

國立交通大學

電機與控制工程學系

碩士論文

多姿態人臉辨識及其在  
機器人與人互動之應用



Pose-Variant Face Recognition and  
Its Application to Human-Robot Interaction

研究生：王仕傑

指導教授：宋開泰 博士

中華民國九十七年九月

多姿態人臉辨識及其在  
機器人與人互動之應用

Pose-Variant Face Recognition and  
Its Application to Human-Robot Interaction

研究生：王仕傑

Student: Shih-Chieh Wang

指導教授：宋開泰 博士

Advisor: Dr. Kai-Tai Song



**Submitted to Department of Electrical and Control Engineering**

**College of Electrical and Computer Engineering**

**National Chiao Tung University**

**in Partial Fulfillment of the Requirements**

**for the Degree of Master**

**in**

**Electrical and Control Engineering**

**September 2008**

**Hsinchu, Taiwan, Republic of China**

中華民國九十七年九月

# 多姿態人臉辨識及其在機器人與人互動之應用

學生:王仕傑

指導教授:宋開泰 博士

國立交通大學電機與控制工程學系

## 摘要

本論文發展一套應用於機器人之多姿態人臉辨識系統。藉由 Active Appearance Model(AAM)的方法找出人臉的形狀模型與紋理模型後，再對輸入的人臉資料利用 Lucas-Kanade 影像校正的改良演算法將形狀模型的特徵點以迭代的方式擷取到多姿態的人臉，並將擷取到的人臉影像資訊透過特徵空間的維度化減後，再進入倒傳遞類神經網路(BPNN)來得出受測者為哪一位家庭成員。整套系統已在實驗室開發的嵌入式數位訊號處理器(DSP)平台予以實現，並且也成功的應用在實驗室所研製的寵物機器人上。實驗結果發現，若以五位受測者做家庭成員多姿態人臉辨識，在 UMIST 資料庫與實驗室內建的人臉資料庫而言，其辨識率分別可達 91%與 95.56%，驗證人臉在不同姿態的情況下，所提出之方法能有效辨識出受測者的人臉。

# **Pose-Variant Face Recognition and Its Application to Human-Robot Interaction**

Student: Shih-Chieh Wang

Advisor: Dr. Kai-Tai Song

Department of Electrical and Control Engineering  
National Chiao Tung University

## **ABSTRACT**

In this thesis, a pose-variant face recognition system has been developed for human-robot interaction. In order to extract the facial feature points from different poses, active appearance model (AAM) is employed to find the position of feature point. The improved Lucas-Kanade algorithm is used to solve the image alignment. After obtaining the location of feature points, the eigenspace of texture model is reduced the dimension and sent to the back propagation neural network (BPNN). By using the BPNN, the proposed recognizes that which family-member is the user. The proposed pose-variant face recognition system has been implemented on an embedded image system of a pet robot. In order to test our method, UMIST and self-built database are both used to evaluate the performance of the proposed algorithm. Experimental results show that the average recognition rate of the UMIST database and self-built database in our lab are 91% and 95.56% respectively. The proposed pose-variant face recognition system is suitable for applying to human-robot interaction.

## 致謝

謹向我的指導教授宋開泰博士致上感謝之意，感謝他兩年來在專業上的指導，以他豐富的學識與經驗，配合理論的應用，使得本論文得以順利完成。感謝口試委員胡竹生教授與莊仁輝教授的指導與意見，讓本論文能夠更加嚴謹。

感謝與我共同奮鬥的同學弘哲、兆戊及煥坤的相互鼓勵及提攜，以及學弟科棟、學妹巧敏及舒涵在生活上帶來的樂趣，同時感謝學長孟儒、嘉豪、格豪、振暘、志昇及裕宏在實作與理論上的指點。

感謝我的朋友們的勉勵與支持，並在研究上相互討論，讓我獲益良多。最後感謝我的父母與家人，由於他們的辛苦栽培，在生活上給予我細心地關懷與照料，使得我才得以順利完成此論文，在此我願以此論文獻給我最感激的父母親。



# 目錄

中文摘要.....	i
英文摘要.....	ii
致謝.....	iii
目錄.....	iv
圖例.....	vi
表格.....	viii
第一章 緒論.....	1
1.1 研究動機.....	1
1.2 相關研究回顧.....	2
1.3 問題描述.....	8
1.4 系統架構與章節說明.....	9
第二章 人臉影像偵測與特徵擷取.....	12
2.1 人臉偵測.....	12
2.1.1 膚色搜尋與選取.....	13
2.1.2 專注式串聯法.....	14
2.1.3 膚色搜尋與選取.....	16
2.2 人臉影像前處理.....	17
2.3 人臉形狀模型.....	18
2.3.1 特徵點的設置.....	18
2.3.2 形狀對齊.....	19
2.3.3 模擬臉部形狀的變化情形.....	22
2.4 紋理模型.....	24
2.4.1 物體的紋理.....	24
2.4.2 Piece-wise Affine Warping.....	25
2.4.3 像素內插.....	27
2.4.4 模擬臉部紋理的變化情形.....	27
2.5 可形變樣板追蹤演算法.....	28
2.5.1 Lucas-Kanade 影像校正演算法.....	29

2.5.2 Global Shape Normalizing Transform (GSNT).....	41
2.5.3 The Normalization Inverse Compositoin (NIC) Algorithm.....	46
2.5.4 結合 GSNT 之 NIC Algorithm.....	46
2.6 結論與討論.....	49
第三章 人臉辨識演算法.....	50
3.1 特徵臉演算法.....	50
3.2 倒傳遞類神經網路.....	53
3.3 資料庫建立流程.....	55
3.4 結論與討論.....	56
第四章 基於人臉辨識之互動系統設計.....	58
4.1 嵌入式影像平台.....	59
4.2 寵物機器人.....	61
4.3 結論與討論.....	62
第五章 實驗結果.....	64
5.1 建立資料庫.....	64
5.2 影像校正演算法迭代結果.....	66
5.3 人臉影像辨識結果.....	69
5.3.1 實驗室內建多姿態人臉影像資料庫.....	70
5.3.2 UMIST 多姿態人臉影像資料庫.....	71
5.4 家庭成員與非家庭成員之人臉辨識結果.....	73
5.5 結論與討論.....	75
第六章 結論與未來展望.....	77
6.1 結論.....	77
6.2 未來展望.....	77
參考文獻.....	79

## 圖例

圖 1-1 一般人臉辨識流程圖 .....	3
圖 1-2 利用 SQI 解決陰影部份的問題.....	5
圖 1-3 系統架構圖 .....	10
圖 2-1 人臉偵測系統流程圖 .....	13
圖 2-2 膚色分割及投影 .....	15
圖 2-3 人臉偵測專注式串聯法 .....	16
圖 2-4 正面人臉偵測實驗結果 .....	16
圖 2-5 影像內插 .....	18
圖 2-6 特徵點的選取 .....	19
圖 2-7 特徵點位置 .....	19
圖 2-8 其他姿態人臉之特徵點位置 .....	20
圖 2-9 未校正的形狀座標 .....	21
圖 2-10 經 Procrustes Analysis 校正過後的形狀座標 .....	22
圖 2-11 mean shape $s_0$ 與其他特徵向量所組成的形狀座標 .....	23
圖 2-12 滿足 Delaunay property 之三角形的外接圓 .....	26
圖 2-13 人臉形狀座標經 Delaunay 後的結果 .....	26
圖 2-14 三角形影像形變法 .....	26
圖 2-15 mean texture $A_0$ 與其他特徵向量所組成的紋理資訊 .....	27
圖 2-16 調整形狀與紋理參數形變得到的其他人臉 .....	28
圖 2-17 Lucas-Kanade 演算法迭代流程 .....	31
圖 2-18 Inverse Compositional 演算法迭代流程 .....	34
圖 2-19 對 $(x_i^0, y_i^0)^T + (\Delta x_i^0, \Delta y_i^0)^T$ 計算座標轉換得 $(x_i, y_i)^T + (\Delta x_i, \Delta y_i)^T$ .....	39
圖 2-20 傳統 AAM 演算法迭代出的正面人臉 .....	41
圖 2-21 傳統 AAM 演算法迭代出的具姿態變化之人臉 .....	41
圖 2-22 結合 GSNT 之 IC 演算法迭代出的正面人臉 .....	45
圖 2-23 結合 GSNT 之 IC 演算法迭代出的具姿態變化之人臉 .....	46
圖 2-24 結合 GSNT 之 IC 演算法迭代出之人臉 .....	48

圖 2-25 結合 GSNT 之 NIC 演算法迭代出之人臉.....	48
圖 3-1 二維影像資料轉換為一維向量.....	51
圖 3-2 資料分佈與其在主分量上的投影.....	51
圖 3-3 形變後之臉部紋理.....	52
圖 3-4 倒傳遞類神經網路架構圖.....	54
圖 3-5 倒傳遞演算法流程.....	55
圖 3-6 資料庫建立流程.....	56
圖 4-1 人臉辨識在機器人與人互動之硬體架構圖.....	58
圖 4-2 DSP 影像平台系統架構圖.....	59
圖 4-3 嵌入式影像平台實體圖.....	60
圖 4-4 實驗室研製之寵物機器人.....	61
圖 4-5 嵌入式人臉辨識系統應用於寵物機器人架構圖.....	61
圖 4-6 機器人與人互動之多姿態人臉辨識系統完整架構圖.....	62
圖 5-1 實驗室內建資料庫範例影像.....	65
圖 5-2 UMIST 資料庫範例影像[39].....	66
圖 5-3 實驗室人臉資料庫之 (a)形狀模型 (b)紋理模型.....	66
圖 5-4 實驗室資料庫之多姿態人臉經影像校正演算法迭代 15 次後結果.....	67
圖 5-5 實驗室資料庫人臉影像經迭代後之 RMS 差值收斂情形.....	68
圖 5-6 UMIST 人臉資料庫之 (a)形狀模型 (b)紋理模型.....	69
圖 5-7 UMIST 資料庫之多姿態人臉經影像校正演算法迭代 15 次後結果.....	69
圖 5-8 UMIST 資料庫人臉影像經迭代後之 RMS 差值收斂情形.....	69
圖 5-9 多姿態人臉辨識之類神經網路架構圖.....	70
圖 5-10 實驗室內建資料庫經類神經網路訓練之 (a)收斂情形 及 (b)初始狀態.....	71
圖 5-11 UMIST 資料庫之測試人臉多姿態影像圖.....	72
圖 5-12 UMIST 資料庫經類神經網路訓練之初始狀態以及收斂情形.....	72
圖 5-13 每位家庭成員的類神經網路架構.....	73
圖 5-14 類神經網路辨識流程圖.....	74
圖 5-15 每位家庭成員經類神經網路訓練之 (a)收斂情形 (b)初始狀態.....	74

## 表格

表 1-1 人臉識別與人臉驗證系統經 Gabor transform 後的辨識率[12].....	6
表 2-1 GSNT+IC 與 GSNT+NIC 迭代影像差值比較 .....	49
表 5-1 家庭成員之多姿態人臉辨識結果 .....	71
表 5-2 UMIST 資料庫經多姿態人臉辨識結果.....	72
表 5-3 家庭成員與非家庭成員之多姿態人臉辨識結果 .....	75



# 第一章 緒論

## 1.1 研究動機

近年來機器人的發展十分蓬勃，無論是在娛樂方面、看護方面及保全方面均步入一般民眾的家庭當中。在娛樂方面，例如恐龍機器人 Pleo[1]，它有喜怒哀樂的情緒反應，能眨眼透過眼神與人互動，作出跳舞、與人握手的動作，突然有巨大聲響，會輕顫露出害怕的神情，輕撫其背部則顯現受用的表情，甚至發出低鳴的聲音；而 NEC 公司發展的第二代產品 PaPeRo 機器人[2] (Partner Personal Robot)，其具有偵測環境的能力，能夠安全的在屋內移動，不發生危險，且能夠進行聲音辨識，可做日常生活的會話溝通；而在多人同時交談的環境下，PaPeRo 可透過多重位置辨識，分辨出不同人的談話內容；在臉部辨識方面，PaPeRo 的眼部內藏有相機，可偵測與物體之間的距離，同時可記憶人類的相貌，因此可以自動找出熟悉的臉孔並與之交談；SONY 公司的 AIBO ERS-210A/R 機器人[3]跟以前的 AIBO 不一樣的地方在於它多了被植入的識別應用晶片，該晶片使得它能夠識別主人的名字、聲音和面貌，並且可以自動充電，在 AIBO 機器狗記錄了主人的名字、聲音和面貌特徵後，即可在人群中找到自己的主人。能量不足時，機器人可自己尋找充電站進行充電，充電完畢後即可自我啟動，並離開充電器。而在看護方面，例如三菱重工所推出的看護機器人[4] Wakamaru，可辨識 10,000 個日常生活用字，且能認得主人，並與人對談。Wakamaru 可依據日常生活中所得的資訊來回應使用者；Wakamaru 經由臉部特徵辨識，總共可以辨識 10 張臉，在說話時，機器人會將自己的臉朝向使用者，並配合機器人的手勢做互動。而 Wakamaru 可記住使用者日常生活節奏，例如睡覺、起床及吃飯時間等等。當使用者在戶外時，可用手機或電腦經由網際網路透過 Wakamaru 來觀看家裡情況。在保全方面，例如皮托科技的住家保全機器人 Sentinel[5]，可透過網際網路能從任何位置遠端監控，使用者可以透過機器人看到或聽到實際情況。整合智能導航系統，機器人可以利用重複或隨機的方式，有效的巡邏一層樓，或者能夠執行任

何點對點的導航(如”到廚房去”),讓遠端使用者不需手動操縱及指揮方向。機器人的攝影鏡頭及麥克風可提供機器人所處附近環境的第一手觀察,而機器人內建的焦電感測器,可以立即感測到任何出現在它附近的人或物。

機器人要能感測從外界得來的資訊,以便可以和人有所互動且自主決定其行為,其中最重要的功能之一是需要一個可靠的人機介面,能從外界擷取重要的訊息,讓機器人知道下一步的行為為何。而要使得機器人和人之間能夠更自然地互動,且也希望機器人能辨認出此人是否為熟悉的人,例如家庭成員等等,則可以藉由人臉辨識讓機器人偵測到人類的身份,並依照人類不同身份有其對應的自主式行為,使得機器人不再是冷冰冰的機器。而寵物機器人可以辨識出互動對象為哪一位家庭成員,就像是一個真實的寵物陪在人類身旁能夠提供娛樂甚至是撫慰人心的功能,使得人與機器人間的互動更為自然。

## 1.2 相關研究回顧



人臉辨識是指給定一個場景的靜態圖像或動態影像,利用儲存有若干已知身份的人臉影像的資料庫驗證與鑑別場景中單個或多個人的身份。根據應用場景的不同,人臉識別可以進一步細分為人臉識別(Face identification)、人臉驗證(Face verification)以及監視列表(Watch list)。在人臉識別系統中,輸入人臉影像跟資料庫中儲存的所有人臉影像進行比對,系統返回資料庫中與輸入影像最相似的那張影像對應的人臉身份作為輸入影像的身份;在人臉驗證系統中,輸入人臉影像與聲稱的身份同時輸入到系統,系統根據輸入影像與資料庫中具有該聲稱身份的人臉影像之間的相似度來決定接受還是拒絕該聲稱的身份;監視列表其場景與人臉識別非常相似,不同之處在於若輸入影像跟資料庫中所有人臉影像的相似度都太低時,則輸入的人臉影像將被識別為身份未知,即拒絕接受。雖然這三種應用場景之間多少都會存在一定的差異,但完成人臉辨識的核心步驟是相同的,圖 1-1 所示為一般人臉辨識流程圖。

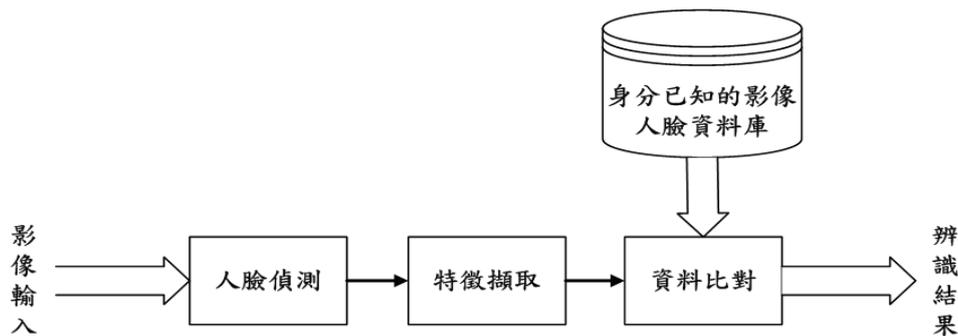


圖 1-1 一般人臉辨識流程圖

圖 1-1 中人臉偵測部份是用來判斷輸入影像是否存在人臉，如果存在則必須找出其位置和大小。經過多年的研究，人臉偵測技術已經取得了很大的進展，出現了許多性能很好的演算法；而特徵擷取部份在不同的文獻中有不同的定義，大部分文獻的特徵擷取實際上均指的是部位的檢測，如眼睛、嘴巴等等；資料比對指的是為了確定輸入人臉影像的身份，從輸入影像中擷取的特徵與資料庫中保存的已知身份的人臉影像的特徵進行比對，求出相似度，再對資料庫中人臉的身份按照相似度進行排序。在人臉識別中，系統會傳送最相似人臉的身份；而在監視列表中，如果資料庫中所有人臉與輸入人臉的相似度都低於某個臨界值，則就會拒絕該輸入人臉，否則就傳送最相似人臉的身份。

對於一個實用的人臉辨識系統來說，上述的三個步驟都是不可或缺的，但人臉偵測已經形成了單獨的研究議題並且也有了很大的進展，而資料比對結果的好壞決定於擷取特徵的正確性，因此在此會注重於特徵擷取的問題。

基本上人臉辨識可以分為兩種主要方法，一是整體特徵方法，一是局部特徵方法。整體特徵方法直接將整張人臉當作單一特徵來做辨識；局部特徵方法先找出臉上的局部特徵，通常是眼睛、鼻子和嘴巴，然後分別根據這些局部特徵做辨識，最後將個別局部特徵的結果統合而得到最後結果。

大部分的人臉辨識方法會先將原始影像資料經過降低維度的程序，以降低運算量，如 Turk *et al.* [6] 或 Belhumeur *et al.* [7] 使用 PCA 及 LDA 的方法來降低維度。在文獻[7]中指出，分別以 PCA 抽取 Eigenface 特徵的方法以及利用 LDA

抽取 Fisherface 特徵的方法來測試當臉部在光源變化、表情變化以及有配件（如眼鏡）的情況下做人臉辨識，其辨識結果是 Fisherface 的錯誤率較低。原因是 PCA 無法有效的區別樣本，且對於光線、角度的效果有限，而 LDA 則是另外加入不同類別間的區分度考量，使得 LDA 更具有辨別力的效果，可是對於同類別中差異太大（光源太亮或太暗或者臉部角度偏轉太多）的資料影響還是有所限制。Hafed *et al.* [8] 則是利用 Discrete Cosine Transform (DCT)來探討，它與 PCA 及 LDA 不同處在於 DCT 抽取的是影像低頻部分當作特徵，不用像 PCA 或 LDA 需要很多樣本來計算轉置矩陣，而是利用單一公式直接對每個樣本降低維度。

在其他眾多人臉辨識研究中，Wiskott *et al.* [9]利用 Elastic Bunch Graph Matching 的方法來解決人臉辨識在角度偏轉上的問題，作者在影像前處理的部份是先用 Gabor Wavelet 轉換來產生許多的 jet，這些 jet 就相當於臉部的基準點，如眼睛、嘴巴等等，再利用這些基準點對人臉做一個圖(graph)，而圖裡面的節點就是利用 jet 所得到的特徵，故在辨識時，若人臉角度有所轉動的話，圖就會變形而使得不同姿態下其臉部的特徵點仍位於臉上的相同位置，因此藉由這些特徵點作人臉辨識便可排除了不同姿態的因素，雖然 Gabor Wavelet 本身有克服光源變化的能力，但就要完全穩定來說，此文獻對於光源變化與複雜的背景還是有一定的缺陷在，且 Gabor Wavelet 是需要龐大計算量的，每多設一個特徵點，就要多計算 24 組頻率的 Gabor jet，並對影像中感興趣區域的每一個像素作摺積(convolution)，導致計算效率降低。

Wang *et al.* [10] 提出了一個解決人臉辨識在光源變化下的問題，利用 SQI (Self Quotient Image) 的方法來解決，其優點是只需用到一些影像處理的技巧，所以不必去做以經驗為依據的學習，另外就是此方法不必去做一些假設即可使用，它的作用在於能將臉部上的不同光線變化給去除掉並且去突顯臉部上的特徵，針對陰影的問題，此方法是最能展現出效果的，如圖 1-2 所示。



圖 1-2 利用 SQI 解決陰影部份的問題[10]

Jadhav *et al.* [11] 一開始先將影像利用兩次的小波轉換分成很多影像形成金字塔圖像特徵來處理光源與臉部表情的變動，再將這些影像所得到的資料利用 KPCA (Kernel Principal Component Analysis) 的方法將輸入圖像資料映射到一非線性映射的特徵空間裡，將空間裡的資料做 KFLD (Kernel Fisher's Linear Discriminant) 使得資料分類較容易，最後再用 WKFC (Wavelet Kernel Fisher Classifier) 將人臉給辨識出來。使用 kernel 領域的方法，其優點在於可以讓非線性的資料投影至特徵空間裡，不需要做線性的假設，進而可讓辨識率提高，而此方法的缺點在於雖然可讓非線性的資料投影至特徵空間裡，但是此非線性的空間是高維度的空間，因此會有計算速度上的考量。

Torres *et al.* [12] 先將影像分割成 108 個點後，再將所取到的點做 Discrete Gabor Transform 來求出特徵向量，而辨識部份分為兩種，一種為人臉識別系統，如前面所述，將抓取到的影像拿去對資料庫的人臉做比對來找出最接近的人臉；另一為人臉驗證系統，與前述不同的是輸入要多出聲稱自己的身份，作者認為使用 Gabor transform 的優點在於可克服一些外在的影響，在此篇文獻中有假設多種外在因素，如表 1-1 所示，例如光源、表情、配件、高斯雜訊、人臉旋轉角度 (5 到 10 度) 及年齡，其在人臉驗證系統都有一定的辨識效果，但在人臉識別系統而言，因為此系統是直接將影像去對資料庫做比對，所以容易對於角度變化或年齡變化產生很大的影響，以至於辨識率非常不佳。

表 1-1 人臉識別與人臉驗證系統經 Gabor transform 後的辨識率[12]

Image variation	Identification rate	Verification rate
Illumination	85.67%	99.33%
Facial expression	83.75%	99.21%
Accessories	83.50%	99.23%
Gaussian filter	84.30%	99.54%
Median filter	81.60%	99.19%
Gaussian noise	84.72%	99.43%
Impulsive noise	84.56%	98.76%
Resize	85.40%	99.30%
Rotation 5°	32.70%	90.20%
Rotation 10°	20.50%	87.20%
Shifting	85.10%	98.61%
Age	X	83.30%

Jia *et al.* [13] 利用 AAM(Active Appearance Model)以迭代的方式找出人臉的形狀(shape)與紋理(texture)，即使在人臉角度變化時都還是可以追蹤到人臉，但是若角度變化太大時，以 2D 影像來說，就必須再建立其他角度上的模型。Xiao *et al.* [14] 為了朝這方面改進，就提出一套利用 2D 的 AAM (Active Appearance Model) 來估算出 3D 的 AAM，再利用 2D 與計算出來的 3D AAM 結合起來去對人臉做迭代匹配的動作，進而在許多角度的變化上有較大的容忍性，並可估算出人臉的朝向角度。藉由 2D 估算出 3D 的 AAM，其優點在於模型的建構較為容易，不一定要找出真正的 3D 影像利用 3DMM(3D Morphable Model)來做分析；Gross *et al.* [15] 也是針對傳統的 AAM 做改良，主要是考慮臉部的遮蔽問題，例如臉部有部份被物品遮住或者是人臉轉動角度過多等。在利用 PCA 建立模型時考慮到 missing data 的部份來得到特徵向量，找出穩定的 Error function，並搭配此文獻提出的 Lucas-Kanade 演算法的改良迭代出遮蔽後的人臉。

Liu *et al.* [16]主要是要解決例如 PCA、LDA 及支持向量機(Support Vector Machine, SVM)分類器所產生的缺點，就 PCA 與 LDA 而言，是將資料轉換到線性子空間去做分析，但因為就資料實質上來說並非是線性的，導致分析時會有許多誤差產生，於是作者就使用 KFDA (Kernel Fisher's Discriminant Analysis)來解決

上述的問題，但此方法若是樣本不足則還是無法發揮效用。至於在 SVM 的分類器方面，其主要有兩個問題產生，一個是對於 outlier 的靈敏度很高，因為 SVM 會對於辨識不好的結果的資料其所給予的權重是一樣的，另一個為 class-boundary-skew 的問題，因為作者在其他一些文獻中發現當 SVM 在處理不平衡的資料量時，若 negative data 在數量上超過 positive data 時就會出現此問題，而作者也提出了 TAF-SVM (Total Margin-Based Adaptive Fuzzy Support Vector Machine) 的方法，主要是針對柔性邊界的 SVM 做改良，進而可在不平衡的訓練集合中找出最佳的分割平面(Optimal Separating Hyperplane, OSH)，且搭配上模糊理論使得所找到的 hyperplane 不會有 overfitting 的問題存在。

Ahonen *et al.* [17] 利用 LBP(Local Binary Pattern)找出人臉裡面各個部位的權重，以 3×3 的遮罩舉例來說，是將中心以外的週遭 8 個 pixel 做二值化的轉換，以中心像素設為臨界值，若大於中心值，則設為 1，反之為 0，然後將這些 0、1 串起來組成一個二進制的編碼，進而來找出權重以便做後續的分析，對於表情變化之人臉影像也不會受到影響，且由於此方法是用在灰階影像來分析，故對於光源變化而言並不會有太大的影響。

陳文俊[18]先是將偵測到的人臉做正規化的動作，包括有特徵(眼睛)的擷取、臉部的調整以及臉部亮度的補償，臉部的調整主要是針對兩眼的距離做一個縮放的功能，且也根據眼睛的連線做一個人臉角度的校正，若要新增資料庫成員時，可利用 on-line 的方式直接藉由新增影像對原有的特徵空間進行更新，不需再取出所有的資料進行計算，解決了時間上的問題，除了在資料方面做了 PCA 的降低維度運算外，作者選擇使用 RBF(Radial Basis Function)類神經網路作為辨識系統，因為作者也在其他文獻中得知此方法具有快速學習的好處，適合用於即時系統，且其最佳化權重值可由 Normal Equation 來求得，減少在訓練時所需花的時間。

汪慶祥[19]認為在日常生活中並不會用到完整三度空間的歐式幾何座標結構，而只對投影幾何中較有限制且簡單的結構來處理。在 3D 影像取得方面，主

要是利用兩張以上對同一場景拍攝而得的相片做投影重建(Projective Reconstruction)而獲得 3D 重建的影像資料，最後則用 2D 對應到 3D 的特徵點拿去對資料庫做比對進而產生辨識結果。

林晁立[20]在特徵擷取方面是先找出臉部各器官的形狀，再以此來找出臉部各器官之間的相對位置及距離，其原因為雖然各器官形狀容易因表情及情緒的變化而改變，但若是用臉部各器官相對位置則無此限制，作為從事人臉辨識的重要依據。而在人臉辨識方面是利用競爭型類神經網路搭配人臉特徵擷取去實現，因為作者發現競爭型類神經網路不需要繁複的訓練和冗長的計算過程，便能迅速的獲得結果。

陳嘉雄[21]是將每位受測者的人臉以不同角度均放入資料庫裡，其分類方式為大分類和細分類，大分類的主要目的是將每位受測者的人臉朝向角度做區隔，例如正面的為一類，轉 30 度角的為一類，以此類推；而細分類則是將同一類別裡的人辨識出來，此方法的缺點是同一個人需要存很多張不同角度的影像進入到資料庫裡，導致比對時，平均一個資料庫成員的比對影像張數增加，影響辨識速度。

梁皓雲[22]主要是用人臉分割特徵區塊的方法來做人臉辨識，分割的區域有眼睛、眉毛及嘴巴的部份，劃分區域可以減少人臉上的膚色區域，其原因是要減少因光線或環境的變化而影響其辨識結果，演算法上面主要是利用 LDA+PCA 的方法，同時在前處理的步驟採用小波轉換，在時間上都有明顯的減少，最後在做比對的辨識上是利用了比較權重的投票法，可將個人的特徵更明顯的表示出來，使得辨識率往上提昇。

### 1.3 問題描述

由於機器人技術逐步成熟，正步入一般民眾的家庭之中，機器人提供娛樂功能成為應用上的發展目標。而要使得機器人和人之間能夠互動地更加自然，希望

讓機器人在家庭之中辨別出對方是哪位家庭成員後，而給予對應的自主性行為娛樂家庭成員。

對於寵物機器人來說，通常機器人與使用者互動時，一般來說其在高度上可能會與人類有所差距，導致機器人可能不知道使用者在面前而造成一些角度上的誤差，或是因使用者未正視機器人上的攝影機進而也造成左右角度上的一些偏差，造成辨識上的困難。故本論文目標為開發一套寵物機器人與人互動之多姿態人臉辨識系統，使寵物機器人能夠在不同姿態的狀況下辨識出對方為哪位家庭成員。

在人臉影像特徵的擷取上，本論文主要是採用整體特徵的方法，雖然在分析上其計算量會較大，但就辨識來說，由於所有的影像資訊都考慮進去，在分析上會較為穩定。若可將搜尋到的人臉中雜訊的部份濾除，例如背景部份，則成功率更可因此而獲得提升。若是採用局部特徵的方法，當有某一局部的資訊發生一些損失時，其對於整體辨識成功率而言會下降許多。

在本論文最主要的貢獻即為不同姿態下之人臉辨識，與文獻[12]比較之下，本論文在多姿態的人臉辨識最多可偏轉 30 度角，其辨識率較文獻[12]的旋轉 5 度還要高出 5% 的辨識率，驗證了使用者在未正視攝影機狀態下依舊能有穩定的辨識率。另外，本論文將演算法均寫入一嵌入式影像系統裡，所有運算都不必透過個人電腦即可讓系統動作，並整合在寵物機器人上，進而與使用者產生互動，增加本系統的實用性。

#### 1.4 系統架構與章節說明

系統架構圖如圖 1-3 所示，其主要分為三個部分，分別是人臉偵測、臉部影像的特徵擷取以及人臉辨識的決策部分。在人臉偵測的部份中，主要是有考慮光源變化的情況，來偵測影像中人臉的所在位置；在人臉已被偵測並且框選出後，由於人臉在影像中其大小並不一定完全一致，且會有其他姿態下所偵測到的人臉

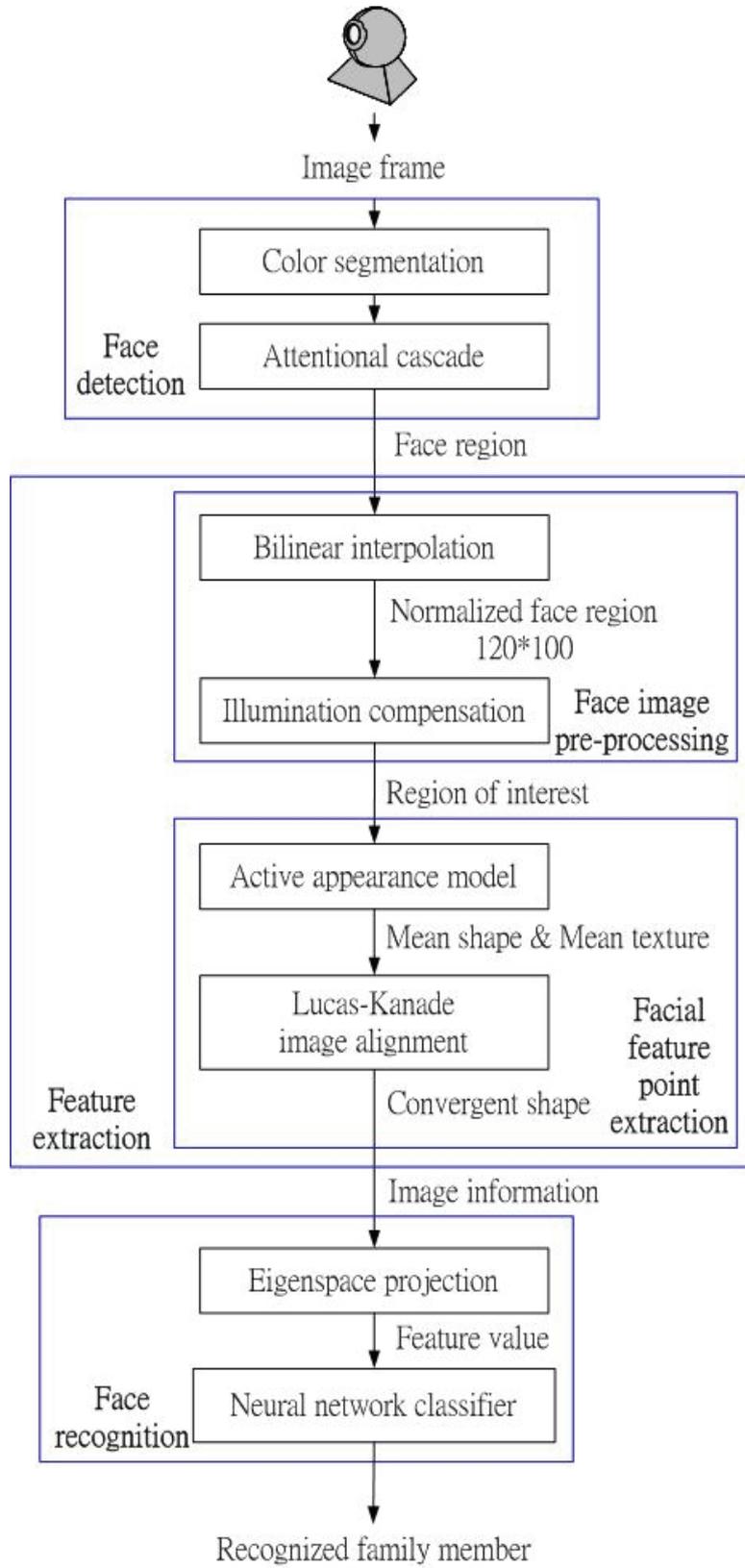


圖 1-3 系統架構圖

情形，另外在環境中的光線由於隨時會變動，造成臉部影像的亮度經常改動，故必須將影像大小做正規化以及照度補償的動作，先讓每張被偵測到的人臉，其大小均一致，再對亮度部份做補償，讓每張測試影像的亮度變化能減緩，並將所有姿態下的人臉放置在框選區域裡正中間的位置，再利用 AAM(Active Appearance Model)演算法與 Lucas-Kanade 影像校正演算法迭代出測試人臉影像，得到最終的人臉影像資訊；最後將所得到的整體影像資訊利用 PCA 投影至特徵空間並作影像維度的化減，將化減過後的資料經由倒傳遞類神經網路(Back Propagation Neural Network, BPNN)分類來得到家庭成員的辨識結果。

本論文共分六章，第一章為介紹相關研究背景及研究動機。第二章介紹人臉影像偵測與特徵擷取。第三章介紹人臉辨識演算法。第四章介紹人臉辨識的互動系統設計。第五章為實驗結果。第六章為結論與未來展望。



## 第二章 人臉影像偵測與特徵擷取

為了使寵物機器人可以對家庭成員進行辨識以提供協助，首先要讓機器人可以偵測出人臉，此部份為機器人視覺系統初始化的狀態。在本章中，除了將詳述如何在環境中找到人臉所在的位置以外，當偵測到人臉之後，並會詳述人臉特徵擷取的過程以進行後面的辨識。

在一些常見的演算法中，通常是先利用人臉偵測將人臉位置框選出來並做一些影像前處理(影像維度大小正規化與照度補償)後，再直接把前置處理過後的影像資訊均當作特徵來做辨識的演算法，假設在框選的人臉區域裡面，以非正面的人臉來說，很容易在框選區域裡面出現人臉以外的背景，若框選到的背景成分太多，則辨識錯誤的可能性則會提高許多。因此，我們採用主動外觀模型(Active Appearance Model, AAM)的方法來做人臉影像上的特徵擷取部份，此演算法主要分成兩個主要部份，第一主要部份是要建出人臉的形狀模型(Shape model)與紋理模型(texture model)，利用主成份分析(Principal Component Analysis, PCA)的方法來找出人臉的兩個模型以及模型的特徵向量；第二主要部份則是利用已建好的模型藉由迭代的方式將人臉影像的特徵給擷取到，再將擷取到的人臉影像利用影像形變的方式將擷取到的影像形變成與模型一樣大小的影像，進而來去做之後的辨識演算法。

### 2.1 人臉偵測

當輸入影像進入本系統時，首先需要做人臉偵測，找出人臉的位置以便作接下來的人臉特徵點擷取，在此我們採用文獻[23]設計的人臉偵測方法，流程如圖 2-1 所示。為了使系統能夠快速的搜尋到人臉位置，在進行人臉偵測之前，我們會先將輸入的 320x240 彩色影像縮小為 160x120 彩色影像，以減少人臉偵測所需的時間。在人臉偵測的過程當中，系統會先將輸入的影像進行膚色的色彩分割，並且透過形態學的閉合運算 (Closing) 填滿膚色區域內空洞及不連續之處。接著

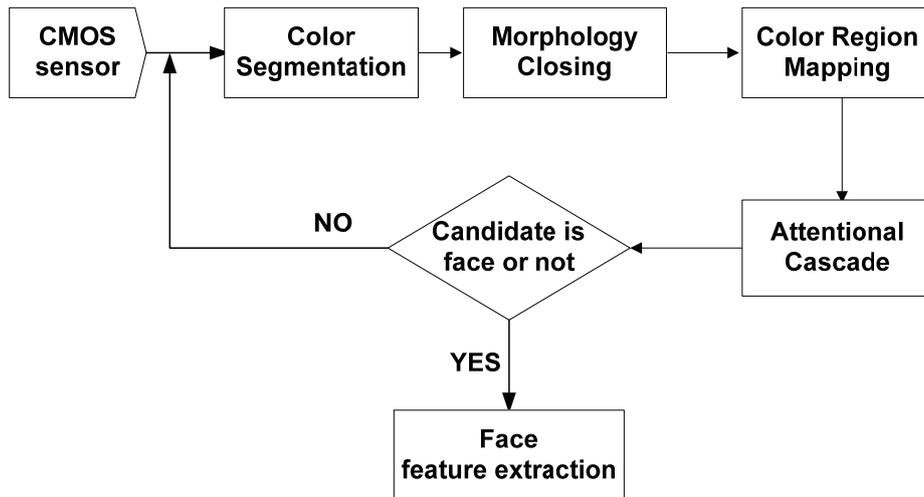


圖 2-1 人臉偵測系統流程圖

我們透過膚色區域投影將屬於膚色的區域框選出來，並且利用投影的長寬比先排除可能不屬於人臉的區域，再將可能屬於人臉的區域交由專注式串聯法 (Attentional Cascade) [24]來判別此區域是否為正面的人臉。以下 2.1 節部份的章節將對圖 2-1 的各元件分別加以說明。

### 2.1.1 膚色搜尋及選取

色彩分割就是找出膚色收斂的範圍並且設定膚色之色彩分佈上下閾值，若設定的範圍太大，則會將不屬於膚色的背景判定為膚色像素。因為膚色為人臉偵測的第一個步驟，必須先找到膚色分佈才可以進行色彩分割，所以我們根據在不同的環境及光源的膚色色彩分佈的情況，調整其膚色分佈閾值。圖 2-2(a)為膚色分割測試圖，圖 2-2(b)將膚色分割出來。接下來透過形態學的閉合運算填滿膚色區域內空洞及不連續之處，如圖 2-2(c)所示。

接著我們希望從這些二值化的影像中找出可能為人臉的區域，以縮小搜尋的範圍，即利用先前所求得色彩分割後的資訊  $I_{skin}$  進行可能人臉區域的判別。我們利用色彩分割後的資訊作水平軸與垂直軸上的投影以找出符合要求的區域。前面我們已求得的特定色彩資訊  $I_{skin}$  為一  $M \times N$  的矩陣，其中  $M$  為影像高，

$N$  為影像寬， $I_{skin}(x, y) = 1$  表示影像中  $(x, y)$  座標上的像素屬於膚色的色彩分佈，反之，則為非膚色。我們將  $I_{skin}$  在  $Y$  軸上進行投影求得  $I_{skin}$  在  $Y$  軸上的投影量  $H(y)$

$$H(y) = \sum_{x=0}^M I_{skin}(x, y) \quad (2-1)$$

從  $H(y)$  中根據設定最小的人臉長，選出  $K$  個  $Y$  方向的區間，接著我們再對這  $K$  個  $Y$  區間內的膚色資訊利用式(2-2)各別進行  $X$  軸上的投影，其投影量各別為  $W_i(x)$ ， $i$  為 1 到  $K$ ， $YS$  為起始投影  $Y$  值， $YE$  為最終投影  $Y$  值。

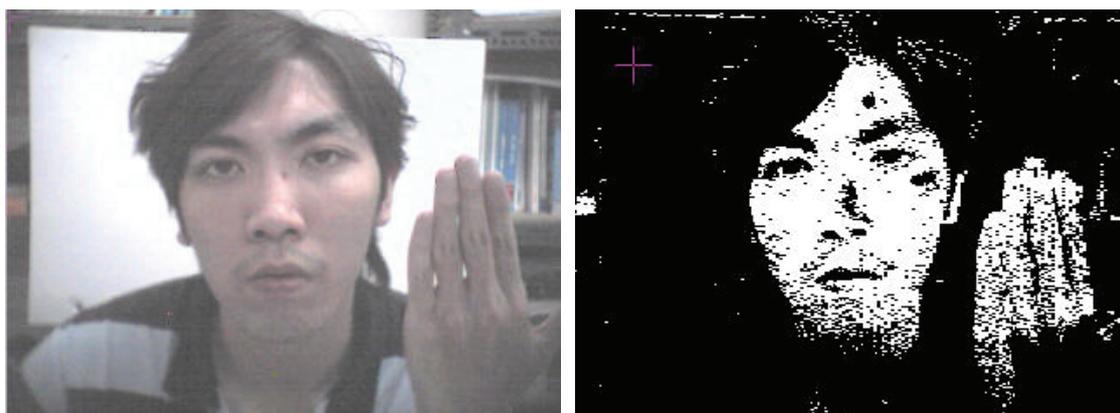
$$W_i(x) = \sum_{y=YS_i}^{YE_i} I_{skin}(x, y), \quad i=1 \sim K \quad (2-2)$$

從  $W_i(x)$  根據設定的最小的人臉寬進行分割，分割出數個可能為人臉的區間，如圖 2-2(d) 所示，經由第一次的投影，框選出  $A1$  及  $A2$  兩個區域，然後再對這兩個區域進行第二次的投影，如圖 2-2(e) 所示，用意在於移除不是膚色的區域。



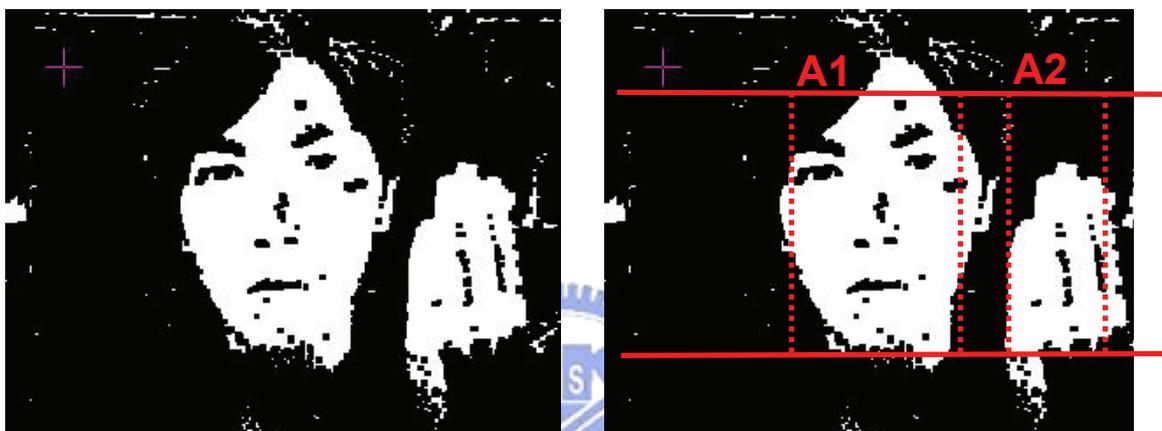
### 2.1.2 專注式串聯法

透過上一節將屬於膚色的區域框選出來，並且利用投影的長寬比先排除可能不屬於人臉的區域，再將可能屬於人臉的區域交由專注式串聯法 (Attentional Cascade) 來判別此區域是否為人臉[24]。專注式串聯法可解決過多的運算量浪費在判斷非人臉影像視窗的問題。在此人臉偵測器主要是使用數個簡單的臉部特徵，稱為簡易分類器 (Simple Classifier)，串聯成為一個複雜分類器 (Complex Classifier)。愈前面的分類器規則愈容易，所需運算量也愈少，但是具有快速判斷影像視窗為人臉或者是背景之效果；愈後面特徵則是愈複雜人臉規則，所以運算量也隨之提高，不過當層數愈多時，人臉偵測的準確性也會提高。以圖 2-3 來說明，在判斷的過程中以  $T$  來表示通過某一層特徵的檢測，若為  $F$  則表示偵測失敗，也就是此影像視窗不是屬於人臉。



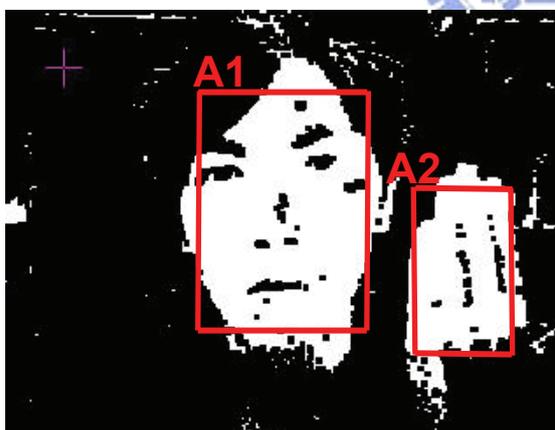
(a)膚色分割測試圖

(b)膚色分割結果



(c)閉合運算

(d)第一次投影



(e)第二次投影

圖 2-2 膚色分割及投影

透過色彩投影及臉部長寬比初步的篩選之後，我們將可能為人臉區域送入串聯法判斷此區域是否為正面的人臉。我們用以下幾點規則做為判斷條件：

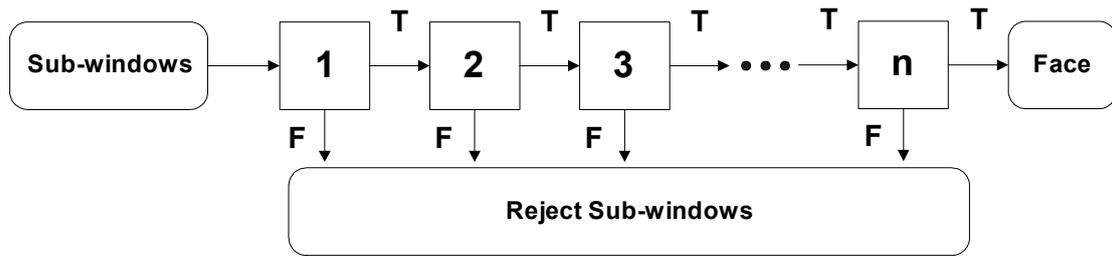


圖 2-3 人臉偵測專注式串聯法

- 1) 人臉投影之長寬比為 1 到 2 倍。
- 2) 人臉上半部兩眼區域與兩眉之間之灰階值總和會小於下半部灰階值的總和。
- 3) 眼睛區域的灰階值總和會小於眉毛中間灰階值的總和。
- 4) 在兩側臉頰相鄰的上下區域其灰階值總和會小於某閾值。

利用串聯的方法來判斷輸入的影像視窗是否為正面的人臉，因為結合了膚色的條件，所以在人臉偵測時不必進行全域的掃描，如此可以減少運算量以達到時偵測人臉的效果。



### 2.1.3 人臉偵測結果

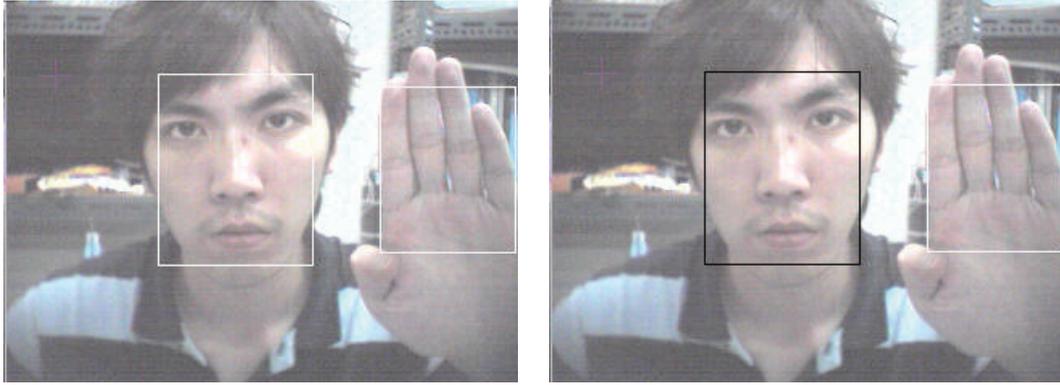
圖 2-4 為正面人臉偵測之結果。圖 2-4(a)為原始之測試影像，利用色彩分割將屬於膚色的範圍分割出來且作閉合運算，圖 2-4 (b)即為運算之後的結果。圖 2-4 (c)為利用膚色投影之後所框選之區域，圖 2-4 (d)為加入人臉特徵判斷之後得到的結果，黑色框框表示找到的人臉位置。



(a) 原始測試影像



(b) 膚色分割



(c) 膚色投影

(d) 加入特徵判斷之處理結果

圖 2-4 正面人臉偵測實驗結果

## 2.2 人臉影像前置處理

當輸入人臉影像進入本系統時，由於經過人臉偵測後，其所框選到的影像大小並不一定完全一樣。為了能夠準確的擷取到人臉影像上的特徵，在影像前置處理時，我們必須將人臉偵測所框選到的影像做正規化的動作，讓每張測試影像的維度大小能夠一樣，在本論文中，其正規化後的影像維度大小為 120x100，而一般常用的影像大小正規化方法為雙線性內插法，故我們也採用此方法來做前置處理。

影像 2 維線性的內插法是利用鄰近已知的影像值  $f(x, y)$  來估測出未知的影像值  $f(x', y')$ ，圖 2-5 上假設白點為已知的影像值，我們可以利用下列式子來估測出未知的黑色點的影像值，其表示式如下：

$$f(x', y') = (1 - \lambda)[(1 - u)f(x, y) + uf(x, y + 1)] + \lambda[(1 - u)f(x + 1, y) + uf(x + 1, y + 1)] \quad (2-3)$$

其中  $\lambda = \frac{x' - x_1}{x_2 - x_1}$  和  $u = \frac{y' - y_1}{y_2 - y_1}$  分別為 x 方向和 y 方向的放大因子

在環境光線部分，因不同的照度對影像所造成的變化是整體性的，也就是

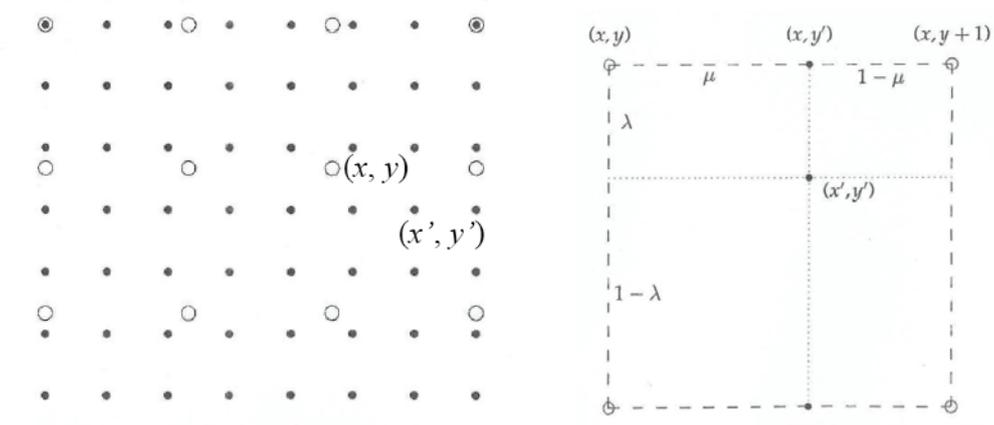


圖 2-5 影像內插

說，當光線變暗，整體的灰階值會偏暗，反之亦然。所以我們利用此特性，對每張擷取到的臉部影像之灰階值進行調整，使其每張影像的平均值為 128。

假設我們有一大小為  $N$  個像素的影像  $I$ ， $I_i$  表示影像中的第  $i$  個像素的灰階值，此影像  $I$  之平均值  $I_{ave}$  可由式(2-4)求出。再利用式(2-5)對影像  $I$  中的所有像素進行調整，以得到照度補償的臉部影像。

$$I_{ave} = \frac{1}{N} \sum_{i=1}^N I_i \quad (2-4)$$

$$I_i = (128 - I_{ave}) + I_i \quad (2-5)$$

## 2.3 人臉形狀模型

人臉形狀模型的主要作法是要做特徵點(landmark)的設置，一般特徵點的取選通常是邊界(boundary)、T 型交界處('T'junction)以及特徵點與特徵點間，如圖 2-6 所示，而點選在特徵點與特徵點間的原因是為了使之後做影像形變時可讓形變後的影像較為平滑(Smoothness)。

### 2.3.1 特徵點的設置

在特徵點的設置部份，本論文的作法是在每張要建模型的人臉影像上設置了

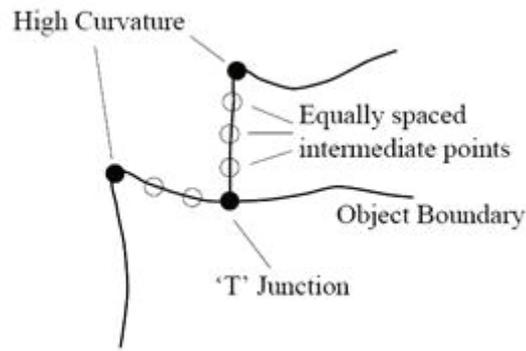


圖 2-6 特徵點的選取[25]

52 個特徵點，如圖 2-7 所示，其特徵點的位置大致上均分佈在嘴巴(8 個點)、鼻子(7 個點)、雙眼(16 個點)、雙眉(8 個點)以及臉部上的輪廓(13 個點)，至於在表情變化或是非正面人臉的設置，其設置的也是與在正面時設置的位置是一樣的，如圖 2-8 所示，就數學式表示，其定義為

$$s = (x_1, y_1, x_2, y_2, \dots, x_v, y_v) \quad (2-6)$$

其中  $(x_i, y_i)$  為每個特徵點的位置。

### 2.3.2 形狀對齊

在將每一張要建模型的人臉的特徵點均設置好後，我們需要將所有標出



圖 2-7 特徵點位置[26]



(a)表情變化之人臉

(b)非正面之人臉

圖 2-8 其他姿態人臉之特徵點位置

的特徵點做形狀對齊，也就是要校正所有的訓練集合(Aligning the training set)。每個形狀的大小、位置與旋轉角度並不一致，為了比較位於不同形狀上相對應的特徵點，則必須透過縮放、旋轉與位移的動作，使得所有形狀能盡量對齊。此處的對齊是對每個形狀向量的大小做正規化，並讓所有形狀的重心相對應，最後調整所有形狀的旋轉角度，而建立起形狀模型。

本論文主要是採用文獻[25]提出的Procrustes Analysis方法來將訓練集合中所有的形狀做對齊，在本例中我們取25張影像做特徵點的對齊，其步驟如下：

1. 選擇第一個形狀作為欲校正形狀的估測值，其定義為  $s_0$
2. 透過旋轉、縮放與位移，讓所有形狀與  $s_0$  對齊。
3. 從所有已對齊的形狀中，再重新估算  $s_0$
4. 若與前一次對齊之誤差值已無明顯化時，則表示已收斂，否則回到步驟2.

此處  $s_0$  的定義是所有形狀中特徵點的平均位置。判斷以上過程是否收斂的方法是當  $s_0$  在某一回合中沒有顯著的變化時，便表示結果已收斂。

令  $x$  為一組  $2v$  維的向量，描述任一形狀的  $v$  個特徵點，如式(2-6)所表示，而  $t = (a, b, c, d, t_x, t_y)$  中的  $a, b, c, d$  代表的是縮放與旋轉的倍率，而  $t_x, t_y$  則代表的是座標位移量，所以對於每個特徵點  $(x_i, y_i)$  的轉換，其計算方式如下：

$$T_i(x) = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x_i \\ y_i \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \quad (2-7)$$

若要得知  $s_0$  在迭代過程中是否已經沒有顯著變化時，則須先找出 Procrustes distance，其計算方式如下：

$$P_d^2 = \sum_{j=1}^v [(x_{j1} - x_{j2})^2 + (y_{j1} - y_{j2})^2] \quad (2-8)$$

可用此方式來判斷此一回合的  $s_0$  與上一回合的變化是否顯著，若不顯著則代表結果已收斂。圖 2-9 為 25 筆未經 Procrustes Analysis 校正過的形狀座標，而圖 2-10 為經此方法所計算出來的收斂結果。

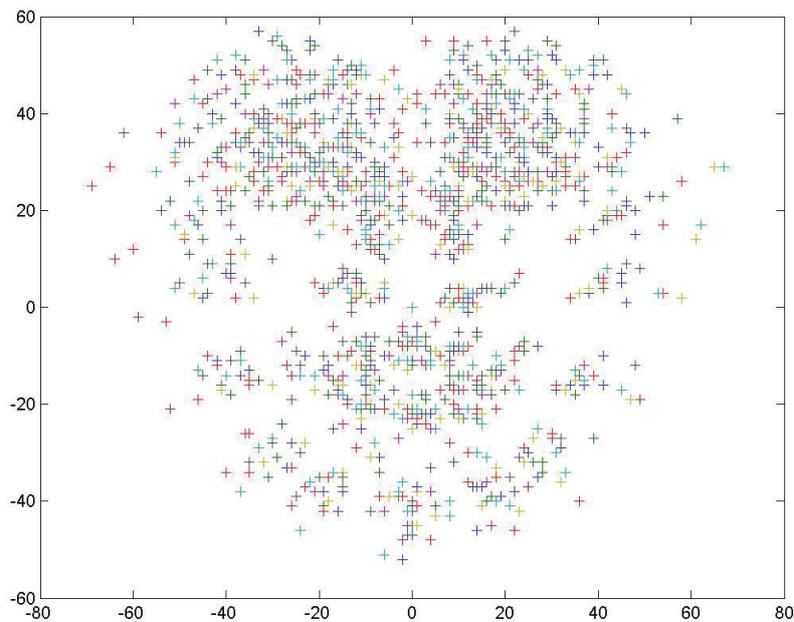


圖 2-9 未校正的形狀座標

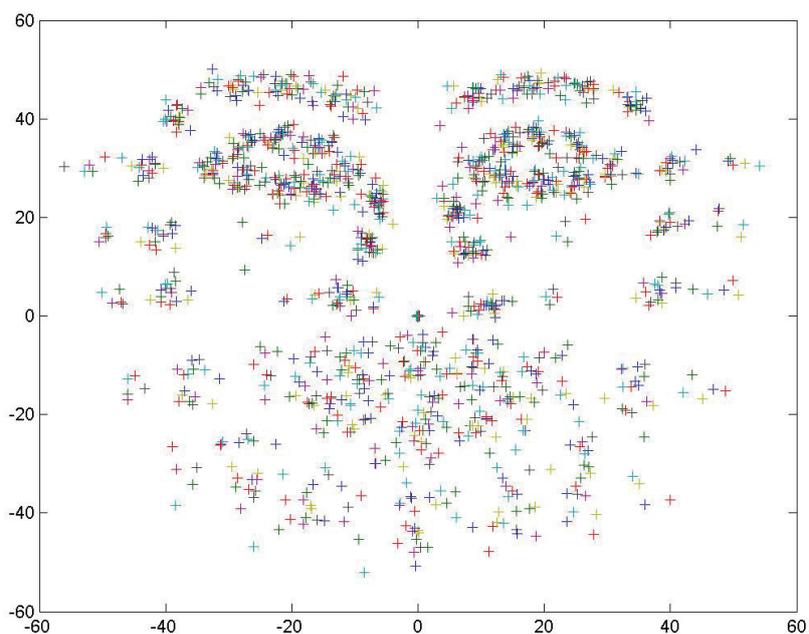


圖 2-10 經 Procrustes Analysis 校正過後的形狀座標

### 2.3.3 模擬臉部形狀的變化情形

在圖2-10中，我們可稱此圖為Point Distribution Model(PDM)[27]，PDM嘗試著去模擬訓練集合中每個特徵點座標的變化情形。而必須注意的是，特徵點並非各自獨立的移動，他們的位置是相關連的，也就是說每一個形狀僅有一組屬於自己的轉換參數。

當訓練集合內的所有形狀與  $s_0$  對齊後，每個形狀皆可被表示成  $2v$  維空間上的一個點，如式(2-6)所表示。因此，對於包含  $n$  個形狀的訓練集合，在  $2v$  維空間上便會形成  $k$  個點，假設此  $k$  個點群聚在空間上的某個範圍內，稱作 Allowable Shape Domain(ASD)[27]。在此範圍裡的每個  $2v$  維的點都代表一組特徵點，且這些特徵點所組成的形狀都會與訓練集合內的形狀相似。

我們透過 Principal Component Analysis(PCA) 來降低  $2v$  維資料的維度 [25]，並近似 ASD 的形狀(即  $2v$  維空間上  $n$  個點的分布情形)。利用前面所描述

的方法，我們可求得k個已對齊形狀的mean shape  $s_0$  (即ASD的中心點)。對2v維空間上的資料作PCA，則可得到ASD的主軸。每一條軸代表所有資料的一種變化模式(Mode of variation)，當人臉形狀改變時，則所有的特徵點會傾向於一起移動的方式在跑。

訓練集中k個臉部形狀的covariance matrix  $C$  定義如下所表示：

$$C = \frac{1}{k} \sum_{i=1}^k (x_i - s_0)(x_i - s_0)^T \quad (2-9)$$

求出此矩陣以後，接著就找出covariance matrix的特徵值與對應的特徵向量，其中特徵向量均為單範正交的形式。這些特徵向量代表了此ASD的主軸，而對應於最大之特徵值的特徵向量則是代表最長軸。

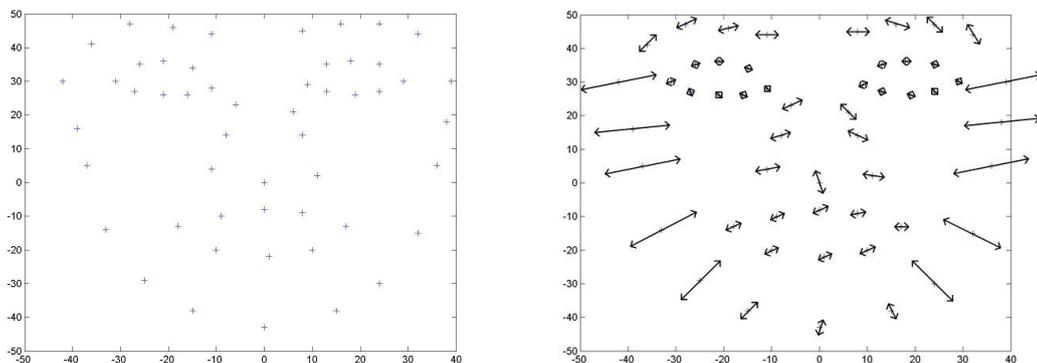
任何在Allowable Shape Domain中的點(即所有真實或者是可被接受的人臉形狀)都可利用mean shape加上特徵向量的線性組合來求得。因此，我們可以藉由  $s_0$  以及計算出來的特徵向量來近似訓練集中任何一個人臉形狀  $s$ ：

$$s = s_0 + \sum_{i=1}^n p_i s_i \quad (2-10)$$

其中  $s_i = (s_{i1} \ s_{i2} \ \dots \ s_{in})^T$  為前n個特徵向量所組成的矩陣，

$p_i = (p_{i1} \ p_{i2} \ \dots \ p_{in})$  為由權重值所組成的向量，稱作shape parameters。

圖2-11所示為mean shape  $s_0$  與其他特徵向量所組成的形狀座標，之後則必須用這些資訊來產生各種不同的形狀座標。



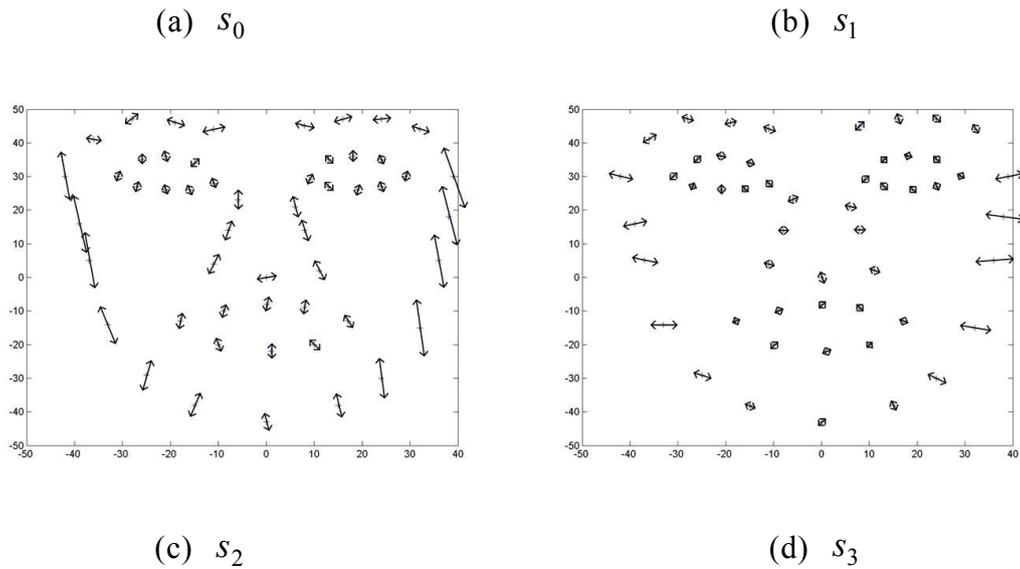


圖2-11 mean shape  $s_0$  與其他特徵向量所組成的形狀座標

## 2.4 紋理模型

一個完整的臉部模型裡除了須考慮人臉形狀以外，還必須涵蓋像素本身的資訊。因為形狀只包含了特徵點間的位置關係，並沒有提供我們像素上的資訊，例如灰階值的大小。

本節將說明如何建立臉部的紋理模型(Texture model)，我們先對紋理下定義，為了讓所有的臉部形狀的取樣數目相等，就利用影像形變(Image warping)將所有臉部形狀形變成 mean face shape，再取得像素值，最後利用 PCA 以說明紋理的變化情形，藉此建立一個臉部的紋理模型。

### 2.4.1 物體的紋理

在形狀模型裡，我們以一組特徵點座標所組成的向量來代表每個形狀，而在紋理模型中，我們將用一組向量來表示每張臉部影像的紋理：

$$g = (g_1, g_2, \dots, g_r) \quad (2-11)$$

其中  $g_i$  表示影像的像素值，而  $r$  代表臉部表面灰階值取樣的總數。

## 2.4.2 Piece-wise Affine Warping

由於臉部形狀的大小不一，其所包含的像素個數也不盡相同，為了讓所有臉部形狀之灰階值取樣的數目相等，則必須先將每個臉部形狀形變成 mean face shape，使兩者的特徵點完全吻合後，才能進行取樣的動作。

最簡單的影像形變方式是假設影像在局部區域內是線性的，為了滿足這個假設，我們利用 Delaunay triangulation [28]將所有點組成的 convex hull 分割成一組三角形的集合。

Delaunay triangulation 是利用許多三角形所組成的網狀(Mesh)來將一組不規則的點連結起來，其中，每個三角形都必須滿足 Delaunay property，即任一點皆不會落在任一個三角形的外接圓內，這表示對任一個三角形來說，其三個頂點都會剛好落在三角形的外接圓上，如圖 2-12 所示。因此也可以將人臉的形狀座標利用 Delaunay 來組成許多的三角形，如圖 2-13 所示。

透過 Affine transformation，像是縮放、位移與偏斜(Skewing)，可將影像  $I$  的點集合對應到影像  $I'$  的點集合。利用此方法我們便可將訓練集中的每個臉部形狀  $s$  都形變成 Mean shape  $s_0$ ，並且進行灰階值的取樣。

假設  $(x_i, y_i)^T$ 、 $(x_j, y_j)^T$  與  $(x_k, y_k)^T$  分別代表影像  $s$  中某一個三角形的三個頂點，如圖 2-14 所示，那麼三角形內部的任一個像素皆可表示成：

$$X = (x, y)^T = (x_i, y_i)^T + \alpha [(x_j, y_j)^T - (x_i, y_i)^T] + \beta [(x_k, y_k)^T - (x_i, y_i)^T] \quad (2-12)$$

其中  $0 \leq \alpha, \beta \leq 1$ ，而  $\alpha$  與  $\beta$  的計算方式如下：

$$\alpha = \frac{(x - x_i)(y_k - y_i) - (y - y_i)(x_k - x_i)}{(x_j - x_i)(y_k - y_i) - (y_j - y_i)(x_k - x_i)} \quad (2-13)$$

$$\beta = \frac{(y - y_i)(x_j - x_i) - (x - x_i)(y_j - y_i)}{(x_j - x_i)(y_k - y_i) - (y_j - y_i)(x_k - x_i)} \quad (2-14)$$

根據 Affine transform， $X$  會對應到影像  $s_0$  中的像素  $X'$ ，因此透過  $\alpha$  與  $\beta$  便

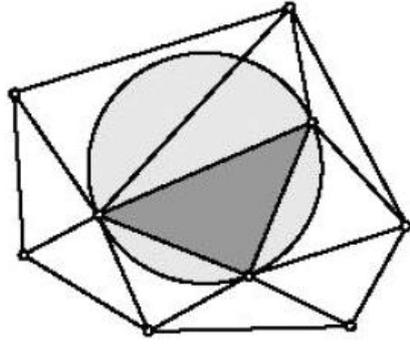


圖 2-12 滿足 Delaunay property 之三角形的外接圓[28]

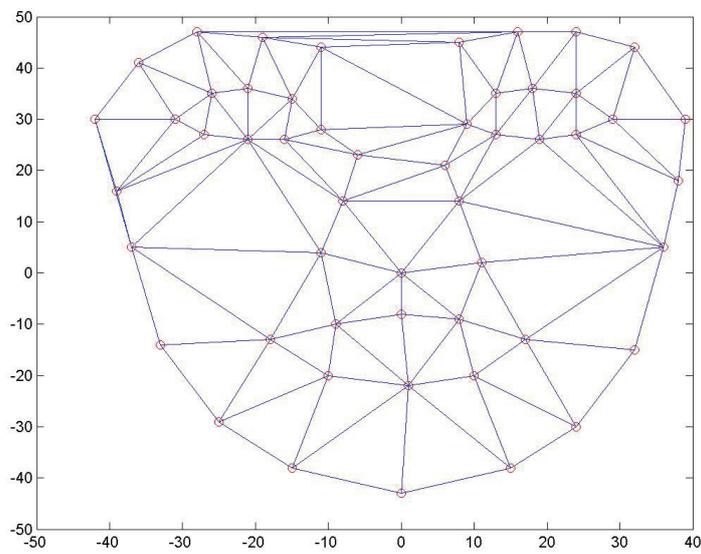


圖 2-13 人臉形狀座標經 Delaunay 後的結果

可決定  $X'$  在影像  $s_0$  中所對應到的三角形內的相對位置：

$$W(x; p) = (x'_i, y'_i)^T + \alpha[(x'_j, y'_j)^T - (x'_i, y'_i)^T] + \beta[(x'_k, y'_k)^T - (x'_i, y'_i)^T] \quad (2-15)$$

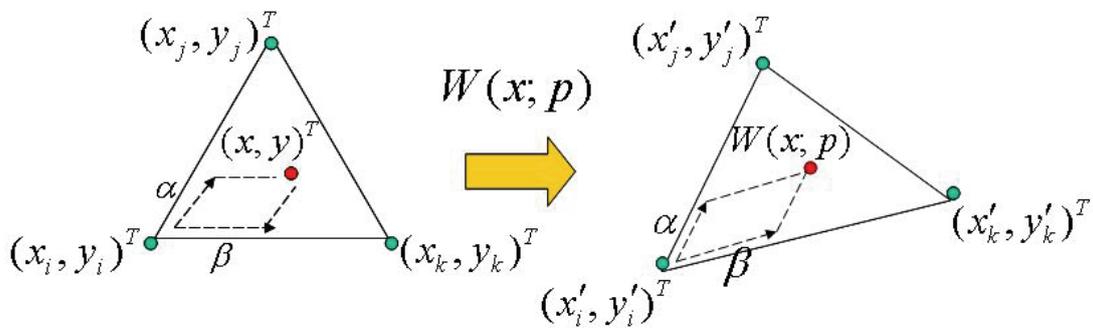


圖 2-14 三角形影像形變法

### 2.4.3 像素內插

在進行灰階值取樣時，由於利用上述之 Piece-wise affine warping 所求的新位置並不會剛好是整數，因此我們採用雙線性內插法來解決此問題。

雙線性內插法在 2.2 節已做了介紹，請參照 2.2 節的雙線性內插法。

### 2.4.4 模擬臉部紋理的變化情形

利用影像形變取得所有臉部形狀的像素值後，為了要去除掉像素間的相關性，我們利用 PCA 將訓練集中所有紋理向量的維度降低，且仍然足夠解釋絕大部分的紋理變化，以便去除多為資料中重複且多餘的關係。

訓練集中  $t$  組臉部紋理的 Covariance matrix  $G$  定義如下：

$$G = \frac{1}{t} \sum_{i=1}^t (g_i - A_0)(g_i - A_0)^T \quad (2-16)$$

其中  $g_i - A_0$  為第  $i$  組臉部紋理與 Mean face texture 之間的差異，而  $t$  組臉部紋理的 Mean face texture 為  $A_0 = \frac{1}{t} \sum_{i=1}^t g_i$ 。求出此矩陣以後，接著就找出 Covariance matrix 的特徵值與對應的特徵向量，其中特徵向量均為單範正交的形式。因此，我們可寫出一線性組合的形式：

$$A(x) = A_0 + \sum_{i=1}^m \lambda_i A_i \quad \forall x \in s_0 \quad (2-17)$$

其中  $A_i = (A_{i1} \ A_{i2} \ \dots \ A_{im})^T$  為前  $m$  個特徵向量所組成的矩陣，

$\lambda_i = (\lambda_{i1} \ \lambda_{i2} \ \dots \ \lambda_{im})$  為由權重值所組成的向量，稱作 Texture parameters。

圖 2-15 所示為 Mean texture  $A_0$  與其他特徵向量所組成的紋理資訊，之後我們必須用此資訊來產生各種不同的人臉。

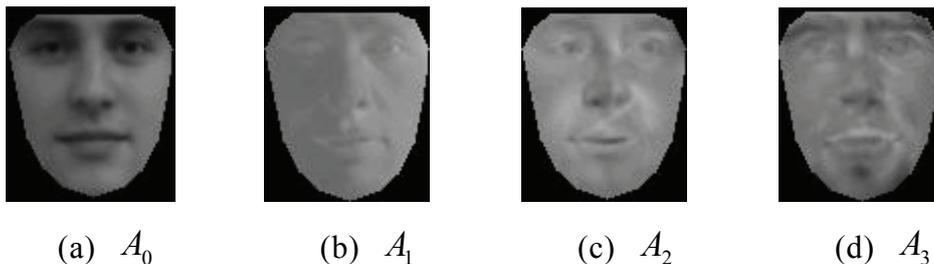


圖 2-15 mean texture  $A_0$  與其他特徵向量所組成的紋理資訊

## 2.5 可形變樣板追蹤演算法

藉由前面所描述的方法得出形狀模型與紋理模型後，我們即可藉由調整模型的參數而形變其他類型的人臉，定義為  $W(x; p)$ ，其代表的是 Mean shape  $s_0$  經過  $p$  的參數調整而得出的人臉模型，也就是座標轉換，舉例如圖 2-16 所示。

由上述可知，當測試人臉影像進來時，我們該如何調整參數以致於可將建好的模型追蹤到測試影像的人臉，此為接下來所要討論的主軸。追蹤的方法目前可分成 Combined AAM [25,29]與 Independent AAM [30]兩種，就 Combined AAM 來說，即是將形狀模型的調整參數與紋理模型的調整參數也一樣是藉由 PCA 的方法將兩組參數結合起來，形成一組可同時改變形狀與紋理的參數  $c = (c_1, c_2, \dots, c_l)^T$ ，所以形狀參數模型與紋理參數模型也會做改變：

$$s = s_0 + \sum_{i=1}^l c_i s_i \quad (2-18)$$

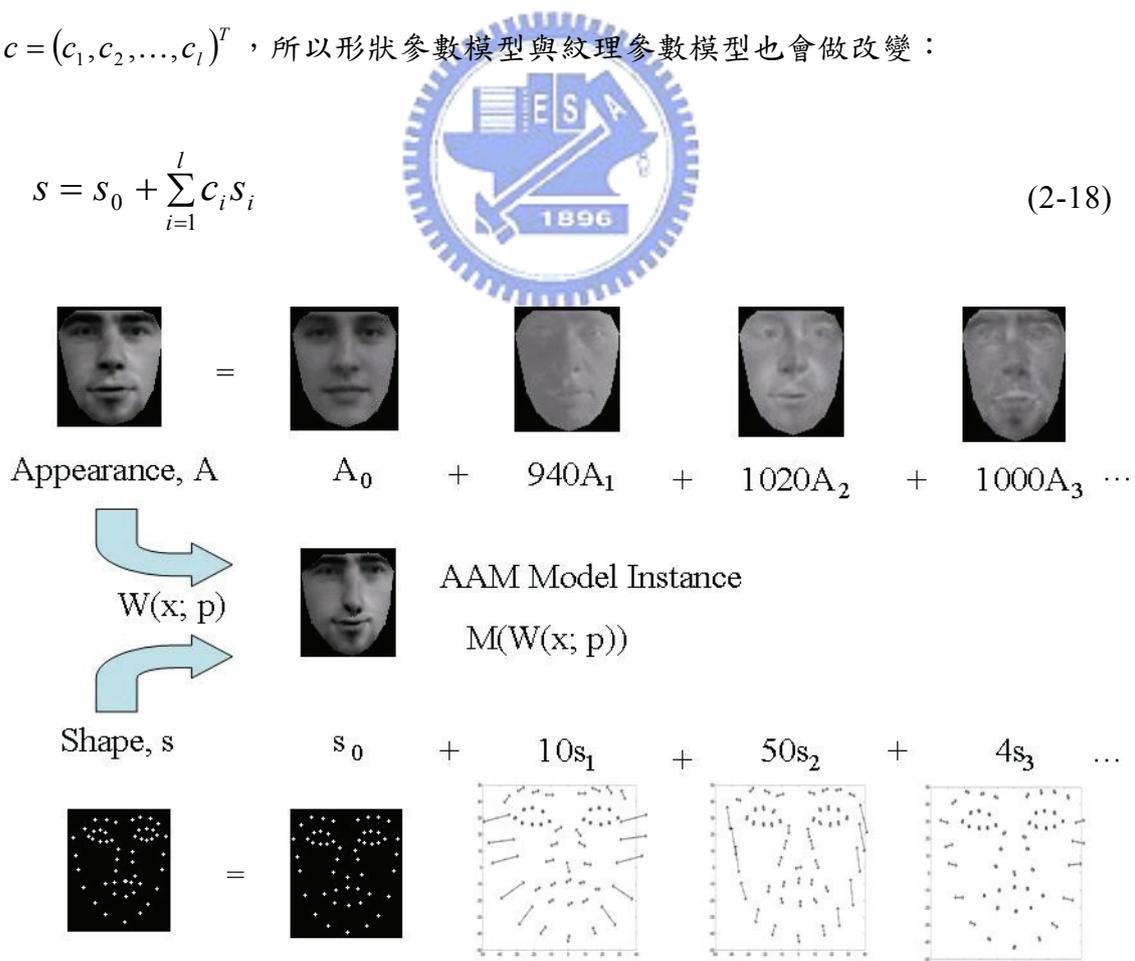


圖 2-16 調整形狀與紋理參數形變得到的其他人臉

$$A(x) = A_0(x) + \sum_{i=1}^l c_i A_i(x) \quad (2-19)$$

在結合原先的兩組參數時，是設定  $c = (p_1, p_2, \dots, p_n, \lambda_1, \lambda_2, \dots, \lambda_m)^T$ ，接著利用 PCA 的方式將維度給化簡，但由於為了保持演算法的正確性，其化簡後的維度不可小於  $n$  與  $m$  的最大值，即  $l \geq \max(n, m)$ ；而在 Independent AAM 中，其主要的調整參數是以形狀參數為主，而紋理參數可在形狀參數經迭代完成後再做計算即可。因此在每次迭代過程的計算量上，Combined AAM 需要較多的計算量，而由於本論文所使用的系統為嵌入式的影像處理平台，需要考慮計算量上面的問題，所以在此我們選擇 Independent AAM 的迭代方法來做可形變樣板的追蹤 [30-32]，即 Lucas-Kanade 演算法。

### 2.5.1 Lucas-Kanade 影像校正演算法

Lucas-Kanade 演算法是以梯度為基礎的影像變動估測演算法，它具有簡單快速的優點，在許多研究與應用上廣泛的被使用，如影像形變校正、物體動態追蹤與立體視覺的計算等。我們將參考[30]所介紹的專屬於 AAM 所用的追蹤演算法，且針對此演算法再做許多進一步的改良，來達成準確的效果。

假設有一樣版影像(Template image)，即  $A_0(x)$ ，代表未形變前的目標影像，另外有一影像  $I$  代表輸入的影像亦為形變後的影像，其中  $I$  與  $A_0$  存在一形變關係使得  $I(W(x; p)) = A_0(x)$ ，若將式(2-20)最小化，則可求得最佳形變參數  $p$ 。

$$\sum_x [A_0(x) - I(W(x; p))]^2 \quad (2-20)$$

由於影像的亮度分佈為非線性，使得形變參數  $p$  不可能直接求出。因此將  $p$  改為以下的形式，利用迭代的方式使  $p$  漸趨於最佳值。

$$\sum_x [A_0(x) - I(W(x; p + \Delta p))]^2 \quad (2-21)$$

$$p \leftarrow p + \Delta p \quad (2-22)$$

在最佳化計算的過程中， $\Delta p$  的估測值是以目前的  $p$  值為依據，因此將式(2-21)的  $I(W(x; p + \Delta p))$  對  $\Delta p \approx 0$  做泰勒展開：

$$I(W(x; p + \Delta p)) = I(W(x; p + 0)) + \nabla I \frac{\partial W(x; p)}{\partial p} \Delta p + \text{High Order Term} \quad (2-23)$$

其中  $\nabla I(x)$  為  $I(x)$  的亮度梯度值， $\frac{\partial W}{\partial p}$  為  $W$  的 Jacobian， $\nabla I \frac{\partial W(x; p)}{\partial p}$  稱為

Steepest descent images。忽略上式的高次項，並將此式代入式(2-21)中來微分求極值：

$$\sum_x \left[ \nabla I \frac{\partial W(x; p)}{\partial p} \right]^T \left[ A_0(x) - I(W(x; p)) - \nabla I \frac{\partial W(x; p)}{\partial p} \Delta p \right] = 0$$

$$\sum_x \left[ \nabla I(x) \frac{\partial W(x; p)}{\partial p} \right]^T \left[ \nabla I(x) \frac{\partial W(x; p)}{\partial p} \right] \Delta p = \sum_x \left[ \nabla I(x) \frac{\partial W(x; p)}{\partial p} \right]^T [A_0(x) - I(W(x; p))]$$

因此可求出  $\Delta p$ ：

$$\Delta p = H^{-1} \sum_x \left[ \nabla I \frac{\partial W(x; p)}{\partial p} \right]^T [A_0(x) - I(W(x; p))] \quad (2-24)$$

其中  $H$  代表 Hessian matrix[30]：

$$H = \sum_x \left[ \nabla I \frac{\partial W(x; p)}{\partial p} \right]^T \left[ \nabla I \frac{\partial W(x; p)}{\partial p} \right] \quad (2-25)$$

所以 Lucas-Kanade 演算法的迭代過程可分九個步驟，且流程圖如圖 2-17 所示：

- (1) 藉由參數  $p$  來形變輸入影像  $I$  去計算  $I(W(x; p))$

- (2) 計算誤差影像  $A_0(x) - I(W(x; p))$
- (3) 藉由參數  $p$  來形變梯度影像  $\nabla I$
- (4) 藉由參數  $p$  來計算 Jacobian matrix
- (5) 計算  $\nabla I \frac{\partial W(x; p)}{\partial p}$
- (6) 由式(2-23)來計算 Hessian matrix
- (7) 計算  $\sum_x \left[ \nabla I \frac{\partial W(x; p)}{\partial p} \right]^T [A_0(x) - I(W(x; p))]$
- (8) 由式(2-24)來計算  $\Delta p$
- (9) 更新參數  $p \leftarrow p + \Delta p$

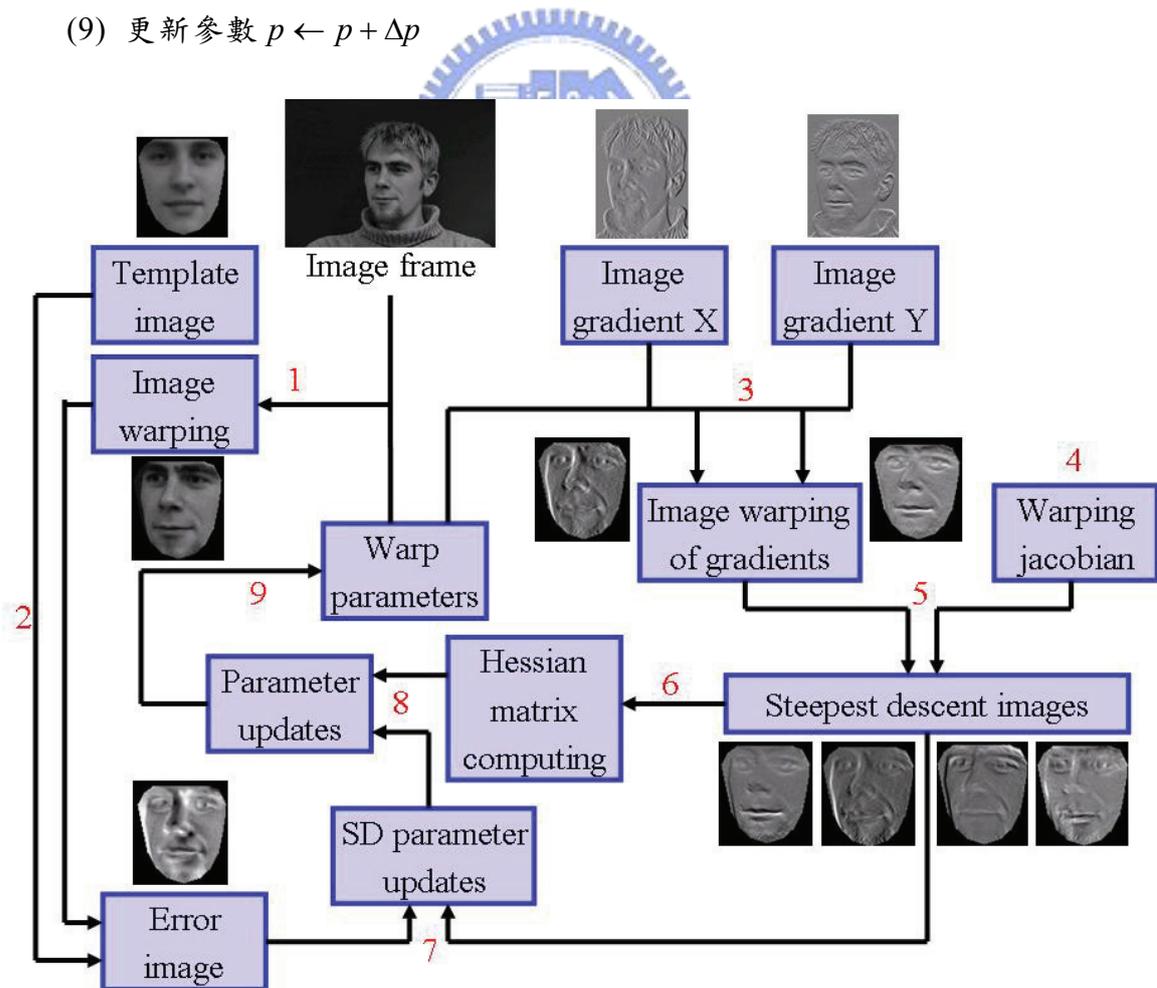


圖 2-17 Lucas-Kanade 演算法迭代流程

不過就此迭代流程的計算量來看，由於步驟(3)與步驟(4)都是與形變參數  $p$  相關的，因此在每次的迭代過程中都還必須再計算一次梯度上的形變以及 Jacobian，以致於後面的 Hessian matrix 也必須重新再計算一次，所以在每次迭代的計算量上面需耗費許多時間，因此必須對此演算法做計算量上面改良[30,32](Inverse Compositional Algorithm, IC)。

Inverse Compositional Algorithm 的重點在於避免讓所有要計算的過程均與形變參數  $p$  相依，因此式(2-21)需改成：

$$\sum_x \left[ I(W(x; p)) - A_0(W(x; \Delta p)) \right]^2 \quad (2-26)$$

$$W(x; p) \leftarrow W(W(x; \Delta p)^{-1}; p) \quad (2-27)$$

同上述方法，我們將  $A_0(W(x; \Delta p))$  做泰勒展開可得：

$$A_0(W(x; \Delta p)) = A_0(W(x; 0)) + \nabla A_0(x) \frac{\partial W}{\partial p} \Delta p + \text{High order term} \quad (2-28)$$

將上式代入式(2-26)可得：

$$\sum_x \left[ I(W(x; p)) - A_0(W(x; 0)) - \nabla A_0 \frac{\partial W}{\partial p} \Delta p \right]^2 \quad (2-29)$$

再將上式對  $\Delta p$  微分求極值可得：

$$\Delta p = H^{-1} \sum_x \left[ \nabla A_0 \frac{\partial W}{\partial p} \right]^T \left[ I(W(x; p)) - A_0(x) \right] \quad (2-30)$$

其中  $H$  同之前所描述，代表為 Hessian matrix，表示為：

$$H = \sum_x \left[ \nabla A_0 \frac{\partial W}{\partial p} \right]^T \left[ \nabla A_0 \frac{\partial W}{\partial p} \right] \quad (2-31)$$

由此兩種演算法比較可看出在傳統的 Lucas-Kanade 演算法是計算輸入影像的梯度值，接著再從此計算出的梯度值來找出 Warping jacobian，每一次迭代就要計算一次此部份；但在 Inverse Compositional Algorithm 的改良演算法可看出梯度值是從樣板影像去計算的，而樣板影像是已知的，所以也就是說在後面的 Warping jacobian 的部份也是已知，故若用此種演算法來做迭代時，只要在迭代前先將樣板影像的梯度值以及 Warping jacobian 計算出來，之後的迭代過程中就不需要再去計算此部份，因此在計算量上，經過改良的演算法確實能加快迭代的時間，且也不會因為速度變快而導致成功率降低，故本系統選擇此改良演算法來做迭代的部份，其迭代的步驟與先前介紹的 Lucas-Kanade 大同小異，只是將其中幾個步驟當成前處理，剩下的步驟即為迭代的步驟，流程圖如圖 2-18 所示：

前處理：



- I. 找出樣板影像  $A_0(x)$  的梯度值
- II. 找出 Jacobian  $\frac{\partial W(x;0)}{\partial p}$
- III. 計算  $\nabla_{A_0} \frac{\partial W}{\partial p}$
- IV. 藉由式(2-31)來計算出 Hessian matrix

迭代直到收斂：

- (1) 藉由參數  $p$  來形變輸入影像  $I$  去計算  $I(W(x; p))$
- (2) 計算誤差影像  $I(W(x; p)) - A_0(x)$

$$(3) \text{ 計算 } \sum_x \left[ \nabla I \frac{\partial W(x; p)}{\partial p} \right]^T [I(W(x; p)) - A_0(x)]$$

(4) 由式(2-30)來計算  $\Delta p$

(5) 更新參數  $W(x; p) \leftarrow W(x; p) \circ W(x; \Delta p)^{-1}$

由此兩演算法比較下來可看出在 Inverse Compositional 演算法中的迭代步驟少做了四項，所以若迭代越多次，相較之下就會比原本的演算法還要少計算越多次。

在前面已經描述計算影像形變的方法，而接下來則是介紹如何去計算 Warping jacobian  $\frac{\partial W}{\partial p}$  的部份[33]，假如  $W(x; p) = (W_x(x; p), W_y(x; p))^T$ ，則：

$$\frac{\partial W}{\partial p} = \begin{pmatrix} \frac{\partial W_x}{\partial p_1} & \frac{\partial W_x}{\partial p_2} & \dots & \frac{\partial W_x}{\partial p_n} \\ \frac{\partial W_y}{\partial p_1} & \frac{\partial W_y}{\partial p_2} & \dots & \frac{\partial W_y}{\partial p_n} \end{pmatrix} \quad (2-32)$$

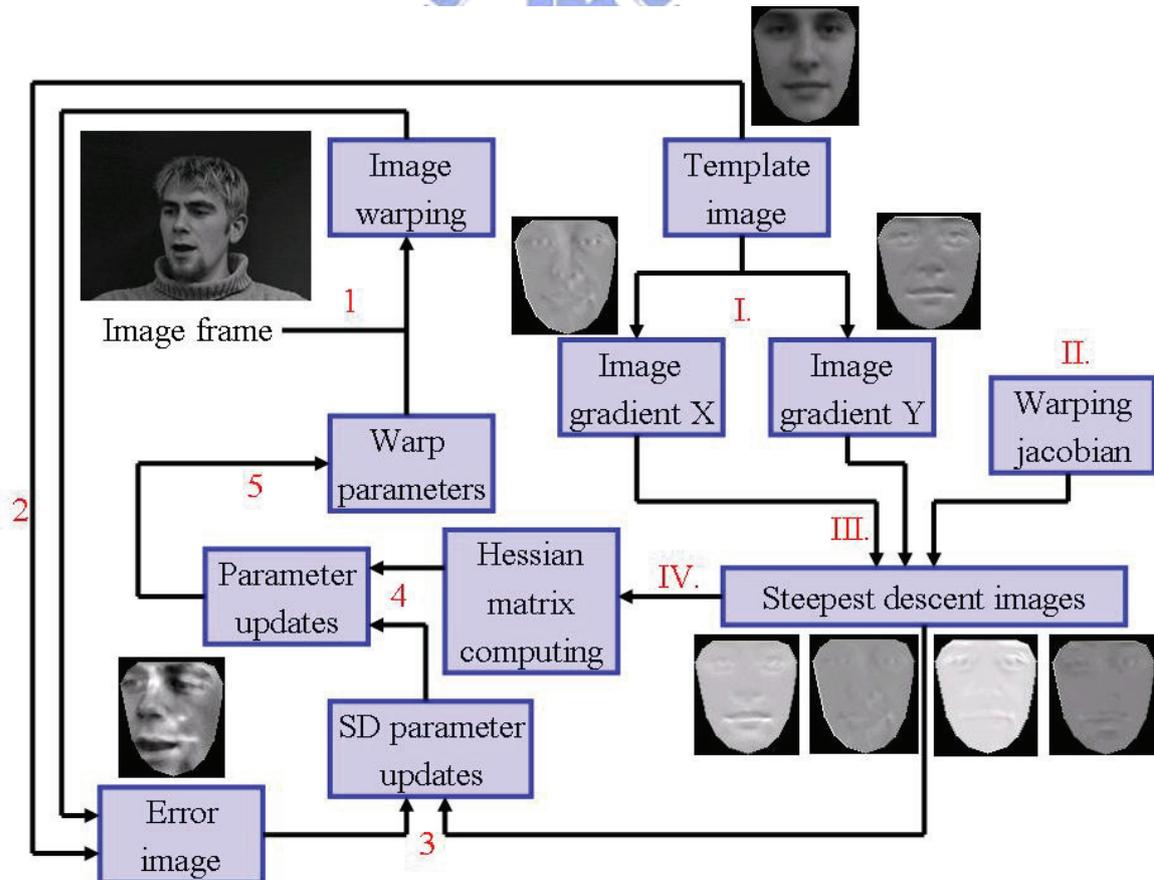


圖 2-18 Inverse Compositional 演算法迭代流程

就此矩陣的第一列來說， $W_x$ 代表的是形變點 $W(x,p)$ 在 $x$ 的座標，而 $p_m$ 則是代表形變參數，利用 Chain rule 可將 $\frac{\partial W_x}{\partial p_m}$ 表示為：

$$\frac{\partial W_x}{\partial p_m} = \sum_{t=1}^v \left[ \frac{\partial W_x}{\partial x_t} \frac{\partial x_t}{\partial p_m} + \frac{\partial W_y}{\partial y_t} \frac{\partial y_t}{\partial p_m} \right] = \sum_{t=1}^v \frac{\partial W_x}{\partial x_t} \frac{\partial x_t}{\partial p_m} \quad (2-33)$$

同理，利用 Chain rule 可將 $\frac{\partial W_y}{\partial p_m}$ 表示為：

$$\frac{\partial W_y}{\partial p_m} = \sum_{t=1}^v \left[ \frac{\partial W_x}{\partial x_t} \frac{\partial x_t}{\partial p_m} + \frac{\partial W_y}{\partial y_t} \frac{\partial y_t}{\partial p_m} \right] = \sum_{t=1}^v \frac{\partial W_y}{\partial y_t} \frac{\partial y_t}{\partial p_m} \quad (2-34)$$

由前面的式(2-9)可得知：

$$X = (x, y)^T = (1 - \alpha - \beta)(x_i^0, y_i^0)^T + \alpha(x_j^0, y_j^0)^T + \beta(x_k^0, y_k^0)^T \quad (2-35)$$

其中 $(x_i^0, y_i^0)^T$ 、 $(x_j^0, y_j^0)^T$ 與 $(x_k^0, y_k^0)^T$ 為三角形的三個頂點且是形變前的影像。

同理，若形變到另一影像時，可表示成：

$$W(x; p) = (1 - \alpha - \beta)(x_i, y_i)^T + \alpha(x_j, y_j)^T + \beta(x_k, y_k)^T \quad (2-36)$$

其中 $(x_i, y_i)$ 、 $(x_j, y_j)$ 與 $(x_k, y_k)$ 代表的是三角形的形變後影像。因此分別對三

角形的三個點做微分可計算出：

$$\begin{aligned} \frac{\partial W_x}{\partial x_i} &= \frac{\partial W_y}{\partial y_i} = 1 - \alpha - \beta \\ \frac{\partial W_x}{\partial x_i} &= \frac{\partial W_y}{\partial y_i} = \alpha \\ \frac{\partial W_x}{\partial x_i} &= \frac{\partial W_y}{\partial y_i} = \beta \end{aligned} \quad (2-37)$$

由上式可得知，當  $z \neq i, j, k$ ， $\frac{\partial W_x}{\partial x_z} = \frac{\partial W_y}{\partial y_z} = 0$ ，其意義為每計算到一個特徵

點時，就可從此特徵點相連到的三角形去計算 Jacobian 在此特徵點所貢獻的值。因此由式(2-33)與式(2-34)可改寫為：

$$\begin{aligned}\frac{\partial W_x}{\partial p_m} &= \sum_{t=1}^v \frac{\partial W_x}{\partial x_t} \frac{\partial x_t}{\partial p_m} = (1 - \alpha - \beta) \frac{\partial x_i}{\partial p_m} + \alpha \frac{\partial x_j}{\partial p_m} + \beta \frac{\partial x_k}{\partial p_m} \\ \frac{\partial W_y}{\partial p_m} &= \sum_{t=1}^v \frac{\partial W_y}{\partial y_t} \frac{\partial y_t}{\partial p_m} = (1 - \alpha - \beta) \frac{\partial y_i}{\partial p_m} + \alpha \frac{\partial y_j}{\partial p_m} + \beta \frac{\partial y_k}{\partial p_m}\end{aligned}\quad (2-38)$$

在  $\frac{\partial x_t}{\partial p_m}$  與  $\frac{\partial y_t}{\partial p_m}$  的計算裡，可從式(2-9)得知：

$$(s_1 \quad s_2 \quad \dots \quad s_n) = \begin{pmatrix} s_1^{x_1} & s_2^{x_1} & \dots & s_n^{x_1} \\ s_1^{y_1} & s_2^{y_1} & \dots & s_n^{y_1} \\ \dots & \dots & \dots & \dots \\ s_1^{x_v} & s_2^{x_v} & \dots & s_n^{x_v} \\ s_1^{y_v} & s_2^{y_v} & \dots & s_n^{y_v} \end{pmatrix}\quad (2-39)$$

其中  $s_m^{x_t}$  與  $s_m^{y_t}$  代表的是第  $m$  個特徵向量的  $x_t$  與  $y_t$  的特徵點位置，所以若

將  $\frac{\partial x_t}{\partial p_m}$  與  $\frac{\partial y_t}{\partial p_m}$  也寫成上式的矩陣形式來表示時，則：

$$\begin{pmatrix} \frac{\partial x_1}{\partial p_1} & \frac{\partial x_1}{\partial p_2} & \dots & \frac{\partial x_1}{\partial p_n} \\ \frac{\partial y_1}{\partial p_1} & \frac{\partial y_1}{\partial p_2} & \dots & \frac{\partial y_1}{\partial p_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial x_v}{\partial p_1} & \frac{\partial x_v}{\partial p_2} & \dots & \frac{\partial x_v}{\partial p_n} \\ \frac{\partial y_v}{\partial p_1} & \frac{\partial y_v}{\partial p_2} & \dots & \frac{\partial y_v}{\partial p_n} \end{pmatrix} = \frac{\partial \begin{pmatrix} p_1 s_1^{x_1} + p_2 s_2^{x_1} + \dots + p_n s_n^{x_1} \\ p_1 s_1^{y_1} + p_2 s_2^{y_1} + \dots + p_n s_n^{y_1} \\ \dots \\ p_1 s_1^{x_v} + p_2 s_2^{x_v} + \dots + p_n s_n^{x_v} \\ p_1 s_1^{y_v} + p_2 s_2^{y_v} + \dots + p_n s_n^{y_v} \end{pmatrix}}{\partial p}$$

$$= \begin{pmatrix} s_1^{x_1} & s_2^{x_1} & \dots & s_n^{x_1} \\ s_1^{y_1} & s_2^{y_1} & \dots & s_n^{y_1} \\ \dots & \dots & \dots & \dots \\ s_1^{x_v} & s_2^{x_v} & \dots & s_n^{x_v} \\ s_1^{y_v} & s_2^{y_v} & \dots & s_n^{y_v} \end{pmatrix} \quad (2-40)$$

最後由式(2-39)與式(2-40)可得出 Warping jacobian 的算法：

$$\begin{aligned} \frac{\partial W_x}{\partial p_m} &= (1 - \alpha - \beta) s_m^{x_i} + \alpha s_m^{x_j} + \beta s_m^{x_k} \\ \frac{\partial W_y}{\partial p_m} &= (1 - \alpha - \beta) s_m^{y_i} + \alpha s_m^{y_j} + \beta s_m^{y_k} \end{aligned} \quad (2-41)$$

在更新參數的步驟中，我們需要藉由  $W(x; \Delta p)$  來計算出  $W(x; \Delta p)^{-1}$ ，因此將做泰勒展開可得：

$$W(x; \Delta p) = W(x; 0) + \frac{\partial W}{\partial p} \Delta p = x + \frac{\partial W}{\partial p} \Delta p + \text{High order term} \quad (2-42)$$

由於  $W(x; 0) = x$  是單位形變，因此可得：

$$W(x; \Delta p) \circ W(x; -\Delta p) = x - \frac{\partial W}{\partial p} \Delta p + \frac{\partial W}{\partial p} \Delta p = x + \text{High order term} \quad (2-43)$$

$$W(x; \Delta p)^{-1} = W(x; -\Delta p) \quad (2-44)$$

當輸入影像進入到系統做處理時，會先將 Mean shape  $s_0$  放入輸入影像中，再根據迭代演算法來做形狀參數的更新，故此更新的方式是從  $s_0$  的形狀來做參數的調整，所以在已知特徵點的情況下，由式(2-9)可得形狀向量的線性組合，同理，式(2-44)也可用線性組合的形式來表示：

$$\Delta s_0 = -\sum_{i=1}^n \Delta p_i s_i \quad (2-45)$$

其中  $\Delta s_0 = (\Delta x_1^0, \Delta y_1^0, \dots, \Delta x_v^0, \Delta y_v^0)^T$ 。為了要更新  $W(x; \Delta p)^{-1}$  到  $W(x; p)$  上，則必須去計算目前特徵點所在的位置  $\Delta s = (\Delta x_1, \Delta y_1, \dots, \Delta x_v, \Delta y_v)^T$ 。一旦知道這些特徵點的位置，就可根據式 (2-9) 來計算  $W(x; p) \circ W(x; \Delta p)^{-1}$  的新形狀參數：

$$p'_i = s_i \cdot (s + \Delta s - s_0) \quad (2-46)$$

其中  $p'_i$  是  $W(x; p) \circ W(x; \Delta p)^{-1}$  的第  $i$  個新的形變參數，此式是做內積的相乘，並以假設  $s_i$  為單範正交的形式。

如何從  $\Delta s_0$  來推算出  $\Delta s$ ，假設考慮人臉形狀中第  $i$  個特徵點，我們需要從 Mean shape  $s_0$  裡面的  $(\Delta x_i^0, \Delta y_i^0)^T$  去計算出目前的形狀  $s$  裡的  $(\Delta x_i, \Delta y_i)^T$ ，圖 2-19 是利用先前所敘述的利用 Delaunay triangulation 的方式做 Piece-wise affine warping 找出座標上的轉換，將  $s_0$  中座標點的  $(x_i^0, y_i^0)^T + (\Delta x_i^0, \Delta y_i^0)^T$  轉換到  $s$  中座標點的  $(x_i, y_i)^T + (\Delta x_i, \Delta y_i)^T$ ，不過在此會產生一個問題，也就是由於第  $i$  個特徵點所相鄰的三角形不只一個，所以在做形變時會不知要挑選哪一個鄰近的三角形，若是形變時所挑選的三角形不同，則  $(x_i, y_i)^T + (\Delta x_i, \Delta y_i)^T$  的結果也會不一樣。所以就一般來說，若遇到此種情況時，就會將與第  $i$  個特徵點相鄰的所有三角形均拿來執行座標轉換的動作，最後再把經由每個鄰近三角形所產生出來的座標結果來取平均，得到此點的新座標值，此方式可讓人臉形狀較為平順，比較符合真實的人臉形狀。

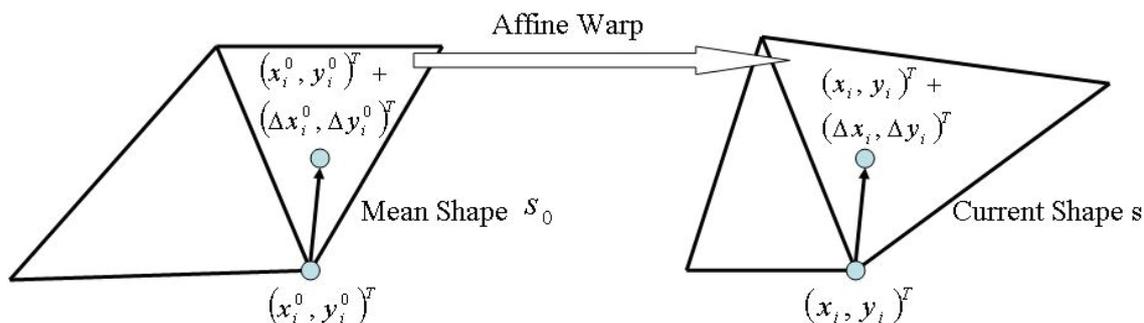


圖 2-19 對  $(x_i^0, y_i^0)^T + (\Delta x_i^0, \Delta y_i^0)^T$  計算座標轉換得  $(x_i, y_i)^T + (\Delta x_i, \Delta y_i)^T$

雖然 Inverse compositional 演算法具有計算量較少的優點，但是它只能解決輸入影像與固定樣板圖像間的匹配問題，由文獻[30]可知，AAM 可具有變動性的樣板存在，其可表示為：

$$\sum_{x \in S_0} \left[ A_0(x) + \sum_{i=1}^m \lambda_i A_i(x) - I(W(x; p)) \right]^2 = \left\| A_0(x) + \sum_{i=1}^m \lambda_i A_i(x) - I(W(x; p)) \right\|^2 \quad (2-47)$$

其中  $\|\cdot\|$  為 L2 norm，接著對上式加以轉化以滿足 Inverse compositional 演算法的定義。因此可轉化為求下式最小化的問題[34]：

$$\left\| A_0(x) - I(W(x; p)) \right\|_{\text{span}(A_i)^\perp}^2 + \left\| A_0(x) + \sum_{i=1}^m \lambda_i A_i(x) - I(W(x; p)) \right\|_{\text{span}(A_i)}^2 \quad (2-48)$$

其中  $\text{span}(A_i)$  表示由紋理的特徵向量  $A_i$  形成的線性子空間，而  $\text{span}(A_i)^\perp$  則是表示具有正交補集的線性子空間。由於式(2-48)的前半部份只與形狀參數  $p$  有關，而後半部份對於任何的形狀參數  $p$  其最小值均為 0，因此可以採用 Inverse compositional 演算法先求出滿足式(2-48)前半部份的形狀參數  $p$ ，且將  $p$  看作常數，進而求出滿足後半部份的紋理參數  $\lambda_i$ 。因為  $A_i(x)$  是相互正交的

關係，故紋理參數  $\lambda_i$  可表示為：

$$\lambda_i = \sum_{x \in S_0} A_i(x) \cdot [I(W(x; p)) - A_0(x)] \quad (2-49)$$

由於 Inverse compositional 演算法是在此部份是藉由  $\text{span}(A_i)^\perp$  的空間中來進行，因此  $\nabla A_0 \frac{\partial W}{\partial p}$  與 Hessian matrix 也都必須要投影至  $\text{span}(A_i)^\perp$  才能做迭代的

計算，其迭代流程如下：

前處理：

I. 找出樣板影像  $A_0(x)$  的梯度值

II. 找出 Jacobian  $\frac{\partial W(x;0)}{\partial p}$

III. 計算  $SD_j(x) = \nabla A_0 \frac{\partial W}{\partial p_j} - \sum_{i=1}^m \left[ \sum_{x \in S_0} A_i(x) \cdot \nabla A_0 \frac{\partial W}{\partial p_j} \right] A_i(x)$

IV. 計算 Hessian matrix  $= \sum_x SD^T(x)SD(x)$

迭代直到收斂：

(1) 藉由參數  $p$  來形變輸入影像  $I$  去計算  $I(W(x; p))$

(2) 計算誤差影像  $I(W(x; p)) - A_0(x)$

(3) 計算  $\sum_x SD^T(x)[I(W(x; p)) - A_0(x)]$

(4) 計算  $\Delta p = H^{-1} \sum_x SD^T [I(W(x; p)) - A_0(x)]$

(5) 更新參數  $W(x; p) \leftarrow W(x; p) \circ W(x; \Delta p)^{-1}$

收斂後：

(1) 藉由式(2-48)計算紋理參數  $\lambda$

## 2.5.2 Global Shape Normalizing Transform (GSNT)

在上一節所描述的傳統 AAM 演算法中，大部分都是用在正面的人臉，如圖 2-20 所示，若是應用在多姿態中，其容忍範圍相當有限，若是人臉偏轉多一些，則可能形狀特徵點就會迭代失敗，如圖 2-21 所示。



圖 2-20 傳統 AAM 演算法迭代出的正面人臉

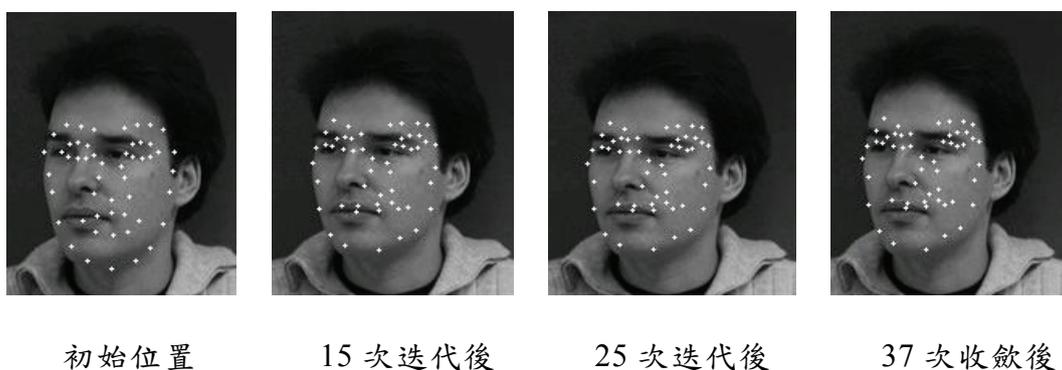


圖 2-21 傳統 AAM 演算法迭代出的具姿態變化之人臉

由於測試人臉影像的形狀  $s$  是定義在模型座標框架下的，是去除平移、縮放和旋轉的轉換之後的人臉形狀，而實際影像中的人臉往往包含了平移、縮放與旋轉的成份，因此必須對  $s$  進行 Global shape normalizing transform，來得到實際影像座標框架下的人臉形狀。假設平移位置為  $(t_x, t_y)$ ，旋轉角度為  $\theta$ ，縮放比例為  $k$ ，姿態參數向量  $q = (a, b, t_x, t_y)^T$ ，其中  $a = k \cos \theta - 1$ ， $b = k \sin \theta$ ，則對於單一點  $(x, y)$ ，其 Global transform 函數  $N(x; q)$  可表示為

$$N(x; q) = \begin{pmatrix} 1 + a & -b \\ b & 1 + a \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix} \quad (2-50)$$

因此最後得到的人臉形狀可以表示為  $N(W(x; p); q), \forall x \in s_0$ 。對於 Mean shape  $s_0$  中的任意一點  $x_0$ ，其在最後人臉形狀中的位置為  $N(W(x_0; p); q)$ ，而  $x_0$  所對應到的紋理值為  $A(x_0)$ ，在  $N(W(x_0; p); q)$  的紋理值也應為  $A(x_0)$ ，因此最後的 AAM 模型實例可表示為：

$$M(N(W(x; p); q)) = A(x), \forall x \in s_0 \quad (2-51)$$

其中  $M$  是一個與形狀  $N(W(x; p); q)$  大小相同的二維影像。

當給定輸入影像  $I(x)$  時，其目的也是為了要讓模型實例  $M(N(W(x; p); q))$  與  $I(x)$  匹配，希望兩者之差能最小化。對於 Mean shape  $s_0$  中的任意一點  $x_0$ ，其在輸入影像  $I$  中的位置為  $N(W(x_0; p); q)$ ，對應到的紋理值為  $A(x_0) = A_0(x_0) + \sum_{i=1}^m \lambda_i A_i(x_0)$ ， $N(W(x_0; p); q)$  所對應到的紋理值為  $I(N(W(x_0; p); q))$ ，因此 AAM 的匹配算法可表示為

$$\sum_{x \in s_0} \left[ A_0(x) + \sum_{i=1}^m \lambda_i A_i(x) - I(N(W(x; p); q)) \right]^2 \quad (2-52)$$

並讓此式最小化。

就 Global transform 函數  $N(x; q)$  而言，為了要使此函數與式(2-9)搭配在一起計算，因此也必須建立一組具有線性組合的係數 [30]。若以  $s_0 = (x_1^0, y_1^0, \dots, x_v^0, y_v^0)^T$  來表示，則可假設  $s_1^* = s_0 = (x_1^0, y_1^0, \dots, x_v^0, y_v^0)^T$ ， $s_2^* = (-y_1^0, x_1^0, \dots, -y_v^0, x_v^0)^T$ ， $s_3^* = (1, 0, \dots, 1, 0)^T$ ，與  $s_4^* = (0, 1, \dots, 0, 1)^T$ ，故可將  $N(x; q)$  表示成：

$$N(x; q) = s_0 + \sum_{i=1}^4 q_i s_i^* \quad (2-53)$$

其中  $q = (q_1, q_2, q_3, q_4)$ ，而個別的參數表示為  $a = q_1$ 、 $b = q_2$ 、 $t_x = q_3$  與  $t_y = q_4$ 。因此就可利用式(2-51)來計算 Piece-wise affine warping 以及 Warping jacobian，其計算的方法與利用式(2-9)時用的計算方法是一樣的[30]。

至於在形狀參數更新的合成方面，由於在上一節已經得知  $W(x; \Delta p)^{-1} = W(x; -\Delta p)$ ，因此若要產生  $N \circ W$  的形狀參數，則參數  $(\Delta q, \Delta p)$  可表示成：

$$N \circ W(x; \Delta q, \Delta p)^{-1} = N \circ W(x; -\Delta q, -\Delta p) \quad (2-54)$$

故合成形式為：

$$(N \circ W)(x; q, p) \circ (N \circ W)(x; \Delta q, \Delta p)^{-1} = (N \circ W)(x; q, p) \circ (N \circ W)(x; -\Delta q, -\Delta p) \quad (2-55)$$

至於如何找出  $(N \circ W)(x; -\Delta q, -\Delta p)$ ，其方法與上一節的作法類似，根據式(2-45)針對此部份可改寫為：

$$\Delta s_0 = -\sum_{i=1}^n \Delta p_i s_i - \sum_{j=1}^4 \Delta q_j s_j^* \quad (2-56)$$

其中  $\Delta s_0 = (\Delta x_1^0, \Delta y_1^0, \dots, \Delta x_v^0, \Delta y_v^0)^T$ 。若定義  $N \circ W(s_0; q, p) = s^+$ ，其表示為：

$$N \circ W(s_0; q, p) = N(s_0 + \sum_{i=1}^n p_i s_i; q) = s_0 + \sum_{i=1}^n p_i s_i + \sum_{j=1}^4 q_j s_j^* \quad (2-57)$$

為了要更新  $(N \circ W)(x; \Delta q, \Delta p)^{-1}$  到  $(N \circ W)(x; q, p)$  上，則必須去計算目前特徵點所在的位置  $\Delta s^+ = (\Delta x_1^+, \Delta y_1^+, \dots, \Delta x_v^+, \Delta y_v^+)^T$ 。若得知特徵點的位置，則可根據式 (2-9) 與式 (2-53) 來計算  $(N \circ W)(x; q, p) \circ (N \circ W)(x; \Delta q, \Delta p)^{-1}$  的新形狀參數：

$$q_i = s_i^* \cdot (s^+ - s_0) \quad (2-58)$$

$$p_i = s_i (N(s^+; q)^{-1} - s_0) \quad (2-59)$$

同樣的也考慮樣板的可變動性，則經過 Global shape normalizing transform 的迭代流程如下：

前處理：

I. 找出樣板影像  $A_0(x)$  的梯度值

II. 找出 Jacobian  $\frac{\partial W(x;0)}{\partial p}$  與  $\frac{\partial N(x;0)}{\partial q}$

III. 計算  $SD_j(x) = \nabla A_0 \frac{\partial N}{\partial q_j} - \sum_{i=1}^m \left[ \sum_{x \in s_0} A_i(x) \cdot \nabla A_0 \frac{\partial N}{\partial q_j} \right] A_i(x)$

與  $SD_{j+4}(x) = \nabla A_0 \frac{\partial W}{\partial p_j} - \sum_{i=1}^m \left[ \sum_{x \in s_0} A_i(x) \cdot \nabla A_0 \frac{\partial W}{\partial p_j} \right] A_i(x)$

$$\text{IV. 計算 Hessian matrix} = \sum_x SD^T(x)SD(x)$$

迭代直到收斂：

- (1) 藉由參數  $p$  與  $q$  來形變輸入影像  $I$  去計算  $I(N(W(x; p); q))$
- (2) 計算誤差影像  $I(N(W(x; p); q)) - A_0(x)$
- (3) 計算  $\sum_x SD^T(x)[I(N(W(x; p); q)) - A_0(x)]$
- (4) 計算  $\Delta p = H^{-1} \sum_x SD^T [I(N(W(x; p); q)) - A_0(x)]$
- (5) 更新參數  $(N \circ W)(x; q, p) \leftarrow (N \circ W)(x; q, p) \circ (N \circ W)(x; \Delta q, \Delta p)^{-1}$

收斂後：

- (1) 計算紋理參數  $\lambda_i = \sum_x A_i(x) \cdot [I(N(W(x; p); q)) - A_0(x)]$

圖 2-22 與圖 2-23 為藉由 IC 演算法結合 GSNT 的方法得出正面人臉與具姿態變化的迭代結果，可看出具姿態變化的迭代結果明顯比圖 2-21 還要來的準確許多。

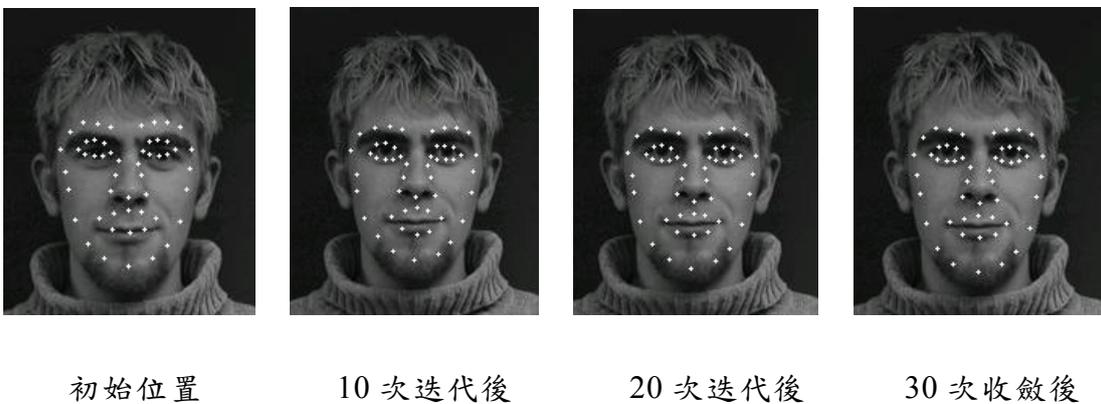


圖 2-22 結合 GSNT 之 IC 演算法迭代出的正面人臉

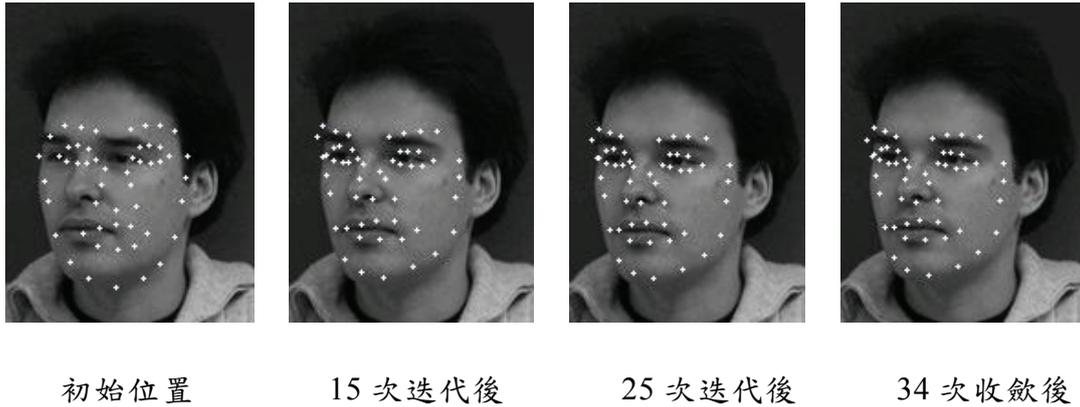


圖 2-23 結合 GSNT 之 IC 演算法迭代出的具姿態變化之人臉

### 2.5.3 The Normalization Inverse Compositional (NIC) Algorithm

在 Inverse compositional 演算法中，當輸入影像  $I(x)$  經過形變後去計算影像的差值  $E(x)$  時，其差值的結果並沒有經過正規化的動作而繼續下一步的迭代流程，此方式可能會造成迭代到人臉上的特徵點產生些許誤差，而導致特徵點無法完全對準人臉的形狀。因此就會在每一次的迭代過程中藉由計算紋理參數來對影像差值的部份做正規化[35]，以補償最後迭代到人臉形狀的特徵點位置，其表示式如下：

$$\lambda_i = -\sum_x A_i(x) E(x) \quad (2-60)$$

$$E_{norm}(x) \leftarrow E(x) + \sum_{i=1}^m \lambda_i A_i(x) \quad (2-61)$$

### 2.5.4 結合 GSNT 之 NIC Algorithm

基於 2.5.2 節及 2.5.3 節的描述下，得知在影像的迭代過程中，若是利用 Global shape normalizing transform 則可以容忍較多姿態的人臉，另外若是利用影像差值補償也可將追蹤的特徵點更準確，由於此兩者均可以讓系統擷取

到較好的特徵，使得後續的辨識演算法的部份更是能發揮其效果。因此我們將 2.5.2 節提出的 Global shape normalizing transform 及 2.5.3 節提出的 Normalization inverse compositional algorithm 做結合，則可產生以下的迭代步驟：

前處理：

I. 找出樣板影像  $A_0(x)$  的梯度值

II. 找出 Jacobian  $\frac{\partial W(x;0)}{\partial p}$  與  $\frac{\partial N(x;0)}{\partial q}$

III. 計算  $SD_j(x) = \nabla A_0 \frac{\partial N}{\partial q_j} - \sum_{i=1}^m \left[ \sum_{x \in S_0} A_i(x) \cdot \nabla A_0 \frac{\partial N}{\partial q_j} \right] A_i(x)$

與  $SD_{j+4}(x) = \nabla A_0 \frac{\partial W}{\partial p_j} - \sum_{i=1}^m \left[ \sum_{x \in S_0} A_i(x) \cdot \nabla A_0 \frac{\partial W}{\partial p_j} \right] A_i(x)$

IV. 計算 Hessian matrix  $= \sum_x SD^T(x)SD(x)$

迭代直到收斂：

(1) 藉由參數 p 與 q 來形變輸入影像 I 去計算  $I(N(W(x; p); q))$

(2) 計算誤差影像  $E(x) = I(N(W(x; p); q)) - A_0(x)$

(2a) 藉由式(2-58)與式(2-59)來計算  $\lambda$  與  $E_{norm}(x)$

(3) 計算  $\sum_x SD^T(x)E_{norm}(x)$

(4) 計算  $\Delta p = H^{-1} \sum_x SD^T E_{norm}(x)$

(5) 更新參數  $(N \circ W)(x; q, p) \leftarrow (N \circ W)(x; q, p) \circ (N \circ W)(x; \Delta q, \Delta p)^{-1}$

收斂後計算：

$$(1) \text{ 計算紋理參數 } \lambda_i = \sum_x A_i(x) \cdot E_{norm}(x)$$

圖 2-24 為只經由 Global shape normalizing transform 所迭代出的人臉影像特徵點，可看出在特徵點並沒有完全追蹤到人臉的輪廓上，但是若結合 GSNT 之 NIC 演算法，如圖 2-25 所示，而表 2-1 所表示為圖 2-24 與圖 2-25 經由每次迭代時所追蹤到的特徵點所形變出的紋理部份與原始影像的紋理部份所取的差值(以 RMS 值表示)可看出在我們提出的演算法中，其與原影像的差值較小，故由此可驗證出原本未能追蹤到人臉輪廓的特徵點，經由影像差值的補償後，其追蹤的效果更好。

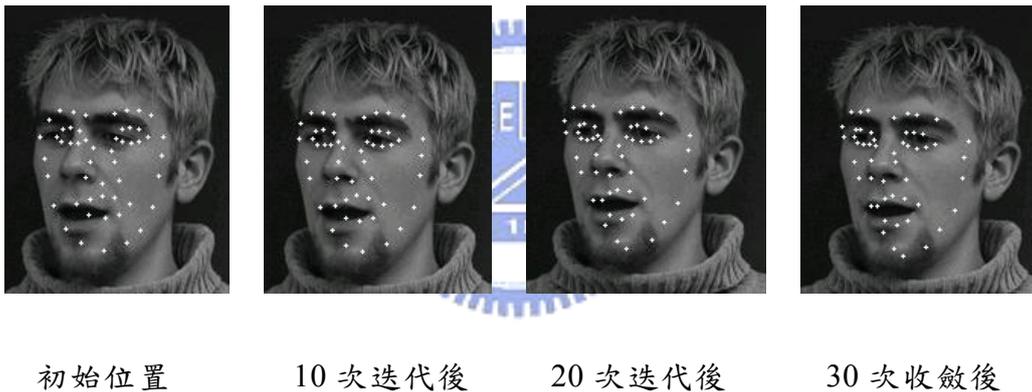


圖 2-24 結合 GSNT 之 IC 演算法迭代出之人臉

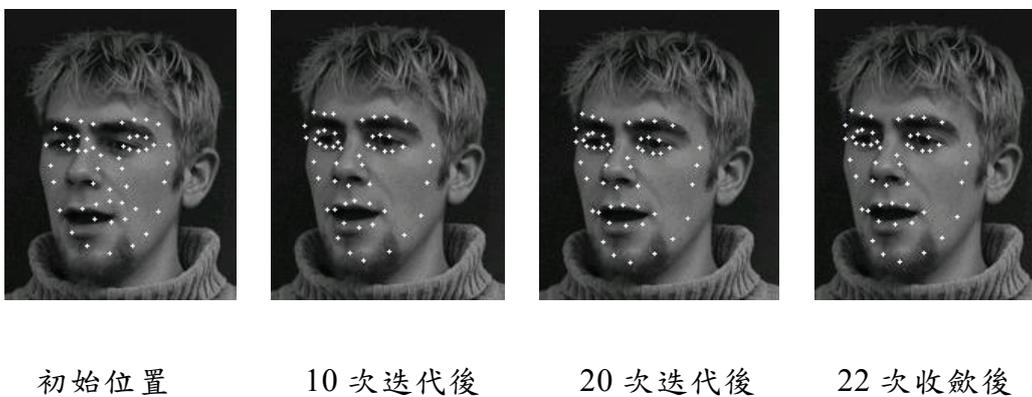


圖 2-25 結合 GSNT 之 NIC 演算法迭代出之人臉

表 2-1 GSNT+IC 與 GSNT+NIC 迭代影像差值比較

Approach Iterative times	GSNT+IC	Proposed method
5	26.4487	23.8958
10	22.6183	18.7853
15	21.84	18.4432
20	21.6116	18.479
25	21.4897	18.3221

## 2.6 結論與討論

本系統利用膚色及特徵判斷完成一即時性的人臉偵測系統，在臉部特徵方面，我們使用區域之灰階值總和與鄰近的區域的灰階值總和來判斷人臉的特徵，因此在人臉的旋轉有一定的容錯率，人臉可左右旋轉 30 度，抬頭與低頭也可達到約 20 度，皆可以準確地及快速地將正面的人臉框選出來。

在多姿態人臉的特徵擷取中，首先必須先定義好人臉上的特徵點，利用定義好的特徵點套用 PCA 的方法來得出 Mean shape 以及形狀的特徵向量。而紋理的部份則是藉由影像形變並也套用 PCA 的方法得出 Mean texture 以及紋理的特徵向量，因而也建立出 Model。

當輸入的人臉影像進入本系統時，在此套用可形變樣板追蹤的 Lucas-Kanade 迭代演算法來追蹤出多姿態人臉所對應的特徵點，不過由於 Lucas-Kanade 迭代演算法其運算量較大，導致系統速度變慢，於是對此演算法做改良，利用 Normalization Inverse Compositional 演算法來取代，此方法可以讓迭代時的特徵點能更加準確，並結合 Global Shape Normalizing Transform 讓人臉可具有較大的姿態容忍度，以便於做後續的辨識效果能更好。

## 第三章 人臉辨識演算法

為了縮短辨識時所需的時間，我們利用特徵臉的方法對輸入的臉部影像進行資料維度的化減，以減少辨識時所需的資料量。最後我們再利用倒傳遞類神經網路對化減後的資訊進行辨識，以得到家庭成員人臉辨識的結果。

### 3.1 特徵臉演算法

本系統主要的目的是為了應用在家用寵物機器人上，且由於本系統的平台是嵌入式系統，所以系統的即時性非常的重要。因此，我們採用主成份分析的特徵臉演算法對臉部影像作維度的化減，並搭配倒傳遞類神經網路做最後的辨識。主成份分析法在第二章雖然已有些許的介紹，但在下面的內容中會針對特徵臉的部分再做一次詳盡的介紹。主成份分析是透過 K-L(Karhunen-Loeve)轉換來達成影像上統計資訊的維度化減的方法，其可將輸入訊號的高維度資訊中所有的相關性消去，以化減資料的維度。而在演算法的運算過程中，將二維矩陣的資訊，以一維的向量方式來表示，如圖 3-1 所示，並且從輸入的所有影像中萃取出臉部特徵空間，臉部特徵空間的維度比原空間維度小，將輸入的人臉影像投影到臉部特徵空間，可以較少的維度的資訊，以達到減少維度的目的。但對原始資料進行縮減，不可避免的會產生資料的遺失，因此適當的選取主成份向量才能盡可能的降低資訊的遺失。

以下我們概略說明主成份分析法，當有二維空間資料如圖 3-2(a)所示，利用下面將介紹的特徵臉方法求得資料主要分布的方向為  $V$ 。則  $V$  為主要特徵向量，當所有資料到以  $V$  為軸的一維空間上時，如圖 3-2(b)，可以看到在二維空間中的兩個不同的類別投影到一維的特徵空間後，還是能明顯的區分出兩個類別。所以適當的選取主成份的軸即能縮減資料量，又不會造成資料分類錯誤。

在求取特徵空間的方法中，一般利用 K-L 轉換求得，我們先取  $M$  張大小為  $P \times Q$  維的人臉影像，由第二章的特徵擷取部份可知，由於人臉形狀經過形變的關

係，其有效的影像資訊並不完全是一整張  $P \times Q$  維的影像，而是屬於臉部紋理的影像，如圖 3-4 所示，所以當輸入影像轉為一維的向量時，僅需將屬於人臉紋理的影像資訊就好，故我們將每張影像轉為  $Z \times 1$  維的向量，以  $i_1, \dots, i_M$  表示。如下為第  $k$  張影像所轉成的行向量  $i_k$

$$i_k = \begin{bmatrix} x_{k1} \\ x_{k2} \\ \vdots \\ x_{kj} \\ \vdots \\ x_{kZ} \end{bmatrix}, k=1 \sim M \quad (3-1)$$

其中  $x_{kj}$  代表  $i_k$  的第  $j$  個分量。M 張影像的平均值  $\bar{i}$  則為

$$\bar{i} = \frac{1}{M} \sum_{j=1}^M i_j \quad (3-2)$$

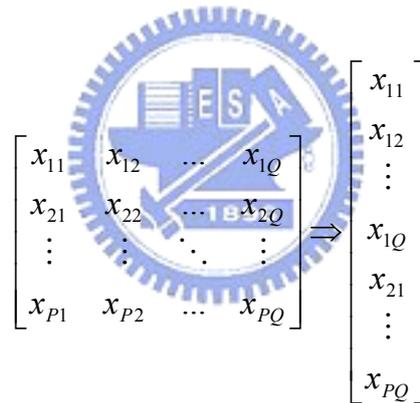
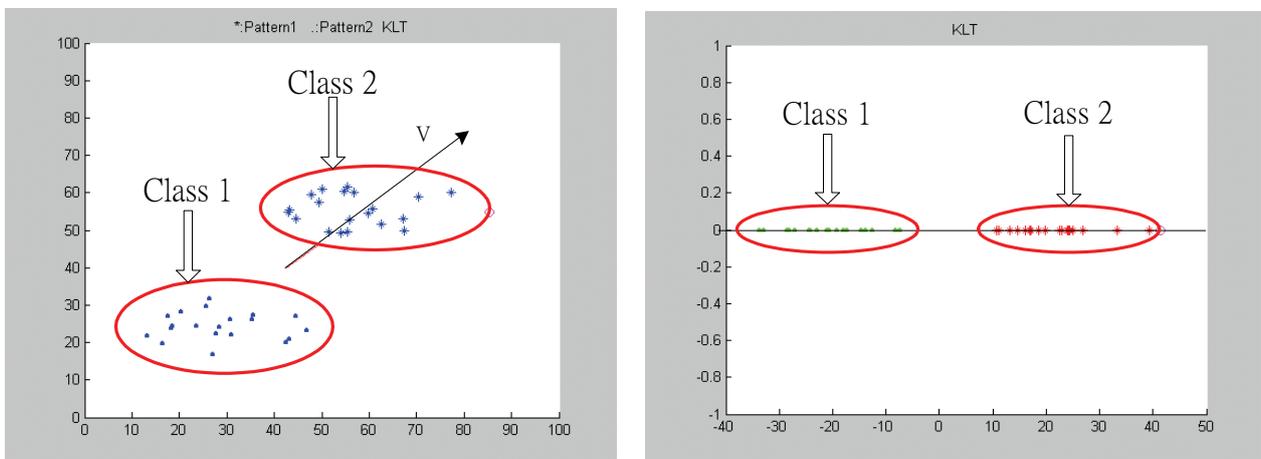


圖 3-1 二維影像資料轉換為一維向量



(a) 資料分佈與其主要分量

(b) 二維資訊在一維空間上投影

圖 3-2 資料分佈與其在主分量上的投影



圖 3-3 形變後之臉部紋理

平均值  $\bar{i}$  代表了  $M$  張影像所共有的特徵成份。為突顯這些影像彼此間不同的成份，必須減去共有的部份，已得到  $M$  張差像(difference image)向量  $\Phi_j$

$$\Phi_j = i_j - \bar{i}, j=1 \sim M \quad (3-3)$$

藉由  $\Phi_j$  向量，我們希望找出一向量  $u_1$ ，使得所有差像  $\Phi_j$  在  $u_1$  上的投影平方和為最大，即：

$$\begin{aligned} \max_{u_1} S &= \sum_{j=1}^M \frac{(\Phi_j^T u_1)^2}{\|u_1\|^2} \\ &= \frac{u_1^T \left( \sum_{j=1}^M \Phi_j \Phi_j^T \right) u_1}{u_1^T u_1} \\ &= \frac{u_1^T C u_1}{u_1^T u_1} \end{aligned} \quad (3-4)$$



其中  $C$  為一協方差矩陣(covariance matrix)，其維度為  $Z \times Z$  維。

$$C = \sum_{j=1}^M \Phi_j \Phi_j^T = B B^T \quad (3-5)$$

$$B = [\Phi_1 \quad \Phi_2 \quad \cdots \quad \Phi_M] \quad (3-6)$$

根據 Rayleigh Quotient 定理[36]，若  $u_1$  為  $C$  矩陣中最大特徵值所對應的最大特徵向量，則此時  $S$  為最大，其餘之特徵向量將使  $S$  有局部極大值(Local Maximun)。為了求得這些特徵向量，對矩陣  $C$  求其特徵值(eigenvalue)  $\lambda_k$  與特徵向量(eigenvector)  $u_k$ ，滿足下列條件：

$$Cu_k = \lambda_k u_k, k=1\sim M \quad (3-7)$$

其中 C 矩陣為  $Z \times Z$  維度，求解其所有的 eigenvalues 及 eigenvectors 需龐大的計算量。若不能有效的降低計算量，將使得在訓練特徵空間時需要花上相當長的時間。

在文獻[37]中提供了一個解決方法，用化減 K-L 轉換的計算，可以減少求 PCA 時所需的計算量。先計算  $B^T B$  矩陣(維度為  $M \times M$ )的特徵向量  $v_k$ ，滿足下式：

$$B^T B v_k = \mu_k v_k \quad (3-8)$$

將上式兩邊同乘 B，可得到：

$$BB^T B v_k = \mu_k B v_k \quad (3-9)$$

令  $B v_k = u_k$ ，則上式改寫為

$$BB^T u_k = \mu_k u_k \quad (3-10)$$

因此當求出  $v_k$  之後，便可以利用  $v_k$  線性組合來求  $C = BB^T$  矩陣之  $u_k$

$$u_k = B v_k \quad (3-11)$$

一般情況， $M \ll Z$ 。因此，求解的  $v_k$  計算量遠比求解  $u_k$  小得多。如此便可以很快的求得臉部特徵空間。

### 3.2 倒傳遞類神經網路

在本論文中，我們使用倒傳遞類神經網路(Back Propagation Neural Network, BPNN)作為辨識系統，其主要的想法是將一組樣本輸入與輸出間映射的問題變成非線性最佳化的問題，加入隱藏層後可使最佳化問題的可調參數增加，並利用迭代運算的方式求解最佳值，因而增強類神經網路的學習能力，使系統錯誤率降至

最低。

倒傳遞類神經網路的架構為多層感知器(Multilayer Perceptron, MLP)，一般常用的 MLP 網路架構如圖 3-4 所示，其中  $X = [x_1, x_2, \dots, x_N]^T$  代表輸入向量；在輸入層與輸出層之間有隱藏層(hidden layer)，之所以稱為隱藏層，是因為該層與外界資訊隔離，網路只透過輸入層與輸出層作為對外的介面。輸入層與輸出層是用來表現輸入項與輸出值，兩者神經元數目依問題的形式而定；隱藏層神經元數目往往需以試誤法決定，隱藏層的層數可以依問題複雜度由一層增加到數層，在本論文中僅需一層隱藏層即可。網路中靠相關權重連結各層之神經元，輸入值由輸入層直接傳入隱藏層， $w_{ji}$  代表輸入層與隱藏層之權重，經加權累加後再透過活化函數轉換可得一輸出值， $y = [y_1, y_2, \dots, y_P]^T$  為輸出層之輸出值，同理再傳入輸出層， $w_{kj}$  代表隱藏層與輸出層之權重。

在本系統倒傳遞類神經網路的訓練中，其演算法流程如圖 3-5 所示，在網路參數設定上，輸入層的資料是投影至特徵空間的人臉資料，若欲辨識之家庭成員

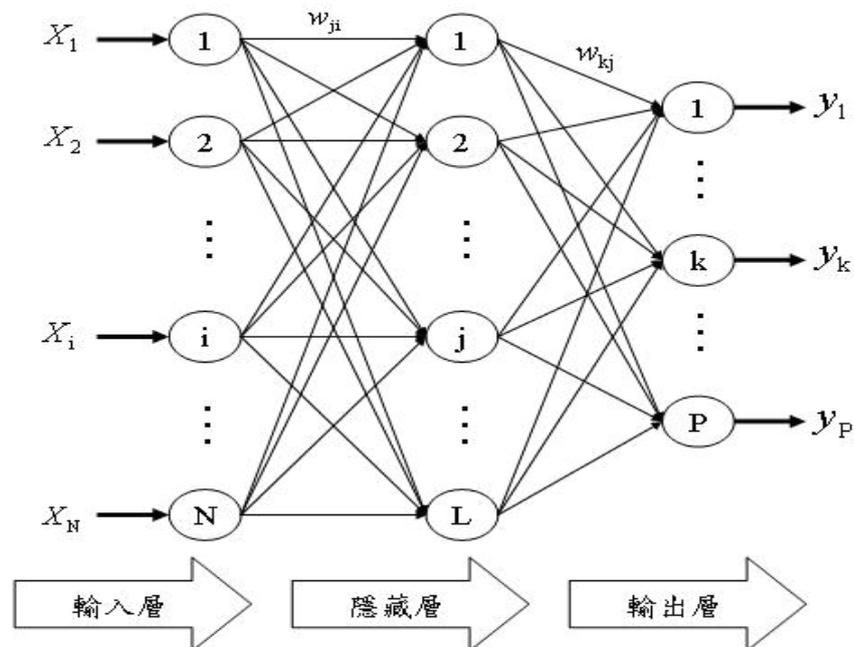


圖 3-4 倒傳遞類神經網路架構圖

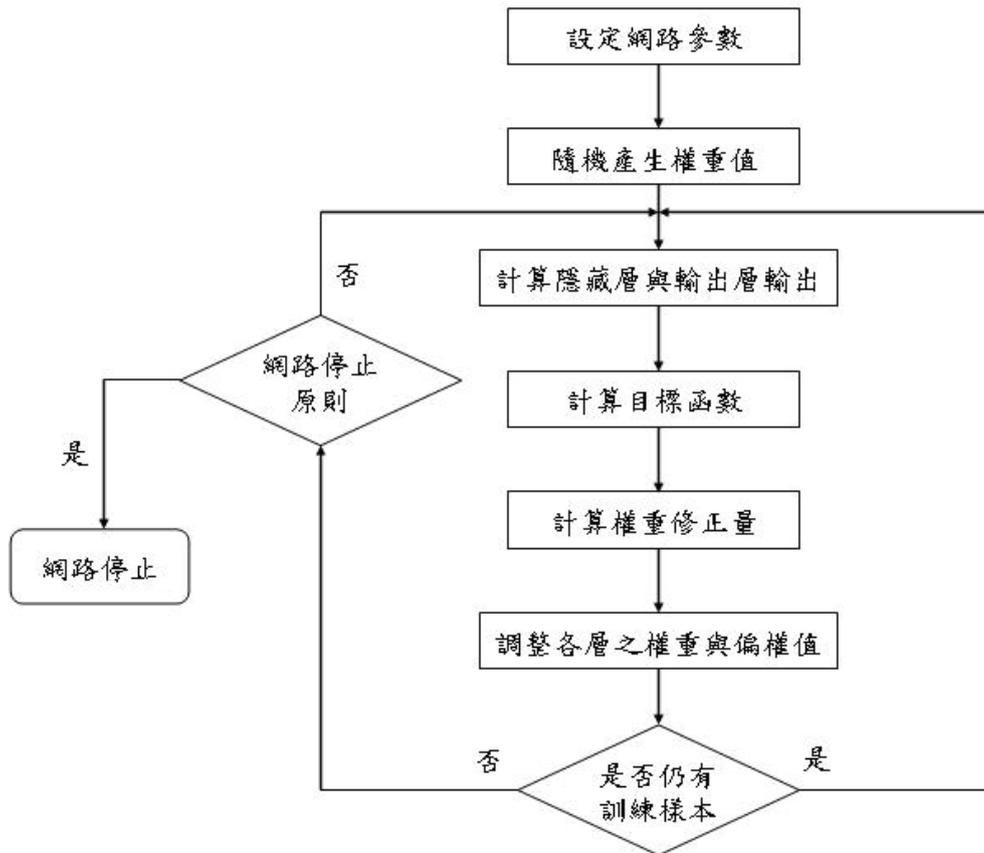


圖 3-5 倒傳遞演算法流程

有  $P$  位，則第一位家庭成員的目標輸出值設為  $[1, -1, -1, \dots, -1]$ ，其中  $-1$  有  $P-1$  個，同理若是第二位家庭成員，則目標輸出值為  $[-1, 1, -1, \dots, -1]$ ，以此類推，直到第  $P$  位家庭成員為止。至於在權重方面，我們是以隨機亂數來初始隱藏層權重  $w_{ji}$  與輸出層權重  $w_{kj}$ ，直到訓練穩定且學習收斂後，即可做線上的測試。

### 3.3 資料庫建立流程

資料庫建立流程如圖 3-6 所示，首先將欲當資料庫的人臉影像以人工方式標定出特徵點，利用影像形變法使原始人臉影像形變至由式(2-3)建立出的 Mean shape  $s_0$  的形狀裡，最後在投影至特徵空間即可放入資料庫裡。由於在第 2 章特徵擷取部

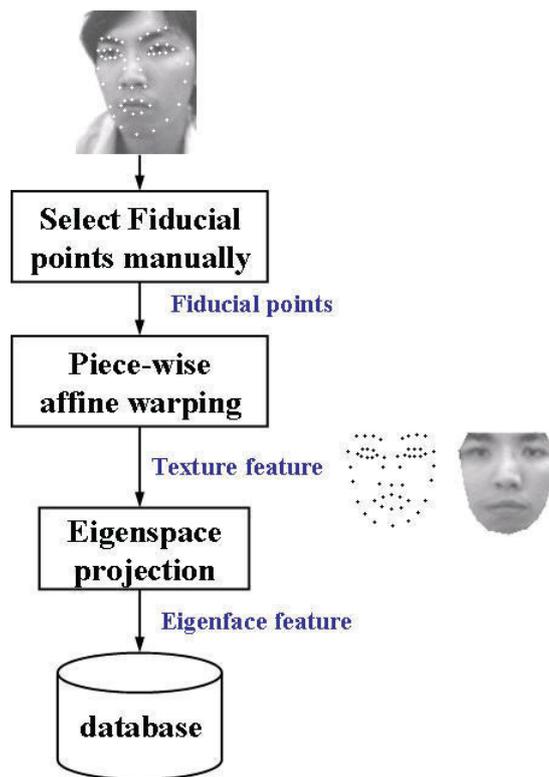


圖 3-6 資料庫建立流程

份已利用 PCA 建立出 Mean shape  $S_0$  以及 Mean texture  $A_0$ ，因為在建立 Appearance model 的過程中需要將欲建立成 model 的人臉影像以人工方式標定特徵點，若在建立特徵臉的資料庫同樣也做標特徵點的動作時，其在離線時所耗費的時間是不小的。因此，若在剛開始建立 Appearance model 時也一併將欲當資料庫之人臉影像以人工方式標上特徵點，則可避免在做特徵臉時需再做標特徵點的動作，以節省建立 Appearance model 與特徵臉資料庫的時間。

### 3.4 結論與討論

在辨識演算法中特徵空間的部份，將高維度資料中所有的相關性消去，以化減高維度資料的資料量。此方法雖然能藉由消去資料間的相關性以提供維度的化減，使原空間中的可分辨的類別在特徵空間內亦可分辨。但還是可能發生在原空

間內相同類別的點，投影到特徵空間後，變成不同類別。或是原空間中不同類別的點，投影到特徵空間後會造成重疊到同一個類別中。使得無法順利分類，因此在特徵空間的選取會影響後面的辨識結果，此為本論文未來需要改進的地方。

在倒傳遞類神經網路部份，網路參數的設定需要做持續的測試，以學習速率的設定來說，若學習速率設定的越小，則收斂過程會越趨緩，但迭代次數會相對地提高；反之若學習速率太大，雖然可以加速搜尋效率，但容易造成搜尋過程中，目標函數產生不穩定的震盪情形，故需要多次的嘗試，才可找出適合此訓練資料的學習速率，同理，其他需設定的網路參數也是需要作多次的測試才可找出穩定收斂的結果。



## 第四章 基於人臉辨識之互動系統設計

本章將設計一套整合人臉辨識系統的寵物機器人與人互動系統，主要是透過 PC 讓嵌入式影像平台與寵物機器人做訊號上的傳輸，圖 4-1 為人臉辨識在機器人與人互動之硬體架構圖，使用嵌入式影像平台之目的是希望利用影像平台即時取像和運算的特性，可以快速地擷取出影像特徵，實行後端辨識的機制。另外，由於在機器人系統整合方面，其包含了機器視覺系統(人臉偵測、追蹤及辨識)與機器人移動系統等等，若每一部分都在 PC 伺服端上處理的話，將會佔據很大的運算量，導致機器人的反應緩慢並且產生驗證困難等問題，若能將每個系統模組化的話，則在系統整合及驗證方面相對而言會較容易實現與解決，因此我們在嵌入式影像平台上實現本人臉辨識系統。本系統之人臉輸入影像是藉由嵌入式影像平台的 CMOS sensor 作取像動作，透過 FPGA 板將影像送進 DSP 後，並計算人臉偵測與辨識演算法產生一輸出訊號，接著透過 PC 來送出動作命令給寵物機器人上的馬達控制卡，並透過手部、耳部與腳部的馬達來產生對應的動作反應。

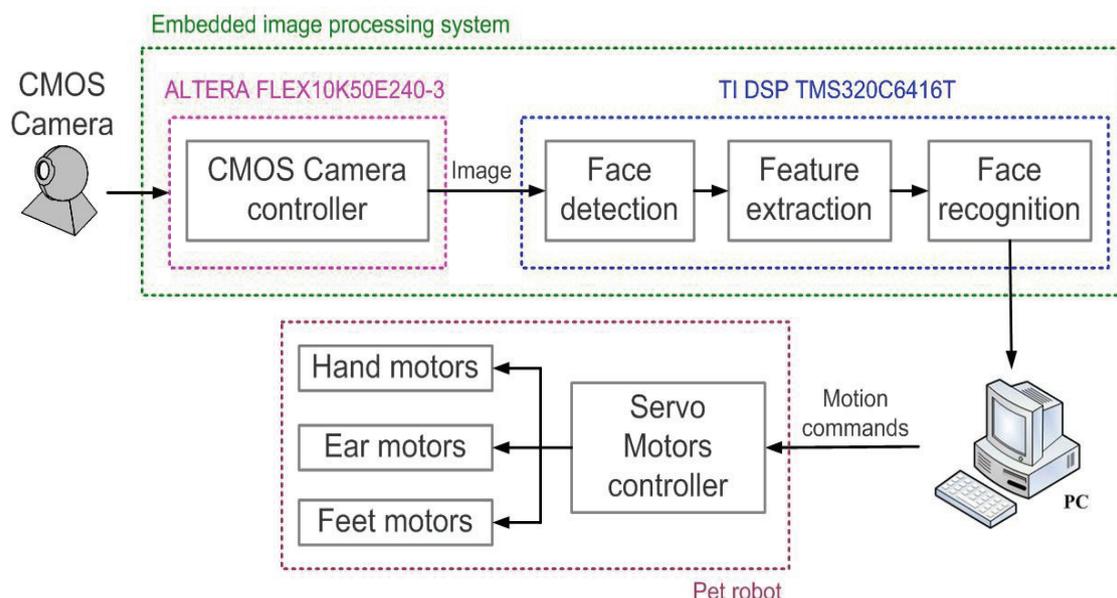


圖 4-1 人臉辨識在機器人與人互動之硬體架構圖

## 4.1 嵌入式影像平台

本論文所採用的影像擷取系統乃為本實驗室所開發之 DSP 嵌入式影像平台 [38]，圖 4-2 為系統架構圖，主要包含了 CMOS 影像感測板、FPGA Board、DSK6416T 影像發展板和一顆緩衝記憶體(Frame Buffer)，其實體圖如圖 4-3 所示。CMOS sensor 其最大有效像素為 640x480，最大更新率為 30 frames/sec，影像輸出格式包含了 8-bit raw data、8/16-bit YCrCb(本系統所選用的即為此)、16-bit RGB 和 24-bit RGB。此外，此 CMOS 感測板尚包含一些影像調整功能，包括了自動白平衡(Auto White Balancing)及 Gamma 參數調整可對整張影像的亮度及色彩飽和度進行調整等功能。在 FPGA 板方面所使用的為 ALTERA 公司生產，型號為 FLEX10K50E240-3，主要功能為處理緩衝記憶體之讀出與寫入的控制訊號、資料排序及產生讓 DSK6416T 讀取影像資料之觸發訊號，另外要透過 I2C 來做 CMOS 感測器的初始化過程，其初始化內容如下：

- 影像更新速率：每秒 30 張 Frames
- 影像輸出格式：8-bit 4:2:2 YCrCb QVGA (320x240 像素大小)
- 輻射校正：Gamma = 2.2
- 色彩飽和度：Saturation=1.5

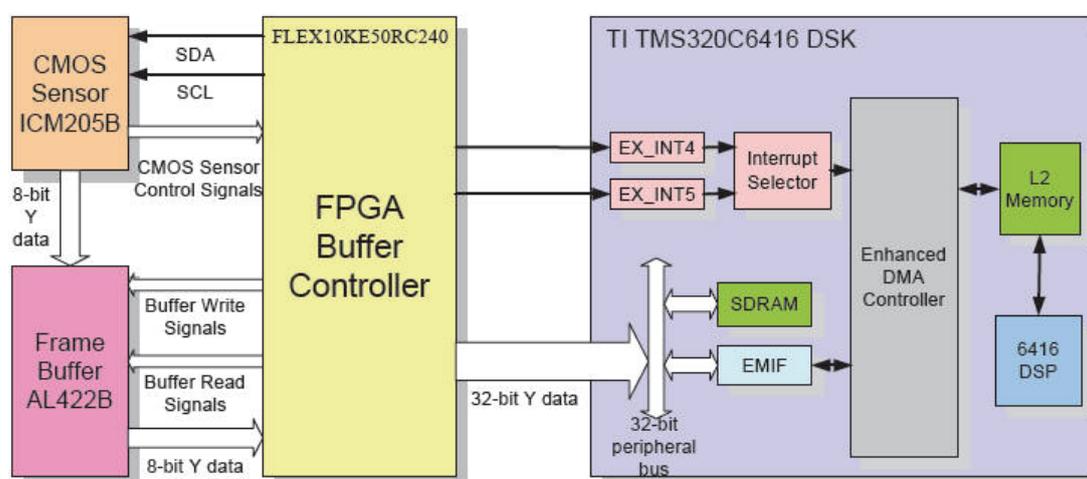


圖 4-2 DSP 影像平台系統架構圖[38]

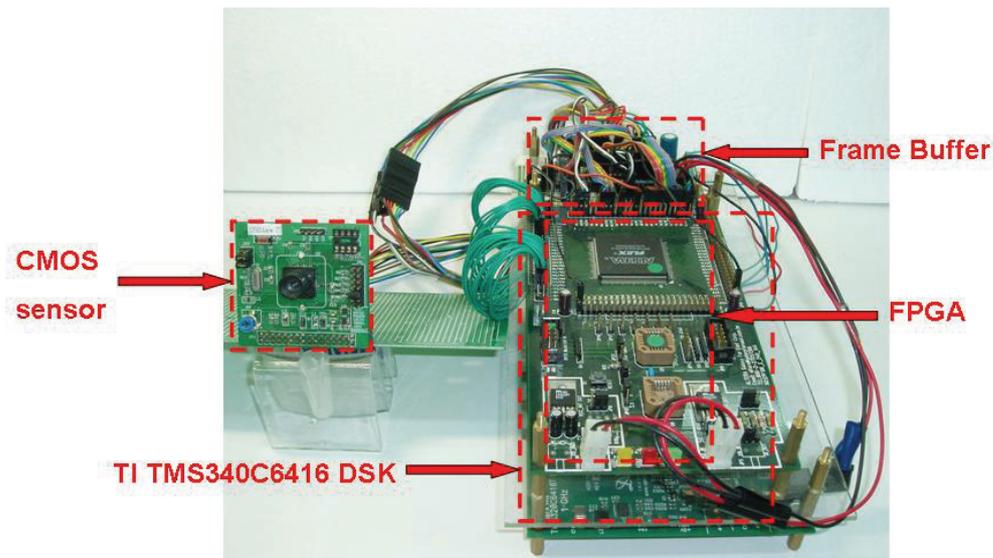


圖 4-3 嵌入式影像平台實體圖

在緩衝記憶體(AL422B)的部分主要做為影像資料儲存之用，其記憶體長度為 384Kbytes 可以完整儲存兩張 QVGA 大小之影像，且資料存取動作為 FIFO(First In First Out)，因為 CMOS sensor 輸出影像資料為連續但卻比較緩慢，無法讓 DSK6416 在短時間內得到一張完整之影像資料，所以使用此緩衝記憶體的目的是在於寫入及讀取資料的速度可以使用不同的工作速度，即使寫入資料的速度慢，我們可以等寫入完整之資料後，再用較快的速度讀取所有的資料(此緩衝記憶體之最快工作時脈為 50MHz)，以節省讀取資料的時間。在影像處理板方面選用的為德州儀器(TI)公司所出產之 TMS320C6416T DSK 發展板，此發展板所使用中央處理器為德州儀器研發之 C6416 處理器，此處理器具有高效能之定點運算、週邊 I/O 擴充容易及內建龐大的記憶體等優點，適合做為嵌入式系統之中央處理控制器，其主要做為影像資料搬運及即時運算的功能，工作頻率為 1GHz。在和外部週邊溝通時，例如馬達運動控制器或電腦，此影像平台利用通用 I/O 訊號線 (GPIO) 完成兩組通用非同步收發傳輸器 (Universal Asynchronous Receiver/Transmitter, UART) 做為通訊介面，此嵌入式影像平台可以直接下達控制命令以控制其他週邊，或者也可將處理過的資訊傳到 PC 端做一整合的動作。

## 4.2 寵物機器人

在我們的寵物機器人中，其內部有 7 顆馬達，可分別獨立控制頭、雙耳、雙手與雙腳，做出動作反應。而負責取像的 CMOS sensor 則是放在機器人的額頭上，如圖 4-4 所示，將所得畫面送入嵌入式影像處理系統，執行辨識軟體以獲得畫面中人臉辨識的輸出結果。而將嵌入式系統應用於寵物機器人之架構圖如圖 4-5 所示，由 FPGA 控制 CMOS sensor 取像存到 frame buffer，再藉由 DSP 的外部記憶體介面(External memory interface, EMIF)將 frame buffer 中的影像取出進行運算，最後 PC 根據辨識結果產生馬達控制訊號，並透過 RS-232 傳送給 RC Servo，使寵物機器人能夠做出反應動作，與使用者進行互動。



圖 4-4 實驗室研製之寵物機器人

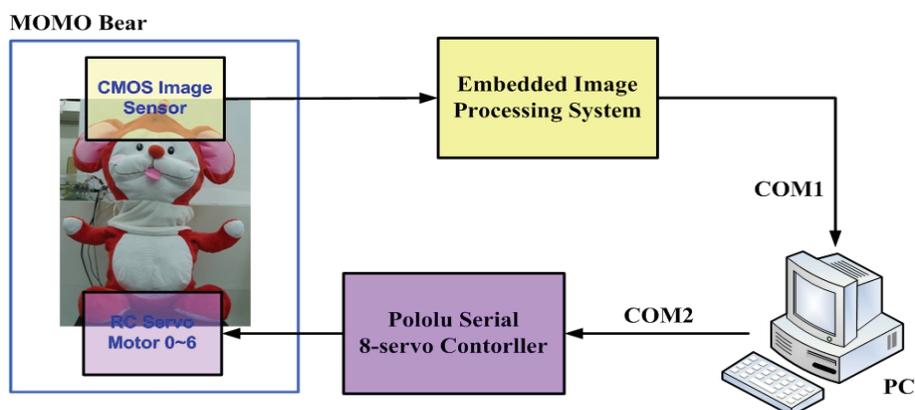


圖 4-5 嵌入式人臉辨識系統應用於寵物機器人架構圖

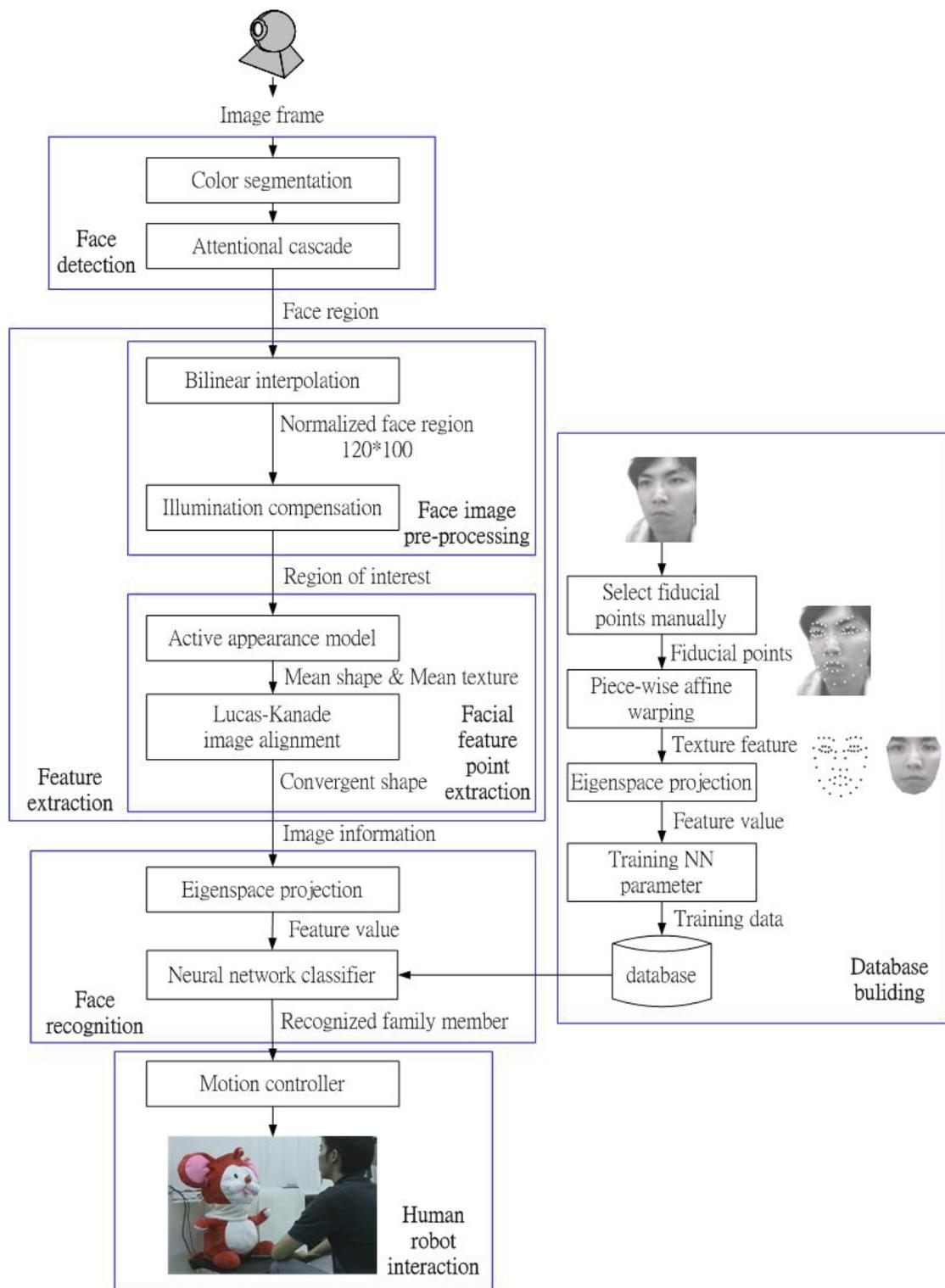


圖 4-6 機器人與人互動之多姿態人臉辨識系統完整架構圖

### 4.3 結論與討論

本論文完成一套寵物機器人與人互動的系統，並利用嵌入式即時影像平台完

成多姿態人臉辨識演算法，透過 PC 送動作命令給裝置在寵物機器人上的 RC Servo，使寵物機器人能夠做出反應動作，並與使用者進行互動。圖 4-6 為機器人與人互動之多姿態人臉辨識系統的完整架構圖，當資料庫的人臉資料建立完成後，透過類神經網路與測試的人臉資料進行比對，藉由得到的比對結果來與機器人進行互動。



## 第五章 實驗結果

為了驗證本論文的人臉影像特徵擷取，及家庭成員人臉辨識系統之效能，我們將藉由以下幾個步驟來驗證：

1. 建立測試用的影像資料庫。
2. 利用 AAM 與 Lucas-Kanade 影像校正演算法迭代出人臉在不同姿態下追蹤結果，其迭代結果分別應用在 UMIST 資料庫[39]以及實驗室內建的資料庫。
3. 利用 1.建立好的資料庫以及 UMIST 資料庫的人臉影像進行辨識後之結果。
4. 家庭成員與非家庭成員之人臉辨識結果。

### 5.1 建立資料庫

為了要驗證本論文提出的人臉影像辨識及其決策，需要建立一套合適之資料庫以供實驗使用。在做特徵擷取演算法時，其資料庫也等同於建立模型的部份。

由於本論文主要的應用環境是在家庭裡，其資料庫的人數並不會太多，所以在我們的實驗中，均假設有五位家庭成員，故在建立資料庫時，也是以此五位家庭成員為主。就實驗室內建資料庫而言，由於實際應用時，使用者在測試時可能會有許多不同的姿態，為了能夠產生各種姿態下的分析，並且也考慮到本論文之人臉偵測方法的極限，所以在初步測試我們總共假設了 9 種姿態，分別為正面 0 度、往左偏轉 10 度、20 度、30 度、往右偏轉 10 度、20 度、30 度、正面朝上 15 度、及正面朝下 20 度。每種角度每人各做 5 次，則有 45 張影像，範例影像如圖 5-1 所示；而就 UMIST 資料庫而言，我們也是從資料庫裡挑選出任意 5 位資料庫成員來做測試。由於但由於本論文演算法在角度容忍度上有一定的限制，若是角度偏差太大，則特徵擷取會失敗，故我們只取資料庫裡頭部左右偏轉 30 度以內的人臉做分析，故總共取了 105 張影像做後續的測試，其部份的範例影像



圖 5-1 實驗室內建資料庫範例影像



圖 5-2 UMIST 資料庫範例影像[39]

如圖 5-2 所示。

## 5.2 影像校正演算法迭代結果

本實驗主要是將測試的人臉影像利用 Lucas-Kanade 影像校正演算法來做臉部的特徵點定位，以擷取出人臉影像的紋理部分，判斷追蹤到的特徵點是否收斂有兩種方式，一種為判斷每次迭代後的形狀增量係數 $\Delta p$ 是否小於一臨界值，但由於每張人臉影像的收斂次數不同，可能需要多次的迭代或是少數幾次的迭代來讓追蹤到的特徵點收斂，但因為本系統所使用的平台為嵌入式即時影像系統，若迭代次數太多，則會嚴重影響系統的運算速度，故我們採用固定迭代次數的方式，其目的是為了要防止迭代次數過多的情形發生。當輸入影像進入到特徵擷取

階段時，會將影像正規化為 120x100 的灰階影像，而就迭代次數而言，我們目前設定的次數為 16 次，其運算時間為 1 秒鐘左右。就實驗室內建的人臉影像而言，每人各有 45 張影像，且包含 9 種不同的姿態，故我們在每個人的每一種姿態取一張影像來建立模型，用以計算出形狀模型與紋理模型，同時也當作資料庫的訓練資料，圖 5-3 所示為實驗室內建資料庫的人臉影像所得出的形狀模型與紋理模型。而剩餘的影像(180 張)則當成是測試資料的部份。圖 5-4 所示為部份的測試人臉資料經模型在 9 種姿態下之迭代 16 次後的結果，其排列依序為正面、左邊偏轉 10 度、右邊偏轉 10 度、左邊偏轉 20 度、右邊偏轉 20 度、左邊偏轉 30 度、

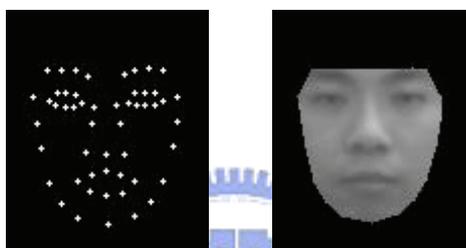


圖 5-3 實驗室人臉資料庫之 (a)形狀模型 (b)紋理模型

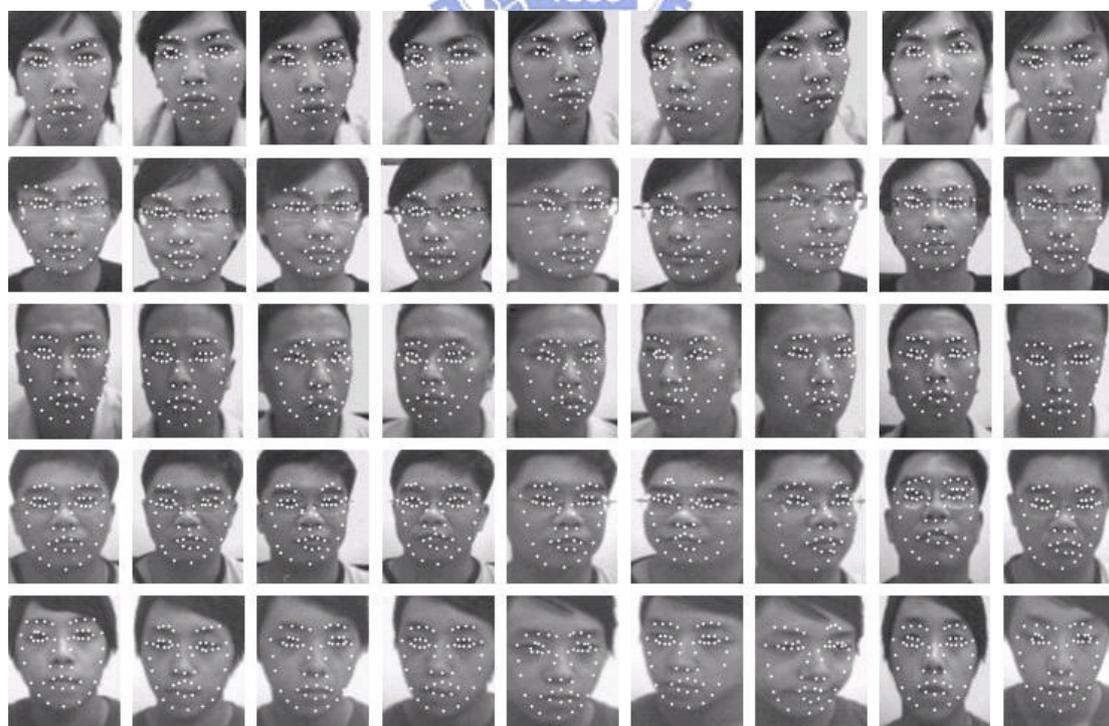


圖 5-4 實驗室資料庫之多姿態人臉經影像校正演算法迭代 15 次後結果

右邊偏轉 30 度、正面朝上 15 度以及正面朝下 20 度，可看出在左右偏轉 30 度角時，其特徵點已經無法完全追蹤到定點上，但因為特徵點均還是落在人臉的輪廓裡面，所以對於處理紋理資訊的影響並不大。圖 5-5 所示為圖 5-4 中第一列人臉影像經迭代後之 RMS 差值收斂情形，橫軸代表迭代次數，縱軸為人臉影像的 RMS 差值，由圖中可證實此演算法確實可找出最匹配的人臉模型並做後續的辨識分析。

就 UMIST 資料庫而言，每人各任取 10 張影像來建立形狀模型與紋理模型，如圖 5-6 所示，剩下的影像(55 張)則當成是測試資料的部份，圖 5-7 為 UMIST 資料庫之部份多姿態人臉經影像校正演算法迭代 15 次後的結果，再由圖 5-8 所示的迭代圖 5-7 第一列影像之 RMS 差值收斂情形可驗證本系統的特徵擷取演算法能成功應用在其他的人臉資料庫上。

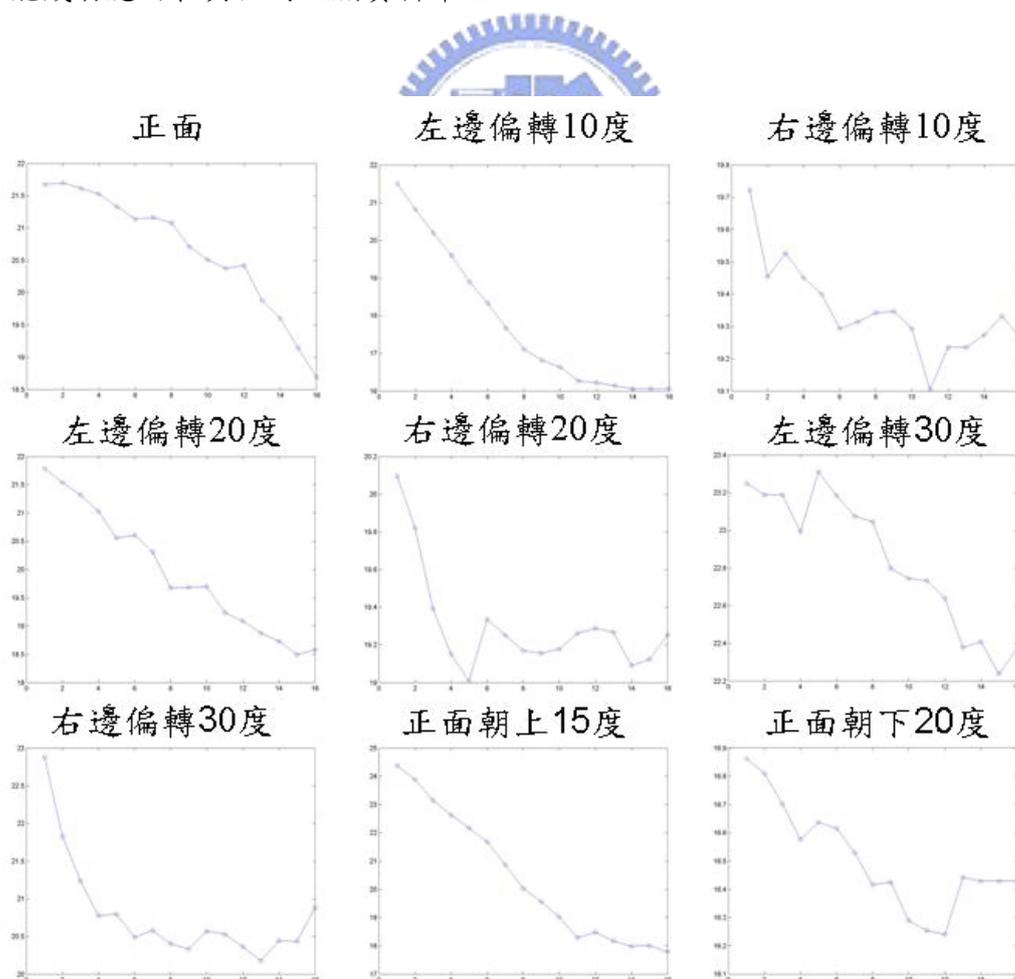


圖 5-5 實驗室資料庫人臉影像經迭代後之 RMS 差值收斂情形

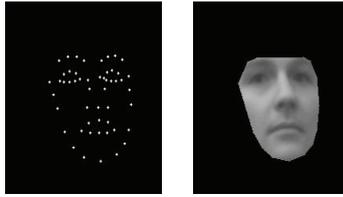


圖 5-6 UMIST 人臉資料庫之 (a)形狀模型 (b)紋理模型



圖 5-7 UMIST 資料庫之多姿態人臉經影像校正演算法迭代 15 次後結果

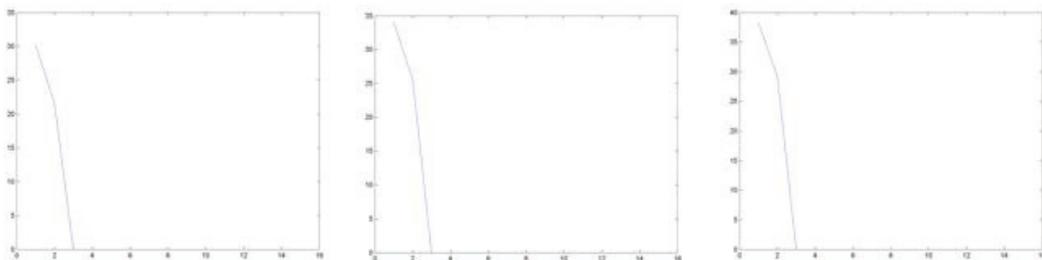


圖 5-8 UMIST 資料庫人臉影像經迭代後之 RMS 差值收斂情形

### 5.3 人臉影像辨識結果

本實驗主要是將經過人臉特徵擷取後的影像做維度上的化減與辨識的動作，在此做了兩個實驗，除了由實驗室裡建立的多姿態人臉影像資料庫外，為求

本系統的穩定性，我們也從目前人臉辨識研究領域裡常用的資料庫中，挑選具有多姿態的 UMIST 資料庫，藉由這兩個資料庫來驗證本系統的性能。

### 5.3.1 實驗室內建多姿態人臉影像資料庫

就實驗室內建的多姿態人臉影像資料庫而言，每人有 45 張人臉影像，成員分別為任傑、孟儒、巧敏、格豪與振暘，我們各取 9 張做訓練資料，剩下的 36 張則為測試資料，分成 4 張正面，往左偏轉 10 度、20 度與 30 度各 4 張，同理，往右偏轉也是各 4 張，正面朝上 4 張以及朝下 4 張。本實驗中，利用 PCA 得到的特徵空間其維度為 53 維，至於在 BPNN 的實驗中，其網路架構如圖 5-9 所示，輸入層的數目即為特徵空間的維度，而本實驗的家庭成員為 5 位，故輸出層有 5 個輸出值，隱藏層神經元數目的選定經多次的測試下是設為 125 個，迭代次數經測試下為 10000 次，圖 5-10 為 5 個輸出值經過學習後之收斂情形，可看出雖然在開始部分以及迭代到中間部分時，都突然有稍微發散的情形產生，但是依迭代到最後時都還是會收斂。表 5-1 為經過多姿態辨識演算法的結果，分別將正面、正面朝上及朝下歸於一類、左邊偏轉為一類以及右邊偏轉為一類，故每一類均會有 12 張影像作辨識，其平均辨識率可達 95.56%。

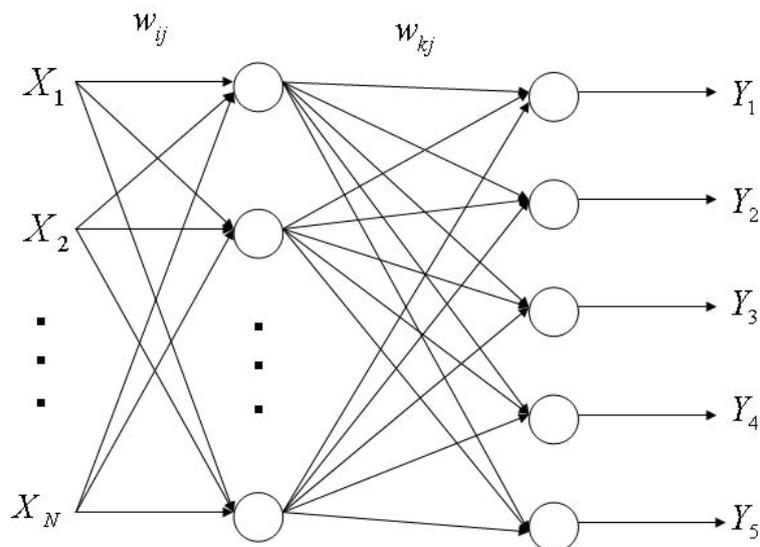


圖 5-9 多姿態人臉辨識之類神經網路架構圖

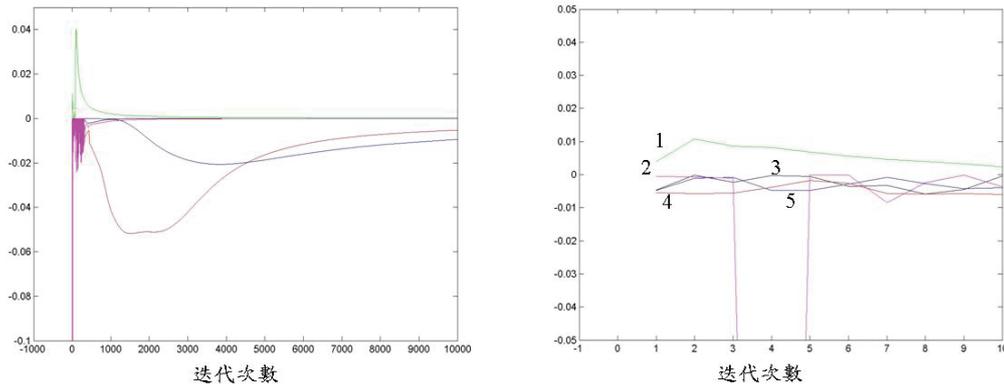


圖 5-10 實驗室內建資料庫經類神經網路訓練之 (a)收斂情形 及 (b)初始狀態

表 5-1 家庭成員多姿態人臉辨識結果

Input	Output	仕傑	孟儒	巧敏	格豪	振暘	辨識率
	正面+朝上+朝下	12	0	0	0	0	97.22%
	左邊偏轉10~30度	11	0	0	0	1	
	右邊偏轉10~30度	12	0	0	0	0	
	正面+朝上+朝下	0	12	0	0	0	97.22%
	左邊偏轉10~30度	0	12	0	0	0	
	右邊偏轉10~30度	1	11	0	0	0	
	正面+朝上+朝下	0	0	12	0	0	97.22%
	左邊偏轉10~30度	0	0	11	0	1	
	右邊偏轉10~30度	0	0	12	0	0	
	正面+朝上+朝下	1	0	0	11	0	91.67%
	左邊偏轉10~30度	0	0	0	12	0	
	右邊偏轉10~30度	0	0	0	10	2	
	正面+朝上+朝下	2	0	0	0	10	94.44%
	左邊偏轉10~30度	0	0	0	0	12	
	右邊偏轉10~30度	0	0	0	0	12	

### 5.3.2 UMIST 多姿態人臉影像資料庫

就 UMIST 資料庫的人臉多姿態影像而言，我們共取了 105 張影像，其中每人各取 10 張做訓練資料，剩餘的 55 張則為測試資料，如圖 5-11 所示，大致上可分析出正面有 20 張，右邊偏轉有 23 張及右邊偏轉有 12 張。利用 PCA 得到特徵空間的維度為 49 維，等同於 BPNN 的輸入層數目，而隱藏層的神經元數目為 150 個，迭代次數為 10000 次，收斂情形如圖 5-12 所示，表 5-2 為 UMIST 資料



圖 5-11 UMIST 資料庫之測試人臉多姿態影像圖

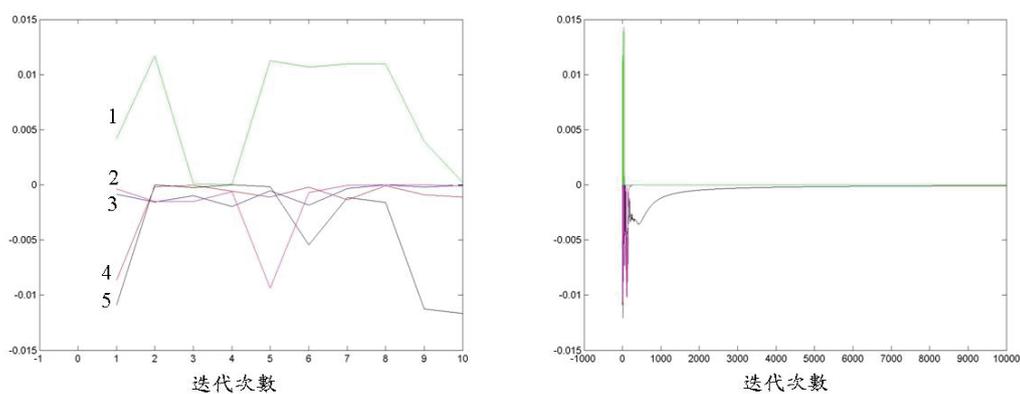


圖 5-12 UMIST 資料庫經類神經網路訓練之初始狀態以及收斂情形

表 5-2 UMIST 資料庫經多姿態人臉辨識結果

Input \ Output	正確	不正確	辨識率
正面	18/20	2/20	90%
右邊偏轉	21/23	2/23	91.3%
左邊偏轉	11/12	1/12	91.67%

庫的多姿態人臉辨識結果，其得出的平均辨識率為 91%。由此可驗證出人臉在多姿態的情形下，本系統能夠辨識出受測者為某一位家庭成員的機率均可達九成以上。

## 5.4 家庭成員與非家庭成員之人臉辨識結果

由於在本論文中，寵物機器人的應用環境是在家庭之中，所以除了讓寵物機器人能夠辨別家庭成員以外，並希望也能夠寵物機器人辨別出使用者是否為家庭成員，而給予不同的互動模式。

在本實驗中，由於要辨認非家庭成員，就 5.3 節之類神經網路來說，當測試的人臉資料進入到類神經網路時，只要取當中輸出層最高的神經元即可辨識出該成員，但是若測試資料為非家庭成員時，發現系統還是會將此非家庭成員辨識成家庭裡的某一位成員，若在實用性來說是相當不足的。於是我們對此採用另一套作法，首先將每一位家庭成員各自訓練出一組類神經網路，其網路架構如圖 5-13 所示，當測試人臉資料進到辨識階段時，此資料會分別送到每個家庭成員的類神經網路並搭配上原空間之距離作權重去判別是否為該家庭成員，若非該家庭成員，則會繼續進入到下一位家庭成員之類神經網路，以此類推，若該人臉資料均無法通過每位家庭成員的類神經網路輸出門檻值時，則可判定該人臉資料為非家庭成員，其流程圖如圖 5-14 所示。

假設有 8 位成員進入到類神經網路作辨識，其中 5 位為家庭成員，分別為仕傑、孟儒、巧敏、格豪及振暘，另外 3 位為非家庭成員，分別為舒涵、裕宏及維峻，訓練部份如同 5.3.1 節所描述，每位家庭成員取 9 張作訓練資料，剩下的 36

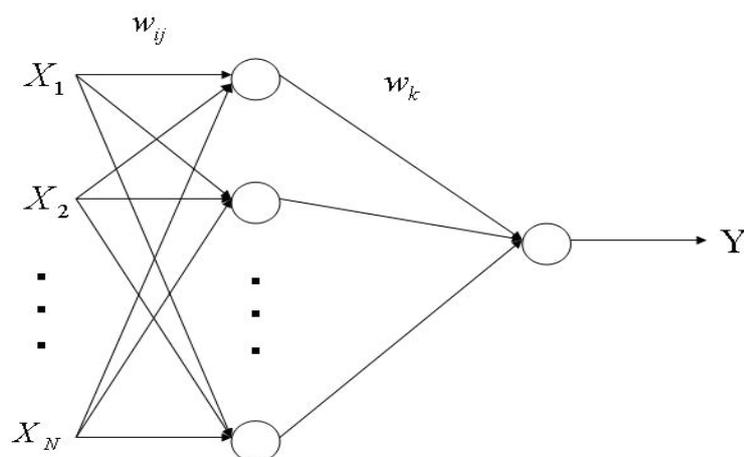


圖 5-13 每位家庭成員的類神經網路架構

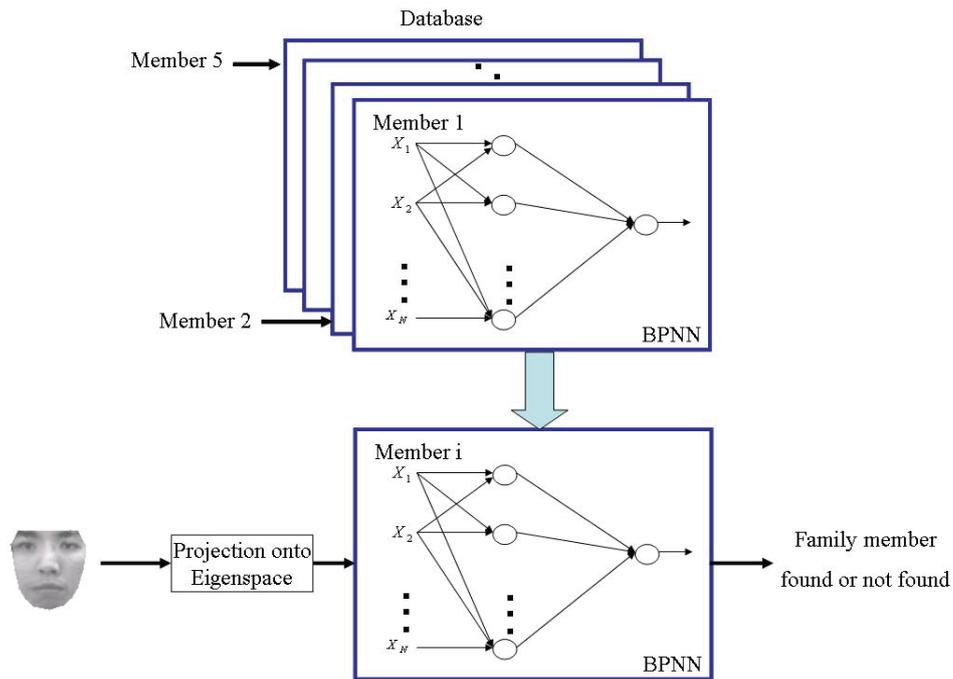


圖 5-14 類神經網路辨識流程圖

張為測試資料，同樣的，每位非家庭成員的測試資料也為 36 張，包含多姿態的人臉影像。在每位家庭成員的類神經網路中，我們所設的隱藏層神經元數目為 75 個，迭代次數為 5000 次，圖 5-15 為每位家庭成員在建立自己的類神經網路的收斂情形，從圖中可看出在每人各自所訓練的類神經網路裡，均還是可以達到收斂。表 5-3 為家庭成員與非家庭成員之人臉辨識結果，其整體平均正確率為 88.19%，可看出其效果並不太好，其原因為非家庭成員的不確定性很大，對於未訓練過的非家庭成員資料作辨識時，無法準確的定出類神經網路的輸出門檻值，以致於非家庭成員辨識不佳，此部份也為本論文未來需要加強的地方。

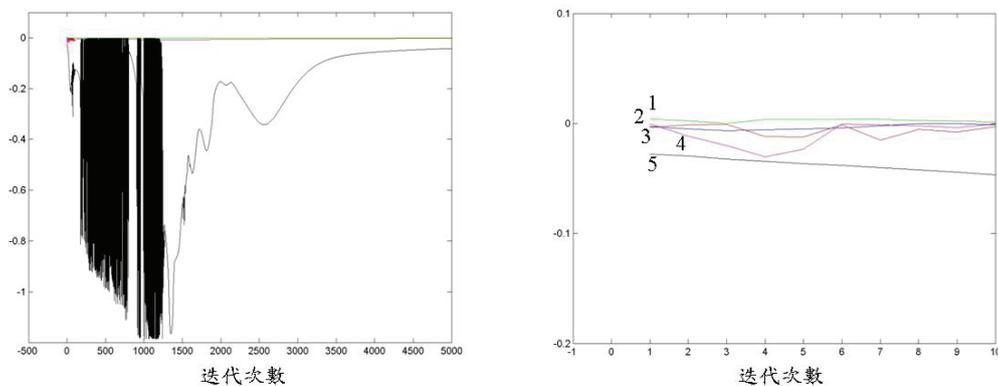


圖 5-15 每位家庭成員經類神經網路訓練之 (a)收斂情形 (b)初始狀態

表 5-3 家庭成員與非家庭成員之多姿態人臉辨識結果

Output Input	仕傑	孟儒	巧敏	格豪	振暘	非家庭成員	正確率
仕傑	35	0	0	0	0	1	97.22%
孟儒	0	36	0	0	0	0	100%
巧敏	0	0	32	0	0	4	88.89%
格豪	0	0	0	31	0	5	86.11%
振暘	0	0	0	0	31	5	86.11%
舒涵	0	0	4	3	1	28	77.78%
裕宏	0	1	2	3	0	30	83.33%
維峻	3	0	2	0	0	31	86.11%

## 5.5 結論與討論

在本章的實驗中，其經由辨識的部份，主要分成兩種情況來做，一種為家庭成員多姿態人臉辨識，利用實驗室內建以及 UMIST 的多姿態人臉資料庫其平均辨識率各別可達 95.56% 以及 91%，驗證人臉在多姿態下均還是能有效的判斷使用者為哪一位家庭成員；而另一種情況則是考慮若有非家庭成員進入系統辨識時，系統能否判斷使用者為非家庭成員。由表 5-3 可看出在非家庭成員的辨識中，其效果並不佳，就目前的可能原因而言，一為雖然類神經網路有經過資料庫成員的訓練，但是對於非資料庫成員卻無比對樣本，以致於在類神經網路架構的辨識上容易誤判；二為在本章實驗中，由於人臉資料樣本數取的不夠多，在做類神經網路訓練時，其訓練的樣本數只有 9 筆的人臉資料，可能由於訓練樣本過少，而導致此類神經網路無法發揮其作用；三為在類神經網路的架構中，我們是取最基本的架構來做分析，也就是此類神經網路架構的隱藏層僅只有一層，導致在空間的類別切割時過於線性化，降低了系統的正確率。因此在家庭成員與非家庭成員的辨識的部份是之後需

要進一步來作加強，提升系統的辨識率。

就本實驗所使用的嵌入式影像平台而言，從經由 CMOS sensor 取像後，經過人臉偵測、特徵擷取以及人臉辨識，在運算的速度上，每張影像的運算時間 1.5 秒，也就是每三秒鐘可分析兩張影像。由於 AAM 演算法中，其運算速度是與影像大小以及迭代次數成反比，因此在本系統的演算法上，我們將人臉影像正規化成 120x100 的大小作分析，另外在迭代次數上，我們固定為 16 次，如此才可讓系統分析影像的時間是固定的，而讓寵物機器人與使用者進行互動。



## 第六章 結論與未來展望

### 6.1 結論

本論文提出了多姿態人臉辨識系統架構，並應用於實驗室所研製的寵物機器人上，利用實驗室發展的嵌入式影像平台擷取人臉影像，辨識出使用者人臉為家庭之中的某一位家庭成員。

在人臉影像特徵擷取方面，經由膚色和人臉特徵完成人臉偵測，並框選出人臉位置，再對框選到的人臉做影像大小的正規化(大小為 120x100)以及照度上的補償，接著利用 PCA 的方法找出 AAM 的形狀模型與紋理模型，再將模型套入經改良的 Lucas-Kanade 影像校正演算法迭代出人臉的形狀，以得到人臉的影像資訊。

在辨識部份，將得到的人臉影像資訊利用 PCA 的方法投影到特徵空間並讓維度化減，接著利用倒傳遞類神經網路進行家庭成員的比對，在資料庫的建立下，以手繪的方式標定出人臉的形狀特徵點，並利用形變以及特徵空間投影的方法將資料庫人臉的影像資訊大小的維度與系統輸出的維度相同，才可做與資料庫上的比對。在實驗室建立的與 UMIST 資料庫建立的人臉影像中，經由初步的實驗結果，其辨識率可達到 95.56%以及 91%，並驗證了本系統在多姿態情況下其辨識率可達九成以上，而在家庭成員與非家庭成員的多姿態人臉影像辨識上，其平均辨識率只有 88.19%，在方法上還需做補強，例如類神經網路的非資料庫樣本、訓練樣本數目以及網路的層數。

### 6.2 未來展望

目前本系統仍有許多可改善的空間：

#### 1. 在特徵擷取方面

在利用影像校正迭代法追蹤人臉時，若人臉光源變化較為劇烈時，則容易讓人臉的形狀特徵點位置移動到錯誤的地方，甚至是人臉以外，且因

光源變化的關係，而讓系統誤以為特徵點已收斂，影響後端的辨識。可以將影像做具有抗光源變化的 Gabor Wavelet 轉換，再去做影像迭代的動作。本系統在人臉左右偏轉角度的限制為 30 度以內，若人臉偏轉超過 30 度時，則必須再重新找出角度偏轉過多的人臉的形狀模型與紋理模型，且特徵點需重新定義，若是利用 2D AAM 與 3D 的做結合，則系統可以知道使用者目前的偏轉角度，並去做迭代校正。

## 2. 在人臉辨識方面

雖然利用 PCA 的方法能夠進行維度的縮減，但相對的也遺失了少許資訊，造成在辨識上的誤差。可以試著找出不需進行化減亦可不會讓辨識速度降低太多的方法。同理，在做人臉辨識的比對時，若資料庫成員變多時，相對的比對的速度就會降低。是否能先從資料庫找出各個家庭成員的關鍵性特徵部份，若能找出每個成員的關鍵性特徵，再對此關鍵性特徵做維度化減，則當進行成員比對時，可以只做關鍵性特徵上的比對即可，讓系統的速度能再加快；另外在辨識人臉時，當受測者為非資料庫之成員時，如何準確的判別受測者為非家庭成員為我們未來的重要議題。

## 參考文獻

- [1] [http://www.robotworld.org.tw/index.htm?pid=10&News\\_ID=342/](http://www.robotworld.org.tw/index.htm?pid=10&News_ID=342/)
- [2] [http://www.nec.co.jp/robot/english/intro/intro\\_01.html](http://www.nec.co.jp/robot/english/intro/intro_01.html)
- [3] <http://www.stcsm.gov.cn/detail/detail.aspx?tbno=202&id=34537>
- [4] <http://www.mhi.co.jp/kobe/wakamaru/english/live/index.html>
- [5] <http://www.pitotech.com.tw/c/6-product06-2.htm>
- [6] M. A. Turk and A. P. Pentland, "Face Recognition using Eigenfaces," *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, USA, 1991, pp. 586-591.
- [7] P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using Class Specific Linear Projection," *Proc. of IEEE Transactions on Pattern Analysis and Machine Intelligence*, USA, vol. 19, no. 7, 1997, pp. 711-720.
- [8] W. Chen, M. J. Er and S. Wu, "Illumination Compensation and Normalization for Robust Face Recognition using Discrete Cosine Transform in Logarithm Domain," *Proc. of IEEE Transactions on Systems, Man, and Cybernetics*, Singapore, vol. 36, no. 2, 2006, pp.458-466.
- [9] L. Wiskott, J. M. Fellous, N. Kruger and V. D. Malsburg, "Face Recognition by Elastic Bunch Graph Matching," *Proc. of IEEE Transactions on Pattern Analysis and Machine Intelligence*, Germany, 1997, pp. 775-779.
- [10] H. Wang, S. Li and Y. Wang, "Face Recognition under Varying Lighting Conditions using Self Quotient Image," *Proc. of IEEE International Conference on Automatic Face and Gesture Recognition*, China, 2004, pp. 819-824.
- [11] D. V. Jadhav and R. S. Holambe, "Multiresolution based Kernel Fisher Discriminant Model for Face Recognition," *Proc. of IEEE International Conference on Information Technology*, India, 2007, pp. 848-853.
- [12] A. T. Gualberto, S. P. Gabriel, N. M. Mariko and P. M. Hector, "Face Recognition Algorithm using The Discrete Gabor Transform," *Proc. of IEEE International Conference on Electronics, Communications and Computers*, Mexico, 2007, pp. 35-40.

- [13]P. Jia and H. Hu, "AAM Based HCI for an Intelligent Wheelchair," *Proceedings of the 1st International Symposium on Multispectral Image Processing and Pattern Recognition*, China, 2007.
- [14]J. Xiao, S. Baker, I. Matthews and T. Kanade, "Real-Time Combined 2D+3D Active Appearance Models," *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, USA, 2004, pp 535-542.
- [15]R. Gross, S. Baker and I. Matthews, "Constructing and Fitting Active Appearance Models with Occlusion," *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition Workshop*, USA, 2004, pp. 72-80.
- [16]Y. H. Liu and Y. T. Chen, "Face Recognition Using Total Margin-Based Adaptive Fuzzy Support Vector Machines", *Proc. of IEEE Transactions on Neural Networks*, Taiwan, vol. 18, no. 1, 2007, pp. 178-192.
- [17]T. Ahonen, A. Hadid, M. Pietikainen, "Face Description with Local Binary Patterns: application to Face Recognition", *Proc. of IEEE Transactions on Pattern Analysis and Machine Intelligence*, Finland, 2006, pp 2037-2041.
- [18]陳文俊, 人臉辨識及其在家用機器人與人互動之應用, 碩士論文, 國立交通大學電機與控制工程學系, 2002.
- [19]汪慶祥, 利用三維資訊做人臉辨識, 碩士論文, 國立交通大學資訊科學系, 2000.
- [20]林晁立, 以臉部器官形狀, 寬度, 相對位置從事人臉影像辨識, 碩士論文, 私立東海大學資訊科學系, 2000.
- [21]陳嘉雄, 多角度人臉之二階段辨識, 碩士論文, 國立東華大學資訊工程系, 2002.
- [22]梁皓雲, 利用區塊人臉特徵為基礎之混合式人臉辨識系統, 碩士論文, 國立中央大學資訊工程系, 2006.
- [23]周崇民, 光源變化下之即時人臉追蹤, 碩士論文, 國立交通大學電機與控制工程學系, 2005.
- [24]P. Viola and M. Jones, "Rapid Object Detection Using a Boosted Cascade of Simple Features," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2001.
- [25]T. F. Cootes and C. J. Taylor, Technical Report: Statistical Models of Appearance for Computer Vision, The University of Manchester School of Medicine, 2004.

- [26]<http://www2.imm.dtu.dk/~aam/>
- [27]T. F. Cootes, C. J. Taylor, D. H. Cooper and J. Graham, “Active Shape Models-Their Training and Application,” *Computer Vision and Image Understanding*, 1995, vol. 61, no.1, pp.38-59.
- [28]J.R. Shewchuk, Triangle: engineering a 2D quality mesh generator and Delaunay triangulator, *In Applied Computational Geometry, FCRC'96 Workshop*, pp. 203-222. Springer-Verlag, 1996.
- [29]T. F. Cootes, G. J. Edwards and C. J. Taylor, “Active Appearance Models,” *Proc. of IEEE Transactions on Pattern Analysis and Machine Intelligence*, UK, 2001, pp. 681-685.
- [30]I. Matthews and S. Baker, “Active Appearance Models Revisited,” *International Journal of Computer Vision*, vol. 60, no.2, 2004, pp.135-164.
- [31]S. Baker, F. Dellaert, and I. Matthews, “Aligning Images Incrementally Backwards,” Technical Report CMU-RI-TR-01-03, Robotics Institute, Carnegie Mellon University, 2001.
- [32] S. Baker and I. Matthews, “Lucas-Kanade 20 Years On: A Unifying Framework: Part 1,” Technical Report CMU-RI-TR-02-16, Robotics Institute, Carnegie Mellon University, 2002.
- [33]P. Jia, “Active Appearance Model,” Online technical report available from [http://www.visionopen.com/members/jiapei/pei\\_aamreport2006.pdf](http://www.visionopen.com/members/jiapei/pei_aamreport2006.pdf).
- [34] S. Baker, R. Gross, I. Matthews and T. Ishikawa, “Lucas-Kanade 20 Years On: A Unifying Framework: Part 2,” Technical Report CMU-RI-TR-03-01, Robotics Institute, Carnegie Mellon University, 2003.
- [35]S. Baker, R. Gross and I. Matthews, “Lucas-Kanade 20 Years On: A Unifying Framework: Part 3,” Technical Report CMU-RI-TR-03-35, Robotics Institute, Carnegie Mellon University, 2003.
- [36]S.J. Leon, *Linear Algebra with Application*, Prentice Hall, 2002.
- [37]H. Gupta, A. K. Agrawal, T. Pruthi, C. Shekhar and R. Chellappa, “An Experimental Evaluation of Linear and Kernel-Based Methods for Face Recognition”, *Proc. of IEEE Workshop on Applications and Computer Vision*, USA, 2002, pp 13-18.

[38]A. Henry and K. T. Song, “Embedded CMOS Imaging System for Real-Time Robotic Vision”, *Proc. of IEEE International Conference on Intelligent Robots and Systems, Edmonton, Alberta, Canada, 2005*, pp.3694-3699..

[39][images.ee.umist.ac.uk/danny/database.html](http://images.ee.umist.ac.uk/danny/database.html).

