A Camera-Vision-based Indoor Navigation and Obstacle Avoidance Wearable Assistive Device for Visually Impaired People

Wei-Jun Lin, Mu-Chun Su, Chun-Hsiang Cheng, Cheng-Yu Tsai, Yi-Hsin Chen
Department of Computer Science & Information Engineering
National Central University, Taoyuan City, Taiwan
emails: {weijen, muchun.su}@g.ncu.edu.tw, {jerrysmoove3, roy1997860328}@gmail.com, yihsinchen@g.ncu.edu.tw

***Abstract*—In the past, if visually impaired people needed to go to an unknown/unfamiliar environment they had never been before, they had to rely on the assistance of relatives or friends and unavoidably caused trouble for both visually impaired people and their helpers. In this paper, we present a wearable assistive device to aid the visually impaired people. We use YOLOv5 to detect signages and Convolutional Recurrent Neural Network (CRNN) to recognize texts embedded in the detected signages. Via a stereo camera, our system can help visually impaired people independently move in unknown environments. In addition, MobileNet was used to detect uneven pavements in front of visually impaired people to alert them whenever dangerous conditions exist on the road.**

Keywords-visually impaired people; indoor navigation; object detection.

I. INTRODUCTION

Mobility is essential for blind and visually impaired people to move safely and efficiently as independently as possible through all environments. The white cane is the most popular navigation tool used by visually impaired people to scan their surroundings for obstacles or orientation marks. However, the detection range of white cane is restricted to about 1.5 meters from the user and the white cane must always be used to tap the surface of the road while walking, which is still limited.

Another well-known method is to adopt a guide dog. A well-trained guide dog will actively look out for hazards and obstacles that the visually impaired person cannot detect, such as a blocked path or an overhead obstruction. However, it takes a long time and cost to train a qualified guide dog. It also takes a long time to match a new handler after training. On average, a guide dog can only serve for six to seven years, which leads to another problem is their retirement.

Although visually impaired people tend to have a better sense of space than ordinary people, they still cannot reach an unknown environment without assistance from others and will resist going out for fear of disturbing others.

For the reasons outlined above, it would be helpful for visually impaired people to develop a wearable assistive device to inform the visually impaired user of any type of danger and help them navigate while moving in unfamiliar environments. In this paper, we present a wearable assistive device that can detect obstacles and different levels of heights. The device can also analyze the information of signages as well as help visually impaired people to navigate indoors.

In Section II, we introduce the related works on indoor navigation and obstacle avoidance and their shortcomings. In Section III, we introduce the hardware system for our work. In Section IV, we describe the system architecture which includes the algorithms and experiments of both obstacle detection and indoor navigation for visually impaired people. Finally, Section V is the conclusion of this paper.

II. RELATED WORK

In the last couple of years, various technologies focusing on helping visually impaired people have been introduced. They aim at increasing mobility of visually impaired users and providing additional information about nearby surroundings.

For obstacle avoidance, [1] introduced a method using infrared sensors to detect the position of obstacles. [2] uses depth cameras to build an indoor map to detect obstacles, and [3][4] build maps based on Device and Application Programming Interface (API) from Google's Tango project to help visually impaired people avoid obstacles.

For indoor navigation, the easiest way to guide a visually impaired person to walk indoors is to use a car-like navigation device. Global Positioning System (GPS) does not work well indoors and walls, ceilings, insulation etc., can absorb the signals making it harder or even impossible for a GPS device to determine its location. Many researchers have proposed various methods for indoor positioning. Yang [5] used the Round-Trip Time (RTT) of WiFi to calculate the distance by sending messages to multiple WiFi access points and using Angle of Arrival (AoA) technology to calculate the angle to the user itself in order to achieve indoor positioning. [6] and [7] presented methods which are based on the Received Signal Strength Indication (RSSI) of Bluetooth to determine the user's location. There are methods based on Radio Frequency Identification (RFID) location tracking system [8] or methods based on Ultra-Wideband (UWB) [9]. All these methods are using signal strength, angle of signal and arrival time to achieve indoor positioning. However, these methods consume a large amount of manpower and resources to pre-built maps in each public place, or the visually impaired people have to walk through these areas once to record the maps. It does not help visually impaired people to visit those places that have not been visited before. Instead of using indoor positioning, we provide a method using signage detection to parse the information on signage and help visually impaired people to navigate indoors.

 ISBN: 978-1-68558-017-9

III. THE HARDWARE SYSTEM

The wearable assistive device made in this paper includes a ZED Stereo Camera and NVIDIA AGX Jetson Xavier.

The ZED Stereo Camera is a depth camera made of two-color lenses with up to 20 meters effective distance. We use ZED for acquiring the depth information of the surroundings and the signages.

We use NVIDIA AGX Jetson Xavier as a wearable computer that computes all the deep learning networks. Although NVIDIA provides other lighter-weight devices, Xavier is one of the few embedded systems that meets the requirements of memory and computing performance.

IV. SYSTEM ARCHITECTURE

For the system architecture, our system provides two major features:

1. Obstacles Detection and Avoidance: To warn the visually impaired people whether there is uneven pavement ahead, so as to avoid kicking or missing their foot. We will discuss the detailed algorithm in Section 4. A.
2. Signage Detection and Navigation: To help visually impaired people navigate to find their destination in an unknown public place by analyzing the information of the signages and parsing the text on these signages. We will discuss the detailed algorithm in Section 4. B.

A. *Obstacles Detection and Avoidance*

We use MoblieNet to detect obstacles. However, because there are some obstacles and uneven pavements cannot be detected by specific shape or color, we use (1) Edge information (2) Depth image (3) Grayscale image as the input of MobileNet (see Figure 1).

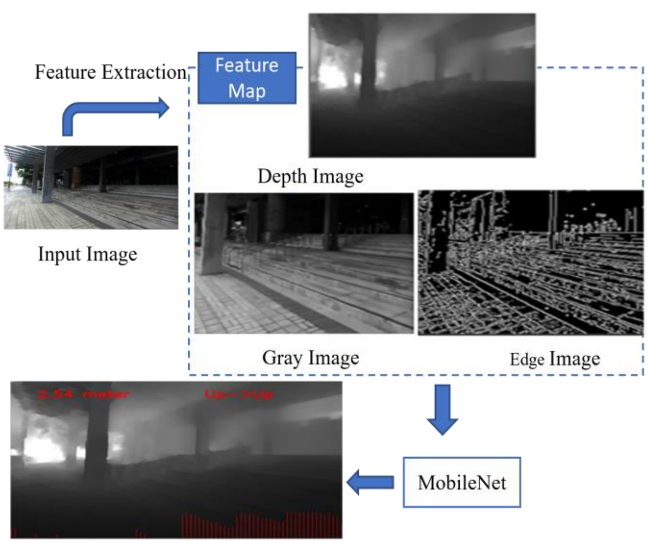

Figure 1. The process of obstacle detection

We use the Canny Edge Detection as the detector to find the edges in the images. It uses changes in color or brightness to find depth discontinuities, texture changes and differences in an image.

However, edge information is not only obtained by the different levels of height of the pavement. It can be more likely obtained by the texture of the pavement. Therefore, we use the depth image as the second input.

Finally, the color image itself is also important information. But, in order to reduce the memory load of the model, we compressed the color image into a grayscale image.

We divide our predicted output as four types: (1) Flat (2) Upstairs (3) Downstairs (4) Obstacles. We use MoblieNet and DenseNet121 to test the results, respectively. The results are shown in Table 1 and Table 2.

TABLE 1. CONFUSION MATRIX FOR OBSTACLE RECOGNITION USING MOBILENET

		Target				
Predicted		Flat	Upstairs	Downstairs	Obstacles	Precision
	Flat	1,502	0	0	0	100%
	Upstairs	0	1,016	34	1	96.67%
	Downstairs	59	11	518	0	88.01%
	Obstacles	133	88	0	1,678	88.36%
	Recall	88.67 %	91.12%	93.84%	99.94%	Accuracy =93.53%

TABLE 2. CONFUSION MATRIX FOR OBSTACLE RECOGNITION USING DENSENET121

		Target				
Predicted		Flat	Upstairs	Downstairs	Obstacles	Precision
	Flat	1,563	54	0	0	96.66%
	Upstairs	0	1,004	14	10	97.67%
	Downstairs	131	0	538	7	79.59%
	Obstacles	0	57	0	1,662	96.68%
	Recall	92.27 %	90.05 %	97.46%	98.99%	Accuracy =94.58%

B. *Signage Detection and Navigation*

We use object detection and text recognition to analyze the information on the signage. We use the distance information detected by the stereo camera to guide the visually impaired people to their destination. The whole process of signage detection is shown in Figure 2.

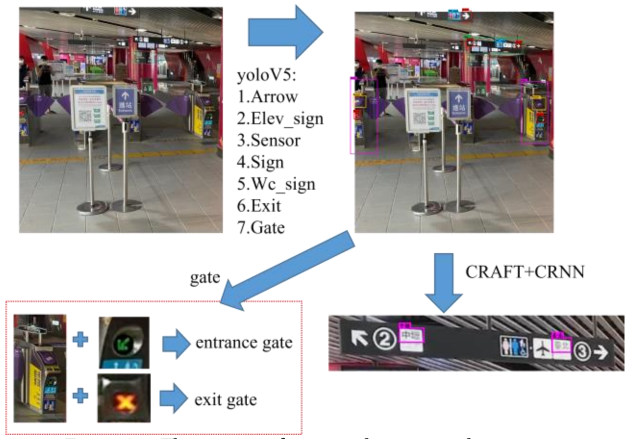

Figure 2. The process of signage detection and navigation

First, we use YOLOv5 to detect the objects (signages, gates, arrow symbols, toilet signs, elevator signs, sensors, no-passing signs etc.) in indoor public places.

Second, because there are a lot of arrow symbols on signages, entry gates and the elevator signs themselves (up and down arrows), some of them do not actually tell the users directions. We want to exclude those arrow symbols that do not follow the signage. For entry gates, both entry and exit gates are detected. We want only entry gates to be shown in our result. We use the following rules to delete the irrelevant arrow symbols and gates:

- If a "no-passing sign" is inside the detection area of an entry gate, the model would label this as a gate users cannot pass. We avoid combining International System of Units (SI) and Centimeter–Gram–Second (CGS) units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If mixed units need to be used, the units need to be clearly states for each quantity that is used in an equation.
- If an "arrow symbol" is inside the detection area of an entry gate, the model would label this as a gate users can pass.
- If an "arrow symbol" is inside the detection area of an elevator sign, it is represented as part of an elevator sign. The model would not label this arrow symbol as a valid input.
- If an "arrow symbol" is followed by a signage, the arrow indicates the orientation of the place. The model would mark its orientation.
- We would exclude all arrow symbols not listed above.

In addition to the graphic signages detected, text signages recognition would also be included in our solution. We take the image of signages detected by YOLOv5 as input. We use CRAFT [10] to detect the text position on the detected image. Then, we use CRNN [11] to extract the text (location) from the images.

Finally, we take the bounded box of the detected signage on the color image to correspond to the same position of the depth image captured by ZED as the estimation of the distance. In order to improve the accuracy of the distance, we shrunk each side of the bounded box inward by 25% to reduce the distance deviation when the detected signage is sloped.

V. CONCLUSION

In this paper, we present a wearable assistive device to aid the visually impaired people. We have asked visually impaired people to test our system to prove the feasibility of the system. The signage detection and navigation algorithm designed in this paper does not accurately locate the visually impaired people's position, but it can help the visually impaired people navigate to a specific location by analyzing the information on the signage. With our obstacles detection and avoidance, visually impaired people can be warned by our system that there may be bumps or depressions ahead. When the visually impaired people are climbing stairs, the detection system can warn that the user is approaching a flat surface or not. Our system would improve the mobility and ability for visually impaired people to walk in an unfamiliar environment and improve their safety during walking.

ACKNOWLEDGMENT

This paper was partly supported by the Ministry of Science and Technology, Taiwan, R.O.C, under 109-2221-E-008 -059 -MY3 and 110-2634-F-008 -005 -.

REFERENCES

[1] D. A. Ross and A. Lightman, "Talking braille: a wireless ubiquitous computing network for orientation and wayfinding," in Assets '05: Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility. New York, NY, USA: ACM, 2005, pp. 98–105.

[2] Y. Lee and G. Medioni, "Wearable RGBD indoor navigation system for the blind", Proc. Eur. Conf. Comput. Vis., pp. 493-508, 2014.

[3] B. Li et al., "Vision-Based Mobile Indoor Assistive Navigation Aid for Blind People," in IEEE Transactions on Mobile Computing, vol. 18, no. 3, pp. 702-714, 1 March 2019, doi: 10.1109/TMC.2018.2842751.

[4] R. Jafri and M. M. Khan, "Obstacle detection and avoidance for the visually impaired in indoors environments using googles project tango device" in Computers Helping People with Special Needs ser. Lecture Notes in Computer Science, Springer International Publishing, pp. 179-185, Jul. 2016.

[5] C. Yang and H. Shao, "WiFi-based indoor positioning," in IEEE Communications Magazine, vol. 53, no. 3, pp. 150-157, March 2015, doi: 10.1109/MCOM.2015.7060497.

[6] N. Pakanon, M. Chamchoy and P. Supanakoon, "Study on Accuracy of Trilateration Method for Indoor Positioning with BLE Beacons," 2020 6th International Conference on Engineering, Applied Sciences and Technology (ICEAST), 2020, pp. 1-4, doi: 10.1109/ICEAST50382.2020.9165464.

[7] E. Essa, B. A. Abdullah and A. Wahba, "Improve Performance of Indoor Positioning System using BLE," 2019 14th International Conference on Computer Engineering and Systems (ICCES), 2019, pp. 234-237, doi: 10.1109/ICCES48960.2019.9068142.

[8] S. Willis and S. Helal, "RFID information grid for blind navigation and wayfinding," Ninth IEEE International Symposium on Wearable Computers (ISWC'05), 2005, pp. 34-37, doi: 10.1109/ISWC.2005.46.

 ISBN: 978-1-68558-017-9

[9] J. Fu, Y. Fu and D. Xu, "Application of an Adaptive UKF in UWB Indoor Positioning," 2019 Chinese Automation Congress (CAC), 2019, pp. 544-549, doi: 10.1109/CAC48633.2019.8996692.

[10] Y. Baek, B. Lee, D. Han, S. Yun and H. Lee, "Character Region Awareness for Text Detection," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 9357-9366, doi: 10.1109/CVPR.2019.00959.

[11] B. Shi, X. Bai and C. Yao, "An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 11, pp. 2298-2304, 1 Nov. 2017, doi: 10.1109/TPAMI.2016.2646371.