

利用 Kinect 平台建構互動式旗語學習系統

紀慶和¹ 張嘉文^{1*} 顧皓翔²

¹銘傳大學 電腦與通訊工程學系

²華夏科技大學 資訊工程學系

cwchang@mail.mcu.edu.tw

摘要

互動式技術的發展，人們可更輕易地操控各式設備，因此許多研究透過 Kinect 技術捕捉使用者全身骨架動作，用肢體語言來進行控制並獲得相關服務。有鑑於此，本研究規劃與設計一互動式旗語學習系統 (Kinect-based Interactive Semaphore-learning System, KISS)，並建構於 Raspberry Pi 開放式平台上，主要包括旗語辨識子系統 (Semaphore Identification Sub-system) 與自我學習子系統 (Self-learning Sub-system)。此外，因雙旗旗語並不會因地區或文化而有所不同，故本研究利用雙旗旗語為人機溝通之語言。而互動式旗語學習系統藉由 Kinect 擷取使用者的骨架資訊後，經由一系列的演算法和類神經網路取得結果，再藉由互動系統與使用者互動，可即時了解使用者自身的姿勢是否正確，進而達到學習雙旗旗語之功效，具有高彈性和容易攜帶之優點。

關鍵詞：互動式旗語學習系統、雙旗旗語、旗語辨識、自我學習系統、骨架資訊。

Abstract

Kinect can capture the user's skeleton data, and then use the data to interact and control systems. Hence, this study designs and proposes a Kinect-based Interactive Semaphore-Learning System, which is called KISS. KISS is built on Raspberry Pi and composed of Semaphore Identification Sub-system and Self-learning Sub-system. Furtherer, Semaphore is a kind of universal signals. It is easy to be a control signal. KISS will get results after analyzing skeleton data from Kinect, and then checking the user's position is correct or not. It with high flexibility and capability can make a user easily learn semaphore using KISS.

Keywords: Interactive learning semaphore system, Semaphore, Kinect, Self-learning and skeleton data.

1. 前言

人機互動介面無論在消費型電子、智慧型家電或工業控制等領域皆為重點研發項目，其目的是方便

使用者與系統之溝通與操控，屬於一門整合機電、自動控制與資訊科技等技術之研究，目前國內外學、業界無不投入大量的研發產能與資源來開發各領域所需的人機互動介面，如：iPhone 的互動就是透過觸碰螢幕來達到人機互動，而微軟所發表的 Kinect 就是透過攝影機和紅外線來擷取人體資訊而做到高親和度的人機互動系統。目前國內外大廠均已發表各式親和度高的人機互動產品，例如由 Leap Motion, Inc 所開發的 Leap Motion, 用手就可以取代滑鼠, 而且可以任意抓取三維資料[6]。各項產品或研究使機器能快速整合至家庭、公共場所，進行保全、防災、看護及各種服務工作，大幅降低學習曲線[4][5]。

傳統人機互動介面會因為開發商或是研究者的不同，而操作動作會有極高的差異性，例如：同為觸碰輸入介面，Apple iOS 和 Google Android 對於手勢動作上的定義有著很多的差異存在。人機互動介面與社會大眾及社群產業的關係日趨密切，如何提高辨識的準確度與溝通語言通用為相當重要之議題。本論文構想設計與規劃一套互動式人機介面系統，未來透過整合及學習即可容易整合至其他各平台，有鑒於此，此類系統應考量以下議題：

- (1) 整合性(Integrity)與互通性(Interoperability)。不同廠商或實驗室會開發出不同人機互動溝通語言，一個平台很難同時整合多種溝通語言，因此透過開源的 OpenKinect 做為感應器的溝通平台，並利用同樣為開源的標準函式庫 Open CV，並使用雙旗旗語作為溝通語言，達到通用互動溝通語言之成效。
- (2) 攜帶性(Portability)。本研究利用英國的樹莓派 (Raspberry Pi)，因該平台不論體積重量都比一般主機輕巧，從而可以輕易地帶往各地點，符合本計畫方便移動之成效。
- (3) 辨識率(Recognition Rate)。本研究利用自組織映射圖網路(Self-Organizing Map)學習網路，而其特色在於模仿人類大腦中，類似功用的神經元會群聚在一起特性，可對大量學習樣本使用聚類分析，進而得到該所屬之群體。
- (4) 互動性(Interactivity)。透過應用軟體的回饋，可以了解自己的姿勢是否正確無誤，如有錯誤也可即時修正，進而增加使用者對於雙旗旗語的記憶能力。
- (5) 親和度(Affinity)。由於現今最直接的最簡單的語言，就是肢體語言，只需要使用雙手，即可

輕鬆表達出 26 個英文字母和數字。且肢體語言不需要太多媒介，如：觸控螢幕、按鍵、滑鼠等設備。即可降低整體硬體設備支出費用。有鑑於此，本研究規劃與設計一互動式旗語學習系統 (Kinect-based Interactive Semaphore-learning System, KISS)，並建構於 Raspberry Pi 開放式平台上，主要包括旗語辨識子系統 (Semaphore Identification Sub-system) 與自我學習子系統 (Self-learning Sub-system)。

在本研究的第 2 節將相關研究做一整理與說明；第 3 節將所提出之動式旗語學習系統架構與設計做一論述；第 4 節將展示相關實作畫面；並於第 5 節做一結論。

2. 文獻回顧與探討

近年來人體辨識系統之研究多數都在提升精準度的方面，而改良方式約略分為三種。(1) 人機互動媒介：透過擷取人體資訊重點的不同，可以達到不同的控制效果，現今的取得人體資訊的重點大概圍繞在語音、觸控、臉部和肢體。對於人機互動方面的研究中許宏榕學者運用 Webcam 作為影像擷取器然後使用 IRBNN 作為分類器，分類各種表情及臉部狀態 [12]。而劉忠信學者利用連續影像相減法來找出前景移動物體，當物體在移動時，前後時序差的影像相減後會偵測出物體移動的邊緣，利用物體移動的邊緣來建構物體外形的資訊 [15]。S.S. HUANG 學者透過結合無線感測器改善 Kinect 先天性的偵測弱點，針對當部分肢體平貼於地面或有肢體重疊的姿態進行姿態辨識 [13]。(2) 人體姿態辨識：在於傳統透過 Webcam 取得 2D 資訊來分析人體姿勢，到現今出現結合 Webcam 和紅外線的 Kinect 來取得 3D 骨架資訊，進而大大增加辨識成功率。對於影像擷取、設備與影像演算法方面的研究中，Y.M. Chen 學者運用 KINECT 取得深度影像 (Depth Image)，取得人體關節結構，由人體各關節轉化為方向 (Orientation) 以作為姿態上特徵 [1]。C.S. Fahn 等學者運用 webcam，並使用粒子濾波器 (particle filters) 演算法來偵測物體的動作，然後藉由移動來排除物體的辨識干擾，然後使用自適應增強演算法 (Adaptive Boosting) 法取得最正確的辨識結果 [2]。(3) 類神經學習演算法：透過學習演算法讓機器可以自我排除誤差，而在近一步的提升辨識的成功率。對於類神經學習演算法的研究中，多數採用徑向基底函數類神經網路 (Radial Basis Neural Network, RBNN) 和自組織映射圖網路 (Self-Organizing Map, SOM) 作為解決途徑，而 RBNN 為前饋式類神經網路，其特色在於模仿人類大腦中的局部調整功能。而架構為三層，輸入層，隱藏層和輸出層，而運作方式是當資料進入輸入層的時候，會經過函數轉換輸入到隱藏層，最後輸出層再將所有的隱藏層相加成後輸出。SOM 為非監督式學習網路，其特色在於模仿人類大腦中，類似功

用的神經元會群聚在一起的特性。因此，只要可以對大量學習樣本使用聚類分析，進而得到該所屬之群體，就適用自組織映射圖網路。如張家鳴學者運用輪廓的水平與垂直投影向量以及離散傅立葉轉換 (DFT) 來取得投影向量的特徵值，並且配合物體的長寬比值，以這組數值代入 SONFIN (自我組織學習模糊類神經網路) 取得結果 [14]。陳高斌學者將語音特徵輸入 SOM，讓其訓練向量在經過學習的過程後，自動產聚類產生若干個聚類中心，然後將這些特徵向量當作是 PNN 的特徵單元，最後以此機率類神經網路分類模型作為語者辨識的分類器 [11]。Y.M. Wu 等學者結合 RBNN (Radial Basis Neural Network) 和 LLE (Locally linear embedding) 非線性降維的方法，不僅可以減少數據維度和計算複雜性，且保留各類型的面部拓撲的良好樣本，以及避免臉部照明，姿勢和其他因素 [8]。

由上述的相關研究中，可以了解大部分的研究對於精確度具有相當深入，但所有平台必須建構在主機上，而傳統主機體積過大且不方便移動，導致限制了平台的使用地點及範圍，且辨識結果會因應各實驗室不同，而有不同的結果。故本研究規劃與設計一互動式旗語學習系統 (Kinect-based Interactive Semaphore-learning System, KISS) 進行人機語言整合和辨識平台化簡，整合使用者 (User)、感測器 (Sensor)、硬體平台 (Hardware platform)、輸出設備 (Output Device) 進行四方的互動情境，達到可以隨地使用辨識平台且降低文化上對辨識平台的影響。

3. 系統架構

互動式旗語學習系統 (Kinect-based Interactive Semaphore-learning System, KISS)，主要使用 Kinect 作為感應器，因該設備整合了攝影機和紅外線，可以快速地得到三維空間影像。而後搭配建構在 Raspberry Pi 平台上的 Open CV 標準函式庫作為影像辨識之用，而後搭配自組織映射圖 (Self-Organizing Map) 類神經網路作為提高辨識率。使用者 (User)、感測器 (Sensor)、硬體平台 (Hardware platform)、輸出設備 (Output Device) 進行四方的互動情境，其架構圖如圖 1 所示。透過此一整合型運作模式，能使人機互動平台在本體結構減少的前提下，由感測器達到姿勢偵測、由硬體平台達到互動推論、由輸出設備達到人機互動之成效。為了達成上述之議題，本計畫將建立一個由 ARM 作為處理晶片的平台，透過 Linux 高互通性和高效率運作環境。主要包括旗語辨識子系統 (Semaphore Identification Sub-system) 與自我學習子系統 (Self-learning Sub-system)。

3.1. 規劃與建構旗語辨識子系統 (Semaphore Identification Sub-system)

旗語辨識子系統主要負責建構建構所使用之旗

語與相關情境，此外並須有效整合於開放式平台之上。故於此議題下，主要規劃與設計(1) 選擇所使用之旗語、(2) 進行 Raspberry Pi 開放式平台開發、

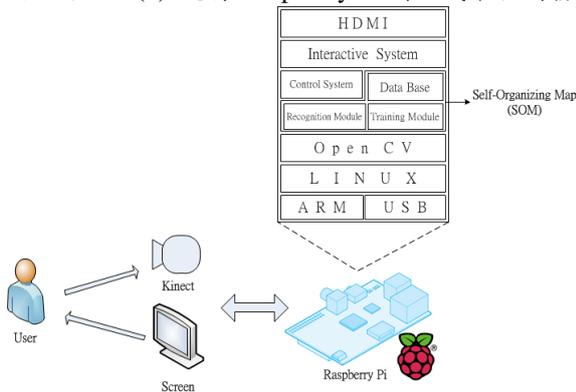


圖 1 互動式旗語學習系統(Interactive Learning Semaphore System)

以及與(3) 互動系統開發。

(1) 旗語選擇

旗語分成單旗旗語與雙旗旗語。單旗旗語的旗手雙手握住一面方旗傾斜四十五度角，而揮動的幅度約是八十度和一百三十度兩種。單旗旗語使用的語言是摩斯密碼，當要傳送點時揮動八十度，而線的時候就是揮動一百三十度。而雙旗旗語的旗手雙手各拿一面方旗，每隻手可指 7 種方向，除了待機信號之外，兩旗不會重疊。藉由兩手所指的方向來辨別出所表示的字母或數字，如圖 2 [9]。旗語在國際上具有高互通性及統一性，不會因為國家或文化上的差別，而有不同的姿勢產生，進而可以簡化學習曲線。

圖 2 雙旗旗語所代表之意義

雙旗旗語具有辨識容易，而單一字母只需要一個動作即可完成，可以大幅減少訊息傳遞的時間，而具有高時效性，且因一個動作即可完成字母傳遞，所以在傳遞連續訊息方面，也較單旗旗語容易。

(2) Raspberry Pi 開放式平台

開發平台為英國樹莓派基金會所開發的 Raspberry Pi，如圖 3 所示[7]。Raspberry Pi 具有(i) 取得門檻較一般大型主機低，容易推廣給教學資源不易取得的地點。(ii) 該平台因體積輕巧很多，可以輕易地帶往各地點，具高可攜帶性。(iii) 該平台為開放式平台，可自行修改參數較一般平台多，能對個別地區做微調。(iv) 具有社群與基金會維護該平台，可容易獲得新資訊。



圖 3 Raspberry Pi 開發平台

而為了解決系統通用性的問題，本研究使用由 Intel 開發的 Open CV 開源電腦視覺庫。由 C 和 C++ 構成，讓電腦視覺和圖像方面很多通用的演算法。Open CV 擁有多個 C 函數的跨平台的中層或高層 API。它不須使用其它的外部函式庫，但如果需要也可以使用外部函式庫。而該函式庫可以裝載在任意的電腦系統上，如：Windows， Mac OS， Linux， FreeBSD， OpenBSD， iOS， Maemo 和 Android 等平台上運行。有鑒於此，結合 Raspberry Pi 開放式平台和 Open CV 函式庫，做為人體辨識系統，將由 Kinect 所取得的影像資訊藉由 Open CV 函式庫做運算，最後得出的結果送往下一步自我學習系統或是互動系統，其流程如圖 4 所示。

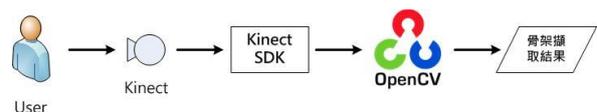


圖 4 辨識系統流程圖

(3) 互動系統開發規劃

圖 5 為互動式情境流程圖，一開始互動系統會要求使用者擺出所指示字母的旗語動作，經由辨識模組辨識完畢後，便傳送到互動模組。而後互動模組就會根據指示字母是否與使用者所擺出的旗語字母相同，若相同就繼續下一題直到使用者結束訓練。若比對結果不相似，則互動系統將視為錯誤動作，而後要求使用者重做或是教學使用者正確動作為何。

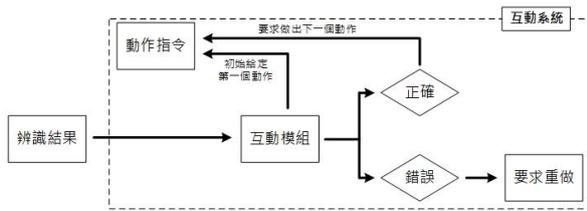


圖 5 互動系統架構圖

3.2. 規劃與建構自我學習子系統(Self-learning Sub-system)

每個人的在動作上都會有細微的誤差，然而如何排除誤差也是重點之一，例如:N, R, M, S 這四個字母的動作相似，所以在判讀上就會有問題。其推論核心為類神經網路，並建構適適合解決誤差的方法，所以利用自組織映射圖網路(Self-Organizing Map, SOM)於姿態識別之差異。因不同人的動作都會有些微誤差，但每一個動作都可以以聚類分析取得該所屬的群組，進而將誤差排除，得到最後所需要的結果，SOM 為非監督式學習網路，該特色在於模仿神經元會群聚在一起的特性。因此，只要可以對大量學習樣本使用聚類分析，進而得到該所屬之群體，所以本計畫就適用自組織映射圖網路做為自我學習網路。SOM 的運作方式分為兩階段，分別為訓練階段和映射階段。由訓練樣本建立一個 SOM 網路，在與輸入樣本相映射，得出最正確的結果。其 SOM 執行相關步驟與程序如下，如圖 6 所示：

- Step 1. 初始化：初始化所有神經元的權重值，給予隨機的極小數值。
- Step 2. 網路連接：將輸入層的所有神經元與 SOM 的神經元做連結。
 - Step 2-1. 使用向量指標，找出輸入向量與 SOM 的神經元之前的權重向量的相似性。
 - Step 2-2. 找出輸入向量與 SOM 之間最接近的神經元。
- Step 3. 藉由輸入向量找出最接近的神經元，調整最佳匹配單元(Best Matching Unit, BMU)與其鄰近範圍內的神經元的權重值，調整公式如下。

$$\omega_v(t+1) = \omega_v(t) + \theta(v,t)\alpha(t)(D(t) - \omega_v(t))$$

- Step 4. 隨時間 t 增加，重複 Step2，直到 $t > \lambda$ ， λ 為閾值。

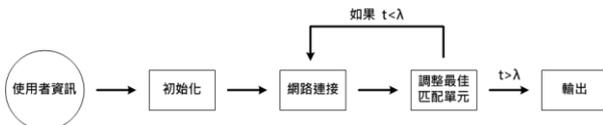


圖 6 自組織映射圖網路執行流程

4. 系統實作

本章節將展示互動式旗語學習系統 (Kinect-based Interactive Semaphore-learning System,

KISS)，本研究利用開放式作業系統 Linux 與 KDevelop 作為互動系統開發軟體。KDevelop 用以開發 GUI 應用軟體整合開發環境 (Integrated Development Environment, IDE)，環境外觀如圖 7 所示。當進入程式時會需要等待 Kinect 連接至電腦系統，但因 Raspberry Pi 效能因素，需要一點時間作初始化動作，等初始化完成後，則立即開始訓練過程，

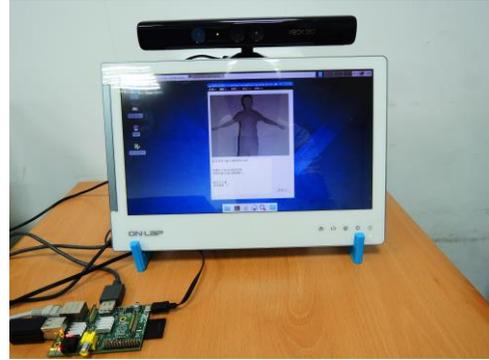


圖 7 互動式旗語學習系統外觀

如圖 8 所示。而互動式旗語學習系統會針對使用者的動作判別是否正確，當使用者的動作不精確或是完全錯誤時，會要求使用者重新動作，如圖 9 所示。



圖 8 互動式旗語學習系統初始化完成



圖 9 學習系統發現使用者動作不正確

5. 結論

本論文結合了開源的 Open Kinect 與方便攜帶的 Raspberry Pi 規劃與建構出一互動式旗語學習系統 (Kinect-based Interactive Semaphore-learning System, KISS)。學習系統藉由 Open Kinect 驅動 Kinect 能在 Raspberry Pi 使用之後，再由 Kinect 紅外線感應器所回傳的使用者影像深度，計算出使用者的雙手與使用者中心的座標，再透過演算法取得最終結果，接下來藉由互動系統即時了解動作是否正確，最終達到學習旗語的功效。有效解決與達到整合性 (Integrity) 與互通性 (Interoperability)、攜帶性 (Portability)、辨識率 (Recognition Rate)、互動性 (Interactivity) 與親和度 (Affinity) 之議題。

而未來的研究方向會繼續朝向(1)系統對於組合旗語的研究規畫以及開發，而不只單單侷限於單一字母的表示。(2)增加互動系統的有趣度，讓使用者增加學習旗語上的興趣。

6. 致謝

本研究承蒙科技部研究計畫予於補助，(計畫編號：MOST 103-2221-E-146 -011 與 103-2815-C-130 -104 -E)，得以順利完成此研究，在此獻上最誠摯的謝意。

參考文獻

- [1] F. Buccolieri, C. Distanto, A. Leone, "Human Posture Recognition Using Active Contours and Radial Basis Function Neural Network," Proceedings of IEEE Conference on Advanced Video and Signal Based Surveillance, pp. 213-218, 2005.
- [2] C.S. Fahn, S.L. Chiang, "Real-Time

Upper-Limbs Posture Recognition Based on Particle Filters and AdaBoost Algorithms," Proceedings of 20th IEEE International Conference on Pattern Recognition, pp. 3854-3857, 2010.

- [3] H. Hikawa, S. Matsubara, "Pseudo RBF Network for Position Independent Hand Posture Recognition System," Proceedings of International Joint Conference on Neural Networks Neural Networks, pp. 1049-1054, 2007.
- [4] N. Iwane, N. , "Arm Movement Recognition for Flag Signaling with Kinect Sensor," Proceedings of IEEE International Conference on Virtual Environments Human-Computer Interfaces and Measurement Systems, pp. 86-90, 2012.
- [5] H.J. Lee, S.H. Hwang, S.M. Lee, Y.G. Lim, K.S. Park, "Estimation of Body Postures on Bed Using Unconstrained ECG Measurement," IEEE Journal of Biomedical and Health Informatics, Vol. 17, No. 6, pp. 985-993, 2013.
- [6] Leap Motion, <https://www.leapmotion.com>, 2014.
- [7] Raspberry Pi, <http://www.raspberrypi.org>, 2014.
- [8] Y.M. Wu, L. Liu, N. Li, "Research of Face Recognition Based on LLE and RBF Neural Network," Proceedings of 2010 IEEE International Conference on Electrical and Control Engineering, pp. 1605-1608, 2010.
- [9] Wikipedia, free encyclopedia. <http://zh.wikipedia.org/wiki/%E6%97%97%E8%AA%9E>, 2014.
- [10] 陳佑銘。應用權重三維骨架關節於動作辨識之研究。碩士論文。台南：國立成功大學電腦與通信工程研究所。2012。
- [11] 陳高斌。應用 SOM-PNN 混合神經網路在語者識別。碩士論文。高雄：義守大學電機工程學系。2012
- [12] 許宏榕。適應性中心點選擇半徑基底函數委員會機器於人臉識別。碩士論文。雲林：雲林科技大學資訊工程研究所 2008。
- [13] 黃秀珊。2013。多模態體感動作辨識系統之人機介面研究。碩士論文。桃園：國立中央大學資訊工程學系
- [14] 張家鳴。使用模糊類神經網路辨識人體姿態以及姿態估測。碩士論文。台中：國立中興大學電機工程學系。2005。
- [15] 劉忠信。肢體動作辨識技術研究。碩士論文。桃園：國防大學理工學院資訊工程學系。2010。