International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 12 (2015) pp. 32711-32721

http://www.ripublication.com

Design of an Economic Voice Enabled Assistive System for the Visually Impaired

D.Radha and Amudha J

Department of Computer Science and Engineering
Amrita Vishwa Vidyapeetham,
Amrita School of Engineering, Bangalore, India

ABSTRACT

Assistive technologies are meant to improve the quality of the life of visually impaired population. However due to factors like income level/economic status of the visually impaired, hand held devices that are not ease of use, ignorance of the existing assistive technologies makes less reachability of these assistive systems for the needy. This paper proposes a system which has been developed to be an economic, simple, Voice Enabled Assistive System (VEAS) for visually impaired to identify and locate the objects in his/her nearby environment. VEAS uses attention theories to easily locate the objects from its background with less computational time. The user interface has been made flexible, by making the human interface system quite comfortable to an ignorant user and assisted with a speech processor for guiding them.

Keyterms: Feature Extraction, Top Down, Bottom up, Visual attention, Fuzzy Inference system, KNN algorithm, Saliency map, Minkowski's metric.

I. Introduction

World Health Organization(WHO) estimated that 285 million people are visually impaired worldwide: 39 million are blind and 246 million have low vision. About 90% of the world's visually impaired live in low-income settings[1]. The WHO Global Action Plan 2014-2019 works for 'Universal Eye Health' to reduce the visual impairment in the world. The gloomy thing is that the recent estimates from the WHO indicate that 90 % of visually impaired people live in the poorest countries of the world. India is homed to be one-fifth of the world's visually impaired people and so, any strategies to battle blindness must consider economic conditions of the people[2]. One side is working on preventing such disabilities. The other side work is in developing assisting aids for visually impaired. But there are issues that often forbid the use of assistive systems such as “lack of information about assistive technology”, "the high cost and severely limited sources of financial assistance for assistive technology", "insufficient numbers of organizations and personnel able to provide instruction in the use of new technology" and "developments such as the graphical user interface which hamper access to new computer and electronic information technology"[3]. The usage of wearable devices and handheld devices were discussed in [4]. Wearable devices may be discomfort in many situations than handheld devices. Regular sensor information and handling certain heavy or tight devices may lead to uneasiness for the users.

So the assistive system for visually impaired should be serving system than burden system. VEAS system concentrates on building a simple, economic and handheld guidance system which can be effectively used at least for the identification of basic objects in their indoor environment.

II. Literature Survey

Many personal guidance system for visually impaired are existing and tested which have their own pros and cons. A system [5] uses navigation system to guide the visually impaired person. It receives information from a Global Positioning System receiver. Global Positioning System (GPS) is used to locate blind travelers in outdoor environments, to guide them along routes to destinations, and to provide them with information about nearby points of interest [5]. The systems use text to express information to the visually impaired by electronic displays or synthesized speech. These kinds of systems also use spatial displays to pass on the information about the environment than synthesizing speech alone. Synthesized speech would be done by a virtual acoustic display and to sense within the auditory space of the user which coincides ideally with the actual locations of the waypoints and off-route landmarks.

Another system has spatial display using a haptic pointer interface (HPI). The HPI is like the functioning of a Talking Signs receiver. It is used along with Talking Signs system of remote infrared signage.

In this system, each "sign" is consisting of an infrared transmitter with a stored "utterance" to identify the location (such as an entrance to a building or a telephone booth) [6]. When hand-held receiver with the user is pointed in the direction of one of the transmitters, an identifying utterance is heard by the user. The direction of the greatest signal strength gives the precise sensing of the direction to the transmitter which is attained by sweeping the hand left and right. The user holds a pointer (a small rectangular stick) to which an electronic compass is attached. Auditory information from the computer is displayed using a shoulder-mounted speaker.

In another version of the HPI, an actual Talking Signs receiver is interfaced to the computer[7]. The electronic compass is mounted on top of the receiver, and the auditory information is displayed using the receiver's speaker. There is a possibility to switch from localizing real Talking Signs to localizing locations in this version. The main aim of these versions is by sensing the orientation of the handheld device, the computer outputs an acoustic signal when the orientation is within some tolerance of the direction to the waypoint or location that is stored in the database. Use of HPI and Talking Signs receiver are similar in experience.

An assisting system named ARGUS [8] focuses onto a satellite based navigation (GNSS/ EDAS) terminal for visually impaired, guiding them along pre-defined tracks, using specifically designed HMI such as acoustic and haptic signals. ARGUS project is based on the provision of a track perception by holophonic technologies which gives an innovative support system for guiding visually impaired. It helps in increasing the safety or enhances the experience in outdoor activities by using web services. Visually impaired people mostly prefer direct percept information about their environment. Several spatial displays were created based on that fact**.**

Drishti system[9] is implemented to work for both familiar and unfamiliar environments. Drishti guide the users to travel independently and safely in indoor and outdoor environment. It uses a precise position measurement system, a wireless connection, a wearable computer and a vocal communication interface for the same. Drishti uses Differential-GPS as its location system, dynamic routing and rerouting ability to provide user with an optimal route in outdoors. There is a possibility for the user to switch the system from an outdoor to indoor environment with a simple voice command. An OEM ultrasound positioning system is used to provide precise indoor location measurements. Experiments show an in-door accuracy of 22 cm. In an indoor environment, the user is prompted with voice to avoid the obstacles and step-step-by walking guidance.

A system that uses the object oriented approach to develop a talking GPS system as navigation system for visually impaired which can be used even for obstacle avoidance[10]. This may not be suited for indoor navigation. Another system using sensors are used for searching tasks within three-dimensional environment. Key press is used to send inquiries to a connected portable computer and they are answered through a text-to speech engine[11].

For any kind of assisting systems available, visually challenged people have their own difficulties in using them as they are much lacking in simplicity. Most of the systems are not much used because of their handling nature, price and complexity. So there is a need for a simple and economic system for basic needs of visually challenged people. The proposed system, Voice Enabled Guidance System (VEAS) overcome the intricacy in the systems discussed above and guide them to identify the position of the object placed, color of the object and then the object itself with voice data**.**

III. Methodology

VEAS assist the visually impaired in finding the objects in front of them. It is mainly used in indoor because training would have done on known objects. It works as follows: A blind person at home can use their mobile for taking snapshots or a video can be taken in which keyframes are taken and sent to the system. VEAS is considered as an application in the mobile. The image is divided into nine quadrants and named as top left, top mid, top right, mid left, center, mid right, bottom left, bottom mid and bottom right. The system identifies the objects in the image using bottom up and top down saliency. Identified object’s color, position(quadrant) and object name are converted into audio output and delivered to the blind person which will help the person to use the required object. The framework of VEAS is shown in Fig 1.

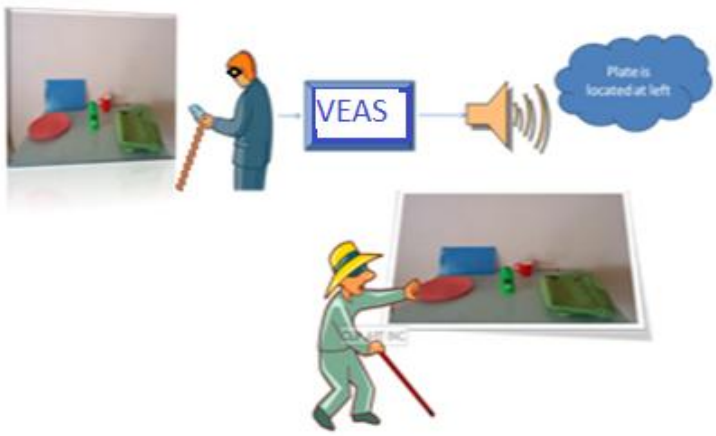

Fig. 1 Framework of VEAS

VEAS works in three phases, Pre-processing phase, Learning phase and Testing phase.

A. Pre-Processing phase:

The pre-processing phase of VEAS is shown in Fig 2. The feature extraction method is applied over the target image to get the required feature maps. This feature vector extracted for the target image is stored dynamically into an excel sheet. As new objects are added the new feature vectors can be appended to the existing feature dataset. These features are utilized later in the testing phase for decision making.

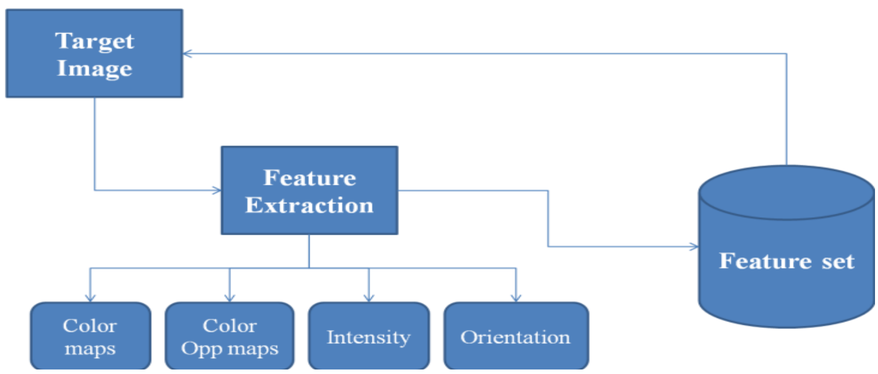

Fig 2. Pre-processing step for the proposed system

There are several feature maps that can be generated from an RGB image. These are the features which help in categorizing an object into different classes. The three main properties are Color, Intensity and Orientation. There are 13 feature maps generated based on the three properties using the conventional method. The first Color property includes two sets of maps. One is for the basic colour maps(R, G, B, and Y) and the other set is color opponency maps (RG, BY, GR and YB). The intensity feature map is obtained by normalizing the intensity values of all the pixels of the image. The third property Orientation involves four orientation feature maps in different directions. The directions are identified by four directional Gabor filters like 0, 40, 90 and 135. The pyramids generated for each orientation gives the corresponding edges on different scales. Also 3 conspicuity maps are generated by summing up the colour feature map, intensity feature map and the orientation feature maps.

B. Learning Phase

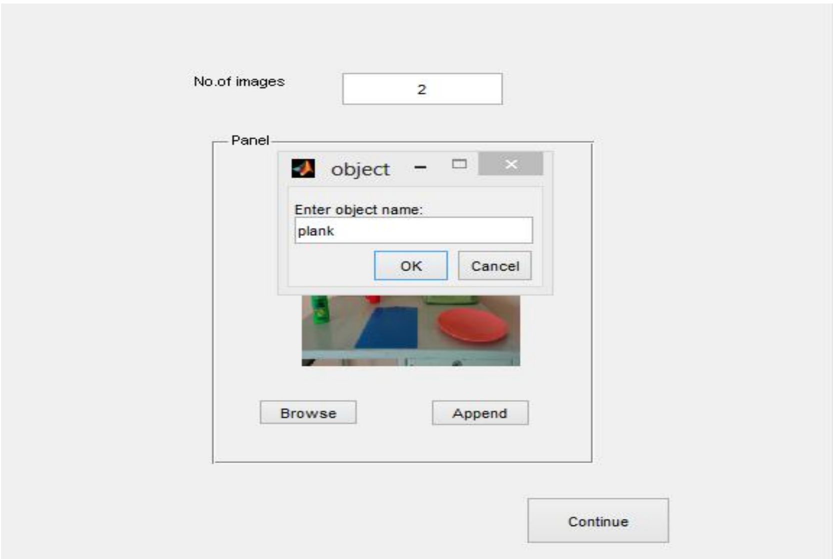

Fig 3. Learning phase of VEAS

Learning phase is depicted in Fig 3 which is used to add new objects and store their features. Feature extraction is done for the Region of Interest in the input image and the weights are calculated for each of the maps.
The weight w_i of map X_i is computed by (1).

$$w_i = m_{i,(MSR)}/m_{i,(image-MSR)} \quad (1)$$

i ϵ [1, … 13]

where, MSR is the most salient region of the map. The feature maps are sorted based on the weights. The sorted array is given as input to the Fuzzy inference system which outputs the excitation and inhibition map based on the 9 rules of the Fuzzy system. The fuzzy system uses the combination of natural language and imprecise data as input to achieve the results. The concept of weight calculation and inhibition of insignificant regions used in VOCUS [10] for computing the top down map is followed. Fuzzy logic is applied on a pixel by pixel basis to highlight the significant pixels and to inhibit the unwanted pixels. These are mapped on to excitation and inhibition maps. The top down saliency map is obtained from the excitation and inhibition map.

The excitation map E is the weighted sum of all feature and conspicuity maps X_i that are important for the target, namely the features with weights greater than 1.

$$E = \sum_{i:w_i>1}(w_i * X_i) \qquad (2)$$

The inhibition map I collects the maps that are not present in the target region, namely the features with weights smaller than 1.

$$I = \sum_{i:w_i>1}((\frac{1}{w_i}) * X_i) \qquad (3)$$

The bottom up section includes the color feature maps, intensity feature map and orientation feature maps. These feature maps are combined to form 3 conspicuity maps which are in turn summed up to form the bottom up saliency map. Finally, global saliency map is computed combining the bottom up and top down saliency maps.

C. Testing phase

The global saliency map is generated for the input image in the testing phase which is depicted in Fig 4.

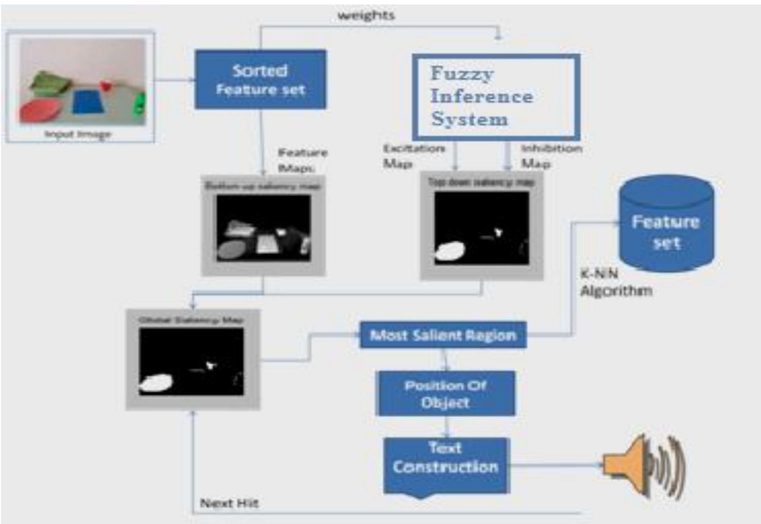

Fig 4. Testing Phase of the proposed system

The Most Salient Region (MSR) is determined. Most salient region is detected using the method of region growing. It is a pixel based image segmentation method where the initial seed points are given as input. For this the pixels having the maximum value in the whole saliency map are selected and are considered as the 'seed points'. Hence the region growing is done with the initial seed points and this region is considered as the first object detected by the system. For the next object (i.e. Next Hit) to be identified, the region grown is masked and the next salient region is calculated. To know which object is identified through the region growing procedure the feature matching is done and the most nearest neighbor is the assumed to be the object. **K-Nearest Neighbour(**KNN) algorithm is applied to identify the object.

KNN is a non parametric lazy learning algorithm. It is a supervised learning method used for classification. This is an example of instance based learning where the unclassified object is compared to the training dataset stored and classified into the most similar class. KNN assumes that data is in a feature space with a notion of distance. The distance is calculated using Minkowski Metric [13] for KNN calculation as shown in (4).

$$d_{st} = \sqrt[p]{\sum_{j=1}^{n} |x_{sj} - y_{tj}|^p} \quad (4)$$

For p=1, the Minkowski metric gives the city block metric also called as Manhattan distance, for special case of p=2, the Minkowski metric gives the Euclidean distance, and for p=∞, the Minkowski metric gives the Chebychev distance. According to the distance obtained, the obtained object is assigned to kth nearest class. The smaller value of k in the algorithm leads to overfitting. If k is not too small then it will tend to smooth out the learning behaviour. If the value of k is too large, the interesting behaviour might be overlooked. So, the balanced value of k will give efficient results.This output is now constructed as text and is given as an input to the text to speech converter to give an audio output to the user.

IV. Results And Analysis

The target objects used in the experiment: A plate, a Cup, a Shampoo, a Bag and a Plank(Fig. 5). VEAS is tested on a dataset of 37 images. The images are considered to be in RGB color space.

Fig 5. The target objects used in the experiment

Experiment is done in two ways: One with images of different color and shape objects. The other one with images of different shape and some are similar color objects.

Results of images with different color and shape objects: The objects considered here are cup, shampoo and plank. Details of success and failure in object detection is shown in the following Table 1.:

Table 1. Success and Failure in object identification with distinct color objects.

Objects to be identified	No. of images with the object to be identified	Successfully identified	Failed to identify	Success rate
Red Cup	15	15	0	100 %
Green Shampoo	12	10	2	92 %
Blue Plank	15	14	1	97 %

Following are the results of identifying the objects which are different in shape and similar in color feature:

The steps involved in identifying the red plate is shown in the Fig. 4. The hit level of identification depends on the saliency of the object in a particular image. The test image can have multiple trained objects and all of them are identified in different levels.

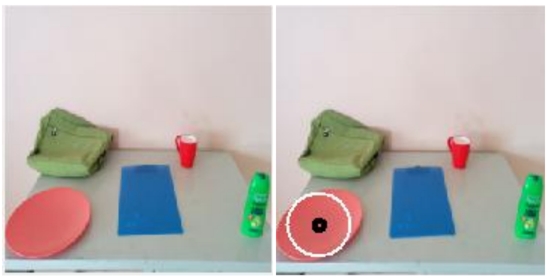

Fig 6. Successful case i/p and o/p of VEAS

Fig 7. Failure case i/p and o/p of VEAS

The success case and failure case are shown in Fig 6. and Fig 7. respectively.

The errors occur as the decision is based on the comparison of the pixel properties with the feature dataset.

Table 2. Success and Failure in object identification with similar color objects.

Objects to be identified	No. of images with the object to be identified	Successfully identified	Failed to identify	Success rate
Cup	33	33	0	100 %
Shampoo	26	24	2	92 %
Plank	34	31	3	91 %
Plate	25	13	12	52%
Bag	28	8	20	28 %

The proposed method identifies all the trained objects in the images that are tested. The failure cases exist because of the similarity between some objects features like the color. This puzzlement can slightly be reduced by considering the aspect ratio of the objects to distinguish the objects. This factor worked well for some objects where the intensity distribution of the objects was even. The system is 96% accurate on an average for the objects without similarity and 73% accurate on an average for objects with similarity are evaluated.

To summarize, the results obtained here are very encouraging and provides an efficient way of detecting and identifying those objects with less similarity. But, by adding the adaptive parameters for inter-and intra-cluster formation for object clustering in VEAS, the above conflicts can be avoided.

V. Conclusion and Future Enhancements

The human eye is the organ which gives us the sense of sight, allowing us to observe and learn more about the surrounding world than we do with any of the other four senses. People without eye sight have more difficulty in their daily routine than most of the disabilities. Many assisting devices are not much of use for them as they are lacking in user-friendliness, lack of technology awareness, high price, discomfort in using etc.. Some may require more attention in sensing the information. But the proposed VEAS system may be very simple to use, cost effective and uses handheld device. VEAS is considered as a mobile application for the visually impaired. The model uses the feature set as the categorizing factor which would increase the quality of search for the objects. Results have demonstrated that the proposed system is able to identify almost all the objects successfully. The model reduces the complexity of analysing a scene for multi-target search. The system is tested for indoor scenarios with images of stationary objects captured with a still camera. The efficiency of the system can further be increased by increasing the number of training samples and slightly increasing the k value of clustering.

The objects considered are distinct in shape, position and some of them have a common color.The system identifies multiple objects in the scenario, accurately and efficiently and gives an audio output to the user about the object and its location. The user interface of VEAS has been made flexible, by making the human interface system quite comfortable to an ignorant user and assisted with a speech processor for guiding them. It can be deployed in a handheld device and can be trained in simpler and easier way.

The VEAS can be extended to detect the objects with lower false positive rate. The adaptive parameters for intra-and inter-cluster formation avoids the conflicts in identifying the objects. The distance of the object from the user can also be calculated to provide better information to the visually impaired. The system works well with the primary coloured objects and can be extended to secondary colours.

Acknowledgments

This research was supported by Divya K V, I Sharanya, B Padma Sravani. We thank them for their support in the implementation.

References

1. http://www.who.int/blindness/AP2014_19_English.pdf
2. Khanna R, Raman U, Rao GN, 'Blindness and poverty in India: the way forward', *Clinical and experimental optometry, Journal of the Australian Optometrical Association, Nov 2007.*
3. Derrick W. Smith, Stacy M. Kelly, and Gaylen Kapperman,'Assistive Technology for Students with Visual Impairments', *A Position Paper of the Division on Visual Impairments, Council of Exceptional Children 2011.*
4. R. Velázquez,' Wearable Assistive Devices for the Blind', *Wearable and Autonomous Biomedical Devices and Systems for Smart Environment: Issues and Characterization, LNEE 75, Springer, pp 331-349, 2010.*
5. Loomis JM, Golledge RG, Klatzky RL, Speigle JM, Tietz J.' Personal guidance system for the visually impaired', *Proceedings of the First Annual International ACM/SIGCAPH Conference on Assistive Technologies (Assets '94), Marina Del Rey, CA, October 31–November 1,New York: 1994.*
6. Loomis, J. M., Golledge, R. G., & Klatzky, R. L.'GPS-based navigation systems for the visually impaired', *In W. Barfield & T. Caudell, (Eds.), Fundamentals of Wearable Computers and Augmented Reality, (pp. 429-446). Mahwah, NJ: Lawrence Erlbaum Associates. (2001).*
7. Jack M. Loomis, James R. Marston, Reginald G. Golledge, Roberta L. Klatzky, 'Personal Guidance System for People with Visual Impairment: A Comparison of Spatial Displays for Route Guidance', *Published in Journal of Visual Impairment and Blindness. 2005;*
8. Oihana Otaegui, Estíbaliz Loyo, Eduardo Carrasco, Claudia Fösleitner, John Spiller, Manfred Schrenk, Daniela Patti, Rafael Olmedo, Markus Dubielzig,'ARGUS: Assisting Personal Guidance System for People with Visual Impairment', *48th ISOCARP Congress 2012.*

9. Ran, L. Helal, S., Moore, S ‘Drishti: an integrated indoor/outdoor blind navigation system and service’, *Pervasive Computing and Communications, 2004, Proceedings of the Second IEEE Annual Conference, March 2004.*
10. Bruce Moulton, Gauri Pradhan, Zenon Chaczko ‘Voice Operated Guidance Systems for Vision Impaired People: Investigating a User-Centered Open Source Model’, *International Journal of Digital Content Technology and its Applications, Volume 3, Number 4, December 2009.*
11. Andreas Hub, Joachim Diepstraten, Thomas Ertl,‘ Design and Development of an Indoor Navigation andObject Identification System for the Blind’, *ASSETS'04,ACM, October 18–20, 2004, Atlanta, Georgia, USA.*
12. S. Frintrop, ‘VOCUS: A Visual Attention System for Object Detection and Goal-Directed Search’. *Heideberg, Germany: Springer-Verlag, 2006, vol.3899, Lecture Notes in Artif. Intell. (LNAI).*
13. Leila M. G. Fonseca, Laércio M. Namikawa and Emiliano F. Castejon, ‘Digital Image Processing in Remote Sensing’. *IEEE 2009.*