

基於 SVM 的手勢辨識

劉家名	陳榮靜	戴紹國*
朝陽科技大學資訊管理系	朝陽科技大學資訊管理系	朝陽科技大學資訊管理系
e-mail : kevin50201a@gmail.com	e-mail : crching@cyut.edu.tw	e-mail : *skdai@cyut.edu.tw

摘要

手勢辨識技術近幾年來在人機介面應用上越來越被重視，有將手勢辨識應用在家電控制、VR 感測、密語傳遞等。本研究目的在分析手部姿勢，並進行影像的預先處理包括擷取影像、影像灰階以及利用 SVM 分類建模並預測，本文的流程為：取得影像，再使用 MATLAB 將手部圖片轉製成二維陣列，並利用 R 語言進行 SVM 做資料分析建立模型，再利用不同手勢資料進行預測並測試，未來將在樹梅派上實現本研究實驗室方法得以達到基於手勢辨識控制之目的。

關鍵詞：手勢辨識、樹梅派、SVM、MATLAB。

Abstract

Gesture recognition is more important to human-computer interaction this day. It can control home electronic device, VR, key delivering. This research aim is to do gesture analyzing, and media preprocessing. The preprocessing includes media collecting, gray-level transformation. Using SVM classifier to build a model and to detect users' gesture. The research steps are: getting media, transform media to gray-level by using MATLAB and use SVM to analyze data and build model by R language, using different gesture data to test and predict. The research will combine with raspberry pi to build gesture recognition system in the future.

Keywords: Gesture recognition, Raspberry pi, SVM, MATLAB

1.前言

手勢辨識技術在人機介面上被大量使用，取代了以往的鍵盤以及滑鼠等輸入裝置，讓使用者可以有更直覺地使用體驗，但是關於手勢的圖像資料庫比起臉部資料庫來說較為貧乏，而 SVM 這項分類方法在大部分的情況下也有較好的表現，此外 MATLAB 對於影像處理方面也有提供了一些方法，例如：傅立葉轉換或將圖片轉置成二元陣列。

本篇論文會先收集大量手部照片當作資料庫，再利用 MATLAB 中將影像轉置成灰階的方法，將所需要分析的目標手勢影像進行分析資料輸入端的預處理，再利用 R 語言裡面的 LIBSVM-e1071 配件包進行 SVM 的分析建立模型，最後達成分析手勢的目的。

隨著資訊科技技術不斷的進步與網際網路的發展迅速，並且資訊產業的發展漸漸地導向以人為本這個方向，進而衍生出智慧家庭的概念，要達到智慧家庭這個想法，物聯網(IOT)扮演著很重要的角色，所謂物聯網(IOT)的意思是萬物皆連上網，只要是家裡的電器都連上網路，就可以進行一連串物聯網(IOT)的應用，例如：環境偵測、控制家電開關，應用到物聯網(IOT)的技術有相當多，比如手勢辨識、人臉辨識等等。

手勢辨識常常應用在人機介面，取代以往的輸入端，例如：鍵盤、滑鼠等等，以達到更直覺的操控，另外手勢辨識也應用在許多地方，例如：多媒體應用控制、視頻遊戲和醫療系統，又因為手部便是會受到許多因素影響，例如：光線明暗度、距離的遠近或是遮擋，所以一般手勢辨識可以分為幾種方式進行，第一種是以軌跡來分辨，藉由手部的移動，來計算出 2D 或 3D 的座標值，但是此方法受限於硬體設備的攝像頭需要可以偵測深度，第二種，配戴電子手套，此方法雖然可以有很高的精確度，第三種分辨輪廓，並計算灰階度來做分類，總體來說分為兩個步驟：先進行機器學習，再來是模板匹

配，來達到手勢辨識的目的，而本文採用的方法為利用彩色攝影機拍攝照片再利用 MATLAB 找出輪廓，並計算灰階度，最後利用 R 語言進行 SVM 建模。

本研究為了該如何將不同的手勢區分出來進行實驗，當中為了將影像資料特徵提取找出最重要的特徵並進行分類，因而使用了影像灰階、去除背景、特徵提取以及 SVM 分類器這四項技術來達到本研究之研究目的。

本文其他部分包括：第二節文獻探討，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，第六節為參考文獻。

2. 文獻探討

手勢辨識常常被應用在生活的許多地方上，例如：張厥煒[1]等學者提出的門禁系統，就是利用手勢解鎖，來達成目的，其手勢辨識的重點為抓取手部的形狀特徵，進行驗證。由 Qiang ZHu[4]等學者提出的膚色檢測法當作擷取手勢特徵的方法，再利用 SVM 進行分類應用。或是 S. Malik[5]等學者提出的利用靜態背景相減法取得手部外型特徵。其他 A. Kuznetsova [2]等學者也提出了將手部姿勢的深度圖變換到點雲圖，其由特徵的形狀函數 (ESF) 描述相符，然後通過多層分類隨機森林 (MLRF)。Z. Ren [7]等學者所提出的是通過其輪廓描述手姿勢形狀，然後使用模板執行手勢分類。Y. Wang [6]等學者提出一個基於遞送的 2D 體積形狀由 SVM 用於使用深度圖像的手姿勢分類的架構。Ana I. Maqueda* [3]等學者提出了一套 VSLBP 的系統，加入時間與空間的資料維度來改善 LBP 增加準確率以及效率，並建立一個手勢資料庫，並公開出來。

MATLAB 常常被應用於影像處理，如 Ana I. Maqueda* [3]等學者利用 MATLAB 來做為研究工具，本研究利用 MATLAB 進行影像處理。R 語言是一個近年來常常被應用於統計的開放軟體，此軟體提供許多演算法比如：回歸分析或 SVM 等等，本研究利用 R 語言進行 SVM 分類實驗。

3. 研究方法

手勢辨識應用在物聯網上是一項常見且具有挑戰性的應用，常見的步驟為以下步驟(1)取得影像(2)影像分析(3)定義手勢涵義並建立

模型(4)將影像輸入模型判斷(5)識別出手勢並執行特定指令，圖 1 為手勢辨識流程。

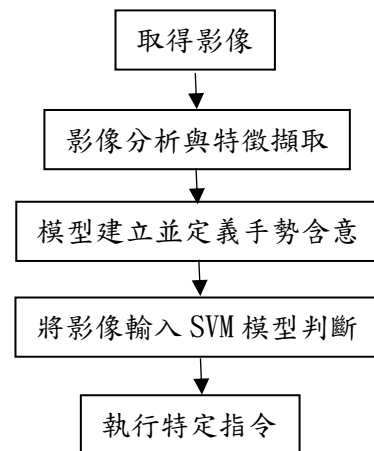


圖 1、手勢辨識流程

一般來說，花最多時間和資源的流程在於影像分析這一個流程，對於電腦來說抓取特徵是相當困難的一項運算，因為影像其實就是陣列，而我們需要讓電腦知道那些影像的特徵可以代表這張影像所以需要界定出不同影像的差別，所以為了減少未來應用在樹梅派上的研究中所需要做的資料預先處理時間與資源，所以本研究之研究動機為將影像資料進行預處理以便於未來在樹梅派應用之研究順利。

本研究將分為四個步驟依序進行，第一步先確定想要解決的問題以及要達成的目標和結果，第二步蒐集有相關文獻例如：手勢辨識、特徵抓取、SVM 分類器、影像處理，並且根據蒐集到的文獻當中所使用方法進行探討，第三步將進行實驗，將所構想的流程實現，最後分析實驗結果並歸納。以下將分別說明各個研究步驟：

- (1)釐清對於本研究之欲解問題、目的與研究的範圍。
- (2)影像資料的收集，例如：不同面向的手部照片，找出實驗過程中所需要技術的相關文獻，以及前人做過的相關研究，藉此可以釐清實驗的流程與方法。
- (3)本研究實驗將進行以下步驟：第一步，收集影像資料，目標為三種手勢，分別是：握拳、拇指、掌心，每一種類型各十張總共有三十張，第二步，影像進行灰階處理，將全彩的原圖轉成灰階，目的是降低資料維度，但是又可以保留原

始資料的特徵，第三步，抓取特徵將經過影像處理過的影像進行特徵抓取，

(4)依據實驗結果，對於實驗的成效是否可以達到實際應用的精確性，或是實驗的流程有哪些不足的地方需要改善，並加以改善讓實驗更加順利。

此外本篇論文會使用支持向量機 Support Vector Machines(SVM)來進行分類建立模型，支持向量機(SVM)是一個常常被用來進行分類的監督式演算法，其主目的是找到超平面上的一條分界線，分界線到各個分類的最短距離要相等並且距離越大越好，希望達到分類明確之目的，但是如果資料的特徵不明顯有時候會發生分類不清的問題，一般來說，會要求輸入訓練資料進行訓練，再利用測試資料來測試模型。

4. 實驗

本次實驗環境作業系統 Windows 10，處理器使用：Intel(R)Core(TM)i5-4200U，CPU@1.60GHz 2.30GHz，64 位元作業系統。

研究方向與目的確定為將影像資料進行處理並分類，實驗部分進行如下：收集影像資料：拇指、掌心、握拳各收集了十張不同人所拍攝的影像，此外未來規劃擴展同手勢不同角度之影像來增加手勢判定準確度，但須先評估實驗結果再作決定，避免 SVM 建立模型產生過度學習的現象，而判定標準為準確率，如果準確率達九十百分比，則會加入新的數據集建立模型並與原始模型準確度做比較。影像處理：使用 MATLAB 進行影像處理，目前暫訂了特徵選取的形狀為圓形，並且逐步進程式撰寫，未來預計加入其他的特徵擷取方式，並互相比較實驗結果，包括：準確度以及時間。

將上述之實驗後續完成，包括資料收集以及影像處理，並進行 SVM，內容如下：
SVM 建模：目前預計使用 RStudio 當作分析軟體，並利用其中的 SVM 套件包(e1071)來進行 SVM，會先將三種手勢分別代表的類別，例如：拇指代表 1、握拳代表 2、掌心代表 3，將影像資料集分成百分之七十做為訓練模型，剩下的百分之三十做為測試，其中的資料集包含三種手勢各十張影像。

4.1 實驗工具

本研究的實驗工具為 RStudio 和 MATLAB，其中 RStudio，是一款對於大數據分析存在著很大的優勢，因為本身其實是一種統計軟體，系統本身就是用於數據分析，並且在擴充套件上

相當方便，也因為是開放的軟體，所以套件資源也相當豐富，不僅是分析資料，更包含分析後的結果可以視覺化的呈現，更讓大數據分析更有意義，而 MATLAB 這個軟體，不管是在學術界或是業界用於影像處理的方便性都是大家認可的。

利用 R 語言的 e1071 套件包來進行 SVM 分類，在第三步驟中會進行 SVM 的分類，將資料庫的資料分為訓練集與測試集，訓練集為總資料庫的百分之七十，測試集為總資料的百分之三十。

4.2 收集影像資料

先收集三種不同種類的手勢，手勢包括：拇指圖 2(a)、握拳圖 2(b)以及掌心圖 2(c)，每種各十張。



(a) 拇指



(b) 握拳



(c) 掌心

圖 2 不同的手勢示意圖



(b) 灰階影像(握拳)

4.3 影像分析

利用 MATLAB 讀取影像，將影像像素尺寸調整成 12×12 ，目的為降低維度，再將影像轉灰階，圖 4(a) 掌心，圖 4(b) 握拳，圖 4(c) 拇指，本文特徵抓取將填滿特徵形狀設定為 disk，將手部特徵填滿，將原圖減去填滿的圖剩下的為雜訊，最後雜訊去除得到特徵。圖 3 為影像處理流程圖。圖 5 為影像處理結果。

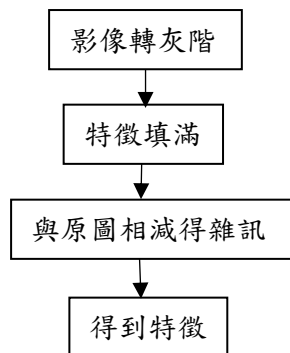


圖 3、影像處理



(a) 灰階影像(掌心)



圖(c)灰階影像(拇指)

圖 4、不同灰階影像

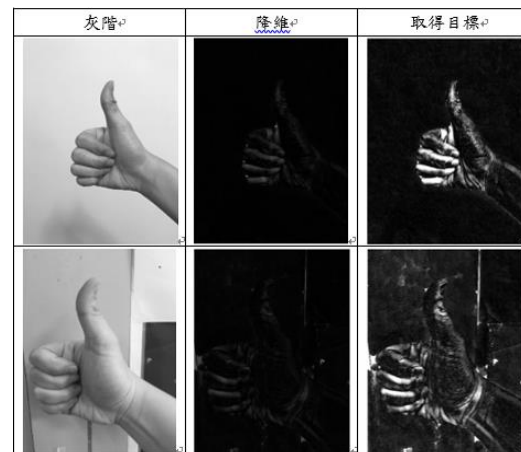


圖 5、影像處理結果

4.4 SVM 分類

影像資料的預先處理，將取得的影像陣列整理成一張圖包含 144 個欄位，其中每一個欄位代表著每一個像素值，而輸入的手勢分別為拇指、握拳以及掌心各十張影像，合計三十張圖，另外將最後一欄定義為手勢代表含意，例如拇指代表 1、握拳代表 2、掌心代表 3，圖 6 為 SVM 資料預先處理，將調整尺寸完成的影

像轉成灰階，再將陣列值依序放入表格，像素的輸入以欄優先依序放入，最後會形成表 1。

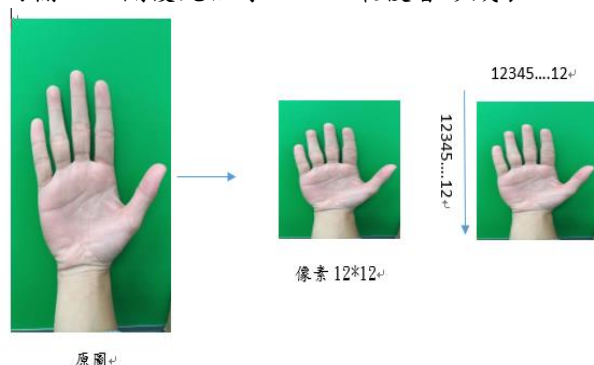


圖 6 像素放入表格以欄優先
表 1、SVM 資料預先處理

	Pix 1	Pix 2	...	Pix 143	Pix 144	Class
Image 1	123	54	...	99	244	1
Image 2	115	76	...	113	213	2
...	...	...	...	...	...	...
Image30	100	46	...	97	77	3

Class 當中的 1 代表拇指、2 代表握拳、3 代表掌心，像素的輸入以欄為優先

rr 進行 SVM 分類，將資料集的百分之七十進行 SVM 訓練模型，將剩下的百分之三十的資料集進行模型測試，本實驗使用的 kernel 為 radial，而最後會利用均方根誤差(RMSE)，公式(1)，來評斷測試結果。

本次實驗將每張影像 144 個像素當作特徵值，共三十張影像，表 2 當中的 class 欄位為正確值，表 2 為測試結果，當中的 pred 欄位為預測值，本實驗檢定方式使用 Accuracy，公式(1)，本實驗結果其正確率為：56%。

表 2、測試結果

Pix1	...	Pix144	class	pred	結果
237	...	217	3	2.638361	3
225	...	44	3	1.821807	2
244	...	216	2	1.795294	2
232	...	141	1	1.998807	2
247	...	85	1	1.088613	1
236	...	41	2	1.634095	2
231	...	226	2	1.444123	1
225	...	206	3	2.561585	3
240	...	110	3	2.138703	2

class 為實際值、pred 為預測值、結果為分類結果

$$Accuracy = \sum_{i=1}^n \frac{T_i}{n} \times 100\% \quad (1)$$

Accuracy 可以算出正確率，其值越大代表越準

確，其中 n 為樣本數， T_i 判定是否分類正確，當分類結果與 class 相等，則 $T_i = 1$ ，否則為 0。

5. 結論

蒐集了三種不同種類的手部影像各十張，以及利用 MATLAB 進行特徵擷取，最後使用 SVM 分類，並得出正確率為：56%。

本次影像分析的結果並不如預期，如果在背景單純的環境下執行，會得到不錯的效果，反之在複雜背景之下，效果並不彰顯，未來規劃會將雜訊降低加入研究重點。

SVM 部分，在未來會將預測結果進行正規畫，並且希望可以加入混淆矩陣來檢定以增加實驗的可靠性。

近年來，獨居老人的人口有逐漸增長的趨勢，為了解決老人因視線不明而跌倒之問題，希望將本實驗於樹梅派上實現，將樹梅派所擷取到的資訊，傳送至伺服器，並進行手勢辨識，回傳特定指令使的樹梅派可以控制燈光，以降低因視線不明造成危險的目的。

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] 張厥煒；林鈺皓，使用深度點雲技術之手勢手形比對門禁系統=A Hand Gesture and Shape Matching Access Control System Using Depth Point Cloud Technology 資訊、科技與社會學報 15:1/2 2015.12[民 104.12] 頁 79-95
- [2] A. Kuznetsova, L. Leal-Taixe, B. Rosenhahn, "Real-time sign language recognition using a consumer depth camera", *IEEE International Conference on Computer Vision Workshops*, 2013, pp. 83–90
- [3] Ana I. Maqueda*, Carlos R. del-Blanco, Fernando Jaureguizar, Narciso García, "Human-computer interaction based on visual hand-gesture recognition using volumetric spatiograms of local binary patterns", *Computer Vision and Image Understanding* 141, 2015, pp.126–137.
- [4] Quing Zhu, Ching-Tung Wu, Kwang-Ting Cheng, Yi-leh Wu, "An adaptive skin modle and its application to objectionable image filtering", *Proceedings of the 12th annual ACM International conference on Multimedia*, 2004

- [5] S. Malik, J. Laszlo, “Visual touchpad: A two-handed gestural input device”, *ICMI 04: 6th International Conference on Multimodal Interfaces*, 2004, pp. 289-296
- [6] Y. Wang, R. Yang, “Real-time hand posture recognition based on hand dominant line using Kinect”, *IEEE International Conference on Multimedia and Expo Workshops*, 2013, pp.1–4
- [7] Z. Ren, J. Yuan, Z. Zhang, “Robust hand gesture recognition based on finger-earth mover’s distance with a commodity depth camera”, *Proceedings of the International Conference on Multimedia, ACM*, 2011, pp. 1093–1096