

運用樹莓派建置情緒辨識控制系統

李鴻德
朝陽科技大學
資訊管理系

e-mail :

rockofsky1@gmail.com

*陳榮靜
朝陽科技大學
資訊管理系

e-mail :

*crching@cyut.edu.tw

王曉棟
朝陽科技大學
資訊管理系

e-mail :

xdwangjsj@xmut.edu.cn

摘要

人臉表情分析對於人機互動，一直都是很重要的課題。本論文研究提出一個使用樹莓派建置情緒辨識系統架構，透過樹莓派之鏡頭模組偵測使用者表情，並運用情緒辨識 API 辨識使用者當時之情緒。人的情緒有很多種類，而不同情緒下，人臉的表情會有不同的特徵，人在情緒為「喜」的時候，嘴角會上揚。「怒」的時候，容易皺眉頭。當系統分析出使用者當時的表情後，會有相對應的動作。辨識出情緒為喜或樂，使用者心情是好的時候，系統會播節奏輕快的曲子，燈光也會調成較為鮮豔的顏色，如果辨識結果為怒或哀，使用者情緒不穩定的時候，系統會播舒緩情緒的曲子，燈光調成柔和的光線，舒緩使用者的情緒。

關鍵詞：情緒辨識 API、樹莓派、表情辨識

Abstract

Emotion recognition is an important part in human-computer interaction. This research is bases on Emotion API to combine with Raspberry Pi to build an emotion recognition system. Using Raspberry Pi's camera model collects user's facial expression. Raspberry Pi will send user's facial expression to a personal computer (PC), and the PC analyzes user's emotion by Emotion API. According to the result, system provides servicing on the right way. People have various emotion, with different emotion, human's facial expression which contains different features. For example, if people is happy, he will smile. If people is angry, his brow will be squeezed. When result is happy, system will give user bright colors and rhythm of light music. When result is angry or sad, system will give user soft light and some songs to soothe emotion.

Keywords: Human-computer interaction, Emotion API, Raspberry Pi, Emotion Recognition

1. 前言

現在全球對於物聯網的討論相當熱烈，由於網路速度的大幅進步，資訊科技的應用範圍大幅提升，我們對於電腦的硬體及軟體的要求越來越高，希望電腦能夠帶給我們更方便的生活品質，所以我們越來越講究人機互動，人類往往會透過行為傳達出自己的情緒，這些除了會顯示出人類的情緒，也夾帶了許多的訊息，這些訊息可以了解人類當時的狀況，比如說：影像照護系統，是透過觀察病人的表情及行為確認病人的病況是否穩定；或者應用在行車駕駛上，當駕駛精神狀況不良時，臉部便會呈現倦容，表情會有所變化，監測系統就能給予適當的處理，提醒駕駛休息。

而電腦視覺對於人機互動的發展中佔有很重要的地位，如果電腦能夠清楚辨識出人類的表情，並給予適當的處理，那麼電腦跟人類的關係就更為靠近，所以近年來臉部表情辨識 (facial expression recognition) 在學術的研究中越來越熱門。根據擷取臉部特徵資訊加以量化並分類，便能夠讓電腦學習辨識人類的情緒狀態，能夠對人類提供更進一步的服務。

現在用臉部影像特徵作為電腦辨識表情的依據，大致分為兩類特徵提取方式。第一類的方式是根據幾何特徵做特徵提取的動作[4-7]，幾何特徵提取是利用臉部動作編碼系統 (Facial Action Coding System, FACS)[10]，臉部動作編碼系統能夠用來敘述臉部在特定區塊各自位移的情形，第二類是根據外觀特徵作提取的方式[8-9]，主要是處理區域紋理特徵；而 Zhang, M. J. Lyons, M. Schuster and S. kamatsu 提出的研究[12]，比較了兩種類行的特徵提取方式，該研究結果顯示，透過區域紋理作特徵提取的方式，相較於幾何特徵提取的效果要來的好。由 P. Viola and M. Jones, 提出的維奧拉-瓊斯目標檢測框架 (Viola-Jones object detection framework) [1][11]是近幾年來在人臉表情辨識的研究上，熱門的分類器之一，其在

運算速度的方面上，有著不錯的表現。

在本論文中，我們提出一個方法將學術上表情辨識的研究，應用於實際生活中；以樹莓派作為蒐集人類臉部影像的工具，透過 R 語言將影像交給情緒辨識 API 作辨識，最後將結果傳回樹莓派，樹莓派再依判斷結果作下達指令的動作。

本文架構如下：第二節回顧本論文相關的影像辨識技術，第三節為本論文所提出的研究方法，第四節為實驗結果討論，第五節為結論與未來方向。

2. 研究背景

2.1 樹莓派在影像上的應用

樹莓派是一種微型電腦，有多樣的擴充模組，樹莓派可以利用擴充鏡頭模組獲取影像資訊；樹莓派支援 Python，同時 Python 也支援跨平台電腦視覺庫(OpenCV)，讓樹莓派有能力在影像辨識的方面上進行相關應用。

2.2 維奧拉-瓊斯目標檢測框架

維奧拉-瓊斯目標檢測框架是一種分類器(Classifier) [1][11]，有三個特點，第一個特點，分類器可以利用每個相鄰的像素(pixel)差找出矩形特徵，便可利用這些矩形特徵定義人類臉部的框架，圖 1. 為利用矩型特徵找出人臉框架的範例，此種方法在計算的速度上，有非常好的表現。第二個特點，此方法具備即時性(real time)，在跨平台電腦視覺庫也有提供函式可供開發者使用。最後一個特點，維奧拉-瓊斯目標檢測框架是一個自適應增強的分類器(AdaBoost)，自適應增強分類器的特性在於分類後被分錯的樣本都會被拿來訓練下一個分類器，即使一開始是弱分類器(Weak Classifier)，經過不斷訓練也是會訓練出一個強分類器(Strong Classifier)。



圖 1 利用矩形特徵辨識人臉框架

2.3 卷積神經網絡

卷積神經網絡 (Convolutional Neural Network, CNN) [2] 為前饋神經網路 (Feedforward Neural Network) 的一種，前饋神經網路是神經網路中最簡易且最早發明的。卷積神經網路是由一個全連通層和一或多個卷積層所組成，同時也包含關聯權重和池化層 (pooling layer)。卷積神經網路在影像跟語音辨識的效能上有著更好的表現。

2.3.1 卷積層

卷積層 (Convolutional layer) [2]，在卷積神經網路中每層卷積層都由若干個卷積單元組成。卷積運算的目標在於得到輸入的不同特徵，第一層卷積層可能只能夠得到初級的特徵像是稜角、邊緣或是線條等級的特徵，透過多層級的網絡能夠從初級的特徵中疊代得到更高級的特徵。

2.3.2 線性整流層

線性整流層 (Rectified Linear Units layer, Re LU layer) [2] 是使用線性整流 (Rectified Linear Units, Re LU)

$$f(x) = \max(0, x) \quad (1)$$

作為這一層神經的激活函數 (Activation function)。它可以優化判定函數跟整個神經網絡中的非線性特性，而且線性整流層不會影響卷積層。另外，別種的函數也能用來改善網絡的非線性特性，例如雙曲正切函數，

$$f(x) = \tanh(x), f(x) = |\tanh(x)| \quad (2)$$

Re LU 函數的優點在於可以提升神經網絡的訓練速度數倍，並能夠不影響模型整體的泛化準確度，所以目前是比較主流的函數。

2.3.3 池化層

池化 (Pooling) [2] 在卷積神經網絡中是重要的一環，池化像是一種模式的向下採樣。非線性池化函數有許多種模式，「池化最大化 (Max pooling)」是較常被人使用的函式。池化最大化是將輸入的圖片分割為多個矩形區域，再對每一個矩形區域輸出最大的值。這種機制

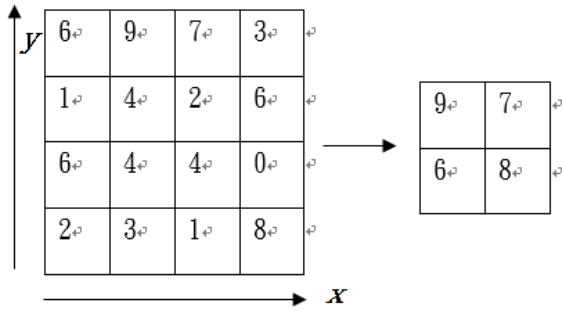


圖.2 池化概念圖

可以有效地原因在於，發現一個特徵後，特徵的精確位置遠不比該特徵和其他特徵的相對位置的關係來的重要。池化層會不間斷地縮減數據的空間大小，因此參數的數量和運算量也會跟著下降，有控制過度擬合的效果。CNN 的卷積層之間都會周期性地加入池化層。

池化層通常會分別運作於每個輸入的特徵並且縮減其特徵的大小。目前較為常使用形式的池化層是每 2 個元素就會從圖像分割出 2*2 的區塊，然後對每個區塊 4 個數值中取出最大值。這能夠縮減四分之三的數據量。圖 2. 為池化之概念圖。

由於池化層縮減數據的大小的速度過快，所以當前研究的趨勢是使用較小的池化濾鏡代替池化層。

2.3.4 損失函數層

損失函數層 (loss layer) [2]是用來判斷訓練的過程，比較的預測結果和真實結果之間的誤差，再決定如何調整下一次的訓練，損失函數層常常是網絡最後的層級。

2.4 情緒辨識 API

情緒辨識 API (Emotion API)[3]，是透過使用者上傳圖檔之人類臉部表情做為輸入資料，給圖檔中每一個人的表情做評分的動作，在評分的過程中同時運用了人臉辨識 API (Face API)，人臉辨識 API 會在圖檔找出人臉位置並用方框做標記。

情緒辨識 API 回傳的情感結果有八種：生氣、輕視、噁心、害怕、開心、無表情、難過和驚訝，這些情感都對應特定的表情，而這些表情也是人類交流中重要的部分。將圖檔上傳至情緒辨識 API 示範網頁所得到的結果如圖.3。

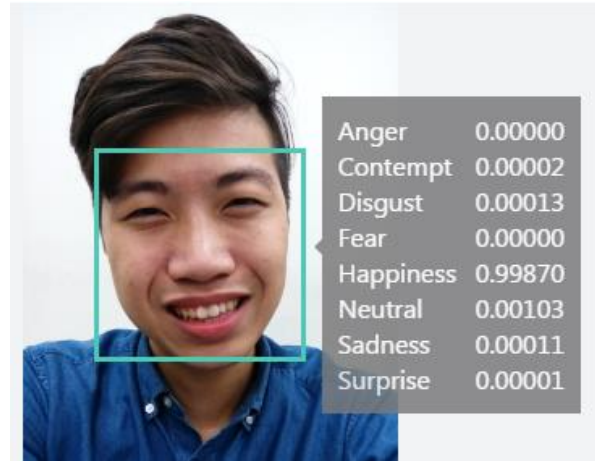


圖 3 情緒 API 示範網頁回傳結果

3. 研究方法

本論文所提的情緒分析應用系統，需具備樹莓派、樹莓派鏡頭模組、個人電腦及需要控制的相關設備，個人電腦需要安裝好 R Studio，此研究提出的架構乃是利用樹莓派之鏡頭模組，擷取使用者臉部影像，樹莓派再透過 Wi-Fi 將影像送至主機，利用 R 語言呼叫情緒辨識 API (Emotion API) 做情緒分析，在將分析後結果回傳至樹莓派，樹莓派再依照結果判斷後控制設備，圖 4. 為本論文所提方法知系統流程圖。

透過情緒辨識 API 可以抓出預先定義好的 27 個臉部標示 (Landmark point) 如圖 5[1]，這 27 個臉部標示分別為左右眉毛的眉頭跟眉尾 4 個點，雙眼眼睛的上下左右及瞳孔 10 個點，鼻子上中下兩側加鼻頭共 7 個點，嘴角兩側 2 個點，上嘴唇上下方 2 個點，下嘴唇上下方 2 個點。

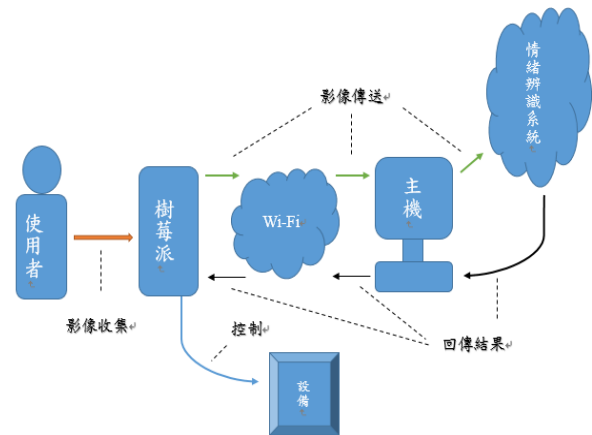


圖.4 系統流程圖



圖.5 27 個臉部標示

再根據人類的表情特徵進行情緒分析，情緒辨識 API 結果有八種，分別為「生氣、輕視、噁心、害怕、開心、無表情、難過、驚訝」。本研究將此八種情緒分為兩類，第一類是好心情，第二類是壞心情，好心情類別只有開心，則壞心情類別包含生氣、輕視、噁心、害怕、無表情、難過、驚訝。辨識完情緒隨後將分析結果送回至樹莓派，樹莓派再依照辨識結果下達指令至相關設備。

4. 初步實驗

4.1 情緒辨識單元

為了確認本研究系統的可行性，我們將利用 R 語言撰寫程式，上傳圖檔，呼叫情緒辨識 API 進行相關實驗。我們測試 8 張拍照圖片測試分析結果如下：

本實驗圖檔來源自智慧型手機之拍照檔，目的是在沒有控制任何因素的實驗下，情緒辨識 API 的表現如何。圖 6.為 R 語言運算後所繪出的結果，會給予八種情緒的情緒分數，情緒分數最高的情緒及是程式判斷的結果，圖 6.顯示開心的情緒分數有 0.997591 是 8 個情緒分數中最高的，及可判定該圖像人的情緒為開心。

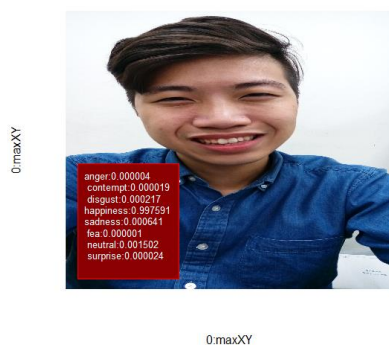


圖.6 R 語言繪出的結果

上傳八種不同表情的圖檔給程式分析後的結果如表格 1.，其中 AN(生氣)、CO(輕視)、DI(噁心)、FE(害怕)、HA(開心)、NE(無表情)、SA(難過)、SU(驚訝)，表格欄位主要分為圖檔輸入欄及運算結果欄，從結果可以觀察出，系統對於生氣、輕視、噁心、害怕、難過和驚訝等情緒分析較不擅長，而在高興跟無表情的方面表現不錯。

表格 1

AN 圖檔輸入	結果	CO 圖檔輸入	結果
	AN: 0.0026 CO: 0.0004 DI: 0.0025 FE: 0.0020 HA: 0.0004 NE: 0.9110 SA: 0.0550 SU: 0.0215		AN: 0.0009 CO: 0.0154 DI: 0.0010 FE: 0.0021 HA: 0.0015 NE: 0.7750 SA: 0.1940 SU: 0.0100
DI 圖檔輸入	結果	FE 圖檔輸入	結果
	AN: 0.0015 CO: 0.3350 DI: 0.0424 FE: 0.0066 HA: 0.0004 NE: 0.4460 SA: 0.1310 SU: 0.0372		AN: 0.0010 CO: 0.0023 DI: 0.0026 FE: 0.0051 HA: 2.94e-05 NE: 0.7810 SA: 0.0408 SU: 0.1670
HA 圖檔輸入	結果	NE 圖檔輸入	結果
	AN: 0.0056 CO: 0.0001 DI: 0.05320 FE: 0.0001 HA: 0.9280 NE: 0.0133 SA: 0.0033 SU: 0.0017		AN: 0.0015 CO: 0.0244 DI: 0.0042 FE: 0.0002 HA: 0.0002 NE: 0.8770 SA: 0.0876 SU: 0.0053
SA 圖檔輸入	結果	SU 圖檔輸入	結果
	AN: 0.0155 CO: 0.0283 DI: 0.0025 FE: 0.0051 HA: 0.0002 NE: 0.5190 SA: 0.4200 SU: 0.0104		AN: 0.0001 CO: 0.0002 DI: 0.0030 FE: 0.0011 HA: 4282e-06 NE: 0.0928 SA: 0.0041 SU: 0.8990

4.2 控制單元

得到情緒辨識結果後，回傳結果至樹莓派做判斷，樹莓派判斷結果後，將會透過藍芽，下達指令控制藍芽智慧光球，圖.7 為藍芽智慧光球設備，可以透過藍芽傳輸指令控制藍芽光球的燈光顏色。



圖.7 藍芽智慧光球設備

表格 2 藍芽智慧光球的情緒顏色對照表

情緒	燈光顏色	情緒	燈光顏色
生氣	 紅色	開心	 橘色
輕視	 湖水綠	無表情	 黃色
噁心	 綠色	難過	 藍色
害怕	 紫色	驚訝	 桃紅

我們將八種情緒結果設定為八種不同顏色，這八種顏色分別為紅色(生氣)、湖水綠(輕視)、噁心(綠色)、害怕(紫色)、開心(橘色)、無表情(黃色)、難過(藍色)、驚訝(桃紅色)，表格 2 是藍芽智慧光球的情緒顏色對照表。根據藍芽智慧光球顯示不同的顏色，我們即可知道使用者當下的情緒為何。

5 結論與未來方向

本研究的實驗皆為初步實驗，還有許多不完善的部分，像是情緒辨識 API，在開心情緒的辨識是上有著不錯的表現，但在生氣、難過、輕視、噁心、驚訝、害怕及無表情等情緒辨識上，無表情的情緒分數都會佔很高的分數，在未來研究上將會著重於處理 API 回傳的數值，過濾後再判斷，或是在影像傳至情緒辨識系統前加入前處理的動作，希望能夠改善情緒辨識的準確率。

未來除了改善情緒辨識的準確率，並會繼續研究如何擴增系統，控制藍芽智慧光球只是初步實驗，證明可以透過樹莓派做為媒介，收集資訊並控制設備。將來希望能夠將系統控制設備範圍擴增至影音設備或是溫度設備，利用控制音樂或是溫度的調節改善使用者的情緒。

致謝

本論文受科技部計畫補助編號 MOST-104-2221-E-324-019-MY2, MOST-103-2632-E-324-001-MY3，特此申謝。

參考文獻

- [1] "不再人心隔肚皮！R 語言實作人臉情感辨識", <https://dataholic.wordpress.com/2016/10/29/不再人心隔肚皮！R語言實作人臉情感辨識>
- [2] "維基百科(卷積神經網絡)", <https://zh.wikipedia.org/wiki/卷積神經網絡>
- [3] "Microsoft Emotion API", <https://www.microsoft.com/cognitive-services/en-us/emotion-api>
- [4] I. Kotsia and I. Pitas, "Using geometric deformation features and support vector machines," *IEEE Trans. Image Processing*, vol. 16, no. 1, pp. 172-187, 2007.
- [5] M. Pantic and I. Patras, "Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences," *IEEE Trans. Systems, Man, and Cybernetics --Part B: Cybernetics*, vol. 36, no. 2, pp. 433-449, 2006.

- [6] M. Pantic and L. J. M. Rothkrantz, "Facial action recognition for facial expression analysis from static face images," *IEEE Trans. Systems, Man, and Cybernetics --Part B: Cybernetics*, vol. 34, no. 3, pp. 1449-1461, 2004.
- [7] M. Pantic and L. Rothkrantz, "Expert system for automatic analysis of facial expression," *Image and Vision Computing*, vol. 18, no. 11, pp. 881-905, 2000.
- [8] M. S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Recognizing facial expression: machine learning and application to spontaneous behavior," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 2, pp. 568-573, 2005.
- [9] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face recognition by independent component analysis," *IEEE Trans. Neural Networks*, vol. 13, no. 6, pp. 1450-1464, 2002.
- [10] P. Ekman and W. V. Friesen, "*Facial Action Coding System*," *Consulting Psychologists Press*, College Avenue, Palo Alto, CA, 1978.
- [11] P. Viola and M. Jones, "Robust real time object detection," in *IEEE ICCV Workshop on Statistical and Computational Theories of Vision*, 2001
- [12] Z. Zhang, M. J. Lyons, M. Schuster and S. Akamatsu, "Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron," in *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition*, pp. 454-459, 1998.