色空間(Fisher color space)，即便是光線產生變化，使用線性識別分析法也能夠將目標物跟背景分離。Y. Tong 等四人[14]結合賈柏小波(Gabor wavelets)跟灰階值特徵，使用切換式猜測量測模型來估測人臉 3D 立體旋轉角度，以達到人臉偵測及追蹤效果。J. Tu, H. Tao 和 T. Huang 兩人[15]使用人臉 3D 立體模型追蹤來模擬個人電腦裡面屬於人機介面的滑鼠，追蹤人臉位置來移動滑鼠位置，偵測嘴巴的形狀來實現滑鼠的打開、關閉和拖曳等功能。M. Kim 等四個人[16]在追蹤人臉上，使用一些視覺限制條件達成適應性外觀追蹤法，最後使用隱藏式馬可夫模型(hidden Markov model)來辨識人臉。Zheng 和 S. M. Bhandarkar 兩人[17]結合適應性粒子群演算法跟適應性推昇分類法(AdaBoost)[1]來偵測及追蹤人臉。M. Balasubramanian 等三人[18]先使用膚色偵測法找出人臉位置，再透過輻狀基底類神經網路找出精確的眼睛跟嘴巴位置。

人臉辨識目前有統計法[19]、人臉子空間辨識法[20]、外觀基底辨識法[21]、3D 立體辨識法[22][23]和稀疏編碼(sparse coding)[24][25]等辨識方法。P. N. Belhumeur 等三人[26]提出費雪人臉(Fisherfaces)，使用費雪線性鑑別分析法來辨識人臉類別，費雪人臉錯誤率

比特徵人臉(eigenfaces)小。A. R. Chowdhury 等四人[22]透過一連串人臉影像，只要使用一般模型，並採用馬可夫鏈蒙地卡羅法(Markov chain Monte Carlo)取樣機制最佳化，就可以重建 3D 立體人臉，以進行不同角度的人臉辨識。S. Baker 和 T. Kanade 兩人[27]提出新的超解析度演算法，將低解析度影像重建為高解析度影像，這有助於人臉偵測及辨識，特別適用於一般解析度不高的監視系統。X. Liu 等三人[28]提出特徵空間(eigenspace)之人臉訓練樣本是由一群衰變參數來控制，以選擇多一點的新樣本和少部份的舊樣本進行訓練，再做人臉辨識。G. Shakhnarovich 等三人[29]使用近期積分(late integration)策略和早期積分(early integration)策略，以及相互子空間(mutual subspace)法來處理長時間所觀察到的同一個人臉資料，之後使用庫爾貝克-萊伯勒差異(Kullback-Leibler divergence)法來衡量待測人臉跟哪一個已知類別比較相似。W. Y. Zhao 和 R. Chellappa 兩人[30]提出遮光人臉的源頭遮光(source-from-shading)新處理法，因為是對稱人臉，所以可以使用自比例(self-ratio)影像法來處理陰影問題，最後合成影像從密度資訊得到的每一個點都存在著唯一整體解，而且還是唯一局部解。Y. Li 等三人[31]使用核心鑑別分析(kernel discriminant analysis)法來辨識不同角度人臉，將高維度特徵空間投影到等效面，然後就可以採用線性鑑別分析法來處理這個非線性辨識問題，目的就是將同一類別的距離拉得更近，不同類別的距離拉得更遠。Y. Zhang 和 A. M. Martinez 兩人[19]提出權重統計法來辨識人臉，將每張人臉切割成  $K$  個小圓形做特徵抽取，仍然可以辨識墨鏡或圍巾等遮掩物的干擾。N. Vaswani 和 R. Chellappa 兩人[32]提出新的主成份零空間(null space)分析法進行分類，計算所有類別之同類別變異量，取最小值之類別視為零空間，也就是趨近零空間的意思，可應用於人臉分類。J. Wright 等五人[24]提出稀疏編碼來處理正面人臉部份被遮掩的辨識問題，使用  $l_1$ -norm 最小化讓訓練人臉線性地與稀疏誤差值相加，得到的就是遭受遮掩後的重建人臉。

人臉光線不均勻的處理方式大致上有濾波器法[33][34]、流型模型(manifold model)統計法[35][36][37]和 3D 立體模型法[23][38]。W. T. Freeman 和 J. B. Tenenbaum 兩人[39]提出雙線性模型學習法來解決色彩明亮度跟未知外形兩大問題。Y. Adini 等三人[40]提出類似賈柏小波濾波器讓人臉光線均勻化。D. W. Jacobs 等三人[41]使用 Hessian 矩陣找



出來自相同人物的兩張人臉梯度比值，來判斷是否屬於同一類別。M. Savvides 等三人[42]提出 Corefaces 進行人臉辨識，使用相關濾波器(correlation filter)降低光線不均勻問題，再使用主成份分析法做人臉分類。S. Du 和 R. Ward 兩人[33]使用小波濾波器正規化每個人臉，光線均勻正規化後再做人臉辨識。T. Ojala 等三人[34]提出局部二位元圖樣(local binary patterns)抽取紋理特徵，不需複雜計算，然後使用多重解析分析法(multiresolution analysis)做人臉分類。X. Tan 和 B. Triggs 兩人[43]提出局部三位元圖樣(local ternary patterns)，是由局部二元圖樣延伸而來的，使用局部距離轉換取代原來直方統計法，來計算相似性距離以做人臉分類。A. S. Georgiades 等三人[21]使用三張固定角度的訓練樣本，以及反照率來重建光線均勻的人臉，然後再採用生成模型(generative models)估計聯合概率分佈，即可進行良好的人臉辨識。T. Vetter 和 T. Poggio 兩人[44]提出一張 2D 平面人臉樣本就可以重建 3D 立體人臉，條件為使用 49 個同一類別的平面人臉做訓練。S. Malassiotis 和 M. G. Strintzis 兩人[45]提出人臉角度跟光線補償方法，利用對稱基底內插法將人臉旋轉到正面位置，使用支持向量機回歸方程式計算近似的光線方向再做補償。R. Gross 等三人[46]使用像素密度做為特徵，而提出特徵光場(eigen light-field)和費雪光場(Fisher light-field)，針對不同人臉角度跟明亮度進行辨識。O. Arandjelovic 和 R. Cipolla 兩人[35]提出的明亮度補償是結合粗略的直方圖統計校正法跟精細的流型模型(manifold model)法，分解每一個外觀，輸入高斯姿態分類器來判斷人臉角度，再將人臉正規化到正面角度進行人臉辨識。O. Arandjelovic 等五人[36]提出流型密度相異值(manifold density divergence)辨識方法，就是透過統計方程式將這個差異值最小化。O. Arandjelovic 和 R. Cipolla 兩人[37]使用三個高斯姿態分類器並結合相似度統計分析法來判斷人臉角度，運用區域射線強度(gamma intensity)稍微補償了明亮度的變化，更精細的補償就是結合已學習之明亮度變化的線性流型，以及人臉樣本分佈的限制條件來完成明亮度正規化。V. Blanz 和 T. Vetter 兩人[38]提出只要單一張 2D 平面人臉影像，經過 3D 立體形變模型(morphable model)訓練，調整外形和紋理，即可重建出 3D 立體人臉，接下來可以把待測人臉旋轉到一個固定正面角度，再做辨識。D. Jiang 等六人[23]提出只要單一張 2D 平面人臉，即可有效率地結合 3D 立體模型，透過主成份分析法，壓縮 3D 立體模型

的外觀，合成後的人臉可以應付姿態角度、明亮度和表情(PIE)這三個大幅度變化的辨識問題。

本論文章節架構安排如下，第二章介紹論文使用到的相關技術和方法，包括人臉位置偵測[6]、人臉正規化[9][10][47][48][49]以及人臉辨識[50][24][25]。第三章則說明我們改善後的辨識系統。第四章呈現實驗結果與分析。最後第五章做一個總結以及後續研究的方向與建議。



## 第二章 相關知識及理論

人臉辨識首要工作就是擷取人臉的部位，本論文採用膚色偵測法[6]縮小搜尋的範圍，透過賈柏小波[52]抽取特徵，再結合類神經網路訓練[2]，以判斷搜尋的區域是否符合人臉的條件。儘可能把所有入臉的位置找出來，對於後續的人臉辨識是有幫助的。

人臉正規化(normalize)是對於訓練樣本和測試樣本有三種一致性處理，包含定位、尺寸縮放和明亮度[48][49][51]三種做法。其中定位[9][10][47]處理通常都是以雙眼為參考點的，譬如主動式形狀模型(active shape model, ASM)演算法[9]或眼部偵測法[5]，更重要的是人臉尺寸縮放後，雙眼的距離仍然維持固定值。此外，生活照的光源本來就是一個不可控制的場域，舉凡室內、室外、晴天、烏雲或光源照射的角度都是影響的因素，所以明亮度一致性也是必備的處理程序。

最後工作才是人臉辨識，Turk-Pentland 在 1991 年提出特徵人臉(eigenfaces)[50]，在人臉辨識上貢獻，就是利用主成分分析法(principal component analysis, PCA)計算一組基底人臉的線性組合，重建人臉可以透過每個不同權重的特徵人臉加總起來，並有效地將影像維度降低，不過，辨識率會隨著人臉表情的變化、髮型改變、角度不同、眼鏡或圍巾遮掩物等因素而降低。

另外賈柏小波(Gabor wavelets)[52][53]在人臉辨識上貢獻，在於賈柏小波使用二維特徵抽取，並結合類神經網路(neural networks)計算神經元的權重，進行人臉分類。通常，賈柏小波的二維特徵向量取自五種尺寸、八個方向，來保留人臉像素前後左右的關係，而重建人臉必須仰賴足夠多的特徵向量才不會大量失真，因此特徵維度很高，計算量也比較大。為了降低特徵維度以及處理特徵向量最佳化，稀疏表述(sparse representation)[54]提供了一個降低特徵維度的方法，針對人臉遭受 50% 遮掩物，仍然有 87% 的辨識率。

稀疏表述或稀疏編碼(sparse coding)[24][25]在辨識上貢獻，就是利用  $l_1$ -norm 最小化，讓重建後影像的每個像素點跟原影像的誤差達到最小值，對於所屬訓練類別誤差最小化，但是也盡量拉開跟其他類別的距離，這樣即使受到光線或遮掩物等干擾，辨識率也不會快速下降[25]，所以計算測試人臉在訓練類別上的每一組係數向量，就顯得很重



要了。此演算法目前在辨識正面人臉、表情些微變化或是明亮度微幅改變都有不錯的辨識率，只不過還有一些限制條件，就是無法精準辨識非正面人臉，尤其是大角度的改變，因為訓練樣本並沒有建構三維的人臉資料[24]。

本章節架構安排如下，2.1 節描述人臉位置偵測。2.2 節說明人臉正規化方法，包括人臉定位、尺寸縮放以及濾波器正規化處理光線明亮度的方法。最後 2.3 節闡述人臉辨識演算法。

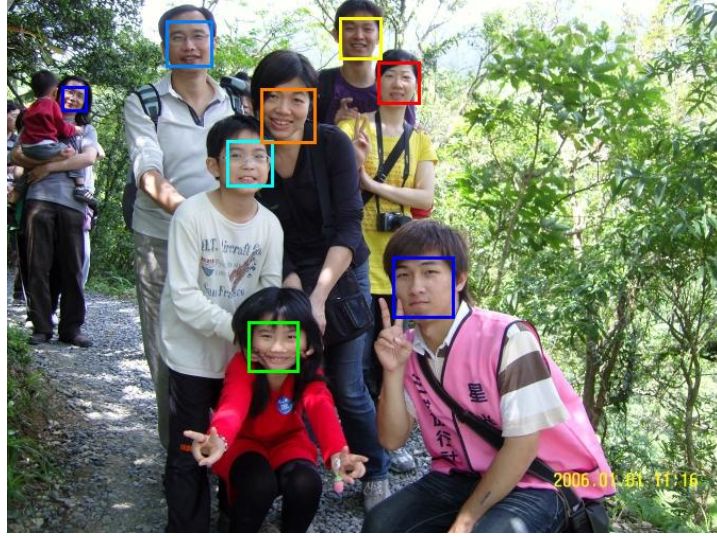
## 2.1 人臉位置偵測

圖 2.1 是一張很典型的生活照，拍照的時候，有些人站在前面，有些人站在後面，光線明暗度不一樣，雖然人臉傾斜一個角度，但還是屬於可辨識的範圍。本系統採用膚色偵測方法[6]縮小人臉掃描範圍，使用類神經網路訓練人臉跟非人臉樣本，透過賈柏小波[52]抽取特徵，找到人臉位置，如圖 2.1(a)長方形框起來的位置，主要有八個人臉，長寬像素都不一樣，圖 2.1(b)是使用主動式外觀模型(active appearance model, AAM)[9][10][47]擷取人臉外形，圖 2.1(c)也是使用主動式外觀模型將人臉旋轉及尺寸縮放到相同比例，最右邊人臉是一個比較不理想的正規化，因為包括了背景雜訊。

### 2.1.1 膚色位置偵測

要在彩色生活照片尋找人臉，通常需要依序掃描每一個位置，檢查不同尺寸大小的區域是否符合人臉條件，這樣的運算是相當耗時的，如果可以去除非膚色區域以減少運算時間，將可增加搜尋速度。膚色特徵隨著光線照射角度、相機參數、人種和身體不同位置而有不同的特性，所以選擇合適的膚色彩色空間才能分類好膚色跟非膚色區域。

*RGB* 用於數位影像的色彩空間，分別代表紅色、綠色和藍色，為了減少明亮度不均勻還有不同人種等因素，通常會把 *RGB* 正規化為  $R + G + B = 1$ ，具有低變化的特性，有助於膚色偵測[6]。



(a)



(b)



(c)

圖 2.1 典型生活照，(a)偵測人臉位置；(b)擷取人臉外形；(c)人臉正規化。

$YC_B C_R$  是一個正交的色彩空間，分別表示流明、藍色色度和紅色色度，因為亮度被獨立出來，也很適合膚色偵測[6]，以下是  $RGB$  轉換為  $YC_B C_R$  的式子，

$$\begin{bmatrix} Y \\ C_B \\ C_R \end{bmatrix} = \begin{bmatrix} 0.257 & 0.504 & 0.098 \\ -0.148 & -0.291 & 0.439 \\ 0.439 & -0.368 & -0.071 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix}, \quad (2.1.1)$$

式子(2.1)的  $R$ 、 $G$  和  $B$  三個數值都是從 0 到 255 的正整數。

$HSV$  偏向於人眼直觀的色彩空間，分別表示色相、飽和度和明度，色相可取 0 到 1 代表不同顏色的數值；飽和度數值越高表示顏色越純，數值越低表示逐漸變灰；明度數值越高表示亮度越高。因為亮度被抽離出來，也適合膚色偵測[6]，以下是  $RGB$  轉換為

HSV 的公式，

$$H_1 = \arccos \left( \frac{0.5[(R-G) + (R-B)]}{\sqrt{(R-G)^2 + (R-B)(G-B)}} \right), \quad (2.1.2)$$

$$H = \begin{cases} \frac{H_1}{360^\circ}, & \text{若 } B \leq G \\ 1 - \frac{H_1}{360^\circ}, & \text{若 } B > G \end{cases}, \quad (2.1.3)$$

$$S = \begin{cases} 0, & \text{若 } \max(R, G, B) = 0 \\ 1 - \frac{\min(R, G, B)}{\max(R, G, B)}, & \text{若 } \max(R, G, B) \neq 0 \end{cases}, \quad (2.1.4)$$

$$V = \frac{\max(R, G, B)}{255}, \quad (2.1.5)$$

其中  $\arccos$  表示反餘弦運算， $\max$  表示取最大值的運算， $\min$  表示取最小值的運算。

膚色跟非膚色的分類有很多方法[6]，本論文使用邊界範圍來構成人臉膚色範圍，分別是， $105 \leq C_R \leq 200$ ， $105 \leq C_B \leq 210$ ， $0.01 \leq H \leq 0.1$ 。膚色範圍比較大的缺點是輸入影像的膚色面積隨之加大，造成搜尋人臉的時間也增加了，但優點是比較不容易遺漏人臉偵測的位置，對於後段的人臉辨識是有幫助的。

## 2.1.2 人臉特徵抽取

二維的賈柏小波在訊號處理這塊領域受到廣泛的應用，兼具時域跟頻域的良好特徵，也經常應用於人臉特徵抽取[53]，式子表示如下，

$$\Psi_{s,d}(\vec{z}) = \frac{\|\vec{k}_{s,d}\|^2}{\sigma^2} [\exp(j\vec{k}_{s,d} \cdot \vec{z}) - \exp(-\frac{\sigma^2}{2})] \exp(-\frac{\|\vec{k}_{s,d}\|^2 \cdot \|\vec{z}\|^2}{2\sigma^2}) \quad (2.1.6)$$

其中  $\vec{z} = (x, y)$ ， $\vec{k}_{s,d} = k_s \exp(j\theta_d)$ ， $k_s = \frac{k_{\max}}{f^s}$ ， $\theta_d = \frac{d}{8}\pi$ ， $\theta_d \in [0, \pi)$  是屬於二維空間

的八個方向， $k_{\max}$  是最大頻率值， $f$  是賈伯濾波器在頻域的間隔係數， $s$  是控制賈伯濾波器的尺寸， $d$  是控制賈伯濾波器的方向， $k_s$  是為了讓不同空間頻率的濾波器有大致相等的能量， $\sigma$  是決定高斯窗(Gaussian window)的寬度和波向量的比例。

本論文採用五種尺寸( $k_{\max} = \pi$  ,  $f = \sqrt{2}$  ,  $s \in 0, \dots, 4$ )、八個方向( $d \in 0, \dots, 7$ )，以及  $\sigma = \pi$  的賈柏濾波器。一張二維的影像使用賈柏小波來抽取特徵的方法，就是執行旋積(convolution)運算，

$$G_{u,v}(\vec{z}) = I * \Psi_{s,d}(\vec{z}), \quad (2.1.7)$$

其中  $I$  表示輸入的二維影像， $\Psi_{s,d}(\vec{z})$  可以從式子(2.1.6)得到的賈伯濾波器。

### 2.1.3 類神經網路

類神經網路有許多相互連結的處理單元，如同人類大腦的神經元，透過樣本的訓練來展現學習、回想和歸類推演的能力，在處理圖形識別、資料分類和最佳化都有不錯的效果。

本論文採用倒傳遞類神經網路，是屬於監督式的前饋(feed forward)類神經網路，網路的所有神經元只有單一方向的向前傳遞，無向後傳遞或側向傳遞。倒傳遞類神經網路的架構屬於多層感知器(multi layer perceptron)，搭配誤差倒傳遞演算法(error back propagation)來訓練人臉跟非人臉樣本，以處理非線性的分類問題。

一般類神經網路可分為三層，如圖 2.2，輸入層表示輸入項，隱藏層的層數可以從問題的複雜程度由一層增加到數層，輸出層表示輸出項，其中隱藏層函數為雙彎曲正切函數  $\frac{1 - e^{-2h}}{1 + e^{-2h}}$ ，輸出層函數為雙彎曲函數  $\frac{1}{1 + e^{-h}}$ ， $h = 1, \dots, N$ 。

倒傳遞演算法的學習法則可分為正向傳遞跟反向傳遞兩種，在正向傳遞過程，輸入訊號從輸入層到隱藏層經過雙彎曲正切函數運算，再傳向輸出層的雙彎曲函數運算，才有輸出值，每一層神經元只影響下一層神經元的結果，若輸出層的輸出向量與目標向量(標準答案)兩者誤差值過大的時候，則兩者的差異轉為反向傳遞，將誤差訊號沿著原來的網路通道回傳，藉此修改每一層神經元的權重值，減小誤差值，再重新輸入訊號，如此不斷地運算直到誤差值達到容忍的誤差範圍而停止運算。倒傳遞類神經網路雖然可以處理非線性分類問題，但是學習結果還是會受到許多參數的影響，譬如隱藏層神經元的數目和神經元非線性函數的選擇等等，通常需要透過幾次的試誤法來決定適當的參數。

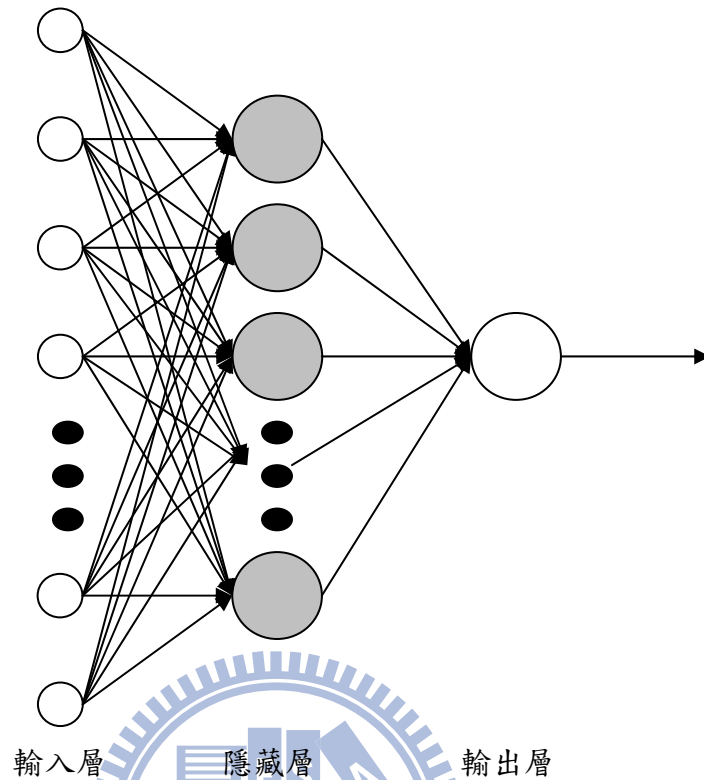


圖 2.2 人臉與非人臉分類之類神經網路架構圖。

## 2.2 人臉正規化

人臉正規化主要包含三個處理程序：定位、尺寸縮放和明亮度一致化，以下兩個小章節將分別說明如何利用主動式外觀模型和濾波器來達到正規化目的。

### 2.2.1 主動式外觀模型演算法

先知道人臉位置，接下來就是把人臉外形擷取下來，使用主動式外觀模型(active appearance model, AAM)[10][47]可達到這個目的。主動式外觀模型的想法一開始來自於 1987 年 Kass 等人提出的 snake 方法，主要應用於邊界檢定與圖像切割，該方法使用一條由  $n$  個控制點組成得連續封閉曲線做為 snake 模型，然後用一個能量函數做為匹配程



度的評量函數，首先將模型設定在待測物預估位置的周圍，透過不斷疊代運算使能量函數最小化，當內外能量達到平衡時即得到待測物的邊界與特徵。1989 年 Yuille 等人提出使用參數化的可變形模板來代替 snake 模型，這也是主動型外觀模型的理論基礎。1995 年 Cootes 等人提出主動式形狀模型(active shape model, ASM)演算法[9]，這是一個可以定位和調整人臉大小的方法，首先準備訓練樣本，以手動方式描繪出人臉的形狀，也就是  $n$  個控制點，經過主成分分析法(principal component analysis, PCA)訓練後，自動調整控制點，以找尋待測物可能的或符合的形狀，除此之外，主動式形狀模型也會考慮到人臉灰階值而調整形狀的控制點，以便找出人臉正確的位置，尤其是眼睛部位。主動式形狀模型演算法[9]的訓練樣本裡總共有  $N$  個形狀，訓練步驟描述如下：

### 主動式形狀模型演算法

**步驟一** 旋轉角度、尺寸縮放和變換每一個形狀，來對齊第一個形狀。

**步驟二** 每一個形狀對齊後，計算全部形狀的平均形狀(mean shape)。

**步驟三** 正規化平均形狀的三個參數：位置、尺寸縮放和旋轉角度。

**步驟四** 所有形狀重新對齊平均形狀，再調整這三個參數：位置、尺寸縮放和旋轉角度。

**步驟五** 回到步驟二，直到收斂條件成立。

**步驟六** 不斷計算合成模型，計算出合成模型跟待測物的最短距離，最後這個合成模型就是待測物最佳的人臉形狀。

1998 年 Cootes 等人在主動式形狀模型演算法基礎上提出主動式外觀模型[10][47]，除了利用待測物形狀(shape)資訊，也包含紋理(texture)資訊。主動式外觀模型利用主成分分析法訓練樣本與平均形狀的形狀模型，訓練所有樣本的形狀因變形而更接近主要形狀(main shape)，將每個灰階點轉成向量，並計算出訓練樣本灰階向量跟平均外觀(appearance)的變異量，主要有這三個參數：位置、尺寸縮放和旋轉角度，在搜尋階段調整此三個參數值，以便找出最佳的人臉形狀。主動式外觀模型演算法的步驟描述如下：

### 主動式外觀模型演算法

**步驟一** 計算灰階誤差向量  $\Delta I$ ，目標就是讓  $|\Delta I|^2$  盡量地小，使用待測物灰階向量  $I_i$  減掉目前模型參數  $c$  的灰階向量  $I_m$  而得  $\Delta I$ ，

$$\Delta I = I_i - I_m. \quad (2.2.1)$$

**步驟二** 計算模型參數變異量  $\Delta c$ ，因模型參數有很多個參數，這是一個高維度最佳化問題，為了節省運算時間，實際上使用逼近法來求得最佳解，透過灰階誤差向量  $\Delta I$  以及矩陣  $A$  即可，而矩陣  $A$  是使用多重多變數線性回歸法計算已知樣本模型得來的，

$$\Delta c = A \Delta I. \quad (2.2.2)$$

**步驟三** 計算新的模型參數  $c$ ，先給定一個外觀模型參數初始值  $c_0$ ，包含形狀跟灰階資訊，由參數變異量  $\Delta c$  替換後得到新的模型參數  $c$ ，

$$c = \Delta c + c_0. \quad (2.2.3)$$

**步驟四** 估算灰階誤差向量  $\Delta g_0$ ，透過新的模型參數  $c$  先計算新的形狀控制  $x$  和灰階控制  $g$ ，才得到正規灰階向量  $g_m$ ；而  $g_s$  則是利用形狀控制  $x$  擷取待測樣本的灰階向量，兩者相減即得灰階誤差向量  $\Delta g_0$ ，

$$\Delta g_0 = g_s - g_m. \quad (2.2.4)$$

**步驟五** 計算目前的灰階誤差純量值，目的就是得到最小的能量函數  $E_0$ ，

$$E_0 = |\Delta g_0|^2. \quad (2.2.5)$$

**步驟六** 估算模型置換函數  $\Delta c$ ，矩陣  $A$  是透過多重多變數線性回歸法計算已知樣本模型得來的，

$$\Delta c = A \Delta g_0. \quad (2.2.6)$$

令  $k = 1$  執行下一步驟，

**步驟七** 計算新的模型參數  $c_1$ ，先給定一個外觀模型參數初始值  $c_0$ ，包含形狀跟灰階資訊，由參數變異量  $\Delta c$  替換後得到新的模型參數，

$$c_1 = c_0 - k \Delta c. \quad (2.2.7)$$

**步驟八** 估算新的灰階誤差向量  $\Delta g_1$ ，透過新的模型參數  $c_1$  先計算出形狀控制  $x_1$  和灰階

控制  $g_1$ ，才得到正規灰階向量  $g_m$ ；而  $g_{s1}$  則是利用形狀控制  $x$  擷取待測樣本的灰階向量，兩者相減即得灰階誤差向量  $\Delta g_1$ ，

$$\Delta g_1 = g_{s1} - g_m. \quad (2.2.8)$$

**步驟九** 若符合收斂條件

$$|\Delta g_1|^2 < E_0.$$

就接受新的模型參數  $c_1$ ；若收斂條件不成立，則更換參數  $k = 1.5$ ， $k = 0.5$ ， $k = 0.25 \dots$  回到步驟七分別重新計算，直到尋找適當的模型參數  $c_1$  為止。

主動式外觀模型演算法在應用上有兩個值得留意的問題是，如何決定初始模型參數  $c_0$ ？新的模型參數  $c_1$  如何能夠收斂得更好？關於這兩個問題，也是一個不錯的研究工作 [55][56]；然而，本系統採取的是適應性人臉偵測位置為初始模型參數  $c_0$  的位置參數。

## 2.2.2 光線明亮度正規化

生活照片受到不同光線角度的影響，人臉光線明暗度通常是不一致的，本系統採用八種濾波器正規化演算法[48][49]，分別是單尺寸視網膜(SSR)正規化、多尺寸視網膜(MSR)正規化、適應性單尺寸視網膜(ASSR)正規化、同形濾波(HOMOF)正規化、單尺寸自商影像(SSSQI)正規化、多尺寸自商影像(MSSQI)正規化、離散餘弦轉換(DCT)正規化和可導引濾波器(SF)正規化等八種演算法，分別描述如下：

### 一．單尺寸視網膜正規化演算法

單尺寸視網膜正規化演算法(single scale retinex normalization algorithm)，是參考 Jobson[51]演算法，輸入影像  $f(x, y)$  正規化分佈於 0 到 255 正整數，其中  $x = 1, \dots, M$ ， $y = 1, \dots, N$ ， $M = N$  為偶數，輸出影像  $R_{SSR}(x, y, \sigma)$  計算如下，

$$R_{SSR}(x, y, \sigma) = \log(f(x, y) + 0.01) - \log(Z(x, y, \sigma)), \quad (2.2.9)$$

其中， $\log$  表示對數函數運算， $Z(x, y, \sigma)$  求解過程如下，先建立濾波器  $G(x, y, \sigma)$ ，

$$G(x, y, \sigma) = \exp(-((M/2 - x)^2 + (M/2 - y)^2) / \sigma^2) / \text{sum}G, \quad (2.2.10)$$

計算正規化參數  $\text{sum}G$ ，



$$sumG = \left( \sum_{x=1}^M \sum_{y=1}^N \exp(-((M/2 - x)^2 + (M/2 - y)^2) / \sigma^2) \right), \quad (2.2.11)$$

本系統取  $\sigma = 15$ ，再求解  $Z(x, y, \sigma) = f(x, y) * G(x, y, \sigma)$ ，符號\*表示旋積(convolution)運算，若  $Z(x, y, \sigma)$  某一個像素等於零，則令為 0.01，以免取對數函數運算得到負無窮大值。最後輸出影像  $R_{SSR}(x, y, \sigma)$  由小到大排序讓每一個像素值都不一樣，因為像素值超過 255，所以再做一次正規化，讓所有像素分佈在 0 到 255 正整數。

## 二．多尺寸視網膜之正規化演算法

多尺寸視網膜之正規化演算法(multi scale retinex normalization algorithm)，也是參考 Jobson[51]演算法，主要由上面單尺寸視網膜正規化演算法延伸而來，本系統多尺寸視網膜的  $\sigma$  取三個值[7 15 21]，單尺寸視網膜的  $\sigma$  只有一個值 15，所以稍微改寫輸出影像，分別計算三次結果加總起來即可，輸出影像  $R_{MSR}(x, y)$  計算如下，

$$R_{MSR}(x, y) = \sum_{i=1}^3 R_{SSR}(x, y, \sigma_i), \quad (2.2.12)$$

其中  $R_{SSR}$  由式子(2.2.9)直接帶入即可，最後輸出影像  $R_{MSR}(x, y)$  由小到大排序讓每一個像素值都不一樣，因為像素值超過 255，所以再做一次正規化，讓所有像素分佈在 0 到 255 正整數。

## 三．適應性單尺寸視網膜之正規化演算法

適應性單尺寸視網膜之正規化演算法(adaptive single scale retinex normalization algorithm)[57]，輸入影像  $f(x, y)$  正規化分佈於 0 到 255 正整數，其中  $x = 1, \dots, M$ ， $y = 1, \dots, N$ ，輸出影像  $R(x, y)$  計算如下，

$$R(x, y) = \log(f(x, y) + 1) - \log(L_n(x, y) + 1), \quad (2.2.13)$$

其中， $\log$  表示自然對數運算。 $L_n(x, y)$  計算說明如下，令初始值  $L_o(x, y) = f(x, y)$ ，疊代十次計算  $L_n(x, y)$ ，

$$L_n(x, y) = L_o(x, y) * W, \quad (2.2.14)$$

符號\*表示旋積(convolution)運算，並取最大值  $L_n(x, y)$ ， $W$  計算方式敘述如後，

$$L_n(x, y) = \max(L_n(x, y), L_o(x, y)), \quad (2.2.15)$$

更新  $L_o(x, y)$  值，

$$L_o(x, y) = L_n(x, y), \quad (2.2.16)$$

疊代十次計算式子(2.2.14)~式子(2.2.16)，

其中  $W$  計算如下，符號  $\times$  表示乘積運算，

$$W = \alpha \times \beta, \quad (2.2.17)$$

$$\alpha = 1 / (1 + \tau^{0.5} / (0.1 \times \exp(-\text{mean}(\tau) / 0.1))), \quad (2.2.18)$$

$$\beta = 1 / (1 + (I / (10 \times \exp(-\text{mean}(I) / 10)))^{0.5}), \quad (2.2.19)$$

$$I = (G_x^2 + G_y^2)^{0.5}, \quad (2.2.20)$$

而上面式子(2.2.18)裡面的  $\exp$  表示指數運算， $\text{mean}$  表示平均值運算，係數  $\tau = \sin(\tau_{CLI} \pi / 2)$ ，先計算  $\tau_{CLI}$ ，由輸入影像  $f(x, y)$  裡面取一個區塊三乘三像素點做運算，對中心點計算出八個差異絕對值的總和，所有區塊計算完畢再除以九做平均。我們舉一個區塊三乘三範例，如圖 2.3 所示，中心點像素值 15，周圍八個像素值減掉中心值，取絕對值再加總起來為 379，除以九得到平均值  $\tau_{CLI} = 42.11$ ，而  $\tau = \sin(\tau_{CLI} \pi / 2) = -0.1719$ 。

100	212	61
3	15	8
41	19	13

圖 2.3 區塊三乘三像素值計算範例。

上面式子(2.2.20)  $G_x$  和  $G_y$  分別是計算  $x$  方向和  $y$  方向的梯度(gradient)，我們舉出一個梯度  $G_x$  範例，如圖 2.4(a)是一個  $f(x, y)$  為三乘三像素點範例，圖 2.4(b)是  $f(x, y)$  往  $x$  方向位移兩個像素，圖 2.4(c)是  $f(x, y)$  在  $x$  方向最右邊兩列像素為零，圖 2.4(d)得出梯度  $G_x$  為圖 2.4 (b)減掉圖 2.4 (c)，但只取中間列，不取第一列跟最後一列。

另外，我們再舉出一個梯度  $G_y$  範例，如圖 2.5(a)是一個  $f(x, y)$  為三乘三像素點範例，圖 2.5(b)是  $f(x, y)$  往  $y$  方向位移兩個像素，圖 2.5(c)是  $f(x, y)$  在  $y$  方向最右邊兩列像素為零，圖 2.5(d)得出梯度為圖 2.5(b)減掉圖 2.5(c)，但只取中間列，不取第一列跟最後一列。

最後輸出影像  $R(x, y)$  由小到大排序讓每一個像素值都不一樣，因為像素值超過 255，所以再做一次正規化，讓所有像素分佈在 0 到 255 正整數。

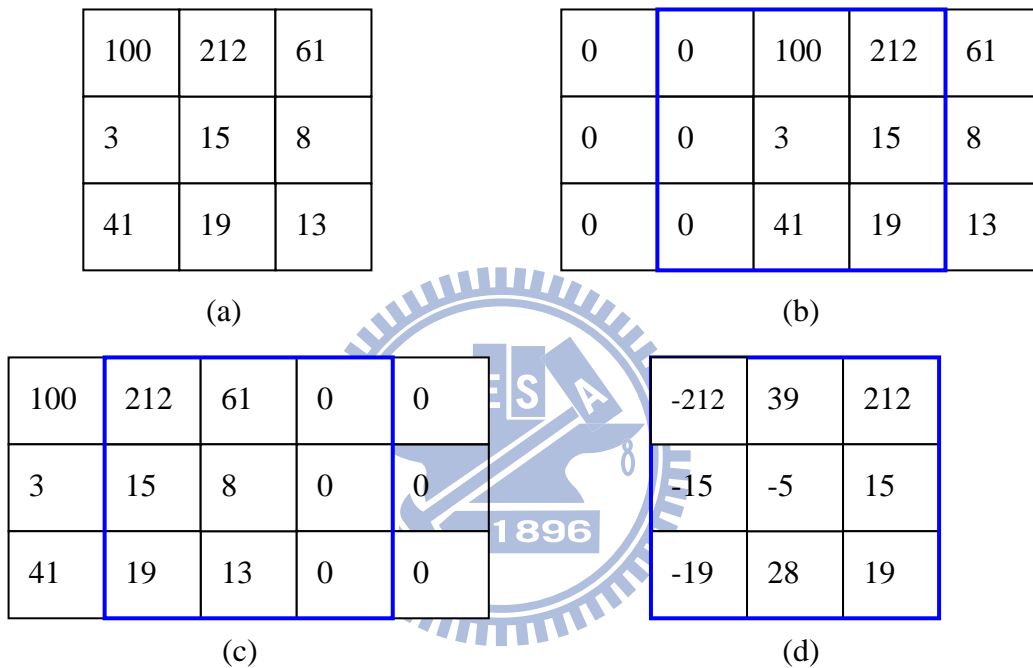


圖 2.4 計算  $x$  方向梯度  $G_x$  範例，(a)  $f(x, y)$  像素點；(b)  $f(x, y)$  往  $x$  方向位移兩個像素；  
(c)  $f(x, y)$  在  $x$  方向最右邊兩列像素為零；(d) 梯度  $G_x$  為(b)減掉(c)。

#### 四．同形濾波之正規化演算法

同形濾波之正規化演算法(homomorphic filtering normalization algorithm)[58]，輸入影像  $f(x, y)$  正規化分佈於  $\{0, 1\}$ ，其中  $x = 1, \dots, M$ ， $y = 1, \dots, N$ ，輸出影像  $R(x, y)$  計算如下，

$$R(x, y) = \exp(\text{real}(\text{IFFT}(F \times H))), \quad (2.2.21)$$

0	0	0
0	0	0
100	212	61
3	15	8
41	19	13

(b)

100	212	61
3	15	8
41	19	13
0	0	0
0	0	0

(c)

100	212	61
3	15	8
41	19	13

(a)

-3	-15	-8
59	193	48
3	15	8

(d)

圖 2.5 計算  $y$  方向梯度  $G_y$  範例，(a)  $f(x, y)$  像素點；(b)  $f(x, y)$  往  $y$  方向位移兩個像素；  
(c)  $f(x, y)$  在  $y$  方向最下面兩行像素為零；(d) 梯度  $G_y$  為 (b) 減掉 (c)。

其中， $\exp$  表示指數運算， $real$  表示只取複數裡面的實部，不取虛部，符號 $\times$ 表示乘積運算， $IFFT$  表示二維反向傅立葉轉換(inverse FFT)， $F$  跟  $H$  的計算說明如下，首先對輸入影像做二維傅立葉轉換(FFT)，

$$F = FFT(\log(f(x, y) + 0.01)), \quad (2.2.22)$$

其中， $\log$  表示自然對數運算， $FFT$  表示二維傅立葉轉換，

$$H = (1 - 1/\alpha) \times HPF + 1/\alpha, \quad (2.2.23)$$

本系統取  $\alpha = 8$ ，高通濾波器  $HPF$  由低通濾波器計算而來，

$$HPF = 1 - LPF, \quad (2.2.24)$$

低通濾波器  $LPF$  計算如下，

$$LPF = IFFTSHIFT(1/(1 + ((X^2 + Y^2)^{0.5} / CutOff)^{2n})), \quad (2.2.25)$$

其中  $IFFTSHIFT$  是二維反向傅立葉轉換位移運算，本系統取  $CutOff = 0.12$ ， $n = 2$ ， $X$

是二維矩陣拷貝  $N$  個一維向量  $X1$  組合起來的，其中  $X1$  是一維向量具有  $1 \sim M$  個值平均分佈在  $\{-0.5, 0.5\}$ ； $Y$  是二維矩陣拷貝  $M$  個一維向量  $Y1$  組合起來的，其中  $Y1$  是一維向量具有  $1 \sim N$  個值平均分佈在  $\{-0.5, 0.5\}$ 。最後輸出影像  $R(x, y)$  由小到大排序讓每一個像素值都不一樣，因為像素值超過 255，所以再做一次正規化，讓所有像素分佈在 0 到 255 正整數。

## 五・單尺寸自商影像之正規化演算法

單尺寸自商影像之正規化演算法(single scale self quotient image normalization algorithm)[59]，輸入影像  $f(x, y)$  正規化分佈於 0 到 255 正整數， $x = 1, \dots, M$ ， $y = 1, \dots, N$ ，輸出影像  $R_{SSQI}(x, y, \sigma, \beta)$  計算如下，本系統取高斯參數  $\sigma = 1$ ，區塊  $\beta = 5$ ，

$$R_{SSQI}(x, y, \sigma, \beta) = f(x, y) / (Z(x, y) + 0.01), \quad (2.2.26)$$

若  $R_{SSQI}(x, y, \sigma, \beta)$  不是有限數值，則使用左邊像素值取代，因為計算過程是由左至右，所以可以保證全部都是有限數值；其中  $Z(x, y)$  計算如下

$$Z(x, y) = \frac{1}{\beta^2} \sum_{k=1}^{\beta} \sum_{h=1}^{\beta} G1(k, h) \times Region(x, y, k, h), \quad (2.2.27)$$

其中區域矩陣  $Region(x, y, k, h)$  表示在複製矩陣  $F(j, h)$  裡面取一個區塊五乘五像素(因為本系統取區塊  $\beta = 5$ )， $k = 1, \dots, \beta$ ， $h = 1, \dots, \beta$ ， $j = 1, \dots, M \times N$ ，而複製矩陣  $F(j, h)$  就是先把二維影像  $f(x, y)$  轉換成一維向量，再拷貝五次(區塊  $\beta = 5$ )形成二維複製矩陣  $F(j, h)$ ；另外  $G1(k, h)$  計算方式如下，

$$G1(k, h) = G2(k, h) / SumG2, \quad (2.2.28)$$

$G2(k, h)$  如下式子取正規化，

$$G2(k, h) = G(k, h) \times Mask(k, h), \quad (2.2.29)$$

$$SumG2 = \sum_{k=1}^{\beta} \sum_{h=1}^{\beta} G2(k, h), \quad (2.2.30)$$

而  $G(k, h)$  為高斯濾波器，本系統取  $\sigma = 1$  的五乘五( $\beta = 5$ )矩陣

$$G = \begin{bmatrix} 0.0030 & 0.0133 & 0.0219 & 0.0133 & 0.0030 \\ 0.0133 & 0.0596 & 0.0983 & 0.0596 & 0.0133 \\ 0.0219 & 0.0983 & 0.1621 & 0.0983 & 0.0219 \\ 0.0133 & 0.0596 & 0.0983 & 0.0596 & 0.0133 \\ 0.0030 & 0.0133 & 0.0219 & 0.0133 & 0.0030 \end{bmatrix}, \quad (2.2.31)$$

遮罩矩陣  $Mask(k, h)$  計算方式如下，先計算區域矩陣  $Region(x, y, k, h)$  裡面 25 個像素點的平均值，若每個像素點大於等於這個平均值，則  $Mask(k, h)$  標記為 1，否則為 0，遮罩矩陣  $Mask(k, h)$  計算範例可參考圖 2.6，圖 2.6(a) 是在複製矩陣  $F(j, h)$  裡面取一個區域矩陣  $Region(x, y, k, h)$ ，計算所有像素的平均值為 27.8，比平均值大的地方標記為 1 而得到圖 2.6(b) 遮罩矩陣  $Mask(k, h)$ ；把複製矩陣  $F(j, h)$  裡面所有區域矩陣  $Region(x, y, k, h)$  計算完畢，由式子(2.2.27)得出  $Z(x, y)$ ，再由式子(2.2.26)推算輸出影像  $R_{SSSQI}(x, y, \sigma, \beta)$ 。最後，輸出影像  $R_{SSSQI}(x, y, \sigma, \beta)$  由小到大排序讓每一個像素值都不一樣，因為像素值超過 255，所以再做一次正規化，讓所有像素分佈在 0 到 255 正整數。

100	212	61	7	15	1	1	1	0	0
3	15	8	14	17	0	0	0	0	0
41	19	13	18	25	1	0	0	0	0
0	12	9	42	8	0	0	0	1	0
1	20	23	9	3	0	0	0	0	0

(a)

(b)

圖 2.6 五乘五遮罩矩陣計算範例，(a)先取一個五乘五區域矩陣；(b)計算後求得的遮罩矩陣。

## 六．多尺寸自商影像之正規化演算法

多尺寸自商影像之正規化演算法(multi scale self quotient image normalization algorithm)，主要由上面單尺寸自商影像(SSSQI)之正規化演算法延伸而來的，多尺寸自

商影像(MSSSQI)有四組高斯參數 $\sigma=[1 \ 1.2 \ 1.4 \ 1.6]$ ，分別對應四組區塊參數 $\beta=[3 \ 5 \ 11 \ 15]$ ，單尺寸自商影像只有一組高斯參數 $\sigma=1$ ，對應一組區塊參數 $\beta=5$ ，所以稍微改寫輸出影像，分別計算四次結果加總起來即可得到輸出影像 $R_{MSSQI}(x,y)$ ，

$$R_{MSSQI}(x,y) = \sum_{i=1}^4 R_{SSQI}(x,y,\sigma_i,\beta_i), \quad (2.2.32)$$

最後，輸出影像 $R_{MSSQI}(x,y)$ 由小到大排序讓每一個像素值都不一樣，因為像素值超過255，所以再做一次正規化，讓所有像素分佈在0到255正整數。

## 七．離散餘弦轉換之正規化演算法

離散餘弦轉換之正規化演算法(discrete cosine transform normalization algorithm)[60]，輸入影像 $f(x,y)$ 先進行自然對數運算 $f(x,y) = \log(f(x,y)+1)$ ，其中，每個像素加一的目的是避免自然對數計算後出現負無窮大值，然後 $f(x,y)$ 再正規化分佈於0到255正整數， $x=1, \dots, M$ ， $y=1, \dots, N$ ，輸出影像 $R(x,y)$ 計算如下，

$$R(x,y) = IDCT(Df), \quad (2.2.33)$$

其中 $IDCT$ 運算，表示二維反向離散餘弦轉換運算；先計算 $f(x,y)$ 的二維離散餘弦轉換運算( $DCT$ 運算)，

$$Df = DCT(f(x,y)), \quad (2.2.34)$$

經過二維離散餘弦轉換運算後得到 $Df$ ，為了讓影像低頻非零係數排列在高頻係數之前，於是使用ZigZag(蜿蜒)排列(範例如圖2.7)，使得低頻非零係數排列在二維影像的左上角，令 $Df$ 前20筆為零(不全然屬於低頻係數)，由式子(2.2.33)再推算輸出影像 $R(x,y)$ 。最後，輸出影像 $R(x,y)$ 由小到大排序讓每一個像素值都不一樣，因為像素值超過255，所以再做一次正規化，讓所有像素分佈在0到255正整數。

## 八．可導引濾波器之正規化演算法

可導引濾波器之正規化演算法(steerable filter normalization algorithm)[61]，輸入影像 $f(x,y)$ 先進行自然對數運算 $f(x,y) = \log(f(x,y)+1)$ ，其中，每個像素加一的目的是避免自然對數計算後出現負無窮大值，然後 $f(x,y)$ 再正規化分佈於0到255正整數， $x=$

1	2	6	7	15	16
3	5	8	14	17	26
4	9	13	18	25	27
10	12	19	24	28	33
11	20	23	29	32	34
21	22	30	31	35	36

圖 2.7 六乘六矩陣之 ZigZag(蜿蜒)排列範例。

$1, \dots, M$  ,  $y = 1, \dots, N$  , 輸出影像  $R(x, y)$  計算如下 ,

$$R(x, y) = \sum_{j=1}^8 f(x, y) * G_j , \quad (2.2.35)$$

其中 , 符號\*表示卷积(convolution)運算 , 濾波器  $G_j$  計算如下

$$G_j = \cos((j-1)\pi/8)G_x + \sin((j-1)\pi/8)G_y , \quad (2.2.36)$$

其中  $j = 1, \dots, 8$  , 在二維平面  $\{0, \pi\}$  取八等份 , 由原先高斯對稱方程式

$G(x, y) = \exp(-(x^2 + y^2))$  , 經過一階微分[61] , 推導出下面兩個方向的基本濾波器  $G_x$  和

$G_y$  ,

$$G_x = -2X \exp(-(X^2 + Y^2)/(2\sigma^2)) , \quad (2.2.37)$$

$$G_y = -2Y \exp(-(X^2 + Y^2)/(2\sigma^2)) , \quad (2.2.38)$$

$$X = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} , \quad (2.2.39)$$

$$Y = \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} , \quad (2.2.40)$$

其中  $X$  和  $Y$  分別是  $x$  軸方向跟  $y$  軸方向的邊緣偵測 , 本系統取  $\sigma = 0.5$  , 透過式子(2.2.36)



則可計算出濾波器  $G_j$ ，再由式子(2.2.35)推算出輸出影像  $R(x, y)$ 。最後，輸出影像  $R(x, y)$  由小到大排序讓每一個像素值都不一樣，因為像素值超過 255，所以再做一次正規化，讓所有像素分佈在 0 到 255 正整數。

## 2.3 人臉辨識

直覺上，稀疏編碼(sparse coding)[24][25][62]的參數求解就是以  $l_0$ -norm 來量測(計算一個向量裡非零的個數)，但這是一個非確定性多項式(NP-hard)問題，所以一般以  $l_1$ -norm 最小值逼近求解  $l_0$ -norm 最小值，成立條件是誤差值  $e$  足夠稀疏(sparse)[25]，也就是讓重建後影像的每個像素點跟原影像整體的誤差達到最小值。

圖 2.8 舉例二十張訓練樣本(共十類)，來說明稀疏編碼的作用，目的就是計算出訓練樣本(圖 2.8(b))二十個係數的每一個係數(圖 2.8(c))，利用每一個係數乘上每一個訓練樣本，線性組合加總起來，就得到重建人臉(圖 2.8(d))。稀疏編碼的係數向量計算就是使用  $l_1$ -norm 最小化逼近求解，透過一張待測人臉跟二十張訓練人臉，得到二十個係數。通常稀疏編碼問題可表示如下，

$$\text{找出 } \min_{\alpha} \|\alpha\|_1, \quad \text{使得 } \|Y - X\alpha\|_2^2 \leq e, \quad (2.3.1)$$

其中  $Y$  是一個已知訊號， $X$  表示資料庫編碼原子， $\alpha$  表示編碼向量， $e > 0$  是一個常數純量。從另外一個角度來看，方程式(2.3.1)裡的  $Y$  也可以表示標準答案， $X$  表示訓練樣本，我們就是要去找到一組盡可能小的編碼向量  $\alpha$ ，讓正數誤差值  $e$  也盡量地小。

$$\text{找出 } \min_{\alpha} \|Y - X\alpha\|_2^2, \quad \text{使得 } \|\alpha\|_1 \leq \sigma, \quad (2.3.2)$$

上面式子(2.3.2)成立的條件是，殘餘值  $\varepsilon = Y - X\alpha$  分佈型態就是高斯(Gaussian)分佈，而最小平方估測(LSE)解其實就是相當於最大粗略估測(MLE)解，其中

$Y = [y_1; y_2; \dots; y_n] \in \mathbb{R}^n$  是一個已被編碼的訊號， $X = [x_1; x_2; \dots; x_m] \in \mathbb{R}^{n \times m}$ ， $x_j$  表示資料庫行向量的第  $j$  個編碼原子， $\alpha$  表示編碼向量， $\sigma > 0$  是一個常數純量。

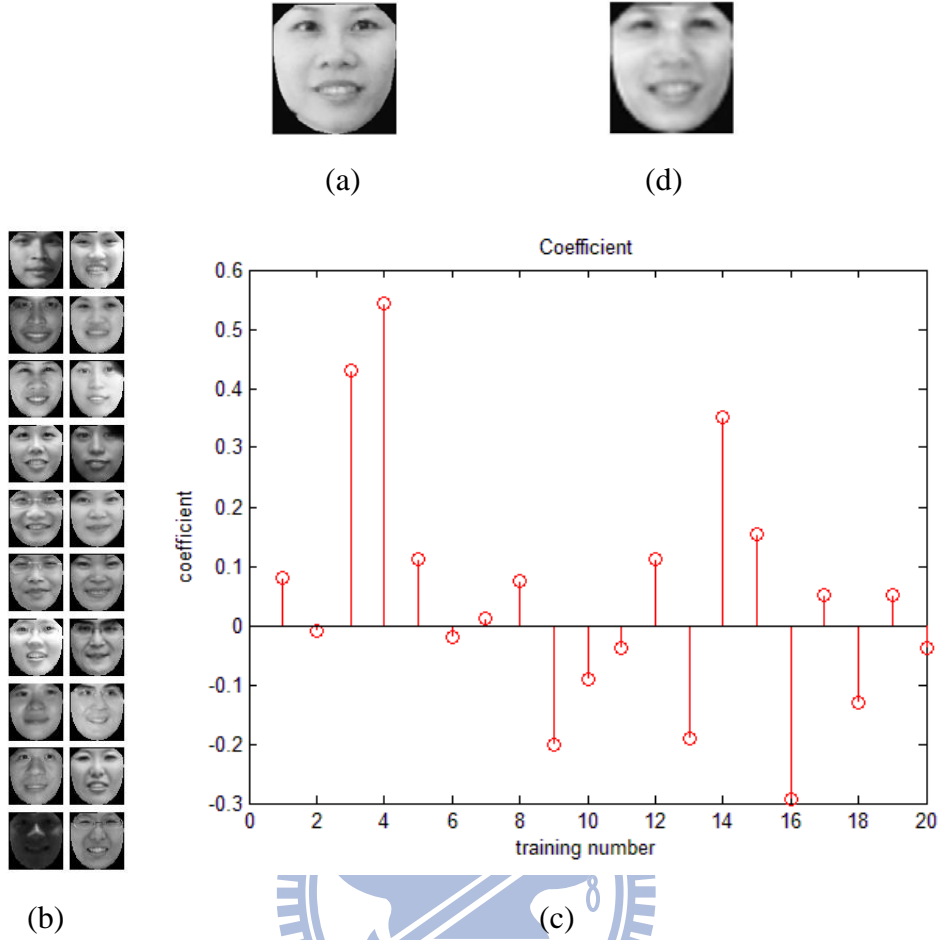


圖 2.8 (a)待測人臉；(b)訓練樣本；(c)稀疏編碼計算二十個係數；(d)重建後的人臉。

下面式子(2.3.3)成立的條件是，殘餘值  $\varepsilon = Y - X\alpha$  分佈型態就是拉氏(Laplacian)分佈，而最小平方估測(LSE)解其實就是最大粗略估測(MLE)解，

$$\text{找出 } \min_{\alpha} \|Y - X\alpha\|_1, \quad \text{使得 } \|\alpha\|_1 \leq \sigma, \quad (2.3.3)$$

實際上，殘餘值  $\varepsilon$  分佈型態既不是高斯分佈，也不是拉氏(Laplacian)分佈，特別是人臉受到遮掩物或光線等干擾的時候，所以光是使用這三個式子(2.3.1) ~ (2.3.3)，辨識率還是不夠高的。

接下來改寫  $X$ ，讓  $X = [d_1; d_2; \dots; d_n] \in \mathbb{R}^{n \times m}$ ，目的就是將式子(2.3.3)改寫為式子(2.3.4)， $d_i$  表示資料庫列向量的第  $i$  個編碼原子，而且  $\varepsilon = Y - X\alpha = [\varepsilon_1; \varepsilon_2; \dots; \varepsilon_n]$ ，所以每一個殘餘值  $\varepsilon$  可表示為  $\varepsilon_i = y_i - d_i\alpha$ ， $i = 1, 2, \dots, n$ ，

$$\text{找出 } \min_{\alpha} \sum_{i=1}^n \rho_{\theta}(y_i - d_i \alpha), \quad \text{使得 } \|\alpha\|_1 \leq \sigma, \quad (2.3.4)$$

上面式子(2.3.4)裡的  $\rho_{\theta}$  是屬於  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$  彼此獨立、分佈均勻的機率密度函數(PDF)，在此並不會直接求解，而是透過疊代運算找出越來越小的權重值  $\omega_{\theta}(\varepsilon_i)$ ，也就是將式子(2.3.4)近似於下面式子(2.3.5)，

$$\text{找出 } \min_{\alpha} \|W^{1/2}(Y - X\alpha)\|_2^2, \quad \text{使得 } \|\alpha\|_1 \leq \sigma, \quad (2.3.5)$$

其中對角複製矩陣  $W_{i,i} = \omega_{\theta}(\varepsilon_{0,i}) = \rho'_{\theta}(\varepsilon_{0,i}) / \varepsilon_{0,i}$ ，權重方程式選擇如下，

$$\omega_{\theta}(\varepsilon_i) = 1 / (1 + \exp(\mu \delta (\varepsilon_i^2 / \delta - 1))), \quad (2.3.6)$$

$$\delta = \frac{1}{n} \sum_{i=1}^n \varepsilon_i^2, \quad (2.3.7)$$

其中  $\mu$  和  $\delta$  都是正數純量，取  $\mu = c / \delta = 1 / \delta$ ，令  $c = 1$  常數純量；上面式子(2.3.7)  $\delta$  取所有殘餘值的平均值，當每個殘餘值都很接近  $\delta$ ， $\omega_{\theta}(\varepsilon_i)$  趨近於 0.5 表示稀疏編碼足夠稀疏；當每個殘餘值遠大於  $\delta$ ， $\omega_{\theta}(\varepsilon_i)$  趨近於 0 表示權重的代表性不足；當每個殘餘值遠小於  $\delta$ ， $\omega_{\theta}(\varepsilon_i)$  趨近於 0.7 也可以表示權重的代表性不足。

接下來說明式子(2.3.5)的由來，如下

令  $F_{\theta}(\varepsilon) = F_{\theta}(Y - X\alpha) = \sum_{i=1}^n \rho_{\theta}(\varepsilon_i)$ ，由第一階泰勒(Taylor)在  $\varepsilon_0$  展開式得到的近似值為

$\tilde{F}_{\theta}(\varepsilon) = F_{\theta}(\varepsilon_0) + (\varepsilon - \varepsilon_0)^T F'_{\theta}(\varepsilon_0) + \mathfrak{R}_1(\varepsilon)$ ，其中  $\mathfrak{R}_1(\varepsilon)$  是高階微分項， $F'_{\theta}(\varepsilon_0)$  是  $F_{\theta}(\varepsilon)$  對  $\varepsilon$  一次微分後在  $\varepsilon_0$  的值，而  $\rho'_{\theta}$  是  $\rho_{\theta}$  的一次微分，所以  $F'_{\theta}(\varepsilon_0) = [\rho'_{\theta}(\varepsilon_{0,1}); \rho'_{\theta}(\varepsilon_{0,2}); \dots; \rho'_{\theta}(\varepsilon_{0,n})]$ ，其中  $\varepsilon_{0,i}$  表示  $\varepsilon_0$  的第  $i$  個元素。

稀疏編碼通常希望是一個任意兩點連線的集合，皆包含在原集合內的問題(strictly convex)，所以高階微分項近似於  $\mathfrak{R}_1(\varepsilon) = 0.5(\varepsilon - \varepsilon_0)^T W(\varepsilon - \varepsilon_0)$ ，其中  $W$  為對角複製矩陣，因為  $\varepsilon_i$  跟  $\varepsilon_j$  彼此獨立，且無交叉變數，其中  $i \neq j$ 。 $F_{\theta}(\varepsilon)$  在  $\varepsilon = \mathbf{0}$  可以達到最小值  $\mathbf{0}$ ，所以  $\tilde{F}_{\theta}(\varepsilon)$  在  $\varepsilon = \mathbf{0}$  也可以達到最小值  $\mathbf{0}$ 。而  $W_{i,i} = \omega_{\theta}(\varepsilon_{0,i}) = \rho'_{\theta}(\varepsilon_{0,i}) / \varepsilon_{0,i}$  屬於正純量，因為  $\rho'_{\theta}(\varepsilon_i)$  跟  $\varepsilon_i$  不是同時為正數，就是同時為負數。 $\tilde{F}_{\theta}(\varepsilon) \approx \mathfrak{R}_1(\varepsilon) \approx 0.5 \|W^{1/2} \varepsilon\|_2^2 + b$ ，

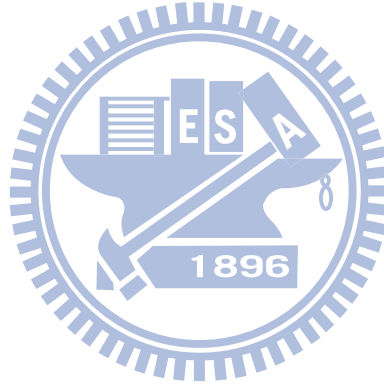
其中純量  $b$  是由  $\varepsilon_0$  決定的，且  $\varepsilon = Y - X\alpha$  代入後就得到式子(2.3.5)。

稀疏編碼在進行疊代演算法前，必須先決定第一個殘餘值  $\varepsilon^{(1)}$ ，以及收斂條件。因為一開始並不知道測試樣本是比較靠近哪一個類別，所以測試樣本值減掉平均訓練樣本值就是第一個殘餘值，表示如下，

$$\varepsilon^{(1)} = y_j - y_r^{(1)} = y_j - y^{(0)} = y_j - \frac{1}{n} \sum_{i=1}^n x_i, \quad (2.3.8)$$

收斂條件除了達到最大疊代次數以外，也可以訂定權重變化量小於門檻值  $\gamma$  就停止疊代運算，表示如下，

$$\|W^{(t)} - W^{(t-1)}\|_2 / \|W^{(t-1)}\|_2 < \gamma. \quad (2.3.9)$$



## 第三章 人臉辨識系統

本系統定位為小型的人臉辨識系統，大約是十個類別，可應用於家庭生活照之分類工作，或是搜尋特定成員的所有生活照。

本章節架構安排如下，3.1 節說明人臉位置偵測的方法。3.2 節提出本系統使用的正規化濾波器。最後 3.3 節闡述人臉辨識使用的稀疏編碼(sparse coding)演算法。

### 3.1 人臉位置偵測方法

以下是人臉偵測演算法的步驟，可參考圖 3.1 的流程圖，

**步驟一** 計算彩色生活照的膚色位置。

**步驟二** 決定多重尺寸視窗搜尋點，使用已知的平均人臉與偵測到的膚色區域做旋積運算，取區域內的極大值當做人臉搜尋的中心點。

**步驟三** 找出可能的人臉尺寸，由搜尋尺寸長寬  $20 \times 20$ ，每次增加 1.2 倍，直到搜尋視窗超過膚色區域為止。

**步驟四** 在判斷人臉與非人臉之前，必須先手動建立人臉與非人臉資料庫當做訓練樣本，使用賈柏小波抽取特徵，再透過倒傳遞類神經網路來訓練已分類好的人臉跟非人臉兩個類別，調整權重讓人臉的目標輸出值為 1，非人臉的目標輸出值為-1，當隱藏層的所有權重值能夠讓均方誤差值小於  $10^{-4}$ ，即表示訓練結果趨於收斂。

**步驟五** 當一張擷取後的影像經過特徵抽取後，輸入到已訓練好的倒傳遞類神經網路，若輸出值越接近目標值 1 表示人臉的可能性越高，反之，若輸出值越接近目標值-1 表示非人臉的可能性越高。

**步驟六** 最後人臉的位置可能有好幾個不同尺寸或好幾個不同位置交錯在一起，所以必須做合併的動作。當小尺寸的位置屬於大尺寸的一小部份，取輸出值比較大的當做人臉位置及對應的尺寸；若兩個相同座標有不同尺寸的人臉，代表人臉可能性越高，可以取尺寸的平均值當做新的人臉尺寸；若相似大小的人臉尺寸交疊在一起，可以使用區



域比較法，剔除輸出值比較小的人臉位置，留下輸出值最大的人臉位置為最後人臉偵測結果。



圖 3.1 人臉位置偵測流程圖。

### 3.1.1 多重尺寸視窗搜尋

利用上述的膚色偵測方法，在這個膚色區域內粗略地標記出可能的人臉位置，使用事先算好的平均人臉(圖 3.2)與膚色區域做旋積運算，取區域內的極大值當做人臉搜尋的中心點。由搜尋視窗長寬  $20 \times 20$ ，每次增加 1.2 倍，直到搜尋視窗超過膚色區域為止。我們的目標就是把所有人臉全部找出來，盡量降低錯誤偵測。



圖 3.2 平均人臉。

### 3.1.2 訓練類神經網路

訓練類神經網路的流程，首先輸入灰階的訓練影像樣本，經過直方圖等化來增加影像特徵的明顯程度；然後與賈伯濾波器做旋積運算，產生一組特徵輸入向量；經過降維後再送進類神經網路做訓練。

本論文使用的影像訓練樣本來自於卡內基美隆大學(Carnegie Mellon university)機器人協會所建立的 PIE 資料庫[63]，此資料庫共有 68 個人、41,368 張人臉照片，包括 13 種拍攝角度(人臉姿態)、43 種不同明亮度的拍攝條件，以及 4 種不同表情。本論文任意選取 PIE 資料庫中的 600 張人臉，包括 15 個人、5 種不同姿態和 8 種照明環境；另外任意選取 AR 資料庫[64]中的 200 張人臉，包括 20 個正面姿態以及 10 種照明環境；在網際網路任意選取 200 張亞洲東方人臉；所以共有 1,000 張人臉，再加上 2,000 張非人臉樣本，總共 3,000 張影像做訓練。非人臉樣本是在網際網路任意選取無人臉的彩色生活照，包括衣服、風景照、建築物等等樣本，經本系統擷取後做類神經網路訓練的，不選取人臉部份五官(耳、眉、眼、鼻、口)當做非人臉樣本。

每張影像都使用賈伯小波抽取特徵向量為  $2430 \times 1$ ，中間隱藏層為一層共 35 個維度，輸出值為一個純量，若輸出值越接近目標值 1 表示人臉的可能性越高，反之，若輸

出值越接近目標值-1 表示非人臉的可能性越高。當隱藏層的所有權重值能夠讓均方誤差值小於 $10^{-4}$ ，即表示訓練結果趨於收斂。讓類神經網路學習得更正確的做法，就是準備正確的訓練樣本、足夠多的樣本數量以及樣本間的差異性越大越好。

## 3.2 光線明亮度正規化濾波器

為了處理人臉受到光線不均勻照射而降低辨識率，本論文使用八種濾波器，分別測試哪一種濾波器可以得到最好的辨識率，分別是單尺寸視網膜(SSR)之正規化演算法[51]、多尺寸視網膜(MSR)之正規化演算法[51]、適應性單尺寸視網膜(ASSR)之正規化演算法[57]、同形濾波(HOMOF)之正規化演算法[58]、單尺寸自商影像(SSSQI)之正規化演算法[59]、多尺寸自商影像(MSSQI)之正規化演算法[59]、離散餘弦轉換(DCT)之正規化演算法[60]和可導引濾波器(SF)之正規化演算法[61]共八種。

圖 3.3 最左邊的原始人臉是從 AR 資料庫[64]取出的一個人臉範例，光線主要從左臉進來，不是從正面進來，尤其可以從圖 3.3 最左邊數過來第四個(ASSR 適應性單尺寸視網膜正規化)跟第五個(HOMOF 同形濾波正規化)看出來，光線的確從左臉進來。



圖 3.3 最左邊是原始人臉，其他經過明亮度濾波器處理後的結果，由左而右分別是 SSR，MSR，ASSR，HOMOF，SSSQI，MSSQI，DCT 和 SF。

## 3.3 稀疏編碼權重更新疊代演算法

人臉辨識使用的稀疏編碼(sparse coding)權重更新疊代演算法，收斂條件除了達到最大疊代次數以外，也可以訂定權重對角矩陣跟上一次的變化量相比較值，小於 5%就停止疊代運算，表示如下，

$$\|W^{(t)} - W^{(t-1)}\|_2 / \|W^{(t-1)}\|_2 < \gamma = 5\% . \quad (3.3.1)$$



為了降低  $l_1$ -ls[62] 運算量，在式子(2.3.6)  $\omega_\theta(\varepsilon_i) = 1/(1 + \exp(\mu\delta(\varepsilon_i^2/\delta - 1)))$ ，對  $\omega_\theta(\varepsilon_i)$  做十等份平均統計，去除最大極端分佈值，只取剩餘九等份的權重做運算，如圖 3.4 最左邊人臉圈圈部分表示選擇權重的位置點，標記為統計法 1；另外，去除頭尾兩個極端分佈值，只取中間八等份的權重做運算，如圖 3.4 中間人臉圈圈部分表示選擇權重的位置點，標記為統計法 2；而圖 3.4 最右邊人臉顯示，刪除權重位置點的條件是  $\omega_\theta(\varepsilon_i) < 0.1$ ，權重數量遠比左邊兩個方式多，標記為門檻法。

圖 3.5 是兩種權重選擇統計法在辨識率上面的比較，整體而言，統計法 2 的辨識率比統計法 1 來得好，尤其是訓練樣本數目在四個以下，而測試條件是使用稀疏編碼加上可導引濾波器正規化。統計法 2 可以減少稀疏編碼的權重數目為原來的 60%，除了降低系統運算量，同時特徵向量仍然具有代表性。

接下來的實驗將會採用圖 3.4 中間的統計法 2 運算來刪除極端值，當每個殘餘值遠大於  $\delta$ ， $\omega_\theta(\varepsilon_i)$  趨近於 0 表示權重的代表性不足，可以去除。當每個殘餘值遠小於  $\delta$ ， $\omega_\theta(\varepsilon_i)$  趨近於 0.7 也可以表示權重的代表性不足，可以去除。若  $\omega_\theta(\varepsilon_i)$  趨近於 0.5 表示殘餘值接近  $\delta$ ，稀疏編碼足夠稀疏，這些權重就相當具有代表性了。

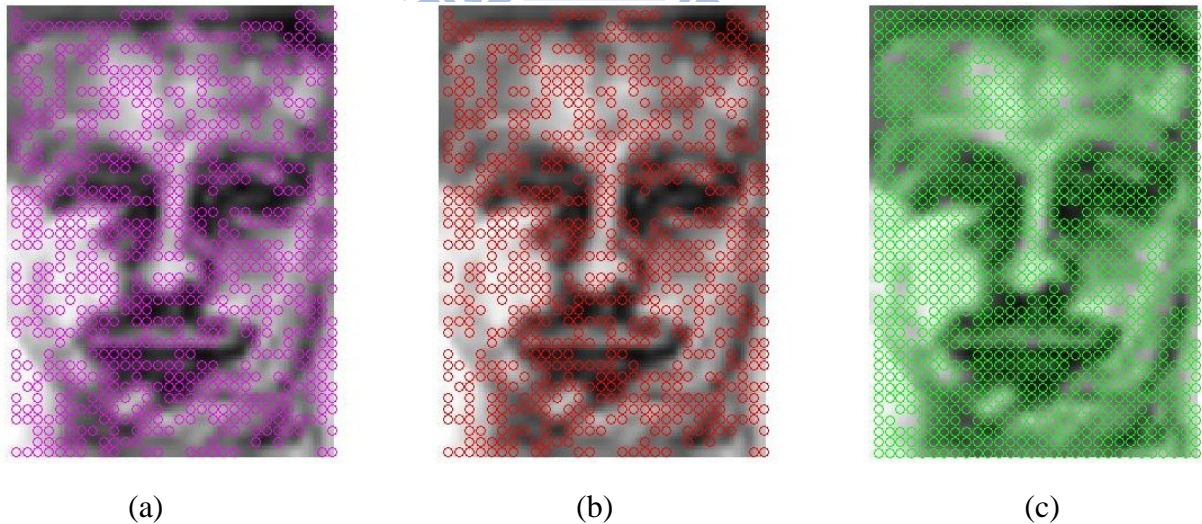


圖 3.4 選擇權重位置比較圖，(a)統計法 1；(b)統計法 2；(c)門檻法。

接下來，我們整理稀疏編碼權重更新疊代演算法，如下所示：

**稀疏編碼權重更新疊代演算法**

輸入：測試樣本  $y_j$  經過  $l_2$ -norm 正規化，訓練樣本  $X$  的每一行都經過  $l_2$ -norm 正規化，其中  $j = 1, \dots, m$ ，共有  $m$  個測試樣本。

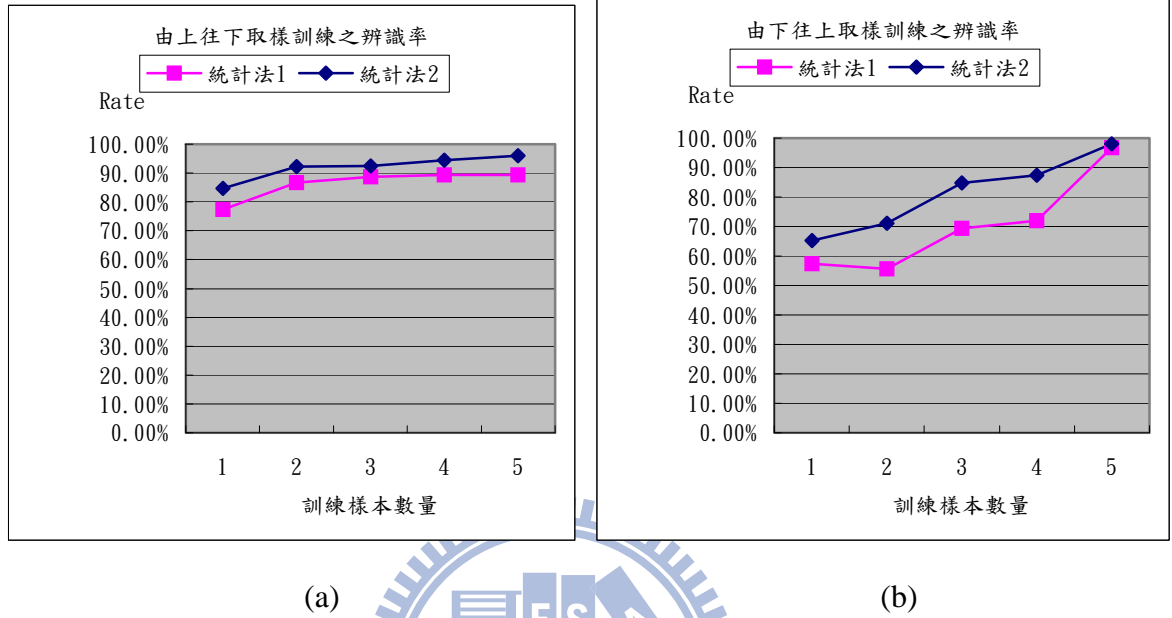


圖 3.5 兩種權重選擇統計法在辨識率上面的比較，(a) 由上往下取樣；(b) 由下往上取樣。

輸出：稀疏編碼向量  $\alpha$ 。

起始值  $t = 1$ ， $y_r^{(1)} = y^{(0)} = \frac{1}{n} \sum_{i=1}^n x_i$ ，演算法步驟描述如下，

步驟一 計算殘餘值

$$\varepsilon_i^{(t)} = y_j - y_r^{(t)}。$$

步驟二 估算權重值

$$\omega_{\theta}(\varepsilon_i^{(t)}) = 1 / (1 + \exp(\mu^{(t)} \delta^{(t)} ((\varepsilon_i^{(t)})^2 / \delta^{(t)} - 1)))， \quad (3.3.2)$$

其中  $\delta^{(t)}$  取所有殘餘值的平均值  $\delta^{(t)} = \frac{1}{n} \sum_{i=1}^n (\varepsilon_i^{(t)})^2$ ，當每個殘餘值都很接近

$\delta^{(t)}$ ， $\omega_{\theta}(\varepsilon_i^{(t)})$  趨近於 0.5 表示稀疏編碼足夠稀疏； $\mu^{(t)} = c / \delta^{(t)}$ ， $\mu^{(t)}$  和  $\delta^{(t)}$  都

是正數純量，取  $\mu^{(t)} = c / \delta^{(t)} = 1 / \delta^{(t)}$ ，令  $c = 1$  純量常數值。

步驟三 計算稀疏編碼向量

找出  $\alpha^{(new)} = \min_{\alpha} \| (W_{i,i}^{(t)})^{1/2} (y_j - X\alpha) \|_2^2$ ，使得  $\|\alpha\|_1 \leq \sigma$ ，

其中  $W_{i,i}^{(t)} = \omega_{\theta}(\varepsilon_i^{(t)})$  是估算權重後的對角複製矩陣；但是為了減少運算量，所以對  $\omega_{\theta}(\varepsilon_i^{(t)})$  做十等份平均統計，去除頭尾兩個極端分佈值，只取中間八等份的權重做運算。

#### 步驟四 更新稀疏編碼向量

當  $t = 1$ ， $\alpha^{(t)} = \alpha^{(new)}$ ；

當  $t > 1$ ， $\alpha^{(t)} = \eta^{(t)}(\alpha^{(new)} - \alpha^{(t-1)}) + \alpha^{(t-1)}$ ，

其中  $0 < \eta^{(t)} < 1$ ，但必須滿足條件

$$\sum_{i=1}^n \rho_{\theta}(\varepsilon_i^{(t)}) < \sum_{i=1}^n \rho_{\theta}(\varepsilon_i^{(t-1)})。$$

#### 步驟五 計算重建後的測試樣本

$y_r^{(t)} = X\alpha^{(t)}$ ，  
疊代次數增加一次  $t = t + 1$ 。

#### 步驟六 判斷收斂條件

若還沒有滿足收斂條件，回到步驟一繼續執行，收斂條件為

$$t < max\_iter \text{ 或者是 } \|W^{(t)} - W^{(t-1)}\|_2 / \|W^{(t-1)}\|_2 < \gamma。$$

若滿足收斂條件，令  $\alpha = \alpha^{(t)}$ ， $\omega_{\theta}(\varepsilon) = \omega_{\theta}(\varepsilon^{(t)})$ 。

#### 步驟七 人臉分類

計算測試樣本  $y_j$  所屬類別，也就是殘餘值乘上權重值的  $l_2$ -norm 最小值即為所屬的類別，

$$\min_k \| \omega_{\theta}(y_j - x_k \alpha_k) \|_2，$$

其中  $k = 1, \dots, K$ ，共有  $K$  個類別。先計算每一個類別的重建人臉  $x_k \alpha_k$ ，殘餘值為  $y_j - x_k \alpha_k$ ，殘餘值乘上權重值則為  $\omega_{\theta}(y_j - x_k \alpha_k)$ ，進行  $l_2$ -norm 運算後找出最小值即為此類別。

## 第四章 實驗結果與分析

本系統所使用的測試環境為，32 位元 Windows 7 作業系統，CPU Intel Core 2 Duo 2.1GHz，2GB RAM，測試生活照片的步驟與條件描述如下，

**步驟一** 先進行生活照的人臉位置偵測，儘可能把所有人脸的位置找出來，對於後續的人臉辨識率是有幫助的。

**步驟二** 經過主動外觀模型(AAM model)[10]將每一張人脸的輪廓擷取下來，讓每一張人脸都有一樣的輪廓跟相同的尺寸。

**步驟三** 每一張人脸都經過濾波器處理明亮度不均勻的問題，包含測試樣本跟訓練樣本。

**步驟四** 使用稀疏編碼(sparse coding)[25]演算法辨識人脸。

**步驟五** 比對輸入人脸在資料庫所屬類別，記錄此人臉出現在哪幾張生活照片，再進行生活照人物檢索或分類功能。

本章節架構安排如下，4.1 節是人脸偵測的實驗結果，4.2 節先提出特徵人脸(eigenfaces)[50]辨識率，做為 4.3 節稀疏編碼辨識率的比較基準，4.4 節呈現不同明亮度處理所得到的辨識率，4.5 節提出可導引濾波器處理後的人臉辨識率，4.6 節測試本系統人脸解析度最小可接受的範圍，4.7 節則分析生活照片的人臉辨識率，最後 4.8 節把實驗結果做個總結與討論。

### 4.1 人脸偵測的實驗結果

本實驗測試用的照片，是從自己拍攝以及同事提供的生活照片，任意挑選 60 張團體照，實際總共有 707 張人脸，根據 3.1 章節描述的人臉偵測方法，進行測試。

圖 4.1 是人脸偵測的部份實驗結果，圖 4.1(a)都是正面人脸，實際是 7 張人脸，系統可以正確偵測到 7 張人脸。圖 4.1(b)則是一個有遮掩物的正面人脸，實際是 33 張人脸，





(a)



(b)

圖 4.1 生活照片人臉偵測的部份實驗結果，(a)正面人臉；(b)有遮掩物的正面人臉；(c)有明亮度不均勻的正面人臉；(d)有大角度變化的人臉。





(c)



(d)

圖 4.1(續) 生活照片人臉偵測的部份實驗結果。



系統可以正確偵測到 32 張人臉，未被偵測到的人臉有 1 張，偵測到錯誤的人臉有 13 張，因人臉受到遮掩物而特徵不明顯時，系統的偵測效果將會受到影響。圖 4.1(c)則是一個明亮度不均勻的正面人臉，實際是 33 張人臉，系統可以正確偵測到 31 張人臉，未被偵測到的人臉有 2 張，偵測到錯誤的人臉有 14 張，因此人臉受到照明影響而干擾特徵時，系統的偵測效能將會下降。圖 4.1(d)則是一個有大角度變化的人臉，實際是 13 張人臉，系統可以正確偵測到 12 張人臉，未被偵測到的人臉有 1 張，偵測到錯誤的人臉有 1 張，由於人臉特徵並不完全，所以系統的偵測效果也會受到影響。

針對 60 張團體生活照，共有 707 張人臉，進行人臉偵測，實驗結果整理於表 4.1，根據影響人臉特徵的因素，可歸類為明亮度、角度和遮掩物三種狀況的組合。人臉偵測率計算方式如式子(4.1)，我們的目標就是要偵測到所有的人臉，以利於下一階段的人臉辨識。整體而言，表 4.1 的正確偵測率將近 82%，正確的個數有 581，遺漏的個數有 126，其中，遮掩物造成無法偵測人臉的影響程度是最大的。

$$\text{生活照片人臉偵測率} = \frac{\text{偵測正確的人臉個數}}{\text{偵測正確的人臉個數} + \text{遺漏的人臉個數}} \quad (4.1)$$

表 4.1 生活照片人臉偵測結果分析。

人臉特徵	正確的個數	遺漏的個數	正確偵測率
(a)正面人臉	201	0	100.0%
(b)明亮度不均勻	123	12	91.1%
(c)角度影響	112	18	86.1%
(d)遮掩物影響	32	20	61.5%
(e)明亮度不均勻及角度影響	72	23	75.7%
(f)角度及遮掩物影響	17	23	42.5%
(g)明亮度不均勻及遮掩物影響	23	21	52.2%
(h)明亮度不均勻及角度和遮掩物影響	1	9	10.0%
合計	581	126	82.1%

圖 4.2 分別展示因為明亮度不均勻、側臉角度變化太大以及遮掩物等三種因素的組合而無法偵測人臉的狀況。圖 4.2(a)中間人臉因為眼鏡反光大及仰角太大而無法偵測，圖 4.2(b)右邊人臉因為仰角太大而無法偵測，圖 4.2(c)中間後面人臉因為遮掩物而無法偵測，圖 4.2(d)下面人臉因為光線不均勻而無法偵測，圖 4.2(e)因為傾斜角度太大及照明問題而無法偵測。

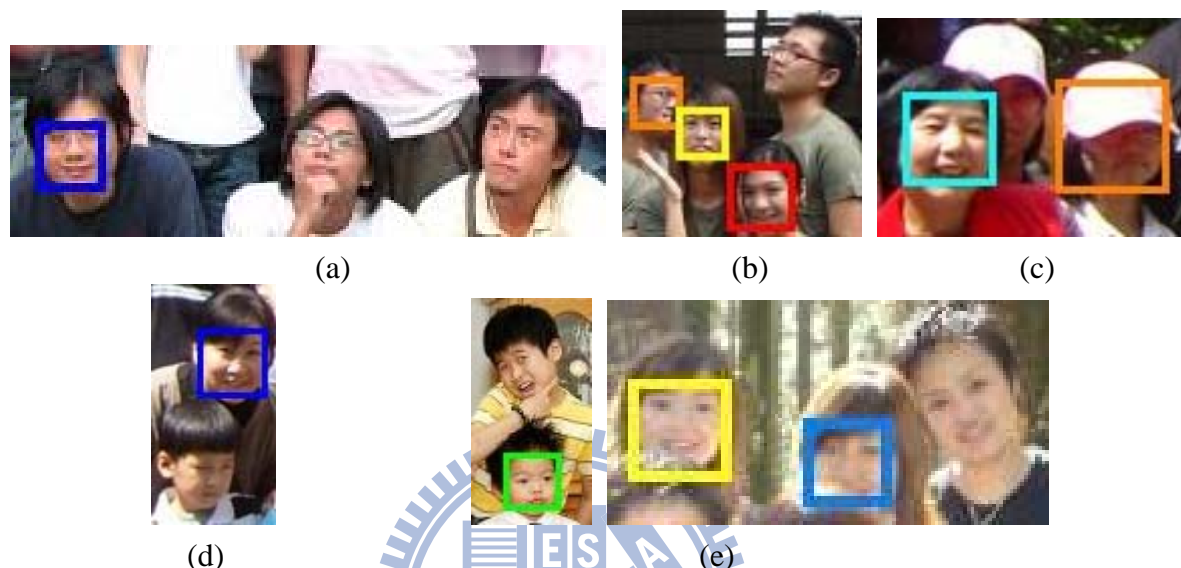


圖 4.2 (a)中間人臉因為眼鏡反光大及仰角太大而無法偵測；(b)右邊人臉因為仰角太大而無法偵測；(c)中間後面人臉因為遮掩物而無法偵測；(d)下面人臉因為光線不均勻而無法偵測；(e)因為傾斜角度太大及照明問題而無法偵測。

## 4.2 特徵人臉辨識率

特徵人臉(eigenfaces)辨識率整理於下面表 4.2，很明顯地，在不包含未知人臉分類條件下，AT&T ORL 資料庫[65]辨識率大約是生活照片的六倍(71% : 13% and 78% : 11%)，訓練樣本數越多，辨識率越高。

測試條件是，任意選取十個類別，每個類別包含十張人臉；第十一類別屬於未知人臉類別，任意挑選 150 張人臉。AT&T ORL 資料庫的人臉像素為  $10,304 (= 112 \times 92)$ ，生活照人臉像素為  $24776 (= 163 \times 152)$ ，所有訓練樣本都使用主成份分析法(principal component analysis)[50]降低維度，找出具有代表性的特徵組成人臉空間(face space)，輸入的測試樣本計算後得到每個特徵人臉的權重，然後使用歐幾里得距離(Euclidean

distance)計算跟人臉空間的距離，若距離大於臨界值(*threshold*，本實驗取 0.5)，表示屬於第十一類別，本實驗不取未知人臉類別來訓練，僅利用歐幾里得距離的臨界值來判斷是否屬於未知人臉類別。

訓練樣本取樣方式總共有兩種，一種是由上往下依序取樣(表 4.2)，另一種是由下往上依序取樣(表 4.3)，兩種取樣訓練方式的辨識率，差異不大(圖 4.3)。

生活照辨識率最好的就是取特徵人臉兩個訓練樣本(圖 4.4)，包含未知人臉類別的辨識率為 62%；但是，不包含未知人臉類別的辨識率卻下降到 11%，這是下一章節必須改善的地方。

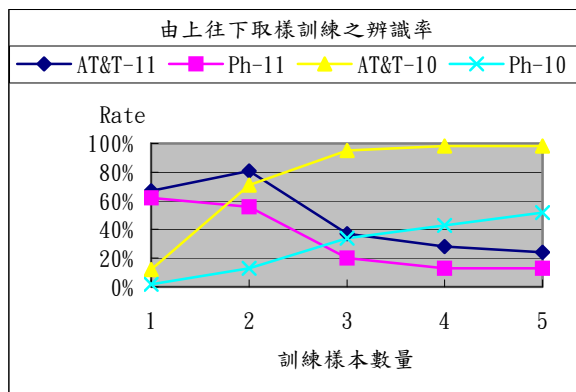
AT&T ORL 資料庫在不包含未知人臉類別測試裡，只要三個訓練樣本(圖 4.5)，辨識率就能高達 95%；但是，若包含未知人臉類別測試，訓練樣本增加到五個(圖 4.6)，辨識率快速下降到 37%，原因是這十個類別不容易跟未知人臉類別區分開來。

表 4.2 由上往下取樣訓練之特徵人臉辨識率在 AT&T ORL 資料庫跟生活照片的比較。

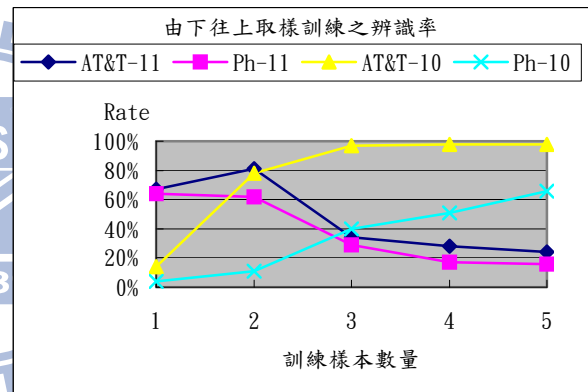
訓練樣本		辨識率 (threshold = 0.5)			
取樣方式	數量	包含未知人臉分類		不包含未知人臉分類	
		AT&T ORL 資料庫	生活照片	AT&T ORL 資料庫	生活照片
由上往下 (編號 1→5)	1	67%	62%	12%	2%
	2	81%	56%	71%	13%
	3	37%	20%	95%	34%
	4	28%	13%	98%	43%
	5	24%	13%	98%	52%

表 4.3 由下往上取樣訓練之特徵人臉辨識率在 AT&T ORL 資料庫跟生活照片的比較。

訓練樣本		辨識率 (threshold = 0.5)			
取樣方式	數量	包含未知人臉分類		不包含未知人臉分類	
		AT&T ORL 資料庫	生活照片	AT&T ORL 資料庫	生活照片
由下往上 (編號 10→6)	1	67%	64%	14%	4%
	2	81%	62%	78%	11%
	3	34%	29%	97%	40%
	4	28%	17%	98%	51%
	5	24%	16%	98%	66%



(a)



(b)

圖 4.3 兩種取樣訓練之特徵人臉辨識率比較，(a)由上往下取樣；(b) 由下往上取樣。



圖 4.4 生活照特徵人臉兩個訓練樣本。

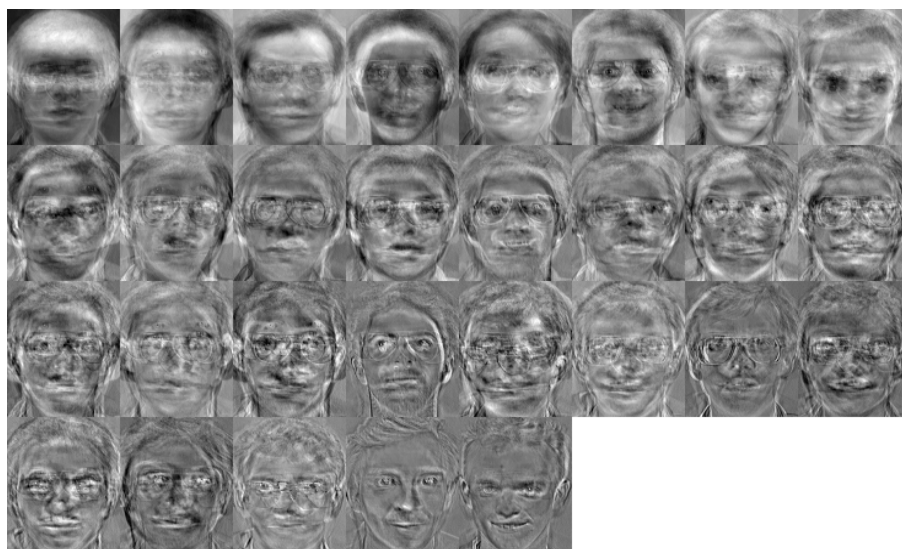


圖 4.5 AT&T ORL 資料庫特徵人臉三個訓練樣本。

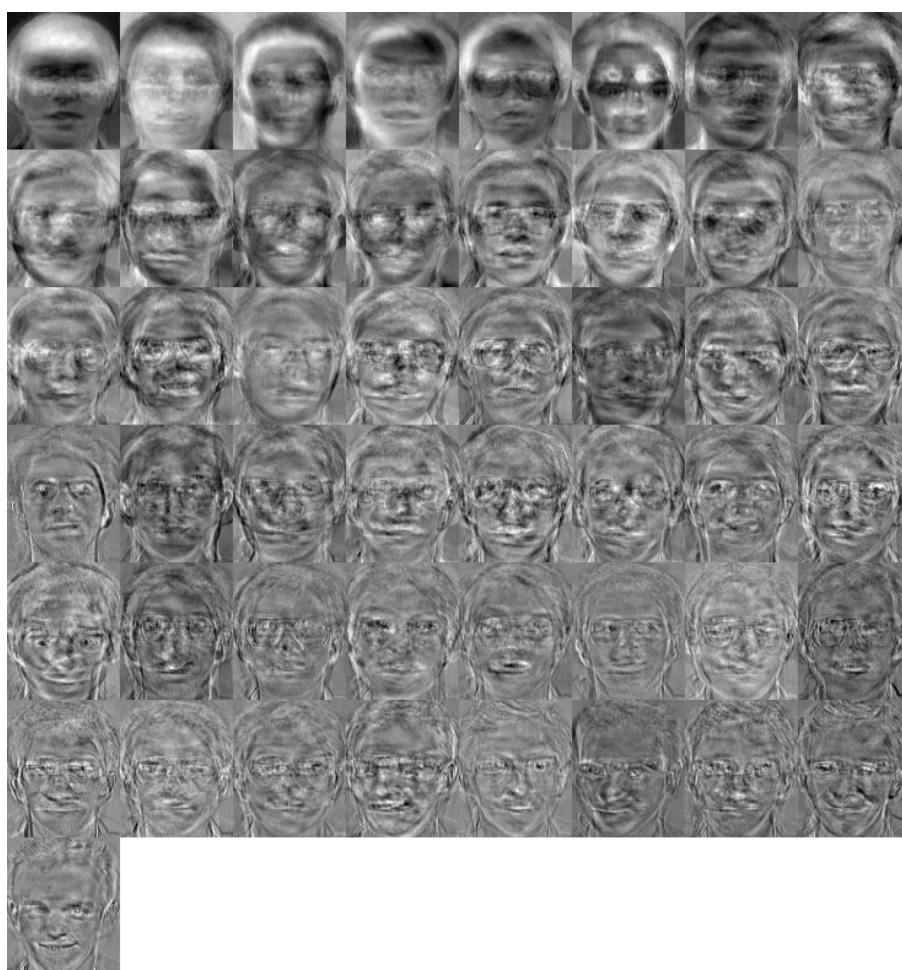


圖 4.6 AT&T ORL 資料庫特徵人臉五個訓練樣本。

### 4.3 稀疏編碼人臉辨識率

稀疏編碼人臉辨識率整理於表 4.4 跟表 4.5，AR 資料庫[64]辨識率分佈於 62.1% ~ 100%，生活照辨識率分佈於 46.7% ~ 85.0%，訓練樣本數越少，辨識率越低。

訓練樣本取樣方式總共有兩種，一種是由上往下依序取樣(表 4.4)，另一種是由下往上依序取樣(表 4.5)，與 4.2 節方式一樣，唯一不同的是，若包含未知人臉類別做分類，則固定取未知人臉類別編號 1~50 來訓練，剩下編號 51~150 為測試樣本；若不包含未知人臉類別做分類，則不取未知人臉類別來訓練或測試。

表 4.4 由上往下取樣訓練之稀疏編碼辨識率在 AR 資料庫跟生活照片的比較。

訓練樣本		辨識率			
取樣方式	數量	包含未知人臉分類		不包含未知人臉分類	
		AR 資料庫	生活照片	AR 資料庫	生活照片
由上往下 (編號 1→5)	1	87.9%	61.1%	97.8%	46.7%
	2	95.0%	68.9%	98.8%	62.5%
	3	97.6%	72.4%	100%	75.7%
	4	95.6%	73.1%	96.7%	85.0%
	5	96.7%	72.0%	96.0%	82.0%

表 4.5 由下往上取樣訓練之稀疏編碼辨識率在 AR 資料庫跟生活照片的比較。

訓練樣本		辨識率			
取樣方式	數量	包含未知人臉分類		不包含未知人臉分類	
		AR 資料庫	生活照片	AR 資料庫	生活照片
由下往上 (編號 10→6)	1	62.1%	60.5%	93.3%	47.8%
	2	62.8%	65.6%	98.8%	57.5%
	3	80.0%	66.5%	100%	65.7%
	4	85.0%	66.3%	100%	76.7%
	5	97.3%	69.3%	100%	82.0%

接下來，分析表 4.5 AR 資料庫辨識率 62.1% ~ 62.8% 偏低的原因，難道是正規化裡



面的明亮度處理得不夠好嗎？這部份我們下一章節 4.4 做實驗再釐清。

圖 4.7 呈現人臉資料 AR 資料庫(圖 4.7(a))跟生活照資料庫(圖 4.7(b))，差別在於生活照的人臉少了髮型跟下巴的資訊；AR 人臉像素為  $19,800(165 \times 120)$ ，生活照人臉像素為  $24776(163 \times 152)$ ，每一張人臉都經過多尺寸視網膜(multi scale retinex, MSR)處理明亮度後，再進行稀疏編碼運算。另外，也可以降低人臉解析度為  $1230(41 \times 30)$  或  $1520(40 \times 38)$ ，再進行稀疏編碼運算，不過這會降低辨識率  $1\% \sim 5\%$ 。

然而，辨識率  $62.1\% \sim 62.8\%$  偏低的原因，初步分析與人臉光線明亮度有關，在包含未知人臉類別情況下，採取編號 1~2(圖 4.8(a))當訓練的辨識率高達  $87.9\% \sim 95.0\%$ ，光線主要從正面人臉進來；採取編號 10~9(圖 4.8(b))當訓練樣本，光線主要從左臉進來，辨識率降低到  $62.1\% \sim 62.8\%$ 。圖 4.8(a)跟(b)只有訓練樣本明亮度的差異，表情變化並不大，圖 4.8(c)未知人臉類別 50 張訓練樣本包含明亮度高與低的人臉。所以，取明亮度不均勻的人臉當作訓練樣本，雖然會降低辨識率，不過，可以增加多張訓練樣本來彌補辨識率下降的問題。

圖 4.9 是 AR 資料庫經過多尺寸視網膜(MSR)正規化後的訓練樣本，明亮度一致性問題得到明顯改善。

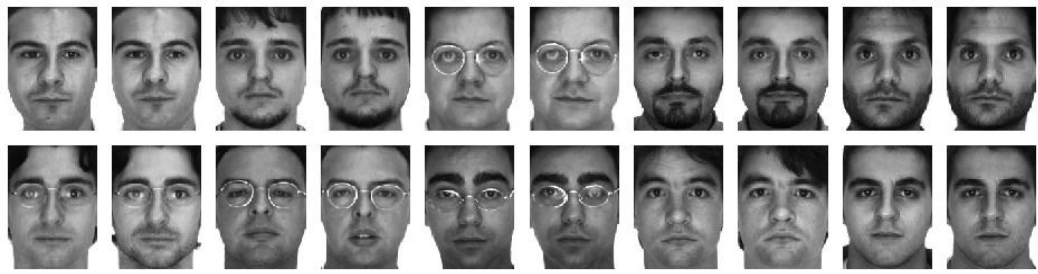
另外，生活照最好的辨識率就是  $73.1\% \sim 85.0\%$ ，條件是選取每一類別編號 1~4(圖 4.10(a))當訓練樣本，增加訓練樣本數目也將提高辨識率；圖 4.10(b)是固定選擇編號 1~50 當作未知人臉類別的訓練樣本，包含明亮度高與明亮度低的樣本。

圖 4.11 是生活照經過多尺寸視網膜(MSR)正規化後的訓練樣本，明亮度一致性問題也得到明顯改善。

圖 4.12 是本章節多尺寸視網膜正規化辨識成功的例子，圖 4.13 是舉出一個辨識失敗的例子，因為重建後的人臉跟錯誤類別有點類似。



圖 4.7 (a)從 AR 資料庫任意取出 250 張人臉做實驗；(b)從生活照資料庫任意取出 250 張人臉做實驗；(a)(b)的第一列到第十列分別代表每一個類別的人臉，每一類別有十張人臉，其餘十五列共 150 張人臉代表未知人臉類別。



(a)



(b)



(c)

圖 4.8 (a)(b)訓練樣本分別代表 AR 資料庫每一類別的編號 1~2 人臉，和 10~9 人臉；(c) 訓練樣本代表未知人臉類別編號 1~50 人臉；(a)跟(b)差異在於明亮度，表情差異不大；(c)包含明亮度高與明亮度低的訓練樣本。





(a)

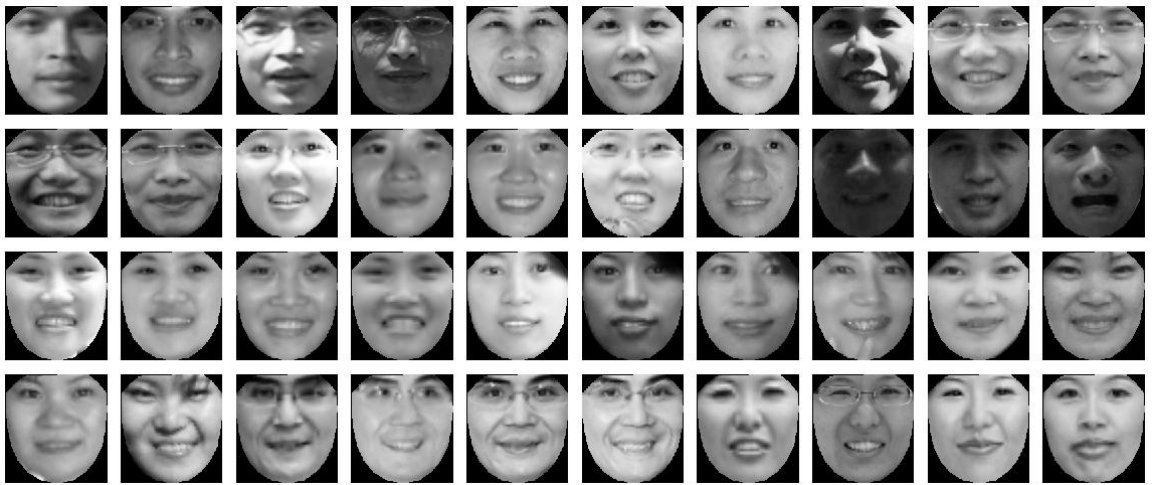


(b)



(c)

圖 4.9 AR 資料庫經過多尺寸視網膜正規化後得到的訓練樣本。



(a)



(b)

圖 4.10 (a)生活照資料庫的訓練樣本每一類別編號 1~4 人臉；(b)未知人臉類別的訓練樣本編號 1~50 人臉；(a)(b)包含明亮度高與明亮度低的訓練樣本。



(a)



(b)

圖 4.11 生活照經過多尺寸視網膜正規化後得到的訓練樣本。





圖 4.12 多尺寸視網膜正規化辨識成功的例子，最左邊是原始人臉，左二是經過正規化的人臉，右二是稀疏編碼重建的人臉，最右邊是最靠近類別的人臉。

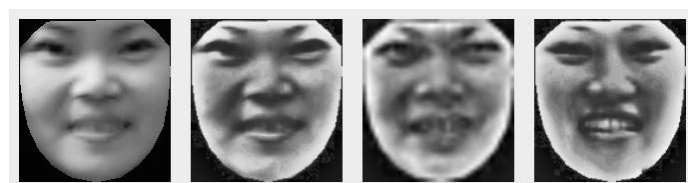
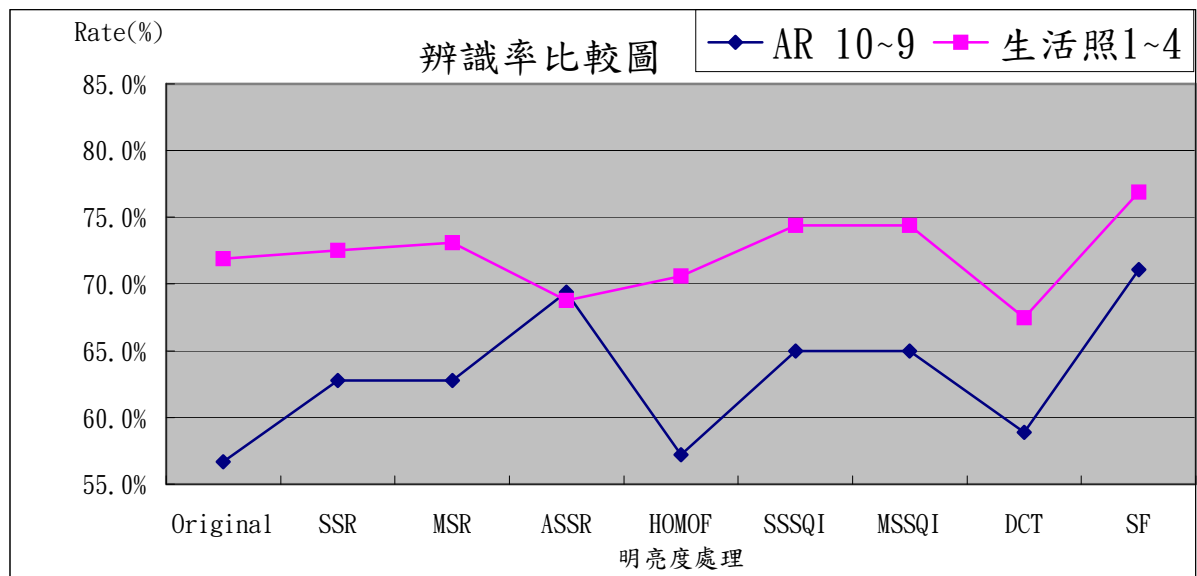


圖 4.13 多尺寸視網膜正規化辨識失敗的例子，最左邊是原始人臉，左二是經過正規化的人臉，右二是稀疏編碼重建的人臉，最右邊是最靠近類別的人臉。

#### 4.4 明亮度正規化對於辨識率的影響

本節實驗的訓練樣本，取樣方式跟 4.3 節一樣，測試條件包含未知人臉類別，固定取前面 50 張訓練，後面 100 張做測試；AR 資料庫採取編號 9 ~ 10 當訓練樣本，其餘編號 1 ~ 8 當測試樣本，取樣的理由是因為 4.3 節的表 4.5 辨識率偏低到只有 62.8%；生活照採取編號 1 ~ 4 當訓練樣本，其餘編號 5 ~ 10 當測試樣本，取樣的理由是因為 4.3 節的表 4.4 辨識率最高為 73.1%；每一張人臉都經過圖 4.14 明亮度八種演算法處理後，才進行稀疏編碼辨識人臉。

圖 4.14 測試結果顯示，經過可導引濾波器(steerable filter, SF)正規化後的辨識率還不錯，對於兩種資料庫都能同時增加辨識率，所以下一章節選擇可導引濾波器正規化，針對 4.3 節的表 4.4 跟表 4.5 相同的測試條件再測試一遍。



(a)



(b)

圖 4.14 (a)八種明亮度處理演算法之辨識率曲線比較圖；(b)最左邊對應的是 Original 為原始人臉，最右邊對應的是 SF 為可導引濾波器處理後的人臉，由左而右分別是 SSR，MSR，ASSR，HOMOF，SSSQI，MSSQI 和 DCT。

## 4.5 可導引濾波器正規化的人臉辨識率

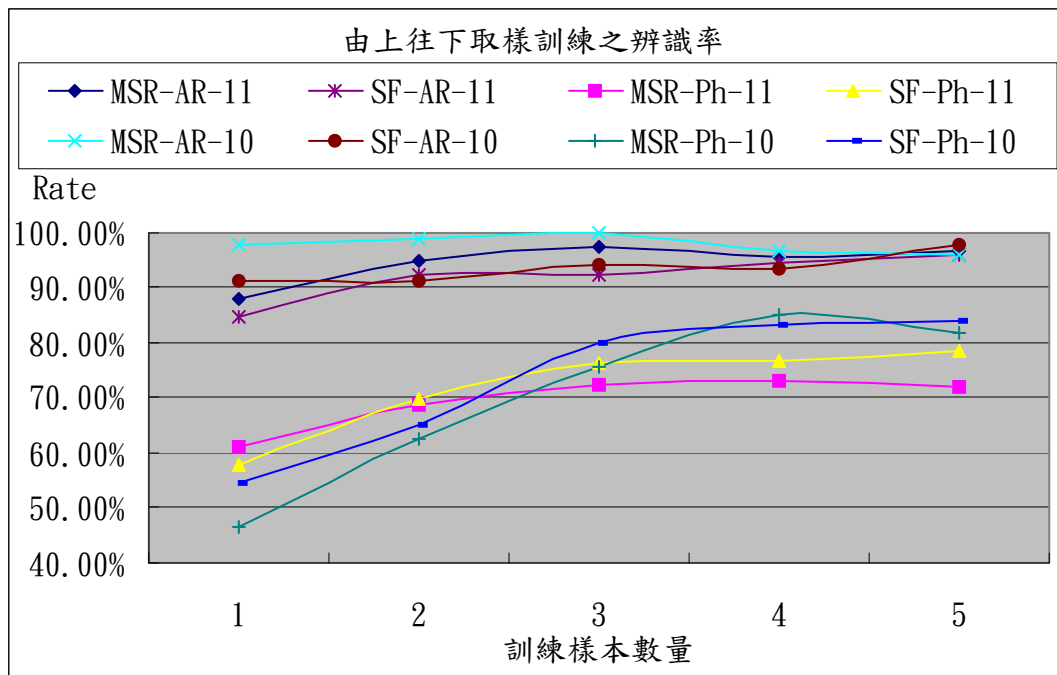
根據 4.3 節的表 4.4 跟表 4.5 的測試條件，只改變明亮度正規化，將多尺寸視網膜 (MSR) 改為可導引濾波器 (SF)，其餘測試條件都一樣，測試後的辨識率整理於表 4.6、表 4.7 和圖 4.15，整體而言可導引濾波器的辨識率比多尺寸視網膜來得好，AR 資料庫辨識率由 62.8% 增加到 71.1%。而圖 4.16 是舉例一個可導引濾波器辨識失敗的例子，圖 4.17 是可導引濾波器辨識成功的例子，重建後的人臉很接近測試人臉 (參考圖 4.16 右二人臉，圖 4.17 右二人臉)。圖 4.18 是生活照 (圖 4.10) 經過可導引濾波器正規化後的訓練樣本；而圖 4.19 是 AR 資料庫 (圖 4.8) 經過可導引濾波器正規化後的訓練樣本。

表 4.6 由上往下取樣訓練之可導引濾波器正規化後的辨識率在 AR 資料庫跟生活照片的比較。

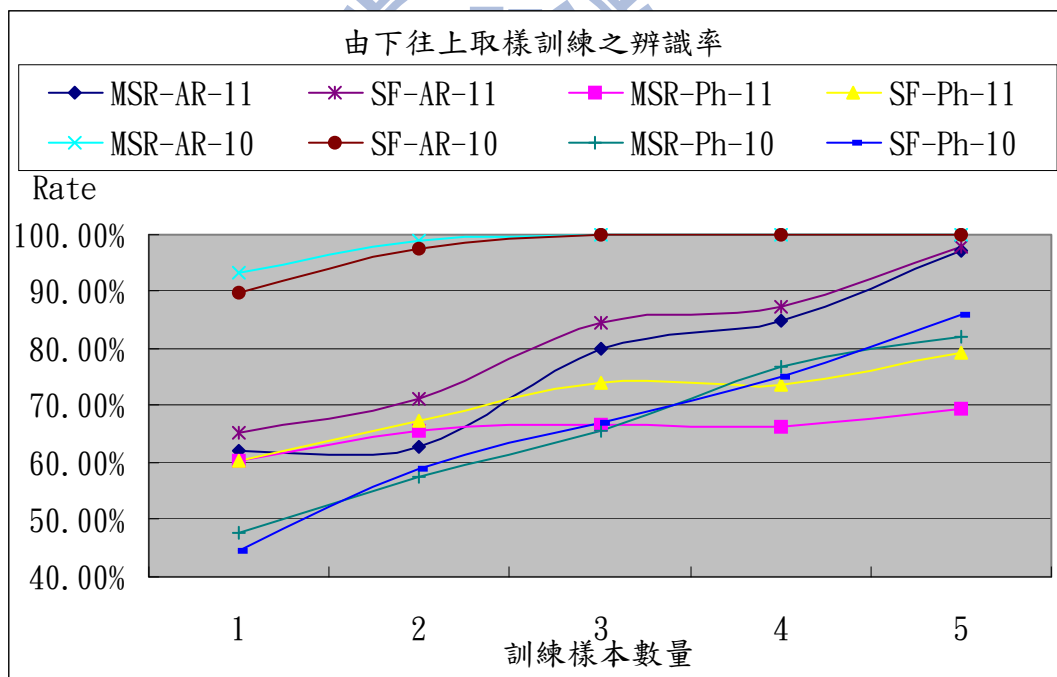
訓練樣本		辨識率			
取樣方式	數量	包含未知人臉分類		不包含未知人臉分類	
		AR 資料庫	生活照片	AR 資料庫	生活照片
由上往下 (編號 1→5)	1	84.7%	57.9%	91.1%	54.4%
	2	92.2%	70.0%	91.3%	65.0%
	3	92.4%	76.5%	94.3%	80.0%
	4	94.4%	76.9%	93.3%	83.3%
	5	96.0%	78.7%	98.0%	84.0%

表 4.7 由下往上取樣訓練之可導引濾波器正規化後的辨識率在 AR 資料庫跟生活照片的比較。

訓練樣本		辨識率			
取樣方式	數量	包含未知人臉分類		不包含未知人臉分類	
		AR 資料庫	生活照片	AR 資料庫	生活照片
由下往上 (編號 10→6)	1	65.3%	60.5%	90.0%	44.4%
	2	71.1%	67.2%	97.5%	58.8%
	3	84.7%	74.1%	100%	67.1%
	4	87.5%	73.8%	100%	75.0%
	5	98.0%	79.3%	100%	86.0%



(a)



(b)

圖 4.15 兩種取樣訓練之可導引濾波器(SF)與多尺寸視網膜(MSR)辨識率曲線比較圖，(a) 由上往下取樣；(b) 由下往上取樣。

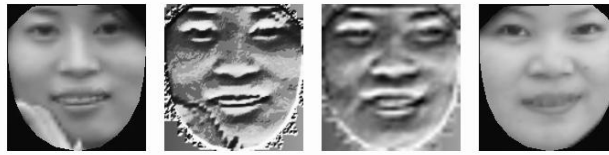


圖 4.16 可導引濾波器辨識失敗的例子，最左邊是原始人臉，左二是經過正規化的人臉，右二是稀疏編碼重建的人臉，最右邊是最靠近類別的人臉。



圖 4.17 可導引濾波器辨識成功的例子，最左邊是原始人臉，左二是經過正規化的人臉，右二是稀疏編碼重建的人臉，最右邊是最靠近類別的人臉。





(a)



(b)

圖 4.18 生活照經過可導引濾波器正規化後的訓練樣本。





(a)



(b)



(c)

圖 4.19 AR 資料庫經過可導引濾波器正規化後的訓練樣本。

## 4.6 人臉解析度對於辨識率的影響

為了分析人臉解析度(像素多寡)對於辨識率的影響，測試條件比照 4.5 節的表 4.7，唯一改變的條件就是縮減長寬比例，從下面表 4.8 測試結果可以發現，生活照人臉像素從 1520 下降到 420，辨識率下降 5.0%(= 76.9% - 71.9%)；AR 人臉像素從 1230 下降到 391，辨識率下降 5.6%(= 94.4% - 88.8%)；所以低於 400 像素的人臉辨識率將低於 70%。

表 4.8 人臉解析度之辨識率比較。

冪次方	0	1	2	3	4	5	6	7	8
縮放比例	1	1/2	1/4	1/6	1/7	1/7.5	1/8	1/16	1/32
AR 像素	19800	4920	1230	540	391	352	300	70	15
AR 1~4 辨識率	95.0%	96.3%	94.4%	92.5%	88.8%	85.6%	79.4%	78.1%	60.6%
生活照 1~4 辨識率	81.9%	80.0%	76.9%	75.6%	71.3%	71.9%	65.6%	65.0%	55.6%
生活照像素	24776	6156	1520	675	483	420	380	90	20

## 4.7 生活照片人臉辨識率分析

因為可導引濾波器結合稀疏編碼，在 AR 正面人臉和生活照兩種資料庫都有 80% 以上辨識率，所以接下來測試系統整體辨識率就採取這種方法。本系統定位為小型的人臉辨識系統，總共有 11 個類別，除了已知人臉有 10 個類別，每個類別任意挑選 5 個訓練樣本；另外還包括第十一類別是未知人臉類別，任意挑選 50 個訓練樣本。根據 4.1 章節實驗結果，生活照片總共偵測到 581 個人臉，全部都進行人臉辨識，測試結果分析條列於表 4.9，根據影響人臉特徵的因素，可歸類為明亮度不均勻、角度變化和遮掩物影響這三種組合。辨識率計算方式如式子(4.2)，整體而言，辨識率將近 83%，正面人臉因為特徵最清楚，所以辨識率最高，其次是明亮度不均勻跟角度兩大條件，辨識率最低的是遮掩物。

$$\text{生活照片人臉辨識率} = \frac{\text{辨識正確的人臉個數}}{\text{辨識正確的人臉個數} + \text{辨識錯誤的人臉個數}}, \quad (4.2)$$

表 4.9 生活照片人臉辨識結果分析。

生活照片辨識率			
人臉特徵	辨識正確的個數	辨識錯誤的個數	辨識正確率
(a)正面人臉	195	6	97.0%
(b)明亮度不均勻	109	14	88.6%
(c)角度影響	100	12	89.2%
(d)遮掩物影響	26	6	81.2%
(e)明亮度不均勻及角度影響	40	32	55.5%
(f)角度及遮掩物影響	9	8	52.9%
(g)明亮度不均勻及遮掩物影響	9	14	39.1%
(h)明亮度不均勻及角度和遮掩物影響	0	1	0.0%
合計	488	93	83.9%

## 4.8 實驗結果與討論

圖 4.20 是本系統的說明，圖 4.20(a)是偵測生活照兩個人臉的位置(藍色框框)；圖 4.20(b)是擷取人臉外形；圖 4.20(c)是主動式外觀模型正規化的人臉；圖 4.20(d)是男生錯誤的分類結果，主要是人臉一部份在樹蔭下，一部份在陽光下所導致的錯誤分類，四張圖由左到右分別是主動式外觀模型正規化的人臉、可導引濾波器正規化的人臉、稀疏編碼重建後的人臉，最右邊是稀疏編碼判斷跟資料庫最接近的人臉；圖 4.20(e)是女生正確的分類結果，四張圖由左至右的說明與圖 4.20(d)相同；圖 4.20(f)是取自資料庫的人臉當做辨識結果。由此例子說明光線分佈不均勻將會降低本系統的辨識率。





(a)



(b)



(c)



(d)

(e)



(f)

圖 4.20 生活照片之人物分類系統說明，(a)人臉偵測結果，(b)擷取的人臉外形，(c)主動式外觀模型正規化的人臉，(d)最右邊人臉是男生錯誤的分類結果，(e)最右邊人臉是女生正確的分類結果，(f)取自資料庫的人臉當做辨識結果。

## 第五章 結論與未來展望

本論文提出一個可以應用在相簿管理或數位相框的人臉偵測與人臉辨識方法，因為一般生活照片的拍攝條件，都不太一樣，這會增加人臉偵測與人臉辨識的困難度。所以在人臉偵測階段，採用膚色偵測法縮小搜尋人臉的範圍，以提升偵測效率。然後使用賈伯小波抽取人臉整體特徵，再輸入倒傳遞類神經網路做訓練，將使得人臉偵測到的位置與尺寸更為準確。

在人臉辨識階段，採用稀疏編碼(sparse coding)結合可導引濾波器(steerable filter)，生活照片辨識率可達 80%，AR 資料庫辨識率為 94.4% ~ 98.0%，基本上，訓練樣本越多，辨識率越高，但前提是，訓練樣本的定位點必須正確，否則，這些不準確的訓練樣本，將成為辨識系統的干擾訊號，會增加誤判的機率。

本篇論文提出直方圖統計法來減少稀疏編碼的權重數目，目的是為了降低系統運算量，同時特徵向量仍然具有代表性。

未來的展望，可以朝人臉辨識三個挑戰來探討，分別是姿態角度改變、明亮度不均勻和表情變化這三大問題: 1.增加正確定位人臉位置的機會，而且精確地切割人臉形狀，不讓特徵點偏移得太嚴重。2.正規化人臉像素的明亮度，並且允許一張明亮度不均勻的訓練樣本而得到良好的人臉辨識率。3.建立 3D 立體稀疏編碼來允許人臉大角度或表情的變化。



# 參 考 文 獻

- [1] B. Moghaddam and A. Pentland, "Probabilistic Visual Learning for Object Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 696-710, 1997.
- [2] H. A. Rowley, S. Baluja, and T. Kanade, "Neural Network-Based Face Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 23-38, 1998.
- [3] B. Heisele, T. Poggio, and M. Pontil, "Face Detection in Still Gray Images," *A.I. memo AIM-1687*, Artificial Intelligence Laboratory, MIT, 2000.
- [4] P. Viola and M. J. Jones, "Robust Real-Time Face Detection," *IJCV*, vol. 57, pp. 137-154, 2004.
- [5] P. Wang, M. B. Green, Q. Ji, and J. Wayman, "Automatic Eye Detection and Its Validation," in *Proc. CVPR*, vol. 3, pp. 164-171, 2005.
- [6] P. Kakumanu, S. Makrogiannis, and N. Bourbakis, "A Survey of Skin-Color Modeling and Detection Methods," *Pattern Recognition*, vol. 40, pp. 1106-1122, 2007.
- [7] S. Birchfield, "Elliptical Head Tracking Using Intensity Gradients and Color Histograms," *Computer Vision and Pattern Recognition*, pp. 232-237, 1998.
- [8] L. Wiskott, J. M. Fellous, N. Kruger, and C. Malsburg, "Face Recognition by Elastic Bunch Graph Matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 775-779, 1997.
- [9] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active Shape Models - Their Training and Application," *Computer Vision and Image Understanding*, pp. 38-59, 1995.
- [10] T. F. Cootes, G. J. Edwards and C. J. Taylor, "Active Appearance Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp 681-685, 2001.
- [11] D. Maio and D. Maltoni, "Real-Time Face Location on Grayscale Static Images," *Pattern Recognition*, vol. 33, pp. 1525-1539, 2000.
- [12] H. S. Lee and D. Kim, "Robust Face Tracking by Integration of Two Separate Trackers: Skin Color and Facial Shape," *Pattern Recognition*, vol 40, pp. 3225-3235, 2007.
- [13] M. N. Francesc, A. Sanfeliu, and D. Samaras, "Integration of Deformable Contours and a Multiple Hypotheses Fisher Color Model for Robust Tracking in Varying Illuminate

Environments,” *Image and Vision Computing*, vol. 25, pp. 285-296, 2007.

[14] Y. Tong, Y. Wang, Z. Zhu, and Q. Ji, “Robust Facial Feature Tracking under Varying Face Pose and Facial Expression,” *Pattern Recognition*, vol. 40, pp. 3195-3208, 2007.

[15] J. Tu, H. Tao, and T. Huang, “Face as Mouse Through Visual Face Tracking,” *Computer Vision and Image Understanding*, vol.108, pp. 35-40, 2007.

[16] M. Kim, S. Kumar, V. Pavlovic, and H. Rowley, “Face Tracking and Recognition with Visual Constraints in Real-World Videos,” *CVPR*, pp. 1-8, 2008.

[17] W. Zheng and S. M. Bhandarkar, “Face Detection and Tracking Using a Boosted Adaptive Particle Filter,” *Journal of Visual Communication and Image Representation*, vol. 20, pp. 9-27, 2009.

[18] M. Balasubramanian, S. Palanivel, and V. Ramalingam, “Real Time Face and Mouth Recognition Using Radial Basis Function Neural Networks,” *Expert Systems with Applications*, vol. 36, pp. 6879-6888, 2009.

[19] Y. Zhang and A. M. Martinez, “A Weighted Probabilistic Approach to Face Recognition from Multiple Images and Video Sequence,” *Image and Vision Computing*, vo. 24, pp. 626-638, 2006.

[20] G. Shakhnarovich and B. Moghaddam, “Face Recognition in Subspaces,” *Handbook of Face Recognition*, Springer, 2004.

[21] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman “From Few to Many: Generative Models for Recognition Under Variable Pose and Illumination,” *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 23, pp. 643-660, 2001.

[22] A. R. Chowdhury, R. Chellappa, S. Krishnamurthy, and T. Vo, “3D Face Recostruction from Video Using a Generic Model,” *ICME*, vol. 1, pp. 449-452, 2002.

[23] D. Jiang, Y. Hu, S. Yan, L. Zhang, H. Zhang, and W. Gao, “Efficient 3D Reconstruction for Face Recognition,” *Pattern Recognition*, vol. 38, pp. 787-798, 2005.

[24] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, “Robust Face Recognition via Sparse Representation,” in *IEEE PAMI*, vol. 31, pp.210-227, 2009.

[25] M. Yang, L. Zhang, J. Yang, and D. Zhang, “Robust Sparse Coding for Face Recognition,” in *CVPR*, pp. 625-632, 2011.

- [26] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition Using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 711-720, 1997.
- [27] S. Baker and T. Kanade, "Limits on Super-Resolution and How to Break Them," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 1167-1183, 2002.
- [28] X. Liu, T. Chen and S. M. Thornton, "Eigenspace Updating for Non-Stationary Process and Its Application to Face Recognition," *Pattern Recognition*, vo. 36, pp. 1945-1959, 2002.
- [29] G. Shakhnarovich, J. W. Fisher, and T. Darrell, "Face Recognition from Long-Term Observations," *Computer Science*, vol. 2352, pp. 851-865, 2002.
- [30] W. Y. Zhao and R. Chellappa, "Symmetric Shape-from-Shading Using Self-ratio Image," *Computer Vision*, vol. 45, pp. 55-75, 2001.
- [31] Y. Li, S. Gong, and H. Liddell, "Constructing Facial Identity Surfaces in a Nonlinear Discriminating Space," *CVPR*, vol. 2, pp. 258-263, 2003.
- [32] N. Vaswani and R. Chellappa, "Principal Components Null Space Analysis for Image and Video Classification," *IEEE Transactions on Image Processing*, vol. 15, pp. 1816-1830, 2006.
- [33] S. Du and R. Ward, "Wavelet-Based Illumination Normalization for Face Recognition," *ICIP*, vol. 2, pp. 954-957, 2005.
- [34] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 971-987, 2002.
- [35] O. Arandjelovic and R. Cipolla, "A Pose-Wise Linear Illumination Manifold Model for Face Recognition Using Video," *Computer Vision and Image Understanding*, vol. 113, pp. 113-125, 2009.
- [36] O. Arandjelovic, G. Shakhnarovich, J. Fisher, R. Cipolla, and T. Darrell, "Face Recognition with Image Sets Using Manifold Density Divergence," *CVPR*, vol. 1, pp. 581-588, 2005.
- [37] O. Arandjelovic and R. Cipolla, "An Illumination Invariant Face Recognition System for Access Control Using Video," in *Proceedings of the British Machine Vision Conference*, pp.

537-546, 2004.

[38] V. Blanz and T. Vetter, "A Morphable Model for the Synthesis of 3D Faces," in *Proceedings of International Conference on Computer Graphics*, pp. 187-194, 1999.

[39] W. T. Freeman and J. B. Tenenbaum, "Learning Bilinear Models for Two-Factor Problems in Vision," *CVPR*, pp. 554-560, 1997.

[40] Y. Adini, Y. Moses, and S. Ullman, "Face Recognition: the Problem of Compensating for Changes in Illumination Direction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp.721-732, 1997.

[41] D. W. Jacobs, P. N. Belhumeur, and R. Barsi, "Comparing Images Under Variable Illumination," *CVPR*, pp. 610-617. 1998.

[42] M. Savvides, B. V. K. V. Kumar, and P. K. Khosla, ""Corefaces"- Robust Shift Invariant PCA based Correlation Filter for Illumination Tolerant Face Recognition," *CVPR*, vol. 2, pp. 834-841, 2004.

[43] X. Tan and B. Triggs, "Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions," *IEEE Transactions on Image Processing*, vol. 19, pp. 1635-1650, 2010.

[44] T. Vetter and T. Poggio, "Linear Object Classes and Image Synthesis from a Single Example Image," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 733-742, 1997.

[45] S. Malassiotis and M. G. Strintzis, "Robust Face Recognition Using 2D and 3D Data: Pose and Illumination Compensation," *Pattern Recognition*, vol. 38, pp. 2537-2548, 2005.

[46] R. Gross, I. Matthews, and S. Baker, "Appearance-Based Face Recognition and Light-Fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 449-465, 2004.

[47] T.F. Cootes, G. Edwards, and C.J. Taylor, "Comparing Active Shape Models with Active Appearance Models," in *Proc. British Machine Vision Conference*, vol. 1, pp. 173-182, 1999.

[48] V. Struc, B. Vesnicer, F. Mihelic, and N. Pavesic, "Removing Illumination Artifacts from Face Images Using the Nuisance Attribute Projection," in *IEEE ICASSP*, pp. 846-849, 2010.

- [49] Toolbox for illumination invariant face recognition (the INface toolbox), [Online]. Available: <http://luks.fe.uni-lj.si/en/staff/vitomir/index.html>
- [50] M. A. Turk and A. P. Pentland, "Face Recognition Using Eigenfaces," *IEEE Proceedings of Computer Vision and Pattern Recognition*, pp. 586-591, 1991.
- [51] D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and Performance of a Center/Surround Retinex," *IEEE Transactions on Image Processing*, vol. 6, no. 3, pp. 451-462, 1997.
- [52] J. G. Daugman, "Complete Discrete 2-D Gabor Transforms by Neural Networks for Image Analysis and Compression," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 7, pp. 1169-1179, 1988.
- [53] L. Shen and L. Bai, "A review on Gabor wavelets for face recognition," *Pattern Analysis and Application*, vol. 9, pp. 273-292, 2006.
- [54] M. Yang and L. Zhang, "Gabor Feature Based Sparse Representation for Face Recognition with Gabor Occlusion Dictionary," *ECCV*, vol. 6, pp. 448-461, 2010.
- [55] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma, "RASL: Robust Alignment by Sparse and Low-rank Decomposition for Linearly Correlated Images," *CVPR*, pp. 763-770, 2010.
- [56] S. Yan, H. Wang, J. Liu, X. Tang, and T. S. Huang, "Misalignment-Robust Face Recognition," *IEEE Transactions on Image Processing*, vol. 19, no. 4, pp. 1087-1096, 2010.
- [57] L. Meylan and S. Susstrunk, "High Dynamic Range Image Rendering with a Retinex-Based Adaptive Filter," *IEEE Transactions on Image Processing*, vol. 15, pp. 2820-2830, 2006.
- [58] R. Gonzalez and R. Woods, *Digital Image Processing*, Prentice Hall, 3<sup>rd</sup> edition, 2008.
- [59] H. Wang, S. Z. Li, and Y. Wang, "Face Recognition under Varying Lighting Conditions Using Self Quotient Image," in *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 819-824, 2004.
- [60] W. Chen, M. J. Er, and S. Wu, "Illumination Compensation and Normalization for Robust Face Recognition Using Discrete Cosine Transform in Logarithmic Domain," *IEEE Transactions on Systems*, vol. 36, pp. 458-466, 2006.
- [61] W. T. Freeman and E. H. Edelson, "The Design and Use of Steerable Filters," *IEEE*



*Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 9, pp. 891-906, 1991.

[62] S. J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An Interior-Point Method for Large-Scale  $\ell_1$ -Regularized Least Squares," *IEEE Journal on Selected Topics in Signal Processing*, pp. 606-617, 2007.

[63] T. Sim, S. Baker, and M. Bsat, "The CMU Pose, Illumination, and Expression Database," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 1615-1618, 2003.

[64] AR Face Database, [Online]. Available:

<http://www4.comp.polyu.edu.hk/~csmyang/Publication.html>

<http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html>

[65] AT&T ORL Face Database, [Online]. Available:

<http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>

