

Easy-Guide 以智慧型手機實現之簡易導盲系統

Easy-Guide: A Simple Smart Phone-based Guiding System for Visually Impaired People

專題組員: 蕭勝興、李政哲、楊朝勳、黃昶翔

專題編號: NTPUCSIE-104-008

執行期間: 2015 年 08 月 至 2016 年 06 月

1. 摘要

如何以科技解決視障人士導航及迴避障礙物是一直以來很熱門的話題。視障人士在一個熟悉的環境行走時，很難知悉前方突發的路況，且當視障人士行走至一個陌生的環境時，更需要有旁人指引，侷限的視野也增加了行走時遭遇障礙物的風險。

本計畫提出一套系統來解決視障人士導航及迴避障礙物的問題，使視障使用者從本校校園內部走至大門公車站，對路途環境有更多的認識。本系統希望透過電腦輔助影像辨識並結合手機 APP，製作一套簡易的輔助導引系統。透過智慧型手機取得使用者面前的影像，並將這些影像傳到後端系統進行處理，經由超快速區域性卷積類神經網路演算法 (Faster Region Convolutional Neural Network, Faster R-CNN) 及你只需要看一次 (You Only Look Once, YOLO) 演算法辨識出前方多種障礙物，最後將結果回饋至智慧手機，透過深度學習演算法，本系統可達到 53% 的辨識率，成功增進視障者對前方環境的認識。

2. 簡介

過去也有不少針對前方環境給予

警示及盲人導航的相關研究，根據裝置類別及辨識方法可分為以下兩類：(a) 深度攝影機辨識系統，以及 (b) 改裝過之導盲杖，此兩類做法各有利弊，以下將針對此部分來介紹。

[1]-[2] 採用 Microsoft Kinect 做為系統的主要辨識硬體，但是 Kinect 無法在光線強烈的地方使用，用 Microsoft Kinect 所提出的相關應用，僅能辨別出前方是否有障礙物，無法識別障礙物種類。

[3]-[4] 則是另一種較常被實作的方式，其中又可分為影像形式與感測形式。在影像形式的例子中，將網路攝影鏡頭裝設在導盲杖末端上，透過辨識攝影鏡頭所抓取的影像，來給予使用者相關警示；在感測形式的例子中，則是在導盲杖末端裝設 RFID 感測模組，並在環境周圍鋪設大量的 RFID 晶片，透過感測相對應的 RFID 晶片來達到定位及導航目的，但此類改裝系統並不普及，且很難被運用於實際生活情況中。影像形式所辨識的影像，僅限於特定場合，若擷取之影像並沒有相對應的導盲物件，系統便無任何作用；感測形式所應用的環境，需事先鋪設大量的 RFID 晶片，當使用者走到一個沒有 RFID 的環境中時，此

系統也無法發揮功能。

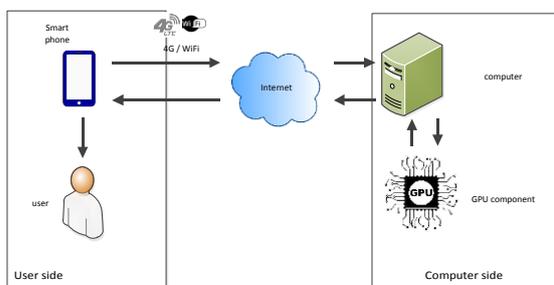
本計畫提出一個以行動裝置為主的視障導盲系統，並透過深度學習演算法，辨識前方的多種障礙物資訊，本系統的使用環境不僅限於特定室內場合，也不須特別鋪設 RFID 晶片，除了增加導盲系統的使用環境外，也提供視障人士更多資訊。

3. 專題進行方式

針對系統架構，本段落將分成系統概況、軟體流程及軟體模組進行介紹。

3.1 系統概況

系統架構圖如圖一所示，本系統分成手機端與電腦端。手機端為配有後相機鏡頭及網路傳輸模組之智慧型手機，本計畫以 Infocus M320 來進行開發與測試；電腦端則配有可運行 CUDA 的 GPU，本計畫以 GTX 640 來進行深度運算。



圖一、系統架構圖

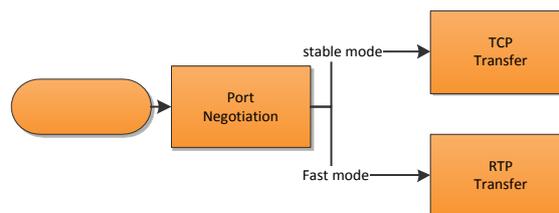
3.2 軟體流程

本系統的運作流程。運作流程如下：1) 智慧型手機會不斷抓取前方的影像，2) 將影像送入手機辨識模組中進行人臉辨識及樓梯辨識，3) 將影像傳送至遠端電腦中，4) 將影像送入深度辨識模組得到辨識結果，並計算出方位及距離，5) 將辨識結果、方位及

距離資訊回傳至手機，6) 透過語音告知使用者。當使用者行經至網路訊號不夠的地區時，手機連線程式會自動斷線，並不斷嘗試重新連線，在此期間手機仍會繼續運行手機辨識模組。

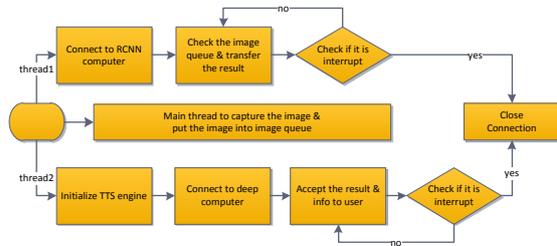
根據視障使用者在本校行走的情形，可能有以下兩種情況：1) 視障使用者剛來到本校，並不熟悉本校環境，則使用本系統時，需要較高的辨識精準度。2) 視障使用者在人潮較多的時間及地點行走時，會有較高機率遇到障礙物，例如學生將腳踏車隨意停在建築物門口，或是狗在路上休息，此時需要較快的速度告知使用者前方狀況。針對以上敘述的兩種情形，本系統提供穩定傳輸模式(Stable Mode)和快速傳輸模式(Fast Mode)，以下將講述本系統的軟體流程，並特別講述兩種模式之異同。

手機運行流程如圖二所示。當使用者開始使用本裝置後，一開始會先執行埠協商程序，之後便根據視障使用者的設定，來執行相對應的協定傳輸。在本系統中，穩定傳輸模式是使用 TCP 協定進行傳輸，快速傳輸模式則是使用 RTP 協定進行傳輸。由於一張圖片的大小太大，無法僅用一個 RTP 封包傳送，因此本系統將圖片拆解成數個 RTP 封包，封包傳送至電腦端後，透過時戳得知封包順序，並還原成原本影像來進行辨識。

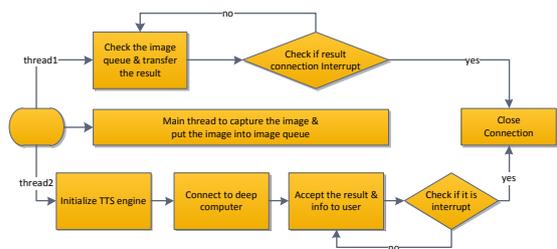


圖二、手機軟體流程圖

切換進入穩定傳輸模式後，手機會對電腦發出影像傳送連線請求和結果接收連線請求，連線建立以後，即開始將影像送至電腦端進行辨識，同時接收電腦回傳的辨識結果，其運作流程如圖三所示；進入快速傳輸模式以後，和穩定傳輸唯一不同的地方是，手機並不會對電腦發出影像傳送連線請求，而是直接傳送 RTP 影像封包。當偵測到手機斷線後，影像辨識程序便會停止，其運作流程如圖四所示。所有連線停止後，手機會重新進行阜協商。



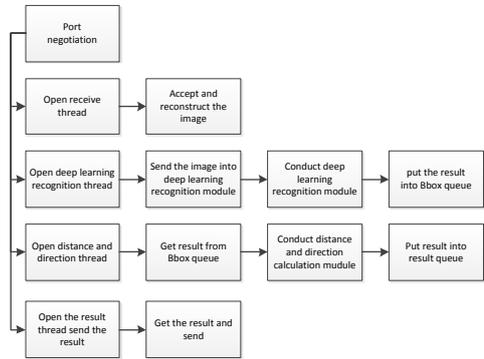
圖三、穩定傳輸模式流程圖



圖四、快速傳輸模式流程圖

圖五為電腦端的軟體流程圖。程式會不斷地監聽遠方手機端的連線請求，當收到請求後，手機會和電腦進行阜協商，決定工作運行的連線阜號碼。結束阜協商後，電腦會新增四個線程，分別接收手機傳送的影像、執行深度辨識模組、執行方向及距離計算模組及傳送辨識結果。當手機斷線

後，電腦將結束新增的四個線程，僅剩下阜協商線程，繼續等待下一個使用者的連線。



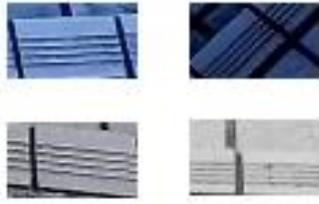
圖五、電腦端軟體流程圖

3.3 軟體模組

針對軟體模組，本段落將分成特徵辨識模組、深度辨識模組及方向距離模組依序進行介紹。

3.3.1 特徵辨識模組

經過了對視障使用者的訪談後，本系統決定以較簡單的影像辨識演算法，針對人、樓梯及墩柱進行辨識。當使用者在使用本系統時遇到異常斷線，手機內部的離線辨識功能將繼續運行。因為手機的運算資源及記憶體遠低於電腦，手機無法順利快速運行深度學習演算法，因此本系統採用哈爾特徵(Haar Feature)分類及方向梯度值方圖(Hog Feature)特徵分類實作離線影像辨識，在快速模式中，為了加快辨識速度及整體運行速度，本系統利用 NDK 實作特徵辨識模組，並透過離散型自適增強演算法 (Discrete Adaboost Algorithm)，訓練哈爾特徵分類器，正樣本如圖六所示。



圖六、訓練之正樣本

當智慧型手機擷取前方影像後，影像會被送入特徵模組，其模組執行過程如下：1) 複製擷取後的影像，原始影像將送往遠端電腦，複製之影像將顯示特徵辨識結果 1) 透過以自適增強演算法訓練而成之分類器辨識樓梯，2) 透過方向梯度值方圖訓練而成之分類器辨識人臉，3) 將結果圈選並繪製在複製影像中，其流程如圖七所示。



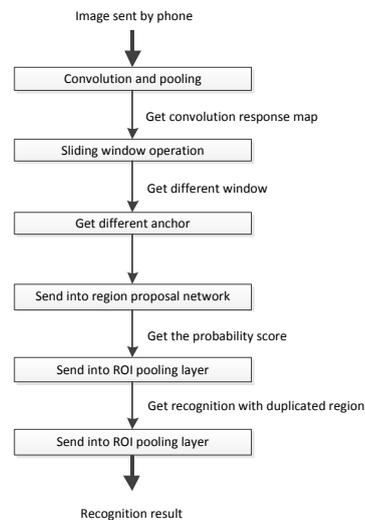
圖七、特徵模組運行流程

3.3.2 深度辨識模組

為了以系統性的方式達到多種物件的辨識目的，本系統採用深度學習(Deep Learning)方式，透過機器自動學習並擷取特徵值，大幅度縮短物件特徵之擷取時間。本系統針對若干個深度學習演算法[8]-[12]進行評估後，決定以 Faster R-CNN 和 YOLO 演算法來進行系統性的物品辨識。本計畫考慮了兩種演算法的優缺點，在穩定傳輸模式將使用 Faster R-CNN 網路架構進行辨識，快速傳輸模式則使用 YOLO 網路架構進行辨識。為了提升卷積神經網路的計算速度，因此本計畫使用統一計算架構 (Computer Unified Device Architecture, 簡稱 CUDA) 來進

行大量矩陣運算，本系統使用的 CUDA 環境為 CUDA 7.5。

在穩定傳輸模式中，深度辨識模組的運行流程如下：1) 影像接收執行接收影像後，會將影像傳送至模組中，2) 將影像送入卷積神經網路模型中，得到此影像之卷積響應圖，3) 對卷積響應圖進行滑動窗口掃描，得到 j 個窗口，4) 對 j 個窗口選取不同大小及方向比例的錨(Anchor)，5) 將每一個錨送入區域分類層，得到候選區域是物體及不是物體的分數。6) 將座標值對應的卷積響應圖送入 ROI 池化層，7) 對辨識結果運行非極大值抑制演算法，最後得到辨識物體的種類及座標值，其運行過程如圖八所示。本系統以 Caffe 作為主要開發的深度學習框架。



圖八、穩定傳輸模式深度辨識運行流程

在快速傳輸模式中，深度辨識模組的運行流程如下：1) 影像接收執行接收影像後，會將影像傳送至模組中，2) 將影像送入卷積神經網路模型中，得到此影像之卷積結果，其運行流程如圖九所示。

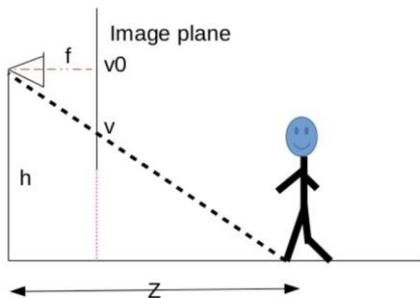


圖九、快速傳輸模式深度辨識運行流程

3.3.3 方向距離模組

影像透過深度辨識模組辨識以後，會得到障礙物的種類編號、物體起始座標及物體寬高。方向距離線程會不斷將辨識結果輸入進方向距離模組，而此模組之輸出為障礙物的名稱、障礙物的方向及距離。

在距離辨識部分，本系統透過幾何學來推測距離，如上圖十所示。其中 f 為相機鏡頭焦距， k_v 為像素密度 (Pixel Density)，單位為 pixel/meter， h 為相機鏡頭高度， v_0 為成像之中心坐標， v 為目標物在成像中與地平面交接之坐標， Z 為障礙物與使用者之距離。此公式利用當相機垂直於地面拍攝時，成像在無窮遠處將收束於中心，所以所有物體與地平面之交接必定位於成像中心以下，通過找出與地平面交接之坐標來推算圖十之相似三角形，本系統得以在不考慮目標物之大小下得出所求距離 Z 。公式(1)為距離計算公式。



圖十、距離計算示意圖

$$Z = \frac{f * k_v * h}{v - v_0} \quad (1)$$

在方向辨識部分，此模組將影像平均分成左中右三個等分，並考慮障礙物的中心點座標，位於哪一個等分之一中，因而會得到左、前或右三種可能的方向輸出，經過此模組後，障礙物的名稱、障礙物的方向及距離會被放置傳送佇列中，由傳輸線程將結果回傳至使用者手機。

4. 主要成果與評估

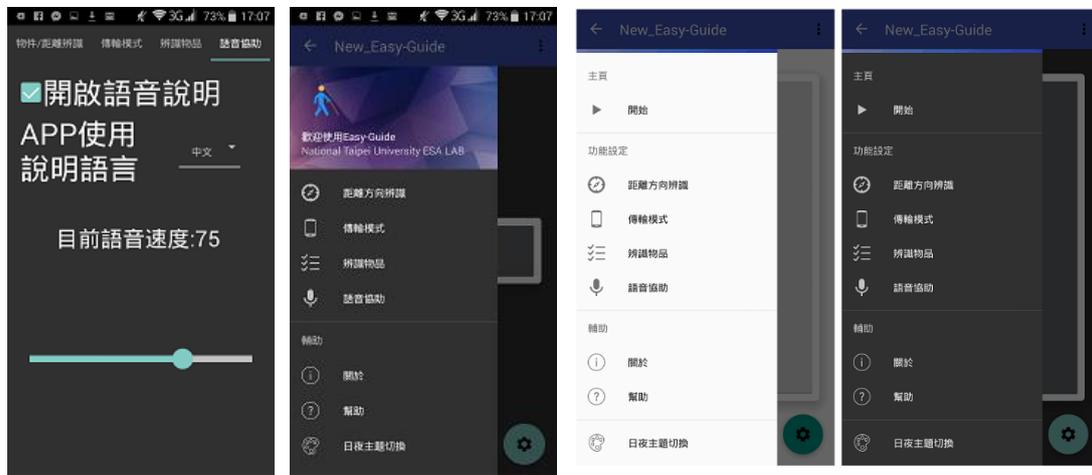
本段將分為手機程式與辨識成果兩個小節進行闡述。

4-1. 手機程式

使用者端使用的介面程式主要在 Android 上以 Java 開發，使用者首次開啟 APP 後，手機會唸出此 APP 的基本使用方式，讓初次使用的使用者不會因為操作不易而浪費時間。進入 APP 主畫面如圖十一左邊所示，由於使用者在進行設定的次數遠低於執行辨識功能的次數，因此將大量空間配置為開始使用按鈕。使用者點選主畫面右下方的藍色按鈕後，隨即進入設定介面，如圖十一中間所示。



圖十一、主畫面與部分設定頁面



圖十二、側邊導引設定與日夜主題選擇畫面

進入設定頁面後，使用者可以左右滑動手機視窗或是點選特定標題，來進行相關設定。在此頁面中可以選擇是否要聽取方向或距離的資訊。傳輸模式頁面中可以選擇快速傳輸模式或穩定傳輸模式。若選擇穩定快速傳輸模式，則手機會透過 TCP 傳輸協定將影像傳輸至影像辨識電腦；若選擇快速傳輸模式，則手機會透過 RTP 傳輸協定將影像傳輸至影像辨識電腦。

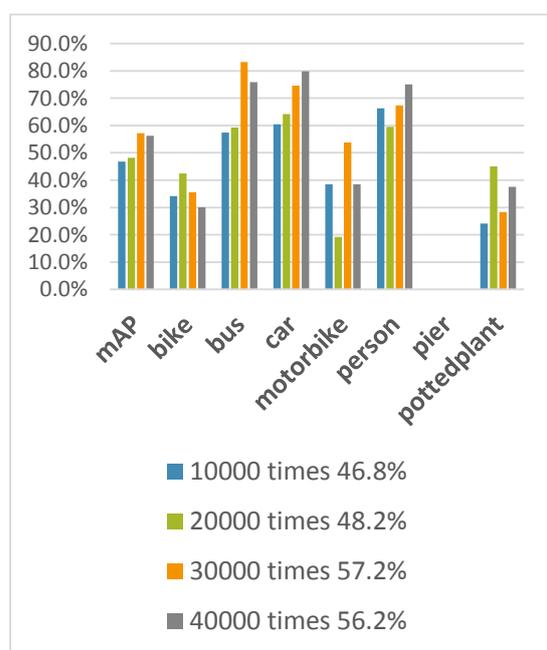
在第三個設定頁面中，使用者可以設定想要了解的障礙物種類名稱；最右邊的設定頁面中，使用者可以設定語音說明的語言，目前本系統提供

中文與英文兩種選擇；在此頁面中也可調整語音設定的說話速度，及是否關閉語音設定。在主畫面中，使用者可以觸及畫面左方向右拖曳，即可開啟側邊導引設定。可點選上面四個選項進行上述的功能調整，其畫面如圖十二所示。為了滿足視障不同使用者的視覺習慣，本系統提供黑底白字和白底黑字兩種顏色介面使用者可依自己適合的習慣來調整。

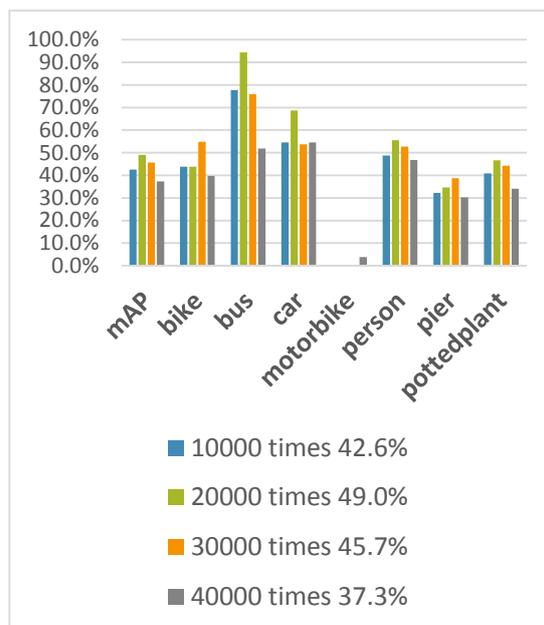
4-2. 辨識成果

為了更符合使用者的需求，本計畫在經過了與視障者訪談以後，期望

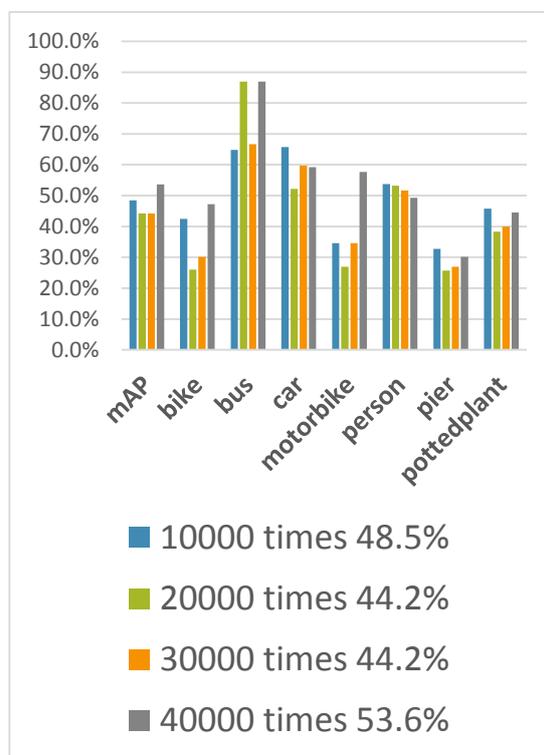
本系統能成功辨識出前方的人、車、公車、摩托車、腳踏車、盆栽、狗、樓梯及墩等九種障礙物。本系統使用一個已用 ImageNet 圖片庫訓練過的網路模型來進行訓練，並計畫透過 PASCAL 圖片庫進行二度訓練，但 PASCAL 圖片庫有兩項缺點：1) PASCAL 中的 20 項類別中並沒有包含墩。2) PASCAL 圖片庫的圖片來自世界各地，但並沒有針對本校周遭環境拍攝，且用手機拍到的照片品質，也會影響辨識結果。基於上述原因，本計畫拍攝校園內部及周圍共 2000 照片，並逐一標示障礙物位置及種類。分成三種情況進行訓練：1) 僅使用 PASCAL 圖片庫進行訓練。2) 僅使用自行拍攝的 2000 張圖片進行訓練。3) 使用 PASCAL 圖片庫及 2000 張圖片進行訓練，其得到的各種物品之辨識率分別如圖十三至十五所示。



圖十三、僅使用 PASCAL 圖片庫訓練之模型測試結果



圖十四、僅使用自行拍攝的 2000 張圖片訓練之模型測試結果



圖十五、PASCAL 圖片庫及 2000 張圖片訓練之模型測試結果

5. 結語與展望

本計畫整合了傳統的哈爾辨識演算法及兩個深度辨識演算法，開發一個可提供更多資訊給視障者的 Android 應用程式。成功的提供視障使用者障礙物資訊，透過本系統，可以幫助使用者在行走時，快速的提醒使用者前方的路況；若視障者不熟悉本校環境，也能透過本系統，快速的了解前方部分障礙物的路況，讓視障者能快速的認識前方環境，並針對結果做出適當的判斷與反應。

6. 銘謝

首先非常感謝我們的專題指導老師。對於專題，老師採取非常開放但卻扎實的態度。一開始老師便沒有限制我們的題目方向，因此我們可以想到並實作如此有意義的題目。從專題的架構規劃、系統規格、需求調查到最後的軟體實作、系統整合與實測，老師堅持要走過整個流程，讓我們對於整個專案的執行流程更加清楚。

再來要感謝 ESA、AIART 的學長姊及 IM 的同學。在我們開始專題規畫時，感謝哲仕學長願意花時間，講解關於專題規劃的架構流程；感謝威任學長及依蓉學姊，在遇到程式瓶頸時，能給我們即時的建議與協助；感謝嘉豪學長及彥禎學長，在遇到效能部分的困難時，能夠給予合適的建議，並提供可能的解決方式；感謝彥霖學長及君豪學長，為我們講解深度學習的原理，及在實作上的困難提供建議；感謝彥威學長及弘亞同學能針對影像處理部分給予協助及指導，並針對效能及實作細節給予建議。

最後感謝四位受測者同學及其他曾經幫助過我們的人。謝謝受測者能抽空測試我們的系統，並給予我們系統很多的建議與回饋。感謝所有人讓我們能順利的完成這項有意義的專題！

7. 參考文獻

- [1] Steve Mann, Jason Huang, Ryan Janzen, Raymond Lo, Valmiki Ramoersadm, Alexander Chen, and Taqveer Doha, "Blind Navigation with a Wearable Range Camera and Vibrotactile Helmet," in *19th ACM International Conference on Multimedia*, Scottsdale, Arizona, USA, November 28-December 1 2011, pp.1325–1328.
- [2] Vítor Filipe, Filipe Fernandes, Hugo Fernandes, António Sousa, Hugo Paredes, and João Barroso, "Blind navigation support system based on Microsoft Kinect," in *2012 International Conference on Software Development for Enhancing Accessibility and Fighting Info-exclusion (DSAI)*, Douro, Portugal, 19-22 July, 2012, pp. 94–101.
- [3] Myat K Khine, Thiri Thandar Aung, Myo Myint, and Zaw Min Naing, "RFID-based Audio Guidance Cane for Blind and Visually Impaired Person," *International Journal of Engineering Research & Technology (IJERT)*, vol. 3, no. 8, pp. 1472–1477, August 2014.
- [4] Bin Ding, Haitao Yuan, Li Jiang, and Xiaoning Zang, "The Research on Blind Navigation System Based on RFID," in *International Conference on*

- Wireless Communications, Networking and Mobile Computing*, Shanghai, China, 21-25 September 2007, pp. 2058–2061.
- [5] Anuar Mohamed Kassima, Takashi Yasunoa, Mohd Shahrieel Mohd Arasb, Ahmad Zaki Shukorb, Hazriq Izzuan Jaafarb, Mohamad Faizal Baharomb, and Fairul Azni Jafarb, “Vision Based of Tactile Paving Detection,” *Jurnal Teknologi (Sciences & Engineering)*, vol. 77, no. 20, pp. 25–32, August 2015.
- [6] P. Espinace, T. Kollar, A. Soto, and N. Roy, “Indoor Scene Recognition through Object Detection,” in *2010 IEEE International Conference on Robotics and Automation (ICRA)*, Anchorage, Alaska, USA, 3-7 May 2010, pp. 1406–1413.
- [7] Sisong Wang, and Han Wang, “2D Staircase Detection Using Real AdaBoost,” in *the International Conference on Information, Communications and Signal Processing (ICICIS)*, Macau, China, 8-10 December, 2009, pp. 1–5.
- [8] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik, “Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation,” in *Computer Vision and Pattern Recognition (CVPR)*, Columbus, Ohio, USA, 24-27 June 2014, pp. 580–587.
- [9] Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, “ImageNet classification with deep convolutional neural networks,” in *Neural Information Processing Systems (NIPS)*, California, USA, 3-8 December, 2012, pp. 4824–4833.
- [10] Koen E. A. van de Sande, Jasper R. R. Uijlings, Theo Gevers, Arnold W. M. Smeulders, “Segmentation as Selective Search for Object Recognition,” in *International Conference on Computer Vision (ICCV)*, Barcelona, Spain, 6-13 November, 2011, pp. 1879–1886.
- [11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, “Spatial Pyramid Pooling in Deep Convolutional Network for Visual Recognition,” in *European Conference on Computer Vision (ECCV)*, Zurich, Switzerland, 6-12 September, 2014, pp. 1–11.
- [12] Ross Girshick, “Fast R-CNN” in *IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, 13-16 December, 2015, pp 1440–1448.
- [13] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” in *Neural Information Processing Systems (NIPS)*, Montreal, Canada, 7-12 December, 2015, pp. 1–9.
- [14] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *Computer Vision and Pattern Recognition (CVPR)*, Nevada, USA, 27-30 June, 2016.