Journal of Electrical Engineering, Electronics, Control and Computer Science – JEEECCS, Volume 6, Issue 22, pages 17-22, 2020

Smart Walking Stick for the Blind with Text Image Recognition

Mojisola Akinsiku[1], Brendan Ubochi[1] and Waliu Apena[1]
Department of Electrical and Electronic Engineering
The Federal University of Technology
Akure, Nigeria.
akinsikumojisola@gmail.com, bcubochi@futa.edu.ng, woapena@futa.edu.ng

***Abstract* – Reduced privacy, mobility and productivity are some of the major concerns for the visually impaired. Currently, most of the widely available assistive technologies for the blind have limited functionality and can be ineffective for everyday use. In this paper, the design of an electronic-based walking stick for the blind with text image recognition capability is discussed. The device uses echolocation to guide the user in order to avoid collision with obstacles. It uses a Raspberry Pi 3 model B+ which operates with Raspbian operating system and a Raspberry Pi camera module to acquire the image of the surroundings. These acquired images are digitally processed in real-time using a combination of basic morphological operations. Text image detection and recognition is performed using the EAST convolutional neural network (CNN) model and the Tesseract OCR model respectively. Finally, a text-to-speech operation is performed on the detected characters and read out to the user through an interface.**

Keywords-Walking Stick; Image processing; Optical Character Recognition; CNN

I. INTRODUCTION

Among the major forms of disability, visual impairment is one of the most severe types and affects many people around the world. The International Classification of Diseases [1] defines blindness as a form of distance vision impairment presenting visual acuity worse than 3/60. Also, according to the World Health Organization (WHO) [2] global estimates, about 2.2 billion people are visually impaired. The cost of the partial and sometimes total exclusion of this large segment of the population in the economy can be enormous [3, 4]. While, there has been huge improvements in the provision of public infrastructure with facilities to cater for the visually impaired, the challenge, however, remains to provide assistive technologies for a fully self-sufficient guidance for the blind.

The functional complexity of existing systems can range from a basic obstacle detection to systems that utilize Google map navigation. While some proposed systems attempt to solve the navigation problem, others cater for image detection and recognition alone. Advanced navigation systems are IoT (internet of things) powered multi-sensor approaches to support blind people with GPS navigation system [5]. These systems have been developed to guide blind people with voice navigated GPS but are unable to help them identify objects and signs in their immediate environment, an important feature for a self-sufficient guidance. Another assistive technology, the "Smart Stick for the Blind" is designed to include an infrared (IR) sensor, an ultrasound sensor and a water sensor [6]. The system gives an alert if any one of the sensors is triggered and it uses a buzzer to alert the blind person. It is, however, limited by the absence of a location identifier and the ability to recognize texts. In [7], an obstacle detection system is developed using a generative adversarial network (GAN) model that is trained with human eye fixations to predict saliency, and the depth information provided by an RGB-D sensor. The proposed solution in [8] developed for use in outdoor environments is a wearable smart glasses that is equipped with cameras. The system interprets video scenes and translates them into speech. As these systems provide examples of some of the currently available assistive systems, there is therefore the need for expand the scope and effectiveness ensuring that the user is guided around obstacles and is able to interact better with the environment.

In this study, we report on the development of a smart walking stick with text image recognition. Its capabilities include obstacle detection and notification; image capture and text recognition; and a text-to-speech conversion.

II. MATERIALS AND METHODS

The methodology is divided into two key parts: 1) Obstacle detection and avoidance 2) Image processing and character recognition. In Fig. 1, all the sensors and modules used are connected to the Raspberry Pi (a minicomputer) which performs operations on the sensed inputs. Also, the power supply is connected to the Raspberry Pi making it easy to power the sensors from its 5V pin.

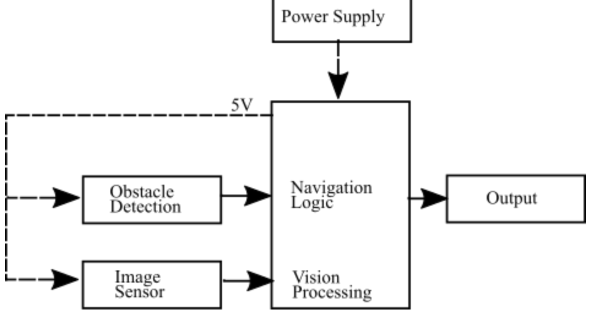


Figure 1. System architecture showing the operational blocks of the device and the direction of signal flow

A. Obstacle Detection and Avoidance System

The design of the obstacle detection system is simple and uses the difference in transmission and reception times of a wave pulse from the ultrasonic sensor and the velocity of sound in air to estimate the distance from an obstacle. The estimated distance is compared with a set threshold (100cm) to alert the user when the distance is less than this threshold. For the voice notification, an open source speech synthesizer, eSpeak software [9] is used. This software is a form of the formant based synthesis technique and has been used for speech synthesis in several languages. Basically, the ultrasonic sensor is used to detect the presence of an obstacle while the eSpeak software is configured to send a voice message to the user through a wired hands-free interface. The obstacle detection and avoidance system is the default operation mode and only over-ridden by the image processing and character recognition system whenever a text is detected.

B. Image Processing and Character Recognition

Fig. 2 shows the steps to perform text detection and recognition. In the figure, the scanned door image is preprocessed and this involves such operations as edge detection and morphological operations. Two key components of most systems are (i) text detection from images and (ii) character recognition.

The camera module utilized in this work is the Raspberry Pi camera manufactured by Raspberry Pi Foundation with board dimensions of 25mm × 20mm × 9mm. It has a resolution of 5 MP and it is a preferred choice due to its compatibility with the Raspberry Pi. The input image given in RGB format is first converted into grayscale in order to simplify the algorithm and thus reduce the computational requirements. An optimal edge detection technique, Canny Edge Detection [10], is used to detect the region of interest (ROI). The detected ROI is further processed and enhanced in order to enable accurate character recognition. Fig. 3 shows the effects of morphological operations of dilation, closing and masking on the image. Text detection is performed before the text can be recognized. In the text detection stage, the EAST (Efficient and Accuracy Scene Text) detection pipeline [11], a fully convolutional neural network model that is trained to predict the existence of text instances and their geometries from full images, is used. The algorithm is fast and is suitable for real-time text detection in natural environments. Observe, that the number 061 is not selected in the ROI. This is because the code only reads alphabets and is configured to read characters on the nameplate. For the character recognition, the Tesseract Optical Character Recognition (OCR) model is used [12]. Finally, the eSpeak software is used to read the text to the user.

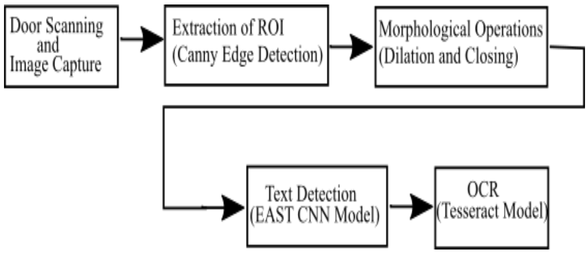


Figure 2. Image processing operations for text detection and recognition.

III SYSTEM ASSEMBLY AND RESULTS

The whole system is powered with a 5V 1800mA DC battery which can be easily charged, providing power to the various sub-components (see Fig. 4). When in continuous operation, the walking stick operates on average for a minimum of 5 hours on a single charge. The camera module is connected carefully to one of the small sockets on the Raspberry Pi that is specially dedicated to the camera. The Raspberry Pi has a Camera Serial Interface Type 2 (CSI-2) designed especially for interfacing the camera. The character recognition is limited by the camera resolution resulting in inaccuracies for distances longer than a meter away from the text. This would, therefore, make the device unsuitable for recognizing distant signs and texts especially in outdoor environments. Additionally, the device does not perform well under poor illumination conditions. The tests were carried out on an office corridor with nameplates of staff on the doors. The combined processes of text recognition and text-to-speech conversion occurs in milliseconds. In the code, however, a delay of 1 second is specified for these processes before executing subsequent operations. For distances below 1 meter, the system showed accurate text recognition. The ultrasonic sensor is connected to the General Purpose Input and Output (GPIO) signal pin of the Raspberry Pi and glued firmly to the walking stick. The system is configured to continuously alert the user when the distance from an obstacle is less than a meter. Fig. 5 shows a readout of the results obtained from the ultra-sonic sensor. The information from the sensors are converted to speech and read to the user through a wired hands-free interface, thus, suppressing significantly the ambient noise.The developed system has the ability to guide people with severe forms of visual impairment. Fig. 6 shows the system sub-components and the attached device on a typical walking stick. Note that the orientation of the camera should be directly opposite to the text for an accurate text detection and recognition. This is an important requirement for an effective operation but could be overcome if the camera is mounted on a servo-motor enabling a frontal scan, both vertically and horizontally, over 180o. Additionally, use of night vision cameras would enable operation at low illumination conditions. Further improvements to the device could be achieved by including GPS functionality either as stand-alone or connected to a GPS-enabled mobile device.

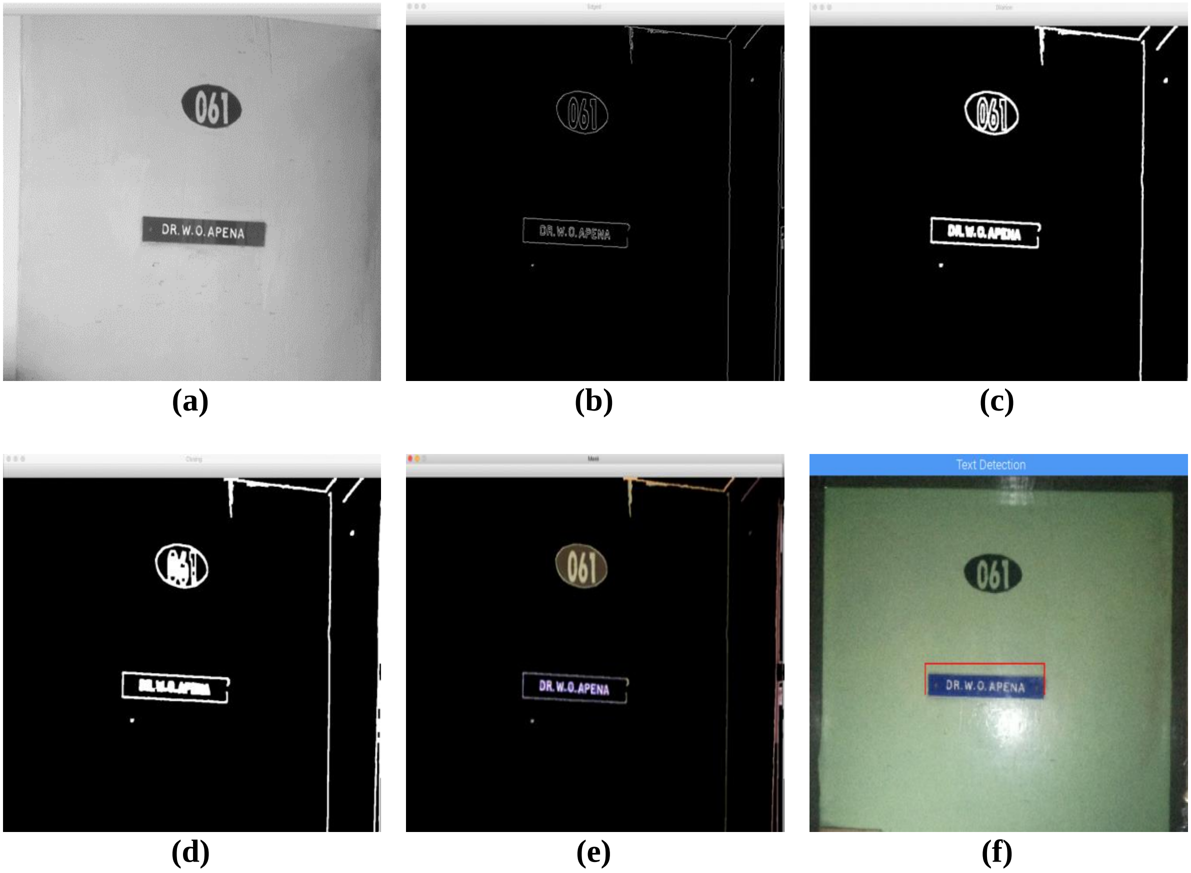


Figure 3. Image processing operations to extract the ROI showing (a) conversion to Grayscale image, (b) edge detection, (c) dilation (d) closing, (e) masking and (f) text detection.

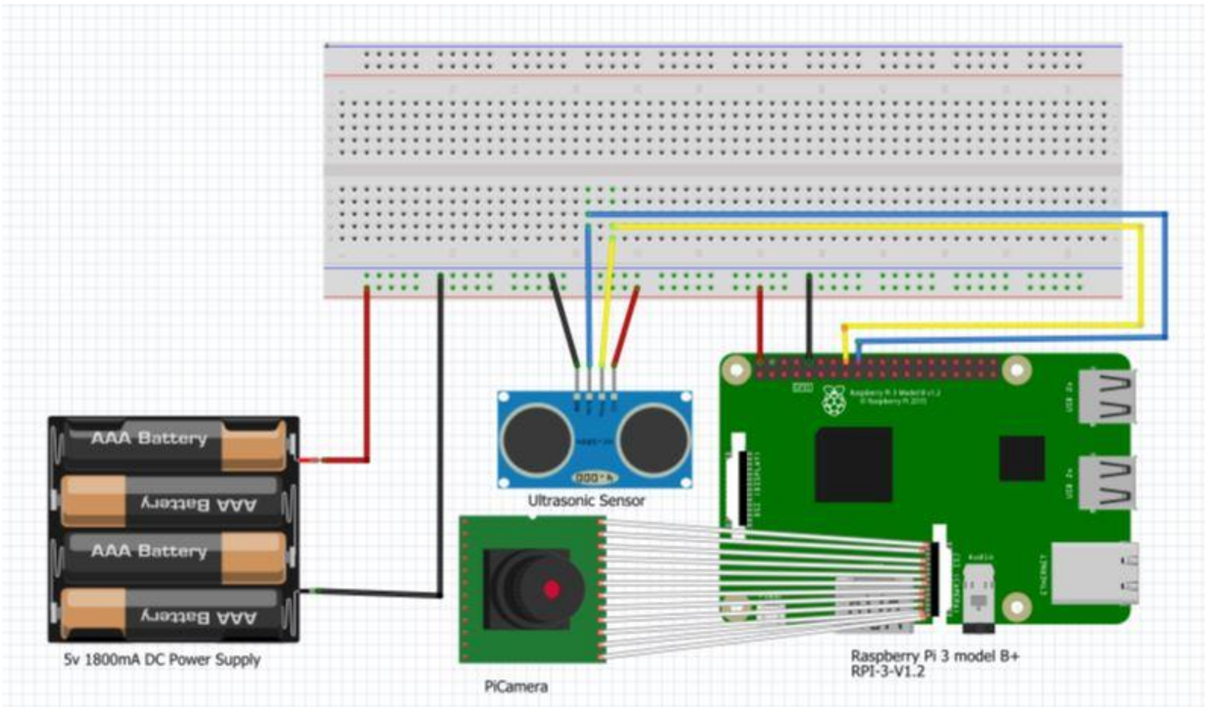


Figure 4. Layout diagram of the system showing the DC power supply and connections of the sensors to the Raspberry Pi.

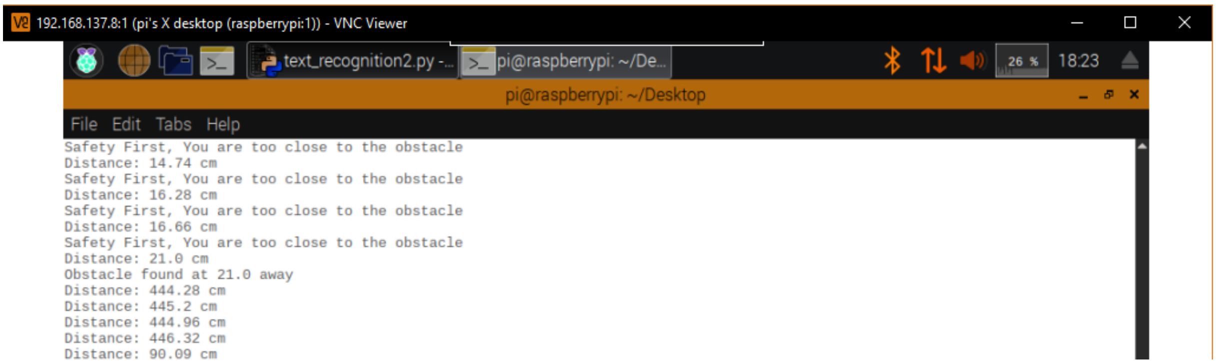
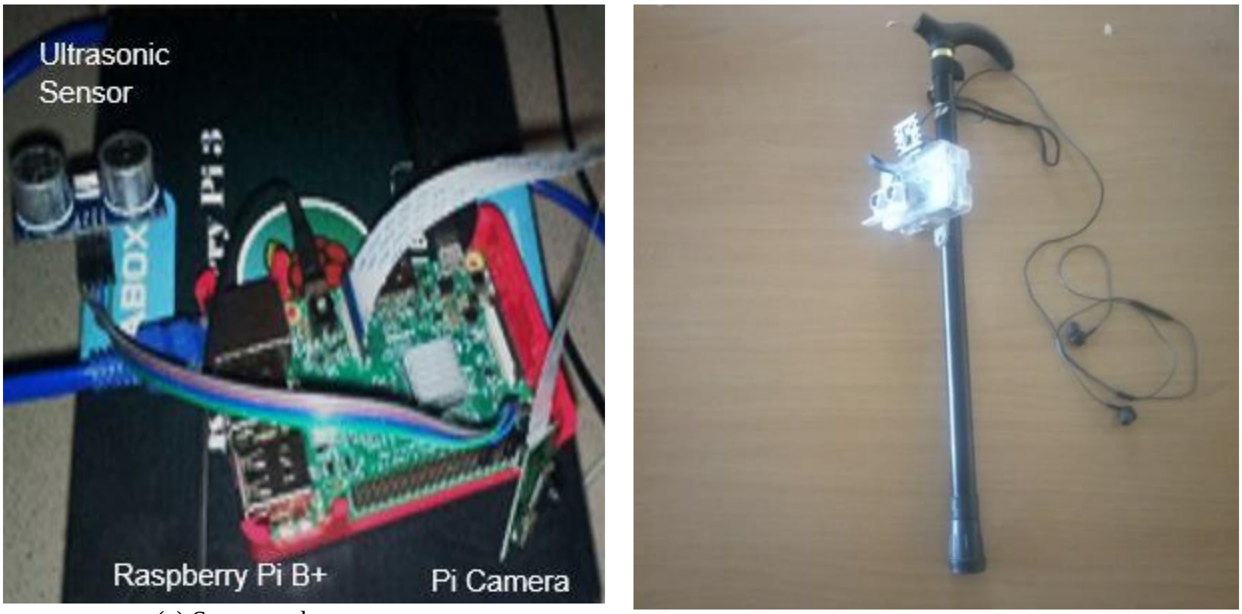


Figure 5. Readings from the ultrasonic sensor when in operation.



(a) System sub-components (b) Walking stick with device attached

Figure 6. The sub-components in the developed device (a) and a typical walking stick with developed device firmly attached (b).

CONCLUSION

In order to provide for an improved walking aid for a self-sufficient guidance, this paper has demonstrated a methodology for the development of a walking stick for the blind with text image recognition and voice guidance. The system includes obstacle detection and text image processing and is capable of reading signs and text and communicating the information to the user. The obstacle detection is accomplished using an ultrasonic sensor which uses the difference in transmission and reception times of a wave pulse from the ultrasonic sensor and the velocity of sound in air to estimate the distance from an obstacle. The text detection and recognition process includes image processing operations on a scanned image to detect a useful region of interest (ROI), extract the text image, and detect the text. The detected text is converted to speech and read out to the user through a hands-free, thus, reducing the effects of ambient noise. Due to the requirement for its use in environments with high illumination, its operation is best suited for sign and nameplate recognition in office buildings and university campuses.

REFERENCES

[1] World Health Organization: International Classification of Diseases. 11th Revision, 2018.

[2] World Health Organization. Blindness and Visual Impairment. Available at:http//who.int/en/news-room/fact-sheets/detail/blindness-and-visual-impairment

[3] J. Köberlein, K. Beifus, C. Schaffert, and R. P. Finger (2013) .The economic burden of visual impairment and blindness: a systematic review. BMJ Open, vol 3 issue 11, pp 1-14

[4] U. Chakravarthy, E. Biundo, R O Saka, C Fasser, R. Bourne, J. A. Little, The economic impact of blindness in Europe, Ophthalmic Epidemiol, vol. 24 issue 4 (2017) pp. 239-247.

[5] N.S. Mala, S.S. Thushara and S. Subbiah,"Navigation gadget for visually impaired based on IoT", 2017 2nd International Conference on Computing and Communications technologies (ICCCT), Chennai, (2017) pp. 334-338.

[6] A.A. Nada, M. A. Fakhr and A.F. Seddik,"Assistive infrared sensor based smart stick for the blind people," 2015 Science and Information Conference (SAI), London, 2015, pp. 1149-1154.

[7] G. Dimas, C. Ntakolia, and D.K. Iakovidis , "Obstacle Detection Based on Generative Adversarial Networks and Fuzzy Sets for Computer-Assisted Navigation", International Conference on Engineering Applications of Neural Networks, (2019) pp. 533-544.

[8] Iakovidis D.K., Diamantis D., Dimas G., Ntakolia C., and Spyrou E., Digital Enhancement of Cultural Experience and Accessibility for the Visually Impaired. In: Paiva S. (Ed.),

Technological Trends in Improved Mobility of the Visually Impaired. Springer International Publishing, (2020), pp. 237-271.

[9] [online] Available: http://espeak.sourceforge.net

[10] J. F. Canny, A computational approach to edge detection, IEEE Trans. On Pattern Analysis and Machine Intelligence 6 (1986) pp 679-698

[11] X. Zhou, C. Yao, H. Wen, Y. Wang, S. Zhou, W. He, and J. Liang, East: An Efficient and Accurate Scene Text Detector, IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[12] R. Smith, An Overview of the Tesseract OCR Engine, Proc. International Conference on Document Analysis and Recognition. 2007.