

國立雲林科技大學
資訊工程研究所碩士班

碩士論文

結合生理訊號之臉部表情辨識



**Facial Expression Recognition with
Consideration of Physiological Signals**

研究生：蔡政勳

指導教授：張傳育 博士

中華民國 九十七 年 六 月

結合生理訊號之臉部表情辨識

**Facial Expression Recognition with Consideration of Physiological
Signals**

研究生：蔡政勳 Jeng-Shiun Tsai

指導教授：張傳育 Chuan-Yu Chang

國立雲林科技大學

資訊工程研究所碩士班



Graduate School of Computer Science and Information Engineering
National Yunlin University of Science & Technology
in Partial Fulfillment of the Requirements
for the Degree of
Master of Science
in
Computer Science and Information Engineering

June 2008

Douliu, Yunlin, Taiwan, Republic of China

中華民國九十七年六月

國立雲林科技大學博、碩士班 論文著作權歸屬與行使同意書

論文名稱：結合生理訊號之臉部表情辨識

依照智慧財產局著作權法相關規定，本論文著作權之歸屬與行使如下：

本論文之學生為著作人

說明：論文之指導教授僅為觀念之指導，並未參與內容表達之撰寫，依著作權法規定，學生為該論文之著作人，並於論文完成時，即享有該論文之著作權。

本論文之學生與指導教授為共同著作人

說明：論文之指導教授不僅為觀念的指導，且參與內容之表達而與學生共同完成論文，且各人之創作，不能分離利用者，則為共同著作，學生與指導教授為論文的共同著作人並共同享有著作權，此等共同著作權（包括著作財產權及著作人格權）的行使，即應取得學生與指導教授之共同同意後，始得為之。

共同著作人行使方式，協議事項如下：

一、 論文不要公開發表。

論文要公開發表，發表論文時作者順序（作者在左者為前，在右者為後）：

張傳育，蔡政勤

二、 論文事後不可修改。

論文事後可以修改。

三、 論文未來不可授權他人利用。

論文未來可以授權他人利用。

本同意書壹式 貳 份，由學生與指導教授雙方各執 壹 份為憑。

學生簽章：蔡政勤

指導教授簽章：張傳育

中華民國 97 年 7 月 31 日

國立雲林科技大學
研究所碩士班
論文口試委員會審定書

本校 資訊工程研究所 碩士班 蔡政勳 君

所提論文 結合生理訊號之臉部表情辨識

合於碩士資格水準，業經本委員會評審認可。

口試委員：

詹寶珠

張傳育

許明華

李岸南

指導教授：

張傳育

研究所所長：

王 以 傳

中華民國 97 年 07 月 07 日

結合生理訊號之臉部表情辨識

學生：蔡政勳

指導教授：張傳育 博士

國立雲林科技大學資訊工程研究所碩士班

摘要

在人與人的互動之間，臉部表情和情緒表達佔有很大部分。一般人可以藉由觀察某人的表情、肢體動作，甚至藉由生理訊號的特徵，來判別這個人的情緒為何，但這過程需要經由我們的經驗去學習而得到認知。所以情緒表達在人與人的社交關係中扮演著相當重要的角色，它幫助人們了解彼此內心的感受。

本篇論文提出一套結合視訊與生理訊號之臉部表情辨識系統。臉部表情的特徵是從不同的表情上擷取出特徵點，並且組成一組特徵向量。生理訊號方面，擷取三種生理訊號：皮膚電位反應、指溫和心率，分析並定義出特徵向量。在分類與辨識方面，以半徑基底函數類神經網路(RBFNN)來分辨四種情緒：愛、歡樂、驚訝和害怕。在所設計的情緒激發實驗中，蒐集並擷取臉部和生理特徵的資料，並且訓練和測試類神經網路。藉由臉部特徵及生理訊號，綜合這兩種特徵我們可以達到臉部表情辨識，實驗結果顯示所提出的方法可以達到不錯臉部表情辨識結果。

Facial Expression Recognition with Consideration of Physiological Signals

Student: Jeng-Shiun Tsai

Advisor: Dr. Chuan-Yu Chang

**Graduate School of Computer Science & Information Engineering
National Yunlin University of Science & Technology**

ABSTRACT

Facial expression and emotion expression is very important in life. We can know what the emotion he is by observing people facial expression, body movement, and the features of physiological signals. However, it should be realized by learning. Emotion expression plays an important role in human interaction, and it helps people realize to what they feel to each other.

A facial expression recognition system with consideration of physiological signals is proposed. In this system, facial features are extracted from specific feature points in facial images. The physiological features vector consists of three physiological signals, skin conductivity, finger temperature and heart rate. These features are set to inputs of radial basis function neural network (RBFNN) is applied to classify four emotions, love, joy, surprise, and fear. By combining the characteristic of facial and physiological features, we can achieve facial expression recognition. The experiment result shows that the proposed method can achieve good performance.

誌謝

光陰似箭，時空飛梭，想想這兩年的碩士生活就像是人生的走馬燈，一眨眼就過去了，讓人不勝追思。能在此實驗室裡度過研究生涯，不管是生活氣氛、同儕間的情感、老師的教悔，都讓我無法忘懷，也感到非常幸運，這些點滴都是讓我有勇氣、動力走完這段路以及完成此篇論文，而讓我的人生有段美好的回憶。

首先，要感謝的是我的指導教授—張傳育博士這兩年的用心指導，不管是課業上的指點、待人處世的道理，使我培養出研究精神及生活經驗，這一切都讓我成長許多，在此對我的指導教授致上最高的敬意。另外，還要特別感謝口試委員—詹寶珠教授、李宗南教授和許明華教授在口試時給予論文許多寶貴的意見及指導，使論文更加完善。

另外，感謝實驗室的博班學長—宏仁及俊錫，以及上屆的學長姐—綺綺、金璜(可達)、詠榮、瑞泰、青諭、宗穎，在研究及生活都能帶給我適時的指導；感謝一起趕進度、一起娛樂玩耍的同窗戰友—小虎、小風、鄭婷、宏榕、小李、洪爺，靠著大家的努力和韌性、無比偉大的潛力，讓我們一起畢業了；感謝學弟妹們—易璉、香意、阿強、威爺、阿誠，及其他實驗室的同學們，有了你們讓生活更是多彩多姿。還有那群 辦高中同學—朱、春榮、勇迪、昭達、嘉莉等，感謝你們陪我走過這段，讓我思考著未來人生的意義。

最後，感謝我的家人，一直支持我到完成碩士學業，給我無後顧之憂的生活，希望能分享我的成果，感謝你們。

蔡政勳 謹致於雲科資工

2008 年 6 月 31 日

目錄

摘要	-----	i
ABSTRACT	-----	ii
誌謝	-----	iii
目錄	-----	iv
表目錄	-----	vi
圖目錄	-----	vii
第一章	緒論 -----	1
1.1	研究動機 -----	1
1.2	相關文獻 -----	2
1.3	研究目標 -----	5
1.4	章節大綱 -----	7
第二章	臉部表情特徵 -----	8
2.1	人臉偵測 -----	8
2.2	人臉特徵點擷取 -----	9
2.2.1	瞳孔特徵點 -----	10
2.2.1.1	Hue 空間 -----	10
2.2.1.2	IOD 灰階二值化 -----	12
2.2.2	眼睛特徵點 -----	14
2.2.3	眉毛特徵點 -----	16
2.2.4	嘴巴特徵點 -----	17
2.3	表情特徵向量 -----	19
第三章	生理訊號特徵 -----	22
3.1	生理訊號 -----	22
3.2	生理訊號前處理及特徵向量 -----	27
第四章	表情分類與辨識 -----	29
4.1	半徑基底函數(RBF)類神經網路 -----	29
4.2	應用半徑基底函數類神經網路於分類特徵向量與辨識表情 -----	32
第五章	實驗結果與討論 -----	36
5.1	資料收集 -----	36
5.2	情緒激發實驗之測試 -----	44

5.3	表情資料庫之測試 -----	46
5.3.1	JAFFE 多人影像測試 -----	46
5.4	實驗比較 -----	49
第六章	結論 -----	52
參考文獻	-----	53
附錄	Facial Expression Recognition with Consideration of Physiological Signals -----	56



表目錄

表 5.1	影片片段之內容描述 -----	38
表 5.2	JAFFE 各表情張數 -----	42
表 5.3	表情正確率 -----	45
表 5.4	生理訊號正確率 -----	45
表 5.5	結合表情和生理訊號之正確率 -----	45
表 5.6	JAFFE 之 Cross-validation 四類表情辨識率 -----	47
表 5.7	JAFFE 之一半訓練、一半測試五類表情辨識率 -----	47
表 5.8	JAFFE 之一半訓練、一半測試四類表情辨識率 -----	48
表 5.9	[15]Cross-validation 之表情辨識率-----	49
表 5.10	[15]一半訓練一半測試之表情辨識率 -----	49
表 5.11	本論文提出方法 Cross-validation 之表情辨識率 -----	49
表 5.12	本論文提出方法一半訓練一半測試之表情辨識率 -----	50
表 5.13	Cheng [20]使用之訓練及張數 -----	50
表 5.14	Cheng [20] 辨識率之比較 -----	50
表 5.15	本論文提出方法之一半訓練一半測試之辨識率 -----	51
表 5.16	Bayshyal [8] 辨識率之比較 -----	51

圖目錄

圖 1.1	運動單元(action unit, AU) -----	4
圖 1.2	表情辨識系統流程圖 -----	6
圖 2.1	人臉偵測範例 -----	9
圖 2.2	臉部之特徵點 -----	9
圖 2.3	H 成分之瞳孔位置可能範圍 -----	11
圖 2.4	以圖 2.3 為例之 H 成分的左眼水平及垂直投影曲線圖 -----	11
圖 2.5	在 H 成分下得到之瞳孔位置 -----	12
圖 2.6	瞳孔上、下邊界之擷取 -----	13
圖 2.7	灰階二值化方法得到之瞳孔位置 -----	14
圖 2.8	最終左、右眼瞳孔位置 -----	14
圖 2.9	眼內角點搜尋方式示意圖 -----	15
圖 2.10	眼睛特徵 -----	16
圖 2.11	眉毛特徵 -----	17
圖 2.12	嘴巴區域經過灰階轉換、等化處理及 IOD 二值化處理的結果 -----	18
圖 2.13	HIS 色彩空間之圓型色彩平面 -----	19
圖 2.14	嘴巴區域之 H 成分經過二值化的結果 -----	19
圖 2.15	嘴巴特徵點 -----	19
圖 2.16	表情特徵向量計算流程圖 -----	20
圖 2.17	臉部特徵點及特徵距離 -----	21
圖 3.1	PQRST 波形型態 -----	23
圖 3.2	肌電圖範例 -----	23
圖 3.3	皮膚電位反應 -----	24
圖 3.4	皮膚溫度 -----	25
圖 3.5	血流量 -----	25
圖 3.6	原始呼吸訊號 -----	26
圖 3.7	心率 -----	26
圖 3.8	生理訊號處理流程圖 -----	27
圖 4.1	半徑基底函數(RBF)類神經網路的架構圖 -----	31
圖 4.2	兩組 RBF 類神經網路分類與辨識流程 -----	32
圖 4.3	情緒辨識之 RBF 網路架構圖 -----	33
圖 4.4	RBFNN1 之網路架構圖 -----	33

圖 4.5	RBFNN2 之網路架構圖 -----	34
圖 4.6	RBFNN3 之網路架構圖 -----	35
圖 5.1	情緒激發實驗示意圖 -----	37
圖 5.2	情緒激發實驗環境之全景 -----	37
圖 5.3	情緒激發影片之片段截圖 -----	39
圖 5.4	表情影像資料之範例 -----	40
圖 5.5	生理訊號在不同情緒之範例 -----	41
圖 5.6	JAFFE 表情範例 -----	42



第一章

緒論

1.1 研究動機

情緒表達，在人與人的社交關係中扮演相當重要角色，並幫助人們瞭解彼此內心的感受。常見的情緒表達有臉部表情、肢體動作、說話語調、講話用字、身體姿勢與身體反應等等。除了這些可以觀察到的情緒表達方式之外，通常也會伴隨著生理訊號的變化。

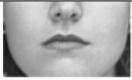





















在許多的情緒反應之中，表情是人類特有的臉部特徵，而且通常是複雜多樣的。我們可以藉由五官位置的變動，如眼睛、眉毛、嘴巴或臉部肌肉來斷定其表情。所以一般人可以輕易地藉由觀察某人的表情來判別此人的情緒，但這過程需要經由我們的經驗去學習而得到認知。除此之外，在情緒的判別上，生理訊號是一個很有效的指標。由生理學可知，當情緒被激發時我們不能去自主控制生理訊號的反應，因此生理訊號的反應是非自主神經的，對於情緒判別，這個特性是非常有用的資訊。常見人類的生理訊號有心電、肌電、腦波、血壓、脈搏、皮膚電位、體溫、呼吸深淺與心跳頻率。要讓電腦學習如何判別臉部表情和生理情緒是件非常困難的工作，電腦不像人類一樣會自動學習、思路複雜，必須要給電腦指令才會有所動作。在人工智慧研究領域，有專家發明機器人，讓電腦學習人類的行為，使其和人類一樣具有自主能力。因此，本論文的目的在提出一套結合臉部表情和基於生理訊號的表情辨識系統，藉由此系統更加瞭解受試者內心真實的情緒。

1.2 相關文獻

多年來，在表情分析的相關研究裡，一直都有許多專家學者在研究、開發。對於人臉動作幾何描述比較知名同時也是最早用於人臉表情辨識的方法之一是人臉動作編碼系統(Facial Action Coding System, FACS)[1]，也有很多辨識方法都以它為基礎。Ekman 和 Friesen 對臉部肌肉的運動及其對表情的控制作用作了深入研究，在 1978 年開發了 FACS 來描述人臉表情和研究人類的認知行為，在人臉的範圍上，一共劃分約 46 個既相互獨立且又相互聯繫的運動單元(action unit, AU)，並分析出這些運動單元的特徵及其控制的主要區域與相關的表情，如圖 1.1 所示。此外，他們也定義出 6 種常見的人臉表情與其範例：害怕(fear)、驚訝(surprise)、悲傷(sad)、生氣(anger)、厭惡(disgust)和快樂(happy) [2]。分析出這些運動單元的方法很直觀，易於理解，但在這過程中需要耗費大量人力和時間來標記影像的特徵運動點 [3]，所以在這方法下的系統編碼速度、可靠性、實用性就成為障礙。Wang *et al.* [4] 所發展出的方法則是建出一個 Active appearance model (AAM) 來追蹤臉部的定位點，再利用這些定位點的位置關係來判定表情種類，但建造一個通用的 AAM 必須利用人工標記的方式來標記事先定義的定位點，而且需要大量的人臉資訊，此步驟相當複雜且繁瑣，在此實驗環境人臉的拍攝背景也必須相當單純。在表情辨識方面，類神經網路常被使用於分類及辨識，Silva [7]設計了改良式半徑基底函數(modified radial basis functions, RBFs)類神經網路在靜態影像(Cohn-Kanade facial expression database)中分類及辨識六種表情(fear、surprise、sad、anger、disgust 和 happy)，Bashyal [8]使用了學習向量量化(Learning Vector Quantization, LVQ)類神經網路在靜態影像(Japanese Female Facial Expression database, JAFFE)中分類及辨識七種表情(neutral、fear、surprise、sad、anger、disgust 和 happy)，實驗結果均顯示出具有相當不錯的辨識率。

在基於生理訊號的情緒辨識的研究中，目前有許多方法被提出來。Takahashi [5] 設計情緒激發實驗來讓受試者激發情緒，收集了三種生理訊號的樣本，計算出訊號的特徵之後，使用一組支援向量機(Support Vector Machine, SVM)類神經網路，分別來分析及辨識五種(歡樂(joy)、生氣(anger)、悲傷(sadness)、害怕(fear)和放鬆(relax))及三種情緒(歡樂(joy)、生氣(anger)和放鬆(relax))，但辨識結果不太理想。Rosalind [6] 在長時間下收集受試者的五種生理訊號，得到計算後的訊號特徵，再經過特徵選取(Feature Selection)的動作，辨識八種情緒(自然(neutral)、生氣(anger)、厭惡(hatred)、悲痛(grief)、柏拉圖的愛(platonic love)、羅曼蒂克的愛(romantic love)、歡樂(joy)和敬畏(reverence))，得到訊號資料是針對同一個人而需花數天來收集，所以將耗費許多時間且過程非常冗長。



NEUTRAL	AU 9	AU 10	AU 12	AU 20
				
Lips relaxed and closed.	The infraorbital triangle and center of the upper lip are pulled upwards. Nasal root wrinkling is present.	The infraorbital triangle is pushed upwards. Upper lip is raised. Causes angular bend in shape of upper lip. Nasal root wrinkle is absent.	Lip corners are pulled obliquely.	The lips and the lower portion of the nasolabial furrow are pulled pulled back laterally. The mouth is elongated.
AU15	AU 17	AU 25	AU 26	AU 27
				
The corners of the lips are pulled down.	The chin boss is pushed upwards.	Lips are relaxed and parted.	Lips are relaxed and parted; mandible is lowered.	Mouth stretched open and the mandible pulled downwards.
AU 23+24	AU 9+17	AU9+25	AU9+17+23+24	AU10+17
				
Lips tightened, narrowed, and pressed together.				
AU 10+25	AU 10+15+17	AU 12+25	AU12+26	AU 15+17
				
AU 17+23+24	AU 20+25			
				
















NEUTRAL	AU 1	AU 2	AU 4	AU 5
				
Eyes, brow, and cheek are relaxed.	Inner portion of the brows is raised.	Outer portion of the brows is raised.	Brows lowered and drawn together	Upper eyelids are raised.
AU 6	AU 7	AU 1+2	AU 1+4	AU 4+5
				
Cheeks are raised.	Lower eyelids are raised.	Inner and outer portions of the brows are raised.	Medial portion of the brows is raised and pulled together.	Brows lowered and drawn together and upper eyelids are raised.
AU 1+2+4	AU 1+2+5	AU 1+6	AU 6+7	AU 1+2+5+6+7
				
Brows are pulled together and upward.	Brows and upper eyelids are raised.	Inner portion of brows and cheeks are raised.	Lower eyelids cheeks are raised.	Brows, eyelids, and cheeks are raised.

圖 1.1 運動單元(action unit, AU)，數字為單元之編號

1.3 研究目標

在本論文中，提出一套結合視訊與生理訊號之臉部表情辨識。首先，在影像中進行人臉偵測和追蹤，並且擷取出人臉範圍。在臉部表情方面，取出人臉區域範圍之後，在此人臉範圍取出臉部特徵點，並且利用這些特徵點計算並組成一組特徵向量。在生理訊號方面，擷取出三種生理訊號：皮膚電位反應(skin conductivity)、指溫(finger temperature)和心率(heart rate)，分析並定義出特徵向量。在得到表情的特徵向量及生理訊號特徵向量之後，將兩種特徵向量輸入到半徑基底函數類神經網路(Radial Basis Function Neural Network, RBFNN)去分類及辨識四種情緒：害怕(fear)、愛(love)、快樂(joy)和驚訝(surprise)。圖 1.2 為整個系統架構的流程圖。



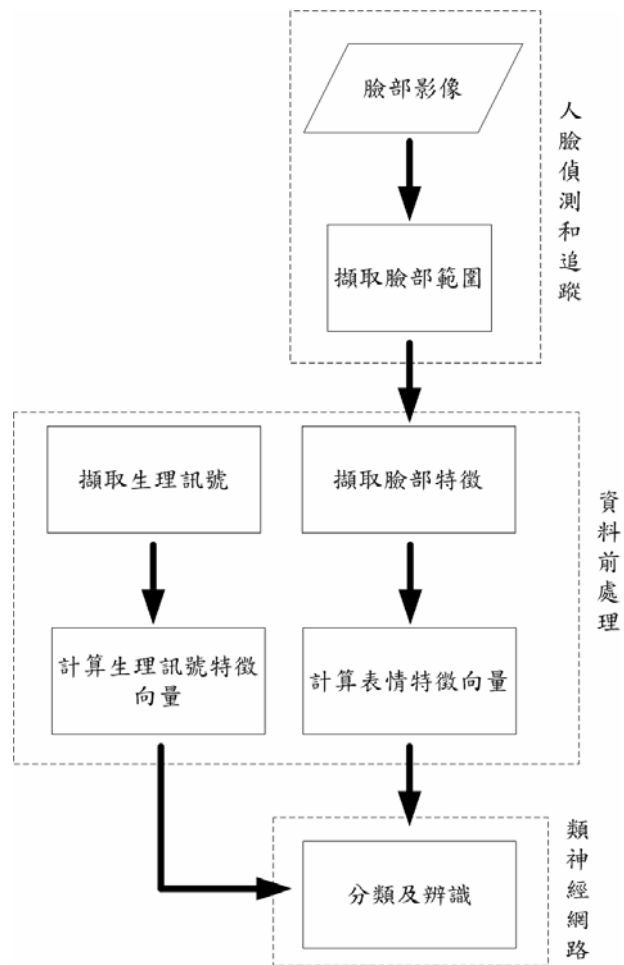


圖 1.2 表情辨識系統流程圖

1.4 章節大綱

本論文的其他部分內容安排如下：在第二章中，將詳細介紹臉部特徵點擷取方法及定義出表情特徵向量。第三章介紹生理訊號及定義出生理訊號特徵向量。使用類神經網路的分類與辨識方法第四章中呈現。第五章為實驗結果與討論。最後，第六章為本論文所提出的方法做一些結論。



第二章

臉部表情特徵

在一個完整的表情辨識系統中，不外乎是從人臉偵測開始，然後再擷取出人臉特徵，到最後的分類與辨識表情，而每一步驟都相當重要。以下各節將針對於人臉偵測到臉部特徵擷取步驟作一個詳細的介紹。

2.1 人臉偵測

目前有許多人臉偵測方法已被提出，且有相當不錯的結果[16-18]。在人臉偵測與追蹤的方法，我們採用 Chang [9]所提出的適應性色彩空間轉換之人臉追蹤，在此方法中，根據環境條件的變動去選擇適合的色彩空間，無論是在靜態影像或動態序列影像，彩色影像或灰階影像，皆能自動偵測人臉的位置及框取出人臉區域。此外，即使環境背景複雜或是有光源影響，或是影像中有多個人臉，偵測效果也是非常的理想。圖 2.1 為人臉偵測的範例。

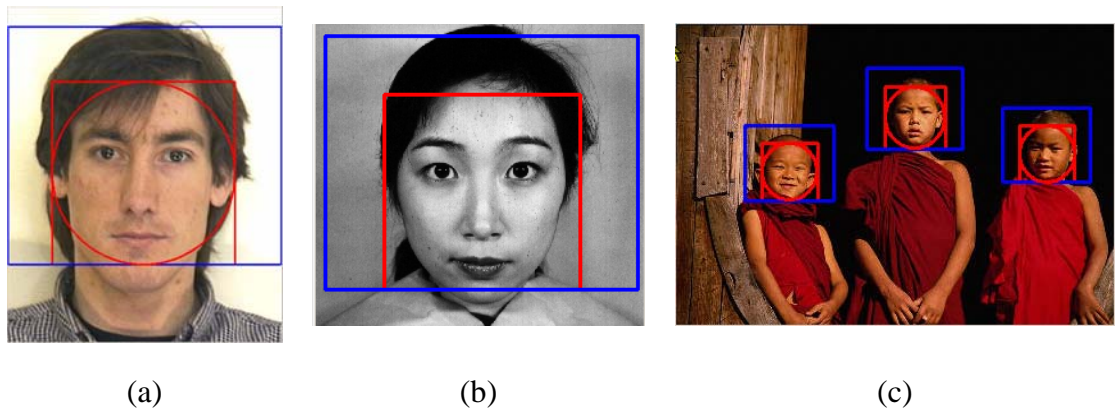


圖 2.1 人臉偵測範例 (a)彩色影像；(b)灰階影像；(c)多人臉彩色影像

2.2 人臉特徵點擷取

由於人臉表情的展現，絕大多數是根據五官的形狀及位置的變化來判斷，所以針對變動量較明顯的五官擷取特徵點，而且這些特徵點必須當作判斷人臉表情的變化的基礎，如此對於表情判斷的計算上會簡單許多。如圖 2.2 所示，我們在眼睛、眉毛、嘴巴這三部份擷取人臉特徵點，即眼睛左右各 3 點，眉毛左右各 2 點，嘴巴 4 點，一共取 14 個點。

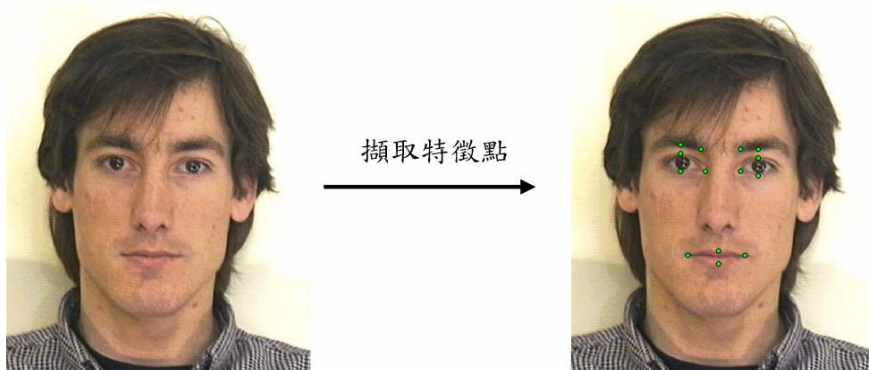


圖 2.2 臉部之特徵點

2.2.1 瞳孔特徵點

我們必須依據某些事實來選取各個特徵點的位置，對於整張人臉來說，瞳孔所在的位置是較好辨認的，所以我們將以瞳孔位置為尋找其他特徵點的依據。以下各小節將介紹利用色調 Hue 成分和 IOD 灰階二值化方法來取出瞳孔特徵點。

2.2.1.1 Hue 成分

在 HSI (hue, saturation, intensity) 色彩空間的 H(hue) 成份上，瞳孔表現是最明顯的色彩，所以我們先取瞳孔區域的大概範圍，亦即定為在 2.1 節中所框取人臉影像區域中從 1/4 到 1/2 高、1/3~11/12 寬為瞳孔大概範圍，即人臉上半部之中間區域範圍，如圖 2.3 所示。若能擷取出瞳孔位置，就能以瞳孔所在位置為依據，取出相對位置來尋找其他臉部特徵點。

在 HSI 色彩空間之 H 成份的定義，我們可以假設給一個以 RGB 色彩表示的影像，在每一個 RGB 像素的 H 成分可由以下方程式獲得：

$$H = \begin{cases} \theta^\circ & \text{if } B \leq G \\ 360^\circ - \theta^\circ & \text{if } B > G \end{cases} \quad (2.1)$$

其中

$$\theta = \cos^{-1} \left\{ \frac{\frac{1}{2}[(R-G) + (R-B)]}{[(R-G)^2 + (R-B)(G-B)]^{1/2}} \right\} \quad (2.2)$$

得到在 H 空間的瞳孔區域後，以左眼來說，我們在此區域的左半部作 H 空間的水平和垂直投影，如圖 2.4 所示，由於瞳孔位置並不會落在靠近區域邊界上，所以我們尋找水平及垂直投影範圍從第 15 個位置至倒數第 15 個位置的範圍取出最

大值水平及垂直投影的位置，即我們所認定的瞳孔位置，如此可以過濾掉不必要的投影值。同樣的方法，亦可找到右眼瞳孔的位置，最終結果如圖 2.5 所示。

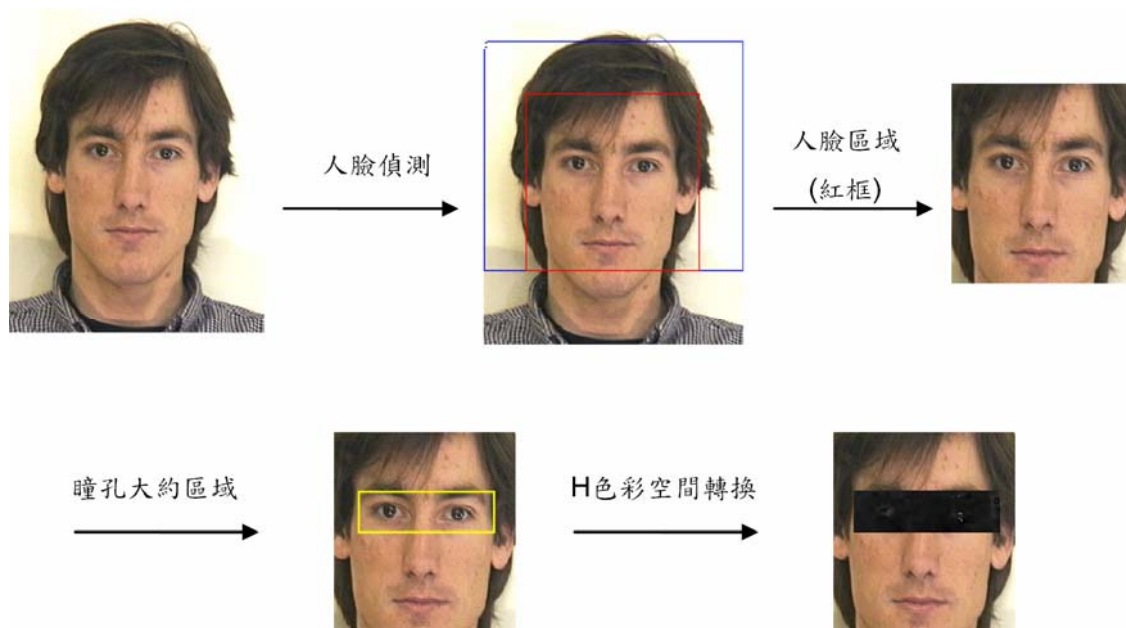


圖 2.3 H 成分之瞳孔位置可能範圍

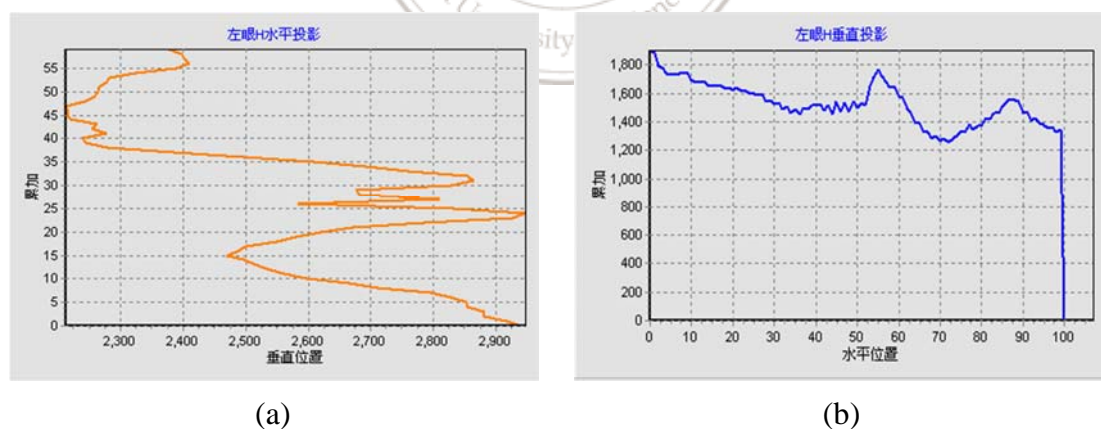


圖 2.4 以圖 2.3 為例之 H 成分的左眼(a)水平及(b)垂直投影曲線圖



圖 2.5 在 H 成分下得到之瞳孔位置

2.2.1.2 IOD 灰階二值化

IOD(intensity optical density)二值化是設一個臨界值把影像給二值化，但臨界值是依照整張影像的全部像素值比例而設定的。假設一個二維灰階影像 M 的函數為 $f(x,y)$ ，經過二值化後的影像為 $B(x,y)$ ，如下式所示：

$$B(x,y) = \begin{cases} 1 & \text{if } f(x,y) > T \\ 0 & \text{if } f(x,y) \leq T \end{cases} \quad (2.3)$$

其中影像 M 的灰階值 $f(x,y)$ 為 0 到 255 的值， T 為一個臨界值。所以我們必須知道一張影像的直方圖(histogram)才能算出全部像素值的比例。假設 $h(r)$ 為一張影像在直方圖的分布， r 的值是在 0 到 255， N 是一張影像所有的像素點總數， I 是介於 0 到 1 之間的比例值，IOD 為一個臨界值，則

$$iod = h^{-1}(I \times N) \quad (2.4)$$

我們把計算出 iod 臨界值拿來做二值化即是 IOD 二值化。

我們以 IOD 灰階二值化方法，再次搜尋瞳孔位置。首先，對於所框出人臉區域作 Sobel 濾波器邊緣偵測[9]，並正規劃到 -1~1 之間，這裡我們取 0.5 作為臨界值，當作二值化影像的依據，若大於 0.5 則設為 0，反之則設為 1，再以此二值化影像 $I(x,y)$ 取水平統計直方圖(histogram) $H(y)$ ，利用此直方圖 $H(y)$ 作為瞳孔區域上邊界的依據， $H(y)$ 定義如下：

$$H(y) = \sum_{x=x_1}^{x=x_2} I(x, y) \quad (2.5)$$

其中 x_1 為人臉區域左邊界， x_2 為人臉區域右邊界。對水平統計直方圖 $H(y)$ 由上往下找，偵測上述定義大約瞳孔區域高度範圍，當找到第一個 $H(y)$ 的值大於人臉區域寬度的 1/15 時，將其 y 座標定為同瞳孔位置上邊界；同樣的，對水平統計直方圖 $H(y)$ 由下往上找，當找到第一個 $H(y)$ 的值大於人臉區域寬度的 1/15 時，將其 y 座標定為同瞳孔位置下邊界。其流程如圖 2.6 所示。

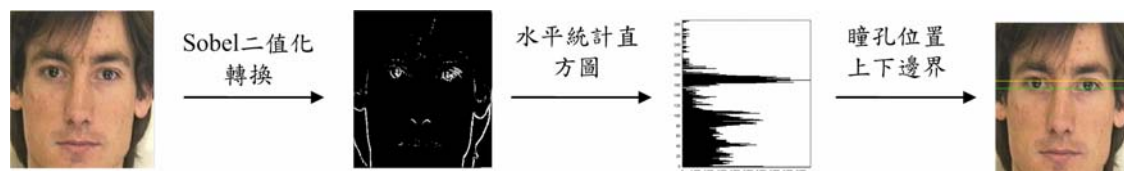


圖 2.6 瞳孔上、下邊界之擷取

接下來，在此上、下邊界之內區域高，由上述所定義瞳孔可能位置的寬長來找出瞳孔位置。由於瞳孔與其附近區域做比較會有較黑的顏色，所以只針對區域的灰階值取直方統計圖(histogram)，再以動態的方式來取臨界值，我們定義瞳孔的顏色會在此區域 I 值為 0.05 時所對應的臨界值內，因此由直方統計圖從灰階值為 0 開始算起，累加到個數為此區域的總像素個數的 I 值為 0.05 時所對應的臨界值，若區域內像數值小於臨界值，則設為 1，反之則設為 0。接下來，以左、右眼個別計算，把兩眼區域二值化的點算出平均位置即為最終瞳孔位置，如圖 2.7 所示。

最後，我們將在 H 空間和 IOD 灰階二值化方法所得到的瞳孔位置，個別計算兩種方法所取得左、右眼瞳孔位置的平均位置，最終得到左、右眼瞳孔的位置，如圖 2.8 所示。



圖 2.7 灰階二值化方法得到之瞳孔位置

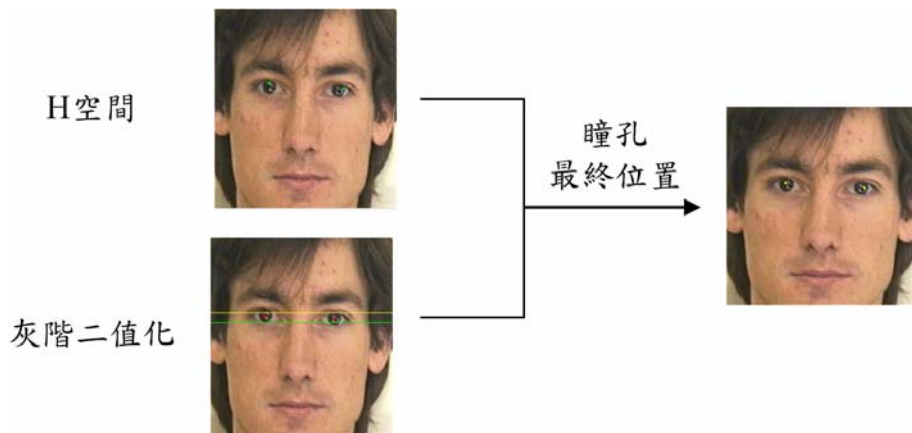


圖 2.8 最終左、右眼瞳孔位置

2.2.2 眼睛特徵點

得到兩眼瞳孔的位置之後，再取其歐幾里德距離，我們定義以瞳孔為中心，往上、下各 $1/7$ 瞳距，往左、右各 $1/3$ 瞳距為眼睛大約所在的區域。把此區域轉換成灰階，對此區域做灰階統計直方圖(histogram)，再針對此直方圖以動態方式取臨界值作 IOD 灰階二值化。這裡我們定義這段區域面積 I 值為 0.35 時所對應的臨界值當作二值化臨界值；換句話說，從直方圖的灰階值為 0 開始算起，累加到個數為此區域的總像素的第 35% 個，將這個第 35% 個的灰階值當作臨界值，若小於此臨界值設為 1 ，反之則設為 0 。圖 2.10(a)為二值化結果。另外，再對此眼睛區域做 Sobel 濾波器邊緣偵測[10]，這裡我們取灰階值 127 當作臨界值作二值化，若小於此臨界值設為 1 ，反之則設為 0 。圖 2.10(b)為 Sobel 邊緣偵測二值化結果。

以左眼來說，從邊界最右下點開始，以圖 2.9(a)所示的方法開始搜尋，直到找到二值化為 1 且 Sobel 邊緣偵測二值化為 1 的點，此點我們可以當作左眼內角的特徵點。另外，以瞳孔為中心，由上往下找水平方向 3 個像素點內，直到找到二值化為 1 且 Sobel 邊緣偵測二值化為 1 的點即為左眼上點特徵點；再由下往上找水平方向 3 個像素點內，直到找到二值化為 1 且 Sobel 邊緣偵測二值化為 1 的點為左眼下點特徵點。同樣的方法，亦可以找到右眼內角特徵點、右眼上點特徵點、右眼下點特徵點，而右眼內角特徵點搜尋方法如圖 2.9(b)所示。圖 2.10(c)為眼睛特徵點的最終結果。

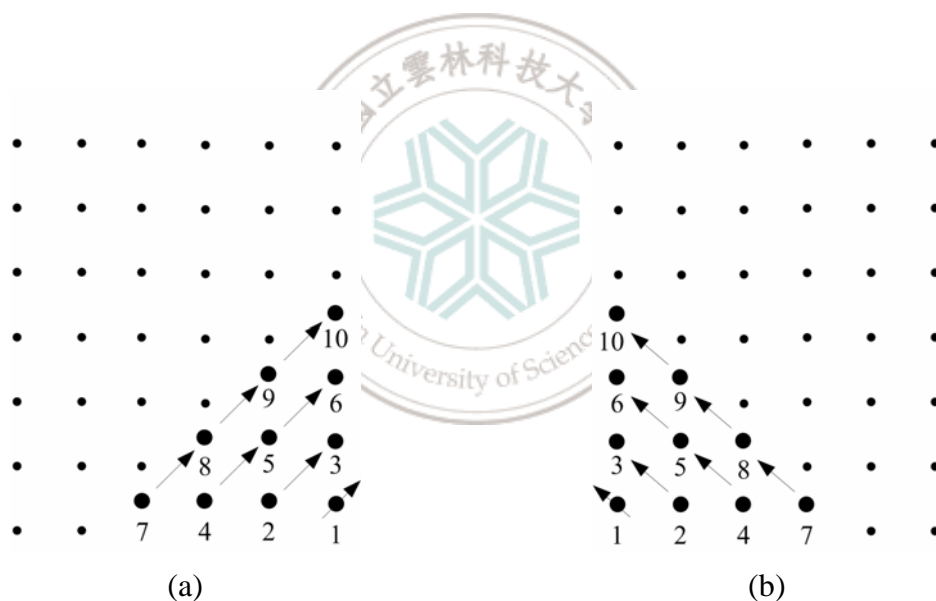


圖 2.9 眼內角點搜尋方式示意圖 (a)左眼從最右下點開始；(b)右眼從最左下點開始

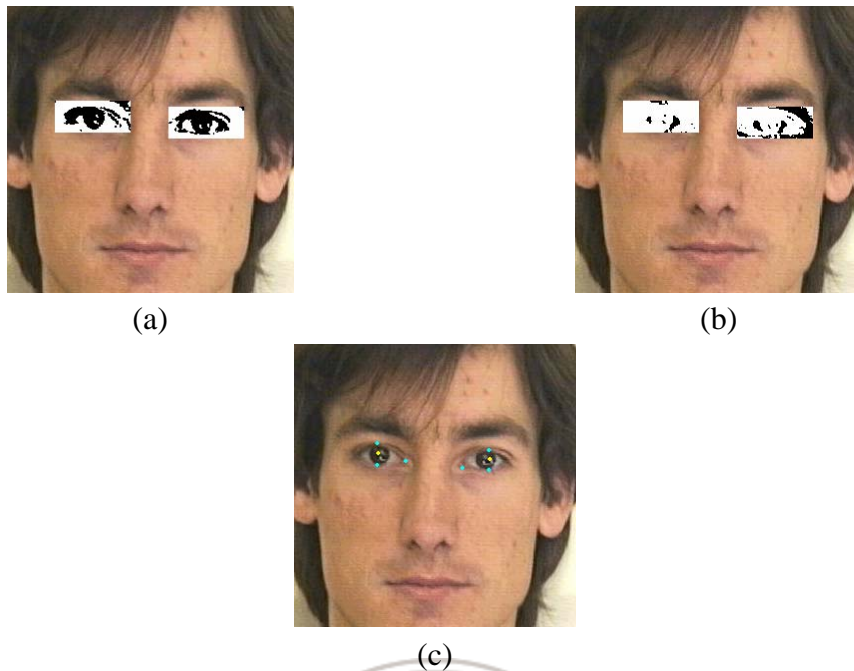


圖 2.10 眼睛特徵 (a) IOD 灰階二值化；(b) Sobel 邊緣偵測二值化；(c) 左、右眼之特徵點

2.2.3 眉毛特徵點

此部份是要擷取眉毛特徵點，左右眉毛各 2 點。我們定義以眼睛區域上邊界往上 2/9 瞳距當作眉毛區域中心點的垂直位置，再取瞳孔的水平位置為眉毛區域中心點的水平位置。

接下來，取眉毛區域中心點往上、下各 1/6 瞳距當作上、下邊界，往左、右各 1/3 瞳距當作左、右邊界，得到左右兩邊眉毛大約所在的區域。同樣的，對此區域做灰階統計直方圖，再針對此直方圖以動態方式取臨界值作 IOD 灰階值二值化， I 值為 0.5 時所對應的臨界值當作二值化臨界值，若小於此臨界值設為 1，反之則設為 0。圖 2.11(a) 為二值化結果。

以左眉來說，從邊界最右下點開始，同時以圖 2.9(a) 所示的方法開始搜尋，直到找到二值化為 1，此點我們可以當作左眉內點的特徵點。接著，以左眉毛區域中

心點為基準，由上往下找水平方向 3 個像素點內，直到找到二值化為 1 即為左眉上點特徵點，再由下往上找水平方向 3 個像素點內，直到找到二值化為 1 即為左眉下點特徵點。同樣的方法，亦可以找到右眉內點特徵點、右眉下點特徵點，而右眉內點特徵點搜尋方法如圖 2.9 (b)所示。圖 2.11(b)為眉毛特徵點的最終結果。

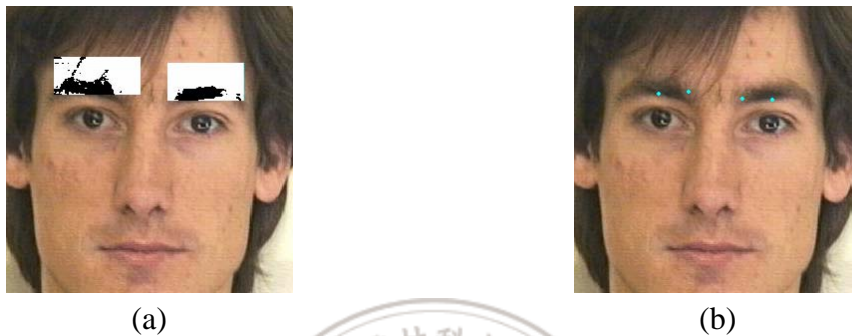


圖 2.11 眉毛特徵 (a) IOD 灰階二值化；(b) 左、右眉毛之特徵點

2.2.4 嘴巴特徵點

此部份要擷取 4 個嘴巴特徵點。我們定義以取左右兩眼瞳孔的水平平均位置為嘴巴區域中心點的水平位置，再以兩眼瞳孔的垂直平均位置往下加上 1.1 倍的瞳距為嘴巴區域中心點的垂直位置。接下來，取嘴巴區域中心點往上 0.15 倍瞳距當作上邊界、往下 0.47 倍瞳距當作下邊界，再往左、右各 0.47 倍瞳距當作左、右邊界，得到嘴巴大約所在的區域。將此區域轉換成灰階，同時對此區域作等化(equalization)及 IOD 灰階二值化處理。所謂的等化(equalization)是假設在一張影像中灰階 r_k 發生的機率可近似為

$$p_r(r_k) = \frac{n_k}{n} \quad k = 0, 1, 2, \dots, L-1 \quad (2.6)$$

其中 n 是影像中的像素總數， n_k 是有灰階 r_k 之像素的數目， L 是影像中可能灰階的總數。(2.5)式所給之轉換函數的離散形式為

$$S_k = T(r_k) = \sum_{j=0}^k p_r(r_j) = \sum_{j=0}^k \frac{n_j}{n} \quad k = 0, 1, 2, \dots, L-1 \quad (2.7)$$

因此，經由(2.6)式將輸入影像中灰階 r_k 的每一像素映射至輸出影像中灰階 s_k 的對應像素中獲得一個處理過的輸出影像。

圖 2.12 為嘴巴區域經過灰階轉換、等化處理及 IOD 灰階二值化處理的結果，這邊的 I 值設為 0.5；處理過後的二值化影像中，取影像中最左、最右邊二值化為 1 作為嘴巴左、右兩側的特徵點。



圖 2.12 嘴巴區域經過灰階轉換、等化處理及 IOD 灰階二值化處理的結果

接下來，要尋找嘴巴上下兩點的特徵點，利用先前所定義出嘴巴所在區域內，把此區域由 RGB 轉換成 HIS 的色彩空間(如圖 2.13 所示)，也是主要利用 HIS 的 H(Hue)成分，取小於 25 度或大於 350 度當作臨界值來做二值化，因為此範圍內的 Hue 較接近嘴唇顏色，介於此範圍設為 1，否則就設為 0，圖 2.14 為二值化結果。以上述所找到嘴巴左右兩點之間的所計算出中點垂直位置為基準，從此二值化影像中嘴巴區域的上邊界開始由上往下、從嘴巴區域的下邊界開始由下往上找，找到兩點二值化為 1 的點即是我們所要的嘴巴上下兩點特徵點。圖 2.15 為嘴巴特徵點的最終結果。

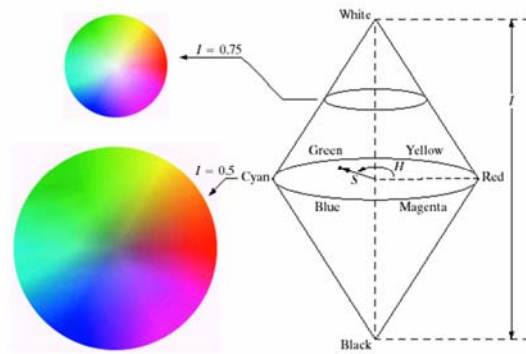


圖 2.13 HIS 色彩空間之圓型色彩平面



圖 2.14 嘴巴區域之 H(Hue)空間經過二值化的結果

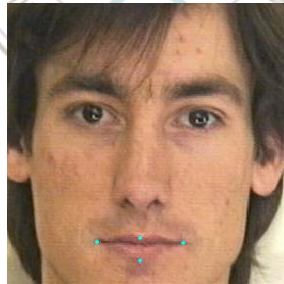


圖 2.15 嘴巴特徵點

2.3 表情特徵向量

將 2.2 節中所得到的 14 個特徵點位置，計算一組特徵距離，將這些特徵距離正規化(normalize)成特徵值，再將有表情影像的正規化距離減去無表情正規化特徵距離，而得到一組特徵向量。圖 2.16 為表情特徵向量計算流程圖。

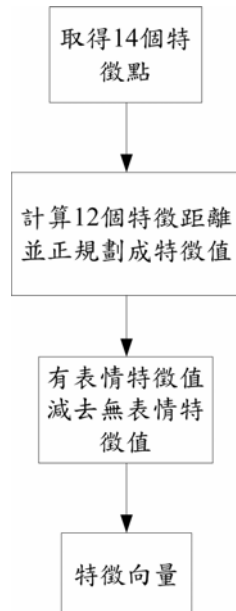


圖 2.16 表情特徵向量計算流程圖

我們對各個特徵點加以定義命名，如圖 2.17(a)所示。其中點 1、6 為左右眉毛內側，點 2、7 為左右眉毛中央下側，點 3、8 為左右眼內側，點 4、9 為左右眼中央上側，點 5、10 為左右眼中央下側，點 11、13 分別為嘴巴上下兩側，點 12、14 分別為嘴巴左右兩嘴角。而點和點的距離定義如圖 2.17(b)所示，各點距離的定義如下：

$$\begin{aligned}
 n &= \text{dist}(3,8) & a &= \text{dist}(2,4) & b &= \text{dist}(1,3) & c &= \text{dist}(1,6) & d &= \text{dist}(6,8) & \\
 e &= \text{dist}(7,9) & f &= \text{dist}(4,5) & g &= \text{dist}(9,10) & h &= \text{dist}(5,12) & & & \\
 I &= \text{dist}(\text{mid}(3,8),11) & j &= \text{dist}(10,14) & k &= \text{dist}(11,13) & l &= \text{dist}(12,14) & & & (2.8)
 \end{aligned}$$

其中 $\text{dist}(x, y)$ 為點 (x, y) 的歐幾里德運算(Euclidean norm)， $\text{mid}(x, y)$ 為點 (x, y) 的中點。

由於每個人的五官所算出來的特徵距離，或是偵測出來的人臉範圍，即使是無表情的狀態下，皆不會完全相同，所以需將這些特徵距離作正規化；在人臉上

兩眼內角的距離是不會變動的，因此以兩眼內角距離(n)當作一個不變量來執行正規化的動作，將得到的特徵距離作正規化後得到特徵值，計算如下：

$$a' = a/n \quad b' = b/n \quad c' = c/n \quad d' = d/n \quad e' = e/n \quad f' = f/n \quad g' = g/n \quad h' = h/n \quad i' = i/n \quad j' = j/n \quad k' = k/n \quad l' = l/n \quad (2.9)$$

最後把這些正規化後的特徵值減去無表情正規化後的特徵值(non)，得到一組特徵向量(off)，計算如下：

$$\begin{aligned} off_a &= a' - non(a') & off_b &= b' - non(b') & off_c &= c' - non(c') \\ off_e &= e' - non(e') & off_f &= f' - non(f') & off_g &= g' - non(g') \\ off_h &= h' - non(h') & off_i &= i' - non(i') & off_j &= j' - non(j') \\ off_k &= k' - non(k') & off_l &= l' - non(l') \end{aligned} \quad (2.10)$$

其中 $non(*) = \frac{non(*)}{non(n)}$ ，為無表情的特徵值。

而特徵向量我們可以改寫成：

$$F = [off_a, off_b, \dots, off_l]^T \quad (2.11)$$

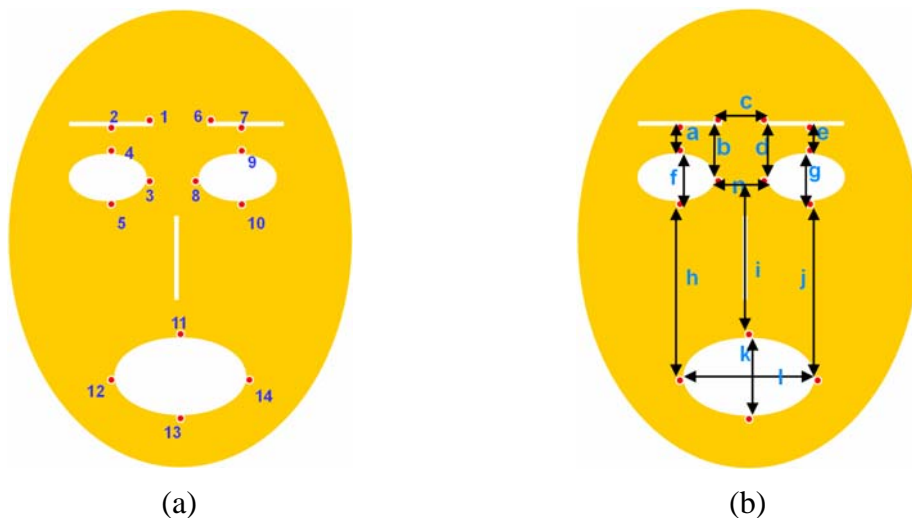


圖 2.17 臉部特徵點及特徵距離 (a) 各點之命名定義；(b) 特徵點距離定義

第三章

生理訊號特徵

由於我們不能自主地控制生理訊號的變動，所以若要了解內心的情緒，觀察生理訊號是一個很好的指標，它將真實且直接的反映出你內在的感覺。本章節將介紹常見生理訊號的種類，以及如何對所需的生理訊號做前處理，最後再決定出生理訊號特徵向量。

3.1 生理訊號

本小節將介紹一些常見的生理訊號種類，如心電圖、肌電、皮膚電位反應、皮膚溫度、血流量、呼吸以及心率。

心電圖

心搏的一個週期，是由右心房上竇房結(SA node)發出電脈衝以漸進波的方式傳遞至左、右心房，造成左右心房的收縮(圖 3.1 P 部分)，電脈衝傳達房室結(AV node)後，接著電脈衝便藉由傳遞纖維將電脈衝傳遞(圖 3.1 Q 部分)至左右心室造成左右心室收縮(圖 3.1 R 部分)，在一連串的电活動之後心臟暫時靜止，心室等待再極化以恢復帶負電狀態(圖 3.1 T 部分)完成一次心搏。心臟本身的電位變化會經過心臟周圍的導電組織與體液反映到身體表面，所謂的心電圖(Electrocardiogram, ECG)是運用微電極技術紀錄心臟微小電脈衝的變化所產生的心肌細胞內外電位差，由儀器放大電活動訊號描繪下的圖形，藉以瞭解心臟是否正常地運作。

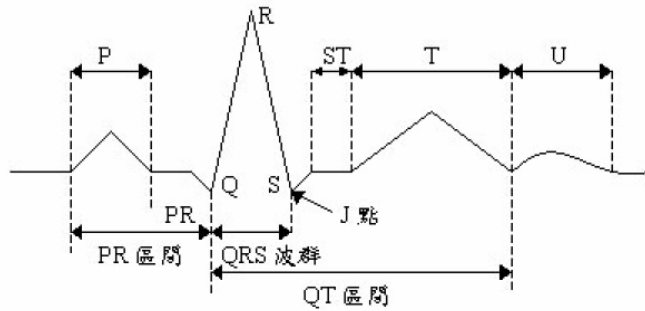


圖 3.1 PQRST 波形型態

肌電訊號

EMG(electromyography)訊號即是肌電訊號，它是在肌肉收縮時，肌肉兩端的電位差所產生出來的訊號，依個人肌肉在動作時收縮強度的多寡，造成不同的電位差異，得到的訊號也因此有不同的強度及特徵。產生電位的肌肉稱為肌纖維群，當活化的肌纖維群的動作電位變化經過肌肉、皮膚組織的空間傳到皮膚表面，此時皮膚表面的電極所收集到的電壓訊號稱之為表面肌電圖(surface EMG)。圖3.2為肌電圖範例。

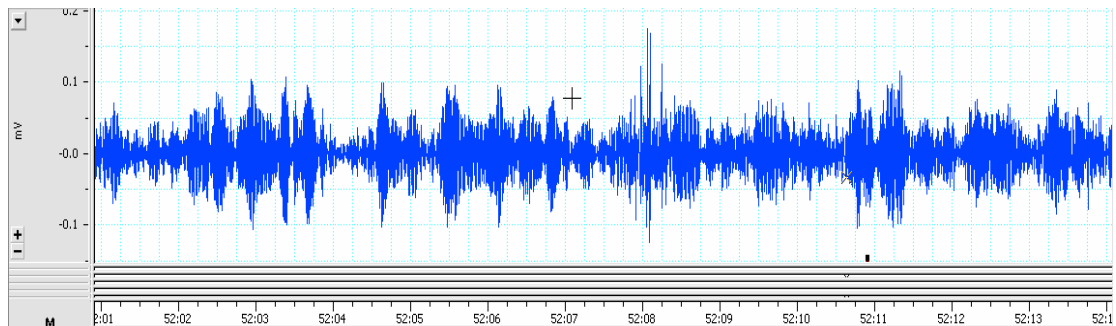


圖 3.2 肌電圖範例

皮膚電位反應

皮膚電位反應 (galvanic skin response, GSR)，又稱為皮膚電鍍反應 (electrodermal response, EDR)、精神電位反射 (psychogalvanic reflex, PGR)，或是皮膚傳導反應 (skin conductance response, SCR)。這指的是皮膚上的電子傳導能力，通常當皮膚上的汗腺分泌時，就會影響皮膚電位的變動，也發生在交感神經系統的活化，我們可以利用偵測器來測量來得到其數值。圖 3.3 是皮膚電位的範例圖。

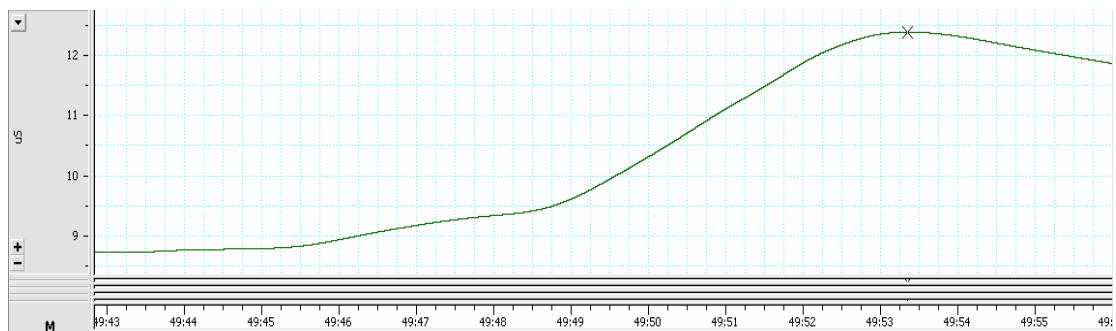


圖 3.3 皮膚電位反應

皮膚溫度

在身體皮膚表面上溫度敬稱之為皮膚溫度 (skin temperature)，皮膚溫度主要是來自於皮膚上的血液的溫度。當情緒或身體狀況有改變跡象，皮膚溫度就會有些許變動，通常皮膚溫度是緩慢且平順的變動，而不會是劇烈的。

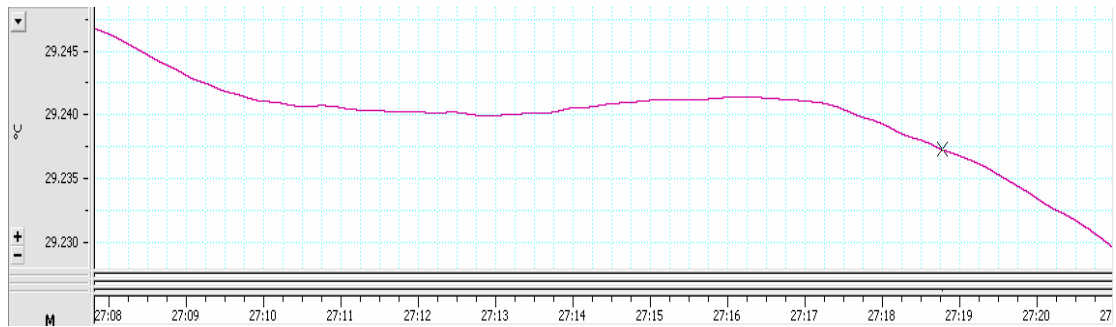


圖 3.4 皮膚溫度

血流量

血流量(blood volume pulse)即單位時間內流過血管某一截面的血量,亦稱容積速度,以每分鐘的毫升數或公升數來表示。血流量的大小主要取決於兩個因素,即血管兩端的壓力差和血管對血流的阻力。而血液在血管中是靠心臟搏動之間來流動。

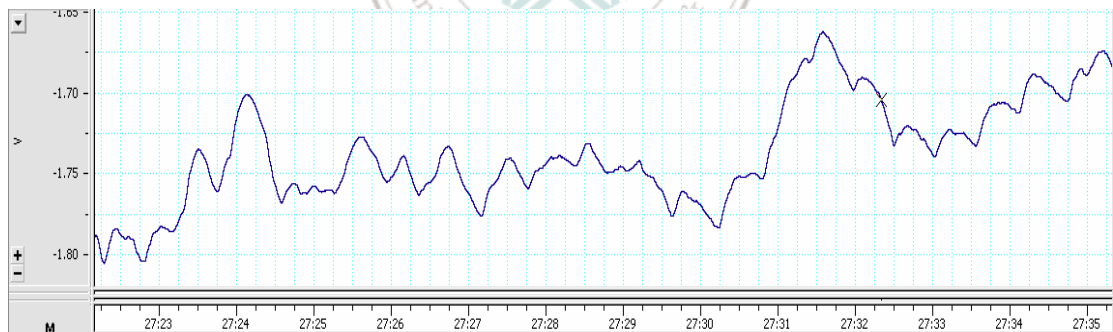


圖 3.5 血流量

呼吸

呼吸(respiration)通常可以根據胸腔擴張來觀察,所以呼吸深淺可以測量胸腔擴張的大小程度來得到。由於每個人呼吸訊號樣本的有很大差異性,所以並沒有

標準的波形圖。圖 3.6 為原始呼吸訊號樣本。

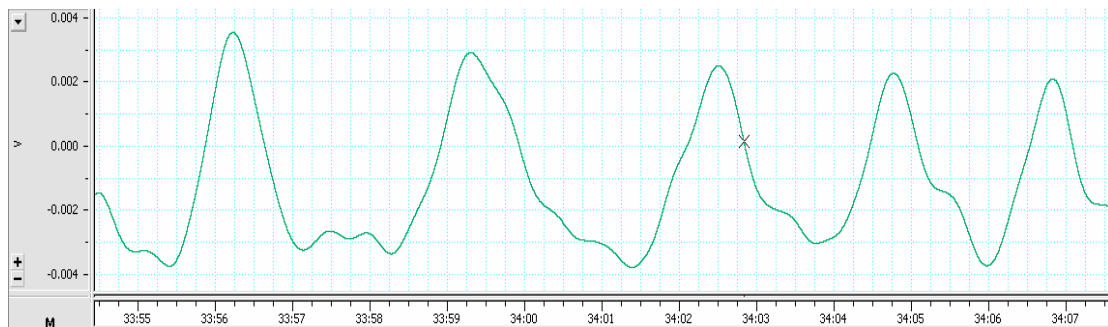


圖 3.6 原始呼吸訊號

心率

心率(heart rate)，即心臟搏動的頻率，通常以每分鐘心跳的次數來表示。

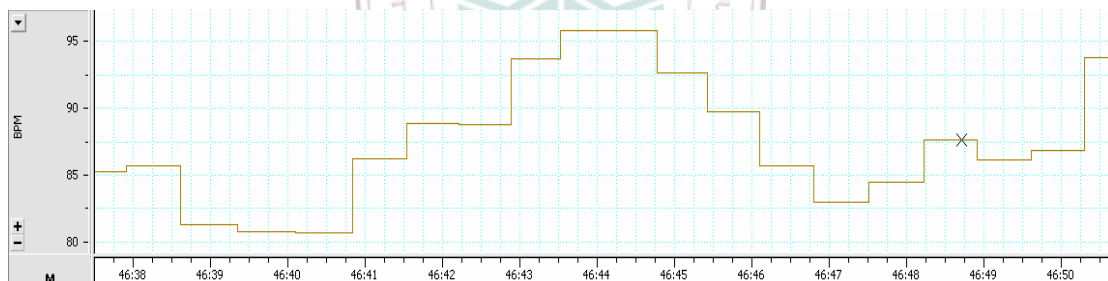


圖 3.7 心率

3.2 生理訊號前處理及特徵向量

對於訊號收集方面，我們擷取三種生理訊號：皮膚電位反應(GSR)、指溫(Temp)和心率(HR)，對這三種生理訊號前處理，最後得到一組生理訊號特徵向量，其流程如圖 3.8 所示。

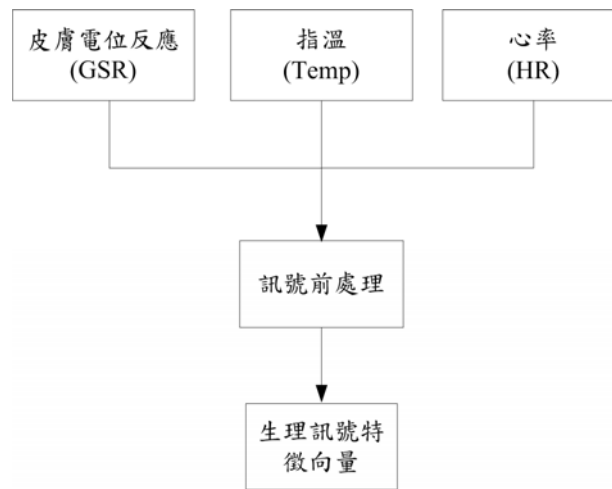


圖 3.8 生理訊號處理流程圖

對於每個人的生理訊號，其影響程度會有些不同，所以必須先經過正規化的動作。假設 δ 為生理訊號種類({GSR、Temp、HR})， S_m^δ 為在 δ 中的第 m 個值，則正規化可以表示成：

$$\tilde{S}_m^\delta = \frac{S_m^\delta - \min(S^\delta)}{\max(S^\delta) - \min(S^\delta)} \quad (3.1)$$

其中 S^δ 為生理訊號的資料集合； \tilde{S}_m^δ 為在 δ 裡第 m 個正規化的值； $\max(S^\delta)$ 和 $\min(S^\delta)$ 分別代表在 S^δ 中的最大值和最小值。在這裡正規化的方式必須以同一個人獨立來計算，也就是說，不能和其他人一起作正規化，以免發生正規化後資料差異性過

大。

接下來，將決定在每個情緒之 \tilde{S}_m^δ 的每個子段的平均。設 μ_k^δ 為在 S^δ 中第 k 個子段的平均值，並且由下式把每個 \tilde{S}_m^δ 中的子段計算出來：

$$\mu_k^\delta = \frac{1}{T} \sum_{m=(k-1)*T+1}^{k*T} \tilde{S}_m^\delta \quad (3.2)$$

其中 T 是每個子段的資料總數。

最後，我們可以決定由 μ_k^δ 所組成的生理訊號特徵向量 P_k ，以(3.3)式來表示：

$$P_k = [\mu_k^{GSR}, \mu_k^{Temp}, \mu_k^{HR}]^T \quad (3.3)$$



第四章

表情分類與辨識

在本章節中，提出一組以半徑基底函數(Radial Basis Function, RBF)類神經網路，分類表情特徵向量及生理訊號特徵向量藉以辨識情緒種類。

4.1 半徑基底函數(RBF)類神經網路

半徑基底函數(Radial Basis Function, RBF)屬於基本的前饋式類神經網路 [12,13]，此網路的呈現是藉由訓練的資料，這些資料包含了輸入空間的向量和網路期望的輸出。網路架構如圖 4.1 所示，包含了輸入層、隱藏層及輸出層。此 RBF 類神經網路的輸出模型如下表示：

$$\begin{aligned} f_i(\mathbf{x}) &= \sum_{k=1}^N w_{ik} \phi_k(\mathbf{x}, c_k) \\ &= \sum_{k=1}^N w_{ik} \phi_k(\|\mathbf{x} - c_k\|_2) \end{aligned} \quad (4.1)$$

其中 $\mathbf{x} \in R^{n \times 1}$ 是輸入向量， ϕ_k 是 RBF 類神經網路的基底函數， w_{ik} 代表第 i 個輸出層神經元與第 k 個隱藏層神經元之間連接的權重， $\|\cdot\|_2$ 表示歐式距離 (Euclidean norm)， N 是代表隱藏層的神經元個數， $c_k \in R^{n \times 1}$ 是表示 RBF 類神經網路輸入向量空間的中心點。若 RBF 類神經網路中的基底函數為高斯(Gaussian)函數，則 (4.1) 式可改寫成如下：

$$\begin{aligned}
f_i(\mathbf{x}) &= \sum_{k=1}^N w_{ik} \phi_k(\|\mathbf{x} - \mathbf{c}_k\|_2) \\
&= \sum_{k=1}^N w_{ik} \exp(-\|\mathbf{x} - \mathbf{c}_k\|_2^2 / \sigma_k^2)
\end{aligned} \tag{4.2}$$

其中 σ_k 是高斯函數的寬度，其定義如下所示：

$$\sigma_k = \frac{d_{\max}}{\sqrt{N}} \tag{4.3}$$

其中 d_{\max} 是所有中心點之間的最大 Euclidean 距離。

因此，使用一種基於監督式學習演算法的序率坡降法(stochastic gradient method)，此誤差花費函數(cost function)被定義如下：

$$\begin{aligned}
J(n) &= \frac{1}{2} |e(n)|^2 \\
&= \frac{1}{2} [y_{\text{desired}}(n) - \sum_{k=1}^N w_k(n) \phi\{\mathbf{x}(n), \mathbf{c}_k(n)\}]^2 \\
&= \frac{1}{2} [y_{\text{desired}}(n) - \sum_{k=1}^N w_k(n) \exp(-\frac{\|\mathbf{x}(n) - \mathbf{c}_k(n)\|_2^2}{\sigma_k^2(n)})]^2
\end{aligned} \tag{4.4}$$

此網路的參數更新方程式由下列式子所更新：

$$\begin{aligned}
\mathbf{w}(n+1) &= \mathbf{w}(n) - \mu_w \left. \frac{\partial J(n)}{\partial \mathbf{w}} \right|_{\mathbf{w}=\mathbf{w}(n)} \\
&= \mathbf{w}(n) + \mu_w e(n) \boldsymbol{\psi}(n)
\end{aligned} \tag{4.5}$$

$$\begin{aligned}
\mathbf{c}_k(n+1) &= \mathbf{c}_k(n) - \mu_c \left. \frac{\partial J(n)}{\partial \mathbf{c}_k} \right|_{\mathbf{c}_k=\mathbf{c}_k(n)} \\
&= \mathbf{c}_k(n) + \mu_c \frac{e(n) w_k(n)}{\sigma_k^2(n)} \phi\{\mathbf{x}(n), \mathbf{c}_k(n), \sigma_k\} [\mathbf{x}(n) - \mathbf{c}_k(n)]
\end{aligned} \tag{4.6}$$

$$\begin{aligned}
\sigma_k(n+1) &= \sigma_k(n) - \mu_\sigma \left. \frac{\partial J(n)}{\partial \sigma_k} \right|_{\sigma_k=\sigma_k(n)} \\
&= \sigma_k(n) + \mu_\sigma \frac{e(n) w_k(n)}{\sigma_k^3(n)} \phi\{\mathbf{x}(n), \mathbf{c}_k(n), \sigma_k\} \|\mathbf{x}(n) - \mathbf{c}_k(n)\|^2
\end{aligned} \tag{4.7}$$

其中，

$\psi(n) = [\phi\{x(n), c_1(n), \sigma_1\}, \dots, \phi\{x(n), c_N(n), \sigma_N\}]^T$ ， $e(n) = y_{desired}(n) - y_i(n)$ ， $y_{desired}(n)$ 是期望的網路輸出， μ_w, μ_c , and μ_σ 為合適的學習速率常數。

整個 RBF 類神經網路演算法步驟可以歸納為如下所示：

- Step1: 先從訓練的資料中隨機選取當成 RBF 的中心點。
- Step2: 根據(4.3)式計算類神經網路的中心點初始寬度。
- Step3: 隨機地去初始化學習速率和輸出層的網路權重。
- Step4: 利用 (4.2) 式計算出網路的輸出。
- Step5: 根據 (4.5)、(4.6)、(4.7)式更新網路的常數。
- Step6: 若網路收斂則停止計算；否則回到 Step4。

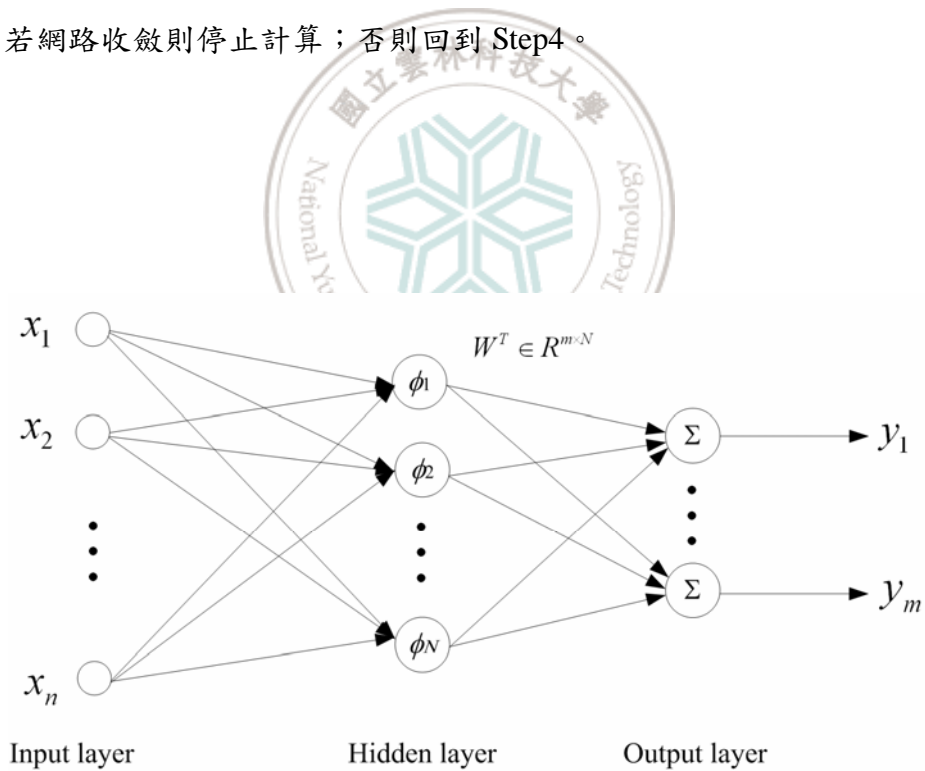


圖 4.1 半徑基底函數(RBF)類神經網路的架構圖

4.2 應用半徑基底函數類神經網路於分類特徵向量與辨識表情

本論文提出三組獨立的半徑基底函數類神經網路(radial basis function neural network, RBFNN)，使用表情特徵向量以及生理訊號特徵向量分別輸入個別的RBFNN(RBFNN1 及 RBFNN2)進行訓練與測試，經過 RBFNN1 及 RBFNN2 的訓練完成後，首先將訓練表情特徵向量以及生理訊號特徵向量分別輸入到 RBFNN1 及 RBFNN2 來測試，經過測試後把個別輸出結果組成一組向量，再把此向量輸入到 RBFNN3 來訓練。完成訓練後，將待測的表情特徵向量以及生理訊號特徵向量分別輸入 RBFNN1 及 RBFNN2，得到一組向量後再輸入 RBFNN3，最後得到最終輸出的分類結果。圖 4.2 為表情辨識之流程圖。

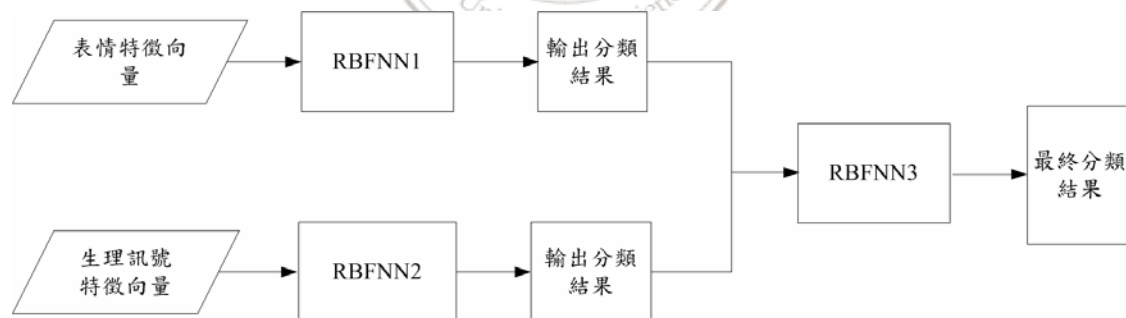


圖 4.2 表情辨識之流程

半徑基底函數類神經網路屬於監督式學習演算法，包含訓練階段和測試階段。圖 4.3 為 RBFNN 架構圖，輸出情緒類別有害怕(fear)、愛(love)、快樂(joy)和驚訝(surprise)，而 RBFNN1 之輸入向量為表情特徵向量(F)，RBFNN2 之輸入向量

為生理訊號特徵向量(P_k)。圖 4.4 和圖 4.5 分別為 RBFNN1 和 RBFNN2 的網路架構圖。在 RBFNN1 訓練階段中，輸入表情特徵向量(F)當作輸入向量，中心點的選擇是從訓練的資料中隨機方法選取獲得，選取之後我們接著計算 d_{\max} 和 σ_k 。網路的初始權重也是隨機選取。待網路收斂後，我們可以使用此網路去檢測表情的種類；相同地，在 RBFNN2 訓練階段中，輸入生理訊號特徵向量(P_k)當作輸入向量，待網路收斂後，我們可以使用此網路去檢測生理訊號之情緒的種類。

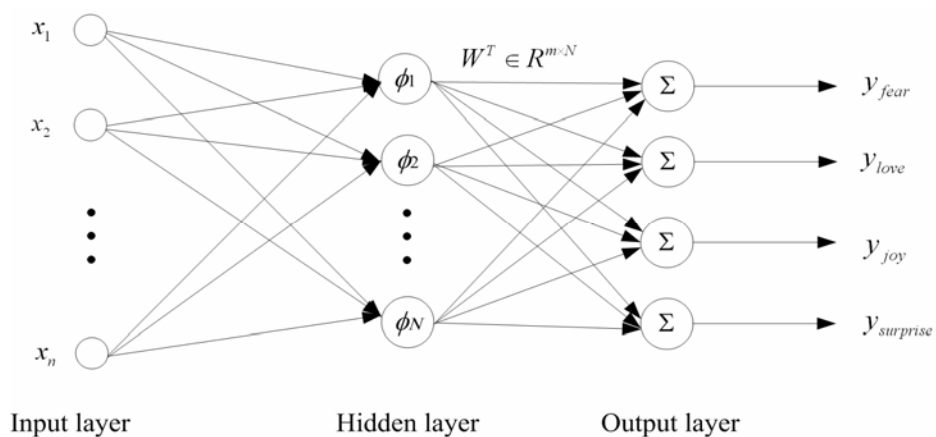


圖 4.3 兩組 RBF 類神經網路分類與辨識流程

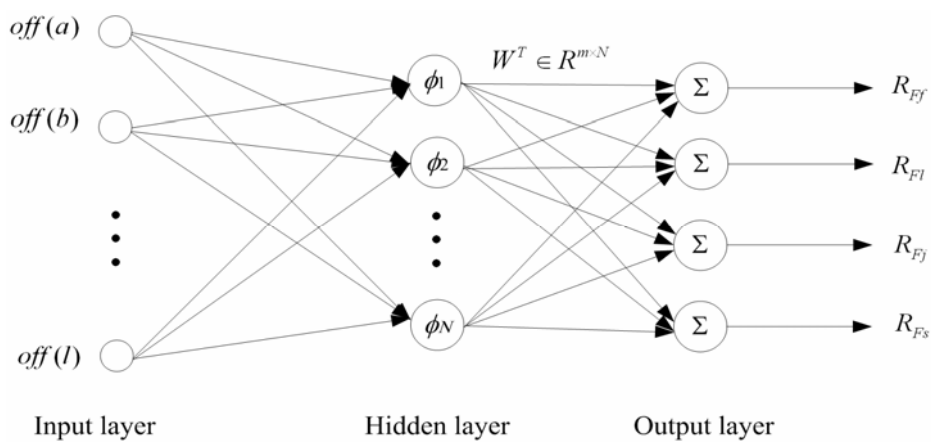


圖 4.4 RBFNN1 之網路架構圖

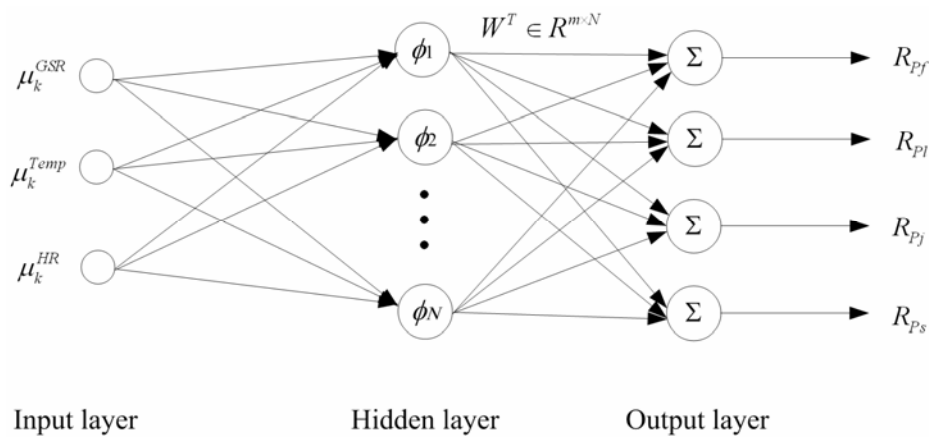


圖 4.5 RBFNN2 之網路架構圖

在 RBFNN1 及 RBFNN2 訓練完成後，分別將訓練的表情特徵向量和生理訊號特徵向量(F 和 P_k)輸入 RBFNN1 及 RBFNN2，分別經過 RBFNN1 及 RBFNN2 作測試，輸出結果 RBFNN1 得到輸出向量 R_F ($R_F = [R_{Ff}, R_{Fl}, R_{Fj}, R_{Fs}]$)，RBFNN2 得到輸出向量 R_P ($R_P = [R_{Pf}, R_{Pl}, R_{Pj}, R_{Ps}]$)。將此兩組輸出向量組成一組向量 R_R ($R_R = [R_{Ff}, R_{Fl}, R_{Fj}, R_{Fs}, R_{Pf}, R_{Pl}, R_{Pj}, R_{Ps}]$)，輸入到 RBFNN3 來作訓練，此目的是用來調整最終輸出分類的權重，以得到最佳的輸出分類結果。圖 4.6 為 RBFNN3 的網路架構圖。

在測試階段中，將待測的表情特徵向量(F)和生理訊號特徵向量和(P_k)分別輸入到 RBFNN1 及 RBFNN2，得到一組測試後的輸出向量 R_F 和輸出向量 R_P ， R_F 和 R_P 再組成 R_R ，把 R_R 輸入到 RBFNN3 來作測試，最後的輸出結果向量 R_R ($R_R = [R_{Rf}, R_{Rl}, R_{Rj}, R_{Rs}]$)，取最大值的元素所屬類別即是最終輸出辨識結果。

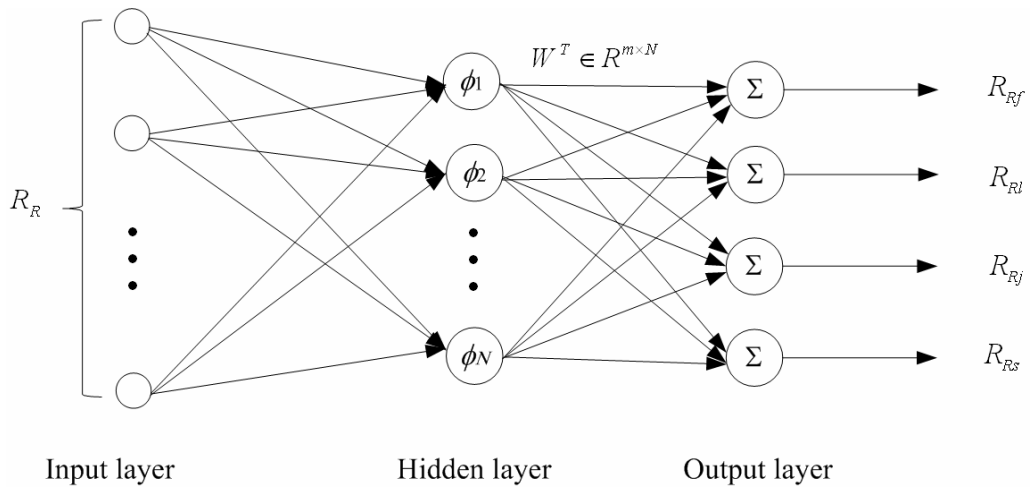


圖 4.6 RBFNN3 之網路架構圖



第五章

實驗結果與討論

5.1 資料收集

我們設計一套情緒激發(emotion exciting)實驗來收集受試者的臉部表情影像和生理訊號資料。圖 5.1 為整個情緒激發實驗的概觀。首先我們在受試者的身上裝設三種生理訊號偵測器，收集皮膚電位反應(galvanic skin response, GSR)、指溫(finger temperature, Temp)及心率(heart rate, HR)等生理訊號資料，同時在受試者面前放置一台攝影機來拍攝錄影受試者的臉部影像。圖 5.2 為整個實驗環境的全景。為了使受試者的情緒激發出來，我們準備了一系列可激發情緒之影片片段包含害怕(fear)、愛(love)、快樂(joy)和驚訝(surprise)從螢幕上播放來給受試者觀看，範例如圖 5.3 所示，藉由這些影片讓受試者在情緒上激發出害怕(fear)、愛(love)、快樂(joy)和驚訝(surprise)，表 5.1 為影片片段的內容描述。當情緒被激發的時候，受試者的臉部表情影像和生理訊號會各自的被記錄下來。圖 5.4 為所收集表情影像資料之範例，圖 5.5 為所收集生理訊號在不同情緒之範例。

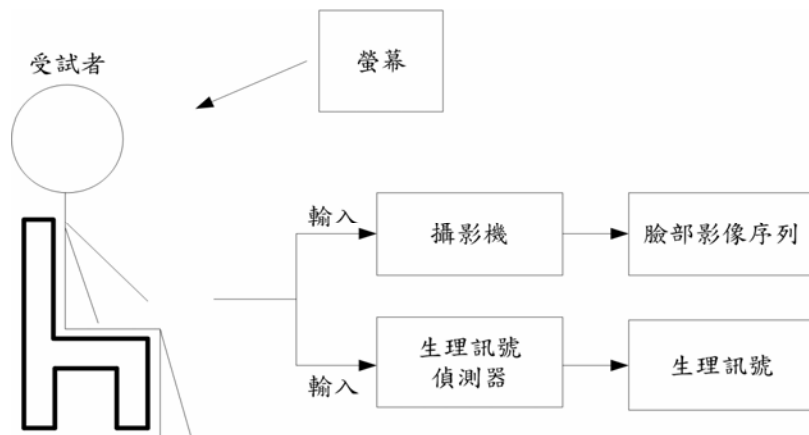


圖 5.1 情緒激發實驗之示意圖



圖 5.2 情緒激發實驗環境之全景

表 5.1

影片片段之內容描述

<p>害怕(fear)</p>	<p>取自電影“嚇破膽”之片段，內容是兇手用一條鋼絲把在船上甲板的人群全部給殺了，人群全部支離破碎，整個事件過程相當的血腥驚悚。</p>
<p>愛(love)</p>	<p>取自“老師的啟示”，故事描述是男孩從小家境清寒，大家不喜歡他且被同學排擠，因而個性孤僻，直到他的老師了解男孩的狀況後，對他開始輔導、關心，陪他唸書等。老師在男孩身上也學到教育的意義。最後男孩長大了，日子也過的不錯，回想起來這位老師是男孩一生中遇到最好的老師，也是生命中的貴人。</p>
<p>快樂(joy)</p>	<p>逗趣歡樂的糗事片段，如小孩走路時突然滑倒、搶新娘拋花而掉入水裡和撞到牆壁、被人歡呼拋起來而撞到天花板和燈柱等。</p>
<p>驚訝(surprise)</p>	<p>魔術師表演一段鱷魚商標的魔術，把商標變大、商標移動到手機內、商標變成一隻小鱷魚等，最後再變回商標。</p>



(a) 害怕



(b) 愛



(c) 快樂



(d) 驚訝

圖 5.3 情緒激發影片之片段截圖



(a)



(b)



(c)



(d)

圖 5.4 表情影像資料之範例 (a)害怕(fear)；(b)愛(love)；(c)快樂(joy)；(d)驚訝(surprise)

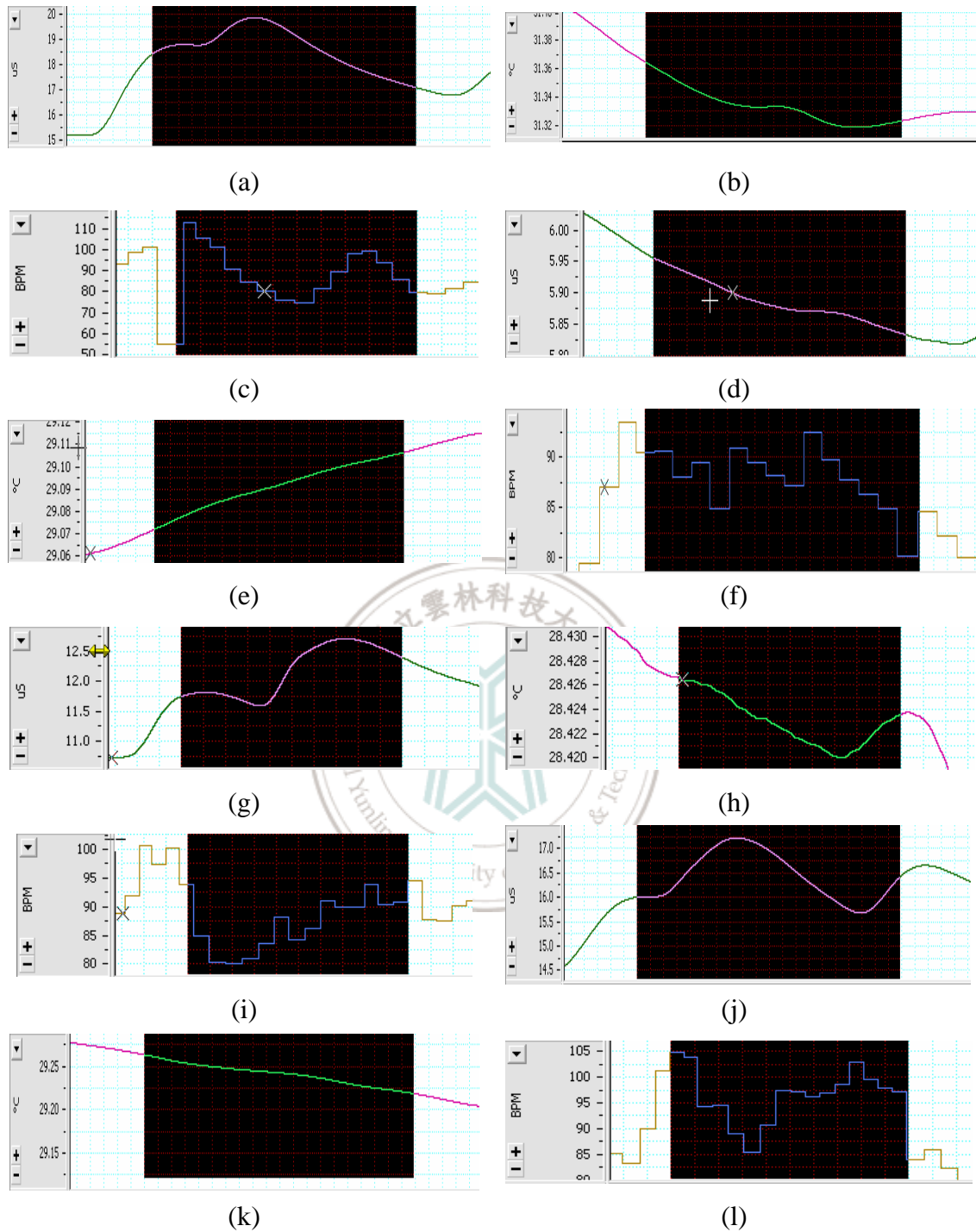


圖 5.5 生理訊號在不同情緒之範例(黑色區域)，(a-c)分別為害怕之 GSR、Temp、HR 的訊號；(d-f)分別為愛之 GSR、Temp、HR 的訊號；(g-i)分別為快樂之 GSR、Temp、HR 的訊號；(j-l)分別為驚訝之 GSR、Temp、HR 的訊號

為了評估本論文提出的半徑基底函數類神經網路之效能，本論文也將測試 JAFFE (Japanese Female Facial Expression)表情資料庫[13]，其中有 10 個日本女性共 149 張影像(圖 5.6)，包含了害怕、無表情、快樂、驚訝、生氣和悲傷，表 5.2 為表情範例張數。

表 5.2 JAFFE 各表情張數

表情	害怕	無表情	快樂	驚訝	生氣	悲傷
張數	28	30	31	30	30	29





圖 5.6 JAFFE 表情範例

5.2 情緒激發實驗之測試

在本測試中，從設計的情緒激發實驗中 6 位受試者上，收集了 240 張臉部影像和其相關的生理訊號，包含了四種情緒，害怕(fear)、愛(love)、快樂(joy)和驚訝(surprise)。對於某種情緒來說，在訊號取樣上我們觀察受試者臉部有表情的時候，或是觀察生理訊號比較其他受試者有著相同趨勢的變動，則取出這部份的訊號。我們在生理訊號取樣時間的長度為 10 秒鐘，每 0.05 秒取 1 次，所以 6 個人在 4 種情緒中的 10 秒鐘的訊號一共有 4800 筆原始資料樣本，而經過生理訊號前處理後的資料樣本為 240 筆資料；在此生理訊號的取樣率中，生理訊號是可能在短時間變動的幅度很大，因此在每秒的訊號中，取樣率定為每秒 20 筆資料，如此能減少資料的計算量又能足以代表訊號波動的變化。在收集的臉部影像中，我們取 10 秒鐘長度的影像資料，每秒鐘取 1 張，因此 10 秒鐘共有 10 張影像；在此臉部影像的取樣率中，因為我們從觀察中所取樣臉部影像，受試者臉部表情變化皆是有連續性的，每秒的表情大都是有一致性的，若取樣率設太密集，可能會增加計算量，因此為了增加計算速度，我們一秒取一張影像作為臉部影像資料樣本且足以代表在某情緒的臉部影像。在臉部影像部份，我們取奇數秒的臉部表情樣本當作訓練樣本輸入到 RBFNN1 來訓練，偶數秒的臉部表情樣本當作測試樣本輸入到 RBFNN1 來測試。在生理訊號部份，取奇數秒的生理訊號樣本當作訓練樣本輸入到 RBFNN2 來訓練，偶數秒的生理訊號樣本當作測試樣本輸入到 RBFNN2 來測試。

在 RBFNN1 及 RBFNN2 的訓練中，因為高斯函數呈現常態分布，所以隱藏層的半徑基底函數採用高斯函數。網路權重的學習速率 μ_w 我們設為 0.1， μ_c 設為 0.09， μ_o 設為 0.05，收斂條件為滿足迭代次數 5000 次或是誤差花費函數小於 10^{-5} ，網路的權重是隨機選取且介於 0.01~0.0001 之間；RBFNN1、RBFNN2 及 RBFNN3 之隱藏層中心點設為 100 顆。在測試部分，我們分成三個部份，對表情和生理訊號

特徵作個別測試 RBFNN1 及 RBFNN2，最後結合兩組特徵測試 RBFNN3。表 5.3 為 RBFNN1 測試之表情正確率，表 5.4 為 RBFNN2 之生理訊號正確率，表 5.5 為 RBFNN3 之結合表情和生理訊號之正確率。

表 5.3 表情正確率

	害怕	愛	快樂	驚訝	正確率%
害怕(30)	20	10	0	0	66.67%
愛 (30)	0	30	0	0	100%
快樂(30)	0	0	30	0	100%
驚訝(30)	0	1	1	28	93%
平均					90%

表 5.4 生理訊號正確率

	害怕	愛	快樂	驚訝	正確率%
害怕(30)	25	0	0	5	83%
愛 (30)	0	29	1	0	96.67%
快樂(30)	1	0	28	1	93%
驚訝(30)	2	0	4	24	80%
平均					88.33%

表 5.5 結合表情和生理訊號之正確率

	害怕	愛	快樂	驚訝	正確率%
害怕(30)	24	0	0	6	80%
愛 (30)	0	30	0	0	100%
快樂(30)	0	0	30	0	100%
驚訝(30)	0	0	0	30	100%
平均					95%

我們可以發現，雖然臉部表情結合生理訊號後，害怕表情的辨識率比只用生理訊號的辨識率稍微降低，不過在其他表情的辨識率上仍有良好的表現，而最後的整體辨識率可達到 95%，結果顯示結合生理訊號的表情辨識率有不錯的效能。比較表情的辨識結果，我們可以歸納出幾點問題因素：觀察受試者的生理訊號有波動的情況，但在同時並沒有做出任何臉部表情。此外，設計影片應該是要激發某種情緒，但是受試者卻作出另一種表情，舉例來說，應該激發驚訝的情緒卻作出歡樂的表情，在害怕的情境中卻作出驚訝的表情，因此在臉部表情辨識率上會顯的不理想。在激發情緒實驗中，雖然受試者觀看相同的激發情緒片段，所激發的情緒程度可能很大，也可能很小，而影響生理訊號的波動也可能有很大的差異。因此，這些無法預期的因素就有可能影響到辨識率。假如受試者的表情或是生理訊號波動的趨勢能有一致性，其辨識率會更理想。

5.3 表情資料庫之測試

對於表情辨識系統中，為了測試半徑基底函數類神經網路效能，我們以 JAFFE 和自拍影像資料庫來做訓練及測試。訓練及測試方法為：以 JAFFE 資料庫為主，取多人影像來做訓練及測試半徑基底函數類神經網路，利用 Cross-validation 以及取一半訓練樣本，另一半為測試樣本。

5.3.1 JAFFE 多人影像測試

在 JAFFE 多人測試部分，在這部份分為兩種測試方式；第一種測試方式為 Cross-validation，以 JAFFE 資料庫為主，取資料庫任九人為訓練樣本，而剩下的一人作為測試樣本，每個人皆測試一次。表情分類分別為五類(生氣、無表情、快樂、驚訝和悲傷)及四類(生氣、無表情、快樂和驚訝)，此外對於兩組分類表情訂

立不同的半徑基底函數類神經網路的誤差花費函數的更新之收斂條件(閾值)，以及調整網路之隱藏層個數來測試，其四類表情辨識結果分別為如表 5.6。第二種測試方式為在 JAFFE 資料庫中隨機取出一半影像為訓練樣本，另一半為測試樣本，表情分類也分別為五類及四類，並且對於兩組分類表情訂立不同的半徑基底函數類神經網路的誤差花費函數之收斂條件(閾值)，以及調整網路之隱藏層個數來測試，其辨識結果分別為如表 5.7 及表 5.8。

表 5.6 JAFFE 之 Cross-validation 四類表情辨識率

閾值 \ 隱藏層個數	0.00001	0.000001
140	89.08%	89.08%
150	82.35%	87.60%
160	86.54%	89.08%
170	84.87%	88.43%
180	85.71%	88.43%

表 5.7 JAFFE 之一半訓練、一半測試五類表情辨識率

閾值 \ 隱藏層個數	0.0001	0.00001	0.000001
90	82.43%	82.43%	81.25%
100	82.43%	82.43%	83.78%
110	82.43%	82.43%	81.25%
120	83.78%	81.07%	81.25%
130	85.13%	83.78%	81.25%

表 5.8 JAFFE 之一半訓練、一半測試四類表情辨識率

閥值 隱藏層個數	0.0001	0.00001	0.000001
80	91.66%	90.00%	91.67%
90	91.66%	90.00%	91.67%
100	91.66%	90.00%	93.33%
110	91.66%	86.67%	91.67%
120	91.66%	85.00%	91.67%

5.4 實驗比較

為了證明系統的有效性，我們以 JAFFE 資料庫為比較基準。取多人表情影像如生氣、無表情、快樂和驚訝這四種表情範例，共 121 張，來做訓練及測試半徑基底函數類神經網路，利用 Cross-validation 以及取一半訓練樣本，另一半為測試樣本。實驗結果將與[15]來進行比較，[15]所使用的類神經網路為倒傳遞類神經網路(Back Propagation neural network)，其實驗結果比較 Cross-validation 如表 5.9，一半訓練一半測試如表 5.10。而本論文提出辨識方法實驗結果 Cross-validation 如表 5.11，一半訓練一半測試如表 5.12。

表 5.9 [15]Cross-validation 之表情辨識率

	生氣	無表情	快樂	驚訝	辨識率%
生氣(30)	28	1	1	0	93.33%
無表情(30)	0	28	3	0	93.33%
快樂(31)	3	0	22	5	73.33%
驚訝(30)	0	2	2	25	83.33%
平均					85.12%

表 5.10 [15]一半訓練一半測試之表情辨識率

	害怕	無表情	快樂	驚訝	辨識率%
生氣(15)	15	0	0	0	100%
無表情(15)	0	15	0	0	100%
快樂(16)	1	0	14	1	87.5%
驚訝(14)	0	2	3	9	64.43%
平均					88.33%

表 5.11 本論文提出方法 Cross-validation 之表情辨識率

	生氣	無表情	快樂	驚訝	辨識率%
生氣(30)	29	0	1	0	96.67%
無表情(30)	1	27	0	2	90%
快樂(31)	0	1	26	4	83.87%
驚訝(30)	0	2	3	25	83.33%
平均					88.43%

表 5.12 本論文提出方法一半訓練一半測試之表情正確率

	生氣	無表情	快樂	驚訝	正確率%
生氣(15)	15	0	0	0	100%
無表情(15)	0	14	0	1	93.33%
快樂(16)	0	1	15	0	93.75%
驚訝(14)	0	0	2	12	85.71%
平均					93.33%

另外，取 JAFFE 表情資料庫之多人表情影像如無表情、快樂、驚訝、生氣和悲傷這五種表情範例，共 150 張，訓練及測試半徑基底函數類神經網路。實驗結果將與 Cheng [20]來進行比較。其訓練及測試張數如表 5.13 所示，而實驗結果比較如表 5.14 所示。

表 5.13 Cheng [20]使用之訓練及張數

表情	無表情	快樂	生氣	悲傷	驚訝
訓練張數	6	14	21	4	8
測試張數	24	17	9	25	22

表 5.14 Cheng [20] 辨識率之比較

	無表情	快樂	生氣	悲傷	驚訝	平均辨識率
Cheng[20]	70.83%	64.71%	55.56%	52%	68.18%	62.4%
The proposed	91.67%	100%	100%	20%	77.27%	72.16%

在表 5.14 我們可以發現，本論文所使用辨識方法在悲傷的辨識率上是不理想的，原因是由於在表 5.13 所示的訓練張數只有 4 張，測試張數有 25 張，此訓練樣本數對於我們所使用的類神經網路是不足的，以至於在辨識率上是不理想，若能提高悲傷的訓練樣本數，其辨識率或許會提高。若改以一半訓練及一半測試，本

本論文所使用辨識方法之辨識率如表 5.15 所示。

表 5.15 本論文提出方法之一半訓練一半測試之辨識率

	無表情	快樂	生氣	悲傷	驚訝	平均辨識率
The proposed	93.33%	80%	93.33%	86.67%	71.43%	85.13%

最後，再取 JAFFE 表情資料庫之多人表情影像如生氣、無表情、快樂、驚訝、生氣、害怕和悲傷這七種表情範例，共 207 張，訓練及測試半徑基底函數類神經網路。實驗結果將與 Bayshyal [8]來進行比較，其測試張數每人每種表情各 1 張，JAFFE 表情資料庫 10 人測試張數共 70 張，而實驗結果比較如表 5.16 所示。

表 5.16 Bayshyal [8] 辨識率之比較

	平均辨識率
Bayshyal [8]	87.51%
The proposed	90%

第六章

結論

在表情辨識上，大部分是直接觀察臉部表情或是擷取生理訊號來分析。然而直接臉部表情有時並不能準確斷定反映在內心的情緒，或者以單純分析生理訊號的波動來斷定情緒種類又不夠強健，而且每個人影響生理訊號的強度是完全不會相似的。所以為了追求更準確地判斷的表情種類，提供一個表情辨識系統是本論文的目標。

因此，本論文提出一套結合生理訊號之臉部表情的辨識系統。在我們所設計的情緒激發實驗中所得到的臉部影像及原始生理訊號，可以經過前處理後得到的臉部特徵及生理訊號特徵，再使用類神經網路訓練和測試這些特徵，最終去判斷真實內心的情緒和臉部表情為何。

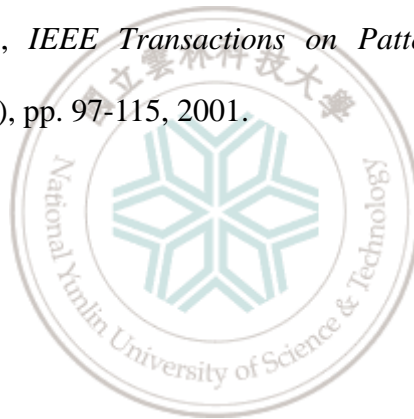
由於每個人是唯一且獨特，可能存在不可預期的因素而影響著生理訊號，而且缺少了“ground truth”，因此沒有一個標準的機制在，使得在辨識上有著極大的挑戰性。假以時日，如果分析更多的資訊，如肢體動作、身體姿勢、聲音語調、用字選擇、不同個性的人、每日及長時間觀察、受試者的直接反應等特徵，或許可以得到更精確的結果。

參考文獻

- [1] P. Ekman, W.V Friesen, *The Facial Action Coding System: A Technique For The Measurement Of Facial Movement*. San Francisco: Consulting Psychologists Press, 1978.
- [2] P. Ekman, W.V Friesen, “Constants across cultures in the face and emotion,” *J Per. Soc. Psychol.* 17 (2) , pp. 124-129, 1971.
- [3] G.Donato, S. Bartlett C. Hager, P. Ekman, J. Sejnowski, “Classifying facial action”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21, 10, pp. 974-989, 1999.
- [4] Peng Wang, Frederick Barrett, Elizabeth Martin, Marina Milonova, Raquel E. Gur, Ruben C. Gur, Christian Kohler, Ragini Verma, “Automated video-based facial expression analysis of neuropsychiatric disorders,” *Journal of Neuroscience Methods*, pp .224–238, 2008.
- [5] Kazuhiko Takahashi, “Remarks on emotion recognition from multi-modal bio-potential signals,” *IEEE International Conference on Industrial Technology*, pp. 1138-1143, 2004.
- [6] Rosalind W. Picard, Elias Vyzas, Jennifer Healey, “Toward machine emotional intelligence: analysis of affective physiological state,” *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 10, pp. 1175-1191, 2001.
- [7] Chathura R. De Silva, Surendra Ranganath, Liyanage C. De Silva, “Cloud basis function neural network: A modified RBF network architecture for holistic facial expression recognition”, *Pattern Recognition*, Vol. 41, pp. 1241-1253, 2008.

- [8] Shishir Bashyal, Ganesh K. Venayagamoorthy, "Recognition of facial expression using Gabor Wavelets and learning vector quantization", *In Press, Corrected Proof, Available online 14 January, 2008.*
- [9] C.Y. Chang, Y.C Tu, and H.H. Chang, "Adaptive Color Space Switching Based Approach for Face Tracking," *Lecture Notes in Computer Science*, Vol. 4223, pp. 244-252, 2006.
- [10] Rafael C. Gonzalez, Richard E. Woods, "Digital Image Processing 2nd", *Prentice-Hall International, Inc, 2002.*
- [11] R. M. Gray, "Vector Quantization," *IEEE Acoustics, Speech, and Signal Processing Magazine*, Vol. 1, pp. 4-29, 1984.
- [12] F. M. Ham I. Kostanic, "Principles of Neurocomputing for Science & Engineering", McGraw-Hill international edition, ISBN 0-07118161-x.
- [13] S. Haykin, "Neural Networks A Comprehensive Foundation second edition" Prentice-Hall, Inc. ISBN 0-13-908385-5.
- [14] M. J. Lyons, Julien Budynek, and Shigeru Akamatsu, "Automatic Classification of Single Facial Images," *IEEE Transaction On pattern analysis and machine intelligence*, Vol. 21 , No. 12, pp. 1357-1362, December, 1999.
- [15] 吳明衛，自動化臉部表情辨識系統，國立成功大學，碩士論文。
- [16] Shiqian Su, Baocai Yin, "A Robust Face Detection Method," *IEEE Proceedings of the Third International Conference on Image and Graphics*, 2004.
- [17] Rein-Lien Hsu, Mohamed Abdel-Mottaleb, Anil K. Jain, "Face Detection In Color Images", *IEEE Image Processing*, pp. 1046-1049, 2001.

- [18] C. Garcia, G. Zikos, G. Tziritas, “Face Detection in Color Images using Wavelet Packet Analysis”, *IEEE International Conference on Multimedia Computing and Systems*, Vol. 1, pp. 703-708, 1999.
- [19] 陳世璋，使用類神經網路之自動化臉部表情辨識系統，國立成功大學，碩士論文。
- [20] Shyi-Chyi Cheng, Ming-Yao Chen, Hong-Yi Chang, Tzu-Chuan Chou, “Semantic-based facial expression recognition using analytical hierarchy process”, *Expert Systems with Applications*, Vol. 33, pp. 86-95, 2007.
- [21] Tian, Y.-L., Kanad, T., & Cohn, J. F. , “Recognition Action Units for Facial Expression Analysis”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23(2), pp. 97-115, 2001.



Facial Expression Recognition with Consideration of Physiological Signals

Chuan-Yu Chang (張傳育) Jeng-Shiun Tsai (蔡政勳)
Institute of Computer Science and Information Engineering,
National Yunlin University of Science & Technology
chuanyu@yuntech.edu.tw

摘要

本論文的目的是提出一套結合視訊與生理訊號之臉部表情辨識系統。臉部表情的特徵是從不同的表情上擷取出特徵點，並且組成一組特徵向量。生理訊號方面，擷取三種生理訊號：皮膚電位反應、指溫和心率，分析並定義出特徵向量。在分類與辨識方面，以學習向量量化類神經網路(LVQNN)來分辨四種情緒：愛、歡樂、驚訝和害怕。在所設計的情緒激發實驗中，蒐集並擷取臉部和生理特徵的資料，並且訓練和測試類神經網路(LVQNN)。實驗結果顯示我們提出的方法在基於臉部表情和生理訊號的情緒辨識上能有不錯的識別。

關鍵詞： 臉部表情辨識、生理訊號、情緒、類神經網路

Abstract

A facial expression recognition system with consideration of physiological signals is proposed in this paper. Facial features extracted from specific feature points in facial expression images are combined as facial feature vectors. The physiological features vector consists of three physiological signals, skin conductivity, finger temperature and heart rate, are used to analysis. A learning vector quantization (LVQ) neural network is applied to classify four emotions, love, joy, surprise, and fear. Facial and physiological features extracted from the specific designed emotion-exciting experiments are used to train and test the classifier. Experimental results show the recognition system can correctly identify facial expressions and physiological emotions.

Keywords: facial expression, physiological signals, emotion, neural networks.

I. Introduction

Human emotional expressions play an important role in human-to-human interaction, and help people realize what they feel to each other. Some common expressions are facial expression, body language, word choices, tone of voice, posture, and physiology response. Besides word choices and tone of voice, posture, body language and human facial expression are visible emotional reactions that accompany with some physiological signal changes.

Among the visible emotional reactions, facial expressions often cause various facial feature changes while emotion are excited. Thus, the conspicuous and detectable face features can be used to detect different emotions. According to face features, such as position of eyebrows, eyes, and mouth, we can determine what the facial expression is because the variation of these face features are more than other face features. A well known facial expression model is given by Ekman and Friesen in [1]. They also identify six basic human emotions: *Fear*, *Surprise*, *Sad*, *Anger*, *Disgust*, *Happy* [2] and their associated facial expressions. Hammal [3] propos a method to extract face feature points from eyes, eyebrows, and mouth in the face image. In facial expression recognition (FER), early attempts by Suwa [6] use pattern recognition to recognize facial expression. Since then, Samal and Iyengar [7], Pantic and Rothkrantz [8], Fasel and Luetin [9], and Wang [10] have produced automatic FER approaches from dynamical video sequences or static images. In addition, Gabor Wavelet has been applied to recognize facial expression in frequency domain [4] [5].

Physiological signals are usually used as features to recognize emotions. Physiological signal pattern of emotion recognition has important applications in medicine, psychology

and physiology, mental disorder, and human-to-computer interaction. These applications can potentially help in assessing and quantifying stress, tension, anger, and other emotion that influence health. Because physiological signals reaction is non-autonomic nerves in physiology, we are not able to control these signals change while emotion is excited. This characteristic can be useful information. Common human physiological signals are skin conductivity, blood volume pressure, respiration, electromyogram (EMG), electrocardiogram (ECG), electroencephalographic (EEG), finger temperature, heart rate, and else. In emotion recognition areas, some researches have used physiological signals as features to determine and classify different kinds of emotions. Rosalind [11] analyzes four kinds of signals to classify eight classes of emotions. Kim [12] proposes a physiological signal-based emotion recognition system using three signals to recognize three and four categories, respectively. Takahashi [13] reports an emotion recognition system identifying five emotions from three bio-potential signals.

In this paper, we propose a novel approach for facial expression recognition by considering human physiological signals. We combine facial expression with physiological signals to recognize four kinds of emotion (fear, love, joy, and surprise). Although the facial expression sometimes is not corresponding to the physiological emotion, we try to find the general relationship between them and fine out what the real emotion is. This paper is organized as follows. In Section 2, the data samples which we collect is described. Our methods are presented in Section 3. The experimental results are provided in Section 4. Finally, the conclusions are given in Section 5.

II. Facial Expression and Physiological Signals Acquisition

An experiment is designed to acquire and record facial expressions and physiological signals. The acquisition procedure is shown in Figure 1. We put three physiological signal sensors on the subjects' body surface to gather the galvanic skin response (GSR), finger temperature (Temp), and heart rate (HR).

Simultaneously, in front of the subjects there is a camera which will catch their face and record face images while they watch video from monitor. In order to excite subjects' emotions (fear, love, joy, and surprise), the subjects are requested to watch several designed video segments. By watching the video segments, emotions of the subjects are excited; physiological signals are recorded by sensors and the facial expressions are also captured by the camera.

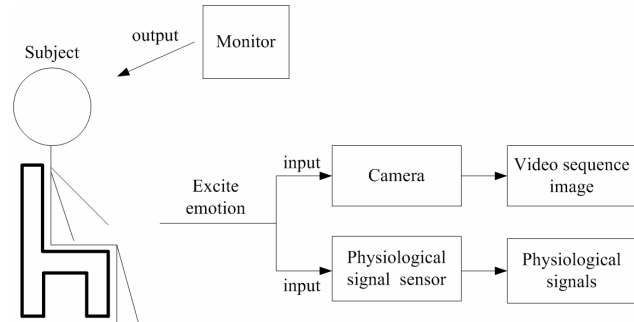


Figure 1. The schematic diagram of the emotion-exciting and emotion acquisition environment.

III. Methods

In this section, we will present our methods in three parts - facial expression, physiological signals and facial expression recognition

1. Facial Expression

To recognize the facial expression, we first detect and track the face region; second extract respective feature points from the detected face region; third calculate respective face features; and last construct the facial expression model and recognize different facial expressions in the detected region. Detail process are shown in Figure 2 and described in the following subsections.

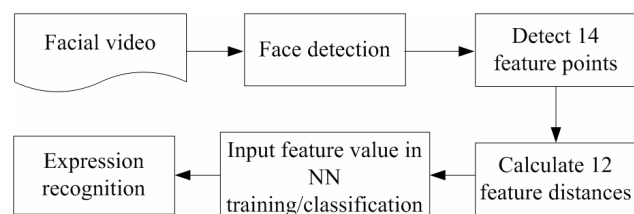


Figure 2. The flow chart of facial expression recognition.

1.1 Face Detection

There are many methods using ellipse template matching strategies to detect the face regions. To detect faces robustly, in our prior work [14], we proposed an adaptive color space switching method, which can automatically detect and mark face regions in video frames. The method adaptively selects an appropriate color space according to varying environment. It can detect multiple faces under complex background environment and variable lighting condition. Figure 3 shows two results as examples of the face detection.

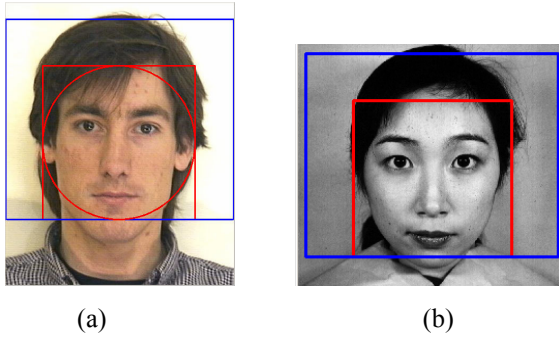


Figure 3. Samples of face detection (a) face detection in color image; (b) face detection in gray scale image.

1.2 Feature Points

To extract enough significant feature points of a face, we define and extract fourteen feature points, from eyes, eyebrows, and mouth. The points are 3 points on each eye, 2 points on each eyebrow, and 4 points on the mouth. These locations of significant feature points are shown in Figure 4(a). These feature points are the basis in determining the facial expression.

1.3 Feature Distances

Based on the 14 feature points extracted from face, we define 12 feature distances. The feature distances are calculated by Euclidean distance as shown in Figure 4(b).

Because the size of the detected face region and the feature distances are varying with different people and different environment instance, the feature distances are normalized by the distance between the inner corners of two eyes in advance. This is based on the fact that the inner

corners of two eyes are stable and invariant for an individual person. The normalized feature distances are calculated by following equation:

$$X_{nor}^{exp} = X^{exp} / n, \quad X \in \{a, b, \dots, l\} \quad (1)$$

where n denotes the distance between the inner corners of two eyes, the X_{nor}^{exp} and X^{exp} represents the normalized distance and the original distance that existing facial expression, respectively. Similarly, we acquire the natural face expression (the face without any expression) as the comparison basis of face expression. The normalized feature distance of the natural face expression is computed by following equation:

$$X_{nor}^{unexp} = X^{unexp} / n, \quad X \in \{a, b, \dots, l\} \quad (2)$$

Where n denotes the distance between the inner corners of two eyes, X_{nor}^{unexp} and X^{unexp} represents the normalized and original distance of the feature distances captured from natural expression, respectively. Accordingly, we get two sets of normalized distances, consisting of expression and natural features.

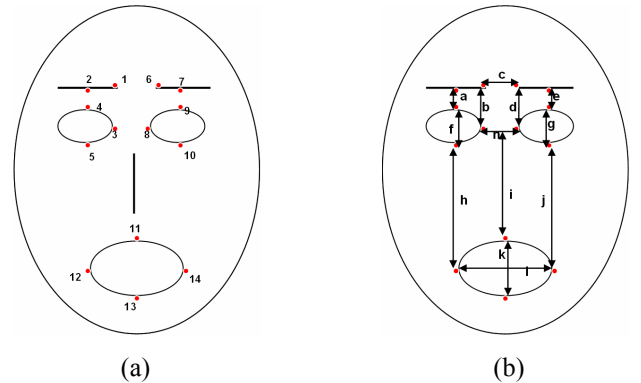


Figure 4. (a) The feature points in the face; (b) the feature distances among the feature points.

1.4 Facial Feature Vector

After obtained two sets of normalized feature distances, we define the feature offset vector of the facial expression. Compared with the natural expression, we can obtain the characteristics of facial expressions. The feature offset vector is defined as:

$$off(X) = X_{nor}^{exp} - X_{nor}^{unexp}, \quad X \in \{a, b, \dots, l\} \quad (3)$$

Therefore, Eq. (3) can be re-described as:

$$F = [off(a), off(b), \dots, off(l)]^T \quad (4)$$

2. Physiological Signals

Physiological signals are more sensitive than facial expression, thus are usually used for emotion recognition. The framework of the emotion recognition comprises following steps: (1) collecting three raw physiological signals; (2) preprocessing these signals; (3) calculating signals as feature vectors; and (4) constructing physiological signal models and recognizing different emotions. The processes are shown in Figure 5. Details of the processes are described in the following sub-sections.

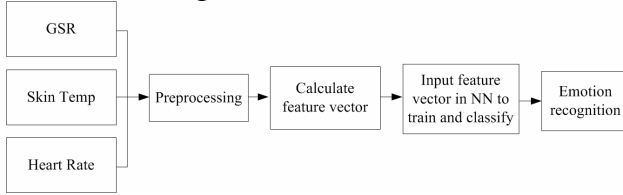


Figure 5. The schematic flow of emotion recognition using physiological signals.

2.1 Signals Extraction

In collecting the raw physiological signals, three physiological signals, skin conductivity (*GSR*), finger temperature (*Temp*), and heart rate (*HR*) are measured. There are two critical moments, when the signal patterns are extracted, first, the subject makes facial expressions and second, the similar trend of large variation happens while comparing to others. Every period of the signal segment is ten seconds, where data patterns are extracted every 0.05 seconds. Therefore, a ten-second signal segment has two hundred data patterns.

2.2 Signals Preprocessing and Signals Feature vector

Because the exciting degree varies and physiological signals are varied with different subjects, the normalization is required on each signals of the subjects. Let δ represents the physiological signals which are *GSR*, *Temp*, and *HR*, then, S_m^δ represents the m th value of δ . The normalized equation is defined as:

$$\tilde{S}_m^\delta = \frac{S_m^\delta - \min(S^\delta)}{\max(S^\delta) - \min(S^\delta)} \quad (5)$$

where S^δ is a data set of physiological signals

GSR, *Temp*, and *HR*; \tilde{S}_m^δ is the m th normalized value of δ ; $\max(S^\delta)$ and $\min(S^\delta)$ are the maximum and the minimum value of S^δ , respectively. The normalization is performed individual on each the physiological signals of the subjects. In other words, the physiological signals should not be normalized with that of the other subjects.

We compute the mean of sub-segment which is extracted from \tilde{S}_m^δ of each emotion. Let μ_k^δ be the mean value which represents k th sub-segment of S^δ and is calculated by each emotion \tilde{S}_m^δ as follow:

$$\mu_k^\delta = \frac{1}{T} \sum_{m=(k-1)*T+1}^{k*T} \tilde{S}_m^\delta \quad (6)$$

where T is total numbers of each sub-segment. The μ_k^δ of each emotion is organized as a feature vector Z_k which is defined as follow:

$$Z_k = [\mu_k^{GSR}, \mu_k^{Temp}, \mu_k^{HR}]^T \quad (7)$$

3. Recognition of Facial Expression and Emotion based Physiological Signals

After the previous steps, we have obtained feature vectors of the facial expression and physiological signals from a specific simulated emotion exciting experiment. These vectors are used to train the neural network, which is adopted to classify the different emotions. After training, new feature vectors, which are different from that are used in the training process, are used to test the trained neural network which determines the emotions of the testing feature vectors.

3.1 Learning Vector Quantization Classifier

We use a learning vector quantization (LVQ) neural network to classify the facial expression and emotion. Learning vector quantization neural network is proposed by Kohonen [15] [16] which is one kind of supervised learning method. The LVQ learning process is that suppose there is input training a group sample of M dimension vector, which can think a group samples in M dimension space, and each sample has its classification which has classified center.

Therefore, LVQ determinates the classified center of each classification by samples in the sample space. LVQ has two layers, which is input and output layer. The neural network architecture for LVQ is shown in Figure 6.

Suppose that i . $\mathbf{x} = \mathbf{F}$ and \mathbf{Z}_k is the input vector in LVQNN1 and LVQNN2, respectively, where $\mathbf{F} = [\text{off}(a), \text{off}(b), \dots, \text{off}(l)]^T$ and $\mathbf{Z}_k = [\mu_k^{GSR}, \mu_k^{Temp}, \mu_k^{HR}]^T$ as defined previously. ii . $\mathbf{y} = \{y_{fear}, y_{love}, y_{joy}, y_{surprise}\}$ is output vector both in LVQNN1 and LVQNN2. The w_j represents the weight vector defined as:

$$\mathbf{w}_j = [w_{j1}, w_{j2}, \dots, w_{jN}]^T \quad (8)$$

where $j = \{\text{fear}, \text{love}, \text{joy}, \text{surprise}\}$; $N = \{N_1, N_2\}$ is input vector dimension which N_1 is 12 which is $\{a, b, \dots, l\}$ of \mathbf{F} in LVQNN1 and N_2 is 3 which is $\{GSR, Temp, HR\}$ of \mathbf{Z} in LVQNN2, respectively.

Each input vector should be classified in terms of its actual category number in advance and represents as C_{x_i} in LVQNN1 and LVQNN2 individually. C_{w_j} which in LVQNN1 and LVQNN2 individually is corresponding category label of each neuron weight vector. The determination of winner neuron is computed by Euclidean norm, is represented as:

$$\min_{\forall j} d(x_i, w_j) = \min_{\forall j} \|x_i - w_j\|_2^2 \quad (9)$$

where $d(x_i, w_j)$ is selected the shortest Euclidean distance which between each neuron and input vector. The neuron which has the shortest distance with input vector is winner neuron. This also indicates that the input vector is associated with the classification of winner neuron. Hence, by comparing the category number of the input vector C_{x_i} and the category label of winner neuron C_{w_j} , we can judge the classified result is correct or not. If it is correct or not, we do two kinds of update process with weight vector w_j :

1. if the category label of winner neuron C_{w_j} is equal with the category number of the input vector C_{x_i} , that is $C_{w_j} = C_{x_i}$, the updated mode is:

$$w_j(k+1) = w_j(k) + \mu(k)[x_i - w_j(k)] \quad (10)$$

2. if the category label of winner neuron C_{w_j} is not equal with the category number of the input vector C_{x_i} , that is $C_{w_j} \neq C_{x_i}$, the updated mode is:

$$w_j(k+1) = w_j(k) - \mu(k)[x_i - w_j(k)] \quad (11)$$

where μ is learning rate, and k is iteration number.

After two kinds of update process, when the classified result is correct, the winner neuron would be close the input vector; on the contrary, when the classified result is not correct, the winner neuron would be far from the input vector. In the end, the weight vector will adjust to the better location which fits in with the category number of input vector after iteration.

The LVQ algorithm steps can be generalized as following:

- Step 1: Randomly selecting the weight vector $w_j(0)$ and initializing. Set up learning rate μ and iteration number k .
- Step 2: Compute the Euclidean distance between input vector x_i and neuron, and decide winner neuron in equation (8).
- Step 3: Use equation (9) and (10) update weight vector.

When the weight vector updates slightly or has no update, the network will be convergent and stop the algorithm, otherwise return Step 2. Furthermore, we can adjust the iteration number and learning rate so that the network learns better.

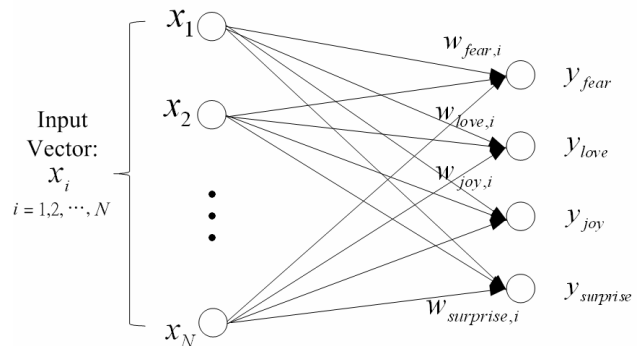


Figure 6. Neural network architecture for learning vector quantization. The output classes $y = \{y_{fear}, y_{love}, y_{joy}, y_{surprise}\}$.

3.2 Pattern Classification using LVQ Neural Network

The obtained face feature vectors and physiological signals feature vectors are classified by two learning vector quantization neural networks (LVQNNs), respectively. The classification process is shown in Figure 7.

Each input feature vector is compared with its LVQNN with the Euclidean distance. For an unknown emotion, the distance between individual classes of two LVQNNs is summed up. The final classification result is decided by the one has the least summation value. For example, the final summation result of class 1 is calculated by sum up the output value of class 1 of LVQNN1 and the output value of class 1 of LVQNN2. Similarly, we can calculate the summation results of class 2, class 3, and class 4. Finally, the class has the minimum summation value is the classification result.

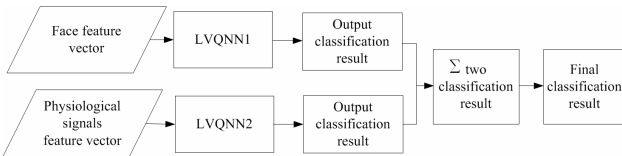


Figure 7. The training and classification processes

IV. Results

From the emotion exciting experiment, we collected 240 static facial images and their corresponding physiological signals including four kinds of emotion (fear, love, joy, and surprise) from 6 subjects. The duration of physiological signal of each emotion is ten seconds. The first five seconds of physiological signal samples are used for training the LVQNNs, and the remaining samples are used for testing. The sampling rate of the physiological signal and the camera is 20Hz. Samples of facial expression image are shown in Figure 8 and three samples of physiological signals are shown in Figure 9.

The training epochs and learning rate of learning vector quantization neural networks are set as 5000 and 0.01, respectively. To evaluate the recognition capability of facial and physiological features, we compared the

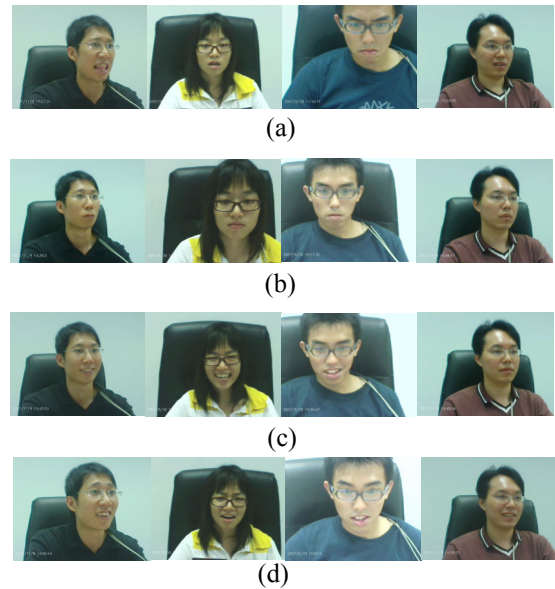


Figure 8. Samples of facial expression images: (a) fear; (b) love; (c) joy and (d) surprise.

recognition rate individually by only use the facial features, physiological signals and combination of the facial and physiological features. Table 1 shows the recognition result by facial features. Table 2 shows the recognition result by physiological signals. The recognition results for combination of the facial and physiological features are given in Table 3.

By comparing results in Table 1-3, we find that the average emotion recognition ratio with is not so high. This may explain as followings. The subjects may realize an event and have emotion response instantly after watching the event clip. But the facial expression may not be significant if they do not want to make an expression. In our acquisition environment, the subjects are watching the video clips along. Another problem is that the video should excite one certain emotion of subjects, but the facial expression of subjects appears another emotion, such as surprise becomes joy. Moreover, exciting emotion degree by watching the same video is not so identical based on physiological signals of each subject. The exciting emotion degree may be high or may be low. These problems could confuse us to recognize, therefore, this part is not so good. If the facial expression is more identical, the recognition rate would be better.

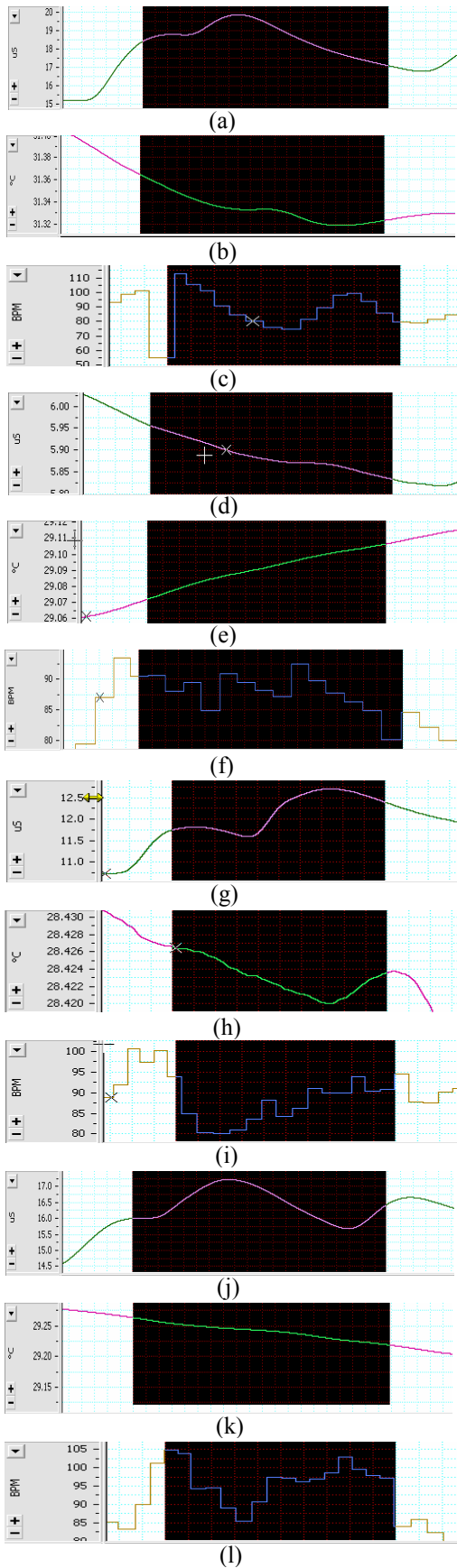


Figure 9. Samples of physiological signals for four emotions in black area:(a-c) the *GSR*, *Temp* and *HR* signals of fear;(d-f) *GSR*, *Temp* and *HR* signals of love;(g-i) the *GSR*, *Temp* and *HR* signals of joy;(j-l) the *GSR*, *Temp* and *HR* signals of surprise.

Table 1. Recognition Result of facial features

	Fear	Love	Joy	Surprise	Acc%
Fear	13	16	1	0	43.3%
Love	1	28	0	1	93.3%
Joy	0	4	20	6	66.7%
Surprise	2	8	7	13	43.3%
Total					55.8%

Table 2. Recognition result of physiological signals

	Fear	Love	Joy	Surprise	Acc%
Fear	15	0	2	13	50%
Love	0	27	3	0	90%
Joy	5	5	15	5	50%
Surprise	4	2	0	24	80%
Total					67.5%

Table 3. Recognition Result of combination of facial and physiological features

	Fear	Love	Joy	Surprise	Acc%
Fear	17	2	2	9	56.7%
Love	0	29	1	0	96.7%
Joy	5	0	16	9	53.3%
Surprise	3	1	0	26	86.7%
Total					73.3%

V. Conclusions

In this paper, we propose a facial expression recognition system with consideration of physiological signals. We analyzed external (facial expression) and internal factors (physiological signals) of human responses to judge what the inherent emotion is. The average recognition rates are 55.8%, 67.5% and 73.3% by using facial expression, physiological signals, and combination of both, respectively. Because everyone is unique and there exist many unexpected factors which affect the raw physiological signals, we conclude that four generalized expression models from 6 subjects. The absence of the “ground truth” makes the emotion recognition directly by using physiological signals seem to be a big challenge.

Possibly a reasonable result will be achieved when the system has dealt with manifold modalities, such as body language, word choices, tone of voice, posture, body movement, and other physiological signals. Even consideration

with more information such as user location, time of day, proximity to different personality people, and the direct response of user's diction analysis should become more effective in building a long term model of human emotional response.

Acknowledgements

This work was supported by the National Science Council, Taiwan, under the grants NSC 96-2221-E-224-070.

Reference

- [1] P. Ekman, W.V Friesen (1978), The Facial Action Coding System: A Technique For The Measurement Of Facial Movement. San Francisco: Consulting Psychologists Press.
- [2] P. Ekman, W.V Friesen (1971), "Constants across cultures in the face and emotion," *J Per. Soc. Psychol.* 17 (2) , pp. 124-129.
- [3] Z. Hammal, N. Eveno (2006), A. Caplier, Py Coulon, "Parametric models for facial feature segmentation," *Signal Processing* 86, pp. 399-413.
- [4] Zhan Yong-zhao, Ye Jing-fu, NIU De-jiao, CAO Peng (2004), "Facial Expression recognition Based on Gabor Wavelet Transformation and Elastic Templates Matching," *Proc. of the Third International Conference on Image and Graphics, IEEE Computer Society*, pp. 254-257.
- [5] Shishir Bashyal, Ganesh K. Venayagamoorthy (2008), "Recognition of facial expression using Gabor Wavelets and learning vector quantization", *Engineering Applications of Artificial Intelligence*, In Press, Corrected Proof, Available online 14 January.
- [6] M. Suwa, N. Sugie, K. Fujimora (1978), "A preliminary note on pattern recognition of human emotional expression," *Proceedings of the Fourth International Joint Conference on Pattern Recognition, Kyoto, Japan*, pp. 408-410.
- [7] A. Samal, P.A. Iyengar (1992), "Automatic recognition and analysis of human faces and facial expression: a survey," *Pattern Recognition* 25 (1), pp. 65-77.
- [8] M. Pantic, L.J.M Rothkrantz (2000), "Automatic analysis of facial expression: the state of the art," *IEEE Trans. Pattern Anal. Mach. Intell.* 22 (12), pp. 1424-1445.
- [9] Pop B. Fasel, J. Luetin (2003), "Automatic facial expression analysis: a survey," *Pattern Recognition* 36 (1), pp. 259-275.
- [10] Peng Wang, Frederick Barrett (2008), Elizabeth Martin, Marina Milonova, Raquel E. Gur, Ruben C. Gur, Christian Kohler, Ragini Verma, "Automated video-based facial expression analysis of neuropsychiatric disorders," *Journal of Neuroscience Methods*, pp. 224-238.
- [11] Rosalind W. Picard, Elias Vyzas, Jennifer Healey (2001), "Toward machine emotional intelligence: analysis of affective physiological state," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 23, NO. 10, pp. 1175-1191.
- [12] K. H. Kim, S. W. Bang, S. R. Kim (2004), "Emotion recognition system using short-term monitoring of physiological signals," *Medical & Biological Engineering & Computing*, Vol. 42, pp. 439-427.
- [13] Kazuhiko Takahashi (2004), "Remarks on emotion recognition from multi-modal bio-potential signals," *Proc. of IEEE International Conference on Industrial Technology*, pp. 1138-1143.
- [14] C.Y. Chang, Y.C Tu, and H.H. Chang (2006), "Adaptive Color Space Switching Based Approach for Face Tracking," *Lecture Notes in Computer Science*, Vol. 4223, pp. 244-252.
- [15] T. Kohonen (1986), "Learning Vector Quantization for Pattern Recognition", Technical Report TKK-F-A601, Helsinki University of Technology, Finland.
- [16] T. Kohonen (1990), "Improved Versions of Learning Vector Quantization", *Proceedings of the International Joint Conference on Neural Networks*, San Diego, CA, Vol. 1, pp. 545-50.