

國立交通大學

電控工程研究所

碩士論文

具表情混合比例及強度之人臉表情辨識系統
Facial Expression Recognition with Mixture Ratio of
Basic Expressions and Expression Intensity Estimation

The logo of National Tsing Hua University is a circular emblem. It features a central shield with a book and a torch, surrounded by a gear-like border. The year '1896' is inscribed at the bottom of the shield.

研究生：簡碩成

指導教授：宋開泰 博士

中華民國一百年七月二十八日

具表情混合比例及強度之人臉表情辨識系統

Facial Expression Recognition with Mixture Ratio of
Basic Expressions and Expression Intensity Estimation

研究生：簡碩成

Student: Shuo-Cheng Chien

指導教授：宋開泰 博士

Advisor: Dr. Kai-Tai Song



電控工程研究所

碩士論文

A Thesis

Submitted to Institute of Electrical 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 Control Engineering

July 2011

Hsinchu, Taiwan, Republic of China

中華民國一百年七月二十八日

具表情混合比例及強度之人臉表情辨識系統

學生:簡碩成

指導教授:宋開泰 博士

國立交通大學電控工程研究所

摘要

本論文發展出一套可用來辨識出人臉表情中各基本表情所占比例以及人臉表情強度之人臉表情辨識系統。首先以主動外觀模型(Active Appearance Model, AAM)訓練出人臉影像的形狀及紋理模型,並採用改良式 Lucas-Kanade 影像校正進行輸入影像校正;然後利用之前訓練出的 AAM 紋理模型取得紋理特徵參數。本論文所使用的紋理特徵參數分為全臉、上半臉、下半臉三種,分別作為強度辨識、上半臉 Action Units (AU)組合辨識、下半臉 AU 組合辨識各自的輸入。強度辨識、上半臉 AU 組合辨識、下半臉 AU 組合辨識都使用倒傳遞類神經網路(Back propagation Neural Network)給出評價分數;接著我們提出一套整合設計可用這些評價分數得到表情強度及表情比例。

Facial Expression Recognition with Mixture Ratio of Basic Expressions and Expression Intensity Estimation

Student: Shuo-Cheng Chien

Advisor: Dr. Kai-Tai Song

Institute of Electrical Control Engineering
National Chiao Tung University

ABSTRACT

In this thesis, a facial expression recognition system which can recognize facial expressions as well as the expression intensity and mixture ratio of basic expressions is developed. Active Appearance Model (AAM) is used to train shape model and texture model. The improved Lucas-Kanade image alignment algorithm is then applied to align the input images to obtain texture features. A novel method is proposed to recognize ratio of basic expressions and intensity of facial expression. Three kinds of texture features are used in this method: 1. texture features of whole face, which are used as inputs of facial expression intensity recognition, 2. texture features of upside face, which are used as inputs of upper face action units recognition, 3. texture features of downside face, which are used as the inputs of lower face action units recognition. Back propagation neural networks are used to obtain the recognition scores, which are then exploited to classify the facial expression results, including basic facial expression ratio and the facial expression intensity.

誌謝

在此謹向我的指導教授宋開泰博士致上誠摯感謝之意，感謝他兩年來對我在課業以及研究上的指導。在論文题目的選定、系統架構的提出以及實驗方法的完善方面都以豐富的學識及經驗加以指導，使得本論文得以順利完成。

感謝實驗室學長姐孟儒、嘉豪、格豪、巧敏、信毅、兆平及允智，在我研究上有疑惑時給予詳盡的建議。感謝與我在這兩年來一起努力的同學仕晟、建宏與上峻的鼓勵及幫忙，以及學弟家昌、章宏不時提供的幫助。還有要特別致謝 Carnegie Mellon University 提供的 Cohn-Kanade AU-Coded Facial Expression Database，它豐富的資料幫助我驗證演算法的正確性。

最後，我要感謝我的父母，因為他們的辛苦栽培以及細心關懷，使得我能夠以健康的身心持續研究並得以順利完成論文，在此願將此論文獻給我敬愛的父母。



目錄

摘要.....	i
ABSTRACT.....	ii
誌謝.....	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.....	3
1.3. 問題描述.....	16
1.4. 系統架構與章節說明.....	17
第二章、人臉影像特徵擷取方法.....	20
2.1. 人臉偵測演算法.....	20
2.1.1. 人臉區域篩選.....	21
2.1.2. 人臉偵測範例.....	22
2.2. 主動外觀模型.....	23
2.3. 人臉形狀模型.....	23
2.3.1. 標定特徵點.....	23
2.3.2. 平均人臉形狀.....	25
2.3.3. 模擬人臉形狀變化.....	27
2.4. 人臉紋理模型.....	27
2.4.1. 分段仿射變形.....	27
2.4.2. 模擬人臉紋理變化.....	29
2.5. 影像校正方法.....	29
2.5.1. Inverse Compositional 演算法.....	29
2.5.2. 總體形狀正規化轉換.....	31
2.5.3. 影像修正.....	32
2.5.4. 影像校正整體流程.....	33
第三章、人臉表情混合比例及強度辨識方法.....	35
3.1. 人臉表情混合比例及強度辨識架構.....	35
3.2. 人臉紋理特徵參數.....	36

3.3. 倒傳遞類神經網路	37
3.4. 面部動作編碼組合偵測	38
3.4.1. 面部動作編碼組合	39
3.4.2. 面部動作編碼組合與表情關係	40
3.4.3. 面部動作編碼組合評分	41
3.5. 人臉表情混合比例辨識	41
3.5.1. 全臉評分	43
3.5.2. 上半臉評分	44
3.5.3. 下半臉評分	44
3.5.4. 整合評分	45
3.6. 人臉表情強度辨識	46
第四章、實驗結果	48
4.1. 整體訓練流程	48
4.2. 類神經網路訓練	49
4.3. 基本表情辨識結果	51
4.4. 表情強度辨識結果	56
4.5. 混合表情比例辨識結果	58
4.6. 結論與討論	67
第五章、結論與未來展望	68
5.1. 結論	68
5.2. 未來展望	68
參考文獻	69

圖例

圖 1-1、驚喜表情	2
圖 1-2、人臉表情辨識一般主要流程	4
圖 1-3、表情連續圖片	4
圖 1-4、考慮時間作為辨識設計之系統架構[3].....	5
圖 1-5、Temporal Patterns Model.....	5
圖 1-6、不同背景與光照下的取像	6
圖 1-7、不同人臉角度下的取像	7
圖 1-8、AAM 標點範例[6]	8
圖 1-9、自動標注 AAM 特徵點之系統架構[7]	8
圖 1-10、利用特徵點中的幾何關係取特徵點範例[8].....	9
圖 1-11、利用 subclass divisions 分類取特徵點範例[9].....	9
圖 1-12、數種 AU 的示意圖	10
圖 1-13、區域分割以及擷取特徵點範例[11].....	11
圖 1-14、FACS 中定義的六種基礎表情	11
圖 1-15、[15]中 unsupervised learning 之例子	12
圖 1-16、SVM 分類範例[15]	13
圖 1-17、HMM 模型分類範例[17].....	13
圖 1-18、使用 Partially-Observed Hidden CRF 的模型範例[14].....	14
圖 1-19、辨識表情分類範例[19].....	14
圖 1-20、表情辨識架構圖[19].....	15
圖 1-21、六種表情之光流場向量[20].....	15
圖 1-22、使用二元分類樹分類範例[21].....	16
圖 1-23、整體系統架構	19
圖 2-1、Haar-like 特徵法	20
圖 2-2、在不同距離及不同人臉角度下的人臉偵測	22
圖 2-3、兩人在不同距離下的人臉偵測	23
圖 2-4、特徵點選取原則[26].....	24
圖 2-5、特徵點標點範例[26].....	24
圖 2-6、不同表情下特徵點標點範例	25
圖 2-7、人臉形狀對齊範例	26
圖 2-8、平均形狀(Mean shape)範例.....	26
圖 2-9、人臉形狀用 Delaunay triangulation 表示	28
圖 2-10、三角形的對應轉換	28
圖 2-11、Inverse Compositional 演算法架構.....	30

圖 2-12、總體形狀正規化轉換	32
圖 2-13、樣板人臉影像	33
圖 3-1、人臉表情混合比例及強度辨識架構(FER-RIE)	35
圖 3-2、上半邊人臉	36
圖 3-3、下半邊人臉	37
圖 3-4、完整人臉	37
圖 3-5、上半邊面部動作編碼組合(AUC_up)	39
圖 3-6、下半邊面部動作編碼組合(AUC_down)	40
圖 3-7、人臉表情混合比例辨識架構	42
圖 3-8、人臉強烈表情範例(Happy, Disgust)	46
圖 3-9、表情強度辨識架構	47
圖 4-1、CK+ 訓練圖像範例	49
圖 4-2、上面部動作編碼組合類神經網路	50
圖 4-3、下面部動作編碼組合類神經網路	50
圖 4-4、表情強度類神經網路	51
圖 4-5、CK+ 測試圖像範例	52
圖 4-6、混合表情測試圖像(1)	59
圖 4-7、混合表情測試圖像(2)	60
圖 4-8、混合表情測試圖像(3)	61
圖 4-9、混合表情測試圖像(4)	62
圖 4-10、混合表情測試圖像(5)	63
圖 4-11、混合表情測試圖像(6)	64
圖 4-12、混合表情測試圖像(7)	65
圖 4-13、混合表情測試圖像(8)	66

表格

表 1-1、內政部戶政司的人口資料	1
表 3-1、上半邊面部動作編碼組合與表情關聯性	40
表 3-2、下半邊面部動作編碼組合與表情關聯性	41
表 3-3、上下面部動作編碼組合搭配與表情關聯性	43
表 3-4、表情搭配檢查表	45
表 4-1、基本表情人臉影像表情辨識結果(Total)	52
表 4-2、基本表情人臉影像表情辨識結果(Angry, Disgust)	53
表 4-3、基本表情人臉影像表情辨識結果(Fear, Happy)	54
表 4-4、基本表情人臉影像表情辨識結果(Sadness, Surprise)	55
表 4-5、基本表情人臉影像表情辨識結果[34]	56
表 4-6、人臉表情強度辨識結果	57
表 4-7、人臉表情強度辨識與問卷比較結果(寬鬆)	58
表 4-8、人臉表情強度辨識與問卷比較結果(嚴格)	58
表 4-9、混合表情測試辨識輸出結果(1)	59
表 4-10、混合表情測試問卷調查結果(1)	59
表 4-11、混合表情測試辨識輸出結果(2)	60
表 4-12、混合表情測試問卷調查結果(2)	60
表 4-13、混合表情測試辨識輸出結果(3)	61
表 4-14、混合表情測試問卷調查結果(3)	61
表 4-15、混合表情測試辨識輸出結果(4)	62
表 4-16、混合表情測試問卷調查結果(4)	62
表 4-17、混合表情測試辨識輸出結果(5)	63
表 4-18、混合表情測試問卷調查結果(5)	63
表 4-19、混合表情測試辨識輸出結果(6)	64
表 4-20、混合表情測試問卷調查結果(6)	64
表 4-21、混合表情測試辨識輸出結果(7)	65
表 4-22、混合表情測試問卷調查結果(7)	65
表 4-23、混合表情測試辨識輸出結果(8)	66
表 4-24、混合表情測試問卷調查結果(8)	66

第一章 緒論

1.1 研究動機

有許多評論認為服務型機器人在不久的將來會走入每個家庭，其普及率會像是今日的家電一般，理由是社會高齡化後家庭的人力會不足導致於會無人照顧年長者與小孩。如表 1-1 所示(內政部戶政司的資料)的年齡結構百分比跟老化指數可以知道台灣的人口正不停老化中[1]。現在的對應做法是從國外引進人力(菲籍外傭、越籍外傭.....等等)，但是這種作法有其缺點(每月必須負擔大筆的傭人薪水、傭人的品質不穩定時有欺凌年長者或小孩的案例)，也有其極限(不可能無限的引進)。因此，就如前文所敘述的服務型機器人在不久的將來會走入每個家庭，如何讓服務型機器人更加完善就成為需要研究的課題。其中，如何讓服務型機器人辨認出家庭成員的情緒就是一個非常重要的課題。

表 1-1、內政部戶政司的人口資料[1]

年別 (民國)	百分比分配			老年 人口 依賴比	老化 指數	扶養比
	0~14 歲	15~64 歲	65 歲以上			
85 年	23.15	68.99	7.86	11.39	33.95	45
86 年	22.60	69.34	8.06	11.62	35.65	44
87 年	21.96	69.79	8.26	11.83	37.59	43
88 年	21.43	70.13	8.44	12.04	39.40	43
89 年	21.11	70.26	8.62	12.27	40.85	42
90 年	20.81	70.39	8.81	12.51	42.33	42
91 年	20.42	70.56	9.02	12.78	44.17	42
92 年	19.83	70.94	9.24	13.02	46.58	41
93 年	19.30	71.20	9.50	13.30	49.00	40
94 年	18.70	71.60	9.70	13.60	52.00	40
95 年	18.12	71.88	10.00	13.91	55.17	39
96 年	17.56	72.24	10.21	14.13	58.13	38
97 年	16.95	72.62	10.43	14.36	61.51	38
98 年	16.34	73.03	10.63	14.56	65.05	37
99 年	15.65	73.61	10.74	14.59	68.64	36



圖 1-1、驚喜表情

如果機器人可辨認出家庭成員的情緒，就有可能依照每個人的情緒給予不同的服務以及互動。可設想的情境有：例如使用者心情愉快時，機器人也跟著有愉快的表情；又例如使用者心情不好時，機器人能夠加以安慰。若機器人認得每位家庭成員的情緒使得機器人與人類的互動將可依照情緒的不同而有不同的回應。這將使機器人與人類的互動變得更為自然，例如裝有語音系統的機器人可以依照辨識出的情緒轉變使用的語氣。在目前市面上也看得到一些寵物機器人在販售，若寵物機器人能夠很精確的估計主人的情緒，那麼就能夠依照主人的情緒做出更人性化的反應。

在現行的研究中，人臉表情辨識多是將偵測到的表情加以分類成六類基本表情，然而卻較少研究著墨於辨認出其他表情以及不同的表情強度，例如收到令人意外的好禮物時表情可能是又驚又喜如圖 1-1 所示，這時就不能只以單一的驚或者喜加以分類。又例如微笑與大笑所對應的情緒強度並不相同，若只將其分類到同一類就不足以描述完整的表情。這種只將人臉分類成六類基本表情的作法會使得機器人對情緒的互動反應也侷限於六種，這將會限制住機器人互動反應的人性

化程度。因此，若希望機器人能夠對情緒有更人性化以及更複雜的反應，更精確的辨識出人臉的表情就是必須要達成的功能。但要如何更精確的辨識人臉表情？首先考慮如何辨識出 6 類基本表情之外的人臉表情，由於我們可將許多表情看成 6 類基本表情的混合例如：驚喜、既傷心又害怕、既厭惡又憤怒……等等，我們可以設法辨識出人臉表情中六種基本表情所占比例以辨識出這些情況。接著我們考慮如微笑、大笑、微怒、大怒……等等情況，這樣的情況下所需要的是表情強度辨識。因此，我們認為要做到更精確的辨識人臉表情需要的是辨識出人臉表情混合比例以及表情強度。

另一方面，參考韓孟儒等人所提出的人臉表情模擬系統[2]，此系統能夠以調整六種基本表情參數的方式模擬出人臉表情。若搭配如[2]這種人臉表情模擬系統，那麼整套系統便能辨識出互動者的人臉表情之後使機器人作出相對應的表情模擬，對於機器人與人類的互動變得更為自然的可能性又更進了一步。

綜合以上所述，服務型機器人在不久的將來會普及化。而精確的辨識人臉表情是使服務型機器人能夠提供更加人性化的服務以及能夠與使用者更加生動的互動的一個非常重要的基本要素。對此我們將設法辨識出人臉表情混合比例以及表情強度以解決此問題。

1.2 相關研究回顧

人臉表情辨識是相當實用的研究課題，也有相當多的人研究過這個課題，由此而衍生出了各式各樣的演算法。雖然演算法數量很多但是它們的主要流程大部分可以大略畫成圖 1-2。各式研究中所使用的影像輸入資料各不相同，常見的是使用單張表情圖片(Image-based)，另外也有一次使用多張連續圖片的例子(Image Sequence)[3]。圖 1-3 為表情從底部(Onset)到頂點(Apex)的示意圖。頂點過後會再回到底部(Offset)，因可視為從底部到頂點的反過程因此常不被考慮。一次使用多張連續圖片作為輸入能夠使用眾多連續人臉表情變化的圖片以擷取比較全

面的資訊。也就是說使用多張連續圖片作為輸入能夠擷取到表情變化的過程，而非只考慮到頂點的部分。使用連續人臉表情變化的圖片作為輸入需要考慮到時間的問題，也就是時間作為一個變數項在演算法模型中為重要參考對象。圖 1-4 是考慮時間作為辨識設計之系統架構的例子，而圖 1-5 所示 Temporal Patterns Model 便是將時間順序考慮進去的分類模型。此模型用 Haar-like 特徵法取出時序中每張的特徵值再加總計算，由於輸入是從底部到頂點的數張圖像，比較一張圖像能夠取得更多特徵而增加辨識率。

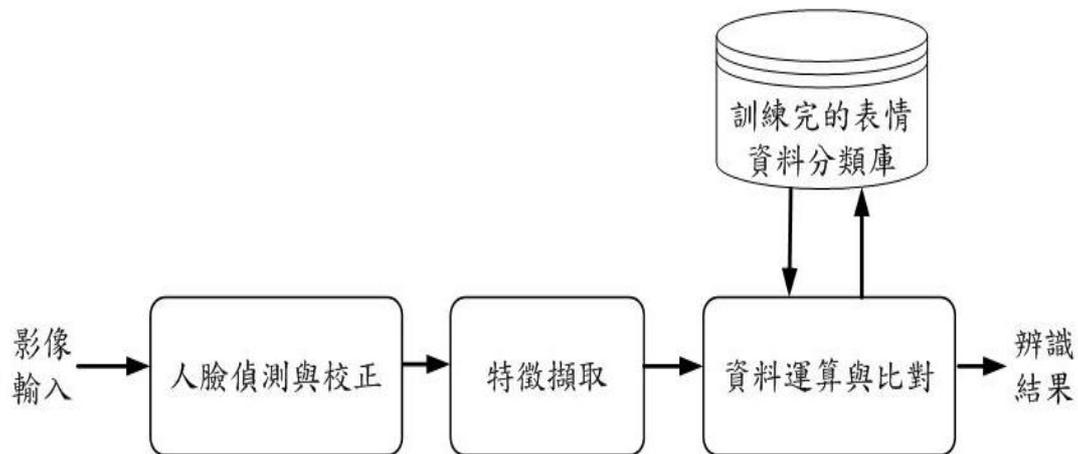


圖 1-2、人臉表情辨識一般主要流程

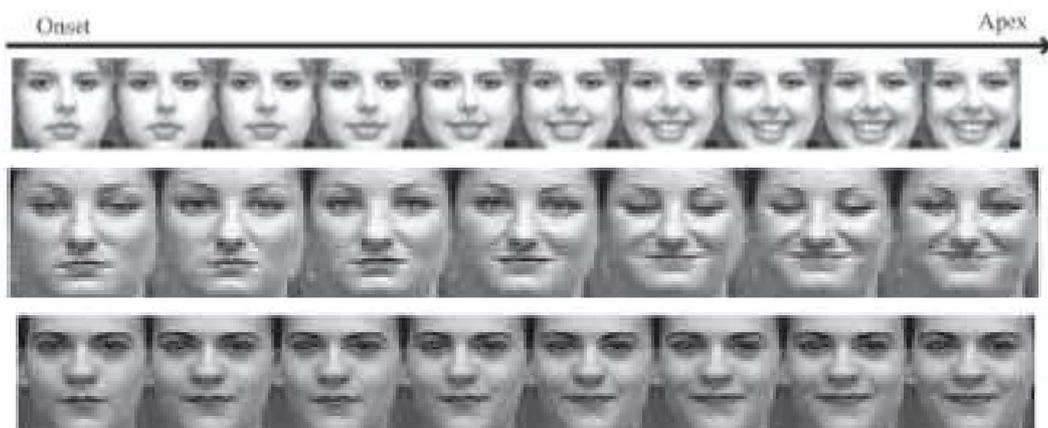


圖 1-3、表情連續圖片[3]

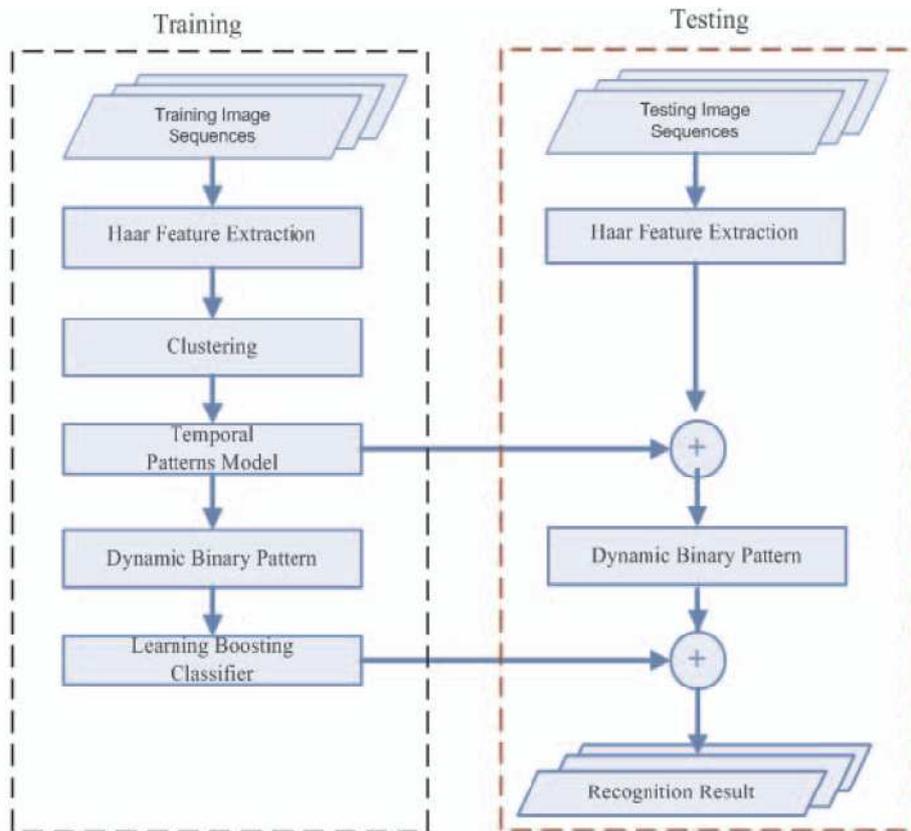


圖 1-4、考慮時間作為辨識設計之系統架構[3]

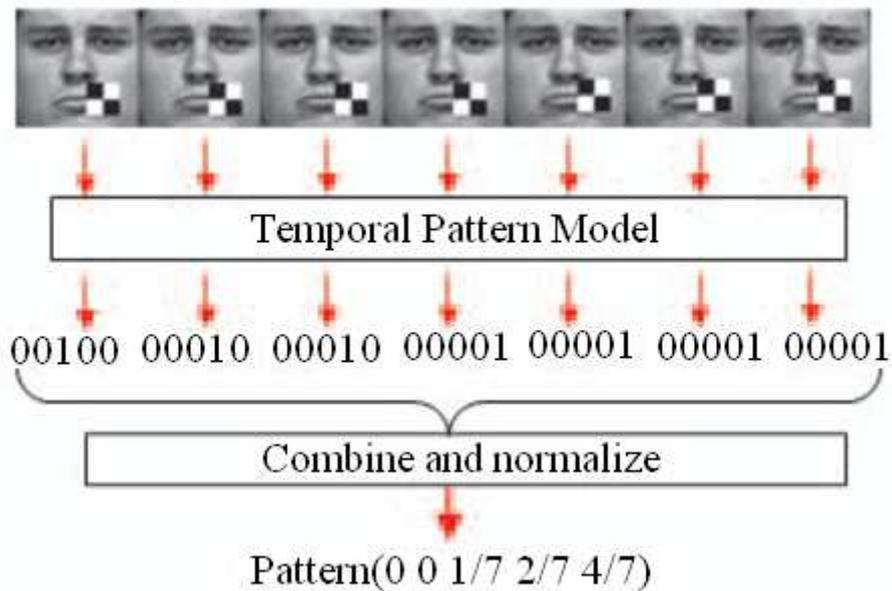


圖 1-5、Temporal Patterns Model[3]

人臉表情辨識在一開始的階段是人臉偵測，其主要的挑戰有下列幾種：1.當時的光照的亮度與角度不同 2.人臉的角度不同 3.人臉的表情不同 4.臉部特徵的改變(例如戴上不同的眼鏡、戴上口罩.....等等)5.複雜的背景。

為了找出人臉的位置，相當廣為採用的方法是特徵比對法，其中常被提到的有 Haar-like feature[4]。除了臉部特徵的改變太大(例如帶上太陽眼鏡、口罩)會使得人臉特徵受到遮擋，很有可能無法認出人臉位置以外，都有不錯的偵測表現。下面的圖像是一些例子[4]，分別為不同背景與光照下的取像以及不同人臉角度下的取像。其中圖 1-6 為不同背景與光照下的取像，圖 1-7 則是不同人臉角度下的取像。

回歸前面所述的基本流程，在人臉偵測階段，將偵測出人的臉孔並將其與背景分割開來，在這個階段許多研究常會加入人臉追蹤的功能[4]，對當下之畫面取像完畢後需要作一些處理，也就是把前面人臉偵測所得到的人臉資料做進一步的處理，通常會先做校正(例如：角度校正、光照校正)，接著再做下一步處理(例如特徵點擷取)。

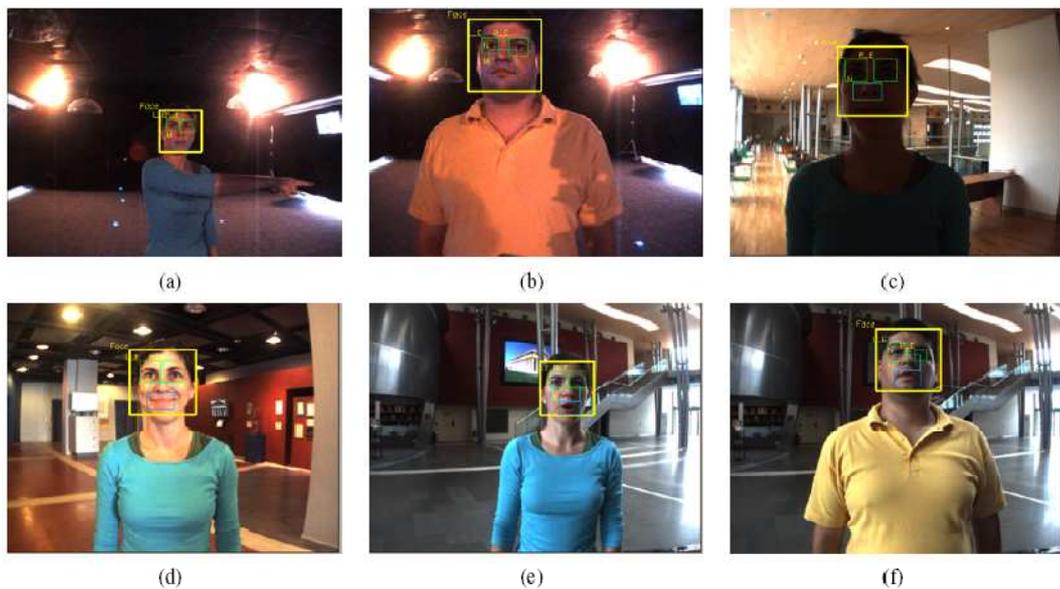


圖 1-6、不同背景與光照下的取像[4]



圖 1-7、不同人臉角度下的取像[4]

在特徵擷取階段，可以將特徵擷取的取向分為整體特徵及局部特徵[5]。整體特徵取整張人臉作為單一特徵；局部特徵法先找出人臉的一些局部特徵，例如眼睛、鼻子和嘴巴，之後將個別局部特徵的結果統整。若再細分還可以分類成顏色特徵、紋理特徵、形狀特徵、空間關係特徵.....等等不同的分類。在[5]中所使用的辨識特徵就是人臉的形狀特徵以及紋理特徵。

特徵點的選擇會因應用情況而有不同，人臉辨識以及表情辨識研究中常被使用的 Active Appearance Model (AAM) 模型的定義裡[6]，特徵點位置為眼睛、鼻子、嘴巴、眉毛以及臉型之輪廓。其特徵點標點的範例如圖 1-8 所示：一般 AAM 特徵標點必須由人手動標點後才能夠訓練系統，所以在更新資料庫的時候相當的麻煩，為了解決此問題有不少研究提出了不同的方法，例如在[7]提出了自動標注 AAM 特徵點的方法。圖 1-9 為[7]之架構圖，它利用已經標注好的圖像，經過此架構的處理可以讓標注特徵點的圖像自動標注 AAM 特徵點。

也有研究是利用特徵點中的幾何關係以及 Support Vector Machine (SVM) 與 Markov random field 自動去找特徵點，圖 1-10 為其範例[8]；另一個研究[9]中則是利用 subclass divisions 分類的概念去自動找到特徵點，圖 1-11 為其範例。

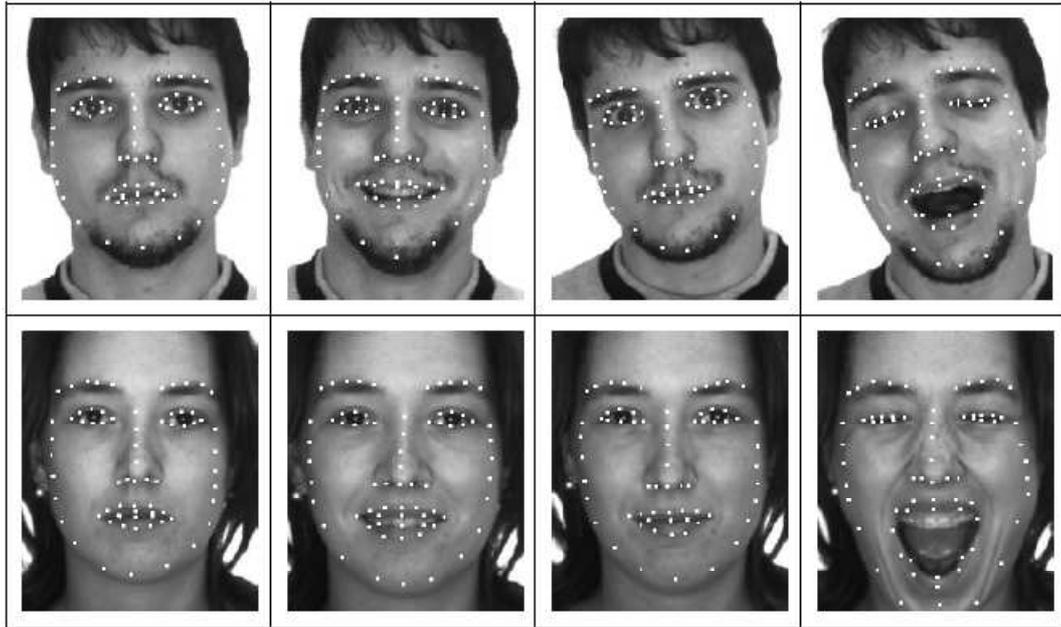


圖 1-8、AAM 標點範例[6]

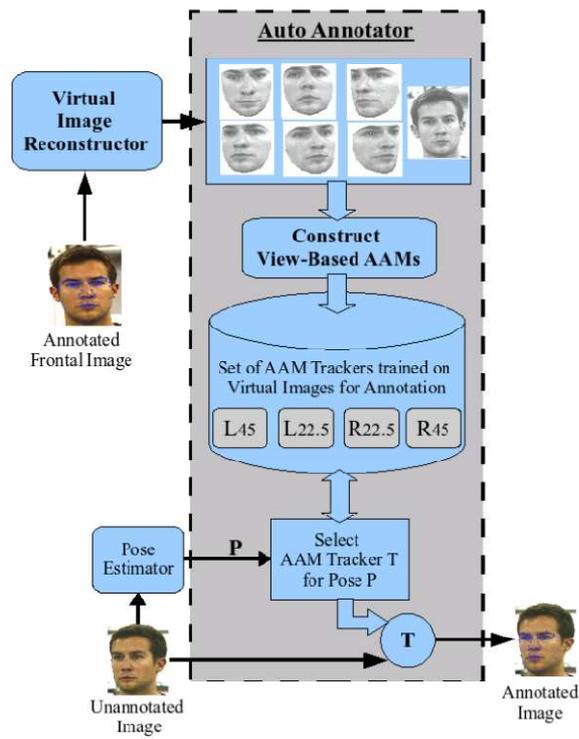


圖 1-9、自動標注 AAM 特徵點之系統架構[7]

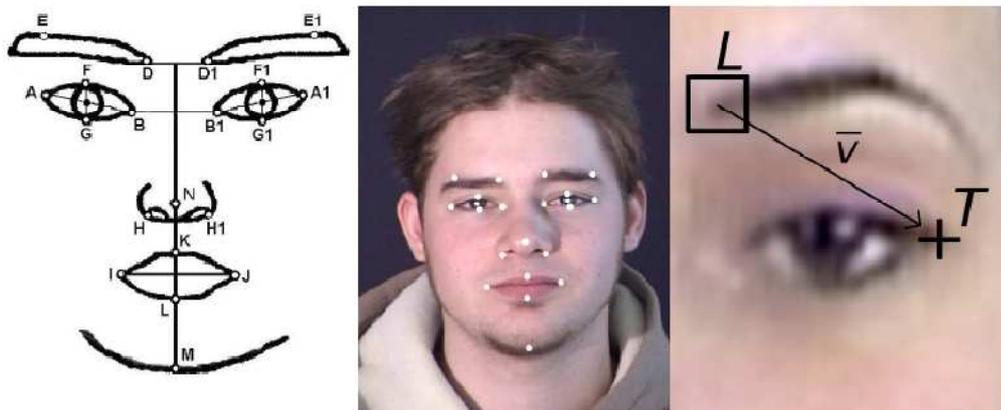


圖 1-10、利用特徵點中的幾何關係取特徵點範例[8]



圖 1-11、利用 subclass divisions 分類取特徵點範例[9]

在許多關於表情辨識的論文中都可看到[10]所定義的 Facial Action Coding System (FACS)，其利用人臉在顯現出不同表情之時會帶動不同臉部肌肉運動的原理，定義了 44 種不同的 Action Unit(AU)，人臉的各種不同表情所對照的是不同的 AU 組合。圖 1-12 是 FACS 中定義的一些 AU 範例。AU 在特徵擷取中衍生出的意義為動態特徵也就是從表情動作開始到完成可以偵測到有數個 AU，通常取從表情動作開始到頂點(變化最大)的數個 AU 作為代表。

AU 概念也可應用於特徵點區域的劃定，如在[11]的研究中，它並不去偵測有哪些 AU 發生而是去依據 AU 會發生的區域作分割以及擷取特徵點，圖 1-13 為其範例。



圖 1-12、數種 AU 的示意圖[10]



圖 1-13、區域分割以及擷取特徵點範例[11]



圖 1-14、FACS 中定義的六種基礎表情[10]

AU 這個概念在另外許多關於表情分類的研究上也常常被運用，如 [12]-[14] 中皆有使用此定義。

在表情分類方面[10]中定義出六種不同的基本表情並被許多研究所沿用，亦即為驚訝(surprise)、生氣(anger)、高興(happy)、傷心(sad)、害怕(fear)、厭惡(disgust)等六種。如圖 1-14 所示：

在資料比對這個階段把資料處理的結果利用已經訓練好的表情分類資料庫或者是分類器對照出人臉的表情。資料庫的來源部分，許多研究採取人為的方式進行訓練(supervised learning)，但也有經由自我分類而不需要人為指定分類的作法(unsupervised learning)。

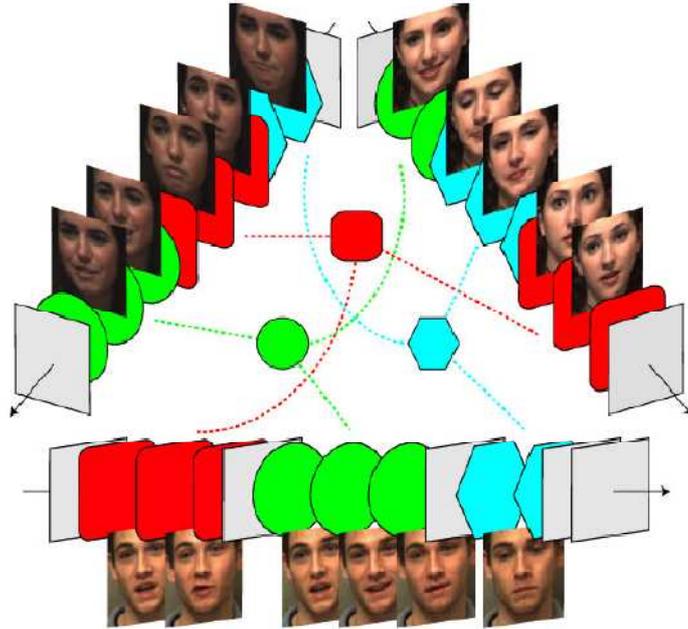


圖 1-15、[15]中 unsupervised learning 之例子

圖 1-15 是一個例子，圖中的不同顏色代表不同表情事件，這些表情事件都是自動分類出的而不是人為去干涉。

常見之訓練分類的方法有：

1. Neural Network[5]: 類神經網路是一種能夠根據每次的訓練時輸出結果與應有結果進行比對並更新的數學模型，它由大量的節點(node)彼此相互連接構成。每個節點都有一特定的輸出函數，即 activation function。每兩個節點間的連接都有一個對於通過該連接信號的加乘值，稱為 weight，這相當於類神經網路的記憶。NN 的輸出則依 NN 的連接方式，weight 和 activation function 的不同而不同。而 NN 自身通常都是對未知的某種演算法或者函數的逼近，也可能是趨近一種邏輯的產生。

2. Ada-Boost[3]: Boosting 是一種將弱算法(弱分類器)融合為強算法(強分類器)的方法。Ada-Boost 是一種迭代型演算法，它根據每次訓練資料中每個樣本的分類是否正確，以及上次的整體分類的準確率，來確定每個樣本的權值。將修改過權值的新數據給下層分類器進行訓練，最後將每次訓練得到的分類器最後融合起來，作為最後的決策分類器。

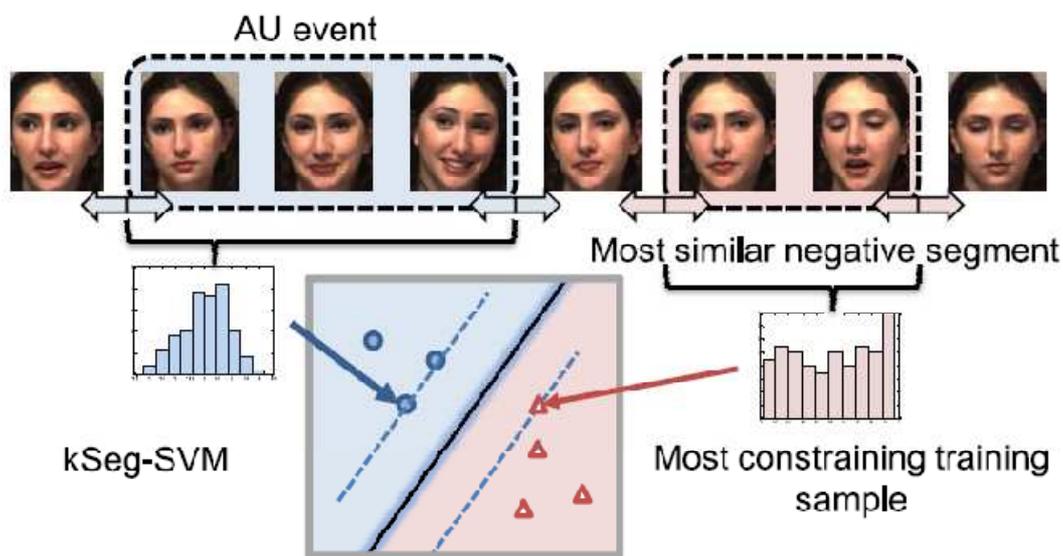


圖 1-16、SVM 分類範例[16]

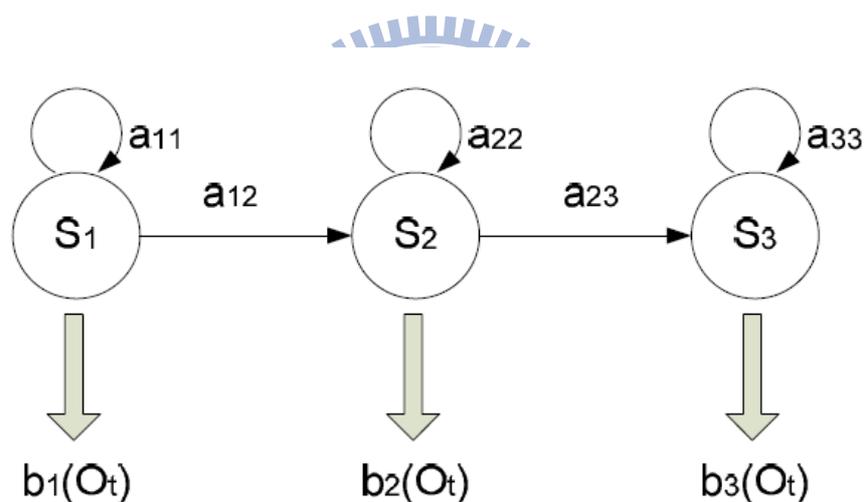


圖 1-17、HMM 模型分類範例[17]

3.Support Vector Machine(SVM): 如圖 1-16 的分類例子[16]，實線為找出的 Hyper-plan，而將兩條虛線稱之為 Support Hyper-plans，目標是找出最佳的 Classification Hyper-plan 使兩 Support Hyper-plans 之間有最大的 Margin。

4. Hidden Markov Model(HMM): Hidden Markov Model(HMM)是統計模型，它用來描述一個含有隱含未知參數的 Markov 過程。其難點是從可觀察的參數中確定該過程的隱含參數。換言之在 HMM 模型中,狀態並不是直接可見的，但受

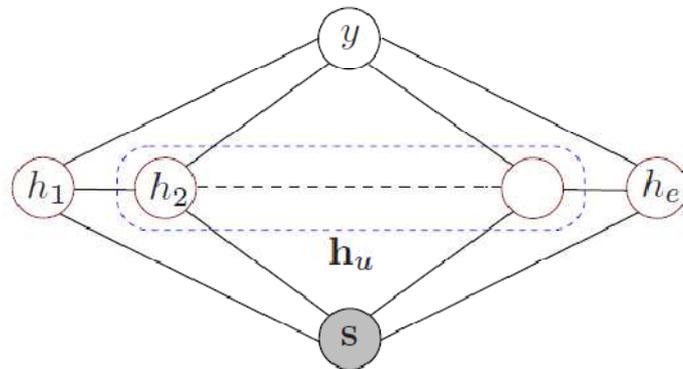


圖 1-18、使用 Partially-Observed Hidden CRF 的模型範例[14]



圖 1-19、辨識表情分類範例[19]

這些狀態影響的某些輸出則是可見的。每一個狀態在可能轉換的狀態上都是以機率決定。因此輸出結果的資料能夠推論出系統處於何種隱含狀態的機率。圖 1-17 為一研究[17]中所使用的 HMM 模型。

5. Markov Random Field(MRF),Conditional Random Field(CRF):這些都是基於機率的數學模型。一些研究會使用它們的變體，例如 Partially-Observed Hidden Conditional Random Fields 在[14]中被使用，圖 1-18 為使用 Partially-Observed Hidden CRF 的模型範例。

有些表情辨識研究與一般表情辨識研究的目的不同，並不只將表情分類到六種基本表情而是將表情分類到各式各樣的自定義表情或者加入表情強度的辨

識。例如研究[19]中目的為辨識出 25 種不同的表情，圖 1-19 為其中幾種表情範例。研究[19]中所使用的架構圖如圖 1-19 所示，相比一般研究多了許多的表情分類器。又例如在研究[20]中以光流場向量為特徵辨識出六種基本表情以及表情強度，其使用統計找出對某種表情較具代表性的區域，提升其參考價值，進而能提升辨識率。

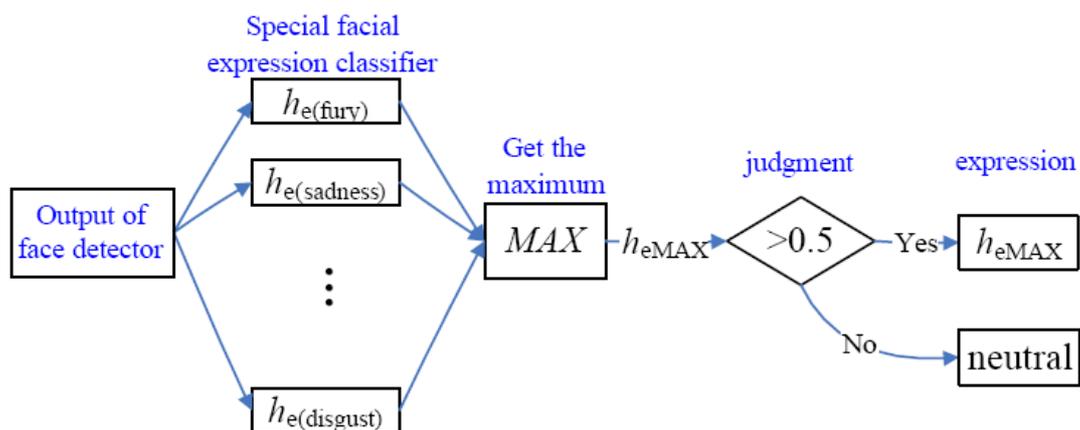


圖 1-20、表情辨識架構圖[19]

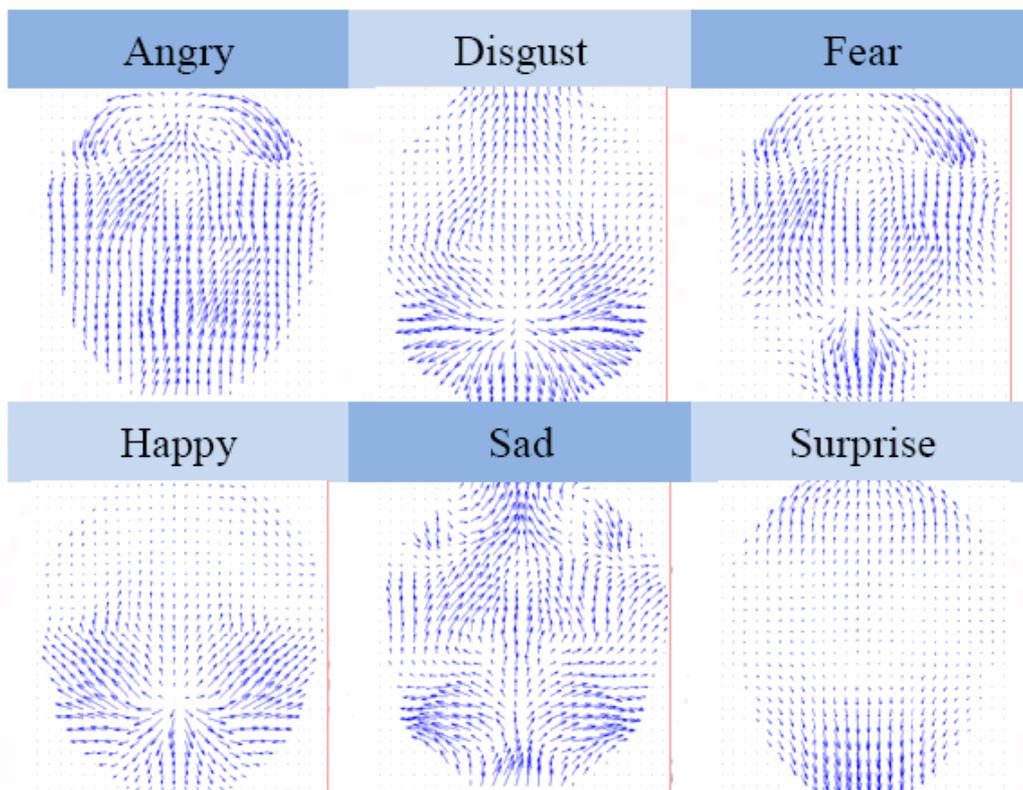


圖 1-21、六種表情之光流場向量[20]

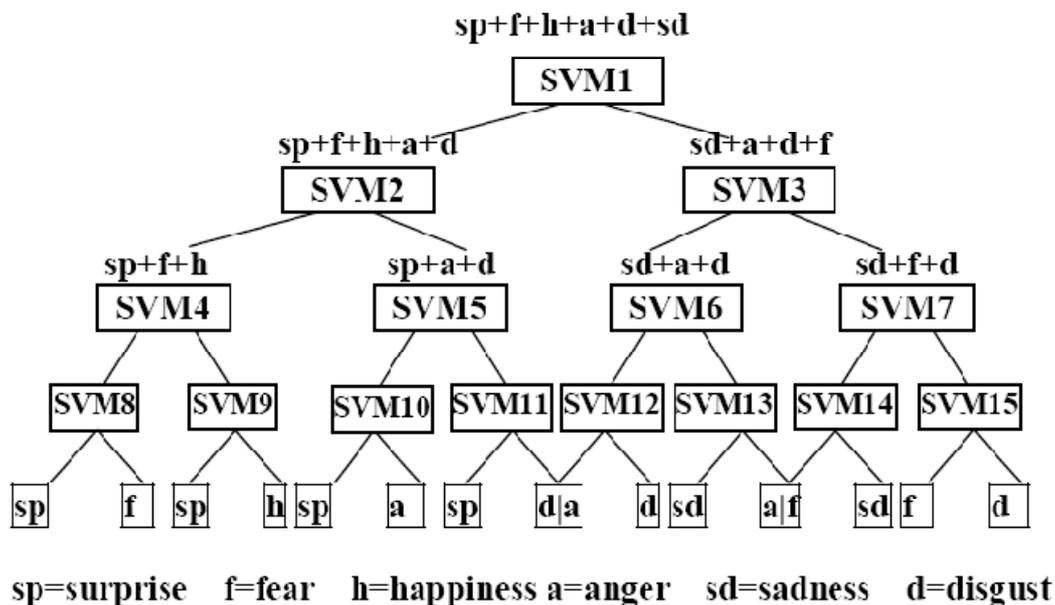


圖 1-22、使用二元分類樹分類範例[21]

在另一個研究[21]中，他們考慮到六種基本表情並不足以應付一般應用情境，因此利用 Fuzzy Kernel Clustering (FKC)、Support Vector Machines (SVM) 以及二元分類樹架構辨識出兩種基本表情混合時的情況。

在上述三個研究中[19]-[21]所擁有的共同點為都不侷限於只將表情辨識為六個基本表情。研究[19]目的為辨識出數目遠多於6的自定義表情，研究[20]之目的在辨識出六種基本表情之外還外加辨識出表情強度，而研究[21]之目的為辨識出兩種基本表情混合時的狀況。本論文與研究[19]-[21]相類似的是不侷限於只將表情辨識為六個基本表情，不同的是本論文想要探討的是更具有一般性的辨識方法。

1.3 問題描述

人臉表情辨識目前所達成的成果大多只集中於依據各種辨識方法將表情歸類在六種基本表情其中一種。但在實務上面，人臉的表情並非只有單純六種，例如其他表情:驚喜、驚恐.....等等不只是單一的基本表情所能涵蓋。如果把強度也

考慮進去的話，就有更多變化例如:大笑、微笑.....等等的表情。簡單來說，人的情緒有多少種，那麼表情就有多少種。

在服務型機器人與人類的互動設計中，機器人越能作出像人類的反應是越好的。為了作出像人類的反應，分辨出人類的情緒是必要的。但是，若只是簡單用基本表情分類將使用者的表情分類在許多時候是不夠精確的，而不精確的辨識表情結果可能會導致不適當的互動反應。考慮到人臉表情的模糊性，我們必須提出一套新的辨識方法辨識出正確的人臉表情。

為了解決辨識人臉表情的模糊性，可以把各種人臉表情視為六種基本表情的比例混合搭配以及最終搭配的表情強度。如此一來待解決的問題就有兩個，分別為表情比例的辨識以及表情強度的辨識。

在本論文中，我們將提出一套整合設計可以分別辨識出表情比例以及表情強度。希望藉由本設計讓未來服務型機器人針對使用者的不同情緒作出最適當的反應，讓機器人更加人性化。



1.4.系統架構及章節說明

本論文提出一個可辨識人臉表情混合比例以及表情強度的系統架構。此架構可略分為三個部分，分別為人臉偵測、人臉特徵值擷取、人臉表情混合比例以及表情強度辨識，其中人臉表情混合比例以及表情強度辨識部分為本論文提出方法的主要重心。

人臉偵測部分是為了將人臉圖像區域從原圖像中取出，此部分主要使用 Haar-like[22][23]人臉偵測方法，人臉特徵值擷取部分是為了要取得能夠在後面人臉表情混合比例以及表情強度辨識中使用的輸入值。此部分則是使用了主動外觀模型(AAM)及Lucas-Kanade 演算法校正人臉然後取出特徵值。人臉表情混合比例以及表情強度辨識部分是為了辨識出使用者的表情混合比例以及表情強度。此部分先將剛剛得到的特徵值輸入倒傳遞類神經網路(Back Propagation Neural

Network, BPNN)得到評分，然後再將這些評分經過我們所提出的基於上下面部動作編碼組合與基本表情的關聯性之運算架構求出人臉表情混合比例以及表情強度。整體架構圖參考圖1-23。

本論文總共分為五章，第一章為緒論，介紹研究動機、相關研究、問題描述以及系統架構。第二章為人臉影像特徵擷取方法，介紹本論文中所使用的人臉特徵值擷取方法。第三章為人臉表情混合比例及強度辨識方法，此章節介紹本論文提出的利用倒傳遞類神經網路以及我們所提出的評分演算法辨識出人臉表情混合比例及強度。第四章為實驗結果。最後第五章則是結論及未來展望。



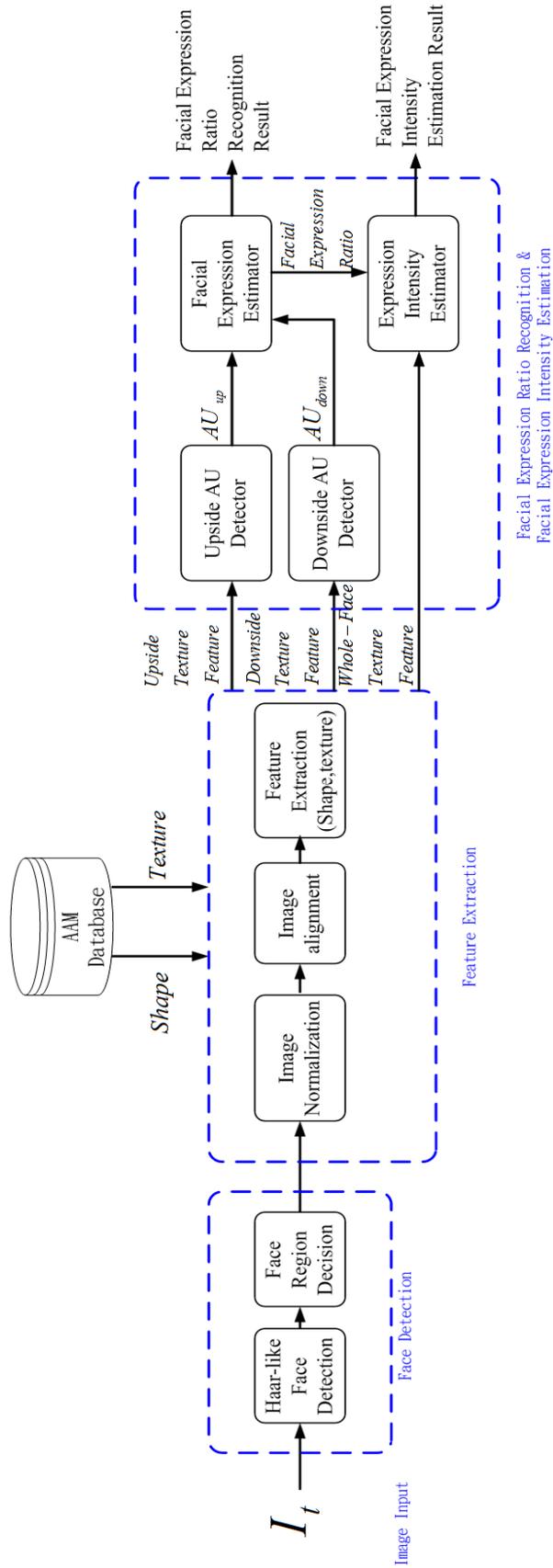


圖 1-23、整體系統架構

第二章 人臉影像特徵擷取方法

在以我們提出的人臉表情辨識演算法辨識人臉表情的混合比例以及表情強度之前，我們需要使用人臉偵測得到人臉區域以及擷取出之後辨識所需的人臉特徵。因此，接下來在本章將介紹從人臉偵測到擷取人臉特徵值之過程。

2.1 人臉偵測演算法

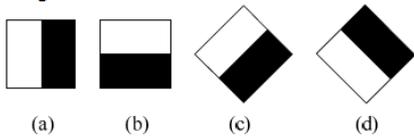
首先，輸入是一張張從攝影機得到的影像圖片。這些影像輸入到系統裡後一開始要先偵測人臉，然後將人臉區域從原圖像中分割出來，這些工作是人臉偵測 (Face Detection) 這個區塊的範圍。

在人臉偵測的部分，參考了[5] [22][23]的研究，採用 Haar-like 人臉偵測來找出可能的人臉位置。Haar-like 人臉偵測方法之特徵可分成三類(如圖 2-1):

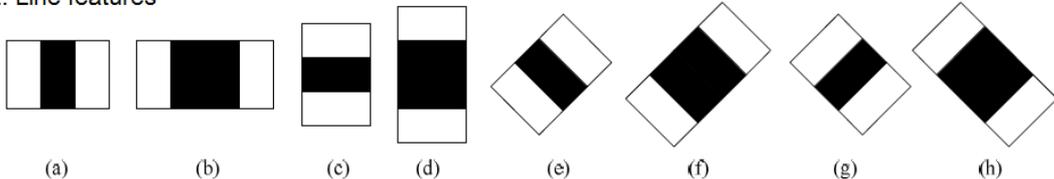
1. 四種邊緣特徵
2. 八種線特徵
3. 兩種中心包圍特徵

計算 Haar-like 特徵時是將圖 2-1 白色部分的影像像素值總和與圖 2-1 黑色部分的影像像素值總和相減得出特徵值，藉由改變矩形特徵的大小和位置，便可在

1. Edge features



2. Line features



3. Center-surround features

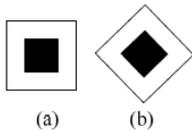


圖 2-1、Haar-like 特徵法

輸入的影像中取出大量 Haar-like 特徵。

取得這些 Haar-like 特徵後便要利用這些特徵值求得人臉區域，這裡採用 Boosting 方法中的 AdaBoost[22][23]以及 Cascade[22][23]方法建構出多階層多分類器的篩選方法，每階層的分類採用 AdaBoost，然後一階一階的篩選出最後留下的區域。

本論文使用 OpenCV(Open Source Computer Vision)所提供的 Haar-like 物體偵測方法來達到上述的人臉偵測。

2.1.1.人臉區域篩選

由於 Haar-like 人臉偵測法會有誤判，因此還要搭配其他的方式，這裡參考了[5]使用的大小限定以及膚色範圍偵測等方法以選出最適合的人臉區域。

首先，應該考慮到當人臉距離攝影機有一段距離時，人臉圖像將會過小而不利於辨識，因此我們設下長寬皆不得低於 50 像數的限制，若換算成距離大約是人不能距離攝影機兩公尺以上。若候選的影像區域長寬低於 50 像數，那就將其刪除掉。

接著，由於選出的影像區域可能有誤判，或者鏡頭前有多個人臉，那在這個時候就要考慮用膚色密度和影像區域面積做進一步的判定。

考慮到光照變化的影響，必須對 RGB 色彩模型作正規化，轉換公式為(2-1)以及(2-2)式[5]:

$$r=R/(R+G+B) \quad (2-1)$$

$$g=G/(R+G+B) \quad (2-2)$$

(2-1)(2-2)中的 R、G、B 為原始 RGB 影像中的紅、綠、藍影像畫素，接著以 r 跟 g 的數值計算膚色範圍[5][24]:

$$g \text{ 的上限[19]: } g_{up} = -1.376r^2 + 1.0743r + 0.1452 \quad (2-3)$$

$$g \text{ 的下限[19]: } g_{down} = -0.776r^2 + 0.5601r + 0.1766 \quad (2-4)$$

$$\text{過濾白色: } W_r = (r - 0.33)^2 + (g - 0.33)^2 \quad (2-5)$$

$$\text{膚色密度: } Skin(r, g) = \begin{cases} 1, & (g < g_{up}) \text{ and } (g > g_{down}) \text{ and } W_r > 0 \\ 0, & \text{otherwise} \end{cases} \quad (2-6)$$

(2-6)中 $Skin(r,g)=1$ 為膚色， $Skin(r,g)=0$ 為非膚色。將所有 $Skin(r,g)$ 的點加總起來便可得出膚色範圍面積，膚色密度可從膚色範圍面積除以總區域面積得到。我們取膚色範圍面積最大的影像區域作為最近的人臉區域。因此我們可總結人臉偵測流程為：

1. Haar-like 人臉偵測法選出候選區域
2. 太小的區域剔除(代表很遠的人臉或是誤判)
3. 最後選擇膚色範圍面積最大的候選區域便可得到最近的人臉區域。

2.1.2 人臉偵測範例

本論文所使用的人臉偵測方法可偵測出最近攝影鏡頭的人臉區域。在可靠度方面，只要不過遠(遠於 2 公尺)過近(近於 20 公分)或者人臉角度偏轉過大，那麼都可準確的抓到人臉區域。圖 2-2 和 2-3 為範例，圖 2-2 為在不同距離及不同人臉角度下的人臉偵測，可看到此人臉偵測方法的可靠度相當高；圖 2-3 為兩人在不同距離下的人臉偵測，可觀察到此人臉偵測方法會取最近的人臉區域。



圖 2-2、在不同距離及不同人臉角度下的人臉偵測

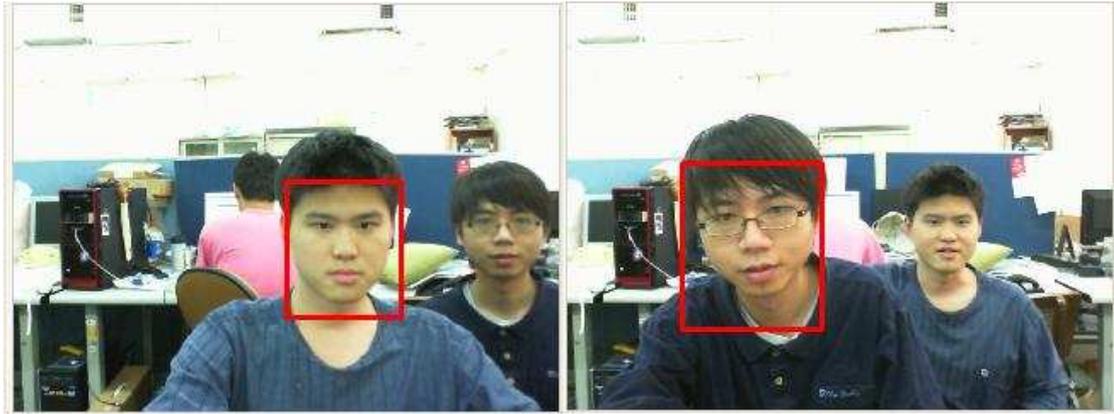


圖 2-3、兩人在不同距離下的人臉偵測

2.2 主動外觀模型

經由人臉偵測取得人臉圖像後，下一步驟是以主動外觀模型(Active Appearance Model, AAM)找出人臉的特徵。主動外觀模型由 Cootes 等人[25]-[27]所提出，主要由形狀模型(Shape Model)與紋理模型(Texture Model)所組成。接下來將說明主動外觀模型的建置方法。

2.3 人臉形狀模型

人臉的主動外觀模型中，人臉的形狀模型由平均人臉形狀(Mean shape)及平均人臉形狀可容許的變化向量組成，利用此模型能夠模擬出不同的人臉形狀。

2.3.1 標定特徵點

主動外觀模型(Active Appearance Model, AAM)在訓練建置時必須手動標定特徵點。在選擇特徵點位置時應定義於曲線的高曲率(High Curvature)部分、曲線間的 T 型交點('T' Junction)，以及中間點，如圖 2-4 所示。

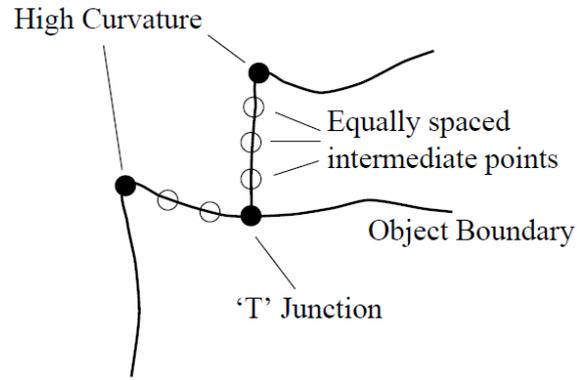


圖 2-4、特徵點選取原則[26]



圖 2-5、特徵點標點範例[26]

在標定特徵點的位置方面，參考[5][26][27]的定義，定義出 70 個人臉的特徵點如圖 2-5，標點位置分配為嘴巴共 18 個點、眼睛共 16 個點、鼻子共 9 個點、眉毛共 10 個點以及人臉輪廓共 17 個點，全部一共標定了 70 個點，圖 2-6 顯示了在不同表情變化下的特徵點位置。



圖 2-6、不同表情下特徵點標點範例

2.3.2 平均人臉形狀

人臉形狀由 70 個特徵座標組成，我們可以由這些不同的特徵座標計算出一組平均形狀。因為輸入人臉形狀的位置、大小、偏轉角度都不相同，我們必須對齊(Align)所有輸入的形狀才能得到我們需要的平均形狀(Mean shape)。

在對齊形狀的部分是使用 Procrustes Analysis[5] [26][27]。因為輸入人臉形狀的位置、大小、偏轉角度都不相同，我們必須對所有輸入的人臉形狀做縮放、位移以及旋轉來對齊， $T_t(\mathbf{x})$ 是經由縮放、位移以及旋轉的輸出， a 、 b 、 c 、 d 是縮放及旋轉的參數，至於 t_x 、 t_y 則是 x 、 y 方向的位移。

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

在對齊過程中，我們先計算出一組平均形狀，接著利用(2-7)的方法計算出新的人臉形狀並算出新的平均形狀，接著重複同樣的步驟直至收斂。在收斂判定方面，我們利用 Procrustes distance(2-8)確定平均形狀的上一次與這一次的差異是否

夠小，若夠小則收斂。而在(2-8)中 P_d 為 Procrustes distance， j 為座標序號， v 是座標序號最大值 70，而 x_{j1}, y_{j1} 則是這一次的平均形狀中特徵點的座標 x_{j0}, y_{j0} 為前一次平均形狀中特徵點的座標。

$$P_d^2 = \sum_{j=1}^v [(x_{j1} - x_{j0})^2 + (y_{j1} - y_{j0})^2] \quad (2-8)$$

我們輸入 120 個人臉形狀做為訓練範例。圖 2-7(a)是原本的人臉形狀重疊後再一起顯示，每一個點都代表 120 個人臉形狀其中一個形狀的 x 、 y 座標，圖 2-7(b)則是經過對齊後將人臉形狀重疊後再一起顯示，圖 2-8 對齊後的所有形狀座標取平均值算出的平均形狀(Mean shape)。

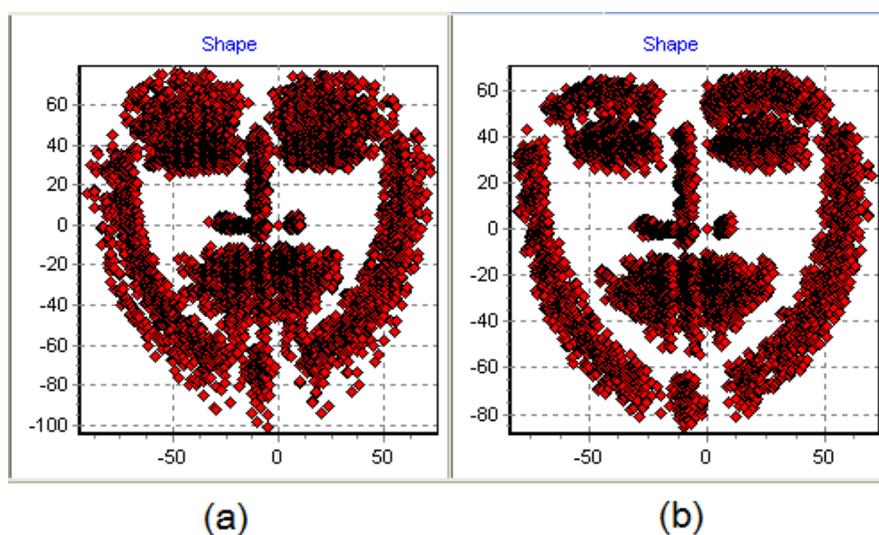


圖 2-7、人臉形狀對齊範例

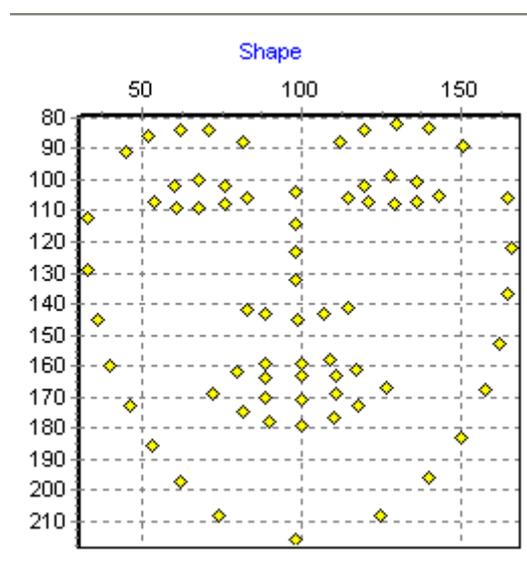


圖 2-8、平均形狀(Mean shape)範例

2.3.3 模擬人臉形狀變化

在前面章節有提到形狀座標總共有 70 個 x 、 y 值。如(2-9)所示，每個 shape 都可視為 $2v$ 維的向量。

$$s = (x_1, y_1, x_2, y_2, \dots, x_v, y_v), v = 70 \quad (2-9)$$

為了建立可模擬人臉形狀變化的模型，我們使用主成分分析(Principal Component Analysis, PCA)以簡化維度資料[27]。平均人臉形狀 s_0 用(2-10)表示，其中 $x_i = \text{training shape set}$ 。

$$s_0 = \frac{1}{n} \sum_{i=1}^n x_i \quad (2-10)$$

$$S = \frac{1}{n} \sum_{i=1}^n (x_i - s_0)(x_i - s_0)^T \quad (2-11)$$

共變異矩陣 S 則可由(2-11)求得特徵向量(eigenvector)和特徵值(eigenvalue)。然後我們可以建立一個形狀變化模型(2-12)，(2-12)內的 s_i $i=1,2,\dots,n$ 為前 n 大的特徵值對應的特徵向量， p_i $i=1,2,\dots,n$ 為權重值所組成的形變參數(shape parameter)。

$$s = s_0 + \sum_{i=1}^n p_i s_i \quad (2-12)$$

2.4 人臉紋理模型

人臉紋理模型也如形狀模型一般，可由平均紋理(Mean texture)與平均紋理可容許之變化所組成，利用人臉紋理模型就能模擬出不同的人臉紋理。人臉的紋理特徵由影像的像素數值所組成。

2.4.1 分段仿射變形

因為每個人臉的大小、形狀、特徵位置都不相同，在建立人臉紋理模型需要用分段仿射變形(Piecewise Affine Warping)讓紋理分布一致化。我們採用 Delaunay triangulation[22]對人臉做分割，再對分割後的各個三角形將每張訓練資料的人臉紋理由原始形狀轉至平均形狀中。Delaunay triangulation 是利

用許多的三角形將輸入的點連接起來，它的特性是每個三角形的外接圓內不會有其他點，我們用 Delaunay triangulation 將人臉形狀進行分割，修改嘴巴部分使得可模擬出嘴唇開合的形狀以及使人臉形狀左右三角形對稱，如圖 2-9 所示。用 Delaunay triangulation 將人臉形狀分割完成後，將訓練的人臉形狀依每個三角形中的紋理轉換到平均人臉形狀中對應的三角形如圖 2-10 所示。

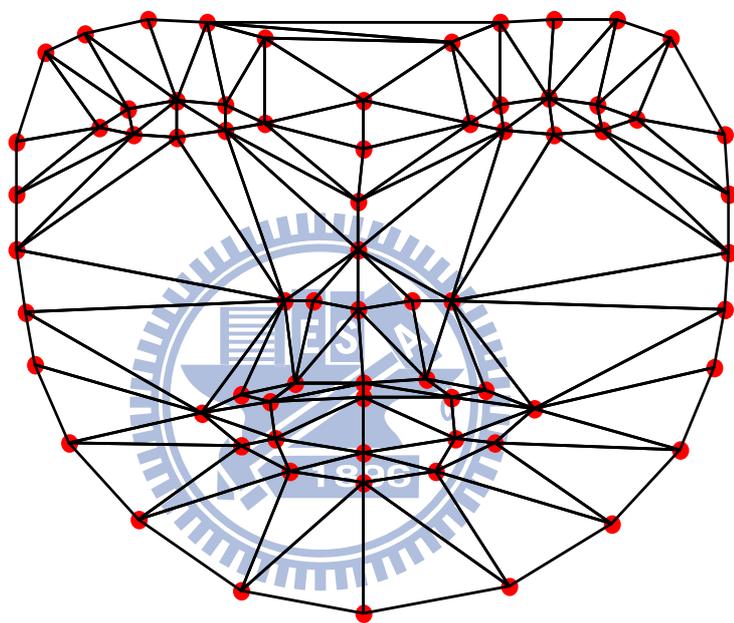


圖 2-9、人臉形狀用 Delaunay triangulation 表示[5]

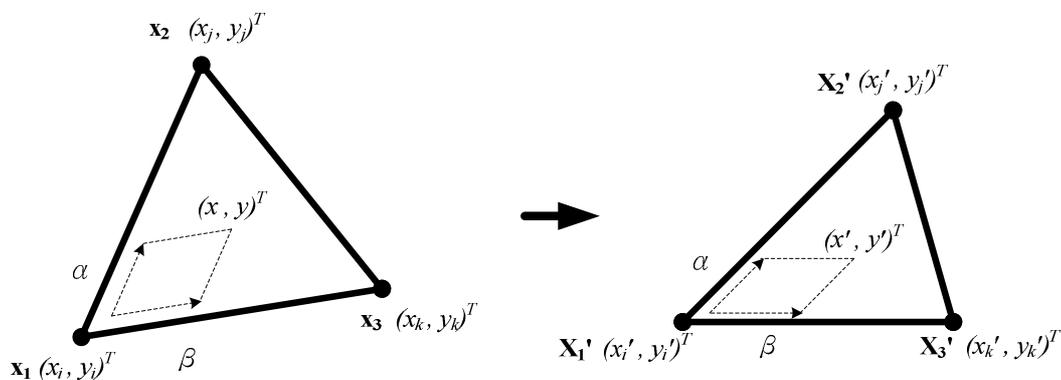


圖 2-10、三角形的對應轉換[5]

2.4.2 模擬人臉紋理變化

我們利用 Principal Component Analysis(PCA)將訓練資料中的人臉紋理資料化簡，並找出人臉紋理的變化。紋理模型組成的向量可表示為(2-13)，其中 $r = \text{texture pixel number}$ 。

$$g = (g_1, g_2, \dots, g_r) \quad (2-13)$$

平均紋理(mean texture) A_0 可從所有轉正的人臉影像紋理相加得出(2-14)，

其中 $g_i = \text{training texture set}$ 。

$$A_0 = \frac{1}{t} \sum_{i=1}^t g_i \quad (2-14)$$

$$G = \frac{1}{t} \sum_{i=1}^t (g_i - A_0)(g_i - A_0)^T \quad (2-15)$$

共變異矩陣 G 的特徵向量(eigenvector)和特徵值(eigenvalue)可從(2-15)計算出，然後可以得到人臉紋理變化模型(2-16)，(2-16)中的 $A_i(x)$, $i=1,2,\dots,n$ 為前 n 大特徵向量(eigenvector)， λ_i 則為對應的特徵值(eigenvalue)。

$$A(x) = A_0(x) + \sum_{i=1}^m \lambda_i A_i(x) \quad \forall x \in s_0 \quad (2-16)$$

2.5 影像校正方法

從前面的章節我們已經掌握了建立主動外觀模型(AAM)中的形狀模型(Shape Model)以及紋理模型(Texture Model)的方法。接下來，我們要找出輸入人臉影像的特徵值。這章將敘述我們如何利用改良過的 Lucas-Kanade 演算法 [28][30] 將輸入人臉影像校正，並利用形狀模型與紋理模型找出輸入人臉影像的形狀參數(shape parameter)以及紋理參數(texture parameter)，然後最後得到我們所需的人臉特徵值。

2.5.1 Inverse Compositional 演算法

Lucas-Kanade 演算法是以影像梯度(Gradient)為基礎的影像變化估測演

算法，常用於影像形變校正、物體追蹤等。Inverse Compositional 演算法[28][30] 為 Lucas-Kanade 演算法之改良方法，比 Lucas-Kanade 演算法更有效減少運算量，原因是 Lucas-Kanade 演算法需要在每次迭代過程中計算出輸入影像的梯度值(gradient)，接著計算 Warping Jacobian 並求出 Steepest descent image 及 Hessian matrix。在 Inverse Compositional 演算法當中，是使用樣版影像計算影像梯度值。由於樣版影像是已知的，Inverse Compositional 演算法只要在迭代開始前計算出樣版影像的梯度值，那麼 Warping Jacobian 以及 Steepest descent image 和 Hessian matrix 就能預先計算，迭代過程就可省去這一部分的計算量，迭代的速率可得以提升。本論文採用 Inverse Compositional 演算法對輸入人臉影像作校正。圖 2-11 中 I~IV 為前處理，1~5 為迭代過程。

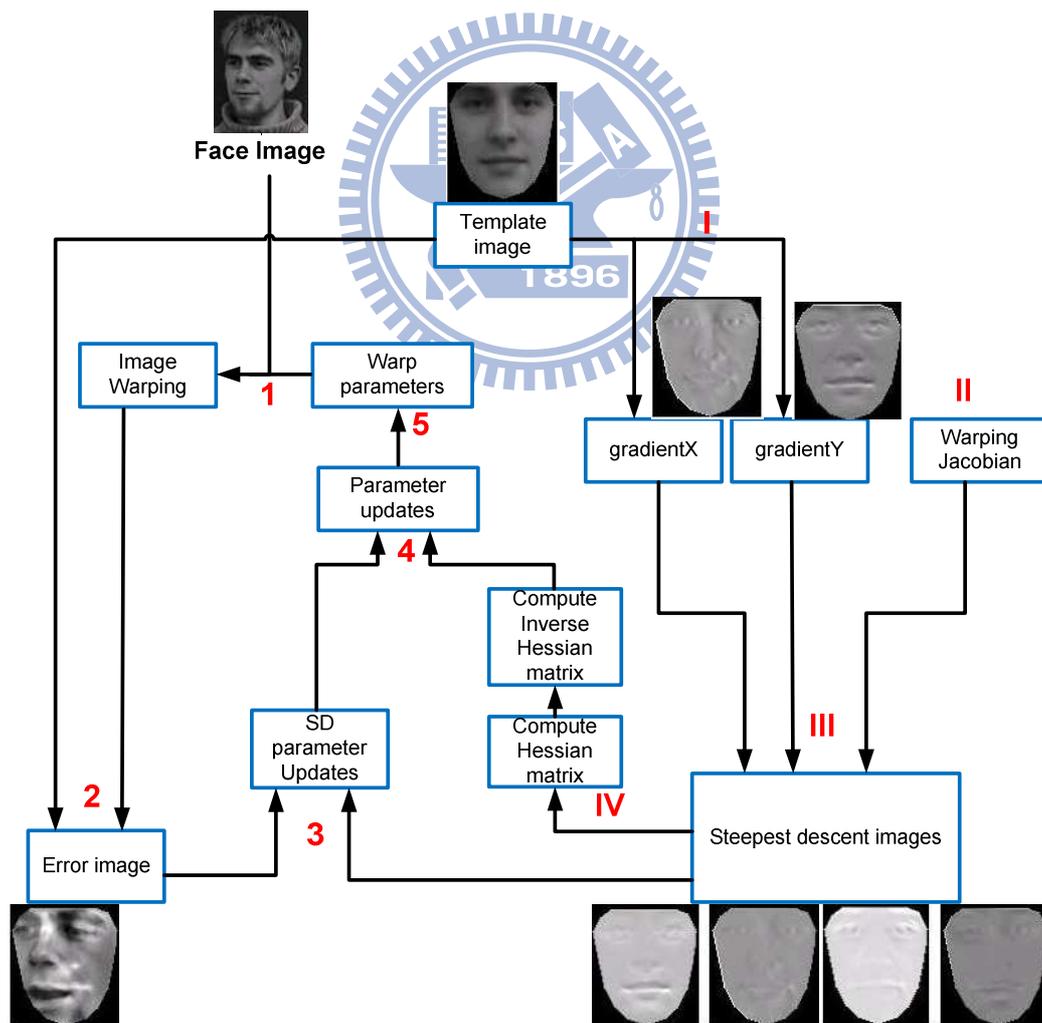


圖 2-11、Inverse Compositional 演算法架構[5]

2.5.2 總體形狀正規化轉換

我們運用總體形狀正規化轉換(Global Shape Normalizing Transform)[28]得到人臉的位移、縮放和旋轉資訊。由於在一般 Inverse Compositional 演算法中，對輸入人臉偏轉的角度容忍度十分有限，因此在校正過程中使用總體形狀正規化轉換之目的為讓輸入人臉偏轉角度之容忍度變大。

我們定義 $N(x; q)$ 為訓練資料的 Global Shape Normalizing Transform[5]，若平移為 (t_x, t_y) ，旋轉角度為 θ ，可用一個參數 $q = (a, b, t_x, t_y)^T$ 代表其形變參數，其中 $a = k \cos \theta - 1$ ， $b = k \sin \theta$ ， $N(x; q)$ 如下所示。

$$N(x; q) = \begin{bmatrix} (1+a) & -b \\ b & (1+a) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \quad (2-17)$$

與之前建立的人臉形狀模型搭配運算，必須建立一組線性組合的係數[25]，這一組係數可模擬形狀模型的變化。若平均形狀 $s_0 = (x_1^0, y_1^0, \dots, x_v^0, y_v^0)^T$ ，則 $s_1^* = s_0 = (x_1^0, y_1^0, \dots, x_v^0, y_v^0)^T$ ， $s_2^* = (-y_1^0, x_1^0, \dots, -y_v^0, x_v^0)^T$ ， $s_3^* = (1, 0, \dots, 1, 0)^T$ ， $s_4^* = (0, 1, \dots, 0, 1)^T$ ，如圖 2-12 所示， $N(x; q)$ 可表示為：

$$N(x; q) = s_0 + \sum_{i=1}^4 q_i s_i^* \quad (2-18)$$

在紋理方面，使用(2-19)計算 (q_1, q_2, q_3, q_4) 的 steepest descent image，這邊的 steepest descent image 會在 Inverse Compositional 演算法中使用。

$$SD_j(x) = \nabla A_0 \frac{\partial N}{\partial q_j} - \sum_{i=1}^m \left[\sum_{x \in S_0} A_i(x) \cdot \nabla A_0 \frac{\partial N}{\partial q_j} \right] A_i(x) \quad (2-19)$$

以及 p 參數的 steepest descent image， $j=1, 2, \dots, n$ ：

$$SD_{j+4}(x) = \nabla A_0 \frac{\partial W}{\partial p_j} - \sum_{i=1}^m \left[\sum_{x \in S_0} A_i(x) \cdot \nabla A_0 \frac{\partial W}{\partial p_j} \right] A_i(x) \quad (2-20)$$

Hessian matrix(會在 Inverse Compositional 演算法中使用)之計算為：

$$H = \sum_x SD(x)^T SD(x) \quad (2-21)$$

紋理參數可表示為：

$$\lambda_i = \sum_{x \in S_0} A_i(x) \cdot [I(N(W(x; p); q)) - A_0(x)] \quad (2-22)$$

而紋理參數 λ_i 將會用來做為我們之後所使用的人臉特徵值。

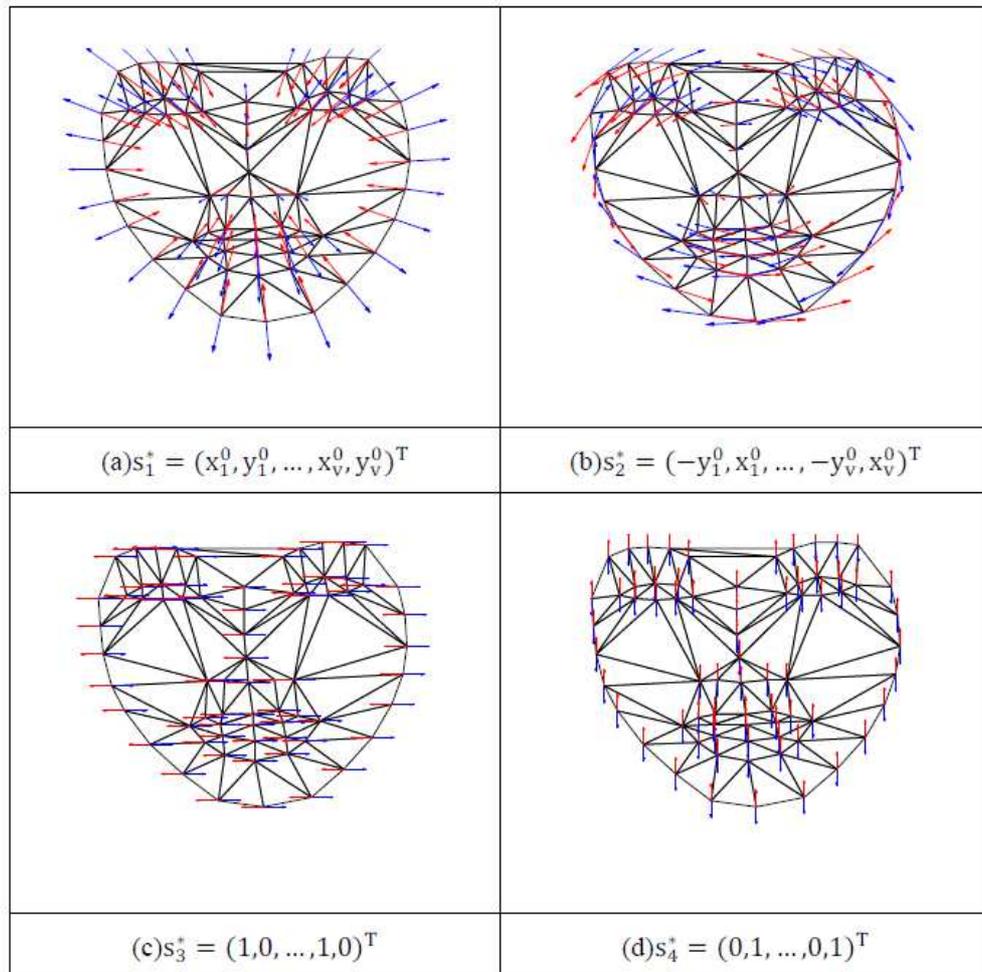


圖 2-12、總體形狀正規化轉換[5]

2.5.3 影像修正

在影像校正過程中使用的影像修正有兩個部分，一為梯度影像修正，二為直方圖等化。首先介紹梯度影像修正：梯度包括垂直方向梯度(gradient X)與水平方向梯度(gradient Y)兩種。由於我們是取樣版影像(如圖 2-13)並算出梯度值，在黑色邊緣的地方會計算出錯誤的梯度值，因此我們將邊緣的梯度值修正為與網內一格相同。



圖 2 -13、樣板人臉影像

再來介紹另一個修正:直方圖等化，其優點是處理後更能凸顯出人臉的特徵部位。在計算直方圖等化時，先將輸入的影像計算出其直方圖，然後由直方圖除以人臉形狀範圍中所有像素之數量算出機率密度函數(Probability Density Function, PDF)，並算出累積分佈函數(Cumulative Distribution Function, CDF)，接著為了對應影像灰階而將累積分佈函數由 0-1 放大到 0-255 的範圍，並將影像原始灰階對應到新的灰階值來取代，就可得到直方圖等化後的影像。

2.5.4 影像校正整體流程

我們結合上述的所有方法並總結流程[5]:

前處理：

- I. 根據樣板影像 $A_0(x)$ 計算梯度影像 ∇A_0
- II. 梯度影像 ∇A_0 進行梯度影像修正
- III. 求出 $(x; 0)$ 的 Jacobian $\frac{\partial W}{\partial p}$ 及 $\frac{\partial N}{\partial q}$
- IV. 由(2-19)與(2-20)計算改變的 steepest descent images $SD_j(x)$ 和 $SD_{j+4}(x)$
- V. 由(2-21)計算 Hessian matrix

迭代過程：

- (1).根據 $N(x; q)$ 與 $W(x; p)$ 形變影像 I ，計算形變影像 $I(N(W(x; p); q))$

- (2). 將形變影像 $I(N(W(x; p); q))$ 進行直方圖等化
- (3). 計算誤差影像(Error Image) $I(N(W(x; p); q)) - A_0(x)$
- (4). 計算 $\sum_x SD(x)^T [I(N(W(x; p); q)) - A_0(x)]$
- (5). 由 inverse Hessian matrix 計算 Δp 與 Δq
- (6). 更新參數 $(N \circ W)(x; p, q) \leftarrow (N \circ W)(x; p, q) \circ (N \circ W)(x; \Delta p, \Delta q)^{-1}$

迭代過程結束：

由(2-22)計算出紋理變化參數 λ_i

本研究在輸入人臉之校正階段時所使用的 Piecewise Affine Warping 是將所有紋理對應轉換到平均形狀，因此在輸入人臉有較強的形狀變化時，在輸出的紋理圖像上較不會有明顯對應。雖然紋理圖像已正規化到平均形狀，還是能夠藉由各部位紋理的變化辨識出表情比例以及強度。由於紋理變化參數保留了人臉很多的資訊，我們選擇紋理變化參數做為我們接下來所使用的人臉特徵參數，並在下一章有更多說明。



第三章 人臉表情混合比例及強度辨識方法

本論文在此章將提出一個基於面部動作編碼組合[10]與表情關聯性的人臉表情混合比例及強度辨識方法(Facial Expression Recognition with Ratio and Intensity Estimation, FER-RIE)，在下面的章節將詳述此方法之運算架構。

3.1 人臉表情混合比例及強度辨識架構

人臉表情辨識目前所達成的成果大多集中於依據各種辨識方法將表情歸類在六種基本表情，也就是 Angry、Disgust、Fear、Happy、Sadness 和 Surprise。但在實務上面，人臉的表情不只是單純六種，例如驚喜、驚恐.....等等的表情即是混合不同比例的二或多種基本表情。如果把強度也考慮進去的話，就有更多變化例如:大笑、微笑.....等等的表情。因此，理想的人臉表情分析系統需要考慮到上述的種種情況。本演算法為了解決辨識人臉表情的模糊性，把各種人臉表情視為六種基本表情的比例混合搭配以及最終搭配的表情強度，並提出以下的架構。

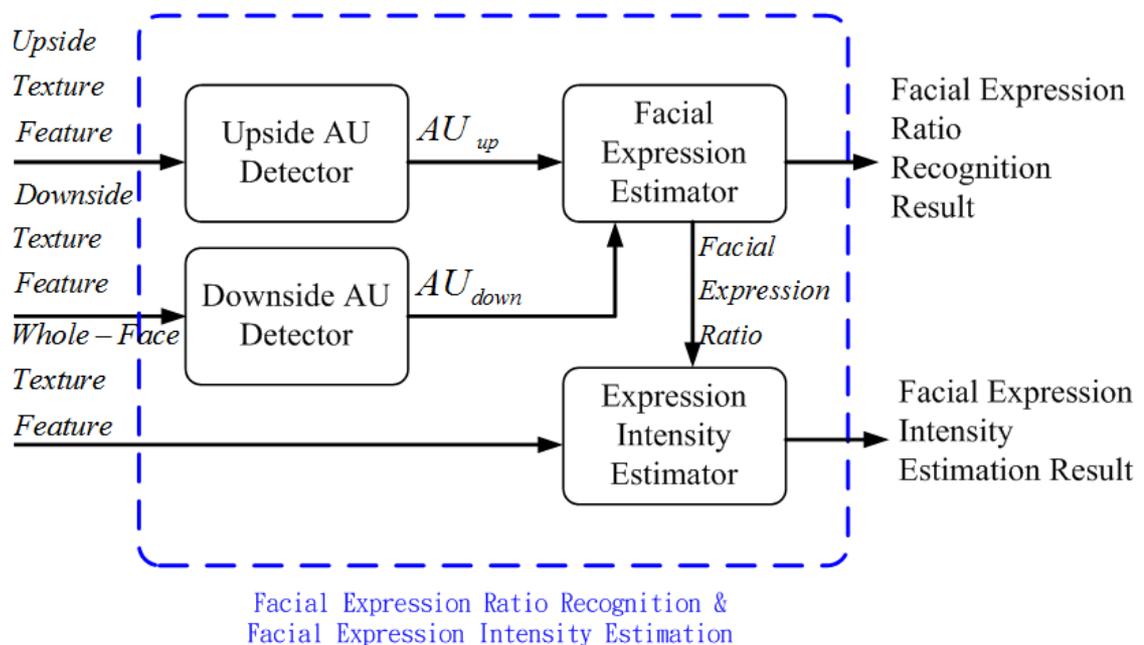


圖 3-1、人臉表情混合比例及強度辨識架構(FER-RIE)

此架構輸入從前一部分(人臉特徵擷取)輸出的人臉紋理參數，輸出人臉表情比例及強度。下面的章節將會詳細說明這個架構的運作方法。

3.2 人臉紋理特徵參數

我們採用了人臉的紋理特徵參數當作我們辨識架構的輸入，由於紋理特徵參數保留了人臉的重要資訊，而且相對於使用影像可以減低不少數據量。

從第二章得知紋理特徵參數為：

$$\lambda_i = \sum_{x \in S_0} A_i(x) \cdot [I(N(W(x; p); q)) - A_0(x)] \quad (3-1)$$

上面(3-1)中 $A_0(x)$ 為樣板影像， $A_i(x)$ 為人臉紋理變化特徵向量， $I(N(W(x; p); q))$ 為形變到人臉平均形狀的人臉影像。因此 $I(N(W(x; p); q)) - A_0(x)$ 為誤差影像(Error Image)， λ_i 為紋理特徵參數， $i=1,2,\dots,m$ ，我們將 λ_i 經過正規化後得到 λ'_i (3-2)作為本章辨識架構的輸入特徵參數。

$$\lambda'_i = \frac{\lambda_i}{\sum_{i=1}^m \lambda_i} \quad i = 1, 2, \dots, m \quad (3-2)$$

表情辨識架構的輸入又可分為三種：

1. 從上半邊人臉(圖 3-2)取得的紋理特徵參數(Upside Texture Feature)。
2. 從下半邊人臉(圖 3-3)取得的紋理特徵參數(Downside Texture Feature)。
3. 從整張人臉(圖 3-4)取得的紋理特徵參數。



圖 3-2、上半邊人臉



圖 3-3、下半邊人臉



圖 3-4、完整人臉

這三種不同紋理特徵參數分別會作為上半邊人臉面部動作編碼組合偵測的輸入(Upside AU Detector)、下半邊人臉面部動作編碼組合偵測的輸入(Downside AU Detector)以及表情強度辨識(Expression Degree Estimator)的輸入。

3.3 倒傳遞類神經網路

在 FER-RIE 架構中，我們需要一個演算法能夠計算出上半臉與上面部動作編碼組合的相似度、下半臉與下面部動作編碼組合的相似度，以及全臉與各種基本表情強度的相似度。由於類神經網路在經由訓練之後能夠藉由輸入的特徵值去

判定輸入之特徵值與各輸出結果之間的相似度，我們就使用類神經網路作為這部分的演算法。

一般類神經網路架構分為輸入層、隱藏層以及輸出層，其層與層之間有權重值(weight)，訓練類神經網路便是藉由調整權重值使輸出入結果與訓練資料同步。在本論文中，我們使用的類神經網路為倒傳遞類神經網路(Back Propagation Neural Network, BPNN)模型[32]。在訓練類神經網路時，運用倒傳遞的方法，將更動值從最後輸出層計算回輸入層，再對於每一層的權重值與偏權值做修改。經過多次迭代訓練後，便可得到最後的權重值與偏權值，而訓練出可使用的類神經網路。我們應用 BPNN 達成三個目的分別是利用上半臉的紋理特徵計算出上半臉與上面部動作編碼組合的相似度、利用下半臉的紋理特徵計算出下半臉與下面部動作編碼組合的相似度，以及利用全臉的紋理特徵計算出全臉與各種基本表情強度的相似度。

3.4 面部動作編碼組合偵測

在 Ekman 等人所提出的研究中[10]中基於人臉肌肉運動所定義的面部動作編碼(Action Units, AU)共有 44 種，而與六種基本表情相關聯的面部動作編碼可以統計出主要的幾種 AU 組合。由於面部動作編碼在 Facial Action Coding System(FACS)[9]中被分為上面部動作編碼(Upper Face Action Units)以及下面部動作編碼(Lower Face Action Units)，我們將 AU 組合分為上半邊面部動作編碼組合(AUC_up)以及下半邊面部動作編碼組合(AUC_down)。這種分類方法能夠有效的減少全部需要考慮的 AU 組合的數目，並且有助於獨立辨識出上半臉與下半臉特徵與表情的關聯程度。本研究藉由將上半邊人臉的紋理特徵參數輸入類神經網路辨認出上半邊人臉與各個上面部動作編碼組合的相似度，以及藉由將下半邊人臉的紋理特徵參數輸入類神經網路辨認出下半邊人臉與各個下半邊動作編碼組合的相似度。然後將類神經網路輸出的相似度評分作為基於 AU 組合與表情關聯度的運算架構之輸入，並最終得到人臉表情混合比例。

3.4.1 面部動作編碼組合

面部表情編碼組合是從 Cohn-Kanade Facial Expression Database 中標有基本表情的人臉圖像資料所對應之面部表情編碼(AU)資料中所選出。目標是選取具有代表性以及區隔性的面部表情編碼組合(AUC)。選取方法則是先選出在各種基本表情中較常出現的上下面部表情編碼組合，也就是過濾掉較不常出現的上下面部表情編碼組合。接下來，從上面選出的上下面部表情編碼組合裡將相似的面部表情編碼組合歸為一類，歸類的基準則是 1.對應的表情種類相同 2.組合內包含之 AU 大部分相同。最後從已歸為一類之組合中選出最常出現的面部表情編碼組合。

我們選擇了 6 組上半邊面部動作編碼組合如圖 3-5 所示，並選擇了 7 組下半邊面部動作編碼組合圖 3-6 所示。例如在圖 3-5 中的 AUC_up1:1+2 代表的是上半邊面部動作編碼組合編號 1 由 Action Unit 1(AU1)以及 Action Unit 2(AU2)所組成，在圖 3-6 中的 AUC_down1:12+25 代表的是下半邊面部動作編碼組合編號 1 由 Action Unit 12(AU12)以及 Action Unit 25(AU25)所組成。

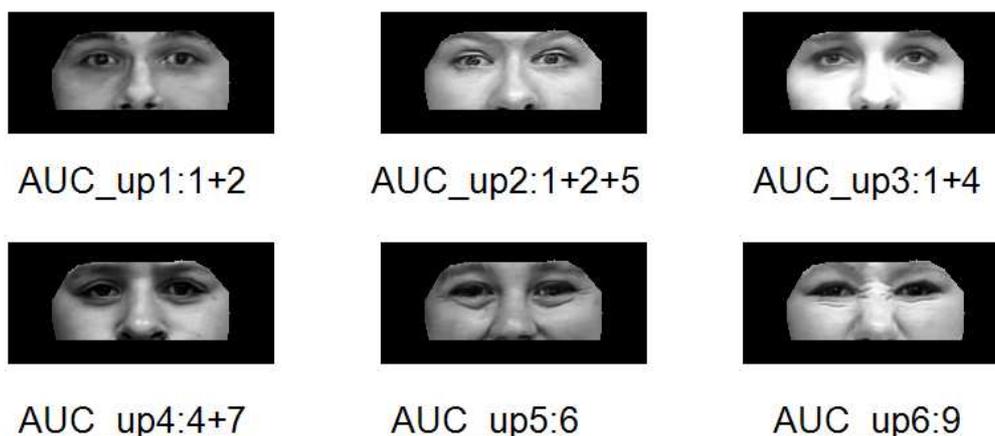


圖 3-5、上半邊面部選出之 6 組動作編碼組合(AUC_up)



圖 3-6、下半邊面部選出之 7 組動作編碼組合(AUC_down)

3.4.2 面部動作編碼組合與表情關係

從 Cohn-Kanade Facial Expression Database [33]選出的訓練資料所整理出上下面部動作編碼組合與表情關聯性統計資料中可觀察出上下面部動作編碼組合與表情關係。有的面部動作編碼組合只與一種表情有關聯，有的就與 2 種表情相關。根據其關聯性可整理出表 3-1 與表 3-2，1 代表相關而 0 代表無關。

表 3-1、上半邊面部動作編碼組合與表情關聯性

	Angry	Disgust	Fear	Happy	Sadness	Surprise
AUC_up1	0	0	1	0	1	0
AUC_up2	0	1	0	0	0	1
AUC_up3	0	0	1	0	1	0
AUC_up4	1	0	0	0	0	0
AUC_up5	0	0	0	1	0	0
AUC_up6	0	1	0	0	0	0

表 3-2、下半邊面部動作編碼組合與表情關聯性

	Angry	Disgust	Fear	Happy	Sadness	Surprise
AUC_down1	0	0	0	1	0	0
AUC_down2	1	1	0	0	0	0
AUC_down3	0	0	0	0	1	0
AUC_down4	1	0	0	0	0	0
AUC_down5	0	1	0	0	1	0
AUC_down6	0	0	1	0	0	0
AUC_down7	0	0	0	0	0	1

3.4.3 面部動作編碼組合評分

本論文將人臉紋理特徵輸入倒傳遞類神經網路得到輸出值，這裡所使用的倒傳遞類神經網路的輸出值為-1 到 1 之間。我們將 0 設為門檻值，超過 0 的輸出值作為對上下半邊面部動作編碼組合與輸入人臉圖像相似度的評分。越靠近 0 代表相似度越低，而越靠近 1 代表相似度越高。這些評分將為作為人臉表情混合比例辨識的輸入值。

3.5 人臉表情混合比例辨識

辨識人臉表情混合比例之目的為辨認出輸入人臉影像的六種基本表情的混合比例。

我們在觀察人臉圖像時，可以獨立的觀察部分人臉或者觀察全部的人臉以估測其表情。因此，我們可依照 FACS[10]的分類法將人臉分成上下兩部分，並藉由獨立觀察上半邊的人臉以及下半邊的人臉辨認其表情為何。為了辨認其表情為何，我們運用上面提到的上下面部動作編碼(AU)組合與表情關聯表(表 3-1、表 3-2)以及上下半邊人臉對上下 AU 組合的相似度評分就可以對表情提出評價。再來，從觀察完整的人臉圖像，能夠更加精確的將表情分類到六種基本表情，因此

我們針對辨認基本表情的情況，利用上下 AU 組合搭配與表情關聯性(表 3-3)就能夠辨認出全臉與基本表情的相似度。如此一來就可兼顧分開觀察部分人臉對複雜表情評價的優勢以及觀察全臉對單純表情評價的優勢。

因此，我們使用由類神經網路輸出的對所有上下臉與 AU 組合之相似度評分作為輸入。我們設計出三部分對表情提出評價:第一部分為全臉評分、第二部分為上半臉評分、第三部分為下半臉評分，分別以上下 AU 組合搭配、獨立以上面部 AU 組合以及獨立以下面部 AU 組合對六種基本表情提出評分。因此，我們結合這三部分輸出評分提出人臉表情混合比例辨識架構如圖 3-7 所示。最後，加總這三部分輸出評分並求得六種基本表情所占比例。接下來我們會詳細敘述第一到第三部分的計算方法。

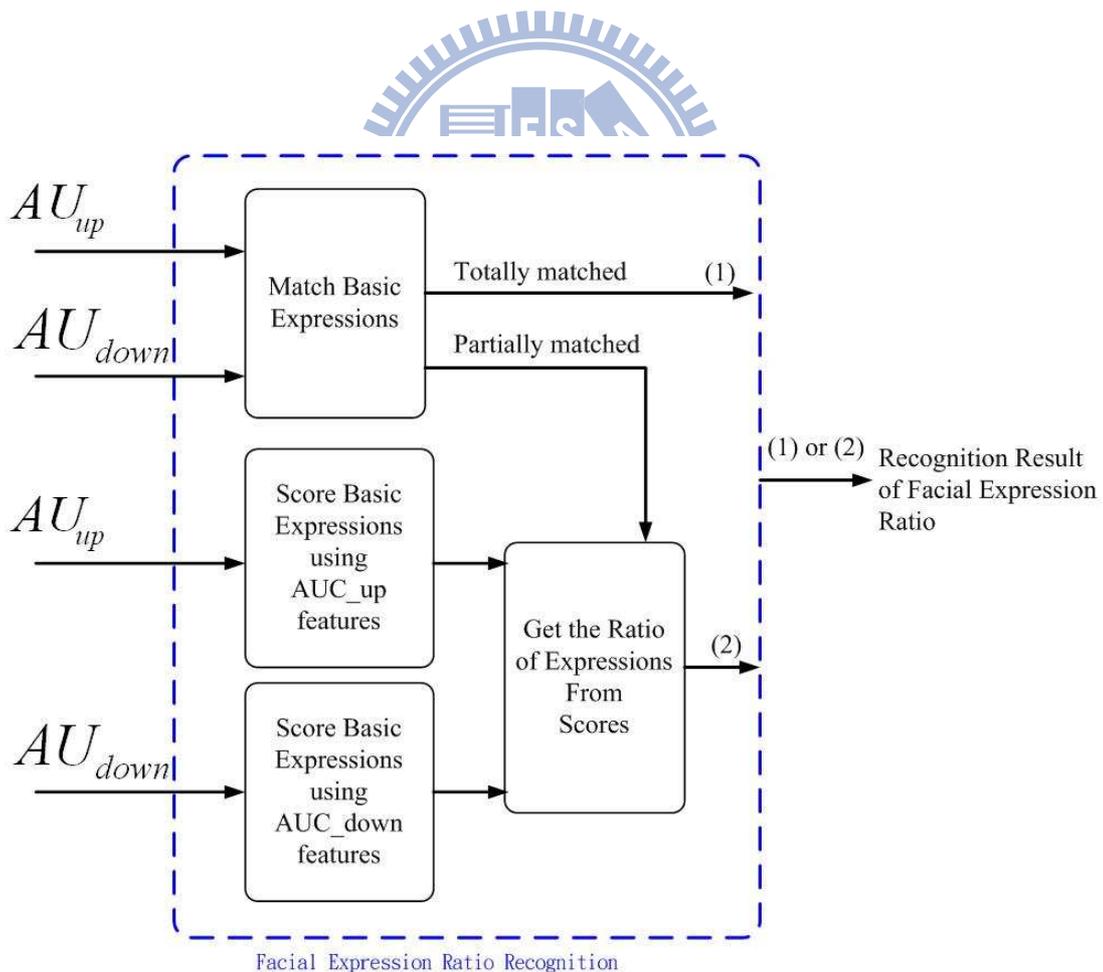


圖 3-7、人臉表情混合比例辨識架構

表 3-3、上下面部動作編碼組合搭配與表情關聯性

		AUC_down						
		1	2	3	4	5	6	7
AUC_up	1			Sa			Fe	
	2						Fe	Su
	3			Sa		Sa	Fe	
	4		An		An			
	5	Ha						
	6		Di			Di		

3.5.1 全臉評分

第一部分的評分觀察的是上下面部動作編碼組合搭配。首先，我們從訓練資料中可得到上下面部動作編碼組合搭配所對應的表情如表 3-3 所示，對應表情編號: An=Angry, Di=Disgust, Fe=Fear, Ha=Happy, Sa=Sadness, Su=Surprise。例如 AUC_up 的編號 5 與 AUC_down 的編號 1 搭配可對應到 Happy。

若上下 AU 組合評分各自中最高的輸入評分項搭配起來可對應到某一基本表情(見表 3-3)，且上下 AU 組合評分各自都只有非零項，那就是完全符合基本表情條件(Totally matched)，對應到的基本表情為 100%其他為 0%。但若上下 AU 組合評分中最高的輸入評分項搭配起來可對應到某一基本表情，但是上下 AU 組合評分有兩個以上的非零項，那麼有可能是部分符合基本表情條件(Partially matched)，還需要確認是否滿足(3-3)條件。

$$\frac{AU_{up_max}}{AU_{up_max} + AU_{up_sec}} > 0.6, \frac{AU_{down_max}}{AU_{down_max} + AU_{down_sec}} > 0.6 \quad (3-3)$$

其中 AU_up_max 為上面部 AU 組合輸入評分最高分，AU_up_sec 為第二高分。AU_down_max 為下面部 AU 組合輸入評分最高分，AU_down_sec 為第二高分。此過濾條件能過濾掉高低分相差太少的情況。若為部分符合，此一部分的評分如

(3-4)所示。這裡的sum_AU_up為上面部AU組合輸入評分總和，而sum_AU_down為下面部AU組合輸入評分總和。

$$\text{score_p1}[i] = \frac{\text{AU_up_max} + \text{AU_down_max}}{\text{sum_AU_up} + \text{sum_AU_down}} \times \sqrt{\text{AU_up_max} \times \text{AU_down_max}} \quad (3-4)$$

其中 $\frac{\text{AU_up_max} + \text{AU_down_max}}{\text{sum_AU_up} + \text{sum_AU_down}}$ 可計算出AU_up_max + AU_down_max占所有輸入評分的比例， $\sqrt{\text{AU_up_max} \times \text{AU_down_max}}$ 則是可計算出兩者之幾何平均，如此便可輸出一個對應某一基本表情(見表 3-3)的評分。此式子(3-4)的特性為當上面部AU組合輸入評分最高分與下面部AU組合輸入評分最高分各自占所有輸入評分總合比例越高時，(3-4)對應某一基本表情的輸出評分會隨之升高。

3.5.2 上半臉評分

第二部分輸出評分觀察的是上面部AU組合，這裡依照其與表情關聯性之強弱給出評分，對照表 3-1、表 3-2 和(3-5)計算出這部分的評分。這裡的score_p2[j]為第j表情編號的評分，AU_up[i]為第i上面部AU組合輸入評分，num[i]為第i上面部AU組合所對應的表情關聯個數(由於關聯個數越多代表對單一表情關聯性越低，因而將輸入評分除以關聯個數作關聯度調整)，u(i,j)則是第i上面部動作編碼組合對第j表情編號關聯值(1或0)。如此便能將所有與某一種基本表情有關聯的上面部AU組合的輸入評分調整後加總。

$$\text{score_p2}[j] = \sum_{i=1}^6 \sum_{j=1}^6 \frac{\text{AU_up}[i]}{\text{num}[i]} \times u(i,j) \quad (3-5)$$

3.5.3 下半臉評分

第三部分輸出評分觀察下面部AU組合，這裡亦依照其與表情關聯性之強弱給出評分，一樣對照表 3-1、表 3-2 和(3-6)計算出這部分的輸出評分。這裡的score_p3[j]為第j表情編號的評分，AU_down[i]為第i下面部AU組合輸入評分，num[i]為第i下面部動作編碼組合所對應的表情關聯個數，d(i,j)則是第i下面部動作編碼組合對第j表情編號關聯值(1或0)。如此便能將所有與某一種基本表

情有關聯的下面部 AU 組合的輸入評分調整後加總。

$$\text{score_p3}[j] = \sum_{i=1}^6 \sum_{j=1}^6 \frac{\text{AU_down}[i]}{\text{num}[i]} \times d(i,j) \quad (3-6)$$

3.5.4 整合評分

然後將三部分的輸出評分加總(3-7)，N、M 皆為常數，若 $\text{score_p2}[i]=0$ 或 $\text{score_p3}[i]=0$ 則 $M=0$ 。N 與 M 都設為大於 1，N 調整第一部分的輸出評分所占比例，M 則是當上下 AU 組合有關聯到同一表情時給予加權。

$$\text{score}[i] = \sum_{i=1}^6 (\text{score_p1}[i] \times N + (\text{score_p2}[i] + \text{score_p3}[i]) \times M) \quad (3-7)$$

接著，因為表情間彼此有搭配可能與否，例如 Happy 不會跟 Sadness 一同出現，利用表 3-4 就可將 Happy 跟 Sadness 一同出現情形視為誤判並加以更正。例如當輸出的最高評分為 Happy 但 Sadness 輸出評分不為零時則將其歸零。

最後將所有輸出評分加總，藉由計算輸出評分除以所有輸出評分總和得到所有表情的混合比例(3-8)。

$$\text{ratio}[i] = \frac{\text{score}[i]}{\sum_{j=1}^6 \text{score}[j]} \quad (3-8)$$



表 3-4、表情搭配檢查表

	Angry	Disgust	Fear	Happy	Sadness	Surprise
Angry	1	1	0	0	1	1
Disgust	1	1	1	0	1	1
Fear	0	1	1	0	1	1
Happy	0	0	0	1	0	1
Sadness	1	1	1	0	1	1
Surprise	1	1	1	1	1	1

3.6 人臉表情強度辨識

表情強度辨識的輸入使用從整張人臉取得的紋理特徵參數並用類神經網路辨識輸出，目的是希望能夠得到一個能夠辨認表情強弱並以0到1的範圍作評分。首先，我們考慮將所有表情混在一起只用不同強度做區別訓練的做法。由於各種表情的分別(如圖 3-8)相當明顯，而且是強度越強差別越大，強行歸在一起分類的做法顯然不妥當。再來，我們考慮訓練分類所用的級數。由於級數越多每一級所分到的訓練張數會隨之降低，並且由於強度具有模糊性難以明確分級，以及不同的人臉強度表現也不相同，若是分級太多對辨識而言沒有幫助且會拉低辨識率。因此我們使用六個分類器，並定義兩種分類對表情強度作訓練。由於表情強度具有模糊性不能單純以二分法區分，因此將分類器之輸出當作對強弱兩端的評分，將結果以從0到1的比例方式計算會較為適當。

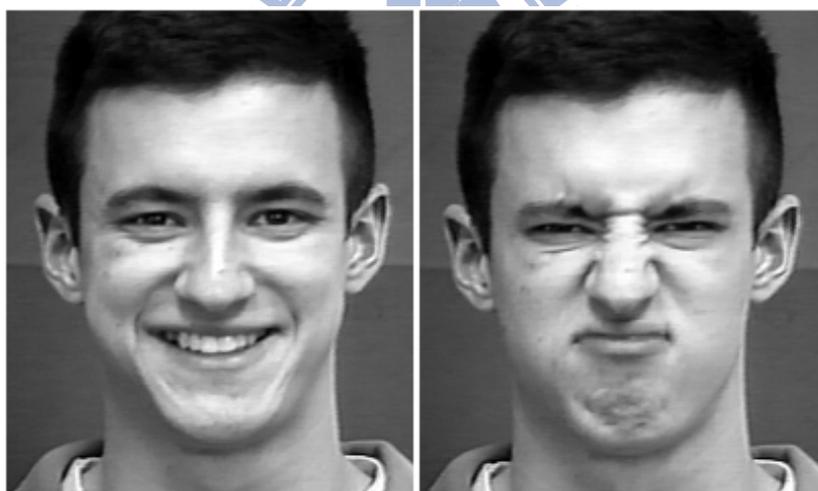


圖 3-8、人臉強烈表情範例(Happy, Disgust)

考慮上面提到的幾點，我們提出一個強度辨識架構如圖 3-9，分別以六種基本表情之強度分類訓練類神經網路，各個類神經網路分類為兩類分別為表情強度強(High)與表情強度弱(Low)。訓練類神經網路完成後，類神經網路輸出為6類各2級的-1到1之值， $High[i], Low[i], i = 1 \sim 6, i$ 對應表情編碼。接著我們引入從人臉表情混合比例辨識輸出的結果 $R[i], i = 1 \sim 6, i$ 對應表情編碼。將類神經網路輸出用式子(3-9)使其範圍在0到1之間。再用式子(3-10)求出最後的表情強度。

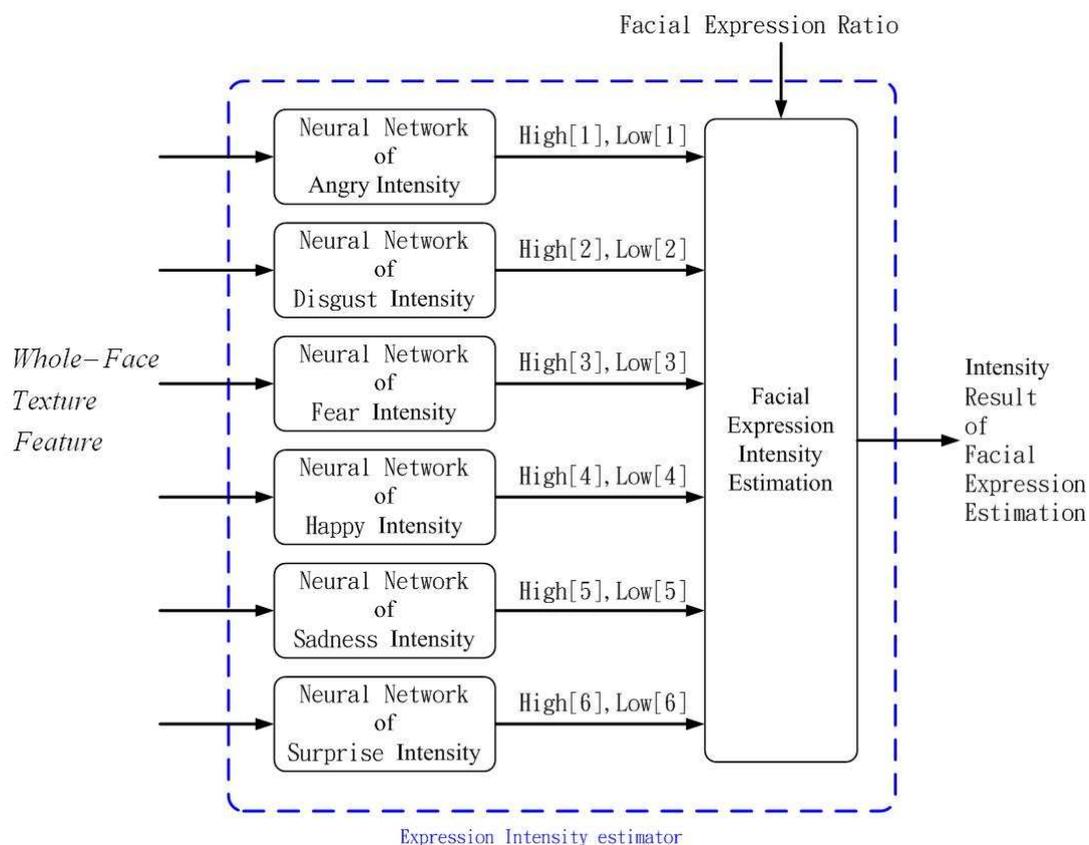


圖 3-9、表情強度辨識架構

$$H[i] = \frac{High[i]+1}{2}, L[i] = \frac{Low[i]+1}{2}, i = \text{facial expression code} \quad (3-9)$$

$$D = \sum_{i=1}^6 \frac{H[i]}{H[i]+L[i]} \times R[i], i = \text{facial expression code} \quad (3-10)$$

在式 3-8 使用的 High 是類神經網路輸出中的高強度項目，Low 則是類神經網路輸出中的低強度項目，High 跟 Low 都各有 6 個值分別對應到 6 個不同表情。在式 3-9 中使用的 R 是從人臉表情混合比例辨識得來的比例，D 則是最終的表情強度輸出結果。式(3-9)的目的為藉由辨識出的表情比例加乘上 6 個類神經網路的輸出找出最適合其表情比例的输出強度。

第四章 實驗結果

為了驗證本論文提出的辨識方法是確實可行的，我們使用 Cohn-Kanade Facial Expression Database (CK+ database) [33]的影像當作訓練以及測試用資料庫。實驗分為 3 個部分: 1.測試是否能正確辨識出六種基本表情 2.測試表情強度之辨識是否正確 3.測試輸入人臉為非單純表情時之表情混合比例是否正確。

4.1 整體訓練流程

我們首先要先挑選人臉圖像作為訓練之輸入，圖像來源可以是由人臉偵測所擷取的人臉圖像或是從資料庫中挑選，本實驗是從 CK+ database 中挑選人臉圖像。

訓練流程:

1. 以挑選好的人臉圖像以及其文件資料(座標、表情、AU 組合、表情強度) 作為 AAM 訓練模型的輸入，並輸出訓練好的 AAM 模型以及各種訓練資料(上半臉特徵值、下半臉特徵值、全臉特徵值、AU 組合與表情關聯、已分類的表情強度資料)。
2. 利用從 1 步驟訓練得到的資料(上下半臉人臉特徵值、AU 組合)訓練出可辨認上下 AU 組合的類神經網路。
3. 利用從 1 步驟訓練得到的資料(AU 組合與表情關聯)訓練出可以用上下 AU 組合評分當作輸入，輸出人臉表情混合比例的演算模型。
4. 利用從 1 步驟訓練得到的資料(全臉特徵值與已分類的表情強度資料)訓練出可以用全臉特徵值當作輸入，輸出人臉表情強度的演算模型。

在建立訓練用資料庫方面，我們選取有適當 AU 組合的圖像序列進行訓練。一個序列由弱到強取七張，標上 AAM 特徵點、表情、AU 組合及表情強度，共取 24 個序列，總共 168 張的人臉影像作為資料庫。圖 4-4 為訓練資料範例 (Surprise)，上面部 AU 組合編號為 5，下面部 AU 組合編號為 1。其中(a),(b),(c)

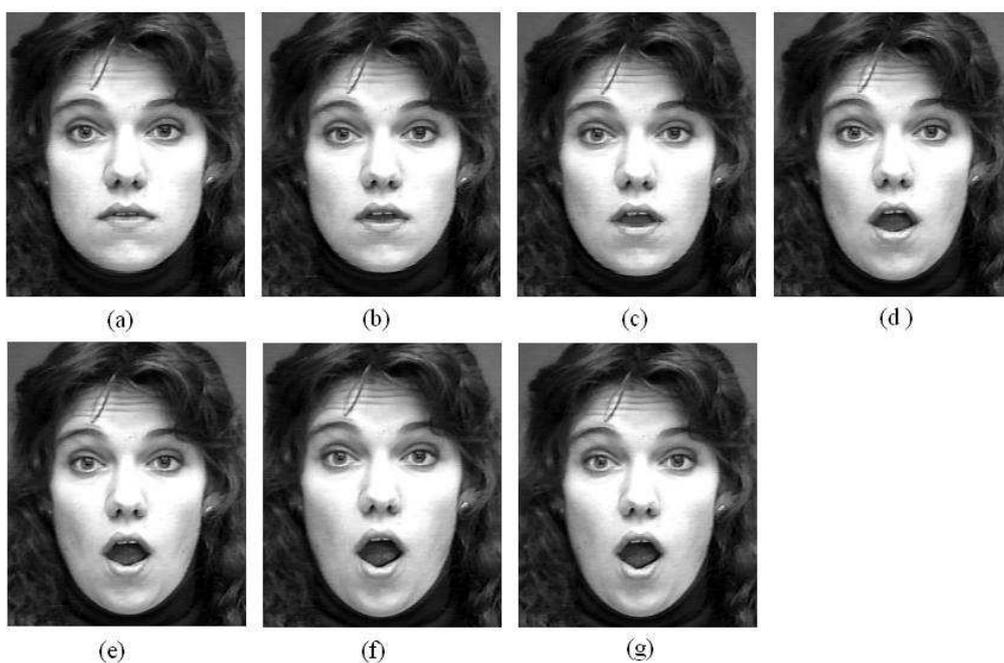


圖 4-1、CK+ 訓練圖像範例

表情強度標注為 1(低強度)，(d) 表情強度標注為 0(不參與表情強度訓練)，(e),(f),(g) 表情強度標注為 2(高強度)。

4.2 類神經網路訓練

本論文之實驗總共使用了 8 個類神經網路並可分為三類用途，第一類用途為辨識上半邊人臉與上面部動作編碼組合之相似度，第二類用途為辨識下半邊人臉與下面部動作編碼組合之相似度，第三類用途為辨識表情強度(六種基本表情以不同的類神經網路加以訓練)。訓練輸入總共有 168 張人臉影像。

首先介紹第一類與第二類的類神經網路，它們的架構如圖 4-2 以及圖 4-3 所示。訓練輸入分別為上臉部的材質特徵以及下臉部的材質特徵，訓練輸出則分別為對應的上面部 AU 組合(6 種組合)及下面部 AU 組合(7 種組合)。

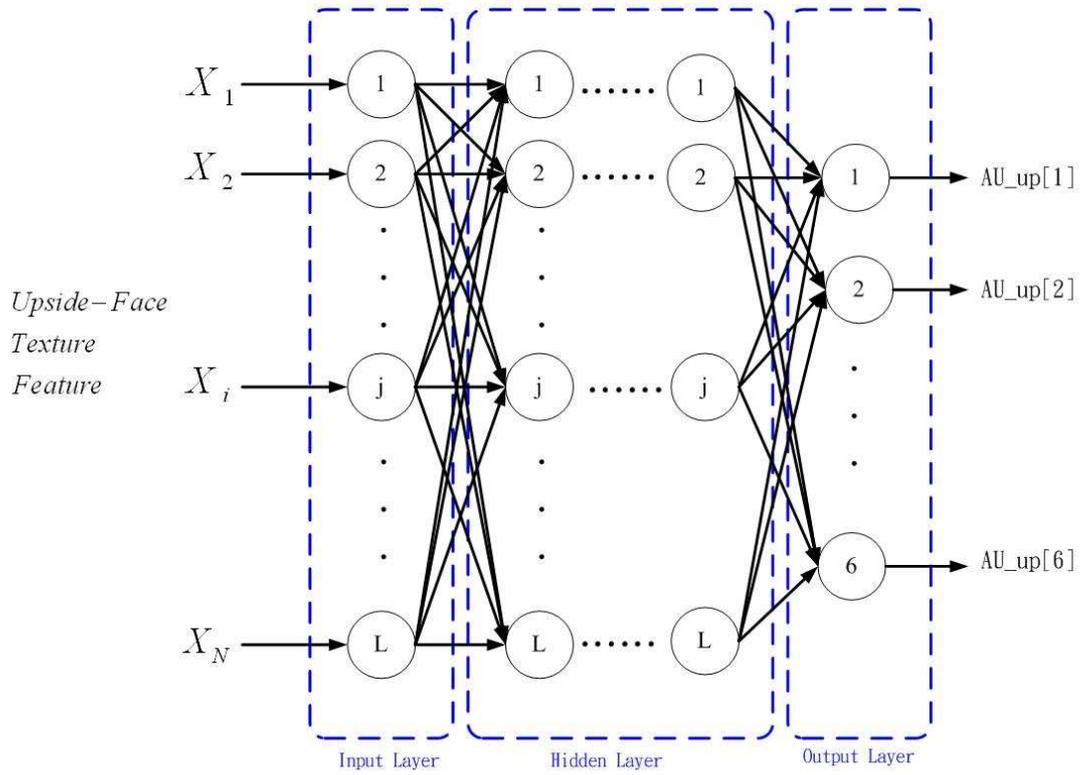


圖 4-2、上面部動作編碼組合類神經網路

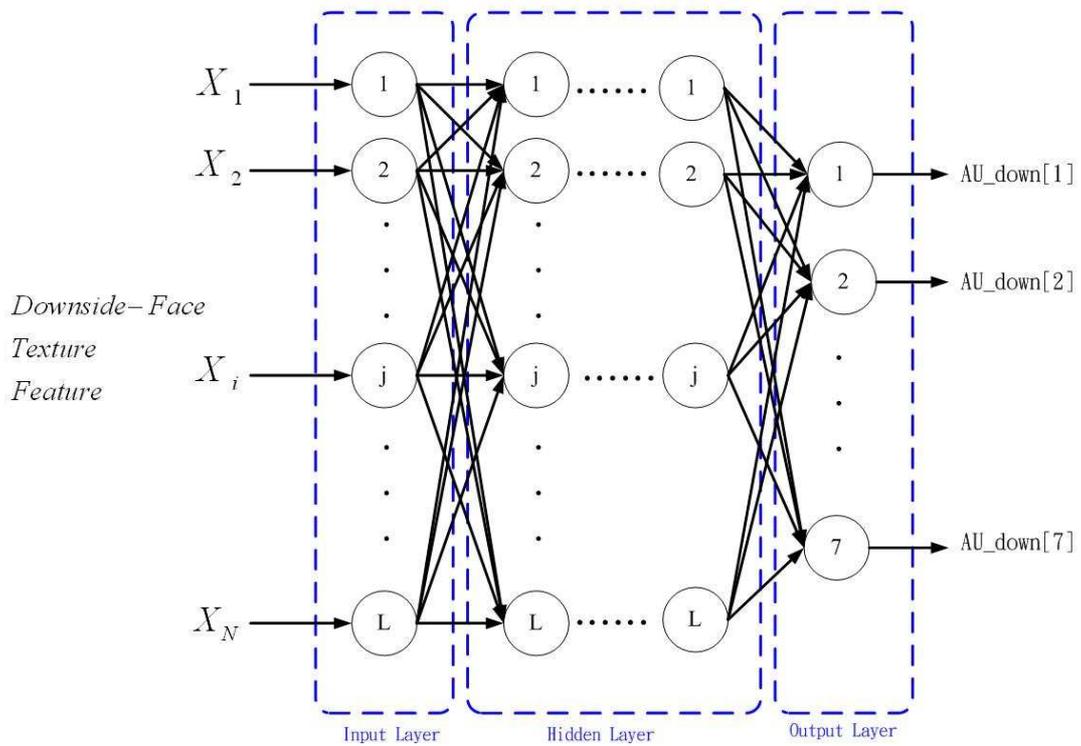


圖 4-3、下面部動作編碼組合類神經網路

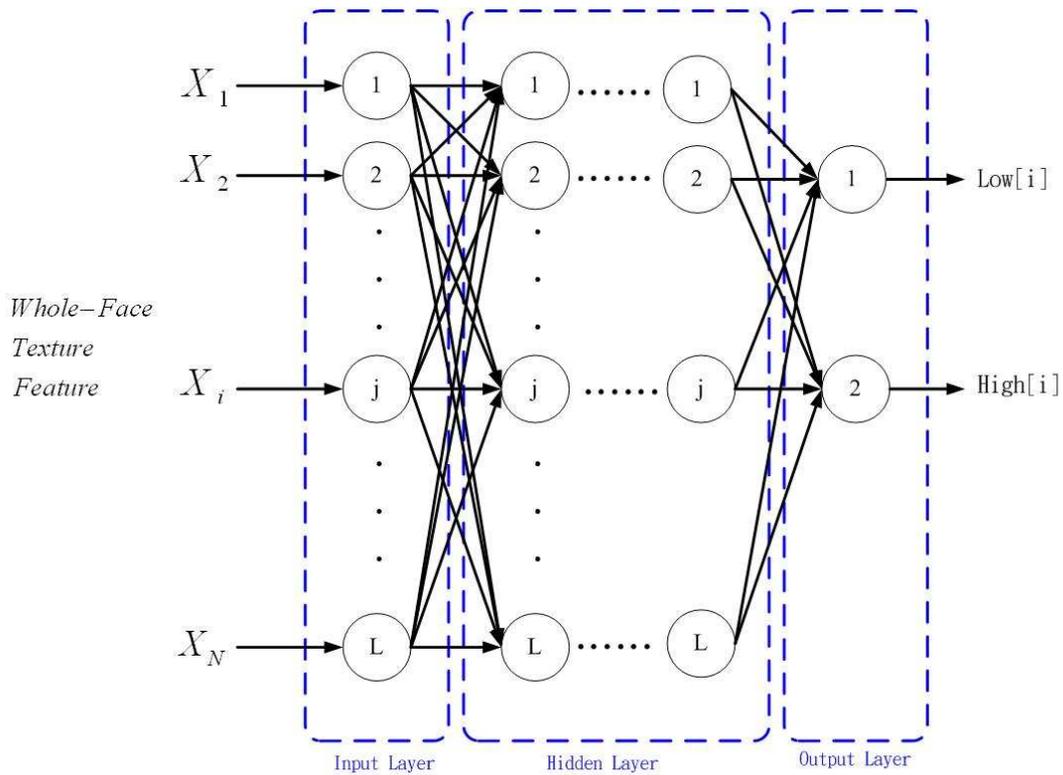


圖 4-4、表情強度類神經網路

最後介紹第三類的類神經網路，如圖 4-4 所示，總共用了六個第三類的類神經網路，訓練輸入為全臉的材質特徵而訓練輸出分為強(High)弱(Low)兩類。

4.3 基本表情辨識結果

在此章節中將使用 CK+ 資料庫[33]驗證本論文所提出的演算法在辨識基本表情(輸入為六種基本表情其中一種)時之表現。此測試使用 18 個序列共 90 張的基本表情人臉影像作為輸入，並藉由調整辨識判斷標準得到不同的實驗結果。圖 4-5 為一個序列的測試資料範例(Angry)。

在基本表情人臉影像表情辨識測試中，將結果以情緒分類依序呈現出來(表 4-2~4-4)。由於結果是以比例呈現因而必須用一套標準統計出辨識率，標準設為最高比例為正確表情並且高於 50%，以及第二高的比例不超過 30%，然後得出表 4-1 的結果。



圖 4-5、CK+ 資料庫測試圖像範例

表 4-1、基本表情人臉影像表情辨識結果(Total)

		INPUT						Total
		Angry	Disgust	Fear	Happy	Sad	Surprise	
OUTPUT	Angry	15	0	0	0	0	0	
	Disgust	0	14	0	0	0	0	
	Fear	0	0	14	2	0	0	
	Happy	0	0	0	13	0	0	
	Sad	0	0	0	0	13	0	
	Surprise	0	0	0	0	0	15	
	Other	0	1	1	0	2	0	
	Correct	15	14	14	13	13	15	84
	Wrong	0	1	1	2	2	0	6
	Total	15	15	15	15	15	15	90
Recognition Rate	100%	93.33%	93.33%	86.67%	86.67%	100%	93.33%	

表 4-2、基本表情人臉影像表情辨識結果(Angry, Disgust)

		Input(Angry)														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	An	100%	94.8%	100%	100%	100%	100%	100%	100%	100%	100%	78.8%	100%	100%	100%	100%
	Di											10.6%				
	Fe															
	Ha															
	Sa		5.2%									10.6%				
	Su															
		Input(Disgust)														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	An						11%	9.5%				27.7%	26.4%	5.8%		
	Di	100%	100%	100%	100%	100%	75.4%	87.5%	55.9%	58.6%	100%	57.1%	58.1%	82.9%	100%	100%
	Fe															
	Ha															
	Sa									14.4%		15.2%	15.5%	11.4%		
	Su						13.6%	3%	44.1%	27%						
Output																
Output																

表 4-3、基本表情人臉影像表情辨識結果(Fear, Happy)

		Input(Fear)														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	An															
	Di															
	Fe	100%	100%	50.3%	77%	87.2%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	Ha															
	Sa			49.7%	23%	12.8%										
	Su															
		Input(Happy)														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	An															
	Di															
	Fe	77.6%				88.5%										
	Ha		100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	Sa	22.4%				11.5%										
	Su															
Output																
Output																

表 4-4、基本表情人臉影像表情辨識結果(Sadness, Surprise)

		Input(Sadness)														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	An															
	Di													0.5%		
	Fe	50%					50%							10.5%		
	Ha															
	Sa	50%	100%	100%	100%	100%	50%	100%	100%	100%	100%	100%	100%	89%	100%	100%
	Su															
		Input(Surprise)														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	An															
	Di															
	Fe															
	Ha															
	Sa															
	Su	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Output																
Output																

表 4-5、其他研究之基本表情人臉影像表情辨識結果[34]

	Angry	Disgust	Fear	Happy	Sad	Surprise	Total
Angry	93.3%	13.3%	3.3%	0.0%	3.3%	0.0%	
Disgust	0.0%	83.3%	23.3%	6.7%	0.0%	0.0%	
Fear	0.0%	0.0%	66.7%	0.0%	0.0%	0.0%	
Happy	0.0%	3.3%	0.0%	93.3%	0.0%	0.0%	
Sad	6.7%	0.0%	0.0%	0.0%	96.7%	0.0%	
Surprise	0.0%	0.0%	6.7%	0.0%	0.0%	100%	
Recognition Rate	93.3%	83.3%	66.7%	93.3%	96.7%	100%	88.9%

在其他研究中[34]，也運用了 CK+資料庫做表情辨識之研究，表 4-5 可跟本研究之基本表情人臉影像表情辨識結果表 4-1 相比較。



4.4 表情強度辨識結果

在此章節中將使用 CK+資料庫[33]驗證本論文所提出的演算法在辨識表情強度時之表現。此測試使用 30 張的基本表情人臉影像以及 15 張的混合表情(CK+資料庫未註明表情者)人臉影像一共 45 張人臉影像作為輸入，並將輸出的測試結果與問卷調查作比較(表 4-5)，其中 N 為影像編號，I 為表情辨識結果，AVG 為問卷結果平均，STD 則為標準差。問卷總共調查了 9 個人的意見並統計結果。首先以測試結果與問卷調查結果之平均值在 1.5 個問卷調查結果之標準差之內 (1.5 Stdev)作為判斷標準得出表 4-7，再以測試結果與問卷調查結果平均值在問卷調查結果之 1 個問卷調查結果之標準差之內(1 Stdev)作為判斷標準得出表 4-8。

表 4-6、人臉表情強度辨識結果

Input															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
N															
I	0.225	0.749	0.810	0.283	0.647	0.316	0.896	0.319	0.765	0.674	0.702	0.890	0.321	0.902	0.868
N	16	17	18	19	20	21	22	23	24	25	26	27	28	28	30
I	0.332	0.897	0.331	0.500	0.902	0.772	0.270	0.810	0.226	0.810	0.500	0.902	0.770	0.380	0.902
N	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
I	0.762	0.366	0.253	0.790	0.368	0.651	0.654	0.903	0.254	0.223	0.227	0.291	0.595	0.636	0.296
Input															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
N															
AVG	0.433	0.678	0.633	0.156	0.333	0.511	0.856	0.456	0.789	0.722	0.600	0.733	0.522	0.767	0.700
STD	0.218	0.244	0.260	0.124	0.218	0.136	0.167	0.207	0.220	0.222	0.200	0.200	0.228	0.187	0.218
N	16	17	18	19	20	21	22	23	24	25	26	27	28	28	30
AVG	0.533	0.767	0.356	0.622	0.756	0.478	0.189	0.589	0.383	0.694	0.522	0.794	0.583	0.650	0.856
STD	0.287	0.235	0.309	0.228	0.207	0.222	0.162	0.252	0.278	0.240	0.205	0.151	0.190	0.184	0.133
N	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
AVG	0.811	0.278	0.189	0.611	0.667	0.689	0.611	0.689	0.433	0.178	0.411	0.456	0.367	0.578	0.500
STD	0.154	0.239	0.252	0.203	0.158	0.190	0.232	0.176	0.240	0.179	0.196	0.219	0.200	0.244	0.218
Result of questionnaires															

表 4-7、人臉表情強度辨識與問卷比較結果(1.5 Stdev)

		INPUT							
		Angry	Disgust	Fear	Happy	Sad	Surprise	Other	Total
	Correct	5	5	5	5	5	5	13	43
	Wrong	0	0	0	0	0	0	2	2
	Total	5	5	5	5	5	5	15	45
	Recognition Rate	100%	100%	100%	100%	100%	100%	86.67%	95.56%

表 4-8、人臉表情強度辨識與問卷比較結果(1 Stdev)

		INPUT								
		Angry	Disgust	Fear	Happy	Sad	Surprise	Other	Total	
	Correct	3	4	5	8	5	4	4	11	36
	Wrong	2	1	0	0	1	1	4	9	
	Total	5	5	5	5	5	5	15	45	
	Recognition Rate	60%	80%	100%	100%	80%	80%	73.33%	80%	

4.5 混合表情比例辨識結果

在此章節中將使用 CK+ 資料庫[33]驗證本論文所提出的演算法在辨識混合表情(資料庫未標注表情)之表情混合比例時表現。此測試使用 8 張的混合表情人臉影像作為輸入，並將輸出的測試結果與問卷調查作比較，此問卷總共調查了 25 個人的意見並統計結果。這 25 人的年齡層分布在 25 歲~55 歲之間，男性有 16 人，女性有 9 人。問卷調查之對象為實驗室的同學以及電機系所辦公室之職

員，在經過一一詢問後得到了總共 25 人的協助。下面列出全部的測試與問卷調查結果，其中 Stdev 項為標準差。



圖 4-6、混合表情測試圖像(1)

表 4-9、混合表情測試辨識輸出結果(1)

Angry	Disgust	Fear	Happy	Sadness	Surprise
0%	50%	0%	0%	50%	0%

表 4-10、混合表情測試問卷調查結果(1)

	Angry	Disgust	Fear	Happy	Sadness	Surprise
Average	23.2%	37.6%	0%	0%	39.2%	0%
Stdev	26.1%	29.45%	0%	0%	39.26%	0%

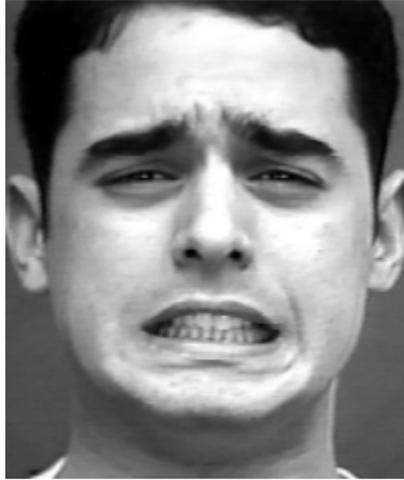


圖 4-7、混合表情測試圖像(2)

表 4-11、混合表情測試辨識輸出結果(2)

Angry	Disgust	Fear	Happy	Sadness	Surprise
0%	0%	80.1%	0%	19.9%	0%

表 4-12、混合表情測試問卷調查結果(2)

	Angry	Disgust	Fear	Happy	Sadness	Surprise
Average	6.8%	19.6%	38%	0%	32.8%	2.8%
Stdev	17.01%	21.89%	41.33%	0%	33.85%	8.43%



圖 4-8、混合表情測試圖像(3)

表 4-13、混合表情測試辨識輸出結果(3)

Angry	Disgust	Fear	Happy	Sadness	Surprise
70%	30%	0%	0%	0%	0%

表 4-14、混合表情測試問卷調查結果(3)

	Angry	Disgust	Fear	Happy	Sadness	Surprise
Average	61.2%	36%	0.8%	0.4%	1.2%	0.4%
Stdev	29.91%	30.55%	2.77%	2%	4.4%	2%



圖 4-9、混合表情測試圖像(4)

表 4-15、混合表情測試辨識輸出結果(4)

Angry	Disgust	Fear	Happy	Sadness	Surprise
0%	0%	50%	0%	50%	0%

表 4-16、混合表情測試問卷調查結果(4)

	Angry	Disgust	Fear	Happy	Sadness	Surprise
Average	1.2%	14.2%	43.6%	2.4%	38.6%	0%
Stdev	3.32%	20.29%	39.36%	10.12%	41.22%	0%

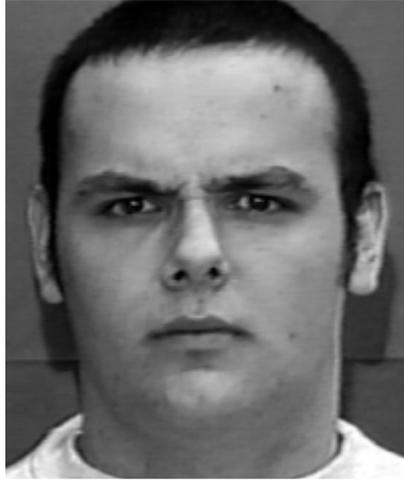


圖 4-10、混合表情測試圖像(5)

表 4-17、混合表情測試辨識輸出結果(5)

Angry	Disgust	Fear	Happy	Sadness	Surprise
100%	0%	0%	0%	0%	0%

表 4-18、混合表情測試問卷調查結果(5)

	Angry	Disgust	Fear	Happy	Sadness	Surprise
Average	55.4%	31.6%	2.8%	0%	7%	2.8%
Stdev	33.97%	34.6%	8.91%	0%	14.29%	10.61%



圖 4-11、混合表情測試圖像(6)

表 4-19、混合表情測試辨識輸出結果(6)

Angry	Disgust	Fear	Happy	Sadness	Surprise
0%	0%	0%	0%	100%	0%

表 4-20、混合表情測試問卷調查結果(6)

	Angry	Disgust	Fear	Happy	Sadness	Surprise
Average	16.2%	31.8%	2%	0%	50%	0%
Stdev	25.71%	37.83%	7.07%	0%	42.91%	0%



圖 4-12、混合表情測試圖像(7)

表 4-21、混合表情測試辨識輸出結果(7)

Angry	Disgust	Fear	Happy	Sadness	Surprise
0%	0%	50%	0%	0%	50%

表 4-22、混合表情測試問卷調查結果(7)

	Angry	Disgust	Fear	Happy	Sadness	Surprise
Average	0%	2.8%	21.6%	2.8%	0.4%	72%
Stdev	0%	6.78%	22.49%	10.61%	2%	21.21%



圖 4-13、混合表情測試圖像(8)

表 4-23、混合表情測試辨識輸出結果(8)

Angry	Disgust	Fear	Happy	Sadness	Surprise
0%	0%	0%	55.7%	0%	44.3%

表 4-24、混合表情測試問卷調查結果(8)

	Angry	Disgust	Fear	Happy	Sadness	Surprise
Average	0.4%	1.2%	2.8%	37.8%	2.4%	55.4%
Stdev	2%	3.32%	10.21%	34.94%	12%	32.4%

4.6 結論與討論

在辨識基本表情測試時，由表 4-1 之結果可以看出辨識系統可以準確的辨認出基本表情的情況。在設定判定標準為最高比例正確且大於 50% 並且第二高比例不大於 30% 時，表 4-1 的平均辨識率可達 93.33%。與其他使用 Cohn-Kanade 資料庫的研究[34]相比較，其他研究[34]的平均辨識率為 88.9%，比本研究的平均辨識率 93.33% 略低。

在辨識表情強度測試時，由表 4-5 可以觀察到在以測試結果與問卷調查結果在 1.5 個標準差之內作為判斷標準之時，平均辨識率可達 95.56%。在以測試結果與問卷調查結果在 1 個標準差之內作為判斷標準之時，如表 4-6 所示，平均辨識率仍有 80%。

在辨識混合表情比例測試時，由問卷調查中標準差的數值可以得知，每個人對表情比例的判斷頗有分歧，但我們比較問卷結果的平均與系統輸出可以觀察到問卷結果與辨識系統輸出有明確的正相關性。



第五章 結論與未來展望

5.1 結論

本論文提出了一個表情辨識系統，其能夠有效的辨識出互動者的人臉表情混合比例與表情強度。這套系統利用主動外觀模型(Active Appearance Model, AAM)擷取出人臉特徵值，接著使用我們所提出的根據面部動作編碼組合與表情關聯性的評分系統以及類神經網路輸出表情比例。最後使用我們所提出的根據人臉特徵值、類神經網路以及表情比例的表情強度辨識系統得到表情強度。

在基本表情辨識測試實驗中，平均辨識率可達 93.33% 比其他研究[34]的平均辨識率 88.9% 為高。在表情強度辨識測試實驗中，平均辨識率可達到 95.56%。而在混合表情比例辨識測試實驗中，問卷結果與系統輸出結果有明顯的正相關性。

因此我們總結出一個結論，本論文提出的這個具表情混合比例及強度之人臉表情辨識系統能夠有效辨識各種表情強度的人臉影像的表情比例以及強度。



5.2 未來展望

本論文提出的表情辨識系統仍有不盡完美之處，以下提出一些可以改進的地方：

在擷取特徵點的部分，在某些情況下會發生特徵點迭代錯誤的結果。錯誤主要發生於表情動作過大還有某些人臉會比較容易發生特徵點迭代錯誤，原因應該是在於其與樣本影像差別過大所導致。或許能夠用一些改良AAM的方法加以改進。

在辨識表情比例的部分，若是可以先個別辨識出臉部所有AU再搭配上一套新的評價機制。這套評價機制必須考慮各AU之間的關聯，以及AU與表情之間的關聯。由於AU可能的組合數目非常多，設計複雜度會相對提升不少，不過或許能夠讓辨識表情比例準確度更加提升。

參考文獻

- [1] 內政部戶政司全球資訊網(<http://www.ris.gov.tw/ch4/static/y1s400000.xls>)
- [2] Meng-Ju Han, Chia-How Lin and Kai-Tai Song, “Autonomous Emotional Expression Generation of a Robotic Face,” in *Proc. of IEEE SMC 2009*, St Antonio, TX, USA, 2009, pp. 2501-2506.
- [3] Peng Yang, Qingshan Liu, Xinyi Cui and Dimitris N. Metaxas, “Facial Expression Recognition Using Encoded Dynamic Features,” in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2008, Anchorage, Alaska, USA, pp.1-8.
- [4] Maria Pateraki, Haris Baltzakis, Polychronis Kondaxakis and Panos Trahanias, “Tracking of Facial Features to Support Human-Robot Interaction,” in *Proc. of IEEE International Conference on Robotics and Automation*, 2009, Kobe, Japan, pp.3755-3760.
- [5] 陳奕廷, 人臉辨識及表情辨識之整合設計, 碩士論文, 國立交通大學電機與控制工程學系, 2010.
- [6] A.M. Martinez and R. Benavente, *The AR Face Database*, CVC Technical Report #24, June 1998.
- [7] Akshay Asthana, Roland Goecke, Novi Quadrianto and Tom Gedeon, “Learning Based Automatic Face Annotation for Arbitrary Poses and Expressions from Frontal Images Only,” in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2009, Miami, Florida, USA, pp.1635-1642.
- [8] Michel Valstar, Brais Martinez and Xavier Binefa, “Facial Point Detection using Boosted Regression and Graph Models,” in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, San Francisco,

CA, USA, pp. 2729-2736.

- [9] Liya Ding and Aleix M. Martinez, "Precise Detailed Detection of Faces and Facial Features," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2008, Anchorage, Alaska, USA, pp.1-7.
- [10] Paul Ekman and Wallace V. Friesen, *Unmasking the face*, Prentice-Hill Inc. , Englewood Cliffs, New Jersey, USA, 1975.
- [11] Peng Yang, Qingshan Liu and Dimitris N. Metaxas, "Exploring Facial Expressions with Compositional Features," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, San Francisco, CA, USA, pp. 2638-2644.
- [12] Peng Yang, Qingshan Liu and Dimitris N. Metaxas, "Boosting Coded Dynamic Features for Facial Action Units and Facial Expression Recognition," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2007, Minneapolis, Minnesota, USA, pp. 1-6.
- [13] Zhi Zhang, Vartika Singh, Thomas E. Slowe, Sergey Tulyakov, and Venugopal Govindaraju, "Real-time Automatic Deceit Detection from Involuntary Facial Expressions," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2007, Minneapolis, Minnesota, USA, pp. 1-6.
- [14] Kai-Yueh Chang, Tyng-Luh Liu and Shang-Hong Lai, "Learning Partially-Observed Hidden Conditional Random Fields for Facial Expression Recognition," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2009, Miami, Florida, USA, pp.1-6.
- [15] Feng Zhou, Fernando De la Torre and Jeffrey F. Cohnz, "Unsupervised Discovery of Facial Events," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, San Francisco, CA, USA, pp. 2574-2581.

- [16] Tomas Simon, Minh Hoai Nguyen, Fernando De La Torre and Jeffrey F. Cohn, “Action Unit Detection with Segment-based SVMs,” in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, San Francisco, CA, USA, pp. 2737-2744.
- [17] Lifeng Shang and Kwok-Ping Chan, “Nonparametric Discriminant HMM and Application to Facial Expression Recognition,” in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2009, Miami, Florida, USA, pp. 2090-2096.
- [18] Paul Viola and Michael Jones, “Robust Real-time Face Detection,” in *Proc. of the Eighth International Conference On Computer Vision*, 2001, Vancouver, British Columbia, Canada, pp.747.
- [19] Yu-Li Xue, Xia Mao, and Fan Zhang, “Beihang University Facial Expression Database and Multiple Facial Expression Recognition,” in *Proc. of the Fifth International Conference on Machine Learning and Cybernetics*, 2006, Dalian, China, pp. 3282 – 3287.
- [20] 莊蕙如, 萃取表情重要特徵進行表情辨識與表情強度分析, 碩士論文, 國立清華大學資訊工程學系, 2009.
- [21] Hui Zhao, Zhiliang Wang, Jihui Men, “Facial Complex Expression Recognition Based on Fuzzy Kernel Clustering and Support Vector Machines,” in *Proc. of the Third International Conference on Natural Computation*, 2007, Haikou, Hainan, China, pp.562-566.
- [22] Paul Viola and Michael Jones, “Rapid Object Detection Using a Boosted Cascade of Simple Features,” in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001, Kauai Marriott, Hawaii, USA, pp. I-511-I-518.
- [23] Rainer Lienhart and Jochen Maydt. “An Extended Set of Haar-like Features for

- Rapid Object Detection,” in *Proc. of International Conference on Image Processing*, 2002, Rochester, New York, USA, pp. 900-903
- [24] C. C. Chiang, W. K. Tai, M. T. Yang, Y. T. Huang and C. J. Huang, “A Novel Method for Detecting Lips, Eyes and Faces in Real Time,” *Real-Time Imaging*, vol. 9, no. 4, pp. 277-287, 2003.
- [25] T. F. Cootes, G. J. Edwards and C. J. Taylor, “Active appearance models.” in *Proc. of 5th European Conference on Computer Vision*, 1998, Springer, Berlin, pp 484–498.
- [26] T. F. Cootes, G. J. Edwards and C. J. Taylor, “Active Appearance Models,” In *IEEE Transactions on Pattern Analysis And Machine Intelligence*, Vol. 23, No. 6, 2001, pp. 681-685
- [27] T. F. Cootes and C. J. Taylor, *Technical Report: Statistical Models of Appearance for Computer Vision*, The University of Manchester, School of Medicine, 2004.
- [28] I. Matthews and S. Baker, “Active Appearance Models Revisited,” *International Journal of Computer Vision*, vol. 60, no.2, 2004, pp.135-164.
- [29] C. Goodall, “Procrustes Methods in the Statistical Analysis of Shape,” *Journal of the Royal Statistical Society B*, Vol 53, no.2, pp.285-339, 1991.
- [30] 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.
- [31] I. Matthews and S. Baker, “Lucas-Kanade 20 Years On: A Unifying Framework,” *International Journal of Computer Vision*, Vol. 56, No. 3, 2004, pp. 221 - 255.
- [32] 羅華強, *類神經網路—MATLAB 的應用*, 高立圖書, 2005.
- [33] Kanade, T., Cohn, J. F., & Tian, Y. (2000). Comprehensive database for facial expression analysis. *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00)*, Grenoble, France, 46-53.

- [34] M. Obaid, R. Mukundany, R. Goeckezx, M. Billinghamurst and H. Seichter “A Quadratic Deformation Model for Facial Expression Recognition” in Proc. of Digital Image Computing: Techniques and Applications, 2009, Melbourne, Australia, pp. 264-270.

