AR Marker Aided Obstacle Localization System for Assisting Visually Impaired

Xinrui Yu, Guojun Yang, Scott Jones and Jafar Saniie
Department of Electrical and Computer Engineering
Illinois Institute of Technology, Chicago, Illinois, USA

***Abstract*— As of in 2017, approximately 3.3% of the people in the world (253 million) are visually impaired. These people face mobility difficulties which impact their quality of life. We propose a system that will assist the visually impaired with their indoor environment mobility. It uses the information from pre-registered AR (Augmented Reality) markers to identify specific accessible facilities, such as hallways, restrooms, staircases, and offices. An RGB-D sensor (a sensor that provides color and depth information of every pixel) captures the scene which includes the AR markers. This scene-based information is processed by a neural network to recognize and localize obstacles and accessible facilities. We present the processing algorithm for image and depth profiling. System performance in terms of obstacle localization and recognition was evaluated inside building.**

Keywords—Visually impaired, obstacle localization, depth profiling

I. Introduction

Difficulties are often encountered by the visually impaired, especially inside unfamiliar buildings/facilities [1]. These difficulties generate the need for an obstacle localization system for visually impaired people. A vast range of sensors, combined with different algorithms, have been examined to come up with such a system, with different levels of successes [2,3,7-9].

Numerous obstacles of different categories will be encountered in daily lives of the visually impaired. It should be taken into notice that obstacles which could be avoided by sighted people subconsciously would hamper the movements of the visually impaired seriously. In many cases, the difficulties for visually impaired navigating lie on the obstacle localization, especially on categorizing the obstacles [3]. Without categorizing an obstacle, it is not hugely beneficial for the visually impaired to know the position of the obstacles. Not knowing what obstacle is being encountered, is the concern of the visually impaired as they move inside an unfamiliar building. Our paper provides a powerful and practical solution.

Using a RGB-D sensor, synchronous visual and depth information can be provided to the program [4,5]. Latest developments of convolutional neural network boost the performance of image classification. With the help from MATLAB [6], obstacle recognition in our algorithm can operate in real-time. Moreover, with the help of AR markers, the identification of different facilities becomes much easier. These conditions made our work possible.

II. System Description

A. System Components and Principles of Operation

The proposed system consists of three pieces of hardware: The RGB-D Sensor (Kinect v2), a PC as a processing unit, and I/O devices for the visually impaired. As to software, the program consists of three modules: AR marker recognition module, obstacle recognition module, and obstacle localization module.

The system operates as shown in Fig. 1. The RGB and depth images generated by the RGB-D sensor will be sent to PC. The AR marker recognition module analyses the RGB image only, while the obstacle recognition module processes both RGB and depth images using a convolutional neural network (CNN). Together, they generate positional information of different entities relative to the point of observing, and information about the property (e.g., category) of the entities. The information is used by the localization module to generate verbal information for the visually impaired, providing them with the position and category information of different obstacles and AR markers.

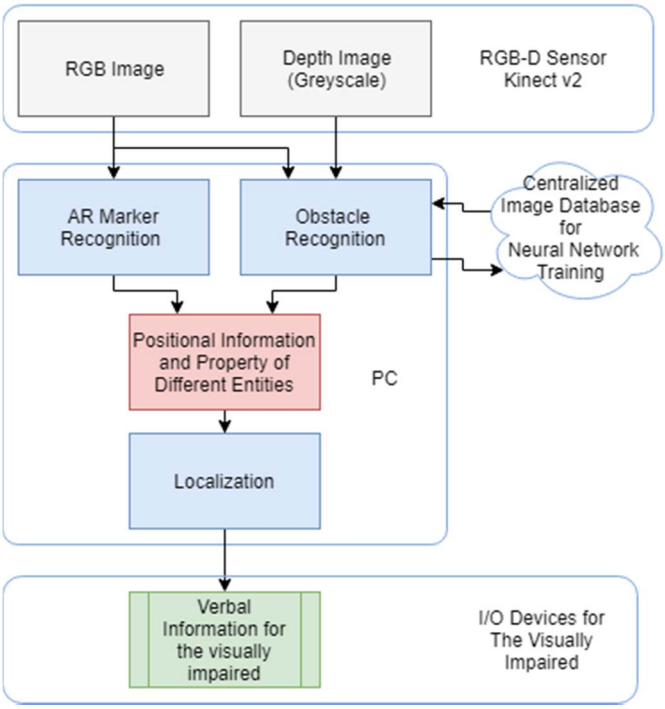

Fig. 1. System Flowchart of Object Recongnition and Localization

978-1-5386-5398-2/18/$31.00 ©2018 IEEE

A centralized image database will be created and used for training the CNN. The database is designed to be able to update itself based on the new images provided by the users. Preferably, the database will be monitored and maintained to further increase the accuracy and robustness of the system.

In cases of emergency like system malfunction, the visually impaired will be noticed via sound alarm. Also, if one of the modules is determined to be faulty, the other module will be able to provide notice, then operate the system in safe mode with sacrifice on performance instead of completely shutting down.

B. Innovative Features

The idea of using an RGB-D sensor to aid the visually impaired has already been studied [2,7]. There are also many projects that use depth information to help the visually impaired [8]. However, restrained by technical and methodological issues, some means of utilizing the sensor have not been thoroughly explored. The system presented in this paper shows following innovative features, and they are very powerful in testing:

1) Image Segmentation using Depth Image

While most image segmentation techniques and algorithms are applied on RGB or grayscale image, or dedicated to images captured by special instruments (X-ray, MRI, etc.), we performed image segmentation on depth images.
Using segmented depth images as masks, the difficulties in pattern filtering are solved, since the surface pattern does not show up on depth images. Therefore, it is possible to separate a certain object from the background easily with good accuracy, even if the object is "camouflaged" at the visible light spectrum. This is very helpful in obstacle detection.

2) Region Extraction using Segment-Overlaid RGB Image

While depth image is very powerful in object detection, the depth image of a certain object, without any pattern or detail, is not helpful in determining the category of the object. Many existing obstacle localization systems for the visually impaired have no ability to determine the category of the obstacle [2,7,9]. This is when the synchronous RGB and depth image frames of a RGB-D sensor came in handy. By overlaying the segmented region of the depth image onto the RGB image, the surface detail of the object is obtained.

3) Image Recongition Using Extracted Image Aided by Convolutional Neural Network (CNN)

Using the extraction technique, the shape of the object can be separated from the background, greatly reducing the unrelated information. Furthermore, with the help of a pre-trained CNN, the accuracy of the image classification can be vastly improved compared with conventional classification techniques, which provides ideal results in terms of speed and accuracy.

4) Obstacle Localization Using Scaled Depth Image

In typical obstacle localization systems, the location of the obstacle is determined by analyzing reflexed waveform [8]. This is a complex process, and susceptible to noise. By reading the pixel intensity of a depth image, the range of an object in this image can be immediately obtained, thus avoiding aforementioned disadvantages. Also, it is no longer required to determine the location of a facility marked by AR marker with only RGB image; the range of the marker can be read directly from corresponding area on the depth image.

III. Object Recongnition and Localization

A. Recognizing AR Markers

AR markers are widely used in mapping and robot applications [10]. Each AR marker carries an ID number associated with certain locations (e.g., offices, elevators, restrooms, etc.) AR markers can be used as labels for vision-based navigation system.

Moreover, AR markers can be used to assist the recognition and localization of different entities. By recognizing AR markers in the field of view, corresponding facilities can be located, and later localized with the localization module.

Floor, staircases, and hallways can be recognized with the help of AR markers. Their positional information helps localization module to generate the information of their whereabouts for the visually impaired.

A MATLAB program based on checkboard detection is used to locate and recognize the AR markers. Its procedures are summarized as follows:

Step 1. The original image goes through edge detection, revealing the perpendicular edges inside the marker;

Step 2. The perpendicular edges of the image are examined to locate a region with a standard pattern of alternating white and black color;

Step 3. The positions of the intersections of the edges in the region are stored.

Step 4. The number and location of the intersections are used to read the information from the AR marker.

The steps of edge detection and AR marker region location (shown in cyan color) is shown in Fig. 2.

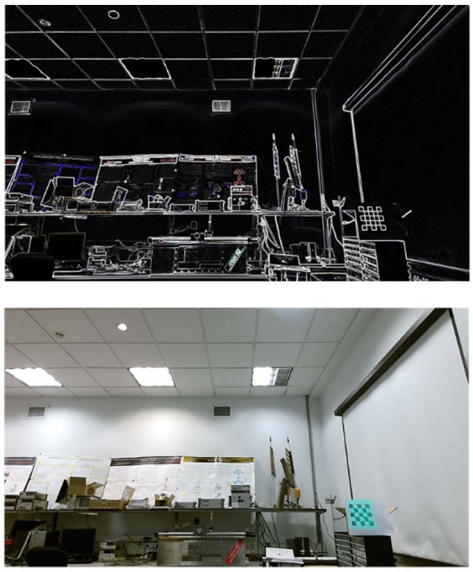

Fig. 2. Edge Detection and AR Marker Region Locating

The region of the AR marker is also marked on the depth using image alignment described in the next section.

B. Obstacle Profiling

In this paper, we focus on obstacles in unfamiliar indoor environments, such as an office building or a shopping mall. The categories of the obstacles may include people, furniture, walls, etc.

The raw input information of our obstacle recognition program is RGB image captured by the colored camera and the depth image frame (in form of a grayscale image) from the ToF (Time of Flight) camera. Using an RGB-D Sensor provided us with a distinct advantage in this field. While most obstacle detection methods are built solely with depth sensors, we can combine the information from the RGB and the depth image, providing a better comprehension of the category of the obstacle.

The procedures of the obstacle profiling are discussed below.

1) Image Acquisition and Alignment

The acquisition of the RGB and depth image is done by the basic color and depth tools provided by the Microsoft Kinect v2 for Windows SDK. The depth tool is modified so it returns a grayscale image with pixel intensity 0~255 mapping to range 0.5~8m. Such clipping reduces the resolution of range on the depth image but allows the pixel intensity corresponding to a certain range continuous along the depth axis, thus making the obstacle profiling program to be implemented easier.

The RGB image frame has a resolution of 1920*1080, while the depth image frame has a resolution of 512*424. This means that the two images must be aligned before being analyzed as a combined information source. The alignment of the images is done by using Control Point Selection Tool in MATLAB as shown in Fig. 3. The points are marked with cyan dots.

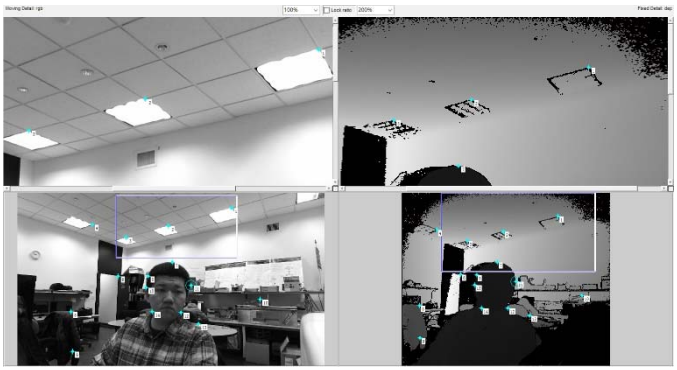

Fig. 3. Control Point Selection in MATLAB

After control point selection, two alignment matrices are generated. The *imwarp* function of MATLAB aligns the two images according to these generated matrices.

Fig. 4. Depth Image Aligned with RGB Image

2) Image Segmentation

To isolate different obstacles from the background, image segmentation is needed. Segmentation is done on the depth image, and then combined with the RGB image, to provide an all-around result.

The depth image is chosen as the basis of the segmentation due to its character. To be more specific, a depth image reflects depth information only, textures on the same object (stripes on clothing, wallpaper decoration, etc.) will be ignored. Such character of depth image greatly reduces the probable factors of false segmentation. The segmentation is done by a watershed transform algorithm on MATLAB, optimized for depth image segmentation. The existing image processing toolbox of MATLAB is powerful enough to perform real-time analysis of the image [6]. Compared with other image segmentation methods, watershed transform is better in terms of its ability of recognizing and generating non-linear and precise boundaries between two regions. As shown in Fig 5, the algorithm of watershed transform is summarized in the following steps [11]:

Step 1. The original noisy depth image is filtered with a 3*3 median 2-D filter, removing abundant noise pixels.

Step 2. The filtered image is calibrated in terms of contrast, making the different regions of the image more recognizable for the algorithm.

Step 3. The image is chopped by a number of straight lines, removing the blind spots from the image.

Step 4. The image further goes through opening, closing, and reconstruction, removing the rest noisy chunks and smooths edges of each region. The result is shown in Fig. 5 (b).

Step 5. Regional maxima areas are obtained with inbuilt MATLAB functions, determining the number of regions and their corresponding location in the image.

Step 6. Regions are labeled according to the calculation of the watershed algorithm using regional maxima area information. The regions are shown in Fig. 5 (c).

Step 7. The labeled regions are mapped to the RGB image. Result shown in Fig. 5 (e).

Step 8. The labeled regions are examined before extraction. The background regions are removed because they are not of immediate concern to the visually impaired. The blind spots of the depth camera can be seen on Fig. 5 (a), on four corners of the original image. Their pixel intensities do not reflect real depth information; thus, they are also removed. A region is determined to be a background region or a blind spot region if it contains one of the four corners of the image.

Step 9. The labeled regions of the RGB image in Fig. 5 (f), now examined, are extracted and await recognition.

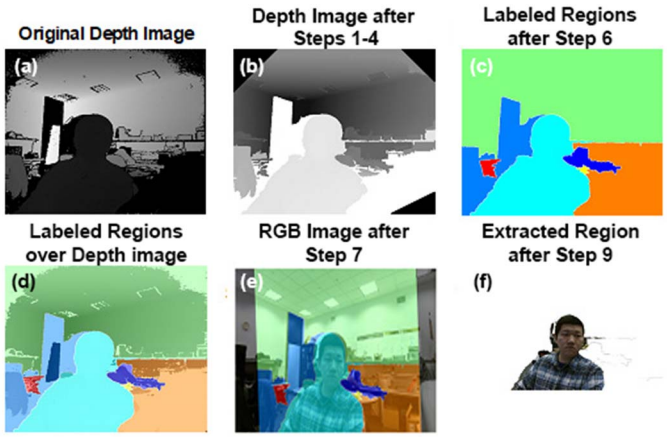

Fig. 5. Steps of Obstacle Profiling

C. *Obstacle Recongition*

After the obstacles are profiled, they will be recognized with image processing library/toolbox. The depth information of the obstacles will be generated and provided to the localization module.

The core of this part is to determine the category of the obstacle. Obviously, the visually impaired can be better informed and assisted if the category of the obstacle can be determined. The classification of the obstacles is performed by a deep learning program on MATLAB, using a pre-trained Convolutional Neural Network (CNN). A Convolutional Neural Network (CNN) is a powerful machine learning technique from the field of deep learning. CNNs are trained using large collections of images. From these large collections, CNNs can learn rich feature representations for a wide range of images [12].

The requirements of an applicable image classification program for the visually impaired can be described as follows:

- Wide range. The classification program should be able to classify most common indoor objects.
- Real-time. The classification should be done within frame separation.
- Reliable. The accuracy of classification should be ≥ 95% to be reliable enough to the visually impaired.
- Sensitivity. The program should be able to distinguish a certain object from all directions.

An image category classifier using CNN can extract features with respect to all these requirements, and with margins to spare depending on the property of the CNN. Conversely a conventional classifier using hand-crafted features falls behind in these terms.

For this paper, a pre-trained CNN provided by MATLAB is used together with selected categories of images from Caltech 101 [12]. The categories with little probability to appear in an indoor environment (e.g., aircraft) are removed for better memory usage and computing speed. More categories can also be added to the dataset for a larger number of recognizable obstacles.

The initialization of the classifier follows the following steps [12]:

Step 1. The images from Caltech 101 are loaded. Only selected categories are loaded; unnecessary categories, like vehicles and animals are dropped to save memory space and reduce execution time.

Step 2. The pre-trained AlexNet network is loaded.

Step 3. The images are normalized for CNN, since the AlexNet network can only process 227 by 227 RGB images.

Step 4. The training and testing on image sets are prepared. 30% of the images from the loaded images will be used for training, and the remainder will be used for validation.

Step 5. Training features are extracted using CNN. The layer right before classification is chosen.

Step 6. A multiclass SVM classifier is trained using CNN.

Step 7. The classifier is evaluated using validation sets. The mean accuracy of the group is close to 1.

After the initialization of the classifier, images can be input to the classifier directly; the classifier will give prediction for the image. The obstacle recognition program will then mark the label according to the prediction for the corresponding

region. After every undetermined region is given a category, the all-labeled picture is given to the localization part.

D. Localization

The outputs from AR marker and obstacle algorithms are used to generate localization information to aid the visually impaired. This part of the program uses the original depth image combined with the regional category information to generate the azimuth and range of different AR markers and obstacles.

To provide accurate positional information to the visually impaired based on the original depth image, the characteristics of the cameras must be known. The following table provided the information we need [13]:

TABLE I. CHARACTERISTICS OF KINECT V2 CAMERAS

Type	Characteristics		
	Resolution in pixels N by M	*Horizontal field of view in degrees* α	*Vertical field of view in degrees* β
Color	1920 x 1080	84.1	53.8
Depth	512 x 424	70.6	60

We are considering 2-D movements only, so only horizontal field of view is taken into consideration. We have the following equation to determine the relative position of an obstacle:

$$\theta = \frac{P\alpha}{N} - \frac{\alpha}{2} \quad (1)$$

where P is the horizontal position of the geometric center of the region, in terms of the number of pixels from the left edge of the image; α is the horizontal field of view of the depth camera in degrees, α = 70.6; N is the resolution of the depth image along the horizontal axis, N = 512; θ is the azimuth of the obstacle in degrees, 0 indicates dead ahead, negative value indicates on the left, and positive value indicates on the right.

The second parameter to be determined is the distance of the obstacle. In the original depth image, the intensity of a pixel indicates the distance of the object, with the following relationship:

$$d = 0.5 + \frac{I}{255} \times 7.5 \quad (2)$$

Where d is the distance of the pixel in meters, I denotes the intensity of the pixel (0 to 255). The schematic is shown in Fig. 6.

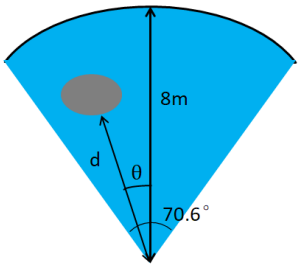

Fig. 6. Schematic of Localization

The steps of the localization program are as follows:

Step 1. The average pixel intensity and the position of the geometrical center of every essential region are calculated with inbuilt MATLAB function.

Step 2. The relative positions of the corresponding entities are calculated with equation (1) and (2). They are arranged in a sequence with the nearest obstacles at the start. The positions of the AR markers are arranged after the obstacles.

Step 3. The relative position, combined with the category of the entity, is translated to verbal information and played to the visually impaired according to the sequence, e.g., "Person three point five meters away 15 degrees to the left, restroom four meters away 23 degrees to the right." This gives the information of the obstacles to the user.

IV. RESULTS AND DISCUSSION

The whole system is tested indoor, to be more specific in a teaching building. Different people and objects are present in the view range of the RGB-D sensor, and the tasks are executed. The results are shown below.

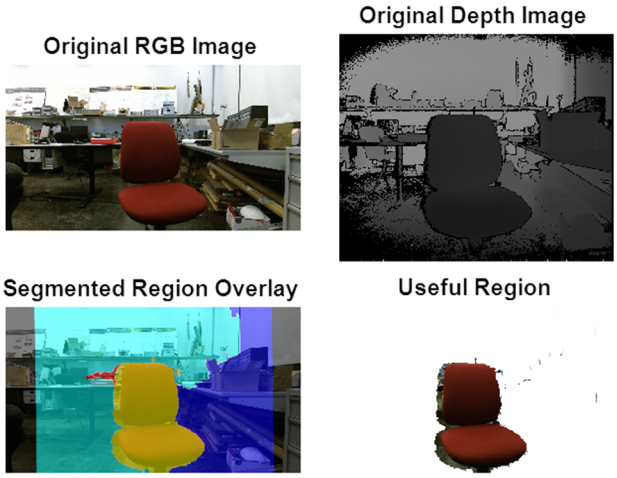

Fig. 7. Segmentation Result Example No. 1

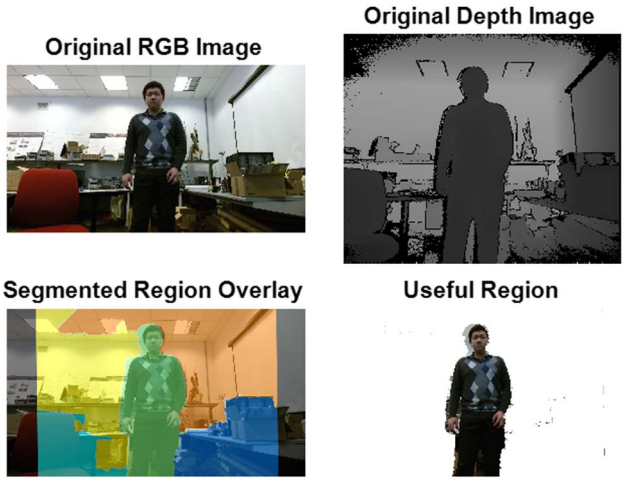

Fig. 8. Segmentation Result Example No. 2

From the above examples, we can see that the image segmentation program can segment regions of interest successfully, with enough accuracy of the image classifier to determine the category of each object in image.

The region of interest is then extracted and input to the image classifier. The classifier is pre-trained, and ready to sort the input images. The classifier distinguished the two images successfully. The chair in the first example is recognized as category "chair," and the person in the second example is recognized as category "faces".

The category information is passed onto the localization program. The relative position of the regional obstacle is calculated and generates verbal information.

For the first example, the average pixel intensity in the region is 32.5018; thus d = 1.4559m. The weighted center of the region is [246.9683, 268.2252], thus θ= -1.2454 degrees. Therefore, the verbal information will be:

"Chair one point five meter away one degree to the left."

For the second example, the average pixel intensity in the region is 46.4639; thus d = 1.8666m. The weighted center of the region is [244.6818, 255.4554], thus θ= -1.5607 degrees. Therefore, the verbal information will be:

"Person one point nine meter away two degrees to the left."

The location of AR marker marked facilities can also be provided with the same manner. The verbal information can be replaced by other preferable means of notification.

V. Conclusion

According to the test results mentioned above, the system performance described in the introduction is achieved. In a word, the information from the RGB-D sensor can be analyzed to generate adequate situational awareness information to help the visually impaired navigate in an indoor environment.

However, some aspects of the system are yet to be perfected. Some of these are due to hardware limitations. As a 5-year-old hardware which is no longer supported, Kinect v2 is somewhat insufficient to achieve perfect performance. For example, blind spots exist in the field of view of the depth camera, resulting in more complicated operation of the image segmentation algorithm. Finally, GPS based guidance system can be provided for temporary outdoor travel.

For future designs, enhanced off-the-shelf hardware can be used, such as the D-series Intel® RealSense™ Depth Camera. As the next generation RGB-D sensors, their specifications are more advanced than Kinect v2 in every aspect, especially in the maximum range, depth stream output resolutions, and the depth field of view. By utilizing this system would probably provide improved results of obstacle localization and recognition. Also, smaller and lighter laptops/tablets can be used, enabling the system better suited for everyday use.

The conclusion obtained from our test results is very promising. Using a common laptop, it is possible to provide information for obstacle recognition and localization to the visually impaired in real-time. Combined with AR markers on different surfaces of an indoor environment, it is possible to create an "Visually Impaired Friendly" building. In such building, it is possible for the visually impaired to navigate around, use various facilities/equipment and avoid obstacles much like normal people, without physical contact with the obstacle. This would be hugely beneficial for the visually impaired.

References

[1] World Health Organization. (2018). *Vision impairment and blindness*. [online] Available at: http://www.who.int/mediacentre/factsheets/fs282/en/ [Accessed 21 Feb. 2018].

[2] S. Wang and Y. Tian, "Detecting stairs and pedestrian crosswalks for the blind by RGBD camera," *2012 IEEE International Conference on Bioinformatics and Biomedicine Workshops*, Philadelphia, PA, pp. 732-739, 2012.

[3] C. R. Prashanth, T. Sagar, N. Bhat, D. Naveen, S. R. Rupanagudi and R. A. Kumar, "Obstacle detection & elimination of shadows for an image processing based automated vehicle," *2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Mysore, pp. 367-372, 2013.

[4] O. Wasenmüller, M. Meyer and D. Stricker, "CoRBS: Comprehensive RGB-D benchmark for SLAM using Kinect v2," *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Lake Placid, NY, pp. 1-7, 2016.

[5] M. Carfagni, R. Furferi, L. Governi, M. Servi, F. Uccheddu and Y. Volpe, "On the Performance of the Intel SR300 Depth Camera: Metrological and Critical Characterization," in *IEEE Sensors Journal*, vol. 17, no. 14, pp. 4508-4519, July15, 15 2017.

[6] C. Lu, J. Shi and J. Jia, "Abnormal Event Detection at 150 FPS in MATLAB," *2013 IEEE International Conference on Computer Vision*, Sydney, VIC, pp. 2720-2727, 2013.

[7] A. Aladrén, G. López-Nicolás, L. Puig and J. J. Guerrero, "Navigation Assistance for the Visually Impaired Using RGB-D Sensor With Range Expansion," in *IEEE Systems Journal*, vol. 10, no. 3, pp. 922-932, Sept. 2016.

[8] S. S. Bhatlawande, J. Mukhopadhyay and M. Mahadevappa, "Ultrasonic spectacles and waist-belt for visually impaired and blind person," *2012 National Conference on Communications (NCC)*, Kharagpur, pp. 1-4, 2012.

[9] F. Ribeiro, D. Florêncio, P. A. Chou and Z. Zhang, "Auditory augmented reality: Object sonification for the visually impaired," *2012 IEEE 14th International Workshop on Multimedia Signal Processing (MMSP)*, Banff, AB, pp. 319-324, 2012.

[10] G. Yang and J. Saniie, "Indoor navigation for visually impaired using AR markers," *2017 IEEE International Conference on Electro Information Technology (EIT)*, Lincoln, NE, pp. 1-5, 2017.

[11] Mathworks. (2018). Marker-Controlled Watershed Segmentation - MATLAB & Simulink Example - MathWorks United Kingdom. [online] Available at: https://www.mathworks.com/help/images/examples/marker-controlled-watershed-segmentation.html [Accessed 21 Feb. 2018].

[12] Mathworks. (2018). Image Category Classification Using Deep Learning - MATLAB & Simulink - MathWorks United Kingdom. [online] Available at: https://www.mathworks.com/help/vision/examples/image-category-classification-using-deep-learning.html [Accessed 19 Feb. 2018].

[13] Intel. (2018). Overview of the Intel® RealSense™ Depth Camera. [online] Available at: https://software.intel.com/en-us/realsense/d400 [Accessed 16 Apr. 2018].