Smart Glasses using Deep Learning and Stereo Camera

Jung-Hwna Kim, Sun-Kyu Kim, Tea-Min Lee, Yong-Jin Lim and Joonhong Lim
Department of Electronic Engineering
Hanyang University
Seoul, Republic of Korea
{ angelkim88, jhlim }@hanyang.ac.kr

***Abstract*—Blind people generally use cane or highly trained guide dogs for safe walking. In the case of the existing blind cane, there is a disadvantage that it cannot cope quickly with each changing situation. There is a cost for managing guide dogs. In this paper, we propose smart glasses using deep learning and stereo camera to overcome these disadvantages. A stereo camera calculates the distance between the user and the obstacle and then informs the dangerous level by the vibration motor and sound. In addition, the location of user is informed using LED to prevent accidents at night. We use YOLO v3 (You Only Look Once version 3), one of the deep learning algorithms, to inform the type of obstacle to the user. The experimental results indicate that the proposed smart glasses overcome the existing blind guidance system disadvantages and may be helpful when the user recognize and avoid obstacles.**

Keywords—YOLO, Blind, Smart Glasses, Stereo Camera, Deep Learning, Wireless Communication

I. Introduction

Recently, researches are being actively carried out to improve the life quality of disabled person by using artificial intelligence[1]. In the case of a walking assist system for the blind, there are development issues related to smart cane and glasses using sensor fusion such as camera, ultrasonic, gyro and others. In the past, the blind people used ordinary cane or trained guide dog. Conventional assistive canes have a low detection rate for obstacles and limited detection area. In the case of guide dog, there are places that can not be accessed, and there is a high cost for managing and training guide dog.

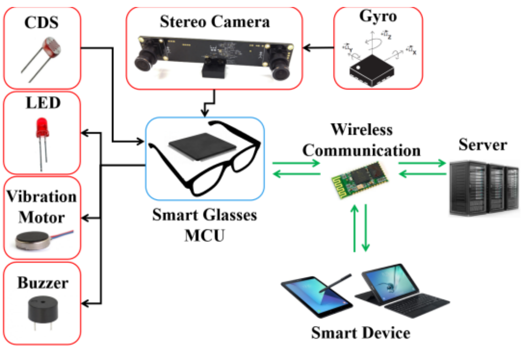

Fig. 1. Block diagram of smart glasses for blind people

To overcome these disadvantages, we propose smart glasses using deep learning and stereo camera in this paper. Fig. 1 shows the block diagram of the proposed system. The stereo cameras are used to calculate the distance between the blind and the obstacle. gyro sensor is used to distinguish between floor and obstacles. The vibration motor and the buzzer are installed in three places on the front, left and right side. The vibration motor and the buzzer operate according to the distance from the obstacle. When the obstacle approaches from the right side, the right vibration motor and buzzer operate. On the other hand, the left vibration motor and the buzzer operate when approaching from the left side. If the obstacle approaches the front, the front vibration motor and the buzzer operate. Day and night are identified using Cds (Cadmium sulfide) sensors. At night, the visibility of the other person may be limited. In order to prevent accidents, the LED operates and informs the other person about the location of the blind people. After the detection of the obstacle, a deep learning algorithm is used to recognize the type of obstacle. In this paper, we use YOLO v3 (You Only Look Once version 3) as a deep learning algorithm. Deep learning algorithm can not be performed on low level MCU of smart glasses since this algorithm compute with a lot of data. To perform fast computation, image data is transmitted to the server using wireless communication. If the server recognizes an obstacle as a vehicle, it calculates the distance to the blind user. If the calculated distance is dangerous, smart devices and buzzer alert blind user to dangerous situations. The driver is informed that the location of the blind user through the buzzer and LED.

II. Distance Detection Algorithm

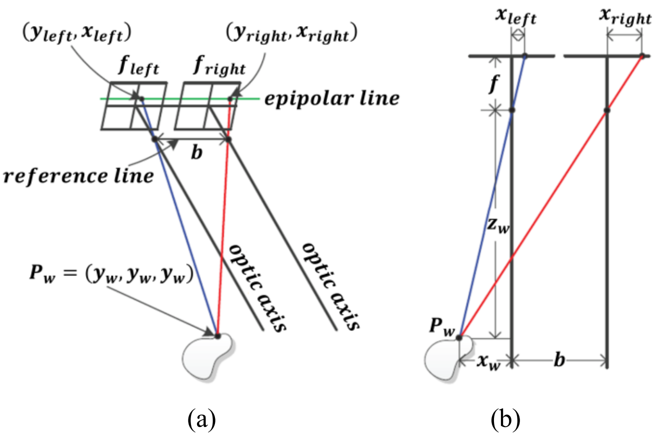

Fig. 2. Geometry of stereo vision (a) Projection processing, (b) Dimension simplification

The depth of the image can be measured using stereo vision as shown in Fig. 2(b). Each image coming into the left and right camera can be expressed as follows[2]:

$$\text{Left image} : \frac{x_{left}}{f} = \frac{x_w}{z_w}, \text{Right image} : \frac{x_{right}}{f} = \frac{b+x_w}{z_w} \quad (1)$$

which is a formula based on the triangle similarity ratio. b is the distance between the two cameras. f is the focal length. x_{right} and x_{left} are the coordinates of the points corresponding to each other in the two images. The obstacle distance can be obtained by finding the corresponding point. The depth information z_w can be obtained as

$$z_w = \frac{bf}{x_{right} - x_{left}} = \frac{bf}{d} \quad (2)$$

where d is disparity. In order to reduce the matching time, we use the rectified epipolar constraint using the calibrated image as shown in Fig. 2(a). The rectification means a process of converting an image so that the epipolar line is parallel.

III. Obstacle Recognition Algorithm

In this paper, we use the YOLO v3 algorithm to recognize obstacles. Conventional deep learning algorithms such as CNN and RNN are slow to recognize since these methods find the object many times using the size of each bounding box. On the other hand, YOLO v3 makes detections at three different scales. In YOLO v3, the detection is done by applying 1 x 1 detection kernels on feature maps of three different sizes at three different places in the network as shown in Fig. 3. This network is the base feature extractor of Darknet-53. It has 53 convolutional layers as shown in Fig. 4. The up-sampled layers concatenated with the previous layers help preserve the fine grained features which help in detecting small objects. In addition, YOLO v3 performs multilabel classification for objects detected in images[3].

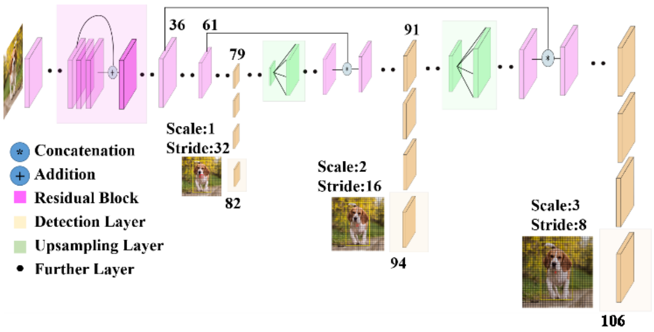

Fig. 3. YOLO v3 algorithm architecture

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Fig. 4. Darknet-53 layer architecture

IV. Experimental Results

We made the frame of smart glasses using a 3D printer as shown in Fig. 4. The experiment has been performed on ROS

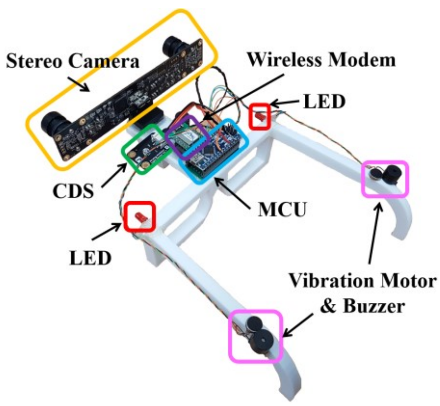

Fig. 5. The develpoed smart glasses using stereo camera and deep learning

Kinetic Kame. The processor specifications of server are Intel(R) Core(TM) i7-8700K and 32GB RAM. CPU clock is 3.7 GHz. Operating systems are Ubuntu 16.04 LTS and Windows 10 64Bit. The dataset used to detect and recognize obstacles is Microsoft COCO. The dataset contains photos of 91 objects types with a total of 2.5 million labeled instances. Fig. 6(a) shows that the front obstacle is close to the user so that the buzzer and the vibration motor operate. Fig. 6(b) is a depth measurement image using a stereo camera. The red tone color means the obstacle is close. Conversely, blue tone color means far away. Fig. 6(c) is an obstacle detection of cloud point type using a gyro sensor. Since the gyro sensor is used, it is possible to distinguish between floor and obstacle. Fig. 6(d) is the obstacle recognition result using YOLO v3.

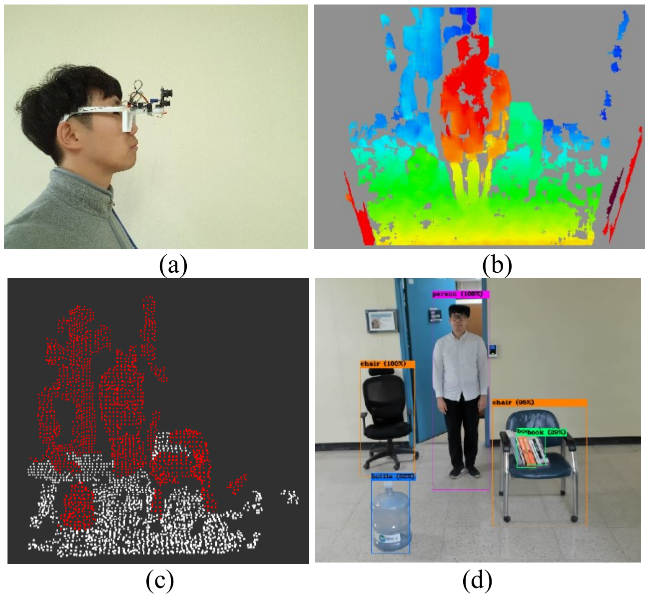

Fig. 6. The develpoed smart glasses experimental result

V. Conclusion

When avoiding obstacle, it is important to know the location as well as the type of obstacle. The existing blind device for the blind people informed only the location. The proposed device informs the blind people by location and type of obstacle using deep learning and stereo camera. The experimental results indicate that the proposed device has detected an obstacle well. It could recognize the kind of overlapping obstacle. The proposed device may be helpful for safe walking of the blind user.

Acknowledgment

This research was supported in part by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (GrantNo.2017R1D1A1B03033806).

References

[1] Josh Jia-Ching Ying, Chen-Yu Li, Guan-Wei Wu, Jian-Xing Li, Wei-Jheng Chen, and Don-Lin Yang, "A Deep Learning Approach to Sensory Navigation Device for Blind Guidance," HPCC / SmartCity / DSS, UK, pp. 1195–1200, June 2018.

[2] Vaddi Chandra Sekhar, Satyajit Bora, Monalisa Das,Pavan Kumar Manchi,Josephine S and Roy Paily, "Design and Implementation of Blind Assistance System Using Real Time Stereo Vision Algorithms," 2016 29th VLSID, India, pp. 421–426, January 2016.

[3] Joseph Redmon, Ali Farhadi, "Yolov3: An incremental improvement," Computer Vision and Pattern Recognition, pp. 1–6, April, 2018.