Towards Independent Navigation with Visual Impairment: A Prototype of a Deep Learning and Smartphone-based Assistant

Bineeth Kuriakose
Oslo Metropolitan University
Oslo, Norway
bineethk@oslomet.no

Raju Shrestha
Oslo Metropolitan University
Oslo, Norway
raju.shrestha@oslomet.no

Frode Eika Sandnes
Oslo Metropolitan University
Oslo, Norway
frodes@oslomet.no

ABSTRACT

People with visual impairments may find it hard to navigate in unfamiliar spaces. Even though several systems have been proposed to make navigation easier, such systems often ignore two essential factors for anyone using a support system: comfort and portability. This work presents a portable navigation support system for people with visual impairments. The system utilizes Smartphones as a portable platform and leverages the application of a deep learning object detection system for guiding the user during navigation. Audio output is used to communicate information about obstacles to the user. The system offers promising accuracy in object detection and distance estimation.

CCS CONCEPTS

• **Human-centered computing → Accessibility systems and tools**.

KEYWORDS

navigation, deep learning, visual impairment, portable, smartphone

ACM Reference Format:
Bineeth Kuriakose, Raju Shrestha, and Frode Eika Sandnes. 2021. Towards Independent Navigation with Visual Impairment: A Prototype of a Deep Learning and Smartphone-based Assistant. In *The 14th PErvasive Technologies Related to Assistive Environments Conference (PETRA 2021), June 29-July 2, 2021, Corfu, Greece.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3453892.3464946

1 INTRODUCTION

According to the World Health Organisation, at least 2.2 thousand million people have a vision impairment or blindness[1]. Several studies had documented that people with visual impairments often find navigation challenging [7]. Conventional navigation aids such as white canes, guide dogs, etc. are being used from a long time. But there exist several limitations with those traditional navigation aids [4]. By identifying those limitations, various navigation support systems have been proposed in the literature for users with visual impairments [1, 2, 6]. Artificial intelligence and machine learning are being explored recently by researchers to design and implement navigation solutions for people with visual impairments. When portability becomes one of the primary notions in navigation support systems, the advantages of smartphones have become a focus in research [3].

In this work, we developed a smartphone-based navigation solution prototype that uses a deep learning-based object detection model and a distance estimation method and informs the user about the detected objects and their distances through audio feedback in real-time.

2 RELATED WORKS

The deep learning-based assistive system proposed by Lin et al. [6] consists of an RGB-D camera and an earphone, a laptop for doing the deep learning processing part, and a smartphone touch-based interaction. Bhowmick et al. [2] presented an assistive navigation system using Microsoft Kinect's onboard depth sensor. The main drawbacks of [2, 6] are associated with the accuracy of object detection. Besides, these systems were not convenient for the users in terms of portability. The navigation assistance system proposed by Bai et al. [1] was based on cloud and vision computing. The system uses a smartphone to interact with the user through voice. But the system requires a data network for processing the environment data and giving instructions to the user. Besides, the system has the limitation of being less accurate in object detection. Several other similar systems and solutions had been reviewed in [4].

3 SYSTEM DESIGN

During the initial design process, informal discussions with a handful of visually impaired people were conducted. This helped with understanding the users' needs as input to the design of the prototype. The information regarding the obstacles, such as its type, distance, position, and a proper channel to give output, are some basic requirements for a navigation system for people with visual impairments [4]. Hence, the prototype consists of four main modules: object detection, distance estimation, position estimation, and text-to-speech that gives the audio output to the user.

For object detection, the You Only Look Once (YOLO) v5m version model[2] is used. YOLO is a single-stage object detector that is used for real-time object detection applications because of performance (FPS), ease of use (setting up and configurations), and versatility (models can be converted to serve different devices). The model is pre-trained with the MS COCO dataset[3]. The dataset

[1]https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment

[2]https://github.com/ultralytics/yolov5/
[3]https://cocodataset.org

PETRA 2021, June 29-July 2, 2021, Corfu, Greece

ACM ISBN 978-1-4503-8792-7/21/06…$15.00
https://doi.org/10.1145/3453892.3464946

consists of 80 objects such as person, car, chair etc. The model is converted to TensorFlow and then to a *tflite* format for deploying in a smartphone using *TensorFlow lite converter*[4].

For distance estimation, the *Rule of 57* method [5] is used. For estimating distance, two parameters, object size (height) and angle from the camera to the object, are needed in *Rule of 57* method. For demonstration, we selected some objects, fixed a value corresponding to their size (height) during the implementation, and used this for distance estimation. To estimate the angular size, two sensors-geomagnetic field and accelerometer sensors present in a smartphone were used[5]. This particular distance estimation method is selected after a detailed analysis done in another work [5].

For position estimation, the output from the object detection model is used. The model can return an array of numbers representing a bounding rectangle that surrounds its position for each detected object. The smartphone display space is divided into three regions such as *left, center*, and *right*. And the coordinates of the detected object is mapped to the display region of the smartphone. The region where the object is detected will be given as its position.

After getting the information about the detected object, its position, and the distance, the information is forwarded to the text-to-speech module to give an audio output to the user. For text-to-speech conversion, Python's TTS library *Pyttsx3* [6] is used. The prototype app can also tell whether the detected objects are 'stationary' or 'moving'. The motion sensors, particularly the accelerometer present in the smartphone, are used to incorporate this feature [3].

4 EXPERIMENT, RESULTS AND DISCUSSIONS

An android application is implemented based on the system design. The rear-end camera of the smartphone can capture real-time video while the user is in navigation. The system was tested with a Huawei P30 Pro smartphone and in an indoor environment. The app displays what kind of object is detected, how far it is located, information regarding its position and whether it is 'stationary' or 'moving' on its screen. When the user touches any part of the screen, the app provides audio feedback regarding the detected object. Two examples from the prototype application are shown in Figure 1. The first example shows the result after detecting a 'person', 'bicycle' and 'potted plant' objects.

We also experienced some false object detections and distance estimation results. The second example in Figure 1 shows a false case, where the app detects the object 'jacket' as 'human', and the distance estimation result is also not correct for the detected object 'jacket'. Some false detections can be anticipated as this is an initial prototype built using only a pre-trained object detection model, which was originally not intended for a navigation application. The issue with inaccuracies might be also due to the lightweight model used in the prototype. Lightweight neural network models generally have less accuracy than the standard models since they have fewer parameters and calculations when there is a constraint of small model size to deploy in a smartphone.

Space limitations prevents us from reporting detailed statistical analyses of the accuracy of the methods herein. These will be included in forthcoming work.

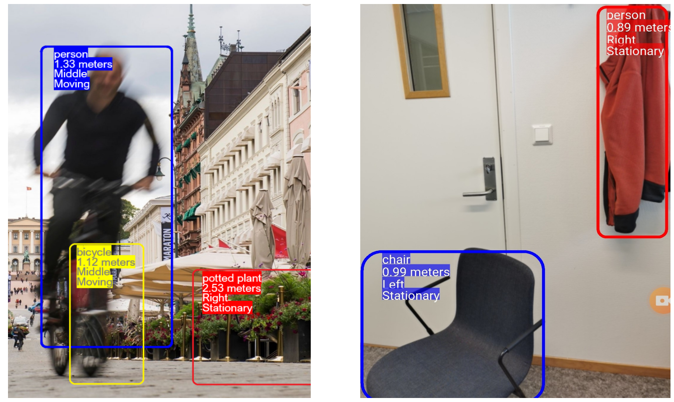

Figure 1: Two example results from the prototype system: (a) Correctly detected objects, distance, position and motion state. (b) Incorrectly detected 'jacket' as 'human' and inaccurately estimated distance.

5 CONCLUSION

The test results show the usefulness of the prototype as a potential solution for navigation support of people with visual impairments. As future work, the model can be trained to support object detection of more relevant objects in navigation scenarios, integrate with scene recognition to better understand the environment, and incorporate multi-modal output support to realize a complete navigation assistant solution. Moreover, we plan to conduct a comprehensive user evaluation and performance analysis of the system with visually impaired participants.

REFERENCES

[1] Jinqiang Bai, Dijun Liu, Guobin Su, and Zhongliang Fu. 2017. A cloud and vision-based navigation system used for blind people. In *Proceedings of the 2017 International Conference on Artificial Intelligence, Automation and Control Technologies (AIACT'17)*. 1–6.

[2] Alexy Bhowmick, Saurabh Prakash, Rukmani Bhagat, Vijay Prasad, and Shyamanta M Hazarika. 2014. IntelliNavi: Navigation for blind based on Kinect and machine learning. In *International Workshop on Multi-disciplinary Trends in Artificial Intelligence (MIWAI'14)*. Springer, 172–183.

[3] Bineeth Kuriakose, Raju Shrestha, and Frode Eika Sandnes. 2020. Smartphone Navigation Support for Blind and Visually Impaired People-A Comprehensive Analysis of Potentials and Opportunities. In *International Conference on Human-Computer Interaction (HCI'20)*. Springer, 568–583.

[4] Bineeth Kuriakose, Raju Shrestha, and Frode Eika Sandnes. 2020. Tools and Technologies for Blind and Visually Impaired Navigation Support: A Review. *IETE Technical Review* (2020), 1–16.

[5] Bineeth Kuriakose, Raju Shrestha, and Frode Eika Sandnes. 2021. Distance Estimation Methods for Smartphone-based Navigation Support Systems [Unpublished]. In *Advances in Intelligent Systems and Computing*. Springer, 000.

[6] Yimin Lin, Kai Wang, Wanxin Yi, and Shiguo Lian. 2019. Deep learning based wearable assistive system for visually impaired people. In *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCV'19)*, Vol. 1. 2549–2557.

[7] Roberto Manduchi and Sri Kurniawan. 2011. Mobility-related accidents experienced by people with visual impairment. *AER Journal: Research and Practice in Visual Impairment and Blindness* 4, 2 (2011), 44–54.

[4] https://www.tensorflow.org/lite/convert
[5] https://developer.android.com/guide/topics/sensors/sensors_position
[6] https://pypi.org/project/pyttsx3