

國立交通大學

電控工程研究所

碩士論文

基於時間強化設計之情緒辨識方法

Robust Emotion Recognition by Using a
Temporal-Reinforced Approach

研究生：林昭宇

指導教授：宋開泰 博士

中華民國一百零二年七月

基於時間強化設計之情緒辨識方法

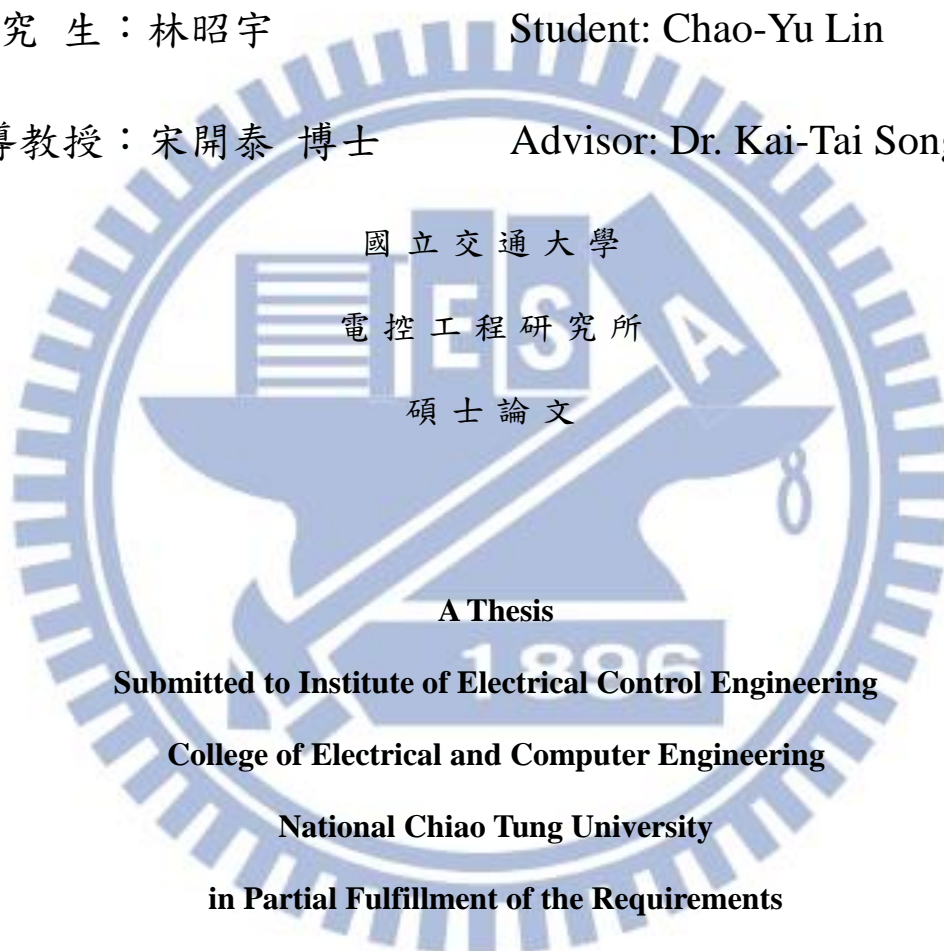
Robust Emotion Recognition by Using a Temporal-Reinforced
Approach

研 究 生：林昭宇

Student: Chao-Yu Lin

指 導 教 授：宋開泰 博士

Advisor: Dr. Kai-Tai Song



國立交通大學

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 2013

Hsinchu, Taiwan, Republic of China

中華民國一百零二年七月

基於時間強化設計之情緒辨識方法

學生:林昭宇

指導教授:宋開泰 博士

國立交通大學電控工程研究所

摘要

本論文之主旨在研究基於連續影像之情緒辨識方法，文中提出一套基於連續時間關聯資訊之情緒辨識與描述之方法。本方法首先透過主動外觀模型(Active appearance model, AAM)產生人臉影像樣本之形狀模型與紋理模型，擷取人臉特徵點及幾何特徵值，再由相關向量機(Relevance vector machine, RVM) 辨識情緒狀態。在辨識設計方面，本研究透過時序分析，辨識情緒類別之可能性(Likelihood)，並將辨識結果轉換至二維 Arousal 與 Valence 平面(Arousal-Valence Plane, A-V Plane)，以利系統之反應設計。所發展之方法能針對情緒程度、類別比例等資訊做更細微之辨識，且能夠分析情緒之轉變過程。經由實驗驗證，所發展之方法確能有效提升情緒辨識之效能，對基本表情之辨識率可達 95% 以上，對複雜情緒亦能做有效之辨識。為驗證本方法線上(on-line)辨識之效果，本論文設計一套基於人臉情緒辨識之智慧音樂選曲系統，此系統可藉由即時人臉情緒辨識，選取適當之音樂進行播放，透過音樂將使用者情緒逐漸導向至目標情緒。

Robust Emotion Recognition by Using a Temporal-Reinforced Approach

Student: Chao-Yu Lin

Advisor: Dr. Kai-Tai Song

Institute of Electrical Control Engineering
National Chiao Tung University

ABSTRACT

In this thesis, a temporal-reinforced approach to enhancing emotion recognition from facial images has been developed. Shape and texture models of facial images are computed by using active appearance model (AAM), from which facial feature points and geometry feature values are extracted. The extracted features are used by relevance vector machine (RVM) to recognize emotional states. In this work, we propose a temporal analysis approach to recognizing likelihood of emotional categories, such that more subtle emotion, such as degree and ratio can be obtained. Furthermore, a method is developed to map the recognition result to the Arousal-Valence Plane (A-V Plane). Experimental results verify that the performance of emotion recognition is enhanced by the proposed method. Furthermore, the A-V values are applied to an intelligent music selection system. With emotion recognition of current A-V values, appropriate songs are selected and played by this system to change a person emotion towards a target emotion.

誌謝

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與研究上的指導，並且教導我待人處事的態度與方法。他以豐富的學識與經驗，配合理論的應用，從論文題目、系統架構至實驗方法、實驗內容的完善性對我加以指導，使本論文得以順利完成。

感謝實驗室學長姊孟儒、嘉豪、格豪、巧敏、信毅、允智、育萱、建宏、仕晟、碩成、上峻、章宏、家昌與 Carlos，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京叡、俊儒與依穎的鼓勵及幫忙，以及學弟明翰、瑋哲、佑霖、奕夫與政輝不時提供幫助，並在生活與課業帶來樂趣。

最後，感謝我的父母與姊姊，由於父母的辛苦栽培，以及家人的細心關懷，並在過程中適時給我支持與鼓勵，使得我能夠以健康的身心持續研究，順利完成論文，在此願將此論文獻給我摯愛的家人。

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.....	4
1.2.1. 人臉情緒辨識方法.....	4
1.2.2. 程度、比例、連續時間人臉情緒辨識方法.....	6
1.2.3. 相關向量機原理.....	8
1.2.4. 可能性分析相關理論.....	11
1.3. 問題描述.....	13
1.4. 系統架構與章節說明.....	14
第二章、人臉偵測與特徵點擷取	16
2.1. 人臉偵測.....	16
2.1.1. 人臉區域決定.....	16
2.1.2. 人臉影像正規化.....	20
2.2. 主動外觀模型.....	20
2.3. 人臉形狀模型.....	21
2.3.1. 標註特徵點.....	21
2.3.2. 平均人臉形狀.....	22
2.3.3. 模擬臉部形狀變化.....	24
2.4. 人臉紋理模型.....	25
2.4.1. 人臉紋理.....	25
2.4.2. 分段仿射變形.....	26
2.4.3. 模擬臉部紋理變化.....	27

2.5. 影像校正演算法	28
2.5.1. Inverse Compositional 演算法	29
2.5.2. 整體形狀正規化轉換	31
2.5.3. 梯度影像修正	33
2.5.4. 直方圖等化	34
2.5.5. 影像校正整體方法	34
第三章、基於機率之連續影像情緒辨識	34
3.1. 人臉特徵擷取	36
3.2. 特定情緒辨識	40
3.3. 類別可能性	41
3.3.1. 類別可能性辨識	42
3.3.2. 類別可能性耦合	47
3.3.3. 類別可能性時序分析	50
3.4. 基於 A-V Plane 之情緒狀態辨識	53
第四章、實驗結果	61
4.1. 特徵點偵測結果	61
4.2. 基本情緒辨識結果	64
4.3. 混合情緒辨識結果	68
4.3.1. 問卷調查結果與辨識結果相關性評估	72
4.3.2. 整體情緒類別 RMSE 評估	72
4.3.3. 混合情緒類別組合評估	74
4.4. 基於 A-V Plane 之情緒辨識結果	74
4.5. 基於情緒辨識之音樂選取實驗	78
4.5.1. 音樂資料庫	78
4.5.2. 音樂選取設計	78
4.5.3. 音樂選取實驗結果	82
第五章、結論與未來展望	92
5.1. 結論	92
5.2. 未來展望	93
參考文獻	94
附錄一、基本人臉情緒混合比例問卷調查樣張	100

圖例

圖 1-1、寵物型機器人.....	2
圖 1-2、照護型機器人.....	2
圖 1-3、應用於健康照護之機器人(Aethon Tug)[12].....	3
圖 1-4、SVM 與 RVM 分類器對兩類資料分類之比較[28].....	10
圖 1-5、SVM 與 RVM 建立 Sinc 函式回歸模型之比較[28].....	10
圖 1-6、系統架構.....	15
圖 2-1、Viola 與 Jones 提出之 Haar-like 矩形特徵.....	17
圖 2-2、人臉偵測之矩形特徵範例.....	18
圖 2-3、Lienhart 與 Maydt 提出之 Haar-like 矩形特徵.....	18
圖 2-4、串接分類器架構圖.....	19
圖 2-5、雙線性內插示意圖.....	20
圖 2-6、特徵點選取原則示意圖.....	21
圖 2-7、7 種基本表情特徵點標註範例.....	22
圖 2-8、訓練平均人臉形狀範例.....	24
圖 2-9、平均人臉形狀.....	24
圖 2-10、Delaunay triangulation 方法對人臉切割示意圖.....	26
圖 2-11、三角形之形變座標轉換示意圖.....	27
圖 2-12、Inverse Compositional 演算法架構[36].....	30
圖 2-13、整體形狀正規化轉換形狀變化模型[36].....	32
圖 2-14、樣板人臉影像.....	34
圖 3-1、18 個人臉特徵點位置.....	37
圖 3-2、6 種基本人臉表情相關之 AUs.....	38
圖 3-3、情緒辨識架構圖.....	41
圖 3-4、情緒間 Likelihood 關係示意圖.....	50
圖 3-5、個別情緒 Likelihood 追蹤之有限狀態機.....	52
圖 3-6、Russell 提出之 A-V Plane.....	53
圖 3-7、A-V 參考值之決定.....	55
圖 4-1、CK+ 資料庫序列影像範例，由中性表情變化至驚訝表情.....	62
圖 4-2、7 種基本表情特徵點偵測結果.....	62
圖 4-3、序列影像中訓練 RVM 分類器與判斷情緒 Likelihood 影像樣本示意圖.....	64
圖 4-4、混合情緒測試樣本(Sample Index : 2).....	69
圖 4-5、混合情緒測試樣本(Sample Index: 3).....	69
圖 4-6、混合情緒測試樣本(Sample Index : 4).....	70
圖 4-7、混合情緒測試樣本(Sample Index : 7).....	70

圖 4-8、混合情緒測試樣本(Sample Index: 8)	71
圖 4-9、混合情緒測試樣本(Sample Index : 10)	71
圖 4-10、10 組序列人臉情緒比例調查結果整體均方根誤差(灰直方)，與辨識結果 之均方根誤差(黑直方).....	73
圖 4-11、10 組序列人臉情緒比例調查結果之情緒組合比例之和(灰直方)，辨識結 果對調查結果之情緒組合 Likelihood 之和(黑直方).....	75
圖 4-12、情緒從中性情緒轉換至 (a)生氣，(b)厭惡，(c)害怕，(d)高興，(e)傷心， (f)驚訝.....	76
圖 4-13、音樂資料庫之音樂分佈示意圖	79
圖 4-14、智慧型音樂選取之系統流程	79
圖 4-15、情緒與音樂關係示意圖	81
圖 4-16、情緒與音樂之決定與轉換	81
圖 4-17、情緒轉換情境設計，由不愉快情緒轉為愉快情緒之實驗	87
圖 4-18、情緒轉換情境設計，由興奮情緒轉為平靜情緒之實驗	91



表格

表 3-1、6 種基本人臉表情之 AUs 組合.....	38
表 3-2、6 種人臉表情之 AUs 組合定義.....	39
表 3-3、訓練樣本 2 情緒之間以 SRVM 辨識結果.....	46
表 3-4、訓練樣本 2 情緒之間以 DRVM 辨識結果.....	46
表 3-5、2 情緒之間分類器之選擇.....	47
表 4-1、18 個特徵點偵測平均<最大>誤差(單位：Pixel).....	63
表 4-2、Likelihood 時序辨識方法之基本情緒辨識率.....	65
表 4-3、Likelihood 非時序辨識方法之基本情緒辨識率.....	66
表 4-4、本研究提出之 Likelihood 時序辨識方法與相關研究之比較.....	67
表 4-5、混合情緒辨識與問卷調查結果(Sample Index：2).....	69
表 4-6、混合情緒辨識與問卷調查結果(Sample Index：3).....	69
表 4-7、混合情緒辨識與問卷調查結果(Sample Index：4).....	70
表 4-8、混合情緒辨識與問卷調查結果(Sample Index：7).....	70
表 4-9、混合情緒辨識與問卷調查結果(Sample Index：8).....	71
表 4-10、混合情緒辨識與問卷調查結果(Sample Index：10).....	71
表 4-11、情緒 Likelihood 辨識結果與問卷調查平均相關性.....	72
表 4-12、時序辨識方法之 A-V 轉換標準差.....	77
表 4-13、非時序辨識方法之 A-V 轉換標準差.....	77

第一章、緒論

1.1. 研究動機

近幾十年來，世界各國皆投入不少人力資源於機器人之研究，從早期於工廠輔助生產之機械手臂，到現今的軍事用機器人、服務型機器人等。不同的應用領域，對於機器人之設計，則需有不同之考量。人口老化與少子化，造成人力資源之短缺；又或是生活、工作忙碌導致小孩缺乏父母之陪伴、年長者缺乏家人之關心。許多研究因而著手於照護型機器人、寵物型機器人、娛樂型機器人等服務型、社交型機器人之發展，以年長者與小孩照護為目的做為機器人發展方向[1]。以機器人代替親友之角色陪伴年長者，代替家人之角色監控並回報年長者之健康狀況，或者是代替父母之角色照顧孩童。然而，諸如此類之人機互動領域，必須顧及社會對於機器人之接受度，以及機器人本身之實用性與安全性。讓機器人了解使用者之意圖、想法，讓使用者願意接近機器人，願意花時間與其相處，已是服務型、社交型機器人研究中不可或缺之問題。服務型、社交型機器人，往往為了提升人機互動之可靠度、接受度與實用性為目的，而做擬人化之設計。機器人利用感測器、資料處理之能力，以及外型之設計，使其具備類人或類動物之行為[2-6]，如圖 1-1 ~ 圖 1-2 所示。

本研究著重於人臉情緒辨識，以應用於人機互動之領域為研究重點。現今已有相當多的人臉情緒辨識相關研究，應用於各種不同之領域[7]。人臉為判斷使用者感受與互動相當直覺且重要的資訊。有些研究將人臉情緒辨識應用於客戶與銷售員之互動上，以提升銷售量[8]，例如：當客戶來到烤箱販賣區，系統辨識客戶看到哪些產品，有最大之正向情緒反應，利用此資訊進一步使銷售員選擇行銷手法，說服客戶購買產品，或提供店家進行特定產品之促銷。相同的情況，也



(a)



(b)

圖 1-1、寵物型機器人 (a) AIST 研發之 Paro[2], (b) Sony 研發之 AIBO[3]。



(a)



(b)

圖 1-2、照護型機器人 (a) Gecko 研發之 CareBot[4], (b) Robosoft 研發之 Kompai 機器人[5]。

可應用於網路商店，藉由情緒辨識系統偵測到當客戶瀏覽特定產品時產生高度興趣，而進一步提供相關產品類別資訊；有些學者則提出將人臉情緒辨識應用於電視廣告對客戶吸引力之分析[9]，例如：在球場電視牆播放廣告時，球隊經營公司利用辨識系統判斷球迷對此廣告的反應，喜悅?驚訝?感到無趣?忽略?利用諸如此類之識別，提供公司改善經營策略。

將人臉情緒辨識系統應用於人與機器人互動上之研究則相對來得少。Robosoft[5]之 Kompai robot，如圖 1-2(b)所示，目前主要是利用語音、觸碰界面與使用者做溝通，在未來則計畫加入與使用者互動為導向之機器人視覺能力，其中結合機器人臉之設計與人臉情緒辨識系統，使機器人與使用者互動時產生適當

之表情變化與情緒反應，即為其發展方向之一[10]；另外，Swangnetr 與 Kaber 之研究提到[11]，在醫療照護方面，由於護理師人力資源日漸短缺，過去 20 年間，像是醫院、護理之家等健康照護相關產業，已引進移動能力之機器人(例如：Aethon Tug [12]，如圖 1-3 所示)，長期協助護理師照護病患。然而，此類機器人目前主要的工作僅限於二地間藥物、餐點、床單等物品之傳遞，且未與病患直接做接觸，仍然必須依賴護理師做為機器人與病患間之橋樑。學者探討未來將藉由照護型機器人直接與病患做接觸，真正減少人力資源，減少護理師的工作量。如此則必須使機器人具備偵測與回應病患狀態(例如：生理狀態、情緒狀態)之能力，以達到正向照護之功效。

描述使用者情緒，包含二種常見之方法，分別為情緒類別分類，以及將情緒映射至 2D 或 3D 之空間，做情緒程度數值分析[13]。情緒類別分類：主要是將情緒分為基本、被普遍認可之類別，一般而言可分為生氣、厭惡、害怕、高興、傷心、驚訝與中性之 7 種基本情緒[14]。基本表情分類，為目前情緒辨識相關研究最普遍之方法；情緒程度數值分析：有些學者認為每天人與人之間之互動，通常為非基本、微妙之情緒變化，而包含複雜之情緒狀態(例如：沉思、尷尬或沮喪)[15]。如此微妙、複雜之情緒狀態，必須藉由多種臉部表情類別、肢體動作、聲音或生理訊號所構成。單一類別之分類，則無法傳達多種資訊所構成之複雜情緒狀態。因此利用不同空間維度，以數值方式對情緒做程度、比例等描述，能夠



圖 1-3、應用於健康照護之機器人(Aethon Tug)[12]。

探索微妙、複雜、連續之情緒變化，有系統的將情緒狀態做結合，而非獨立之類別定義。

本研究設計重點在於使機器人根據使用者情緒呈現適當之反應。如上所述，即使機器人設計多種程度之情緒呈現，但未能提供適當之情緒資訊做為輸入，仍無法達成其功效。也就是說，若只將人臉表情分類成 7 類基本情緒，機器人所呈現之情緒反應也會侷限於七種，限制互動人性化之程度。在設計上情緒辨識系統必須對情緒程度加以判斷，使機器人能夠對不同程度之情緒呈現適當反應，並隨著使用者情緒之轉換而轉變互動狀態。本研究以 Russell [15]所提出之心理學模型(2D model, Arousal-Valence model)為根據，進行情緒 Arousal 與 Valence(A-V)之分析，合理描述情緒間之程度與不確定性，以及情緒間之轉換，做為機器人行為反應呈現之依據。能夠表達機器人不同程度反應之方式有相當多種，像是機器人表情程度，語調高低、強弱，或播放不同種類之音樂使人心情高興等設計。本研究設計一套情緒辨識之音樂選取系統，藉由系統辨識使用者情緒，決定播放之音樂類型，依據不同場合，設法將使用者情緒轉換至適當之目標情緒，以達到人機互動之目的。

1.2. 相關研究回顧

本章節將針對一般人臉情緒辨識方法，以及以程度、比例、連續時間辨識人臉情緒方法做相關研究之深入探討。另外，本章節亦會對相關向量機原理，與可能性(Likelihood)辨識相關理論做說明。

1.2.1. 人臉情緒辨識方法

Ekman 與 Friesen 之研究指出[14]，所有複雜之情緒皆可藉由生氣、厭惡、害怕、高興、傷心、驚訝與中性 7 種基本情緒所構成，現今情緒辨識之研究亦多

以區分這七種情緒為研究重點。

一般人臉情緒辨識相關研究，以提升基本情緒類別之辨識率為主要訴求，往往著手於特徵擷取之方法，或分類器之設計進行研究探討[16 - 17]。Gu 等人之研究 [16]，透過人眼視覺之分析，提出多層分類器組合之情緒辨識架構。作者將輸入影像切割為 49 個區塊，每個區塊皆有 50% 與相鄰區塊重疊，利用不同大小與方向之賈伯濾波(Gabor filter)對影進行處理，並對處理過後之影像，模仿人類視網膜之結構做放射狀編碼。此碼利用主成份分析(Principal component analysis, PCA)與 Fisher 線性判別(Fisher linear discriminant, FLD)進行分析並結合 K 最近鄰法(K-nearest neighbor, KNN)進行分類產生區域特徵，最後再進行整合產生全域特徵做為情緒類別決定之用。此研究除了能夠對一般完整之人臉進行辨識，亦能夠對遺失部份資訊之人臉(例如：眼睛或嘴巴被遮蔽)做有效分類。然而，此研究必須有準確之人臉區域偵測，即影像必須只包含人臉資訊而盡可能的移除背景資訊，否則無法取得準確之辨識特徵。Song 等人[17]提出結合皮膚形變參數(Skin deformation parameters, SDPs)與臉部動態參數(Facial animation parameters, FAPs)之分析，擷取影像強度比例特徵與位移特徵判斷人臉形變，並利用支持向量機(Support vector machine, SVM)進行情緒分類。作者以主動形狀模型(Active shape model, ASM)擷取人臉特徵點，以部份特徵點所圍出之區域影像，做為臉部紋理分析之範圍。紋理特徵之擷取主要是計算測試樣本與中性表情二者區域影像之強度比例，而幾何特徵則是計算測試樣本相對於中性表情特徵點距離之位移比例(例如：嘴巴寬度變化)，研究提出之影像比例特徵有效克服情緒辨識時光線之影響。然而，此方法在應用時必須預先知道使用者之中性表情，方能進行特徵之擷取。

同樣以提升基本情緒類別辨識率為目的，Hong 等人[18]則是提出一套能夠學習新臉孔之情緒辨識系統，使系統能夠適應新人臉之表情。作者以 Gabor filter 為基礎擷取 16 個距離特徵，做為情緒辨識之特徵，並以 SVM 做情緒類別之訓

練。研究中作者提出重新學習新特徵之訓練機制，利用重新訓練錯誤之表情資料與一定數量之關鍵舊集合，以調整 SVM 切割平面，產生新的 SVM 分類器。此方法不僅可以辨識原本無法辨識之新臉孔，且能夠保持對舊有資料之辨識率。

諸如上述之情緒辨識系統，皆以人臉表情對使用者情緒進行 5~7 種類別之辨識，提出不同的特徵擷取之方法、情緒特徵之選擇，或分類器之設計與訓練，以提升情緒辨識之準確性。然而，實際應用時，使用者之情緒往往包含不同的程度，甚至在同一時間點，人臉表情可能具有多種情緒之混合(例如：又驚又喜)，以有限之類別則不足以描述人臉情緒之微妙變化。

1.2.2. 程度、比例、連續時間人臉情緒辨識方法

為了更精準的描述使用者之情緒，有些研究進一步判斷使用者人臉表情之情緒程度、比例。Song 與 Chien [19]提出一套辨識人臉表情基本情緒類別比例與強度之方法。作者以主動外觀模型(Active appearance model, AAM)訓練人臉影像之形狀與紋理模型，以建立人臉模型，並由紋理模型擷取紋理特徵參數，做為情緒辨識之特徵。此研究將臉部分為全臉、上半臉，與下半臉，分別利用紋理特徵參數搭配倒傳遞類神經網路(Back propagation neural network, BPNN)之設計，進行強度辨識、上半臉 Action units(AUs)組合辨識，與下半臉 AUs 組合辨識，並給予評價分數。作者提出整合評價分數設計辨識得到人臉表情之情緒強度與比例。另外，有些研究則將人臉情緒映射至心理學家所提出之 2D 或 3D 模型[15], [20 - 21]，同樣對情緒做程度之分析辨識。Hupont 等人之研究[22]利用 Whissell[20]所提出之 2D 模型，對人臉情緒做分析，Whissell 提出之 2D 模型以 Evaluation 與 Activation 所構成，用以描述情緒更細部之狀態。此研究以臉部之特徵點距離與角度做為辨識特徵，作者藉由分析 5 種分類器對於不同情緒之辨識效能，附予各分類器相對於各種情緒辨識之權重，並提出一套演算法結合 5 種分類器，計算基本情緒類別之信心指數。最後研究以基本情緒在 2D 模型所在位置之 Evaluation 與 Activation

值做為權重，與信心指數做加乘，將基本情緒辨識結果映射至 2D 模型，取得更細微之情緒狀態。

諸如上述之情緒辨識相關研究，主要以單張人臉影像樣本，對情緒進行程度、比例，或結合心理學模型描述細微情緒狀態之辨識。然而，有些研究則提出藉由序列、連續影像之辨識，能夠提升系統之穩健性與辨識準確性。Yang 等人之研究[23]認為 Harr-like 特徵做為人臉情緒辨識特徵，能夠最簡單的擷取，達到最有效之辨識。研究擷取數張序列影像之 Harr-like 特徵，考量特徵在序列影像中之時序關係，並以二進制之數值(0, 1)進行編碼，藉由序列影像之編碼組合，做為基本情緒之辨識特徵。此研究以 Boosting 學習法，學習情緒之特徵，建構情緒辨識分類器，辨識使用者情緒。Hommel 與 Handmann 之研究提到[24]，每個人對於特定情緒之呈現方式不盡相同，擁有個人獨特之特徵，容易影響情緒辨識效能，故此研究以連續時間之分析，計算個人特定之平均人臉，以平均人臉消除 AAM 形狀與紋理不必要之特徵參數。作者以 SVM 進行回歸分析，將情緒映射至 1D 空間中，辨識人臉情緒之正向程度。另外，作者亦提到，在序列影像中，人臉情緒不可能產生瞬間之變化，因此在建立回歸模型(Regression model)時，研究合理的消除瞬間之誤差。此研究在進行即時辨識時，個人特定之平均人臉則必須預先做定義。Zhang 與 Ji[25]提出動態貝式網路(Dynamic Bayesian networks, DBNs)架構，分析人臉表情序列影像之時間關係。作者以 Ekman 提出之 Facial Action Coding System (FACS)，有系統的選擇人臉表情幾何特徵與紋理特徵，其系統機率架構包含視覺資料層(Visual data layer)、AU 層(AU layer)，與分類層(Classification)三層，並結合影像之時序做情緒 Likelihood 辨識，除了擷取當前測試樣本之人臉特徵，進行人臉 AUs 之辨識，再利用 AUs 之組合結果辨識情緒之 Likelihood，亦與先前之辨識結果結合。研究中，作者以統計方式取得人臉特徵與 AUs 之機率關係，並依據 AUs 機率，利用最大可能性估測(Maximum likelihood estimation, MLE)決定分類層中情緒類別之 Likelihood 參數，以辨識使用者情緒之

Likelihood。Nicolaou 等人[26]提到在多維情緒辨識時，情緒變化包含時間與空間相依之關係在輸出向量間，以及輸入與輸出關係之間。研究提出輸出聯想之相關向量機(Output-Associative Relevance vector machine, OA-RVM)回歸分析架構，OA-RVM 透過預先定義之 Temporal window，分析過去、現在、未來情緒間之關係，在每次辨識時依據 window 之大小，將前後數個 RVM 之輸出，即 A-V，與現在之幾何輸入特徵做結合，共同辨識當前情緒之 A-V，使得輸入與輸出具有線性之相依性。

無論在辨識基本情緒類別，或情緒程度、比例之關係，藉由連緒影像之分析，能夠從影像中擷取更多人臉資訊，考量影像樣本時序間之關係，有效提升情緒辨識之正確性。由於連續影像之情緒辨識，系統需要同時處理數張影像進行分析，對於即時情緒辨識，處理速度則需納入考量之中。

1.2.3. 相關向量機原理

核方法(Kernel method)常被用來做為機器學習(Machine learning)之設計，用來解決分類、回歸分析等相關研究問題，Kernel method 主要概念在於將無法分割之低維度樣本特徵，試圖藉由核函數(Kernel function)將樣本特徵映射至較高維度之特徵空間，使樣本能夠有效被分類，常見之 Kernel method 包含支持向量機(Support vector machine, SVM) [27]、相關向量機(Relevance vector machine, RVM)[28]與高斯過程(Gaussian Processes, GP)。這個領域之相關研究往往著手於研究、改良現有之機器學習方法，以提高其準確性、運算速度等為目的[18], [26]，其中在生物辨識中又以 SVM 與 RVM 較為常見。支持向量機(Support vector machine, SVM)為目前最為普遍、受到重視的方法之一，SVM 以最小風險，即最大化兩類交界(Margin)之原則學習 Kernel 參數，對兩類訓練樣本利用學習取得的決定邊界(Decision boundary)進行切割，如此學習法則能有效避免過度訓練(Over-fitting)之問題，其僅挑選接近 Decision boundary 之訓練樣本做為 SVM 的

支持向量(Support vector, SV)，亦即僅使用到部份的 Kernel function，故其解相對來得稀疏，而其分類準確性或運算速度，往往相較於一般分類、分群演算法(例如：類神經網路)還要來得高。雖然，SVM 在實作上具備即時分類、準確性高等優點，但其在應用上仍包含些許問題[28]。

- 雖然 SVM 取得相對稀疏之解，其 SV 數量明顯少於訓練樣本數，但 SVM 仍會因大量訓練資料時，雜訊與類別訓練樣本重疊，而導致 SV 成線性成長，往往造成 Over-fitting，除了使 Decision boundary 運算複雜度變高，且訓練過程中，亦會將與訓練樣本重疊之雜訊一併訓練，雖然完美的對訓練樣本做分類，但對於測試樣本則可能分類錯誤。
- SVM 僅可做類別之判定，而缺乏機率之預測，無法容忍類別間不確定性之問題，可能造成錯誤分類。
- 由於 SVM 對雜訊相當的敏感，在求解時嚴格限制所有訓練樣本必須完美的被切割。利用誤差參數 C 可以放寬限制，然而，此參數必須自行調整，參數之決定對結果將造成相當大的影響。
- Kernel function 必須符合梅西定理(Mercer's condition)，其必須為連續對稱正整數之 Kernel。

Tipping[28]提出之相關向量機(Relevance vector machine, RVM)，則改良 SVM 之上述問題。RVM 藉由貝式(Bayesian)結構進行 Kernel 參數之學習，RVM 選取最能代表特定類別之訓練樣本做為 RV，而非最接近 Decision boundary 之訓練樣本，其 RV 不僅更為稀疏，使得運算速度大幅度的下降，且保有 SVM 分類之準確性。如圖 1-4 為 SVM 與 RVM 分類之比較，圖中包含兩類之訓練資料，共 100 筆訓練樣本，SVM 與 RVM 皆選擇 Gaussian kernel，即 Radial basis function(RBF) kernel 為基礎做學習，由圖可看出 RVM(圖 1-4(b))之 RV(4 個)明顯少於 SVM 之 SV(38 個)，且 RVM 分類結果仍不亞於 SVM。如圖 1-5 為 SVM 與 RVM 回歸分析之比較，SVM 與 RVM 皆選擇 linearspline kernel 為基礎做學習，建立 Sinc 函

式之回歸模型，其訓練樣本由 Sinc 函式加入 $\epsilon = \pm 0.01$ 之誤差取得，同樣可看出 RVM 回歸模型(圖 1-5(a))之 RV (6 個)少於 SVM 回歸模型(圖 1-5(b))之 SV(29 個)，且二者所建立之回歸模型皆相當接近 Sinc 函式，SVM 與 RVM 回歸模型均方根誤差(Root-mean-square-error, RMSE)分別為 0.0291 與 0.0245。另外，RVM 之另一特性，則是可以做機率之預測，對不確定之樣本，做較為合理之辨識。

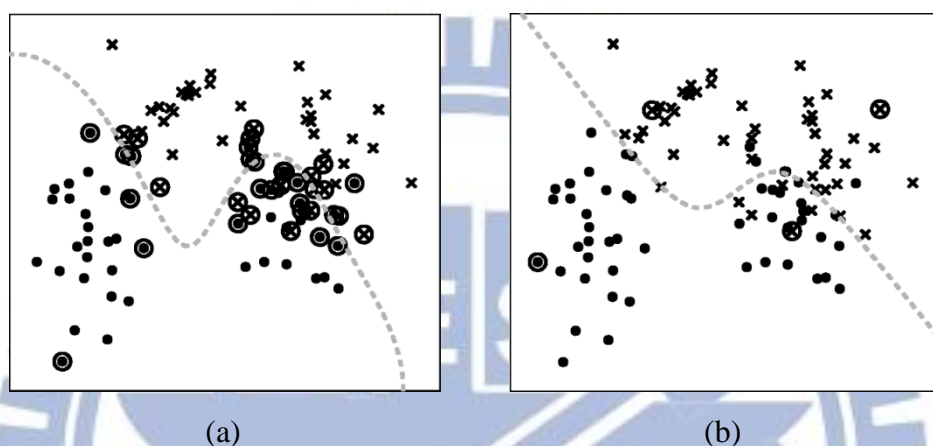


圖 1-4、SVM 與 RVM 分類器對兩類資料分類之比較[28]，(a)SVM 之學習結果，共包含 38 個 SV，(b)RVM 之學習結果，共包含 4 個 RV。

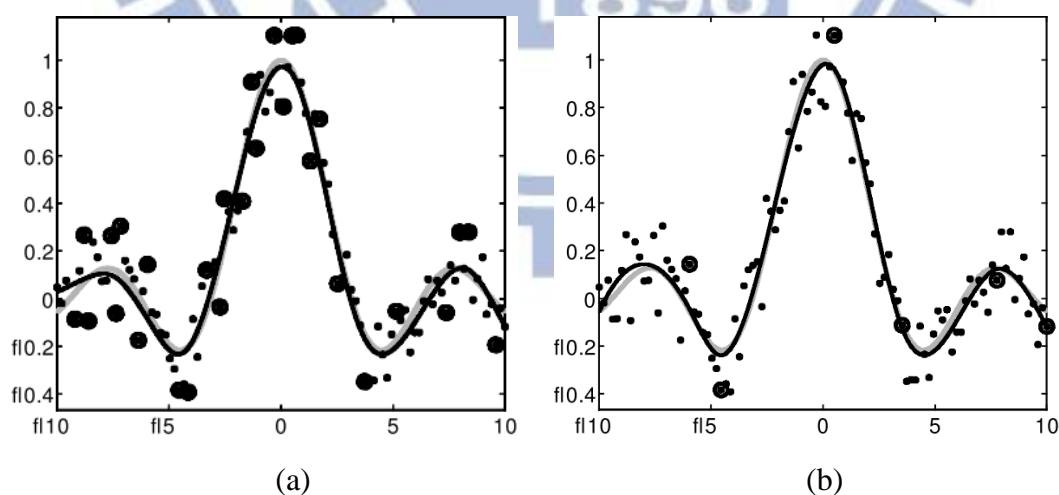


圖 1-5、SVM 與 RVM 建立 Sinc 函式回歸模型之比較[28]，(a)SVM 之回歸模型，共包含 29 個 SV，其 RMSE 為 0.0291，(b)RVM 之回歸模型，共包含 6 個 RV，其 RMSE 為 0.0245。

RVM 屬於監督式學習法(Supervised learning)，Supervised learning 給定一組特徵向量 $\{\mathbf{x}_n\}_{n=1}^N$ 與其對應之輸出 $\{t_n\}_{n=1}^N$ 做訓練。若應用於特徵分類，則 t_n 即為類別標籤；若應用於回歸分析，則 t_n 為任意實數。訓練之主要目的，在於學習一個模型，以預測新特徵向量之輸出。實際應用上，若應用於資料分類，必須克服類別間特徵重疊(over lap)之問題；若應用於回歸分析，則必須注意雜訊之問題。

一般而言，Supervised learning 使用如(1-1)之基本形式，描述輸入特徵向量與輸出間之關係，而所謂訓練即是在學習(1-1)中權重 \mathbf{w} 之過程。

$$y(\mathbf{x}; \mathbf{w}) = \sum_{i=1}^M w_i \phi_i(\mathbf{x}) = \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}), \quad (1-1)$$

其中 輸出 $y(\mathbf{x}; \mathbf{w})$ 為 權重 $\mathbf{w} = (w_1, w_2, \dots, w_M)^T$ 與 基底函式 (Basis function) $\boldsymbol{\phi}(\mathbf{x}) = (\phi_1(\mathbf{x}), \phi_2(\mathbf{x}), \dots, \phi_M(\mathbf{x}))^T$ 之線性組合。RVM 使用與 SVM 相同之函式，將輸入與輸出之關係，由(1-1)改寫為(1-2)，

$$y(\mathbf{x}; \mathbf{w}) = \sum_{i=1}^N w_i K(\mathbf{x}, \mathbf{x}_i) + w_0, \quad (1-2)$$

其中 Basis function 由 Kernel function, $K(\mathbf{x}, \mathbf{x}_i)$ 做定義。RVM 藉由貝式機率架構(Bayesian probabilistic framework)學習(1-2)權重，以建立輸入與輸出之模型。其主要目標在於建立一個盡可能稀疏之模型，在學習過程中大多數之權重被設為零，僅保留能夠有效分類、相對重要之 Kernel function，其學習演算法於本論文 3.3.1 節詳細介紹。

1.2.4. 可能性分析相關理論

在無法確定樣本之類別時，常會以 Likelihood 做分析，判斷樣本在每一類之可能性，使得樣本類別之描述更為精準[25]。Likelihood 辨識藉由對訓練樣本進行統計推論，推論過程假設訓練樣本遵循某個已知之分佈模型(例如：Gaussian)，並對訓練樣本進行參數化(Parametric)。最後，利用最大可能性估測(Maximum

likelihood estimation, MLE)取得訓練樣本之分佈參數，如此估測之參數將使樣本之估測分佈附加於假定分佈模型之上，以此分佈做為決策(Decision)模型，以決定測試樣本之 Likelihood [29]。

給定一組獨立相同分佈(independent and identically distributed, iid)之訓練樣本 $X = \{x^t\}_{t=1}^N$ ，並假設 x^t 屬於 $p(x|\theta)$ 之分佈，即 $x^t \sim p(x|\theta)$ 。由於 x^t 各自獨立，因此由樣本 X 所構成參數 θ 之 Likelihood，可以視為各個點 Likelihood 之乘積，如(1-3)所示，

$$l(\theta|X) = p(X|\theta) = \prod_{t=1}^N p(x^t|\theta), \quad (1-3)$$

藉由 MLE 估測(1-3)之參數 θ ，以取得 x^t 之分佈，使 x^t 之分佈盡可能的接近 $p(x|\theta)$ 。為了計算方便，通常會將(1-3)以 \log 形式表示，其所求得之 θ 仍保持不變，如(1-4)所示，

$$L(\theta|X) = \log l(\theta|X) = \sum_{t=1}^N \log p(x^t|\theta). \quad (1-4)$$

Likelihood 辨識問題中最常見之假定分佈包含伯努利分佈 (Bernoulli distribution)、多項式分佈 (Multinomial distribution)，與高斯分佈 (Gaussian distribution)三種。Bernoulli distribution 主要用來解兩類辨識之問題，其僅用來判斷 2 者發生之機率，以(1-5)表示，

$$P(x) = p^x(1-p)^{1-x}, x \in \{0, 1\}, \quad (1-5)$$

其中 p 代表某一事件發生之機率； $1-p$ 代表另一事件發生之機率。Multinomial distribution 可視為 Bernoulli distribution 之概括，其考量多類發生之機率，且所有類別發生之機率和為 1，以(1-6)表示，

$$P(x_1, x_2, \dots, x_K) = \prod_{i=1}^K p_i^{x_i}, \quad (1-6)$$

其中 當結果代表狀態 i 之機率，則 x_i 為 1，否則為 0。Gaussian distribution 又

稱為常態分佈(Normal distribution)，為最常被用來做為假定之分佈。假設 Gaussian distribution 之平均為 $E[X] = \mu$ ，變異數為 $\text{Var}(X) = \sigma^2$ ，記作 $N(\mu, \sigma^2)$ ，Gaussian distribution 以(1-7)表示，

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], -\infty < x < \infty, \quad (1-7)$$

實作時，藉由 MLE 解得上述 3 者假定分佈之參數，取得訓練樣本之估測分佈。本論文 1.2.3 節提到 RVM 能夠對類別做機率式之預測，即是以 Bernoulli distribution 做為訓練樣本假定分佈，估測樣本分佈參數，取得類別之 Likelihood，其細節於本論文 3.3.1 節詳細介紹。

1.3. 問題描述

大多數情緒辨識之相關研究中，主要以區分基本情緒類別為目的做辨識系統之設計。事實上，一般情況下，人臉表情產生基本情緒之頻率相當的低，其通常具有微妙的情緒變化，或多種情緒之組合，僅使用基本情緒類別做分類，則不足以精確描述使用者情緒，亦會影響人機互動之效能，使互動之形式受限於基本情緒類別。基於上述之觀點，情緒辨識時應加入情緒程度、比例對情緒做分析。常見之情緒描述方法，則是將人臉情緒映射至心理學情緒 2D 模型，藉由心理學模型之映射，能夠呈現使用者不同的情緒狀態，例如：Russell[15]所提出之心理學情緒 2D 模型(A-V Plane)，能夠藉由代表放鬆與興奮程度之 Arousal，與代表愉快與不愉快程度之 Valence，對使用者情緒做更細微之辨識。

另外，由於人臉情緒變化相當微妙，有時甚至可能受到環境光線、人臉角度或人臉表情本身之不確定性等因素影響，僅使用單張影像對情緒做辨識，可能造成情緒之誤判，導致辨識系統對於人臉情緒轉換無法做出合理之判斷。現今攝影裝置通常能夠在 1 秒內擷取數十張影像，在考量辨識系統處理速度，與妥善之演算法設計下，以連續影像為依據，進行情緒分析，能夠取得更多辨識資訊，提升

辨識之效能。

綜合以上所述，為了使人機互動系統更為人性化、更被使用者所接受，故必須考量不同情緒之程度、比例，設法將情緒映射至心理學情緒 2D 模型，對情緒做精準之辨識；另外，由於人機互動系統通常是根據使用者情緒之轉換而轉換互動模式，故辨識系統必須對使用者情緒之轉換做合理判斷，在處理速度考量設計下，以連續影像進行時序分析對情緒做辨識，提升情緒辨識之準確性，與互動之效能。

1.4. 系統架構與章節說明

本研究提出之系統架構如圖 1-6 所示，可略分為人臉影像前處理、人臉特徵擷取、連續時間情緒程度辨識三個部份，其中在連續時間提取適當之特徵，判斷使用者情緒程度與轉換，為研究重心所在。

本研究以 Viola[30]與 Lienhart[31]提出之 Harr-like 人臉偵測，擷取影像中人臉之位置，取得人臉區域影像。根據人臉區域影像，以 AAM 演算法[32 - 34]建出人臉模型，並擷取其中之特徵點，計算人臉情緒之距離特徵與位移特徵，做為情緒辨識之幾何特徵。辨識過程中，藉由擷取之幾何特徵，結合時序之分析，辨識情緒之 Likelihood，並配合 Russell[15]提出之 A-V Plane，進行情緒 A-V 之辨識，分析使用者目前情緒之程度與其轉換過程。A-V Plane 上之 A-V 辨識結果將做為人機互動系統互動轉換之依據，使互動系統與使用者之互動更為合理、更為人性化。

本論文共分為 5 章節，第一章為緒論，介紹本研究之研究動機、相關研究、問題描述，以及系統架構；第二章為人臉偵測與特徵點擷取，說明本研究感興趣之人臉特徵點擷取方法；第三章為基於機率之連續影像情緒辨識，此章節介紹本研究提出之特定情緒辨識、Likelihood 時序分析，以及 Likelihood 轉換至 A-V 方法，做為智慧型音樂選取之依據；第四章為實驗結果，包含特徵點偵測結果、情

緒辨識結果，與音樂選取實驗；最後，第五章為結論以及未來展望。

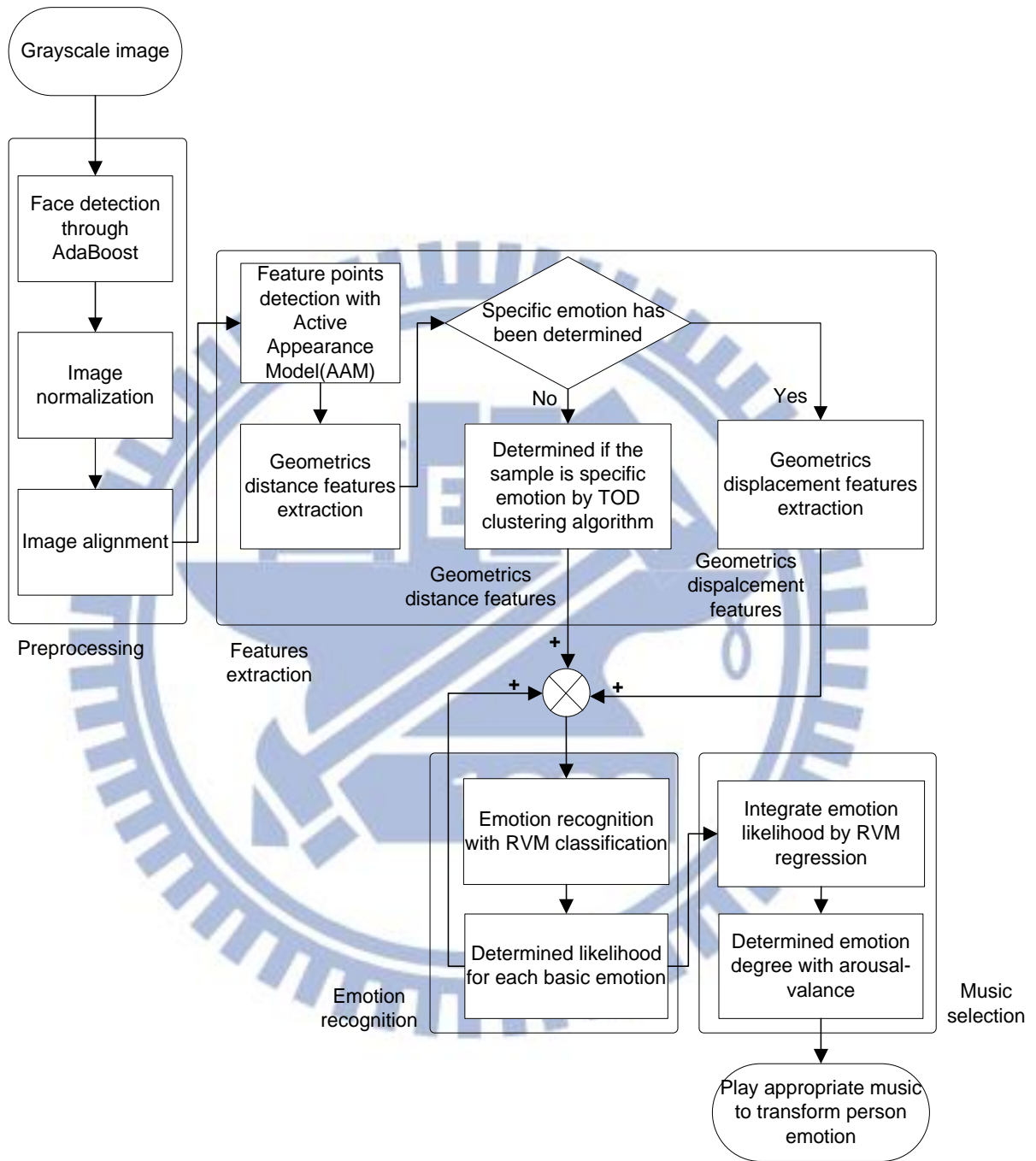


圖 1-6、系統架構。

第二章、人臉偵測與特徵點擷取

特徵點偵測之準確性，對於情緒辨識之正確率有絕對的影響。本研究藉由 Adaboost[30]演算法偵測影像中之人臉，以擷取人臉區域，減少背景雜訊之影響，進而對人臉影像進行特徵點偵測與情緒辨識。特徵點偵測部份，本研究則是利用主動外觀模型(Active appearance model, AAM) [32 - 34]建立人臉模型，以擷取其中之特徵點，進行情緒辨識特徵之計算。本章節將說明人臉偵測之 Adaboost 演算法，以及特徵點偵測之 AAM 演算法。

2.1. 人臉偵測

本研究之人臉偵測包含人臉區域決定，與人臉區域正規化二個部份。藉由攝影機擷取解析度為 640×480 之影像，偵測其中之人臉，以取得人臉區域。為了避免人臉大小遠近影響擷取之特徵，本研究對人臉區域進一步正規化至 200×240 之人臉影像。

2.1.1. 人臉區域決定

本研究人臉區域之決定，使用 Viola 與 Jones 所提出名為 AdaBoost[30]之方法從輸入影像中，取得人臉區域範圍，此範圍即為情緒辨識特徵擷取之依據。AdaBoost 演算法主要概念在於透過多個弱分類器組成一個強分類器，藉由數個分類特徵對目標物做分析，分析哪些特徵能夠清楚分類目標物，分類目標物正確率愈高之特徵則被選定為弱分類器，經由數個弱分類器之組合，若分類目標物之正確率達到需求以上，則停止尋找弱分類器。因此欲利用 AdaBoost 演算法分類人臉與非人臉，則必須找出適當之輸入特徵。

a. Haar-like 矩形特徵：

利用矩形特徵可以簡單描述圖形的結構，當然也可以用來當作檢測人臉之特徵模版，最早由[30]所提出之矩形特徵，如圖 2-1 所示。矩形特徵主要藉由分析物體表面亮度變化會造成明亮區塊與灰暗區塊，對物件做偵測。因此利用已知偵測目標之矩形特徵與目標做比對，可用來判斷是否為目標物。應用於人臉偵測方面，矩形特徵可以分析臉部之陰影變化，有效分類人臉與非人臉，例如：眼睛與眼皮產生之陰影如圖 2-2(b)所示，鼻梁與眼窩產生的陰影如圖 2-2(c)所示，又或者是嘴巴與其周圍產生的陰影如圖 2-2(d)所示等，亦即可利用白色區域像素與黑色區域像素的差值，做為偵測人臉特徵。

Lienhart 與 Maydt[31]將[30]所提出之矩形特徵推廣至旋轉 45 度之 Haar-Like 矩形特徵，共包含 15 個矩形特徵模版如圖 2-3 所示，本研究以此特徵做為人臉偵測之依據，Adaboost 訓練與分類之方法如下所述。

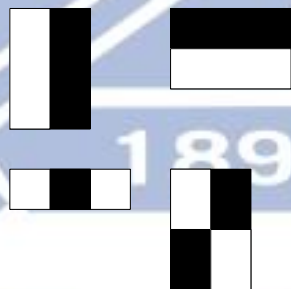


圖 2-1、Viola 與 Jones 提出之 Haar-like 矩形特徵。



(a)



(b)



圖 2-2、人臉偵測之矩形特徵範例。(a)原始人臉影像樣本，(b)眼睛與眼皮灰階值色差，(c)眼窩與鼻樑灰階值色差，(d)嘴巴周圍灰階值色差。

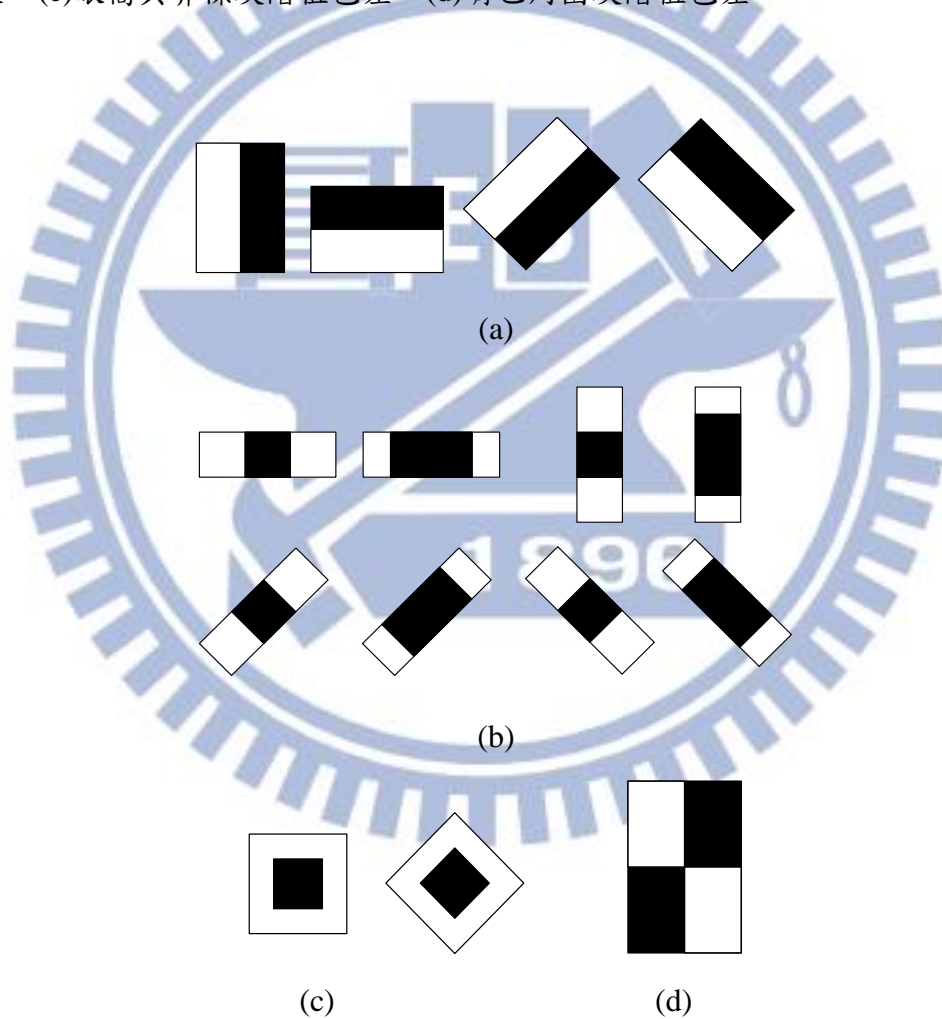


圖 2-3、Lienhart 與 Maydt 提出之 Haar-like 矩形特徵。(a)邊緣特徵，(b)線特徵，(c)中心環繞特徵，(d)特殊對角特徵。

b. Adaboost 訓練演算法：

Adaboost 之訓練可從一個很大的特徵，集中選擇很小一部分之關鍵特徵，產生有效之分類器。透過提高數個簡單之弱分類器，最後組成強分類器，其訓練錯誤率可以說幾乎為零，具備良好之推廣性。

Adaboost 應用於人臉偵測方面，需要有足夠數量之樣本，進行人臉特徵之搜尋。此處樣本影像分為人臉與非人臉兩種，非人臉影像數量需大於人臉影像數量，使搜尋之結果收斂到辨識率較佳之特徵上。

c. 串接分類器：

Adaboost 演算法雖然可藉由強分類器，有效區分人臉與非人臉，但事實上大多數的影像皆為非人臉，且大部份非人臉皆可利用更簡單的方法做區別，故並非所有影像皆需由強分類器做人臉與非人臉之判斷。

由[30]所提出的串聯式偵測架構，如圖 2-4 所示，透過分級分類器對影像逐步做分析，級數愈前面，可用愈少的計算量排除大多數非人臉，級數較後者則愈難判斷人臉與非人臉，必須使用強分類器做分類，如此之系統架構則可以有效提高辨別速度。

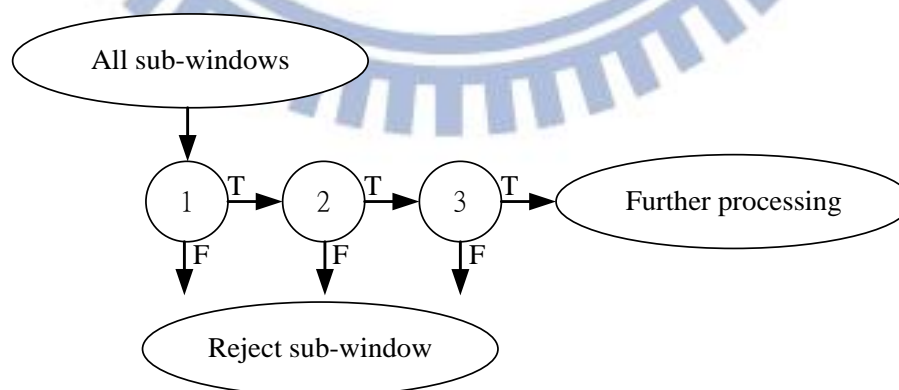


圖 2-4、串接分類器架構圖。

2.1.2. 人臉影像正規化

本研究藉由雙線性內插之方法[35]，將人臉區域影像正規化至 200×240 ，重新計算人臉影像之像素(Pixel)值，以降低人臉遠近大小等因素之影響。雙線性內插法利用鄰近已知 Pixel 值 $f(x, y)$ ，計算新影像之 Pixel 值 $f(x', y')$ ，其計算方式如(2-1)所示。雙線性內插方法示意如圖 2-5 所示，白點為原解析度 Pixel 值，黑點為預估測新解析度 Pixel 值。

$$f(x', y') = (1 - \lambda) \times [(1 - u) \times f(x, y) + u \times f(x, y + 1)] + \lambda \times [(1 - u) \times f(x + 1, y) + u \times f(x + 1, y + 1)],$$

$$\lambda = \frac{x' - x}{(x + 1) - x}, u = \frac{y' - y}{(y + 1) - y}, \quad (2 - 1)$$

2.2. 主動外觀模型

Cootes 等人所提出之 AAM[32 - 34]是由主動形狀模型(Active Shape model, ASM)[36]發展而來。AAM 由形狀模型與紋理模型所構成，藉由此二模型對輸入之人臉影像進行模擬，以建立人臉模型。模擬之人臉，能夠克服人臉不同角度之偏轉與形變，並經過適當之影像修正以抵抗光線變化，藉由考量人臉形狀與人臉紋理，完整模擬之人臉擷取特徵，對於人臉特徵擷取之穩定度則有所提升。建立

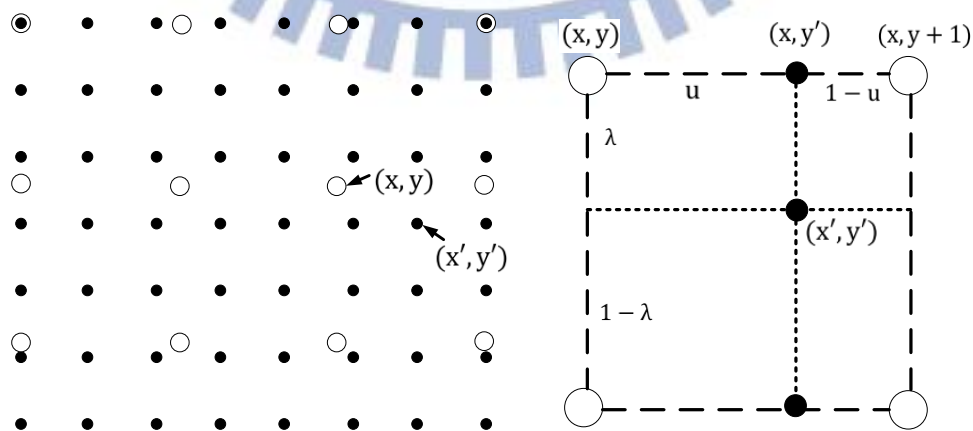


圖 2-5、雙線性內插示意圖。白點代表已知 Pixel，黑點代表預估測之 Pixel。

之人臉模型能夠擷取人臉幾何特徵或紋理特徵，做為情緒辨識之特徵，本研究主要以 AAM 擷取之特徵點計算幾何特徵進行情緒辨識。

2.3. 人臉形狀模型

AAM 之形狀模型包含平均人臉形狀，與平均人臉形狀可容許之變化向量，藉由人臉形狀模型，以模擬不同的人臉形狀。

2.3.1. 標註特徵點

建立人臉形狀模型時，必須先以手動方式標註特徵點，特徵點所在位置為曲線之高曲率(High curvature)，或曲線之 T 型交點(T Junction)所在處。另外，定義上述二者之中間點(Intermediate point)對邊界做描述，如圖 2-6 所示。

本研究參考[34], [36- 37]特徵點之定義，定義出 70 個人臉特徵點，分別為二邊眉毛各 5 個點、鼻子 9 個點、二邊眼睛輪廓各 8 個點，人臉輪廓 19 個點。另外，本研究在嘴巴外緣設置 12 個點、內緣 6 個點，以模擬嘴巴之開闔。7 種基本情緒之特徵點標註範例，如圖 2-7 所示。

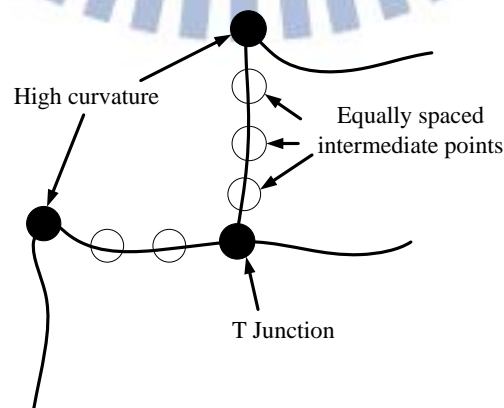


圖 2-6、特徵點選取原則示意圖。

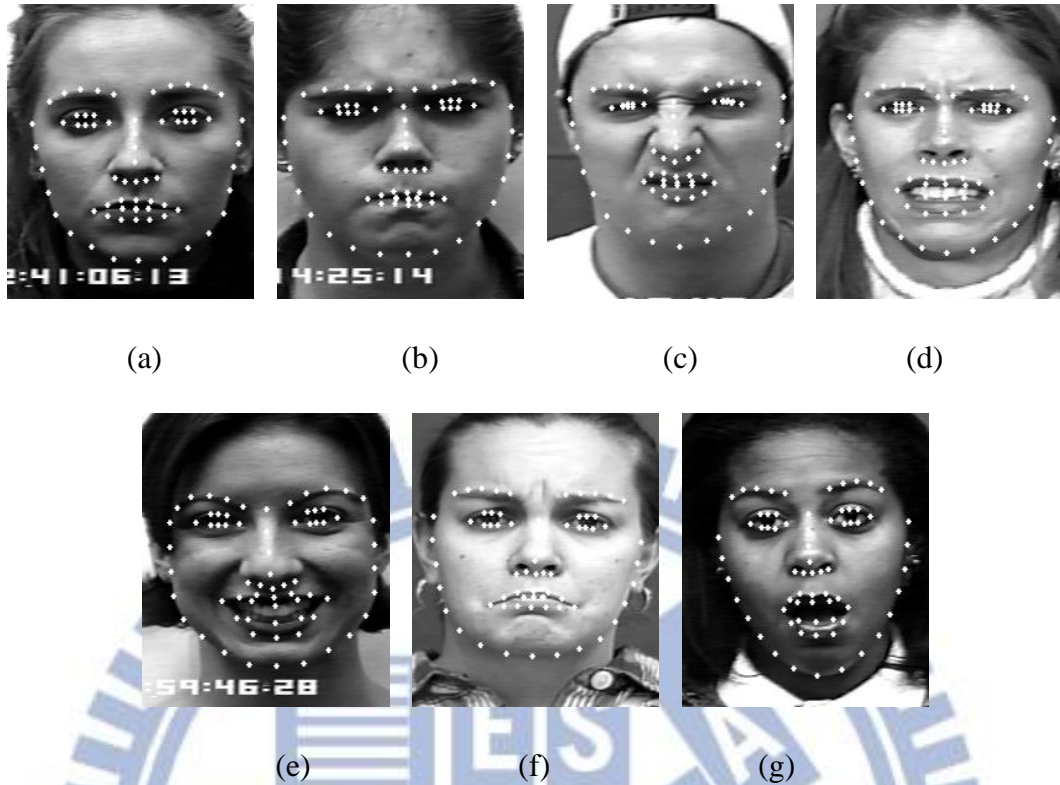


圖 2-7、7 種基本表情特徵點標註範例，(a)中性，(b)生氣，(c)厭惡，(d)害怕，(e)高興，(f)傷心，(g)驚訝。

2.3.2. 平均人臉形狀

本研究藉由輸入之人臉特徵點座標，計算平均人臉形狀，做為輸入人臉形狀之基底。根據本論文 2.3.1 節標註之 70 個特徵點，可由(2-2)表示之，

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

其中 s 代表人臉形狀； (x_v, y_v) 代表特徵點之座標。由於每個輸入人臉形狀之大小、位置與旋轉角度皆不一致，因此必須對齊所有人臉形狀，以取得正確之平均人臉形狀。本研究使用 Procrustes Analysis 進行形狀對齊[34], [38]，Procrustes Analysis 主要利用對所有輸入形狀做縮放、位移，以及旋轉之方式，以實現形狀對齊之目的，如(2-3)所示，

$$T_t(\mathbf{x}) = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \mathbf{x} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}, \quad (2-3)$$

其中 $T_t(\mathbf{x})$ 代表經由縮放、位移，以及旋轉之輸出； a, b, c, d 代表縮放與旋轉之參數； t_x, t_y 代表 x, y 方向之位移。

假設平均人臉形狀為 (x'_i, y'_i) ，與任一個輸入人臉形狀 (x_i, y_i) ，其中 $i = 1, 2, \dots, 70$ ，利用二者之線性轉換，即可解得 a, b, c, d, t_x, t_y 之矩陣參數，如(2-4)所示，

$$\begin{bmatrix} S_{xx} & S_{xy} & S_x \\ S_{xy} & S_{yy} & S_y \\ S_x & S_y & n \end{bmatrix} \begin{bmatrix} a & c \\ b & d \\ t_x & t_y \end{bmatrix} = \begin{bmatrix} S_{xx'} & S_{xy'} \\ S_{yx'} & S_{yy'} \\ S_{x'} & S_{y'} \end{bmatrix},$$

$$S_{xx} = \sum x_i^2, S_{yy} = \sum y_i^2, S_x = \sum x_i, S_y = \sum y_i,$$

$$S_{xx'} = \sum x_i x'_i, S_{yy'} = \sum y_i y'_i, S_{x'} = \sum x'_i, S_{y'} = \sum y'_i,$$

$$n = 70, \quad (2-4)$$

將解得參數帶入(2-3)以取得轉正之輸入形狀。經由上述方法消除縮放、旋轉與位移之情況，計算得到新的平均人臉形狀，重複同樣的步驟直至收斂，以得到最終之平均人臉形狀。本研究藉由計算 Procrustes distance，如(2-5)所示，判斷現在平均人臉形狀與前一次平均人臉形狀之差異，當差異足夠小時即收斂。

$$P_d^2 = \sum_{j=1}^v [(x_{j1} - x_{j0})^2 + (y_{j1} - y_{j0})^2], \quad (2-5)$$

其中 (x_{j1}, y_{j1}) 代表現在之平均人臉形狀座標； (x_{j0}, y_{j0}) 代表前一次之平均人臉形狀座標。

圖 2-8 為 120 個人臉訓練平均人臉形狀範例。圖 2-7(a) 代表原本的人臉形狀重疊情況，其中每一個點皆代表 120 個人臉形狀之 x, y 座標。利用 Procrustes Analysis 消除縮放、位移、旋轉之人臉形狀重疊情況，如圖 2-7(b) 所示。對齊後所求得之平均人臉形狀如圖 2-9 所示。

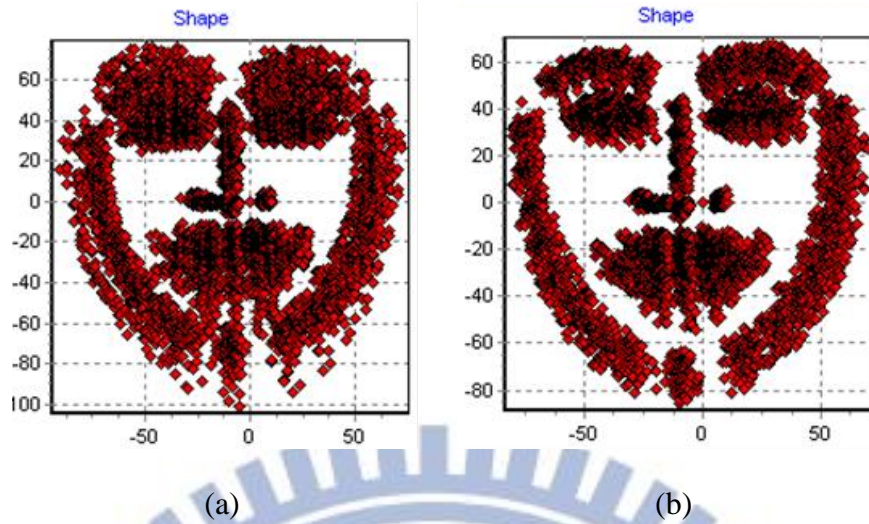


圖 2-8、訓練平均人臉形狀範例，(a)人臉形狀對齊前重疊情況，(a)人臉形狀對齊後重疊情況。

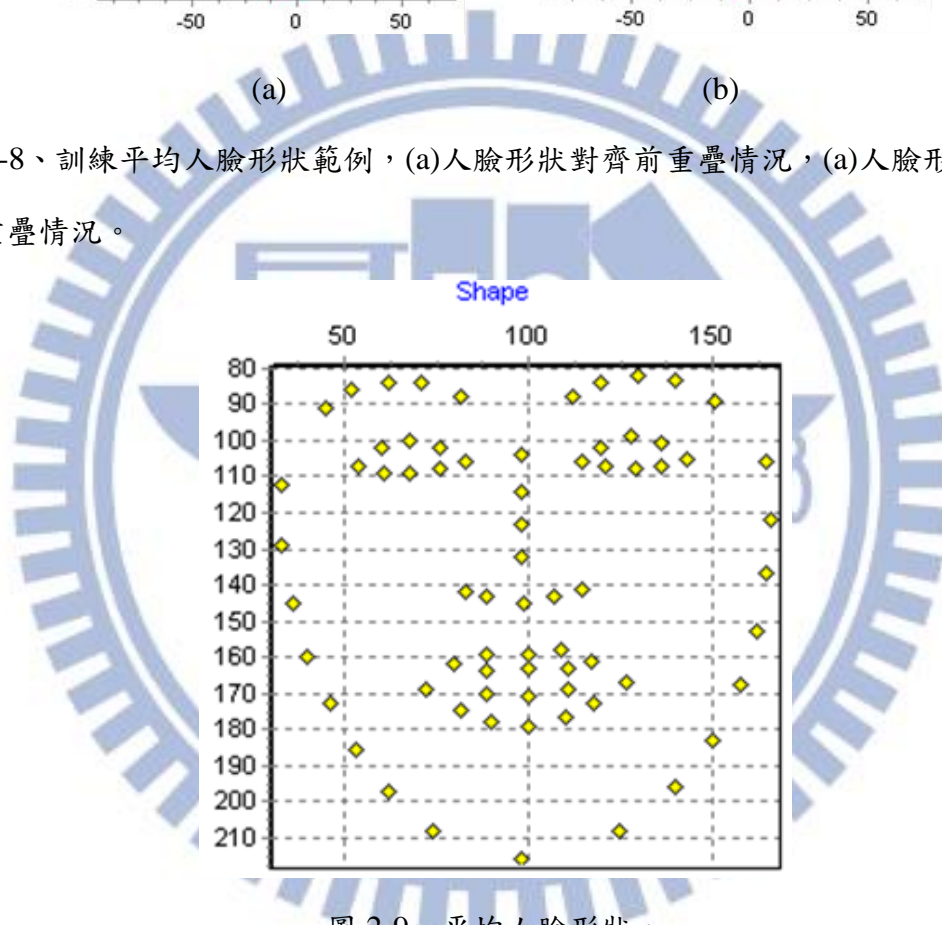


圖 2-9、平均人臉形狀。

2.3.3. 模擬臉部形狀變化

對齊後之人臉形狀，可由(2-2)表示之，即每個人臉形狀皆視為分佈於 $2v$ 維向量空間之一點。為了建立人臉形狀變化模型，本研究使用主成份分析(Principal component analysis, PCA)化簡 $2v$ 維度資料，以取得所有人臉形狀之間的變化情

況[34]。在 $2v$ 維空間中，平均人臉形狀 s_0 ，即是將所有人臉形狀相加取平均，如(2-6)所示，

$$s_0 = \frac{1}{n} \sum_{i=1}^n x_i, \quad (2-6)$$

其中 x_i 代表訓練人臉形狀。共變異矩陣 S 則為所有人臉形狀減去平均人臉形狀，如(2-7)所示，

$$S = \frac{1}{n} \sum_{i=1}^n (x_i - s_0)(x_i - s_0)^T, \quad (2-7)$$

由共變異矩陣之計算，即可解得特徵向量(Eigenvector)與特徵值(Eigenvalue)，以建立形狀變化模型，如(2-8)所示，

$$s = s_0 + \sum_{i=1}^t p_i s_i, \quad (2-8)$$

其中 s_i 代表前 t 大 Eigenvalue 對應之 Eigenvector； p_i 代表權重值所組成之形變參數。

2.4. 人臉紋理模型

AAM 之紋理模型同樣包含平均人臉紋理，與平均人臉紋理可容許之變化向量，藉由人臉形狀模型，以模擬不同的人臉紋理。

2.4.1. 人臉紋理

人臉紋理特徵可以為原始影像之 RGB Pixel、灰階 Pixel，或經由影像處理後之 Pixel 所組成，紋理模型所組成之向量如(2-10)所示，

$$g = (g_1, g_2, \dots, g_r), \quad (2-10)$$

其中 g_r 代表每個 Pixel 之紋理數值； r 代表紋理模型總共之 Pixel 數，本研究

使用平均人臉形狀範圍內所包含之 Pixel 數決定 r 值。

2.4.2. 分段仿射變形

由於訓練人臉影像樣本，每個人臉形狀、位置與大小都有所差異，因此必須對影像樣本進行分段仿射變形(Piecewise affine warping)。本研究參考 Delaunay triangulation[39]利用多個三角形將人臉輸入之特徵點連接起來，其所形成的每個三角形外接圓內不會有其他特徵點存在，以此方法對人臉形狀做切割，並修改嘴巴部份以容許模擬出嘴巴開闔形狀，以及將人臉形狀左右三角形對稱[36]，如圖 2-10 所示。切割後的各個三角形將人臉紋理由原始人臉形狀轉換至平均人臉形狀中。

利用 Delaunay triangulation 將人臉形狀切割完成後，則需將人臉紋理轉正，本研究以人臉形狀切割取得之每個三角形與平均人臉形狀對應之三角形進行轉換，亦即將每個人臉形狀三角形內之 Pixel 值轉換至平均人臉形狀對應之三角形，如(2-11)所示，

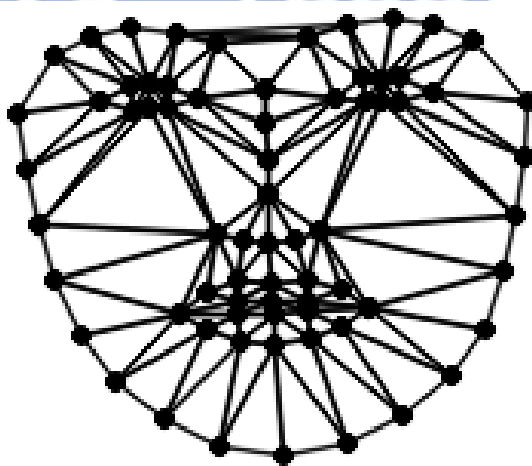


圖 2-10、Delaunay triangulation 方法對人臉切割示意圖。

$$\mathbf{x} = \mathbf{x}_1 + \alpha(\mathbf{x}_2 - \mathbf{x}_1) + \beta(\mathbf{x}_3 - \mathbf{x}_1), 0 \leq \alpha, \beta \leq 1,$$

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

$$\beta = \frac{(y - y_i)(x_j - x_i) - (x - x_i)(y_j - y_i)}{(x_j - x_i)(y_k - y_i) - (y_j - y_i)(x_k - x_i)}, \quad (2-11)$$

其中 $\mathbf{x}_1(x_i, y_i)^T$, $\mathbf{x}_2(x_j, y_j)^T$, $\mathbf{x}_3(x_k, y_k)^T$ 代表任意一個三角形之三個頂點座標； $\mathbf{x}(x, y)^T$ 代表三角形中的任意一點座標。藉由 α 與 β 之座標轉換，即可以計算紋理 Pixel 值在平均人臉形狀中的位置 $\mathbf{x}'(x', y')^T$ ，如(2-12)所示，

$$\mathbf{x}' = \mathbf{x}'_1 + \alpha(\mathbf{x}'_2 - \mathbf{x}'_1) + \beta(\mathbf{x}'_3 - \mathbf{x}'_1), \quad (2-12)$$

三角形之形變座標轉換示意圖，如圖 2-11 所示。

2.4.3. 模擬臉部紋理變化

人臉紋理轉正後，必須對所有訓練人臉影像求得其紋理之變化，以建立臉部紋理模型。本研究同樣藉由 PCA 對訓練之人臉紋理資料做化簡，以求得人臉紋理之變化。

平均紋理 A_0 ，即是將所有轉正之人臉影像相加取平均計算而得，如(2-13)所示，

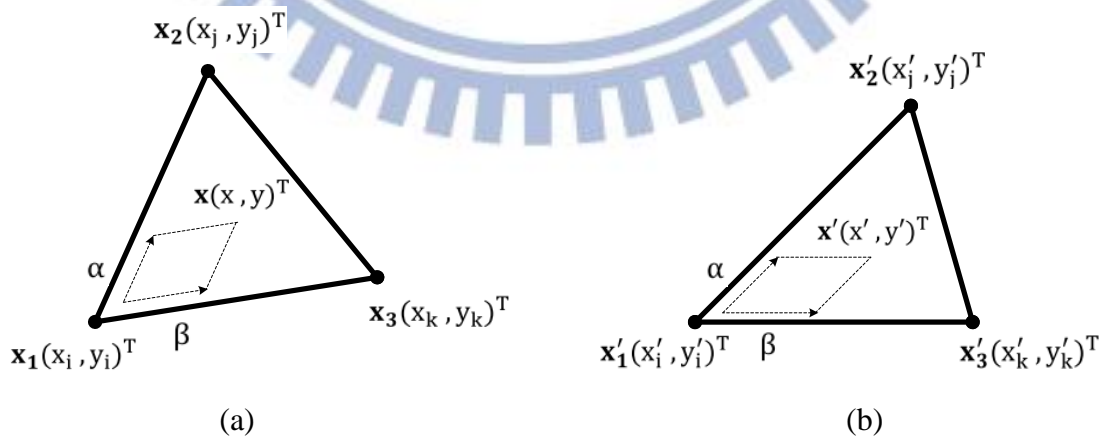


圖 2-11、三角形之形變座標轉換示意圖，(a)三角形轉正前，(b)三角形轉正後。

$$A_0 = \frac{1}{n} \sum_{i=1}^n g_i, \quad (2-13)$$

其中 g_i 代表訓練紋理。共變異矩陣 G ，則是將所有訓練紋理減去平均紋理 A_0 求得，如(2-14)所示，

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

由共變異矩陣之計算，即可解得 Eigenvector 與 Eigenvalue，以建立紋理變化模型，如(2-15)所示，

$$A(x) = A_0(x) + \sum_{i=1}^t \lambda_i A_i(x) \quad \forall x \in s_0, \quad (2-15)$$

其中 $A_i(x)$ 代表前 t 大 Eigenvalue 對應之 Eigenvector； λ_i 代表權重值所組成之形變參數。

2.5. 影像校正演算法

由上述之 AAM 演算法已取得其形狀模型與紋理模型，本節將介紹藉由輸入人臉影像擷取特徵之方法。利用 AAM 之形狀模型與紋理模型即可建立人臉模型，對輸入之人臉影像做模擬，並得到對準之人臉特徵位置。

AAM 校正演算法，可分為 Independent AAM[37]與 Combine AAM[2.3]二種方法。Independent AAM 將形狀模型與紋理模型分開，分別計算形狀參數與紋理參數，即可取得模擬結果；而 Combine AAM 則是將形狀模型與紋理模型整合為一個模型，透過整合參數之計算，取得模擬結果。然而，Independent AAM 主要以形狀參數之計算為主，因此計算量上 Combine AAM 所需之時間較多。本研究選擇 Independent AAM 進行影像校正。

2.5.1. Inverse Compositional 演算法

Independent AAM 主要是以 Lucas-Kanade 人臉校正演算法[37 - 38]進行影像校正。Lucas-Kanade 校正演算法以影像梯度為基礎做影像變化之估測，其常用於影像形變校正、物體追蹤等。然而，Lucas-Kanade 校正演算法在迭代過程中，皆與形變參數 p 有關，必須重複計算 Warping Jacobian, Steepest descent image 以及 Hessian matrix，運算量因而大幅提升。Inverse Compositional 演算法[37 - 38]改善上述之問題，其與 Lucas-Kanade 主要差異在於，

- Lucas-Kanade 校正演算法，在每次迭代過程皆需重複計算輸入影像之梯度值，以及 Warping Jacobian, Steepest descent image 與 Hessian matrix。
- Inverse Compositional 校正演算法利用樣板影像 A_0 計算影像之梯度值，由於樣板影像已知，因此 Warping Jacobian, Steepest descent image 與 Hessian matrix 亦已知。

Inverse Compositional 演算法迭代時即可省去梯度值，以及 Warping Jacobian, Steepest descent image 與 Hessian matrix 之計算，速度因而提升，且能夠維持住其影像校正之效能。本研究採用 Inverse Compositional 演算法，對輸入人臉影像做校正。如圖 2-12 所示，Inverse Compositional 演算法分為前處理之 4 個步驟(I ~ IV)與迭代之 5 個步驟(1 ~ 5)[34], [36]，如下所述，

前處理：

- I. 根據樣板影像 $A_0(x)$ 計算梯度影像 ∇A_0 。
- II. 求出 $W(x; 0)$ 即樣板人臉模型的 Warping Jacobian $\frac{\partial W}{\partial p}$ 。
- III. 計算出 Steepest descent images $\nabla A_0 \frac{\partial W}{\partial p}$ 。

IV. 由 Steepest descent images 之結果，進一步解得 Hessian matrix。

迭代過程：

1. 根據 $W(x; p)$ 形變影像 I ，計算出形變影像 $I(W(x; p))$ 。
2. 計算誤差影像 $I(W(x; p)) - A_0(x)$ 。
3. 計算 $\sum_x \left[\nabla A_0 \frac{\partial W}{\partial p} \right]^T [I(W(x; p)) - A_0(x)]$ 。
4. 計算形變參數變化量 Δp 。
5. 更新參數 $W(x; p) \leftarrow W(x; p) \circ W(x; \Delta p)^{-1}$ 。

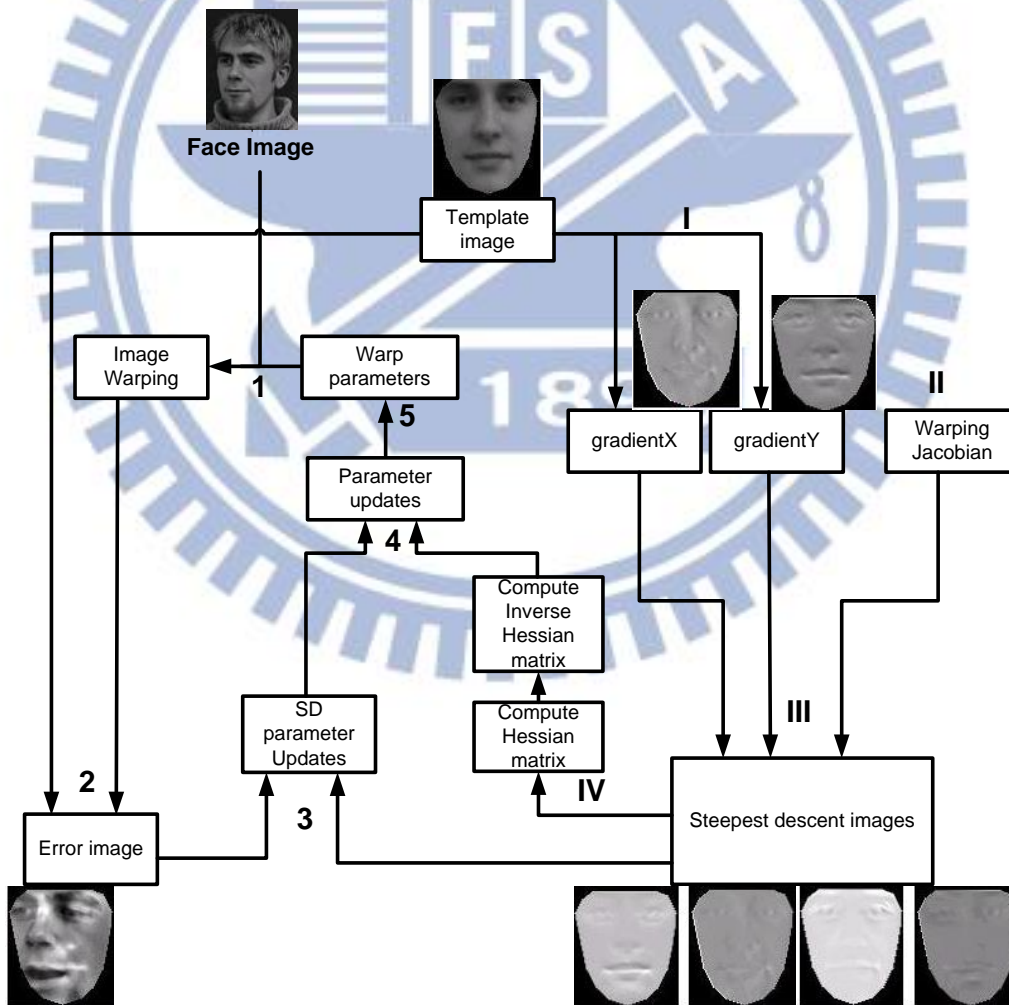


圖 2-12、Inverse Compositional 演算法架構[36]。

2.5.2. 整體形狀正規化轉換

由於一般而言 AAM 對於人臉偏轉之角度容忍度十分有限，容易導致偏轉角度之人臉迭代錯誤，且在訓練人臉形狀變化模型前，經由 Procrustes Analysis 對人臉做對齊，而移除某些縮放、位移與旋轉的資訊。因此，本研究藉由整體形狀正規化轉換[34], [36]，以取得實際人臉擁有之縮放、位移與旋轉的資訊。

本研究定義 $N(x; q)$ 為訓練資料的整體形狀正規化轉換，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(2-16)所示，

$$N(x; q) = \begin{bmatrix} (1+a) & -b \\ b & (1+a) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}, \quad (2-16)$$

為了與之前建立的人臉形狀模型(2-8)搭配運算，必須建立一組線性組合之係數[34]，用以模擬形狀模型之變化。假若平均形狀 $s_0 = (x_1^0, y_1^0, \dots, x_v^0, y_v^0)^T$ ，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 2-13 所示， $N(x; q)$ 可以表示為(2-17)之形式，

$$N(x; q) = s_0 + \sum_{i=1}^4 q_i s_i^*, \quad (2-17)$$

在形狀參數更新方面，此處必須求解 $N \circ W$ 之形狀參數。參考[36]於 Inverse Compositional 演算法中得知 $W(x; \Delta p)^{-1} = W(x; -\Delta p)$ ，因此參數 $(\Delta q, \Delta p)$ 可表示為(2-18)，

$$N \circ W(x; \Delta q, \Delta p)^{-1} = N \circ W(x; -\Delta q, -\Delta p), \quad (2-18)$$

其合成形式如(2-19)所示，

$$\begin{aligned} & (N \circ W)(x; q, p) \circ (N \circ W)(x; \Delta q, \Delta p)^{-1} \\ & \approx (N \circ W)(x; q, p) \circ (N \circ W)(x; -\Delta q, -\Delta p), \end{aligned} \quad (2-19)$$

當輸入人臉影像時，本研究先將平均人臉形狀 s_0 加入影像中，藉由更新形變參數之方式，以平均人臉形狀做參數之調整，故(2-18)可由(2-8)之線性組合

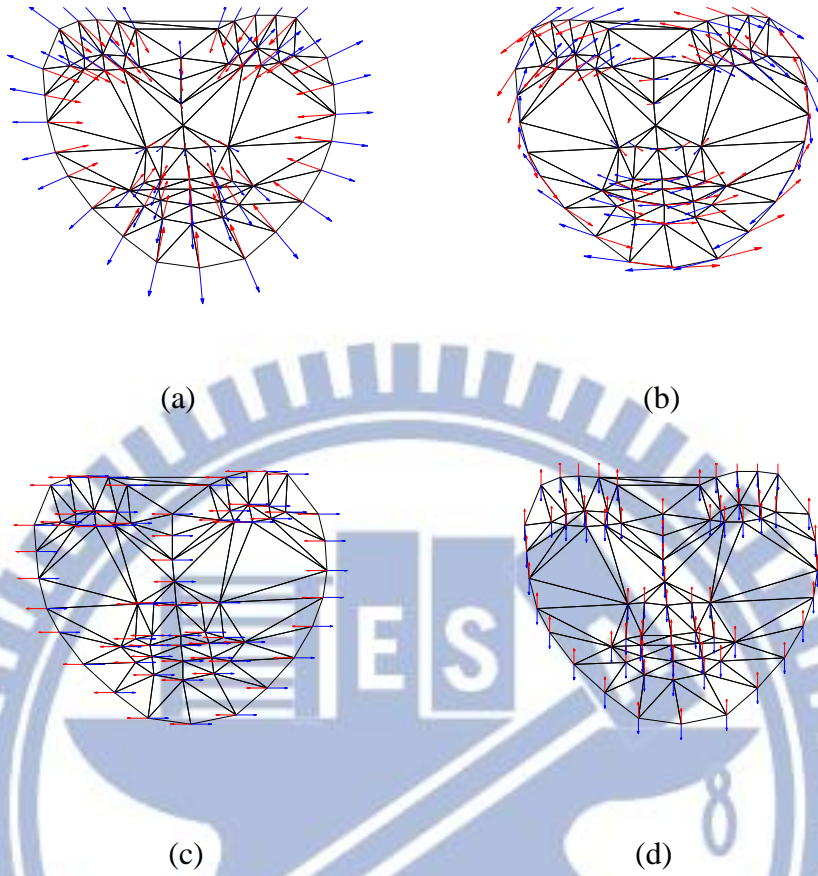


圖 2-13、整體形狀正規化轉換形狀變化模型[36]，(a) $s_1^* = (x_1^0, y_1^0, \dots, x_v^0, y_v^0)^T$ ，
 (b) $s_2^* = (-y_1^0, x_1^0, \dots, -y_v^0, x_v^0)^T$ ，(c) $s_3^* = (1, 0, \dots, 1, 0)^T$ ，(d) $s_4^* = (0, 1, \dots, 0, 1)^T$ 。

方式表示，如(2-20)所示，

$$\Delta s_0 = - \sum_{i=1}^n \Delta p_i s_i - \sum_{j=1}^4 \Delta q_j s_j^*, \quad (2-20)$$

建立之形狀變化模型，如(2-21)所示，

$$N \circ W(s_0; q, p) = N \left(s_0 + \sum_{i=1}^n p_i s_i; q \right) = s_0 + \sum_{i=1}^n p_i s_i + \sum_{j=1}^4 q_j s_j^*, \quad (2-21)$$

利用正交之關係，即可計算求得 $(N \circ W)(x; q, p) \circ (N \circ W)(x; \Delta q, \Delta p)^{-1}$ 之形狀參數，如(2-22) ~ (2-23)所示，

$$q_i = s_i^* \cdot (s^\dagger - s_0), \quad (2-22)$$

$$p_i = s_i \cdot (N(s^\dagger; q)^{-1} - s_0), \quad (2-23)$$

其中 s^\dagger 定義為 $s^\dagger = N \circ W(s_0; q, p)$ 。

在紋理方面，藉由(2-24)計算 (q_1, q_2, q_3, q_4) 之 Steepest descent image，用於 Inverse Compositional 演算法中，

$$SD_j(x) = \nabla A_0 \frac{\partial N}{\partial q_j} - \sum_{i=1}^m \left[\sum_{x \in S_0} A_i(x) \cdot \nabla A_0 \frac{\partial N}{\partial q_j} \right] A_i(x). \quad (2-24)$$

另外，藉由(2-25)計算 p 參數的 Steepest descent image，

$$SD_{j+4}(x) = \nabla A_0 \frac{\partial W}{\partial p_j} - \sum_{i=1}^m \left[\sum_{x \in S_0} A_i(x) \cdot \nabla A_0 \frac{\partial W}{\partial p_j} \right] A_i(x), \quad (2-25)$$

其中 $j=1,2,\dots,n$

而用於 Inverse Compositional 演算法之其計算如(2-26)所示，

$$H = \sum_x SD(x)^T SD(x), \quad (2-26)$$

紋理參數如(2-27)所示，

$$\lambda_i = \sum_{x \in S_0} A_i(x) \cdot [I(N(W(x; p); q)) - A_0(x)], \quad (2-27)$$

2.5.3. 梯度影像修正

在 2.5.2 節介紹之整體形狀正規化轉換中，必須由樣板影像 $A_0(x)$ 計算梯度影像 ∇A_0 ，其包含垂直方向之梯度(Gradient X)與水平方向之梯度(Gradient Y)。然而，在計算梯度影像時，會以人臉樣板整張影像做計算，如此會計算到非平均人臉影像部份，如圖 2-14 邊緣黑色處，而造成人臉形狀邊緣之梯度值產生錯誤。因此，本研究將邊緣錯誤之梯度值，以平均人臉形狀往內一個 Pixel 之梯度值取代。



圖 2-14、樣板人臉影像。

2.5.4. 直方圖等化

直方圖等化之影像修正，其優點在於修正過後之影像更能夠凸顯人臉之特徵。計算時，先產生輸入影像之直方圖，再由直方圖除以人臉形狀範圍中所有 Pixel 之數量，以求得該平均人臉影像之機率密度函數(Probability Density Function, PDF)，進一步算出累積分佈函數(Cumulative Distribution Function, CDF)。為了對應影像灰階值，而將累積分佈函數由 0-1 放大到 0-255 的範圍，並將影像原始灰階值以對應到新的灰階值取代，如此即可得到直方圖等化後之影像。

2.5.5. 影像校正整體方法

結合上述之所有方法並總結其流程[36]，如下所述，

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 $W(x; 0)$ 即樣板人臉模型的 Jacobian $\frac{\partial W}{\partial p}$ 及 $\frac{\partial N}{\partial q}$ 。

IV. 由(2-24)與(2-25)計算改變之 steepest descent images $SD_j(x)$ 和 $SD_{j+4}(x)$ 。

V. 由(2-26)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)與(2-23)計算出形狀變化參數 q_i 與 p_i ；由(2-27)計算出紋理變化參數 λ_i 。

藉由計算求得之形狀變化參數以及紋理變化參數，即可模擬出人臉形狀模型與紋理模型，以建立 AAM 人臉模型，擷取適當之人臉特徵點，進而計算本研究選擇之情緒辨識幾何特徵。事實上，透過建立 AAM 人臉模型，亦可擷取紋理特徵，然而，為了減少特徵之維度，以降低運算複雜度，根據相關研究之分析，擷取適當之幾何特徵，已足以做為情緒辨識之依據[18], [26]。

第三章、基於機率之連續影像情緒辨識

情緒辨識之結果，影響著人機互動之效能，相較於基本情緒類別辨識，對基本情緒進行程度、比例等情緒資訊辨識，能夠提供更精準之人臉情緒狀態，做為人機互動之依據。本研究利用人臉距離特徵與位移特徵做為幾何特徵，經由相關向量機(Relevance vector machine, RVM)辨識情緒之可能性(Likelihood)為主要目的，提出以連續影像對情緒做時序分析之方法，對序列影像間情緒 Likelihood 之關係做考量，提升人臉情緒辨識之正確性。另外，為了達成人機互動之目的，本研究提出將情緒 Likelihood 轉換為 A-V 之方法，讓使用者之情緒狀態能夠與同樣使用 A-V 設計機器情緒狀態之系統做對應，以設計一套智慧型人機互動系統。本章節將分別對人臉特徵擷取、RVM 分類器之 Likelihood 時序辨識設計，與 RVM 回歸模型轉換 Likelihood 至 A-V 之方法進行說明。

3.1. 人臉特徵擷取

根據 Ekman 定義的 Facial Action Coding System (FACS) [40]，其提供能夠表示人臉微妙肌肉變化之 44 個 Action units (AUs)(例如：嘴唇緊縮、眉毛上揚等資訊)，本研究將此類資訊轉換為幾何特徵做描述。藉由本論文第二章介紹之 AAM 建立人臉模型，從中擷取 18 個人臉特徵點，如圖 3-1 所示，而這 18 個特徵點結合 AUs 之人臉特徵分析，有系統的歸納出足以代表人臉表情之 16 個幾何特徵，其中本研究之幾何特徵包含距離特徵與位移特徵。

[18], [25], [41]之研究分別使用不同資料庫對人臉情緒與 AUs 之組合做分析。然而，每位使用者呈現基本情緒之人臉表情不盡相同，本研究以 Cohn-Kanade 資料庫[41]分析之情緒 AUs 組合為基礎，並與使用不同資料庫[18], [25]之情緒分



圖 3-1、18 個人臉特徵點位置。

析結果做整合，以歸納取得較為一般化之情緒 AUs 組合。本研究歸納整合 6 種基本人臉情緒之 AUs 組合如表 3-1 與圖 3-2 所示，而基本情緒相關 AUs 與其對應之特徵點位移關係則如表 3-2 所示，以 AU1 為例，其描述內側眉頭上揚之臉部肌肉變化特徵，可由特徵點 P7 與 P15 之距離 ($\|P7P15\|$) 變化，以及特徵點 P9 與 P16 之距離 ($\|P9P16\|$) 變化觀查得知。利用表 3-1、表 3-2 與圖 3-2 之人臉表情 AUs 組合與定義，本研究由 18 個特徵點分析歸納出 16 個幾何特徵取代 AUs，16 個幾何特徵中的距離特徵分別為 $\|P1P3\|$ ， $\|P2P4\|$ ， $\|P1P2\|$ ， $\|P2P3\|$ ， $\|P3P4\|$ ， $\|P1P4\|$ ， $\|PcP3\|$ ， $\|P4P5\|$ ， $\|P6P8\|$ ， $\|P7P15\|$ ， $\|P8P14\|$ ， $\|P15P16\|$ ， $\|P2P11\|$ ， $\|P10P12\|$ ， $\|P9P16\|$ ， $\|P12P17\|$ ，其中 Pc 為內眼角 P7 與 P9 之中點。 $\|P7P9\|$ 則做為正規化因子，對上述之 16 個距離特徵進行正規化，如(3-1)所示，以降低人臉大小等因素之影響。

$$\text{Normalized distance features} = \frac{\text{Distance features}}{\|P7P9\|}, \quad (3-1)$$

16 個幾何特徵中的位移特徵，則為上述 16 個距離特徵相對於特定情緒之變化量。

表 3-1、6 種基本人臉表情之 AUs 組合。

Emotional category	Visual Cues				
Anger	AU4	AU7	AU23	AU24	
Disgust	AU9	AU10			
Fear	AU1	AU2	AU4	AU7	AU20
Happy	AU12	AU25			
Sadness	AU1	AU4	AU15		
Surprise	AU1	AU2	AU5	AU27	








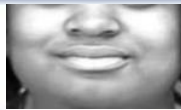






AU1	AU2	AU4	AU5	AU7
				
Inner brow raiser	Outer brow raiser	Brow lower	Upper lid raiser	Lid tightener
AU9	AU10	AU12	AU15	AU20
				
Nose wrinkle	Upper lip raiser	Lip corner puller	Lip corner depressor	Lip stretcher
AU23	AU24	AU25	AU27	
				
Lip tightener	Lip pressor	Lips part	Mouth stretch	

圖 3-2、6 種基本人臉表情相關之 AUs。

表 3-2、6 種人臉表情之 AUs 組合定義。

AUs	Facial Visual Cues	Description
AU1	$\ \overline{P7P15}\ (\ \overline{P9P16}\)$ increased	Inner brow raiser
AU2	$\ \overline{P8P14}\ (\ \overline{P12P17}\)$ increased	Outer brow raiser
AU4	$\ \overline{P8P14}\ (\ \overline{P12P17}\)$, $\ \overline{P7P15}\ (\ \overline{P9P16}\)$ and $\ \overline{P15P16}\ $ decreased, or wrinkle in R1 increased	Brow lower
AU5	$\ \overline{P6P8}\ (\ \overline{P10P12}\)$ increased	Upper eyelid raiser
AU7	$\ \overline{P6P8}\ (\ \overline{P10P12}\)$ decreased	Eyelid tighter
AU9	$\ \overline{P7P15}\ (\ \overline{P9P16}\)$ decreased, wrinkle in R2 increased	Nose wrinkle
AU10	$\ \overline{P4P5}\ (\ \overline{P2P11}\)$ decreased	Upper lip raiser
AU12	$\ \overline{P4P5}\ (\ \overline{P2P11}\)$ decreased, $\ \overline{P2P4}\ $ increased, $\ \overline{P6P8}\ (\ \overline{P10P12}\)$ decreased	Lip corner puller
AU15	$\ \overline{P4P5}\ (\ \overline{P2P11}\)$ increased, $\ \overline{P2P4}\ $ increased	Lip corner depressor
AU20	$\ \overline{P4P5}\ (\ \overline{P2P11}\)$ non-change, $\ \overline{P2P4}\ $ increased	Lip stretcher
AU23	$\ \overline{P1P3}\ $, $\ \overline{P2P4}\ $ decreased	Lip tightener
AU24	$\ \overline{P1P3}\ $ decreased, $\ \overline{P2P4}\ $ non-change	Lip presser
AU25	$\ \overline{P1P3}\ $ increased	Lips part
AU27	$\ \overline{P1P3}\ $ increased, $\ \overline{P2P4}\ $ decreased	Mouth Stretch

3.2. 特定情緒辨識

距離特徵可以直接以單張影像樣本，計算特徵點距離取得。至於位移特徵擷取方面，本研究則提出以分群演算法，對特定情緒進行辨識，並以特定情緒為依據做特徵點位移之計算，此特定情緒必須為使用者最強程度之基本情緒(例如：生氣、厭惡等情緒)，以確保使用位移特徵辨識之效能。由於每個人所呈現之特定情緒之表情有所不同，若未做身分辨識，在辨識特定情緒上有其難度。本研究提出對特定情緒本身做分群之方法，主要目的在於將表情特徵看起來相似之特定情緒分為一群，例如：若指定中性情緒做為特定情緒，藉由分群之方法將看起來像生氣的中性情緒分為一群，看起來像厭惡的中性情緒又分為另一群等。本研究利用 Threshold order-dependent (TOD)分群演算法[42]，對特定情緒進行分群，在本研究中選擇中性情緒做為特定情緒，以此情緒為根據計算位移特徵。

TOD 分群演算法，假設輸入特徵向量 $\{\mathbf{x}_n\}_{n=1}^N$ ，類群代表中心 $\{\mathbf{y}_n\}_{n=1}^k$ ，特徵向量與類群代表中心距離臨界值 t 。TOD 演算法初始選擇一組特徵向量做為類群代表中心，即 $\mathbf{y}_1 = \mathbf{x}_1$ 。其他特徵向量則與目前所有的代表中心做距離估測，當距離大於預測臨界值 t ，則自行形成另一群，產生新的代表中心，整體 TOD 演算法分群步驟如下。

Step 1. 初始代表中心與類群數，並設定臨界值 t 。

$$\mathbf{y}_1 = \mathbf{x}_1; k = 1 \text{ 其中 } k \text{ 為目前的類群數, } \mathbf{x}_1 \in C_1; t = 0.1。$$

Step 2. 計算訓練樣本與代表中心之均方根誤差(Root mean square error, RMSE)，判斷特徵向量所屬類群。

本研究利用 RMSE 如(3-2)，做為判斷特徵向量所屬類群依據。

$$\text{RMSE}_k = \sqrt{\frac{\sum_{i=1}^N (x_i - y_{ki})^2}{N}}, \quad (3-2)$$

其中 N 代表 N 個特徵數； x_i 為觀察到的特徵值； y_{ki} 為第 k 個代表中心。

假若 $RMSE_k \geq t, \forall k$ ，則產生新的代表中心， $y_{k+1} = x_n, k = k + 1$ ，否則挑選 $\min(RMSE_k)$ 做為 x_n 所屬類群，即 $x_n \in C_k$ 。

當從訓練樣本中之特定情緒，利用 TOD 演算法取得 k 個代表中心後，在辨識時則可以此 k 個代表中心，計算其與測試樣本之 RMSE 如(3-2)，判斷測試樣本是否為使用者之特定情緒，本研究特定情緒 RMSE 之臨界值設為 0.1，若 RMSE 小於臨界值則代表偵測到使用者之特定情緒，在往後情緒之辨識，可以進一步結合位移特徵做辨識，以提升情緒辨識之準確性。

3.3. 類別可能性

本論文利用 Tipping[28]所提出之 RVM 進行情緒類別 Likelihood 辨識與回歸分析，並參考[43]實現 RVM 演算法，其架構如圖 3-3 所示。一般情緒辨識方法[19],[25]以單張影像樣本資訊對情緒做辨識，如此可能受到環境光線、人臉角度或人臉情緒之表情本身的不確定性等因素影響，而造成情緒辨識不穩定。本研究提出連續情緒辨識，對情緒進行時序之分析，主要目的在於提升情緒辨識之準確性，並對情緒轉換做出合理之判斷。本研究在辨識基本情緒 Likelihood 時導入追蹤之方法，每次辨識 Likelihood 時，除了以 3.1 節介紹之幾何特徵做為辨識資訊，亦會與先前辨識之 Likelihood 資訊做結合判斷，亦即前一次辨識到的 Likelihood 會進行回授，做為此次 Likelihood 辨識資訊之一。如此，當某個情緒轉換至另一情

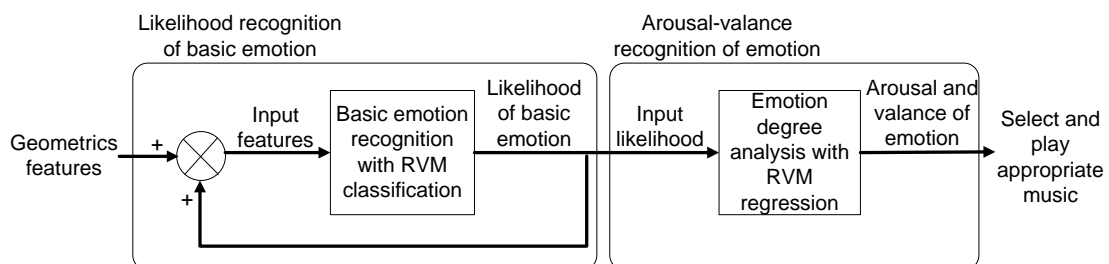


圖 3-3、情緒辨識架構圖。

緒之過程中，情緒能夠連續被辨識，確保轉換結果之合理性與正確性，適時抑制情緒 Likelihood 判斷錯誤之情況，與不合理之瞬間情緒轉變。

藉由 Likelihood 辨識與 A-V Plane 之映射，不僅對基本情緒做分類，亦對情緒間之不確定性提供更多資訊，達到情緒比例、程度辨識之目的，以設計更為人性化、更被使用者接受之人機互動系統。本節(3.3 節)將對情緒類別 Likelihood 辨識做介紹，回歸分析辨識情緒 A-V 則將在下一節(3.4 節)做詳細說明。

3.3.1. 類別可能性辨識

RVM 保有 SVM 分類準確性高與運算速度快之優點，其亦改善 SVM 之缺點，如本論文 1.2.3 節所述之 RVM 其中兩個特性：使用 Bayesian probabilistic framework 學習法則，以迭代的方式學習(1-2)辨識模型之權重，得到對應之訓練樣本做為 RV 的 RVM，能夠建立比 SVM 更為稀疏之模型，大多數之權重將被設為零，使得 RVM 之有效 kernel function 數遠少於 SVM，執行速度方面明顯的提升，且辨識率仍不亞於 SVM；另外，RVM 可以利用適當之分佈模型為基礎，取得 Likelihood 函數，決定每一個類別之 Likelihood，對於不確定類別之特徵，能夠有較精準之判斷。基於上述 RVM 之特性，本研究將分別以距離特徵與位移特徵，訓練兩種不同之 RVM 分類器，最後再將二者進行融合，如此能夠有效提升情緒辨識之效能。本節(3.3.1 節)將詳細說明 RVM 分類器應用於情緒 Likelihood 辨識之演算法，以及分類器融合之方法與策略。

a. RVM 分類器演算法

事實上，RVM 本質為兩類分類器，訓練分類器時以兩兩類別做訓練，最後再將訓練結果進行耦合(Coupling)，形成多類之 RVM 分類器。以下就 RVM 訓練兩類分類器之方法做說明與推導。

假設給定一組訓練資料為 $\{\mathbf{x}_n, t_n\}_{n=1}^N$ ，其中 \mathbf{x}_n 代表幾何特徵之特徵向量， t_n 代

表情緒類別編號， $t_n \in \{0, 1\}$ ，本研究應用分類問題常用之 Logistic sigmoid link function $\sigma(\cdot)$ ，將原線性模型如(1-2)，改寫為(3-3)之形式，使模型輸出介於 0 ~ 1 之機率值。

$$\sigma[y(\mathbf{x}_n; \mathbf{w})] = \frac{1}{1 + e^{-y(\mathbf{x}_n; \mathbf{w})}}, \quad (3-3)$$

其中 本研究選擇 Radial basis function (RBF) 做為 Kernel function，如(3-4)所示，其能夠對非線性高維度訓練樣本做有效之處理，且此 Kernel function 僅有參數 σ 需做調整，計算較為容易。

$$K(\mathbf{x}, \mathbf{x}_i) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{2\sigma^2}\right), \quad i = 1, \dots, N, \quad (3-4)$$

其中 本研究以交叉驗證之方式對訓練資料做測試，取得最佳之 σ ，以此 kernel function 將訓練資料映射至可切割之維度空間做分類。藉由 Bernoulli distribution 取得 Likelihood function $P(\mathbf{t}|\mathbf{w})$ ，對特徵向量做情緒 Likelihood 之辨識，如(3-5)所示，

$$P(\mathbf{t}|\mathbf{w}) = \prod_{n=1}^N \sigma[y(\mathbf{x}_n; \mathbf{w})]^{t_n} \{1 - \sigma[y(\mathbf{x}_n; \mathbf{w})]\}^{1-t_n}, \quad (3-5)$$

其中 $t_n \in \{0, 1\}$ 。

RVM 以(3-5)對 \mathbf{w} 做估測，然而，若直接使用 Maximum likelihood 之方法估測 \mathbf{w} ，往往導致 Over-fitting 之情況發生，亦即使大部份之權重非零。為了避免此情況發生，必須對 \mathbf{w} 加以限制。在此限制 \mathbf{w} 機率分佈應為落在 0 周圍之高斯分佈 (Zero-mean Gaussian distribution) $\mathcal{N}(\cdot)$ ，如(3-5)所示，

$$p(\mathbf{w}|\boldsymbol{\alpha}) = \prod_{i=0}^N \mathcal{N}(w_i | 0, \alpha_i^{-1}) = \prod_{i=0}^N \frac{\alpha_i}{\sqrt{2\pi}} \exp\left(-\frac{\alpha_i w_i^2}{2}\right), \quad (3-6)$$

其中 $\boldsymbol{\alpha}$ 為事前權重 (prior weight) 分佈。事後權重 (posterior weight) \mathbf{w} 可由以下過程解之。

Step 1. 將 $p(\mathbf{w}|\mathbf{t}, \boldsymbol{\alpha})$ 以 log 形式表示。

假設 $\boldsymbol{\alpha}$ 已知，事後權重(posterior weight)條件機率分佈，可由貝式定理，改寫為如(3-7)所示，

$$p(\mathbf{w}|\mathbf{t}, \boldsymbol{\alpha}) = \frac{P(\mathbf{t}|\mathbf{w})p(\mathbf{w}|\boldsymbol{\alpha})}{p(\mathbf{t}|\boldsymbol{\alpha})}, \quad (3-7)$$

由(3-7)得知 $p(\mathbf{w}|\mathbf{t}, \boldsymbol{\alpha}) \propto P(\mathbf{t}|\mathbf{w})p(\mathbf{w}|\boldsymbol{\alpha})$ ，可藉由解(3-8)取得 \mathbf{w}_{MP} ，

$$\begin{aligned} \mathbf{w}_{MP} &= \arg \max_{\mathbf{w}} p(\mathbf{w}|\mathbf{t}, \boldsymbol{\alpha}) \\ &= \arg \max_{\mathbf{w}} P(\mathbf{t}|\mathbf{w})p(\mathbf{w}|\boldsymbol{\alpha}) \\ &= \arg \max_{\mathbf{w}} \log[P(\mathbf{t}|\mathbf{w})p(\mathbf{w}|\boldsymbol{\alpha})], \end{aligned} \quad (3-8)$$

取 log 後之 $P(\mathbf{t}|\mathbf{w})p(\mathbf{w}|\boldsymbol{\alpha})$ ，如(3-9)所示，

$$\begin{aligned} &\log[P(\mathbf{t}|\mathbf{w})p(\mathbf{w}|\boldsymbol{\alpha})] \\ &= \sum_{n=1}^N [t_n \log y_n + (1 - t_n) \log(1 - y_n)] - \frac{1}{2} \mathbf{w}^T \mathbf{A} \mathbf{w}, \end{aligned} \quad (3-9)$$

其中 $y_n = \sigma[y(\mathbf{x}_n; \mathbf{w})]$ ； $\mathbf{A} = \text{diag}(\alpha_0, \alpha_1, \dots, \alpha_N)$ 。由於(3-9)為 logistic log-likelihood function，故必須以迭代近似之方式解 \mathbf{w}_{MP}

Step 2. 拉普拉斯近似(Laplacian approximation) [44 - 45]。

使用 Second-order Newton method 對(3-9)做二次微分，取得 Hessian，如(3-10)

~ (3-11)

$$\nabla_{\mathbf{w}} \log[P(\mathbf{t}|\mathbf{w})p(\mathbf{w}|\boldsymbol{\alpha})] = \boldsymbol{\Phi}^T (\mathbf{t} - \mathbf{y}) - \mathbf{A} \mathbf{w}, \quad (3-10)$$

$$\nabla_{\mathbf{w}} \nabla_{\mathbf{w}} \log[P(\mathbf{t}|\mathbf{w})p(\mathbf{w}|\boldsymbol{\alpha})] = -(\boldsymbol{\Phi}^T \mathbf{B} \boldsymbol{\Phi} + \mathbf{A}), \quad (3-11)$$

其中 $\boldsymbol{\Phi} = [\boldsymbol{\phi}(\mathbf{x}_1), \boldsymbol{\phi}(\mathbf{x}_2), \dots, \boldsymbol{\phi}(\mathbf{x}_N)]^T$,

$$\boldsymbol{\phi}(\mathbf{x}_n) = [1, K(\mathbf{x}_n, \mathbf{x}_1), K(\mathbf{x}_n, \mathbf{x}_2), \dots, K(\mathbf{x}_n, \mathbf{x}_N)]^T;$$

$$\mathbf{y} = (y_1, y_2, \dots, y_N)^T, \quad y_n = \sigma[y(\mathbf{x}_n; \mathbf{w})];$$

$$\mathbf{B} = \text{diag}(\beta_1, \beta_2, \dots, \beta_N), \quad \beta_n = y_n(1 - y_n), \quad y_n = \sigma[y(\mathbf{x}_n; \mathbf{w})].$$

(3-8)以 Laplacian approximation 推導之結果將極為近似高斯函數(Gaussian

function)，其平均值即為 \mathbf{w}_{MP} ，共變矩陣(covariance) Σ 即為(3-11)代入 \mathbf{w}_{MP} 取負號再取倒數之結果。藉由假設(3-10)等於 0，得到 \mathbf{w}_{MP} ，再進一步取得 Σ ，整理過後可以得到如(3-12) ~ (3-13)之結果。

$$\Sigma = (\Phi^T \mathbf{B} \Phi + \mathbf{A})^{-1}, \quad (3-12)$$

$$\mathbf{w}_{MP} = \Sigma \Phi^T \mathbf{B} \mathbf{t}. \quad (3-13)$$

Step 3. 更新 α 。

對 α 進行初始，並藉由[46]所提出之方法更新 α ，如(3-14)所示，

$$\alpha_i^{\text{new}} = \frac{\gamma_i}{\mu_i^2}, \quad (3-14)$$

其中 初始 $\alpha_i = (1/N)^2, i = 1, \dots, N$ ， N 代表訓練樣本數；

$$\mu = \mathbf{w}_{MP};$$

$$\gamma_i = 1 - \alpha_i N_{ii}, N_{ii} \text{ 代表 } \Sigma \text{ 第 } i \text{ 項在對角線上之元素。}$$

藉由(3-14)對 α 之更新，再配合(3-12) ~ (3-13)更新 Σ 與 \mathbf{w}_{MP} 。

經過數次之迭代，大多數之 α_i 會趨近於無限大，亦即 w_i 受到限制，使其值趨近於 0，使得 $N_{ii} \approx \alpha_i^{-1}$ 而 $\gamma_i \approx 0$ ，取得相對稀疏之解。其餘之 α_i 則會趨近於定值，使得 $\gamma_i \approx 1$ ，所對應到之 \mathbf{x}_i 即為相對向量(relevance vector, RV)。當求得 \mathbf{w}_{MP} 後，則可以對新的資料 \mathbf{x}_* 利用(3-15)計算 $P(\mathbf{t}_* | \mathbf{w})$ ，亦即本研究情緒之 Likelihood。

$$P(\mathbf{t}_* | \mathbf{w}) = \sigma[y(\mathbf{x}_*; \mathbf{w})]^{t_*} \{1 - \sigma[y(\mathbf{x}_*; \mathbf{w})]\}^{1-t_*}, \quad (3-15)$$

b. 靜態 RVM 與動態 RVM 整合

本研究將 RVM 分類器分為以距離特徵建立之分類器即 Static RVM (SRVM)，以及以位移特徵建立之分類器即 Dynamic RVM (DRVM)。研究統計發現，有些情緒之間使用距離特徵較容易做分類，有些情緒之間則以位移特徵較容易分類。以表 3-3 與表 3-4 為例，此二表分別為使用 SRVM 與 DRVM，對已訓練過之影像樣本，情緒間分類準確性之結果。舉例來說，由二表可看出生氣與厭惡以位移特

表 3-3、訓練樣本 2 情緒之間以 SRVM 辨識結果。

	Neutral	Anger	Disgust	Fear	Happy	Sadness	Surprise
Neutral	X	90.74%	92.59%	93.51%	100.00%	84.26%	100.00%
Anger	90.74%	X	86.11%	100.00%	100.00%	86.11%	100.00%
Disgust	92.59%	86.11%	X	100.00%	94.44%	91.67%	94.44%
Fear	93.51%	100.00%	100.00%	X	100.00%	86.11%	100.00%
Happy	100.00%	100.00%	94.44%	100.00%	X	94.44%	94.44%
Sadness	84.26%	86.11%	91.67%	86.11%	94.44%	X	100.00%
Surprise	100.00%	100.00%	94.44%	100.00%	94.44%	100.00%	X

表 3-4、訓練樣本 2 情緒之間以 DRVM 辨識結果。

	Neutral	Anger	Disgust	Fear	Happy	Sadness	Surprise
Neutral	X	99.07%	100.00%	97.22%	100.00%	100.00%	100.00%
Anger	99.07%	X	91.67%	94.44%	100.00%	94.44%	100.00%
Disgust	100.00%	91.67%	X	100.00%	100.00%	100.00%	100.00%
Fear	97.22%	94.44%	100.00%	X	91.67%	91.67%	100.00%
Happy	100.00%	100.00%	100.00%	91.67%	X	100.00%	100.00%
Sadness	100.00%	94.44%	100.00%	91.67%	100.00%	X	100.00%
Surprise	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	X

徵較容易做區分，相較於使用距離特徵做分類，其誤判率較低；另外，由二表亦可看出利用 DRVM 做情緒辨識，其誤判率較低，對於整體辨識率有明顯之提升。經過 10 組資料之交叉驗證，本研究歸納如表 3-5 情緒特徵與分類器間之關係，提出整合二分類器之方法，將表 3-5 做為分類器選擇之依據。其中表 3-5 中，S

表 3-5、2 情緒之間分類器之選擇。

	Neutral	Anger	Disgust	Fear	Happy	Sadness	Surprise
Neutral	X	S	S	S	S	S	S
Anger	S	X	D	D	D	D	D
Disgust	S	D	X	D	D	D	D
Fear	S	D	D	X	S	S	S
Happy	S	D	D	S	X	D	D
Sadness	S	D	D	S	D	X	D
Surprise	S	D	D	S	D	D	X

代表 SRVM；D 代表 DRVM。由於每個人臉部所呈現之特定情緒不見得相同，常常會有混淆之情況產生，其中又以中性情緒與生氣、厭惡、傷心之情緒最容易造成混淆。為了避免使用 TOD 判斷特定情緒時，非判斷到使用者最強程度之情緒，導致使用位移特徵影響到整體辨識結果，故本研究以前一次情緒辨識之結果為依據，當前一次情緒辨識結果為特定情緒，此次情緒辨識則使用 SRVM 做辨識，在此本研究以中性情緒做為特定情緒。

無論是 SRVM 或 DRVM，經由上述之 RVM 分類器演算法，已辨識求得兩兩類別間之情緒 Likelihood，接下來則必須對所有 Likelihood 進行整合，計算某個情緒相對於 7 種基本情緒之 Likelihood。

3.3.2. 類別可能性耦合

本質為兩類分類之 RVM 分類器，若應用於多類別之分類，常應用「平行 (Parallel)」或「分層樹(Hierarchical Tree-Based)」之方法判斷輸入特徵所屬類別[47]。然而，使用 Parallel 方法進行整體類別判斷之過程中，有較高之準確性，因其不易受到如 Hierarchical Tree-Based 方法，由上至下分層判斷所包含誤差傳遞之風

險。

Parallel 方法又區分為一對一(One-against-one)與一對所有(One-against-all)之設計。One-against-one 之設計透過二個類別間做比對，再將比對之結果進行整合(例如：投票決定)；One-against-all 之設計則是將某一類別與其他所有類別做比對，再將最大輸出，例如：數值最高者，決定為最後類別。事實上對於 RVM 而言，One-against-all 之做法可視為將原先以 Bernoulli distribution 為基礎取得 Likelihood 的 RVM，改以 Multinomial distribution 為基礎學習(1-2)的權重決定 Likelihood[28]，如(3-16)所示，

$$P(\mathbf{t}|\mathbf{w}) = \prod_{n=1}^N \prod_{k=1}^K \sigma\{y_k(\mathbf{x}_n; \mathbf{w}_k)\}^{t_{n,k}}, \quad (3-16)$$

其中 N 代表 N 組訓練資料； K 代表 K 類。然而，相較於 One-against-one 而言，直接將 RVM 改以 One-against-all 的方法解多類別之 Likelihood，由於 One-against-all 必須同時考量 K 類之權重做計算，因此有相當高的運算複雜度，不適宜即時辨識之應用。本研究採用 One-against-one 之方法，對使用 Bernoulli distribution 為基礎判斷到類別間之 Likelihood 進行耦合(Coupling)[48]，取得 7 種情緒個別之 Likelihood。

假設特徵向量 \mathbf{x} 對於所有類別(K 類， $K > 2$)之 Likelihood 為 $\mathbf{p}(\mathbf{x}) = (p_1(\mathbf{x}), p_2(\mathbf{x}), \dots, p_K(\mathbf{x}))$ ，二個類別間之條件機率(Conditional likelihoods) $r_{ij} = \text{Prob}(i|j)$ 。估測 r_{ij} 之模型如(3-17)所示，

$$\mu_{ij} = \frac{p_i}{p_i + p_j}, \quad (3-17)$$

在此希望找到一個 p_i 使得 μ_{ij} 相當接近 r_{ij} 。由於包含 $K-1$ 個獨立參數， $K(K-1)/2$ 方程式，故無法直接由一般之數值方法解之，必須以迭代的方式求解。本研究使用 Kullback-Leibler 距離(Kullback-Leibler distance, KLD)演算法[49]，求解 μ_{ij} 與 r_{ij} 間的差距，如(3-18)所示，

$$\begin{aligned}\ell(\mathbf{p}) &= \sum_{i \neq j} n_{ij} r_{ij} \log\left(\frac{r_{ij}}{\mu_{ij}}\right) \\ &= \sum_{i < j} n_{ij} \left[r_{ij} \log\left(\frac{r_{ij}}{\mu_{ij}}\right) + (1 - r_{ij}) \log\left(\frac{1 - r_{ij}}{1 - \mu_{ij}}\right) \right],\end{aligned}\quad (3-18)$$

其中 n_{ij} 為第 i 類與第 j 類之訓練樣本數。在此必須找到一組 \mathbf{p} ，使得 $\ell(\mathbf{p})$ 為最小。對(3-18)做一次微分，如(3-19)，並設其為 0，可以得到如(3-20)之關係式。

$$\frac{\partial \ell(\mathbf{p})}{\partial p_i} = \sum_{j:j \neq i} n_{ij} \left(-\frac{r_{ij}}{p_i} + \frac{1}{p_i + p_j} \right), \quad (3-19)$$

$$\begin{aligned}\sum_{j:j \neq i} n_{ij} \mu_{ij} &= \sum_{j:j \neq i} n_{ij} r_{ij}, \quad i = 1, 2, \dots, K, \\ \text{and } \sum_{i=1}^K p_i &= 1,\end{aligned}\quad (3-20)$$

經由以下之步驟進行迭代，則可以取得 p_i 。

Step 1. 初始 p_i ，並找到對應的 μ_{ij} 。

p_i 初始可藉由取所有 Likelihood 之平均，再將某一類本身相對於其他類之所有 Likelihood 做加總取得，如(3-21)所示，

$$p_i = \frac{2 \sum_{i \neq j} r_{ij}}{K(K-1)}, \quad (3-21)$$

事實上，(3-21)為對整體 p_i 做最簡單之估測結果，其亦符合 $\sum_{i=1}^K p_i = 1$ 。

Step 2. 更新 p_i 。

藉由(3-22)更新 p_i ，對更新後之 p_i 進行正規化，使其符合 $\sum_{i=1}^K p_i = 1$ ，並重新計算 μ_{ij} 。

$$p_i^{\text{new}} = p_i \frac{\sum_{j:j \neq i} n_{ij} r_{ij}}{\sum_{j:j \neq i} n_{ij} \mu_{ij}}, \quad (3-22)$$

Step 3. 判斷是否收斂。

每次更新 p_i ，重新計算得到之 $\ell(\mathbf{p}^{\text{new}})$ 會大於更新前之 $\ell(\mathbf{p})$ ，即 $\ell(\mathbf{p}^{\text{new}}) - \ell(\mathbf{p}) \geq 0$ ，故上述之 p_i 更新過程為逐步收斂。當等號成立，即符合(3-20)，則此 \mathbf{p} 即為欲求之解。

Prove :

$$\ell(\mathbf{p}^{\text{new}}) - \ell(\mathbf{p}) = \left(\sum_{i \neq j} n_{ij} r_{ij} \right) \log \frac{\sum_{j: j \neq i} n_{ij} r_{ij}}{\sum_{j: j \neq i} n_{ij} \mu_{ij}} - \sum_{i \neq j} n_{ij} \log \left(\frac{\sum_{j: j \neq i} n_{ij} r_{ij}}{\sum_{j: j \neq i} n_{ij} \mu_{ij}} \mu_{ij} + 1 - \mu_{ij} \right)$$

假設 $a = \sum_{j: j \neq i} n_{ij} r_{ij}$ ， $b = \sum_{j: j \neq i} n_{ij} \mu_{ij}$ ，

$$\begin{aligned} \ell(\mathbf{p}^{\text{new}}) - \ell(\mathbf{p}) &= a \log \frac{a}{b} - \sum_{i \neq j} n_{ij} \log \left[\left(\frac{a}{b} - 1 \right) \mu_{ij} + 1 \right] \\ &\geq a \log \frac{a}{b} - \sum_{i \neq j} n_{ij} \left(\frac{a}{b} - 1 \right) \mu_{ij} \\ &= a \log \frac{a}{b} - (a - b) \geq 0 \end{aligned}$$

3.3.3. 類別可能性時序分析

本研究以連續影像樣本情緒應平滑之轉換為前提，提出 Likelihood 辨識時序分析之策略，以提升辨識準確性為目的。每次辨識 Likelihood 時，皆會以先前辨識 Likelihood 為根據，結合幾何特徵進行此次 Likelihood 之辨識，如(3-23)，以及圖 3-4 所示。

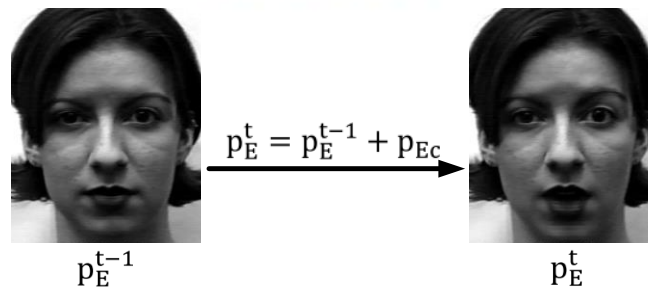


圖 3-4、情緒間 Likelihood 關係示意圖。 p_E^t 代表此次情緒之 Likelihood； p_E^{t-1} 代表前一次情緒之 Likelihood； p_{Ec} 代表 Likelihood 變化量。

$$p_E^t = p_E^{t-1} + p_{Ec}, \quad (3-23)$$

其中 p_E^t 代表此次情緒之 Likelihood，其符合 $\sum p_E^t = 1$ ，下標 E 為情緒類別，上標 t 代表第 t 次； p_{Ec} 代表 Likelihood 變化量，如(3-24)，

$$p_{Ec} = \begin{cases} p_E^t - p_E^{t-1}, & |p_E^t - p_E^{t-1}| \leq T_{Ec} \\ T_{Ec}, & |p_E^t - p_E^{t-1}| > T_{Ec} \end{cases}, \quad (3-24)$$

其符合 $\sum p_{Ec} = 0$ ，且變化量受臨界值 T_{Ec} 之限制，而 T_{Ec} 之決定如(3-25)所示，

$$T_{Ec} = \begin{cases} T_E \times 2^n, & n = 1, 2, 3 \\ T_E \times 10, & n > 3 \end{cases}, \quad (3-25)$$

其中 T_E 代表變化量限制基底，以系統整體處理速度做決定；n 代表 Likelihood 第 n 次同向(正向或負向)變化。 T_{Ec} 影響著 Likelihood 之變化速度，當 T_{Ec} 值愈大，情緒之 Likelihood 收斂愈快，但正確性也可能隨之下降；當 T_{Ec} 值愈小，情緒之 Likelihood 收斂愈慢，亦會影響系統辨識效能。為了確保 Likelihood 辨識之準確性 T_{Ec} 以 2 之倍數成長，且其最大值受到限制於 $T_E \times 10$ 。本研究以 Cohn Kanade (CK+) 資料庫[41]做測試，此資料庫影像擷取速度為 30 fps (frame per second)。研究統計由某一情緒變化至另一情緒平均為 10 張影像樣本，亦即可以假設某個情緒之 Likelihood 由 0 變化至 1 需要 10 張影像樣本，以此為基礎推導出如(3-26) 決定 T_E ，

$$T_E = \left[\sum_{n=1}^3 2^n + 10 \times (10 \times f/30 - 4) \right]^{-1}, \quad f \geq 20, \quad (3-26)$$

其中 式中第 1 項代表前 3 個變化量限制以 2 之倍數成長，第 2 項則代表第 4 個變化量限制在 $T_E \times 10$ ；10 代表 10 張影像樣本；30 代表 30 fps；f 代表目前系統整體處理速度(fps)。事實上，當前 4 張影像 Likelihood 走勢皆相同，前 3 個變化量穩定成長時，即代表使用者情緒 Likelihood 走勢已確定。在走勢不變的條件下，欲判斷之後某張影像最終之 Likelihood，可直接以同一張影像持續丟入 RVM 分類器，並結合(3-23)~(3-26)做計算，進行 Likelihood 辨識，直到收斂($p_{Ec} = 0$) 為止，此即為該影像最終之 Likelihood 值。

情緒之 Likelihood 整體時序辨識與修正過程如圖 3-5 所示。例如：當首次偵測到 Likelihood 走勢為正向之變化($p_{Ec} > 0$)，其變化量限制最大不會超過 $T_{Ec} = T_E \times 2$ ，若經過數次觀察，依舊為正向變化，限制則逐步的放寬，但最大仍受MaxT所限制。在過程中，若 Likelihood 間的走勢轉為負向($p_{Ec} < 0$)，則變化量重新由 $T_{Ec} = T_E \times 2$ 開始做限制，反之亦是如此。經過上述之追蹤設計策略，使情緒間轉換過程更為合理，在某些不確定人臉情緒之情況下，有效提升情緒辨識之準確性。

需特別注意，當個別決定每個 p_E^t 後，必須確認是否符合機率基本定理，即 $\sum p_E^t = 1$ ，進一步推導得知 $\sum p_{Ec} = 0$ ，若無法符合此條件，則必須由(3-27)對 p_{Ec} 進

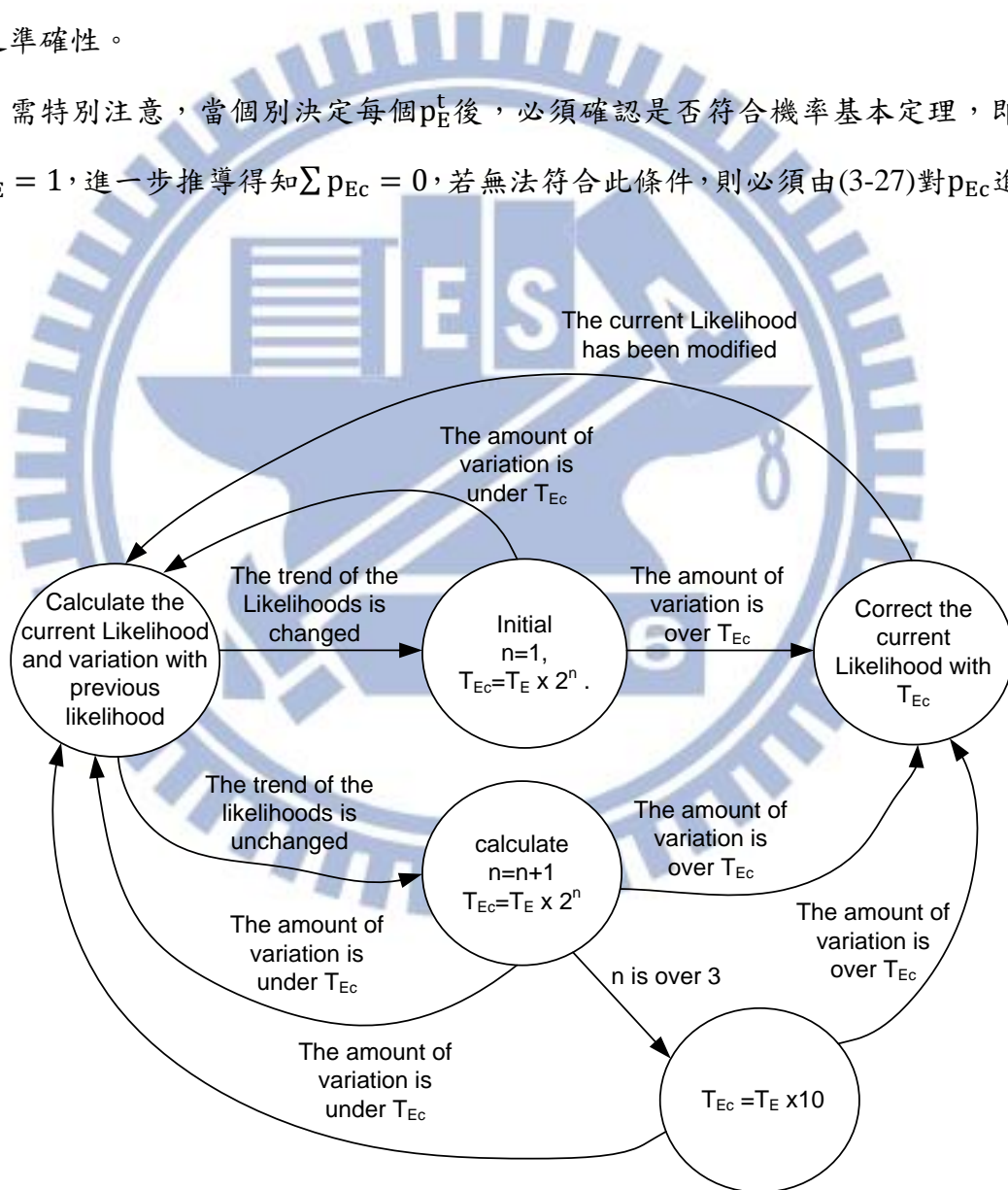


圖 3-5、個別情緒 Likelihood 追蹤之有限狀態機。

行調整取得 p_{Ec}^{new} ，再將 p_{Ec}^{new} 代入(3-28)取得修正後之 p_E^t 即 $p_E^{t, new}$ ，此即為最後情緒類別之 Likelihood。

$$p_{Ec}^{new} = \frac{\sum p_{Ec}}{\# \text{ of category}} + p_{Ec}, \quad (3-27)$$

$$p_E^{t, new} = p_E^t + p_{Ec}^{new}, \quad (3-28)$$

3.4. 基於 A-V Plane 之情緒狀態辨識

人機互動相關研究，常以代表情緒程度之數值(例如：Arousal 與 Valence)，對人機互動系統之機器情緒狀態做設計[50 - 51]，因此本研究希望將使用者情緒狀態，同樣轉換至情緒程度數值，與機器情緒狀態做對應，進行人機互動之設計。本研究以 Russell [15], [52]之 A-V Plane 做為情緒 A-V 辨識依據，如圖 3-6 所示。Russell 提出之 A-V Plane，由代表放鬆與興奮程度之 Arousal，與代表愉快與不愉快程度之 Valence 所構成，其程度值介於-1 ~ 1 之間。A-V Plane 之 A-V 能夠清楚描述情緒之程度與比例等資訊，其亦代表著情緒間之關係，可視為不同情緒之組合[22]，對情緒做更細微之分析。

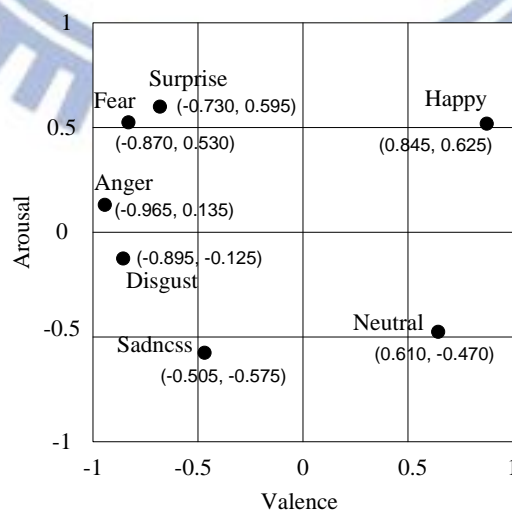


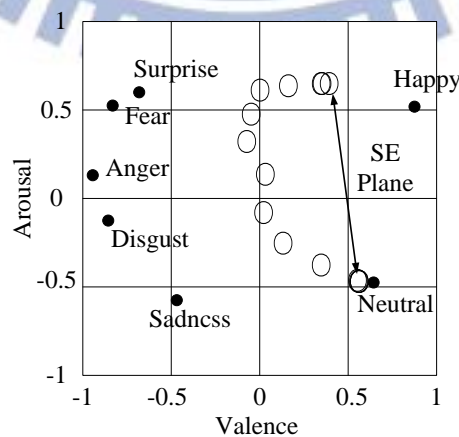
圖 3-6、Russell 提出之 A-V Plane。

本研究以情緒 Likelihood 為根據，進行 RVM 回歸模型(Regression model)之設計，將情緒映射至 A-V Plane 上。本研究分別訓練 Arousal 與 Valence 之 Regression model，對二者進行辨識。在訓練 Regression model 前，必須決定 Likelihood 之組合與其對應之 A-V，本研究參考[22]提出以心理學 2D 模型上基本情緒位置做為權重，計算情緒 A-V 之方法，對情緒 A-V 做初步之估測，如(3-29)所示，其結果如圖 3-7(a)所示，在此將資料庫序列中端點影像 A-V 之連線，視為此序列所在之平面，定義為序列情緒平面(Sequence emotion plane, SE Plane)。

$$\text{Valence} = \frac{\sum_{i=1}^7 \text{Valence}_i \times \text{Likelihood}_i}{\sum_{i=1}^7 \text{Likelihood}_i},$$

$$\text{and Arousal} = \frac{\sum_{i=1}^7 \text{Arousal}_i \times \text{Likelihood}_i}{\sum_{i=1}^7 \text{Likelihood}_i}, \quad (3-29)$$

其中 i 代表基本情緒類別； Valence_i 與 Arousal_i 分別代表 A-V Plane 上特定情緒之座標值，如圖 3-6 所示； Likelihood_i 代表影像樣本由 RVM 分類器辨識之特定情緒 Likelihood。由於 CK+ 資料庫皆為由中性情緒變化至特定情緒之序列影像所構成，因此其端點影像樣本之 A-V，應位於 A-V Plane 之特定情緒位置。本研究將端點影像之 A-V 移至特定位置，其餘 A-V 則修正至相對應位置，以取得不同 Likelihood 組合對應之 A-V 參考值，上述 A-V 之修正可以利用如下過程實現。



(a)

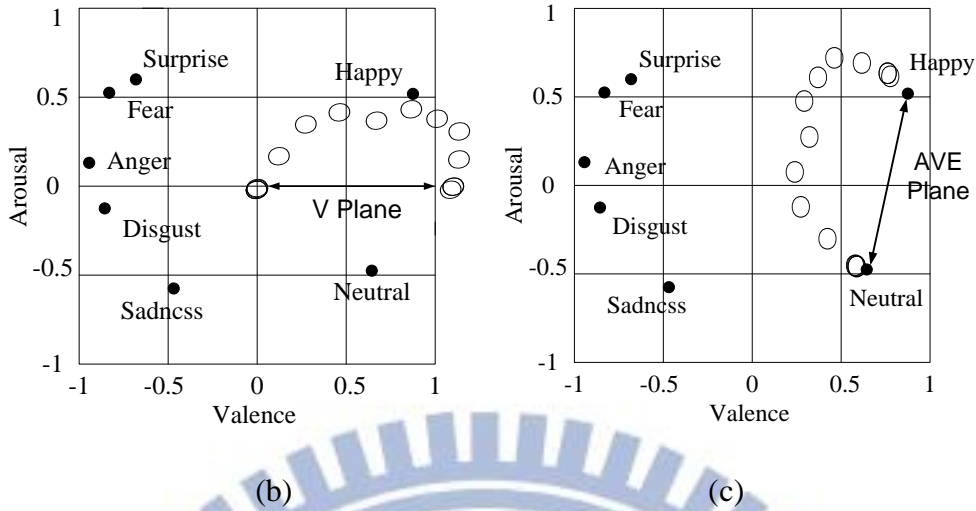


圖 3-7、A-V 參考值之決定，(a)使用[22]提出之基本情緒權重初步計算 A-V 之結果，(b)將初步取得在 SE Plane 之 A-V，旋轉、平移並正規化至 V Plane，(c)將 V Plane 之 A-V 旋轉、平移至 AVE Plane，取得端點影像之 A-V 修正至中性情緒與特定情緒位置之 A-V 參考值。

Step 1. 序列中所有影像在 SE Plane 之 A-V，以序列中第一張影像(中性情緒)之 A-V 做為原點進行旋轉，並平移至 Valence 平面(Valence plane, V Plane)，其中修正後之原點(第一張影像之 A-V)位於 A-V Plane 之原點(A:0, V:0)，如圖 3-7(b)與(3-30)所示，

$$\begin{bmatrix} V_V \\ A_V \end{bmatrix} = \begin{bmatrix} \cos(\theta_{SEV}) & -\sin(\theta_{SEV}) \\ \sin(\theta_{SEV}) & \cos(\theta_{SEV}) \end{bmatrix} \begin{bmatrix} V_{SE} - V_{SE,F} \\ A_{SE} - A_{SE,F} \end{bmatrix}, \quad (3-30)$$

其中 V_V 代表旋轉與平移至 V Plane 之 Valence； A_V 代表旋轉與平移至 V Plane 之 Arousal； V_{SE} 代表在 SE Plane 之 Valence； A_{SE} 代表在 SE Plane 之 Arousal； $V_{SE,F}$ 代表在 SE Plane 中性情緒之 Valence； $A_{SE,F}$ 代表在 SE Plane 中性情緒之 Arousal； θ_{SEV} 代表 SE Plane 至 V Plane 之角度。

Step 2. 在此先將 A-V Plane 上，中性情緒所在位置 $P_{AVE,N}$ ，與資料庫序列相同之特定情緒所在位置 $P_{AVE,S}$ ，兩點形成之連線，定義為 Arousal 與 Valence 情緒平面(Arousal and valence emotion plane, AVE Plane)。藉由(3-31)正規化 V

Plane 之序列 A-V，其結果如圖 3-7(b)所示，

$$V'_V = \frac{(V_V - V_{V,F})}{(V_{V,E} - V_{V,F})} \times \|\overline{P_{AVE,S}P_{AVE,N}}\|,$$

and $A'_V = A_V,$ (3-31)

其中 V'_V 代表正規化後之 V_V ； A'_V 代表正規化後之 A_V ； $V_{V,F}$ 代表在 V Plane 第一張影像之 Valence； $V_{V,E}$ 代表在 V Plane 最後一張影像之 Valence。

Step 3. 利用(3-32)將 V Plane 上，正規化後之序列 A-V，以第一張影像之 A-V 為原點進行旋轉，並平移至 AVE Plane，其中修正後之原點(第一張影像之 A-V)位於 A-V Plane 上 Neutral 所在之 A-V 位置，如圖 3-7(c)所示。

$$\begin{bmatrix} V_{AVE} \\ A_{AVE} \end{bmatrix} = \begin{bmatrix} \cos(\theta_{VAVE}) & -\sin(\theta_{VAVE}) \\ \sin(\theta_{VAVE}) & \cos(\theta_{VAVE}) \end{bmatrix} \begin{bmatrix} V'_V - V'_{V,F} \\ A'_V - A'_{V,F} \end{bmatrix} + \begin{bmatrix} V_{AV,F} \\ A_{AV,F} \end{bmatrix}, \quad (3-32)$$

其中 V_{AVE} 代表旋轉與平移至 AVE Plane 之 Valence； A_{AVE} 代表旋轉與平移至 AVE Plane 之 Arousal； $A'_{V,F}$ 代表在 V Plane 第一張影像正規化後之 A_V ； θ_{VAVE} 代表 V Plane 至 AVE Plane 之角度。

經由上述之步驟，即可取得本研究情緒 Likelihood 對應之 A-V 參考值，如圖 3-7(b) 所示，以訓練 A-V 之 RVM Regression model。以下就 RVM Regression model 訓練演算法做詳細說明與推導。

RVM 回歸分析與 RVM 分類演算法相似，只是多了誤差項 ϵ 之估測，如(3-33) 所示，

$$t_n = y(\mathbf{x}_n; \mathbf{w}) + \epsilon_n, \quad (3-33)$$

其中 $\{\mathbf{x}_n, t_n\}_{n=1}^N$ 代表給定之一組訓練資料， \mathbf{x}_n 代表 Likelihood，其做為輸入特徵向量； t_n 代表 Likelihood 對應之 Arousal 或 Valence 加上誤差項之估測； $y(\mathbf{x}_n; \mathbf{w})$ 之定義如(1-2)所示。進一步將上式改寫為機率之形式，以求解權重 \mathbf{w} 與誤差項 ϵ ，如(3-34)所示，

$$p(t_n | \mathbf{x}) = \mathcal{N}(t_n | y(\mathbf{x}_n; \mathbf{w}), \sigma^2), \quad (3-34)$$

其中 $\mathcal{N}(\cdot)$ 代表高斯分佈函數，上式代表平均值為 $y(\mathbf{x}_n; \mathbf{w})$ ，變異數即誤差項為 σ^2 ，即 $\epsilon_n \sim \mathcal{N}(0, \sigma^2)$ 之 t_n 高斯分佈。假設 $\{t_n\}_{n=1}^N$ 為彼此獨立之隨機變數，在已知 \mathbf{w} 與 σ^2 之情況下，完整 \mathbf{t} 的 Likelihood 分佈可寫為如(3-35)所示，

$$\begin{aligned} p(\mathbf{t}|\mathbf{w}, \sigma^2) &= \prod_{n=1}^N \mathcal{N}(t_n|y(\mathbf{x}_n; \mathbf{w}), \sigma^2) \\ &= (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left\{-\frac{1}{2\sigma^2} \|\mathbf{t} - \Phi\mathbf{w}\|^2\right\}, \end{aligned} \quad (3-35)$$

其中 $\mathbf{t} = (t_1, t_2, \dots, t_N)^T$;

$\mathbf{w} = (w_1, w_2, \dots, w_N)^T$;

$\Phi = [\phi(\mathbf{x}_1), \phi(\mathbf{x}_2), \dots, \phi(\mathbf{x}_N)]^T$,

$\phi(\mathbf{x}_n) = [1, K(\mathbf{x}_n, \mathbf{x}_1), K(\mathbf{x}_n, \mathbf{x}_2), \dots, K(\mathbf{x}_n, \mathbf{x}_N)]^T$ 。

由於 $p(\mathbf{t}|\mathbf{w}, \sigma^2)$ 為高斯分佈，故可利用定積分求解。在已知 \mathbf{t} 機率分佈之情況下，所要求解之條件機率如(3-36)所示，

$$\begin{aligned} p(t_*|\mathbf{t}) &= \int p(t_*|\mathbf{w}, \sigma^2) p(\mathbf{w}, \sigma^2|\mathbf{t}) d\mathbf{w} d\sigma^2 \\ &= \int p(t_*|\mathbf{w}, \sigma^2) \frac{p(\mathbf{t}|\mathbf{w}, \sigma^2) p(\mathbf{w}, \sigma^2)}{p(\mathbf{t})} d\mathbf{w} d\sigma^2, \end{aligned} \quad (3-36)$$

其中 t_* 為預測目標值。通常在求解 \mathbf{w} 與 σ^2 ，會希望從(30)做 \mathbf{w} 與 σ^2 的最大可能性估測(Maximum likelihood estimation, MLE)，但如此往往造成 \mathbf{w} 中大部份的元素非 0，而導致 Over-fitting 的情況產生。為了避免此情況發生，與 RVM 分類演算法相同需限制 \mathbf{w} 機率分佈為 Zero-mean Gaussian distribution，如(3-6)所示。經由此限制，可將(3-36)改寫為(3-37)，

$$p(t_*|\mathbf{t}) = \int p(t_*|\mathbf{w}, \alpha, \sigma^2) p(\mathbf{w}, \alpha, \sigma^2|\mathbf{t}) d\mathbf{w} d\alpha d\sigma^2, \quad (3-37)$$

由於多了 α 之條件限制，故無法直接從(3-37)求解 \mathbf{w} 與 σ^2 。前者 $p(t_*|\mathbf{w}, \alpha, \sigma^2)$ ， t_* 只與 \mathbf{w} ， σ^2 有直接關係，在積分上沒問題。然而，後者 $p(\mathbf{w}, \alpha, \sigma^2|\mathbf{t})$ 則必須利用貝式定理進行拆解，如下過程所示，

$$\begin{aligned}
p(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2 | \mathbf{t}) &= \frac{p(\mathbf{w}, \mathbf{t}, \boldsymbol{\alpha}, \sigma^2)}{p(\mathbf{t})} \\
&= \frac{p(\mathbf{w}, \mathbf{t}, \boldsymbol{\alpha}, \sigma^2)}{p(\mathbf{t}, \boldsymbol{\alpha}, \sigma^2)} \cdot \frac{p(\mathbf{t}, \boldsymbol{\alpha}, \sigma^2)}{p(\mathbf{t})} \\
&= p(\mathbf{w} | \mathbf{t}, \boldsymbol{\alpha}, \sigma^2) p(\boldsymbol{\alpha}, \sigma^2 | \mathbf{t}), \\
p(\mathbf{w} | \mathbf{t}, \boldsymbol{\alpha}, \sigma^2) &= \frac{p(\mathbf{w}, \mathbf{t}, \boldsymbol{\alpha}, \sigma^2)}{p(\mathbf{t}, \boldsymbol{\alpha}, \sigma^2)} \\
&= \frac{p(\mathbf{t} | \mathbf{w}, \boldsymbol{\alpha}, \sigma^2) p(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2)}{p(\mathbf{t} | \boldsymbol{\alpha}, \sigma^2) p(\boldsymbol{\alpha}, \sigma^2)} \\
&= \frac{p(\mathbf{t} | \mathbf{w}, \boldsymbol{\alpha}, \sigma^2) p(\mathbf{w} | \boldsymbol{\alpha}, \sigma^2) p(\boldsymbol{\alpha}, \sigma^2)}{p(\mathbf{t} | \boldsymbol{\alpha}, \sigma^2) p(\boldsymbol{\alpha}, \sigma^2)} \\
&= \frac{p(\mathbf{t} | \mathbf{w}, \boldsymbol{\alpha}, \sigma^2) p(\mathbf{w} | \boldsymbol{\alpha}, \sigma^2)}{p(\mathbf{t} | \boldsymbol{\alpha}, \sigma^2)},
\end{aligned}$$

由於 \mathbf{t} 只與 \mathbf{w} , σ^2 有直接關係，而 \mathbf{w} 只與 $\boldsymbol{\alpha}$ 有直接關係，故可將 $p(\mathbf{w} | \mathbf{t}, \boldsymbol{\alpha}, \sigma^2)$ 改寫如下所示，

$$p(\mathbf{w} | \mathbf{t}, \boldsymbol{\alpha}, \sigma^2) = \frac{p(\mathbf{t} | \mathbf{w}, \sigma^2) p(\mathbf{w} | \boldsymbol{\alpha})}{p(\mathbf{t} | \boldsymbol{\alpha}, \sigma^2)},$$

由貝式定理可將 $p(\mathbf{w} | \mathbf{t}, \boldsymbol{\alpha}, \sigma^2)$ 改寫為

$$p(\mathbf{w} | \mathbf{t}, \boldsymbol{\alpha}, \sigma^2) = \frac{p(\mathbf{t} | \mathbf{w}, \sigma^2) p(\mathbf{w} | \boldsymbol{\alpha})}{\int p(\mathbf{t} | \mathbf{w}, \sigma^2) p(\mathbf{w} | \boldsymbol{\alpha}) d\mathbf{w}'}$$

由於 $p(\mathbf{t} | \mathbf{w}, \sigma^2)$ 與 $p(\mathbf{w} | \boldsymbol{\alpha})$ 皆為高斯函數，故可求解二者乘積之定積分，如(3-38)所示，

$$\begin{aligned}
p(\mathbf{t} | \boldsymbol{\alpha}, \sigma^2) &= \int p(\mathbf{t} | \mathbf{w}, \sigma^2) p(\mathbf{w} | \boldsymbol{\alpha}) d\mathbf{w}, \\
&= (2\pi)^{-\frac{N}{2}} |\sigma^2 \mathbf{I} + \boldsymbol{\Phi} \mathbf{A}^{-1} \boldsymbol{\Phi}^T|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \mathbf{t}^T (\sigma^2 \mathbf{I} + \boldsymbol{\Phi} \mathbf{A}^{-1} \boldsymbol{\Phi}^T)^{-1} \mathbf{t} \right\}, \quad (3-38)
\end{aligned}$$

其中 $\mathbf{A} = \text{diag}(\alpha_0, \alpha_1, \dots, \alpha_N)$ 。整理過後可得(3-39) ~ (3-41)，

$$p(\mathbf{w} | \mathbf{t}, \boldsymbol{\alpha}, \sigma^2) = (2\pi)^{-\frac{N+1}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{w} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{w} - \boldsymbol{\mu}) \right\}, \quad (3-39)$$

$$\boldsymbol{\Sigma} = (\sigma^{-2} \boldsymbol{\Phi}^T \boldsymbol{\Phi} + \mathbf{A})^{-1}, \quad (3-40)$$

$$\boldsymbol{\mu} = \sigma^{-2} \boldsymbol{\Sigma} \boldsymbol{\Phi}^T \mathbf{t}, \quad (3-41)$$

經過上述之推導，可將(3-37)改寫為(3-42)。

$$p(t_*|\mathbf{t}) = \int p(t_*|\mathbf{w}, \boldsymbol{\alpha}, \sigma^2)p(\mathbf{w}|\mathbf{t}, \boldsymbol{\alpha}, \sigma^2)p(\boldsymbol{\alpha}, \sigma^2|\mathbf{t}) d\mathbf{w}d\boldsymbol{\alpha}d\sigma^2, \quad (3-42)$$

Tipping[28]對(3-42)做如(3-43)之假設，作者提到以點估測(Point-estimate)之機率分佈，與完整式子估測之機率分佈幾乎相同，因此預測出來之結果也將幾乎一致。

$$\int p(t_*|\boldsymbol{\alpha}, \sigma^2)\delta(\boldsymbol{\alpha}_{MP}, \sigma_{MP}^2) d\boldsymbol{\alpha}d\sigma^2 \approx \int p(t_*|\boldsymbol{\alpha}, \sigma^2)p(\boldsymbol{\alpha}, \sigma^2|\mathbf{t}) d\boldsymbol{\alpha}d\sigma^2, \quad (3-43)$$

藉由(3-43)之假設，剩下的問題則是解 $\boldsymbol{\alpha}_{MP}$ 與 σ_{MP} 。

$$\begin{aligned} (\boldsymbol{\alpha}_{MP}, \sigma_{MP}^2) &= \arg \max_{\boldsymbol{\alpha}, \sigma^2} p(\boldsymbol{\alpha}, \sigma^2|\mathbf{t}) \\ &= \arg \max_{\boldsymbol{\alpha}, \sigma^2} p(\mathbf{t}|\boldsymbol{\alpha}, \sigma^2)p(\boldsymbol{\alpha})p(\sigma^2) \\ &= \arg \max_{\boldsymbol{\alpha}, \sigma^2} p(\mathbf{t}|\boldsymbol{\alpha}, \sigma^2), \end{aligned} \quad (3-44)$$

由於從(3-38)無法直接利用MLE求解 $\boldsymbol{\alpha}_{MP}$ 與 σ_{MP} ，因此必須使用近似的方式求解。利用(3-38)分別對 $\boldsymbol{\alpha}$ 與 σ^2 偏微分，並假設其等於0，可得到如(3-45)~(3-46)更新 $\boldsymbol{\alpha}$ 與 σ^2 的式子[46]。

$$\alpha_i^{\text{new}} = \frac{\gamma_i}{\mu_i}, \quad (3-45)$$

$$\sigma^{2\text{new}} = \frac{\|\mathbf{t} - \Phi\mathbf{w}\|^2}{N - \sum_i \gamma_i}, \quad (3-46)$$

其中 $\gamma_i = 1 - \alpha_i N_{ii}$ ， N_{ii} 代表 $\boldsymbol{\Sigma}$ 之第*i*項在對角線上之元素； N 代表訓練資料樣本數。對 $\boldsymbol{\alpha}$ 與 σ^2 先進行初始，藉由(3-45)~(3-46)分別對二者進行更新，再配合(3-40)~(3-41)更新 $\boldsymbol{\Sigma}$ 與 $\boldsymbol{\mu}$ 。同樣的，經過數次的迭代，大多數的 α_i 會趨近於無限大，亦即 w_i 受到限制，使其值趨近於0，使得 $N_{ii} \approx \alpha_i^{-1}$ 而 $\gamma_i \approx 0$ ，取得相對稀疏之解。其餘的 α_i 則會趨近於定值，使得 $\gamma_i \approx 1$ ，所對應到的 \mathbf{x}_i 即為相對向量(relevance vector, RV)。

當求得 $\boldsymbol{\alpha}_{MP}$ 與 σ_{MP} ，即可從給定之Likelihood得到對應之Arousal或Valence值。對於新的資料 \mathbf{x}_* 而言，藉由(3-43)之假設，可以將(3-42)改寫為(3-47)，預測其分佈。

$$p(t_*|\mathbf{t}) = \int p(t_*|\mathbf{w}, \boldsymbol{\alpha}_{MP}, \sigma_{MP}^2) p(\mathbf{w}|\mathbf{t}, \boldsymbol{\alpha}_{MP}, \sigma_{MP}^2) d\mathbf{w}, \quad (3-47)$$

在上式積分式中之二項皆為高斯分佈，因此定積分結果亦為高斯分佈，如(3-48)所示。

$$p(t_*|\mathbf{t}) = \mathcal{N}(t_*|y_*, \sigma_*^2), \quad (3-48)$$

其中

$$y_* = \boldsymbol{\mu}^T \boldsymbol{\phi}(\mathbf{x}_*), \quad (3-49)$$

$$\sigma_*^2 = \sigma_{MP}^2 + \boldsymbol{\phi}(\mathbf{x}_*)^T \boldsymbol{\Sigma} \boldsymbol{\phi}(\mathbf{x}_*), \quad (3-50)$$

其中 \mathbf{x}_* 代表新的測試樣本 Likelihood。然而，若僅欲求解新的測試樣本對應之 Arousal 或 Valence，而不進行其分佈之預測，則可直接由(3-49)解之。

藉由 RVM 分類器進行基本情緒 Likelihood 辨識，並利用 RVM 回歸模型將 Likelihood 轉換為 A-V，進行情緒程度與情緒轉換分析，本研究設計一套情緒辨識之音樂選取系統，依據不同場合與使用者情緒，選擇適當之音樂做播放，設法改變使用者情緒，實現人機互動之目的。

第四章、實驗結果

本研究之實驗結果主要包含三個部分之驗證，分別為特徵擷取準確性、情緒 Likelihood 辨識正確率，以及 Likelihood 轉換 A-V Plane 之合理性與正確性。人臉特徵擷取為人臉情緒辨識中相當重要之一環，特徵擷取準確性將影響系統辨識之效能；而完善之人機互動系統具備準確之辨識率，方能以合理、受使用者接受之互動方式與使用者互動。最後，在本章節中亦會對本研究提出之情緒辨識應用於智慧型音樂選取之系統設計進行實驗。

本實驗以 Cohn-Kanade (CK+) 資料庫[41]進行情特徵擷取與情緒類別分析，資料庫中共包含 123 位，年齡介於 18 到 50 歲之間人仕之樣本，其中 69% 為女生；81% 來自歐美國家、13% 與非裔美洲人，其餘 6% 來自不同的族群。資料庫中每種情緒皆由中性人臉情緒變化至特定人臉情緒之序列影像所構成，每組序列影像約 10 到 60 張影像，特定人臉情緒包含生氣、厭惡、害怕、高興、傷心與驚訝，影像解析度為 640×490 ，其中一組序列影像範例如圖 4-1 所示。另外，CK+ 資料庫包含建立 AAM 人臉模型所需之標註特徵點，如本論文 2.3.1 節所述。

4.1. 特徵點偵測結果

本研究以 AAM 建立人臉模型，從中擷取 18 個人臉特徵點。特徵點偵測準確性方面，本實驗從 CK+ 資料庫中，每種情緒各取 100 張測試影像樣本，共 600 張測試影像樣本，其中包含生氣、厭惡、害怕、傷心、高興與驚訝 6 種情緒序列影像樣本，而序列中包含不同程度之基本情緒，以此進行 18 個特徵點偵測準確性判斷，18 個特徵點如圖 3-1 所示。圖 4-2 為其中 7 張測試影像樣本，分別代表 7 種基本情緒(含中性)之特徵點偵測結果。



圖 4-1、CK+ 資料庫序列影像範例，由中性表情變化至驚訝表情。

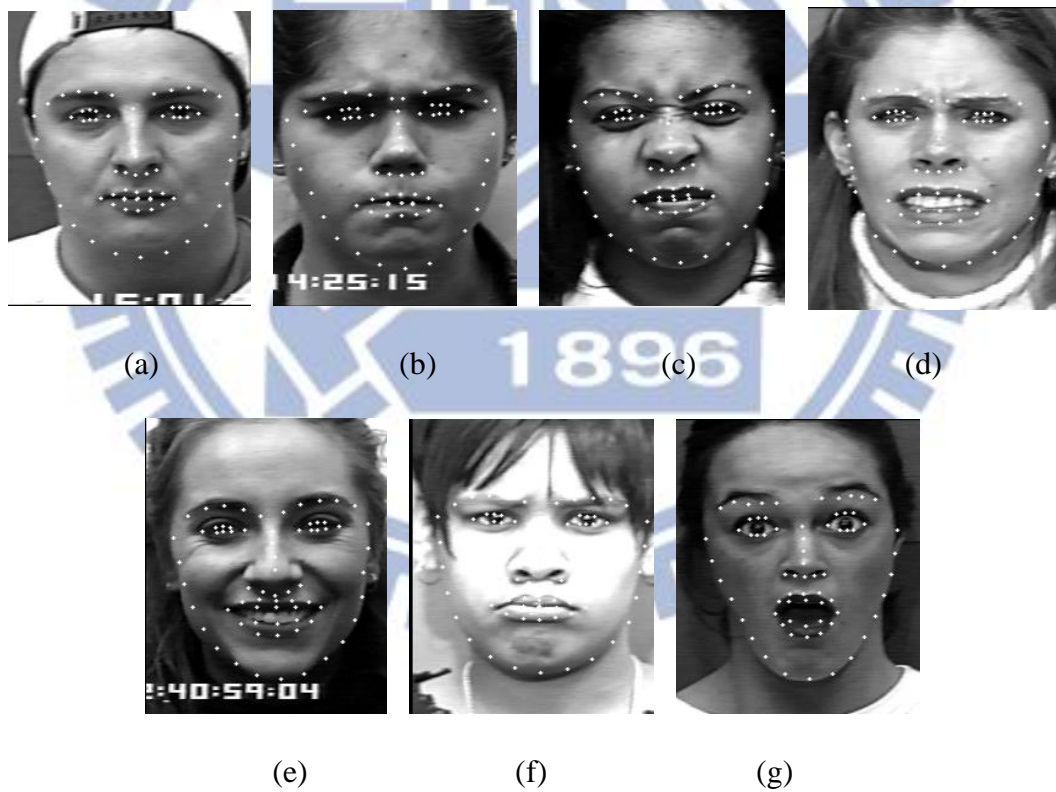


圖 4-2、7 種基本表情特徵點偵測結果，(a)中性，(b)生氣，(c)厭惡，(d)害怕，(e)高興，(f)傷心，(g)驚訝。

測試樣本 18 個特徵點偵測準確性如表 4-1 所示，其結果為從 AAM 建立之人臉模型中，擷取本研究感興趣之 18 個特徵點，與手動標點之結果，比較其誤差之像素(Pixel)。由表 4-1 可看出，P1 ~ P12 特徵點之偵測平均誤差約為 1 ~ 3

表 4-1、18 個特徵點偵測平均<最大>誤差(單位：Pixel)。

	Anger	Disgust	Fear	Happy	Sadness	Surprise	Average
P1	1.22 <3.16>	1.44 <4.12>	1.46 <5.00>	1.11 <2.24>	1.41 <3.61>	1.16 <4.00>	1.30 <5.00>
P2	2.36 <6.08>	2.78 <7.28>	2.42 <6.32>	2.60 <6.08>	1.98 <5.39>	2.04 <7.28>	2.36 <7.28>
P3	1.90 <5.00>	2.76 <7.07>	2.17 <6.32>	2.67 <5.10>	2.21 <7.07>	1.98 <7.00>	2.28 <7.07>
P4	3.25 <7.07>	2.11 <7.62>	2.12 <5.10>	1.77 <7.07>	1.85 <4.47>	2.21 <5.39>	2.22 <7.62>
P5	3.12 <6.71>	3.21 <7.81>	2.45 <7.00>	1.92 <4.00>	2.84 <6.40>	2.17 <6.32>	2.62 <7.81>
P6	1.90 <6.40>	2.73 <6.40>	2.27 <7.28>	1.66 <4.00>	2.98 <6.40>	2.15 <7.07>	2.28 <7.28>
P7	2.47 <5.83>	2.40 <7.28>	2.09 <4.47>	1.89 <5.39>	2.13 <5.66>	2.04 <5.10>	2.16 <7.28>
P8	2.35 <5.10>	2.05 <6.71>	1.94 <4.24>	1.70 <4.12>	2.14 <7.07>	1.67 <3.61>	1.97 <7.07>
P9	2.21 <6.71>	3.01 <7.62>	2.22 <5.39>	0.89 <2.24>	2.34 <7.07>	2.13 <5.39>	2.13 <7.62>
P10	2.05 <4.24>	2.48 <7.00>	2.32 <7.07>	1.06 <4.47>	3.01 <7.21>	2.07 <6.32>	2.16 <7.21>
P11	2.63 <7.07>	3.44 <7.07>	1.90 <5.10>	1.88 <5.10>	2.23 <6.00>	2.04 <6.40>	2.35 <7.07>
P12	2.30 <5.39>	1.87 <8.25>	1.69 <4.12>	1.36 <4.24>	1.74 <7.07>	1.68 <5.39>	1.77 <8.25>
P13	4.38 <7.00>	5.86 <8.25>	3.83 <8.49>	4.90 <7.21>	4.63 <7.62>	4.66 <10.20>	4.71 <10.20>
P14	2.84 <6.71>	5.16 <8.94>	4.05 <8.06>	3.81 <7.62>	4.06 <7.81>	4.95 <11.70>	4.15 <11.70>
P15	4.21 <7.07>	4.61 <8.54>	3.95 <7.62>	3.83 <7.81>	4.58 <7.81>	5.31 <11.70>	4.41 <11.70>
P16	3.85 <6.71>	4.45 <9.90>	3.86 <7.62>	4.08 <7.81>	5.33 <7.81>	4.98 <10.82>	4.43 <10.82>
P17	2.75 <6.71>	4.29 <8.60>	3.31 <7.81>	3.94 <7.81>	3.51 <8.60>	4.13 <10.44>	3.65 <10.44>
P18	3.03 <6.08>	4.63 <8.06>	3.69 <7.28>	4.21 <7.07>	3.73 <7.81>	3.76 <11.00>	3.84 <11.00>

個 Pixel。而 P13 ~ P18 為眉毛之特徵點，其平均誤差相對來得大，約 3 ~ 6 個 Pixel 之誤差，主要原因在於本研究所使用之 AAM 演算法，對於不同距離、不同人臉大小之測試樣本，其建立之人臉模型會有些許誤差，尤其是人臉外圍之輪廓，或者是形變過大之表情變化(例如：驚訝)。然而，P1 ~ P18 平均約有 5 個 Pixel 以內之誤差，仍在情緒辨識可接受之範圍內。

4.2. 基本情緒辨識結果

當由 AAM 擷取特徵點後，即可計算情緒特徵值，進行情緒 Likelihood 辨識。在本實驗過程中，為了避免特徵點偵測之誤差，而造成特徵擷取不準確，因此在本論文 4.2 ~ 4.4 節之實驗，利用手動標示之 18 個特徵點，擷取情緒辨識之特徵，如本論文 3.3 節所述之 16 個幾何特徵，以驗證本研究情緒 Likelihood 辨識與 A-V 辨識之效能。

本實驗從 CK+ 資料庫中已標註情緒類別之序列影像，每種情緒(生氣、厭惡、害怕、高興、傷心與驚訝)各取 10 組由中性情緒變化至特定基本情緒之序列影像樣本，共 60 組序列影像樣本做為基本情緒類別之辨識效果驗證。在本實驗中每組序列影像用來訓練 RVM 分類器，進而判斷端點影像之情緒 Likelihood。圖 4-3 以 10 張序列影像樣本為例，序列中，前 3 張影像(No. 1-3)選其中 1 張做為中性情緒之訓練樣本，其餘 2 張則做為判斷基本情緒類別辨識結果之端點影像樣本；

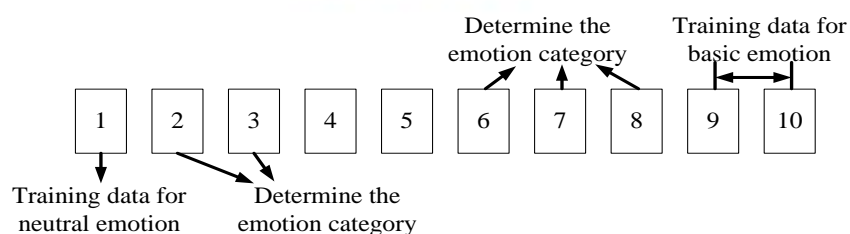


圖 4-3、序列影像中訓練 RVM 分類器與判斷情緒 Likelihood 影像樣本示意圖。

最後 5 張影像(No. 6-10)選其中 2 張做為特定情緒之訓練樣本，其餘 3 張則做為判斷基本情緒類別辨識結果之端點影像樣本。本實驗以上述之訓練樣本，利用距離特徵與位移特徵分別訓練 3.3 節所述之 RVM 靜態分類器(SRVM)與動態分類器(DRVM)。

實驗中分別以本研究提出之 Likelihood 時序辨識，與情緒 Likelihood 非時序辨識，進行 7 種基本情緒類別(含中性)辨識比較，而情緒類別之判定，以影像樣本中 Likelihood 值最大者決定之，然而，為了與 Likelihood 非時序辨識進行基本情緒類別之比較，必須確保 Likelihood 時序辨識其端點影像 Likelihood 已完全收斂，因此當(3-25)限制之變化量臨界值到達 $T_E \times 10$ 時，即 Likelihood 走勢已確定，本實驗將端點影像持續丟入 RVM 分類器做 Likelihood 辨識直到收斂為止，如本論文 3.3.3 節之說明，以此結果判斷端點影像之基本情緒類別。其中本實驗 Likelihood 非時序辨識，其辨識結果為直接由本論文 3.3.1a 節之 RVM 演算法做情緒辨識，而未加入 DRVM(位移特徵訓練之 RVM，如 3.3.1b 節之說明)，以及

表 4-2、Likelihood 時序辨識方法之基本情緒辨識率。

Observation Truth	Neutral	Anger	Disgust	Fear	Happy	Sadness	Surprise
Neutral	95.83%	1.67%	0.56%	0.28%	0.00%	1.67%	0.00%
Anger	3.33%	91.11%	2.22%	0.00%	0.00%	3.33%	0.00%
Disgust	5.56%	0.00%	94.44%	0.00%	0.00%	0.00%	0.00%
Fear	3.53%	0.00%	0.00%	90.68%	3.49%	2.30%	0.00%
Happy	0.00%	0.00%	0.00%	2.22%	97.78%	0.00%	0.00%
Sadness	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%
Surprise	0.00%	0.00%	0.00%	1.11%	0.00%	0.00%	98.89%

Total average: 95.64%

表 4-3、Likelihood 非時序辨識方法之基本情緒辨識率。

Observation Truth	Neutral	Anger	Disgust	Fear	Happy	Sadness	Surprise
Neutral	96.67%	0.28%	1.67%	0.00%	0.00%	1.39%	0.00%
Anger	11.11%	83.33%	0.00%	0.00%	0.00%	5.56%	0.00%
Disgust	14.44%	2.22%	83.33%	0.00%	0.00%	0.00%	0.00%
Fear	17.78%	0.00%	0.00%	82.22%	0.00%	0.00%	0.00%
Happy	2.22%	0.00%	0.00%	4.44%	93.33%	0.00%	0.00%
Sadness	20.00%	0.00%	0.00%	0.00%	0.00%	80.00%	0.00%
Surprise	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%

Total average: 90.89%

Likelihood 時序分析(如本論文 3.3.3 節)之情緒辨識方法。表 4-2 為本研究提出情緒辨識方法之基本情緒辨識率，表 4-3 為 Likelihood 非時序辨識方法之基本情緒辨識率。由此二表可看出，個別情緒辨識率由本研究提出之方法整體而言比 Likelihood 非時序辨識方法來得好，甚至有大幅度之改善，而本研究提出之情緒辨識方法平均辨識率亦高於 Likelihood 非時序辨識方法，且對於較不易與中性情緒區分之生氣、厭惡、害怕與傷心情緒，受到中性情緒之影響則有所改善，傷心情緒辨識率更是大幅的提升。

另外，與其他同樣使用 CK+ 資料庫，對情緒做 7 種基本情緒分類之研究比較如表 4-4 所示，其中 Jain 等人[53]利用序列影像資訊對人臉情緒做辨識，此研究結合隱藏變量動態條件隨機域模型(Latent-dynamic conditional random fields, LDCRFs)與條件隨機域模型(Conditional random fields, CRFs)，經由序列影像建立描述人臉形狀微妙變化之機率模型，有效區分相似人臉表情之動態形狀特徵，藉由人臉形狀時序變化資訊對情緒類別做時序分析與情緒類別辨識。[17, 54]則以

單張影像辨識基本情緒為主，Song 等人[17] 提出結合皮膚形變參數(Skin deformation parameters, SDPs)與臉部動態參數(Facial animation parameters, FAPs)之分析，藉由已知的使用者中性表情為基礎，擷取影像強度比例特徵與位移特徵判斷人臉形變，最後以支持向量機(Support vector machine, SVM)進行情緒分類。Silva 等人[54]則以局部二值模式(Local binary pattern, LBP)為基礎進行改良，對人臉影像以像素之梯度大小做編碼，擷取人臉情緒辨識特徵，並以卡方度量(Chi-squared metric)為基礎，設計人臉情緒分類器，對人臉情緒做分類。

本研究提出之情緒辨識方法，利用時序分析對序列情緒做辨識，最後以最強程度之端點影像做基本情緒類別判斷與比較。經由相關研究之比較可以看出，本研究提出之方法，個別情緒辨識率皆可達到 90% 以上，相較於同樣以連續人臉資訊做為辨識特徵之研究[53]，由於特徵之選擇與分類器之設計有所不同，造成個別情緒辨識效能之優劣有所差異。然而，整體平均辨識率皆較相關研究之方法還要來得高。主要原因包含兩個部份，其一為本研究時序分析之方法，如本論文

表 4-4、本研究提出之 Likelihood 時序辨識方法與相關研究之比較。

Category	Song et. al. [17]	Jain et. al. [53]	Silva et. al. [54]	Proposed method
Neutral	90.07%	73.46%	91.5%	95.83%
Anger	90.57%	76.71%	78.2%	91.11%
Disgust	86.05%	81.51%	92.5%	94.44%
Fear	84.62%	94.37%	71.7%	90.68%
Happy	93.62%	98.55%	93.9%	97.78%
Sadness	90.24%	77.22%	70.4%	100.00%
Surprise	92.31%	99.06%	85.2%	98.89%
Average	89.64%	85.54%	85.3%	95.83%

3.3.3 節，能夠避免單張影像不確定之情緒特徵而造成誤判；其二則是本研究於序列影像中取得特定情緒特徵為基準，將距離特徵結合位移特徵進行情緒辨識，如本論文 3.2 節之說明，有效提升情緒辨識之效能。

4.3. 混合情緒辨識結果

本實驗挑選 CK+資料庫中未標註情緒類別之序列影像，進行混合情緒問卷調查，更進一步驗證本方法對混合情緒類別辨識性能做驗證。資料庫中有未標註情緒類別之序列影像，意味著其端點影像具有不確定性，可能為數種情緒之混合，故本實驗選擇此類序列之端點影像做問卷調查。本問卷共挑選 CK+資料庫未標註情緒類別 10 組序列之端點影像供 13 位受測者觀看。受測者根據本身認知，判斷影像包含哪些基本表情，以百分比形式填入該影像 7 種基本表情(含中性)之比例，問卷之樣張如附錄 A 所示。

本實驗經由 Likelihood 時序辨識，判斷序列之端點影像基本情緒類別 Likelihood 組合之辨識結果，以驗證本系統能夠對混合情緒 Likelihood 做適當之判斷。其中 6 組序列影像之端點影像如圖 4-4 ~ 圖 4-9 所示，辨識結果與問卷調查結果如表 4-5 ~ 表 4-10。在表 4-5 ~ 表 4-10 中間卷調查之平均為，13 位受測者判斷該影像每個情緒類別之比例平均，而問卷調查每個情緒之均方根誤差 (Root-mean-square-error, RMSE)($RMSE_k$)，可由(4-1)計算得知，

$$RMSE_k = \sqrt{\frac{\sum_{i=1}^N (x_{i,k} - y_k)^2}{N}}, \quad k = 1, 2, \dots, 7, \quad (4-1)$$

其中 $N = 13$ (13 位受測者問卷調查結果); $x_{i,k}$ 代表每位受測者對該影像樣本情緒比例填寫結果; y_k 代表第 k 個情緒類別，由 13 位受測者對於該影像樣本情緒比例判斷結果之間卷調查平均。本實驗對辨識結果與問卷調查結果進行相關性 (Correlation)、均方根誤差 (Root-mean-square-error, RMSE)，與情緒類別組合三個部份之評估，以驗證辨識結果之正確性為目的。



圖 4-4、混合情緒測試樣本(Sample Index : 2)。

表 4-5、混合情緒辨識與問卷調查結果(Sample Index : 2)。

	Neutral	Anger	Disgust	Fear	Happy	Sadness	Surprise
Proposed method	0.00%	35.32%	64.66%	0.00%	0.00%	0.01%	0.00%
Average of survey result	6.92%	28.46%	54.62%	4.23%	0.77%	4.62%	1.92%
RMSE of survey result	11.82%	29.11%	28.76%	7.03%	2.77%	9.46%	4.80%



圖 4-5、混合情緒測試樣本(Sample Index: 3)。

表 4-6、混合情緒辨識與問卷調查結果(Sample Index : 3)。

	Neutral	Anger	Disgust	Fear	Happy	Sadness	Surprise
Proposed method	23.16%	3.83%	4.06%	4.45%	56.30%	4.35%	3.84%
Average of survey result	33.46%	0.00%	0.00%	1.54%	64.23%	0.00%	0.77%
RMSE of survey result	21.15%	0.00%	0.00%	3.76%	21.59%	0.00%	2.77%



圖 4-6、混合情緒測試樣本(Sample Index：4)。

表 4-7、混合情緒辨識與問卷調查結果(Sample Index：4)。

	Neutral	Anger	Disgust	Fear	Happy	Sadness	Surprise
Proposed method	12.27%	1.35%	13.66%	57.25%	6.56%	7.51%	1.4%
Average of survey result	8.46%	13.85%	33.46%	35.00%	0.00%	3.08%	6.15 %
RMSE of survey result	12.14%	22.56%	24.27%	32.02%	0.00%	6.30%	14.46%



圖 4-7、混合情緒測試樣本(Sample Index：7)。

表 4-8、混合情緒辨識與問卷調查結果(Sample Index：7)。

	Neutral	Anger	Disgust	Fear	Happy	Sadness	Surprise
Proposed method	0.00%	0.00 %	0.00 %	0.00%	0.00%	100.00%	0.00%
Average of survey result	2.31%	4.23%	13.85%	7.31%	0.00%	71.92%	0.38%
RMSE of survey result	4.39%	8.13%	15.02%	9.71%	0.00%	16.27%	1.39%

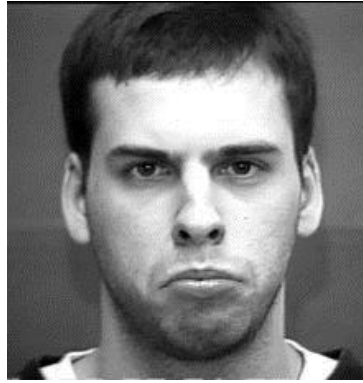


圖 4-8、混合情緒測試樣本(Sample Index: 8)。

表 4-9、混合情緒辨識與問卷調查結果(Sample Index : 8)。

	Neutral	Anger	Disgust	Fear	Happy	Sadness	Surprise
Proposed method	39.72%	28.85%	30.85%	0.00%	0.00%	0.56%	0.00%
Average of survey result	37.69%	17.31%	21.92%	0.00%	0.00%	21.92%	1.92%
RMSE of survey result	22.70%	15.89%	16.01%	0.00%	0.00%	32.76%	6.93%

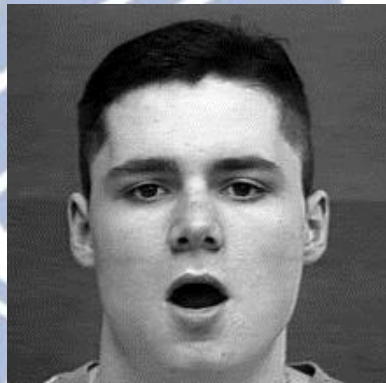


圖 4-9、混合情緒測試樣本(Sample Index : 10)。

表 4-10、混合情緒辨識與問卷調查結果(Sample Index : 10)。

	Neutral	Anger	Disgust	Fear	Happy	Sadness	Surprise
Proposed method	0.00%	0.00%	0.01%	1.57%	0.00%	0.00%	98.41%
Average of survey result	39.23%	0.77%	1.15%	4.23%	0.00%	2.69%	52.31%
RMSE of survey result	28.42%	2.77%	3.0%	9.09%	0.00%	6.65%	28.62%

4.3.1. 問卷調查結果與辨識結果相關性評估

本實驗以統計學 Correlation 定義，以(4-2)評估問卷調查個別情緒比例與辨識結果之相關性，

$$\begin{aligned} \text{Corr}_k(X^P, X^Q) &= \frac{\text{cov}(X^P, X^Q)}{\sigma_{X^P} \sigma_{X^Q}} \\ &= \frac{\sum_{i=1}^N (x_{i,k}^P - y_k^P)(x_{i,k}^Q - y_k^Q)}{\sqrt{\sum_{i=1}^N (x_{i,k}^P - y_k^P)^2 \sum_{i=1}^N (x_{i,k}^Q - y_k^Q)^2}}, k = 1, 2, \dots, 7, \end{aligned} \quad (4-2)$$

其中 $N = 10$ (10 組序列影像)； X^P 代表 Likelihood 辨識結果， $x_{i,k}^P$ 代表 10 組序列影像第 k 個情緒類別之 Likelihood， y_k^P 代表 10 組序列影像第 k 個情緒類別之 Likelihood 平均； X^Q 代表問卷調查結果， $x_{i,k}^Q$ 代表 10 組序列影像第 k 個情緒類別之比例， y_k^Q 代表 10 組序列影像第 k 個情緒類別之比例平均； $\text{Corr}_k(X^P, X^Q)$ 值介於 $-1 \sim 1$ 之間，由表 4-11 結果可看出，所有情緒類別之平均相關性(Correlation)皆大於 0.5，而其中 6 種基本情緒之平均相關性可達 0.8 以上，高興情緒之平均相關性達到 0.98，因此整體而言，情緒類別 Likelihood 辨識結果與問卷調查之平均結果有明顯的正相關。

4.3.2. 整體情緒類別 RMSE 評估

本實驗以整體情緒類別 RMSE 分析問卷調查結果與辨識結果，整體情緒比例差異之合理性。圖 4-10 為 10 組序列辨識結果之 RMSE(黑直方)，以及問卷調查中每位受測者判定其中一張端點影像情緒類別整體 RMSE 之平均(灰直方)，RMSE 公式如(4-3)所示，

表 4-11、情緒 Likelihood 辨識結果與問卷調查平均相關性。

	Neutral	Anger	Disgust	Fear	Happy	Sadness	Surprise
Correlation	0.53	0.84	0.86	0.95	0.98	0.95	0.97

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (x_i - y_i)^2}{N}}, \quad (4-3)$$

其中 N 代表資料樣本數； x_i 代表資料樣本； y_i 代表資料樣本平均。辨識結果 $RMSE(RMSE_P)$ 計算時， $N = 7$ (7類基本情緒)； x_i 代表辨識系統辨識之各種情緒 Likelihood； y_i 代表各種情緒問卷調查平均。受測者 $RMSE(RMSE_Q)$ 計算時， $N = 7$ ； x_i 代表受測者判斷各種情緒之情緒比例； y_i 代表各種情緒問卷調查平均，所有受測者 $RMSE_Q$ 加總取平均即為圖 4-10 灰直方之值。由圖 4-10 可看出，調查結果平均情緒比例與辨識結果之人臉情緒 Likelihood 大致吻合，整體而言二者關係一致， $RMSE$ 介於 5% ~ 20% 之間。其中， $RMSE_P$ 於樣本編號 1 與編號 10 有 20% 之誤差，主要原因在於，13 位受測者或系統本身對於人臉表情認知差異所造成，可由問卷調查結果 $RMSE_k$ 看出，大多數情況 $RMSE_k$ 相當大。例如：圖 4-9 與表 4-10 中，問卷調查結果平均以中性情緒與驚訝情緒之比例較高，而辨識系統辨識之 Likelihood 則以驚訝情緒最高，如此之差異，導致系統辨識 Likelihood 之整體

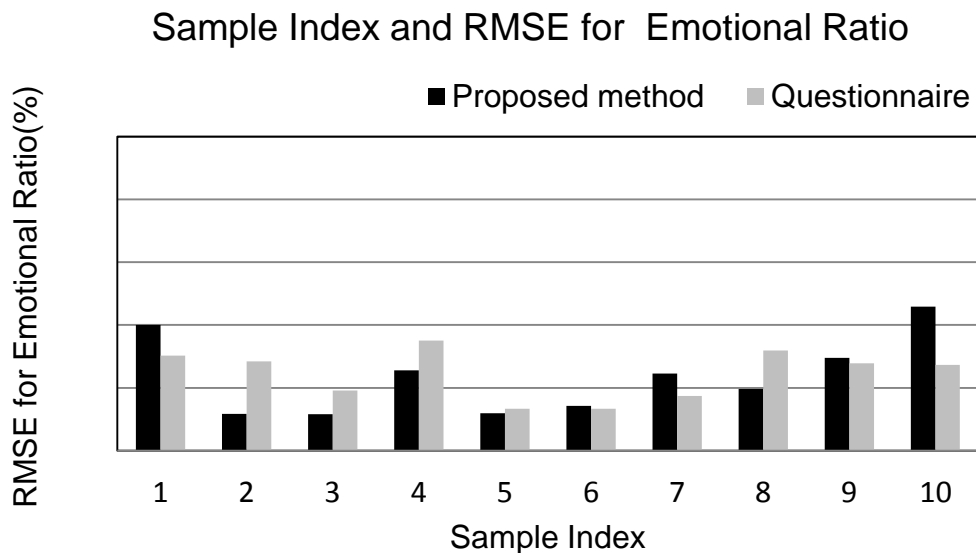


圖 4-10、10 組序列人臉情緒比例調查結果整體均方根誤差(灰直方)，與辨識結果之均方根誤差(黑直方)。

RMSE_P偏高。另外，若與RMSE_Q相比，由圖 4-10 可看出，RMSE_P與RMSE_Q差距皆不到 5%，除了編號 10 有較大之差距，但其仍在 10% 以內，本實驗以此評估判定辨識結果之誤差在合理範圍內，系統辨識結果與每位受測者判斷影像樣本情緒之認知差異一致。

4.3.3. 混合情緒類別組合評估

本實驗另外以情緒類別組合做分析，評估問卷調查結果之情緒類別組合，與辨識結果之情緒類別組合之一致性。在此以(4-4)判斷影像樣本之情緒組合，

$$SR = \sum_i^t \text{Ratio}_i, \quad t < 7, \quad (4-4)$$

其中 SR 代表問卷調查平均人臉情緒比例之和。當該影像前 t 大問卷調查平均人臉情緒比例之和 SR 大於 70%，則影像樣本呈現之情緒，視為此 t 個情緒之組合，通常為 1~3 類情緒組合，例如：圖 4-6 與表 4-7 所示，該影像樣本生氣、厭惡、害怕三者之情緒比例和為 82.31%，故該人臉情緒視為此三者之組合。所有影像統計結果如圖 4-11 所示，圖 4-11 灰直方為人臉影像經問卷調查判定之主要情緒組合比例之和，黑直方為由辨識系統對該影像情緒組合之 Likelihood 加總結果，同樣可以看出辨識結果與問卷調查結果呈現正相關性，辨識結果情緒組合 Likelihood 之和，遠大於其餘 Likelihood 之和。

4.4. 基於 A-V Plane 之情緒辨識結果

本研究藉由情緒之 Likelihood 建立 Regression model，將情緒轉換至 A-V Plane，進行情緒 A-V 辨識。實驗同樣從 CK+ 資料庫，每種情緒各取 10 組序列影像，共 60 組序列影像，做 10 次之交叉驗證(10-fold cross-validation)，每次每

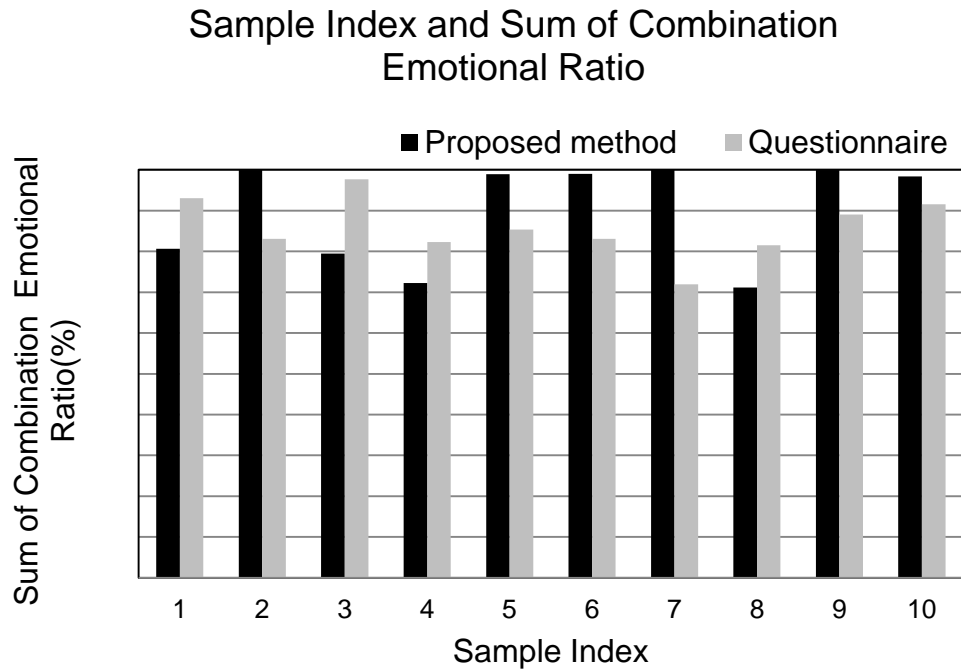


圖 4-11、10 組序列人臉情緒比例調查結果之情緒組合比例之和(灰直方)，辨識結果對調查結果之情緒組合 Likelihood 之和(黑直方)。

種情緒各取 9 組序列影像進行 Regression model 訓練，剩下的 1 組則做為測試 (Leave-one-subject out)。實驗對本研究所提出之情緒辨識方法取得 Likelihood 後辨識之 A-V，與 Likelihood 非時序辨識(直接使用本論文 3.3.1a 節之 RVM 分類器進行 Likelihood 辨識)方法取得 Likelihood 後辨識之 A-V，進行二者辨識結果之驗證比較。

圖 4-12 為其中 6 組利用 Regression model 辨識序列影像 A-V 之結果，分別由中性情緒辨識至 6 種基本情緒。由結果可以看出，使用本研究提出之情緒辨識方法，有較合理之情緒轉換過程。由於每組序列影像皆由連續之影像樣本所構成，轉換過程不應有不合理之瞬間轉換產生，如圖 4-12(c)所示。另外，資料庫之序列影像皆由中性情緒直接轉換至特定情緒，除了端點必須收斂至特定情緒，轉換過程亦應朝特定情緒進行轉換，而不會經過其他情緒類別，如圖 4-12(a)所示。

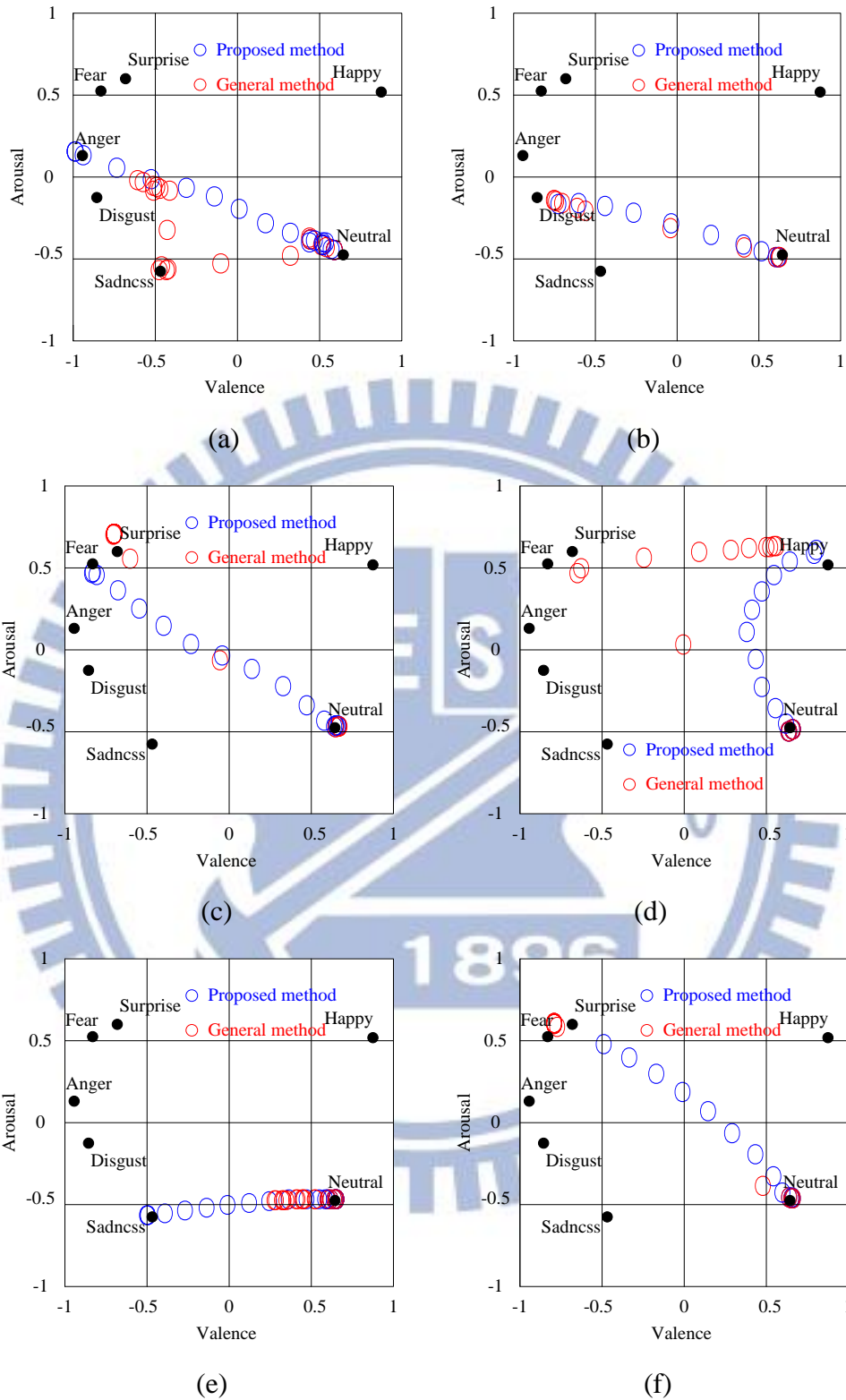


圖 4-12、情緒從中性情緒轉換至 (a)生氣，(b)厭惡，(c)害怕，(d)高興，(e)傷心，(f)驚訝。藍圈代表本研究提出之 Likelihood 時序情緒辨識方法之辨識結果，紅圈為 Likelihood 非時序辨識方法(一般情緒辨識方法)之辨識結果。

本研究藉由(4-5)~(4-6)計算影像樣本間之變化量，

$$\text{Variation}_A = \frac{A_{\text{present}} - A_{\text{previous}}}{T}, \quad (4-5)$$

$$\text{Variation}_V = \frac{V_{\text{present}} - V_{\text{previous}}}{T}, \quad (4-6)$$

其中 A 代表 Arousal；V 代表 Valence；T 代表取樣時間(T = 1/30 sec)。藉由此二式分析轉換過程中是否有不合理之瞬間變化產生，利用(4-7)將此結果以標準差(Standard deviation) σ ，針對 60 組序列影像由本研究提出之 Likelihood 時序辨識，與 Likelihood 非時序辨識轉換 Likelihood 至 A-V 之結果進行評估。

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - y_i)^2}{N}}, \quad (4-7)$$

其中 N 代表某個序列影像之 A-V 變化次數； x_i 代表影像樣本間變化之 A-V 量； y_i 代表影像樣本間變化 A-V 量之平均。由本研究提出情緒辨識之情緒轉換分析結果如表 4-12 所示，由 Likelihood 非時序辨識之情緒轉換分析結果如表 4-13 所示。由二表可看出無論是 Arousal 或 Valence 由本研究提出情緒辨識方法，情緒轉換過程中之瞬間變化程度較小，其標準差平均小於 Likelihood 非時序辨識方

表 4-12、時序辨識方法之 A-V 轉換標準差。

	Anger	Disgust	Fear	Happy	Sadness	Surprise	Average
Arousal	0.81	0.63	1.79	2.01	0.36	1.61	1.20
Valence	2.07	2.14	1.81	1.40	1.58	1.91	1.82

表 4-13、非時序辨識方法之 A-V 轉換標準差。

	Anger	Disgust	Fear	Happy	Sadness	Surprise	Average
Arousal	0.97	0.99	3.51	3.27	0.64	4.75	2.35
Valence	2.54	3.87	4.26	4.56	2.61	6.30	4.02

法，6種基本情緒轉換個別之結果亦是如此。另外，情緒轉換結果收斂至特定情緒之情況，可由本論文 4.3 節之實驗結果做說明。因 Regression model 是依據情緒 Likelihood 所建立，故其收斂情況亦會與 Likelihood 之端點影像基本情緒辨識分類結果相符，如表 4-2 所示。

4.5. 基於情緒辨識之音樂選取實驗

本研究藉由 4.4 節之實驗，驗證情緒 Likelihood 轉換至 A-V 之情緒辨識效能，而本研究另外設計智慧型音樂選取系統，驗證 A-V 辨識於人機互動之實用性。智慧型音樂選取完整音樂選取系統，利用 AAM 演算法建立人臉模型，偵測人臉特徵點(本論文第二章)，擷取人臉情緒辨識之幾何特徵(本論文 3.1 ~ 3.2 節)，並藉由 RVM 演算法對情緒進行情緒 Likelihood 時序辨識(本論文 3.3 節)，最後再將使用者之情緒 Likelihood 轉換為 A-V，以計算相對應之音樂，設法轉換使用者情緒。本節介紹本實驗選擇之音樂資料庫，與音樂選取設計策略，最後以兩組不同情緒轉換目標之設定，驗證音樂選取與情緒 A-V 轉換分析辨識之效能。

4.5.1. 音樂資料庫

Yang 等人的研究中[49]，挑選 60 首著名之英文歌曲，進行音樂影響使用者情緒相關研究。作者以手動方式將 60 首音樂修減為 25 秒之長度，每首歌皆由 40 人標註 Arousal 與 Valence 至 A-V Plane 上，其值介於[-1, 1]之間，資料庫之音樂分佈如圖 4-13 所示。本研究則以此做為音樂資料庫，進行系統之設計。

4.5.2. 音樂選取設計

音樂選取系統設定情緒轉換之目標(V_T, A_T)，以及轉換過程預備播放之所有歌曲數目 N_T ，利用辨識取得之情緒 A-V，經由圖 4-14 之智慧音樂選取之系統流

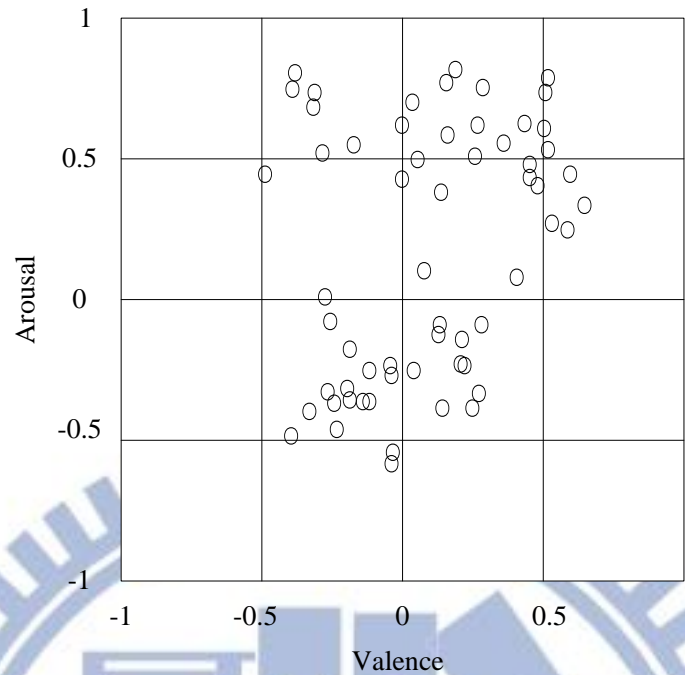


圖 4-13、音樂資料庫之音樂分佈示意圖。

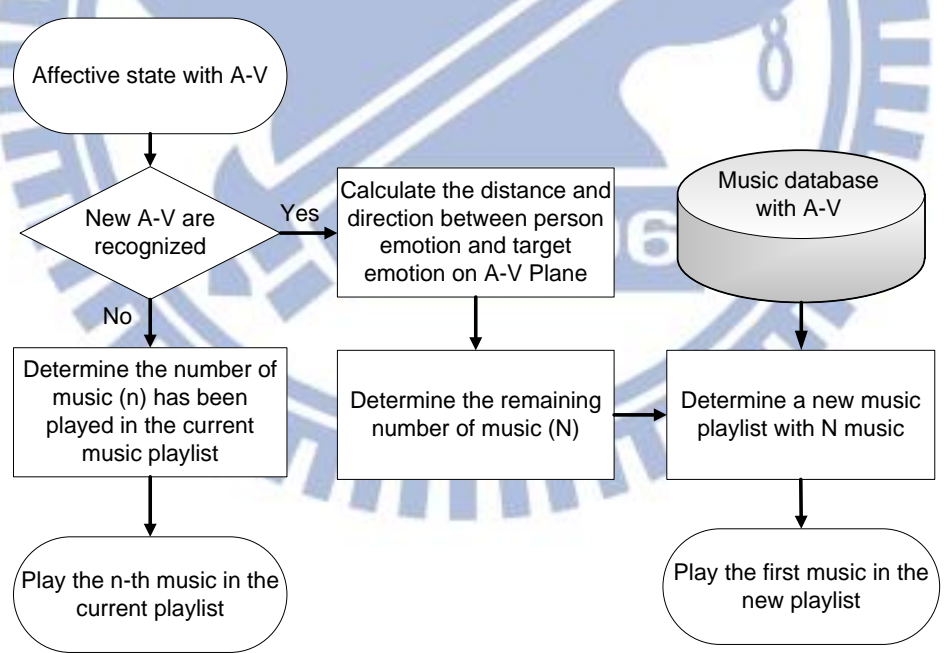


圖 4-14、智慧型音樂選取之系統流程。

程，選擇適當之音樂進行播放。每當辨識到新的 A-V 時，其可能為使用者情緒有所轉變，或首度取得使用者情緒，則根據系統設定之情緒轉換之目標(V_T, A_T)，

以及轉換過程預備播放之所有歌曲數目 N_T ，計算情緒轉換目標與使用者情緒在 A-V Plane 上距離與角度關係，並決定 N 首期望播放之歌曲位置，如 (4-8) 所示，

$$A_E^n = A_S + \frac{d_{TS}}{N} \times \sin(\theta_{TS}) \times n,$$

$$\text{and } V_E^n = V_S + \frac{d_{TS}}{N} \times \cos(\theta_{TS}) \times n, \quad n = 1, 2, \dots, N,$$

$$-\pi \leq \theta_{TS} \leq \pi, \quad (4-8)$$

其中 d_{TS} 代表情緒轉換目標 A-V (V_T, A_T) 與使用者情緒 A-V (V_S, A_S) 之距離； θ_{TS} 代表以 Valence 軸為基準，Valence 軸與 (V_T, A_T), (V_S, A_S) 連線 ($\overline{P_T P_S}$) 之夾角；(V_E^n, A_E^n) 代表期望播放之音樂位置； N 代表目前使用者情緒至情緒轉換目標間，預備播放之歌曲數目； n 代表預備播放之 N 首音樂中第 n 首音樂。藉由 (4-8) 取得期望播放之歌曲位置周圍之音樂，選擇最接近之音樂做播放，取得 N 首歌之音樂播放清單，如 (4-9) 所示，

$$(V_M, A_M) = \arg \min_{V_M, A_M} (\sqrt{(V_E^n - V_M)^2 + (A_E^n - A_M)^2} \times w),$$

$$w = \frac{1}{\cos(\theta_{EM}) + 2}, \quad -\pi \leq \theta_{EM} \leq \pi, \quad (4-9)$$

其中 (V_M, A_M) 代表音樂資料庫之音樂位置； w 代表權重值， θ_{EM} 代表以 (V_T, A_T), (V_E^n, A_E^n), 連線 ($\overline{P_T P_E}$) 為基準， $\overline{P_T P_E}$ 與 (V_M, A_M), (V_E^n, A_E^n) 連線 ($\overline{P_M P_E}$) 之夾角，此權重值主要目的在於，選擇接近預備播放歌曲位置與情緒轉換目標位置之歌曲做播放，然而，假若選擇之音樂已播放過，則選擇第二近的音樂做播放，以此類推。(4-8) ~ (4-9) 之參數位置關係如圖 4-15 所示。使用者情緒至情緒轉換目標間預備播放之歌曲數目 N ，根據系統之設定 N_T ，以及已播放之歌曲數目做決定，其初始為 $N = N_T$ ，隨著情緒之轉換與已播放之歌曲數目而減少。每當情緒有所轉變，重新決定期望播放之歌曲位置 (V_E^n, A_E^n) 與數目 N 時，此次期望播放歌曲數目 N ，皆會扣除前一個音樂播放清單已播放之歌曲數目 n 決定之，即 $N = N - n$ 。例如：圖 4-16 顯示目前辨識到使用者情緒之 A-V 位置 (V_S^1, A_S^1) 與轉換情緒目標 (V_T, A_T) 計算距離 d_{TS}^1 ，決定預期播放之 5 首音樂位置 ($N = 5$)；當播放完第 1 首音樂 ($n = 1$)，

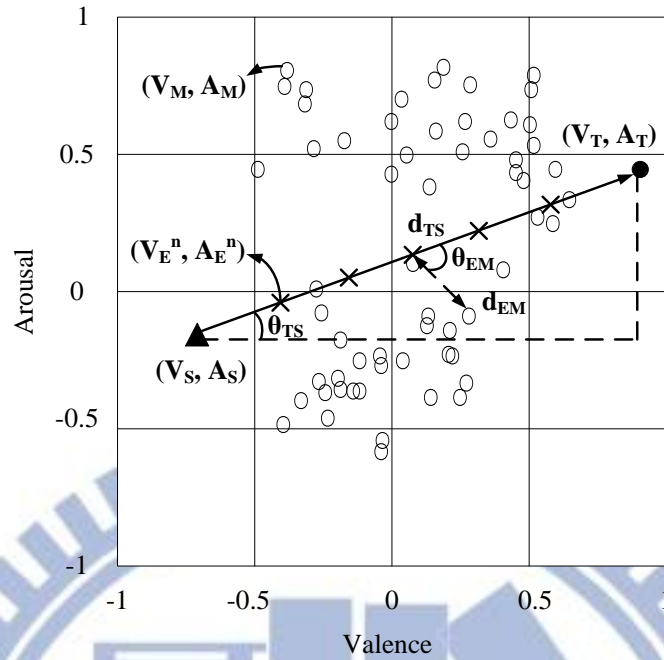


圖 4-15、情緒與音樂關係示意圖。此圖在情緒轉換目標(黑圓)與使用者情緒(黑三角)，共設定 5 首期望音樂位置(叉叉)進行情緒轉換，並選取最接近期望音樂位置之音樂(白圓)做播放。

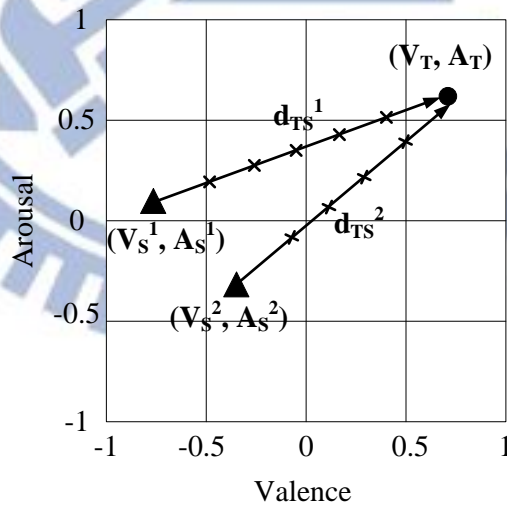


圖 4-16、情緒與音樂之決定與轉換。轉換情緒目標位於 (V_T, A_T) ，以及期望播放之音樂位置以叉叉表示之；目前辨識到使用者情緒之 A-V 位置 (V_S^1, A_S^1) 與轉換情緒目標 (V_T, A_T) 之距離為 d_{TS}^1 ；下次辨識到使用者情緒之 A-V 位置 (V_S^2, A_S^2) 與轉換情緒目標 (V_T, A_T) 之距離為 d_{TS}^2 。

且辨識到使用者情緒位置轉換至 (V_S^2, A_S^2) ，則重新計算使用者情緒與轉換情緒目標 (V_T, A_T) 之距離 d_{TS}^2 ，以及決定預期播放之4首音樂位置 $(N = 5 - 1)$ 。然而，假若播放完一首歌曲，使用者情緒仍未轉換，則不需重新改變預期播放之音樂位置 (V_E^0, V_E^0) 及數目 N ，只需以(4-8) ~ (4-9)選擇程度較強之下一首歌曲做播放，即 $n = n + 1$ 。

當播放完系統設定預備播放之所有音樂數目 $(N = 0)$ ，或播放完當前使用者情緒與轉換情緒目標間預備播放之音樂數目 $(N = n)$ ，亦或是使用者情緒相當接近設定之情緒轉換目標 $(d_{TS} \leq 0.2)$ 時，則系統停止選擇音樂做播放。

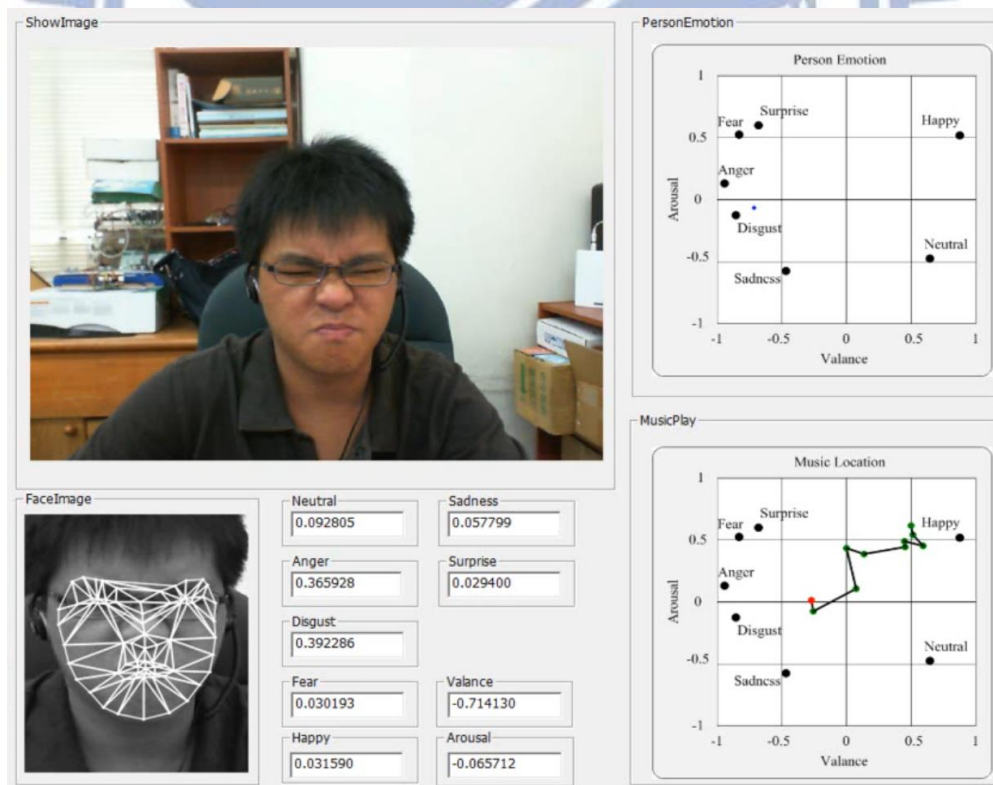
4.5.3. 音樂選取實驗結果

此實驗之目的在於驗證本論文設計之人臉情緒辨識與音樂選曲系統之效能，經由設定不同之音樂轉換目標，驗證情緒轉換過程與選曲之音樂是否合宜，以下包含兩種實驗之驗證，情緒轉換目標分別設定為高興情緒 $(A : 0.63, V : 0.85)$ ，以及中性情緒 $(A : -0.47, V : 0.61)$ 之位置。

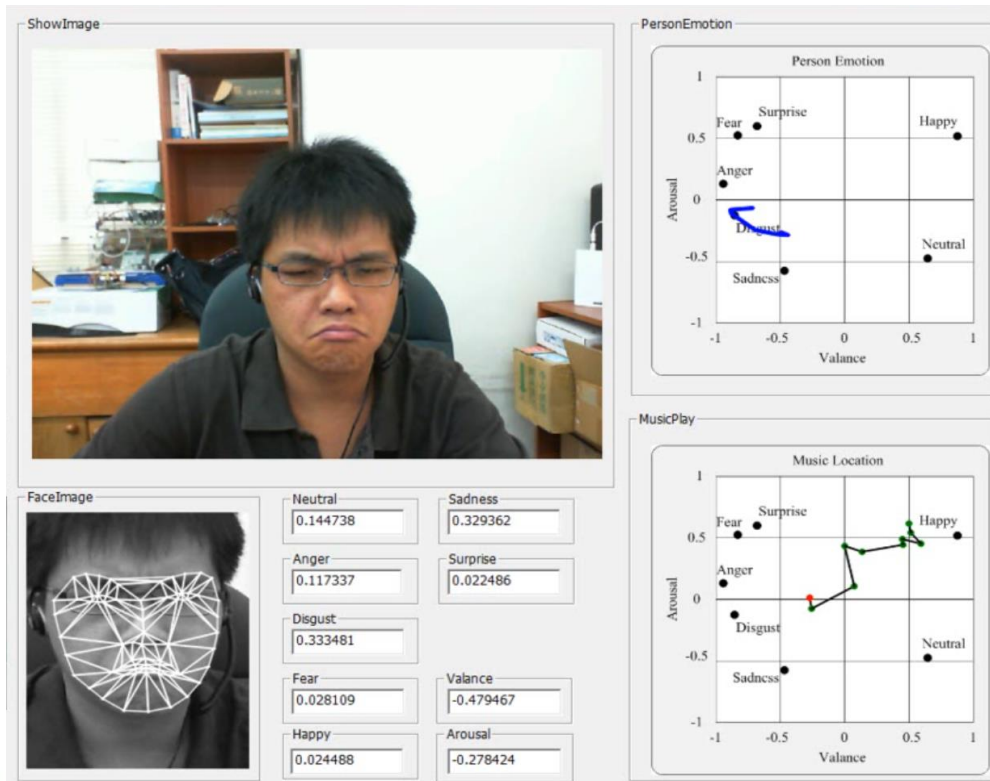
a. 情緒轉換目標 $(A : 0.63, V : 0.85)$

此實驗中，系統設定之情緒轉換目標為 $(A : 0.63, V : 0.85)$ 即高興情緒，使用者情緒與情緒轉換目標之播放音樂數目設定為10首，主要目的在於希望將使用者情緒由不愉快轉為愉快。系統一開始辨識使用者情緒為 $(A : -0.07, V : -0.71)$ 即既生氣又厭惡之情緒，如圖4-17(a)右上A-V Plane藍色軌跡所示，系統藉由計算使用者目前情緒與情緒轉換目標位置之距離，播放第1首音樂位置為 $(A : 0.02, V : -0.27)$ ，並取得預備播放之9首音樂位置，分別如圖4-17(a)右下A-V Plane紅點與綠點所示；透過第1首音樂之播放，使用者情緒逐漸往Valence(愉快程度)為正向的情緒做轉換，如圖4-17(b)所示；系統經過一段時間之情緒辨識，使用者A-V收斂至 $(A : -0.54, V : -0.45)$ 即傷心情緒，第1首音樂播放結束，系統重新計算使

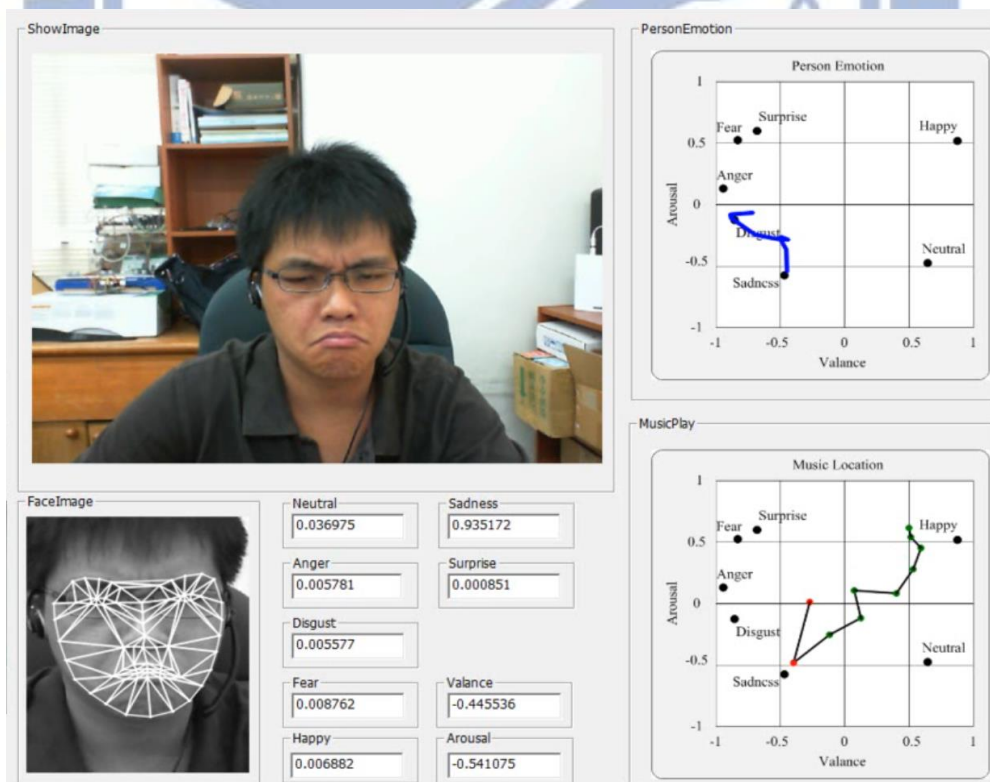
用者情緒與情緒轉換目標之距離，播放第 2 首音樂(A：-0.48, V：-0.40)，並取得其餘預備播放之 8 首音樂位置，如圖 4-17(c)所示；透過第 2 首音樂之播放，使用者情緒依舊往 Valence 正向轉換，如圖 4-17(d)所示；第 2 首音樂播放結束時，使用者情緒轉換至(A：-0.35, V：0.17)，此時中性情緒占大多數之比例，Valence 已轉為正向，系統計算取得第 3 首音樂播放位置(A：-0.38, V：0.14)，以及預備播放之 7 首音樂位置，如圖 4-17(e)所示；第 4 首音樂透過轉換至(A：-0.46, V：0.62)之使用者情緒即中性情緒，計算取得其播放位置(A：-0.33, V：0.27)與預備播放之 6 首音樂，如圖 4-17(f)所示；透過第 4 首音樂之播放，使用者情緒逐漸轉變收斂至轉換目標之高興情緒，如圖 4-17(g)所示；最終使用者情緒轉換至接近目標之情緒位置(A:0.57, V:0.82) (使用者情緒與情緒轉換目標之距離 $d_{TS} \leq 0.2$)，計算位於目標情緒附近之音樂(A：0.45, V：0.59)進行最後一首歌曲之播放，如圖 4-17(h)所示。



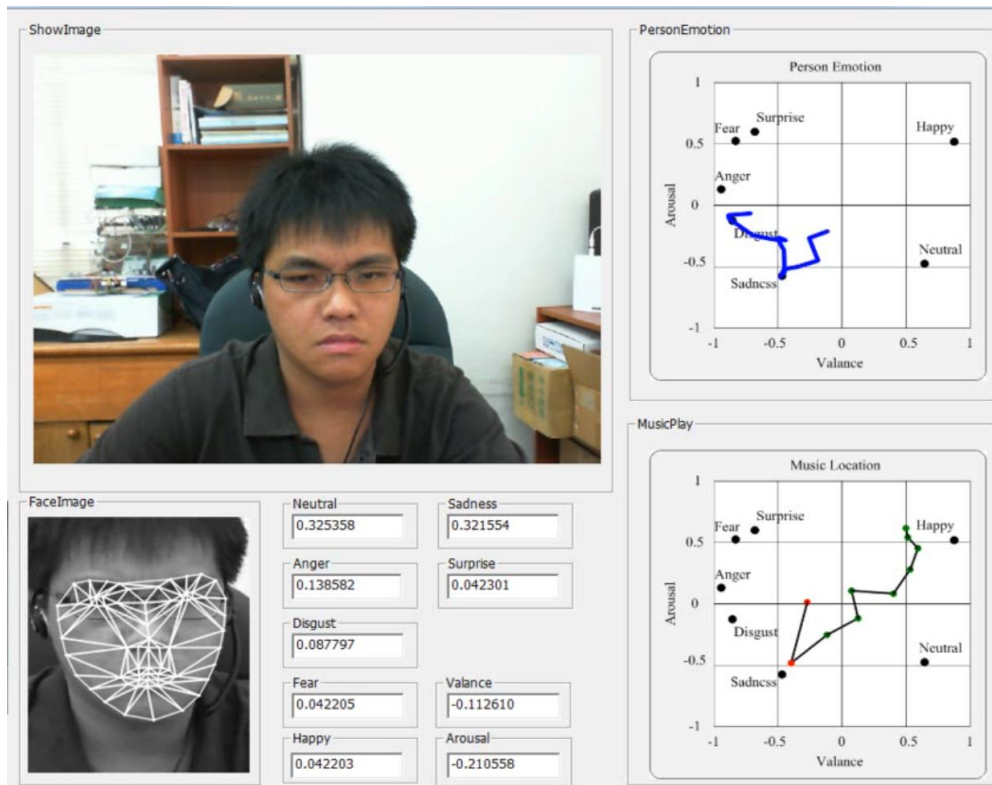
(a)



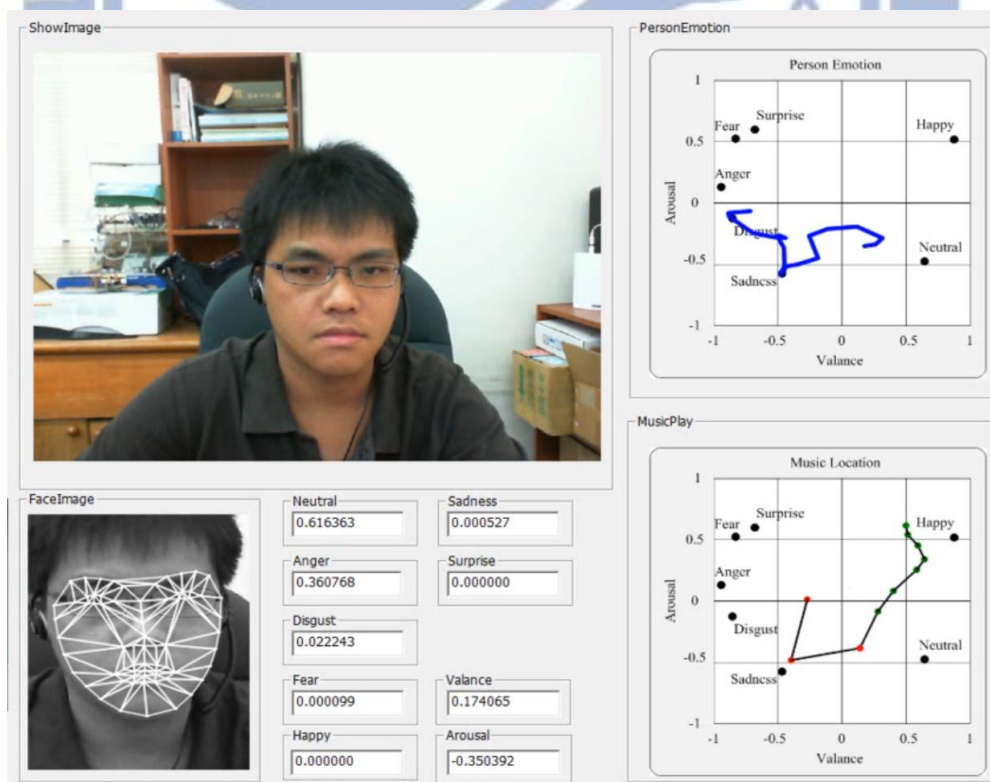
(b)



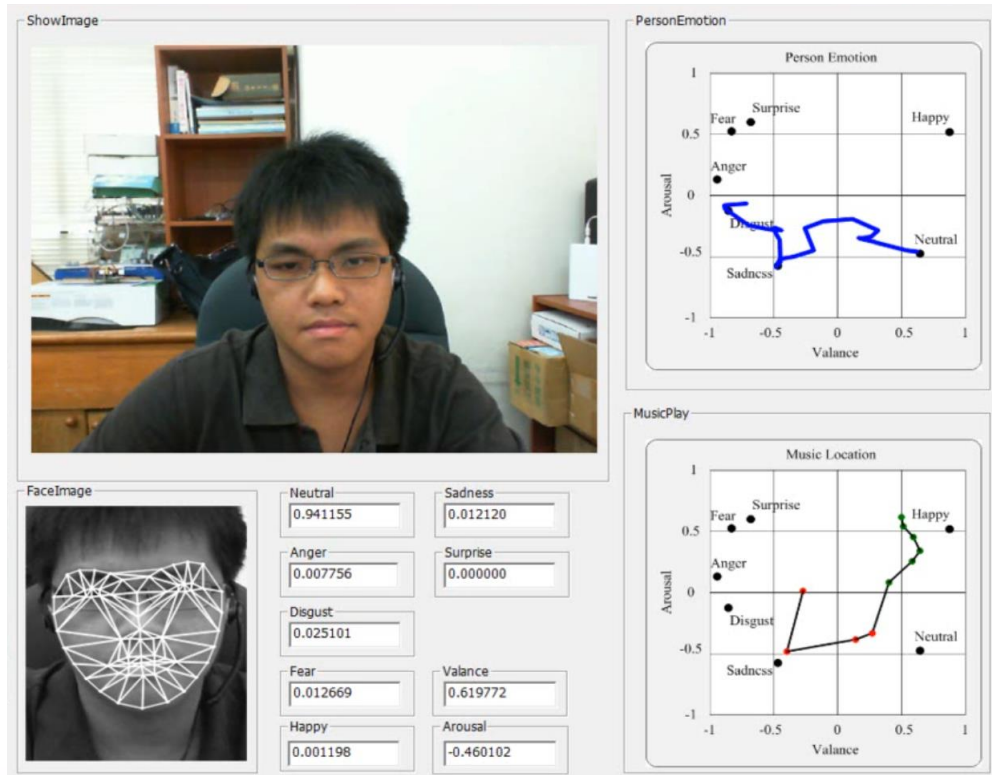
(c)



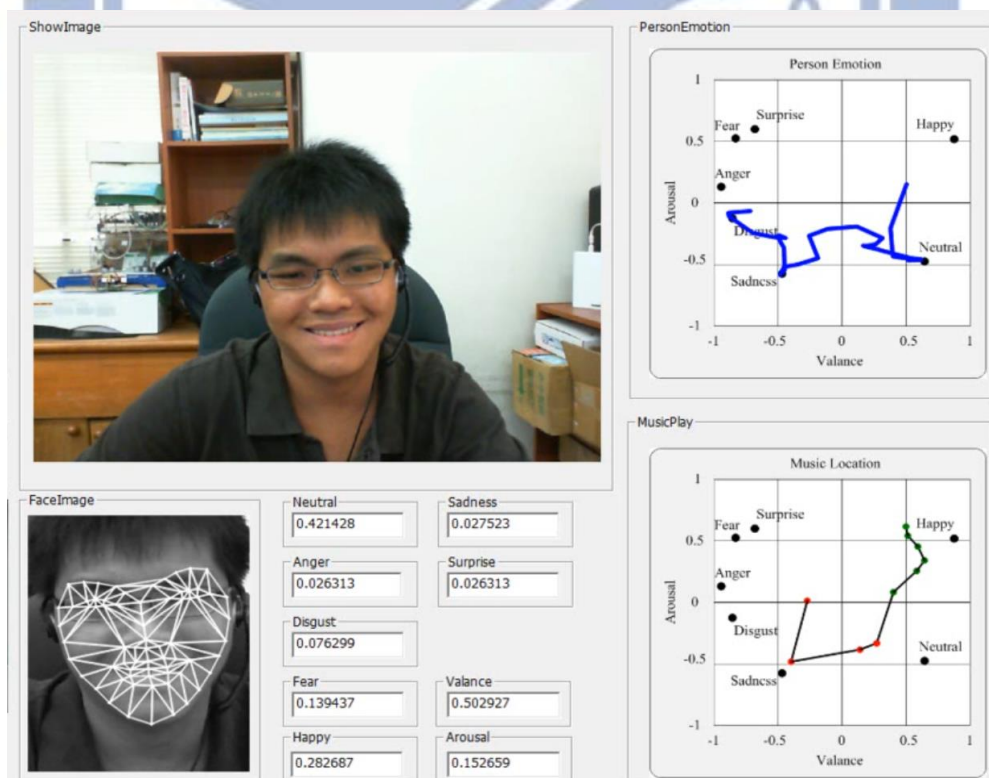
(d)



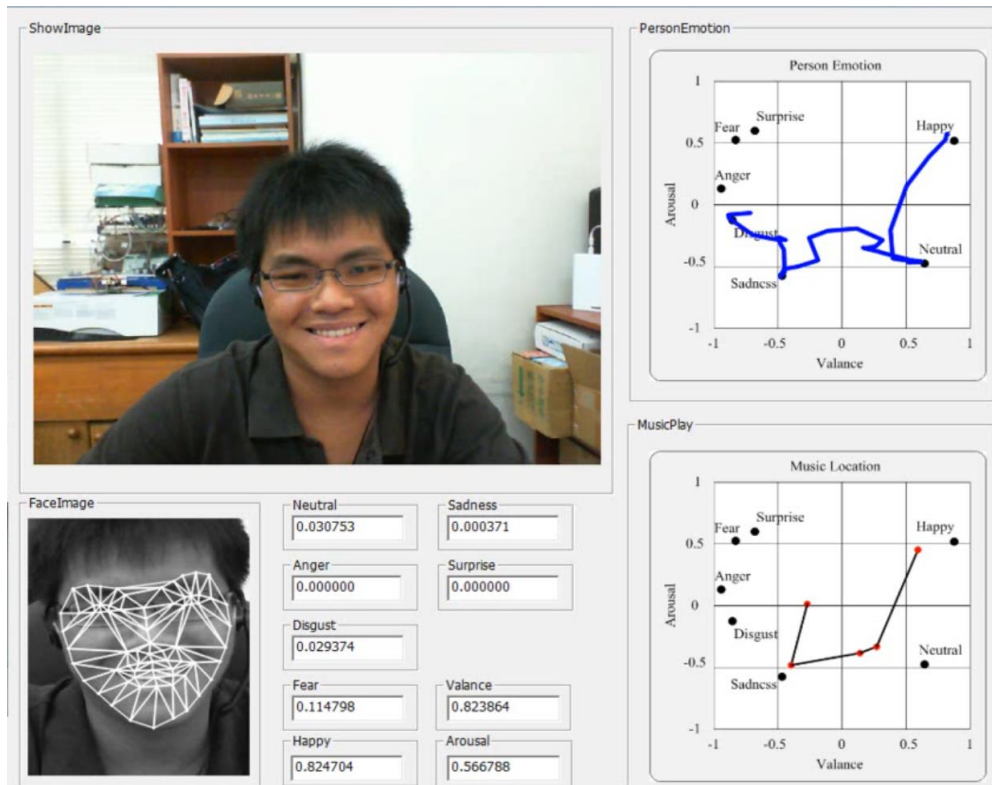
(e)



(f)



(g)

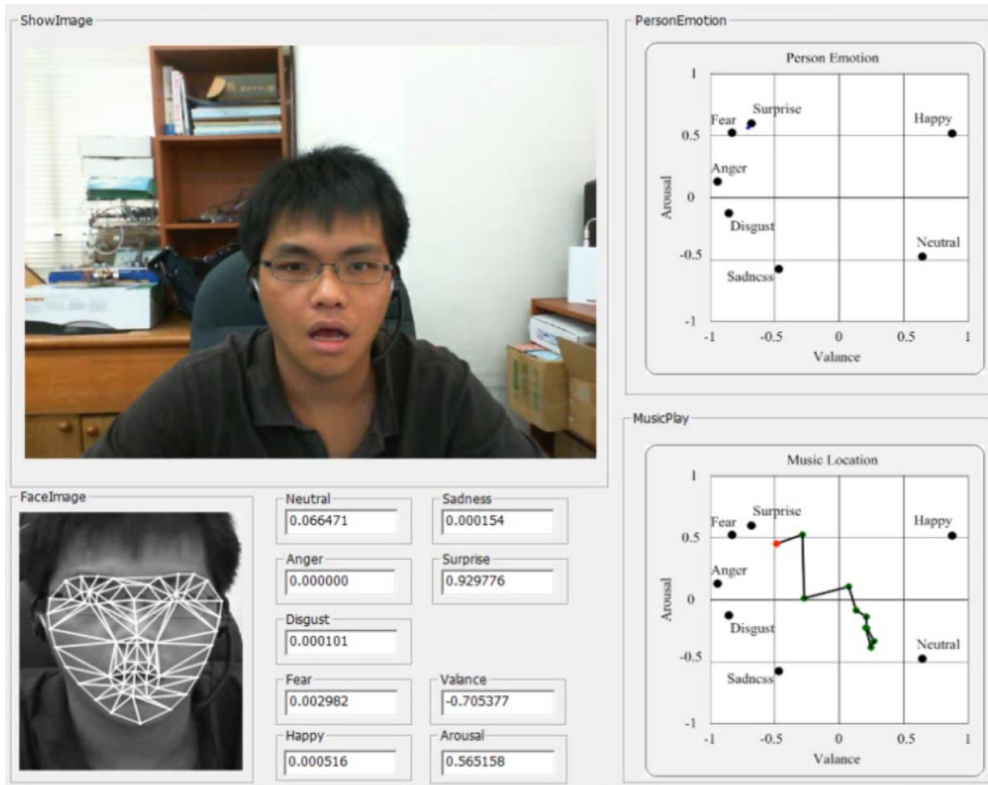


(h)

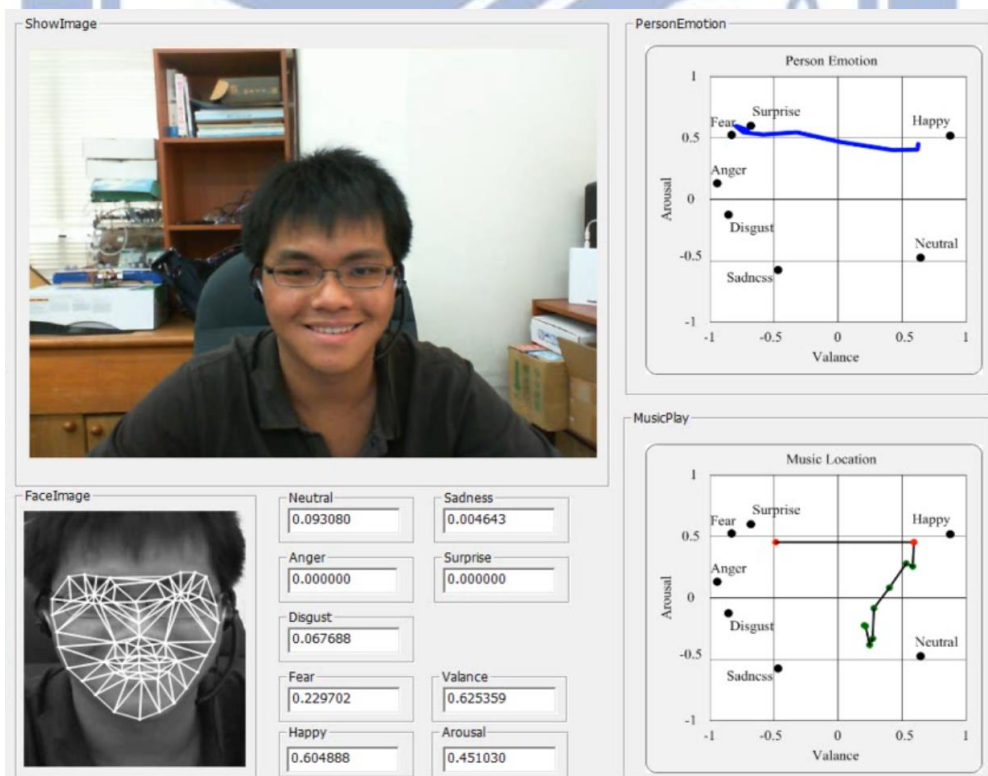
圖 4-17、情緒轉換情境設計，由不愉快情緒轉為愉快情緒之實驗，其中右上 A-V Plane 代表使用者情緒 A-V 之軌跡，右下 A-V Plane 紅點代表已播放或正在播放之音樂，綠點為預備播放之音樂。(a)使用者情緒(A: -0.07, V: -0.71)，第 1 首音樂播放位置(A: 0.02, V: -0.27)，(b)使用者情緒往 Valence 為正方向做轉換，(c)第 1 首音樂播放結束，使用者情緒(A: -0.54, V: -0.45)，播放第 2 首音樂(A: -0.48, V: -0.40)，(d)使用者情緒持續往 Valence 為正方向做轉換，(e)使用者情緒(A: -0.35, V: 0.17)，播放第 3 首音樂(A: -0.38, V: 0.14)，(f)使用者情緒轉換至(A: -0.46, V: 0.62)，播放第 4 首音樂(A: -0.33, V: 0.27)，(g)使用者情緒逐漸轉變收斂至轉換目標之高興情緒，(h)最終使用者情緒轉換至(A: 0.57, V: 0.82)，播放最後一首音樂(A: 0.45, V: 0.59)。

b. 情緒轉換目標 (A : -0.47, V : 0.61)

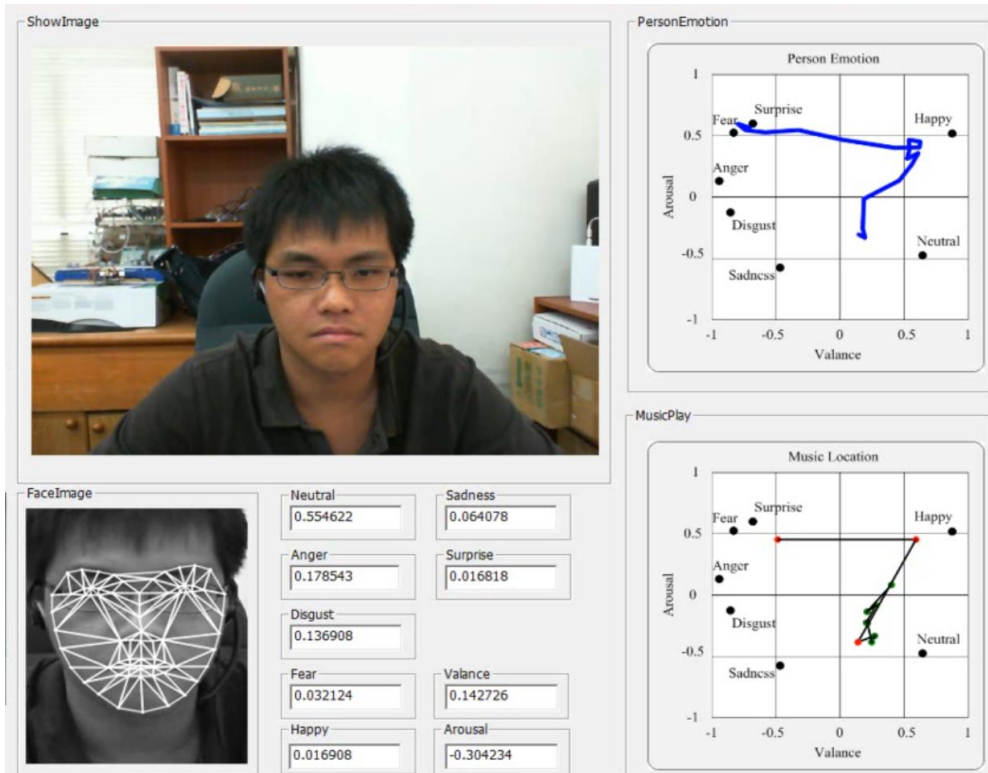
此實驗中，系統設定之情緒轉換目標為(A : -0.47, V : 0.61)即中性情緒，使用者情緒與情緒轉換目標之播放音樂數目設定為 10 首，主要目的在於希望將使用者情緒由興奮轉為平靜。系統一開始辨識到使用者情緒為(A : 0.57, V : -0.71)即驚訝情緒，如圖 4-18(a)右上 A-V Plane 藍色軌跡所示，以此計算與情緒轉換目標之距離，播放第 1 首音樂(A : 0.45, V : -0.49)，並取得預備播放之 9 首音樂位置，分別如圖 4-18(a)右下 A-V Plane 紅點與綠點所示；第 1 首音樂播放結束時，使用者情緒轉換至(A : 0.45, V : 0.63)高興情緒占大多數之比例，計算取得第 2 首音樂播放之位置為(A : 0.45, V : 0.59) 與其餘 8 首音樂之位置，如圖 4-18(b)所示；經由 2 首不同音樂之播放，使用者情緒轉換至(A : -0.30, V : 0.14)，已逐漸改變為中性情緒，系統計算取得播放之第 3 首音樂位置(A : -0.38, V : 0.25)，以及預備播放 7 首音樂之位置，由圖 4-18(c)可看出由於音樂資料庫在中性情緒附近的音樂較為稀疏，如圖 4-13 所示，故在選擇音樂時會有集中於某一區域跳動的情況產生；最終使用者情緒逐漸轉換至目標情緒，使用者之 Arousal(興奮程度)在轉換過程中逐漸下降，回復至平靜之情緒(A : -0.44, V : 0.56) (使用者情緒與情緒轉換目標之距離 $d_{TS} \leq 0.2$)，計算取得最後一首音樂播放之位置，為情緒轉換目標附近之音樂(A : -0.33, V : 0.27) ，如圖 4-18(d)所示。



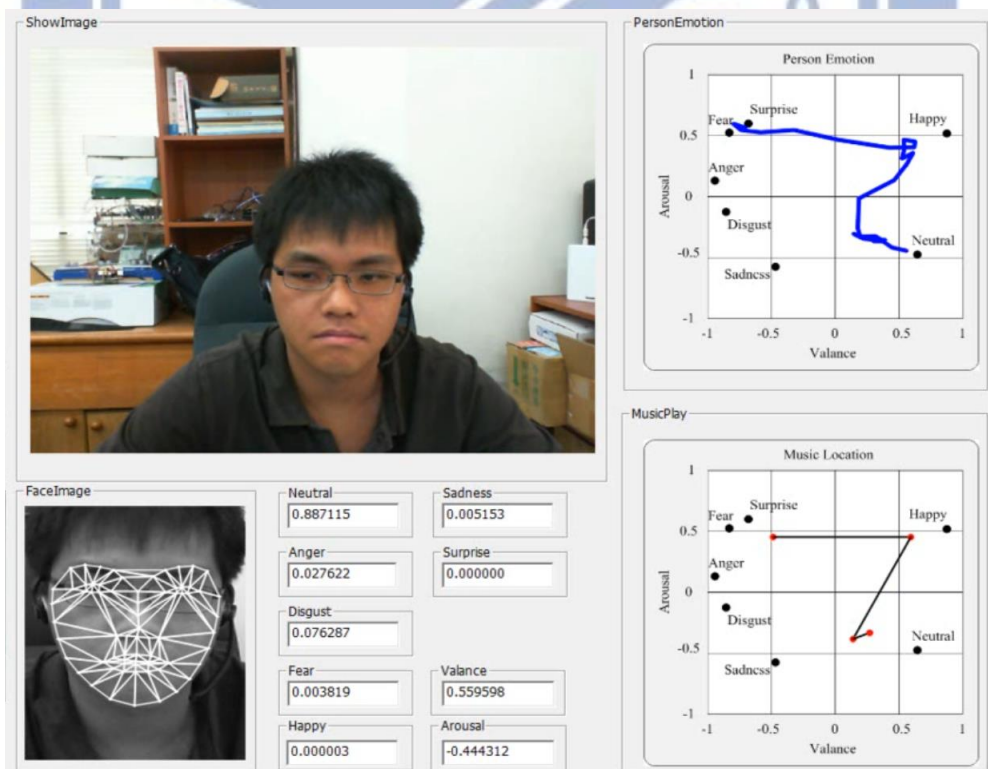
(a)



(b)



(c)



(d)

圖 4-18、情緒轉換情境設計，由興奮情緒轉為平靜情緒之實驗，其中右上 A-V Plane 代表使用者情緒 A-V 之軌跡，右下 A-V Plane 紅點代表已播放或正在播放之音樂，綠點為預備播放之音樂。(a)使用者情緒(A：0.57, V：-0.71)，第 1 首音樂播放位置(A：0.45, V：-0.49)，(b)第 1 首音樂播放結束，使用者情緒(A：0.45, V：0.63)，播放第 2 首音樂(A：0.45, V：0.59)，(c)使用者情緒(A：-0.30, V：0.14)，播放第 3 首音樂(A：-0.38, V：0.25)，(e)最終使用者情緒轉換至接近目標情緒之位置(A：-0.44, V：0.56)，播放最後一首音樂(A：-0.33, V：0.27)。



第五章、結論與未來展望

5.1. 結論

本論文提出一套以時序分析強化情緒辨識之方法，辨識使用者情緒可能性 (Likelihood)，並將其結果映射至 A-V Plane，對人臉情緒做更細微之分析，並透過連續影像之辨識，對情緒進行時序之分析，以提升人機互動之效能。本研究以 AAM 演算法建立人臉模型，以擷取人臉之幾何特徵，藉由距離特徵與位移特徵之整合，以及時序之分析，設計 RVM 分類器與回歸模型，實現能夠分析程度、比例等情緒資訊之情緒辨識系統。本研究以此辨識系統，應用於音樂之選取，透過音樂之播放，適時轉換使用者情緒，達到人機互動之目的。

本論文之實驗，以本研究提出之 Likelihood 時序辨識系統，與 Likelihood 非時序辨識系統(直接使用本論文 3.3.1a 節 RVM 分類器進行 Likelihood 辨識)之辨識結果進行比較，對於 Likelihood 辨識分類至 7 種基本情緒而言，可以發現本研究提出之辨識系統辨識率達 95.83%，相較於 Likelihood 非時序辨識系統之 90.89% 來得高；另外，本研究以問卷調查對 Likelihood 進一步做混合情緒比例之辨識效能驗證，利用相關性做判定，其辨識結果與問卷調查成正相關(Correlation)，而利用 RMSE 與類別組合對結果做判定，同樣可以看出辨識結果與問卷調查結果情緒比例關係一致，二者辨識結果相符。對於 A-V 辨識而言，加入時序分析之情緒轉換過程較為合理，未有瞬間情緒辨識之情況，以標準差進行效能評估，本研究提出之辨識系統 A-V 之標準差分別為 1.20 與 1.82，相較於 Likelihood 非時序辨識之 2.35 與 4.02 來得低。

5.2. 未來展望

本論文目前提出之辨識系統，藉由時序分析，能夠對情緒做更細緻之描述與辨識，但其仍有幾點問題必須解決。

特徵點擷取方面，本研究所使用之 AAM 演算法，對於人臉情緒形變過大之表情無法準確擷取人臉特徵，主要原因在於其與樣板影像差別過大所致，或許可藉由建立更豐的資料庫，或加入局部特徵做改善。另外，就運算速度而言，AAM 建立人臉模型之速度仍有加強之空間，若前述之 AAM 準確性有所改善，可透過影像追蹤之方式，每次計算形變模型與紋理模型時，以前一次之人臉模型為基礎，迭代計算此次之人臉模型參數，降低迭代次數與運算量。

情緒辨識方面，本研究雖然加入時序對情緒做分析，但其仍以靜態情緒辨識為主，對話情境之動態情緒辨識則未能達到其辨識效能，主因仍在於運算速度與時序分析之完善性。在系統運算量納入考量情況下，透過更深入之時序分析，與人臉變化頻率分析，設法擷取人臉表情關鍵之辨識特徵，可以做為未來之探討方向，以求更自然、更廣範之人機互動為目標。

参考文献

- [1] A. Sharkey and N. Sharkey, "Children, and Elderly, and Interactive Robots Anthropomorphism and Deception in Robot Care and Companionship," *IEEE Robotics & Automation Magazine*, vol. 18, no. 1, pp. 32-38, 2011.
- [2] "PARO," available: www.parorobots.com.
- [3] "Sony-AIBO," available: www.sony-aibo.co.uk.
- [4] "Gecko System - Mobile Robot Solution for Safety, Security and Service," available: www.geckosystems.com/markets/CareBot.php.
- [5] "Robosoft," available: www.robosoft.com/robotic-solutions/healthcare/kompai/index.html.
- [6] D. Hanson, "Hanson Robotics Inc.," available: www.hansonrobotics.com.
- [7] A. Vinciarelli, M. Pantic, D. Heylen, C. Pelachaud, I. Poggi, F. D. Errico and M. Schroeder, "Bridging the Gap between Social Animal and Unsocial Machine: A Survey of Social Signal Processing," *IEEE Trans. Affective Computing*, no. 1, vol. 3, pp. 69-87, 2012.
- [8] G. S. Shergill, A. Sarrafzadeh, O. Diegel and A. Shekar, "Computerized Sales Assistants: The Application of Computer Technology to Measure Consumer Interest – A Conceptual Framework," *Journal of Electronic Commerce Research*, vol. 9, no. 2, pp. 176-191, 2008.
- [9] S. Gregory, "Spy on Sports Fans," *TIME Ideas*, 2013, available: <http://ideas.time.com/2013/03/14/10-big-ideas/slide/spy-on-sports-fans/>.
- [10] B. T. Horowitz, "Cybercare: Will Robots Help the Elderly to Live at Home for Longer?" *Scientific American*, June 21, 2010.
- [11] M. Swangnetr and D. B. Kaber, "Emotional State Classification in Patient-Robot

- Interaction Using Wavelet Analysis and Statistics-Based Feature Selection,” *IEEE Trans. Human-Machine Systems*, vol. 1, no. 43, pp. 63-75, 2013.
- [12] “Aethon,” available: www.aethon.com.
- [13] A. A. Salah and T. Gevers (eds.), “Computer Analysis of Human Behavior,” *Springer*, 2011, chapter 10.
- [14] P. Ekman and W. V. Friesen, “Unmasking the Face: A Guide to Recognizing Emotion from Facial Clues,” *Prentice Hall*, New Jersey, 1975.
- [15] J. A. Russell, “A Circumplex Model of Affect,” *Journal of Personality & Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.
- [16] W. Gu, C. Xiang, Y. V. Venkatesh, D. Huang and H. Lin, “Facial Expression Recognition Using Radial Encoding of Local Gabor Features and Classifier Synthesis,” *Pattern Recognition*, vol.45, no.1, pp.80-91, 2012.
- [17] M. Song, D. Tao, Z. Liu, X. Li and M. Zhou, “Image Ratio Features for Facial Expression Recognition Application,” *IEEE Trans. System Man and Cybernetics Part B-Cybernetics*, vol. 42, no. 3, pp. 779-788, 2010.
- [18] K. T. Song, M. J. Han and J. W. Hong, "Online Learning Design of an Image-Based Facial Expression Recognition System," To appear in *Intelligent Service Robotics*, Vol. 3, No. 3, pp. 151-162, 2010.
- [19] K. T. Song and S. C. Chien, “Facial Expression Recognition Based on Mixture of Basic Expression and Intensities,” in *Proc. IEEE Int. Conf. Systems, Man, and Cybernetics*, Seoul, South Korea, pp. 3123-3128, 2012.
- [20] C. M. Whissell, “The Dictionary of Affect in Language,” *Emotion: Theory, Research and Experience*, New York: *Academic Press*, 1989.
- [21] R. E. Thayer, “The Biopsychology of Mood and Arousal,” *New York: Oxford University Press*, 1989.

- [22] I. Hupont, E. Cerezo and S. Baldassarri, "Sensing Facial Emotions in A Continuous 2D Affective Space," *IEEE Int. Conf. Systems, Man and Cybernetics*, Istanbul, Turkey, pp. 2045-2051, 2010.
- [23] P. Yang, Q. Liu, X. Cui and D. N. Metaxas, "Facial Expression Recognition Using Encoded Dynamic Features," *IEEE Conf. Computer Vision and Pattern Recognition*, Anchorage, AK, USA, pp.1-8, 2008.
- [24] S. Hommel and U. Handmann, "AAM Based Continuous Facial Expression Recognition for Face Image Sequences," in *Proc. IEEE Int. Symp. Computational Intelligence and Informatics*, Budapest, Hungary, pp. 189-194, 2011.
- [25] Y. Zhang and Q. Ji, "Active and Dynamic Information Fusion for Facial Expression Understanding from Image Sequences," *IEEE Trans. Pattern Analysis and Machine Intelligent*, vol. 27, no. 5, pp. 699-714, 2005.
- [26] M. A. Nicolaou, H. Gunes and M. Pantic, "Output-Associative RVM Regression for Dimensional and Continuous Emotion Prediction," in *Proc. IEEE Int. Conf. Automatic Face & Gesture Recognition*, Santa Barbara, CA, USA, pp. 21-24, 2011.
- [27] N. Cristian and J. S. Taylor, "An Introduction to Support Vector Machines and Other Kernel-based Learning Methods," *New York: Cambridge University Press*, 2000.
- [28] M. E. Tipping, "Sparse Bayesian Learning and the Relevance Vector Machine," *Journal of Machine Learning Research*, vol.1, no.3, pp. 211–244, 2001.
- [29] A. Ethem, "Introduction to Machine Learning," Second Edition. *The MIT Press*, Cambridge, Massachusetts, London, England, 2010.
- [30] P. Viola and M. Jones, "Rapid Object Detection Using a Boosted Cascade of

- Simple Features,” *IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, Kauai, HI, pp. 511- 518, 2001.
- [31] R. Lienhart and J. Maydt, “An Extended Set of Haar-Like Features for Rapid Object Detection,” in *Proc. Int. Conf. Image Processing*, Rochester, NY, USA, pp. 900-903, 2002.
- [32] T. F. Cootes, G. J. Edwards and C. J. Taylor, “Active Appearance Models.” in *Proc. European Conf. Computer Vision*, Springer, Berlin, pp 484–498, 1998.
- [33] T. F. Cootes, G. J. Edwards and C. J. Taylor, “Active Appearance Models,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 6, pp. 681-685, 2001.
- [34] T. F. Cootes and C. J. Taylor, “Technical Report: Statistical Models of Appearance for Computer Vision,” *The University of Manchester School of Medicine*, 2004.
- [35] A. McAndrew, “Introduction to Digital Image Processing with Matlab,” *Thomson Course Technology*, 2004.
- [36] 陳奕堯, 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.
- [37] I. Matthews and S. Baker, “Active Appearance Models Revisited,” *Int. Journal of Computer Vision*, vol. 60, no.2, pp.135-164, 2004.
- [38] 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.
- [39] J.R. Shewchuk, “Triangle: Engineering a 2D Quality Mesh Generator and Delaunay Triangulator,” *First Workshop on Applied Computational Geometry, Proceedings*, Philadelphia, pp. 124-133, 1996.
- [40] P. Ekman and W.V. Friesen, “Facial Action Coding System (FACS): A Technique

for the Measurement of Facial Movement,” *Palo Alto, Calif: Consulting Psychologists Press*, 1978.

- [41] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews, “The Extended Cohn-Kanade Dataset (CK+): A Complete Dataset for Action Unit and Emotion-Specified Expression,” *IEEE Computer Society Conf. Computer Vision and Pattern Recognition Workshops*, San Francisco, CA, USA, pp. 94-101, 2010.
- [42] M. Frieddman and A. Kandel, “Introduction to Pattern Recognition,” *World Scientific*, 1999.
- [43] D. E. King, “Dlib-ml: A Machine Learning Toolkit,” *Journal of Machine Learning Research* 10, pp. 1755-1758, 2009.
- [44] D. J. C. MacKay, “The Evidence Framework Applied to Classification Networks,” *Neural Comput.*, vol. 4, no. 5, pp. 720–736, 1992.
- [45] L. Tierney and J. B. Kadane, “Accurate Approximations for Posterior Moments and Marginal Densities,” *Journal of the American Statistical Association*, vol. 81, no. 393, pp.82-86, 1986.
- [46] D. J. C. MacKay, “Bayesian Interpolation,” *Neural Comput.*, vol. 4, no. 3, pp. 415–447, 1992a.
- [47] F. Melgani and L. Bruzzone, “Classification of Hyperspectral Remote Sensing Images with Support Vector Machines,” *IEEE Trans. Geoscience and Remote Sensing Symposium*, vol. 42, no. 8, pp. 1778–1790, 2004.
- [48] T. Hastie and R. Tibshirani, “Classification by Pairwise Coupling,” *Annals of Statisitcs*, vol. 26, no. 2, pp. 451-471, 1998.
- [49] S. Kullback and R. A. Leibler, “On Information and Sufficiency,” *Annals of Mathematical Statistics*, vol. 22, no.1, pp. 79-80, 1951.
- [50] Y. H. Yang, Y. F. Su, Y. Ch. Lin and H. Chen, “Music Emotion Recognition: the

Role of Individuality,” in *Proc. the International workshop on Human-centered Multimedia*, Augsburg, Bavaria, Germany, pp.13-22, 2007.

- [51] M. J. Han, C. H. Lin, and K. T. Song, “Robotic emotional expression generation based-on mood transition and personality model,” *IEEE Trans. Cybernetics*, 2012.
- [52] J. A. Russell and M. Bullock, “Multidimensional Scaling of Emotional Facial Expressions: Similarity from Preschoolers to Adults,” *Journal of Personality and Social Psychology*, vol. 48, no. 5, pp. 1290–1298, 1985.
- [53] S. Jain, H. Changbo and J. K. Aggarwal , “Facial Expression Recognition with Temporal Modeling of Shapes,” *IEEE Int. Conf. Computer Vision Workshops*, Barcelona, Spain, pp. 1642-1649, 2011.
- [54] E. Silva, C. Esparza and Y. Mejia , “POEM-based Facial Expression Recognition, a New Approach,” *Image, Signal Processing, and Artificial Vision*, Antioquia, Colombia, pp. 162-167, 2012.

附錄一、基本人臉情緒混合比例問卷調查樣張

受測者基本資料：

性別：女 年齡：28

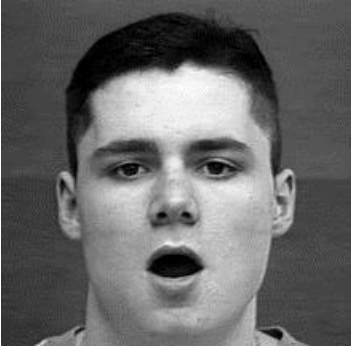





問卷說明：

請依序填入各影像所含基本情緒 Neutral(中性, Ne), Anger(生氣, An), Disgust(憎惡、厭惡, Di), Fear(害怕, Fe), Happy(高興, Ha), Sadness(傷心, Sa), Surprise(驚訝, Su)之百分比，百分比加總必須為 100%。

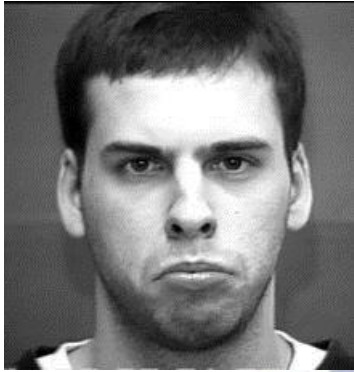
例如：

中性: 20	生氣: 40	厭惡: 10	害怕: 15
高興: 0	傷心: 15	驚訝: 0	

S002				S026			
							
指出 S002 影像所包含基本情緒比例之百分比				指出 S026 影像所包含基本情緒比例之百分比			
中性: 0	生氣: 70	厭惡: 30	害怕: 0	中性: 40	生氣: 30	厭惡: 30	害怕: 0
高興: 0	傷心: 0	驚訝: 0		高興: 0	傷心: 0	驚訝: 0	

<p style="text-align: center;">S074</p>  <p>指出 S074 影像所包含基本情緒比例之百分比</p>	<p style="text-align: center;">S015</p>  <p>指出 S015 影像所包含基本情緒比例之百分比</p>																
<table border="1"> <tr> <td>中性: 90</td> <td>生氣: 0</td> <td>厭惡: 0</td> <td>害怕: 0</td> </tr> <tr> <td>高興: 0</td> <td>傷心: 0</td> <td>驚訝: 10</td> <td></td> </tr> </table>	中性: 90	生氣: 0	厭惡: 0	害怕: 0	高興: 0	傷心: 0	驚訝: 10		<table border="1"> <tr> <td>中性: 70</td> <td>生氣: 0</td> <td>厭惡: 0</td> <td>害怕: 0</td> </tr> <tr> <td>高興: 30</td> <td>傷心: 0</td> <td>驚訝: 0</td> <td></td> </tr> </table>	中性: 70	生氣: 0	厭惡: 0	害怕: 0	高興: 30	傷心: 0	驚訝: 0	
中性: 90	生氣: 0	厭惡: 0	害怕: 0														
高興: 0	傷心: 0	驚訝: 10															
中性: 70	生氣: 0	厭惡: 0	害怕: 0														
高興: 30	傷心: 0	驚訝: 0															
<p style="text-align: center;">S013</p>  <p>指出 S013 影像所包含基本情緒比例之百分比</p>	<p style="text-align: center;">S028</p>  <p>指出 S028 影像所包含基本情緒比例之百分比</p>																
<table border="1"> <tr> <td>中性: 20</td> <td>生氣: 0</td> <td>厭惡: 60</td> <td>害怕: 0</td> </tr> <tr> <td>高興: 0</td> <td>傷心: 0</td> <td>驚訝: 20</td> <td></td> </tr> </table>	中性: 20	生氣: 0	厭惡: 60	害怕: 0	高興: 0	傷心: 0	驚訝: 20		<table border="1"> <tr> <td>中性: 30</td> <td>生氣: 0</td> <td>厭惡: 0</td> <td>害怕: 10</td> </tr> <tr> <td>高興: 0</td> <td>傷心: 0</td> <td>驚訝: 60</td> <td></td> </tr> </table>	中性: 30	生氣: 0	厭惡: 0	害怕: 10	高興: 0	傷心: 0	驚訝: 60	
中性: 20	生氣: 0	厭惡: 60	害怕: 0														
高興: 0	傷心: 0	驚訝: 20															
中性: 30	生氣: 0	厭惡: 0	害怕: 10														
高興: 0	傷心: 0	驚訝: 60															
<p style="text-align: center;">S073</p>  <p>指出 S073 影像所包含基本情緒比例之百分比</p>	<p style="text-align: center;">S009</p>  <p>指出 S009 影像所包含基本情緒比例之百分比</p>																
<table border="1"> <tr> <td>中性: 10</td> <td>生氣: 0</td> <td>厭惡: 0</td> <td>害怕: 20</td> </tr> <tr> <td>高興: 0</td> <td>傷心: 0</td> <td>驚訝: 70</td> <td></td> </tr> </table>	中性: 10	生氣: 0	厭惡: 0	害怕: 20	高興: 0	傷心: 0	驚訝: 70		<table border="1"> <tr> <td>中性: 10</td> <td>生氣: 0</td> <td>厭惡: 20</td> <td>害怕: 0</td> </tr> <tr> <td>高興: 0</td> <td>傷心: 70</td> <td>驚訝: 0</td> <td></td> </tr> </table>	中性: 10	生氣: 0	厭惡: 20	害怕: 0	高興: 0	傷心: 70	驚訝: 0	
中性: 10	生氣: 0	厭惡: 0	害怕: 20														
高興: 0	傷心: 0	驚訝: 70															
中性: 10	生氣: 0	厭惡: 20	害怕: 0														
高興: 0	傷心: 70	驚訝: 0															

S025



指出 S025 影像所包含基本情緒比例之百分比

S004



指出 S004 影像所包含基本情緒比例之百分比

中性: 80	生氣: 10	厭惡: 10	害怕: 0	中性: 0	生氣: 40	厭惡: 60	害怕: 0
高興: 0	傷心: 0	驚訝: 0		高興: 0	傷心: 0	驚訝: 0	

