Computer Vision based Assistive Technology For Blind and Visually Impaired People

Shankar Sivan.*
Sree Buddha College Of Engineering
Pattoor, Alappuzha
Kerala, India
shankarsivan20@gmail.com

Gopu Darsan †
Sree Buddha College Of Engineering
Pattoor, Alappuzha
Kerala, India
gops601@gmail.com

ABSTRACT

The computer vision based assistive technology for the blind and visually impaired is a developing area. The assistive technology helps the visually impaired by providing them with a greater independence. By enabling them with their day-to-day activities like indoor and outdoor navigation, obstacle detection, locating the doors and lost objects, etc. Even though different assistive technologies are available for the blind, most of them have complex designs which are developed for a specific purpose and are expensive for the commercial production. Rather than depending on a traditional white cane, the blind and visually impaired people can make use of the cheaper assistive device proposed in this paper. The proposed system incorporates several assistance features in a device which will be an asset for them according to their needs.

CCS Concepts

•**Computing methodologies** → *Computer vision;*

Keywords

Assistive technology; Computer vision; Image processing; electronic travel aids; wearable systems

1. INTRODUCTION

Computer vision is a field that deals with acquiring, processing, examining and understanding the images. Output is in form of description or an interpretation or some quantitative measurements to obtain an understanding high-dimensional data from the real world in order to produce numerical or symbolic information to make a decision.Computer vision is an area for duplicating the abilities of human vision. Which also known as Image Analysis, Scene Analysis, Image Understanding, Robotics ,Artificial Intelligence , Computer Graphics, Pattern Recognition. The computations are done by electronically comprehending and apprehending an image. As a scientific discipline, computer vision is concerned with the theory behind artificial systems that extract information from images. The image data can take many forms, such as video sequences, views from multiple cameras, or multi-dimensional data from a medical scanner. As a technological discipline, computer vision seeks to apply its theories and models to the construction of computer vision systems.

The World Health Organization estimates that there are around 39 million blind people around the globe. Thus there is a need for assistive and rehabilitative devices. The most popular aid for the blind is by using the white cane with a guide dog to avoid obstacles. The brain plasticity enable the blind to use their occipital lobe to perceive the object through other sensory modalities, thus the blind people localizes the dynamic obstacles through the sense of hearing but in an unknown environment this tent to be a challenge to determine the object. Over years different commercial applications were developed and among them the most popular application's are the GPS powered applications like Mobile Geo, Braille Note GPS, MoBIC etc., computer vision based application like The vOICe, NAVI, ENVS, TVS etc. and several other prototypical applications. The past few decades saw the tremendous growth in the computer system hardware. This has lead to cheaper and compact sized high performance computers which enables the scientists and researchers to create handheld and wearable devices for the assistance of blind and visually impaired people. This paper describes the various computer vision based assistive technology which was developed for them and proposes a cheaper and efficient system.

2. RELATED WORK

2.1 The vOICe

The vOICe [1] [1] vision technology is a sensory substitution system for the totally blind which gives a visual experience through live camera views by image-to-sound renderings. The sensory substitution technology make use of the neural plasticity of the human brain i.e., sensory substitution make use of the human brains ability of cortical remapping [2]. Which enhances the senses of other sensory modalities, for

*Post Graduate Scholar
†Assistant professor

[1] www.seeingwithsound.com

ICCCNT '16, July 06-08, 2016, Dallas, TX, USA
© 2016 ACM. ISBN 978-1-4503-4179-0/16/07…$15.00
DOI: http://dx.doi.org/10.1145/2967878.2967923

Figure 1: The vOICe system

Figure 2: The ENVS components

example the blind people can use their occipital lobe to comprehend objects through the use of other sensory modalities. The vOICe technology constitutes a live feed from a head-mounted camera, the image is scanned from left to right of the video frame which is in turn converted into soundscapes. The audio mapping is done by associating height to pitch and brightness to loudness. The vOICe requires a minimum amount of training and effort. EyeMusic [3] and PSVA [4] are sensory substitution systems similar to The vOICe for aiding blind and visually impaired people.

2.2 Electro-Neural Vision System

Simon Meers and Koren Ward developed a visual substitution system called Electro-Neural Vision system(ENVS) [5]. The ENVS provide a virtual perception of the three-dimensional profile and colour of the surroundings through electrical pulses. The ENVS system comprises of a stereo camera which can capture the image and calculate the disparity depth map indicating the distance to each point of the image. Special gloves with electrodes are used to deliver the electrical pulses to the fingers. Transcutaneous Electro-Neural Stimulation unit in the ENVS system samples the depth value from the computer and converts them into an electro-neural pulse. Then these pulses are delivered to the fingers through electrodes in gloves. The intensity of the pulse varies with distance sampled corresponding to the depth map region. Colors of the surroundings is determined from the pulse frequency. ENVS system enables a blind person to have a virtual experience by creating a mental map of the 3D profile of his/her surroundings.

2.3 Clear Path Guidance for Blind

Volodymyr Ivanchenko, James Coughlan, William Gerrey and Huiying Shen [6] developed a system to assist the navigation of blind people who uses wheelchair. For the visually impaired wheelchair riders, it is extremely difficult to travel. As they are unaware about the hazards until it is too late. Also it is difficult to maintain the orientation while traveling on a straight line. The clear path guidance system informs user about the terrain information using the computer vision based range sensor. The two stereo camera mounted above riders head which connected to a computer analyses the terrain by tracking the traditional white cane used by the blind people. System alerts the user vocally regarding walls and obstacles in the direction to which the white cane is pointed.

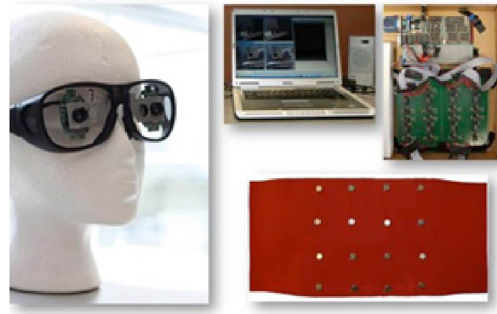

Figure 3: Tyflos Navigator Prototype

2.4 Tyflos

Tyflos [7] is an electronic travel aid which was initially developed by Dr. Bourbakis and latter additional features were incorporated to the system. Tyflos prototype integrates a portable computer, cameras and GPS sensors, microphones, text-to-speech converter, language processor, a 2D vibration vest, a speech synthesizer and an audio recorder. The tyflos system has a stereo vision module which is attached on conventional eyeglasses. This stereo vision system captures environmental data and process them. From the data acquired, the system creates a depth map of the 3D environment of the surroundings. Tyflos system has a vibratory belt which is worn by the blind person on his/her abdomen. The vibratory belt has a two dimensional array of 16 elements. The depth map is mapped to a tactile vocabulary and the user can sense them through the vibratory belt and locate obstacles for a safe navigation.

2.5 Virtual White Cane

Roberto Manduchi and Dan Yuan [8, 9] developed a laser-based mobility device which make use of the computer vision technology. The hand held device can be used as an alternative to the traditional white cane used by the blind people for navigation. The user receives feedback about his/her surrounding through a tactile interface and audio signals, from which the blind person can make a mental image of the scene. Device scans surroundings with a laser pointer which is combined with a digital camera and a computer processor. The surroundings spacial information are gathered and analyzed as the user moves around. The system produce special sounds for steps, curb or a drop off thus making the navigation more comfortable.

2.6 FingerReader

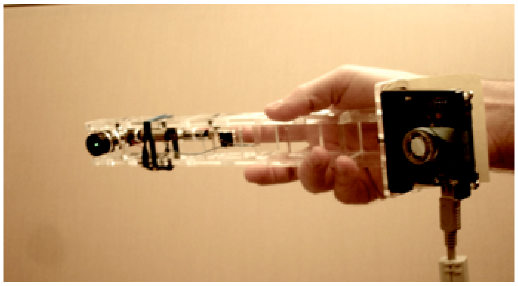

Figure 4: The "virtual white cane"

Figure 5: FingerReader

FingerReader [11] is a wearable text reading device to assist the blind people and dyslexic readers. FingerReader is an accurate and efficient system which make use of the computer vision technology to scan the printed text. System has a scene and finger detection, which tracks the fingertip to localize a horizontal focus region, which can be adjusted as a parameter. After the text line is extracted, the Tesseract Optical character recognition(OCR) engine is used to extract the words. OCR extracts one word at a time and uttered to the user. Limitation of the technology is that the camera does not auto focus and continuous feedback is needed. Hindsight [10] is similar to the FingerReader, which also uses Tesseract OCR engine to detect the text. Computer vision algorithms are used for deblurring and stabilizing the image to maximize the reading speed.

3. FUNCTIONALITIES AND ALGORITHMS

Computer vision has various image processing capabilities which analyses the image, data acquired is used for various computations. The functionalities and algorithms which were used for development of the proposed system are described in following sections.

3.1 Histogram Equalization

Histogram is the graphical representation of distribution of intensities in an image. The histogram quantifies number of pixels for each value that is considered. What histogram equalization does is stretching the intensity of an image there by enhancing contrast of the image. Mapping of one distribution to another such that, the intensity value of the image spreads over a much larger range. Remapping is done using cumulative distribution function. i.e, Let histogram be, $\mathbf{H}(\mathbf{i})$ then the cumulative distribution is given by

$$\mathbf{H}'(\mathbf{i}) = \sum_{j=0}^{j<i} \mathbf{H}(\mathbf{j})$$

such that the maximum value of $\mathbf{H}'(\mathbf{i})$ is 255.

3.2 Canny Edge Detector

Canny Edge detector [12] was developed in the year 1986 by John F. Canny. The Canny Edge detector has lower error, good localization and minimal response. The Canny Edge detector algorithm goes through following steps, initially noises in the image are filtered out using Gaussian filter [14]. In the next step we find the intensity gradient of the image here after the computation we obtain the gradient strength and direction. Next step the non-maximum suppression is applied i.e., the removal of pixels which are not considered as the part of an edge. The final step is called hysteresis where thresholding of the image is done, in this step the pixels with higher gradients are accepted and the pixels with lower threshold gets rejected. The pixels in-between threshold will only be accepted if they are connected to the pixel that is above the upper threshold.

3.3 BRISK:Binary Robust Invariant Scalable Keypoints

BRISK [15] algorithm is used for the description matching for the object detection module of the system. BRISK was preferred over SURF [13] and SIFT [16] due to its high performance, also offers a much easily configurable circular sampling pattern which computes th brightness comparisons to form a binary descriptor string. BRISK is useful in real time projects with limited computation power. The keypoint detection in BRISK is inspired by the work of E. Mair [17]. During the key point description BRISK make search for the maxima not only in image plane but also for saliency using FAST [18] scores, thus BRISK estimates the true scale of keypoints in continuous scale-space. For the keypoint description BRISK select the brightness comparisons for a much better descriptiveness. Also BRISK make use of the orientation-normalized descriptors to identify the direction of each keypoint and the descriptor matching is done using BRIEF [19].

4. PROPOSED ARCHITECTURAL MODEL OF THE ASSISTIVE DEVICE

In this section key stages of the proposed system are explained. The proposed system incorporates text recognition, object identification, door detection and security system into a single device. The architecture constitutes an image capturing system which is used to capture images for the real time applications. A processing system which is a raspberry pi system running GNU-Linux operating system. The user activates different modules through an input device. The output is provided to the user through a speech system.

4.1 Object Detection Module

The object detection module in the system will help the blind by locating the lost object or while searching for a particular object. The system consist of a video camera capturing the scene. When user activates the object detection module, a frame is taken. The captured frame goes

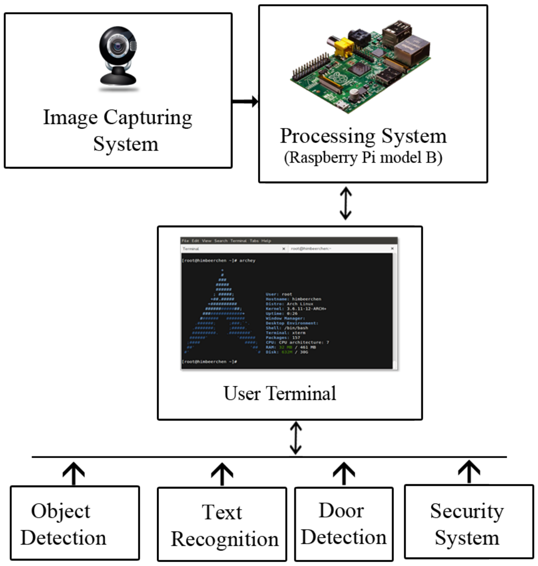

Figure 6: Architecture Of The Proposed System

into the initial preprocessing of the image, where the image enhancement is done. The frame taken at time t is compared with the previous image frame, if they are same we ignore the frame and provide the user with previous output. Thus reducing unwanted computation and improving performance. Now the enhanced image passes through the BRISK keypoint detection where the true scale of each keypoint in continuous scale-space is estimated. Now the BRISK keypoint descriptor samples after the analyses of keypoints, containing the sub-pixel image location and associated floating point scale values. To avoid the aliasing effect on image intensity in the pattern Gaussian smoothening is applied. Due to deterministic sampling pattern a uniform sampling-point is generated around the keypoint. BRISK also uses fewer sampling-points rather than pairwise comparisons thereby reducing the complexity. Now keypoint matching is done using BRIEF with the database containing features extracted from registered objects. If the object is detected the user is provided an audio output of the object name.

4.2 Text Recognition

The text recognition module of the system enables blind people during navigation, reading warning boards, door signs etc., The text recognition module consist of a camera which captures the image, this goes through an initial preprocessing. During this stage the image enhancement like deblurring and stabilization of the images is done. Now based on background color of the image it undergoes two stages. Gray scale conversion and binarization [21] of the image. If the text is written in a lighter color and a dark background then the image goes through a gray level conversion and then binary conversion of the image occurs. But if the text is dark in a light background then, the image initially go through an inverted gray level conversion and then an inverted binary conversion of the image. Now after the few more enhancement the final binary image is selected and passed on to the Tesseract OCR engine [20]. The output obtained is provided to the speech system.

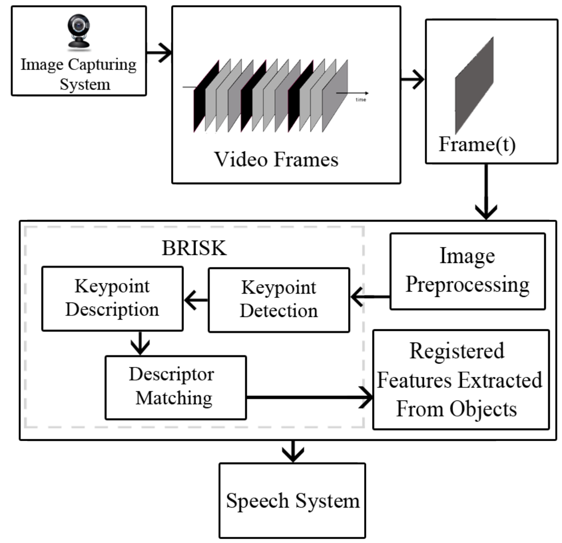

Figure 7: Object Detection Module

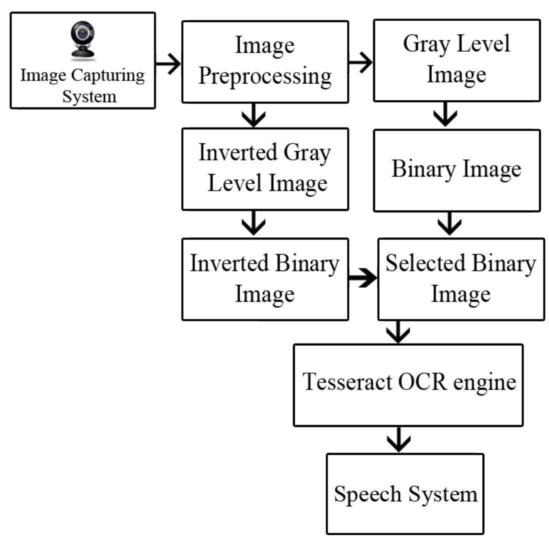

Figure 8: Text Recognition System

Figure 9: The frame(t) and Registered objects in the database.

4.3 Door Detection

The door detection is a useful feature for the blind during navigation. System captures the scene and after initial preprocessing, to reduce the noise and unnecessary corners. Now the Canny edge detector is used to create a binary edge map. Based on the corner detector proposed by Nelson H. C. Yung, Xiao Chen He [22] an edge map based on global and local curvature is made. Even if there is any hidden surface the four corners of the door-frame is detected. Thus on detection of a door the blind is provided with an audio alert.

4.4 Security System

When the security system module is activated, it alert the user if an intruder enters into the scene and saves the image. Camera in the system monitors scene in constant intervals of time and the frame captured goes into the processing system by making use of a simple background subtraction between the initial and current frame the system monitors for a change and by using Haar Feature-based Cascade Classifier [23] [24], if the intruder(s) enters current frame, the blind person is alerted through a speech system.

5. EXPERIMENTAL RESULTS

The proposed system was tested on Raspberry Pi 3 Model B device. The processor used was ARM Cortex A53 with C.P.U. speed of 1.2 GHz, running a GNU-Linux distribution Archlinux-ARM operating system. The system had 1 G.B. RAM and 8 G.B. ROM. The image capturing system used Logitech C310 webcam with 5 mega-pixel camera. The project was implemented in C++ programming language using Intel IPP - Open Source Computer Vision Library

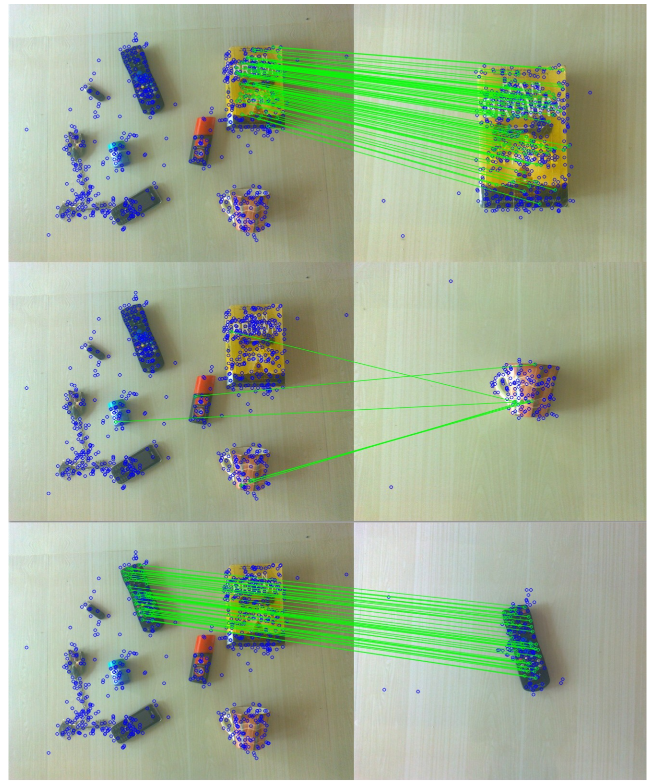

Figure 10: Object Identification : Using SURF

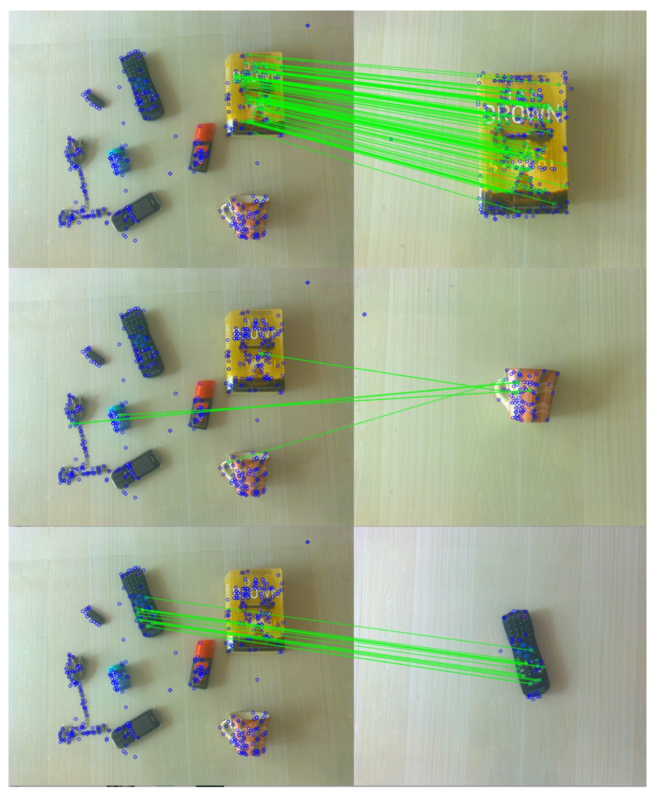

Figure 11: Object Identification : Using SIFT

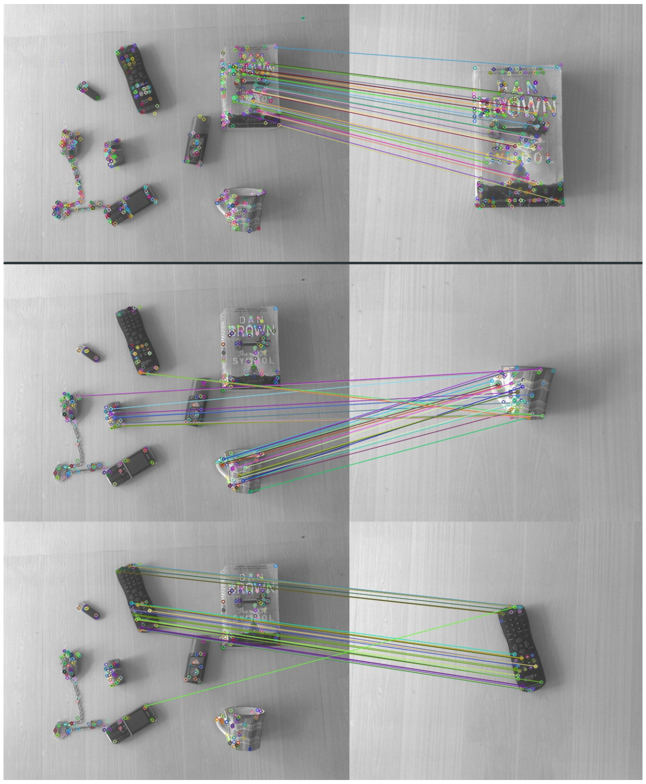

Figure 12: Object Identification: Proposed System

(OpenCV) version 2.4 [2]. For the speech synthesizing an opens source software eSpeak[3] was used.

During experimentation, when the object detection module was activated camera captured the scene containing different object. Figure 9 shows the frame which was captured at time t and the objects that are registered in the database. Existing system make use of SURF and SIFT algorithms for object identification. So the proposed system was evaluated against them. Table 3 reveals that in real-time the proposed system making use of BRISK is faster than the other SURF and SIFT.

Figure 10, 11 shows the object identification using SURF and SIFT, Figure 12 gives results of the proposed system. Table 1 and 2 reveals that overall output performance of the proposed system in identifying object A and C is similar. But the performance in identifying object B is better in proposed system. Further experiment revealed that the device is efficient in identifying the objects. After object identification output is provided to the user through speech synthesizer.

Figure 13 shows the execution and result of the text recognition module, the captured image of the road sign (bottom left) and the image after computation (bottom right). The terminal shows output "ROAD ENDS" after OCR computation. The result is piped to the speech synthesizer to provide output for the user.

Figure 14 gives result of the door detection module, when the blind search for door in his surrounding the corresponding frames are searched for a door and if found the user is

[2]http://opencv.org/
[3]http://espeak.sourceforge.net/

Figure 13: Text Recognition

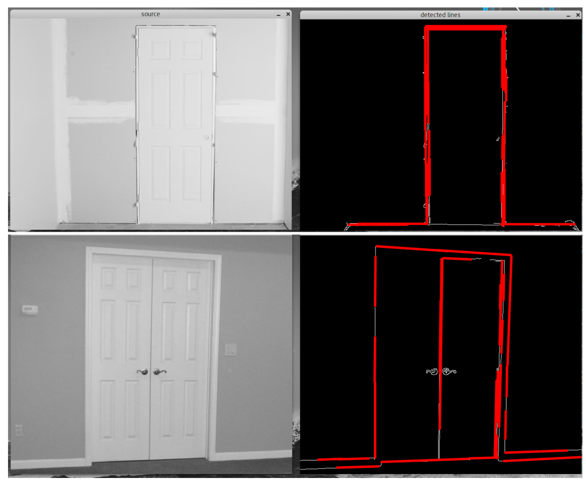

Figure 14: Door Detection

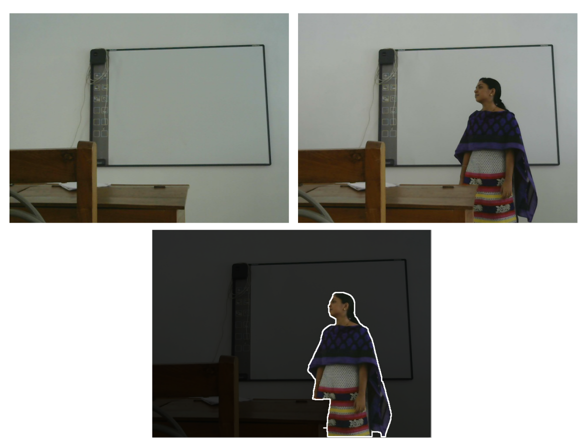

Figure 15: Security system: Intruder Detection

alerted. Figure 15 gives the results security system when it is activated by the user. When an intruder is detected the user gets alerted.

During experimentation it was discovered that their are certain limitation to the system. While capturing the images sometimes problem arises with an unusable image, due to blurring caused by focus and improper holding of the camera and dept of the field. To a limit this can be solved by using a deblurring and an auto focus system. But few of the images were not good enough and thus had to be discarded. System performance was also compromised with the processor speed. As the technology developing exponentially this could be solved in the future, with a much better performing compact and cheaper processor.

Table 1: Mathching Lines in Object Identification

	OBJECT A	OBJECT B	OBJECT C
SURF	42	6	39
SIFT	88	7	14
BRISK	44	23	27

Table 2: Correct Match in Object Identification

	OBJECT A	OBJECT B	OBJECT C
SURF	42	2	39
SIFT	88	1	14
BRISK	44	16	26

Table 3: Total Processing Time

	SURF	SIFT	BRISK
Processing Time(s)	3.602	2.8146	1.992

The cost for a Raspberry Pi device is 35$ and 15$ for the camera. System make use of open source softwares which are free of cost. BRISK can be used commercially without cost, unlike SURF and SIFT which are patent protected. Hence the total cost for the system can be minimized to 50$ - 60$. Thus the proposed system turned out to be a cheaper mobile assistive device.

6. CONCLUSIONS

The paper proposed a cheaper and efficient assistive device for the blind and visually impaired. The device has an object detection module which make use of BRISK, along with text recognition, door detection and a security system which detects intruders. The device can be used for in real-time thus it can be used as a mobile assistive device for the blind and visually impaired. The system is implemented in C++ using OpenCV library. Further researches will be done to improve and also to include additional features to the system.

7. ACKNOWLEDGMENTS

The authors would like to acknowledge Dr. S Suresh Babu, Anil A. R., Minu Lalitha Madhavu for their valuable inputs and Sree Buddha College of Engineering for providing the facilities for completion of the project.

8. REFERENCES

[1] Auvray M., Hanneton S., O'Regan J. K. (2007). Learning to perceive with a visuo-auditory substitution system: localisation and object recognition with "the vOICe". Perception 36 416-430. 10.1068/p5631.

[2] Muller M. M., Elbert T., Rockstroh B., Pantev C., Taub E. (June 1, 1998). "Perceptual correlates of changes in cortical representation of fingers in blind multifinger Braille readers". Journal of Neuroscience 18 (11): 4417-4423.

[3] Hornof, A., & Sato, L. (2004). EyeMusic: Making Music with the Eyes. Proceedings of the 2004 Conference on New Interfaces for Musical Expression (NIME04), Hamamatsu, Japan, June 3-5, 185-188.

[4] Capelle C, Trullemans C, Arno P, Veraart C (1998). "A real-time experimental prototype for enhancement of vision rehabilitation using auditory substitution.". IEEE Transactions Biomedical Engineering, 45: 1279-1293.

[5] Meers, S.; Ward, K., "A vision system for providing 3D perception of the environment via transcutaneous electro-neural stimulation," in Information Visualisation, 2004. IV 2004. Proceedings. Eighth International Conference on , vol., no., pp.546-552, 14-16 July 2004 doi: 10.1109/IV.2004.1320198

[6] Volodymyr Ivanchenko, James Coughlan, William Gerrey, and Huiying Shen. 2008. Computer vision-based clear path guidance for blind wheelchair users. In Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility (Assets '08). ACM, New York, NY, USA, 291-292. DOI=http://dx.doi.org/10.1145/1414471.1414543

[7] Bourbakis, N.; Keefer, R.; Dakopoulos, D.; Esposito, A., "A Multimodal Interaction Scheme between a Blind User and the Tyflos Assistive Prototype," in Tools with Artificial Intelligence, 2008. ICTAI '08. 20th IEEE International Conference on , vol.2, no., pp.487-494, 3-5 Nov. 2008 doi: 10.1109/ICTAI.2008.52

[8] Ilstrup, D., Elkaim, G., "Low Cost, Low Power Structured Light Based Obstacle Detection," ION/IEEE Position, Location, and Navigation Symposium, ION/IEEE PLANS 2008, Monterey, CA, May 5-8, 2008, pp. 865-870

[9] Yuan, D., and Manduchi, R., "A Tool for Range Sensing and Environment Discovery for the Blind", IEEE Workshop on Real-Time 3-D Sensors and Their Use, 2004.

[10] Stearns, L., Du, R., Oh, U., Wang, Y., Chellappa, R., Findlater, L., & Froehlich, J. (2014) The Design and Preliminary Evaluation of a Finger-Mounted Camera and Feedback System to Enable Reading of Printed Text for the Blind. European Conference on Computer Vision (ECCV) 2014 Workshops. pp. 615–631. Springer.

[11] Roy Shilkrot, Jochen Huber, Wong Meng Ee, Pattie Maes, and Suranga Chandima Nanayakkara. 2015. FingerReader: A Wearable Device to Explore Printed Text on the Go. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15). ACM, New York, NY, USA, 2363-2372. DOI=http://dx.doi.org/10.1145/2702123.2702421

[12] J Canny. 1986. A Computational Approach to Edge

Detection. IEEE Trans. Pattern Anal. Mach. Intell. 8, 6 (June 1986), 679-698. DOI=http://dx.doi.org/10.1109/TPAMI.1986.4767851

[13] H. Bay, A. Ess, T. Tuytelaars, and L.V. Gool. "SURF: Speeded Up Robust Features". Computer Vision and Image Understanding (CVIU), 110:346-359, 2008.

[14] Ian T. Young and Lucas J. van Vliet. 1995. Recursive implementation of the Gaussian filter. Signal Process. 44, 2 (June 1995), 139-151. DOI=10.1016/0165-1684(95)00020-E

[15] Stefan Leutenegger, Margarita Chli, and Roland Y. Siegwart. 2011. BRISK: Binary Robust invariant scalable keypoints. In Proceedings of the 2011 International Conference on Computer Vision (ICCV '11). IEEE Computer Society, Washington, DC, USA, 2548-2555. DOI=http://dx.doi.org/10.1109/ICCV.2011.6126542

[16] David G. Lowe. 2004. Distinctive Image Features from Scale-Invariant Keypoints. Int. J. Comput. Vision 60, 2 (November 2004), 91-110.

[17] E. Mair, G. D. Hager, D. Burschka, M. Suppa, and G. Hirzinger. Adaptive and generic corner detection based on the accelerated segment test. In Proceedings of the European Conference on Computer Vision (ECCV), 2010.

[18] L. Guo, J. Li, Y. Zhu and Z. Tang, "A novel Features from Accelerated Segment Test algorithm based on LBP on image matching," Communication Software and Networks (ICCSN), 2011 IEEE 3rd International Conference on, Xi'an, 2011, pp. 355-358. doi: 10.1109/ICCSN.2011.6013732

[19] M. Calonder, V. Lepetit, C. Strecha, and P. Fua. BRIEF: Binary Robust Independent Elementary Features. In Proceedings of the European Conference on Computer Vision (ECCV), 2010.

[20] R. Smith, "An Overview of the Tesseract OCR Engine," 2013 12th International Conference on Document Analysis and Recognition, pp. 629-633, Ninth International Conference on Document Analysis and Recognition (ICDAR 2007) Vol 2, 2007

[21] B. Gatos, I. Pratikakis and S.J. Perantonis, "An adaptive binarisation technique for low quality historical documents",IARP Workshop on Document Analysis Systems (DAS2004), Lecture Notes in Computer Science (3163), September 2004, pp. 102-113.

[22] He, X., Yung, N.: Corner detector based on global and local curvature properties. Optical Engineering 47(5) (2008) DOI: http://dx.doi.org/10.1117/1.2931681

[23] Paul Viola and Michael J. Jones. Rapid Object Detection using a Boosted Cascade of Simple Features. IEEE CVPR, 2001.

[24] Rainer Lienhart and Jochen Maydt. An Extended Set of Haar-like Features for Rapid Object Detection. IEEE ICIP 2002, Vol. 1, pp. 900-903, Sep. 2002.