Asian Journal of Research in Computer Science

4(3): 1-8, 2019; Article no.AJRCOS.52396
ISSN: 2581-8260

A Smart Vision Based Navigation Aid for the Visually Impaired

Benjamin Kommey[1*], Kumbong Herrman[1] and Ernest Ofosu Addo[1]

[1]Department of Computer Engineering, Faculty of Electrical and Computer Engineering, College of Engineering, Kwame Nkrumah University of Science and Technology (KNUST), Kumasi, Ghana.

Authors' contributions

This work was carried out in collaboration among all authors. Author BK brought the idea and designed the study, the system architecture, block diagrams, wrote the abstract, introduction and conclusion, as well as the first draft of the manuscript. Author KH performed all the literature reviews, experimental setups, data collection and analysis. Author EOA designed the proposed model, managed the analyses of the study and the final write up. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJRCOS/2019/v4i330114
Editor(s):
(1) Dr. Stephen Mugisha Akandwanaho, Department of Information Systems and Technology, University of KwaZulu-Natal, South Africa.
Reviewers:
(1) Nadia Kanwal, Lahore College for Women University, Pakistan.
(2) Farooq Aftab, University of Science and Technology Beijing, China.
Complete Peer review History: http://www.sdiarticle4.com/review-history/52396

Original Research Article

Received 12 September 2019
Accepted 16 November 2019
Published 22 November 2019

ABSTRACT

Due to the ever increasing number of blind and visually impaired people in the world, there has been a great amount of research dedicated to the design of assistive technologies to support them. The various assistive technologies apply different techniques including laser, ultrasonic sensors and image processing. Autonomous navigation is a significant challenge for the visually impaired, it makes life uncomfortable for them and poses serious safety issues. In this paper we review the progress made so far in vision based systems and propose an approach for developing navigation aids through techniques used in other autonomous systems like self-driving vehicles. The proposed system uses a front camera to capture images and then produces commensurate guiding audio signals that allow the user freely move in their environment. An extra rear camera is included to allow the user to obtain more information about the scene. Care is taken however not to overload the user with information. The proposed method is tested both in indoor and outdoor scenes and is effective in notifying the user for any obstacles. The goal of this paper is to propose a model for and to develop subsystems for an intelligent, high performance, affordable and easy to use image based navigation aid for the visually impaired.

**Corresponding author: E-mail: bkommey.coe@knust.edu.gh;*

Keywords: Assistive technologies; autonomous; navigation aid; visually impaired; YOLO.

1. INTRODUCTION

As of 2015, 36.0 million people were estimated to be blind, 216.6 million people had moderate to severe visual impairment and 188.5 million had mild visual impairment [1]. This is a significant increase from previous years. It is also projected that 114.6 million people would be blind in 2050 [1]. The large number and potential increase in the number of visually impaired has motivated a lot of research in the field of Assistive Technologies.

The term "Assistive Technologies" (AT), in the broadest sense, refers to any set of scientific achievements (products, environmental modifications, services and processes) useful to overcome limitations and/or improve function for an individual [2]. ATs run the gamut of applications from personal mobility, to recreation and sports. The specific design considerations and type of technology used are dictated by the AT's application. While there exists a plethora of ATs that are efficient in specific tasks, a generalized AT is too complex a concept to be fully formalized. Even though the overall aim of AT design is to provide as much autonomy as possible for the users, it is hard to describe what a fully-autonomous AT would be so we settle for a semi-autonomous AT. In this paper we refer to a semi-autonomous assistive technology for visually impaired (SAATVI) as one that can:

(i) Provide visual self-localization and mapping.
(ii) Detect and track objects in the user's environment.
(iii) Adapt to a new environment.
(iv) Recognize human activity.
(v) Appropriately inform the user about their environment.
(vi) Initiate or influence the user to take appropriate decisions based on the above points.
(vii) Permit the user to navigate their environment with minimal assistance.

Electronic travel aids (ETAs) are a specific type of AT that provide the visually impaired with a means of navigating on their own. A number of different technologies have since been used to develop ETAs for the visually impaired. B Ando [3] roposed a multi-sensor approach to assist visually impaired people in specific urban navigation tasks. Sung et al. [4] have also designed a laser cane to replace the traditional guide cane. SWAN [5] is a Wearable Audio Navigation System developed by Georgia Tech. vOICe [6] developed by Meijer ushered in a new paradigm in the design of ETAs for the visually impaired; that of using computer vision and guiding the user through auditory signals. A more recent trend however is to use a combination of one or more of these (laser, sensors, computer vision, GPS).

A seemingly parallel yet strikingly similar engineering endeavor is that of self-driving cars. The progress achieved in the design of electronic travel aids for the visually impaired does not match that attained by autonomous vehicles. The 2005 DARPA Grand Challenge and the 2007 DARPA Urban Challenge demonstrated a lot of potential in the field of autonomous systems. In the 2007 DARPA Urban Challenge, 6 out of the 11 autonomous vehicles in the finals successfully navigated an urban environment to reach the finish line [7]. The goal of this paper is to discuss the adoption and application of techniques used in these semiautonomous vehicles as well as other autonomous systems for use in designing a navigation aid for the visually impaired. We begin by drawing analogies between the two systems and then systematically build model that meets our criterion for a SAATVI.

2. RELATED WORKS

Computer vision techniques for developing ETAs generally produce guiding audio signals from images of the scene. Semantic as well as non-semantic approaches have been developed so far. vOICe [6] is a non-semantic image processing technique that uses the "Piano Transform" conversion principle to map an image to a particular sound. In his method, the image is first divided into a 2 dimensional array of pixels. A mapping function uses the column, row and gray level of each pixel to produce a corresponding sine wave. The audio signal sent to the user is then a superposition of these sinusoids. A similar system to vOICe is NAVI (Navigational Assistance for the Visually Impaired) [8] that plumbs image processing methodologies applied to navigation assistance for visually impaired. In NAVI, a 2D image of the scene is resized to 32 x 32 and the gray scale is reduced to 4 levels. The image is then differentiated into objects and background. Conversion of the processed image into stereo sound is done by selecting the amplitude of the

stereo sound to be directly proportional to the intensity of image pixels.

There is one common difficulty in both the vOICe and NAVI systems, the distance between the user and the obstacle cannot be obtained directly by the users.

Due to the high computational complexity of vOICe, Xuan Zhang et al. [9] proposed a modification that uses Inverse Fast Fourier Transform (IFFT) to perform image-sound conversion.

In general, the drawback of non-semantic techniques is that, users of the system have to be trained to identify the different sound patterns.

Semantic techniques map sound to the image based on the information on the image. Rui et al designed a model that detects objects from a scene, represents them with their names and then converts them to 3D bin-aural sound [10]. This method however, fails when objects are too close or too far. It also causes information overload by trying to notify the user of too many objects.

Bogdan et al. [11] designed an automatic cognition system that is able to detect, track and recognize, in real-time, all relevant objects existent in the scene without any prior knowledge about shape, position or dynamics. Acoustic warning signals are transmitted to the user via bone conducting headphones. Just like in the case of vOICE and NAVI, the user needs appropriate training to make sense of the warnings.

Tyflos [12] is a state of the art wearable assistive technology designed by Bourbakis. It provides the visually impaired user with reading and navigation capabilities. It integrates a wireless portable computer, cameras, range and GPS sensors, microphones, natural language processor, text-to-speech device, an ear speaker, a speech synthesizer, a 2D vibration vest and a digital audio recorder. Data collected by the Tyflos sensors is processed by appropriate modules, each of which is specialized in one or more tasks. The system is however large and may be tedious for the user to carry around. In addition, the system does not provide the capability for the user to make sense of the activities of other humans thus providing little or no support for interaction with others.

LookTel [13] is a mobile application built for assisting the visually impaired. In order to provide extra computational power, it uses micro-cloud computing. LookTel employs the SIFT algorithm for object and location detection. It has an XML based meta-information system database and an OCR engine helps the visually impaired user in reading. The system also has GPS capabilities for navigation purposes. While it is a promising system, it struggles to adapt to new environments and its size may pose problems.

3. MATERIALS AND METHODS

3.1 SmartVisionNavi System Architecture

The SmartVisionNavi system architecture (Fig. 1) and the proposed model, as shown in Fig. 2 consists of an image acquisition unit, an image processing unit, an earphone and an external memory module.

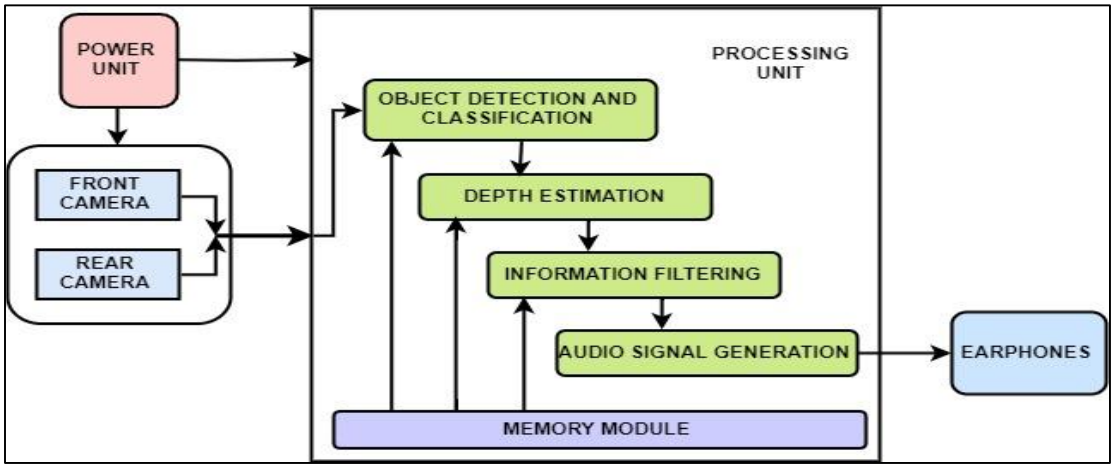

Fig. 1. SmartVisionNavi system architecture

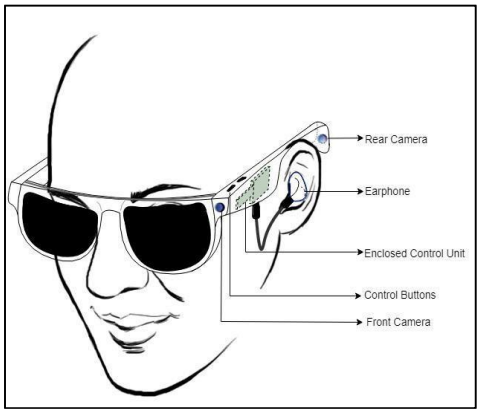

Fig. 2. SmartVisionNavi proposed model

The image acquisition unit is responsible for providing a comprehensive display of the user's environment. To do this we use a front and a rear camera to ensure a thorough view. The main role of our embedded control unit is to detect and classify objects. For the purpose of this project we utilize an ATmega328. Since a number of objects can be detected, it is important to filter the information outputted by the control unit to ensure that it only provides the user with information relevant to their navigation. Also key to our model is the method of generating guiding sounds from the detected objects. An extra memory module stores information relevant for image processing and audio generation as well as system parameters. Below we discuss the functioning of some of each of the specific modules.

3.2 Object Detection and Classification

Visually impaired users are usually required to navigate through a constantly changing environment. For instance it could be a street with cars and people constantly moving. To enable the user to rapidly apprise and rightly react to changes or signals from their environment, it is required that object detection should be fast, accurate, and able to recognize a wide variety of objects. By using neural network detection frameworks, we can gain significant speed and accuracy. However, the small set of objects these can detect limits their application in such a scenario. In this project we employ YOLO9000, a state-of-the-art, real-time object detection system that can detect over 9000 object categories. In the YOLO model [14], a single neural network predicts bounding boxes and class probabilities directly from full images in one evaluation. YOLO9000 is jointly trained on the COCO detection dataset and the ImageNet classification dataset. Fig. 3 shows the convolutional neural network model for YOLO. The object detection time $t_o < 1s$.

3.3 Depth Estimation

Just like a robot, a visually impaired user can plan its movement with precision if it can estimate the distance between it and objects in its environment. Therefore, an accurate calculation of the distance of the user from the nearby objects of interest is key to our model.

Most depth estimation schemes make use of vision sensors. However, a single camera can only give a 2-D projection of a real world 3-D projection. This results in information loss and therefore it is difficult to estimate depth information In order to estimate the 3-D coordinates of the object of interest we need to obtain additional information. Precision, speed and accuracy are important for whatever model we adopt for depth estimation. That said, we can approach the problem by using multiple sensors or by using an algorithmic oriented solution.

The easiest approach is to use dual camera separated by a small distance (usually a few millimeters) to capture images from different viewpoints. These two images form a stereo pair, and is used to compute depth information. This is however expensive since we will require two extra cameras, one for the rear and another for the front.

We employ a depth estimation technique developed by Liaquat et al. [15]. The proposed depth estimation technique uses only a single camera. The proposed algorithm exploits the fact that an object which is close to the camera looks bigger in the 2-D image plane and keeps getting smaller as the camera moves away from the object and is divided into two major stages. First the objects of interest are detected and classification is performed. After classification, the bounding boxes of the objects are used to calculate their areas. A curve fitting equation derived from trained data is used in conjunction with the area to estimate the distance from the camera. The depth estimation time $t_d < 0.5s$.

3.4 Information Filtering

YOLO outputs the top classes and their probability for each frame. Any probability above 25% is considered a confident detection result.

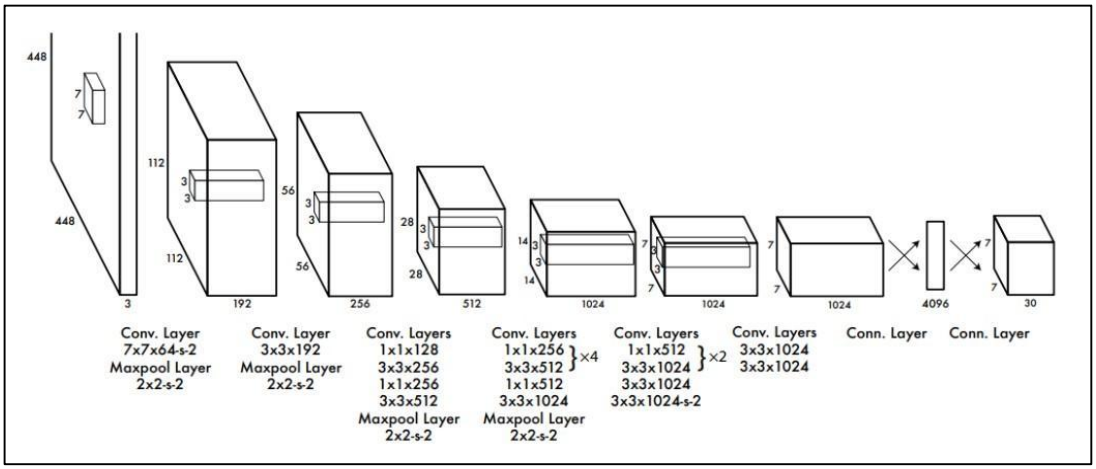

Fig. 3. Convolutional neural network of the YOLO model

Since multiple objects may be detected above the 25 % threshold, we need to present only those objects that could obstruct the user or are important in their navigation. Furthermore presenting multiple objects within a short time span may overload the user with information. Based on the depth estimate for each object, we only present the user with information about objects that are within 4 meters of the user's location. In addition, we also only provide the user with information every two seconds.

3.5 Audio Signal Generation

We have a wide variety of audio methods to notify the user. Non semantic techniques like that proposed by Meijer [6] require extensive training of the user, so we opt for a semantic approach. Rui et al. [10] proposed a method of audio signal notification that uses 3D sound based on the Unity 3D game engine called 3DCeption. The 3DCeption renders the binaural sound effect with the help of the Head-Related Transfer Function (HRTF) to simulate the reflection of the sound on human body (head, ear, etc.) and obstacles (such as wall and floor). This method while accurate, requires extra processing and may increase the size of our system.

Since we can obtain the distance from the user and the name of the objects in the user's location, we will rather use these to notify the user. A visually impaired user may not be able to accurately measure 4 meters but can intuitively make an estimate. The name and position (in front or behind) of each object within 4 meters of the user are sent to a text to speech converter. The audio is then sent to the user via the earphone to serve as a guiding signal. The visually impaired user can be guided to make better decisions in planning their path based on this information about the environment provided. The audio generation time $t_g < 0.5s$. The overall processing time for the system is bounded by:

$$t_o + t_d + t_g < 2s$$

3.6 System Work Flow

The user has to manually turn on the system for it to begin execution. As soon as the system is on, the front and the rear cameras capture images which are fed to the object detection and depth calculation units. Both a front and a rear camera are used so as to provide a better view of the user's environment. The output of these stages is filtered to ensure that the user is only provided with relevant information necessary for their travel. An audio signal is generated based on the need of the system to notify the user or not. The system continues execution until it is stopped by the user. Fig. 4 demonstrates the system work flow.

4. TEST AND RESULTS

In order to validate the proposed model, the system was tested in both indoor (Fig. 5) and outdoor (Fig. 6) environment. Particular attention was paid and investigated in the object detection, depth estimation, information filtering and audio generation. The information filtering and audio generation follow from the detection and depth estimation. All modules were well behaved and functioned as expected. Tables 1 and 2 contains some sample test results. In each case we detect

the objects in the scene and determine if the user is notified of their presence or not. Our motive is to ensure that the user is notified of all objects within 1 meter of their location.

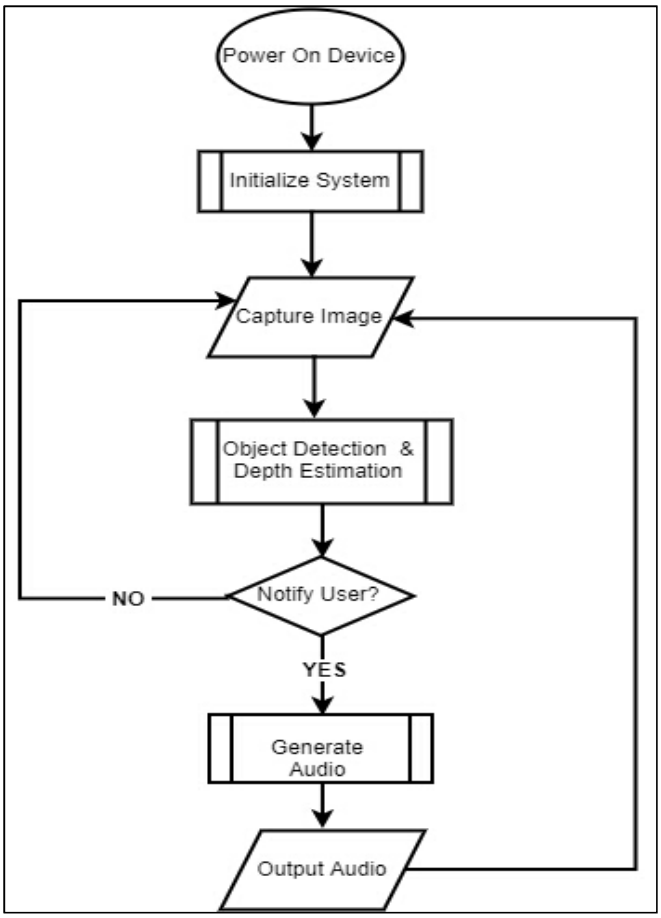

Fig. 4. SmartVisionNavi system work flow

Fig. 5. Outdoor settings detection

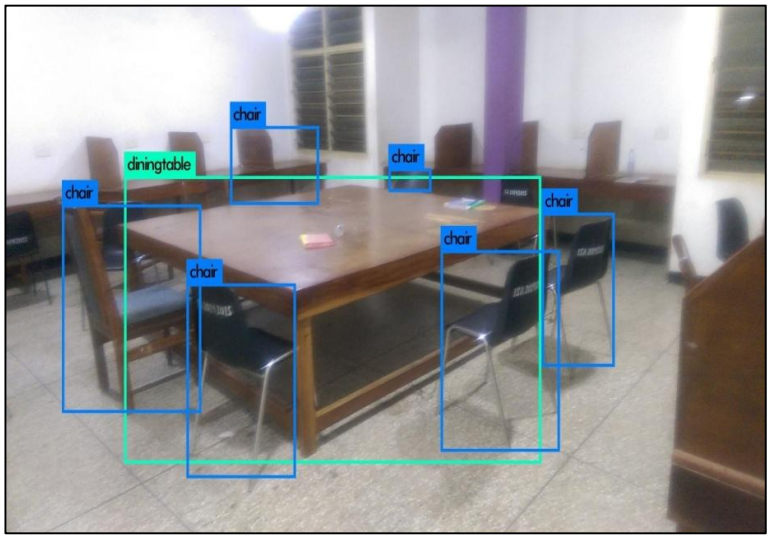

Fig. 6. Indoor settings detection

Table 1. Detection in outdoor settings

Object	Probability	Distance(m)	Should notify	Notified
Person	1.0	0.40	Yes	Yes
Person	1.0	1.65	Yes	Yes
Handbag	0.89	1.64	Yes	Yes
Car	0.98	5.12	No	No
Car	0.96	9.13	No	No
Person	0.99	8.97	No	No
Person	0.99	8.92	No	No
Person	0.99	8.94	No	No
Person	0.99	11.26	No	No
Person	0.98	11.73	No	No
Person	0.98	12.14	No	No
Person	0.97	12.56	No	No
Person	0.92	12.78	No	No
Person	0.91	12.92	No	No
Person	0.91	13.10	No	No
Person	0.90	13.25	No	No
Person	0.87	14.01	No	No
Person	0.84	14.15	No	No

Table 2. Detection in indoor settings

Object	Probability	Distance(m)	Should notify	Notified
Dining table	0.83	0.79	Yes	Yes
chair	0.99	0.81	Yes	Yes
chair	0.99	0.97	Yes	Yes
chair	0.97	1.23	Yes	Yes
chair	0.96	1.52	Yes	Yes
chair	0.72	3.97	Yes	Yes

5. CONCLUSION

In this work, we proposed a model for a travel aid for the blind and visually impaired people. We presented a functional requirements and analysis of each of the modules. We consistently borrowed techniques from other autonomous systems to enhance our system.

A large amount of capital is currently being invested in new technology for autonomous driving, accelerating research progress in this field. We expect that such results can be highly beneficial in the research of assistive tools to help navigate blind and visually impaired people, and improve their quality of life. We have indeed demonstrated that this is possible. In the future, we will consider the motion direction, velocity and acceleration of objects in the user's environment as well as their position. In addition, work will be done to develop more accurate depth estimation algorithms and efficient methods of outputting information. Furthermore, work will be done on the design of more efficient algorithms for information filtering. Finally, attention will also be paid to extending the work to take input from multiple sensor types.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Bourne, et al. Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: A systematic review and meta-analysis. Lancet Glob Health. 2017;5: e888–97.
   (Published Online August 2)
2. Cook, et al. Assistive technologies: Principles and practice. Elsevier Health Sciences; 2014.
3. Ando B. A smart multisensor approach to assist blind people in specific urban navigation tasks. IEEE Transactions Neural Systems and Rehabilitation Engineering. 2008;16(6):592-594.
4. Sung, et al. Development of an intelligent guide-stick for the blind. IEEE International Conference on Robotics and Automation Seoul, Korea, May 21- 26; 2001.
5. Ivanov R. Real-time GPS track simplification algorithm for outdoor navigation of visually impaired. Journal of Network and Computer Applications; 2012. DOI: 10.1016/j.jnca.2012.02.002
6. Peter B, Meijer L. An experimental system for auditory image representations. IEEE Transactions on Biomedical Engineering. 1992;39(2).
7. Lex Fridman. Large-scale deep learning based analysis of driver behavior and interaction with automation in MIT autonomous vehicle technology study; 2017.
8. Aladrén, et al. Navigation assistance for the visually impaired using RGB-D sensor with range expansion. In IEEE Systems Journal. 2016;10(3):922-932. DOI: 10.1109/JSYST.2014.2320639
9. Zhang, et al. An efficient method of image-sound conversion based on IFFT for vision aid for the blind. Lecture Notes on Software Engineering. 2014;2(1).
10. Rui, et al. Let blind people see: Real-time visual recognition with results converted to 3D audio in Stanford CS231n reports; 2018.
11. Bogdan, et al. Seeing without sight – An automatic cognition system dedicated to blind and visually impaired people. 2017 IEEE International Conference on Computer Vision Workshops; 2017.
12. Bourbakis, et al. A multimodal interaction scheme between a blind user and the Tyflos assistive prototype. In Tools with Artificial Intelligence, ICTAI '08. 20th IEEE International Conference on. 2008;2:487-494.
13. Sudol, et al. LookTel — Computer vision applications for the visually impaired. UCLA: Computer Science 0201; 2013. Available:http://escholarship.org/uc/item/57q1b167
14. Redmon J, et al. YOLO9000: Better, faster, stronger. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI. 2017;6517-6525. DOI: 10.1109/CVPR.2017.690
15. Liaquat, et al. Object detection and depth estimation of real world objects using single camera. Fourth International Conference on Aerospace Science and Engineering (ICASE), Islamabad; 2015.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sdiarticle4.com/review-history/52396